For this lab, in order to work with my data efficiently and create the 8 visualizations, I had to adjust the data accordingly. Some visualizations required optimized data in order to load and work properly. For others, all of my data was efficiently processed. Certain visualizations were better in showcasing my data over others.

The bivariate scatterplot was the perfect way of showing all of my data. I did not adjust anything from the previous lab. It properly showed the different categories of apps available by color coding. The only down side was that because there were too many apps to account for, it looked cluttered in certain graphs and some of the axis intervals didn't give the data justice, specifically "Number of Installations."

The correlation matrix gave a very general view of the data, but it's difficult to identify specific trends because the data depends heavily on categories as well which cannot be shown as a numerical value.

The scatter plot matrix and parallel coordinate display provided a better understanding of the data and allowed for color coding. The only downside was that due to the processing of these programs, I had to downsize my data and remove any apps with similar attribute values. At the end, I made sure I had 15 apps per category to ensure the user can see the different trends within each category. The scatterplot matrix made it easier to compare the various combinations of axes. The parallel coordinate allowed for the user to see the data for each app individually and compare them with other apps, while also being able to see all attributes at once for more comparisons and observations.

The PCA plot, biplot, and MDS displays did not clearly show the trends because it is highly dependent on numerical values, which in the case of my data, wasn't as compatible. Although, it did display some numerical values, it didn't do the full job. Some of my data

numbers are largely different, for example number of installations vs. app rating – the difference being 1million vs. 4.5. The varying numbers and large scale difference affected the output of the 4 visualizations mentioned. For the MDS display, perhaps clustering will make a big difference, but the display for this lab did not help all that much.

Moving forward with my data, I want to find a better way to identify the categories when graphing and creating visualizations. I also want to create something more interactive so that the user can better understand certain trends.