2016 Election Analysis

Dana Laufer

December 17, 2020

1 Abstract

In this study, I address the problems that arose when predicting the 2016 presidential election, and search for the significant demographics in predicting a county's majority vote. In addition, I investigate the nature of counties whose majority vote changed parties from the 2012 to 2016. Using data on the 2016 presidential election and demographics, I first perform principal component analysis and hierarchical clustering to see any natural groupings of counties based on census data. Then, I use various classification methods to predict a county's vote in the 2016 election based on census data, concluding that the most significant predictors were the percent of population who are white and percent of population who commute on public transportation. Finally, I use those same classification methods only on counties who voted Democratic in the 2012 election to predict which counties would switch parties in the 2016 election. I concluded that the most significant predictors in whether a formerly Democratic county stayed Democratic or flipped was that county's proportion of white population and proportion of individuals working in particular job sectors.

2 Introduction

Predicting the outcome of elections has always been a complicated and difficult issue, largely because Voter behavior is so hard to predict. The simplest and most intuitive way to predict voter behavior would be to take a random sample, and generalize that data to the rest of the population. So, if 10% of the polled individuals stated that they will vote for Trump, then 10% of the United States population will vote for Trump. However, predicting voter behavior this way is often inaccurate, as there are many other factors to consider that will complicate the model. For one, there are the usual complications that involve random samples: ensuring that the samples are truly random and representative of the population. For example, an easy way to conduct a random sample is to call randomly generated phone numbers. However, this still may not be representative of the population, because some types of individuals may be more likely to ignore a call from an unknown number than others. Further, some portions of the population don't have phones, such as Amish communities and much of the homeless population. In addition, voter behavior is further complicated because there are many individuals who will change their minds many times before an election, particularly if both candidates are undesirable to them. As an individual takes in new information from the news, political ads, their friends and their family, they may change their mind on the candidate they want as president. Since new information comes out almost every day, the general public's favored candidate likely changes day to day, making it hard to predict the true election results. Finally, voter behavior is difficult to predict because in certain situations, individuals may lie because they are too embarrassed to share who they're voting for, and so a large portion of a candidate's support base may be unrepresented in polls.

In 2012, statistician Nate Silver successfully predicted the outcome of every single state in the presidential election[4]. Nate Silver's prediction approach was unique because for each day, he calculated the probability of a wide range of different voter proportions, rather than simply focusing on the voter proportion with the highest likelihood[3]. For example, looking at several poll surveys, he could calculate the probability of different voter distributions on a given day. Then, using Bayes' Theorem, he could calculate the probability of different voter distributions on the next day, given that the previous day's voting distribution was some particular proportion.

Despite Nate Silver's unique approach, he still didn't correctly predict Trump's victory in 2016, along with almost every other reputable predictions. Systematic polling error resulted in almost every single poll largely overestimating Clinton's lead over Trump. Many factors could have contributed to this systematic error, and they all link back to why voter behavior is so difficult to predict. For one, it may be that Trump's supporters, particularly women, were embarrassed to share with pollsters that they were voting for Trump. In addition, it's possible that Trump supporters tend to be more distrustful of institutions, and so they are less likely to respond to poll calls. Finally, it seems that many people who were either undecided or were planning on voting third-party decided last minute to vote for Trump. To make future predictions better, I believe there has to be more weight given to the immeasurable variation that is bound to occur in any prediction of human behavior. In addition, I think more money should be dedicated to ensuring that polls are carried out in a diligent, accurate manner.

Considering America's two party system, it would seem to me on the surface that our elected president's party shouldn't change much from election to election. However, there are many voters who don't strictly follow party lines, but rather vote for whichever candidate they prefer in a given election. In the end, these voters are truly the ones who decide the election results. Since a Democrat won in 2012 but a Republican won in 2016, I would guess that a big deciding factor of the 2016 election was voters who flipped from Democrat to Republican. My aim in this project is to investigate the 2016 presidential election and determine trends among voters, as well as investigate the nature of voters who voted for Obama in 2012 but voted for Trump in 2016.

3 Questions of Interest

The first question of interest that I will investigate is which significant factors can be used to predict a county's 2016 presidential winner.

The second question of interest that I will investigate is whether there are distinct differences between counties who voted Democratic in 2012 and Republican in 2016, and counties who voted Democratic in both years. How easily can one predict the change in a county's party from any given election to the next, and could it be useful in predicting future elections?

4 Data and Methods

4.1 Data

I will be using the two provided datasets on census data and 2016 presidential election data, as well as a dataset on presidential election data from 2000 to 2016. The 2000 to 2016 presidential election data that I found was collected by the MIT Election Data and Science Lab[2]. Data is provided on a sub-county level, giving the votes per candidate for every presidential election spanning 2000 to 2016, as well as additional information such as the party of the candidate and the FIPS value of the sub-county. Since I was only interested in the 2012 election data, I filtered out all of the data from other years before loading my data into R. In addition, in my 2016 election data, I filtered out all rows with a FIPS value of 2000, as according to the US Department of Agriculture, this FIPS value does not correspond to any county[1]. This left me with 18,345 rows in my 2016 election dataframe. In addition, I filtered out all state level and country level data, so that I was left with only county level data on the 2016 presidential election. Before fitting any classification models, I combined the census and election data together for each county.

All three of these datasets are from non-random data collection by the United States government. Because of this, they all offer precise measurements that do not distort the system in question. However, the census data, although the most recent available, was collected in 2010, while the election data in question was collected in 2012 and 2016. Because of this, there may be discrepancies between the estimated demographics of a county during an election and their true demographics.

As neither of these datasets provide any identifying information, it is very unlikely that analysis of the data could cause harm to the people represented in it. Neither of the datasets seem to have been compiled with nefarious purposes in mind, so it is unlikely that analysis of any of the data could raise ethical concerns.

4.2 Methods

For my first question of interest, I will fit a decision tree, a logistic regression, a logistic regression with a LASSO penalty, a KNN model, a boosting model and finally a random forest to predict a county's winning candidate (calculated as the candidate with the majority of votes), based on census data of said county. I will then decide on the best of the six for predicting the election results, and use this model to find the most significant predictors of winning candidate.

For my second question of interest, I will take only the counties who voted Democratic in the 2012 election, and perform the exact same methods as for my first question of interest. My main goal is in determining the differences between these two models, and to find the significant predictors in determining whether a county flipped from Democrat to Republican.

4.3 Exploratory Analysis

4.3.1 Visualization

Before delving into my questions of interest, I first produced preliminary visualizations to get an understanding of the relationship between my variables. First, I drew a bar chart of all named candidates and the total votes they received, to get an idea of the proportion of votes in the 2016 presidential election. The resulting plot is featured below.

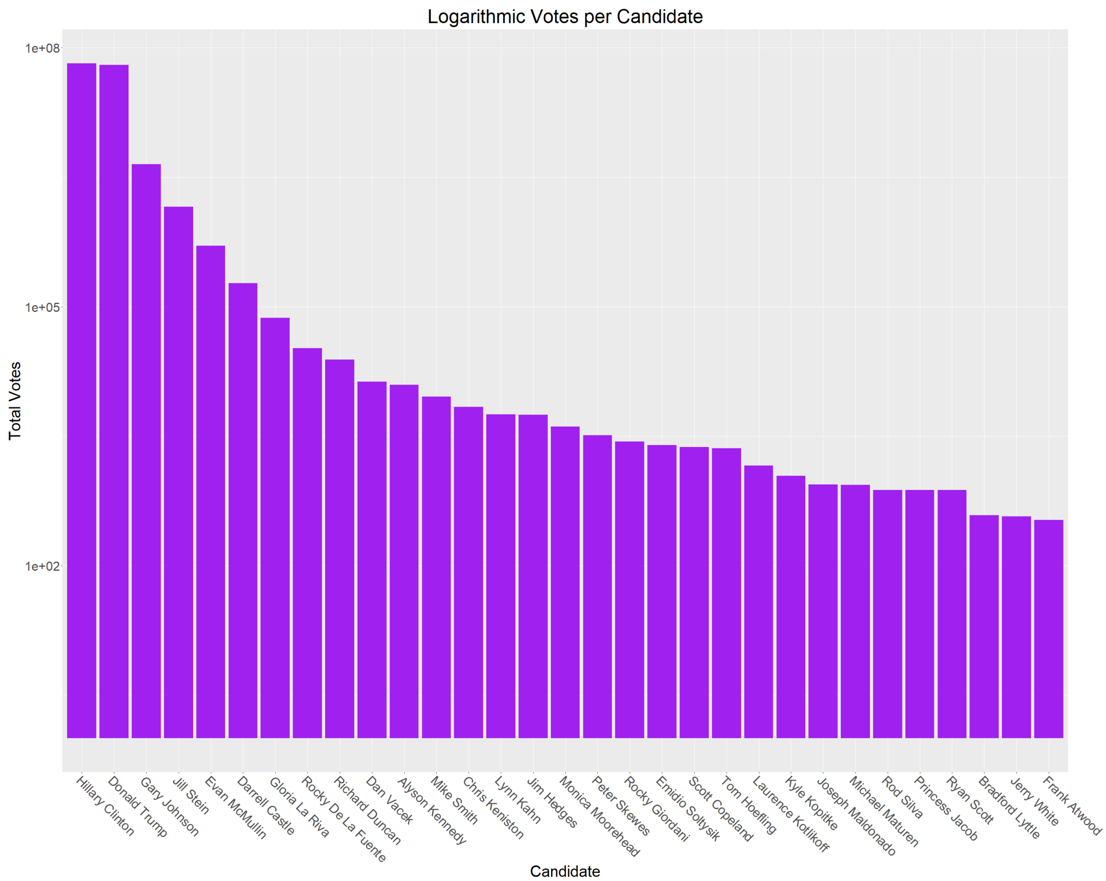

Figure 1: A logarithmic barplot of total votes received by each presidential candidate in the 2016 election.

