Pymaceuticals

Introduction

Figure 1. Illustration by author

Pymaceuticals Inc., a burgeoning pharmaceutical company based in San Diego, specializes in anti-cancer pharmaceuticals. Recently, Pymaceuticals Inc. began screening for potential treatments for squamous cell carcinoma (SCC), a commonly occurring form of skin cancer.

In their most recent animal study, 250 mice identified with SCC tumor growth were treated through a variety of drug regimens. Over the course of 45 days, tumor development was observed and measured. The purpose of the study was to compare the performance of Pymaceuticals' drug of interest, ***Capomulin***, versus the other treatment regimens. As tasked by the executive team, this report provides a top-level summary and the tables and figures needed for the technical report of the study.

Business Problem:

The project is based on the following business problem:

What are the meaningful insights regarding the performance of an anti-cancer drug Capomulin gained from a recent animal study?

Target Audience:

This report is targeted to the ***(a) business stakeholders (b) executive team*** and ***(c) scientists*** developing anti-cancer drugs in Pymaceuticals Inc.

Data source & cleaning:

1. The data for drug regimen applied to each mouse (Mouse_metadata.csv) and time evolution of the tumor (Study_results.csv) available as csv files in the Resources subdirectory.

2. No discrepancy was found in the data types of each column. There were no missing data in either data-frames. Both the data-frames were "outer" merged on *"Mouse ID"* column.

Data Statistics:

- Ten drug regimens were studied. They are: Capomulin, Ketapril, Naftisol, Infubinol, Stelasyn, Ramicane, Propriva, Zoniferol, Placebo and Ceftamin
- Number of times a drug regimen was administered **varied from one to 13**. A drug administered once had only one timepoint (i.e. timepoint = 0).
- **Table 1**. Shows the statistics of tumor volume for each drug regimen encompassing all mice samples in the study.
- Drug regimen ***Ramicane*** has the smallest mean tumor volume followed by ***Capomulin***.
- Since the starting volume of all tumor is 45mm^3, ***Ramicane*** and ***Capomulin*** are only two drugs that resulted in decrease in tumor volume. Average tumor volume increased for all other drug regimens.
- Let's select a subset of the data where mice ***have been exposed to a drug regimen more than once.***
- **Table 2**. Shows the statistics of tumor volume for each drug regimen for such a subset (i.e. samples with only one time point removed).
- By comparing the tables, we see that variance decreased for: ***Ketapril, Naftisol, Stelasyn, Zoniferol, Placebo, Ceftamin*** and ***Propiva.*** The mice samples not considered in Table 2 belonged to those drug regimens.

Drug Regimen	Mean	Median	Variance	Std_dev	SEM
Capomulin	40.68	41.56	24.95	4.99	0.33
Ceftamin	52.59	51.78	39.29	6.27	0.47
Infubinol	52.88	51.82	43.13	6.57	0.49
Ketapril	55.24	53.70	68.55	8.28	0.60
Naftisol	54.33	52.51	66.17	8.13	0.60
Placebo	54.03	52.29	61.17	7.82	0.58
Propriva	52.32	50.85	42.35	6.51	0.51
Ramicane	40.22	40.67	23.49	4.85	0.32
Stelasyn	54.23	52.43	59.45	7.71	0.57
Zoniferol	53.24	51.82	48.53	6.97	0.52

Table 1. Statistics table of all mice samples

Drug Regimen	Mean	Median	Variance	Std_dev	SEM
Capomulin	40.68	41.56	24.95	4.99	0.33
Ceftamin	52.77	51.91	38.84	6.23	0.47
Infubinol	52.88	51.82	43.13	6.57	0.49
Ketapril	55.35	53.74	68.15	8.26	0.61
Naftisol	54.43	52.54	65.93	8.12	0.60
Placebo	54.08	52.32	61.05	7.81	0.58
Propriva	52.37	50.91	42.28	6.50	0.51
Ramicane	40.22	40.67	23.49	4.85	0.32
Stelasyn	54.28	52.49	59.30	7.70	0.57
Zoniferol	53.28	51.83	48.42	6.96	0.52

Table 2. Statistics table for mice samples with more than one timepoint

Bar Plots:

Figures 2 & 3 are Bar Plots showing number of data points for each treatment regimen.

