Exploring election results with few questions

Jibin

3/06/2020

```
#Loading libraries

library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.6.3
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v tibble  3.1.0     v dplyr   1.0.5
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
## v purrr   0.3.4

## Warning: package 'tibble' was built under R version 3.6.3

## Warning: package 'tidyr' was built under R version 3.6.3

## Warning: package 'readr' was built under R version 3.6.3

## Warning: package 'purrr' was built under R version 3.6.3

## Warning: package 'dplyr' was built under R version 3.6.3

## Warning: package 'forcats' was built under R version 3.6.3

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(readr)
library(readxl)

#Reading and viewing all files from the working directory


county_facts <- read_csv("county_facts.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   .default = col_double(),
##   area_name = col_character(),
##   state_abbreviation = col_character()
## )
## i Use `spec()` for the full column specifications.
```

```
View(county_facts)

county_facts_dictionary <- read_excel("county_facts_dictionary.xlsx")
View(county_facts_dictionary)

primary_results <- read_csv("primary_results.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   state = col_character(),
##   state_abbreviation = col_character(),
##   county = col_character(),
##   fips = col_double(),
##   party = col_character(),
##   candidate = col_character(),
##   votes = col_double(),
##   fraction_votes = col_double()
## )
```

```
View(primary_results)


#Q.1) Does higher level of education have a subsequent effect on the voting or the number of votes rece

#To answer this question, we need a graph which would compare the percent of degree holders and the vot


# Creating a new dataset which represents only the percentage of bachelor degree holders in all the sta

(degree_holders <- county_facts %>% select(fips,EDU685213) %>% arrange(desc(EDU685213)))
```

```
## # A tibble: 3,195 x 2
##     fips EDU685213
##    <dbl>     <dbl>
##  1 51610      74.4
##  2 51013      71.7
##  3 35028      63.4
##  4 51510      61.4
##  5 24027      60
##  6 36061      58.9
##  7 51059      58.6
##  8  8013      58.3
##  9 51107      57.9
```

```
## 10 24031      57.1
## # ... with 3,185 more rows
```

```
(degree_holders <- degree_holders %>% rename("Percentage_Degree_Holders"=EDU685213 ))
```

```
## # A tibble: 3,195 x 2
##     fips Percentage_Degree_Holders
##    <dbl>                     <dbl>
##  1 51610                      74.4
##  2 51013                      71.7
##  3 35028                      63.4
##  4 51510                      61.4
##  5 24027                      60
##  6 36061                      58.9
##  7 51059                      58.6
##  8  8013                      58.3
##  9 51107                      57.9
## 10 24031                      57.1
## # ... with 3,185 more rows
```

