In this module we were asked to perform data analytics on mouse data from a study. We were given the Mouse ID, Drug regimen, Sex, Age, Weight, Tumor Volume, Timepoints, and Metastatic datapoints. We had to clean the data and noticed there were two entries for the same mouse ID so we have to drop some of those values to have a clean data frame to work with. I pulled the overall statistical summary of mean, median, variance, standard deviation, and standard error from each drug regimen. From those data points, it looks like Capomulin and Ramicane had the lowest tumor volume stats across the board. From an initial look, it may seem that those two drug regimens would be the most successful and treating tumors. If you split the data by sex, it looks like those datapoints don’t have much effect on the drug regime. The male and female distribution is almost evenly split with 51% males and 49% females. There were not many outliers with the data either other than Infubinol, which had one outlier. Overall, it does look like Capomulin is an effective drug regimen because we took one mouse sample and performed a scatter plot. The plot showed a dramatic decrease in the tumor size over time. I did have an issue trying to perform a correlation on this mouse data, so I still must figure that piece out. I was still able to get the formula for the regression as y=0.95x + 21.55. If I were to continue the study, I would perform a scatter plot and regression analysis with a Ramicane treated mouse to see if it had the same effects as Capomulin since they had similar values.