

Assignment 1

Manuel Romero

Run of the program

```
Opening file Boston.csv
Reading Line 1
HEADING: rm,medv
New Length : 506
Closing file Boston.csv
Number of records: 506

Stats for rm
The sum of the vector data is :      3180.03
The mean of the vector data is :      6.28463
The median of the vector data is :      6.209
The range of the vector data is :      5.219

Stats for medv
The sum of the vector data is :      11401.6
The mean of the vector data is :      22.5328
The median of the vector data is :      21.2
The range of the vector data is :      45

Covariance = 4.49345

Corralation = 0.69536

Program terminated.Fernandos-MBP-2:~ fernandoromero$ |
```

Using the built in functions in R is much easier and faster than having to create my own code. The code is not hard to write but is time consuming especially compared to just using the built in functions. Furthermore, being able to open the Boston.csv file was difficult even with the given code. I had to go and specify the path for the file otherwise my code would not find it which is an error on my behalf.

The mean, median, and range are all useful for data analysis since it allows the user to fill in missing data that would be close to the actual data and allows the user to predict future data entries with formulas such as the best fit line that allows a user to plug in a value and get an expected value in return, they also include how close the actual data would be to the line. Covariance gives us how much a change in x changes y and correlation simply scales down to a smaller scale from -1 to 1. If it is either 1 or -1 that means the correlation is perfect, anything near 1 or -1 is pretty good but the closer to 0 the less correlation between the data if the correlation is 0 that means that the data points are independent from each other. This is useful in machine learning because it allows us to predict future data events and if the data we are using is actually able to predict them.