1. For this assignment, I intended to use as much starter code as possible. All starter code was left intact.
2. The initial data cleaning relied heavily on the previous Pandas exercise.
3. Cells 1 and 3 are original code written by me. In order to simplify Syntax and avoid errors encountered in cell 3, I wanted to rename the data frame columns that contained irregular characters. I used the "Rename columns by mapping:" technique found at https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.rename.html to rename the columns in question.
4. For cells 5, 6 and 7, I used Professor Booth's code to find and drop the duplicate mouse from the data.
5. I referred to Prof.'s example to get the variance, standard deviation, and SEM functions for my code.
6. I used the line `avg_tumor = df2.groupby("Drug Regimen")["Tumor Volume (mm3)"].mean()` from Prof.'s code as a basis for calculating the mean, median, variance, standard deviation, and SEM of tumor sizes for each drug regimen in cell 9.
7. I also used Prof's example code in cell 10 to create a data frame using aggregation
8. I used Prof.'s example code in cell 11 to create a visualization using Pandas.
9. To create a visualization in cell 12, I used the "Data Viz Recipe" example code that has been used many times during our in-class activities.
10. I used Prof's "option1" code to eliminate duplicate mice in cell 13. I used slightly different syntax from Prof.'s example to count the number of male and female mice in cell 14.
11. I also used Prof.'s code in cell 15 to create a pie chart using Pandas,
12. To create a visualization in cell 16, I used the "Data Viz Recipe" example code that has been used many times during our in-class activities
13. I used Prof.'s example in cell 17 to find the last timepoint for each mouse and append it to my data frame. I also used his example to create a data frame that dropped mice which received placebo treatment.
14. I used Prof.'s example, which was taken from Xpert AI, entirely in cell 19.
15. To create a visualization in cell 21, I used the "Data Viz Recipe" example code that has been used many times during our in-class activities. I did copy the line `plt.boxplot(tumor_volumes, labels=treatments)` from Prof.'s code.
16. I used Prof.'s example code to isolate data for a single mouse in cell 23.
17. In cell 25, I used the "Data Viz Recipe".
18. I reused the method from cell 23 to isolate the drug Capomulin, then I used Prof.'s line `df6 = df5.groupby("Mouse ID").agg({"Tumor Volume (mm3)": "mean", "Weight (g)": "mean"}).reset_index()` to find the average weights and tumor volumes of each mouse in cell 27.
19. In cell 32, I used the "Data Viz Recipe".
20. I copied Prof.'s example code to create a regression in cell 29. I also added lines from his example into the "Data Viz Recipe" that I copied down from cell 28 to re-plot my graph with a line of best fit in cell 33. The lines I added into cell 33 were `# Annotate - plt.annotate(line_eq, (22, 37), fontsize=15, color="black")` and `plt.plot(avg_cap_tumor.Weight, regress_values, color="black", linewidth=6) # the regression line`.