Visualising BIG Data

Carolyn King

This report looks at the disruptive effect big data is having on the visualisation of information, by first investigating the objectives of visualising data, then examining the attributes, uses and risks of big data, and finally analysing the opportunities in this area of visual communication design.

Turning data into wisdom

In itself, data means nothing. Data consists of raw values, and pieces of data (such as the number 39.9 or the symbol °) are useless without context.

Once put in context, data becomes **information**. So 39.9°F in a weather report indicates a cold day, while 39.9°C on a patient's thermometer indicates a high-body temperature. The same data in different contexts gives different meanings.

When information is processed cognitively, organised and validated (implicitly through experience, or explicitly through guidance and education) it becomes **knowledge** (Cooper, 2017). Knowledge tells us that a patient whose temperature is 39.9°C is most likely feverish and in need of medical attention. The original data now has meaning and can be acted on.

The final step in this transition is to turn knowledge into **wisdom**, by applying ethical principles and judgment. Wisdom might suggest in some cases that a patient with a thermometer reading of 39.9°C who does not show other signs of fever should be retested, especially if the patient's history shows they have been known to dip their thermometer in a cup of tea to get attention. Data can lead to false assumptions and unwise actions if the full context is not known.

What is data visualisation?

Data visualisation is the presentation of data in a visual format. Visualisation can help make data more accessible by representing it in way that enhances understanding, putting it into a human context. Data can be visualised as maps, charts, tables, infographics, interactive models, films, animations, 3D objects or even performances.
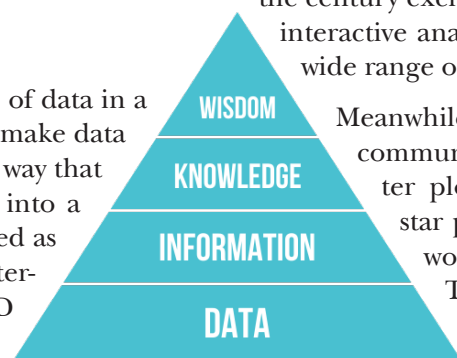
The aim of **data visualisation** is to enable wisdom—convert data into knowledge that can be applied to achieve a **purpose**. Data visualisation tells a story, so its design should consider who will access the data, when/where/how they will use it, and most importantly why they need it and what **insights** they are looking for.

A brief history of data visualisation

Humans have been translating their perception of the world into graphical forms and visualising quantitative information in the form of maps, diagrams and symbols since early civilisation. During the Renaissance period, the humanist approach of scholars, artists and scientists triggered development of many techniques for observing, interpreting and displaying data.

The 17th to 19th centuries saw the emergence of new theoretical approaches such as probability theory and demographic statistics, and data relating to social, scientific and economic statistics was gathered in a more systematic way. The value of such data became widely recognised by governments, health professionals and economists (Friendly & Denis, 2001).

The advent of computers in the 20th century enabled rapid computation of statistics and new ways of analysing and displaying information, and by the end of the century exchange of data via the Internet allowed interactive analysis and visualisation of data from a wide range of sources.

Meanwhile, theories and techniques for visual communication of data emerged, such as scatter plots, bivariate analyses, Gantt charts, star plots, timelines, interactive maps and word clouds (Friendly & Denis, 2001). These have become part of the visual language and toolsets now used for data visualisation.
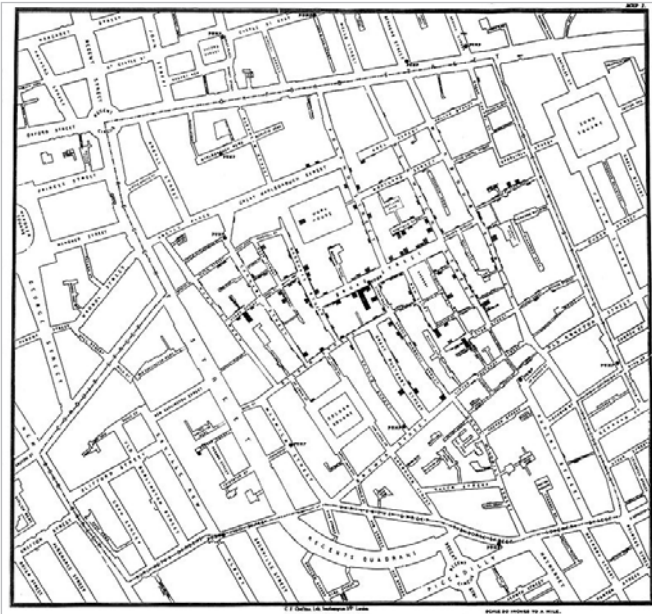

Figure 1 DIKW pyramid (Longlivetheux, 2015)

Figure 2 John Snow's dot map of cholera case locations was able to pinpoint a water pump causing deaths and show the link between water quality and cholera (Snow, 1854).

The growth of digital data

Digital data has grown exponentially in recent years, at the rate of 2.5 exabytes a day (an exabyte is a billion billion bytes). According to IBM, 90% of all the world's data was created in the last two years (IBM, n.d.).

A turning point in data growth was the emergence in the early 2000s of 'Web 2.0' services based on sharing and collaboration by users—social media, online networks, wikis and communication tools. This meant a lot of data was now coming direct from consumers and could be used by service providers to customise content, target messages, get instant feedback and identify trends (Peltier-Davis, 2015).

Another paradigm shift was the mobile revolution from the mid 2000's onwards, with people accessing and transmitting all sorts of data on the move via apps, websites and other usage/location data from mobile devices. There are currently over 5.1 billion mobile subscribers, which means nearly 70% of the world's population is producing and consuming mobile data in some form (GSMA Intelligence, 2018).

