US Presidential Election – Data Analysis & Visualization

Markus Stellwag

Contents

1. INTRODUCTION

In this Project we are going to analyse the U.S.A. presidential elections. In particular, we will focus on the presidential election in years 2008, 2012 and 2016, in order to provide some advice to the next Democratic candidate to let him/her get the most efficient electoral campaign.

The U.S.A. presidential election have a complex system, so we are going to explain how to become a President to have a better understanding of what we are talking about.

Lots of people dream of becoming President of the United States. But to officially run for office, a person needs to meet three basic requirements established by the U.S. Constitution (Article 2, Section 1).

A Presidential candidate must be:

- A natural born citizen (U.S. citizen from birth)
- At least 35 years old and
- A U.S. resident (permanently lives in the U.S.) for at least 14 years.

People with similar ideas usually belong to the same political party. The two main parties in the U.S. are Republican and Democrat.

Many people want to be President. They campaign around the country and compete to try to win their party's nomination.

In caucuses (candidate selection meeting), party members meet, discuss, and vote for who they think would be the best party candidate.

In primaries, party members vote in a state election for the candidate they want to represent them in the general election.

After the primaries and caucuses, each major party, Democrat and Republican, holds a national convention to select a Presidential nominee.

The party's Presidential nominee announces his or her choice for Vice President.

The Presidential candidates campaign throughout the country to win the support of the general population.

On election day, people in every state cast their vote.

When people cast their vote, they are actually voting for a group of people called electors.

The number of electors each state gets is equal to its total number of Senators and Representatives in Congress. A total of 538 electors form the Electoral College.

Each elector casts one vote following the general election. The candidate who gets 270 votes or more wins.

The newly elected President and Vice President are then inaugurated on January 20th.
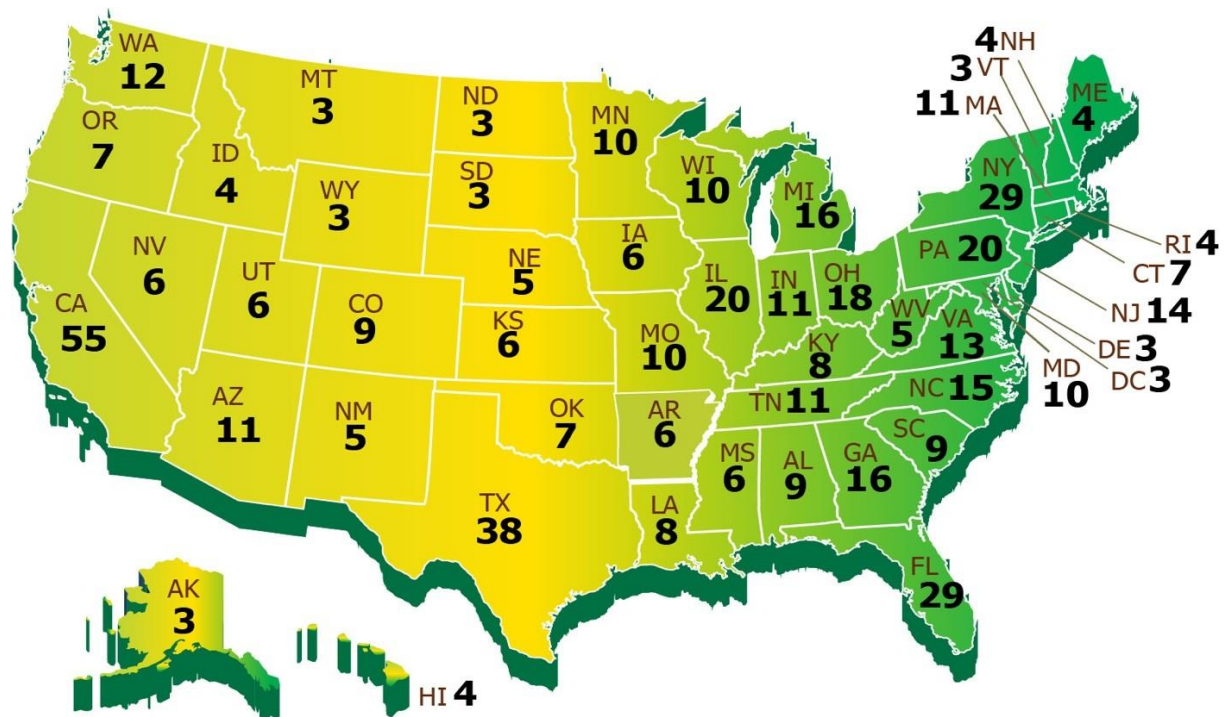
An election for President of the United States occurs every four years on Election Day, held the first Tuesday after the first Monday in November. The next Presidential election will be

held on November 3, 2020. In the event no candidate receives the majority, the House of Representatives chooses the President and the Senate chooses the Vice President.

Unlike in other U.S. elections, the President and Vice President are not elected directly by the people. Instead, they're chosen by "electors" through a process called the Electoral College.

The idea of using electors comes from the Constitution. The nation's founders saw it as a compromise between electing the President by a popular vote among citizens and electing the President in Congress.

The number of electors each state gets is determined by how many members of Congress (House and Senate) the state has. Including Washington, D.C.'s three electors, there are a total of 538 electors in all. U.S. territory residents don't vote in the Presidential election and are not represented in the Electoral College. View the distribution of electors by state.

WA 12
MT 3
ND 3
MN 10
WI 10
MI 16
4 NH
3 VT
11 MA
ME 4
OR 7
ID 4
WY 3
SD 3
IA 6
NY 29
RI 4
NV 6
UT 6
CO 9
NE 5
IL 20
IN 11
OH 18
PA 20
CT 7
NJ 14
CA 55
KS 6
MO 10
KY 8
WV 5
VA 13
DE 3
MD 10
DC 3
AZ 11
NM 5
OK 7
AR 6
TN 11
NC 15
SC 9
MS 6
AL 9
GA 16
TX 38
LA 8
FL 29
AK 3
HI 4

Each state's political parties choose their own slate of potential electors. After you cast your ballot for President, your vote goes to a state-wide tally. In 48 states and Washington, D.C., the winner gets all of the electoral votes for that state. This means his or her party's electors in that state will vote in the Electoral College. Maine and Nebraska assign their electors using a proportional system called the Congressional District Method.

A candidate needs the vote of at least 270 electors—more than half—to win the Presidential election

Winning the Popular Vote but Losing the Election

Though uncommon, it is possible to win the Electoral College, but lose the popular vote. That means that a candidate can win a combination of states and reach the 270 electors mark without winning the majority of votes across the country. This has happened five times in American elections, most recently in 2016.

What Happens if No Candidate Gets 270 Electoral Votes?

In the rare event that no candidate gets the needed 270 electoral votes, the decision would go to the House of Representatives, who would vote to elect the new President from among the top three candidates. A similar process would take place in the Senate to elect the Vice President from among the top two candidates. The only time this has happened was during the 1824 election when John Quincy Adams received the most votes in the House of Representatives after no candidate won a majority of the Electoral College.

Once we know the electoral system, we now decided to analyse the data coming from the previous elections (2008,2012 and 2016). Also we considered some extra data that we found interesting regarding the Reported Voting and Registration of the Voting-Age Population, by Sex, Race and Hispanic Origin, for States in year 2008, 2012 and 2016 in consideration of that we will have to suggest to the new candidate a clear path to follow for the next presidential election.

2. THE PROBLEM ANALYSIS

To give an idea to our Democrat candidate, we have to analyse the problem as best as we can, and to do so, we started taking the data files from the government website and Wikipedia.

Once we have got all the data we wanted to work on, we started to think about what the best approach was. To define that, we needed to have a better idea of what the data were referring to, so as first step, we started opening the data and checking them, visualizing, trying to find some indicating factors that could give us a clear figure of what we are dealing with.

Firstly, we decided to have a wide look at the data, so we started checking how many votes each candidate did obtain and which states they got in the past elections. We saw that the major two candidates in 2012 have been Obama and Romney, this factor gave us the idea of checking also the data from 2008, where Obama was candidate too.

Then we compared these votes by candidates and marked out some differences.

The following step has been to check how many people voted in total each year, cause, knowing that, we have an idea how many people are involved in the election program and how many have been left out or not completely convinced by the candidates to express their vote.

After that we calculated how many electoral colleges each candidate won, knowing that the electoral colleges are the fundamental ones in order to be elected and that we will have to reach at least the number of 270 in order to win.

The following step was to check how many votes each candidate received in the elections in 2012 and visualize the difference.

Same thing we did for the election in 2016, but, in this case, the candidates were Clinton and Trump. Once we compared the votes, saying that, we mean the number of people that voted each candidate, we realized that the votes that Trump received were less, in total number, of the ones that Clinton received, but, as we know, this can happen. So, from that, we decided to go deeper in the analysis and try to get some more figures to determine what a better strategy for the Democrat candidate is.

We started checking the states that each candidate won in each election to see what the tendency of each state has been, to check if it is the same or changed in order to see why a change occurred and in which places. In this way we can visualise properly and discuss about where to focus our future campaign. We will refer to the states that have a thin difference or changed their opinion during the years as swing states and will analyse them with particular attention, giving as candidate base state, the ones that have been stable during the the last three elections.

As a final analysis we decided to have a look on difference in vote per races, gender, family income, community size to check differences and tendencies.

3. DATA ANALYSIS

Now we are going to analyse the data that we have. We decided to take into consideration, as extra Data, the results of the elections of 2008, the data of the votes referred to the different races, divided by state and a data referring to the 2016 Presidential vote by demographic group.

The decision of taking the data file of 2008 has been made because the democrats candidate was the same that won the elections in 2012 (Barack Obama). Concerning the file regarding the races, we decided that was interesting to have a better idea of what differences regarding the reflexes of voting at the democrats' candidate change and to check what the reaction has been after 2012. The last data file seemed necessary in order to explain the uncommon results we had in last elections of 2016.

3.1 CHECKING AND PREPARING THE DATA

This project will include different functions, plots and maps. To have the possibility of using all the tools we need, we will install and load some packages. The packages that we are going to use are: "readr", "dplyr", "ggpubr", "ggplot2", "lattice", "maps", "gridExtra", "webshot", "plotly". The first step is to load the libraries.

```r
library("readr")
library("dplyr")
library("ggpubr")
library("ggplot2")
library("lattice")
library("maps")
library("gridExtra")
library("webshot")
library("plotly")
```

Here we are dealing with files .csv that separates the values with a ";". In order to load them successfully in R, we use the function read.csv() that contains the arguments header and fill set as TRUE by default. For the separation of the variables we use the argument sep=";", to take out the decimals we use dec=",". The files we are going to use are files coming from the election result of 2008, 2012 and 2016. We also have a data file that contains the states names and their respectively acronyms. So what we do now is to load the files calling them "Ele2008" for the data file that contains the data from election of 2008. "Ele2012" for the election of 2012 and "Ele2016" for the ones of 2016. Finally, we call the file "state" implying the states information.

```r
Ele2008<-read.csv("USA_2008_election.csv",header = TRUE, sep = ";", dec = ",")
Ele2012<-read.csv("data_presidential_elections_2012.csv",header = TRUE, sep = ";", dec = ",")
Ele2016<-read.csv("data_presidential_elections_2016.csv",header = TRUE, sep = ";", dec = ",")
states=read.csv("states.csv", header=TRUE, sep = ";")
```

Now we have a look at the data frame to see if all the characteristics are the ones we needed. To do so we use the function str() that allows us to see the data frame characteristics, then

nrow() to check the numbers of rows present in the file and if they are corresponding to the numbers that we are expecting. Besides, we use ncol() to do the same, but for the columns, head() and tail() to check the top and the bottom of our files and see if there is any part of them that does not make sense. Then we used the unique() function in order to check how many candidates we have and which ones they are, for that we referred just to the column where our candidates' names are.

