

# gattoMakingResearchUseful2015

## Making research useful : current challenges and good practices in data visualisation

Info ▾

Bibliography: Gatto, M. (2015) 'Making research useful: current challenges and good practices in data visualisation'. Available at: <https://ora.ox.ac.uk/objects/uuid:526114c2-8266-4dee-b663-351119249fd5> (Accessed: 14 March 2023).

Page-no: 4

PDF: [Full Text PDF](#)

[Open in Zotero](#)

[Open DOI](#)

Abstract ▾

This report advocates for the increased use of data visualisation techniques to illuminate research findings and provides suggestions to overcome some of the challenges academics currently face in using them.

### Notes

Imported on 2023-03-14 16:10

#### Important

- 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. (p. 4)
- So data visualisation does not replace narrative (p. 5)
- Purpose of data viz and its relevance to a non-technical audience:
  - 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. (p. 5)
- The case for the broader use of data visualisation:
  - 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. (p. 5)
  - 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 (p. 5)
- Highlighting the lack of norms despite the widespread use:
  - 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 (p. 5)

- (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).

- Assumptions and projected need:
  - ===([p. 7](#))
- The need for more evidence--based decision--making ([p. 7](#))
- Making case for evidence based decision-making and the use of visual data analysis in this decision-making process:
  - Evidence--based news and data visuals are increasingly expected by media consumers ([p. 8](#))
- 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 ([p. 8](#))
- A note of skepticism and a warning on the use of data visualisation to mislead:
  - 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. ([p. 9](#))
- Data visualisation for reduced cognitive load and increased sensemaking (link to paper on pre-cognitive processing of visual elements):
  - 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) ([p. 9](#))
- Data viz in sense making and story:
  - In sum, data visualisation seeks to simplify the representation of hypotheses, theories, or stories, ([p. 9](#)). #sensemaking , #story
- 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 \(p. 11\)](#)
- Stephen Few ([p. 11](#))
- The case for data viz in visual analysis, sensemaking and insight "instinctive analysis":
  - 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. ([p. 11](#))
- Data viz for actionable insight:
  - (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 ([p. 13](#)).

# 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.

- Summary of the case for data vis:
  - == (p. 15)
- A note about audience requirements and the bias towards certain type of visualisations in academia based on a familiarity with a specific set of software tools:
  - 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). (p. 16)
- EDI element:
  - 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). (p. 16)
- Chen describes the following as the main types of data used in the social sciences: (p. 17).

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

- The issue of software lock in for data visualisation and the barrier it then presents for open sharing of the data sets. (case for open source, open standards):
  - 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 nongovernmental organisations will not ([p.17](#)) #open
- Most of these challenges boil down to one simple truth: academic training and scientific production have often been incompatible with nonacademic demands. We hope that a greater understanding of the potential uses of data visualisation may help to overcome this incompatibility. ([p.18](#))
- Misuse of data visualisation:**
  - 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) ([p.19](#))

## 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*

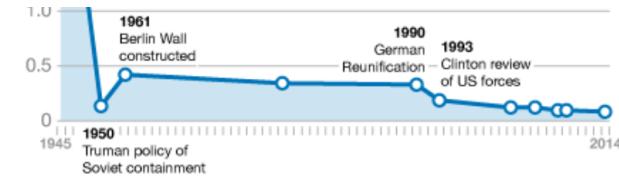
Source: *The Times*, 4 June 2014



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.

- Having care for your reader. Data viz and its reliability (and abuse thereof):
  - ==== ([p. 20](#))

- (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.

- Tufte Rules:

  - ==== ([p. 21](#))

- Design to anticipate problems. I suggest prototype solutions to expose problems.:

  - 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). ([p. 21](#))

- McCandless Information is Beautiful:

  - 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). ([p. 22](#))

- Data visualisation in decision making, insight and gaining a competitive advantage:

  - 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 ([p. 23](#)).

- Overarching Good Practice? ([p. 23](#))

  - (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. ([p. 23](#))

  - Know your audience and understand their needs: tell people what is relevant to them, and, when possible, make it personal.<sup>10</sup> ([p. 23](#))

  - 10 For good practices pertaining to specific types of graphs, visit: <https://www.gov.uk/service-manual/user-centred-design/data-visualisation.html#gds-examples>. ([p. 23](#))

  - (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?<sup>11</sup> ([p. 24](#))

  - (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.<sup>12</sup> ([p. 24](#))

  - (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). ([p. 24](#))

  - (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) ([p. 24](#))

  - (7) Properly label all axes in simple language and ensure that scales are correctly illustrated. ([p. 24](#))

  - (8) Provide clear, interesting titles to graphics. ([p. 24](#))

  - (9) Unclutter: get rid of different fonts, colours, and information that may detract from your main 'story'. ([p. 24](#))

  - 10) Illustrate how data disaggregates throughout different levels of analysis (micro and macro), when possible. ([p. 24](#))

  - 11) Always provide underlying data, as a means of allowing validation and replication. ([p. 24](#))

- Platform supporting data visualisation:

  - Tableau Public ([p. 25](#))

- Resources to allow upskilling in data visualisation:

- 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. [\(p. 26\)](#)

- **Gov usage of existing platforms:**

- 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 [\(p. 26\)](#)

- **No substitute for having a data visualisation specialist on the team:**

- The assistance of IT professionals is also often used in more formal ways. [\(p. 27\)](#)

- **Data visualisation supported by narrative and the input data:**

- 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.) [\(p. 31\)](#)

- **Extensive bibliography:**

- [Bibliography \(p. 35\)](#)

## Query

- **Query. Is there an Irish equivalent of GDS:**

- [GDS \(p. 26\)](#)