EXTRACTING CONDITIONAL PROPORTIONS FROM BAR GRAPHS

Stephanie Budgett, Maxine Pfannkuch, and Malia Puloka
The University of Auckland, New Zealand
s.budgett@auckland.ac.nz

Bar graphs are often touted as a presentation format that improves reasoning when faced with conditional probability scenarios. However, research suggests that extracting information from bar graphs in order to make conditional proportion comparisons may be challenging. Furthermore, the configuration of bar graphs can influence their underlying message. The purpose of this paper is to explore two undergraduate students' reasoning when they are asked to select appropriate bar graph representations to answer comparison questions that they posed, involving two categorical variables. Results from this small pilot study suggest that there was some confusion in deciphering the comparison of interest, and bar graph configuration contributed to conflict in conclusions drawn.

INTRODUCTION

Members of a statistically literate society must be able to engage with, and interpret, data encountered in the media and elsewhere in order to make meaningful conclusions about the situation being described. Social phenomena can be multifaceted, often requiring the consumer to wrangle the resulting numerical and graphical summaries to decode the relevant information to make informed decisions. Today's citizens have access to a myriad of online information dashboards displaying financial data, health-related data, employment data, and data from many other sources. For example, in the current pandemic climate, the public are inundated with dashboards or portals providing numeric and visual displays of data related to case numbers, vaccination rates, and variant dominance. Not all dashboards are designed equally. It is conjectured that a certain degree of statistical, data and graphical literacy is required for effective and appropriate interpretation.

Bar graphs are one of the most commonly used graphs in information dashboards, with the purpose of allowing the viewer to make visual comparisons between groups and over time (Srinivasan et al., 2018). For example, Figure 1 shows representations of COVID-19 case numbers and status, classified according to gender and ethnicity, and was downloaded from the Statistics New Zealand (STATS NZ) COVID-19 data portal that is readily available to the public.

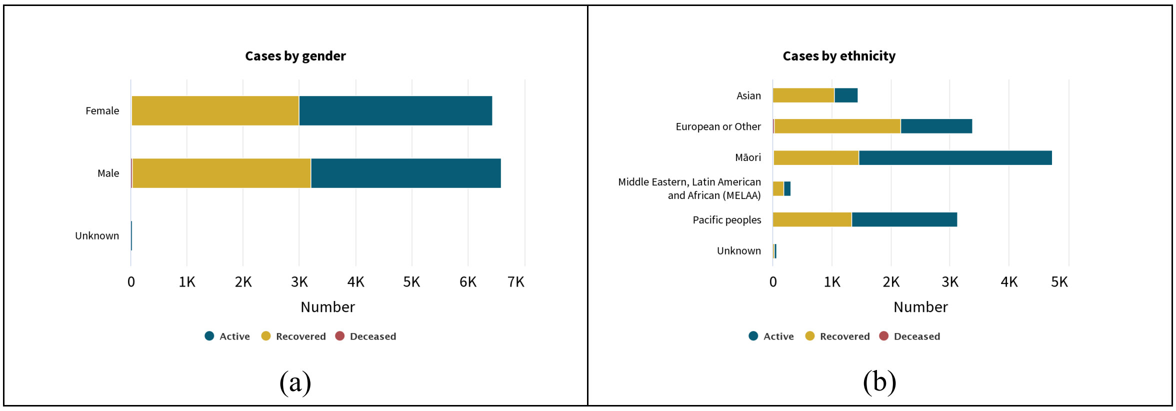


Figure 1. Screenshots from the Stats NZ COVID-19 data portal (https://www.stats.govt.nz/experimental/covid-19-data-portal, captured 15 December 2021)

