

Assignment Week 1

rachelgreenlee

8/29/2020

Overview

In June of 2020 FiveThirtyEight published an article discussing how voter registration started out strong in early 2020, but dropped dramatically once COVID hit. This data set compares 2016 and 2020 voter registration for January through April or May (depending on locale) in 11 states and Washington DC. FiveThirtyEight obtained the data from the Center for Innovation and Research.

Link to the article: <https://fivethirtyeight.com/features/voter-registrations-are-way-way-down-during-the-pandemic/>

The Work

First I import the dataset that I've uploaded to my GitHub repository. These variables are well-named, but I'll rename two for practice.

```
reg_data <- read.csv(url("https://raw.githubusercontent.com/rachel-greenlee/538voter_registration/master/reg_data.csv"))
colnames(reg_data) <- c("State", "Year", "Month", "New Reg Voters")
colClasses = c("factor", "factor", "factor", "numeric")
```

Lets also write out the month names instead of the abbreviations for practice. At first I got some errors using this code, but then realized it was because these imported as factors after using the class function to check, so I switched the class to character and then could rename the columns. Here is a peek at the first few rows with these changes.

```
reg_data$Month <- as.character(reg_data$Month)
class(reg_data$Month)
```

```
## [1] "character"
```

```
reg_data$Month[reg_data$Month == "Jan"] <- "January"
reg_data$Month[reg_data$Month == "Feb"] <- "February"
reg_data$Month[reg_data$Month == "Mar"] <- "March"
reg_data$Month[reg_data$Month == "Apr"] <- "April"

head(reg_data)
```

```
##      State Year      Month New Reg Voters
## 1 Arizona 2016   January      25852
## 2 Arizona 2016  February      51155
## 3 Arizona 2016    March      48614
## 4 Arizona 2016    April      30668
## 5 Arizona 2020   January      33229
## 6 Arizona 2020  February      50853
```

Since there are only 4 variables, I don't need to select a subset of columns as the data would be pointless without all 4, but for practice let's remove the State variable and create the "sillysubset".

```
sillysubset <- subset(reg_data, select = c("Year", "Month", "New Reg Voters"))
head(sillysubset)
```

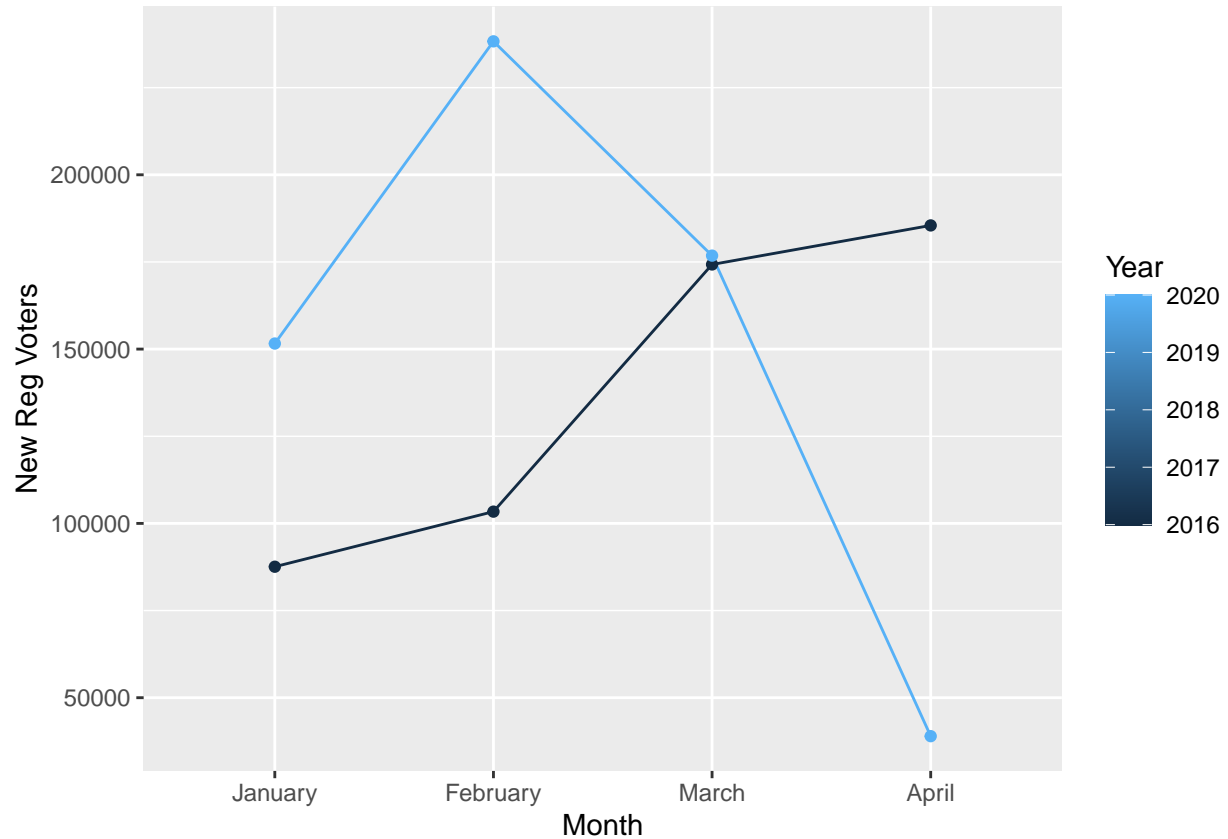
```
##      Year      Month New Reg Voters
## 1 2016   January      25852
## 2 2016  February      51155
## 3 2016    March      48614
## 4 2016    April      30668
## 5 2020   January      33229
## 6 2020  February      50853
```

Just to see if I could figure it out, I did a California only subset of the data and graphed it with 2 lines, one representing each year. I used ggplot2. I had to figure out how to put the months in order, which I found I could do by setting the levels of these factors. On the graph, you can see with the light blue line that in 2020 voter registration drops dramatically below 2016 from February to April.

```
reg_dataCA2020 <- subset(reg_data, State == "California")
head(reg_dataCA2020)
```

```
##      State Year      Month New Reg Voters
## 9 California 2016   January      87574
## 10 California 2016  February     103377
## 11 California 2016    March     174278
## 12 California 2016    April     185478
## 13 California 2020   January     151595
## 14 California 2020  February     238281
```

```
reg_dataCA2020$Month <- factor(reg_dataCA2020$Month, levels = c("January", "February", "March", "April"))
ggplot(data=reg_dataCA2020, aes(x=Month, y='New Reg Voters', group=Year)) +
  geom_line(aes(color=Year)) +
  geom_point(aes(color=Year))
```



Conclusions

When this article was published in June I'm sure the April/May data was the most recent available, but now that we are much closer to the election it would be a natural additions to add on the New Voter Registration numbers for these same states up until July/August if possible. Expanding to include more states could also be beneficial.

Considering that COVID is still limiting many traditional voter registration efforts even as we enter September I'd imagine the registrations are still dampened when compared to 2016, though at the same time there has been many political protests in the past months that may have motivated more people to register. I'd be very vurious to see an updated dataset!