Technologically, growth in data has been driven by:

- The dramatic increase in data-generating devices (not just phones, but Internet-connected devices with sensors transmitting data)
- The wide availability of networks and technologies to connect devices and transmit data
- A greatly enhanced capacity to store massive amounts of data.

Commercially, data growth has come from businesses' drive to:

- Optimise or streamline operations
- Monetise digital assets
- Exploit market intelligence
- Improve customer experiences (Penkler, 2018).

Growth in visualisation tools

Coupled with these growth factors, there is now a proliferation of tools for dissecting, analysing and displaying data. Data mining has developed into a science, and data visualisation has become the art of explaining that science to people, from researchers through to consumers. Data visualisations are shared across a range of channels, from mainstream media to education platforms, business software, professional networks and social media.
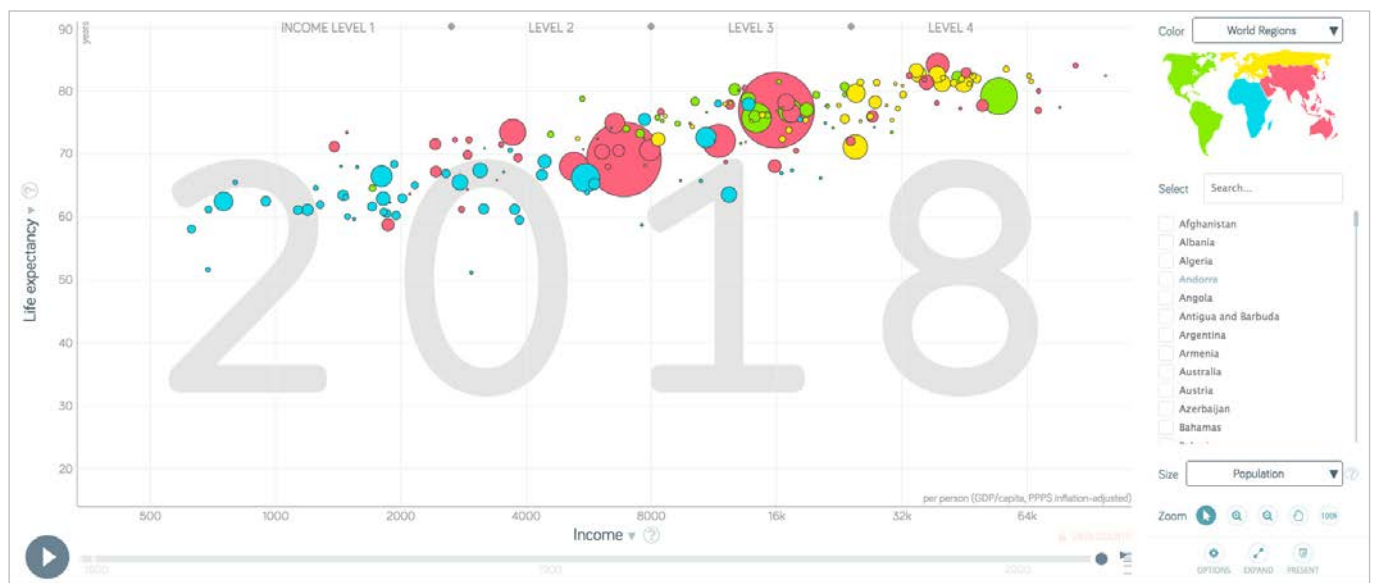

Figure 3 Visualisation linking income to life expectancy, developed by Hans Rosling and now available as an interactive plot. The bubble chart assigns four variables to each country: life expectancy (y-axis), GDP (x-axis), continent (colour), and population (bubble size). Clicking the play button shows changes over time from 1800 to the current day (Gapminder, 2018).

What characterises big data?

The term big data refers to the sheer scale of digital data available, and is often described using the three Vs of Volume, Velocity and Variety (Mediratta, 2015).

- **Volume** refers to the amount of data collected, stored and processed. Advances in data storage and processing systems allow for extremely high volumes of data and large files to be stored and exchanged.
- **Velocity** means the speed with which data is produced, transferred and analysed. Data is generated 24/7, no longer just in business hours, and can be monitored in real time via 'streaming analysis'.
- **Variety** is another key aspect of big data. Data comes not just from databases and online applications, but also from a variety of real world and digital contexts: social media, emails, audio, video, GPS systems and countless Internet-connected devices.

While the three Vs help explain technical aspects of big data, three further V's describe business and communication perspectives of big data:

- **Variability:** Is the nature, context and structure of the data consistent and suitable for analysis against a relevant model?
- **Veracity:** Is the data trustworthy, valid and fit for purpose? Can its source be verified?
- **Value:** What benefits or insights will come from analysing the data? Does that meet ethical and compliance requirements? What is the desired outcome?

One more 'V' could be added to these: **Visualisation**. When data is put in a visual context, it becomes information that **tells a story**, which can build knowledge and enable informed decision-making.

How big data is used and visualised

Big data feeds directly into decision support systems used by commercial and government organisations to analyse trends, plan strategies and adjust tactics in response to real time or historic market information. It is used by industry to monitor equipment and resources, and by consumers to monitor fitness, health, finances and other aspects of everyday life.

Visualisation of data for these purposes can take many forms, but is often in the form of a dashboard showing key metrics, highlighting trends or issues. Within a dashboard, data can be displayed using charts, tables, icons, images or interactive visualisations that can be manipulated by users. The design of these dashboards requires an understanding of user needs and business requirements. It has been suggested that a 'design thinking' approach to dashboard design, where the needs of users are at the centre, is the best way to approach this. (Cahyadi & Prananto, 2015)

In science and academia, big data is used to inform research projects. In their paper '*Why Big Data Isn't Enough*', Chai and Shih say there has been a trend for researchers to delve into data looking for patterns, relationships and interesting stories, rather than first creating a hypothesis and then examining data to find causal relationships. They argue that data-mining research methods have risks and should be seen as supplemental to existing methods (Chai & Shih, 2017).

