JSM 2021 STUDENT PAPER AWARD (ASA SECTIONS ON STATISTICAL COMPUTING AND STATISTICAL GRAPHICS)

Perception of exponentially increasing data displayed on a log scale

Emily A. Robinson[a], Reka Howard[a], Susan VanderPlas[a]

[a]Department of Statistics, University of Nebraska - Lincoln,

ARTICLE HISTORY
Compiled October 25, 2020

ABSTRACT
Log scales are often used to display data over several orders of magnitude within one graph. During the COVID pandemic, we've seen both the benefits and the pitfalls of using log scales to display data. This paper aims to...

KEYWORDS
Exponential; Log; Visual Inference; Perception

Emily and Reka, when we get into the editing stage, I've found this strategy to be useful: basically, when you add new text, use your color (feel free to change the command, for now yours are set to \er and \rh. The way this usually works is that when e.g. I read over a document that Emily has recently edited, I will remove her flagged text to indicate that I've seen/accepted the changes (and vice versa - I'll edit text and highlight it with my color, and you can accept/modify and flag yours too) – sometimes modifications happen first and then all of the color in a paragraph gets taken out once we've moved on. This not only leads to a nice rainbow effect, but you can quickly spot changes, too. If you're changing some slight phrasing/wording that doesn't change meaning, it's not necessary to highlight those changes - highlight content changes, not e.g. verb tenses. If something is a comment and has been addressed, comment it out initially and then delete the line after a couple of weeks.

1. Introduction and Background

- Why Graphics (communication to the public, technological advances, need for research on graphics)

Graphics are a useful tool for displaying and communicating information. Researchers include graphics to communicate their results with other researchers and news sources rely on graphics to convey news stories to the public. Better software has meant easier and more flexible drawing, consistent themes, and higher standards (Unwin (2020)).

- Introduce Log Scales (what are they used for, where are they used (ecological

CONTACT Emily A. Robinson. Email: emily.robinson@huskers.unl.edu, Reka Howard. Email: rekahoward@unl.edu, Susan VanderPlas. Email: susan.vanderplas@unl.edu

data, covid, etc.))

One common graphical display choice is the use of log scales used to display data over several orders of magnitude within one graph. Logarithms convert multiplicative relationships to additive ones, providing an elegant way to span many orders of magnitude, to show elasticities and other proportional changes, and to linearize power laws (Menge et al. (2018)). We have recently experienced the use of log scales first-hand as case counts data from the novel coronavirus, COVID-19, were displayed on both the log and linear scale. PUT SOME DASHBOARD REFERENCES HERE

- Previous exponential (log/linear scale) studies (literature review).
- Visual Inference (what is it? how do we use it? etc.)
- What is new in this paper.

Buja et al. (2009)

2. Data Generation

2.1. *Model Generation and Simulation*

2.2. *Parameter Selection*

- Use of lack of fit statistic.
- Mapping parameter selections to what we see visually.
- Curvature (Easy/Medium/Hard)

3. Study Design

3.1. *Lineup Setup*

3.2. *Participant Recruitment*

3.3. *Task Description*

- Lineup
  - The goal of this is to test an individuals ability to perceptually differentiate exponentially increasing data with differing rates of change on both the linear and log scale.

4. Results

4.1. *Effect of Curvature*

4.2. *Effect of Variability*

4.3. *Linear vs Log*

4.4. *Participant Reasoning*

5. Discussion

5.1. *Conclusion*

5.2. *Future Research*

- What we learned from lineups but what we still want to learn.
- You draw it
  - Mosteller et al. (1981) designed and carried out an empirical investigation to explore properties of lines fitted by eye. The researchers found that students tended to fit the slope of the first principal component or major axis (the line that minimizes the sum of squares of perpendicular rather than vertical distances) and that students who gave steep slopes for one data set also tended to give steep slopes on the others. Interestingly, the individual-to-individual variability in slope and in intercept was near the standard error provided by least squares for the four data sets.
  - The goal of this task is to test an individuals ability to make predictions for exponentially increasing data.
  - Previous literature suggests that we tend to underestimate predictions of exponentially increasing data. *find reference*
  - The idea for this task was inspired by the New York Times "You Draw It" page which is fun to check out.
- Estimation
  - This tests an individuals ability to translate a graph of exponentially increasing data into real value quantities. We then ask individuals to extend their estimates by making comparisons across levels of the independent variable.
  - Friel, Curcio, and Bright (2001) emphasize the importance of graph comprehension proposing that the graph construction plays a role in the ability to read and interpret graphs.

Supplementary Materials

Acknowledgement(s)

References

Buja, Andreas, Dianne Cook, Heike Hofmann, Michael Lawrence, Eun-Kyung Lee, Deborah F. Swayne, and Hadley Wickham. 2009. "Statistical inference for exploratory data analysis and model diagnostics." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 367 (1906): 4361–4383. Accessed 2020-10-06. `https://royalsocietypublishing.org/doi/10.1098/rsta.2009.0120`.

Friel, Susan N., Frances R. Curcio, and George W. Bright. 2001. "Making Sense of Graphs: Critical Factors Influencing Comprehension and Instructional Implications." *Journal for Research in Mathematics Education* 32 (2): 124. Accessed 2020-05-29. https://www.jstor.org/stable/749671?origin=crossref.

Menge, Duncan N. L., Anna C. MacPherson, Thomas A. Bytnerowicz, Andrew W. Quebbeman, Naomi B. Schwartz, Benton N. Taylor, and Amelia A. Wolf. 2018. "Logarithmic scales in ecological data presentation may cause misinterpretation." *Nature Ecology & Evolution* 2 (9): 1393–1402. Accessed 2020-08-18. http://www.nature.com/articles/s41559-018-0610-7.

Mosteller, Frederick, Andrew F. Siegel, Edward Trapido, and Cleo Youtz. 1981. "Eye Fitting Straight Lines." *The American Statistician* 35 (3): 150–152. Accessed 2020-05-29. http://www.tandfonline.com/doi/abs/10.1080/00031305.1981.10479335.

Unwin, Anthony. 2020. "Why is Data Visualization Important? What is Important in Data Visualization?" *Harvard Data Science Review* Accessed 2020-04-27. https://hdsr.mitpress.mit.edu/pub/zok97i7p.