

Chapter eleven

Jonathan Gray, Liliana Bounegru, Stefania Milan and Paolo Ciuccarelli

Ways of seeing data: Towards a critical literacy for data visualizations as research objects and research devices

Introduction

‘Every image embodies a way of seeing’, wrote the British art critic John Berger in his 1972 classic *Ways of Seeing* (Berger, 1972, p. 10). Through this book and accompanying television series, he proposed elements of a critical literacy for making sense of the visual landscapes that we inhabit – from reproductions of art historical masterpieces to the advertising which adorns our cities and media environments. As well as guiding the attention of his viewers and readers around specific images, he also sought to examine the way in which images are reproduced and mediated (drawing on the work of the German philosopher and critic Walter Benjamin), as well as the broader social, cultural, economic and political contexts around them.

In the digital age, data visualizations are becoming an increasingly prominent genre for the visual representation and mediation of collective life – from digital analytics dashboards to interactive news graphics. Similarly, data visualizations are becoming more and more popular in media and communications studies, and in the humanities and social sciences more broadly. According to Stephen Few, ‘data visualization is the graphical display of abstract information for two purposes: sense-making (also called data analysis) and communication’ (2014, n.p.). This might include, for example, the representation of information about numbers, words, relations, times or locations. But why have data visualizations become so prevalent? And what does it mean to approach data visualization as a research device? What might it offer media and communications scholars? Data

visualizations promise to assist us in making sense of complex data and complex phenomena, allowing us to simplify and bring order to dense information, for explanatory and communicative purposes. They may help us to analyze, filter, browse and explore complex information. Paraphrasing Marshall McLuhan (1962), data visualizations are said to amplify our senses and our abilities to make sense of the world around us. But data visualizations are not a neutral tool. They come with particular 'ways of seeing', particular analytical, mediation and narrative regimes regarding which we ought to be attentive as we use them to do research and tell stories.

Taking inspiration from Berger's agenda, in this chapter we argue that the use of data visualizations as both instruments and objects of study requires commensurate forms of critical literacy. We also draw on Philip Agre's (1997) notion of 'critical technical practice' for the social study of artificial intelligence and from Bernhard Rieder and Theo Röhle's (2012) notion of 'methodological reflexivity' in relation to digital methods in the humanities and social sciences, both of which gesture towards forms of engagement with new research methods and technologies that are accompanied by critical reflection on their uses. We seek to extend their proposition that research practices ought to 'oscillate [...] between concrete technical work and methodological reflexivity' (Rieder and Röhle, 2012, p. 80) to the practice of working with data visualizations.

We argue that data visualizations are not only bright adornments to our twenty first century information environments. They embody and engender not only particular ways of *seeing* (as Berger comments), but also ways of *knowing* and ways of *organizing* collective life in our digital age. In other words, data visualizations reflect and articulate their own particular modes of rationality, epistemology, politics, culture and experience. It is precisely to emphasize these 'world-making' capacities, that in this chapter we prefer the term 'device' to the more commonly utilized term 'tool' to refer to data visualization. While the term 'tool'

establishes the object as possessing coherence and connotes an instrumental relationship between user and object, this chapter proposes to develop an account of data visualizations as devices in order to draw attention to their capacities to ‘assemble and arrange the world in specific social and material patterns’ (Law and Ruppert, 2013, p. 230). While arriving to media and communication studies with particular inscriptions as described above, following Law and Ruppert (2013) we argue that data visualizations as device are at the same time indeterminate and open to multiple and diverse forms of usage. For this reason, to follow Marres and Gerlitz (2015), deploying data visualization in media and communication research may be thought of as a process of developing mechanisms to align its affordances (and limitations) with the discipline-specific conventions of our fields. Given the growing role of data visualizations in our information environments, we think it is vital to develop a critical literacy to read, understand and work with them.

We propose a three part heuristic framework for what should be taken into account when reading, working with and conducting research about data visualizations. In doing so this chapter does not aim to develop a practical guide for the effective use of visualization in media and communication research, *per se*. Instead we propose a framework to sensitize researchers to forms of mediation embedded in data visualization as research devices in order to support their critical and reflexive use.

This framework is organized around three forms of mediation that can be studied in relation to data visualizations: (i) the mediation from *world to data* of the sources of information that underpin visualizations; (ii) the mediation from *data to image* of the graphical representations of this information; and (iii) the mediation from *image to eye* and the mind in the socially, culturally and historically specific ‘ways of seeing’ engendered in the data visualization. Each of these three forms of mediation can be studied with a broad range of methods – from more familiar qualitative and quantitative approaches (such as visual and

textual analysis, interviews or surveys), to emerging digital and computational methods. We describe and illustrate these different forms of mediation, and ways of studying them, with reference to a collection of over two hundred data visualizations about public finances.

While for heuristic purposes our framework proposes the study of mediations between 'world', 'data', 'image' and 'eye', our intention is to provide a starting point to inform and broaden inquiry rather than to propose a neat and rigid distinction between these different elements. They are in fact mutually constitutive such that data constitute as much as they represent the societal dynamics which they measure (Espeland and Stevens, 2008) and that regimes of measurement and visualization are generative of specific publics, practices and cultures (Ruppert, 2015). The forms of mediation which we propose are also not exhaustive. For example, drawing on research on the reactivity of metrics (Espeland and Sauder, 2007; Gerlitz and Lury, 2014), a fourth layer of mediation could be formulated, concerned with the study of actions to intervene, respond to and modify the dynamics captured by data visualizations. In addition to this, as the following sections will show, each of the three forms of mediation is actually constituted by multiple sub-layers of mediation or inscription. These will vary according to the nature of the 'research apparatus' in which visualizations are embedded, in that different types of data and methods are accompanied by different types of inscription (Ruppert, 2013). We also do not hold a strict order with regard to the proposed layers of mediation.

The past few decades have seen the development of a body of literature dedicated to data visualization and information graphics. In particular there is a growing body of books and articles offering practical guidance as well as showcasing different examples,

techniques and approaches.¹ This literature is itself interesting not only from a practical perspective, but also as a way to understand the forms of mediation involved in the composition of data visualizations – both those that receive attention and those that remain neglected – as well as the aesthetics, cultures, values, ideals and practices associated with their production. These resources can be useful as a source to disassemble and understand the making of data visualization projects.

¹ See, for example: Cairo, 2012; Card et al., 1999; Cleveland, 1993, 1994; Few, 2009, 2012, 2013; Heller and Landers, 2014; Katz, 2012; Krum, 2013; Lima, 2011; McCandless, 2012, 2014; Meirelles, 2013; Munzner, 2014; Murray, 2013; Rendgen, 2012, 2014; Spence, 2014; Steele and Iliinsky, 2010; Tactical Technology Collective, 2014; Tufte, 1990, 1997ab, 2001; Wong, 2013; Yau, 2011, 2013.

While many of these works focus on a single layer of mediation – that of data to the image, looking at how information is translated into graphical form – we suggest that in using and studying data visualizations it is essential to grasp not just the production of images from data but also the datasets and data infrastructures that data visualizations draw on, as well as the cultural practices and ideals implicated in the composition of visualizations which invite a particular way of seeing. There have been a number of very informative works on this latter topic (for example Drucker, 2014; Halpern, 2015), which we will explore further below. Here our focus is less on data visualization as a field, but rather on developing the elements of a critical reflexivity that would accompany and inform the practice of data visualization in research and other contexts. This critical reflexivity is important not only in the study of data visualizations as objects, but also to improve our abilities to deploy them as research devices. Below we outline research outlooks and methods for studying all three forms of mediation – as well as pointing to further resources that may be useful for each one.

An Example: Visualizing Information about Public Finances

To illustrate our proposed framework we have chosen to work with a collection of data visualization projects about public finances. These include data visualization projects from media organizations, journalists, civil society organizations and public institutions. This collection has been gathered in the context of research to empirically map how information about public finances is used in the service of democratic engagement with fiscal policy (Gray, 2015ab).

