My Growth as a Data Scientist

I create a ton of data in my research work at cal poly and as a result of this class I think a lot about my data entry now. I now create data sets with the mind set of how I could process them using Rstudio later. One thing that has changes a lot is I know use more variable so I can subset and analyzing more easily later. With more variables it easier pull specific data out of data set, and it makes a huge difference in the number of visualizations I am able to create when I am analyzing data. One hypothetical example would be if I am looking dependent and independent variable I would now also add what sample batch they are in. With the added variable of the sample batch I can use the group\_by( ) function to subset it by batch and create a column visualization to ensure there is no wide spread variation between batches. Also, think about what the the data is not showing now and how an additional visualization could give the reader more data. An example of this is adding a jitter aesthetic behind a box plot to help the reader see how the statistics visualized by the box plot is being influenced.