Output:

Text

Description automatically generatedOpening file Boston.csv

For rm:

Sum of rm is: 3180.03

Mean of rm is: 6.28463

Median of rm is: 6.2085

Range of rm is: 5.219

For medv:

Sum of rm is: 11401.6

Mean of rm is: 22.5328

Median of rm is: 21.2

Range of rm is: 45

Covariance: 4.48457

Correlation: 0.693986

Mean is the average out of a given set. This can help in making estimations as well as setting a base for a certain criterion for further scrutiny. This base can tell us if a child is over or under the average height for that age group for example. Median is the number in the middle of the sorted set and can be good for comparing with the mean to see if the data set is skewed more one way versus another. The range is the difference between the highest and lowest number in the dataset and is used to show how expansive a data set is. These can all be used to create decent estimations for future entities in the dataset as well as understand the current one better.

Covariance tells us how two variables or datasets differ from one another while correlation shows if they are related. Basically, the covariance determines if one variable is growing alongside the other, getting smaller while the other grows or vice versa, and the rate at which this is happening. Correlation tells us if one influences the other or if they are likely influenced by the same dataset or variable. These can be incredibly powerful and useful when analyzing large datasets and trying to gleam as much information from them as possible.

Coding my own functions in C++ was a little more tedious than using the built-in functions with R. It’s much easier to give R a dataset and use the functions given than to code your own functions for each file.