Data Visualization

CATHERINE D'IGNAZIO
Emerson College, USA

Data visualizations are an increasingly powerful way to make knowledge claims about the world. While the visual display of quantitative information has a long history, data visualization has become a more mainstream form of communication than ever before. Data visualization can be defined as the graphic presentation of information that has been systematically collected. Forms of visualizations range from the familiar pie charts, timelines, and maps to more specialized forms such as treemaps, cartograms, streamgraphs, and chord diagrams. While much of contemporary data visualization is two-dimensional and distributed online, data visualization may also occur in print publications, motion graphics, or take on 3-D forms such as sculptures, public art, murals, and performances.

Whereas visualization was previously a niche form of communication for scientists, statisticians, and other professions, the past 20 years have witnessed a proliferation of data visualization more broadly in professional cultures as well as in popular culture. Data visualizations are routinely created by corporations, news organizations, nonprofit and nongovernmental organizations (NGOs), policymakers and government, independent designers and bloggers, and everyday people. Visualization has accompanied an increased interest in "data-driven decision-making" and it is often asserted that visualizations aid in gaining insight from large and complex data sets. Visualization publishers and designers often state that they are trying to make data more accessible and actionable to members of the general public and see data visualization serving a democratizing function, though scholars have questioned whether it truly does or not (Kennedy, Hill, Aiello, & Allen, 2016).

History of data visualization

While this entry focuses on visualization in the contemporary context, it is important to know and to teach learners that the form is not new. Scholars have shown that map projections date as early as Ptolemy (c. 85–c. 165) and there are examples of charts as early as the 10th century (Tufte, 1983). Data visualization witnessed a golden age in Europe from 1850 to 1900, with the growth of statistical theory, wide-scale systematic record-keeping by nation-states, and development of more sophisticated visual methods (Friendly, 2008). Until recently, however, data visualization was not seen as a single form or field of inquiry, with disciplines such as medicine, geography, statistics, and astronomy developing visual display methods independently of each other. There is

The International Encyclopedia of Media Literacy. Renee Hobbs and Paul Mihailidis (Editors-in-Chief), Gianna Cappello, Maria Ranieri, and Benjamin Thevenin (Associate Editors).

DOI: 10.1002/9781118978238.ieml0095

much research to be done in this realm, for example, tracing the use of data visualization in a nonwestern context. Additionally, historians have challenged prior accounts of the evolution of data visualization, in particular in the way they write women's work with data out of the historical record (Klein, 2014).

Context of data visualization

There are three interrelated phenomena that have led to the popularization of data visualization. The first is the rise of "Big Data" defined as an interplay of technology, analysis, and mythology (boyd & Crawford, 2012). Because of advances in technological capacity and analytical methods, states and corporations are able to collect, store, and process more data than ever before. Data visualization is commonly seen as a path toward gaining insight from large and complicated data sets that would be impossible to understand in tabular form. It is worth highlighting the third part of boyd and Crawford's definition: "the widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy" (2012, p. 663). Data visualization is often the visual enactment of this "aura of truth." Learning to be a critical reader and designer entails learning productive ways to challenge that aura.

The second phenomenon is the rise of the open data movement. Open data is the idea that certain data should be freely available to everyone to use and republish as they wish, without restrictions from copyright, patents, or other mechanisms of control. The movement started with governments in the United States and the United Kingdom but has now gone international with more than 50 nations across six continents committed to opening some of their data as well as the participation by some NGOs and corporations. The ideals of the open data movement include government transparency and accountability, increased trust, democratic access, faster scientific progress, economic development, and citizen empowerment. The practical outcome of the open data movement is that there are a proliferating number of ways to download data about everything from congressional voting records to agricultural statistics. For new learners these repositories can be both a tremendous realm of possibility and appear deceptively complete. Finding the data one *needs* remains a key challenge in data-driven inquiry projects.

The third and final phenomenon is "datafication," which can be defined as the transformation of social action into quantified data, thus allowing for real-time tracking and predictive analysis. For example, a personal health tracker uses an accelerometer to make a rough approximation of an individual's steps during the day. This constitutes "datafying" steps: counting and recording them through an analog-to-digital conversion process. Datafication is happening at many scales and becoming a norm in many industries, particularly as new economic models for successfully surveilling populations and monetizing data become apparent (van Dijck, 2014). For example, at the city scale, cities are experimenting with sensors in urban spaces to datafy traffic patterns, energy consumption, pedestrian activity, waste streams, and air quality, among many other things. Learning about data visualization entails understanding how and why institutions go about datafying various aspects of the world.

