Data Visualisation as an Emerging Tool for Online Research

In: The SAGE Handbook of Online Research Methods

By: Helen Kennedy & William Allen
Edited by: Nigel G. Fielding, Raymond M. Lee & Grant Blank
Pub. Date: 2017
Access Date: September 16, 2019
Publishing Company: SAGE Publications Ltd
City: 55 City Road
Print ISBN: 9781473918788
Online ISBN: 9781473957992
DOI: https://dx.doi.org/10.4135/9781473957992
Print pages: 307-326

Data Visualisation as an Emerging Tool for Online Research

Helen Kennedy William Allen

Introduction

This chapter focuses on data visualisation, an increasingly important method in the online research toolset and a means of communicating research results to peers and the wider public. This is not a 'how-to' chapter; it does not guide the reader through the process of making data visualisations. That project would require more words than are available here and, anyway, is best undertaken by professional data visualisers. There are several good, how-to books written by visualisation practitioners, such as Cairo (2013), Few (2012), Kirk (2016), Tufte (1983) and Yau (2013), which readers can turn to for guidance on the visualisation process, and we draw on some of them here. We describe ourselves as academics who are researchers *of* and researchers *with* data visualisation who, in the process of doing our research, have witnessed and reflected on a growth in academic visualisation. Responding to this phenomenon and other issues relating to the spread of data and visualisations, this chapter focuses on how online researchers might think critically about them, something which, we suggest, is a pre-requisite to producing good visualisations.

The chapter starts with a brief note about what data visualisation is, before moving to a longer discussion of claims about what data visualisations can and cannot do. We consider it vital to foreground these questions before moving on to a discussion of the processes of creating and engaging with data visualisations. The subsequent section on visualisation tools, techniques and processes aims to point readers in the direction of resources and highlight key principles and approaches, rather than cover this subject matter comprehensively. The chapter then moves to focus on examples of using data visualisation within social science research, drawing on our own experiences. It concludes by summarising what we consider to be the key issues for online researchers seeking to use data visualisation in their research, noting the importance of attending to audiences, their needs and the contexts of their visualisation use.

We have been researching (and researching with) data and their visualisation for several years. Together with Rosemary Lucy Hill and Andy Kirk, in 2014 and 2015 we worked on Seeing Data (http://seeingdata.org/), a research project that explored the factors in visualisation consumption and production processes that affect user engagement. Before that, Will worked with data visualisation through his work for the Migration Observatory (http://www.migrationobservatory.ox.ac.uk/) at the University of Oxford and Helen researched the spread of data mining (Kennedy, 2016), and we both continue doing these things. We draw on our research in this chapter and on some of the publications that have resulted from it, which we have authored in collaboration with Hill, Kirk and others.

A Note on what Data Visualisation is (and is not)

A data visualisation is a visual representation of data, often in charts and graphs. It shows statistical, numerical data in visual ways in order to help people make sense of data. Experts believe representing data visually makes it possible to communicate data effectively and gives people the opportunity to analyse and examine large datasets which would otherwise be difficult to understand (for example, Few, 2008). In *Data Visualisation: A Handbook for Data Driven Design*, a publication aimed at social science researchers unfamiliar with the area, Andy Kirk defines data visualisation as 'the representation and presentation of data to facilitate understanding' (2016: 19). He then breaks down this definition with reference to each of its core elements. *Representation* refers to the choices made about the visual form in which the data will be portrayed, such as decisions about which chart types to use, whereas *presentation* refers to decisions about the visualisation design, such as colour choice, composition, level of interactivity and annotation. For Kirk, and in the view of other visualisers, there are two main ways in which visualisations 'facilitate understanding': the first is to communicate data and the second is to enable their exploration or analysis (Kennedy, 2014). Both modes are relevant to online researchers: we may use visualisation to communicate our research data to expert peers or non-expert publics, or we may visualise our data in order to explore and analyse them. For example, a powerful tool like Tableau makes it possible to 'see' data (and identify patterns within them) in ways that are simply not possible with large datasets presented in tabular form. We may also produce visualisations which enable our audiences to do this, as seen in the case study examples discussed later.

Of course, a defining feature of a data visualisation is that it has data at its heart. This differentiates it from an infographic, which is traditionally static, made for print-based consumption and explains phenomena graphically but may contain no data, or data in charts which exist alongside other illustrations like photographs. A data visualisation is also different from an information visualisation: in the latter, information is the output; whereas in the former, data are the input, although these terms are often used interchangeably. There is, of course, much more to data visualisation than these simple definitions suggest, as we demonstrate in the next section.

What Data Visualisation Can (and Cannot) Do

What are data visualisations used for, what can they do and how might online researchers integrate them into their practice? A number of visualisation professionals assert that visualisations can promote greater understanding of data by making them accessible and transparent (Few, 2008; Zambrano and Engelhardt, 2008). Experts and practitioners often express a belief that, through visualisation, they can 'do good with data', the trademarked tagline of US-based visualisation agency Periscopic (Periscopic, 2014). This view that visualisation is a way of 'doing good with data' was widespread amongst visualisation designers who we interviewed for our Seeing Data research (Kennedy, 2014; Kennedy *et al.,* 2016a).

The idea that visualisation can promote awareness can be traced back to the work of Otto and Marie Neurath in the mid-nineteenth century and their development of the graphical language Isotype, a visual way of representing quantitative information via icons (Zambrano and Engelhardt, 2008). The Neuraths believed that

'visual education is related to the extension of intellectual democracy within single communities and within mankind' (Neurath *et al.*, 1973: 247). They put their ideas into practice in museums they directed, where they used charts to enable the general public to develop understanding of 'the problems the community of Vienna had to tackle' (Neurath, quoted in Zambrano and Engelhardt, 2008: 283).

Zambrano and Engelhardt link the ideas of the Neuraths to contemporary projects like GapMinder (http://www.gapminder.org/world), which describes itself as 'a modern "museum" that helps making the world understandable, using the Internet' and aims to promote global sustainable development by visualising related statistics (Stiftelsen Gapminder, n.d.). The efforts of other contemporary visualisers can also be seen in this vein, such as Stefanie Posavec's 'Open Data Playground' (http://www.stefanieposavec.co.uk/data/#/open-data-playground/), a set of floor-based games that provide people with the opportunity to play with materialisations of open datasets and make sense of the data for themselves. These projects, in different ways, reflect the belief that visualisations make data transparent, summed up in the words of Stephen Few (2008): 'infovis can make the world a better place'.

However, as we note in an article co-authored with Hill and Aiello (Kennedy *et al.,* 2016a), critical commentators argue that data visualisations can privilege certain viewpoints, perpetuate existing power relations and or create new ones, and they often draw on examples of visualisations in the media as evidence of this view. These include the US Republican party's visualisation of the Democrats' proposed reforms to healthcare, described by Valarakis (2014) as an over-complicated visualisation which serves to make the proposed reforms seem over-complicated too, and the UK newspaper *Daily Express*'s use of visualisations to communicate an anti-trade union ideology, studied by Dick (2015).

These and other commentators observe that data visualisations are not neutral windows onto data; rather, visualisations are the result of numerous choices: as Ambrosio points out, 'visual manifestations [of data] are themselves informed by judgement, discernment and choice' (2015: 137). Yet although there are many subjective processes involved in visualising data, some critics argue that the resulting visualisation often 'pretends to be coherent and tidy' (Ruppert, 2014). Visualisations and the data within them seem objective, even though they are not. This appearance has a number of origins. First, they report numbers, historically trusted because they appear universal, impersonal and neutral, as Porter (1995) and others have argued. Second, data and visualisation are often associated with science, also seen to be objective and therefore trustworthy. Third, as we argue with Hill and Aiello (Kennedy *et al.*, 2016a), the conventions that have been established over time also work to imbue visualisations with the quality of objectivity, producing the impression that visualisations are 'showing the facts, telling it like it is, offering windows onto data' (Kennedy *et al.*, 2016a: 716).

The shape that visualised data take is the result not only of the decisions and priorities of the data visualiser and the data gatherer, but also of the makers of the visualisation and data gathering software used. Human decisions influence and shape the design, development, arrangement and implementation of data and their visualisation in many ways. Consequently, data are never 'raw' – the very concept of 'raw data', as Bowker (2005) puts it, is an oxymoron. Data, like their visualisation, are generated through processes

which necessarily involve interpretation. These interpretations are in turn biased by the subjective filters that individual humans apply as they make them (Bollier, 2010). To understand how visualisations turn out the way that they do, it is necessary to acknowledge the roles of the people, software packages and processes that produce them.

Most good data visualisers recognise that both of the perspectives discussed in this section (that visualisation can make data accessible and that visualisation involves manipulating data) are valid. Kirk's book (2016) includes extensive discussion of the ways in which visualisation involves decision making, about what to prioritise, what to leave out, how to present and represent data, all of which influence how visualisations – and data – look. He argues that for a visualisation to be trustworthy, all data treatments and transformations – including smoothing, cleaning, converting and adjusting – must be noted and shared with users. Doing this means making transparent the perspective that has influenced the visualisation design. A good example of how different perspectives on the same data can lead to different design decisions and therefore different messages can be seen in ‘Iraq's Bloody Toll’ and responses to it. This visualisation, produced in 2011 by Simon Scarr for *the South China Morning Post* and reproduced in Figure 18.1, is deliberately evocative. The use of an upside-down bar chart with rounded rather than square ends, the colour red (visible at the original URL) and the visualisation's title all communicate a clear message: ‘too many deaths in Iraq’ (Hill, 2014). In 2014, Andy Cotgreave of Tableau drew on this visualisation to respond to an article in The *Guardian* newspaper's datablog entitled ‘Why you should never trust a data visualisation’ (Burn-Murdoch, 2013)[1] in which the author expressed concern about the credibility that is often attached to data visualisations. In Cotgreave's (2014) response, shown in Figure 18.2, he shows how the same data can have a very different effect with three simple changes to ‘Iraq's Bloody Toll': he flips the bar chart up, changes the title to ‘Iraq: Deaths on the Decline’ and makes the visualisation blue, not red (Figure 18.2, colour change visible at original URL). In so doing, he points to the impossibility of neutrality in data visualisation. The same data can be represented in different ways to create different messages, all of which are ostensibly trustworthy.