> *"The assumption is: The bigger the data, the more powerful the findings. As appealing as this viewpoint may be, we think it's misguided."*
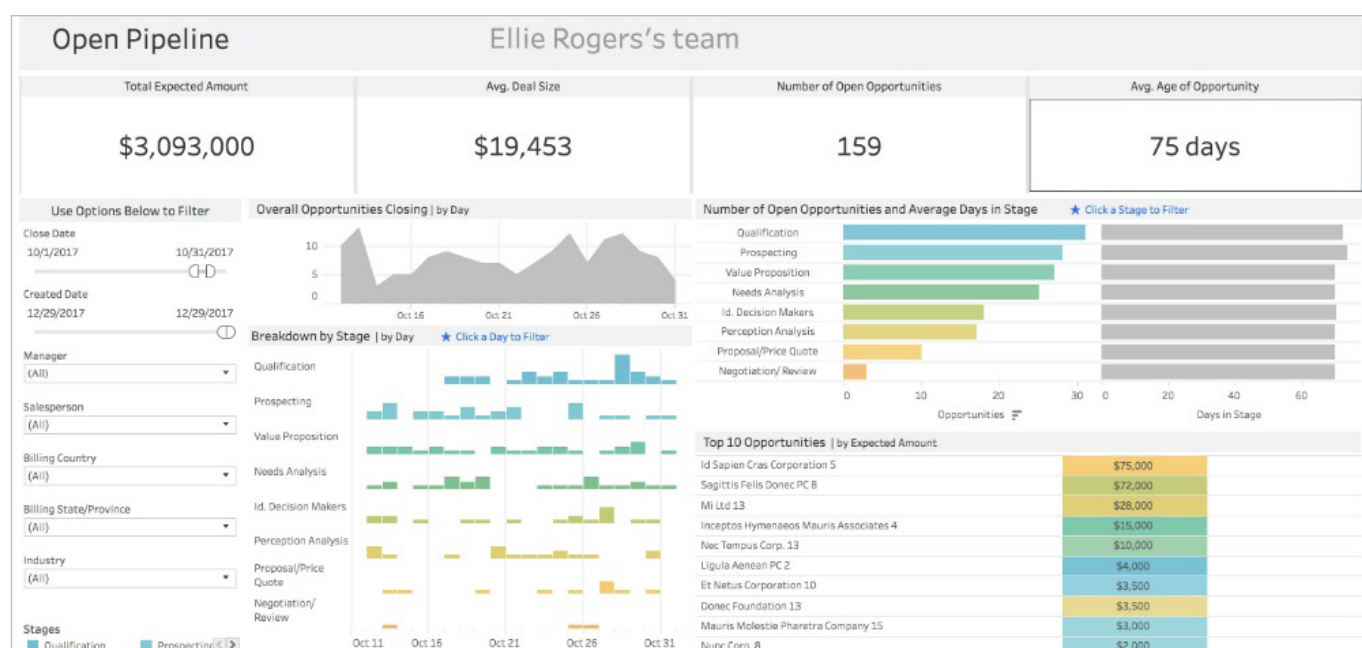
(Chai & Shih, 2017, p. 67)

Figure 4 Example of a business data dashboard (Tableau, n.d.).

Consumers are also generators and users of big data (whether or not they are aware of it). They use big data to support everyday tasks such as finding the best price for a product, comparing ratings for accommodation, or deciding whether to watch a video. For many online services, the consumer becomes the product, as it is the data they create that shapes the service, becoming a highly valuable currency.

