STAT 892 - Final Paper

Ryan Lalicker

December 11, 2024

WRITE ABSTRACT BEFORE SUBMITTING (Beatles (1967)) (Lauer and O'Brien (2020))

Introduction

Many areas of research rely on visualizations to express their findings to both experts in the field and the general public alike. This applies to other areas of life as well such as news, entertainment, and public health that reach the largest audience possible with easily-digestible visualizations and quick explanations over in-depth, domain specific reports. (Unwin (2020)). Although this practice goes back many centuries (Friendly, Valero-Mora, and Ibanez Ulargui (2010)), in today's fast moving-technology driven world creating visualizations has never been easier. Yet, visualizations need to be effective in order to have the desired outcome of educating the public.

A key issue in creating effective visualizations is that of bias. Sometimes bias comes from the presenter of information. Mistakes in visualizations can lead others to draw the wrong conclusions. Other times, the presenter may take advantage of an audience's trust to mislead them with purposely biased plots. Bias does not come from just the presenter though. Audience members with long-held beliefs on certain subjects may resist new information that contradicts their previous opinion.

Another issue with effectively communicating through visualization is audience understanding. While good visualizations are meant to be accessible, sometimes the best way to show data can be largely unknown to the public. This diminishes the effect of the visualization.

In this paper I will propose an experiment with the intent to account for several facets of audience interactions with plots. The experiment will measure how a general audience perceives biased plots versus more neutral plots while measuring performance across different plot types and topics. The files used in the paper can be found here.

Motivation

To understand how bias can show up in creating and interpreting visualizations we first need to discuss what makes an effective visualization. In my opinion the components of effective visualizations can be grouped into two categories: accurate representations of the data and making a visualization readable for the public. The former is an ethical responsibility for a researcher since, whether intentional or not, misleading or outright false visualizations hinder the true findings of the research. The second category has many components to it. These include taking into account accessibility concerns such as avoiding plots based in red and green together and ensuring axes and labels are readable for the viewer. Note, the latter point falls into both categories.

Perhaps the greatest challenge of making effective visualizations is considering the public's graphical literacy. An example of this occurred during the first year of the COVID-19 pandemic. Many American media outlets chose to use a logarithmic scale to illustrate the total number of deaths. As seen in Figure 1, the logarithmic scale on the right does a better job of showing how the rate of deaths is changing than the linear scale on the left. However, a study by Romano et al. (2020) found on average both experienced scientists and the general public that took part in the study understand the linear scale far more than the logarithmic scale. This is likely due to a lack of experience with log functions and illustrates how choosing the plot that one feels best represents the data is not always the one an audience understands.
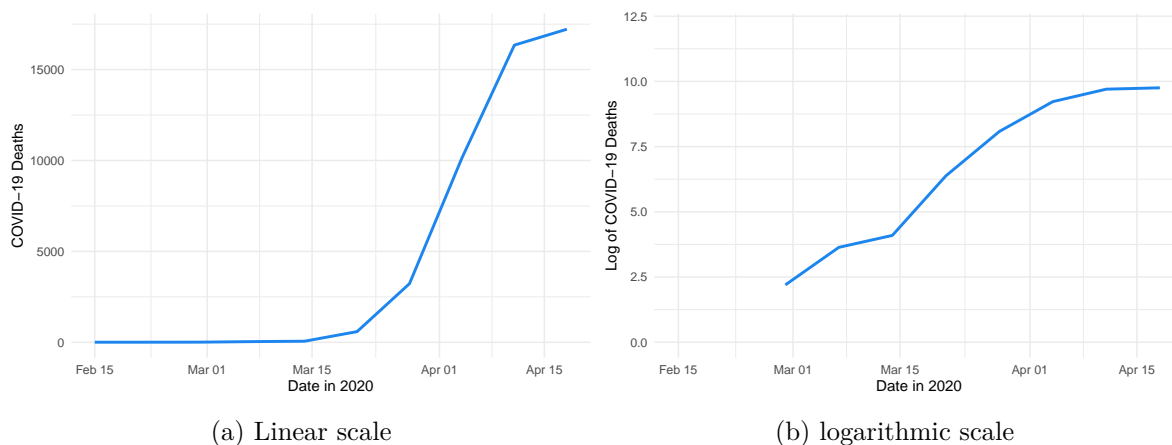

(a) Linear scale   (b) logarithmic scale

Figure 1: COVID-19 deaths in the United States. Data from Febrary 9 through April 18, 2020. Data provided by NCHS/DVS (2024).

The example above shows how the creator of a visualization can do what they believe is right and still have difficulty getting their point across due to audience misunderstanding. One element I want to look at is what types of basic visualizations do people understand best.

While it seems likely the public would struggle to understand non-linear scales, knowing how pie charts compare to bar charts in terms of audience understanding would be useful.