Unpacking some of the key messages conveyed in Figure 1 involves comparisons *between* or *across* groups, meaning that the viewer needs to condition on one variable and consider the distribution of another. For example, to compare the Case status distribution for males and females requires conditioning on gender and comparing the distribution of Case Type for each level (Figure 1(a)). Other key messages require comparisons *within* groups, where the viewer compares counts or proportions within all levels of one variable. For example, to reason that there are more Active than Recovered cases in each of the Māori and Pacific ethnic groups requires a comparison of the counts of Active cases

In S. A. Peters, L. Zapata-Cardona, F. Bonafini, & A. Fan (Eds.), *Bridging the Gap: Empowering & Educating Today's Learners in Statistics. Proceedings of the 11th International Conference on Teaching Statistics (ICOTS11 2022), Rosario, Argentina.* International Association for Statistical Education. iase-web.org

versus Recovered cases in both the Māori ethnicity and the Pacific ethnicity (Figure 1(b)). Thus, the purpose of this small study is to focus on two first-year university statistics students' choice of bar graph and their reasoning as they made comparisons *across* and *between* groups in response to the comparison question that they posed about given categorical variables.

BACKGROUND LITERATURE

Research in the area of risk perception has found that people are much better at interpreting information if it is presented in frequency format rather than in probability format and perform best when the information is accompanied with visual representations such as bar graphs, tree diagrams, unit squares, or icon arrays (e.g., Böcherer-Linder et al., 2017; Garcia-Retamero & Hoffrage, 2013; Ghosh et al., 2008). Participants in these studies comprised university students, medical professionals, and patients. With the exception of the Ghosh et al. study, participants were required to solve Bayesian-type problems. In the Ghosh et al. study, the participants were women who were at increased risk of breast cancer and were asked to estimate their breast cancer risk with information presented either in a bar graph only, or in a bar graph accompanied by frequency format information. Importantly in these studies, participants were provided with appropriately designed visual displays that would contribute to solving the posed questions, and it was found that a combination of numeric and visual information led to more accurate interpretations, with some visualizations being more effective than others.

However, according to Xiong et al. (2002), the arrangement of bar graphs has an effect on the messages they convey, and the messages perceived may not correspond to the messages intended. For example, they presented a series of simple bar graphs, arranged in four different formats, to 76 participants and asked for their 'takeaways,' that is, what patterns or stories they perceived in the graphs. The four formats were vertical, overlaid, stacked, and adjacent (Figure 2). The participant 'takeaways' were then classified by comparison-type. It was found that bar graph arrangement had an impact on the comparisons made. That is, "visual arrangements can afford different visual comparisons in bar charts, and viewers most readily compare bars that are visually aligned and spatially proximate" (p. 962). In terms of understanding how the Active cases were distributed across ethnic groups (see Figure 1(b)), an '*across group* (ethnicity)–*within element* (Active)' comparison using Xiong et al.'s classification, either adjacent or vertical arrangements may be more informative than the stacked or overlaid arrangement.

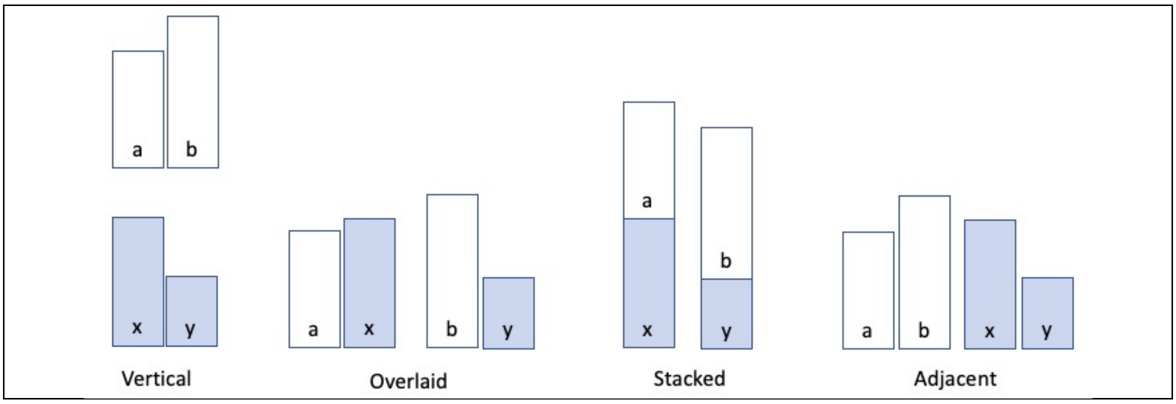


Figure 2. Bar graph arrangements considered by Xiong et al. (2022, p. 956)

