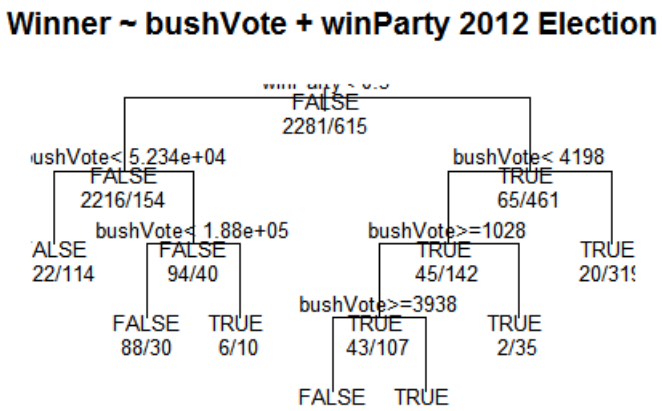
HW 8 - Classification Trees with Recursive Partitioning

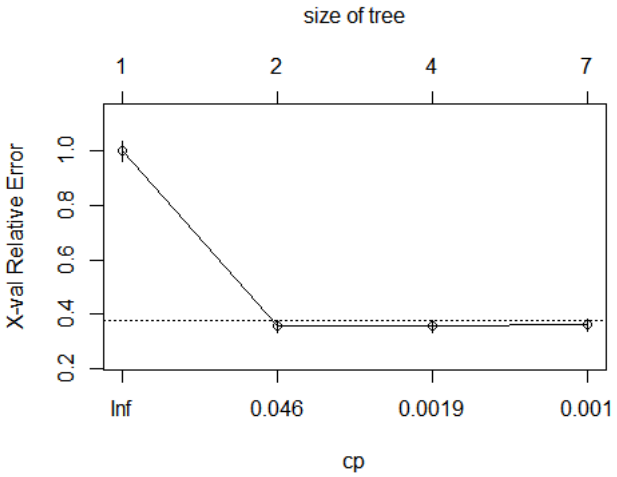
Shiwei Chen

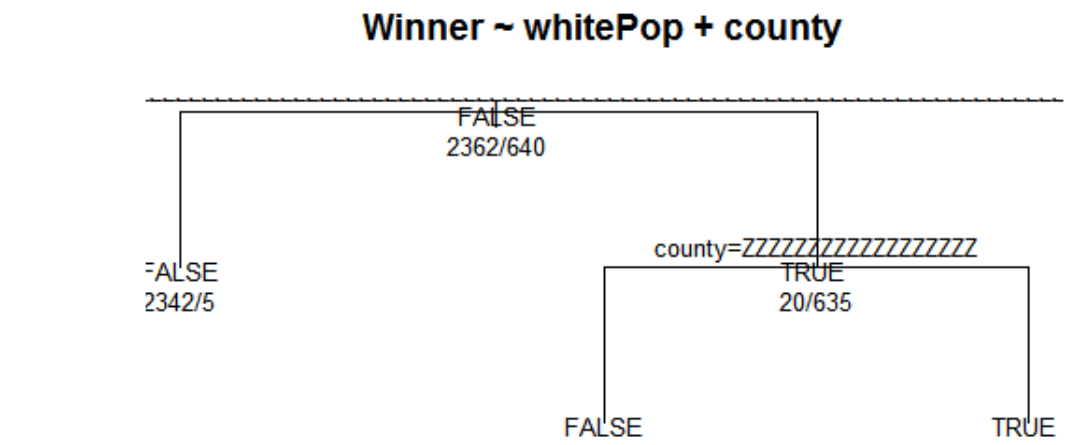
December 6, 2014

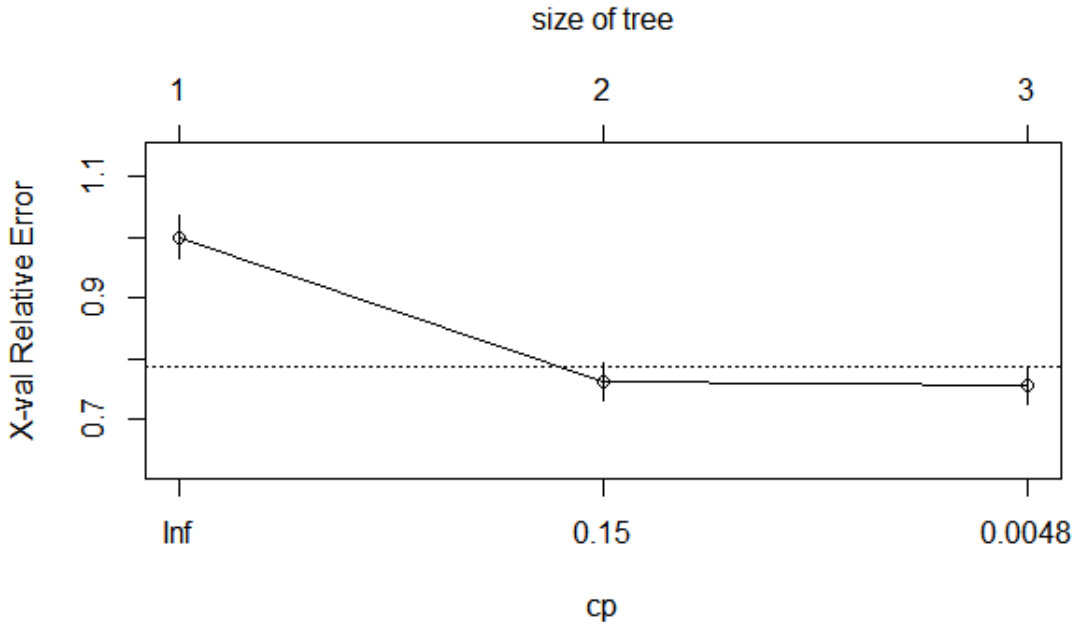
1 Recursive Partitioning with Rpart

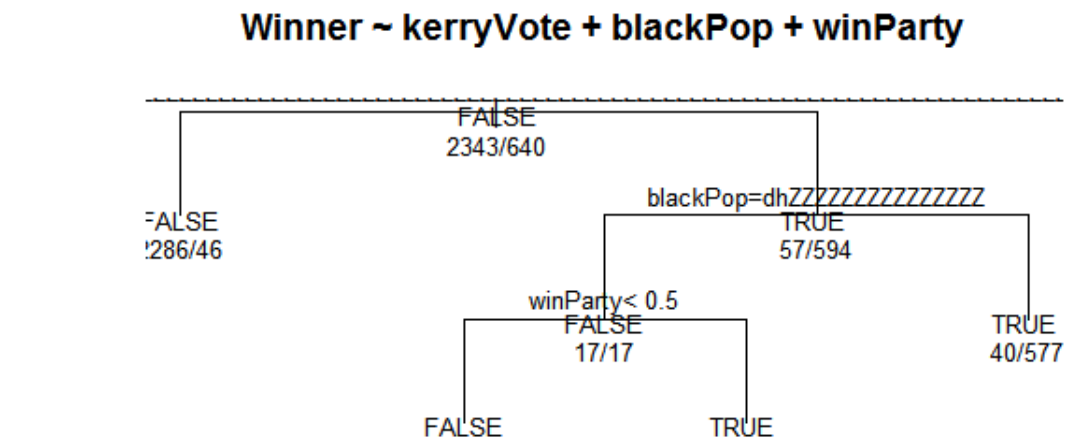
I played around with 4 different formulas to use for predictors to predict the winner of the 2012 election, but my best results would never go less than 0.4 relative error during the cross validation. This makes me wonder, if there is a combination of predictors out there that will significantly reduce relative error. The worst formula to predict winner was using county and whitePop only, which gave relative cross_validation error of 0.7. Also, it was interesting to see in the third classification tree that blackPop was a better predicator than winParty because it was chosen to split higher on the tree. Race in conjunction to other factors like socio-economic class and age can often be a good predication of vote.

