Practical Machine Learning Course Project

First, I split the training set into 2 parts, 80% going to training and 20% to validating. This was done using the createDataPartition function and so this was a random division based on the “classe” variable

Then I trained rf models to start. I de-selected any columns that had NA, since many of them had 15000+ NA’s out of under 16000 entries, so rather than work out the headache of cleaning the NA’s I omitted them.  
  
I then used a subset of the remaining columns, partially out of a mistake because I didn’t realize how many columns were clear of NAs. The resultant RF model had a 82% accuracy when validated against the previously mentioned validation subset.  
  
Next, I built a model using 53 columns that have no NAs. This model reached 99% accuracy, which leads me to feel that there is some over-fitting going on. I then calculated the MeanDecreasedGini to find out which variables that the model was most sensitive to change. I then selected the top 10 of these to fit another model.

This model also received 99% accuracy. In an effort to reduce possible over fitting, I deselected the lower 5 variables.

This model received