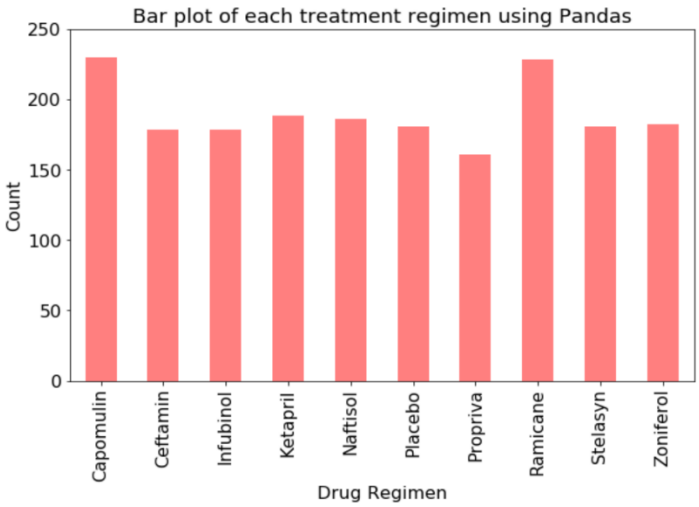

Figure 3. Bar Plot using Pandas.

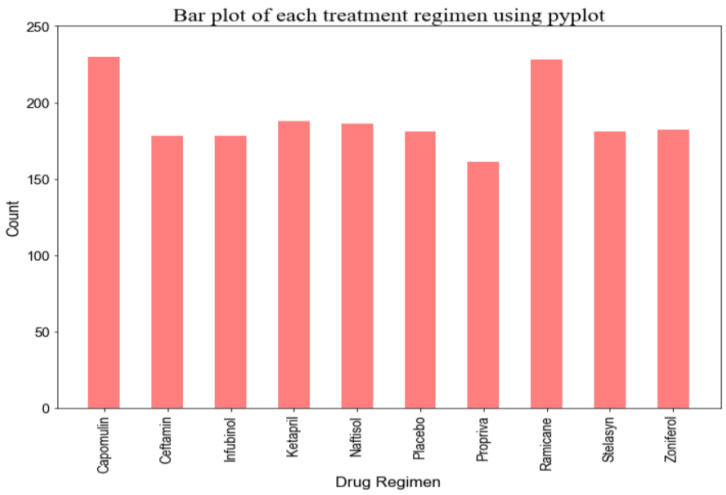

Figure 2. Bar Plot using Pyplot.

