

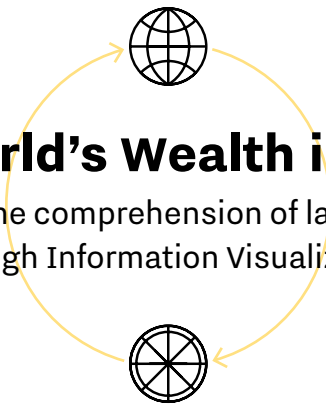


The World's Wealth in Pizza:

Improving the comprehension of large numbers
through Information Visualization



Lisa Staudinger



The World's Wealth in Pizza:

Improving the comprehension of large numbers
through Information Visualization

Lisa Staudinger

Master's thesis submitted for the degree of Master of Arts

Aalto University
School of Arts, Design and Architecture
Department of Media
Visual Communication Design

April 2018

Supervisor: Professor Rupesh Vyas, Aalto University
Advisor: Docent Tomi Kauppinen, Ph.D., PD

Abstract

Author: Lisa Staudinger

Title of thesis: The World's Wealth in Pizza: Improving the comprehension of large numbers through Information Visualization

Department: Department of Media

Degree programme: Visual Communication Design

Year: 2018

Number of pages: 102 + 16

Language: English

Extreme numerical magnitudes are part of our daily lives, from science to economics to politics. Specifically for large monetary measures, however, there are no comprehensive models for visualization practitioners to promote their understanding. Previous works on this topic have provided a framework for the visual depiction of complex measures but did not assess its effectiveness in communicating the real magnitude of the presented measures. In this thesis I bring together findings from Information Visualization and numerical cognition to extend the existing framework and assess the effects of different strategies, with a focus on monetary measures. For this, I created three visualization prototypes and conducted a series of user tests focused on insight creation. User tests highlighted advantages and disadvantages for different strategies and yielded various findings for their implementation in Information Visualisation.

Keywords: information visualization, numerical cognition, visual strategies, large numbers, monetary measures, insight-based evaluation

Acknowledgements

I want to thank the following people who have contributed in their various ways to this thesis:

My advisor Tomi Kauppinen for his enthusiasm, support and ideas throughout the whole thesis process. This thesis would not been able without his help. My supervisor Rupesh Vyas, Masood Masoodian and Markku Reunanen for their help and comments. Marija for laughing, crying, loving, hating, dancing, hugging and punching some leaves with me. Maja for being my shining thesis role model and for proofing the print files. My boyfriend for supporting and encouraging me in so many ways and for proofreading my thesis. All my former classmates and teachers at Aalto University who made these last few years such an enjoyable experience. All my test participants for their time and enthusiasm. To all the persistent dwellers of the 5th floor: you can do it. Lastly I want to deeply and gratefully acknowledge the support and funding I received for the completion of this thesis from Aalto Online Learning (A!OLE) (onlinelearning.aalto.fi).

Vor Allem möchte ich meinen Eltern für ihre liebevolle Unterstützung während der letzten vier Jahre danken, die es mir ermöglicht hat, hier in Finnland zu studieren. Diese Erfahrung hat mich als Person wachsen lassen und hat mir völlig neue Blickwinkel auf meinen Beruf und die Welt eröffnet.

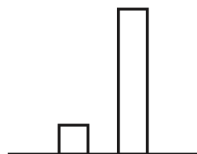
ABSTRACT	3
ACKNOWLEDGEMENTS	5
1 INTRODUCTION	9
1.1 Previous & related research	11
1.2 Research question & scope	12
1.3 Thesis structure	13
2 THEORETICAL BACKGROUND	15
2.1 Information Visualization	17
What is Information Visualization?	17
Why present information visually?	18
Visualization for comprehension	21
2.2 Number Cognition	30
Number processing	30
Cognitive limits	34
2.3 Summary	36
3 FRAMEWORK	37
3.1 Concrete Scales	38
3.2 Reflections on Concrete Scales	40
3.3 Adapted Framework	44
3.4 Summary	60
4 PROTOTYPES	61
4.1 Stacking Bills	63
4.2 Money Clock	65
4.3 Buying Power	67
5 EVALUATION	69
5.1 Evaluating insight in Information Visualization	70
5.2 User testing	73
5.3 Results	78
5.4 Discussion of results	86
6 CONCLUSION	89
6.1 Contribution	90
6.2 Limitations	90
6.3 Future Research	91
7 BIBLIOGRAPHY	95
8 IMAGE SOURCES	101
9 APPENDIX	103

1 Introduction

»To people whose minds go blank when they hear something ending in ›illion‹, all big numbers are the same, so that exponential explosions make no difference. Such an inability to relate to large numbers is clearly bad for society. It leads people to ignore big issues on the grounds that they are incomprehensible.« (Hofstadter 1982, p.117)

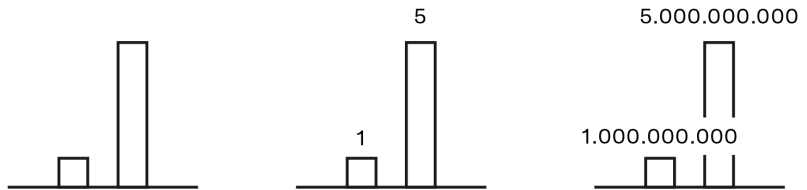
Large numbers of all types rule our lives. From trillions of debt to millions of views and hundreds of lightyears. Undoubtedly, as these numbers shape our day-to-day life in crucial ways, handling them should be second nature to all of us. However, as numerous examples, and our own personal experiences show, most of us have a poor understanding of the true size of these phenomena. Measures outside our immediate personal experience carry little to no meaning for the layman, leading to misinterpretations, misjudgments, or in the worst case, complete disregard of important issues. We fall prey to what Saul Wurman calls the »uh-huh« syndrome: Instead of admitting our lack of understanding, we nod our heads, repeating »uh-huh, uh-huh«, desperately trying to hide our shortcomings. »This only prevents us from learning and exacerbates our suspicion that everyone else knows more than we do« (Wurman 2001, p.56). Traditional Information Visualization often falls short in explaining the true meaning of big numbers, assuming that statistically correct visualizations satisfy the requirements of truthfulness and are therefore enough to create understanding in readers/viewers.

Consider the following bar chart:



A chart like this can only present the relative difference between two numbers, but not express their absolute size.

It makes no difference if the bars represent the numbers one and five or one-billion and five-billion. They are completely interchangeable.



Consider this chart in the context of real life. Imagine those numbers represented the victims of an imaginary war. It makes a huge difference whether 5 people died or 5 billion. The visualization can, however, not express this fact.

To make numbers more understandable, Michael Blastland, journalist for the Guardian, host of the BBC Radio 4 program *More or Less* and author of *The numbers game*, recommends reconnecting numbers with experiences from real life, and urges that the human scale is an important factor for making numbers meaningful. Examples like the *Dictionary of Numbers* and *Is That A Big Number?* are attempts at making measures more relatable by giving relatable context. The chrome browser extension *Dictionary of Numbers*¹ tries to help readers make sense of numbers they encounter, by automatically inserting explanations of numbers in »human terms« into a website's text. *Is That A Big Number?*² is a website that »puts numbers into a context we can understand«. Upon entering a number, the website gives examples of magnitudes that are smaller, bigger and equivalent. Both of these examples are praiseworthy efforts, but limited to textual representation. Furthermore, there is no study on their effectiveness.

¹ dictionaryofnumbers.com

² isthatabignumber.com

Despite its importance, the Information Visualization community has, as of yet, paid little attention to this disconnect between large numbers and our mental capabilities.

1.1 Previous & related research

Aside from Information Visualization, the problem of presenting extreme magnitudes beyond the human experience has been extensively studied in connection to science, technology, engineering and mathematics (STEM) education. Different disciplines deal with unfamiliar measures on different scales, from the extremely small (e.g. nanoscopic sizes) to the extremely big (e.g. geological timeframes).

Studies such as Resnick et al.'s (2017a) examined analogy, active prediction and corrective feedback as methods to learn about geological times. Similarly, Parker (2011) proposes a timeline analogy based on Google Earth to represent both relative and absolute time. Focusing on nanoscopic sizes, Song and Quintana (2012) investigated whether the understanding of nanoscopic sizes can be aided through multimodal *temporal-aural-visual* representations (TAVR), and which modalities are the most effective in supporting students' conceptualization of sizes. Studying people's preferences for the depiction of »small things«, Ma (2007) describes strategies derived from a study conducted with visitors of a science museum. Categorizing the collected examples, Ma identified seven different types of strategies: Object-to-Object Comparison, Actual size, Analogy, Notation (anchoring), Zoom (containment relationships), (use of) Technology and Other.

Related to the context of journalism, Barrio et al.'s study (2016) investigated how the comprehension of unfamiliar numbers in online news articles can be improved. They developed a framework of comparative sentences, termed *perspectives*, that use percentages and ranks to give context for numerical measures. These perspectives are composed of a scaling factor, an attribute and a reference entity, e.g. »To put this into perspective, 3,400 miles is about 1.27 times larger than the width of the continental United States of America« (Barrio, Goldstein & Hofman 2016). Barrio et al. found that their sentence structures substantially improved people's performance in remembering previously presented measures, estimate unfamiliar values, and detect errors in manipulated measures. This study focuses only on textual strategies, however, and is limited in the types of translation it investigates.

Most closely related to the context of this thesis is Chevalier et al.'s (2013) work on *concrete scales*. Concrete scales explain unfamiliar or extreme measures by re-expressing or comparing them through physical objects that are more familiar to the reader/viewer. For this, Chevalier et al. introduced a practical framework for the design and analysis of such compositions. While this work concentrates on making unfamiliar measures understandable through different visual strategies, it offers no evaluation on the effectiveness of these strategies. Furthermore, the framework shows several inconsistencies and general shortcomings. As such, Chevalier et al.'s framework was used as the basis for an adapted, more-inclusive framework. Both the concrete scale framework and the adapted framework are discussed in further depth in the Framework chapter.

1.2 Research question & scope

As research on this topic is sparse, the goal of this thesis is to take a first step towards exploring possibilities for improving the comprehension of large numbers through Information Visualization methods and critically evaluating different translation techniques. The focus of this work is not on developing novel visualization methods, but on creating an understanding of the underlying processes of (large) number comprehension and how these can be supported through the tools of Information Visualizations. I believe that such an effort can enable readers to truly understand the information presented in our visualizations and help tackle problems such as the rise of »fake news«.

Consequently, my research question for this thesis is:
 In the context of Information Visualization, **how effective are various strategies in supporting the comprehension of large (monetary) numbers?**

To answer this question I produced three visualization prototypes based on theoretical and practical research and tested those strategies in an insight-based evaluation approach, comparing:

- how many instances of insight occur
- what types of insights occur
- when they occur

I chose to focus on monetary measures for my prototypes and evaluation, since this is the area in which most people would commonly encounter »large numbers« such as millions, billions and trillions. Furthermore, money exists both as a theoretical concept and a physical object, which makes it an interesting and challenging subject.

1.3 Thesis structure

This thesis is structured as follows:

- In the theoretical background chapter, I first introduce theories on information comprehension from both Information Visualization and numerical cognition literature.
- Building upon these theories, in the next chapter I review and extend a framework for presenting extreme magnitudes by Chevalier et al. (2013).
- In the prototypes chapter, I present three visualization prototypes that I designed based upon the revised framework and findings from a corpus of existing visualizations.
- The effectiveness of these prototypes in regards to insight creation is then evaluated in the next chapter.
- I close the thesis with a summary of my findings, a discussion on my work and an outlook for future research.

Note: The pronoun »they« is purposefully used throughout this thesis as a gender-neutral indicator for both a single person and a group of people.

2 Theoretical Background

In this chapter I discuss theories on information comprehension from both Information Visualization and numerical cognition literature. I start with defining the meaning of »large numbers« and the meaning and goals of Information Visualizations. I then elaborate on three concepts from Information Visualization (Storytelling, Empathy and Interactivity) that have the potential to support the comprehension of complex information. Focusing not only on the presentation of information, but also the cognitive processes behind, I then introduce relevant theories from number cognition and processing.

LARGE NUMBERS?

When talking about large numbers and our problems in regard to their comprehension, it is important to define what a *large number* is first.

On the most basic level, the largeness of a number is relative to its context. Ten is a large number in a set containing otherwise only single digit numbers. Fifty is an incredibly big number for a preschooler, who has only learned numbers up to ten. Disregarding these numbers that are relatively large, it seems that humans have problems grasping the magnitude of numbers once they exceed a certain absolute size. We might still be able to understand and manipulate these numbers in an abstract, mathematical sense, but our brains fail to visualize the actual quantity of the number any longer. The Encyclopædia Britannica defines *large numbers* as »numbers above one million that are usually represented either with the use of an exponent such as 10^9 or by terms such as billion or thousand millions that frequently differ from system to system« (Encyclopædia Britannica 2011). Professor of cognitive science Douglas R. Hofstadter (Hofstadter 1982, p.124) notes in his 1982 essay *On number numbness*, that quantities exceeding 10^4 or 10^5 (ten or one-hundred thousand) surpass his ability to visualize. In their study on large number magnitude estimation Landy et al. (2017) define their *large numbers* as numbers in the range of around 10^6 to 10^{12} , i.e. one million to one trillion. Large enough to fall outside (most) direct human experience, but small enough to play a crucial role in our everyday life, from science to politics.

In the context of this thesis and its focus on monetary measures, I am defining *large numbers* as values in the range of around 10^6 to 10^{12} (million to trillion). I arrived at this specific range considering the above-cited references, as well as personal observations of articles from major U.S and U.K newspapers (theguardian.com, nytimes.com, etc.). Monetary numbers that are big enough to be newsworthy seem to lie in the range from one million to several trillions (in english-language news pieces)³. The highest monetary numbers I encountered in this context reported the U.S national debt, which currently stands at about \$20 trillion, with an annual budget deficit of around \$1 trillion (nytimes.com, 2015).

³ *This range does not necessarily hold true to countries outside the U.S. and Europe. Dealing with numbers in the millions is quite common, for example, in Indonesia, where the rent for a reasonably priced apartments can be several million Indonesian Rupiah per month. (Cheek 2017)*

2.1 Information Visualization

What is Information Visualization?

The definitions of *Information Visualization* are manifold and dramatically diverse. Although the field has existed for a considerable amount of time, no clear, universal definition of the term has emerged. The reason being that Information Visualization touches many disciplines, and different stakeholders study it through different lenses. As a result, every piece of writing on Information Visualization has to start with a definition of the term itself, to clarify the viewpoint they are applying. This thesis is no exception.

A rather popular definition of Information Visualization is Ben Shneiderman's: Information Visualization is »[t]he use of computer-supported, interactive, visual representations of abstract data to amplify cognition« (Card, Mackinlay & Shneiderman 1999, p.7) This definition is very narrow, excluding printed, non-interactive material, and presentation of concrete phenomena. Consequently, several authors have attempted to find more general definitions that encompass a wider range of approaches, or have tried to demarcate those approaches even further. Luca Masud and colleagues (2010), for example, defined six disciplines of visualization⁴, each discipline being concerned with slightly different aspects of »information« and using different tools to achieve their goals.

1. Data Visualization
2. Information Visualization
3. Scientific Visualization
4. Information Aesthetics
5. Infographics
6. Knowledge Visualization

Robert Spence offers a view that is less tied to the specifics of the producer, the artifact or the audience. According to Spence (2014b),

- ⁴ *The six disciplines of visualization according to Masud et al. (2010).*

Visualization is the formation of a mental model of something.

Visualization is not what is presented on a computer screen or on paper-visualization happens in the mind of the viewer. Whatever is graphically displayed is solely meant to support a viewer's insight or understanding. This aligns with Shneiderman's statement: »The purpose of visualization is insight, not pictures« (Card, Mackinlay & Shneiderman 1999, p.6). Consequently, Information Visualization should be defined as the mental process of deriving information (insight, understanding etc.) from encoded data.

This description of visualizing information – to support human mental process – is most relevant to this thesis, and in the context of this thesis I will define Information Visualization as (mostly) **visual presentations that aim at supporting the understanding of data.**

Why present information visually?

As Barrio et al.'s (2016) studies have shown, merely using text to elaborate on a magnitude is already enough to show significant improvements in recall, estimation, and error-detection (all of which they define as aspects of comprehension). So why graphically visualize these strategies? Why go through the trouble of visualizing these strategies further? What is the added value?

In *Information Visualization: Perception for Design*, Colin Ware (2004) discusses the relationship between visual and textual information; From the distinct cognitive systems used to process these different types of information, to their interplay in composite visualizations. Ware provides a list of key advantages for each type of information (Ware 2004, p.304):

Images are better for	Text is better for
showing structural relationships	describing abstract concepts
presenting detail and appearance	conveying procedural information
localization information	communicating program logic
recollection	formulating conditions

As can be seen from this list, natural language and images address different needs that can even be complementary in some cases. The same holds true not only for textual, but also for numerical information, henceforth simply referred to as data.

An example often employed to show how images can help viewers reveal certain aspects of data is *Anscombe's Quartet*. Developed by Francis Anscombe in 1973, the quartet consists of four datasets that have nearly identical summary statistics, but when plotted, reveal enormous differences between the datasets (Wikipedia):

I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

FIG.1. Anscombe's quartet as a table. Note how the summary statistics are (nearly) identical for all four datasets.

Mean of x = 9.0 Sample variance of x = 11 Correlation between x and y = 0.816
 Mean of y = 7.50 Sample variance of y = 4.125 Linear regression line: $y = 3.00 + 0.500x$

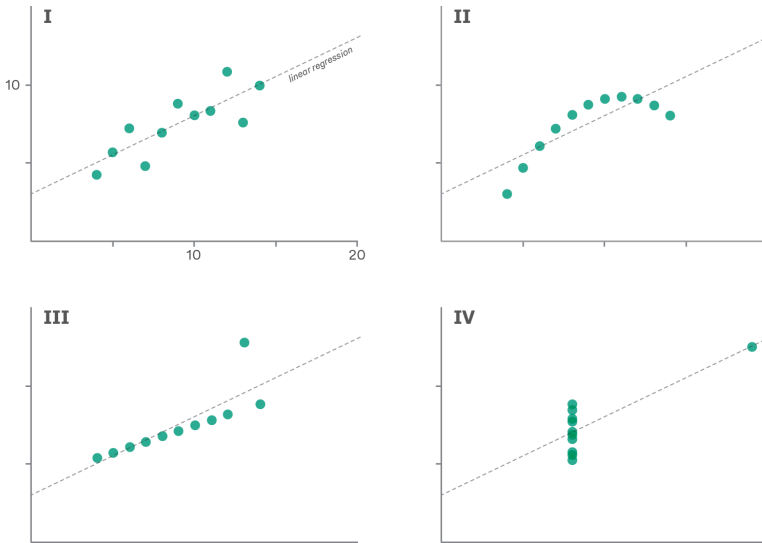


FIG.2. Anscombe's quartet plotted as scatterplots.

Just as summary statistics can obscure the true nature of data, number words (e.g. million, billion, trillion), which »sum up« the value of a number, can also conceal the true magnitude of those numbers. Consider this example, which one might encounter in the news on any given day:



FIG.3. Extract from 1000 Times by Randall Munroe.

An enormous difference in scale like this (170 billion = 170.000 million) might be easily missed when reading a news piece. One might be skimming the text and simply mis-read the words, or one might be unaware of the powers these words represent. By visualizing these two numbers, however, their difference in size becomes painfully apparent⁵:

165 million |
170 billion |

⁵ *The opposite can be true as well, however, as we have seen by the example I gave in the introduction.*

All these examples illustrate how visualizations act as metaphors that can »decrease the gap between the data and user's mental model of it« (Yi et al. 2008). They facilitate the integration of presented data into existing real-world experiences and knowledge, thereby easing the viewer's cognitive load.

These examples should not, however, lead to the conclusion that one mode of presentation (visual, textual, numerical etc.) is inherently better than another. Rather it shows that each has its own advantages and can compensate for the weaknesses of other modes. This coincides with the claim of multimedia, which states that giving information in more than one medium of communication leads to better comprehension (Ware 2004, p.306). Eventually, providing multiple presentations of information, be it by visualizing a high-dimensional data set or just explaining a single number through a metaphor, supports the comprehension

of this information and allows for the accommodation of readers/viewers with varying skill-sets.

Visualization for comprehension

The disconnect between big numbers and our personal experience is generally not a prominent topic of discussion in the Information Visualization literature (or community). The issue might be mentioned briefly with some concern, but is usually not further addressed⁶. Therefore this section will describe wider principles employed in the field of Information Visualization that improve comprehension of data.

⁶ *The problem is much more discussed in the context of journalism, and several books have tried to give recommendations for presenting numbers in more relatable ways see Blastland, Dilnot (2009) and Paulos (2001).*

STORYTELLING

»[...] I've always been very sensible to the way stories, parables, vignettes, and sometimes even jokes help put formal mathematics into context, illustrate its limitations, and emphasize what should be a truism: that numbers and statistics always require interpretation.« (Paulos 2001, p.xiv)

In recent years storytelling, story structure, and narrative have become topics of interest in Information Visualization research and discourse. A story is »an ordered sequence of steps, each of which can contain words, images, visualizations, video, or any combination thereof« (Kosara, Mackinlay 2013). Stories can explain and enlighten, they create images and emotion, and they connect the storyteller (designer) with the audience (reader/viewer) (Lidwell, Holden & Butler 2010, p.230). Some have argued that stories can provide meaningful, relatable context for hard data, thereby supporting the understanding and retention of information (Kosara, Mackinlay 2013). As such, storytelling is a method for explaining and presenting existing knowledge rather than analyzing unknown data. This also means that narrative visualizations are necessarily the product of individual, editorial decisions, much more, and much more openly than exploratory or »neutral« charts (Hullman, Diakopoulos 2011).

