Example output:

rm mdev

sum 3180.03 11401.6

mean 6.28463 22.5328

median 6.209 21.2

range 5.219 45

covariance 4.49345

correlation 0.696737

Creating these outputs in C++ took exponentially more time than it did in R since R has these functions prebuilt. One problem I had when programming in C++ was that I had to make temporary vectors to find the range and median because I needed to sort them but that would mess with covariance and correlation result.

The mean measures the average value, it gives us a center point to make a well-educated guess on the data. The median gives us the center value and can help us understand if the majority of values are above or below the average. The range gives us the scope of how far apart we can expect the data to be, the larger the range the harder it is to predict a value.

Both correlation and covariance are the measure of how related the 2 data sets are from one another. Correlation is the covariance scaled to be between -1 and 1 but for both cases the further away from zero they are the more related the data sets will be. This is very useful for machine learning because we can tell if data sets are related before we even train our machine.