FACULDADE DE CIÊNCIAS SOCIAIS E HUMANAS
UNIVERSIDADE NOVA DE LISBOA

How to tell stories using visualization: strategies towards Narrative Visualization

Ana Raquel de Ponte Figueiras

Doctorate Dissertation in

Digital Media

June, 2016

Thesis submitted in accordance with the requirements for the degree of Doctor of Philosophy in Digital Media, under scientific supervision of Prof. António Granado.

Tese apresentada para cumprimento dos requisitos necessários à obtenção do grau de Doutor em Media Digitais, realizada sob a orientação científica de Prof. António Granado.

Financial support from:

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

UT Austin | Portugal
International Collaboratory for Emerging Technologies, CoLab

Declaro que esta tese é o resultado da minha investigação pessoal e independente. O seu conteúdo é original e todas as fontes consultadas estão devidamente mencionadas no texto, nas notas e na bibliografia.

O candidato,

____

Lisboa, _____ de _______________ de __________

Declaro que esta tese se encontra em condições de ser apreciada pelo júri a designar.

O orientador,

____

Lisboa, _____ de _______________ de __________

How to tell stories using visualization:
strategies towards Narrative Visualization

This document was created unsing the (pdf)LaTeX processor, based in the "unlthesis" template[1], developed at the Dep. Informática of FCT-NOVA [2]. [1] https://github.com/joaomlourenco/unlthesis [2] http://www.di.fct.unl.pt

To my parents, family, and friends.

Acknowledgements

Completing a thesis is a great challenge and thanking all those who contributed to it is an even greater one. Many people influenced me, shaped my thinking, and my work. Some of them are well aware of it and some are not, either way I want to thank all of them, including those not mentioned here by name.

First, I would like to thank Professor António Granado from the Department of Communication Sciences for being my thesis supervisor: thank you for reviewing my work and always pushing me to do better. Second, I would like to thank Fundação para a Ciência e a Tecnologia and the UT Austin|Portugal Program for funding my research and giving me the opportunity to spend a semester at UT Austin (SFRH / BD / 51739 / 2011). I also have to thank Professor Rui Cádima for, in 2010, encouraging me in my plans to enroll in this PhD program; Professor Sharon Strover, Professor Nuno Correia, Professor Peter Hall, Professor Riley Triggs, and Professor Carol King for the feedback on my research; my colleagues; all the people that reviewed my work; and the focus group participants for their helpful insights and for accommodating this study in their deadline-driven schedules, without them this research would not be possible. I also need to thank Professor Bruno Giesteira, from Universidade do Porto, for being the jury of my research project proposal and for his valuable comments on my initial work.

I also have to thank my mother for encouraging me to continue my studies, always giving me the support I needed, never doubting I was going to succeed, and being a role model of strength and character. To my father for teaching me the importance of hard work and my grandmother for being a great role model of resilience. Their support all these years was essential. I would also like to thank my extended family and my close friends (the ones that have known me forever and the ones that I made later in life): thank you all for everything.

Lastly, I would like to thank my partner Flávio Martins for always challenging me to learn more, for his support, for his amazing proof-reading skills, and for always encouraging me to go beyond my capabilities. He is responsible for many of the skills I acquired and was a very patient teacher.

The greatest value of a picture is when it forces us to notice what we never expected to see.

John Tukey

Abstract

The benefits of storytelling are long-known and its potential to simplify concepts, convey cultural values and experiences, create emotional connection, and capacity to help retain information has been explored in different areas, such as journalism, education, marketing, and others. Narratives not only have been the main way people make sense of the world, but also the easiest way humans found out to share complex information.

Due to its potential narratives have also recently been approached in the area of Information and Knowledge Visualization, several times being referred to as *Narrative Visualization*. This matter is also particularly important for news media, one of the areas that has been pushing the research on *Narrative Visualization*. The necessity to incorporate storytelling in visualizations arises from the need to share complex data in a way that is engaging. Nowadays we also have the challenge of the high amount of information available, which can be hard to cope with. Advances in technology have enabled us to go beyond the traditional forms of storytelling and representing data, giving us more attractive and sophisticated means to tell stories.

In this dissertation, I explore the benefits of infusing visualizations with narratives. In addition I also present ways of combining storytelling with visualization and efficient methods to represent and make sense of data in a way that allows people to relate with the information. This research is closely related to journalism, but these techniques can be applied to completely different areas (education, scientific visualization, etc.). To further explore this topic a mixed-method evaluation that consists of a typology, several case studies and a focus group study was chosen, as well as design studies and techniques review. This dissertation is intended to contribute to the evolving understanding of the field of narrative visualization.

Keywords: Information visualization; narrative visualization; storytelling

Resumo

Os benefícios da utilização das narrativas são desde há muito conhecidos e o seu potencial para simplificar conceitos, transmitir valores culturais e experiências, criar ligações emocionais e capacidade para ajudar a reter a informação tem sido explorado em diferentes áreas. As narrativas não são só a principal forma como as pessoas obtêm o sentido do mundo, mas também a forma mais fácil que encontrámos para partilhar informações complexas.

Devido ao seu potencial, as narrativas foram recentemente abordadas na área da Visualização de Informação e do Conhecimento, muitas vezes apelidada de *Visualização Narrativa*. Esta questão é particularmente importante para os media, uma das áreas que tem impulsionado a investigação em *Visualização Narrativa*. A necessidade de incorporar histórias nas visualizações surge da necessidade de partilhar dados complexos de um modo envolvente. Hoje em dia somos confrontados com a elevada quantidade de informação disponível, um desafio difícil de resolver. Os avanços da tecnologia permitiram ir além das formas tradicionais de narrativa e de representação de dados, dando-nos meios mais atraentes e sofisticados para contar histórias.

Nesta tese, exploro os benefícios da introdução de narrativas nas visualizações. Adicionalmente também exploro formas de combinar histórias com a visualizações e métodos eficientes para representar e dar sentido aos dados de uma forma que permite que as pessoas se relacionem com a informação. Esta investigação está bastante próxima da área do jornalismo, no entanto estas técnicas podem ser aplicadas em diferente áreas (educação, visualização científica, etc.). Para explorar ainda mais este tema foi adotada um avaliação que utiliza diferentes metodologias como a tipologia, vários casos de estudo, um estudo com grupos de foco, e ainda estudos de design e análise de técnicas.

Palavras-chave: Visualização de informação; visualização narrativa; narrativa

Contents

List of Figures xxiii

List of Tables xxvii

Glossary xxix

Acronyms xlv

1 Introduction 1
- 1.1 Motivation 3
- 1.2 Approach 7
- 1.3 Research questions 8
- 1.4 Methodology 9
- 1.5 Overview 10
- 1.6 Contributions 11

2 Visualization analysis 13
- 2.1 Visualization history and its most influential examples 15
  - 2.1.1 War Mortality – Florence Nightingale 16
  - 2.1.2 Figurative Map of the successive losses in men of the French Army in the Russian campaign 1812-1813 - Charles Minard 17
  - 2.1.3 London Cholera Map – John Snow 19
  - 2.1.4 Gapminder – Rosling 20
  - 2.1.5 Visualization on the Web 22
  - 2.1.6 Wikileaks and open government data 24
- 2.2 The present and future of visualization 25
  - 2.2.1 Principles borrowed from other areas 27
  - 2.2.2 The importance of the insights and knowledge 34
  - 2.2.3 The importance of function, user engagement, and target audience 41
  - 2.2.4 Functionality over aesthetics 43

2.3 Chapter Summary 48

3 The many names of visualization 51
3.1 Defining Data, Information, and Knowledge 53
3.2 Visualization of Data, Information, Knowledge, and everything in between . 54
3.2.1 Scientific visualization 54
3.2.2 Visual analytics 56
3.2.3 Data visualization 56
3.2.4 Knowledge visualization 57
3.2.5 Casual, social, collaborative, and artistic visualization 58
3.3 Chapter Summary 64

4 The role of interactivity 67
4.1 Interactivity versus interaction 68
4.2 Previous interaction techniques taxonomies 69
4.3 A new Interaction Techniques Taxonomy 70
4.3.1 Filtering and selecting 71
4.3.2 Abstract/Elaborate: Zoom, details-on-demand, and linking 71
4.3.3 Overview and explore 74
4.3.4 Connect/relate 75
4.3.5 History and extraction of features 75
4.3.6 Reconfigure and encode 76
4.3.7 Complex forms of interaction: participation/collaboration and gamification 78
4.4 Animation 80
4.4.1 Arguments in favor and against of the use of animation 84
4.5 Chapter Summary 86

5 A new Visualization Typology 89
5.1 Existing Typologies 90
5.2 The proposed Typology 91
5.2.1 Sequential Graphic 93
5.2.2 Slide Show 93
5.2.3 Chart/Diagram 94
5.2.4 Map 94
5.2.5 Tag Clouds 95
5.2.6 Model 95
5.2.7 Drawing 96
5.2.8 Video/Animation 96
5.2.9 Photograph 97
5.2.10 Poster 97
5.2.11 Game 98
5.3 ReThinking Visualization 98
5.4 Chapter Summary 101

6 Typology Case Studies **105**
6.1 England Riots . . . 105
6.2 The World of Seven Billion . . . 107
6.3 Death penalty statistics, country by country . . . 108
6.4 Home and Away: Iraq and Afghanistan War Casualties . . . 110
6.5 What Does China Censor Online? . . . 112
6.6 Ground Zero Now . . . 113
6.7 How Many Households Are Like Yours? . . . 115
6.8 A Week of Check-ins on the Path to One Billion . . . 117
6.9 100 Years of World Cuisine . . . 118
6.10 How Local News Is Going Mobile Infographic . . . 119
6.11 SPENT . . . 120
6.12 Chapter Summary . . . 121

7 Focus group study **123**
7.1 Procedures . . . 123
7.2 Results . . . 126
7.2.1 The most discussed . . . 126
7.2.2 Links as context . . . 128
7.2.3 Non-playable . . . 131
7.3 Chapter Summary . . . 131

8 Using visualization to tell stories **133**
8.1 Narratives and New Forms of Storytelling . . . 136
8.1.1 Narrative visualization . . . 138
8.1.2 History of storytelling in information/knowledge visualization . . . 138
8.2 Chapter Summary . . . 141

9 Narrative strategies **143**
9.1 Interactivity and its relation with context . . . 144
9.2 Empathy . . . 146
9.3 The Relation Between Time and Narrative . . . 146
9.4 Gamification as a Way to Have Storytelling in Visualizations . . . 147
9.5 Case Study . . . 151
9.5.1 *How Many Households are Like Yours?* enhanced with empathy . . . 152
9.5.2 *What Does China Censor Online?* enhanced with context . . . 153
9.5.3 *Death Penalty Statistics, Country by Country* enhanced with time . . . 154
9.6 Chapter Summary . . . 155

10 Techniques for visualization on the web **157**
10.1 Tools for information visualization . . . 159
10.1.1 Tableau . . . 160
10.1.2 D3.js . . . 161
10.1.3 Ellipsis . . . 163

10.1.4 Other tools: Many Eyes, Google Charts, and Gephi 164
10.2 Choosing the right visualization for a given function 170
10.2.1 A visualization selection matrix 174
10.3 Chapter Summary 177

11 Conclusions **179**
11.1 Thesis contributions 180
11.1.1 Research objectives 181
11.1.2 Summary of major contributions 185
11.1.3 Summary of minor contributions 186
11.2 Future work 186
11.2.1 Work on narrative visualization evaluation 187
11.3 Closing remarks 190

Bibliography **193**

I Publications **209**

II Typology initial tables **243**

III Focus group rating sheet **249**

IV Focus group semistructured questions **251**

List of Figures

2.1 Playfair's time-series chart of *Exports and Imports to and from Denmark and Norway from 1700 to 1780*, published in his Commercial and Political Atlas 15
2.2 Florence Nightingale's coxcomb diagram on mortality in the British army 16
2.3 Minard's *Carte figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813* 17
2.4 John Snow's map of the clusters of cholera cases in the epidemic of 1854 19
2.5 Hans Rosling's Trendalyzer, the information visualization software for animation of statistics 21
2.6 Gestalt laws referenced by Ware (2004) and Nesbitt and Friedrich (2002) 27
2.7 Example of proximity in a bar chart 28
2.8 Example of similarity and connectedness in a flowchart 29
2.9 Example of continuity in a line chart 29
2.10 Example of symmetry in a pyramid bar chart 30
2.11 Example of closure in a bar chart 31
2.12 Example of common fate in *Trendalyzer* 32
2.13 Example of simplicity in a bar chart 33
2.14 Example of familiarity in a bar chart 34
2.15 Example of Semantic Zoom 36
2.16 Evaluation levels for Visual Analytics (Figure 6.1 (Cook and Thomas, 2005)) . . . 38
2.17 Scale (Figure 1 (Kosara, 2007)) 44
2.18 Model of information aesthetics (Figure 4 and 5 (Lau and Vande Moere, 2007)) . 45

3.1 The Data Information Knowledge Wisdom (DIKW) hierarchy and definitions of these terms in perceptual and cognitive space by Zeleny (1987) and by Ackoff (1989). 52
3.2 Examples of scientific visualization 55
3.3 Examples of casual visualization 59
3.4 Four different collaborative visualization scenarios. 61
3.5 Examples of artistic information visualization 63

3.6 Taxonomies comparison . . . 64

4.1 D3 zoomable treemap . . . 72
4.2 D3 collapsible tree . . . 73
4.3 *The view from the Shard: a new and expanded panorama of London* . . . 74
4.4 Examples of the use of the connect interaction technique . . . 75
4.5 Encode interactive options in *As the Oscars age, so do the nominees* . . . 76
4.6 Encode interactive options in *Bloomberg Billionaires* . . . 77
4.7 Examples of gamification in news media . . . 80
4.8 *Bloomberg Millionaires* map view . . . 81
4.9 Taxonomies comparison . . . 87

5.1 Notes about categories and the visual and interactive elements . . . 92
5.2 Examples of a Sequential Graphic visualization and a Slide Show visualization . . . 93
5.3 Examples of a Chart/Diagram visualization and a Map visualization . . . 94
5.4 Examples of a Tag Cloud visualization and a Model visualization . . . 95
5.5 Examples of a Drawing visualization and a Video/Animation visualization . . . 96
5.6 Examples of a Photograph visualization and a Poster visualization . . . 97
5.7 Examples of Game visualizations . . . 98
5.8 Filtering on *ReThinking Visualization* homepage . . . 99
5.9 *ReThinking Visualization* visualization page . . . 100
5.10 Typologies comparison . . . 102

6.1 *England Riots* timeline by *The Guardian* . . . 106
6.2 *The World of Seven Billion* by *National Geographic* . . . 107
6.3 *Death penalty statistics, country by country* by *The Guardian* . . . 108
6.4 *Home and Away* by *CNN* . . . 110
6.5 *What Does China Censor Online?* by David McCandless . . . 112
6.6 *Ground Zero Now* by *The New York Times* . . . 114
6.7 *How Many Households Are Like Yours?* by *The New York Times* . . . 115
6.8 *A Week of Check-ins on the Path to One Billion* by *Foursquare* . . . 117
6.9 *100 Years of World Cuisine* by *100 Years of World Cuisine* . . . 118
6.10 *How Local News Is Going Mobile* infographic by *Column Five* . . . 119
6.11 *SPENT* by *McKinney* . . . 120

7.1 Characteristics of the visualizations presented to the participants of the focus groups 125
7.2 Visualizations used in the focus group study and the elements that compose them radial matrix . . . 125
7.3 Ranking of the average ratings for each visualization in terms of comprehension, likability, and navigation . . . 126
7.4 Scores given by the participants to each visualization in terms of comprehension, likability, and navigation . . . 127
7.5 *Death Penalty in the US Mapped* by *The Guardian* . . . 127
7.6 *How Much CO2?* by David McCandless . . . 128

7.7 *Faces of the dead* by *The New York Times* . . . 129
7.8 *British troops killed in Afghanistan* by *The Guardian* . . . 130
7.9 *Evolution of the Web* by *Google* . . . 130

8.1 Examples of early storytelling . . . 134
8.2 *Snow Fall* by *The New York Times* . . . 137

9.1 Two systems for narrative visualization annotation . . . 145
9.2 Two screens of *Gauging Your Distraction* by *The New York Times* . . . 148
9.3 Two systems for narrative visualization annotation . . . 150
9.4 *How many households are like yours?* enhanced with storytelling . . . 152
9.5 *What does China censor online?* enhanced with storytelling . . . 153
9.6 *Death penalty statistics, country by country* enhanced with storytelling . . . 154

10.1 Storytelling process (Figure 2 (Lee et al., 2015)) . . . 158
10.2 Visualization tools (D3, Google Charts, Ellipsis, Tableau, Gephi, and Many Eyes) 160
10.3 Visualization with Tableau . . . 161
10.4 Visualization with D3 . . . 162
10.5 Ellipsis . . . 164
10.6 Many Eyes . . . 165
10.7 Gephi . . . 166
10.8 Google Charts . . . 167
10.9 Chart selector guide (Figure 7.1 (Abela, 2008, p. 99)) . . . 171
10.10Jason Davies's D3 example of parallel sets . . . 173

11.1 SUM Model (Figure 1 (Sauro and Kindlund, 2005b)) . . . 189

List of Tables

1.1 Table of contents 10

2.1 Trends in the usage of client-side programming languages for websites 22

4.1 Interaction techniques taxonomies 70
4.2 Proposed taxonomy 70

5.1 Existing visualization typologies 90

6.1 *England Riots* genre, elements, and reading/viewing order 106
6.2 *The World of Seven Billion* genre, elements, and reading/viewing order 108
6.3 *Death penalty statistics, country by country* genre, elements, and reading/viewing order 109
6.4 *Home and Away* genre, elements, and reading/viewing order 111
6.5 *What Does China Censor Online?* genre, elements, and reading/viewing order . . 113
6.6 *Ground Zero Now* genre, elements, and reading/viewing order 114
6.7 *How Many Households Are Like Yours?* genre, elements, and reading/viewing order 116
6.8 *A Week of Check-ins on the Path to One Billion* genre, elements, and reading/viewing order 117
6.9 *100 Years of World Cuisine* genre, elements, and reading/viewing order 118
6.10 *How Local News Is Going Mobile* genre, elements, and reading/viewing order . . . 120
6.11 *SPENT* genre, elements, and reading/viewing order 121
6.12 Typology case studies and respective typology categories 122

10.1 Visualization selection suggestion matrix (at least one quantitative variable) 175
10.2 Visualization selection suggestion matrix (only categorical data) 177

11.1 Published papers on the topic of information visualization and storytelling 180

III.1 Focus group rating sheet 250

Glossary

accompanying article An article that was written to support a visualization or that has the visualization as supporting information. It is usually presented in a separate page and can be accessed through a link.

Adobe Flash A software platform used for creating animated content, vector graphics, web browser games, and other web applications.

aesthetics A branch of philosophy that gives a set of principles concerned with the nature of art, beauty, and taste, and with the creation and appreciation of beauty.

ambient visualization A field closely related to information aesthetics that aims to communicate information in the periphery of attention, resorting to engaging displays. These are "information visualization applications that do not reside on the screen of a desktop computer, but in the environment or periphery of the user" (Skog et al., 2003, p. 234).

animation Animation is the change of a visual representation over time through the rapid display of sequential static images that minimally differ from each other, resulting in an illusion of movement or shape change.

annotations Annotations are bits of textual information that are presented as a support for the information presented in the visualization. Annotations are a promising way to complement articles since they have the capacity to add context that otherwise would be very difficult to provide.

Apple Keynote A presentation software that is part of Apple's productivity suite, iWork.

application In computing, an application (app) is a program or piece of software designed to fulfill a particular purpose (a group of coordinated functions, tasks, or activities). Applications are designed to run on computers, on mobile devices such as smartphones and tablet computers (known as mobile apps), or in a web browser (known as web apps).

arc diagram A style of string visualization, introduced by Wattenberg (2002), in which nodes are places along a line and arcs connect the nodes in one of the two halfplanes. The thickness of the arcs can be used to represent frequencies.

area chart A chart that is based on the line chart. The area between axis and line are commonly emphasized with colors, textures and hatchings. Commonly one compares with an area chart two or more quantities.

artistic information visualization See artistic visualization.

artistic visualization "Artistic visualizations are visualizations of data done by artists with the intent of making art" (Viégas and Wattenberg, 2007, p. 183). It is also known as information or data art and it is an area on the edges of information visualization more concerned with aesthetics and the sublime. An example are the weather data LEGO pieces of sculptor Nathalie Miebach (available at `http://nathaliemiebach.com/gulf.html`).

audio narration An audible content usually in the form of a story or a description of events that acts as support information for the visualization.

bar graph See bar chart.

bar chart A chart or graph that uses narrow columns of different heights to show and compare different amounts.

Bejeweled A series of tile-matching online skill-based games created by PopCap Games. It was first developed in 2001 as a web-based Flash game named Diamond Mine.

big data Extremely large data sets that may be analyzed computationally to reveal patterns, trends, and associations, especially relating to human behaviour and interactions.

Bokeh A Python interactive visualization library that intends to be the Python alternative to D3. More in Subsection 10.1.4.1.

box plot Also known as a box and whisker diagram, this is a way to display graphically groups of numerical data through their quartiles. The vertical lines (or whiskers) indicate variability outside the upper and lower quartiles.

bubble map A map where the quantity of a variable for a specific geographic location is represented by the size of a bubble.

bubble chart A variation of the scatter plot that uses Cartesian coordinates to display values for two variables and in which a third variable is represented by the size of the bubble.

calendar A chart that presents the register of days according to a particular system.

canvas element An element consisting of a drawable region with height and width attributes available in HTML5, which allows for dynamic, scriptable rendering of 2D shapes and bitmap images, and can be accessed by JavaScript through a set of drawing functions.

caption A caption is a small bit of text used to shortly describe, in a couple of words, an element of the visualization. It can be a name, a keyword, etc.

cartogram A map in which some thematic variable is mapped on a geographic map, but the geometry or space of the map is distorted in order to convey the information of a quantitative variable. For instance, when mapping world population data by country, more populated countries such as China would appear larger than less populated countries.

casual information visualization See casual visualization.

casual visualization According to Pousman et al. (2007, p. 1149) it is "the use of computer mediated tools to depict personally meaningful information in visual ways that support everyday users in both everyday work and non-work situations".

chart A graphical representation of data. The term is used for diagrams or graphs, that organize and represent numerical or qualitative data; sheets giving information in tabular form; and maps that contain extra information, such as nautical or aeronautical charts.

Chart/Diagram A genre that only comprises visualizations in which the chart or diagram is the main focus of the visualization. It includes every type of chart/diagram, from the common bar charts to Venn diagrams.

chord diagram A type of diagram used to display relationships between data in a matrix. These relationships are represented by arranging the data as nodes around a circle and connecting the data points that share a relationship with a ribbon. The thickness of the ribbon represents the value assigned to that connection, and color can be used to group data points into categories.

Chrome A freeware web browser developed by Google and available since 2008. It is available for Windows, Mac OS, and Linux operating systems and there is a mobile version for Android devices. It is currently the most popular web browser for desktop computers and smartphones.

circle graph In the mathematical area of graph theory, it is an intersection graph that represents the pattern of intersections of a family of sets on a circle. This term is also often used in relation to several different circular charts such as pie charts and doughnut charts (and their respective exploded versions), sunburst charts, and polar-area charts.

circular bar chart A bar chart, also known as radial bar chart, plotted on a polar coordinate system, rather than on a cartesian one. Although more visually appealing, it is often misleading because the outside bars take up more area than the bars on the inside.

click highlight The interactive action of highlighting or emphasizing a content (text, image, pictogram, etc.) when a computer user moves the pointer to a certain location on a screen and presses a button on a mouse, usually the left button (click), or other pointing device.

click detail The interactive action of showing extra information when a computer user moves the pointer to a certain location on a screen and presses a button on a mouse, usually the left button (click), or other pointing device.

cloropleth A map in which areas are shaded or patterned in proportion to the value of a variable that corresponds to each geographical location.

color matrix It is similar to a matrix (a rectangular arrangement of elements into rows and columns), but instead of each cell having a number, symbol, or expression, it has a color that was previously assigned to a variable.

combo box A combo box is a commonly used graphical user interface widget or control. Traditionally, it is a combination of a drop-down list or list box and a single-line editable textbox, allowing the user to either type a value directly into the control or choose from a list of existing options by scrolling.

concept map A diagram that depicts suggested relationships between concepts. It is used to organize and represent knowledge, linking concepts, usually enclosed in a circle or box. The relationships between concepts are indicated by a line links the concepts and there can also be words associated to the connecting lines, which specify the relationship between the two concepts.

coxcomb diagram See polar-area chart.

cycle graph It is a graph that consists of a number of vertices connected in a closed chain. It is very common in graph theory.

D3 A JavaScript library to create SVG graphics from data. More in Subsection 10.1.2.

data A collection of qualitative or quantitative information that can be used to calculate, report, and/or analyze. The word data has generated considerable controversy on whether it is singular, plural of datum, or an uncountable noun. Nowadays, despite the complaints of traditionalists, who consider it to be the plural of datum, data is considered a mass noun, similar to information, therefore not having a plural and taking a singular verb. The concept of data is defined in more detail in Section 3.1.

data visualization See Subsection 3.2.3.

decision tree This is a type of visualization that supports decision making. Decision trees use a tree-like graph to present scenarios and their possible consequences and are commonly used in research, specifically in decision analysis, to help choosing between several courses of action.

diagram See chart.

dot plot Also known as a dot chart, strip chart or stripplot, this chart consists of data points plotted on a fairly simple scale, typically using filled in circles. The number of dots represented for each variable on the scale represent the value for that variable. Its most common use is to display distribution/frequency.

dot map Also known as a dot distribution map or dot density map, this is a kind of map that uses dots as symbols to show the occurrence of a feature or phenomenon. Dot maps are specially useful to visualize spatial patterns.

doughnut chart A chart similar to a pie chart, with the exception of a blank center and the ability to support multiple statistics at once.

drag objects In computer graphical user interfaces, drag and drop is a pointing device gesture in which the user selects a virtual object by *grabbing* it and dragging it to a different location or onto another virtual object.

Drawing This is a type of visualization that combines information and illustration. In order to be effective and become a visualization and not a mere drawing, it has to combine the illustration with another type of visualization such as a chart or a map.

Ellipsis "A system that combines a domain-specific language for storytelling with a graphical interface for story authoring" Satyanarayan and J. Heer (2014, p. 361). More in Subsection 10.1.3.

exploded view A diagram, picture, or technical drawing of an object, that shows the relationship or order of assembly of various parts.

external link An external link is a link in another page or another website that can be clicked in order to access extra information that supports the visualization.

filtering Filtering is a function that processes a data structure (typically a list) in order to produce a new data structure containing exactly those elements of the original data structure which have in common the characteristic that was chosen in the filter.

Firefox A free and open source web browser developed by the Mozilla Foundation and the Mozilla Corporation since 2003. It is available not only for Windows and Mac OS but also for Linux operating systems. There are also mobile versions available for Android and Firefox OS.

flowchart A type of diagram that represents an algorithm, workflow or process, showing the steps as boxes of various kinds, and their order by connecting them with arrows. Flowcharts are used in analyzing, designing, documenting or managing a process or program in various fields.

Fusion Tables A data management web service provided by Google very similar to Google Spreadsheets. More in Subsection 10.1.4.

Game This genre is the least common. It comprises visualizations that use formal elements of games such as rules, goals, scores, competition, and the notion of winning.

gamification Refers to the application of game mechanics and/or elements of game design in non-game contexts to engage users in the presented situation. Common game elements include scoring, achievements, competition with other users, etc. Gamification has been studied and applied in several domains, such as online marketing to promote products or services.

Gephi A visualization and data analysis software written in Java. More in Subsection 10.1.4.

Gestaltism From the German word *Gestalt*, which means form, this is a theory of mind of the Berlin School of experimental psychology developed in the 1900s that, among other things, deals with the study of perception.

ggplot2 One of the many R packages and is widely known for making it easier to use R for creating statistical graphics while still taking advantage of its power. More in Subsection 10.1.4.1.

Google Scholar A freely accessible web search engine by Google that indexes scholarly literature from several different publications, publishing formats, and disciplines..

Google Charts A library provided by Google that allows users to create charts and embed them in a web page. It contains prebuilt and ready to use charts, ranging from the very basic bar chart to the more complex treemap, which can be customized to the user's needs.

graph A graph is a representation of a set of objects where some pairs of objects are connected by links.

grouped political map In opposition to physical maps, which focus on the geography of the area, these were designed to show governmental boundaries of countries and states. They can also indicate the location of major cities, and they usually include significant bodies of water.

grouped bar chart A bar chart where, for each categorical group, there are two or more bars colored to represent a particular grouping.

heat map A geographical map that uses color to represent quantities. It differs from a cloropleth by not having a well defined path and representing the areas in a free form. It is commonly used to visualize weather phenomena.

heat map matrix A graphical representation of data where the individual values contained in a matrix are represented as colors.

histogram A representation of a frequency distribution by means of rectangles whose widths represent class intervals (continuous variables) and whose areas are proportional to the corresponding frequencies.

hover highlight This is the action of highlighting or emphasizing a content (text, image, pictogram, etc.) on mouseover or mouse hover (raised when the user moves or *hovers* the pointer over a particular area).

hover details Extra bits of information that pop up on mouseover or mouse hover (raised when the user moves or *hovers* the pointer over a particular area).

HTML The abbreviation of HyperText Markup Language. HTML is the standard markup language to create web pages.

hyperlink In computing, it is a highlighted word or picture in a document or Web page providing direct access to a whole document or to a specific element within a document.

illustration A visualization or a depiction, such as a drawing, sketch, or painting. It uses examples in order to make something easier to understand. An illustration can be used to elucidate, decorate or represent scientific images, processes or technical information on how to use something.

information art See artistic visualization.

information visualization See Section 3.2.

input box A text field or text box is a widget (or control) that has the purpose of allowing the user to input text information to be used by the visualization. Text boxes usually display a text cursor (commonly a blinking vertical line), indicating the current region of text being edited.

interactivity In computing, it is the dialog that occurs between the user and a computer. Interactive applications are designed to respond to the actions/commands of the user and not to run automatically without immediate user involvement. Particularly in visualization interactivity refers to the quality of interaction among the components that comprise the visualization.

Internet Explorer A series of web browsers developed by Microsoft, included as part of the Windows operating systems since 1995.

introductory text A short text that explains what will follow in the visualization or states the purpose and goals of the visualization.

JavaScript JavaScript is a high-level, dynamic, object-oriented, lightweight, interpreted programming language. It is one of the world's most popular programming languages and, although JavaScript is a general-purpose programming language, it is most well-known as the scripting language for the Web.

knowledge A collection of facts, information, and skills acquired either by experience, education, or transmission from another who has it, by instruction, or by extracting it from experience. The concept of knowledge is defined in more detail in Section 3.1.

knowledge visualization See Subsection 3.2.4.

library In software, a library is a collection of functions and methods used to add or implement functionality to a software.

Likert scale A psychometric scale developed by Rensis Likert in 1932 to represent people's attitudes towards a topic. It is commonly involved in research employing questionnaires and is usually represented as a five (or seven) point scale. Common examples of likert scales access agreement (from *strongly agree* to *strongly disagree*), frequency (from *very* *frequently* to *never*), likelihood (from *almost always true* to *almost never true*), or importance (from *very important* to *unimportant*).

line graph See line chart.

line chart A graph, also known as line graph, in which points representing values of a variable for suitable values of an independent variable are connected by a broken line.

link to the raw data A link to the raw data allows the reader to see the original data that gave origin to the visualization. This data can be in provided in the same page of the visualization, in another page of the same website or even in another website. This raw data can be provided in a spreadsheet, a JSON, etc.

link to external article A link to an external article allows the reader to directly follow an article in another page or even another website by clicking. This hyperlink can point to a whole document or to a specific element within a document.

logo A graphic mark or emblem commonly used by commercial enterprises, organizations and even individuals to aid and promote instant public recognition.

Machine Learning A subfield of computer science that develops algorithms to provide computers the ability to learn from and make predictions on data.

Many Eyes "A public web site where users may upload data, create interactive visualizations, and carry on discussions" (Viégas, Wattenberg, Ham, et al., 2007, p. 1121). More in Subsection 10.1.4.

map A symbolic depiction highlighting relationships between elements of a space. It can be tangible, mimicking the world as truthfully as it can, or as Minard's flow map of Napoleon's March an intangible map, representing the physical place in a way that accentuates other information about it.

Map A classic type of visualization that can be tangible (it represents where things are placed and tries to mimic as truthfully as it cans the real world) or intangible (it represents not only information about physical places but also about events that occurred on those places).

matplotlib It is the most popular option for visualizing data in Python. More in Subsection 10.1.4.1.

matrix A rectangular array of numbers, symbols, or expressions, that are arranged in rows and columns.

Microsoft Powerpoint A presentation software that is part of the Microsoft Office suite.

Microsoft Excel A spreadsheet software that is part of the Microsoft Office suite.

mind map A diagram used to organize information, beginning with a main concept, usually enclosed in a circle or box, that branches out to more specific or connected topics. These additional concepts are also usually enclosed in circles or boxes.

model A structural design; a miniature representation of something; a pattern of something to be made.

Model This is a more technical visualization and particularly good to show projects of buildings or to describe complex processes.

Mosaic A discontinued early web browser, produced in 1993 by NCSA, which was later renamed Netscape Navigator.

narrative An account of connected events that is transmitted either by being spoken, written, or visually represented. It can be fictional or non-fictional.

narrative visualization See Subsection 8.1.1.

Natural Language Processing A field of computer science that researches strategies for computers to process and make sense of written and spoken words. It is a computer science field, integrated in the domain of artificial intelligence.

navigation button A user interface element that provides the user a simple way to trigger an navigation event. These buttons allow the user to navigate the information back and forward in a certain order.

Netscape Navigator The dominant web browser in terms of usage share in the 1990s. It was discontinued in 2008 because the user base declined in the late 90s, partly because of Internet Explorer web browser's popularity and partly because of the lack of innovation.

network diagram A drawing of a graph or network diagram is a pictorial representation of the vertices and edges of a graph.

non-ribbon chord diagram Similarly to the chord diagram, it is a type of diagram used to display relationships between data in a matrix, with the data arranged as nodes around a circle and with the data points that share a relationship connected with a line (instead of a ribbon of which the thickness represents the value assigned to that connection).

object size See size representing quantity.

object react to mouse movement This interaction happens when the mouse movement provided by the user influences the objects on the visualization. For instance, when the mouse approaches the objects these are repelled like if the mouse was pushing the objects.

open data Data that can be freely used and distributed by anyone without restrictions from copyright, patents or other control mechanisms. Usually the use of this data is subject to the requirement to attribute authorship and permit re-use/redistribution.

Opera A web browser developed by Opera Software first released 1995. It is available for numerous different computer platforms (Windows, Mac OS, and Linux operating systems) and is a popular alternative for mobile phones (working on devices running Android, iOS, Windows Phone/Mobile, Symbian, Maemo, Bada, and BlackBerry).

parallel sets A ways to visualize multivariate categorical data with a layout similar to parallel coordinates, that instead of having individual data points substitutes them by a frequency-based representation.

parallel coordinates A type of visualization for multivariate data, that usually consists of a set of axis parallel to each other and equally spaced (typically vertical). Each line that connects the different axis is a data point that has a corresponding value for each axis.

perception The process by which humans translate sensory information in order to represent and understand the world around them. Perception is influenced by the stimulation of physical or chemical of the sense organs, which are then translated into signals in the nervous system and interpreted according to the individuals education, memories, and expectations. The concept of perception is defined in more detail in Subsection 2.2.1.

photograph An image produced by the action of radiant energy and especially light on a sensitive surface, usually photographic film or an electronic medium such as a CCD or a CMOS chip. It depicts or records visual perception with a similar appearance to the subject photographed.

Photograph This genre comprises the visualizations in which one or more photographs are the main part of the visualization.

pictogram Also called a pictogramme, pictograph, or icon. It is an ideogram that conveys its meaning through its pictorial resemblance to a physical object.

pie chart A chart consisting of a circle that is divided into parts to show the size of the different amounts that are a part of a whole amount.

platform A technology infrastructure that can be customized by outside developers or users, having the possibility to be adapted to different needs that platform's original developers did not previously contemplated.

player controls Player or media controls are user interface elements typically associated with media such as video and sound. These controls are used to enact and change or adjust the process of watching film or listening to audio and commonly consist of buttons such as *play*, *pause*, *stop*, etc.

Plotly A web-based tool that combines both analytics and visualization. More in Subsection 10.1.4.

polar grid A grid plotted on polar coordinates. A well known example of a polar grid is David McCandless' *Colours in culture* available at `http://www.informationisbeautiful.net/visualizations/colours-in-cultures/`.

polar-area chart A chart similar to a usual pie chart, where the sectors are equal angles but differ in how far they stretch from the center of the circle out. These charts, credited to Florence Nightingale, are also referred to as *coxcombs*.

population pyramid Also known an age pyramid or age picture diagram, it is a chart that represents the distribution of various age groups in a population through horizontal bars, usually separated by sex, forming the shape of a pyramid when the population is growing.

Poster These visualizations are generally static mimicking the structure of a vertical poster so common on magazines and marketing campaigns. It conjugates information and graphic elements in order to be appealing and eye-catching and charts and diagrams play an important part of the visualization.

pragmatic visualization According to Kosara (2007, p. 633) it is "what we term the technical application of visualization techniques to analyze data." The comprehension of the data is the main goal of these visualizations.

Processing A programming language and integrated development environment (IDE) built for visual arts. More in Subsection 10.1.4.1.

Processing.js A JavaScript version of the popular *Processing* visual programming language. It not only allows users to create visualizations but also any type of interactive content (including games). It is only supported on browsers that implemented the canvas element.

programming language A formal constructed language used to communicate a set of detailed instructions to a machine.

pyramid It is used to show proportional, interconnected, or hierarchical relationships with the largest component on the bottom and narrowing up.

Python A general-purpose, interpreted, object-oriented, dynamic programming language.

R A programming language and software environment for statistical computing and graphics. More in Subsection 10.1.4.1.

radar chart A line chart or area chart plotted on polar coordinates, on which the y-axis is the radius and the x-axis the angle. Although it is a common method to display multivariate data, it is difficult to interpret.

radial tree A type of tree diagram that expands outwards, radially.

Safari A web browser developed by Apple, based on the WebKit engine, first released in 2003 with the Mac OS operating system. A mobile version has been included in iOS devices since 2007, when the first iOS device, the iPhone, was released.

sankey arc A visualization based on arc diagrams, proposed by Nagel et al. (2012), "which extends the arc diagram technique by laying out the weighted edges of a node adjacent to each other".

sankey diagram A display of flows, in which the width of the arrows shows the flow quantity.

scalability The capability of a computer application or product (hardware or software) to handle the growth of the volume of work. It includes not only the ability to function well, but also to take full advantage of the rescaled situation.

scale Commonly used in physics, geography, and other sciences, a scale is something graduated used as a measure or rule. It has a series of marks at known intervals used to measure distances (as the height of the mercury in a thermometer) or used as a scheme of rank or order (a scale of taxation).

scatter plot Also known as scatterplot or scattergraph, it is a type of mathematical diagram using Cartesian coordinates to display values for two variables for a set of data.

scenario matrix A matrix divided in four quadrants where the beginning and the end of both axis represent two opposed qualitative variables. Concepts are placed in either of the quadrants according to how they would be placed along each axis.

scientific visualization See Subsection 3.2.1.

scroll activated animations This is a relatively new trend in web design. These consist on animations or effects that unfold and slide across the screen triggered by touchscreen, computer mouse motion or a keypress.

scrollbar An object in a graphical user interface with which continuous text, pictures or anything else can be scrolled and viewed even if it does not fit into the space in the display, window, or viewport.

search A search box or search field is a common GUI element. It is usually a single-line text box with the dedicated function of accepting user input to be searched for in a database. It helps the user to find the information he is looking for by allowing him/her to introduce the terms.

sensemaking This is the process by which people give meaning to experience and understanding in high complexity or uncertain situations.

sensor A device that measures or detects a physical property and then records or converts them into signals which can be read by the user. It can also respond to what was detected. Some well known examples of sensors are: thermometers, pressure sensors, barometer, acceleration sensors, motion sensors, etc.

Sequential Graphic Graphics with a chronological order of events, for example timelines.

size representing quantity This visual representation usually consists on an object of a geometrical shape or a pictogram which its size represents a quantity. There are commonly more than one and are used for comparing quantities.

slider An object in a Graphical User Interface, also known as track bar in Microsoft literature, with which a user may set a value by moving an indicator, usually in a horizontal fashion. In some cases the user may also click on a point on the slider to change the setting.

Slide Show A structure that follows a typical slide show format. It can incorporate interaction within the confines of each slide, allowing the user to explore particular points before moving ahead to the next stage. It has an order imposed by the author but it is not necessarily a chronological sequence.

social visualization See Subsection 3.2.5.

span chart This chart is used to display ranges between a minimum value and a maximum value, focusing the attention on the extreme values. It is similar to box plot in appearance but it does not include the additional lines. the It is ideal for making comparisons of ranges labeled with categories.

speech balloon A graphic convention used most commonly in comic books, comic strips and cartoons to allow words to be understood as representing the speech or thoughts.

stacked bar chart A bar chart with different groups represented on top of each other.

stacked area chart A chart based on the area chart, with the different areas represented on top of each other in order for the final shape to represent the overall total.

storytelling The transmission of events/occurrences in words, sound and/or images, comprising a story/narrative . Stories/narratives have been shared in every culture as a means of entertainment, to educate, and to instill moral values. It is deeply rooted in oral tradition, in other words, in the information that is passed down through the generations by word of mouth.

streamgraph A type of stacked area chart, on which the areas are displaced around a central axis. This visualization type was created by Lee Byron and later analyzed in detail by its author in the paper *Stacked Graphs – Geometry & Aesthetics* (Byron and Wattenberg, 2008).

streaming In computing, it is the transmission of data in a continuous stream while earlier parts are being used.

sublime In aesthetics, or the philosophy of art, the sublime is the quality of greatness. The term refers to what "inspires awe, grandeur, and evokes a deep emotional and/or intellectual response" (Kosara, 2007, p. 633).

sunburst chart Also known as a multilevel pie chart, it is used to visualize hierarchical data, depicted by concentric circles. The circle in the center represents the root node, with the hierarchy moving outward from the center.

table It consists of an ordered arrangement of data usually in rows and columns for ready reference. The use of tables is pervasive throughout all communication, research and data analysis. Tables appear in print media, handwritten notes, computer software, traffic signs and many other places.

Tableau It is one of the most popular visualization products and includes both data analysis and visualization tools. Its popularity probably comes from the fact that is such an easy tool to use, requiring no programming skills. More in Subsection 10.1.1.

Tag Cloud This type of visualization is a representation for text data, more specifically keywords or tags. Tag Clouds are useful to show which words occur more often, the size of the word being the differentiating factor.

tag cloud Typically represented by the ordering of tags or words inline, in alphabetical order or randomly, and usually manipulated so the the font size represents number of occurrences.

text It is usually a larger information than the annotations. It is commonly more descriptive and more story like. The term text is also used for unstructured text, mere disconnected words.

timeline A way of displaying a list of events in chronological order. It is typically represented by a long bar labelled with dates alongside itself and usually events labelled on points where they would have happened.

timetable A representation of a plan that sets out the times at which events are intended to occur. A example of a timetable is a schedule that shows the times when transport (such as a bus or train) is expected to leave and/or arrive.

title The name given to the visualization. Almost every visualization has it.

toolkit A set of software tools that are used to develop other applications.

tooltip A common graphical user interface element used in conjunction with a cursor. A small box with information that appears when the user hovers or clicks a certain point or item.

transit map A kind of topological map (a simplified map where all the unnecessary detail has been removed and only vital information remains) used to represent the routes and stations of a transport system. It is a schematic diagram with color coded lines to indicate each line or service and named icons to indicate stops.

tree diagram A way of representing the hierarchical nature of a structure. It resembles a tree, even though the chart is generally upside down compared to an actual tree, with the *root* at the top and the *leaves* at the bottom.

treemap A treemap is a display of hierarchical data by using nested rectangles. Each branch is given a rectangle, tiled with smaller rectangles representing sub-branches. A leaf node's rectangle has an area proportional to a specified dimension on the data and is often colored to show a separate dimension.

usability A quality attribute that assesses the ease of use and learnability of an object (a software application, tool, machine, etc.). Usability is defined by 5 quality components: learnability, efficiency, memorability, errors, and satisfaction.

user contribution When a visualization allows user contribution it allows the user to enter information that will be displayed and become part of the visualization.

venn diagram A graph that employs closed curves and especially circles to represent logical relations between and operations on sets and the terms of propositions by the inclusion, exclusion, or intersection of the curves.

vernacular The term vernacular is commonly used to refer to the use of a nonstandard language or a dialect from a region in opposition to the use of a formal or official language. It is also used to refer to the normal spoken form of a language.

video A video consists of moving images that have been recorded. It can be a movie, television show, event, etc., that has been recorded onto a analogical or digital support so that it can be watched on a television or computer screen.

video narration Video narration is an audible content usually in the form of a story or a description of events that acts as support information for the visualization and is included in the form of video. This video can be the visualization itself or a separate media that is part of the visualization.

Video/Animation This genre comprises the visualizations in which a video or animation is the main part of the visualization. It also depends on other types of visualizations in other to be considered a Video/Animation visualization.

virtual reality The computer-generated simulation of a three-dimensional image or environment that can be interacted with by a person using special electronic equipment, such as Oculus Rift, Samsung Gear, or with a mobile phone and Google Cardboard (a set of goggles made from cardboard, with plastic lenses).

visual analytics A field of information and scientific visualization that focuses on the use of automated analysis techniques combined with interactive visual interfaces to promote understanding, reasoning, and facilitate decision making.

visualization The process of representing abstract data or information graphically, in a way that can aid in understanding the meaning of the data or information.

weighted network A network diagram where the edges among the vertices have weights assigned.

wheel A wheel chart is a circular chart divided into sectors, similar to a pie chart, on which the size of the sectors is always the same and does not represent a numerical proportion. It is commonly composed of several rings that can have the same size or represent a numerical proportion.

zoom (visual) Zoom as a visual element is an amplification of an image. It can be an amplification on which the original image has extra detail or simply an amplified version of the exact same image.

zoom (interactive) A zooming user interface or zoomable user interface is a graphical environment where users can change the scale of the viewed area in order to see more detail or less, and browse through different documents.

Acronyms

API Application Program Interface.

CSS Cascading Stylesheets.

D3 Data Driven Documents.

DIKW Data/Information/Knowledge/Wisdom.

DOM Document Object Model.

DSL domain-specific language.

EDA exploratory data analysis.

HCI Human Computer Interaction.

IDE integrated development environment.

InfoVis IEEE Information Visualization.

SDA social data analysis.

SUM Single Usability Metric.

SVG Scalable Vector Graphics.

TED Technology, Entertainment, Design.

VAST IEEE Visual Analytics Science and Technology.

VDSP visual data storytelling process.

VML Vector Markup Language.

WWW World Wide Web.

Chapter 1

Introduction

The digital revolution has created new media and transformed the field of communication. This revolution presents challenges not only for these new media that are born in cyberspace, but also for traditional media such as newspapers, radio stations and television channels. The Web is stimulating the worldwide dissemination of information. Therefore, traditional media are now trying to take advantage of the possibilities of the Web in order to become true online media. This migration of traditional media to the online environment has been slow and difficult. Although the media are moving each day farther way from the simple reproduction of pages of the printed version of a newspaper on a Web adapted layout, journalists are not yet using the full capabilities of the Web. At least, most online newspapers/radio stations/television channels now produce news using a new syntax consisting of words, sounds, videos, graphics and hyperlinks, all combined for the user to choose their own path of scanning information. This, however, is not enough. It is imperative for media to find new ways of storytelling that are appealing to the public.

Three decades after the creation of the Internet, it has become important to reflect on the impact of the Web on journalism in order to provide the public interesting ways to absorb information. Interactive visualizations and other visually appealing ways to present information seem to be a way to catch the attention of a public that is starting to buy fewer newspapers, and watching less television. Moreover, these new techniques allow the public to access information that it could formerly not be included on traditional media articles, mainly because of physical limitations. More tools (such as visualization, interactivity, maps, live streaming, and mobile applications) exist than ever before. The possibilities for using and mixing all these different techniques seem endless, and some online journalists have already begun integrating them into their narratives. However, various techniques used in other areas can also be applied to journalism. It is important to understand the range of techniques available, no matter what area they come from, in order for journalists to select which techniques work better for news storytelling and to discover new ways to digitally tell news stories.

One valuable option for the area of journalism is the visualization of information, which lately has become an essential tool for data comprehension, analysis, and exploration. With the explosion of the amount of information available (from sensors, governmental open data, etc.) "gathering, filtering, and visualizing what is happening beyond what the eye can see has a growing value" (Gray et al., 2012). Specially when we are talking about digital information, the increase was from scarce to superabundant, however the ability to extract wisdom and insight of the data is still low.

"As the eye is the best judge of proportion, being able to estimate it with more quickness and accuracy than any other of our organs, it follows that wherever relative quantities are in question, a gradual increase or decrease of any revenue, receipt or expenditure of money, or other value, is to be stated, this mode of representing it is peculiarly applicable; it gives a simple, accurate, and permanent idea, by giving form and shape to a number of separate ideas, which are otherwise abstract and unconnected" (Playfair, 1801). Since the processing power of the human visual system is immense, information visualization has revealed to be an amazing tool for understanding and retaining large amounts of complex information. Visualization also facilitates the recognition of patterns, reduces search times, and aids hypothesis formulation. Its power has long been proven (Card, Mackinlay, and Shneiderman, 1999; C. Chen and Yu, 2000; Lankow et al., 2012; Ware, 2004) and it has been taken advantage of since the man felt the urge to communicate.

There are several examples of some kind of data visualization in early 16th century. However only in the 18th century, after the birth of the probability theory and demographic statistics, we begin to see more formal visualizations (Friendly, 2008). The most well-known examples must be William Playfair's early line graphs and bar charts, and more elaborate visualizations such as Florence Nightingale's polar-area charts and Charles Minard's chart displaying flows. The last is also a pretty good example (and the actual pioneer) of the introduction of storytelling in visualizations.

In this thesis I depart from the premise raised with Minard's *Carte figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813*, that visualizations are able to tell stories, and explore what strategies can be used to introduce storytelling in visualizations. These strategies range from adding short stories or narrative elements using annotations and using time to introduce the feeling of storytelling or story-flow, to more complex strategies such as gamification. In this thesis the focus will be online visualizations that can include or not some degree of interactivity.

Over the last few years, storytelling has been a hot topic in the area of information/data/knowledge visualization. The interest in this topic, although it had its genesis in the area of Data and Information visualization with two of the pioneers being Gershon and Page (2001), also sparked outside of the research community due to its prospect of use in areas such as journalism, marketing, or education. Even some of the research approaches to this topic are closely related to the industry, for example Segel and J. Heer (2010), who coined the term narrative visualization and studied it in the area of Journalism.

With the qualities associated with information visualization and storytelling, narrative visualization can become very successful. However, establishing a correct balance between narrative and visualization is vital. We have to maintain the rigor and accuracy associated with visualization and not introduce narrative elements that could hamper the apprehension and assimilation of the information.

1.1 Motivation

Due to the explosion of information and knowledge we have been witnessing in the last few years, potentially sparked by the blooming of the open data movement, it is imperative that we discover better ways to understand information and to reduce the complexity of the information available. This boom of information also contributed to the crescent need of more appealing ways to represent the data, otherwise the public will not be interested in the point that the data is meant to illustrate.

This research started with the goal to understand what new types of news articles could be created for online media, how traditional narratives could be fused with sophisticated techniques from different fields, and which kinds of techniques the public likes best. Towards these goals, in early stages of the research, I started to see visualization popping out among the several strategies as a viable option. And the idea of having visualizations as a stand alone medium to do storytelling seemed appealing.

The challenge here is not only discovering ways to highlight the potential stories that exist within the data, but also to transform visualizations in such a way that they adopt several narrative characteristics and eventually become a form of storytelling itself. If visualizations get successfully infused with narrative we can overcome both the limitations of textual and visual representation.

The benefits in using visualization for supporting users in coping with the complexity in knowledge- and information-rich scenarios has already been proven (Keller and Tergan, 2005). However there are still opportunities to make a contribution, specially in the sub-genre of narrative visualization (visualizations intended to convey stories (Segel and J. Heer, 2010)). Another problem in information visualization research where there is also room for new contributions, and that can be closely related with narrative visualization, is the visual metaphors that shape the way the information is structured (Ziemkiewicz and Kosara, 2008). Visual metaphors have long been a concern, for information visualization researchers as well as for the creators of visualizations, because their interpretation depends too much on the user's internal knowledge representations. The question about "how can visualization systems be tailored to accommodate human perception and information processing" (Gershon and Page, 2001) continues to be one of the problems in the field of information visualization that has not yet been entirely solved.

This thesis was driven by questions such as:

- *What elements of traditional storytelling can be embedded in data-driven visualization?*

- *How do we balance the narrative flow without disturbing the experience of discovery?*
- *What elements of design and interactivity help us to better tell these data-driven stories?*

Having this in mind I examine the benefits of adding storytelling to visualizations and explore possible strategies to do so. Achieving an answer to these questions is important because the answers could be the starting point for building a set of guidelines for narrative visualization. Working towards the establishment of these guidelines would benefit both researchers and visualization creators, enabling new synergies.

This research aims to contribute to the field of information visualization not by presenting new visual representations but by transforming visual representations that already exist in a way that allows these visualizations to serve as better means of storytelling. I believe that this topic is greatly relevant and the questions posed are important not only for the information visualization community but also to every area that wishes to elevate visualizations to a more complex form of dissemination of information/data. Empirical study is much needed for the field to move forward.

Problems and needs in the research area of Information Visualization

Nowadays, most research on visualizations is still based on the time it takes to complete a task (Kosara and Mackinlay, 2013). However this makes little sense when the goal is to produce engaging visualizations, that people spend time on. This does not mean that functional visualizations which have the sole purpose of conveying the data as simply and clearly as possible will cease to exist, it only means that a new kind of visualizations is blooming and that its goals may differ slightly from the goals of traditional visualizations.

This new kind of visualization has recently gained a world-wide popularity boost by the fact that technology provided us with new tools to convey information, inclusively in a story-like fashion (Gershon and Page, 2001). People get excited with good visualizations and the proof of this fact is that people are sharing these visualizations online, often not even caring if they have an article associated with it.

For people to get engaged with the visualizations these must be both appealing and informative. However there is a fine balance between functionality and aesthetics that should by all means be preserved. According to C. Chen (2010), information visualization is close to science and has to maintain patterns of rigor, accuracy, and faithfulness. Nevertheless, nowadays there is a misunderstanding about what visualization should be and sometimes the creators spend too much time on looks and forget the real purpose of a visualization. Although it is important to have aesthetically appealing visualizations their main purpose is to inform and it is only when they excel at their main purpose that they reach true beauty. Visualizations that are not successful in providing access to the information are failed visualizations (Steele and Iliinsky, 2010).

According to C. Chen (2005), the "Top 10 Unsolved Information Visualization Problems" were: usability; understanding elementary perceptual–cognitive tasks; prior knowledge;

education and training; intrinsic quality measures; scalability; aesthetics; paradigm shift from structures to dynamics; causality, visual inference, and predictions; and knowledge domain visualization. Some of these problems even cross to other areas such as visual analytics (Keim, Mansmann, et al., 2008), which, like visualization, is built upon methods from scientific analytics, geospatial analytics and information analytics.

Other concerns approached by researchers in the field of information visualization are the development of new visual representations and new interaction techniques. These are not, however, the approaches taken in this dissertation. Although this research skims through some of these issues that are relevant for the area of information visualization, such as aesthetics and quality measures the focus is really on the users comprehension and satisfaction. Having this in mind this research also tries to shed a light on possible visual encodings for certain types of information. Although I also try to shed a light on what types of visualizations work for narrative visualization I believe that trying to find specific types of visual encoding for a kind of information is not the approach that should be taken in information visualization research. This is because every data set is different and the stories within the data must be highlighted according to the targeted audience. Furthermore the visualization will always be affected by the quality of the data representation (Thomas and Cook, 2006).

This area of research is still very new and there is still little information on how to introduce storytelling in visualizations and even less research on what techniques work for the audience. It would benefit from rigorously studying and measuring the impact of visualizations. According to Keim, Mansmann, et al. (2008), in the area of visual analytics "user acceptability is a further challenge; many novel visualization techniques have been presented, yet their wide-spread deployment has not taken place, primarily due to the users' refusal to change their working routines". However, what it has been seen lately in the field of information visualization is that the users are getting more and more familiar with most types of visualizations due to its heavy use in the media, and their visualization literacy is improving.

Problems and needs in Online Journalism

The impact of the Information Age on traditional mass media was massive. The newspaper industry, television and radio were completely transformed by this revolution, specially by its most iconic breakthrough: the Internet. In a way it is possible to say that the old media have undergone a process of metamorphosis. This is a phenomenon that Fidler (1997) named as mediamorphosis: "The transformation of communication media, usually brought about by the complex interplay of perceived needs, competitive and political pressures, and social and technological innovations." However, this transformation was not a complete rupture with old mass media, instead the new and the old media coexist in a sometimes difficult relationship.

Newspapers have been for some time under the threat of becoming an endangered species and broadcast television and radio, although in a more comfortable position, are struggling too. However, this is not a adage to the end of news. Possibly more than ever, news will thrive. In 2010, the Pew Research Center revealed that "Americans are spending more time with the news than was the case a decade ago" (PRC, 2010). However this fact is due to the increase of the online and mobile audience. Even television news viewership is starting to suffer with only about a third (34%) of those younger than 30 who participated in the 2012's PRC (2012) News Consumption Survey saying they watched TV news the day before, while in 2006, nearly half of young people (49%) said they watched TV news the prior day. Overall television continues to be the preferred media for news gathering and the one which people spend more time with.

Meanwhile, online news audience grows every day. This audience is generally young, between the ages of 18 and 29, and for them the Web is the main source of their daily news (PRC, 2012). However, even for people between the ages of 30 and 49 the Internet is an important news resource: "48% named it as a main source for news, twice the percentage that rely on newspapers (22%). Among this age cohort, television is still first (63%)" (PRC, 2012). They are also active participants. New survey data found that "half (50%) of social network users share or repost news stories, images or videos while nearly as many (46%) discuss news issues or events on social network sites" (PRC, 2014).

Even though the Internet is nowadays an important source of information and the old media are struggling to maintain their supremacy in the diffusion of news, Internet has not and will not completely kill newspapers, television, and radio stations. Some of these new media were specifically made for the Web and do not have any representation outside the online environment, but the more traditional ways of doing journalism such as newspapers, magazines, television, and radio stations have migrated and continue to migrate to the Web. While some stopped existing in the real world, others maintain their old media and an online version. As a matter of fact, top news websites belong to old media corporations and most are newspapers, the ones who suffered the most with the threat of online media (PRC, 2012).

Nevertheless, a simple transcription of the old media format to the Web is not enough to catch the audience's attention anymore. Online citizens are developing new cognitive skills that do not fit the verticality of the old media industry. Therefore, it is imperative that the media, which were previously governed by the paradigm of mass communication (vertical and unidirectional), adapt to this more open and horizontal communication system. Journalists have to adapt to this new level of immediacy and ubiquity, where news get old even quicker than in the traditional media. Deuze (2001) suggests that online news require an important shift in the culture of journalism. Some changes already occurred and continue to be implemented.

Nowadays, journalists are confronted with more tools than ever before: data visualization, interactivity, maps, live streaming, mobile applications, etc. The possibilities for using and mixing all this different techniques seem endless, and some online journalists have already began integrating them into their narratives. However, there are lots of techniques, used in other areas, which can also be applied to journalism. According to Bogost et al. (2010), "at its core, news is comprised of ideas. It is not made of folded newsprint, broadcast studios, or Web pages (...) It is a practice in which research combines with a devotion to the public interest, producing materials that help citizens to make choices about their private lives and their communities. There is nothing medium-specific about journalism, no reason that its output must take the familiar form of text, image, or video." This is why it is so important to understand the range of techniques available, no matter what area they come from (art, gaming, etc.), to select which techniques work better for news storytelling, and to discover new techniques to digitally tell news stories. "Those who excel at coming to the point of a story will have to learn to stretch the story, to develop it in the most imaginative and complete possible way, using all of these different new possibilities which appeal to all of our senses" (Giussani, 1997).

It is not enough to simply be online. The consumers are driving this cultural change that is happening in online newsrooms. They expect a variety of features, such as the possibility to discuss, to participate, to contribute, to explore freely without an imposed sequence, and if they do not get what they want they simply look for it in another website. The simple existence of links changes the whole interaction the reader has with the news. According to Hall (2001), "each one that is encountered by the reader forces a decision as to whether to follow the link or stay within the anchor text. The process insists that the reader thinks about the text in a way that print and broadcast texts do not." The journalist has the responsibility to provide multiple directions for the consumers to follow, almost as the author of a computer game that gives various possibilities to the player and encourages him/her to interact with the world, but also makes sure to facilitate the interaction, influencing him/her decisions (Burton, 2005).

1.2 Approach

With this scope, the research topic of this dissertation lies at the intersection of two main fields: Information visualization and Journalism. "Visualization has proven to be an effective strategy for supporting users in coping with complexity in knowledge- and information-rich scenarios" (Keller and Tergan, 2005), specially when the challenge is to represent large sets of data. Journalists have long been using visualization to present complex data to the public but have to rely on the public's ability to understand the visualizations. Moreover, dissecting massive data sets, such as census data, and transforming it in bit size information that can be understood by the majority of the public is an even bigger challenge. Introducing storytelling in visualizations might be a way to ease the interpretation of the visualization and this would be a great benefit for journalism.

Most of the findings can also be applied to other areas where storytelling can play a part on the explanation of data. The clearest examples of areas that would benefit with the introduction of storytelling in visualizations would be education and scientific visualization (as Ma et al. (2012) reported). In both these areas, the challenge is to produce visualizations that are understandable, even when composed of information that the final consumer might not have previous knowledge on, therefore having difficulties in interpreting it.

The research started with the analysis of professionally-produced visualizations. This analysis is available at a website, entitled ReThinking Visualization [1], created for this research. After this qualitative analysis of several visualizations, it became clear that visualizations are still very much used as a support of the written format. This review also allowed to gain empirical knowledge on visualization structures that facilitate the introduction of storytelling in a literal

[1] http://rethinkingvis.com/

way or in a semantic way. Doing storytelling in a semantic way means mimicking the qualities of storytelling, for instance having the sequential flow of a story, in the actual structure of the visualization. This is the case of timelines.

"Visualizations naturally perform a contextualizing function when they present data to accompany news" (Hullman, Diakopoulos, and Adar, 2013). Although infographics are very good at it, they are not only useful to convey huge sets of data, but they can also be used to tell stories, and lately this potential has been hugely discussed. Introducing storytelling in the visualizations is still a challenge and it is usually easier to associate the visualization with a story by having a link from one to the other than to incorporate the story on the visualization. However, several strategies to incorporate storytelling in visualizations have been used by news media, advertising agencies, and other industries, unfortunately without any reasoning about its efficacy. Part of this research consisted in identifying these strategies and analyzing them. This analysis led to a reflection on the benefits of the incorporation of narrative in visualizations and the possible public's reaction to these new visualization strategies.

"For as long as people have been around, they have used stories to convey information, cultural values, and experiences" (Gershon and Page, 2001). The benefits of using stories for enhancing comprehension and memory have already been proven. However, there is not enough evidence on how these benefits translate to narrative visualization (Hullman, Diakopoulos, and Adar, 2013) and which kinds of sequences are effective both in maintaining the explorative qualities of information visualization and in conveying a narrative. In this thesis, I approach this issue having in mind the inputs received from the participants of a focus group study, in which the participants were asked to evaluate several visualizations and give their opinion in terms of comprehension, likability and navigation. Based on the results of this study and confronting it with the literature on this subject, I identify effective sequences and several elements that reflect the user preferences.

1.3 Research questions

The fundamental hypothesis of this research is that it is possible to tell stories using visualization, either by having the visualizations structured in a way that resembles the storytelling flow or by incorporating narrative moments in the visualizations. This study addresses three general framing questions:

- RQ1: Which techniques can be used to tell stories using visualization?
- RQ2: Which elements can be used to have an effective storytelling?
- RQ3: Which types of visualizations might appeal to the public?

These questions guide the overall research presented in this dissertation. The questions were formulated in an open-ended manner in order to have the shifting that the research process implies. Although these research questions were formulated in a broadly fashion, they are intended to answer questions in the specific case of journalism. However, since the same findings apply to different areas, the research was intentionally opened to other areas of knowledge. Achieving the answer to these questions is crucial to the understanding of the progress that has been done on the field of visualization and of the possibilities that the Web provides to this area. It is important to understand the range of techniques available, no matter which area they come from, to be able to evaluate and select which techniques work better for storytelling and to discover new ways to digitally tell stories.

1.4 Methodology

To achieve these goals, I used four different methodologies, at different stages of the research: literature review, mixed-method evaluation, design studies, and critical reflection. These stages correlate with the general structure of the dissertation. The methodologies chosen are a combination of both quantitative and qualitative methods.

Literature Review. To introduce the research subject and conceptualize the steps that have been given towards the establishment of narrative visualization as a research field, I first and foremost present an extensive review of the literature. I thoroughly study prior work on information visualization and other disciplines close to this field of research such as knowledge visualization, scientific visualization, and Human Computer Interaction (HCI). In addition, I synthesize evidence from studies on the particular sub category of narrative visualization to derive a theoretical approach that drives the work presented in this dissertation. I also survey some literature on online journalism with a focus on visualization and other types of visual forms of storytelling. Although most literature review is concentrated in the first three chapters, relevant research is cited through out the rest of the dissertation.

Mixed-method Evaluation. My research strategy has been to first present a mixed-method evaluation composed of a new interaction techniques taxonomy, a visualization typology, case-studies, and a focus group study. The typology proposed is a classification system taken from patterns in visualizations that have already been done by various online newspapers, magazines, and other online media. Related to this typology, I will present several case studies that illustrate it and compare it to several other typologies. As part of this mixed-method evaluation a focus group study was also done as an exploratory evaluation to understand which techniques the public likes best. The interaction techniques taxonomy derives from the interactive elements that are currently being used in the field and that were observed in the visualizations that were studied. It was created to guide the research on the actual benefits of interactivity in visualization and to serve as a framework to help discuss and evaluate interaction techniques.

Design Studies. Having the previous mixed-method evaluation results in mind, I focused on design studies and techniques review. At this stage, I present some mock-ups to demonstrate three of the strategies of storytelling approached in this thesis: Context; Empathy; The Relation Between Time and Gamification. For the fourth strategy presented, gamification, I opted for a review of different examples that already exist instead of the mock-up. This decision was based on the fact that game-y visualizations are composed of several layers of interaction, and transposing them to a mock-up would be difficult. In the review of the gamified visualizations I present both the strong and weak points of the examples presented and compare them. Linked to the design studies, I also review several tools for building visualizations, emphasizing the ones that better enable the introduction of the narrative strategies previously presented. In the same section, I also present some guidelines for choosing the most effective type of visualization for different types of data.

Critical Reflection. The final part of this thesis is a critical perspective on narrative visualization, and all the techniques and tools commonly used in this research area. I explore the different roles involved in the creation of these visualizations and reflect on the use of visualization to engage audiences. I also present an overview and a comprehensive discussion of the concepts, tools, and techniques presented in this thesis, and reveal the limitations of this work, pointing out opportunities for future work.

1.5 Overview

As summarized in Table 1.1, the dissertation has four parts: State of the art, Methodology, Possibilities for narrative visualization and Reflecting on narrative visualization.

Part	Chapter
State of the art	chapter 2 – Visualization analysis chapter 3 – The many names of visualization chapter 8 – Using visualization to tell stories
Methodology	chapter 4 – The role of interactivity chapter 5 – A new Visualization Typology chapter 6 – Typology Case Studies chapter 7 – Focus group study
Possibilities for Narrative Visualization	chapter 9 – Narrative strategies chapter 10 – Techniques for visualization on the web
Reflecting on Narrative Visualization	chapter 11 – Conclusions

Table 1.1: Table of contents

Chapter 2, Chapter 3, and Chapter 8: Part I composes the literature review both of information visualization as a research area and as a valuable tool for journalism. I give an overview of the new forms of storytelling that emerged with the popularization of the Web and establish a relationship between these and strategies of storytelling for visualization. I also highlight the specificities of the area of Information/Knowledge/Data visualization and establish some possible connections between the blooming of the Open Data Movement and the increase of use of visualization in several areas. Moreover, I reflect on the whole history of the use of visualization, storytelling, and the combination of the two. This is important to understand the possible uses of narrative visualization and to position the goals of this research.

Chapter 4 – Chapter 7: Part II forms the core of this dissertation, presenting both the methodologies used in this research and the findings. I present a new typology for characterizing visualizations and provide several case studies to exemplify every type of visualization. I also report on an exploratory focus group study conducted with the purpose of collecting information on the narrative elements in a collection of visualizations and the possible inclusion of storytelling elements in those. Moreover, I highlight the importance of interactivity and propose eleven categories of interaction techniques. The ReThinking Visualization website — a project built with the intent to help building a better understanding of visualization through the dissection of all the pieces that compose a visualization in order to detect patterns — is also presented here.

Chapter 9 – Chapter 10: Part III offers a critical perspective on information visualization and possible strategies for the introduction of storytelling in visualizations. In this part, I highlight several techniques, tools, and guidelines to build narrative visualizations.

Chapter 11: I then conclude with Part IV, which presents a summary of the whole research and present new avenues of research that still can be explored in narrative visualization research. This thesis ends with a more in-depth reflection about how these strategies can contribute to the growing efforts to change the usual approach of integrating visualizations into storytelling.

1.6 Contributions

The research developed in this thesis has three major contributions:

- a new taxonomy of interaction techniques for visualization, explained in detail in Section 4.3;
- a new visualization typology, reviewed in Chapter 5 and further explained in Chapter 6 through case study analysis;
- and a set of narrative strategies for visualization, analyzed in depth in Chapter 9.

Over the course of this dissertation, I also present other contributions that helped shape the whole research. This is the case of the **ReThinking Visualization website**, presented in section 5.3, which includes the analysis of the elements that compose each of the 298 visualizations in the collection. Another minor contribution worth mentioning is the review of several **techniques, tools, and guidelines to build narrative visualizations**, presented in chapter 10.

Materials, ideas, tables, and figures in this thesis have appeared previously in the publications below. After each reference, I note the chapters in which the material is used.

- Figueiras, Ana. "A Typology for Data Visualization on the Web." In Information Visualisation (IV), 2013 17th International Conference, pp. 351-358. IEEE, 2013. Material from this publication appears in Chapter 5, entitled A new Visualization Typology, and Chapter 6, entitled Typology Case Studies.

- Figueiras, Ana. "How to Tell Stories Using Visualization." In Information Visualisation (IV), 2014 18th International Conference on, pp. 18-18. IEEE, 2014. Material from this publication appears in Chapter 7, entitled Focus group study.
- Figueiras, Ana. "Narrative Visualization: A Case Study of How to Incorporate Narrative Elements in Existing Visualizations." In Information Visualisation (IV), 2014 18th International Conference on, pp. 46-52. IEEE, 2014. Material from this publication appears in Chapter 9, entitled Narrative strategies.
- Figueiras, Ana. "Towards the Understanding of Interaction in Information Visualization." In Information Visualisation (IV), 2015 19th International Conference on, pp. 46-52. IEEE, 2015. Material from this publication appears in Chapter 4, entitled The role of interactivity.

Pre-print versions of these publications are included in Appendix I.

Chapter 2

Visualization analysis: its history and paradigms

Today's society is surrounded by data. This data is increasing in numbers, in complexity and in terms of accessibility (Keller and Tergan, 2005). Recently, citizens have gained more and more access to information on their country and population, which was previously only accessible to governments. "Public bodies are among the largest creators and collectors of data in many different domains, e.g., geographic data, tourist information, statistical and business data, weather information, and so on" (Janssen, 2011, p. 446). Moreover, people have developed an appetite for information and are constantly looking for the most recent data, in order to make better decisions. Consequently, companies and advertisers are always trying to push their information (Krum, 2013). As reported by Krum (2013) people have become *informavores* — a term coined by Miller (1983) — and depend on information to feed the mind as the body depends on food. Information is what enables their mind to survive (Miller, 1983).

Although humans have always been curious beings, this curiosity has grown, possibly fostered by the developments that enabled us to have easier access to information and to process large amounts of information more easily. The World Wide Web (WWW) was definitely one of the driving forces of this data consumption and producing trend. Moreover, the developments in the area of big data have facilitated the storage, analysis, and processing of information, allowing us not only to better quantify the world but also to get insights that were heretofore difficult to encounter. This data that before was unmanageable and completely unstructured (mere words, numbers, images, sensor data, etc.) can now be properly processed and we can extract knowledge and insights. Techniques of artificial intelligence such as Natural Language Processing and Machine Learning are currently being used to tackle this issue.

The fact that there are new and better ways to process data also encouraged many entities to collect more data in order to make better decisions, because having more data usually reflects in an easier and better extraction of patterns and meaning (Bradshaw, 2014). "Decisions that previously were based on guesswork, or on painstakingly handcrafted models of reality, can now be made using data-driven mathematical models" (Jagadish et al., 2014, p. 86). And this has become so popular that nearly everyone is doing it: governments, retail, researchers, etc. Furthermore, this multitude of data has become extremely valuable. Companies, such as Facebook and Google, which have access to the users preferences and habits have a huge advantage and can tailor their ads to each user using the data he/she gives them. "Data has become a new class of economic asset, like currency or gold" (Lohr, 2012, Website).

The quantity of data leads to a problem that is hard to solve: the feeling of information overload that is growing on people. According to Dörk (2012), both in their personal and professional lives, people become overwhelmed with the amount of information available and that can harm their informed decision making process. "Today's abundance of digital information — as exemplified by our email inboxes, news feeds, and Web search results — can be viewed similarly, as both overwhelming information overload and fascinating information access" (Dörk, 2012, p. 2). The volume of information is not the only reason for this overload. Its diversity is also problematic. "Diversity may occur both in the nature of the information itself, and in the format in which it appears, with a typical business user having to deal with paper, e-mail, voicemail, traditional websites, and so on, to which the newer blogs, wikis and the like must be added" (Bawden and Robinson, 2009, p. 184). Often these resources also refer onwards to other content through the use of hyperlink (Dörk, 2012). Moreover, the fact that information can also change over time, be updated, or disappear can increase information anxiety. Poor organization and presentation are also common causes of information anxiety.

According to Bawden and Robinson (2009), improving information literacy is clearly part of the solution to the information overload issue. In order to prevent the feeling of information anxiety, users have to be able to locate, evaluate, use, and communicate information in the most efficient and accurate way. Although practice ends up diminishing the feeling of anxiety, creating visual representations that make the data easier to understand can also help. Being able to represent this plethora of data in a way that is easy for humans to process has become a major concern in a wide variety of fields, from scientific research to news media, and specially in the field of HCI.

Information visualization has bloomed as a good option for representing data , and has lately become a particularly popular option for representing data on the Web mainly because this medium provides the conditions for it to thrive (such as interactivity, connectivity, and scalability). The newest browsers support rich multimedia content and interactive features that allow a more engaging exploration of the visualizations, without third-party plugins (Dörk, 2012). However, we cannot expect that simply representing the data in a visualization will be good enough to represent all of the insight that the data can unravel. It is vital that the data is represented in a way that is both appealing and understandable in order to maintain the users interested. Moreover, the information has to be pre-structured in order to be easily accessible by users (Keller and Tergan, 2005). Nonetheless, how much of the data is shown and to what depth the users can go if they wish to must be balanced. Interactivity also has to be introduced thoughtfully. Several creative opportunities arise here.

2.1 Visualization history and its most influential examples

Who thinks that information/data/knowledge visualization is a new fad is completely wrong. Its use is nothing new. Even the earliest cave paintings can be understood as some form of visualization (Lankow et al., 2012), just not as focused on large sets of data as modern-day visualization. Moreover, maps, which are also a type of visualization, have been extensively used since 16,500 BC. It is possible to find several examples of graphs and timelines during the 1600s. But, according to Rogers (2013), William Playfair's charts mark the birth of infographics, in the late 18th century, if considering modern charts. One of his well known charts, shown in Figure 2.1, is a time-series charts of *Exports and Imports to and from Denmark and Norway from 1700 to 1780*, published in his Commercial and Political Atlas in 1786. This time-series shows England's balance of trade with Denmark and Norway, presenting the relation between time and units. He also makes use of color to distinguish between the balance in favor (light brown) and against (light pink) England, and between exports (red) and imports (yellow).

William Playfair, in addition to being a pioneer in the use of graphical displays, is also credited with the invention of several charts/diagrams, such as line charts, bar charts, pie charts and circle graphs (Friendly, 2005). His first bar chart dates back to 1786 and was the first to compare two discrete quantitative variables other than space or time (as the bars in Joseph Priestley's timeline, which were used to represent the life span of a person).

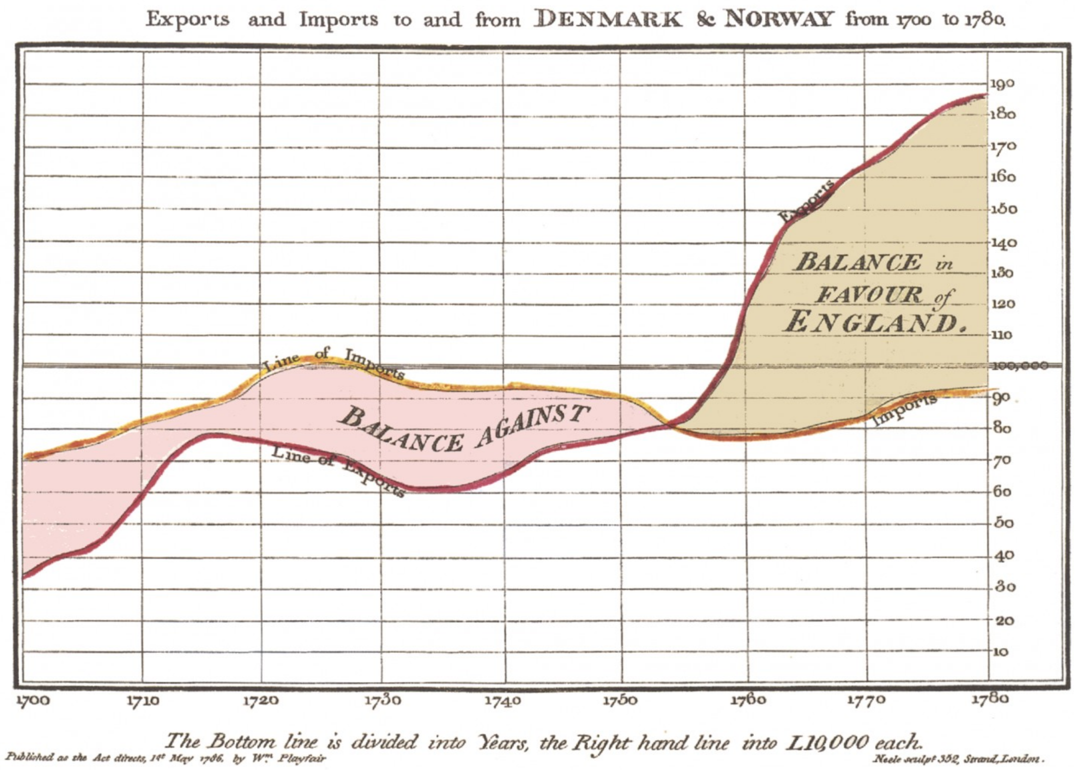


Figure 2.1: Playfair's time-series chart of *Exports and Imports to and from Denmark and Norway from 1700 to 1780*, published in his Commercial and Political Atlas

At first, visualizations were almost exclusive of the scientific domain with the purpose of validating experiments, communicating findings, exploring datasets and seeing data that can be invisible at first sight (Ma et al., 2012). There are also several examples of its early use in engineering and economics. In these domains, the visualizations are usually created for an audience that wont find difficult to interpret the visualization. Even some of the most popular early visualization examples, such as Florence Nightingale report on the *Mortality of the British Army*, were created having in mind an educated audience (explained in detail in Subsection 2.1.1). However, since the late 1930s visualization has been an usual presence in news media, and these have a broader audience in mind. The *USA Today* and *Fortune* Magazine were some of the early adopters of visualizations. Data visualization has a long history and several different uses, however some are more memorable than others. Some visualizations had such an impact that were able to change mentalities and policies.

2.1.1 War Mortality – Florence Nightingale

One of the most influential visualizations is certainly Florence Nightingale's 1858 key report to the British Parliament. In this report, she resorted to infographics to show the causes of death in the British Army during the Crimean War. She intended to show the English Parliament that diseases had a bigger impact in the number of war fatalities than anything else. Therefore, improvements in health services for the British troops were needed. The Parliament had previously been warned about the lack of health and hygiene of the troops but was unresponsive.

Florence Nightingale's presentation included statistical graphics, pie charts (developed

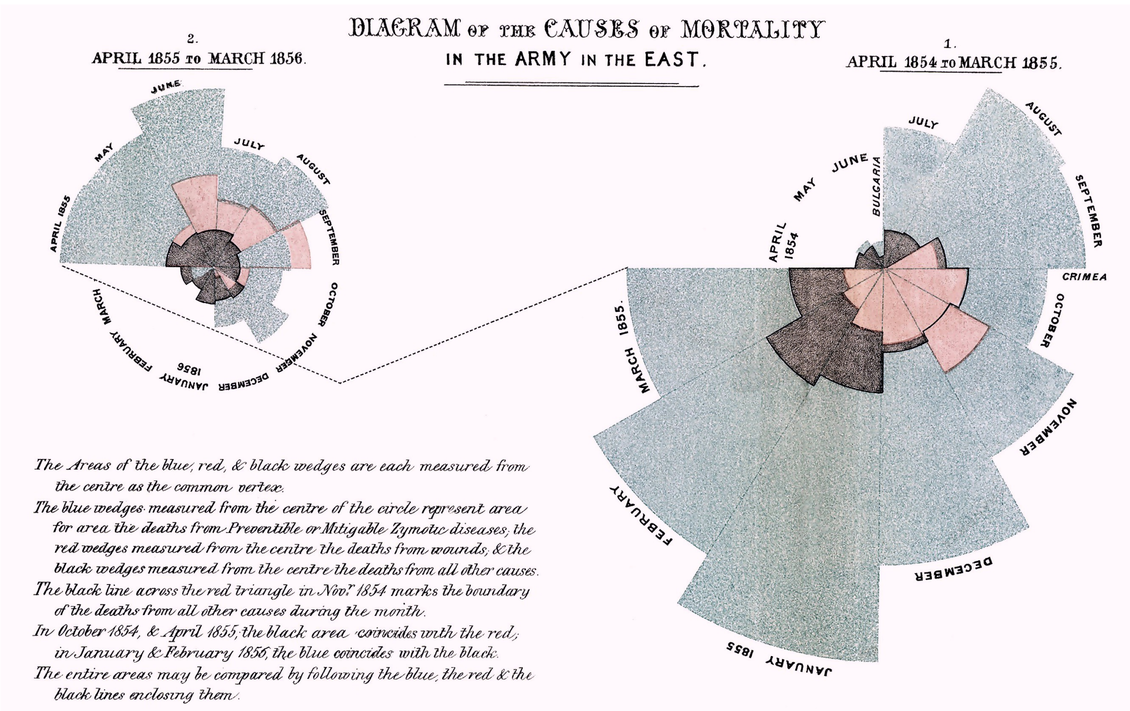


Figure 2.2: Florence Nightingale's coxcomb diagram on mortality in the British army

by William Playfair in 1801), and her famous coxcomb diagram, shown in Figure 2.2. She used the coxcomb to represent 12 months, one from April 1855 to March 1856 and one from April 1854 to March 1855. Both diagrams show the overall mortality due to deficient sanitary measures in the British army in Russia, Bulgaria and Turkey (Nightingale et al., 1859). For each month she drew a wedge that represents the deaths from preventable or mitigable Zymotic diseases (in blue), the deaths directly from wounds (in red), and the deaths that occurred from all of the other causes (in black).

The two diagrams illustrate how the sanitary reform, initiated in March 1855 and which included cleaning, clothing, nutrition, medication, general sanitation of the building and other measures, was able to dramatically reduced the death rate. Florence Nightingale was able to tell her story with data , successfully representing the amount of deaths due to preventable disease. The impact of this visualization was so big that sanitation became a major priority for the British Army.

2.1.2 Figurative Map of the successive losses in men of the French Army in the Russian campaign 1812-1813 - Charles Minard

For Edward Tufte this "may well be the best statistical graphic ever drawn" (Tufte, 1983, p. 28) and for Chevallier this chart "defied the pen of the historian" (Friendly, 2002, p. 45) because it is such a brutal portrait of the events. Charles Minard's *Figurative Map of the successive losses in men of the French Army in the Russian campaign 1812-1813*, shown in Figure 2.3, is often (by many) regarded as one of the best visualizations of all time. The original way Minard was able to represent Napoleon's disastrous losses suffered during the Russian campaign of 1812 is efficient, visually interesting and able to tell a story by it self.

Using the data available in the works of Messrs. Chiers, de Ségur, de Fezensac, de Chambray and the unpublished diary of Jacob (pharmacist of the Army since October 28th),

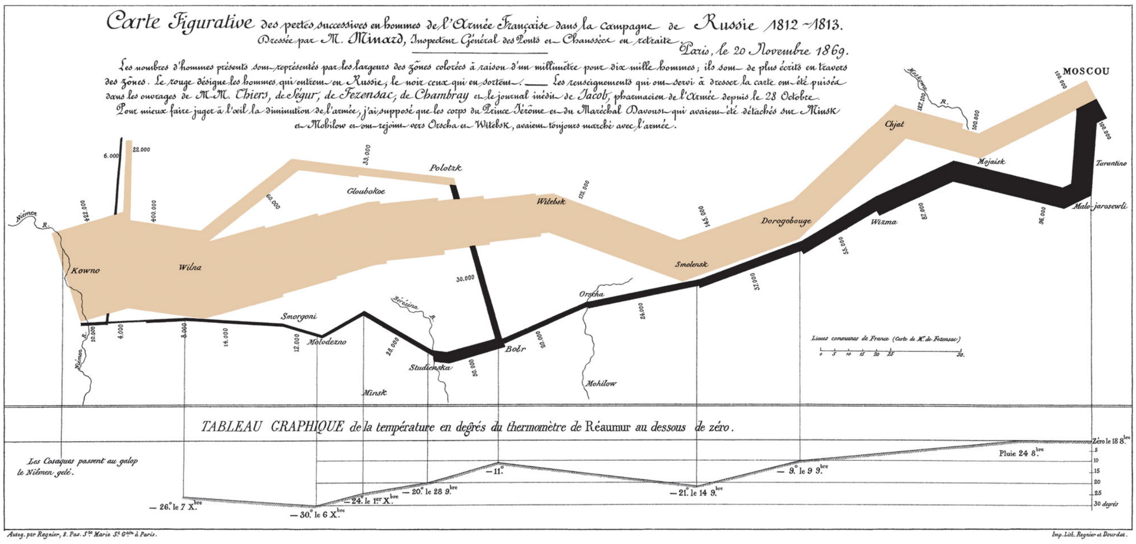


Figure 2.3: Minard's *Carte figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813*

Minard drew a map/timeline/flowchart visualization to represent the journey of Napoleon's army from the Polish border with Russia towards Moscow. The width of the brown band represents the size of the French army moving to Moscow, at a scale of one millimeter for ten thousand men. The black band represents the size of the army on retreat. The number of men is also written next to the band. The position of the band depicts the geographic location of the army. In order to facilitate the representation of the number of men, Minard displays the size of the army as if the troops under Prince Jèrôme and under Marshal Davoust, who were sent to Minsk and Mobilow and who rejoined near Orscha and Witebsk, were always part of the army that marched towards Moscow. It is also important to highlight the significance of rivers in this particular visualization because, in sub-zero temperatures, crossing a river had an impact on the number of causalities. In Minard's map it is clear that many men lost their lives crossing of the Berezina river, since the map shows around 50 thousand soldiers approaching the river and only 28 thousand being able to continue the retreat to France from there.

