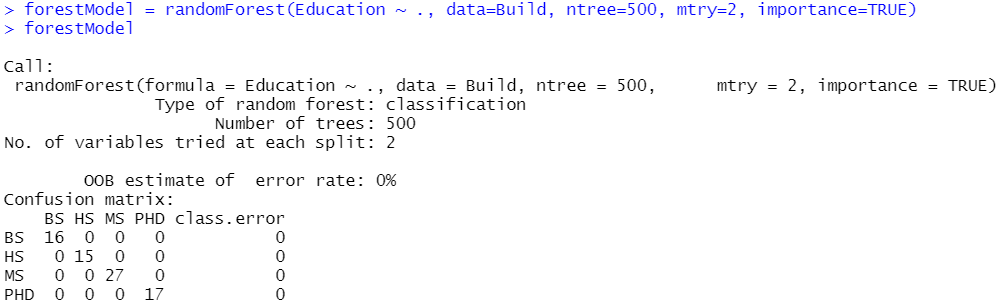
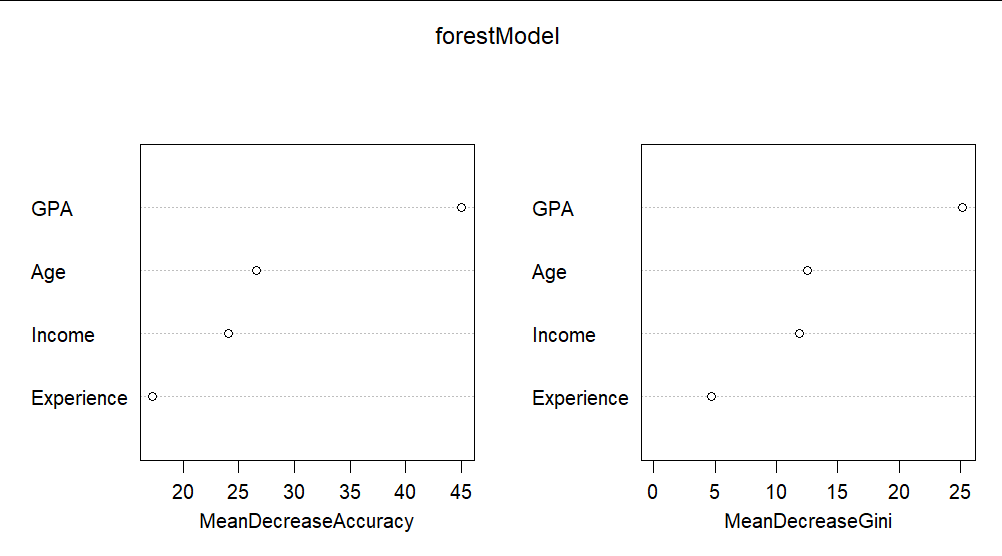
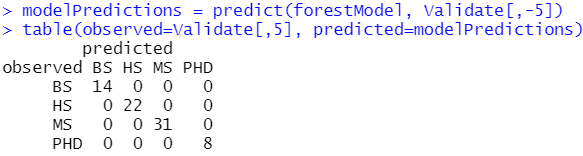


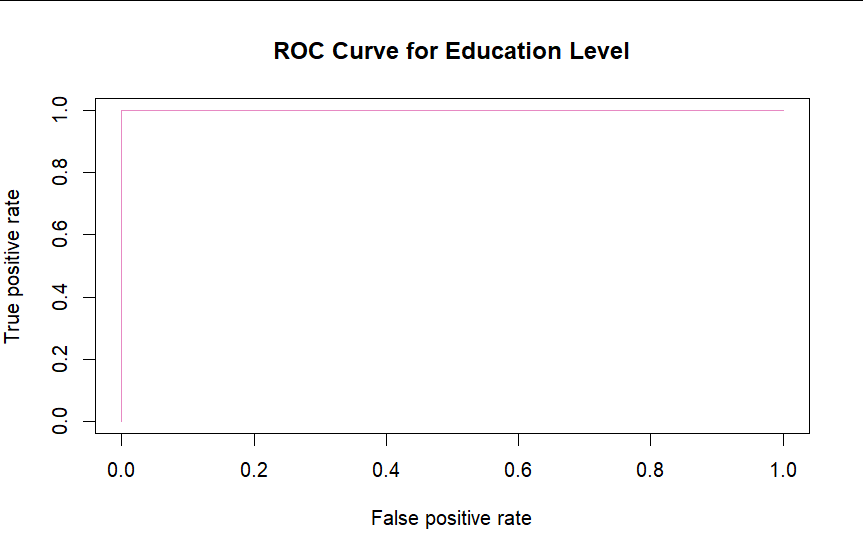
Starting off, we had to use bootstrapping to increase the number of samples in the data set from 21 samples to 150 samples. The number of samples needed was determined by eNf, or e4 = 54.598, in which 75 > 54.598 so we can use 75 samples for the build and validate data sets, or 150 samples total.







Then we used the generated build and validate data sets to create a random forest model for predicting education levels based on the generated values for the other four variables (Experience, GPA, Age, Income). The forest model had a OOB error rate of 0%, meaning the model was 100% correct. We further verified this by looking at the importance our model assigned for each of the variables, and through testing if the model could predict the validate data set, in which the forest model was 100% correct on predicting the validate data set.



When creating an ROC curve graph for the forest model, it turns out like the graph above because the model is perfect for all variables.



