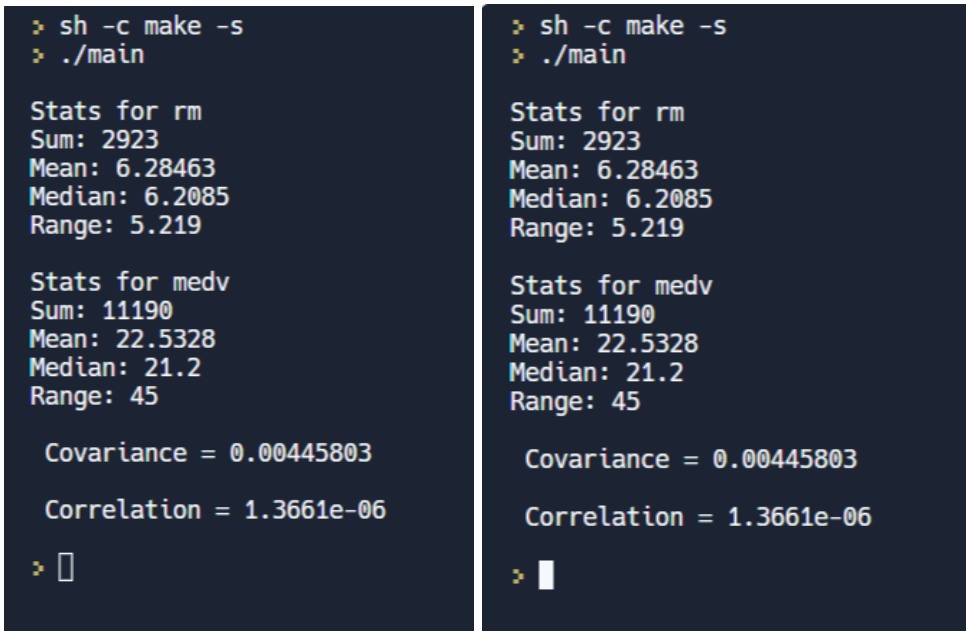
Source Code Outputs



Explain your experience using R's built-in functions versus writing your own C++ functions

I prefer to use built-in functions. It can get quite complicated to work with the logic of the program as well as attempt to incorporate the math component of it. I had a difficult time using the language tools (c++ features) to mimic the mathematical formulas.

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

The mean, median, and mode are measures of statistical distributions.

- Mean is the average of the data set.
- Median is the middle of the set.
- Range is the difference between the largest and smallest numeric values of the data set.

This type of information can be useful when working with a huge data set and might be beneficial to represent the entire set with a single value that describes the entire set. It can also be useful for algorithms because most algorithms are based on statistical models, linear models, and so on.

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

Covariance measures how much two random variables vary together, whereas correlation measures when a change in one variable can cause a change in another. The direction of the

linear relationship between attributes is indicated by covariance. Correlation assesses the strength and direction of a two-attribute linear connection. In machine learning, covariance is used to assess how closely two random variables are related, whereas correlation is used to discover if a change in one variable affects another one.