Minard was 80 years old and long retired from the French government's *Council des Ponts et Chaussées* (of which he was general inspector) when he created this visualization using innovative techniques for displaying flows of people. Minard was able to successfully illustrate several different types of data in the same visualization: the path and distance traveled, the number of men in the army at any given point, the dates on which the army was at a determined location, the temperatures registered on the return journey (shown along the bottom), and major battles that the army faced (of which the impact in number of causalities is shown by the strangulation of the band). With this visualization he was able to tell the story of this catastrophic war where 412 thousand men (98% of the army) lost their lives (only 100 thousand reached Moscow and only 10 thousand returned home). Moreover, he was able to fascinatingly display the impact of the cold weather in the number of deaths registered.

The Figurative Map was so important in Minard's life that Chevallier, his obituarist, from all his graphical innovations, singled out this particular visualization because "the image is gripping; and, especially today, inspires bitter reflections on the cost to humanity of the madnesses of conquerors and the merciless thirst of military glory" (Tufte, 2006, p. 134). According to Friendly (2000), he was a true pioneer not only in statistical graphics but also in what geographers call *thematic cartography*. This map is also one of the first examples of narrative visualization, because it is able to narrate a journey, making use of a *semantic symmetry* that gives the sense of story-flow and temporal sequence. "It is immediately obvious that the declining width of the wider gray line, moving in a direction from left to right, represents the declining number of soldiers as the campaign proceeded" (Goebel et al., 2014, p. 27). The fact that Minard opted for a map visualization to tell the story of this journey is another reason for its success. He successfully translated the familiar action of following a path with a finger on a map and relied on the human capability to fill in the blanks.

According to Friendly (2002), Charles Minard was not *one-hit wonder* and through his work we can observe not only the ingenious use of graphic forms invented by others (such as Minard's use of Playfair's pie charts for the first time in a map to show both amounts and relative proportions) but also some inventions. However his last work was his finest achievement and "influenced several generations of statisticians and cartographers" (Friendly, 2002, p. 49). This type of band graph, for instance, was later named *Sankey diagram*, and is a specific type of flow diagram in which the width of the band is proportional to the flow quantity.

2.1.3 London Cholera Map – John Snow

Another visualization that had a significant impact was John Snow's map showing the clusters of cholera cases in the London epidemic of 1854 for his work *On the Mode of Communication of Cholera*, which can be seen in Figure 2.4. The English physician, in addition to being known as a leader in the adoption of anesthesia and medical hygiene, is known for his clever dot map that allowed him to show that the water supply was spreading the disease. Back in the 1850's there was little information on microorganisms, even less about cholera (which spread through London's Soho in a sudden outbreak), and none at all about bacteria. Cholera was believed to be an airborne disease.

No one was able to understand the source of this outbreak that killed more than 500 people in just 10 days. John Snow obtained from the General Register Office a detailed list of 83 deaths from cholera and plotted it on a map (and found out the common interactions among the people who had died). He then came to the conclusion that the higher death counts were clustered around the Broad Street water pump (Tufte, 1997). This fact alerted Snow to the possible causes of the outbreak. Consequently, it became apparent that the water of the Broad Street pump was polluted by sewage. "Thus the theory implicating the particular pump was

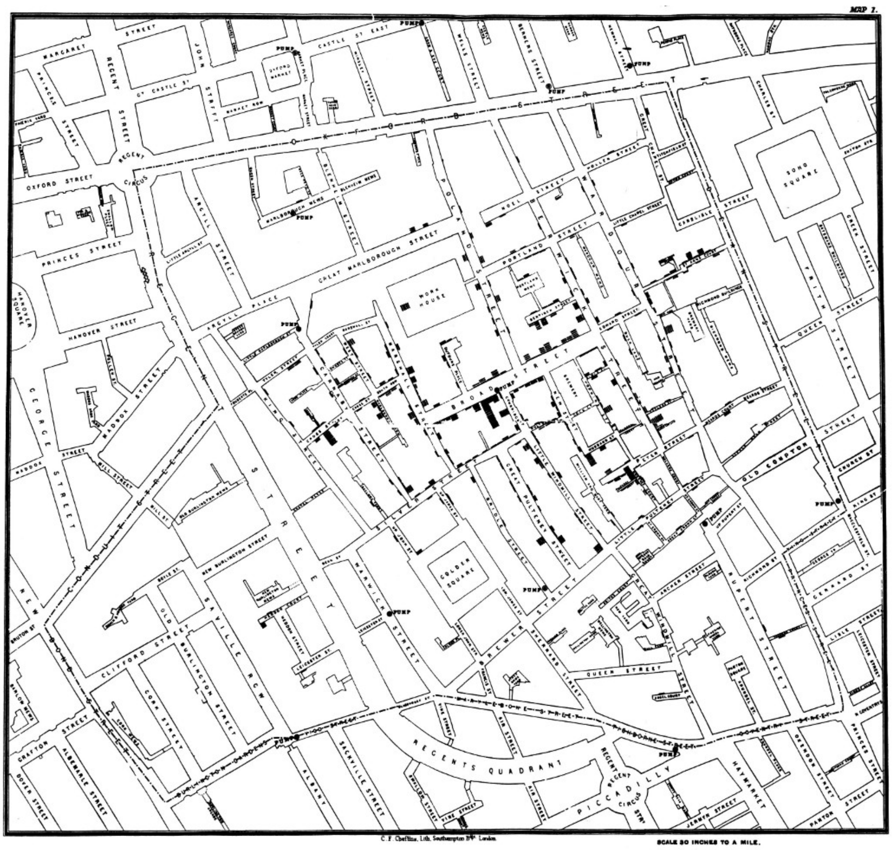


Figure 2.4: John Snow's map of the clusters of cholera cases in the epidemic of 1854

confirmed by the observed covariation: in this area of London, there were few occurrences of cholera exceeding the normal low level, except among those people who drank water from the Broad Street pump" (Tufte, 1997, p. 6). Nowadays it is known that the lack of treatment of human feces and drinking water enables cholera to spread, however John Snow was the first to establish this correlation, decades before the discovery of the bacterium *Vibrio cholerae* which confirmed his theory.

"Snow described his findings to the authorities responsible for the community water supply, the Board of Guardians of St. Jame's Parish, on the evening of September 7, 1854" (Tufte, 1997, p. 6). In consequence the pump-handle was removed and the epidemic soon ended. Moreover, due to Snow's report, which included the map showing the clusters of cholera cases and showed that there was a decrease on the number of deaths once the pump handle was removed, the authorities were convinced that a better sewage system was needed. The map was not the only reason why his report was effective, it was effective because Snow was able to produce and justify a detailed statistical analysis of the event. According to Tufte (1997) Snow's map is so memorable because it places the data in a context for assessing cause and effect. Moreover the map enabled to make comparisons, alternative explanations, and assessment of errors. "Even in the face of issues raised by a modern statistical critique, it remains wonderfully true that John Snow did, after all, show exactly how cholera was transmitted and therefore prevented." (Tufte, 1997, p. 15).

2.1.4 Gapminder – Rosling

In 2006, Swedish scientist Hans Rosling presented at the Technology, Entertainment, Design (TED) Conference, creating a buzz among those interested in information visualization. Rosling made an ambitious promise to the audience: to show the best stats they've ever seen[1]. Hans Rosling is a licensed physician, however he is best known for his background on public health and statistics. That background motivated him to create *Gapminder World*[2] and the Trendalyzer software. Trendalyzer is an information visualization software for animation of statistics developed in 2005 by the Gapminder Foundation (founded by Hans together with Ola Rosling and Anna Rosling Rönnlund).

In his presentation at TED, thanks to Gapminder, Rosling was able to show the relationship of different dimensions, such as income and life expectancy, and also to do comparisons between different countries over the last two centuries. Being able to compare countries that might seem disparate gives us useful insight. For instance, at the presentation Rosling compared data from the United States of America and Vietnam: "1964: America had small families and long lives; Vietnam had large families and short lives. And this is what happens: the data during the war indicates that even with all the death, there was an improvement of life expectancy. By the end of the year, the family planning started in Vietnam and they went for smaller families. And the United States up there is getting longer lives, keeping family size. And in the '80s now, they give up communist planning and they go for market economy, and it moves faster even than social life. And today, we have in Vietnam the same life expectancy and the same

[1] Presentation available at https://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen
[2] http://www.gapminder.org/world

family size here in Vietnam, 2003, as in United States, 1974, by the end of the war" [Video]. According to Rosling, only the data is able to show the tremendous social change that Asia went through and how this change began before the economical change.

Rosling wanted to tell several stories and his visual storytelling was what allowed him to transform the data, which was overlooked by others, into captivating information. Moreover, his way to present the data was also easier to understand. Each country is represented by a circle and its color represents its geographic region: America is yellow, Europe and Central Asia is orange, East Asia and Pacific is red, Middle East and North Africa is green, Sub-Saharan Africa is purple, and South Asia is blue. The size of the circle represents the country's population, the y-axis is the life expectancy at birth, and the x-axis is the fertility rate (births per woman). The forth dimension is time and it is presented through the use of animation. The position of the bubbles in a given intersection of the two axis changes as time passes and the year is shown in the background. "Both the size and locations of bubbles smoothly animate as time passes. This technique appears to be very effective in presentations, where a presenter tells the observer where to focus attention. It makes the data come to life, and emphasizes the critical results of an analysis" (Robertson, Fernandez, et al., 2008, p. 1325). This visualization is one of the first examples of the successful use of animation in a visualization. The inclusion of animation in this bubble chart makes all the difference in terms of storytelling. "When Hans Rosling uses it, he is telling a story about the data and at key points in the presentation primes the observer

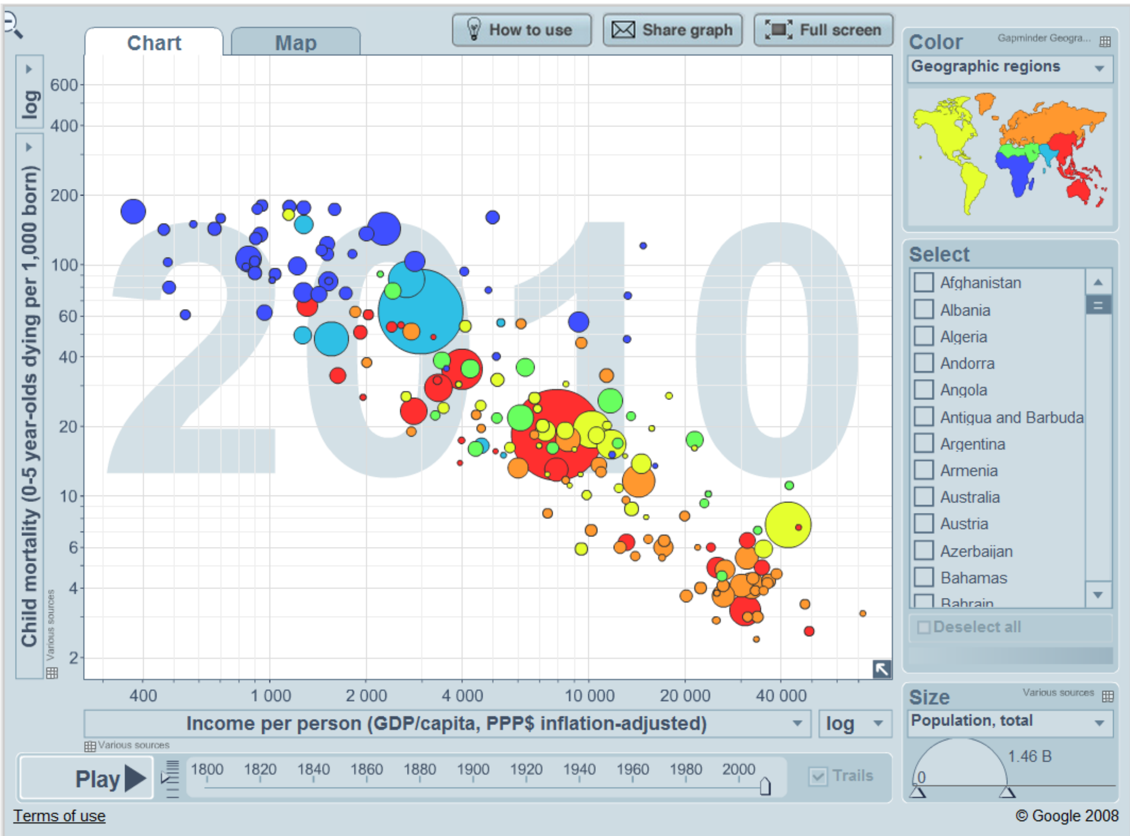


Figure 2.5: Hans Rosling's Trendalyzer, the information visualization software for animation of statistics

to look at a particular part of the bubble chart before some significant event occurs. The effect adds a sense of excitement to the data: the movement of the bubbles becomes a critical part of the story" (Robertson, Fernandez, et al., 2008, p. 1325). For instance, Rosling was able to tell the story, among many others, of how Mao Tse-tung brought health to China, what happened when he died, and how when Deng Xiaoping became paramount leader of China replacing Hua Guofeng in 1978 China became one of the fastest growing economies in the world and its standard of living of hundreds of millions of Chinese rose.

2.1.5 Visualization on the Web

The Web is a thriving environment for visualization. At the end of the 20th century, Gershon, Eick, et al. (1998) predicted that the developments in hardware and software, followed by decreasing prices, and the widespread and more sophisticated use of the Web, were going to drive a wide adoption of visualization by diverse groups. In fact, this prognosis was confirmed and we have witnessed an increase in research and industry produced visualizations, and also in amateur, independently produced visualizations. Moreover, as predicted, the population's visual literacy also increased and developers started to care for HCI issues and investing in producing more user friendly visualizations (Gershon, Eick, et al., 1998). The Web allows more interactivity and flexibility in linking to other sources of information (Rohrer and Swing, 1997), allowing visualizations to be more engaging for both proficient users and beginners.

According to Murray (2013, p. 3) "visualizations aren't truly visual unless they are *seen*" and distribution on the Web is a quick way to reach a global audience. Therefore, visualization creators are getting more and more concerned with the tools that used to make these visualizations available. Until recently most Web-based visualization relied on browser extensions such as Flash and Java (Dörk, 2012), which limited the number of people that could see them. According to Murray (2013) these proprietary software and plugin-ins limit the accessibility of the visualization, which should be available to the widest audience possible. However, lately visualization creators have put a lot of effort into creating visualizations using solutions provided natively by most browsers: "HTML for page content, CSS for aesthetics, JavaScript for interaction, SVG for vector graphics, and so on" (Bostock et al., 2011, p. 2301). JavaScript solutions for instance have increased in popularity and solutions such as Flash are slowly losing its appeal, as can be seen in Table 2.1. Because it runs in widely varying environments (cross browser) and is accessible to people with disabilities, JavaScript has gradually became the de facto standard language for providing rich interactive content in Web pages.

Languages	2011 Jan	2012 Jan	2013 Jan	2014 Jan	2015 Jan	2015 Sep
JavaScript	88.4%	91.1%	92.4%	88.6%	88.1%	90.0%
Flash	28.5%	25.6%	21.1%	15.7%	12.1%	10.1%
Silverlight	0.3%	0.3%	0.2%	0.2%	0.1%	0.1%
Java	0.1%	0.2%	0.2%	0.1%	0.1%	<0.1%

Table 2.1: Trends in the usage of client-side programming languages for websites

Several JavaScript libraries, tailored for visualizations, have been developed in the last five years, the most popular being Data Driven Documents (D3)[3], an open source JavaScript library usually used to generate Scalable Vector Graphics (SVG) graphics, a vector image format long supported by browsers but underutilized. Its popularity might be linked to the fact that SVG graphics perform well in high-resolution displays, and these are now becoming increasingly popular. The reasons for its large user base might be the fact that D3 is an open source project that allows developers to build on top and improve it, and its use of Web standards. "By building on key standards, D3 keeps pace with the evolving technological ecosystem of the web, improving expressiveness and accessibility" (Bostock et al., 2011, p. 2308).

Web-based visualizations even ignited new trends in research. A good example of a type of visualization that blossomed on the Web is what Pousman et al. (2007) named casual information visualization, an emerging subdomain for information visualization research that does not study visualizations tailored for expert users who have previous knowledge and experience. Visualizations that belong to this subdomain are also characterized by having more personal data (less work motivated); having momentary and repeatable usage, or contemplative usage; and supporting insights that are different from the ones in traditional information visualization, that are not as analytical. For Pousman et al. (2007) ambient visualization, social visualization, and artistic information visualization are all types of Casual Information Visualization. However, even though they are far from traditional information visualization research, they are still studied in this research area. A well-known example is *NameVoyager* (Wattenberg, 2005), a web-based visualization of historical trends in baby naming.

Another intrinsic part of Web-based visualizations is sharing and discussing, which according to J. Heer, Viégas, et al. (2009) is an important accompaniment to data visualization, since sensemaking is often a social process. The Web is the perfect environment for collective data analysis because it allows people to easily communicate with each other. Nowadays, whenever a new visualization is published online, hoards of people instantly line up to share and comment, usually in the various social networks. The fact that so many different people, with different backgrounds and cultures, are able to have access to the visualization and share their different interpretations of data makes the analysis of the data much richer.

The Web has also been an invaluable source of data (Dörk, 2012; Rohrer and Swing, 1997). There is an astonishing amount of user-generated data being produced and shared continuously on the Web (across multiple networks and in various formats): people share pictures, review movies, comment on real-time events, support brands, etc. According to Statista (2016), Facebook currently has around 1.59 billion monthly active users and Twitter has circa 320 million active users. In addition, the English Wikipedia includes 5,146,826 articles and in 2015 it averaged 995 new articles per day (Wikipedia, 2016). This data, so large and complex that traditional data processing cannot cope with it, is often referred to by the broad term big data. Even though big data plays an important role in marketing, "the impact of data abundance extends well beyond business" (Lohr, 2012, Website). Lohr (2012) highlights the example of Justin Grimmer, an assistant professor at Stanford, who developed a

[3] https://d3js.org/

system to look for insights into how political ideas are spread by automatically analyzing blog posts, Congressional speeches, press releases, and news articles. According to Jagadish (2015, p. 49) "big data now impacts nearly every aspect of our modern society, including business, government, health care, and research in almost every discipline: life sciences, engineering, natural sciences, art & humanities". Big data encompasses several challenges, and although it may not seem so, dealing with the computational processing of the data is not the main one (Jagadish, 2015). Although capture, storage, transfer, and even search and analysis are issues that still must be tackled (specially since the amount of data keeps increasing and there is still room for improvement in managing real-time data), the biggest problem concerns human interaction, because "human ability does not scale" (Jagadish, 2015, p. 50).

Visualization can be used to ease this problem and possibly help humans identify patterns in the data and consequently help to acquire relevant insights. However, simply using visualization is not the answer to the problem of the amount of data. Jagadish (2015) warns that large volumes of data can be hard to plot in a way that is easy for a human to understand. Several developments in the area of HCI have been made in order to better visually represent large sets of data, however, according to Dörk (2012) ensuring responsiveness and interactivity when datasets are distributed across the Web still constitutes a major challenge.

2.1.6 Wikileaks and open government data

The accessibility of the information is key for enabling citizen input into decision-making, and this trend of information accessibility is inviting citizens to participate in the public sphere. According to Dahlgren (2005), the Internet is a major development in the contemporary history of Western democracy. Without well-informed citizens there is no democracy, and citizens cannot be informed if there is no transparency in the data. Data transparency also promotes accountability, because it allows citizens to control their government, preventing corruption. "Opening up government data enables the citizens to learn about the activities of their government, to hold their government accountable for its actions and its spending and to participate in the political process" (Janssen, 2011, p. 446).

However, having the data available is not enough to promote this culture of participation and transparency. If the data is not analyzed it is almost impossible for citizens to make sense of it. Although the popularization of tools such as spreadsheets and visualization tools have enabled most people to take raw data and make sense of it (Rogers, 2013), data curation is still necessary. This is where journalism comes into play. The data journalism trend reflects this movement that is spreading across the globe and this intent to make sense of the available data. According to Rogers (2013, Ebook), four factors were propitious to this outcome:

- "the widespread availability of data via the internet;"
- "easy-to-use spreadsheet packages on every home computer;"
- "a growing interest in visualizing data, to make it easier to understand;"
- "huge news stories that would not have existed without the statistics behind them."

According to Rogers (2013), it would be impossible to report complex stories such as the Wikileaks releases on Afghanistan, Iraq, and the US embassy cables without resorting to math and visualization tools. Although data journalism (or computer-assisted reporting) already existed (Gray et al., 2012) and was used extensively in the reporting of elections, the coverage of Wikileaks was a challenge for news media. For the first time it was completely impossible to make sense and to present these complex data sets to the readers without the use of illustrations and interactive visualizations.

The Wikileaks data, together with the open data movement, were two game changers that fueled the worldwide demand for the intelligibility of the available data. With some differences between them, Wikileaks and Open Data are both about transparency and democracy. Wikileaks is focused on the accountability of errors made by governments and the awakening of citizens to these errors. Moreover, the legitimacy of the release of several confidential documents is questionable and Wikileaks legal status is complex. Open data, on the other hand, is about prevention of errors and abuses. It is about allowing citizens to collaborate with government officials and helping them make decisions. Furthermore, the data released is previously curated before it is released and the organizations usually make sure that the released data will not harm anyone. For instance, the data released rarely contains any personal information that would allow individuals to be identified. Nowadays, several national governments distribute some of their data online[4] and, since the 1980s, the European Commission has been pushing the disclosure of some data that could benefit citizens if it becomes more accessible (Janssen, 2011, p. 446). In fact, in 2003, the European Parliament and Council proposed the adoption of Directive 2003/98/EC on the re-use of public sector information and, in 2010, the European Commission reemphasized in its *Digital Agenda* the importance of supplying the data.

As Janssen (2011, p. 446) states, "these data are indispensable for public policy development and service delivery, but they are also very valuable to citizens, organizations, and businesses for public participation, for decision-making, and for creating innovative products and services." This improved access to information is also boosting the information/data/-knowledge visualization trend. Although big data and data journalism are not graphics and visualizations, as Rogers (2013) states, these play a big part in helping making sense of and representing the data. Visualization is proficient in presenting large sets of data, allowing comparisons, while also being appealing for the users.

2.2 The present and future of visualization

Over the last years the popularity of visualization has exploded. According to Dörk (2012, p. 1), the Web "has become a vast information space containing rich content and semantic relationships, and a platform for sophisticated interactivity and graphics." And these visualizations are not exclusively produced by academics or by the media. They come from every area, for instance marketing, health, and some are not even professionally produced. "It's a time when the mainstream media is embracing sophisticated techniques born in university research labs — a time when you can open *The New York Times* and see complex treemaps and

[4] http://datacatalogs.org

network diagrams" (Viégas and Wattenberg, 2008, p. 49). These visualizations that originate outside of the research community are often called vernacular, "in a nod to Tibor Kalman's admiration of *low* art" (Viégas and Wattenberg, 2008, p. 49). According to Dörk, Feng, et al. (2013, p. 2192), "the rise of casual and vernacular visualization is a testament to the growing significance of visualization beyond professional confines."

The field of Information/Data/Knowledge visualization was for a very long time locked away from the public eye, only making an occasional appearance during elections and more data heavy events. However, with this visualization boom, people became fascinated with visualizations, in such a way that they share them on social media all the time. Their interest is not surprising. The brain is sensitive to visual stimuli and since birth humans are more than accustomed to having a huge amount of data coming into the brain through the eyes. At roughly the rate of an Ethernet connection (Koch et al., 2006), the human retina is able to transmit data to the brain. Moreover, the brain has the ability to rapidly process that data, see it in context, recognize patterns, make comparisons, and make sense of it (J. Heer, Viégas, et al., 2009), which is one of the factors that makes humans receptive to information visualization. However, this ability of aiding sensemaking is a social process (J. Heer, Viégas, et al., 2009) and therefore the same data set, visualized in the same way, can have multiple interpretations.

This growing fascination with information visualization intensified the demand for new visualizations, which consequently attracted more people to the field, particularly from the design, art, and journalism communities. The impact of these different communities in the field of information visualization was very positive and raised awareness to the need for visualizations to be more visually appealing. However, this also made the number of visualizations that fail in terms of functionality rise, which is one of the major concerns of this research area. Although C. Chen (2005, p. 15) identifies aesthetics as one of the *Top 10 Unsolved Information Visualization Problems*, he clarifies that "the purpose of information visualization is the insights into the data that it provides, not just pretty pictures". And this is one of the problems that arises with the blurred lines between information visualization and what is often called information art. The problem begins with the concept of aesthetics: "a concept that relates to the beauty in both nature and art" (Cawthon and Vande Moere, 2007, p. 637), that often implies judgments of sentiment and taste. Therefore it is subjective and varies between cultures.

Even though, aesthetics has been approached in several fields that regard functionality over beauty, such as architecture, user experience and HCI (the last two being closer to the information visualization area). The objective of adding aesthetics to these fields is "to stimulate the desire, positively influence the first impression, encourage repeated usage or even overwhelm its audience" (Cawthon and Vande Moere, 2007, p. 637). According to C. Chen (2005), to create visually appealing visualizations it is important to find a balance between aesthetics and insights. However, and because it is difficult to research this topic, there are not enough studies that reveal (or at least shed sufficient light on) what are the visual properties that enable users to characterize a visualization as visually appealing.

Having a deep understanding of the differences between information and information art (or, as Kosara (2007) discerns, between pragmatic and artistic visualization) might be useful in terms of research. According to Dörk, Feng, et al. (2013, p. 2191), "artistic visualizations often involve personal experiences, individual opinions, and the context of the viewing experience in the interpretation", and several of its specificities could be explored.

The dissemination of information visualization and this introduction of aspects from different fields of visual representation has sparked the need for principles, guidelines, and methodologies. In order to produce better visualizations, it is imperative to have in mind some key guidelines. According to Gershon, Eick, et al. (1998), the evolution of a discipline consists of four stages: it initially emerges as a craft practiced by practitioners, then the researchers start analyzing processes in order to develop principles and theories, after that engineers (or in this case visualization producers) refine the principles and give insights about the creation process, and finally the technology is spread and becomes widely available. However, in the area of information visualization everything evolved simultaneously and, although most of the principles of this area are often referenced in information visualization literature, there is not yet a clear set of standard principles. Having sound design principles will be vital for the further development of the field.

2.2.1 Principles borrowed from other areas

The principles that are often referenced in the literature are usually borrowed from other areas such as the study of perception in Gestalt psychology (Nesbitt and Friedrich, 2002; Ware, 2004) and HCI (Dykes et al., 2010; I. Herman et al., 2000; Sears and Jacko, 2009). Gestaltism (from the German word *Gestalt* which means form) is a theory developed in the 1900s that, among other things, deals with the study of perception. According to the *Gestalt Principles of Organization* there are factors that impact the perception of form and on how parts are grouped into structural forms. These principles concern the process of humans identifying patterns in visual displays. According to Ware (2004), these laws of pattern perception easily translate to the domain of information displays. The principles that Ware (2004) mentions are proximity, similarity, connectedness, continuity, symmetry, closure, relative size, and common fate. Nesbitt and Friedrich (2002), on the other hand, leave out the laws of symmetry, closure and relative size but add the laws of simplicity and familiarity to their list of Gestalt principles that can be applied to animated visualizations of network data. The principles cited by Ware (2004) and Nesbitt and Friedrich (2002) can be seen in Figure 2.6.

Some of the Gestalt laws of pattern perception are also borrowed from the area of HCI.

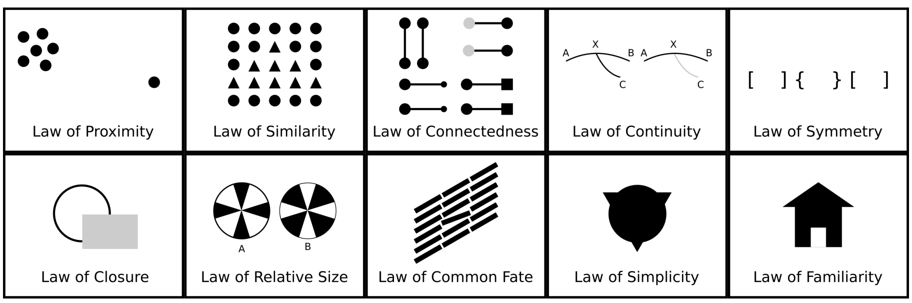
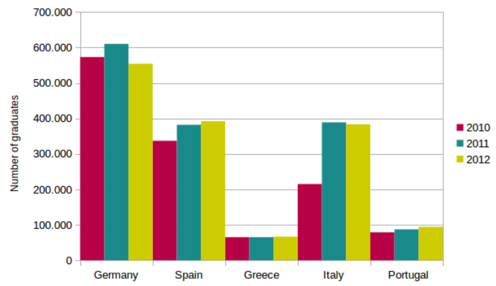
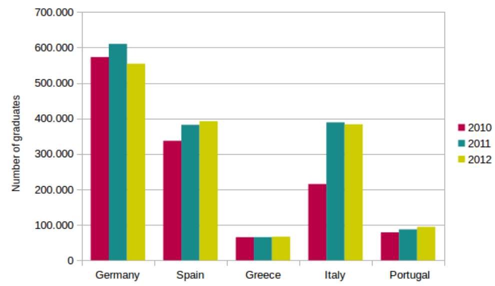


Figure 2.6: Gestalt laws referenced by Ware (2004) and Nesbitt and Friedrich (2002)



(a) Bar chart on which the bars are not grouped



(b) Bar chart with bars grouped by country

Figure 2.7: Example of proximity in a bar chart

According to (Sears and Jacko, 2009) the HCI principles of visual communication are harmony (the logical arrangement of similar or related elements), balance (the equilibrium between all of the elements) and simplicity (the elimination of every unnecessary decoration in order to provide clarity and eliminate ambiguity). The principle of balance aligns with the Gestalt law of symmetry. The principle of harmony by definition is similar to the laws of proximity and similarity, but is closer to the Gestalt law of unity, which is related to the quality of congruity (the arrangement between the elements that look as through they belong together). The law of simplicity is also referenced in the Gestalt theory, and follows the same guidelines as the HCI principle of simplicity. Neither the principle of unity or of simplicity however is part of the list of laws that Nesbitt and Friedrich (2002) referenced as useful for information displays.

Law of Proximity

The law of proximity states that viewers instinctively assume that items that are closer are related, and that items that are further apart are not related. Items that are closer to each other are perceived as being aggregated in the same group, even if the shapes and sizes are radically different. Humans tend to do this because it is easier and faster to process few sets that contain several elements than to process a large number of smaller stimuli. This means that, in a visualization, this principle of proximity can be used as a visual metaphor to let the viewer know that there is something in common between those items that are closer, as seen in Figure 2.7b. Since there is no difference in terms of grouping in the example shown in Figure 2.7a the viewer barely realizes that there are three bars for each country.

Law of Similarity

According to the principle of similarity, stimuli that physically resemble each other will be perceived as belonging to the same group. Therefore, when observing the second panel of Figure 2.6, the viewer will be able to rapidly identify the shape of a triangle, formed of nine smaller triangles, inside the square, because they perceive the smaller triangles as part of a whole. The similarity between items can be in color, shape, orientation, etc.

This principle in visualization can be used to draw the viewers' attention or to help him/her make interpretations. For instance, in a flowchart, which is a type of diagram typically used to represent a work-flow or a process, the different shapes of the boxes are used to represent

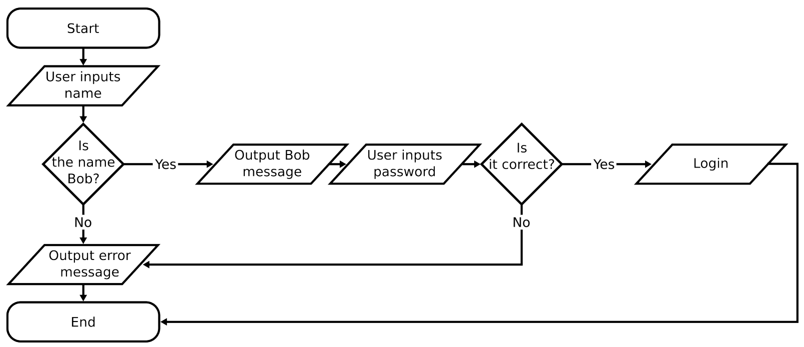


Figure 2.8: Example of similarity and connectedness in a flowchart

specific actions. Therefore, when the viewer sees the flowchart he/she instinctively recognizes that two boxes with the same shape are the same type of action. This allows the viewer to better understand the visualization. For example, in Figure 2.8 the viewer understands that the diamond shapes are conditions or decisions, that the parallelogram represents input/output actions, and that the rounded corner rectangles are generic processing steps.

Law of Connectedness

Although it was overlooked by the Gestalt psychologists, connectedness is one of the fundamental principles, and is often more powerful than proximity, color, size, or shape (Figure 2.6, panel three). This principle is based on the fact that having a line connecting items is a powerful way of expressing that these have a relationship. According to Ware (2004), this principle is specially relevant for visualizations such as node-link diagrams. If a box is drawn around some elements this is also interpreted as connectedness. In Figure 2.8 the arrows help the viewer establish a link between the items that compose the flowchart.

Law of Continuity

According to the Gestalt principle of continuity, humans tend to perceive items as singular, uninterrupted entities and this fact allows the differentiation of the stimuli even when

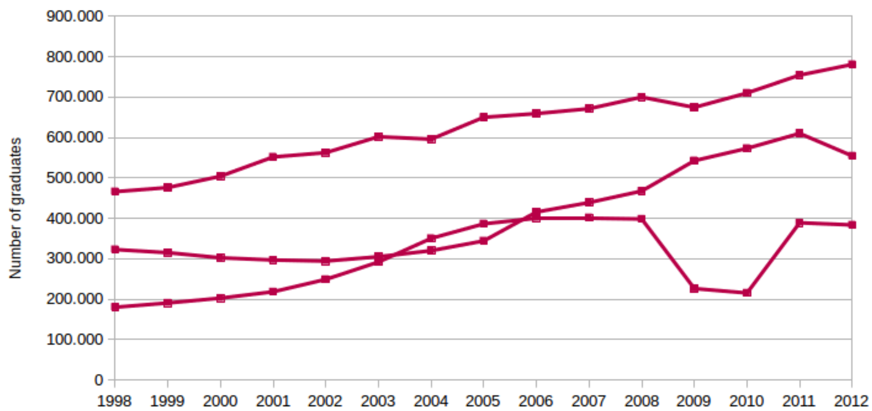


Figure 2.9: Example of continuity in a line chart

they overlap. As seen in the forth panel of Figure 2.6, the arc that connects point A to point B is perceived as a continuous line and the arc from point X to point C as another. The viewer does not consider the line started in point A to finish in point C or even in point X. The viewer also does not consider the item from B to C to be a continuous line. Having the segment XC colored differently is not necessary to maintain the interpretation cited above. The principle also applies if the item consists of various elements arranged along lines.

Due to the principle of continuity, in Figure 2.9, the viewer is able to understand the direction followed by the first and second lines even though they intersect. The viewer is able not only to identify the orientation of each line but also perceive each line as a whole. This happens because humans have a tendency to follow lines, curves, and other forms even if they have abrupt changes in direction. However it is easier to interpret the line chart if the lines have different colors or shapes, and this is the standard procedure in creating line chart.

Law of Symmetry

The principle of symmetry states that an item will be perceived as incomplete if it is unbalanced or is not symmetrical. This principle is also often referred to as the law of balance because it is achieved when the visual weight is evenly placed on each side of an axis. According to Ware (2004), symmetry is a powerful organizing principle because symmetrically arranged pairs are perceived more strongly as forming a visual whole.

Therefore, even when two symmetrical elements, such as the brackets in the fifth panel of Figure 2.6, are apart the viewer is able to connect them in order to form a coherent shape. This is why the viewer will never perceive them as four square brackets and two curly brackets but as two pairs of symmetrical square brackets and a pair of curly brackets. The viewer will also rarely interpret it as being a pair of square bracket with two individual square brackets and a pair of curly brackets in between.

Ware (2004, p. 192) states that "a possible application of symmetry is in tasks in which data analysts are looking for similarities between two different sets of time-series data." In the

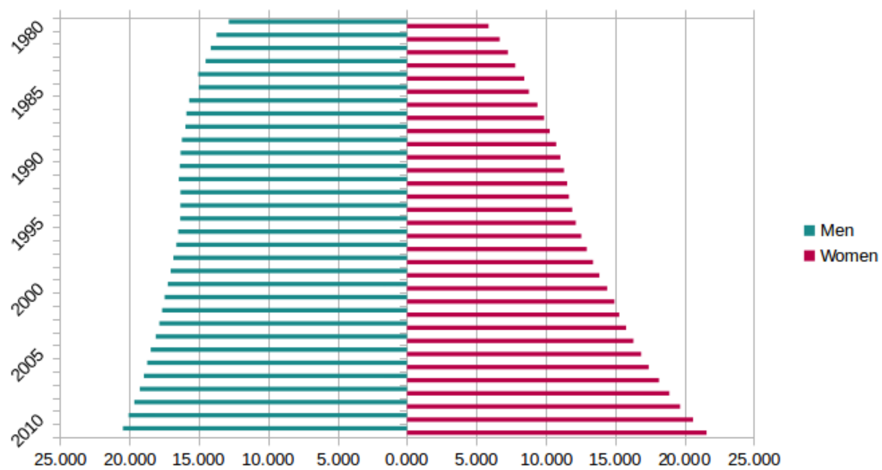


Figure 2.10: Example of symmetry in a pyramid bar chart

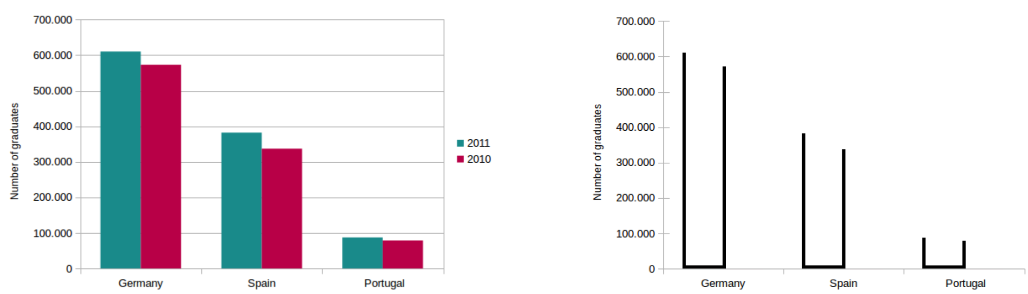
example shown in Figure 2.10, the male population data are plotted to the left (on apparent negative x values) and the female population data to the right, creating a pyramid–like shape. Having the data displayed in a pyramid barchart allows viewers to perceive similarities that would be harder to understand if the data was represented, for example, in parallel plots.

Law of Closure

The Gestalt law of closure states that humans perceive open shapes as incomplete shapes and consequently tend to close the gaps, interpreting the item as a if it was a complete form. When the item is partially hidden behind another, the viewer will also perceive that item as a complete item, as can be seen in panel six of Figure 2.6. This happens because the human mind has a natural tendency to recognize familiar patterns. Similarly, a human will also be able to complete a phrase that is missing some letter if the words are familiar and will be able to complete a sequence of numbers where one or more are missing. For example, if this word *perc_ption* is presented to the user he/she would be able to understand that the word presented is perception; and if the user is presented with the sequence *10 12 _ 16 18* he/she will instantly see that the missing number is *14*.

This principle is present in the redesign of the bar chart by Tufte (1983). According to Tufte (1983) there is much that can be erased in a traditional bar chart without compromising the data. In this minimalist bar chart, which can be seen in Figure 2.11b, the bars were replaced by vertical lines. A horizontal line connects each pair, maintaining the link that allows viewers to understand them as a group, as in Figure 2.11a. In the minimalist version the only lines kept were the x and y axis, instead of the full frame. The grid, used to aid in the visual alignment of data, was also removed.

Tufte (1983) advocates that although decorations can be used to help editorialize, they should never distort the data and therefore most ink that is not used to present data should be removed. The problem is that people are not familiar with these minimalist charts and although they are able to fill in the gaps they have to make a bigger effort to interpret these. The traditional bar chart on the other hand is very familiar and therefore simple to interpret. Moreover, there are several examples of the successful use of decorations in information charts. Bateman et al. (2010) found in a previous study that embellished charts were understood just



(a) Traditional bar chart (b) The same bar chart in Tufte's minimalist format

Figure 2.11: Example of closure in a bar chart

as accurately as plain charts. Nevertheless, participants recalled embellished charts significantly better and preferred them to plain charts.

Law of Relative Size

According to the principle of relative size, the viewer tends to perceive the smaller components of a pattern as objects. Therefore, as shown in the seventh panel of Figure 2.6, the viewer identifies the pattern in A as being a black propeller in a white background. The effect is the same if the colors are inverted (B), meaning that the color is not responsible for this perception. According to the law of relative size patterns that are oriented vertically or horizontally are easier to perceive as objects Ware (2004).

Law of Common Fate

The principle of common fate (represented in the eighth panel of Figure 2.6 and in Figure 2.12) declares that the viewer will tend to group together items that appear to be moving, at the same rate, in the same direction. This effect of interpreting the items as the same stimulus when they move in the same direction is maintained even when the shape of the items is not the same. It is possible to see this effect when observing flocks of birds. A single flock of birds is perceived as an unified whole because every element of the flock moves in the same direction at the same velocity. And even when a second flock intersects the first one, the viewer is still able to distinguish the two as separate flocks because each element moves in the direction common to its group.

This principle is particular important in user-interface design. One of the most well known examples of the gestalt principle of common fate in web design, for instance, is the cascading drop-down menu: it provides additional layers of navigation by presenting a sub-menu next to the entry of the main menu that the user clicked or hovered, fitting more navigation options onto the screen without obstructing the view of the main menu. When presented with this kind of menu the user instinctively knows where the second menu option comes from and perceives it as part of the option he/she selected from the main menu.

Consequently it is also very useful in information visualization. The Gestalt principle of common fate applies to line charts, on which the lines that seem to be heading in the same direction are instinctively grouped by the user and the lines that move in a different direction are perceived as unrelated to the general trend of the group. Similarly, Hans Rosling takes advantage of this principle to show trends in his TED presentation. Using movement he is

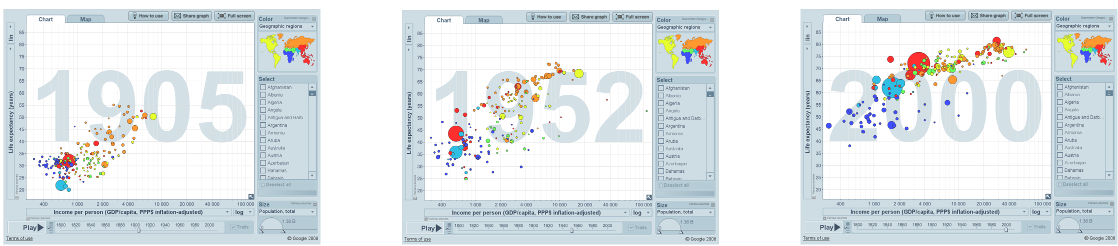


Figure 2.12: Example of common fate in *Trendalyzer*

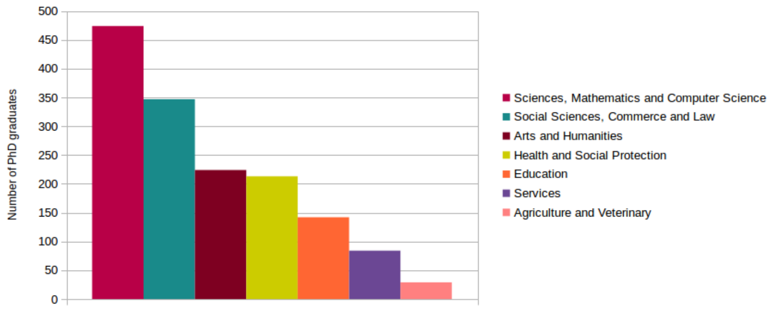


Figure 2.13: Example of simplicity in a bar chart

able to show which countries do not follow the trend of increase in life expectancy conjugated with the increase of income. He is also able to spot the outliers.

Law of Simplicity

The law of simplicity is often referred to as the law of *Prägnanz*, which in German means *good figure* and for that this principle is also often referred to as the principle of good continuation. This principle states that humans tend to perceive stimuli in a the simplest way possible. For instance, when presented with a complex structure such as the one present in the ninth panel of Figure 2.6, viewers tend to perceive the structure as two simpler shapes superimposed (a triangle and a circle overlapping) and not as a whole more complex polygon. If the object is already simplified the interpretation will be easier due to this unconscious ability to simplify things. However the process of simplification also occurs in complex scenarios.

As seen in Figure 2.13, unless there is a specific order for the bars (in ascending or descending numeric order, alphabetic order, etc.) they should be organized in a way that is visually simpler. However, it is important to make sure that the simplification process does not lead to unintended conclusions. For a human it is easier to perceive elements arranged on a line or a curve. Accordingly, the bar chart presented above (organized from lowest to highest) will be easier to read because it suggests a continuous line. Therefore, to allow a more intuitive and rapid interpretation, one should always seek to simplify visualizations and arrange the data logically and systematically.

Law of Familiarity

According to the Gestalt law of familiarity, humans perceive structures that appear familiar and tend to form groups if the items feel familiar or meaningful. The viewer's interpretation of a stimuli will always be influenced by his/her prior experiences.

For instance, in the last panel of Figure 2.6, the viewer will hardly see two squares and a triangle, he/she will perceive it as a house, and depending on the context will add extra interpretations: a home, a family, etc. The individual elements will be perceived as a group, because the grouping is more meaningful, creating a more recognizable form than the parts individually. The human eye searches constantly for significance even in random or vague stimuli. This extreme search for familiarity is known as *pareidolia*: the tendency to perceive a

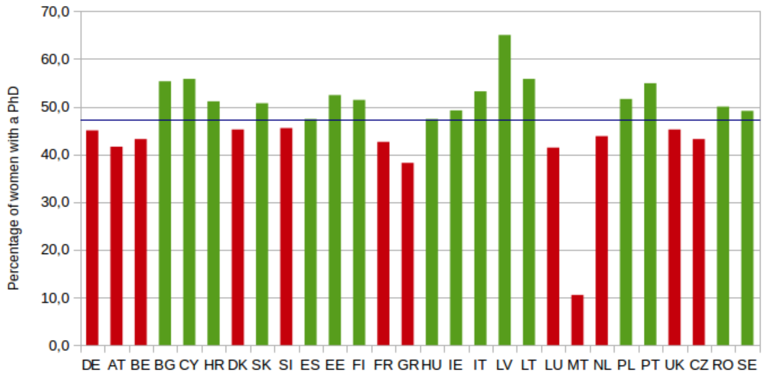


Figure 2.14: Example of familiarity in a bar chart

familiar and meaningful pattern in something where none actually exists, for instance seeing a face on the Moon or the face of Jesus on burnt toast, hearing human voices in static noise, or identifying shapes in the clouds or smoke.

Visualization can take advantage of the users search for familiarity, but not to this extreme. An example of the use of familiarity in visualization can be seen in Figure 2.14. In this bar chart the colors are used to convey a message that is familiar to many cultures: green means good and red means bad. The countries that are above average are colored green (positive meaning) and the ones that are below average are colored red (negative meaning). Using colors to convey positivity and negativity is a popular convention. Another example of the use of color to convey familiarity can be seen in Figure 2.10, in which the color pink is used to represent women and the color blue to represent men. A lot of color meaning arises from past experience and is deeply influenced by culture.

2.2.2 The importance of the insights and knowledge

Being able to translate information into knowledge is one of the core goals of information visualization. Knowledge is the circumstance or condition of apprehending truth or fact through reasoning and can be acquired through experience or education. Having knowledge as an ultimate goal supports the assumption that visualization facilitates understanding and cognition (the process of acquiring knowledge through thought, experience, and senses).

Closely linked to knowledge is the idea of *insight*, which can be defined as the capacity to gain an accurate and deep understanding of something or someone. *Insight* has been commonly stated as one of the purposes of information visualization, if not the main purpose, by many authors (Card, Mackinlay, and Shneiderman, 1999; C. Chen, 2005; Spence, 2001; Wijk, 2005; Yi et al., 2008). For C. Chen (2010) it is the holy grail of information visualization and for Card, Mackinlay, and Shneiderman (1999) it is the true purpose of visualization (not pictures as some may think). The main goal of these insights is "discovery, decision-making, and explanation" (Card, Mackinlay, and Shneiderman, 1999, p. 6).

North (2006) characterizes insights in information as being complex, qualitative, deep, unexpected, and relevant. Generally, the process of acquiring insight is already complex, but in information visualization this complexity is increased by the fact that the amount of data from which to extract insights is usually large. Moreover, some insights might require the exploration of most or even the entirety of the dataset. Due to the amount of data, the process of obtaining insights in visualizations is also hampered because insights are uncertain and subjective, and built over time, "accumulating and building on itself to create depth" (North, 2006, p. 6). However, "insight is deeply embedded in the data domain, connecting the data to existing domain knowledge and giving it relevant meaning" (North, 2006, p. 6). Therefore, the amount of data is both a challenge and a strength.

The process of gaining insight

In order to shed light into how visualizations can enable users to gain insights that will turn into knowledge, Yi et al. (2008) present four largely distinctive processes of gaining insight often discussed in information visualization literature: Provide Overview, Adjust, Detect Pattern, and Match Mental Model. Although these are not categories of interaction or presentation techniques, these are processes that can be originated by using particular interaction and presentation techniques.

When a visualization provides *overview* features, it allows the users to understand the big picture of a dataset. Providing overview is the first stage of the *Visual Information Seeking Mantra* (Shneiderman, 1996) and although it does not guarantee that the user will gain insights by being able to get an overview of the data in the first place, it seems that it enables the user to find portions of the data that he/she needs to investigate more, "thereby promoting further exploration of the dataset" (Yi et al., 2008, p. 3). It was once believed that the ability to overview the data was vital to discover the connections and patterns in the data. However, more recently, it has been found that features that allow the user to interact with the data are more effective in increasing the will to search for insight.

In opposition, the process of *Adjusting* refers to the process of tuning the level of abstraction and/or the range of selection in order to enable the user to change his/her perception. There are several different strategies to enable the user to adjust the view. The two interaction techniques described by Yi et al. (2008) are filtering and *grouping*. The first is also part of the *Visual Information Seeking Mantra*. Filtering allows the user to explore a large dataset by reducing the amount of information he/she has to deal with, choosing the data that he/she is interested in and excluding the rest permanently or temporarily. "By allowing users to control the contents of the display, users can quickly focus on their interests by eliminating unwanted items" (Shneiderman, 1996, p. 339). The second allows the user to explore the data by abstracting the dataset into more manageable pieces. It is difficult to understand a large, unorganized dataset, but by grouping and aggregating we can reduce users' search and working memory load (Yi et al., 2008). The information should be gathered, simplified, organized, and labeled.

Zooming, although not in the set of interaction techniques presented by Yi et al. (2008), is also an adjusting technique. It is included in the *Visual Information Seeking Mantra* and consists of allowing the user to focus on a portion of a collection that he/she is interested in. According to Shneiderman (1996, p. 339), zooming, which could be done "on one dimension

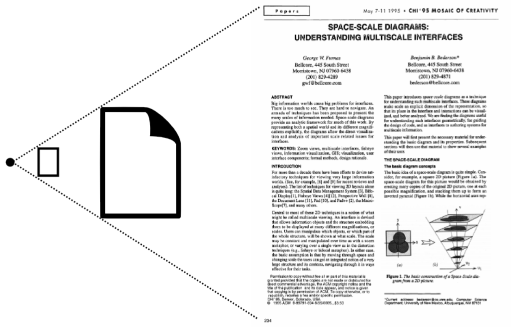


Figure 2.15: Example of Semantic Zoom

at a time by moving the zoom bar controls or by adjusting the size of the field-of -view box", prevents the users getting lost by preserving their sense of position and context. More recently information visualization and visual analytics researchers have also been studying another zooming technique called *semantic zooming*. This more complex and less common zooming technique differs from geometric zoom, with which the objects are simply magnified, by providing a different visual representation of the objects depending on the amount of real estate available for them to be displayed (Büring and Reiterer, 2005; Furnas and Bederson, 1995). According to Perlin and Fox (1993, p. 58), "as the magnification of an object changes, the user generally finds it useful to see different types of information about that object", and this is not possible with geometric zooming where objects change only in size and not in shape when magnified. However, with semantic zooming an object could be represented, as it can be seen in figure 2.15, as a dot in a broader view, and as the user zooms in it could be represented as a rectangle, then as an icon for a document and finally as a page of text. Or, for instance, in a broader view a text document could be represented with just its title, but if magnified, that view may be augmented into a short summary or outline, and at some point the entire text would be revealed (Perlin and Fox, 1993).

The third process identified by Yi et al. (2008) is *detect pattern* and consists of finding specific distributions, trends, frequencies, outliers, or structures in the dataset. Moreover, detect pattern techniques can also allow users to detect relationships, trade offs or anomalies. According to Yi et al. (2008) a pattern itself could be an insight.

Match mental model is the last process. The ability that a visual representation of data has of decreasing the gap between the actual data and user's mental model of it is one of the recognized benefits of information visualization. It helps by reducing cognitive load, amplifying human cognition of familiar presences, and linking the presented visual information with real-world knowledge. Metaphors facilitate more effective mapping of data.

Visualization literacy and evaluation

"Ultimately InfoVis is about harnessing human's remarkable visual perception capabilities to help identify trends, patterns, and unusual occurrences in datasets"(Yi et al., 2008, p. 1). Therefore the process of getting insights is closely linked to the user's visualization literacy. According to Boy, Rensink, et al. (2014, p. 1964) *visualization literacy* is "the ability to confidently use a given data visualization to translate questions specified in the data domain into visual queries in the visual domain, as well as interpreting visual patterns in the visual domain as properties in the data domain." However this ability can be impaired by the user's perceptual abilities, knowledge of the data, value assigned to various insights, and what he/she is willing to spend in terms of time and effort (Wijk, 2005).

Recently researchers have put a lot of effort in developing methodologies for accessing visualization literacy (Boy, Rensink, et al., 2014) and several visualization evaluation studies have been published (Burkhard, 2005; J. Heer, Mackinlay, et al., 2008; Kapler and Wright, 2005; Plaisant, Grosjean, et al., 2002; Pousman et al., 2007; Shi et al., 2005; Willett et al., 2011). However, there is still little information on how to effectively measure insights and how to evaluate the value of a visualization. The struggle to measure how much insight is acquired is not the only obstacle for visualization evaluation. Assessing of the value of the insights acquired is also an issue (Wijk, 2005).

Even though evaluation in information visualization is no longer only focused on measuring time spent on a task or in other usability and efficiency metrics, there are still several problems that are often encountered. Kim et al. (2014) state that the most common types of failures in evaluating visualizations are:

- confounding factors (this could be anything from visual or interaction elements to the amount of data shown which adversely affects the factors being studied);
- insincere participants;
- obsession with the p-Value (in other words, blindly applying statistical hypotheses tests seeking for statistical significance);
- the search for a unicorn ("searching for perfection in tasks, datasets, participants, baseline comparisons, and so on can be akin to finding a unicorn, the legendary horned horse from mythology" (Kim et al., 2014, p. 144));
- boring dominance (which is when having the *perfect* evaluation study might be a sign of failure of evaluation methodology).

In the *Visual Analytics research agenda* (Cook and Thomas, 2005), three levels of evaluation are described: component, system, and work environment. The specificities of evaluation efforts targeted at these different levels can be seen in figure 2.16.

At the *component level* Cook and Thomas (2005) put the evaluation studies that aim at determining the benefits of visual representations, interaction techniques, and overall designs. The metrics used are the standard ones for usability evaluation, such as effectiveness (for instance, the number of errors committed on a task or the number of incomplete tasks), efficiency (for instance, the time taken to complete a task), and user satisfaction. According to Plaisant, Fekete, et al. (2008, p. 121) "data analysis algorithms can often be evaluated with metrics that can

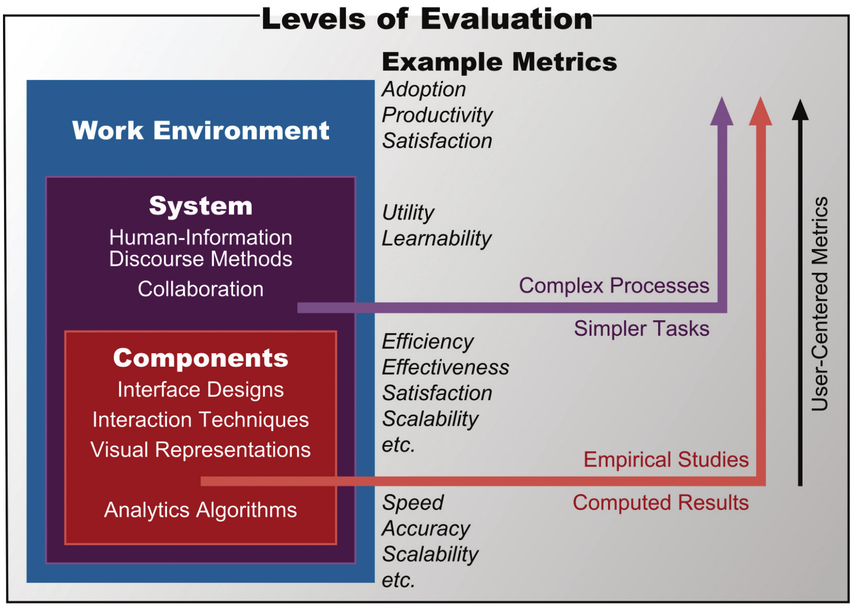


Figure 2.16: Evaluation levels for Visual Analytics (Figure 6.1 (Cook and Thomas, 2005))

be observed or computed (e.g. speed or accuracy), while other components require empirical user evaluation to determine their benefits."

Evaluation studies focused at the *system level* are concerned with the combination and integration of multiple components and often compare the proposed system with the technology currently used by target users. In addition to user satisfaction, metrics need to address the learnability and utility of the system, and the "evaluations often take place in the laboratory using surrogate scenarios but addressing complex tasks conducted over longer periods of time than component-level evaluations" (Cook and Thomas, 2005, p. 153). Insight-based evaluations are also included at the system level, where participants are given several months to explore the data set and are then asked to report on the insights gained.

Last but not least is the level where evaluation studies concerned with the *work environment* are concentrated. At this level Cook and Thomas (2005) put the evaluation studies that address issues influencing technology adoption, and metrics can include adoption rate, trust, and productivity. For this level of evaluation, case studies and ethnographic studies are very useful. However, researchers in information visualization and analytics rarely opt for these kinds of studies because they are time-consuming and may not be easy to replicate or generalize to other application domains.

As can be seen from figure 2.16, user-centered metrics can and should be applied across all levels. Cook and Thomas (2005) add that usability evaluation remains a cornerstone of user-centered design and usability engineering, however, it cannot by it self provide sufficient insight to guide visual analytics research in promising directions.

Instead of identifying the failures of the evaluation studies in the field, as Kim et al. (2014) do, Cook and Thomas (2005) decided to approach some of the issues evaluating visual analytics as challenges. Cook and Thomas (2005) separate the component level challenges from the system level challenges. *At the component level*, the main challenge is to change the evaluation efforts from isolated evaluations into ones that can be turned into new guidelines for the field. Due to the fact that there are myriads of different approaches to visualize data it is important to achieve guidelines that can help visualization creators selecting techniques based on the tasks and data characteristics. According to Cook and Thomas (2005, p. 154), "studies would be aided by the development of comprehensive task taxonomies and benchmark repositories of datasets, tasks, and results."

Another problem at this level is the fact that most studies are concerned with evaluating simple tasks, for instance asking participants to *locate* or to *identify* things on the visualization. Tasks requiring users to compare, associate, distinguish, rank, cluster, correlate, or categorize are rarely approached because the studies that evaluate complex tasks are difficult to design. Moreover, the analysis process of the components of visual analytics is rarely an isolated, short-term process, which is for itself a considerable challenge. Recruiting subjects for longer periods of time can be very difficult.

At the system level, Cook and Thomas (2005) characterize evaluation as a daunting challenge. Due to its complexity system-level tasks are difficult to emulate in a laboratory environment. Besides, the results will also be affected by the users' motivation and expertise, which likewise are difficult to predict and to be controlled for the experimental results to be accurate. According to Cook and Thomas (2005), using domain experts will lead to more realistic results, yet if the system being evaluated is not intended to be used by experts this might lead to inaccurate results that will not be useful. Furthermore it is often difficult to gain access to domain experts for extended periods of time.

Similarly to what happens at the component level, at the system level is hard to deal with studies that would require users to look at the same data from different perspectives and over a long period of time. Considering that "discovery is seldom an instantaneous event but instead requires the study and manipulation of data repetitively from multiple perspectives and possibly using multiple tools" (Cook and Thomas, 2005, p. 155), the challenge of time is one of the toughest in evaluation studies that aim to measure insights gained with a visualization.

In order to shed new light on possible improvements in the evaluation of visualizations, Plaisant, Fekete, et al. (2008) mention some lessons learned from three contests: InfoVis 2003, InfoVis 2004, and InfoVis 2005. These contests were based on other large scale evaluation campaigns, such as TREC (the Text REtrieval Conference), a conference created in 1992 for the purpose of evaluating different information retrieval systems in various tasks. The three contests had a similar general organization: after the dataset and tasks were posted, participants had four months to send their submission (a two page summary; a video illustrating the interactive techniques used; a detailed web page describing the tool and how it was used to accomplish the tasks; and information about the team and tool provenance). External judges with information visualization expertise and application domain experts then reviewed the submissions and evaluated the submissions in terms of quality, appropriateness, usefulness, flexibility, and quality of the supporting materials.

As argued by Plaisant, Fekete, et al. (2008), although the contests were an artificial testing situation, they encouraged participants to thoroughly exercise their systems over a long period of time, mimicking a fairly realistic analysis process. Similarly to what happens in evaluation studies outside of the contest scenario, benchmarks are difficult to create, promote and use. Nonetheless, these are necessary to push the technology curve, and ultimately to shed light on guidelines for visualization.

Insight at the edges of Information Visualization

With recent developments such as casual (Pousman et al., 2007), artistic (Viégas and Wattenberg, 2007), participatory (Viégas, Wattenberg, and Feinberg, 2009) / collaborative (J. Heer, Viégas, et al., 2009; Wood et al., 1997), and narrative (Segel and J. Heer, 2010) visualization, the traditional boundaries of information visualization have started to expand (Dörk, Feng, et al., 2013). Here, at what Pousman et al. (2007) call the edges of information visualization, evaluation becomes an even more challenging task. Nevertheless, "looking at the margins of the field may provide insight into problems and opportunities at the core of the field, and vocabulary that we use to describe these edge cases may be applicable to the entirety of information visualization" (Pousman et al., 2007, p. 1148).

Insight on non-traditional information visualization can be even more complex and heterogeneous. According to Pousman et al. (2007), there may be multiple types of insights, which can occur in completely different situations: it can be intentionally sought out and happen from focused work, or it can be a serendipitous encounter without a fixed or explicit task behind it. Both Saraiya et al. (2005) and Pousman et al. (2007) propose a list of types of insights, but coming from different backgrounds: the first from traditional visualization (genomics) and the second from casual visualization.

Saraiya et al. (2005) chose to group insights into four main categories: overview, patterns, groups, and details. These categories were created based on a study of five popular microarray visualization tools and the evaluation of the insights gained by the participants in the experiment. The authors characterize *overview insights* as insights that are gained while comparing the overall expression distributions. *Detail insights*, on the other hand, are acquired through the exploration of focused information, normally about subjects that they were familiar with. *Pattern insights* are acquired by providing ways for the user to identify or compare across data attributes. Lastly, *Grouping insights* are derived by the identification or comparison of groups. Some of these categories were also included in the processes through which people gain insights identified by Yi et al. (2008), namely overview insights and pattern insights. Due to the fact that the visualizations utilized in this study were traditional visualizations the categories also reflect that aspect.

Pousman et al. (2007), on the other hand, identify categories of insight that are not traditional. They divide insight in four, not mutually exclusive, categories: analytical insight, awareness insight, social insight, and reflective insight.

Acquired through exploratory analysis and extrapolation, *analytical insight* is very traditional, but when present in non-traditional visualizations can be broadened. There are several low-level tasks (or *cognitive tasks*) that can result in an analytic insight. Wehrend and C. Lewis (1990) identified eleven: identify, locate, distinguish, categorize, cluster, distribute, rank, compare within entities, compare between relations, associate, and correlate. These tasks can result in large or small eureka moments.

Awareness insight comes from maintaining awareness of a particular data stream, being on the lookout for fluctuations in the stream or for shifting patterns. "Staying aware of information does not help in some analytical task, but instead helps people communicate socially, keep on top of cultural trends and memes, and make connections between domains in informal ways" (Pousman et al., 2007, p. 1150). This kind of insight is often obtained in ambient visualization (Pousman et al., 2007) and information art (Stasko, McColgin, et al., 2005).

Social insight is insight about social situations, networks, and social life in general. This type of insight is subject to revision and reinterpretation, and it commonly occurs as a confirmation of suspicions or preconceived notions that the user previously had. As reported by Pousman et al. (2007, p. 1150), "while not having a particularly analytical character, gaining insight into social workings can open new hypotheses, invite reflection, and even change social or individual behavior".

If social insight is deeply connected with the knowledge that the user already has, this is even more true in relation to *reflective insight*. This is the kind of insight that the user gets about him/herself, the world, and his/her place in it. According to Pousman et al. (2007), this type of insight is often accomplished by providing a new perspective to the user or to enable him/her to distance him/herself from the situation.

2.2.3 The importance of function, user engagement, and target audience

Regardless of all of the pros that come with the increased concern with producing *beautiful visualizations*, we cannot risk the trivialization and marginalization of visualization. According to Steele and Iliinsky (2010, p. 1), beauty in visualization is not merely aesthetics: "for a visual to qualify as beautiful, it must be aesthetically pleasing, yes, but it must also be novel, informative, and efficient". One of the top priorities should continue to be to create visualizations that excel at exposing the data efficiently, in order to be easily understood by its target audience.

A visualization should never add more noise to the data and make it more difficult to comprehend. Therefore, function and purpose have to always be taken into account when designing a visualization. As stated by Kirk (2012, p. 15), it is not easy to effectively merge form and function, but "our aim should be to hit that sweet-spot where something is aesthetically inviting and functionally effective." Unless we are talking about more artistic or casual visualizations, which give more relevance to aesthetics, concerns with function should be the top priority. Nonetheless, it is also true that attractive visualizations usually get favorable emotional and mental responses and that attractive form enhances function, thus the ideal would be not having to sacrifice one for the other.

Although data definitely plays a role in the final shape of the visualization there are other variables that end up influencing the form of the visualization a lot more. This is the case of purpose/function. The popularization of the personal computer, and specially widespread use of the Internet, led to the adoption of visualization strategies in domains that are not directly related with computer science (the field where information visualization was born). According to Pousman et al. (2007), information visualization now has a wide range of users, from a wide variety of domains, who use information visualization in their work tasks and everyday situations, and these users have a substantial impact in the visualization techniques. The fact that information visualization started being targeted at non-domain experts rose the concern with engagement: the power to attract and hold the users' interest in order for him/her to further explore the visualization. Consequently, we need to understand which visualization techniques can be used by non-expert users and which techniques are flexible enough to be adapted to different types of users and tasks.

Engagement is particularly important in fields that benefit not only from the time the user spends exploring the visualization, but also from users sharing the visualization. A notable example is news media: if users spend more time on the website exploring a visualization that translates into more valuable space for advertisers (who prefer to have their adds featured in pages where users spend more time); the fact that users are so interested in the visualization that they chose to share it with their friends, either through social media or by word of mouth, means that the page where the visualization is featured will reach more people therefore generating more revenue from advertising. Another business that benefits from the use of information visualization is marketing: if users are engaged in exploring the visualization, that means that they will be exposed to the information about the product for a longer period than they would if they were exposed to a traditional add (static or video); similarly to what happens in news media, shares promote the content and make it reach new audiences, occasionally even reaching audiences that would not come across that visualization. However, these visualizations, targeted at a wider (often inexperienced) audience, are not the only kind of visualizations that are said to benefit from techniques that improve engagement. For instance, engagement has been identified as a particularly important feature for collaborative/participatory visualization, because users have to be invested in the visualization in order to be willing to contribute.

Previous research has identified aesthetics (Lang, 2009; Vande Moere, Tomitsch, et al., 2012), narratives (Boy, Detienne, et al., 2015; Figueiras, 2014a; Ma et al., 2012; Satyanarayan and J. Heer, 2014; Segel and J. Heer, 2010; Waldner et al., 2014), and interaction (Diakopoulos et al., 2011; Dix and Ellis, 1998; Liang et al., 2010; Yi et al., 2007) as possible factors for engagement. However, there is little evidence that these strategies really benefit the exploration of the visualization and, if they do make the user feel more engaged, how can they be applied to the visualization. According to Healey and Enns (2012, p. 17), "the exact mechanisms behind engagement are currently not well understood". Healey and Enns (2012) argue that engagement can be used to direct the users' attention to elements that the visualization designer believes are important. However, by directing the users' attention too much we can discourage him/her to freely explore and find insights that the designer could not predict. According to Dörk, Feng, et al. (2013), omission and emphasis can be used to reveal patterns in the dataset, and these decisions are in the hands of the designer. "Choosing what is highlighted will be affected by the designer's values and intentions" (Dörk, Feng, et al., 2013, p. 2190).

There are several variables behind the success or failure of a visualization, hence it is difficult to understand which elements help and which disrupt the exploration of the visualization. Can engagement get in the way of the understanding of the visualization? Achieving the answer to this and other questions regarding engagement is vital to understand if we can achieve a balance between the pros and cons (distraction) of the several strategies to engage the user in the visualization. The heterogeneity of characteristics in visualization makes it difficult to measure the effectiveness of each of the user engagement strategies typically used in information visualization (aesthetics, narratives, and interaction). The ideal scenario would be to evaluate each element by itself. Several studies already attempted to do that: Liang et al. (2010) for interaction techniques, Satyanarayan and J. Heer (2014) for narrative, and Bateman et al. (2010) for visual embellishments, to name a few. However, most visualizations use more than one strategy therefore making it difficult to understand which element is responsible for the highest level of user engagement. Nonetheless, it is also possible to evaluate a visualization as a whole. For this effect, researchers usually resort to common website evaluation metrics: average session duration, page depth, comments and shares, returning users, etc. However, evaluating visualizations as a whole rarely gives insights on how to build future visualizations, because it is difficult to understand if there are elements that help more than others. Similarly to what happens in web design, it is important to follow the good practices of HCI when creating a visualization.

2.2.4 Functionality over aesthetics

Although aesthetics and art are often approached by the field of information visualization, it is always stressed that the main objective driving visualization design is to provide insights and aesthetics are secondary. However, it is also known that aesthetic plays an important role in user engagement, thus being a potentially positive influence even on more traditional forms of visualization stimulating desire, positively influencing the first impression, encouraging repeated use, or overwhelming the audience — being memorable — (Cawthon and Vande Moere, 2007). Considering that visualization has for some time moved from being exclusively for highly skilled experts to being of interest to the masses, achieving solutions to issues such as user engagement has become more and more essential. According to Vande Moere and Purchase (2011), the three potential characteristics that affect engagement are: design quality (in which aesthetics is included), data focus, and user interaction.

The concern with aesthetics does not have a long tradition within traditional information visualization. Visualization is frequently seen as a part of human-centered computing and most visualizations are "designed to remove any sublimity, and instead foster immediate understanding" (Kosara, 2007, p.633). According to Kosara (2007), there is a division between the more technical, analysis-oriented side of information visualization and the artistic approaches to the field. This division forces visualizations more concerned with aesthetics to a separate category (Art Visualization) or to other edges of information visualization (such as Casual information visualization, social visualization, ambient visualization) and prevents a more integrated approach. Nevertheless, the contribute that these edges of the field can have to the *core* are valuable and can push the field of Information Visualization forward. In the

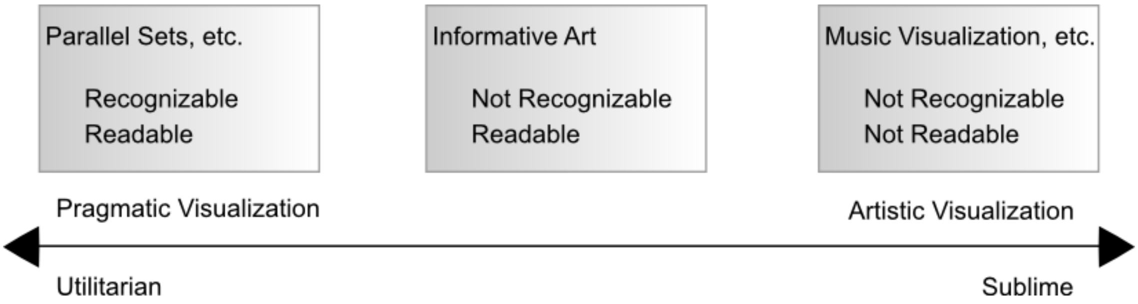


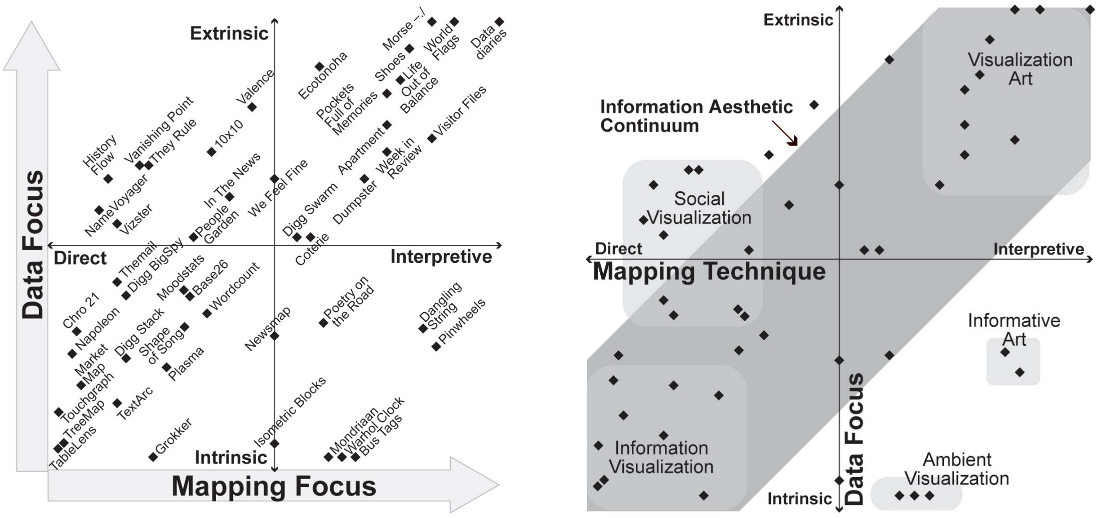
Figure 2.17: Scale (Figure 1 (Kosara, 2007))

opinion of Vande Moere and Purchase (2011), the different perspectives of other communities can facilitate the introduction of complex and relevant issues to a large audience through popular media.

Recently though, it seems that the information visualization community is leaning towards a more integrated approach. This was perhaps sparked by the growing interest that people from outside of this research area (artists, journalists, designers, etc.) have demonstrated in visualization, mainly as a tool for visual expression on the Web. Consequently, there was also a boom of user-friendly and sophisticated visualization toolkits that have aesthetics in mind. "This online practice seems to purposefully use striking visual styles, for instance to attract the attention of a sizable audience, to compel potential users to engage with the visualization, or to share the visualization experience with other" (Vande Moere, Tomitsch, et al., 2012, p. 2739).

According to Kosara (2007), visualization can be placed on a scale ranging from utilitarian to sublime), being more pragmatical or more artistic respectively. The scale can be seen in Figure 2.17. The primary goal of pragmatic visualization is for the user to understand the data, hence it has to allow the user to explore and analyze the data. Pragmatic visualization on the scale proposed by Kosara (2007) is more utilitarian, readable, and recognizable. It is a more objective portrayal of facts and is not based on personal opinions or subjective points of view. Having in mind the model of information aesthetics proposed by Lau and Vande Moere (2007), it would be considered more *direct* in terms of mapping techniques (the user is able to infer the underlying data) and more *intrinsic* in terms of data focus (it facilitates the acquisition of insights by the user). In contrast, the goal of artistic visualization is for the user to understand a basic concern, thus rather than making the data readable, the objective is to make something visible in an interesting way. On the scale proposed by Kosara (2007), artistic visualization is more sublime, in other words it inspires awe, grandeur, and evokes a deep emotional and/or intellectual response. Normally, it is more subjective and the information is more obscured, so the underlying values are unclear and it is not readily recognizable as a visualization. In comparison to pragmatic visualization, it would be on the opposite side of both scales in the model proposed by Lau and Vande Moere (2007, p. 90): it would be considered more *interpretative* in terms of mapping techniques ("the visualization design may be stylized, adopted from cross-disciplinary inspirations") and more *extrinsic* in terms of data focus (it facilitates the communication of meaning implied by the data).

The model of information aesthetics by Lau and Vande Moere (2007), which can be



(a) Some of the visualizations analyzed placed of the model according their data and mapping focus

(b) Visualization fields identified on the model of information aesthetics

Figure 2.18: Model of information aesthetics (Figure 4 and 5 (Lau and Vande Moere, 2007))

seen in figure 2.18, shows that there is a correlation between mapping technique and data focus. In other words, "techniques that are based on direct mapping often focus on intrinsic patterns, whilst interpretive mapping highlights extrinsic data meaning" (Lau and Vande Moere, 2007, p. 90).

According to this proposed model, based on the analysis of existing visualization examples, traditional information visualization examples tend to use more direct mapping techniques and have a more intrinsic data focus (bottom left corner), and visualization art examples tend to use more interpretative mapping techniques and have more extrinsic data focus (top right corner). Lau and Vande Moere (2007) also realized it is possible to find other types of visualization such as social visualization (top left corner), and ambient visualization and informative art (bottom right corner) outside of the spectrum between the two extremes (Information Visualization and Visualization Art). This means that social visualization, also known as collaborative or participatory visualization, uses more direct mapping techniques, but more extrinsic data focus. In order words, since engagement is important for social/collaborative/participatory visualization the data focus is more extrinsic, however it is important that the users fully understand the visualization so direct mapping techniques are common used. Consequently, it also means that ambient visualization and informative art use more interpretative mapping techniques, but have more intrinsic data focus. In order words, since ambient visualization and informative art are close to visualization art interpretative mapping techniques are typically used, however it is important that the data is easily understood so the data focus is more intrinsic. According to Lau and Vande Moere (2007, p. 91), informative art, ambient and social visualization "are nevertheless part of and most probably formed the foundation for the information aesthetic movement."

However, a lot of visualizations do not fit in any of these extremes, pragmatic or artistic, and are scattered along the scale, being either closer to utility or to sublimity. This is the case of social visualization, casual information visualization, ambient visualization, informative art, etc.

Evaluating aesthetics

Excluding the case of Artistic Visualization, in which utility seems to play a less crucial role, aesthetics cannot be the main focus information visualization. For all the types of visualization that are neither fully pragmatical nor fully artistic, it is vital that the ideal balance between function and aesthetics is achieved. However this is not an easy task. Similarly to what happens with measuring the success of most subjective and complex components of visualization outside of traditional usability metrics, the first obstacle is the fact that is hard to measure aesthetics (Vande Moere and Purchase, 2011). The concept of aesthetics is related to beauty which is very subjective, but it has been successfully introduced in other functionality driven areas such as architecture and product design, as something other than a reflection of personal judgment. The fact that the users may not even understand why they find a visualization aesthetically pleasing or be able to explain what makes it attractive is also a limitation in the process of evaluating aesthetics.

According to Norman (2005), in order to analyze the attractiveness of the product, its behavior, and the image it presents to the user, one needs to understand the user's response to the product at three different levels:

- Visceral: the first impact to the overall appearance, touch, and feel; subconscious reactions to certain experiences;
- Behavioral: the experience with the product; function, performance, and usability;
- Reflective: thoughts that come afterwards, how it makes the user feel, the message it tells others, etc.

These consequently translate into three different kinds of design.

Following the train of thought by Norman (2005) on the relationship between emotion, beauty and usability, several other researchers in the field of HCI (Sonderegger and Sauer, 2010; Tractinsky et al., 2000) have pointed out that aesthetics and usability can be positively correlated, supported by the fact that there is no usability principle that would inherently conflict with the usability principles.

In graph-drawing, aesthetics has been successfully studied and measured as a quantifiable entity, and is considered a positive influence on task effectiveness. In this area of study, researchers have found that there are benefits in minimizing bends and edge crossings, and maximizing angles, orthogonality and symmetry (Cawthon and Vande Moere, 2007).

In the same vein as the studies in graph-drawing, Cawthon and Vande Moere (2007) researched the relationship between aesthetic and measures of effectiveness and efficiency in data visualization. Having task abandonment, erroneous response time, and aesthetics as metrics, this study presented to its participants 7 of the 11 visualizations of the study (TreeMap, IcicleTree, SpaceTree, Windows Explorer, BeamTrees, StarTree, Dendogram Tree, Polar View, StepTree, Botanical Viewer, and SunBurst) and asked them structure and attribute-related questions, to rate the perceived beauty of the visualization, and to compare visualizations that used different techniques and rank them. The study showed that the SunBurst method was associated with the highest level of perceived beauty and that the BeamTrees method was the least favorite. Additionally, the study found a correlation between a favorable or unfavorable aesthetic ranking and metrics of task abandonment and erroneous response.

There are also some approaches that aim to measure and quantify features related to the concept of *order*, such as proportionality, complexity, and variety Vande Moere and Purchase (2011). However, according to C. Chen (2005, p. 15), in the graph-drawing community "much of the aesthetics wisdom consists more of heuristics than empirical evidence at the elementary level of perceptual-cognitive tasks" and most researchers are not focused on the semantics associated with the data, rather focusing mostly on graph-theoretical properties.

Although much could be learned by the study of successful examples of visualization, there is not much research moving in that direction. The reason for this is that, although it would be useful to understand what makes some popular visualizations so successful, it is also hard to understand if their success arises from the way the data is visualized or from the data itself. It would also be useful to know more about the design reasoning behind these successful examples. However, the design process behind the creation a lot of these popular visualizations is the *genius design* approach, in which the designer is the only one responsible for all the design decisions and is moved mostly by his/her natural instincts. Although other approaches such as user-centered design (which involves capturing the users' needs and preferences beforehand) and system design (which is based on a structured roadmap of decisions) are also used, according to Vande Moere and Purchase (2011) most approaches are a balance between engineering design and creative design.

Another obstacle to the evaluation of information visualization is the fact that the sub types of visualization at the edges of the field might provide different types of insights. Inspired by the hypothesis that casual information visualizations can provide other kinds of insight (Pousman et al., 2007), a comparative study of three visualization demonstrators with the corresponding exemplars was carried out byVande Moere, Tomitsch, et al. (2012): an Analytical Style Demonstrator (Gapminder, Many Eyes, and OECD eXplorer), a Magazine Style Demonstrator (We Feel Fine, Digg Labs, and remap) and an Artistic Style Demonstrator (Bitalizer, Texone, and Poetry on the Road 2004). Without much emphasis on usability metrics, such as task performance, the evaluation study intended to observe the impact of different levels of visual and interactive embellishment had on the same data presented using a scatter plot technique. Vande Moere, Tomitsch, et al. (2012) found that style seems to have a big impact on usability. According to the authors, the participants considered that the analytical style visualization, without visual or interactive embellishments, was more clear, informative, useful, usable, functional, easy to understand, and surprisingly more engaging and enjoyable. The participants also considered the insights from an analytical style as deeper. The researchers also observed that all stylistic approaches had the same depth, confidence, and difficulty of insights. The kinds of insights generated by the different stylistic approaches were different though.

According to the patterns observed in this evaluation study, because embellishments tend to *hide* or distract from visual patterns, the more embellished styles produced more insights derived from reasoning, reflection, or interpretation. Nonetheless, there was not much difference between the two embellished styles in terms of both usability and insights. In fact, a previous study by Bateman et al. (2010) found that there was no difference in comprehension between embellished and plain charts. In contrast to what happened in the study by Vande Moere, Tomitsch, et al. (2012), the participants of the study carried by Bateman et al. (2010) said that the embellished chart was more enjoyable. Additionally, the embellished chart was also the participants' favorite, considered the most attractive, the easiest to describe and to remember, the easiest to remember details, the most accurate to describe and to remember, and the fastest to describe and to remember (Bateman et al., 2010). Although this evaluation study by Vande Moere, Tomitsch, et al. (2012) sheds light on possible benefits or disadvantages of having visual and interactive embellishments on visualizations, its findings are still not solid enough and more research is needed. One of the issues with this study might be that the demonstrators were not the best representative samples.

Since in visualization there is a strong relationship between functionality and user satisfaction, possibly studies like the experiment conducted by Tractinsky et al. (2000) to test the relationships between users' perceptions of beauty and usability can be successfully replicated for the particular case of visualization. The study carried by Tractinsky et al. (2000, p. 141) revealed that there is a "tight relationships between users' initial perceptions of interface aesthetics and their perceptions of the system's usability". It would be interesting to see if the same would happen in the study of visualization and if users would perceive a beautiful visualization that does a poor job in promoting insights as useful just because it is aesthetically pleasing.

2.3 Chapter Summary

In this thesis chapter 2 — Visualization analysis — is the chapter with the most emphasis on literature review and forms the first of a three-part literature review. This chapter itself is divided in two parts: the past of visualization, and the present and future of visualization.

Reviewing the past history of visualization is important, as it helps to understand where the field is going. According to Friendly (2008), "It is common to think of statistical graphics and data visualization as relatively modern developments in statistics", however visualization have deep roots in the earliest map-making and visual depictions. Some of these earlier examples of visualization such as Minard's *Carte figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813* are seen as valuable inspiration for newer trends such as narrative visualization. Therefore it is important to understand visualization's history and how it has impacted the present and the future of the field. This first part of the chapter focuses on:

- the history of visualization — referencing some of the most iconic examples, such as Florence Nightingale's coxcomb diagram on mortality in the British army (subsection 2.1.1) and Hans Rosling's Trendalyzer (subsection 2.1.4);

- the impact of the Web and how it shapes visualization (subsection 2.1.5);
- and the impact that key events such as the Wikileaks releases on Afghanistan, Iraq and US embassy cables, and the open government data trend had on the popularization of visualization on the Web (subsection 2.1.6).

The second part of chapter 2 includes a reflection upon the present and future of the field of information visualization by focusing on:

- the principles used in visualization that originated in other areas — such as the Gestalt laws referenced by Ware (2004) and Nesbitt and Friedrich (2002) (subsection 2.2.1);
- the importance of the insights and knowledge (subsection 2.2.2) —- describing the process of gaining insights, the role of visualization literacy and evaluation, and how the process of gaining insights occurs at the edges of Information visualization;
- the role of function, user engagement, and target audience (subsection 2.2.3);
- and the importance of giving priority to functionality over aesthetics, while notwithstanding the role that aesthetics plays in user engagement and the necessity to evaluate it (subsection 2.2.4).

In the next chapter, I establish how some key terms will be approached throughout this dissertation and the definitions adopted, in order to avoid mixing up concepts.

Chapter 3

The many names of visualization: Information/Knowledge/Data

Where is the wisdom that we have lost in knowledge?
Where is the knowledge that we have lost in information?
— T. S. Eliot, *The Rock*

There is no consensus regarding the many names of visualization. Information visualization, data visualization, infographics, and other terms are often used interchangeably, and the confusion is usually extended to their definitions, which often overlap. "For instance, data and information are often interchangeable in computing (for example, data processing and information processing or data management and information management)" (M. Chen et al., 2009, p. 12). The answer to this issue seems to lay in the definition of key terms used in visualization research: data, *information*, and knowledge. Several definitions of data, information, and knowledge were proposed over the years, coming from disciplines such as psychology, management sciences, and epistemology (M. Chen et al., 2009). However the terms were seen in a new light when computer science gained more notoriety.

