Data Visualization

This quarter, I participated in research on data visualization with my mentor, Kayla Irish. With data visualization being a highly ubiquitous piece of statistics, and something that makes its way into every report, every writeup, and every presentation, we wanted to research how to make our data visualizations. Our approach was two-pronged: first we read a book that gave some general theory about effective data visualizations, then we worked to create data visualizations that were visually appealing and that followed this guidance. The book we read was *The Visual Display of Quantitative Information,* by Edward R. Tufte.

The first part of Tufte's book focuses on graphical excellence, which he defines as the things that make a data visualization effective in presenting data. He gives three general guidelines that should be followed. First, it should be remembered that the purpose of Data Visualization is to reveal data, not to invent data. Data visualizations have the benefit of being able to quickly share a good amount of data with the consumer. Our purpose in creating them is to ensure that the patterns that exist in the data are echoed in the graphic in a way that is readily visible. Second, data visualizations should present interesting data in a well-designed way. If we have data that is not terribly indicative of any pattern or otherwise depicts variables that are not related in any way that allows us to form a narrative, that is uninteresting data. The notion of being well-designed, however, is not terribly prescriptive. Tufte gives some general guidelines and rules of thumb but also states that creating a visualization is as much a statistical endeavor as it is an artistic one. We have creativity, and there are infinite effective ways to display quantitative information visually. Lastly, Tufte urges that a good visual is one which gives the audience "as many ideas as possible…with the least ink in the least time". So, visualizations should be clear in the patterns they depict, but should also not be too complex, lest the audience not be able to understand what is going on

The second part of Tufte's book focuses on graphical integrity, which refers to the common belief that graphics and charts are used to deceive and lie. Tufte argues that data visualization is not more susceptible to lies and deception than normal speech, but he does give some suggestions to ensure that the integrity of a visualization is preserved. First, he talks about the overinflation of graphic size. He defines an index called the Lie Factor, which he defines as the quotient between the visually depicted size of effect, and the measured size of effect. If for example, a visual showed a measured increase of 25%, but displayed it in such a way that it appeared twice the size, that would be a lie factor of 100%/25% = 4. We should strive for a Lie Factor as close to 1 as possible. Second, Tufte dispels the idea that if we seek more visually appealing, highly decorated visuals, that will inherently lead to a bit of overexaggeration. He remakes overexaggerating visuals, and shows that a cool design is no excuse for inflating the effect we desire the audience to see. Lastly, Tufte asserts the importance of context. When we use numbers out of context, such as presenting costs in nominal dollars instead of real dollars, we present a narrative with our design that may not actually be indicative of what is going on. Just as in regular statistical analysis, we must account for confounding variables.

Based on these ideals, I was able to create three visuals from a dataset of 20,000 games. I created a density plot of the number of moves to completion based on the result of the game, as well as two visualizations depicting the relative usage of each kind of piece, both across the different ranks, as well as across the flow of a game. I very much enjoyed this DRP experience and learned so much about data visualizations. I feel very inspired to take the opportunities I have to create data visualizations that are effective, sleek, and well-designed.