When it comes to reasoning with categorical data, the fact that bar graphs can be arranged in different ways may be seen as both beneficial and detrimental. As noted by Xiong et al., although certain bar graph formats may prompt particular comparisons, conclusions made about the comparisons are not guaranteed to be accurate. In a small exploratory eye-tracking study in which participants were given comparison tasks accompanied by four bar graphs, four spider graphs, or four line graphs, more time was spent locating the appropriate bar graph to answer the question than was spent locating the appropriate spider or line graphs (Goldberg & Helfman, 2010). Easy tasks were defined as those requiring a '*within group–across element*' comparison, for example, in Figure 1(b) a comparison of Case numbers between two levels (e.g., Māori and Pacific) of one variable (ethnicity). In order to make such a comparison, two areas of interest (AOIs) need to be located, notably the heights of the Māori ethnicity bar and the height of the Pacific ethnicity bar, and subsequently compared. In easy tasks, Goldberg and Helfman (2010) found there was twice as much uncertainty time—a measure of the time taken between locating the second AOI and task completion—associated with bar graphs (around 8 seconds on average) than with spider graphs (5 seconds) and line graphs (3 seconds). Furthermore, scanpath analysis, where scanpath refers to eye-movement data captured by the eye-tracking software, indicated that all four bar graphs tended to be viewed before reaching a conclusion, whereas only one spider graph and one line graph was viewed, suggesting hesitation in deciding which was the most appropriate bar graph to use to answer the question posed. Although the authors acknowledged that the easy task bar graphs were presented to the participants first, and that this could have impacted performance, it is nevertheless important to consider how novices may naturally reason with bar graphs and make decisions about which bar graph to use when making a comparison.

Although bar graphs and other visualizations have been shown to contribute to enhanced performance in the interpretation of risk information (e.g., Böcherer-Linder et al., 2017; Garcia-Retamero & Hoffrage, 2013), these studies did not require participants to select an appropriate representation, or to articulate its interpretation. Rather, the study participants either answered multiple-choice questionnaire items or estimated numerical probability and frequency information. Conversely, viewers of dashboards are not required to answer tasks based on the information provided. Rather, they intuitively process the dashboard display to make a judgment about the key message being communicated. Whereas it may seem natural to associate bar graphs with discrete comparisons (Zacks & Tversky, 1999), knowing which comparisons are relevant in terms of decoding key messages can be problematic. For example, the fact that most of the Māori cases are classified as Active is a message conveyed by Figure 1(b) because the portion of the relevant bar length corresponding to Active is more than half of the total length. However, due to confusion of the inverse, a common misconception related to the interpretation of conditional probabilities (Villejoubert & Mandel, 2002), it is easy to see how a takeaway message such as "most of cases in the Māori group are classified as Active" could be interpreted as "most of the cases classified as Active are in the Māori group." In fact, at the time at which the screenshots were captured, less than one-half of Active cases were associated with the Māori ethnic group. Based on the representations in Figure 1, it is not easy to determine and compare the proportion of Active cases associated with each of the ethnic groups. The viewer would need to compare the Active portion of the bar associated with the Māori ethnic group with the Active portions of bars associated with the other ethnic groups, judgements not easily facilitated by the stacked arrangement.