3.1.a. CHECK ON 2008

str(Ele2008)

```
## 'data.frame':    102 obs. of  6 variables:
##  $ Acronym          : Factor w/ 51 levels "AK","AL","AR",..: 2 2 1 1
4 4 3 3 5 5 ...
##  $ State.or.District : Factor w/ 51 levels "Alabama","Alaska",..: 1 1
2 2 3 4 4 5 5 ...
##  $ Candidate        : Factor w/ 2 levels "Mc Cain","Obama": 2 1 2 1 2
1 2 1 2 1 ...
##  $ People.Votes     : int  813479 1266546 123594 193841 1034707 12301
11 422310 638017 8274473 5011781 ...
##  $ Percentage.of.Votes: Factor w/ 39 levels "27%","30%","33%",..: 9 30
8 29 15 24 9 29 31 7 ...
##  $ Electoral.Votes  : int  0 9 0 3 0 10 0 6 55 0 ...
```

nrow(Ele2008)

```
## [1] 102
```

ncol(Ele2008)

```
## [1] 6
```

head(Ele2008)

```
##    Acronym Stat.or.Dist Candidate  People.Vot  Perc.of.Vot  Electoral.Votes
## 1     AL     Alabama     Obama      813479       39%            0
## 2     AL     Alabama     Mc Cain    1266546      60%            9
## 3     AK     Alaska      Obama      123594       38%            0
## 4     AK     Alaska      Mc Cain    193841       59%            3
## 5     AZ     Arizona     Obama      1034707      45%            0
## 6     AZ     Arizona     Mc Cain    1230111      54%            10
```

tail(Ele2008)

```
##     Acronym Stat.or.Dist Candidate People.Vot Perc.of.Vot  Electoral.Votes
## 97     WV    West Virginia  Obama     303857     43%            0
## 98     WV    West Virginia  Mc Cain   397466     56%            5
## 99     WI    Wisconsin      Obama     1677211    56%            10
## 100    WI    Wisconsin      Mc Cain   1262393    42%            0
## 101    WY    Wyoming        Obama     82868      33%            0
## 102    WY    Wyoming        Mc Cain   164958     65%            3
```

unique(Ele2008$Candidate)

```
## [1] Obama    Mc Cain
## Levels: Mc Cain Obama
```

3.1.b   CHECK ON 2012

str(Ele2012)

```
## 'data.frame':    204 obs. of  5 variables:
## $ State.or.district   : Factor w/ 51 levels "AK","AL","AR",..: 2 1 4 3
5 6 7 9 8 10 ...
## $ Candidate           : Factor w/ 4 levels "Barack Obama",..: 1 1 1 1
1 1 1 1 1 1 ...
## $ People.Votes        : int   795696 122640 1025232 394409 7854285 1323
102 905083 242584 267070 4237756 ...
## $ Percentatge.of.Votes: Factor w/ 179 levels "0,00%","0,02%",..: 95 10
3 114 89 163 132 156 159 179 128 ...
## $ Electoral.Votes     : int   0 0 0 0 55 9 7 3 3 29 ...
```

nrow(Ele2012)

```
## [1] 204
```

ncol(Ele2012)

```
## [1] 5
```

head(Ele2012)

##	Stat.or.dist	Candidate	People.Vot	Perc.of.Vot	Electoral.Votes
## 1	AL	Barack Obama	795696	38,36%	0
## 2	AK	Barack Obama	122640	40,81%	0
## 3	AZ	Barack Obama	1025232	44,59%	0
## 4	AR	Barack Obama	394409	36,88%	0
## 5	CA	Barack Obama	7854285	60,24%	55
## 6	CO	Barack Obama	1323102	51,49%	9

tail(Ele2012)

##	Stat.or.dist	Candidate	People.Vot	Perc.of.Vot	Electoral.Votes
## 199	VT	Jill Stein	594	0,20%	0
## 200	VA	Jill Stein	8627	0,22%	0
## 201	WA	Jill Stein	20928	0,67%	0
## 202	WV	Jill Stein	4406	0,66%	0
## 203	WI	Jill Stein	7665	0,25%	0
## 204	WY	Jill Stein	0	0,00%	0

unique(Ele2012$Candidate)

```
## [1] Barack Obama Mitt Romney  Gary Johnson Jill  Stein
## Levels: Barack Obama Gary Johnson Jill Stein Mitt Romney
```

3.1.c   CHECKING 2016

```
str(Ele2016)

## 'data.frame':      224  obs. of  5 variables:
## $ State.or.district   : Factor w/ 56 levels "AK","AL","AR",..: 2 1 4 3
5 6 7 9 8 10 ...
## $ Candidate            : Factor w/ 4 levels "Donald Trump",..: 3 3 3 3
3 3 3 3 3 3 ...
## $ People.Votes         : int   729547 116454 1161167 380494 8753788 1338
870 897572 235603 282830 4504975 ...
## $ Percentatge.of.Votes: Factor w/ 211 levels "","0,19%","0,26%",..: 96
102 136 93 202 153 180 176 211 149 ...
## $ Electoral.Votes      : int   0 0 0 0 55 9 7 3 3 0 ...

nrow(Ele2016)

## [1] 224

ncol(Ele2016)

## [1] 5

head(Ele2016)

##     Stat.or.dist     Candidate     People.Vot Perc.of.Vot Electoral.Votes
## 1    AL          Hillary  Clinton    729547    34,36%         0
## 2    AK          Hillary  Clinton    116454    36,55%         0
## 3    AZ          Hillary  Clinton    1161167   45,13%         0
## 4    AR          Hillary  Clinton    380494    33,65%         0
## 5    CA          Hillary  Clinton    8753788   61,73%        55
## 6    CO          Hillary  Clinton    1338870   48,16%         9


tail(Ele2016)

##       Stat.or.dist  Candidate    People.Vot Perc.of.Vot Electoral.Votes
## 219   VT          Jill Stein      6758      2,11%          0
## 220   VA          Jill Stein     27638      0,69%          0
## 221   WA          Jill Stein     58417      1,82%          0
## 222   WV          Jill Stein      8075      1,13%          0
## 223   WI          Jill Stein     31072      1,04%          0
## 224   WY          Jill Stein      2515      0,97%          0

unique(Ele2016$Candidate)

## [1] Hillary Clinton  Donald  Trump     Gary Johnson     Jill  Stein
## Levels: Donald Trump Gary Johnson Hillary Clinton Jill Stein
```

From reading the characteristic of the files and the head and tail, we noticed that the % sign was present in one of the column of the file, that will create problems in using that specific column, so to avoid that we removed that symbol. To do so, we used the function sub() saying to switch the symbol with a blank space. After that we replaced the comma present in the numbers decimal division with a dot, since we could face problems otherwise. At the same time, we gave also a numeric characteristic to the column of the Percentage of Votes to be

able to use it as numeric and operate with them. Finally, we checked the structure in order to have everything as we want.

REMOVING %TAGS FROM COLUMNS 2008 - 2016

2008
```
rem.pe08 <- sub("%", "", as.character(Ele2008$Percentage.of.Votes))
rem.pe08 <- as.numeric(sub(",", ".", rem.pe08))
Ele2008$Percentage.of.Votes=rem.pe08
```

2012
```
rem.pe12 <- sub("%", "", as.character(Ele2012$Percentatge.of.Votes))
rem.pe12 <- as.numeric(sub(",", ".", rem.pe12))
Ele2012$Percentatge.of.Votes=rem.pe12
```

2016
```
rem.pe16 <- sub("%", "", as.character(Ele2016$Percentatge.of.Votes))
rem.pe16 <- as.numeric(sub(",", ".", rem.pe16))
Ele2016$Percentatge.of.Votes=rem.pe16
```

REPLACING NAs WITH 0

Another thing we noticed has been that in the Ele2016 file there were present some NA values that needed to be replaced by 0 in order to use the file properly, so do we with the following code.

```
Ele2016[is.na(Ele2016)] = 0
```

The file from 2008 was already prepared with the states names and their Acronyms, as we see, but that is not true for 2012 and 2016 files. So here we are going to rename the column of the dataframes "Ele2012" and "Ele2016" that was called "State.or.disctrict" as "Acronym". We did that in order to have the same name for the column that is also present in the "states" file because in this way we can merge the states file without having issues. So the next step is to merge the states file with the files "Ele2012" and "Ele2016". We do that to be able to compare data, and have a better idea of what the numbers are referring to in our analysis. Last step for getting our files comparable was calling the columns of the files with the same names, as we said, the file "Ele2008" did not required any consideration as it was already in this way. Our new columns will have the following names: "Acronyms" for the acronyms of the states, "State.or.District", for the name of the state or the district, "Candidate" for the candidate's name, "People.Votes" for the numbers of people votes that the candidates obtained, "Percentage.of.Votes" for the percentage of votes that each candidate obtained, "Electoral.Votes" for the electoral votes each candidate received per state.

For 2012 file:

```
Ele2012 <- rename(Ele2012, Acronym = State.or.district)
Ele2012 <- merge(Ele2012, states, by="Acronym")
Ele2012 <- Ele2012[c("Acronym", "State.or.District", "Candidate", "People.
Votes","Percentatge.of.Votes",
                    "Electoral.Votes")]
```

Same for 2016:

```
now states <> Ele2016
Ele2016 <- rename(Ele2016, Acronym = State.or.district)
Ele2016 <- merge(Ele2016, states, by="Acronym")
Ele2016 <- Ele2016[c("Acronym", "State.or.District", "Candidate", "People.
Votes","Percentatge.of.Votes",
                     "Electoral.Votes")]
```

3.2 Visualization of election results

As first approach we decided to have a glance to what the situation of the percentages of votes by candidate is. Here we analyzed the data frame of "Ele2008", we identified two major candidates, Mc Cain for the Republicans and Obama for the Democrats. Furthermore, we also took into consideration the data from "Ele2012" and Ele2016, were the candidates were 4 in those cases. We analyzed the percentages of the American Election System, because as we know, a candidate could win the election without having the popular majority. In this way we just avoid that problem and visualize who between the two candidates received more votes in percentage. Our first approach was to change the order of the candidates. For that we used the function factor to change the characteristic of the variable "Candidate". Hence, we can change the order by calling the argument levels and put in the desired order. To create the plot we used the package "ggplot", in particular the function ggboxplot. This function allows us to obtain a plot straight from the file. We insert the dataset and located on the "x-axis" the Candidate column and on the "y-axis" the Percentage of votes column. Then we allocate different colors to visualize the candidates' values and defined the type of color we wanted. Finally, we defined the labels and decided to insert an intercept at the value 50 in order to make the comprehension of the graph easier.

Comparison Percentage of Votes by Candidate – Boxplot and Pie Chart

3.2.a Elections of 2008

In the following boxplot, we see the mean of percentage of votes by candidate

```
Ele2008$Candidate<- factor(Ele2008$Candidate, levels = c("Obama", "Mc Cain
"))
ggboxplot(Ele2008, x = "Candidate", y = "Percentage.of.Votes",
          color = "Candidate", palette = c("navy", "red2"),
          ylab = "Percentage of votes", xlab = "Candidates")+
geom_hline(yintercept = 50)
```
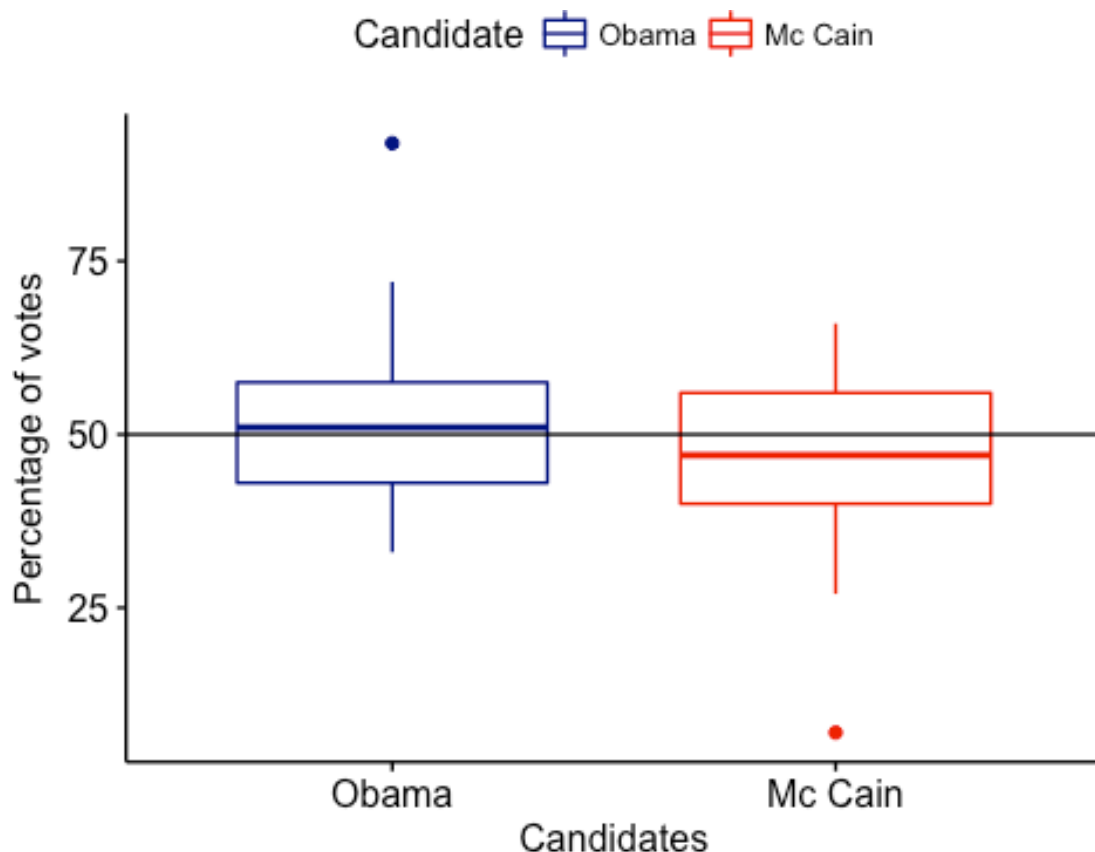
In the following part we want to check the total amount of people that voted each candidate. In this way we determine how many votes each candidate actually received. We used the function filter() in the data "Ele2008" in order to get all the amount of votes that Obama received per state. After that we did the sum of the column containing the people votes. We used the same approach for the other candidate, in this case, Mc Cain.

```
obama08.all  =  filter(Ele2008,    Candidate == "Obama")
obama08.sum = sum(obama08.all$People.Votes)

mccain.all  =  filter(Ele2008,    Candidate == "Mc Cain")
mccain.sum = sum(mccain.all$People.Votes)
```

The next step was to check how many electoral votes each candidate received. To do so, we filtered the data "Ele2008" by candidate and used a pipeline to a sequential command, filter in this case, asking to take in consideration just the values greater than zero. In this way we have a clear idea of how many electoral votes each candidate received.

```
obama08  =  filter(Ele2008,    Candidate == "Obama") %>% filter(Electoral.Vote
s > 0)
mccain  =  filter(Ele2008,    Candidate == "Mc Cain")  %>% filter(Electoral.Vot
es > 0)
```

At this point we wanted to show the results that we have, and to do that, we decided to create another matrix, to recall the data for plotting. To do that we used the function matrix() and inserted the names of the candidates and the sum of the electoral votes they obtained.

