Assignment 1 - Segmenting Clinton and Obama Voters

Sarah Mansoor

2022-10-20

The Data

What data do you have access to? Any adjustments/changes you made to prepare the data for your analysis. What other data you may need to improve your analysis?

The data consists of 41 variables and 2868 observations. The data contains variables relating to number of votes, geographic details, election details, age details, education details, race details, employment and income details, and population details (see Appendix A). The columns that had missing rows were handled by replacing the missing values with the mean of those columns. The data that is available is sufficient to conduct a meaningful analysis, but one factor that might influence Clinton and Obama's votes are any other parties that may win in some Countys. It is possible that someone besides Clinton and Obama may win in another County, so having information about other votes would improve the analysis. (medicare vs medicarerate, socialsecutiry vs socialsecutiryrate, disabilities vs disabilityrate ??).

Variable	Variable Description
County	County Name
State	State Code
Region	Region Name
FIPS	Federal Information Processing Standards
ElectionDate	Election Date
ElectionType	Election Type
TotalVote	Total Votes
Clinton	Clinton's Votes
Obama	Obama's Votes
MalesPer100Females	Percentage of Males
AgeBelow35	Percentage below 35 years of age
Age35to65	Percentage between 35 and 65 years of age
Age65andAbove	Percentage above 65 years of age
White	Percentage White Individuals
Black	Percentage of Black Individuals
Asian	Percentage of Asian Individuals
AmericanIndian	Percentage of American Indian Individuals
Hawaiian	Percentage of Hawaiian Individuals
Hispanic	Percentage of Hispanic Individuals
HighSchool	Percentage with a Highschool Diploma
Bachelors	Percentage with a Bachelors Degree
Poverty	Percentage in Poverty
IncomeAbove75K	Percentage with Income Above 75K
MedianIncome	Median Income of County
AverageIncome	Average Income of County
UnemployRate	Unemployment Rate
ManfEmploy	
SpeakingNonEnglish	Percentage Speaking Non-English
Medicare	Number of those with Medicare
MedicareRate	Medicare Rate
SocialSecurity	Number of those with Social Security
SocialSecurityRate	Social Security Rate
RetiredWorkers	Number of Retired Workers
Disabilities	Number with Disabilities
DisabilitiesRate	Disabilities Rate
Homeowner	Percentage who are Homeowners
SameHouse1995and2000	Percentage in the same house in 2000 as 1995
Pop	Population
PopDensity	Population Density
LandArea	Land Area of County
FarmArea	Farm Area in County

```
# Read Data into R
setwd("~/Desktop/Fall 2022 MMA/MGTA 601/a1")
clinton_obama <- read.csv("Clinton-Obama.csv")

# Next use the function summary to find out the number of missing data points in the demographic/county
summary(clinton_obama)
```

```
##     County             State              Region              FIPS
##  Length:2868        Length:2868        Length:2868        Min.   : 1001
```

```
##  Class :character   Class :character   Class :character   1st Qu.:18102
##  Mode  :character   Mode  :character   Mode  :character   Median :30110
##                                                          Mean   :31029
##                                                          3rd Qu.:46124
##                                                          Max.   :56045
##
##  ElectionDate       ElectionType         TotalVote          Clinton
##  Length:2868        Length:2868        Min.   :     13   Min.   :     4
##  Class :character   Class :character   1st Qu.:    732   1st Qu.:   329
##  Mode  :character   Mode  :character   Median :   2330   Median :  1106
##                                        Mean   :  12864   Mean   :  5974
##                                        3rd Qu.:   6418   3rd Qu.:  3200
##                                        Max.   :1413869   Max.   :771700
##                                        NA's   :1131      NA's   :1131
##      Obama        MalesPer100Females   AgeBelow35      Age35to65
##  Min.   :     4   Min.   : 76.20     Min.   : 4.50   Min.   :18.60
##  1st Qu.:   254   1st Qu.: 94.90     1st Qu.:42.20   1st Qu.:38.00
##  Median :   878   Median : 97.60     Median :45.40   Median :39.80
##  Mean   :  6178   Mean   : 99.08     Mean   :45.66   Mean   :39.59
##  3rd Qu.:  2749   3rd Qu.:100.50     3rd Qu.:48.70   3rd Qu.:41.40
##  Max.   :743686   Max.   :200.90     Max.   :75.80   Max.   :79.20
##  NA's   :1131
##  Age65andAbove       White            Black           Asian
##  Min.   : 2.20   Min.   :  6.70   Min.   : 0.10   Min.   : 0.100
##  1st Qu.:12.20   1st Qu.: 81.90   1st Qu.: 0.60   1st Qu.: 0.300
##  Median :14.40   Median : 93.70   Median : 2.60   Median : 0.500
##  Mean   :14.76   Mean   : 86.76   Mean   : 9.90   Mean   : 1.065
##  3rd Qu.:17.00   3rd Qu.: 97.50   3rd Qu.:11.93   3rd Qu.: 0.900
##  Max.   :33.70   Max.   :100.00   Max.   :86.00   Max.   :46.600
##                                   NA's   :80      NA's   :94
##  AmericanIndian      Hawaiian           Hispanic        HighSchool
##  Min.   : 0.100   Min.   : 0.00000   Min.   : 0.100   Min.   :34.7
##  1st Qu.: 0.200   1st Qu.: 0.00000   1st Qu.: 1.200   1st Qu.:70.6
##  Median : 0.400   Median : 0.00000   Median : 2.400   Median :78.7
##  Mean   : 1.689   Mean   : 0.07706   Mean   : 7.349   Mean   :76.9
##  3rd Qu.: 0.800   3rd Qu.: 0.10000   3rd Qu.: 6.800   3rd Qu.:83.6
##  Max.   :91.800   Max.   :30.60000   Max.   :97.600   Max.   :97.0
##  NA's   :99                                          NA's   :1
##    Bachelors        Poverty      IncomeAbove75K   MedianIncome
##  Min.   : 4.90   Min.   : 2.60   Min.   : 1.50   Min.   :16868
##  1st Qu.:11.10   1st Qu.:10.10   1st Qu.: 9.20   1st Qu.:32314
##  Median :14.30   Median :13.20   Median :11.90   Median :37272
##  Mean   :16.45   Mean   :13.95   Mean   :14.02   Mean   :39107
##  3rd Qu.:19.30   3rd Qu.:16.90   3rd Qu.:16.40   3rd Qu.:43430
##  Max.   :63.70   Max.   :39.40   Max.   :57.30   Max.   :98245
##  NA's   :1       NA's   :1       NA's   :2       NA's   :1
##  AverageIncome    UnemployRate      ManfEmploy      SpeakingNonEnglish
##  Min.   : 5148   Min.   : 1.500   Min.   : 0.2024   Min.   : 0.400
##  1st Qu.:23250   1st Qu.: 3.800   1st Qu.: 4.7194   1st Qu.: 3.000
##  Median :26222   Median : 4.700   Median : 8.9509   Median : 4.600
##  Mean   :27281   Mean   : 4.899   Mean   :10.6564   Mean   : 8.617
##  3rd Qu.:29840   3rd Qu.: 5.700   3rd Qu.:14.9613   3rd Qu.: 8.600
##  Max.   :93377   Max.   :15.300   Max.   :57.2291   Max.   :92.100
##  NA's   :30      NA's   :1        NA's   :293       NA's   :1
```

```
##     Medicare        MedicareRate   SocialSecurity    SocialSecurityRate
##  Min.   :      7   Min.   :   60   Min.   :     20   Min.   : 1518
##  1st Qu.:   2162   1st Qu.:14108   1st Qu.:   2548   1st Qu.:16827
##  Median :   4514   Median :16863   Median :   5360   Median :20040
##  Mean   :  13724   Mean   :17088   Mean   :  15616   Mean   :19939
##  3rd Qu.:  10613   3rd Qu.:19854   3rd Qu.:  12845   3rd Qu.:23032
##  Max.   :1059297   Max.   :60845   Max.   :1047590   Max.   :41943
##  NA's   :1         NA's   :1       NA's   :1         NA's   :1
##  RetiredWorkers    Disabilities      DisabilitiesRate   Homeowner
##  Min.   :    15   Min.   :     4.0   Min.   :  125    Min.   :19.60
##  1st Qu.:  1490   1st Qu.:   274.8   1st Qu.: 1380    1st Qu.:70.50
##  Median :  3095   Median :   698.5   Median : 2192    Median :75.20
##  Mean   :  9869   Mean   :  2381.4   Mean   : 2693    Mean   :73.89
##  3rd Qu.:  7722   3rd Qu.:  1671.2   3rd Qu.: 3436    3rd Qu.:78.90
##  Max.   :681035   Max.   :400125.0   Max.   :20185    Max.   :89.60
##  NA's   :1        NA's   :8          NA's   :8        NA's   :2
##  SameHouse1995and2000      Pop            PopDensity          LandArea
##  Min.   :15.40        Min.   :     60   Min.   :    0.10   Min.   :    2
##  1st Qu.:54.30        1st Qu.:  12335   1st Qu.:   19.80   1st Qu.:  434
##  Median :59.40        Median :  26968   Median :   47.25   Median :  626
##  Mean   :58.74        Mean   :  99226   Mean   :  265.18   Mean   : 1004
##  3rd Qu.:64.00        3rd Qu.:  68372   3rd Qu.:  118.03   3rd Qu.:  950
##  Max.   :90.50        Max.   :9948081   Max.   :70190.80   Max.   :20105
##  NA's   :1                                                 NA's   :1
##     FarmArea
##  Min.   :   1.0
##  1st Qu.:  84.0
##  Median : 182.0
##  Mean   : 293.7
##  3rd Qu.: 348.0
##  Max.   :4595.0
##  NA's   :87
```

```
# Get the structure of the data
str(clinton_obama)
```

```
## 'data.frame':    2868 obs. of  41 variables:
##  $ County              : chr  "Adair" "Adams" "Allamakee" "Appanoose" ...
##  $ State               : chr  "IA" "IA" "IA" "IA" ...
##  $ Region              : chr  "Midwest" "Midwest" "Midwest" "Midwest" ...
##  $ FIPS                : int  19001 19003 19005 19007 19009 19011 19013 19015 19017 19019 ...
##  $ ElectionDate        : chr  "2008-01-03" "2008-01-03" "2008-01-03" "2008-01-03" ...
##  $ ElectionType        : chr  "Caucuses" "Caucuses" "Caucuses" "Caucuses" ...
##  $ TotalVote           : int  75 50 80 60 48 80 420 140 100 150 ...
##  $ Clinton             : int  22 18 25 17 16 23 117 49 28 47 ...
##  $ Obama               : int  24 7 33 10 17 23 179 43 35 53 ...
##  $ MalesPer100Females  : num  96.7 96.8 104.5 94 94.7 ...
##  $ AgeBelow35          : num  37.7 37.1 41.5 42 37.4 45.1 49.6 43.7 44.6 45.9 ...
##  $ Age35to65           : num  40.3 40.8 40.8 39.2 40.2 40.7 36 40.5 38.9 39.6 ...
##  $ Age65andAbove       : num  21.9 22.1 17.7 18.9 22.5 14.3 14.3 15.7 16.5 14.4 ...
##  $ White               : num  99.2 99.2 99 98.5 99.3 99.1 89.6 98.6 98.4 98.6 ...
##  $ Black               : num  0.1 NA 0.2 0.7 0.2 0.2 7.8 0.7 0.5 0.3 ...
##  $ Asian               : num  0.5 0.2 0.2 0.4 0.2 0.2 1.2 0.3 0.6 0.4 ...
##  $ AmericanIndian      : num  NA 0.5 0.3 0.1 0.1 0.1 0.2 0.2 0.1 0.2 ...
```

```
##  $ Hawaiian            : num  0 0 0 0 0 0 0.1 0 0 0 ...
##  $ Hispanic            : num  0.8 0.6 6.9 1.1 0.6 0.6 2.5 1 0.7 0.7 ...
##  $ HighSchool          : num  87.8 84.5 81.4 81.4 82.5 87.8 86.5 89 87.7 84.6 ...
##  $ Bachelors           : num  11.2 12 14.4 12.2 12.3 13.9 23 18.8 21.5 12.7 ...
##  $ Poverty             : num  9.7 10.6 10.8 14.9 9.2 7.6 13.7 8.8 7 9.9 ...
##  $ IncomeAbove75K      : num  8.6 7.8 10.7 6.6 9.7 15.9 15.6 14.2 15.7 13.6 ...
##  $ MedianIncome        : int  39568 35296 36822 31764 38941 48713 40502 45009 46665 43627 ...
##  $ AverageIncome       : int  29222 27297 25579 23705 31100 28649 30406 31499 32245 27778 ...
##  $ UnemployRate        : num  3.2 4.6 4.6 5.4 4.1 4 3.9 3 3.2 4.2 ...
##  $ ManfEmploy          : num  12.92 6.96 16.47 15.08 8.33 ...
##  $ SpeakingNonEnglish  : num  2 1.6 8.2 2.5 2.7 2.8 6.5 2.6 3.8 6 ...
##  $ Medicare            : int  1551 978 2838 3072 1472 4055 20362 4431 4356 3523 ...
##  $ MedicareRate        : int  19735 22936 19294 22479 22797 15019 16174 16657 18398 16761 ...
##  $ SocialSecurity      : int  1825 1110 3270 3415 1740 4825 23100 5370 4840 3885 ...
##  $ SocialSecurityRate  : int  23222 26032 22231 24989 26947 17870 18349 20186 20442 18483 ...
##  $ RetiredWorkers      : int  1230 720 2245 2050 1150 3105 14960 3490 3280 2485 ...
##  $ Disabilities        : int  90 78 186 453 79 269 2835 359 170 277 ...
##  $ DisabilitiesRate    : int  1145 1829 1265 3315 1223 996 2252 1350 718 1318 ...
##  $ Homeowner           : num  75 74.7 76.4 74 79.2 79.4 68.9 75.8 78.2 78.2 ...
##  $ SameHouse1995and2000: num  66 65.4 64.1 60.2 70.4 60.8 54.5 60.6 62 62.9 ...
##  $ Pop                 : int  7714 4192 14796 13422 6278 26962 126106 26584 23837 21045 ...
##  $ PopDensity          : num  13.5 9.9 23.1 27 14.2 ...
##  $ LandArea            : int  570 425 659 516 444 718 572 574 440 573 ...
##  $ FarmArea            : int  373 238 326 236 261 401 275 313 255 340 ...
```

```
# Correctly identify categorical variables as such
clinton_obama$County <- as.factor(clinton_obama$County)
clinton_obama$State <- as.factor(clinton_obama$State)
clinton_obama$Region <- as.factor(clinton_obama$Region)
clinton_obama$ElectionType <- as.factor(clinton_obama$ElectionType)
str(clinton_obama)
```

```
## 'data.frame':    2868 obs. of  41 variables:
##  $ County              : Factor w/ 1691 levels "Abbeville","Acadia",..: 5 6 20 41 68 125 139 157 179
##  $ State               : Factor w/ 46 levels "AL","AR","AZ",..: 11 11 11 11 11 11 11 11 11 11 ...
##  $ Region              : Factor w/ 4 levels "Midwest","Northeast",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ FIPS                : int  19001 19003 19005 19007 19009 19011 19013 19015 19017 19019 ...
##  $ ElectionDate        : chr  "2008-01-03" "2008-01-03" "2008-01-03" "2008-01-03" ...
##  $ ElectionType        : Factor w/ 2 levels "Caucuses","Primary": 1 1 1 1 1 1 1 1 1 1 ...
##  $ TotalVote           : int  75 50 80 60 48 80 420 140 100 150 ...
##  $ Clinton             : int  22 18 25 17 16 23 117 49 28 47 ...
##  $ Obama               : int  24 7 33 10 17 23 179 43 35 53 ...
##  $ MalesPer100Females  : num  96.7 96.8 104.5 94 94.7 ...
##  $ AgeBelow35          : num  37.7 37.1 41.5 42 37.4 45.1 49.6 43.7 44.6 45.9 ...
##  $ Age35to65           : num  40.3 40.8 40.8 39.2 40.2 40.7 36 40.5 38.9 39.6 ...
##  $ Age65andAbove       : num  21.9 22.1 17.7 18.9 22.5 14.3 14.3 15.7 16.5 14.4 ...
##  $ White               : num  99.2 99.2 99 98.5 99.3 99.1 89.6 98.6 98.4 98.6 ...
##  $ Black               : num  0.1 NA 0.2 0.7 0.2 0.2 7.8 0.7 0.5 0.3 ...
##  $ Asian               : num  0.5 0.2 0.2 0.4 0.2 0.2 1.2 0.3 0.6 0.4 ...
##  $ AmericanIndian      : num  NA 0.5 0.3 0.1 0.1 0.1 0.2 0.2 0.1 0.2 ...
##  $ Hawaiian            : num  0 0 0 0 0 0 0.1 0 0 0 ...
##  $ Hispanic            : num  0.8 0.6 6.9 1.1 0.6 0.6 2.5 1 0.7 0.7 ...
##  $ HighSchool          : num  87.8 84.5 81.4 81.4 82.5 87.8 86.5 89 87.7 84.6 ...
##  $ Bachelors           : num  11.2 12 14.4 12.2 12.3 13.9 23 18.8 21.5 12.7 ...
```

```
##  $ Poverty             : num  9.7 10.6 10.8 14.9 9.2 7.6 13.7 8.8 7 9.9 ...
##  $ IncomeAbove75K      : num  8.6 7.8 10.7 6.6 9.7 15.9 15.6 14.2 15.7 13.6 ...
##  $ MedianIncome        : int  39568 35296 36822 31764 38941 48713 40502 45009 46665 43627 ...
##  $ AverageIncome       : int  29222 27297 25579 23705 31100 28649 30406 31499 32245 27778 ...
##  $ UnemployRate        : num  3.2 4.6 4.6 5.4 4.1 4 3.9 3 3.2 4.2 ...
##  $ ManfEmploy          : num  12.92 6.96 16.47 15.08 8.33 ...
##  $ SpeakingNonEnglish  : num  2 1.6 8.2 2.5 2.7 2.8 6.5 2.6 3.8 6 ...
##  $ Medicare            : int  1551 978 2838 3072 1472 4055 20362 4431 4356 3523 ...
##  $ MedicareRate        : int  19735 22936 19294 22479 22797 15019 16174 16657 18398 16761 ...
##  $ SocialSecurity      : int  1825 1110 3270 3415 1740 4825 23100 5370 4840 3885 ...
##  $ SocialSecurityRate  : int  23222 26032 22231 24989 26947 17870 18349 20186 20442 18483 ...
##  $ RetiredWorkers      : int  1230 720 2245 2050 1150 3105 14960 3490 3280 2485 ...
##  $ Disabilities        : int  90 78 186 453 79 269 2835 359 170 277 ...
##  $ DisabilitiesRate    : int  1145 1829 1265 3315 1223 996 2252 1350 718 1318 ...
##  $ Homeowner           : num  75 74.7 76.4 74 79.2 79.4 68.9 75.8 78.2 78.2 ...
##  $ SameHouse1995and2000: num  66 65.4 64.1 60.2 70.4 60.8 54.5 60.6 62 62.9 ...
##  $ Pop                 : int  7714 4192 14796 13422 6278 26962 126106 26584 23837 21045 ...
##  $ PopDensity          : num  13.5 9.9 23.1 27 14.2 ...
##  $ LandArea            : int  570 425 659 516 444 718 572 574 440 573 ...
##  $ FarmArea            : int  373 238 326 236 261 401 275 313 255 340 ...
```

