

# Project Description

## 1 Overview

Statistical graphics and models are powerful tools to summarize data and 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. Chart design guidelines are often based on opinion, not empirical study, rendering many scientific communications sub-optimal or ineffective. This is alarming: effective science communication is critical for cultivating public trust in the scientific process and ensuring that decision makers accurately interpret supporting information. Addressing these challenges, my long-term career goal is to examine statistical graphics with the goal of *helping people use data more effectively*, and to apply this research to educate and inspire a new generation of scientists while supporting science literacy among the general public.

This CAREER proposal addresses a fundamental research question underpinning this problem: *How do design decisions impact the use, design, and perception of data visualizations?* Three research objectives support this goal:

- **RO1:** Create a framework for comprehensive graphical testing across multiple levels of user engagement.
- **RO2:** Assess the impact of measurement methods on experiments evaluating statistical graphics.
- **RO3:** Empirically validate common chart design guidelines, measuring the impact of design decisions on task performance.

Integrated with these research efforts, the overall **education goal** is to leverage visualization research to motivate statistical learning and improve data-driven 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 graduate course modules for K-12 educators that connect ongoing research to engaging, hands-on classroom activities for teaching statistics, math, and science.
- **EO3:** Improve the penetration of visualization research beyond academia by incorporating summaries of empirical studies in resources used by data scientists, industry analysts, and researchers in STEM disciplines.

Experiential learning activities will connect graphics research to critical concepts within statistics courses at the undergraduate level as well as in K-12 activities provided during graduate coursework for STEM educators. In addition, incorporating research summaries into general visualization resources will not only connect data visualization creators with research; improving these resources will improve teaching materials for statistical computing and will involve undergraduates in research and outreach in graphics and science communication.

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 Intellectual Merit

This work will expand our understanding of graphical perception and communication by empirically and systematically examining chart design through comprehensive, task-based testing. The proposed studies will be used to generate a framework relating evaluation methods to user engagement with graphics, establish the impact of different experimental design decisions on results, and promote integration of multiple evaluation methods to provide a holistic assessment of visualization effectiveness. Additionally, this project will prioritize inclusion neurodiverse and disabled individuals, ensuring that design guidelines account for accessibility concerns. The results of the systematic examination of different experimental design and testing methods will not only ground design guidelines in empirical results; if successful, the experiments will also help reconcile the results from historical studies with conflicting results. While there are task-based taxonomies for *selection of chart types*, a systematic framework for selecting *testing methods* based on levels of engagement and critical tasks is innovative; we expect that this framework will facilitate well-rounded experiments that examine chart design and use from multiple perspectives, providing nuanced results focused on audience use of graphics. The education activities proposed in this project are closely tied to the research objectives, providing avenues for dissemination of research results as well as inclusion of audiences in graphics research. As a result, education and research activities will combine to support new pedagogical research in experiential learning. This new research will examine the use of statistical graphics as an entry-point to quantitative subjects for individuals who are not traditionally interested in pursuing STEM careers. Previous collaborative research projects have established new and re-imagined old methods for testing statistical graphics; when combined with training and experience in statistics at the intersection of computer science, psychology, and communication, I am well equipped to complete this project supported by collaborations with researchers in cognitive psychology and statistical education.

My long-term career goal is to examine statistical graphics with the goal of *helping people use data more effectively*, and to apply this research to educate and inspire a new generation of scientists while supporting science literacy among the general public.

## 3 Research Plan

### 3.1 Background

**3.1.1 Overview** Scientific graphics transform quantitative data into image representations that can make use of the human visual system, leveraging our ability to take in and process huge quantities of information with minimal cognitive effort. However, unlike many mathematical data transformations, the transformation to visual space incurs loss both in the rendering of data to image and the transition from image to cognitive representation. That is, when creating data visualizations, we have to be concerned not only with the accuracy of the rendered image, but also with how that image is perceived by the viewer. It is easy to find entire books filled with situations in which the transition from data to image produces results which are misleading [1–3]; identifying scenarios where the transition from image to cognitive representation is suboptimal is more challenging and requires user studies. There have been empirical studies of graphics for at least 100 years [4–6], but the foundational work in graphical perception is [7], which established viewer’s ability to accurately estimate information from simple visual displays. While this work is important, and valuable, it has been synthesized into recommendations and rankings which go far beyond the original experiments [8, 9] with limited empirical verification, though in many cases these extrapolations are based in part on cognitive and perceptual research that is not specific to scientific visualization. In addition, it is easy to forget that [7] examined charts with respect to the direct numerical accuracy of quantitative estimates; the results do not necessarily apply if we are interested instead in determining whether differences between quantities can be perceived, ordered, remembered [10], or used to reach a reasonable real-world decision. While each of