Examples of data visualization: personal, professional, public

Data visualization happens in a variety of contexts ranging from personal to professional to public. In the personal context, there have been a proliferation of self-tracking apps and personal health management technologies. Individuals can measure and track their steps, heart rates, GPS (Global Positioning System) location, posture, golf swing, income and spending, home climate control systems, and many more aspects of their everyday lives.

This data is presented back to the individual in the form of data visualizations that show current statistics, progression toward a goal, or compare data across time. For example, personal health trackers present users with a personal health dashboard, a series of small visualizations on one screen that are intended to be checked daily. On this dashboard, we see visualizations of heart rate, sleep patterns, weight, and weekly heart rate. From a critical perspective, it is important to remember that these visualizations have been designed to show some things and obscure others. Personal health trackers, particularly those used at scale, have access to aggregated health information of many people. Missing from the personal dashboard are the ways in which personal data is then stored, aggregated with millions of other records, sold to third-party brokers, and otherwise used by the corporate host to drive sales and revenue. This is not to imply malicious or exploitative intent. It is simply to state that visualizations, like many other media, obscure the material conditions of their making in favor of a localized experience.

Data visualizations are prevalent in a variety of professional contexts, too many to fully describe in detail in this entry. While financial and scientific professions have used data visualizations for a long time, they are on the rise in realms such as municipal government, food safety, and nonprofit management. Data visualizations are used for internal monitoring and communication purposes as well as for external communication between an institution and its constituents or customers. For example, Figure 1 shows a network visualization of all of the data available via the New York City Open Data Portal. Each node is a single data set and the graphic icons allude to the data's general topic area, so a leaf icon represents environmental data. Further research is needed into the purpose, context, and reception of data visualization across extremely diverse professional contexts.

Finally, data visualizations have become extremely popular in more public forms of communication such as marketing, advertising, journalism, and social media. The "computer-assisted reporting" that took shape in the 1970s and 1980s now more often goes by the term "data journalism." This work takes numerous forms for the reader or end user. The most ubiquitous are simple, static charts that support news stories in print, on the web, or that are broadcast. More complex forms include interactive experiences that allow users to explore a question through visualizing data, such as "512 Paths

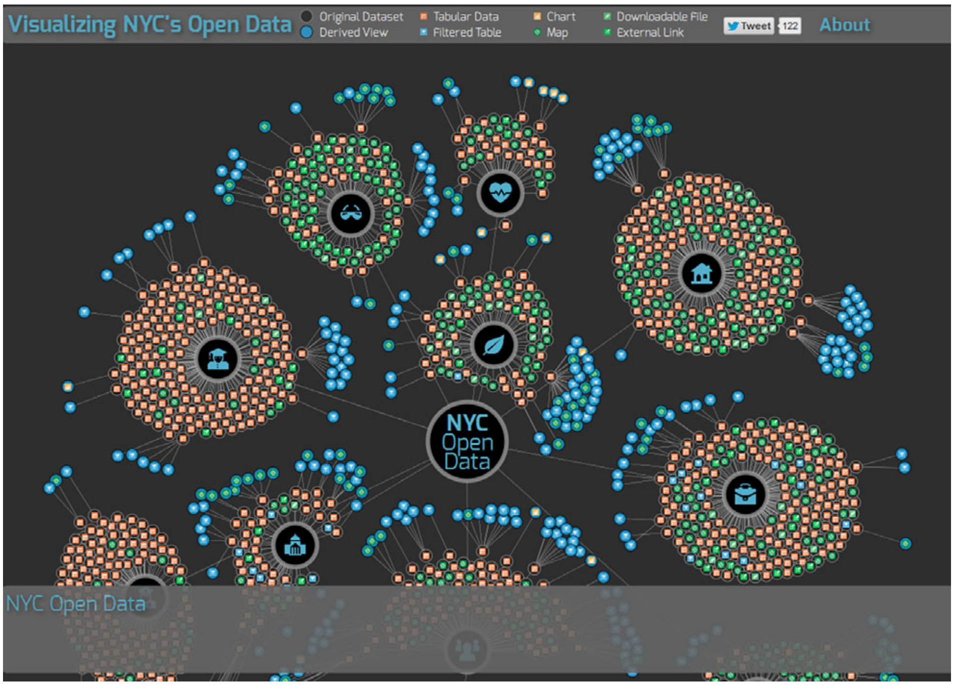


Figure 1 All of the data sets available on the NYC Open Data Portal. Source: nycopendata.socrata.com, October 2016.

to the White House" (Bostock & Carter, 2012) that explores alternate presidential outcomes depending on how US states vote in a national election. And news organizations are increasingly experimenting with hybrid forms that combine narrative storytelling with visualizations. The *Guardian* story in Figure 2 is a good example of this hybrid approach which uses short narrative statements punctuated by a series of visualizations to describe how Australian women won the 4×100 freestyle relay. The piece includes an animated graphic that demonstrates race order, a timeline showing when the record had previously been broken, and three photos annotated with data points.