Some of these proposed definitions are intimately related to the Data/Information/-Knowledge/Wisdom (DIKW), also referred to as Knowledge Hierarchy, Information Hierarchy, or Knowledge Pyramid, which can be seen in Figure 3.1. The DIKW pyramid is so important for the clarification of these terms that is "often quoted, or used implicitly in definitions of data, information and knowledge" (Rowley, 2007, p. 164). The DIKW is a well known model for classifying human understanding in the perceptual and cognitive space that implies that data can be used to create information, that information can be used to create knowledge, and that knowledge can be used to create wisdom.

	M. Zeleny	R.L. Ackoff
Wisdom	Know-why	Evaluated understanding
Knowledge	Know-how	Application of data and information; answers how questions
Information	Know-what	Data that are processed to be useful; provides answers to who, what, where and when questions
Data	know-nothing	Symbols

Figure 3.1: The Data Information Knowledge Wisdom (DIKW) hierarchy and definitions of these terms in perceptual and cognitive space by Zeleny (1987) and by Ackoff (1989).

Although the origin of the DIKW hierarchy dates back to the poem *The Rock* by T. S. Eliot, published in 1934, it was the organizational theorist Russel Ackoff (1989) who brought it to prominence. For Ackoff (1989) wisdom is at the top of what he calls the hierarchy of types. As a fourth level he adds *understanding*, between knowledge and wisdom, followed by information, and data at the base. Understanding is the appreciation of why. It is an analytical and interpolative process that relates previously held knowledge to the new knowledge that is being acquired. In other words, it differs from knowledge by not only making possible the transformation of information into instructions but also by being able to answer why questions. Understanding is often omitted from most DIKW hierarchies and that fact seems to support the popular idea that understanding is not a separate level but the thing that supports the transitions from each stage to the next (Rowley, 2007).

According to Ackoff (1989), understanding has an aura of permanence but only wisdom is permanent in the true sense. However there can be no wisdom without understanding, and also no understanding without knowledge. For Ackoff (1989) wisdom requires judgment and is intimately related with the ethical and aesthetic values of the individual. Although wisdom is commonly included as a level in the DIKW hierarchy it is not often considered outside of the discussion of this model, even if the DIKW model is used as the base for the study.

Zeleny (1987), who approached the DIKW hierarchy a couple of years before Ackoff (1989), also proposes an addition to the model. He maintains data at the base of the pyramid, followed by information, knowledge, and wisdom. At the very top of the pyramid Zeleny (1987) puts *enlightenment*, which according to him is related with the concept of *wisdom* (with answering or understanding why) and also with the idea of truth, of right and wrong, and the sense of social acceptance, respect, and sanctioning. In other words, on the model by Zeleny (1987) wisdom is closer to what Ackoff (1989) considers understanding (both related to knowing why), and enlightenment is somehow closer to Ackoff's definition of wisdom. Although Zeleny (1987) considers that wisdom refers to explicability, his definition of wisdom is related to ethics, in the same way that for Ackoff (1989) wisdom is. In Figure 3.1 the definitions that Ackoff (1989) has for each term can be seen and compared to the model proposed by Zeleny (1987).

3.1 Defining Data, Information, and Knowledge

Data is usually on the base of the DIKW pyramid. Keller and Tergan (2005) describe data as being raw, symbols or isolated and non-interpreted facts. In fact, most authors (Ackoff, 1989; M. Chen et al., 2009; Rowley, 2007) consider that data are symbols. Data can come in a variety of forms: as numbers or words; being measurements, observations, or descriptions of things; written in paper, stored in an electronic memory as bits and bytes, or kept as facts in a person's mind. Data has no significance beyond its existence and according to Rowley (2007) are of no use until they are in a usable form, because they do not convey any specific meaning. A usable form is given by context, interpretation, and organization. When data are well-formed and meaningful they become information, and the difference between the two is functional and not structural.

Information has been defined in several different ways. It is a polysemic concept that can be associated with several different explanations depending on the field, hence a single definition of what is information would hardly satisfy the numerous applications of the term (Shannon, 1953). The most well known definition of information comes from the field of *Information Theory* and from the *mathematical theory of communication*, which is primarily concerned with efficient ways of encoding and transferring data. To Shannon (1953), information is viewed as a purely quantitative measure of communication exchanges and is unrelated to qualitative properties such as meaningfulness.

However, over the last three decades, a bipartite definition that consists of a conjugation of data and meaning has become the most prominent, specially in fields that tend to approach data and information as more concrete (less abstract) entities, such as Information Science; Information Systems Theory, Methodology, Analysis and Design; Information (Systems) Management; Database Design; and Decision Theory (Floridi, 2005). This definition of information encompasses data as the raw material which has to be processed and refined. Once this happens meaning is attached to the data and this data becomes information. If an individual can use the data to answer questions beginning with *who*, *what*, *where*, and *when* then that data he has is in fact information.

Data is transformed into information, which in turn provides knowledge. It is a collection of information that "can be obtained either by transmission from another who has it, by instruction, or by extracting it from experience" (Rowley, 2007, p. 166). Knowledge consists of the application of data and information, using it to answer *how* questions (M. Chen et al., 2009). Information ages rapidly, but knowledge, on the other hand, has a longer life-span (Ackoff, 1989). Although knowledge involves understanding patterns, and combines the acquired information with experience and other accumulated information, it does not provide a profound understanding of why. For instance, when an individual memorizes information, such as learning the multiplication table, he/she is acquiring knowledge and can answer questions directly related to that memorized information (how much is 2 x 2 or 5 x 6). However, without the next step in the hierarchy — wisdom for Zeleny (1987) and understanding for Ackoff (1989) — which requires cognitive and analytical ability, he/she will not be able to answer something that is not on the table that he/she memorized (how much is 145 x 34).

3.2 Visualization of Data, Information, Knowledge, and everything in between

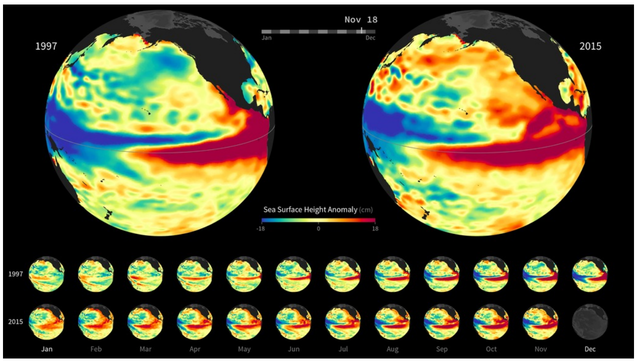
Several attempts have been made in the visualization community to clarify the different terms used for visualization. Sometimes information visualization is considered to be an umbrella term that covers other many subtypes of visualization. This is the case of Card and Mackinlay (1997), who consider information visualization to be a broader area that encompasses several major types of visualization, one of them being scientific visualization. The other types identified by Card and Mackinlay (1997) are GIS-based visualizations, multidimensional scatter graphs, multidimensional tables, information landscapes and spaces, node-link diagrams, trees, and text transforms. According to Chi (2000), this data-oriented taxonomy was later expanded in the book *Readings in Information Visualization: Using Vision to Think*(Card, Mackinlay, and Shneiderman, 1999). This categorization also inspired Chi (2000) to taxonomize various information visualization techniques. Similarly to Card and Mackinlay (1997), his categories include scientific visualization, geographical-based info visualization, multidimensional plots, information landscapes and spaces, trees, and text, but Chi (2000) adds 2D, network, and web visualization. However, it is also very common to see information and scientific visualization as two completely separate areas under the umbrella term *visualization*. This perspective of the visualization field as a whole is supported by Tory and Moller (2004).

When comparing information visualization and knowledge visualization, on the other hand, the two are mostly seen as separate disciplines. The research in these two areas was primarily focused on techniques of visualization but are now starting to move towards strategies to make visualizations that are independent and eschew other types of narratives. Storytelling is one of the factors that is enabling the approximation of these two areas of research that have been historically separated and developed independently. With the introduction of storytelling the differences between information and knowledge visualization are getting blurred. According to Bertschi et al. (2011, p. 330) "in order for information to transform into knowledge, one must share some context, some meaning, in order to become encoded and connected to preexisting experience." Storytelling is one of the tools that is capable of introducing context and meaning in the visualization, not only helping users to establish connections between the complex data represented,but also introducing an important component, particular to knowledge visualization, that is tacit knowledge. This category of knowledge, that encompasses things such as intuitions and subjective insights, is difficult to communicate (Keller and Tergan, 2005). However it is known that storytelling has the power to engage and make people relate, and possibly facilitating the sharing of tacit knowledge.

3.2.1 Scientific visualization

Information visualization has been historically closer to the scientific fields (Ma et al., 2012) and often to computer science (Keller and Tergan, 2005). However scientific visualization often is interested that the visual representation of the data has some kind of correspondence with reality: "this area is primarily concerned with the visualization of 3D phenomena

(a) 3D plot of the various states of a fictitious substance with water-like properties (Plate III in Maxwell (1990))



(b) Comparisons of Pacific Ocean sea surface height anomalies in 2015 and during El Niño in 1997 by NASA (Jentoft-Nilsen, 2015)

Figure 3.2: Examples of scientific visualization

(architectural, meteorological, medical, biological, etc.), where the emphasis is on realistic renderings of volumes, surfaces, illumination sources, and so forth, perhaps with a dynamic (time) component" (Friendly, 2005, p. 2). According to Keim, Andrienko, et al. (2008), current scientific visualization research is focused on interactive exploration and the efficiency of the visualization techniques to enable it, but progressively the interest of the scientific visualization research community in methods to automatically derive relevant visualization parameters has grown. Common techniques in scientific visualization include computer animation and simulation, surface and volume rendering, and volume visualization, among others. Two examples of scientific visualization can be seen in Figure 3.2: the first, one of the earliest examples of scientific visualization, is a thermodynamic surface sculpted in clay by James Clerk Maxwell in 1874, which can be seen in figure 3.2a; the second is a visualization by NASA that provides side-by-side comparisons of Pacific Ocean sea surface height anomalies in 2015 and during the 1997 El Niño-Southern Oscillation, which can be seen in figure 3.2b.

According to Tory and Moller (2004) and their proposed taxonomy, scientific visualization tends to occupy the top left area of their high-level visualization taxonomy. This means that for scientific visualization tends to follow continuous design models, which assume that data can be interpolated, and have given display attributes, such as spatialization, colour, transparency, and time. On the other hand, information visualization tends to occupy the the bottom right corner, in other words, their design model is discrete (assumes data cannot be interpolated) and have chosen display attributes. As reported by Tory and Moller (2004, p. 155), "middle areas are ambiguous and belong to both categories or neither", showing that the two areas can overlap.

The names historically given to the two research areas are rather unfortunate. Information visualization is not unscientific and scientific visualization is not uninformative. Both often share the same goals, can focus on similar subjects, and aim to amplify cognition. Unlike scientific visualization, which is inherently the visual representation of physically-based data, information visualization often makes use of visual metaphors instead. Information visualization is about manipulating typically abstract raw data and representing it in a structured way that enables users to extract simple and exact information that will clarify questions of what, where, who, when, and why. Its main goal is to help users understand and analyze the data. Most of the times, including when it was originally defined by Robertson, Card, et al. (1993), the term information visualization refers to computer-supported interactive graphical representations of information (Card, Mackinlay, and Shneiderman, 1999; C. Chen, 2010).

3.2.2 Visual analytics

There is also a common misunderstanding between information visualization and visual analytics. According to Keim, Andrienko, et al. (2008) the origin of the confusion between the two areas lies in the fact that there are a lot of examples of information visualization highly related to visual analytics. However, visualization does not have to "necessarily deal with an analysis tasks nor does it always also use advanced data analysis algorithms" (Keim, Andrienko, et al., 2008, p. 158).

Although the public often mixes the two, most authors seem to agree that visual analytics is an area that makes use of information visualization, therefore being a separate area of research that deals with visualization of data. It integrates scientific visualization, information visualization, and other disciplines such as data management and analysis, spatio-temporal data, human perception and cognition, cognitive and decision science, and statistical analysis, combining "automated analysis techniques with interactive visualizations for an effective understanding, reasoning and decision making on the basis of very large and complex datasets" (Keim, Andrienko, et al., 2008, p. 157). Nonetheless, in the past, most of the research in visual analytics has originated in the visualization community, however that fact is slowly changing with the increasing complexity of knowledge discovery algorithms. That fact has contributed to the gradual diversion of the two research areas.

Visual analytics and information visualization share common goals: improve understanding of the data (specially when dealing with high volumes of data), promote discovery and insights, etc. However, the primary goal of visual analytics is actually the analytical reasoning, and visualization is the facilitator of the data analysis and consequent reasoning (Keim, Andrienko, et al., 2008; Thomas and Cook, 2006). More than allowing users to detect expected events, Cook, Earnshaw, et al. (2007) suggests that it is crucial to allow users to discover the unexpected: to discover anomalies in the data, hidden patterns, obscure relationships. Therefore, the visual analytics community is much more focused in researching how the users can interact with the data and turn that information "into intelligence to tune underlying analytical processes" (Keim, Andrienko, et al., 2008, p. 158). Another important research topic in visual analytics is automated analysis algorithms.

3.2.3 Data visualization

The term data visualization has become very popular in the last few years. There is no consensus about what the term means actually means and the attempts to define are usually too close to the definition of information visualization. Some people consider data visualization to be closer to scientific visualization for historical reasons or by assuming that scientific visualization deals with more data. However, if "information is data that has been given meaning through interpretation by way of relational connection and pragmatic context" (Keller and Tergan, 2005, p. 3), and if a successful visualization turns data into information, then a successful data visualization is actually information visualization.

Not all information visualization thus is a visualization of data, because structured information can also be visually represented using information visualization. Notwithstanding, even when the volume of data is taken into account, most literature consider the correct term to be information visualization, and data visualization to be no more than sub category of information visualization that deals specifically with high volumes of data. However, there are a few examples (Fayyad et al., 2002; Han et al., 2011; Soukup and Davidson, 2002) of the use of the term data visualization in disciplines such as data mining, probably because the main focus of this area is raw data. In this area, and in other literature that uses this term, data visualization is focused primarily in the identification of patterns.

3.2.4 Knowledge visualization

Contrary to what happens in other areas of visualization, which are closer to computer science, knowledge visualization is closer to social sciences focusing on abstract information related to the human cognition (Keller and Tergan, 2005), such as knowledge and experience (Bertschi et al., 2011) rather than discrete/numerical information. Although knowledge visualization also answers question of how and why, these answers are more grounded to the understanding of the visualization presented rather than the visualization itself.

According to Bertschi et al. (2011), knowledge visualization is a way to map information and concepts graphically in order to convey knowledge. Knowledge visualization aims to overcome the limitations of text and visuals by representing both in a structured and meaningful way, and is often more easy for the user to process more effectively than propositional visualizations (Keller and Tergan, 2005). Since information needs to be processed in order to become knowledge, it is vital to be able to represent it in ways that promote this transmission of knowledge. This form of visualization is particularly useful in collaborative work situations and other tasks that require knowledge-intensive communication between individuals or within groups. As argued by Bertschi et al. (2011, p. 332) "mapping the group dialogue can facilitate the integration of knowledge and it can surface misunderstandings more prominently than text". Well known examples of knowledge visualization include knowledge mind mapping, conceptual diagrams, and knowledge maps. Tergan et al. (2006) argue that the choice of way to visualize knowledge is influenced by the domain by which is used: in the domain of psychology researchers would be more concerned with clarifying the inherent aspects of mental representations of knowledge; in educational scenarios the main concern is facilitating the process of learning, solving problems, generating ideas, and giving instructions; in business the most common knowledge visualization task is to represent knowledge owned by a person, a group, or an enterprise in order to be transferred, rather fast, to different stakeholders.

Existing research recognizes a number of similarities between information visualization and knowledge visualization, and stress how both areas are grounded in similar theoretical assumptions (Bertschi et al., 2011; C. Chen, 2010; Keller and Tergan, 2005; Tergan et al., 2006). One of the most cited resemblances is the fact that both disciplines are focused on organizing information and knowledge so it can be accessed easily, in order for users to get insights. However, most agree that the two disciplines diverge in terms of the complexity of the insights they want to provide (C. Chen, 2010). For information visualization the insights can be either simple or complex, however, for knowledge visualization the primary goal is the transfer of knowledge (Burkhard, 2005), therefore the insights have to be more complex. Knowledge visualization not only aims to provide insights, but also share experiences, values, or perspectives, in a way that can be easily assimilated.

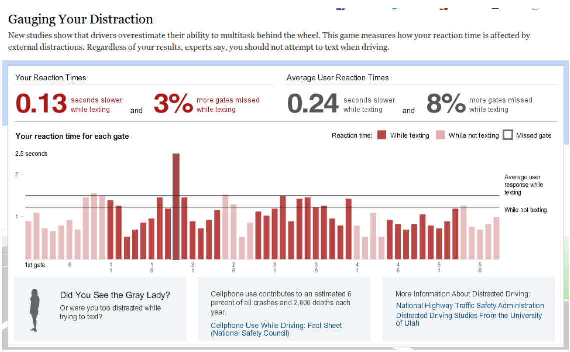
Another fact that sets knowledge visualization apart from information and all of the other forms of visualization is the emphasis given to knowledge (as defined in Section 3.1) rather than to numerical information in high-dimensional datasets. According to Masud et al. (2010), knowledge communicates procedural knowledge (how to do something) and conditional knowledge (when and why to use the acquired knowledge), giving the user the ability to use the acquired knowledge to take action. Because knowledge has to be recreated in the mind of each individual that interprets the visualization, in the process of creation there has to be an increased concern with who the target audience (Burkhard, 2005), due to the different cognitive capacities that they can have. In order to improve the knowledge acquisition things such as the distinction between the different types of knowledge present in the visualization (if it is know-why or know-how) and, if necessary, different visualization methods that complement the main one can be provided to the user. Moreover, knowledge visualization makes use of strategies such as spatial layout and highlight of contextual relationships to improve the knowledge acquisition (Keller and Tergan, 2005).

3.2.5 Casual, social, collaborative, and artistic visualization

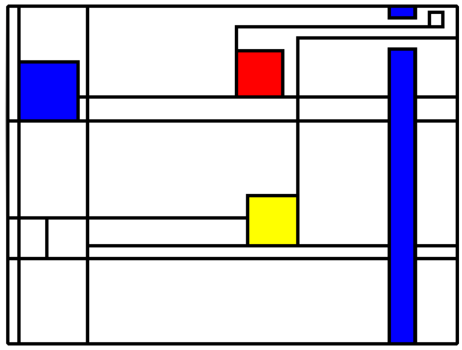
Non-traditional visualization examples have become more prominent with the popularization of information visualization, specially on the Web, and the growing concern about user engagement. While Pousman et al. (2007) call these the edges of information visualization, Dörk, Feng, et al. (2013) prefer to call them the alternative voices of information visualization and identify four different categories: artistic visualization, narrative visualization, participatory visualization, and casual visualization. Dörk, Feng, et al. (2013) leave out ambient visualization that for Lau and Vande Moere (2007) is an independent category that is neither traditional information visualization nor visualization art.

Casual or vernacular visualization

Casual (Pousman et al., 2007) and vernacular (Viégas and Wattenberg, 2008) visualization, which do not have analytical insight as their primary goal, played an important role on the transition of visualization from a *only for experts* domain into the mainstream scene. The term casual information visualization was coined by Pousman et al. (2007, p. 1151) and refers to visualizations that convey "increased focus on activities that are less task driven, datasets that are personally meaningful, and built for a wider set of audiences." It is a branch of traditional information visualization that would fit most visualizations produced by news media, marketing



(a) *Gauging Your Distraction*, New York Times visualization to raise awareness about cellphone use while driving



(b) A bus schedule visualization, inspired by Mondrian (Figure 7 (Skog et al., 2003))

Figure 3.3: Examples of casual visualization

initiatives, and non-professionally produced visualization in general. Often casual information visualizations are made with the intent of raising awareness to a topic (*Gauging Your Distraction*[1], an example can be seen in Figure 3.3a), sharing participatory content or sharing reflections about oneself, and making engaging data displays (an example can be seen in Figure 3.3b). According to Pousman et al. (2007), casual information visualization is an umbrella term that encompasses other domains such as ambient, social, and artistic information visualization. In the views of Pousman et al. (2007), other visualization types at the edges of information visualization would also fit in the broader domain of casual information visualization.

Casual information visualizations are not defined by the visualization techniques used and can even employ very traditional techniques such as treemaps or node-link diagrams. What makes it different from traditional information visualization, according to Pousman et al. (2007), are the following key aspects: user population (includes many different types of users, from experts to novices), usage pattern (in contrast to what happens with traditional information visualization, which users explore intensely for a period of time with few interruptions, usage patterns in casual information visualization are more contemplative and extend for longer periods of time), data type (usually personal and not work-motivated), and insight (not analytical).

However, the term *casual* can be seen as negative and some authors, such as Viégas, Wattenberg, and Feinberg (2009), have expressed their concerns with the term. According to Viégas, Wattenberg, and Feinberg (2009, p. 1142), “the word “casual” itself seems inaccurate in describing many of the cases where people have used personally meaningful data”. Moreover, some of these visualizations that Pousman et al. (2007) would consider casual information visualization are not created using personally meaningful data. Viégas, Wattenberg, and Feinberg (2009) highlight as an example the situation in which students use Wordle[2] (a web-based tool for visualizing text in a manner similar to a tag-cloud, with an added attention to typography, color, and composition) in homework assignments. Moreover, Viégas, Wattenberg, and Feinberg (2009) also discuss that the word *casual* does not describe well the situations in which

[1] http://www.nytimes.com/interactive/2009/07/19/technology/20090719-driving-game.html
[2] http://www.wordle.net/

personally meaningful data is used.

In a previous paper, Viégas and Wattenberg (2008) referred to visualizations that originate from outside of the traditional information visualization environment as vernacular visualizations. The term *vernacular* is commonly used to refer to the use of a nonstandard language or a dialect from a region in opposition to the use of a formal or official language. It is also used to refer to the normal spoken form of a language. The fact that the term vernacular encapsulates this notion of common, of something that is used by a lot of people in a more relaxed scenario, of something that is part of an individuals personality, makes it a better fit for the definition that was given to casual information visualization. Viégas and Wattenberg (2008) base their notion of vernacular visualization in Tibor Kalman's idea of vernacular design. Vernacular design is design produced on the edges of mainstream design. It is often used as synonym of popular or folkloric design, or an antonym of erudite or formal design. Billboards, packages, and several other objects that are associated with the idea of popular culture are frequently categorized as vernacular design. In the same way that formal design appropriates vernacular design, traditional information visualization is also incorporating elements of vernacular or casual information visualization. Both vernacular design and vernacular visualization are not in any way seen as inferior or non-professional (even when produced by non-professionals).

Social and collaborative visualization

Wattenberg (2005) and Pousman et al. (2007) talk about social visualization, Viégas and Wattenberg (2007) and J. Heer, Ham, et al. (2008) talk about collaborative visualization, and Viégas, Wattenberg, and Feinberg (2009) and Dörk, Feng, et al. (2013) talk about participatory visualization. However, in general, the definition of all these terms share more or less the same features. The definition that seems to be farther away is social visualization, by Pousman et al. (2007). Although Pousman et al. (2007, p. 1146) begin their explanation of what is social information visualization by explaining how social information is all around us and how "articles are collaboratively written and images and songs are shared, sampled, and remixed", they later focus their definition of social information visualization on the visualization of social processes, social networks, and social situations. The examples cited are visualizations of communities and social networks (PeopleGarden (Xiong and Donath, 1999) and Vizster[3]), and of personal data or data typically produced by one user (Themail (Viégas, Golder, et al., 2006) and tag clouds). Likewise, Wattenberg (2005) describes a web-based visualization applet, named NameVoyager[4], which allows users to explore information on the baby name trends in the United States since 1900. However, Wattenberg (2005) also references the fact that users explored the visualization collaboratively and mainly used the comments section to do it.

This description of what is social visualization by Wattenberg (2005) is closer to most definitions of *collaborative visualization*. In 2005, in the research and development agenda for visual analytics Cook and Thomas (2005), collaboration was mentioned as one of the grand challenges in the field. Collaboration has become a necessity because the datasets are getting bigger, more complex, ill-defined, and broadly scoped (Isenberg et al., 2011). The common

[3] http://vis.stanford.edu/jheer/projects/vizster/
[4] http://www.babynamewizard.com/voyager

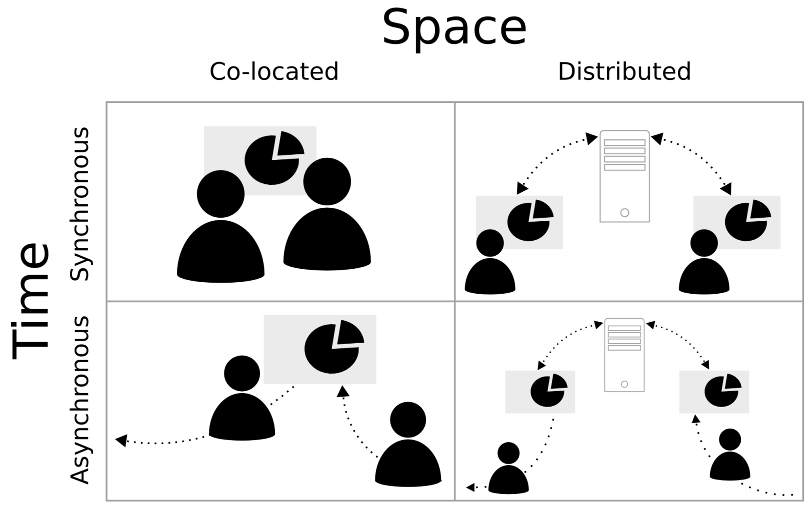


Figure 3.4: Four different collaborative visualization scenarios.

characteristics of modern day datasets made it very difficult for a single individual to fully understand all the insights that the data can provide. Therefore, being able to share what one learned about the data and to have access to what patterns others found is important for the full exploration of the dataset.

One of the earliest definitions of collaborative visualization, by Raje et al. (1998), has experts as the target audience. However, the collaborators in collaborative visualization do not need to be experts to have a valid contribution to the data analysis. Not only non-experts are able to learn from expert collaborators, but they also can provide a different and valid point of view on the data. Likewise, there are definitions that limit the use of collaborative visualizations to the field of scientific visualization, but there are several other communities that do collaborative exploration of visualizations, and that contribute to the subfield of collaborative visualization with new techniques that improve collaboration.

Based on the general definition of information visualization, that focus on computer-supported visual representations, Isenberg et al. (2011, p. 312) chose to define collaborative visualization as "shared use of computer-supported, (interactive,) visual representations of data by more than one person with the common goal of contribution to joint information processing activities." According to Isenberg et al. (2011) and J. Heer and Agrawala (2008), collaborative visualization can occur in two different *space* scenarios (collocated or distributed/remote), and two *time* scenarios (synchronous and asynchronous), as seen in Figure 3.4. As argued by J. Heer and Agrawala (2008), most research on this topic is focused on synchronous scenarios, both co-located and remote, but Isenberg et al. (2011) refutes that out of the 34 papers on collaborative visualization published in the three IEEE VisWeek conferences until 2010 (VIS[5] since 1990, IEEE Information Visualization (InfoVis) since 1995, and IEEE Visual Analytics Science and Technology (VAST) since 2005) only nine cover co-located scenarios. Both **synchronous and asynchronous co-located collaborative visualization** usually involve shared displays (J. M. Heer, 2008), however it can also be done using single-display technology.

[5] http://ieeevis.org/

According to (Isenberg et al., 2011), synchronous co-located collaborative visualization is commonly done using single-display technologies such as interactive walls and multi-touch tabletop displays but these can be transformed into multi-display environments when integrated with mobile and wireless devices. Although these scenarios of collaboration could be done using non-tech/non-interactive visualizations, in the views of J. Heer, Ham, et al. (2008, p. 92) "novel display and interactive technologies, including wall-sized and tabletop interfaces, introduce new possibilities and challenges for co-located collaborators."

J. Heer, Viégas, et al., 2009, p. 88 report that remote collaboration primarily focused on synchronous interaction "such as shared virtual workspaces and augmented reality systems that enable multiple users to interact concurrently with visualized data." Similarly **synchronous distributed collaborative visualization** also makes use of shared virtual workspaces and video conference systems.

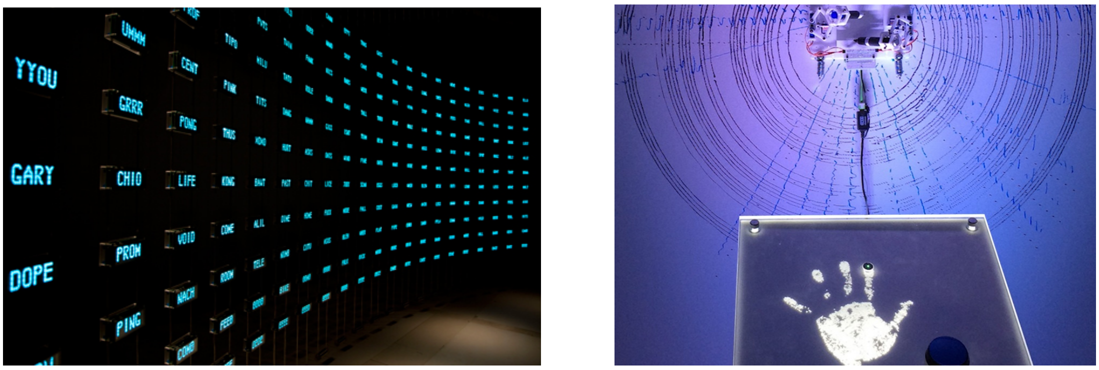
In *Voyagers and Voyeurs: Supporting Asynchronous Collaborative Visualization* J. Heer, Viégas, et al. (2009) focus on **remote (or distributed) asynchronous collaboration**, which is the most common kind of collaboration over the Web. J. Heer, Viégas, et al. (2009) see great potential in asynchronous collaborative visualization, specially in group-oriented analysis, which seems to improve decision-making. J. Heer (2006) observed that when people interacted with *Vizster* (a visualization of the social network Friendster) in a group they would spend more time exploring the visualization and the exploration would be deeper and more nuanced. They would even "issue challenges to each other, such as finding the path to a particular shared friend from the current view" (J. Heer, 2006, p. 2). According to J. Heer (2006), Wattenberg (2005) also reached similar conclusions while measuring the use of *NameVoyager*. Isenberg et al. (2011, p. 5) believes that *Many Eyes*, with its web-based design, is also "an example of an asynchronous, distributed collaborative visualization tool: collaborators access the website using their browsers through the Internet from different places and at different times."

Despite its popularity, until 2009, there was still not enough research on asynchronous collaboration for interactive visualization (J. Heer, Viégas, et al., 2009; Viégas and Wattenberg, 2006). According to (J. Heer, Viégas, et al., 2009, p. 88), contrarily to what would be desirable, "images of the visualization are transferred as printouts or screenshots, or included in word-processing or presentation documents". Although this scenario that (J. Heer, Viégas, et al., 2009) describe is still common, meanwhile "a trend towards collaborative data analysis and exploration has emerged in information visualization" (Isenberg, 2009, p. 2). For Isenberg (2009) visualizations such as *ManyEyes*, *iCharts*, *Verifiable.com*, or *Swivel* are a few examples of the boom in collaborative data analysis, taking advantage of a tendency that humans have had for a long time: social interaction around data.

Artistic visualization

Artistic information visualization (also known as data art) has become quite popular in the last 20 years, being featured in museums and as art installations. *Listening Post*[6], a visualization by Mark Hansen and Ben Rubin that presents a grid of screens that show the

[6] http://modes.io/listening-post-ten-years-on/



(a) *Listening Post* by Mark Hansen and Ben Rubin

(b) *Heart Bot* by Sid Lee

Figure 3.5: Examples of artistic information visualization

content of conversations happening on the Internet in real-time, and *Heart Bot*[7], an interactive visualization that collected the heart rate of the people who interacted with the visualization and drawn this data on a wall, are some examples of artistic information visualization.

Because it is based on actual data, it is sometimes seen as the new realism, the artistic representation of current reality. However, this is not an easy concept to define. The definition proposed by Viégas and Wattenberg (2007, p. 183) seems to be the most appropriate: "artistic visualization are visualizations of data done by artists with the intent of making art." Contrary to want happens with more pragmatical kinds of visualization, artistic information visualization is more sublime, being meant to have a deep impact in the user (Kosara, 2007) and often involving more intangible data such as personal experiences and individual opinions (Dörk, Feng, et al., 2013). For Manovich (2002) the representation of personal subjective experience is even more interesting and challenging than turning abstract and impersonal data into a beautiful and meaningful representation. Furthermore, artistic information visualization does not even need to be beautiful or aesthetically pleasing, and does not need to be efficient nor useful. Usually the main goal is to spark the curiosity of the user or raise awareness to an issue, and the data is used as the vehicle. According to Kosara (2007), the data also provides proof that the issue the artist wants to expose is real and valid.

Sometimes it is very hard to distinguish artistic information visualization from other types of visualization that simply have aesthetic corners when visualizing the data. This distinction has ignited several discussions online among the information visualization and the visual analytics communities (*Teradata, David McCandless, and yet another detour for analytics*[8]; *Does GE Think We're Stupid?*[9]; *What Is the Best Response to Bad Practices?*[10]). More vibrant or aesthetic approaches to information visualization are seen as a threat to the interpretation of the data and as harmful practices. However, the truth is that these visualizations that are closer to art or to infotainment are often very successful with the users, even if they are less effective in presenting the data. Moreover, these approaches can also point towards strategies that can

[7]http://sidlee.com/en/work/HEARTBOT/HEARTBOT
[8]http://www.perceptualedge.com/blog/?p=935
[9]http://www.perceptualedge.com/blog/?p=995
[10]https://www.perceptualedge.com/blog/?p=2154

be applied to more strict types of visualization that support decision-making, such as scientific visualization and visual analytics, in order to make traditional information visualization more appealing.

3.3 Chapter Summary

There is no consensus regarding the many names of *visualization* and the term is used to mean different things, in different contexts. There are also several terms (Information visualization, data visualization, infographics, etc.) that are used interchangeably, with definitions that overlap. Mixing up all these terms and descriptions can lead to dubiety.

In chapter 3 — The many names of visualization — I attempt to clarify the use of the terms Information, Knowledge and Data Visualization as well as all of the other terms that fall under the umbrella of visualization. I begin by defining *data*, *information* and *knowledge* based on previous work (Ackoff, 1989; M. Chen et al., 2009; Rowley, 2007) carried on the many variants of research on visualization. Through the analysis of this literature, I formulated the general idea that guides the use of these three terms throughout the thesis, which is: *Data is transformed into information, which in turn provides knowledge.*

I then reviewed the differences between the different visualization communities and the terms they chose to use: Scientific visualization (subsection 3.2.1), Visual analytics (subsection 3.2.2), Data visualization (subsection 3.2.3), Knowledge visualization (subsection 3.2.4), and terms used at the edges of visualization such as casual, social, collaborative, and artistic visualization (subsection 3.2.5). As can be seen in Figure 3.6, there are several terms used, they all have different characteristics, and there are possibly many other more that have not yet been formally identified. These terms represent types of visualization that range between *not readable/not recognizable* and *readable/recognizable*. These visualizations can be *sublime*, *utilitarian*, or something in between.

Figure 3.6: Taxonomies comparison

Resorting to a principle associated with 20th century modernist architecture and industrial design, Lima (2009), in his "Information Visualization Manifesto" argues that "Form Follows Function" and adds that, due to the incongruent nature of data, form has to follow a purpose or revelation and not the data. What all of these types or ways of doing visualization have in common is the fact that they all have their specific purpose and that is what shapes the way they visualize the data.

Chapter 4

The role of interactivity

If we interpret information visualization as "the use of computer-supported, interactive, visual representations of abstract data to amplify cognition" as does Card, Mackinlay, and Shneiderman (1999, p. 7), it is almost impossible to discard the role of interactivity. Nonetheless, many support a different definition that stands on the fact that interactivity is not always necessary to have a successful visualization and that interactivity can sometimes negatively affect the understanding of the data. Few however would deny that interactivity has several benefits specially when the datasets are quite large. Moreover, taking into account that information visualization is deep-rooted in the computer science community it makes sense to approach interactivity as a key element and to explore its possible benefits to the field.

Interactivity has been utilized in information visualization with several purposes. The more common are: 1) making the data more engaging or playful and 2) showing the data in manageable portions, for instance by partitioning it, either by browsing or by querying. According to Keim (2002) having the data in smaller portions is particularly important when exploring large datasets. Doing so facilitates both the understanding and the analysis of the data because the degree of complexity is reduced. By employing interactivity techniques, visualization creators try to give the users the ability to properly explore the data and find appropriate answers to their questions. Providing ways for the users to independently find the answers (exploratory visualization) often seems to be a better option than presenting answers to what the creator believes are the users' questions (explanatory visualization), not only because it is difficult to predict what the questions will be but also because visualization is a discovery tool and limiting its potential to provide insights is a mistake.

As argued by Dörk, Carpendale, et al. (2011) serendipity is one of the important aspects of the information seeking process. Serendipity is a term coined by Horace Walpole in 1754 to describe the discoveries done by the Peregrinaggio di tre giovani figliuoli del re di Serendippo (The Three Princes of Serendip), the characters of a Persian fairy tale translated into Italian by

Christophero Armeno. According to Walpole the princes were “always making discoveries, by accidents and sagacity, of things which they were not in quest of” Remer (1965, p. 20). Having this in mind the modern definition of the word is: the phenomenon of making pleasant discoveries or finding valuable things by change.

People often find relevant information without actively seeking, either through friends and family or skimming through media. Additionally, they seem to enjoy these serendipitous encounters and usually are driven by curiosity into exploring further. Keyword search and filtering interfaces can be a threat to serendipitous information encounters, but strategies such as “similarity-based suggestions and visual information surrogates” (Dörk, Carpendale, et al., 2011) can be used to promote these encounters. Studies such as the one by Thudt et al. (2012) hint to the fact that information visualization can enhance serendipity and even point to visualization design goals for promoting serendipity through information visualizations: multiple visual access points, highlighting adjacencies, flexible visual pathways, enticing curiosity, and playful exploration.

Interactivity is a complex issue and there is still little empirical evidence about its efficacy in terms of improving understanding of the data. Moreover, there is still few research that points out guidelines of how to incorporate it successfully and that proves that playable visualizations are indeed more enjoyable and popular among users. In order to study the impact that interactivity has in information visualization it is important to understand what types of interaction techniques are currently being used in the field and to have a framework to help discuss and evaluate interaction techniques. Although several interaction techniques taxonomies have the specific case of information visualization in mind, most existing taxonomies do not include new interaction techniques such as gamification. Therefore, I propose a new taxonomy based on previous research (Figueiras, 2015). After conducting an extensive review of popular visualizations and their interactive capabilities, I proposed eleven categories of interaction techniques: filtering, selecting, abstract/elaborate, overview and explore, connect/relate, history, extraction of features,reconfigure, encode, participation/collaboration, and gamification.

4.1 Interactivity versus interaction

Before starting the discussion about interactivity it is important to understand the distinction between interactivity and *interaction*. The two terms are often used interchangeably and loosely, but there is a subtle difference. According Aigner (2011), although interaction is considered as a valuable asset concise definitions of it and of interactivity are rarely ever given because it is often assumed that these are simple concepts.

According to Sedig et al. (2012, p. 13), because “the suffix “ity” is used to form nouns that denote the quality or condition of something”, interactivity is the quality or condition of interaction. In other words, interactivity is the capability to act on what will show up on the screen and interactions (and interaction techniques) are the actions that the user performs and the feedback that the system gives to those actions (Liang et al., 2010). Additionally, interaction techniques will affect interactivity. However, defining interactivity becomes even more complex because interactivity can be seen as a product or as a process.

Stromer-Galley (2004) argues that interactivity-as-process entails human interaction and is a process of communication. According to Stromer-Galley (2004), this kind of interactivity can occur between people or between people through mediated channels. For Aigner (2011) it is the intangible concept of a process that takes into account the process of active discourse, the user's tasks and goals, and the interaction context.

On the other hand, interactivity-as-product occurs "between people and computers, and between computers through software, hardware, and networks" (Stromer-Galley, 2004, p. 391). Interactivity-as-product is concerned with the features of a tool or medium that provide user interaction (Aigner, 2011). For instance, if a system provides de ability to click on an hyperlink in order to retrieve a web page this is interactivity-as-product. Stromer-Galley (2004) suggests that the study of interactivity-as-product encompasses both the evaluation of the quality of the tool/website/app and the interactive features provided, and the way the users engage with those features. "Measurement of interactivity-as-product can focus on the range of interactive experiences afforded by the medium; observation of the speed or time taken to complete a task; subjective measurements of how users understand or experience such features; and influence of interactive features on perceptions of site producers or control over the information experience, or on the effects such features might have on cognitive processing, including information acquisition, memory and recall, user attention, and so on" (Stromer-Galley, 2004, p. 392).

The definition of interactivity that Sedig et al. (2012) give only concerns interactivity-as-product. Sedig et al. (2012, p. 18) divide interactivity between internal and external, internal interactivity referring "to the quality of interaction among representation space, computing space, and information space" and external interactivity referring to "to the quality of interaction among mental space, representation space, and interaction space." External interactivity, which is the target of most research in information visualization, is split between macro and micro level interactivity. Interactivity at macro level is related to how interactions are combined in order to allow the user to perform more complex tasks. According to Sedig et al. (2012, p. 23) the macro-level interactivity considerations are: "what interactions should be made available to users within a particular context?; how do interactions complement one another to accomplish tasks?; do interactions correspond with users' mental models of how interactions should work?; and should constraints be placed on the order in which interactions can be performed?" In contrast, interactivity at micro level is related to individual interactions and how these affect cognitive processes. Therefore the considerations that Sedig et al. (2012, p. 23) identify are: "what are the structural elements of an interaction?; how should action elements be operationalized?; how should reaction elements be operationalized?; and how does the operationalization of structural elements affect perceptual and cognitive processes?"

4.2 Previous interaction techniques taxonomies

The *Visual Information-Seeking Mantra* by Shneiderman (1996) is the most well known general interaction techniques taxonomy. However when we seek for a more extensive taxonomy for Infovis we find a multitude of studies (Amar et al., 2005; Dix and Ellis, 1998; Keim,

Shneiderman (1996)	Overview, zoom, filter, details-on-demand, relate, history, and extract
Keim (2002)	Dynamic projections, interactive filtering, interactive zooming, interactive distortion, and interactive linking and brushing
Yi et al. (2007)	Select, explore , reconfigure, encode, abstract/elaborate, filter, and connect

Table 4.1: Interaction techniques taxonomies

2002; Yi et al., 2007) showing that there is not a taxonomy that is consensual. According to Yi et al. (2007) defining a taxonomy is challenging and they can easy get obsolete if a new interaction technique that does not fit any of the taxonomic units is discovered.

A careful analysis of recent visualizations reveals that current taxonomies do not include newer interaction techniques that are now being introduced such as participation or gamification. Therefore, there was the necessity to evaluate the existing literature in order to propose a better taxonomy. Table 4.1 summarizes the studies that were taken into account while developing our proposed taxonomy, two that only concern interaction techniques for information visualization (Keim, 2002; Yi et al., 2007) and a more general approach (Shneiderman, 1996).

4.3 A new Interaction Techniques Taxonomy

In order to more systematically explore the purposes of interactivity in information visualization, I began with the goal of building a comprehensive list of interaction techniques. Backed up by the existing literature (Keim, 2002; Shneiderman, 1996; Yi et al., 2007), 232 visualizations popular on the web were evaluated and the types of interaction they use were

Filtering	only show me the data in which I am interested
Selecting	mark or track items in which I am interested
Abstract/Elaborate	adjust the level of abstraction of the data
Overview and Explore	overview first, zoom and filter, then details-on-demand
Connect/Relate	show me how this data is related
Reconfigure	give me a different arrangement of the data
Encode	give me a different representation of the data
History	allow me to retrace the steps I take in the data exploration
Extraction of features	allow me to extract data in which I am interested
Participation/Collaboration	allow me to contribute to the data
Gamification	show me the data in a more playful way

Table 4.2: Proposed taxonomy

studied. The visualizations are available at ReThinking Visualization [1]. From this study eleven categories emerged: filtering, selecting, abstract/elaborate, overview and explore, connect/relate, history, extraction of features, reconfigure, encode, participation/collaboration, and gamification (Figueiras, 2015).

4.3.1 Filtering and selecting

Reducing complexity is one of the major goals of introducing interactivity in visualizations. A common way to achieve this effect is by filtering the data. Filtering out uninteresting items, either by specifying a range or a condition, is a natural method of requesting data.

The most successful way to filter data is through the use of dynamic filters that allow the users to quickly see how the data representation is affected when the items of no interest to him/her are eliminated or deemphasized. The data remains unchanged and can be shown whenever the users wishes by resetting the criteria (Yi et al., 2007). Card, Mackinlay, and Shneiderman (1999) found empirical evidence of the efficacy of dynamic queries referring its advantages and disadvantages. In 1999, one of the disadvantages was that the dynamic queries approach was poorly matched with the hardware and software systems available back then. Nowadays this has been overcome and therefore dynamic queries have become extremely popular, not only in information visualization. The most successful filtering implementations are the ones that allow the immediate update of the display (Craft and Cairns, 2005). The advances in technology permitted improvements in terms of performance, and these filtering systems have become incredibly more responsive.

A simplified way to filter data and selectively hide and reveal items is a way to aid cognition that enables users to quickly focus on what really matters to them. However, long delays between the user's input and the system's response can negatively affect the whole experience and inclusively the final interpretation of the visualization.

Select functionalities can also be used to aid cognition. Being able to mark and track items or sets that are interesting becomes particularly useful when there is the possibility of changing the visual representation of the data (Yi et al., 2007) or when the data is dynamic and constantly updated. According to Yi et al. (2007, p. 1226) "rather than acting as a standalone technique, Select interaction is coupled with other interaction techniques to enrich user exploration and discovery." Select techniques also act as a filter, which instead of hiding the remaining data puts in evidence the data of interest and allows the user to see it in contrast with the other items.

4.3.2 Abstract/Elaborate: Zoom, details-on-demand, and linking

Several abstract/elaborate interactivity techniques are used in information visualization. These interaction capabilities allow the user to easily adjust the level of abstraction of the data representation to his/her interpretation needs (Yi et al., 2007). The user regulates the amount of stimuli that the visualization provides him/her by varying the amount of information that displayed or emphasized.

[1] http://rethinkingvis.com/

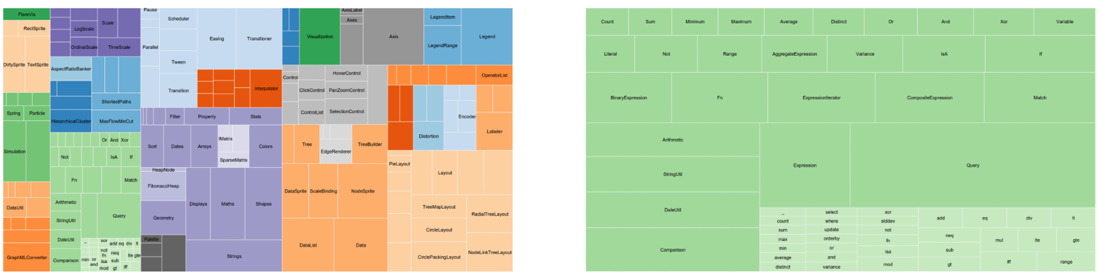
Zooming is a very common and well-known example of abstract/elaborate interactivity technique (Yi et al., 2007). Often there is some confusion with the term *zooming* due to its use as a term for generic scalar changes, rather than adjustments of vantage point. According to Craft and Cairns (2005, p. 111) it "refers to the adjustment by the user of the size and position of data elements on the screen." Zooming allows the user to see an overview of the visualization (through zoom-out) or to see a smaller, more detailed, view without fundamentally altering the representation(as it can be seen on Figure 4.1). This technique acts as a kind of filter by navigation, allowing the user to apply the technique on items of interest, simultaneously removing from view or reducing the size of items that are not of interest. As it happens with filtering, zooming helps in reducing complexity.

The use of zooming techniques in visualizations facilitate two distinct cognitive tasks:

- when zooming-in the user is being aided with the organization of the information into meaningful patterns, which is enabled by the removal of extraneous information from his/her visual field;
- when zooming-out the user is presented with hidden contextual information that was presented to him/her upon the start of the exploration but that he/she probably cannot recall.

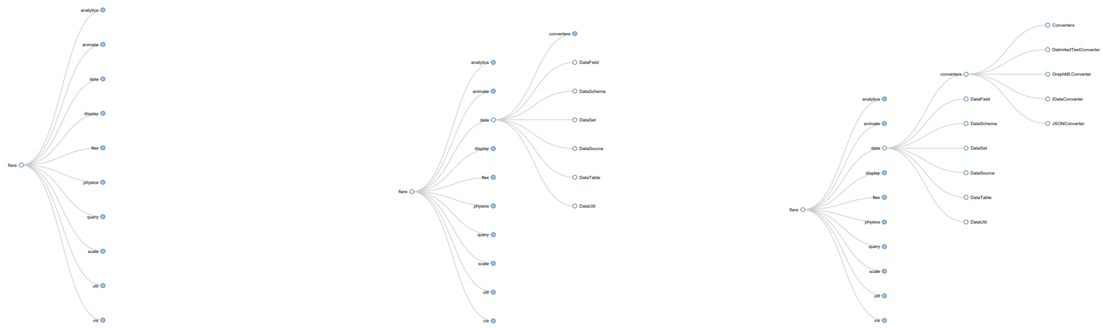
Although with different implications for cognition, these two actions are procedurally and visually symmetrical (Craft and Cairns, 2005). In other words the zooming-in action enlarges smaller data elements and the zooming-out action produces the opposite result (reduces larger data elements). *Zooming-in* enlarges small data elements in which the user is interested, removing from view or reducing the size of large uninteresting data. *Zooming out* has the opposite effect. The results are procedurally and visually symmetrical however the implications to cognition are very different.

Specially when dealing with large sets of data, it is important to provide the user with both representations. The highly compressed representation of the data (Keim, 2002) will provide an overview that will reveal the position of the data he/she is interested within the whole information space, will reveal outliers and patterns, etc. The more detailed view will provide the data in manageable inputs (Craft and Cairns, 2005), without the noise of data



(a) Overview of the treemap (b) View once the light green area is clicked

Figure 4.1: D3 zoomable treemap



(a) Tree with two levels (b) Tree with three levels (c) Tree with four levels

Figure 4.2: D3 collapsible tree

that is not of interest for the user. Having zooming options allows the user to have the best of both types of representations in the same visualization. "The objects may, for example, be represented as single pixels on a low zoom level, as icons on an intermediate zoom level, and as labeled objects on a high resolution" (Keim, 2002, p. 105).

However, zooming is only successful when it preserves the user's sense of position and context. If there is not a smooth transition between levels of zooming or if the user's input does not translate adequately his/her interpretation may be affected. According to Shneiderman (1996, p. 339), "a very satisfying way to zoom in is by pointing to a location and issuing a zooming command, usually by clicking on a mouse button for as long as the user wishes."

Details-on-demand is another type of abstract/elaborate interaction. This technique consists of getting additional details upon the selection of an item or group. As stated by Craft and Cairns (2005, p. 112), "the details-on-demand technique provides this additional information on a point-by-point basis, without requiring a change of view."

There are several ways to provide the user details-on-demand on a visualization but one of the most common techniques is by providing drill-down options. Drill-down operations are very common in tree visualizations, to which they provide the functionality of only showing the levels or sub-trees that are of interest to the user (as seen in Figure 4.2). This functionality allows the limitations of screen space and visual complexity to be overcome, while maintaining the general representational context.

Another popular details-on-demand technique is the use of tool-tips or pop-ups. This interactivity technique, often provided on mouse-hover or click, allows the user to access detailed information about an item (Yi et al., 2007), which usually would not be easily shown in the visualization. According to Segel and J. Heer (2010), details-on-demand is one of the types of interactivity common in narrative visualization. These annotations, often overlooked in information visualization evaluation despite of its important role, can be textual, graphical, and even social/participatory (Hullman and Diakopoulos, 2011). They can provide backstories that not only help in the level of engagement of the user but also provide relevant details. Annotations are also useful to focus the users attention on a specific area of the visualization (Hullman and Diakopoulos, 2011).

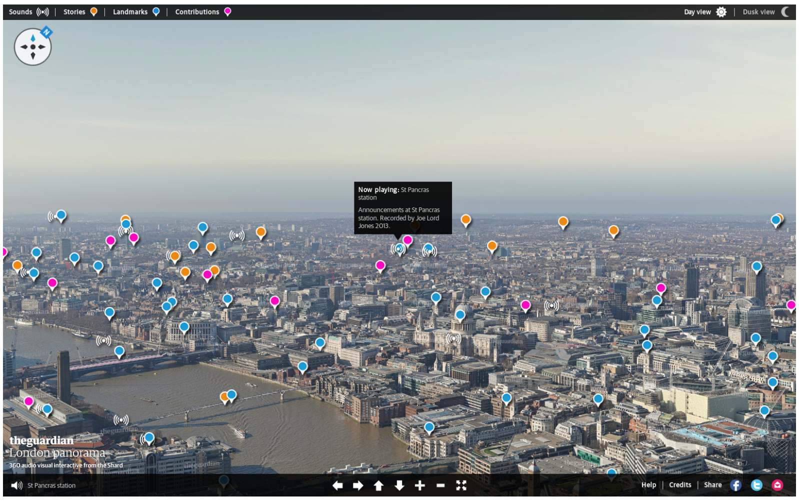


Figure 4.3: *The view from the Shard: a new and expanded panorama of London*

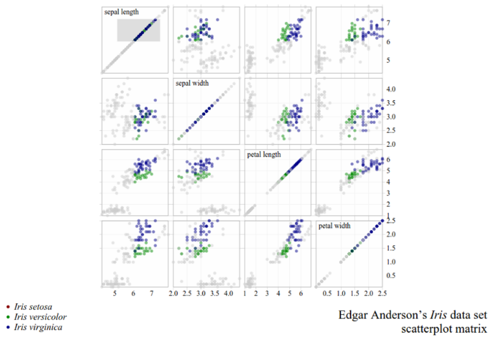
Linking is a technique that is not often regarded as a details-on-demand operation. Linking can be used to give access to external information, as it is the case of links (which reference data that the reader can access directly by clicking on it), or (as referenced by (Keim, 2002)) to give access to a different visualization method.

4.3.3 Overview and explore

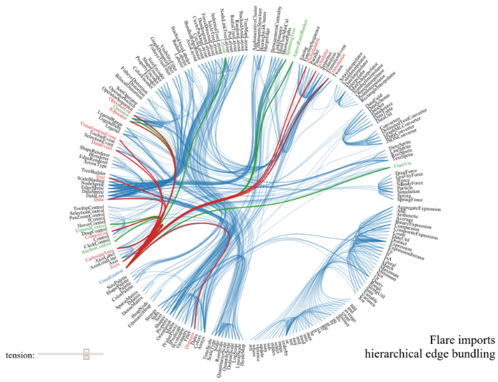
Although it is useful to provide the user with detailed information it is also important to allow the user to have an **overview** of the entire collection. Actually, according to the *Visual Information Seeking Mantra* (Shneiderman, 1996) it is better to overview first, because the overview gives the user the general context necessary to understand the dataset as a whole. That will allow the user to more easily identify patterns and themes in the data (Craft and Cairns, 2005). According to Craft and Cairns (2005), even the overall shape of the visualization can give insights about the information that is encoded. Further examination possibilities can be added by introducing any of the abstract/elaborate techniques cited in Subsection 4.3.2.

Due to the complexity and size of most datasets, visualization creators often opt for showing only a limited number of items at a time. View/screen limitations and fundamental perceptual and cognitive limitations in human information processing also force creators to reduce the amount of information shown (Yi et al., 2007). However, this information should still be available for exploration in order to enable users to examine a different subset of data and consequently get insights derived from the comparison of data.

Explore interactions provide this possibility. According to Yi et al. (2007) explore techniques show new data by making these enter the view and removing other, instead of making complete changes. As reported in the survey by Yi et al. (2007), the most common type of explore interactions is *panning*. This interactivity technique consists of the movement of a camera across a scene or the opposite, and in computer assisted visualizations "is often achieved by a special mode where the user grabs the scene and moves it with a mouse or by simply altering the view via scrollbars" (Yi et al., 2007, p. 1227). An example of this technique can



(a) D3 scatterplot



(b) D3 hierarchical edge bundling

Figure 4.4: Examples of the use of the connect interaction technique

be seen in The Guardian's *The view from the Shard: a new and expanded panorama of London* [2] visualization. In this visualization, which can be seen in Figure 4.3, the viewer is able to smoothly move the viewing focus from a position to the other, either by using the direction buttons on the bottom of the screen or by grabbing the image and moving the mouse.

4.3.4 Connect/relate

Connect, also referred to as relate, is an interactivity technique that enables viewing relationships between the data items. These relationships can be shown by highlighting links between the items that are already represented in the visualization or even by showing items that are relevant to an item that the user has interest in and that were previously hidden (Yi et al., 2007). According to Craft and Cairns (2005, p. 112), "supporting discovery of relationships is particularly important where comparisons need to be made among the characteristics of different data objects in the display."

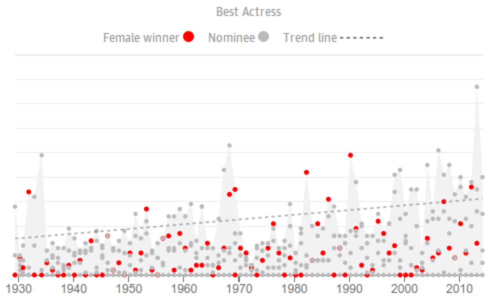
In Figure 4.4a the user is able to compare the data of interest for him/her by selecting specific data items in the first scatter plot for example. The same data items will be highlighted in the other scatter plots and the items that were not selected will be deemphasized. Even though the color coding helps in finding the data of interest in the different views displayed, it would be difficult for the user to do comparisons if he was not able to highlight the data of interest. There would be too much noise.

Connect interactions can also be applied in visualizations that consist of a single view (Yi et al., 2007). For instance, in a chord diagram, such as the one in Figure 4.4b, connect interactions can be used to enable the user to highlight the connections that he/she is interested in and easily set them apart from other relationships in the matrix.

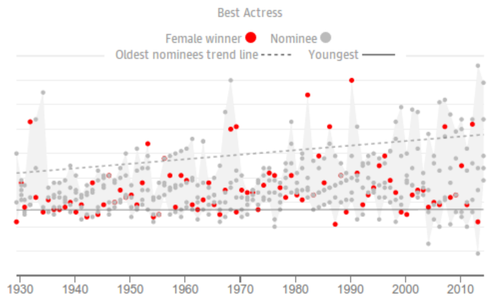
4.3.5 History and extraction of features

"Information exploration is inherently a process with many steps, so keeping the history of actions and allowing users to retrace their steps is important" (Shneiderman, 1996,

[2] gu.com/p/3d4q6



(a) Age difference of nominees



(b) Actual age

Figure 4.5: Encode interactive options in *As the Oscars age, so do the nominees*

p. 340). Providing ways for the user to undo and replay his/her actions allows him/her to not only recover from mistakes in the data exploration, but also to progressively refine the exploration (Craft and Cairns, 2005). In 1996, Shneiderman (1996) pointed this interaction technique as one that is frequently disregarded in information visualization. The **history** feature is still often forgotten by visualization creators nowadays.

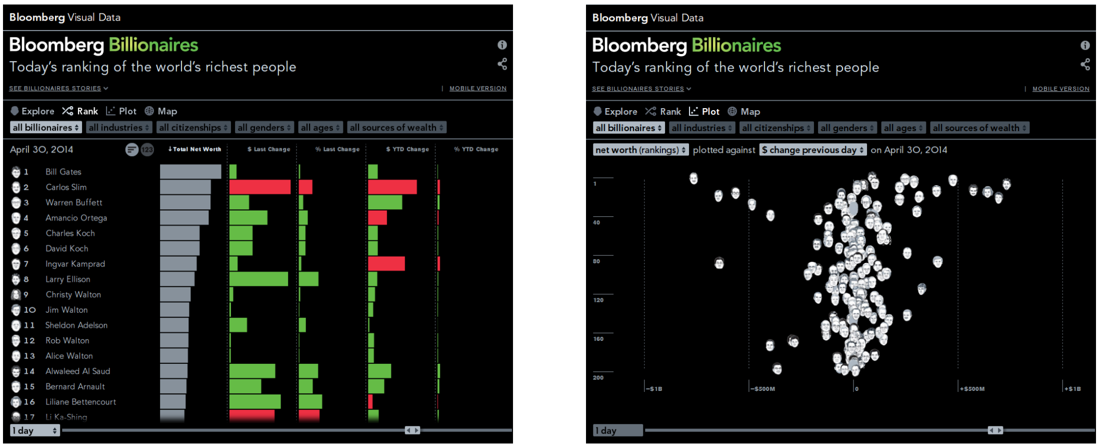
Another technique that is less common is the capability of **extraction** of important findings. Exploring the data often becomes a lengthy and complex task, therefore allowing the users to extract the data so it can be shared, dissected, or even seen in other visual representations, can reduce that complexity and result in better insights (Craft and Cairns, 2005; Shneiderman, 1996). Allowing the query parameters to be extracted can also benefit the data exploration preventing the need to repeat actions.

4.3.6 Reconfigure and encode

The **reconfigure** interactive technique provides the users with different perspectives about the dataset by changing the spatial arrangement of the representation (Yi et al., 2007). This can be done, for instance, by allowing the user to rearrange the order of columns or the rows, or by allowing the change of the attributes presented on the axis of a graph.

For example, in the visualization *As the Oscars age, so do the nominees*[3], The Guardian plots the ages of Oscar winners and nominees on a series of charts for different Oscar categories, allowing the user to filter by age difference and actual age. According to the data plotted, it is possible to see that, in recent years, the Academy has recognized an ever-broader range of ages as the gap between the youngest and oldest nominees has grown wider. The Guardian used the reconfigure technique to allow the user to choose between seeing the age difference plotted Figure 4.5a or the actual age Figure 4.5b. The first view allows to instantly perceive the trend of an ever-broader range of ages of nominees. The view by actual age allows to easily perceive the gap between the youngest and oldest nominees, which has grown wider in the last few years. The rearrangement of the data allows the user to have different perspectives that he/she probably would not have with a single representation.

[3] gu.com/p/3n7c6



(a) Rank view (b) Plot view

Figure 4.6: Encode interactive options in *Bloomberg Billionaires*

Another way to provide different perspectives on the data is by providing completely new representations. According to Yi et al. (2007, p. 1228), "in Infovis systems, visual elements serve an important role not only because they can affect pre-attentive cognition but also because they are directly related to how users understand relationships and distributions of the data items." Therefore, providing encoding techniques that allow the user to fundamentally change the visual representation can facilitate the discovery of new insights and patterns in the data. The changes in **encode** can be in terms of color, size, and even shape.

In Figure 4.6 it is possible to see the encode interaction technique applied in a visualization by the media company Bloomberg. The visualization entitled *Bloomberg Billionaires*[4] is a dynamic ranking of the world's richest people that creates a top 100 billionaires based on changes in markets, the economy, and personal assets. It allows the user to see a rank view of the billionaires on a given date and the last change in their net worth (seen in Figure 4.6a) and the same data in a plot view (seen in Figure 4.6b). While the rank view emphasizes the order of the rank, the plot view emphasizes the last change in their net worth, therefore the user will more easily see that Carlos Slim, for instance, lost a lot of money on May 23 2014 (-$520.3M). However, in this view it is more difficult to see small net worth losses or gains, such as the ones that Bill Gates had (+$110.1M). Without this technique it would be more difficult for the user to come across these insights.

The use of reconfigure and encoding techniques can be combined in the same visualization. An example can be seen in Figure 4.6a, where it is shown that, in the *Bloomberg Billionaires* visualization, the user is able to order each of the different columns by ascending or descending order, due to the use of the reconfigure interactive technique. The user can opt to see the net worth ordered by total, by last change in dollars, by last change in percentage, by year to date change (from January 1st of the current year up until the chosen date) in dollars, and by year to date change in percentage.

[4] http://www.bloomberg.com/billionaires/

4.3.7 Complex forms of interaction: participation/collaboration and gamification

Participation and collaboration are relatively new trends in information visualization. Both build on the growing will to empower users and building on the participatory culture. This neologism, which was first explored by Henry Jenkins (Jenkins, 2006), opposes to the consumer culture by transforming the user in a *produser* (Bruns, 2008) who not only participates as a consumer of content but also as a contributor to the content they consume, shaping that content. Participatory culture began as an alternative phenomenon, often seen as a parallel subculture, however it is "anything but fringe or underground today" (Jenkins, 2006, p. 2) and is being embraced by most institutions, from education and politics to media and advertising. It grew out of the blogs, forums, and mailing lists and is now an integrated feature in different domains, visualization being one of them.

In information visualization research, this inclusion of participatory culture is referred to as participation or collaboration. Mostly the different terms converge to the same definition, however both terms can also be used to characterize slightly different types of interaction. The most common definitions center on the fact that there is more than one person — usually geographically separated (Li et al., 2006) — contributing to the visualization interpretation/understanding, sharing their insights (Isenberg et al., 2011; Raje et al., 1998). A concept that usually accompanies these definitions is social data analysis (SDA), which, according to Wattenberg (2005), concerns the social interaction around data analysis. It is a version of exploratory data analysis (EDA): a rich data analytical approach to analyzing datasets, recommended as a complement to confirmatory methods, that often relies on visual methods, based on the work of John Tukey. Similarly to EDA, SDA focus on the exploration of the data beyond the formal modeling and the confirmation of previous assumptions, but "relies on social interaction as source of inspiration and motivation" (Wattenberg and Kriss, 2006, p. 551). In the analysis of NameVoyager Wattenberg (2005), found that its success might have been related with the social nature of the exploration of the web-based visualization. NameVoyager plots historical trends in baby naming and cause a buzz even among who do not find the data interesting. The creators found that the users were engaging in an intense dialogue about the visualization deeply exploring the data, helping each other discovering outliers and making causal relations, and even challenging each other to find patterns in the data. Since sensemaking is often a social process (J. Heer, Viégas, et al., 2009) — done in person or resorting to telecommunication devices — and data interpretation is frequently a group activity, it was also expected that data visualization exploration became a social activity if the means necessary to support data analysis as a social process were provided. Even if the visualization itself does not allow this sharing of insight, it might still occur separately in social networks, chats, and even offline. For J. Heer (2006, p. 1), "the immersive and compelling nature of many social visualizations arise not only from the nature and presentation of the data under consideration, but also from the social interactions, both implicit and explicit, surrounding the use of the visualization."

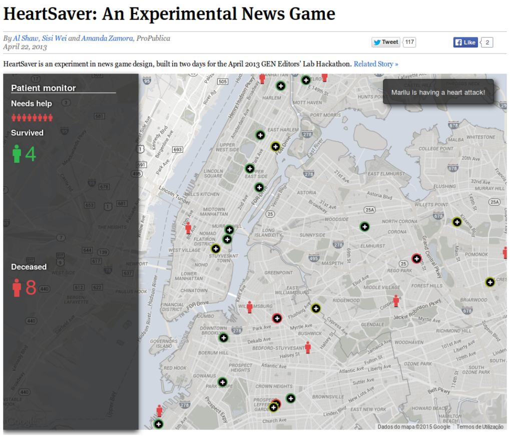
This phenomenon of wanting to explore visualizations in a social, collaborative fashion (which has inclusively been an important factor for the adoption of visualization) has been identified by several other authors (J. Heer, Viégas, et al., 2009; Kamvar and Harris, 2011; Segel and J. Heer, 2010; Viégas, Wattenberg, and Feinberg, 2009; Viégas, Wattenberg, Ham, et al., 2007; Willett et al., 2011). They have pointed out various strategies that better allow social insights, for instance tags, links, bookmarks, doubly linked discussions, graphical annotations, the traditional comments, etc. One of the biggest challenges with sharing insights specially about an interactive visualization is to share a specific state of the visualization, which is usually defined by a determined setting of filters or search parameters. Bookmarks for instance can identify a fixed state of the visualization (J. Heer, Viégas, et al., 2009; Viégas, Wattenberg, Ham, et al., 2007) so that the user can share directly with other users or even include it in their comments along with their insights. Another convenient feature is the possibility to do annotations on the visualization. This can be done by adding textual annotations that feature interesting insights communicated by the users (J. Heer, Viégas, et al., 2009), which is a very familiar action since it resembles the activity of annotating paper documents (Bargeron and Moscovich, 2003), or by highlighting and selecting specific items to include in their comments (graphical annotations).

In spite of all the perceptual and cognitive benefits that better social interactions provide, most visualizations continue to rely on simple text comments to allow users to share their insights (Satyanarayan and J. Heer, 2014). According to Satyanarayan and J. Heer (2014), although there is evidence that users are eager to share their own data stories most collaborative visualization tools provide minimal support for reusing visualizations and other types of more intense collaboration. Unfortunately, collaborative features that take full advantage of the opportunities that the web brings tend to be harder to implement, therefore techniques such as user-generated annotations and bookmarks are rare.

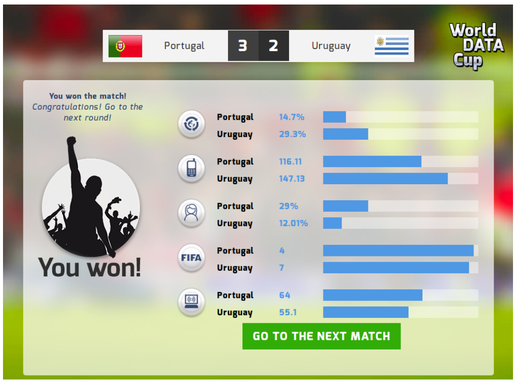
Participation/collaboration can also have a bigger impact on the visualization itself. For example, *Home and Away: Iraq and Afghanistan War Casualties*[5], the web-based visualization by CNN that maps the fallen soldiers in the wars in Afghanistan and Iraq (explored in more detail in Section 6.4), allows the users to add information about each soldier. Using iReport, CNN's citizen journalism tool that allows users to contribute pictures and videos of news stories, the users can add memories and messages about a certain soldier that they know. The fact that the users' contributions are about a subject of the dataset, and less about insights on the data as a whole or about the visual representation, makes this kind of contribution different. This kind of participation/contribution becomes part of the visualization itself, shaping it in a permanent way with changes to the data that will be visible to other users.

Gamification is one of the most complex interaction techniques that can be added to a visualization. Gamification "is an informal umbrella term for the use of video game elements in non-gaming systems to improve user experience (UX) and user engagement"(Deterding, Sicart, et al., 2011, p. 2426) and comprises a panoply of elements such as narrative context, ranks and reputations, time constraints, levels, goals, etc. This type of interaction is the least common because its production is time consuming. Even if gamified visualizations do not need to be as complex as a commercial computer game — and according to Deterding, Dixon, et al.

[5] http://edition.cnn.com/SPECIALS/war.casualties/



(a) *Heart Saver* by ProPublica



(b) *World Data Cup* by La Stampa

Figure 4.7: Examples of gamification in news media

(2011) this is what distinguishes *gamification* from entertainment and serious games — nor does the data used need to be ever changing as it happened with *Salubrious Nation* by Diakopoulos et al. (2011), the time, effort and skills required to make them are stopping its spreading. Most of the *game-y* information graphics, or playable infographics — the alike term coined by Bogost et al. (2010) — have been produced by news media (*HeartSaver*[6] by ProPublica, shown in Figure 4.7a; *World Data Cup*[7] by La Stampa, shown in Figure 4.7b; *Budget Hero*[8] by American Public Media; etc.) or marketing initiatives (*SPENT*[9] by McKinney), organizations that depend on deadlines and usually cannot invest too much time developing these projects.

Although gamified visualizations can include most of the traditional interaction techniques that were discussed previously, what makes them different is the inclusion of game mechanics or game design patterns. According to Deterding, Dixon, et al. (2011, p. 12) "neither game mechanics nor game design patterns refer to (prototypical) implemented solutions; both can be implemented with many different interface elements."

4.4 Animation

Interactivity is often paired with animation, which traditionally is seen as the change of a visual representation over time through the rapid display of sequential static images that minimally differ from each other, resulting in an illusion of movement or shape change (Chevalier, Dragicevic, and Franconeri, 2014; Robertson, Fernandez, et al., 2008). According to Chevalier, Dragicevic, and Franconeri (2014) these changes usually occur in the spatial parameters (object trajectories) and temporal parameters (object speed or pacing).

However, this is the classical definition of animation though and, according to C. Gonzalez (1996, p.27), if we want to support the idea that animation might act as an aid to understanding and decision making we cannot follow the notion that animation is passive and

[6] http://projects.propublica.org/graphics/heartsaver
[7] http://www.lastampa.it/medialab/webdoc/la-stampa-academy/world-data-cup/eng
[8] http://www.publicinsightnetwork.org/budgethero/
[9] http://playspent.org

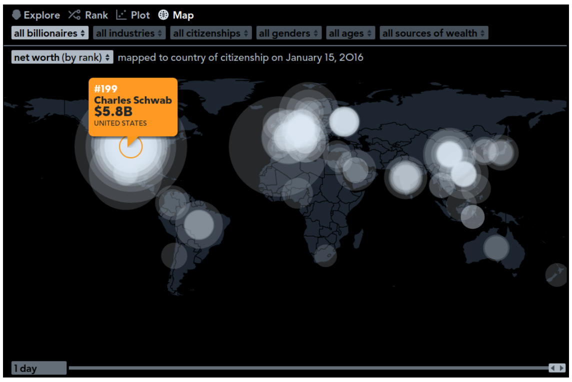


Figure 4.8: *Bloomberg Millionaires* map view

solely the "change in the positioning of the objects on a screen". C. Gonzalez (1996) sees interactive or real-time animation as a different state of animation that is dissociated with the classical/passive definition of animation. A kind that varies according to the user's actions and helps him/her to develop a more appropriate mental model of the task, not just a presentation technique. Moreover, C. Gonzalez (1996) found that animated environments that are paired with interactivity are more enjoyable and improves the accuracy of decision making tasks.

In graphical user interfaces, animation has been used with several different purposes such as maintaining relationships between different views when zooming (Bederson, Hollan, et al., 1996; Shanmugasundaram and Irani, 2008), facilitating the transition between focus and context — detailed and contextual views — (Robertson et al., 2002c; Robertson, Card, et al., 1993; Stasko and Zhang, 2000), to smooth the revelation of more content (collapse and expand) (Bederson, Clamage, et al., 2004; Bladh et al., 2005; Plaisant, Grosjean, et al., 2002; Robertson, Mackinlay, et al., 1991; Schaffer et al., 1996; Shi et al., 2005; S. Zhao et al., 2005), to help understanding the transition between different views of the same data (Bezerianos et al., 2010; Chevalier, Dragicevic, and Hurter, 2012; Elmqvist, Dragicevic, et al., 2008; J. Heer and Robertson, 2007), and even to animate changes in the content (Chevalier, Dragicevic, Bezerianos, et al., 2010). This has led to the dissemination of the idea that, in addition to helping with user engagement (Tversky et al., 2002; Wattenberg and Kriss, 2006), animation makes interfaces easier to use and more understandable (C. Gonzalez, 1996; Robertson, Card, et al., 1993).

J. Heer and Robertson (2007) identified seven different types of animated transitions commonly used in data graphics and visualization: view transformation (such as panning and zooming), substrate transformation (such as axis rescaling and graphical fisheye distortions), filtering, ordering (such as attribute values sorting and manual re-ordering), timestep, visualization change, and data schema change. While visualization transitions includes changes in color, size, and shape encodings, filtering does not and only adds and removes items in the visual representation. data schema changes can be accompanied by changes in shape encodings, because some data cannot be represented by some visual representations. For example, in the visualization *Bloomberg Millionaires*, presented previously in Subsection 4.3.6, if the user is interested in knowing the countries of the various millionaires he/she has to change the representation to the map representation (which can be seen in Figure 4.8), otherwise he/she wont be able to see the information he/she is interested in. With timestep transitions the data schema is not changed, what changes is the portion of the data that is shown (all time, a particular year, from day x to day y, etc.).

Animation's benefits in terms of directing attention to points of interest are supported by the well known notion that motion is highly effective at attracting attention due to peripheral vision's capability to easily perceive it (J. Heer and Robertson, 2007). According to J. Heer and Robertson (2007), animation's benefits in terms of inducing perceptions of causality and intentionality are also suggested by perceptual literature. Moreover there are also evidences of benefits in facilitating the users perception of object constancy when objects are transformed by changing position, size, and color (Robertson, Mackinlay, et al., 1991). There is also the common belief that smoothly animated transitions can be the solution the problems in understanding large changes in displays.

J. Heer and Robertson (2007), in their article entitled *Animated Transitions in Statistical Data Graphics*, propose a set of design recommendations for the use of animation in visualization and statistical data graphics, based on the previous work by Tversky et al. (2002) on the principles of congruence and apprehension. To ensure the principle of congruence, "according to which the content and format of the graphic should correspond to the content and format of the concepts to be conveyed" (Tversky et al., 2002, p. 247), J. Heer and Robertson (2007) recommend that:

- valid data graphics are maintained during the animated transitions by, for instance, avoiding uninformative animation (because unwarranted attributions to the data should be minimized whenever it is possible);
- consistent semantic-syntactic mappings are used, in order to aid understanding (in other words, even across different types of data graphics similar transitions should be used with similar semantic operators for the users to be able to recognize patters);
- semantic correspondence is respected, in order to avoid misinterpretations (syntax – *the visual marks and their composition* – cannot violate semantics – *the meaning of the graphic*);
- and ambiguity is avoided (in other words, the same animation should not be used for unrelated variables if this can cause confusion and "ideally, semantic operators should have noticeably different transitions" (J. Heer and Robertson, 2007, p. 1242)).

To preserve the principle of apprehension, J. Heer and Robertson (2007) recommend that:

- similar transitions are grouped, respecting the Gestalt principle of common fate;
- occlusion is minimized, because users tend to have difficulty tracking objects that get occluded during animated transitions;

- predictability is maximized, because strategies that allow the user to predict the target state of an item after viewing just a fraction of its trajectory, for example, by slowing down the animation as the ending state approaches, can reduce the user's cognitive load and improve his/her tracking capabilities;
- simple transitions are used, in order to alleviate confusion, reduce cognitive load, and improve predictability;
- staging is used for complex transitions (in other words, complex transitions should be broken up into a set of simpler sub-transitions);
- and transitions are made as long as needed, but not longer than they need to be.

However, it is still often pointed out that there is not enough research on the actual benefits of animation and that it can also bring a handful of problems. In the 90s C. Gonzalez (1996) stated the theoretically-based guidelines to design and effectively use animation were still missing. Nowadays, neither these evidences that animation makes interfaces easier to use and understand are stronger, nor, as Chevalier, Dragicevic, and Franconeri (2014) point out, is there empirical research comparing the effectiveness of different types of designs or guidelines that are not too general. Existing studies are divided between the ones that favor the advantages of animation and the ones against it.

According to Multiple Object Tracking test results from perceptual psychology literature, tests in which observers are required to mentally track specific objects moving among other objects placed with the intent to distract the observer, people can only track a maximum of 7 or 8 objects under carefully controlled situations (Chevalier, Dragicevic, and Franconeri, 2014). Contradicting the belief that object tracking performance with smoothly animated transitions is better than harsh transitions, these tests also reveal that, with the exception of extreme cases, speed has minimal impacts on performance and, unless there is extreme crowding, "object occlusion is surprisingly undisruptive when the occluding surface is clearly distinguishable from the tracked objects" (Chevalier, Dragicevic, and Franconeri, 2014, p. 2242).

Even so the studies that propose that the disadvantages overpower the benefits, for instance Tversky et al. (2002), are also able to identify advantages in specific situations such as visualization. For the particular case of animated transitions in visualization, J. Heer and Robertson (2007) show evidence that suggests that staggering, a pacing technique that introduces an incremental delay in start and stop times in order to provide a reduction in inter-elements occlusion and less overwhelming visual transitions, is an animation technique that presents benefits on the users' graphical perception of changes between static data graphics. However, more recently Chevalier, Dragicevic, and Franconeri (2014, p. 2241) presented a study with proof that introducing staggering has a negligible impact on visual tracking performance and its few benefits may be reduced by its harmful effects: "a loss of common-motion grouping information about which objects travel in similar paths, and less predictability about when any specific object would begin to move" (Chevalier, Dragicevic, and Franconeri, 2014).

There are a lot of contradictions in animation research. Even in terms of user engagement, research by Robertson, Fernandez, et al. (2008) points to the fact that (at least in trend visualization) users only slightly preferred the animated version to the static depiction. Notwithstanding, this study also points out the virtues of the use of animation for presentation.

In the particular case of visualization, animation has been used with several different purposes: to visualize trends/tendencies (Robertson, Fernandez, et al., 2008) (as it is the case of Gapminder, described previously in Subsection 2.1.4, which was designed to show trends over time), to view transformations (which consist of a change in viewpoint) (Bederson and Boltman, 1999; J. Heer and Robertson, 2007), to view how things work (Smith and Platt, 1987), to visualize the transition of data from one state to another (J. Heer and Robertson, 2007; Robertson et al., 2002b), to filter the data (J. Heer and Robertson, 2007), to switch between different temporal snapshots (Chevalier, Dragicevic, and Franconeri, 2014), etc.

Although one of the first uses of animation in information visualization was in trend visualization, animation has mainly been used in small doses to support interactivity. It has received particular attention within the study of tree visualization and Polyarchy Visualization, the visualization of multiple intersecting hierarchies (Plaisant, Grosjean, et al., 2002; Robertson et al., 2002a; Robertson et al., 2002c; Robertson, Mackinlay, et al., 1991). However, it does not matter in what proportions animation is used because its use has always been controversial, with several opinions in favor and against.

4.4.1 Arguments in favor and against of the use of animation

Arguments in favor focus on the evidence that animation may be used to improve interaction (J. Heer and Robertson, 2007), reduce task time (Robertson et al., 2002c), can help to keep users oriented (Bladh et al., 2005; Robertson, Mackinlay, et al., 1991; Tversky et al., 2002), and that in some cases it might even facilitate learning and decision making (Bederson and Boltman, 1999; C. Gonzalez, 1996). The argument that visualizations which incorporate animation are popular among users and more engaging than static visualizations is also used in favor of the use of animation, even if there is still not enough research to prove this intuition (Chevalier, Dragicevic, and Franconeri, 2014; J. Heer and Robertson, 2007).

C. Gonzalez (1996) performed one of the first studies on how animation could have benefits in terms of decision-making and although it advocates in favor of animation it shows that its use depends greatly on the user's experience, on the visual representation itself, on the quality of the animation and on its realism. Animation seems to be only one of the dimensions of information visualization that can lead to better decision making accuracy, ease of use, and enjoyability, together with other structural parts of the visualization such as its form of representation (and consequent adequacy to the data) and inclusion of interactive elements. When correctly used there is evidence of its use increasing the level of engagement (J. Heer and Robertson, 2007; Tversky et al., 2002) and it has been proven to be very popular among users. One of the reasons for animations popularity is its apparent effectiveness in attracting attention and its ability to allow users to better understand changes in objects (Robertson, Mackinlay, et al., 1991). One of the frequently cited examples is Rosling's bubble chart, because it is easier for the user to understand how much a country evolved when using animation than to have several bubble charts for each year. It is impossible for the user to notice subtle changes when he/she has to look to more than one chart. According to Ware (2004), the brain has a tendency to group moving objects and can perceive causality through animation/motion. This ability allows the brain to find patterns and to interpret the meaning of these associations.

However, several researchers (Bladh et al., 2005; Tversky et al., 2002) also point out that the use of animation can be problematic even in visualizations. One of the arguments against animation are that its inclusion is no guarantee of improved performance, mainly because "involves issues of timing and complexity that static depictions avoid, and may mislead if the animations violate the underlying data semantics" (J. Heer and Robertson, 2007, p. 1240). According to Bladh et al. (2005), animation is a double-edged sword. In their exploratory study of *The Effect of Animated Transitions on User Navigation in 3D Tree-Maps*, they observed that even though users using animation were more likely to take shortcuts and were able to complete a task in fewer steps, they were also more likely to get lost and do severe navigational errors. In another study, Bederson and Boltman (1999) explored the hypothesis that animation would not only help users navigate the information but also recall the information seen in order to later reconstruct the information space. Nevertheless, the results indicated that by using animation the users indeed remember some relationships more easily, for instance spatial location of the data, however animation did not help users to learn more complex relationships.

Even J. Heer and Robertson (2007), supporters of the use of animation in data graphics, are able to point some weaknesses: animation can be a distraction from the data/information and make the wrong information stand out, when not done carefully it can lead to false relations and incorrect interpretations, or it can be a form of what Bateman et al. (2010) categorize as *chart junk*. Moreover, they emphasize how difficult is to estimate the optimal duration of the animation, which when too slow can become boring or degrade task times and when too fast may result in increased errors. Smoothly-animated transitions have often been used in information visualization backed by the fact that perceptual constancy, introduced by Robertson, Card, et al. (1993), might "help the user maintain a sense of the true nature of the information despite the visual changes that occur during view transformations" (Shanmugasundaram, Irani, and Gutwin, 2007, p. 71). Most designers resource to their intuition when making decisions about how to introduce smooth transitions.

However, there is also little empirical evidence about its efficacy. Shanmugasundaram, Irani, and Gutwin (2007) conducted two experiments to learn more about the effect of smooth transitions on perceptual constancy in node-link diagrams. Their results show that, for the specific case of node-link diagrams, smooth transitions, in comparison with fast transitions or no transitions at all, improve the perception of connectivity and assist in maintaining structural information in viewpoint changes. Shanmugasundaram, Irani, and Gutwin (2007) were also able to find an ideal duration (0.5 seconds) for viewpoint changes of node-link diagrams that facilitates perceptual constancy.

The reality is that there is not yet enough information about how animated transitions affect perception and understanding when used in information visualization, neither effective guidelines to successfully introduce it. However, a better understanding of its impact is specially important for interactive information visualization, which requires the user to navigate in the midst of all the data, sometimes having to switch between different views. According to Chevalier, Dragicevic, and Franconeri (2014, p. 2242), "understanding how views relate to each other is an integral part of the visual exploration activity."

4.5 Chapter Summary

Interactivity and interaction have been for a long time referenced as important elements by the several research communities that study visualization (Visual Analytics (Aigner, 2011), Information Visualization (C. Chen, 2010), and Data Visualization (Murray, 2013)). Therefore, it would be almost impossible to write about visualization without mentioning it. In this chapter I write about the role of interactivity and establish how the terms *interactivity* and *interaction* will be used throughout this thesis.

However, the main contribution in chapter 4, which is also one of major contributions of the thesis, is the new Interaction Techniques Taxonomy, presented in section 4.3. This taxonomy was built in order to more systematically explore the purposes of interactivity in information visualization and consequently understand how storytelling can be incorporated in visualization. Since, interactivity seemed to be such a crucial part on the introduction of storytelling in visualization, I began with the goal of building a comprehensive list of interaction techniques that can be found in sophisticated visualizations.

Backed up by the existing literature, I evaluated 232 visualizations and studied the interaction techniques that compose them. Then I tried to apply the existing taxonomies (Keim, 2002; Shneiderman, 1996; Yi et al., 2007) to the examples analyzed However these were not exhaustive enough to classify some of the more sophisticated types of interaction techniques found. The taxonomy proposed by Shneiderman (1996) was somewhat generic, while later taxonomies that were more concerned with information visualization, such as the ones proposed by Keim (2002) and Yi et al. (2007), were still not exhaustive enough.