The research question forming the basis of this paper is therefore: *Having posed questions involving categorical data, how does bar graph configuration, overlaid or adjacent, impact students' answers?*

METHOD

The data forming the basis of this paper comes from a small exploratory pilot study exploring university students' reasoning with categorical data representations. Two first-year statistics students, Sera and Tara, were the only ones, from a class of approximately 200 students, who volunteered to participate. They were familiar with proportional reasoning and conditional probability scenarios and had experience with exploring relationships between categorical variables through the use of two-way tables of counts and bar graphs. Over two 2-hour sessions, they worked together on a variety of tasks in which they were provided with scenarios involving categorical data. A 'think-aloud' protocol was adopted, where the participants were encouraged to discuss their thinking processes and resulting actions. On occasion, the researchers would intervene in order to clarify what the participants were thinking. The research method is similar to that used in a pre-clinical trial in which a proposed intervention is explored and adapted in a laboratory setting prior to implementation in humans (Schoenfeld, 2007).

In the first session, prior to viewing any data summaries or representations, the participants were asked to pose investigative questions relating to two categorical variables Gender (Male, Female) and Student Loan (Yes, No). Their questions were written on a whiteboard and then grouped, in collaboration with the researchers, into those sharing similar structures, for example simple, joint, conditional, and comparison questions. Following this, a variety of pre-prepared representations of the data, taken from an online survey voluntarily completed by first year introductory statistics students, were made available. Tara and Sera were then asked to interpret the representations and to decide which, if any, would answer the questions they posed. The second session involved exploration of Gender and the categorical variable Social Media Usage per day (none, < 1 hour, 1–3 hours, 3–6 hours, > 6 hours). The format of the second session was similar to that of the first.

The sessions were video-recorded and transcribed for analysis. A task oriented qualitative analysis approach (Bakker & van Eerde, 2015) was conducted on the session transcripts with the aim of identifying salient features of the students' thinking and reasoning that would inform the research question. In order to capture non-verbal communication, video excerpts were used to support the transcripts because the students frequently gestured to their chosen bar graph.

