Project Description

1 Objectives

Statistical graphics are powerful and efficient tools to convey data to support human decision making; however, empirical research on graphical perception is sparse relative to the number of decisions necessary to make a good chart; when relevant studies are available, they often use incomparable methods and produce conflicting results. Many recommendations are based on opinion rather than empirical study, rendering scientific communications sub-optimal or ineffective. This is alarming, as effective science communication is critical for cultivating public trust in the scientific process and ensuring that decision makers accurately interpret information when making choices which impact people's lives.

Our visual systems are uniquely optimized for efficient assessment of visual information, which is why data visualizations are so effective. Some disciplines, such as forensic pattern analysis (shoe-prints, firearms, handwriting) take this further, using subjective visual evaluations in place of quantitative methods. Fundamentally, these comparisons are similar to data visualization: the image encodes numerical data and actionable information is extracted via human perception. We can use the insights we gain by studying perception of graphics to "close the loop", building empirical, explainable features for pattern evidence that replicate visual comparisons performed by forensic examiners, producing machine-learning algorithms to provide objective, quantitative support for examiner testimony in court.

This CAREER proposal addresses a fundamental research question underpinning both of these problems: *How do humans make decisions supported by data visualizations?* Three research objectives (ROs) will support the overall **research goal** of advancing our understanding of the role of data visualizations in human decision making:

- **RO1:** Examine effectiveness of charts across different tasks, such as visual search, comparison, prediction, and estimation, by developing multi-modal methods for graphical testing. If successful, this will produce testing methods that focus on specific cognitive tasks related to use of scientific visualizations, and will facilitate comparison of results across experiments and methods.
- **RO2:** Empirically validate visualization design guidelines, measuring the impact of design decisions on performance of different tasks. If successful, this will produce nuanced guidelines that are task-focused and empirically derived.
- **RO3:** Produce quantitative statistical models for explainable, reproducible decision support using images by engineering statistical features to mimic human perception.

Integrated with these research efforts, the overall **education goal** is to cultivate statistical learning and scientific decision making in society. Three education objectives (EOs) address this goal:

- **EO1:** Develop and implement experiential learning activities in graphics for undergraduate introductory statistics courses.
- **EO2:** Create and implement open educational resources (OER) to introduce reproducible science and open-source software development in statistical programming courses.
- **EO3:** Engage with forensic scientists, lawyers, and judges, evaluating the scientific support for forensic disciplines and promoting open, reproducible scientific practices.

The research, educational activities, and outreach conducted in support of these objectives will advance our understanding of the design, perception, and use of visual data displays. New suites of task-focused testing methods will facilitate precise, targeted experiments in data visualization, and applying multiple task-focused testing methods to examine design guidelines will facilitate better scientific communication and dissemination of results, building trust in science and leveraging scientific principles. Building statistical models to mimic human perception takes a new approach to explainable machine learning by creating understandable features during the model design process; this facilitates public trust in the model by through analogies to the human experience.

Measuring the pedagogical effectiveness of experiential learning in undergraduate statistics courses will provide new understanding of effective strategies for STEM outreach and engagement. Incorporating students into the research process with experiential learning and using graphics to motivate statistical concepts will engage students at a deeper level and open up STEM topics to math-phobic students in a tangible way. Weaving scientific communication and reproducibility into courses on statistical programming will produce a new generation of scientists equipped with the tools to conduct better, more reproducible research and communicate the results to other scientists and to the public.

The research and educational activities described in this project have the potential to significantly improve how scientists communicate scientific results to each other as well as to the general public, increasing public trust in science and facilitating public decision making based on experimental data and results.

