Module 5 Challenge – Pymaceuticals

William Davis | [william.c.davis@nasa.gov](mailto:william.c.davis@nasa.gov)

----

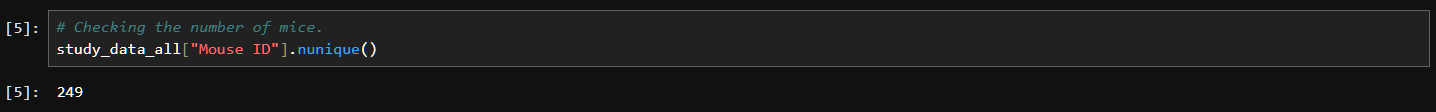
Summary Analysis

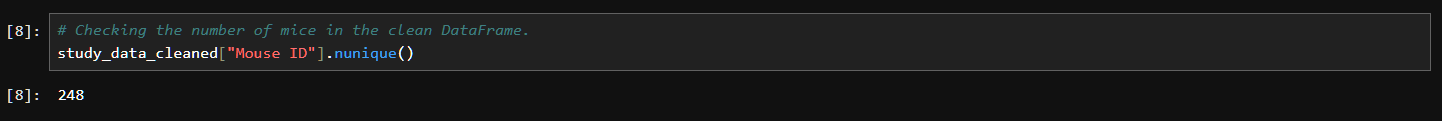
* The final tumor volume was notably lower for mice on the Capomulin and Ramicane drug regimens compared to those on Infubinol and Ceftamin.
* An analysis of tumor volume for Mouse b128 showed a tumor volume reduction from 45 to 39 mm3.
* There is a good correlation (~84%) between tumor volume and mouse weight for those mice on the Capomulin drug regimen.

**Prepare the Data (20 points)**

* The datasets are merged into a single DataFrame. (6 points)
* The number of mice are shown from the merged DataFrame. (2 points)
* Each duplicate mice is found based on the Mouse ID and Timepoint. (6 points)
* A clean DataFrame is created with the dropped duplicate mice. (4 points)
* The number of mice are shown from the clean DataFrame. (2 points)

See attached file: Pymaceuticals.ipynb





**Generate Summary Statistics (15 points)**

* The mean of the tumor volume for each regimen is calculated using groupby. (2 points)
* The media of the tumor volume for each regimen is calculated using groupby. (2 points)
* The variance of the tumor volume for each regimen is calculated using groupby. (2 points)
* The standard deviation of the tumor volume for each regimen is calculated using groupby. (2 points)
* The SEM of the tumor volume for each regimen is calculated using groupby. (2 points)
* A new DataFrame is created with using the summary statistics. (5 points)

See attached file: Pymaceuticals.ipynb



**Create Bar Charts and Pie Charts (15 points)**

* A bar plot showing the total number of timepoints for all mice tested for each drug regimen using Pandas is generated. (4.5 points)
* A bar plot showing the total number of timepoints for all mice tested for each drug regimen using pyplot is generated. (4.5 points)
* A pie chart showing the distribution of unique female versus male mice using Pandas is generated. (3 points)
* A pie chart showing the distribution of unique female versus male mice using pyplot is generated. (3 points)

