Khang Thai
9/11/22
CS 4375.003

```
Opening the file Boston.csv
Reading line 1
heading: rm,medv
new length 506
Closing file Boston.csv
Number of records: 506

Stats for rm
Sum: 3180.03
Mean: 6.28463
Median: 6.209
Range: 5.219

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

Covariance = 4.49345

Correlation = 0.69536

Program terminated
Process finished with exit code 0
```

Coding in my own functions is a lot harder than using R since I have to take into account how many variables there are, what I am looking for, and many other variables to make sure the function does exactly what I want. R allows me to simply enter in one statement and it will sort out the data immediately. I haven't coded in C++ in a while so it did take some time to remember what I needed to do even though the project was fairly easy. I just needed to remember how to make functions again and have the statement printed out so that I can figure out the rest.

Finding Mean, Median, and Range are useful in data exploration because Range allows us to see what the max and minimum values of the data are after sorting, the Mean allows us to find the average of the entire data set but also allows us to find the correlation and the covariance of the data, and the Median will give us the exact middle point of the data after being sorted.

Covariance is calculated by finding the mean of the 2 data sets individually and then taking each element and subtracting it from the mean for each set individually. After that, multiply the 2 data sets then add them together. Finally divide it by the total number of data sets minus 1. Covariance is basically correlation but from -1 to 1 scale. Correlation is measured by the covariance of 2 variables by dividing the standard deviation of x multiplied by the standard deviation of y. Covariance and Correlation are important to machine learning because it allows the machine to be able to predict future data. If the 2 data sets have a strong correlation then the machine can predict the data without actually recording the data.