Throughout history, various types of visualizations have been used to present, prove, explain and educate. In *Envisioning Information*, Edward Tufte (1997, p.106) describes how small multiples are used to show narrative sequences in Christiaan Huygens work *Systema Saturnium*, published in 1659. In his treatise, the Dutch physicist, mathematician, astronomer, and inventor Christiaan Huygens interweaves small diagrammatic images and notes about his observations in an almost comic-like fashion to report and explain his new discoveries about the structure of Saturn and its movement around the sun.

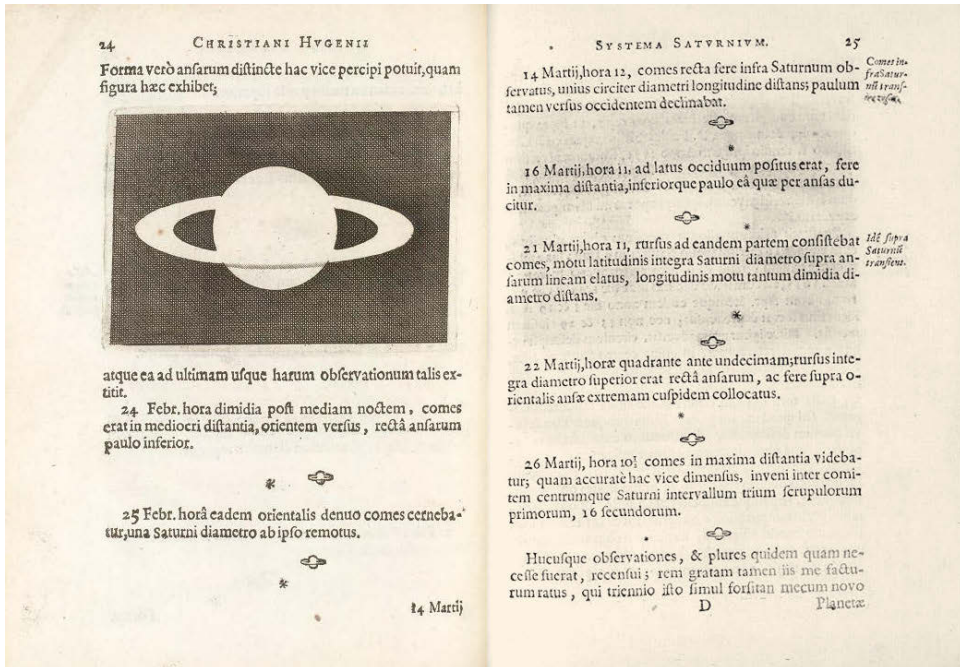


FIG. 4. A page from Christiaan Huygens' *Systema Saturnium*.

In a similar manner, two of the most praised visualizations in information visualization history, Florence Nightingale's rose chart of deaths during the Crimean War, and Charles Minard's maps of Napoleon's Russian campaign in 1812, were created to explain complex subjects, rather than to analyze the underlying data (Kosara, Mackinlay 2013).

More recently, visualizers have begun to analyze the underlying structures of narrative visualizations. Segel and Heer (2010), for example, classified different genres, visual narratives tactics and supportive mechanisms of *data stories*, i.e. data-driven

journalistic pieces. They present seven genres of narrative visualizations (Magazine Style, Annotated Chart, Partitioned Poster, Flow Chart, Comic Strip, Slide Show, Film/Video/Animation) and discuss three story-structures for interactive visualizations that balance author-driven and reader-driven approaches, such as the now famous *Martini Glass* structure.

Hullman and Diakopoulos (2011) explored rhetorical techniques in narrative visualizations, and how these interact with »extra-representational influences«, such as the reader/viewer's personal interpretation of the information. They further discuss different rhetoric techniques connected to editorial choices that are made in regards to the data, visual representation, textual annotations, and interactivity.

In their paper *Storytelling: The Next Step for Visualization*, Kosara and Mackinlay (2013) describe the current state of storytelling in information visualizations and furthermore propose several research directions. Consequently, they touch upon some principal aspects of narrative visualizations. These aspects include: memory, context, embellishments, interaction, annotations, and collaboration.

Most recently, Robert Kosara (2017) presented a new argument structure for narrative visualizations, consisting of claim, evidence, explanation, and conclusion. By recombining and/or repeating several of these elements, different structures and levels of complexity can be achieved in a story.

EMPATHY

The issue of connecting a concrete meaning with large numbers has recently gained a moderate amount of attention in the visualization community in connection with social, political, and activist topics. Sarah Slobin (2014) appropriately articulated the problem in the title of her article: *What If the Data Visualization Is Actually People?*. In this context visualizers put forth empathy as a tool to connect readers/viewer with the human reality represented by numbers. It should be noted, however, that information visualizers generally use the term *empathy* rather informally, without going deeper into its various scientific meanings. Generally, in this context, the goal of empathy in visualization is to convey emotion and/or allow personal connection to the data.

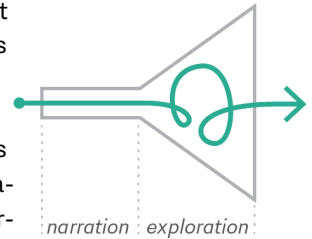


FIG. 5. The Martini Glass structure. After an author-driven introduction, the reader/viewer can freely explore the data.

Although not specifically targeted at Information Visualization, Paul and Scott Slovic (2015) argue in *Numbers and Nerves* that affect (in this case translated to emotions or compassion) is essential to making judgements and decisions about any type of information, especially when it concerns social, environmental or political issues. When we experience statistical numbing and compassion fatigue, affect has the power to convey the meaning of numbers and motivate action. According to Slovic and Slovic one way to achieve affect or compassion is through emotional, relatable framing, storytelling, and the use of personal stories. Similar techniques have been proposed in the context of Information Visualization:

Jacob Harris (2015), software architect at the New York Times Newsroom, recommends the use of a strategy he calls *Putting People First*. When the data is about people, rather than showing aggregated numbers or statistics, show the individual person through images, personal details or stories. Though not an Information Visualization in the traditional sense, the *Stolpersteine* project initiated by German artist Gunter Demnig uses an approach like this to commemorate the victims of National Socialism in Europe during World War II.



FIG. 6. Stolpersteine, literally »stumbling stones«, are cobblestone-sized brass plates commemorating the names and lives of victims of Nazi persecution and extermination. This one reads: Here lived Else Liebermann von Wahlendorf. Native Dutch. Year 1876. Humiliated-Vilified. Dead 8.1.1943

Another strategy Harris proposes is the use of so-called *Wee-people*; using human figures instead of abstract shapes. This strategy has similarities with a chart type known as *Unit Chart*, *Pictograph Chart*, *Dot Plot*, *Tally* or *Pixel Chart*, in which an icon or graphic form represents a concrete unit. The most prominent example for this chart type is the *ISOTYPE system*. The ISOTYPE system likewise demands the use of human shapes when humans are the focus of the data, but those shapes mostly represent a larger number of units (e.g. each man icon stands for 100 million men) (Neurath, Neurath 1980). In contrast, Harris argues that, to evoke empathy, figures should not stand for aggregates. One figure should always equal one person.

To evoke empathy in readers/viewers, Harris lastly recommends the idea of *Far and Near*. This concept involves presenting data at different levels of depth, from a general (far) and a specific (near) view. While the far view can give context and broader meaning to the data, the near view connects with the reader/viewer on a personal level. It can also mean breaking down the data into smaller parts that are easier to understand. Both methods allow the reader/viewer to understand the overall issue through a more relatable vantage point.

Mushon Zer-Aviv (2015) elaborates this idea through the example of Perisopic's visualization of gun murders in the U.S.⁷ In this interactive visualization, the lives of gun and firearms victims are presented as bright dots that shoot up in an arc, dropping suddenly to the ground, as they die. The »life-arc« itself, however, continues, giving a glimpse into how much longer each person could have lived. The visualization begins by drawing a handful of individual life-arcs, labeled with details such as »Alexander Lipkins, killed at 29« and »could have lived to be 93«. It then starts rapidly amassing severed life-arcs, adding up lives lost and the potential number of stolen years. Once it reaches the total (11,419 people killed in 2013), it allows for a deeper exploration of the individual lines, linking to news stories about the cases. In this visualization, the single, specific life-arc functions as an emotional lead-in to the bigger picture of gun-related deaths in America. It also invites comparison to one's own personal »life-arc«. It is an example of how empathic framing can be used to introduce *far view* data and topics.

⁷ guns.perisopic.com

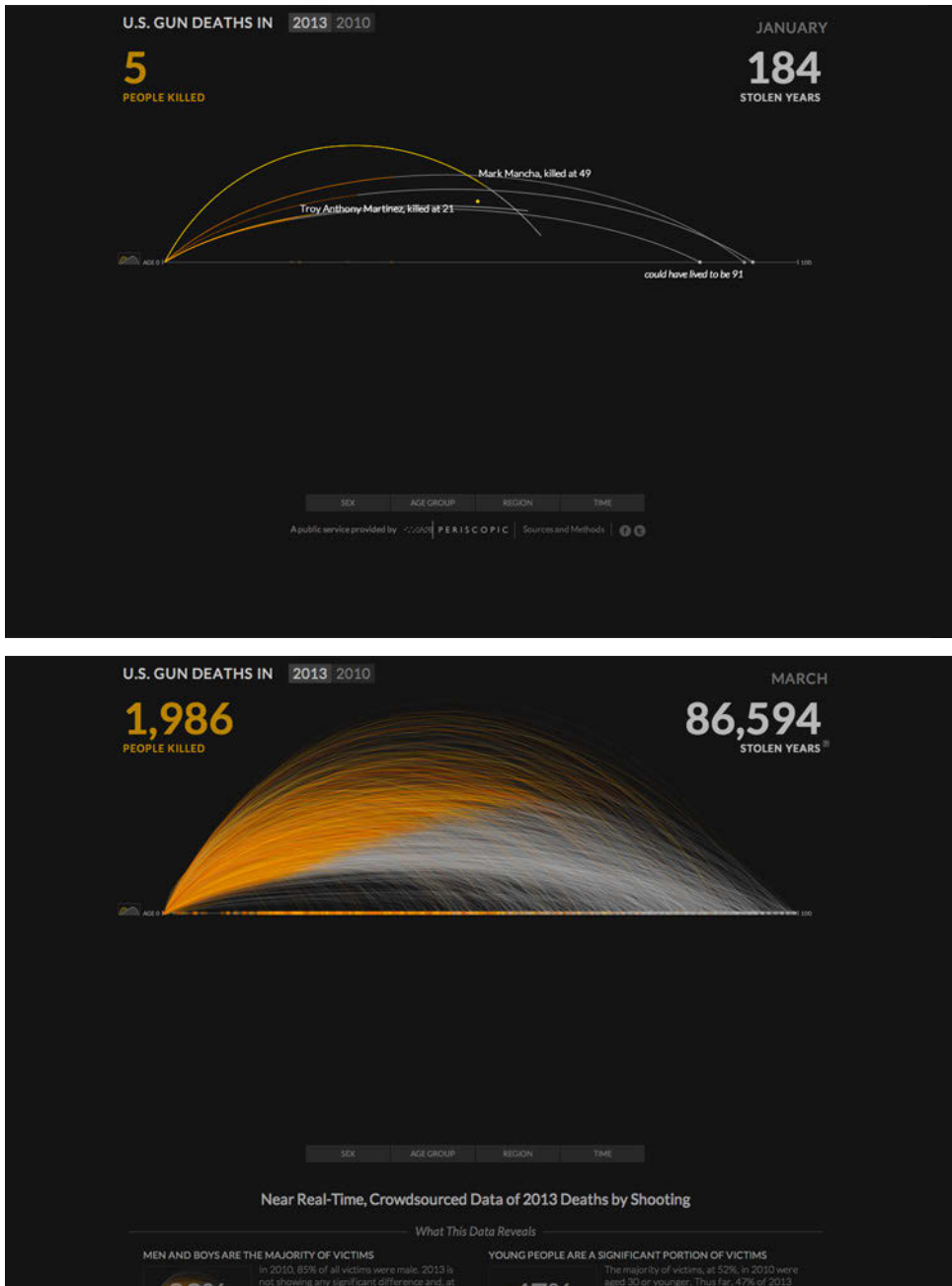


FIG.7. Periscope's interactive visualization of gun murders in the U.S.

The use of empathy, however, has also been criticized by visualizers such as Alberto Cairo. In an epilogue draft for his book *The Truthful Art*, Cairo (2016) argues that relying on emotions can prevent fair and sensible judgement on important, public issues. Furthermore, emotional narratives might mislead readers/viewers when presented without data.

INTERACTIVITY

In contrast to classical static visualizations, interactive visualizations allow the user to dynamically update and change aspects of the data or the view. As a result, interactive visualizations/interfaces have the ability to enhance our mental capacities by offloading certain cognitive processes to a computer program (Ware 2008, p.17). A computer program can store, calculate and show a plethora of data simultaneously and nearly instantaneously, as well as link this information to even more information. This enables humans to find meaningful patterns and make decisions based on their discoveries. Tasks involved in interface interactions have been discussed in detail by authors such as Shneiderman (1996), and Spence (2014a), as well as the design implications for such systems (Ware 2008, Shneiderman, Plaisant 2005). While this knowledge is crucial for the technical implementation of any type of interaction, it does not explain why and how interaction can support the understanding of a (pre-defined) message, such as the magnitude of a number.

One particular feature that emerges when people speak about interactivity, is its power to provide personal views and experiences (emphasis added):

*»Those who acquire to gain insights must explore, interactively, subsets of that corpus to find **their way** towards the view that triggers the ›ah ha!‹ experience.« (Spence 2007, p.136)*

*»Being able to flexibly change perspective on the dataset allows people to make sense of various aspects and test different hypotheses **they have generated**.« (Yi et al. 2008)*

Interactivity can create understanding by allowing to the reader/viewer to create their own, personalized view on the data. Alberto Cairo calls this act of locating oneself in the data, the »me-factor« (Cairo 2016). A personal angle, a personal »ah ha« moment acts as an entry point to comprehend the broader phenomena.

Can You Live on the Minimum Wage?

By JEREMY ASHKENAS FEB. 8, 2014

More than 4.8 million workers now earn the lowest legal pay. This calculator, for a single childless worker, shows the hard choices that have to be made living on the smallest paychecks.

Start by choosing your state:

The minimum wage in New York is **\$8.00** an hour. Tally your living expenses by entering what you think is the least you need for each item. You may find that even your rock-bottom expenses aren't met and that you have to work more hours, pay your bills late, borrow money or do without. [Related Editorial >](#)

Housing
Enter monthly rent or mortgage payments (and don't forget insurance, if this is paid separately) that you would expect to pay for a modest apartment or home in your area. Monthly

Utilities and Services
Electricity, gas, heating oil, telephone, cable, Internet — add them up. Monthly

Transportation
Your car comes with fuel, maintenance and repair costs, as Monthly

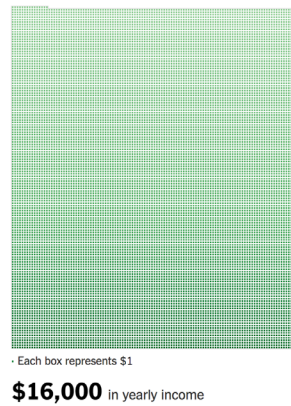


FIG. 8. Can you live on the Minimum Wage? by the New York Times.

An example for this strategy is the New York Times' *Can You Live on the Minimum Wage?*. In this interactive visualization the reader/viewer is challenged to allocate an average minimum wage to every-day expenses such as housing, transportation, health care, food and taxes. When an amount for a specific item is assigned, a unit chart, representing the overall yearly income, is automatically updated to show the remaining dispensable assets. By selecting a different U.S. state the reader can try to make ends meet in their own area, thus getting a perspective on the issue through their own experience. This interactive visualization/calculator is a companion piece to an editorial arguing for the rise of the minimum wage in the U.S.⁸ This rather playful trail-and-error approach to communicating a difficult issue correlates with the idea of *Explorable Explanations*.

⁸ nytimes.com/2014/02/09/opinion/sunday/the-case-for-a-higher-minimum-wage.html

The term Explorable Explanations was coined by Bret Victor (2011) in his groundbreaking article of the same name. Explorable Explanations are essentially interactive systems that allow the reader to learn by exploring and developing an understanding of a phenomena. Other than traditional data-exploration tools, games or simulations, however, Explorable Explanations are always part of an author-driven narrative. As Niki Case puts it: »The author must guide the reader, and provide a structure for the learning experience. Only then can the reader respond, by asking and answering the questions that the author provokes« (Case 2014). Most existing works that have been collected under the Explorable Explanations term, deal with complicated mathematical, programmatic and social systems, such *The Evolution of Trust* a »game about the theory of social trust & cooperation« (explorabl.es).

Connected to the idea of Explorable Explanations are *Reactive Documents* which allow the reader to change variables in/outside a text, thereby automatically updating the text, and *Contextual Information*. Reactive Documents enable the reader to experiment with the author's arguments and see the consequences. Contextual Information work like annotations, insofar as they provide related material directly in/aside the text. However, their content can be potentially controlled by the reader, rather than the author, enabling the former to critically assess the author's claims.

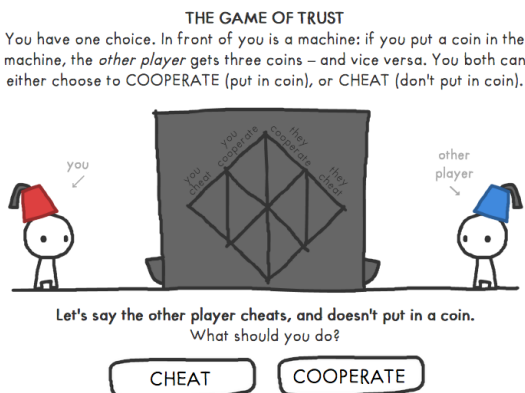


FIG.9. In the *Evolution of Trust* the player chooses whether to cooperate with or cheat their partner to win coins.

2.2 Number Cognition

Number Cognition is an area of cognitive sciences that is concerned with the neural and cognitive mechanisms of numerical skills and mathematic abilities (Cohen Kadosh, Dowker 2016, p.vvii). Research in this discipline focuses on topics such as: perception of numerosity in animals and humans, numerical abilities of human infants, development of mathematical abilities in children, cognitive disabilities, the effects of culture and language on number understanding and the underlying mechanisms of our ability to process numbers (Dehaene 1999). Number cognition involves and touches upon multiple complex processes, from psychology to neuroscience and philosophy. Even the core systems of our numerical abilities are still debated amongst experts (Feigenson, Dehaene & Spelke). As such, I cannot claim to give an exhaustive account of all aspects of these phenomena. I will, however, attempt to present a general overview of those systems that might be relevant to the topic of this thesis, because I believe that is essential to understand the mechanisms and limitations of our number perception to be able to communicate large numbers in an effective way.

Note: There are immense differences in the number abilities of adult and children. As the target group of visualization is generally adult, the following review will concentrate only on adult number cognition. Furthermore, in this review I concentrate on number representation through Arabic numerals.

Number processing

What happens when we try to read/process a number? The most influential number processing model is Dehaene's *Triple Code Model* of number representation (Dehaene 1992; Dehaene, Cohen 1995). According to this model, numbers are mentally represented and manipulated in three different forms. The *visual number form*, the *auditory-verbal representation* and the *numerical magnitude representation*.

VISUAL NUMBER FORM

The visual number form refers to the symbolic representation of a number through Arabic numerals as a string of digits (e.g. 345).

When processing large (symbolic) numbers, this model is especially crucial, as a string has to be correctly identified, before it can be connected to a quantity. In the case of Arabic Numerals that also includes correctly decoding the number notation that describes the structural logic of a string; the place-value principle. Following the place-value principle, the quantity that a digit represents is decided by its position in a string of digits. Arabic numerals employ the base 10, which means that with each position to the left, the value of the digits increase by one power of ten: the rightmost position in a string of digits has a value of 10^0 , moving one position to the left, the digit has a value of 10^1 , one more to the left has a value of 10^2 , and so forth. Hence 345 is constructed from $3 \times 10^2 + 4 \times 10^1 + 5 \times 10^0$.

AUDITORY-VERBAL REPRESENTATION

In the auditory-verbal representation a number is built up from a word sequence (e.g. THREE-HUNDRED-AND-FORTY-FIVE). In contrast to the visual number form, processing of the auditory-verbal form is intimately tied to language specific number naming conventions. Fairly straightforward number words constructions in, for example, Chinese and Japanese (345 is coded as THREE HUNDRED - FOUR TEN - FIVE), stand in a stark contrast to more obscure number word systems found in other languages. In German (and Dutch, Maltese, Arabic etc.), the order of two-digit numbers is inverted, meaning that 345 is spoken as »drei-hundert-fünf-und-vierzig«, literally THREE-HUNDRED-FIVE-AND-FOURTY. French is even more erratic, 70 for instance is spoken as SIXTY-TEN and 98 would be FOUR-TWENTY-TEN-EIGHT. As such, connecting symbolic and verbal representations of numbers is a big challenge in numerical development (Nuerk, Moeller & Willmes 2016).