The new interaction taxonomy was built having in mind the two previous taxonomies that only concern interaction techniques for information visualization (Keim, 2002; Yi et al., 2007) and the more general approach (Shneiderman, 1996). As it can be seen in Figure 4.9, the new proposed taxonomy has drawn inspiration from the previous taxonomies. It includes all the types of interaction techniques identified by Shneiderman (1996) and by Yi et al. (2007), and also includes a reinterpretation of the types of interaction identified by Keim (2002). Some were combined and gave origin to a single category, which was the case of:

- Filter by Shneiderman (1996), Interactive Filtering by Keim (2002), and Filter by Yi et al. (2007) — which gave origin to **Filtering** in the new taxonomy
- Overview by Shneiderman (1996) and Explore by Yi et al. (2007) — which gave origin to **Overview and Explore** in the new taxonomy
- Connect by Yi et al. (2007) and Relate by Shneiderman (1996) — which gave origin to **Connect/Relate** in the new taxonomy

Others were incorporated in a category but as a subtype of interaction techniques, which was the case of Zoom and Details-on-demand by Shneiderman (1996), and Interactive Linking and Brushing by Keim (2002), which were all incorporated in the new category **Abstract/Elaborate**. Abstract/Elaborate is also one of the categories in the typology proposed by Yi et al. (2007). Zoom is also linked to the category **Overview and Explore** because it is often used to switch between the overview of the entire collection and the further exploration of the data points in which the user is interested. However, zoom was not considered a subtype of Overview and Explore.

I decided to combine some of the categories found in the existing taxonomies because in the examples analyzed I found that some of these would always be together: the action of Overview would always be paired with the option to further explore, and the action of connecting would always enable the user to relate the connected data points. Since I could not find the use of Dynamic Projections and Interactive Distortion in the visualization examples that I analyzed I also chose to leave out these categories from the new proposed interaction techniques taxonomy.

The main lacuna that I found in the existing taxonomies is that these do not include more sophisticated techniques that are now being introduced, such as participation or gamification. Since these are more high-level forms of interaction (in opposition to low-level interactions such as clicking or hovering) I could choose not include these forms of interaction in the new taxonomy. However, these are forms of interaction nonetheless and definitely shape the way the user interacts with the visualization, therefore it felt wrong to leave these out and reduce them to the lower forms of interaction used.

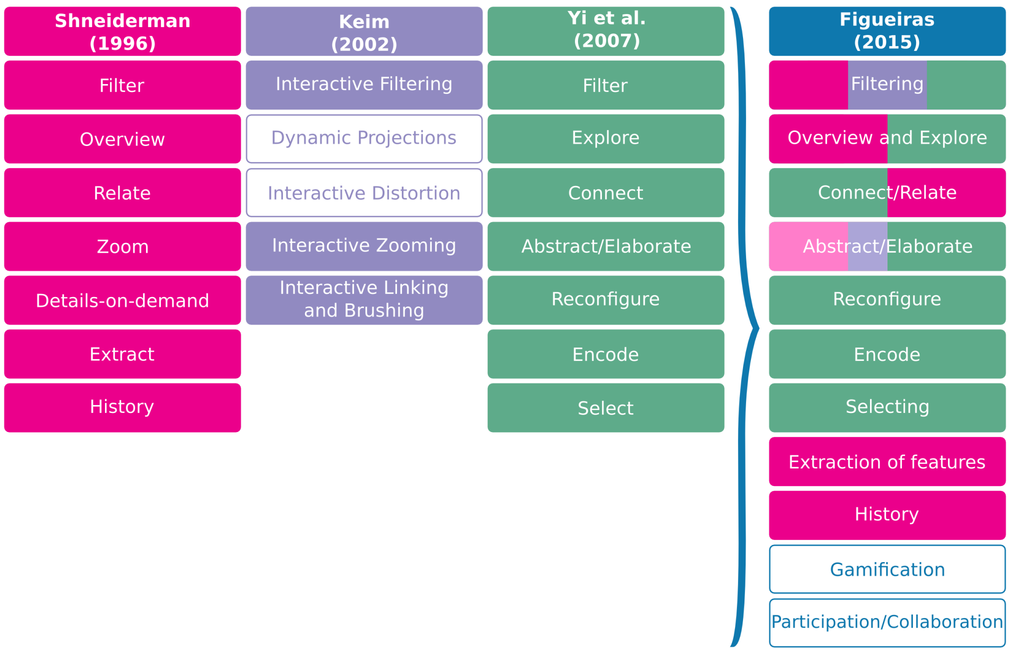


Figure 4.9: Taxonomies comparison

The new eleven categories of interaction proposed are: filtering, selecting, abstract/elaborate, overview and explore, connect/relate, history, extraction of features, reconfigure, encode, participation/collaboration, and gamification.

In section 4.4 I also talk about animation and present arguments in favor and against its use. According to Robertson, Fernandez, et al. (2008, p. 1325), animation is a "a very dramatic way to show trends, especially in a presentation" and evoke the well known presentation done by Hans Rosling, where he resorts to animation to tell a story about how a country evolved throughout the years. "The effect adds a sense of excitement to the data: the movement of the bubbles becomes a critical part of the story" (Robertson, Fernandez, et al., 2008, p. 1325). Studies have found in animation benefits in terms of memorization and knowledge acquisition (Bederson and Boltman, 1999), task time reduction and user satisfaction (Robertson et al., 2002b), and improvements in navigation (Bladh et al., 2005).

However, some research has also pointed problematic aspects of animation (Bladh et al., 2005; Tversky et al., 2002) and that it is in no way a guarantee of improved performance. According to J. Heer and Robertson (2007) issues of timing and complexity may induce users in error and cause them to make wrong interpretations of the data.

The most important conclusion regarding this topic is that animation should always be used thoughtfully because there is still not enough information on how its use affects perception. In order to study the effects of animated transitions on graphical perception, J. Heer and Robertson (2007) crafted a transitions taxonomy that can be used to better inform the design of animated transitions. The seven transition types identified by J. Heer and Robertson (2007) were: view transformation; substrate transformation; filtering; ordering; timestep; visualization change; and data schema change.

Chapter 5

A new Visualization Typology

There is a whole panoply of ways to visualize data: some more traditional such as trees, maps, and bar graph, others that are closer to art. Nevertheless, there are lots of creative and fascinating, ways to visualize data. On the Web, these visualizations can explode in terms of creativity: they can be animated, interactive and multimedia. In order to be able to understand how to tell stories with visualizations it is imperative to profoundly understand all the pieces that compose a visualization. Some work has already been carried out in this area (Bogost et al., 2010; Nichani and Rajamanickam, 2003; Segel and J. Heer, 2010), however a careful analysis of recent online visualizations revealed that the available classification schemes are not exhaustive enough to classify some of the examples of visualizations that are being done nowadays. Although it is more or less possible to try to fit a visualization on an existing category, forcing the visualizations into a category that does not fully correspond to its characteristics it is not desirable in terms of research.

Therefore, I present a typology for visualization on the web that aims to fill this lacuna that exists in this area of research. The proposed typology was elaborated through an empirical analysis and a comparative study of existing data visualizations. These examples were not randomly chosen, they were chosen through an extensive research of what is currently being done on online newspapers and magazines, blogs, scientific videos, visualization research websites, and even publicity campaigns, and more importantly what is popular and shared by Internet users. The conception of this typology also required reviewing the related work already published around this theme and this related work served as a foundation to this new classification scheme. This new typology consists on eleven different types or genres of data visualization that are not mutually exclusive: Sequential Graphic, Slide Show, Chart/Diagram, Map, Tag Cloud, Model, Drawing, Video/Animation, Photograph, Poster, and Game. For better understanding of this new typology, I also present eleven case studies that were selected for demonstrating the specificity of each genre.

5.1 Existing Typologies

The conception of this typology required reviewing the related work already published around this theme. This related work served as a foundation to this new classification scheme. The authors that are used as support for this new typology are Bogost et al. (2010), Nichani and Rajamanickam (2003), and Segel and J. Heer (2010), and their types of visualizations can be seen in Table 5.1.

Segel and J. Heer (2010) identify seven basic genres of narrative visualization that are not mutually exclusive, and that can be combined giving origin to visualizations that are more complex: *Magazine Style*, *Annotated Chart*, *Partitioned Poster*, *Flow Chart*, *Comic Strip*, *Slide Show*, and *Film/Video/Animation*. Additionally, these genres can also have messaging to provide additional information about the visualization (in the form of headlines, captions, labels, and annotations) and/or interactivity (buttons for navigation, hover highlighting and details, time sliders, ability to filter, search, drill-down the content or zoom). "There are many possible types and degrees of interactivity, though common forms in narrative visualization include navigation buttons, hover highlighting, hover details-on-demand, filtering, searching, drill-down, zooming, and time sliders" (Segel and J. Heer, 2010, p. 1146).

Segel and J. Heer (2010) also have an interesting classification for the experience the reader/viewer has while interacting with the visualization. They argue that most visualization does not fit the author-driven versus reader-driven dichotomy, and is somewhere in the middle. Therefore, they identified three categories for these visualizations: *martini glass structure*, *interactive slideshow*, and *drill-down story*. The martini glass visualization structure begins with an author-driven approach, initially using questions, observations, or written articles to introduce the visualization. Once the authors' intended narrative is complete, the visualization opens up to a reader-driven stage where the user is free to interactively explore the data. Segel and J. Heer (2010) called this structure a martini glass because a single path is given by the author of the visualization (stem), but this path gets wider and wider with the multiplicity of available paths that appear after the main story is told. This possibility to have multiple reading paths is made possible through reader-driven interactivity: linking, highlighting, filtering, etc. The interactive slideshow structure, on the other hand, has a more linear path. Segel and J. Heer (2010) emphasize that this visualization genre follows the typical structure of a slideshow but also includes some interaction on each individual slide. The authors also pointed out that this type of structure allows the user to explore particular points of the overall visualization before moving forward on the author-driven part of the visualization. Finally they present the

Segel and J. Heer (2010)	*Magazine Style*, *Annotated Chart*, *Partitioned Poster*, *Flow Chart*, *Comic Strip*, *Slide Show*, and *Film/Video/Animation*.
Nichani and Rajamanickam (2003)	*Narrative*, *Instructive*, *Explorative*, and *Simulative*.
Bogost et al. (2010)	*Graphs*, *Sequential Graphics*, *Maps*, and *Diagrams*.

Table 5.1: Existing visualization typologies

most reader-driven structure, the drill-down story, which has a structure that enables the user to choose to explore different details of the story freely in spite of maintaining the user in a general framework. Although this type of visualization gives the user freedom to choose his favorite backstory, the author of the visualization still has the responsibility to point the user the possible paths.

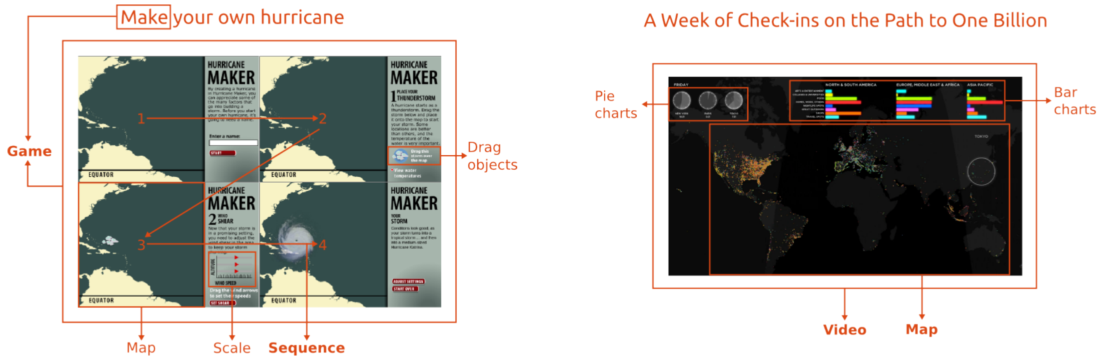
Nichani and Rajamanickam (2003) offer four categories just for interactive graphics: *Narrative*, *Instructive*, *Explorative*, and *Simulative*. "Narratives are used for telling straightforward stories, instructives provide step-by-step directions to reach a single goal, exploratives allow the user to engage in their own processes of sense-making, and simulatives allow the reader to grasp the process of a system" (Nichani and Rajamanickam, 2003, Website). The classification provided is specially tailored for interactive visualizations.

Bogost et al. (2010), in *Newsgames: Journalism at Play*, elaborate a categorization for interactive (or playable) infographics that can also be applied to non-interactive visualizations. According to the authors playable infographics can be *Graphs*, *Sequential Graphics*, *Maps*, and *Diagrams*. This last categorization seems to be the most interesting although there is not much information in the book about how the authors chose this categories and what elements make the examples that they give belong to each category. The typology offered is, as it happens in the categories proposed by Nichani and Rajamanickam (2003), only for interactive graphics, and they have an interesting way of terming as *playable data*. Their notion of playable data versus non-playable data will also be used in the typology proposed here.

5.2 The proposed Typology

The necessity to classify data visualizations emerged in the context of this larger research about how to introduce more storytelling elements in visualizations. In an early stage of the research it was evident that the classification schemes mentioned above were not broad enough to classify every example that was encountered. Therefore, it was necessary to create a classification that fitted the highest number of possible cases.

To elaborate the classification, 200 visualization examples were analyzed according to their narrative elements, reading/viewing order, visual elements, and interactive elements. Four matrices were elaborated in order to be able to mark the different elements that were being encountered in the visualization examples. In Appendix II it is possible to see the initial matrices that were done when the typology started to be developed, with some of the examples that were analyzed. In terms of narrative, the 8 elements found in the analyzed visualizations were: accompanying article, annotations, audio narration, captions, introductory text, text, title, and video narration. The 27 visual elements identified were: animation, area chart, bar chart, bubble chart, bubble map doughnut chart, exploded view, histogram, line chart, logo, map, model, network diagram, photograph, pictogram, pie chart, pyramid, scale, size representing quantity, speech balloon, table, tag cloud, timeline, tree diagram, venn diagram, video, and drawing (which was later changed to illustration). In terms of interactivity the 16 elements that were present in the analyzed visualizations were: click details, click highlights, drag objects, filtering, hover details, hover highlight, input box, link to external article, navigation buttons, object



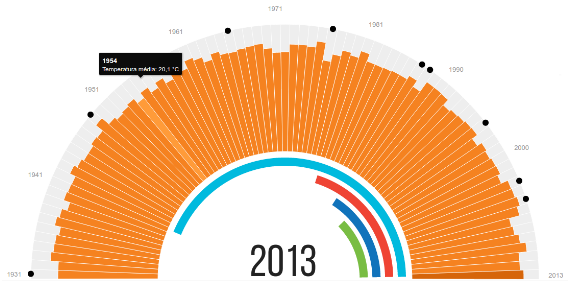
(a) *Make your own hurricane* analysis (b) *A Week of Check-ins* analysis

Figure 5.1: Notes about categories and the visual and interactive elements

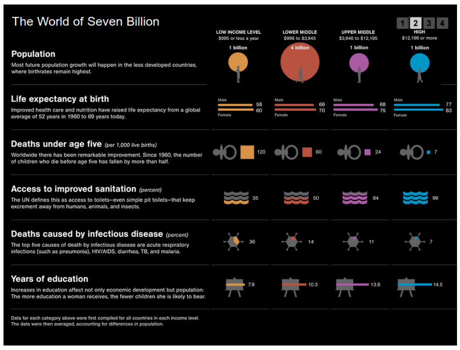
react to mouse movement, scrollbar, search, user contribution, zoom (interactive), and game mechanics (now referred to as gamification). Slide show was removed as an interactive element and the visualizations that had it were changed to include just navigation buttons. All these narrative, visual, and interactive elements are clearly defined in the glossary at the beginning of this thesis. The two reading/viewing orders possible were linear (the reading/viewing path is previously decided by the author of the visualization) and user directed path (the user has complete control on how he/she views the visualization and the author of the visualization has minimal control of the path that the user takes, therefore providing several options).

It was measured which elements were more prominent and which influenced more the interaction that the user has with the visualization (see examples in Figure 5.1). In other words, for each visualization, after all the different elements that compose that visualization were identified, it was studied which of these elements occupied a larger area of the visualization. For instance in Figure 5.1b, the map occupies almost 80 percent of the whole visualization area and therefore it is one of the most prominent visual elements. However, not only the visual elements have a big impact in the visualization. Interaction also plays a big part in how the user sees the visualization and therefore the fact that the visualization *A Week of Check-ins on the Path to One Billion*[1] is also a video influences greatly the interaction that the user can have with the visualization: he/she can only play, pause, or stop what he/she is seeing. Taking these two aspects into account the visualization, *A Week of Check-ins on the Path to One Billion* was categorized both as a Map and as a Video/Animation visualization. The visualization could be categorized simply as a Map, however the video/animation has such a big impact in the visualization that it was decided that this visualization should have a mixed categorization. The example shown in Figure 5.1a, on the other hand, although the map also occupies a large area of the visualization it was decided that the game component was far more prominent than the map visual element. It is very clear that the gamification interactive element is the highlight of the whole visualization and a sign of that if the fact that the title of the visualization, *Make Your Own Hurricane*[2] emphasizes game aspect of the visualization. Therefore, the map becomes a mere vehicle for the main characteristic of the visualization: the game.

[1] http://rethinkingvis.com/visualizations/13
[2] http://rethinkingvis.com/visualizations/44



(a) *80 anos de Verão by Público*



(b) *The World of Seven Billion* by National Geographic

Figure 5.2: Examples of a Sequential Graphic visualization and a Slide Show visualization

After taking all these elements into account and identifying the most prominent elements in each of the 200 visualization initially studied, a new typology, which is intended to be exhaustive, was formulated. The eleven different types or genres of visualizations are Sequential Graphic, Slide Show, Chart/Diagram, Map, Tag Cloud, Model, Drawing, *Video/Animation*, Photograph, Poster, and Game. The categories were based on some of the work that has already been carried out in this area and referred previously as related work in Section 5.1 by borrowing some of the types that seemed unavoidable. Still this categorization and the analysis of the examples have some subjectivity. These genres vary mostly in terms of visual and interactive elements that the genre has and are not mutually exclusive, being possible to combine genres to classify more complex visualizations. From the 200 visualizations that were studied for the development of the typology only 18 percent are classified as being mixed typed and most of the cases were visualizations classified as video mixed with other type. In most of these examples the video plays such a huge part on the way that the viewer/user accesses the information that it was impossible not to classify them as a video also.

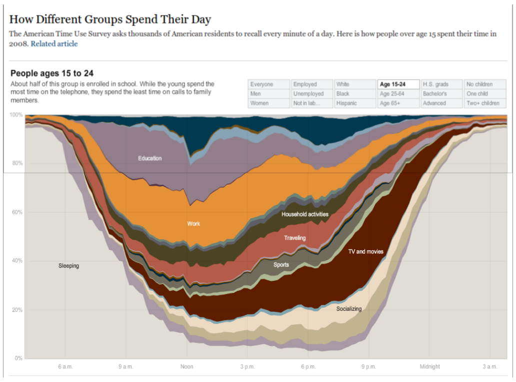
5.2.1 Sequential Graphic

A Sequential Graphic, a category also in the classification scheme of Bogost et al. (2010), is a chronological graphic. This type of visualization is usually represented through a timeline, as Público's visualization entitled *80 anos de Verão*[3], which can be seen in Figure 5.2a, but it can also be a cause/effect kind of sequence, as the South Florida Sun Sentinel's *Make your own hurricane* (that was not categorized as a Sequential Graphic solely because the gamification was a more prominent characteristic), which was presented previously in Figure 5.1a. This kind of visualization is very useful to show the user events that are influenced by previous actions.

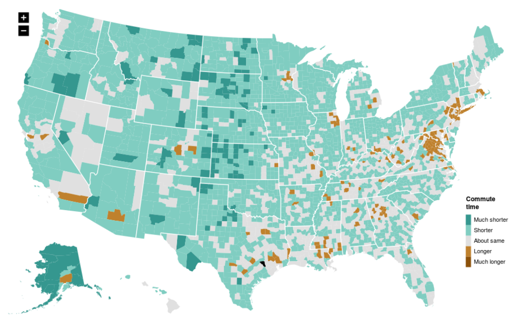
5.2.2 Slide Show

A Slide Show has an order imposed by the author but it is not necessarily a chronological sequence. An example of a Slide Show is National Geographic's *The World of Seven Billion*[4]

[3] http://rethinkingvis.com/visualizations/200
[4] http://rethinkingvis.com/visualizations/3



(a) *How Different Groups Spend Their Day* by The New York Times



(b) *Compare Worst and Best Commutes in America* by FlowingData

Figure 5.3: Examples of a Chart/Diagram visualization and a Map visualization

visualization, which can be seen in Figure 5.2b. Slide Shows can be composed of photographs, information, or even charts and the only interaction that it provides is moving forward or backwards. The fact that the Slide Show imposes a strict order on the reading/viewing order of the visualization has a great impact on the interaction. Therefore, even if the visualization has other visual elements such as charts or maps, it is classified as a Slide Show. Even though a slide show is mostly a container, it is responsible for the overall impact of the visualization.

5.2.3 Chart/Diagram

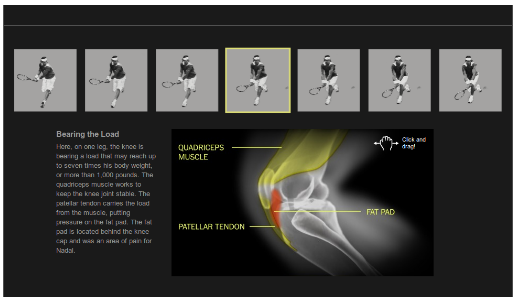
Chart/Diagram is a classic visualization used extensively on the media, on research, etc. Its heavy usage may be related to the fact that these visualizations are very easy to understand and the public is very familiar with them. This category includes every type of chart/diagram, from the common bar charts to the Venn diagrams. An example of Chart/Diagram visualization is The New York Times' *How Different Groups Spend Their Day*[5] visualization, which can be seen in Figure 5.3a. A visualization will only fit this category if the main focus of the visualization is the chart/diagram, because although there are many visualizations that include these visual elements as an added value, for example, on Foursquare's video/animation *A Week of Check-ins on the Path to One Billion*, which will be explored in more detail in Section 6.8, most of them have an element that is more prominent than the charts/diagrams and these are used just to

5.2.4 Map

A Map is also a classic visualization. It can be tangible (it represents where things are placed and tries to mimic as truthfully as it cans the real world), such as the visualization entitled *Compare Worst and Best Commutes in America*[6] by FlowingData (which can be seen in Figure 5.3b), or like Minard's *Carte figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813*, explored in detail in Subsection 2.1.2, an intangible map that represents not only information about physical places but also about events that occur on those places. In this map, Minard captured not only geographical information, like the

[5] http://rethinkingvis.com/visualizations/25
[6] http://rethinkingvis.com/visualizations/219

(a) *Every Noise at Once*



(b) *Nadal's Knee* by The New York Times

Figure 5.4: Examples of a Tag Cloud visualization and a Model visualization

direction taken by the army as they traveled and the location the troops passed through, but also non-geographical information, like the size of the army as troops died from hunger and wounds, and the temperatures they experienced. Intangible maps can be fictional maps or interpretations of the world that do not resemble a traditional map.

As in the previous category, sometimes visualizations have a map but this is not the main element, therefore that visualization cannot be considered a Map. Of all the visualizations analyzed the one that closely resembles an intangible map in the visualization *Lisbon's Blood Vessels*[7], a visualization on which the traffic of Lisbon is portrayed exploring metaphors of living organisms with circulatory problems – the thickness, the color, and the length of the vessels are excited by the number of vehicles and average velocity in each road.

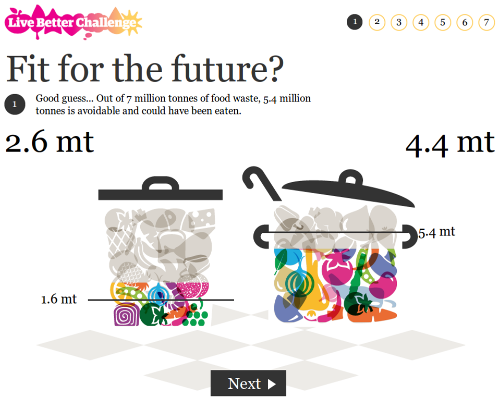
5.2.5 Tag Clouds

Tag clouds, such as *Every Noise at Once*[8], which can be seen in Figure 5.4a, are very popular online, being sometimes used as navigation on blogs and websites using hyperlinks. This type of visualization is a representation for text, more specifically keywords or tags, and can be useful to show which occur more often, the size of the word being the differentiating factor. However, according to Harris (2011, Website), tag clouds can lead to fake conclusions and be harmful: "When looking at the word cloud of the War Logs, does the equal sizing of the words *car* and *blast* indicate a large number of reports about car bombs or just many reports about cars or explosions? How do I compare the relative frequency of lesser- used words? Also, doesn't focusing on the occurrence of specific words instead of concepts or themes miss the fact that different reports about truck bombs might be use the words *truck*, *vehicle*, or even *bongo* (since the Kia Bongo is very popular in Iraq)?"

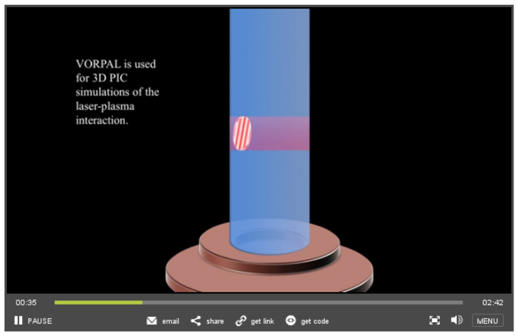
5.2.6 Model

A Model is a more technical visualization that was previously almost exclusive to scientific visualization and now is being used by news media to explain complex topics. This is the

[7] http://rethinkingvis.com/visualizations/107
[8] http://rethinkingvis.com/visualizations/166



(a) *Fit for the future?* by The Guardian



(b) *Laser-Plasma Accelerator* by Wired

Figure 5.5: Examples of a Drawing visualization and a Video/Animation visualization

case of The New York Times' Nadal's Knee[9] visualization, which can be seen in Figure 5.4b. The visualization sheds a light on the details of the Rafael Nadal's style of play and how his two-handed backhand put a lot of stress on his injured left knee by presenting 3D models of his knee that the user can rotate to better understand the pressure points.

Nowadays Model visualizations are getting more and more popular because there are more people with the expertise to create 3D models. This type of visualization is particularly good to show projects of buildings or to describe complex processes. It usually includes exploded views of the object or detailed instructions about processes.

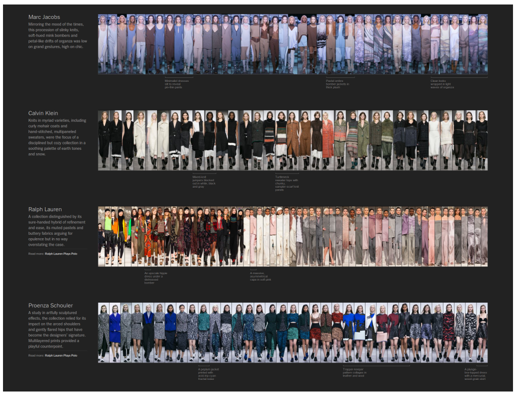
5.2.7 Drawing

A Drawing is a type of visualization that combines information and illustration. It is a very popular visualization on the printed press and it is quite common online too. In order to be effective and become a data visualization and not a mere drawing, this type of visualization has to combine the illustration with another type of visualization such as, for example, charts/diagrams like the visualization *Fit for the future?*[10] by The Guardian, which can be seen in Figure 5.5a, or even videos, like in the New York Times' visualization *Three Generations of a Family Under One Roof*[11], that shows the lives of three generations that live in a building in Chinatown. The visualization consists of a drawing of the building and videos of each family positioned on the floor where that family lives.

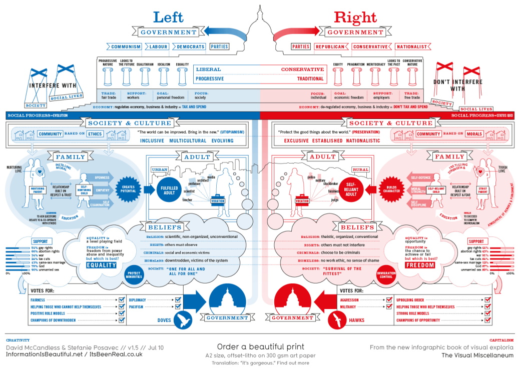
5.2.8 Video/Animation

The Video/Animation is obviously the category of the types of visualization in which there is a video or animation that is the main part of the visualization. As in the previous category this type of visualization has to include other types of visualizations in other to be

[9] http://rethinkingvis.com/visualizations/89
[10] http://rethinkingvis.com/visualizations/143
[11] http://rethinkingvis.com/visualizations/36



(a) *Front Row to Fashion Week* by The New York Times



(b) *Left vs Right* by Information is Beautiful

Figure 5.6: Examples of a Photograph visualization and a Poster visualization

considered a Video/Animation visualization. An example of a Video/Animation visualization is *Laser-Plasma Accelerator*[12] by Wired, which can be seen in Figure 5.5b.

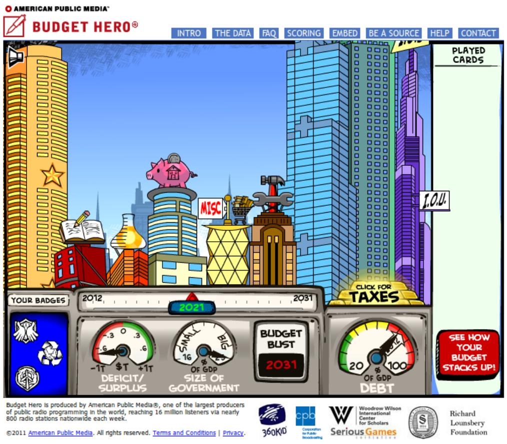
5.2.9 Photograph

Another category that depends on other types of visualizations in order to be considered one is the Photograph. This is the least common type of visualization. One example of this type of visualization that was analyzed in this investigation, and therefore one of the case studies presented in Chapter 6, is the *100 Years of World Cuisine*[13] visualization, which is explained in more detail in Section 6.9. Another example of a Photograph visualization is The New York Times' *Front Row to Fashion Week*[14], which can be seen in Figure 5.6a. The photos are really thin, only revealing a glimpse of the main colors of the outfit and on mouse hover the user can see the full look. Additionally another representation is presented: the color hues were abstracted to create small swatches of different designers, which is an interesting way to compare the shows.

5.2.10 Poster

The Poster genre is inspired on the Partitioned Poster category by Edward Segel and Jeffrey Heer. These are usually static visualizations that include both textual and graphic elements, although it may be either wholly graphical or wholly text. Since posters are typically both eye-catching and informative, it is commonly used to advertise products. An example of a Poster visualization is *Left vs Right*[15] by Information is Beautiful, which is shown in Figure 5.6b. This visualization is a concept-map exploring the Left vs Right political spectrum, which tries to shed a light on what do the left and right actually stand for: their views on family, society, culture, beliefs, and government.

[12] http://rethinkingvis.com/visualizations/41
[13] http://rethinkingvis.com/visualizations/9
[14] http://rethinkingvis.com/visualizations/75
[15] http://rethinkingvis.com/visualizations/42



(a) *Budget Hero* by American Public Media

(b) *Could you be a medallist* by The Guardian

Figure 5.7: Examples of Game visualizations

5.2.11 Game

The Game category might be the least common of all categories but it is a type of visualization that can potentially be very appealing to the public. According to Ian Bogost, Simon Ferrari and Bobby Schweizer, "even if they are not games quite like Pac-Man or The Sims, infographics can become game-like, exploiting the properties of games in numerous ways: to encourage the manipulation of information for replayability, to allow pleasurable engagement with a system, or to invite exploration" (Bogost et al., 2010). Although there are not many visualizations of this type there are some interesting examples such as *SPENT*[16](one of the case studies presented in Section 6.11), *Budget Hero*[17] (shown in Figure 5.7a), and *Could you be a medallist*[18] (shown in Figure 5.7b). *Budget Hero* is a game-y visualization that allows the user to control where tax dollars go. The user can see how his/her priorities line up with the realities of managing billions of dollars of federal spending and see the impact of his/her decisions in the balance between policy choices and the financial stability of the country. *Could you be a medallist* is a retro 8-bit style game that allows the users to compete in a game version of the 100m, 10km, 100m freestyle swim, and bicycle road race, and to compare his/her results against the all-time greats. The user can also see if his/her time would have ever earned a place on the podium.

5.3 ReThinking Visualization

ReThinking Visualization is a website built with the intent to be a resource for anyone interested in information visualization. The project's main goal is to help building a better understanding of all the pieces that compose a visualization and to help detecting patterns in visualizations across different areas or disciplines. It was built as the ground work for this

[16] http://rethinkingvis.com/visualizations/59
[17] http://rethinkingvis.com/visualizations/38
[18] http://rethinkingvis.com/visualizations/38

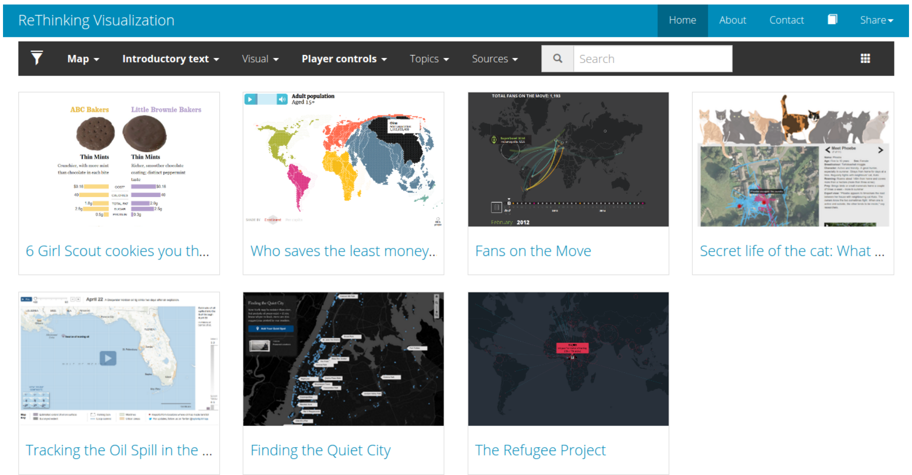


Figure 5.8: Filtering on *ReThinking Visualization* homepage

exhaustive typology that intends to be sufficiently exhaustive to classify all the different examples of visualizations that are being created nowadays. Being able to classify visualizations is important to evaluate the progress of the maturing visualization field, to help focus and direct future research, and to help creating better visualizations that make use of the elements that are essential for a visualization to be successful.

ReThinking Visualization consists of a collection of 298 visualizations (including the case study examples shown in Chapter 6) chosen through an extensive research of what is currently being done on online newspapers and magazines, blogs, scientific visualization, visualization research websites, even advertising campaigns, and more importantly what is popular and shared by Internet users. The analysis of these successful examples not only can help understand the reasons for its success but it can also inform the creation of new visualizations.

In the visualizations that were studied after the typology was created new elements were identified. Whenever a new element was found all of the visualizations previously studied were checked to see if that element was present. However, it was not necessary to alter the typology because the visualizations that were studied afterwards fitted the categories in the typology perfectly.

The narration element that was added to the list of elements that compose a visualization was external link. This narrative element was found in 65 of the analyzed visualizations and played an important role in providing context to the information provided by the visualization. To the list of visual elements 30 new elements were added: arc diagram (1 visualization found), cartogram (2 visualizations found), chord diagram (1 visualization found), circle graph (1 visualization found), circular bar chart (2 visualizations found), cloropleth (22 visualizations found), color matrix (1 visualization found), decision tree (1 visualization found), dot map (21 visualizations found), dot plot (5 visualizations found), flowchart (3 visualizations found), graph (1 visualization found), heat map matrix (4 visualizations found), non-ribbon chord diagram (1 visualization found), parallel coordinates (1 visualization found), parallel sets (1 visualization found), polar-area chart (1 visualization found), population pyramid (1 visualization found), radar chart (1 visualization found), scatter plot (18 visualizations found), stacked area chart (7

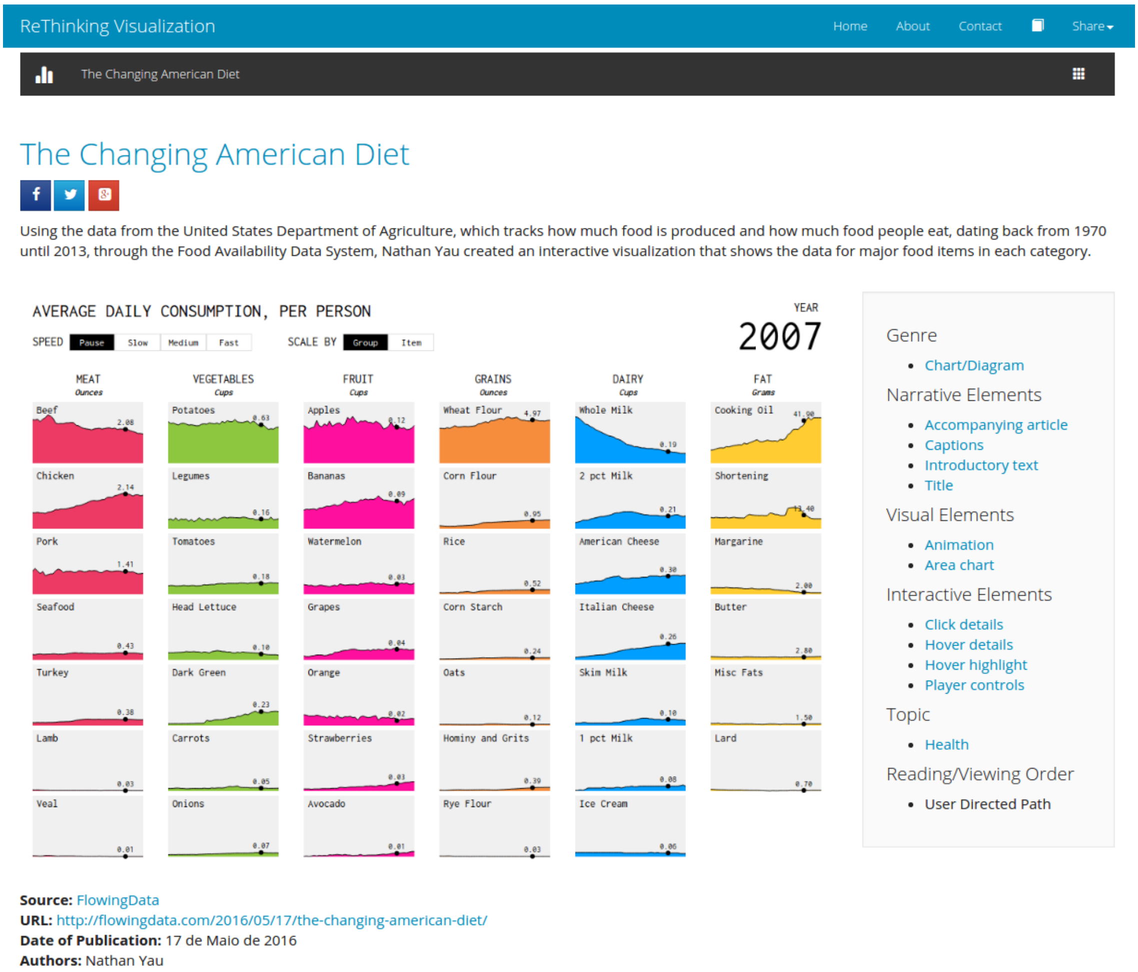


Figure 5.9: *ReThinking Visualization* visualization page

visualizations found), stacked bar chart (1 visualization found), streamgraph (1 visualization found), sunburst chart (2 visualizations found), timetable (1 visualization found), tooltip (114 visualizations found), transit map (1 visualization found), treemap (2 visualizations found), wheel (5 visualizations found), and zoom (visual) (58 visualizations found). In the list of interactive elements 6 new elements were included: combo box (37 visualizations found), link to the raw data (19 visualizations found), (36 visualizations found), player controls (30 visualizations found), scroll activated animations (8 visualizations found), slider (26 visualizations found), and virtual reality (1 visualization found). The definitions of these narrative, visual, and interactive elements were also added to glossary at the beginning of this thesis.

The user can check out the most recent visualizations added on the home page. However he/she can also use the navigation to filter and check only the visualizations with certain characteristics. As it can be seen in Figure 5.8, the user can, for instance, filter all the visualizations that are of the genre Map, that have the narrative element introductory text, that have the interactive element click detail, and that are on the topic death. Being able to view several visualizations that use the same elements promotes the discovery of patterns. For instance, by filtering the visualizations choosing to see only visualizations of the genre Map and experimenting to filter by the different interactive elements, it is possible to see that hover details is the most common interactive element in Map visualizations. The user can also user the search to find the visualization by its title. This facilitates the search for visualizations that the user already knows.

As it can be seen in Figure 5.9, the individual page of every visualization presented in the website includes an analysis of its components:

- the genre (according to the proposed typology);
- the narrative, visual, and interactive elements;
- the topic;
- the reading/viewing order;
- the source, the URL, the publication date, and the authors.

A short description of the visualization is also provided, in order for the user to quickly understand what the visualization is about. The visualization page also includes share buttons so that the user can share the visualization through Facebook, Twitter, or Google Plus.

5.4 Chapter Summary

The main contribution in this chapter, and one of the thesis major contributions, is the new Visualization Typology. This typology was built because, in order to be able to understand how to tell stories with visualizations, I believed it was necessary to understand all the pieces that compose a visualization.

There were previous attempts to build classification schemes for visualizations but when I tried to apply these to the visualization examples that I collected these classification schemes revealed either to be too generic (Segel and J. Heer, 2010) or not exhaustive enough (Bogost et al., 2010; Nichani and Rajamanickam, 2003) to classify some of the more sophisticated examples of visualizations that are created nowadays. For instance, none of the existing typologies had a category that would be adequate to classify the visualization *100 Years of World Cuisine*, which uses different sized kitchen containers filled with blood to represent the amount of deaths that resulted from 25 conflicts (Khmer genocide, Sudanese civil war, Biafran war, etc.) from 1915 to the visualization's date of publication. The same can be said for the visualization *Cruise Control*[19], which raises awareness to the precautions to have in order to ensure the safety of both passengers and crew members by identifying on the drawing of a boat the potential health and safety hazards for each section (stateroom, pool, kitchen, stairs, and decks).

Although if hard pressed it would be possible to fit a visualization to an existing category, forcing the visualizations into a category that does not fully describes its characteristics is not desirable for research. For instance, using the generic typology by Nichani and Rajamanickam (2003) one could fit model-like visualizations, such as *Nadal's Knee* visualization or *A Final*

[19] http://rethinkingvis.com/visualizations/242

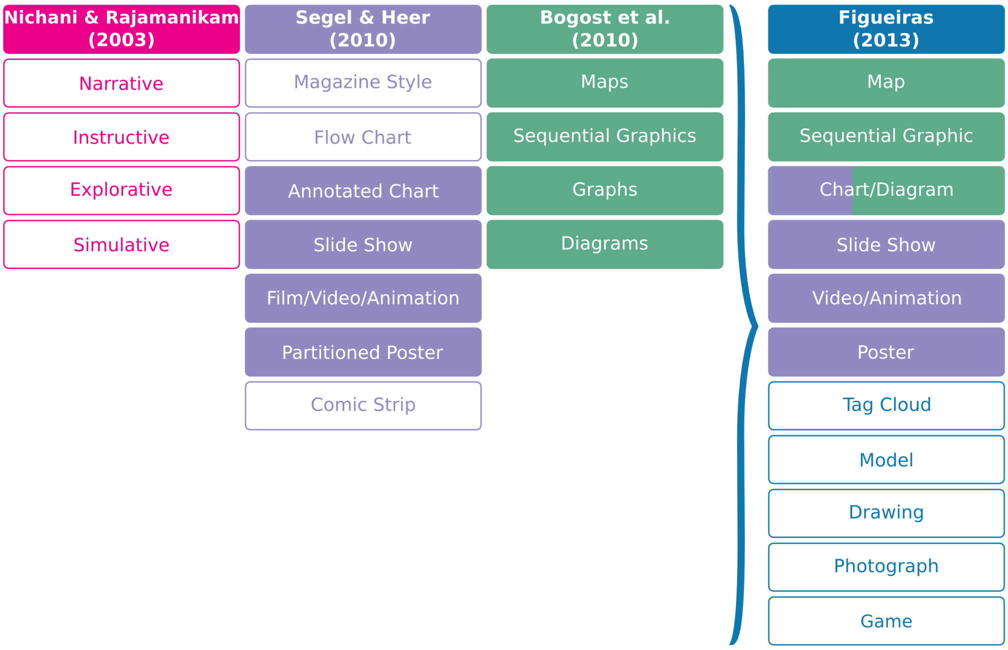


Figure 5.10: Typologies comparison

Visit[20] by The New York Times, in the *simulative* or *instructive* category, but what could be done with visualizations that do not fit any of these categories, such as the *Atlas of emotions*[21] which would barely fit the *explorative* category because it does not provide that many options to explore?

Therefore, I present a typology for visualization on the web, built through an empirical analysis and a comparative study of existing visualizations, that aims to be as complete as possible to be able to fit non-traditional visualization examples. The new visualization typology consists on eleven different types or genres of data visualization that are not mutually exclusive: Sequential Graphic, Slide Show, Chart/Diagram, Map, Tag Cloud, Model, Drawing, Video/Animation, Photograph, Poster, and Game. This new typology was inspired by the existing typologies (Bogost et al., 2010; Nichani and Rajamanickam, 2003; Segel and J. Heer, 2010) and in Figure 5.10 is represented which categories were incorporated in the new one. The typology is analyzed in detail in chapter 6 through the use of case studies for each category in the typology.

In this chapter I also present the ReThinking Visualization project. When I presented the typology at the *17th International Conference on Information Visualisation* someone asked me if the matrices that I created to study the visualization examples were available online, because they thought it could be a good resource for other researchers to identify patterns in popular visualizations. Therefore, I built the ReThinking Visualization website following the same classification used in the typology and I now have a collection of 298 visualizations analyzed and available online.

[20] http://rethinkingvis.com/visualizations/195
[21] http://rethinkingvis.com/visualizations/290

On ReThinking Visualization the user can see the most recent visualizations on the main page but can also chose to filter and only check the visualizations with a certain combination of elements. This allows the user to easily find patterns. For instance, by filtering by the genre Map and experimenting to filter by the different interactive elements, it is possible to see that hover details is the most common interactive element in this type of visualization. The user can filter by genres; narrative, visual, and interactive elements; topics; and sources.

Chapter 6

Typology Case Studies

In order to better understand what different kinds of visualizations exist and to try to understand what makes a visualization a Sequential Graphic, Slide Show, Chart/Diagram, Map, Tag Cloud, Model, Drawing, Video/Animation, Photograph, Poster, or Game, eleven examples were gathered for a more exhaustive analysis. This analysis consisted of the study of the impact of all the individual elements that compose these visualizations. The examples of visualizations in the case study come from very different sources such as online journalism web sites, and specialized blogs: *The New York Times*, *CNN*, *The Guardian*, *National Geographic*, *Information is Beautiful*, etc. There is one visualization example for each category in the typology and they were chosen because they are clear representatives of their category.

6.1 England Riots

After the shooting of Mark Duggan in Tottenham, England suffered widespread rioting between August 6 and 10 2011. To provide better understanding of the sequence of events *The Guardian* created an interactive timeline (shown in Figure 6.1), showing the most important incidents and how they spread over the different neighborhoods. *England Riots*[1] enables the user to scroll through the events, ordered by hour, and watch how the riots unfolded.

The main visual element is a vertical timeline with pictograms along, identifying different types of events (Riot, Police, Statement, Court, Fire, or Cleanup). When the visualization loads, the bottom of the vertical timeline presents the first events that occurred and the most recent events are seen far away at the top of the vertical timeline. The user can scroll towards the most recent events by dragging that timeline. However, there is also an horizontal timeline at the bottom which shows the day of the events, highlights the date of the event currently being shown in the vertical timeline, and allows the user to jump to a particular date.

[1] http://rethinkingvis.com/visualizations/2

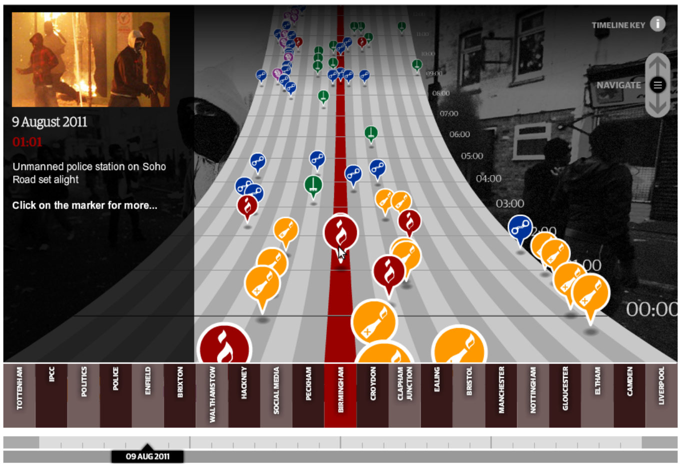


Figure 6.1: *England Riots* timeline by *The Guardian*

In order to explain the events the timeline also includes an introductory text, captions, annotations, and it also has external links to articles about single events. Therefore, taking into consideration the categories for reader/viewer experience by Segel and J. Heer (2010) this visualization would a Drill-Down Story, because the user gets extra details about the events when he/she hovers the pictograms. These details include photographs and short descriptions of the events. All of the individual elements of this visualization can be seen in Table 6.1.

England Riots, according to the proposed typology, is a Sequential Graphic and, although it leads the user to a sequential order of reading/viewing, it enables the user to explore the events in the order he/she intends, because the user can pick a date on the horizontal timeline or follow the chronological order of the events by dragging the main timeline.

England Riots	
Genre	Sequential Graphic
Narrative Elements	Annotations Captions External link Introductory text Title
Visual Elements	Photograph Pictogram Timeline
Interactive Elements	Hover details Hover highlight Link to external article Scrollbar Slider
Reading/Viewing Order	User Directed Path

Table 6.1: *England Riots* genre, elements, and reading/viewing order

6.2 The World of Seven Billion

National Geographic's *The World of Seven Billion*[2], shown in Figure 6.2, is a Slide Show visualization. However, as it is common practice in this type of visualization, it requires the use of other visualization elements, in this case a heat map matrix and several charts. Although this is a playable visualization, the interaction is very limited: the user can click to see the introductory text, click to see the different slides, and the navigation buttons are highlighted on mouse-hover. As most slide shows the order of reading/viewing is linear because,although the user can choose to view the slides in a different order, the fact that the user has no idea what slide 1, 2, 3, and 4 are about will force the user to click in the order imposed by the author of the visualization. *The World of Seven Billion* visualization only works because it only has four slides. Otherwise it would be boring for the user and he/she probably would not see every slide.

The charts are a very prominent part of this visualization and most of them are not really charts but chart-like objects representing quantities. There are pictograms representing the number of cars, of personal computers, of children, etc., divided by type of income. This information is organized in a structure resembling a table on which the columns are the different levels of income (low income level, lower middle, upper middle, and high), the first row is population, and the subsequent rows are different things that characterize the lives of the world population: life expectancy (represented through the length of a bar); deaths under age five (represented by a pacifier icon); access to improved sanitation (represented by three waves); deaths caused by infectious disease (represented by a microbe icon); years of education (represented by a blackboard icon); literacy rate (represented by a book); fertility rate (represented by a child pictogram); rate of natural population increase (represented by the inclination of a line); net migration rate (represented by an arrow pointing out or in depending on the direction of the migration); urban population (represented by a city icon) phone subscription

[2] http://rethinkingvis.com/visualizations/3

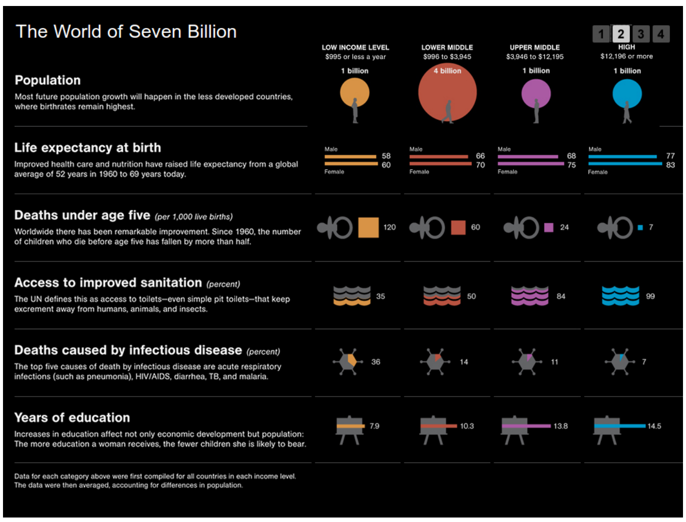


Figure 6.2: *The World of Seven Billion* by *National Geographic*

The World of Seven Billion	
Genre	Slide Show
Narrative Elements	Accompanying article Annotations Captions Introductory text Title
Visual Elements	Heat map matrix Pictogram Size representing quantity
Interactive Elements	Click details Navigation buttons
Reading/Viewing Order	User Directed Path

Table 6.2: *The World of Seven Billion* genre, elements, and reading/viewing order

(represented by a phone and a mobile phone pictogram); internet users (represented by a computer); personal computers (represented by a mouse icon); cars (represented by a car icon), and carbon dioxide emissions (represented by a factory pictogram).

In terms of narration elements, this visualization, besides the introductory text, it has captions and annotations. This types of narrations are vital on a visualization that resorts to images to expose the data. All the elements of this visualization can be seen in Table 6.2.

6.3 Death penalty statistics, country by country

Death penalty statistics, country by country[3], shown in Figure 6.3, is a visualization by *The Guardian* that accompanies an article about countries that maintain the death penalty, flowing the execution of Kim Jong-un's uncle. The visualization shows Amnesty International data on

[3] http://rethinkingvis.com/visualizations/5

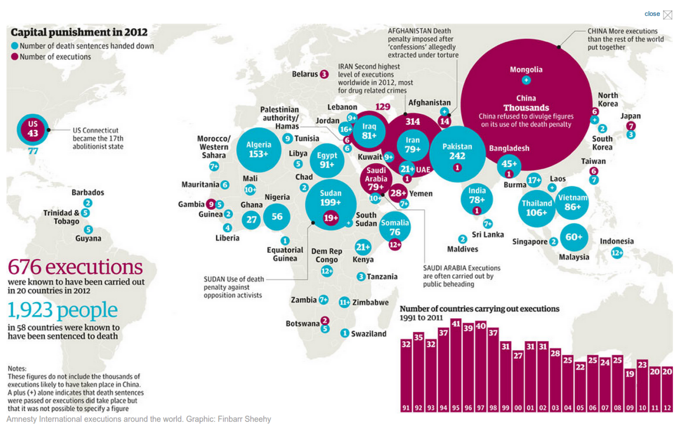


Figure 6.3: *Death penalty statistics, country by country* by *The Guardian*

Death penalty statistics, country by country	
Genre	Chart/Diagram Map
Narrative Elements	Accompanying article Annotations Captions Introductory text Title
Visual Elements	Bar chart Bubble map
Interactive Elements	No Interactive Elements
Reading/Viewing Order	User Directed Path

Table 6.3: *Death penalty statistics, country by country* genre, elements, and reading/viewing order

executions around the world. This visualization can be classified as a Map, but it can also be considered a Chart/Diagram because it is also composed of diagrams. On the bottom, there is a bar chart / timeline representing the number of abolitionist countries in contrast with the number of executing countries, since 1991 till 2010.

Since *Death penalty statistics, country by country* is a non-playable visualization it has no interactive elements. The use of some interactivity such as hover or click details would be useful on a visualization of this type, because it could add some interesting information making the data more meaningful. A hyperlink to particular execution stories would be very effective to increase the empathy between the reader/viewer and the data. The use of interactivity would also make the visualization less cluttered and consequently more appealing to the public because it would not generate so much information overload. All of the elements present in the visualization can be seen in Table 6.3.

The fact that this visualization accompanies an article makes the fact that there are not more narration elements understandable. *Death penalty statistics, country by country* has an introductory text, captions, and annotations that indicate small information like names of countries, dates, and some trivia (such as the fact that china has more executions than the whole world put together, or the fact that in Saudi Arabia executions are often carried out by public beheading). However, there are several examples of successful Chart/Diagram visualizations that accompany articles but still include several narrative elements in the visualization, making them interesting even when the user does not read the article. These narrative elements are usually provided through the use of interactivity. This is the case of *Across U.S. Companies, Tax Rates Vary Greatly*[4]. In this Chart/Diagram visualization by *The New York Times* when the user hovers the circles in the visualization he/she not only sees what company that circle represents but also the effective tax rate, taxes paid, and earnings of that company. It would be very difficult to represent all this extra information without resorting to interactivity.

[4] http://rethinkingvis.com/visualizations/88

6.4 Home and Away: Iraq and Afghanistan War Casualties

Some American media have done, since the beginning, an extensive coverage of the wars in Afghanistan and Iraq. With the immense amount of data that they were able to gather about the deaths of soldiers in both wars, *CNN* and *Stamen Design* have launched *Home and Away: Iraq and Afghanistan War Casualties*[5], shown in Figure 6.4, an aesthetically-pleasing interactive data visualization that enables viewers to track each trooper's birth places against a map with the location where they died in Afghanistan or Iraq. Through the careful analysis of this data visualization it is possible to see that it fits the Map category, in which the size of the bubble represents the number of deaths on that place. Instead of viewing the map the user can opt to view a table with the deceased soldiers, but this option of visualization it is clearly not the main way to visualize this, being only useful if the user is someone that is looking for a soldier in particular. This visualization allows the audience to learn about soldiers not only from the U.S but also from other countries.

There are two separate lists of casualties, Afghanistan and Iraq, and the data can be browsed either by map or table. In the map view two parallel maps are presented: a dot map of the places of birth of the soldiers and bubble mapof either Afghanistan or Iraq with the causalities. Complementary graphics are provided along the bottom to show trends of age, location, and date of death. Casualties can be filtered using the criteria on these complementary graphics. It is also possible to search soldiers by name through a search box.

This data visualization allows viewers to click on the points that represent a soldier that died in one of these wars and learn more details about the life of each soldier on their profile page. The visualization is integrated with the *iReport*[6] platform, *CNN*'s user-generated news community, allowing family and friends of the deceased to tell their personal stories, share memories, and pay tribute. It goes beyond news reporting: it is a platform for participation.

[5] http://rethinkingvis.com/visualizations/6
[6] http://edition.cnn.com/specials/opinions/cnnireport

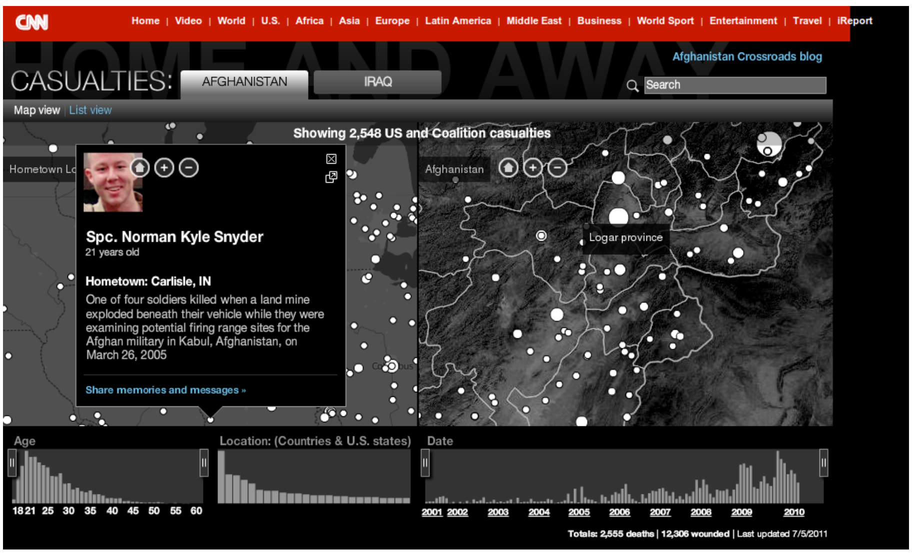


Figure 6.4: *Home and Away* by *CNN*

Home and Away	
Genre	Map
Narrative Elements	Annotations Captions Introductory text Text Title
Visual Elements	Bar chart Bubble map Dot map Photograph Table Tooltip Zoom (visual)
Interactive Elements	Click detail Drag objects Filtering Hover details Hover highlight Search User contribution Zoom (interactive)
Reading/Viewing Order	User Directed Path

Table 6.4: *Home and Away* genre, elements, and reading/viewing order

This possibility of social interaction that *Home and Away: Iraq and Afghanistan War Casualties* provides to the audience is notable, since it provides more annotations that help building the story of each soldier's life and death. The content produced by the family and friends of the deceased enrich the visualization, saving journalists a huge amount of work that would take probably years to do. These personal stories make the user feel more empathy with the data. Once a user clicks on a dot, that dot is no longer just a marking on a map or part of the data, it becomes a story. The additional navigation through the graphics of age and location also make viewers feel more connected to the data because people will instinctively click on those who were the same age or from the same town. This feeling of connection with the people that compose the visualization greatly enhances the viewers experience with the visualization as a whole. "The user identifies with stories the map traces, constructing relevant meaning from fragments" (Bogost et al., 2010, p. 54).

In terms of visual elements, which can be seen in Table 6.4, *Home and Away: Iraq and Afghanistan War Casualties* has photographs of the deceased soldiers, small bar charts with their ages, places where they are from, and years of death. It also has small pictograms to indicate the zoom, and home. There are also tooltips that pop up whenever the user clicks on one of the bubbles in the map (with information about that fallen soldier) and a simple animation that appears while the map is loading.

Since this is a playable visualization it has some interactive elements: navigation buttons, scrollbar, search, filtering, zoom, click-able details, highlighting and popping details while hovering, and the possibility to drag objects (in this case dragging the map to move it around). *Home and Away: Iraq and Afghanistan War Casualties* fits perfectly the Visual Information-Seeking Mantra by Shneiderman (1996): Overview first, zoom and filter, then details-on-demand. It also respects most of his seven tasks of a good visualization (Overview, zoom, filter, details-on-demand, relate, history, and extract), although I believe that it fails on the last two.

Home and Away: Iraq and Afghanistan War Casualties also has important narration elements that imply that this visualization is a good example of a narrative visualization, a visualization that is able to tell stories with data. It has a title like most visualizations, captions, annotations, and an introductory text that helps the user to understand the main story (the wars in Afghanistan and Iraq).

In terms of reading/viewing order, *CNN*'s visualization of the Iraqi and Afghan war is an event reporting via geographical visualization that is not intended to be experienced in any particular order and that it does not require the viewer to interact with the whole data, the reading/viewing path is completely driven by the user. Maps like this one encourage the viewer to explore the data, picking the level of detail he/she wants to know about, and constructing narratives as they go. Segel and J. Heer (2010, p. 1146) characterize the structure of visualizations like this one as Drill-Down Story, because it "presents a general theme and then allows the user to choose among particular instances of that theme to reveal additional details and backstories", it has a reader-driven approach.

6.5 What Does China Censor Online?

What Does China Censor Online?[7], which can be seen in Figure 6.5, is a simple tag cloud that only has a title and text, in this case mere disconnected words. In terms of visual elements and although it is not visible at first, this visualization has a map, because the shape that

[7]http://rethinkingvis.com/visualizations/63

Figure 6.5: *What Does China Censor Online?* by David McCandless

What Does China Censor Online?	
Genre	Tag Cloud
Narrative Elements	Annotations Text Title
Visual Elements	Map Tag cloud
Interactive Elements	No Interactive Elements
Reading/Viewing Order	User Directed Path

Table 6.5: *What Does China Censor Online?* genre, elements, and reading/viewing order

the words form is the map of China. However the map is secondary, a mere detail on the Tag Cloud (it might not even be noticed unless the user is familiar with the shape of the country), therefore it would not be considered a Map visualization. This visualization is not playable, therefore has no interactive elements. Table 6.5 presents all the elements of this visualization.

Although this kind of visualization is considered bad (Harris, 2011), since it has so many problems in giving the right emphasis to the data, What Does China Censor Online? works well for its purpose. Maybe this is because it was man made and not automatically generated like most tag or word clouds. However, this visualization would probably benefit if a new overlay of information was added to it through the use of interactivity. Consequently it would be way more effective as a visualization and not merely beautiful.

6.6 Ground Zero Now

The *Ground Zero Now* visualization[8], which can be seen in Figure 6.6, is part of a huge collection of articles about 9/11 entitled *9/11: The Reckoning*[9], that is divided in: *The Decade*, *That Day*, *War Abroad*, *War at Home*, *Remembrance*, *Rebuilding*, *Muslims Now*, *9/11 State of Mind*, and *Portraits Redrawn*. This *New York Times* interactive graphic, part of the *Rebuilding* segment, is at the same time a Model and a Video/Animation visualization.

In terms of visual elements it has two animation videos that, amongst other elements, include a map of the Ground Zero area. Like it would be expectable in a Model visualization it has models of buildings, pictograms and illustrations, in this particular case, engineering drawings. Additionally it also provides an expanded view to better understand how the irrigation system on the Ground Zero memorial will work. Although it is considered playable since it is a Video/Animation, *Ground Zero Now* does not have any interactivity, because clicking a play button is not considered proper interactivity.

Ground Zero Now can be considered a narrative visualization since it has many narration elements. The story is told not only by the captions, annotations, and video narration, but also

[8]`http://www.nytimes.com/interactive/2011/08/30/us/sept-11-reckoning/ground-zero.html`
[9]`http://www.nytimes.com/interactive/us/sept-11-reckoning/viewer.html`

Figure 6.6: *Ground Zero Now* by *The New York Times*

by the article that accompanies the visualization. Nevertheless, the visualization can work on its own, and it did not need to be associated with an article.

Although normally videos have a linear narrative, because of the fact that this visualization is composed by two different videos and an accompanying article, considering the visualization as a whole it is possible to consider the ordering an user directed path because the user can choose what to read or view first. This visualization does not really fit any of categories that Segel and J. Heer (2010) have for the reader/viewer experience. All the elements of the visualization can be seen in Table 6.6.

Ground Zero Now	
Genre	Model Video/Animation
Narrative Elements	Annotations Captions Title Video narration
Visual Elements	Animation Illustration Map Model Pictogram
Interactive Elements	Player controls
Reading/Viewing Order	User Directed Path

Table 6.6: *Ground Zero Now* genre, elements, and reading/viewing order

6.7 How Many Households Are Like Yours?

Following the article *Baby Makes Four, and Complications*[10], which tells the story an unconventional Brooklyn family composed by a woman, her son, her sperm donor, and his lover, *The New York Times* published an interactive visualization, shown in Figure 6.7, for exploring different types of American households. Upon entering the page of the visualization, the viewer is able to choose the primary residents of his/hers (or other) household to see how the entered household compares to the rest of America's households. Pictograms are used to represent the elements chosen by the user to compose a household. The audience is first presented with pictograms that represent the set of primary residents (married couple; male/female unmarried partners; single male; single female; male unmarried partners; and female unmarried partners) that they chose and can then add secondary members of the household (child under 18; child over 18; child-in-law; foster child; parent or parent-in-law; siblings or siblings-in-law; grandchild; other relative; housemate or roommate; Roomer, boarder or lodger; and other non-relative) that will also be represented as pictograms. Complementary graphics, such as bar and area charts, are provided along the bottom to show the viewer how the number of households like the one he/she selected have changed over time, which races have more households of that kind, and what is the income of those households. The graphics update on the fly whenever the user adds or subtracts a household member.

How Many Households Are Like Yours?[11] fits two of the eleven proposed genres, although one of the types plays a major role on the visualization than the other. The visualization can be considered a Drawing, but as a form to add information it also includes charts, so it could also be considered a Chart/Diagram. However, since the charts are just complements of the main

[10] http://nyti.ms/1Be3ZrZ
[11] http://rethinkingvis.com/visualizations/1

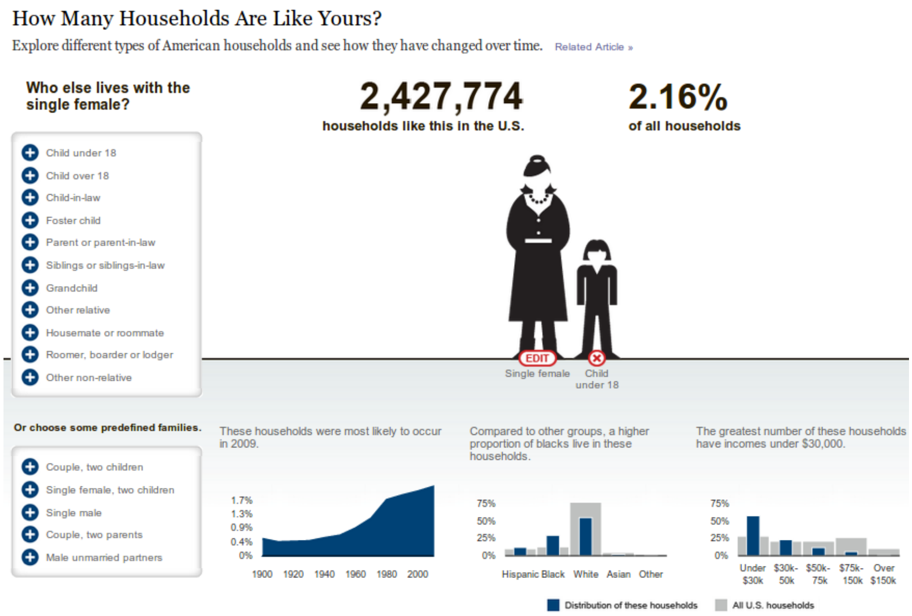


Figure 6.7: *How Many Households Are Like Yours?* by *The New York Times*

How Many Households Are Like Yours?	
Genre	Drawing
Narrative Elements	Accompanying article Annotations Captions Introductory text Title
Visual Elements	Area chart Bar chart Pictogram Tooltip
Interactive Elements	Filtering Hover details
Reading/Viewing Order	User Directed Path

Table 6.7: *How Many Households Are Like Yours?* genre, elements, and reading/viewing order

visualization, *How Many Households Are Like Yours?* was classified as a Drawing visualization.

As it can be seen in Table 6.7, the *How Many Households Are Like Yours?* visualization is composed of many different narrative elements: title, captions, annotations, and an introductory text. However, the narration element that really helps turning this visualization into a narrative visualization is the article that accompanies the visualization. Still the visualization can work on its own, and it did not need to be associated with an article. Since there is a written article to introduce the visualization, *How Many Households Are Like Yours?* is considered to have a Martini Glass visualization structure, according to Segel and Heer's categories for structure.

The reading/viewing order, according to this new typology, is directed by the user. However, the use of the question is an indicator of the author driven approach that this visualization have: the author wants the reader to try his household type first.

In terms of interactivity elements, *How Many Households Are Like Yours?* enables the user to filter the data and to access details by hovering the content. However the New York Times visualization has a main problem. One of the purposes of having data visualizations is so that people can do visual comparisons easily, but in this visualization the data cannot be compared. Only one kind of household is displayed at a time, so there is no easy way to check which kind of household is more common, to compare the viewer's family with their friends' households, etc. Most of the percentages vary so little that the number does not really mean anything to the viewer. In addition, the pictograms end up having no use other than aesthetics, their size does not mean the number of American families that are like that, or their income, or anything. The first rule that Tufte (1983) presents in *The Visual Display of Quantitative Information* is to "above all else show the data". The viewer must be given enough information in order to have answers to his/her questions, and he/she should not have to make a great effort to find the answer to their questions. The viewer must be able to filter the information but that filter should not limit his/her access to all the other data.

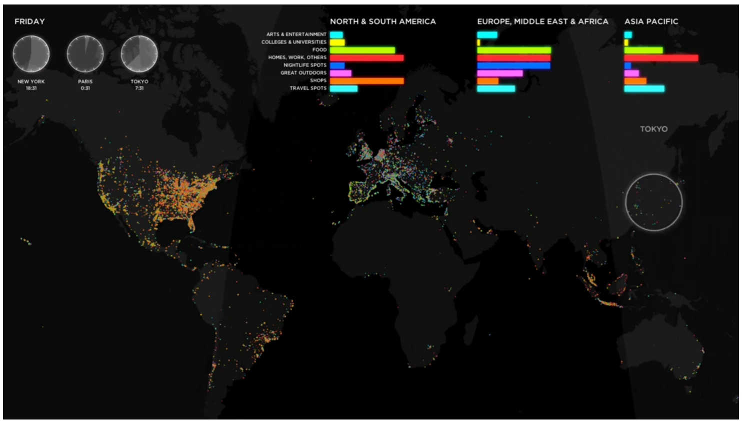


Figure 6.8: *A Week of Check-ins on the Path to One Billion* by *Foursquare*

6.8 A Week of Check-ins on the Path to One Billion

A Week of Check-ins on the Path to One Bilion[12], shown in Figure 6.8, is a Map, Video/Animation type of visualization, created to promote *Foursquare*. The visualization consists of an animated time-lapse video that shows locations around the world light up on a map as people there check-in at different times of the day. Since it is a video it is considered playable, but it does not have interactivity, and its reading/viewing order is completely linear. In terms of narrative elements this visualization does not have much, just a title and captions.

Since it is a promotional video, the visualization includes the company logo. In terms of visual elements it also has three bar charts (North & South American; Europe, Middle East and Africa; Asia Pacific) with the types of locations of the check-in enabling the viewer to know which kind of check-in happens the most, three pie chart like clocks with different time zones, and obviously video, animation and a map.

[12]http://rethinkingvis.com/visualizations/13

A Week of Check-ins on the Path to One Billion	
Genre	Map Video/Animation
Narrative Elements	Captions Title
Visual Elements	Bar chart Dot map Pie chart Video Zoom (visual)
Interactive Elements	No Interactive Elements
Reading/Viewing Order	Linear

Table 6.8: *A Week of Check-ins on the Path to One Billion* genre, elements, and reading/viewing order

Figure 6.9: *100 Years of World Cuisine* by *100 Years of World Cuisine*

6.9 100 Years of World Cuisine

100 Years of World Cuisine[13], shown in Figure 6.9, is a shocking photograph that shows the number of deaths on 25 conflicts from 1915 till the date of publication of the visualization, on which the number of deaths of each conflict is represented by the quantity of blood in different kitchen containers. As it can be seen in Table 6.9 this is a non-playable visualization with few narrative elements. There are also additional charts for the amount of deaths by continent, the amount of deaths by decade, and the total of deaths in the 20th century that appear of the visualization and the amount that does not appear. *100 Years of World Cuisine* has a user driven reading/viewing order.

100 Years of World Cuisine would benefit a lot if it used interactivity. If it was possible to click on each container of blood present in the picture and have access to additional data or to a story, this visualization would be even more interesting and it would better fulfill its purpose: to create awareness to the number of people that die because of these conflicts.

[13] http://rethinkingvis.com/visualizations/9

100 Years of World Cuisine	
Genre	Photograph
Narrative Elements	Accompanying article Captions Title
Visual Elements	Line chart Photograph Pie chart Size representing quantity
Interactive Elements	No Interactive Elements
Reading/Viewing Order	User Directed Path

Table 6.9: *100 Years of World Cuisine* genre, elements, and reading/viewing order

6.10 How Local News Is Going Mobile Infographic

How Local News Is Going Mobile[14], which can be seen in Figure 6.10, is a perfect example of a Poster like visualization, where the charts and diagrams also play an important part. The *How Local News Is Going Mobile: Could the iPad Be the New Sunday Press* visualization seeks to answer the newspaper industry's biggest question – can the iPad resuscitate this dying industry - through the use of a series of charts combined with strong graphic elements. As a typical poster it conjugates information and graphic elements in order to be appealing and eye-catching. As most posters this visualization is non-playable.

As it can be seen in Table 6.10, in terms of visual elements this visualization uses a lot of illustrations and pictograms, some of them even used to present data like an umbrella that is used as a pie-chart. There is also a reinterpretation of a bar chart on which the background behind the bars is a watermark flag and the bar in front shows that drawing in bolder colors. There are also regular bar charts, circular bar charts, doughnut charts, and pie charts.

In terms of narrative elements, *How Local News Is Going Mobile Infographic* has an introductory text to introduce the user to the topic, captions, and annotations that complement the graphics and illustrations. The order of reading is optional, the user can choose to read it from top to bottom or just check the elements that pop out or that are more appealing for.

[14] http://rethinkingvis.com/visualizations/27

Figure 6.10: *How Local News Is Going Mobile* infographic by *Column Five*

How Local News Is Going Mobile	
Genre	Poster
Narrative Elements	Annotations Captions Introductory text Title
Visual Elements	Bar chart Circular bar chart Doughnut chart Illustration Pictogram Pie chart
Interactive Elements	No Interactive Elements
Reading/Viewing Order	User Directed Path

Table 6.10: *How Local News Is Going Mobile* genre, elements, and reading/viewing order

6.11 SPENT

SPENT[15], shown in Figure 6.11, is a Game style visualization launched in February 2011 by *McKinney*, an advertising agency, and the Urban Ministries of Durham, a private non-profit organization. The motto is simple (or not): Could you live on $1,000 a month? This Game lets the user, in this case the player, make the everyday choices necessary to get by on a tight income. First the user has to choose a job, like a waiter at a restaurant or a temporary typist, and that means different rates of pay. Then the user is presented with everyday choices: what food to buy, whether or not to pay for the car insurance, etc.

In terms of visual elements, which can be seen along with the other elements of the visualization in Table 6.11, this visualization uses a timeline for the user to see in which day of the month he/she is, pictograms, and tooltips. In addition to an introductory text, annotations

[15] http://rethinkingvis.com/visualizations/59

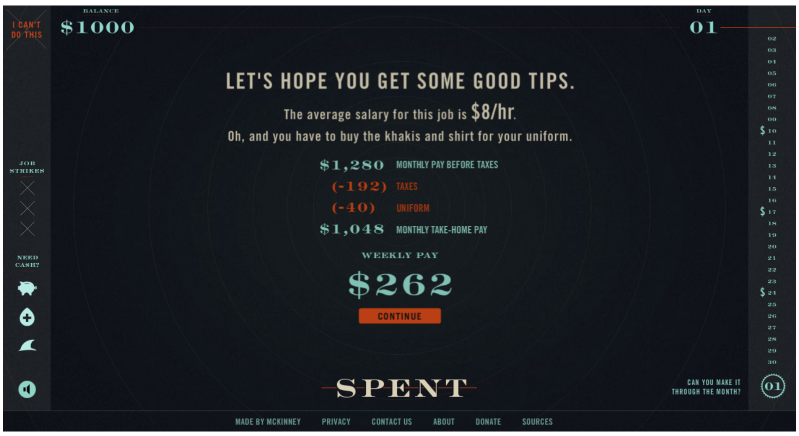


Figure 6.11: *SPENT* by *McKinney*

SPENT	
Genre	Game
Narrative Elements	Annotations Captions Introductory text Title
Visual Elements	Pictogram Timeline Tooltip
Interactive Elements	Click detail Gamification Hover highlight Input box Navigation buttons Player controls Scrollbar
Reading/Viewing Order	User Directed Path

Table 6.11: *SPENT* genre, elements, and reading/viewing order

and captions provide not only the information necessary for the user to navigate through the game, but also gives additional data like how many families choose not to go to the dentist because it is too expensive, etc. It is this amount of additional data that makes this not a just a game but a collection of information that is able to raise awareness to the problem of poverty.

The interactive elements are very important in this kind of visualization, because these elements are the key to transform the interaction of this visualization into a game. In addition to the typical game mechanics, there are navigation buttons, scrollbar, player controls to control the sounds of the game, details that pop-up on click, and details that can be highlighted. In the beginning of the visualization there is also an input box for users to type in some text. As in any game, the path is directed by the user.

6.12 Chapter Summary

In this chapter the new visualization typology presented in chapter 5 is illustrated through the use of case studies. The analysis of these case studies was essential to better understand what makes a visualization a Sequential Graphic, Slide Show, Chart/Diagram, Map, Tag Cloud, Model, Drawing, Video/Animation, Photograph, Poster, or Game.

I gathered eleven examples, each chosen because it is a clear representative of its category in the typology, and did a more exhaustive analysis of the elements that compose them. The impact that each component has in the definition of the respective category was also studied. The examples of visualizations gathered come from very different sources such as online news media, specialized visualization websites, etc.

Visualization title	Source	Typology category
England Riots	The Guardian	Sequential Graphic
The World of Seven Billion	National Geographic	Slide Show
Death penalty statistics, country by country	The Guardian	Chart/Diagram
Home and Away: Iraq and Afghanistan War Casualties	CNN	Map
What Does China Censor Online?	David McCandless	Tag Cloud
Ground Zero Now	The New York Times	Model
How Many Households Are Like Yours?	The New York Times	Drawing
A Week of Check-ins on the Path to One Billion	Foursquare	Video/Animation
100 Years of World Cuisine	100 Years of World Cuisine	Photograph
How Local News Is Going Mobile	Column Five	Poster
SPENT	McKinney	Game

Table 6.12: Typology case studies and respective typology categories

Table 6.12 shows all the case study examples, their sources, and respective category that they are used to illustrate. Some examples can be categorized into more than one category. This is the case of: Death penalty statistics, country by country; Ground Zero Now; and A Week of Check-ins on the Path to One Billion. However, in table 6.12 it is only presented the category that each of those case studies are illustrating.

In the next chapter, I present the results of a focus group study where some of the types of visualization identified in this chapter were tested.

Chapter 7

Focus group study

The focus group was conceptualized and structured as way to gather information about factors such as comprehension, likability, and navigation. The method was used because it fosters the discussion between the participants and enables to obtain qualitative and affective information from participants easily. These focus group sessions were conceptualized as an exploratory exercise to obtain an emotional response from the participants, an aspect that could not be evaluated using the survey method.

7.1 Procedures

The study took place at *Universidade Nova de Lisboa* and used the focus group method to collect data. The location for the sessions was a classroom with a computer with Internet connection for each participant at the *Faculdade de Ciências Sociais e Humanas* campus. There were a total of 16 participants, divided into 2 groups of 8 elements each, and no personal information about them was collected. The groups were mixed in terms of gender and age.

The source of the participant population was university students, either from the *New Media and Web Practices* or the *Communication Sciences* masters program, that were willing to participate in this research. They were invited to take part in the focus group study by their Professor. Although the participants belong to the same cohort, they have different backgrounds in the area of communication: journalism, design, marketing, etc. This diversity of backgrounds guarantees diversity of responses also ensuring that they had enough previous knowledge to fully understand the stimulus provided during the sessions. The focus group sessions were done on the first day of classes therefore most participants did not knew each other prior to the experiment. This fact reduces the chances of the participants being influenced by the opinion of others.

The focus group sessions began with a standard introduction and explanation of the purpose of the research. The participants received a rating sheet (shown in Appendix III) to rate the visualizations, and could see and interact with the visualizations through a computer. The moderator asked the participants to rate each visualization presented on a scale from 1 to 10 immediately after they interacted with the visualization and before the discussion about that visualization started. The rating was not done as a group activity because it was important that this reflection was independent of the group-think that could be generated by the discussion. The visualizations were rated in terms of comprehension:

- Was the information presented in a clear, comprehensible way?
- Was the purpose easy to understand?

... likability:

- Was the visualization interesting and engaging?
- Was the interaction enjoyable?

... and navigability:

- Was the data easy to navigate?
- Was it clear how to interact with it?

The sheets were returned to the moderator at the end of the session.

The participants were given a few minutes to explore and interact with the first visualization and afterwards the moderator asked some semi structured questions about it (presented in Appendix IV). This process was repeated for each visualization in this study. These questions were asked in order to start the discussion between the participants. They were asked to explain their answer and to provide their views about the visualizations, discussing them with the other participants. The sessions were recorded for record keeping and their answers were later transcribed.

The moderator also asked some questions regarding two or more visualizations at the same time: Which of the two do you prefer? Which one do you think is more attractive visually? Which do you think tells the story better? The questions about comparisons were asked using similar visualizations or comparable topics.

The discussions took about one hour and both groups were shown eleven examples of visualizations of different types and different characteristics (playable, non-playable, introductory text, accompanying article, and audio narration). The group was shown the examples on Figure 7.1. Only three of the visualizations (marked on the figure with a dark gray background) are non-playable. These examples were chosen through a previous research (previously presented in Chapter 5) of what is currently being done on online newspapers and magazines, blogs, scientific videos,visualization research websites, and even publicity campaigns, and more importantly what is already popular and shared by the users of the Internet.

Sequential Graphic	Evolution of the Web	Introductory Text
Map	Home and away	
	Death penalty in the US mapped	
Poster	How local news is going mobile	
Photograph	Faces of the dead	
	British troops killed in Afghanistan	
Tag Cloud	What does China censor online?	Little or No Text
Drawing	How much CO2?	
	How many households are like yours?	Accompanying Text
Chart/Diagram	Death penalty statistics, country by country	
Model	Ground Zero now	

Figure 7.1: Characteristics of the visualizations presented to the participants of the focus groups

As it is present in Figure 7.2, most visualizations share common elements, specially in terms of narrative. Most visualizations chosen for this study also share the same reading/viewing order, user directed path, however some examples with a linear reading/viewing order were also shown in order for the participants to state which type they preferred. In terms of interactive elements most of the visualizations analyzed share two types of interaction: click detail and filtering. Every other element (navigation buttons, hover highlight, hover details, link to external article, scrollbar, search, zoom (interactive), and drag objects) is common to at least two visualizations, except for the interactive element object react to mouse movement. This

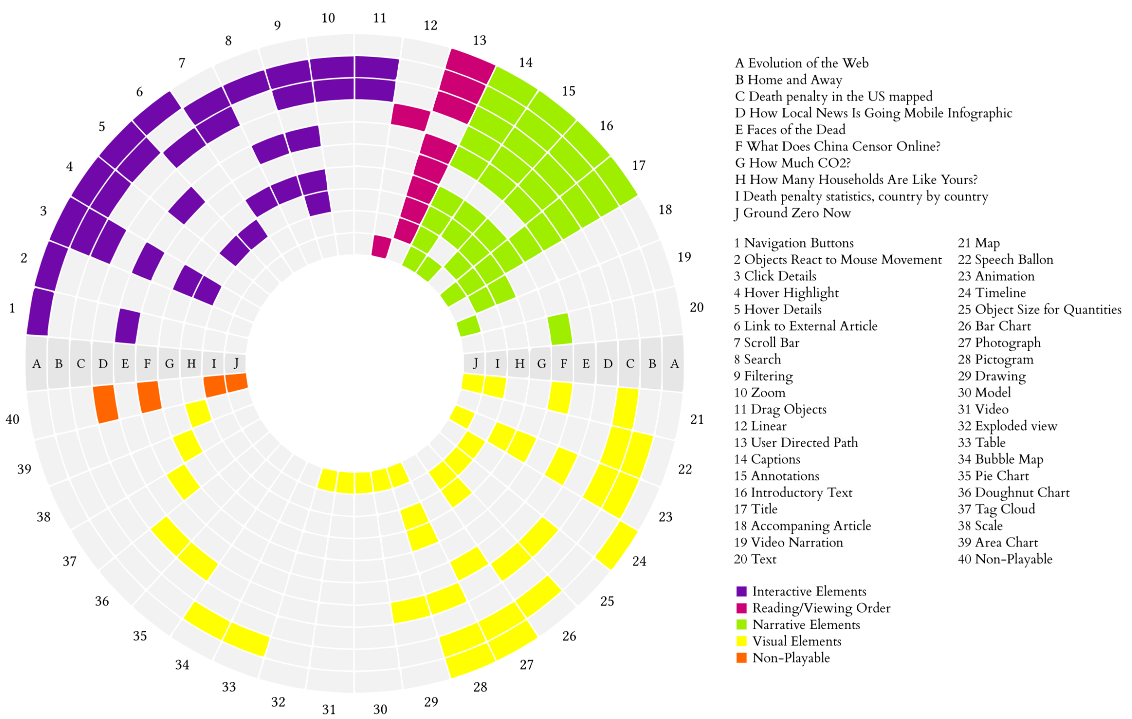


Figure 7.2: Visualizations used in the focus group study and the elements that compose them radial matrix

fact occurs because this type of interactive element is less common and did not really influenced the example too much.