Figure 18.1‘Iraq's Bloody Toll’ by Scarr (2011)

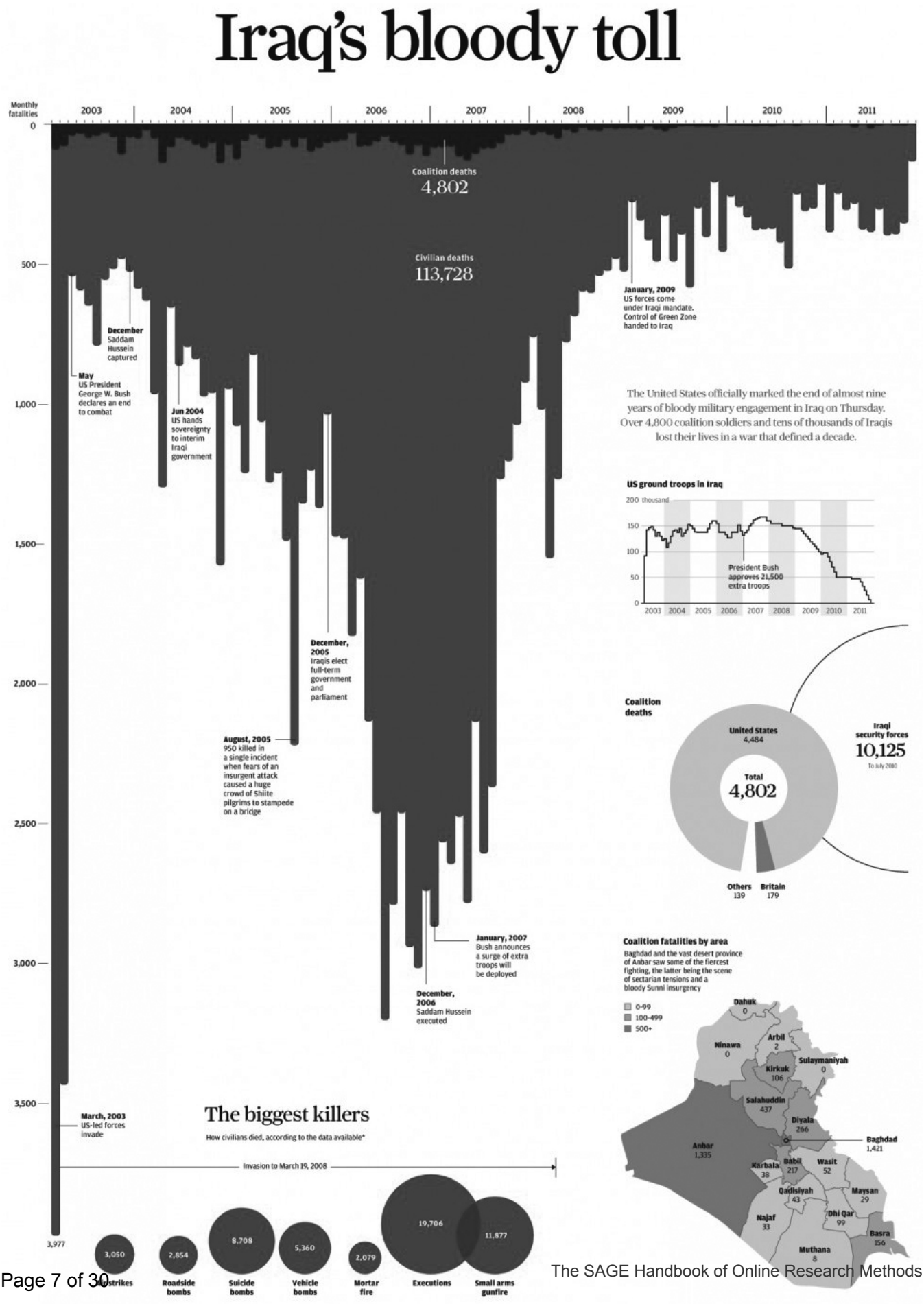
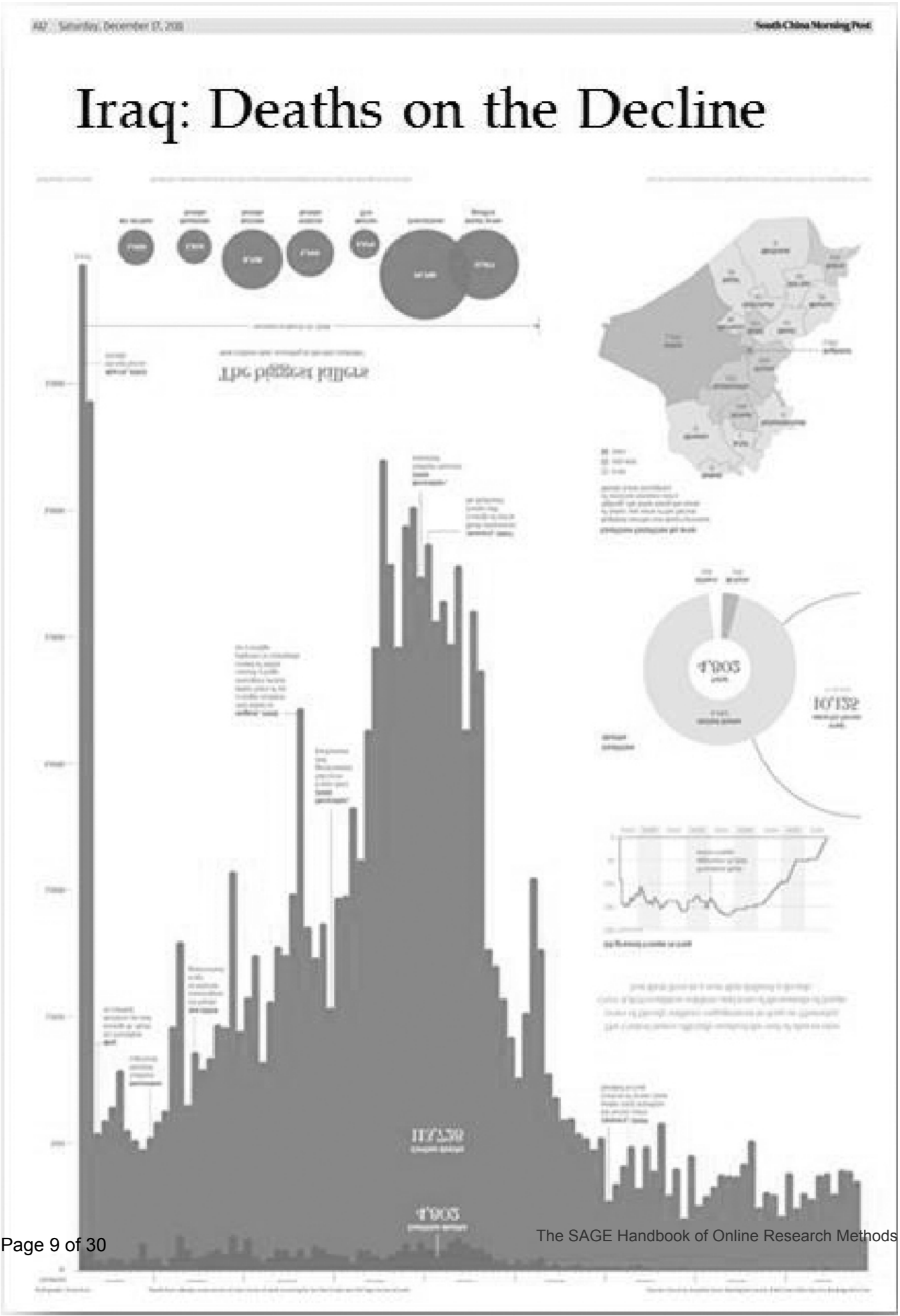


Figure 18.2‘Iraq: Deaths on the Decline’ by Cotgreave (2014)



These points are relevant to online researchers because considering how data visualisation can be used in different contexts is an important component of good, reflective visualisation practice. Researchers using and producing data visualisations need to understand them sociologically in order to do so well. As we note with Hill and Aiello, almost 30 years ago Fyfe and Law (1988) urged sociologists to take the visual seriously in the study of social life, because '[d]epiction, picturing and seeing are ubiquitous features in the process by which most human beings come to know the world as it really *is* for them' (Fyfe and Law, 1988: 2, cited in Kennedy *et al.,* 2016a: 732). This need remains, especially for researchers working with data visualisation.

Doing Data Visualisation I: Tools, Techniques, Process

In this section, we provide some brief commentary on the tools, techniques and process of data visualisation. In doing so, we draw heavily on Kirk's book and website.

Tools: Software

There is a huge and wide range of software for data visualisation. The resources page of Kirk's Visualising Data website (http://www.visualisingdata.com/resources/) lists 298 tools, applications and platforms that can be used in the data visualisation process. These range from software which is specifically for visualisation (such as Big Picture, Chart Builder, D3, Graphviz and others which specialise in particular chart types, such as e-Sankey), to tools to help with the aesthetic aspects of visualisation (such as 0 To 255 and other colour-selection tools) and programming languages like Python. Kirk categorises these resources into tools for data handling; charting tools; programming-based tools; multivariate, mapping and web-based tools; specialist tools; and resources for working with colour. We comment on four tools below, because they are widely used (Tableau and NVivo), important (D3) or freely available and not complex to use (Raw). These are just a few examples of available tools, and this brief discussion is far from comprehensive. Interested readers are encouraged to visit Kirk's site for more detail.