Knowledge domains

For the purposes of media literacy learners, there are five distinct knowledge domains that comprise learning about data visualization. They draw from creative and technical fields, so learners and instructors can often feel overwhelmed by the amount of disparate knowledge to assimilate. Learners need to understand basic principles of *graphic design*, including typography and visual relations such as color theory, contrast, balance, alignment, and repetition. *Interaction design and user experience* play a role when data visualizations are interactive. There are now established conventions for displaying summary views, zooms, and details (Kennedy et al., 2016; Shneiderman, 1996). Reading, interpreting, and creating data visualizations requires at least passing knowledge of *statistics* in order to characterize patterns from aggregated observations about

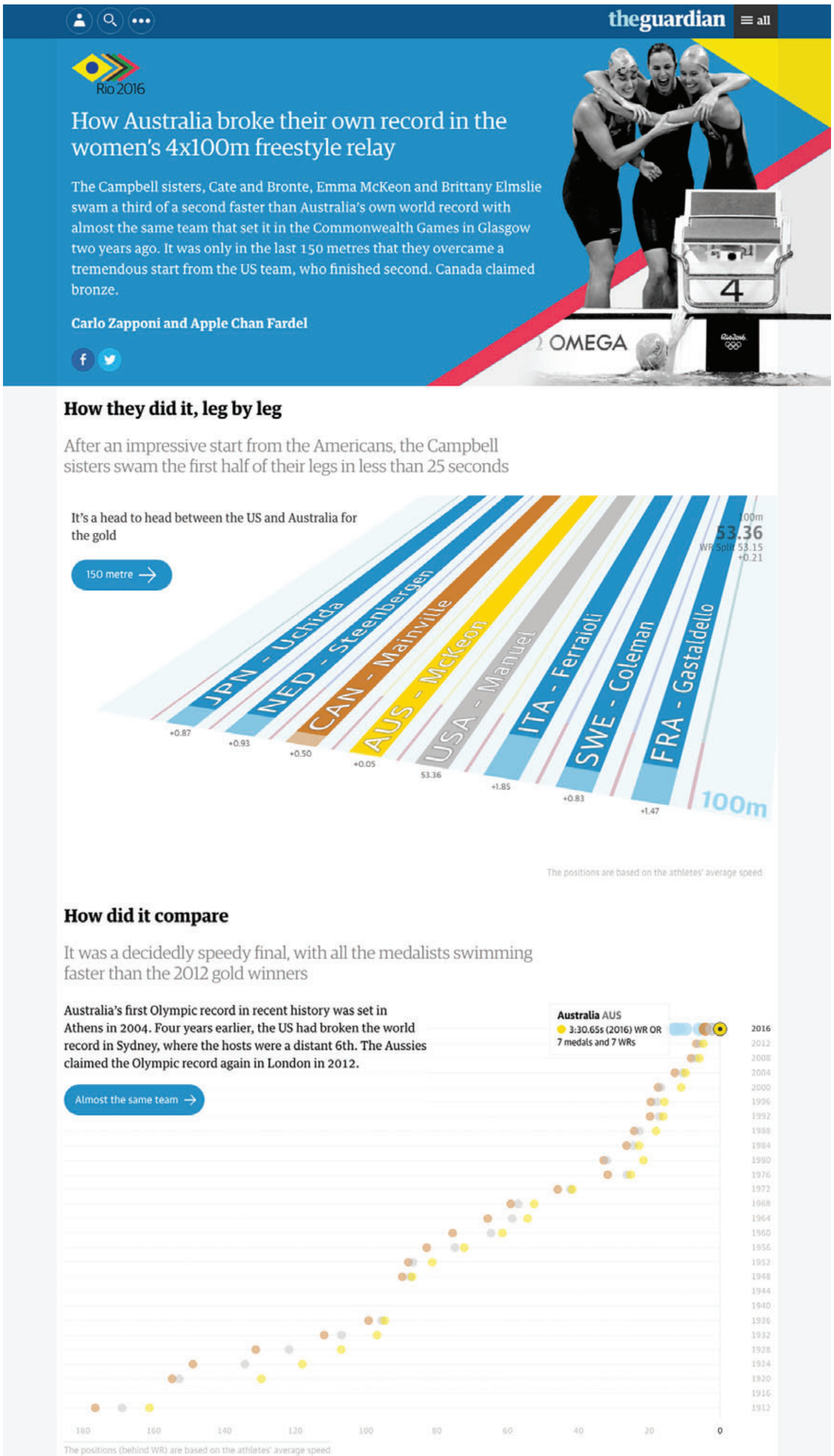
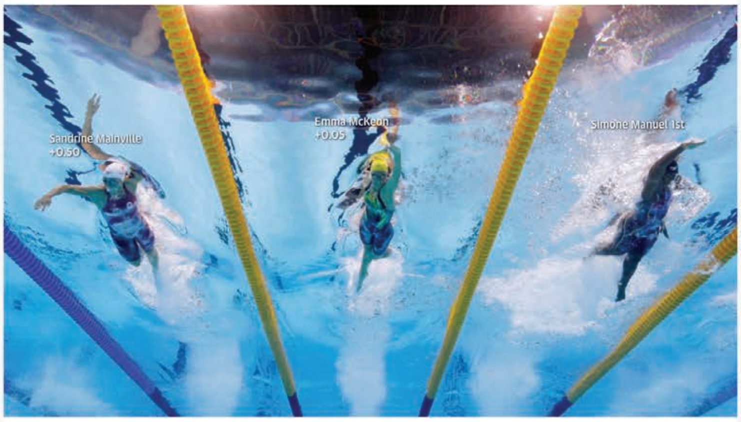