Since Clinton and Trump received so many more votes than the other candidates, I plotted this on a logarithmic scale to better compare the votes that other, third party candidates received. We can see from this plot that there were 31 total named candidates in the 2016 presidential election.

Next, I found the candidate with the majority vote in each county and each state, and used this data to visualize the distribution of votes in America on a geographic level. (Figures 1 and 2 in the Appendix).

Then, I sought out to visualize the relationships between certain census demographics to determine how clear trends may be among counties. To do this, I created a scatterplot matrix among select variables (Figure 3 of the Appendix). While a few of the variables showed clear linear relationships, most seemed to have very little pattern. This told me that most of the variables should be fairly independent, and offer unique information to aid in predicting the winning candidate.

I was primarily interested in seeing whether any of these variables had a relationship with the voting distributions of counties. To do this, I chose a few variables to focus on, and for each one compiled all counties whose value of that variable was particularly high (I chose high to indicate a value equal to or greater than the 75th percentile). Then for each one, I found the proportion of votes for each candidate and graphed them all side by side. The resulting plot is provided below.

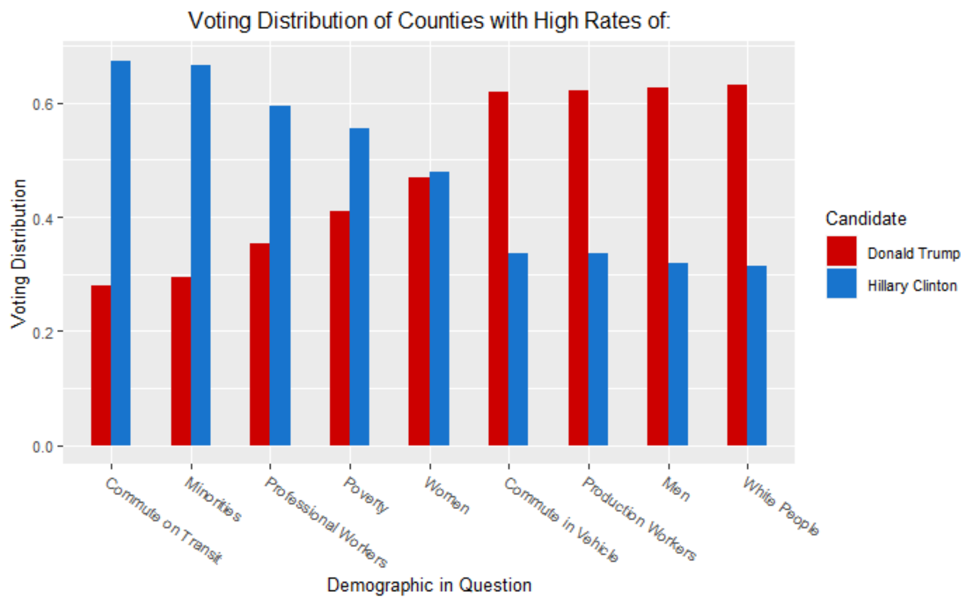

Figure 2: A barplot of the voting proportions of counties with high rates of various demographics.

Clearly there is some trend here among variables and voting distributions. We can see that counties with higher percentages of men voted proportionally for Trump much more than counties with higher percentages of women. This indicates that men may have been more likely to vote for Trump than women. In addition, counties with higher percentages of white people voted proportionally for Trump much more than counties with high percentages of minorities. Again, this implies that white people may have been more likely to vote for Trump than minorities. Interestingly, counties with higher rates of individuals who commute on public transportation voted proportionally for Trump much less than counties with higher rates of individuals who commute alone in a car, truck or van. I can think of two reasons for this. One, people who commute on public transportation are likely to be less wealthy than those who commute in a vehicle, and it would make sense that those with less wealth would be less likely to vote for Trump, as indicated by the voting distribution of counties with high levels of poverty. On the other hand, this could be because counties with high levels of commuters on public transportation are likely big cities such as New York or San Francisco, which historically vote much more democratically than rural areas.

Finally, using the 2012 and 2016 presidential election data, I visualized which counties changed parties from 2012 to 2016. Counties who flipped from Democrat to Republican are colored red, while counties who flipped from Republican to Democrat are colored blue.

Clearly, the large majority of counties remained unchanged. However, we can see that many more counties flipped Republican than flipped Democrat. This makes sense given that the winner of the 2021 election was a Democrat, while the winner of the 2016 election was

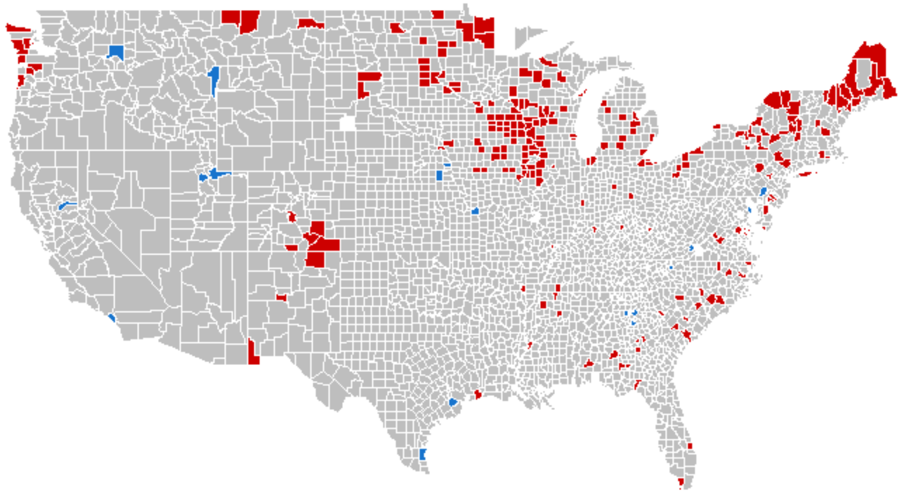

Figure 3: A map of counties whose winning presidential candidate's party flipped from 2012 to 2016.

a Republican. This reinforces my previous thought that counties who flipped to Republican were vital in the 2016 election, and tells me that I should largely focus on these counties in my analysis.

4.3.2 Dimensionality Reduction

After investigating the data through visualizations, I performed principal component analysis as a dimensionality reduction tactic to get a better sense of the most important variables. First, I ran a principal component analysis on the county and sub-county level 2016 presidential election data, being sure to scale and center the data. I scaled the data because while many of the attributes are given on the same scale (percentages), there are some that are on a much different scale than the rest. For example, all of the data on income is given in dollars, whose values are all much greater than any of the percentages. In addition, the attribute MeanCommute is given in minutes, which is on a much smaller scale than the income data is. By scaling the features, I ensure that they are all on the same scale and therefore can be compared fairly. In addition, I center the data in order to ensure that they are all spread out around zero. Since none of our attributes can be negative, none of the attributes can currently be centered around zero,so we should center them before running principal component analysis.

Upon running principal component analysis, I see that for the county level data, per capita income, percent of children under poverty level, and percent of population under poverty level are the three features with the largest absolute values of first principal component. The first principal component value for per capita income is positive, while percent of child poverty and percent of poverty are negative. This means that counties with a high PC1 value will have lower than average percentages of child poverty and general poverty, and higher than average income per capita. This implies that income per capita is negatively correlated with poverty and child poverty. In addition, by calculating the cumulative proportion of variance explained, I found that 18 principal components are needed to explain 90% of the variation in the county level census data. The plot of proportion of variance explained and cumulative proportion of variance explained are Figures 4 and 5 in the appendix.

On the other hand, per capita income, percent of professional workers and median income are the three features with the biggest absolute values of first principal component on the sub-county level. Here, all three of these attributes have positive PC1 values. This means that sub-counties with a high PC1 values have higher than average income per capita, professional workers and median income. This implies that income per capita, percent of professional workers and median income are all positively correlated. In addition, by calculating the cumulative proportion of variance explained, I found that 21 principal components are needed to explain 90% of the variation in the sub-county level census data. The plot of proportion of variance explained and cumulative proportion of variance explained are Figures 6 and 7 in the Appendix.

‘

4.3.3 Clustering

As a final step of exploratory analysis, I ran hierarchical clustering with complete linkage on both the original county level data and the county level data reduced to five dimensions using principal component analysis. I cut both into 10 clusters and plotted a dendrogram (Figures 8 and 9 in the Appendix). The clustering based on the dimensioned-reduced county level data seemed to place a very large portion of the counties in either a pink cluster or a green cluster, while the clustering on the original data seemed to place counties in clusters slightly more evenly. In the original county level clustering, San Mateo County is clustered with 26 other counties, almost all of which come from historically blue states. However, in the dimension-reduced clusters, I see that now San Mateo County is clustered with 225 other counties. This time, many of the other counties in the cluster are from historically Republican states, such as Utah, Texas and Wyoming. It seems as though the original data clustering put San Mateo county in a more accurate cluster, since San Mateo county is in California, a historically blue state. It makes sense that the clusters observed from hierarchical clustering on the first five principal components would be larger and less accurate than that from the full dataset. While the first five components account for much of the variation, it doesn't account for all of it in the same way that the full dataset does.

