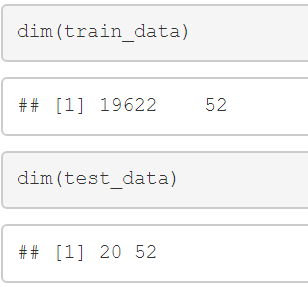
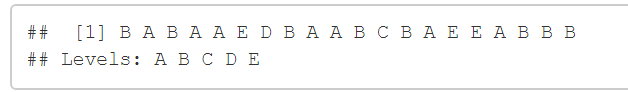


The original training and testing data have 160 variables. I removed columns with NA entries and removed near zero value variable's , which brought the number of variables down to 59. I then removed 7 additional variables which contained information that I deemed not useful: X , User\_names, raw\_timestamp\_part\_1, raw\_timestamp\_part\_2, cvtd\_timestamp, new\_window, num\_window. (Prior to removing these variables, I was achieving perfect accuracy on my training and validation sets, but my model was predicting all of the test cases to be of classe A.)



From here, I split the training data into two sets: “train\_data” for training the model (60%) and “test\_data” for testing of the model (40%). I trained a random forest on “train\_data” using the default parameters.I chose Decision Tree model to train my model and to analyze the accuracy.Then I chose a random forest model because they tend to be very accurate and the data set was small enough that using a random forest was feasible.

I predicted the classes on “train\_data” and found that the accuracy was 100%. I then used this model to predict the values on the “test\_data” set and found the accuracy to be 98.9%.



Like this I get best level of accuracy for data and submitted my answers, and it correctly identified all 20 cases.