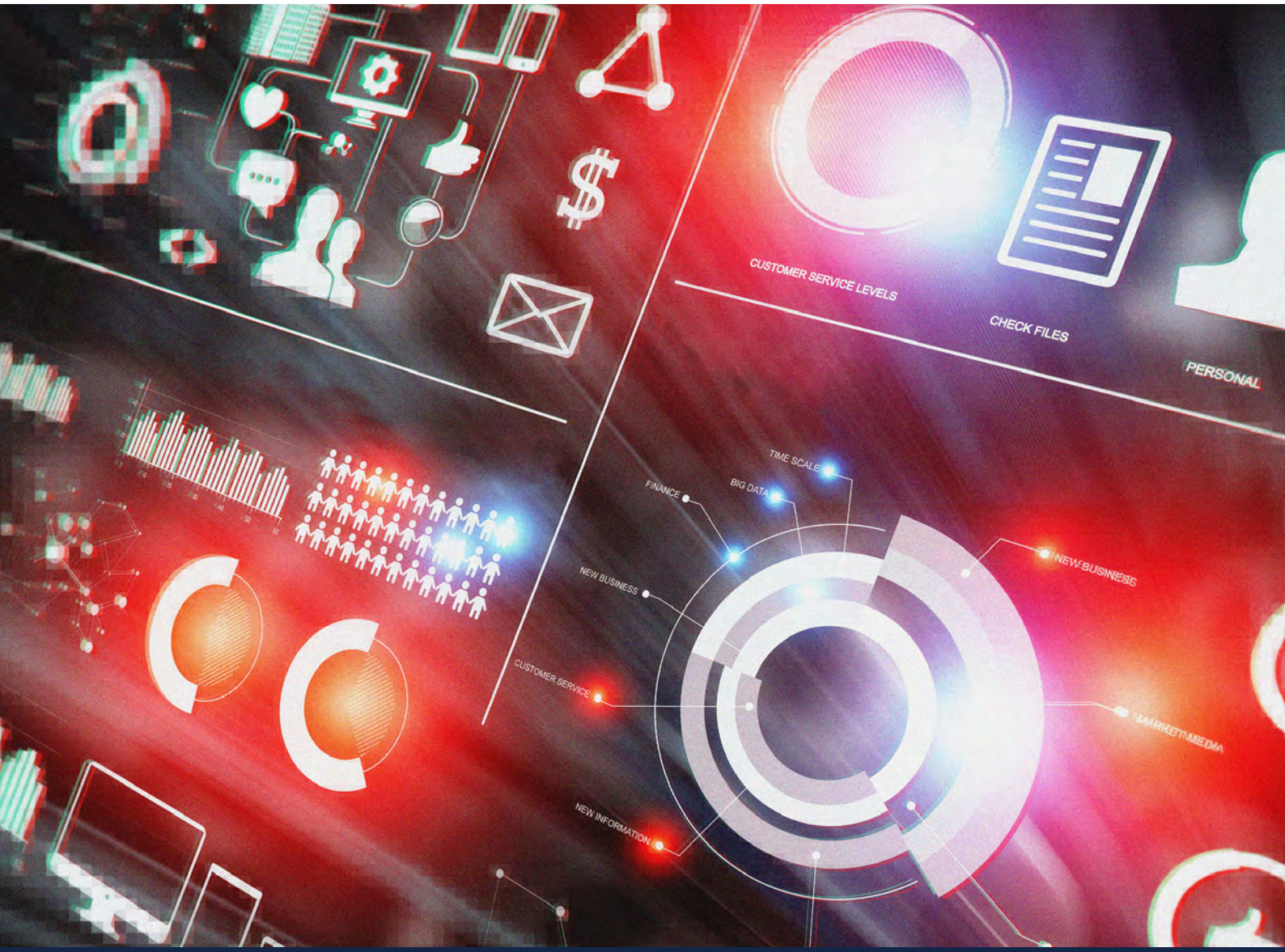
UNIVERSITY OF OXFORD
REUTERS INSTITUTE for the STUDY of JOURNALISM

REPORT

Making Research Useful: Current Challenges and Good Practices in Data Visualisation

Malu A. C. Gatto

May 2015



Cover image: Messenger Marketing

Published by the Reuters Institute for the Study of Journalism with the support of the University of Oxford's ESRC Impact Acceleration Account in partnership with Nesta and the Alliance for Useful Evidence

Produced in partnership with

Nesta

Alliance for Useful Evidence

University of Oxford

DPIR Department of Politics & International Relations

Foreword

This publication results from a project on how to better visualise academic research conducted by the Reuters Institute in a partnership with Nesta and the Alliance for Useful Evidence. The project was funded by the University of Oxford's ESRC Impact Acceleration Account, and supported by the Knowledge Exchange office of the Department of Politics and International Relations.

The project explored the demand for visualisation and what is necessary to improve the communication of research results and academic knowledge in visual form and to make information available to the public, policy-makers, media, businesses, and other stakeholders.

The project brought together policy-makers, journalists, and academics to identify needs for visualisation, gaps in its provision, good practice in creating visualisations, and challenges in providing effective visualisations.

I am particularly grateful for the assistance and support provided throughout the process by Liz Greenhalgh, Knowledge Exchange officer in the Department of Politics and International Relations, University of Oxford; the support of Jonathan Breckon, head of the Alliance for Useful Evidence; and the work of Malu A. C. Gatto, a DPhil student at St Antony's College, University of Oxford, who wrote this report and conducted research throughout the project.

This publication is intended to provide readers with an overview of the concept and issues of visualisation and inspire readers to consider options that may help develop their uses of visualisation in conveying complex information.

Robert G. Picard
Project Principal Investigator
Reuters Institute
March 2015

Contents

Executive Summary

This report is the culmination of the exploratory phase of a project conducted by the Reuters Institute for the Study of Journalism, the Department of Politics and International Relations at the University of Oxford, and the Alliance for Useful Evidence. It summarises and further develops the discussions brought up during three workshops carried out between June and November 2014.

Current developments suggest that the practice of relying on data for decision-making is here to stay: governments, the media, organisations, and businesses are becoming more reliant on it, and consumers are now also starting to demand products and services that are based on data analysis. Effective communication of data through data visualisation is becoming more necessary.

Despite this trend in other sectors, academics have often struggled to share their data with other actors and to disseminate their research findings to broader audiences. We suggest that data visualisation is an excellent tool to advance research, initiate communication with actors from other sectors, as well as promote dissemination and increase the impact of research findings.

Knowledge gaps, the variety of types and forms of data used, the availability, training, and limitations of specific software typically learned and used by social scientists, as well as expectations traditional to academia have often prevented academics from engaging with and promoting their research to leaders from other sectors. Based on knowledge derived from our workshops, we share suggestions of good practice in producing and using data visualisation in academia, journalism, public policy, and business. These practices include: preparing visuals for specific audiences, identifying your 'story' and the appropriate chart to use, displaying relationships that the brain can process more quickly, and uncluttering so as to not detract from the main 'story'.

Recognising that professionals from different fields have varying levels of training in data visualisation, we present four potential pathways for producing good-quality visualisations that have been tested by academics, government institutions, organisations, or firms (many of which were represented in our workshops). The suggested pathways are: (1) do it yourself; (2) collaborate with cross-sector or cross-departmental professionals; (3) seek the help of IT professionals; and (4) provide the data to others for them to produce the visualisation.

Becoming more knowledgeable about good practices in data visualisation, and seeking to enhance visualisation capabilities through formal training and collaboration, could prove beneficial for social scientists' academic work and its impact beyond academia.

1. Introduction

'Data Visualisation' refers to the visual representation of statistical and other types of numeric and non-numeric data through the use of static or interactive pictures and graphics. So data visualisation does not replace narrative, but is often used in combination with it to improve understanding. Whether accompanied by text or not, the main goal of visualisation is to elucidate patterns, gaps, schemes, regularities, and connections that may not be easily identified by rapidly reading raw data or long texts. Data visualisation improves the understanding of data for experienced researchers, statisticians, and academics, as well as for a much broader non-specialist audience.

In sum, data visualisation reduces knowledge gaps – especially quantitative skills gaps. Verbal and numerical data are currently being produced at unprecedented rates and used in a variety of ways: it is estimated that over 90% of existing data has been produced since 2010 (*Science Daily*, 2013). While up to 2003, five exabytes of data had been produced *in total*, in 2013 the same amount of data was produced *every day* (Gunelius, 2014).

This 'data explosion' may be more easily grasped when analysing internet and social media usage. In 2012, Google received 2 million search queries per minute. By 2014, this number had *doubled*. Data produced by intelligence company DOMO, and cited by Gunelius, show that in every minute of 2014, roughly 2.5 million pieces of content were shared by Facebook users, 277,000 tweets were made, 216,000 new photos were posted on Instagram, 72 hours of video were uploaded to YouTube, 61,141 hours of music were played on Pandora, 48,000 apps were downloaded from the Apple store, and over 204 million emails were sent (Gunelius, 2014).

Although, as a society, we have become increasingly exposed to and reliant on data, the meaning and understanding of the numbers we see still often remain lost in columns and rows. Data visualisation allows information consumers to digest large amounts of data more easily by organising data in a way that highlights relationships, patterns, or gaps. As knowledge becomes increasingly reliant on large amounts of data, visualisation becomes not only a practical tool in improving understanding, but also an essential one. Furthermore, people are becoming more and more acquainted with visual displays of data. While production of and exposure to graphics has previously been limited to academics, scientists, and statisticians, most people are now exposed to at least some type of data visualisation on a daily basis: charts are now constantly used in TV shows (think weather reporting and election coverage, for example) and online and print media (see Appendix B for more examples). Children are also consuming this, directly (in school or through infographic books targeted at them: Rogers, 2014) and indirectly (by watching TV and using the internet). As such, familiarity with representations of data is expected to increase. Despite this, the norms guiding the appropriate and effective uses of data visualisation remain unclear.

This report seeks to be a guide for good practice and possible pathways to producing good-quality data visuals. It hopes to do this by shedding light on existing debates on the applicability of data visualisation in the academic, government, policy, media, and business spheres. In doing so, it also urges academics to consider data visualisation as a research *and* dissemination tool. It also advocates collaborative work between academic and non-academic spheres to increase the visibility (and impact) of academic research.

To gain further understanding of how academic, government, policy, media, and business leaders currently use data visualisation, the Reuters Institute for the Study of Journalism and the Department of Politics and IR of the University of Oxford, together with the Alliance for Useful Evidence (a partnership between Nesta, the Big Lottery Fund, and the Economic and Social Research Council) carried out the project with the same title as this report 'Making Research Useful: Current Challenges and Good Practices in Data Visualisation'. The project, funded by an ESRC Impact Acceleration Award, and led by Professor Robert G. Picard at the Reuters Institute, was structured around three workshops that brought together data visualisation experts from the media, policy, and business spheres, as well as a number of academics, and other individuals eager to learn more about the theory and practice of data visualisation.

120 people signed up for our London-based workshop on 20 June 2014, and a total of 57 people for our 17 October and 14 November Oxford-based sessions.

This report is a summary of what we learned through workshops and research, but it is only the first step in what we hope to be a longer process of learning and advocating for the use of data visualisation.

The report is structured as follows. First, we convey the benefits of employing data visualisation as a research and a dissemination tool; secondly, we introduce the current challenges to the practice of data visualisation, especially as it pertains to communication across different sectors; thirdly, we elucidate good current practices of data visualisation applied by social scientists, journalists, public policy-makers, and business leaders. In the fourth section we cover potential pathways to producing good-quality data visualisation. Finally, we identify some of the remaining challenges for academics in engaging with data visualisation, and conclude by presenting some of the ways in which the University of Oxford and the Alliance for Useful Evidence will continue to promote data visualisation as an effective tool for collaboration, dissemination, and research.[1]

[1] Also note that a full list of workshop presenters can be found in the Appendices, where we also offer further explanation of types of data visualisation and provide suggestions for relevant further readings, useful software and platforms, and free online courses.

2. Why Use Data Visualisation?

The rapid expansion in the amount of data produced daily has led to related relevant developments in a number of areas, including:

(1) accessibility – the abundance of available data requires more accessible ways of absorbing and processing information;
(2) literacy – individuals' constant exposure to design through different forms of media has allowed society, and especially younger generations, to be more aware and knowledgeable about interpreting and producing visualisations of information;
(3) quick analysis and decision-making – companies, governments, and businesses are now employing data to understand policy outcomes and client preferences (McCandless, 2010).

All of these parallel and linked developments promote the use of data visualisation.

Visualisation can improve understanding and has the potential to increase the uses of research evidence. According to the definition provided in 1987 at the National Science Foundation's Visualisation in Scientific Computing Workshop report: '[v]isualization offers a method for seeing the unseen. It enriches the process of scientific discovery and fosters profound and unexpected insights' (quoted in Hansen and Johnson, 2005: p. xiv). This is because 'all people with normal perceptual abilities are predominantly visual' (Few, 2014b), so visualising data improves our collective understanding of research findings. Furthermore, graphics summarise data into manageable formats, thus allowing our brains to process comparisons, patterns, and differences much more rapidly (Koch et al., 2006).

Availability of Data, Increase in Demand for Visualisations

Evidence has always been the basis for academic research, and, given the amount and availability of fresh data across industries, it has become crucial for other sectors as well. The Alliance for Useful Evidence is proactive in promoting evidence as the basis for better social policy. The UK government is also beginning to follow this trend. Sir Jeremy Heywood, Cabinet Secretary and Head of the Civil Service, described the What Works Network, an initiative launched in March 2013 that promoted the use of evidence and research to inform policy as 'world-leading' (Heywood, 2014). The Network is a platform that brings together existing evidence on the outcomes of government projects/services, as a means of identifying whether public spending and policy are effective. As such, the Network seeks not only to check on the efficiency of existing public policy, but also to inform future initiatives.[2] The need for more evidence-based decision-making was also one

[2] For more information: <https://www.gov.uk/what-works-network>.

of the key conclusions of 'Challenges of Government: Flourishing Cities', a conference recently held at Oxford's Blavatnik School of Government, where CNRS Director of Research, Patrick Le Galès, called for greater collaboration between academics and policy-makers, as a means of improving public policy (Lindoso, 2014). For decisions to be made based on evidence, however, datasets have to be made available. This is precisely what a number of governments, organisations, and businesses are doing: publicising their data in readily accessible formats. For instance, the US government's open data project has made roughly 140,000 datasets public to date (Data.gov, 2014). Efforts are still limited, however. Large numbers of data produced by governments and companies are still inaccessible, as is some data collected by academics.