5 Results

5.1 Question 1

In order to answer my first question of interest, I fit split the 2016 election data into a randomly selected training and testing set, and fit six classification models on the training data. For each model I saved the test and training error, so that I can compare the models at the end.

The first classification model that I fit on the 2016 data was a decision tree, pruned using 10-fold cross validation to minimize misclassification error. The plot of the decision tree before pruning can be found at Figure 10 in the Appendix, while the pruned decision tree is plotted below.

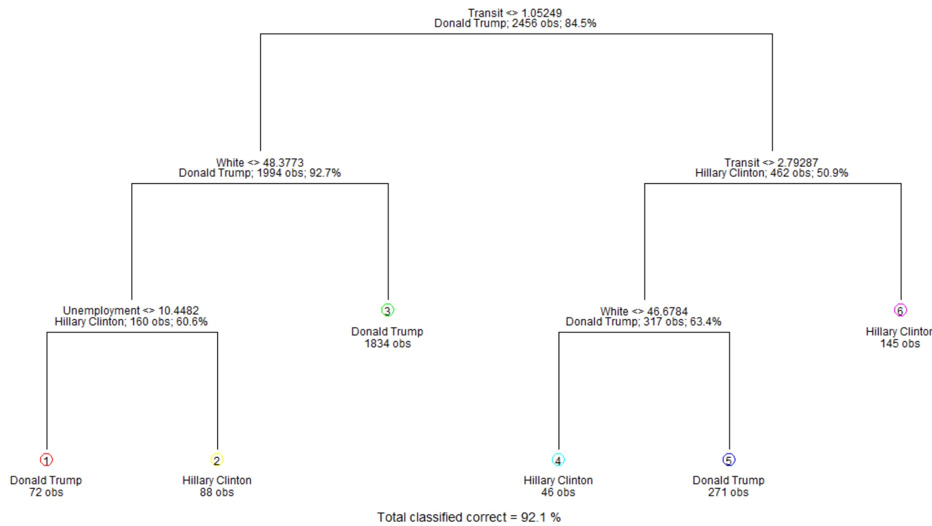

Figure 4: A pruned decision tree of the 2016 election data.

From the pruned decision tree, we can see that the most important variables to predict a county's winner are Transit, White, and Unemployment. It looks like having a high Transit variable leads a county to be predicted to vote for Clinton. This goes back to the chart I made earlier, where counties with high transit tended to have a higher proportion of Clinton votes. In addition, it looks like if a county has a high White variable it will be more likely to have Trump predicted as the winner. Finally, if a county has a high 17 unemployment rate, Clinton is more likely to be predicted as the winner. In all, this implies that counties with higher White variables are correlated with Trump winning, while counties with higher unemployment and Transit variables are correlated with Clinton winning.

Next, I fit a logistic regression model. From the summary of the model, I see that there are many significant variables in the logistic regression model, but a few of them are White, Unemployment, Drive, IncomePerCap, and Citizen. Two of these are consistent with the decision tree, but one of the important predictors in the decision tree, Transit, is not significant in this model. It's possible that in this model, since Drive and Carpool are both significant predictors, they take on the same role that Transit did in the decision model, since Transit is almost surely negatively correlated with Drive and Carpool.

Remember that in a logistic regression, the odds, $p1-p$, are equal to e to the linear combination of predictors and coefficients given. So, a single unit increase in any of the predictors will increase the odds multiplicatively by e^{β}, where *beta* is the coefficient of the predictor. So, for example, if a county has a one unit increase in percent of white population, the estimated odds that that county will vote for Clinton increase multiplicatively by $e^{-1.388e-01} = 0.87$, since the coefficient for White is -1.388e-01. Since this number is less than 1, the odds will decrease. In addition, a one unit increase in percent of population that is unemployed will multiplicatively increase the estimated odds that that county will vote for Clinton by $e^{2.105e-01} = 1.234$. Since this number is greater than one, the estimated odds will increase. These two estimations both make sense given the decision tree.

Next, I fit a lasso logistic regression model, and see that the MSE is minimized when the regularization parameter lambda is set to 50e-4. When fitting the regression with this optimal value of lambda, all of the coefficients are non-zero except for Income, IncomeErr, Office, Transit, and WorkAtHome. So, most of the important predictors from the last two classification models are still significant in this model. However, Income was one of the significant predictors in the logistic regression, which is now zero. This predictor was removed from this model as a way to prevent overfitting, but likely increased the bias of the model.

Next I fit a K nearest neighbors model, and found the optimal K value to be 50 using 10-fold cross validation. The KNN model has no way of finding importance of predictors, so in this model all predictors were given equal weight.

Finally, I fit both a boosting model and a random forest model. The most significant predictors for both of these models were by far White and Transit. The plot of variable importance for the both models are provided as figures 11 and 12 in the Appendix.

In order to compare these six models, I charted an ROC curve, as well as displayed the testing and training error for each model below.

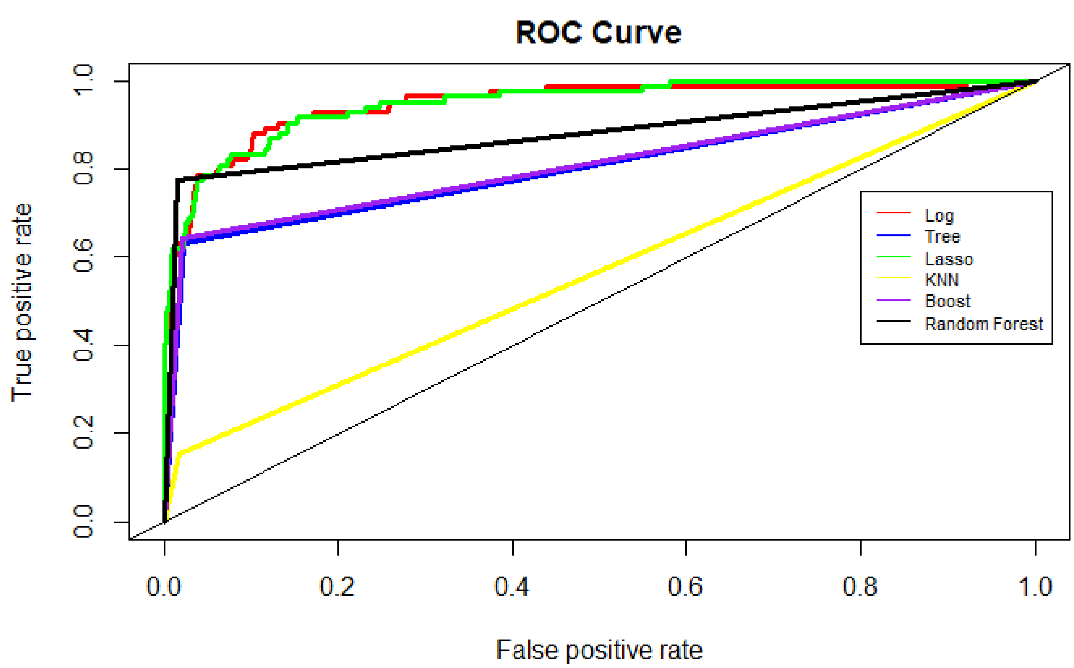

Figure 5: The Receiver Operating Curve of all six classification models for 2016 election data.

Model Type	Training Error	Testing Error
Decision Tree	0.07899	0.06991
Logistic	0.07858	0.06504
Lasso	0.07369	0.06504
KNN	0.14047	0.13008
Boosting	0.06474	0.0667
Random Forest	0.00081	0.04390

Table 1: A table giving the testing and training misclassification error of each model.

It looks as though both the logistic regression and the lasso regression have a much larger area under the curve than the other classification methods. In computing the areas under the curve for each model, I find that the lasso logistic regression has the greatest area at 0.949. So, if one's goal is to minimize the false positive rate and maximize the true positive rate, they should likely model the 2016 election using a lasso logistic regression. However, looking at our recorded misclassification rates, we can see that the random forest model has the lowest training error and the lowest testing error by far. So, if one's goal is to reduce misclassification error, one should likely use a random forest to model the 2016 election. In choosing between the lasso regression model and random forest model, the lasso regression model is much more difficult to visualize than the random forest tree would be, so the results may be more difficult to interpret if one chooses the lasso model. On the other hand, since the lasso logistic model gives probabilities rather than classifications, this model may be better in answering certain questions about the proportion of votes within a county. For example, it would be interesting to see if the probability of Clinton winning times a county's population would give a number close to the number of votes for Clinton in that county.

5.2 Question 2

In order to answer my second question of interest, I found the 2016 election data for only counties who voted for Obama in the 2012 election, split the data into training and testing sets, and fit the same classification models as before, once again saving the test and training error for each.

The first classification model that I fit was a decision tree, pruned using 10-fold cross validation to minimize misclassification error. The plot of the decision tree before pruning can be found at Figure 13 in the Appendix, while the pruned decision tree is plotted below.