2 Research Plan

2.1 Background Information

Statistical visualizations are a primary way scientists communicate results to other scientists as well as the general public, but we have been arguing about best practices for the visual display of data almost since these graphical forms were created [1–3]. Different individuals and organizations have released guidelines for graphical design [4–8], but many of these are based on aesthetic preferences, individual experience, or extrapolation from limited empirical studies; for an overview, see [9]. Even when there are empirical studies to support guidelines for choosing one graphical form over another [10–12, see 13 for a review], these studies measure different quantities in different ways ; the design and implementation differences have a potentially large effect on the generalizability of the results. Complicating this problem, the design space of visualization user studies is incredibly large [14], and studies can address charts from multiple perspectives, including perception, task-based interaction (with different tasks producing different results), structural characteristics, and more [15]. In addition, the test population is an important and under-examined aspect of the graphics experiment design space: we know that expertise as well as disorders such as dyslexia, dyscalculia, and ADHD affect perception, numeracy, and other processes involved in graph comprehension [16–18]; while these subpopulations can be difficult to test compared to the general population, empirical design recommendations must also address accessibility and audience considerations, as designers already consider these issues [19] but have no firm guidance about best practice. Relatively few studies directly compare the effect of different approaches for testing graphics, but minor changes to data collection methods can impact accuracy of even simple numerical estimation tasks [20]; it is reasonable to expect that the specifics of the graphical testing procedure used may impact the results, potentially explaining conflicting empirically based recommendations [2, 3, 12]. Without the ability to integrate results from different types of studies, it is hard to generate scientifically accurate recommendations from existing research. To bridge this gap, information visualization needs a suite of easily implemented, validated graphical testing methods that are systematically applied to test existing design guidelines in order to produce nuanced, empirically grounded recommendations.

2.2 Preliminary Data

During the early stages of the pandemic, log scales were fairly common because they allow comparison of different orders of magnitude on the same chart; however, it was difficult to find studies evaluating how successfully log scales were read and interpreted, though there were studies suggesting that people systematically under-predict exponential growth [21, 22]. In order to address this gap in the literature, I worked with a graduate student, Emily Robinson, to design and execute a series of three studies addressing different tasks related to log scales: could participants detect a difference in rates of exponential growth? Could they predict future values? Could they estimate additive and multiplicative differences from an exponential time series? We examined these questions by testing charts with log and linear scales using the lineup protocol [**buja_statistical_2009?**], participant drawn trends, and questions that required numerical calculation in a single omnibus study conducted using Prolific. Each set of results was analyzed separately [23, additional publications in progress], with additional publications discussing the implementation [24] and validating the new method [25] to gather data using participant-drawn trend lines. As expected when examining a single graphical choice using different methods, the results were mixed: using lineups, we determined that overall we have a hard time detecting changes in curvature, though we can detect a curve among straight lines more easily than we can detect a line (or less-steep curve) among curves. When assessing predictions, we found that the use of log scales reduced under-prediction of exponential growth. Finally, when examining participants' ability to estimate quantities from both types of charts, we found that participa Future work related to this study will analyze the differences in participant performance across methods to assess the importance of individual characteristics such as skill level and educational background.

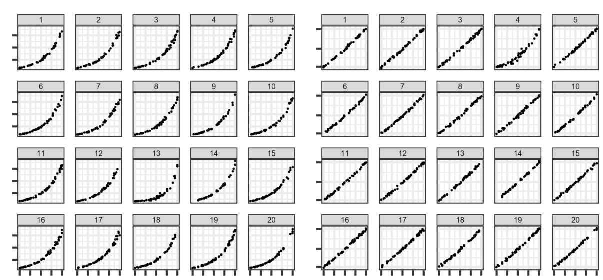

Figure 1: Lineups (log and linear scale) test detection of differences in exponential growth rates.

While there are many graph design and scientific communication questions which have not yet been experimentally tested, there is another problem with current literature: some comparisons were tested experimentally using graphics which were state-of-the-art at the time, but have not been re-examined in light of technological developments in the past 40-60 years. For instance, [10] examined the use of fixed 3D projections of 2D data, but developments since the 1980s mean that analysts can now create 3D digital environments with appropriate shading and interactivity, and can even print these creations to create 3D charts that can be physically manipulated, touched, and considered from all angles. Are the guidelines that suggest 3D renderings are less accurately read and misleading still valid under these conditions?

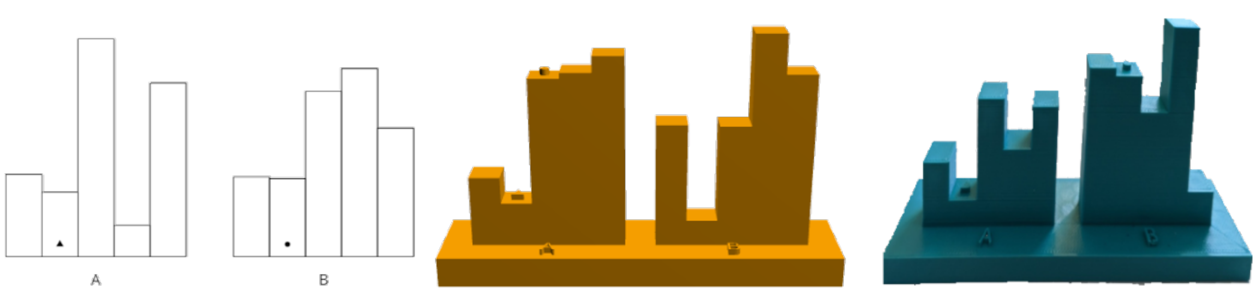

Figure 2: Three renderings of bar charts: 2D (left), 3D digital render (middle), and 3D printed (right).

