Optimus\_Project

Karan Dassi

4/19/2019

Table of Contents

[Loading Libraries and Dataset: 1](#_Toc6787717)

[Data Transformation into useable format! 4](#_Toc6787718)

[Applying logistic model to the whole dataset and I am trying to test a logistic model here for vh12g and then after cross validations will go ahead to find values for vh14g: 7](#_Toc6787719)

[Predictions and Confusion matrix Cross Validation for original dataset: 13](#_Toc6787720)

[Let’s go ahead and use ROC Curve to check the accuracy of our test: 15](#_Toc6787721)

[Confusion matrix with threshold = 0.40. 16](#_Toc6787722)

[Looking at the summary of Logit, let us try to remove some variables with a lot of nan values and not significant: 18](#_Toc6787723)

[Applying logistic model to the reduced dataset: 21](#_Toc6787724)

[Let us go ahead and make predictions and cross validate for our reduced model: 24](#_Toc6787725)

[Let’s go ahead and use ROC Curve to check the accuracy of our reduced test: 26](#_Toc6787726)

[Confusion matrix with threshold = 0.40 for our reduced model. 27](#_Toc6787727)

[Let’s go ahead and now try to apply variable selection method and I am choosing Stepwise: 29](#_Toc6787728)

[Applying logistic model to the reduced dataset from Stepwise and alread removed variables: 52](#_Toc6787729)

[Let us go ahead and make predictions and cross validate for our reduced model from Stepwise: 55](#_Toc6787730)

[Confusion matrix with threshold = 0.40. 57](#_Toc6787731)

[Check for multicollinearity: 58](#_Toc6787732)

[Durbin Watson test to check serial correlation: 59](#_Toc6787733)

[Finally, let us predict the voter probabilities for voting in 2014 but we will also have to add vh14p and g12\_precint\_turnout: 59](#_Toc6787734)

[Final output csv file: 64](#_Toc6787735)

## Loading Libraries and Dataset:

suppressPackageStartupMessages(library(tidyverse))  
suppressPackageStartupMessages(library(corrplot))  
suppressPackageStartupMessages(library(Hmisc))  
suppressPackageStartupMessages(library(ROCR))  
suppressPackageStartupMessages(library(ggplot2))  
suppressPackageStartupMessages(library(GGally))  
suppressPackageStartupMessages(library(lmtest))  
suppressPackageStartupMessages(library(car))  
suppressPackageStartupMessages(library(perturb))  
  
voter <- read.csv("..\\voterfile\\voterfile.csv")  
str(voter)

## 'data.frame': 50000 obs. of 39 variables:  
## $ optimus\_id : int 861681 1084850 644435 57683 167371 974034 660415 313964 720804 547190 ...  
## $ age : num 69 20 28 78 68 69 53 36 53 30 ...  
## $ party : Factor w/ 8 levels "American Independent",..: 8 1 6 1 2 2 8 2 2 6 ...  
## $ ethnicity : Factor w/ 6 levels "African-American",..: 3 3 3 3 5 3 3 1 3 2 ...  
## $ maritalstatus : Factor w/ 3 levels "Married","nan",..: 1 2 2 1 2 2 1 2 1 1 ...  
## $ dwellingtype : Factor w/ 4 levels "Large mult wo/Apt number",..: 3 2 2 2 2 3 2 2 3 3 ...  
## $ income : Factor w/ 6 levels "0-35k","125k-200k",..: 5 6 6 6 6 5 6 6 4 3 ...  
## $ education : Factor w/ 12 levels "Bach Degree - Extremely Likely",..: 1 9 9 9 9 6 9 9 5 10 ...  
## $ cd : num 4 2 3 3 4 2 3 4 3 3 ...  
## $ dma : Factor w/ 4 levels "LAS VEGAS DMA (EST.)",..: 1 3 1 1 1 3 1 1 1 1 ...  
## $ occupationindustry : Factor w/ 19 levels "Civil Servant",..: 13 15 15 15 15 15 15 15 12 15 ...  
## $ vh14p : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ vh12g : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ vh12p : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ vh10g : int 1 0 0 0 1 0 0 1 1 1 ...  
## $ vh10p : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ vh08g : int 1 0 0 0 1 1 0 0 0 1 ...  
## $ vh08p : int 0 0 0 0 1 0 0 0 0 0 ...  
## $ vh06g : int 1 0 1 0 1 1 0 1 0 1 ...  
## $ vh06p : int 0 0 0 0 1 1 0 0 0 0 ...  
## $ vh04g : int 1 0 1 0 1 1 0 1 0 0 ...  
## $ vh04p : int 0 0 0 0 1 1 0 0 0 0 ...  
## $ vh02g : int 1 0 0 0 1 1 0 1 0 1 ...  
## $ vh02p : int 0 0 0 0 1 0 0 0 0 0 ...  
## $ vh00g : int 1 0 0 0 1 1 1 1 0 1 ...  
## $ vh00p : int 0 0 0 0 1 0 0 0 0 0 ...  
## $ net\_worth : Factor w/ 9 levels "$1-4999","$10000-24999",..: 3 9 5 9 9 5 9 9 3 9 ...  
## $ petowner\_dog : Factor w/ 2 levels "nan","Yes": 1 1 1 1 1 1 1 1 2 1 ...  
## $ intrst\_nascar\_in\_hh : Factor w/ 2 levels "nan","Yes": 1 1 2 1 1 1 1 1 1 1 ...  
## $ intrst\_musical\_instruments\_in\_hh: Factor w/ 2 levels "nan","Yes": 1 1 1 1 1 1 1 1 1 1 ...  
## $ donates\_to\_liberal\_causes : Factor w/ 2 levels "nan","Yes": 1 1 1 1 1 1 1 1 1 1 ...  
## $ donates\_to\_conservative\_causes : Factor w/ 2 levels "nan","Yes": 1 1 1 1 1 1 1 1 1 1 ...  
## $ home\_owner\_or\_renter : Factor w/ 3 levels "Likely Homeowner",..: 1 3 3 3 3 1 3 3 1 3 ...  
## $ g08\_precinct\_turnout : num 0.56 0.84 0.49 0.84 0.71 0.69 0.75 0.67 0.75 0.62 ...  
## $ g10\_precinct\_turnout : num 0.54 0.82 0.34 0.79 0.66 0.6 0.64 0.56 0.6 0.47 ...  
## $ g12\_precinct\_turnout : num 0.75 0.92 0.7 0.91 0.81 0.8 0.81 0.79 0.79 0.73 ...  
## $ p08\_precinct\_turnout : num 0.17 0.47 0.04 0.24 0.19 0.16 0.16 0.12 0.13 0.06 ...  
## $ p10\_precinct\_turnout : num 0.32 0.62 0.09 0.46 0.37 0.23 0.3 0.24 0.25 0.14 ...  
## $ p12\_precinct\_turnout : num 0.24 0.47 0.06 0.3 0.34 0.17 0.17 0.15 0.13 0.09 ...

head(voter)

## optimus\_id age party ethnicity maritalstatus  
## 1 861681 69 Republican European Married  
## 2 1084850 20 American Independent European nan  
## 3 644435 28 Non-Partisan European nan  
## 4 57683 78 American Independent European Married  
## 5 167371 68 Democratic nan nan  
## 6 974034 69 Democratic European nan  
## dwellingtype income education cd  
## 1 Single Family Dwelling Unit 75k-125k Bach Degree - Extremely Likely 4  
## 2 nan Unknown nan 2  
## 3 nan Unknown nan 3  
## 4 nan Unknown nan 3  
## 5 nan Unknown nan 4  
## 6 Single Family Dwelling Unit 75k-125k HS Diploma - Likely 2  
## dma occupationindustry vh14p vh12g vh12p vh10g vh10p  
## 1 LAS VEGAS DMA (EST.) Medical 0 0 0 1 0  
## 2 RENO DMA (EST.) nan 0 0 0 0 0  
## 3 LAS VEGAS DMA (EST.) nan 0 0 0 0 0  
## 4 LAS VEGAS DMA (EST.) nan 0 0 0 0 0  
## 5 LAS VEGAS DMA (EST.) nan 0 0 0 1 0  
## 6 RENO DMA (EST.) nan 0 0 0 0 0  
## vh08g vh08p vh06g vh06p vh04g vh04p vh02g vh02p vh00g vh00p  
## 1 1 0 1 0 1 0 1 0 1 0  
## 2 0 0 0 0 0 0 0 0 0 0  
## 3 0 0 1 0 1 0 0 0 0 0  
## 4 0 0 0 0 0 0 0 0 0 0  
## 5 1 1 1 1 1 1 1 1 1 1  
## 6 1 0 1 1 1 1 1 0 1 0  
## net\_worth petowner\_dog intrst\_nascar\_in\_hh  
## 1 $100000-249999 nan nan  
## 2 nan nan nan  
## 3 $250000-499999 nan Yes  
## 4 nan nan nan  
## 5 nan nan nan  
## 6 $250000-499999 nan nan  
## intrst\_musical\_instruments\_in\_hh donates\_to\_liberal\_causes  
## 1 nan nan  
## 2 nan nan  
## 3 nan nan  
## 4 nan nan  
## 5 nan nan  
## 6 nan nan  
## donates\_to\_conservative\_causes home\_owner\_or\_renter g08\_precinct\_turnout  
## 1 nan Likely Homeowner 0.56  
## 2 nan nan 0.84  
## 3 nan nan 0.49  
## 4 nan nan 0.84  
## 5 nan nan 0.71  
## 6 nan Likely Homeowner 0.69  
## g10\_precinct\_turnout g12\_precinct\_turnout p08\_precinct\_turnout  
## 1 0.54 0.75 0.17  
## 2 0.82 0.92 0.47  
## 3 0.34 0.70 0.04  
## 4 0.79 0.91 0.24  
## 5 0.66 0.81 0.19  
## 6 0.60 0.80 0.16  
## p10\_precinct\_turnout p12\_precinct\_turnout  
## 1 0.32 0.24  
## 2 0.62 0.47  
## 3 0.09 0.06  
## 4 0.46 0.30  
## 5 0.37 0.34  
## 6 0.23 0.17

## Data Transformation into useable format!

#Transforming some variables to factors which were not by default.   
  
voter %>%  
 mutate(vh14p = as.factor(vh14p)) %>%  
 mutate(vh12p = as.factor(vh12p)) %>%  
 mutate(vh10p = as.factor(vh10p)) %>%  
 mutate(vh08p = as.factor(vh08p)) %>%  
 mutate(vh06p = as.factor(vh06p)) %>%  
 mutate(vh04p = as.factor(vh04p)) %>%  
 mutate(vh02p = as.factor(vh02p)) %>%  
 mutate(vh00p = as.factor(vh00p)) %>%  
 mutate(vh12g = as.factor(vh12g)) %>%  
 mutate(vh10g = as.factor(vh10g)) %>%  
 mutate(vh08g = as.factor(vh08g)) %>%  
 mutate(vh06g = as.factor(vh06g)) %>%  
 mutate(vh04g = as.factor(vh04g)) %>%  
 mutate(vh02g = as.factor(vh02g)) %>%  
 mutate(vh00g = as.factor(vh00g)) ->  
 voter  
  
#Removing vh14p, optimus ID and g12\_precint\_turnout because of logic as to test our model we are using only data till 2012.  
  
voter %>%  
 select(-vh14p, -optimus\_id, -g12\_precinct\_turnout) ->  
 voter.rm  
str(voter.rm)

## 'data.frame': 50000 obs. of 36 variables:  
## $ age : num 69 20 28 78 68 69 53 36 53 30 ...  
## $ party : Factor w/ 8 levels "American Independent",..: 8 1 6 1 2 2 8 2 2 6 ...  
## $ ethnicity : Factor w/ 6 levels "African-American",..: 3 3 3 3 5 3 3 1 3 2 ...  
## $ maritalstatus : Factor w/ 3 levels "Married","nan",..: 1 2 2 1 2 2 1 2 1 1 ...  
## $ dwellingtype : Factor w/ 4 levels "Large mult wo/Apt number",..: 3 2 2 2 2 3 2 2 3 3 ...  
## $ income : Factor w/ 6 levels "0-35k","125k-200k",..: 5 6 6 6 6 5 6 6 4 3 ...  
## $ education : Factor w/ 12 levels "Bach Degree - Extremely Likely",..: 1 9 9 9 9 6 9 9 5 10 ...  
## $ cd : num 4 2 3 3 4 2 3 4 3 3 ...  
## $ dma : Factor w/ 4 levels "LAS VEGAS DMA (EST.)",..: 1 3 1 1 1 3 1 1 1 1 ...  
## $ occupationindustry : Factor w/ 19 levels "Civil Servant",..: 13 15 15 15 15 15 15 15 12 15 ...  
## $ vh12g : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ vh12p : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ vh10g : Factor w/ 2 levels "0","1": 2 1 1 1 2 1 1 2 2 2 ...  
## $ vh10p : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ vh08g : Factor w/ 2 levels "0","1": 2 1 1 1 2 2 1 1 1 2 ...  
## $ vh08p : Factor w/ 2 levels "0","1": 1 1 1 1 2 1 1 1 1 1 ...  
## $ vh06g : Factor w/ 2 levels "0","1": 2 1 2 1 2 2 1 2 1 2 ...  
## $ vh06p : Factor w/ 2 levels "0","1": 1 1 1 1 2 2 1 1 1 1 ...  
## $ vh04g : Factor w/ 2 levels "0","1": 2 1 2 1 2 2 1 2 1 1 ...  
## $ vh04p : Factor w/ 2 levels "0","1": 1 1 1 1 2 2 1 1 1 1 ...  
## $ vh02g : Factor w/ 2 levels "0","1": 2 1 1 1 2 2 1 2 1 2 ...  
## $ vh02p : Factor w/ 2 levels "0","1": 1 1 1 1 2 1 1 1 1 1 ...  
## $ vh00g : Factor w/ 2 levels "0","1": 2 1 1 1 2 2 2 2 1 2 ...  
## $ vh00p : Factor w/ 2 levels "0","1": 1 1 1 1 2 1 1 1 1 1 ...  
## $ net\_worth : Factor w/ 9 levels "$1-4999","$10000-24999",..: 3 9 5 9 9 5 9 9 3 9 ...  
## $ petowner\_dog : Factor w/ 2 levels "nan","Yes": 1 1 1 1 1 1 1 1 2 1 ...  
## $ intrst\_nascar\_in\_hh : Factor w/ 2 levels "nan","Yes": 1 1 2 1 1 1 1 1 1 1 ...  
## $ intrst\_musical\_instruments\_in\_hh: Factor w/ 2 levels "nan","Yes": 1 1 1 1 1 1 1 1 1 1 ...  
## $ donates\_to\_liberal\_causes : Factor w/ 2 levels "nan","Yes": 1 1 1 1 1 1 1 1 1 1 ...  
## $ donates\_to\_conservative\_causes : Factor w/ 2 levels "nan","Yes": 1 1 1 1 1 1 1 1 1 1 ...  
## $ home\_owner\_or\_renter : Factor w/ 3 levels "Likely Homeowner",..: 1 3 3 3 3 1 3 3 1 3 ...  
## $ g08\_precinct\_turnout : num 0.56 0.84 0.49 0.84 0.71 0.69 0.75 0.67 0.75 0.62 ...  
## $ g10\_precinct\_turnout : num 0.54 0.82 0.34 0.79 0.66 0.6 0.64 0.56 0.6 0.47 ...  
## $ p08\_precinct\_turnout : num 0.17 0.47 0.04 0.24 0.19 0.16 0.16 0.12 0.13 0.06 ...  
## $ p10\_precinct\_turnout : num 0.32 0.62 0.09 0.46 0.37 0.23 0.3 0.24 0.25 0.14 ...  
## $ p12\_precinct\_turnout : num 0.24 0.47 0.06 0.3 0.34 0.17 0.17 0.15 0.13 0.09 ...