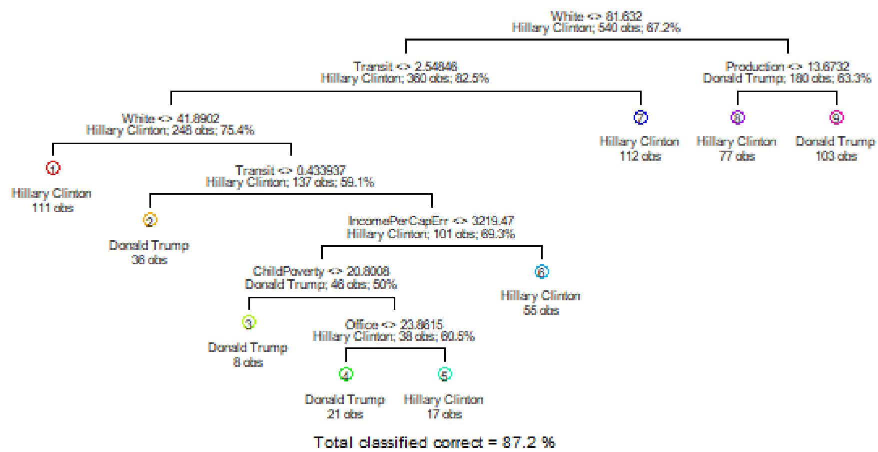

Figure 6: A pruned decision tree of the 2012 Democratic counties, predicting the outcome of the 2016 election.

From the pruned decision tree, we can see that the most important variables to predict a county's winner are Transit, White, Production, IncomePerCapErr, ChildPoverty and Office. Just like the previously constructed decision tree, it looks like having a high Transit variable leads a county to be predicted to vote for Clinton, whereas having a high White variable leads a county to be predicted to vote for Trump. In contrast to the previous decision tree, there are many more variables at play here. The most interesting one I see is IncomePerCapErr, which tracked the error of the income per capita. If a county has high income per capita error, it is predicted that Clinton will win. It seems puzzling that the error term would have any correlation with a county's winning candidate. In addition, with both Office and Production variables on this tree, it looks like what type of jobs a county tends to have might have more influence on who wins the election than it did previously.

Next, I fit a logistic regression model. From the summary of the model, I see that there are many fewer significant predictors in this model than in the previous logistic model. The three most significant predictors are White, Professional and PrivateWork. This seems somewhat consistent with the important predictors from the decision tree, but it seems curious that the decision tree had more significant predictors than the general decision tree did, but this logistic model has much fewer significant predictors than the more general logistic model did. Again, we see that race and type of job has a significant influence on who won the election in any given county.

In this model, if a county has a one unit increase in percent of white population, the estimated odds that that county will vote for Clinton increase multiplicatively by $e^{-9.965e-02}$ $= 0.905$, since the coefficient for White is -9.965e-02. While this number means that an increase in white population will still decrease the county's likelihood of voting for Clinton, the magnitude of decrease is slightly smaller than in the general logistic model.

Next, I fit a lasso logistic regression model, and see that the MSE is minimized when the regularization parameter lambda is set to 10e-4. When fitting the regression with this optimal value of lambda, all of the coefficients are non-zero except for ChildPoverty, OtherTransp and SelfEmployed. This lasso model has many more significant predictors than the more generalized lasso model, just as the decision tree did.

Next I fit a K nearest neighbors model, and found the optimal K value to be 30 using 10-fold cross validation. Again, the KNN model has no way of finding importance of predictors, so in this model all predictors were given equal weight.

Finally, I fit both a boosting model and a random forest model. The most significant predictors for both of these models were by far White and Production. The plot of variable importance for the both models are provided as figures 14 and 15 in the Appendix.

In order to compare these six models, I charted an ROC curve and displayed the testing and training error for each model below.

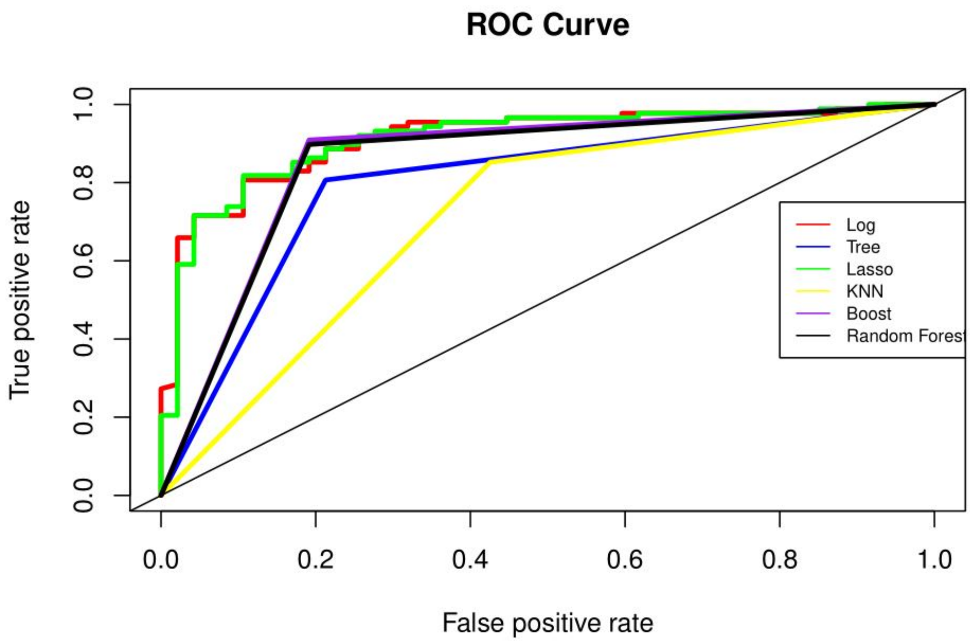

Figure 7: ROC curve of all six classification models for Democratic 2012 counties.

Model Type	Training Error	Testing Error
Decision Tree	0.12778	0.20000
Logistic	0.1463	0.15556
Lasso	0.15556	0.15556
KNN	0.25741	0.24444
Boosting	0.23333	0.23704
Random Forest	0.00185	0.13333

Table 2: A table giving the training and testing misclassification error for each model.

It looks as though once again, both the logistic regression and the lasso regression have a much larger area under the curve than the other classification methods. Upon calculating the areas under the curve, I find that this time the logistic regression has the greatest area at 0.913. However, the boosting and KNN models seem to have larger areas under the curve in this ROC than in the previous. In computing the areas under the curve for each model, I find that the lasso logistic regression has the greatest area at 0.949. So, if one's goal is to minimize the false positive rate and maximize the true positive rate, they should likely model the Democratic 2012 counties using a logistic regression. However, looking at our recorded misclassification rates, we can see that once again the random forest model has the lowest training error and the lowest testing error. So, if one's goal is to reduce misclassification error, one should likely use a random forest to model the Democratic 2012 counties in the 2016 election. The pros and cons for the logistic regression and random forest model are the same as those for the lasso logistic regression and random forest model, since the nature of the logistic regression and lasso logistic regression are so similar.

The random forest model was one of the ideal models for both sets of data. If we were to compare both of these models, the two most significant predictors for the general data are White and Transit, while the most significant predictor for the Democratic data was by far White, followed by Production. Again we can see that White is an extremely predictor for who a county voted for, whether they were blue or red in 2012. However, we also see that while percent of population who commutes on public transportation is a significant predictor for general data, it was not very significant for the Democratic counties.

In addition, the lasso logistic regression model was one of the ideal models for the generalized 2016 data, while a general logistic regression model was one of the ideal models for the 2012 Democratic data. If we were to compare both of these models, we see that all of the predictors for the generalized election data are significant except for Income, IncomeErr, Office, Transit, and WorkAtHome. On the other hand, for the data of only the counties who were Democratic in 2012, the most significant predictors were White, Professional and PrivateWork. Therefore, while there's many predictors that are significant in predicting who a county voted for in the 2016 election, there are much fewer significant predictors in predicting whether a Democratic county would flip in 2016. By calculating the effect of a unit change in these variables on the odds of Clinton winning, we can see that a higher white population decreases the odds of a Democratic county still voting blue in 2016. On the other hand, a higher population working in either a professional job or in private work increased the odds of a Democratic county still voting blue in 2016. So higher white populations and lower professional and private workers were the main factors in making a blue county more likely to flip to Republican in the 2016 presidential election.

6 Conclusion

Extensive statistical analysis and classification models revealed that two of the most important variables in predicting which candidate received majority vote in a county are the percent of population that is white and percent of population who commutes on public transportation. On the other hand, if a county voted blue in 2012, the most important variables in predicting whether they flipped or remained blue in 2016 are percent of population who is white and percent of population who works in a production job. In the future, it's possible that focusing on these significant variables when predicting election outcomes could improve the prediction accuracy. In addition, I believe that focusing on which counties will flip in an election rather than simply which counties will vote for which candidate could help improve prediction models. Just like Nate Silver based his prediction models on every day up until the election rather than simply the election day, this could be a way to focus on the big picture rather than solely the final result and achieve more accurate predictions.

Future work could be focused on investigating the nature of counties who flipped from Republican to Democrat from the 2012 to 2016 presidential election. In addition, the classification models, particularly the logistic and lasso regression models, could likely be fitted better to the data. In my findings, I used a simple 0.5 probability threshold to determine class labels. However, there may have been other probability thresholds that would have resulted in a more accurate model.

I believe the results found in this study are trustworthy, and apply to the entirety of the United States. However, it is important to note that the findings only apply to the 2016 presidential election. It's very possible that the significant variables in predicting the 2020 presidential election are very different. However, this study is a good starting point for future investigations of election prediction methods.

References

[1] United States Department of Agriculture. *County FIPS Codes*. URL: https://www.nrcs.usda.gov/wps/portal/nrcs/detail/national/home/?cid=nrcs143_013697.

[2] MIT Election Data and Science Lab. *County Presidential Election Returns 2000-2016*. Version V6. 2018. DOI: 10.7910/DVN/VOQCHQ. URL: https://doi.org/10.7910/DVN/VOQCHQ.

[3] Bob O'Hara. *How did Nate Silver predict the US election?* URL: https://www.theguardian.com/science/grrlscientist/2012/nov/08/nate-sliver-predict-us-election.