7.2 Results

The participants had very strong opinions about each of the visualizations and, with few exceptions, most visualizations received high scores in terms of comprehension, likability, and navigation (see Figure 7.4). The overall average rating of the visualizations according to the ratings from 1 to 10 can be seen on Figure 7.3.

7.2.1 The most discussed

Two visualizations caused a lot of discussion and achieved high scores in all of the categories (comprehension, likability, and navigation): *The New York Times*' *How many households are like yours?* and *The Guardian*'s *Death penalty in the US mapped*[1].

The first had the most participants giving it a score of 10 in terms of comprehension and likability. It was also the visualization for which more participants gave a 9 in likability, and navigation. This visualization also had the highest average rating (8.3 for comprehension, 8.1 for likability, and 8.2 for navigation) and no individual ratings under 4. This interactive visualization, previously explored in detail in Section 6.7, allows the users to explore different types of American households by choosing the primary residents of a type of household and seeing how it compares to other American households.

Death Penalty in the US Mapped allows users to see which US states have carried out the death penalty most often and which have proportionately the most executions. *The Guardian*'s visualization is part of a larger article entitled *Death penalty statistics from the US: which state*

[1] http://rethinkingvis.com/visualizations/20

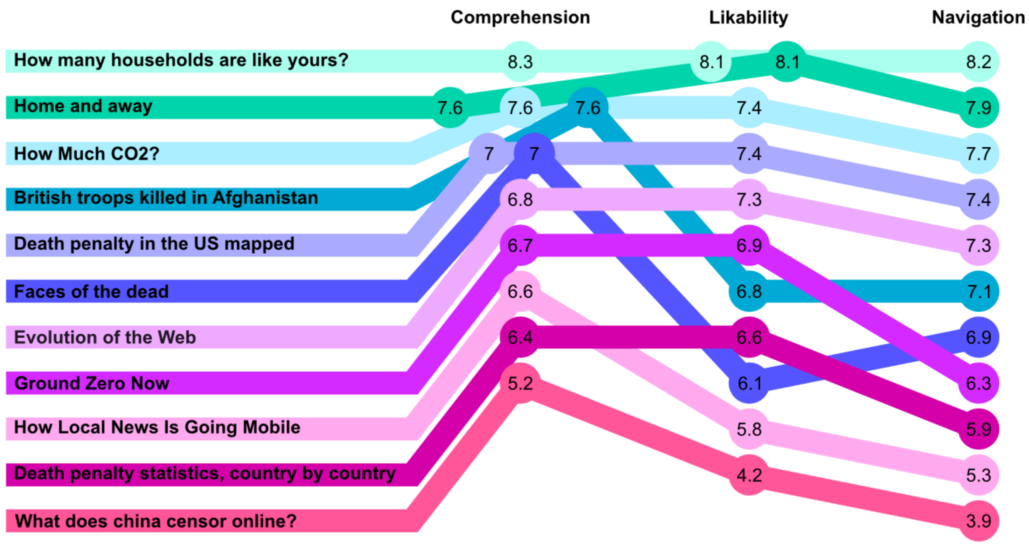


Figure 7.3: Ranking of the average ratings for each visualization in terms of comprehension, likability, and navigation

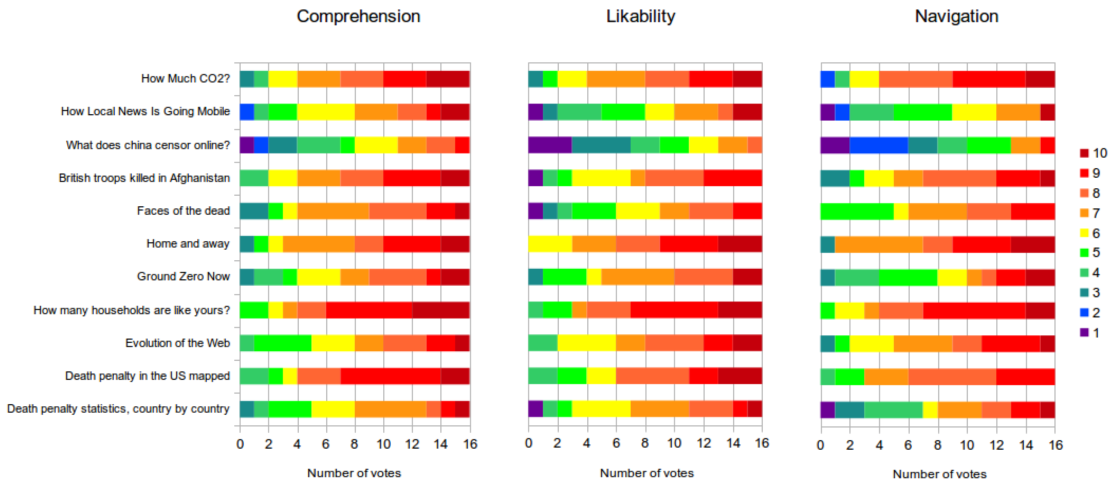


Figure 7.4: Scores given by the participants to each visualization in terms of comprehension, likability, and navigation

executes the most people?[2] and consists of a map where the states are colored according to the number of executions. It is possible to click on a state to explore it. When the user clicks on a state, a bubble pops up with further information for different years. The user is also able to choose on the dropdown menu to see the executions since 1976, in 2011, which state has more inmates on death row (in January 2011), which have more death row inmates per million pop, the number of executions per million pop since 1976 in each state, and the total number of executions from 1608 to 2011 for each state.

Death penalty in the US mapped, which can be seen in Figure 7.5, had the same number of participants giving it a 10 in terms of likability as *The New York Times*' *How many households are like yours?*, but had more participants giving it a 9 in comprehension. On average, the visualization ranked 3rd in terms of comprehension and likability and 4th in terms of navigation. The overall scores were not as high as the ones given to *The New York Times*' visualization.

[2] http://www.theguardian.com/news/datablog/2011/sep/21/death-penalty-statistics-us

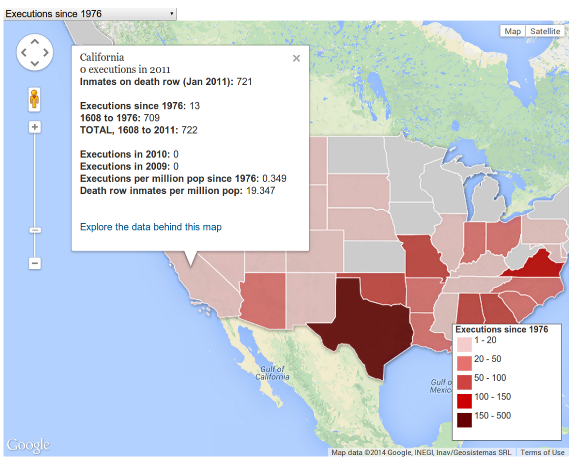


Figure 7.5: *Death Penalty in the US Mapped* by *The Guardian*

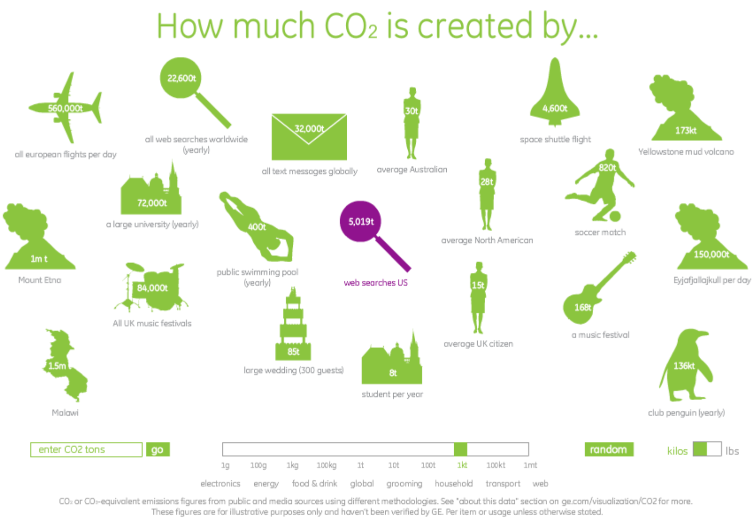


Figure 7.6: *How Much CO2?* by David McCandless

In the sessions, the *Death penalty in the US mapped* visualization was compared with a visualization of a similar subject: *Death penalty statistics, country by country*, previously described in Section 6.3. This visualization by *The Guardian* accompanies an article about countries that maintain the death penalty. The visualization is composed of a map with bubbles of different sizes that represent the number of individuals executed and the number of death sentences handed down in 2012. On the bottom, there is a timeline representing the number of countries carrying out executions, from 1991 till 2012.

The participants were very vocal about the these visualizations, mainly because the first was playable and the second was not. Most participants preferred the one that allowed interactivity even though they noticed that the other provided more information on the subject. Some participants stated that the fact that the information was immediately presented caused confusion and they would prefer if it showed that information only on click.

Another popular example among the participants, having an average of 7.6 for comprehension, 8.1 for likability, and 7.9 for navigation (as seen in Figure 7.3), was *CNN*'s *Home and away: Iraq and Afghanistan war casualties* visualization (presented in Section 6.4). This Map visualization is composed of two maps where the audience can find the birth place of a trooper that has fallen either in Afghanistan or Iraq and relate with the location where he/she died. By clicking on the points that represent a fallen soldier the audience can learn more details about them on their profile page. This visualization, along with *Death penalty in the US mapped* and *How many households are like yours?*, had the highest number of participants giving it a 10 for likability and had the highest number of participants giving it a 10 for navigation.

7.2.2 Links as context

How Much CO2 is Created By...[3], a visualization created by David McCandless for *General Electric* that can be seen in Figure 7.6, was also one of the most popular visualizations

[3] http://rethinkingvis.com/visualizations/26

among the participants. As it can be seen in Figure 7.3, this visualization ranked 2nd in terms of comprehension and likability and 3rd in terms of navigation. *How Much CO2 is Created By...* is an interactive visualization that shows the amount of carbon that different activities, entities, or events emit. Although the participants liked this visualization they thought that something was missing: text. One of the participants stated that if instead of the visualization just showing the fact that web searches in the US produce 5,019 tons of CO2 he would prefer that they also explained how this happens. According to him "it is lacking support stories."

Another visualization that the participants said suffered from the lack of storytelling was the *Faces of the dead*[4]. This visualization ranked 4th in terms of comprehension and likability and 5th in terms of navigation. It was also the one that had a higher number of participants giving it an 8 in terms of comprehension. This visualization consists of a picture of a soldier that fell in Iraq or Afghanistan composed by little squares that represent other soldiers that also died in these wars. Every time we click on one of the squares the big picture becomes the photo of that soldier that we clicked on. One of the participants said that this visualization looks more like a work of digital art, but does not give sufficient information to become interesting as a visualization. Furthermore the fact that the pictures were in black and white made it more difficult for the participants to relate with it. One of the participants even stated that because the features were blended by the lack of color, the soldiers were too homogenized.

Most participants preferred another visualization with a similar topic that was compared to *Faces of the dead*: *The Guardian*'s *British troops killed in Afghanistan*[5] visualization. This visualization ranked 3rd in terms of comprehension and likability and 4th in navigation, as it can be seen in Figure 7.3. *The Guardian*'s visualization also uses pictures of fallen soldiers but shows them in color and includes a hyperlink to the story of how that soldier died. Because the

[4]http://rethinkingvis.com/visualizations/45
[5]http://rethinkingvis.com/visualizations/18

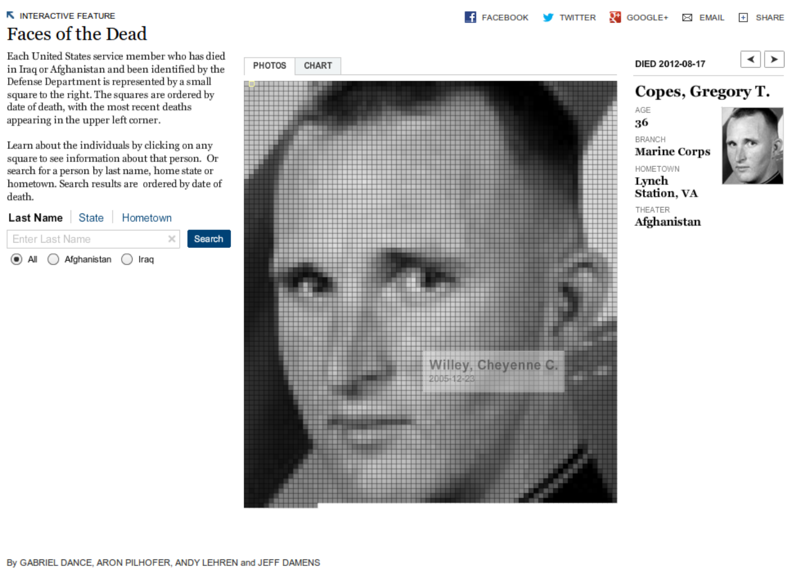


Figure 7.7: *Faces of the dead* by *The New York Times*

Figure 7.8: *British troops killed in Afghanistan* by *The Guardian*

pictures were in full color the participants felt closer to those people and felt that they could relate to them more easily. The fact that they could also read the story also contributed for that sense of closeness with the visualization. The participants felt that in this visualization the soldiers were not a number, a dot, or a square, they were real people.

One visualization that also plays on this idea of the stories appearing as support information is the *Evolution of the Web*[6] visualization. This visualization by Google allows the user to visualize the history of the Web and its browsers, and at the same time understanding the pace at which these technologies were developed. *Evolution of the Web* includes several lines across the main timeline that the authors call web technology *strands* and 7 separate timelines for each browser (Mosaic, Netscape Navigator, Opera, Internet Explorer, Safari, Firefox, and Chrome), where timelines the major developments on the web platform are represented. There is also a secondary visualization, shown when the user clicks a button on the top left corner that says *The Growth of the Internet*, to provide additional information on the number of users and traffic.

[6]http://rethinkingvis.com/visualizations/22

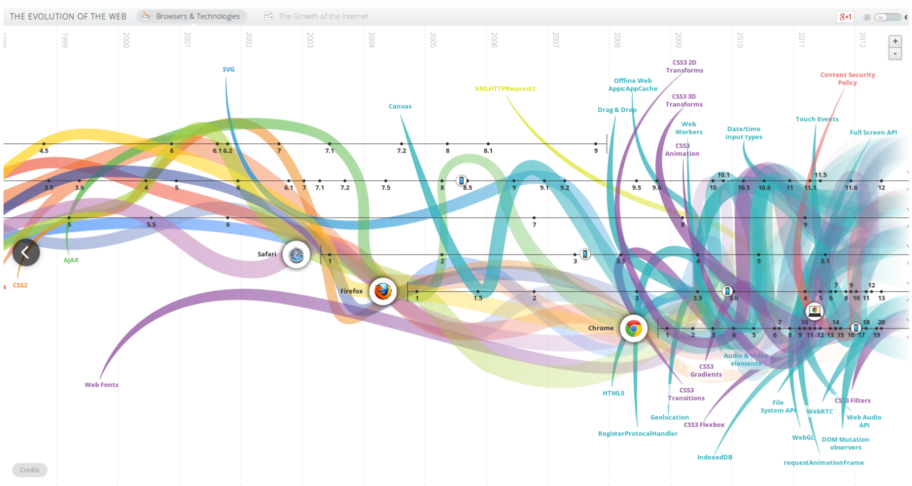


Figure 7.9: *Evolution of the Web* by *Google*

This was also one of the participants' favorite visualizations, being in the top five for all dimensions of evaluation (see Figure 7.3). The average rating for this visualization was 6.8 for comprehension, 7.3 for likability, and 7.3 for navigation. One of the participants said that this visualization was almost perfect. "First it is a timeline and timelines give this idea of story, of flow, and moreover we are automatically placed at present time, which is probably what we have the most interest in, then being able to then navigate backwards," he added. Other participant referred that she enjoyed the fact that the visualization did not present the information all at once but allowed them to explore by clicking on links available, and that sparked her curiosity.

7.2.3 Non-playable

The participants were also shown two other visualizations that did not have any kind of interactivity: *What does china censor online?*,previously analyzed in Section 6.5 and *How Local News Is Going Mobile: Could the iPad Be the New Sunday Press?*, explained in detail in Section 6.10. The first is a simple non-playable tag cloud in which the shape that the words make the map of China. As it can be seen in Figure 7.3, this visualization had the lowest overall scores. The average rating for this visualization was 5.2 for comprehension, 4.2 for likability, and 3.9 for navigation. One of the participants even stated that she took a long time to realize that there was no interactivity in the visualization and that she though that she just did not understand where she was supposed to click. When inquired about what could have been done to improve this visualization and make it more interesting, some participants responded that additional information could be shown on mouse hover and that could make the visualization much better. The overall feeling was that they did not learn anything new.

How Local News Is Going Mobile, the second non-playable visualization shown was a poster like visualization. Although the ratings given were not completely bad (6.6 for comprehension, 5.8 for likability, and 5.3 for navigation) the participants were almost indifferent to the visualization *How Local News Is Going Mobile: Could the iPad Be the New Sunday Press?* One of the participants said that this visualization was not very stimulating and that after seeing visualizations with so much interactivity this one just looked even worst. The overall opinion was that they lost interest in the visualization because there was nothing more to discover.

The participants were also shown a video based visualization and surprisingly, although the video had no interactivity, most participants showed an overall positive response to it. *The Ground Zero Now* visualization has two animation videos that include models of buildings, pictograms and drawings, but does not have any interactivity. About the audio and video narration the opinions were divided and although some participants said that they really enjoyed this kind of storytelling the majority did not express much interest in it.

7.3 Chapter Summary

In chapter 7 I present the focus group study that I conceptualized and structured as way to gather information about factors such as comprehension, likability, and navigation. This method was chosen because it fosters the discussion between participants and enables to obtain qualitative and affective responses easily.

In this study that took place at *Universidade Nova de Lisboa*, the participants saw 11 visualizations of different types and with distinct characteristics (some were playable, some non-playable, and all had different types of narration). The participants received a rating sheet to rate the visualizations and the moderator asked them to rate the visualization from 1 to 10 immediately after they interacted with it and before the discussion started. Visualizations were rated in terms of comprehension, likability, and navigability. For the discussion the moderator asked semi-structured questions to stimulate the discussion, such as:

- Which visualization do you prefer?
- Which visualization do you think is more visually attractive?
- Which visualization do you think tells the story better?
- etc.

Overall, the participants gave a high score in terms of comprehension, likability, and navigation to most visualizations. However they had very strong opinions about each of the visualizations. The results in section 7.2 are sectioned into three parts: the most discussed, links as context, and non-playable. The results showed what participants value in a visualization:

- links that provide some context;
- empathy between the users and the data;
- interactivity (participants generally preferred playable visualizations and they felt that some of the visualizations they liked the least could be fixed by adding interactivity);
- support stories that provide some context;
- visual metaphors that give the feeling of story flow, such as timelines.

In regards to the results from the study presented in this chapter, I must say that these are not definitive answers to the research question (RQ3) “which types of visualizations might appeal to the public?”, however the results definitively shed a light on what may be good strategies to improve enjoyability. Nonetheless, to be able to make design recommendations for visualization, further evaluation is needed. This issue will be later discussed in more detail in section 11.2 — Future work.

Chapter 8

Using visualizations to tell stories

Narrative is first and foremost a prodigious variety of genres, themselves distributed amongst different substances – as though any material were fit to receive man's stories. Able to be carried by articulated language, spoken or written, fixed or moving images, gestures, and the ordered mixture of all these substances; narrative is present in myth, legend, fable, tale, novella, epic, history, tragedy, drama, comedy, mime, painting (think of Carpacio's Saint Ursula), stained glass windows, cinema, comics, news item, conversation. Moreover, under this almost infinite diversity of forms, narrative is present in every age, in every place, in every society; it begins with the very history of mankind and there nowhere is nor has been a people without narrative. All classes, all human groups, have their narratives, enjoyment of which is very often shared by men with different, even opposing, cultural backgrounds. Caring nothing for the division between good and bad literature, narrative is international, transhistorical, transcultural: it is simply there, like life itself.

— Barthes (2004, p. 65), *The Rock*

According to Satyanarayan and J. Heer (2014, p. 361), "stories are a pervasive aspect of human culture; they convey information in a memorable form that can engage and establish causal links". Storytelling is an ancient art deeply rooted in our common human culture. Whether if it is conveyed through spoken or written words, music, or images, storytelling is present in most human activities. In fact, according to neuroscientist António Damásio (1999), storytelling is also a vital operation of the brain because the process by which the human brain constructs consciousness is based on storytelling. When interacting with an object, the brain constructs "a simple narrative without words" (Damásio, 1999, p. 168) that has all the elements of a narrative: characters (the organism and the object) and a sequence of events that unfold in time (a beginning, a middle, and an end).

(a) Tablet V of the *Epic of Gilgamesh*

(b) Illustration of La Fontaine's fable *The Tortoise and the Hare*

Figure 8.1: Examples of early storytelling

Although we cannot pinpoint the exact moment when Man started this habit of telling stories, we know that their origin is ancient. In addition to oral storytelling there is proof that ancient cultures registered stories in order to immortalize them. Cave painting seems to be the oldest proof of registered storytelling and the paintings in the Lascaux Caves (France), the earliest example, depicts a series of events that tell us about prehistoric hunting practices.

Another early example of registered storytelling is the *Epic of Gilgamesh*, a poem from ancient Mesopotamia (see Figure 8.1a). The earliest surviving *printed* story, carved in clay tablets, tells the story of Gilgamesh, king of Uruk, and the Great Flood. The fact that the story was registered in a medium that can be transported allowed the story to spread from Mesopotamia to Europe and Asia very quickly.

Aesop's and La Fontaine's (see Figure 8.1b) fables are also well-known examples of storytelling. Fables are short fictional stories that usually resort to anthropomorphization (the attribution of human characteristics to non-human entities) to illustrate moral lessons. For instance, Aesop's fable *The Tortoise and the Hare*, later adapted by La Fontaine who published it in the first volume of the collection *Fables Choisies*, in 1668, teaches that being arrogant and underestimating your opponents has its consequences. The hare was sure that it would win the race, so it decided to stop for a nap during the race, while the tortoise continued slow but without stopping eventually winning the race. Fables like this one were spread and retold, and continue to be remembered, still teaching valuable lessons today.

However, narrative is not exclusive to orality, or to textual and visual forms of representation. At the light of musical semiotics (the study of the signifying ability of music) music has also been seen as a form of narrative (Maus, 1991). According to Maus (1991, p. 6), "listeners can hear musical successions as story-like because they can find something like actions, thoughts, and characters in music". Take for instance Vivaldi's *Le quattro stagioni*: it is one of the earliest examples of program music, a term applied to instrumental musical compositions that have extramusical meaning (that follows a story, or is intended to evoke an atmosphere, or describes a scenery) and illustrates the sequence of events that occur with the passing of the seasons. One is able to understand which season corresponds to each concerto and imagine the events that occur in the season that is currently being described in the piece (flowers blooming in the spring, leaves falling in the autumn, etc.).

Stories were (and still are) spread to instill values, educate, entertain, spread news, and register important events, gaining a huge importance is every society. Recently, this potential to transmit values has been explored in robotics. In their paper *Using Stories to Teach Human Values to Artificial Agents*, Riedl and Harrison (2016) explore the usage of short stories for artificial intelligence to learn social conventions. According to Riedl and Harrison (2016, p. 1) "In this paper, we hypothesize that an artificial intelligence that can read and understand stories can learn the values tacitly held by the culture from which the stories originate". The system learns by getting a reward every time it reacts as a human would in a respective social situation and penalized when it does not, which was previously learned from the short stories.

Storytelling even has a healing role, as Aristóteles (1993) pointed out in *Poetics*. Although, in his work of dramatic and literary theory, Aristóteles (1993) refers to the concept of catharsis having poetics in mind (in this case also including drama), even non literary storytelling can have this cathartic power freeing people from their fears and unwanted emotions. It is up to the storyteller to encapsulate a compelling narrative and explore the potential of the story in order to awake feelings in the audience, and good storytellers are able to do it. Therefore, the raconteur became a respected figure in the community, being responsible for the perpetuation of traditions and the preservation of culture.

From cave painting to books, from movies to games, stories have fascinated mankind and each society impresses in its stories marks of their identity and their culture. However, in spite of having thousands of different forms of storytelling present in our every day life, what still automatically springs to mind to most of us when we hear the term storytelling is the image of an elder narrating an old fairy tale to children (Ma et al., 2012).

Although it may not appear so, modern storytelling still maintains many of the characteristics of this traditional idea. Technological advances are actually helping to introduce in modern storytelling more of these characteristics that we appreciate so much in traditional storytelling. For instance, with the use of interactivity the audience can feel the joy of the moment of discovery so typical of the live narration of a story.

To truly understand narrative and its application in several different media it is important to understand what a narrative/story is. Usually a narrative is seen as an account of connected events that is transmitted either by being spoken, written, or visually represented. It can be fictional or non-fictional. According to Lee et al. (2015), although a story may not have a predefined temporal or narrative structure, there is always components that are typical of stories such as structures, elements, and concepts. Nonetheless a narrative is commonly associated with the idea of something that has a beginning, a middle, and an end (Liu et al., 2013), which can be linear or non-linear (through the use of *analepsis*, *prolepsis*, and other narrative techniques). According to D. Herman (2011), other basic elements for a narrative are: *situatedness* (discourse context or occasion for telling), event sequencing (structured time-course of events), world making/world disruption (disruption of a state of equilibrium), and what it is like (the feelings of living through the situation and a foregrounding of human experience). Some of these elements of the narrative can be easily applied to different media in order to introduce this feeling of storytelling.

8.1 Narratives and New Forms of Storytelling

New forms of media are creating new ways of storytelling but also the traditional media is transforming their old means of storytelling. The medium continues shaping the information (Gershon and Page, 2001) and the way it is represented. However, computation and the Internet are part of a *New Medium* that has given us the possibility to employ characteristics typical from other media, creating a multimedia narratives more complex and sophisticated than ever. **Interactive storytelling**, for instance, is becoming very popular on news casting and documentaries, areas that were up until recently holding on to traditional forms of storytelling. Interactive storytelling refers to the act of telling stories enhanced with technological features, from simple interaction features to social or collaborative features, that give to the user control over the story that is being told, either by changing the course of the narrative or by choosing which of the parts of the story to see.

Nonetheless it is not only on the web that stories were transformed and the combination of many different media is a strong trend. Nowadays, what once were obscure and alternative forms of storytelling, such as crossmedia and transmedia storytelling, are becoming more and more popular. Meanwhile, the older term multimedia is slowly losing its popularity, however there are still prominent examples of multimedia storytelling being produced. These three terms, although seeming quite similar, differ on how they use media form (the type of language used, which can be text, film, illustrations, audio, etc.) and media channel (the Web, television, newspapers, public installations, etc.).

The term **transmedia** storytelling refers to many stories, with many forms, spread through many channels, but that belong to the same storyworld. Although each story is complete in itself, when all the stories are put together they reveal a larger/more complex story and allow the user to expand his/her understanding of a larger subject. An interesting example of transmedia storytelling is *The Hunger Games* movie trilogy. Together with the movie a transmedia storytelling strategy that included a website and a social media campaign was created to engage the audience. RED Interactive Agency and Microsoft created a portal entitled *The Hunger Games Explorer*[1] that aggregates content about the movie, produced by fans, that is spread through social media (Facebook, Twitter, Instagram, YouTube, and Tumblr) and combines it with promotional content about the movie (trailers, interviews, photos, etc.). The portal also encourages fans to share and create content, and enter competitions to win not only virtual badges but also *real world* prizes. The engagement with the portal, all the traditional media (posters, trailers, etc.), and the movie itself separately increases the audience enjoyment and understanding of the story. However, even though each separate feature probably is satisfying on its own, the sum of interaction with all the media enhances even more the enjoyment of the user.

[1] http://ff0000.com/work/view/hunger-games-explorer

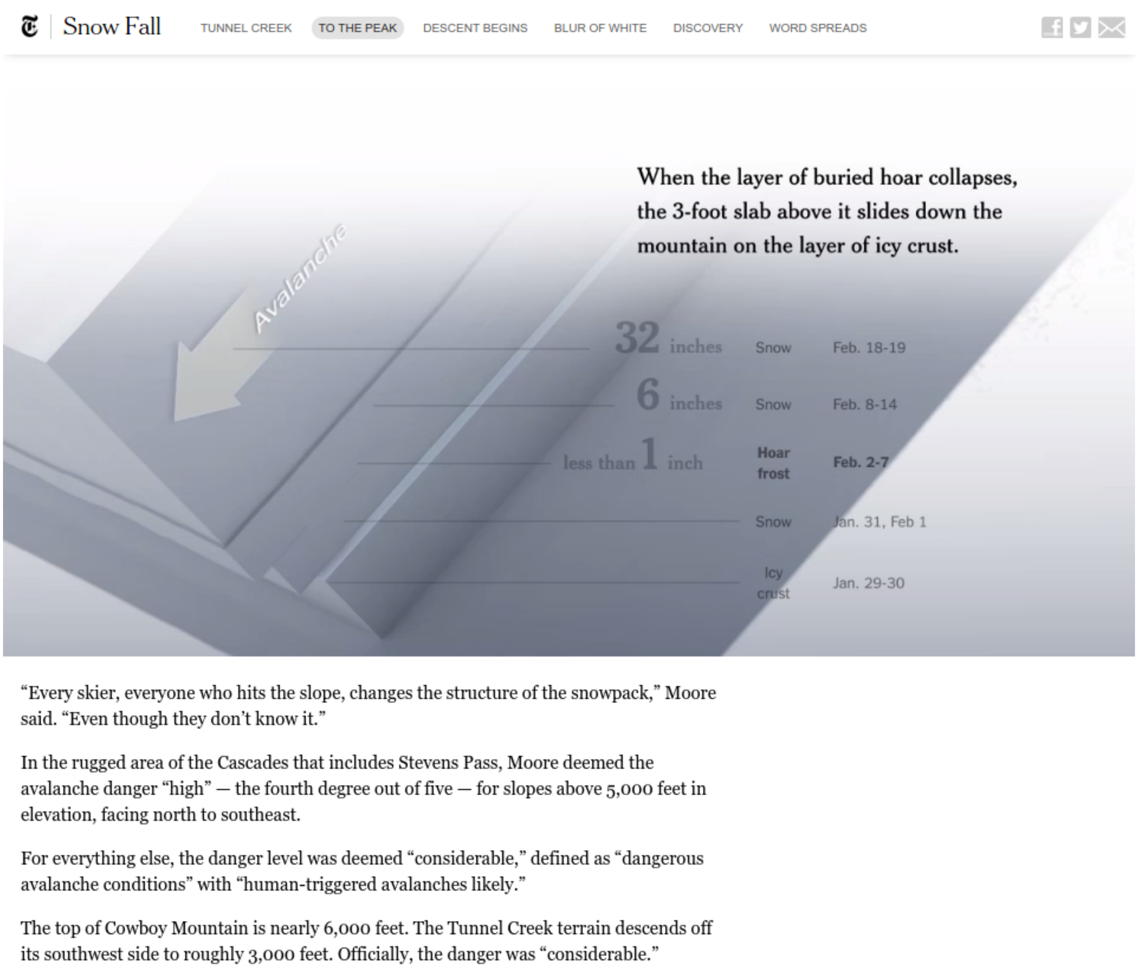


Figure 8.2: *Snow Fall* by *The New York Times*

Crossmedia storytelling can seem very similar to transmedia storytelling but sum of all the parts is not meant to be seen as a full story that is enhanced by all the parts that compose it. The term crossmedia storytelling can be summarized as: one story that is told through many channels (for instance, simultaneously sharing the same story in the same way on television, newspaper, and Internet). It is commonly used in marketing campaigns where, although the specificities of the media shape the way the story is presented, the same story is replicated on the different media.

On the other hand, the term multimedia storytelling refers to one story, with many forms, that is spread through only one channel. A great example of multimedia storytelling is *Snow Fall*[2] (See Figure 8.2) by *The New York Times*, the story of the 2012 Tunnel Creek Avalanche in Washington, that resulted in three fatalities and one injured. It is a multimedia piece composed by six chapters (entitled *Tunel Creek*, *To the Peak*, *The Descent Begins*, *Blur of White*, *Discovery*, and *Word Spreads*) that resorts to text, large striking photos, scroll activated animations, videos, virtual models, and slide shows to tell its story.

This story had such a big impact and created such a buzz that the executive director

[2] http://www.nytimes.com/projects/2012/snow-fall/

of *The New York Times* Jill Abramson, on her keynote in the 14th International Symposium on Online Journalism, said that Snow Fall has become a verb. According to Jill Abramson "*To snowfall* means to tell a story with fantastic graphics and video and every kind of multimedia, and that is absolutely organic to the storytelling itself" (Martinez, 2013, Website).

8.1.1 Narrative visualization

Until recently visualizations have been used to support traditional forms of storytelling as extra information or supporting evidence Segel and J. Heer, 2010. Nonetheless, there has been a great effort lately to transform visualizations in an independent form of storytelling that can exist by itself without support of a traditional form of storytelling such as a video or text. The research about the ways of doing this is being carried out in various different areas but mainly in journalism Cairo, 2012; Segel and J. Heer, 2010, an area in which there has been a great effort to create multidimensional stories composed of other media besides text, and information/knowledge visualization Gershon and Page, 2001; Hullman and Diakopoulos, 2011; Kosara and Mackinlay, 2013, two disciplines which were primarily focused on techniques of visualization but are now starting to research strategies to make visualizations that are independent eschew other types of narratives.

Narrative visualization is an emerging genre focused on the effective communication of complex data to an audience using visualization, in a way that is both engaging and that promotes the sharing of insights. According to Dove and Jones (2012, p. 1) "One of the distinguishing features of narrative visualization is the use of interactive exploratory techniques to enhance the communication of ideas and promote insight through discovery." Using narrative elements in visualizations often help create a structured interpretation path that usually does not exist in traditional information visualization.

8.1.2 History of storytelling in information/knowledge visualization

Information visualization is much more than a visual representation of data. It is the complex process of dissecting raw data that by itself has little meaning and presenting it in a way that it is no longer complex. Although it has been used for a long time Information/Knowledge visualization has blossomed with the emergence of the new media. "These new technologies truly allow us to do things we never could with paper, so we should expect it to take awhile to gain sufficient understanding of them before we can apply them as effectively as we would like" (Gershon and Page, 2001, p. 33).

Moved by the rising of these new media Gershon and Page (2001) were the first to notice the valuable contribution that storytelling could give to information visualization. However, according to Kosara and Mackinlay (2013), they fail to describe actual visualizations and focus mainly on map views without numerical data. In other words, they focus more on simple visual representation, without however truly describing examples of actual information visualization/data visualization and contributing with strategies *de facto* to introduce storytelling.

Later, in 2010, the theme sparked again when Segel and J. Heer (2010) reinvented this notion of using storytelling in visualizations naming it narrative visualization. The authors state that these *data stories* are an emerging class of visualizations. However this fact can be argued because it is not easy to know if storytelling is just enhancing regular information/-knowledge visualizations or if it is another area of research completely independent. Even Segel and J. Heer (2010) do not compare narrative visualization with information visualization/knowledge visualization and simply state that this form of storytelling is different than traditional storytelling. For Rodríguez et al., 2015, *Narrative Visualization: Telling Stories with Data* by Segel and J. Heer (2010) continues to be the flagship contribution in the intersection of visualization, data, and storytelling.

According to Lee et al. (2015), the fact that these data stories are getting more and more popular increased the interest of the research community in this topic. The symbiotic relationship between information visualization/data visualization and storytelling has revealed to be one of the more prevalent topics in visualization in the last few years (Fisher et al., 2008; Hullman and Diakopoulos, 2011; Ma et al., 2012; Segel and J. Heer, 2010).

Although the first paper on the use of storytelling in visualization was published in 2001 (Gershon and Page, 2001), it was not until 2011 that the topic gained widespread attention from the research community. Based on my research on this topic, I can speculate that least 50 papers on this topic were published on major visualization conferences since 2011. Along with the growth of the number of narrative visualizations *in the wild*, the number of new tools for creating more complex visualizations might be in the genesis of this interest from the research community.

In an analysis of the use of storytelling in visualization, Lee et al. (2015) consider that there is not a consensus on what a data story (or a narrative visualization) encompasses and most research is concerned with how the different visualization components are able to improve storytelling. According to Lee et al. (2015, p. 84) "Narrowing the scope of what is termed a data story, for instance, by distinguishing between a visual data story and a data visualization, helps us open the door for a more detailed examination, covering the aspects of the visual data storytelling process that have so far received less research attention." However, although I believe that focusing only on narrative visualizations can have its benefits, I came to the conclusion that it is also important to understand how more traditional visualizations can be infused with narrative elements for these to become more narrative without becoming data stories and consequently becoming more casual. In this way more complex visualizations will be easier to understand and will reach a wider audience without being less serious or scientific. Consequently, this can also lead to a better visualization literacy, because a wider audience will be able to get use to complex visualizations.

With the incorporation of storytelling into visualizations these will be able to give explanations about the subject and wont depend too much on the audience to be able to interpret the data correctly. This is because using narrative elements in visualizations often helps create a structured interpretation path that usually does not exist in traditional information visualization (Diakopoulos, 2010). Moreover, they can be entirely independent of other means of storytelling being able to get the point across easily and in sufficient detail for the audience to understand it. The bits of storytelling in these visualizations do not need to be over informative and descriptive because, with the help of the information represented in visual form, the audience is able to fill in the gaps in the story.

All of these elements make this kind of visualization pleasing, not only because it does not require a lot of time and effort by the audience to assimilate the information but also because it sparks the audience's curiosity and transforms the task of acquiring information into a fun activity. According to Ma et al. (2012, p. 12), "they leave a lasting impression, either by piquing the audience's curiosity and making them want to learn more or by conveying a deeper meaning than your everyday run-of-the-mill sequence of causally related events".

Nonetheless Lee et al. (2015, p. 85) are right when they say that a definition that is too broad and that would consider "any images containing even simple charts with little explanations or reading aids" a data story cannot be accepted. Therefore, they propose three characteristics that have to be present in data stories:

- a set of stories or specific facts;
- annotations (such as labels, and text) or narration that should help to highlight and emphasize the data;
- a meaningful order or connection between the narrative elements.

This means, for instance, that a visualization that allows completely unguided exploration or that has no extra information that frames the data cannot be considered a narrative visualization. Visualizations that are too exploratory with little author-driven guidance can undermine comprehensibility and engagement, resulting in a user that is under-informed or even misinformed. However, this does not imply that exploratory visualizations that are thoughtfully designed cannot be engaging and considered narrative visualizations. This just means that the user, in more exploratory visualizations, is always a volatile variable. On the other hand, visualizations that are over-curated and too story-driven also tend to be boring, specially for proficient users.

Achieving a equilibrium between exploratory and expository is important if we want to have visualizations that are easy to interpret, appealing, and that still leave possibilities for exploration. Also, if we are able to successfully introduce storytelling we will be able to produce visualizations that can be entirely independent of other means of storytelling.

Storytelling can be introduced through the use of persuasive/rhetorical techniques and exploratory/dialectic strategies (Hullman and Diakopoulos, 2011). As it is pointed out in Section 9.1 strategies such as annotations, that add context to the data, can transform traditional visualizations such as simple charts into data stories.

There is an intense discussion (Diakopoulos, 2010) in visualization research about whether or not introducing storytelling is beneficial. However, most seem to agree that, when done right, it can be a powerful way to create a structured interpretation path (Diakopoulos, 2010). Good narrative visualizations allow the user to engage with the data, makes the insight jump out, and helps users to cope with their short attention spans and lack of data literacy.

Notwithstanding, there has also been an increasing concern with how much the incorporation of narrative will impact the exploration of the data and whether or not this will distract the user from the data Diakopoulos, 2010. Although having a direction will help users that are less familiar with the subject an undirected exploration can help proficient users find new interpretations of the data and even discover meanings that were not foreseen by the creators of the visualization. It's important to understand every building block of the visualization in order to create a narrative that does not overpower the data.

Research (Diakopoulos, 2010) has revealed that having flexible narratives that point out particular landmarks for the user to explore, but still allowing the free exploration of the in-between landmarks, is a good option. Nonetheless there is still research to be done on how the narrative influences the interpretation process (Hullman and Diakopoulos, 2011) and how to effectively create these narratives. More research is needed to understand which rhetorical techniques can be used and if it is possible to build a set of techniques that works for different sets of data.

8.2 Chapter Summary

In chapter 8, I approach the use of storytelling in visualization. I begin by reviewing the new forms of storytelling (section 8.1) that appeared with the rise of the new media. I briefly review new forms of storytelling such as interactive storytelling, transmedia, and crossmedia in order to pave the reflection upon the subfield of narrative visualization (subsection 8.1.1).

Narrative visualization as a subfield of information visualization is a relative new research area. However, although it might not seem so, the use of storytelling in visualization is not exactly new, and some examples of what we now call narrative visualization were created way before the term was coined in 2010 by Segel and J. Heer (2010). One of the most memorable examples is Minard's map of the Napoleonic invasion of Russia. It tells the story of a failed invasion, where many men died as the troops moved towards Moscow, and shows the impact that the rivers had on the death toll. Understanding the history of storytelling in information/knowledge visualization is essential to understand the field, therefore I also briefly review this history in subsection 8.1.2, providing hints to the narrative strategies that will be explored in detail in chapter 9 — Narrative strategies.

Chapter 9

Narrative strategies

Several narrative strategies have been approached by different researchers in the past years, particularly approaches closer to semiotics, critical theory, and journalism. Authors such as Segel and J. Heer (2010) and Hullman and Diakopoulos (2011) proposed narrative strategies for visualization based on visual rhetoric. The approach by Segel and J. Heer (2010) aimed towards structure and generalized advice for designing narrative visualization. Hullman and Diakopoulos (2011) go farther proposing an analytical framework for visualization rhetoric that cross editorial layers (data, visual representation, textual annotation, and interactivity) and a set of techniques for visualization rhetoric (omission, metonymy, data provenance, representing uncertainty, identification, obscuring, contrast, classification, redundancy, typographic emphases, irony, similarity, individualization, anchoring, filtering). Their objective went towards the constitution of a guide to how much visualization rhetoric should be used on the design of visualizations. They also give insights about the impact of these rhetorical aspects influence the user's interpretation of the original data.

The focus group results shed a light on what may be good strategies for storytelling in visualizations. According to the participants' responses interactivity seemed to be the most important strategy. One the participants even said anecdotally: "With a video you retain about 8 to 15% of the information, if the video has some interactivity that percentage skyrockets to about 70%. That says a lot about the power of interactivity."

Although the level of interactivity was pointed over and over again as an important feature other characteristics also stood out. What made *The New York Times*' *How many households are like yours?* visualization the overall favorite was the fact that it enabled the participants to relate with the visualization. One of the participants referred that "the fact that the visualization allows us to identify with the subject instantly sparks an additional interest. In fact, the article that is linked to the visualization is about a family that is so different from mine that it probably would not even catch my attention, but since I'm already interested in the subject because of the visualization I would probably read the article also."

In this chapter, I analyze three particular approaches that were identified in the focus group study and that were previously discussed in Information visualization research: context (closely linked to annotation), empathy, and temporality (its relation with story-flow). Other strategies such as gamification are also approached in relation with the previous strategies, specially with empathy. All of these aspects are approached in relation with interactivity.

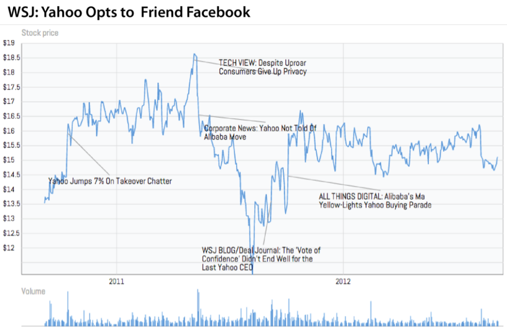
9.1 Interactivity and its relation with context

The benefits of using interactivity in visualizations have been long known. According to Kosara and Mackinlay (2013, p. 47), "being able to not just see the data, but quickly change the view, add different data, etc., makes analyzing it much faster and more effective." However, Ma et al. (2012) consider that it should be carefully balanced, otherwise the creator of the visualization loses control over how the story is told.

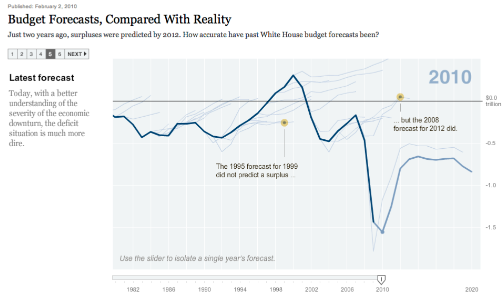
Interactivity opened up the possibility of adding new layers of content to information visualization. Thanks to this additional layer of content, most of the times in the form of annotations, visualizations can, in addition to the data itself, provide content that is able to add context. This content has the potential to help a user make sense of the data (Hullman, Diakopoulos, and Adar, 2013). In addition interactivity offers the possibility to show the content on demand, giving the user a sense of freedom.

As it can be seen in Chapter 7, the opinions collected in the focus group study seem to hint that the audience prefers short moments of storytelling that they can access if they feel the urge, rather than having a dense storytelling that they have to carefully follow. This fact is not surprising since some research about how information visualizations with annotations are a promising way to complement articles has already been done (Hullman and Diakopoulos, 2011; Hullman, Diakopoulos, and Adar, 2013; Satyanarayan and J. Heer, 2014). Hullman and Diakopoulos (2011, p. 2231) analyzed 51 professionally-produced narrative visualization from news media, political outlets, and independent graphic designers, "in order to deepen understanding of how common design techniques represent rhetorical strategies that make certain interpretations more probable." In this analysis they found, as part of their *Visualization Rhetoric Framework* to guide discussion of the rhetorical aspects of information visualization, four editorial layers that can be used to convey meaning: data, visual representation, interaction, and textual annotations. The fact that annotations, them being textual, graphical, or social, play such an important part in many presentations that include visualization is a testimony of their efficacy in focusing a user's attention. However, annotations have often been overlooked in information visualization evaluation research (Hullman and Diakopoulos, 2011).

Among the research done on annotations in narrative visualization two researches stand out. The first, shown in Figure 9.1a, is *Contextifier*, an approach Hullman, Diakopoulos, and Adar (2013) to automatically generate annotations for narrative visualization meant to accompany online news. The system "consists of four main components: a news article corpus, a query generator, an annotation selection engine (which is comprised of three feature generators



(a) *Contextifier* (Figure 6 (Hullman, Diakopoulos, and Adar, 2013))



(b) *Ellipsis* (Figure 2 (Satyanarayan and J. Heer, 2014))

Figure 9.1: Two systems for narrative visualization annotation

and an integrator), and a graph generator" (Hullman, Diakopoulos, and Adar, 2013, p. 2710). The second, shown in Figure 9.1b, is a system named *Ellipsis*, developed by Satyanarayan and J. Heer (2014, p.361) building on top of the research carried by Hullman and Diakopoulos (2011), which "combines a domain-specific language (DSL) for storytelling with a graphical interface for story authoring", enabling users without programming experience to create narrative visualization. The annotations in *Ellipsis* consist of simple shapes such as rectangles, ellipses and arrows, and text with configurable properties such as position, color, and size.

Annotations have the capacity to add context that otherwise would be very difficult to provide, easing the user's interpretation, suggesting conclusions, and generally guiding the user's interaction with the visualization. This context information is easier to assimilate than a dense article and can serve as little moments of storytelling. Moreover, these annotations enable the free exploration of the data and its context stories allow the user to follow just the information he/she is most interested in, also improving his/her enjoyability.

The participants showed a lot of interest in exploring the visualizations freely and seem to prefer to be moved by their own curiosity. However, the common strategy in most visualizations continues to be having a dense article with context information and only short annotations, something that is not recommendable since several users do not read the articles that accompany the visualizations. Nonetheless, according to Hullman, Diakopoulos, and Adar (2013), annotations are a promising way to complement articles since they have the capacity to add context that otherwise would be very difficult to provide. The context does not need to always be in the form of a dense storytelling and it can also be given in the form of external links and short annotations.

Another interesting feedback regarding interactivity is that the participants believed that the visualizations they liked the least could probably be fixed with an overlay of information, presented through the use of interactivity. According to them, the use of some interactivity such as hover or click details would change these visualizations completely, because it would make the data more meaningful. The participants also stressed that they enjoyed the visualizations that provided hyperlinks to other content and that this possibility does not prevent them from returning to the visualization. Although none of the examples shown had this feature, when confronted with the possibility of having external links, for example for Wikipedia pages, the participants referred that they would value this option. One of the participants even said that "the immediate reaction is to click on the links available."

9.2 Empathy

Storytelling can also add another dimension to the visualization: empathy. Together with emotion, empathy is a concept that is not often associated with information visualization, specially because emotion and empathy are usually associated with chaos and visualization is associated with objectivity. However emotive/empathic information visualizations revealed to be often more memorable (Kosara and Mackinlay, 2013) and even, at times, more enjoyable. This sense of empathy can be achieved by making the user relate to the topic or to individuals represented in the data (by allowing the user to see him/her represented or by putting the user in someone else's shoes).

Kosara and Mackinlay (2013) asked in *Storytelling: The Next Step for Visualization* what makes a visualization memorable. Everything seems to point out exactly to what they refer to as a possible cause: that the visualization is memorable when people relate to it. The issue of empathy kept coming up in the focus group discussion. The participants kept talking about how much they related or not to the visualizations' subjects, sometimes not even knowing the reason why they related with the visualization.

One of the best examples of this is *The New York Times' How many households are like yours?* The focus group participants selected this visualization as their overall favorite and in their responses they stressed the fact that what made this visualization so interesting was the fact that they could chose to explore families similar to theirs. They also said that they would like if there were smaller articles about each type of family so they could see if they have a similar lifestyle. One of the participants stated that when he chose a kind of family he was declaring an intention, therefore he would be interest in reading an article about the type of family that he chose and not another type of family even if they had a more interesting lifestyle.

In a test focus group we have done in the *University of Texas at Austin*, in the USA, the *Home and away: Iraq and Afghanistan war casualties* visualization was one of the most popular examples, mainly because the participants could relate with the subject. Most of the participants either had family members in the military or had friends that were stationed in Iraq or Afghanistan, therefore, they not only related more with this visualization but also felt very passionate about it. Not all the responses were positive though. Some of the participants felt almost offended to see that the visualization had pictures of the fallen soldiers and that it contained a lot of information about the soldiers, etc. Nonetheless, this response was a product of the participants close relation with the subject.

9.3 The Relation Between Time and Narrative

Another characteristic that the focus group participants appreciated was when the visualizations provided some temporal structure. In fact, according to Kosara and Mackinlay (2013, p. 47) "one of the fundamental features of stories is that they provide a temporal structure, even if not necessarily linear." Temporality is a major structural factor in our lives and it is closely related with narrativity (Ricoeur, 1990). Narratives are able to represent the human experience of time in its two different modes: the linear succession time (the sequence of minutes, hours, days) and the phenomenological time(the past, the present, and the future, which do not necessarily correspond to the linear structure of before and after, in other words, a narrative may begin with a culminating event or the temporality that is lived in the narrative may not concur with the time of the events the the story is said to depict).

Therefore if visualizations are able to not only introduce storytelling elements but also have a story flow they will be more successful and then can be considered narrative visualizations. A temporal structure is something that can give visualizations a sense of story-flow and this often appeals to users, because it gives them the ability to navigate their way to particular information. Structures such as timelines are very efficient in giving this temporal sequence feel. Nonetheless, this sense of temporality does not need to be expressed as a linear structure and stories are useful way to do so because they do not always have a linear temporal structure (Kosara and Mackinlay, 2013). A key research problem is to discover new visual metaphors for representing information and to understand what analytical tasks they support (Gershon, Eick, et al., 1998).

9.4 Gamification as a Way to Have Storytelling in Visualizations

Most visualization design nowadays follows the *Visual Information-Seeking Mantra* by Shneiderman (1996): Overview first, zoom and filter, then details on demand. In other words, the user should be able to see the big picture first (summarization), then to filter the information (browsing) in a way that does not limit his/her access to all the rest of the data, and be able to shift between this macro and micro levels of information, obtaining details on demand. Although this mantra is very useful in most visualization scenarios it does not mean that other alternatives do not exist. For instance when using strategies such as gamification to introduce storytelling it can sometimes be useful to adopt a strategy of showing a specific case and afterwards presenting the information about the big picture.

Gamification, previously seen in Subsection 4.3.7, consists of the application of strategies common to games (and not play or playfulness) design in non-gaming contexts in order to improve user engagement. According to Deterding, Dixon, et al. (2011), although the term has been around since 2008 it was only in 2010 that it seen a widespread adoption and, due to the criticism of the term by several researchers (Werbach, 2014), there are several other terms used to represent this same concept, such as *exploitationware* (Bogost, 2007) and *gamefulness* (Deterding, Dixon, et al., 2011; McGonigal, 2011).

Gamification is one of the strategies that can help to introduce storytelling and also create empathy in visualizations. After all computer games are the most popular example of interactive storytelling (Kosara and Mackinlay, 2013) and maybe their form of storytelling can be successfully replicated in visualizations. Nowadays, gamification is used in several different

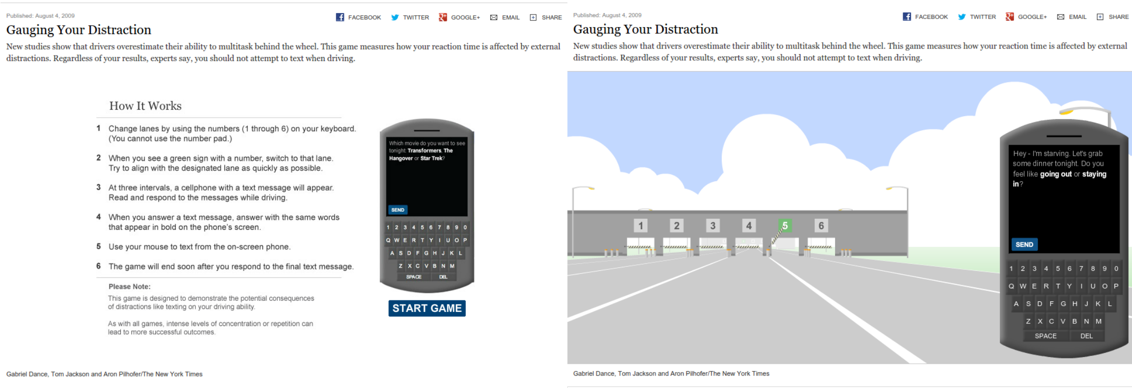


Figure 9.2: Two screens of *Gauging Your Distraction* by *The New York Times*

domains, from education to health, as a mean to encourage people to participate in an activity (using a system of points, badges, and/or leaderboards to convince people to do exercise and get in shape) or to facilitate understanding of an issue (by the use of simulation, as it is the case of *Gauging Your Distraction*[1], a game-y visualization by *The New York Times* that asks users to drive and text simultaneously, requiring the user to do both tasks well, shown in Figure 9.2).

SPENT (seen previously in Section 6.11) is one of the attempts of merging visualization and games. The Game lets the player make the everyday choices necessary to get by on a tight income: choosing a job, food to buy, pay the for the car insurance or take the son to the dentist, etc. SPENT has a timeline for the player to see in which day of the month he/she is, pictograms, animations, and tooltips. Once in a while it also gives additional data such as how many families choose not to go to the dentist because it is too expensive.

This additional data makes SPENT more than just a game, but is this enough to transform it into a visualization? According to Bogost et al. (2010, p. 47), "even if they are not games quite like Pac-Man or The Sims, infographics can become game-like, exploiting the properties of games in numerous ways: to encourage the manipulation of information for replayability, to allow pleasurable engagement with a system, or to invite exploration".

However, even though there are some successful examples of the use of gamification in information visualization, such as *SPENT*, *Budget Hero*, *Make your own hurricane*, and *Could you be a medallist?* (previously seen in Chapter 5), *Gauging Your Distraction*, *World Data Cup*[2], *Can You Live on the Minimum Wage?*[3], *The Great British Class Calculator*[4], *Rio+20 interactive: is the world getting better or worse?*[5], *The Sexperience 1000*[6], *HeartSaver*[7], and *Survive 125*[8]) there is not much research on this topic. One of the few attempts to approach this subject was carried by Macklin et al. (2009). To better understand the value of using gamification in visualization, Macklin et al. (2009) built three Game prototypes that are intended to illustrate different types

[1] http://rethinkingvis.com/visualizations/226
[2] http://rethinkingvis.com/visualizations/178
[3] http://rethinkingvis.com/visualizations/119
[4] http://rethinkingvis.com/visualizations/149
[5] http://rethinkingvis.com/visualizations/130
[6] http://rethinkingvis.com/visualizations/109
[7] http://rethinkingvis.com/visualizations/226
[8] http://rethinkingvis.com/visualizations/229

of data: *Kimono Colors*, *Mannahatta: The Game*, and *Trees of Trade*. The authors wanted to know if data is able to create play and if play is able to enlighten data, and if when the user plays with the data he/she is actually understanding the data and learning new things. The first prototype is a Game to learn the relationships between the materials and the dyes used to make traditional Japanese kimonos, where the user *fishes* for colors using a *lure* that represents a material used to make kimonos. "The lures, when placed on the screen by the player's mouse, attract the colors that are created with those ingredients" (Macklin et al., 2009, p. 3). The second prototype is a location-based mobile Game that makes the players walk around Manhattan interacting living and non-living elements that represent the ecology of the place 400 year ago. "The goal of *Mannahatta: The Game* is to achieve and maintain the status of *Eco-Master* across different blocks on the island of Manhattan by revealing *eco nodes* and forming sustainable *links* between them" (Macklin et al., 2009, p. 4). The final prototype is a Game that invites the user to "take apart the man-made products, find the trees from which they came and recreate them, un-kill the species that lived with that tree and leave the world with a complete and balanced forest ecosystem" (Macklin et al., 2009, p. 5).

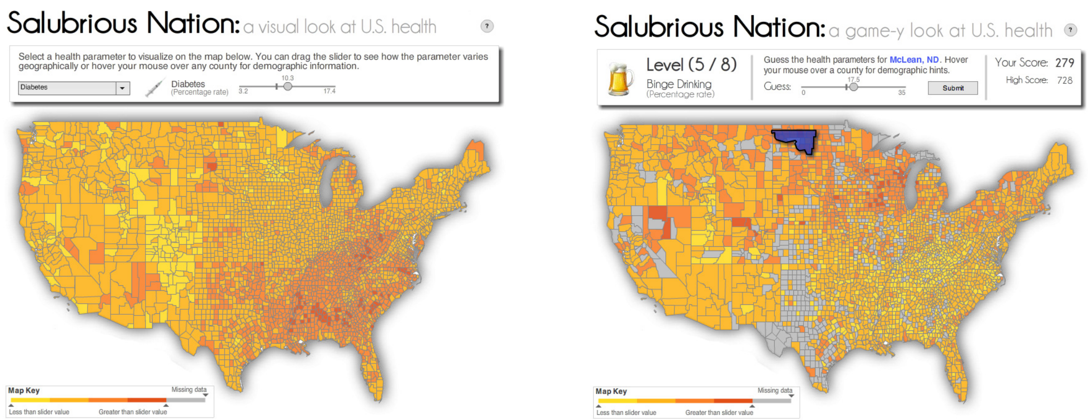
Although they found that some of the gamified versions were less straight forward for users than in the traditional visualization, Macklin et al. (2009) still conclude that certain game mechanics can be successfully applied to visualization. However this process takes time and has to be done thoughtfully. Nonetheless, they believe that "the next generation's scatter plots, bar charts, and sociograms may be better understood, not in a newspaper, but in a game" (Macklin et al., 2009, p. 7).

Another approach to this topic was carried by Nicholas Diakopoulos in a joint effort with other researchers (Diakopoulos, 2010; Diakopoulos, 2011; Diakopoulos et al., 2011). Diakopoulos (2011) points out that there are several challenges with the gamification of information graphics, specially when these deal with data that is variable through time (updated, refreshed, dynamic). He compares *game-y* information graphics (information graphics that include formal elements of games such as goals, scores, competition, and the notion of *winning* (Diakopoulos, 2010)) with traditional games, which usually benefit from a carefully developed design component and consequently take time to be released.

Most of the *game-y* information graphics, or playable infographics (the alike term coined by Bogost et al. (2010)), have been produced by news media (for example *Budget Hero*, by American Public Media, and *Can You Live on the Minimum Wage?*[9], by *The New York Times*) or marketing initiatives (*SPENT* by *McKinney*). The fact is that these organizations depend on deadlines and usually cannot invest too much time developing these types of visualizations. However the level of gamification used does not need to be as complex as in commercial computer games, nor the data ever changing as in *Salubrious Nation* by Diakopoulos et al. (2011). Specially in news media the information has a short life cycle and is often preferable to have a stable visualization than to have one that will forever be up-to-date.

Diakopoulos et al. (2011) conducted one of the few researches about the reaction to

[9] http://rethinkingvis.com/visualizations/119



(a) Salubrious Non-Game-y (Figure 2.a (Diakopoulos et al., 2011))

(b) Salubrious Guess for binge drinking (Figure 2.b (Diakopoulos et al., 2011))

Figure 9.3: Two systems for narrative visualization annotation

more game-like visualizations. They tested a *game-y* version of *Salubrious Nation*[10] against a *non-game-y* version of the same infographic. The objective was not only to explore strategies to introduce game mechanics in visual analytics, but also to the impact that these strategies have in user behavior.

Both versions (seen in Figure 9.3) consist of a map where, in addition to the demographic data acquired from the U.S. Census Bureau (population, poverty rate, life expectancy, proportion of people over age 65, etc.), eight health indicators (adult smoking rate, binge drinking rate, teen birth rate, fast food restaurants per capita, soda consumed per capita, diabetes rate, obesity rate, and heart attack death rate) are geographically coded by county, after being extracted from various online sources such as County Health Rankings and the Department of Health and Human Services' Community Health Status Indicators. The *non-game-y* version of *Salubrious Nation* (seen in Figure 9.3a) is still interactive but does not include the guessing game component. *Salubrious Guess* (seen in Figure 9.3b), the *game-y* version of *Salubrious Nation* that is still available online, uses geographically tagged public health data to create a Game where the goal is to accurately guess the extent of the given health parameter for a randomly selected target county. The data helps the player to do an informed guess. There is a third version that was also tested, *Salubrious Eliminate*. In this version, similarly to what happens in popular games such as Bejeweled where "the game mechanic entails matching groups of colors in order to eliminate game units, such as blocks or balls" (Diakopoulos et al., 2011, p. 1720), the user has to color the blank areas of the map with the current color creating a contiguous color area. "When a contiguous color area is created it is eliminated from play and the player receives points based on the size of the area cleared (i.e. the number of counties in the contiguous area)" (Diakopoulos et al., 2011, p. 1721).

The three versions were evaluated online relying on post-task recollection of insights. The participants were randomly assigned to one of the versions and were free to interact with the visualization for as long as they wanted. The users interactions (such as time spent on the

[10] http://salubriousnation.com/

task and the number of clicks, hovers, and drags) were recorded and afterwards they were asked to fill out a questionnaire where demographic information was collected along with "Likert scale ratings of subjective impressions of the experience such as enjoyability and learning, and what we expected to be covariates such as degree of interest in public health and experience with online graphics and casual games" (Diakopoulos et al., 2011, p. 1721).

The authors have concluded that sometimes the game features steers the attention away from the actual data. However it can also "successfully motivate interactions and cause users to explore and bias both the exploration of parameters and the nature of insights in interesting ways" (Diakopoulos et al., 2011, p. 1726). They were also able to conclude that *game-y* infographics are as enjoyable as *non-game-y* infographics and that it might be useful in helping to structure the interaction (Diakopoulos, 2010). Further research can provide insights on which types of gamification limit the users attention deficits or play with this fact in order to channel the users attention to particular data.

It would be interesting to see how the participants of the focus group study, presented Chapter 7, would react to more game like visualizations. After all computer games are the most popular example of interactive storytelling Kosara and Mackinlay, 2013 and maybe their form of storytelling can be successfully replicated in visualizations. However it would be very difficult to analyze this kind of visualization in a focus group environment, because the participants would need more time to explore it to get strong insights. They would probably even need to interact with the visualization more than once. The fact that evaluation is so difficult with gamified visualizations is one of the probable reasons why this type of visualization is rarely a topic approached by Information Visualization researchers.

9.5 Case Study

Taking an empirical approach, three professionally-produced visualizations and their utilization of narrative elements are analyzed, and it is explored how these could possibly be redesigned to better introduce narrative components. In order to illustrate this approach, I present three simple prototypes of the introduction of storytelling in the selected case studies. These three case studies were used to demonstrate three of the strategies of storytelling approached in this thesis. The first example, *How many households are like yours?* highlights how it is possible to introduce short stories to add empathy. The second, *What does china censor online?* illustrates how it is possible to add context. The last example, *Death penalty statistics, country by country* demonstrates the benefits of time as a form to introduce the feeling of narrative.

This approach aims to be a contribution somehow between a design study and a model, therefore the implications of these strategies are discussed, trying to shed a light on the impact they will have on the interpretation and level of enjoyability of the visualization. I was driven by the motivation to pave the way to future research on the impact of these strategies and on the establishment of design conventions for narrative visualization.

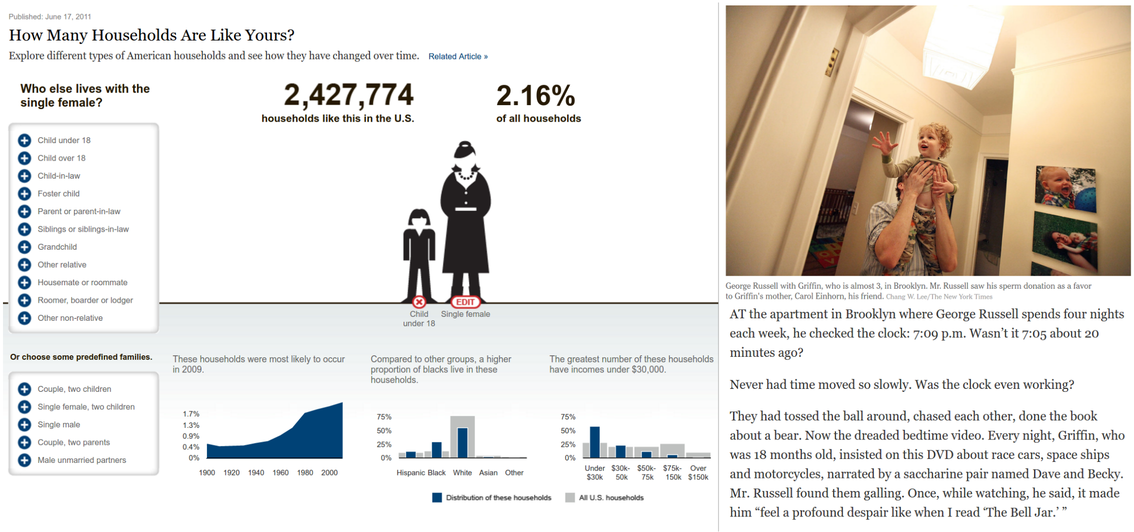


Figure 9.4: *How many households are like yours?* enhanced with storytelling

9.5.1 *How Many Households are Like Yours?* enhanced with empathy

In *How many households are like yours?*, previously presented in Section 6.7, the user is first presented with the option to choose the primary residents of a household (married couple; male/female unmarried partners; single male; single female; male unmarried partners; and female unmarried partners), represented through pictograms. Afterwards the user can add secondary members of the household (child under 18; child over 18; child-in-law; foster child; parent or parent-in-law; siblings or siblings-in-law; grandchild; other relative; housemate or roommate; Roomer, boarder or lodger; and other non-relative), also represented as pictograms. The graphic updates on the fly and simultaneously shows how the entered household compares to the rest of America's households. The visualization shows the total of households in the US that are like the one the user selected and the respective percentage. On the bottom there is a breakdown by time, race, and household income.

In terms of interactivity *The New York Times* visualization enables the user to click and hover details and filter the data. The narrative elements we identified in it were title, captions, annotations, Introductory text, and accompanying article. Considering that *How many households are like yours?* has a written article to introduce the visualization,it is considered to have a Martini Glass visualization structure, according to the categories for structure by Segel and J. Heer (2010).

The user is presented with the possibility to choose any kind of household that he/she wishes however the visualization challenges the user to try his/her own family. This creates a sense of proximity between the user and that data. The user will consequently relate to the data presented because the results will shown him/her information about families that are similar, and possibly that share the same characteristics (in terms of income and ethnicity) of his/her family. This feeling of relatability is very appealing to the user, however it can be enhanced with storytelling.

It is possible still to improve the sense of relatability that the user feels with the data if, instead of having only one long article about one type of family introducing the visualization, short stories about the different kinds of families are introduced for every type of family possible. In Figure 9.4 it is presented how this can be done without changing the visualization too much. Basically, similarly to what happens with the graphics on the bottom of the visualization I propose that the visualization includes also a short article characterizing the type of family that the user selected. I used the main article that accompanies the original visualization as the example for household with a single female with a child under 18. Having stories for each type of household helps the user to see that data not just as a type of household but as real people, real families just like the user's family.

Relatability is a factor that helps the user enjoy the visualization. It is one of the characteristics that makes it memorable (Kosara and Mackinlay, 2013) and one that is able to make the user feel empathy with the subject or the individuals represented in the data, and which will probably make the visualization more successful.

9.5.2 *What Does China Censor Online?* enhanced with context

The visualization, by David McCandless, *What Does China Censor Online?*, previously presented in Section 6.5, is a simple Tag Cloud that only has a title and text, in this case mere disconnected words. The tag cloud has the shape of the map of China, but this is not immediately visible to the user, unless he/she knows the shape of the country. *What Does China Censor Online?* is a static visualization.

This visualization would benefit greatly with the addition of extra information to add

Figure 9.5: *What does China censor online?* enhanced with storytelling

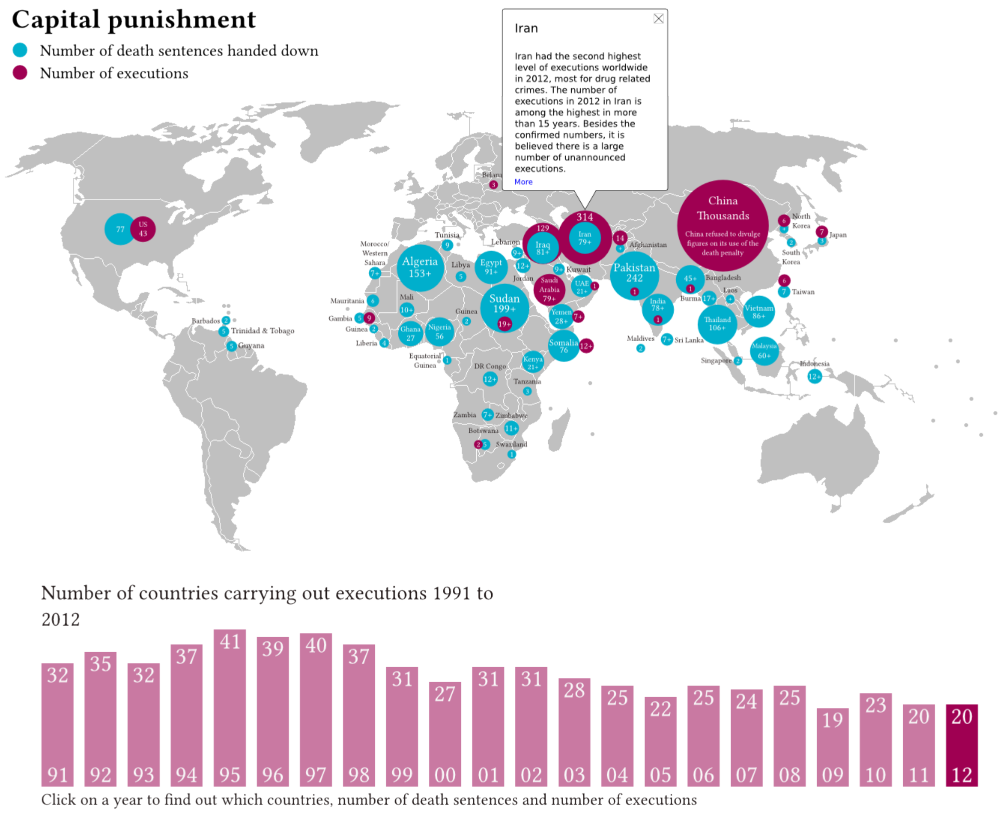


Figure 9.6: *Death penalty statistics, country by country* enhanced with storytelling

context. In Figure 9.5, I present how this can be done maintaining most of the original design. I propose the introduction of small *tooltips* that pop-up when the user clicks on one of the websites censored. This would help the user realize the possible reasons for the censorship and get additional insights. Moreover, it avoids that the user feels forced to exit the visualization to learn more about the topics he/she is interested. There should also be external links to the actual websites that are being censored.

Context information can work as little moments of storytelling. This kind of short stories can be more easily interpreted by the user than a dense article. Context could also be introduced as external links, for instance to Wikipedia pages, related articles, or short annotations.

This context information is beneficial for providing information that otherwise would be difficult to provide (Hullman, Diakopoulos, and Adar, 2013). Without this extra information the visualization does not have much utility and it is just an eye-catching but shallow visualization.

9.5.3 *Death Penalty Statistics, Country by Country* enhanced with time

Death penalty statistics, country by country, previously seen in detail in Section 6.3, is a visualization by *The Guardian* that accompanies an article about countries that maintain the death penalty. The map/diagram static visualization has bubbles of different sizes to represent the number of death sentences handed and executions in countries that are still carrying executions. On the bottom there is also a timeline representing the number of abolitionist countries for each year between 1991 till 2012. The timeline resembles a bar graph. Apart from the large article of which the visualization is part of, *Death penalty statistics, country by country* in terms of narrative elements only has an Introductory text and captions that indicate the short information such as names of countries and dates.

This visualization would probably benefit if the timeline would actually function as a navigation and when the user clicks a certain year the map would show the number of death sentences handed and executions of that year, as it can be seen in Figure 9.6. This representation of time and, specially the evolution of events, often appeals to users, as it was mentioned in the focus group study presented in Chapter 7.

The use of interactivity elements such as hover or click details would also be useful on a visualization such as this one, because it could add extra information making the data more meaningful. In the prototype shown in Figure 9.6, I propose adding *tooltips* with extra information about the executions and death sentences when the user clicks each countries' bubble, in this way adding more context to the data. This *tooltip* could have general information about the subject for each country in a given year or a particular execution story, which could increase the empathy between the user and the data.

9.6 Chapter Summary

In this chapter I approach four strategies to introduce storytelling in visualizations: 1) interactivity and its relation with context (section 9.1), 2) empathy (section 9.2), 3) the relation between time and narrative (section 9.3), and 4) gamification (section 9.4). Each of these strategies, with the exception of *gamification*, are illustrated with case studies. There is no case study of the *gamification* strategy due to its complexity. However, section 9.4 includes the review of several existing visualizations that use gamification and the analysis of one of the few examples of evaluation of *game-y* visualizations: the study and comparison of the *game-y* and *non-game-y* versions of *Salubrious Nation*.

The case studies consist of mock-ups that illustrate the inclusion of each strategy in an existing visualization. This is an approach somehow between a design study and a model and tried to shed a light on the impact the introduction of each narrative strategy will have on the interpretation and level of enjoyability of the visualization. The motivation behind this approach was the intent to pave the way to future research on the impact of these strategies and on the establishment of design conventions for narrative visualization. One possible direction for future research is, similarly to what Diakopoulos et al. (2011) did for *Salubrious Nation*, to evaluate a version with the inclusion of a narrative strategy and a version devoid of it, in order to compare both versions in terms of the impact that the narrative strategies have in the users' experience.

Chapter 10

Techniques for visualization on the web

Most research on narrative visualization has either focused on general guidelines to develop this kind of visualizations, in the description of the development of specific visualizations tailored to tell stories, or in developing tools to create narrative visualization. The review of general guidelines is important to choose which narrative elements work for information visualization but it does not help visualization creators to actually build narrative visualizations. Similarly, research that describes the creation of a particular visualization can also help to pinpoint difficulties in the process of creating a narrative visualization but commonly ends up just describing a process that would be difficult to replicate . Research that focus in the development of new tools for narrative visualizations is probably the kind of research that benefits visualization creators the most, because it provides them tools to do it. However, learning new tools is time consuming and, unless the new tool is very easy to use or provides new visualization possibilities that no other tool provides, visualization creators will usually prefer to stick to tools they already know.

Therefore, there is the need for more research on the whole process of creating a narrative visualization that can serve as guidelines to build new visualizations and that facilitates this process. Lee et al. (2015) also identified research opportunities on: how to make compelling stories and guidelines on how to chose a plot; tools that facilitate the process of telling a story (tools that allow enough flexibility to tell more complex stories that support sophisticated interactions and that are easy enough for people with less programming skills to use); on holistic tools that "combine data analysis, scripting, editing, and presenting functionality all into one tool or a suite of tools" (Lee et al., 2015, p. 89); understanding in which situations to resort to narrative visualization; and identifying emerging scenarios for which narrative visualization can be a good option to visualize the data. Evaluation is also a research topic

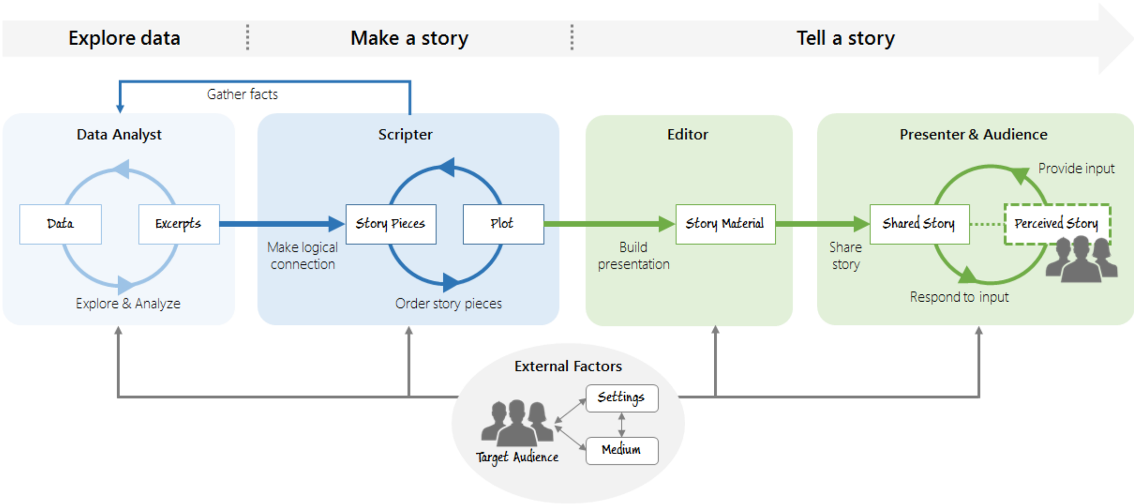


Figure 10.1: Storytelling process (Figure 2 (Lee et al., 2015))

that both in narrative visualization and in traditional information visualization still does not get enough attention, as it was pointed out in Subsection 2.2.2. The evaluation of particular visualizations, tools, and algorithms has been approached by several researchers in traditional information visualization (Burkhard and Meier, 2005; Plaisant, Grosjean, et al., 2002; Saraiya et al., 2005; Shanmugasundaram, Irani, and Gutwin, 2007; Shi et al., 2005; Vande Moere, Tomitsch, et al., 2012; Willett et al., 2011) and there is also some research how to do the evaluation (J. Heer and Agrawala, 2008; J. Heer, Mackinlay, et al., 2008; Kim et al., 2014; Lam et al., 2012; Pandey et al., 2014; Plaisant, Fekete, et al., 2008; Yi et al., 2008), but there are still several research opportunities to evaluate visualizations that are at the edges of traditional information visualization.

To better understand what is needed to build a narrative visualization, Lee et al. (2015) propose a model for visual data storytelling process (VDSP), which is presented in Figure 10.1. "The VDSP summarizes the main roles and activities that visualization storytellers engage in as they turn raw data into a visually shared story, along with the types of artifacts that result from these activities" (Lee et al., 2015, p. 86). The process is divided in three parts: explore data (which also includes the analysis of the data), make a story (which concerns the process of assembling the data in an interesting and compelling way by organizing the data, establishing connections, etc.), and tell a story (which includes building final presentation, sharing it, and handling the feedback). This process does not necessarily need to start at the part that corresponds to the exploration of the data and the creator of the visualization can revisit the same part of the process multiple times.

In order to build Ellipsis[1], a visualization tool to support narrative devices, Satyanarayan and J. Heer (2014) also conducted a study to understand how narrative visualizations are crafted in order to build a tool for storytelling. The study included interviews to two journalism graduate students and two Knight Journalism Fellows, both with experience in authoring data-driven stories but with web development expertise ranging from beginner to intermediate.

[1] http://idl.cs.washington.edu/projects/ellipsis/

They were asked about about the process they follow when crafting data-driven stories and their answers do not fall far from the steps that Lee et al. (2015) describe. According to Satyanarayan and J. Heer (2014, p. 362) "despite different working styles and expertise, all four respondents described a three-phase design process: exploration to uncover interesting stories in data sets, drafting to prototype ways of communicating the stories they found, and production to develop the final interactive."

The interviewees were also inquired about which tools they used to build the narrative visualizations and referred that different tools were used in different stages of the development (Satyanarayan and J. Heer, 2014). Tools such as Microsoft Excel, Tableau[2] or R/ggplot2 were used to build static visualizations that help to discover compelling stories. For prototyping Microsoft Powerpoint or Apple Keynote are used to test annotations, animations, and sequences. Finally, tools such as Fusion Tables, Tableau, Adobe Flash, and D3 are the favorites for the production phase.

Nonetheless, the interviewees also pointed some challenges that each tool poses. Although they complain that easier tools such as Fusion Tables and Tableau do not fully support storytelling, they also think that tools such as Adobe Flash and D3 require too much knowledge to be used. More sophisticated tools usually require an experienced programmer, but "in smaller companies, there may not be a visualization developer, leaving journalists little recourse beyond Fusion Tables and Tableau" (Satyanarayan and J. Heer, 2014, p. 362).

10.1 Tools for information visualization

Research in information visualization has for long been focused in developing sophisticated tools for creating visualizations and "as the field of information visualization matures, the tools developed in our research laboratories are reaching users" (Shneiderman and Plaisant, 2006, p. 2). The positive reception of these tools by the general public (outside academia) has sparked the interest in researching more user friendly tools and evaluating the efficiency of existing tools.

Nowadays, there is a panoply of tools to visualize data which vary in type (library, web or desktop application, Application Program Interface (API), toolkit, platform, and programming language), in the level of coding that is necessary to build the visualizations (ranging from a lot of coding to no coding at all), and in technology used (JavaScript, Adobe Flash, Python, Processing, R, HTML to name a few). This variety also contributed to the ever-growing popularity of visualization. The fact that there is something for every type of user, from experts to novices, and for every desired outcome, from traditional information visualization to narrative visualization, made it possible that this visual representation technique, which has already proven to aid cognition and understanding, started to be applied to different fields where data and information has to be pre-processed to be easier to interpret. According to Masud et al. (2010, p. 445) "this ever-growing diffusion of visualization tools even in non-expert contexts and during decision making processes and planning phases (opposed to an exclusively

[2] http://www.tableau.com/

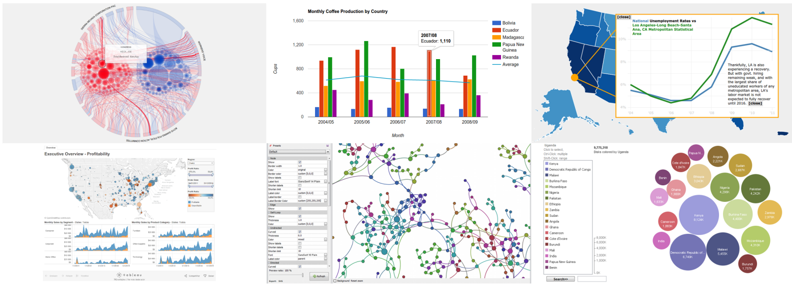


Figure 10.2: Visualization tools (D3, Google Charts, Ellipsis, Tableau, Gephi, and Many Eyes)

analytical approach) requires a more open way of thinking about visualizations, able to take in account the *world out there*, the *real world*."

In Figure 10.2 it is possible to see six popular tools of different types, level of coding knowledge, and technology: D3, which is a JavaScript library for data visualization that requires programming skills; Google Charts[3], which is a free web-based tool to create web-based charts, graphs, and tables; Ellipsis, which is a graphical interface for storytelling; Tableau, which is a set of tools for data analysis and visualization, targeted for business intelligence; Gephi[4], which is a desktop application for data analysis and visualization that requires no coding; and Many Eyes, which is a collaborative data visualization web application that requires no programming or technical expertise (discontinued on June 12 2015 but included in IBM's *Watson Analytics*[5]). These and other visualization tools will be reviewed in more detail next.

10.1.1 Tableau

Tableau is one of the most popular visualization tools probably because it is so easy to use. Being targeted for business intelligence, Tableau is a set data analysis and visualization tools that supports a wide variety of charts, graphs, maps, etc.

The company has free versions but also sales software licenses and charges maintenance/services fees. Tableau offers two types of tools (developer and sharing tools) divided in six main products (Tableau Desktop, Tableau Server, Tableau Online, Tableau Reader, Tableau Mobile, and Tableau Public). Tableau Desktop and Tableau Public are developer tools. The Desktop version is a full developer software to analyze and visualize data, available in a personal and a professional version, and is not free to use. The data sources can be spreadsheets, a SQL databases, Hadoop datasets, or the cloud. The Professional edition of Tableau Desktop is compatible with Tableau Online and Tableau Server, which are complementary products specially targeted for businesses. Tableau Server, is an analytics application that provides all of the features of Tableau Desktop in a browser. On the other hand, Tableau Online is the hosted version of Tableau Server and allows its users to share insights in the cloud. This version of Tableau enables that the data is shared securely and only authorized users can interact with the data and dashboards. Both Tableau Online and Tableau Server require Tableau Desktop Professional Edition.

[3] https://developers.google.com/chart/
[4] https://gephi.org/
[5] http://www.ibm.com/analytics/watson-analytics/

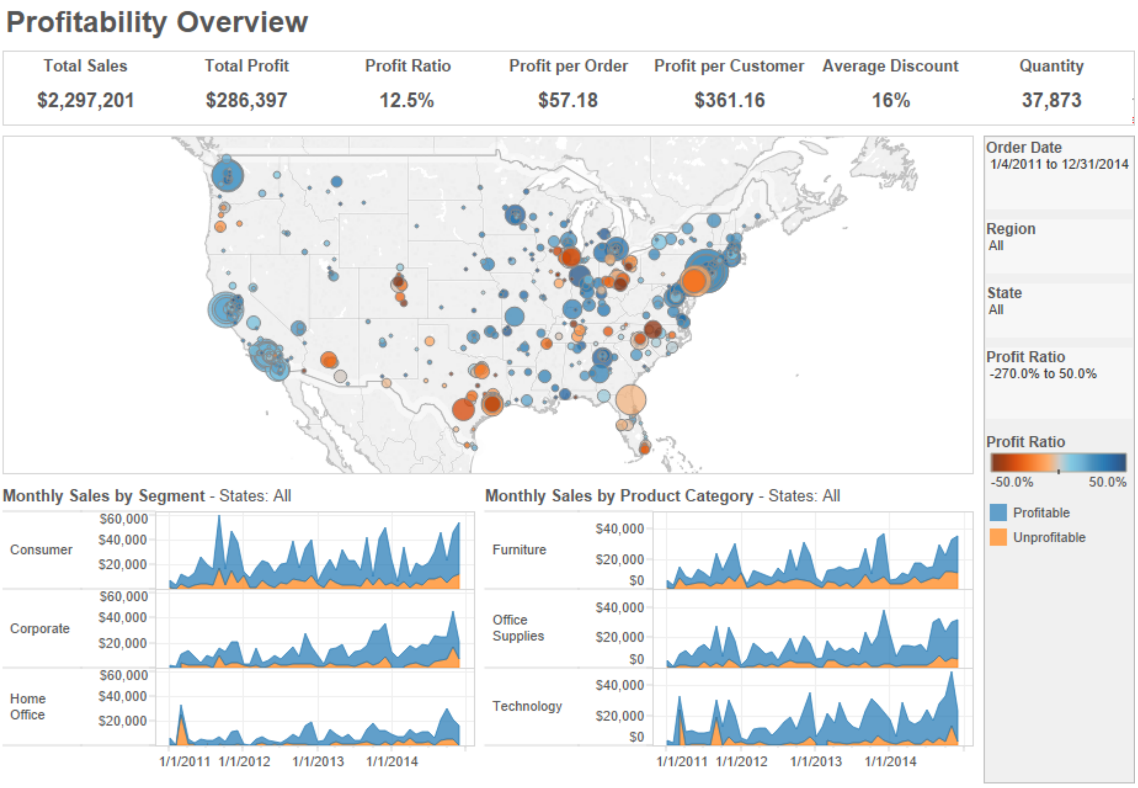


Figure 10.3: Visualization with Tableau

The Public version is free, but its name stands for the fact that the visualizations created with this version are public and can be seen by anyone. These visualizations can only be saved and shared through Tableau Public and the data sources must be Microsoft Excel and/or text files. Some users complain that the Public version includes a big footer in the visualizations that shows that these were created through Tableau.