## Applying logistic model to the whole dataset and I am trying to test a logistic model here for vh12g and then after cross validations will go ahead to find values for vh14g:

#Frequency of 0's and 1's in vh12g (voter history for 2012 general election)  
voter.rm %>%  
 group\_by(vh12g) %>%  
 summarise (n = n()) %>%  
 mutate(freq = n / sum(n))

## # A tibble: 2 x 3  
## vh12g n freq  
## <fct> <int> <dbl>  
## 1 0 36805 0.736  
## 2 1 13195 0.264

#We can easily see that there are a lot of 0's (No's) i.e. not a lot of people came out to vote  
  
voter\_logit <- glm(vh12g ~ ., data=voter.rm, family=binomial(link="logit"), na.action = na.exclude)  
  
summary(voter\_logit)

##   
## Call:  
## glm(formula = vh12g ~ ., family = binomial(link = "logit"), data = voter.rm,   
## na.action = na.exclude)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.1168 -0.2483 -0.1410 0.1521 3.2337   
##   
## Coefficients:  
## Estimate  
## (Intercept) -6.528e+00  
## age 1.835e-02  
## partyDemocratic 2.242e-01  
## partyGreen 9.238e-01  
## partyLibertarian 6.834e-01  
## partyNatural Law 5.443e-01  
## partyNon-Partisan 2.454e-01  
## partyOther 2.106e-02  
## partyRepublican 2.497e-01  
## ethnicityAsian -4.290e-01  
## ethnicityEuropean 7.786e-02  
## ethnicityHispanic -2.015e-01  
## ethnicitynan -7.694e-03  
## ethnicityOther -1.309e-01  
## maritalstatusnan -1.433e-01  
## maritalstatusNon-Traditional 7.385e-03  
## dwellingtypenan -4.039e-03  
## dwellingtypeSingle Family Dwelling Unit -1.574e-01  
## dwellingtypeSmall Mult or large mult w/apt number -2.077e-01  
## income125k-200k 4.812e-02  
## income200k+ 1.619e-01  
## income35k-75k -1.571e-01  
## income75k-125k -3.584e-02  
## incomeUnknown 5.153e-03  
## educationBach Degree - Likely -5.752e-02  
## educationGrad Degree - Extremely Likely 2.071e-03  
## educationGrad Degree - Likely 2.319e-02  
## educationHS Diploma - Extremely Likely 2.618e-02  
## educationHS Diploma - Likely 4.152e-02  
## educationLess than HS Diploma - Ex Like -1.103e+01  
## educationLess than HS Diploma - Likely -1.712e-01  
## educationnan -2.297e-01  
## educationSome College - Likely 1.707e-05  
## educationSome College -Extremely Likely 1.233e-01  
## educationVocational Technical Degree - Extremely Likely 2.194e-01  
## cd -6.301e-02  
## dmaLOS ANGELES DMA (EST.) 9.525e-01  
## dmaRENO DMA (EST.) 4.905e-02  
## dmaSALT LAKE CITY DMA (EST.) 4.211e-01  
## occupationindustryClerical/Office 4.630e-01  
## occupationindustryComputer Professional 5.309e-01  
## occupationindustryCreative Arts 1.578e+00  
## occupationindustryEducation 8.444e-01  
## occupationindustryEngineering 2.347e-01  
## occupationindustryFinancial Services 4.547e-01  
## occupationindustryFood Services 7.143e-01  
## occupationindustryLegal 8.724e-01  
## occupationindustryMaintenance Services -7.103e-01  
## occupationindustryManagement 6.211e-01  
## occupationindustryManufacturing 5.633e-01  
## occupationindustryMedical 6.983e-01  
## occupationindustryMilitary 7.328e-01  
## occupationindustrynan 5.503e-01  
## occupationindustryOther 5.499e-01  
## occupationindustrySales/Marketing 9.929e-01  
## occupationindustryScientific -1.266e+00  
## occupationindustrySkilled Trades 3.797e-01  
## vh12p1 3.224e+00  
## vh10g1 2.505e+00  
## vh10p1 4.716e-01  
## vh08g1 9.665e-01  
## vh08p1 1.544e-01  
## vh06g1 3.845e-01  
## vh06p1 -1.502e-01  
## vh04g1 3.327e-01  
## vh04p1 9.518e-02  
## vh02g1 -7.466e-03  
## vh02p1 -1.142e-01  
## vh00g1 2.664e-01  
## vh00p1 4.582e-02  
## net\_worth$10000-24999 -9.122e-02  
## net\_worth$100000-249999 1.321e-01  
## net\_worth$25000-49999 -1.541e-01  
## net\_worth$250000-499999 -2.416e-02  
## net\_worth$499999+ -1.262e-01  
## net\_worth$5000-9999 -1.058e-01  
## net\_worth$50000-99999 1.577e-01  
## net\_worthnan 1.671e-03  
## petowner\_dogYes 2.839e-01  
## intrst\_nascar\_in\_hhYes -2.532e-02  
## intrst\_musical\_instruments\_in\_hhYes 2.945e-01  
## donates\_to\_liberal\_causesYes -1.428e-01  
## donates\_to\_conservative\_causesYes -2.581e-01  
## home\_owner\_or\_renterLikely Renter -7.464e-02  
## home\_owner\_or\_renternan -1.548e-02  
## g08\_precinct\_turnout 1.210e+00  
## g10\_precinct\_turnout 1.613e-01  
## p08\_precinct\_turnout 8.869e-01  
## p10\_precinct\_turnout -5.121e-01  
## p12\_precinct\_turnout -9.549e-01  
## Std. Error z value  
## (Intercept) 1.262e+00 -5.173  
## age 1.264e-03 14.519  
## partyDemocratic 8.705e-02 2.576  
## partyGreen 6.626e-01 1.394  
## partyLibertarian 2.268e-01 3.014  
## partyNatural Law 1.690e+00 0.322  
## partyNon-Partisan 9.295e-02 2.640  
## partyOther 2.392e-01 0.088  
## partyRepublican 8.731e-02 2.860  
## ethnicityAsian 1.878e-01 -2.284  
## ethnicityEuropean 1.501e-01 0.519  
## ethnicityHispanic 1.555e-01 -1.295  
## ethnicitynan 1.572e-01 -0.049  
## ethnicityOther 1.879e-01 -0.697  
## maritalstatusnan 3.577e-02 -4.006  
## maritalstatusNon-Traditional 6.379e-02 0.116  
## dwellingtypenan 1.148e+00 -0.004  
## dwellingtypeSingle Family Dwelling Unit 1.146e+00 -0.137  
## dwellingtypeSmall Mult or large mult w/apt number 1.148e+00 -0.181  
## income125k-200k 7.749e-02 0.621  
## income200k+ 9.674e-02 1.674  
## income35k-75k 6.193e-02 -2.537  
## income75k-125k 6.491e-02 -0.552  
## incomeUnknown 9.206e-02 0.056  
## educationBach Degree - Likely 9.066e-02 -0.635  
## educationGrad Degree - Extremely Likely 7.314e-02 0.028  
## educationGrad Degree - Likely 9.900e-02 0.234  
## educationHS Diploma - Extremely Likely 6.284e-02 0.417  
## educationHS Diploma - Likely 8.355e-02 0.497  
## educationLess than HS Diploma - Ex Like 1.159e+02 -0.095  
## educationLess than HS Diploma - Likely 1.187e-01 -1.442  
## educationnan 8.806e-02 -2.609  
## educationSome College - Likely 7.226e-02 0.000  
## educationSome College -Extremely Likely 8.603e-02 1.434  
## educationVocational Technical Degree - Extremely Likely 3.503e-01 0.626  
## cd 1.748e-02 -3.604  
## dmaLOS ANGELES DMA (EST.) 6.044e-01 1.576  
## dmaRENO DMA (EST.) 4.555e-02 1.077  
## dmaSALT LAKE CITY DMA (EST.) 1.169e-01 3.601  
## occupationindustryClerical/Office 4.544e-01 1.019  
## occupationindustryComputer Professional 6.937e-01 0.765  
## occupationindustryCreative Arts 1.044e+00 1.511  
## occupationindustryEducation 4.889e-01 1.727  
## occupationindustryEngineering 5.233e-01 0.448  
## occupationindustryFinancial Services 4.569e-01 0.995  
## occupationindustryFood Services 5.586e-01 1.279  
## occupationindustryLegal 5.784e-01 1.508  
## occupationindustryMaintenance Services 7.165e-01 -0.991  
## occupationindustryManagement 4.500e-01 1.380  
## occupationindustryManufacturing 4.626e-01 1.218  
## occupationindustryMedical 4.522e-01 1.544  
## occupationindustryMilitary 5.313e-01 1.379  
## occupationindustrynan 4.436e-01 1.241  
## occupationindustryOther 4.478e-01 1.228  
## occupationindustrySales/Marketing 7.518e-01 1.321  
## occupationindustryScientific 1.330e+00 -0.953  
## occupationindustrySkilled Trades 4.636e-01 0.819  
## vh12p1 7.555e-02 42.669  
## vh10g1 5.273e-02 47.510  
## vh10p1 4.450e-02 10.598  
## vh08g1 4.080e-02 23.687  
## vh08p1 4.363e-02 3.539  
## vh06g1 6.005e-02 6.404  
## vh06p1 5.248e-02 -2.861  
## vh04g1 4.696e-02 7.083  
## vh04p1 4.304e-02 2.211  
## vh02g1 6.061e-02 -0.123  
## vh02p1 5.067e-02 -2.253  
## vh00g1 3.834e-02 6.947  
## vh00p1 4.691e-02 0.977  
## net\_worth$10000-24999 1.812e-01 -0.503  
## net\_worth$100000-249999 1.365e-01 0.968  
## net\_worth$25000-49999 1.557e-01 -0.989  
## net\_worth$250000-499999 1.392e-01 -0.174  
## net\_worth$499999+ 1.480e-01 -0.853  
## net\_worth$5000-9999 2.048e-01 -0.517  
## net\_worth$50000-99999 1.607e-01 0.981  
## net\_worthnan 1.332e-01 0.013  
## petowner\_dogYes 5.379e-02 5.279  
## intrst\_nascar\_in\_hhYes 6.882e-02 -0.368  
## intrst\_musical\_instruments\_in\_hhYes 1.378e-01 2.138  
## donates\_to\_liberal\_causesYes 3.457e-01 -0.413  
## donates\_to\_conservative\_causesYes 2.674e-01 -0.965  
## home\_owner\_or\_renterLikely Renter 1.671e-01 -0.447  
## home\_owner\_or\_renternan 5.398e-02 -0.287  
## g08\_precinct\_turnout 5.659e-01 2.139  
## g10\_precinct\_turnout 6.545e-01 0.246  
## p08\_precinct\_turnout 5.702e-01 1.555  
## p10\_precinct\_turnout 6.114e-01 -0.837  
## p12\_precinct\_turnout 4.718e-01 -2.024  
## Pr(>|z|)   
## (Intercept) 2.31e-07 \*\*\*  
## age < 2e-16 \*\*\*  
## partyDemocratic 0.009989 \*\*   
## partyGreen 0.163235   
## partyLibertarian 0.002582 \*\*   
## partyNatural Law 0.747369   
## partyNon-Partisan 0.008287 \*\*   
## partyOther 0.929845   
## partyRepublican 0.004235 \*\*   
## ethnicityAsian 0.022359 \*   
## ethnicityEuropean 0.603895   
## ethnicityHispanic 0.195168   
## ethnicitynan 0.960973   
## ethnicityOther 0.486113   
## maritalstatusnan 6.19e-05 \*\*\*  
## maritalstatusNon-Traditional 0.907832   
## dwellingtypenan 0.997192   
## dwellingtypeSingle Family Dwelling Unit 0.890744   
## dwellingtypeSmall Mult or large mult w/apt number 0.856352   
## income125k-200k 0.534569   
## income200k+ 0.094205 .   
## income35k-75k 0.011175 \*   
## income75k-125k 0.580787   
## incomeUnknown 0.955360   
## educationBach Degree - Likely 0.525725   
## educationGrad Degree - Extremely Likely 0.977413   
## educationGrad Degree - Likely 0.814808   
## educationHS Diploma - Extremely Likely 0.676944   
## educationHS Diploma - Likely 0.619233   
## educationLess than HS Diploma - Ex Like 0.924212   
## educationLess than HS Diploma - Likely 0.149203   
## educationnan 0.009085 \*\*   
## educationSome College - Likely 0.999812   
## educationSome College -Extremely Likely 0.151695   
## educationVocational Technical Degree - Extremely Likely 0.531095   
## cd 0.000313 \*\*\*  
## dmaLOS ANGELES DMA (EST.) 0.115055   
## dmaRENO DMA (EST.) 0.281597   
## dmaSALT LAKE CITY DMA (EST.) 0.000317 \*\*\*  
## occupationindustryClerical/Office 0.308219   
## occupationindustryComputer Professional 0.444091   
## occupationindustryCreative Arts 0.130747   
## occupationindustryEducation 0.084108 .   
## occupationindustryEngineering 0.653842   
## occupationindustryFinancial Services 0.319734   
## occupationindustryFood Services 0.201020   
## occupationindustryLegal 0.131506   
## occupationindustryMaintenance Services 0.321476   
## occupationindustryManagement 0.167506   
## occupationindustryManufacturing 0.223388   
## occupationindustryMedical 0.122545   
## occupationindustryMilitary 0.167813   
## occupationindustrynan 0.214781   
## occupationindustryOther 0.219468   
## occupationindustrySales/Marketing 0.186593   
## occupationindustryScientific 0.340827   
## occupationindustrySkilled Trades 0.412852   
## vh12p1 < 2e-16 \*\*\*  
## vh10g1 < 2e-16 \*\*\*  
## vh10p1 < 2e-16 \*\*\*  
## vh08g1 < 2e-16 \*\*\*  
## vh08p1 0.000402 \*\*\*  
## vh06g1 1.52e-10 \*\*\*  
## vh06p1 0.004221 \*\*   
## vh04g1 1.41e-12 \*\*\*  
## vh04p1 0.027002 \*   
## vh02g1 0.901963   
## vh02p1 0.024275 \*   
## vh00g1 3.73e-12 \*\*\*  
## vh00p1 0.328694   
## net\_worth$10000-24999 0.614620   
## net\_worth$100000-249999 0.333017   
## net\_worth$25000-49999 0.322431   
## net\_worth$250000-499999 0.862201   
## net\_worth$499999+ 0.393626   
## net\_worth$5000-9999 0.605414   
## net\_worth$50000-99999 0.326642   
## net\_worthnan 0.989987   
## petowner\_dogYes 1.30e-07 \*\*\*  
## intrst\_nascar\_in\_hhYes 0.713000   
## intrst\_musical\_instruments\_in\_hhYes 0.032534 \*   
## donates\_to\_liberal\_causesYes 0.679497   
## donates\_to\_conservative\_causesYes 0.334361   
## home\_owner\_or\_renterLikely Renter 0.655122   
## home\_owner\_or\_renternan 0.774270   
## g08\_precinct\_turnout 0.032432 \*   
## g10\_precinct\_turnout 0.805393   
## p08\_precinct\_turnout 0.119841   
## p10\_precinct\_turnout 0.402324   
## p12\_precinct\_turnout 0.042978 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 57682 on 49977 degrees of freedom  
## Residual deviance: 25234 on 49888 degrees of freedom  
## (22 observations deleted due to missingness)  
## AIC: 25414  
##   
## Number of Fisher Scoring iterations: 11

**The AIC here is 25414 and there are also a lot non significant values**

## Predictions and Confusion matrix Cross Validation for original dataset:

attach(voter.rm)   
set.seed(1)  
train <- sample(1:nrow(voter.rm), 0.7\*nrow(voter.rm))  
test <- seq(1:nrow(voter.rm))[-train]  
  
voter.logit.train <- glm(vh12g ~ ., family=binomial(link="logit"), data=voter.rm[train,])  
  
# Predicted values using the fitted train model and the test data for vh12g  
  
voter.logit.test=predict(voter.logit.train, voter.rm, type="response")[test]  
  
voter.logit.test[1:10]

## 3 4 8 13 14 20   
## 0.010304041 0.017687321 0.207924999 0.203661416 0.188985589 0.946314658   
## 23 28 31 33   
## 0.557745632 0.412599800 0.031182637 0.009040557

#Convert proportions to actual 0's or 1's  
  
voter.pred.test = ifelse(voter.logit.test>0.5, 1,0)  
voter.pred.test[1:10] # List first 10

## 3 4 8 13 14 20 23 28 31 33   
## 0 0 0 0 0 1 1 0 0 0

# Cross tabulate Prediction with Actual  
  
conf.mat <- table("Predicted"=voter.pred.test, "Actual"=vh12g[test])   
  
colnames(conf.mat) <- c("No", "Yes")  
rownames(conf.mat) <- c("No", "Yes")  
  
#Final Confusion Matrix  
conf.mat

## Actual  
## Predicted No Yes  
## No 10207 931  
## Yes 883 2973

#We can see that our model does a very good job at predicting Yes's and No's  
  
# Computing Fit Statistics  
  
TruN <- conf.mat[1,1] # True negatives  
TruP <- conf.mat[2,2] # True positives  
FalN <- conf.mat[1,2] # False negatives  
FalP <- conf.mat[2,1] # False positives  
TotN <- conf.mat[1,1] + conf.mat[2,1] # Total negatives  
TotP <- conf.mat[1,2] + conf.mat[2,2] # Total positives  
Tot <- TotN+TotP # Total  
  
# Now let's use these to compute accuracy and error rates  
  
Accuracy.Rate <- (TruN + TruP) / Tot  
  
#It is quite clear that our accuracy overall is pretty good.  
Accuracy.Rate

## [1] 0.8790183

Error.Rate <- (FalN + FalP) / Tot  
#The overall error rate is quite bad  
Error.Rate

## [1] 0.1209817

# Sensitivity -- rate of correct positives  
  
Sensitivity <- TruP / TotP   
Sensitivity

## [1] 0.7615266

#Prediction for true positives is quite bad  
  
# Specificity -- rate of correct negatives  
  
Specificity <- TruN / TotN   
Specificity

## [1] 0.9203787

#Predicting true negatives is really good with this model.  
  