[4] Daniel Terdiman. *Obama's win a big vindication for Nate Silver, king of the quants*. Nov. 6, 2012. URL: https://www.cnet.com/news/obamas-win-a-big-vindication-for-nate-silver-king-of-the-quants/.

7 Appendix

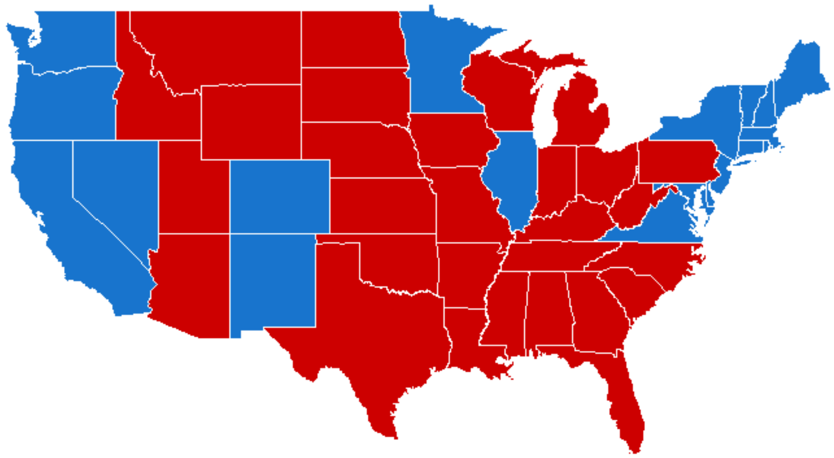

Figure 1: A map of states colored by winning candidate, where red corresponds to Trump and blue to Clinton.

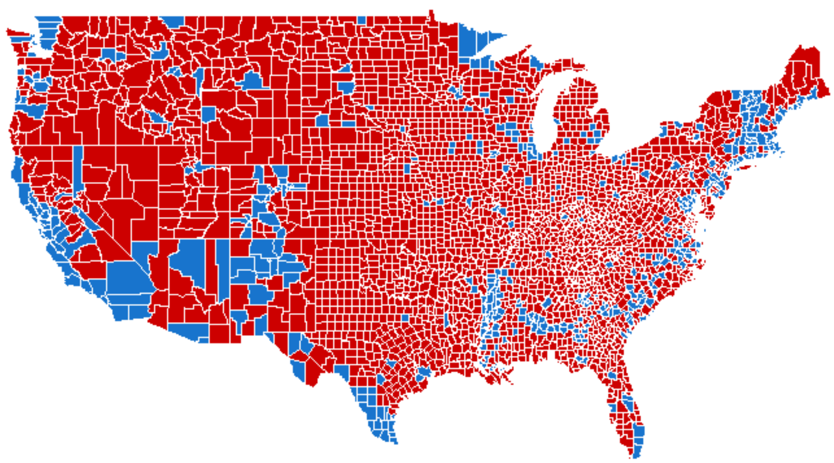

Figure 2: A map of counties colored by winning candidate, where red corresponds to Trump and blue to Clinton.

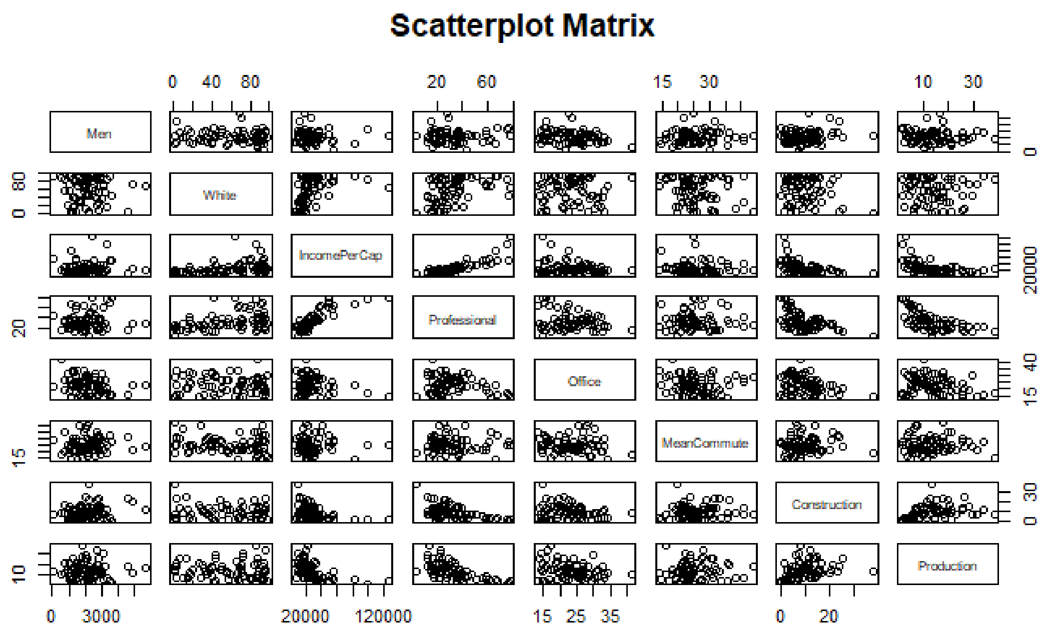

Figure 3: A scatterplot matrix of various variables from the census data.

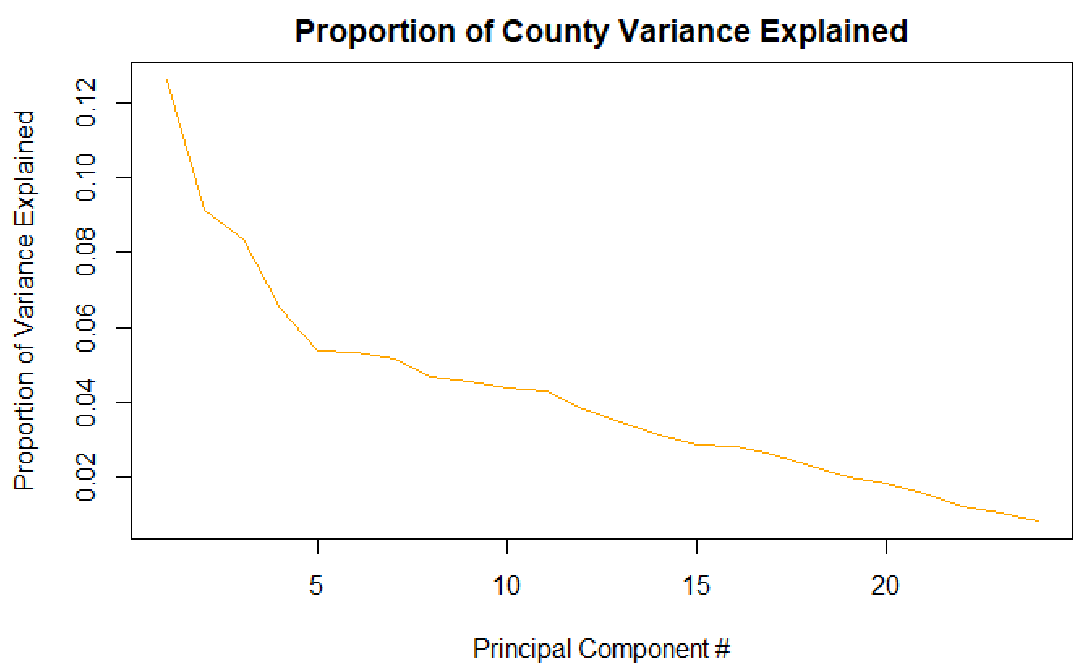

Figure 4: A plot of the proportion of variance accounted for by each principal component for county data.

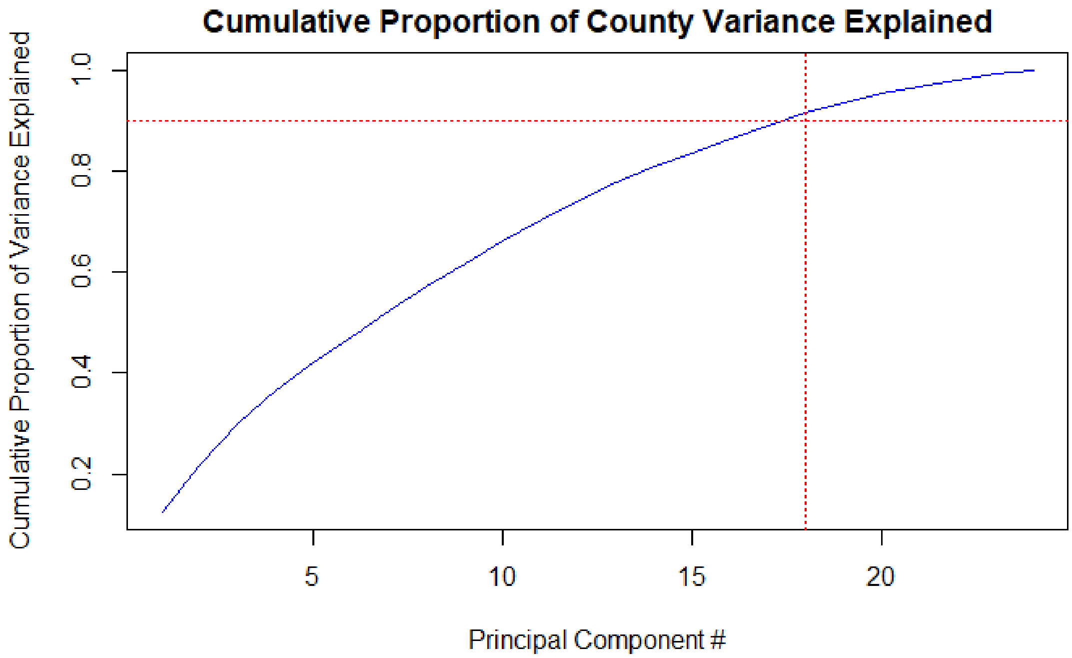

Figure 5: A plot of the cumulative proportion of variance accounted for by each principal component for county data. The principal component at which 90% of the variance is explained is indicated with red lines.