Missing data

```
meanBlack <- mean(clinton_obama$Black[!is.na(clinton_obama$Black)])
clinton_obama$Black <- clinton_obama$Black
clinton_obama$Black[is.na(clinton_obama$Black)] <- meanBlack

meanAsian <- mean(clinton_obama$Asian[!is.na(clinton_obama$Asian)])
clinton_obama$Asian <- clinton_obama$Asian
clinton_obama$Asian[is.na(clinton_obama$Asian)] <- meanAsian

meanAmericanIndian <- mean(clinton_obama$AmericanIndian[!is.na(clinton_obama$AmericanIndian)])
clinton_obama$AmericanIndian <- clinton_obama$AmericanIndian
clinton_obama$AmericanIndian[is.na(clinton_obama$AmericanIndian)] <- meanAmericanIndian

meanHighSchool <- mean(clinton_obama$HighSchool[!is.na(clinton_obama$HighSchool)])
clinton_obama$HighSchool <- clinton_obama$HighSchool
clinton_obama$HighSchool[is.na(clinton_obama$HighSchool)] <- meanHighSchool

meanBachelors <- mean(clinton_obama$Bachelors[!is.na(clinton_obama$Bachelors)])
clinton_obama$Bachelors <- clinton_obama$Bachelors
clinton_obama$Bachelors[is.na(clinton_obama$Bachelors)] <- meanBachelors

meanPoverty <- mean(clinton_obama$Poverty[!is.na(clinton_obama$Poverty)])
clinton_obama$Poverty <- clinton_obama$Poverty
clinton_obama$Poverty[is.na(clinton_obama$Poverty)] <- meanPoverty

meanIncomeAbove75K <- mean(clinton_obama$IncomeAbove75K[!is.na(clinton_obama$IncomeAbove75K)])
clinton_obama$IncomeAbove75K <- clinton_obama$IncomeAbove75K
clinton_obama$IncomeAbove75K[is.na(clinton_obama$IncomeAbove75K)] <- meanIncomeAbove75K

meanMedianIncome <- mean(clinton_obama$MedianIncome[!is.na(clinton_obama$MedianIncome)])
```

```
clinton_obama$MedianIncome <- clinton_obama$MedianIncome
clinton_obama$MedianIncome[is.na(clinton_obama$MedianIncome)] <- meanMedianIncome

meanAverageIncome <- mean(clinton_obama$AverageIncome[!is.na(clinton_obama$AverageIncome)])
clinton_obama$AverageIncome <- clinton_obama$AverageIncome
clinton_obama$AverageIncome[is.na(clinton_obama$AverageIncome)] <- meanAverageIncome

meanUnemployRate <- mean(clinton_obama$UnemployRate[!is.na(clinton_obama$UnemployRate)])
clinton_obama$UnemployRate <- clinton_obama$UnemployRate
clinton_obama$UnemployRate[is.na(clinton_obama$UnemployRate)] <- meanUnemployRate

meanManfEmploy <- mean(clinton_obama$ManfEmploy[!is.na(clinton_obama$ManfEmploy)])
clinton_obama$ManfEmploy <- clinton_obama$ManfEmploy
clinton_obama$ManfEmploy[is.na(clinton_obama$ManfEmploy)] <- meanManfEmploy

meanSpeakingNonEnglish <- mean(clinton_obama$SpeakingNonEnglish[!is.na(clinton_obama$SpeakingNonEnglish
clinton_obama$SpeakingNonEnglish <- clinton_obama$SpeakingNonEnglish
clinton_obama$SpeakingNonEnglish[is.na(clinton_obama$SpeakingNonEnglish)] <- meanSpeakingNonEnglish

meanMedicare <- mean(clinton_obama$Medicare[!is.na(clinton_obama$Medicare)])
clinton_obama$Medicare <- clinton_obama$Medicare
clinton_obama$Medicare[is.na(clinton_obama$Medicare)] <- meanMedicare

meanMedicareRate <- mean(clinton_obama$MedicareRate[!is.na(clinton_obama$MedicareRate)])
clinton_obama$MedicareRate <- clinton_obama$MedicareRate
clinton_obama$MedicareRate[is.na(clinton_obama$MedicareRate)] <- meanMedicareRate

meanSocialSecurity <- mean(clinton_obama$SocialSecurity[!is.na(clinton_obama$SocialSecurity)])
clinton_obama$SocialSecurity <- clinton_obama$SocialSecurity
clinton_obama$SocialSecurity[is.na(clinton_obama$SocialSecurity)] <- meanSocialSecurity

meanSocialSecurityRate <- mean(clinton_obama$SocialSecurityRate[!is.na(clinton_obama$SocialSecurityRate
clinton_obama$SocialSecurityRate <- clinton_obama$SocialSecurityRate
clinton_obama$SocialSecurityRate[is.na(clinton_obama$SocialSecurity)] <- meanSocialSecurityRate

meanRetiredWorkers <- mean(clinton_obama$RetiredWorkers[!is.na(clinton_obama$RetiredWorkers)])
clinton_obama$RetiredWorkers <- clinton_obama$RetiredWorkers
clinton_obama$RetiredWorkers[is.na(clinton_obama$RetiredWorkers)] <- meanRetiredWorkers

meanDisabilities <- mean(clinton_obama$Disabilities[!is.na(clinton_obama$Disabilities)])
clinton_obama$Disabilities <- clinton_obama$Disabilities
clinton_obama$Disabilities[is.na(clinton_obama$Disabilities)] <- meanDisabilities

meanDisabilitiesRate <- mean(clinton_obama$DisabilitiesRate[!is.na(clinton_obama$DisabilitiesRate)])
clinton_obama$DisabilitiesRate <- clinton_obama$DisabilitiesRate
clinton_obama$DisabilitiesRate[is.na(clinton_obama$Disabilities)] <- meanDisabilitiesRate

meanHomeowner <- mean(clinton_obama$Homeowner[!is.na(clinton_obama$Homeowner)])
clinton_obama$Homeowner <- clinton_obama$Homeowner
clinton_obama$Homeowner[is.na(clinton_obama$Homeowner)] <- meanHomeowner

meanSameHouse1995and2000 <- mean(clinton_obama$SameHouse1995and2000[!is.na(clinton_obama$SameHouse1995a
clinton_obama$SameHouse1995and2000 <- clinton_obama$SameHouse1995and2000
```

```
clinton_obama$SameHouse1995and2000[is.na(clinton_obama$SameHouse1995and2000)] <- meanSameHouse1995and20

meanLandArea <- mean(clinton_obama$LandArea[!is.na(clinton_obama$LandArea)])
clinton_obama$LandArea <- clinton_obama$LandArea
clinton_obama$LandArea[is.na(clinton_obama$LandArea)] <- meanLandArea

meanFarmArea <- mean(clinton_obama$FarmArea[!is.na(clinton_obama$FarmArea)])
clinton_obama$FarmArea <- clinton_obama$FarmArea
clinton_obama$FarmArea[is.na(clinton_obama$FarmArea)] <- meanFarmArea
```

```
summary(clinton_obama)
```

```
##         County          State            Region          FIPS
##  Washington:  30   TX     : 251   Midwest  : 814   Min.   : 1001
##  Franklin  :  25   GA     : 159   Northeast: 217   1st Qu.:18102
##  Jefferson :  25   VA     : 134   South    :1419   Median :30110
##  Lincoln   :  23   KY     : 120   West     : 418   Mean   :31029
##  Jackson   :  22   MO     : 115                    3rd Qu.:46124
##  Madison   :  20   IL     : 102                    Max.   :56045
##  (Other)   :2723   (Other):1987
##  ElectionDate         ElectionType    TotalVote          Clinton
##  Length:2868        Caucuses: 310   Min.   :     13   Min.   :     4
##  Class :character   Primary :2558   1st Qu.:    732   1st Qu.:   329
##  Mode  :character                   Median :   2330   Median :  1106
##                                     Mean   :  12864   Mean   :  5974
##                                     3rd Qu.:   6418   3rd Qu.:  3200
##                                     Max.   :1413869   Max.   :771700
##                                     NA's   :1131      NA's   :1131
##      Obama        MalesPer100Females   AgeBelow35      Age35to65
##  Min.   :     4   Min.   : 76.20     Min.   : 4.50   Min.   :18.60
##  1st Qu.:   254   1st Qu.: 94.90     1st Qu.:42.20   1st Qu.:38.00
##  Median :   878   Median : 97.60     Median :45.40   Median :39.80
##  Mean   :  6178   Mean   : 99.08     Mean   :45.66   Mean   :39.59
##  3rd Qu.:  2749   3rd Qu.:100.50     3rd Qu.:48.70   3rd Qu.:41.40
##  Max.   :743686   Max.   :200.90     Max.   :75.80   Max.   :79.20
##  NA's   :1131
##  Age65andAbove       White            Black          Asian
##  Min.   : 2.20   Min.   :  6.70   Min.   : 0.1   Min.   : 0.100
##  1st Qu.:12.20   1st Qu.: 81.90   1st Qu.: 0.6   1st Qu.: 0.300
##  Median :14.40   Median : 93.70   Median : 3.0   Median : 0.500
##  Mean   :14.76   Mean   : 86.76   Mean   : 9.9   Mean   : 1.065
##  3rd Qu.:17.00   3rd Qu.: 97.50   3rd Qu.:11.4   3rd Qu.: 1.000
##  Max.   :33.70   Max.   :100.00   Max.   :86.0   Max.   :46.600
##
##  AmericanIndian      Hawaiian           Hispanic        HighSchool
##  Min.   : 0.100   Min.   : 0.00000   Min.   : 0.100   Min.   :34.7
##  1st Qu.: 0.200   1st Qu.: 0.00000   1st Qu.: 1.200   1st Qu.:70.6
##  Median : 0.400   Median : 0.00000   Median : 2.400   Median :78.7
##  Mean   : 1.689   Mean   : 0.07706   Mean   : 7.349   Mean   :76.9
##  3rd Qu.: 0.900   3rd Qu.: 0.10000   3rd Qu.: 6.800   3rd Qu.:83.6
##  Max.   :91.800   Max.   :30.60000   Max.   :97.600   Max.   :97.0
##
##    Bachelors        Poverty      IncomeAbove75K   MedianIncome
```

```
##  Min.   : 4.90   Min.   : 2.60   Min.   : 1.50   Min.   :16868
##  1st Qu.:11.10   1st Qu.:10.10   1st Qu.: 9.20   1st Qu.:32316
##  Median :14.30   Median :13.20   Median :11.90   Median :37282
##  Mean   :16.45   Mean   :13.95   Mean   :14.02   Mean   :39107
##  3rd Qu.:19.30   3rd Qu.:16.90   3rd Qu.:16.40   3rd Qu.:43425
##  Max.   :63.70   Max.   :39.40   Max.   :57.30   Max.   :98245
##
##  AverageIncome    UnemployRate      ManfEmploy      SpeakingNonEnglish
##  Min.   : 5148   Min.   : 1.500   Min.   : 0.2024   Min.   : 0.400
##  1st Qu.:23287   1st Qu.: 3.800   1st Qu.: 5.1209   1st Qu.: 3.000
##  Median :26292   Median : 4.700   Median :10.1175   Median : 4.600
##  Mean   :27281   Mean   : 4.899   Mean   :10.6564   Mean   : 8.617
##  3rd Qu.:29812   3rd Qu.: 5.700   3rd Qu.:14.0054   3rd Qu.: 8.604
##  Max.   :93377   Max.   :15.300   Max.   :57.2291   Max.   :92.100
##
##     Medicare        MedicareRate   SocialSecurity    SocialSecurityRate
##  Min.   :      7   Min.   :   60   Min.   :     20   Min.   : 1518
##  1st Qu.:   2163   1st Qu.:14109   1st Qu.:   2549   1st Qu.:16827
##  Median :   4514   Median :16868   Median :   5360   Median :20040
##  Mean   :  13724   Mean   :17088   Mean   :  15616   Mean   :19939
##  3rd Qu.:  10619   3rd Qu.:19850   3rd Qu.:  12861   3rd Qu.:23032
##  Max.   :1059297   Max.   :60845   Max.   :1047590   Max.   :41943
##                                                      NA's   :1
##  RetiredWorkers    Disabilities      DisabilitiesRate   Homeowner
##  Min.   :    15   Min.   :     4.0   Min.   :  125    Min.   :19.60
##  1st Qu.:  1490   1st Qu.:   276.0   1st Qu.: 1380    1st Qu.:70.50
##  Median :  3098   Median :   700.5   Median : 2192    Median :75.20
##  Mean   :  9869   Mean   :  2381.4   Mean   : 2693    Mean   :73.89
##  3rd Qu.:  7728   3rd Qu.:  1690.2   3rd Qu.: 3436    3rd Qu.:78.90
##  Max.   :681035   Max.   :400125.0   Max.   :20185    Max.   :89.60
##                                      NA's   :8
##  SameHouse1995and2000      Pop            PopDensity          LandArea
##  Min.   :15.40        Min.   :     60   Min.   :    0.10   Min.   :    2
##  1st Qu.:54.30        1st Qu.:  12335   1st Qu.:   19.80   1st Qu.:  434
##  Median :59.40        Median :  26968   Median :   47.25   Median :  626
##  Mean   :58.74        Mean   :  99226   Mean   :  265.18   Mean   : 1004
##  3rd Qu.:64.00        3rd Qu.:  68372   3rd Qu.:  118.03   3rd Qu.:  951
##  Max.   :90.50        Max.   :9948081   Max.   :70190.80   Max.   :20105
##
##     FarmArea
##  Min.   :   1.0
##  1st Qu.:  87.0
##  Median : 192.0
##  Mean   : 293.7
##  3rd Qu.: 344.0
##  Max.   :4595.0
##
```

