# 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.

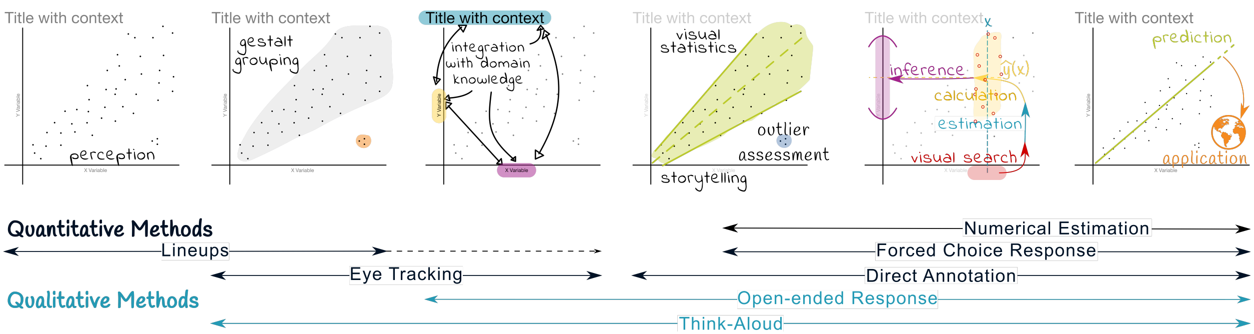


Figure 1: Levels of cognitive engagement with charts, roughly ordered by complexity, time, and effort. Methods which effectively measure (or could be extended to measure) each stage are shown below the charts. Text annotations show examples of the types of operations which involved in each stage.

The research objectives proposed here are designed to lay a foundation for more consistent evaluation and testing of scientific visualizations. We focus on the integrated cognitive complexity and temporal evolution of user-chart interaction, which is roughly illustrated in Figure 1. Previous hierarchies have focused on the complexity of single graphical tasks [7, 11, 12]; while this is a useful way to determine which chart to use to display data, it does not approach different ways users engage with a single chart: are they perceiving the graphical forms without engaging with the underlying symbolic meaning? Using the chart to understand the underlying natural phenomenon? Doing statistical inference (e.g. visually estimating parameter values from the graph)? Making decisions based on their understanding of the data? Each of these use cases involves different cognitive tasks, and as a result, different graphical testing methods must be used to assess the effectiveness of charts under each type of engagement.

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. @fig-cognition-hierarchy shows some of the methods we intend to assess and compare, along with the rough stages of cognition these methods target. Next, we will establish the impact of different experimental configurations and ways of measuring and recording users' answers. We expect that this will not only help graphics researchers design and implement new user studies, but we hope to also facilitate comparison of results from past studies, providing context to conflicting conclusions. Finally, we will empirically validate common chart design guidelines, testing whether extrapolated results and aesthetic opinions hold up under critical user studies.

The results from these objectives, taken together, are intended to build a user-focused foundation for measuring and assessing the design and use of data visualizations. The choice to approach testing graphics from the perspective of how the user is interacts with and makes decisions based on the visual representation of the data places this research firmly at the intersection of statistics, cognitive science, measurement, and communication. While previous researchers [7, 11, 13] have assessed graphics from the perspective of different estimation or user interaction tasks, the present project is focused on **measurement** methods for different stages of user interaction with graphics. Thus, this project will develop methodology for measuring the functional cognition underlying data driven decision making using visual aids. The results from the proposed research will also allow integration of conflicting historical results, hopefully leading to a robust set of empirical evidence that can be integrated to produce more robust, task-focused design guidelines for statistical graphics.

**3.1.2 Review of Relevant Literature** The first studies experimentally examining the effectiveness of statistical graphics took place approximately 100 years ago; since then, the quantity of charts created, the methods available for creating charts, and the technology available for measuring and evaluating comprehension have evolved in remarkable ways. [14] provides a comprehensive review of studies that experimentally examine the use of statistical graphics as well as the underlying research in cognitive psychology topics such as perception, memory, attention, and executive function which influence our ability to use statistical graphics effectively.

What is remarkable given the ubiquity of statistical graphics in scientific communication is that even after 100 years of empirical graphics research, we still have relatively little empirical evidence to support of some common design guidelines and heuristics; where there are empirical studies, they often conflict or have been over-extrapolated from the design and goal of the original experiments. For example, Tufte's data-ink ratio [15] has been thoroughly tested [11, 12, 16–18], but results have been decidedly mixed, suggesting that the data-ink ratio is too simplistic; even so, it is still part of the common vernacular and makes its way into many different design guidelines [19]. Another common recommendation is to locate the most important variables along position axes (e.g. $x$ and $y$ in a scatterplot) rather than encoding quantitative information in color; this is because [7] found higher levels of accuracy in these comparisons, but accuracy of numerical estimation is not the only important way people use charts [20], and in fact, it is relatively uncommon for individuals to directly estimate one specific numerical quantity from a chart: for these tasks, a table would be much more appropriate [21]. At a fundamental level, we know that graphics are useful for communicating scientific results and for exploring our data; whether the target audience is ourselves, peers, or the general public, graphics are an invaluable tool. So why do we assess graphics based on things like estimation accuracy, and then extrapolate the results to tasks and situations that don't revolve around estimation accuracy? A better alternative to the patchwork testing of individual questions with highly specific methods might be to try to assemble a functional picture of how specific charts (or design decisions) hold up under different levels of use, and then create guidelines which are more nuanced but can be applied more widely than current mantras.