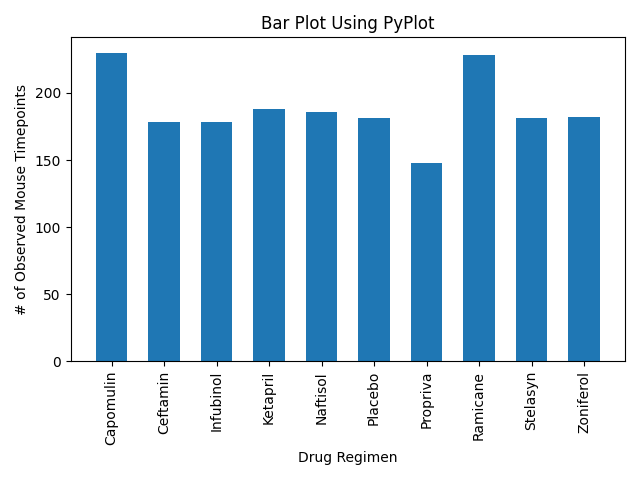
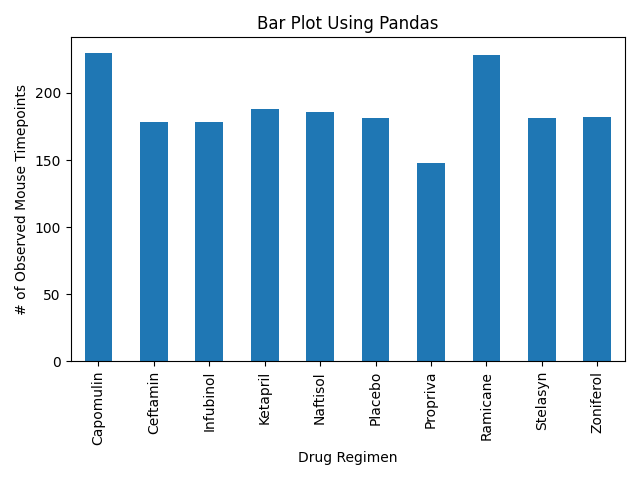
See attached file: Pymaceuticals.ipynb

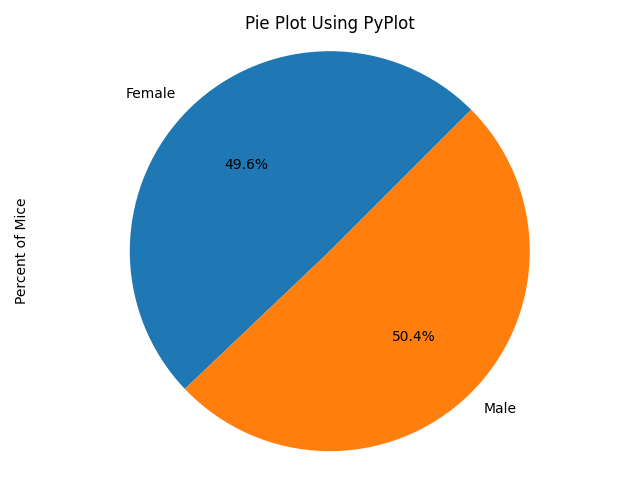
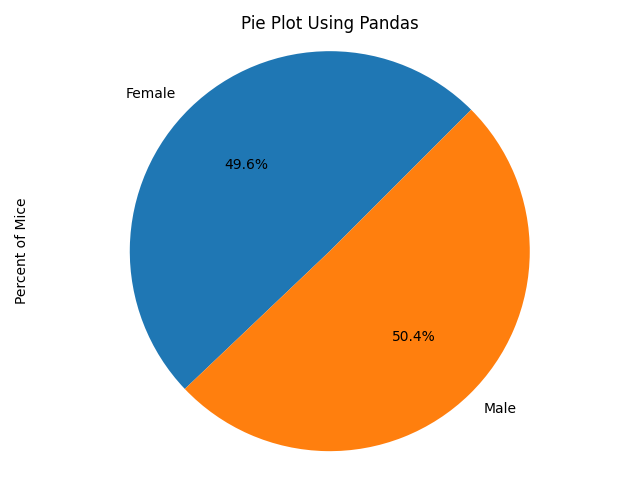
See attached figures: bar\_plot\_using\_pandas.png