RESULTS

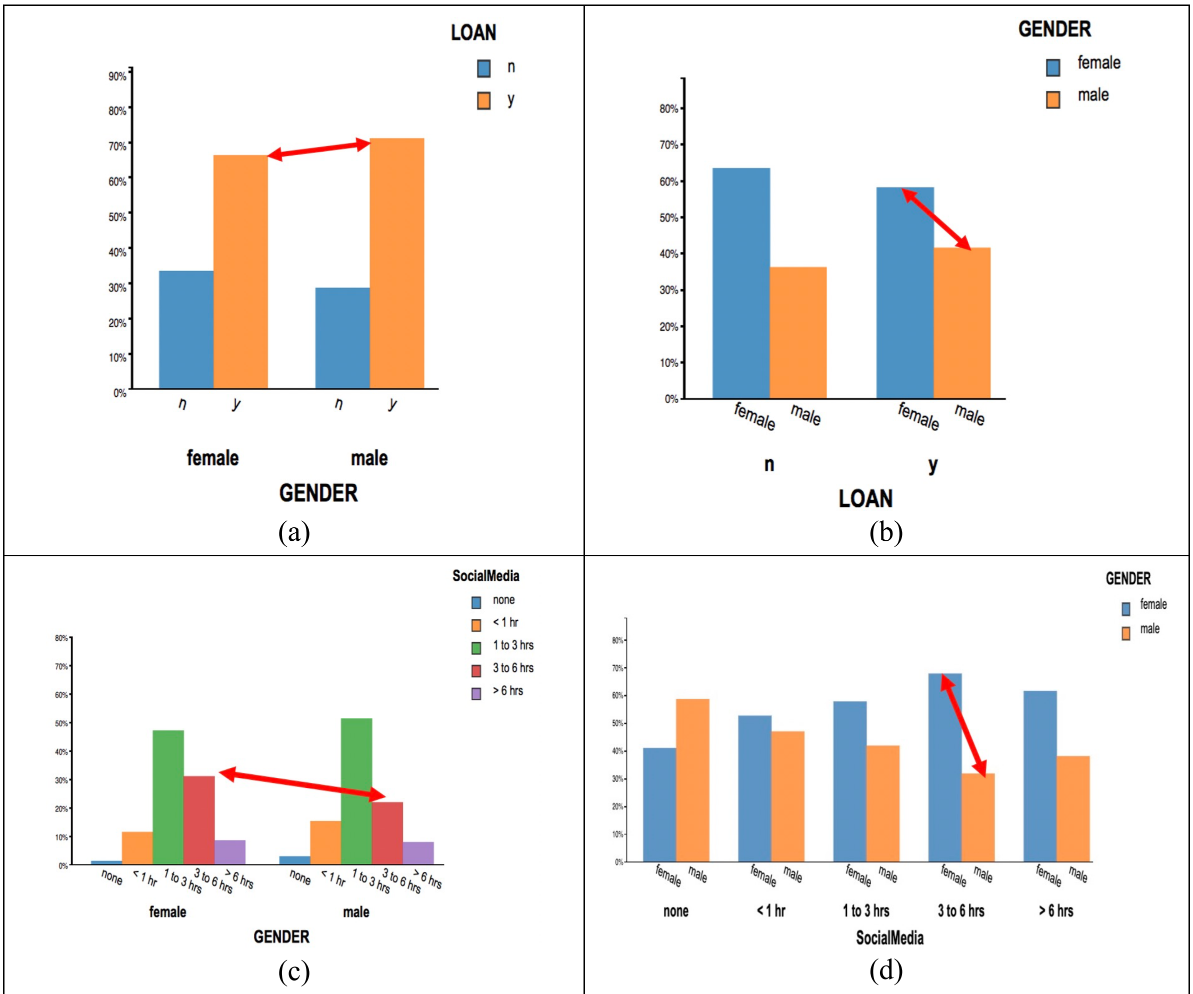
The students were given a variation of a stacked bar graph, the findings of which have been reported (Budgett & Puloka, 2019). Hence, the focus of this paper will be on the overlaid and adjacent bar graph representations used by Sera and Tara in order to answer the following two questions that they posed:

- Question 1: *Who is more likely to have a student loan, males or females?*
- Question 2*: Are females more likely to spend 3–6 hours on social media than males?*

The bar graph representations in Figure 3 were created using a prototype software tool developed at the University of Auckland (Pfannkuch & Budgett, 2017) and include both overlaid and adjacent arrangements. In order to answer Question 1, Sera stated: "*Males are more likely to have a student loan."* She pointed to the two orange bars in the representation shown in Figure 3(a) and, when probed by the researcher, "*What you indicated is to compare the height of this bar* [female, y] *with the height of that bar* [male, y]*?*", she agreed. However, when she considered the representation in Figure 3(b), she stated: *"it's more likely for females to have a student loan"* and justified her answer by pointing to the blue and orange bars indicated in Figure 3(b) and commenting that, *"if you compare these two together [it] is more likely for females to have a student loan."* Thus, Sera's answer to Question 1 varied according to the bar graph arrangement she interpreted.

When using Figure 3(a), Sera made an '*across group–within element*' comparison with grouping variable Gender having two elements (Student Loan = No, Yes) in each of its two levels (Female, Male). She visually compared the Yes element of the variable Student Loan across the two levels of the grouping variable Gender (Female/Yes vs Male/Yes) and concluded that it is more likely for males to have a student loan than it is for females. However, when she interpreted Figure 3(b), she made a '*within group–across element*' comparison, with grouping variable Student Loan having two elements in each group (No/Female, No/Male, Yes/Female, Yes/Male). She visually compared across the Male and Female elements within the Yes level of the grouping variable Student Loan (Yes/Female vs Yes/Male), concluding that it is more likely for those with a student loan to be female than male. However, these two comparisons are not equivalent. Interestingly, Sera did not seem surprised by the conflicting conclusions and appeared unaware that she had made different comparisons. She stated: *"If you compare these two together,"* referencing Figure 3(a), *"[it] is more likely for females to have a student loan, and then compare these two,"* referencing Figure 3(b), *"to say that [it] is more likely for males than females to have a student loan."*

