Incorporating Data Visualisation Into Teaching and Learning

Meng Li
The University of Queensland
m.li9@uqconnect.edu.au

The profound advancements in technology have rendered novel forms of data and data visualisation increasingly accessible to individuals within society, thereby influencing daily decision-making processes. To address this change, this study sets out to review recent research on data-driven inquiries at the K–12 level from two perspectives: innovative data visualisation and non-traditional data sources. Our findings indicate that transnumeration of multiple data representations, along with data moves throughout the process of data visualisation, can potentially enhance the development of visual reasoning and data modelling skills.

Over the last two decades, there has been an increasing and ongoing demand for *data-literate* citizens who are capable of using data as sound evidence to make decisions on a wide range of topics (Engel, 2017). Data literacy, which can be broadly described as “learning to read and write the world with data” (National Academies of Sciences, Engineering and Medicine, 2023, p. 5), is not only a requirement for entering into the workforce but also a requisite skill for navigating everyday life. One approach to foster students’ data literacy could be implementing a *data science* course that is designed for students to learn about the world through data (Gould, 2021). Defined as “the science of learning from data”, data science is an interdisciplinary subject which focuses on using evidence-based approaches as well as technology to make sense of data (Donoho, 2017, p. 763). From this perspective, Donoho (2017) classified all the activities associated with data science into six divisions with *data visualisation* included in half of these divisions. In broad terms, *data visualisation* is both a *process* and a *product* that are embedded in data-based inquires rather than the equivalence of graphics (Arcavi, 2003). Moreover, the rapid development of technology has expanded the data that are available for classroom instructions (Gould & Çetinkaya-Rundel, 2013), and this change has fundamentally shaped data visualisations. This inclusion of visualisations (not just graphs) has started to appear in curricula. For example, in the latest Australian Curriculum Mathematics (AC:M), data visualisation is first introduced in Year 4 statistics as follows, “represent data using many-to-one pictographs, column graphs and other displays or visualisations” (AC9M4ST01).

As can be interpreted from this content descriptor, data visualisation is considered as supplementary graphical representations for numerical information in addition to conventional graphs such as pictographs and column graphs. Although the new version of the Australian curriculum vaguely mentions the use of “other display or visualisations”, it fails to address data with “non-standard sources such as those with multiple characteristics or ambiguity (e.g., photographs, text, mobile apps)” (Makar et al., 2023, p. 976). The term *data visualisation* included in Grade 4 statistics and data display is used across Year 7 to Year 10 curriculum. This example indicates that more explicit guidelines for data visualisation need to be provided to classroom teachers and so they can incorporate this concept into everyday teaching practice. Therefore, the aim of this paper is to discuss opportunities for embedding data visualisation in the mathematics curriculum within the context of K–12 data science education. The research question guiding this literature review is:

- How can data visualisations be incorporated into classroom teaching and learning?

(2024). In J. Višňovská, E. Ross, & S. Getenet (Eds.), *Surfing the waves of mathematics education. Proceedings of the 46th annual conference of the Mathematics Education Research Group of Australasia* (pp. 327–334). Gold Coast: MERGA.

What Constitutes Data

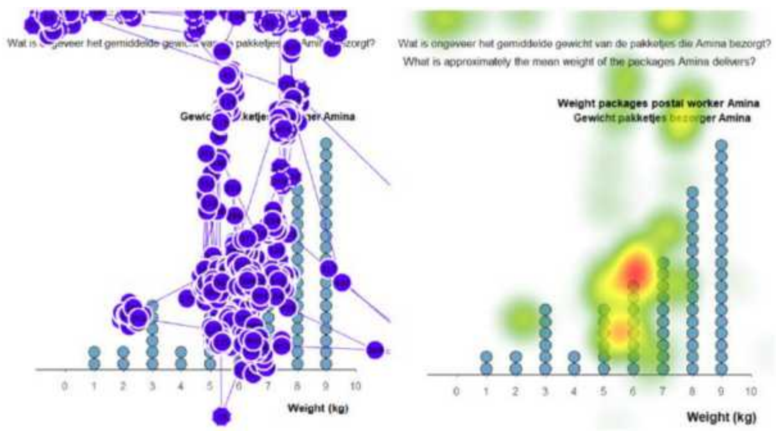
Back to the 17th century, data were exclusively produced, managed, and owned by scientists (Leonelli, 2019). During the last few decades, the development of technology has made large data sets accessible to the public. Consequently, people are immersed in statistical data within their daily experiences and are thus compelled to cultivate a discerning aptitude for data consumption (Watson, 2002). Moreover, these radical transformations have also shaped the conceptualisation of data. In other words, the categories of data considered by non-statisticians have been expanded from only *numerical* or *categorical* to a broader group which includes non-traditional data such as images, sounds videos, texts, and geospatial data. As such, data can be defined as any "information collected or generated from the world from which inferences about various phenomena can be made" (Wise, 2020, p. 165). This definition exhibits a breadth that encompasses not only the diverse forms of data but also the purpose for its manipulation and analysis. Stated differently, data serve as entities utilised by individuals to comprehend, explain, communicate, and even predict diverse phenomena within the physical world as well as our society. This calls for the need to incorporate modern data visualisations which focus on large, complex, and newer forms of data into teaching and learning practice (Nolan & Perrett, 2016).