Analysis

Independent vars

```
# Create some possible independent variables (things we might like to predict in a regression using the
clinton_obama$Obama_margin <- clinton_obama$Obama - clinton_obama$Clinton
clinton_obama$Obama_margin_percent <- clinton_obama$Obama_margin/clinton_obama$TotalVote
clinton_obama$Obama_wins <- ifelse(clinton_obama$Obama_margin >0, 1,0)
names(clinton_obama)
```

```
##  [1] "County"               "State"                "Region"
##  [4] "FIPS"                 "ElectionDate"         "ElectionType"
##  [7] "TotalVote"            "Clinton"              "Obama"
## [10] "MalesPer100Females"   "AgeBelow35"           "Age35to65"
## [13] "Age65andAbove"        "White"                "Black"
## [16] "Asian"                "AmericanIndian"       "Hawaiian"
## [19] "Hispanic"             "HighSchool"           "Bachelors"
## [22] "Poverty"              "IncomeAbove75K"       "MedianIncome"
## [25] "AverageIncome"        "UnemployRate"         "ManfEmploy"
## [28] "SpeakingNonEnglish"   "Medicare"             "MedicareRate"
## [31] "SocialSecurity"       "SocialSecurityRate"   "RetiredWorkers"
## [34] "Disabilities"         "DisabilitiesRate"     "Homeowner"
## [37] "SameHouse1995and2000" "Pop"                  "PopDensity"
## [40] "LandArea"             "FarmArea"             "Obama_margin"
## [43] "Obama_margin_percent" "Obama_wins"
```

```
table(clinton_obama$Obama_wins)
```

```
##
##   0   1
## 913 824
```

```
percent_Clin <- 913/(913+824)*100
percent_Obama <- 824/(824+913)*100

percent_Clin
```

```
## [1] 52.56189
```

```
percent_Obama
```

```
## [1] 47.43811
```

Training, test, validation sets

```
# Partition Data - counties that have voted = Training, counties that have not voted = Test
set.seed(1)
```

```r
voted <- which(!is.na(clinton_obama$TotalVote))
train.df <- clinton_obama[voted, ]
test.df <- clinton_obama[-voted, ]

# training 60% and validation 40%
train_index <- sample(c(1:dim(train.df)[1]), round(dim(train.df)[1]*0.6))
train.df <- train.df[train_index, ]
valid.df <- train.df[-train_index, ]
```

Predicting who will win

Logistic Regresion

```r
options(scipen=999)
clinton_obama.glm <- glm(Obama_wins ~ .-County -State -ElectionDate -TotalVote -Clinton -Obama -FIPS -O
summary(clinton_obama.glm)
```

```
##
## Call:
## glm(formula = Obama_wins ~ . - County - State - ElectionDate -
##     TotalVote - Clinton - Obama - FIPS - Obama_margin_percent -
##     Obama_margin, family = "binomial", data = train.df, trace = 0)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -3.11817  -0.55574  -0.09616   0.55289   2.44878
##
## Coefficients:
##                          Estimate    Std. Error z value         Pr(>|z|)
## (Intercept)          -68.430852042 102.802216894  -0.666         0.505631
## RegionNortheast       -2.619577769   0.479566849  -5.462 0.00000004697854 ***
## RegionSouth           -2.690447688   0.410125575  -6.560 0.00000000005379 ***
## RegionWest             1.933180546   0.411038186   4.703 0.00000256158821 ***
## ElectionTypePrimary   -1.416076684   0.276074622  -5.129 0.00000029078171 ***
## MalesPer100Females    -0.011759176   0.011942775  -0.985         0.324807
## AgeBelow35             1.049759702   1.027904437   1.021         0.307130
## Age35to65              1.061352387   1.026293527   1.034         0.301061
## Age65andAbove          0.820439932   1.026575689   0.799         0.424174
## White                 -0.249106037   0.056133175  -4.438 0.00000908962911 ***
## Black                 -0.055922184   0.052528350  -1.065         0.287053
## Asian                 -0.471450752   0.098289464  -4.797 0.00000161418197 ***
## AmericanIndian        -0.241707989   0.064854074  -3.727         0.000194 ***
## Hawaiian              -0.939119458   0.892025786  -1.053         0.292435
## Hispanic              -0.029336150   0.027787730  -1.056         0.291095
## HighSchool            -0.092526873   0.032029891  -2.889         0.003868 **
## Bachelors              0.236718233   0.034690902   6.824 0.00000000000888 ***
## Poverty               -0.266316585   0.070594796  -3.772         0.000162 ***
## IncomeAbove75K        -0.035601100   0.056047350  -0.635         0.525300
## MedianIncome          -0.000051357   0.000044081  -1.165         0.243993
## AverageIncome         -0.000012231   0.000036414  -0.336         0.736960
## UnemployRate          -0.076625075   0.092833047  -0.825         0.409140
```

```
## ManfEmploy            -0.006350655   0.014952441  -0.425         0.671038
## SpeakingNonEnglish    -0.021535297   0.035876954  -0.600         0.548337
## Medicare               0.000071464   0.000070729   1.010         0.312308
## MedicareRate           0.000010054   0.000040194   0.250         0.802483
## SocialSecurity        -0.000197722   0.000073541  -2.689         0.007175 **
## SocialSecurityRate     0.000022841   0.000088744   0.257         0.796889
## RetiredWorkers         0.000157306   0.000095106   1.654         0.098124 .
## Disabilities           0.000074871   0.000058208   1.286         0.198347
## DisabilitiesRate      -0.000491616   0.000183727  -2.676         0.007455 **
## Homeowner             -0.011932615   0.024436123  -0.488         0.625324
## SameHouse1995and2000   0.086392240   0.025699189   3.362         0.000775 ***
## Pop                    0.000001844   0.000002574   0.716         0.473816
## PopDensity            -0.000259644   0.000128626  -2.019         0.043529 *
## LandArea               0.000081337   0.000083078   0.979         0.327559
## FarmArea              -0.000958023   0.000404896  -2.366         0.017977 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1440.43  on 1040  degrees of freedom
## Residual deviance:  783.78  on 1004  degrees of freedom
##   (1 observation deleted due to missingness)
## AIC: 857.78
##
## Number of Fisher Scoring iterations: 6
```

```
summary(train.df)
```

Double check for NAs

```
##         County        State           Region         FIPS
##  Lincoln   : 13   GA     : 95   Midwest  :289   Min.   : 1003
##  Jefferson : 11   VA     : 76   Northeast: 85   1st Qu.:13118
##  Franklin  : 10   MO     : 73   South    :478   Median :24028
##  Washington: 10   IL     : 63   West     :190   Mean   :25588
##  Jackson   :  9   IA     : 62                   3rd Qu.:36084
##  Union     :  9   TN     : 59                   Max.   :53077
##  (Other)   :980   (Other):614
##  ElectionDate         ElectionType   TotalVote            Clinton
##  Length:1042        Caucuses:177   Min.   :     17.0   Min.   :     4.0
##  Class :character   Primary :865   1st Qu.:    775.5   1st Qu.:   333.2
##  Mode  :character                  Median :   2327.0   Median :  1090.5
##                                    Mean   :  12895.5   Mean   :  5830.4
##                                    3rd Qu.:   6505.0   3rd Qu.:  3483.8
##                                    Max.   :1079079.0   Max.   :314634.0
##
##      Obama          MalesPer100Females   AgeBelow35      Age35to65
##  Min.   :     4.0   Min.   : 76.20     Min.   :29.10   Min.   :18.60
##  1st Qu.:   253.2   1st Qu.: 94.80     1st Qu.:42.10   1st Qu.:38.02
##  Median :   876.5   Median : 97.40     Median :45.40   Median :39.50
```

```
##  Mean   :  6414.3   Mean   : 99.11     Mean   :45.69   Mean   :39.48
##  3rd Qu.:  2840.5   3rd Qu.:100.38     3rd Qu.:48.70   3rd Qu.:41.10
##  Max.   :743686.0   Max.   :200.90     Max.   :72.40   Max.   :56.50
##
##  Age65andAbove       White           Black           Asian
##  Min.   : 3.80   Min.   :19.70   Min.   : 0.10   Min.   : 0.100
##  1st Qu.:12.12   1st Qu.:77.90   1st Qu.: 0.60   1st Qu.: 0.300
##  Median :14.40   Median :92.00   Median : 3.80   Median : 0.500
##  Mean   :14.84   Mean   :84.97   Mean   :11.44   Mean   : 1.244
##  3rd Qu.:17.30   3rd Qu.:97.08   3rd Qu.:15.05   3rd Qu.: 1.100
##  Max.   :33.70   Max.   :99.80   Max.   :79.70   Max.   :32.900
##
##  AmericanIndian      Hawaiian          Hispanic        HighSchool
##  Min.   : 0.100   Min.   :0.00000   Min.   : 0.100   Min.   :53.30
##  1st Qu.: 0.200   1st Qu.:0.00000   1st Qu.: 1.400   1st Qu.:70.92
##  Median : 0.400   Median :0.00000   Median : 2.600   Median :78.90
##  Mean   : 1.641   Mean   :0.06296   Mean   : 6.506   Mean   :77.36
##  3rd Qu.: 0.900   3rd Qu.:0.10000   3rd Qu.: 6.600   3rd Qu.:84.00
##  Max.   :76.800   Max.   :1.40000   Max.   :80.500   Max.   :96.30
##
##    Bachelors        Poverty      IncomeAbove75K  MedianIncome   AverageIncome
##  Min.   : 5.40   Min.   : 3.10   Min.   : 3.5   Min.   :19407   Min.   : 8184
##  1st Qu.:11.12   1st Qu.:10.10   1st Qu.: 9.1   1st Qu.:32459   1st Qu.:22900
##  Median :14.60   Median :13.15   Median :11.9   Median :37511   Median :26108
##  Mean   :17.02   Mean   :13.65   Mean   :14.5   Mean   :39822   Mean   :27605
##  3rd Qu.:19.90   3rd Qu.:16.60   3rd Qu.:16.8   3rd Qu.:44011   3rd Qu.:29960
##  Max.   :57.10   Max.   :36.00   Max.   :54.6   Max.   :94173   Max.   :93377
##
##   UnemployRate      ManfEmploy      SpeakingNonEnglish    Medicare
##  Min.   : 2.200   Min.   : 0.2437   Min.   : 0.400     Min.   :    65
##  1st Qu.: 3.600   1st Qu.: 5.2002   1st Qu.: 2.900     1st Qu.:  2214
##  Median : 4.400   Median : 9.5982   Median : 4.500     Median :  4578
##  Mean   : 4.689   Mean   :10.3890   Mean   : 7.996     Mean   : 15662
##  3rd Qu.: 5.400   3rd Qu.:13.0836   3rd Qu.: 8.300     3rd Qu.: 11247
##  Max.   :14.700   Max.   :57.2291   Max.   :80.500     Max.   :667017
##
##   MedicareRate   SocialSecurity   SocialSecurityRate RetiredWorkers
##  Min.   : 3903   Min.   :    85   Min.   : 1518      Min.   :    55
##  1st Qu.:13992   1st Qu.:  2591   1st Qu.:16658      1st Qu.:  1565
##  Median :16987   Median :  5355   Median :20184      Median :  3135
##  Mean   :17167   Mean   : 17678   Mean   :19961      Mean   : 11346
##  3rd Qu.:19961   3rd Qu.: 13480   3rd Qu.:23149      3rd Qu.:  8280
##  Max.   :58686   Max.   :722160   Max.   :41943      Max.   :468230
##
##   Disabilities      DisabilitiesRate   Homeowner     SameHouse1995and2000
##  Min.   :     4.0   Min.   :  125    Min.   :19.60   Min.   :28.00
##  1st Qu.:   273.5   1st Qu.: 1373    1st Qu.:69.80   1st Qu.:53.42
##  Median :   713.5   Median : 2192    Median :74.90   Median :58.50
##  Mean   :  2635.3   Mean   : 2588    Mean   :73.49   Mean   :57.97
##  3rd Qu.:  1746.8   3rd Qu.: 3439    3rd Qu.:78.60   3rd Qu.:63.10
##  Max.   :154868.0   Max.   :13349    Max.   :89.60   Max.   :78.80
##                     NA's   :1
##       Pop            PopDensity         LandArea          FarmArea
##  Min.   :    578   Min.   :    0.2   Min.   :    6.0   Min.   :   1.0
```

```
##  1st Qu.:  12312   1st Qu.:   20.6   1st Qu.:  441.2   1st Qu.:  82.0
##  Median :  27428   Median :   44.5   Median :  616.5   Median : 194.5
##  Mean   : 114085   Mean   :  405.0   Mean   : 1029.5   Mean   : 266.2
##  3rd Qu.:  72532   3rd Qu.:  124.0   3rd Qu.:  907.0   3rd Qu.: 317.8
##  Max.   :5288655   Max.   :70190.8   Max.   :18661.0   Max.   :3777.0
##
##   Obama_margin      Obama_margin_percent   Obama_wins
##  Min.   :-50300.0   Min.   :-0.79692     Min.   :0.000
##  1st Qu.:  -788.5   1st Qu.:-0.25032     1st Qu.:0.000
##  Median :    -3.0   Median :-0.02000     Median :0.000
##  Mean   :   583.8   Mean   :-0.03348     Mean   :0.475
##  3rd Qu.:   321.8   3rd Qu.: 0.18059     3rd Qu.:1.000
##  Max.   :429052.0   Max.   : 0.69726     Max.   :1.000
##
```

```
clinton_obama_step.glm <- step(clinton_obama.glm, direction="both", trace=0)
summary(clinton_obama_step.glm)
```

