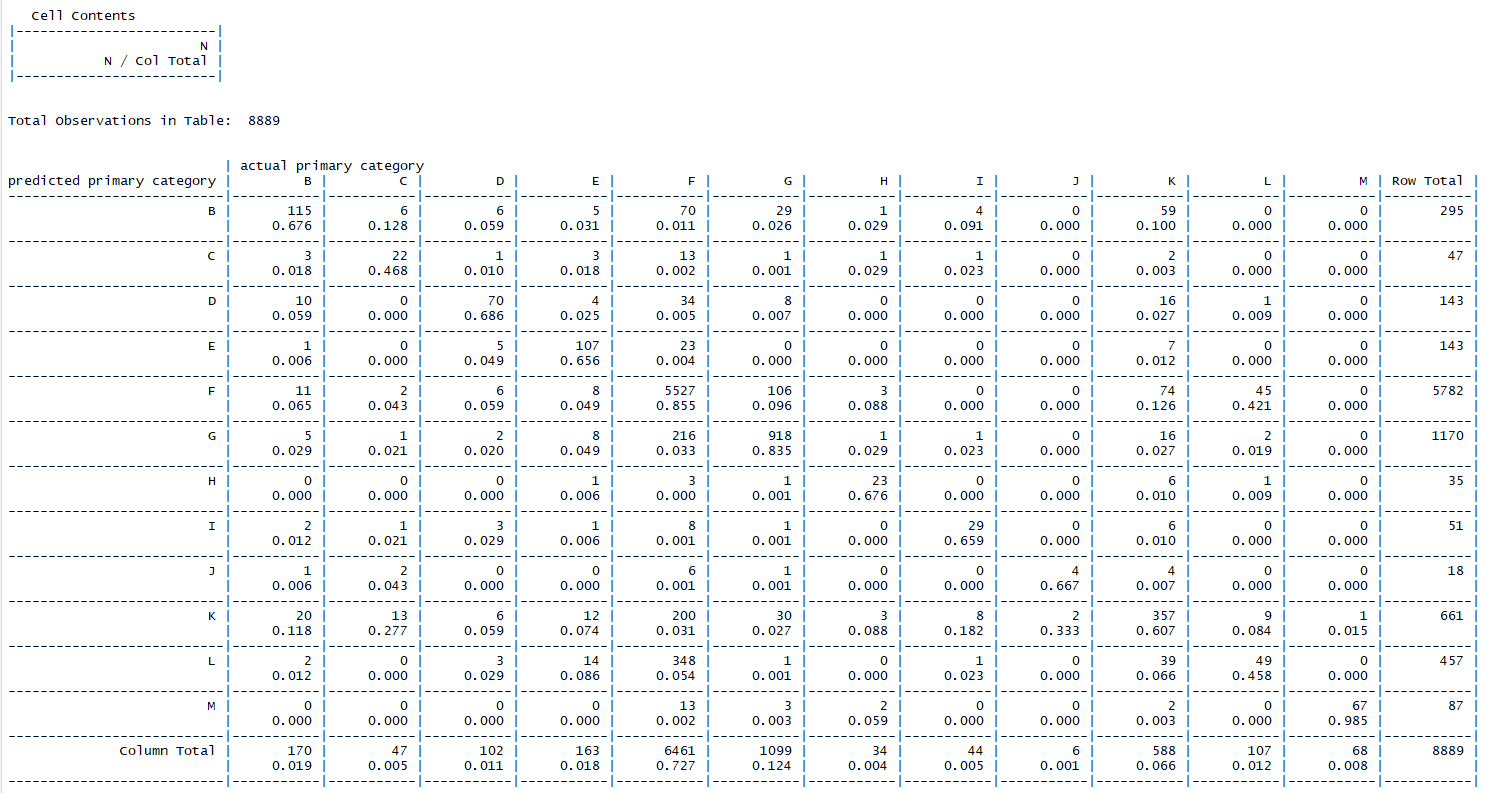
**Module 5**

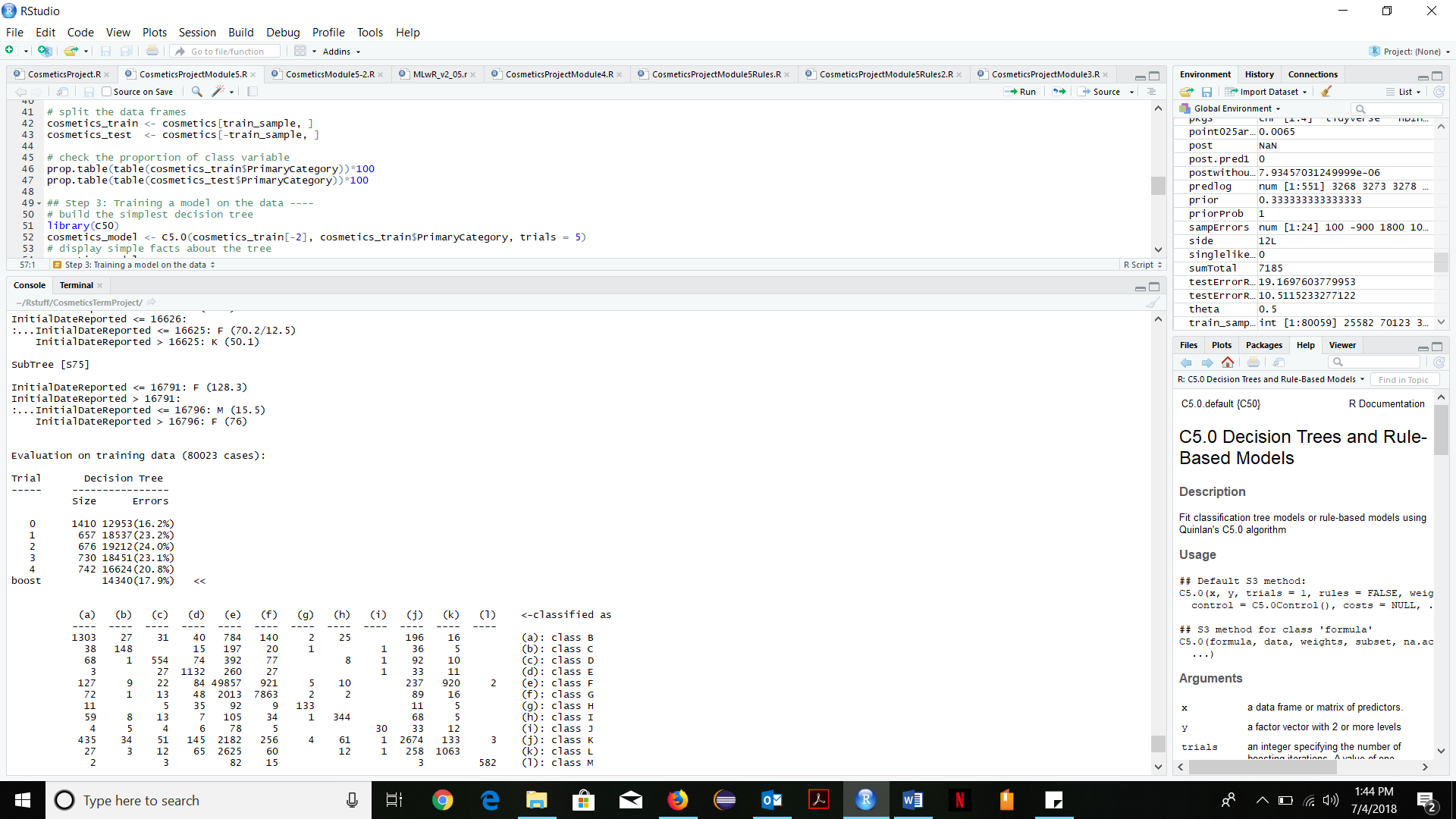
For both, decisiosn trees and rule learning, I used 90% of the data as training data and 10% as testing data. I randomized the data, setting the seed to 123 for repeatable results. The proportions of all levels of the target factor in the original, training and testing dataset were about equal. I removed one of the levels from the PrimaryCategory target factor because there was too small a percentage of it in the data and it so happened that there was either zero in the training data or zero in the test data.

**Decision Tree**

I predicted the PrimaryCategory type of the cosmetic product based on the factor: IsDiscontinued