Nonetheless, the focus on evidence-based decision-making has permeated both the public and private labour markets, increasing the demand for more data scientists. Evidence-based news and data visuals are increasingly expected by media consumers. As a result, journalists are currently encouraged to use (and make available) hard data to back up their work, and to provide accompanying visuals that summarise their main findings.[3] In the business sector, however, evidence-based decision-making is still the exception - although the trend is beginning to permeate. In fact, recent findings suggest that companies that rely on evidence-based decision-making positively benefit and are, on average, 8% more productive (Bakhshi et al., 2014).

The momentum in evidence-based decision-making is also increasing the demand for related jobs. A recent article published by the World Economic Forum shows that over 61,000 new jobs posted on LinkedIn require data collection, analysis, and presentation skills (Patil, 2014) and a recent McKinsey report estimated that, by 2018, the United States alone could have a shortage of up to 190,000 data scientists and '1.5 million managers and analysts with the know-how to use the analysis of big data to make effective decisions' (Manyika et al., 2011). The Harvard Business School Blog (Patil, 2013) has deemed 'data scientist' the 'sexiest profession alive' and reports from McKinsey (Manyika et al., 2011), NPR (Noguchi, 2011), and articles from a number of mainstream media outlets, such as the *New York Times* (Hardy, 2012), have also claimed data scientist to be the most in-demand job.

Some, however, are sceptical of these rapid developments. There is significant reluctance to accept data visualisation as a reliable tool for analysis and decision-making. For the *Guardian*, John Burn-Murdoch writes: 'Data presented in any medium is a powerful tool and must be used responsibly, but it is when information is expressed visually that the risks are highest' (Burn-Murdoch, 2013). Precisely because we can process visual information

[3] This was brought up multiple times by media representatives attending our workshops and has also been the topic of a 2012 event hosted by Nesta titled 'The Rise of Data Journalism'. For more information: <http://www.nesta.org.uk/event/rise-data-journalism>.

more quickly, we are also more likely to be influenced (and fooled) by visuals. In this way, something as simple as a change in colour schemes or the inclusion of one type of image (e.g. icon, emoticon, picture) over another, may impact our understanding and interpretation of a visual. The rise in data and visualisation literacy could suggest that, in the near future, people might not fall so easily for 'bad' or 'deceitful' data visuals. Burn-Murdoch is not so optimistic, however: he suggests that one of the reasons why we continue to be more critical of text than of data visuals is because we are generally taught to critique text and be exposed to text in different stages of writing, but rarely see the progression of visuals apart from in a more polished form as a representation of final results and findings (2013). To overcome this issue, literacy in data analysis and visualisation ought to be developed not only through increased exposure, but through proper training.

Despite scepticism from some, current developments suggest that the practice of increasingly relying on data is here to stay: businesses, organisations, governments, and the media are becoming more and more reliant on them and consumers are now also demanding products that are based on data analysis. With this, the effective communication of data through data visualisation becomes crucial.

Visualisations Improve Understanding

It is not only the increasingly rapid production of data that is pushing for visualisation to be more commonly used, but also the understanding that visualising data increases perceptive and cognitive understanding of both quantitative *and* verbal information. As Stephen Few explains:

> *Data visualization is effective because it shifts the balance between perception and cognition to take fuller advantage of the brain's abilities. Seeing (i.e. visual perception), which is handled by the visual cortex located in the rear of the brain, is extremely fast and efficient. We see immediately, with little effort. Thinking (i.e. cognition), which is handled primarily by the cerebral cortex in the front of the brain, is much slower and less efficient. Traditional data sensemaking and presentation methods require conscious thinking for almost all of the work. Data visualization shifts the balance toward greater use of visual perception, taking advantage of our powerful eyes whenever possible.* (Few, 2014a)

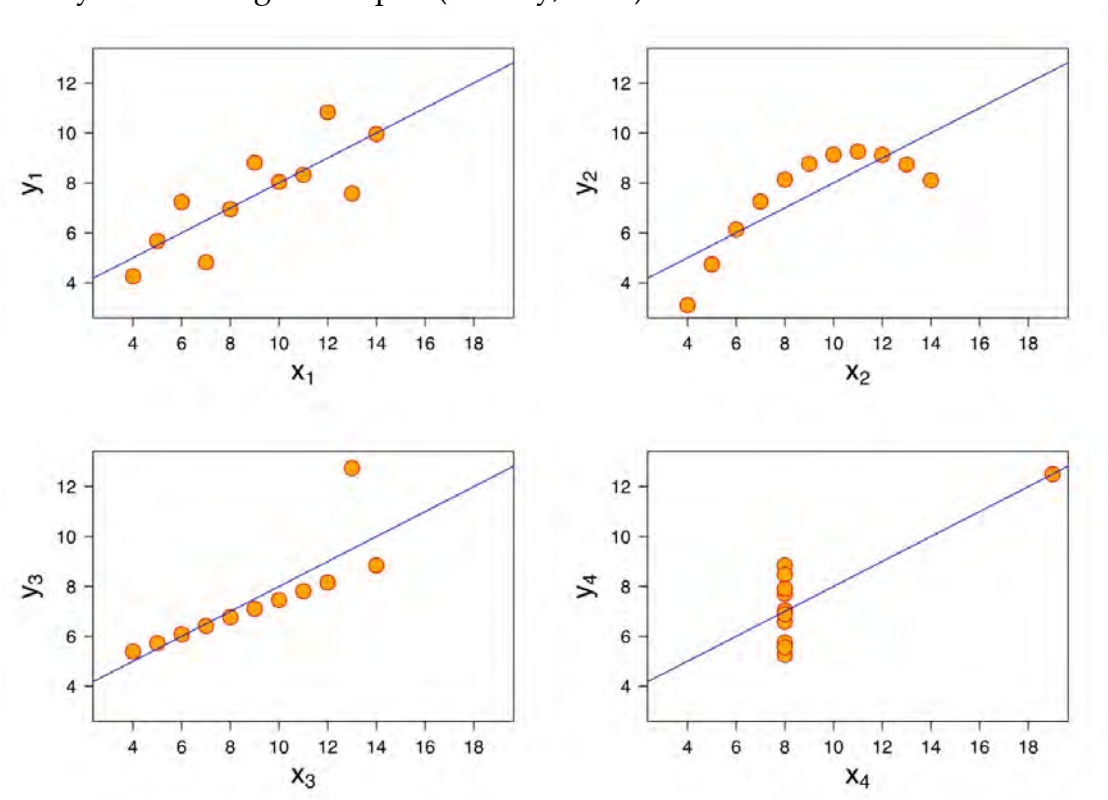
In sum, data visualisation seeks to simplify the representation of hypotheses, theories, or stories, which are often not clear from raw data (i.e. a quick look at a spreadsheet or a table may give someone the wrong impression about general data trends/patterns, while visuals easily overcome this problem; see Anscombe's quartet in the next section). This means that visuals allow for simple and powerful communication of data, while also serving as a tool for research development.

Academics Can Benefit

There are many reasons as to why academics should adopt data visualisation. Min Chen, Professor of Scientific Visualisation at the University of Oxford, identifies four main uses for data visualisation: (1) invention; (2) analytics; (3) dissemination; and (4) observation.[4] For academics, these four uses of data visualisation can be understood as belonging to two broader categories: that of research, theory building and analysis, and that of dissemination and impact.

Improving Research Practice: Theory and Analysis

Although the emphasis on data visualisation as an important tool of *dissemination* is gaining momentum now, academics have long demonstrated the importance of graphics to the process of *research*. In 1973, statistician Francis Anscombe presented the 'Anscombe's quartet', a set of four datasets with nearly identical values (often to a second decimal point) in mean of x and y, variance of x and y, correlation, and linear regression. Despite having the same values when displayed in a table of descriptive statistics, Anscombe showed that the datasets looked completely different when plotted (see figure 1). Anscombe used this example to support the notion that data visualisation should not only be used at the end of research projects (to illustrate research findings), but also at the initial stage of data exploration and analysis, when theory is still being developed (Iliinsky, 2012).



Source: <http://blog.visual.ly/why-is-data-visualization-so-hot>.

Figure 1.

[4] For Professor Min Chen's presentations and articles on the matter, see: <http://www.oerc.ox.ac.uk/people/min-chen>.

Chen agrees. He characterises data visualisation as a tool that enhances the process of hypothesis-testing, model-fit, and appropriateness of methods and systems applied, therefore serving at an analytical level. He also believes data visualisation is important at the 'inventive level', where visualisation becomes a developmental aid to assist in theory-building and refining.[5] The notion that data visualisation should be used for research is not exclusive to academia. Stephen Few, founder of Perceptual Edge, a firm that provides data visualisation consultancy to businesses and other types of organisations, argues that data visualisation can aid in 'discovery', which he defines as accomplished through a process of exploration that searches 'for significant facts' (Few, 2014b).

During one of our workshops Félix Krawatzek, DPhil student in Politics at the University of Oxford, shared how data visualisation has been a methodological and analytical tool in his research; while investigating issues pertaining to discourse network analysis, he found that visualisation allows for a clearer presentation of his data and a more instinctive analytical platform. Although visualisation has the potential of enhancing understanding of data patterns or gaps and enhancing the 'analysis stage' of research for all types of data and theory, the tool may be particularly useful for some types of data and theory, namely those that involve complex relationships between and across variables and observations.

Beyond being a tool for better understanding data patterns/gaps, data visualisation can benefit the research stage in a number of other ways. For instance, by illustrating expected causal relationships and their mechanisms, mind-maps and flow-charts can be helpful for theory-refining. Recently, Peter Kraker, researcher in Social Computing at the Know-Center, Graz University of Technology, showed that data visualisation can also assist academics in better identifying contributions and gaps in the content and methodologies of bodies of literature. In his 2015 work, Kraker developed source code that allows users to produce visualisations of given research fields using data from the online reference system Mendeley (Kraker, 2015).

Although traditionally data visualisation in academia has been used to summarise research findings, visualisations can also be helpful in the process of developing research: be it by identifying gaps in the literature that need further studying, by clarifying theories and hypotheses, or by providing preliminary understanding of data.

[5] For Professor Min Chen's presentations and articles on the matter, see: < http://www.oerc.ox.ac.uk/people/min-chen>.

Robert G. Picard, Director of Research[6] at the Reuters Institute for the Study of Journalism, suggests that data visualisation is useful in 'increasing public understanding of academic research' – and as academia becomes more concerned with research impact,[7] data visualisation can be a helpful tool in disseminating research and achieving this goal.

In a paper recently presented by Ramon Bauer and Nikola Sander, scholars of Demography and Human Capital and Geography and Regional Research, respectively, the authors contend that academic ideas and complex findings can be translated to jargon-free language in data visualisations, in a way that communicates research to a broader audience without compromising or misrepresenting scientific findings (and 'keeping control over how science is portrayed'). In this context, Bauer and Sander also point to data visualisation as an effective tool in promoting the dialogue between academics, policy-makers, and the wider public – creating 'mutual benefits for science and the public' (Sander and Bauer, 2015). This is particularly important as research shows that stakeholders in the public and corporate sectors often have the perception that university curricula are 'too theoretical', thus not properly training students for the type of data analysis required to solve problems in the 'real world' (Bakhshi et al., 2014: 29).

There are a number of ways data visualisation can support research communication and engagement. One of them is to employ visuals in a number of different academic and non-academic settings, where interaction with an audience is important. Opportunities to use data visuals could be discipline-specific or broader academic conferences; invited lectures; personal academic websites and/or blogs; social media outlets such as Twitter, Facebook, LinkedIn, and Academia.edu; and media outlets (university blogs and/or non-academic media). These opportunities are only expected to increase as online and open sources become more commonly used means of seeking impact and engagement (Anselmo, 2015).

Dr Ruth Dixon's experience of using charts on social media to share her work (see Box 1 on page 13) illustrates how the tool benefits both academics and research consumers.

Yet one more way to support research impact is by using data visualisation for advocacy. This is precisely what Andrew Steele, bioinformatician and creator of Scienceogram, does. At one of the workshops he explained how Scienceogram provides visual insight into government spending as it relates to investment in science (Scienceogram.org). Noah Iliinsky contends that visuals are more effective than textual information in

[6] Robert G. Picard was Director of Research at the Reuters Institute at the time of this project but stepped down from that role in December 2014.

[7] The UK's Research Excellence Framework (REF) broadly defines 'impact' by assessing how academic research conducted has 'an effect on, change or benefit to the economy, society, public policy, culture and the quality of life, beyond academia' (Hill, 2014).

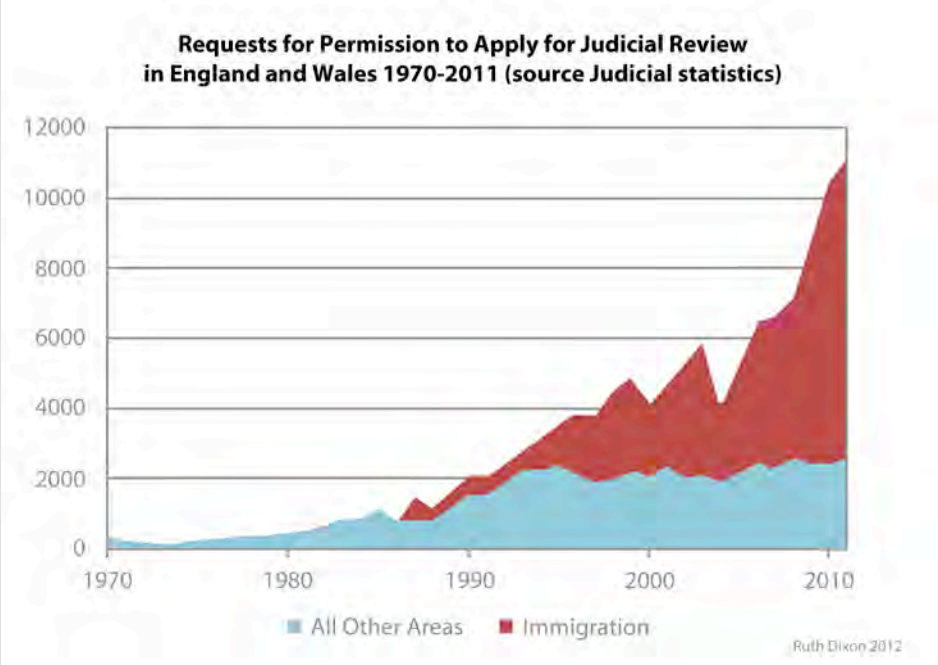
'changing people's minds' and in reducing 'incorrect beliefs among potentially resistant subjects' (Iliinsky, 2012). For providing access to 'actionable insight' data visualisation is not only an important communication and research tool, but also a valuable instrument of advocacy (ibid.).