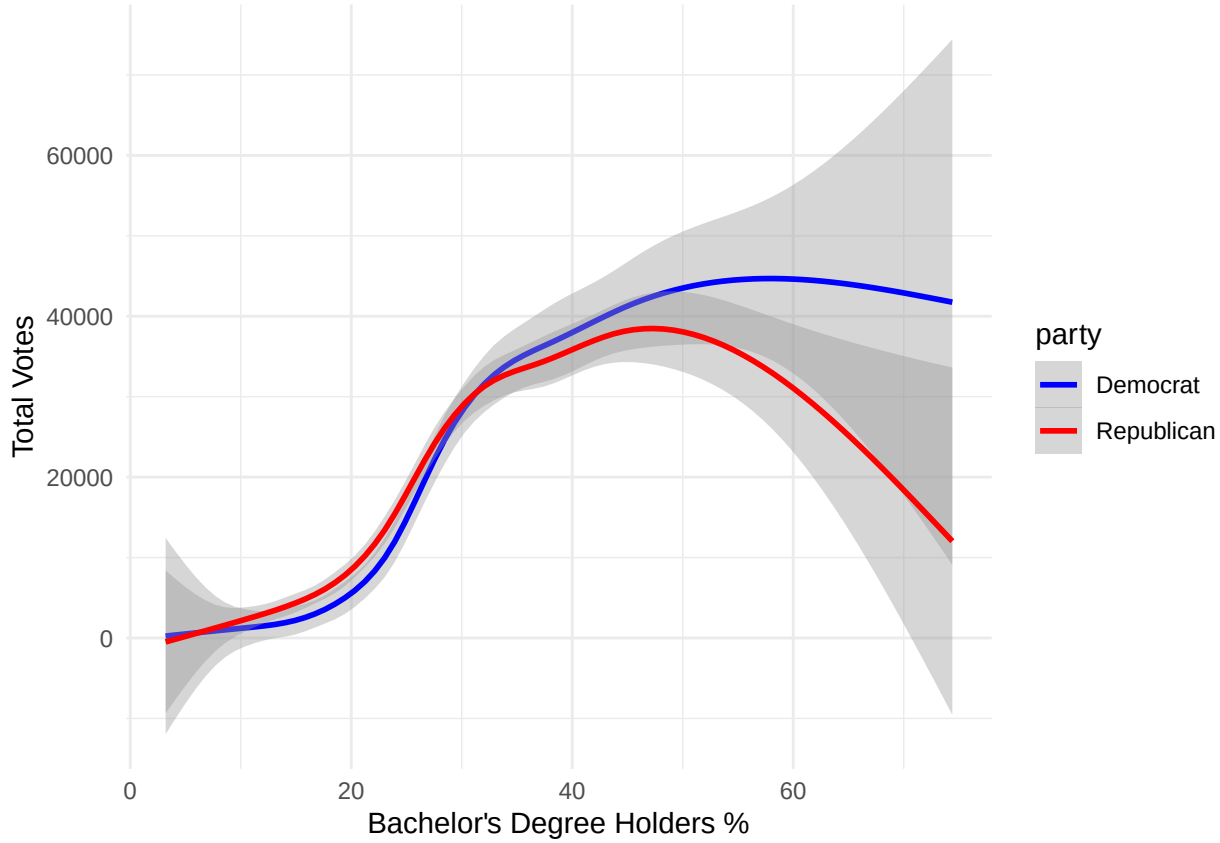
```
#Calculating the total votes received by a party in a state and county using fips which serves as a uni

(total_votes <- primary_results %>% group_by(fips,party) %>% summarise(votes_total=sum(votes)))
```

```
## `summarise()` has grouped output by 'fips'. You can override using the `.groups` argument.
```

```
## # A tibble: 7,773 x 3
## # Groups:   fips [4,208]
##     fips party      votes_total
##    <dbl> <chr>            <dbl>
##  1  1001 Democrat          2931
##  2  1001 Republican       11839
##  3  1003 Democrat          7984
##  4  1003 Republican       49100
##  5  1005 Democrat          2789
##  6  1005 Republican        3357
##  7  1007 Democrat          1188
##  8  1007 Republican        3891
##  9  1009 Democrat           959
## 10  1009 Republican       14791
## # ... with 7,763 more rows
```

```
# Merging datasets: degree_holders and total_votes by fips

merged_dataset <- merge(x = degree_holders ,y= total_votes, by = "fips")


#Plotting a graph to compare total votes received and the percentage of population with degree holders.

merged_dataset %>% ggplot(aes(x=Percentage_Degree_Holders,y=votes_total,color=party)) + geom_smooth() +
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```




```
#Answer: It can be observed that as the percentage of population with a bachelors degree increases, the

#A possible likely explanation could be that a person with a higher level of education would have views



#Q.2) Compare votes received by a candidate with respect to the hispanic population in that region?