# False Positive Rate   
FalseP.Rate <- 1 - Specificity  
FalseP.Rate

## [1] 0.07962128

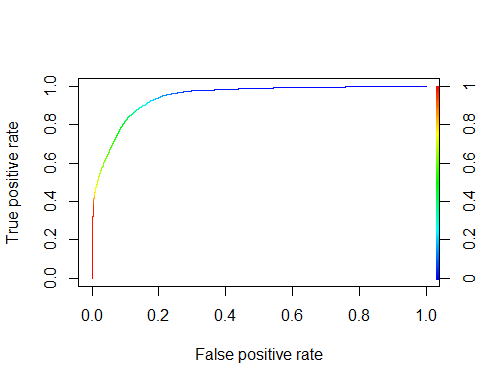
logit.rates.50 <- c(Accuracy.Rate, Error.Rate, Sensitivity, Specificity, FalseP.Rate)  
  
names(logit.rates.50) <- c("Accuracy Rate", "Error Rate", "Sensitivity", "Specificity", "False Positives")  
  
print(logit.rates.50, digits=2)

## Accuracy Rate Error Rate Sensitivity Specificity   
## 0.88 0.12 0.76 0.92   
## False Positives   
## 0.08

**From the cross validation, we can easily see that this model does a pretty good job at predicting both True Negatives and True Positives and has a very good accuracy rate!**  
**Also, I believe we should not change lambda values (threshold) because 0.5 threshold is very logical and gives us a good model for prediction, though we could reduce the threshold values from a 0.50 to a 0.40, let us test that out as well.**

## Let’s go ahead and use ROC Curve to check the accuracy of our test:

pred <- prediction(voter.logit.test, vh12g[test])   
  
perf <- performance(pred,"tpr","fpr")  
plot(perf, colorize=T)



auc <- performance(pred,"auc")  
  
c(auc@y.name[[1]], auc@y.values[[1]])

## [1] "Area under the ROC curve" "0.943670984604399"

**The result from ROC is pretty good with area under curve being closer to 0.95 which is considered very good!**

## Confusion matrix with threshold = 0.40.

voter.pred.test = ifelse(voter.logit.test>0.40, 1,0)  
voter.pred.test[1:10] # List first 10

## 3 4 8 13 14 20 23 28 31 33   
## 0 0 0 0 0 1 1 1 0 0

# Cross tabulate Prediction with Actual  
  
conf.mat <- table("Predicted"=voter.pred.test, "Actual"=vh12g[test])   
  
colnames(conf.mat) <- c("No", "Yes")  
rownames(conf.mat) <- c("No", "Yes")  
  
#Final Confusion Matrix  
conf.mat

## Actual  
## Predicted No Yes  
## No 9845 606  
## Yes 1245 3298

# Computing Fit Statistics  
  
TruN <- conf.mat[1,1] # True negatives  
TruP <- conf.mat[2,2] # True positives  
FalN <- conf.mat[1,2] # False negatives  
FalP <- conf.mat[2,1] # False positives  
TotN <- conf.mat[1,1] + conf.mat[2,1] # Total negatives  
TotP <- conf.mat[1,2] + conf.mat[2,2] # Total positives  
Tot <- TotN+TotP # Total  
  
# Now let's use these to compute accuracy and error rates  
  
Accuracy.Rate <- (TruN + TruP) / Tot  
  
#It is quite clear that our accuracy overall is pretty good.  
Accuracy.Rate

## [1] 0.8765506

Error.Rate <- (FalN + FalP) / Tot  
#The overall error rate is quite bad  
Error.Rate

## [1] 0.1234494

# Sensitivity -- rate of correct positives  
  
Sensitivity <- TruP / TotP   
Sensitivity

## [1] 0.8447746

#Prediction for true positives is quite bad  
  
# Specificity -- rate of correct negatives  
  
Specificity <- TruN / TotN   
Specificity

## [1] 0.8877367

#Predicting true negatives is really good with this model.  
  
# False Positive Rate   
FalseP.Rate <- 1 - Specificity  
FalseP.Rate

## [1] 0.1122633

logit.rates.40 <- c(Accuracy.Rate, Error.Rate, Sensitivity, Specificity, FalseP.Rate)  
  
names(logit.rates.40) <- c("Accuracy Rate", "Error Rate", "Sensitivity", "Specificity", "False Positives")  
  
print(logit.rates.40, digits=2)

## Accuracy Rate Error Rate Sensitivity Specificity   
## 0.88 0.12 0.84 0.89   
## False Positives   
## 0.11

logit.fit.stats.compare <- rbind(logit.rates.50, logit.rates.40)  
print(logit.fit.stats.compare, digits=2)

## Accuracy Rate Error Rate Sensitivity Specificity  
## logit.rates.50 0.88 0.12 0.76 0.92  
## logit.rates.40 0.88 0.12 0.84 0.89  
## False Positives  
## logit.rates.50 0.08  
## logit.rates.40 0.11

**And to my surprise, it actually gives a better prediction than what we got at threshold = 0.5, this one seems to be much more balanced by improving a lot on sensitivity but reducing specificity a bit and accuracy is still better.**

## Looking at the summary of Logit, let us try to remove some variables with a lot of nan values and not significant:

#Marital Status  
voter.rm %>%  
 filter(maritalstatus == "nan") %>%  
 count()

## # A tibble: 1 x 1  
## n  
## <int>  
## 1 30638

#Marital status has a lot of nan values over 60 percent and the nan in marital status is quite significant leading to a biased (wrong) model, lets remove this!  
  
voter.rm %>%  
 select(-maritalstatus) ->  
 voter.clean.var  
  
#Dwelling Type  
voter.clean.var %>%  
 filter(dwellingtype == "nan") %>%  
 count()

## # A tibble: 1 x 1  
## n  
## <int>  
## 1 26083

#Dwelling type has a lot of nan values about 50 percent and is not that significant, lets remove this!  
  
voter.clean.var %>%  
 select(-dwellingtype) ->  
 voter.clean.var  
  
#education  
voter.clean.var %>%  
 filter(education == "nan") %>%  
 count()

## # A tibble: 1 x 1  
## n  
## <int>  
## 1 22410

#Education has a lot of nan values about 50 percent and is not that significant, lets remove this!  
  
voter.clean.var %>%  
 select(-education) ->  
 voter.clean.var  
  
#Occupation Industry  
voter.clean.var %>%  
 filter(occupationindustry == "nan") %>%  
 count()

## # A tibble: 1 x 1  
## n  
## <int>  
## 1 41808

#OccupationIndustry has a lot of nan values about 85 percent and is not that significant, lets remove this!  
  
voter.clean.var %>%  
 select(-occupationindustry) ->  
 voter.clean.var  
  
#Replacing all nan for the following variables to No as it seems the most logical way to me.  
voter.clean.var %>%  
 mutate(petowner\_dog = str\_replace\_all(petowner\_dog, "nan", "No")) %>%  
 mutate(petowner\_dog = as.factor(petowner\_dog)) %>%  
 mutate(intrst\_nascar\_in\_hh = str\_replace\_all(intrst\_nascar\_in\_hh, "nan", "No")) %>%  
 mutate(intrst\_nascar\_in\_hh = as.factor(intrst\_nascar\_in\_hh)) %>%  
 mutate(intrst\_musical\_instruments\_in\_hh = str\_replace\_all(intrst\_musical\_instruments\_in\_hh, "nan", "No")) %>%  
 mutate(intrst\_musical\_instruments\_in\_hh = as.factor(intrst\_musical\_instruments\_in\_hh)) %>%  
 mutate(donates\_to\_liberal\_causes = str\_replace\_all(donates\_to\_liberal\_causes, "nan", "No")) %>%  
 mutate(donates\_to\_liberal\_causes = as.factor(donates\_to\_liberal\_causes)) %>%  
 mutate(donates\_to\_conservative\_causes = str\_replace\_all(donates\_to\_conservative\_causes, "nan", "No")) %>%  
 mutate(donates\_to\_conservative\_causes = as.factor(donates\_to\_conservative\_causes))->  
 voter.clean.var  
  
str(voter.clean.var)

## 'data.frame': 50000 obs. of 32 variables:  
## $ age : num 69 20 28 78 68 69 53 36 53 30 ...  
## $ party : Factor w/ 8 levels "American Independent",..: 8 1 6 1 2 2 8 2 2 6 ...  
## $ ethnicity : Factor w/ 6 levels "African-American",..: 3 3 3 3 5 3 3 1 3 2 ...  
## $ income : Factor w/ 6 levels "0-35k","125k-200k",..: 5 6 6 6 6 5 6 6 4 3 ...  
## $ cd : num 4 2 3 3 4 2 3 4 3 3 ...  
## $ dma : Factor w/ 4 levels "LAS VEGAS DMA (EST.)",..: 1 3 1 1 1 3 1 1 1 1 ...  
## $ vh12g : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ vh12p : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ vh10g : Factor w/ 2 levels "0","1": 2 1 1 1 2 1 1 2 2 2 ...  
## $ vh10p : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ vh08g : Factor w/ 2 levels "0","1": 2 1 1 1 2 2 1 1 1 2 ...  
## $ vh08p : Factor w/ 2 levels "0","1": 1 1 1 1 2 1 1 1 1 1 ...  
## $ vh06g : Factor w/ 2 levels "0","1": 2 1 2 1 2 2 1 2 1 2 ...  
## $ vh06p : Factor w/ 2 levels "0","1": 1 1 1 1 2 2 1 1 1 1 ...  
## $ vh04g : Factor w/ 2 levels "0","1": 2 1 2 1 2 2 1 2 1 1 ...  
## $ vh04p : Factor w/ 2 levels "0","1": 1 1 1 1 2 2 1 1 1 1 ...  
## $ vh02g : Factor w/ 2 levels "0","1": 2 1 1 1 2 2 1 2 1 2 ...  
## $ vh02p : Factor w/ 2 levels "0","1": 1 1 1 1 2 1 1 1 1 1 ...  
## $ vh00g : Factor w/ 2 levels "0","1": 2 1 1 1 2 2 2 2 1 2 ...  
## $ vh00p : Factor w/ 2 levels "0","1": 1 1 1 1 2 1 1 1 1 1 ...  
## $ net\_worth : Factor w/ 9 levels "$1-4999","$10000-24999",..: 3 9 5 9 9 5 9 9 3 9 ...  
## $ petowner\_dog : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 2 1 ...  
## $ intrst\_nascar\_in\_hh : Factor w/ 2 levels "No","Yes": 1 1 2 1 1 1 1 1 1 1 ...  
## $ intrst\_musical\_instruments\_in\_hh: Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...  
## $ donates\_to\_liberal\_causes : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...  
## $ donates\_to\_conservative\_causes : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...  
## $ home\_owner\_or\_renter : Factor w/ 3 levels "Likely Homeowner",..: 1 3 3 3 3 1 3 3 1 3 ...  
## $ g08\_precinct\_turnout : num 0.56 0.84 0.49 0.84 0.71 0.69 0.75 0.67 0.75 0.62 ...  
## $ g10\_precinct\_turnout : num 0.54 0.82 0.34 0.79 0.66 0.6 0.64 0.56 0.6 0.47 ...  
## $ p08\_precinct\_turnout : num 0.17 0.47 0.04 0.24 0.19 0.16 0.16 0.12 0.13 0.06 ...  
## $ p10\_precinct\_turnout : num 0.32 0.62 0.09 0.46 0.37 0.23 0.3 0.24 0.25 0.14 ...  
## $ p12\_precinct\_turnout : num 0.24 0.47 0.06 0.3 0.34 0.17 0.17 0.15 0.13 0.09 ...

## Applying logistic model to the reduced dataset:

voter\_logit <- glm(vh12g ~ ., data=voter.clean.var, family=binomial(link="logit"), na.action = na.exclude)  
  
summary(voter\_logit)

##   
## Call:  
## glm(formula = vh12g ~ ., family = binomial(link = "logit"), data = voter.clean.var,   
## na.action = na.exclude)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.1349 -0.2476 -0.1418 0.1547 3.2263   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) -6.331449 0.272543 -23.231 < 2e-16  
## age 0.018357 0.001238 14.826 < 2e-16  
## partyDemocratic 0.221748 0.086872 2.553 0.010692  
## partyGreen 0.858399 0.668844 1.283 0.199350  
## partyLibertarian 0.683382 0.226179 3.021 0.002516  
## partyNatural Law 0.510473 1.703332 0.300 0.764413  
## partyNon-Partisan 0.243805 0.092776 2.628 0.008592  
## partyOther 0.011852 0.239407 0.050 0.960515  
## partyRepublican 0.261429 0.087107 3.001 0.002689  
## ethnicityAsian -0.413521 0.187164 -2.209 0.027146  
## ethnicityEuropean 0.106696 0.149456 0.714 0.475291  
## ethnicityHispanic -0.180874 0.154934 -1.167 0.243041  
## ethnicitynan 0.021961 0.156577 0.140 0.888456  
## ethnicityOther -0.099897 0.187198 -0.534 0.593586  
## income125k-200k 0.073842 0.075474 0.978 0.327892  
## income200k+ 0.202285 0.094418 2.142 0.032158  
## income35k-75k -0.127228 0.061120 -2.082 0.037377  
## income75k-125k 0.002578 0.063626 0.041 0.967679  
## incomeUnknown -0.046892 0.068872 -0.681 0.495962  
## cd -0.057839 0.017369 -3.330 0.000868  
## dmaLOS ANGELES DMA (EST.) 0.932651 0.604024 1.544 0.122573  
## dmaRENO DMA (EST.) 0.049044 0.045437 1.079 0.280415  
## dmaSALT LAKE CITY DMA (EST.) 0.415953 0.116676 3.565 0.000364  
## vh12p1 3.215800 0.075451 42.621 < 2e-16  
## vh10g1 2.504833 0.052649 47.576 < 2e-16  
## vh10p1 0.471741 0.044363 10.634 < 2e-16  
## vh08g1 0.972868 0.040677 23.917 < 2e-16  
## vh08p1 0.157444 0.043476 3.621 0.000293  
## vh06g1 0.388325 0.059949 6.478 9.32e-11  
## vh06p1 -0.156566 0.052334 -2.992 0.002775  
## vh04g1 0.338950 0.046832 7.238 4.57e-13  
## vh04p1 0.094406 0.042936 2.199 0.027895  
## vh02g1 0.004996 0.060399 0.083 0.934082  
## vh02p1 -0.115449 0.050542 -2.284 0.022357  
## vh00g1 0.275681 0.038109 7.234 4.69e-13  
## vh00p1 0.044558 0.046789 0.952 0.340939  
## net\_worth$10000-24999 -0.093364 0.181199 -0.515 0.606376  
## net\_worth$100000-249999 0.149575 0.136228 1.098 0.272213  
## net\_worth$25000-49999 -0.155044 0.155496 -0.997 0.318719  
## net\_worth$250000-499999 -0.007364 0.138936 -0.053 0.957731  
## net\_worth$499999+ -0.114499 0.147636 -0.776 0.438015  
## net\_worth$5000-9999 -0.125648 0.204992 -0.613 0.539915  
## net\_worth$50000-99999 0.159310 0.160330 0.994 0.320397  
## net\_worthnan -0.003261 0.132627 -0.025 0.980384  
## petowner\_dogYes 0.306752 0.052016 5.897 3.70e-09  
## intrst\_nascar\_in\_hhYes -0.013902 0.068172 -0.204 0.838418  
## intrst\_musical\_instruments\_in\_hhYes 0.298610 0.137570 2.171 0.029962  
## donates\_to\_liberal\_causesYes -0.144045 0.345257 -0.417 0.676524  
## donates\_to\_conservative\_causesYes -0.260046 0.266200 -0.977 0.328627  
## home\_owner\_or\_renterLikely Renter -0.101717 0.165443 -0.615 0.538674  
## home\_owner\_or\_renternan -0.037280 0.052755 -0.707 0.479776  
## g08\_precinct\_turnout 1.165386 0.564643 2.064 0.039024  
## g10\_precinct\_turnout 0.295263 0.651750 0.453 0.650527  
## p08\_precinct\_turnout 0.832841 0.566932 1.469 0.141824  
## p10\_precinct\_turnout -0.492404 0.608791 -0.809 0.418617  
## p12\_precinct\_turnout -0.980158 0.471185 -2.080 0.037507  
##   
## (Intercept) \*\*\*  
## age \*\*\*  
## partyDemocratic \*   
## partyGreen   
## partyLibertarian \*\*   
## partyNatural Law   
## partyNon-Partisan \*\*   
## partyOther   
## partyRepublican \*\*   
## ethnicityAsian \*   
## ethnicityEuropean   
## ethnicityHispanic   
## ethnicitynan   
## ethnicityOther   
## income125k-200k   
## income200k+ \*   
## income35k-75k \*   
## income75k-125k   
## incomeUnknown   
## cd \*\*\*  
## dmaLOS ANGELES DMA (EST.)   
## dmaRENO DMA (EST.)   
## dmaSALT LAKE CITY DMA (EST.) \*\*\*  
## vh12p1 \*\*\*  
## vh10g1 \*\*\*  
## vh10p1 \*\*\*  
## vh08g1 \*\*\*  
## vh08p1 \*\*\*  
## vh06g1 \*\*\*  
## vh06p1 \*\*   
## vh04g1 \*\*\*  
## vh04p1 \*   
## vh02g1   
## vh02p1 \*   
## vh00g1 \*\*\*  
## vh00p1   
## net\_worth$10000-24999   
## net\_worth$100000-249999   
## net\_worth$25000-49999   
## net\_worth$250000-499999   
## net\_worth$499999+   
## net\_worth$5000-9999   
## net\_worth$50000-99999   
## net\_worthnan   
## petowner\_dogYes \*\*\*  
## intrst\_nascar\_in\_hhYes   
## intrst\_musical\_instruments\_in\_hhYes \*   
## donates\_to\_liberal\_causesYes   
## donates\_to\_conservative\_causesYes   
## home\_owner\_or\_renterLikely Renter   
## home\_owner\_or\_renternan   
## g08\_precinct\_turnout \*   
## g10\_precinct\_turnout   
## p08\_precinct\_turnout   
## p10\_precinct\_turnout   
## p12\_precinct\_turnout \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 57682 on 49977 degrees of freedom  
## Residual deviance: 25296 on 49922 degrees of freedom  
## (22 observations deleted due to missingness)  
## AIC: 25408  
##   
## Number of Fisher Scoring iterations: 7

**The AIC has decreased to 25408 from 25414 but it is still not a very bad result.**

## Let us go ahead and make predictions and cross validate for our reduced model:

attach(voter.clean.var)   
set.seed(1)  
train <- sample(1:nrow(voter.clean.var), 0.7\*nrow(voter.clean.var))  
test <- seq(1:nrow(voter.clean.var))[-train]  
  
voter.logit.train.red <- glm(vh12g ~ ., family=binomial(link="logit"), data=voter.clean.var[train,])  
  
voter.logit.test.red=predict(voter.logit.train.red, voter.clean.var, type="response")[test]  
  
voter.logit.test.red[1:10]

## 3 4 8 13 14 20   
## 0.011033314 0.016209313 0.213912616 0.187410173 0.173346217 0.944566160   
## 23 28 31 33   
## 0.565271249 0.424188395 0.026580405 0.008427496

voter.pred.test.red = ifelse(voter.logit.test.red>0.5, 1,0)  
voter.pred.test.red[1:10] # List first 10

## 3 4 8 13 14 20 23 28 31 33   
## 0 0 0 0 0 1 1 0 0 0

conf.mat <- table("Predicted"=voter.pred.test.red, "Actual"=vh12g[test])   
  
colnames(conf.mat) <- c("No", "Yes")  
rownames(conf.mat) <- c("No", "Yes")  
  