Ultimately, data visualisation should be used by academics because it enhances the understanding of information for both research developers and research consumers. We suggest that data visualisation is an excellent tool to initiate communication with other actors, and to promote dissemination of research findings. Figure 2 on page 14 uses a visual to summarise the discussions of this report so far. Nonetheless, for various reasons detailed below, academics have often struggled to share their data with other actors, and to disseminate their research findings to broader audiences.

Box 1. Data Visualisation and Research Impact: An Example

Dr Ruth Dixon recounts how data visualisation increased the exposure of her research with Professor Christopher Hood (Department of Politics and IR, University of Oxford)

On 19 November 2012 I heard on the BBC *Today Programme* that David Cameron had said that applications for judicial review (JR) had tripled since the 1990s, holding up important projects and slowing economic growth. I have looked at the number of JR applications for my work on 'Reshaping Executive Government' with Professor Christopher Hood (funded by the Leverhulme Trust). I knew that the increase since 1995 was almost entirely due to immigration and asylum cases (and not topics such as infrastructure projects). However, no one on the *Today Programme* made that point, nor did any others who commented on the story on Twitter. So later that morning, I tweeted our graph (sadly, the original on Twitpic is no longer available) and sent the link to a couple of people who had commented on the story.



A few hours later I had a phone call from the Guardian Data Blog, asking if they could put the data on their site. We had assembled the data from the publicly available Judicial Statistics reports so we were happy to share this dataset. I linked our spread sheet to our project webpage, from which it was copied to the Data Blog. A week or so later I also had a phone call from the R4 *Today Programme* to ask about the data – information that they then used in interviews with Lord Woolf and Chris Grayling on 15 December.

Interest in these statistics has continued. In January 2013, Maurice Sunkin and Varda Bondy of the Public Law Project wrote an article on JR statistics and success rates, which also linked to our graph. In 2014, we were asked by an Oxford professor for permission to use an updated version of the graph in the forthcoming edition of his textbook on administrative law.

Source of image from Box 1: Ruth Dixon, 2012.

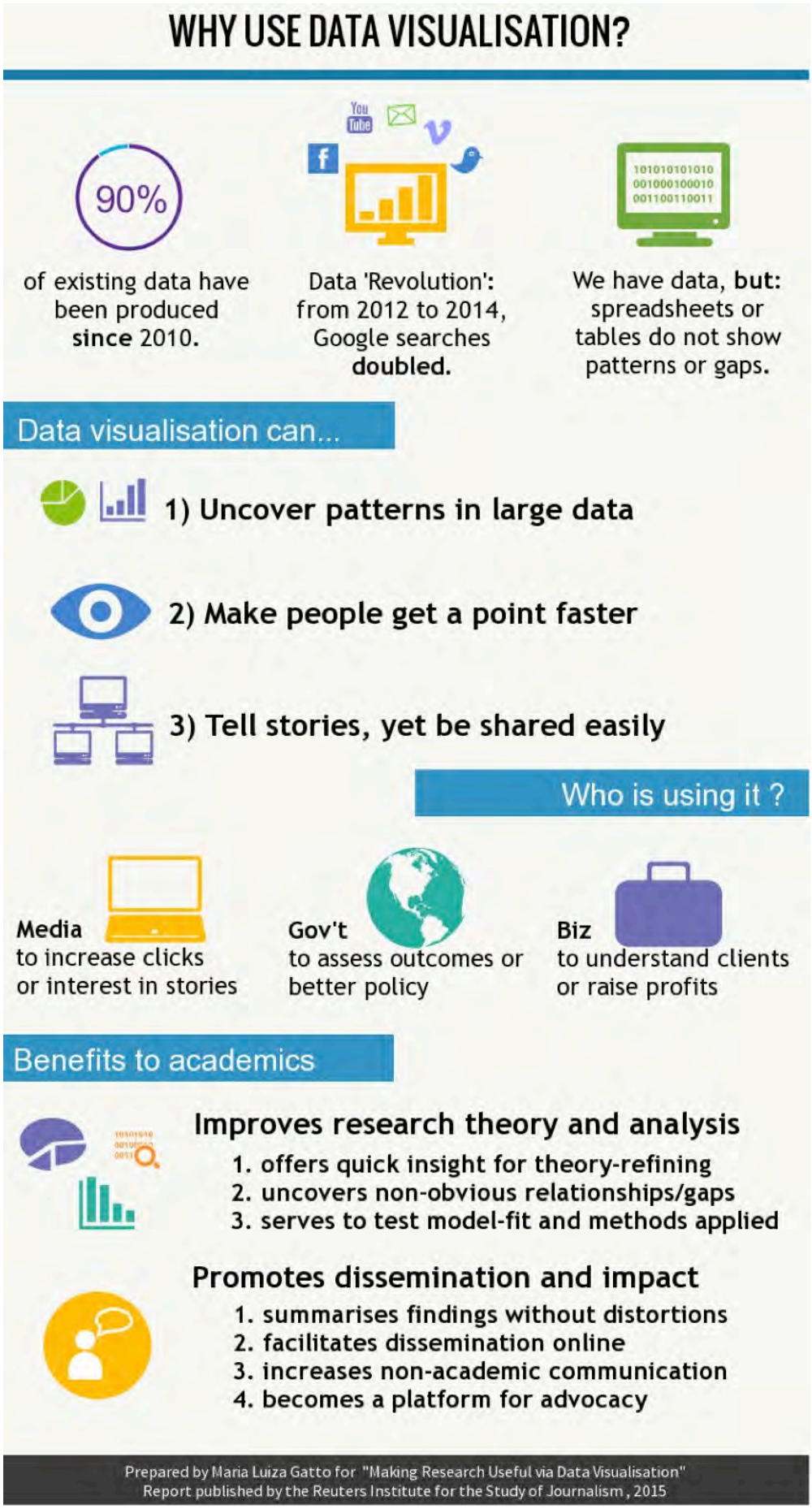

WHY USE DATA VISUALISATION?
90%
of existing data have been produced **since** 2010.
Data 'Revolution': from 2012 to 2014, Google searches **doubled.**
We have data, **but:** spreadsheets or tables do not show patterns or gaps.
Data visualisation can...
1) Uncover patterns in large data
2) Make people get a point faster
3) Tell stories, yet be shared easily
Who is using it ?
Media to increase clicks or interest in stories
Gov't to assess outcomes or better policy
Biz to understand clients or raise profits
Benefits to academics
Improves research theory and analysis
1. offers quick insight for theory-refining
2. uncovers non-obvious relationships/gaps
3. serves to test model-fit and methods applied
Promotes dissemination and impact
1. summarises findings without distortions
2. facilitates dissemination online
3. increases non-academic communication
4. becomes a platform for advocacy
Prepared by Maria Luiza Gatto for "Making Research Useful via Data Visualisation"
Report published by the Reuters Institute for the Study of Journalism , 2015

Figure 2.

3. Social Scientists and Current Challenges to Data Visualisation

This section is largely based on discussions that took place during our workshops and refers to points raised by participants in regards to current challenges social scientists may face in using data visualisation, as well as in sharing research findings and collaborating with journalists, policy-makers, and business leaders. Data visualisations produced by social scientists have the potential of being conducive to applications in policy, media, and non-profit sectors. Understanding the challenges that social scientists currently face in adopting data visualisation as a research and dissemination tool is thus particularly relevant.

Knowledge Gap

One aspect that may prevent academics from using data visualisation is lack of knowledge about software and platforms that produce visuals. Most scholars do not receive formal training in data visualisation and therefore rely on statistical packages with limited visualisation capabilities, such as SPSS or STATA, or spreadsheets and word processing programmes to develop visualisations. This often means that researchers are limited by the little or no training they receive, as well as by what the software to which they have access allows them to do. As a result, many social scientists continue to rely on regression and correlation tables, or simple bar and plot charts to display quantitative data, and word clouds, or no visualisation at all, to summarise qualitative data. This (re)production of ready-made graphics, however, while often appropriate for journals, does not necessarily enhance quick audience understanding of data in presentations or posters, and may not maximise the potential gains of theory-refining. Furthermore, the production of data visualisation in specific statistical packages may limit the scope of communication between academics and other stakeholders (e.g. graphics produced for a publication may not synthesise the information that is most important for public policy-makers, or may not be as interestingly displayed for journalistic use).

At this stage, there also seems to be a significant issue of self-selection in the field that is preventing women from being exposed to (and gaining knowledge about) data visualisation at the same rate as men. Although data on this are scarce, it is estimated that only 23.1% of data visualisation practitioners are women. This is consistent with the proportion of women presenting at data visualisation conferences, which could potentially be reinforcing the low entrance rates of women into the field (Stefaner, 2013). As the current report shows, data visualisation is increasingly important as a tool of research and dissemination; as such, gender-based knowledge gaps may lead to even further obstacles for women in academia, especially as it relates to the impact of their work (Campbell, 2014).

Types and Format of Data

Many scholars find it particularly challenging to visualise textual data, as well as relational data – which partly derives from a lack of knowledge of data visualisation, and partly from the limiting tabular format in which social scientists continue to display their data.[8] The range and variety of data types used in social scientific research was one of the points raised as creating potential barriers for data visualisation. Chen describes the following as the main types of data used in the social sciences:

- textual,
- network,
- tabular (the most popular),
- software,
- volume,
- vector,
- tensor field, and
- geo-information.

One of the challenges that results from the variety of data types is that academics often produce datasets in formats that can only be read by specific software, or that are structured in ways that prevent use by other researchers or stakeholders. Another issue that arises is that the huge variation in data types means that academics need to learn to produce visualisations that can more appropriately illustrate specific types of data.

Software Diversity

Department, generation, university, data type, and personal preferences are just some of the factors that may influence academics' choice of data analysis software. The most common include SPSS, STATA, MatLab, R, and, NVivo, all of which have different graphics packages. For instance, R and its 'add-ons' are known for their potential in illustrating network and other complex relationships, while NVivo is recognised for its capacity to produce graphics that display relationships between non-numerical data. More importantly, however, the fact that most academics rely on only one form of data analysis software means that the graphics package available in that particular software represents the array of data visualisation possibilities for a given researcher.

The training in and use of specific statistical packages also creates an extra obstacle for communication between academics and public policy practitioners. While universities may provide access to a range of paid software, most media outlets, think tanks, government branches, and non-governmental organisations will not. Something as simple as not having access to a specific type of software therefore hinders the possibility of non-

[8] For more on this, see Professor Min Chen's presentations and articles: <http://www.oerc.ox.ac.uk/people/min-chen>.

academics replicating findings for non-academic audiences and engaging with academic research in other ways.

Journal Expectations and Discouragement in Sharing New Data

Academic journals in the social sciences currently discourage the use of visualisation; given that most journals are published in print, visualisation is still difficult to handle in black and white, and costly to reproduce. This is likely to change as online publication becomes more popular and as more journals start giving researchers the possibility of publishing additional supporting materials online. Nonetheless, so far academics' use of data visualisation often does not fit journal standards.

Another current challenge academics face is that of 'timing': journalists and policy-makers have to produce material with a quick turnaround; academics, however, work on datasets for a long period of time, generally only making them public after using them for several publications.

Most of these challenges boil down to one simple truth: academic training and scientific production have often been incompatible with non-academic demands. We hope that a greater understanding of the potential uses of data visualisation may help to overcome this incompatibility.

4. Overcoming Current Challenges: Good Practice in Using Data Visualisation

Precisely because data visualisation only presents snapshots of larger research projects, it often raises scepticism about the reliability of findings. For many, visuals have been used deceptively (e.g. to convince consumers of a point that is not representative of the data). That could be achieved by either structuring or coding data in a way that is not clear to audiences; by applying a methodology that (willingly or not) hides something; by using unreliable sources; or by changing the range of the axes, playing with sizes of bars and colours to influence consumers' perception and reading of a given graph (for more on this, see Parikh, 2014). During one of our workshops, Alan Smith, principal methodologist at the Office for National Statistics, presented an example involving a graph produced by *The Times*, revised by Full Fact, and re-revised by him (see Box 2 on page 19). He concluded: 'people who get paid to make graphs can make very different things with the same 11 points' – which is precisely what leads stakeholders to be sceptical about relying on data visualisation as a tool for decision-making.

Because visuals can be used deceptively (purposefully or not), there has been a lot of debate as to whether data visualisation is indeed an aid that enhances understanding. In a piece for the *Creative Review*, Patrick Burgoyne asks:

> *Yes, graphical invention can be used to explain complex ideas and present detailed data in digestible form in the cause of an argument or political position, but this will not necessarily aid understanding. As newspapers have known for decades, a graph is just another way of telling a story. But whose story?* (Burgoyne, 2010)

Resistance to using data visualisation has also been a consequence of visualisation that is not deceptive but is poor in accomplishing better understanding of data. A number of journalists, academics, and policy-makers have become engaged in this discussion, often writing articles, blog posts, or even dedicating entire website sections or Tumblrs to display examples of bad visualisations (see e.g. Viz.wtf, n.d.; Yau, n.d.; Limer, 2013; Sonderman, 2014; Diaz, 2011). Nonetheless, as Alberto Antoniazzi from the *Guardian* concludes, this 'shows that data analysis is part of all our lives now, not just the preserve of a few trained experts handing out pearls of wisdom' (quoted in Rogers, 2014).

This points to a very important aspect of regulating the use of effective data visualisation: quality matters. In this section, we discuss best practice in data visualisation, as developed by social scientists, journalists, public policy and business leaders, and conclude by deriving a set of overarching guidelines for the elaboration of quality data visuals.

