

Susan VanderPlas

Statement of Research Interests

802 17th St.

Auburn, NE 68305

☎ (515) 509-6613

✉ svanderplas@gmail.com

📄 www.srvanderplas.github.io

🌐 [srvanderplas](https://www.srvanderplas.github.io)

Statistical graphics are important during every stage of the statistical process: exploratory analysis, examining model assumptions, and presenting statistical results. Graphics serve as a form of guided external cognition: by summarizing the data in an effective visual format, they free up cognitive resources, allowing the viewer to consider the implications of the data, rather than trying to organize the data internally. As graphics serve as cognitive aids, it is important that they are carefully designed to display important features of the data in a way that facilitates visual inference. I study this interplay between statistical graphics and the brain, seeking to understand how the brain processes statistical graphics and to apply that knowledge to create intuitive, effective, and interactive visual representations of complex data. This research is highly interdisciplinary, integrating work from statistics, cognitive psychology, education, neuroscience, and human-computer interaction in order to communicate statistical information more effectively through graphics.

Perception of Statistical Graphics My dissertation examines the impact of several facets of human perception of statistical graphics: optical illusions, visual aptitude, and the hierarchy of features in graph perception, arguing that it is important to consider human perception in the design and evaluation of statistical graphics.

In the first chapter, I explore the sine illusion, which affects the graphical perception of data along nonlinear trend lines, in two separate user studies. In the first study, I derive two transformations which break the illusion, and evaluate those transformations with user studies, establishing both the presence of the illusion and the efficacy of the transformations. In the second study, I explore the psychological underpinnings of the illusion in order to establish that the illusion is pervasive and occurs as a result of binocular depth perception (with evidence from a case study of an individual lacking that ability); I also examine the distribution of the distortion in the general population with another user study. The sine illusion research highlights a neglected issue in statistical graphics: how do we ensure that information is perceived by the brain as it is presented graphically? This work has been well received: the first study has been submitted to and accepted by JCGS; the second study was selected for the Graphics section student paper award at JSM in 2014, and an expanded paper is in preparation for publication. The papers are largely designed to raise awareness of the illusion and its effect on a wide array of oft-used graphics, such as scatter plots, time-series charts, hammock plots, stream graphs, ribbon charts, and candlestick plots.

The remaining chapters are more focused on visual inference, exploring the use of plots to conduct formal graphical inference; this approach provides a framework for determining the “graphical power” of different types of charts. In the second chapter, I explore the interaction between tests of spatial ability and visual inference using statistical lineups (sets of 20 plots, 19 generated by permutation and one composed of real data); the results suggest that lineups are ultimately a classification task,

though the accuracy of visual inference is also a function of the type of chart, overall visual ability, and demographic factors such as STEM training, research experience, age, and gender. This study is currently in preparation for publication.

In the final chapter, I examine the hierarchy of visual features in statistical graphics in the context of the lineup protocol; that is, which graphical features are most visually salient - are we more sensitive to outliers or skew in q-q plots? While this experiment is still in the design and execution stage, I expect the results to reinforce the necessity of carefully highlighting important components of statistical graphics using aesthetics such as color, shape, and size, freeing cognitive resources for higher-level tasks.

Interactive Data Visualization In addition to my research on the psychological underpinnings of statistical graphics, I also collaborate with researchers in other disciplines to create effective graphics for large data sets or multidimensional data. I have contributed to the development of the `animint` package for R, which extends the grammar of graphics (as implemented in `ggplot2`) to create interactive, animated graphics for the web. I have also worked with soybean researchers at the USDA to analyze and visualize soybean genetic data at the population and individual level; the result is a series of topic-specific interactive applets (using `animint` and RStudio's `Shiny` package) which integrate the data and analysis for use by biologists. Working with co-authors, I have also created interactive applets for use in the statistics classroom; these applets were designed to intuitively demonstrate difficult statistical concepts.

Future Research As data sets grow larger, interactive visualizations become more critical because user interaction can be used to create more complex and nuanced visualizations that display slices of a data set in sequence. With this increasing complexity, it becomes even more critical to consider the perceptual system in visualization design. I would like to investigate the relationship between the user experience and different types of interactivity, utilizing tools such as eye tracking, mouse tracking, and open-ended questions about the data; these tools provide insight into both the perceptual process and the information communicated by the interactive visualization. This research would build on existing literature examining static graphics, but would also need to incorporate sensory integration research, as the motion of transitions between different states of interactive graphics triggers specific perceptual organization schemes. In parallel with this research, I would also like to explore methods for reducing the complexity of interactive visualizations (including binning techniques that are useful for static plots); this research is much more directly applicable to data visualization in the field, as current interactive plots typically rely on JavaScript and related libraries, which render each graphical object separately (and consequently do not scale well to "big data").

Statisticians can assemble huge databases of information, but in order to communicate findings (particularly outside of the field), one of our most effective tools is still a well-designed chart. Understanding the interaction between human perception and statistical graphics provides an essential foundation for optimizing graphics to more effectively communicate with each other and with those outside of the field.