One particularly popular visualisation tool is Tableau, available in both Desktop (paid) and Public (free) versions. Featuring many different chart types, as well as drag-and-drop interfaces, this tool offers a range of options for visualising data and publishing the results. (Some of the case study visualisations within this chapter were created using Tableau Public). A large user-community provides answers to common questions as well as guidance for combining datasets in the Tableau format. Users should note that using the Public, free version of Tableau means making data publically available, hence the name.

D3 is an important programming library for creating dynamic and interactive data visualisations. It is a JavaScript library for manipulating documents based on data using HTML (HyperText Markup Language), SVG (Scalable Vector Graphics) and CSS (Cascading Style Sheets). In using these web-based languages, it produces visualisations which adhere to web standards, meaning that they can be embedded in webpages which will function across browsers and devices, now and into the future.

Raw is an open web application to create custom visualisations on top of the D3.js library through a simple interface. Primarily conceived as a tool for designers and visualisation experts, Raw enables visualisations to be exported and embedded in webpages. Even though it is a web application, Raw does not store data. Therefore, data are only available to the person who originally uploads them. Raw is open, customisable and free to download.

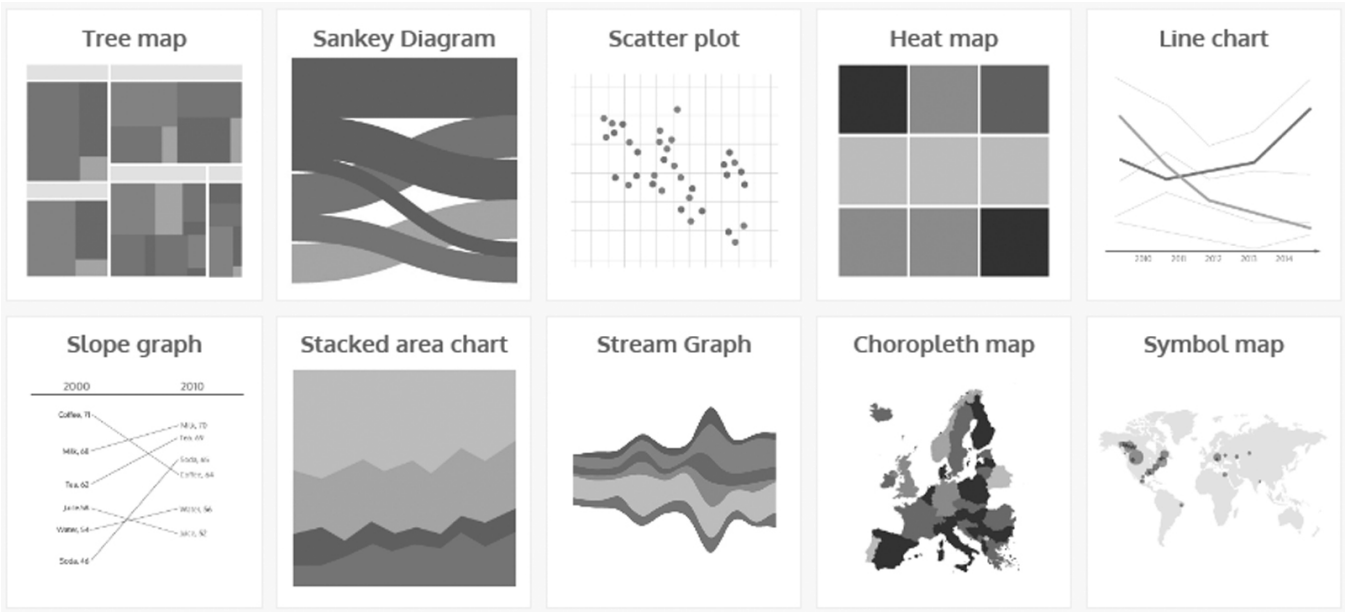
Many qualitative researchers will already use NVivo, an application for collecting, organising and analysing data gathered through interviews, focus groups and other methods. Like other software for handling qualitative and quantitative data, NVivo also offers the option of visualising data in a range of chart types, such as bar charts, hierarchical clusters, word trees, cluster diagrams and geographical maps. These can be used by researchers both to explore their own data and to communicate them to their audiences. Other available software for qualitative data include QDA Miner and MaxQDA. QDA Miner provides options to visualise where codes generated by the researcher appear in textual data, which can reveal patterns of co-occurrence. MaxQDA contains similar features, including the ability to create word clouds of most frequently occurring words, but it is, not it is worth specifically highlighting one particular tool called the ‘Document Portrait'. After dividing selected texts into equal segments that are represented by a square or circle as the user wishes, this portrait function assigns a colour to each symbol that corresponds with its given code. The researcher can sort and display these coloured symbols to get a sense of the proportions of different codes in a set of texts, or to identify which codes are used most frequently. Newer versions of MaxQDA allow users to click through the resulting ‘portrait’ of symbols to access the underlying textual data, too.

Techniques: Graph and Chart Types

On our Seeing Data project website, we include a section entitled ‘Understanding Data Visualisations', which aims to help people who are interested in data visualisations but not experts to make sense of them. One subsection, Inside The Chart (http://seeingdata.org/sections/inside-the-chart/), introduces what we see as 14 of the most commonly used graph and chart types, explaining what they show, how they should be read, their limitations and alternative charts which can be used to show similar data. Amongst these are familiar charts such as the ubiquitous bar chart, stacked bar chart, pie chart and line chart; fairly common and easy-to-read visualisations such as the choropleth map and the symbol map; and more complex forms including the radar chart, tree map, Sankey diagram, scatter plot, heat map, slope graph, stacked area chart and stream graph. All of these are shown in Figure 18.3 below.

Figure 18.3Screenshot of the Seeing Data website showing 14 common chart types

Source: http://seeingdata.org/sections/inside-the-chart/



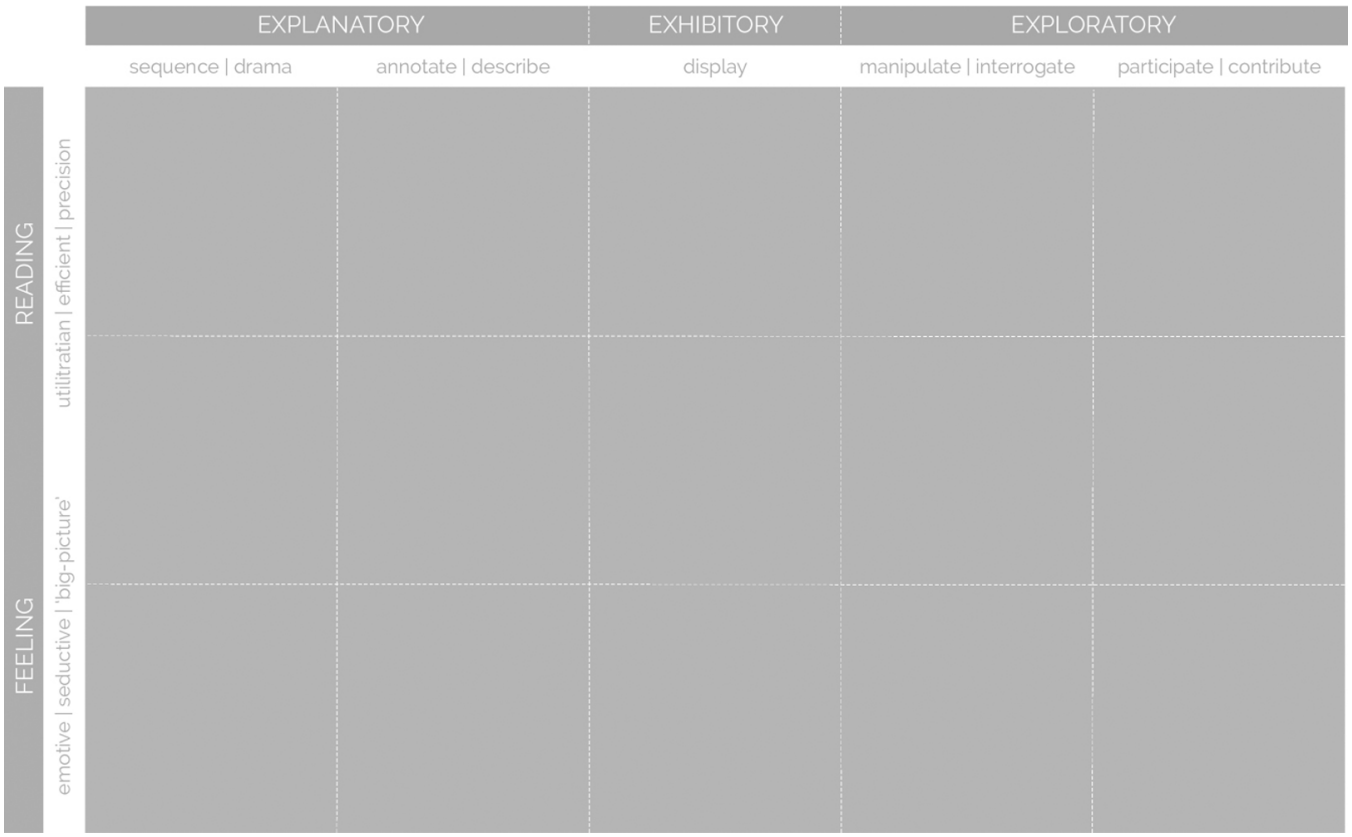
Kirk's book (2016) identifies 50 common chart types. There, he provides this useful categorisation to assist readers in understanding the types of data and relationships between data that they commonly represent:

- **Categorical chart types:** used for comparing categories and distributions of quantitative values (for example, bar charts)
- **Hierarchical chart types:** used for comparing part-to-whole relationships and hierarchies (for example, pie charts and tree maps)
- **Relationship chart types:** used for graphing relationships through correlations and connections (for example, a scatter plot or a Sankey diagram)
- **Temporal chart types:** used for showing trends and activities over time (for example, a line chart or a stream graph)
- **Spatial data chart types:** used for mapping spatial data (for example, in a symbol or chloropleth map) (Kirk, 2016: 158)

Process: Principles, Purpose and other Professional Practices

In addition to discussing tools and chart types, Kirk proposes three guiding principles for visualisation design. Visualisations should be seen to be trustworthy, accessible and elegant, argues Kirk, and designers need to consider how to produce visualisations that comply with these principles. Kirk presents ‘the purpose map', reproduced in Figure 18.4, as a way of ensuring that all thinking and decision making is aligned to these principles, as well as adhering to the desired outcomes of the visualisation.