these alternate tasks has been addressed in user studies of graphics, it is extremely difficult to synthesize the total graphics literature across many different disciplines (including psychology, computer science, statistics, design, and communication) in order to derive empirically driven guidelines for creating graphs which not only accurately transform the data into an image but also present the data in a form which can be effectively used by the intended audience.

The research objectives proposed here are designed to lay a foundation for more consistent evaluation and testing of scientific visualizations. We focus on the level of user engagement with the chart: while there are hierarchies and taxonomies of graphical tasks that break down intended to help **designers** select the best graphical representation, we focus instead on how the ultimate target of the graph is intended to interact with the image and the information presented within. XXX this is not yet quite clear as to what the difference is between this and e.g. task by taxonomy carswell 1992 XXX

We will first identify and evaluate methods for graphical testing across multiple levels of user engagement, comparing methods which examine equivalent stages of graph comprehension and use. Next, we will establish the impact of different experimental configurations and ways of measuring and recording users' answers, with the goal of establishing the impact of different measurement methods on the underlying experiment results. We expect that this will not only help graphics researchers design and implement new user studies, but will also facilitate comparison of results from past studies, hopefully allowing us to explain conflicting results. Finally, we will empirically validate common chart design guidelines, testing whether extrapolated results and aesthetic opinions hold up under critical user studies.

**3.1.2 Review of Relevant Literature** 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 [4–6]. Different individuals and organizations have released guidelines for graphical design [11–15], but many of these are based on aesthetic preferences, individual experience, or extrapolation from limited empirical studies; for an overview, see [16]. Even when there are empirical studies to support guidelines for choosing one graphical form over another [7, 17, 18, see 19 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 [20], and studies can address charts from multiple perspectives, including perception, task-based interaction (with different tasks producing different results), structural characteristics, and more [21]. 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 [22–24]; 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 [25] 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 [26]; it is reasonable to expect that the specifics of the graphical testing procedure used may impact the results, potentially explaining conflicting empirically based recommendations [4, 5, 18]. 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.

### 3.1.3 Preliminary Studies

### 3.2 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 [27, 28]. 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 [29, additional publications in progress], with additional publications discussing the implementation [30] and validating the new method [31] 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 participants tended to make serious inferential errors with regards to interpreting the language of the questions: questions asking for additive differences were often answered with multiplicative estimates, and vice versa, indicating that participants had trouble both with understanding the questions and visually estimating the answers. 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. This COVID-era project has directly inspired the comparisons across different testing methods proposed here: we experimentally validated performance across a number of levels of engagement with similar data, varying the y-axis scale transformation. It would be extremely helpful to systematically examine the impact of experimental testing decisions, such as estimating the ratio ( $A/B$ ) vs. the proportion ( $A/(A+B)$ ), data collection methods (numeric estimation vs. slider with tick marks vs. slider without tick marks), and task-based testing methods: this would allow us to compare the impact of these decisions, selecting the most accurate testing method for the experimental question and providing some context that may allow us to reconcile the results of conflicting historical studies.

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, [7] 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?

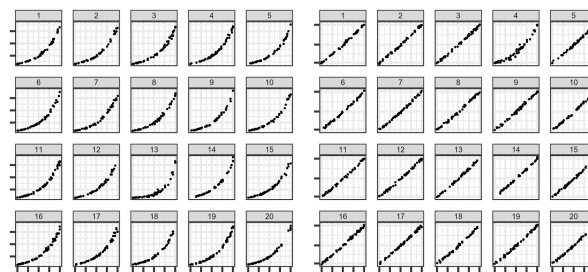


Figure 1: Lineups (log and linear scale) test detection of differences in exponential growth rates.