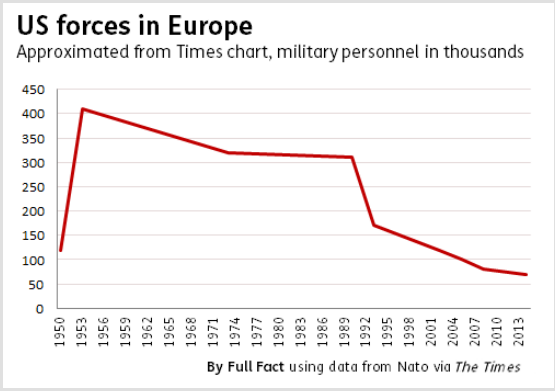
Box 2. Data Visualisation and Reliability

Alan Smith clarified the importance of doing quality data visualisation

On 4 June 2014, *The Times* wrote an article which contained a graphical display of the absolute number of US military personnel acting in Europe across the years, from 1945 to 2014 (see graph A). Alan Smith noted that the X-axis on the graph contained eleven points, which, although unevenly distributed in terms of numerical intervals, had been evenly distributed on the graph. As such, the illustration depicted the distance between the first two points on the graph, of five years, the same way it depicted the distance between its consecutive points (of three years, twenty(!) years, seventeen years, three years, and so on). On Twitter, Smith wrote: 'Graphic in The Times today needs a little dose of x-axis medication…'. The Tweet opened up a debate about common misuses of data visualisation and many Twitter users got involved in the conversation. Full Fact picked up the discussion, corrected the X-axis and produced graph B, which conveys very clearly that the change in the correction of the axis makes a big difference in terms of visual perception and graphical interpretation. The X-axis was not the only problem, however: the first point on the Y-axis (1945) was annotated in the original as having a value of 3 million, while all other points were below 450,000 (where the Y-axis ended). Alan Smith corrected this, and published graph C, with scales adjusted for both axes – and which sharply contrasts with graph A.

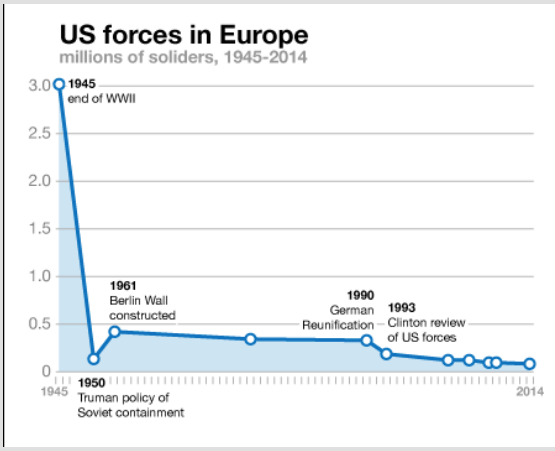
Graph A) From *The Times*

Source: *The Times*, 4 June 2014



Graph B) From Full Fact

Source: O'Brien, 2014



Graph C) From Alan Smith

Source: Alan Smith, via Twitter on 4 June 2014, and later presented at our first workshop.

Data Visualisation in the Social Sciences

Edward Tufte was one of the first academics to engage in the systematic study of data visualisation and to produce guidelines for best practice in using graphic displays. In his 1983 book, *The Data Display of Quantitative Information,* he argues that graphical excellence 'gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space' (Tufte, 1983: 51). Tufte's tradition of simple graphics that fairly represent data have followed the scholar throughout his career, granting him the title of 'Leonardo da Vinci of data'. Although Tufte provides guidance that is specific to types of data and graphics, he also poses a number of general principles, which should be applied as a means of achieving presentation excellence. These can be summarised as:

(1) Maximise the ink/space devoted to actual data and not to other decorative aspects ('chart junk'); in other words, less is more when it comes to non-data information.
(2) Do not omit detail that may be important for graphic interpretation/understanding – this includes clear specification of axes, labelling, etc.
(3) When possible, use graphics to achieve a number of functions.
(4) Do not use graphical representation to deceive your audience in regards to what the actual data point to.
(5) Do not use graphics when not necessary – for instance, when data can be summarised using a simple table.

Departing from the standpoint of human cognition, Isabel Meirelles, an information designer and associate professor of graphic design in the US, understands data visualisation as a complex process that is grounded on perception, human–computer interaction, personal knowledge, and experiences. According to her, it is the responsibility of the designer of data presentation to anticipate the potential challenges that data visualisation may pose to the reader and produce designs that solve these problems (Meirelles, 2013).

Beyond making data visualisation easily shareable, another way in which academics could strengthen communication across sectors is by using visualisation as a tool that can, beyond summarising information, be a compelling hook. William Allen, from the Centre on Migration, Policy and Society (COMPAS at the University of Oxford), suggested this can be done by incorporating relevant colours, shapes, and images into data visualisation.[9] The Observatory also allows interested individuals to build their own charts, based on specific interests, which is also a useful way of allowing interaction and interest in an organisation's work.

[9] It is important to highlight, however, that while colours, shapes, and images can be used to enhance clarity and make data more persuasive, they are also used pervasively to manipulate audiences.

Data Visualisation in Journalism

Data journalism is becoming more common, a trend that is confirmed by the rapid rise in data-driven projects led by major and niche media outlets. The *New York Times*'s The Upshot, ESPN's FiveThirtyEight, led by Nate Silver, the *Guardian*'s DataBlog and Graphics page, and the BBC's Interactives, are just a few examples. Even media outlets that do not have specific sections on their sites specifically dedicated to data-driven journalism and/or data visualisation have engaged in using data visuals as a tool for greater impact. In a recent talk at the Reuters Institute, Amanda Farnsworth, Editor for Visual Journalism at the BBC, conveyed that readers of the BBC website are more likely to click on article links that included a data visual instead of a picture.

Consistent with the advice of Tufte, David McCandless, data journalist and author of the website and book *Information is Beautiful*, suggests: (a) 'distilling' data and using the fewest possible number of words to present information in a useful, aesthetically pleasing, manner; (b) understanding that patterns and connections are the main aspects that should be clearly illustrated by data; (c) compressing huge amounts of information into coherent small frames (McCandless, 2010). Similar principles are also developed by Andy Kirk who, using actual graphic illustrations from the *New York Times*, points to ten important elements of data visualisation in the media industry: clarity of context and purpose; respect for the reader; editorial rigour and integration; clarity of questions; data research and preparation; visual restraint; layout and placement; diversity of techniques; technical execution; and annotation (Kirk, 2012a).

Journalists, however, face some challenges that academics may not. One of them is that the dynamism of the sector presses for quick turnaround. In this context, Claire Miller, Senior Journalist at Trinity Mirror Regionals, explained during one of our workshops that data visualisation for the media needs to be produced quickly, which often means elaborating something simple. Without referring to Tufte, Kate Day, Data Journalist at the *Telegraph*, also refers to some of his now second-nature principles: 'less can be more powerful', which, for a number of practical reasons, has also become a motto for data visualisation in journalism.

Although always present in print media, data visualisation has become even more prominent in online publications, which often use social media to promote stories. As Day mentioned in one of our workshops, 'every click is a choice [made by the reader]' and data visualisation is one of the tools currently used by journalists to help readers make that choice – by summarising interesting information into an easily understandable and 'shareable' illustration. Although speaking of 'better practices' in journalism, Day offered helpful advice to academics as well. Knowing your audience, thinking about language that may be more appropriate to this audience, providing a catchy headline, and focusing on the main point of your 'story'

(research) were some of Day's tips of which academics should take note.

Finally, Miller also pointed out that data visuals are not only a way of summarising an accompanying story, but are often what trigger the curiosity of a writer to pick up and further investigate a story. More than the end of a project, data visualisation can also be the point of departure for research.

Data Visualisation in Public Policy and Business

Visualisation for public policy and business may have different purposes, including: raising awareness of a certain issue, persuading a group of stakeholders to take a certain position, or evaluating the outcomes of a policy. For example, Aleks Collingwood of the Joseph Rowntree Foundation uses the presentation and visualisation of data to draw attention to social policy questions. One way in which her organisation is able to do this is by having established an online platform that allows interested parties to easily access up-to-date data, use, and share their graphics.

Tufte's recommendations for data presentation in academic work and journalism are also in line with the guidance businesses currently receive. Stephen Few, business consultant and founder of Perceptual Edge, advises his clients, which include a substantial number of universities, technology firms, government bodies, financial institutions, and other organisations, that to use data visualisation to gain a business edge, they should: (a) use graphics to present information in a concise and dense manner; and (b) add value and power to information through complexity (Few, 2009).

Unlike academics, who currently use data visualisation overwhelmingly for the purpose of displaying final results, policy-makers and business leaders may also be interested in using data visualisation to initiate investigations of what they do not know – for example, by asking 'why has a given policy/plan not yielded the expected results?' – and using visuals to provide potential answers.

Overarching Good Practice?

Although practitioners from different areas have analysed data visualisation through distinct lenses, social scientists, journalists, policy-makers, and business leaders have reached one clear conclusion: data visualisation should be used (by consumers or producers) to enhance the understanding of large amounts of numerical or textual data. Practical applications of this includes:

(1) Start early: think about data visualisation as an exploration *and* dissemination tool that should be used throughout a project, not only at the end.
(2) Know your audience and understand their needs: tell people what is relevant to them, and, when possible, make it personal.[10]

[10] For good practices pertaining to specific types of graphs, visit: <https://www.gov.uk/service-manual/user-centred-design/data-visualisation.html#gds-example>.

(3) Consider whether your data would be better understood if summarised visually. Ask yourself: Do I have a lot of data that need to be summarised? Or are my data structurally complex? Do I have a story to tell?[11]
(4) Identify the appropriate type of chart to use (see Appendix B). Ask yourself: What is the story I want to tell? What do I want to show: hierarchies, distributions, processes, trends, correlations, etc.?[12]
(5) Display relationships that the brain can understand easily, which include:
    a. Difference (same–not the same; alike–different)
    b. Sizes
    c. Positions/locations
    d. Sequences (order; pattern; continuity)
    e. Time and timelines
    f. Series (grouping; arranged; occurring in a certain order).
(6) Present data in categories that are meaningful to your audience. For example:
    a. Choose between metric system or US customary units depending on where you are presenting
    b. Translate currency to values that are more commonly understood by your audience
    c. Contextualise very large amounts (e.g. £1 billion) by comparing them to a unit of measurement that your audience can more easily understand (i.e. showing what government programmes £1 billion could cover, or how many units of a common household utility £1 trillion could buy).
(7) Properly label all axes in simple language and ensure that scales are correctly illustrated.
(8) Provide clear, interesting titles to graphics.
(9) Unclutter: get rid of different fonts, colours, and information that may detract from your main 'story'.
(10) Illustrate how data disaggregates throughout different levels of analysis (micro and macro), when possible.
(11) Always provide underlying data, as a means of allowing validation and replication.

[11] Guidelines from Gov.uk (2014) suggest the following: 'If there are very few data points (eg top rate income tax down 5%, all other rates unchanged), it's clearer to write a sentence than draw a picture. If every item must be read precisely (to several decimal places) then a table is best.'
[12] Severino Ribecca's *Data Visualisation Catalogue* (n.d.) offers explanation of the different uses of a variety of types of charts and helps you decide on a visualisation method based on what story you want to tell. The data company Visage also offers a comprehensive free resource that explains how to choose charts based on your data. The e-book can be downloaded at: <http://visage.co/data-visualization-101-design-charts-graphs>.

5. Data Visualisation in Practice: Different Pathways

Knowing the good practices and expectations of professionals from different fields in regards to data visualisation may certainly help in producing better graphics. Nonetheless, challenges associated with gaps of knowledge may remain. In a July 2014 report based on 45 interviews with business leaders on data talent, Nesta found that a 'perfect data analyst' would have the following qualities, as summarised by figure 3.

	Description
Core skills	• **Analytical:** strong mathematics and statistics foundation. • **Technical:** programming and database skills.
Domain and business knowledge	• **Knowledge of the sector:** understanding data sources and real-world situations and processes behind the data. • **Awareness of business goals and processes:** knowing the business questions that matter, and how data fits with the organisation.
Soft skills	• **Storytelling:** Ability to transform analytical insights into actionable – and compelling – business recommendations. • **Team–working:** Enjoy working with people from different disciplines.
Competencies	• **Analytical mindset:** Being able to re–formulate complex questions as analytical tasks. • **Creativity:** Knack for generating unexpected solutions to problems, and exploring data from different angles. • **Curiosity:** Wanting to understand how the world works.

Source: Bakhshi et al., 2014: 22.
Figure 3.

Given the many desired qualities, it is thus not surprising that the companies surveyed have struggled to find individuals with the 'right mix of skills' (Bakhshi et al., 2014). In producing high-quality data visualisation, many individuals may encounter similar struggles. For this reason we suggest that, although not always feasible or appropriate, a more combinatorial and collaborative approach to data visualisation could be considered as a way of overcoming skill gaps. In this section, we suggest five potential pathways to producing and disseminating high-quality data visualisation.

Do It Yourself

The first, and most obvious, pathway to data visualisation is to do it yourself, either with your existing knowledge of data analysis software, using simple platforms that do not require coding, or by undertaking further training that would allow the production of data visualisation that can be more broadly used.

Tableau, for instance, is a platform that does not require knowledge of coding and that is made available for free through Tableau Public (Tableau Software is also free to students from anywhere in the world and for

instructors at a number of universities) (Tableau Software, n.d.). Additionally, training in Java Script platforms, such as *D3.js*, for instance, would be a good investment for someone who would like to produce their own graphics, yet make them accessible to broader audiences and/or be interactive. GovLab Academy offers a number of related instructional videos, while Dashing D3.js (Gutierrez, n.d.) and Scott Murray (Murray, 2012) offer free online tutorials on D3.js. Appendices C, D, and E to this report include three potentially useful lists for those who prefer producing data visualisation themselves: a list of popular blogs on data visualisation curated by experts on the topic, a list of relevant free software and platforms, and a list of free online courses on data visualisation.

Collaborate with Cross-Sector, Cross-Departmental Professionals

Collaborations, either cross-departmentally within the same organisation or done across organisations and/or sectors, is one way of overcoming potential skill gaps. During one of our workshops, John Walton, Senior Broadcast Journalist for the BBC Visual Journalism Unit, shared his experience and conveyed that the collaboration between academics and journalists is already taking place in journalism, and that it generally leads to fruitful results. In fact, the most popular pages in 2013 from the BBC and *New York Times*, were 'The Great British Class Calculator' and 'How Y'all, Youse and You Guys Talk' (respectively), both data visuals and both products of collaborations between academics and journalists. Walton says that this type of collaboration has become more and more common. To complete its recent 'NHS Winter Project' the BBC put together a team of journalists, academics, data scientists, and web-developers.