NUMERICAL MAGNITUDE REPRESENTATION

While both visual and verbal representations can work without the activation of the magnitude representation, for example during mathematical manipulations, neither of them carry any semantic meaning (Dehaene, Cohen 1995). The meaning of a number - the quantity directly associated with it - is represented only through the numerical magnitude representation. It processes the approximate numerical quantity of a number.

The numerical magnitude representation is also assumed to be connected with the *approximate number sense* (ANS) and *sub-*

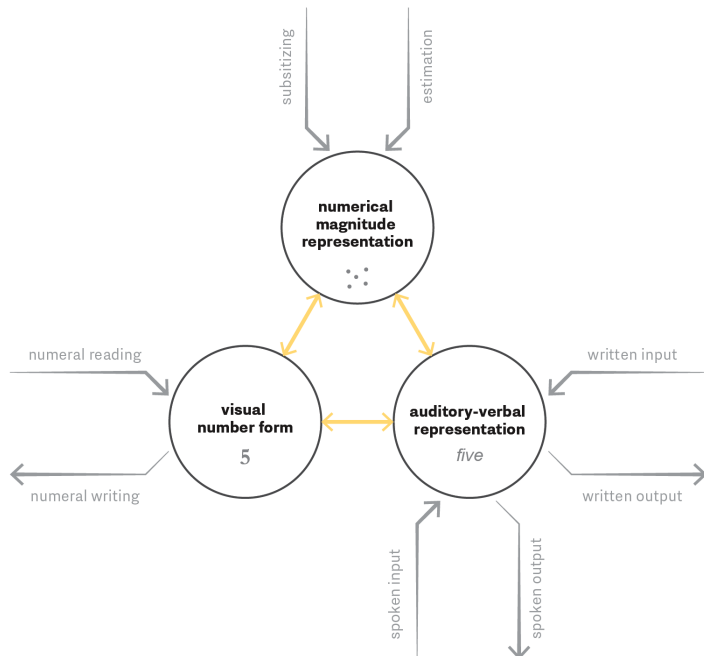
sitizing processes. *Subitizing* is the process of identifying the number of objects when there are less than four objects. Subitizing is a very precise and fast process, this means we can, without counting, (almost) instantly and correctly identify groups of 1 to 4 objects (Dehaene 1999, p.68). Beyond this number we are not able to enumerate objects instantaneously anymore, and we have to rely on counting or on the estimation, for example through the approximate number sense.⁹

⁹ This also explains why it is so hard to distinguish between 1000000000 and 10000000000. Separating long numbers into units of three is a common recommendation in typographical-setting (Forssman, de Jong 2004).

The *approximate number sense* is described as the ability to approximately estimate the magnitude or numerosity of a large set (more than 4) of objects or events (Dehaene 1999). The approximate number sense is inherent to humans and animals and is independent from words or symbols (Sasanguie et al. 2013). When we have to distinguish two sets of objects, our ability to estimate their numerosity follows Weber’s law. It is just as easy (quick) to distinguish between 1 and 2, as it is to tell 4 and 8 apart. The bigger sets get, the bigger their difference has to be, to correctly distinguish them. This is called the »difference effect« (Dehaene 1999, 72).

According to the Triple Code Model, the visual, verbal and magnitude representation of numbers are accessed through different formats.

FIG.10. The Triple Code Model. Note how written numerals cannot be connected to the numerical magnitude estimation directly, but have to be processed in the visual or verbal form first.



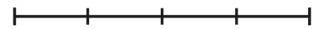
Access to the visual identification happens through reading or writing of (Arabic) numerals. The magnitude of a set of visual or auditory items can be processed through subitizing or estimation processes. Lastly, the verbal presentation is linked to written/spoken in- and outputs. The triple Code model however also assumes that the three modes are connected, and that inputs can be translated from one code into another.

Aside from the Triple Code Model, anecdotal observations and formal experiments have given strong indications that space is strongly connected with the processing of numbers. One such indicator is the *SNARC* (Spatial-Numerical Association of Response Codes) effect. This effect occurs when test-subjects are asked to process numerical information in combination with left or right lateralized response. For example, when test-subjects have to decide whether a number is smaller or larger than 50, they tend to respond faster if the left hand is assigned to the »smaller« response and the right to the »larger« (Dehaene 1999, p.81; Dehaene, Bossini & Giraux 1993). This apparent association of space and numerical information in humans is traditionally explained through the metaphor of a mental *number line*¹⁰. The number line is a mental model of numbers, in which adjoining numbers are positioned on an imaginary line, with smaller numbers on the left and larger on the right (Dehaene 1999, p.76)¹¹.

The number line model has also been used as a tool to assess people's ability to correctly estimate numerical magnitudes. In a typical *number line estimation task*, study participants are presented with a line with labeled endpoint (e.g. 0–100, thousand–billion) and are asked to estimate the location of a particular number (Landy, Silbert & Goldin 2013). Estimation tasks such as these have revealed that depending on age, range of numbers, and familiarity with the numbers, people place numbers in a logarithmic-like or linear¹² fashion (Dehaene 1999, p.76). Logarithmic-like estimation behavior is mostly observed in small children and is therefore considered as a sign of undeveloped magnitude perception. This behavior seems to yield to a more »adult« linear estimation as children improve their numerical skills. Most children are capable of linearly estimating numbers from 0 to 100 by age 7, and numbers from 0 to 1000 range by age 9 (Booth, Siegler 2006).

¹⁰ A growing number of studies have challenged that hypothesis. See van Dijk et al. (2016) for an in-depth discussion.

¹¹ The direction is reversed in cultures where the dominant reading direction is from right to left (Dehaene 1999, p.76).



linear number line



logarithmic-like number line

¹² Linear in this case means that numbers are spaced correctly.

Cognitive limits

Where are the limits of our number processing abilities? Related to the topic of this thesis (comprehension of large numbers), two different behaviors have been observed in adults, overestimation of large written numbers and underestimation of large sizes.

In a study designed to research the mechanisms of the adult number line, Landy et al. (2013) recruited 200 participants for an online number line estimation task with endpoints of one-thousand and one-billion. Around 40% of all participants incorrectly overestimated the position of 1 million, placing it nearly in the middle of the line.

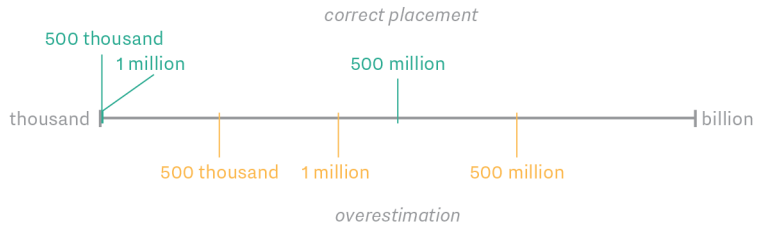


FIG.11. Correct linear placement (green) and incorrect overestimation placement (yellow) on a number line from thousand to billion.

Regarding the magnitude of numbers, rather than their relation to each other, Tretter, Jones & Minogue (2006) found clear evidence for an underestimation behavior for large magnitudes. In their study, 215 participants ranging from elementary school to phd students were asked to list corresponding objects for sizes from one nanometer to one billion meters. Their results showed that, besides vastly overestimating nanoscopic sizes, participants furthermore tended to name objects too small by several powers for large sizes. This decline in accuracy started at around 10 to 100 meters and was most prominent in younger participants. Errors got more pronounced with increasing size even for the older participants, with around 40% of phd students listing objects too small for 1 billion meters.

This type of behavior might be connected to the way different numerical magnitudes are organized in our mind. It has been hypothesized that the numerical magnitude representation is directly connected to locations on the mental number line. At least

for smaller numbers it is assumed that their magnitude representation is directly associated with a specific spot on this imaginary line. While research has not yet found a definite upper limit to the capacity of the number-line model, Landy et al. (2017) argue that it is reasonable to assume that numbers bigger than one million might fall outside of its natural capabilities. If one million, however, constitutes the limit of our natural mental »ruler«, how do we handle numbers beyond that? Landy et al. propose that we construct independent, linear number-lines for different categories of numbers (e.g. million, billion, small, large), which are then combined to form a single number-line.

A similar categorization behavior was observed by Tretter, Jones, Andre et al. (2006) in relation to scale cognition. In an experiment designed to examine underlying concepts of spatial distance and scale boundaries, participants ranging from elementary school to phd students were asked to estimate the correct size of various objects and distances. Results from this task suggested distinct boundaries between different size categories as well as the existence of personal anchor objects for these categories. In a subsequent study investigating the accuracy of spatial scale concepts among students and experts of different ages, Tretter, Jones & Minogue (2006) furthermore noted that, to understand scales that cannot be directly experienced, experts tended to mentally jump into disconnected »scale-worlds« and situate measures within those distinct categories.

While it has been hypothesized that these categories might be the result of association with number scale words (such as thousand, million, billion etc.), studies have rather rejected such a connection. No significant performance difference has been shown, for example, in number-line estimation task, when numbers are presented through scale words or pure numerals (Landy, Charlesworth & Ottmar 2017; Cheek 2017).

2.3 Summary

In this chapter I have demonstrated that visualizations can support and enhance our natural mental capacities and as such contribute to the comprehension of complex issues. Narrative frameworks and emotional connections can furthermore support this by connecting the author or data with the reader/viewer. Interactivity can be used to offer varying levels of detail to the reader/viewer and enable personalized views.

Regarding numerical cognition I have presented several mechanisms that might influence our perception of »large numbers«, such as the effect of language and different number naming conventions on number processing. This chapter also highlighted the importance of providing clear and easily discernible visual presentations of numbers, maybe even using both the visual and verbal number form, as this is the way through which we eventually judge the magnitude of a number. Furthermore, an apparent connection between numerical and spatial cognition was discussed, which shows itself in a lateral preference when processing numbers. Finally, I discussed theories regarding the existence of separated »scale world« in our mind. Here I see the potential of Information Visualization in assisting the connection these disconnected number lines.

3 Framework

Building upon the learning from the previous chapter, in this chapter I review and extend a framework for presenting extreme magnitudes by Chevalier, Vuillemot and Gali (2103).

I present the adapted framework and give examples for existing visualizations using the different strategies I describe.

To be able to evaluate different translation strategies, it is necessary to know what types of strategies exist, how these strategies are constructed and which cognitive mechanisms are involved when they are used. Creating a framework is one way to expose the basic structures and mechanisms of a phenomena, by creating clear definitions and categories for different aspects of that phenomena. Considering the scope of a Master's thesis, I decided to built upon an existing framework by Chevalier, Vuillemot and Gali. The objective of their research matches the aim of my work, but the framework itself is applicable for my purposes only to a certain extent. I will further elaborate upon these limits in the Reflections section.

3.1 Concrete Scales

In their paper *Using Concrete Scales: A Practical Framework for Effective Visual Depiction of Complex Measures*, Chevalier et al. (2013) introduce a new framework for designing and analyzing *concrete scale* compositions. In this context concrete scales means the visual representations of physical objects in real world settings. Using keyword combinations such as »infographics, scale, the size of, equivalent to, how much, as large as, what is the size of, how worthy is« etc., Chevalier et al. collected around 300 graphic compositions from various online outlets. They deconstructed and analyzed the collected images based on the object-to-object relations and the object-to-space relations. From this corpus they developed their taxonomy of *object types* and *measure relations*, and a novel markup language called *CSML* (Concrete Scale Markup Language) to describe objects and their relations in concrete scale compositions.

Depending on their relationship with other objects, objects in concrete scales can have different roles. In the concrete scales framework, there are *base objects*, *anchors*, and *containers*. The relationship between the different objects in a composition is described through the measure relations. According to Chevalier et al. there exist three such measure relations; *Comparison*, *Containment* and *Unitization*. These measure relations can further be combined to form strategies such as *Analogy*, *Zoom*, *Lock* and *Small Multiples*.

Comparison juxtaposes the magnitude of a »shared property« (Chevalier, Vuillemot & Gali 2013) of two or more objects. The length of an unfamiliar object, for example, could be compared to the more familiar length of paper clip. Chevalier et al. point out that comparisons can be exact or approximate, depending on whether the relation between objects can be accurately defined. Comparison was the most common relation measure in Chevalier et al.'s corpus, being used by 83% of all compositions.

Containment describes the relation of two (or more) objects by placing one object within another. A containment relation can also help with comparison, as it groups several objects into a single unit. As with the object type Container, the measure relation *Containment* not only applies to physical containers but also virtual geometric volumes and the nesting of visual elements in a composition.

Unitization involves assessing the magnitude of a base object in terms of a different object, which then becomes the new unit of measurement in the composition. Unitization is helpful as it breaks down an extreme magnitude into smaller, possibly easier to understand units. Chevalier et al. recommend keeping objects organized in containers to remain countable.

Analogy conveys the size difference between two objects (the target) by relating it to the size difference of a more familiar pair (the base). Chevalier et al. categorize analogy as a form of »simultaneous pairwise comparison« (Chevalier, Vuillemot & Gali 2013).



FIG.12. Physical visualization showing how many Earths could fit inside the Sun.

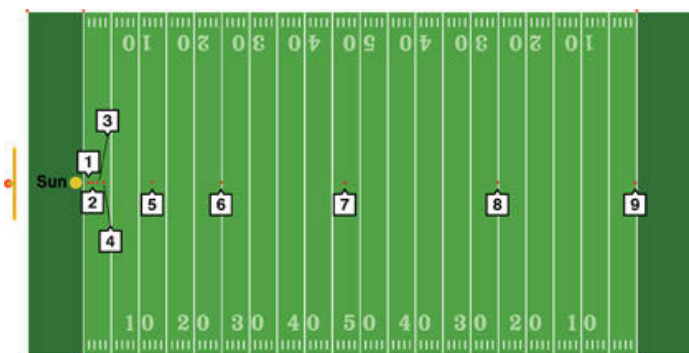


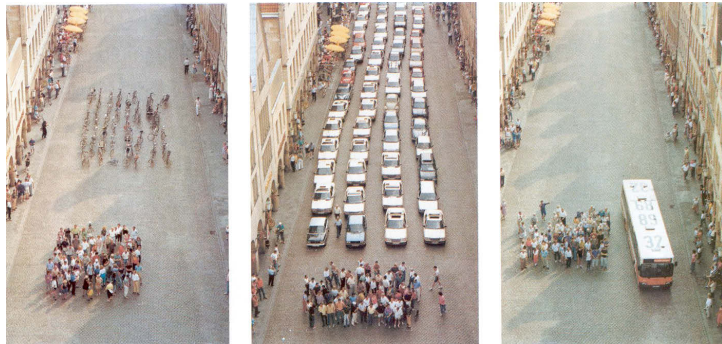
FIG.13. Scale model of the planets' distance to the Sun. The distance between the Sun(1) and Pluto(9) is mapped onto the length of an American Football field.

Zoom can be used when the magnitude is too large to show both overview and detail in the composition. On paper, zoom is shown through side-by-side views, in dynamic media effect can be achieved through animation or interaction. Chevalier et al. classify zoom as a comparison or containment relation.

Small Multiples refers to a composition that features many small elements.

Lock is a comparison constrained to a specific dimension. Locking the dimension that is being compared can be used to create a collection of vastly different objects that share one property.

FIG.14. A 1991 poster commissioned by the City of Münster's planning department showing the space necessary to transport 72 people by bicycle, car and bus. The composition explores the relation of the different objects locked along the dimension of space.



3.2 Reflections on Concrete Scales

Through their work on concrete scale compositions, Chevalier et al. managed to bring together strategies from both theoretical research and practical work into a single practical framework. As such, their framework offers a thorough list of measure relations and strategies that can be employed to improve the comprehension of extreme or unfamiliar magnitudes. In the context of my particular research, however, some aspects of the concrete scales paper and framework have to be reconsidered:

SCOPE

The idea of concrete scales is built upon a concept by Adam Nieman, called *Concrete Visualizations*. Concrete Visualizations depict matter/data in the context of physical space, using »the real world as a canvas for data visualization« (Nieman 2011). Simi-

larly, concrete scale compositions are limited to representing »tangible counterpart[s] with proper spatial properties« (Chevalier, Vuillemot & Gali 2013). This means that the framework is explicitly limited to the visual representations of objects with a spatial property, placed within a real world scenario. Presenting in-visible, in-tangible measures such as time or monetary value effectively contradicts the narrow definitions of concrete scales. And furthermore excludes the use of abstract symbols or (excessive) text in concrete scale compositions. This strict definition severely limits both the types of measures that can be represented, as well as the form of the representation.

Chevalier et al. recognize these limitations in their paper, but cannot offer a satisfying solution to this discrepancy. Rather they give contradictory advice, suggesting to use (arbitrary) objects, even if the measures do not possess a direct physical counterpart. As an example they describe a scenario in which a yearly salary is represented through a human figure.

DEFINITIONS

In their paper, Chevalier et al. define three measure relations and further describe the four most common strategies that can be built by combining these measure relations. Critically analyzing these relations and strategies, however, reveals that many of them overlap significantly and that, furthermore, their definitions are not clear enough to allow a replicable categorization of compositions.

The most evident example for this is *Lock*. While *Lock* is essentially *Comparison* in the Münster space waste example (comparing a shared property, in this case area), this example —————> (also tagged as *Lock* by Chevalier et al.), shows an $A = n \times B$ relation, which according to their framework is *Unitization*. Although Chevalier et al. note that representation can use several strategies simultaneously, the analyzed corpus and the given definition do not give enough evidence that *Lock* is actually an independent strategy.

A similar issue can be observed in regard to *Small Multiples*. The strategy's definition is rather vague, and there is no consistency to the examples that are tagged as *Small Multiples* in Chevalier et al.'s corpus.

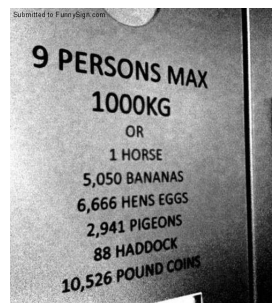


FIG.15. Example from Chevalier et al.'s corpus.



FIG.16. Examples from Chevalier et al.'s corpus tagged as Small Multiples.

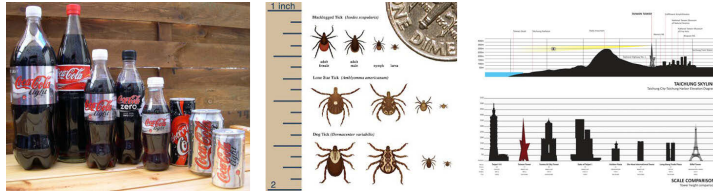
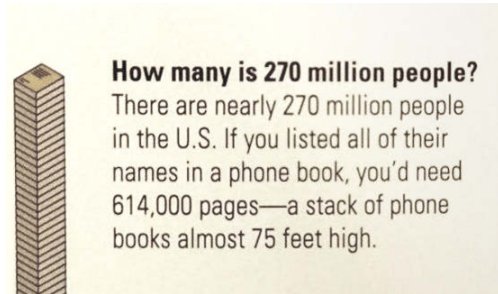


FIG.17. A selection of examples that show several small elements, but are not tagged as Small Multiples.

CONVERSION

While Chevalier et al. discuss other issues in relation to the framework, they do not reflect on the effects of conversion, i.e. the re-expression of one dimension through another. This is especially relevant when dealing with in-material measures, such as money value or time, but can also apply to more tangible measures, such as a quantity of people.

FIG.18. Close-up of a visualization expressing the population of the U.S. through the height of a stack of phone books containing their names.



The question is how these conversions affect the understanding of the target value and construction of mental models. Answering this lies outside the scope of this thesis, but it is an interesting issue that needs further research and an aspect that should be considered when creating visualizations.

EVALUATION

Chevalier et al.'s work focuses on exposing the underlying structures of concrete scale compositions and giving recommendations for their design. As such, their work does not assess the effectiveness of different representations in accurately depicting measures or supporting their comprehension.

Based on these critical reflections I found the need to re-formulate and extend the concrete scale framework. In the following section I describe an adapted framework that takes into consideration additional findings from Information Visualization research and numerical cognition research.

3.3 Adapted Framework

The following framework is a distillation and extension of Chevalier et al.'s concrete scale framework. I will describe a total of six different strategies that can be applied to improve the comprehension of large numbers. Each description is illustrated through several visualization examples, concentrating mainly on large monetary numbers, but also featuring other unfamiliar measures, in order to show a wider range of applications. While presented here as separate approaches, most visualizations are combinations of several strategies. Examples of such combinations are given at the appropriate points.

Terminology (adapted from (Chevalier, Vuillemot & Gali 2013)):

- **measure** is composed of a numerical value and a unit
- **unit** is the type of measurement
- **target measure** is the unfamiliar measure, that is being explained

Strategy	Description
Unitization	$A = n \times B$
Comparison	$A \leq \geq B$
Analogy	$A : B = C : D$
Zoom	$A > B > C$
One-to-One Correspondence	$A = B$
Integration of space/time/sound	

UNITIZATION

Unitization follows the formula $A = n \times B$. The magnitude of the target measure *A* is equal to *n* instances of a new unit *B*. Simply put, Unitization breaks down an extreme magnitude into smaller, more relatable chunks that are easier to grasp. Unitization can be applied on the conceptual level (changing the unit) or merely on the visual level (same unit, but the symbols represent multiple instances). The latter supports visual perception, but not necessarily understanding, since the magnitude stays the same and the »conversion-rate« might be mis-understood by the reader/viewer. If the unit is changed, it is recommendable that the new unit is in some way related to the target measure, either through a logical (kilometers-meters, years-seconds) or contextual connection. For example, dividing a budget cut by the number of people it is affecting (Blastland, Dilnot 2009).