and the numerical varaibles: InitialDateReported, MostRecentDateReported, ChemicalCreatedAt, ChemicalUpdatedAt, ChemicalDateRemoved, and ChemicalCount. ****

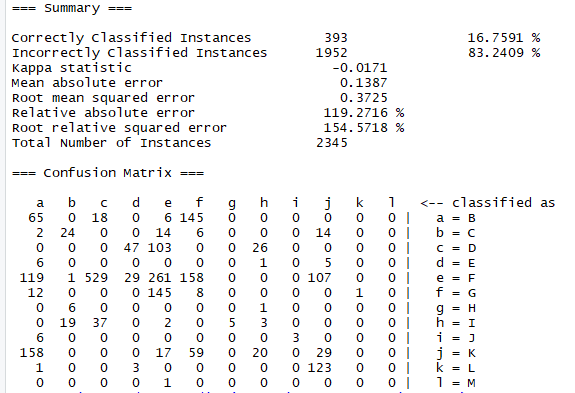
The training data had an error rate of 16.2% while the test data had a close error rate of 18.0%.

Quite interestingly when I later set the trials to 5, to boost performance, I had a higher trial error rate (17.9%) and higher test error rate (19.2%).



**Rule Learning**

**1R**: I had an 83% error rate in the training data and an unpredictable 0.42% error rate in the testing data.

****

**JRip:** there was a 72% error rate in the training data and again an unpredictable 2.2% error rate in the testing data.