Amanda Farnsworth further attested to this in her recent talk at the Reuters Institute. For her, one of the main things is to ensure that the data being used in BBC reports are solid and reliable; this often means collaborating with academics, scientists, and other experts. To develop a map on corruption, for instance, the BBC partnered with Transparency International; for a project about hypothetically placing a human on Mars, they worked with scientists from Imperial College. The strategy adopted by the BBC for the dissemination of their visuals is also worth noting: Farnsworth emphasised that oftentimes material presented online has also been mentioned in TV reports, thus serving as a means of generating content for television *and* further bringing attention to (increasing the impact of) online publications.

Gov.uk used its existing staff from the Government Digital Service (GDS) as well as the free tools Google Analytics and Tableau Public to produce and publish data visualisations to track how the content on 26 departments' websites is being viewed. Although the GDS has been leading these efforts, they have been the collaborative result of a team of statisticians, policy-makers, and creative producers within the GDS and in other

departments. This strategy comes from an understanding that, through data visualisation, the GDS is providing department leaders with something they can easily understand and act upon. As such, the GDS has been very transparent about how it has produced and continues to produce its data visuals; it has detailed its use of specific platforms and software, as well as provided departments with very clear guidelines on how to produce their own visuals. In this way, the GDS has encouraged other government departments to rely on data visualisation for research (and policy changes) as well as to produce quality data visualisations for their own specific needs – all of this without the need for constant cross-department coordination. In sum, by relying on free and technically accessible platforms, and by providing easy-to-follow guidelines, the GDS managed to impact policy and promote data visualisation across departments through formal and informal collaborations.

Seek the Help of IT Professionals (Formally or Informally)

Collaborating with web-developers/IT experts and designers, has also proven an effective way for organisations to produce data visualisation. For the Brazilian Chamber of Deputies this meant relying on volunteers to transform previously inaccessible big data into structured data that could be easily visualised. 'Hacking Marathons' (*Maratonas Hacker*) became the solution: by sponsoring hacking competitions while making its data available, the chamber managed to find teams of experts (often composed of hackers, activists and scholars) who transformed large amounts of data on a number of subjects into interactive platforms and visuals (Câmara dos Deputados, 2014). The United Nations has also engaged in similar efforts. In a partnership with Visualising.org, the UN Global Pulse sponsored a visualisation competition using data results from the UN Global Pulse's 2010 Mobile Survey (Wiederkehr, 2011). Similar efforts are becoming increasingly common and manage to accomplish communication between different sectors in two stages: that of visualisation production and of dissemination.

The assistance of IT professionals is also often used in more formal ways. This is precisely what the Joseph Rowntree Foundation did to build a data visualisation tool on its website. As Aleks Collingwood, their Programme Manager and Statistics and Quantitative Specialist, explained during one of our workshops, the Foundation wanted to become an important source for data on poverty, place, and ageing. Before launching the JRF Data, the Foundation created new roles and hired individuals with statistics and data visualisation backgrounds to fulfil the required tasks, including Collingwood's role. Developing a system based on Google Fusion, the Foundation consulted with a number of technical experts (from places including the *Guardian* Data Blog), and conducted focus groups with their potential data consumers (key stakeholders in academia, non-profit, media, and government sectors). Currently, their website features data visualisations

for 100 indicators, and offers over 120 new pages with visualisations. On each page, an interactive graph or map is accompanied by the embed code for the image displayed, as well as buttons that allow for easily saving it, as well as a link to the raw data used to develop the graphic.[13]

Not many organisations have the capacity to hire a new team (or train an existing team) on analysing and presenting data, however. Another way of formally seeking the assistance of data visualisation professionals is by outsourcing the work. The rapid increase in interest in data visualisation created a market for data professionals, and a number of firms that provide 'data solutions' emerged. Companies such as Periscopic, Visually, ScraperWiki, and MetaLayer provide a variety of services on data gathering, analysis, and visualisation, and often work on a per-project basis.

Provide Data and Allow Unique Visualisation

Finally, the last pathway we suggest is that of allowing consumers to do the work. This means providing data and platforms that facilitate the production of visuals, and allowing individuals to explore relationships on their own (and later disseminate visuals as they please). This is now done by a number of institutions and organisations. The United Nations and the World Bank, for example, provide data for their indicators and allow interested individuals to easily select the population of observations (e.g. countries, years) and the variables that interest them before producing graphics that can be displayed online and/or saved to one's computer (UNData, 2015; World Bank, n.d.).

First developed by Professor Hans Rosling and currently supported by the Gapminder Foundation, Gapminder World is an online platform that allows individuals to choose from a diverse range of datasets (from sources such as the United Nations, the World Bank, the International Labour Organisation, and the Organisation for Economic Co-operation and Development) and build interactive visualisations that display cross-country time series relationships (Gapminder, 2014).

Many Eyes is yet another website that allows people to use existing datasets and create their own visualisations. Originally created by Fernanda Viégas and Martin Wattenberg, Many Eyes is different from other aforementioned platforms because it also serves as a depository for datasets; this means that researchers and other practitioners can upload their datasets into the system and allow others to use them to produce visualisations of their own.[14]

Research centres are also beginning to make their data available for unique visualisations. Oxford University's COMPAS is one of the centres leading this trend. On their website, the Centre not only provides readily available visualisations of data on migration in the UK, but also a platform for

[13] The presentation of data visuals here is very similar to the one by the Information Geographies at the Oxford Internet Institute, which can be accessed at: <http://geography.oii.ox.ac.uk/?page=home>.
[14] See how it works here: <http://www-969.ibm.com/software/analytics/manyeyes/>.

individuals to use the data to create their own charts (and later share them) (Migration Observatory, n.d.). According to Willian Allen, the Centre sees this as a way of directly informing individuals and getting them interested in the work of Centre, but also as a tool for advocacy. In this manner, the Centre thus increases the impact of its research both directly and indirectly (COMPAS, n.d.).

6. Conclusion and Recommendations

Data are now abundant, and governments, companies, media, and organisations are constantly seeking to reap the benefits of these data – whether to identify internal issues and consider potential solutions, to improve transparency, or to communicate findings to clients or other consumers. Data visualisation is thus an important research and dissemination tool. Although evidence has always been the basis of academic research, disseminating findings with non-academics has not. Academics have been falling behind in training on data visualisation, as well as in efforts to increase the dissemination and impact of their research. Becoming more knowledgeable about good practice in data visualisation and seeking to enhance their visualisation capabilities through training and collaboration could thus prove beneficial for social scientists' academic work and beyond-academia impact.

For this to happen, universities need to establish data visualisation support, and provide training to their students. Nesta suggests that one of the ways to begin addressing the skills gap is by incorporating training on free tools of data analysis, visualisation, and open datasets (Bakhshi et al., 2014: 29). On this note, it is important to highlight that training should not *only* take place in IT departments; while technical proficiency is important, the vast array of codeless platforms and software currently allows non-technical experts to produce visuals. Training on effective communication through data visualisation, including the role of social media and artistic efficiency (think Colour Brewer,[15] for instance) should thus also be considered. Furthermore, collaboration is also often useful in allowing for high-quality data visualisation. Interdisciplinary and inter-sector communication (i.e. across humanities, social, natural sciences departments, and beyond academia) seems more important than ever and should be promoted. Finally, issues of diversity and inclusion should also be considered. Data visualisation is still perceived as heavily dependent on technological knowledge (a perception this report seeks to challenge) (Quick et al., 2013); women are disproportionately underrepresented in the tech industry, and debates about gender-balanced contributions to data visualisation are already taking place (Stefaner, 2013; Kirk, 2012b).

Furthermore, engaging in efforts to promote and produce data visualisation is costly. Even if an institution invests in training, for an individual to learn software or coding takes time – as generally do attempts to reach out to potential collaborators. Some of these efforts may end up producing underwhelming outcomes: a given software may not have the capacity to create visuals for all data types, coding may have to be learned and relearned depending on usage, and collaborations may fall apart or delay the dissemination of findings. Experimenting with different formats and

[15] See: <http://colorbrewer2.org>.

pathways to data visualisation may thus be a good way of assessing what best works for individual academics or teams of researchers. Furthermore, assessing the costs and benefits of investing in data visualisation for each project is also essential. Answering some questions may be helpful in this assessment:

- Do I have data that need to be summarised?
- Can my findings be displayed in a unit of analysis that can be easily grasped? If not, can I break it down to a level that inherently makes more sense?
- Are my findings interesting and/or do they have the potential to impact others?
- If so, can individuals/companies/organisations relate to or use my findings to further understand something relevant to them?
- Are my findings more useful for consumption if static or if presentation allows for interaction?
- Can my findings be easily shared and, if so, will they generate further interest in my research?

Given that answers to these questions may vary from project to project, costs associated with training and collaboration may be higher or lower depending on the expected rate of return data visualisation may provide (be it in the form of enhancing theory and hypothesis testing, or in the form of communication and dissemination).

Finally, before concluding this report, it is important to highlight that we advocate for the use of data visualisation in combination with narrative and interpretation (and by providing supporting materials, e.g. data files, codebooks, etc.). Data visualisation is an important tool in improving our inherent understanding of long texts and numbers, and the patterns, trends, and gaps, embedded within large datasets. On its own, however, visualisation can be taken out of context and manipulated to suggest something other than what the larger evidence indicates. Data visualisation is a crucial tool – and increasingly so – for research dissemination, but it should not replace entire datasets, research method outlines, or in-depth interpretations of data.

7. What Now?

The current report concludes the first phase of a project that sought to both, understand and promote data visualisation at Oxford and beyond. Our efforts to advocate for the use of data visualisation, however, do not stop here. This section briefly summarises some of the work the University of Oxford and the Alliance for Useful Evidence will continue to do throughout 2015 to ensure that more people become aware of the benefits and challenges of using data visualisation, and receive the proper training to use this research and communication tool.

Oxford-Based Initiatives

One of the issues pointed out throughout the report is that universities seem to be perpetuating a skills gap by not training students in what is required by the labour market. Oxford currently benefits from hosting one of 15 Q-Step Centres, considered by Nesta as an important step in closing the existing skills gap (Bakhshi et al., 2014). Funded by the Nuffield Foundation and hosted by the Department of Politics and International Relations in collaboration with the Department of Sociology, the Centre is part of a £19.5 million investment for the development and promotion of quantitative training in social science departments across the UK. The Centre provides lectures, data-labs, summer school programmes, and open-access online teaching to undergraduates of the University of Oxford and beyond.[16] As part of the programme, undergraduate students at Oxford are already learning how to make data visuals using R.

The IT services are also doing their part. In January 2015, Howard Noble, Research Support Service Manager at the IT Services, secured funding for a project that seeks to investigate how IT can better 'support researchers who want to engage the public, and academics in other fields, by publishing data visualisations'. The project proposal also includes the creation of 'a prototype data visualisation web infrastructure' and will conclude by defining 'the cost of a full data visualisation support service in terms of staff resource and technical infrastructure'. While undergoing this research, IT services are also currently offering a series of lectures on data visualisation (Patrick, 2015).

Also seeking to promote capacity-building in data visualisation, the Doctoral Training Centre of the Social Sciences Division and the Economic and Social Research Council have recently co-sponsored a student conference, in which data visualisation was also a focus, as illustrated by its choice of keynote speakers (Alan Smith, Head of Digital Content and Data Visualisation at the UK's Office for National Statistics, and Professor Danny Dorling, from Oxford University's School of Geography and the Environment). The conference also offered two sections of a practical

[16] See <www.oqc.ox.ac.uk/about/about.html>.

workshop on using data visualisation to communicate academic research, which were led by Malu A. C. Gatto. Other institutions at the University are also investing in data visualisation training. For instance, the Department of Continuing Education recently ran a four-day workshop on data visualisation for doctoral students in Ecology,[17] and All Souls College also hosted a popular workshop last April.[18]

Furthermore, the Social Sciences Division and the Department of Politics and International Relations seek to continue their efforts to promote data visualisation and research impact. They will continue to provide workshops on how to design academic posters, which often contain elements of data visualisation, and to organise sessions to give students the opportunity of presenting their posters. The Department will also promote relevant uses and examples of data visualisation on its new website, showcasing examples of data visuals that use data produced by DPIR faculty and students, highlighting collaborations for research-based data visualisation and opportunities for training in this area.

Other departments and centres are also working towards the engagement between academic research and civil society. The Reuters Institute for the Study of Journalism has done this through the current project, as well as by hosting a number of talks on data journalism, all of which are available (in podcasts) on its website (RISJ, 2014). COMPAS and the Oxford Internet Institute are doing this by producing and publishing interactive and easily shareable data visuals of their research projects (Oxford Internet Institute, n.d.).

Nesta-Based Initiatives

The UK's innovation charity Nesta is working with the Royal Statistical Society, Creative Skillset, and Universities UK to identify data analyst skills shortages (important skills for data visualisation) in the UK, and develop policies and interventions to address this gap.

Nesta has recently begun to use data visualisation to communicate its research. The first visualisations were popular. The interactive visualisation of the UK creative economy <http://www.nesta.org.uk/blog/interactive-data-visualisations-uks-creative-economy> received 1,400 tweets and international interest. Over the next year a range of others will be published. A variety of approaches to constructing the visualisations will be trialled, from in-house production to outsourcing. And any lessons learned from the process will be shared. Nesta would like to hear about others' experiences too.

Visualisation is an effective tool because both it lets Nesta e*xplain* their insights to readers and allows readers to *explore* the data to draw their own insights. It makes efficient use of space, and effective use of our large visual

[17] For more information: <https://www.conted.ox.ac.uk/courses/details.php?id=C180-14&utm_source=datavis&utm_campaign=envman1415&utm_medium=email>.
[18] For more information: <https://storify.com/digipanoptic/digital-panopticon-data-visualisation-workshop>.

system. Nesta's audience is particularly diverse (not just academics, but policy-makers, business, general public, etc.). They have different interests – and so it is important that they can look at the data in different ways and extract the insights most useful for them.

Research on innovation often focuses on *new* or *less visible* types of economic activities (e.g. alternative finance, sharing economy, innovation), which many readers will not be familiar with or which are more difficult to picture. Also, Nesta's data are often novel, and so it's important for them to be shared with others.

The Alliance for Useful Evidence, based at Nesta, is also conducting a systematic review of 'what works' in research uptake which will look at the different approaches to exchanging research, including visualisation. Evidence masterclasses are being run for policy-makers and charity leaders, called the 'Live Issue Simulator'. These include the best ways to present and visualise research for impact.