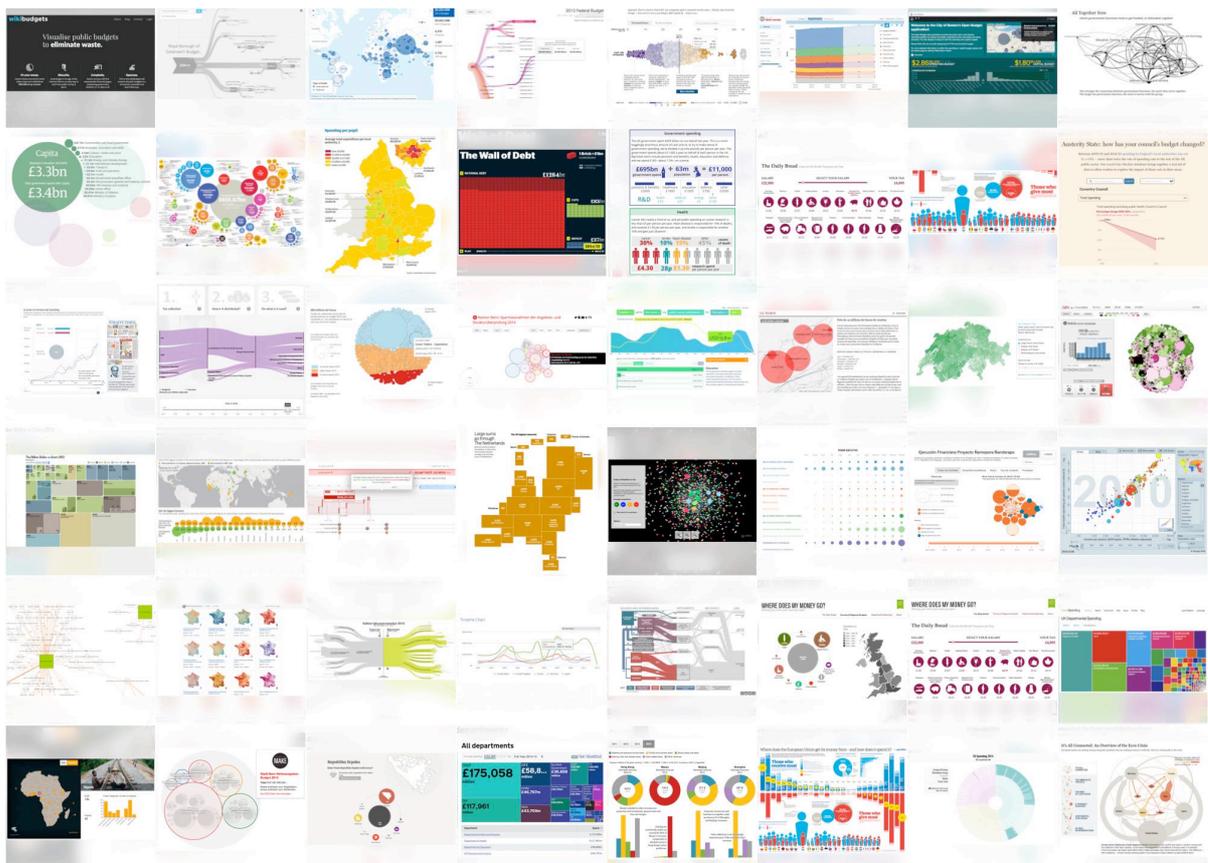


Figure 11.1: A selection from collection of examples of fiscal data visualizations (Gray, 2015b).

There are two main reasons why we consider this thematic focus suitable to make our case.

Firstly, there is a long tradition of work exploring public finances with information graphics.

Edward Tufte uses public finances to illustrate discussion of different techniques in his

classic *The Visual Display of Quantitative Information* citing a venerable tradition of

information graphics which ‘nearly always create the impression that spending and debt are

rapidly increasing’ (Tufte, 2001, p. 65). He alludes to the fiscal information graphics of the

Scottish engineer, economist and pioneer of statistical graphics William Playfair, such as this

one:

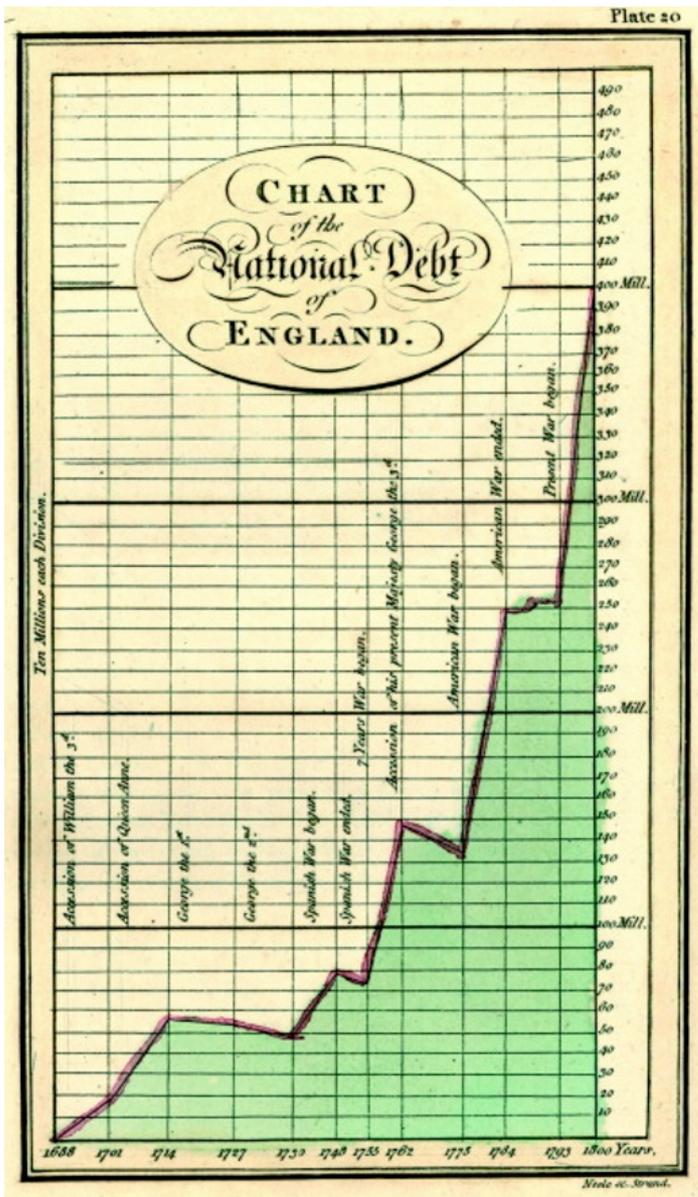


Figure 11.2: 'Chart of the National Debt of England' from William Playfair's *The Commercial and Political Atlas* (Playfair, 1801).

The second reason is that there is a heterogeneous constellation of different issues and concerns that are associated with public finances and fiscal policy. As well as political and economic questions from who in society pays how much tax and how public resources are allocated – many policy areas are underpinned by discussions about public finances, from international development to climate change. The complexity and competing narratives

around this topic makes it well suited to illustrate different approaches for studying data visualizations.

Three Forms of Mediation and How to Study Them

In this section we will propose a research outlook and methods for studying each of the three forms of mediation implicated in the creation of data visualizations that we have outlined. This may serve as a checklist of questions and a menu of different methods that can be used.

1. From World to Data

As we shall see in the next section, many classic works on information visualization focus on the mediation of data to image – in particular focusing on the avoidance of misrepresentation. For example, Tufte suggests that ‘graphical excellence requires telling the truth about data’ (Tufte, 2001, p. 51) and that data visualization designers should ‘let the data speak for itself’ (1997a, p. 45), championing an aesthetic program of ‘graphical integrity’ that we shall challenge and explore further below. But what about the data itself? How does the making of data shape the making of knowledge through visualizations?

Our first form of mediation to be studied is how the data used in the visualization is generated and how this process inscribes itself in the knowledge produced through data visualizations. This might include asking questions such as:

- What information or data is being represented in the visualization?
- What are the sources for this information? Where does the data come from?
- How is the data generated? What are the rationales, methods and standards

inscribed in the data infrastructures through which the data is generated?

- How is the data transformed or prepared?

- Which data sources are combined and how?
- How does the data selectively prioritize certain things over others?

The first step will often be establishing the data sources. In some cases details about the sources of the information will be explicitly referenced or linked to. In other cases additional work might be needed in order to identify these sources. This might be matching the precise tables or datasets that are used from a number of possible contenders (for example, in cases where the institution or database is given, but not the exact table or dataset); looking in the software source code or documentation of the visualization to look for data sources; or conducting interviews with the creators of a data visualization to establish what data was used.

