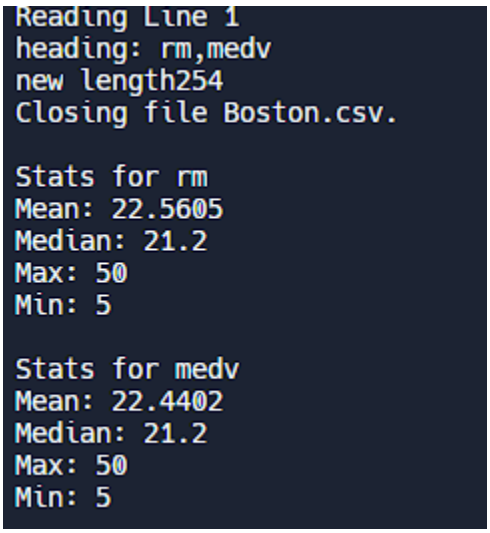
Data Exploration Report

A. copy/paste runs of your code showing the output

a.



B. describing your experience using built-in functions in R versus coding your own functions in C++

a. When comparing and contrasting the built-in functions in R compared to that of C++ I can definitely say that with R it is way easier and less syntax. C++ has some functions that are similar to R but they require more work to setup.

C. describe the descriptive statistical measures mean, median, and range, and how these values might be useful in data exploration prior to machine learning.

a. Three commonly used descriptive statistics in data exploration are mean, median, and range. These measures provide insight into the central tendency and spread of the data. By examining these values, analysts can identify potential outliers or skewness that could impact machine learning results.

D. describe the covariance and correlation statistics, and what information they give about two attributes. How might this information be useful in machine learning?

a. Covariance and correlation are two measures of the relationship between two attributes in a dataset. Covariance shows the linear relationship between two variables, while correlation indicates the strength and direction of that relationship. Knowing the relationship between the target variable and the independent variables can help in the selection of relevant features for a machine learning model.