Bibliography

Anselmo, K. (2015). 'With Academia Moving in a Digital Direction, Sustained Investment in Media Training Would Benefit All.' <http://blogs.lse.ac.uk/impactofsocialsciences/2015/01/28/digital-academia-sustained-investment-media-training> [Accessed Feb. 2015].

Bakhshi, H., Mateos–Garcia, J., and Whitby, A. (2014). *Model Workers: How Leading Companies are Recruiting and Managing their Data Talent*. London: Nesta. <www.nesta.org.uk/sites/default/files/model_workers_web_2.pdf> [Accessed Feb. 2015].

Bakhshi, H., Bravo-Biosca, A., and Mateos-Garcia, J. (2014). *Inside the Datavores: Estimating the Effect of Data and Online Analytics on Firm Performance*. London: Nesta. <www.nesta.org.uk/sites/default/files/inside_the_datavores_technical_report.pdf> [Accessed Feb. 2015].

Burgoyne, P. (2010). 'Data Visualisation: An Aid to Understanding?' Creative Review Blog. <www.creativereview.co.uk/cr-blog/2010/august/data-visualisation-aiding-understanding> [Accessed Jan. 2015].

Burn-Murdoch, J. (2013). 'Why you Should Never Trust a Data Visualisation.' <www.theguardian.com/news/datablog/2013/jul/24/why-you-should-never-trust-a-data-visualisation> [Accessed Jan. 2015].

Câmara dos Deputados, Brasil (2014). 'Conheça Mais.' <www2.camara.leg.br/responsabilidade-social/edulegislativa/educacao-legislativa-1/educacao-para-a-democracia-1/hackathon/2014> [Accessed Feb. 2015].

Câmara dos Deputados, Brasil (2014). 'Educação para a democracia – Hackaton Gênero e Cidadania.' <www2.camara.leg.br/responsabilidade-social/edulegislativa/educacao-legislativa-1/educacao-para-a-democracia-1/hackathon/2014> [Accessed Jan. 2015].

Campbell, J. (2014). 'Women Trapped in Universities' Ivory Basements.' <www.independent.co.uk/news/education/education-news/women-trapped-in-universities-ivory-basements-9908457.html> [Accessed Jan. 2015].

COMPAS (n.d.). 'Big Data, Big Visions.' <https://www.compas.ox.ac.uk/research/the-migration-observatory/big-data-big-visions> [Accessed Feb. 2015].

Data.gov (2014). 'Data.gov Homepage.' <https://www.data.gov> [Accessed Jan. 2015].

Diaz, J. (2011). *Stop Already with the F*cking 'Infographics'*. <http://gizmodo.com/5846087/stop-already-with-the-fcking-infographics> [Accessed Jan. 2015].

Few, S. (2009). *Now you See it*. Oakland, CA: Analytics Press.

Few, S. (2014a). 'Data Visualization for Human Perception.' *The Encyclopedia of Human-Computer Interaction*, 2nd edn. <https://www.interaction-design.org/encyclopedia/data_visualization_for_human_perception.html> [Accessed Feb. 2015].

Few, S. (2014b). 'Why do we Visualize Quantitative Data?' Perceptual Edge. <www.perceptualedge.com/blog/?p=1897> [Accessed Jan. 2015].

Gapminder, (2014). 'Unveiling the Beauty of Statistics for a Fact Based World View.' <www.gapminder.org> [Accessed Feb. 2015].

Gov.uk (2014). 'Data Visualisation: Government Service Design Manual.' <https://www.gov.uk/service-manual/user-centred-design/data-visualisation.html> [Accessed Feb. 2015].
Gunelius, S. (2014). 'The Data Explosion in 2014 Minute by Minute: Infographic.' <http://newstex.com/2014/07/12/the-data-explosion-in-2014-minute-by-minute-infographic> [Accessed Jan. 2015].
Gutierrez, S. (n.d.). 'D3 Tutorials.' <https://www.dashingd3js.com/table-of-contents> [Accessed Jan. 2015].
Hansen, Charles D., and Johnson, Chris R., eds (2005). *The Visualization Handbook*. Burlington, MA, and Oxford: Academic Press.
Hardy, Q. (2012). 'What are the Odds that Stats Would Be This Popular?' *New York Times Bits Blog*. <http://bits.blogs.nytimes.com/2012/01/26/what-are-the-odds-that-stats-would-get-this-popular/?_r=0> [Accessed Feb. 2015].
Heywood, J. (2014). 'Embracing What Works.' <https://openpolicy.blog.gov.uk/2014/12/09/embracing-what-works> [Accessed Jan. 2015].
Hill, S. (2014). 'REF Impact: Much More than Just Economic.' <http://blog.hefce.ac.uk/2014/10/15/ref-impact-much-more-than-just-economic> [Accessed Jan. 2015].
Iliinsky, N. (2012). 'Why is Data Visualization So Hot?' <http://blog.visual.ly/why-is-data-visualization-so-hot> [Accessed Feb. 2015].
Kirk, A. (2012a). *Data Visualisation*. Birmingham: Packt Pub.
Kirk, A. (2012b). 'Discussion: Is Data Visualisation Gender Blind?' <www.visualisingdata.com/index.php/2012/07/discussion-is-data-visualisation-gender-blind> [Accessed Feb. 2015].
Koch, K., McLean, J., Segev, R., Freed, M., Berry, M., Balasubramanian, V., and Sterling, P. (2006). 'How Much the Eye Tells the Brain', *Current Biology*, 16(14): 1428–34. <www.sciencedirect.com/science/article/pii/S0960982206016393> [Accessed Jan. 2015].
Kraker, P. (2015). The researcher's guide to literature: Visualising crowd-sourced overviews of knowledge domains'. *The Impact Blog*. 16 February 2015. http://blogs.lse.ac.uk/impactofsocialsciences/2015/02/16/crowd-sourced-overview-visualizations-of-knowledge-domains/.
Limer, E. (2013). '8 Horrible Data Visualizations that Make No Sense.' <http://gizmodo.com/8-horrible-data-visualizations-that-make-no-sense-1228022038> [Accessed Jan. 2015].
Lindoso, V. (2014). 'Bringing in the Doctors: Can Academia Save Government from Itself?' <http://blogs.bsg.ox.ac.uk/2014/12/19/bringing-in-the-doctors-can-academia-save-government-from-itself> [Accessed Jan. 2015].
LSE Impact of Social Sciences Blog (2011). 'Continual Publishing across Journals, Blogs and Social Media Maximises Impact by Increasing the Size of the "Academic Footprint".' <http://blogs.lse.ac.uk/impactofsocialsciences/2011/10/26/academic-footprint> [Accessed Feb. 2015].
McCandless, D. (2010). 'The Beauty of Data Vizualisation.' <www.ted.com/talks/david_mccandless_the_beauty_of_data_visualization?language=en> [Accessed Jan. 2015].

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., and Hung Byers, A. (2011). 'Big Data: The Next Frontier for Innovation, Competition, and Productivity.' <www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation> [Accessed Feb. 2015].
Meirelles, I. (2013). *Design for Information*. Gloucester, MA: Rockport.
Migration Observatory (n.d.). 'Data and Resources.' <http://migrationobservatory.ox.ac.uk/data-and-resources> [Accessed Feb. 2015].
Murray, S. (2012). 'D3 Tutorials.' <http://alignedleft.com/tutorials/d3> [Accessed Jan. 2015].
Nesta.org.uk (n.d.). 'Case Study: Evidence-Based Policing.' <www.nesta.org.uk/develop-your-skills/developing-and-testing/case-study-evidence-based-policing#sthash.XwNzWI28.dpuf> [Accessed Jan. 2015].
Noguchi, Y. (2011). 'The Search for Analysts to Make Sense of "Big Data".' <www.npr.org/2011/11/30/142893065/the-search-for-analysts-to-make-sense-of-big-data> [Accessed Feb. 2015].
O'Brien, Laura (2014). 'Bad Charts: Times has Trouble with its Times', Full Fact, 4 June. <https://fullfact.org/live/2014/jun/times_bad_chart_military_personnel-32903>.
Oxford Internet Institute (n.d.). 'Information Geographies – Homepage.' <http://geography.oii.ox.ac.uk/?page=home> [Accessed Feb. 2015].
Parikh, R. (2014). 'How to Lie with Data Visualization.' <http://gizmodo.com/how-to-lie-with-data-visualization-1563576606> [Accessed Feb. 2015].
Patil, D. (2013). 'Still the Sexiest Profession Alive.' <https://hbr.org/2013/11/still-the-sexiest-profession-alive> [Accessed Feb. 2015].
Patil, D. (2014). 'Three Ways Data Science is Changing the World.' <https://agenda.weforum.org/2014/03/three-ways-data-science-changing-world> [Accessed Feb. 2015].
Patrick, M. (2015). 'Data Visualisation: A Series of Lunchtime Talks.' <http://blogs.it.ox.ac.uk/acit-rs-team/advice/software-selection/vistalks> [Accessed Feb. 2015].
Quick, M., Bergamaschi, F., and McCandless, D. (2013). 'Diversity in Tech.' <www.informationisbeautiful.net/visualizations/diversity-in-tech> [Accessed Feb. 2015].
Reuters Institute for the Study of Journalism (2014). 'Visualisation of Academic Research.' <http://reutersinstitute.politics.ox.ac.uk/publication/visualisation-academic-research> [Accessed Feb. 2015].
Ribecca, S. (n.d.) *The Data Visualisation Catalogue*. <www.datavizcatalogue.com> [Accessed Feb. 2015].
Rogers, S. (2014). 'Infographics for Children: What they Can Learn from Data Visualisations.' <www.theguardian.com/news/datablog/2014/mar/07/infographics-for-children-can-learn-from-data-visualisations> [Accessed Feb. 2015].

Sander, N., and Bauer, R. (2015). 'How Data Visualisation Enhances the Impact and Visibility of Science', Princeton Population Association of America. <http://paa2015.princeton.edu/uploads/151205> [Accessed Feb. 2015].
Science Daily (2013). 'Big Data, for Better or Worse: 90% of World's Data Generated over Last Two Years.' <www.sciencedaily.com/releases/2013/05/130522085217.htm> [Accessed Jan. 2015].
Scienceogram UK (2015). http://scienceogram.org.
Sonderman, J. (2014). 'People are Tired of Bad Infographics, So Make Good Ones.' <www.poynter.org/how-tos/visuals/149636/people-are-tired-of-bad-infographics-so-make-good-ones> [Accessed Jan. 2015].
Stefaner, M. (2013). 'Gender Balance in Conferences on Data Visualization, Creative Code, Information Graphics', *Truth and Beauty*. <http://moritz.stefaner.eu/projects/gender-balance/#RATIO/NUM> [Accessed Jan. 2015].
Tableau Software (n.d.). 'Tableau for Students: Free Access to Tableau Desktop.' <www.tableausoftware.com/academic/students> [Accessed Jan. 2015].
Ted.com (n.d.). 'David McCandless', *Ted Speaker*. <https://www.ted.com/speakers/david_mccandless> [Accessed Feb. 2015].
Tufte, E. (1983). *The Visual Display of Quantitative Information*. Cheshire, CT (Box 430, Cheshire 06410): Graphics Press.
UNData, (2015). 'Population Using Improved Drinking Water Sources.' <http://data.un.org/visualization/SH_H2O_IMPR> [Accessed Feb. 2015].
Viz.wtf (n.d.). *WTF Visualizations*. <http://viz.wtf> [Accessed Jan. 2015].
Wall, M. (2014). 'Big Data: Are you Ready for Blast-off?' <www.bbc.com/news/business-26383058> [Accessed Jan. 2015].
Walton, J. (2014). 'Getting Started with Data Journalism.' <www.slideserve.com/branden-farmer/getting-started-with-data-journalism> [Accessed Feb. 2015].
Wiederkehr, B. (2011). 'Visualizing Voices of the Vulnerable on datavisualization.ch.' <http://datavisualization.ch/notes/visualizing-voices-of-the-vulnerable> [Accessed Jan. 2015].
World Bank (n.d.). 'Indicators.' <http://data.worldbank.org/indicator>.
Yau, N. (n.d.). 'Mistaken Data.' <http://flowingdata.com/category/statistics/mistaken-data> [Accessed Jan. 2015].

Appendix A. Project Workshops

Workshop 1

Date: Friday 20 June, 2014.
Venue: Nesta, 1 Plough Place, London, EC4A 1DE

13:00 Welcome

- Jonathan Breckon (Head of the Alliance for Useful Evidence, Nesta)

13:05 Introduction

- Dr David Levy (Director, RISJ)

13:10 Keynote Speech: Data Visualisation and the Fourth Technological Revolution?

- Professor Luciano Floridi (Director of Research, Oxford Internet Institute, University of Oxford)

13:25 Q&A session

13:45 Panel: What are the Needs and Challenges for Data Visualisation?

- Chris Hemingway (Head of Analytics, Fraud, Error and Debt Programme, Cabinet Office)
- Alan Smith (Principal Methodologist, Data Visualisation at Office for National Statistics)
- Aleks Collingwood (Programme Manager, Statistics and Quantitative Specialist, Joseph Rowntree Foundation)
- Claire Miller (Senior Data Journalist, Trinity Mirror Regionals)

15:00 Coffee break

15:15 Panel: What is Currently Being Done in the Academy?

- Chair: Professor Robert G. Picard (Director of Research, RISJ)
- William Allen (Migration Observatory, Compass, University of Oxford) and Rob McNeil (Head of Media and Communications, Migration Observatory)
- Dr Mark Jones (Data Visualisation and Mobile Software, Dept of Computer Science at Swansea)
- Simon Walton and Alfie Abdul-Rahman (both e-Research Centre, University of Oxford)

16:05 Facilitated discussion, including audience Q&A: Risks and Opportunities: What Next?

- Chair: Geoff Mulgan, Chair (Chief Executive, Nesta)
- Claire Miller (Senior Data Journalist, Trinity Mirror Regionals)
  Stephen Khan (Editor, The Conversation) User Perspective
  Olly Arber (Director of Digital, Nesta) User Perspective
  Alan Smith (Principal Methodologist, Data Visualisation at Office for National Statistics)

16.55 Concluding comments and next steps

- Geoff Mulgan (Chief Executive, Nesta) and Professor Robert G. Picard (Director of Research, RISJ)

Workshop 2

Date: Friday 17 October, 2014
Venue: Department of Politics and IR, University of Oxford, Manor Road, Oxford, OX1 3UQ.