Once datasets are established for a given visualization, there are different approaches to studying and analyzing them. For example, the study of data sources might draw on literature about ‘sourcing practices’ from media studies and journalism studies (Cottle, 2003; Manning, 2001; Hall et al., 1978; Berkowitz, 2009). Research can also be undertaken on the ‘data infrastructures’ implicated in the production of the datasets that are used in the visualizations. Drawing on previous work in this area, we take the phrase ‘data infrastructure’ to designate socio-technical systems implicated in the creation, processing and distribution of data (see, for example, Akrich, 1992). This might include elements such as standards bodies, software systems, administrative procedures, committees, consultancy processes and many other things (Gray and Davies, 2015).² Here it may be useful to draw on approaches from science and technology studies (STS) to trace the politics embedded in these systems – for example through what has been called ‘infrastructure ethnography’ (see, for example, Star and Ruhleder, 1996; Star, 1999; Bowker and Star, 2000). The composition

² Gray and Bounegru are currently working on another project on ‘data infrastructure literacy’, which includes further suggestions on approaches to studying and working with data infrastructures.

of data infrastructures will lead to the production of different types of data – from statistical data to transactional ‘digital traces’ extracted from digital platforms.

How might we operationalise these approaches in relation to our collection of data visualizations about public money? Firstly, can we *identify* the datasets that are used in the visualizations? The different ways of citing or linking to data sources is itself something that can be studied – as it may reflect different kinds of ideals, norms or practices of knowledge production. A cursory look at different data sourcing practices reveals a wide range of ways in which data is obtained and prepared, as well as varying approaches to publishing details about datasets, software and methodology.

	A	B	C	D	E	F	G	H
	OFFICIAL		source	UNOFFICIAL		source	source	
1	42 bn	tax gap -the difference between tax collected and tax that should in theory or in law be collected.	Guardian	102 bn		http://www.taxresearch.org.uk	http://www.guardian.co.uk/society/2010/sep/17/liberal-democrats-clamp	
2								
3								
4	5.8	Inaccurate self-assessment returns from individuals	HMRC				http://www.telegraph.co.uk/finance/personalfinance/capital-gains-tax/78	
5	0.8	Inaccurate self-assessment returns from large partnerships	HMRC				http://www.guardian.co.uk/business/2010/mar/25/budget-2010-tax-avoid	
6	1.2	Inaccurate returns from small and medium businesses	HMRC				http://www.taxresearch.org.uk/Blog/2008/02/01/25-billion-the-cost-of-tax	
7	2	Inaccurate self-assessment returns from large businesses	HMRC				http://www.taxresearch.org.uk/Blog/2008/02/01/25-billion-the-cost-of-tax	
8	1.4	Tax avoidance by individuals & businesses	HMRC				http://www.hmrc.gov.uk/stats/measuring-tax-gaps-2010.htm.pdf	
9	0.3	Non-declaration of income	HMRC				http://www.hmrc.gov.uk/stats/measuring-tax-gaps-2010.htm.pdf	
10	3.2	Undeclared personal earnings & hidden cash economy	HMRC				http://www.hmrc.gov.uk/stats/measuring-tax-gaps-2010.htm.pdf	
11								
12	6.9	Missing Corporation Tax	HMRC				http://www.hmrc.gov.uk/stats/measuring-tax-gaps-2010.htm.pdf	
13								
14	0.1	Inheritance Tax	HMRC				http://www.hmrc.gov.uk/stats/measuring-tax-gaps-2010.htm.pdf	
15	0.8	Stamp Duties	HMRC				http://www.hmrc.gov.uk/stats/measuring-tax-gaps-2010.htm.pdf	
16	0.2	Petroleum Revenue Tax	HMRC				http://www.hmrc.gov.uk/stats/measuring-tax-gaps-2010.htm.pdf	
17								
18								
19	2	Illegal cigarette & spirits market	HMRC					
20	0.7	Illegal diesel market	HMRC				http://www.hmrc.gov.uk/stats/measuring-tax-gaps-2010.htm.pdf	
21	15.2	VAT gap	HMRC					
22	0.9	Other indirect taxes	HMRC					
23				1	lost to personal tax avoidance using capital gains tax	Daily Telegraph		
24				0.8	lost to overseas tax havens	Guardian		
25				3.2	Income Shifting - moving income to spouses or private companies		http://www.hmrc.gov.uk/stats/measuring-tax-gaps-2010.htm.pdf	
26				3	Domicile tax Avoidance	http://www.taxresearch.org.uk		
27				12.1	Corporate Tax Avoidance	http://www.taxresearch.org.uk	http://www.hmrc.gov.uk/stats/measuring-tax-gaps-2010.htm.pdf	
28				1.1	Capital Gains shifting	http://www.taxresearch.org.uk	http://www.hmrc.gov.uk/stats/measuring-tax-gaps-2010.htm.pdf	
29						http://www.taxresearch.org.uk		
30				1.8	National Insurance avoidance	http://www.taxresearch.org.uk		
31				3	Other tax planning e.g. Inheritance taxes, stamp duty etc.	http://www.taxresearch.org.uk		
32					Based on 37% tax not paid on the World Bank-estimated £197bn shadow economy	http://www.taxresearch.org.uk		
33								
34								
35	41.5	TOTAL		61				

Figure 11.3: Data sources for ‘The Tax Gap’ visualization from the *Guardian Datablog* and Information is Beautiful.

In our collection, some of the projects simply cite the name of the department from which the data was derived. One example from the *New York Times* simply says ‘Source: Office of

Management and Budget’, with no further details.³ Many others give much more detail – including directly linking to datasets, or even providing documentation on methodology about how the data was sourced and transformed, what assumptions were made, and so on. For example a piece from the US investigative news outlet *ProPublica* on ‘The Millions New York Counties Coulda Got’ is accompanied by a short methodological walkthrough titled ‘How We Analyzed New York County Tobacco Bonds’ explaining how they obtained the data, which the cash flow models that they used and why, and the rationales behind other assumptions they made in the application.⁴

Publishing data sources has become *de rigueur* amongst some data journalism outlets. A piece from the *South China Morning Post* explains how data was extracted from PDF files and provides links to the ‘raw data’ used in the visualizations.⁵ One data visualization from the *Guardian Datablog* gives an accompanying spreadsheet which lists a separate source URL for every figure that they have used.⁶ If visualization projects publish the source code of their software, this can be studied in order to see the data files which have been used. For example, the Budget Key project by Israeli non-profit Public Knowledge Workshop link to their source code on the social software sharing platform GitHub which makes visible the changes and transformations that have been made on the data.⁷

As well as looking at the sourcing and transformation of the data, we might also look at the selection of particular datasets and tables to highlight or narrate certain aspects of public

³ http://www.nytimes.com/interactive/2010/02/02/us/politics/20100201-budget-porcupine-graphic.html?_r=0 (accessed 10 March 2016).

⁴ See <https://projects.propublica.org/graphics/ny-millions> and <https://www.propublica.org/article/how-we-analyzed-new-york-county-tobacco-bonds> (accessed 10 March 2016).

⁵ <http://widgets.scmp.com/infographic/20140304/budget2014/data/> (accessed 10 March 2016).

⁶ <http://www.theguardian.com/news/datablog/2010/sep/22/tax-gap-information-beautiful> (accessed 10 March 2016).

⁷ <https://github.com/OpenBudget/> (accessed 10 March 2016).

finances such as spending, revenue or debt. Rather than simply ‘telling the truth’ about public finances – these selections emphasize and de-emphasize different aspects of fiscal policy. We shall further examine the affordances of different visual forms in the next section. Suffice to say here that the selection of different datasets and tables is an important step that should be taken into account in the study of data visualizations – prior to their translation into graphical form. We can compare the data mediated in the visualization to the other data which is made available. Which indicators, subtables or items of data have been selected, which have been left out and how does this guide our attention towards some things and not others?

An interactive news application from *The Times* (UK) called ‘The Wall of Debt’ shows a dominant wall of red bricks depicting ‘national debt’, and a comparatively small green wall called ‘cuts’.⁸ The visual editorial decision to select these two items is an important one in reading this interactive graphic, engaging with a particular political economic narrative about tackling public debt through spending cuts. Another short video clip called ‘Debris’ by David McCandless presents a wide variety of other fiscal data points, ultimately highlighting how the global cost of the credit crisis dwarfs other sums such as the Organization of the Petroleum Exporting Countries’ climate change fund, the budget of the UN, African debt to the West and the funds needed to ‘save the Amazon rainforest’.⁹ As well as looking at the selection of data on a case by case basis for individual visualizations, comparative analysis may be undertaken across a larger collection.

