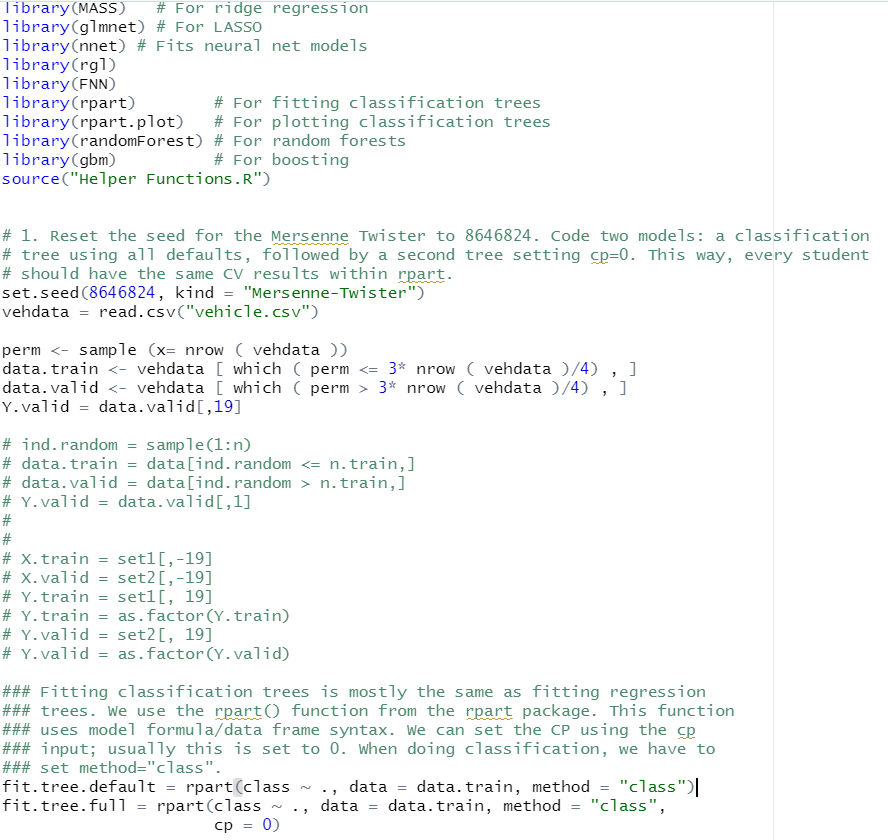
We will fit single trees and random forests to the training data and test them at the end

using the test set.

1. Reset the seed for the Mersenne Twister to 8646824. Code two models: a classification

tree using all defaults, followed by a second tree setting cp=0. This way, every student

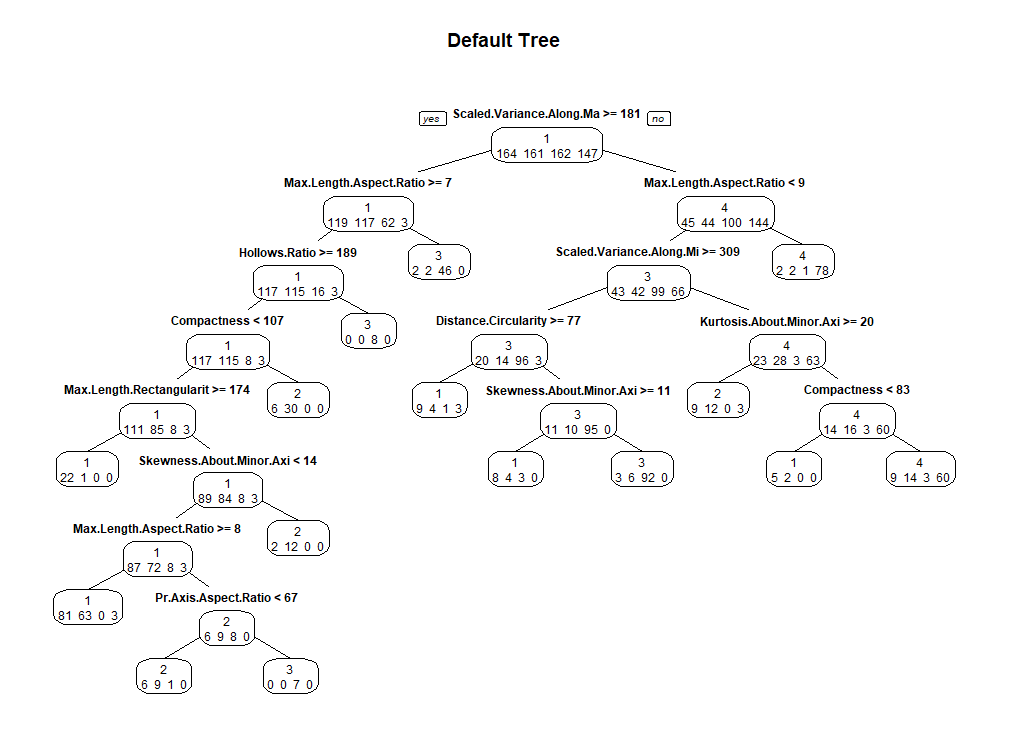
should have the same CV results within rpart.



(a) **Print out the CP table for the default tree.** Is there any evidence that larger trees are needed than what the defaults allow? **Explain.**

We will no longer use this run. All remaining questions refer to the cp=0 object.