Data Visualisation and Its Characteristics

In the realm of mathematical visualisation, data visualisation emerges as a distinct subset, characterised by its capacity to serve as both a procedural methodology (*process*) and an outcome (*product*). Prior to the work of Tukey (1977), who promoted the use of data visualisation for exploratory data analysis (EDA), the role of data visualisation was largely unknown. According to Tukey (1990), the use of data visualisation throughout EDA allows us to search, interpret, and analyse the data to explore phenomena that have occurred or that may occur. Likewise, Unwin (2020) pointed out that data visualisations for EDA were created to reveal new information. termed as "seeing the unseen" by Arcavi, (2003, p. 216). In light of this assertion, it is imperative to regard data visualisation as a facilitative instrument possessing inherent qualities conducive to discovery, rather than merely functioning as a graphical representation. The evolution of technology has substantially augmented the scale, accessibility, and modalities of data. This significant development has shifted data visualisation from conventional graphs for analysing quantitative information to instruments for reasoning about both traditional and non-traditional data.

Figure 1

The Gaze Plot (Left) and Heat Map (Right) (Boels, 2023)

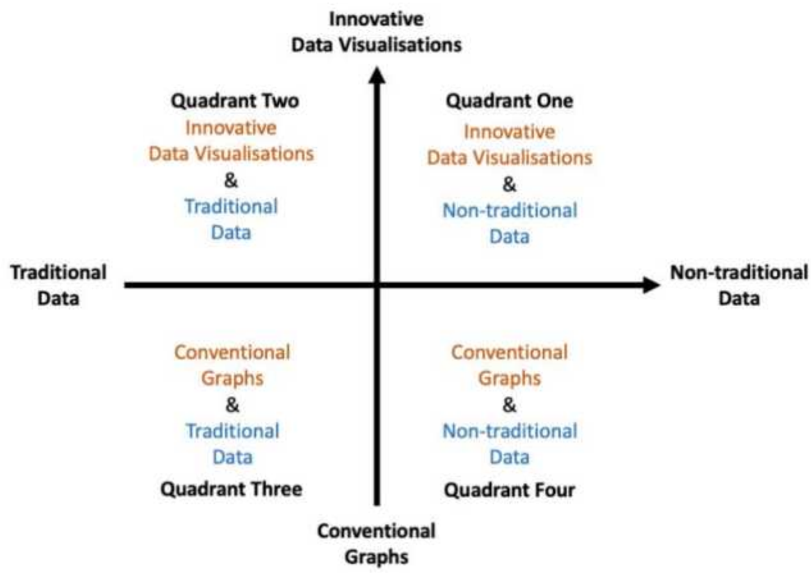


Although newer forms of data often contain valuable insights that might not be seen in traditional data, unstructured forms also make non-traditional data more challenging to visualise. That is, relying solely on conventional graphs may prove inadequate for comprehensively making sense of data, necessitating the utilization of innovative data visualisation techniques. This is evident in the study undertaken by Boels (2023), in which innovative data visualisations such as gaze plots and heat maps were used to investigate gaze data (eye-tracking data). The employment of these data visualisations facilitated researchers in understanding the methodologies that students applied into the interpretation of graphs, revealing strategies previously unnoticed by both educators and students. Additionally, novel data visualisations can be utilised to depict traditional data. A recent commentary by Lim et al. (2023) discussed the possibility of extending conventional graphs to innovative data visualisation. For example, adding a sensory dimension by attaching an audio clip to a time-series graph which depicted the sound level of New York City (numerical data) before and after the COVID lockdown. It is believed that the integration of different elements such as audio, video, interactive features as well as art is an efficient way of enhancing data storytelling. As posted by Clark and Paivio (1991), directing students to integrate events or objects into narratives will foster the establishment of mental imageries, thereby augmenting comprehension in learning.