Tableau Online, Tableau Reader and Tableau Server are sharing tools. Tableau's sharing tools allow users to share the visualizations and dashboards with others in order to enable interaction and shared exploration of the data. Tableau Reader is a free desktop application that allows users to open and interact with visualizations built in Tableau Desktop. This version has no security and anyone that receives the visualization can use Tableau Reader to open it.

Tableau is considered to be very intuitive and does not have a steep learning curve. One of its strongest points is the fact that Tableau is a complete solution that allows data analysis and visualization.

10.1.2 D3.js

Like Tableau, D3 is a very popular tool to create visualizations, however it takes a very different approach to the creation process. Instead of its first goal being to be easy to use, D3 intends to be a sophisticated and powerful tool that allows flexibility both in terms of creation and in terms of distribution. An example of a visualization made using D3 can be seen in Figure 10.4.

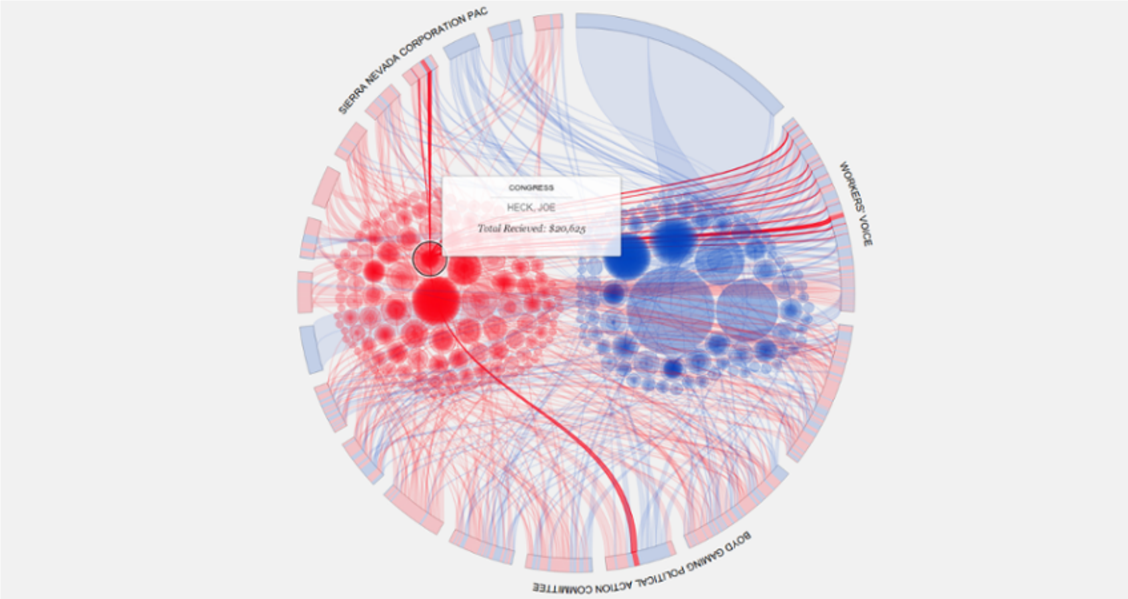


Figure 10.4: Visualization with D3

First released by Mike Bostock in 2011 as a successor to the *Protovis* toolkit, which was a JavaScript library to create SVG graphics from data, its name stands for Data Driven Documents (D3) and it uses existing web technologies such as HTML, Cascading Stylesheets (CSS), and SVG to create sophisticated visualizations. Although it can be used to create charts and maps, D3 is not a chart library or a map library, it is a general purpose visualization library. Moreover, it not only supports the use of large datasets, but also allows the visualization creator to use dynamic behaviors for interaction and animation.

According to Dewar (2012, p. 2) "a huge benefit of how D3 exposes the designer to the web page is that the existing technology in the browser can be leveraged without having to create a whole new plotting language." Because D3 uses existing browser technologies the visualizations created using it are not limited to a small region of the web page like it happens with canvas element based libraries or tools such as Processing.js[6]. Therefore it is fairly simple to use if the user has web development experience. Furthermore it is an open source tool and the user can not only use it but also contribute for its improvement.

Although its gallery provides a nice selection of examples that include the code to build those visualizations, to create more elaborate visualizations with D3 the user has to have a deep proficiency in JavaScript. For novice users D3 can have a steep learning curve, but D3 has a lot of documentation[7] available and nice community support, thus users with less experience will still be able to build a visualization as long as he/she sticks to less complex visualization examples. Document Object Model (DOM) manipulation can also be a challenge since it can be slow for large sets of data because SVG has performance limitations when it comes to drawing lots of elements on the screen. However, humans also have a lot of difficulty dealing with too many elements, therefore a good visualization should limit the number of elements shown at a time.

[6] http://processingjs.org/
[7] https://github.com/mbostock/d3/wiki

D3's biggest downside must be the fact that it has problems with older browsers. Backwards compatibility can with older browsers can be arranged resorting to tools such as *Sizzle*[8], but it can be hard for non-expert users to come up with these solutions. Nevertheless, older browsers will slowly decrease in number of users and hopefully soon all web users will stop using less stable, insecure, slow browsers and this will stop being a limitation for D3.

10.1.3 Ellipsis

Ellipsis is a visualization tool built with the concern of providing ways to support storytelling techniques. The creation of this tool was informed by a study which included interviews with two journalism graduate students and two Knight Journalism Fellows, both with experience in authoring data-driven stories but with web development expertise ranging from beginner to intermediate. This study was conducted in order to understand how narrative visualization are crafted. The respondents emphasized the fact that most visualization applications provide little support for narratives and the ones that support storytelling techniques often require significant programming ability. Consequently, journalists are often removed from a large part of the visualization creation process: "they are responsible for initial designs, but then pass their content to a developer who controls the production phase" (Satyanarayan and J. Heer, 2014, p. 361).

Therefore, having in mind the narrative devices identified by Segel and J. Heer (2010) (*tacit tutorials*, *semantic consistency*, and *matching on content*) and the four editorial layers where rhetorical decisions occur identified by Hullman and Diakopoulos (2011) (*data*, *visual representation*, *textual annotations*, and *interaction*), Satyanarayan and J. Heer (2014, p. 361) intended "to provide tools to support these narrative devices" without programming, consequently allowing less experienced users to build narrative visualizations. They wished to develop "design tools for narrative visualization that support this process could improve efficiency and empower journalists to collaborate with developers" (Satyanarayan and J. Heer, 2014, p. 361).

According to Satyanarayan and J. Heer (2014), Ellipsis combines a JavaScript-based domain-specific language (DSL) with a direct-manipulation interface, in order to minimize programming. Ellipsis was designed to add the storytelling elements to existing visualizations, therefore the visualizations have to be registered within the DSL. Afterwards "the other components of the storytelling model can be instantiated through the UI" (Satyanarayan and J. Heer, 2014, p. 365) but are then translated into DSL statements. Nonetheless, the visualization and the narrative are treated as two independently editable layers.

Ellipsis allows the author of the visualization to create scenes (which also have transitions defined using a syntax of "if this, then that") and draw annotations. The annotations can be simple shapes such as rectangles, ellipses or arrows, and text or labels with configurable properties such as position, color, size, and other style rules. These annotations can be bound to data. In other words, if a circle annotation is bound to its data value, the radius of the circle will reflect the quantity of data. These annotations can also be added to the scene only when

[8] http://sizzlejs.com/

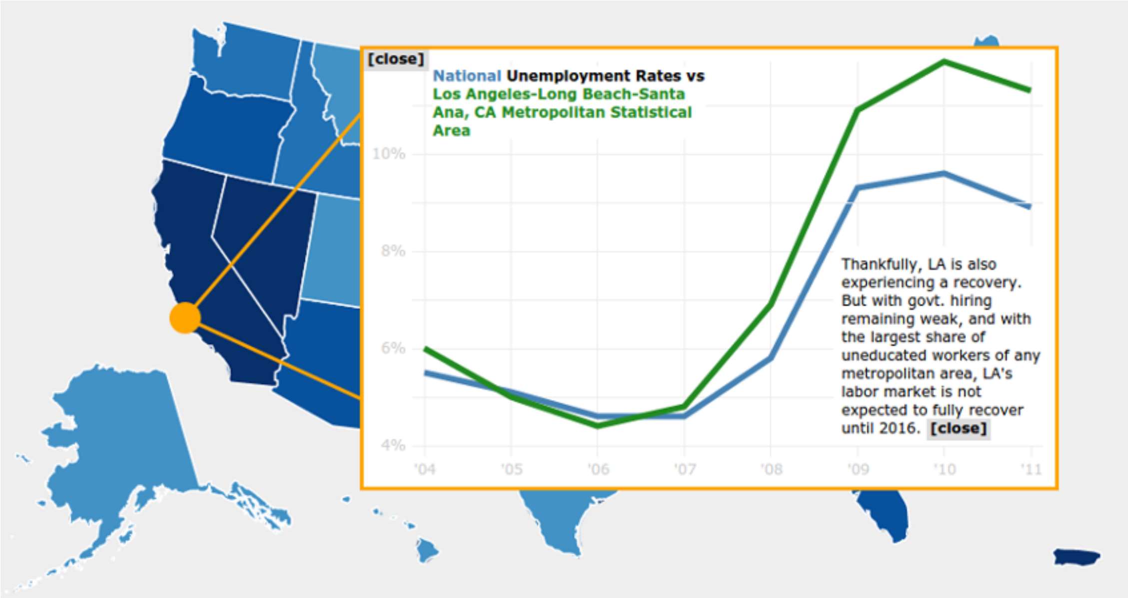


Figure 10.5: Ellipsis

triggered by a certain action of the user. An example of a visualization made using Ellipsis can be seen in Figure 10.5.

Satyanarayan and J. Heer (2014) evaluated Ellipsis by conducting user studies with eight professional journalists and found out that Ellipsis can be useful as a rapid prototyping tool. The participants were able to build their narratives without much guidance and appreciated the visual cues that the tool provided in order to facilitate their task. "They authored stories composed of 2–3 scenes, added annotations to highlight and label points of interest, added timer triggers to animate these annotations, and click event triggers to navigate between scenes" (Satyanarayan and J. Heer, 2014, p. 369). However, they also pointed out some weaknesses: the interfaced could be improved, and highlighting using shapes can be heavy-handed. The truth is that this tool is still very recent and its interface is still not as expressive as its the DSL. Nevertheless, in the views of Satyanarayan and J. Heer (2014, p. 369), "improving the expressivity of the GUI poses a challenge that relates to larger issues of specifying visualizations without programming."

10.1.4 Other tools: Many Eyes, Google Charts, and Gephi

A myriad of visualization tools have been released in the last ten years, but not all survived or had much success. **Many Eyes** is the example of a pretty successful tool that eventually faded away and now is discreetly incorporated in IBM's new analytics product *Watson Analytics*[9] but without the core characteristics that attracted so many users in its golden years.

Launched in 2007 and closed in 2015, Many Eyes was "a public web site where users may upload data, create interactive visualizations, and carry on discussions" (Viégas, Wattenberg, Ham, et al., 2007, p. 1121). In previous research Viégas, Golder, et al. (2006) found that

[9] http://www.ibm.com/analytics/watson-analytics/

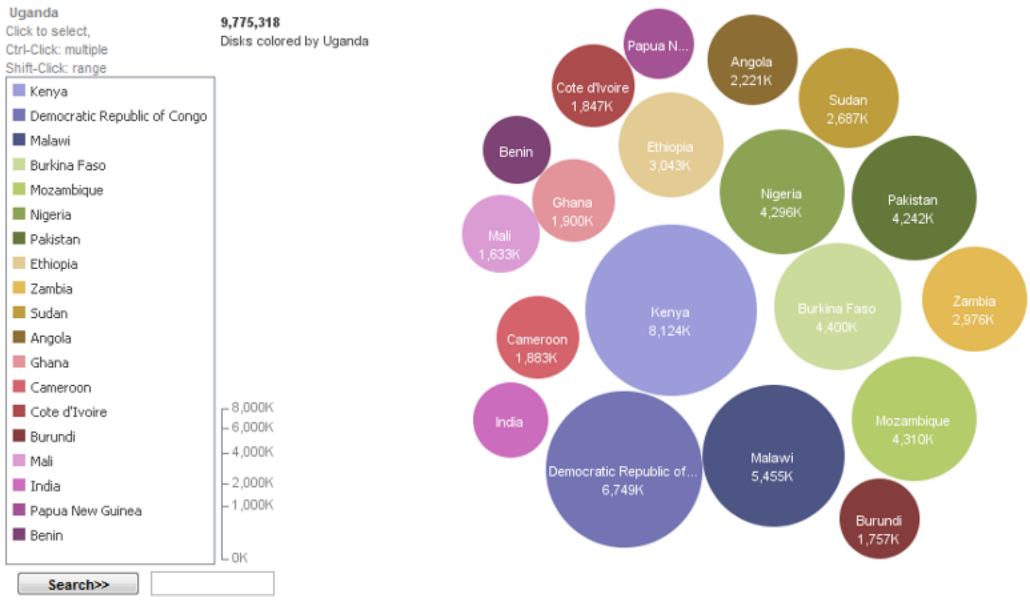


Figure 10.6: Many Eyes

visualization had a strong social component and that a big factor in user engagement seems to be the capability of discussing findings, setting each other puzzles, and drawing inspiration from other users (Wattenberg, 2005). The goals of Many Eyes were to simplify the process of creating the visualization and promoting social data analysis. Users could: upload their own data, which once uploaded would become public and available for everyone to download it and reuse it; create interactive visualizations, being able to choose from a considerable variety of visualization techniques (world maps and maps of the US; classic charts such as line charts, bar charts, bubble charts, and scatter plots; tree diagrams; tag clouds, etc.); and discuss the data, allowing users to comment and include a snapshot of a particular state of the visualization, which takes other users to the exact same configuration in order to let others know what they have discovered. An example of a visualization made using Many Eyes can be seen in Figure 10.6.

Another reason for Many Eyes' popularity was the fact that it was simple to use. Users were able to create interactive visualizations and make them available online with a couple of clicks because the system very user-friendly. Users did not have to have any programming or technical expertise. According to "while visualizations have existed on the web for more than a decade, these have been constructed offline and then separately published to the web" and Many Eyes was one of the first solutions to rely on a pure web-based model. This fact also allowed for it to reach a large audience.

In 2005, Wattenberg (2005) saw in the understanding of the patterns of social data analysis a promising area for future research. However, despite of the success that the social component of Many Eyes had, most visualization tools nowadays still discard this feature. With the exception of tools strongly focused on analytics (mainly business analytics) not many visualization tools still invest in providing strong means for collaboration and participation. However, Satyanarayan and J. Heer (2014, p. 362) believe that the success of social data analysis systems such as Many Eyes indicates "that some users are eager to explore datasets and share their own data stories", and that is why they believe that they should integrate more features that promote social data analysis in Ellipsis in the future.

Gephi, similarly to Many Eyes, is also a visualization software written in Java. However, Gephi is focused on graphs and network diagrams and it is not a web tool, being available for download for Mac OS X, Windows, and Linux. As it happened with D3, this visualization software was developed in an academic environment and was released as an open source project, allowing contributions from the community.

According to Bastian et al. (2009), because it is an attractive and technically accurate visual representation, for many years there has been an interest in using graphs to better understand networks. Therefore, "network exploration tools must head toward real-time visualizations and analysis to improve the user's exploratory process" (Bastian et al., 2009, p. 361). Interactivity is also an important aid to the exploration of large networks.

Gephi was built having all of this in mind: it allows data analysts and scientists to create dynamic, real-time graph-based visualizations, in order to get insights from the data. The software allows the users to interactively explore the graphs, manipulating the structures, shapes, and colors, isolating structure singularities, filtering, clustering, and analyzing additional statistics and metrics. The data can be introduced in a variety of formats even if the data is dynamic: "for instance a web-crawler can be connected to Gephi in order to see the network construction over time" (Bastian et al., 2009, p. 362). Gephi allows the visualization of networks with up to 100,000 nodes and 1,000,000 edges but it is still very fast and efficient.

Even though it does not require programming skills, Gephi requires the user to deeply understand network analysis in order to be able to take full advantage of the tool. Gephi is a tool that would be more suitable for researchers or data analysts, however it has also been used in data-driven journalism.

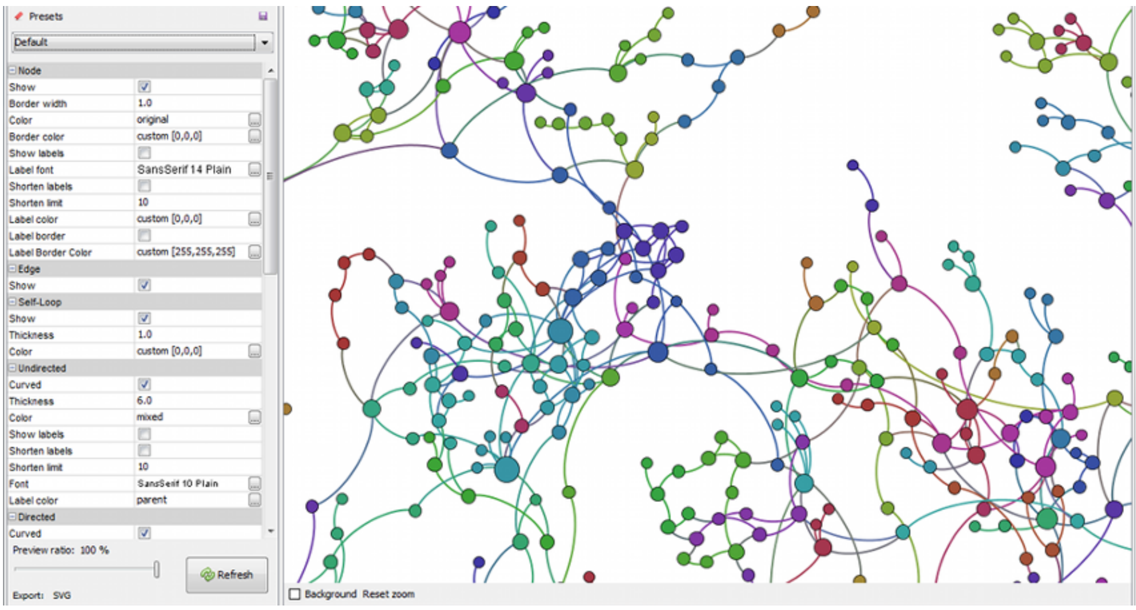


Figure 10.7: Gephi

Google Charts and **Fusion Tables** are two different tools provided by Google to visualize data. The first is a library that allows users to create charts and embed them in a web page. It contains pre-built and ready to use charts, ranging from the very basic bar chart to the more complex treemap, which can be customized to the user's needs. The charts are rendered in HTML5/SVG which provides cross-browser compatibility and cross-platform portability (allowing them to work on mobile as well) without any plugins. It also allows for backwards compatibility using Vector Markup Language (VML) for older versions of Internet Explorer. An example of a visualization made using Google Charts can be seen in Figure 10.8.

Google Charts is very user-friendly and can be used by users with little programming experience (some notions of JavaScript and HTML), unlike tools such as D3. It also provides some simple interactivity features, such as small labels that show-up when the user hovers the chart. However, Google Charts does not provide the same variety of visualizations as sophisticated visualization tools such as D3 and the user will only have access simpler charts. Even though Google Charts provides some styling parameters these are not as flexible as, for instance, styling elements via CSS. One of the most interesting features of Google Charts is the fact that it allows the user to connect to dynamic data.

One of the data sources that can be used with Google Charts, in addition to *Google Spreadsheets* and *SalesForce*, is Google Fusion Tables. This is a data management and visualization tool that not only allows users to work with more complex data sets, but also provides them more intricate and innovative visualization options. Although it is primarily geared towards mapping, it also allows users to create other types of visualizations. Fusion Tables "was originally designed for organizations that are struggling with making their data available internally and externally, and for communities of users that need to collaborate on data management across multiple enterprises" (H. Gonzalez et al., 2010, p. 175) but it has seen a wide adoption namely in

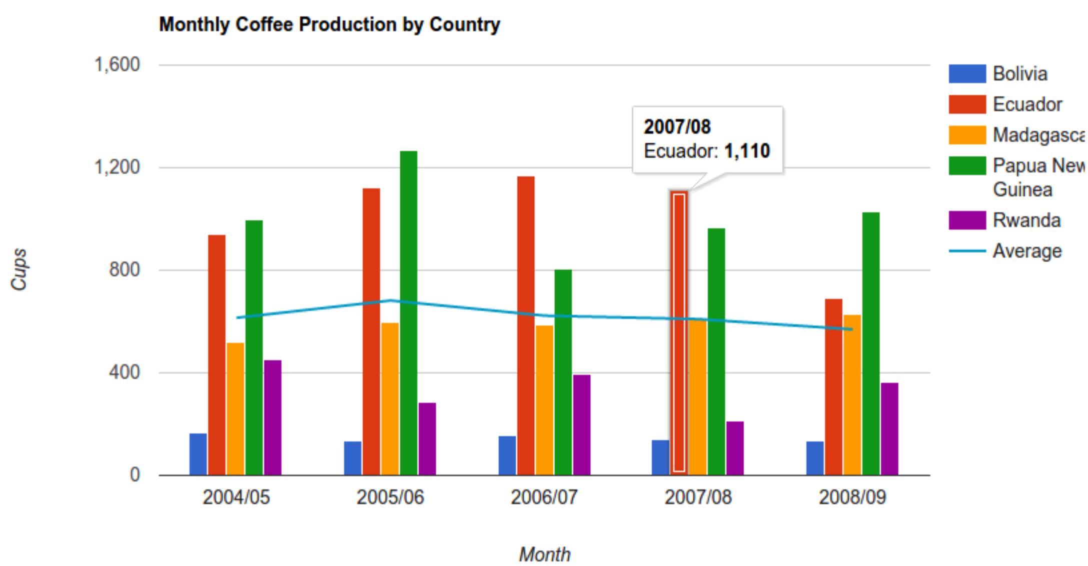


Figure 10.8: Google Charts

journalism, for instance in The Guardian[10].

Google Fusion Tables is very similar to *Google Spreadsheets* and its data is structured in the same way. However, Fusion Tables is more convenient with bigger datasets (providing bulk operations that are usually not necessary for smaller data collections, such as filtering, aggregation, and merges), does not have the same flexibility as *Spreadsheets* (not allowing the user to put any value in any cell or customize cells and columns), is better for quick changing data (facilitating these changes and propagating them everywhere that data is being used), and facilitates collaboration (allowing to quickly find or set apart data from different parties if necessary). According to H. Gonzalez et al. (2010) one of the strengths of Fusion Tables is that it facilitates the users job by limiting the visualization types available to the ones that can be build with the data types found in the user's table. "For example, a scatter plot is available only if at least two numeric columns exist, one for the x axis and one for the y axis. Similarly, a map is available if we detect a location column, e.g., a column with street addresses, or a column with latitude and longitude values" (H. Gonzalez et al., 2010, p. 178).

10.1.4.1 Tools that are not specific for visualization

There are lots of tools that are not built specifically for visualization but that can still be used to create great visualizations. This is the case of statistical graphics software packages and tools such as R, ggplot2, Plotly and matplotlib. These might not be the names that first come to mind when thinking about data visualization, but their inherent rigour is appealing, specially for data analysis. Moreover, these are often used for the initial data analysis and exploration before moving to more interactivity oriented tools such as D3.

R is a programming language and software environment for statistical computing and graphics. It is freely released under a *GNU General Public License* and it can be used on Linux, Mac OS, and Windows machines. R is a powerful tool, however with great power comes great complexity. R's learning curve is considerably steep, nonetheless its strong community provides good support and the several package libraries makes it an interesting tool to use for visualization. According to a recent survey[11] by RJMetrics (2015), R is still the most listed skill by data scientists on *LinkedIn*. It is listed by 48.01 percent of the 11,400 data scientists currently employed by companies known to LinkedIn that were identified in the study.

ggplot2 is one of the many R packages and is widely known for making it easier to use R for creating statistical graphics while still taking advantage of its power. While base R graphics can be plain and not easy to customize, ggplot2 allows its users to create aesthetically pleasing graphics by taking care of decisions about things such as dimensions or colors, nonetheless still allowing its users to customize the graphics if they want. According to Wickham (2009, p. 1) "a carefully chosen set of defaults means that most of the time you can produce a publication-quality graphic in seconds, but if you do have special formatting requirements, a comprehensive theming system makes it easy to do what you want." It was created by Hadley Wickham in 2005 and sets itself apart from most R graphics packages by being an implementation in R of *the*

[10] http://www.theguardian.com/world/datablog/interactive/2010/oct/23/wikileaks-iraq-deaths-map

[11] https://rjmetrics.com/resources/reports/the-state-of-data-science/

Grammar of Graphics, a general scheme for data visualization by statistician Leland Wilkinson. It allows users to easily create most common plots and, with a bit more effort, sophisticated visualizations.

ggplot2 is for R as **matplotlib** (first developed by John Hunt in 2002 as a patch to allow interactive MatLab-style plotting in *iPython* via gnuplot) is for Python, because it is the most popular option for visualizing data in Python. However it does not implement *the Grammar of Graphics* (there is *ggplot* for that, a plotting system for Python based on ggplot2 and *the Grammar of Graphics*). matplotlib is powerful enough to allow implementation of some of the best practices of *the Grammar of Graphics*, but according to users its plotting commands remain rather verbose and its outputs too simple. Similarly to ggplot2, matplotlib is both powerful and complex, but its flexibility has led to a large user base and active developer base. Nonetheless, more sophisticated alternatives are now gaining wide acceptance. This is the case of Bokeh, a Python interactive visualization library that intends to be the Python alternative to D3 by providing high-performance interactivity. When a user creates a visualization with Bokeh, this produces a JSON file as an input for a specific JavaScript library called *BokehJS*, which is able to present the data in modern web browsers. Bokeh can be a good alternative for data scientists that are already familiar with using Python for data analytics and have little experience with JavaScript. In the survey by RJMetrics (2015), Python is a close second, with 46.41 percent of the data scientists currently employed by companies known to *LinkedIn* listing it as a skill. Moreover it is the skill most listed by both senior and junior data scientists. However, R still has the highest percentage of chief data scientists, which seems to point to the fact that R is a more established software package.

Plotly is a web-based tool that combines both analytics and visualization, and allows users to work individually or collaboratively. It also allows users to choose their favorite tools, providing APIs for Python, R, MATLAB, Microsoft Excel, JavaScript, etc. For instance, Plotly's R API allows users to directly create the graphics through R and refine these with Plotly's online tool. Plotly is user-friendly and easily customizable, therefore it is used by some news outlets, such as *The Washington Post*[12] and Wired[13]. It allows users to create both static and interactive visualizations, and makes it easy to share the final output, which can be embedded in any website. Plotly is not a free tool, but has a "community version" available for free. However this version only allows users to create one private chart, import data from some sources (Microsoft Excel, CSV, and XML), export charts in PNG and JPEG, 1000 views per chart per day, limited API calls, and other minor features. To really take advantage of the full potential of the tool the user has to have access to one of the paid versions, which, among other features, allows users to create unlimited private and public charts, and export them in higher quality formats.

When the main objective is to create beautiful and interactivity rich visualizations, **Processing** and **Processing.js** can also be a choice and there are several examples populate the

[12] http://wapo.st/14djWTm
[13] http://www.wired.com/2015/05/math-starbucks-new-mini-sized-frappuccino/

web, such as *Ville Vivant*[14], *On the Origin of Species - The Preservation of Favoured Traces*[15], and *Visualizing The Guardian*[16]. However, neither Processing nor Processing.js are tools specifically built for visualizing data. Processing is a programming language and IDE built for visual arts. The project started in 2001 with the intention of being a tool to teach programming fundamentals by providing visual feedback. Nonetheless, the project developed by Casey Reas and Benjamin Fry evolved to become a tool for professionals. Processing is an open source project and runs on Linux, Mac OS, and Windows machines. Processing.js is a JavaScript version of the popular Processing visual programming language, which not only allows users to create visualizations but also any type of interactive content (including games). It was created in 2008 by John Resig and it uses the canvas element, available on modern web browsers, without Java applets. Maintaining the initial purpose of Processing, Processing.js is used to teach programming on *Khan Academy*[17].

10.2 Choosing the right visualization for a given function

It is not simple to provide a guide to build visualizations. Every dataset is unique, so to be able to visualize it the creator of the visualization has to have a deep understanding of what should be shown, emphasized, and what are the best practices. There is a set of principles and concepts that have to be understood produce a good visualization and reducing the selection of how to visualize a dataset to a simple set of guidelines can be dangerous.

In his book *Advanced Presentations by Design: Creating Communication that Drives Action*, Abela (2008) proposes a chart selection guide (seen in Figure 10.9), which he believes can help visualization creators select a good chart depending on what he/she wants the data to demonstrate. The diagram asks "what would you like to show?" and provides four options: Comparison, Relationship, Distribution, and Composition. Although these are all valid options related to data, what comprises them is not as clear and reducing what you can do with a data set to these four options in over simplistic.

For instance, for Abela (2008) a comparison seems to be the consideration or estimation of the similarities or dissimilarities between two or more categorical variables. However, because it is not clearly stated in the book we cannot be sure if he is talking about categorical data and in fact it is possible to draw comparisons with almost, if not all, types of charts. Nevertheless, there are charts that facilitate the comparison of categorical data more than others, but to understand this it is vital to know the difference between categorical data and quantitative data.

In opposition to quantitative data, which is data where the values can change continuously and the number of different values cannot be counted (for example the number of visitors of a museum during a year), categorical data is data which you can count a small number of different categories (for example the number of visitors of a few museums in a year, the

[14] https://villevivante.ch/
[15] http://fathom.info/traces
[16] https://www.flickr.com/photos/blprnt/sets/72157615061041180/
[17] https://www.khanacademy.org/

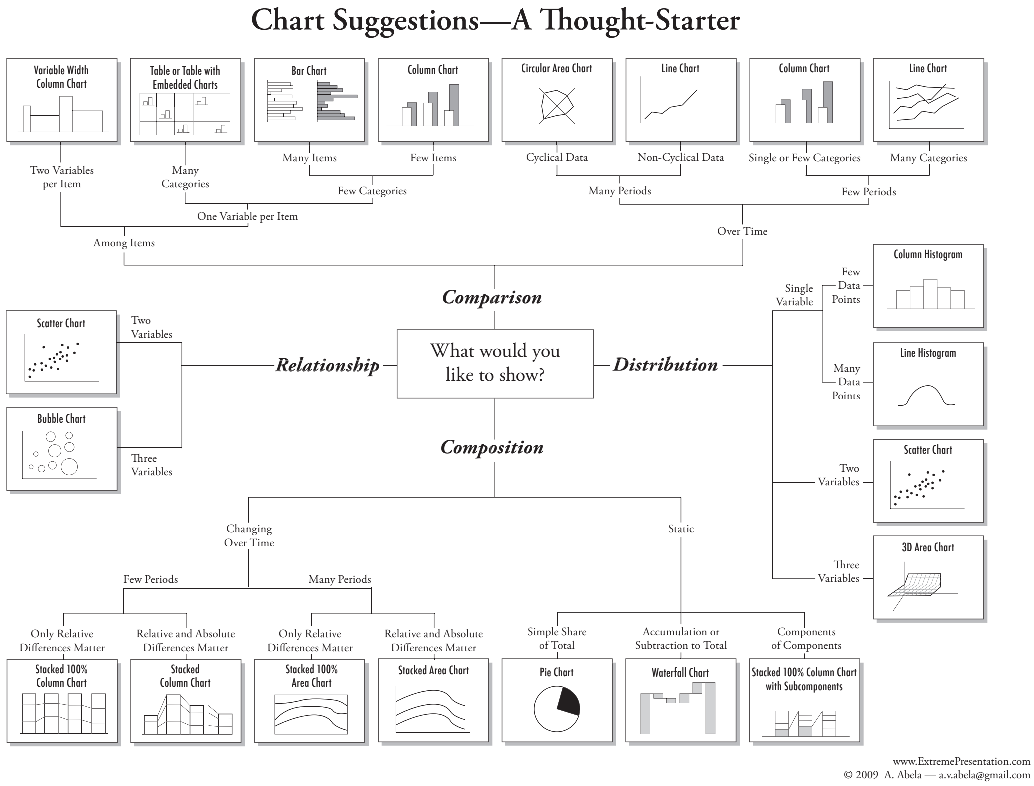


Figure 10.9: Chart selector guide (Figure 7.1 (Abela, 2008, p. 99))

different museums being different categories). Moreover,“ there is often no implicit relation among these classes (whereas it is for numerical variables in terms of ordering and distances among values)” (Bendix et al., 2005, p. 133).

Time, however can be both continuous and categorical data. If the goal is to represent the evolution of the number of visitors throughout the year a line graph would be appropriate because it would show if the visits are increasing or declining, or if there is a period where the number of visits changes dramatically. However, if the goal is to pin point the month with more visits, a bar chart would be more appropriate, and in this case the months would be the categories. There are three main reasons why traditional information visualization techniques, such as scatter plots and parallel coordinates, are better fitted for continuous data variables: “(1) there is a natural one-to-one mapping of data values to visualization parameters like positions and colors and (2) these continuous parameters better match the continuous characteristics of the screen (in the spatial, temporal, and chromatic dimensions)” (Bendix et al., 2005, p. 133).

Most data sets are not just composed of categorical data or of just quantitative data. However, if that is case, for purely quantitative data there is the parallel coordinates visualization and for purely categorical data the parallel sets visualization. Parallel coordinates is type of visualization for multivariate data. Usually it consists of a set of axis parallel to each other and equally spaced (typically vertical). Each line that connects the different axis is a data point that has a corresponding value for each axis. According to Bendix et al. (2005, p. 134), “this view is capable of displaying high-dimensional data, because the axes are visually independent from each other". Take for instance Mike Bostock's parallel coordinates example[18]. It represents a long list of car models from the 70s and 80s, and for each model its mileage, number of cylinders, displacement, horsepower, weight, the time (in seconds) it takes to go from 0 to 60mph, and year. If the data was mapped in a table the models would be the rows and the other variables would be the columns. In this visualization each value for each car model is mapped along each axis and the line is what connects all the values for each model. All the axis, except the number of cylinders, have several different possible values, which is what is standard for this type of visualization. The number of cylinders only varies from 3 to 8 (whole numbers) and in this dataset there are only 5 different values for the number of cylinders, therefore the lines only pass through a very small number of points. This is possible as long as there are not a lot of variables like these on the same visualization.

With parallel coordinates it is possible to identify patterns and correlations. However, it does not always work well with very large datasets and requires a bit of experience to be able to draw significant conclusions from this type of visualization.

On the other hand, parallel sets is one of the most useful ways to visualize multivariate categorical data. It was first introduced by Bendix et al. (2005, p. 133) in a paper entitled *Parallel sets: visual analysis of categorical data* where these are described as "a new visualization method that adopts the layout of parallel coordinates, but substitutes the individual data points by a frequency-based representation". This visualization looks like a cross-over between a sankey diagram and a parallel coordinates visualization, because it represents proportions and multiple variables mapped on more than one axis. Take for instance Jason Davies's D3 example of parallel sets[19] that represents the people that traveled in the Titanic and again trying to translate it to a table: the data could be translated to a multi-way table that included columns with subcategories (a main column survived, with two sub-columns survived and perished; a main column sex, with two sub-columns male and female; a main column age, with two sub-columns child and adult; and a main column class, with four sub-columns first class, second class, third class and crew;) (Bendix et al., 2005), but it would be easier to translate it to a table with four columns (survived, sex, age, and class) on which the lines represent each passenger. In this case the value in each cell of the survived column would be survived or perished, in the sex column would be male or female, in the age column would be child or adult, and in the class column would be first class, second class, third class or crew. Each column would be represented as one axis of which the total length would represent the total number of people that traveled in the Titanic and the *ribbons* would be subdivided according to the number of people that have the same value for each category. In other words, as it can be seen in Figure 10.10, (1) the blue ribbon starts with a certain width that represents the number of survivors, (2) then is subdivided in two having a determined width for the number of women that survived and a different width for the number of men that survived, (3) the part of the ribbon that represents women is again subdivided in two categories (child and adult) in the third axis and so is the part of the ribbon that represents men, (4) finally the part of the ribbon that represents female

[18] http://bl.ocks.org/mbostock/1341021

[19] https://www.jasondavies.com/parallel-sets/

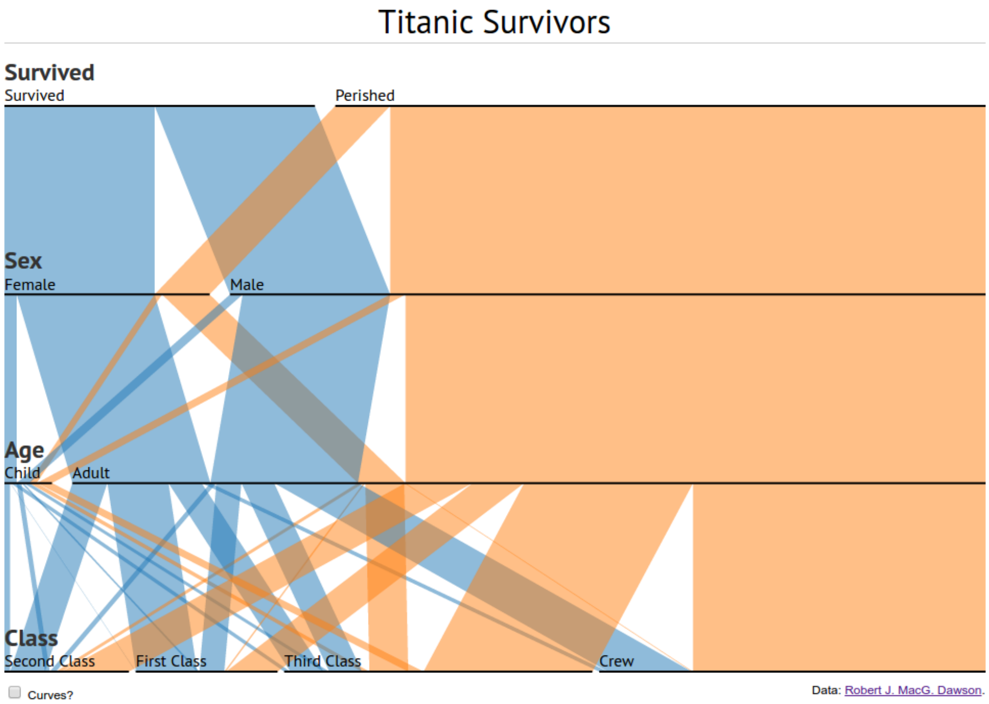


Figure 10.10: Jason Davies's D3 example of parallel sets

children is subdivided in the number of parts that represent the class in which they traveled, and the same happens with the parts that represent the adult females, the male children, and the adult males. The same happens with the orange ribbon that represents the people who perished when the Titanic sunk.

Kosara, Bendix, et al. (2006, p. 559) also developed a visualization application specifically for categorical data called *Parallel Sets*, which supports "interactive visual exploration and analysis" and this was later ported to D3 by software developer Jason Davies with similar sorting and rearranging interactions.

The problems with the comparison option in Abela (2008) chart selection guide do not end with the clarification of what is categorical and quantitative data. The second layer of options once we pick comparison asks to choose between comparison *among items* and *over time*. It is clear that comparison over time would mean that the creator of the chart has the intention of comparing the evolution of a variable over time. However, seeing the evolution of something over time is not an activity exclusive to comparison: when creating a chart, a person can have the intention of seeing the relationship between two or more sets of values over time to understand correlations (for which he/she can use a scatter plot), he/she can have the intention of seeing the distribution — for instance how the students in a class are distributed by grades, from lowest to highest grade — over time using a line chart or a histogram, etc. Moreover, it is not clear what is the difference between *over time* and *changing over time* which later appears in the composition option.

If in the second step of comparison, in the chart selection guide, the user chooses comparison among items, the following options are available: two variables per item (which gives online one chart type option) and one variable per item (which leads to a choice between many categories or few categories). Here the main problem is the fact that apparently tables are only useful when there are many categories, showing that Abela (2008) forgets the value of the use of tables when precise values are needed or when its vital to search for a specific data point.

The options distribution and relationship seem to be the clearest options in the chart selection guide by Abela (2008). If we assume that relationships are correlations the only two options given by the chart (scatter plot for a correlation between two quantitative variables and bubble chart for a correlation between three quantitative variables) are valid options. In this case correlation is the degree to which the values of two or more variables are linearly associated. For instance, if we observe that more hours of study leads to better grades it means that there is a correlation between these two variables. However, if the definition of relationship used by Abela (2008) means a mutual relationship or connection between two or more things, then there are several other types of visualizations that could be used.

Although it is never clearly defined in the book, distribution seems to refer to the display of the frequency of occurrence of various outcomes in a sample: for instance how the students in a class are distributed by grades, from lowest to highest grade. However, it only provides options for displaying the distribution of quantitative values: of a single variable with few or many data points, two variables or three variables. For observing the distribution of categorical data parallel sets would be the best option. Moreover, scatter plots and 3D area charts are not the most appropriate types of charts to display distributions.

Finally, the chart gives the option of composition, that seems to refer to displays that show the relationship between the parts and the whole. The first decision that the chart asks the user to make is to decide if the data is changing over time or static, which seems inappropriate when most of the charts displayed seem to work both for static and for data that changes over time. For example, a pie chart can be used for presenting the percentages of sales in the different months, each slice representing a month, and a stacked bar chart can be used to display categorical data in each of the bars. As it happens for some of the other options there is Moreover, all of the chart options here can be difficult to interpret: stacked charts are hard to understand if the differences between the sections of the bar are small), and the same happens with pie charts, which are only useful when comparing 2 or 3 variables with extremely different values. Pie charts specially are both one of the most popular kinds of charts and one of the most hated. While they are familiar to most users because as children we all learn fractions by looking at sliced pies (quarters, halves, thirds, etc.), but pie charts only facilitate the understanding of percentages close to 0%, 25%, 50%, 75%, and 100%. Recommending the use of such controversial charts every time a user wants to display relations of part-to-whole is not wise, because it will perpetuate the mindless use of these.

10.2.1 A visualization selection matrix

Instead of a guide that limits what the visualization creator has to do, it is better to give visualization creators quick reminders of what can and cannot be done with the data, which

	Points	Colors	Shapes	Lines	Bars
Geographical	bubble map	cloropleth heat map	cartogram	sankey diagram	x
Temporal	scatter plot bubble chart	x	object size area chart stacked area chart streamgraph box plot pie chart doughnut chart	sankey diagram line chart sankey arc	bar chart histogram circular bar chart grouped bar chart stacked bar chart span chart
Relationships	x	heat map matrix	x	sankey arc chord diagram weighted network parallel sets parallel coordinates	x
Part-to-Whole	x	heat map matrix	stacked area chart streamgraph tag cloud treemap radar chart pie chart doughnut chart polar-area chart	sankey diagram line chart weighted network parallel sets	bar chart histogram circular bar chart grouped bar chart stacked bar chart span chart population pyramid
Item comparison	scatter plot bubble chart dot plot	heat map matrix	object size tag cloud treemap pie chart polar-area chart	sankey arc chord diagram weighted network parallel sets parallel coordinates	bar chart histogram circular bar chart grouped bar chart stacked bar chart span chart population pyramid
Distribution	bubble chart dot plot	heat map matrix	object size area chart stacked area chart streamgraph tag cloud radar chart box plot	sankey diagram line chart sankey arc chord diagram weighted network parallel sets	bar chart histogram circular bar chart grouped bar chart stacked bar chart span chart population pyramid

Table 10.1: Visualization selection suggestion matrix (at least one quantitative variable)

does not eliminate the need to understand what the visualization methods comprise. Therefore I propose a matrix, more or less inspired by the *Graph selection matrix* provided by Few (2012) in his book *Show Me the Numbers: Designing Tables and Graphs to Enlighten*. I extend this matrix by including visualization methods commonly used in visualization nowadays and reinterpret the relationships identified by Few (2012). For each pair of dimensions in the matrix there are suggestions of visualizations types that can represent those dimensions.

The columns of the matrix, shown in Table 10.1, present visual elements that can be used to represent quantitative variables: points, colors, shapes, lines, and bars. A visualization type can have many different types of visual elements, however there are elements that stand out more. Therefore, in this matrix the visualization types are not repeated in the different columns, because the more visually striking aspect of each visualization was picked and that was what determined its column. For instance, a bubble chart uses both points and colors, but the colors are not the most prominent visual aspect.

The lines in the matrix correspond to aspects of the data that can be highlighted: geographical, temporal, relationships, part-to-whole, item comparison (categorical data), and distribution. The line for geographical only provides options for the visualization creator to highlight data that deals with the description, distribution, and interaction of geographic locations, including how they are affected by human presence (population distribution, political divisions, etc.). The line for temporal provides options for highlighting events that succeed one another in a period of time. Relationships includes options for both the intention of visualizing the traditional notion of correlation, defined in the previous section, and any intention of highlighting relationships between two or more data points. Similarly to the option of composition by Abela (2008), part-to-whole provides visualizations for highlighting the role that the parts pay in the whole. The line for item comparison provides visualization options for datasets that have categorical data but still include quantitative data, not being purely categorical. Finally, distribution provides options for visualizing the display of the frequency of occurrence of various outcomes in a sample.

Contrary to what happens in the columns, there are several aspects that can be highlighted in a visualization and is up to the visualization creator to choose which aspect he/she thinks is more interesting to highlight. The only exception here is the visualization types that are characterized as geographical, because the geographical aspect shapes the visualization in such a way that, when a visualization of this type is used, the geographical aspects are what will be more highlighted. For instance, a geographical sankey such as Minard's *Carte figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813* (seen in detail in Subsection 2.1.2) highlights both the geographical and the distribution aspects, however the geographical aspect is so striking that we do not refrain to call it a map.

Although the list of types of visualization intends to be as complete as possible it is obviously not complete, because there are slight variations for most visualization types. However the types listed in the matrix are some of the most popular and available in most visualization tools cited in Section 10.1.

There is also the possibility of combining several types of visualization to create a more elaborate final visualization. This combination of several kinds of visualization techniques can be a type of visualization in itself such as the *small multiples display*. Small multiples (often also called lattice chart or grid chart) were popularized by Tufte (1990), according to who these can be the best design solution to a wide range of problems. This is because comparison is at the heart of quantitative reasoning and "small multiple designs, multivariate and data bountiful, answer directly by visually enforcing comparisons of changes, of the differences among objects, of the scope of alternatives" (Tufte, 1990, p. 67).

I also provide a matrix for the representation of exclusively categorical data such as visualization types for hierarchies (network diagrams, sunburst charts, etc.) or for interactions (models), which can be seen in Table 10.2. The columns of this matrix present the visual elements and the lines correspond to the aspects of the data that can be highlighted.

Most of the visual elements present in the visualization suggestion matrix for quantitative data (shown in Table 10.1) are also present in the visualization suggestion matrix for

	Points	Lines	Colors	Bars	Shapes	Grids
Geographical	dot map	transit map	grouped political map	x	x	x
Temporal	x	timeline	x	x	x	calendar timetable
Relationships	x	network diagram tree diagram non-ribbon chord diagram radial tree cycle graph arc diagram mind map concept map decision tree	x	x	venn diagram	matrix
Part-to-Whole	x	x	x	x	pyramid model	calendar timetable
Grouping	x	network diagram mind map non-ribbon chord diagram	polar grid color matrix	sunburst chart	venn diagram	scenario matrix
Hierarchy	x	tree diagram radial tree mind map concept map	x	sunburst chart	x	x
Interactions	x	x	x	x	model	x

Table 10.2: Visualization selection suggestion matrix (only categorical data)

categorical data: points, lines, colors, and bars. Shapes was left out because, for exclusively categorical data, no types of visualization with this visual element was found. Although the visual element grid can also be seen in visualization types included in the visualization suggestion matrix for quantitative data, namely in the heat map matrix matrix, it was not the most visually striking element in any of the types, therefore it was excluded in that matrix. However, the visual element grid was added in the visualization suggestion matrix for categorical data because it was the most visually prominent element in some of the visualizations, namely in the calendar, timetable, matrix, and scenario matrix.

In the aspects that can be highlighted there are also some differences. In addition to geographical, temporal, relationships, and part-to-whole, there is grouping, hierarchy, and interactions. Grouping includes options for visualizing groups of items by similarity, location, classification, etc. The line for location provides visualization examples for displaying ranked or graded series. Finally, the line for iterations provides a visualization type that allows users to see how objects are affected by reciprocal action or influence: model. A model also can highlight the role of the parts in the whole if presented in an exploded view. No visualization types that mainly provide item comparison or distribution were found.

10.3 Chapter Summary

In this chapter I approach techniques for visualization on the web and present some options for tools and visual representations. This was done because more research on the whole process of creating a narrative visualization that can serve as guidelines to build new visualizations and that facilitates this process is needed. The guidelines described in this chapter can inform the work of both information visualization researchers and visualization creators.

From all the available **tools for information visualization** (section 10.1) the ones that I highlight are Tableau, D3.js, and Ellipsis. The first two are probably the most popular tools available but are targeted to creators with different levels of expertise. Tableau has more than 35,000 customer accounts, which range from companies to private users. It is very easy to use, thus targeted for business intelligence for which time is a major concern. Tableau accounts for the analysis and visualization of the data, supporting a wide variety of charts, graphs, maps, etc. Tableau's major limitation is the fact that some features are only available in paid versions.

On the other hand, D3 has a steeper learning curve, but also allows users to be more creative in the way they visualize their data. D3 is a sophisticated and powerful tool that allows the user to choose any style, manipulate it freely and add as much interactivity as he/she wishes. It not only allows flexibility in terms of creation but also in terms of distribution. Due to its power and flexibility D3 has been used by The New York Times to build interesting visualizations such as *Is It Better to Rent or Buy?*[20], *512 Paths to the White House*[21], and *The Most Detailed Maps You'll See From the Midterm Elections*[22].

I also highlight a tool with a much smaller user base, just because it is one of the few available tools that is concerned with supporting storytelling techniques: Ellipsis. According to its creators (Satyanarayan and J. Heer, 2014), Ellipsis is a system that combines a domain-specific language (DSL) for storytelling with a graphical interface that allows story authoring without programming. Examples of visualizations built with Ellipsis include *Budget Forecasts, Compared With Reality*[23] and *Map: Catholics, cardinals by country*[24]

I also briefly review other popular tools such as Many Eyes, Google Charts and Fusion Tables, and Gephi, and some tools that are not specific for visualization but that can also be used to create them such as R, ggplot2, matplotlib, Processing and Processing.js, and Plotly.

In this chapter I also provide guidelines on how to **choose the right visualization for a given function** (section 10.2). I first explain how having a chart selector guide like the one provided by Abela (2008) in his book *Advanced Presentations by Design: Creating Communication that Drives Action* is not ideal for choosing the right visualization for a given function.

Then I present two visualization selection suggestion matrices (for datasets with at least one quantitative variable and for datasets with only categorical data), which can inform visual representation selection without eliminating the need to understand what the visualization methods comprise. These matrices were inspired by the *Graph selection matrix* created by Few (2012) for his book *Show Me the Numbers: Designing Tables and Graphs to Enlighten.*

[20] http://rethinkingvis.com/visualizations/190

[21] http://www.nytimes.com/interactive/2012/11/02/us/politics/paths-to-the-white-house.html

[22] http://www.nytimes.com/interactive/2014/11/04/upshot/senate-maps.html

[23] http://www.nytimes.com/interactive/2010/02/02/us/politics/20100201-budget-porcupine-graphic.html

[24] http://www.cbc.ca/news2/interactives/catholics-cardinals/

Chapter 11

Conclusions

In the same way that a well-told story is able to convey a large amount of information in a more compelling way — enabling the audience to assimilate and retain the information transmitted (Gershon and Page, 2001) — a well structured visualization can also be more compelling for the audience. Technology provides us new tools to convey information in a story-like fashion (Gershon and Page, 2001) and that is clearly transforming our preferences. People get excited with good visualizations and the fact is that people are sharing these visualizations online, even if they do not read the associated articles, might be a lead that the visualization alone is central.

Therefore we have to be able to understand the audience's preferences. We have to be able to introduce storytelling in these visualizations and to tailor visualization systems to accommodate storytelling, because we know the audience might not go beyond the information provided by the visualization.

Although there has been an effort to introduce new evaluation techniques in the field, current research in the area of visualization is still based on task completion, on the time it takes to complete a task (Kosara and Mackinlay, 2013). However, if we want to produce engaging visualizations, that people want to spend time on, that metric might be insufficient. Research in narrative visualization is still very new and there is still little information on how to successfully introduce storytelling in visualizations and even less research on what techniques work for different types of audiences. In 2015, Rodríguez et al. (2015) reviewed the historical background of the research on the intersection of visualization and storytelling. According to Rodríguez et al. (2015, p. 10) "in order to categorize visualization strategies as successful or unsuccessful in its intention to convey a message, an empirical evaluation should be conducted to determine if the user was *invested* in the data of the story, or *distracted* by interactive features or visual constructs that will consequently hinder the core message." Rigorous study and measurement of the impact of visualizations would benefit several areas that use or are beginning to use visualization to convey information in a more appealing way. The research area of narrative visualization has not yet consolidated a solid set of research methods because it is very difficult to evaluate visualizations, therefore it is very hard to be certain in relation to what the users really like and what in fact improves the process of getting new valuable insights.

11.1 Thesis contributions

Narrative visualization is of great relevance and the questions posed here are important not only for the information visualization community but also to every area that wishes to elevate visualizations to a more elaborate form of dissemination of information/data. More empirical study is needed for the field to move forward. The domain of storytelling in visualization is just only starting to take shape and there are still much ongoing discussions. There are ample opportunities to make an impact and this thesis tries to contribute to the ongoing research on the intersection of storytelling and visualization.

Nonetheless, research in storytelling for information visualization has increased. The interest in this subfield of information visualization has sparked not only in areas with a previous interest in storytelling, such as journalism, but also in areas that previously were only interested in using scientific visualization, which in the visualization spectrum proposed by Kosara (2007) and presented in Subsection 2.2.4 is much more utilitarian and closer to what he calls pragmatic visualization. An example of an area that was previously more interested in scientific visualization and is now showing interest in storytelling is biology and biomedical research. This is the case of the research carried by Gratzl et al. (2016, p. 2), who developed CLUE, "a model for reproducing, annotating, and presenting visualization-driven data exploration based on automatically captured provenance data." Gratzl et al. (2016) were inspired by previous research on this topic, namely the narrative strategies explored in this research (Figueiras, 2014a; Figueiras, 2014b), to develop a model targeted for biomedical data. According to Gratzl et al. (2016), the collection and representation of data provenance is still one of the major challenges of data-driven biomedical research, and visual data stories can be an important tool for

Paper title	Number of citations	Citation
What storytelling can do for information visualization	253	Gershon and Page (2001)
Storytelling: its role in information visualization	22	W. Wojtkowski and G. Wojtkowski (2002)
Story telling for presentation in volume visualization	48	Wohlfart and Hauser (2007)
Graphical histories for visualization: Supporting analysis, communication, and evaluation	196	J. Heer, Mackinlay, et al. (2008)
Narratives: A visualization to track narrative events as they develop	51	Fisher et al. (2008)
Narrative visualization: Telling stories with data	335	Segel and J. Heer (2010)
Visualization rhetoric: Framing effects in narrative visualization	99	Hullman and Diakopoulos (2011)
Scientific storytelling using visualization	49	Ma et al. (2012)
A deeper understanding of sequence in narrative visualization	26	Hullman, Drucker, et al. (2013)
StoryFlow: Tracking the Evolution of Stories	56	Liu et al. (2013)
Storytelling: The next step for visualization	89	Kosara and Mackinlay (2013)
Storytelling in visual analytics tools for business intelligence	9	Elias et al. (2013)
How to tell stories using visualization	6	Figueiras (2014a)
Narrative visualization: A case study of how to incorporate narrative elements in existing visualizations	2	Figueiras (2014b)
Data Comics: Sequential Art for Data-Driven Storytelling	2	Z. Zhao et al. (2015)
More than Telling a Story: A Closer Look at the Process of Transforming Data into Visually Shared Stories	6	Lee et al. (2015)
Storytelling in Information Visualizations: Does it Engage Users to Explore Data?	12	Boy, Detienne, et al. (2015)
From Visual Exploration to Storytelling and Back Again	3	Gratzl et al. (2016)

Table 11.1: Published papers on the topic of information visualization and storytelling

communicating multi-step analysis in order to facilitate the reproducibility of findings. The CLUE library integrates into an open source scientific software named Caleydo Web, a visual analysis framework for biomolecular data.

Table 11.1 presents some of the papers that have been published specifically on information visualization with storytelling. These are scientific publications that either reference the research area of narrative visualization in its title (Figueiras, 2014b; Hullman and Diakopoulos, 2011; Hullman, Drucker, et al., 2013; Segel and J. Heer, 2010) or keywords (Boy, Detienne, et al., 2015; Figueiras, 2014a; Lee et al., 2015; Z. Zhao et al., 2015), or clearly references in its title, abstract, or keywords that it explores the use of storytelling or narrative in visualization or visual analytics (Elias et al., 2013; Gershon and Page, 2001; Gratzl et al., 2016; J. Heer, Mackinlay, et al., 2008; Kosara and Mackinlay, 2013; Liu et al., 2013; Ma et al., 2012; Wohlfart and Hauser, 2007; W. Wojtkowski and G. Wojtkowski, 2002). The table also gives a hint of the publication's impact by presenting the number of citations (at time of publishing), which was collected through Google Scholar. The publications were chosen by searching for the following keywords on Google Scholar: narrative; visualization; storytelling.

An analysis of the publications presented in the table reveals how this topic has garnered more interest from the information visualization community. Specially since 2010, not only did the number of published papers about storytelling for information visualization increase on top information visualization conferences and journals (and other related areas such as HCI), but also the number of citations of these papers has also increased.

11.1.1 Research objectives

My work focused on substantiating our understanding of the role that storytelling can play in visualizations and in making these more accessible to everyone, from expert users to novices, and on providing a better understanding of how to best design visualizations that incorporated strong narrative qualities.

Moreover I wanted to understand which techniques add a story feel to the visualizations without affecting the free exploration of the data. Narrative visualization should not become a lean back format, accordingly the quest to add storytelling has to be weighted so that it does not lead to a linear, author-driven interpretation path for the user. Stories in visualization should be used as starting points for data exploration or moments of insight about the data, rather than a predigested narrative.

Providing free access to the data seems to help address the need to expose the intricacy of the information, however this might be confusing specially for non-proficient users. Narrative elements can possibly help frame the inner contradictions of the data and lead the users to their own interpretations of the information, guiding their attention in subtle ways.

I worked towards these objectives by trying to address three essential research questions:

- RQ1: Which techniques can be used to tell stories using visualization?
- RQ2: Which elements can be used to have an effective storytelling?

- RQ3: Which types of visualizations might appeal to the public?

In the following subsections, I provide summaries on how I addressed each research challenge.

The role of interactivity in narrative visualization

The first major contribution presented in this dissertation (chapter 4) is the new interaction techniques taxonomy. As pointed out in previous work (Hullman, Diakopoulos, and Adar, 2013), interactivity is a crucial part on the introduction of storytelling in visualization. Therefore, I began with the goal of building a comprehensive list of interaction techniques, in order to more systematically explore the purposes of interactivity in information visualization in general but having narrative visualization in mind. Backed up by the existing literature, I evaluated 232 visualizations and studied their interaction techniques.

With the new interaction techniques taxonomy and the exploration of the importance of the use of interactivity in visualization, I worked towards an answer to my first research question (RQ1): Which techniques can be used to tell stories using visualization? By systematically studying the use of interaction techniques in popular visualizations it is possible to see the role that these play in narrative strategies such as the introduction of context. In section 9.1 — Interactivity and its relation with context — I explore the relation between interactivity and context as a possible strategy to introduce storytelling in visualizations. Interactivity opened up the possibility of adding new layers of content and thanks to this additional layer of content, most of the times in the form of annotations, visualizations can, in addition to the data itself, provide content that is able to add context.

Additionally, the opinions collected in the focus group study, shown in chapter 7, seem to hint that the audience prefers short moments of storytelling that they can access if they feel the urge, rather than having a dense storytelling that they have to carefully follow. The focus group participants believed that the visualizations they liked the least could probably be fixed with an overlay of context, presented through the use of interaction techniques, such as hover, click details, or even hyperlinks to other content on the same page or external links, for example to Wikipedia pages. According to them the context would be necessary to give meaning to the data.

The context is commonly given by using interaction techniques such as abstract/elaborate, and overview and explore. However, this context can also be given through participation/collaboration interaction techniques. This is the case of the visualization *Home and Away: Iraq and Afghanistan War Casualties*, explored in detail in section 6.4. The visualization is integrated with the CNN *iReport* platform and allows family and friends of the deceased soldiers represented in the visualization to write personal stories and share memories on his/her profile. The context given by these annotations help building the story of each soldier's life and death, and provide the user information that journalists would probably not get access to.

Narrative strategies for visualization

One of the main contributions in this thesis is the narrative strategies proposed, shown in chapter 9. The possible strategies for the introduction of storytelling in visualizations presented were 1) interactivity and its relation with context, 2) empathy, 3) the relation between time and narrative, and 4) gamification — and address the second research question (RQ2): Which elements can be used to have an effective storytelling?

Based on the literature review and the studies conducted some key elements of storytelling were identified. *Context* is for sure the main one and is explored throughout this thesis. Everything seems to point out towards the fact that visualizations that do not provide context rarely succeed in providing the feeling of storytelling. Context is used to fill in the gaps between the data points and allow the plot of the visualization to unfold just like traditional storytelling would do. As books use descriptions to make the readers imagine the full picture, annotations do the same for visualizations.

Another element that can be used to have effective storytelling is storytelling metaphors such as *visual representations of time*. One of the characteristics that focus group participants appreciated was when the visualizations provided some temporal structure, therefore, one of the participants' top five visualizations was *Evolution of the Web*, a timeline visualization by *Google*. One of the participants said that this visualization was almost perfect, because timelines give this feeling of flow. This feeling of flow described by the participants is due to the fact that the structure of a timeline resembles the structure of a traditional story: it has a beginning, a middle, and an end. A key research problem now is to discover new visual metaphors for time, other than timelines, and to understand which analytical tasks these could support.

Although gamification could not be tested in the focus group study due to time constraints, it presents itself as a possible element that can be used to have an effective storytelling. Specially when used to simulate actions that have specific effects, gamification can help to introduce the feeling of a traditional narrative. This can easily be seen in examples such as *SPENT* (analyzed in detail in section 6.11, *Budget Hero*, and *Gauging Your Distraction*. In these visualizations gamification is used to show the impact that player's actions have on the final outcome by simulating the situation: *SPENT* shows that it is very difficult to live in the US earning $1,000 a month and that people have to make difficult choices to get by on such a tight income (whether or not to pay for the car insurance, whether or not to take your child to the dentist, etc.); *Budget Hero* shows that it is difficult to choose where tax dollars go and that it is hard to keep a balance between policy choices and the financial stability of the country; and *Gauging Your Distraction* shows how difficult it is to text and drive at the same time.

A visualization does not need to have all these strategies to have the feeling of storytelling. However, if a visualization incorporates one or more of these elements it will probably enhance the feeling of storytelling.

This research question was also addressed by conducting a literature analysis (also included in chapter 9) of work on narrative visualization based on the theory of visual rhetoric (Hullman and Diakopoulos, 2011). Similarly to what was done for this thesis, Hullman and Diakopoulos (2011) identified, via a systematic analysis of fifty-one professionally-produced narrative visualizations, classes of rhetorical techniques used in InfoVis. These were later characterized according to their rhetorical contribution to the visualization: information access rhetoric (which includes omission and metonymy techniques), provenance rhetoric (which includes data provenance strategies, techniques for representing uncertainty, and author identification techniques), mapping rhetoric (which includes obscuring techniques, visual metaphors and metonymy, contrast techniques, classification, and redundancy techniques), linguistic-based rhetoric (which includes typographic emphases, irony, similarity techniques, and individualization techniques), and procedural rhetoric (which includes anchoring and filtering techniques). These rhetorical techniques can be applied (ones more easily than others) to four different editorial layers: data, visual representation, annotations, and interactivity. As I did, Hullman and Diakopoulos (2011) noticed the utilization of visual metaphors at the level of the visual representation, annotations, and interactivity for promoting the feeling of storytelling.

Although there was previous research on the importance of context in narrative visualization (Hullman and Diakopoulos, 2011; Hullman, Diakopoulos, and Adar, 2013; Satyanarayan and J. Heer, 2014; Stasko and Zhang, 2000; Waldner et al., 2014), on *time* as a storytelling affordance (a feature of a visualization that provides a narrative structure) (Kosara and Mackinlay, 2013), and on gamification as a storytelling strategy for visualization (Diakopoulos, 2010; Diakopoulos et al., 2011; Kosara and Mackinlay, 2013), this analysis was the first general summarization of a set of strategies for introducing storytelling in visualization.

Focus group study

The third and final research question (RQ3) guiding this research was which types of visualizations might appeal to the public? I have addressed this research question by conducting a focus group study (chapter 7), with the intention to gather information about factors such as comprehension, likability, and navigation, in a way that allowed me to obtain an emotional response from the participants. The exploratory focus group study had the purpose of collecting information on the narrative elements in a collection of visualizations, and of approaching the participants about the possible inclusion of storytelling elements in those.

The results showed that participants: valued the existence of links that provide some context; enjoy when the visualization enables the feeling of empathy between the users and the data; generally prefer playable visualizations and feel that some of the visualizations they liked the least could be fixed by adding interactivity; wanted more support stories that could give some context; and that metaphors that give the feeling of story flow such as timelines are appreciated. These are techniques that have long been intuitively understood by others (visualization designers, data journalism teams, etc.), but it is still useful to observe how the audience relates to these visualizations, and what their suggestions for improvement are.

This final research question could better be answered by conducting further evaluation.

However, evaluation in information visualization is still an issue. In 2004, Komlodi et al. (2004) did a survey and identified four thematic areas of evaluation in information visualization:

- comparison of design elements resorting to controlled experiments;
- tool usability evaluation;
- comparison of two or more tools resorting to controlled experiments;
- case studies of tools in realistic settings

Later that year, Plaisant (2004) identified the challenges of doing evaluation in information visualization and pointed the three main ones: matching tools with users, tasks and real problems; improving user testing; and addressing universal usability.

Twelve years later, not much has changed and doing evaluation in visualization is still considered to be difficult "even to the point where it is seen as a black art consisting of equal parts prior experience and trial-and-error" (Elmqvist and Yi, 2015, p. 250). According to Elmqvist and Yi (2015), this happens because evaluating visualizations involves evaluating high-level cognitive tasks which are difficult to isolate, characterize, and measure, resonating in a lower incidence of evaluation in information visualization papers. Therefore, until 2008, from the over 800 papers that Lam et al. (2011) surveyed in the four major visualization venues (EuroVis, InfoVis, Vast, and IVS journal), they could only find 345 papers that included any kind of evaluation. In comparison, according to Elmqvist and Yi (2015, p. 250), "even a cursory read of the proceedings of leading HCI conferences, such as the ACM CHI conference, will show that the vast majority of HCI papers do include at least some form of evaluation."

Therefore, as future work (section 11.2), I intend to work towards finding better evaluation methods specifically for information visualization, in order to reach more meaningful reasoning regarding user engagement and the process of acquiring insights.

11.1.2 Summary of major contributions

In the course of this dissertation, I presented a **new interaction techniques taxonomy** (section 4.3), backed up by the existing literature. The necessity to build a new taxonomy came from the fact that some recent visualizations can hardly be analyzed using the existent taxonomies, because these do not include newer interaction techniques that are now being introduced such as participation or gamification. Although, these are more high-level forms of interaction (in opposition to low-level interactions such as clicking or hovering) these are forms of interaction nonetheless and that is why they are represented in this new interaction taxonomy. The new interaction taxonomy was built having in mind two previous taxonomies that only concern interaction techniques for information visualization (Keim, 2002; Yi et al., 2007) and a more general approach (Shneiderman, 1996).

I analyzed 232 visualizations and studied their types of interaction. From this study emerged eleven categories of interaction: filtering, selecting, abstract/elaborate, overview and explore, connect/relate, history, extraction of features, reconfigure, encode, participation/collaboration, and gamification.

I also presented a **new typology for categorizing visualizations** (chapter 5) and provided several case studies to exemplify every type of visualization. Similarly to what happened with the interaction taxonomy, there were already existing typologies, however these were also not exhaustive enough to classify all the examples studied.

The new categories were chosen according to which elements were more prominent and which influenced more the interaction, and were inspired by the existing classification schemes (Bogost et al., 2010; Nichani and Rajamanickam, 2003; Segel and J. Heer, 2010). This new typology intends to be an updated classification scheme, broad enough to classify all the most recent visualization examples. The new eleven different types or genres of visualizations are: Sequential Graphic; Slide Show; Chart/Diagram; Map; Tag Cloud; Model; Drawing; *Video/Animation*; Photograph; Poster; and Game.

The final major contribution is the possible **strategies for the introduction of storytelling in visualizations** (chapter 9): 1) interactivity and its relation with context, 2) empathy, 3) the relation between time and narrative, and 4) gamification.

11.1.3 Summary of minor contributions

During this dissertation, I also created the **ReThinking Visualization website** (section 5.3), a project developed with the intent of helping to build a better understanding of visualization by dissecting the pieces that compose a visualization and detect patterns, which can guide future research on information and narrative visualization. I also highlighted several **techniques, tools, and guidelines to build narrative visualizations** (chapter 10) through the use of a visualization selection matrix.

11.2 Future work

The contributions presented in this thesis have made considerable progress towards answering the proposed research questions. I believe that I have provided a methodology to introduce storytelling in information visualization, information on techniques and tools that can be used, and insights on what can potentially captivate the audience. The work presented also raises a number of new research questions that can lead to future work.

The findings presented in this thesis and in the published papers (Figueiras, 2013; Figueiras, 2014a; Figueiras, 2014b; Figueiras, 2015) are still not enough to fully understand how to use storytelling in visualization. In this dissertation, I have however identified the requirements and main challenges in narrative visualization, which can be a starting point for future work.

To make further progress on narrative visualization we need to:

- **have clear knowledge about what works and what does not work, which can only be revealed by further evaluation of visualization examples:** evaluation in narrative visualization shares some of the same problems identified in the evaluation of aesthetics (seen in Subsection 2.2.4) because it is a sub type of visualization at the edges of the Information Visualization field;

- **know the narrative visualizations that are being produced in news media, advertising, research, education, etc. that are achieving the desired effect on users and how/why:** evaluating existing examples is not enough to isolate the several techniques used because the users interpret the visualization as a whole and do not always understand what impacted most their exploration of the visualization. Therefore, new evaluation methods need to be developed to understand the benefits and weaknesses of each technique;

- **understand how and where narrative elements should be placed:** most visualizations are created using the method of trial and error which is a fundamental method of solving problems, characterized by repeated iterations until success is achieved. However, many visualizations are also created on a deadline and there is not enough time to iterate and polish things, therefore a clear guideline of when, where and how to introduce storytelling is paramount;

- **learn how the story should be structured** and be able to tell the difference between a story-like visualization and a visualization that uses narrative elements as explanations for the data;

- **know what is the impact of these stories on the users:** here again evaluation seems to be a big issue.

We are at an inflection point where the understanding of design dimensions is enough to start working towards the construction of models for narrative visualization. Once these models are built they should be employed in the design of visualizations, tested, and maybe then we will achieve some answers to the questions that persist in narrative visualization. A systematic study of these narrative visualization is the most surefire way to further amplify our understanding on this subject.

11.2.1 Work on narrative visualization evaluation

This thesis raised concerns on evaluation in information visualization, more precisely in narrative visualization. While I believe that I was able to confirm the fundamental hypothesis that carried this research — it is possible to tell stories using visualization — there is still much work to be done regarding the three general framing questions — What are the best techniques to tell stories in a more visual way? What elements have to be present in order to have an effective storytelling? What types of visualizations appeal to the public? — and this is due to the fact that the field still needs to find better evaluation techniques to be able to reach more meaningful conclusions regarding user engagement and the process of acquiring insights.

As future work, I would like to continue the search for definitive answers to these questions, mainly by trying to establish solid evaluation methods for the study of user engagement and the process of acquiring insights in information visualization. I believe that one of the first steps that has to be done is to search for more evidence on the efficacy of the use of interactivity. Most of the narrative strategies that are presented in this thesis rely on interactivity, however, although it seems that interactivity is important to user engagement and prevention of information overload, there is still little empirical evidence about its efficacy in terms of improving understanding of the data. There is also few research that points out guidelines of how to incorporate it successfully and that confirms that playable visualizations are indeed more enjoyable and popular among users.

The use of interactivity and animation has been discussed extensively in information visualization research, but there has been controversy regarding its benefits with several opinions in favor and against. Most of the published research on this subject evaluates one interactive versus one static visualization — or in the case of Salubrious Nation, a game-y version against a non-game-y version of the same visualization (Diakopoulos et al., 2011). However, by comparing only one type of visualization we cannot be certain that those findings (pro or against interactivity) would be the same if the type of visualization was changed.

My hypothesis is that we not only need to test and compare the users' reaction to a static version, an interactive version, and possibly the raw data, but also test the several types identified in the suggested typology (seen in Chapter 5). The ideal scenario would be to use the same datasets and representing it resorting to different visual representations. My goal is to conduct a task based evaluation where each participant is presented with one version and is asked to complete some tasks specifically conceived for each version. The objective is to follow the Single Usability Metric (SUM) first proposed by Sauro and Kindlund (2005a) in the paper *A Method to Standardize Usability Metrics into a Single Score* and presented in Figure 11.1.

SUM follows the steps previously given by others that have tried to create evaluation methods that use users' subjective assessments of recently completed tasks, such as SUMI (Kirakowski and Corbett, 1993), PSSUQ (J. R. Lewis, 1992), QUIS (Chin et al., 1988) and SUS (Brooke, 1996). According to Sauro and Kindlund (2005a, p. 401), "while the authors of these questionnaires do not necessarily intend for the questionnaires to act as a single measure of usability (...) they are often used by practitioners as a way to measure usability with one number", because this practice is not discourage by the questionnaire itself.

Nonetheless, SUS probably continues to be one of the most popular questionnaire for measuring the perception of usability, but it is only efficient when used along side other usability tests to understand the participants impression of the overall system. SUS uses a 10 item questionnaire with 5 response options: 1) I think that I would like to use this system frequently; 2) I found the system unnecessarily complex; 3) I thought the system was easy to use; 4) I think that I would need the support of a technical person to be able to use this system; 5) I found the various functions in this system were well integrated; 6) I thought there was too much inconsistency in this system; 7) I would imagine that most people would learn to use this system very quickly; 8) I found the system very cumbersome to use; 9) I felt very confident using the system; and 10) I needed to learn a lot of things before I could get going with this system. The response format ranges from *Strongly Disagree* to *Strongly Agree*

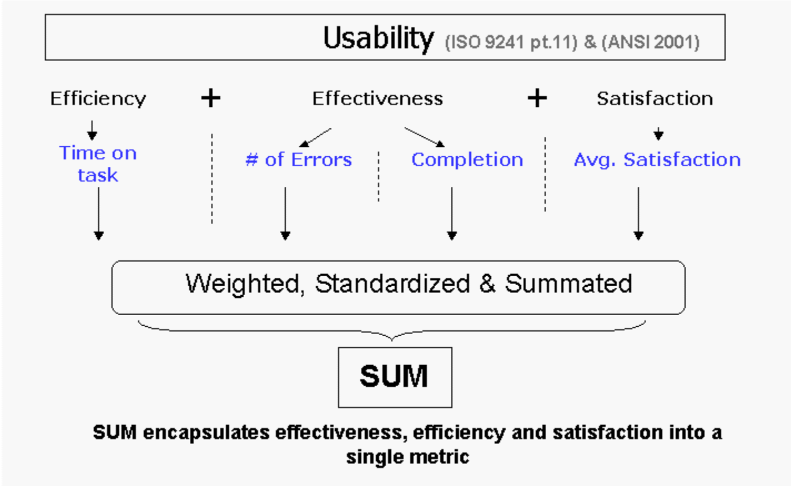


Figure 11.1: SUM Model (Figure 1 (Sauro and Kindlund, 2005b))

Metrics such as completion rates, UI problems, task time, task level satisfaction, errors, expectation, page views and clicks, and conversion rates, can be useful for usability evaluation. However, methods such as SUMI and SUS (Brooke, 1996) are too focused on user satisfaction. Therefore, SUM presents itself as an alternative to these methods because it combines several metrics into a single score (task completion rates, task time, satisfaction, and error counts), facilitating the comparison of different systems in terms of usability. Even when using SUM, post-test questionnaires, such as the SUS could still be applied to gauge user experience.

According to Sauro and Kindlund (2005b), the Single Usability Metric (SUM) is able to evaluate the three dimensions of usability standardized by the International Organization for Standardization Technical Committee in ISO 9241 pt.11: effectiveness, efficiency, and satisfaction. "The four metrics that are used to derive the SUM score of a system in a summative evaluation are: task completion rates, average number of errors, average time on task, and post-task satisfaction" (Sauro and Kindlund, 2005b, p. 10). Although time on task in visualization might not be evaluated in the same way as in a common usability evaluation scenario. For instance, if a user spends a lot of time exploring the visualization is not necessarily an evidence that he/she is failing in the exploration. However, this metric is related to post-task satisfaction metrics, therefore it is still important.

In this task based evaluation there will be some tasks (preferably not more than two) that will be the same — with the same expected responses — for the three versions of the same visualization, in order to be able to collect information on whether or not interactivity does facilitate the process of exploring the visualization and finding information; some that will be the same for the different types of representations, in order to be able to collect information on which type of visual representation works better; and additional tasks that can reveal particular aspects of interactive versions. After completing each task the participant will be asked to complete a post-task questionnaire and give a response (likert-type scale) 5-point semantic distance scales with the end points labeled. The questions will be the following:

- How would you describe the level of difficulty of this task? (1:Very Difficult to 5:Very

Easy)

- How satisfied are you with using this visualization to complete this task? (1:Very Unsatisfied to 5:Very Satisfied)
- How would you rate the amount of time it took to complete this task? (1:Too Much Time to 5:Very Little Time)
- How would you rate the overall experience? (1:Boring to 5:Very Enjoyable)

This study could hopefully provide confirmation of some of the ideas presented in this thesis and even reveal new valuable strategies for narrative visualization. Additionally, it will also give the desired answers to the three general framing questions: What are the best techniques to tell stories in a more visual way? Which elements have to be present in order to achieve an effective storytelling? What types of visualizations appeal more to the public?

11.3 Closing remarks

We reached a high point in the maturation of the Information Age and digital information is finally part of the cultural fabric of most of the developed world. Although there is still a digital divide and not everyone has equal access to the available information and communication technologies, in the Western societies "the diffusion of the Internet is reaching a level between 80% and 90%" (Friemel, 2016, p. 313), only subsisting a *grey divide* that excludes older seniors of age 65+ years, and even in developing countries Internet penetration has been slowly rising and now stands at 35%, according to the ITU (the United Nations specialized agency for information and communication technologies) *ICT Facts and Figures – The world in 2015* report[1]. Therefore, and adding the fact that the amount of data being produced, collected and becoming available is growing, research on how to access, process and represent this data is becoming more important. These are issues that are also being approached outside academia.

According to Danziger (2008), our information literacy is increasing. "As information becomes more ubiquitous in a variety of forms (though primarily as digital content delivered via the internet), we, as a general public, are becoming more comfortable with our ability to navigate these information spaces as part of our everyday lives: we use Google to search the web, we look for audiovisual content on sites like YouTube and Flickr, we make informed purchasing decisions based on sophisticated analysis of information-rich commerce sites such as Amazon, we exchange information and ideas across complex networks of blogs and news aggregators" (Danziger, 2008, p. 73). Therefore, our ability to understand visual representations of data is also slowly increasing, because we are also more exposed to visualization now. Consequently, Information visualization is increasing in popularity and will probably grow to play a significant role in the way we communicate data. However, for Danziger (2008), there has to be an effort to change information visualization in order to allow people that are not experienced data analysts to be able to make sense of this information.

[1] http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2015.pdf

Over the course of my research, I have seen increasing interest in narrative visualization and in using storytelling in visualization. Storytelling is one of the strategies that can be used in information visualization to ease the comprehension of complex data in a way that is engaging. With this dissertation, I believe that I contributed to the understanding of some of the requirements to achieve the feel of storytelling in visualization, and this can not only push research in this field forward but can also help practitioners build new more effective visualizations. Moreover, I identified some of the current challenges in narrative visualization and in the general field of information visualization.

Bibliography

Abela, A. (2008). *Advanced Presentations by Design: Creating Communication that Drives Action.* Pfeiffer essential resources for training and HR professionals. Wiley.

Ackoff, R. (1989). “From data to wisdom”. In: *Journal of Applied Systems Analysis 16* 16, pp. 3–9.

Aigner, W. (2011). “Understanding the Role and Value of Interaction: First Steps”. In: *Proceedings of International Workshop on Visual Analytics (EuroVA 2011).* Ed. by S. Miksch and G. Santucci. Vortrag: International Workshop on Visual Analytics (EuroVA 2011), Bergen, Norway; 2011-05-31. The Eurographics Association, pp. 17–20.

Amar, R., J. Eagan, and J. Stasko (2005). “Low-level components of analytic activity in information visualization”. In: *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*, pp. 111–117.

Aristóteles, P. (1993). “tradução de Eudoro de Souza”. In: *São Paulo: Ars Poética.*

Bargeron, D. and T. Moscovich (2003). “Reflowing Digital Ink Annotations”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* CHI ’03. Ft. Lauderdale, Florida, USA: ACM, pp. 385–393. DOI: 10.1145/642611.642678.

Barthes, R. (2004). “Introduction to the structural analysis of narratives”. In: *Narrative theory: Critical concepts in literary and cultural studies*, pp. 65–116.

Bastian, M., S. Heymann, M. Jacomy, et al. (2009). “Gephi: an open source software for exploring and manipulating networks.” In: *ICWSM* 8, pp. 361–362.

Bateman, S., R. Mandryk, C. Gutwin, A. Genest, D. McDine, and C. Brooks (2010). “Useful junk?: the effects of visual embellishment on comprehension and memorability of charts”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* CHI ’10. Atlanta, Georgia, USA: ACM, pp. 2573–2582. DOI: 10.1145/1753326.1753716.

Bawden, D. and L. Robinson (2009). “The dark side of information: overload, anxiety and other paradoxes and pathologies”. In: *Journal of information science* 35.2, pp. 180–191.

Bederson, B. and A. Boltman (1999). “Does animation help users build mental maps of spatial information?” In: *Information Visualization, 1999. (Info Vis ’99) Proceedings. 1999 IEEE Symposium on*, pp. 28–35. DOI: 10.1109/INFVIS.1999.801854.

Bederson, B., A. Clamage, M. Czerwinski, and G. Robertson (2004). "DateLens: A Fisheye Calendar Interface for PDAs". In: *ACM Trans. Comput.-Hum. Interact.* 11.1, pp. 90–119. DOI: 10.1145/972648.972652.

Bederson, B., J. Hollan, K. Perlin, J. Meyer, D. Bacon, and G. Furnas (1996). "Pad++: A Zoomable Graphical Sketchpad For Exploring Alternate Interface Physics". In: *Journal of Visual Languages & Computing* 7.1, pp. 3–32. DOI: 10.1006/jvlc.1996.0002.

Bendix, F., R. Kosara, and H. Hauser (2005). "Parallel sets: visual analysis of categorical data". In: *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*, pp. 133–140. DOI: 10.1109/INFVIS.2005.1532139.

Bertschi, S., S. Bresciani, T. Crawford, R. Goebel, W. Kienreich, M. Lindner, V. Sabol, and A. Vande Moere (2011). "What is Knowledge Visualization? Perspectives on an Emerging Discipline". In: *Information Visualisation (IV), 2011 15th International Conference on*, pp. 329–336. DOI: 10.1109/IV.2011.58.

Bezerianos, A., F. Chevalier, P. Dragicevic, N. Elmqvist, and J.-D. Fekete (2010). "GraphDice: A System for Exploring Multivariate Social Networks". In: *Computer Graphics Forum* 29.3, pp. 863–872. DOI: 10.1111/j.1467-8659.2009.01687.x.

Bladh, T., D. Carr, and M. Kljun (2005). "The effect of animated transitions on user navigation in 3D tree-maps". In: *Information Visualisation, 2005. Proceedings. Ninth International Conference on*, pp. 297–305. DOI: 10.1109/IV.2005.122.

Bogost, I. (2007). *Persuasive games: The expressive power of videogames*. Mit Press.

Bogost, I., S. Ferrari, and B. Schweizer (2010). *Newsgames*. The MIT Press.

Bostock, M., V. Ogievetsky, and J. Heer (2011). "D^3 data-driven documents". In: *Visualization and Computer Graphics, IEEE Transactions on* 17.12, pp. 2301–2309.

Boy, J., F. Detienne, and J.-D. Fekete (2015). "Storytelling in Information Visualizations: Does It Engage Users to Explore Data?" In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. CHI '15. Seoul, Republic of Korea: ACM, pp. 1449–1458. DOI: 10.1145/2702123.2702452.

Boy, J., R. Rensink, E. Bertini, and J.-D. Fekete (2014). "A Principled Way of Assessing Visualization Literacy". In: *Visualization and Computer Graphics, IEEE Transactions on* 20.12, pp. 1963–1972. DOI: 10.1109/TVCG.2014.2346984.

Bradshaw, L. (2014). "The Great Data Revolution". In: *The Future Of*, p. 21.

Brooke, J. et al. (1996). "SUS-A quick and dirty usability scale". In: *Usability evaluation in industry* 189.194, pp. 4–7.

Bruns, A. (2008). *Blogs, Wikipedia, Second Life, and Beyond: From Production to Produsage*. Digital formations. Peter Lang.

Büring, T. and H. Reiterer (2005). "ZuiScat: Querying and Visualizing Information Spaces on Personal Digital Assistants". In: *Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices &Amp; Services*. MobileHCI '05. Salzburg, Austria: ACM, pp. 129–136. DOI: 10.1145/1085777.1085799.

Burkhard, R. (2005). "Knowledge Visualization: The Use of Complementary Visual Representations for the Transfer of Knowledge: a Model, a Framework, and Four New Approaches". PhD thesis. Swiss Federal Institute for Environmental Science and Technology.