The main focus of this experiment, however, is how the public perceives misleading visualizations compared to accurate plots. These misleading visualizations can of course be honest mistakes by the creator or publisher of the visualization, but in some cases it may be intentionally. Some of the ways this can happen include ill-fitting chart types, the y-axis not starting at zero, data being plotted incorrectly, or even exaggerated plot and axis titles. When plots exhibiting one or more of these issues are published, viewers could consider the misleading information factual, which presents many problems for society.

An example of this is in the plot below which shows the unemployment rate through the year 2011. The photo was taken from a segment aired by Fox News. This plot is misleading in two ways. First, the y-axis goes from 8% to 10%, leading to any changes in the unemployment rate from one month to another to be more pronounced. Additionally, the November percentage of 8.6%, which appears to be the lowest of the year, is instead shown to be equivalent to the previous month's percentage of 9.0%. (Leonardo (2021)).
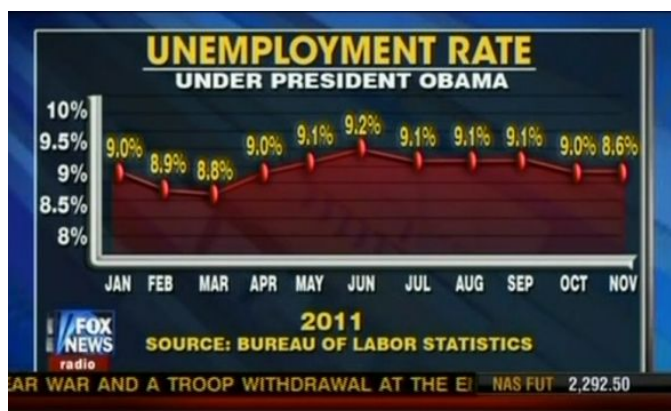

Figure 2: Plot of 2011 unemployment rate with misleading y-axis. Published by Fox News in 2011.

While this plot is not necessarily intentionally misleading, one could easily walk away with an opinion not supported by the underlying data. It is also important to keep in mind the timing and source as well. Economic issues were among the key issues Americans were concerned about following the Great Recession. This means plots that make the economic situation seem worse than it is could have negative consequences. ("Most Important Problem" (2024)). On top of this, Fox News averaged 1.87 million prime-time viewers for the year 2011, meaning many people could have been mislead at once. (Bauder (2011)).

Knowing how the public perceives misleading graphics as compared to good graphics can give researchers a way to quantify the effects of misinformation. Previous work, such as that by Lauer and O'Brien (2020) did this while controlling for the plot type. While the researchers

attempted to account for potential biases from the creators of visualizations through mixing biased and unbiased plots in the experiment, they purposely avoided the other place bias can come from: the viewer. The researchers made all plots related to noncontroversial topics.

In this experiment I will account for potential audience bias by both including plot topics that are considered controversial along with noncontroversial topics and by taking demographic information from each participant. Both groups of topics will have a mix of biased and unbiased plots, as well as different types of graphs. Together, this will allow one to see how audience members with different opinions evaluate different types of graphics depending on the sensitivity of the topic and whether they can detect misleading information.

Literary Review

As previously alluded to, the main inspiration for this experiment comes from Lauer and O'Brien (2020). Their experiment showed each participant four different kinds of plots in a random order. These plots were a bar chart, a line graph, a pie chart, and a bubble graph. Each plot had a deceptive version and a nondeceptive version, called the control, based on the same data. Note, the deceptive plot used many of the tactics discussed previously in this paper. On top of the plots themselves, two titles for each topic were written, again with one being deceptive and one not. This made it so each graph type had four different possible ways to be presented. As previously mentioned, the researchers intentionally chose noncontroversial topics.

The participants were then asked a questions about their perception of each plot and provided answers on a scale from one to six. These questions focused on potential changes or differences from one state to another. This allowed for a lower score to indicate a small change, while larger scores corresponded to larger changes. After completing a question for each of the four plot types, participants were then asked to rank the graphs from easiest to hardest to understand and finally state how comfortable they were with each plot type through multiple choice questions.

In total, the experiment had 329 participants. These participants were either first-year students in psychology classes at Arizona State University or instructors in similar fields at different universities. While this was convenient for the researchers, it is unlikely this group is representative of the population at large, since all participants are either college students or faculty.