Figure 18.4‘The purpose map’ by Kirk (2016)



The purpose map brings together what Kirk defines as the experience and the tone of a visualisation. As the map shows, Kirk argues that there are three types of intentions with regard to user experience, which he defines as follows:

- **Explanatory:** in which visualisers ‘will provide the viewer with a visual portrayal of the subject's data *and* will also take some responsibility to bring key insights to the surface, rather than leave the prospect of interpreting the meaning of the information entirely to the viewer’ (Kirk, 2016: 77).
- **Exploratory:** in which visualisers help users find their own insights, usually through digital, interactive and participatory experiences which allow them to interrogate and manipulate data. Users are free to interact and explore, but might ask themselves: what do you want me to do with this?
- **Exhibitory:** neither explanatory nor exploratory. Kirk describes visualisations which fall into this category as ‘simply visual displays of data'. Viewers have to do the work of interpreting meaning (unlike in explanatory visualisations) – like artworks, they depend on ‘the interpretative capacity of the viewer’ and so are suited to audiences with subject knowledge who can do their own interpreting. They may support explanation given elsewhere, for example in accompanying text or in a presentation.

Although these three intention types are neither mutually exclusive nor exhaustive, the purpose map provides a useful rule-of-thumb guide to visualisation design. Also, particular intentions do not always equate to the use

of particular chart types, although simple charts like bar charts and stacked bar charts are more explanatory, and interactive visualisations using a range of chart types are likely to be more exploratory. Researchers using data visualisation to communicate with expert peers and non-expert publics will often have more explanatory than exhibitory or exploratory purposes. However, we are likely to move towards exploratory approaches as we increasingly share datasets with our audiences and invite them to explore and analyse data for themselves.

The vertical axis of the purpose map relates to the intended tone of the visualisation, which for Kirk exists on a spectrum from reading to feeling. A visualisation which is intended to be *read* prioritises perceptual accuracy, is utilitarian and pragmatic – 'no-frills', Kirk calls it – for example, a simple bar chart. Readable visualisations like bar charts facilitate trustworthiness and accessibility, he argues. In contrast, visualisations that are intended to be *felt* (like a tree map, for example) are used when visualisers 'place more importance on extracting a *gist* of the big, medium and small values and a general *sense* of the relationships that exist. Sometimes an "at-a-glance" sense of scale is simply the most suitable way to portray a subject's values' (Kirk, 2016: 84). It was this sense that visualisations are *felt* as much as they are *read* that motivated us to explore the factors that affect visualisation engagement on Seeing Data, including feelings and emotions. The tone adopted in a visualisation depends on its purpose, but the choices made need to be compatible with attributes of the data.

Two other important decisions in the visualisation production process relate to the extent to which annotation and interactivity will be included. Alan Smith, data visualisation editor at the *Financial Times* newspaper, claims that people are afraid of writing on graphs (Smith, 2016). However, as Smith shows through his own examples, annotations can be extremely useful in helping users navigate. But there are many kinds of annotation available. Kirk distinguishes project annotation, such as titles and subtitles, introductions, user guides and footnotes (which can include links to data sources and credits) from chart annotations, which include labels (axis labels and value labels), legends, reading guides and captions. The challenge, writes Kirk, is to know how much annotation is the right amount: too much might result in a cluttered chart and a patronised audience, whereas too little might leave users struggling to find their own way around a visualisation. In our research, we found that annotation was hugely valued by users unfamiliar with visualisation as a communication form.

The amount of interactivity in a visualisation is another important consideration in the design process. Interactive features usually allow users to adjust the data they are shown or how it is presented, and so support the accessibility principle, according to Kirk. At the time of writing, there is much enthusiasm about the personalisation and gamification of visualisation that interactivity enables, but it is worth noting that in our research, we found that without a clear purpose, these features were not always appreciated by visualisation users. For online researchers using visualisation to communicate data in a journal article or conference presentation, questions about how much and what kind of interactivity are likely to be redundant, but once we start to communicate our research online to non-expert as well as expert audiences, the question gains relevance.

Of course, there are many more considerations than these when producing a data visualisation, such as decisions about axes, scales and which graphical symbols to use. These presentation choices relate to the 'how-to' of data visualisation, addressed comprehensively in the guidebooks we referenced at the start of this chapter. Colour is a particularly important dimension of visualisation because it is a powerful sensory cue and therefore an influential visual property which can have an immediate impact on users and audiences: the different emotional impacts of Figures 18.1 and 18.2 are tied directly to colour. As Kirk notes, every feature of a visualisation has colour properties, and designers should primarily use colour to establish meaning, not to provide decoration. In a similar way, decisions about axes, scales and graphical symbols (and their related forms and areas) should be driven by an understanding of the meaning that is to be conveyed and should emerge from the data in order that the most appropriate combination for presenting data is chosen (see Kirk (2016) for extensive discussion of these considerations).

There is clearly a lot to think about in order to produce a good data visualisation: which tools to use, which chart type is appropriate, whether to include annotation and interactivity. These are in addition to expertise in doing statistical analysis, handling large datasets and comprehending the ideological work that visualisations do to make and shape the data. As Helen argues elsewhere with Hill (Kennedy and Hill, 2016), the growing availability of data and concomitant expectation that researchers will gather, mine, analyse and visualise could be seen as what Gill describes as 'the hidden injuries of neo-liberal academia' (Gill, 2009). In other words, enthusiasm about big data translates into pressure on researchers to engage with them and visualise them, despite sometimes having neither the requisite skills nor the time to acquire them. Neoliberal regimes mean that academic researchers individually shoulder the responsibility of struggling to adapt to ever-changing pressures, of which learning how to make good data visualisations is the most recent example. We consider this to be a serious issue for online researchers wishing to keep their skills current, so we acknowledge it here, if only to break the silence around these hidden injuries, as Gill suggests we should.

Doing Data Visualisation II: Examples from Migration Studies

This section demonstrates how the process of visualisation unfolds in a research setting by drawing on examples from The Migration Observatory where Will works. The Observatory is an independent organisation based at the University of Oxford that aims to inform public debate about immigration through original and secondary research. In an article written by Will about how British civil society organisations perceive data and research, he uses a critical realist orientation (Bhaskar, 1975) to highlight how factors such as the presence of diverse audiences, organisational objectives and available skills contribute to how these groups think about what 'useful' evidence looks like (Allen, 2016). In a similar way here, we attend to the ways in which such contextual factors shaped the representation and presentation of data in the visualisations we discuss. Because of the importance of such factors, we start with a discussion of the context, values and objectives of the Migration Observatory.

The Migration Observatory: Context, Values, Objectives

The Migration Observatory was founded as a politically independent body that brings data and research evidence about immigration and its impacts into public discussions. These discussions happen in the media, policy and government, and civil society, which includes charities and voluntary groups working with migrants, asylum seekers or refugees. The Observatory tries to work in ways that match its stated values of authoritativeness, independence, clarity, comprehensiveness and engagement. Most people encounter its materials, many of which are text-based, through its website (www.migrationobservatory.ox.ac.uk). Migration Observatory materials aim to put potentially complicated statistics into clear, simple summaries, sometimes turning textual insights into visual forms such as charts and maps. These outputs are important in the UK context for several reasons: immigration is a significant issue for the British public (Ipsos MORI, 2016); policy activity which aims to deal with the issue has real impacts on immigrants and also on UK citizens; and media and civil society groups, who are increasingly vocal players in the debate about immigration, increasingly turn to data to build their stories and cases. In this context, there is a pressing need to communicate information about migration and its socioeconomic impacts – and visualisations are a crucial way of meeting this need.

These details about the Migration Observatory and its work matter because visualisations exist in particular informational and political contexts. The Observatory makes visualisations for users in journalism, policy, civil society and for interested members of the public. These audiences have different kinds of skills, available time and end goals in accessing data – factors that we, along with Hill and Kirk, argue influence how visualisations are received (Kennedy *et al.*, under revision). These circumstances and audiences inform our broader arguments about what visualisations can do and why critical thinking about visualisation matters.