Stepwise selection

```
##
## Call:
## glm(formula = Obama_wins ~ Region + ElectionType + AgeBelow35 +
##     Age35to65 + White + Asian + AmericanIndian + Hispanic + HighSchool +
##     Bachelors + Poverty + MedianIncome + SocialSecurity + RetiredWorkers +
##     Disabilities + DisabilitiesRate + SameHouse1995and2000 +
##     PopDensity + FarmArea, family = "binomial", data = train.df,
##     trace = 0)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.2639  -0.5750  -0.1084   0.5803   2.5754
##
## Coefficients:
##                          Estimate  Std. Error z value             Pr(>|z|)
## (Intercept)           9.53128914  5.11009224   1.865             0.062155 .
## RegionNortheast      -2.59497096  0.43971844  -5.901 0.000000003603475316 ***
## RegionSouth          -2.55920874  0.37178948  -6.883 0.000000000005840416 ***
## RegionWest            1.90946490  0.36210259   5.273 0.000000134013848035 ***
## ElectionTypePrimary  -1.45810483  0.25450245  -5.729 0.000000010088336688 ***
## AgeBelow35            0.19375392  0.03874935   5.000 0.000000572752880348 ***
## Age35to65             0.18815341  0.05717320   3.291             0.000999 ***
## White                -0.18659773  0.01611589 -11.578 < 0.0000000000000002 ***
## Asian                -0.43315318  0.05958189  -7.270 0.000000000000359806 ***
## AmericanIndian       -0.18138016  0.02265532  -8.006 0.000000000000001184 ***
## Hispanic             -0.04641877  0.01440247  -3.223             0.001269 **
## HighSchool           -0.08308367  0.02955865  -2.811             0.004942 **
## Bachelors             0.23206869  0.02847697   8.149 0.000000000000000366 ***
## Poverty              -0.28615223  0.06435441  -4.447 0.000008727858481633 ***
## MedianIncome         -0.00008073  0.00002388  -3.380             0.000724 ***
## SocialSecurity       -0.00013793  0.00005679  -2.429             0.015146 *
```

```
## RetiredWorkers        0.00017388  0.00007967   2.182             0.029080 *
## Disabilities          0.00011507  0.00004268   2.696             0.007011 **
## DisabilitiesRate     -0.00041683  0.00015273  -2.729             0.006350 **
## SameHouse1995and2000  0.07738088  0.01980217   3.908 0.000093180276644235 ***
## PopDensity           -0.00024525  0.00011832  -2.073             0.038199 *
## FarmArea             -0.00077452  0.00035661  -2.172             0.029863 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1440.43  on 1040  degrees of freedom
## Residual deviance:  792.15  on 1019  degrees of freedom
##   (1 observation deleted due to missingness)
## AIC: 836.15
##
## Number of Fisher Scoring iterations: 6
```

```
clinton_obama_back.glm  <- step(clinton_obama.glm, direction="backward",  trace=0)
summary(clinton_obama_back.glm)
```

Backward selection

```
##
## Call:
## glm(formula = Obama_wins ~ Region + ElectionType + AgeBelow35 +
##     Age35to65 + White + Asian + AmericanIndian + Hispanic + HighSchool +
##     Bachelors + Poverty + MedianIncome + SocialSecurity + RetiredWorkers +
##     Disabilities + DisabilitiesRate + SameHouse1995and2000 +
##     PopDensity + FarmArea, family = "binomial", data = train.df,
##     trace = 0)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.2639  -0.5750  -0.1084   0.5803   2.5754
##
## Coefficients:
##                         Estimate  Std. Error z value             Pr(>|z|)
## (Intercept)           9.53128914  5.11009224   1.865             0.062155 .
## RegionNortheast      -2.59497096  0.43971844  -5.901 0.000000003603475316 ***
## RegionSouth          -2.55920874  0.37178948  -6.883 0.000000000005840416 ***
## RegionWest            1.90946490  0.36210259   5.273 0.000000134013848035 ***
## ElectionTypePrimary  -1.45810483  0.25450245  -5.729 0.000000010088336688 ***
## AgeBelow35            0.19375392  0.03874935   5.000 0.000000572752880348 ***
## Age35to65             0.18815341  0.05717320   3.291             0.000999 ***
## White                -0.18659773  0.01611589 -11.578 < 0.0000000000000002 ***
## Asian                -0.43315318  0.05958189  -7.270 0.000000000000359806 ***
## AmericanIndian       -0.18138016  0.02265532  -8.006 0.000000000000001184 ***
## Hispanic             -0.04641877  0.01440247  -3.223             0.001269 **
## HighSchool           -0.08308367  0.02955865  -2.811             0.004942 **
## Bachelors             0.23206869  0.02847697   8.149 0.000000000000000366 ***
```

```
## Poverty              -0.28615223  0.06435441  -4.447 0.000008727858481633 ***
## MedianIncome         -0.00008073  0.00002388  -3.380             0.000724 ***
## SocialSecurity       -0.00013793  0.00005679  -2.429             0.015146 *
## RetiredWorkers        0.00017388  0.00007967   2.182             0.029080 *
## Disabilities          0.00011507  0.00004268   2.696             0.007011 **
## DisabilitiesRate     -0.00041683  0.00015273  -2.729             0.006350 **
## SameHouse1995and2000  0.07738088  0.01980217   3.908 0.000093180276644235 ***
## PopDensity           -0.00024525  0.00011832  -2.073             0.038199 *
## FarmArea             -0.00077452  0.00035661  -2.172             0.029863 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1440.43  on 1040  degrees of freedom
## Residual deviance:  792.15  on 1019  degrees of freedom
##   (1 observation deleted due to missingness)
## AIC: 836.15
##
## Number of Fisher Scoring iterations: 6
```

```
clinton_obama_for.glm  <- step(clinton_obama.glm, direction =
"forward",  trace=0)
summary(clinton_obama_for.glm)
```

Forward selection

```
##
## Call:
## glm(formula = Obama_wins ~ (County + State + Region + FIPS +
##     ElectionDate + ElectionType + TotalVote + Clinton + Obama +
##     MalesPer100Females + AgeBelow35 + Age35to65 + Age65andAbove +
##     White + Black + Asian + AmericanIndian + Hawaiian + Hispanic +
##     HighSchool + Bachelors + Poverty + IncomeAbove75K + MedianIncome +
##     AverageIncome + UnemployRate + ManfEmploy + SpeakingNonEnglish +
##     Medicare + MedicareRate + SocialSecurity + SocialSecurityRate +
##     RetiredWorkers + Disabilities + DisabilitiesRate + Homeowner +
##     SameHouse1995and2000 + Pop + PopDensity + LandArea + FarmArea +
##     Obama_margin + Obama_margin_percent) - County - State - ElectionDate -
##     TotalVote - Clinton - Obama - FIPS - Obama_margin_percent -
##     Obama_margin, family = "binomial", data = train.df, trace = 0)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -3.11817  -0.55574  -0.09616   0.55289   2.44878
##
## Coefficients:
##                          Estimate    Std. Error z value         Pr(>|z|)
## (Intercept)          -68.430852042 102.802216894  -0.666         0.505631
## RegionNortheast       -2.619577769   0.479566849  -5.462 0.00000004697854 ***
## RegionSouth           -2.690447688   0.410125575  -6.560 0.00000000005379 ***
```

```
## RegionWest             1.933180546   0.411038186   4.703 0.00000256158821 ***
## ElectionTypePrimary   -1.416076684   0.276074622  -5.129 0.00000029078171 ***
## MalesPer100Females    -0.011759176   0.011942775  -0.985         0.324807
## AgeBelow35             1.049759702   1.027904437   1.021         0.307130
## Age35to65              1.061352387   1.026293527   1.034         0.301061
## Age65andAbove          0.820439932   1.026575689   0.799         0.424174
## White                 -0.249106037   0.056133175  -4.438 0.00000908962911 ***
## Black                 -0.055922184   0.052528350  -1.065         0.287053
## Asian                 -0.471450752   0.098289464  -4.797 0.00000161418197 ***
## AmericanIndian        -0.241707989   0.064854074  -3.727         0.000194 ***
## Hawaiian              -0.939119458   0.892025786  -1.053         0.292435
## Hispanic              -0.029336150   0.027787730  -1.056         0.291095
## HighSchool            -0.092526873   0.032029891  -2.889         0.003868 **
## Bachelors              0.236718233   0.034690902   6.824 0.00000000000888 ***
## Poverty               -0.266316585   0.070594796  -3.772         0.000162 ***
## IncomeAbove75K        -0.035601100   0.056047350  -0.635         0.525300
## MedianIncome          -0.000051357   0.000044081  -1.165         0.243993
## AverageIncome         -0.000012231   0.000036414  -0.336         0.736960
## UnemployRate          -0.076625075   0.092833047  -0.825         0.409140
## ManfEmploy            -0.006350655   0.014952441  -0.425         0.671038
## SpeakingNonEnglish    -0.021535297   0.035876954  -0.600         0.548337
## Medicare               0.000071464   0.000070729   1.010         0.312308
## MedicareRate           0.000010054   0.000040194   0.250         0.802483
## SocialSecurity        -0.000197722   0.000073541  -2.689         0.007175 **
## SocialSecurityRate     0.000022841   0.000088744   0.257         0.796889
## RetiredWorkers         0.000157306   0.000095106   1.654         0.098124 .
## Disabilities           0.000074871   0.000058208   1.286         0.198347
## DisabilitiesRate      -0.000491616   0.000183727  -2.676         0.007455 **
## Homeowner             -0.011932615   0.024436123  -0.488         0.625324
## SameHouse1995and2000   0.086392240   0.025699189   3.362         0.000775 ***
## Pop                    0.000001844   0.000002574   0.716         0.473816
## PopDensity            -0.000259644   0.000128626  -2.019         0.043529 *
## LandArea               0.000081337   0.000083078   0.979         0.327559
## FarmArea              -0.000958023   0.000404896  -2.366         0.017977 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1440.43  on 1040  degrees of freedom
## Residual deviance:  783.78  on 1004  degrees of freedom
##   (1 observation deleted due to missingness)
## AIC: 857.78
##
## Number of Fisher Scoring iterations: 6
```

```
logit.valid.pred <- predict(clinton_obama_step.glm, valid.df, type = "response")
logit.valid.fit <- ifelse(logit.valid.pred >0.5,"1","0")

library(caret)
```

Accuracy

```
## Loading required package: ggplot2
```

```
## Loading required package: lattice
```

```
confusionMatrix(as.factor(logit.valid.fit), as.factor(valid.df$Obama_wins))
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   0   1
##          0 182  30
##          1  32 159
##
##                Accuracy : 0.8462
##                  95% CI : (0.8072, 0.88)
##     No Information Rate : 0.531
##     P-Value [Acc > NIR] : <0.0000000000000002
##
##                   Kappa : 0.6913
##
##  Mcnemar's Test P-Value : 0.8989
##
##             Sensitivity : 0.8505
##             Specificity : 0.8413
##          Pos Pred Value : 0.8585
##          Neg Pred Value : 0.8325
##              Prevalence : 0.5310
##          Detection Rate : 0.4516
##    Detection Prevalence : 0.5261
##       Balanced Accuracy : 0.8459
##
##        'Positive' Class : 0
##
```

```
library(pROC)
```

```
## Type 'citation("pROC")' for a citation.
```

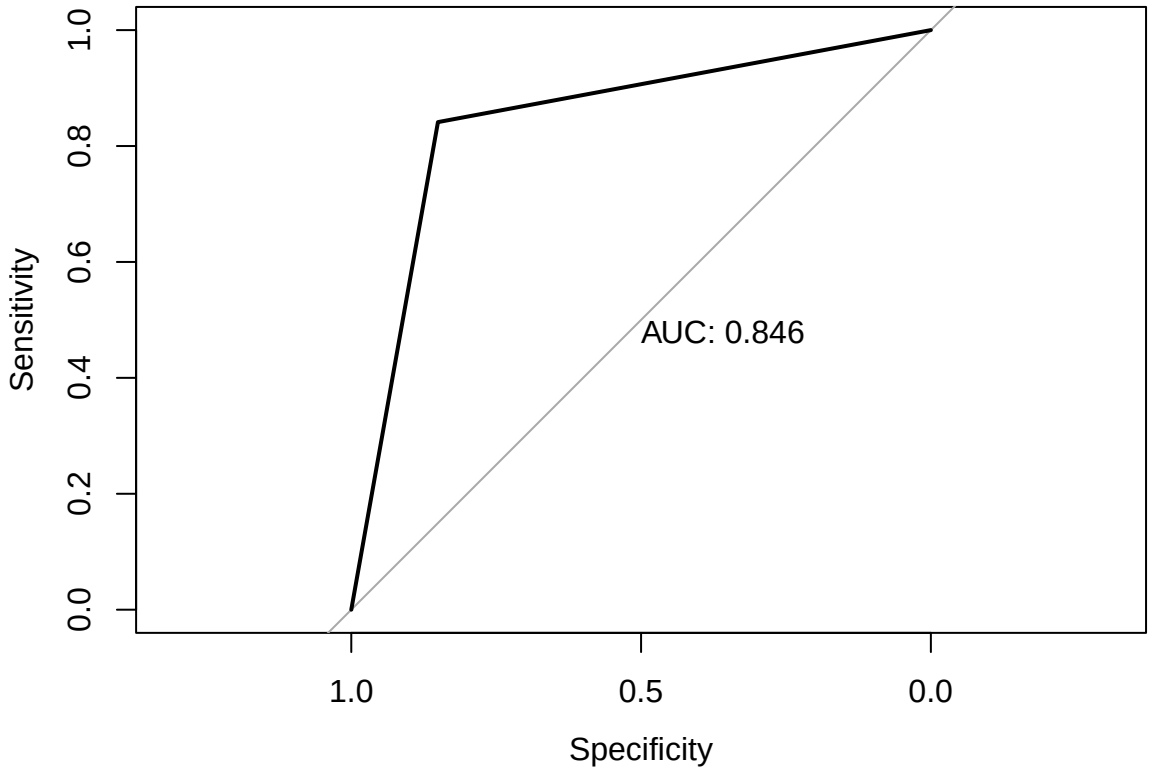
```
##
## Attaching package: 'pROC'
```

```
## The following objects are masked from 'package:stats':
##
##     cov, smooth, var
```

```
test_roc = roc(ifelse(valid.df$Obama_wins=="1",1,0) ~
                ifelse(logit.valid.fit=="1",1,0), plot = TRUE, print.auc = TRUE)
```

```
## Setting levels: control = 0, case = 1
```

```
## Setting direction: controls < cases
```




```
auc(test_roc)
```

```
## Area under the curve: 0.8459
```

```
forecast.Obama_wins <- predict(clinton_obama_step.glm,newdata=test.df,type = "response")
test.df$Obama_win_probability <- forecast.Obama_wins

forecast.Obama_wins<- ifelse(forecast.Obama_wins >0.5,"1","0")
test.df$Obama_wins <- forecast.Obama_wins

valid.df$Obama_win_probability <- valid.df$Obama_margin_percent
train.df$Obama_win_probability <- train.df$Obama_margin_percent

clinton_obama_predicted <- rbind(valid.df, test.df, train.df)

library(dplyr)

clinton_obama_predicted  %>%
  filter(Obama_wins == 1) %>%
  count(Obama_wins)
```

so who wins

```
##   Obama_wins    n
## 1          1 1028
```

```
2868 - 1028
```

```
## [1] 1840
```

```
# 1840
```

