Week 2 Journal

Aaron Schram

Grammar of Graphics

My own experience with the scope of graphics extends from a Data Science Visualizations class from my undergraduate college. We discussed the "Gestalt" principles with a source that I cannot remember, but my take-away was that simpler graphs were better, and there existed a hierarchy of preferences to create graphs. I remembered this during my reading chapter 5 of *Visualization Analysis and Design* (Munzer et al., 2015). The main discussion of interest was the idea of visual channel rankings and psychophysics idea of "visual accuracy". To me, these fell in line with the "Gestalt" principles that I had already learned. It is easier for people to understand simpler representations of data on lower dimensional cartesian planes (smaller marks). By this, I mean we like to see things that are smaller in spatial dimensions (smaller marks), things that are in alignment with one another, and groupings by spatial region or color. To me this intuitively makes sense, as a Statisticians, I feel like the easiest ways to represent data is usually with a bar pot, histogram, or a box-and-whisker plot. These fall into these highly effective marks and visual channels described.

Although the concept of visualizations having a ranking or ordering is not new to me, the ideas of the seven orthogonal classes presented by the *Grammar of Graphics* (Wilkinson, 2010) were completely new to me. I would argue that the first two to three classes (variables, algebras, and scales) are the most irrelevant to my job as a statistical consultant, since I have very little say in the collection process of the data, the storing of the data, or the "true" scale of interest. So, my job starts at a pre-defined scale of interest, and then I build visualizations and create statistics of interest to the client. Honestly, I feel the same is true about the abstraction and entity relationship schema from *Visualisation Design as Language Transformations - From Conceptual Models to Graphics Grammars* (Filonik, 2019). These are ways of storing and representing data that I am not accustomed to, since I get pre-processed data from clients, and most statistical software requires a simple schema to be processed. I have seen clients formulate extended schema in Excel as well as other more complicated structures.

As for myself, regarding the use of the Grammar of Graphics in my own work, I would use default plots in R and lsmeans plots in SAS. These would range from statistics to aesthetics. Although, the coordinates class could be pure ignored since everything is plotted on 2-dimensional cartesian plane for most of my college time. Therefore, my focus has typically

been on geometry with bar plots and line plots and aesthetics with positioning and color. From a visualization analysis perspective, my graphics generated are typically 1-dimensional marks, horizontal positioning, an aligned common scale, and sometime color to denote distinct groups. While this is not a terrible thing, the graphs I typically make in R or SAS are typically lack luster. This is because I'm typically doing ANOVA type tests for clients and sometimes looking at distributional assumptions or interactions. I could use a further incorporation of visual channels and aesthetics. This would be especially useful in the case of higher dimensional regression problems or machine learning problems.

Overall, my understanding of the Grammar of Graphics is more rooted in the idea that simpler is better and that the graph should explain itself. It might be due to my own lack of creativity or through the lens of lower dimensional problems, but I have not paid much heed to the actual hierarchy of marks and visual channels. I typically try to do the most basic things and move on. My work could use some serious improvement with incorporation of channels based on their effectiveness and looking at the data through the lens of effectiveness principle to better express the data. As far as grouping or discrimination and grouping of the data, I typically do this using color. I could, however, incorporate size changes in lines or points, blurring, or similar shapes when groupings become too large or complicated to justify colors or if magnitude of observations needs to be expressed in some way.