Data Viz Project

2016 US general Elections

Daphine Rose Nakiberu

EXECUTIVE SUMMARY

The United States 2016 presidential election was the 58th American Presidential election held on November 8th, 2016. There were two major parties that participated in the US general Elections that is the Republican and Democrat. The Republican's flag bearer Donald Trump defeated the Democratic party flag bearer the former secretary of state Hillary Clinton. Trump emerged as the front-runner despite amidst a wide field of Republican primary candidates while Clinton defeated Senator Bernie Sanders and became the first female presidential nominee of a major political party (Democratic Party). The basis of this report is to assess whether the Republicans won the election free and fair since there are repeated allegations of voter fraud.

Most Americans have felt bad for the Republican party since 1992, only one of its presidential nominees has gotten the most votes for president and that is George W. Bush in 2004. Donald Trump was the latest also ran, loosing nationally to Hillary Clinton by more than 200,000 votes.

Too many of Clinton's votes came in places like California, New York and Illinois, and too few from Wisconsin and Florida. The result is that Trump outdid her in the Electoral College, which is the only place that counts.

Ever since the Republicans entered the White house, there has always been allegations that the voters voted for Clinton, and the Electoral College gave them Trump. Most politicians and sympathetic commentators endlessly claim that the Republicans won the Elections and the Democrats proved that they are out of step with most

Americans. However, neither of this is true and the myth should be nipped in the bud.

Dataset description

The analysis of this executive summary uses a data set of the US 2016 Elections voter with preferences of their states, political parties, counties and candidates. The 2016 US Election dataset was conducted by Ben Hamner 2 years ago(7th version) and can be accessed on https://www.kaggle.com/benhamner/2016-us-election/version/7. Among the verified voters, the overall vote preference mirrors the election results very closely as Hillary Clinton had the largest distribution of votes compared to Donald Trump

load and summarize packages

library(tidyverse)

primary_results <- read_csv("C:/Users/nakibeda/Desktop/primary_results.csv")

View(primary_results)

glimpse(primary_results)

summary(primary_results)

```
library(tidyverse)
primary_results <- read_csv("C:/Users/nakibeda/Desktop/primary_results.csv")
View(primary_results)
glimpse(primary_results)
```

```
## Observations: 24,611
## Variables: 8
## $ state              <chr> "Alabama", "Alabama", "Alabama", "Alabama",...
## $ state_abbreviation <chr> "AL", "AL", "AL", "AL", "AL", "AL", "AL", "...
## $ county             <chr> "Autauga", "Autauga", "Baldwin", "Baldwin",...
## $ fips               <int> 1001, 1001, 1003, 1003, 1005, 1005, 1007, 1...
## $ party              <chr> "Democrat", "Democrat", "Democrat", "Democr...
## $ candidate          <chr> "Bernie Sanders", "Hillary Clinton", "Berni...
## $ votes              <int> 544, 2387, 2694, 5290, 222, 2567, 246, 942,...
## $ fraction_votes     <dbl> 0.182, 0.800, 0.329, 0.647, 0.078, 0.906, 0...
```

```
summary(primary_results)
```

```
##     state           state_abbreviation    county
##  Length:24611       Length:24611       Length:24611
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##       fips             party             candidate             votes
##  Min.   :    1001   Length:24611       Length:24611       Min.   :     0
##  1st Qu.:   21091   Class :character   Class :character   1st Qu.:    68
##  Median :   42081   Mode  :character   Mode  :character   Median :   358
##  Mean   :26671525                                        Mean   :  2306
##  3rd Qu.:90900125                                        3rd Qu.:  1375
##  Max.   :95600036                                        Max.   :590502
##  NA's   :100
##  fraction_votes
##  Min.   :0.0000
##  1st Qu.:0.0940
##  Median :0.2730
##  Mean   :0.3045
##  3rd Qu.:0.4790
##  Max.   :1.0000
##
```