When Unitization is displayed visually, Unit charts are a common choice. As Chevalier et al. noted, however, once the new unit exceeds a certain number, it is advisable to group the units somehow. This can be done through a change in mapping (1 symbol = *n* units), through containment (100 people = 1 bus) or visual grouping (refer to Gestalt principles). Depending on the topic of the visualization it can make sense to use organic or gridded grouping (Boy et al. 2017).

Lastly, it should be noted that the supportive effect of Unitization is eliminated once the number of »relatable« objects becomes an extreme magnitude itself. In this case, *Zoom* can be helpful to provide intermediate steps.

»I tried converting the prices into pizzas, to put it in more familiar terms, and it just became a hard-to-think-about number of pizzas.« (Munroe)

How long a typical worker will have to work to earn what her company's CEO makes in a single hour
assuming no holidays or overtime

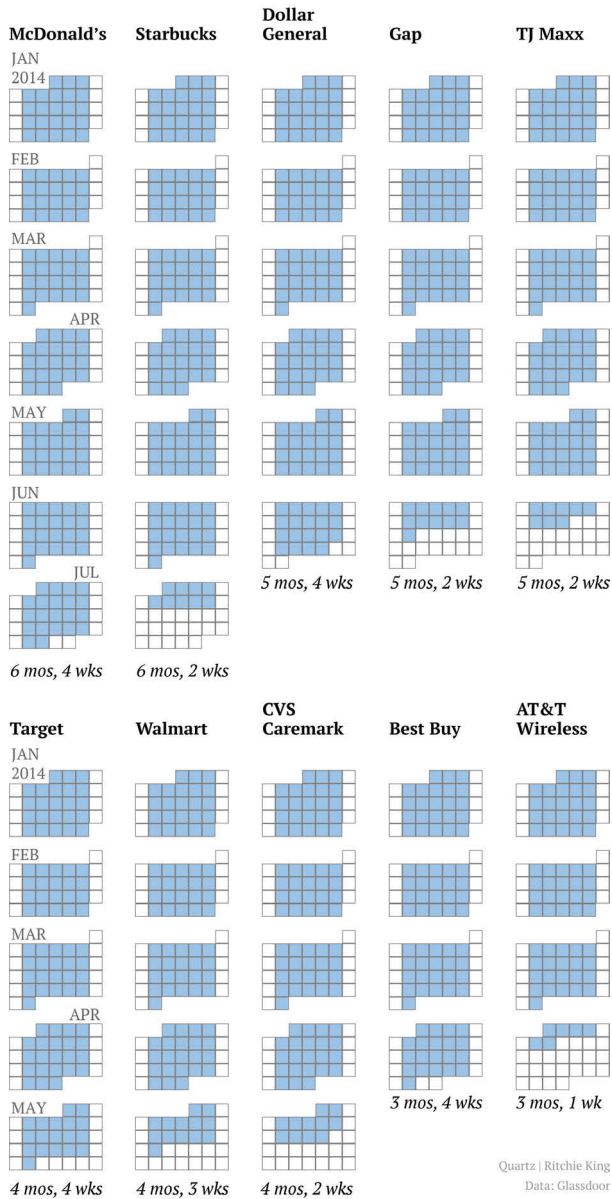
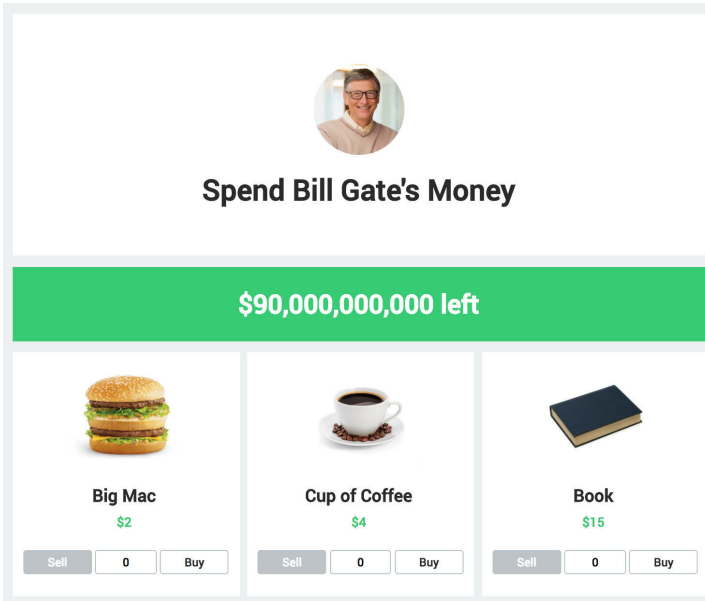


FIG.19. How many month it takes an average worker to earn what the CEO makes in an hour. A visualization by Quartz using unitization and small multiples to highlight the difference in pay between CEO's and average workers of ten major U.S. companies. The CEO's hourly pay is re-expressed in a new unit, namely *n* instances of the daily income of an average worker.



Spend Bill Gate's Money. This website lets the user choose the new unit of measurement, by asking them to spend \$90,000,000,000 on items such as Big Macs and Basketball teams.

FIG.20.



US Debt Visualized in \$100 Bills. This website uses stacks of \$100 Bills to show the scale of the U.S. debt in a rather provocative way. It also uses Zoom to progressively show bigger and bigger amounts, starting from \$100 and Comparison to create anchors for the different values.

FIG.21.

COMPARISON

Comparison is the simplest strategy to visualize the magnitude of a measure. In Comparison, a familiar object acts as an anchor/yardstick through which the magnitude of the unfamiliar target object can be judged and adjusted. It provides a sense of magnitude. For Comparison to work, it is important that the reader/viewer is familiar with the object of comparison. Typical comparison objects in American-centric visualizations are the Statue of Liberty or the Empire State Building, whose heights are probably unfamiliar to people outside the U.S. However, when a good anchor is found for Comparison, the use of labels or scales can become unnecessary, since the object itself acts as a scale. Edward Tufte calls these objects of known size »self-representing scales« (Tufte 1997, p.13). There are three basic types of visual comparison (Gleicher et al. 2011):

- **Juxtaposition:** Showing objects side by side. Objects can also be juxtaposed in time, which is achieved through animation. The positioning of juxtaposed objects has to be carefully designed, to support human perception (Tufte 1991, p.76). Juxtaposition relies on the readers/viewer's memory for comparison.
- **Superposition:** Placing objects on top of each other. To effectively detect similarities and differences in the superimposed objects, these objects have to be sufficiently similar. Superposition depends on the visual system for comparison.
- **Explicit encoding of relationships:** Showing the relative difference between objects. The objects themselves are not shown in the visualization, but rather a new, computed object that describes their relationship. This category utilizes computation for comparison.

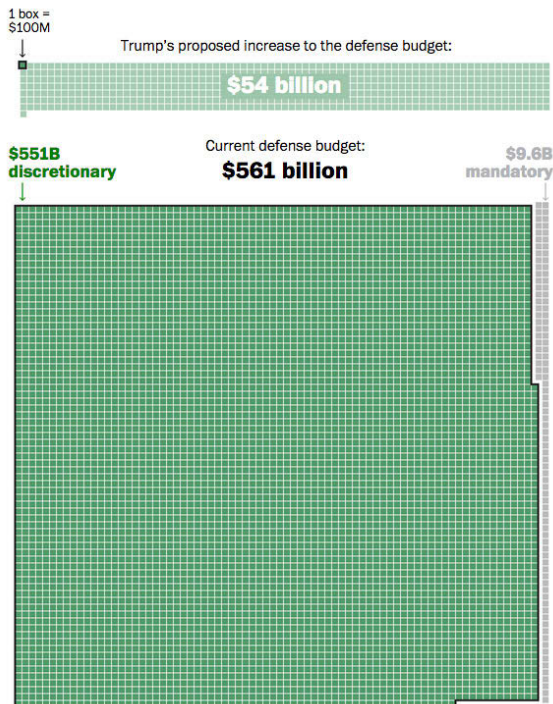
Both interaction and animation can be helpful for Comparison. Ordering, filtering, and changing the type of comparison (e.g. from juxtaposition to superposition) make the comparison between different elements easier/possible. Animation can effectively show temporal comparisons.

Comparison is not only a visual, but also a mental technique. Having a mental catalogue of different sizes (e.g. 1 thousand is the number of seats in a particular stadium; 10 thousand is the number of bricks in my garden wall) helps us to make sense of unfamiliar numbers (Hofstadter 1982; Blastland, Dilnot 2009; Paulos 2001).



All the world's wealth in pizza. This visualization makes use of both Unitization and Comparison. First the world's wealth is translated into pizzas, then the resulting pizza-area is superimposed on Niger for comparison.

FIG. 22.



Trump's defense spending increase isn't extraordinary, but its impact could be. Visualization by the Washington Post juxtaposing President Trump's proposed defense budget increase to the current budget in 2017.

FIG. 23.

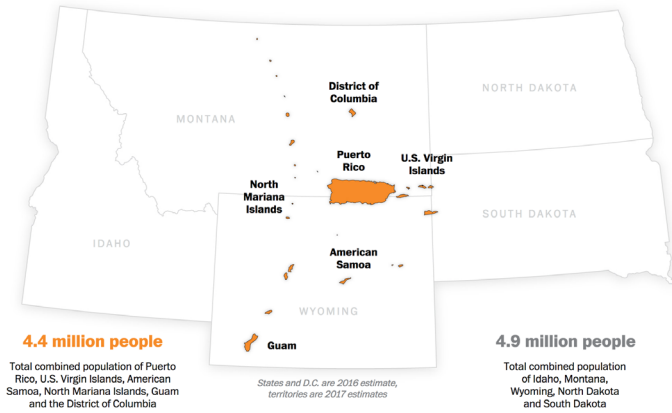


FIG.24. More than 4 million Americans don't have anyone to vote for them in Congress. Visualization by the Washington Post comparing the population and size of the U.S. territories which were ravaged by Hurricane Maria in September 2017 with those of five U.S. mainland states through superposition.

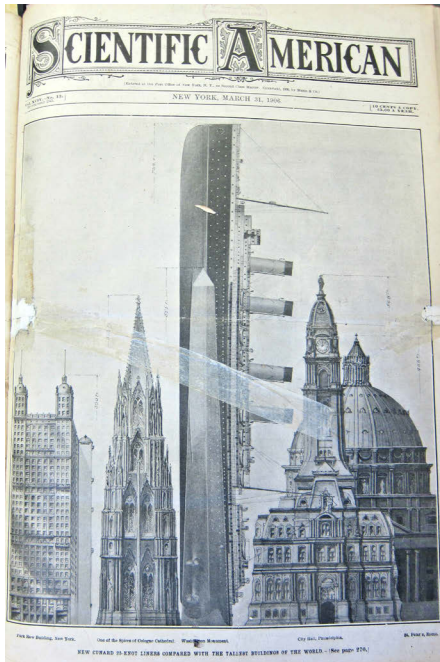


FIG.25. Composition on the title page of Scientific American (March 31, 1906), juxtaposing the length of a new transatlantic ocean liner with the tallest buildings in the world.

ANALOGY

Analogies take advantage of relational reasoning to visualize magnitudes that can not be directly experienced (Resnick et al. 2017b). *Analogy* describes the relation of two (or more) objects by mapping an unfamiliar target concept onto a more familiar base concept. The relationship can be described as $A : B = C : D$, where the relation between the unfamiliar target measures A and B is *equal* to the more familiar relation between the C and D . The base concept visualizes the target concept, thereby supporting understanding and recall (Resnick et al. 2017b). As such, any chart could be described as an analogy. A bar chart, for example, maps the magnitude of a number onto a length that is representable on a paper/screen. Analogical mapping does not have to direct, but can consist of several intermediate steps or concepts (Resnick et al. 2017a). In combination with *Zoom*, the familiar concept can be progressively extended and aligned with the unfamiliar target concept. Such a progressive alignment of concepts can also work to describe the proportional hierarchy/ relation between different intermediate steps.

It is important that the base concept is chosen carefully, and with a clear understanding of the target audience, as this choice can heavily affect the success of the analogy. Several factors have to be taken into consideration when choosing a base concept (Resnick et al. 2017b):

- **Familiarity with the base concept.** The base concept might not be familiar to the reader/viewer. Using the length of an American Football field as a base concept is not helpful to most reader/viewers outside the U.S.
- **Structural differences.** Structural difference between the base and target might lead to confusion. Mapping the unidirectional geological time onto a cyclic clock might lead to wrong conclusion in readers/viewers.
- **Personal connotations.** Personal memories and associations with the base concept can influence the understanding. The size of a specific building, for example, might be unproportionately big in the reader/viewer's head because they visited it as a child.
- **Practical limitations.** The base concept might be constrained by practical issues, e.g. screen size and resolution for digital pieces, class room size for educational demonstrations.

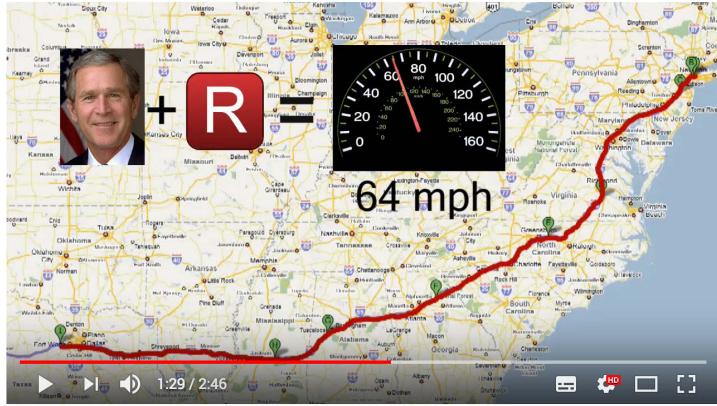


FIG.26. The National Debt Road Trip. This video expresses the U.S. national debt under different presidents as miles per hour and compares the respective distances travelled. The Analogy it uses to calculate this speed is 1 mile = \$5.8 billion and 1 hour = 1 year.



FIG.27. If the Moon were only 1 pixel in size. An interactive visualization of the distances of planets in the solar system, scaled in such a way that the moon is only 1 pixel in size. The user can scroll through the whole length of the solar system.

»One NHS year is about £113bn, so the deficit so far this year [£630 million] could be expressed as a little over two NHS days.«

This extract from *Headline numbers* uses Analogy to exemplify the difference in size between the National Health Service's (NHS) budget and deficit in 2014 by mapping the budget onto the length of a year (Reuben 2014).

ZOOM OR PROGRESSIVE RELATION

If the magnitude of the target measure is too big to make a meaningful comparison, and unitization would result in yet another large number, intermediate steps can be used to progressively work towards a base object. In *Zoom* each step acts as an anchor that gives a reference for the consecutive step. Consequently, Zoom can also illustrate hierarchical or containment relations between these steps. As Tretter, Jones & Minogue (2006) and Landy et al. (2017) have argued, our understanding of magnitudes is fractured, thus it is important to help readers/viewers to correctly combine these separated »scale-worlds«.

Zoom incorporates elements of storytelling, as it consists of a sequence of connected steps. While it is possible to show zooming in a static layout (e.g. through a side-by-side view), interaction and animation can greatly benefit this strategy. The highly successful 1977 educational video *Powers of Ten* by Charles and Ray Eames uses animation to explain the exponential powers of ten. Their video progressively zooms in and out from an aerial view of a man sleeping in a park (10^0 meter) at a rate of 10^{10} meters per second, going as small as 10^{-16} meters and as large as 10^{24} meters (eamesoffice.com). While not using the short scale word »billion« for a »million million«, the video also uses different units of length such as lightyears and ångströms, thereby explaining and highlighting their relationship to meters.

Zoom can be combined with all other strategies, to give the reader/viewer a better understanding of the relationship between different magnitudes and mentally connect different scale-worlds.

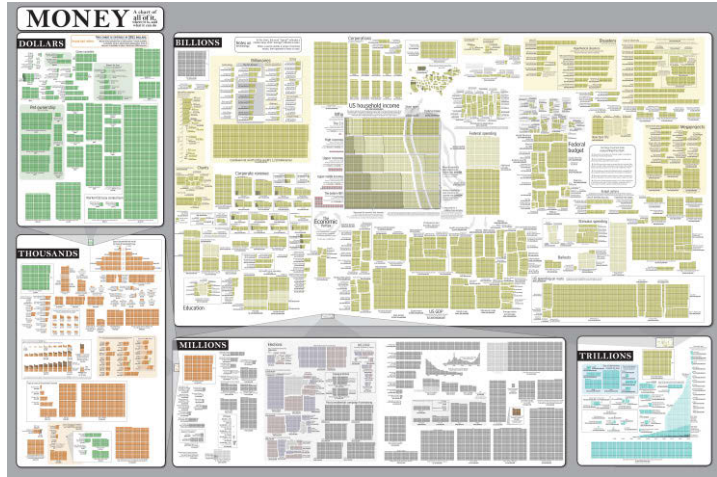


FIG.28. MONEY. A chart of almost all of it, where it is, and what it can do. A massive visualization by Randall Monroe showing the value of different amounts of money as well as the proportional relation between the steps of the short scale though a static zoom layout.

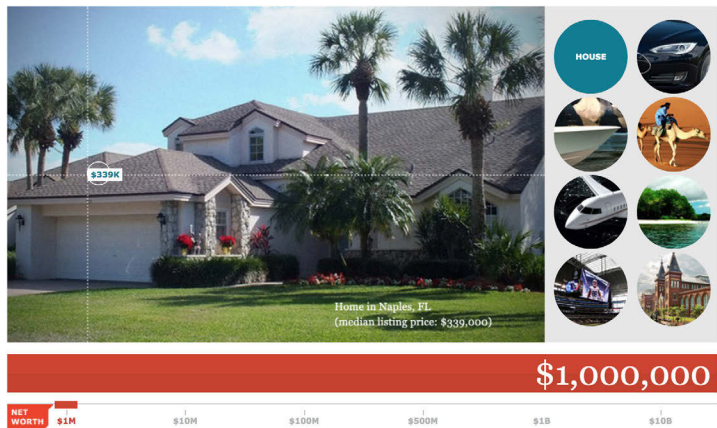
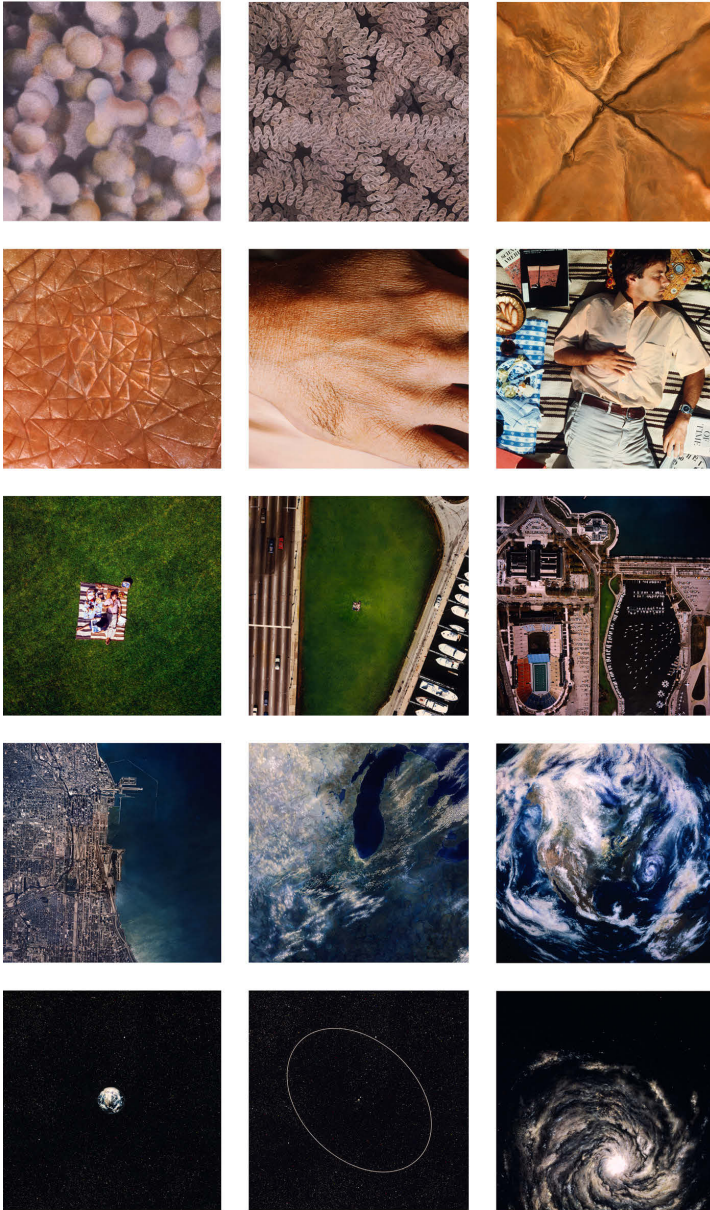


FIG.29. What a billion dollar buys you. An interactive piece by Forbes magazine. This visualization shows which 8 objects together (house, car, boat, vacation, plane, property, sports team and charity) could be purchased for a certain amount of money, ranging from \$1 million to \$10 billion. The reader can compare similar items for different price ranges, giving them a feel for the buying power. A stacked bar chart at the bottom furthermore highlights the difference in size between the steps. The slider itself, unfortunately, is spaced uniformly, therefore somewhat counteracting the chart's effect.