Data from or for consumers is visualised in many ways, from simple graphics such as time/date counters and online poll charts, to more complex visualisations like Socialab's LinkedIn network diagrams (generated from a user's account data) or Strava's 'fly-by' interactive maps (generated from data uploaded by users' GPS-enabled fitness devices).
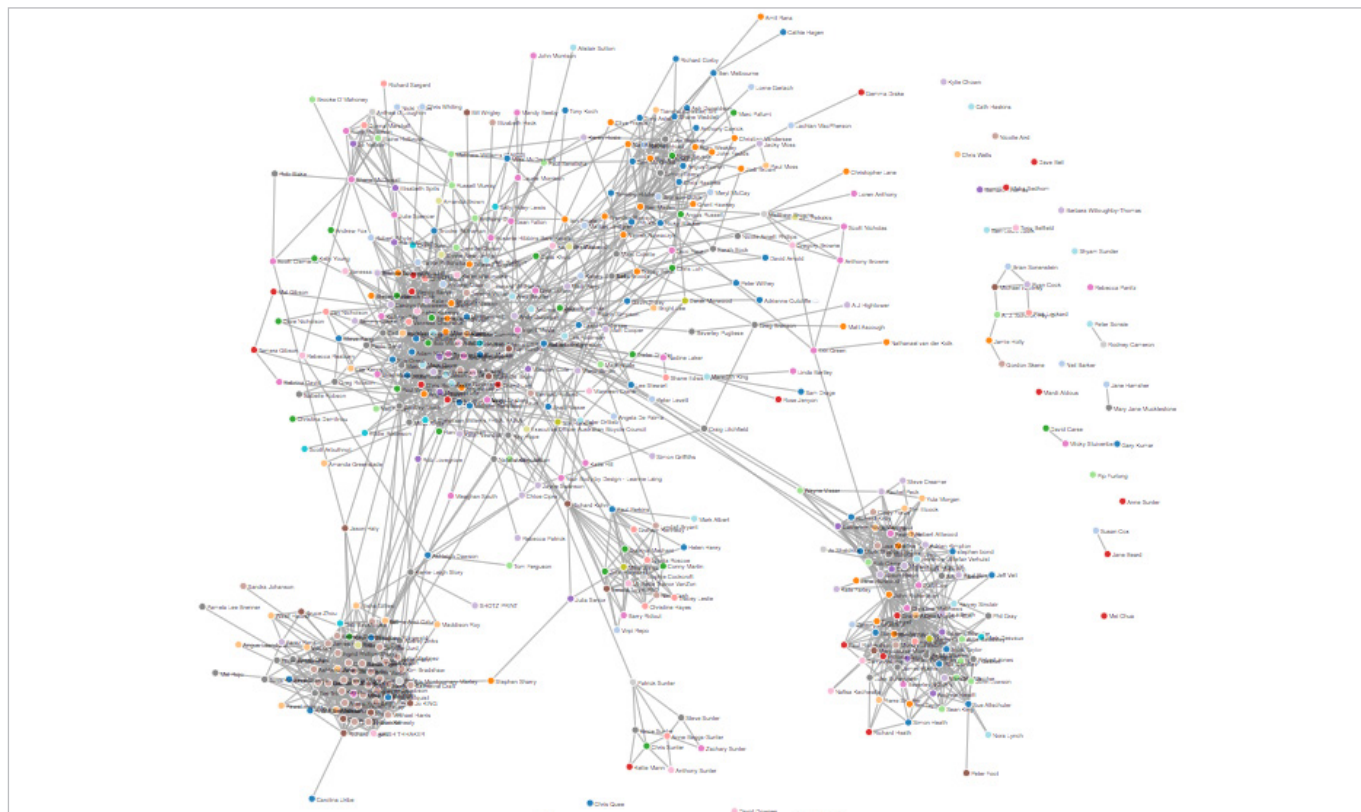

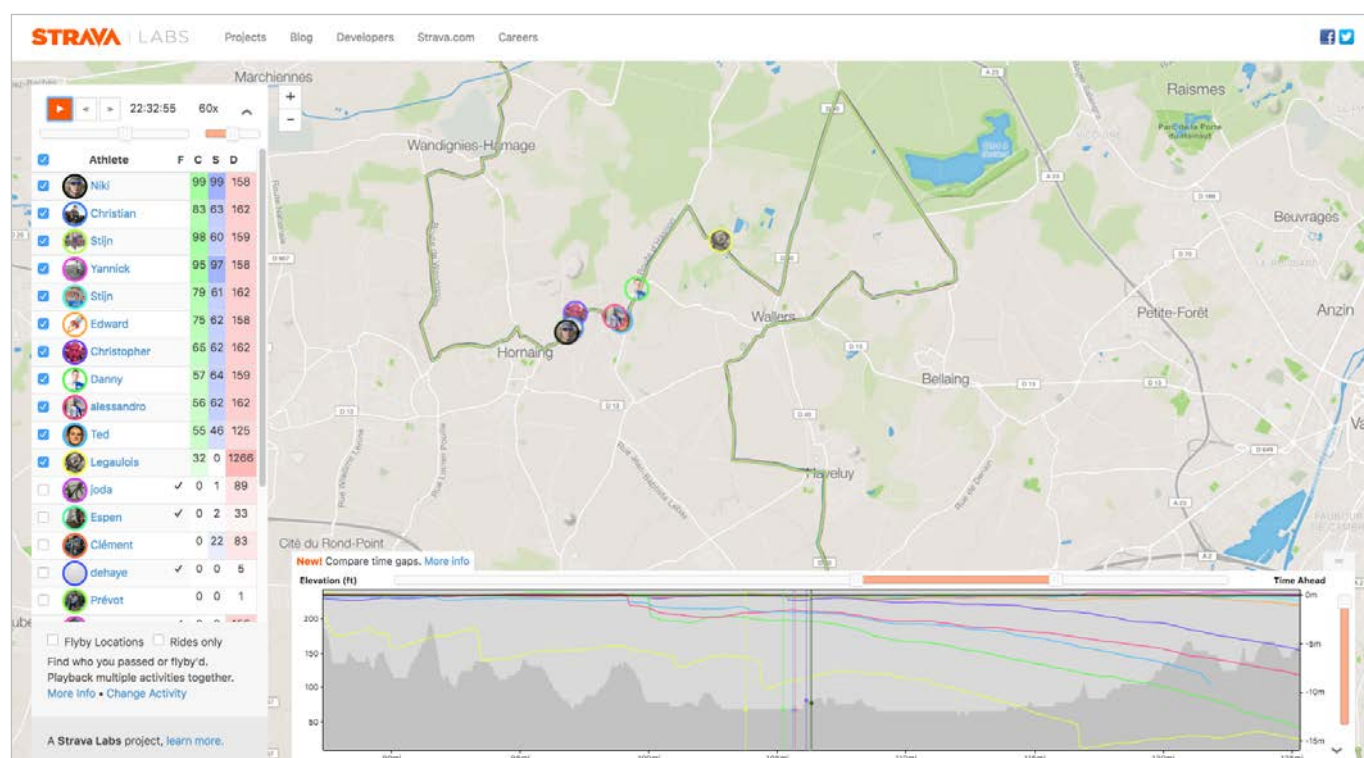**Figure 5** Visualisation of a user's LinkedIn network (Socialab, 2018).

Figure 6 'Fly-by' map tracking movement of cyclists along a route (Strava, 2018).