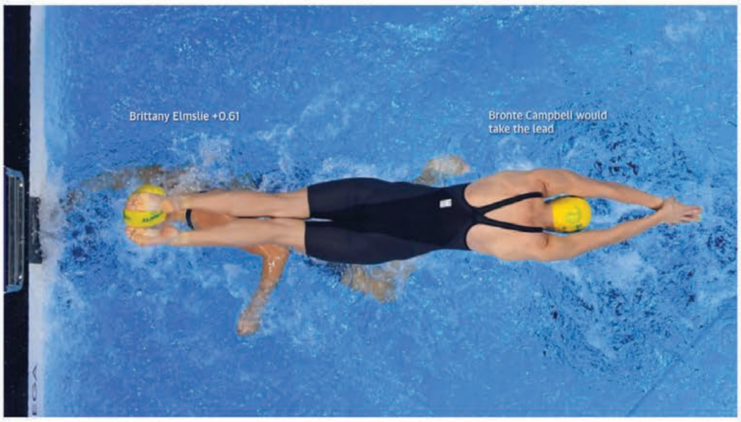
Figure 2 How Australia broke their own record in the women's 4×100m freestyle relay. Source: www.theguardian.com (Zapponi & Chan Fardel, 2016, August 7). (*continued overleaf*)

Coming from behind

Off the platform, only the Dutch were slower that the Aussies. Nonetheless Emma McKeon finished her leg only inches behind American Simone Manuel.



After Brittany Elmslie lost over half a second to Abbey Weitzeil, it was up to Bronte Campbell to swim faster than American Dana Vollmer. And she did, more than a second faster.



Big sister Cate was already half a body ahead of the world record when she jumped into the pool, she then swam her first 50 metres in just 24.15 seconds to lock it.

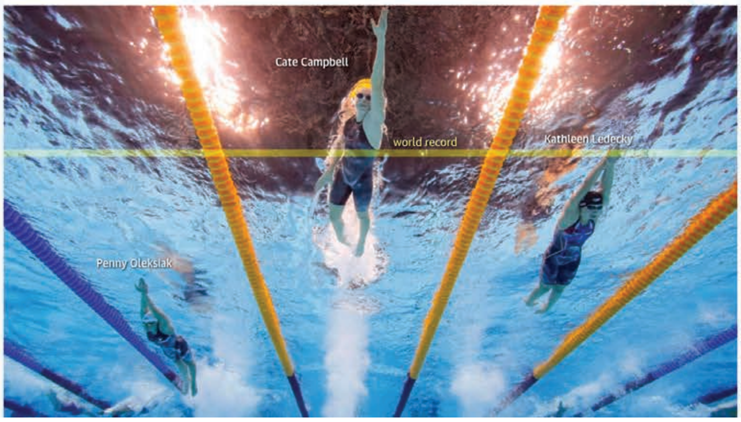


Figure 2 (*Continued*)

