Visualization and Data Analytics

Andrew Golightly, Pranav Rajan, Jared Amen
School of Computing
University of Utah

INTRODUCTION

From infographics that show proper handwashing technique to charts showing business trends, visualizations have the capability to condense complicated information into easily digestible graphics. Despite the extensive research into how people perceive and relate to these visualizations [4], research concerning the use of the tools that make them is sparse. Countless artifacts have been published [1, 2], but there are few, if any, studies available that evaluate the effectiveness of interpreting and creating visualizations from the perspective of a researcher and developer. Furthermore, existing studies focus on the relationship between the visualization tool and the developer, rather than the comparatively complex relationship between the visualization tool, developer, and the resulting visualizations. To that end, we pose the following research questions:

- RQ1: How do visualization and data wrangling tools help to create user-centric visualizations?
- RQ2: What qualifies a visualization as user-centric (as according to the guidelines established by Kelleher et. al. [3])?
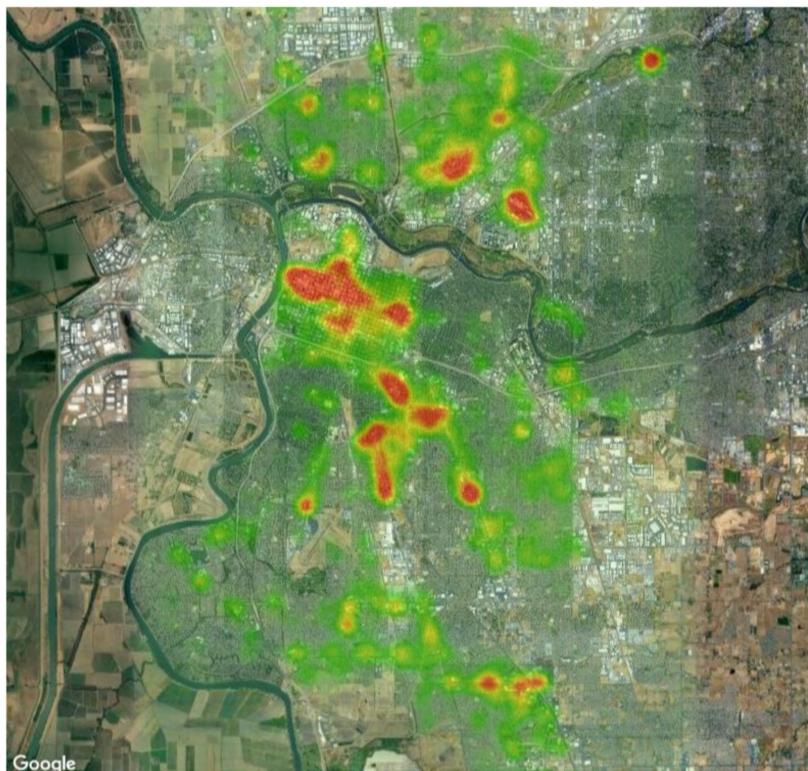

Figure 1: Highest ranked visualization for location task.

METHODS

- To answer RQ1, we conducted a small-scale interview of 6 data science experts. They were asked to describe their process when visualizing data, including tools and heuristics used.
- To answer RQ2, we conducted a large-scale Amazon Mechanical Turk (AMT) study with N=24 workers. They were asked to rank three visualizations for two different visualization tasks and give an explanation for their top choice.

RQ1:
- Python, as a language, and the Python Data Stack were the tools of choice.
- Ad-hoc visualization demanded more end-user-friendly tools.
- Participants were largely unaware of visualization-specific heuristics.

RQ2:
- Simple and conventional visualizations were more useful compared to novel or more complicated visualizations.
- Most participants cited ease of use and clear labeling as reasons for their top choice.
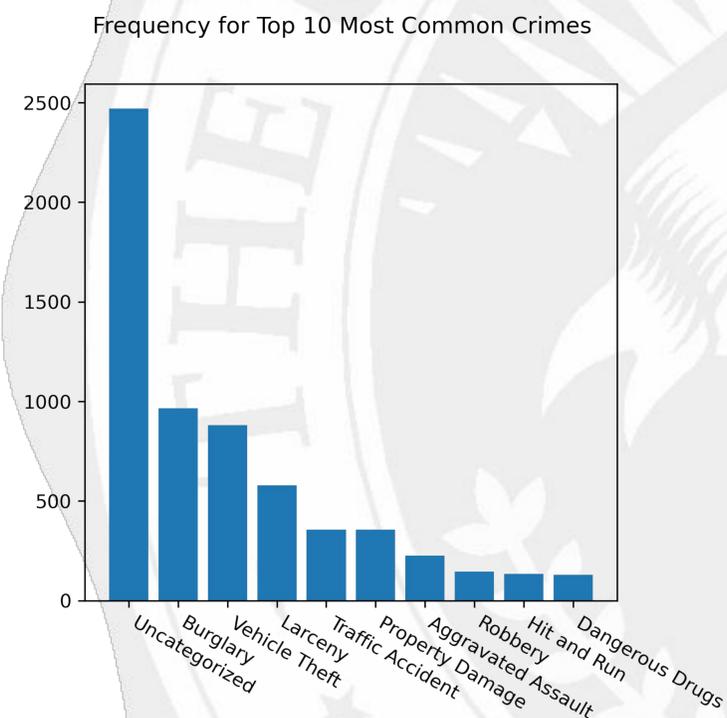- Participants overwhelmingly used their favorite visualization when asked to identify a feature of the data.

Figure 2: Highest ranked visualization for crime type task

RESULTS

- Visualization heuristics are well designed but not explicitly used. Experts tended to apply aspects of the heuristics innately.
- Experts preferred a small range of "Swiss Army Knife" tools for processing and visualizing data, with little to no consideration of other tools for tasks.
- Accessibility and focusing on chart interpretability improved user ability to perform data summarization.
- Users chose visualizations that best helped them perform data analytics tasks even if they were missing bells and whistles such as lots of color.
- Among similarly useful visualizations, color and visual appeal have an influence on the preferred visualization of a dataset.

RECOMMENDATIONS

- Visualization designers and developers should focus on developing visualizations that emphasize encodings and layouts that enable users to extrapolate important details for performing data analytic tasks.
- Tool developers should focus on developing tools that integrate well with one another and the different parts of the data visualization.
- Accessibility and data interpretability of visualizations should be considered when designing visualizations, rather than after visualizations are created.

REFERENCES

[1] Bostock, M., Ogievetsky, V., & Heer, J. (2011). D³ data-driven documents. *IEEE transactions on visualization and computer graphics*, *17*(12), 2301-2309.

[2] Kandel, S., Paepcke, A., Hellerstein, J., & Heer, J. (2011, May). Wrangler: Interactive visual specification of data transformation scripts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3363-3372).

[3] Kelleher, C., & Wagener, T. (2011). Ten guidelines for effective data visualization in scientific publications. *Environmental Modelling & Software*, *26*(6), 822-827.

[4] Peck, E. M., Ayuso, S. E., & El-Etr, O. (2019, May). Data is personal: Attitudes and perceptions of data visualization in rural pennsylvania. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-12).

[5] Wu, K., Petersen, E., Ahmad, T., Burlinson, D., Tanis, S., & Szafir, D. A. (2021, May). Understanding data accessibility for people with intellectual and developmental disabilities. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1-16).