Case Study One: Migration in the Census

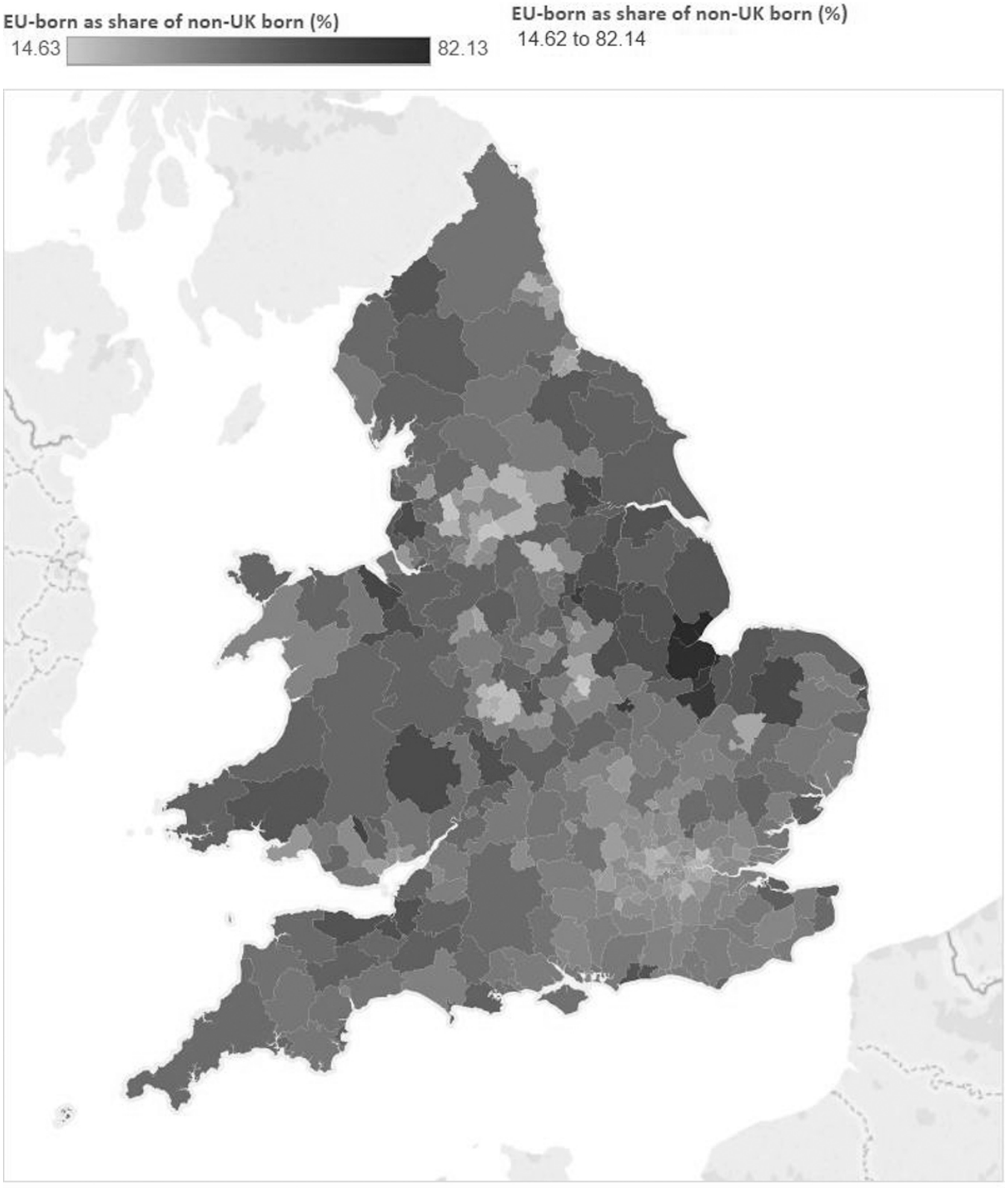
'Migration in the Census' aimed to analyse portions of the 2011 UK Census data (http://www.ons.gov.uk/census/2011census), released by the ONS, that related to the foreign-born population in the UK, and make key points within them accessible and available to audiences at national, regional and local levels.[2] To achieve this, the Observatory produced a series of briefings that provided summaries of results for each of the 12 regions across the UK, as well as summaries for England and Great Britain as wholes. These summaries include interactive maps, generated with Tableau Public, that allow users to filter and customise the visualisations according to their needs and interests. Figure 18.5, depicting the proportion of all foreign-born people who are EU-born in each English and Welsh local authority illustrates the style of these maps. In the online version, when users hover over a particular local authority, they see precise details including the number of EU-born people recorded in that area, drawn directly from the Census. As the proportion of EU-born people (among all foreign-born people) increases in a given local authority, the darker blue that area becomes.

Figure 18.5EU-born residents as share of local non-UK born, England and Wales

Source: http://public.tableau.com/views/MAPEU-bornresidents2011/EU-bornasshareoflocalnon-UKborn?:embed=y&&:loadOrderID=1&:display_count=yes

EU-born (excl. UK-born) residents as share of local non-UK born: England & Wales 2011

Map provided by www.migrationobservatory.ox.ac.uk



Source: England & Wales Census 2011, ONS.

Note: 'Old' EU refers to member countries of the EU in 2001. EU Accession countries refers to countries that joined the EU between 2001 and 2011, namely Bulgaria, Czech Republic, Cyprus (EU part), Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Romania, Slovakia and Slovenia.

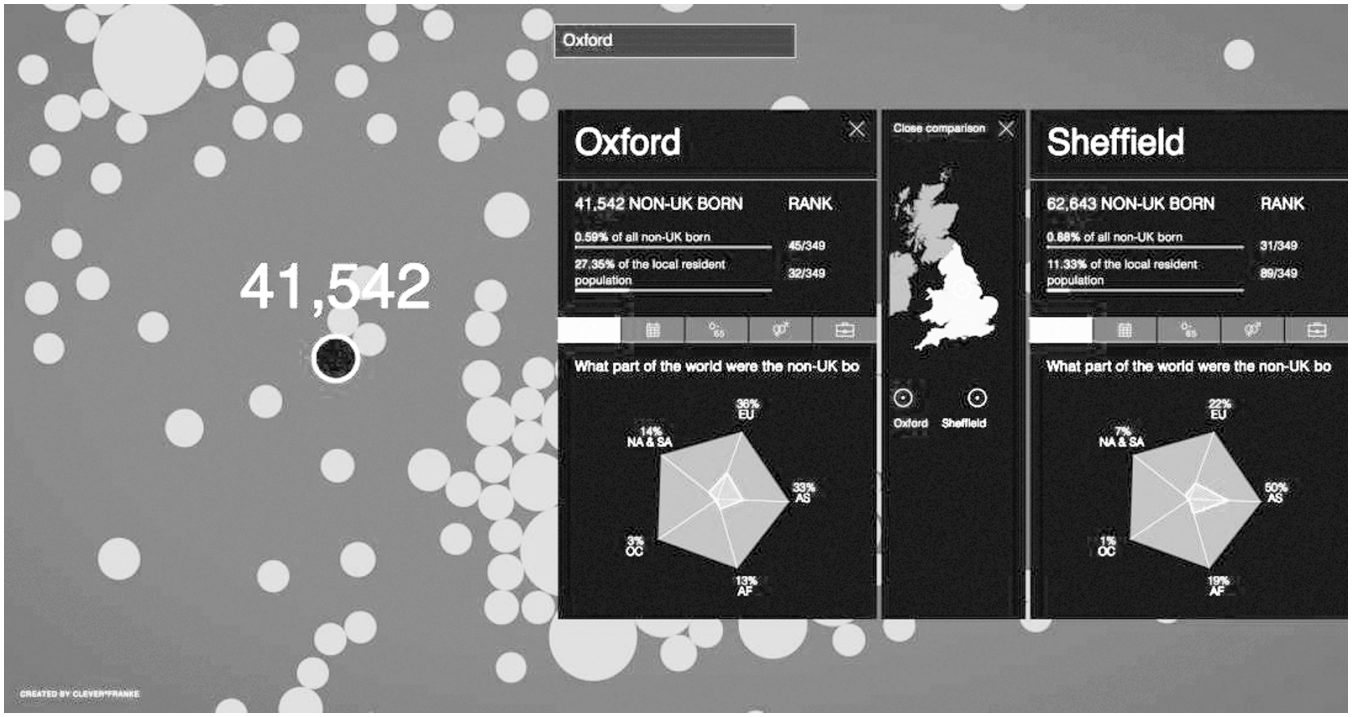
These features are illustrative of the Observatory's visualisation practice: given the potentially wide-ranging audiences noted earlier and the Observatory's value of independence, the goal is to give users a means to engage with the visualisation without suggesting a particular endpoint or conclusion. In Kirk's (2016) terms, this kind of map facilitates *exploration* rather than *explanation*: viewers can metaphorically fly around the country as their curiosities dictate. The maps enable users to *read* specific values by hovering over certain areas; they do not attempt to get viewers to *feel* the issue in charged or emotive ways. That the image is rendered in blue (visible in the online version) is in keeping with the Observatory's branding colour palette. As we found on Seeing Data, these colours evoke authority and objectivity. Furthermore, as we argue elsewhere (Kennedy *et al.,* 2016a), the convention of viewing a country from above appears to remove one kind of perspective and offers the user an ‘unemotional’ or ‘objective’ viewpoint. This advances the Observatory's mission of evidence-based, ‘neutral’ interventions within a politicised arena.

By dividing the data by region and highlighting patterns specific to particular areas, the Observatory also aimed for a high degree of personalisation in the visualisation, which also fits with its context and values. First, enabling audiences to locate information relevant to the places with which they are familiar is one way of engaging with them and making large datasets accessible. Second, instead of imposing an arbitrary cut-off point based on population sizes, all English and Welsh local authorities are included, and this adheres to the value of comprehensiveness. Third, the interactive maps are standalone items, with little editorial or annotation that supports a single interpretation or perspective, which adheres to the value of independence. Thus the Observatory handles a politicised issue as transparently as possible (even though data are never really neutral, as we noted previously).

