1. **copy/paste runs of your code showing the output**

电脑萤幕的截图

描述已自动生成

**b. describing your experience using built-in functions in R versus coding your own**

**functions in C++**

In R, data import, cleaning, and various functions are very complete, and you can quickly obtain relevant information. In C++, before importing the data, we had to spend a lot of time analyzing the various data, and we had to pay attention to whether there were any programming errors. So in terms of ML, R is much better than C++

**c. describe the descriptive statistical measures mean, median, and range, and how these**

**values might be useful in data exploration prior to machine learning**

**M**ean is the average of all data. It sums all the measurements, and divides by the number of measurements. Median is the number at half of measurements. Range is the measure from the smallest measurement to the largest measurement. In data exploration prior to machine learning, mean and median can measuring central tendency, and median can measure the centrality better than mean if there are skewed in measurements. Range is the basic measure of statistical dispersion.

**d. describe the covariance and correlation statistics, and what information they give about**

**two attributes. How might this information be useful in machine learning?**

Covariance is a measure of the joint variability of two random variables. It can give the overall error of the two variables. In ML, it evaluate the effects of their corresponding variables due to their correlation.

Correlation is a statistical measure that expresses the extent to which two variables are linearly related. In ML, it used to measure the distance between two variables relative to their independence