Within the context of education, incorporating innovative forms of data and novel data visualisations may broaden the opportunities for teaching and learning of data science. The Cartesian plane shown below (see Figure 2) categorises different types of data and data visualisation into four quadrants. Too many classroom instructions and textbooks still focus on the combination of conventional graphs and traditional data (see Quadrant Three in Figure 2). As a result, further research is needed to investigate the remaining three quadrants, which encompass either innovative data, novel data visualizations, or both.

Figure 2

The Different Relationships Between Data and Data Visualisation



Review of the Literature on Innovative Data Visualisation and/or Data

The section below reviews the literature related to innovative data visualisation and/or newer forms of data. It commences with an exploration of studies involving innovative data visualisations (see Quadrant One and Two in Figure 2), subsequently progressing to an analysis of studies centred on non-traditional data (Quadrant Four).

Innovative Data Visualisations (Quadrant One and Two)

Within the context of K–12 education, there is a prevalent consensus that data visualisation ideally belongs within the discipline of probability and statistics. However, what data scientists actually do in the real world differs from what is taught in schools. That is, data visualisation is used as a tool to explore, analyse, and communicate data collected from different aspects of our life. This calls the need to design school-based learning experiences that resemble how data visualisation is used in practice. Some researchers have advocated for the integration of data science across different subjects (Jiang et al., 2022). In recent years, there has been an increasing interest in how to incorporate innovative forms of data visualisation into classroom teaching and learning across different disciplines. An example of this is the study conducted by Matuk et al. (2024) in which 15 to 30 middle-school students reasoned about both qualitative and quantitative data via *data-art inquiries*. In the *Photoessay* unit, students collected air quality data from a public map-based database, after making a scatterplot to investigate the association between air quality and life expectancy. Finally, students transcended traditional data by employing photography as a means to explain the diminished air quality within their neighbourhood area (see Figure 3). The use of photo-essay promoted "learners' existing relationship with data" and this facilitated students in gaining insights into the data (Wilkerson & Polman, 2020, p. 5).

Figure 3

A Student's Photoessay (Matuk et al., 2024)

Students in the preceding study utilised diverse forms of data representations, such as scatter plots and photoessays, across the duration of the instructional module. The process of "forming and changing data representations" is described as *transnumeration* by Wild and Pfannkuch (1999, p. 227). To be more specific, transnumeration is a data-visualisation process through which data are re-expressed and re-classified for the purpose of seeking new insights. In other words, it is a process as well as a technique that allows us to investigate the same data set via different perspectives so new meanings may be revealed. The shift from scatterplot to photoessay allowed students to visualise the potential impact of construction on environment in their local area (Matuk et al., 2024). This discovery aligns with the research by Roth and MacGinn (1997), who posited that the transition from *experience-distant* visual representations (e.g., scatterplot) to *experience-near* visual representations (e.g., photoessays) may enhance students' contextualization of data.

Another example of transnumeration in the process of data visualisation is the study undertaken by Makar (2016) who investigated the emergence of children's informal statistical inference. In this study, students used experience-near data visualisations (e.g., pictures) to record data about their classmates' shoe sizes (see Figure 4). Then, they constructed experience-distant visual representations with stacked and ordered white cards to model the shoe distribution of the class. The use of multiple types of data visualisations created opportunities for younger students to visualise the variability as well as the distribution associated with a data set. The change from *experience-near* to *experience-distant* data representation allowed young children to visualise abstract statistical concepts such as variability and distribution of data sets via the process of data visualisation.