In collaboration with visualisation agency Clever Franke and as part of Seeing Data research, the Observatory subsequently expanded on these qualities of exploration and personalisation to produce a more comprehensive visualisation of the same data. This second visualisation added features such as searching by local authority name and comparing two areas side-by-side. It also represented other aspects of foreign-born people that the Census asked about, such as region of origin, sex, time period of arrival and employment status. Figure 18.6 illustrates how the visualisation displays key data about the size of the foreign-born population in each authority. Users can also choose to compare other demographic dimensions, or click around the circles that represent different local authorities and explore their curiosities. These extended exploratory and customizable features also fit with the Observatory's values and aims of providing comprehensive and independent information: the visualisation provides users with an opportunity to engage with the data in a self-directed manner and to access precise values relating to whichever local authority they wished.[3] ‘Migration in the Census’ shows how the process of visualising a large and complicated dataset is governed by motivating factors that influence the appearance of the final visualisation.

Figure 18.6Screenshot of ‘Migration in the Census’ (comparative view)

Source: www.compas.ox.ac.uk/migrationinthecensus



Case Study Two: Migration in the News

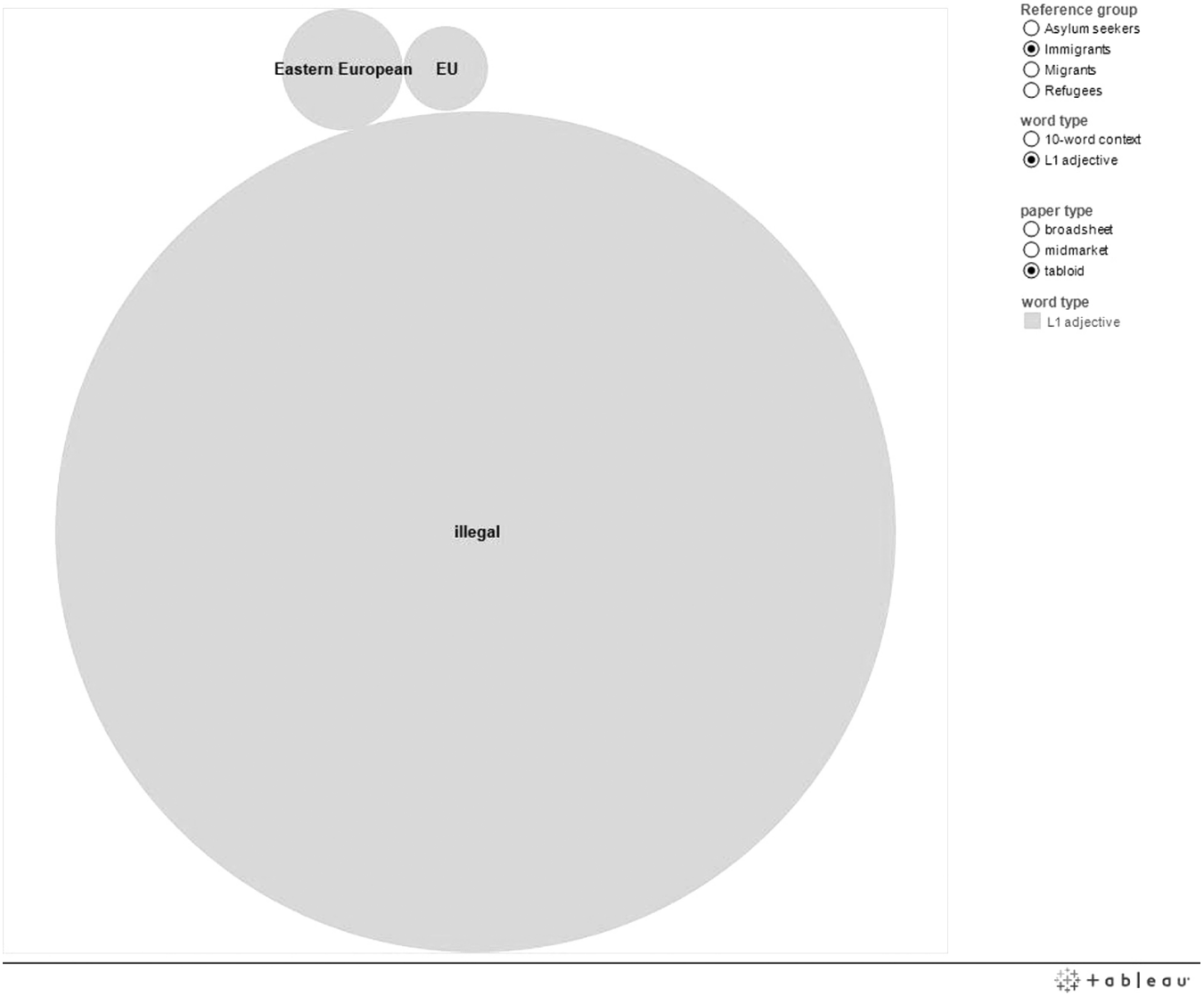
The second case study relates to Observatory research into the ways in which the British press describes migrant groups, including asylum seekers and refugees. Based on a landmark project in linguistics that examined a similar question over the 1996–2005 period (Gabrielatos and Baker, 2008), ‘Migration in the News’ extended this research to look first at the 2010–2012 period, and then later at 2006–2013. This section describes how visualisations were developed on this project, the kinds of decisions made along the way and their impacts on eventual outputs.[4]

The rationale for the ‘Migration in the Media’ project relates to the Observatory value of comprehensiveness: that is, a belief in the importance of understanding how the press as a whole describes migrant groups (Allen, 2014). In its first stage, the project aimed to collect, as far as possible, all news coverage mentioning key migration terms like ‘immigrant', ‘asylum', ‘deportation’ and ‘refugees'. This totalled 58,351 items, comprising nearly 44 million words. Then, using statistical tests, it identified which words consistently described the terms ‘immigrant(s)', ‘migrant(s)', ‘asylum seeker(s)’ and ‘refugee(s)'. These descriptors are known as ‘modifiers'. The two main findings were that immigrants were most consistently described as ‘illegal’ during this period, while asylum seekers were described as ‘failed’ (Blinder and Allen, 2016).

Visualising these results presented some challenges. Qualitative data about modifiers needed to be linked with quantitative data about their frequencies. Furthermore, these results were different across three types

of publication: tabloids, which tend to focus on entertainment or celebrity news; broadsheets which tend to represent ‘traditional’ news reporting or journalism; and mid-markets which typically have a mix of both. For example, how could a visualisation show that ‘illegal’ was far and away the most frequent, consistent modifier for ‘immigrants’ in the tabloid press? Figure 18.7 shows an early attempt to visualise this finding, using the ‘bubble chart’ feature in Tableau Public.

Figure 18.7Modifiers of ‘immigrants’ in the tabloid press, 2010–2012



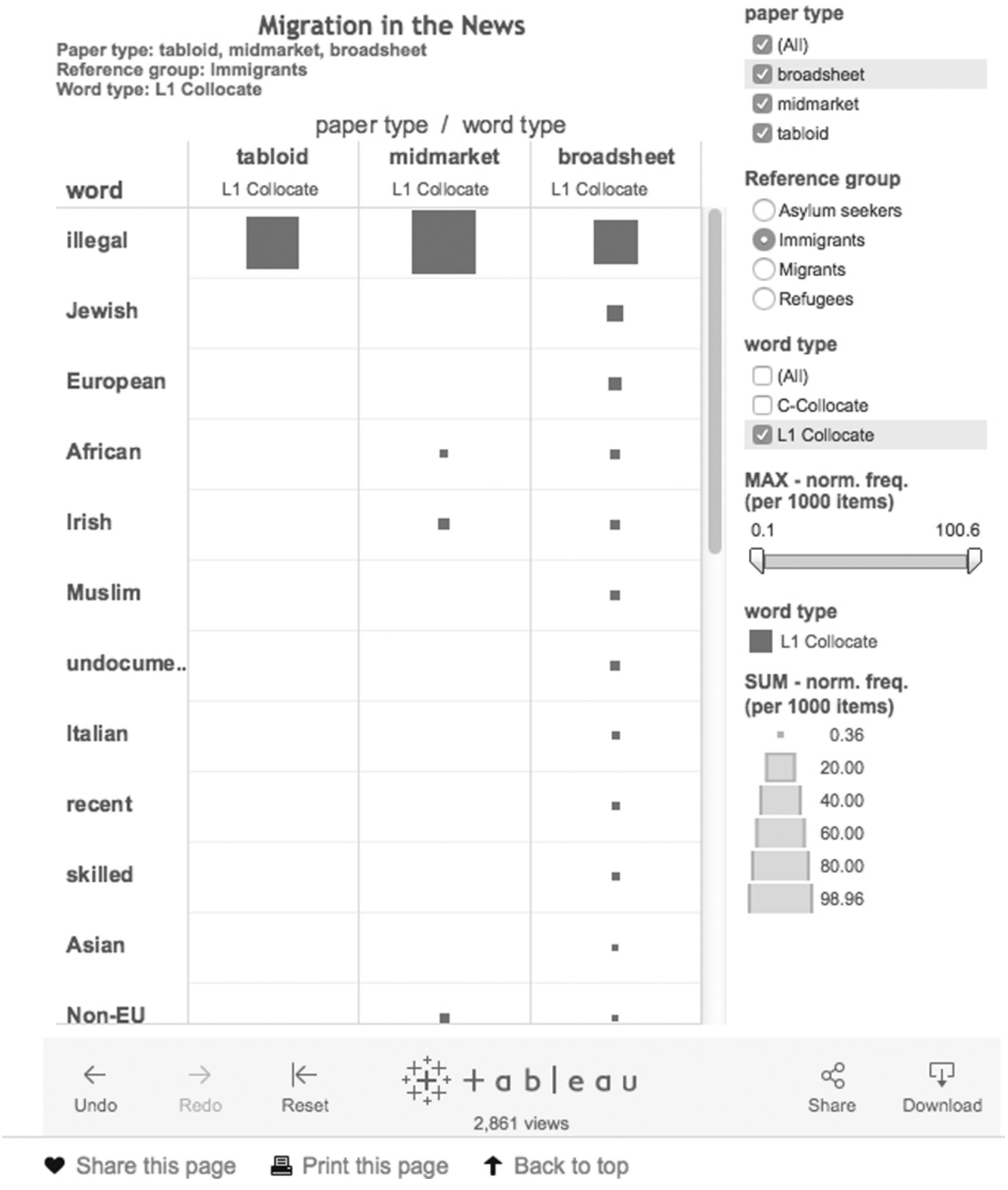
On the one hand, this visualisation clearly makes its point, a point so large as to be unmissable. The word ‘illegal’ visibly modifies ‘immigrants’ much more than ‘Eastern European’ or ‘EU'. But on the other hand, there are some problems with it. The size of the ‘illegal’ bubble suggests that this term is remarkably frequent, but comparison is limited to two other terms. What's more, the visualisation did not allow simultaneous comparison across publication types. In addition, the human brain has difficulty in determining the area of circles compared to the features of other shapes like the lengths of bars (Spence and Lewandowsky, 1991). So, although this cluster of circles had some appeal, the visualisation could potentially introduce some