We might also study the composition of the data itself. We could study the headers, categories and classifications within the datasets to understand which variables and indicators are selected and prioritized. Content analysis of documentation and associated

⁸ <http://appliedworks.co.uk/work/the-times-defining-a-new-era-of-data-journalism/> (accessed 10 March 2016).

⁹ <https://www.youtube.com/watch?v=K7Pahd2X-eE> (accessed 10 March 2016).

materials may provide further details about the units, forms of analysis, methodology, software, standards and other details about how the data was generated. This might also be supplemented by interviews with those involved in the creation of the data about their decisions. It may be that datasets are generated from other more complex database systems which may be of interest to study. In the case of public financial data, we might look into politics, rationales and ways of knowing inscribed into the financial management information systems through which datasets about public finance are generated. These may be studied as socio-technical systems through, for example, document analysis and interviews. Sometimes there may be documentation manuals which explain how these systems function. In the case of public sector data infrastructures, if details are not already published, then there may be routes for formally requesting them through access to information laws. Researchers and civil society groups interested in studying what was the most detailed source of information about public spending in the UK (the Combined Online Information System, or COINS) submitted official ‘freedom of information’ requests about the database. While their requests for its contents were initially turned down, they submitted follow up requests asking about the database and training materials for the database, which were successful.¹⁰

It may also be fruitful to study the categories and classification systems articulated by the data. For example, any of the data visualizations about public money use different categories to describe different areas of expenditure. Where do these categories come from? How did they become the way that they are? On the one hand we can look at the administrative contexts in which the datasets are generated – such as the organization of public sector bodies. We might also look at which kinds of data standards shape public sector data systems. For example, the UN COFOG (Classifications of Functions of

¹⁰ https://www.whatdotheyknow.com/request/schema_for_coins_database (accessed 10 March 2016).

Government) standard is widely used as a reference in order to facilitate comparability between different national budgeting processes.¹¹ This is particularly relevant when it comes to looking at visualizations including multiple countries – such as international development or climate finance. The genesis of these data standards can also be studied through analysis of (historical) document collections, the study of information systems (for example, using approaches from software or platform studies), interviews and ethnographic studies.

2. From Data to Image

The second form of mediation in our heuristic framework is how visualizations mediate the data sources they draw on into graphical form. This might include addressing questions such as:

- How is the data mediated into graphical form?
- What kinds of graphical techniques, methods and technologies have been used?
- What are their affordances? How do they guide our attention towards different

aspects of the data?

- What design decisions have been taken? What are their consequences?

These concerns are prominent in both research and practical literature around data visualizations. Edward Tufte talks of ‘graphical methods that *organize and order the flow of graphical information* presented to the eye’ (Tufte, 2001, p. 154). He places a significant premium on ‘graphical integrity’ of information visualizations such that they ‘defeat graphical distortion and ambiguity’ (Tufte, 2001, p. 77). While this is one important way of looking at how data is mediated into graphical form, there are many other important aspects to study. Rather than just looking at accuracy, fidelity, and how graphics may be truthful or untruthful about data – we can also step back and look at their affordances – such as how they

¹¹ <http://unstats.un.org/unsd/cr/registry/regcst.asp?CI=4> (accessed 10 March 2016).

articulate structures and relationships or how they organize space, time, quantity and categories in relation to the data.

In order to study these things we may learn a great deal from contemporary literature about the creation of data visualizations, as well as classic literature on information graphics. These can be used in order to understand the composition of data visualizations, or to ‘reverse engineer’ them, which helps us to develop a critical and reflexive approach to the use of data visualization as a research device. Jacques Bertin’s 1967 *Semiology of Graphics* proposes a series of ‘retinal variables’ (1983, p. 9) including size, value, texture, color, orientation, shape as a starting point for analyzing and working with different types of information graphics. This work provides an extensive and richly illustrated overview of how different ‘components’ of information can be mapped onto different visual forms. In a similar vein, Tufte’s *The Visual Display of Quantitative Information* contends that ‘data graphics visually display measured quantities by means of the combined use of points, lines, a coordinate system, numbers, symbols, words, shading and color’ (2001, p. 9).

While Bertin’s work is explicitly limited to print graphics which fit on a ‘sheet of white paper’ (1983, p. 42) and many of Tufte’s seminal works also focus on print graphics, many of the elements they discuss have been adopted and developed in relation to digital and interactive data visualizations. Thus we may read these influential works alongside other more recent works that cover digital tools and methods in order to understand how different visual forms are implicated in the visualizations under study – including table graphics, maps, timelines, sparklines, networks, graphs, flow diagrams, small multiples, motion charts, bubble charts and treemaps. These different forms can also be combined to highlight different aspects of the data. We may also look at the affordances of the software or the platforms that were used to create data visualizations (see, for example, Wright, 2008; Manovich, 2002, 2011, 2014). As well as desktop software applications, there is a growing

number of visualization tools which enable visualizations to be generated and embedded.

There is also a growing number of software libraries and components which are widely used to translate data into different graphical forms.

How might we use these approaches to study our collection of data visualizations about public money? We could look at the different graphical forms used in data visualizations to organize attention around different aspects of public finances. For our collection, this could include examining how different graphical elements are used in order to:

- Show a 'bigger picture' of breakdowns of totals into different categories;
- Put different figures into context;
- Show the geographical distribution of funds;
- Show trends or developments over time;
- Show breakdown of funds by sector or recipient;
- Show networks of financial flows;
- Compare different parts of the budgeting cycle (for example, commitments and actual expenditure);
- Compare revenues, expenditures and debts;
- Show allocations per capita.

The formal characteristics of different visualizations in a collection can be studied with reference to either a pre-defined vocabulary of elements or through an 'open coding' or 'emergent coding' process (see Strauss & Corbin, 1998). As similar visual elements can play different roles in different contexts, it is crucial to note not only their presence or absence, but their *relations* with other elements in the visualization as a whole. For example, compare the use of bubbles in the following visualizations:

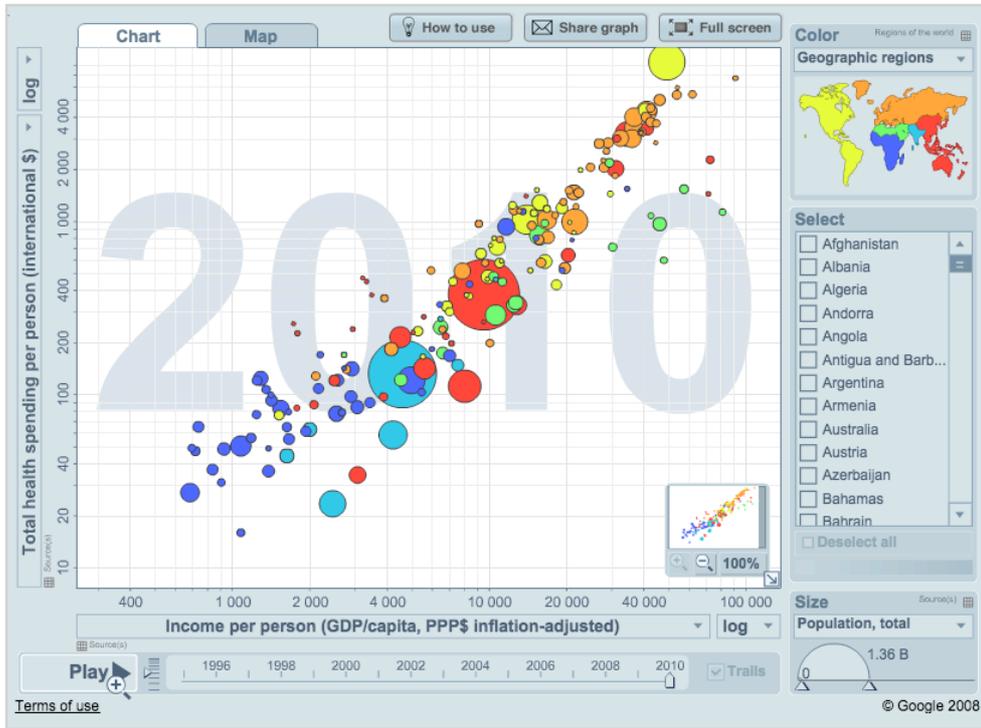


Figure 11.6: Health spending per person compared with income per person over time (Gapminder World); <http://www.gapminder.org/world/> (accessed 10 March 2016).

