Data.World ArizonaVoters

Lev Williams

October 16, 2018

Connecting to API to obtain data

```
ArizonaVoters <- download_file_as_data_frame(
  "vlandry/arizona-registered-voters-by-county-january-2017",
  "Arizona voter registration jan 2017.csv")
```

```
##
  |
  |                                                                      |   0%
  |
  |======================================================================| 100%
```

Loading essential libraries

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 3.4.4
```

```
## -- Attaching packages --------------------------------- tidyverse 1.2.1 --
```

```
## v ggplot2 3.0.0     v purrr   0.2.5
## v tibble  1.4.2     v dplyr   0.7.4
## v tidyr   0.8.1     v stringr 1.2.0
## v readr   1.1.1     v forcats 0.3.0
```

```
## Warning: package 'ggplot2' was built under R version 3.4.4
```

```
## Warning: package 'tibble' was built under R version 3.4.4
```

```
## Warning: package 'tidyr' was built under R version 3.4.4
```

```
## Warning: package 'readr' was built under R version 3.4.4
```

```
## Warning: package 'purrr' was built under R version 3.4.4
```

```
## Warning: package 'dplyr' was built under R version 3.4.3
```

```
## Warning: package 'stringr' was built under R version 3.4.2
```

```
## Warning: package 'forcats' was built under R version 3.4.4
```

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

```
library(ggplot2)
library(lubridate)
```

```
## Warning: package 'lubridate' was built under R version 3.4.4
```

```
##
## Attaching package: 'lubridate'
```

```
## The following object is masked from 'package:base':
##
```

```
##     date
```

Data Cleanup

```
glimpse(ArizonaVoters)
```

```
## Observations: 45
## Variables: 9
## $ County      <fct> Apache, Apache, Apache, Cochise, Cochise, Cochise,...
## $ Precincts   <int> 45, 45, 45, 49, 49, 49, 71, 71, 71, 39, 39, 39, 22...
## $ Date.Period <fct> P.E. 2016, G.E. 2016, 17-Jan, P.E. 2016, G.E. 2016...
## $ Democratic  <fct> 27,114, 27,638, 28,363, 19,546, 20,121, 20,344, 29...
## $ Green       <fct> 42, 53, 55, 129, 147, 154, 231, 292, 307, 26, 34, ...
## $ Libertarian <fct> 145, 160, 162, 471, 560, 578, 722, 866, 864, 161, ...
## $ Republican  <fct> 8,070, 8,241, 8,534, 27,359, 28,296, 28,757, 18,89...
## $ Other       <fct> 12,277, 12,148, 12,361, 25,168, 26,033, 26,457, 23...
## $ TOTAL       <fct> 47,648, 48,240, 49,475, 72,673, 75,157, 76,290, 72...
```

Unfortunately the data in its current form has all the numerical data as factors, so we will need to transform them into numerical values before we can proceed with further analysis.

```
library(varhandle)
```

```
## Warning: package 'varhandle' was built under R version 3.4.4
```

```
facttonum <- function(x){
  return(as.numeric(gsub(",","",unfactor(x))))
}

for(i in 4:9){
  ArizonaVoters[,i] <- facttonum(ArizonaVoters[,i])
}
```

We can now procede with our initial data analysis. It will likely be better to segregate the data by Date.Period as it is essentially time data formatted as a factor variable, in fact it might be better to turn the factor variable into a date variable. According to the provider of the data, P.E. 2016 and G.E 2016 are the Presidental (im assuming Primary) and General elections for the year which happened on 3/22/2016 and 11/8/2016.

```
ArizonaVoters$Date.Period <- unfactor(ArizonaVoters$Date.Period)
a <- ArizonaVoters[1,3]
b <- ArizonaVoters[2,3]
for(i in 1:length(ArizonaVoters[,3])){
  if(ArizonaVoters[i,3] == a){
    ArizonaVoters[i,3] = "03-22-2016"
  }
  else if(ArizonaVoters[i,3] == b){
    ArizonaVoters[i,3] = "11-06-2016"
  }
  else{
    ArizonaVoters[i,3] = "01-01-2017"
  }
}
ArizonaVoters[1,3]
```

```
## [1] "03-22-2016"
```

```
ArizonaVoters$Date.Period <- as.Date(ArizonaVoters$Date.Period,format= "%m-%d-%Y")
```

In order to carry out an analysis properly we will need to turn the data into a "long form" format.

