Aloksai Choudari

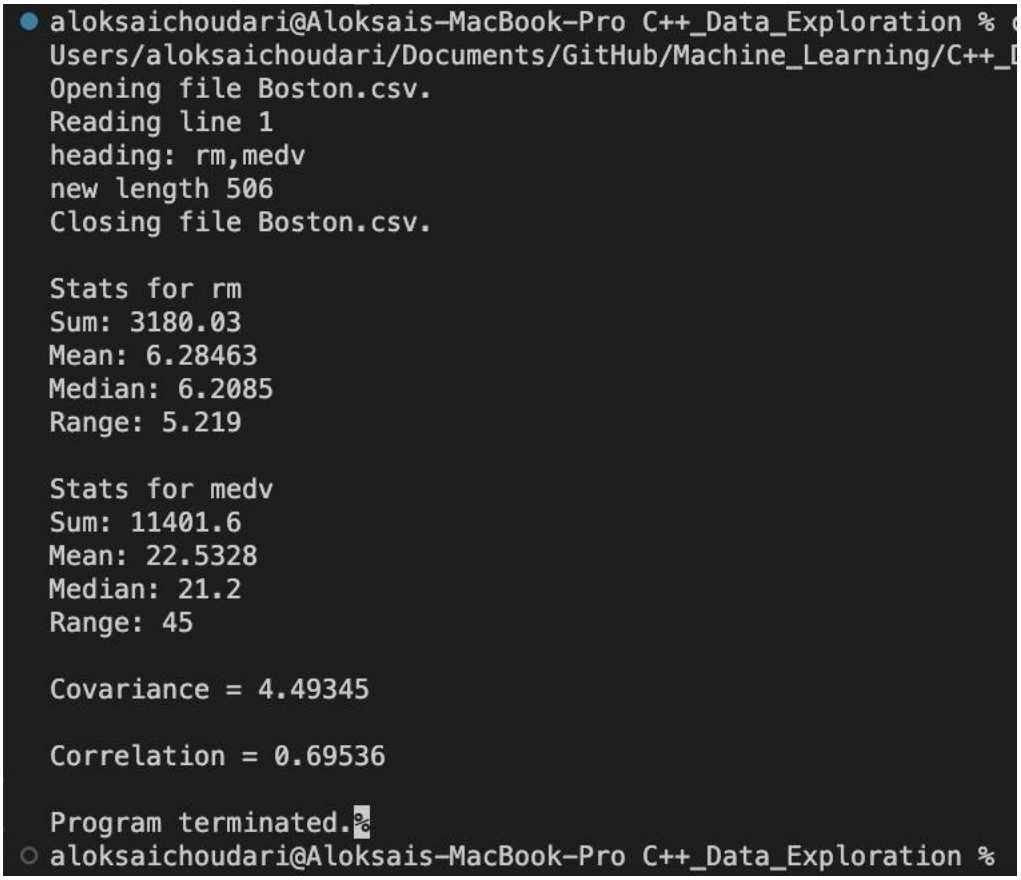
Dr. Karen Mazidi

CS 4375.003

11 September 2022

Portfolio 1: C++ Data Exploration

a.



b. Although coding the functions for sum, mean, median, range, covariance, and correlation in C++ was a great learning experience, using the built-in functions in R is much easier and faster to get data statistics. Writing the functions for sum, mean, median, and range weren't too tedious because they require basic knowledge of math, but creating functions to match the formulas of covariance and correlation were a little bit trickier. Doing this for more formulaic functions would definitely be a time-consuming task, which makes the built-in functions in R far more efficient.
c. Mean, median, and range are measures of tendency, which are useful for summarizing current data sets and predicting future data sets as well. Mean is the average of all of

the numbers in the data set. Median is the middle number in the data set, after it has been organized in ascending order. Finally, the range is the largest number of the data set minus the smallest number in the data set. As mentioned before, these statistics can be used for future predictions, along with measures for graphical use to construct visual representations, such as graphs, charts, models, etc.

d. Covariance represents how two variables in a data set change together, or how they differ, while correlation shows how two variables in a data set coordinate with each other, or how they are related. This information is useful in machine learning because it gives insight on regression and statistics that represent models and can predict future outcomes for various objectives, such as house prices, car prices, trends, etc.