#Final Confusion Matrix  
conf.mat

## Actual  
## Predicted No Yes  
## No 10209 931  
## Yes 881 2973

TruN <- conf.mat[1,1] # True negatives  
TruP <- conf.mat[2,2] # True positives  
FalN <- conf.mat[1,2] # False negatives  
FalP <- conf.mat[2,1] # False positives  
TotN <- conf.mat[1,1] + conf.mat[2,1] # Total negatives  
TotP <- conf.mat[1,2] + conf.mat[2,2] # Total positives  
Tot <- TotN+TotP # Total  
  
Accuracy.Rate <- (TruN + TruP) / Tot  
  
Accuracy.Rate

## [1] 0.8791517

Error.Rate <- (FalN + FalP) / Tot  
Error.Rate

## [1] 0.1208483

Sensitivity <- TruP / TotP   
Sensitivity

## [1] 0.7615266

Specificity <- TruN / TotN   
Specificity

## [1] 0.9205591

FalseP.Rate <- 1 - Specificity  
FalseP.Rate

## [1] 0.07944094

logit.rates.50.red <- c(Accuracy.Rate, Error.Rate, Sensitivity, Specificity, FalseP.Rate)  
  
names(logit.rates.50.red) <- c("Accuracy Rate", "Error Rate", "Sensitivity", "Specificity", "False Positives")  
  
print(logit.rates.50.red, digits=2)

## Accuracy Rate Error Rate Sensitivity Specificity   
## 0.879 0.121 0.762 0.921   
## False Positives   
## 0.079

logit.fit.stats.compare <- rbind(logit.rates.50, logit.rates.40)  
print(logit.fit.stats.compare, digits=2)

## Accuracy Rate Error Rate Sensitivity Specificity  
## logit.rates.50 0.88 0.12 0.76 0.92  
## logit.rates.40 0.88 0.12 0.84 0.89  
## False Positives  
## logit.rates.50 0.08  
## logit.rates.40 0.11

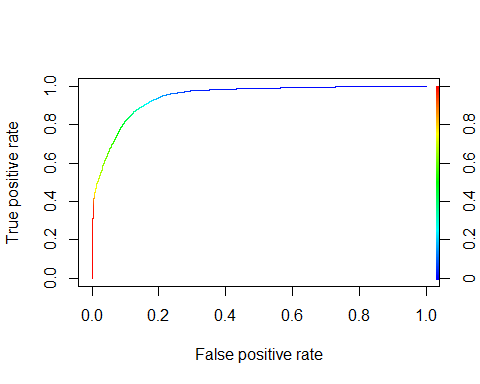
logit.fit.stats.compare.red <- rbind(logit.fit.stats.compare, logit.rates.50.red)  
print(logit.fit.stats.compare.red, digits=3)

## Accuracy Rate Error Rate Sensitivity Specificity  
## logit.rates.50 0.879 0.121 0.762 0.920  
## logit.rates.40 0.877 0.123 0.845 0.888  
## logit.rates.50.red 0.879 0.121 0.762 0.921  
## False Positives  
## logit.rates.50 0.0796  
## logit.rates.40 0.1123  
## logit.rates.50.red 0.0794

**As we can see that our reduced model is even giving us round about the same output even after removing most insignificant variables and the ones with a lots of nas. Thus, this should not cause any problems.**

## Let’s go ahead and use ROC Curve to check the accuracy of our reduced test:

pred <- prediction(voter.logit.test.red, vh12g[test])   
  
perf <- performance(pred,"tpr","fpr")  
plot(perf, colorize=T)



auc <- performance(pred,"auc")  
  
c(auc@y.name[[1]], auc@y.values[[1]])

## [1] "Area under the ROC curve" "0.943845206507121"

**The result from ROC is pretty much the same and still pretty good!**

## Confusion matrix with threshold = 0.40 for our reduced model.

attach(voter.clean.var)   
set.seed(1)  
  
train <- sample(1:nrow(voter.clean.var), 0.7\*nrow(voter.clean.var))  
test <- seq(1:nrow(voter.clean.var))[-train]  
  
voter.logit.train.red <- glm(vh12g ~ ., family=binomial(link="logit"), data=voter.clean.var[train,])  
  
voter.logit.test.red=predict(voter.logit.train.red, voter.clean.var, type="response")[test]  
  
voter.logit.test.red[1:10]

## 3 4 8 13 14 20   
## 0.011033314 0.016209313 0.213912616 0.187410173 0.173346217 0.944566160   
## 23 28 31 33   
## 0.565271249 0.424188395 0.026580405 0.008427496

voter.pred.test.red = ifelse(voter.logit.test.red>0.40, 1,0)  
voter.pred.test.red[1:10] # List first 10

## 3 4 8 13 14 20 23 28 31 33   
## 0 0 0 0 0 1 1 1 0 0

conf.mat <- table("Predicted"=voter.pred.test.red, "Actual"=vh12g[test])   
  
colnames(conf.mat) <- c("No", "Yes")  
rownames(conf.mat) <- c("No", "Yes")  
  
conf.mat

## Actual  
## Predicted No Yes  
## No 9838 609  
## Yes 1252 3295

TruN <- conf.mat[1,1] # True negatives  
TruP <- conf.mat[2,2] # True positives  
FalN <- conf.mat[1,2] # False negatives  
FalP <- conf.mat[2,1] # False positives  
TotN <- conf.mat[1,1] + conf.mat[2,1] # Total negatives  
TotP <- conf.mat[1,2] + conf.mat[2,2] # Total positives  
Tot <- TotN+TotP # Total  
  
Accuracy.Rate <- (TruN + TruP) / Tot  
Accuracy.Rate

## [1] 0.8758837

Error.Rate <- (FalN + FalP) / Tot  
Error.Rate

## [1] 0.1241163

Sensitivity <- TruP / TotP   
Sensitivity

## [1] 0.8440061

Specificity <- TruN / TotN   
Specificity

## [1] 0.8871055

FalseP.Rate <- 1 - Specificity  
FalseP.Rate

## [1] 0.1128945

logit.rates.40.red <- c(Accuracy.Rate, Error.Rate, Sensitivity, Specificity, FalseP.Rate)  
  
names(logit.rates.40.red) <- c("Accuracy Rate", "Error Rate", "Sensitivity", "Specificity", "False Positives")  
  
print(logit.rates.40.red, digits=2)

## Accuracy Rate Error Rate Sensitivity Specificity   
## 0.88 0.12 0.84 0.89   
## False Positives   
## 0.11

logit.fit.stats.compare.red1 <- rbind(logit.fit.stats.compare.red, logit.rates.40.red)  
print(logit.fit.stats.compare.red1, digits=3)

## Accuracy Rate Error Rate Sensitivity Specificity  
## logit.rates.50 0.879 0.121 0.762 0.920  
## logit.rates.40 0.877 0.123 0.845 0.888  
## logit.rates.50.red 0.879 0.121 0.762 0.921  
## logit.rates.40.red 0.876 0.124 0.844 0.887  
## False Positives  
## logit.rates.50 0.0796  
## logit.rates.40 0.1123  
## logit.rates.50.red 0.0794  
## logit.rates.40.red 0.1129

**We can see that all the models perform well and are balanced and we can easily go ahead with our reduced model.**

## Let’s go ahead and now try to apply variable selection method and I am choosing Stepwise:

voter.logit.cor.null <- glm(vh12g ~ 1, data=voter.clean.var, family=binomial(link="logit"))  
  
summary(voter.logit.cor.null)

##   
## Call:  
## glm(formula = vh12g ~ 1, family = binomial(link = "logit"), data = voter.clean.var)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.7828 -0.7828 -0.7828 1.6323 1.6323   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.02580 0.01015 -101.1 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 57710 on 49999 degrees of freedom  
## Residual deviance: 57710 on 49999 degrees of freedom  
## AIC: 57712  
##   
## Number of Fisher Scoring iterations: 4

voter.logit.cor.full <- glm(vh12g ~ ., data=voter.clean.var, family=binomial(link="logit"))  
  
summary(voter.logit.cor.full)

##   
## Call:  
## glm(formula = vh12g ~ ., family = binomial(link = "logit"), data = voter.clean.var)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.1349 -0.2476 -0.1418 0.1547 3.2263   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) -6.331449 0.272543 -23.231 < 2e-16  
## age 0.018357 0.001238 14.826 < 2e-16  
## partyDemocratic 0.221748 0.086872 2.553 0.010692  
## partyGreen 0.858399 0.668844 1.283 0.199350  
## partyLibertarian 0.683382 0.226179 3.021 0.002516  
## partyNatural Law 0.510473 1.703332 0.300 0.764413  
## partyNon-Partisan 0.243805 0.092776 2.628 0.008592  
## partyOther 0.011852 0.239407 0.050 0.960515  
## partyRepublican 0.261429 0.087107 3.001 0.002689  
## ethnicityAsian -0.413521 0.187164 -2.209 0.027146  
## ethnicityEuropean 0.106696 0.149456 0.714 0.475291  
## ethnicityHispanic -0.180874 0.154934 -1.167 0.243041  
## ethnicitynan 0.021961 0.156577 0.140 0.888456  
## ethnicityOther -0.099897 0.187198 -0.534 0.593586  
## income125k-200k 0.073842 0.075474 0.978 0.327892  
## income200k+ 0.202285 0.094418 2.142 0.032158  
## income35k-75k -0.127228 0.061120 -2.082 0.037377  
## income75k-125k 0.002578 0.063626 0.041 0.967679  
## incomeUnknown -0.046892 0.068872 -0.681 0.495962  
## cd -0.057839 0.017369 -3.330 0.000868  
## dmaLOS ANGELES DMA (EST.) 0.932651 0.604024 1.544 0.122573  
## dmaRENO DMA (EST.) 0.049044 0.045437 1.079 0.280415  
## dmaSALT LAKE CITY DMA (EST.) 0.415953 0.116676 3.565 0.000364  
## vh12p1 3.215800 0.075451 42.621 < 2e-16  
## vh10g1 2.504833 0.052649 47.576 < 2e-16  
## vh10p1 0.471741 0.044363 10.634 < 2e-16  
## vh08g1 0.972868 0.040677 23.917 < 2e-16  
## vh08p1 0.157444 0.043476 3.621 0.000293  
## vh06g1 0.388325 0.059949 6.478 9.32e-11  
## vh06p1 -0.156566 0.052334 -2.992 0.002775  
## vh04g1 0.338950 0.046832 7.238 4.57e-13  
## vh04p1 0.094406 0.042936 2.199 0.027895  
## vh02g1 0.004996 0.060399 0.083 0.934082  
## vh02p1 -0.115449 0.050542 -2.284 0.022357  
## vh00g1 0.275681 0.038109 7.234 4.69e-13  
## vh00p1 0.044558 0.046789 0.952 0.340939  
## net\_worth$10000-24999 -0.093364 0.181199 -0.515 0.606376  
## net\_worth$100000-249999 0.149575 0.136228 1.098 0.272213  
## net\_worth$25000-49999 -0.155044 0.155496 -0.997 0.318719  
## net\_worth$250000-499999 -0.007364 0.138936 -0.053 0.957731  
## net\_worth$499999+ -0.114499 0.147636 -0.776 0.438015  
## net\_worth$5000-9999 -0.125648 0.204992 -0.613 0.539915  
## net\_worth$50000-99999 0.159310 0.160330 0.994 0.320397  
## net\_worthnan -0.003261 0.132627 -0.025 0.980384  
## petowner\_dogYes 0.306752 0.052016 5.897 3.70e-09  
## intrst\_nascar\_in\_hhYes -0.013902 0.068172 -0.204 0.838418  
## intrst\_musical\_instruments\_in\_hhYes 0.298610 0.137570 2.171 0.029962  
## donates\_to\_liberal\_causesYes -0.144045 0.345257 -0.417 0.676524  
## donates\_to\_conservative\_causesYes -0.260046 0.266200 -0.977 0.328627  
## home\_owner\_or\_renterLikely Renter -0.101717 0.165443 -0.615 0.538674  
## home\_owner\_or\_renternan -0.037280 0.052755 -0.707 0.479776  
## g08\_precinct\_turnout 1.165386 0.564643 2.064 0.039024  
## g10\_precinct\_turnout 0.295263 0.651750 0.453 0.650527  
## p08\_precinct\_turnout 0.832841 0.566932 1.469 0.141824  
## p10\_precinct\_turnout -0.492404 0.608791 -0.809 0.418617  
## p12\_precinct\_turnout -0.980158 0.471185 -2.080 0.037507  
##   
## (Intercept) \*\*\*  
## age \*\*\*  
## partyDemocratic \*   
## partyGreen   
## partyLibertarian \*\*   
## partyNatural Law   
## partyNon-Partisan \*\*   
## partyOther   
## partyRepublican \*\*   
## ethnicityAsian \*   
## ethnicityEuropean   
## ethnicityHispanic   
## ethnicitynan   
## ethnicityOther   
## income125k-200k   
## income200k+ \*   
## income35k-75k \*   
## income75k-125k   
## incomeUnknown   
## cd \*\*\*  
## dmaLOS ANGELES DMA (EST.)   
## dmaRENO DMA (EST.)   
## dmaSALT LAKE CITY DMA (EST.) \*\*\*  
## vh12p1 \*\*\*  
## vh10g1 \*\*\*  
## vh10p1 \*\*\*  
## vh08g1 \*\*\*  
## vh08p1 \*\*\*  
## vh06g1 \*\*\*  
## vh06p1 \*\*   
## vh04g1 \*\*\*  
## vh04p1 \*   
## vh02g1   
## vh02p1 \*   
## vh00g1 \*\*\*  
## vh00p1   
## net\_worth$10000-24999   
## net\_worth$100000-249999   
## net\_worth$25000-49999   
## net\_worth$250000-499999   
## net\_worth$499999+   
## net\_worth$5000-9999   
## net\_worth$50000-99999   
## net\_worthnan   
## petowner\_dogYes \*\*\*  
## intrst\_nascar\_in\_hhYes   
## intrst\_musical\_instruments\_in\_hhYes \*   
## donates\_to\_liberal\_causesYes   
## donates\_to\_conservative\_causesYes   
## home\_owner\_or\_renterLikely Renter   
## home\_owner\_or\_renternan   
## g08\_precinct\_turnout \*   
## g10\_precinct\_turnout   
## p08\_precinct\_turnout   
## p10\_precinct\_turnout   
## p12\_precinct\_turnout \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 57682 on 49977 degrees of freedom  
## Residual deviance: 25296 on 49922 degrees of freedom  
## (22 observations deleted due to missingness)  
## AIC: 25408  
##   
## Number of Fisher Scoring iterations: 7

voter.step.backward <- step(voter.logit.cor.full, scope=list(lower=voter.logit.cor.null, upper=voter.logit.cor.full), direction="both", test="F")

