Susan Vanderplas
Statistics Department
susan.vanderplas@unl.edu

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

Statistical graphics are powerful and efficient tools to convey numerical information for human decision making; however, research on graphical perception is a patchwork of inconsistent methods. Most recommendations are based on opinion rather than empirical study, rendering many scientific communications sub-optimal or ineffective. This is alarming, as effective science communication is critical for cultivating public trust in the scientific process and ensuring that decision makers accurately interpret information when making choices which impact people's lives. Addressing these challenges, the PI's long-term career goal is to understand statistical graphics and data visualization, *helping humans and algorithms work together more efficiently and effectively*, and apply this research to educate a new generation of scientists and the broader public.

Toward this vision, the overall **research goal** of this CAREER program is to advance understanding of the use and perception of data displays and the impact of their design on humans' ability to use them successfully. Toward this goal, three research objectives (ROs) will be pursued. RO1: Examine effectiveness of graphics across different tasks, such as comparison, prediction, and estimation, by developing multi-modal methods for testing charts. RO2: Empirically validate guidelines for graphical creation, assessing the effect of design decisions on the utility of charts across different uses. RO3: Produce validated, empirical, quantitative tools for reproducible decision-making in forensic science by engineering statistical features that mimic human visual perception in when examining forensic evidence.

Integrated with these research efforts, the overall **education goal** is to cultivate statistical learning and scientific decision making in society. Three education objectives (EOs) will address this goal: EO1: Develop and implement experiential learning activities in graphics for undergraduate introductory statistics courses. EO2: Create and implement open educational resources (OER) to introduce reproducible science and open-source software development in statistical programming courses. EO3: Work with graduate students to engage with forensic scientists, lawyers, and judges, evaluating the scientific support for forensic disciplines and promoting the importance of open, reproducible forensic science.

Intellectual Merit This work will expand our understanding of applied perception and scientific communication; in particular, this research will include studies focused on expert, disabled, and neurodivergent subpopulations, increasing our understanding of accessible graphics in different target groups. Novel methods for graphical study will uncover new research directions in InfoVis and will produce additional tools for evaluating scientific communication. Research in forensic perception will advance our understanding of visual evidence comparisons, produce better error rate estimates, and approach explainable machine learning in a new way. Incorporating experiential learning and graphics research into intro statistics courses will advance knowledge of pedagogy and barriers to interest and engagement in STEM. Finally, assessing the OER curricula developed for reproducibility and open-source science will allow us to measure and explore the development of students' attitudes toward ethical science and responsible conduct of research.

Broader Impacts The statistical graphics research will produce empirical guidelines for better science communication. Applying our improved understanding of visual perception, we will develop algorithms for quantitative, reproducible evaluation of forensic pattern evidence. Engaging with the justice system will emphasize the failings of current subjective evaluation methods, motivate the importance of these algorithms, and facilitate conversations between graduate students in quantitative fields and judges, lawyers, law students, and forensic scientists. Experiential learning and research engagement will produce students who are more engaged in STEM coursework and less averse to STEM careers (while accelerating research productivity). Incorporating reproducibility and open-source software development into statistical coursework will foster a well-trained future generation of scientists who will contribute to the open-science community, emphasize the importance of reproducible and open science, and build new infrastructure while conducting better science.