#Gathering percentage of hispanics from the county_facts dataset and storing it in a new dataframe

(hispanic_ds <- county_facts %>% select(fips,RHI725214) %>% arrange(desc(RHI725214)))
```

```
## # A tibble: 3,195 x 2
##     fips RHI725214
##    <dbl>     <dbl>
##  1 48427      95.8
##  2 48479      95.2
##  3 48323      95.1
##  4 48505      93.9
##  5 48507      93
##  6 48247      91.6
##  7 48215      91.2
```

```
##  8 48047      89.6
##  9 48061      88.7
## 10 48131      88.5
## # ... with 3,185 more rows
```

```
#Calculating the total votes received by a candidate

(total_votes1 <- primary_results %>% group_by(fips,party,candidate) %>% summarise(votes_total=sum(votes)
```

```
## `summarise()` has grouped output by 'fips', 'party'. You can override using the `.groups` argument.
```

```
## # A tibble: 24,521 x 4
## # Groups:   fips, party [7,773]
##     fips party      candidate       votes_total
##    <dbl> <chr>      <chr>                 <dbl>
##  1  1001 Democrat   Bernie Sanders          544
##  2  1001 Democrat   Hillary Clinton        2387
##  3  1001 Republican Ben Carson             1764
##  4  1001 Republican Donald Trump           5387
##  5  1001 Republican John Kasich             421
##  6  1001 Republican Marco Rubio            1785
##  7  1001 Republican Ted Cruz               2482
##  8  1003 Democrat   Bernie Sanders         2694
##  9  1003 Democrat   Hillary Clinton        5290
## 10  1003 Republican Ben Carson            4221
## # ... with 24,511 more rows
```

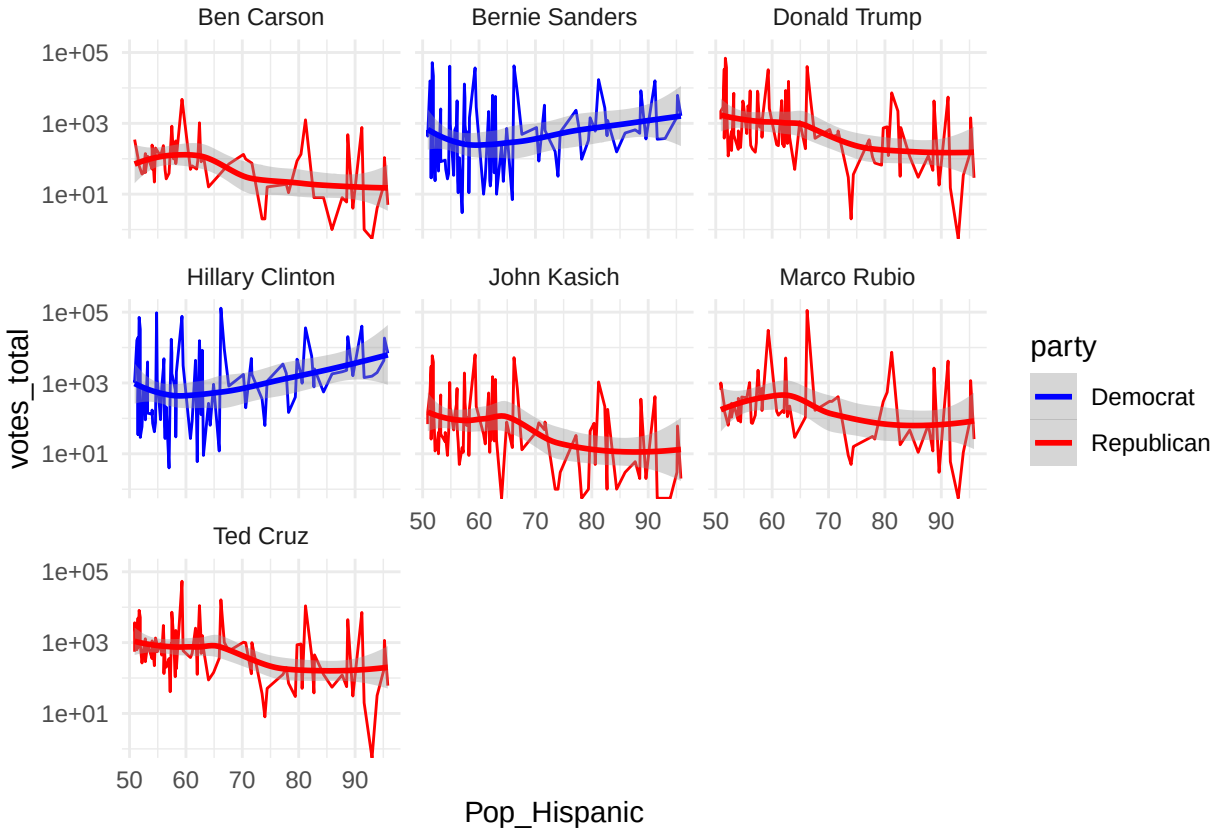
```
#Merging the hispanic_ds with the total_votes datasets.

