From the C++ Data Exploration program:

```
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: 6.209
Range: 5.219

Stats for medv:
Sum: 11401.6
Mean: 22.5328
Median: 21.2
Range: 45

 Covariance = 4.49345

 Correlation = 0.69536

Program terminated.
```

R vs. C++

Creating the functions in C++ is a lot more tedious than simply entering in the built-in functions in R. For one, by creating the functions yourself, you're more likely to make mistakes in your calculations or create bugs that are common in a low-level language like C++. In fact, I had to use RStudio to double check my outputs and make sure they were correct (which, on my first attempt, they were not). My program in C++ has 158 lines while recreating the same program in R would've taken less than 15 lines. Without a doubt, R is far more convenient language to use than C++ for data scientists.

Mean, Median, and Range

Mean is the average value of the dataset, which is found from dividing the sum of the values by the number of values. Median is the value in the middle of the dataset. If there are an odd number of values, then you'll take the average of the middle two values. Both mean and median are useful in data exploration because it can give you an idea of where the center of the data lies. Range is the difference between the highest and the lowest values of the dataset. The range of a dataset can give you an idea of which values are in the dataset. It can also show you how far the outliers are away from each other

Correlation and Covariance

Correlation measures the linear relationship between two variables and is shown at the scale of -1 to +1. -1 demonstrates a negative correlation, +1 demonstrates a positive correlation, and 0 demonstrates no correlation at all. Covariance performs very similarly to correlation minus the fact that it isn't on the [-1,1] scale. Both correlation and covariance show how changing one variable may affect a second variable. This is useful in algorithms such as linear regression, where you'll want to know how related two variables are. You can also use them to test out how well your predicted data measures up to real test data.