The use of R for statistical analysis is very powerful. Its many built-in functions allow the user to easily request information. Coding just some of these functions in another language like C++ requires a little higher understanding of statistics and the formulas or methods used to get certain values. Some of these functions like mean, median and range are fairly simple as they are used to find general information about one attribute. This can be a measure where the center of the data is (mean and median) or a measure of the spread of data (range). These values can be used to find the “expected” value or, for in ML, what something *should* be like. More difficult functions are ones like covariance or correlation which are based on 2 attributes and checking how one may be connected to other. Correlation shows how 2 attributes are related (often times checked as a linear relationship). If the attributes are strongly related, the closer the correlation will be to 1 or -1 dependent on the relationship. If they are weakly related, the closer the correlation will be to 0. Covariance is used to see how 2 attributes are dependent on each other. This shows how a change in 1 attribute will change the other one. These functions are helpful for ML because it allows a computer to find relationships between things that humans might not see instantly or allow them to predict certain behaviors based on previous data.

Here is an example run of some of these functions using a program I made in C++ on the Boston.csv file.

```
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.
```