We specified that the matrix will have two rows and two columns, the first column will be the one with the candidates' name and the second one with the total votes candidate number.

creating vectors with name; then creating a matrix; giving columns names
```
model3 = matrix(c("Obama", obama08.sum,
                  "Mc Cain", mccain.sum), nrow = 2, ncol = 2, byrow = TRUE)
colnames(model3) = c("Candidate", "Total.Votes.Candidate")
```

After that we are going to visualize the results obtained. Firstly, we converted the matrix into a dataframe in order to recall it with ggplot() function. We gave a name to the matrix, in this case tot.votes.08, and changed the columns of the total votes per candidate in a numeric column. Then called the plot ppl.plot08. With ggplot we created an histogram, having the candidate's names on the x-axis and the number of votes on the y-axis. The red color is referred to Mc Cain and the blue to Obama.

converting matrix into dataframe in order to visualize with ggplot
```
tot.votes.08 = as.data.frame(model3)
tot.votes.08$Total.Votes.Candidate<-as.numeric(as.character(tot.votes.08$Total.Votes.Candidate))

t <- list(size = 25,color = "black")

export(plot_ly(labels = c("Obama","Mc Cain"), values = ~ tot.votes.08$Total.Votes.Candidate, type = 'pie', textinfo='label+percent',
               insidetextfont = list(color = '#FFFFFF', size=25),
               outsidetextfont=list(size=25), showlegend=TRUE,
               marker = list(colors=c("navy", "red"))) %>%
        layout(title ="POPULAR VOTES IN 2008 OBAMA: 69456897 – MC CAIN:59934814",
               xaxis = list(showgrid = FALSE, zeroline = FALSE, showticklabels = F),
               yaxis = list(showgrid = FALSE, zeroline = FALSE, showticklabels = F)))
```

POPULAR VOTES IN 2008
**OBAMA: 69456897 – MC CAIN: 59934814**
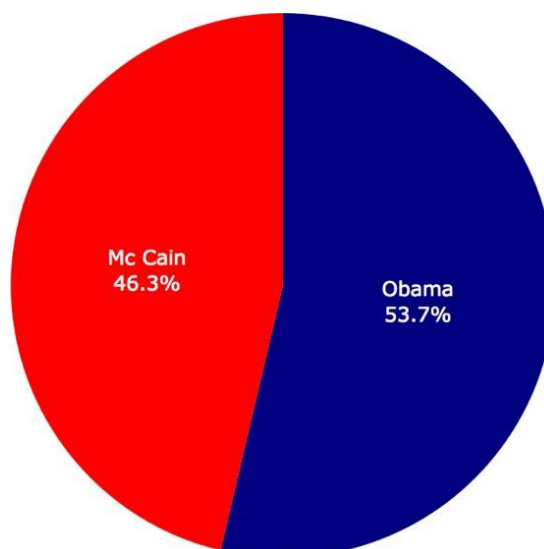
So as we see, Obama gained the majority of popular votes, but let's see what happened in the electoral colleges. To find out what happened we are going to sum all the electoral votes for each candidates, create a matrix as we did before, define the numeric values of the matrix, compile the data frame in order to plot it.

```r
obama08.ele = sum(obama08$Electoral.Votes)
mccain.ele = sum(mccain$Electoral.Votes)

e.colleges.tot08 = matrix(c("Barack Obama", obama08.ele,
"John Mc Cain", mccain.ele), nrow = 2, ncol = 2, byrow = TRUE)
colnames(e.colleges.tot08) = c("Candidate", "Total.electoral.votes")

tot.elvotes.08=as.data.frame(e.colleges.tot08)
tot.elvotes.08$Total.electoral.votes<-as.numeric(as.character(tot.elvotes.
08$Total.electoral.votes))
el.plot08 = ggplot(tot.elvotes.08,
        aes(Candidate, Total.electoral.votes, fill = Candidate)) + scale_y
_continuous(limits = c(0,538))+
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Total electoral votes by candidate - 2008",
        x = "Candidate",
        y = "Votes") + scale_fill_manual(values = c("navy","red2"))
el.plot08
```


Total electoral votes by candidate - 2008
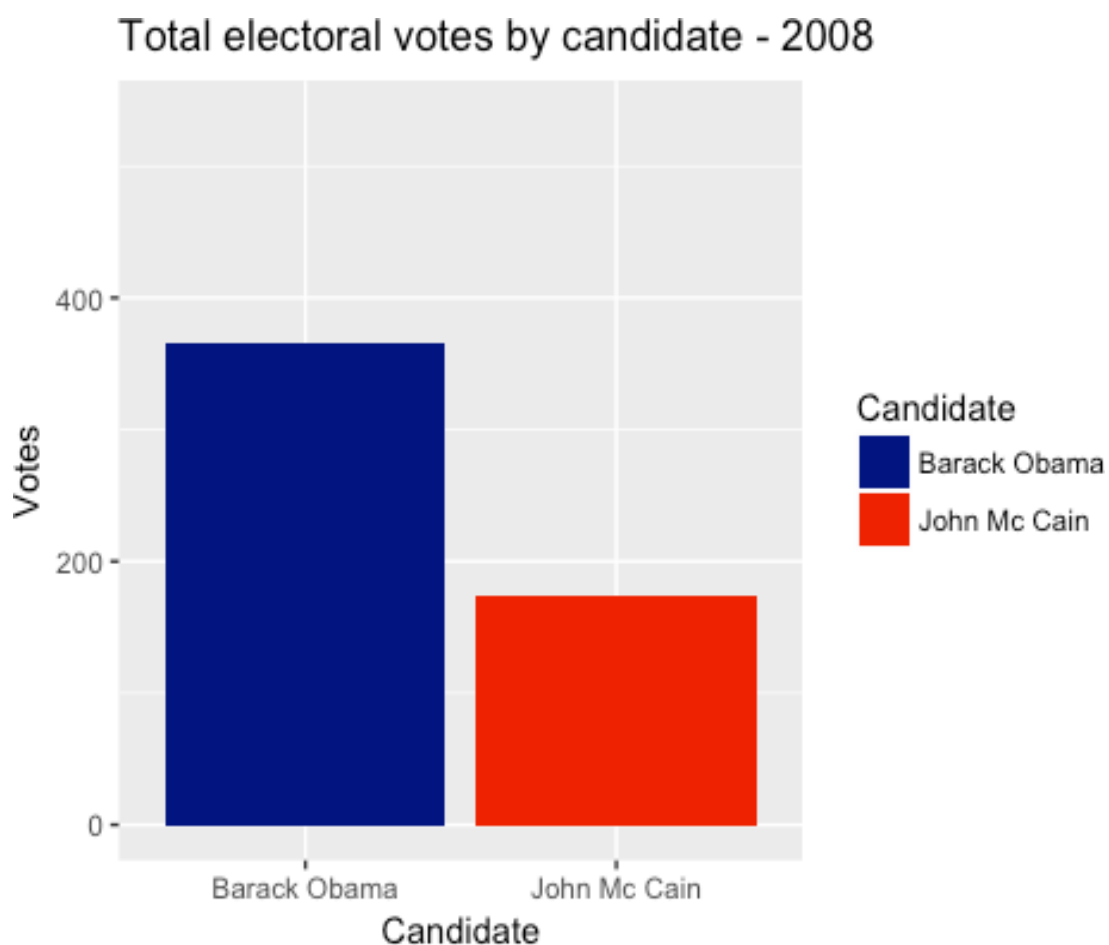
As we can see in this election, the difference in the electoral votes is really high, with Obama getting almost twice of the amount of votes that Mc Cain got. Our first approach was to change the order of the candidates. For that we used the function factor to change the characteristic of the variable "Candidate". Hence, we can change the order by calling the argument levels and put in the desired order of the vectors.

3.2.b Elections of 2012

```
o = "Barack Obama"
r =  "Mitt  Romney"
j  = "Gary Johnson"
s = "Jill   Stein"
Ele2012$Candidate<- factor(Ele2012$Candidate, levels = c(o, r, j,  s))
ggboxplot(Ele2012,x = "Candidate", y = "Percentatge.of.Votes",
          color = "Candidate", palette = c("navy",  "red2","gold","green4"),
           ylab = "Percentage of votes", xlab = "Candidates") + theme(legen
d.position  = c(0.8,0.7))
```
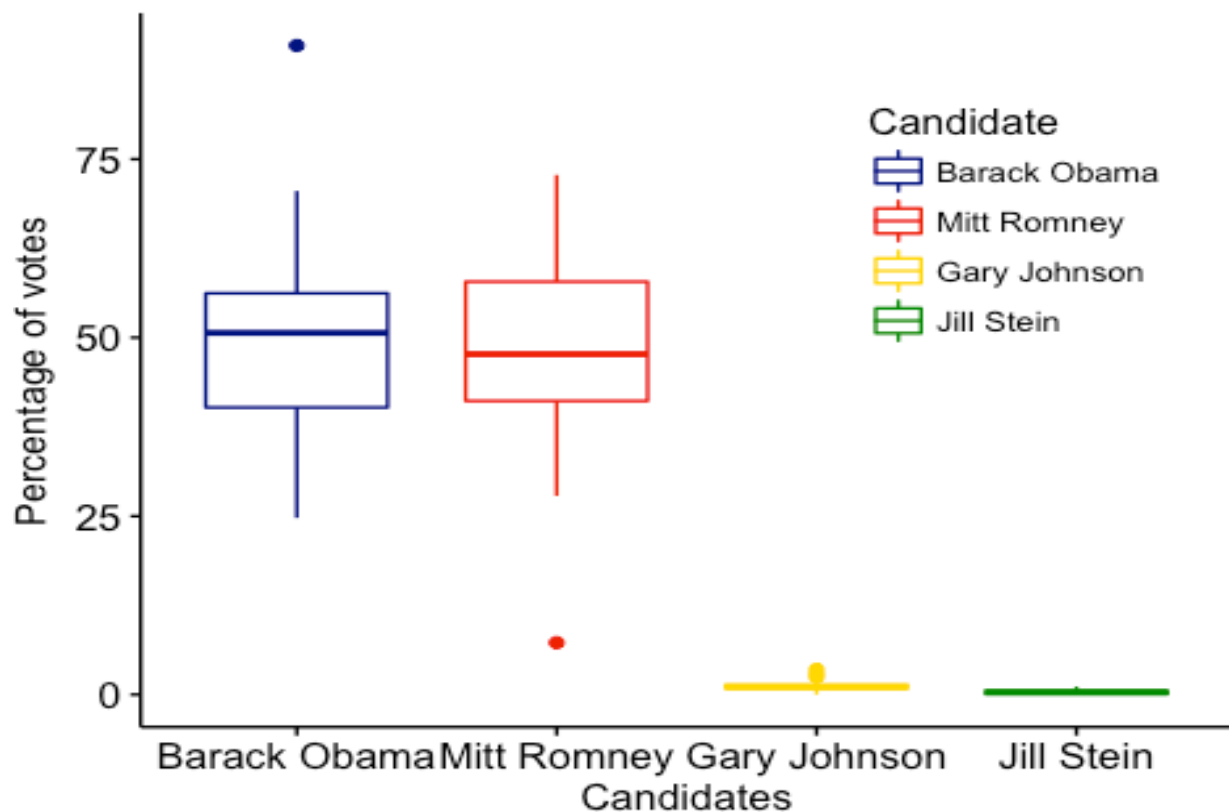
Now we start to analyze the data that we have from the election of 2012. As we know, in this election we had four candidates, so the first thing that we do is to review how many votes these candidates obtained to see which ones are the most relevant. To do it we use the filter() function on the data Ele2012 in order to get each candidate's portion of votes. After that we sum all the values present in the column "People Votes".

```
johnson = filter(Ele2012, Candidate == "Gary Johnson")
johnson.sum = sum(johnson$People.Votes)

stein = filter(Ele2012, Candidate == "Jill Stein")
stein.sum = sum(stein$People.Votes)

obama.all = filter(Ele2012, Candidate == "Barack Obama")
obama.sum = sum(obama.all$People.Votes)

romney.all = filter(Ele2012, Candidate == "Mitt Romney")
romney.sum = sum(romney.all$People.Votes)
```