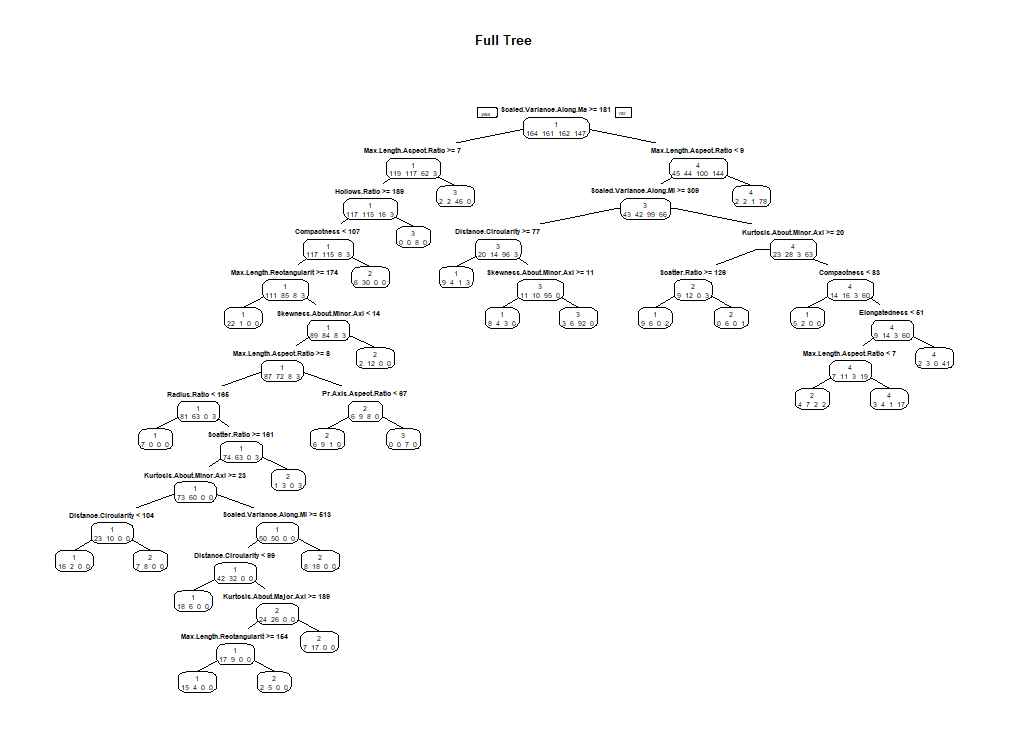


-> It is showing that some nodes are not separated even though they have many number of different classes. For example, the child node of compactness < 107, It has 6 class 1 and 30 class 2.

(b) **Print out the CP table for the tree with cp=0**. **How many splits are**

**suggested as best?**

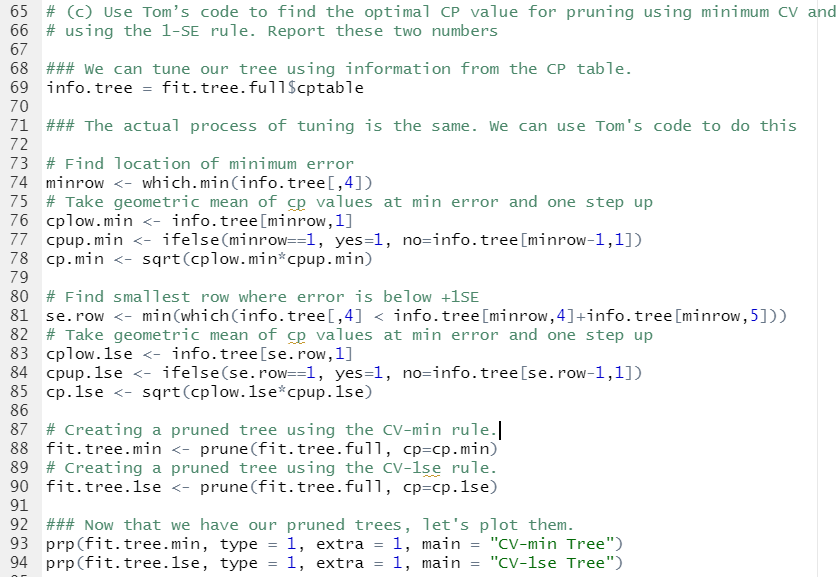




The graph is showing that 25 splits are good.

(c) Use Tom’s code to find the optimal CP value for pruning using minimum CV and

using the 1-SE rule. **Report these two numbers**

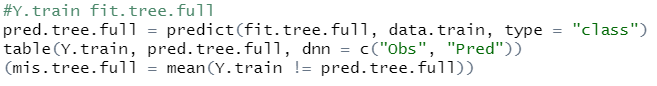




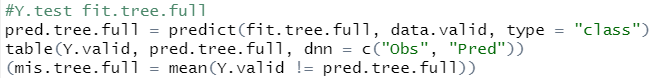


(d) **Report training and test error of all three of these trees, full and two**

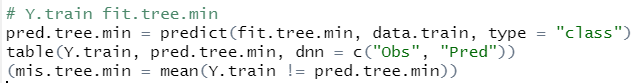
**pruned.**



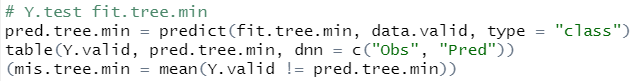




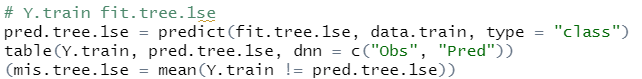




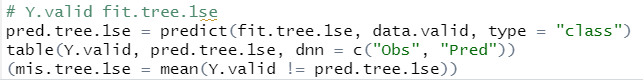














(e) **How does test error compare to other models?**

-> The test error doesn’t show the lowest value (<0.2) but it shows better result than kernel density.