When considering Question 2 in consultation with Figure 3(c), Tara stated: *"looking at the red bars only, out of the three-to-six-hour group, females are more likely to spend three-to-six hours in comparison to males, but only out of the three-to-six-hour group."* She then used Figure 3(d) to make the same conclusion and, estimating the height of the blue bar for the 3–6 hours/Female combination stated that "*out of the people who spend three-to-six hours on social media, 68% of them are females*." Again, these are two different comparisons. Tara made an '*across group–within element*' comparison when comparing the two red bars in Figure 3(c), comparing the same element of the variable Social Media (3–6 hours) across the two levels of the grouping variable Gender (Female/3–6 hours vs Male/3–6 hours). When comparing the blue and orange bars indicated in Figure 3(d), she made a '*within group–across element*' comparison, with grouping variable Social Media having two elements (Female, Male) in each of its five levels (none, < 1 hour, 1–3 hours, 3–6 hours, > 6 hours), visually comparing across the Male and Female elements within the 3–6 hours level of Social Media (3–6 hours/Female vs 3–6 hours/Male). She again concluded that it is more likely for those who spend 3–6 hours on Social Media to be female than male. Although Tara reached the same conclusion, these two comparisons are not equivalent. Tara's verbalization as she interpreted Figures 3(c) and (d) indicated that she was considering two different quantities, but it is unclear whether she recognized this, possibly due to the fact that the conclusion was the same for both comparisons.



Comparisons (a) and (c): Across group–within element [adjacent arrangement]	Comparisons (b) and (d): Within group–across element [overlaid arrangement]

Figure 3. Bar graph representations of Gender and: Student Loan (a), (b); Social Media (c), (d)

DISCUSSION AND CONCLUSION

In line with previous research (e.g., Goldberg & Helfman, 2010), the two participants in this small exploratory study exhibited difficulties in locating the appropriate bar graph to answer two posed comparison questions. Furthermore, comparisons made by reasoning with one bar graph arrangement did not correspond with comparisons made by reasoning with another bar graph arrangement. The bar graph arrangement appeared to influence these two participants' responses, even when the comparison being made was not the one intended. For example, when considering Question 1, the adjacent positioning of the groups Male and Female in Figure 3(a) appeared to facilitate an '*across group (Gender)–within element (Student Loan)*' comparison for Sera, whereas the overlaid arrangement seen in Figure 3(b) resulted in her making a '*within group–across element*' comparison (Xiong, et al., 2021). A similar finding was also observed in Tara's attempts to answer Question 2. Notably, Tara and Sera appeared unaware that their conclusions varied according to the bar graph used to answer the question posed.

Confusing '*across group–within element*' with '*within group–across element*' may be considered a form of confusion of the inverse. To determine which gender is more likely to have a student loan, one would compare the probability of having a student loan for males, and the corresponding probability of having a student loan for females, a comparison of P(SL|Male) and P(SL|Female), an '*across gender–within student loan*' comparison. However, in order to determine if those with a student loan were more likely to be males or females, a '*within student loan–across gender*' comparison of P(Male|SL) and P(Female|SL) is required, which swaps the conditioning variable from Gender to Student Loan.