## Start: AIC=25408.04  
## vh12g ~ age + party + ethnicity + income + cd + dma + vh12p +   
## vh10g + vh10p + vh08g + vh08p + vh06g + vh06p + vh04g + vh04p +   
## vh02g + vh02p + vh00g + vh00p + net\_worth + petowner\_dog +   
## intrst\_nascar\_in\_hh + intrst\_musical\_instruments\_in\_hh +   
## donates\_to\_liberal\_causes + donates\_to\_conservative\_causes +   
## home\_owner\_or\_renter + g08\_precinct\_turnout + g10\_precinct\_turnout +   
## p08\_precinct\_turnout + p10\_precinct\_turnout + p12\_precinct\_turnout  
##   
## Df Deviance AIC F value Pr(>F)  
## - home\_owner\_or\_renter 2 25297 25405 0.7353 0.479371  
## - vh02g 1 25296 25406 0.0135 0.907489  
## - intrst\_nascar\_in\_hh 1 25296 25406 0.0821 0.774519  
## - donates\_to\_liberal\_causes 1 25296 25406 0.3421 0.558644  
## - g10\_precinct\_turnout 1 25296 25406 0.4051 0.524476  
## - p10\_precinct\_turnout 1 25297 25407 1.2931 0.255487  
## - vh00p 1 25297 25407 1.7888 0.181078  
## - donates\_to\_conservative\_causes 1 25297 25407 1.8773 0.170648  
## <none> 25296 25408   
## - p08\_precinct\_turnout 1 25298 25408 4.2844 0.038469  
## - party 7 25312 25410 4.4523 5.813e-05  
## - g08\_precinct\_turnout 1 25300 25410 8.4410 0.003670  
## - p12\_precinct\_turnout 1 25300 25410 8.5904 0.003381  
## - intrst\_musical\_instruments\_in\_hh 1 25301 25411 9.3201 0.002268  
## - vh04p 1 25301 25411 9.5307 0.002022  
## - vh02p 1 25301 25411 10.3158 0.001320  
## - vh06p 1 25305 25415 17.6987 2.593e-05  
## - dma 3 25311 25417 9.8587 1.701e-06  
## - cd 1 25307 25417 21.8675 2.929e-06  
## - net\_worth 8 25323 25419 6.6138 1.135e-08  
## - vh08p 1 25309 25419 25.8565 3.691e-07  
## - income 5 25319 25421 9.0798 1.214e-08  
## - petowner\_dog 1 25331 25441 68.8016 < 2.2e-16  
## - ethnicity 5 25347 25449 19.9207 < 2.2e-16  
## - vh06g 1 25339 25449 84.2884 < 2.2e-16  
## - vh00g 1 25348 25458 103.0111 < 2.2e-16  
## - vh04g 1 25349 25459 103.9902 < 2.2e-16  
## - vh10p 1 25410 25520 224.3679 < 2.2e-16  
## - age 1 25518 25628 438.2940 < 2.2e-16  
## - vh08g 1 25881 25991 1154.3649 < 2.2e-16  
## - vh10g 1 28276 28386 5880.9596 < 2.2e-16  
## - vh12p 1 28460 28570 6243.7183 < 2.2e-16  
##   
## - home\_owner\_or\_renter   
## - vh02g   
## - intrst\_nascar\_in\_hh   
## - donates\_to\_liberal\_causes   
## - g10\_precinct\_turnout   
## - p10\_precinct\_turnout   
## - vh00p   
## - donates\_to\_conservative\_causes   
## <none>   
## - p08\_precinct\_turnout \*   
## - party \*\*\*  
## - g08\_precinct\_turnout \*\*   
## - p12\_precinct\_turnout \*\*   
## - intrst\_musical\_instruments\_in\_hh \*\*   
## - vh04p \*\*   
## - vh02p \*\*   
## - vh06p \*\*\*  
## - dma \*\*\*  
## - cd \*\*\*  
## - net\_worth \*\*\*  
## - vh08p \*\*\*  
## - income \*\*\*  
## - petowner\_dog \*\*\*  
## - ethnicity \*\*\*  
## - vh06g \*\*\*  
## - vh00g \*\*\*  
## - vh04g \*\*\*  
## - vh10p \*\*\*  
## - age \*\*\*  
## - vh08g \*\*\*  
## - vh10g \*\*\*  
## - vh12p \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Step: AIC=25404.79  
## vh12g ~ age + party + ethnicity + income + cd + dma + vh12p +   
## vh10g + vh10p + vh08g + vh08p + vh06g + vh06p + vh04g + vh04p +   
## vh02g + vh02p + vh00g + vh00p + net\_worth + petowner\_dog +   
## intrst\_nascar\_in\_hh + intrst\_musical\_instruments\_in\_hh +   
## donates\_to\_liberal\_causes + donates\_to\_conservative\_causes +   
## g08\_precinct\_turnout + g10\_precinct\_turnout + p08\_precinct\_turnout +   
## p10\_precinct\_turnout + p12\_precinct\_turnout  
##   
## Df Deviance AIC F value Pr(>F)  
## - vh02g 1 25297 25403 0.0175 0.894684  
## - intrst\_nascar\_in\_hh 1 25297 25403 0.0858 0.769573  
## - donates\_to\_liberal\_causes 1 25297 25403 0.3661 0.545120  
## - g10\_precinct\_turnout 1 25297 25403 0.4490 0.502802  
## - p10\_precinct\_turnout 1 25298 25404 1.3184 0.250893  
## - vh00p 1 25298 25404 1.7927 0.180608  
## - donates\_to\_conservative\_causes 1 25298 25404 1.9409 0.163582  
## <none> 25297 25405   
## - p08\_precinct\_turnout 1 25299 25405 4.3133 0.037820  
## - party 7 25313 25407 4.4540 5.785e-05  
## - g08\_precinct\_turnout 1 25301 25407 8.4029 0.003748  
## - p12\_precinct\_turnout 1 25301 25407 8.7982 0.003017  
## - intrst\_musical\_instruments\_in\_hh 1 25302 25408 9.3227 0.002265  
## - vh04p 1 25302 25408 9.5804 0.001968  
## - vh02p 1 25302 25408 10.2303 0.001382  
## + home\_owner\_or\_renter 2 25296 25408 0.7353 0.479371  
## - vh06p 1 25306 25412 17.7254 2.556e-05  
## - dma 3 25312 25414 9.7742 1.924e-06  
## - cd 1 25308 25414 21.7493 3.115e-06  
## - vh08p 1 25310 25416 25.9354 3.543e-07  
## - income 5 25321 25419 9.3661 6.211e-09  
## - net\_worth 8 25329 25421 7.9228 1.024e-10  
## - petowner\_dog 1 25332 25438 69.0498 < 2.2e-16  
## - ethnicity 5 25347 25445 19.8214 < 2.2e-16  
## - vh06g 1 25340 25446 84.3540 < 2.2e-16  
## - vh00g 1 25349 25455 103.1899 < 2.2e-16  
## - vh04g 1 25350 25456 104.8368 < 2.2e-16  
## - vh10p 1 25410 25516 224.1900 < 2.2e-16  
## - age 1 25519 25625 438.3605 < 2.2e-16  
## - vh08g 1 25882 25988 1154.9139 < 2.2e-16  
## - vh10g 1 28277 28383 5882.1229 < 2.2e-16  
## - vh12p 1 28460 28566 6243.3999 < 2.2e-16  
##   
## - vh02g   
## - intrst\_nascar\_in\_hh   
## - donates\_to\_liberal\_causes   
## - g10\_precinct\_turnout   
## - p10\_precinct\_turnout   
## - vh00p   
## - donates\_to\_conservative\_causes   
## <none>   
## - p08\_precinct\_turnout \*   
## - party \*\*\*  
## - g08\_precinct\_turnout \*\*   
## - p12\_precinct\_turnout \*\*   
## - intrst\_musical\_instruments\_in\_hh \*\*   
## - vh04p \*\*   
## - vh02p \*\*   
## + home\_owner\_or\_renter   
## - vh06p \*\*\*  
## - dma \*\*\*  
## - cd \*\*\*  
## - vh08p \*\*\*  
## - income \*\*\*  
## - net\_worth \*\*\*  
## - petowner\_dog \*\*\*  
## - ethnicity \*\*\*  
## - vh06g \*\*\*  
## - vh00g \*\*\*  
## - vh04g \*\*\*  
## - vh10p \*\*\*  
## - age \*\*\*  
## - vh08g \*\*\*  
## - vh10g \*\*\*  
## - vh12p \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Step: AIC=25402.8  
## vh12g ~ age + party + ethnicity + income + cd + dma + vh12p +   
## vh10g + vh10p + vh08g + vh08p + vh06g + vh06p + vh04g + vh04p +   
## vh02p + vh00g + vh00p + net\_worth + petowner\_dog + intrst\_nascar\_in\_hh +   
## intrst\_musical\_instruments\_in\_hh + donates\_to\_liberal\_causes +   
## donates\_to\_conservative\_causes + g08\_precinct\_turnout + g10\_precinct\_turnout +   
## p08\_precinct\_turnout + p10\_precinct\_turnout + p12\_precinct\_turnout  
##   
## Df Deviance AIC F value Pr(>F)  
## - intrst\_nascar\_in\_hh 1 25297 25401 0.0864 0.768741  
## - donates\_to\_liberal\_causes 1 25297 25401 0.3667 0.544821  
## - g10\_precinct\_turnout 1 25297 25401 0.4501 0.502299  
## - p10\_precinct\_turnout 1 25298 25402 1.3213 0.250362  
## - vh00p 1 25298 25402 1.7980 0.179958  
## - donates\_to\_conservative\_causes 1 25298 25402 1.9390 0.163784  
## <none> 25297 25403   
## - p08\_precinct\_turnout 1 25299 25403 4.3120 0.037850  
## - party 7 25313 25405 4.4519 5.821e-05  
## + vh02g 1 25297 25405 0.0175 0.894684  
## - g08\_precinct\_turnout 1 25301 25405 8.4080 0.003737  
## - p12\_precinct\_turnout 1 25301 25405 8.7999 0.003014  
## - intrst\_musical\_instruments\_in\_hh 1 25302 25406 9.3210 0.002267  
## - vh04p 1 25302 25406 9.5753 0.001973  
## - vh02p 1 25302 25406 10.2145 0.001394  
## + home\_owner\_or\_renter 2 25296 25406 0.7373 0.478401  
## - vh06p 1 25306 25410 17.7588 2.512e-05  
## - dma 3 25312 25412 9.7895 1.881e-06  
## - cd 1 25308 25412 21.7322 3.143e-06  
## - vh08p 1 25310 25414 25.9198 3.572e-07  
## - income 5 25321 25417 9.3685 6.176e-09  
## - net\_worth 8 25329 25419 7.9267 1.009e-10  
## - petowner\_dog 1 25332 25436 69.0535 < 2.2e-16  
## - ethnicity 5 25347 25443 19.8257 < 2.2e-16  
## - vh06g 1 25341 25445 87.6849 < 2.2e-16  
## - vh00g 1 25351 25455 106.2288 < 2.2e-16  
## - vh04g 1 25352 25456 109.5833 < 2.2e-16  
## - vh10p 1 25410 25514 224.1787 < 2.2e-16  
## - age 1 25519 25623 438.5968 < 2.2e-16  
## - vh08g 1 25882 25986 1155.5911 < 2.2e-16  
## - vh10g 1 28282 28386 5890.7031 < 2.2e-16  
## - vh12p 1 28461 28565 6243.9208 < 2.2e-16  
##   
## - intrst\_nascar\_in\_hh   
## - donates\_to\_liberal\_causes   
## - g10\_precinct\_turnout   
## - p10\_precinct\_turnout   
## - vh00p   
## - donates\_to\_conservative\_causes   
## <none>   
## - p08\_precinct\_turnout \*   
## - party \*\*\*  
## + vh02g   
## - g08\_precinct\_turnout \*\*   
## - p12\_precinct\_turnout \*\*   
## - intrst\_musical\_instruments\_in\_hh \*\*   
## - vh04p \*\*   
## - vh02p \*\*   
## + home\_owner\_or\_renter   
## - vh06p \*\*\*  
## - dma \*\*\*  
## - cd \*\*\*  
## - vh08p \*\*\*  
## - income \*\*\*  
## - net\_worth \*\*\*  
## - petowner\_dog \*\*\*  
## - ethnicity \*\*\*  
## - vh06g \*\*\*  
## - vh00g \*\*\*  
## - vh04g \*\*\*  
## - vh10p \*\*\*  
## - age \*\*\*  
## - vh08g \*\*\*  
## - vh10g \*\*\*  
## - vh12p \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Step: AIC=25400.84  
## vh12g ~ age + party + ethnicity + income + cd + dma + vh12p +   
## vh10g + vh10p + vh08g + vh08p + vh06g + vh06p + vh04g + vh04p +   
## vh02p + vh00g + vh00p + net\_worth + petowner\_dog + intrst\_musical\_instruments\_in\_hh +   
## donates\_to\_liberal\_causes + donates\_to\_conservative\_causes +   
## g08\_precinct\_turnout + g10\_precinct\_turnout + p08\_precinct\_turnout +   
## p10\_precinct\_turnout + p12\_precinct\_turnout  
##   
## Df Deviance AIC F value Pr(>F)  
## - donates\_to\_liberal\_causes 1 25297 25399 0.3769 0.539261  
## - g10\_precinct\_turnout 1 25297 25399 0.4579 0.498623  
## - p10\_precinct\_turnout 1 25298 25400 1.3348 0.247963  
## - vh00p 1 25298 25400 1.8001 0.179704  
## - donates\_to\_conservative\_causes 1 25298 25400 1.9656 0.160925  
## <none> 25297 25401   
## - p08\_precinct\_turnout 1 25299 25401 4.3145 0.037794  
## - party 7 25313 25403 4.4546 5.774e-05  
## + intrst\_nascar\_in\_hh 1 25297 25403 0.0864 0.768741  
## + vh02g 1 25297 25403 0.0182 0.892796  
## - g08\_precinct\_turnout 1 25301 25403 8.3899 0.003775  
## - p12\_precinct\_turnout 1 25301 25403 8.7871 0.003035  
## - intrst\_musical\_instruments\_in\_hh 1 25302 25404 9.3128 0.002277  
## - vh04p 1 25302 25404 9.5815 0.001967  
## - vh02p 1 25302 25404 10.2177 0.001392  
## + home\_owner\_or\_renter 2 25296 25404 0.7392 0.477479  
## - vh06p 1 25306 25408 17.7482 2.526e-05  
## - dma 3 25312 25410 9.7892 1.882e-06  
## - cd 1 25308 25410 21.7544 3.107e-06  
## - vh08p 1 25310 25412 25.8775 3.651e-07  
## - income 5 25321 25415 9.3781 6.040e-09  
## - net\_worth 8 25329 25417 7.9167 1.047e-10  
## - petowner\_dog 1 25334 25436 72.5007 < 2.2e-16  
## - ethnicity 5 25347 25441 19.8108 < 2.2e-16  
## - vh06g 1 25341 25443 87.7046 < 2.2e-16  
## - vh00g 1 25351 25453 106.2886 < 2.2e-16  
## - vh04g 1 25352 25454 109.6012 < 2.2e-16  
## - vh10p 1 25411 25513 224.3222 < 2.2e-16  
## - age 1 25519 25621 438.5203 < 2.2e-16  
## - vh08g 1 25882 25984 1155.5408 < 2.2e-16  
## - vh10g 1 28282 28384 5890.7244 < 2.2e-16  
## - vh12p 1 28461 28563 6244.1533 < 2.2e-16  
##   
## - donates\_to\_liberal\_causes   
## - g10\_precinct\_turnout   
## - p10\_precinct\_turnout   
## - vh00p   
## - donates\_to\_conservative\_causes   
## <none>   
## - p08\_precinct\_turnout \*   
## - party \*\*\*  
## + intrst\_nascar\_in\_hh   
## + vh02g   
## - g08\_precinct\_turnout \*\*   
## - p12\_precinct\_turnout \*\*   
## - intrst\_musical\_instruments\_in\_hh \*\*   
## - vh04p \*\*   
## - vh02p \*\*   
## + home\_owner\_or\_renter   
## - vh06p \*\*\*  
## - dma \*\*\*  
## - cd \*\*\*  
## - vh08p \*\*\*  
## - income \*\*\*  
## - net\_worth \*\*\*  
## - petowner\_dog \*\*\*  
## - ethnicity \*\*\*  
## - vh06g \*\*\*  
## - vh00g \*\*\*  
## - vh04g \*\*\*  
## - vh10p \*\*\*  
## - age \*\*\*  
## - vh08g \*\*\*  
## - vh10g \*\*\*  
## - vh12p \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Step: AIC=25399.03  
## vh12g ~ age + party + ethnicity + income + cd + dma + vh12p +   
## vh10g + vh10p + vh08g + vh08p + vh06g + vh06p + vh04g + vh04p +   
## vh02p + vh00g + vh00p + net\_worth + petowner\_dog + intrst\_musical\_instruments\_in\_hh +   
## donates\_to\_conservative\_causes + g08\_precinct\_turnout + g10\_precinct\_turnout +   
## p08\_precinct\_turnout + p10\_precinct\_turnout + p12\_precinct\_turnout  
##   
## Df Deviance AIC F value Pr(>F)  
## - g10\_precinct\_turnout 1 25297 25397 0.4665 0.494599  
## - p10\_precinct\_turnout 1 25298 25398 1.3446 0.246226  
## - vh00p 1 25298 25398 1.7835 0.181722  
## - donates\_to\_conservative\_causes 1 25298 25398 2.0695 0.150278  
## <none> 25297 25399   
## - p08\_precinct\_turnout 1 25299 25399 4.3309 0.037431  
## - party 7 25313 25401 4.4499 5.855e-05  
## + donates\_to\_liberal\_causes 1 25297 25401 0.3769 0.539261  
## + intrst\_nascar\_in\_hh 1 25297 25401 0.0967 0.755854  
## + vh02g 1 25297 25401 0.0188 0.891078  
## - g08\_precinct\_turnout 1 25301 25401 8.3341 0.003893  
## - p12\_precinct\_turnout 1 25302 25402 8.7648 0.003072  
## - intrst\_musical\_instruments\_in\_hh 1 25302 25402 9.3141 0.002275  
## - vh04p 1 25302 25402 9.5975 0.001950  
## - vh02p 1 25302 25402 10.2666 0.001355  
## + home\_owner\_or\_renter 2 25296 25402 0.7516 0.471601  
## - vh06p 1 25306 25406 17.7351 2.543e-05  
## - dma 3 25312 25408 9.7646 1.950e-06  
## - cd 1 25308 25408 21.7292 3.148e-06  
## - vh08p 1 25310 25410 25.8029 3.795e-07  
## - income 5 25321 25413 9.3802 6.009e-09  
## - net\_worth 8 25329 25415 7.9112 1.068e-10  
## - petowner\_dog 1 25334 25434 72.1692 < 2.2e-16  
## - ethnicity 5 25347 25439 19.7829 < 2.2e-16  
## - vh06g 1 25342 25442 87.7121 < 2.2e-16  
## - vh00g 1 25351 25451 106.1563 < 2.2e-16  
## - vh04g 1 25353 25453 109.7387 < 2.2e-16  
## - vh10p 1 25411 25511 224.4316 < 2.2e-16  
## - age 1 25519 25619 438.3106 < 2.2e-16  
## - vh08g 1 25883 25983 1155.5036 < 2.2e-16  
## - vh10g 1 28282 28382 5891.9647 < 2.2e-16  
## - vh12p 1 28461 28561 6244.4917 < 2.2e-16  
##   
## - g10\_precinct\_turnout   
## - p10\_precinct\_turnout   
## - vh00p   
## - donates\_to\_conservative\_causes   
## <none>   
## - p08\_precinct\_turnout \*   
## - party \*\*\*  
## + donates\_to\_liberal\_causes   
## + intrst\_nascar\_in\_hh   
## + vh02g   
## - g08\_precinct\_turnout \*\*   
## - p12\_precinct\_turnout \*\*   
## - intrst\_musical\_instruments\_in\_hh \*\*   
## - vh04p \*\*   
## - vh02p \*\*   
## + home\_owner\_or\_renter   
## - vh06p \*\*\*  
## - dma \*\*\*  
## - cd \*\*\*  
## - vh08p \*\*\*  
