Susan VanderPlas

Statement of Research Interests

802 17th St.
Auburn, NE 68305
✆ *(515) 509-6613*
✉ *srvanderplas@gmail.com*
✐ *www.srvanderplas.github.io*
❂ *srvanderplas*

Statistical graphics are useful at every stage of the statistical process; during exploratory analysis, for examining model assumptions, and most importantly, for 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 the data itself. It is important that graphics are carefully designed, displaying important features of the data in a way that facilitates visual inference. I am interested in this interface between statistical graphics and the brain, and I seek to understand how the brain processes statistical graphics and to apply that knowledge to create intuitive, effective, and interactive visual representations of complex, multidimensional, and large datasets. This research is highly interdisciplinary, integrating research from statistics, cognitive psychology, education, neuroscience, and human-computer interaction to communicate information more effectively.

Perception of Statistical Graphics My dissertation examines the impact of several perceptual heuristics as they relate to statistical graphics. I begin by exploring the sine illusion (or line-width illusion), which commonly appears in data with nonlinear trends and affects the perception of the variance/mean relationship. In the first study, I propose 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 a second study of the same illusion, I explore the psychological underpinnings of the illusion in order to establish that the illusion is persistent and occurs due to the wiring in the brain (with a supporting case study) and examine the size of the distortion and the variation in the general population with another user study. 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 experimental results suggest that lineups are ultimately a classification task and that accuracy of visual inference depends on demographic factors such as STEM training or research experience, completion of calculus 1, gender, and age, as well as underlying visual ability. In the final chapter, I explore the hierarchy of visual features in statistical graphics; that is, which graphical features are most visually salient (for example, do viewers notice outliers or cluster separation?) in the context of the lineup protocol. This work has implications for the design of new types of graphics, as graphics which highlight important features of the data intuitively free cognitive resources for higher-level tasks.

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? The first study has been accepted to JCGS; the second study was selected for the Graphics section student paper award at JSM in 2014; an expanded paper will be submitted to ACM Transactions on Applied Perception later

this year. These papers are intended to raise awareness of the illusion and its effect on a wide array of oft-used graphics, such as scatterplots, time-series charts, hammock plots, stream graphs, ribbon charts, and candlestick plots. The other studies are more focused on visual inference; that is, exploring the use of plots to conduct formal statistical inference and examining the power of different types of statistical charts for making visual inference. Understanding how demographic characteristics and visuospatial ability impact the perception of statistical graphics (and understanding which features of a graph are perceived most strongly) is important for statisticians, educators, and scientists, who may wish to tailor graphics to emphasize specific features of the data.

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 datasets 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 dataset 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 amass huge databases of information, but in order to communicate findings (particularly outside of the field), our most effective tool is still a well-designed chart (in conjunction with one or more models). Understanding the interaction between human perception and statistical graphics an essential foundation for communicating with each other and with those outside of the field.