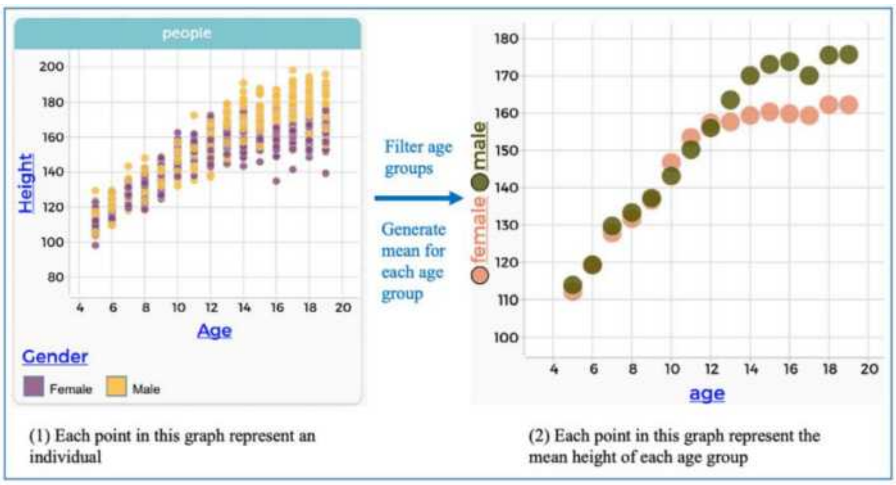
Figure 4

Transnumeration Between Different Types of Data Representations (Makar, 2016)

In addition to changing between different types of data representations, transnumeration can also occur within the same type of data visualisation via *data move*. For Erickson et al. (2019), a data move is “an action that alters a dataset’s contents, structures, or values” (p. 5). The study by Erickson and Chen (2021) offers a comprehensive description of how data moves (e.g., filtering, grouping, and summarizing) can be used to investigate large and unruly data via technology-driven data visualisation tool (e.g., CODAP). As shown in Figure 5, students *filtered* each age group and then plotted *mean height* against all ages. The transformation of the scatterplot unveils the association between age and height, a relationship previously obscured within the original data representation. It should be emphasised that the adaptation of innovative data visualisation, exemplified by the dynamic visual display on CODAP, enhances the efficacy of the data visualisation process. Moreover, transnumeration does not happen spontaneously which means the pivotal role of guided discovery by educators during the process of data visualisation should not be underscored.

Figure 5

(1) The Data Representation Before Filtering; (2) The Data Representation After Filtering (Erickson & Chen, 2021)



The aforementioned study provides an illustrative instance of leveraging an innovative data visualisation tool such as CODAP to elucidate large and unstructured traditional data sets. Also, the integration of CODAP with additional online visualisation tools has the potential to streamline the decision-making process relating to machine learning. An example is the study undertaken by Erickson and Engel (2023) who proposed a data science project for high school students to construct and apply *classification trees*. In the outlined project, students explored how *the tree* worked in a real-life context of breast cancer diagnoses. Concurrently, supplementary data representations, including dot plots and scatterplots, are employed to aid students in identifying the threshold distinguishing positive from negative diagnoses. This procedure further supports the idea of using transnumeration as a way throughout the exploration and analysis of data. Although using *trees* as a mean to visualise data is new to most students, it is believed that such experience will provide students with a solid and profound understanding of how advanced statistics work in data science and machine learning (Erickson & Engel, 2023). Further, educators play a critical role in this data-based inquiry since guidance is needed for students to construct, understand, and apply *trees*.

Conventional Graphs and Non-traditional Data (Quarant Four)

Recently, a number of studies have begun to explore how to use driven visualisation to investigate non-traditional data. These studies may be broadly categorised into two groups: those employing primary data as data sources (Fergusson & Pfannkuch, 2022; Podworny et al., 2022) and those utilising secondary data from data sources that are open to the public (Jiang et al., 2022; 2023; Rao et al., 2023).

The acquisition and processing of primary innovative data frequently requires technology-driven methodologies and instruments. For example, the study by Podworny et al. (2022) offers the possibility of using sensors to collect real-time and authentic data from the local community to investigate environmental factors. In another example, the researchers designed a a step-by-step protocol to teach the participants how to use *R* code to collect and analyse data from a movie-rating website (Fergusson & Pfannkuch, 2022). In both investigations, code-driven tools (e.g., *R* code and *Python*) played essential roles in the transformation of raw data into conventional graphs such as linear regression graphs or line graphs. Put differently, students would not be able to visualise the data undergoing collection without the integration of coding. On one hand, these studies provided insights on how to "merge computational and statistical thinking" via coding in data science education (Gould, 2021, s17). Alternatively, it is uncertain whether computer programming imposes an excessive cognitive burden on learners as well as classroom teachers. Namely, students might allocate a significant portion of their time to acquiring coding skills rather than engaging in exploration and analysis throughout a data-centric inquiry. Additionally, classroom teachers may encounter challenges due to potential inexperience designing and implementing data science projects reliant on code-driven tools.