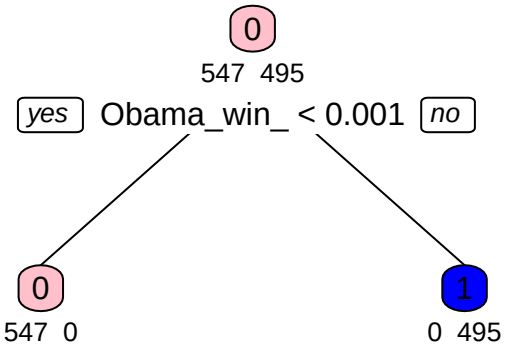
Classification tree

```
library(rpart)
#install.packages("rpart")
# install.packages("rpart.plot")
library(rpart.plot)

clinton_obama.ct <- rpart(Obama_wins ~ .-County -State -ElectionDate -TotalVote -Clinton -Obama -FIPS -
length(clinton_obama.ct$frame$var[clinton_obama.ct$frame$var == "<leaf>"])
```

```
## [1] 2
```

```
#summary(clinton_obama.ct)
prp(clinton_obama.ct, type = 2, extra = 1, under = TRUE, split.font = 1, varlen = -10,  box.palette=c("
```




```
library(caret)
clinton_obama.ct.predict <- predict(clinton_obama.ct,valid.df, type = "class")
confusionMatrix(as.factor(clinton_obama.ct.predict), as.factor(valid.df$Obama_wins))
```

Accuracy

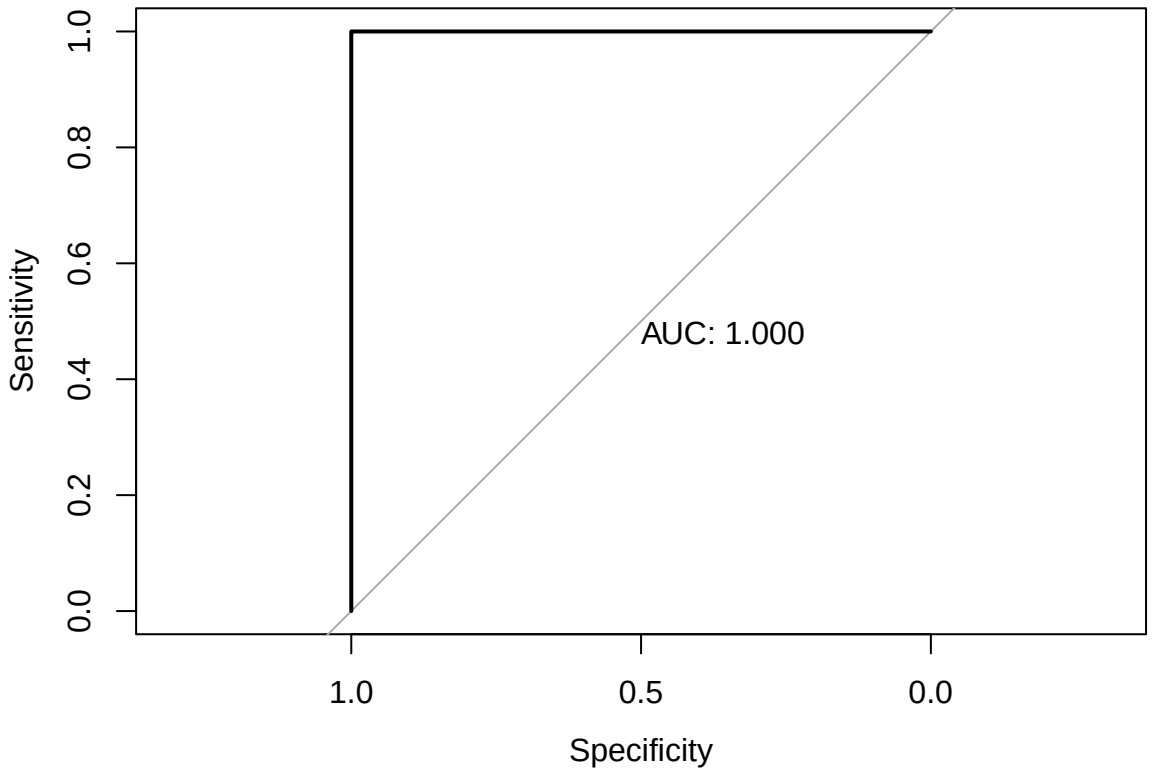
```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   0   1
##          0 214   0
```

```
##          1   0 189
##
##                Accuracy : 1
##                  95% CI : (0.9909, 1)
##     No Information Rate : 0.531
##     P-Value [Acc > NIR] : < 0.00000000000000022
##
##                   Kappa : 1
##
##  Mcnemar's Test P-Value : NA
##
##             Sensitivity : 1.000
##             Specificity : 1.000
##          Pos Pred Value : 1.000
##          Neg Pred Value : 1.000
##              Prevalence : 0.531
##          Detection Rate : 0.531
##    Detection Prevalence : 0.531
##       Balanced Accuracy : 1.000
##
##        'Positive' Class : 0
##
```

```
library(pROC)
test_roc = roc(ifelse(valid.df$Obama_wins=="1",1,0) ~
                 ifelse(clinton_obama.ct.predict=="1",1,0), plot = TRUE, print.auc = TRUE)
```

```
## Setting levels: control = 0, case = 1
```

```
## Setting direction: controls < cases
```




```
auc(test_roc)
```

```
## Area under the curve: 1
```

Predicting the margin

```
options(scipen=999)
clinton_obama.lm <- lm(Obama_margin_percent ~ .-County -State -ElectionDate -TotalVote -Clinton -Obama -
summary(clinton_obama.lm)
```

```
##
## Call:
## lm(formula = Obama_margin_percent ~ . - County - State - ElectionDate -
##     TotalVote - Clinton - Obama - FIPS - Obama_margin - Obama_wins,
##     data = train.df)
##
## Residuals:
##                      Min                       1Q                   Median
## -0.000000000000000098100 -0.000000000000000272  0.000000000000000066
##                       3Q                      Max
##  0.000000000000000406  0.000000000000000007528
##
## Coefficients:
##                                          Estimate
## (Intercept)            0.000000000000006593757239800
## RegionNortheast       -0.000000000000000390922682347
```

```
## RegionSouth           -0.000000000000000515692907547
## RegionWest             0.000000000000000239698165240
## ElectionTypePrimary   -0.000000000000000213039787974
## MalesPer100Females    -0.000000000000000000543713280
## AgeBelow35            -0.000000000000000022888709313
## Age35to65             -0.000000000000000025117666467
## Age65andAbove         -0.000000000000000045657036787
## White                 -0.000000000000000037761386835
## Black                  0.000000000000000001333883611
## Asian                 -0.000000000000000057746916309
## AmericanIndian        -0.000000000000000034473766656
## Hawaiian              -0.000000000000000126107158170
## Hispanic              -0.000000000000000000242742038
## HighSchool            -0.000000000000000002119805933
## Bachelors              0.000000000000000030030180928
## Poverty               -0.000000000000000043007756356
## IncomeAbove75K        -0.000000000000000002749403770
## MedianIncome          -0.000000000000000000009224082
## AverageIncome         -0.000000000000000000006139654
## UnemployRate          -0.000000000000000036765388300
## ManfEmploy            -0.000000000000000003409894020
## SpeakingNonEnglish     0.000000000000000001913165209
## Medicare              -0.000000000000000000006584326
## MedicareRate           0.000000000000000000015153925
## SocialSecurity        -0.000000000000000000006417223
## SocialSecurityRate    -0.000000000000000000006217339
## RetiredWorkers         0.000000000000000000014938543
## Disabilities           0.000000000000000000009516247
## DisabilitiesRate      -0.000000000000000000061580794
## Homeowner              0.000000000000000005904247499
## SameHouse1995and2000   0.000000000000000008591595575
## Pop                    0.000000000000000000000002181
## PopDensity            -0.000000000000000000010907134
## LandArea               0.000000000000000000010594942
## FarmArea              -0.000000000000000000139568332
## Obama_win_probability  0.999999999999998112620858137
##                                        Std. Error              t value
## (Intercept)            0.000000000000011498039760471                0.573
## RegionNortheast        0.000000000000000052380307049               -7.463
## RegionSouth            0.000000000000000042150955209              -12.234
## RegionWest             0.000000000000000044732493268                5.358
## ElectionTypePrimary    0.000000000000000033137941810               -6.429
## MalesPer100Females     0.000000000000000001227791305               -0.443
## AgeBelow35             0.000000000000000114820110355               -0.199
## Age35to65              0.000000000000000114752665403               -0.219
## Age65andAbove          0.000000000000000114813205303               -0.398
## White                  0.000000000000000006197630615               -6.093
## Black                  0.000000000000000005924195432                0.225
## Asian                  0.000000000000000010429233011               -5.537
## AmericanIndian         0.000000000000000007234375393               -4.765
## Hawaiian               0.000000000000000093847174158               -1.344
## Hispanic               0.000000000000000003013490245               -0.081
## HighSchool             0.000000000000000003177820020               -0.667
## Bachelors              0.000000000000000003450388083                8.703
```

```
## Poverty                0.00000000000000006948016723                -6.190
## IncomeAbove75K         0.00000000000000005736884699                -0.479
## MedianIncome           0.00000000000000000004577501                -2.015
## AverageIncome          0.00000000000000000003442139                -1.784
## UnemployRate           0.00000000000000008732821399                -4.210
## ManfEmploy             0.00000000000000001607426168                -2.121
## SpeakingNonEnglish     0.00000000000000003673809560                 0.521
## Medicare               0.00000000000000000005898693                -1.116
## MedicareRate           0.00000000000000000004211084                 3.599
## SocialSecurity         0.00000000000000000006685249                -0.960
## SocialSecurityRate     0.00000000000000000008293378                -0.750
## RetiredWorkers         0.00000000000000000009481062                 1.576
## Disabilities           0.00000000000000000004861428                 1.958
## DisabilitiesRate       0.00000000000000000016632109                -3.703
## Homeowner              0.00000000000000002554155683                 2.312
## SameHouse1995and2000   0.00000000000000002628015259                 3.269
## Pop                    0.00000000000000000000224703                 0.010
## PopDensity             0.00000000000000000006560009                -1.663
## LandArea               0.00000000000000000008782143                 1.206
## FarmArea               0.00000000000000000040330644                -3.461
## Obama_win_probability  0.00000000000000062728386322 15941745972421100.000
##                                Pr(>|t|)
## (Intercept)                    0.566456
## RegionNortheast          0.000000000000183 ***
## RegionSouth            < 0.0000000000000002 ***
## RegionWest               0.000000104115423 ***
## ElectionTypePrimary      0.000000000198487 ***
## MalesPer100Females             0.657978
## AgeBelow35                     0.842034
## Age35to65                      0.826784
## Age65andAbove                  0.690963
## White                    0.000000001578835 ***
## Black                          0.821902
## Asian                    0.000000039285892 ***
## AmericanIndian           0.000002164266411 ***
## Hawaiian                       0.179333
## Hispanic                       0.935814
## HighSchool                     0.504885
## Bachelors              < 0.0000000000000002 ***
## Poverty                  0.000000000876104 ***
## IncomeAbove75K                 0.631865
## MedianIncome                   0.044161 *
## AverageIncome                  0.074779 .
## UnemployRate             0.000027826955941 ***
## ManfEmploy                     0.034138 *
## SpeakingNonEnglish             0.602650
## Medicare                       0.264589
## MedicareRate                   0.000336 ***
## SocialSecurity                 0.337333
## SocialSecurityRate             0.453626
## RetiredWorkers                 0.115429
## Disabilities                   0.050566 .
## DisabilitiesRate               0.000225 ***
## Homeowner                      0.021000 *
```

```
## SameHouse1995and2000              0.001115 **
## Pop                               0.992257
## PopDensity                        0.096691 .
## LandArea                          0.227941
## FarmArea                          0.000562 ***
## Obama_win_probability < 0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0000000000000003187 on 1003 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:      1,  Adjusted R-squared:      1
## F-statistic: 2.527e+31 on 37 and 1003 DF,  p-value: < 0.00000000000000022
```

```
clinton_obama_step.lm <- step(clinton_obama.lm, direction="both", trace=0)
```