merged_dataset1 <- merge(x = hispanic_ds,y= total_votes1, by = "fips")

#Filtering the dataset to have a look of regions where the hispanic community is dominant.

merged_dataset1 <- merged_dataset1 %>% filter(RHI725214 > 50) %>% rename("Pop_Hispanic"=RHI725214)

#Plotting a graph for total votes received versus total population of hispanic community.


merged_dataset1 %>% ggplot(aes(x=Pop_Hispanic,y=votes_total,color=party))  + geom_line() + geom_smooth()
```

```
## Warning: Transformation introduced infinite values in continuous y-axis
```

```
## Warning: Transformation introduced infinite values in continuous y-axis
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

```
## Warning: Removed 9 rows containing non-finite values (stat_smooth).
```




```
#Answer: So from the graph its clearly evident that as the population of Hispanic community increases,

#It can be interpreted from the graph that the Hispanic community do have a preference for the Democrat


#Q.3) Does the female population have an inclination towards female candidates?

#To answer this question, we need a graph which would compare the percent of female population and the

#Gathering percentage of females from the county_facts dataset and storing it in a new dataframe

(female_ds <- county_facts %>% select(fips,SEX255214) %>% arrange(desc(SEX255214)))
```

```
## # A tibble: 3,195 x 2
##     fips SEX255214
##    <dbl>     <dbl>
##  1 13235      56.8
##  2 29117      55.9
##  3 54089      55.2
##  4  1119      55
##  5 51790      55
##  6 51620      54.9
##  7 29007      54.5
##  8 51630      54.5
```

```
##  9 51065      54.3
## 10 56027      54.3
## # ... with 3,185 more rows
```

```
(female_ds <- female_ds %>% rename("Female_pop"=SEX255214))
```

```
## # A tibble: 3,195 x 2
##     fips Female_pop
##    <dbl>      <dbl>
##  1 13235       56.8
##  2 29117       55.9
##  3 54089       55.2
##  4  1119       55
##  5 51790       55
##  6 51620       54.9
##  7 29007       54.5
##  8 51630       54.5
##  9 51065       54.3
## 10 56027       54.3
## # ... with 3,185 more rows
```