A graduate student and I have been investigating this question in two different populations: a population comparable to that used in [10] (statistics faculty and graduate students and their roommates), and a population of introductory statistics students as part of an experiential learning project; the latter population is being tested in Summer 2023, but preliminary data is available from the former population. We have created charts rendered using 2D graphics, 3D digital renderings [26], and 3D-printed graphics [27], as shown in Figure 2. Our initial investigation found few differences between 2D, 3D digital, and 3D printed charts in comparison accuracy; as a result, we are expanding the study to determine whether this is a technological difference and the decreased accuracy in [10] is a result of the 3D fixed-angle projection that does not allow for realistic manipulation. Misapplied depth perception has been implicated in other graphical mis-perceptions [28, 29], and it is entirely possible that 3D charts that are interactive can be accurately perceived while artificial 3D fixed-angle projections into 2D space lead to inaccurate perceptions. If this is the case, the guidelines to avoid 3D graphics may be entirely misguided given the rendering algorithms available today.

2.3 Methods

Link to Excel Sheet

This project is designed to leverage human perception and statistics to facilitate better communication about data and statistical models, as shown in Figure 3.

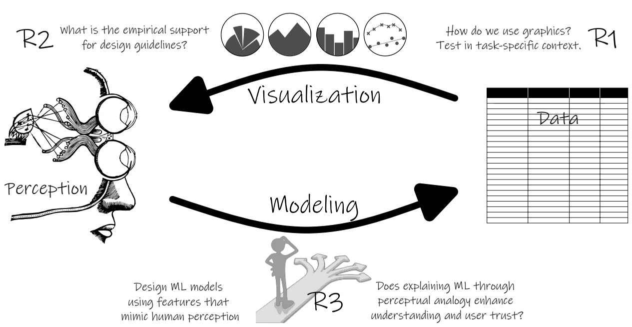

Figure 3: Relationship between research aims.

To address this broad goal, we will benchmark methods for testing statistical graphics that correspond to typical uses for statistical graphics. We will then use these methods to assess design guidelines and recommendations to provide a more nuanced, task-focused, empirical evaluation of these design criteria; this will provide clarity in the importance and impact of design guidelines and, if successful, will allow us to reconcile conflicting historical experiments as well. While these two goals are related, they are not dependent: there are already sufficient methods in the literature for testing graphics to allow us to complete a task-focused evaluation of design guidelines; our contribution in the first aim is to modernize historical methods, fill in gaps in task-focused analysis methods, provide direct comparisons between methods, and create guidelines for the use of each method when testing statistical graphics.

3 Education Plan

3.1 Overview

3.2 Design and methods

3.3 Evaluation

3.4 Integration of research and Education

4 Broader Impacts

5 Timeline

References

1. Wickham H (2013) Graphical criticism: Some historical notes. *Journal of Computational and Graphical Statistics*, 22(1):38–44. https://doi.org/10.1080/10618600.2012.761140

2. Croxton FE, Stryker RE (1927) Bar Charts Versus Circle Diagrams. *Journal of the American Statistical Association*, 22(160):473–482. https://doi.org/10.2307/2276829

3. Eells WC (1926) The Relative Merits of Circles and Bars for Representing Component Parts. *Journal of the American Statistical Association*, 21(154):119–132. https://doi.org/10.2307/2277140

4. (1915) Joint committee on standards for graphic presentation. *Quarterly Publications of the American Statistical Association*, 14(112):790–797. https://doi.org/10.1080/15225445.1915.10503668

5. Brewer CA (1994) Guidelines for use of the perceptual dimensions of color for mapping and visualization. *Color hard copy and graphic arts III*, 2171:54–63. https://doi.org/10.1117/12.175328

6. Kosslyn SM (2006) Graph design for the eye and mind.

7. Tufte E (2001) The visual display of quantitative information.

8. Kelleher C, Wagener T (2011) Ten guidelines for effective data visualization in scientific publications. *Environmental Modelling & Software*, 26(6):822–827. https://doi.org/10.1016/j.envsoft.2010.12.006