Benefits of big data

The analysis and sharing of big data has the potential to improve society, the environment and the economy by monitoring important issues, predicting problems, shortening disaster response times and improving personal wellbeing or economic outcomes for individuals, businesses and governments. Here are three examples:

- **Health and fitness:** Real time wearable devices can capture a patient's health data (such as heart rate) and upload it automatically so healthcare providers can monitor the patient remotely. Fitness tracking devices let an individual monitor their exercise activities, promoting a proactive approach to health and fitness. On a wider scale, medical and fitness data can be combined with data from social media and census information to research correlations between diseases, environment and lifestyle factors (Mediratta, 2015). This data can also support social and urban planning.

- **Education:** Schools and universities can monitor student progress using up-to-the-minute data such as frequency of access to class materials and make interventions aimed at improving outcomes, by adjusting the learning approach or materials. At a broader level, cross-institution data can be analysed to get insights on the effectiveness of the education system, to suggest improvements.

- **Citizen science:** Non-specialists can contribute to scientific research by participating in mass experiments and tasks aimed at collecting or sorting data (National Geographic, 2015).
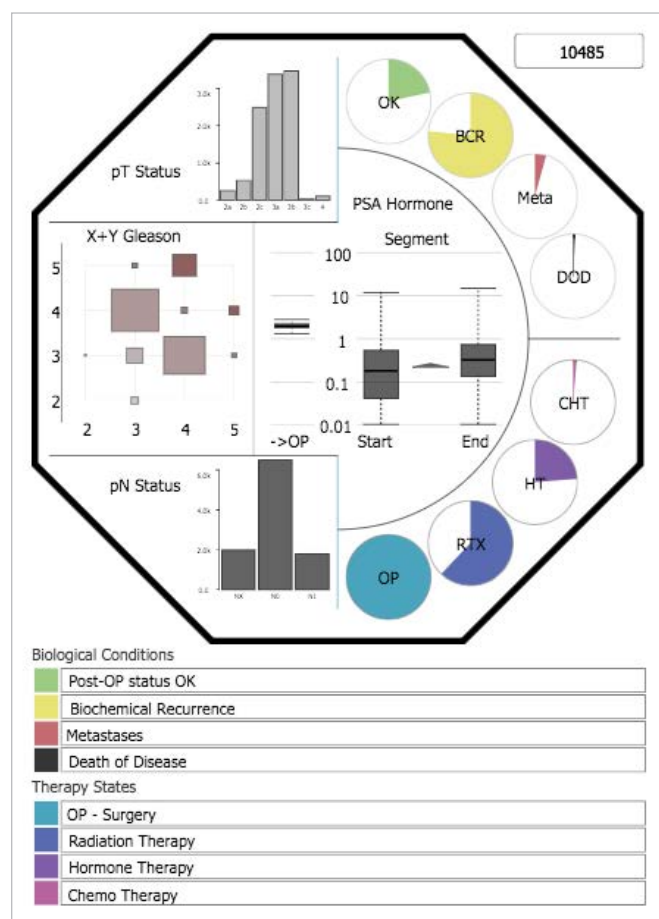

Figure 7 This visual dashboard provides an overview of over 10,000 patient histories in a way that can be analysed and understood by health specialists (Bernard et al., 2018).
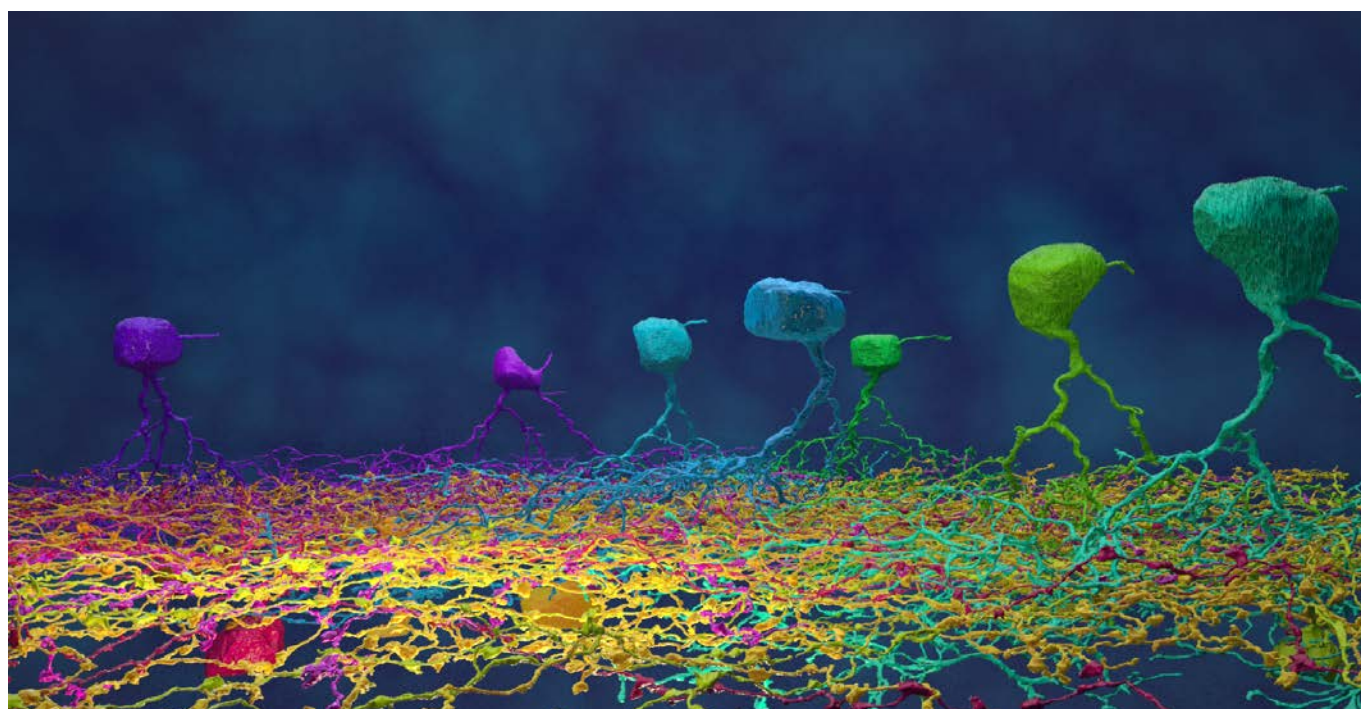

Figure 8 MIT's EyeWire project recruited thousands of gamers to help map neurones in the brain, and the resulting data has produced some stunning visualisations of neural networks (EyeWire, 2014).