It may be conjectured that, when exploring categorical data through the lens of bar graphs, the comparisons that we naturally make may be influenced by their arrangement. The prevalence of confusion of the inverse may therefore partly be a consequence of the fact that while we might recognise that an '*across group–within element*' comparison is required, bar graph configuration may inadvertently lead us to a '*within group–across element*' comparison. Today's citizens are bombarded every day with a myriad of data-derived information, often in the form of information dashboards that typically include bar graphs and other data visualizations. Therefore, today's students need to be able to produce, select, and interpret appropriate visual representations to accompany key messages from underlying data. As we move towards a more visual world, more effort needs to be devoted to facilitating students' reasoning with graphical representations of data and in particular extracting conditional proportions.

REFERENCES

Bakker, A., & van Eerde, D. (2015). An introduction to design-based research with an example from statistics education. In A. Bikner-Ahsbahs, C. Knipping, & N. Presmeg (Eds.), *Approaches to qualitative research in mathematics education* (pp. 429–466). Springer. https://doi.org/10.1007/978-94-017-9181-6_16

Böcherer-Linder, K., Eichler, A., & Vogel, M. (2017). The impact of visualization on flexible Bayesian reasoning. *Avances de investigacion en educacion matematica, 11*, 25–46. https://doi.org/10.35763/aiem.v1i11.169

Budgett, S., & Puloka, M. (2019). Making sense of categorical data–question confusion. In S. Budgett (Ed.), *Decision making based on data. Proceedings of the Satellite conference of the International Association for Statistical Education*, (pp. 1–6). ISI/IASE. https://iase-web.org/Conference_Proceedings.php?p=2019_Decision_Making_Based_on_Data

Garcia-Retamero, R., & Hoffrage, U. (2013). Visual representation of statistical information improves diagnostic inferences in doctors and patients. *Social Science and Medicine, 83*, 27–33. https://doi.org/10.1016/j.socscimed.2013.01.034

Ghosh, K., Crawford, B. J., Pruthi, S., Williams, C. I., Neal, L., Sandhu, N. P., Johnson, R. E., Wahner-Roedler, D., Britain, M. K., Cha, S. S., & Ghosh, A. K. (2008). Frequency format diagram and probability chart for breast cancer risk communication: A prospective, randomized trial. *BMC Women's Health, 8*(1), 1–8. https://doi.org/10.1186/1472-6874-8-18

Goldberg, J. H., & Helfman, J. I. (2010). Comparing information graphics: A critical look at eye tracking. *Proceedings of the 3rd BELIV'10 Workshop: BEyond time and errors: Novel evaLuation methods for Information Visualization* (pp. 71–78). Association for Computing Machinery. https://doi.org/10.1145/2110192.2110203

Pfannkuch, M. & Budgett, S. (2017). Reasoning from an eikosogram: An exploratory study. *International Journal of Research in Undergraduate Mathematics Education*, *3*(2), 283–310. https://doi.org/10.1007/s40753-016-0043-0

Schoenfeld, A. (2007). Method. In F. Lester, Jr. (Ed.), *Second handbook of research on the teaching and learning of mathematics* (pp. 96–107). Information Age.

Srinivasan, A., Brehmer, M., Lee, B., & Drucker, S. M. (2018). What's the difference? Evaluating variants of multi-series bar charts for visual comparison tasks. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1–12). https://doi.org/10.1145/3173574.3173878

Villejoubert, G., & Mandel, D. R. (2002). The inverse fallacy: An account of deviations from Bayes theorem and the additivity principle. *Memory & Cognition, 30*(2), 171–178. https://doi.org/10.3758/BF03195278

Xiong, C., Setlur, V., Bach, B., Lin, K., Koh, E., & Franconeri, S. (2022). Visual arrangements of bar charts influence comparisons in viewer takeaways. *IEEE Transactions on Visualization and Computer Graphics, 28*(1), 955–965. https://doi.org/10.1109/TVCG.2021.3114823

Zacks, J., & Tversky, B. (1999). Bars and lines: A study of graphic communication. *Memory & Cognition, 27*(6), 1073–1079. https://doi.org/10.3758/BF03201236