With this goal in mind, there are several factors that must be considered and evaluated: the measurement methods and variables used to assess charts are of obvious interest, but other factors are also important. Measurement of numerical information that has passed through the human brain in one form or another can be complicated by the method used to record the information. Let us consider the relatively simple case where a participant is asked to estimate the length of a specified bar in a bar chart: the experimenter must determine how this estimate is recorded. Modern web design (assuming our experiment is conducted online or at least that data is entered via a computer interface) provides multiple options: the user can directly enter a number in a text box or indicate the number on a slider (with or without anchor points); the former requires translation into an explicitly numerical domain, where the latter requires that the participant map the chart onto a spatial domain but does not necessarily require explicit formation of a numerical estimate. Direct entry is subject to rounding effects that increase with participant uncertainty [22, 23]; while these effects can be mitigated [24] through modeling, it might be preferable to make use of numerical inputs that might not trigger rounding, such as slider inputs. Unfortunately, slider inputs are not entirely simple either: they can contain anchor points (or not) that participants may latch on to; the inclusion of these additional annotations may reduce cognitive load, but may provide the opportunity for additional anchoring effects that must be considered and possibly modeled. Most research in this area has examined sliders as inputs for categorical variables[25–29] and suggests that using sliders instead of radio button inputs changes the observed distribution of responses in important ways; while the comparison to radio buttons is not relevant to continuous data, the results of these studies suggest that there is a need to explicitly examine the effects of input methods on participant responses. This is just one example of the series of decisions experimenters make about the process of elucidating and recording data from participants which do not directly relate to the hypotheses under investigation but that may well impact the results. Even numerical estimation (regardless of input method) can be tricky: [12] asks participants for the ratio between the part and the whole (e.g. estimating the ratio $A/B$ where $A + B = 1$), while [7] asked participants to directly estimate the value of $A$; it seems possible that the different conclusions regarding the accuracy of pie charts might in fact be a result of the quantity participants were asked to investigate.

Specific methods for testing experimental graphics relevant to this project will be assessed as part of the methods section below, as will relevant preliminary studies that contributed to the development of this project.

### 3.2 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. Carswell CM (1992) Choosing specifiers: An evaluation of the basic tasks model of graphical perception. *Human factors*, 34(5):535–554. https://doi.org/10.1177/001872089203400503
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. Shneiderman B (1996) The eyes have it: A task by data type taxonomy for information visualizations. *Proceedings 1996 IEEE Symposium on Visual Languages*, :336–343. https://doi.org/10.1109/VL.1996.545307
14. 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
15. Tufte E (2001) The visual display of quantitative information.
16. Kelly JD (1988) The Data-Ink Ratio and Accuracy of Information Derived from Newspaper Graphs: An Experimental Test of the Theory. *Visual Communication Division of the Association for Education in Journalism and Mass Communication*,
17. Gillan DJ, Richman EH (1994) Minimalism and the Syntax of Graphs. *Human Factors*, 36(4):619–644. https://doi.org/10.1177/001872089403600405
18. Gillan D, Sorensen D (2009) Minimalism and the Syntax of Graphs: II. Effects of Graph Backgrounds on Visual Search. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, 53:1096–1100. https://doi.org/10.1518/107118109X12524443344998
19. Ajani K, Lee E, Xiong C, Knaflic CN, Kemper W, Franconeri S (2022) Declutter and Focus: Empirically Evaluating Design Guidelines for Effective Data Communication. *IEEE Transactions on Visualization and Computer Graphics*, 28(10):3351–3364. https://doi.org/10.1109/TVCG.2021.3068337
20. Bertini E, Correll M, Franconeri S (2020) Why Shouldn't All Charts Be Scatter Plots? Beyond Precision-Driven Visualizations. *2020 IEEE Visualization Conference (VIS)*, :206–210. https://doi.org/10.1109/VIS47514.2020.00048
21. Gelman A, Wainer H, Briggs WM, Friendly M, Kwan E, Wills G (2011) Why Tables Are Really Much

Better Than Graphs [with Comments and Rejoinder]. *Journal of Computational and Graphical Statistics*, 20(1):3–40. https://doi.org/10.1198/jcgs.2011.09166
22. Ruud PA, Schunk D, Winter JK (2014) Uncertainty causes rounding: An experimental study. *Experimental Economics*, 17(3):391–413. https://doi.org/10.1007/s10683-013-9374-8
23. 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
24. Wang B, Wertelecki W (2013) Density estimation for data with rounding errors. *Computational Statistics & Data Analysis*, 65:4–12. https://doi.org/10.1016/j.csda.2012.02.016
25. Thomas M, Kyung EJ (2019) Slider Scale or Text Box: How Response Format Shapes Responses. *Journal of Consumer Research*, 45(6):1274–1293. https://doi.org/10.1093/jcr/ucy057
26. Liu M, Conrad FG (2019) Where Should I Start? On Default Values for Slider Questions in Web Surveys. *Social Science Computer Review*, 37(2):248–269. https://doi.org/10.1177/0894439318755336
27. Funke F (2016) A Web Experiment Showing Negative Effects of Slider Scales Compared to Visual Analogue Scales and Radio Button Scales. *Social Science Computer Review*, 34(2):244–254. https://doi.org/10.1177/0894439315575477
28. DeCastellarnau A (2018) A classification of response scale characteristics that affect data quality: A literature review. *Quality & Quantity*, 52(4):1523–1559. https://doi.org/10.1007/s11135-017-0533-4
29. Couper MP, Tourangeau R, Conrad FG, Singer E (2006) Evaluating the Effectiveness of Visual Analog Scales: A Web Experiment. *Social Science Computer Review*, 24(2):227–245. https://doi.org/10.1177/0894439305281503