2016 candidates that scored the most votes

```
library(tidyverse)
primary_results %>% arrange(fraction_votes)
```

```
## # A tibble: 24,611 x 8
##    state state_abbreviat~ county   fips party candidate votes
##    <chr> <chr>            <chr>   <int> <chr> <chr>     <int>
##  1 Alas~ AK               State~ 9.02e7 Demo~ Hillary ~     0
##  2 Alas~ AK               State~ 9.02e7 Demo~ Hillary ~     0
##  3 Alas~ AK               State~ 9.02e7 Demo~ Hillary ~     0
##  4 Alas~ AK               State~ 9.02e7 Demo~ Hillary ~     0
##  5 Alas~ AK               State~ 9.02e7 Demo~ Hillary ~     0
##  6 Alas~ AK               State~ 9.02e7 Demo~ Hillary ~     0
##  7 Alas~ AK               State~ 9.02e7 Demo~ Hillary ~     0
##  8 Alas~ AK               State~ 9.02e7 Demo~ Hillary ~     0
##  9 Alas~ AK               State~ 9.02e7 Demo~ Hillary ~     0
## 10 Alas~ AK               State~ 9.02e7 Demo~ Hillary ~     0
## # ... with 24,601 more rows, and 1 more variable: fraction_votes <dbl>
```

a histogram showing the distribution of votes between the Democrats and the Republicans

Before drawing conclusions on whether the elections were not free and fair, it is important to explore which political party had the largest

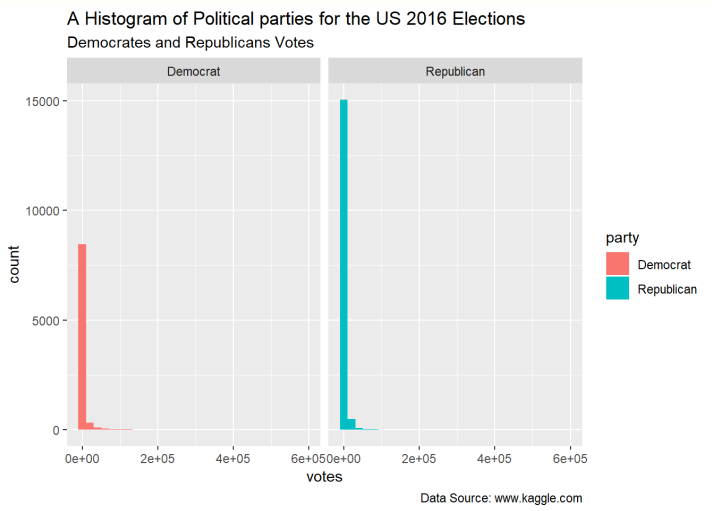
number of votes. The 2016 US Elections, voter choice and party affiliation were nearly synonymous. In the histogram presents that the Republican validated voters were reported choosing Trump by a margin of 15000 votes, while Democrats supported Clinton by 7000 votes. The results justify that the Republican won the 2016 general Elections.

```
library(tidyverse)

ggplot(data = primary_results) + aes(x = votes, fill = party) + geom_histogram() + facet_wrap(~party, ncol = 100) + labs(title="A Histogram of Political parties for the US 2016 Elections",subtitle="Democrates and Republicans Votes",caption = "Data Source: www.kaggle.com")
```

```
library(tidyverse)
ggplot(data = primary_results) + aes(x = votes, fill = party) + geom_histogram() + facet_wrap(
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Histograms showing 2016 US presidential candidates' votesHowever, the histograms below seek to visualize the candidates' correlation in the 2016 US general Election and the results are as follows. For the democrats that is Clinton and Donald Trump for the Republican received majority of the votes showed a normal distribution of votes and this has showed a stiff competition among the candidates.

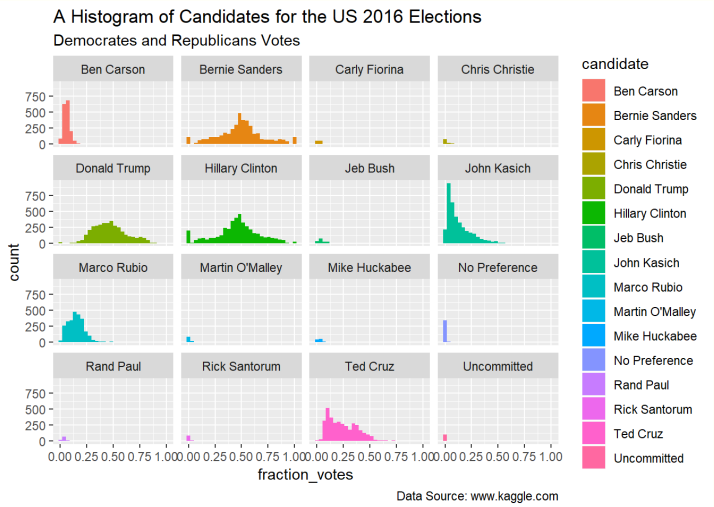
While un committed and no preference voters were politically disengaged, there are far more nonvoters than voters who fall into the

un committed voters who hold a set of political values with a distinct ideological orientation, those with generally liberal values considerably outnumbered those with generally conservative values. In addition to that, candidates like Donald Trump, Hillary Clinton and Bernie Sanders show the normal distribution of votes throughout the United states for example, Clinton's chart show that the Democrats got majority of the votes which justifies the fact that the people voted Clinton, and the Electoral College gave them Trump.

ggplot(data = primary_results) + aes(x = fraction_votes, fill = candidate) + geom_histogram() + facet_wrap(~candidate, ncol = 4) + labs(title="A Histogram of Candidates for the US 2016 Elections",subtitle="Democrates and Republicans Votes",caption = "Data Source: www.kaggle.com")