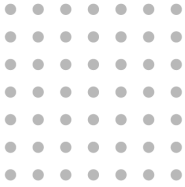


Powers of Ten. A highly successful 1977 educational video by Charles and Ray Eames, which aims to explain exponential powers by progressively zooming in and out from an aerial view 1m^2 in size of a man sleeping in a park.

FIG.30.

ONE-TO-ONE CORRESPONDENCE

Often used in artistic or physical projects, *One-to-One Correspondence* is a way of showing the absolute, sheer size of a magnitude. When using One-to-One Correspondence, one unit of a measure is represented by exactly one object or sign. This direct correlation between symbol and object harkens back to early numeration systems such as clay tokens or notches (Dehaene 1999, p.95).



One-to-One Correspondence is most useful for discrete units, such as a group of objects or beings. Particularly when the target measure describes humans, One-to-One Correspondence is an effective way of sparking interest and evoking empathy (Slovic, Slovic 2015). The power of One-to-One Correspondence can also come from »bodily« aspects, such as collection or producing physical tokens or scrolling/clicking through a visualization.

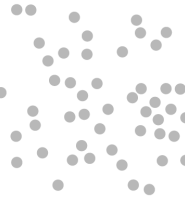


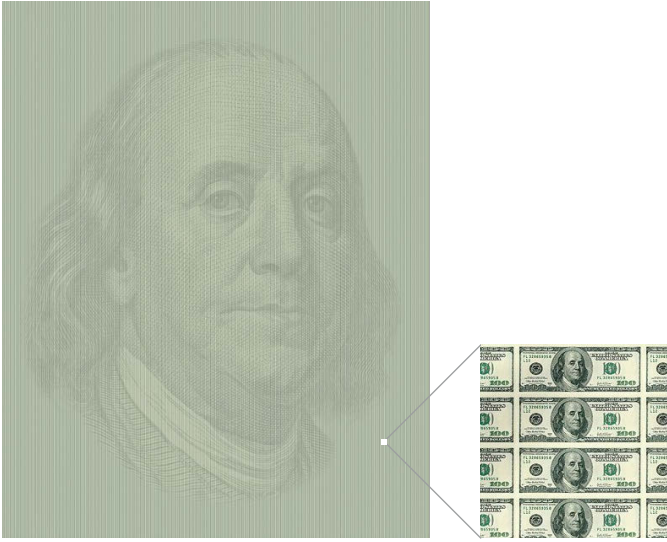
FIG.31. The regularly spaced dots seem more numerous than randomly placed ones. Each group consists of 49 dots.

One-to-One Correspondence might share some visual similarities with Unitization when icons are used to represent the target measure. However, while Unitization changes the unit of the target measure and thereby the absolute number of objects displayed, One-to-One Correspondence only changes its visual appearance.



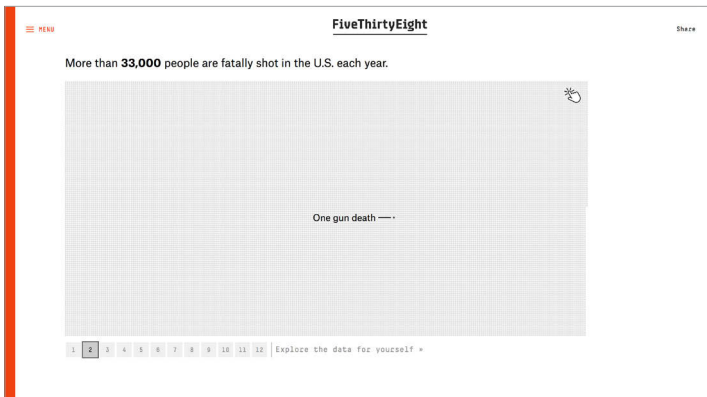
FIG.32. The Solitaire illusion. The cluster of grey dots seems more numerous than the number of green dots, although there are 16 of both.

However, as the magnitude of the target measure is not really simplified, One-to-One Correspondence might not actually ease the cognitive problems that come with trying to understand big numbers. As we have seen, the ability of the human visual system to estimate the numerosity of objects drops dramatically once the set exceeds 3 objects (Dehaene 1999, p.97). Furthermore, our perception of numerosity is extremely subjective to begin with; The type and size of tokens, their placement and density can significantly influence how we perceive them (Dehaene 1999, p.72). One example for this subjectivity is the *Solitaire illusion*. The Solitaire illusion is an illusion of numerosity, in which the subjective number of objects is effected by their spatial arrangement (Frith, Frith 1972). Even when items are controlled for density and area, the ability to discriminate the magnitude of two set of objects is restricted by Weber’s law.



Ben Franklin. A 26×32 meter big, composite image of 125,000 photos of one-hundred dollar bills. This equals the \$12.5 million that the U.S. government spent every hour on the war in Iraq. This work is part of Chris Jordan's *Running the Numbers: An American Self-Portrait series*, in which he uses one-to-one correspondence to create art that »looks at contemporary American culture through the austere lens of statistics«.

FIG.33.



Gun Deaths in America. This interactive visualization by FiveThirtyEight plots an average of 33,000 annual gun deaths in America as individual dots. Interaction reveals further information about the different types of incidents.

FIG.34.

INTEGRATION OF TIME, SPACE, AND SOUND

In his book *Information visualization : design for interaction*, Robert Spence (2007, p.5) writes that, »[Visualization] need not involve a visual experience, as might be suggested by the term ›visual‹ [...], because sound and other sensory modalities - not only graphics - can be employed to represent data«. Indeed, incorporating directly experienceable modalities such as time, space, and sound can help with creating connections to units without direct physical properties or simply reinforce the information through another medium.

Time can be a means to break down a measure that is directly connected with an action e.g. spending or producing per second. Also it is often used to describe the duration of a hypothetical and sequential action, e.g. greeting all the people in NYC, dropping water into a bucket, etc. In some instances, time becomes the new unit of measurement, for example in this analogy, which tries to clarify the differences in magnitude between the steps of the *short scale*¹³ (Blastland, Dilnot 2009, p.17; Paulos 2001, p.12):

1 second = 1 second
 1 million seconds ≈ 11,5 days
 1 billion seconds ≈ 32 years
 1 trillion seconds ≈ 32.000 years

Space is the most directly experienceable, most familiar, and therefore maybe the most powerful modality of those presented here. Experience with different volumes, areas, and distances is essential for the understanding of scales (Tretter et al. 2006). Furthermore, phenomena such as the SNARK effect hint at an inherent connection between numbers and space in the human mind. Using space to express non-spatial units is a common strategy in *Analogy*, for example mapping a geological time frame unto a route on Google Maps (Parker 2011). Space can also be used to compare or organize units.

Concrete aspects of **sound**, such as pitch, duration, loudness, and timbre, can be hard to quantify for a layman, but sound in general can be useful as a supportive tool. It can evoke emotion (e.g. *A sonic memorial to the victims at Orlando's Pulse nightclub*¹⁴), act as a mnemonic device (e.g. *Lightyear.fm*¹⁵) or help to drive a point (e.g. *Visualization of nuclear detonations from 1945 to present*¹⁶).

¹³ *The short scale is a system for naming numbers based on the powers of one thousand. This system is used in most English-speaking countries.*

¹⁴ *revealnews.org/blog/a-sonic-memorial-to-the-victims-at-orlandos-pulse-nightclub*

¹⁵ *lightyear.fm*

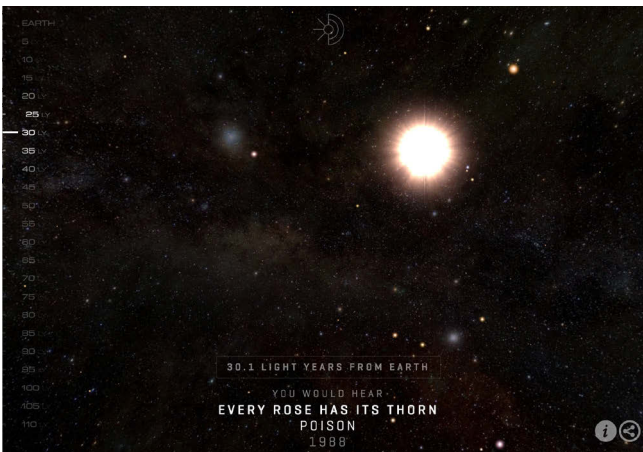
¹⁶ *vimeo.com/135580602*

The use of other modalities, besides the presented, is also possible. Comprehension could also be supported through taste (e.g. data-cuisine.net), smell, texture and temperature. However, most of these require physical presence, which is impossible in traditional paper or display-bound Visualizations.



The National Debt Clock in New York City is a real-time counter of the U.S. national debt. It was erected by Seymour Durst in 1989, when the U.S. national debt stood at (only) \$2.7 trillion. The constant increase of debt can be observed even by casual passersby and brings a sense of urgency to the issue of debt. Similar debt clocks can now be found online for every country.

FIG.35.



Lightyear FM. This visualization uses popular music to explain the speed of light and the vast distances of space. »Since radio broadcast leave Earth at the speed of light, the farther away one gets, the longer the waves take to travel there-and the older the music one hears.«

FIG.36.

3.4 Summary

In this chapter I presented six strategies that can be used to present extreme magnitudes in ways that facilitate their understanding. Those strategies included: *Unitization*, *Comparison*, *Analogy*, *Zoom*, *One-to-One Correspondence* and *Integration of space/time/sound*. Working on this framework highlighted the potential pitfalls of making a visualization for big number comprehension. Many internal and external aspects have to be considered, such as the type of measure and the cultural background of the reader/viewer, and great care has to be taken to create representations that do not misrepresent the data or create more confusion than insight. Furthermore, the importance of having a familiar reference point became evident, as well as the beneficial effects of using multi-modal representations. While I discussed the strategies as separate approaches, a combination of several strategies seems to be necessary to achieve the best results.

One strategy that was not mentioned, is the use of scientific notation ($3 \times 10^4 = 30,000$) to describe big (and small) numbers. Authors such as John Allen Paulos (Paulos 2001, p.11) and Douglas Hofstadter (Hofstadter 1982, p.130) recommend the use of scientific notation as they are more legible and make it easier to spot large jumps in magnitude, which might otherwise be missed.

With the exception of projects related to the death or suffering of human-beings, not a lot of examples could be found from »established« visualization outlets that aimed at making big numbers meaningful. This might be read as an indicator that this matter is unfortunately disregarded by outlets that aim at an »educated« audience. Although by no means complete, I believe that, just as Chevalier et al.'s work was for me, this adapted framework can be a useful starting point for designers that seek to produce visualizations that make the presented numbers meaningful and manageable to their readers/viewers.

4 Prototypes

This chapter describes the design process and functionalities of three visualization prototypes, named Stacking Bills, Money Clock, and Buying Power, that I created based on the revised framework and a corpus of existing visualizations.

Before creating the prototypes, I collected a corpus of representations explaining the magnitude of monetary measures through various strategies. I gathered examples by searching online platforms (Google, Google Images, Twitter, and Youtube), using different topic-related terms and phrases (e.g. [to put this] into perspective, to visualize, million/billion/trillion dollar, big numbers, Bill Gate's worth, etc.) and browsing information visualization books, blogs and major online news outlets such as The New York Times, Washington Post and Bloomberg.

This search resulted in a substantial corpus of examples, ranging from written to visual and physical. The full list of examples can be found at bit.ly/2Hbyvio. It has to be noted that the number of individual strategies is higher than the number of examples, since videos especially can make use of several representations throughout.

I further analyzed this corpus, to determine which strategies were the most commonly used. I then followed the three most popular strategies in the design of the prototypes. I decided to use this approach as I was interested to see if and how effective these established strategies would be. Although only a few of the collected examples were dynamic, I decided to create interactive prototypes. Based on my research and personal experience and finding a personal angle seemed essential for the understanding of large numbers, both of which is hard to do in a static graphic. Finally, I want to mention that the visual design of the prototypes was intentionally kept basic and bare-boned, so that their design would not influence the outcomes of the evaluations.

4.1 Stacking Bills

Stacking Bills is based on a strategy common in »explainer videos«. Following this strategy, monetary value is converted into a corresponding amount of bills, usually \$100 or \$1 bills. This quantity of bills is then re-expressed as a spatial unit, by either stacking, bundling or laying the bills flat alongside each other. Height, volume or area therefore becomes the new unit of the target measure. This conversion connects the intangible, and therefore hard to understand, monetary value with a more directly perceivable experience - that of space. As Tretter, Jones & Minogue (2006) and Jones & Taylor (2009) have noted in their studies on scale cognition, a sense of scale is often developed through physical experiences of moving through space, such as walking, driving a car, or flying in an airplane.

Stacking Bills presents the target measure in the form of stacked €100 bills, and states the height of the resulting stack. This height is transformed into distance and placed on a map for comparison. I chose to use distance rather than height or volume (which were much more common in the corpus), because 1) I assume that long distances are more directly experienceable and therefore easier to relate to than heights and 2) according to Cleveland and McGill (1984) and Mackinlay (1986), lengths are perceptually easier to judge than volumes in a 2-dimensional medium.

The initial view of the prototype primes the question »How much is 346 billion Euro?« and introduces the conversion metaphor through a short introduction text and visual.

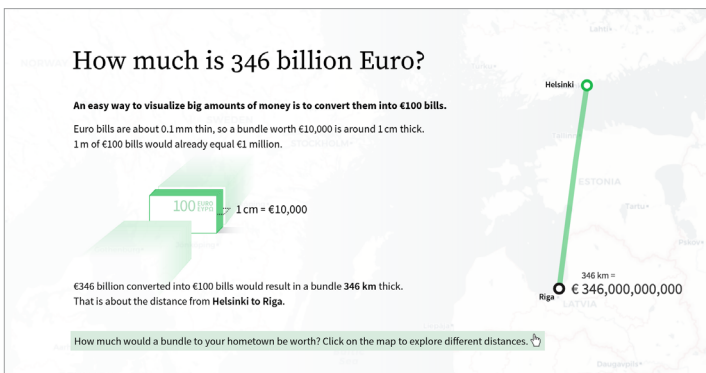


FIG. 37. The initial view of the *Stacking Bills* prototype.

In the same view the user sees the target measure re-expressed as a »stack« (i.e. a line) stretching from Helsinki to Riga on a world map. Helsinki was chosen as an anchor considering that the test users are located in, and therefore familiar with the Helsinki area. Following the initial view, the user is then able to drag the endpoints of the stack to change the length of the resulting stack, thereby adjusting the represented monetary measure. The user can also change the direction of the stack, enabling them to compare different city pairs. The length and value of the stack are displayed next its endpoint and change dynamically as the user interacts with the prototype.

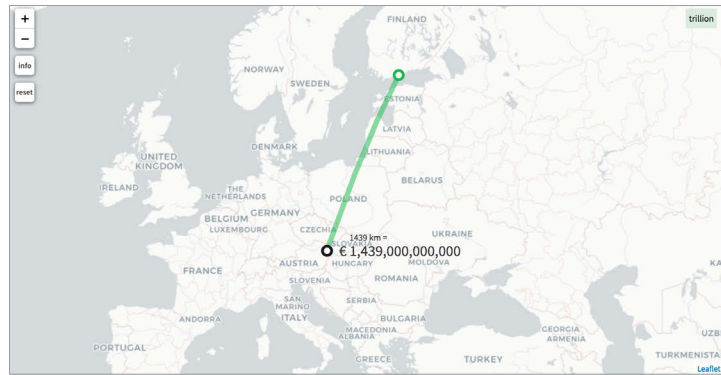


FIG.38. A stack reaching from Helsinki to Vienna, worth about €1.5 trillion.

¹⁷ leafletjs.com

¹⁸ github.com/henrythasler/Leaflet.Geodesic

The prototype was build using *Leaflet*¹⁷, an »open-source JavaScript library for mobile-friendly interactive maps« (Agafonkin 2017) and *Leaflet.Geodesic*¹⁸ by Henry Thasler, an add-on for Leaflet that allows to draw geodesic lines and great circles. A geodesic line shows the shortest distance between two points on the earth’s surface, taking into account the distortion that appears when the (nearly) spherical 3D surface of the earth is projected onto a 2D plane (developers.arcgis.com).

The full prototype can be found at www.lisastaudinger.com/stacking_bills.

4.2 Money Clock

Money Clock uses time as a familiar unit to re-express the target measure. Examples that utilize time as a unit often express the target measure in terms of a rate (e.g. \$1 per second) and a duration (e.g. for 1 year). The target measure itself can be achieved by spending, receiving, paying off, shredding etc., depending on the context. For example:

*»If the debt [\$5.7 trillion] was paid down at a rate of \$1 dollar per second it would take 130,000 years.«
(Holmes, Bagby 1999b)*

Previous studies on nanoscopic sizes have found that temporal representations in combination with visual representations are highly effective in refining the mental model of imperceivable magnitudes (Song, Quintana 2012).

As with *Stacking Bills*, *Money Clock* starts off with the question »How much is 346 billion Euro?«. It then re-expresses the target measure as a (approximate) rate and duration:

If €346,000,000,000 was paid off at a rate of €100 per second, it would take about 109 years to pay it all.

The visualization then starts »paying off« the target measure, symbolized through a counter going down at a rate of €100 per second. Below the counter a progress bar shows the progression and displays the elapsed and remaining duration of the pay off.



FIG.39. Initial view of *Money Clock*.

The user can simply watch the numerical value going down at the preset rate, or they can interactively adjust the variables of the visualization. This allows them to explore how the parameters effect the outcome and the relationship between them. The variables are linked in the following way: Adjusting the initial money value updates the duration time and counter. Setting a new duration time (here the user can change both number and unit) recalculates the money value that could be paid off in that time frame and updates the counter.



FIG.40. Changing one of the input variables impacts the rest of the visualization.

A »skip forward« button furthermore allows the user to skip ahead one quarter of the total duration. This was implemented since the duration of bigger money values are too long to see any visual progress in the bar.

The prototype was build in Javascript and JQuery and can be found at www.lisastaudinger.com/money_clock.

4.3 Buying Power

In contrast to the previous prototypes, *Buying Power* expresses the target measure in terms of its intrinsic value i.e. buying power, rather than its numerosity. To achieve this, *Buying Power* applies Unitization; a relatable object acts as a new unit of measurement through which the magnitude of the target measure is evaluated. In the case of monetary measures, these objects are most commonly consumer goods or services. In this prototype Unitization is supported through the One-to-One Correspondence strategy, visually showing the magnitude differences between the target measure and the new unit.

The initial view of *Buying Power* introduces the idea of unitization and presents the target measure, 346 billion Euro to be equal to 915,344 single family houses in Helsinki (worth about €378,000). This conversion is visually supported through a pictorial unit chart, juxtaposing one instance of U.S. national debt with a responding amount of pictograms symbolizing the instances of the new unit. The chart is fifty units wide, meaning the user has to scroll a considerable amount to see the full length of the chart.

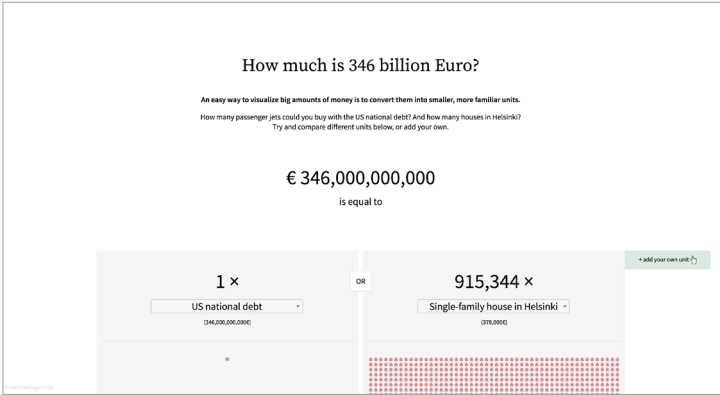


FIG.41. Initial view of *Buying Power*.

The user can interact with the visualization to change the unit of measurement, either through choosing from a list of provided »objects«, or adding their own object through an input interface. If a new object is inputed, the chart is updated with that value and it is added to the list of available options.

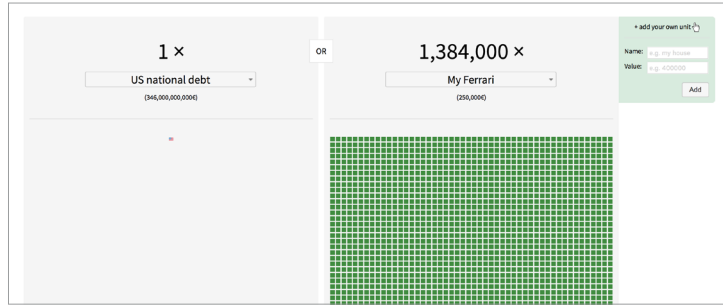


FIG.42. Updated view of the prototype with a custom unit of measurement, plus the expanded input field.

