What Do You See? Perception, Algorithms, and Statistical Graphics

Susan Vanderplas

# Overview

Statistical graphics and other scientific images are incredibly powerful and efficient ways to convey numerical information in a way that humans can use for decision making. However, research on visual perception of statistical graphics and other data visualizations is a patchwork of inconsistent methods, and most recommendations are based primarily on experience rather than empirical study.

Over the course of my career, I want to leverage my research in statistical graphics and data visualization to help humans and algorithms work together more efficiently and effectively. This involves teaching my students to enjoy working with data, to program with data in mind, to give back to the community by publishing open-source tools for reproducible science, and to think about the audience when they create statistical analyses and visualizations. I will bring these focuses to my research as well, conducting experiments which are designed to understand how humans read charts, interact with data, and draw conclusions from statistical graphics, and using this newfound understanding of human perception to enrich statistical modeling and machine learning algorithms by engineering features designed to mimic human perception in a repeatable and reproducible way.

My research goal is to understand how we use and perceive charts, and how the design of statistical graphics impacts our ability to use them successfully. To address this goal, I will (1) develop multimodal methods for testing statistical graphics in order to examine how the same chart might be used in different judgment tasks, such as comparison, prediction, and estimation, and (2) apply these methods to assess the decisions scientists make when creating statistical graphics for exploration or communication of results. I will apply these findings to (3) develop new statistical features that mimic human perception in tasks which are currently conducted through human visual comparison, such as forensic pattern evidence, in order to develop statistically validated, empirical tools to assist scientists with decision-making in a reproducible, auditable way.

My teaching goal is to develop tools and curricula to support statistical learning and scientific decision making in undergraduate, graduate, and legal contexts. To address this goal, I will (1) develop experiential learning activities for introductory statistics courses that involve participating in research in statistical graphics, (2) integrate principles of scientific reproducibility and open-source software into undergraduate and graduate statistical programming courses, and (3) work with forensic scientists, lawyers, and judges to discuss scientific and statistical validity, assess the support for different forensic evidence evaluation techniques, and promote the importance of open, reproducible science in forensics and the legal system.

# Intellectual Merit

# Broader Impacts