bar\_plot\_using\_pyplot.png

pie\_plot\_using\_pandas.png

pie\_plot\_using\_pyplot.png



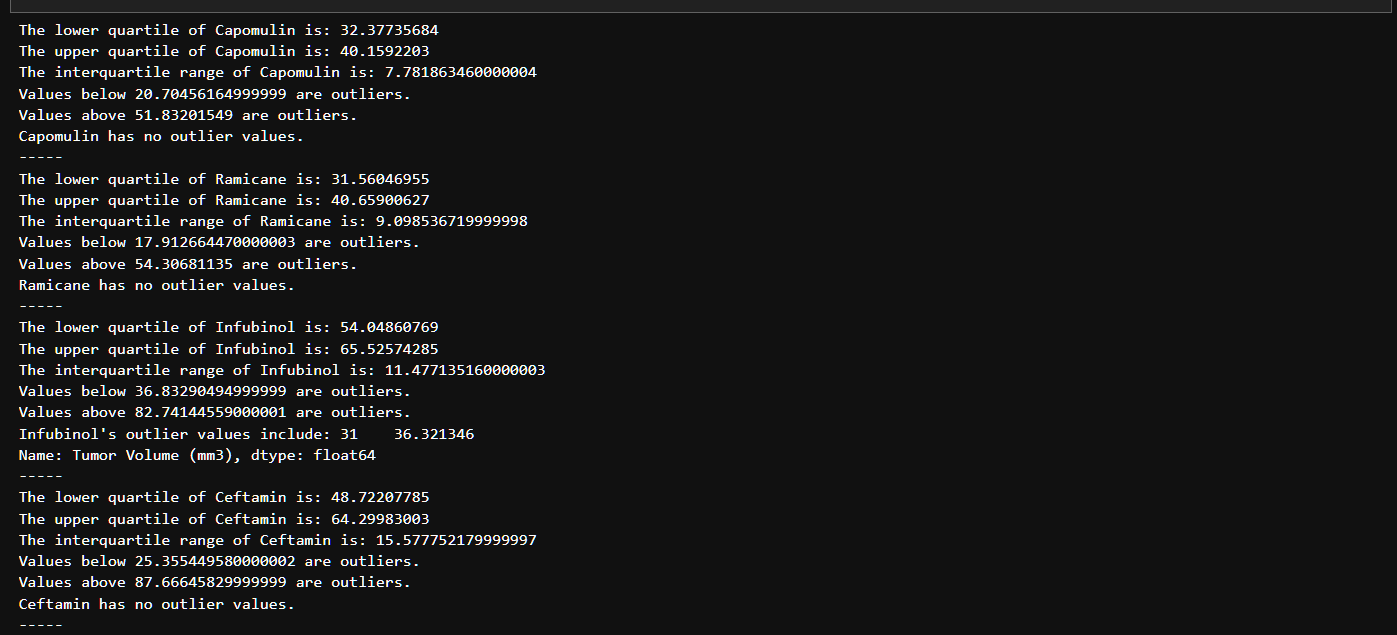


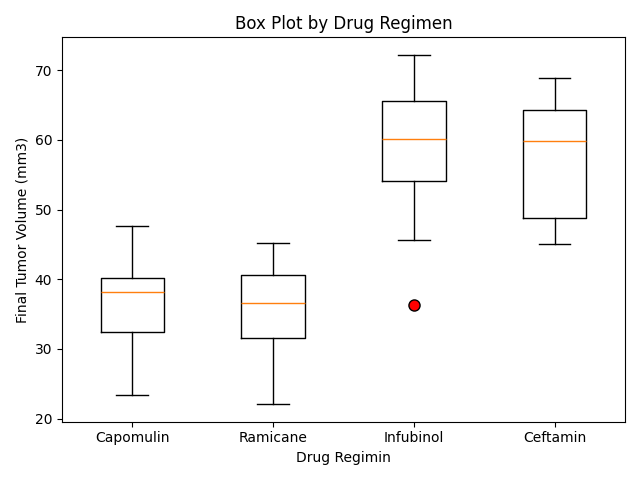
**Calculate Quartiles, Find Outliers, and Create a Box Plot (30 points)**

* A DataFrame that has the last timepoint for each mouse ID is created using groupby. (5 points)
* The index of the DataFrame is reset. (2 points)
* Retrieve the maximum timepoint for each mouse. (2 points)
* The four treatment groups, Capomulin, Ramicane, Infubinol, and Ceftamin, are put in a list. (3 points)
* An empty list is created to fill with tumor volume data. (3 points)
* A for loop is used to display the interquartile range (IQR) and the outliers for each treatment group (10 points)
* A box plot is generated that shows the distribution of the final tumor volume for all the mice in each treatment group. (5 points)