Stepwise selection

```
## Warning: attempting model selection on an essentially perfect fit is nonsense

## Warning: attempting model selection on an essentially perfect fit is nonsense

## Warning: attempting model selection on an essentially perfect fit is nonsense
```

```
summary(clinton_obama_step.lm)
```

```
## Warning in summary.lm(clinton_obama_step.lm): essentially perfect fit: summary
## may be unreliable

##
## Call:
## lm(formula = Obama_margin_percent ~ Region + ElectionType + MalesPer100Females +
##     AgeBelow35 + Age35to65 + Age65andAbove + White + Black +
##     Asian + AmericanIndian + Hispanic + HighSchool + Bachelors +
##     Poverty + IncomeAbove75K + MedianIncome + AverageIncome +
##     UnemployRate + ManfEmploy + SpeakingNonEnglish + Medicare +
##     MedicareRate + SocialSecurity + SocialSecurityRate + RetiredWorkers +
##     Disabilities + DisabilitiesRate + Homeowner + SameHouse1995and2000 +
##     Pop + PopDensity + LandArea + FarmArea + Obama_win_probability,
##     data = train.df)
##
## Residuals:
##                       Min                        1Q                    Median
## -0.00000000000000054196 -0.00000000000000006125 -0.00000000000000000438
##                        3Q                       Max
##  0.00000000000000005666  0.00000000000000249729
##
## Coefficients:
##                                          Estimate
## (Intercept)          -0.000000000000000226494622120
```

```
## RegionNortheast        0.00000000000000043247207648
## RegionSouth            0.00000000000000044529763087
## RegionWest            -0.00000000000000029584052344
## ElectionTypePrimary    0.00000000000000020825513336
## MalesPer100Females     0.00000000000000000169085909
## AgeBelow35            -0.00000000000000000554595480
## Age35to65             -0.00000000000000000347384931
## Age65andAbove          0.00000000000000000466757718
## White                  0.00000000000000002920427349
## Black                 -0.00000000000000001438913466
## Asian                  0.00000000000000006849730228
## AmericanIndian         0.00000000000000003847461514
## Hispanic              -0.00000000000000000430102503
## HighSchool            -0.00000000000000000828443522
## Bachelors             -0.00000000000000003226038897
## Poverty                0.00000000000000005054721062
## IncomeAbove75K         0.00000000000000000976770285
## MedianIncome           0.00000000000000000001043242
## AverageIncome          0.00000000000000000000372119
## UnemployRate           0.00000000000000005539314854
## ManfEmploy             0.00000000000000000379946690
## SpeakingNonEnglish    -0.00000000000000000064190042
## Medicare               0.00000000000000000000664702
## MedicareRate          -0.00000000000000000001696803
## SocialSecurity         0.00000000000000000000813365
## SocialSecurityRate     0.00000000000000000000702406
## RetiredWorkers        -0.00000000000000000001738517
## Disabilities          -0.00000000000000000001167941
## DisabilitiesRate       0.00000000000000000007436146
## Homeowner             -0.00000000000000000985183430
## SameHouse1995and2000  -0.00000000000000000999566703
## Pop                    0.00000000000000000000001713
## PopDensity             0.00000000000000000001199429
## LandArea              -0.00000000000000000001046211
## FarmArea               0.00000000000000000013938717
## Obama_win_probability  1.00000000000000222044604925
##                                        Std. Error                t value
## (Intercept)            0.00000000000000496942275522                -0.456
## RegionNortheast        0.00000000000000002263211419                19.109
## RegionSouth            0.00000000000000001821591232                24.446
## RegionWest             0.00000000000000001910256312               -15.487
## ElectionTypePrimary    0.00000000000000001432256859                14.540
## MalesPer100Females     0.00000000000000000053015322                 3.189
## AgeBelow35             0.00000000000000004962666076                -0.112
## Age35to65              0.00000000000000004959629932                -0.070
## Age65andAbove          0.00000000000000004962372801                 0.699
## White                  0.00000000000000000267029268                10.937
## Black                  0.00000000000000000255362713                -5.635
## Asian                  0.00000000000000000447535683                15.305
## AmericanIndian         0.00000000000000000311601390                12.347
## Hispanic               0.00000000000000000130058812                -3.307
## HighSchool             0.00000000000000000137130667                -6.041
## Bachelors              0.00000000000000000148062453               -21.788
## Poverty                0.00000000000000000300104635                16.843
```

```
## IncomeAbove75K         0.000000000000000000247955530                   3.939
## MedianIncome           0.000000000000000000000197733                   5.276
## AverageIncome          0.000000000000000000000148121                   2.512
## UnemployRate           0.000000000000000000377284221                  14.682
## ManfEmploy             0.000000000000000000069446636                   5.471
## SpeakingNonEnglish     0.000000000000000000158744473                  -0.404
## Medicare               0.000000000000000000000254550                   2.611
## MedicareRate           0.000000000000000000000181863                  -9.330
## SocialSecurity         0.000000000000000000000287976                   2.824
## SocialSecurityRate     0.000000000000000000000358450                   1.960
## RetiredWorkers         0.000000000000000000000409211                  -4.248
## Disabilities           0.000000000000000000000210114                  -5.559
## DisabilitiesRate       0.000000000000000000000718860                  10.344
## Homeowner              0.000000000000000000110326246                  -8.930
## SameHouse1995and2000   0.000000000000000000113489430                  -8.808
## Pop                    0.000000000000000000000009695                   0.177
## PopDensity             0.000000000000000000000283279                   4.234
## LandArea               0.000000000000000000000379431                  -2.757
## FarmArea               0.000000000000000000001740374                   8.009
## Obama_win_probability  0.000000000000000002707768882 369307737726429696.000
##                                  Pr(>|t|)
## (Intercept)                      0.648649
## RegionNortheast       < 0.0000000000000002 ***
## RegionSouth           < 0.0000000000000002 ***
## RegionWest            < 0.0000000000000002 ***
## ElectionTypePrimary   < 0.0000000000000002 ***
## MalesPer100Females               0.001470 **
## AgeBelow35                       0.911041
## Age35to65                        0.944174
## Age65andAbove                    0.484958
## White                 < 0.0000000000000002 ***
## Black                  0.00000002276026589 ***
## Asian                 < 0.0000000000000002 ***
## AmericanIndian        < 0.0000000000000002 ***
## Hispanic                         0.000976 ***
## HighSchool             0.00000000215146781 ***
## Bachelors             < 0.0000000000000002 ***
## Poverty               < 0.0000000000000002 ***
## IncomeAbove75K         0.00008737151086194 ***
## MedianIncome           0.00000016172047470 ***
## AverageIncome                    0.012152 *
## UnemployRate          < 0.0000000000000002 ***
## ManfEmploy             0.00000005648717881 ***
## SpeakingNonEnglish               0.686034
## Medicare                         0.009155 **
## MedicareRate          < 0.0000000000000002 ***
## SocialSecurity                   0.004830 **
## SocialSecurityRate               0.050323 .
## RetiredWorkers         0.00002352573851017 ***
## Disabilities           0.00000003484725834 ***
## DisabilitiesRate      < 0.0000000000000002 ***
## Homeowner             < 0.0000000000000002 ***
## SameHouse1995and2000  < 0.0000000000000002 ***
## Pop                              0.859789
```

```
## PopDensity             0.00002505261319551 ***
## LandArea                          0.005934 **
## FarmArea               0.00000000000000319 ***
## Obama_win_probability < 0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.00000000000000001377 on 1004 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:      1,  Adjusted R-squared:      1
## F-statistic: 1.39e+34 on 36 and 1004 DF,  p-value: < 0.00000000000000022
```

```
clinton_obama_back.lm  <- step(clinton_obama.lm, direction="backward",  trace=0)
```

Backward selection

```
## Warning: attempting model selection on an essentially perfect fit is nonsense
```

```
## Warning: attempting model selection on an essentially perfect fit is nonsense
```

```
summary(clinton_obama_back.lm)
```

```
## Warning in summary.lm(clinton_obama_back.lm): essentially perfect fit: summary
## may be unreliable
```

```
##
## Call:
## lm(formula = Obama_margin_percent ~ Region + ElectionType + MalesPer100Females +
##     AgeBelow35 + Age35to65 + Age65andAbove + White + Black +
##     Asian + AmericanIndian + Hispanic + HighSchool + Bachelors +
##     Poverty + IncomeAbove75K + MedianIncome + AverageIncome +
##     UnemployRate + ManfEmploy + SpeakingNonEnglish + Medicare +
##     MedicareRate + SocialSecurity + SocialSecurityRate + RetiredWorkers +
##     Disabilities + DisabilitiesRate + Homeowner + SameHouse1995and2000 +
##     Pop + PopDensity + LandArea + FarmArea + Obama_win_probability,
##     data = train.df)
##
## Residuals:
##                       Min                        1Q                    Median
## -0.0000000000000000054196 -0.0000000000000000006125 -0.0000000000000000000438
##                        3Q                       Max
##  0.0000000000000000005666  0.0000000000000000249729
##
## Coefficients:
##                                        Estimate
## (Intercept)           -0.000000000000000002264946221 20
## RegionNortheast        0.000000000000000000432472076 48
## RegionSouth            0.000000000000000000445297630 87
## RegionWest            -0.000000000000000000295840523 44
## ElectionTypePrimary    0.000000000000000000208255133 36
```

```
## MalesPer100Females     0.00000000000000000169085909
## AgeBelow35            -0.00000000000000000554595480
## Age35to65             -0.00000000000000000347384931
## Age65andAbove          0.00000000000000003466757718
## White                  0.00000000000000002920427349
## Black                 -0.00000000000000001438913466
## Asian                  0.00000000000000006849730228
## AmericanIndian         0.00000000000000003847461514
## Hispanic              -0.00000000000000000430102503
## HighSchool            -0.00000000000000000828443522
## Bachelors             -0.00000000000000003226038897
## Poverty                0.00000000000000005054721062
## IncomeAbove75K         0.00000000000000000976770285
## MedianIncome           0.00000000000000000001043242
## AverageIncome          0.00000000000000000000372119
## UnemployRate           0.00000000000000005539314854
## ManfEmploy             0.00000000000000000379946690
## SpeakingNonEnglish    -0.00000000000000000064190042
## Medicare               0.00000000000000000000664702
## MedicareRate          -0.00000000000000000001696803
## SocialSecurity         0.00000000000000000000813365
## SocialSecurityRate     0.00000000000000000000702406
## RetiredWorkers        -0.00000000000000000001738517
## Disabilities          -0.00000000000000000001167941
## DisabilitiesRate       0.00000000000000000007436146
## Homeowner             -0.00000000000000000985183430
## SameHouse1995and2000  -0.00000000000000000999566703
## Pop                    0.00000000000000000000001713
## PopDensity             0.00000000000000000001199429
## LandArea              -0.00000000000000000001046211
## FarmArea               0.00000000000000000013938717
## Obama_win_probability  1.00000000000000222044604925
##                                        Std. Error                t value
## (Intercept)            0.00000000000000496942275522                -0.456
## RegionNortheast        0.00000000000000002263211419                19.109
## RegionSouth            0.00000000000000001821591232                24.446
## RegionWest             0.00000000000000001910256312               -15.487
## ElectionTypePrimary    0.00000000000000001432256859                14.540
## MalesPer100Females     0.00000000000000000053015322                 3.189
## AgeBelow35             0.00000000000000004962666076                -0.112
## Age35to65              0.00000000000000004959629932                -0.070
## Age65andAbove          0.00000000000000004962372801                 0.699
## White                  0.00000000000000000267029268                10.937
## Black                  0.00000000000000000255362713                -5.635
## Asian                  0.00000000000000000447535683                15.305
## AmericanIndian         0.00000000000000000311601390                12.347
## Hispanic               0.00000000000000000130058812                -3.307
## HighSchool             0.00000000000000000137130667                -6.041
## Bachelors              0.00000000000000000148062453               -21.788
## Poverty                0.00000000000000000300104635                16.843
## IncomeAbove75K         0.00000000000000000247955530                 3.939
## MedianIncome           0.00000000000000000000197733                 5.276
## AverageIncome          0.00000000000000000000148121                 2.512
## UnemployRate           0.00000000000000000377284221                14.682
```

```
## ManfEmploy             0.00000000000000000069446636                  5.471
## SpeakingNonEnglish     0.00000000000000000158744473                 -0.404
## Medicare               0.00000000000000000000254550                  2.611
## MedicareRate           0.00000000000000000000181863                 -9.330
## SocialSecurity         0.00000000000000000000287976                  2.824
## SocialSecurityRate     0.00000000000000000000358450                  1.960
## RetiredWorkers         0.00000000000000000000409211                 -4.248
## Disabilities           0.00000000000000000000210114                 -5.559
## DisabilitiesRate       0.00000000000000000000718860                 10.344
## Homeowner              0.00000000000000000110326246                 -8.930
## SameHouse1995and2000   0.00000000000000000113489430                 -8.808
## Pop                    0.00000000000000000000009695                  0.177
## PopDensity             0.00000000000000000000283279                  4.234
## LandArea               0.00000000000000000000379431                 -2.757
## FarmArea               0.00000000000000000001740374                  8.009
## Obama_win_probability  0.00000000000000000002707768882 369307737726429696.000
##                                  Pr(>|t|)
## (Intercept)                      0.648649
## RegionNortheast       < 0.0000000000000002 ***
## RegionSouth           < 0.0000000000000002 ***
## RegionWest            < 0.0000000000000002 ***
## ElectionTypePrimary   < 0.0000000000000002 ***
## MalesPer100Females               0.001470 **
## AgeBelow35                       0.911041
## Age35to65                        0.944174
## Age65andAbove                    0.484958
## White                 < 0.0000000000000002 ***
## Black                  0.00000002276026589 ***
## Asian                 < 0.0000000000000002 ***
## AmericanIndian        < 0.0000000000000002 ***
## Hispanic                         0.000976 ***
## HighSchool             0.00000000215146781 ***
## Bachelors             < 0.0000000000000002 ***
## Poverty               < 0.0000000000000002 ***
## IncomeAbove75K         0.00008737151086194 ***
## MedianIncome           0.00000016172047470 ***
## AverageIncome                    0.012152 *
## UnemployRate          < 0.0000000000000002 ***
## ManfEmploy             0.00000005648717881 ***
## SpeakingNonEnglish               0.686034
## Medicare                         0.009155 **
## MedicareRate          < 0.0000000000000002 ***
## SocialSecurity                   0.004830 **
## SocialSecurityRate               0.050323 .
## RetiredWorkers         0.00002352573851017 ***
## Disabilities           0.00000003484725834 ***
## DisabilitiesRate      < 0.0000000000000002 ***
## Homeowner             < 0.0000000000000002 ***
## SameHouse1995and2000  < 0.0000000000000002 ***
## Pop                              0.859789
## PopDensity             0.00002505261319551 ***
## LandArea                         0.005934 **
## FarmArea               0.00000000000000319 ***
## Obama_win_probability < 0.0000000000000002 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.00000000000000001377 on 1004 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:      1,  Adjusted R-squared:      1
## F-statistic: 1.39e+34 on 36 and 1004 DF,  p-value: < 0.00000000000000022
```

```
clinton_obama_for.lm  <- step(clinton_obama.lm, direction =
"forward",  trace=0)
summary(clinton_obama_for.lm)
```