the world. Learners find they can get rather far with descriptive statistics (e.g., mean, median, mode, max., and min.) but will be able to do more if they understand inferential statistics and advanced methods of data science such as building predictive models with machine learning. Basic understanding of *visual cognition* augments and integrates the prior three domains, as it deals with how the brain processes information, detects patterns, and uses visual search queries as part of its everyday architecture. Finally, data visualizations need to be assembled by some means. This is typically done through *software and web development*, that is, through writing code. Coding is not an absolute requirement because out-of-the-box software packages for data visualization are becoming more advanced; however, it does afford the ability to process data more effectively, tailor data visualizations to a particular project, and customize interactions.

Data visualization as process and product of inquiry

While data visualization is most commonly seen and considered as a polished communicative product, it is important to note that it is part of a broader process of inquiry, analysis, and communication. Learning data visualization is learning an inquiry process that is methodical and data-driven and setting expectations for what should be happening at each stage of the process. For this reason, many data visualization modules happen in research methods courses where students are learning how to formulate research questions and which methods are useful for which kinds of knowledge claims. Research methods courses introduce many of the critical skills that students need for working with data, including standardized methods for collecting data, designing surveys, and calculating uncertainty. Data visualization may be taught outside of the context of a formal research methods course, but educators may need to work harder to cultivate critical perspectives on data (as described below).

When data visualization is the goal of a data-driven inquiry project, it is important to note that visualizing data happens throughout the process, not just at the end as the product. While it is most common as a reader to encounter data visualizations as finished products of the data analysis process, data visualization creators will use visualization at every stage in order to understand questions like: Does my data answer my question? Where is the missing data? Do I need to find and combine another data set to answer my question? Do I need to clean and standardize my data? What is the right method of analysis? (See Figure 3.)

Critical perspectives and metaskills

Because data visualization is a relatively new phenomenon as a form of popular communication, there is not a great deal of scholarship on critical analysis of these artifacts as media. Scholarship about visualization has tended to come from more technical fields such as *information visualization*, which concerns itself with the design of interactive systems that deal with large amounts of data, as well as *computer science* where the layout algorithms and methods for processing and displaying large amounts of data

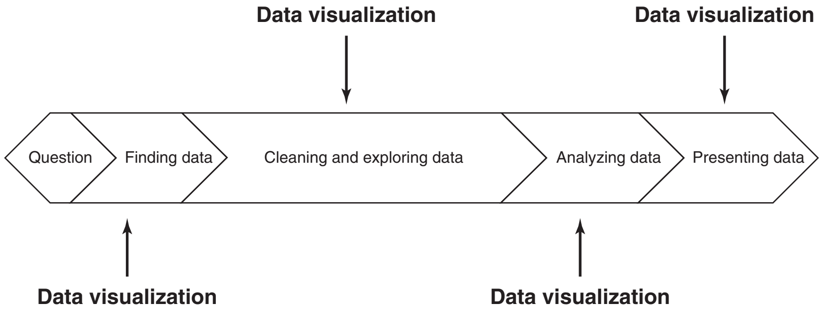


Figure 3 Visualization happens throughout the data exploration and analysis process, not just at the end. Source: The author.

are invented. For this reason, data visualization has tended to have a more positivist inflection. But this is changing. In recent years, with the increasing attention to the intersection between data and society, scholars have borrowed from fields like *critical cartography*, *GIS*, *human–computer interaction* and the *digital humanities* to propose ways to critically read and create data visualizations. These proposals go by various names such as critical information visualization (Dörk, Feng, Collins, & Carpendale, 2013), public participatory GIS (Elwood, 2006), and feminist data visualization (D'Ignazio & Klein, 2016). In general, critical perspectives take issue with the way that data visualizations appear to present a neutral, objective "view from nowhere," note the inequalities in the data ecosystems from which these visualizations emerge, and highlight the power imbalances between who collects, stores, and analyzes data and who is collected, stored, and analyzed. Some designers have advocated incorporating other senses such as touch and even taste into experiential displays of data so that people might "feel" data, "visceralize" data, or "physicalize" data. Educators have asserted that these more creative methods can be used to inspire learners, help them feel less intimidated, and provide examples for designers to create more inclusive visualizations for particular audiences (Bhargava, 2014).