See attached file: Pymaceuticals.ipynb

See attached figures: box\_plot\_by\_drug\_regimen.png





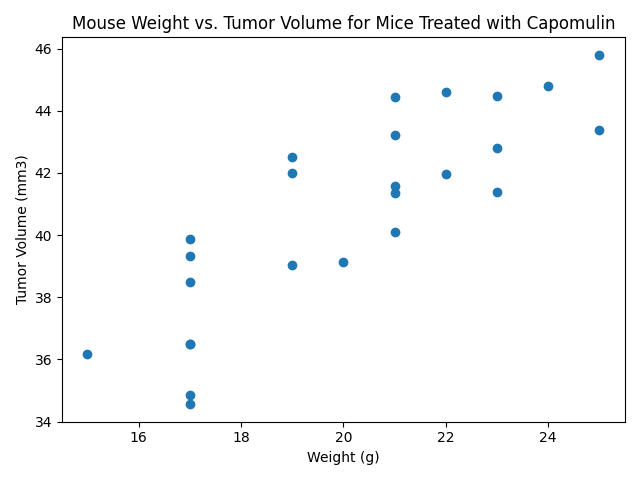
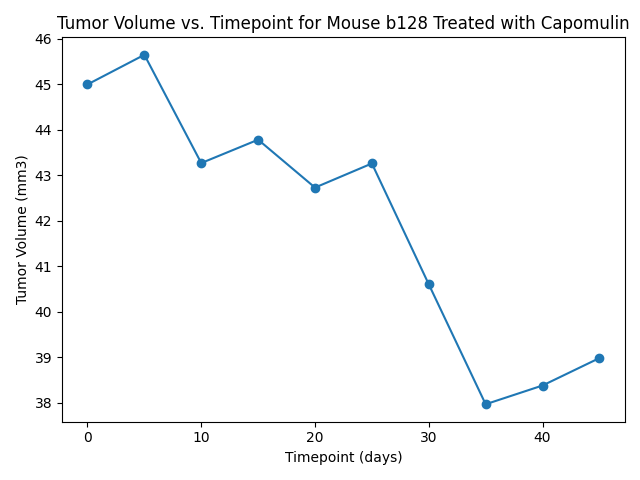
**Create a Line Plot and a Scatter Plot (10 points)**

* A line plot is generated that shows the tumor volume vs. time point for one mouse treated with Capomulin. (5 points)
* A scatter plot is generated that shows average tumor volume vs. mouse weight for the Capomulin regimen. (5 points)

See attached file: Pymaceuticals.ipynb

See attached figures: line\_plot\_tumor\_vol\_vs\_timepoint.png

scatter\_plot\_mouse\_weight\_vs\_tumor\_vol.png

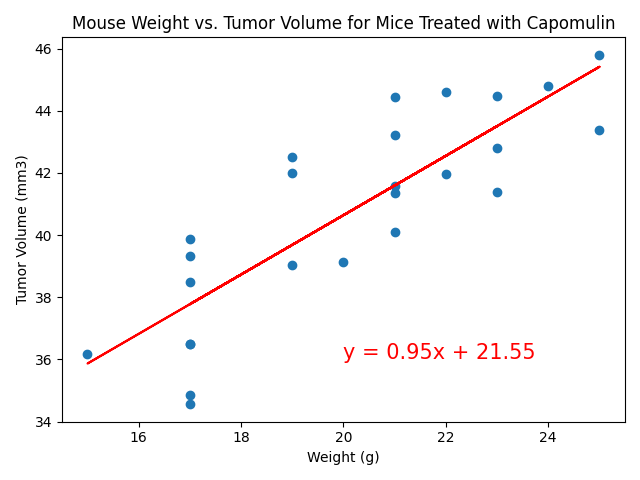


**Calculate Correlation and Regression (10 points)**

* The correlation coefficient and linear regression model are calculated for mouse weight and average tumor volume for the Capomulin regimen. (10 points)

See attached file: Pymaceuticals.ipynb

See attached figures: scatter\_plot\_mouse\_weight\_vs\_tumor\_vol\_regression.png

****