confusion, which goes against the Observatory's value of clarity.

In a subsequent iteration, squares replaced circles, a legend gave users information about how differently sized squares related to the frequency of modifiers, and features such as toggles for publication types enabled users to compare different subsets of the press. The Observatory added annotation, which it believed would help users make sense of some of its representational decisions. Figure 18.8 shows a screenshot of the visualisation as it currently exists, showing the prevalence of ‘illegal’ as a descriptor of migrants across the British press.[5] At the same time, as in the first ‘Migration in the Census’ visualisation, this output also aimed to give users control over its presentational form, depending on their interests. And, as with the local authority-level data, it also reports the precise frequencies of each modifier when the user hovers over the relevant square, allowing users to read values efficiently.

Figure 18.8Modifiers of ‘immigrants', all publication types, 2010–2012

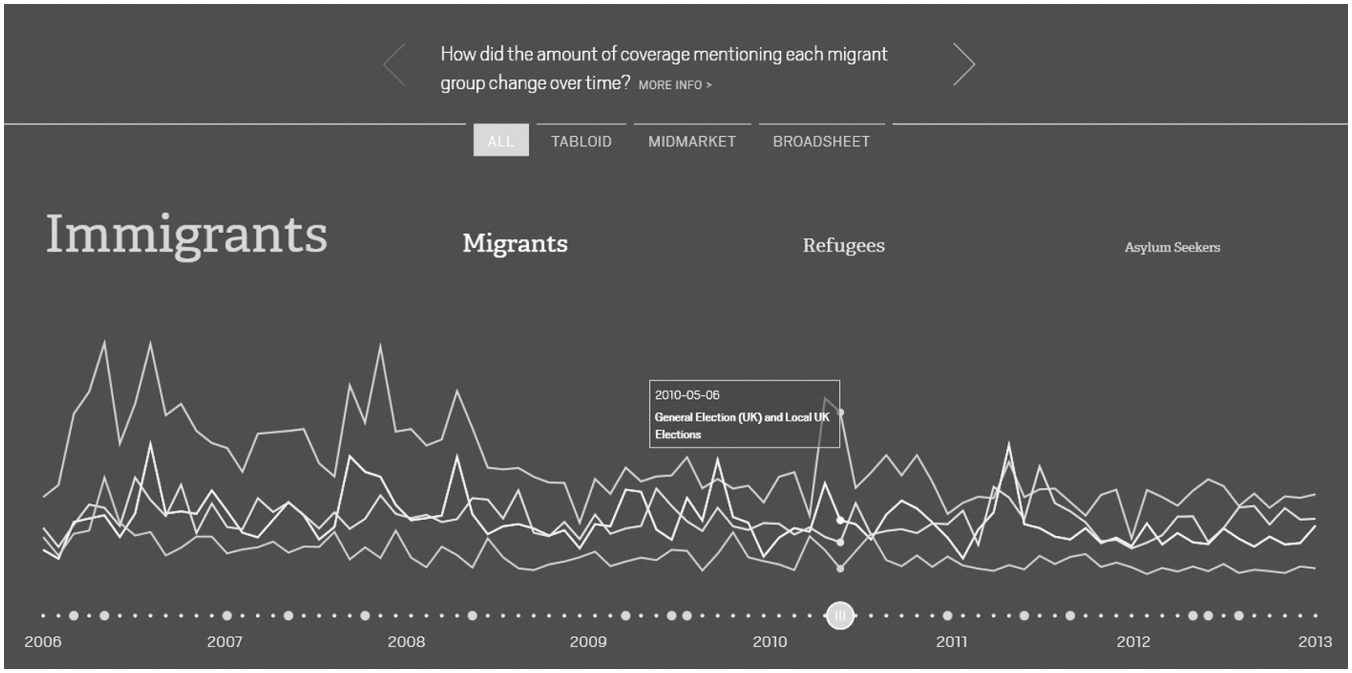
Source: http://public.tableau.com/views/MigrationintheNews/MigrationintheNewsinteractive?:embed=y&:loadOrderID=0&:display_count=yes



After completing this study of the 2010–2012 period, the Observatory expanded its perspective longitudinally to include all available national UK newspaper coverage mentioning the same set of key terms from 2006 to 2013 – totalling about 90 million words. Again in collaboration with Clever Franke, this stage aimed to show how the frequency of mentions of each migrant group changed in the press over time, as well as how the kinds of modifiers used to describe each group differed among subsets of the press. Unlike the original visualisation in Figure 18.8 that mainly emphasised *reading* data, this newer visualisation attended to *feeling* data by communicating overall impressions about the nature of press coverage. The goal was to show how press coverage had changed over time, rather than show static, aggregated results, and so seeing the direction and general scale of these changes was more important than reading precise figures.

Figure 18.9 shows a portion of the visualisation. It depicts the changing frequencies of each key term – ‘immigrants', ‘migrants', ‘asylum seekers’ and ‘refugees’ – over the seven-year period. As the user moves along the timeline, the relative size of each key word increases and decreases according to its frequency in the dataset. These key words are differentiated by colour using the same palette from the Observatory's other publications: the colouration of each word signals which line in the underlying chart it corresponds with, allowing the viewer, at a glance, to get a sense of which terms were used more frequently without reading specific data points. Users can also choose different subsets of the press, or get more information about the methods by which the data were generated by clicking ‘more info'. Additional contextual information is available through annotations connected to nodes on the timeline. These signal, for example, when elections occurred or when particularly important policies involving immigration became law. Some of these project and chart annotations aim to be explanatory, but they also aim to make the visualisation trustworthy and engaging, given that explaining data handling processes is an essential component of good, professional and honest visualisation practice (Kirk, 2016).

Figure 18.9Screenshot of frequency analysis with annotation, all publications (a work-in-progress version of the interactive 'Migration in the News' visualisation can be found at www.compas.ox.ac.uk/migrationinthenews)



Another section of the visualisation attempts to give a sense of how different modifiers are associated with each key term. In the earlier visualisation, frequency was indicated through the size of squares and users could pick out specific values by hovering over them. This visualisation, in contrast, depicts frequency through the saturation of colour, as seen in Figure 18.10 (and more clearly online www.compas.ox.ac.uk/migrationinthenews). Throughout the visualisation, the same colours express findings about the same key terms, to guide the user through the different parts. Also, because the project focused on differences over time, the top 100 modifiers were included in the analysis where possible. The intended effect was a 'wall of words', where frequent modifiers would stand out by glowing more brightly. Here again the difference between feeling and reading numbers is important: although the visualisation does not show precisely how much more frequently 'failed' modifies 'asylum seekers' compared to, say, 'child' or 'genuine', users can immediately sense which words are most frequent without referring to precise figures.

Figure 18.10Screenshot of modifiers of ‘asylum seekers', all publications

These two case study projects illustrate how the Observatory uses visualisations to achieve its aims, and how decisions and thinking about some of the issues discussed earlier in this chapter align with the Observatory's values. Enabling exploration of extended datasets aligns with comprehensiveness, and presenting region-specific snapshots is a way of engaging with audiences who might have more local rather than national interests. Highlighting how these datasets are limited – what the data can and cannot say – aligns with authoritativeness, and project annotations that make complex statistics understandable aim to achieve clarity. The overall project of including lots of different kinds of data, in ways that are customisable yet clearly marked, aims to present an independent view of this complicated issue. These examples illustrate how key questions about visualisation such as ‘what works', ‘for whom’ and ‘under what circumstances’ are addressed in practice. Thinking critically about its own visualisation practice, iterating and revising visualisations accordingly (and if feasible) are integral to the ways in which the Observatory makes and shares its visualisations.

Towards Good Data Visualisation Practice for Online Researchers

One aspect of data visualisation that we have not discussed in any length in this chapter is users and audiences. This is somewhat ironic given that our main research interest in this field relates to the views of users, who we consider to be either overlooked or decontextualised in studies of information and data visualisation. By this we mean that visualisation research which *does* involve users provides little information about them and ignores socio-cultural and contextual factors of the kinds discussed in the previous section, which play a significant role in visualisation consumption and engagement. Our research was therefore informed by our critique of psychology-influenced studies of visualisation reception which aim to separate out perceptual processes from the messy contexts in which they take place. The research that we undertook confirmed that a range of socio-cultural factors affect engagement, including the subject matter of the

visualisation, its original location or media source, users' beliefs and opinions, the time that users have at their disposal to explore visualised data, users' confidence in the skills that they think they need to make sense of a visualisation (such as statistical and visual literacy, language skills and critical thinking skills) and the emotional dimensions of engaging with diverse elements of a visualisation (aesthetics, the data themselves, the subject matter and the source). We report extensively on these findings elsewhere (Kennedy, 2015; Kennedy *et al.,* under revision).