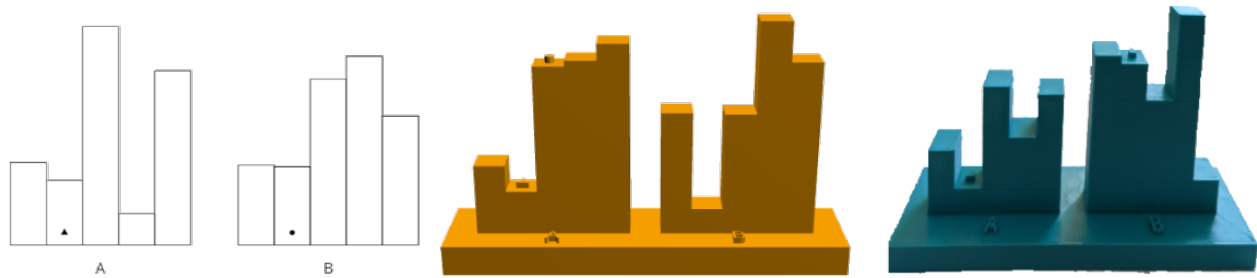


Figure 2: Three renderings of bar charts: 2D (left), 3D digital render (middle), and 3D printed (right).

A graduate student and I have been investigating this question in two different populations: a population comparable to that used in [7] (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 [32], and 3D-printed graphics [33], 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 [7] 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 [34, 35], 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.

### 3.3 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.

To address this broad goal, we will first 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, providing a more nuanced, task-focused, empirical evaluation of these design criteria to provide clarity in the importance and impact of design guidelines; if successful this will improve design criteria guidelines, but will also allow us to reconcile the results from conflicting historical experiments. 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.

## **4 Education Plan**

### **4.1 Overview**

### **4.2 Design and methods**

### **4.3 Evaluation**

### **4.4 Integration of Research and Education**

## **5 Timeline**

## **6 Broader Impacts**

## **7 Results from Prior NSF Support**

## References

1. Cairo A (2016) The truthful art: Data, charts, and maps for communication.
2. Cairo A (2019) How charts lie: Getting smarter about visual information.
3. Huff D (1954) How to lie with statistics.
4. 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>
5. 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>
6. 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>
7. 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>
8. Mackinlay J (1986) Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics*, 5(2):110–141. <https://doi.org/10.1145/22949.22950>
9. Franconeri SL, Padilla LM, Shah P, Zacks JM, Hullman J (2021) The science of visual data communication: What works. *Psychological Science in the Public Interest*, 22(3):110–161. <https://doi.org/10.1177/15291006211051956>
10. Borkin MA, Bylinskii Z, Kim NW, Bainbridge CM, Yeh CS, Borkin D, Pfister H, Oliva A (2016) Beyond memorability: Visualization recognition and recall. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):519–528. <https://doi.org/10.1109/TVCG.2015.2467732>
11. (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>
12. 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>
13. Kosslyn SM (2006) Graph design for the eye and mind.
14. Tufte E (2001) The visual display of quantitative information.
15. 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>
16. 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>
17. 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>
18. 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>
19. 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>
20. 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](https://doi.org/10.1007/978-3-030-34444-3_7)
21. 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](https://doi.org/10.1007/978-3-030-34444-3_3)
22. 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

23. 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.
24. 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>
25. 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>
26. 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>
27. Timmers H, Wagenaar WA (1977) Inverse statistics and misperception of exponential growth. *Perception & Psychophysics*, 21(6):558–562. <https://doi.org/10.3758/bf03198737>
28. Wagenaar WA, Sagaria SD (1975) Misperception of exponential growth. *Perception & Psychophysics*, 18(6):416–422. <https://doi.org/10.3758/BF03204114>
29. Robinson EA (2022) Human Perception of Exponentially Increasing Data Displayed on a Log Scale Evaluated Through Experimental Graphics Tasks.
30. 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>
31. 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>
32. Murdoch D, Adler D (2023) Rgl: 3D visualization using OpenGL.
33. Marius Kintel (2023) OpenSCAD documentation. OpenSCAD. <https://openscad.org/documentation.html>
34. 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>
35. 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>