Risks and problems of big data

Information is power, and power can be misused. This has always been the case with data. Mark Twain famously quoted Benjamin Disraeli as saying "There are three kinds of lies: lies, damned lies, and statistics" (Twain, 1906, p. 471). The difference with big data is the scale of the data and the speed with which it is shared, as the impact of errors will be magnified.

Information overload resulting from the constant stream of data in everyday life makes it very difficult for people to identify and select data that is relevant, actionable and trustworthy.

Potential problems of big data visualisation include:

- **Misinterpretation:** If data is shown out of its original context or the visual presentation is not clear, viewers may misunderstand the story being told. When visualisations are shared via blogs or social media they are often viewed out of context.

- **Unquestioning belief in data:** If people see a visual image, they want to believe it. "Seeing is believing because seeing is seduction", says Hepworth in a paper discussing promises and pitfalls of big data visualisation, "the experience of seeing is strongly correlated with truth" (Hepworth, 2017, p. 7). Hepworth argues the seductiveness of data visualisations should be considered carefully because this power to suggest truth can impact research quality.

- **Misrepresentation of data:** Data can be misrepresented, either intentionally or unintentionally, through the way it is visualised, the range of data selected, the scale or units of data compared, or the commentary attached to it.

- **Privacy, security and ethics:** Individuals have limited control over how their data is used or shared. Data crosses national boundaries and each country has different laws in relation to privacy and data use.

- **Selective views of data:** Not all data is created equal. Those who distribute data act as mediators. "Facebook does not neutrally relay messages. It collects, organizes, and relays posts and advertisements based on internal analytics that maximize engagement and positive responses" (Schrock, 2017, p. 73). Therefore any visualisations may unwittingly present a skewed subset of the original data.

- **Data integrity and quality:** If data is not gathered via formal research methods, it may be missing contextual metadata, incomplete or unverified. Decisions made on the basis of big data are harder to audit and could lead to incorrect assumptions (Clarke, 2016).

> *"Many inferences from big data are currently being accorded greater credibility than they actually warrant"* (Clarke, 2016, p. 86)
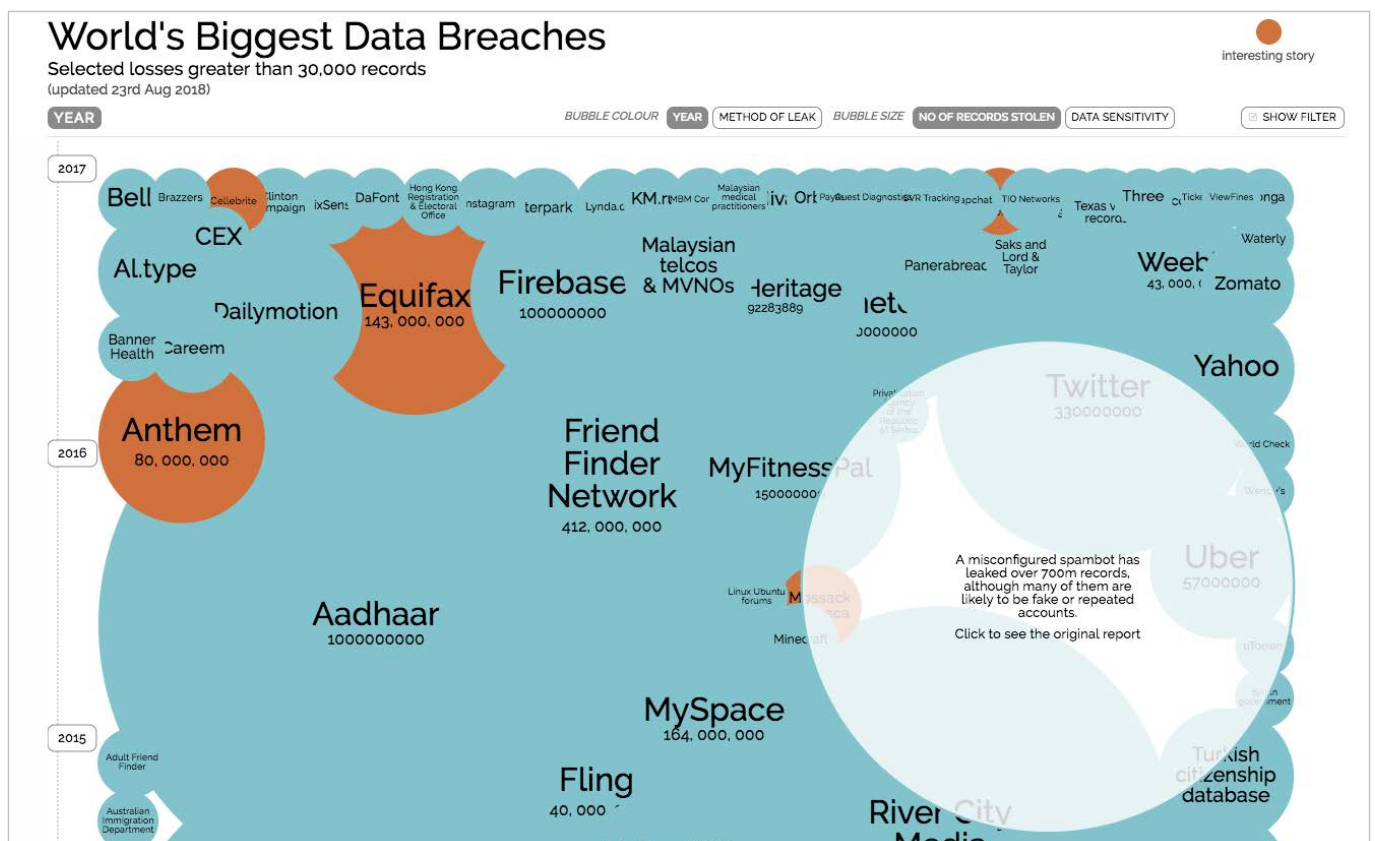

