Data Exploration

The built-in R functions are very simple and easy to use. Implementing those functions using C++ was relatively straightforward after refreshing myself with C++ syntax and style and using the relevant formulas. The mean of a vector is its average and can be viewed as an expected value for new or predicted values. The median is the middle value of a vector can be seen as an average of sorts that excludes outliers. The range is the minimum and maximum value together and can give a general sense of how spread out the data is. Covariance between two vectors measures how closely related changes between vectors are and, while it can be useful, is often used to calculate correlation. Correlation is covariance scaled to fit between -1 and 1. Values close to the extremes indicate high correlation while values close to zero indicate low correlation.

Program output:

```
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: 6.2085
Range: 3.561 8.78

Stats for medv
Sum: 11401.6
Mean: 22.5328
Median: 21.2
Range: 5 50

Covariance = 4.49345

Correlation = 0.69536

Program terminated.
```