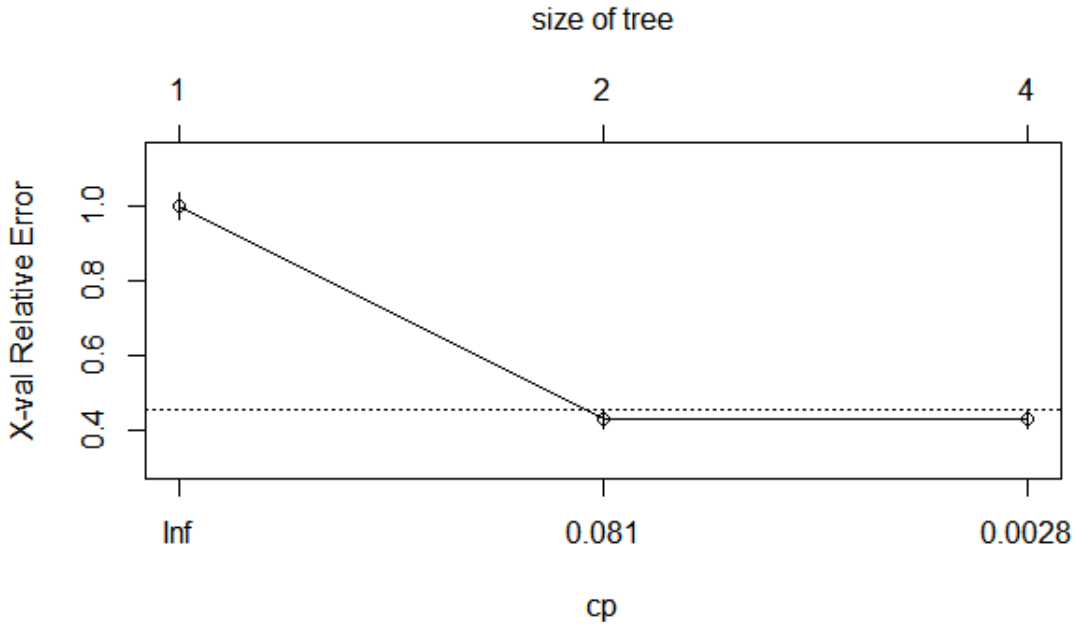


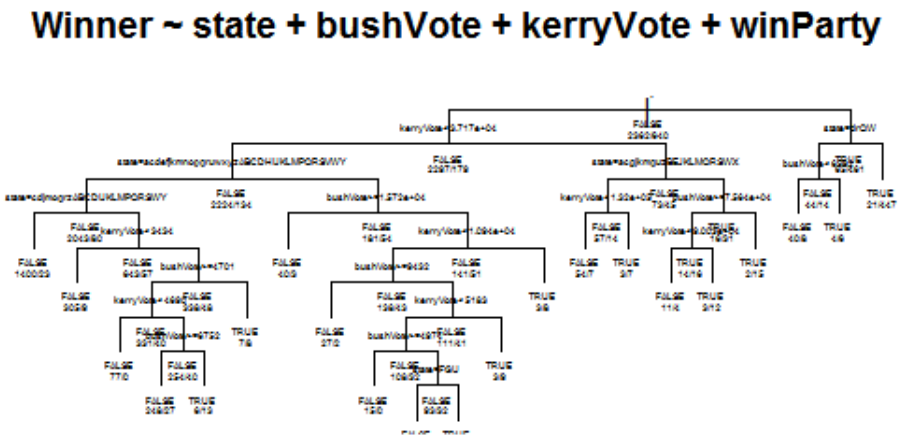



Winner ~ whitePop + county
FALSE
2362/640
county=ZZZZZZZZZZZZZZZZZ
FALSE
2342/5
TRUE
20/635
FALSE
TRUE


size of tree
1
2
3
X-val Relative Error
0.7
0.9
1.1
Inf
0.15
0.0048
cp


Winner ~ kerryVote + blackPop + winParty
FALSE
2343/640
FALSE
1286/46
blackPop=dhZZZZZZZZZZZZZZZ
TRUE
57/594
winParty< 0.5
FALSE
17/17
TRUE
40/577
FALSE
TRUE