The results of the plot interpretation are shown in the next figure. (Lauer and O'Brien (2020), pg.333). In each plot type, the largest average scores, which again indicate the greater perceived change from one state to another, come from the deceptive plots. This gap is somewhat pronounced in all plot types except the bubble graphs. Meanwhile, exaggerated plot titles seems to have played far less of a role in participant answers. This could suggest that viewers are better equipped to disregard poor headlines than poor graphs.
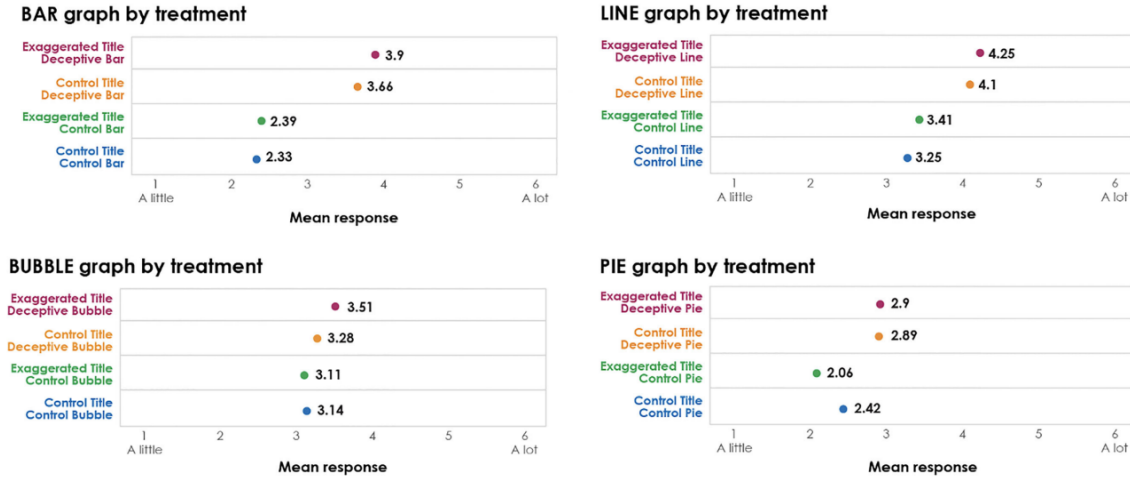
Figure 3: Mean responses to graphs in the experiment by Lauer and O'Brien (2020).

While attempting to replicate the title effect would be interesting, for this experiment I will not be focusing on plot titles. In its place, each plot type will have a deceptive and control plot related to a controversial topic with a neutral headline for all plots. This is what I believe the experiment by Lauer and O'Brien (2020) is lacking in regards to my own research interests of the role bias plays in the perception of visualizations.

For a better understanding of how to conduct an experiment that measured bias I considered a study by Schwalbe et al. (2024). The study was focused on how American individuals from different backgrounds perceive headlines and corresponding articles. Half of the articles were nonpolitical while the other half were all related to Donald Trump. These articles were graded on a scale from against to for Trump. The participants were shown an equal number of each. The articles were also real news, fake news, and a mix of both in equal numbers.

Participants were asked questions about the articles. These included whether the article was truthful and if they would share it with others. These answers taken on a five-point scale ranging from negative to positive. At the end of the study participants were asked several demographic questions, including their opinions of Donald Trump on a seven-point scale. On this scale a one implied strong opposition, a four was neutral, and a seven was strong support. Researchers chose to combine participants with scores of one to three in one group, five to seven in another, and leave scores of four as their own category.

For the articles involving Trump, researchers labeled cases concordant when the article skew matched the participants beliefs and discordant when they did not. Researchers used these labels along with participant answers to generate a political bias score. The results of this metric highlight the importance of this research as the authors found real news was more effected by bias than fake news.

Participants were recruited through a platform called Lucid. The service used quota sampling based off demographic targets to find participants on multiple websites. Researchers used quota sampling to ensure their sample reflected the U.S. Census for many demographic variables. 2,180 participants were recruited, though 371 were excluded for failing to follow survey rules or missing attention checks built into the survey.

While this study focused on articles instead of visualizations, it shows ways of measuring bias that I believe could be applied to a graphics study. Additionally, quota sampling presents an interesting approach to building a sample.

Experimental Methods

Sets of Plots

The survey used in this experiment will consist of two sets of plots: one set related to noncontroversial topics and one for controversial topics. For simplicity the controversial topics will be related to politics, health, or the environment. Specifically, the topics will be chosen if they tend to be interpreted differently based on ones position on the left-right political spectrum. The noncontroversial set will focus on topics less likely to fit onto this spectrum. The topics in both sets should be based in the country of the participants in an attempt to mimic a visualization they could see in their daily lives. For the purpose of this paper we can assume the participants are all living in the United States.

Noncontroversial	Controversial
Movies from a particular studio in a year	Changes in unemployment rate
Graduation rates at public universities	Government spending
Ticket sales of minor league baseball team	Changes in average temperatures over time
Adoption rates at a pet shelter	Gun deaths in a major city

Figure 4: Potential topics used for generating plots sorted into the sets outlined in the paper.

