STAT 892 - Final Paper

Ryan Lalicker

December 10, 2024

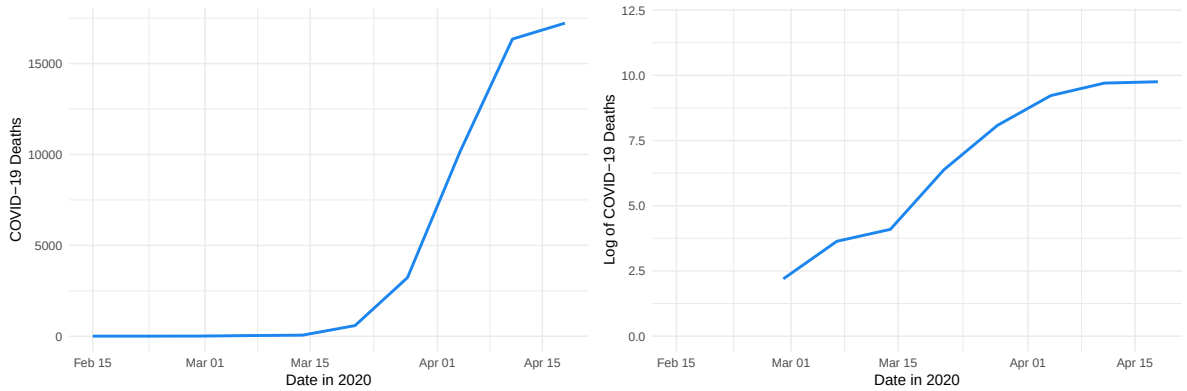
WRITE ABSTRACT BEFORE SUBMITTING (Beatles (1967)) (Lauer and O'Brien (2020))

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.

What makes a visualization effective though? In my opinion the components of effective visualizations can be grouped into two categories: accurate representations of the data and making the visualization readable for the public. The former is an ethical job 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. 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 visualizations for the public though is taking into account their graphical literacy. An example of this is 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, as a study by Romano et al. (2020) finds, people understood the linear scale far more than the logarithmic scale due to their lack of experience with the latter. This 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 accompanying files to this project can be found here, or by reaching out to me.

Motivation

Literary Review

In

Experimental Methods

Data Collection

Analysis Plan

Conclusion

References

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.

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.

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.