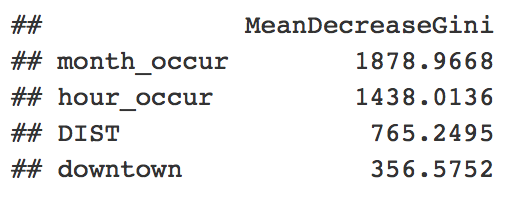
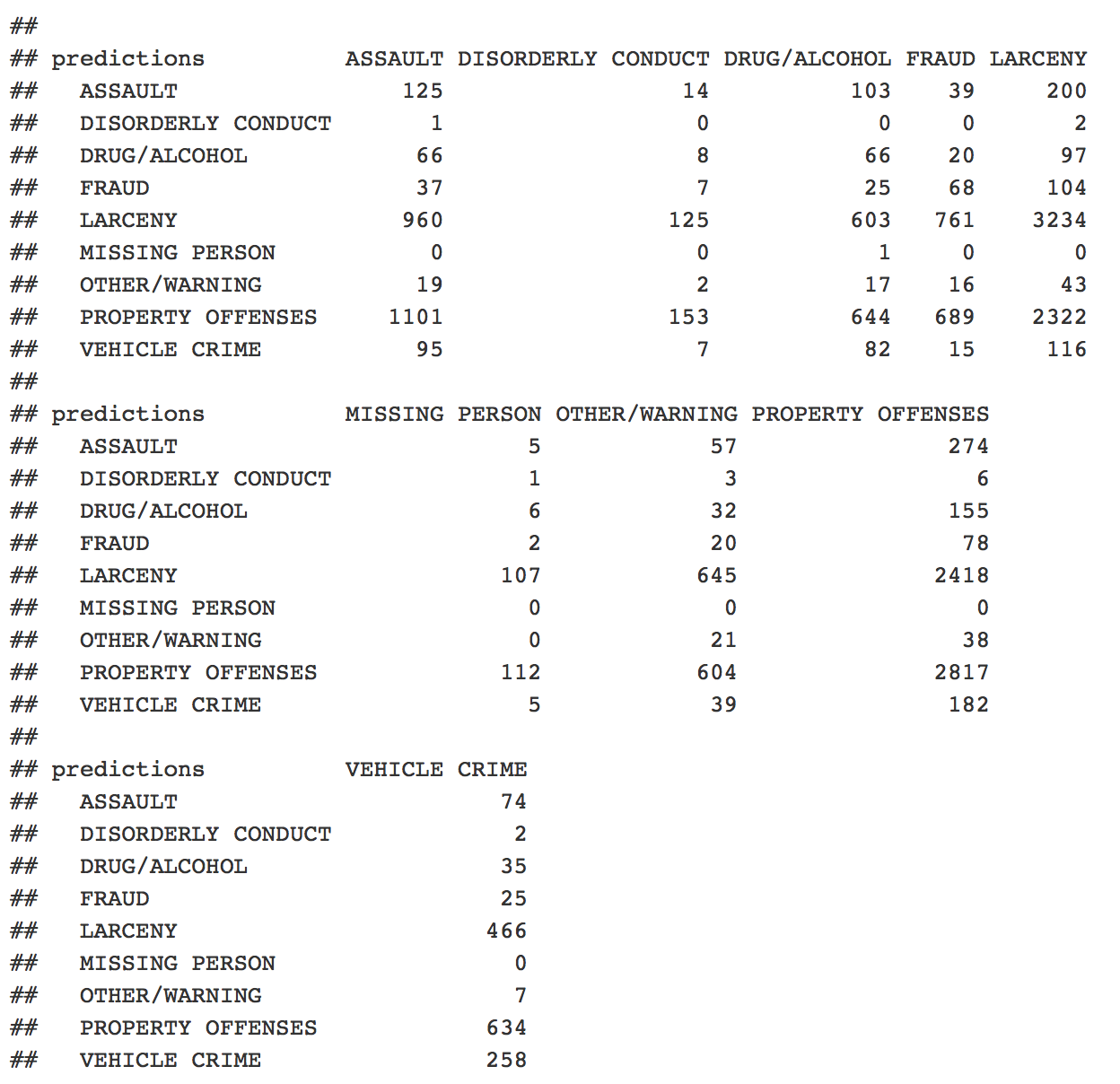
Performing random forest classifying charge:



* Month and hour of occurence have a much higher influence on charge than district and whether or not it occured downtown

a

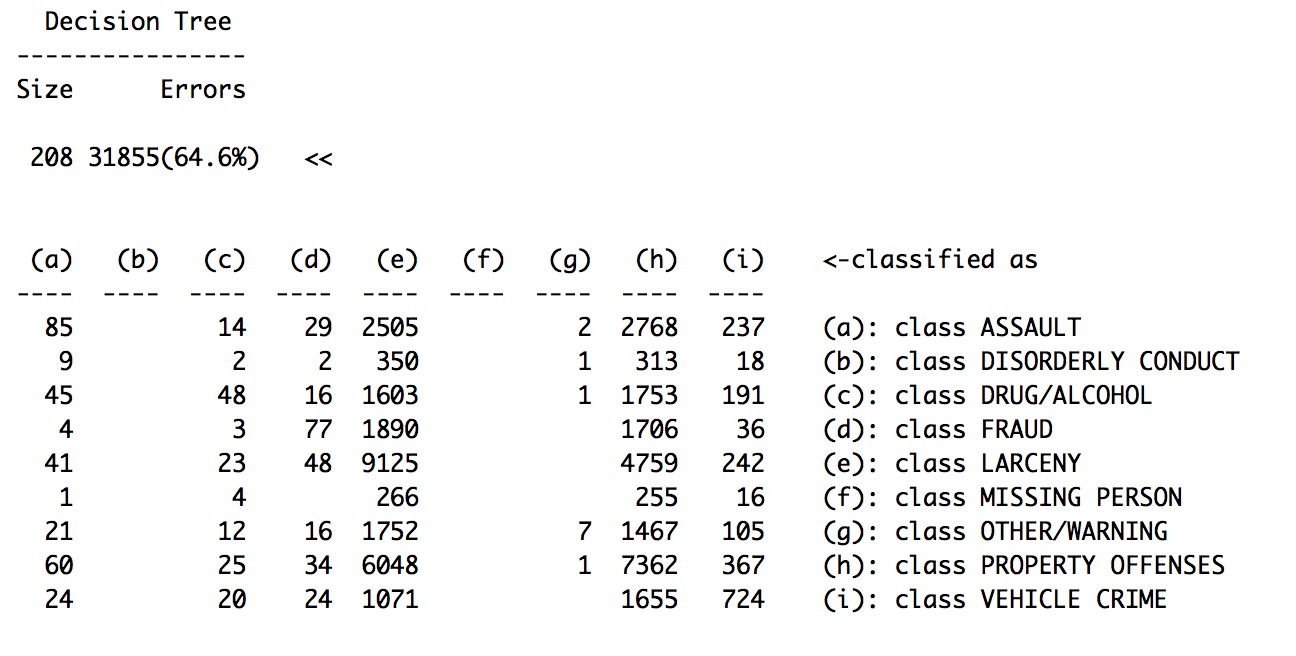
Performing bagging classifying charge:





l

Using C5.0 classifying charge:



C5.0 is an improvement on C4.5. At each node of the tree, C4.5 chooses the attribute of the data that most effectively splits its set of samples into subsets enriched in one class or the other. The splitting criterion is the normalized [information gain](https://en.wikipedia.org/wiki/Information_gain_in_decision_trees) (difference in [entropy](https://en.wikipedia.org/wiki/Entropy_(information_theory))). The attribute with the highest normalized information gain is chosen to make the decision.

----------------------------------------

Using classification trees to predict downtown without using classifiers like DIST, BIG\_ZONE, or area is basically useless and the tree doesn’t grow past one node: independent variables do not provide enough information to grow the tree.