UNIVERSITY OF Nebraska Lincoln

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 data to support human decision making; however, research on graphical perception is a patchwork of inconsistent methods and conflicting results. 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 examine statistical graphics with the goal of *helping humans and algorithms work together more efficiently and effectively*, and to 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 design the charts effectiveness. Toward this goal, three research objectives (ROs) will be pursued. RO1: Examine effectiveness of graphics across different tasks by developing methods for testing charts at different levels of user engagement. RO2: Empirically validate visualization design guidelines, measuring the impact of design decisions on task performance. RO3: Produce quantitative statistical models for reproducible decision support using images by engineering statistical features to mimic human perception.

Integrated with these research efforts, the overall **education goal** is to cultivate statistical learning and scientific 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 and implement open educational resources (OER) to introduce reproducible science and open-source software development in statistical programming courses. EO3: Engage with forensic scientists, lawyers, and judges, evaluating the scientific support for forensic disciplines and promoting open, reproducible scientific practices.

Intellectual Merit This work will expand our understanding of scientific communication with graphics by reconciling sparse and conflicting research and opinions with empirical testing to produce nuanced guidelines for chart design. New methods of empirical graphical testing will be developed to cover a wider range of engagement with charts. In addition, this work will specifically examine the effects of expertise and disability (visual and cognitive), increasing our understanding of accessibility and audience considerations in graphics. While at first glance forensic pattern evidence seems unrelated to graphical perception, both graphics and pattern evidence images are visual tools used for making decisions; studying how visual patterns are perceived and used for decision-making in graphics allows us to reverse-engineer the process to develop reproducible, quantitative features that capture visual patterns in evidence that is currently evaluated subjectively. 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 This research will produce nuanced, empirically driven guidelines to support better science communication, enhancing scientists' ability to disseminate their research to the general public through data visualization. In addition, this project leverages explainable machine learning (EML) to reduce cognitive demands of decision-making based on statistical models while building trust in statistical decision tools through perceptual analogies. One specific application of this approach to EML is the proposed work with forensic pattern evidence, which will emphasize the subjectivity and lack of reproducibility of current methods; the work will also require 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 to support future research.