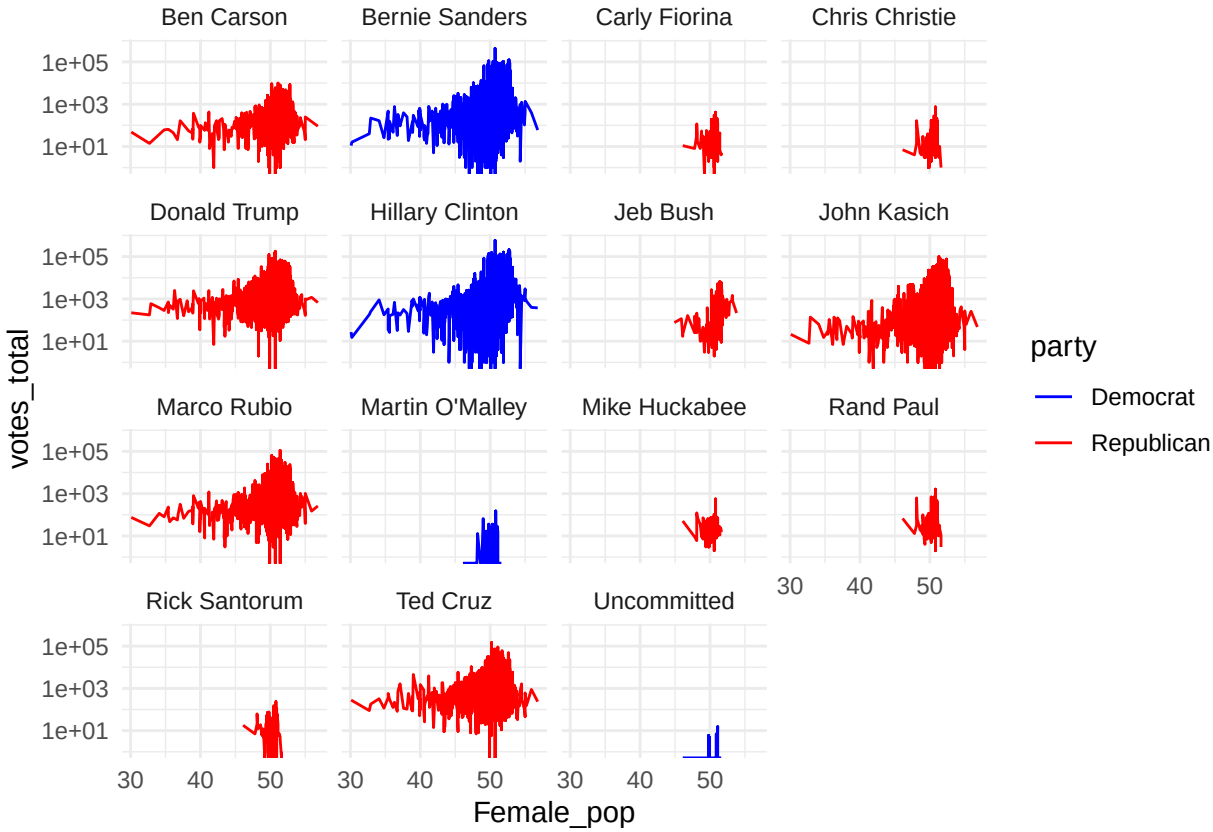
```
#Calculating the total votes received by a candidate

(total_votes2 <- primary_results %>% group_by(fips,party,candidate) %>% summarise(votes_total=sum(votes)
```

```
## `summarise()` has grouped output by 'fips', 'party'. You can override using the `.groups` argument.
```

```
## # A tibble: 24,521 x 4
## # Groups:   fips, party [7,773]
##     fips party      candidate       votes_total
##    <dbl> <chr>      <chr>                 <dbl>
##  1  1001 Democrat   Bernie Sanders          544
##  2  1001 Democrat   Hillary Clinton        2387
##  3  1001 Republican Ben Carson             1764
##  4  1001 Republican Donald Trump           5387
##  5  1001 Republican John Kasich             421
##  6  1001 Republican Marco Rubio            1785
##  7  1001 Republican Ted Cruz               2482
##  8  1003 Democrat   Bernie Sanders         2694
##  9  1003 Democrat   Hillary Clinton        5290
## 10  1003 Republican Ben Carson             4221
## # ... with 24,511 more rows
```

```
#Merging the total votes received by a candidate and the respective female population in that region

merged_dataset2 <- merge(x = female_ds,y= total_votes2, by = "fips")

#Facetting based on votes received by each candidate

merged_dataset2 %>% ggplot(aes(x=Female_pop,y=votes_total,color=party))  + geom_line() + facet_wrap(~ca
```

```
## Warning: Transformation introduced infinite values in continuous y-axis
```




```
#Answer: There is no evident pattern as male candidates have received similar number of votes as female

#The notion that female voters are more likely to vote for a female candidate was not accurate as per t
```