Figure 9 Interactive visualisation of data breaches; allowing user exploration and access to source data (McCandless, 2017)

Challenges and opportunities for designers

Big data provides big challenges, but also big opportunities for designers and creators of data visualisations.

Key trends and opportunities in data visualisation include:

- **Interactivity:** users filter or manipulate their view of the information according to their role or the insights they are looking for.
- **Real-time updating:** live data processed on-the-fly means users have a dynamic, current view rather than a static snapshot.
- **Augmented or virtual reality:** mobile devices that overlay virtual images onto real environments allow users to access visual data directly in context using augmented reality (Byun et al, 2016). With virtual reality, users could potentially immerse themselves in a visualisation and explore or manipulate it.
- **Sustainability:** data visualisations and infographics can be a powerful way of communicating messages relating to social or environmental data, helping people gain empathy and understand important issues.
- **Multi-disciplinary collaboration:** IT specialists, business managers and research scientists configure data visualisation tools and dashboards. Results could be improved by collaboration with graphic designers and user experience (UX) designers together with input from statisticians, psychologists, journalists and business analysts (Hepworth, 2017).

- **Innovation:** dynamic, large scale data sets are hard for humans to visualise, and systems and techniques for handling big data visualisations are in their infancy. Latest techniques include hierarchical, multi-layered approaches to data exploration, in-situ processing that adapts to user responses, and predictive displays that match visualisation style to user needs/aims (Bikakis, 2018). There are real opportunities for visual communication designers to shape these developments.

> *"…advances in human-computer interaction have created completely new paradigms for exploring graphical information in a dynamic way, with flexible user control."*

(Friendly & Denis, 2001)

Summary

Big data has disrupted visual communication practices and introduced problems such as information overload, difficulties of dealing with complex data sets, and associated risks of data quality and security.

However, with those problems come big opportunities for visual communication designers to work with researchers and industry to help make sense of this massive amount of data, creating highly usable and compelling data visualisations that tell stories, build knowledge and shape futures.
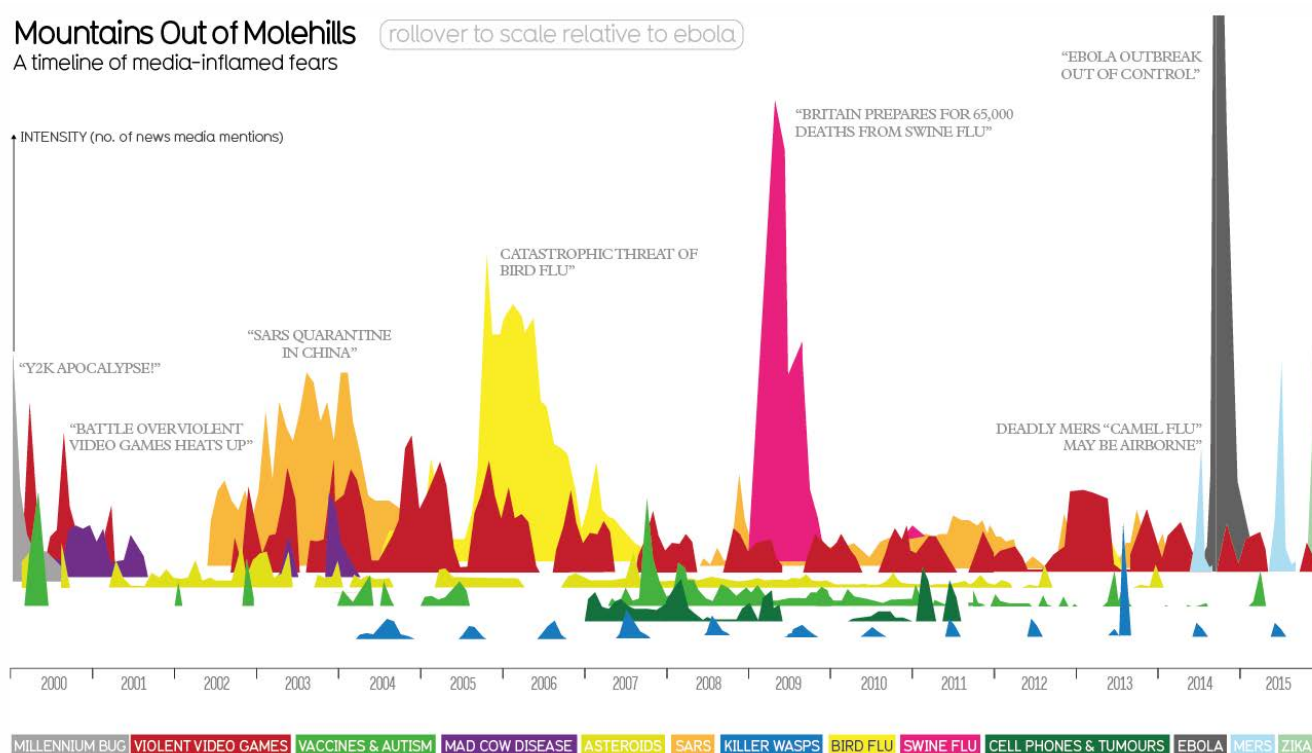

Figure 10 Interactive timeline visualisation of media-inflamed fears based on data from Google news (McCandless, 2018).