Figure 6

Using Multiple Data Representations to Evaluate Machine Learning Model (Jiang Et Al., 2022)

In contrast to data-based inquiries employing primary innovative data, the necessity of coding as a tool is less pronounced in investigations utilizing secondary innovative data. Instead, the focus of these data science projects is often on wrangling, analysis and evaluation of unstructured data. A recent study by Jiang et al., (2022) examined how high school students built machine learning models for classification of text data (e.g., positive and negative reviews of restaurants). In this study, multiple data visualisations such as dot plots and bar graphs were used for the evaluation of predictive features (e.g., keywords from the reviews) associated with the models (see Figure 6). Another example of this is a recent study undertaken by Rao et al. (2023) who designed a sequence of activities to guide learners how to use multiple data representations to make sense of real-world data. In the first activity, participants constructed network graphs to visualise the interactions between different characters in a movie. As can be seen in both studies, students constructed experience-distant representations (e.g., dot plots and network graphs) based on experience-near data (e.g., restaurant reviews and narratives of a movie). The transition from experience-distant to experience-near allowed learners to visually process innovative data and so built models to represent the data-driven phenomena.

Conclusion and Implications

The aim of this literature review was to explore how to incorporate data visualisations both as a process and a product into classroom teaching and learning. From the perspective of the four-quadrant relationship between data and data visualisation (see Figure 2), we examined recent studies focusing on data-driven inquiries. The literature view has identified that *transnumeration* of multiple data representations as well as *data moves* during the process of data visualisation is a way that may foster contextualisation, reasoning, and modelling with data. Moreover, the transition between experience-near and experience-distant, namely the data visualisation continuum could be used as a tool enhance students' visual reasoning in everyday teaching practice. However, most recent studies in this area are descriptive in nature and so more empirical studies are needed. Another uncertainty identified by this literature review is whether coding should be incorporated into data-based inquires involving innovative data. At last, there have been few attempts to investigate non-traditional data and innovative data visualisations (Quadrant One, Figure 2). Taken together, these results suggest that there is a need for researchers and educators to design and implement data-based inquiries which focus on using multiple data representations and data moves to investigate newer forms of data via the process of data visualisation.

References

Arcavi, A. (2003). The role of visual representations in the learning of mathematics. *Educational Studies in Mathematics*, *52*(3), 215–241. http://www.jstor.org/stable/3483015

Australian Curriculum, Assessment and Reporting Authority (ACARA). (2022). *Australian curriculum (v9.0)*. ACARA. https://v9.australiancurriculum.edu.au/

Boels, L. (2023). Reflections on gaze data in statistics education. *Teaching Statistics*, *45*(1), S40-S51. http://doi.org/10.1111/test.12340

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, *3*(3), 149–210. https://doi.org/10.1007/BF01320076

Donoho, D. (2017). 50 years of data science. *Journal of Computational and Graphical Statistics*, *26*(4), 745–766. http://doi.org/10.1080/10618600.2017.1384734

Engel, J. (2017). Statistical literacy for active citizenship: A call for data science education. *Statistics Education Research Journal*, *16*(1), 44–49. http://doi.org/10.52041/serj.v16i1.213

Erickson, T., & Chen, E. (2021). Introducing data science with data moves and CODAP. *Teaching Statistics*, *43*(1), S124–S132. http://doi.org/10.1111/test.12240

Erickson, T., & Engel, J. (2023). What goes before the CART? Introducing classification trees with Arbor and CODAP. *Teaching Statistics*, *45*(1), S104–S113. http://doi.org/10.1111/test.12347

Erickson, T., Wilkerson, M., Finzer, W., & Reichsman, F. (2019). Data moves. *Technology Innovations in Statistics Education*, *12*(1). http://doi.org/10.5070/T5121038001

Fergusson, A., & Pfannkuch, M. (2022). Introducing high school statistics teachers to predictive modelling and APIs using code-driven tools. *Statistics Education Research Journal*, *21*(2). http://doi.org/10.52041/serj.v21i2.49

Gould, R. (2021). Toward data-scientific thinking. *Teaching Statistics*, *43*(S1), S11–S22. http://doi.org/10.1111/test.12267

Gould, R., & Çetinkaya-Rundel, M. (2013). Teaching statistical thinking in the data deluge. In T. Wassong, D. Frischemeier, P. R. Fischer, R. Hochmuth, & P. Bender (Eds.), *Using tools for learning mathematics and statistics* (pp. 377–392). Springer.

