**Setup / Initial Data Review**

Before analyzing the data into various dataframes or plots, I first reviewed the combined data for any duplicates. The duplicate\_mice dataframe provides the duplicate data. Thereafter, this item was dropped, and value counts were re-validated.

**Summary statistics**

Summary statistics were provided in 3 different ways. The first was utilizing the describe() function. This was helpful in determining that my code for future steps were indeed correct. The second summary table was created through grouping items by Drug Regimen and calculating each requested statistical value, where these were then placed into a dataframe. The last summary of statistics was utilizing the aggregation method and you will see that only 1 line of code was needed to capture requested statistics.

**Bar and Pie Charts**

Prior to calling information into charts, I thought it would be much cleaner to create a dataframe for the data requested. This was also helpful to see that information selected is correctly applied to the charts and was able to sort the values for an easier to read chart. The first bar chart was done through the pandas call "plotdata.plot(kind="bar")". The second bar chart was created through use of pyplot (matplotlib). Initially I was having challenges with the data matching the previous chart and values on my dataframe, but I was able to troubleshoot and adjust (I was trying to pull individual variables, rather than call the dataframe).

Similar to the creation of the bar charts, I also created a simple dataframe for creating the pie charts on the sex of mice in this drug trial. Again, the first chart was created through pandas and the second was through pyplot.

**Quartiles, Outliers and Boxplots**

This was a fairly challenging section, as I didn't get much practice. However, I continued to circle back to examples worked on through class.

**Line and Scatter Plots**

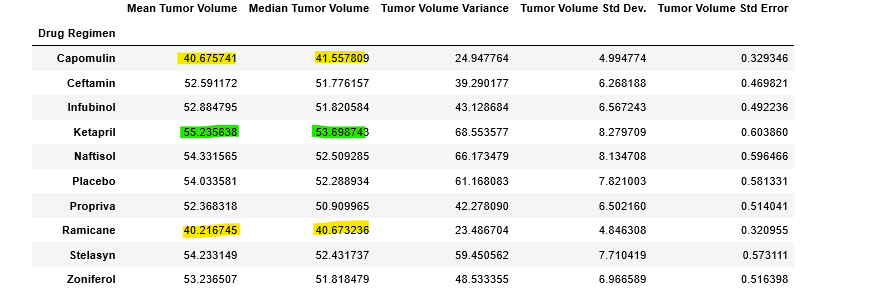
Similar to work done with the bar and pie charts, I opted to create a dataframe with only the data needed to plot the line and scatter plots. This helps show the values on the dataframe match that of the chart. I am not sure why, but even though I have the code to add chart titles (including for the axes) these did not appear in my output.

**Correlation and Regression**

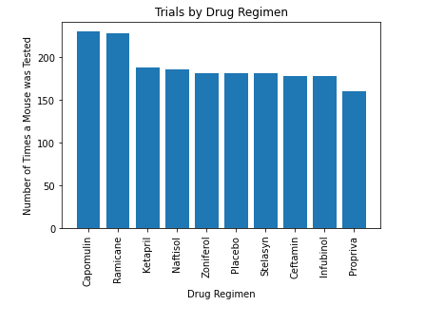
Again, this was challenging as I did not practice this more than I would have liked. I referred to class examples and replayed class recordings to assist with completing this section. I ran into the same issue with the scatter plot titles, however was able to show the slope on the graph. I am guessing that I may be missing a print call for the y='slope'+ x + 'intercept', but my print call does not print within the graph.

**Overall analysis**

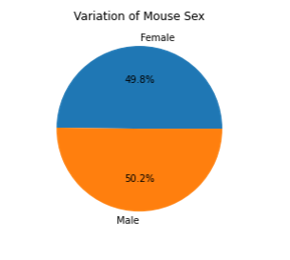
In reviewing the summary statistic data, it appears there are a couple drug regimens in the trial that have shown lower tumor volumes than other treatments. These are for Capomulin and Ramicane. Both these regimens have means and medians in the low 40s, where all the other regimens are in the 50s. The worst performing regimen is surprisingly not the placebo, but Ketapril.



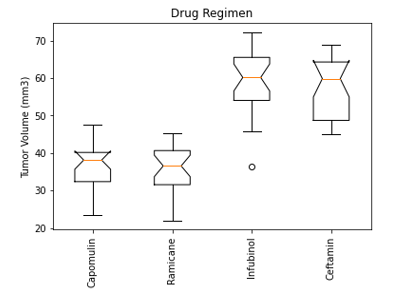
This bar chart of trials by regimen shows the consistency of how many times a mouse was tested for a unique drug. In analyzing this data, it may be likely that the scientists performing testing may wish to increase testing with drugs other than Capomulin and Ramicane so that there is an even distribution of testing.



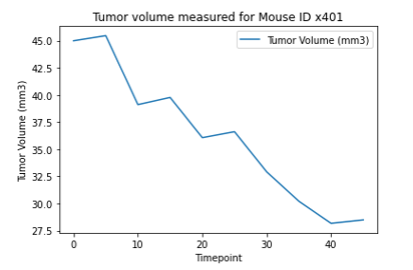
This pie chart shows that the sex of mice tested were evenly distributed.



When I ran code to determine quartiles, this helped identify potential outliers in this trial. Additionally, the boxplot below helps to highlight the outlier related to the Infubinol drug regimen. This would require investigation on whether there was an error with data capture or measurement of the tumor volume.



I selected a mouse (x401) to measure its tumor volume over time. This mouse was treated with Capomulin and the tumor volume showed a decrease in size over the timepoint in this study.



The scatter charts below show the correlation of the average weight of mice (x axis) and the tumor volume (y axis). This evidences a positive correlation, which measures at 0.8419.