¹⁹ github.com/twemoji

The prototype was built in Javascript and JQuery, some pictograms were adapted from Twemoji¹⁹. The prototype can be found at www.lisastaudinger.com/buying_power.

5 Evaluation

In order to answer my research question *How effective are various strategies in supporting the comprehension of large (monetary) numbers?* I first discuss the current state of evaluation in the field of Information Design, and argue for, and describe an insight-based evaluation approach. The rest of this chapter details the setup of the user test sessions and discusses the results in depth.

5.1 Evaluating insight in Information Visualization

Dealing with Information Visualization, one eventually arrives at the question: What is a good Information Visualization? According to Edward Tufte (2001), maybe the most prominent (and controversial) authority in the field, »graphical excellence« can be judged by measures such as the *data-ink ratio*²⁰ and the *lie-factor*²¹. However, while this approach might be feasible for some instances, it cannot be applied to every type of visual representation of information, as different visualizations are meant to serve different purposes. While some visualizations are intended for exploration and discovery, others explain and contextualize information, and some others are produced to prove a single argument. The quality or effectiveness of a visualization depends on its goal, and can (or should) not be measured with one simple, universal yardstick. The following section will therefore dive deeper into different methods of evaluation used in Information Visualization.

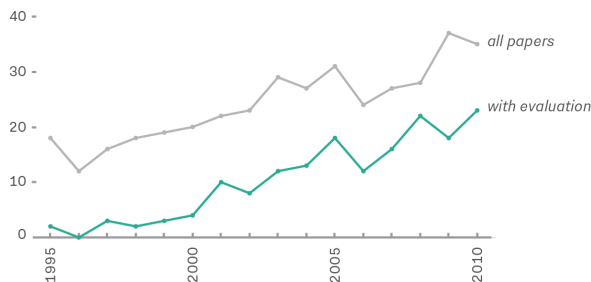
²⁰ *Data-ink ratio is the proportion of a graphic that carries actual information. (Tufte 2001, p.93)*

²¹ *The lie factor measures the relationship between a data point and its depiction in a graphic. (Tufte 2001, p.57)*

Evaluation in Information Visualization

During the last decades, the evaluation of (information) visualization has become a hot topic in visualization research (van Wijk 2013). Not only are new visualization tools and methods being studied and evaluated, but researchers also strive to validate »established« truths and assumptions in the field (see for example Skau and Kosara's (2016) paper on how people read data from pie charts). This trend towards empirical research and validation can also be seen by the steady increase in papers submitted to visualization conferences that feature some kind of evaluation (Lam et al. 2012).

FIG.43. Number of papers submitted to InfoVis from 1995-2010, and the share with at least some kind of evaluation.



While initially visualization researchers adapted their research methods mostly from related fields such as human computer interaction (HCI), computer graphics (CG) (Lam et al. 2012; Carpendale 2008), researchers now have access to a wide range of evaluation approaches, as well as a sizable amount of recommendations for their application (van Wijk 2013; Lam et al. 2012). Some common evaluation approaches to test the effectiveness of a visualization include (Lam et al. 2012; Saraiya, North & Duca 2005; Plaisant 2004):

- **Controlled Experiments & Laboratory Observation** (often testing response accuracy and response time for specific design elements or tools)
- **Formative Usability Testing** (e.g. think-aloud, interviews)
- **Metrics, Heuristics, and Models** (evaluation based on measures such as data-ink or cognitive models)
- **Case and Field Studies** (evaluating tools in realistic, natural environments)

Despite the wide range of evaluation methods available to researchers, there is a dramatic unbalance in the use of these methods. In a review of more than 300 visualization papers, Lam et al. (2012) found that 85% of research assessed either user performance, user experience or a visualization algorithm. This means that an overwhelming majority of evaluation is centered on assessing visualizations rather than underlying processes or user understanding.

As stated in the introduction of this thesis, the focus of this work lies on investigating the effectiveness of different visualization strategies in improving the understanding of big numbers. But what is *understanding*? According to Bloom's taxonomy of educational learning objectives, understanding is a cognitive processes that involves constructing meaning from messages and integrating new knowledge into existing schemas (Bloom, Anderson & Krathwohl 2001, p.70). As such, understanding is a type of learning. Assessing understanding quantitatively, however, is difficult since the process of understanding and its outcome are mostly very personal and open-ended. In Information Visualization research, this type of personal learning has been studied under the term of *insight*.

Evaluating insight

Insight has been declared the principle goal of Information Visualization by many authors. A well known quote by Card, Mackinlay and Shneiderman states that, »the purpose of visualization is insight, not pictures« (Card, Mackinlay & Shneiderman 1999, p.6). In his paper *Toward Measuring Visualization Insight* North (2006) argued that insight is composed of five important features: it is complex, deep, qualitative, unexpected, and relevant. Chang et al. (2009) identified two different types of insight: spontaneous insight and knowledge-building insight. Spontaneous insight describes a moment of enlightenment (an »aha« moment) and is a form of problem-solving. This type of insight is commonly researched in the cognitive sciences. Knowledge-building insight, on the other hand, means gaining deep, complex knowledge that can create novel mental connections and support hypothesis generation. These two types of insight are related, as revelations created by spontaneous insight can serve as the building ground for knowledge building insights in the future. The resulting knowledge-building insight in turn can lead to new spontaneous insights.

Most relevant to this thesis is the work of Saraiya, North & Duca (2005), who have defined insight as »an individual observation about the data by the participant, a unit of discovery.« In their study on bioinformatics visualizations Saraiya et al. merged aspects of usability testing and the controlled experiment method to create a novel method and set of measures for evaluating insight. Their procedure follows the think-aloud protocol but further analyzes the information gathered through this based on a list of quantifiable characteristics. By coding occurrences of insight by category, qualitative information can be translated into quantitative data. This allows to detect patterns of insight in the participants' interaction with the prototypes (North 2006). The think-aloud protocol can also hint at specific features of the visualizations that support insights for the users or cause problems for them.

5.2 User testing

The user tests were conducted as a between-subjects study with one independent variable, the visualization strategies/prototypes, which had three treatments:

- Stacking Bills
- Money Clock
- Buying Power

The dependent variables for the user tests consisted of:

- pre-test results
- total time spent with the tool
- amount, type and category of insights
- benchmark test results
- usability issues

MEASURES

Insights recorded during the sessions were encoded according to the following quantifiable characteristics, adapted from Saraiya, North & Duca (2005):

- **Observation:** The actual remark about the data.
- **Time:** The amount of time till an insight is reached.
- **Complexity:** The complexity of the insight. A general observation about size (*that's a huge pile of bills*) earns 1 point, exemplifying a value earns 2 points, interpreting, comparing or correlating a magnitude earns 3 points.
- **Directed versus Unexpected:** Directed insights are the result of intentional actions/questions by the user. Unexpected insights are other findings that the user did not specifically search for.
- **Correctness:** Incorrect observations can occur when the visualization is misinterpreted.
- **Category:** Based on the observations that were collected during the test sessions, the insights were divided into six different categories: personal, summarizing, inferring, revelation, interpreting, reading, and hypotheses. These categories are described in more detail in the Results section.

DATASET

The dataset used for the prototypes is the U.S. budget deficit in 2015, which at that time stood at \$425 billion. The number was taken from a New York Times online article, titled *U.S. Budget Deficit Rose in July, but 8-Year Low Is Expected for Year*, published on August 12, 2015. As most of the participants are expected to be more familiar with the value of Euro than U.S. dollars, the measure was converted into 2018 Euro. As such, \$425 billion equals around €346 billion.

PROTOTYPES

Three prototypes were tested: *Stacking Bills*, *Money Clock*, and *Buying Power*. Refer to the Prototypes chapter for detailed descriptions.

PARTICIPANTS

Nine participants (three per prototype) were recruited through an email communication and a Facebook post aimed at current and former students and staff from the Department of Media. According to Nielsen (2000), three to five users are enough for a qualitative user testing, as the likelihood of getting new learnings levels out after about five test sessions. The final participant group consisted of five Master students and four Master degree holders from various design related fields. The average age of the participants was 29.8 years. As origin is an important factor for number understanding, participants were asked to report their country of origin and their native language. Two participants came from countries that used the short scale number naming system, four came from countries that used the long scale, and three came from countries that used neither (India and Taiwan). One participant had previously interacted with a prototype during a personal communication, so they were assigned to a tool they hadn't seen before. Otherwise participants were randomly assigned to a prototype.

PILOT TEST

One pilot test was conducted to check the test material and set-up. The participant was a student from the Department of Media. The participant performed the test in the same set-up as described below, interacting with the *Stacking Bills* prototype. Accordingly, the test material was minimally adjusted (some errors were corrected, some tests extended). Based on the par-

participant's comments during the think-aloud protocol I identified several quantifiable characteristics of insight which were later used to analyze the insights from the real tests.

SET-UP AND PROCEDURE

All except for one test were conducted in the same room on the University premises. A camera on a tripod was used to record the participants and screen during the session. The participants received the test-tasks on paper and the prototypes were presented on a MacBook Pro (Retina, 15-inch, Early 2013) equipped with a mouse. Before evaluating the actual prototypes, test participants were given a pretest assignment to assess their individual baseline-proficiency with big numbers. The participants were then presented with a short piece of text, that acted as an introduction to the monetary measure. After reading the text participants received a short questionnaire. Participants then watched a short video example of the think-aloud protocol, in order to make sure that they understood what to do. After this, participants were presented with one of the three prototypes. They were then asked to interact with the prototypes on a laptop as long as they wanted and voice their insights while doing so. The session was recorded on video for later analysis. After participants stopped interacting with the visualization, they were again asked to answer a question and perform a benchmark task related to the prototype. Lastly participants had the opportunity to comment on their experience with the strategy.

TASKS

All test material can be found in the appendix.

- **Pretest:** Participants were given a pretest to assess their individual baseline-proficiency with big numbers. Similar to Tretter, Jones and Minogue's (2006) *Scale Anchoring Objects assessment* test in which participants name objects representing scales from one nanometer to one billion meters, participants were asked to provide examples for objects/services/measures that matched the value of a given number from 10 to one billion.
- **Inspection of the news article:** Participants were handed a short news article about the U.S. budget deficit, acting as a primer for the prototype.

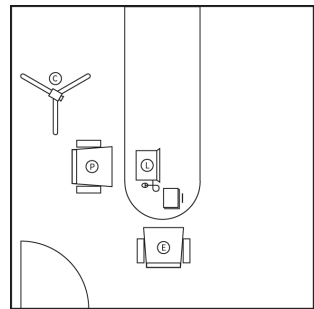


FIG.44. Test setup.
 E-Experimenter
 P-Participant
 C-Camera
 L-Laptop

- **Questionnaire:** The questionnaire consisted of three questions:

- *Have you heard/read about this number (\$425 billion / €346 billion) before?* This was asked to ensure that participants were unfamiliar with the number.

- *How actively do you follow economic issues?* Participants answered 5-point itemized scale, from *Not at all - I am not interested* to *I follow economic issues extensively and even research about them further*. This was asked to exclude participants that were too proficient in economic issues.

- *How confident are you that you understand the magnitude of the number you just saw (€346 billion)?* Participants were asked to evaluate themselves on a 5-point itemized scale from *I am not at all confident* to *I am very confident*. This was done to check their personal confidence beforehand, find out if the strategy improved their confidence and if their confidence matched their benchmark-results. The results from this question were, however not used in the final analysis, as they yielded no usable results.

- **Interaction with the prototypes + Think-aloud:** Participants were instructed to interact with the prototypes as long as they wanted. No tasks were assigned, since I was also interested in seeing their natural interaction with the prototypes. For the think-aloud protocol, I followed a »loose« version of the technique, as it is commonly practiced in usability testing. In such a think-aloud test, users interact with the system while continuously verbalizing their thoughts, feelings, insights, inferences and explanations. The experimenter stays silent, but can intervene if the user falls silent. Boren and Ramey (2000) give an in-depth analysis of how think-aloud in usability testing does not follow the strict rules originally intended by the inventors of the method, Ericsson and Simon, who forbid any type of reflection on the subject's part.

- **Self-evaluation:** After the interaction, participants were asked how confident they now felt, again on a 5-point itemized scale.

- **Benchmark Test:** For the benchmark test, participants were instructed to estimate an unfamiliar monetary measure in terms

of a particular strategy. This task is inspired by Barrio et al.'s (Barrio, Goldstein & Hofman 2016) estimation task. Result from this test were meant to give insight into whether participants could apply the strategies they saw. Depending on the prototype they were assigned to, their question was:

- *Using the same strategy you just saw (bundling €100 bills), how long would the bundles for the following values be?*
 - *Using the same strategy you just saw (paying off a sum at a rate of €100 per second), how long would it take to pay off the following values?*
 - *Using the same strategy you just saw (converting a sum into smaller, more familiar units), what could you buy with the following values?*
- **Discussion:** The last part of the test session consisted of a free form interview, in which I asked participants for opinions and comments on the visualization and the strategy. I also asked whether they had seen similar visualizations or visualization strategies before and whether they could remember any specific values. Furthermore, I specifically invited them to reflect on the aspects of interactivity and personalization in the prototypes.

5.3 Results

In my analysis I am following Saraiya, North & Duca's recommendation, who note that »[s]ince this evaluation method is more qualitative and subjective than quantitative and the number of participants is limited, a general comparison of tendencies in the results is most appropriate« (Saraiya, North & Duca 2005).

	Stacking Bills	Money Clock	Buying Power
Number of insights	17	11	18
Average number of insights	5.7	3.7	6
Average time (sec) to first insight	110	167	109
Average total time (sec) spend with prototype	386.7	347	480
Aggregated complexity of insights	25	17	40
Average complexity of insights	1.47	1.55	2.22
Number of unexpected insights	2	3	7
Number of incorrect insights	0	3	1

Overview of the collected metrics.

Evaluation of insight characteristics

Number of insights: Buying Power and Stacking Bills yielded very similar results in the overall (18/17) and average (6/5.7) number of insights. Money Clock performed much worse, leading to only 11 distinct observations overall and an average of 3.7 per participant. Variance, however, was relatively high within each prototype due to the small number of tests.

The lower number of insights for Money Clock could stem from the lower degree of interactivity and playfulness that this prototype offered. This seems to have led to less engagement and shorter time spend with the prototype.

Time: The average total time reports the time users spend with the prototype until they lost interest or thought they had understood the principle. On average, participants spend most time with the Buying Power prototype (about 8 minutes), a bit less time with the Stacking Bills prototype (around 6.5 minutes) and

the least amount of time with the Money Clock prototype (around 5.8 minutes). This lower time either indicates a more efficient visualization (personal observations from the test show no indications of this, actually the opposite), less interest in the strategy or that participants gave up due to complications with the interface. A reverse pattern can be seen in the average time needed to reach the first insight. Participants using the Money Clock prototype reached their first insight on average 2.8 minutes after starting the interaction. The participants using Buying Power and Stacking Bills, on the other hand, needed an average of 1.9 minutes to reach their first insights about the data. Lower times suggest that participants could find interesting issues faster and therefore understood the premise of the prototypes more easily.

Complexity: 1 to 3 points were given according to the complexity of the individual observations. The aggregated complexity of Buying Power was the highest with 40 points, second highest for Stacking Bills with 25 points and lowest for Money Clock with 17 points. Averaging the complexity in regards to the total number of insights per prototype shows that Money Clock and Buying Power produced relatively similar complexity with 1.55 and 1.47 points respectively. Buying Power yielded slightly more complex insights with an average of 2.2 points.

1 point: *Okay, I could get a lot of sail boats with this.*

2 points: *Three blocks in Helsinki is already €400 million.*

3 points: *So, if I could save a hundred Euros per second I still wouldn't be a millionaire in 99 days.*

Direct and unexpected insights: Participants using the Buying Power prototype overall reached more insights that were unrelated to the main goal of the respective strategy. Most commonly these unexpected insights concerned the price of the comparison objects and their differences.

»I thought passenger jets might be more expensive than cruise ships, but obviously not.«

Correctness: Overall, only four incorrect insights were recorded, one for Buying Power and three for Money Clock. All incorrect observations for Money Clock were reported by the same participant however, and originated from a misreading or mis-

understanding of the short scale. The same issue was observed several times with other participants during the pre-test and benchmark-test. Other participants also commented on the issue of number-naming systems while interacting with the prototype, but could ultimately name the numbers correctly in regards to the short scale naming conventions.

»If I wanna pay it in 99 days, it would take me 855 thousand [actually 855 million], I would have less than a million.«

Insight categories

The following seven general categories were identified from the insights collected during the think-aloud. Note: Insights can be assigned to several categories at the same time.

	Stacking Bills	Money Clock	Buying Power
personal	7	4	3
revelation	2	1	2
facts	6	3	1
summary	7	2	3
interpretation	1	4	4
inference	1	0	3
hypotheses	1	1	4

*Distribution of insights
acc. to category.*

Personal: Insights in this category are related to a personal experience of the participant. As all prototypes included some degree of interaction, all of them allowed users to situate themselves and their experience in the visualization. Not all participants took this opportunity, however. The results and personal observations during the test suggest that the Stacking Bills prototype was the most successful in encouraging connections with personal experiences. Participants intuitively tested the prototype with familiar locations, such as their home countries and their study places. Both Money Clock and Buying Power yielded

about the same amount of personal insights. Personal insights are especially important/interesting, as it can be assumed that participants attach the new information to an already existing mental model.

»I've flown that distance [from Helsinki to Riga] many times, that's familiar to me.«

Revelation: Revelations are observations that correct a previous misconception or correct a mental model. Revelations are most closely related to spontaneous insights (aha-moments), and sudden problem-solving. These types of insights are valuable as they can lead to future knowledge-building insights (Chang et al. 2009). However, this type of insight was fairly rare, and overall only five instances were recorded during the sessions.

»I didn't know that billion would have 9 zeros. I thought it would be more zeros.«

Facts: Often times participants were voicing singular facts extracted from the visualization, such as the time needed to pay of a certain amount of money. These observations come directly from the prototypes and therefore indicate that participants understood the premise of the visualization. They also give some understanding of which aspects participants found worthy to note aloud. These types of insights were interestingly rare for Buying Power, as most observations were of a summarizing nature. The reason for this could be that the numbers presented in Buying power were generally bigger than in the other prototypes. This leads to the question whether participants paid attention to the actual numbers.

»Oh - to Iran it would be €3,360 billion.«

Summary: Nearly all participants reached some kind of summarizing insight, summing up a process, relationship or fact in a very general, superficial way. Summarizing observations might be considered low level insights, as they do not reveal whether deeper or even any understanding was achieved or information was read correctly.

»That's a huge pile of bills.«

Interpretation: Rather than merely extracting or generalizing a fact, some participants tried to interpret those facts by explaining them in their own words and connecting them to other factors not shown in the visualizations. Interpretation can be desirable as it hints to the integration of new knowledge into existing schemas. Both Buying Power and Money Clock produced more insights of this type than Stacking Bills. This is not surprising since Stacking Bills uses a rather specific translation strategy that cannot easily be re-expressed, and as such might be less versatile.

»That's more than the population of Finland.«

Inference: Inference describes extrapolation on the basis of underlying patterns. The initial assumption was that participants would get a concrete understanding of the conversion strategies used (e.g. money to meter, money to time), and be able to make new predictions based on this. However, only one participant (interacting with Buying Power) seemed to show active understanding and tried to give extrapolations from the available examples during the interaction phase. This could also be a result of the think-aloud method used, which might not be suitable to prompt such an act of reflection. The benchmark test administered after the interaction directly tested the participants' ability to infer from the data.

»I guess I have to scroll a lot till one ends and the other one goes on for 10 more times.«

Hypotheses: Interestingly, interacting with the prototypes lead some participants to new questions and observations that were not intended by the visualization. Two participants assigned to the Buying Power prototype were especially interested in extending the context of the data and making connections to bigger social and economic issues. This behavior might stem from personal interests of the particular participants, but could also be caused by the prototype itself, as it is the only visualization that considers the actual value of the presented measure.

»That [amount of money] could keep the economy of Finland going for one or two years maybe.«

Test results

Pre-test: The pre-test was initially intended to assess the baseline numerosity of participants and their ability to visualize the magnitude of various values. For this, participants were asked to provide a »object« matching the value of seven numbers from 10 to one-billion. In the real testing, however, it came to show that the result could not be used for that purpose. This could either be a result of poorly worded instructions or the test itself. Out of 60 collected answers (2 participants did not provide answers to all values), roughly one half were non-quantifiable. This included answers such as »coins in a treasure chest«, »DNA chain« or »square meters of a mansion«. One could interpret answer such as these as a sign that the participant did not have a clear image of the magnitude in their minds, however this would be mostly guesswork. As such, the results of the pre-test were dismissed as a dependable variable. A full list of all answers can be found in the appendix.