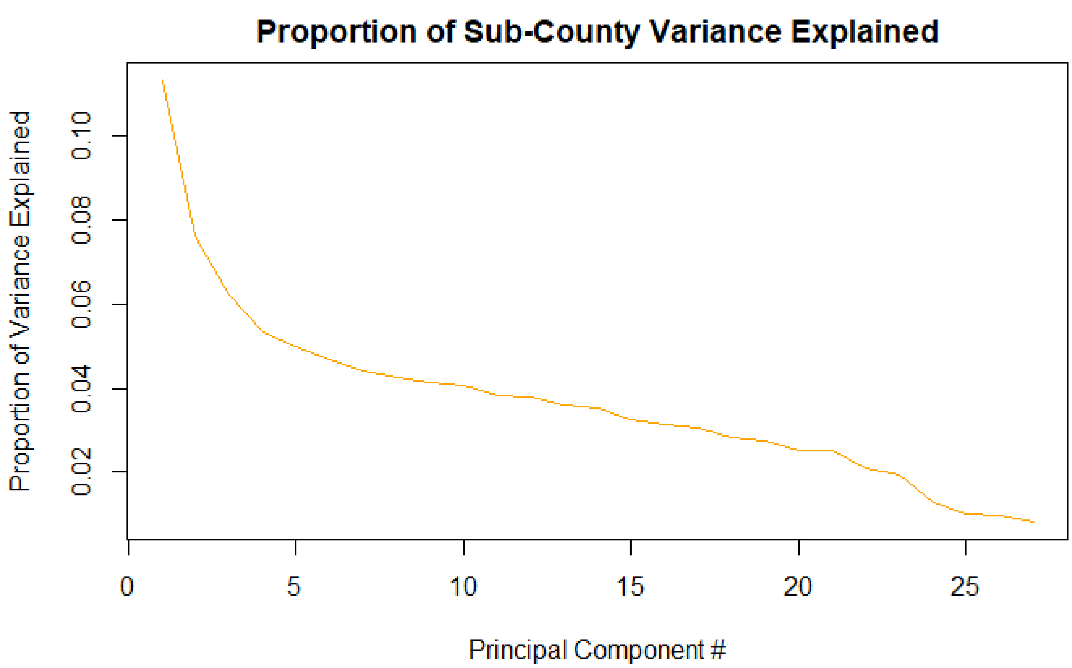

Figure 6: A plot of the proportion of variance accounted for by each principal component for sub-county data.

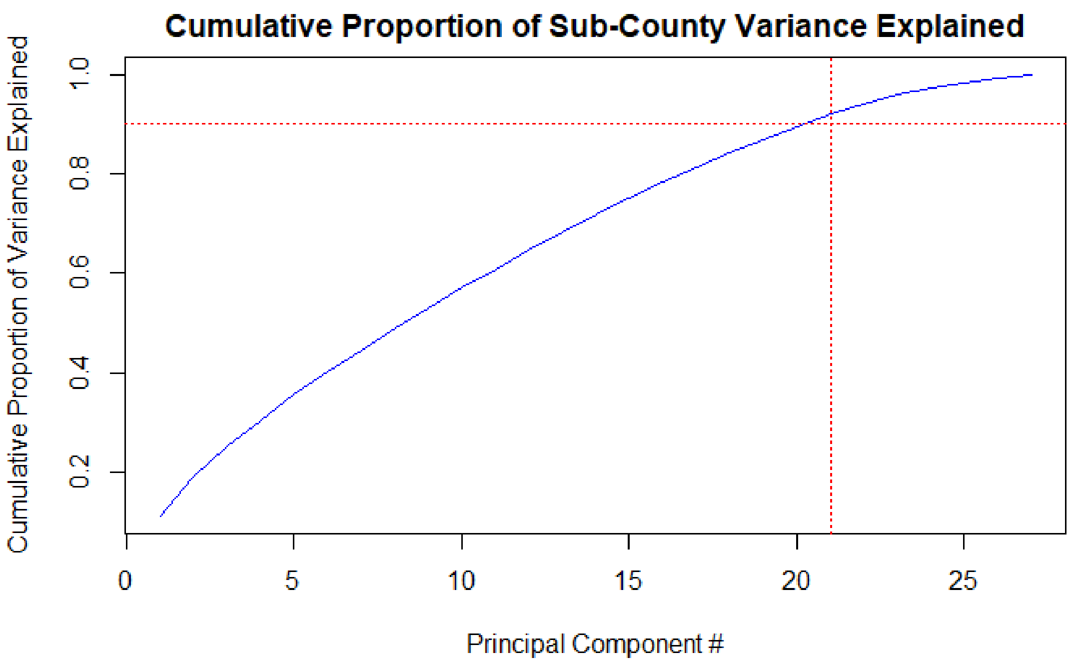

Figure 7: A plot of the cumulative proportion of variance accounted for by each principal component for sub-county data. The principal component at which 90% of the variance is explained is indicated with red lines.

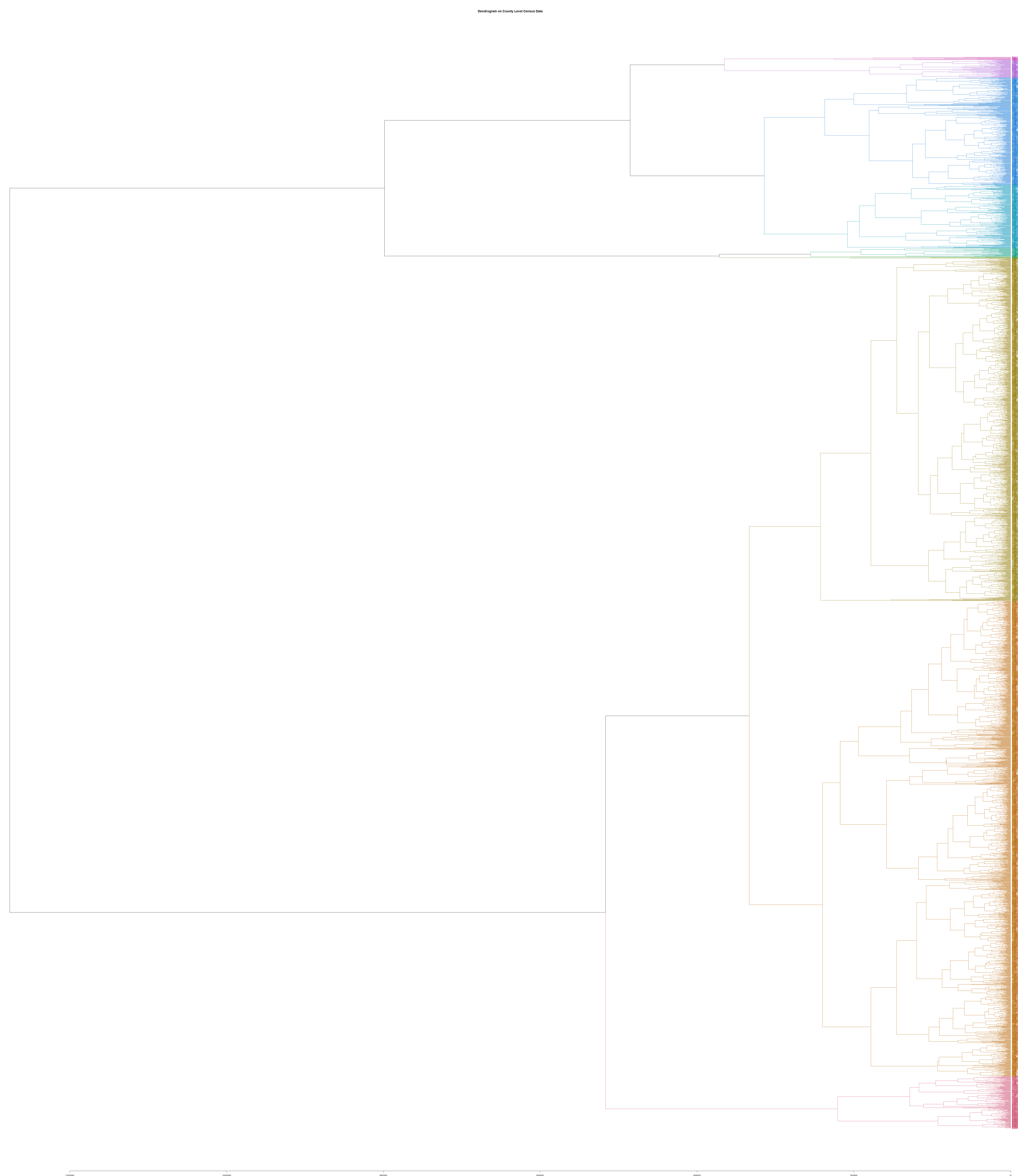

Figure 8: A dendrogram of the county level data, colored into 10 clusters.

