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 26, 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

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 in scientific publications and news sources rely on graphics to convey news stories to the public. During the onset of the novel coronavirus - COVID19 - pandemic, we saw an influx of dashboards being developed to displaying case counts, transmission rates, and outbreak regions (Charlotte 2020). As a result, people began subscribing to news sources involved in graphically tracking the coronavirus (example John Burn-Murdoch Financial Times - SITE THIS) and gaining more exposure to the use of graphics. Many of these graphics helped guide decision makers to implement policies such as shut-downs or mandated mask wearing. Better software has meant easier and more flexible drawing, consistent themes, and higher standards. As a consequence, principles are needed on how to decide which of many possible graphics to draw (Unwin 2020).

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

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

laws (Menge et al. 2018). When presenting log-scaled data, it is possible to use either untransformed values (for example, values of 1, 10 and 100 are equally spaced along the axis) or log-transformed values (for example, 0, 1, and 2). We have recently experienced the benefits and pitfalls of using log-scales as COVID-19 dashboards displayed case count data on both the log and linear scale (Fagen-Ulmschneider 2020). INSERT BENEFITS AND PITFALLS OF LOG SCALES HERE. While COVID-19 is the most well known example, log-scales have been used to display data in ecological research, etc. PUT OTHER AREAS HERE.

- Previous exponential (log/linear scale) studies (literature review).
    - Our default is on the log scale and the linear scale is a learnt behavior (Varshney and Sun 2013)
        - whole number magnitude representations progress from a compressive, approximately logarithmic distribution to an approximately linear one. Transitions occur earlier for smaller than for larger ranges of whole numbers, corresponding both to the complexity of the numbers and to the ages when children gain experience with them. In summary, estimation proceeds logarithmically initially and transitions to linear later in development, for several different numerical ranges. (Siegler and Braithwaite 2017)
        - in American children, logarithmic mapping does not disappear all at once, but vanishes first for small numbers and much later for larger numbers from 1 to 1000 (up to fourth or sixth grade in some children). (Dehaene et al. 2008)
    - (Jones 1979), (Jones 1977), (Wagenaar and Timmers 1978)
    - misconceptions (Menge et al. 2018)
    - descrimination between curve types is possible (Best, Smith, and Stubbs 2007)
        - Compared many factors: exponential, asymptotic, and linear trends increasing or decreasing bar, suspended bar, scatter, and line plots number of points high, medium, low variability
        - Asked to identify the type of curve (exponential, asymptotic, linear; increasing, decreasing)
        - hypothesis is 2-stage estimation: first, identify the type of curve and direction, then use that information for prediction
        - this experiment is examining whether discrimination between curve types is possible
        - Results
            - accuracy higher when nonlinear trends presented (e.g. it's hard to say something is linear, but easy to say that it isn't)
            - accuracy higher with low variability – variability was additive, e.g. constant variance around mean function it appears that participants examined curvature to make the determination of type
- Visual Inference (what is it? how do we use it? etc.)
    - lineup protocol (Buja et al. 2009, Wickham et al. (2010), Hofmann et al. (2012), Majumder, Hofmann, and Cook (2013), VanderPlas and Hofmann (2017)) REREAD/SKIM ALL THESE
        - Statistical lineups have previously been utilized in graphical experiments to quantify the...
    - visual p-values (VanderPlas et al. n.d.)

- What is new in this paper.

2. Data Generation

The most common type of lineup used in graphical experiments is a standard lineup containing one "target" dataset embeded within a set of null datasets. One way to generate the null datasets when working with real data is through the use of permutation. In this study, the target dataset was generated by model A while the null datasets were generated by model B. FIX WORDING HERE

2.1. *Exponential Model*

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*

Participants were recruited from Reddit. GIVE SUMMARY DESCRIPTIVE STATISTICS OF PARTICIPANT DEMOGRAPHICS.

