

Ziyi Li

Mentor: Michael Pearce

My DRP project was titled “History and Practice of Data Communication,” mentored by Michael Pearce. During the first three weeks, we focused on the history of the development of data communication techniques from medieval to modern times. While exploring these readings, I developed a better understanding of the intellectual history of data visualization. Specifically, I learned about what motivated each new development, their goals, and how they related to developments in other fields. We traced the process of how data communication techniques were developed from the 14th century to now. Moreover, I explored how to properly use various types of plots, such as pie charts, stacked bar charts, scatterplot matrices, bubble charts, etc. For instance, I learned how the perception of visualizations can be influenced by the creator’s choice of color, size, opacity, shape, density, axes, confidence intervals, labels, and many other characteristics.

After gaining a basic understanding of the traits and characteristics of different visualizations, I was able to apply what I learned to modern problems. During the next few weeks, we discussed visualizations about the 2020 US presidential election and Covid-19.

While reading articles about the 2020 US presidential election, we discussed how the media can mislead people and how difficult it is to present data without misleading. The most personally impactful article for me was Nate Silver’s “The Media Has A Probability Problem,” which mainly talked about errors of analysis and how the media often misinterprets and blames data. Journalists and readers tend only to believe what they want to believe. When voters were asked to guess the probability of Trump or Clinton winning the 2016 election, the average voter gave Clinton only a 53 percent chance of winning and gave Trump a 43 percent chance — so while respondents slightly favored Clinton, it was not with much confidence. In the article, Silver talked about how most journalists did not believe that someone like Trump could become president four years ago. Furthermore, Silver argued that while making predictions, the world needs less certainty. He concluded with the quote, “in journalism...a story can be 1. fast, 2. interesting and/or 3. true — two out of the three — but it’s hard for it to be all three at the same time.”

Furthermore, we looked at data visualization for Covid-19, such as UW’s Covid Case Tracking Dashboard, the Worldwide Dashboard, and US State-Level Dashboards. Specifically, we discussed how poor Covid-19 data visualizations can mislead the public. For example, cluttered numbers and missing context make charts easy for people to interpret data incorrectly. We critiqued some bad visualizations from the Texas State Covid-19 Dashboard, Vox, and other sites. My main takeaway from this was that people who create these visualizations may have good intentions to correctly communicate information, but they can still be interpreted differently by the audience.

As someone who creates visualizations, I learned that the only thing we can do is try our best to make sure what we create would not lead to any bias or misinterpretation. We need to make careful choices when visualizing data to fight against data misinterpretation.