Then, before visualizing the results, we are going to create a matrix. The matrix contains the vectors (o= Barack Obama, r= Mitt Romney, j= Gary Johnson, s= Jill Stein) containing the votes of the candidates and the respective sum of votes that we calculate previously. We gave also the names to the columns "Candidates" and "Total.Votes.Candidate".

```
o = "Barack Obama"
r = "Mitt Romney"
j = "Gary Johnson"
s = "Jill Stein"

model = matrix(c(o, obama.sum, r, romney.sum, j, johnson.sum, s, stein.sum
), nrow = 4, ncol = 2, byrow = TRUE)
colnames(model) = c("Candidate", "Total.Votes.Candidate")
```

So now we are ready to see the results that they received in 2012. We created the matrix having it as a data frame firstly, in this way we can use it with ggplot() function. Again we have to change the order of the candidates. For that we used the function factor to change the characteristic of the variable "Candidate". Hence, we can change the order by calling the argument levels and put in the desired order of the vectors. Then we called the graph ppl.plot12 and we plot it with the function ggplot() having on the x-axis the candidates' names and on the y-axis total people votes for 2012.

```
tot.votes.12 = as.data.frame(model)
tot.votes.12$Total.Votes.Candidate<-as.numeric(as.character(tot.votes.12$T
otal.Votes.Candidate))
tot.votes.12$Candidate<-factor(tot.votes.12$Candidate, levels = c(o, r, j
, s))
ppl.plot12 = ggplot(tot.votes.12,
        aes(Candidate, Total.Votes.Candidate, fill = Candidate)) +
```

```
   geom_bar(stat = "identity", position = "dodge") + scale_y_continuous(lim
its = c(0,100000000), labels = scales::comma)+
   labs(title = "Total people  votes  by  candidate  - 2012",
        x = "Candidate",
      y = "Total  People  Votes") + scale_fill_manual(values = c("navy","red2"
, "gold", "green4")) + scale_x_discrete(labels = c("Barack Obama" = "B.Oba ma",
"Gary Johnson" = "G. Johnson", "Jill Stein" = "J. Stein", "Mitt Romne y" =
"M.Romney"))
ppl.plot12
```


Total people votes by candidate - 2012
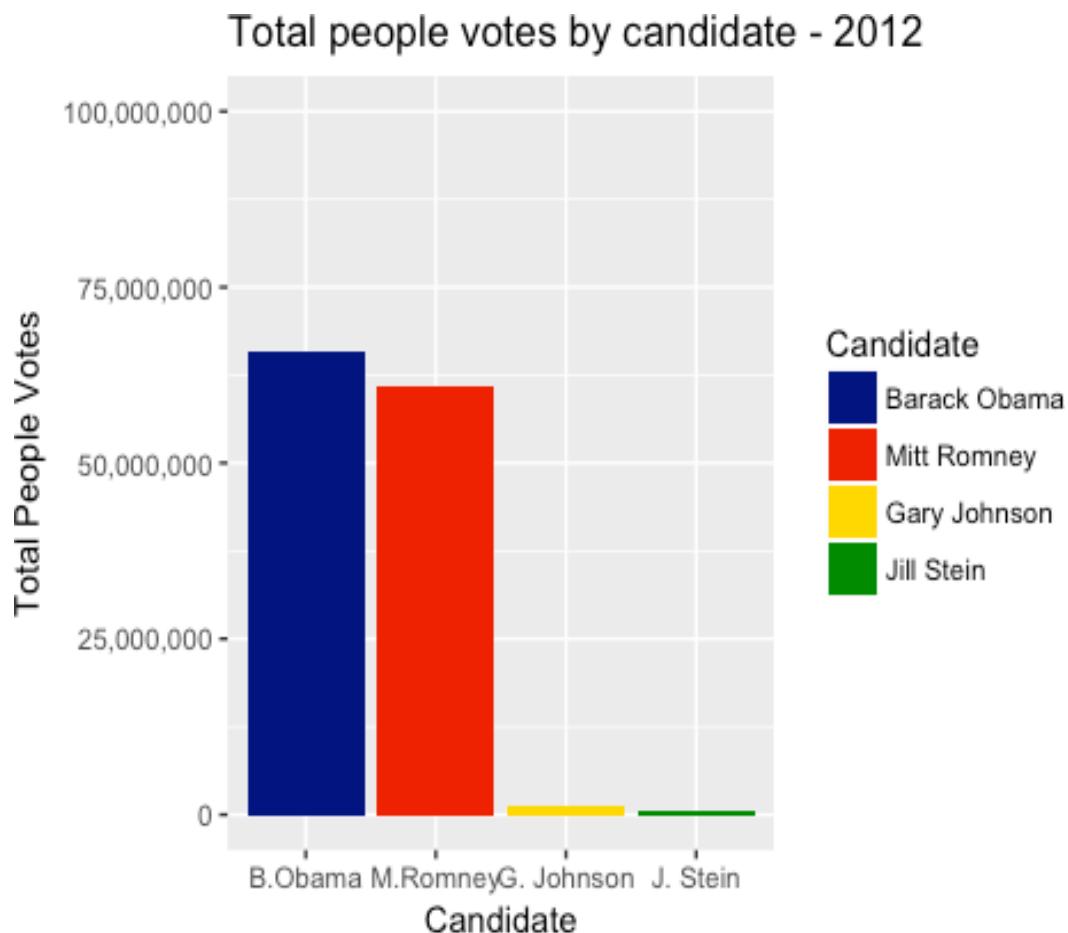
From this histogram, we can see that Barack Obama and Mitt Romney had got more votes, so it does not make sense to take into consideration the other candidates for further analysis. For this reason, we are going to exclude from the file "Ele2012" the data related to the other candidates. We do it creating a vector containing the names of Barack Obama and Mitt Romney, and filter the file "Ele2012".

```
dm.rp12 = c("Barack Obama", "Mitt Romney")
obama.romney = filter(Ele2012, Candidate %in% dm.rp12)
```

Once we determined which ones the most important candidates are, we are going to check how many electoral votes they have got, using the filter that gives us the values of electoral votes greater than zero.

```
obama = filter(Ele2012,   Candidate == "Barack Obama") %>% filter(Electoral
.Votes > 0)
```

```
romney   =   filter(Ele2012,     Candidate == "Mitt Romney") %>% filter(Electoral
.Votes > 0)
```

Then we summed the amount of electoral votes they obtained and created a matrix in order to plot it afterwards. The column of the matrix will be the candidates and the respective sums of votes that gained and their names are: "Candidate" and "Total.electoral.votes". In the plot we set up as y-axis limit the total amount of votes that we can reach, so it is 538. By doing that we called the function of ggplot2 package scale_y_continuous().

```
obama.ele = sum(obama$Electoral.Votes)
romney.ele = sum(romney$Electoral.Votes)
e.colleges.tot = matrix(c(o, obama.ele, r, romney.ele), nrow = 2, ncol = 2
, byrow = TRUE)
colnames(e.colleges.tot) = c("Candidate", "Total.electoral.votes")
```

```
converting matrix into dataframe in order to visualize with ggplot; comparing electoral votes
tot.elvotes.12=as.data.frame(e.colleges.tot)
tot.elvotes.12$Total.electoral.votes<-as.numeric(as.character(tot.elvotes.
12$Total.electoral.votes))
el.plot12 = ggplot(tot.elvotes.12,
        aes(Candidate, Total.electoral.votes, fill = Candidate)) + scale_y_
continuous(limits = c(0,538)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Total electoral votes by candidate - 2012",
       x = "Candidate",
       y = "Votes") + scale_fill_manual(values = c("navy","red2"))
el.plot12
```


Total electoral votes by candidate - 2012
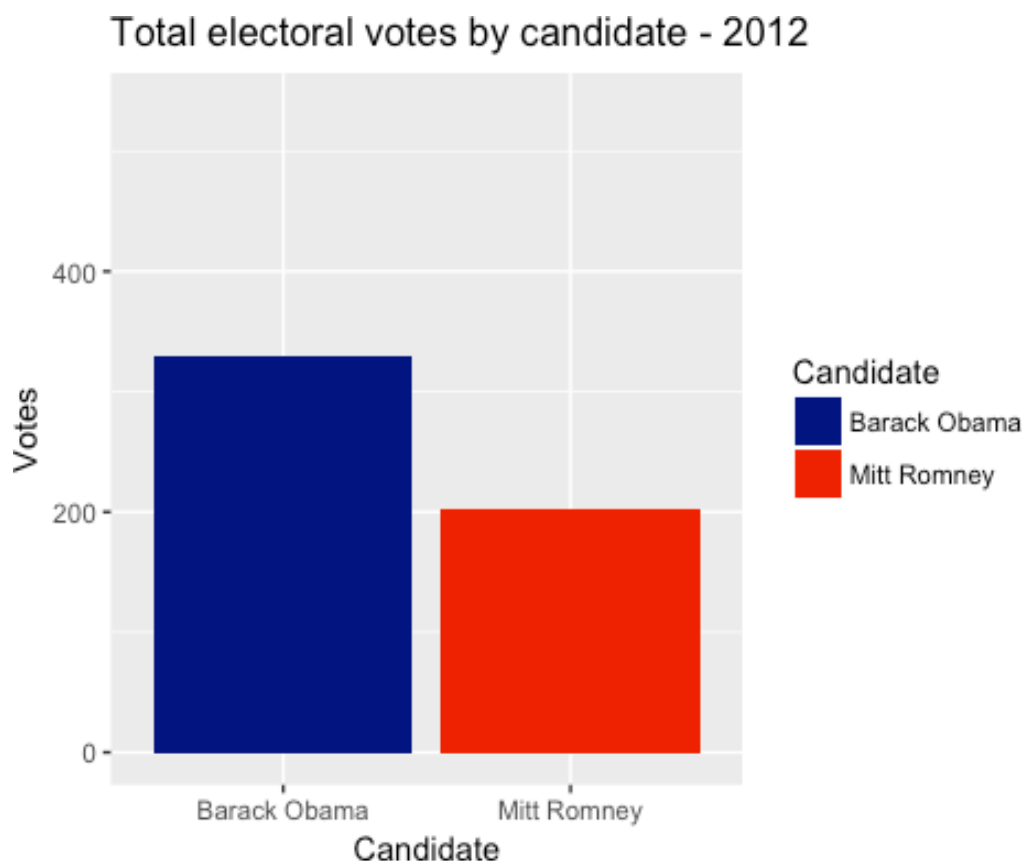
3.2.c Elections of 2016