3.3. *Task Description*

- Lineup Task
  - 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

Best, Lisa A., Laurence D. Smith, and D. Alan Stubbs. 2007. "Perception of Linear and Nonlinear Trends: Using Slope and Curvature Information to Make Trend Discriminations." *Perceptual and Motor Skills* 104 (3): 707–721. Publisher: SAGE Publications Inc, Accessed 2020-07-06. https://doi.org/10.2466/pms.104.3.707-721.

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`.

Charlotte, Lisa. 2020. "You've informed the public with visualizations about the coronavirus. Thank you." Jul. `https://blog.datawrapper.de/datawrapper-effect-corona/`.

Dehaene, Stanislas, Véronique Izard, Elizabeth Spelke, and Pierre Pica. 2008. "Log or Linear? Distinct Intuitions of the Number Scale in Western and Amazonian Indigene Cultures." *Science* 320 (5880): 1217–1220. 00651 Publisher: American Association for the Advancement of Science Section: Report, Accessed 2020-05-19. `https://science.sciencemag.org/content/320/5880/1217`.

Fagen-Ulmschneider, Wade. 2020. "91-DIVOC." `http://91-divoc.com/pages/covid-visualization/`.

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`.

Hofmann, Heike, Lendie Follett, Mahbubul Majumder, and Dianne Cook. 2012. "Graphical Tests for Power Comparison of Competing Designs." *IEEE Transactions on Visualization and Computer Graphics* 18 (12): 2441–2448. Accessed 2020-04-06. `http://ieeexplore.ieee.org/document/6327249/`.

Jones, Gregory V. 1977. "Polynomial perception of exponential growth." *Perception & Psychophysics* 21 (2): 197–198. Accessed 2020-06-25. `http://link.springer.com/10.3758/BF03198726`.

Jones, Gregory V. 1979. "A generalized polynomial model for perception of exponential series." *Perception & Psychophysics* 25 (3): 232–234. Accessed 2020-05-19. `http://link.springer.com/10.3758/BF03202992`.

Majumder, Mahbubul, Heike Hofmann, and Dianne Cook. 2013. "Validation of Visual Statistical Inference, Applied to Linear Models." *Journal of the American Statistical Association* 108 (503): 942–956. Accessed 2020-02-28. `http://www.tandfonline.com/doi/abs/10.1080/01621459.2013.808157`.

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`.

Siegler, Robert S., and David W. Braithwaite. 2017. "Numerical Development." *Annual Review of Psychology* 68 (1): 187–213. Accessed 2020-05-19. `http://www.annualreviews.org/doi/10.1146/annurev-psych-010416-044101`.

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`.

VanderPlas, Susan, and Heike Hofmann. 2017. "Clusters Beat Trend!? Testing Feature Hierarchy in Statistical Graphics." *Journal of Computational and Graphical Statistics* 26 (2): 231–242. Accessed 2020-02-28. `https://www.tandfonline.com/doi/full/10.1080/10618600.2016.1209116`.

VanderPlas, Susan, Christian Rottger, Dianne Cook, and Heike Hofmann. n.d. "Statistical Significance Calculations for Scenarios in Visual Inference." Preprint, Accessed 2020-09-29. `https://github.com/srvanderplas/visual-inference-alpha`.

Varshney, Lav R., and John Z. Sun. 2013. "Why do we perceive logarithmically?" *Significance*

10 (1): 28–31. Accessed 2020-05-07. http://doi.wiley.com/10.1111/j.1740-9713.2013.00636.x.

Wagenaar, W. A., and H. Timmers. 1978. "Extrapolation of exponential time series is not enhanced by having more data points." *Perception & Psychophysics* 24 (2): 182–184. Accessed 2020-05-19. http://link.springer.com/10.3758/BF03199548.

Wickham, Hadley, Dianne Cook, Heike Hofmann, and Andreas Buja. 2010. "Graphical inference for infovis." *IEEE Transactions on Visualization and Computer Graphics* 16 (6): 973–979.