```
LongVoters <- ArizonaVoters %>% gather(key = Party, value = Members,
    Democratic, Green, Libertarian, Republican, Other)
```

Data Analysis

In order to see the overall trend of the state's voting population, we need to merge together the total voting population for each of the parties and see how they compare to one another from time stamp to time stamp.

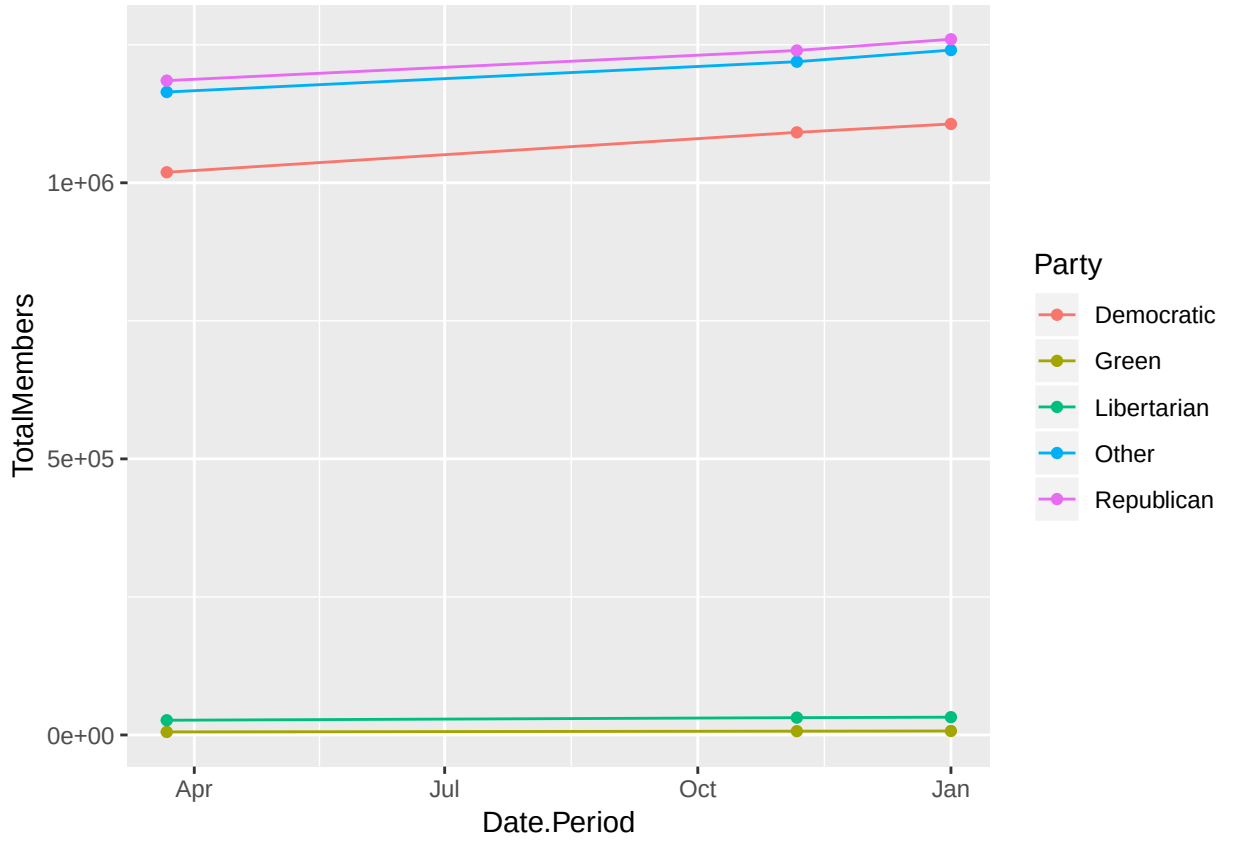
```
PartyTotal <- LongVoters %>% select(Date.Period, Party, Members) %>%
  group_by(Date.Period, Party) %>% summarise(TotalMembers = sum(Members))

PartyTotal
```

```
## # A tibble: 15 x 3
## # Groups:   Date.Period [?]
##    Date.Period Party       TotalMembers
##    <date>      <chr>              <dbl>
##  1 2016-03-22  Democratic      1019050.
##  2 2016-03-22  Green              5512.
##  3 2016-03-22  Libertarian       26653.
##  4 2016-03-22  Other           1164373.
##  5 2016-03-22  Republican      1185023.
##  6 2016-11-06  Democratic      1091323.
##  7 2016-11-06  Green              6894.
##  8 2016-11-06  Libertarian       31358.
##  9 2016-11-06  Other           1219277.
## 10 2016-11-06  Republican      1239614.
## 11 2017-01-01  Democratic      1106514.
## 12 2017-01-01  Green              7198.
## 13 2017-01-01  Libertarian       32191.
## 14 2017-01-01  Other           1240263.
## 15 2017-01-01  Republican      1259956.
```

From here we can construct a line graph to visually track the growth of each party.

```
PartyTotal %>% ggplot(aes(x=Date.Period,y=TotalMembers,color=Party)) +
  geom_point() + geom_line()
```



As it seems, the larger parties grow more quickly than the smaller ones however we may need to look into just how much each party grew relative to its size during the time from the Presidental Primary to January 1st, 2017.

```
propinc <- numeric(5)
for (i in 1:5) {
    propinc[i] <- (PartyTotal[i + 10, 3]/PartyTotal[i, 3]) -
        1
    propinc[i] <- round(propinc[[i]], 3) * 100
    print(paste("The", PartyTotal[i, 2], "Party grew by", propinc[i],
        "percent from the Presidental Primary to January 1, 2017."))
}
```

```
## [1] "The Democratic Party grew by 8.6 percent from the Presidental Primary to January 1, 2017."
## [1] "The Green Party grew by 30.6 percent from the Presidental Primary to January 1, 2017."
## [1] "The Libertarian Party grew by 20.8 percent from the Presidental Primary to January 1, 2017."
## [1] "The Other Party grew by 6.5 percent from the Presidental Primary to January 1, 2017."
## [1] "The Republican Party grew by 6.3 percent from the Presidental Primary to January 1, 2017."
```

So although it seemed like the larger parties grew even more quickly than the smaller ones, proportionally speaking, the smaller parties grew more quickly relative to their size.

Special Thanks to Victor Landry on data.world for providing the dataset.