Data Visualization Challenge - Observations

The Data Visualization challenge analyzed data from Pymaceuticals, Inc. and the purpose of the study was to compare different drug treatment regimens. The data came from two data sets that were merged together into a new DataFrame. Then, the duplicate mouse was found and another DataFrame was created. Bar charts, pie charts, quartiles, outliers, box plots, line plots, scatter plots, correlation and regression were all used in this assignment.

The bar charts that were created using Pandas and Pyplot showed the total number of rows (Mouse ID and Timepoints) for each drug regimen used. Capomulin and Ramicane had more observed timepoints than the others and based on the aggregations, they both had the lowest variance comparatively.

The distribution of female versus male mice was visualized in a pie chart using both Pandas and Pyplot. 51% or 126 of the subjects were male and 49% or 122 were female. Since they were fairly equal, it could be assumed that the results of the analysis would be true for both sexes.

Four Treatment groups were used in a box plot to show the distribution of the tumor volume for each group. Ramican had a lower median, but Capomulin had a smaller box which means that the data was not spread out and more condensed than Ramican. Infubinol, had an outlier in the lower range.

The line chart depicted Tumor Volume versus Time points for Mouse ID s185. This chart shows that the Tumor Volume decreased on Capomulin from 45 mm3 to 23 mm3 over the course of 45 days.

Finally, Correlation and regression were analyzed and the data was displayed in a scatter plot with the line of best fit and the r-squared value was calculated as .71. This means there is a strong correlation between weight and tumor volume for Capomulin.