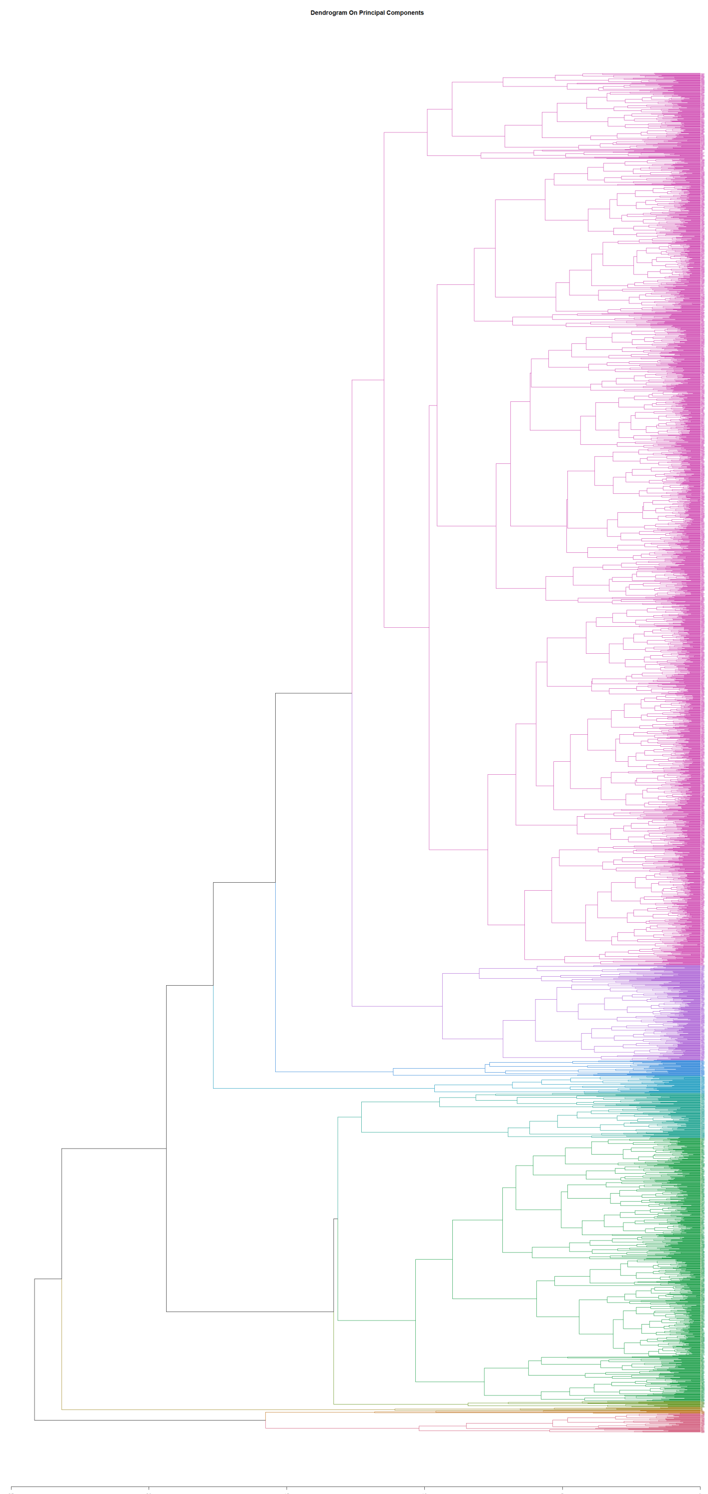

Figure 9: A dendrogram of the five dimension county level data, colored into 10 clusters.

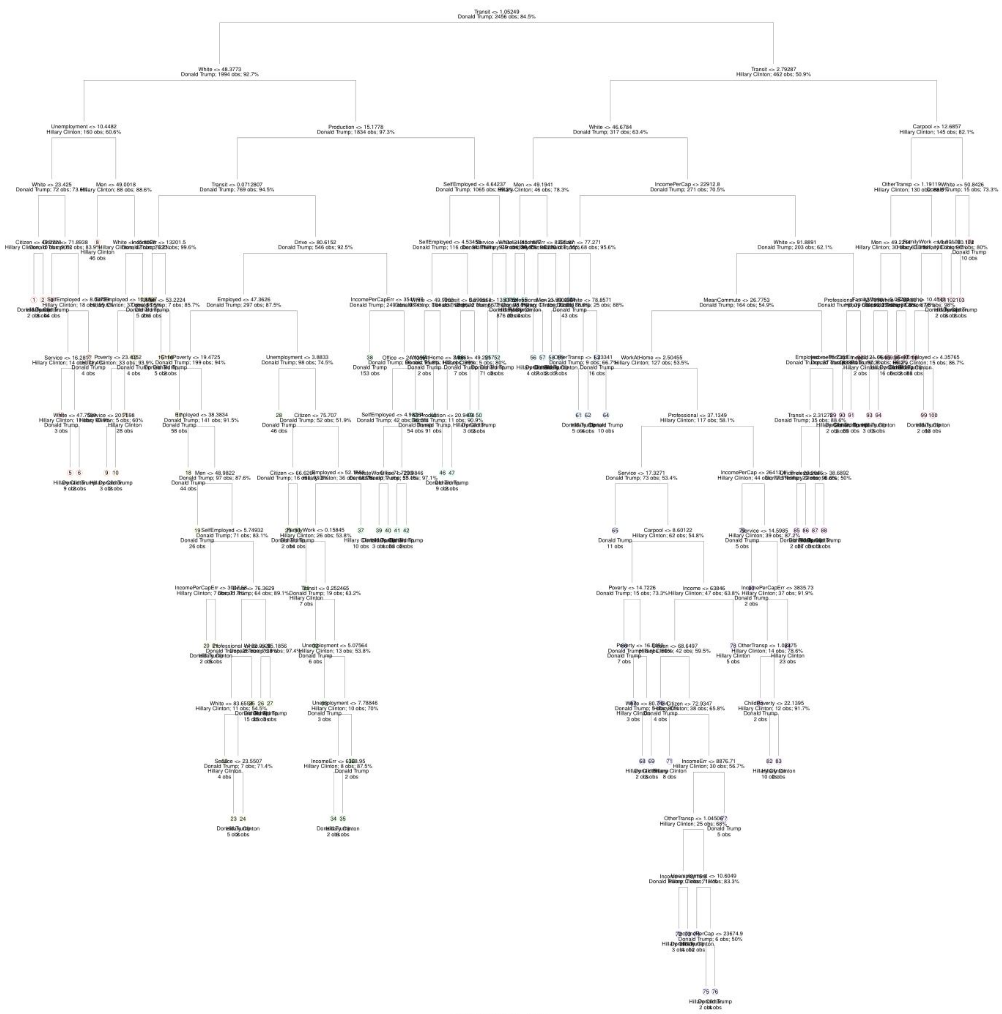

Figure 10: The unpruned decision tree of the 2016 election data.

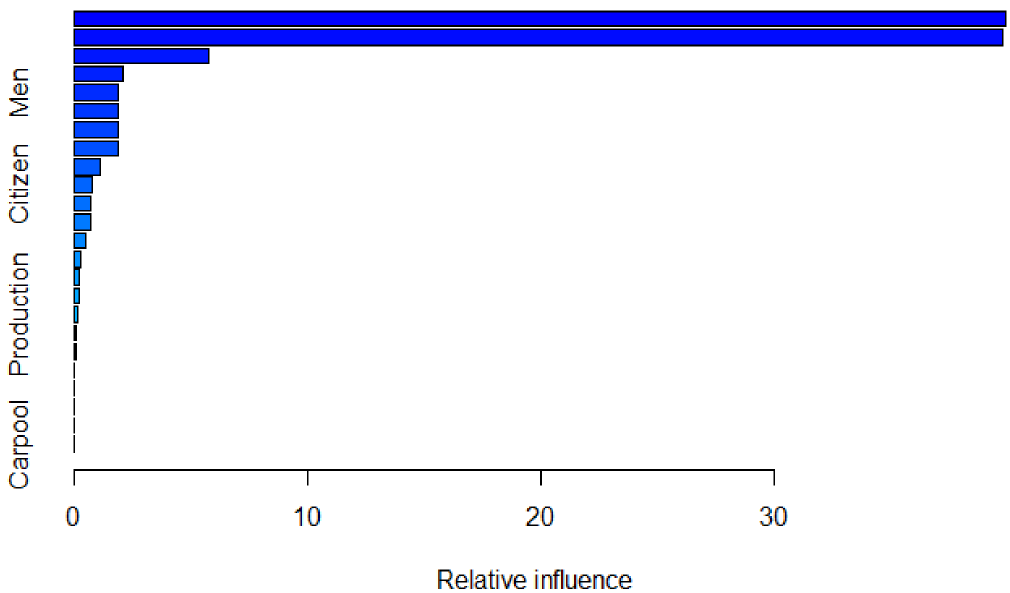

Figure 11: A plot of variable importance of the boosting model, performed on 2016 election data.