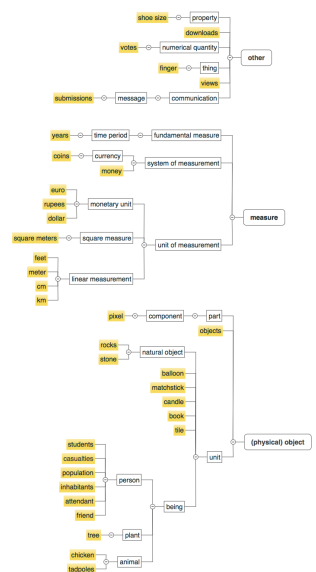
Nevertheless the result yielded another interesting finding, in the form of the units that participants choose for the different values. Using *WordNet*²², a lexical database developed at the Cognitive Science Laboratory of Princeton University, that provides a semantic network of words based on sets of cognitive synonyms, I categorizing the units according to their hypernyms²³. The following patterns emerged:

Units could be grouped into three big categories: Measures, Objects, and Other. In the chart, prominent subgroups of these categories were highlighted, such as linear measures, monetary measures, and persons. The dominance of monetary measures in the one hundred and one million category was interesting is for example. The 10 thousand category is the least clearly defined, as was evident from the participants' comments during the test. Most identified this as the hardest category, and could not immediately recall any connection with that number, in contrast to numbers such as 1 million, which is commonly encountered in f.e. news media.

²² wordnet.princeton.edu

²³ *hypernym: An umbrella term for words with more specific meanings.*

FIG.45. Sorting of results from the pre-test according to WordNet hierarchies.



Values in the pre-test

	10	100	1,000	10,000	100,000	1 million	1 billion
Units listed by participants	feet	Euro	Euro	Euro	miles	rupees	money
	centimeter	Euro	Euro	feet	meter	rupees	money
	meter	Euro	squaremeters	squaremeters	kilometer	Euro	Euro
	kilometer	Dollar	years	coins	books	Dollar	youtube views
	shoe size	meter	entries	typhoons	trees	votes	dna
	fingers	years	tadpoles	chicken	inhabitants	tiles	stones
	rocks	baloons	pixels	trees	inhabitants	downloads	population
	students	matchsticks	casualties	inhabitants		objects	population
	friends	candles	attendants			inhabitants	population

TYPES OF UNITS



Units provided by participants during the pre-test, categorized by types.

Benchmark-test: The benchmark-test measured 1) how well participants remembered the initially shown conversion and 2) how well they could apply the conversion strategy, i.e. how well they could infer. Since the sample is extremely small, statistical analysis of this data is not possible, but an individual analysis can be attempted. Results from the benchmark-test were coded after a seven-point accuracy scale modeled after Tretter et al.'s (2006) *Accuracy Codes*. The scale measures how close the provided answer is to the real value within different factors of ten.

- 3: object listed was too small by at least a factor of 1000.
- 2: object listed was too small by at least a factor of 100 (but not by 1000).
- 1: object listed was too small by at least a factor of 10 (but not by 100).
- 0: object listed was within factor of 10 (larger or smaller) of requested size.
- 1: object listed was too big by at least a factor of 10 (but not by 100).
- 2: object listed was too big by at least a factor of 100 (but not by 1000).
- 3: object listed was too big by at least a factor of 1000.

In general, performance on the test varied greatly, even within prototypes. In regards to recall, most participants managed to remember the initially shown value (after interacting with the prototype), except for two participants, one of which provided highly inconsistent and incorrect answers all throughout the test. As for inferring unknown values, the majority of participants performed very poorly on the first value (€1 million), with more than half giving overestimations by a factor of 1000. This is espe-

cially surprising for the participants interacting with the Stacking Bills prototype, since the thickness of a bundle worth €1 million was explicitly mentioned in the introductory text. The enormous mis-estimation is less surprising for the Money Clock users, as participants had to extrapolate from 346 billion = 109 years to 1 million = 2 hours, which is a calculation not many people can do on the spot. The results for the second value (€120 million) were equally mixed, and no pattern emerged. Since the third value was €700 billion, about half of the participants arrived at their answer by simply doubling their estimate for €350 billion. This is a reasonable step for the Money Clock users, which had to give their estimate as a time unit. While this logic could have been applied for all strategies, two participants from Stacking Bills and Buying Power, respectively, did not follow this logic and rather tried to give new estimates.

Participants' comments on the prototypes

General: Most participants agreed that personalization was interesting and important to understand the magnitude of the numbers. Some participants noted that they would not, in a realistic setting, read the introduction texts of the prototypes, but would start interacting immediately. One participant made an interesting comment regarding the differences between short and long scale, saying that he would like to see a clear statement of which number naming system was used. When asked whether they had encountered similar attempts at explaining large number before, most participants could list at least one other instance. Many were related to social issues such as CO₂ production, food waste or poverty.

Stacking Bills: Participants found Stacking Bills fun and easy to use. One feature that particularly helped participants was the »translation« of the numerals into short scale names. Some participants wished for more elaborate interaction functionality, but found that optional.

Money Clock: Two out of three participants found Money Clock not very engaging, while one felt distressed and anxious at the sight of the countdown. All participants wished for an option to change the rate of pay-off rather than the time frame. They also

noted that a visual depiction of the shrinking amount of money would have been helpful.

Buying Power: Most participants wished for more context in the Buying Power prototype, either to external topics or between the available options. One usability issue led to momentary insecurity and confusion in the participants.

5.4 Discussion of results

The results from the user test showed that the most numerous insights occurred with the Stacking Bills and Buying power prototypes, likely because they allowed for more interaction and exploration on the user's side. As expected, the ability to personalize the prototypes was important for most participants. This became evident during the interactions and the subsequent discussions. Through this feature, some participants seemed to experience, at least momentarily, an »aha« moment, in which the magnitude of the numbers gained some personal meaning for them. Most successful in providing personal insights, was the Stacking Bills prototype, as users intuitively dragged the »money stack« to familiar places. Following this logic, it would have been beneficial if users of the Money Clock prototype would have been able to change the »pay off« rate to a personal number, such as their own »income rate«.

Users of both the Stacking Bills and Buying Power prototype reached their first insights at round 1.9 minutes, much faster than the users of the Money Clock prototype. Users also spent the least amount of time with this prototype. From observations it seems like the concept of »paying off« was harder to understand and sometimes led to confusions as to what was being counted. Using money as a value seemed to lead to more critical and complex insights, as can be seen from the Buying Power results. This poses the question if abstraction, as it was done in Stacking Bills, is advisable. Buying Power, however, had the issue of explaining large numbers through more large, hard-to-understand numbers, as some participants noted. Using intermediate steps and applying the Zoom strategy could have helped with this issue.

Although the design of the prototypes was kept purposefully simple, interaction and interface components proved to be a crucial factor. Some usability issues seemed to severely interfere with the users' ability to gain insights from the prototypes. In the Buying Power prototype, for example, users had a hard time comparing the lengths of the unit charts, and only few tried to scroll to the end of the chart. A possible solution for this could be to provide a miniature overview of the whole chart, similar to the mini locator maps of text-editors such as *Sublime*. Adding little annotations or process indicators, such as in *If the Moon was Only 1 Pixel in Size*, might also encourage users to continue scrolling till the end, getting the full experience. Especially in the Money Clock prototype, usability and interface issues hindered the acquisition of insight. The skip forward button attracted users' attention, but did not deliver any interesting results from the users' perspective.

A factor that should not be underestimated in the acquisition of insights is user motivation. Users that noted in advance that they were not interested in »money things«, generally interacted for a shorter period of time and gained less insights from the prototypes. Clearly, an initial interest in the topic is essential for the strategies to be effective. This problem is obviously hard to avoid in such a controlled test setting with forced interactions. In real life, however, an engaging pull-in can be created, for example, through the use of narrative or emotional framing.

Unfortunately, the estimation benchmark test could not provide conclusive evidence as to whether participants were able to adopt and adapt the presented metaphors.

Overall, the evaluation highlighted the importance of interactivity and personalization. Although the results are not conclusive and the test setup had some shortcomings, sessions gave some indication that the concept of spending or paying-off did not support participants' understanding. Visualizing the actual value of the monetary number seems to be the most promising strategy, but a lot of care has to be put into designing the visualization. Using familiar spaces as a reference point appears to be a good method to compellingly illustrate the size differences between numbers. This strategy on its own, however, might not allow for much inference or interpretation on the reader/viewer's side.

6 Conclusion

The goal of this thesis was to compare the effectiveness of different strategies for improving the comprehension of large numbers through Information Visualization. The following will summarize the most important findings of this thesis, list its main contributions, discuss its limitations and give an outlook for possible future research.

6.1 Contribution

In order to investigate the effectiveness of strategies aiming at making large numbers meaningful to the average reader/viewer in an Information Visualization context, I first described theories from Information Visualization and number cognition. This review yielded important visual, structural, and cognitive implications for the design of visual strategies. Furthermore, to be able to accurately describe and evaluate these strategies, I reviewed and extended an existing framework for the visual depiction of extreme measures. The final adapted framework consisted of seven approaches: *Unitization*, *Comparison*, *Analogy*, *Zoom*, *One-to-One Correspondence*, and *Integration of time/space/sound*. This framework was further elaborated by findings from the theoretical review and several examples of existing visualizations. Based on this theoretical framework and a corpus of visualizations, I created three visualization prototypes (*Stacking Bills*, *Money Clock*, and *Buying Power*) specifically for the topic of monetary measures. Following an insight-based approach I evaluated and compared these prototypes in terms of amount, category, and amount of insights created in a series of user tests.

This thesis contributes to the field of Information Visualization by connecting theories from Information Visualization and number cognition regarding the presentation and understanding of large numbers. By outlining a framework of strategies and gathering a sizable collection of examples for existing visualizations, it gives guidelines and a source of inspiration for designers who want to create visualizations that support the comprehension of such numbers. This thesis also serves as an example of an insight-based evaluation approach, and through it different categories of insight were defined, which can be valuable for future research.

6.2 Limitations

I tried to provide both a comprehensive and relevant overview of all topics related to the topic of this thesis, however, both Information Visualization and especially numerical cognition are extensive, interdisciplinary, and highly complex fields. As such, a review of these fields can only ever be a narrow view based on a personal decision. Regarding the focus on monetary measures,

it proved to be a somewhat difficult unit, as its value depends very much on the context it is in and, moreover, its physicality and its worth have no natural connection between them. One limitation of the evaluation component of this thesis is the relatively small number of participants in the user tests and the big individual differences in cultural background and motivation between participants. As Hullman and Diakopoulos (2011) note, individual differences between readers/viewers can majorly affect the interpretation and perception of visualizations. Based on these factors it is questionable if a repeated test would yield the same results.

Aside from the reliability of the results, their validity also has to be critically considered. The heterogeneity of the prototypes, for example, is problematic, especially in regards to the varying levels of interactivity and refinement. Results from the Money Clock prototype in particular suffered from mis-designed interactions. Both the pre- and benchmark-test also had some flaws, which became evident after the testing session; The aforementioned differences between the prototypes, for example, limited the comparability between the benchmark-test results. The small number of questions furthermore made the test prone to fleeting errors such mis-reading the numerals (such a mistake was directly observed during the testing sessions). A more extensive test with more balanced preconditions would be needed to arrive at statistically significant results from this type of benchmark-test. Regarding the recorded measures, using *time till first insights* might be problematic. Although it was measured in the original study by Saraiya et al., such measures are not reliable when users have to verbalize their thoughts while interacting with a prototype (usability.gov). Lastly, the reported measures were the result of a rather short term usage, as interactions generally lasted only from 3 to 12 minutes. It is questionable if any deep insights can be won in such a short time frame.

6.3 Future Research

In the attempt to find ways to express large numbers in more comprehensible ways, this thesis has provided some answers, but in the process, has raised even more questions for future research. Results from the evaluation test could be verified with

a bigger study involving more test subjects and an improved evaluation approach. It would also be interesting to study the effects of these strategies over a longer period of time and in a more natural setting.

The literature discussed here also opens up further questions on the formation of mental models, the roles of metaphors, and the mechanisms of learning. Additional research into these areas could prove beneficial to the further development of the issue.

Looking ahead, the next step in research on this topic could be to investigate less invasive techniques, i.e. strategies that can be incorporated into existing visualization types. Examples of such unobtrusive inclusions can be found occasionally, but have not been formalized in any way.

All in all, improving the comprehension of large numbers is an important issue, as the amount of information and the magnitude thereof is growing exponentially. Our ability to process these numbers, however, is slow to catch up and needs to be supported in any way possible. Only if we truly understand can we make meaningful decisions.

7 Bibliography

- Agafonkin, V. 2017, , *Leaflet*. Available: <http://leafletjs.com>.
- Agnew, R. & Moyer, D. 1999, »Age & Race« in *Understanding USA*, ed. R.S. Wurman, TED Conferences.
- Barrio, P.J., Goldstein, D.G. & Hofman, J.M. 2016, »Improving Comprehension of Numbers in the News«, *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* ACM, New York, NY, USA, pp. 2729.
- Blastland, M. & Dilnot, A.W. 2009, *The numbers game : the commonsense guide to understanding numbers in the news, in politics, and in life*, Penguin Group, New York, N.Y.
- Bloom, B.S., Anderson, L.W. & Krathwohl, D.R. 2001, *A taxonomy for learning, teaching, and assessing : a revision of Bloom's Taxonomy of educational objectives*, Abridged ed edn, Longman, New York.
- Booth, J.L. & Siegler, R.S. 2006, »Developmental and individual differences in pure numerical estimation«, *Developmental psychology*, vol. 42, no. 1, pp. 189-201.
- Boren, T. & Ramey, J. 2000, »Thinking aloud: reconciling theory and practice«, *IEEE Transactions on Professional Communication*, vol. 43, no. 3, pp. 261-278.
- Boy, J., Pandey, A.V., Emerson, J., Satterthwaite, M., Nov, O. & Bertini, E. 2017, »Showing People Behind Data: Does Anthropomorphizing Visualizations Elicit More Empathy for Human Rights Data?«, *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* ACM, New York, NY, USA, pp. 5462.
- Cairo, A. 2016, February 3-last update, *Draft for the Epilogue of The Truthful Art*. Available: <http://www.thefunctionalart.com/2016/02/empathy-and-visualization.html> [2017, 6/13].
- Card, S.K., Mackinlay, J.D. & Shneiderman, B. 1999, *Readings in information visualization : using vision to think*, Morgan Kaufmann, San Francisco (CA).
- Carpendale, S. 2008, »Evaluating Information Visualizations« in *Information Visualization*, eds. A. Kerren, J.T. Stasko, J. Fekete & C. North, Springer-Verlag, Berlin, Heidelberg, pp. 19-45.
- Case, N. 2014, Sept 8-last update, *Explorable Explanations*. Available: <http://blog.ncase.me/explorable-explanations/> [2018, 4/2].
- Chang, R., Ziemkiewicz, C., Green, T.M. & Ribarsky, W. 2009, »Defining Insight for Visual Analytics«, *IEEE Comput.Graph.Appl.*, vol. 29, no. 2, pp. 14-17.
- Cheek, K.A. 2017, »Do Indonesian Children's Experiences with Large Currency Units Facilitate Magnitude Estimation of Long Temporal Periods?«, *Research in Science Education*, vol. 47, no. 4, pp. 889-911.
- Chevalier, F., Vuillemot, R. & Gali, G. 2013, »Using Concrete Scales: A Practical Framework for Effective Visual Depiction of Complex Measures«, *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2426-2435.
- Cleveland, W.S. & McGill, R. 1984, »Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods«, *Journal of the American Statistical Association*, vol. 79, no. 387, pp. 531-554.

- Cohen Kadosh, R. & Dowker, A. (eds) 2016, *The Oxford Handbook of Numerical Cognition*, 1st edn, Oxford University Press, New York, NY.
- Dehaene, S. 1992, »Varieties of numerical abilities«, *Cognition*, vol. 44, no. 1-2, pp. 1-42.
- Dehaene, S. 1999, *The number sense: how the mind creates mathematics*, Penguin Books, London.
- Dehaene, S., Bossini, S. & Giraux, P. 1993, »The mental representation of parity and number magnitude«, *Journal of Experimental Psychology: General*, vol. 122, no. 3, pp. 371-396.
- Dehaene, S. & Cohen, L. 1995, »Towards an Anatomical and Functional Model of Number Processing«, *Mathematical cognition*, vol. 1, no. 1, pp. 83-120.
- developers.arcgis.com, *Create geodesic geometries* [Homepage of Esri], [Online]. Available: <https://developers.arcgis.com/java/10-2/guide/create-geodesic-geometries.htm> [2018, 3/1].
- eamesoffice.com, *POWERS OF TEN AND THE RELATIVE SIZE OF THINGS IN THE UNIVERSE*. Available: <http://www.eamesoffice.com/the-work/powers-of-ten/> [2018, 4/4].
- Encyclopædia Britannica 2011, April 15-last update, *Large numbers* [Homepage of Encyclopædia Britannica, inc.], [Online]. Available: <https://www.britannica.com/topic/large-numbers-1765137> [2018, February 12].
- explorabl.es, *Explorable Explanations*. Available: <http://explorabl.es/> [2018, 4/4].
- Feigenson, L., Dehaene, S. & Spelke, E. »Core systems of number«, *Trends in cognitive sciences*, vol. 8, no. 7, pp. 307-314.
- Forssman, F. & de Jong, R. 2004, *Detailtypografie : Nachschlagewerk für alle Fragen zu Schrift und Satz*, 2.th edn, Schmidt, Mainz.
- Frith, C.D. & Frith, U. 1972, »The solitaire illusion: An illusion of numerosity«, *Perception & Psychophysics*, vol. 11, no. 6, pp. 409-410.
- Gleicher, M., Albers, D., Walker, R., Jusufi, I., Hansen, C.D. & Roberts, J.C. 2011, »Visual comparison for information visualization«, *Information Visualization*, vol. 10, no. 4, pp. 289-309.
- Harris, J. 2015, January 15-last update, *Connecting with the Dots*. Available: <https://source.opennews.org/articles/connecting-dots/> [2018, 3/23].
- Hofstadter, D.R. 1982, »On number numbness« in *Metamagical Themas: Questing for the Essence of Mind and Pattern* Penguin, London, pp. 115-135.
- Holmes, N. & Bagby, M. 1999a, »Crime« in *Understanding USA*, ed. R.S. Wurman, TED Conferences.
- Holmes, N. & Bagby, M. 1999b, »Federal Debt« in *Understanding USA*, ed. R.S. Wurman, TED Conferences.
- Hullman, J. & Diakopoulos, N. 2011, »Visualization Rhetoric: Framing Effects in Narrative Visualization«, *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 12, pp. 2231-2240.
- Huygens, C. 1659, *Cristiani Hugenii Zulichemii, Const. f. Systema Saturnium : sive, De causis mirandarum Saturni phaenomenon, et comite ejus Planeta Novo*, Hagae-Comitis: Ex typographia Adriani Vlacq.
- Jones, M.G. & Taylor, A.R. 2009, »Developing a sense of scale: Looking backward«, *Journal of Research in Science Teaching*, vol. 46, no. 4, pp. 460-475.
- Kosara, R. 2017, »An Argument Structure for Data Stories«, *Proceedings of the Eurographics/IEEE VGTC Symposium on Visualization (EuroVis)*, eds. B. Kozlíková, T. Schreck & T. Wischgoll.

- Kosara, R. & Mackinlay, J. 2013, »Storytelling: The Next Step for Visualization«, *Computer*, vol. 46, no. 5, pp. 44-50.
- Lam, H., Bertini, E., Isenberg, P., Plaisant, C. & Carpendale, S. 2012, »Empirical Studies in Information Visualization: Seven Scenarios«, *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 9, pp. 1520-1536.
- Landy, D., Charlesworth, A. & Ottmar, E. 2017, »Categories of Large Numbers in Line Estimation«, *Cognitive science*, vol. 41, no. 2, pp. 326-353.
- Landy, D., Silbert, N. & Goldin, A. 2013, »Estimating Large Numbers«, *Cognitive Science*, vol. 37, no. 5, pp. 775-799.
- Lidwell, W., Holden, K. & Butler, J. 2010, *Universal Principles of Design, Revised and Updated : 125 Ways to Enhance Usability, Influence Perception, Increase Appeal, Make Better Design Decisions, and Teach through Design*, Rockport Publishers, Osceola.
- Ma, J. 2007, , *Visitor's Drawings of Small*. Available: http://www.nisenet.org/catalog/evaluation/visitors_drawings_small_2007_formative_evaluation [2017, 12/11].
- Mackinlay, J. 1986, »Automating the Design of Graphical Presentations of Relational Information«, *ACM Trans.Graph.*, vol. 5, no. 2, pp. 110-141.
- Masud, L., Valsecchi, F., Ciuccarelli, P., Ricci, D. & Caviglia, G. 2010, »From Data to Knowledge - Visualizations As Transformation Processes Within the Data-Information-Knowledge Continuum«, *Proceedings of the 2010 14th International Conference Information Visualisation* IEEE Computer Society, Washington, DC, USA, pp. 445.
- Munroe, R., *Real Estate*. Available: <https://xkcd.com/1894/> [2018, 3/1].
- Neurath, O. & Neurath, M. 1980, *International picture language / Internationale Bildersprache*, Department of Typography & Graphic Communication, University of Reading, Reading.
- Nielsen, J. 2000, March 19-last update, *Why You Only Need to Test with 5 Users* [Homepage of Nielsen Norman Group], [Online]. Available: <https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/> [2018, 2/28].
- Nieman, A. 2011, »Concrete vs abstract visualisation: The real world as a canvas for data visualisation« in *Proceedings of the ADS-VIS 2011: Making visible the invisible: art, design and science in data visualisation*, ed. M. Hohl, University of Huddersfield, Huddersfield, pp. 49-56.
- North, C. 2006, »Toward Measuring Visualization Insight«, *IEEE Comput.Graph.Appl.*, vol. 26, no. 3, pp. 6-9.
- Nuerk, H., Moeller, K. & Willmes, K. 2016, »Multi-digit Number Processing: overview, conceptual clarifications, and language influences« in *The Oxford Handbook of Numerical Cognition*, eds. C.K. Roi & A. Dowker, 1st edn, Oxford University Press, New York, NY, pp. 106-139.
- nytimes.com 2015, August 18-last update, *U.S. Budget Deficit Rose in July, but 8-Year Low Is Expected for Year* [Homepage of The New York Times Company], [Online]. Available: <https://nyti.ms/2BQzJje> [2018, 1/11].
- Parker, J.D. 2011, »Using Google Earth to Teach the Magnitude of Deep Time«, *Journal of College Science Teaching*, vol. 40, no. 5, pp. 23-27.
- Paulos, J.A. 2001, *Innumeracy: mathematical illiteracy and its consequences*, Hill and Wang, New York (N.Y.).