## - income \*\*\*  
## - net\_worth \*\*\*  
## - petowner\_dog \*\*\*  
## - ethnicity \*\*\*  
## - vh06g \*\*\*  
## - vh00g \*\*\*  
## - vh04g \*\*\*  
## - vh10p \*\*\*  
## - age \*\*\*  
## - vh08g \*\*\*  
## - vh10g \*\*\*  
## - vh12p \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Step: AIC=25397.27  
## vh12g ~ age + party + ethnicity + income + cd + dma + vh12p +   
## vh10g + vh10p + vh08g + vh08p + vh06g + vh06p + vh04g + vh04p +   
## vh02p + vh00g + vh00p + net\_worth + petowner\_dog + intrst\_musical\_instruments\_in\_hh +   
## donates\_to\_conservative\_causes + g08\_precinct\_turnout + p08\_precinct\_turnout +   
## p10\_precinct\_turnout + p12\_precinct\_turnout  
##   
## Df Deviance AIC F value Pr(>F)  
## - p10\_precinct\_turnout 1 25298 25396 0.8993 0.342984  
## - vh00p 1 25298 25396 1.8134 0.178112  
## - donates\_to\_conservative\_causes 1 25298 25396 2.0773 0.149512  
## - p08\_precinct\_turnout 1 25299 25397 3.9243 0.047599  
## <none> 25297 25397   
## + g10\_precinct\_turnout 1 25297 25399 0.4665 0.494599  
## - party 7 25313 25399 4.4538 5.788e-05  
## + donates\_to\_liberal\_causes 1 25297 25399 0.3856 0.534648  
## + intrst\_nascar\_in\_hh 1 25297 25399 0.1051 0.745846  
## + vh02g 1 25297 25399 0.0199 0.887843  
## - p12\_precinct\_turnout 1 25302 25400 8.3012 0.003964  
## - intrst\_musical\_instruments\_in\_hh 1 25302 25400 9.2886 0.002307  
## - vh04p 1 25302 25400 9.4922 0.002065  
## + home\_owner\_or\_renter 2 25297 25401 0.7745 0.460940  
## - vh02p 1 25303 25401 10.3419 0.001301  
## - vh06p 1 25306 25404 17.6599 2.646e-05  
## - dma 3 25313 25407 10.0904 1.215e-06  
## - cd 1 25309 25407 22.5019 2.105e-06  
## - vh08p 1 25310 25408 25.7593 3.881e-07  
## - income 5 25321 25411 9.4523 5.075e-09  
## - net\_worth 8 25329 25413 7.8694 1.243e-10  
## - g08\_precinct\_turnout 1 25320 25418 44.8818 2.115e-11  
## - petowner\_dog 1 25334 25432 72.0612 < 2.2e-16  
## - ethnicity 5 25348 25438 19.8094 < 2.2e-16  
## - vh06g 1 25342 25440 87.4634 < 2.2e-16  
## - vh00g 1 25351 25449 106.6856 < 2.2e-16  
## - vh04g 1 25353 25451 110.2944 < 2.2e-16  
## - vh10p 1 25411 25509 224.2158 < 2.2e-16  
## - age 1 25520 25618 438.8965 < 2.2e-16  
## - vh08g 1 25883 25981 1156.0313 < 2.2e-16  
## - vh10g 1 28283 28381 5892.5886 < 2.2e-16  
## - vh12p 1 28461 28559 6244.1940 < 2.2e-16  
##   
## - p10\_precinct\_turnout   
## - vh00p   
## - donates\_to\_conservative\_causes   
## - p08\_precinct\_turnout \*   
## <none>   
## + g10\_precinct\_turnout   
## - party \*\*\*  
## + donates\_to\_liberal\_causes   
## + intrst\_nascar\_in\_hh   
## + vh02g   
## - p12\_precinct\_turnout \*\*   
## - intrst\_musical\_instruments\_in\_hh \*\*   
## - vh04p \*\*   
## + home\_owner\_or\_renter   
## - vh02p \*\*   
## - vh06p \*\*\*  
## - dma \*\*\*  
## - cd \*\*\*  
## - vh08p \*\*\*  
## - income \*\*\*  
## - net\_worth \*\*\*  
## - g08\_precinct\_turnout \*\*\*  
## - petowner\_dog \*\*\*  
## - ethnicity \*\*\*  
## - vh06g \*\*\*  
## - vh00g \*\*\*  
## - vh04g \*\*\*  
## - vh10p \*\*\*  
## - age \*\*\*  
## - vh08g \*\*\*  
## - vh10g \*\*\*  
## - vh12p \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Step: AIC=25395.72  
## vh12g ~ age + party + ethnicity + income + cd + dma + vh12p +   
## vh10g + vh10p + vh08g + vh08p + vh06g + vh06p + vh04g + vh04p +   
## vh02p + vh00g + vh00p + net\_worth + petowner\_dog + intrst\_musical\_instruments\_in\_hh +   
## donates\_to\_conservative\_causes + g08\_precinct\_turnout + p08\_precinct\_turnout +   
## p12\_precinct\_turnout  
##   
## Df Deviance AIC F value Pr(>F)  
## - vh00p 1 25299 25395 1.8524 0.173514  
## - donates\_to\_conservative\_causes 1 25299 25395 2.0684 0.150382  
## - p08\_precinct\_turnout 1 25299 25395 3.0562 0.080438  
## <none> 25298 25396   
## + p10\_precinct\_turnout 1 25297 25397 0.8993 0.342984  
## - party 7 25314 25398 4.4364 6.095e-05  
## + donates\_to\_liberal\_causes 1 25298 25398 0.3882 0.533276  
## + intrst\_nascar\_in\_hh 1 25298 25398 0.1121 0.737729  
## + vh02g 1 25298 25398 0.0219 0.882227  
## + g10\_precinct\_turnout 1 25298 25398 0.0211 0.884387  
## - vh04p 1 25302 25398 9.0836 0.002580  
## - intrst\_musical\_instruments\_in\_hh 1 25303 25399 9.3282 0.002258  
## - vh02p 1 25303 25399 10.0717 0.001507  
## + home\_owner\_or\_renter 2 25297 25399 0.7692 0.463394  
## - p12\_precinct\_turnout 1 25306 25402 16.7045 4.375e-05  
## - vh06p 1 25307 25403 17.4960 2.884e-05  
## - cd 1 25309 25405 21.9258 2.841e-06  
## - dma 3 25313 25405 9.9809 1.425e-06  
## - vh08p 1 25311 25407 25.5213 4.391e-07  
## - income 5 25321 25409 9.3536 6.397e-09  
## - net\_worth 8 25330 25412 7.9454 9.433e-11  
## - g08\_precinct\_turnout 1 25324 25420 52.3062 4.817e-13  
## - petowner\_dog 1 25334 25430 71.6149 < 2.2e-16  
## - ethnicity 5 25348 25436 19.7082 < 2.2e-16  
## - vh06g 1 25342 25438 87.7118 < 2.2e-16  
## - vh00g 1 25352 25448 106.5921 < 2.2e-16  
## - vh04g 1 25354 25450 110.3066 < 2.2e-16  
## - vh10p 1 25412 25508 224.8674 < 2.2e-16  
## - age 1 25520 25616 438.0012 < 2.2e-16  
## - vh08g 1 25883 25979 1155.1988 < 2.2e-16  
## - vh10g 1 28288 28384 5901.5319 < 2.2e-16  
## - vh12p 1 28462 28558 6244.6482 < 2.2e-16  
##   
## - vh00p   
## - donates\_to\_conservative\_causes   
## - p08\_precinct\_turnout .   
## <none>   
## + p10\_precinct\_turnout   
## - party \*\*\*  
## + donates\_to\_liberal\_causes   
## + intrst\_nascar\_in\_hh   
## + vh02g   
## + g10\_precinct\_turnout   
## - vh04p \*\*   
## - intrst\_musical\_instruments\_in\_hh \*\*   
## - vh02p \*\*   
## + home\_owner\_or\_renter   
## - p12\_precinct\_turnout \*\*\*  
## - vh06p \*\*\*  
## - cd \*\*\*  
## - dma \*\*\*  
## - vh08p \*\*\*  
## - income \*\*\*  
## - net\_worth \*\*\*  
## - g08\_precinct\_turnout \*\*\*  
## - petowner\_dog \*\*\*  
## - ethnicity \*\*\*  
## - vh06g \*\*\*  
## - vh00g \*\*\*  
## - vh04g \*\*\*  
## - vh10p \*\*\*  
## - age \*\*\*  
## - vh08g \*\*\*  
## - vh10g \*\*\*  
## - vh12p \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Step: AIC=25394.66  
## vh12g ~ age + party + ethnicity + income + cd + dma + vh12p +   
## vh10g + vh10p + vh08g + vh08p + vh06g + vh06p + vh04g + vh04p +   
## vh02p + vh00g + net\_worth + petowner\_dog + intrst\_musical\_instruments\_in\_hh +   
## donates\_to\_conservative\_causes + g08\_precinct\_turnout + p08\_precinct\_turnout +   
## p12\_precinct\_turnout  
##   
## Df Deviance AIC F value Pr(>F)  
## - donates\_to\_conservative\_causes 1 25300 25394 2.1180 0.145584  
## - p08\_precinct\_turnout 1 25300 25394 3.1705 0.074986  
## <none> 25299 25395   
## + vh00p 1 25298 25396 1.8524 0.173514  
## + p10\_precinct\_turnout 1 25298 25396 0.9383 0.332731  
## + donates\_to\_liberal\_causes 1 25299 25397 0.3713 0.542319  
## + intrst\_nascar\_in\_hh 1 25299 25397 0.1146 0.734974  
## - party 7 25315 25397 4.5013 5.026e-05  
## + vh02g 1 25299 25397 0.0280 0.867058  
## + g10\_precinct\_turnout 1 25299 25397 0.0238 0.877288  
## - vh02p 1 25303 25397 8.5699 0.003419  
## - intrst\_musical\_instruments\_in\_hh 1 25303 25397 9.2552 0.002350  
## + home\_owner\_or\_renter 2 25298 25398 0.7718 0.462179  
## - vh04p 1 25304 25398 10.3181 0.001318  
## - vh06p 1 25307 25401 16.3199 5.358e-05  
## - p12\_precinct\_turnout 1 25307 25401 16.8521 4.047e-05  
## - dma 3 25314 25404 9.9389 1.515e-06  
## - cd 1 25310 25404 22.0590 2.651e-06  
## - vh08p 1 25312 25406 26.3362 2.879e-07  
## - income 5 25322 25408 9.3417 6.577e-09  
## - net\_worth 8 25331 25411 7.9289 1.001e-10  
## - g08\_precinct\_turnout 1 25325 25419 52.1840 5.126e-13  
## - petowner\_dog 1 25335 25429 71.7557 < 2.2e-16  
## - ethnicity 5 25349 25435 19.7237 < 2.2e-16  
## - vh06g 1 25343 25437 87.2152 < 2.2e-16  
## - vh04g 1 25355 25449 110.1436 < 2.2e-16  
## - vh00g 1 25358 25452 116.6671 < 2.2e-16  
## - vh10p 1 25414 25508 227.5094 < 2.2e-16  
## - age 1 25522 25616 440.0670 < 2.2e-16  
## - vh08g 1 25884 25978 1156.0807 < 2.2e-16  
## - vh10g 1 28291 28385 5904.9795 < 2.2e-16  
## - vh12p 1 28483 28577 6283.8579 < 2.2e-16  
##   
## - donates\_to\_conservative\_causes   
## - p08\_precinct\_turnout .   
## <none>   
## + vh00p   
## + p10\_precinct\_turnout   
## + donates\_to\_liberal\_causes   
## + intrst\_nascar\_in\_hh   
## - party \*\*\*  
## + vh02g   
## + g10\_precinct\_turnout   
## - vh02p \*\*   
## - intrst\_musical\_instruments\_in\_hh \*\*   
## + home\_owner\_or\_renter   
## - vh04p \*\*   
## - vh06p \*\*\*  
## - p12\_precinct\_turnout \*\*\*  
## - dma \*\*\*  
## - cd \*\*\*  
## - vh08p \*\*\*  
## - income \*\*\*  
## - net\_worth \*\*\*  
## - g08\_precinct\_turnout \*\*\*  
## - petowner\_dog \*\*\*  
## - ethnicity \*\*\*  
## - vh06g \*\*\*  
## - vh04g \*\*\*  
## - vh00g \*\*\*  
## - vh10p \*\*\*  
## - age \*\*\*  
## - vh08g \*\*\*  
## - vh10g \*\*\*  
## - vh12p \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Step: AIC=25393.74  
## vh12g ~ age + party + ethnicity + income + cd + dma + vh12p +   
## vh10g + vh10p + vh08g + vh08p + vh06g + vh06p + vh04g + vh04p +   
## vh02p + vh00g + net\_worth + petowner\_dog + intrst\_musical\_instruments\_in\_hh +   
## g08\_precinct\_turnout + p08\_precinct\_turnout + p12\_precinct\_turnout  
##   
## Df Deviance AIC F value Pr(>F)  
## - p08\_precinct\_turnout 1 25301 25393 3.1474 0.076055  
## <none> 25300 25394   
## + donates\_to\_conservative\_causes 1 25299 25395 2.1180 0.145584  
## + vh00p 1 25299 25395 1.9019 0.167869  
## + p10\_precinct\_turnout 1 25299 25395 0.9294 0.335031  
## + donates\_to\_liberal\_causes 1 25300 25396 0.4759 0.490287  
## - party 7 25316 25396 4.4836 5.296e-05  
## + intrst\_nascar\_in\_hh 1 25300 25396 0.1480 0.700477  
## + g10\_precinct\_turnout 1 25300 25396 0.0262 0.871379  
## + vh02g 1 25300 25396 0.0258 0.872370  
## - vh02p 1 25304 25396 8.7867 0.003036  
## - intrst\_musical\_instruments\_in\_hh 1 25304 25396 9.1839 0.002443  
## + home\_owner\_or\_renter 2 25299 25397 0.8078 0.445847  
## - vh04p 1 25305 25397 10.2869 0.001341  
## - vh06p 1 25308 25400 16.4282 5.060e-05  
## - p12\_precinct\_turnout 1 25308 25400 16.8228 4.110e-05  
## - cd 1 25311 25403 21.9385 2.823e-06  
## - dma 3 25315 25403 10.0009 1.384e-06  
## - vh08p 1 25313 25405 26.3149 2.911e-07  
## - income 5 25324 25408 9.3686 6.175e-09  
## - net\_worth 8 25332 25410 7.9068 1.085e-10  
## - g08\_precinct\_turnout 1 25326 25418 52.1218 5.290e-13  
## - petowner\_dog 1 25335 25427 70.3887 < 2.2e-16  
## - ethnicity 5 25350 25434 19.7398 < 2.2e-16  
## - vh06g 1 25344 25436 87.1857 < 2.2e-16  
## - vh04g 1 25356 25448 109.9733 < 2.2e-16  
## - vh00g 1 25359 25451 116.7354 < 2.2e-16  
## - vh10p 1 25415 25507 227.8683 < 2.2e-16  
## - age 1 25523 25615 439.9366 < 2.2e-16  
## - vh08g 1 25885 25977 1155.4188 < 2.2e-16  
## - vh10g 1 28293 28385 5908.2621 < 2.2e-16  
## - vh12p 1 28483 28575 6281.7740 < 2.2e-16  
##   
## - p08\_precinct\_turnout .   
## <none>   
## + donates\_to\_conservative\_causes   
## + vh00p   
## + p10\_precinct\_turnout   
## + donates\_to\_liberal\_causes   
## - party \*\*\*  
## + intrst\_nascar\_in\_hh   
## + g10\_precinct\_turnout   
## + vh02g   
## - vh02p \*\*   
## - intrst\_musical\_instruments\_in\_hh \*\*   
## + home\_owner\_or\_renter   
## - vh04p \*\*   
## - vh06p \*\*\*  
## - p12\_precinct\_turnout \*\*\*  
## - cd \*\*\*  
## - dma \*\*\*  
## - vh08p \*\*\*  
## - income \*\*\*  
## - net\_worth \*\*\*  
## - g08\_precinct\_turnout \*\*\*  
## - petowner\_dog \*\*\*  
## - ethnicity \*\*\*  
## - vh06g \*\*\*  
## - vh04g \*\*\*  
## - vh00g \*\*\*  
## - vh10p \*\*\*  
## - age \*\*\*  
## - vh08g \*\*\*  
## - vh10g \*\*\*  
## - vh12p \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Step: AIC=25393.33  
## vh12g ~ age + party + ethnicity + income + cd + dma + vh12p +   
## vh10g + vh10p + vh08g + vh08p + vh06g + vh06p + vh04g + vh04p +   
## vh02p + vh00g + net\_worth + petowner\_dog + intrst\_musical\_instruments\_in\_hh +   
## g08\_precinct\_turnout + p12\_precinct\_turnout  
##   
## Df Deviance AIC F value Pr(>F)  
## <none> 25301 25393   
## + p08\_precinct\_turnout 1 25300 25394 3.1474 0.0760548  
## + donates\_to\_conservative\_causes 1 25300 25394 2.0949 0.1478007  
## + vh00p 1 25300 25394 2.0174 0.1555114  
## + donates\_to\_liberal\_causes 1 25301 25395 0.4837 0.4867675  
## - party 7 25317 25395 4.4718 5.486e-05  
## + intrst\_nascar\_in\_hh 1 25301 25395 0.1345 0.7137657  
## + g10\_precinct\_turnout 1 25301 25395 0.1000 0.7517904  
## + p10\_precinct\_turnout 1 25301 25395 0.0315 0.8591760  
## + vh02g 1 25301 25395 0.0207 0.8857363  
## - intrst\_musical\_instruments\_in\_hh 1 25306 25396 9.3255 0.0022611  
## - vh02p 1 25306 25396 9.6846 0.0018592  
## + home\_owner\_or\_renter 2 25301 25397 0.8140 0.4430911  
## - vh04p 1 25307 25397 10.5745 0.0011473  
## - vh06p 1 25309 25399 15.1100 0.0001016  
## - p12\_precinct\_turnout 1 25309 25399 15.6446 7.653e-05  
## - dma 3 25317 25403 10.4267 7.455e-07  
## - cd 1 25314 25404 24.7474 6.557e-07  
## - vh08p 1 25315 25405 26.0807 3.286e-07  
## - income 5 25325 25407 9.2275 8.595e-09  
## - net\_worth 8 25333 25409 7.8360 1.403e-10  
## - petowner\_dog 1 25337 25427 71.0595 < 2.2e-16  
## - g08\_precinct\_turnout 1 25341 25431 78.1253 < 2.2e-16  
## - ethnicity 5 25351 25433 19.7477 < 2.2e-16  
## - vh06g 1 25345 25435 86.7453 < 2.2e-16  
## - vh04g 1 25357 25447 109.8216 < 2.2e-16  
## - vh00g 1 25360 25450 116.0865 < 2.2e-16  
## - vh10p 1 25417 25507 229.1532 < 2.2e-16  
## - age 1 25524 25614 439.6593 < 2.2e-16  
## - vh08g 1 25888 25978 1157.0784 < 2.2e-16  
## - vh10g 1 28294 28384 5906.0942 < 2.2e-16  
## - vh12p 1 28496 28586 6304.4345 < 2.2e-16  
##   
## <none>   
## + p08\_precinct\_turnout .   
## + donates\_to\_conservative\_causes   
## + vh00p   
## + donates\_to\_liberal\_causes   
## - party \*\*\*  
## + intrst\_nascar\_in\_hh   
## + g10\_precinct\_turnout   
## + p10\_precinct\_turnout   
## + vh02g   
## - intrst\_musical\_instruments\_in\_hh \*\*   
## - vh02p \*\*   
## + home\_owner\_or\_renter   
## - vh04p \*\*   
## - vh06p \*\*\*  
## - p12\_precinct\_turnout \*\*\*  
## - dma \*\*\*  
## - cd \*\*\*  
## - vh08p \*\*\*  
## - income \*\*\*  
## - net\_worth \*\*\*  
## - petowner\_dog \*\*\*  
## - g08\_precinct\_turnout \*\*\*  
## - ethnicity \*\*\*  
## - vh06g \*\*\*  
## - vh04g \*\*\*  
## - vh00g \*\*\*  
## - vh10p \*\*\*  
## - age \*\*\*  
## - vh08g \*\*\*  
## - vh10g \*\*\*  
## - vh12p \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