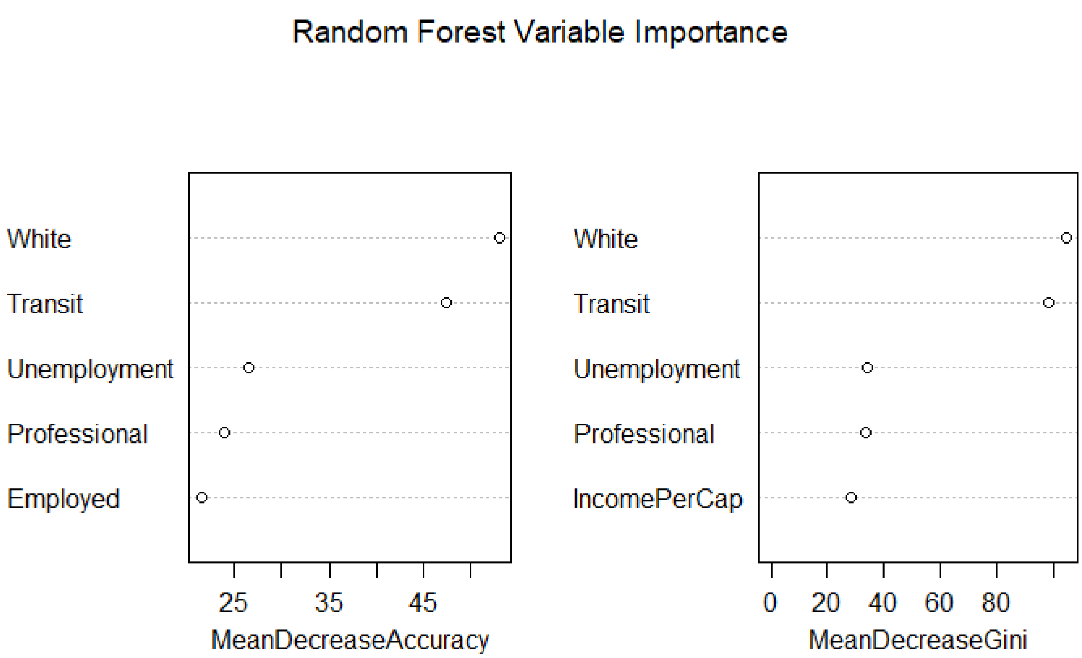

Figure 12: A plot of variable importance of the random forest model, performed on 2016 election data.

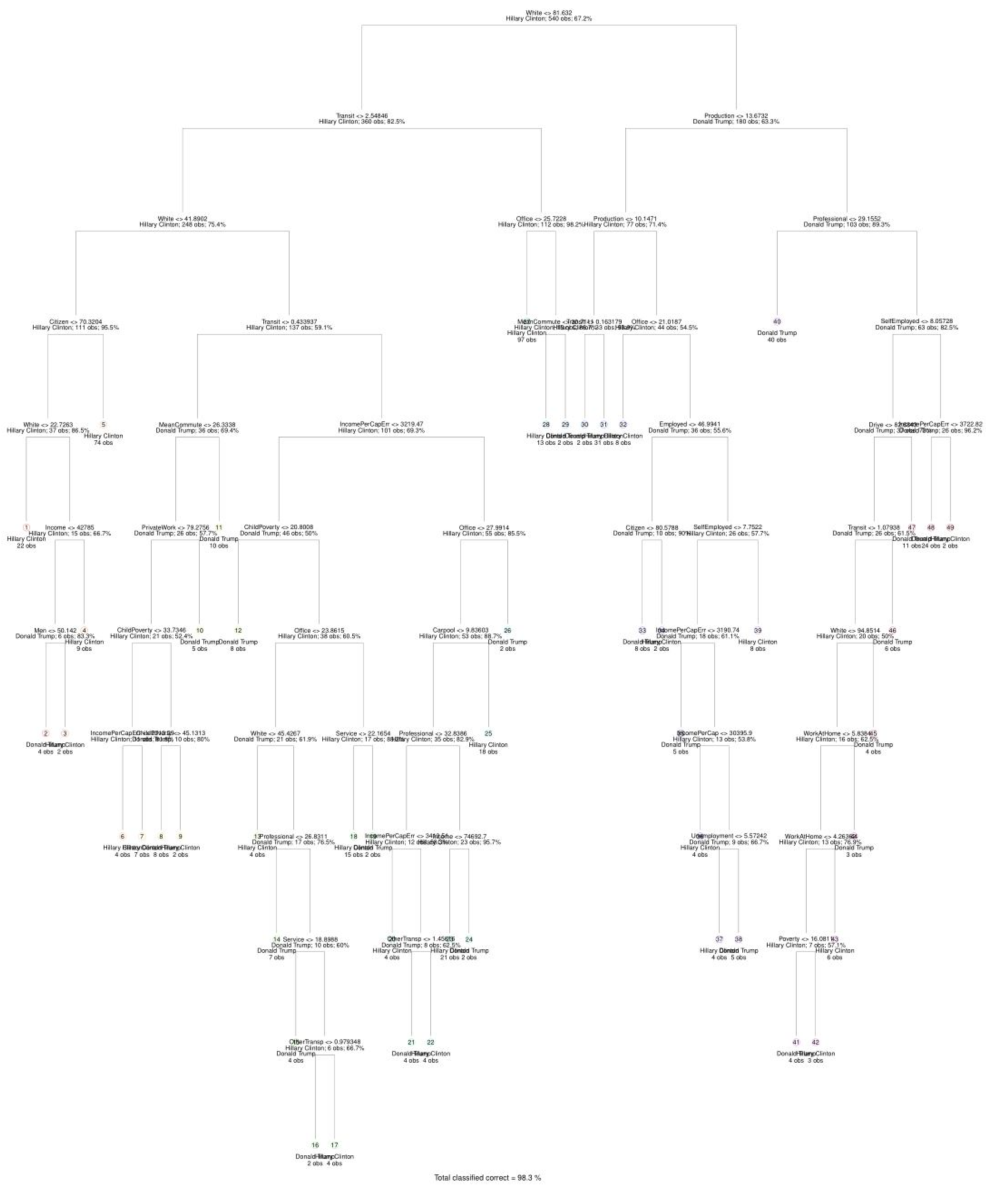

Figure 13: A plot of the unpruned decision tree of the 2012 Democratic counties.

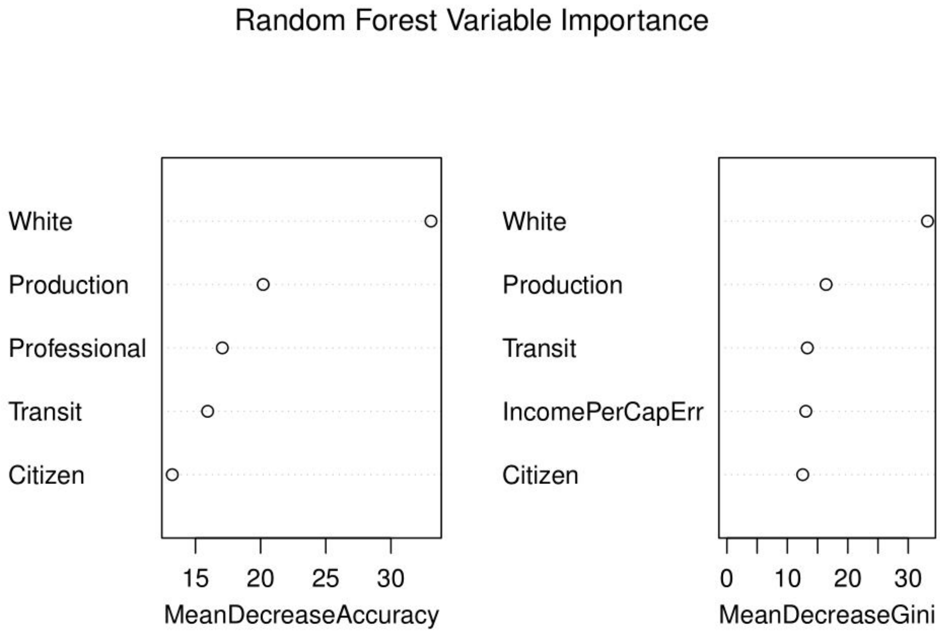

Figure 14: A plot of the variable importance of the random forest model, on the Democratic 2012 counties.

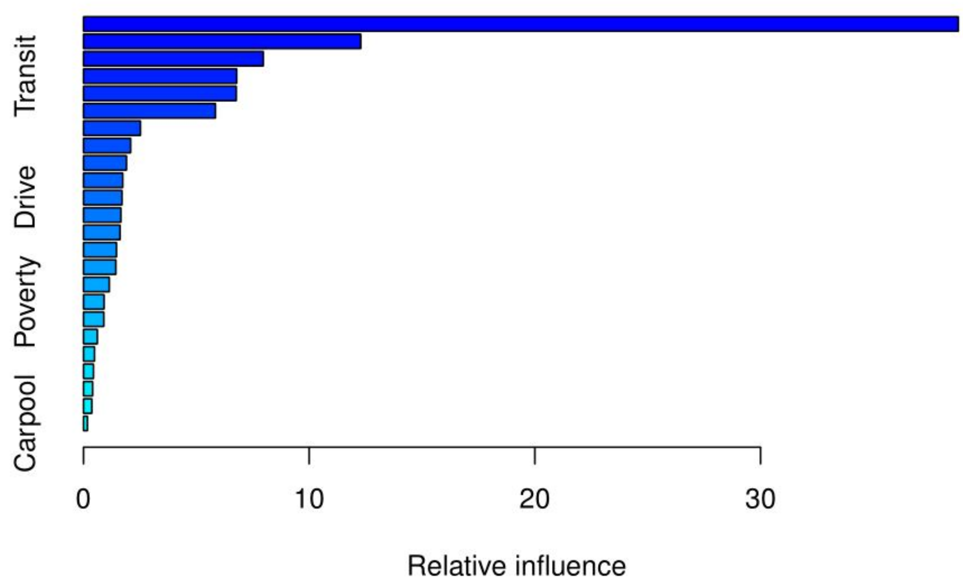

Figure 15: A plot of variable importance of the boosting model, performed on Democratic 2012 counties.