9:40 Introduction: What is Data Visualisation? Examples of Good and Bad Data Visualisation

- Professor Robert G. Picard (Director of Research, RISJ)

10:10 Making Data Intelligible: An Introduction to the Scienceogram

- Andrew Steele (Computational Biologist at Cancer Research UK, Science Communicator and Co-founder of Scienceogram)

10:45 Tea and coffee break
11:00 Approaches to Visualisation and Practical Tools

- Kate Day (Director of Digital Content, the *Telegraph*)

11:45 Discussion and consideration of participants of data visualisation needs

Workshop 3

Date: Friday 14 November 2014.
Venue: Department of Politics and IR, University of Oxford, Manor Road, Oxford, OX1 3UQ.

9.40 Introduction: What is Data Visualisation? Examples of Good and Bad Data Visualisation

- Professor Robert G. Picard (Director of Research, RISJ)

10.10 Visualising Migration: Experiences from Migration Observatory

- William Allen (Centre on Migration, Policy, and Society, University of Oxford)

10:45 Tea and coffee break
11.00 Engaging a Global Audience

- John Walton (Senior Broadcast Journalist, BBC Visual Journalism Unit)

11.30 Visualising Texts through Networks

- Félix Krawatzek (DPhil Student in Politics, University of Oxford)

11.45 Discussion and consideration by participants of data visualisation needs

Appendix B. Data Visualisation: Types of Application

One of the basic foundations of data visualisation is the understanding that the choice of charts should follow from both the type of data used and the theory/goal in place. This section discusses the applicability of some of the most common types of graphics and their respective common usages.[19] In the spirit of McCandless, of showing that the basic principles of data visualisation are now commonplace, we first discuss types of graphics to which we are the most exposed on a daily basis. We then move into ways of commonly expressing quantitative findings in static graphics. We continue by explaining what are 'infographics' and how they relate to the forms previously discussed. Finally, we show how static graphics can be, and have been, transformed into animated forms of data display.

Static Graphics for 'Commonplace' Activities

Although graphics are generally associated with bars, lines, and numbers, data visualisation can take many forms. Per definition, the practice seeks to transform information into illustrations that can be more easily and rapidly grasped. Although we may not recognise this, we are exposed to (and use) data visualisation on a daily basis. Calendars, maps, timelines, mind-maps, and illustration diagrams are just a few types of graphics that we may constantly use or be exposed to.

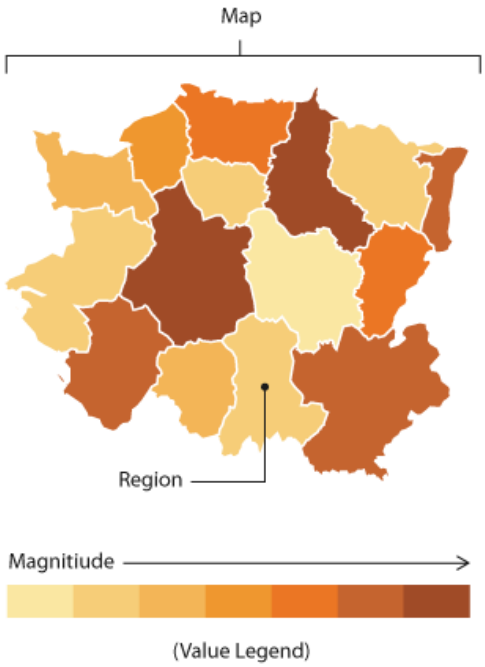
Among these, calendars may be the type that we most regularly use. But how does a calendar comply with the 'principles' of data visualisation? First of all, calendars visually display periods of time that chronically organise events. These events may be given a specific time-slot, in a specific day, which is part of a week, month, and year. While, individually, events may 'tell a story', if analysed in their entirety, calendars can show potential patterns (e.g. classes every Tuesday at 2pm; or bi-weekly meetings on Fridays, etc.), gaps (e.g. weekends, holidays), and other relationships.

Source: Data Visualisation Catalogue, <http://www.datavizcatalogue.com>

Figure 4 – example of a calendar

[19] For extensive explanations on chart applications, see: http://seeingdata.org/sections/inside-the-chart/.

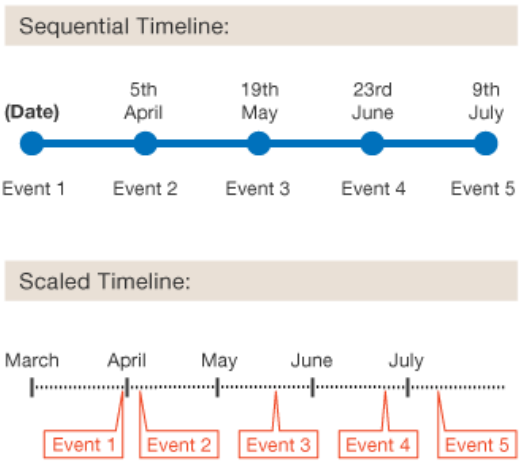
We also use maps frequently. This type of graphic enables us to gain a visual understanding of large geographical spaces by dividing areas into sub-units (e.g. states, cities, school districts) usually using colours to differentiate between them. Colours within maps may also be used to depict the distribution of a variable of interest across a given geographical space. This type of graphic is commonly used to illustrate electoral results and voting patterns, for example. Maps are tools that can assist in our understanding of both spatial and concentration distributions, thus serving more than one purpose (something advocated by Tufte). Used in sequence, maps can also show the impact of time, for example, on a given value of interest, such as voter-party alignment.



Source: Data Visualisation Catalogue, <http://www.datavizcatalogue.com>

Figure 5 – example of a map

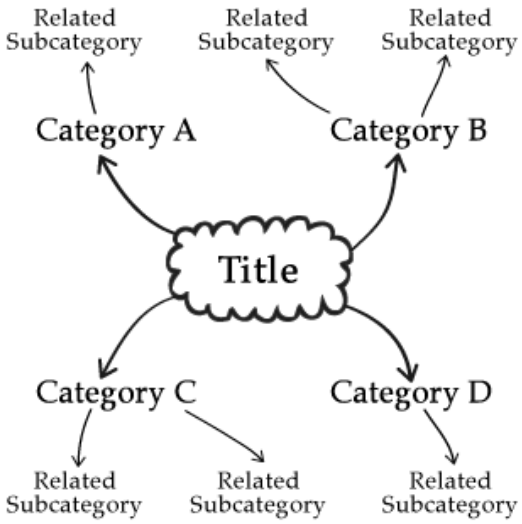
Many people become exposed to timelines early in their school years, given that they are often used as a tool to help students understand sequences of historical events. Timelines can also be frequently found in museums to explain the beginning and end of art periods, and in company project reports to pinpoint different steps and interim deadlines that should lead to final outcomes.



Source: Data Visualisation Catalogue, <http://www.datavizcatalogue.com>

Figure 6 – example of a timeline

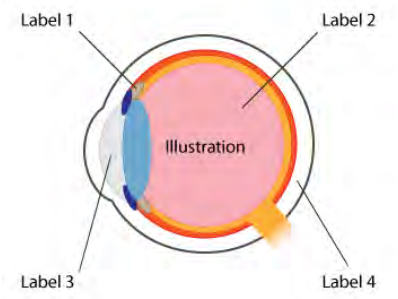
Mind-maps are yet another common tool used to visualise ideas (e.g. plan a paper, project, etc.) and relationships (e.g. between employees in a firm. Similar to mind-maps, tree diagrams are also characterised by small categorical boxes (or circles, etc.) linked by lines. Tree diagrams, however, are unique in that they establish a hierarchy. A common use of this type of diagram is for family trees, which generally go from the most senior known member of the family to the youngest ones.



Source: Data Visualisation Catalogue, <http://www.datavizcatalogue.com>

Figure 7 – example of a mind-map

Finally, illustration diagrams are frequently displayed in textbooks (e.g. to show the different parts of the human body, the different layers of the Earth, the organisation of the solar system, etc.) and manuals (e.g. on how to put together a piece of furniture, on how to play a game, etc.). Illustration diagrams generally depict smaller-scaled images of a given object and explain the different parts of the object through labels that often include name, description, and purpose of individual parts.

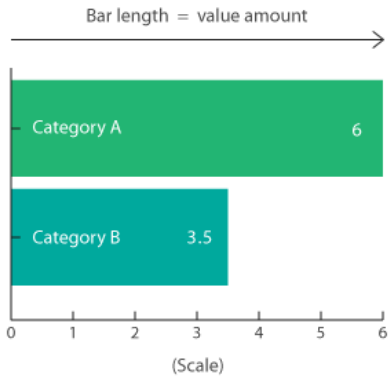


Source: Data Visualisation Catalogue, <http://www.datavizcatalogue.com>
Figure 8 – example of an illustration diagram

Static Graphics for Numerical/Statistical Data Display

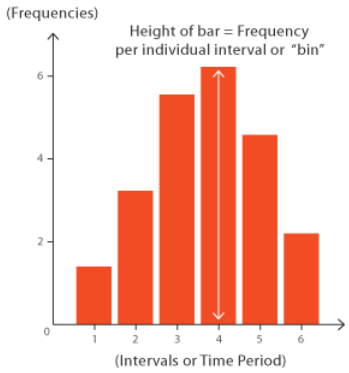
The preceding section shows that data visualisation is something that we are all exposed to in greater or smaller extents on a daily basis. In this section, we show the most commonly used graphic forms for the display of numerical and statistical information.

Bar charts use either horizontal or vertical columns to compare values of Y for defined categories of X. For instance, a chart may depict household income and X may pertain to different ethnic backgrounds. As such, each bar pertains to a specific ethnic background. The size of the line pertains to the average (or other measure) of household income held by a particular ethnic group. This type of graphic allows for the assessment of relationships between variables (X and Y), comparisons across value of X, as well as potential patterns.



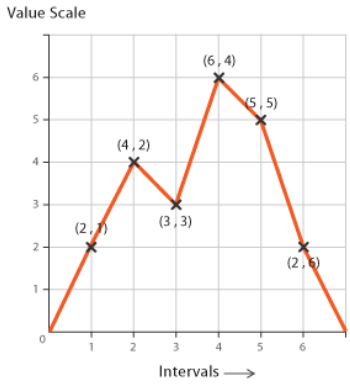
Source: Data Visualisation Catalogue, <http://www.datavizcatalogue.com>
Figure 9 – example of a barchart

Histograms look very similar to bar charts but have a different purpose. While the Y-axis on histograms refers to frequency (or density, proportion, etc.) of a given variable, X always refers to time (e.g. intervals, such as days, years, decades, journal cycles, etc.). This type of graphic allows for the assessment of patterns, distributions, and comparisons that assist in the understanding of how data evolve over time.



Source: Data Visualisation Catalogue, <http://www.datavizcatalogue.com>
Figure 10 – example of a histogram

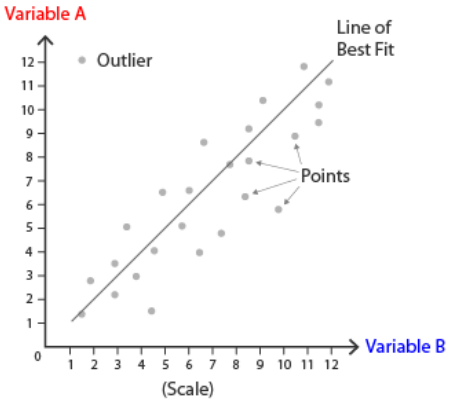
Line graphs are also generally used to display the values of a given variable Y across time. Line graphs can also serve the purposes of bar charts and histograms at the same time. This is because different lines, representing categories of a third variable, may be added to the same plot. This means that line graphs may depict, for example, GDP growth over time for various countries. This type of graphic thus allows not only for the comparison of data over time, but also for how this longitudinal data compares across different categories of a variable other than time.



Source: Data Visualisation Catalogue, <http://www.datavizcatalogue.com>
Figure 11 – example of a line graph

Scatter-plots are also a commonly used type of graphic; they are generally applied to clarify and illustrate relationships between two different variables. As such, each axis represents the scale of a given variable, and each point on

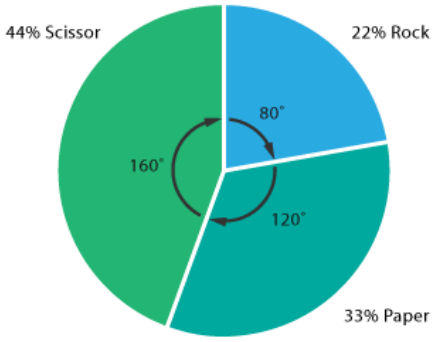
the plot marks the intersection of an observation's values on both variables. Together, points on the plot can show the level of correlation between the two variables of interest.



Source: Data Visualisation Catalogue, <http://www.datavizcatalogue.com>

Figure 12 – example of a scatter-plot

Pie charts are not very commonly used in the academic world, although they are very much employed in both media and business spheres. Pie charts are used to show how different proportions of a whole are allocated. This type of graphic is commonly employed to display how a company's profits are being reinvested, how budget is allocated, or the demographic composition of a firm.