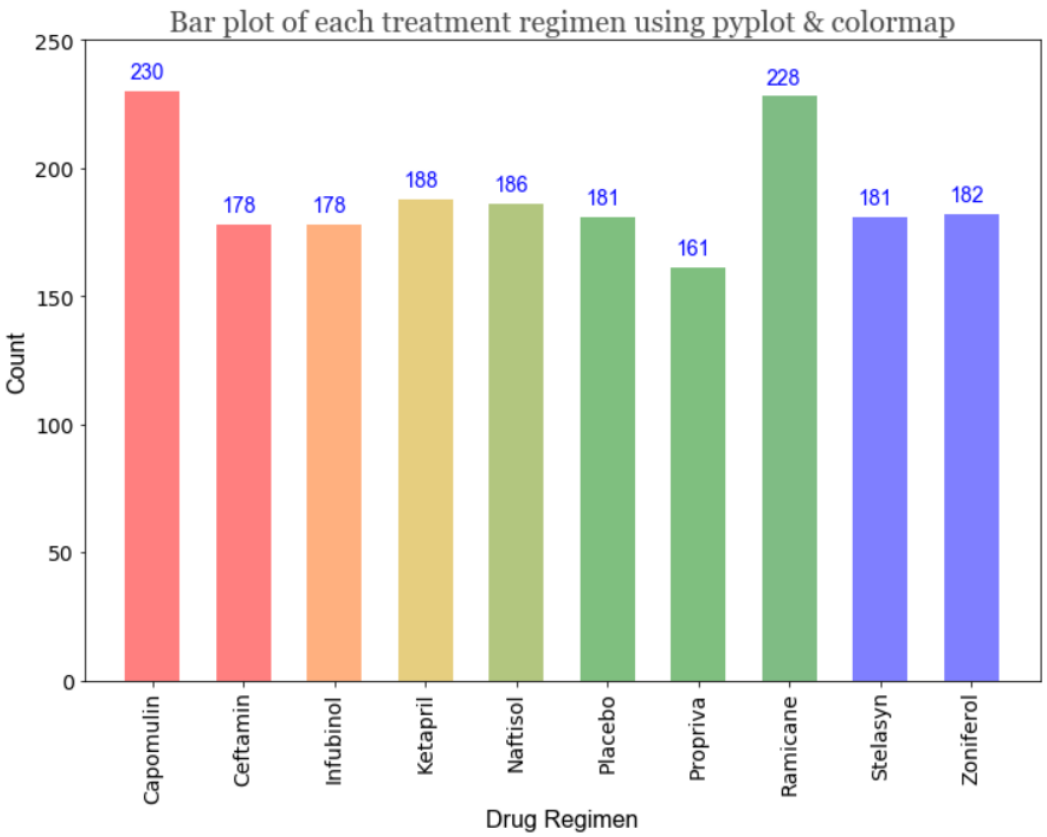

Figure 4. Bar Plot using Pyplot & Colormap

Figure 4 is the same plot using Pyplot and Colormap

Capomulin & ***Ramicane*** were tested 230 and 228 times respectively. About 21% more than any other drug regimen.

Pie Plots:

There are 124 (49.8%) female mice and 125 (50.2%) male mice in the study

Sex	Count	Percent Count
Female	124	49.8
Male	125	50.2

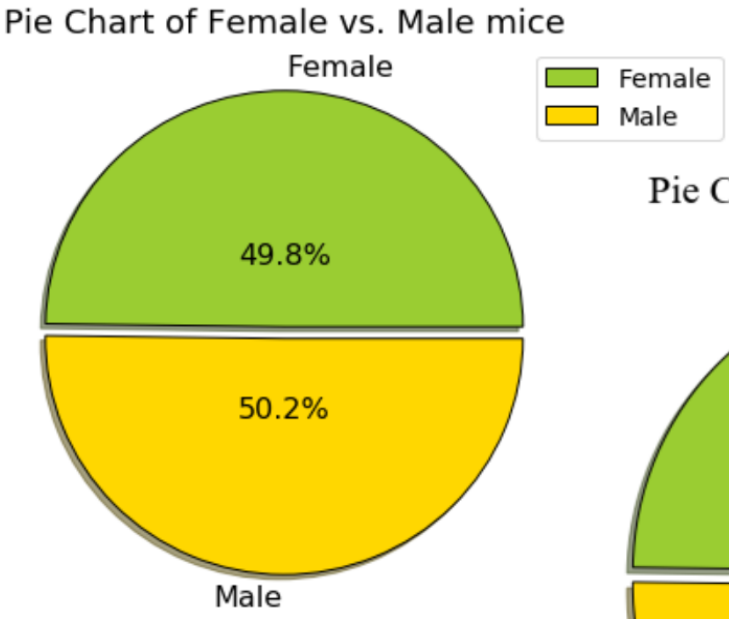

Figure 5. Pie chart using Pandas

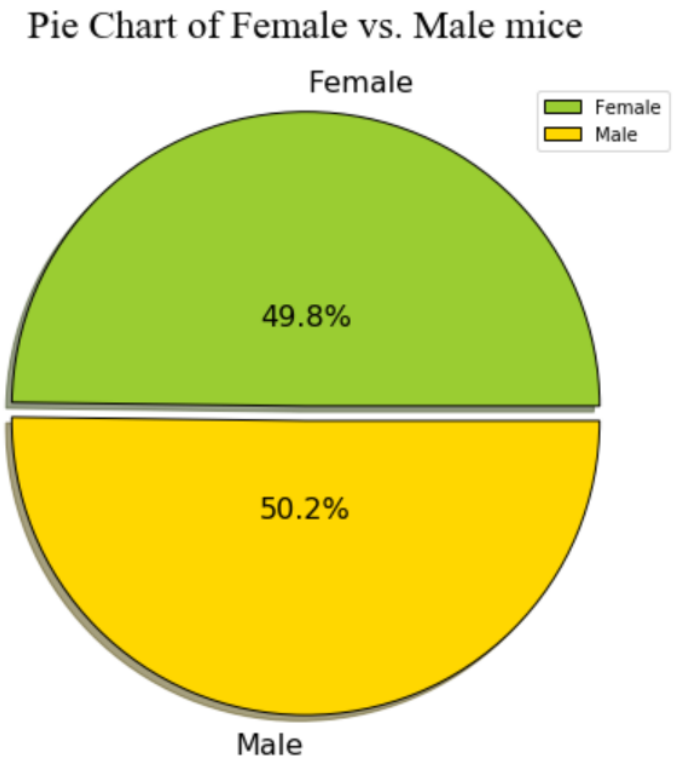

Figure 6. Pie chart using Pyplot

The study was almost equally distributed between male and female gender.

Selecting the most promising treatments:

We selected the most promising drug regimens by comparing two dataframes:

a. df_treatment_start: Contains start tumor volume and start time for each regimen
b. df_treatment_end: Contains end tumor volume and end time for each regimen

The average study duration is calculated as an average of differentials between treatment end and start timepoints. The candidates with null study duration were dropped from the dataframe before calculating the average.

Figure 7 shows the average study duration for each drug regimen.

Capomulin and ***Ramicane*** were studied the longest among all drug regimens.

The average change in tumor volume is calculated as an average of differentials between start and end tumor volumes

Figure 8 shows the average change in tumor volume for each drug regimen.