Burkhard, R. and M. Meier (2005). "Tube Map Visualization: Evaluation of a Novel Knowledge Visualization Application for the Transfer of Knowledge in Long-Term Projects". In: *j-jucs* 11.4, pp. 473–494.

Burton, J. (2005). "News-game journalism: History, current use and possible futures". In:

Byron, L. and M. Wattenberg (2008). "Stacked Graphs: Geometry and Aesthetics". In: *IEEE Transactions on Visualization and Computer Graphics* 14.6, pp. 1245–1252. DOI: 10.1109/TVCG.2008.166.

Cairo, A. (2012). *The Functional Art*. Peachpit Press.

Card, S. and J. Mackinlay (1997). "The structure of the information visualization design space". In: *Information Visualization, 1997. Proceedings., IEEE Symposium on*, pp. 92–99. DOI: 10.1109/INFVIS.1997.636792.

Card, S., J. Mackinlay, and B. Shneiderman (1999). *Readings in information visualization: using vision to think*. Morgan Kaufmann.

Cawthon, N. and A. Vande Moere (2007). "The Effect of Aesthetic on the Usability of Data Visualization". In: *Information Visualization, 2007. IV '07. 11th International Conference*, pp. 637–648. DOI: 10.1109/IV.2007.147.

Chen, C. (2005). "Top 10 unsolved information visualization problems". In: *Computer Graphics and Applications, IEEE* 25.4, pp. 12–16. DOI: 10.1109/MCG.2005.91.

— (2010). "Information visualization". In: *Wiley Interdisciplinary Reviews: Computational Statistics* 2.4, pp. 387–403. DOI: 10.1002/wics.89.

Chen, C. and Y. Yu (2000). "Empirical studies of information visualization: a meta-analysis". In: *International Journal of Human-Computer Studies* 53.5, pp. 851–866.

Chen, M., D. Ebert, H. Hagen, R. Laramee, R. Van Liere, K.-L. Ma, W. Ribarsky, G. Scheuermann, and D. Silver (2009). "Data, Information, and Knowledge in Visualization". In: *Computer Graphics and Applications, IEEE* 29.1, pp. 12–19. DOI: 10.1109/MCG.2009.6.

Chevalier, F., P. Dragicevic, A. Bezerianos, and J.-D. Fekete (2010). "Using Text Animated Transitions to Support Navigation in Document Histories". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '10. Atlanta, Georgia, USA: ACM, pp. 683–692. DOI: 10.1145/1753326.1753427.

Chevalier, F., P. Dragicevic, and S. Franconeri (2014). "The Not-so-Staggering Effect of Staggered Animated Transitions on Visual Tracking". In: *Visualization and Computer Graphics, IEEE Transactions on* 20.12, pp. 2241–2250. DOI: 10.1109/TVCG.2014.2346424.

Chevalier, F., P. Dragicevic, and C. Hurter (2012). "Histomages: Fully Synchronized Views for Image Editing". In: *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*. UIST '12. Cambridge, Massachusetts, USA: ACM, pp. 281–286. DOI: 10.1145/2380116.2380152.

Chi, E. (2000). "A taxonomy of visualization techniques using the data state reference model". In: *Information Visualization, 2000. InfoVis 2000. IEEE Symposium on*, pp. 69–75. DOI: 10.1109/INFVIS.2000.885092.

Chin, J. P., V. A. Diehl, and K. L. Norman (1988). "Development of an Instrument Measuring User Satisfaction of the Human-computer Interface". In: *Proceedings of the SIGCHI*

Conference on Human Factors in Computing Systems. CHI '88. Washington, D.C., USA: ACM, pp. 213–218. DOI: 10.1145/57167.57203.

Cook, K., R. Earnshaw, and J. Stasko (2007). "Guest Editors' Introduction: Discovering the Unexpected". In: *Computer Graphics and Applications, IEEE* 27.5, pp. 15–19. DOI: 10.1109/MCG.2007.126.

Cook, K. and J. Thomas (2005). *Illuminating the path: The research and development agenda for visual analytics*. Tech. rep. Pacific Northwest National Laboratory (PNNL), Richland, WA (US).

Craft, B. and P. Cairns (2005). "Beyond guidelines: what can we learn from the visual information seeking mantra?" In: *Information Visualisation, 2005. Proceedings. Ninth International Conference on*, pp. 110–118. DOI: 10.1109/IV.2005.28.

Dahlgren, P. (2005). "The Internet, public spheres, and political communication: Dispersion and deliberation". In: *Political communication* 22.2, pp. 147–162.

Damásio, A. (1999). *The feeling of what happens: Body and emotion in the making of consciousness*. New York: Harcourt Brace.

Danziger, M. (2008). "Information visualization for the people". PhD thesis. Massachusetts Institute of Technology.

Deterding, S., D. Dixon, R. Khaled, and L. Nacke (2011). "From Game Design Elements to Gamefulness: Defining "Gamification"". In: *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*. MindTrek '11. Tampere, Finland: ACM, pp. 9–15.

Deterding, S., M. Sicart, L. Nacke, K. O'Hara, and D. Dixon (2011). "Gamification. Using Game-design Elements in Non-gaming Contexts". In: *CHI '11 Extended Abstracts on Human Factors in Computing Systems*. CHI EA '11. Vancouver, BC, Canada: ACM, pp. 2425–2428.

Deuze, M. (2001). "Online journalism: Modelling the first generation of news media on the World Wide Web". In: *First Monday* 6.10.

Dewar, M. (2012). *Getting Started with D3*. Oreilly and Associate Series. O'Reilly Media.

Diakopoulos, N. (2010). "Game-y Information Graphics". In: *CHI '10 Extended Abstracts on Human Factors in Computing Systems*. CHI EA '10. Atlanta, Georgia, USA: ACM, pp. 3595–3600. DOI: 10.1145/1753846.1754024.

— (2011). "Design Challenges in Playable Data". In: *CHI Workshop on Gamification*.

Diakopoulos, N., F. Kivran-Swaine, and M. Naaman (2011). "Playable data: characterizing the design space of game-y infographics". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '11. Vancouver, BC, Canada: ACM, pp. 1717–1726. DOI: 10.1145/1978942.1979193.

Dix, A. and G. Ellis (1998). "Starting Simple: Adding Value to Static Visualisation Through Simple Interaction". In: *Proceedings of the Working Conference on Advanced Visual Interfaces*. AVI '98. L'Aquila, Italy: ACM, pp. 124–134.

Dörk, M. (2012). "Visualization for Search: Exploring Complex and Dynamic Information Spaces". PhD thesis. University of Calgary.

Dörk, M., S. Carpendale, and C. Williamson (2011). "The Information Flaneur: A Fresh Look at Information Seeking". In: *Proceedings of the SIGCHI Conference on Human Factors* *in Computing Systems*. CHI '11. Vancouver, BC, Canada: ACM, pp. 1215–1224. DOI: 10.1145/1978942.1979124.

Dörk, M., P. Feng, C. Collins, and S. Carpendale (2013). "Critical InfoVis: Exploring the Politics of Visualization". In: *CHI '13 Extended Abstracts on Human Factors in Computing Systems*. CHI EA '13. Paris, France: ACM, pp. 2189–2198. DOI: 10.1145/2468356.2468739.

Dove, G. and S. Jones (2012). "Narrative Visualization: Sharing Insights into Complex Data". In: *Interfaces and Human Computer Interaction (IHCI 2012)*.

Dykes, J., J. Wood, and A. Slingsby (2010). "Rethinking Map Legends with Visualization". In: *Visualization and Computer Graphics, IEEE Transactions on* 16.6, pp. 890–899. DOI: 10.1109/TVCG.2010.191.

Elias, M., M.-A. Aufaure, and A. Bezerianos (2013). "Human-Computer Interaction – INTERACT 2013: 14th IFIP TC 13 International Conference, Cape Town, South Africa, September 2-6, 2013, Proceedings, Part III". In: ed. by P. Kotzé, G. Marsden, G. Lindgaard, J. Wesson, and M. Winckler. Berlin, Heidelberg: Springer Berlin Heidelberg. Chap. Storytelling in Visual Analytics Tools for Business Intelligence, pp. 280–297. DOI: 10.1007/978-3-642-40477-1_18.

Elmqvist, N., P. Dragicevic, and J.-D. Fekete (2008). "Rolling the Dice: Multidimensional Visual Exploration using Scatterplot Matrix Navigation". In: *Visualization and Computer Graphics, IEEE Transactions on* 14.6, pp. 1539–1148. DOI: 10.1109/TVCG.2008.153.

Elmqvist, N. and J. S. Yi (2015). "Patterns for visualization evaluation". In: *Information Visualization* 14.3, pp. 250–269.

Fayyad, U., A. Wierse, and G. Grinstein (2002). *Information Visualization in Data Mining and Knowledge Discovery*. The Morgan Kaufmann series in data management systems. Morgan Kaufmann.

Few, S. (2012). *Show Me the Numbers: Designing Tables and Graphs to Enlighten*. Analytics Press.

Fidler, R. (1997). *Mediamorphosis: Understanding new media*. Sage Publications.

Figueiras, A. (2013). "A Typology for Data Visualization on the Web". In: *Information Visualisation (IV), 2013 17th International Conference*, pp. 351–358. DOI: 10.1109/IV.2013.45.

— (2014a). "How to Tell Stories Using Visualization". In: *Information Visualisation (IV), 2014 18th International Conference on*, pp. 18–18. DOI: 10.1109/IV.2014.78.

— (2014b). "Narrative Visualization: A Case Study of How to Incorporate Narrative Elements in Existing Visualizations". In: *Information Visualisation (IV), 2014 18th International Conference on*, pp. 46–52. DOI: 10.1109/IV.2014.79.

— (2015). "Towards the Understanding of Interaction in Information Visualization". In: *2015 19th International Conference on Information Visualisation*, pp. 140–147. DOI: 10.1109/iV.2015.34.

Fisher, D., A. Hoff, G. Robertson, and M. Hurst (2008). "Narratives: A visualization to track narrative events as they develop". In: *Visual Analytics Science and Technology, 2008. VAST '08. IEEE Symposium on*, pp. 115–122. DOI: 10.1109/VAST.2008.4677364.

Floridi, L. (2005). "Is Semantic Information Meaningful Data?" In: *Philosophy and Phenomenological Research* 70.2, pp. 351–370. DOI: 10.1111/j.1933-1592.2005.tb00531.x.

Friemel, T. N. (2016). “The digital divide has grown old: Determinants of a digital divide among seniors”. In: *new media & society* 18.2, pp. 313–331.

Friendly, M. (2000). “Re-visions of Minard”. In: *Statistical Computing & Statistical Graphics Newsletter* 11.1, p. 1.

— (2002). “Visions and re-visions of Charles Joseph Minard”. In: *Journal of Educational and Behavioral Statistics* 27.1, pp. 31–51.

— (2005). “Milestones in the History of Data Visualization: A Case Study in Statistical Historiography”. English. In: *Classification — the Ubiquitous Challenge*. Ed. by C. Weihs and W. Gaul. Studies in Classification, Data Analysis, and Knowledge Organization. Springer Berlin Heidelberg, pp. 34–52. DOI: 10.1007/3-540-28084-7_4.

— (2008). “A Brief History of Data Visualization”. English. In: *Handbook of Data Visualization*. Springer Handbooks Comp.Statistics. Springer Berlin Heidelberg, pp. 15–56. DOI: 10.1007/978-3-540-33037-0_2.

Furnas, G. and B. Bederson (1995). “Space-scale Diagrams: Understanding Multiscale Interfaces”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’95. Denver, Colorado, USA: ACM Press/Addison-Wesley Publishing Co., pp. 234–241. DOI: 10.1145/223904.223934.

Gershon, N., S. Eick, and S. Card (1998). “Information Visualization”. In: *interactions* 5.2, pp. 9–15. DOI: 10.1145/274430.274432.

Gershon, N. and W. Page (2001). “What storytelling can do for information visualization”. In: *Commun. ACM* 44.8, pp. 31–37. DOI: 10.1145/381641.381653.

Giussani, B. (1997). “A new media tells different stories”. In: *First Monday* 2.4.

Goebel, R., S. Wei, and Y. Tanaka (2014). “The Challenge of Semantic Symmetry in Visualization”. In: *Information Visualisation (IV), 2014 18th International Conference on*, pp. 27–33. DOI: 10.1109/IV.2014.32.

Gonzalez, C. (1996). “Does Animation in User Interfaces Improve Decision Making?” In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’96. Vancouver, British Columbia, Canada: ACM, pp. 27–34. DOI: 10.1145/238386.238396.

Gonzalez, H., A. Halevy, C. S. Jensen, A. Langen, J. Madhavan, R. Shapley, and W. Shen (2010). “Google Fusion Tables: Data Management, Integration and Collaboration in the Cloud”. In: *Proceedings of the 1st ACM Symposium on Cloud Computing*. SoCC ’10. Indianapolis, Indiana, USA: ACM, pp. 175–180. DOI: 10.1145/1807128.1807158.

Gratzl, S., A. Lex, N. Gehlenborg, N. Cosgrove, and M. Streit (2016). “From Visual Exploration to Storytelling and Back Again”. In: *bioRxiv*, p. 049585.

Gray, J., L. Chambers, and L. Bounegru (2012). *The data journalism handbook*. " O’Reilly Media, Inc."

Hall, J. (2001). *Online journalism: a critical primer*. Pluto Press.

Han, J., M. Kamber, and J. Pei (2011). *Data Mining: Concepts and Techniques*. The Morgan Kaufmann Series in Data Management Systems. Elsevier Science.

Harris, J. (2011). “Word clouds considered harmful”. In: *New York Times* 13.

Healey, C. and J. Enns (2012). “Attention and Visual Memory in Visualization and Computer Graphics”. In: *Visualization and Computer Graphics, IEEE Transactions on* 18.7, pp. 1170–1188. DOI: 10.1109/TVCG.2011.127.

Heer, J. (2006). “Socializing visualization”. In: *In CHI 2006 Workshop on Social Visualization*.

Heer, J. M. (2008). *Supporting asynchronous collaboration for interactive visualization*. ProQuest.

Heer, J. and M. Agrawala (2008). “Design Considerations for Collaborative Visual Analytics”. In: *Information Visualization* 7.1, pp. 49–62. DOI: 10.1057/palgrave.ivs.9500167. eprint: http://ivi.sagepub.com/content/7/1/49.full.pdf+html.

Heer, J., F. van Ham, S. Carpendale, C. Weaver, and P. Isenberg (2008). “Creation and Collaboration: Engaging New Audiences for Information Visualization”. English. In: *Information Visualization*. Ed. by A. Kerren, J. Stasko, J.-D. Fekete, and C. North. Vol. 4950. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 92–133. DOI: 10.1007/978-3-540-70956-5_5.

Heer, J., J. Mackinlay, C. Stolte, and M. Agrawala (2008). “Graphical Histories for Visualization: Supporting Analysis, Communication, and Evaluation”. In: *Visualization and Computer Graphics, IEEE Transactions on* 14.6, pp. 1189–1196. DOI: 10.1109/TVCG.2008.137.

Heer, J. and G. Robertson (2007). “Animated Transitions in Statistical Data Graphics”. In: *Visualization and Computer Graphics, IEEE Transactions on* 13.6, pp. 1240–1247. DOI: 10.1109/TVCG.2007.70539.

Heer, J., F. Viégas, and M. Wattenberg (2009). “Voyagers and Voyeurs: Supporting Asynchronous Collaborative Visualization”. In: *Commun. ACM* 52.1, pp. 87–97. DOI: 10.1145/1435417.1435439.

Herman, D. (2011). *Basic elements of narrative*. John Wiley & Sons.

Herman, I., G. Melancon, and M. Marshall (2000). “Graph visualization and navigation in information visualization: A survey”. In: *Visualization and Computer Graphics, IEEE Transactions on* 6.1, pp. 24–43. DOI: 10.1109/2945.841119.

Hullman, J. and N. Diakopoulos (2011). “Visualization Rhetoric: Framing Effects in Narrative Visualization”. In: *Visualization and Computer Graphics, IEEE Transactions on* 17.12, pp. 2231–2240. DOI: 10.1109/TVCG.2011.255.

Hullman, J., N. Diakopoulos, and E. Adar (2013). “Contextifier: Automatic Generation of Annotated Stock Visualizations”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’13. Paris, France: ACM, pp. 2707–2716. DOI: 10.1145/2470654.2481374.

Hullman, J., S. Drucker, N. H. Riche, B. Lee, D. Fisher, and E. Adar (2013). “A Deeper Understanding of Sequence in Narrative Visualization”. In: *Visualization and Computer Graphics, IEEE Transactions on* 19.12, pp. 2406–2415. DOI: 10.1109/TVCG.2013.119.

Isenberg, P. (2009). “Collaborative information visualization in co-located environments”. PhD thesis. PhD thesis, University of Calgary.

Isenberg, P., N. Elmqvist, J. Scholtz, D. Cernea, K.-L. Ma, and H. Hagen (2011). “Collaborative visualization: definition, challenges, and research agenda”. In: *Information Visualization* 10.4, pp. 310–326.

Jagadish, H. V. (2015). “Big Data and Science: Myths and Reality”. In: *Big Data Research* 2.2. Visions on Big Data, pp. 49–52. DOI: 10.1016/j.bdr.2015.01.005.

Jagadish, H. V., J. Gehrke, A. Labrinidis, Y. Papakonstantinou, J. Patel, R. Ramakrishnan, and C. Shahabi (2014). "Big Data and Its Technical Challenges". In: *Commun. ACM* 57.7, pp. 86–94. DOI: 10.1145/2611567.

Janssen, K. (2011). "The influence of the {PSI} directive on open government data: An overview of recent developments". In: *Government Information Quarterly* 28.4, pp. 446–456. DOI: 10.1016/j.giq.2011.01.004.

Jenkins, H. (2006). *Fans, Bloggers, and Gamers: Exploring Participatory Culture*. New York University Press.

Jentoft-Nilsen, M. (2015). *El Niño Watch 2015*. Ed. by b. o. p. p. b. A. K. NASA/JPL OSTM/Jason-2 and N. O. S. T. T. Bill Patzert. URL: http://svs.gsfc.nasa.gov/goto?30629.

Kamvar, S. and J. Harris (2011). "We Feel Fine and Searching the Emotional Web". In: *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*. WSDM '11. Hong Kong, China: ACM, pp. 117–126. DOI: 10.1145/1935826.1935854.

Kapler, T. and W. Wright (2005). "GeoTime Information Visualization". In: *Information Visualization* 4.2, pp. 136–146. DOI: 10.1057/palgrave.ivs.9500097. eprint: http://ivi.sagepub.com/content/4/2/136.full.pdf+html.

Keim, D. (2002). "Information visualization and visual data mining". In: *Visualization and Computer Graphics, IEEE Transactions on* 8.1, pp. 1–8. DOI: 10.1109/2945.981847.

Keim, D., G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon (2008). "Visual Analytics: Definition, Process, and Challenges". English. In: *Information Visualization*. Ed. by A. Kerren, J. Stasko, J.-D. Fekete, and C. North. Vol. 4950. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 154–175. DOI: 10.1007/978-3-540-70956-5_7.

Keim, D., F. Mansmann, J. Schneidewind, J. Thomas, and H. Ziegler (2008). "Visual Analytics: Scope and Challenges". English. In: *Visual Data Mining*. Ed. by S. Simoff, M. Böhlen, and A. Mazeika. Vol. 4404. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 76–90. DOI: 10.1007/978-3-540-71080-6_6.

Keller, T. and S.-O. Tergan (2005). "Visualizing Knowledge and Information: An Introduction". In: *Knowledge and Information Visualization*. Ed. by S.-O. Tergan and T. Keller. Vol. 3426. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 1–23. DOI: 10.1007/11510154_1.

Kim, S.-H., J. S. Yi, and N. Elmqvist (2014). "Oopsy-daisy: Failure Stories in Quantitative Evaluation Studies for Visualizations". In: *Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*. BELIV '14. Paris, France: ACM, pp. 142–146. DOI: 10.1145/2669557.2669576.

Kirakowski, J. and M. Corbett (1993). "SUMI: The software usability measurement inventory". In: *British journal of educational technology* 24.3, pp. 210–212.

Kirk, A. (2012). *Data Visualization: a successful design process*. Packt Publishing Ltd.

Koch, K., J. McLean, R. Segev, M. Freed, M. Berry II, V. Balasubramanian, and P. Sterling (2006). "How Much the Eye Tells the Brain". In: *Current Biology* 16.14, pp. 1428–1434. DOI: 10.1016/j.cub.2006.05.056.

Komlodi, A., A. Sears, and E. Stanziola (2004). *Information visualization evaluation review*. Tech. rep. ISRC Tech. Report, Dept. of Information Systems, UMBC.

Kosara, R. (2007). "Visualization Criticism - The Missing Link Between Information Visualization and Art". In: *Information Visualization, 2007. IV '07. 11th International Conference*, pp. 631–636. DOI: 10.1109/IV.2007.130.

Kosara, R., F. Bendix, and H. Hauser (2006). "Parallel Sets: interactive exploration and visual analysis of categorical data". In: *IEEE Transactions on Visualization and Computer Graphics* 12.4, pp. 558–568. DOI: 10.1109/TVCG.2006.76.

Kosara, R. and J. Mackinlay (2013). "Storytelling: The Next Step for Visualization". In: *Computer* 46.5, pp. 44–50. DOI: 10.1109/MC.2013.36.

Krum, R. (2013). *Cool Infographics: Effective Communication with Data Visualization and Design*. Wiley.

Lam, H., E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale (2011). *Seven Guiding Scenarios for Information Visualization Evaluation*. Research Report 2011-992-04. Superseded by and improved in a follow-up journal article.

Lam, H., E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale (2012). "Empirical Studies in Information Visualization: Seven Scenarios". In: *Visualization and Computer Graphics, IEEE Transactions on* 18.9, pp. 1520–1536. DOI: 10.1109/TVCG.2011.279.

Lang, A. (2009). "Aesthetics in information visualization". In: *Trends in Information Visualization*, p. 8.

Lankow, J., J. Ritchie, and R. Crooks (2012). *Infographics: The power of visual storytelling*. John Wiley & Sons.

Lau, A. and A. Vande Moere (2007). "Towards a Model of Information Aesthetics in Information Visualization". In: *Information Visualization, 2007. IV '07. 11th International Conference*, pp. 87–92. DOI: 10.1109/IV.2007.114.

Lee, B., N. Henry Riche, P. Isenberg, and S. Carpendale (2015). "More than Telling a Story: A Closer Look at the Process of Transforming Data into Visually Shared Stories". In: *IEEE Computer Graphics and Applications* 35.5, pp. 84–90. DOI: 10.1109/MCG.2015.99.

Lewis, J. R. (1992). "Psychometric Evaluation of the Post-Study System Usability Questionnaire: The PSSUQ". In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 36.16, pp. 1259–1260. DOI: 10.1177/154193129203601617. eprint: http://pro.sagepub.com/content/36/16/1259.full.pdf+html.

Li, L., F. Li, and R. Lau (2006). "A Trajectory-Preserving Synchronization Method for Collaborative Visualization". In: *Visualization and Computer Graphics, IEEE Transactions on* 12.5, pp. 989–996. DOI: 10.1109/TVCG.2006.114.

Liang, H.-N., P. Parsons, H.-C. Wu, and K. Sedig (2010). "An Exploratory Study of Interactivity in Visualization Tools: Flow of Interaction". In: *Journal of Interactive Learning Research* 21.1, pp. 5–45.

Lima, M. (2009). *Information Visualization Manifesto*. URL: http://www.visualcomplexity.com/vc/blog/?p=644.

Liu, S., Y. Wu, E. Wei, M. Liu, and Y. Liu (2013). "StoryFlow: Tracking the Evolution of Stories". In: *Visualization and Computer Graphics, IEEE Transactions on* 19.12, pp. 2436–2445. DOI: 10.1109/TVCG.2013.196.

Lohr, S. (2012). "The age of big data". In: *New York Times* 11.

Ma, K.-L., I. Liao, J. Frazier, H. Hauser, and H.-N. Kostis (2012). "Scientific Storytelling Using Visualization". In: *Computer Graphics and Applications, IEEE* 32.1, pp. 12–19. DOI: 10.1109/MCG.2012.24.

Macklin, C., J. Wargaski, M. Edwards, and K. Y. Li (2009). "DATAPLAY: Mapping game mechanics to traditional data visualization". In: *Proceedings of DiGRA*. Vol. 2009.

Manovich, L. (2002). "The anti-sublime ideal in data art". In: *URL= http://www. manovich. net/DOCS/data_art. doc*.

Martinez, A. (2013). *New York Times' Jill Abramson @ ISOJ: "To Snow Fall" now means to tell a great multimedia story*. URL: https://knightcenter.utexas.edu/blog/00-13616-new-york-times%E2%80%99-jill-abramson-isoj-snow-fall-now-means-tell-great-multimedia-story.

Masud, L., F. Valsecchi, P. Ciuccarelli, D. Ricci, and G. Caviglia (2010). "From Data to Knowledge - Visualizations as Transformation Processes within the Data-Information-Knowledge Continuum". In: *Information Visualisation (IV), 2010 14th International Conference*, pp. 445–449. DOI: 10.1109/IV.2010.68.

Maus, F. E. (1991). "Music As Narrative". In: *Indiana Theory Review* 12, pp. 1–34.

Maxwell, J. C. (1990). *The Scientific Letters and Papers of James Clerk Maxwell: 1846-1862*. CUP Archive.

McGonigal, J. (2011). *Reality is broken: Why games make us better and how they can change the world*. Penguin.

Miller, G. (1983). "Information Theory in Psychology". In: *The Study of Information: Interdisciplinary Messages*. Ed. by F. Machlup. Wiley, pp. 493–496.

Murray, S. (2013). *Interactive data visualization for the Web*. " O'Reilly Media, Inc."

Nagel, T., E. Duval, A. Vande Moere, K. Kloeckl, and C. Ratti (2012). "Sankey arcs–visualizing edge weights in path graphs". In: *The Eurographics Conference on Visualization*. Eurographics Association, pp. 55–59.

Nesbitt, K. and C. Friedrich (2002). "Applying Gestalt principles to animated visualizations of network data". In: *Information Visualisation, 2002. Proceedings. Sixth International Conference on*, pp. 737–743. DOI: 10.1109/IV.2002.1028859.

Nichani, M. and V. Rajamanickam (2003). *Interactive Visual Explainers–A Simple Classification.*

Nightingale, F., W. Farr, and A. Smith (1859). *A contribution to the sanitary history of the British army during the late war with Russia*. John W. Parker and Son.

Norman, D. A. (2005). *Emotional design: Why we love (or hate) everyday things*. Basic books.

North, C. (2006). "Toward measuring visualization insight". In: *Computer Graphics and Applications, IEEE* 26.3, pp. 6–9.

Pandey, A. V., A. Manivannan, O. Nov, M. Satterthwaite, and E. Bertini (2014). "The Persuasive Power of Data Visualization". In: *Visualization and Computer Graphics, IEEE Transactions on* 20.12, pp. 2211–2220. DOI: 10.1109/TVCG.2014.2346419.

Perlin, K. and D. Fox (1993). "Pad: An Alternative Approach to the Computer Interface". In: *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH '93. Anaheim, CA: ACM, pp. 57–64. DOI: 10.1145/166117.166125.

Plaisant, C. (2004). “The Challenge of Information Visualization Evaluation”. In: *Proceedings of the Working Conference on Advanced Visual Interfaces*. AVI ’04. Gallipoli, Italy: ACM, pp. 109–116. DOI: 10.1145/989863.989880.

Plaisant, C., J.-D. Fekete, and G. Grinstein (2008). “Promoting insight-based evaluation of visualizations: From contest to benchmark repository”. In: *Visualization and Computer Graphics, IEEE Transactions on* 14.1, pp. 120–134.

Plaisant, C., J. Grosjean, and B. Bederson (2002). “SpaceTree: supporting exploration in large node link tree, design evolution and empirical evaluation”. In: *Information Visualization, 2002. INFOVIS 2002. IEEE Symposium on*, pp. 57–64. DOI: 10.1109/INFVIS.2002.1173148.

Playfair, W. (1801). *The commercial and political atlas: representing, by means of stained copper-plate charts, the progress of the commerce, revenues, expenditure and debts of england during the whole of the eighteenth century*. T. Burton.

Pousman, Z., J. Stasko, and M. Mateas (2007). “Casual Information Visualization: Depictions of Data in Everyday Life”. In: *Visualization and Computer Graphics, IEEE Transactions on* 13.6, pp. 1145–1152. DOI: 10.1109/TVCG.2007.70541.

PRC (2010). *Americans Spending More Time Following the News*. The PEW Research Center For The People & The Press. URL: http://www.people-press.org/files/legacy-pdf/652.pdf.

— (2012). *2012 News Consumption Report*. URL: http://www.people-press.org/files/legacy-pdf/2012%20News%20Consumption%20Report.pdf.

— (2014). *State od News Media*. URL: http://www.journalism.org/files/2014/03/Overview.pdf.

Raje, R., M. Boyles, and S. Fang (1998). “CEV: collaborative environment for visualization using Java RMI.” In: *Concurrency - Practice and Experience* 10.11-13, pp. 1079–1085.

Remer, T. (1965). *Serendipity and the three princes: From the Peregrinaggio of 1557*. Norman, U. Oklahoma P.

Ricoeur, P. (1990). *Time and narrative*. Vol. 3. University of Chicago Press.

Riedl, M. and B. Harrison (2016). “Using Stories to Teach Human Values to Artificial Agents”. In: *Proceedings of the 2nd International Workshop on AI, Ethics and Society, Phoenix, Arizona, 2016*. AAAI-16. Phoenix, Arizona, USA.

RJMetrics (2015). *The State of Data Science*. URL: https://rjmetrics.com/resources/reports/the-state-of-data-science/.

Robertson, G., K. Cameron, M. Czerwinski, and D. Robbins (2002a). “Animated Visualization of Multiple Intersecting Hierarchies”. In: *Information Visualization* 1.1, pp. 50–65. DOI: 10.1057/palgrave.ivs.9500002. eprint: http://ivi.sagepub.com/content/1/1/50.full.pdf+html.

— (2002b). “Animated visualization of multiple intersecting hierarchies”. In: *Information Visualization* 1.1, pp. 50–65.

— (2002c). “Polyarchy Visualization: Visualizing Multiple Intersecting Hierarchies”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’02. Minneapolis, Minnesota, USA: ACM, pp. 423–430. DOI: 10.1145/503376.503452.

Robertson, G., S. Card, and J. Mackinlay (1993). "Information visualization using 3D interactive animation". In: *Commun. ACM* 36.4, pp. 57–71. DOI: 10.1145/255950.153577.

Robertson, G., R. Fernandez, D. Fisher, B. Lee, and J. Stasko (2008). "Effectiveness of Animation in Trend Visualization". In: *Visualization and Computer Graphics, IEEE Transactions on* 14.6, pp. 1325–1332. DOI: 10.1109/TVCG.2008.125.

Robertson, G., J. Mackinlay, and S. Card (1991). "Cone Trees: Animated 3D Visualizations of Hierarchical Information". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '91. New Orleans, Louisiana, USA: ACM, pp. 189–194. DOI: 10.1145/108844.108883.

Rodríguez, M. T., S. Nunes, and T. Devezas (2015). "Telling Stories with Data Visualization". In: *Proceedings of the 2015 Workshop on Narrative & Hypertext*. NHT '15. Guzelyurt, Northern Cyprus: ACM, pp. 7–11. DOI: 10.1145/2804565.2804567.

Rogers, S. (2013). *Facts are Sacred: Text only ebook*. Faber & Faber.

Rohrer, R. and E. Swing (1997). "Web-Based Information Visualization". In: *IEEE Computer Graphics and Applications* 17.4, pp. 52–59. DOI: 10.1109/38.595269.

Rowley, J. (2007). "The wisdom hierarchy: representations of the DIKW hierarchy". In: *Journal of Information Science* 33.2, pp. 163–180.

Saraiya, P., C. North, and K. Duca (2005). "An insight-based methodology for evaluating bioinformatics visualizations". In: *Visualization and Computer Graphics, IEEE Transactions on* 11.4, pp. 443–456. DOI: 10.1109/TVCG.2005.53.

Satyanarayan, A. and J. Heer (2014). "Authoring Narrative Visualizations with Ellipsis". In: *Computer Graphics Forum* 33.3, pp. 361–370. DOI: 10.1111/cgf.12392.

Sauro, J. and E. Kindlund (2005a). "A Method to Standardize Usability Metrics into a Single Score". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '05. Portland, Oregon, USA: ACM, pp. 401–409. DOI: 10.1145/1054972.1055028.

— (2005b). "Using a Single Usability Metric (SUM) to Compare the Usability of Competing Products". In: *in Proceeding of the Human Computer Interaction International Conference (HCII 2005), Las Vegas, USA Making Sense of Usability Metrics: Usability and Six Sigma: (Sauro & Kindlund) p.10*.

Schaffer, D., Z. Zuo, S. Greenberg, L. Bartram, J. Dill, S. Dubs, and M. Roseman (1996). "Navigating Hierarchically Clustered Networks Through Fisheye and Full-zoom Methods". In: *ACM Trans. Comput.-Hum. Interact.* 3.2, pp. 162–188. DOI: 10.1145/230562.230577.

Sears, A. and J. Jacko (2009). *Human-Computer Interaction: Design Issues, Solutions, and Applications*. Human Factors and Ergonomics. Taylor & Francis.

Sedig, K., P. Parsons, and A. Babanski (2012). "Towards a Characterization of Interactivity in Visual Analytics." In: *JMPT* 3.1, pp. 12–28.

Segel, E. and J. Heer (2010). "Narrative Visualization: Telling Stories with Data". In: *Visualization and Computer Graphics, IEEE Transactions on* 16.6, pp. 1139–1148. DOI: 10.1109/TVCG.2010.179.

Shanmugasundaram, M. and P. Irani (2008). “The Effect of Animated Transitions in Zooming Interfaces”. In: *Proceedings of the Working Conference on Advanced Visual Interfaces*. AVI ’08. Napoli, Italy: ACM, pp. 396–399. DOI: 10.1145/1385569.1385642.

Shanmugasundaram, M., P. Irani, and C. Gutwin (2007). “Can Smooth View Transitions Facilitate Perceptual Constancy in Node-link Diagrams?” In: *Proceedings of Graphics Interface 2007*. GI ’07. Montreal, Canada: ACM, pp. 71–78. DOI: 10.1145/1268517.1268531.

Shannon, C. (1953). “The lattice theory of information”. In: *Information Theory, Transactions of the IRE Professional Group on* 1.1, pp. 105–107. DOI: 10.1109/TIT.1953.1188572.

Shi, K., P. Irani, and B. Li (2005). “An evaluation of content browsing techniques for hierarchical space-filling visualizations”. In: *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*, pp. 81–88. DOI: 10.1109/INFVIS.2005.1532132.

Shneiderman, B. (1996). “The eyes have it: A task by data type taxonomy for information visualizations”. In: *Visual Languages, 1996. Proceedings., IEEE Symposium on*. IEEE, pp. 336–343.

Shneiderman, B. and C. Plaisant (2006). “Strategies for Evaluating Information Visualization Tools: Multi-dimensional In-depth Long-term Case Studies”. In: *Proceedings of the 2006 AVI Workshop on BEyond Time and Errors: Novel Evaluation Methods for Information Visualization*. BELIV ’06. Venice, Italy: ACM, pp. 1–7. DOI: 10.1145/1168149.1168158.

Skog, T., S. Ljungblad, and L. E. Holmquist (2003). “Between aesthetics and utility: designing ambient information visualizations”. In: *Information Visualization, 2003. INFOVIS 2003. IEEE Symposium on*, pp. 233–240. DOI: 10.1109/INFVIS.2003.1249031.

Smith, R. and L. Platt (1987). “Cover story:. Benefits of animation in the simulation of a machining and assembly line”. In: *Simulation* 48.1, pp. 28–30.

Sonderegger, A. and J. Sauer (2010). “The influence of design aesthetics in usability testing: Effects on user performance and perceived usability”. In: *Applied Ergonomics* 41.3. Special Section: Recycling centres and waste handling – a workplace for employees and users., pp. 403–410. DOI: http://dx.doi.org/10.1016/j.apergo.2009.09.002.

Soukup, T. and I. Davidson (2002). *Visual Data Mining: Techniques and Tools for Data Visualization and Mining*. Wiley.

Spence, R. (2001). *Information visualization*. Vol. 1. Springer.

Stasko, J., D. McColgin, T. Miller, C. M. Plaue, and Z. Pousman (2005). “Evaluating the infocanvas peripheral awareness system: A longitudinal, in situ study”. In:

Stasko, J. and E. Zhang (2000). “Focus+context display and navigation techniques for enhancing radial, space-filling hierarchy visualizations”. In: *Information Visualization, 2000. InfoVis 2000. IEEE Symposium on*, pp. 57–65. DOI: 10.1109/INFVIS.2000.885091.

Statista (2016). *Leading social networks worldwide as of April 2016, ranked by number of active users (in millions)*. URL: http://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/.

Steele, J. and N. Iliinsky (2010). *Beautiful Visualization: Looking at Data through the Eyes of Experts*. " O’Reilly Media, Inc."

Stromer-Galley, J. (2004). “Interactivity-as-Product and Interactivity-as-Process”. In: *The Information Society* 20.5, pp. 391–394. DOI: 10.1080/01972240490508081. eprint: http://dx.doi.org/10.1080/01972240490508081.

Tergan, S.-O., T. Keller, and R. A. Burkhard (2006). “Integrating knowledge and information: digital concept maps as a bridging technology”. In: *Information Visualization* 5.3, pp. 167–174.

Thomas, J. and K. Cook (2006). “A visual analytics agenda”. In: *Computer Graphics and Applications, IEEE* 26.1, pp. 10–13. DOI: 10.1109/MCG.2006.5.

Thudt, A., U. Hinrichs, and S. Carpendale (2012). “The Bohemian Bookshelf: Supporting Serendipitous Book Discoveries Through Information Visualization”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’12. Austin, Texas, USA: ACM, pp. 1461–1470. DOI: 10.1145/2207676.2208607.

Tory, M. and T. Moller (2004). “Rethinking Visualization: A High-Level Taxonomy”. In: *Information Visualization, 2004. INFOVIS 2004. IEEE Symposium on*, pp. 151–158. DOI: 10.1109/INFVIS.2004.59.

Tractinsky, N., A. Katz, and D. Ikar (2000). “What is beautiful is usable”. In: *Interacting with Computers* 13.2, pp. 127–145. DOI: 10.1016/S0953-5438(00)00031-X. eprint: http://iwc.oxfordjournals.org/content/13/2/127.full.pdf+html.

Tufte, E. (1983). *The visual display of quantitative information*. Vol. 2. Graphics press Cheshire, CT.

— (1990). *Envisioning Information*. Cheshire, CT, USA: Graphics Press.

— (1997). *Visual explanations: images and quantities, evidence and narrative*. Vol. 36. Graphics Press Cheshire, CT.

— (2006). *Beautiful evidence*. Vol. 1. Graphics Press Cheshire, CT.

Tversky, B., J. Bauer Morrison, and M. Betrancourt (2002). “Animation: can it facilitate?” In: *International Journal of Human-Computer Studies* 57.4, pp. 247–262. DOI: 10.1006/ijhc.2002.1017.

Vande Moere, A. and H. Purchase (2011). “On the Role of Design in Information Visualization”. In: *Information Visualization* 10.4, pp. 356–371. DOI: 10.1177/1473871611415996.

Vande Moere, A., M. Tomitsch, C. Wimmer, C. Boesch, and T. Grechenig (2012). “Evaluating the Effect of Style in Information Visualization”. In: *Visualization and Computer Graphics, IEEE Transactions on* 18.12, pp. 2739–2748. DOI: 10.1109/TVCG.2012.221.

Viégas, F., S. Golder, and J. Donath (2006). “Visualizing Email Content: Portraying Relationships from Conversational Histories”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’06. Montréal, Québec, Canada: ACM, pp. 979–988. DOI: 10.1145/1124772.1124919.

Viégas, F. and M. Wattenberg (2006). “Communication-minded visualization: A call to action”. In: *IBM Systems Journal* 45.4, p. 801.

Viégas, F. and M. Wattenberg (2007). “Artistic Data Visualization: Beyond Visual Analytics”. English. In: *Online Communities and Social Computing*. Ed. by D. Schuler. Vol. 4564. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 182–191. DOI: 10.1007/978-3-540-73257-0_21.

Viégas, F. and M. Wattenberg (2008). "TIMELINES: Tag Clouds and the Case for Vernacular Visualization". In: *interactions* 15.4, pp. 49–52. DOI: 10.1145/1374489.1374501.

Viégas, F., M. Wattenberg, and J. Feinberg (2009). "Participatory Visualization with Wordle". In: *Visualization and Computer Graphics, IEEE Transactions on* 15.6, pp. 1137–1144. DOI: 10.1109/TVCG.2009.171.

Viégas, F., M. Wattenberg, F. van Ham, J. Kriss, and M. McKeon (2007). "ManyEyes: a Site for Visualization at Internet Scale". In: *Visualization and Computer Graphics, IEEE Transactions on* 13.6, pp. 1121–1128. DOI: 10.1109/TVCG.2007.70577.

Waldner, M., M. Le Muzic, M. Bernhard, W. Purgathofer, and I. Viola (2014). "Attractive Flicker: Guiding Attention in Dynamic Narrative Visualizations". In: *Visualization and Computer Graphics, IEEE Transactions on* 20.12, pp. 2456–2465. DOI: 10.1109/TVCG.2014.2346352.

Ware, C. (2004). "Information Visualization: Perception for Design (Interactive Technologies)". In:

Wattenberg, M. (2002). "Arc diagrams: visualizing structure in strings". In: *Information Visualization, 2002. INFOVIS 2002. IEEE Symposium on*, pp. 110–116. DOI: 10.1109/INFVIS.2002.1173155.

— (2005). "Baby names, visualization, and social data analysis". In: *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*, pp. 1–7. DOI: 10.1109/INFVIS.2005.1532122.

Wattenberg, M. and J. Kriss (2006). "Designing for social data analysis". In: *Visualization and Computer Graphics, IEEE Transactions on* 12.4, pp. 549–557. DOI: 10.1109/TVCG.2006.65.

Wehrend, S. and C. Lewis (1990). "A problem-oriented classification of visualization techniques". In: *Proceedings of the 1st Conference on Visualization'90*. IEEE Computer Society Press, pp. 139–143.

Werbach, K. (2014). "(Re) Defining Gamification: A Process Approach". In: *Persuasive Technology*. Springer, pp. 266–272.

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Use R! Springer New York.

Wijk, J. van (2005). "The value of visualization". In: *Visualization, 2005. VIS 05. IEEE*, pp. 79–86. DOI: 10.1109/VISUAL.2005.1532781.

Wikipedia (2016). *Wikipedia: Size of Wikipedia*. URL: https://en.wikipedia.org/wiki/Wikipedia:Size_of_Wikipedia.

Willett, W., J. Heer, J. Hellerstein, and M. Agrawala (2011). "CommentSpace: Structured Support for Collaborative Visual Analysis". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '11. Vancouver, BC, Canada: ACM, pp. 3131–3140. DOI: 10.1145/1978942.1979407.

Wohlfart, M. and H. Hauser (2007). "Story telling for presentation in volume visualization". In: *Proceedings of the 9th Jointt Eurographics / IEEE VGTC Conference on Visualization*. EUROVIS'07. Norrköping, Sweden: Eurographics Association, pp. 91–98. DOI: 10.2312/VisSym/EuroVis07/091-098.

Wojtkowski, W. and G. Wojtkowski (2002). "Storytelling: its role in information visualization". In: *European Systems Science Congress*. Citeseer.

Wood, J., H. Wright, and K. Brodie (1997). “Collaborative visualization”. In: *Visualization ’97., Proceedings*, pp. 253–259. DOI: 10.1109/VISUAL.1997.663890.

Xiong, R. and J. Donath (1999). “PeopleGarden: Creating Data Portraits for Users”. In: *Proceedings of the 12th Annual ACM Symposium on User Interface Software and Technology*. UIST ’99. Asheville, North Carolina, USA: ACM, pp. 37–44. DOI: 10.1145/320719.322581.

Yi, J. S., Y. a. Kang, J. Stasko, and J. Jacko (2007). “Toward a Deeper Understanding of the Role of Interaction in Information Visualization”. In: *Visualization and Computer Graphics, IEEE Transactions on* 13.6, pp. 1224–1231. DOI: 10.1109/TVCG.2007.70515.

— (2008). “Understanding and Characterizing Insights: How Do People Gain Insights Using Information Visualization?” In: *Proceedings of the 2008 Workshop on BEyond Time and Errors: Novel evaLuation Methods for Information Visualization*. BELIV ’08. Florence, Italy: ACM, 4:1–4:6. DOI: 10.1145/1377966.1377971.

Zeleny, M. (1987). “Management support systems: Towards integrated knowledge”. In: *Human systems management* 7, pp. 59–70.

Zhao, S., M. McGuffin, and M. Chignell (2005). “Elastic Hierarchies: Combining Treemaps and Node-Link Diagrams”. In: *Information Visualization, IEEE Symposium on*, p. 8. DOI: 10.1109/INFOVIS.2005.12.

Zhao, Z., R. Marr, and N. Elmqvist (2015). *Data Comics: Sequential Art for Data-Driven Storytelling*. Tech. rep. Technical Report. Human Computer Interaction Lab, University of Maryland.

Ziemkiewicz, C. and R. Kosara (2008). “The Shaping of Information by Visual Metaphors”. In: *Visualization and Computer Graphics, IEEE Transactions on* 14.6, pp. 1269–1276. DOI: 10.1109/TVCG.2008.171.

Appendix I

Publications

A Typology for Data Visualization on the Web

Ana Figueiras
FCSH – Universidade Nova de Lisboa
{mail@anafigueiras.com}

Abstract

The need to visualize data has originated in the research field, where it has been a useful tool to the study of scientific problems. However, the truth is that data visualization is a great way to present data for any area dealing with information, because visually presented information is more appealing due to its use pictograms and colours and also more efficient in conveying large amounts of information. Throughout the years, there have been efforts to develop a classification for these visualizations, in order to provide a better understanding of this way to present data. There are many different classifications proposed, but none of them are complete.

This paper discusses and develops a typology for online data visualization and infographics. Such a typology will be relevant for a better understanding of what kinds of visualizations exist and to better identify in further research which elements compose a good visualization that pleases the public.

Keywords--- **Data visualization, typology, classification, genres, case study**.

1. Introduction

Today’s society has been transformed by the rise of the Information Age and mankind has reached a point where the amount of information generated is so large that making sense of it can be an overwhelming experience. Using visualization techniques to understand big sets of data is not just a prettier way to represent information, it has become essential. The reason for this is that [8] visualizing information makes it easier to understand, to scheme, to recognize patterns and to make inferences upon the data. Therefore, information graphics have been increasingly used in a wide variety of areas, from the research field, where it emerged, to new areas like journalism and advertising. Nowadays data visualization is everywhere and it is particularly popular on the Internet, where it is used as an eye catcher.

There is a whole panoply of ways to visualize data: some more traditional like tables, pie charts, and bar graphs, others that are closer to art or to mapping. Nevertheless, there are lots of creative and fascinating, ways to visualize data. On the Web these visualizations can explode in terms of creativity: they can be animated, interactive and multimedia. “Creating an infographic is no longer just a matter of making the data visual. Instead, it involves the creation of a tool to help understand that visual data by synthesizing it through play.” [4]

However, nowadays there is a misunderstanding about what an infographic should be and sometimes the creators spend too much time on looks and forget the real purpose of an infographic. Although it is important to have aesthetically appealing visualizations their main purpose is to inform and it is only when they excel at their main purpose that they reach true beauty. Visualizations that are not successful in providing access to the information are failed visualizations. [6]

Although infographics are very good at conveying huge sets of data, they can also be very useful to tell stories, and lately this potential has been hugely discussed. Introducing storytelling in the visualizations is a challenge and usually it is easier to associate the visualization with a story by having a link from one to the other than to incorporate the story on the visualization.

In order to be able to understand how to tell stories with data visualizations it is imperative to profoundly understand all the pieces that compose a visualization. Some work has already been carried out in this area. However a careful analysis of recent online visualizations revealed that the available classification schemes are not exhaustive enough to classify examples of visualizations that are being created nowadays and although we can try to fit a visualization on an existing category, forcing the visualizations into a category that does not fully correspond to its characteristics is not good in terms of research.

This paper discusses and presents a typology for data visualization on the web that aims to fill this lacuna that exists in this area of research. The proposed typology was elaborated through an empirical analysis and a comparative study of existing data visualizations. These examples were chosen through an extensive research of what is currently being done on online newspapers and magazines, blogs, scientific videos, visualization research websites, and even publicity campaigns, and more importantly what is popular and shared by Internet users.

The conception of this typology also required reviewing the related work already published around this theme and this related work served as a foundation for this new classification scheme. This new typology consists on eleven different types or genres of data visualization that are not mutually exclusive: Sequential Graphic, Slide Show, Chart/Diagram, Map, Tag Cloud, Model, Drawing, Video/Animation, Photograph, Poster, and Game.

For better understanding of this new typology in this paper, I also present eleven case studies that were selected for demonstrating the specificity of each genre.

2. Related Work

Edward Segel and Jeffrey Heer [2] identify seven basic genres of narrative visualization that are not mutually exclusive: magazine style, annotated chart, partitioned poster, flow chart, comic strip, slide show, and film/video/animation. Additionally any of these genres can also have messaging (in the form of headlines, captions, labels, and annotations) and/or interactivity. “There are many possible types and degrees of interactivity, though common forms in narrative visualization include navigation buttons, hover highlighting, hover details-on-demand, filtering, searching, drill-down, zooming, and time sliders.” [2]

Edward Segel and Jeffrey Heer also have an interesting classification for the experience the reader/viewer has while interacting with the visualization. They argue that most visualization does not fit the author-driven versus reader-driven dichotomy, and is somewhere in the middle. Therefore, they identified three categories for these visualizations: Martini Glass Structure, Interactive Slideshow, and Drill-Down Story. The Martini Glass visualization structure begins with an author-driven approach, initially using questions, observations, or written articles to introduce the visualization. Once the author’s intended narrative is complete, the visualization opens up to a reader-driven stage where the user is free to interactively explore the data. Segel and Heer called this structure a martini glass because a single path is given by the author of the visualization (stem), but this path gets wider and wider with the multiplicity of available paths that appear after the main story is told. This possibility to have multiple reading paths is made possible through reader-driven interactivity: linking, highlighting, filtering, etc. The Interactive Slideshow structure, on the other hand, has a more linear path. Segel and Heer [2] emphasize that this visualization genre follows the typical structure of a slideshow but also includes some interaction on each individual slide. The authors also pointed out that this type of structure allows the user to explore particular points of the overall visualization before moving forward on the author-driven part of the visualization. Finally they present the most reader-driven structure, The Drill-Down Story, which has a structure that enables the user to choose to explore different details of the story freely in spite of maintaining the user in a general framework. Although this type of visualization gives the user freedom to choose his favourite backstory, the author of the visualization still has the responsibility to point the user the possible paths.

Nichani and Rajamanickam offer four categories just for interactive graphics: narrative, instructive, explorative, and simulative. “Narratives are used for telling straightforward stories, instructives provide step-by-step directions to reach a single goal, exploratives allow the user to engage in their own processes of sense-making, and simulatives allow the reader to grasp the process of a system.” [7]

Ian Bogost, Simon Ferrari and Bobby Schweizer, in Newsgames: Journalism at Play, elaborate a categorization for interactive (or playable) infographics that can also be applied to non-interactive infographics. According to the authors [4] playable infographics can be graphs, sequential graphics, maps, and diagrams. This last categorization seems to be the most interesting although there is not much information in the book about how the authors chose this categories and what particular elements make the examples that they give belong to that particular category. The typology offered is, like Nichani and Rajamanickam's categories, only for interactive graphics, and they have an interesting way of terming as Playable Data. Their notion of Playable data versus Non-playable data will also be used in the typology proposed here.

3. A new data visualization typology

Since the classification schemes mentioned above were not broad enough to classify every example that was encountered I tried to create a classification that fits the highest number of possible cases that compose my body of work. The 200 visualization examples were analysed according to their narrative elements, reading/viewing order, visual elements, and interactive elements. In terms of narrative, the 8 elements that were found in the analysed visualizations were captions, annotations, introductory text, accompany article, text, title, audio narration, and video narration. The 27 visual elements identified were timeline, photograph, bar chart, pie chart, doughnut chart, line chart, bubble chart, area chart, histogram, network diagram, Venn diagram, tree diagram, object size representing quantities, map, bubble map, pictogram, drawing, speech balloon, model, table, logo, video, scale, exploded view, tag cloud, pyramid, and animation. In terms of interactivity the 16 elements that were present in the analysed visualizations were input box, user contribution, slide show, navigation buttons, scroll bar, objects reaction to mouse movement, search, filtering, zoom, click details, click highlight, hover highlight, hover details, link to external article, drag objects, and game mechanics.

Taking all these elements into account, a new typology, which is intended to be exhaustive, is proposed. The eleven different types or genres of data visualization are Sequential Graphic, Slide Show, Chart/Diagram, Map, Tag Cloud, Model, Drawing, Video/Animation, Photograph, Poster, and Game. The categories were based on some of the work that has already been carried out in this area and referred previously as related work. Still this categorization and the analysis of the examples have some subjectivity. These genres vary mostly in terms of visual and interactive elements that the genre has and are not mutually exclusive, being possible to combine genres to classify more complex visualizations. Each category will be explained in detail bellow.

A sequential graphic, a category also in the classification scheme of Ian Bogost et al., is a chronological graphic. This type of visualization can be a timeline, like BBC's *British History Timeline*, or a cause/effect kind of sequence, like the South Florida Sun Sentinel's *Make your own hurricane*. This kind of visualization is very useful to show the user events that are influenced by previous actions.

A Slide Show on the other hand has an order imposed by the author of the visualization but it is not necessarily a chronological sequence. It can be composed of photographs, information, or even charts and the only interaction that it provides to the user is moving forward or backwards. An example of a slide show is The Guardian's *Hurricane Katia visualization*.

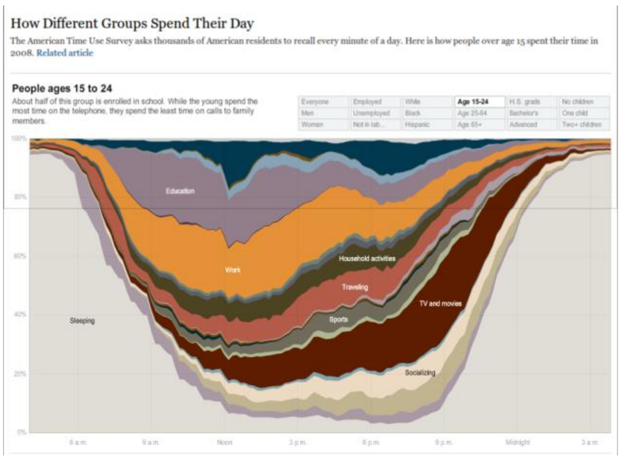


Figure 1 "How Different Groups Spend Their Day", a Chart/Diagram visualization by The New York Times

Chart/Diagram is a classic visualization used extensively on the media, on research, etc. Its heavy usage may be related to the fact that these visualizations are very easy to understand and the public is very familiar with them. This category includes every type of Chart or Diagram, from the common bar charts to the Venn diagrams. A visualization will only fit this category if the main focus of the visualization is the chart/diagram, because although there are many visualizations that include these visual elements as an added value, for example, on Foursquare's video/animation *A Week of Check-ins on the Path to One Billion* most of them have an element that is more prominent than the charts and diagrams and these are used just to give extra information to the user.

A Map is also a classic visualization. It can be tangible (it represents where things are placed and tries to mimic as truthfully as it cans the real world) or like Minard's *Carte figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813* an intangible map that represents not only information about physical places but also about events that occur on those places. In this map Minard captured not only geographical information, like the direction taken by the army as they travelled and the location the troops passed through, but also non-geographical information, like the size of the army as troops died from hunger and wounds, and the temperatures they experienced. Intangible maps can be fictional maps or interpretations of the world that do not resemble a traditional map. As in the previous category, sometimes visualizations have a map but this is not the main element, therefor that visualization cannot be considered a Map.

Tag Clouds are very popular online, being sometimes used on blogs and websites as navigation using hyperlinks. This type of visualization is a representation for text data, more specifically keywords or tags. Tag Clouds are useful to show which words occur more often, the size of the word being the differentiating factor. However, as Jacob Harris mentions, tag clouds can lead to fake conclusions and can be harmful [5]:

> *When looking at the word cloud of the War Logs, does the equal sizing of the words "car" and "blast" indicate a large number of reports about car bombs or just many reports about cars or explosions? How do I compare the relative frequency of lesser- used words? Also, doesn't focusing on the occurrence of specific words instead of concepts or themes miss the fact that different reports about truck bombs might be use the words "truck," "vehicle," or even "bongo" (since the Kia Bongo is very popular in Iraq)?*

A Model is a more technical visualization. Nowadays it is getting more and more popular because there are more people with the expertise to create 3D models. This visualization is particularly good to show projects of buildings or to describe complex processes.

A Drawing is a type of visualization that combines information and illustration. It is a very popular visualization on the printed press and it is quite common online too. In order to be effective and become a data visualization and not a mere drawing, this type of visualization has to combine the illustration with another type of visualization like, for example, charts/diagrams or even videos, like in the New York Times visualization *Three Generations of a Family Under One Roof*.

The Video/Animation is obviously the category of the types of visualization in which there is a video or animation that is the main part of the visualization. As in the previous category this type of visualization has to include other types of visualizations in other to be considered a Video/Animation visualization.

Another category that depends on other types of visualizations in order to be considered one is the Photograph. This is the least common type of visualization. One example of this type of visualization that was analysed in this investigation, and therefore one of the case studies in this paper, is the *100 Years of World Cuisine* visualization.

The Poster genre is inspired on the Partitioned Poster category by Edward Segel and Jeffrey Heer. These are usually static visualizations that include both textual and graphic elements, although it may be either wholly graphical or wholly text. Since posters are typically both eye-catching and informative, it is commonly used to advertise products.

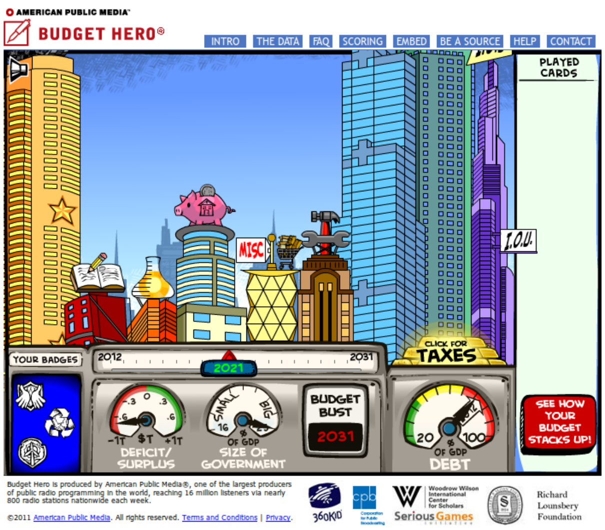


Figure 2 "Budget Hero" a Game visualization by American Public Media

The Game category might be the least common of all categories but it is certainly a type of visualization that is very appealing to the public. According to Ian Bogost, Simon Ferrari and Bobby Schweizer, "even if they are not games quite like Pac-Man or The Sims, infographics can become game-like, exploiting the properties of games in numerous ways: to encourage the manipulation of information for replayability, to allow pleasurable engagement with a system, or to invite exploration." [4] Although there are not many visualizations of this type there are some interesting examples like *SPENT*, one of the case studies in this proposal.

4. Case Studies

To understand what different kinds of visualizations exist and to try to understand what makes a visualization a Sequential Graphic, Slide Show, Chart/Diagram, Map, Tag Cloud, Model, Drawing, Video/Animation, Photograph, Poster, or Game, eleven examples were gathered for a more exhaustive analysis and classification. These examples come from very different sources such as online journalism web sites, and specialized blogs: The New York Times, CNN, The Guardian, National Geographic, Information is Beautiful, etc.

4.1. UK riots timeline

Between 6 and 10 August 2011, England suffered widespread rioting that started with the shooting of Mark Duggan in Tottenham. For better understanding the sequence of events The Guardian created an interactive timeline that shows the most important incidents and how they spread over the different neighbourhoods. The animated timeline enables the user to scroll through the events, ordered by hour, and watch how the riots unfolded.

The pictograms along the timeline identify the subject of that important mark on the sequence of events (Riot, Police, Statement, Court, Fire, or Cleanup). In order to explain the events the timeline also includes an introductory text, captions, annotations, and it also links to external articles that enable the user to learn about single events. Therefore, taking into consideration the special categories that Segel and Heer have for the reader/viewer experience, this visualization would a Drill-Down Story. The user also gets extra details of the events when hovers the pictograms. Other interactive elements present in this visualization include hover highlighting, navigation buttons, and a scroll bar.

The *UK Riots Timeline*, according to the typology I am proposing is a sequential graphic and, although it leads the user to a sequential order of reading/viewing, it enables the user to explore the events in the order he/she intends.

4.2. The World of Seven Billion

National Geographic's *The World of Seven Billion* is a slide show type of visualization. However it also includes a map and several charts. Although this visualization is considered playable, the interaction the user has with the visualization is very limited: the user can click to see the introductory text and the navigation buttons are highlighted on mouse-hover. As most slide shows the order of reading/viewing is linear.

The charts are a very prominent part of this visualization and most of them are not really charts but chart-like objects representing quantities. There are pictograms representing the number of cars, of personal computers, of children, etc., divided by type of income.

In terms of narration elements this visualization, besides the introductory text, it has captions and annotations. This types of narrations are vital on a visualization that resources to images to expose the data.

The *World of Seven Billion* visualization only works because it only has four slides. Otherwise it would be boring for the user and he/she probably would not see every slide.

4.3. Death penalty statistics, country by country

Death penalty statistics, country by country is a visualization by The Guardian that accompanies an article about countries that maintain the death penalty. The visualization is composed mainly by diagrams, therefor fitting the chart/diagram category perfectly. The diagrams in this case are mainly bubbles of different sizes to represent proportions, but there are also man pictograms representing the number of individuals executed. On the bottom there is also a timeline representing the number of abolitionist countries in contrast with the number of executing countries, since 1991 till 2010. The timeline also uses the technique of using objects, in this case circumferences, to represent quantities.

Since this is a non-playable visualization it has no interactive elements. The use of some interactivity such as hover or click details would be useful on a visualization of this type, because it could add some interesting information making the data more meaningful. A link to particular execution stories would be very effective to increase the empathy between the reader/viewer and the data.

However, since this visualization accompanies an article it is understandable that there are not more narration elements. *Death penalty statistics, country by country* has only an introductory text and captions that indicate the small information like names of countries and dates.

4.4. Home and Away: Iraq and Afghanistan War Casualties

Most American media has done, since the beginning, an extensive coverage of the wars in Afghanistan and Iraq. With the immense amount of data that they were able to gather about the deaths of soldiers in both wars CNN and Stamen Design have launched *Home and Away: Iraq and Afghanistan War Casualties*, an aesthetically-pleasing interactive data visualization that enables viewers to track each trooper's birth places against a map with the location where they died in Afghanistan or Iraq. The visualization allows the audience to learn about soldiers not only from the U.S but also from other countries.

There are two separate lists of casualties, Afghanistan and Iraq, and the data can be browsed either by map or table. In the map view two parallel maps are presented: one of the whole world where the places of birth of the soldiers appear and another with either the map of Afghanistan or Iraq. Complementary graphics are provided along the bottom to show trends of age, location, and date of death and through these complementary graphics casualties can be searched using those criteria. It is also possible to search soldiers by name through a search box.

This data visualization allows viewers to click on the points that represent a soldier that died in one of these wars and learn more details about the life of each casualty on their profile page. The visualization is integrated with the iReport platform, CNN's user-generated news community, allowing family and friends of the deceased to tell their personal stories, share memories, and pay tribute. This visualization goes beyond news reporting, it is a platform for participation.

It is notable this possibility of social interaction that *Home and Away: Iraq and Afghanistan War Casualties* provides to the audience, since it provides more annotations that help building the story of each soldier life and death. The content produced by the family and friends of the deceased enrich the visualization, saving journalists a huge amount of work that would take probably years to do. These personal stories make the user feel more empathy with the data. Once a user clicks on a dot, that dot is no longer just a marking on a map or part of the data, it becomes a story. The additional navigation through the graphics of age and location also make viewers feel more connected to the data because people will instinctively click on those who were the same age or from the same town. This feeling of connection with the people that compose the visualization greatly enhances the viewers experience with the visualization as a whole. “The user identifies with stories the map traces, constructing relevant meaning from fragments.” [4]

In terms of visual elements *Home and Away: Iraq and Afghanistan War Casualties* has photographs of the deceased soldiers, small bar charts with their ages, places where they are from, and years of death. It also has small pictograms to indicate the zoom, and home. There are also speech balloons that pop up whenever the user clicks on one of the bubbles in the map and a simple animation that appears while the map is loading.

Since this is a playable visualization it has some interactive elements: navigation buttons, scroll bar, search, filtering, zoom, click-able details, highlighting and popping details while hovering, and the possibility to drag objects (in this case dragging the map to move it around). *Home and Away: Iraq and Afghanistan War Casualties* fits perfectly Ben Shneiderman’s Visual Information-Seeking Mantra [1]: Overview first, zoom and filter, then details-on-demand. It also respects most of his seven tasks of a good visualization (Overview, zoom, filter, details-on-demand, relate, history, and extract), although I believe that it fails on the last two.

Home and Away: Iraq and Afghanistan War Casualties also has important narration elements that imply that this visualization is a good example of a narrative visualization, a visualization that is able to tell stories with data. It has a title like most visualizations, captions, annotations, and an introductory text that helps the user to understand the main story (the wars in Afghanistan and Iraq).

In terms of reading/viewing order, CNN's visualization of the Iraqi and Afghan war is an event reporting via geographical infographic that is not intended to be experienced in any particular order and that it does not require the viewer to interact with the whole data, the reading/viewing path is completely driven by the user. Maps like this one encourage the viewer to explore the data, picking the level of detail he/she wants to know about, and constructing narratives as they go. Edward Segel and Jeffrey Heer characterize the structure of visualizations like this one as Drill-Down Story, because [2] it "presents a general theme and then allows the user to choose among particular instances of that theme to reveal additional details and backstories", it has a reader-driven approach.

Through the careful analysis of this data visualization it is possible to see that it fits the map category, although it is a specific kind of map, a bubble map, where the size of the bubble represents the number of deaths on that place. Instead of viewing the map the user can opt to view a table with the deceased soldiers, but this option of visualization it is clearly not the main way to visualize this, being only useful if the user is someone that is looking for a soldier in particular.

4.5. What Does China Censor Online?

What Does China Censor Online? is a simple tag cloud that only has a title and text, in this case mere disconnected words. In terms of visual elements and although it is not visible at first, this visualization has a map, because the shape that the words form is the map of China. This visualization is not playable, therefor has no interactive elements at all.

Although this kind of visualization is considered bad, since it has so many problems in giving the right emphasis to the data, *What Does China Censor Online?* works well for its purpose. Maybe this is because it was man made and not automatically generated like most tag or word clouds. However, this visualization would probably benefit if a new overlay of information was added to it through the use of interactivity. Consequently it would be way more effective as a visualization and not merely beautiful.

4.6. Ground Zero Now

The *Ground Zero Now* visualization is part of a huge article about 9/11, that is divided in: The Decade, That Day, War Abroad, War at Home, Remembrance, Rebuilding, Muslims Now, 9/11 State of Mind, and Portraits Redrawn. This New York Times visualization, part of the Rebuilding segment of this bigger article entitled The Reckoning, is at the same time a model and a video/animation. In terms of visual elements it has tree animation videos that, amongst other stuff, include a map of the Ground Zero area. Like it would be expectable in a model it has models of buildings, pictograms and drawings, in this particular case, engineering drawings. Additionally it also provides an expanded view to better understand how the irrigation system on the Ground Zero memorial will work. Although it is considered playable since it is a video/animation, Ground Zero Now does not have any interactivity, because clicking a play button is not considered proper interactivity.

Ground Zero Now can be considered a narrative visualization since it has many narration elements. The story is told not only by the captions, annotations, and video narration, but also by the article that accompanies the visualization. Nevertheless, the visualization can work on its own, and it did not need to be associated with an article.

Although normally videos have a linear narrative, the fact that this visualization is composed by three different videos and an accompanying article, it is possible to consider the ordering an user directed path because the user can choose what to read or view first. This visualization does not really fit any of categories that Segel and Heer have for the reader/viewer experience.

4.7. How Many Households Are Like Yours?

Following the article Baby Makes Four, and Complications, which tells the story an unconventional Brooklyn family composed by a woman, her son, her sperm donor and his lover, The New York Times published an interactive infographic for exploring different types of American households.

Upon entering the page of the visualization, the viewer is able to choose the primary residents of his/hers (or other) household to see how the entered household compares to the rest of Americas households. Animated pictograms are used to represent the elements chosen by the user to compose a household. The audience is first presented with pictograms that represent the set of primary residents (married couple; male/female unmarried partners; single male; single female; male unmarried partners; and female unmarried partners) that they chose and can then add secondary members of the household (child under 18; child over 18; child-in-law; foster child; parent or parent-in-law; siblings or siblings-in-law; grandchild; other relative; housemate or roommate; Roomer, boarder or lodger; and other non-relative) that will also be represented as pictograms. Complementary graphics, bar and area charts, are provided along the bottom to show the viewer how the number of households like the one he/she selected have changed over time, which races have more households of that kind, and how is the income of those households. The graphics update on the fly whenever the user adds or subtracts a household member.

How Many Households Are Like Yours? fits two of the eleven proposed genres, although one of the types plays a major role on the visualization than the other. The visualization can be considered a drawing, but as a form to add information it also includes charts, so it can also be considered a chart/diagram.

The *How Many Households Are Like Yours?* visualization is composed of many different narrative elements: title, captions, annotations, and an introductory text. However the narration element that really helps turning this visualization into a narrative visualization is the article that accompanies the visualization. Still the visualization can work on its own, and it didn't need to be associated with an article. Since there is a written article to introduce the visualization, *How Many Households Are Like Yours?* is considered to have a Martini Glass visualization structure, according to Segel and Heer's categories for structure.

The reading/viewing order, according to this new typology that this paper introduces, is directed by the user. However, the use of the question is an indicator of the author driven approach that this visualization have, the author wants the reader to try his household type first.

In terms of interactivity elements *How Many Households Are Like Yours?* Enable the user to click and hover details and filter the data. However the New York Times visualization has a main problem. One of the purposes of having data visualizations is so that people can do visual comparisons easily, but in this visualization the data can't be compared. Only one kind of household is displayed at a time, so there is no easy way to check which kind of household is more common, to compare the viewer's family with their friends' households, etc. Most of the percentages vary so little that the number does not really mean anything to the viewer. In addition, the pictograms end up having no use other than aesthetics, their size doesn't mean the number of American families that are like that, or their income, or anything. Edward Tufte's [3] first rule is "above all else show the data." The viewer must be given enough information in order to have answers to his/her questions, and he/she should not have to make a great effort to find the answer to their questions. The viewer must be able to filter the information but that filter shouldn't limit his/her access to all the other data.

4.8. A Week of Check-ins on the Path to One Billion

A Week of Check-ins on the Path to One Bilion is a map, video/animation type of visualization, created to promote Foursquare. Since it is a video it is considered playable, but it does not have interactivity, and its reading/viewing order is completely linear. In terms of narrative elements this visualization does not have much, just a title and captions.

Since it is a promotional video, the visualization includes the company logo. In terms of visual elements it also has three bar charts (North & South American; Europe, Middle East and Africa; Asia Pacific) with the types of locations of the check-in enabling the viewer to know which kind of check-in happens the most, three pie chart like clocks with different time zones, and obviously video, animation and a map.

4.9. 100 Years of World Cuisine

100 Years of World Cuisine is a shocking, gory photograph that shows the number of deaths on 25 conflicts that happened from 1915 till the present. This visualization is non-playable. The amount of deaths that happened in each conflict is represented by the quantity of blood. There is an additional pie chart to represent the amount of deaths by continent, a line chart to visualize the amount of deaths by decade, and another pie chart to represent the amount of deaths in the 20th century that appear of the visualization and the amount that does not appear.

In terms of narrative elements the 100 Years of World Cuisine visualization only has a title and caption. The order of reading/viewing is completely driven by the reader/viewer.

100 Years of World Cuisine is another visualization that would benefit a lot if it used interactivity. If it was possible to click on each container of blood present in the picture and have access to additional data or, even better, to a story, this visualization would be even more interesting and it would fully its purpose better: to create awareness to the number of people that die because of these conflicts.

4.10. How Local News Is Going Mobile Infographic

How Local News Is Going Mobile is a perfect example of a poster like visualization, where the charts and diagrams also play an important part. As a typical poster it conjugates information and graphic elements in order to be appealing and eye-catching.

In terms of visual elements this visualization uses drawings and pictograms, some of them even used to present data like an umbrella that is used as a pie-chart. There are also bar charts, doughnut charts, and the size of objects used to represent quantities.

In terms of narrative elements How Local News Is Going Mobile Infographic has an introductory text to introduce the reader/viewer to the topic, and captions and annotations that complement the graphics and drawings. The other of reading is optional, the user can choose to read it from top to bottom or just check the elements that are more appealing. As most posters this visualization is non-playable.

4.11. SPENT

SPENT is a Game style visualization launched in February 2011 by McKinney, an advertising agency, and the Urban Ministries of Durham, a private non-profit organization. The motto is simple, or not: Could you live on $1,000 a month? This game lets the user, in this case the player, make the everyday choices necessary to get by on a tight income. First the user has to choose a job, like a waiter at a restaurant or a temporary typist, and that means different rates of pay. Then the user is presented with everyday choices: what food to buy, whether or not to pay for the car insurance, etc.

In terms of visual elements this visualization uses a timeline for the user to see in which day of the month he/she is, pictograms, animations, and speech balloons. In addition to an introductory text, annotations and captions provide not only the information necessary for the user to navigate through the game, but also gives additional data like how many families choose not to go to the dentist because it is too expensive, etc. It is this amount of additional data that makes this not a just a game but a collection of information that is able to raise awareness to the problem of poverty.

The interactive elements are very important in this kind of visualization, because these elements are the key to transform the interaction of this visualization into a game. In addition to the typical game mechanics there are navigation buttons, scroll bars, details that pop-up on click, and details that can be highlighted. As in any game, the path is directed by the user.

Conclusions

In this paper, I present the conclusions that I was able to achieve through the exhaustive analysis of a wide corpus of 200 collected examples from specialized blogs, online journalism, advertising, scientific research, etc. This analysis highlighted patterns on visualizations that allowed me to identify distinct genres of visualizations and elements that compose them.

Understanding the reading/viewing order and the narrative, visual and interactive elements of the visualizations that have been done for only media and other online purposes is a vital part of understanding how we can in the future extend the potential of data visualization. How can we introduce more and more storytelling in visualizations? How can we make visualizations that are easier to understand and at the same time as aesthetically pleasing? Achieving the answer to these questions will allow us not only to create better visualizations but also to better study them and their impact on users. I consider that one of the building blocks to understand data visualization is to have a way to classify them. Without a good classification scheme it is impossible to develop further research in order to understand what types of visualization are better for each type of data and which ones the public likes or not.

Having a classification scheme does not give us a recipe to the perfect visualization and the perfect combination to make a Sequential Graphic, Slide Show, Chart/Diagram, Map, Tag Cloud, Model, Drawing, Video/Animation, Photograph, Poster, or Game, does not seem to exist. Probably we will never achieve this perfect tuning of all the elements that compose visualizations, but through the persistent study of visualizations we will be able to recognize what elements are essential to each type of visualization. We already know that there is no Map visualization without geographical references, or Game visualizations without game mechanics. However the research seems to point out that the mixture of various types of visualizations on a single visualization can be very interesting and may be the key to a more engaging experience for the user.

Also the type of visualization also depends on the type of data. “Choosing the appropriate genre depends on a variety of factors, including the complexity of the data, the complexity of the story, the intended audience, and the intended medium. There are clear cases in which a genre is more appropriate for a particular purpose.” [4] For a visualization to work it has to have the full package: it has to present the data efficiently, to permit the free exploration of the data (to filter it and to compare it), and to be visually appealing. “A “cool” visualization with a strong graphic design will just as readily spread as one that illuminates something fascinating and important about the data.” [1] A good balance of looks and content seems to be the best approach to achieve effective visualizations.

This analysis is important, not only as an attempt to provide a more exhaustive typology that can be of good use for future research, but also to identify possible combinations of elements that can improve certain types of visualizations. However, this investigation is too closely related to the examples used, therefor new visualizations will bring new variables to this study.

References

[1] B. Shneiderman. The eyes have it: a task by data type taxonomy for information visualizations. In *Visual Languages, 1996.* VL, IEEE, 336-343. September 1996.

[2] E. Segel and J. Heer. Narrative visualization: Telling stories with data. In *Visualization and Computer Graphics 2010.* IEEE Transactions,16.6, 1139-1148. November-December 2010.

[3] Tufte, E. (1983). The Visual Display of Quantitative Information. Graphics Pr, 2nd edition (May, 2001), USA.

[4] I. Bogost, S. Ferrari and B. Schweizer. Newsgames: Journalism at play. Cambridge, MA, MIT Press. 2010

[5] J. Harris (2011). Word clouds considered harmful. Nieman Journalism Lab. Retrieved from http://www.niemanlab.org/2011/10/word-clouds-considered-harmful/

[6] J. Steele and N. Iliinsky. Beautiful Visualization: Looking at Data Through the Eyes of Experts. Sebastipol, CA, O'Reilly. 2010

[7] M. Nichani and V. Rajamanickam. Interactive visual explainers - A simple classification. *Elearning post 2003*. url:http://www.elearningpost.com/articles/archives/interactive_visual_explainers_a_simple_classification

[8] T. Kapler and W. Wright. GeoTime Information Visualization. *Information Visualization, 2004.* INFOVIS 2004. INFVIS, IEEE, 25-32. October 2004.

How to tell stories using visualization

Ana Figueiras
FCSH - Universidade Nova de Lisboa
mail@anafigueiras.com

Abstract

Storytelling's benefits are long-known and its potential to simplify concepts, create emotional connection, and capacity to help retain information has been explored in different areas, such as journalism, education, etc. The necessity to incorporate storytelling in visualizations arises from the need to share complex data in a way that it is engaging. Advances in technology have enabled us to go beyond the traditional forms of storytelling and representing data, giving us more attractive and sophisticated means to tell stories.

In this paper, we present the results of a focus group study that was conducted with the purpose of collecting information about the narrative elements in a collection of visualizations and the possible inclusion of storytelling elements in those. In this study was also collected information about the visualizations in terms of comprehension, navigation, and likability with the intent of identifying appealing elements in the visualizations. Furthermore, we suggest strategies to include storytelling in visualizations.

Keywords—**Storytelling, narrative visualization.**

1 Introduction

Since the beginning of time man has used stories to entertain, educate, and instill moral values. Stories prevail in comparison to other forms of presenting information not only due to their power to help assimilate and retain information, but also because stories are in fact more compelling. Predating writing, storytelling is still more frequently associated with oral lore even though it is present in our everyday life in written form and also film and audio. However storytelling has come a long way from its traditional forms and is now a useful tool for education, information, entertainment, and other areas.

Like storytelling information visualization has come a long way beyond classical mathematical graphs and representations. Nowadays, new techniques allow access to information that could formerly not be included on traditional forms of visualization, mainly due to technological limitations. As it happens with storytelling the power of information visualization is also long known. After all, we all began our lives getting most of our information visually [8] and we are so familiarized with the process of interpreting visual representations that this has become one of our favorite ways to consume information.

All technological advancements arising from the emergence and development of computers and the Internet are not only benefiting the field of information visualization, enabling the limitations of textual and visual representation to be overcome, but also all the areas related to storytelling that are now able to infuse their traditional narratives with the sophisticated visualization techniques. Moreover, this marriage between visualization and storytelling is helping to feed information-hungry audiences that do not always want to engage in time-consuming activities such as reading a long article or watching a documentary.

Nowadays the audience is presented with colossal amounts of complex information that would be almost impossible to grasp resorting only to traditional forms of presenting information. Amid the chaos of information that we have nowadays, boosted by the open data trend, it is imperative to provide ways to make sense of information, which is often abstract, non-spatial, and sometimes even non-visual. Visualization has proved to be very effective in these scenarios [12], however introducing storytelling in this equation can not only help the interpretation of the visualization but also help the visualization become even more appealing. But how can we fuse visualization and storytelling? In what ways can the new forms of media transform and contribute to this? What strategies of storytelling in visualizations are more appealing to the public?

In order to achieve the answers to some of these questions we carried out an investigation that tries to shed a light on the narrative elements that could introduce storytelling in visualizations. We also tried to understand the preferences of the users regarding this topic. We conducted a focus group study in which participants looked at several visualization examples collected from various sources such as news media web sites, marketing initiatives, and specialized blogs. In this study, in addition to the general perception about the visualization, we also collected information about the visualizations regarding three specific dimensions: comprehension, navigation, and likability.

2 Related Work

Until recently visualizations have been used to support traditional forms of storytelling as extra information or supporting evidence [15]. Nonetheless, there has been a great effort lately to transform visualizations in an independent form of storytelling that can exist by itself without support of a traditional form of storytelling such as a video or text. The research about the ways of doing this is being carried out in various different areas but mainly in journalism [15, 3], an area in which there has been a great effort to create multidimensional stories composed of other media besides text, and information/knowledge visualization [8, 9, 10, 13, 4], two disciplines which were primarily focused on techniques of visualization but are now starting to research strategies to make visualizations that are independent eschew other types of narratives.

Storytelling is also enabling the approximation of two areas of research that have been historically separated and developed independently: information and knowledge visualization. With the introduction of storytelling the differences between these two areas are getting more and more blurred. According to Bertschi et al. "in order for information to transform into knowledge, one must share some context, some meaning, in order to become encoded and connected to preexisting experience" [1]. Storytelling is one of the tools that is capable of introducing context and meaning in the visualization, not only helping users to establish connections between the complex data represented,but also introducing an important component, particular to knowledge visualization, that is tacit knowledge. This category of knowledge, that encompasses things such as intuitions and subjective insights, is difficult to communicate [12]. However it is known that storytelling has the power to engage and make people relate, therefore possibly facilitating the sharing of tacit knowledge.

2.1 Narrative visualization

Information visualization is much more than a visual representation of data. It's the process of dissecting raw data that by itself has little meaning and presenting it in a way that it's no longer complex. Although it's not new visualization has blossomed with the emergence of the new media. "These new technologies truly allow us to do things we never could with paper, so we should expect it to take awhile to gain sufficient understanding of them before we can apply them as effectively as we would like" [8].

Moved by the rising of these new media Gershon and Page were the first to notice the valuable contribution that storytelling could give to Information visualization. However, according to Kosara and Mackinlay [13], they fail to describe actual visualization and focus mainly on map views without numerical data. In other words, they focus more on simple visual representation.

Later, in 2010, Segel and Heer[15] reinvented this notion of using storytelling in visualizations naming it narrative visualization. By studying the elements of existing visualizations Segel and Heer were able to identify some patterns and structures that news media uses to introduce storytelling in visualizations: Martini Glass Structure, Interactive Slideshow, and Drill-Down Story. The first structure begins with an author-driven approach and only once the author's intended narrative is complete, the visualization opens up to a reader-driven stage where the user is free to interactively explore the data. The Interactive Slideshow approach has a completely linear path with some interactivity within the limits of each slide. Finally there is the Drill-Down Story: completely reader-driven, allowing the user to choose any reading/viewing order possible.

2.2 The benefits of storytelling in visualization

Using narrative elements in visualizations often help create a structured interpretation path that usually does not exist in traditional information visualization [4]. Without storytelling visualizations are not able to give explanations about the subject and depend too much on the audience's ability to interpret the data correctly.

Moreover they can be entirely independent of other means of storytelling, being able to get the point across easily and in sufficient detail for the audience to understand it. The bits of storytelling in these visualizations don't need to be over informative and descriptive because the audience is able to fill in "the gaps in the story with their imagination, experiences, and expectations" [16]. Storytelling can be introduced through the use of persuasive/rhetorical techniques and exploratory/dialectic strategies [9].

All of these elements make narrative visualization pleasing, not only because it doesn't require a lot of time and effort to assimilate the information but also because it sparks the audience's curiosity and transforms the task of acquiring information into a fun activity. According to Ma et al. [14], "they leave a lasting impression, either by piquing the audience's curiosity and making them want to learn more or by conveying a deeper meaning than your everyday run-of-the-mill sequence of causally related events."

However there has been an increasing concern with how much the incorporation of narrative will impact the exploration of the data and whether or not this will distract the user from the data [4]. Although having a direction will help users that are less familiar with the subject an undirected exploration can help proficient users find new interpretations of the data and even discover meanings that were not foreseen by the creators of the visualization. It's important to understand every building block of the visualization in order to create a narrative that doesn't overpower the data.

Research [4] has revealed that having flexible narratives that point out particular landmarks for the user to explore, still allowing the user to freely the in-between landmarks, is a good option. Nonetheless there is still research to be done on how the narrative influences the interpretation process [9] and how to effectively create these flexible narratives. More research is needed for understanding which rhetorical techniques can be used and if it's possible to build a set of techniques that works for different sets of data.

3 Focus group

The focus group was conceptualized and structured as way to gather information about factors such as comprehension, likability, and navigation. This method was used because it fosters the discussion between the participants and enables us to easily obtain qualitative and affective information from participants. This focus group sessions were conceptualized as an exploratory exercise to obtain an emotional response from the participants, an evaluation that could not be done using a survey.

3.1 Procedures

The study took place at *Universidade Nova de Lisboa* and used the focus group method to collect data. The location for the sessions was a room with a computer with Internet connection for each participant at the *Faculdade de Ciências Sociais e Humanas* campus. There were a total of 16 participants, divided into 2 groups of 8 elements each, and no personal information about them was collected.

The source of the participant population was university students, either from the New Media or the Communication masters program, that were willing to participate in this research. They were invited to take part in the focus group study by their teachers. Although the participants belong to the same cohort, they have different backgrounds in the area of communication: journalism, design, marketing, etc. This variance of backgrounds guarantees variance of responses also ensuring that they had enough previous knowledge to fully understand the stimulus provided during the sessions. The groups were also mixed in terms of gender and age. The focus group sessions were done on the first day of classes therefor most participants did not knew each other previously to the experiment. This fact reduces the chances of the participants being influenced by the other participants opinion.

The focus group session began with a standard introduction and explanation of the purpose of the research. The participants received a rating sheet to rate the visualizations and had a computer where they could see and interact with the visualizations. The moderator asked the participants to rate, from 1 to 10, each visualization presented immediately after they interacted with and before the discussion about that visualization started. The rating was not done as a group activity because we wanted them to be independent of the group-think that could be generated by the discussion. The visualizations were rated in terms of comprehension (Was the information presented in a clear, comprehensible way? Was the purpose easy to understand?), likability (Was the visualization interesting and fun? Was the interaction with it enjoyable?), and navigability (Was the data easy to navigate? Was it clear how you were supposed to interact wth it?). These sheets were returned to the moderator at the end of the session.

The participants were given a few minutes to explore and interact with the first visualization and afterwards the moderator asked some semi structured questions about it. The process was repeated for each visualization shown. These questions were asked in order to start the discussion between the participants. They were asked to explain their answer and to provide their views about the visualizations, discussing them with the other participants. The focus group was recorded for record keeping and their answers were later transcribed.

The moderator also asked some questions regarding 2 or more visualizations at the same time: Which of the two do you prefer? Which do you think is more visually attractive? Which do you think tells the story better? The questions about comparisons were asked using similar visualizations or with comparable topics.