Forward selection

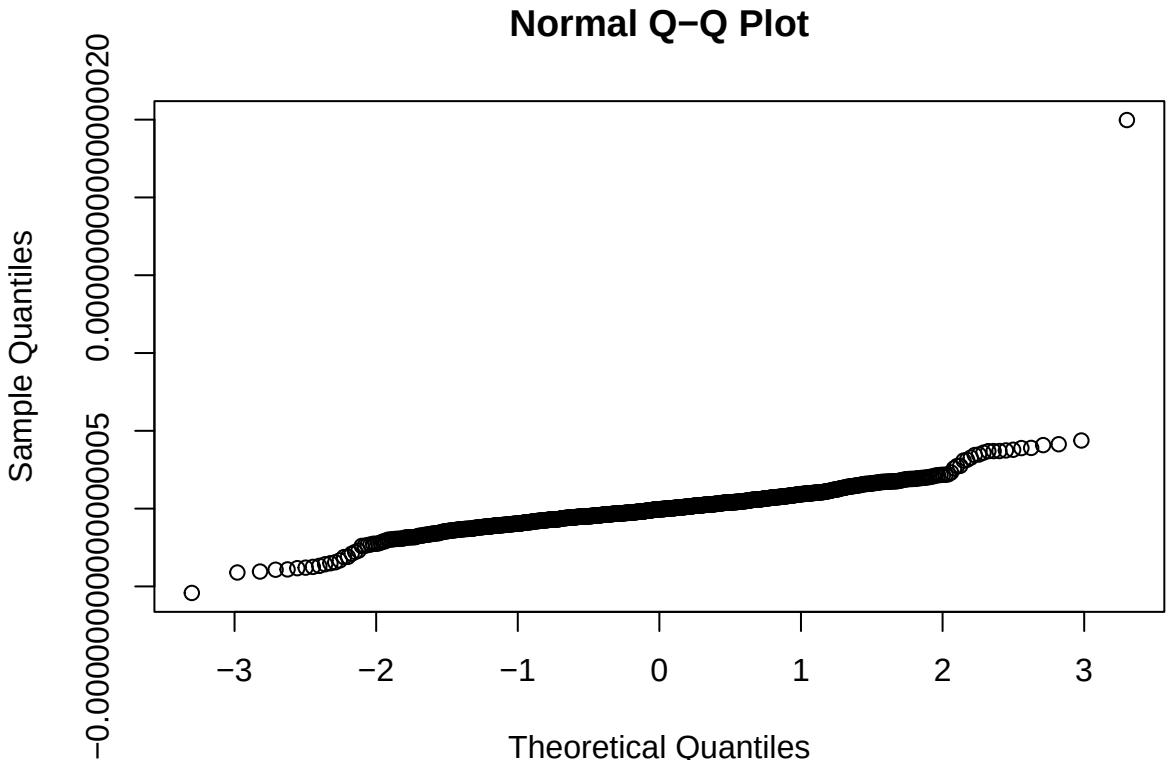
```
##
## Call:
## lm(formula = Obama_margin_percent ~ (County + State + Region +
##     FIPS + ElectionDate + ElectionType + TotalVote + Clinton +
##     Obama + MalesPer100Females + AgeBelow35 + Age35to65 + Age65andAbove +
##     White + Black + Asian + AmericanIndian + Hawaiian + Hispanic +
##     HighSchool + Bachelors + Poverty + IncomeAbove75K + MedianIncome +
##     AverageIncome + UnemployRate + ManfEmploy + SpeakingNonEnglish +
##     Medicare + MedicareRate + SocialSecurity + SocialSecurityRate +
##     RetiredWorkers + Disabilities + DisabilitiesRate + Homeowner +
##     SameHouse1995and2000 + Pop + PopDensity + LandArea + FarmArea +
##     Obama_margin + Obama_wins + Obama_win_probability) - County -
##     State - ElectionDate - TotalVote - Clinton - Obama - FIPS -
##     Obama_margin - Obama_wins, data = train.df)
##
## Residuals:
##                      Min                       1Q                   Median
## -0.000000000000000098100 -0.000000000000000272  0.000000000000000066
##                       3Q                      Max
##  0.000000000000000406  0.000000000000007528
##
## Coefficients:
##                                          Estimate
## (Intercept)            0.000000000000006593757239800
## RegionNortheast       -0.000000000000000390922682347
## RegionSouth           -0.000000000000000515692907547
## RegionWest             0.000000000000000239698165240
## ElectionTypePrimary   -0.000000000000000213039787974
## MalesPer100Females    -0.000000000000000000543713280
## AgeBelow35            -0.000000000000000022888709313
## Age35to65             -0.000000000000000025117666467
## Age65andAbove         -0.000000000000000045657036787
## White                 -0.000000000000000037761386835
## Black                  0.000000000000000001333883611
## Asian                 -0.000000000000000057746916309
## AmericanIndian        -0.000000000000000034473766656
## Hawaiian              -0.000000000000000126107158170
```

```
## Hispanic              -0.000000000000000000242742038
## HighSchool            -0.000000000000000002119805933
## Bachelors              0.000000000000000030030180928
## Poverty               -0.000000000000000043007756356
## IncomeAbove75K        -0.000000000000000002749403770
## MedianIncome          -0.000000000000000000009224082
## AverageIncome         -0.000000000000000000006139654
## UnemployRate          -0.000000000000000036765388300
## ManfEmploy            -0.000000000000000003409894020
## SpeakingNonEnglish     0.000000000000000001913165209
## Medicare              -0.000000000000000000006584326
## MedicareRate           0.000000000000000000015153925
## SocialSecurity        -0.000000000000000000006417223
## SocialSecurityRate    -0.000000000000000000006217339
## RetiredWorkers         0.000000000000000000014938543
## Disabilities           0.000000000000000000009516247
## DisabilitiesRate      -0.000000000000000000061580794
## Homeowner              0.000000000000000005904247499
## SameHouse1995and2000   0.000000000000000008591595575
## Pop                    0.000000000000000000000002181
## PopDensity            -0.000000000000000000010907134
## LandArea               0.000000000000000000010594942
## FarmArea              -0.000000000000000000139568332
## Obama_win_probability  0.999999999999998112620858137
##                                          Std. Error              t value
## (Intercept)            0.000000000000011498039760471                0.573
## RegionNortheast        0.000000000000000052380307049               -7.463
## RegionSouth            0.000000000000000042150955209              -12.234
## RegionWest             0.000000000000000044732493268                5.358
## ElectionTypePrimary    0.000000000000000033137941810               -6.429
## MalesPer100Females     0.000000000000000001227791305               -0.443
## AgeBelow35             0.000000000000000114820110355               -0.199
## Age35to65              0.000000000000000114752665403               -0.219
## Age65andAbove          0.000000000000000114813205303               -0.398
## White                  0.000000000000000006197630615               -6.093
## Black                  0.000000000000000005924195432                0.225
## Asian                  0.000000000000000010429233011               -5.537
## AmericanIndian         0.000000000000000007234375393               -4.765
## Hawaiian               0.000000000000000093847174158               -1.344
## Hispanic               0.000000000000000003013490245               -0.081
## HighSchool             0.000000000000000003177820020               -0.667
## Bachelors              0.000000000000000003450388083                8.703
## Poverty                0.000000000000000006948016723               -6.190
## IncomeAbove75K         0.000000000000000005736884699               -0.479
## MedianIncome           0.000000000000000000004577501               -2.015
## AverageIncome          0.000000000000000000003442139               -1.784
## UnemployRate           0.000000000000000008732821399               -4.210
## ManfEmploy             0.000000000000000001607426168               -2.121
## SpeakingNonEnglish     0.000000000000000003673809560                0.521
## Medicare               0.000000000000000000005898693               -1.116
## MedicareRate           0.000000000000000000004211084                3.599
## SocialSecurity         0.000000000000000000006685249               -0.960
## SocialSecurityRate     0.000000000000000000008293378               -0.750
## RetiredWorkers         0.000000000000000000009481062                1.576
```

```
## Disabilities           0.00000000000000000004861428                 1.958
## DisabilitiesRate       0.00000000000000000016632109                -3.703
## Homeowner              0.00000000000000002554155683                 2.312
## SameHouse1995and2000   0.00000000000000002628015259                 3.269
## Pop                    0.00000000000000000000224703                 0.010
## PopDensity             0.00000000000000000006560009                -1.663
## LandArea               0.00000000000000000008782143                 1.206
## FarmArea               0.00000000000000000040330644                -3.461
## Obama_win_probability  0.00000000000000062728386322 15941745972421100.000
##                                Pr(>|t|)
## (Intercept)                    0.566456
## RegionNortheast          0.000000000000183 ***
## RegionSouth            < 0.0000000000000002 ***
## RegionWest               0.000000104115423 ***
## ElectionTypePrimary      0.000000000198487 ***
## MalesPer100Females             0.657978
## AgeBelow35                     0.842034
## Age35to65                      0.826784
## Age65andAbove                  0.690963
## White                    0.000000001578835 ***
## Black                          0.821902
## Asian                    0.000000039285892 ***
## AmericanIndian           0.000002164266411 ***
## Hawaiian                       0.179333
## Hispanic                       0.935814
## HighSchool                     0.504885
## Bachelors              < 0.0000000000000002 ***
## Poverty                  0.000000000876104 ***
## IncomeAbove75K                 0.631865
## MedianIncome                   0.044161 *
## AverageIncome                  0.074779 .
## UnemployRate             0.000027826955941 ***
## ManfEmploy                     0.034138 *
## SpeakingNonEnglish             0.602650
## Medicare                       0.264589
## MedicareRate                   0.000336 ***
## SocialSecurity                 0.337333
## SocialSecurityRate             0.453626
## RetiredWorkers                 0.115429
## Disabilities                   0.050566 .
## DisabilitiesRate               0.000225 ***
## Homeowner                      0.021000 *
## SameHouse1995and2000           0.001115 **
## Pop                            0.992257
## PopDensity                     0.096691 .
## LandArea                       0.227941
## FarmArea                       0.000562 ***
## Obama_win_probability < 0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.000000000000003187 on 1003 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:      1,  Adjusted R-squared:      1
```

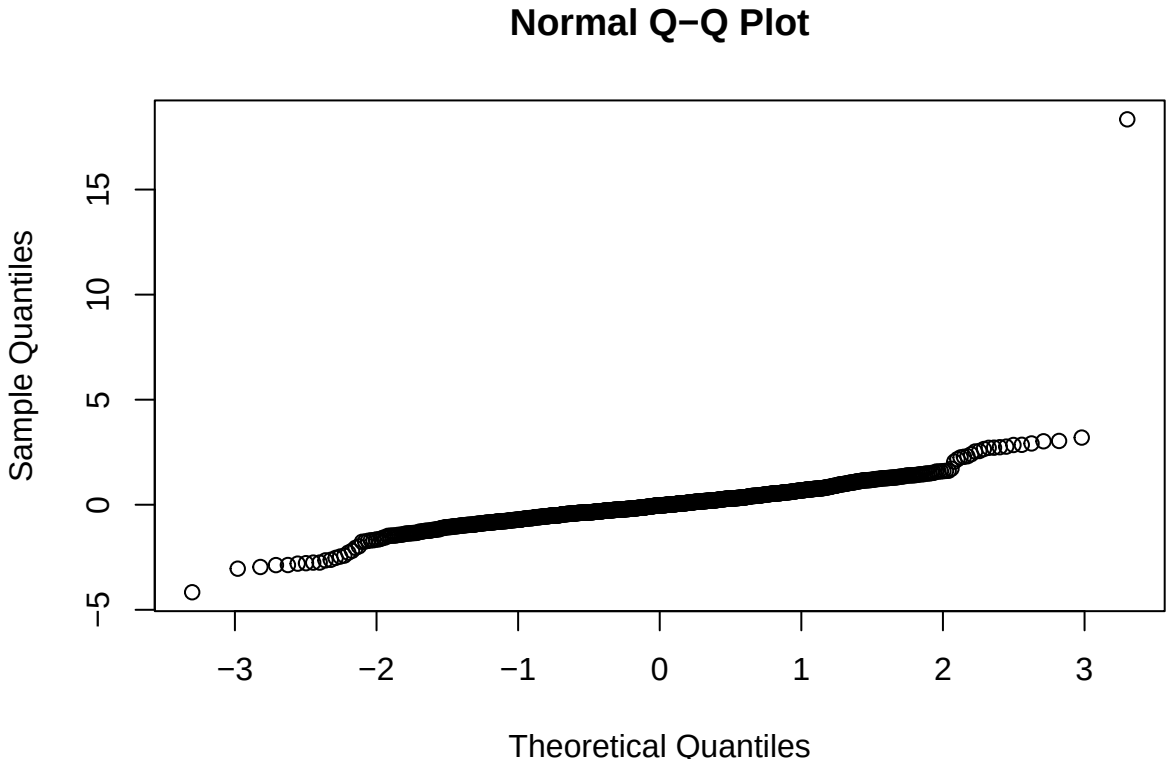
```
## F-statistic: 2.527e+31 on 37 and 1003 DF,  p-value: < 0.00000000000000022
```

```
qqnorm(resid(clinton_obama_step.lm))
```

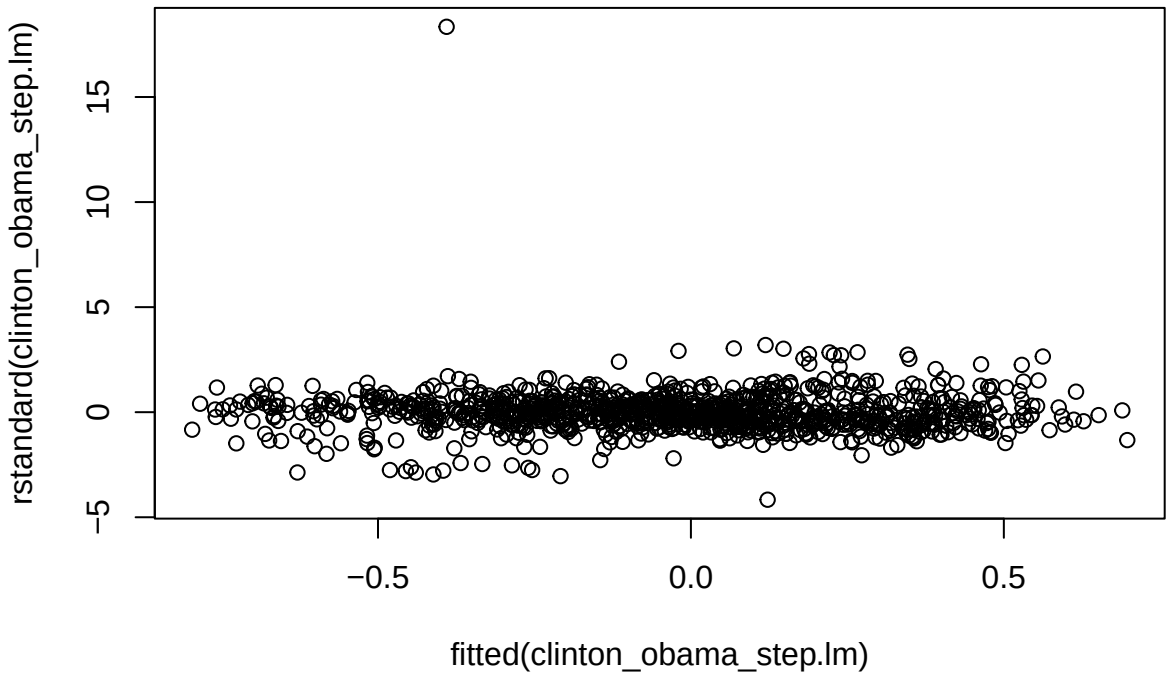


Stepwise model fit

```
qqnorm(rstandard(clinton_obama_step.lm))
```




```
plot(fitted(clinton_obama_step.lm),rstandard(clinton_obama_step.lm))
```




```
library(forecast)
```

Accuracy

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```
valid_predict <- predict(clinton_obama_step.lm, valid.df)
accuracy(valid_predict, valid.df$Obama_margin_percent)
```