Jiang, S., Lee, V. R., & Rosenberg, J. M. (2022). Data science education across the disciplines: Underexamined opportunities for K-12 innovation. *British Journal of Educational Technology*, *53*(5), 1073–1079. https://doi.org/http://doi.org/10.1111/bjet.13258

Jiang, S., Nocera, A., Tatar, C., Yoder, M. M., Chao, J., Wiedemann, K., . . . Rosé, C. P. (2022). An empirical analysis of high school students' practices of modelling with unstructured data. *British Journal of Educational Technology*, *53*(5), 1114–1133. http://doi.org/10.1111/bjet.13253

Jiang, S., Qian, Y., Tang, H., Yalcinkaya, R., Rosé, C. P., Chao, J., & Finzer, W. (2023). Examining computational thinking processes in modeling unstructured data. *Education and Information Technologies*, *28*(4), 4309–4333. http://doi.org/10.1007/s10639-022-11355-3

Leonelli, S. (2019). Data governance is key to interpretation: Reconceptualizing data in data science. *Harvard Data Science Review*, *1*(1). http://doi.org/10.1162/99608f92.17405bb6

Lim, V. Y., Peralta, L. M. M., Rubel, L. H., Jiang, S., Kahn, J. B., & Herbel-Eisenmann, B. (2023). Keeping pace with innovations in data visualizations: A commentary for mathematics education in times of crisis. *ZDM Mathematics Education*, *55*(1), 109–118. http://doi.org/10.1007/s11858-022-01449-0

Makar, K. (2016). Developing young children's emergent inferential practices in statistics. *Mathematical Thinking and Learning*, *18*(1), 1–24. http://doi.org/10.1080/10986065.2016.1107820

Makar, K., Fry, K., & English, L. (2023). Primary students' learning about citizenship through data science. *ZDM Mathematics Education*, *55*(5), 967–979. http://doi.org/10.1007/s11858-022-01450-7

Matuk, C., Vacca, R., Amato, A., Silander, M., DesPortes, K., Woods, P. J., & Tes, M. (2024). Promoting students' informal inferential reasoning through arts-integrated data literacy education. *Information and Learning Science*, *125*(3/4), 163–189. http://doi.org/10.1108/ILS-07-2023-0088

National Academies of Sciences, Engineering, & Medicine. (2023). *Foundations of data science for students in grades K–12: proceedings of a workshop.* National Academy Press.

Nolan, D., & Perrett, J. (2016). Teaching and learning data visualization: Ideas and assignments. *The American Statistician*, *70*(3), 260–269. http://doi.org/10.1080/00031305.2015.1123651

Podworny, S., Hüsing, S., & Schulte, C. (2022). A place for a data science introduction in school: Between statistics and programming. *Statistics Education Research Journal*, *21*(2), http://doi.org/10.52041/serj.v21i2.46

Rao, V. N. V, Legacy, C., Zieffler, A., & delMas, R. (2023). Designing a sequence of activities to build reasoning about data and visualization. *Teaching Statistics*, *45*(S1), S80–S92. https://doi.org/http://doi.org/10.1111/test.12341

Roth, W.-M., & MacGinn, M. K. (1997). Graphing: Cognitive ability or practice? *Science Education*, *81*(1), 91–106. https://doi.org/10.1002/(SICI)1098-237X(199701)81:1<91::AID-SCE5>3.0.CO2-X

Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley Pub. Co.

Tukey, J. W. (1990). Data-based graphics: Visual display in the decades to come. *Statistical Science*, *5*(3), 327–339. http://doi.org/10.1214/ss/1177012101

Unwin, A. (2020). Why is data visualization important? What is important in data visualization? *Harvard Data Science Review*, *2*(1). https://doi.org/10.1162/99608f92.8ae4d525

Wilkerson, M. H., & Polman, J. L. (2020). Situating data science: Exploring how relationships to data shape learning. *The Journal of the Learning Sciences*, *29*(1), 1–10. https://doi.org/10.1080/10508406.2019.1705664

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, *67*(3), 223–248. https://doi.org/http://doi.org/10.1111/j.1751-5823.1999.tb00442.x

Wise, A. F. (2020). Educating data scientists and data literate citizens for a new generation of data. *Journal of the Learning Sciences*, *29*(1), 165–181. http://doi.org/10.1080/10508406.2019.1705678