A Comparison of Good and Bad Graphs in the Statistics Pedagogy Literature

Victoria Celio

Department of Psychology, York University

PSYC 6135: Psychology of Data Visualization

Dr. Michael Friendly

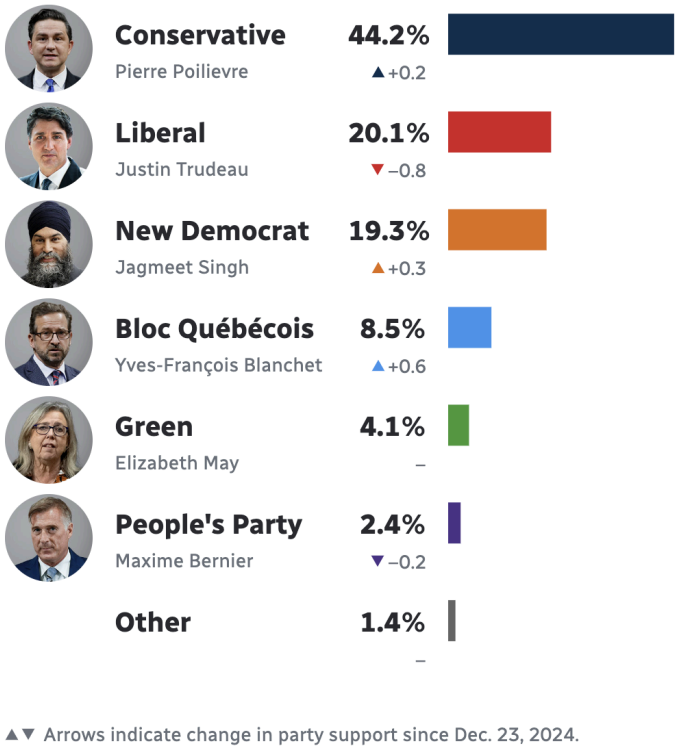
January 16, 2024

A Comparison of Good and Bad Graphs in the Statistics Pedagogy Literature

When finding examples of good and bad graphs, I looked for graphs that fit (or did not fit) my definition of "perfect." I was looking for graphs that clearly and precisely displayed the data and optimized the space and information presented in the graph through visually captivating techniques. I quickly realized that finding a graph that fits my definition of perfect would be nearly impossible. This challenge helped me recognize a central theme of last week's lecture: a good graph is defined by how well it achieves its purpose. Borrowing an aphorism from statistical modelling, one can say that *all graphs are imperfect, but some are useful*. No single formula exists to make a perfect graph in all situations. An aspect of a graph considered fundamental in one setting may hinder its effectiveness in another.

For instance, according to my original definition, Figure 1 would constitute a bad graph (Grenier, 2025). This visualization aimed to highlight how the conservative party started 2025 with a considerable lead over the other parties. Based on my original definition, this would be a bad graph because there is no y-axis scale (leading to potential distortion of the data), and the axes are not labelled. Yet, this graph effectively communicates the article's primary message despite these flaws. Ordering the parties in terms of percentage emphasizes the lead the conservative party has over the other parties. The arrows highlight how the gap between the liberal and conservative parties is increasing. The simplicity and aesthetic details (e.g., the colours of the bars correspond with each political party) make the graph visually appealing and ultimately less intimidating for readers to interpret. These points are critical, considering that this article is intended for a general audience with varying levels of graphical literacy. Although the imprecision of the graph would likely make it inappropriate for an academic setting, incorporating details that may have increased its precision (e.g., error bars) may have weakened the message for the intended audience. Therefore, what makes a good (or bad) graph is how well (or poorly) it achieves its purpose.

Figure 1



I chose examples of good and bad graphs from the statistics pedagogy literature for this assignment. Figure 2 displays an example of a bad graph (Rode & Ringel, 2019). The purpose of this graph was to highlight the statistically significant two-way interaction between output (R and SPSS output) and time (pre and post). Despite the graph containing positive characteristics (e.g., clearly labelled, precise scaling of the y-axis, no clutter, etc.), this graph is considered bad because it does not effectively communicate the two-way interaction. Viewers cannot make the required comparisons to visualize how the magnitude of the effect of output on anxiety is influenced by the level of time.