After having checked what the situation was in 2008 and 2012 now we are going to do the same for 2016. In this year the Democrats candidate changed, from Barack Obama to Hillary Clinton so, we will not have the same that won in the last two elections. Anyway the first step is to check how many votes the four candidates obtained, knowing that the candidates in this year are: Clinton, Stein, Johnson, Trump. As always, we are filtering the data file by candidate and sum the total amount of votes. Then we make the matrix with the results in order to plot them.

```
h =  "Hillary Clinton"
t = "Donald Trump"
j  = "Gary Johnson"
s = "Jill Stein"
Ele2016$Candidate<- factor(Ele2016$Candidate, levels = c(t, h, j, s))
ggboxplot(Ele2016, x = "Candidate", y = "Percentatge.of.Votes",
        color = "Candidate", palette = c("red2", "navy", "gold", "green4"),
          ylab = "Percentage of votes", xlab = "Candidates")+theme(legend.
position = c(0.8,0.7))
```
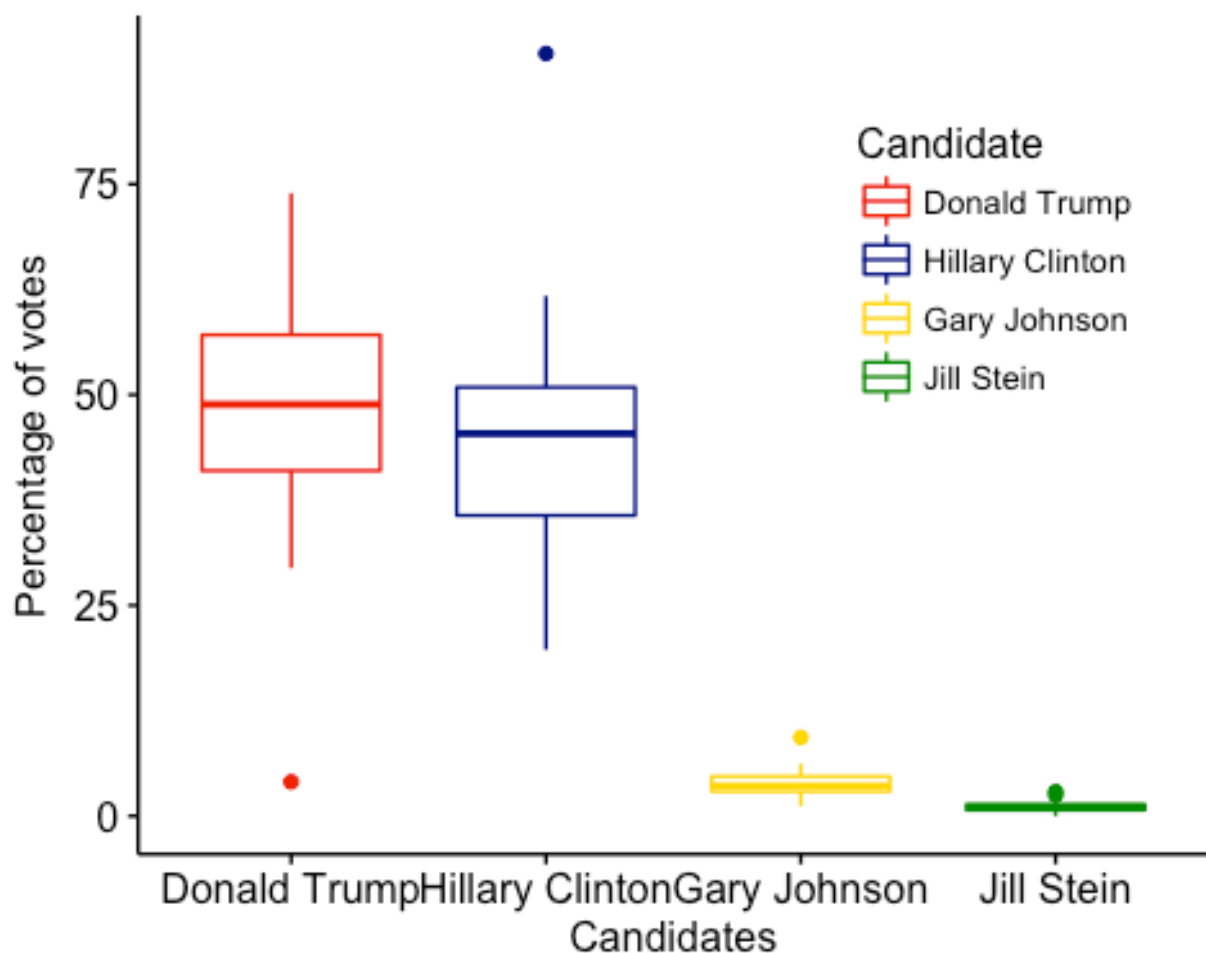
```r
hillary.all   =   filter(Ele2016,    Candidate == "Hillary Clinton")
hillary.sum = sum(hillary.all$People.Votes)

stein16  =  filter(Ele2016,    Candidate == "Jill Stein")
stein16.sum = sum(stein16$People.Votes)

johnson16   =   filter(Ele2016,    Candidate == "Gary Johnson")
johnson16.sum = sum(johnson16$People.Votes)

trump.all   =   filter(Ele2016,    Candidate == "Donald Trump")
trump.sum = sum(trump.all$People.Votes)
```

```r
for 2016 - creating vectors with name; then creating a matrix; giving columns names
h = "Hillary Clinton"
t = "Donald Trump"
j = "Gary Johnson"
s = "Jill Stein"

model2 = matrix(c(h, hillary.sum, t, trump.sum, j, johnson16.sum, s, stein
16.sum), nrow = 4, ncol = 2, byrow = TRUE)
colnames(model2) = c("Candidate", "Total.Votes.Candidate")
```

```r
converting matrix into dataframe in order to visualize with ggplot
tot.votes.16 = as.data.frame(model2)
tot.votes.16$Total.Votes.Candidate<-as.numeric(as.character(tot.votes.16$T
otal.Votes.Candidate))
tot.votes.16$Candidate<- factor(tot.votes.16$Candidate, levels = c(h, t, j
, s))
ppl.plot16 = ggplot(tot.votes.16,
        aes(Candidate, Total.Votes.Candidate, fill = Candidate)) +
  geom_bar(stat = "identity", position = "dodge") + scale_y_continuous(lim
its = c(0,100000000), labels = scales::comma) +
  labs(title = "Total people votes by candidate - 2016",
        x = "Candidate",
        y = "Total People Votes") + scale_fill_manual(values = c("navy","re
d2", "gold", "green4")) + scale_x_discrete(labels = c("Donald Trump" = "D.
Trump", "Gary Johnson" = "G. Johnson", "Hillary Clinton" = "H.Clinton", "J
ill Stein" = "J. Stein"))
ppl.plot16
```

Total people votes by candidate - 2016
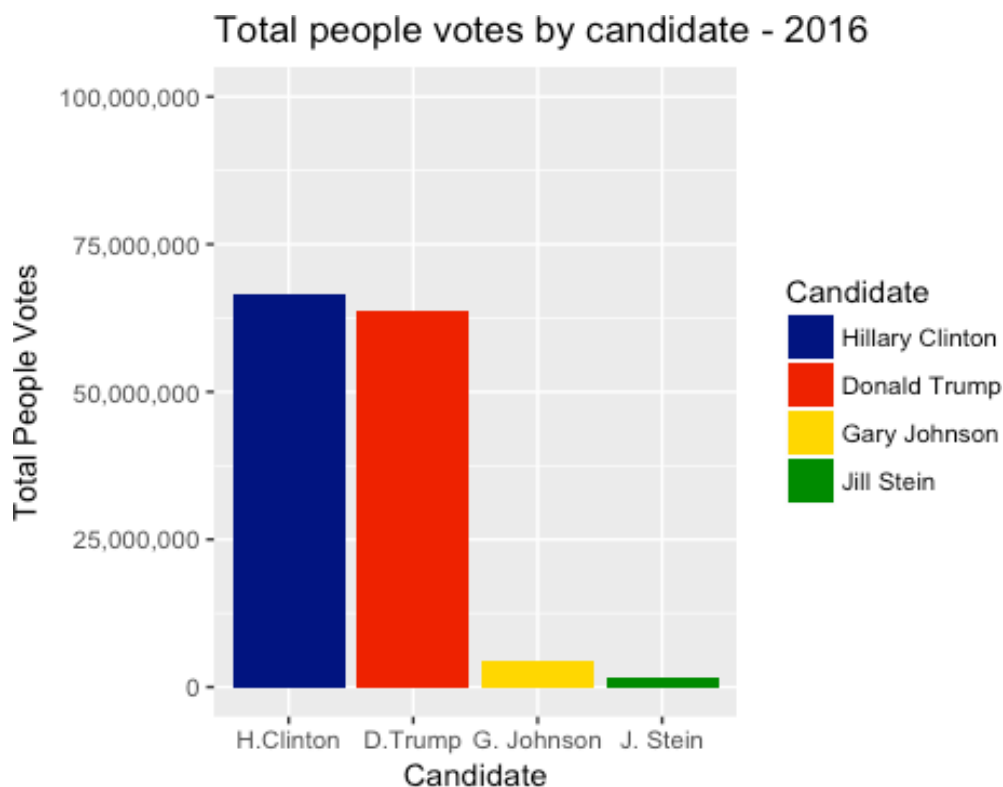
From this plot we can see that the candidate who got more votes has been Hillary Clinton, after her there is Donald Trump, and then there are the last two Johnson and Stein, but noticing the votes that the last two candidates have received, we decided not to consider them for the further analysis.

After we saw what happened in the popular vote, we are going to filter the data file "Ele2016" with the name of candidates that we are interested into. To do so we created a vector containing the candidates' name and use the function filter() adding the command %in% that gives us the possibility to take in consideration just what is presented in the vector that we created.

```
dm.rep16 = c("Hillary Clinton", "Donald Trump")
clinton.trump = filter(Ele2016, Candidate %in% dm.rep16)
```

So now we are ready to analyze the electoral votes, to do so, as we did before, we are going to filter the electoral votes greater than zero and sum the total amount. The name that we gave to the results is "clinton.ele" and "trump.ele". After that we created the matrix containing the candidates and their result and afterwards we plotted it to have a visualization of what the results are.

```
hillary = filter(Ele2016, Candidate == "Hillary Clinton") %>% filter(Electoral.Votes > 0)
clinton.ele = sum(hillary$Electoral.Votes)

trump = filter(Ele2016, Candidate == "Donald Trump") %>% filter(Electoral.Votes > 0)
trump.ele = sum(trump$Electoral.Votes)
```

```
e.colleges.tot16 = matrix(c("Hillary Clinton", clinton.ele, "Donald Trump"
, trump.ele), nrow = 2, ncol = 2, byrow = TRUE)
colnames(e.colleges.tot16) = c("Candidate", "Total.electoral.votes")
```

2016 - converting matrix into dataframe in order to visualize with ggplot; comparing electoral votes

```
tot.elvotes.16 = as.data.frame(e.colleges.tot16)
tot.elvotes.16$Total.electoral.votes<-as.numeric(as.character(tot.elvotes.
16$Total.electoral.votes))
tot.elvotes.16$Candidate <- factor(tot.elvotes.16$Candidate, levels = c("H
illary Clinton","Donald Trump"))
el.plot16 = ggplot(tot.elvotes.16,
        aes(Candidate, Total.electoral.votes, fill = Candidate)) + scale_y_
continuous(limits = c(0,538))+
   geom_bar(stat = "identity", position = "dodge") +
   labs(title = "Total  electoral  votes  by  candidate  -  2016",
        x = "Candidate",
        y = "Votes") + scale_fill_manual(values = c("navy","red2"))
el.plot16
```


Total electoral votes by candidate - 2016
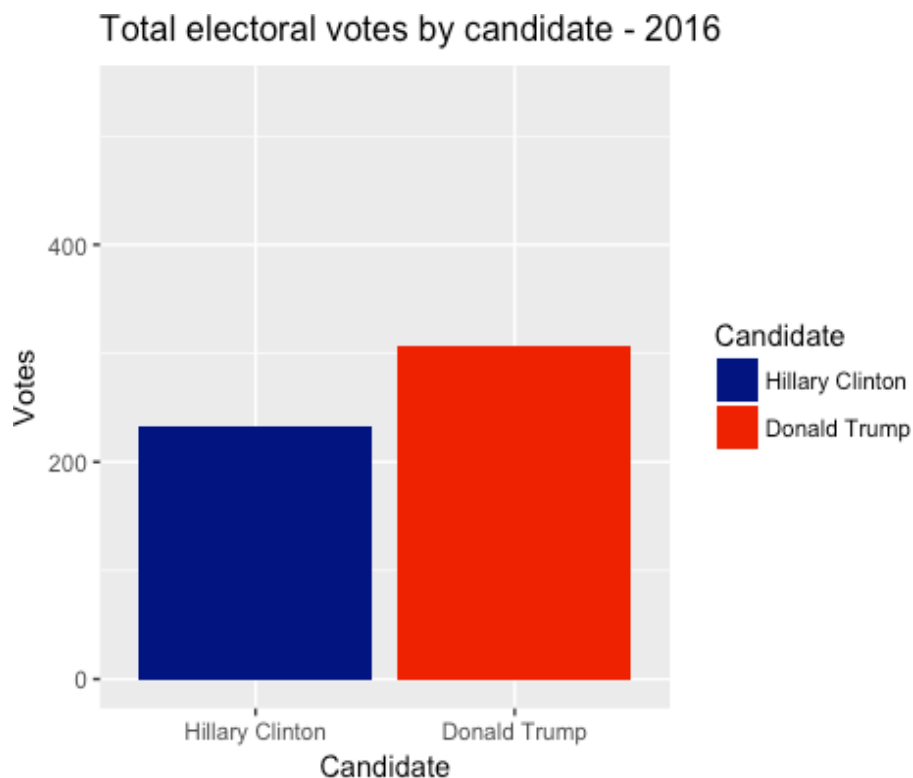
As we can read from this plot, the electoral votes split in a different way compare the popular votes, so here we have that Trump has got more electoral votes then Clinton, which means that Trump won the elections having less popular support than Clinton. This result can occur as we know from the previous information regarding the elections in America. Seeing that we have to understand why this can happen and how our candidate can get a better result either what Democrats have to do in order to improve.

```
grid.arrange(el.plot08, el.plot12, el.plot16, nrow=3, ncol=1)
```