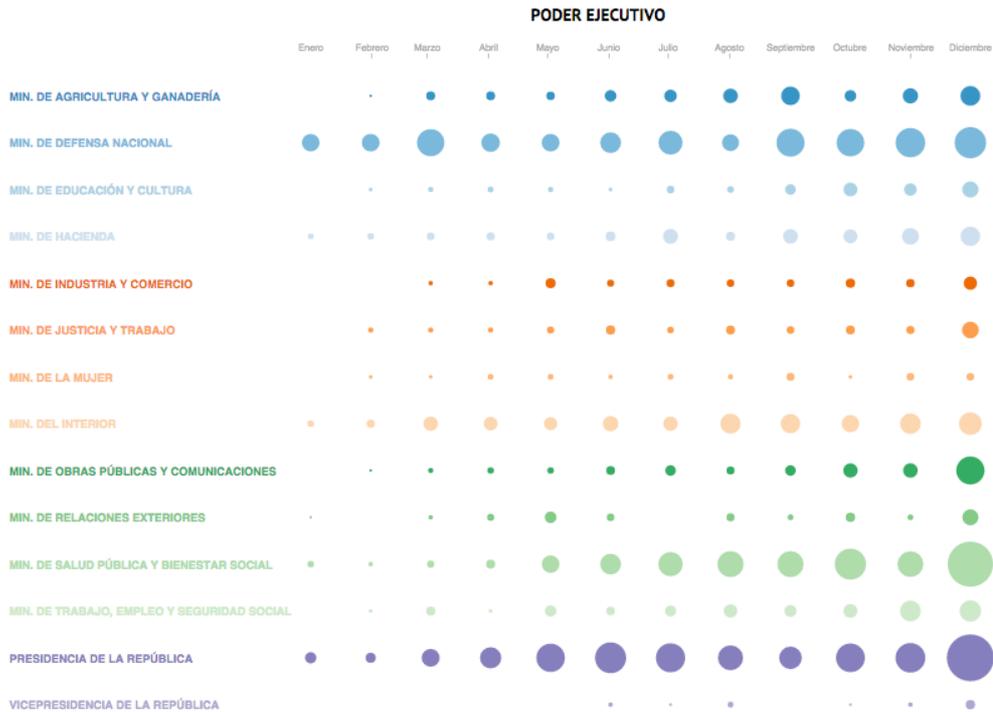


Figure 11.7: Spending by different government departments over time (Government of Paraguay); <https://www.contrataciones.gov.py/datos/visualizaciones/contratos> (accessed 10 March 2016).

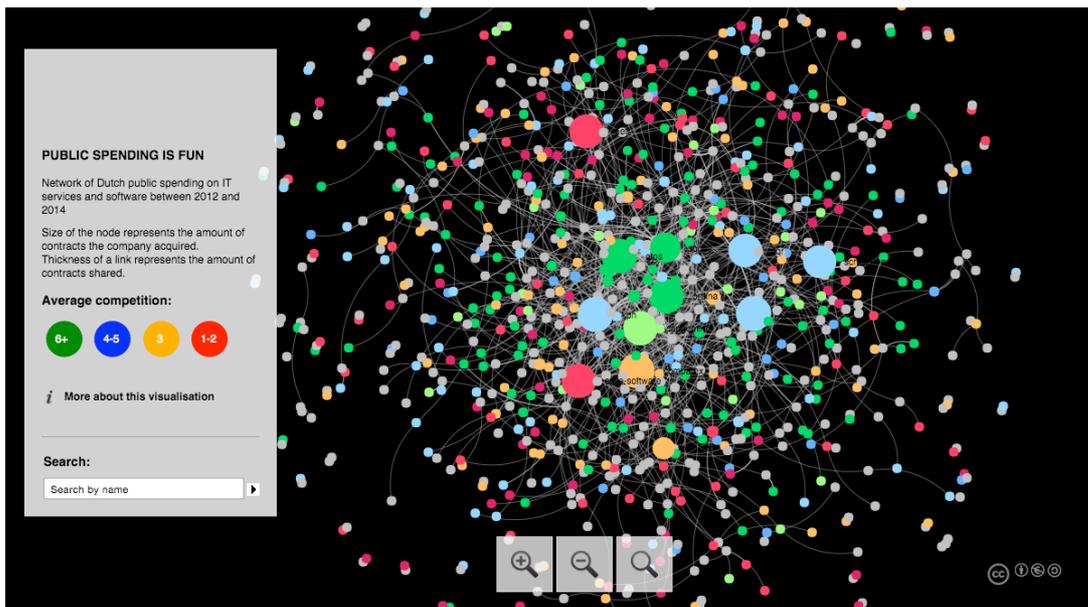


Figure 11.8: Network of Dutch public spending on IT services and software (Adriana Homolova); <http://www.homolova.sk/dh/it/#> (accessed 10 March 2016).

תקציב המדינה לשנת 2014 הוא 435 מיליארד ₪
 מה הם השינויים שחלו בו מתחילת השנה?
 162 התחומים שגדלו מופיעים בצבעים חמים, 48 אלה שקטנו מופיעים בצבעים קרים

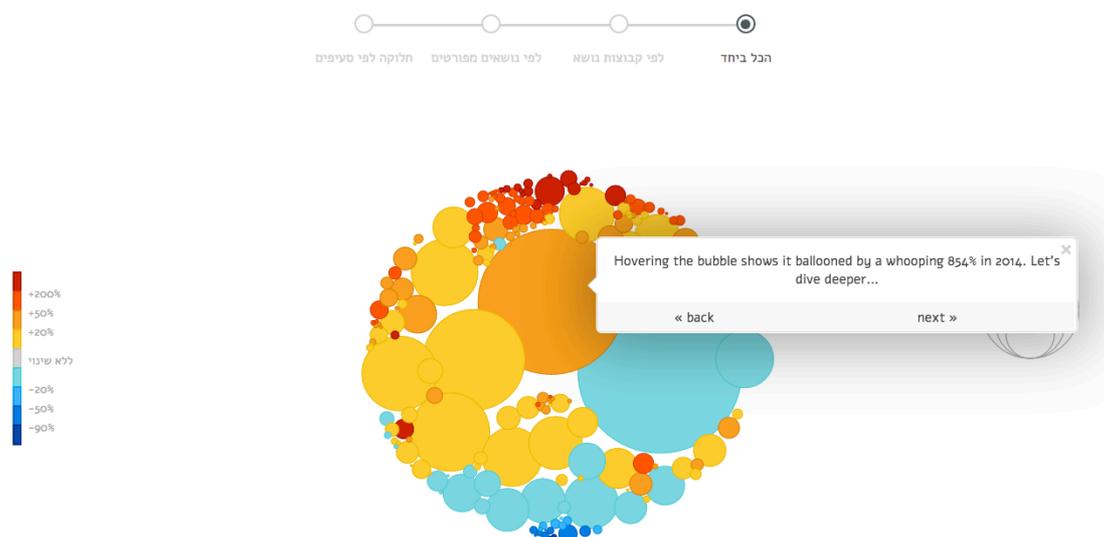


Figure 11.9: Comparing commitments and spending in Israeli budget (Budget Key);

<http://www.obudget.org> (accessed 10 March 2016).

With these examples we can see how the bubbles are arranged, scaled and colored in order to draw out different characteristics of public finances – from representing hierarchical relationships in national budgets, to the locations of development funds, to health spending per capita over time, to spending per department over time, to relationships between recipients of public contracts, to the differences between commitments and spending. In each case, the bubbles are used in different ways to represent different things – with the position, size and color having varying significance for each graphic. These different ways of organizing visual phenomena thus articulate and emphasize different ways of organizing knowledge about public finances.

As well as studying the characteristics of visualizations, their creators could be interviewed regarding their techniques, methods, software tools and design choices. Where software repositories are available these can also be analyzed in order to understand how

data sources are being mediated into graphics – whether for individual repositories or for collections. We might, for example, study repositories associated with fiscal data visualization projects on GitHub.¹²

3. From Image to Eye

The third and final form of mediation in our heuristic framework is how the image appears to the eye and to the mind. As we saw in the last section, visualizations emphasize different aspects of data through design choices, techniques and software which organize data into graphical form. But these graphical forms also engender and depend on different socially, culturally and historically contingent *ways of seeing data*. Hence we might ask:

- What kinds of visual cultures and practices are implicated or reflected in the data visualization? Where do these come from?
- What forms of usage are inscribed in the visualization? Who are the publics of the data visualization? How is it circulated, cited and shared?