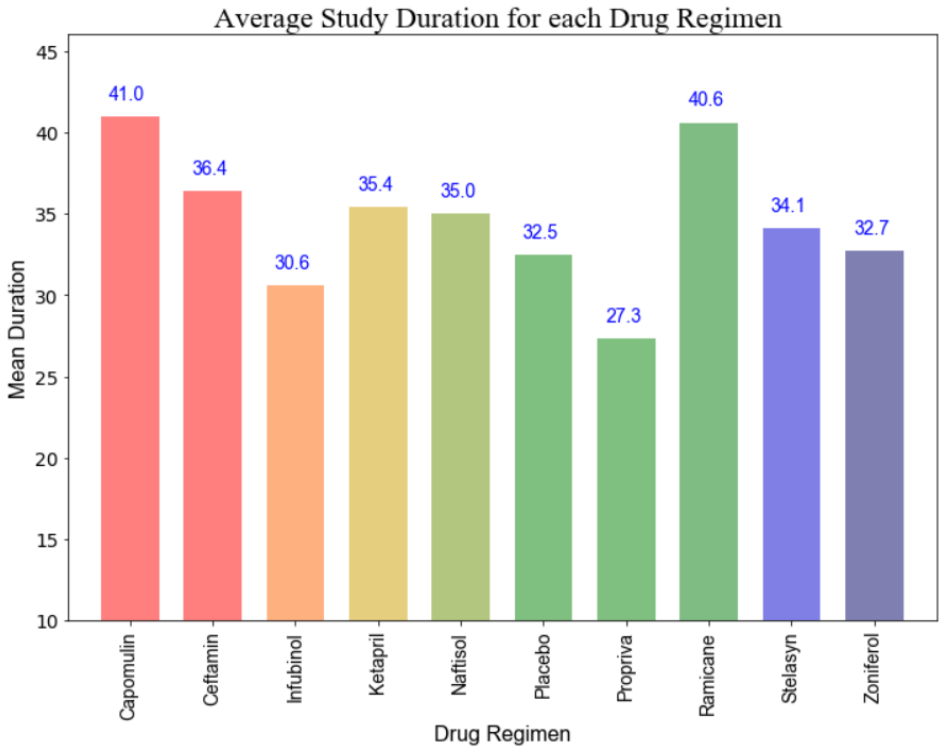

Figure 7.Average study duration for each drug regimen

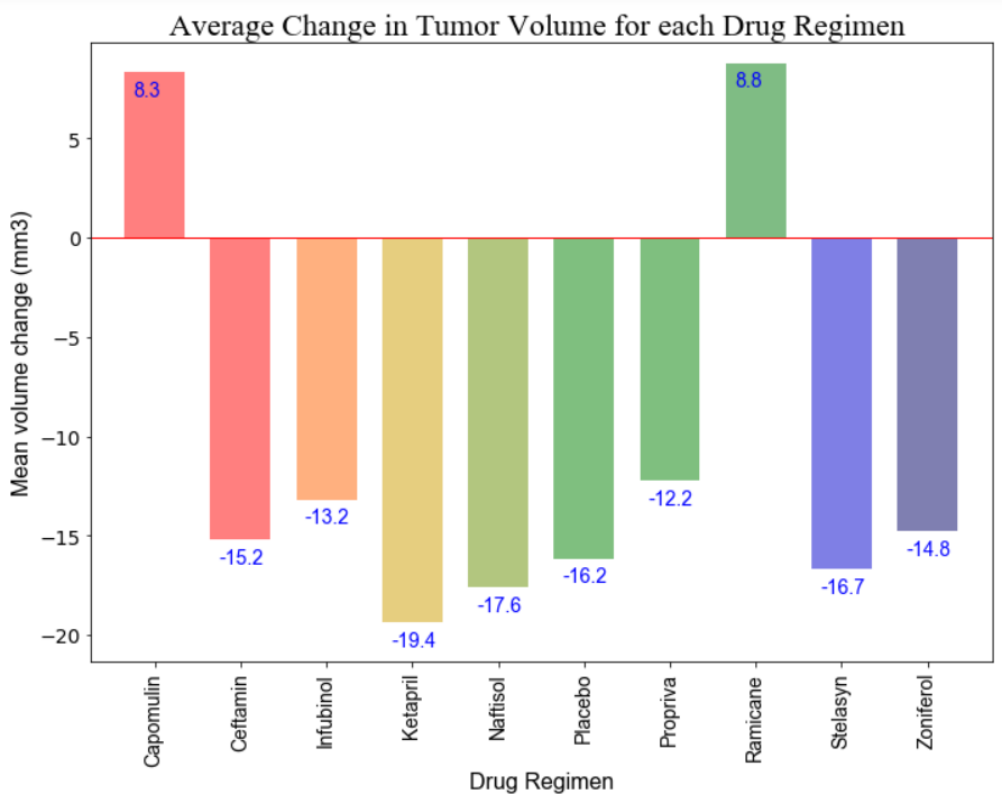

Figure 8. Average Change in Tumor Volume for each drug regimen

Only ***Capomulin*** and ***Ramicane*** resulted in positive change in tumor volume (mm^3).

Which implies that the tumor ***decreased in volume*** over the study duration. This is in sharp contrast with other drug regimens, where the tumor increased in volume.

Highest mean volume change was obtained by:

Ramicane **(8.8 mm^3)** closely followed by ***Capomulin*** **(8.3mm^3)**

	Drug Regimen	Mean Duration	Mean Vol Change	Rate of Change	Est Vol Change
0	Capomulin	41.00	8.332	0.203	5.075
1	Ceftamin	36.43	-15.183	-0.417	-10.425
2	Infubinol	30.60	-13.178	-0.431	-10.775
3	Ketapril	35.43	-19.355	-0.546	-13.650
4	Naftisol	35.00	-17.615	-0.503	-12.575
5	Placebo	32.50	-16.155	-0.497	-12.425
6	Propriva	27.29	-12.226	-0.448	-11.200
7	Ramicane	40.60	8.809	0.217	5.425
8	Stelasyn	34.13	-16.697	-0.489	-12.225
9	Zoniferol	32.71	-14.772	-0.452	-11.300

Table 3. Analysis of Drug Regimens performance with respect to average tumor volume change

After estimating the average study duration and the average tumor volume change, we can calculate the rate of change of tumor volume for each drug regimen. We can further estimate the tumor volume change if the time end point was same (say, 25) for all the regimens. The results (***mean study duration***, ***mean volume change***, ***rate of change*** and ***estimated volume change***) are tabulated in Table 3. Sorting Table 3 in descending order with respect to ***"Estimated Volume Change"*** results in Table 4.

