

Portfolio Component 1: Data Exploration

a. Code Output:

```
File loaded successfully.  
  
Heading: rm,medv  
New Vector Length: 506  
Number of Records: 506  
  
Stats for RM:  
Sum: 3180.03  
Mean: 6.28463  
Median: 6.209  
Range: 5.219 (3.561 to 8.78)  
  
Stats for MedV:  
Sum: 11401.6  
Mean: 22.5328  
Median: 21.2  
Range: 45 (5 to 50)  
  
Covariance for RM and MedV: 4.49345  
  
Correlation for RM and MedV: 0.696737  
  
We're done here.  
  
Process returned 0 (0x0)   execution time : 0.092 s  
Press any key to continue.
```

b. R vs C++:

Obviously, coding new functions in C++ for these formulas is more difficult and time consuming than using the functions in R that are already built-in. However, I do prefer using the functions in C++ but this is likely due to my familiarity with the language.

c. Statistical measures:

- Mean: This is the average of all elements. The mean can give a general sense of a typical value for a data element.

- Median: This is the middle element when the list of elements is sorted. If the data has outliers then the median can be used to gain a better understanding of the data than just the mean alone.
- Range: This is the difference between the upper and lower bounds of the data. Paired with median and mean it can be used to understand how the data is spread across the elements.

d. Covariance and Correlation

- Covariance is the measure of how each variable changes and the degree to which a change in one variable will correspond to a change in the other. Covariance can be useful in that if two variables have a strong covariance we can predict how a variable will change based on the observed change in the other variable.
- Correlation is the same as covariance but scaled down to $[-1,1]$, where 1 means positive correlation, -1 means negative correlation, and the closer the value is to zero the weaker the correlation between the two variables. Correlation serves the same purpose as covariance but is reduced to a simple coefficient.