Rather than seeing the visual forms as neutral instruments for making evidence visible – or as Tufte puts it, ‘instruments for reasoning about quantitative information’ (Tufte, 2001, p. 9) – we can study their genealogies, aesthetics and epistemological affordances to situate them in relation to other ideals, values and practices. For example, visualizations can be read against the background of histories of science, technology and modernity which explore the relationship between vision, knowledge and image making practices – from accounts of occidental ‘ocularcentricism’ (see Jay, 1988) to the role of visual cultures in the development of conceptions of objectivity (see Daston and Galison, 2010).

¹² Bounegru and Gray are currently involved in developing tools and methods for working with metadata from GitHub as part of other research projects.

There have been several recent works which make the case for drawing on hermeneutical approaches from the humanities to enhance the study of data visualizations. For example, in her book *Graphesis* Johanna Drucker advocates ‘critical study of visuality from a humanistic perspective’ in order to ‘de-naturalize the increasingly familiar interface that has become so habitual in daily use’ (Drucker, 2014, p. 9–10). She explores the emergence of contemporary ‘visual epistemology’ with reference to a broad range of developments in art, architecture, design, industry, philosophy and computer science – from the emergence of graphical design to interdisciplinary deliberations about visual abstraction between artists and designers associated with the Bauhaus school in Weimar Germany. In her *Beautiful Data*, Orit Halpern also similarly aims to ‘denaturalize and historically situate’ ideals and practices of data visualization (Halpern, 2015). She traces the genesis of what she calls ‘communicative objectivity’ that has come to be associated with data visualizations drawing on a different disciplinary constellation focusing on cybernetics, communication science, behavioral science, engineering, management studies, urban planning and military research. There is also a growing body of literature focusing on the aesthetics of data visualization (see Manovich, 2002; Jevbratt, 2004; Lau and Vande Moere, 2007; Whitelaw, 2008; Sack, 2011; Cubitt, 2015) as well as on the development of different visual forms such as the timeline (see Rosenberg and Grafton, 2013) or those associated with network analysis (see Freeman, 2004). The narrative dimension of data visualizations can also be studied (see Segel and Heer, 2010; Venturini et al., 2016).

How might these kinds of approaches be adopted to study specific data visualizations – such as our collection about public finances? We could study the aesthetics of these visualizations – for example, the clean, minimalistic style adorned with primary color palettes and icons that has become so widely adopted in many contemporary information graphics.

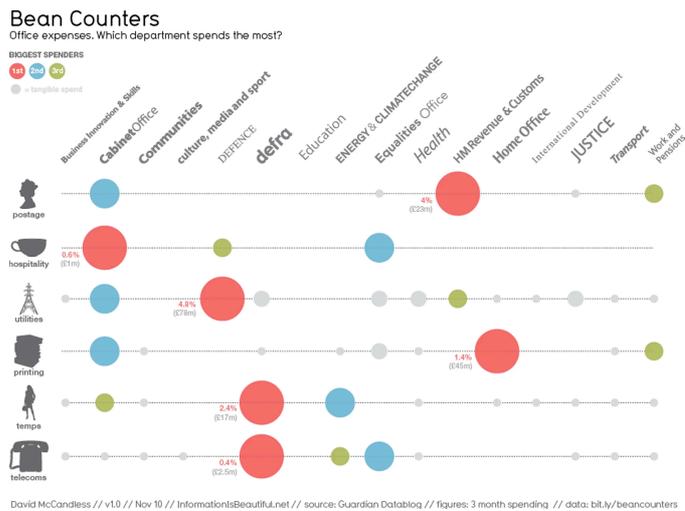


Figure 11.10: Government spending over £25,000 (The Guardian);

<http://www.theguardian.com/news/datablog/2010/nov/19/government-spending-information-beautiful> (accessed 10 March 2016).

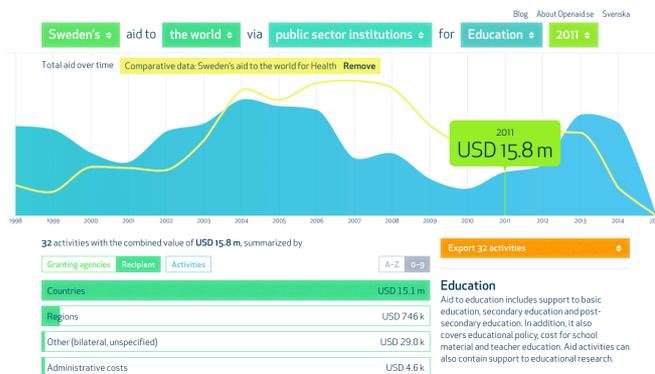


Figure 11.11: Openaid.se (Swedish International Development Cooperation Agency);

<http://www.openaid.se/aid/2014/> (accessed 10 March 2016).

This is similar to the aesthetic that is advocated by Tufte: championing efficiency and parsimony, maximizing the 'data-ink' ratio and eliminating 'chartjunk'. He proposes that 'the design of statistical graphics is a universal matter [...] like mathematics' and that insight into

the design of visualizations may be obtained through the study of ‘excellence in art, architecture and prose’ (Tufte, 2001, p. 10). Peter Galison (1990) has previously studied the links between architectural modernism and the universal aspirations of logic, mathematics and philosophy in the first few decades of the twentieth century. Many of the fiscal data visualizations in our collection look to share this aesthetic. Several of them incorporate icons or pictograms reminiscent of those associated with the Isotype Institute of Marie and Otto Neurath, which has exercised an important influence on the contemporary aesthetics of data visualization (Zambrano and Engelhardt, 2008; Mayr and Schreder, 2014; Headrick, 2000; Rayward, 2008).

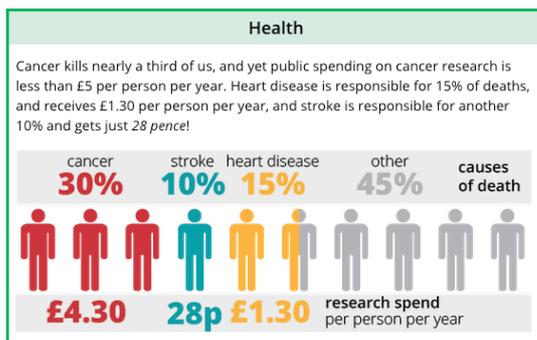


Figure 11.12: Science spending in the UK (Scienceogram);

<http://scienceogram.org/summary/> (accessed 10 March 2016).



Figure 11.13: 'Home and Factory Weaving in England' (Neurath, 1939).

Genealogical study may help to enrich research about what has given rise to the contingent forms that data visualizations may take – identifying a range of different influences and origins of different visual forms. Starting points may be provided through interviews with designers as well as content analysis of relevant design materials – which may be complemented with historical texts and archival research.

We might also study the specific visual forms which are transposed to mediate public money. Many visualizations are described and/or organized as dashboards – giving users an overview of multiple key indicators and trends over time in a single viewing pane, inviting narratives of oversight, optimization, balance and control (see Tkacz, 2015).

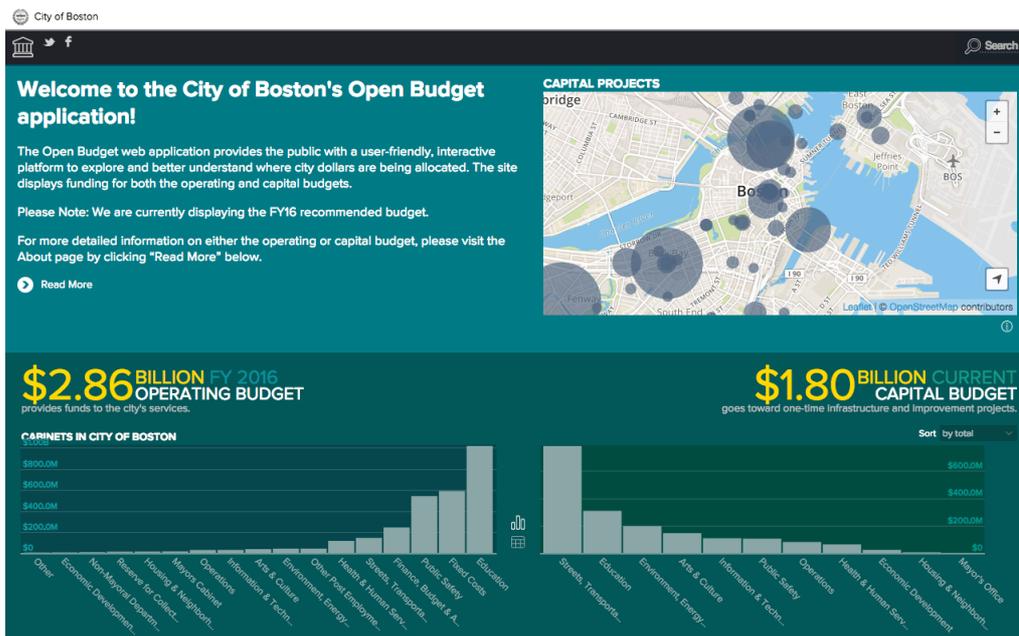


Figure 11.14: City of Boston's Open Budget Application;
<http://budget.data.cityofboston.gov/#/> (accessed 10 March 2016).