summary(voter.step.backward)

##   
## Call:  
## glm(formula = vh12g ~ age + party + ethnicity + income + cd +   
## dma + vh12p + vh10g + vh10p + vh08g + vh08p + vh06g + vh06p +   
## vh04g + vh04p + vh02p + vh00g + net\_worth + petowner\_dog +   
## intrst\_musical\_instruments\_in\_hh + g08\_precinct\_turnout +   
## p12\_precinct\_turnout, family = binomial(link = "logit"),   
## data = voter.clean.var)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.1242 -0.2473 -0.1419 0.1557 3.2212   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) -6.389727 0.255645 -24.995 < 2e-16  
## age 0.018329 0.001234 14.849 < 2e-16  
## partyDemocratic 0.224810 0.086812 2.590 0.009608  
## partyGreen 0.851044 0.668387 1.273 0.202919  
## partyLibertarian 0.681048 0.226324 3.009 0.002620  
## partyNatural Law 0.514643 1.709258 0.301 0.763345  
## partyNon-Partisan 0.244940 0.092744 2.641 0.008266  
## partyOther 0.012755 0.239251 0.053 0.957483  
## partyRepublican 0.264203 0.087068 3.034 0.002410  
## ethnicityAsian -0.425923 0.186473 -2.284 0.022366  
## ethnicityEuropean 0.091822 0.148605 0.618 0.536645  
## ethnicityHispanic -0.193902 0.154296 -1.257 0.208868  
## ethnicitynan 0.006433 0.155678 0.041 0.967038  
## ethnicityOther -0.114222 0.186509 -0.612 0.540260  
## income125k-200k 0.076825 0.075204 1.022 0.306992  
## income200k+ 0.204465 0.094068 2.174 0.029737  
## income35k-75k -0.123747 0.061038 -2.027 0.042623  
## income75k-125k 0.004243 0.063488 0.067 0.946718  
## incomeUnknown -0.059464 0.066071 -0.900 0.368121  
## cd -0.060455 0.017062 -3.543 0.000395  
## dmaLOS ANGELES DMA (EST.) 0.983936 0.599745 1.641 0.100883  
## dmaRENO DMA (EST.) 0.062773 0.044059 1.425 0.154229  
## dmaSALT LAKE CITY DMA (EST.) 0.416871 0.116135 3.590 0.000331  
## vh12p1 3.219319 0.075334 42.734 < 2e-16  
## vh10g1 2.503270 0.052524 47.660 < 2e-16  
## vh10p1 0.475754 0.044275 10.746 < 2e-16  
## vh08g1 0.973196 0.040641 23.946 < 2e-16  
## vh08p1 0.157712 0.043362 3.637 0.000276  
## vh06g1 0.387136 0.058912 6.571 4.98e-11  
## vh06p1 -0.142471 0.051534 -2.765 0.005700  
## vh04g1 0.341207 0.045837 7.444 9.77e-14  
## vh04p1 0.098222 0.042408 2.316 0.020551  
## vh02p1 -0.107775 0.048694 -2.213 0.026876  
## vh00g1 0.283190 0.036855 7.684 1.54e-14  
## net\_worth$10000-24999 -0.095390 0.181116 -0.527 0.598417  
## net\_worth$100000-249999 0.180088 0.131129 1.373 0.169639  
## net\_worth$25000-49999 -0.142259 0.154691 -0.920 0.357764  
## net\_worth$250000-499999 0.024567 0.133777 0.184 0.854293  
## net\_worth$499999+ -0.078532 0.142570 -0.551 0.581749  
## net\_worth$5000-9999 -0.133946 0.204777 -0.654 0.513041  
## net\_worth$50000-99999 0.201079 0.155078 1.297 0.194756  
## net\_worthnan 0.014517 0.130963 0.111 0.911735  
## petowner\_dogYes 0.300129 0.050073 5.994 2.05e-09  
## intrst\_musical\_instruments\_in\_hhYes 0.298647 0.137543 2.171 0.029909  
## g08\_precinct\_turnout 1.424487 0.227700 6.256 3.95e-10  
## p12\_precinct\_turnout -0.784028 0.279140 -2.809 0.004974  
##   
## (Intercept) \*\*\*  
## age \*\*\*  
## partyDemocratic \*\*   
## partyGreen   
## partyLibertarian \*\*   
## partyNatural Law   
## partyNon-Partisan \*\*   
## partyOther   
## partyRepublican \*\*   
## ethnicityAsian \*   
## ethnicityEuropean   
## ethnicityHispanic   
## ethnicitynan   
## ethnicityOther   
## income125k-200k   
## income200k+ \*   
## income35k-75k \*   
## income75k-125k   
## incomeUnknown   
## cd \*\*\*  
## dmaLOS ANGELES DMA (EST.)   
## dmaRENO DMA (EST.)   
## dmaSALT LAKE CITY DMA (EST.) \*\*\*  
## vh12p1 \*\*\*  
## vh10g1 \*\*\*  
## vh10p1 \*\*\*  
## vh08g1 \*\*\*  
## vh08p1 \*\*\*  
## vh06g1 \*\*\*  
## vh06p1 \*\*   
## vh04g1 \*\*\*  
## vh04p1 \*   
## vh02p1 \*   
## vh00g1 \*\*\*  
## net\_worth$10000-24999   
## net\_worth$100000-249999   
## net\_worth$25000-49999   
## net\_worth$250000-499999   
## net\_worth$499999+   
## net\_worth$5000-9999   
## net\_worth$50000-99999   
## net\_worthnan   
## petowner\_dogYes \*\*\*  
## intrst\_musical\_instruments\_in\_hhYes \*   
## g08\_precinct\_turnout \*\*\*  
## p12\_precinct\_turnout \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 57682 on 49977 degrees of freedom  
## Residual deviance: 25301 on 49932 degrees of freedom  
## (22 observations deleted due to missingness)  
## AIC: 25393  
##   
## Number of Fisher Scoring iterations: 7

**The AIC has been reduced to 25393 and now we can go ahead and remove some insignificant variables and rebuild our model.**  
## Removing insignificant variables found from stepwise:

voter.clean.var %>%  
 select(-vh02g, -vh00p, -intrst\_nascar\_in\_hh, -donates\_to\_conservative\_causes, -donates\_to\_liberal\_causes, -home\_owner\_or\_renter, -g10\_precinct\_turnout, -p08\_precinct\_turnout, -p10\_precinct\_turnout) ->  
 voter.clean.var.step  
  
str(voter.clean.var.step)

## 'data.frame': 50000 obs. of 23 variables:  
## $ age : num 69 20 28 78 68 69 53 36 53 30 ...  
## $ party : Factor w/ 8 levels "American Independent",..: 8 1 6 1 2 2 8 2 2 6 ...  
## $ ethnicity : Factor w/ 6 levels "African-American",..: 3 3 3 3 5 3 3 1 3 2 ...  
## $ income : Factor w/ 6 levels "0-35k","125k-200k",..: 5 6 6 6 6 5 6 6 4 3 ...  
## $ cd : num 4 2 3 3 4 2 3 4 3 3 ...  
## $ dma : Factor w/ 4 levels "LAS VEGAS DMA (EST.)",..: 1 3 1 1 1 3 1 1 1 1 ...  
## $ vh12g : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ vh12p : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ vh10g : Factor w/ 2 levels "0","1": 2 1 1 1 2 1 1 2 2 2 ...  
## $ vh10p : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ vh08g : Factor w/ 2 levels "0","1": 2 1 1 1 2 2 1 1 1 2 ...  
## $ vh08p : Factor w/ 2 levels "0","1": 1 1 1 1 2 1 1 1 1 1 ...  
## $ vh06g : Factor w/ 2 levels "0","1": 2 1 2 1 2 2 1 2 1 2 ...  
## $ vh06p : Factor w/ 2 levels "0","1": 1 1 1 1 2 2 1 1 1 1 ...  
## $ vh04g : Factor w/ 2 levels "0","1": 2 1 2 1 2 2 1 2 1 1 ...  
## $ vh04p : Factor w/ 2 levels "0","1": 1 1 1 1 2 2 1 1 1 1 ...  
## $ vh02p : Factor w/ 2 levels "0","1": 1 1 1 1 2 1 1 1 1 1 ...  
## $ vh00g : Factor w/ 2 levels "0","1": 2 1 1 1 2 2 2 2 1 2 ...  
## $ net\_worth : Factor w/ 9 levels "$1-4999","$10000-24999",..: 3 9 5 9 9 5 9 9 3 9 ...  
## $ petowner\_dog : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 2 1 ...  
## $ intrst\_musical\_instruments\_in\_hh: Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...  
## $ g08\_precinct\_turnout : num 0.56 0.84 0.49 0.84 0.71 0.69 0.75 0.67 0.75 0.62 ...  
## $ p12\_precinct\_turnout : num 0.24 0.47 0.06 0.3 0.34 0.17 0.17 0.15 0.13 0.09 ...

## Applying logistic model to the reduced dataset from Stepwise and alread removed variables:

voter.logit.step <- glm(vh12g ~ ., data=voter.clean.var.step, family=binomial(link="logit"))  
  
summary(voter.logit.step)

##   
## Call:  
## glm(formula = vh12g ~ ., family = binomial(link = "logit"), data = voter.clean.var.step)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.1242 -0.2473 -0.1419 0.1557 3.2212   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) -6.389727 0.255645 -24.995 < 2e-16  
## age 0.018329 0.001234 14.849 < 2e-16  
## partyDemocratic 0.224810 0.086812 2.590 0.009608  
## partyGreen 0.851044 0.668387 1.273 0.202919  
## partyLibertarian 0.681048 0.226324 3.009 0.002620  
## partyNatural Law 0.514643 1.709258 0.301 0.763345  
## partyNon-Partisan 0.244940 0.092744 2.641 0.008266  
## partyOther 0.012755 0.239251 0.053 0.957483  
## partyRepublican 0.264203 0.087068 3.034 0.002410  
## ethnicityAsian -0.425923 0.186473 -2.284 0.022366  
## ethnicityEuropean 0.091822 0.148605 0.618 0.536645  
## ethnicityHispanic -0.193902 0.154296 -1.257 0.208868  
## ethnicitynan 0.006433 0.155678 0.041 0.967038  
## ethnicityOther -0.114222 0.186509 -0.612 0.540260  
## income125k-200k 0.076825 0.075204 1.022 0.306992  
## income200k+ 0.204465 0.094068 2.174 0.029737  
## income35k-75k -0.123747 0.061038 -2.027 0.042623  
## income75k-125k 0.004243 0.063488 0.067 0.946718  
## incomeUnknown -0.059464 0.066071 -0.900 0.368121  
## cd -0.060455 0.017062 -3.543 0.000395  
## dmaLOS ANGELES DMA (EST.) 0.983936 0.599745 1.641 0.100883  
## dmaRENO DMA (EST.) 0.062773 0.044059 1.425 0.154229  
## dmaSALT LAKE CITY DMA (EST.) 0.416871 0.116135 3.590 0.000331  
## vh12p1 3.219319 0.075334 42.734 < 2e-16  
## vh10g1 2.503270 0.052524 47.660 < 2e-16  
## vh10p1 0.475754 0.044275 10.746 < 2e-16  
## vh08g1 0.973196 0.040641 23.946 < 2e-16  
## vh08p1 0.157712 0.043362 3.637 0.000276  
## vh06g1 0.387136 0.058912 6.571 4.98e-11  
## vh06p1 -0.142471 0.051534 -2.765 0.005700  
## vh04g1 0.341207 0.045837 7.444 9.77e-14  
## vh04p1 0.098222 0.042408 2.316 0.020551  
## vh02p1 -0.107775 0.048694 -2.213 0.026876  
## vh00g1 0.283190 0.036855 7.684 1.54e-14  
## net\_worth$10000-24999 -0.095390 0.181116 -0.527 0.598417  
## net\_worth$100000-249999 0.180088 0.131129 1.373 0.169639  
## net\_worth$25000-49999 -0.142259 0.154691 -0.920 0.357764  
## net\_worth$250000-499999 0.024567 0.133777 0.184 0.854293  
## net\_worth$499999+ -0.078532 0.142570 -0.551 0.581749  
## net\_worth$5000-9999 -0.133946 0.204777 -0.654 0.513041  
## net\_worth$50000-99999 0.201079 0.155078 1.297 0.194756  
## net\_worthnan 0.014517 0.130963 0.111 0.911735  
## petowner\_dogYes 0.300129 0.050073 5.994 2.05e-09  
## intrst\_musical\_instruments\_in\_hhYes 0.298647 0.137543 2.171 0.029909  
## g08\_precinct\_turnout 1.424487 0.227700 6.256 3.95e-10  
## p12\_precinct\_turnout -0.784028 0.279140 -2.809 0.004974  
##   
## (Intercept) \*\*\*  
## age \*\*\*  
## partyDemocratic \*\*   
## partyGreen   
## partyLibertarian \*\*   
## partyNatural Law   
## partyNon-Partisan \*\*   
## partyOther   
## partyRepublican \*\*   
## ethnicityAsian \*   
## ethnicityEuropean   
## ethnicityHispanic   
## ethnicitynan   
## ethnicityOther   
## income125k-200k   
## income200k+ \*   
## income35k-75k \*   
## income75k-125k   
## incomeUnknown   
## cd \*\*\*  
## dmaLOS ANGELES DMA (EST.)   
## dmaRENO DMA (EST.)   
## dmaSALT LAKE CITY DMA (EST.) \*\*\*  
## vh12p1 \*\*\*  
## vh10g1 \*\*\*  
## vh10p1 \*\*\*  
## vh08g1 \*\*\*  
## vh08p1 \*\*\*  
## vh06g1 \*\*\*  
## vh06p1 \*\*   
## vh04g1 \*\*\*  
## vh04p1 \*   
## vh02p1 \*   
## vh00g1 \*\*\*  
## net\_worth$10000-24999   
## net\_worth$100000-249999   
## net\_worth$25000-49999   
## net\_worth$250000-499999   
## net\_worth$499999+   
## net\_worth$5000-9999   
## net\_worth$50000-99999   
## net\_worthnan   
## petowner\_dogYes \*\*\*  
## intrst\_musical\_instruments\_in\_hhYes \*   
## g08\_precinct\_turnout \*\*\*  
## p12\_precinct\_turnout \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 57682 on 49977 degrees of freedom  
## Residual deviance: 25301 on 49932 degrees of freedom  
## (22 observations deleted due to missingness)  
## AIC: 25393  
##   
## Number of Fisher Scoring iterations: 7

**The AIC has decreased to 25393.**

## Let us go ahead and make predictions and cross validate for our reduced model from Stepwise:

attach(voter.clean.var.step)   
set.seed(1)  
train <- sample(1:nrow(voter.clean.var.step), 0.7\*nrow(voter.clean.var.step))  
test <- seq(1:nrow(voter.clean.var.step))[-train]  
  
voter.logit.train.red.step <- glm(vh12g ~ ., family=binomial(link="logit"), data=voter.clean.var.step[train,])  
  
voter.logit.test.red.step=predict(voter.logit.train.red.step, voter.clean.var.step, type="response")[test]  
  
voter.logit.test.red.step[1:10]

## 3 4 8 13 14 20   
## 0.011018623 0.016004812 0.217246520 0.184599421 0.161237566 0.947369966   
## 23 28 31 33   
## 0.565915508 0.416760707 0.024919832 0.007978673

voter.pred.test.red.step = ifelse(voter.logit.test.red.step>0.5, 1,0)  
voter.pred.test.red.step[1:10] # List first 10

## 3 4 8 13 14 20 23 28 31 33   
## 0 0 0 0 0 1 1 0 0 0

conf.mat <- table("Predicted"=voter.pred.test.red.step, "Actual"=vh12g[test])   
  
colnames(conf.mat) <- c("No", "Yes")  
rownames(conf.mat) <- c("No", "Yes")  
  
conf.mat

## Actual  
## Predicted No Yes  
## No 10204 927  
## Yes 886 2977

TruN <- conf.mat[1,1] # True negatives  
TruP <- conf.mat[2,2] # True positives  
FalN <- conf.mat[1,2] # False negatives  
FalP <- conf.mat[2,1] # False positives  
TotN <- conf.mat[1,1] + conf.mat[2,1] # Total negatives  
TotP <- conf.mat[1,2] + conf.mat[2,2] # Total positives  
Tot <- TotN+TotP # Total  
  