size of tree
1
2
4
X-val Relative Error
0.4
0.6
0.8
1.0
Inf
0.081
0.0028
cp


Winner ~ state + bushVote + kerryVote + winParty

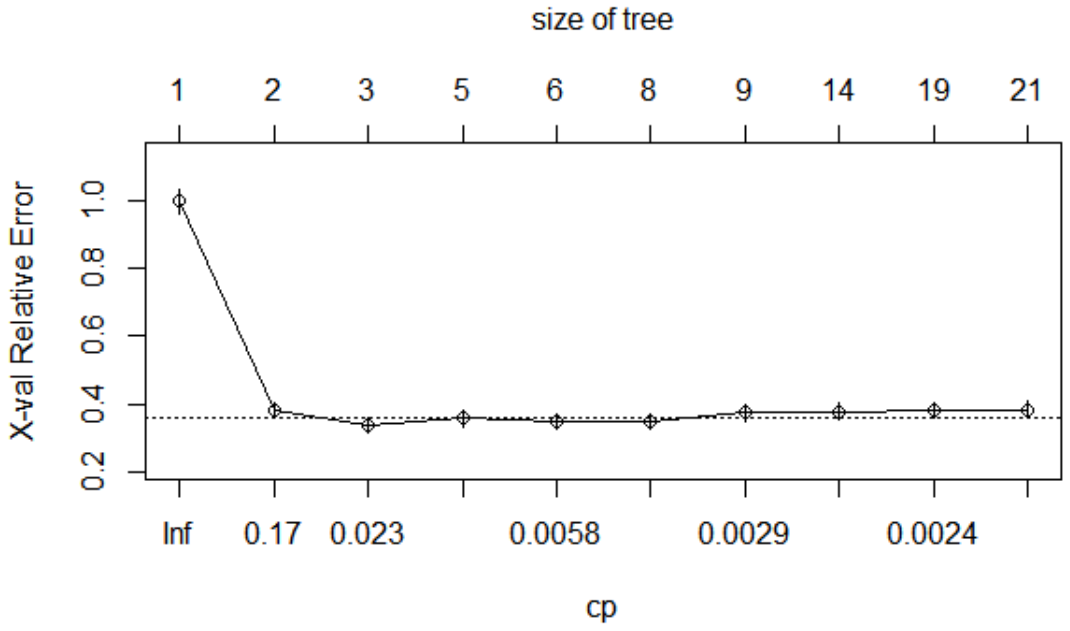

size of tree
1
2
3
5
6
8
9
14
19
21
X-val Relative Error
0.2
0.4
0.6
0.8
1.0
Inf
0.17
0.023
0.0058
0.0029
0.0024
cp

Figure 1.1:

2 VOTE SHIFT

I generated two plots to calculate vote shift because I did not exactly know what defined vote shift. The first plot interprets vote shift by coloring arrows based on party of 2012 election winner and offsetting the arrow by the change in number of voters per county for the winning party. This led to some arrows pointing the wrong way, but a correct color spread similar to the plot in the given pdf.

The second plot was colored based on the party which had greater shift in voters. If the difference in number of repVoters and bushVoters was greater than the difference of demVoters and kerryVoters the arrow for that county would be colored red. The offset is proportional to the difference between the deltas. This made the coloring look incorrect, the arrows pointed in the correct locations. Perhaps if I incorporated both these techniques, it would look "right", but the data it is representing would be incorrect. Also, I was disatisfied with the resolution of image produced by the map function and in the future, I would probably use the ggplot2 package to produce geographical plots.

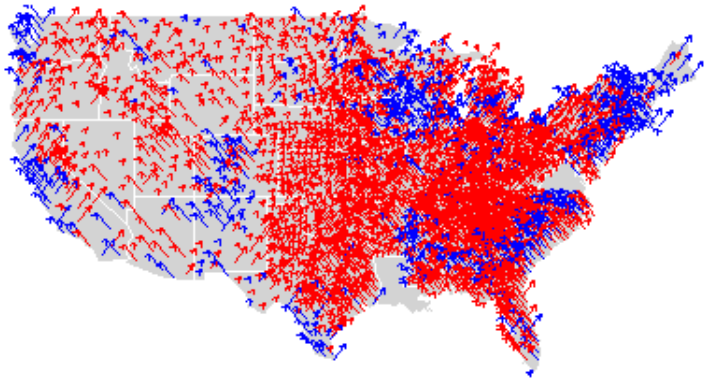


Figure 2.1: Color by Winner

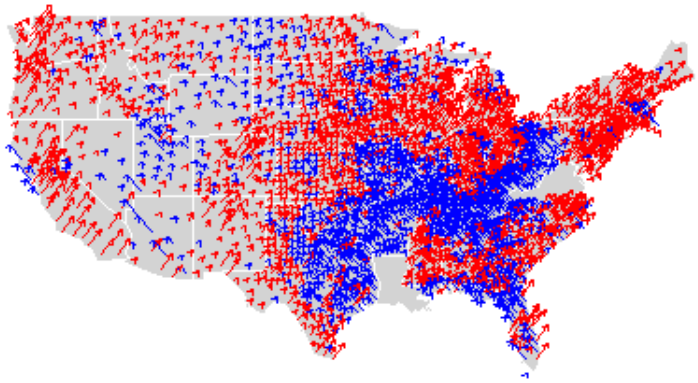


Figure 2.2: Color by Delta 2004 to 2012