```
##                                      ME                      RMSE
## Test set -0.000000000000000000008361398 0.00000000000000003730835
##                                  MAE MPE MAPE
## Test set 0.00000000000000002497594 Inf  Inf
```

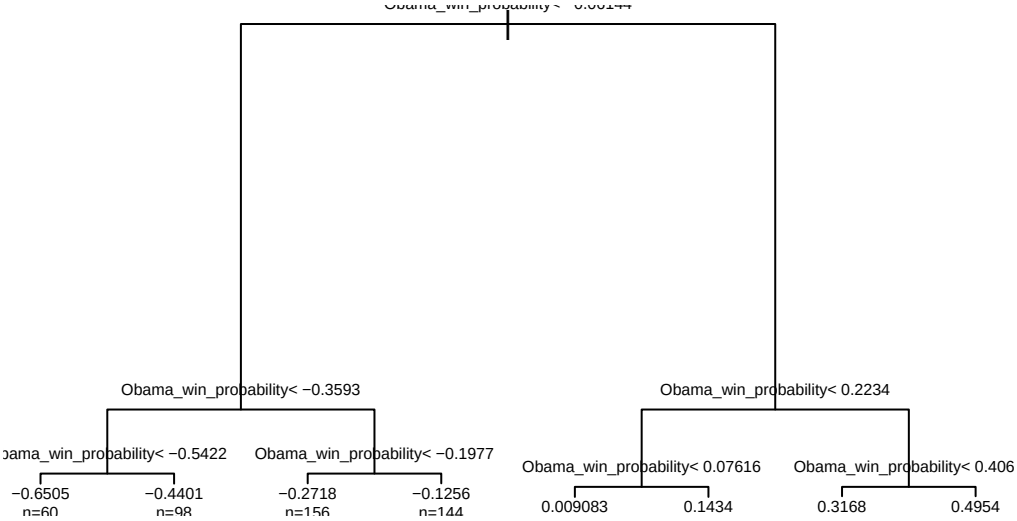
Regression Tree

```
library(dplyr)
clinton_obama_RT <- select(train.df, -County, -State, -ElectionDate, -TotalVote, -Clinton, -Obama, -FIPS

library(rpart)
clinton_obama_reg_tree <- rpart(Obama_margin_percent ~ ., data=clinton_obama_RT)
print(clinton_obama_reg_tree)
```

```
## n= 1042
##
## node), split, n, deviance, yval
##       * denotes terminal node
##
##  1) root 1042 95.0957200 -0.033478190
##    2) Obama_win_probability< -0.06144005 458 14.7392900 -0.311469900
##      4) Obama_win_probability< -0.3593005 158  2.1422450 -0.520010300
##        8) Obama_win_probability< -0.5422066 60  0.2549426 -0.650532600 *
##        9) Obama_win_probability>=-0.5422066 98  0.2393249 -0.440098700 *
##      5) Obama_win_probability>=-0.3593005 300  2.1068940 -0.201638700
##       10) Obama_win_probability< -0.1977241 156  0.3024870 -0.271818700 *
##       11) Obama_win_probability>=-0.1977241 144  0.2037065 -0.125610300 *
##    3) Obama_win_probability>=-0.06144005 584 17.2048600  0.184535900
##      6) Obama_win_probability< 0.223429 367  2.2742920  0.071293110
##       12) Obama_win_probability< 0.07616471 197  0.3255575  0.009083124 *
##       13) Obama_win_probability>=0.07616471 170  0.3028344  0.143383500 *
##      7) Obama_win_probability>=0.223429 217  2.2645480  0.376057100
##       14) Obama_win_probability< 0.4064146 145  0.4040828  0.316797600 *
##       15) Obama_win_probability>=0.4064146 72  0.3258101  0.495399100 *
```

```
plot(clinton_obama_reg_tree)
text(clinton_obama_reg_tree, use.n=TRUE, cex=0.5)
```



Accuracy

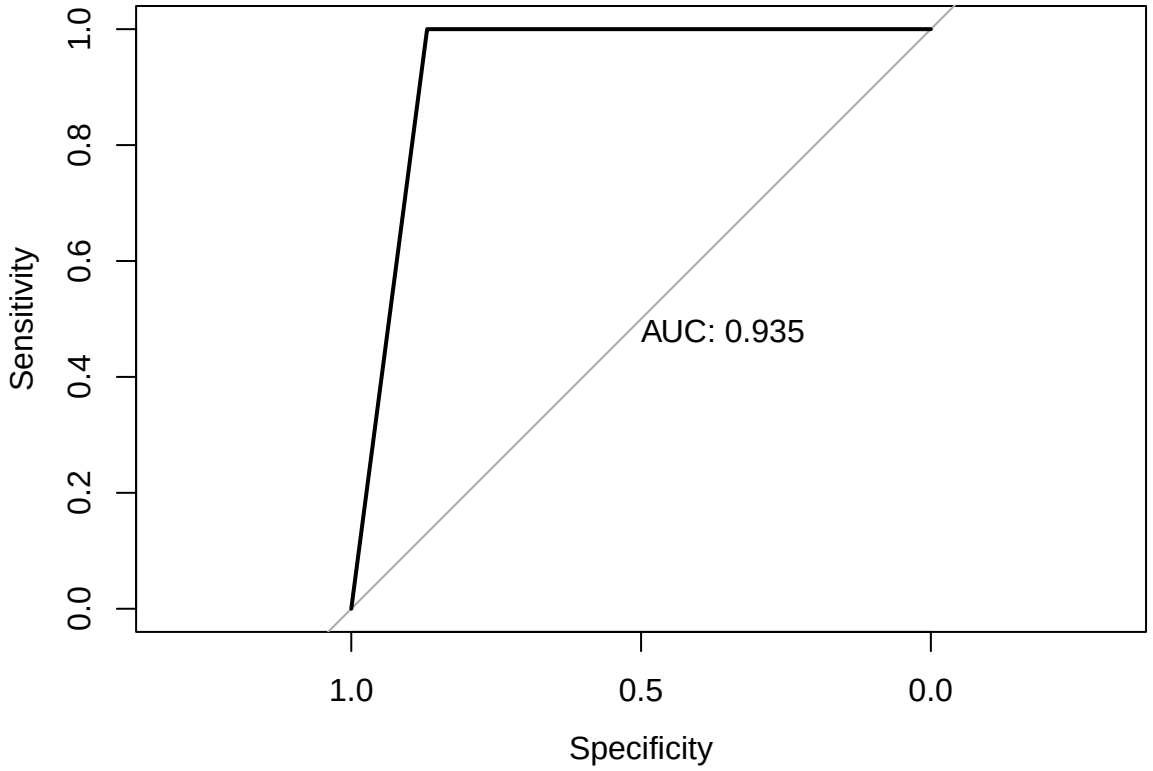
```
RT_obama <- ifelse(predict(clinton_obama_reg_tree,valid.df) >0, 1,0)
confusionMatrix(as.factor(RT_obama), as.factor(valid.df$Obama_wins))
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   0   1
##          0 186   0
##          1  28 189
##
##                Accuracy : 0.9305
##                  95% CI : (0.9011, 0.9533)
##     No Information Rate : 0.531
##     P-Value [Acc > NIR] : < 0.00000000000000022
##
##                   Kappa : 0.8617
##
##  Mcnemar's Test P-Value : 0.0000003352
##
##             Sensitivity : 0.8692
##             Specificity : 1.0000
##          Pos Pred Value : 1.0000
##          Neg Pred Value : 0.8710
##              Prevalence : 0.5310
##          Detection Rate : 0.4615
##    Detection Prevalence : 0.4615
##       Balanced Accuracy : 0.9346
##
##        'Positive' Class : 0
##
```

```
library(pROC)
test_roc <- roc(ifelse(valid.df$Obama_wins=="1",1,0) ~
                ifelse(RT_obama=="1",1,0), plot = TRUE, print.auc = TRUE)
```

```
## Setting levels: control = 0, case = 1
```

```
## Setting direction: controls < cases
```




```
auc(test_roc)
```

```
## Area under the curve: 0.9346
```

```
test.df$Obama_wins <- as.numeric(test.df$Obama_wins)
Obama_MP_pred <- predict(clinton_obama_step.lm, test.df)

test.df$Obama_margin_percent <- Obama_MP_pred

summary(test.df)
```

so who wins

```
##         County         State           Region         FIPS
##  Washington:  11   TX     :251   Midwest  :318   Min.   :15001
##  Jackson    :  10   KY     :120   Northeast: 86   1st Qu.:28134
##  Jefferson :  10   NC     :100   South    :608   Median :41047
##  Lincoln   :   9   IN     : 92   West     :119   Mean   :38894
##  Franklin  :   8   OH     : 88                   3rd Qu.:48268
##  Union     :   8   MS     : 82                   Max.   :56045
##  (Other)   :1075   (Other):398
##  ElectionDate         ElectionType    TotalVote       Clinton
##  Length:1131        Caucuses:  27   Min.   : NA    Min.   : NA
```

```
##  Class :character   Primary :1104   1st Qu.: NA    1st Qu.: NA
##  Mode  :character                   Median : NA    Median : NA
##                                     Mean   :NaN    Mean   :NaN
##                                     3rd Qu.: NA    3rd Qu.: NA
##                                     Max.   : NA    Max.   : NA
##                                     NA's   :1131   NA's   :1131
##      Obama      MalesPer100Females   AgeBelow35      Age35to65
##  Min.   : NA    Min.   : 76.20     Min.   : 4.50   Min.   :24.90
##  1st Qu.: NA    1st Qu.: 94.90     1st Qu.:42.20   1st Qu.:38.00
##  Median : NA    Median : 97.70     Median :45.40   Median :40.10
##  Mean   :NaN    Mean   : 98.83     Mean   :45.51   Mean   :39.74
##  3rd Qu.: NA    3rd Qu.:100.20     3rd Qu.:48.40   3rd Qu.:41.70
##  Max.   : NA    Max.   :194.60     Max.   :68.80   Max.   :79.20
##  NA's   :1131
##  Age65andAbove       White            Black            Asian
##  Min.   : 5.10   Min.   :  6.70   Min.   : 0.100   Min.   : 0.1000
##  1st Qu.:12.45   1st Qu.: 88.10   1st Qu.: 0.500   1st Qu.: 0.3000
##  Median :14.40   Median : 95.20   Median : 2.300   Median : 0.4000
##  Mean   :14.75   Mean   : 88.72   Mean   : 8.104   Mean   : 0.8364
##  3rd Qu.:16.60   3rd Qu.: 97.80   3rd Qu.: 9.400   3rd Qu.: 0.8000
##  Max.   :30.70   Max.   :100.00   Max.   :86.000   Max.   :46.6000
##
##  AmericanIndian      Hawaiian         Hispanic        HighSchool
##  Min.   : 0.100   Min.   : 0.000   Min.   : 0.100   Min.   :34.70
##  1st Qu.: 0.200   1st Qu.: 0.000   1st Qu.: 1.000   1st Qu.:70.40
##  Median : 0.400   Median : 0.000   Median : 2.000   Median :78.10
##  Mean   : 1.931   Mean   : 0.103   Mean   : 8.631   Mean   :75.95
##  3rd Qu.: 0.900   3rd Qu.: 0.100   3rd Qu.: 7.950   3rd Qu.:82.70
##  Max.   :91.800   Max.   :30.600   Max.   :97.600   Max.   :94.70
##
##    Bachelors        Poverty      IncomeAbove75K   MedianIncome
##  Min.   : 4.90   Min.   : 3.90   Min.   : 1.50   Min.   :16868
##  1st Qu.:10.90   1st Qu.:10.65   1st Qu.: 9.30   1st Qu.:31680
##  Median :13.60   Median :13.60   Median :11.80   Median :36638
##  Mean   :15.57   Mean   :14.68   Mean   :13.18   Mean   :37696
##  3rd Qu.:18.30   3rd Qu.:17.50   3rd Qu.:15.75   3rd Qu.:42458
##  Max.   :51.50   Max.   :39.40   Max.   :46.80   Max.   :79927
##
##  AverageIncome    UnemployRate      ManfEmploy      SpeakingNonEnglish
##  Min.   :12197   Min.   : 1.500   Min.   : 0.2024   Min.   : 0.700
##  1st Qu.:23348   1st Qu.: 4.200   1st Qu.: 5.1739   1st Qu.: 3.000
##  Median :26421   Median : 5.100   Median :10.6564   Median : 4.700
##  Mean   :26942   Mean   : 5.288   Mean   :11.2173   Mean   : 9.603
##  3rd Qu.:29738   3rd Qu.: 6.000   3rd Qu.:15.3143   3rd Qu.: 9.450
##  Max.   :89028   Max.   :14.800   Max.   :49.0580   Max.   :92.100
##
##     Medicare       MedicareRate   SocialSecurity   SocialSecurityRate
##  Min.   :    21   Min.   : 2716   Min.   :    20   Min.   : 6842
##  1st Qu.:  1991   1st Qu.:14382   1st Qu.:  2388   1st Qu.:17166
##  Median :  4569   Median :16832   Median :  5465   Median :20070
##  Mean   : 10786   Mean   :17139   Mean   : 12502   Mean   :20040
##  3rd Qu.: 10201   3rd Qu.:19718   3rd Qu.: 11855   3rd Qu.:22822
##  Max.   :314989   Max.   :60845   Max.   :365550   Max.   :39065
##                                                    NA's   :1
```

```
##  RetiredWorkers    Disabilities   DisabilitiesRate   Homeowner
##  Min.   :    15   Min.   :    4   Min.   :  205    Min.   :34.3
##  1st Qu.:  1315   1st Qu.:  294   1st Qu.: 1468    1st Qu.:71.3
##  Median :  3105   Median :  712   Median : 2247    Median :75.5
##  Mean   :  7704   Mean   : 1779   Mean   : 2909    Mean   :74.3
##  3rd Qu.:  7262   3rd Qu.: 1641   3rd Qu.: 3492    3rd Qu.:78.9
##  Max.   :212745   Max.   :94548   Max.   :20185    Max.   :87.7
##                                   NA's   :4
##  SameHouse1995and2000      Pop            PopDensity          LandArea
##  Min.   :32.10        Min.   :     60   Min.   :    0.10   Min.   :   45.0
##  1st Qu.:55.70        1st Qu.:  11730   1st Qu.:   18.35   1st Qu.:  418.0
##  Median :60.50        Median :  27363   Median :   49.40   Median :  637.0
##  Mean   :59.86        Mean   :  75982   Mean   :  144.61   Mean   :  943.1
##  3rd Qu.:64.70        3rd Qu.:  62519   3rd Qu.:  111.45   3rd Qu.:  959.5
##  Max.   :90.50        Max.   :3886207   Max.   :10721.70   Max.   :10491.0
##
##     FarmArea       Obama_margin  Obama_margin_percent   Obama_wins
##  Min.   :   1.0   Min.   : NA    Min.   :0.000001     Min.   :0.0000
##  1st Qu.:  94.5   1st Qu.: NA    1st Qu.:0.021663     1st Qu.:0.0000
##  Median : 184.0   Median : NA    Median :0.228862     Median :0.0000
##  Mean   : 333.7   Mean   :NaN    Mean   :0.336941     Mean   :0.3052
##  3rd Qu.: 435.5   3rd Qu.: NA    3rd Qu.:0.602036     3rd Qu.:1.0000
##  Max.   :2986.0   Max.   : NA    Max.   :0.999958     Max.   :1.0000
##                   NA's   :1131   NA's   :4            NA's   :4
##  Obama_win_probability
##  Min.   :0.000001
##  1st Qu.:0.021663
##  Median :0.228862
##  Mean   :0.336941
##  3rd Qu.:0.602036
##  Max.   :0.999958
##  NA's   :4
```