	Drug Regimen	Mean Duration	Mean Vol Change	Rate of Change	Est Vol Change
0	Ramicane	40.60	8.809	0.217	5.425
1	Capomulin	41.00	8.332	0.203	5.075
2	Ceftamin	36.43	-15.183	-0.417	-10.425
3	Infubinol	30.60	-13.178	-0.431	-10.775
4	Propriva	27.29	-12.226	-0.448	-11.200

Table 4. Top 5 Drug Regimens based on average tumor volume change

The top 4 drug regimens as showed in Table 4 are: Ramicane, Capomulin, Ceftamin and Infubinol

Quartiles, Outliers and Boxplots:

Table 5 shows the statistics of the final tumor volume for the top 4 drug regimens:

Drug Regimen	Mean	Median	Lower Quartile	Upper Quartile	IQR	Upper Bound	Lower Bound
Capomulin	36.67	38.13	32.38	40.16	7.78	51.83	20.71
Ceftamin	57.75	59.85	48.72	64.30	15.58	87.67	25.35
Infubinol	58.18	60.17	54.05	65.53	11.48	82.75	36.83
Ramicane	36.19	36.56	31.56	40.66	9.10	54.31	17.91

Table 5. Statistics table of the average final tumor volume for the top 4 drug regimens

Checking for outliers (i.e. data > Upper Bound or data < Lower Bound):

Among the top four regimens, only Infubinol has one outlier

A Boxplot of the final tumor volume for the top four drug regimens is shown in Figure 9.