- Plaisant, C. 2004, »The Challenge of Information Visualization Evaluation«, *Proceedings of the Working Conference on Advanced Visual Interfaces* ACM, New York, NY, USA, pp. 109.
- Resnick, I., Davatzes, A., Newcombe, N.S. & Shipley, T.F. 2017a, »Using analogy to learn about phenomena at scales outside human perception«, *Cognitive Research*, vol. 2, no. 1, pp. 21. Epub 2017 Mar 20 doi:10.1186/s41235-017-0054-7.
- Resnick, I., Davatzes, A., Newcombe, N.S. & Shipley, T.F. 2017b, »Using Relational Reasoning to Learn About Scientific Phenomena at Unfamiliar Scales«, *Educational Psychology Review*, vol. 29, no. 1, pp. 11-25.
- Reuben, A. 2014, November 28-last update, *Headline numbers: Help with coping with big figures* [Homepage of BBC], [Online]. Available: <http://www.bbc.com/news/business-30206857> [2018, 4/8].
- Saraiya, P., North, C. & Duca, K. 2005, »An insight-based methodology for evaluating bioinformatics visualizations«, *IEEE Transactions on Visualization and Computer Graphics*, vol. 11, no. 4, pp. 443-456.
- Sasanguie, D., Göbel, S.M., Moll, K., Smets, K. & Reynvoet, B. 2013, »Approximate number sense, symbolic number processing, or number-space mappings: What underlies mathematics achievement?«, *Journal of experimental child psychology*, vol. 114, no. 3, pp. 418.
- Segel, E. & Heer, J. 2010, »Narrative Visualization: Telling Stories with Data«, *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, no. 6, pp. 1139-1148.
- Shneiderman, B. 1996, »The eyes have it: a task by data type taxonomy for information visualizations«, *Proceedings 1996 IEEE Symposium on Visual Languages*, pp. 336.
- Shneiderman, B. & Plaisant, C. 2005, *Designing the user interface : strategies for effective human-computer interaction*, 4th ed edn, Pearson/Addison Wesley, Boston.
- Skau, D. & Kosara, R. 2016, »Arcs, Angles, or Areas: Individual Data Encodings in Pie and Donut Charts«, *Proceedings of the Eurographics / IEEE VGTC Conference on Visualization* Eurographics Association, Goslar Germany, Germany, pp. 121.
- Slobin, S. 2014, April 24-last update, *What If the Data Visualization Is Actually People?*. Available: <https://source.opennews.org/articles/what-if-data-visualization-actually-people/> [2018, 3/23].
- Slovic, S. & Slovic, P. (eds) 2015, *Numbers and nerves : information, emotion, and meaning in a world of data*, Oregon State University Press, Corvallis.
- Song, M. & Quintana, C. 2012, »Representing Too Small to See As Too Small to See with Temporal Representation«, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* ACM, New York, NY, USA, pp. 1441.
- Spence, R. 2014a, »Interaction« in *Information Visualization: An Introduction* Springer International Publishing, Cham, pp. 173-223.
- Spence, R. 2014b, »Issues« in *Information Visualization: An Introduction*, ed. R. Spence, Springer International Publishing, Cham, pp. 21-39.
- Spence, R. 2007, *Information visualization : design for interaction*, 2. ed edn, Pearson/Prentice Hall, Harlow.
- Tretter, T.R., Jones, M.G., Andre, T., Negishi, A. & Minogue, J. 2006, »Conceptual boundaries and distances: Students' and experts' concepts of the scale of scientific phenomena«, *Journal of Research in Science Teaching*, vol. 43, no. 3, pp. 282-319.

- Tretter, T.R., Jones, M.G. & Minogue, J. 2006, »Accuracy of scale conceptions in science: Mental maneuverings across many orders of spatial magnitude«, *Journal of Research in Science Teaching*, vol. 43, no. 10, pp. 1061-1085.
- Tufte, E.R. 2001, *The visual display of quantitative information*, 2nd ed edn, Graphics Press, Cheshire (CN).
- Tufte, E.R. 1997, *Visual explanations : images and quantities, evidence and narrative*, 2. pr. with rev edn, Graphic Press, Cheshire (CT).
- Tufte, E.R. 1991, *Envisioning information*, 2nd pr., with rev edn, Graphics Press, Cheshire (CT).
- usability.gov , *Running a Usability Test* [Homepage of U.S. Department of Health & Human Services], [Online]. Available: <https://www.usability.gov/how-to-and-tools/methods/running-usability-tests.html> [2018, 4/6].
- van Dijck, J., Ginsburg, V., Girelli, L. & Gevers, W. 2016, »Linking Numbers to Space: From the Mental Number Line Towards a Hybrid Account« in *The Oxford Handbook of Numerical Cognition*, eds. R. Cohen Kadosh & A. Dowker, 1st edn, Oxford University Press, New York, NY, pp. 89-105.
- van Wijk, J.J. 2013, »Evaluation: A Challenge for Visual Analytics«, *Computer*, vol. 46, no. 7, pp. 56-60.
- Victor, B. 2011, March 10-last update, *Explorable Explanations*. Available: <http://worrydream.com/ExplorableExplanations/> [2018, 4/2].
- Ware, C. 2008, *Visual thinking for design*, Morgan Kaufmann, Burlington, Mass.
- Ware, C. 2004, *Information visualization : perception for design*, 2nd ed edn, Morgan Kaufmann, San Francisco (CA).
- Wikipedia, *Anscombe's quartet*. Available: https://en.wikipedia.org/wiki/Anscombe%27s_quartet [2018, 4/5].
- Wurman, R.S. 2001, *Information anxiety. 2*. Que, Indianapolis (IN).
- Yi, J.S., Kang, Y., Stasko, J.T. & Jacko, J.A. 2008, »Understanding and Characterizing Insights: How Do People Gain Insights Using Information Visualization?«, *Proceedings of the 2008 Workshop on BEyond Time and Errors: Novel evaLuation Methods for Information Visualization* ACM, New York, NY, USA, pp. 4:1.
- Zer-Aviv, M. 2015, October 19-last update, *DataViz—The UnEmpathetic Art*. Available: <https://responsibledata.io/2015/10/19/dataviz-the-unempathetic-art/> [2018, 3/23].

8 Image Sources

- FIG.1: own material. Data from https://en.wikipedia.org/wiki/Anscombe's_quartet
- FIG.2: own material. Data from https://en.wikipedia.org/wiki/Anscombe's_quartet
- FIG.3: Extract from <https://xkcd.com/558/>
- FIG.4: Retrieved from <https://archive.org/details/CristianiHugeni00Huyg>
- FIG.5: Adapted from Segel, Heer (2010)
- FIG.6: © Axel Mauruszat, https://commons.wikimedia.org/wiki/File:Stolperstein_Else_Liebermann_von_Wahlendorf_Berlin_Budapester_Strasse.jpg
- FIG.7: <http://www.periscopic.com/our-work/more-than-400000-stolen-years-an-examination-of-u-s-gun-murders-in-2010>
- FIG.8: <https://www.nytimes.com/interactive/2014/02/09/opinion/minimum-wage.html>
- FIG.9: <http://ncase.me/trust/>
- FIG.10: Adapted from Dehaene (1992)
- FIG.11: Adapted from Landy, Silbert & Goldin (2013)
- FIG.12: <https://www.strudel.org.uk/blog/astro/001012.shtml>
- FIG.13: <http://science.jrank.org/kids/pages/243/Models-in-Astronomy.html>
- FIG.14: <https://www.flickr.com/photos/carltonreid/7999178447>
- FIG.15: http://farm4.staticflickr.com/3768/9142576908_b268405d97_o.jpg
- FIG.16: Examples from Chevalier et al.'s corpus. https://docs.google.com/spreadsheets/d/1CwiiDMynLP8bcTs5d9smO_QqE_xskpZWjiKOx01dCiY/pub?single=true&gid=2&output=html
- FIG.17: Examples from Chevalier et al.'s corpus. https://docs.google.com/spreadsheets/d/1CwiiDMynLP8bcTs5d9smO_QqE_xskpZWjiKOx01dCiY/pub?single=true&gid=2&output=html
- FIG.18: Reed, Moyer (1999)
- FIG.19: <https://qz.com/156522/how-many-months-it-takes-an-average-worker-to-earn-what-the-ceo-makes-in-an-hour>
- FIG.20: <http://neal.fun/spend>
- FIG.21: http://demonocracy.info/infographics/usa/us_debt/us_debt.html
- FIG.22: <https://waitbutwhy.com/2014/03/combined-wealth-world.html>
- FIG.23: <https://www.washingtonpost.com/graphics/politics/trump-presidential-budget-2018/?tid=graphics-story>

- FIG.24: https://www.washingtonpost.com/graphics/2017/national/fair-representation/?utm_term=.53e0f26e510e
- FIG.25: <http://longstreet.typepad.com/thesciencebookstore/2016/03/measuring-things-with-ships.html>
- FIG.26: <https://www.youtube.com/watch?v=P5yxFtWdDcc&feature=youtu.be>
- FIG.27: http://joshworth.com/dev/pixelspace/pixelspace_solarsystem.html
- FIG.28: <https://xkcd.com/980/huge>
- FIG.29: <https://www.forbes.com/special-report/2013/what-a-billion-dollars-buys-you.html>
- FIG.30: <http://www.eamesoffice.com/the-work/powers-of-ten> & <https://www.museum-joanneum.at/en/press/press-current/events/event/3700/measuring-the-world-3>
- FIG.31: Adapted from Dehaene (1999)
- FIG.32: Adapted from Frith and Frith (1972)
- FIG.33: <http://www.chrisjordan.com/gallery/rtn/#ben-franklin>
- FIG.34: <https://fivethirtyeight.com/features/gun-deaths/>
- FIG.35: <http://content.time.com/time/business/article/0,8599,1850269,00.html>
- FIG.36: <http://www.lightyear.fm/>
- FIG.37: own material
- FIG.38: own material
- FIG.39: own material
- FIG.40: own material
- FIG.41: own material
- FIG.42: own material
- FIG.43: Data from www.vispubdata.org/site/vispubdata
- FIG.44: own material
- FIG.45: own material

9 Appendix

Initials

Age

Country of origin

Native language

Highest degree or level of school you have completed

Profession / field of study

Your personal information will only be used to validate the test data for my thesis. If I publish any of your information in my thesis, it will be in an anonymized form or as part of a general statistic. I will handle your information confidentially and not share it with any third party. Likewise I will only use your video recording to analyze this session and will not publish any excerpts or stills without your explicit permission.

Please write an 'object' that matches the value of each number listed.
An 'object' could be a measure, amount, price, distance etc.

number	'object' matching the number
10 (ten)	
100 (1 hundred)	
1,000 (1 thousand)	
10,000 (10 thousand)	
100,000 (100 thousand)	
1,000,000 (1 million)	
1,000,000,000 (1 billion)	

Aug. 12, 2015

U.S. Budget Deficit Rose in July, but Forecasts See Progress

WASHINGTON — The United States government ran a much higher budget deficit in July than a year ago, but it is still on track for the lowest full-year deficit in eight years.

The Congressional Budget Office estimates that the year-end total will drop to around \$425 billion (**€346 billion**), making it the lowest deficit since 2007.

Have you heard/read about this number (\$425 billion / €346 billion) before?

- yes
- no

How actively do you follow economic issues?

- Not at all – I am not interested.
- I hear about general issues sometimes.
- I might read an article if I come across it, but I don't follow it further.
- I actively read articles about certain economic topics.
- I follow economic issues extensively and even research about them further.

How confident are you that you understand the magnitude of the number you just saw (€346 billion)?

- I am not at all confident.
- I am a bit confident.
- I am somewhat confident.
- I am quite confident.
- I am very confident.

How confident are you **now** that you understand the magnitude of the number you just saw (€346 billion)?

- I am much less confident.
- I am a bit less confident.
- I am equally confident.
- I am a bit more confident.
- I am much more confident.

Using the same strategy you just saw (bundling €100 bills),
how long would the bundles for the following values be?

You can use concrete numbers or describe the distance geographically.

value

distance

€ 1,000,000 (1 million)

€ 120,000,000 (120 million)

€ 350,000,000,000 (350 billion)

€ 700,000,000,000 (700 billion)

Using the same strategy you just saw
(paying off a sum at a rate of €100 per second),
how long would it take to pay off the following values?

value

time

€ 1,000,000 (1 million)

€ 120,000,000 (120 million)

€ 350,000,000,000 (350 billion)

€ 700,000,000,000 (700 billion)

Using the same strategy you just saw
(converting a sum into smaller, more familiar units),
what could you buy with the following values?

value

objects

€ 1,000,000 (1 million)

€ 120,000,000 (120 million)

€ 350,000,000,000 (350 billion)

€ 700,000,000,000 (700 billion)

Pre-test answers						
10	100	1,000	10,000	100,000	1 million	1 billion
10 cm on a ruler	100 €	a nice bike	(area of) a football field	miles of a 4 year old car	(population of) Glasgow	(magnitude of) state debt (in the Czech republic)
closer circle of friends	(price of a high-end) keyboard	attendance at a music festival	(population of) Kansas	in habitants of a town in India	downloads of an app (Woit)	DNA chain
cloth	balloons	casualties in 2 weeks in Gaza	bank savings	height of the space station in meters	house (price)	population of China
feet across the street	candles in a cemetery	entries to a design competition	chicken in an industrial chicken coop	distance to the moon	objects in a city (seen from a satellite)	population of India
fingers	distance of a sprint	m2 of a mansion	coins in a treasure chest	dense forrest (finland)	price money on who wants to be a millionaire	population of India
my way to school by bike	euro bill	medium price for an average laptop	feet under the sea	books in a library	price money on who wants to be a millionaire	startup culture
rocks	matchsticks in a big matchbox	pixel width of a desktop layout (-24px)	not so dense forrest (mediterranean)	(population of) Lahti	stones on the ground	views on gangnam style (youtube)
students in a room	price of a (conference) ticket	tadpoles in a pond	thundering typhoons'		tiles	what governments deal with
US shoe size	years	years	- captain Haddock		votes here or there in a general election (in germany)	
					what you win in a game show	

Benchmark-test answers			
1 million - answers	Value	correct	AC
2,5 houses	1.000.000	1.000.000	0
1000 VR headsets	275.000	1.000.000	0
2 cruiseships	2.000.000.000	1.000.000	3
23h	82.800	3.600	1
6 monts	15.768.000	3.600	3
99 days	8.553.600	3.600	3
10.000 km	10.000.000	1	3
Helsinki to Espoo	11.000	1	3
100 m	100	1	2
120 million - answers	Value	correct	AC
1 cruise ship	1.000.000.000	120.000.000	-3
120 houses in Kaskisaari	360.000.000	120.000.000	0
8 metros	16.000.000.000	120.000.000	2
40h	144.000	1.123.200	0
53 years	1.671.620.000	1.123.200	3
1.5 years	47.310.000	1.123.200	1
distance to the bottom of the pacific ocean	4.000	120	1
Helsinki to Lahti	100.000	120	2
120 m	120	120	0
350 billion - answers	Value	correct	AC
1 us budget deficit OR 350 cruiseships	350.000.000.000	350.000.000.000	0
health care for 350 million people in the U.S.	1.400.000.000.000	350.000.000.000	0
1 us budget deficit	350.000.000.000	350.000.000.000	0
100 days	8.640.000	3.469.400.000	-2
111 years	3.500.940.000	3.469.400.000	0
100 years	3.154.000.000	3.469.400.000	0
circumference of earth	40.075.017.000	350.000	3
Helsinki to Riga	346.000	350.000	0
350 km	350.000	350.000	0

Benchmark-test answers *cont.*

700 billion - answers	Value	correct	AC
2 us budget deficits	700.000.000.000	700.000.000.000	0
sending 10 satellites into space	5.000.000.000	700.000.000.000	-2
1 iPhone X for every citizen of helsinki	1.000.000.000	700.000.000.000	-2
200 days	17.280.000	6.970.340.000	-2
222 years	7.001.880.000	6.970.340.000	0
200 years	6.308.000.000	6.970.340.000	0
1 light year	9.500.000.000.000.000	700.000	3
Helsinki to Lithuania	600.000	700.000	0
700 km	700.000	700.000	0

Stacking Bills						
Observation	Complex.	Direct/ Unexpected	Correct.	Category	Time	
oh no, Helsinki to Vantaa is only 11 billion €	2	d	y	personal; reading; revelation	03:50	
in terms of distance that's a shitload of money (helsinki to sweden)	1	d	-	summarizing		
from helsinki to India – that's a shitton of money	1	d	-	personal; summarizing		
1m is 1million. Ah. That's a cool fun-fact.	2	d	y	reading	00:50	
I think about bank robbery now – you could have 10 cm and it would already be ... A lot of money. And it's easy to carry.	1	u	-	hypotheses; interpreting		
346 km – wow that's a lot	1	d	y	summarizing		
Oh – to Iran it would be 3,360 billion	2	d	y	personal; reading		
that's really a lot	1	d	-	summarizing		
I didn't know that billion would have 9 zeros. I thought it would be more zeros.	1	u	y	revelation		
1m of bills is a lot, in both money and size	1	d	y	summarizing	00:50	
So i've flown that distance many times, that's familiar to me.	1	d	y	personal		
that's a huge pile of bills.	1	d	y	summarizing		
2 billion euros to my home	2	d	y	personal; reading		
so that's almost 7 billion (Helsinki to Otaniemi)	2	d	y	personal; reading		
that's a ton of money	1	d	y	summarizing		
So it gets less when i zoom in – would i even be able to get it back to million?	3	d	y	inferring		
yeah, so from my house to the super market is 100 million	2	d	y	personal; reading		

Money Clock					
Observation	Complex.	Direct/ Unexpected	Correct.	Category	Time
that's a lot of years and days and hours and seconds to go	1	d	y	summarizing	00:51
okay so 200,000,000 – ah, only 23 days	2	d	y	reading	
oh, that's a 100 euros per second. That's quite a lot	2	u	y	revelation	
i have to find a job that pays a 100€/second	1	u	-	personal, interpreting	
how long would it take me to pay it off with my income rate, i wonder – probably a lot	1	d	-	personal	
i put in 1 million – 2 hours, okay	2	d	y	reading	
who would even be able to pay a rate of 100€ per second? That's impossible.	1	u	-	hypotheses	04:11
I changed the amount of years to 20, but I can't even read the number that came. It's so large.	1	d	-	summarizing	
So i put in 99 years , the years i hope to live, and ... so i have a huge debt already	1	d	n	personal; interpreting	03:20
if i wanna pay it in 99 days, it would take me 855 thousand (actually 855 million), i would have less than a million	2	d	n	reading; interpreting	
so if i could save a hundred a second i wouldn't be a millionaire in 99 days	3	d	n	personal; interpreting	

Buying Power						
Observation	Complex.	Direct/ Unexpected	Correct.	Category	Time	
okay, i could get a lot of sail boats with this	1	d	y	summarizing	01:21	
wow cruise ships are fucking expensive	1	u	-	summarizing		
it would be interesting to see different countries	1	d	-	hypotheses		
or compared to the gdp, because it's relative to the country size	3	d	-	hypotheses		
I would like to know how many cruise ships you could get for a passenger jet	3	d	-	inferring		
this tells me that a cruise ship is 5 times more expensive than a passenger jet	3	u	n	inferring		
I guess i have to scroll a lot till one ends and the other one goes on for 10 more times	3	d	y	inferring		
so it's like 346 million i phone x	2	d	y	reading	00:55	
that's more than the population of finland (the amount of average finnish yearly income)	3	u	y	interpreting		
that could keep the economy in finland going for one or two years maybe	3	u	-	hypotheses		
this is a little easier to visualize (the passenger jets)	1	d	-	personal		
i thought passenger jets might be more expensive than cruise ships, but obviously not	3	u	y	revelation		
okay, that's an insane number	1	d	-	summarizing		
so these many people (1.5 million) can pay for the (budget deficit)	2	d	y	interpreting		
in the us it could be a long way to understand extreme poverty there	3	u	-	hypotheses		
They (passenger jets) are so much more expensive than a single family house	1	u	y	revelation	03:12	
I could eat bisquits for my whole life (entered digestive biquit price 2€)	3	d	y	personal; interpreting		
you could travel around the world for many times (with the amount of flights to Taiwan)	3	d	y	personal; interpreting		

Fonts: Nitti, Nitti Grotesk
Print: Picaset Oy, Helsinki