```
ggplot(data = primary_results) + aes(x = fraction_votes, fill = candidate) + geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



A scatter plot showiing the relationship between the Republicans and the Democrats

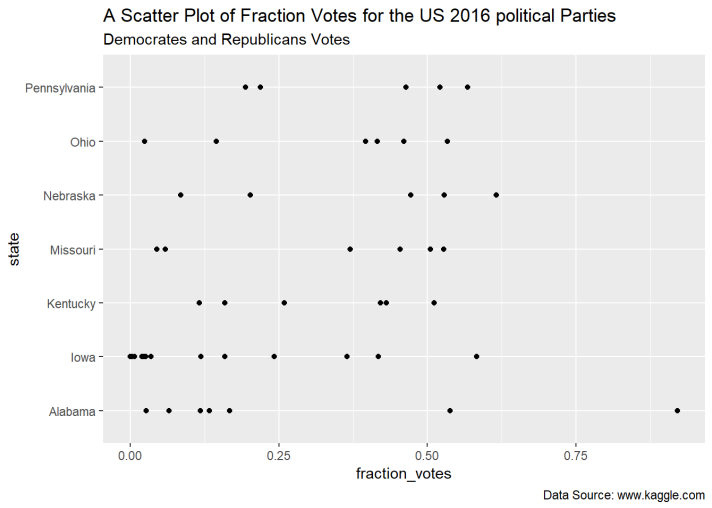
The scatter graph below presents the number of voters from states with Butler County. The scatter plot shows a strong positive correlation between the Republicans and the Democrats in Pennsylvania, Ohio, Nebraska, Missouri, Kentucky, Iowa and Alabama.

library(tidyverse)

primary_results %>% filter(county == "Butler") %>% ggplot() + aes(x = fraction_votes, y = state) + geom_point() + labs(title="A Scatter Plot of Fraction Votes for the US 2016 political Parties",subtitle="Democrates and Republicans Votes",caption = "Data Source: www.kaggle.com")

```
library(tidyverse)
```

```
primary_results %>% filter(county == "Butler") %>% ggplot() + aes(x = fraction_votes, y = stat
```