As well as studying data visualizations as cultural forms from a humanistic perspective, we can also trace their circulation, reception and how they are used and viewed by different publics (see Kennedy et al., 2016). In addition to interviews and workshops, digital methods could be utilized to extract and analyze traces from digital platforms and online spaces in order to review the contexts in which visualizations are being used and shared (Rogers, 2013). For example, we could query for names or URLs of projects on social media, search engine results, news media, or collections of documents. This may also help to chart the particular publics and context of usage of the visualizations – which can assist with their study as social and cultural forms.

Conclusion

In this chapter we have proposed a heuristic framework that may be used to develop critical reflexivity around the use of data visualization as an object of research or a communicative

or analytical device in research. We have illustrated this framework with reference to a collection of data visualizations about public money, and suggested some methods and approaches that could be used to study them.

Firstly we looked at the study of data sources, focusing on research approaches to examine how they selectively articulate and mediate different aspects of the world – including how to identify sources, tracing how they have been transformed and studying data infrastructures. Secondly we looked at ways to study how the data sources are mediated into graphical form – including through the analysis of their visual properties, software and design choices. Thirdly and finally, we looked at ways of studying graphical forms as socially, culturally and historically contingent forms engendering different ways of seeing – including by tracing their diverse influences, and by analyzing their circulation and contexts of usage.

We hope that the study of these three forms of mediation through some of the approaches that we have discussed in this chapter may provide a useful starting point for researchers who wish to use data visualization as a research object or device. As data visualizations become more and more central, prominent and familiar as ways of knowing and organizing phenomena, we think it becomes imperative to develop a richer understanding of the ways of seeing and ways of knowing that they engender. Just as Berger's *Ways of Seeing* helped to advance broader awareness of the critical study of images and visual culture, so we hope that further research in this area will advance literacy around ways of seeing data and ways of seeing with and through data visualizations. As visualization tools and practices become more and more ubiquitous, this might include not only the development of a critical *hermeneutics*, but also new kinds of self-reflexive *praxis* for the creation and reconfiguration of visualizations which are attentive to the forms of mediation that we have outlined. Experimentation in this direction might be informed by calls

for non-reductive visualization (Manovich, 2002), humanistic interfaces (Drucker, 2014), feminist data visualization (D'Ignazio, 2015), inventive methods (Lury and Wakeford, 2012), rethinking dashboards (Tkacz, 2015), and critical analytics (Rogers, 2015).

References

- Agre, P. E. (1997) Toward a Critical Technical Practice: Lessons Learned in Trying to Reform AI. In G. C. Bowker, S. Star, L. Gasser et al. (Eds.), *Social Science, Technical Systems, and Cooperative Work: Beyond the Great Divide* (Mahwah: Lawrence Erlbaum), pp. 131–58.
- Akrich, M. (1992) The De-Description of Technical Objects. In W. Bijker and J. Law (Eds.), *Shaping Technology/Building Society: Studies in Sociotechnical Change* (Cambridge, MA: MIT Press), pp. 205–24.
- Berger, J. (1972) *Ways of Seeing* (London: Penguin).
- Berkowitz, D. (2009) Reporters and Their Sources. In K. Wahl-Jorgensen and T. Hanitzsch (Eds.), *The Handbook of Journalism Studies* (New York: Routledge), pp. 102–15.
- Bertin, J. (1983) *Semiology of Graphics: Diagrams, Networks, Maps*. (Madison: University of Wisconsin Press).
- Bowker, G., and Star, S. (1999) *Sorting Things Out: Classification and its Consequences* (Cambridge, MA: MIT Press).
- Cairo, A. (2012) *The Functional Art: An Introduction to Information Graphics and Visualization* (Berkeley: New Riders).
- Card, S. K., Mackinlay, J., and Schneiderman, B. (Eds.). (1999) *Readings in Information Visualization: Using Vision to Think* (San Francisco: Morgan Kaufmann).
- Cleveland, W. S. (1993) *Visualizing Data* (Murray Hill: Hobart Press).
- Cleveland, W. S. (1994) *The Elements of Graphing Data* (Murray Hill: Hobart Press).

- Cottle, S. (2003). News, Public Relations and Power: Mapping the Field. In S. Cottle (ed.), *News, Public relations and Power* (London: Sage), pp. 3–24.
- Cubitt, S. (2015) Data Visualization and the Subject of Political Aesthetics. In D. Berry and M. Dieter (Eds.), *Postdigital Aesthetics: Art, Computation And Design* (London: Palgrave Macmillan), pp. 179–90.
- D'Ignazio, C. (2015) What Would Feminist Data Visualization Look Like? MIT Center for Civic Media 20 December: <https://civic.mit.edu/feminist-data-visualization> (accessed 10 March 2016).
- Daston, L. and Galison, P. (2010) *Objectivity* (Cambridge, MA: MIT Press).
- Drucker, J. (2014) *Graphesis: Visual Forms of Knowledge Production* (Cambridge, MA: Harvard University Press).
- Espeland, W. N. and Stevens, M. L. (2008) A Sociology of Quantification. *European Journal of Sociology* 49(3): 401–36.
- Espeland, W. N. and Sauder, M. L. (2007) Rankings And Reactivity: How Public Measures Recreate Social Worlds. *American Journal of Sociology* 113(1): 1–40.
- Few, S. (2009) *Now You See It: Simple Visualization Techniques for Quantitative Analysis* (Oakland: Analytics Press).
- Few, S. (2012) *Show Me the Numbers: Designing Tables and Graphs to Enlighten*. 2nd ed (Burlingame: Analytics Press).
- Few, S. (2013) *Information Dashboard Design: Displaying Data for At-a-Glance Monitoring* (Burlingame: Analytics Press).
- Few, S. (2014) Data Visualization for Human Perception. In M. Soegaard and R. Dam (Eds.), *The Encyclopedia of Human-Computer Interaction*. 2nd Ed. (Aarhus: The Interaction Design Foundation), available at: <https://www.interaction->

design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/data-visualization-for-human-perception (accessed 20 March 2016).

Freeman, L. C. (2004) *The Development of Social Network Analysis* (Vancouver: Booksurge).

Galison, P. (1990) Aufbau/Bauhaus: Logical Positivism and Architectural Modernism. *Critical Inquiry* 16(4): 709–52.

Gerlitz, C and Lury, C. (2014) Social Media and Self-Evaluating Assemblages: On Numbers, Orderings and Values. *Distinktion* 15(2): 174–88.

Gray, J. (2015a) Open Budget Data: Mapping the Landscape. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2654878 (accessed 15 March 2016).

Gray, J. (2015b) Examples of Fiscal Data Visualisations. *figshare*. Available at: https://figshare.com/articles/Examples_of_Fiscal_Data_Visualisations/1548331 (accessed 15 March 2016).

Gray, J. and Davies, T. (2015) Fighting Phantom Firms in the UK: From Opening Up Datasets to Reshaping Data Infrastructures? Working paper presented at the Open Data Research Symposium at the 3rd International Open Government Data Conference (Ottawa). Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2610937 (accessed 10 March 2016).

Hall, S., Critcher, C., Jefferson, T. et al. (1978) *Policing the Crisis: Mugging, the State, and Law and Order* (London: Macmillan).

Halpern, O. (2015) *Beautiful Data: A History of Vision and Reason Since 1945* (Durham: Duke University Press).

