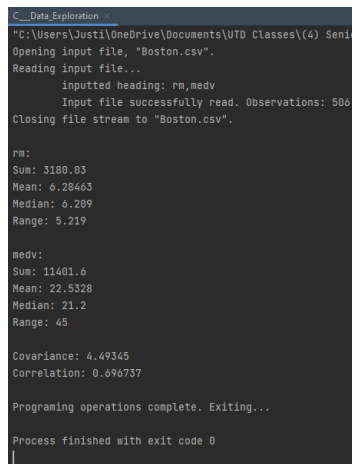


## Data Exploration Writeup



```
C:\Data_Exploration>
"C:\Users\Justi\OneDrive\Documents\UTD Classes\4) Senior\
Opening input file, "Boston.csv".
Reading input file...
  inputted heading: rm,medv
  Input file successfully read. Observations: 506.
Closing file stream to "Boston.csv".

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

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

Covariance: 4.49345
Correlation: 0.696737

Programing operations complete. Exiting...

Process finished with exit code 0
```

After creating this program, it's clear that using built-in R functions was a lot easier than having to write up the actual functions myself. R seems to make statistical computations such as this way easier on the programmer.

Of course, we calculate the mean, median and range. As prompted by the instructions for this assignment, I'll explain what each is. The mean is simply the average of the specific data attributes of an element in the data set. The median is the middle-value of the attributes of the element. And the range is the spread of the data attributes of the element in the data set. Prior to machine learning, each were useful in statistics as they allowed for greater interpretation of data, and analysis of trends in the data.

Covariance and correlation were helpful in interpretation of data as well. In statistics, covariance helped us determine the relationship between two variable's variances. Likewise, correlation helped us determine if their relationship was linearly related. Both are useful in finding out if two variables are related in any way. Furthermore, this is useful in machine learning as the whole point of machine learning is to analyze data to find relationships between data attributes to form accurate predictions.