For online researchers, the primary audience for our data visualisations will often be peers who encounter them in conference presentations, journal articles or other scholarly publications. But visualisation is also a method for communicating our research to broader, non-specialist audiences, something that is increasingly significant in the UK context because having an impact beyond our scholarly communities is an ever more important measure of research excellence. Researchers producing data visualisations for both expert peers and non-expert wider publics need to be attentive to who their audiences are and what needs they have. They also need to navigate the ways in which the factors discussed in the previous paragraph might impact on the user groups' engagement with visualisations.

Data visualisation promises to make data accessible and transparent to broad audiences, and it certainly has the potential to do that. But it is not a simple window onto data: online researchers need to be alert to what might be called the politics of data visualisation – that is, the work that visualisations do to communicate data in certain ways – and in so doing, to produce visualisations reflectively, in ways that are aware of contexts and audiences, as well as more pragmatic questions of what tools to use and what chart types to include. Being aware of all of these things and so producing good visualisations does not come easily, and the provision of resources to enable researchers to develop these skills is also a political issue. We have suggested some solutions to some of these problems, such as abiding by professional principles or engaging in practices which help users become aware of the choices made in producing a visualisation (for example, including annotations such as a link to a data source or an explanation of the ways in which data have been treated). If we have one message for readers of this chapter, it is to be attentive to what it is that we are doing when we visualise our data, for whom, in what contexts and with what effects.

Notes

1. This article was itself a response to a blogpost called, 'Why you should never trust a data scientist' (Warden 2013).

2. Full details of the project and the whole suite of materials are available at http://www.migrationobservatory.ox.ac.uk/projects/census

3. As long as that local authority was English or Welsh: a key limitation of this visualisation was its exclusion of Scotland and Northern Ireland.

4. For more details about the 2010–2012 dataset, methods and results, see Allen (2014) and Blinder and

Allen (2016).

5. Full details of the initial project (covering 2010–2012) and the accompanying visualisation itself can be found at http://migrationobservatory.ox.ac.uk/projects/media

References

Allen, W. (2014) 'Does comprehensiveness matter? Reflections on analysing and visualising UK press portrayals of migrant groups'. Paper presented at the Computation Journalism Symposium, 24–25 October, Columbia University, New York.

Allen, W. (2016) 'Factors that impact how civil society intermediaries perceive evidence', *Evidence & Policy: A Journal of Research, Debate and Practice*. DOI: http://dx.doi.org/10.1332/174426416X14538259555968

Ambrosio, C. (2015) *'Objectivity and representative practices across artistic and scientific visualization'*. In **A. Carusi**, **A. S. Hoel**, **T. Webmoor** and **S. Woolgar** (eds.), *Visualization in the Age of Computerization* (pp. 118–44). London: Routledge.

Bhaskar, R. (1975) *A Realist Theory of Science*. London: Verso.

Blinder, S. and **Allen, W.** (2016) 'Constructing immigrants: portrayals of migrant groups in British national newspapers, 2010–2012', *International Migration Review*, 50: 3–40.

Bollier, D. (2010) *The Promise and Peril of Big Data*. Available at http://www.aspeninstitute.org/publications/promise-peril-big-data

Bowker, G. (2005) *Memory Practices in the Sciences.* Cambridge, MA: MIT Press.

Burn-Murdoch, J. (2013) 'Why you should never trust a data visualisation', *Guardian.* Available at http://www.theguardian.com/news/datablog/2013/jul/24/why-you-should-never-trust-a-data-visualisation

Cairo, A. (2012) *The Functional Art: An Introduction to Information Graphics and Visualization*. San Francisco, CA: New Riders.

Cotgreave, A. (2014) 'Controlling your message with title, colour and orientation', Gravy Anecdote, 16 October. http://gravyanecdote.com/uncategorized/should-you-trust-a-data-visualisation

Crawford, K. (2013) 'Algorithmic illusions: hidden biases of big data'. Presentation at the Strata Conference, February 26–28, 2013. Available at https://www.youtube.com/watch?v=irP5RCdpilc

Dick, M. (2015) 'Just fancy that: an analysis of infographic propaganda in *The Daily Express*, 1956–1959', *Journalism Studies,* 16(2), 152–74.

Few, S. (2008) 'What ordinary people need most from information visualization today', *Perceptual Edge: Visual Business Intelligence Newsletter*. Available at: http://www.perceptualedge.com/articles/visual_business_intelligence/what_people_need_from_infovis.pdf

Few, S. (2012) *Show Me The Numbers: Designing Tables and Graphs to Enlighten.* Burlingame, CA: Analytics Press.

Fyfe, G. and **Law, J.** (eds.) (1988) *Picturing Power: Visual Depiction and Social Relations*. London: Routledge.

Gabrielatos, C. and **Baker, P.** (2008) 'Fleeing, sneaking, flooding: A corpus analysis of discursive constructions of refugees and asylum seekers in the UK press, 1996–2005', *Journal of English Linguistics*, 36(1), 5–38.

Gill, R. (2009) *'Breaking the silence: the hidden injuries of neo-liberal academia'*. In **R. Ryan-Flood** and **R. Gill** (eds.), *Secrecy and Silence in the Research Process: Feminist Reflections.* London: Routledge, pp. 228–44.

Hill, R. L. (2014) 'How social semiotics can help us understand big data visualisations'. Seeing Data. Available at http://seeingdata.org/how-social-semiotics-can-help-us-understand-big-data-visualisations/

Ipsos MORI. (2016) 'Economist/Ipsos MORI January 2016 Issues Index: concern about immigration returns to prominence'. Available at: https://www.ipsos-mori.com/researchpublications/researcharchive/3691/Economist-Ipsos-MORI-January-2016-Issues-index.aspx

Kennedy, H. (2014) 'How do visualization professionals think about users?'. Seeing Data. Available at http://seeingdata.org/visualization-professionals-think-users/

Kennedy, H. (2015) 'Seeing data: visualisation design should consider how we respond to data emotionally as well as rationally'. LSE Impact Blog. Available at http://blogs.lse.ac.uk/impactofsocialsciences/2015/07/22/seeing-data-how-people-engage-with-data-visualisations

Kennedy, H. (2016) *Post, Mine, Repeat: Social Media Data Mining Becomes Ordinary.* Basingstoke, UK: Palgrave Macmillan.

Kennedy, H. and **Hill, R.L.** (2016) 'The pleasure and pain of visualising data in times of data power', *Television and New Media.* Available at: http://tvn.sagepub.com/content/early/2016/09/07/1527476416667823.abstract.

Kennedy, H., **Hill, R. L.**, **Aiello, G.** and **Allen, W.** (2016a) 'The work that visualisation conventions do', *Information, Communication and Society*, 19(6): 715–735.

Kennedy, H., **Hill, R. L.**, **Allen, W.** and **Kirk, A.** (under revision) '(Big) data visualisations and their users:

how socio-cultural factors challenge existing definitions of effectiveness', *First Monday*.

Kirk, A. (2016) *Data Visualisation: A Handbook for Data Driven Design*. London: Sage Publications.

Korzybski, A. (1931) 'A non-Aristotelian system and its necessity for rigour in mathematics and physics'. Presentation at the American Mathematical Society/American Association for the Advancement of Science, New Orleans, Louisiana, 28 December 1931. Reprinted in *Science and Sanity*, 1933: 747–61.

Neurath, O., **Neurath, M.** and **Cohen, R. S.** (1973) *Empiricism and Sociology.* Vol. 1. Dordrecht: Reidel.

Periscopic. (2014) Homepage. Available at http://www.periscopic.com/

Porter, T. M. (1995) *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, NJ: Princeton University Press.

Ruppert, E. (2014) 'Visualising a journal: big data and society'. Paper presented at the ICS Visual & Digital Cultures Research Seminar, Leeds. 13th May 2014.

Scarr, S. (2011) *Iraq's Bloody Toll'*. South China Morning Post, 17 December. Available at www.scmp.com/infographics/article/1284683/iraqs-bloody-toll

Smith, A. (2016) 'Crafting charts that can withstand the data deluge', *Financial Times*. Available at http://www.ft.com/cms/s/0/3f195d40-b851–11e5-b151–8e15c9a029fb.html#axzz3yFKKUQ7M

Spence, I. and **Lewandowsky, S.** (1991) 'Displaying proportions and percentages', *Applied Cognitive Psychology*, 5(1): 61–77.

Stiftelsen Gapminder (n.d.) Gapminder. Available at http://www.gapminder.org/

Tufte, E. R. (1983) *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press.

Valarakis, A. (2014) *On Data Visualization: Rhetoric and the Revival of the Body Politic.* (MA), University of Amsterdam.

Warden, P. (2013) 'Why you should never trust a data scientist', Pete Warden's blog. Available at http://petewarden.com/2013/07/18/why-you-should-never-trust-a-data-scientist

Yau, N. (2013) *Data Points: Visualization That Means Something.* New York, NY: John Wiley & Sons.

Zambrano, R. N. and **Engelhardt, Y.** (2008). *'Diagrams for the masses'*. In **G. Stapleton**, **J. Howse** and **J. Lee** (eds.), *Diagrams 2008*. Berlin: Springer Verlag, pp. 282–92.

http://dx.doi.org/10.4135/9781473957992.n18