9. Kandogan E, Lee H (2016) A Grounded Theory Study on the Language of Data Visualization Principles and Guidelines. *Electronic Imaging*, 28:1–9. https://doi.org/10.2352/ISSN.2470-1173.2016.16.HVEI-132

10. Cleveland WS, McGill R (1984) Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387):531–554. https://doi.org/10.1080/01621459.1984.10478080

11. Hofmann H, Follett L, Majumder M, Cook D (2012) Graphical tests for power comparison of competing designs. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2441–2448. https://doi.org/10.1109/TVCG.2012.230

12. Spence I (1990) Visual psychophysics of simple graphical elements. *Journal of Experimental Psychology: Human Perception and Performance*, 16:683–692. https://doi.org/10.1037/0096-1523.16.4.683

13. Vanderplas S, Cook D, Hofmann H (2020) Testing statistical charts: What makes a good graph? *Annual Review of Statistics and Its Application*, 7(1):61–88. https://doi.org/10.1146/annurev-statistics-031219-041252

14. Abdul-Rahman A, Chen M, Laidlaw DH (2020) A survey of variables used in empirical studies for visualization. *Foundations of data visualization*, :161–179. https://doi.org/10.1007/978-3-030-34444-3_7

15. Bolte F, Bruckner S (2020) Measures in visualization space. *Foundations of data visualization*, :39–59. https://doi.org/10.1007/978-3-030-34444-3_3

16. Cheng D, Xiao Q, Chen Q, Cui J, Zhou X (2018) Dyslexia and dyscalculia are characterized by common visual perception deficits. *Developmental Neuropsychology*, 43:497–507. https://doi.org/10.1080/87565641.2018.1481068

17. Chity N, Harvey J, Quadri S, Pete S, Stein S (2012) Thinking DIfferently. Assistive Technology as a Complement to the Learning Style of Post-Secondary Students with ADHD: Recommendations for Design Opportunities.

18. Hokken MJ, Krabbendam E, van der Zee YJ, Kooiker MJG (2023) Visual selective attention and visual search performance in children with CVI, ADHD, and Dyslexia: A scoping review. *Child Neuropsychology*, 29(3):357–390. https://doi.org/10.1080/09297049.2022.2057940

19. Bako HK, Liu X, Battle L, Liu Z (2023) Understanding how Designers Find and Use Data Visualization Examples. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):1048–1058. https://doi.org/10.1109/TVCG.2022.3209490

20. Honda H, Kagawa R, Shirasuna M (2022) On the round number bias and wisdom of crowds in different

response formats for numerical estimation. *Scientific Reports*, 12(1):8167. https://doi.org/10.1038/s41598-022-11900-7

21. Timmers H, Wagenaar WA (1977) Inverse statistics and misperception of exponential growth. *Perception & Psychophysics*, 21(6):558–562. https://doi.org/10.3758/bf03198737

22. Wagenaar WA, Sagaria SD (1975) Misperception of exponential growth. *Perception & Psychophysics*, 18(6):416–422. https://doi.org/10.3758/BF03204114

23. Robinson EA (2022) Human Perception of Exponentially Increasing Data Displayed on a Log Scale Evaluated Through Experimental Graphics Tasks.

24. Robinson EA, Howard R, VanderPlas S (2023) "You Draw It": Implementation of Visually Fitted Trends with R2d3. *Journal of Data Science*, 21(2):281–294. https://doi.org/10.6339/22-JDS1083

25. Robinson EA, Howard R, VanderPlas S (2022) Eye Fitting Straight Lines in the Modern Era. *Journal of Computational and Graphical Statistics*, 0(0):1–8. https://doi.org/10.1080/10618600.2022.2140668

26. Murdoch D, Adler D (2023) Rgl: 3D visualization using OpenGL.

27. Marius Kintel (2023) OpenSCAD documentation. OpenSCAD. https://openscad.org/documentation.html

28. VanderPlas S, Hofmann H (2015) Signs of the sine illusion—why we need to care. *Journal of Computational and Graphical Statistics*, 24(4):1170–1190. https://doi.org/10.1080/10618600.2014.951547

29. Hofmann H, Vendettuoli M (2013) Common angle plots as perception-true visualizations of categorical associations. *IEEE Transactions on Visualization & Computer Graphics*, (12):2297–2305. https://doi.org/10.1109/TVCG.2013.140