Some examples of potential topics are shown in Figure 4. Plots related minor league baseball teams and pet adoptions in a random U.S. town are unlikely to come with the prior biases that accompany climate change or gun violence visualizations. One key point related to some of the controversial takes is timing. In terms of where answers fall on a political spectrum, opinions on a statistic like the unemployment rate change based on who is in power. This implies that the survey portion of the experiment should not take place between a national election and the inauguration of any new president or congress.

Plot Types

For this experiment, participants will be shown four types of graphs. These will be bar charts, line plots, pie charts, and bubble graphs. Each chart type will appear once per set in a random order. As was the case in the experiment by Lauer and O'Brien (2020), each plot will have a control version and a deceptive version. The version shown to the audience will be decided through randomization.

The control and deceptive plots must follow certain guidelines. For bar charts, the control graphs must have an equal width for each bar and an axis that starts at zero. While there are many ways to make a deceptive bar chart, this experiment will focus on a y-axis starting at some number other than zero.
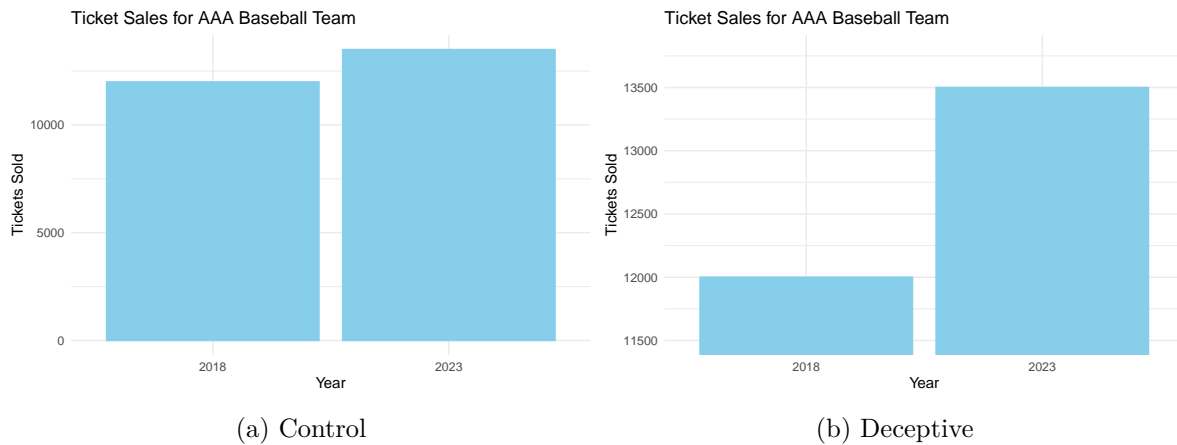

(a) Control   (b) Deceptive

Figure 5: Example of a bar chart related to ticket sales of a fictional AAA baseball team in 2018 and 2023.

Demographic Questions

Data Collection

Analysis Plan

Potential Flaws and Improvements

Conclusion

References

Bauder, David. 2011. "Nielsen Ratings: Despite Losing Viewers, Fox Dominates Cable News." Times Herald-Record. 2011.

Beatles, The. 1967. "A Day in the Life." From the album *Sgt. Pepper's Lonely Hearts Club Band.*

Friendly, Michael, Pedro Valero-Mora, and Joaquín Ibanez Ulargui. 2010. "The First (Known) Statistical Graph: Michael Florent van Langren and the 'Secret' of Longitude." *The American Statistician*, 1–11.

Lauer, Claire, and Shaun O'Brien. 2020. "How People Are Influenced by Deceptive Tactics in Everyday Charts and Graphs." *IEEE Transactions on Professional Communication* 63 (4): 327–39.

Leonardo, A. 2021. "Misleading Graphs: Real Life Examples." Statistics How To. 2021.

"Most Important Problem." 2024. Gallup. 2024.

NCHS/DVS. 2024. "Provisional COVID-19 Death Counts by Week Ending Date and State." Center for Disease Control; Prevention. https://data.cdc.gov/NCHS/Provisional-COVID-19-Death-Counts-by-Week-Ending-D/r8kw-7aab/about_data.

Romano, Alessandro, Chiara Sotis, Goran Dominioni, and Sebastián Guidi. 2020. "The Public Do Not Understand Logarithmic Graphs Used to Portray COVID-19." LSE. 2020.

Schwalbe, Michael C., Katie Joseff, Samuel Woolley, and Geoffrey L. Cohen. 2024. "When Politics Trumps Truth: Political Concordance Versus Veracity as a Determinant of Believing, Sharing, and Recalling the News." *Journal of Experimental Psychology: General* 153 (10): 2524–51. https://doi.org/10.1037/xge0001650.

Unwin, Antony. 2020. "Why Is Data Visualization Important? What Is Important in Data Visualization?" *Harvard Data Science Review* 2 (1).

Appendix A: R Code

Appendix B: AI Prompts and Answers

Prompt: How many R's are in the word strawberry?

Answer: The word strawberry contains three R's.