Data Exploration

```
fireb@DESKTOP-TVV6QUH MINGW64 ~/Desktop/GitHub Code/ML_Portfolio/ML_Portfolio/Data_Exploration (main)
$ ./outfile
Opening csv..
Header: rm,medv
Size: 506
Closing csv..
-----
Sum: 3180.03
Mean: 6.28463
Median: 9.3125
Range: 5.219
-----
Sum: 11401.6
Mean: 22.5328
Median: 31.8
Range: 45
=====
Covariance: 4.49345
Correlation: 0.69536
```

I think the built-in functions of R were very easy to learn and implement. Compared to writing these functions in C++, R offers lots of useful operations that are very common in analyzing data sets. The four basic statistical measures of sum(), mean(), median(), and range() have been easy and intuitive to use in R. It may be the lack of use in recent times, but I found programming the functions in C++ manually to be a chore, and I'm very grateful to have everything ready for me in R right from the beginning. I had some troubles interpreting the covariance and correlation coefficient computations, but I've managed somehow.

Mean is the average of all number from a set. Median is the "central tendency," and shows the midpoint value of a set. Range is the difference between the greatest datapoint value and the smallest datapoint value. These measurements can be used to generalize datasets, such as in overall trends in the distribution.

Covariance measures how related two variables are. A positive covariance indicates that the two variables tend to increase together, while a negative covariance indicates one decreasing as the other increases. Correlation measures how strong the relationship between two variables are as they change values. It ranges from -1 to 1, where -1 shows a perfect negative relationship, 0 shows virtually no relationship, and 1 shows a perfect positive relationship. These two measurements are important for measuring association between variables. Observing strong correlation, for example, may indicate we can predict one variable using the other. Machine learning utilizes this, as it calls for finding patterns that models can predict in the future beyond a given dataset.