Week 5 Journal

Aaron Schram

Evaluating Graphics

For comparing or evaluating graphics, it is easy to start with the papers by Elavsky *et al* (2022) and Burns *et al* (2020), since they use function accessibility POUR+CAF (Perceivable, Operable, Understandable, and Robust + Compromising, Assistive, and Flexible) and Bloom's Taxonomy for knowledge acquisition (Knowledge, Comprehension, Application, Analysis, Synthesis, and Evaluation), respectively. With Burns *et al* (2020) having a high focusing on trying to ask the correct set of questions for an observer to evaluate a graphic, while Elavsky *et al* (2022) focus on a wide array of accessibility criteria for a rater to evaluate a graphic.

Starting with the least interesting group, Burns *et al* (2020) and Bloom's Taxonomy can be naively described as just asking the participant a series of questions that are hierarchical or scaling in nature to help define the usefulness of the graphic. These questions are designed to ask the participants to answer questions about basic facts, contextualize information, apply these to a problem, understand the relationship, create something new, and judge the value of the graphic. This is the least interesting of the two groups since the authors are essentially quizzing the participant with the last questions being to judge the graphic. I get that they want to assess more of if the image is conveying the correct idea, but the method does not provide a robust in-depth guide for a single user to evaluate any graphic. Without a guide telling them or hinting at relationship for each question, since there are wrong answers answered in the article, the final question of judging the graphic is already biased by perceived incorrect information. At the best, the participant would answer neutrally as not knowing the true answers. Therefore, the only real use for a comprehensive question-based approach would be in cases where a group designs a graphic with an intent or goal in mind, and they then quiz participants to understand how close a given person understands the graphic made. Then, Bloom's Taxonomy would be useful.

Now for the more interesting of the two groups presented, Elavsky *et al* (2022) presents Chartibility using the function accessibility POUR+CAF to answer questions about the level of accessibility of graphics and graphical packages/functions. Chartibility is broken into five different components with the goal of testing the statistical graphics perceivability, operability, robustness, assitiveness, understandability, compromisability, and flexibility (Elavsky *et*

al, 2022). These are measured through visual testing, keyboard probing, screen reader inspecting, checking cognitive barriers, and evaluating context. Naively speaking, visual testing is associated with color, contrast, and text size, keyboard probing is just how easy is the visual to navigate, screen reader is just reading the text associated with each graphic, cognitive barriers is the clarity and understandability of the text, and context is kind of everything else.

Honestly, visual testing and cognitive barriers are the most important things going on here. We need our visuals to be color blind friendly, so that anyone can see the contrasts created, and we also need legible and viewable text to understand what is being observed with the graphics. These two things are so important since they explicitly define the graph being generated and viewed. Both of these applies to color being used on the graphics, legends created, data dictionaries presented, and summaries generated along with the graph. However, this primarily applies to people with dyslexia or color deficiencies.

My biggest critique is that Chartibility is supposed to be all about accessibility, but all of the accessibility for low vision or vision impaired people is primarily located to the visual reader category, and the evaluating context category. I personally think that there should be several distinct categories to access how user friendly each graphic is for various disabilities, such as color blindness, dyslexia, dyscalculia, low vision, blindness, etc. For instance, sonification, tactile, textual summaries, and semantic evaluations should not all fall into the final category. I would have like to see sonification and textile graphics in their own categories, so we would have something like visual testing, sonification testing, and textile testing, while cognitive barriers could have then been broken up into interpretability for each of these groups.