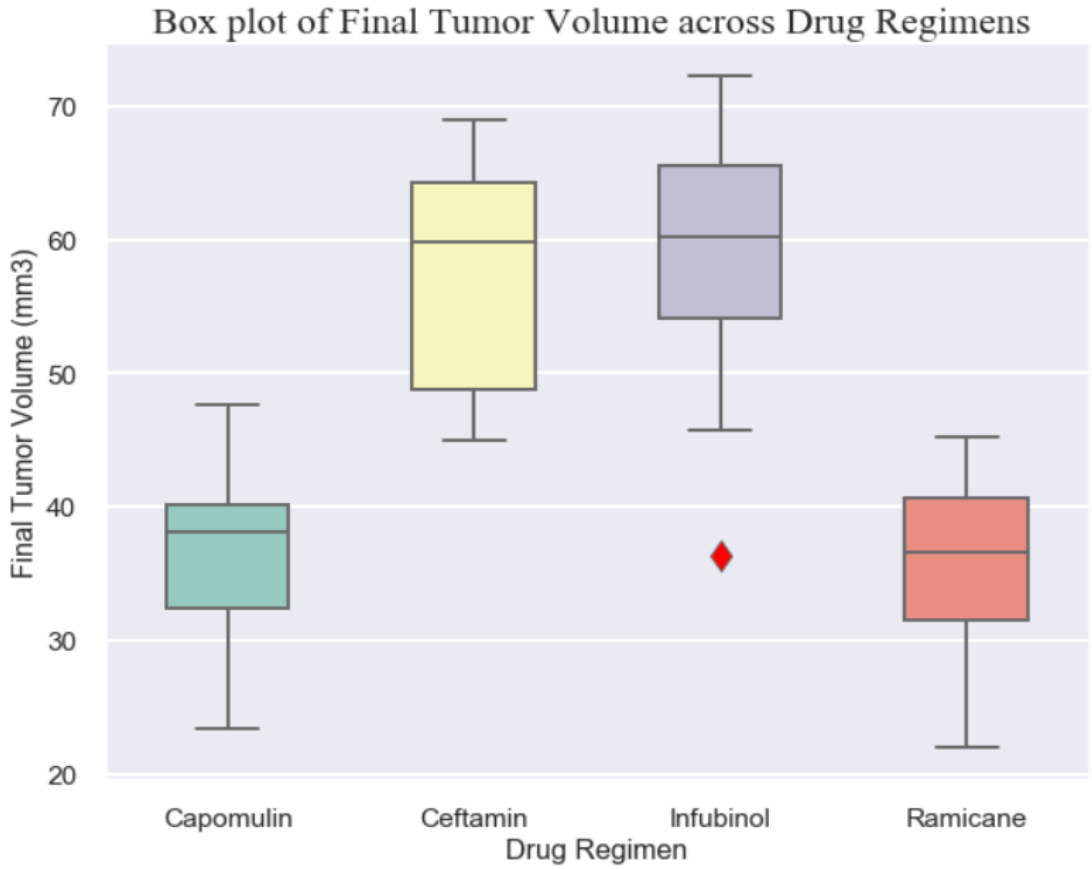

Figure 9. Box plot of the top four drug regimens

Line and Scatter plots:

Figure 10 shows the ***evolution of tumor volume*** for mouse ID ***b128*** treated with ***Capomulin*** regimen. Initial tumor volume is 45mm^3. A ***large decrease in tumor volume*** can be observed from time point 25 to 35. Interestingly, tumor volume increases as the study is continued beyond the timepoint of 35. ***The average weight (g) and average tumor volume is shown in Table 6***. The data from Table 6 is visualized in the form of a scatter plot as shown in Figure 11. The ***exists a strong positive correlation*** between the average tumor volume and average weight of mice samples under the drug regimen Capomulin.

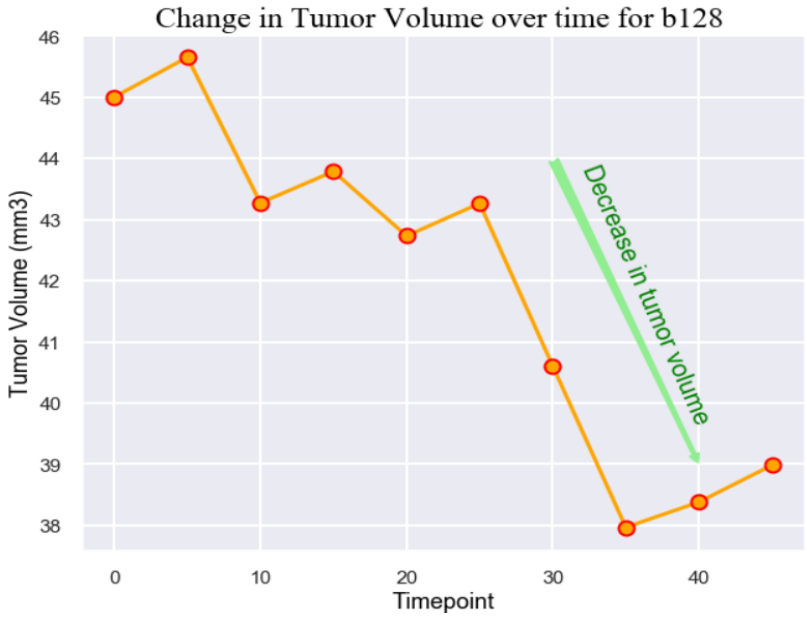

Figure 10. Evolution of tumor volume under drug regimen Capomulin

	Weight (g)	Mean Vol (mm3)
0	15	36.182
1	17	37.214
2	19	41.182
3	20	39.141
4	21	42.089
5	22	43.288
6	23	43.341
7	24	44.806
8	25	44.062

Table 6. Average Weight vs. Average Tumor Volume (Capomulin)

