Ryan Banafshay

Data Exploration in C++

Example of running code with Boston.csv data:

```
Opening 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 = 8.78

Stats for medv

Sum = 11401.6
Mean = 22.5328
Median = 21.2
Range = 50

 Covariance = 4.77386

 Correlation = 0.682338
```

C++ feels counterintuitive when it comes to calculating different statistical values. It is much more convenient to use R's built in functions. R is a language that is designed to provide this type of information so it is not surprising. The different data points we calculated are useful for a variety of different applications. Truthfully I do not know any real use of mean, as the value can easily be manipulated by outliers. Median on the other hand gives the middle value of a sorted data set, which can be used to generalize the average number in the set. Covariance and correlation are probably the most important of the data we calculated. Covariance is the measure of the relationship between two random variables. This can then be used to calculate correlation. Correlation is important as it shows how closely related the two variables are.