Component-1

1. Pasted output of code section

```
PS C:\Users\Eric\repo\CS-4375-Machine-Learning\Component-1> ./a.exe
Attempting to open default file Boston.csv
Reading line 1
Headings: rm,medv
There are 506 observations in the file

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

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

Covariance = 4.48457

Correlation = 0.69536

Program Terminated
```

2. Description of experience writing the functions for myself in c instead of using the R functions.

The experience was fairly intuitive. Once I understood exactly how correlation and covariance were calculated it wasn't to difficult to write the code for their respective sections. The weirdest part was when I forgot that you needed to sort the vector for median to be able to be properly calculated.

3. Describe mean, range, median, and sum.

Sum is the addition of all observations single feature. Mean is the sum when divided by the number of elements. Range is finding the highest and lowest value and subtracting low from high. Finally median is finding the middle value amongst a set of values. Or the mean of the 2 middle values if there are an even number of values. These are used to see the basic structure of a dataset and get an idea of what type of values you are working with.

4. Describe covariance and correlation.

Covariance is the measure to what degree higher values in one dataset correspond with higher or lower values in another dataset. If higher values in one pertain to higher in another you get a positive covariance. Likewise, if higher values in one correspond with lower values in another you get a negative covariance. If this rule doesn't generally hold true, your covariance will approach 0. The same rules apply for correlation coefficient. The greater or lower the correlation coefficient the more one value can be correlated to another high or low value. The coefficient will always be in the range of -1 to 1. These values are used to determine if two datasets have a relationship in either a positive or negative manner.