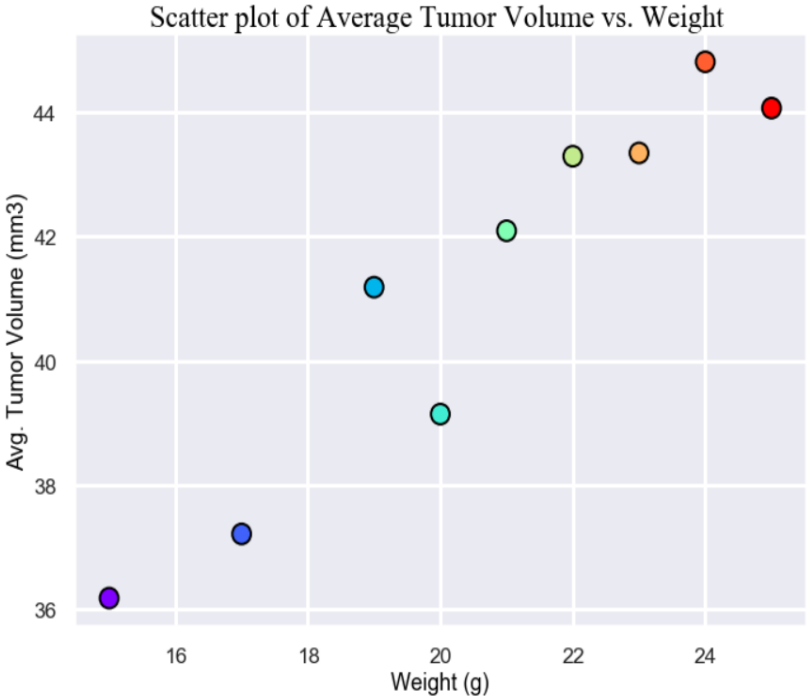

Figure 11. Scatter plot based on the data from Table 6 for drug regimen Capomulin

Linear Regression Model:

Figure 12 shows the output from a ***linear regression model*** applied to the data in Table 6. The model output is as follows:

$$Avg.Tumor\ vol\ (mm^3) = 0.89 * Weight(g) + 22.76$$

```
Correlation Coefficient (R-squared):0.951
p-value (level of statistical significance):8.45e-05
Standard error of estimate:0.111
```

R^2 is a statistical measure of fit that indicates how much variation of a dependent variable is explained by the independent variable(s) in a regression model. R^2 value of 0.951 implies 95% of the observed variation can be explained by model's inputs. The p-value indicates that the data is statistically significant (only 0.008% chance of being generated randomly)

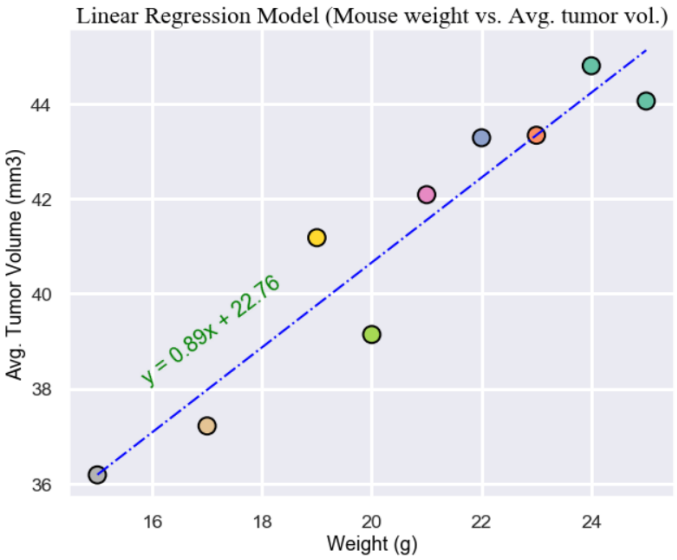

Figure 12. Linear Regression model of average tumor volume with respect to weight (g)

Figure 13. Illustration by Rob Donnely (www.slate.com)

Conclusions:

1. Dataset has almost equal number of male (125) and female (124) mice
2. Only Capomulin and Ramicane regimens resulted in decrease in tumor volume. Tumor volume increased for other drug regimens.
3. Topmost treatment regimens based on mean decrease in tumor volume are: Ramicane, Capomulin, Ceftamin and Infubinol
4. Average final tumor volume was lowest for Ramicane regimen and Capomulin regimen exhibited the least variability in final tumor volume
5. For Capomulin regimen, a strong positive correlation between average final tumor volume and average mouse weight was observed.