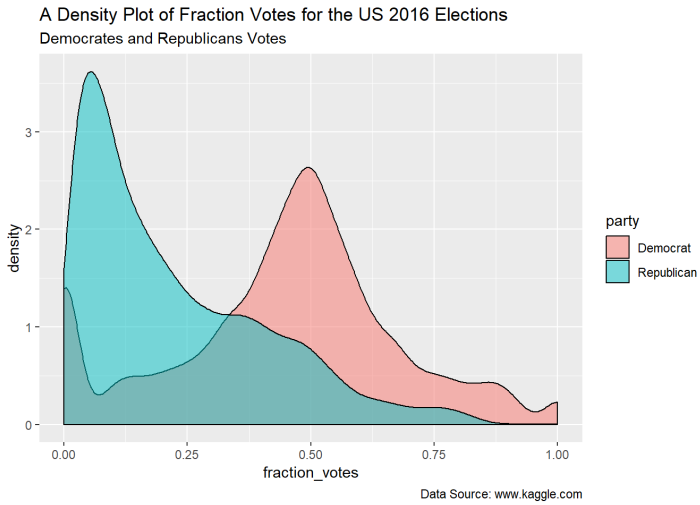


density plot showing the destribution of votes between the Democrats and the Repubicans

A density plot below presents the distribution of data for both republicans and democrats over a continuous interval of votes during the Elections. This chart has focused to display which party had the highest concentration of votes during the 2016 US Elections. The density plot below has justified that the Democrats have the largest distribution of votes compared to Democrats, hence justifying that the elections were not free and fair. However, the republicans' curve is positively skewed since the mean is greater than the median.

ggplot(primary_results, aes(x = fraction_votes, fill = party)) + geom_density(alpha = 0.5) + labs(title="A Density Plot of Fraction Votes for the US 2016 Elections",subtitle="Democrates and Republicans Votes",caption = "Data Source: www.kaggle.com")

```
ggplot(primary_results, aes(x = fraction_votes, fill = party)) + geom_density(alpha = 0.5) + l
```

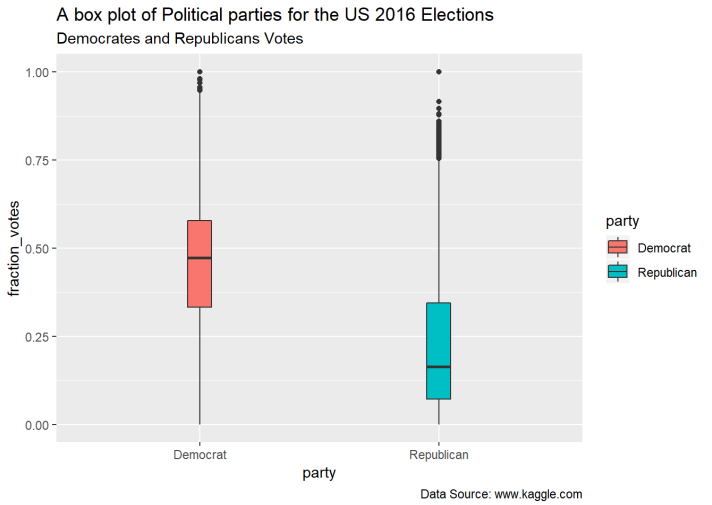


A box plot showing the distributon votes between the Democrats and the Republicans

The box plot below illustrates the distribution of votes among the voters. In addition to that, the box chart below seeks to distribution of votes between the Republicans and the Democrats. The democrats have the constitutes of the highest median of fraction votes which is 0.48 with a maximum of 60 and a minimum of 30 votes. While the republicans validated the smaller mean with a majority of 25votes and a maximum of 32 votes and a minimum of 7 votes.

ggplot(data = primary_results) + aes(x = party, y = fraction_votes, fill = party) + geom_boxplot(width = 0.1) + labs(title="A box plot of Political parties for the US 2016 Elections",subtitle="Democrates and Republicans Votes",caption = "Data Source: www.kaggle.com")

```
ggplot(data = primary_results) + aes(x = party, y = fraction_votes, fill = party) + geom_boxpl
```



Conclusion

Twice in this century, the Electoral College has overruled the American people, handing the White House to someone they rejected. The outcome of this election is ominous based on the visualizations, and it's not the voters to blame.

REFERENCES

https://www.chicagotribune.com/news/opinion/commentary/ct-donald-trump-wins-clinton-electoral-college-20161109-story.html

https://www.kaggle.com/benhamner/2016-us-election

https://en.wikipedia.org/wiki/2016_United_States_presidential_electi