Learning about data visualization, therefore, represents an opportunity to become a critical reader and a critical data designer. These two roles go hand in hand. Informed reading of data visualization necessitates knowledge gained through hands-on experience of the data collection, analysis, and design process. While the section on *knowledge domains* described the fields that visualization draws from, those do not encompass the metaskills needed to operationalize this knowledge. These come from cultivating critical data literacy and include learning objectives such as (i) understanding what questions can be answered with data; (ii) understanding which aspects of the world data represents; (iii) understanding of institutional data collection and publication practices, including the role of data collection in surveillance, advertising, and corporate profit; and (iv) understanding that no data is "raw", that is, no data is an unfiltered representation of reality. Fostering critical data literacy should be seen as the core work of any process of teaching data visualization.

SEE ALSO: Critical Information Literacy; Data Literacy

References

Bhargava, R. (2014). Speaking data. Retrieved from https://datatherapy.org/2014/07/09/speaking-data/
Bostock, M., & Carter, S. (2012). 512 Paths to the White House. *New York Times*. Retrieved from http://www.nytimes.com/interactive/2012/11/02/us/politics/paths-to-the-white-house.html?_r=0
boyd, d., & Crawford, K. (2012). Critical questions for Big Data. *Communication & Society, 15*(5), 662–679. doi: 10.1080/1369118X.2012.678878
D'Ignazio, C., & Klein, L.F. (2016). Feminist data visualization. Retrieved from http://www.kanarinka.com/wp-content/uploads/2015/07/IEEE_Feminist_Data_Visualization.pdf
Dörk, M., Feng, P., Collins, C., & Carpendale, S. (2013). Critical InfoVis: Exploring the politics of visualization. *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (pp. 2189–2198). doi: 10.1145/2468356.2468739
Elwood, S. (2006). Critical issues in participatory GIS: Deconstructions, reconstructions, and new research directions. *Transactions in GIS, 10*(5), 693–708. doi: 10.1111/j.1467-9671.2006.01023.x
Friendly, M. (2008). A brief history of data visualization. In C.-H. Chen, W. Härdle, & A. Unwin (Eds.), *Handbook of data visualization* (pp. 15–56). Berlin, Germany: Springer.
Kennedy, H., Hill, R.L., Aiello, G., & Allen, W. (2016). The work that visualisation conventions do. *Information, Communication & Society, 19*(6), 715–735. doi: 10.1080/1369118X.2016.1153126
Klein, L. (2014). Visualization as argument. Retrieved from http://lklein.com/2014/12/visualization-as-argument/
Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. *Proceedings 1996 IEEE Symposium on Visual Languages* (pp. 336–343). doi: 10.1109/VL.1996.545307
Tufte, E. (1983). *The visual display of quantitative information*. Cheshire, CT: Graphics Press.
van Dijck, J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance and Society, 12*(2), 197–208.
Zapponi, C., & Chan Fardel, A. (2016, August 7). How Australia broke their own record in the women's 4×100m freestyle relay. *The Guardian*. Retrieved from https://www.theguardian.com/sport/ng-interactive/2016/aug/07/how-australia-broke-their-own-record-in-the-womens-4x100-freestyle-relay

Further reading

Bhargava, R. (2016). Data therapy – tips for empowering people with data. Retrieved from https://datatherapy.org/
Kirk, A. (2012). *Data visualization: A successful design process: A structured design approach to equip you with the knowledge of how to successfully accomplish any data visualization challenge efficiently and effectively*. Birmingham, England: Packt.
Meirelles, I. (2013). *Design for information: An introduction to the histories, theories, and best practices behind effective information visualizations*. Beverly, MA: Rockport.
Ware, C. (2008). *Visual thinking for design*. Burlington, MA: Morgan Kaufmann.

Catherine D'Ignazio is an assistant professor of civic media and data visualization at Emerson College, a principal investigator at the Engagement Lab, and a research affiliate at the MIT Center for Civic Media. Her work focuses on data literacy, feminist

technology, and civic art. D'Ignazio has authored articles on critical data literacy as well as co-developed a platform called DataBasic.io. Her art and design projects have won awards from the Tanne Foundation, Turbulence.org, and the LEF Foundation. Her work has been exhibited at the Eyebeam Center for Art & Technology, Museo d'Antiochia of Medellín, and the Venice Biennial.