- Headrick, D. R. (2000) *When Information Came of Age: Technologies of Knowledge in the Age of Reason and Revolution, 1700-1850* (Oxford: Oxford University Press).
- Heller, S. and Landers, R. (2014) *Raw Data: Infographic Designers' Sketchbooks* (London: Thames and Hudson).
- Jay, M. (1988). The Rise of Hermeneutics and the Crisis of Ocularcentrism. *Poetics Today* 9(2): 307–26.
- Jevbratt, L. (2004) The Prospect of the Sublime in Data Visualizations. `
- Katz, J. (2012) *Designing Information: Human Factors and Common Sense in Information Design* (Hoboken: Wiley).
- Kennedy, H., Hill, L., Aiello, G. et al. (2016) The Work that Visualisation Conventions Do. *Information, Communication & Society* 19(6): 715–35.
- Krum, R. (2013) *Cool Infographics: Effective Communication with Data Visualization and Design* (Indianapolis: Wiley).
- Lau, A. and Vande Moere, A. (2007) Towards a Model of Information Aesthetics in Information Visualization. In *Proceedings of the 11th International Conference Information Visualization* (Washington, DC: IEEE Computer Society), pp. 87–92.
- Law, J. and Ruppert, E. (2013) The Social Life of Methods: Devices. *Journal of Cultural Economy* 6(3): 229–40.
- Lima, M. (2011) *Visual Complexity: Mapping Patterns of Information* (New York: Princeton Architectural Press).
- Lury, C. and Wakeford, N. (Eds.) (2012) *Inventive Methods: The Happening of the Social* (London: Routledge).
- Manning, P. (2001) *News and News Sources: A Critical Introduction* (London: Sage).

- Manovich, L. (2002). Data Visualisation as New Abstraction and Anti-sublime. In B. Hawk, D. Rieder and O. Oviedo (Eds.), *Small Tech: The Culture of Digital Tools, Electronic Mediations* (Minneapolis: University of Minnesota Press), pp. 3–9.
- Manovich, L. (2011) What is Visualisation? *Visual Studies* 26(1): 36–49.
- Manovich, L. (2014) Visualization Methods for Media Studies. In C. Vernallis, A. Herzog and J. Richardson (Eds.), *Oxford Handbook of Sound and Image in Digital Media* (Oxford: Oxford University Press), pp. 253–78.
- Marres, N. and Gerlitz, C. (2015) Interface Methods: Renegotiating Relations Between Digital Social Research, STS and Sociology. *The Sociological Review* 64(1): 21–46.
- Mayr, E. and Schreder, G. (2014) Isotype Visualizations. A Chance for Participation & Civic Education. *JeDEM – eJournal of eDemocracy and Open Government* 6(2), 136–50.
- McCandless, D. (2012) *Information is Beautiful* (New York: Collins).
- McCandless, D. (2014) *Knowledge is Beautiful* (New York: Collins).
- McCosker, A. and Wilken, R. (2014) Rethinking ‘Big Data’ as Cvisual Knowledge: The Sublime and the Diagrammatic in Data Visualisation. *Visual Studies* 29(2): 155–64
- McLuhan, M. (1962) *The Gutenberg Galaxy: The Making of Typographic Man* (Toronto: University of Toronto Press).
- Meirelles, I. (2013) *Design for Information: An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations*. Beverly, MA: Rockport Publishers.
- Munzner, T. (2014) *Visualization Analysis and Design* (Boca Raton: CRC Press).
- Murray, S. (2013) *Interactive Data Visualization for the Web* (Sebastopol: O’Reilly).
- Neurath, O. (1939) *Modern Man in the Making* (New York: Knopf).

- Playfair, W. (1801) *The Commercial and Political Atlas, Representing, by Means of Stained Copper-Plates, the Progress of the Commerce, Revenues, Expenditure, and Debt of England during the Whole of the Eighteenth Century*. 3rd ed (London: Wallis).
- Rayward, W. B. (Ed.). (2008) *European Modernism and the Information Society: Informing the Present, Understanding the Past* (Aldershot: Ashgate).
- Rendgen, S. (2012) *Information Graphics* (Köln: Taschen).
- Rendgen, S. (2014) *Understanding the World: The Atlas of Infographics* (Köln: Taschen).
- Rieder, B. and Röhle, T. (2012) Digital Methods: Five Challenges. In D. M. Berry (Ed.), *Understanding Digital Humanities* (London: Palgrave Macmillan), pp. 67–84.
- Rogers, R. (2013) *Digital Methods* (Cambridge, MA: MIT Press).
- Rogers, R. (2015) From Dashboards to Critical Analytics. Paper presented at Data Power Conference (University of Sheffield).
- Rosenberg, D. and Grafton, A. (2013) *Cartographies of Time: A History of the Timeline* (New York: Princeton Architectural Press).
- Ruppert, E. (2015) Doing the Transparent State: Open Government Data as Performance Indicators. In: R. Rottenburg, S. E. Merry, S.-J. Parkand and J. Mugler (Eds.), *A World of Indicators: The Making of Governmental Knowledge through Quantification*. (Cambridge: Cambridge University Press), pp. 1–18.
- Ruppert, E., Law, J. and Savage, M. (2013) Reassembling Social Science Methods: The Challenge of Digital Devices. *Theory, Culture & Society* 30(4): 22–46.
- Sack, W. (2011) Aesthetics of Information Visualization. In M. Lovejoy, C. Paul and V. Vesna (Eds.), *Context Providers: Conditions of Meaning in Media Arts* (Bristol: Intellect), pp. 123–50)..
- Schoffelen, J., Claes, S., Huybrechts, L. et al. (2015) Visualising Things. Perspectives on How to Make Things Public Through Cisualisation. *CoDesign* 11(3–4): 179–92.

- Segel, E. and Heer, J. (2010) Narrative Visualization: Telling Stories with Data. *IEEE Transactions on Visualization and Computer Graphics* 16(6): 1139–48.
- Spence, R. (2014) *Information Visualization: An Introduction* (New York: Springer).
- Star, S. (1999) The Ethnography of Infrastructure. *American Behavioral Scientist* 43(3): 377–91.
- Star, S. and Ruhleder, K. (1996) Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces. *Information Systems Research* 7(1): 111–34.
- Steele, J. and Iliinsky, N. (Eds.) (2010) *Beautiful Visualization: Looking at Data through the Eyes of Experts* (Sebastopol: O'Reilly).
- Strauss, A. and Corbin, J. (1998) Basics of Qualitative Research. Techniques and Procedures for Developing Grounded Theory. 2nd ed (London: Sage).
- Tactical Technology Collective (2014) *Visualising Information for Advocacy*.
- Tkacz, N. (2015) Connection Perfected: What the Dashboard Reveals. Keynote Address at Digital Methods Winter School 2015 (University of Amsterdam).
- Tufte, E. R. (1990) *Envisioning Information* (Cheshire: Graphics Press).
- Tufte, E. R. (1997a) *Visual Explanations: Images and Quantities, Evidence and Narrative* (Cheshire: Graphics Press).
- Tufte, E. R. (1997b) *Visual and Statistical Thinking: Displays of Evidence for Making Decisions* (Cheshire: Graphics Press).
- Tufte, E. R. (2001) *The Visual Display of Quantitative Information*. 2nd ed (Cheshire: Graphics Press).
- Tukey, J. (1977) *Exploratory Data Analysis* (Reading: Addison-Wesley).
- Venturini, T., Bounegru, L., Jacomy, M. and Gray, J. (forthcoming) How to Tell Stories with Networks: Exploring the Narrative Affordances of Network Graphs with the Iliad. In

- M. T. Schaefer and K. van Es (Eds.), *The Datafied Society: Social Research in the Age of Big Data* (Amsterdam: University of Amsterdam Press).
- Whitelaw, M. (2008) Art Against Information: Case Studies in Data Practice. *Fibreculture* 11: <http://eleven.fibreculturejournal.org/fcj-067-art-against-information-case-studies-in-data-practice/> (accessed 10 March 2016).
- Wong, D. M. (2013) *The Wall Street Journal Guide to Information Graphics: The Dos and Don'ts of Presenting Data, Facts, and Figures* (New York: Norton).
- Wright, R. (2008) Data Visualization. In M. Fuller (ed.), *Software Studies: A Lexicon*. Cambridge, MA: MIT Press), pp. 78–87.
- Yau, N. (2011) *Visualize This: The FlowingData Guide to Design, Visualization, and Statistics* (Indianapolis: Wiley).
- Yau, N. (2013) *Data Points: Visualization That Means Something* (Indianapolis: Wiley).
- Zambrano, R. N. and Engelhardt, Y. (2008) Diagrams for the Masses: Raising Public Awareness – From Neurath to Gapminder and Google Earth. In G. Stapleton, J. Howse and J. Lee (Eds.), *Diagrammatic Representation and Inference* (Berlin: Springer), pp. 282–92.