Figure 2

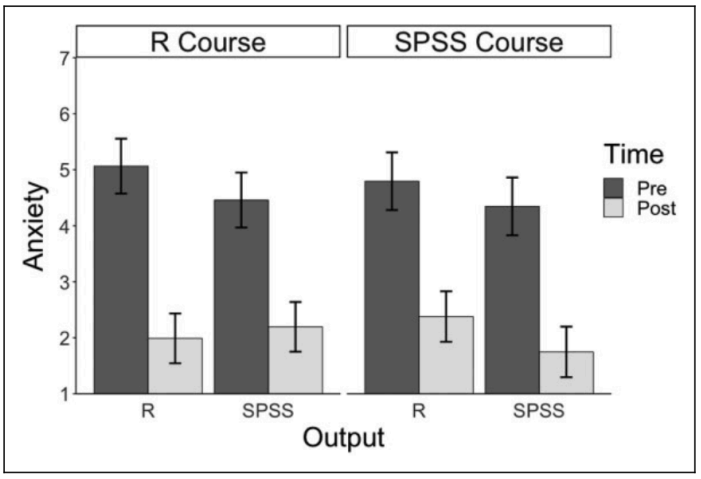


Figure 1. Students' anxiety toward learning statistical output reported at the beginning and end of the R and SPSS Courses. Error bars are 95% confidence intervals.

A good example of a visualization of an interaction is provided in Figure 3 (Perchtold-Stefan et al., 2024). These researchers were interested in visualizing the two-way interaction between time (pre-test and post-test) and group (statistics anxiety training, general anxiety training and control). The amount of information displayed in Figures 2 and 3 is the same: we see the means of each condition and measures of error; however, how the information is displayed in Figure 3 is more effective for visualizing an interaction. This visualization allows readers to see how the magnitude of the effect of time on the relative share of positive re-interpretations is impacted by the type of intervention. I also like the additional detail of using colour *and* shape to denote the different conditions for accessibility.

Figure 3

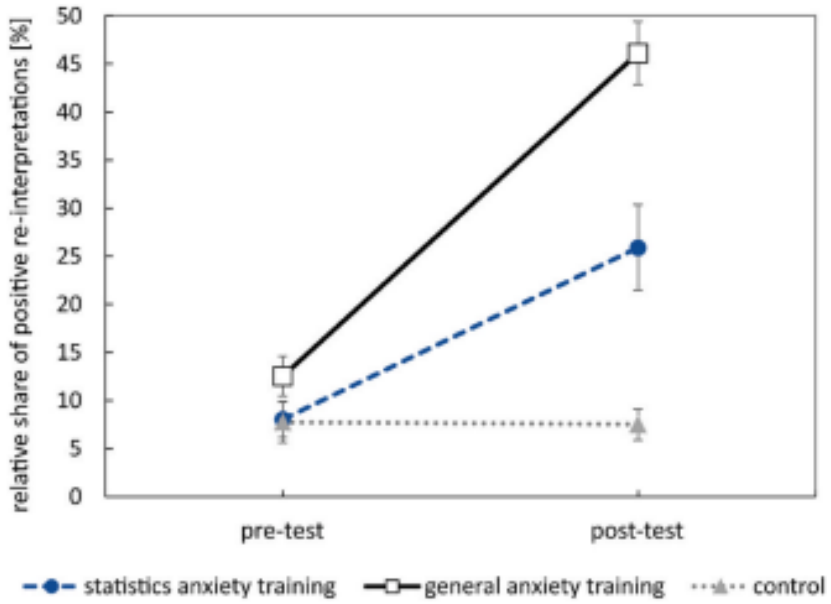


Fig. 3. Changes in relative number of generated positive Re-interpretations from pre-to post-test.

Note. Significant increases in both reappraisal training groups. Error bars denote standard errors.

References

Grenier, É. (2025, January 13). *Poll Tracker*. CBC News. https://newsinteractives.cbc.ca/elections/poll-tracker/canada/

Perchtold-Stefan, C. M., Schertler, M., Paechter, M., Fink, A., Weiss, E. M., & Papousek, I. (2024). Learning to be inventive in the face of statistics: A positive reappraisal intervention for statistics anxiety, *Journal of Behavior Therapy and Experimental Psychiatry*, *82*, 101913. https://doi.org/10.1016/j.jbtep.2023.101913

Rode, J. B., & Ringel, M. M. (2019). Statistical software output in the classroom: A comparison of R and SPSS. *Teaching of Psychology*, *46*(4), 319-327. https://doi.org/10.1177/0098628319872605