Optimus\_Project

Karan Dassi

4/19/2019

##Loading Libraries and Dataset:

suppressPackageStartupMessages(library(tidyverse))  
suppressPackageStartupMessages(library(ggplot2))  
  
  
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 ...

voter.result <- read.csv("optimus\_output.csv")  
str(voter.result)

## 'data.frame': 50000 obs. of 6 variables:  
## $ optimus\_id: int 861681 1084850 644435 57683 167371 974034 660415 313964 720804 547190 ...  
## $ age : int 69 20 28 78 68 69 53 36 53 30 ...  
## $ vh14p : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ vh12g : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ vote : int 1 0 0 0 1 0 0 0 0 0 ...  
## $ prob : num 0.53088 0.00564 0.01224 0.01489 0.45202 ...

##Joining the two Data Frames:

voter.combine <- cbind(voter, "vh14g" = voter.result$vote, "vh14prob" = voter.result$prob)

##Visualize my resultd compared to previous years general voting:

voter.combine %>%  
 summarise("Votes\_2000" = sum(vh00g, na.rm = T),"Votes\_2002" = sum(vh14g, na.rm = T), "Votes\_2004" = sum(vh04g, na.rm = T),"Votes\_2006" = sum(vh06g, na.rm = T), "Votes\_2008" = sum(vh08g, na.rm = T),"Votes\_2010" = sum(vh10g, na.rm = T),"Votes\_2012" = sum(vh12g, na.rm = T), "Votes\_2014" = sum(vh14g, na.rm = T)) %>%  
 gather(Votes\_2000:Votes\_2014, key = Year, value = Sum)->  
 votes.sum  
  
votes.sum %>%  
ggplot(mapping = aes(x = Year, y = Sum, fill = Year)) +  
 geom\_col() +  
 theme\_bw() +  
 ylab("Sum of total votes")