Sequential Graphic	Evolution of the Web	Introductory Text
Map	Home and away	
	Death penalty in the US mapped	
Poster	How Local News Is Going Mobile	
Photograph	Faces of the dead	
	British Troops Killed in Afghanistan	
Tag Cloud	What does china censor online?	Little or No Text
Drawing	How Much CO2?	
	How Many Households Are Like Yours?	Accompanying Text
Chart/Diagram	Death penalty statistics, country by country	
Model	Ground Zero Now	

Figure 1: Characteristics of the visualizations presented to the focus groups' participants

The discussions took about one hour and both groups were shown eleven examples of visualizations of different types and with different characteristics (Playable, Non-Playable, Introductory text, Accompanying article, and Audio narration). The group was shown the examples on Figure 1. Only 3 of the visualizations, marked on the figure with a dark gray background, are non-playable. These examples were chosen through a previous research [7] of what is currently being done on online newspapers and magazines, blogs, scientific videos,visualization research websites, and even publicity campaigns, and more importantly what is popular and shared by Internet users.

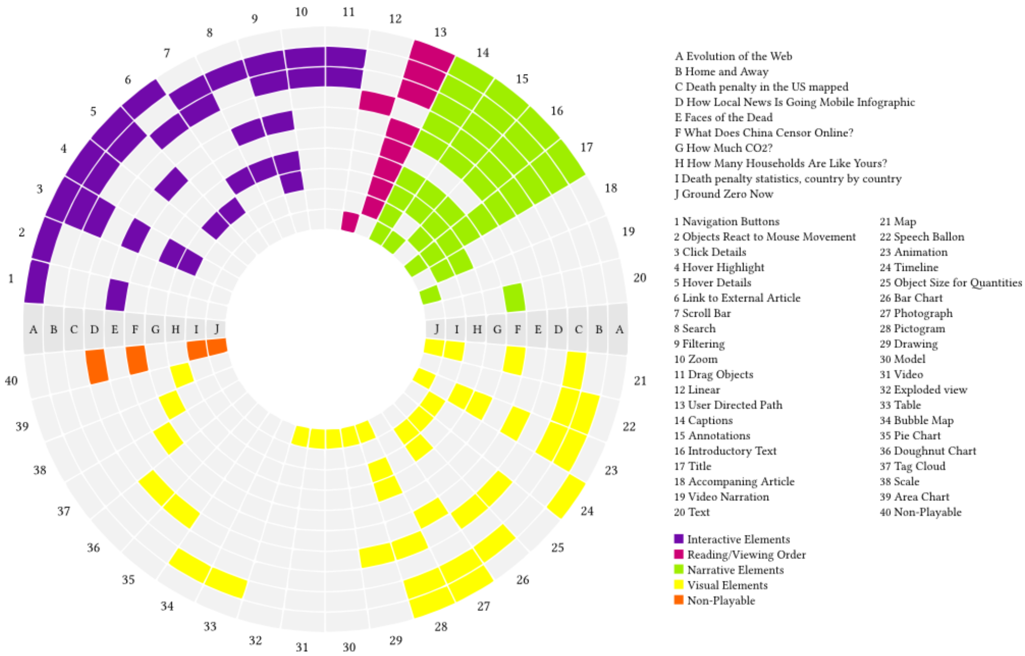


Figure 2: Visualizations used in the focus group study and the elements that compose them

As we present in Figure 2 most visualizations share common elements, specially in terms of Narrative. Most visualizations chosen for this study also share the same reading/viewing order, *User Directed Path*, however some examples with a *Linear* reading/viewing order were also shown in order for the participants to state which type they preferred. In terms of Interactive Elements most of the visualizations analyzed share two types of interaction: *Click Details* and *Filtering*. Every other element (*Navigation Buttons, Hover Highlight, Hover Details, Link to External Article, Scroll Bar, Search, Zoom,* and *Drag Objects*) is common to at least two visualizations, except for the Interactive Element *Objects React to Mouse Movement*. This fact occurs because this type of Interactive Element is less common and did not really influenced the example too much.

3.2 Results

The participants had very strong opinions about each of the visualizations and, with few exceptions, most visualizations received high scores in terms of comprehension, likability, and navigation (see Figure 3).

3.2.1 The most discussed

Two visualizations caused a lot of discussion and achieved high scores in all of the categories (comprehension, likability, and navigation): the New York Times' "How many households are like yours?"[1] and The Guardian's "Death penalty in the US mapped".[2]

The first had the most participants giving it a score of 10 in terms of comprehension and likability. It was also the visualization for which more participants gave a 9 in likability and navigation. This visualization also had the highest average rating (8.3 for comprehension, 8.1 for likability, and 8.2 for navigation) and no individual ratings under 4. This interactive visualization for exploring different types of American households is associated with the article "Baby Makes Four, and Complications." It allows the audience to choose the primary residents of a type of household to see how it compares to other American households. There are animated pictograms used to represent the elements chosen by the user to compose a household: the primary residents (married couple; male/female unmarried partners; single male; single female; male unmarried partners; and female unmarried partners) and the secondary members of the household (child under 18; child over 18; child-in-law; foster child; parent or parent-in-law; siblings or siblings-in-law; grandchild; other relative; housemate or roommate; Roomer, boarder or lodger; and other non-relative). Complementary graphics that update on-the-fly are provided along the bottom to show how the number of households like the one selected have changed over time, which races have more households of that kind, and what is

[1] http://www.nytimes.com/interactive/2011/06/19/nyregion/how-many-households-are-like-yours.html
[2] http://www.theguardian.com/news/datablog/interactive/2011/sep/21/death-penalty-us-map

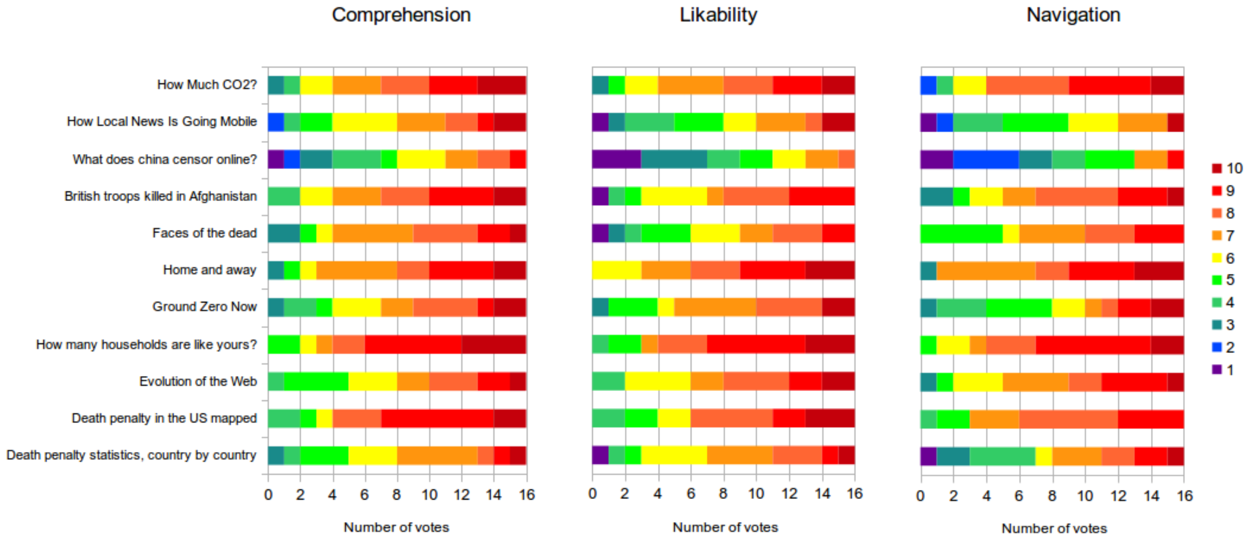


Figure 3: Scores given by the participants to each visualization in terms of comprehension, likability, and navigation

the income of those households.

The Guardian's visualization consists of a map where the states are colored according to the number of executions. When we click on a state, a bubble pops up with further information for different years. This interactive is also part of a larger article. "Death penalty in the US mapped" had the same number of participants giving it a 10 in terms of likability but had more participants giving it a 7 in comprehension. On average, the visualization ranked 3rd in terms of comprehension and likability and 4th in terms of navigation. The overall scores were not as high as the ones given to the New York Times' visualization.

In the sessions the "Death penalty in the US mapped" visualization was compared with a visualization of a similar subject: "Death penalty statistics, country by country"[3]. This visualization by the Guardian accompanies an article about countries that maintain the death penalty. The visualization is composed of a map with bubbles of different sizes that represent proportions and man pictograms representing the number of individuals executed. On the bottom there is a timeline representing the number of abolitionist countries in contrast with the number of executing countries, since 1991 till 2010.

The participants were very vocal about the these visualizations mainly because the first was playable and the second wasn't. Most participants preferred the one that allowed interactivity even though they noticed that the other gave more information on the subject. Some participants stated that the fact that the information was immediately presented caused confusion and they would prefer if it only showed that information on click.

Another popular example among the participants (having an average of 7.6 for comprehension, 8.1 for likability, and 7.9 for navigation) was CNN's "Home and away: Iraq and Afghanistan war casualties"[4] visualization. This map visualization is composed of two maps where the audience can find birth place of a trooper that died either in Afghanistan or Iraq and relate it the location where he/she died. By clicking on the points that represent a fallen soldier the audience can learn more details about their life on their profile page. There are also complementary graphics at the bottom to show trends of age, location, and date of death. The data can be navigated through these complementary graphics or through a search box. This visualization, along with "Death penalty in the US mapped", had the highest number of participants giving it a 10 in terms of likability and had a higher number of participants giving it a 10 for navigation.

3.2.2 Links as context

The "How Much CO2?"[5] visualization done by David McCandless for GE was also one of the most popular visualizations among the participants. This visualization ranked

[3] http://www.theguardian.com/news/datablog/2011/mar/29/death-penalty-countries-world
[4] http://edition.cnn.com/SPECIALS/war.casualties/
[5] http://visualization.geblogs.com/visualization/co2/

2nd in terms of comprehension and likability and 3rd in terms of navigation. “How Much CO2?” is an interactive visualization that shows the amount of carbon that different activities, entities, or events emit. Although the participants liked this visualization they thought that something was missing: text. One of the participants stated that if instead of the visualization just showing that web searches in the US produce 5,019 tons of CO2 he would prefer that they also explained how this happens. According to him “it is lacking support stories.”

Another visualization that the participants said it suffered from the lack of storytelling was the “Faces of the dead”[6]. This visualization ranked 4th in terms of comprehension and likability and 5th in terms of navigation. It was also the one that had a higher number of participants giving it an 8 in terms of comprehension. This visualization consists in a picture of a soldier that died in Iraq or Afghanistan formed by little squares that represent other soldiers that also died in these wars. Every time we click on one of the squares the big picture becomes the photo of that soldier that we clicked on. One of the participants said that the this visualization looks more like a work of digital art, but does not give enough information to become interesting as a visualization. Furthermore the fact that the pictures were in black and white made it more difficult for the participants to relate with it. One of the participants even stated that because the features were blended by the lack of color the soldiers were too homogenized.

Most participants preferred another visualization with a similar topic that was compared to “Faces of the dead”: The Guardian’s “British troops killed in Afghanistan”[7] visualization. This visualization ranked 3rd in terms of comprehension and likability and 4th in navigation. The Guardian’s visualization also uses pictures of fallen soldiers but shows them in color and includes a link to the story of how that soldier died. Because the pictures were in color the participants felt closer to those people and felt that they could relate more to them. The fact that they could also read the story also contributed for that sense of closeness with the visualization. The participants felt that in this visualization the soldiers were no longer a number, a dot, or a square, they were real people.

One visualization that also plays on this idea of the stories appearing as support information is the“Evolution of the Web”[8] visualization. This was also one of the participants’ favorite visualizations. The average rating for this visualization was 6.8 for comprehension, 7.3 for likability, and 7.3 for navigation One of the participants said that this visualization was almost perfect. “First it is a timeline and timelines give this idea of story, of flow, and moreover we are automatically placed at present time, which is probably what we have the most interest, then being able to then navigate backwards,” he added. Other participant referred that she enjoyed the fact that the visualization didn’t present all the information at once but allowed to explore by clicking on links, and that sparked her curiosity.

3.2.3 Non-playable

The participants were also shown two other visualizations that did not have any kind of interactivity:“What does china censor online?”[9] and “How Local News Is Going Mobile: Could the iPad Be the New Sunday Press?”[10]. The first is a simple non-playable tag cloud in which the shape that the words form is the map of China. This visualization had the lowest overall scores. The average rating for this visualization was 5.2 for comprehension, 4.2 for likability, and 3.9 for navigation. One of the participants even stated that she took a long time to realize that there was not interactivity in the visualization and that she probably just did not understand where she was supposed to click. When inquired about what could have been done to this visualization, to make it more interesting, some participants responded that if on mouse hover additional information was shown that would already make the visualization a hundred times better. The overall feeling was that they did not learn anything with that visualization

The second non-playable visualization shown is a poster like visualization with drawings, pictograms, bar charts, and doughnut charts. Although the ratings given were completely bad (6.6 for comprehension, 5.8 for likability, and 5.3 for navigation) the participants were almost indifferent to the visualization “How Local News Is Going Mobile: Could the iPad Be the New Sunday Press?” One of the participants said that this visualization was not very stimulating and that after seeing visualizations with so much interactivity this one just looked even worst. The overall opinion was that they lost the interest in the visualization because there was nothing more to discover.

The participants were also shown a video based visualization and surprisingly, although the video had no interactivity, most participants showed a overall positive response to it. The “Ground Zero Now”[11] visualization is part of the Rebuilding segment of a huge article about 9/11 entitled The

[6] http://www.nytimes.com/interactive/us/faces-of-the-dead.html
[7] http://www.theguardian.com/world/interactive/2011/sep/20/british-troops-killed-in-afghanistan-interactive
[8] http://www.evolutionoftheweb.com/
[9] http://www.informationisbeautiful.net/visualizations/what-does-china-censor-online/
[10] http://old.columnfivemedia.com/work-items/how-local-news-is-going-mobile-infographic-could-the-ipad-be-the-new-sunday-press/
[11] http://www.nytimes.com/interactive/2011/08/30/us/sept-11-reckoning/ground-zero.html

Reckoning. The visualization has three animation videos that include models of buildings, pictograms and drawings, but does not have any interactivity. About the audio or video narration the opinions were divided and although some of the participants said that they enjoy this kind of storytelling the majority did not express that much interest in it.

4 Analysis

The focus group results can shed a light on what may be good strategies for storytelling in visualizations. Interactivity seemed to be the most important strategy. One the participants even said: "With a video you retain about 8 to 15% of the information, if the video has some interactivity that percentage skyrockets to about 70%. That says a lot about the power of interactivity."

Although the level of interactivity was pointed over and over again as an important feature another characteristic stood out. What made the New York Times' "How many households are like yours?" visualization be the overall favorite was the fact that it enabled the participants to relate with the visualization. "The fact that the visualization allows us to identify with the subject instantly sparks an additional interest. In fact, the article that is linked to the visualization is about a family that is so different from mine that probably wouldn't even catch my attention, but since I'm already interested in the subject because of the visualization I would probably read the article also."

4.1 Interactivity and its relation with context

The benefits of using interactivity in visualizations have been long known. According to Kosara and Mackinlay [13], "being able to not just see the data, but quickly change the view, add different data, etc., makes analyzing it much faster and more effective." However, Ma et al. [14] consider that should be carefully balanced otherwise the creator of the visualization loses control over how the story is being told.

The opinions collected in this focus group study seem to point out that the audience prefers short moments of storytelling that they can access if they feel the urge to do it than having a dense storytelling that they have to carefully follow. The participants showed a lot of interest in exploring the visualization freely and seem to prefer to be moved by their own curiosity.

Another interesting feedback regarding interactivity is that the participants believed that the visualizations they liked the least could probably be fixed with an overlay of information, presented through the use of interactivity. According to them, the use of some interactivity such as hover or click details would change these visualizations completely, because it would make the data more meaningful. The participants also stressed that they enjoyed when the visualizations provided links to other content and that this fact doesn't prevent them from returning to the visualization. Although none of the examples shown had this feature, when confronted with the possibility of having external links, for example for Wikipedia pages, the participants referred that they would value this option.

Some research about how information visualizations with annotations are a promising way to complement articles has already been done [11]. These annotations have the capacity to add context that otherwise would be very difficult to provide, easing the user's interpretation and suggesting conclusions. Hullman et al. [11] developed an approach to automatically generate these annotations: Contextifier. In this case the narrative visualizations created were meant to accompany online news.

It would be interesting to see how the participants would react to more game like visualizations. After all computer games are the most popular example of interactive storytelling [13] and maybe their form of storytelling can be successfully replicated in visualizations.

SPENT[12] is one of the few attempts of merging visualization and games. It was launched in February 2011 by McKinney and the Urban Ministries of Durham. The objective is for the audience to understand if it is possible to live on $1,000 a month. The game lets the player make the everyday choices necessary to get by on a tight income: choosing a job, food to buy, pay the for the car insurance or take the son to the dentist, etc. SPENT has a timeline for the player to see in which day of the month he/she is, pictograms, animations, and speech balloons. Once in a while it also gives additional data like how many families choose not to go to the dentist because it is too expensive, etc.

This additional data makes SPENT more than just a game, but is this enough to transform it into a visualization? According to Bogost et al. [2], "even if they are not games quite like Pac-Man or The Sims, infographics can become game-like, exploiting the properties of games in numerous ways: to encourage the manipulation of information for replayability, to allow pleasurable engagement with a system, or to invite exploration."

Nicholas Diakopoulos [5] points out that there are several challenges with the gamification of information graphics, specially when these deal with data that is variable through time (updated, refreshed, dynamic). He compares *game-y* information graphics (information graphics that include formal elements of games such as rules, goals, scores, competition, and the notion of "winning" [4]) with traditional games, which usually benefit from a carefully developed design component and consequently take a long time to be released.

Most of the *game-y* information graphics, or playable

[12] http://playspent.org/

infographics (the alike term coined by Bogost et al. [2]), have been produced by news media (for example "Budget Hero"[13] , created by American Public Media) or marketing initiatives ("SPENT" by McKinney). The fact is that these organizations depend on deadlines and usually cannot invest too much time developing these types of visualizations. However the level of gamification used doesn't need to be as complex as in commercial computer games, nor does the data used need to be ever changing as it happened with Salubrious Nation by Diakopoulos et al. [6]. Specially in news media the information has a short life cycle and is often preferable to have a stable visualization than to have one that will forever be up-to-date.

Diakopoulos et al. [6] conducted one of the few researches about the reaction to more game-like visualizations. They tested a *game-y* version of Salubrious Nation against a *Non-Game-y* version of the same infographic. The *game-y* version of Salubrious Nation uses geographically tagged public health data to create a game where the goal is to accurately guess the extent of the given health parameter for a randomly selected target county. The data helps the player to do an informed guess. The *non-game-y* version of Salubrious Nation is still interactive but does not include the guessing game component.

The authors have concluded that sometimes the game features takes the attention away from the actual data. However it can also "successfully motivate interactions and cause users to explore and bias both the exploration of parameters and the nature of insights in interesting ways." [5] They were also able to conclude that *game-y* infographics are as enjoyable as a non-game infographics and that it might be useful in helping to structure the interaction [4]. Further research can provide insights on which types of gamification limit the users attention deficits or even play with this fact in order to channel the users attention to particular data.

4.2 Empathy and Temporality

Another issue that kept coming up in the focus group discussion was empathy. The participants kept talking about how much they related or not to the visualizations' subjects, sometimes not even knowing exactly why.

Kosara and Mackinlay [13] asked in "Storytelling: The Next Step for Visualization" what makes a visualization memorable. Everything seems to point out exactly to what they refer to as a possible cause, that the visualization is memorable when people relate to it.

One of the best examples of this is the New York Times' "How many households are like yours?" visualization. The participants in the focus group elected this visualization as their overall favorite and in their responses they stressed the fact that what made this visualization so interesting was the fact that they could chose to explore families similar to theirs. They also said that they would like if there were smaller articles about each type of family so they could see if they have a similar lifestyle. One of the participants stated that when he chose a kind of family he was declaring an intention, therefore he would be interest in reading an article about the type of family that he chose and not another type of family even if they had a more interesting lifestyle.

In a test focus group We have done in the University of Texas at Austin, in the USA, the "Home and away: Iraq and Afghanistan war casualties" visualization was one of the most popular examples, mainly because the participants could relate with the subject. Most of the participants either had family members in the military or had friends that were stationed in Iraq or Afghanistan, therefore, they not only related more with this visualization but also felt very passionate about it. Not all the responses were positive though. Some of the participants felt almost offended to see that the visualization had pictures of the fallen soldiers and that it contained a lot of information about the soldiers, etc. Nonetheless, this response was a product of the participants close relation with the subject.

A final characteristic that the participants appreciated was when the visualizations provided some temporal structure. In fact, according to Kosara and Mackinlay [13] "one of the fundamental features of stories is that they provide a temporal structure, even if not necessarily linear." Therefore if visualizations are able to not only to introduce elements of storytelling but also have a story like flow they will be more successful and maybe then they can be considered narrative visualizations.

Conclusion

In the same way that a well-told story is able to convey a large amount of information in a simpler and more compelling way, enabling the audience to assimilate and retain the information transmitted [8], a well structured visualization can also captivate the audience. Technology provided us with new tools to convey information in a story-like fashion[8] and that is clearly transforming our preferences. People get excited with good visualizations and the proof of this fact is that people are sharing these visualizations online and most of the times don't even care if they have an article associated with it. They just care about the visualizations alone.

Therefore we have to be able to understand the audience's preferences. We have to be able to successfully introduce storytelling in these visualizations, to tailor visualization systems to accommodate storytelling, because we do not know if the audience will go beyond the information that the visualization provides.

We believe that this topic is greatly relevant and the ques-

[13] http://www.marketplace.org/topics/economy/budget-hero

tions we pose by are important not only for the Information Visualization community but also to every area that wishes to elevate visualizations to a more complex form of dissemination of information/data. Empirical study is much needed for the field to move forward.

Nowadays, most research on visualizations is merely based on the time it takes to complete a task [13], but this makes little sense when what we want to produce is engaging visualizations, that people spend time on. This area of research is still very new and there is still little information on how to introduce storytelling in visualizations and even less research on what techniques work for the audience. It would benefit from rigorously studying and measuring the impact of visualizations. Although the study presented is mostly preliminary and exploratory we believe that it still provides some insights on what the audience cares about. The results of this focus group study showed that interactivity, drilling-down, additional context, and the ability to create a sense of relatability are important factors for the users to feel engaged. These have already been intuitively understood by others (visualization designers, data journalism teams, etc.) but it's still useful to observe what the public thinks of it, how they relate to these visualizations, and what they think could be improved.

The domain of storytelling in visualization is only just starting to take shape and, although quite a few research contributions have appeared recently on this subject, there are still ongoing discussion. There are ample opportunities to make in impact and this study hopes to give its contribution.

A specific aim for future work is to test other visualizations with the focus group method and probably focus on one particular design element to conduct an in-depth evaluation. Moreover we want to continue researching the elements that make a good visual storytelling in order to create a set of techniques and conventions for designing visualizations with a strong storytelling/narrative component.

References

[1] S. Bertschi, S. Bresciani, T. Crawford, R. Goebel, W. Kienreich, M. Lindner, V. Sabol, and A.V. Moere. What is knowledge visualization? perspectives on an emerging discipline. In *Information Visualisation (IV), 2011 15th International Conference on*, pages 329–336, 2011.

[2] Ian Bogost, Simon Ferrari, and Bobby Schweizer. *Newsgames*. The MIT Press, 2010.

[3] Alberto Cairo. *The Functional Art*. Peachpit Press, 2012.

[4] Nicholas Diakopoulos. Game-y information graphics. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '10, pages 3595–3600, New York, NY, USA, 2010. ACM.

[5] Nicholas Diakopoulos. Design challenges in playable data. In *CHI Workshop on Gamification*, 2011.

[6] Nicholas Diakopoulos, Funda Kivran-Swaine, and Mor Naaman. Playable data: characterizing the design space of game-y infographics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 1717–1726, New York, NY, USA, 2011. ACM.

[7] A. Figueiras. A typology for data visualization on the web. In *Information Visualisation (IV), 2013 17th International Conference*, pages 351–358, July 2013.

[8] Nahum Gershon and Ward Page. What storytelling can do for information visualization. *Commun. ACM*, 44(8):31–37, August 2001.

[9] J. Hullman and N. Diakopoulos. Visualization rhetoric: Framing effects in narrative visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):2231–2240, 2011.

[10] J. Hullman, S. Drucker, N.H. Riche, Bongshin Lee, D. Fisher, and E. Adar. A deeper understanding of sequence in narrative visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2406–2415, Dec 2013.

[11] Jessica Hullman, Nicholas Diakopoulos, and Eytan Adar. Contextifier: Automatic generation of annotated stock visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 2707–2716, New York, NY, USA, 2013. ACM.

[12] Tanja Keller and Sigmar-Olaf Tergan. Visualizing knowledge and information: An introduction. In Sigmar-Olaf Tergan and Tanja Keller, editors, *Knowledge and Information Visualization*, volume 3426 of *Lecture Notes in Computer Science*, pages 1–23. Springer Berlin Heidelberg, 2005.

[13] R. Kosara and J. Mackinlay. Storytelling: The next step for visualization. *Computer*, 46(5):44–50, 2013.

[14] Kwan-Liu Ma, I. Liao, J. Frazier, H. Hauser, and H.-N. Kostis. Scientific storytelling using visualization. *Computer Graphics and Applications, IEEE*, 32(1):12–19, 2012.

[15] E. Segel and J. Heer. Narrative visualization: Telling stories with data. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):1139–1148, 2010.

[16] Wita Wojtkowski and W Gregory Wojtkowski. Storytelling: its role in information visualization. In *European Systems Science Congress*. Citeseer, 2002.

Narrative Visualization: A Case Study of How to Incorporate Narrative Elements in Existing Visualizations

Ana Figueiras
FCSH - Universidade Nova de Lisboa
mail@anafigueiras.com

Abstract

Stories have long been used to convey information, cultural values, and experiences. Narratives not only have been the main way people make sense of the world, but also have been the easiest way humans found out to share complex information. However, today we are confronted with the problem of the amount of information available, which sometimes is hard to cope with. Combining storytelling with visualization has been pointed out as an efficient method to represent and make sense of data, at the same time allowing people to relate with the information.

In this paper, we explore the benefits of adding storytelling to visualizations. Drawing on case studies from news media to visualization research websites, we identified possible strategies to introduce storytelling in visualizations such as adding short stories or narrative elements using annotations and using time to introduce the feeling of storytelling or story-flow.

Keywords—**Storytelling, narrative visualization, case study.**

1 Introduction

In recent years many have researched the potential of the use of storytelling in information/data visualization. Moreover the interest in the area has also sparked outside of the research community due to its prospect of use in areas such as journalism, marketing, or education. The challenge is not only discovering ways to highlight the potential stories that exist within the data but also to transform visualizations in such a way that they adopt several narrative characteristics and eventually become a form of storytelling it self.

If visualizations get successfully infused with narrative we can overcome both the limitations of textual and visual representation. Due to the explosion of information and knowledge we have been witnessing in the last few years, potentially sparked by the blooming of the open-data movement, it is imperative that we discover better ways to understand information and to reduce the complexity of the information available. The benefits in using visualization for supporting users in coping with the complexity in knowledge- and information-rich scenarios has already been proven [9]. However there are still opportunities to make a contribution, specially in the sub-genre of narrative visualization (visualizations intended to convey stories [14]).

With the qualities associated with information visualization and storytelling, narrative visualization can become very successful. Establishing a correct balance between narrative and visualization however is vital. We have to maintain the rigor and accuracy associated with visualization and not introduce narrative elements that could hamper the apprehension and assimilation of the information.

Driven by questions such as *What elements of traditional storytelling can be embedded as part of the data-driven visualization? How do we balance the narrative flow of the visualization without disturbing the experience of discovery? What elements of design and interactivity help us to better tell these data-driven stories?*, in this work we examine the benefits of adding storytelling to visualizations and explore possible strategies to do so. We collected examples of professionally-produced visualizations and used them as case studies. We take an empirical approach, analyzing three visualizations, their use of narrative elements, and how they could be redesigned to better introduce storytelling elements. We explored three narrative strategies that could become relevant dimensions of narrative visualization: context, empathy, and temporality. In order to illustrate our approach we present three simple prototypes of the introduction of storytelling in the selected case studies. Finally, we discuss the implications of these strategies and pave the way to future research on the impact of these strategies and on the establishment of design conventions for narrative visualization.

2 Related Work

"Narrative is first and foremost a prodigious variety of genres, themselves distributed amongst different substances – as though any material were fit to receive man's stories. Able to be carried by articulated language, spoken or written, fixed or moving images, gestures, and the ordered mixture of all these substances; narrative is present in myth, legend, fable, tale, novella, epic, history, tragedy, drama, comedy, mime, painting (think of Carpacio's Saint Ursula), stained glass windows, cinema, comics, news item, conversation. Moreover, under this almost infinite diversity of forms, narrative is present in every age, in every place, in every society; it begins with the very history of mankind and there nowhere is nor has been a people without narrative. All classes, all human groups, have their narratives, enjoyment of which is very often shared by men with different, even opposing, cultural backgrounds. Caring nothing for the division between good and bad literature, narrative is international, transhistorical, transcultural: it is simply there, like life itself. [1]"

Storytelling is an ancient art deeply rooted in our common human culture. In spite of having thousands of different forms what still automatically springs to mind to most of us when we hear the term storytelling is the image of an elder narrating an old fairy tale to children [12].

Although it may not appear so, modern storytelling still maintains many of the characteristics of its traditional form. Technological advances are actually helping to introduce in modern storytelling more of these characteristics that we appreciate so much in traditional storytelling. For instance, the use of interactivity makes the user feel the joy of the moment of discovery typical of the live narration of a story.

2.1 Narratives and New Forms of Storytelling

There are basic elements for a narrative [6]: *situatedness* (discourse context or occasion for telling), event sequencing (structured time-course of events), world making/world disruption (disruption of a state of equilibrium), and what it is like (the feelings of living through the situation and a foregrounding of human experience).

Another notion typically associated with narratives is the idea of beginning, middle, and end [11]. In the beginning we have the set up or base reality, that may include background information and a change or conflict. The middle is usually composed of a struggle, complication, or development, pointing towards a climax. With the end comes the resolution and sometimes developments that are left open.

The medium shapes the information [5] and the way it is represented. Computation and the Internet are part of a *New Medium* that has given us the possibility to employ characteristics typical from other media, creating multimedia narratives more complex and sophisticated than ever. Interactive storytelling, for instance, is becoming very popular on news casting and documentaries, areas that were up until recently holding on to traditional forms of storytelling.

2.2 Using Visualizations to Tell Stories

The symbiotic relationship between information/data visualization and storytelling has revealed to be one of the more prevalent topics in visualization in the last few years [4, 7, 12, 14]. Gershon and Page [5] were the first to notice that storytelling could give a valuable contribution to the area of Information visualization, without however truly describing examples of actual information/data visualization and contributing with strategies *de facto* to introduce storytelling.

In 2010 the theme sparked again when Segel and Heer[14] re-approached it, naming it narrative visualization. The authors state that these *data stories* are an emerging class of visualizations. In addition to providing a typology for classifying visualizations and generalized advice for designing narrative visualizations, the authors also identified patterns and structures that news media use to introduce storytelling: Martini Glass Structure, Interactive Slideshow, and Drill-Down Story. These structures vary in terms of how much author-driven paths and how it is structured: the first beginning with an author-driven approach that opens up to a reader-driven stage once the author's narrative is over, the second being a completely linear path with some interactivity within the limits of each slide, and the last one being completely reader-driven. Segel and Heer also argue that most visualizations do not fit the author-driven (Explanatory) versus reader-driven (Exploratory) dichotomy, so commonly established among visualization creators, and is somewhere in the middle. Even narrative dense visualizations can include data that can be freely explored by users and let them draw their own insights.

There is an intense discussion [2] in the visualization area about whether or not introducing storytelling is beneficial. However most seem to agree that, when done right, it can be a powerful way to create a structured interpretation path [2]. Good narrative visualizations allow the user to engage with the data, makes the insight jump out, and helps users to cope with their short attention spans and lack of data literacy. Nowadays most visualizations depend on different media to provide explanations about the data, usually text. However, often people try to interpret the visualization by it self and do not care about the extra information necessary for its interpretation. Visualizations that are too exploratory with little author-driven guidance can undermine comprehensibility and engagement, resulting in a user that is under-informed or even misinformed. This does not imply that exploratory visualizations that are thoughtfully designed cannot be engaging, however the user is always a volatile variable. On the other hand, visualizations that are over-curated and too story-driven also tend to be boring, specially for proficient users.

Achieving a equilibrium between exploratory and expository is important if we want to have visualizations that are easy to interpret, appealing, and that still leave possibilities for exploration. Also, if we are able to successfully introduce storytelling we will be able to produce visualizations that can be entirely independent of other means of storytelling.

Even tough there is a considerable amount of literature on narrative structuring techniques for visualization [14], there considerably less on clear guidelines recipes that creators can use to find the best narrative strategies for different types of visualizations. Research [2] has already revealed that having flexible narratives with landmarks and spaces for the user to freely the in-between is a good option. However, this research towards design strategies and rhetorical techniques is still much needed.

2.3 Narrative Strategies

Several narrative strategies have been approached by different researchers in the past years, particularly approaches closer to semiotics, critical theory, and journalism. Authors such as Segel and Heer [14] and Hullman and Diakopoulos [7] proposed narrative strategies for visualization based on visual rhetoric. Segel and Heer's approach aimed towards structure and generalized advice for designing narrative visualizations. Hullman and Diakopoulos go farther proposing an analytical framework for visualization rhetoric that cross editorial layers (data, visual representation, textual annotation, and interactivity) and a set of techniques for visualization rhetoric(omission, metonymy, data provenance, representing uncertainty, identification, obscuring, contrast, classification, redundancy, typographic emphases, irony, similarity, individualization, anchoring, filtering). Their objective went towards the constitution of a guide to how much visualization rhetoric should be used on the design of visualizations. They also give insights about the impact of these rhetorical aspects influence the user's interpretation of the original data.

In this paper we analyze three particular approaches that were previously discussed in this field of study: context (closely linked to annotation), empathy, and temporality (its relation with story-flow). All of these aspects will be approached in relation with interactivity.

2.3.1 Context and Empathy

Interactivity opened up the possibility of adding new layers of content to information visualization. Thanks to this additional layer of content, most of the times in the form of annotations, visualizations can, in addition to the data itself, provide content that is able to add context. This content has the potential to help a user make sense of the data [8]. In addition interactivity offers the possibility to show the content on demand, giving the user a sense of freedom.

The free exploration of the data and its context stories allows the user to follow just the information he/she is most interested in, also improving his/her enjoyability.

According to Hullman et al. [8] annotations are a promising way to complement articles since they have the capacity to add context that otherwise would be very difficult to provide. Annotations with context information are easier to assimilate than a dense article and can serve as little moments of storytelling. However the context does not need to always be in the form of storytelling, it can also be given in the form of external links and short annotations.

These moments of storytelling can add another dimension to the visualization: empathy. Empathy and emotion are concepts that were not often associated with information visualization, speacially because the first concept are usually associated with chaos and the last with objectivity. However emotive/empathic information visualizations revealed to be often more memorable [10] and even, at times, more enjoyable. This sense of empathy can be achieved by making the user relate to the topic or to the individuals represented in the data (by allowing the user to see him/her represented or by putting the user in other people's shoes).

2.3.2 The Relation Between Time and Narrative

Temporality is a major structural factor in our lives and it is closely related with narrativity[13]. Narratives are able to represent the human experience of time in its two different modes: the linear succession time (the sequence of minutes, hours, days) and the phenomenological time(the past, the present, and the future, which do not necessarily correspond to the linear structure of before and after, in other words, a narrative may begin with a culminating event or the temporality that is lived in the narrative may not concur with the time of the events the the story is said to depict).

Temporal structure is something that can give visualizations a sense of story-flow and this often appeals to users, because it gives them the ability to navigate their way to particular information. Structures such as timelines are very efficient in giving this temporal sequence feel. Nonetheless, this sense of temporality does not need to be expressed as a linear structure and stories are useful way to do so because they do not always have a linear temporal structure [10]

3 Case Study

Three case studies were used to demonstrate the strategies of storytelling approached in this paper. The first example, *How many households are like yours?*[1] highlights how it is possible to introduce short stories to add empathy. The second, *What does china censor online?*[2] illustrates how it's possible to add context. The last example, *Death penalty statistics, country by country*[3] demonstrates the benefits of time as a form to introduce the feeling of narrative.

[1] http://www.nytimes.com/interactive/2011/06/19/nyregion/how-many-households-are-like-yours.html
[2] http://www.informationisbeautiful.net/visualizations/what-does-china-censor-online/
[3] http://www.theguardian.com/news/datablog/2011/mar/29/death-penalty-countries-world

3.1 Home Many Households are Like Yours?

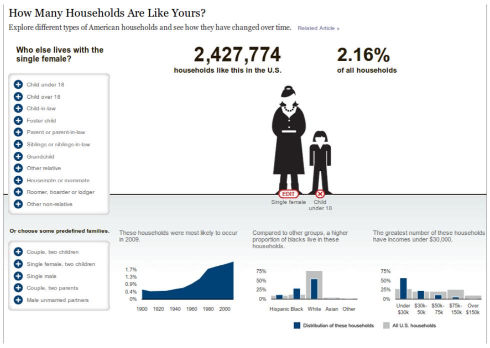


Figure 1: Original How many households are like yours?

Accompanying the article *Baby Makes Four, and Complications* on the changing family dynamic, The New York Times published an interactive visualization that lets users explore different types of American households: *How many households are like yours?*, shown in Figure 1.

The user is first presented with the option to choose the primary residents of a household (married couple; male/female unmarried partners; single male; single female; male unmarried partners; and female unmarried partners), represented through pictograms. Afterwards the user can add secondary members of the household (child under 18; child over 18; child-in-law; foster child; parent or parent-in-law; siblings or siblings-in-law; grandchild; other relative; housemate or roommate; Roomer, boarder or lodger; and other non-relative), also represented as pictograms. The graphic updates on the fly and simultaneously shows how the entered household compares to the rest of America's households. The visualization shows the total of households in the US that are like the one the user selected and the respective percentage. On the bottom there is a breakdown by time, race, and household income.

The user is presented with the possibility to choose any kind of household that he/she wishes however the visualization challenges the user to try his/her own family. This creates a sense of proximity between the user and that data.

In a previous research [3] we analyzed the different elements that compose this visualization. In terms of interactivity the New York Times visualization enables the user to click and hover details and filter the data. The narrative elements we identified in it were title, captions, annotations, introductory text, and accompanying article.

It is possible still to improve the sense of relatability that the user feels with the data if we introduce short stories about the different kinds of families instead of having only one long article about one type of family introducing the visualization. In Figure 2 we present how this can be done without changing the visualization too much. Basically, similarly to what happens with the graphics on the bottom of the visualization we propose that the visualization includes also a short article characterizing the type of family that the user selected. We used the main article that accompanies the original visualization as the example for household with a single female with a child under 18. Having stories for each type of household helps the user to see that data not just as a type of household but as real people.

Relatability is a factor that helps the user enjoy the visualization. It is one of the characteristics that makes it memorable [10] and one that is able to make the user feel

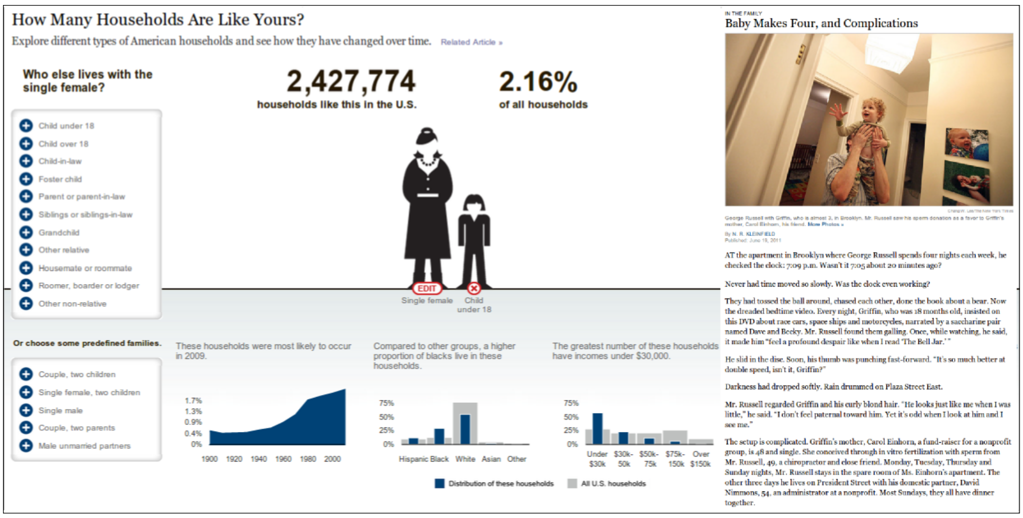


Figure 2: How many households are like yours? enhanced with storytelling

empathy with the subject or the individuals represented in the data, and which will probably make the visualization more successful.

Figure 3: Original What does China censor online?

3.2 What Does China Censor Online?

The visualization, by David McCandless, *What Does China Censor Online?*, shown in Figure 3, is a simple tag cloud that only has a title and text, in this case mere disconnected words.

The user is presented with a non-playable visualization that, although it is not visible at first, it is in fact a map. The tag cloud is shaped as a map of China.

This visualization would benefit greatly with the addition of extra information to add context. In Figure 4 we present how this can be done maintaining most of the original design. We propose the introduction of small *tooltips* that pop-up when the user clicks on one of the websites censored. This would help the user realize the possible reasons for the censorship. There should also be external links to the actual websites.

Context information can work as little moments of storytelling. This kind of short stories can be more easily interpreted by the user than a dense article. Context could also be introduced as external links, for instance Wikipedia links or related articles, or short annotations.

This context information is beneficial for providing information that otherwise would be difficult to provide [8].

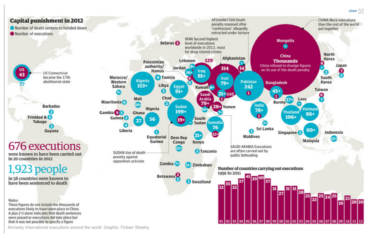


Figure 5: Original DPS, country by country

3.3 Death Penalty Statistics, Country by Country

Death penalty statistics (DPS), country by country, shown in Figure 5, is a visualization by The Guardian that accompanies an article about countries that maintain the death penalty.

Figure 4: What does China censor online? enhanced with storytelling

The map/diagram static visualization has bubbles of different sizes to representing the number of death sentences handed and executions in countries that are still carrying executions. On the bottom there is also a timeline representing the number of abolitionist countries for each year between 1991 till 2012. The timeline resembles a bar graph. Apart from the large article of which the visualization is part of, *Death penalty statistics, country by country* in terms of narrative elements only has an introductory text and captions that indicate the short information such as names of countries and dates.

This visualization would probably benefit if the timeline would actually function as a navigation and when the user clicks a certain year the map would show the number of death sentences handed and executions of that year. This representation of time and, specially the evolution of events, often appeals to users.

The use of interactivity elements such as hover or click details would also be useful on a visualization such as this one, because it could add extra information making the data more meaningful. In our prototype, shown in Figure 6, we propose adding *tooltips* with extra information about the executions and death sentences when the user clicks each countries' bubble. This *tooltip* could have general information about the subject for each country in a given year or a particular execution story to increase the empathy between the user and the data.

4 Conclusion

In this paper we explore the benefits of adding storytelling to visualizations and propose three strategies that still leave the way open for free exploration of the data: adding context, empathy, and temporal references. These techniques can be combined.

We take an empirical approach, analyzing three professionally-produced visualizations, their utilization of narrative elements, and how they could possibly be redesigned to better introduce narrative components. In order

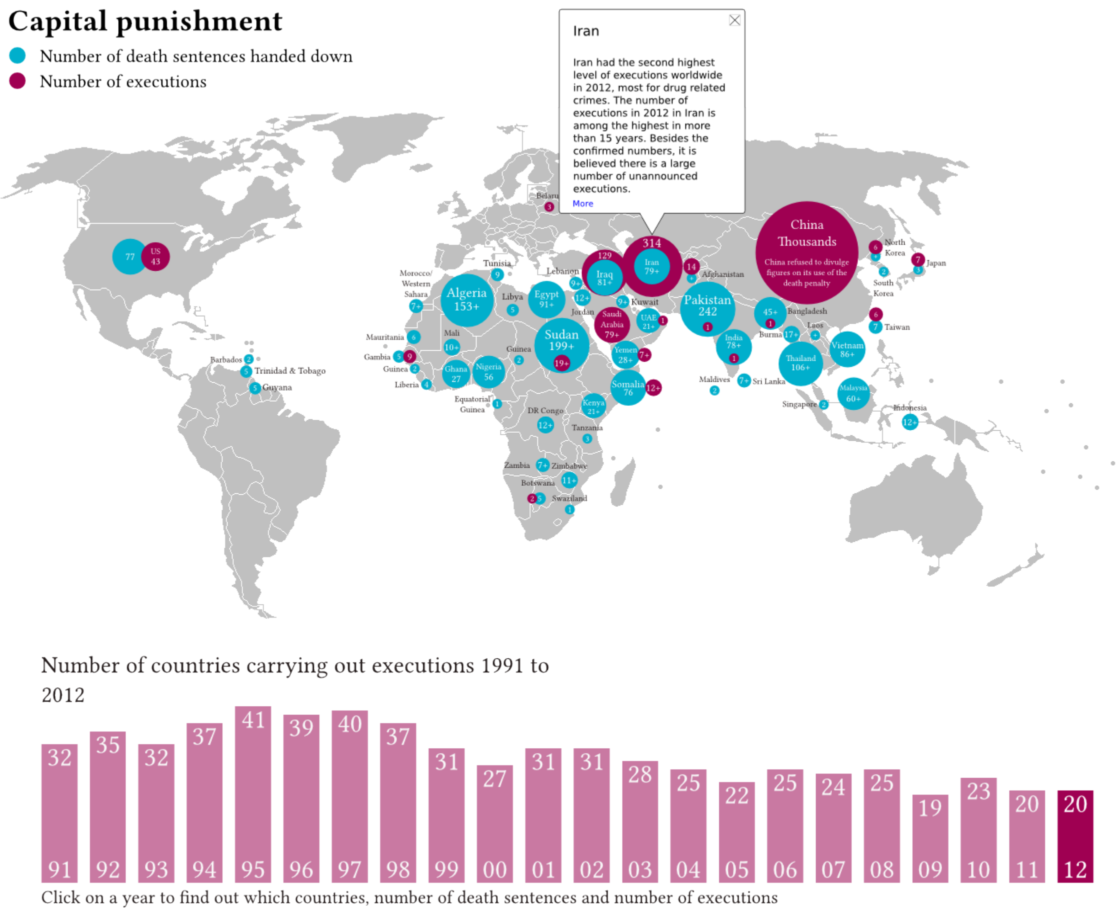


Figure 6: Death penalty statistics, country by country enhanced with storytelling

to illustrate our approach we presented three simple prototypes of the introduction of storytelling in the selected case studies. This paper aims to be a contribution somehow between a design study and a model, therefore we discuss the implications of these strategies and try to shed a light on the impact they will have on the interpretation and level of enjoyability of the visualization. We were driven by the motivation to pave the way to future research on the impact of these strategies and on the establishment of design conventions for narrative visualization.

One of our main goals was to research techniques that add a story feel to the visualizations without however preventing the free exploration of the data. Narrative visualization shouldn't become a lean-back format, accordingly the quest to add storytelling has to be weighted so that it does not lead to a linear, too author-driven interpretation path for the user. Stories in visualization should be used as starting points for data exploration or short moments of insight about the data, rather than a predigested narrative.

Providing free access to the data seems to help address the need to expose the intricacy of the information, however this might be confusing specially for non-proficient users. Possibly narrative elements can help frame the inner contradictions of the data and lead the users to their own interpretations of the information, guiding their attention in subtle ways.

To make further progress on narrative visualization we still need to know:

- what makes it work;
- which of the narrative visualizations that are being produced in news media, advertising, research, education, etc. are having the desired effect on users;
- how and where should narrative elements be placed;
- how should the story be structured;
- what is the impact of these stories on the users.

We're at an inflection point where we understand the design dimensions enough to start working towards the construction of models for narrative visualizations. Once these are built they should be employed in the design of visualizations, tested, and maybe then we will achieve some answers to these questions. The systematic study of these narrative visualizations is the only way to further amplify our understanding on this subject.

References

[1] Roland Barthes. Introduction to the structural analysis of narratives. 1988.

[2] Nicholas Diakopoulos. Game-y information graphics. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '10, pages 3595–3600, New York, NY, USA, 2010. ACM.

[3] A. Figueiras. A typology for data visualization on the web. In *Information Visualisation (IV), 2013 17th International Conference*, pages 351–358, July 2013.

[4] D. Fisher, A. Hoff, G. Robertson, and M. Hurst. Narratives: A visualization to track narrative events as they develop. In *Visual Analytics Science and Technology, 2008. VAST '08. IEEE Symposium on*, pages 115–122, Oct 2008.

[5] Nahum Gershon and Ward Page. What storytelling can do for information visualization. *Commun. ACM*, 44(8):31–37, August 2001.

[6] David Herman. *Basic elements of narrative*. John Wiley & Sons, 2011.

[7] J. Hullman and N. Diakopoulos. Visualization rhetoric: Framing effects in narrative visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):2231–2240, 2011.

[8] Jessica Hullman, Nicholas Diakopoulos, and Eytan Adar. Contextifier: Automatic generation of annotated stock visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 2707–2716, New York, NY, USA, 2013. ACM.

[9] Tanja Keller and Sigmar-Olaf Tergan. Visualizing knowledge and information: An introduction. In Sigmar-Olaf Tergan and Tanja Keller, editors, *Knowledge and Information Visualization*, volume 3426 of *Lecture Notes in Computer Science*, pages 1–23. Springer Berlin Heidelberg, 2005.

[10] R. Kosara and J. Mackinlay. Storytelling: The next step for visualization. *Computer*, 46(5):44–50, 2013.

[11] Shixia Liu, Yingcai Wu, Enxun Wei, Mengchen Liu, and Yang Liu. Storyflow: Tracking the evolution of stories. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2436–2445, Dec 2013.

[12] Kwan-Liu Ma, I. Liao, J. Frazier, H. Hauser, and H.-N. Kostis. Scientific storytelling using visualization. *Computer Graphics and Applications, IEEE*, 32(1):12–19, 2012.

[13] Paul Ricoeur. *Time and narrative*, volume 3. University of Chicago Press, 1990.

[14] E. Segel and J. Heer. Narrative visualization: Telling stories with data. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):1139–1148, 2010.

Towards the Understanding of Interaction in Information Visualization

Ana Figueiras
FCSH - Universidade Nova de Lisboa
mail@anafigueiras.com

Abstract

Over the past few years the web has been responsible for the rise in popularity of visualizations and it seems that interactive or playable visualizations have become more popular and end up standing out more. The use of interactivity and animation has been extensively discussed in information visualization research, but there has been some controversy in relation to its benefits. Additionally, there is still little empirical evidence about its efficacy in terms of improving understanding of the data and there is few research that points out guidelines of how to incorporate it successfully and that proves that playable visualizations are indeed more enjoyable and popular among users.

In order to guide future research on the actual benefits of interactivity in visualization it is important to understand what types of interactivity are currently being used in the field and to have a framework to help discuss and evaluate interaction techniques. After conducting an extensive review of popular visualizations and their interactive capabilities, we propose eleven categories of interaction techniques: filtering, selecting, abstract/elaborate, overview and explore, connect/relate, history, extraction of features,reconfigure, encode, participation/collaboration, and gamification.

Keywords—**Visualization, interaction, taxonomy.**

1 Introduction

If we interpret information visualization as "the use of computer-supported, interactive, visual representations of abstract data to amplify cognition" as does Card et al. [5], it is almost impossible to discard the role of interactivity. Nonetheless, many support a different definition that stands on the fact that interactivity is not always necessary to have a successful visualization and that interactivity can sometimes negatively affect the understanding of the data. Few however would deny that interactivity has several benefits specially when the data sets are quite large. Moreover, taking into account that information visualization is deep-rooted in the computer science community it makes sense to approach interactivity as a key element.

Interactivity has been utilized in information visualization with several purposes. The more common are: 1) making the data more engaging or playful and 2) showing the data in manageable portions. According to Keim [19] having the data in smaller portions is particularly important when exploring large data sets. Doing so facilitates both the understanding and the analysis of the data because the degree of complexity is reduced. By employing interactivity techniques, visualization creators try to give the users the ability to properly explore the data and find appropriate answers to their questions. Providing ways for the users to independently find the answers (exploratory visualization) often seems to be a better option than presenting answers to what the creator believes are the users' questions (explanatory visualization), not only because it is difficult to predict what the questions will be but also because visualization is a discovery tool and limiting its potential to provide insights is a mistake.

In order to study the impact of interactivity in Infovis it is important to study and understand the possible interaction techniques, and most existing taxonomies do not include new interaction techniques such as gamification. Therefore, we propose a new taxonomy based on previous research.

2 Background

The *Visual Information-Seeking Mantra* by Shneiderman [24] is the most well known general interaction techniques taxonomy. However when we seek for a more extensive taxonomy for Infovis we find a multitude of studies [1, 10, 19, 30] showing that there is not a taxonomy that is consensual. According to Yi et al. [30] defining a taxonomy is challenging and they can easy get obsolete if a new interaction technique that does not fit any of the taxonomic units is discovered.

A careful analysis of recent visualizations reveals that current taxonomies do not include newer interaction techniques that are now being introduced such as participation or gamification. Therefore, we saw the necessity to evaluate the existing literature in order to propose a better taxonomy. Table 1 summarizes the studies that were taken into account while developing our proposed taxonomy, two that only concern interaction techniques for information visualization [19, 30] and a more general approach [24].

Shneiderman [24]	Overview, zoom, filter, details-on-demand, relate, history, and extract
Keim [19]	Dynamic projections, interactive filtering, interactive zooming, interactive distortion, and interactive linking and brushing
Yi et al. [30]	Select, explore , reconfigure, encode, abstract/elaborate, filter, and connect

Table 1: Interaction techniques taxonomies

3 Categories

In order to more systematically explore the purposes of interactivity in information visualization, we began with the goal of building a comprehensive list of interaction techniques. Backed up by the existing literature [19, 24, 30], we evaluated 232 visualizations that were popular on the web and studied the types of interaction they use. The visualizations are available at *www.rethinkingvis.com* and belong to the same corpus of visualizations studied on previous research[11]. From this study eleven categories emerged: filtering, selecting, abstract/elaborate, overview and explore, connect/relate, history, extraction of features, reconfigure, encode, participation/collaboration, and gamification.

Filtering	only show me the data in which I am interested
Selecting	mark or track items in which I am interested
Abstract/Elaborate	adjust the level of abstraction of the data
Overview and Explore	overview first, zoom and filter, then details-on-demand
Connect/Relate	show me how this data is related
Reconfigure	give me a different arrangement of the data
Encode	give me a different representation of the data
History	allow me to retrace the steps I take in the exploration of the data
Extraction of features	allow me to extract data in which I am interested
Participation/Collaboration	allow me to contribute to the data
Gamification	show me the data in a more playful way

Table 2: Proposed taxonomy

Each type of interaction will be discussed in more detail in the following subsections.

3.1 Filtering and selecting

Reducing complexity is one of the major goals of introducing interactivity in visualizations. A common way to achieve this effect is by **filtering** the data. Filtering out uninteresting items, either by specifying a range or a condition, is a natural method of requesting data.

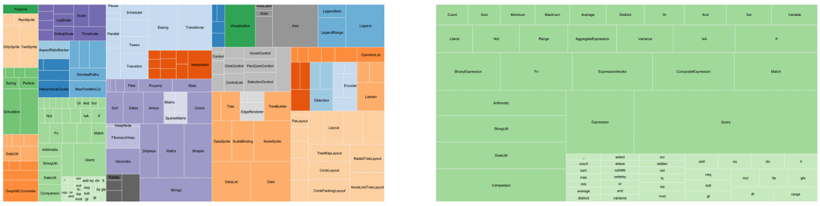
The most successful way to filter data is through the use of dynamic filters that allow the users to quickly see how the data representation is affected when the items of no interest to him/her are eliminated or deemphasized. The data remains unchanged and can be shown whenever the users wishes by reseting the criteria [30]. Card et al. [5] found empirical evidence of the efficacy of dynamic queries referring its advantages and disadvantages. In 1999 one of the disadvantages was that the dynamic queries approach was poorly matched with the hardware and software systems available back then. Nowadays this has been overcome and therefore dynamic queries have become extremely popular, not only in information visualization. The most successful filtering implementations are the ones that allow the immediate update of the display [6]. The advances in technology permitted improvements in terms of performance, and these filtering systems have become incredibly more responsive.

A simplified way to filter data and selectively hide and reveal items is a way to aid cognition that enables users to quickly focus on what really matters to them. However, long delays between the user's input and the system's response can negatively affect the whole experience and inclusively the final interpretation of the visualization.

Select functionalities can also be used to aid cognition. Being able to mark and track items or sets that are interesting becomes particularly useful when there is the possibility of changing the visual representation of the data [30] or when the data is dynamic and constantly updated. According to Yi et al. [30] "rather than acting as a standalone technique, Select interaction is coupled with other interaction techniques to enrich user exploration and discovery." Select techniques also act as a filter, which instead of hiding the remaining data puts in evidence the data of interest and allows the user to see it in contrast with the other items.

3.2 Abstract/Elaborate

Several abstract/elaborate interactivity techniques are used in information visualization. These interaction capabilities allow the user to easily adjust the level of abstraction of the data representation to his/her interpretation needs [30]. The user regulates the amount of stimuli that the visualization provides him/her by varying the amount of information that displayed or emphasized.



(a) Overview of the treemap (b) After click on the light green area

Figure 1: D3 zoomable treemap

Zooming is a very common and well-known example of abstract/elaborate interactivity technique [30]. Often there is some confusion with the term *zooming* due to its use as a term for generic scalar changes, rather than adjustments of vantage point. According to Craft and Cairns [6] it "refers to the adjustment by the user of the size and position of data elements on the screen." Zooming allows the user to see an overview of the visualization (through zoom-out) or to see a smaller, more detailed, view without fundamentally altering the representation(as it can be seen on Figure 1). This technique acts as a kind of filter by navigation, allowing the user to apply the technique on items of interest, simultaneously removing from view or reducing the size of items that are not of interest. As it happens with filtering, zooming helps in reducing complexity.

The use of zooming techniques in visualizations facilitate two distinct cognitive tasks:

- when zooming-in the user is being aided with the organization of the information into meaningful patterns, which is enabled by the removal of extraneous information from his/her visual field;
- when zooming-out the user is presented with hidden contextual information that was presented to him/her upon the start of the exploration but that he/she probably cannot recall.

Although with different implications for cognition, these two actions are procedurally and visually symmetrical [6]. In other words the zooming-in action enlarges smaller data elements and the zooming-out action produces the opposite result (reduces larger data elements). *Zooming-in* enlarges small data elements in which the user is interested, removing from view or reducing the size of large uninteresting data. *Zooming out* has the opposite effect. The results are procedurally and visually symmetrical however the implications to cognition are very different.

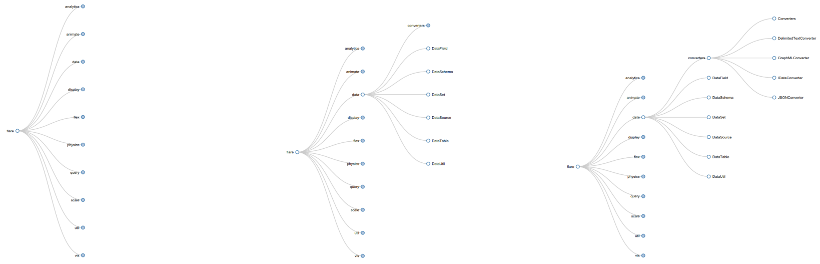
Specially when dealing with large sets of data, it is important to provide the user with both representations. The highly compressed representation of the data [19] will provide an overview that will reveal the position of the data he/she is interested within the whole information space, will reveal outliers and patterns, etc. The more detailed view will provide the data in manageable inputs [6], without the noise of data that is not of interest for the user. Having zooming options allows the user to have the best of both types of representations in the same visualization. "The objects may, for example, be represented as single pixels on a low zoom level, as icons on an intermediate zoom level, and as labeled objects on a high resolution" [19].

However, zooming is only successful when it preserves the user's sense of position and context. If there is not a smooth transition between levels of zooming or if the user's input does not translate adequately his/her interpretation may be affected. According to Shneiderman [24] "a very satisfying way to zoom in is by pointing to a location and issuing a zooming command, usually by clicking on a mouse button for as long as the user wishes."

Details-on-demand is another type of abstract/elaborate interaction. This technique consists of getting additional details upon the selection of an item or group. As stated by Craft and Cairns [6] "the details-on-demand technique provides this additional information on a point-by-point basis, without requiring a change of view."

There are several ways to provide the user details-on-demand on a visualization but one of the most common techniques is by providing drill-down options. Drill-down operations are very common in tree visualizations, to which they provide the functionality of only showing the levels or sub-trees that are of interest to the user (as seen in Figure 2). This functionality allows the limitations of screen space and visual complexity to be overcome, while maintaining the general representational context.

Another popular details-on-demand technique is the use of tool-tips or pop-ups. This interactivity technique, often provided on mouse-hover or click, allows the user to access detailed information about an item [30], which usually would not be easily shown in the visualization. According to Segel and Heer [23], details-on-demand is one of the types of interactivity common in narrative visualization. These annotations, often overlooked in information visualization evaluation despite of its important role, can be textual, graphical, and even social/participatory [15]. They



(a) Tree with two levels (b) Tree with three levels (c) Tree with four levels

Figure 2: D3 collapsible tree

can provide backstories that not only help in the level of engagement of the user but also provide relevant details. Annotations are also useful to focus the users attention on a specific area of the visualization [15].

Linking is a technique that is not often regarded as a details-on-demand operation. Linking can be used to give access to external information, as it is the case of hyperlinks (which reference data that the reader can access directly by clicking on it), or (as referenced by [19]) to give access to a different visualization method.

3.3 Overview and explore

Although it is useful to provide the user with detailed information it is also important to allow the user to have an **overview** of the entire collection. Actually, according to the *Visual Information Seeking Mantra* [24] it is better to overview first, because the overview gives the user the general context necessary to understand the data set as a whole. That will allow the user to more easily identify patterns and themes in the data [6]. According to Craft and Cairns [6] even the overall shape of the visualization can give insights about the information that is encoded. Further examination possibilities can be added by introducing any of the abstract/elaborate techniques cited in Section 3.2.

Due to the complexity and size of most data sets, visualization creators often opt for showing only a limited number of items at a time. View/screen limitations and fundamental perceptual and cognitive limitations in human information processing also force creators to reduce the amount of information shown [30]. However, this information should still be available for exploration in order to enable users to examine a different subset of data and consequently get insights derived from the comparison of data.

Explore interactions provide this possibility. According to Yi et al. [30] explore techniques show new data by making these enter the view and removing other, instead of making complete changes. As reported in the survey by Yi et al. [30], the most common type of explore interactions is *panning*. This technique consists of the movement of a camera across a scene or the opposite, and in computer assisted visualizations "is often achieved by a special mode where the user grabs the scene and moves it with a mouse or by simply altering the view via scrollbars" [30].

3.4 Connect/relate

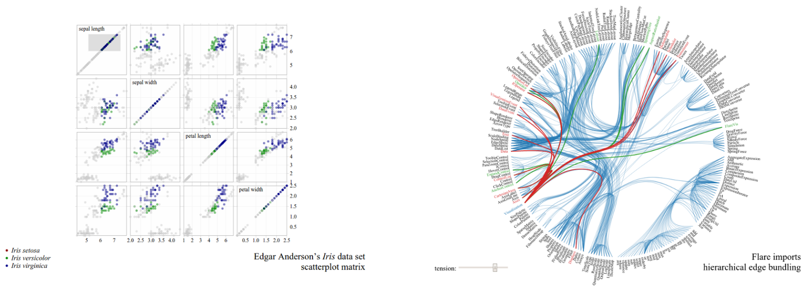
Connect, also referred to as relate, is an interactivity technique that enables viewing relationships between the data items. These relationships can be shown by highlighting links between the items that are already represented in the visualization or even by showing items that are relevant to an item that the user has interest in and that were previously hidden [30]. According to Craft and Cairns [6] "supporting discovery of relationships is particularly important where comparisons need to be made among the characteristics of different data objects in the display" [6].

In Figure 3a the user is able to compare the data of interest for him/her by selecting specific data items in the first scatter plot for example. The same data items will be highlighted in the other scatter plots and the items that were not selected will be deemphasized. Even though the color coding helps in finding the data of interest in the different views displayed, it would be difficult for the user to do comparisons if he was not able to highlight the data of interest. There would be too much noise.

Connect interactions can also be applied in visualizations that consist of a single view [30]. For instance, in a chord diagram, such as the one in Figure 3b, connect interactions can be used to enable the user to highlight the connections that he/she is interested in and easily set them apart from other relationships in the matrix.

3.5 History and extraction of features

"Information exploration is inherently a process with many steps, so keeping the history of actions and allowing users to retrace their steps is important" [24]. Providing ways for the user to undo and replay his/her actions allows him/her to not only recover from mistakes in the data exploration, but also to progressively refine the exploration [6]. In



(a) D3 scatterplot (b) D3 hierarchical edge bundling

Figure 3: Examples of the use of the connect interaction technique

1996 Shneiderman [24] pointed this interaction technique as one that is frequently disregarded in information visualization. The **history** feature is still often forgotten by visualization creators nowadays.

Another technique that is less common is the capability of **extraction** of important findings. Exploring the data often becomes a lengthy and complex task, therefore allowing the users to extract the data so it can be shared, dissected, or even seen in other visual representations, can reduce that complexity and result in better insights [6, 24]. Allowing the query parameters to be extracted can also benefit the data exploration preventing the need to repeat actions.

3.6 Reconfigure and encode

The **reconfigure** interactive technique provides the users with different perspectives about the data set by changing the spatial arrangement of the representation [30]. This can be done, for instance, by allowing the user to rearrange the order of columns or the rows, or by allowing the change of the attributes presented on the axis of a graph.

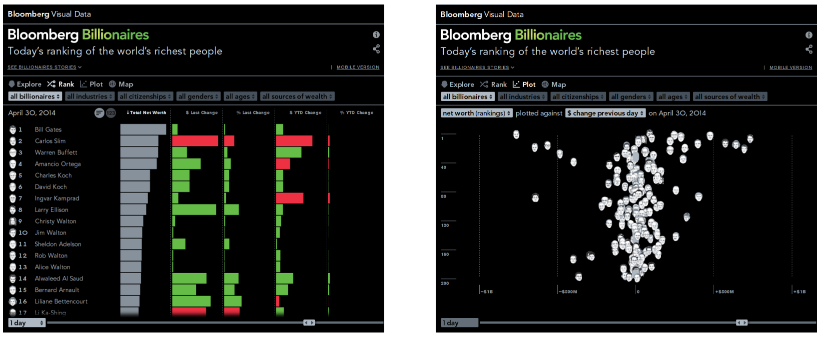
For example, in *As the Oscars age, so do the nominees*[1] The Guardian plots the ages of Oscar winners and nominees on a series of charts for different Oscar categories, allowing the user to filter by age difference and actual age. It is possible to see that, in recent years, the Academy has recognized an ever-broader range of ages as the gap between the youngest and oldest nominees has grown wider. The Guardian used the reconfigure technique to allow the user to choose between seeing the age difference plotted or the actual age. The first view allows to instantly perceive the trend of an ever-broader range of ages of nominees. The view by actual age allows to easily perceive the gap between the youngest and oldest nominees, which has grown wider in the last few years. The rearrangement of the data allows the user to have different perspectives that he/she probably would not have with a single representation.

Another way to provide different perspectives on the data is by providing completely new representations. According to Yi et al. [30] "in Infovis systems, visual elements serve an important role not only because they can affect pre-attentive cognition but also because they are directly related to how users understand relationships and distributions of the data items." Therefore, providing encoding techniques that allow the user to fundamentally change the visual representation can facilitate the discovery of new insights about the data. The changes in **encode** can be in terms of color, size, and even shape.

In Figure 4 it is possible to see the encode interaction technique applied in a visualization by the media company Bloomberg. The visualization entitled *Bloomberg Billionaires*[2] allows the user to see a rank view of the billionaires on a given date and the last change in their net worth (seen in Figure 4a) and the same data in a plot view (seen in Figure 4b). While the rank view emphasizes the order of the rank, the plot view emphasizes the last change in their net worth, therefore the user will more easily see that Carlos Slim, for instance, lost a lot of money on May 23 2014 (-$520.3M). However, in this view it is more difficult to see small net worth losses or gains, such as the ones that Bill Gates had (+$110.1M). Without this technique it would be more difficult for the user to come across these insights.

The use of reconfigure and encoding techniques can be combined in the same visualization. An example can be seen in Figure 4a, where it is shown that, in the *Bloomberg Billionaires* visualization, the user is able to order each of the different columns by ascending or descending order, due to the use of the reconfigure interactive technique. The user can opt to see the net worth ordered by total, by last change in dollars, by last change in percentage, by year to date change (from January 1st of the current year up until the chosen date) in dollars, and by year to date change in percentage.

[1] http://www.theguardian.com/film/interactive/2014/mar/02/oscars-award-nominees-age-best-actress-actor
[2] http://www.bloomberg.com/billionaires/



(a) Rank view

(b) Plot view

Figure 4: Encode interactive options in *Bloomberg Billionaires*

3.7 More complex forms of interactivity

Participation and collaboration are relatively new trends in information visualization. Both build on the growing will to empower users and building on the participatory culture. This neologism, which was first explored by Henry Jenkins [17], opposes to the consumer culture by transforming the user in a *produser* [4] who not only participates as a consumer of content but also as a contributor to the content they consume, shaping that content. Participatory culture began as an alternative phenomenon, often seen as a parallel subculture, however it is “anything but fringe or underground today” [17] and is being embraced by most institutions, from education and politics to media and advertising. It grew out of the blogs, forums, and mailing lists and is now an integrated feature in different domains, visualization being one of them.

In Infovis research, this inclusion of participatory culture is referred to as participation or collaboration. Mostly the different terms converge more or less to the same definition, however both terms can also be used to characterize slightly different types of interaction. The most common definitions center on the fact that there is more than one person (usually geographically separated [20]) contributing to the visualization interpretation/understanding, sharing their insights [16, 21]. A concept that usually accompanies these definitions is *social data analysis* (SDA), which, according to Wattenberg [27], concerns the social interaction around data analysis. It is as version of *exploratory data analysis* (EDA), which is a rich data analytical approach to analyzing data sets, recommended as a complement to confirmatory methods, that often relies on visual methods, based on the work of John Tukey. Similarly to EDA, SDA focus on the exploration of the data beyond the formal modeling and the confirmation of previous assumptions, but “relies on social interaction as source of inspiration and motivation” [28]. In the analysis of NameVoyager Wattenberg [27] found that its success might have been related with the social nature of the exploration of the web-based visualization. NameVoyager plots historical trends in baby naming and cause a buzz even among who do not find the data interesting. The creators found that the users were engaging in an intense dialogue about the visualization deeply exploring the data, helping each other discovering outliers and making causal relations, and even challenging each other to find patterns in the data. Since sensemaking is often a social process [14] (done in person or resorting to telecommunication devices) and data interpretation is frequently a group activity, it was also to expect that data visualization exploration became a social activity if the means necessary to support data analysis as a social process are provided. Even if the visualization itself does not allow this sharing of insight, it might still occur separately in social networks, forums, chats, and even offline. According to Heer [13], “the immersive and compelling nature of many social visualizations arise not only from the nature and presentation of the data under consideration, but also from the social interactions, both implicit and explicit, surrounding the use of the visualization.”

This phenomenon of wanting to explore visualizations in a social, collaborative fashion (which has inclusively been an important factor for the adoption of visualization) has been identified by several other authors [14, 18, 23, 25, 26, 29]. They have pointed out various strategies that better allow social insights, for instance tags, links, bookmarks, doubly linked discussions, graphical annotations, the traditional comments, etc. One of the biggest challenges with sharing insights specially about an interactive visualization is to share a specific state of the visualization, which is usually defined by a determined setting of filters or search parameters. Bookmarks for instance can identify a fixed state of the visualization [14, 25] so that the user can share directly with other users or even include it in their comments along with their insights. Another convenient feature is the possibility to do annotations on the visualization. This can be done by adding textual annotations that feature interesting insights communicated by the users [14], which is a very familiar action since it resembles the activity of annotating paper documents [2], or by highlighting and selecting specific items to include in their comments (graphical annotations).

In spite of all the perceptual and cognitive benefits that better social interactions provide, most visualizations continue to rely on simple text comments to allow users to share their insights [22]. According to Satyanarayan and Heer [22], although there is evidence that users are eager to share their own data stories most collaborative visualization tools provide minimal support for reusing visualizations and other types of more intense collaboration. Unfortunately, collaborative features that take full advantage of the opportunities that the web brings tend to be harder to implement, therefore techniques such as user-generated annotations and bookmarks are rare.

Participation/collaboration can also have a bigger impact on the visualization itself. For example, *Home and Away: Iraq and Afghanistan War Casualties*[3], the web-based visualization by CNN that maps the fallen soldiers in the wars in Afghanistan and Iraq, allows the users to add information about each soldier to their profile in the visualization. Using iReport, CNN's citizen journalism tool that allows users to contribute pictures and videos of news stories, the users can add memories and messages about a certain soldier that they know. The fact that the users' contributions are about a subject of the data set, and less about insights on the data as a whole or about the visual representation, makes this kind of contribution different from the ones cited above. This kind of participation/contribution becomes part of the visualization itself, shaping it in a permanent way with changes to the data that will be visible to other users.

Gamification is one of the most complex interaction techniques that can be added to a visualization. According to Deterding et al. [8] this "is an informal umbrella term for the use of video game elements in non-gaming systems to improve user experience (UX) and user engagement" and comprises a panoply of elements such as narrative context, ranks and reputations, time constraints, levels, goals, etc. This type of interaction is the least common because its production is time consuming. Even if gamified visualizations do not need to be as complex as a commercial computer game (and according to Deterding et al. [7] this is what distinguishes *gamification* from entertainment and serious games), nor does the data used need to be ever changing as it happened with Salubrious Nation by Diakopoulos et al. [9], the time, effort and skills required to make them are stopping its spreading. Most of the *game-y* information graphics, or playable infographics (the alike term coined by Bogost et al. [3]), have been produced by news media (*Budget Hero*[4] by American Public Media, *HeartSaver*[5] by ProPublica, *World Data Cup*[6] by La Stampa, etc.) or marketing initiatives (*SPENT*[7] by McKinney), organizations that depend on deadlines and usually cannot invest too much time developing these types of visualizations [12].

Although gamified visualizations can include most of the traditional interaction techniques that were discussed previously, what makes them different is the inclusion of game mechanics or game design patterns. According to Deterding et al. [7] "neither game mechanics nor game design patterns refer to (prototypical) implemented solutions; both can be implemented with many different interface elements."

Conclusions and Future work

In this paper, we proposed eleven different categories of interaction techniques, which resulted from the exhaustive analysis of a large corpus of 232 visualizations collected from specialized blogs, online journalism, advertising, scientific research, etc. This profound analysis led to the identification of patterns that correspond to types of interactions. Its main contribution is a steeping stone to the future study of individual interaction techniques, which consequently will allow a deeper understanding of how interactivity should be used in information visualization and which techniques have a bigger impact both on enjoyability and understanding.

The claims that interactivity can augment the user's understanding of the data and overcome some of the limitations of the representation still lack concrete evidence. One of the reasons for this is the complexity generated by all of the different interaction techniques that can be used. Therefore, we need a solid base that allows us to guide future studies that hopefully will establish guidelines for the effective use of interactivity in visualization.

This paper also aims to renew the interest on interactivity as a complex component that should not be forgotten by the Infovis domain.

References

1. R. Amar, J. Eagan, and J. Stasko. Low-level components of analytic activity in information visualization. In *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*, pages 111–117, Oct 2005.

2. D. Bargeron and T. Moscovich. Reflowing digital ink annotations. In *Proceedings of the SIGCHI Conference on Human*

[3] http://edition.cnn.com/SPECIALS/war.casualties/
[4] http://www.publicinsightnetwork.org/budgethero/
[5] http://projects.propublica.org/graphics/heartsaver
[6] http://www.lastampa.it/medialab/webdoc/la-stampa-academy/world-data-cup/eng
[7] http://playspent.org

Factors in Computing Systems, CHI '03, pages 385–393, New York, NY, USA, 2003. ACM. ISBN 1-58113-630-7.

3. I. Bogost, S. Ferrari, and B. Schweizer. *Newsgames*. The MIT Press, 2010.

4. A. Bruns. *Blogs, Wikipedia, Second Life, and Beyond: From Production to Produsage*. Digital formations. Peter Lang, 2008. ISBN 9780820488660.

5. S. K. Card, J. D. Mackinlay, and B. Shneiderman. *Readings in information visualization: using vision to think*. Morgan Kaufmann, 1999.

6. B. Craft and P. Cairns. Beyond guidelines: what can we learn from the visual information seeking mantra? In *Information Visualisation, 2005. Proceedings. Ninth International Conference on*, pages 110–118, July 2005.

7. S. Deterding, D. Dixon, R. Khaled, and L. Nacke. From game design elements to gamefulness: Defining "gamification". In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, MindTrek '11, pages 9–15, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0816-8.

8. S. Deterding, M. Sicart, L. Nacke, K. O'Hara, and D. Dixon. Gamification. using game-design elements in non-gaming contexts. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '11, pages 2425–2428, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0268-5.

9. N. Diakopoulos. Design challenges in playable data. In *CHI Workshop on Gamification*, 2011.

10. A. Dix and G. Ellis. Starting simple: Adding value to static visualisation through simple interaction. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, AVI '98, pages 124–134, New York, NY, USA, 1998. ACM.

11. A. Figueiras. A typology for data visualization on the web. In *Information Visualisation (IV), 2013 17th International Conference*, pages 351–358, July 2013. doi: 10.1109/IV.2013.45.

12. A. Figueiras. How to tell stories using visualization. In *Information Visualisation (IV), 2014 18th International Conference on*, pages 18–18, July 2014.

13. J. Heer. Socializing visualization. In *In CHI 2006 Workshop on Social Visualization*, 2006.

14. J. Heer, F. B. Viégas, and M. Wattenberg. Voyagers and voyeurs: Supporting asynchronous collaborative visualization. *Commun. ACM*, 52(1):87–97, Jan. 2009. ISSN 0001-0782.

15. J. Hullman and N. Diakopoulos. Visualization rhetoric: Framing effects in narrative visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):2231–2240, 2011. ISSN 1077-2626.

16. P. Isenberg, N. Elmqvist, J. Scholtz, D. Cernea, K.-L. Ma, and H. Hagen. Collaborative visualization: definition, challenges, and research agenda. *Information Visualization*, 10 (4):310–326, 2011.

17. H. Jenkins. *Fans, Bloggers, and Gamers: Exploring Participatory Culture*. New York University Press, 2006. ISBN 9780814742853.

18. S. D. Kamvar and J. Harris. We feel fine and searching the emotional web. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, WSDM '11, pages 117–126, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0493-1.

19. D. Keim. Information visualization and visual data mining. *Visualization and Computer Graphics, IEEE Transactions on*, 8(1):1–8, 2002. ISSN 1077-2626.

20. L. Li, F. Li, and R. Lau. A trajectory-preserving synchronization method for collaborative visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 12(5):989–996, Sept 2006. ISSN 1077-2626.

21. R. R. Raje, M. Boyles, and S. Fang. Cev: collaborative environment for visualization using java rmi. *Concurrency - Practice and Experience*, 10(11-13):1079–1085, 1998.

22. A. Satyanarayan and J. Heer. Authoring narrative visualizations with ellipsis. *Computer Graphics Forum*, 33(3):361–370, 2014. ISSN 1467-8659.

23. E. Segel and J. Heer. Narrative visualization: Telling stories with data. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):1139–1148, 2010. ISSN 1077-2626.

24. B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Visual Languages, 1996. Proceedings., IEEE Symposium on*, pages 336–343. IEEE, 1996.

25. F. Viegas, M. Wattenberg, F. van Ham, J. Kriss, and M. McKeon. Manyeyes: a site for visualization at internet scale. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6):1121–1128, Nov 2007. ISSN 1077-2626.

26. F. Viegas, M. Wattenberg, and J. Feinberg. Participatory visualization with wordle. *Visualization and Computer Graphics, IEEE Transactions on*, 15(6):1137–1144, Nov 2009. ISSN 1077-2626.

27. M. Wattenberg. Baby names, visualization, and social data analysis. In *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*, pages 1–7, Oct 2005.

28. M. Wattenberg and J. Kriss. Designing for social data analysis. *Visualization and Computer Graphics, IEEE Transactions on*, 12(4):549–557, July 2006. ISSN 1077-2626.

29. W. Willett, J. Heer, J. Hellerstein, and M. Agrawala. Commentspace: Structured support for collaborative visual analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 3131–3140, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0228-9.

30. J. S. Yi, Y. ah Kang, J. Stasko, and J. Jacko. Toward a deeper understanding of the role of interaction in information visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6):1224–1231, 2007. ISSN 1077-2626.

Appendix II

Typology initial tables

Visualization Title	Source	Playable	Non-Playable
007, la venganza	El Mundo	•	
100 Years of World Cuisine	100yearsofworldcuisine.com		•
9/11 anniversary: 10 years of terror attacks	The Guardian	•	
2012: The End ofthe World?	Information is Beautiful		•
Articles of War	Information is Beautiful		•
A Week of Checkins on the Path to One Billion	Foursquare	•	
Battle of the Book Worms	Good.is		•
Because Every Country is the Best at Something	Information is Beautiful	•	
Boundary changes: the first map of England's new constituencies	The Guardian	•	
British History Timeline	BBC	•	
British troops killed in Afghanistan	The Guardian	•	
Confira a evolução da população do mundo desde 1950	Época	•	
Conversation Cloud	The Economist	•	
Death penalty in the US mapped	The Guardian	•	
Death penalty statistics, country by country	The Guardian		•
Education 2011	The Guardian		•
Everything Makes Them Sick	The New York Times	•	
Evolution of the Web	Google	•	
Faces of the Dead	The New York Times	•	
Global cities of the future	McKinsey Quarterly	•	
Ground Zero Now	The New York Times	•	
Home and Away: Iraq and Afghanistan War Casualties	CNN	•	
How Different Groups Spend Their Day	The New York Times	•	

Visualization Title	Source	Narration Elements								Ordering	
		Captions	Annotations	Introductory Text	Accompaning Article	Text	Title	Audio Narration	Video Narration	User Directed Path	Linear
007, la venganza	El Mundo	●	●				●				●
100 Years of World Cuisine	100yearsofworldcuisine.com	●					●			●	
9/11 anniversary: 10 years of terror attacks	The Guardian	●	●				●				●
2012: The End of the World?	Information is Beautiful	●	●				●			●	
Articles of War	Information is Beautiful	●	●				●			●	
A Week of Check-ins on the Path to One Billion	Foursquare	●					●				●
Battle of the Book Worms	Good.is	●		●			●			●	
Because Every Country is the Best at Something	Information is Beautiful	●					●			●	
Boundary changes: the first map of England's new constituencies	The Guardian	●		●			●			●	
British History Timeline	BBC	●	●				●			●	
British troops killed in Afghanistan	The Guardian	●		●			●			●	
Confira a evolução da população do mundo desde 1950	Época	●		●			●			●	
Conversation Cloud	The Economist	●	●				●			●	
Death penalty in the US mapped	The Guardian	●	●	●			●			●	
Death penalty statistics, country by country	The Guardian	●		●	●		●			●	
Education 2011	The Guardian	●	●		●		●			●	
Everything Makes Them Sick	The New York Times	●	●	●			●				●
Evolution of the Web	Google	●	●	●			●			●	
Faces of the Dead	The New York Times	●	●	●			●			●	

Visualization Title	Visual Elements																										
	Timeline	Photograph	Bar chart	Pie chart	Doughnut chart	Line chart	Bubble chart	Area chart	Histogram	Network diagram	Venn diagram	Tree diagram	Object size representing quantities	Map	Bubble map	Pictogram	Drawing	Speech Balloon	Model	Table	Logos	Video	Scale	Exploded view	Tag cloud	Pyramid	Animation
007, la venganza	●	●	●																								●
100 Years of World Cuisine		●		●		●							●														
9/11 anniversary: 10 years of terror attacks	●														●												●
2012: The End of the World?			●					●						●		●	●										
Articles of War	●												●														
A Week of Check-ins on the Path to One Billion			●	●										●							●	●					●
Battle of the Book Worms									●							●	●										
Because Every Country is the Best at Something														●													
Boundary changes: the first map of England's new constituencies														●				●									●
British History Timeline	●	●																●									●
British troops killed in Afghanistan		●														●											●
Confira a evolução da população do mundo desde 1950	●					●								●													
Conversation Cloud										●	●		●														●
Death penalty in the US mapped														●				●									●
Death penalty statistics, country by country	●												●			●											
Education 2011			●										●			●											

Visualization Title	Source	Genres										
		Sequential Graphic	Slide Show	Chart/Diagram	Map	Tag Cloud	Model	Drawing	Video/Animation	Photograph	Poster	Game
007, la venganza	El Mundo	●										
100 Years of World Cuisine	100yearsofworldcuisine.com									●		
9/11 anniversary: 10 years of terror attacks	The Guardian	●			●							
2012: The End of the World?	Information is Beautiful			●							●	
Articles of War	Information is Beautiful										●	
A Week of Check-ins on the Path to One Billion	Foursquare				●				●			
Battle of the Book Worms	Good.is			●							●	
Because Every Country is the Best at Something	Information is Beautiful				●							
Boundary changes: the first map of England's new constituencies	The Guardian				●							
British History Timeline	BBC	●										
British troops killed in Afghanistan	The Guardian									●		
Confira a evolução da população do mundo desde 1950	Época	●			●							
Conversation Cloud	The Economist			●								
Death penalty in the US mapped	The Guardian				●							
Death penalty statistics, country by country	The Guardian			●								
Education 2011	The Guardian			●							●	
Everything Makes Them Sick	The New York Times									●		
Evolution of the Web	Google	●										
Faces of the Dead	The New York Times									●		

Appendix III

Focus group rating sheet

	Visualizations	Comprehension										Likability										Navigation									
1	Death penalty statistics, country by country	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
2	Death penalty in the US mapped	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
3	Evolution of the Web	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
4	How Many Households Are Like Yours?	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
5	Ground Zero Now	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
6	Home and away	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
7	Faces of the dead	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
8	British troops killed in Afghanistan	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
9	What does china censor online?	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
10	How Local News Is Going Mobile	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
11	How Much CO2?	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10

Table III.1: Focus group rating sheet

Appendix IV

Focus group semistructured questions

1. Which one do you prefer?
   - Non-Playable (Death penalty statistics, country by country)
   - Playable (Death penalty in the US mapped)
2. Which of the following elements do you think helps to tell the story better? And which did you enjoy more?
   - Introductory text (Evolution of the Web)
   - Accompanying article (How Many Households Are Like Yours?)
   - Audio narration (Ground Zero Now)
3. Taking in to account these 3 visualizations about the same topic, which do you think tells the story better?
   - Home and away: Iraq and Afghanistan war casualties
   - Faces of the dead
   - British troops killed in Afghanistan
4. Which one is more visually attractive?
   - Home and away: Iraq and Afghanistan war casualties
   - Faces of the dead
   - British troops killed in Afghanistan
5. What story is being told? Can you understand it? Did you enjoy it? What would you change?
   - What does china censor online?

- How Local News Is Going Mobile: Could the iPad Be the New Sunday Press?
- How Much CO_2?

6. Which visualization/s did you like the most and why?

2016

How to tell stories using visualization:
strategies towards Narrative Visualization

Ana Figueiras