Total electoral votes by candidate - 2008

Total electoral votes by candidate - 2012

Total electoral votes by candidate - 2016
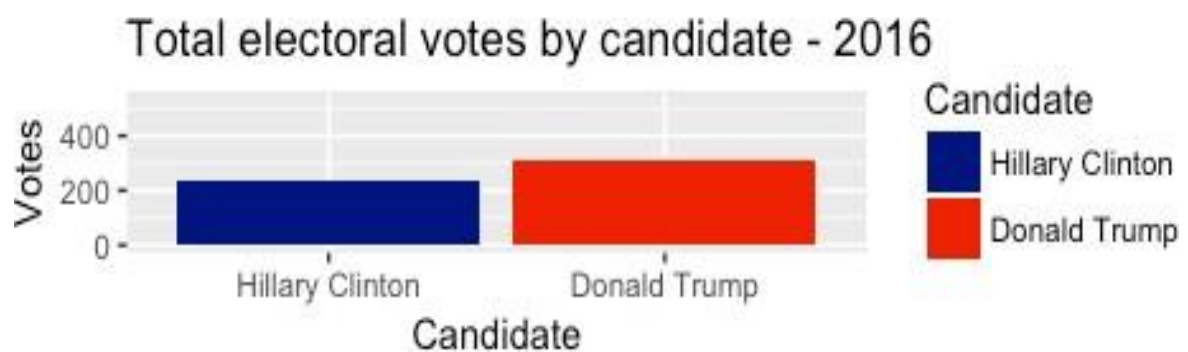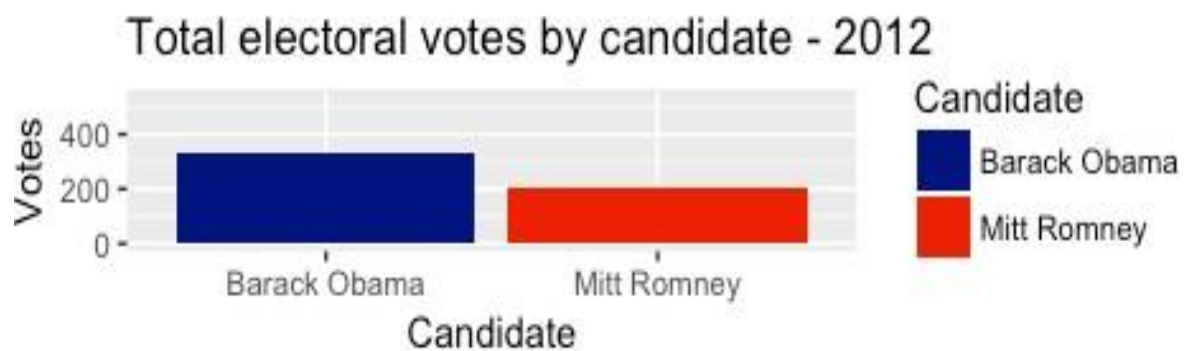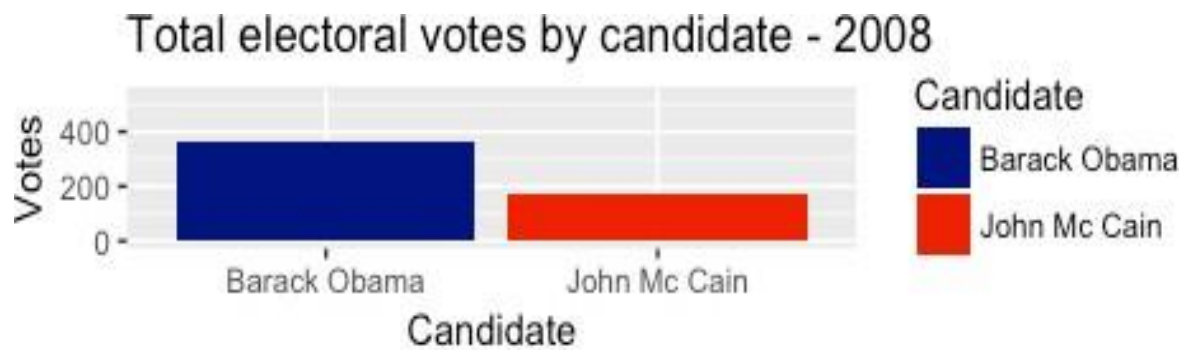
4. DETAILS OF ELECTIONS STATE BY STATE, WITH MAP VISUALIZATION

In this section our objective is to create a colored map of the United States of America (USA) in order to highlight and visualize the election results of the major candidates. After 2008 there will be iterations of this code implementation for the years 2012 and 2016, as we would like to see whether there are key states that are essential for the election. Hence, it was necessary to create two separate vectors (win.obama08 and win.mccain) containing all states which have been won by democrats(Obama) and republicans(Mccain). For the further approach we call at the beginning a blank map of the USA by using "state" for the USA map. At this point we have to remark that Alaska and Hawaii are not included in the maps database "state", accordingly information for either states will be manually included in the conclusion section. In the next step we fit in the created vectors with the winning states of out candidates, and allocate them colors, in this case blue for the democrats and red for the republican party. A descriptive title finalizes the mapping objective.

```r
win.obama08 = c( "California", "Colorado", "Connecticut", "District of Columbia",
         "Delaware", "Florida", "Hawaii", "Indiana", "Iowa",
         "Illinois", "Massachusetts",  "Maryland", "Maine" ,
         "Michigan",  "Minnesota",  "Nebraska",   "New Hampshire", "New Jersey",
         "New Mexico", "Nevada",    "New York", "North Carolina", "Ohio",
         "Oregon",  "Pennsylvania",  "Rhode  Island",  "Virginia",
         "Vermont", "Washington", "Wisconsin")
win.mccain = c("Alaska", "Alabama", "Arkansas", "Arizona", "Georgia", "Idaho",
         "Kansas", "Kentucky", "Louisiana", "Missouri", "Mississippi",
         "Montana", "North Dakota", "Oklahoma", "South Carolina",
         "South Dakota", "Tennessee", "Texas", "Utah", "West Virginia", "Wyoming")
map(database = "state")
map(database = "state", regions = win.obama08, col = "blue", fill=T, add=TRUE)
map(database = "state", regions = win.mccain, col = "red2", fill=T, add=TRUE)
title(main = "WINNING STATES BY CANDIDATE - 2008")
text(x=state.center$x, y=state.center$y,state.abb, col="white", cex = 0.5)
```

WINNING STATES BY CANDIDATE - 2008
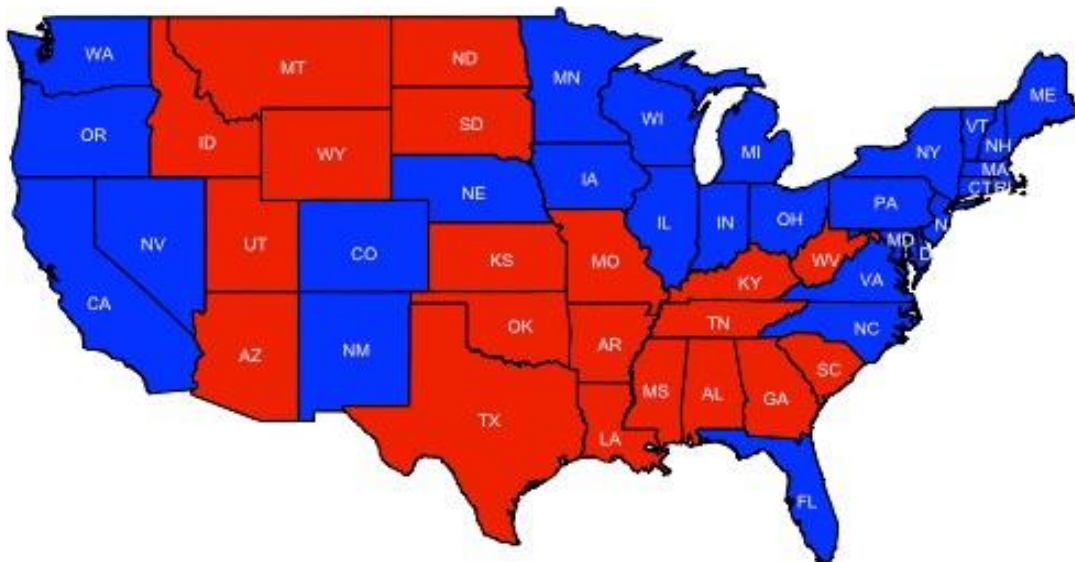
Here we have the first iteration as pointed out in the prior part.

```
win.obama = c( "California", "Colorado", "Connecticut", "District of Colum
bia",
          "Delaware", "Florida", "Hawaii", "Iowa" ,
          "Illinois", "Massachusetts",  "Maryland", "Maine" ,
          "Michigan", "Minnesota",   "New Hampshire", "New Jersey",
          "New Mexico", "Nevada",    "New York",    "Ohio",
          "Oregon", "Pennsylvania", "Rhode Island", "Virginia",
          "Vermont", "Washington", "Wisconsin")
win.romney = c("Alaska", "Alabama", "Arkansas", "Arizona", "Georgia", "Ida
ho",
          "Indiana", "Kansas", "Kentucky", "Louisiana", "Missouri", "Missi
ssippi",
          "Montana", "North Carolina", "North Dakota", "Nebraska", "Oklaho
ma", "South Carolina",
          "South Dakota", "Tennessee", "Texas", "Utah", "West Virginia", "
```

```
Wyoming")
map(database = "state")
map(database = "state", regions = win.obama, col = "blue", fill=T, add=TRU
E)
map(database = "state", regions = win.romney, col = "red2", fill=T, add=TR
UE)
title(main = "WINNING STATES BY CANDIDATE - 2012")
text(x=state.center$x, y=state.center$y, state.abb, col="white",cex = 0.5)
```

WINNING STATES BY CANDIDATE - 2012

As already mentioned at the beginning of this section, we refer here as well to our approach
for the year in 2008.

```
win.clinton = c("California", "Colorado", "Connecticut", "District of Colu
mbia",
          "Delaware", "Hawaii",
          "Illinois",  "Massachusetts",  "Maryland", "Maine" ,
          "Minnesota",    "New Hampshire", "New Jersey",
          "New Mexico", "Nevada",   "New York",
          "Oregon", "Rhode  Island", "Virginia",
```
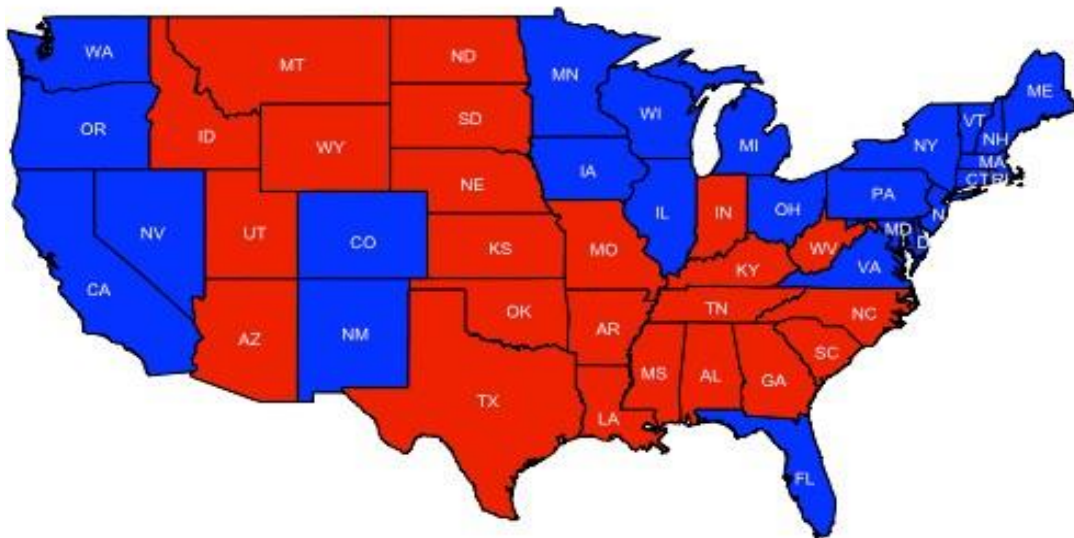
```
                 "Vermont", "Washington")
win.trump = c("Alaska", "Alabama", "Arkansas", "Arizona", "Florida", "Geor
gia", "Idaho","Indiana", "Iowa", "Kansas", "Kentucky", "Louisiana", "Michi
gan", "Missouri", "Mississippi",
                  "Montana", "North Carolina", "North Dakota", "Nebraska", "Oh
io", "Oklahoma", "Pennsylvania", "South Carolina",
                  "South Dakota", "Tennessee", "Texas", "Utah", "West Virginia
", "Wisconsin", "Wyoming")
map(database = "state")
map(database = "state", regions = win.clinton, col = "blue", fill=T, add=T
RUE)
map(database = "state", regions = win.trump, col = "red2", fill=T, add=TRU
E)
title(main = "WINNING STATES BY CANDIDATE - 2016")
text(x=state.center$x, y=state.center$y, state.abb, col="white",cex = 0.5)
```


WINNING STATES BY CANDIDATE - 2016
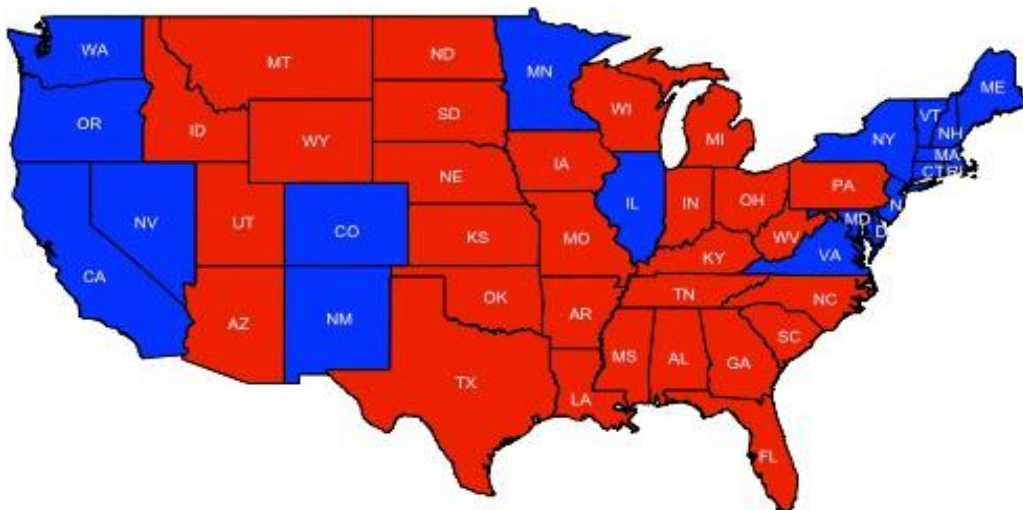
5. SWING STATES

For the upcoming part we decided to have a closer look at the swing states and analyze them, since they are a crucial part in the US elections.