Source: Data Visualisation Catalogue, <http://www.datavizcatalogue.com>

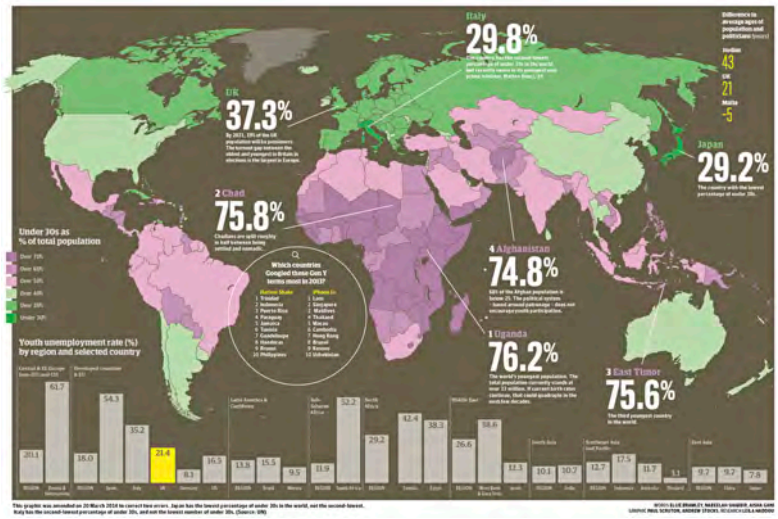
Figure 13 – example of a pie chart

Infographics

Infographics are a static form of data presentation, which generally combine more than one type of graphic and illustrations to tell an encompassing story. Infographics generally combine numerical and non-numerical elements, making them very accessible to non-academic audiences. Media outlets have largely employed infographics, but this form of graphic has also been very popular with international organisations, foundations, non-profits, and businesses that seek to summarise the findings of long reports in easily readable formats. Another benefit of infographics is that they are easily distributed and shared through social media. While reports may remain

largely unread, their main findings may spread through the sharing of their respective infographics. The *Guardian* has used many infographics in its publications; so much so that they have a page dedicated to news stories communicated through them.[20]

Infographics are generally based on the other types of graphics just presented, but combined with other types of illustrations and text to produce a form of presentation that looks less academic and more accessible. This is precisely what figure 14 does. This infographic combines a traditional bar chart, a colour-coded map, and numerical and textual information to cover a number of aspects about youth around the world. Using a colour scheme that identifies the proportion of a given population that is composed of people under 30 years of age (dark green represents less than 30%, pink represents over 50%, and dark purple more than 70%), the graphic allows readers to quickly grasp global and regional trends about youth populations, while also gaining information about individual countries. The infographic also offers information on youth unemployment for select countries (divided by regions) by placing a bar chart at the bottom of the infographic, and highlighting the UK in yellow – thus making it easier for its traditional readership to compare their home country to others. The map also includes information on online use/interaction by adding a list of countries that most Googled the 'General Y terms', 'Harlem Shake' and 'iPhone 5s' (the list is contained in the circle placed between South America and Africa). At the top right corner, the infographic also depicts countries that have the highest, average, and lowest values for 'difference in average ages of population and politicians'. Finally, arrows also point to some countries, providing further youth-related facts that are specific to those countries.



Source: <http://www.theguardian.com/world/graphic/2014/mar/19/world-map-of-youth-youth-bulge>

Figure 14

[20] See: <http://www.theguardian.com/graphics>.

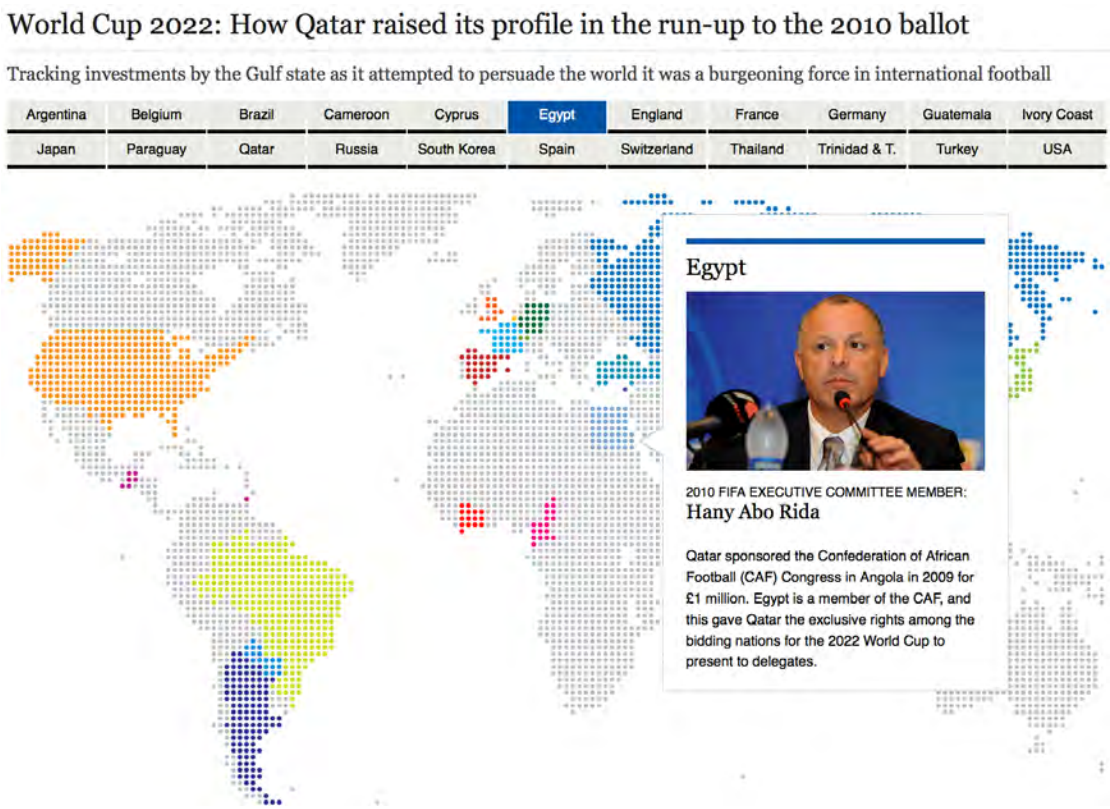
Interactive Graphics

Technological developments and the rise of online media as a popular form of information distribution have also led to an increase in the use of interactive graphics as a form of data presentation. Animated graphics are generally based on one (or more) of the graphic forms explained so far in this Appendix. The difference is that these graphics generally allow interactivity (i.e. readers are capable of choosing which aspects of the graph they would like more information on, often being able to choose the variables that they would like to visualise, including specific time periods, regions, etc.). This type of graphic is therefore extremely powerful: it is capable of presenting a large amount of information in a limited amount of space, while having the potential to satisfying unique individual interests. And more: these graphics tell a full story, often without the need for many words.

Precisely because of their potential power and effectiveness in quickly informing readers, a number of mainstream media outlets have adopted animated graphics as a journalistic tool. The BBC and the *Telegraph* are only a few among the many media outlets that have online sections fully dedicated to displaying their interactive graphics.[21] Examples of this may clarify some of the possible uses of animated graphics.

The *Telegraph* recently published an interactive graphic that uses a map as the base for its presentation on how Qatar raised its profile in the run-up to the 2010 World Cup ballot. This graphic allows the reader to place the mouse over a country of interest to see more information on the type of investment Qatar made in this country. This allows readers to decide the amount of information they want, as well as to focus on what is relevant to them. As the illustration shows, although this type of graphic is interactive and holds more information than the other ones discussed, it is still based on a simple type of data visualisation: a map.

[21] See <http://news.bbc.co.uk/1/hi/in_depth/interactives/default.stm>.



Source: The animated graphic can be accessed at <http://www.telegraph.co.uk/sport/football/world-cup/10871498/World-Cup-2022-How-Qatar-raised-its-profile-in-the-run-up-to-the-2010-ballot.html>.

Figure 15

Appendix C. Blogs: Experts Display Data Visualisation in Practice

- Alberto Cairo's 'The Functional Art' <http://www.thefunctionalart.com>
- Andrew Vande Moere's 'Information Aesthetics' <http://infosthetics.com/>
- Andy Kirk's 'Visualising Data' <http://visualisingdata.com>
- Chris Twigg's 'Stories through Data' <http://storiesthroughdata.blogs.lincoln.ac.uk>
- Cole Nussbaumer's 'Storytelling with Data' <http://www.storytellingwithdata.com>
- Jorge Camões's 'Excel Charts' <http://www.excelcharts.com/blog>
- Lulu Pinney's 'Telling Information' <http://lulupinney.co.uk>
- Matt Stiles's 'The Daily Viz' <http://thedailyviz.com>
- Michal Babwahsingh's webpage <http://michaelbabwahsingh.com>
- Nathan Yau's 'Flowing Data' <http://flowingdata.com>
- Robert Kosara's 'EagerEyes' <http://eagereyes.org>
- Stephen Few's 'Perceptual Edge' <http://www.perceptualedge.com/blog>
- Xaquín González Veira's 'Xocas' <http://www.xocas.com/blog/en>
- Seeing Data <http://seeingdata.org>

Appendix D. Free Software and Online Platforms for Data Visualisation

Basic Charts

- Google Fusion Tables
  <https://support.google.com/fusiontables/answer/2571232>
- Tableau Public
  <http://www.tableau.com/public>
- IBM Many Eyes
  <https://www-01.ibm.com/software/analytics/many-eyes>
- DY Graphs
  <http://dygraphs.com>
- Axis
  <http://www.axis.org>
- VIDI
  <http://www.dataviz.org>
- Google Charts (coding required)
  <https://developers.google.com/chart>
- D3.js (coding required)
  <http://d3js.org>
- Highcharts JS (coding required)
  <http://www.highcharts.com>
- Visualization Toolkit (coding required)
  <http://www.vtk.org>
- VisIt (coding required)
  <https://wci.llnl.gov/simulation/computer-codes/visit>
- Flot Chart (coding required)
  <http://www.flotcharts.org>
- Tangle (coding required)
  <http://worrydream.com/Tangle>
- Gephi (networks, coding required)
  <http://gephi.github.io>

Specific Charts

- Modest Maps (maps, coding required)
  <http://modestmaps.com>
- Dipity (timelines)
  <http://www.dipity.com>
- Tag Crowd (word clouds)
  <http://tagcrowd.com>
- Wordle (word clouds)
  <http://www.wordle.net>
- Gephi (networks, coding required)
  <http://gephi.github.io>

Infographics

- Easel.ly
  <http://easel.ly>
- Venngage
  <http://venngage.com>
- Visme
  <http://www.visme.co>
- Infogr.am
  <http://infogr.am>

Animated and Interactive Charts

- Google Fusion Tables
  <https://support.google.com/fusiontables/answer/2571232>
- Gapminder
  <http://www.gapminder.org>
- Trend Compass
  <http://epicsyst.com/trendcompass/TrendCompass.aspx?home=1>
- Tableau Public
  <http://www.tableau.com/public>
- IBM Many Eyes
  <http://www-01.ibm.com:software:analytics:many-eyes>
- VIDI
  <http://www.dataviz.org>
- Google Charts (coding required)
  <https://developers.google.com/chart>
- Google Motion Charts (coding required)
  <https://developers.google.com/chart/interactive/docs/gallery/motionchart>
- VisIt (coding required)
  <https://wci.llnl.gov/simulation/computer-codes/visit/
- Highcharts JS (coding required)
  <http://www.highcharts.com>

Appendix E. Free Online Courses on Data Visualisation

- Information Visualisation
  <http://ivmooc.cns.iu.edu>
- Data Visualisation and D3.js
  <https://www.udacity.com/course/ud507>
- Tableau Training
  <http://www.tableau.com/learn/training>
- Google Fusion Tables Help
  <https://support.google.com/fusiontables/?hl=en#topic=1652595>
- How to Process, Analyse and Visualise Data
  <http://ocw.mit.edu/resources/res-6-009-how-to-process-analyze-and-visualize-data-january-iap-2012/>
- Scott Murray's D3 tutorials
  <http://alignedleft.com/tutorials/d3>
- Data Visualisation: Theory and Practice
  <http://ocw.usu.edu/instructional-technology-learning-sciences/data-visualization-theory-practice>
- Data Visualisation
  <https//www.coursera.org/course/datavisualization>
- Doing Journalism with Data: First Steps, Skills and Tools
  <http://datajournalismcourse.net/index.php>
- Doing Journalism with Data
  <http://datajournalismcourse.net>

Appendix F. Online Presentations from the Project Workshops

Presentations from Workshop 1: The case for better visualisation and challenges in producing it

<https://reutersinstitute.politics.ox.ac.uk/events/making-research-useful-visualisation>

Presentations from Workshop 2: The current challenges and best practices in policy-oriented and political visualisation

<http://reutersinstitute.politics.ox.ac.uk/events/making-research-useful-visualisation-0>

Presentations from Workshop 3: The current trends in social, relational and human behavioural visualisation

<https://reutersinstitute.politics.ox.ac.uk/events/making-research-useful-visualisation-1>

SELECTED RISJ PUBLICATIONS

Wendy N. Wyatt (ed.)
The Ethics of Journalism: Individual, Institutional and Cultural Influences
(published jointly with I.B.Tauris)

Raymond Kuhn and Rasmus Kleis Nielsen (eds)
Political Journalism in Transition: Western Europe in a Comparative Perspective
(published jointly with I.B.Tauris)

Nigel Bowles, James T. Hamilton, David A. L. Levy (eds)
Transparency in Politics and the Media: Accountability and Open Government
(published jointly with I.B.Tauris)

Julian Petley (ed.)
Media and Public Shaming: Drawing the Boundaries of Disclosure
(published jointly with I.B.Tauris)

James Painter
Poles Apart: The International Reporting of Climate Scepticism

Lara Fielden
Regulating for Trust in Journalism: Standards Regulation in the Age of Blended Media

David A. L. Levy and Robert G. Picard (eds)
Is there a Better Structure for News Providers? The Potential in Charitable and Trust Ownership

David A. L. Levy and Rasmus Kleis Nielsen (eds)
The Changing Business of Journalism and its Implications for Democracy

Tim Gardam and David A. L. Levy (eds)
The Price of Plurality: Choice, Diversity, and Broadcasting Institutions in the Digital Age
published in association with Ofcom

CHALLENGES

John Lloyd and Laura Toogood
Journalism and PR: News Media and Public Relations in the Digital Age
(published jointly with I.B.Tauris)

John Lloyd and Cristina Marconi
Reporting the EU: News, Media and the European Institutions
(published jointly with I.B.Tauris)

James Painter
Climate Change in the Media: Reporting Risk and Uncertainty
(published jointly with I.B.Tauris)

Suzanne Franks
Women and Journalism
(published jointly with I.B.Tauris)

Naomi Sakr
Transformations in Egyptian Journalism
(published jointly with I.B.Tauris)

Nick Fraser
Why Documentaries Matter

Nicola Bruno and Rasmus Kleis Nielsen
Survival is Success: Journalistic Online Start-ups in Western Europe

John Lloyd
Scandal! News International and the Rights of Journalism

Richard Sambrook
Are Foreign Correspondents Redundant? The Changing Face of International News

James Painter
Summoned by Science: Reporting Climate Change at Copenhagen and Beyond

John Kelly
Red Kayaks and Hidden Gold: The Rise, Challenges and Value of Citizen Journalism

Stephen Whittle and Glenda Cooper
Privacy, Probity, and Public Interest

Stephen Coleman, Scott Anthony, and David E Morrison
Public Trust in the News: A Constructivist Study of the Social Life of the News

Nik Gowing
'Skyful of Lies' and Black Swans: The New Tyranny of Shifting Information Power in Crises

Andrew Currah
What's Happening to Our News: An Investigation into the Likely Impact of the Digital Revolution on the Economics of News Publishing in the UK