a. copy/paste runs of your code showing the output

```
Opening fileBoston.csv
Reading line1
Closing the file
Number of Records: 506

Stats for rm:
sum:    3180.03
mean:   6.28463
median: 6.209
range:  5.219

Stats for medv:
sum:    11401.6
mean:   22.5328
median: 21.2
range:  45

Covariance between rm and medv = 4.49345
Correlation between rm and medv = 0.69536
Program Terminated.
```

b. describing your experience using built-in functions in R versus coding your own functions in C++

The difference was appalling. The results that can be achieved with a single function call in R needed to be coded out into multiple lines in C++. I would much rather use R for statistical analysis rather than C++.

c. describe the descriptive statistical measures mean, median, and range, and how these values might be useful in data exploration prior to machine learning

Ans. mean is the average over all the values in the given dataset. Median is the middle value of a dataset sorted from minimum to maximum. Range is the difference between maximum value and minimum value of a given dataset. The reason these values are helpful in data exploration is because they help us understand what is the current trend and what can be expected out of it. For example if a buyer wants to buy a house he could check the mean of house price in a certain geographical location and make a choice accordingly. The median would help a buyer determine if the price of a house he is buying lies above or below the 50% . Range would help the buyer determine what was the lowest and highest price in that location.

d. describe the covariance and correlation statistics, and what information they give about two attributes. How might this information be useful in machine learning?

Ans. statistically covariance can be found by using:

$$\text{cov}(x,y) = \frac{\Sigma(x_i - \mu_x)(y_i - \mu_y)}{N-1}$$

Here:

x_i = values in dataset of x

y_i = values in dataset of y

μ_x = mean of x

μ_y = mean of y

N = number of observations

Covariance is a measure that if a covariance between x and y is positive it is a positive relation meaning if x increases so does y and if the covariance between x and y is negative then the relation is negative meaning if x increases then y decreases.

Statistically correlation can be found by using:

$$\text{cor}(x,y) = \frac{cov(x,y)}{\sigma_x \sigma_y}$$

cov(x,y) = covariance between x and y

σ_x = standard deviation in dataset values of x

σ_y = standard deviation in dataset values of y

There would be correlation in $-1 \leq cor(x,y) \leq 1$ so it would be perfectly correlated if the calculated value comes out to be 1 and perfectly negatively correlated if it comes out to be -1