The states that we consider as swing states are:

- Colorado, Florida, Iowa, Nevada, New Hampshire, North Carolina, Ohio, Virginia, Wisconsin

According to our researches we could identify that there are 9 major swing states. Main objective is to compare the results of the percentages of votes by candidate in each of the 9 swing states, and have them at on glance. Firstly, we created a vector "sw.st08" by using the acronyms of the 9 identified swing states. Then we built a new table containing just these 9 states with the results of our two major candidates by filtering. Finally, we generated a diagram(sw.plot08) considering the function ggplot. Here we have to consider the argument "aes" which is used to build the content of diagram, in this case "Acronym" for the x-axis, "Percentage.of.Votes" for the y-axis and the argument "fill" creates the histograms referring to the candidates. Furthermore, adding the argument "geom_bar" provided the desired result to have the results of the two different candidates next to each other. The function "labs" allows to give the diagram a suitable main title as well as a description of both axis. Eventually, "scale_fill_manual" gives each candidate a specific color, in which red refers to the republican and blue to the democratic party.

```
sw.st08 = c("CO", "FL", "IA", "NV", "NH", "NC", "OH", "VA", "WI")
swing.states08 = filter(Ele2008, Acronym %in% sw.st08)
sw.plot08 = ggplot(swing.states08,
        aes(Acronym, Percentage.of.Votes, fill = Candidate)) +scale_y_conti
nuous(limits = c(0,60))+
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Results by state - 2008",
       x = "State/district",
       y = "Votes %") + scale_fill_manual(values = c("navy","red2"))
```

As described in in the prior part, our approach did not change due to the application of the same codes. The main objective for the iteration will be a multiple visualization of all plots.

```
sw.st12 = c("CO", "FL", "IA", "NV", "NH", "NC", "OH", "VA", "WI")
swing.states = filter(obama.romney, Acronym %in% sw.st12)
sw.plot12 = ggplot(swing.states,
        aes(Acronym, Percentatge.of.Votes, fill = Candidate)) +scale_y_cont
inuous(limits = c(0,60))+
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Results by state - 2012",
       x = "State/district",
       y = "Votes %") + scale_fill_manual(values = c("navy","red2"))
```

For the year 2016 we refer like in 2012 to the description of approach in 2008.

```
sw.st16 = c("CO", "FL", "IA", "NV", "NH", "NC", "OH", "VA", "WI")
swing.states16 = filter(clinton.trump, Acronym %in% sw.st16)
```

```
swing.states16$Candidate = factor(swing.states16$Candidate, levels = c("Hi
llary Clinton", "Donald Trump"))
sw.plot16 = ggplot(swing.states16,
        aes(Acronym, Percentatge.of.Votes, fill = Candidate)) +scale_y_cont
inuous(limits = c(0,60))+
    geom_bar(stat = "identity", position = "dodge") +
    labs(title = "Results by state - 2016",
        x = "State/district",
        y = "Votes %") + scale_fill_manual(values = c("navy","red2"))
```

Here we compiled with the function "grid.arrange" a multiple visualization of the histograms we created in the prior parts for the years 2008 - 2016. For that we have to call "sw.plot08 - sw.plot16", and add "nrow" and "ncol" in order to define the output. The result will provide all plots at one glance so that the margins for each state and years can be detected easily.

```
grid.arrange(sw.plot08, sw.plot12, sw.plot16, nrow=3, ncol=1)
```


Results by state - 2008

Results by state - 2012

Results by state - 2016
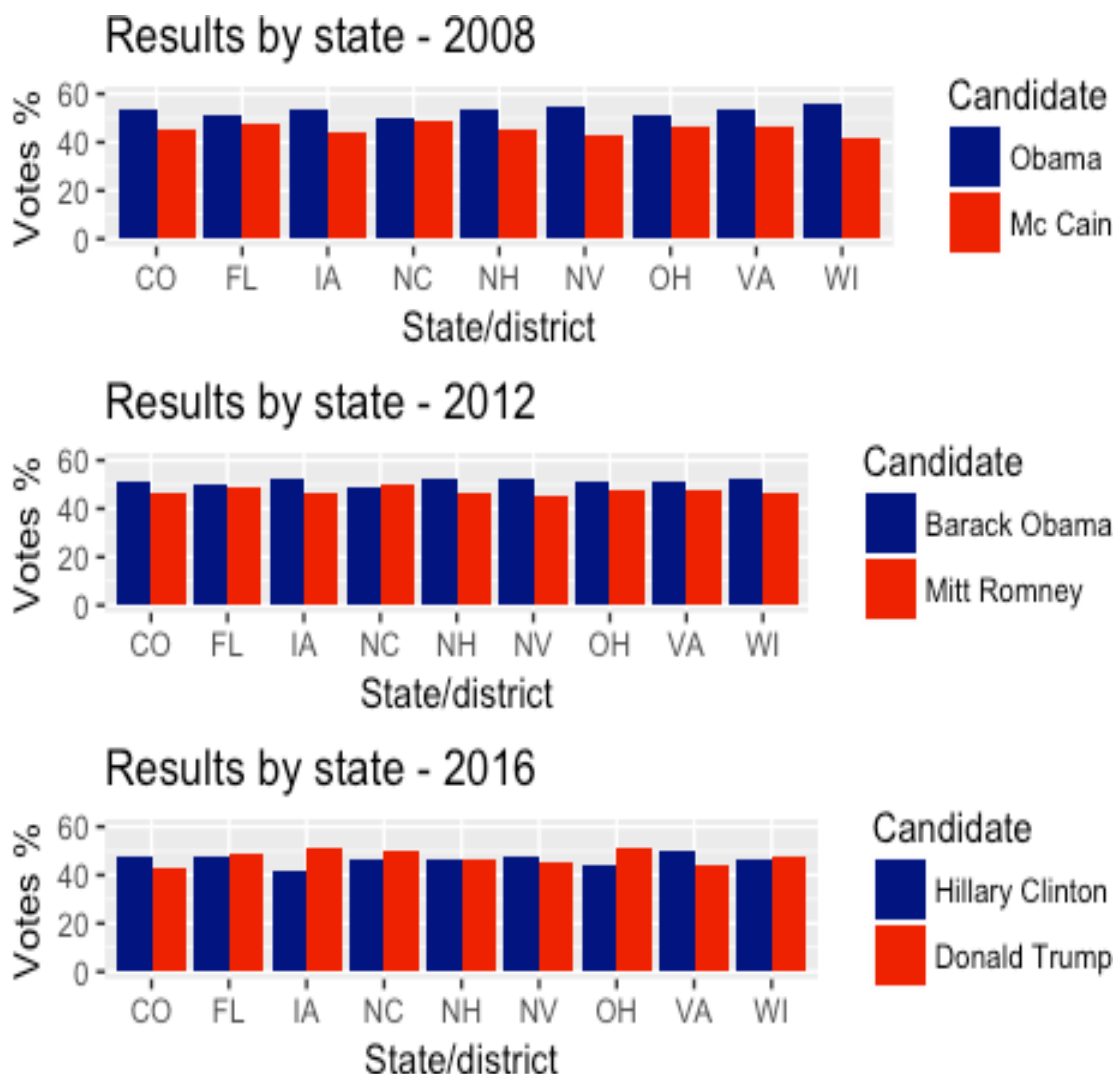
6. ALTERNATIVE APPROACH TO ELECTIONS RESULTS IN TERMS OF RACE AND GENDER

In this paragraph, we are introducing another way we chose to further analyze the information we got previously. We are going to introduce three more data sets regarding the data of the votes divided by races. We called this files Races08, Races12, Races16 respectively for the election of 2008 2012 and 2016.

```
races08 <- read.csv ("race2008END.csv")
races08
races12 <- read.csv ("race2012END.csv")
races12
races16 <- read.csv ("race2016END.csv")
races16
```

We filtered datasets in terms of total values, to see the general US approach towards elections.

```
races08.total <- filter(races08, State == "US")
races12.total <- filter(races12, State =="US")
races16.total <- filter(races16, State  =="US")
```

We created some figures in terms of percentage of difference between registered voters and people voted, to see how many people went for voting, how many of them is not convinced enough.

```
races08.total$total.diff <- races08.total$Percent.registered..Citizen. - races08.total$Percent.voted..Citizen.
d=races08.total$total.diff

races12.total$total.diff <- races12.total$Percent.registered..Citizen. - races12.total$Percent.voted..Citizen.
e=races12.total$total.diff

races16.total$total.diff <- races16.total$Percent.registered..Citizen.- races16.total$Percent.voted..Citizen.
f=races16.total$total.diff

x=c(d,e,f)
```

6.a. Examining Difference of Voted People and Registered Voters

We had an approach of examining total voted people, in terms of registered voters. Our aim was to see some drastic figures throughout years to examine people reflexes. Furthermore, we applied some filters to the latest election file.

```
male16=filter(races16.total, races16.total$Race.and.Hispanic.origin=='Male')
female16=filter(races16.total, races16.total$Race.and.Hispanic.origin=='Female')
male16$Total.registered

## [1] 73761000

female16$Total.registered

## [1] 83835000
```

For 2016, we can see that, there are 10 million more voters, when we compare both sexes. there is 73 millions of voters for females, 63 millions for males. What we can see from having datasets is, for 2016, there were 10 million of people didn't go for voting in terms of females. So, democrats should produce a meaningful campaign to courage women voting, to speak out their voices regarding sexist incidents.

```
differenceof16=mean(races16.total$total.diff)
differenceof12=mean(races12.total$total.diff)
differenceof08=mean(races08.total$total.diff)
```

To See Differences

We checked the mean values of differences between voted and registered voters. So, we decided to create subsets related with these mean values that we extracted.

```
subset(races16.total, races16.total$total.diff > 9)
```

##	State	Race.and.Hispanic.origin	Total.Population
## 2	US	Male	118488000
## 6	US	Black alone	30608000
## 8	US	Hispanic (of any race)	38990000
## 10	US	Black alone or in combination	32211000

##	Total.Citizen.Population	Total.registered	Percent.registered..Total.
## 2	107554000	73761000	62.3
## 6	28808000	19984000	65.3
## 8	26662000	15267000	39.2
## 10	30326000	2 0935000	65.0

##	Percent.registered..Citizen.	Total.voted	Percent.voted..Total.
## 2	68.6	63801000	53.8

31

```
## 6                    69.4              17119000               55.9
## 8                    57.3              12682000               32.5
## 10                   69.0              17875000               55.5
##      Percent.voted..Citizen.    total.diff
## 2                    59.3           9.3
## 6                    59.4           10.0
## 8                    47.6           9.7
## 10                   58.9           10.1
```

At 2016 elections, with a lower limit of 9% difference, we can see that Black people and Hispanic people were not motivated enough to go for voting.

subset(races12.total, races12.total$total.diff > 9.5)

```
##      State       Race.and.Hispanic.origin              Total.Population.
## 2     US                            Male                      113243000
## 4     US                     White  alone                     187084000
## 5     US          White non-Hispanic  alone                   155615000
## 8     US            Hispanic (of any race)                    35204000
## 9     US  White alone or in combination                       190396000
##      Total.Citizen.Population  Total.registered   Percent.registered..Total.
## 2            103022000              71414000                   63.1
## 4            173466000             124697000                   66.7
## 5            152862000             112706000                   72.4
## 8             23329000              13697000                   38.9
## 9            176612000             126822000                   66.6
##      Percent.registered..Citizen.   Total.voted    Percent.voted..Total.
## 2               69.3                  61551000            54.4
## 4               71.9                 107846000            57.6
## 5               73.7                  98041000            63.0
## 8               58.7                  11188000            31.8
## 9               71.8                 109603000            57.6
##      Percent.voted..Citizen.        total.diff
## 2                    59.7            9.6
## 4                    62.2            9.7
## 5                    64.1            9.6
## 8                    48.0            10.7
## 9                    62.1            9.7
```

At 2012 elections, with a lower limit of 9.5% difference, we can get that there were white people and Hispanic people, in majority, didn't go to ballot, and there were 65 million of unregistered white people, which is significant.

subset(races08.total, races08.total$total.diff > 8)

```
##      State Race.and.Hispanic.origin Total.Populati Total.Citiz.Population
## 8     US      Hispanic (of any race)     30852000              19537000
##      Total.registered Percent.registered..Total. Percent.registered..Citiz
## 8        11608000               37.6                              59.4
##      Total.voted Percent.voted..Total. Percent.voted..Citizen. total.diff
## 8        9745000          31.6                    49.9             9.5
```

At 2008 elections, with a lower limit of 8% difference, we see that there were Hispanic people who didn't go for vote that much, nearly 19 millions of registered people around the whole US.