References

Bernard, J., Sessler, D., Kohlhammer, J. & Ruddle, R.A. 2018, 'Using Dashboard Networks to Visualize Multiple Patient Histories: A Design Study on Post-operative Prostate Cancer', *IEEE Transactions on Visualization and Computer Graphics*, pp. 1-1.

Bikakis, N. 2018, 'Big Data Visualization Tools', *Encyclopedia of Big Data Technologies*, Springer, 2018

Byun, H., Park, J.H. & Jeong, Y. 2016, 'Optional frame selection algorithm for adaptive symmetric service of augmented reality big data on smart devices', *Symmetry*, vol. 8, no. 5, pp. 37.

Cahyadi, A. & Prananto, A. 2015, 'Reflecting design thinking: a case study of the process of designing dashboards', *Journal of Systems and Information Technology*, vol. 17, no. 3, pp. 286-306.

Chai, S. & Shih, W. 2017, 'Why big data isn't enough', *MIT Sloan Management Review*, vol. 58, no. 2, pp. 57-61.

Clarke, R. 2016, 'Big data, big risks', *Information Systems Journal*, vol. 26, no. 1, pp. 77-90.

Cooper, P. 2016;2017;, 'Data, information, knowledge and wisdom', *Anaesthesia and Intensive Care Medicine*, vol. 18, no. 1, pp. 55-56.

Eyewire 2014, *Neurons mapped by gamers*, image, viewed on 3 September 2018, <https://plus.google.com/photos/104083442409040201683/album/5862119614724765073/5971457791570736386>

Friendly, M., & Denis, D. J. 2001, *Milestones in the history of thematic cartography, statistical graphics, and data visualization: Timeline*, viewed 31 August 2018 <http://www.datavis.ca/milestones/>

Friendly, M., & Denis, D. J. 2001, *Milestones in the history of thematic cartography, statistical graphics, and data visualization: Introduction*, viewed 31 August 2018, <http://www.datavis.ca/milestones/index.php?page=introduction>

Gapminder 2018, *Life expectancy vs Income bubble chart*, viewed on 4 September 2018, https://www.gapminder.org/tools/#$chart-type=bubbles

GSMA Intelligence 2018. *Global data*, viewed 3 September 2018, <https://www.gsmaintelligence.com/>

Hepworth, K. 2017, "Big data visualization: promises & pitfalls", *Communication Design Quarterly Review*, vol. 4, no. 4, pp. 7-19.

IBM. (n.d.) *Big Data Solutions*, viewed 3 September 2018, <https://www.ibm.com/it-infrastructure/solutions/big-data>

Longlivetheux 2015. *DIKW Pyramid*, image. Wikimedia Commons, viewed 31 August 2018 <https://commons.wikimedia.org/w/index.php?curid=37705247> CC BY-SA 4.0

McCandless, D. 2016, *World's Biggest Data Breaches & Hacks*, Information is Beautiful, viewed 4 September 2018, <http://informationisbeautiful.net/visualizations/worlds-biggest-data-breaches-hacks/>

McCandless, D. 2016, *Mountains Out Of Molehills*, Information is Beautiful, viewed 4 September 2018, <https://informationis-beautiful.net/visualizations/mountains-out-of-molehills/>

Mediratta, A. & Books24x7, I. 2015, *Big Data: Terms, Definitions and Applications*, EMC, viewed 30 August 2018, <http://common.books24x7.com.libraryproxy.griffith.edu.au/toc.aspx?-bookid=97487>

National Geographic 2015, *People Powered Data Visualization*, viewed 31 August 2018, < https://news.nationalgeographic.com/2015/08/150825-data-points-citizen-science/>

Peltier-Davis, C.A. 2015, *Cybrarian's Web 2*, Information Today, Inc. pp xxi-xxii.

Penkler, D. 2018, "Big data: technology and business challenges of big data in the digital economy", *Ubiquity*, vol. 2018, no. January, pp. 1-9.

Schrock, A. 2017, "What Communication Can Contribute to Data Studies: Three Lenses on Communication and Data", *INTERNATIONAL JOURNAL OF COMMUNICATION*, vol. 11, pp. 701-709.

Snow, J . 1854 *Cholera map*, image, retrieved from <http://matrix.msu.edu/~johnsnow/images/online_companion/chapter_images/fig12-5.jpg >

Strava 2018, *Fly-by Paris-Roubaix*, viewed 3 Sep 2018, https://labs.strava.com/flyby/viewer/#130432764?c=u0ced3gj&z=A&t=-1JIewG&a=_D7GB3L8xQd9u8MIgPjkDCtYvwdfccQH&x-=mp-v

Socialab 2018, *LinkedIn Network Visualization and Analysis*, viewed 2 September 2018, <http://socilab.com/#home>

Tableau, n.d., *Salesforce Starters Dashboard*, viewed 31 August 2018, <https://cdnl.tblsft.com/sites/default/files/salesforce_template_screenshot_-_v2.jpg>

Twain, M. 1906, *Chapters from my autobiography*. *Project Gutenberg*. London: Harper, viewed 3 September 2018 <http://www.gutenberg.org/files/19987/19987-h/19987-h.htm>