Accuracy.Rate <- (TruN + TruP) / Tot  
  
Accuracy.Rate

## [1] 0.879085

Error.Rate <- (FalN + FalP) / Tot  
Error.Rate

## [1] 0.120915

Sensitivity <- TruP / TotP   
Sensitivity

## [1] 0.7625512

Specificity <- TruN / TotN   
Specificity

## [1] 0.9201082

FalseP.Rate <- 1 - Specificity  
FalseP.Rate

## [1] 0.07989179

logit.rates.50.red.step <- c(Accuracy.Rate, Error.Rate, Sensitivity, Specificity, FalseP.Rate)  
  
names(logit.rates.50.red.step) <- c("Accuracy Rate", "Error Rate", "Sensitivity", "Specificity", "False Positives")  
  
print(logit.rates.50.red.step, digits=2)

## Accuracy Rate Error Rate Sensitivity Specificity   
## 0.88 0.12 0.76 0.92   
## False Positives   
## 0.08

logit.fit.stats.compare.step <- rbind(logit.fit.stats.compare.red1, logit.rates.50.red.step)  
print(logit.fit.stats.compare.step, digits=2)

## Accuracy Rate Error Rate Sensitivity Specificity  
## logit.rates.50 0.88 0.12 0.76 0.92  
## logit.rates.40 0.88 0.12 0.84 0.89  
## logit.rates.50.red 0.88 0.12 0.76 0.92  
## logit.rates.40.red 0.88 0.12 0.84 0.89  
## logit.rates.50.red.step 0.88 0.12 0.76 0.92  
## False Positives  
## logit.rates.50 0.080  
## logit.rates.40 0.112  
## logit.rates.50.red 0.079  
## logit.rates.40.red 0.113  
## logit.rates.50.red.step 0.080

**As we can see here the reduced stepwise model does even perform somewhat the same as our previous models but helps reduce insignificant variables.**

**The result from ROC is pretty much the same and still pretty good!**

## Confusion matrix with threshold = 0.40.

voter.pred.test.red.step = ifelse(voter.logit.test.red.step>0.40, 1,0)  
voter.pred.test.red.step[1:10]

## 3 4 8 13 14 20 23 28 31 33   
## 0 0 0 0 0 1 1 1 0 0

conf.mat <- table("Predicted"=voter.pred.test.red.step, "Actual"=vh12g[test])   
  
colnames(conf.mat) <- c("No", "Yes")  
rownames(conf.mat) <- c("No", "Yes")  
  
conf.mat

## Actual  
## Predicted No Yes  
## No 9831 610  
## Yes 1259 3294

TruN <- conf.mat[1,1] # True negatives  
TruP <- conf.mat[2,2] # True positives  
FalN <- conf.mat[1,2] # False negatives  
FalP <- conf.mat[2,1] # False positives  
TotN <- conf.mat[1,1] + conf.mat[2,1] # Total negatives  
TotP <- conf.mat[1,2] + conf.mat[2,2] # Total positives  
Tot <- TotN+TotP # Total  
  
Accuracy.Rate <- (TruN + TruP) / Tot  
Accuracy.Rate

## [1] 0.8753501

Error.Rate <- (FalN + FalP) / Tot  
Error.Rate

## [1] 0.1246499

Sensitivity <- TruP / TotP   
Sensitivity

## [1] 0.84375

Specificity <- TruN / TotN   
Specificity

## [1] 0.8864743

FalseP.Rate <- 1 - Specificity  
FalseP.Rate

## [1] 0.1135257

logit.rates.40.red.step <- c(Accuracy.Rate, Error.Rate, Sensitivity, Specificity, FalseP.Rate)  
  
names(logit.rates.40.red.step) <- c("Accuracy Rate", "Error Rate", "Sensitivity", "Specificity", "False Positives")  
  
logit.fit.stats.compare.step1 <- rbind(logit.fit.stats.compare.step, logit.rates.40.red.step)  
print(logit.fit.stats.compare.step1, digits=3)

## Accuracy Rate Error Rate Sensitivity Specificity  
## logit.rates.50 0.879 0.121 0.762 0.920  
## logit.rates.40 0.877 0.123 0.845 0.888  
## logit.rates.50.red 0.879 0.121 0.762 0.921  
## logit.rates.40.red 0.876 0.124 0.844 0.887  
## logit.rates.50.red.step 0.879 0.121 0.763 0.920  
## logit.rates.40.red.step 0.875 0.125 0.844 0.886  
## False Positives  
## logit.rates.50 0.0796  
## logit.rates.40 0.1123  
## logit.rates.50.red 0.0794  
## logit.rates.40.red 0.1129  
## logit.rates.50.red.step 0.0799  
## logit.rates.40.red.step 0.1135

**There is not much of a difference and now let us predict the voter values for 2014 general election**

## Check for multicollinearity:

vif(voter.logit.step)

## GVIF Df GVIF^(1/(2\*Df))  
## age 1.331964 1 1.154108  
## party 1.150021 7 1.010034  
## ethnicity 1.099394 5 1.009521  
## income 1.841815 5 1.062979  
## cd 1.222834 1 1.105818  
## dma 1.649886 3 1.087032  
## vh12p 1.041498 1 1.020538  
## vh10g 1.263335 1 1.123982  
## vh10p 1.220690 1 1.104849  
## vh08g 1.438336 1 1.199307  
## vh08p 1.437362 1 1.198900  
## vh06g 1.398817 1 1.182716  
## vh06p 1.388496 1 1.178345  
## vh04g 1.444523 1 1.201883  
## vh04p 1.546490 1 1.243580  
## vh02p 1.434997 1 1.197914  
## vh00g 1.259349 1 1.122207  
## net\_worth 2.380398 8 1.055700  
## petowner\_dog 1.107032 1 1.052156  
## intrst\_musical\_instruments\_in\_hh 1.015281 1 1.007611  
## g08\_precinct\_turnout 2.070431 1 1.438899  
## p12\_precinct\_turnout 2.268814 1 1.506258

**We can see that the Variable Inflation Factor for all the variables is well under acceptable values, thus we can say that our model passes the multicollinearity test.**

## Durbin Watson test to check serial correlation:

dwtest(voter.logit.step)

##   
## Durbin-Watson test  
##   
## data: voter.logit.step  
## DW = 2.0025, p-value = 0.6101  
## alternative hypothesis: true autocorrelation is greater than 0

**With DW value of 2.0025 we can easily say that there is almost no serial correlation and our model seems fine with regard to serial correlation!**

## Finally, let us predict the voter probabilities for voting in 2014 but we will also have to add vh14p and g12\_precint\_turnout:

#Adding the two variables to our reduced variables till now!  
  
voter %>%  
 select(age, party, ethnicity, income, cd, dma,vh14p, vh12g, vh12p, vh10p, vh10g, vh08g, vh08p, vh06g, vh06p, vh04g, vh04p, vh02p, vh00g, net\_worth, petowner\_dog, intrst\_musical\_instruments\_in\_hh, g08\_precinct\_turnout, p12\_precinct\_turnout, vh14p, g12\_precinct\_turnout ) ->  
 voter.final  
   
str(voter.final)

## 'data.frame': 50000 obs. of 25 variables:  
## $ age : num 69 20 28 78 68 69 53 36 53 30 ...  
## $ party : Factor w/ 8 levels "American Independent",..: 8 1 6 1 2 2 8 2 2 6 ...  
## $ ethnicity : Factor w/ 6 levels "African-American",..: 3 3 3 3 5 3 3 1 3 2 ...  
## $ income : Factor w/ 6 levels "0-35k","125k-200k",..: 5 6 6 6 6 5 6 6 4 3 ...  
## $ cd : num 4 2 3 3 4 2 3 4 3 3 ...  
## $ dma : Factor w/ 4 levels "LAS VEGAS DMA (EST.)",..: 1 3 1 1 1 3 1 1 1 1 ...  
## $ vh14p : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ vh12g : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ vh12p : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ vh10p : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ vh10g : Factor w/ 2 levels "0","1": 2 1 1 1 2 1 1 2 2 2 ...  
## $ vh08g : Factor w/ 2 levels "0","1": 2 1 1 1 2 2 1 1 1 2 ...  
## $ vh08p : Factor w/ 2 levels "0","1": 1 1 1 1 2 1 1 1 1 1 ...  
## $ vh06g : Factor w/ 2 levels "0","1": 2 1 2 1 2 2 1 2 1 2 ...  
## $ vh06p : Factor w/ 2 levels "0","1": 1 1 1 1 2 2 1 1 1 1 ...  
## $ vh04g : Factor w/ 2 levels "0","1": 2 1 2 1 2 2 1 2 1 1 ...  
## $ vh04p : Factor w/ 2 levels "0","1": 1 1 1 1 2 2 1 1 1 1 ...  
## $ vh02p : Factor w/ 2 levels "0","1": 1 1 1 1 2 1 1 1 1 1 ...  
## $ vh00g : Factor w/ 2 levels "0","1": 2 1 1 1 2 2 2 2 1 2 ...  
## $ net\_worth : Factor w/ 9 levels "$1-4999","$10000-24999",..: 3 9 5 9 9 5 9 9 3 9 ...  
## $ petowner\_dog : Factor w/ 2 levels "nan","Yes": 1 1 1 1 1 1 1 1 2 1 ...  
## $ intrst\_musical\_instruments\_in\_hh: Factor w/ 2 levels "nan","Yes": 1 1 1 1 1 1 1 1 1 1 ...  
## $ g08\_precinct\_turnout : num 0.56 0.84 0.49 0.84 0.71 0.69 0.75 0.67 0.75 0.62 ...  
## $ p12\_precinct\_turnout : num 0.24 0.47 0.06 0.3 0.34 0.17 0.17 0.15 0.13 0.09 ...  
## $ g12\_precinct\_turnout : num 0.75 0.92 0.7 0.91 0.81 0.8 0.81 0.79 0.79 0.73 ...

#Replacing all nan for the following variables to No as it seems the most logical way to me.  
voter.final %>%  
 mutate(petowner\_dog = str\_replace\_all(petowner\_dog, "nan", "No")) %>%  
 mutate(petowner\_dog = as.factor(petowner\_dog)) %>%  
 mutate(intrst\_musical\_instruments\_in\_hh = str\_replace\_all(intrst\_musical\_instruments\_in\_hh, "nan", "No")) %>%  
 mutate(intrst\_musical\_instruments\_in\_hh = as.factor(intrst\_musical\_instruments\_in\_hh)) ->  
 voter.clean.var.final  
  
#Final Logistic prediction model starts here:  
  
voter.logit.final <- glm(vh12g ~ ., data=voter.clean.var.final, family=binomial(link="logit"))  
summary(voter.logit.final)

##   
## Call:  
## glm(formula = vh12g ~ ., family = binomial(link = "logit"), data = voter.clean.var.final)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.3849 -0.2441 -0.1427 0.0991 3.2402   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) -5.865260 0.297081 -19.743 < 2e-16  
## age 0.016913 0.001250 13.530 < 2e-16  
## partyDemocratic 0.185031 0.087113 2.124 0.033668  
## partyGreen 0.845906 0.668945 1.265 0.206037  
## partyLibertarian 0.653311 0.227360 2.873 0.004060  
## partyNatural Law 0.601026 1.673309 0.359 0.719457  
## partyNon-Partisan 0.224197 0.093034 2.410 0.015960  
## partyOther -0.015082 0.241161 -0.063 0.950135  
## partyRepublican 0.224486 0.087391 2.569 0.010206  
## ethnicityAsian -0.294319 0.189780 -1.551 0.120940  
## ethnicityEuropean 0.182689 0.152421 1.199 0.230690  
## ethnicityHispanic -0.110114 0.158074 -0.697 0.486056  
## ethnicitynan 0.109619 0.159505 0.687 0.491929  
## ethnicityOther -0.017848 0.190874 -0.094 0.925500  
## income125k-200k 0.050575 0.076575 0.660 0.508957  
## income200k+ 0.196031 0.095583 2.051 0.040276  
## income35k-75k -0.131772 0.062022 -2.125 0.033620  
## income75k-125k 0.006002 0.064530 0.093 0.925891  
## incomeUnknown -0.068809 0.067015 -1.027 0.304532  
## cd -0.049895 0.017535 -2.845 0.004435  
## dmaLOS ANGELES DMA (EST.) 1.141995 0.586915 1.946 0.051684  
## dmaRENO DMA (EST.) 0.099027 0.047790 2.072 0.038253  
## dmaSALT LAKE CITY DMA (EST.) 0.604344 0.118319 5.108 3.26e-07  
## vh14p1 1.539686 0.071126 21.647 < 2e-16  
## vh12p1 2.975424 0.076367 38.962 < 2e-16  
## vh10p1 0.371810 0.045350 8.199 2.43e-16  
## vh10g1 2.461494 0.053012 46.433 < 2e-16  
## vh08g1 0.963629 0.041084 23.455 < 2e-16  
## vh08p1 0.113597 0.044220 2.569 0.010203  
## vh06g1 0.401871 0.059328 6.774 1.25e-11  
## vh06p1 -0.199825 0.052909 -3.777 0.000159  
## vh04g1 0.328257 0.046242 7.099 1.26e-12  
## vh04p1 0.080233 0.043086 1.862 0.062581  
## vh02p1 -0.134220 0.049796 -2.695 0.007031  
## vh00g1 0.283688 0.037333 7.599 2.99e-14  
## net\_worth$10000-24999 -0.095292 0.181496 -0.525 0.599558  
## net\_worth$100000-249999 0.152684 0.131211 1.164 0.244565  
## net\_worth$25000-49999 -0.193300 0.155615 -1.242 0.214175  
## net\_worth$250000-499999 0.017306 0.133971 0.129 0.897217  
## net\_worth$499999+ -0.064356 0.143076 -0.450 0.652851  
## net\_worth$5000-9999 -0.182288 0.206513 -0.883 0.377400  
## net\_worth$50000-99999 0.106680 0.156506 0.682 0.495467  
## net\_worthnan -0.011036 0.131062 -0.084 0.932895  
## petowner\_dogYes 0.298275 0.050802 5.871 4.32e-09  
## intrst\_musical\_instruments\_in\_hhYes 0.299222 0.139288 2.148 0.031697  
## g08\_precinct\_turnout 2.412974 0.396629 6.084 1.17e-09  
## p12\_precinct\_turnout -0.782390 0.283557 -2.759 0.005794  
## g12\_precinct\_turnout -1.528982 0.498817 -3.065 0.002175  
##   
## (Intercept) \*\*\*  
## age \*\*\*  
## partyDemocratic \*   
## partyGreen   
## partyLibertarian \*\*   
## partyNatural Law   
## partyNon-Partisan \*   
## partyOther   
## partyRepublican \*   
## ethnicityAsian   
## ethnicityEuropean   
## ethnicityHispanic   
## ethnicitynan   
## ethnicityOther   
## income125k-200k   
## income200k+ \*   
## income35k-75k \*   
## income75k-125k   
## incomeUnknown   
## cd \*\*   
## dmaLOS ANGELES DMA (EST.) .   
## dmaRENO DMA (EST.) \*   
## dmaSALT LAKE CITY DMA (EST.) \*\*\*  
## vh14p1 \*\*\*  
## vh12p1 \*\*\*  
## vh10p1 \*\*\*  
## vh10g1 \*\*\*  
## vh08g1 \*\*\*  
## vh08p1 \*   
## vh06g1 \*\*\*  
## vh06p1 \*\*\*  
## vh04g1 \*\*\*  
## vh04p1 .   
## vh02p1 \*\*   
## vh00g1 \*\*\*  
## net\_worth$10000-24999   
## net\_worth$100000-249999   
## net\_worth$25000-49999   
## net\_worth$250000-499999   
## net\_worth$499999+   
## net\_worth$5000-9999   
## net\_worth$50000-99999   
## net\_worthnan   
## petowner\_dogYes \*\*\*  
## intrst\_musical\_instruments\_in\_hhYes \*   
## g08\_precinct\_turnout \*\*\*  
## p12\_precinct\_turnout \*\*   
## g12\_precinct\_turnout \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 57682 on 49977 degrees of freedom  
## Residual deviance: 24726 on 49930 degrees of freedom  
## (22 observations deleted due to missingness)  
## AIC: 24822  
##   
## Number of Fisher Scoring iterations: 7

voter.clean.var.final %>%  
 select(-vh12g) ->  
 test.final  
  
voter.logit.test=predict(voter.logit.final, test.final, type="response")  
  
voter.logit.test[1:10]

## 1 2 3 4 5 6   
## 0.530882160 0.005638879 0.012239701 0.014889362 0.452015508 0.105136605   
## 7 8 9 10   
## 0.016802072 0.170222264 0.186934953 0.274374421

voter.pred.test = ifelse(voter.logit.test>0.4, 1,0)  
length(voter.pred.test)

## [1] 50000

sum(voter.pred.test, na.rm = TRUE)

## [1] 15060

**This is the final prediction for 2014 assuming that we do not have 2012 general voting in our test data set and supposing this gives the result for 2014 general election.**

## Final output csv file:

voter %>%  
 select(optimus\_id, age, vh14p, vh12g) ->  
 voter.final.output  
  
str(voter.final.output)

## 'data.frame': 50000 obs. of 4 variables:  
## $ optimus\_id: int 861681 1084850 644435 57683 167371 974034 660415 313964 720804 547190 ...  
## $ age : num 69 20 28 78 68 69 53 36 53 30 ...  
## $ vh14p : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ vh12g : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...

voter.final.output <- cbind(voter.final.output, "vote" = voter.pred.test, "prob" = voter.logit.test)  
head(voter.final.output)

## optimus\_id age vh14p vh12g vote prob  
## 1 861681 69 0 0 1 0.530882160  
## 2 1084850 20 0 0 0 0.005638879  
## 3 644435 28 0 0 0 0.012239701  
## 4 57683 78 0 0 0 0.014889362  
## 5 167371 68 0 0 1 0.452015508  
## 6 974034 69 0 0 0 0.105136605