6.b   Examining Swing States

Since swing states are basically defining the winner of the election. So, with that information, we decided to examine swing states. For example, in the last 120 years, all election winners won with the swing state Ohio, just two times without winning Ohio.

That's why we started to examine swing states and we pointed out meaningful figures in data.

We created a new matrix related with swing states and after that, we again created a difference value, percentage of difference between registered voters and people voted, to see how many people went for voting.

We set the difference of percentage limit with 10%, to have more versatile output.

```
target <- c("COLORADO", "FLORIDA", "IOWA", "NEVADA", "NEW HAMPSHIRE", "NORTH CAROLINA","OHIO", "VIRGINIA", "WISCONSIN", "PENNSYLVANIA")
swing08 <- filter(races08, State %in% target)
swing08$total.diff <- (swing08$Percent.registered..Citizen.-swing08$Percent.voted..Citizen.)
subset(swing08, swing08$total.diff > 10)
```

##	State	Race.and.Hispanic.origin	Total.Population
## 44	NEVADA	Asian alone or in combination	160000
## 62	NORTH CAROLINA	Asian alone	151000
## 66	NORTH CAROLINA	Asian alone or in combination	162000
## 88	PENNSYLVANIA	Asian alone or in combination	166000
## 99	VIRGINIA	Asian alone or in combination	337000

##	Total.Citizen.Population	Total.registered	Percent.registered..Total.
## 44	121000	48000	29.8
## 62	80000	53000	35.4
## 66	92000	61000	37.5
## 88	114000	47000	28.3
## 99	228000	164000	48.6

##	Percent.registered..Citizen.	Total.voted	Percent.voted..Total.
## 44	39.4	35000	21.7
## 62	66.4	43000	28.4
## 66	66.2	50000	31.0
## 88	41.4	35000	21.0
## 99	71.7	138000	41.1

##	Percent.voted..Citizen.	total.diff
## 44	28.6	10.8
## 62	53.3	13.1
## 66	54.7	11.5
## 88	30.6	10.8
## 99	60.6	11.1

We see that there is a result like specifically points out that Asian people didn't go to vote with an average of 11.46 of whole registered voters in Nevada, North Carolina, Pennsylvania and Virginia.

We did another filtering for 2008 elections, which is related with the knowledge females and the black people are the most discriminated groups in current US Presidential campaign.

```
target2 <- c("Female", "Black alone")
swing08.female <- filter(swing08, Race.and.Hispanic.origin %in% target2)
swing08.female$total.diff <- swing08.female$Percent.registered..Citizen.-s
wing08.female$Percent.voted..Citizen.
subset(swing08.female, swing08.female$total.diff > 8)
```

```
##              State        Race.and.Hispanic.origin        Total.Population
## 11  NORTH CAROLINA                       Female                  3595000
##     Total.Citizen.Population Total.registered Percent.registered..Total.
## 11           3425000               2671000                  74.3
##     Percent.registered..Citizen. Total.voted Percent.voted..Total.
## 11              78                2364000              65.8
##     Percent.voted..Citizen.      total.diff
## 11              69                    9
```

With the information that we get from 08 elections, we examined the 2012 elections in the same way that we did for 08 elections.

This time we applied Female and Black alone matrix to see the results.

```
swing12 <- filter(races12, State %in% target)
swing12.female <- filter(swing12, Race.and.Hispanic.origin %in%target2)
swing12.female$total.diff <- swing12.female$Percent.registered..Citizen.-s
wing12.female$Percent.voted..Citizen.
subset(swing12.female, swing12.female$total.diff > 8)
```

```
##              State      Race.and.Hispanic.origin        Total.Population.
## 4          FLORIDA                Black alone             2244000
## 11  NORTH CAROLINA                    Female             3826000
## 15    PENNSYLVANIA                    Female             5120000
##     Total.Citizen.Population Total.registered Percent.registered..Total.
## 4           1919000               1262000                  56.3
## 11          3587000               2894000                  75.7
## 15          4919000               3578000                  69.9
##

Percent.registered..Citizen.        Total.voted      Percent.voted..Total.
## 4              65.8                1104000              49.2
## 11             80.7                2521000              65.9
## 15             72.7                3084000              60.2
##     Percent.voted..Citizen.      total.diff
## 4              57.6                  8.2
## 11             70.3                  10.4
## 15             62.7                  10.0
```

It turns out that, with a 9.6% of registered voters with filter of Black alone and Female, they did not intend for voting in Florida, North Carolina, Pennsylvania.

we had the same approach again for 2016 elections, to see what happened and why Democrats have lost.

```
swing16 <- filter(races16, State %in% target)
swing16.female <- filter(swing16, Race.and.Hispanic.origin %in% target2)
swing16.female$total.diff <- swing16.female$Percent.registered..Citizen.-s
wing16.female$Percent.voted..Citizen.
subset(swing16.female, swing16.female$total.diff > 8)
```

```
##                State   Race.and.Hispanic.origin      Total.Population
## 5              IOWA              Female                   1221000
## 7             NEVADA             Female                   1136000
## 14            OHIO            Black alone                 1031000
## 15   PENNSYLVANIA             Female                   5172000
## 16   PENNSYLVANIA          Black alone                 1071000
##      Total.Citizen.Population   Total.registered   Percent.registered..Total.
## 5            1181000                 877000                  71.8
## 7            1001000                 708000                  62.4
## 14            977000                 722000                  70.0
## 15           5007000                3709000                  71.7
## 16           1007000                 751000                  70.1
##      Percent.registered..Citizen.   Total.voted   Percent.voted..Total.
## 5              74.2                  781000                63.9
## 7              70.8                  625000                55.0
## 14             73.9                  637000                61.8
## 15             74.1                 3215000                62.2
## 16             74.5                  637000                59.5
##      Percent.voted..Citizen.        total.diff
## 5              66.1                    8.1
## 7              62.4                    8.4
## 14             65.2                    8.7
## 15             64.2                    9.9
## 16             63.2                   11.3
```

The 2016 results show that with a 9.25% of difference, in states Iowa, Nevada, Ohio and Pennsylvania, there is a significant difference.

The crucial thing could be the difference in Ohio, since we know that who gets Ohio, also gets the election.

There is 9 millions of Black people who registered but didn't vote, which is a critic detail for elections.

```
obama08.all   =   filter(Ele2008,    Candidate == "Obama")
obama08.sum = sum(obama08.all$People.Votes)


mccain.all   =   filter(Ele2008,    Candidate == "Mc Cain")
mccain.sum = sum(mccain.all$People.Votes)


dm.rp08 = c("Obama", "Mc Cain")
```

7. *CONCLUSION*

To build and create some arguments and suggestions for Democrats, it should be wiser to concentrate what happened at last election, so, here is the facts for last election.

Hillary Clinton and Democrats have lost the election even they had the popular vote, by the difference of 3 millions of votes. This happened 5 times now through all U.S history. In fact, Clinton wins US popular vote by widest margin of any losing presidential candidate

Trump won the Elections with electoral college, his performance is ranked 46th amongst 58 elections. He managed to win it because of some factors happened around country during campaign weeks.
He got the indecisive voters, also he has gained power, somehow, from people that did not go to vote.
The main people who didn't go for vote was possible Democrat voters. You can find this information above in the **Races** and **Gender** part, which exactly points the situation.

There were also lack of campaigns for Democrats in terms of Swing states, which are the states that they've won at last two elections, 2008 and 2012. They have lost Ohio for example, with 8% difference, which was obvious also before elections, as polls have shown.

Swing States

Starting a comparison of the election years 2008 and 2012, Obama lost two states, in this case Nebraska and Indiana. If we have a look at both election years of Nebraska we can determine that about 6% less people voted in 2012, even though population increased over the years. We assume from this information that people who did not vote were a crucial part for the democrats in this state.

Checking on what happened between 2012 and 2016, the difference in States, as we can se from the map, denotes that Republicans have got the electoral votes of Iowa, Wisconsin, Florida, Minnesota, Ohio, and Pennsylvania that were before belonging to Democrats.

Race and Gender

Since the Asian community in terms of voting is between 3-4%, we ignored that fact, and dealt with the rest of 97%.

First, we checked Female and Male distribution in registered voters.

In 2016 Election:

There was 63 million of registered male voters, 53 millions of them voted.
There was also 73 million of registered female voters, 63 million of them voted.

Firstly, we can say that, by checking the latest election, 2016, Democrats should convince and courage Females to vote and shout their words harder; since there is 10 more million voters. They can manipulate the results in their favor, if they can get organized.

Then, we started to check other significant outputs coming from the dataset.

For example, the most crucial swing state is maybe Ohio: with the fact that, for last 120 years, except two elections, Ohio went always for winners. Also, Ohio is demographically very similar to the U.S. average in socio economics and race/ethnic allocation.

 Then, if you lose it, there is a highly possibility to lost the elections. So, when we checked the results on Ohio, we saw that there are 9 million of black people who didn't vote, even they are registered for voting. If there is a campaign for them to go for voting, for example pointing out the equality manners, then it would be a game-changer for elections.

Edison Research, 2016 Elections

We've analysed one of the most reliable Researches conducted related with American Elections; we've found out some significant outputs. You may also see it above the conclusion as filtered datasets.

Those outputs can be used by Democrat Campaign coordinators to develop new strategies for gaining more support from critical states, which tend to vote for Republicans.

The outputs are:

- 14% of LGBT people are voted for Trump, for Republicans in 2016 Elections. Republicans' main vote pool is conservative people and they are strongly against LGBT community and its activities; also, Republican Trump campaign did not put any positive outline related with LGBT rights, they tend to not to make any argument-builder comments for protecting their conservative support.

- For the different races, there are some significant outcomes too. For example, nearly 1.5 million of Asian people are voted for Trump, which that candidate was expressing some opinions for creating a stricter immigrant policy. Also for Hispanics, there is 29%

of them voted for Trump, that is a huge contribution for Republicans. They had that amount of vote by declaring a possible wall for Mexican border.

- Women are the other stress point, since 41 % of them voted for Republicans. There should be an equality campaign to attract them with the most crucial argument related with gender equality.

- Age is another bullet, which needs to examined with sociologic arguments also; U.S voters tend to vote for Republicans as they get old.

- Education is a one of the key points of that Demographic research. As Education level goes higher, Republican votes are decreasing. So, Democrats can create a methodology for reaching the less educated people.

- Family income is another key factor, as the income gets higher, tendency for voting Republicans gets higher too. So, in order to stress out the income equality and also stressing Liberal Economy, Democrats should produce a new argument related with them.

- Community size and State structures is the most vital factor amongst all research; it points out huge figures. 17 % of the voters in U.S are living in Rural area and more then 60% of them voted for Republicans. From Cities to Suburbs, also from Suburbs to Rural Areas, the trending goes through from Democrats to Republicans.

That trending is in fact really crucial, Democrats' votes are decreasing from 59% to 34%, from Cities to Rural Areas. So there should be a welcoming approach for the Rural Area people as well, in order to change the Results for further elections, especially for the states located at Middle of U.S

8. REFERENCES

- Hadley Wickham, Jim Hester and Romain Francois (2017). readr: Read Rectangular Text Data. R package version
https://CRAN.R-project.org/package=readr

- Hadley Wickham, Romain Francois, Lionel Henry and Kirill Müller (2017). dplyr: A Grammar of Data Manipulation. R
package version 0.7.4. https://CRAN.R-project.org/package=dplyr

- Alboukadel Kassambara (2017). ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.1.6.
https://CRAN.R-project.org/package=ggpubr

- H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.

- Sarkar, Deepayan (2008) Lattice: Multivariate Data Visualization with R. Springer, New York. ISBN
978-0-387-75968-5

- Original S code by Richard A. Becker, Allan R. Wilks. R version by Ray Brownrigg. Enhancements by Thomas P Minka and Alex Deckmyn. (2017). maps: Draw Geographical Maps. R package version 3.2.0.
https://CRAN.R-project.org/package=maps

- Baptiste Auguie (2017). gridExtra: Miscellaneous Functions for "Grid" Graphics. R package version 2.3.
https://CRAN.R-project.org/package=gridExtra

- Winston Chang (2017). webshot: Take Screenshots of Web Pages. R package version 0.4.2.
https://CRAN.R-project.org/package=webshot

- Carson Sievert, Chris Parmer, Toby Hocking, Scott Chamberlain, Karthik Ram, Marianne Corvellec and Pedro Despouy (2017). plotly: Create Interactive Web Graphics via 'plotly.js'. R package version 4.7.1.
https://CRAN.R-project.org/package=plotly

- https://www.usa.gov/election

- http://www.telegraph.co.uk/women/politics/donald-trump-sexism-tracker-every-offensive-comment-in-one-place

- https://theintercept.com/2017/08/15/donald-trump-has-been-a-racist-all-his-life-and-he-isnt-going-to-change-after-charlottesville

- https://newrepublic.com/minutes/144992/donald-trump-will-never-stop-racist-president

- https://www.billboard.com/articles/news/7572638/electoral-college-hillary-clinton-donald-trump-2016-president-election-popular-vote