Anurag Diwate
CS 4375.004
3 February 2023

Data Exploration

A. Run of my code:

```
Opening file Boston.csv...
File opened!
Closing file Boston.csv.

Stats for rm:
Sum = 3180.03
Mean = 6.28463
Median = 6.2085
Range = 5.219

Stats for medv:
Sum = 11401.6
Mean = 22.5328
Median = 21.2
Range = 45

The covariance between rm and medv is 4.49345
The correlation between rm and medv is 0.69536

Process finished with exit code 0
```

B. I think having to program these functions in C++ really helped me understand just how convenient it is to have this data about a set built into the programming language. I think it is very useful for a language used by statisticians and data scientists to have these basic functions built-in.

C. Mean is the average value of a dataset. When the data is arranged from the minimum to the maximum value, median is the value in the middle of the dataset. The range of a dataset is calculated by subtracting the minimum value from the maximum value. These values can be very useful for data exploration since they can provide an insight into the spread of the data.

D. Covariance of two datasets is the measure of the direction of their relationship. This means that a positive covariance implies that both datasets are high or low at the same time, and a negative covariance implies that when a dataset tends high, the other tends to go lower. Correlation is the extent to which two datasets are linearly related. This information can be crucial in machine learning to find patterns that can indicate links in the dataset, and can be used to better train our ML algorithms to look for the observed patterns.