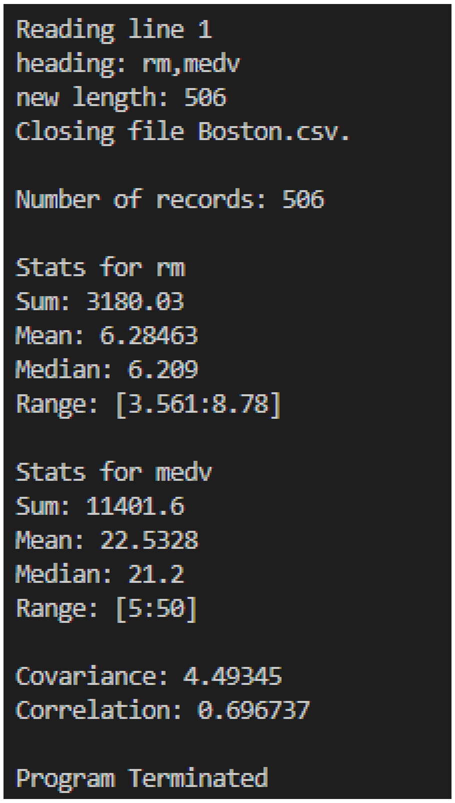
C++ Data Exploration



```
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: [3.561:8.78]

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

Covariance: 4.49345
Correlation: 0.696737

Program Terminated
```

Figure 1: Sample Console Output

This screen capture shows the output of the program when reading the "Boston.csv" file.

It will display the number of records read (observations) and display stats for each variable.

The stats that are displayed are their sum, mean, median, and range.

Since there are only two variables in this specific .csv file, we can figure out their covariance and correlation.

The program will then exit.

My experience writing these functions was tedious, it brought me back to freshman year when we would code functions to only realize there is built-in functionality for it already. Using the built-in functions in R makes it so much easier to read data and pull different information from it. R already has so many different functions you can run with a set of data, so much so that there no point in trying to write them yourself.

Mean is the average of a set of observations, median is the middle value of the set of observation, and range is the minimum and maximum values of a set of observations. These statistical measures can help with data exploration by helping us identify how diverse a dataset is.

Covariance is the measure of how changes in one variable can affect the other variable. Correlation is measured in a range of [-1,1] and shows how negative or positive the correlation between the two variables is. A correlation close to zero indicates that there is little correlation between the two variables. This information can be useful in machine learning because it can help determine if the variables are even worth drawing conclusions from. If two variables have a correlation of -0.005, their observations are most likely not related and should not be compared.