My analysis revolved around streaming tweets based on fire words like “feelthebern” or “bernie2016”. By choosing strictly these tweets we gather a sense of comfort on whoever the author of the tweets must be a supporter of the candidate in their hashtag. So I wanted to prove or gain insight on where each candidate was more popular in the southern states or northern states.

My thoughts are pretty straightforward and simple but still a good way to establish a test. Going by American history class in high school, it’s pretty obvious how some states will most likely come out voting for GOP in primary elections. States like South Carolina, Texas are firm GOP states which are strongly red states. Likewise states in the north like NY, Michigan, and New Hampshire are firmly Democratic, blue states.

I wanted to take equal amount of tweets for Bernie supporters and Trump supporters and map them by longitude and latitude on the U.S Map. First they appeared to be overlapped, then I went ahead and did two more maps separately.

$H_0 : p \leq 0.5$
$H_\mathrm{a} : p > 0.5$

Where Bernie is popular in the Northern States.

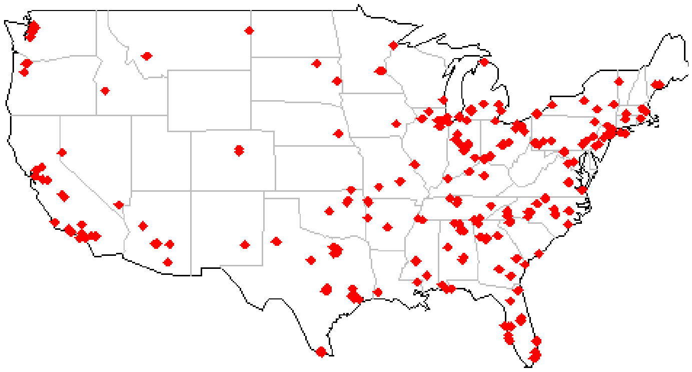
$H_0 : p \leq 0.5$
$H_\mathrm{a} : p > 0.5$

Where Donald is popular in the Southern States.

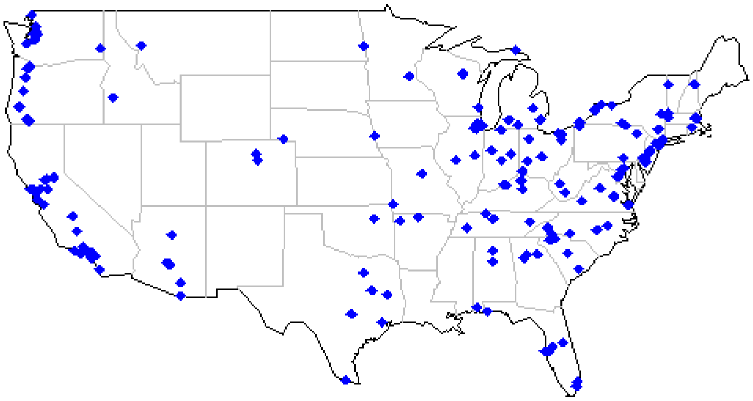
I will use R and Python languages to stream twitter for tweets with a set of hashtags for Bernie and for Donald.

After gaining around 500 tweets each, I know the population is kind of small. However, given the time frame I managed to gather this many from the program I wrote.

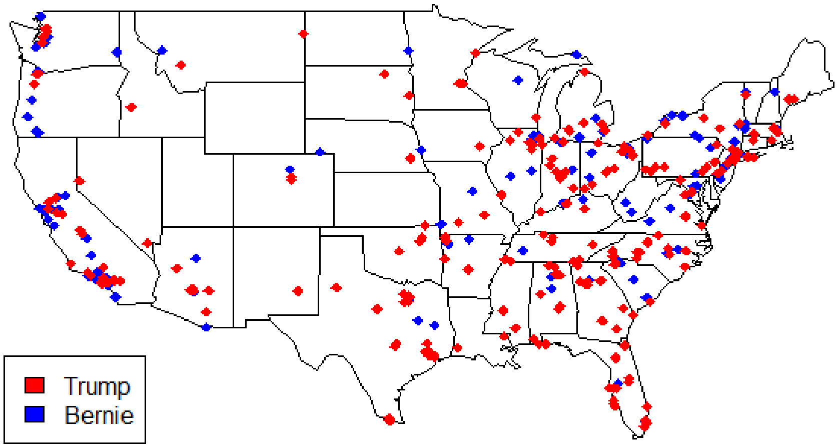
Trump, 2016

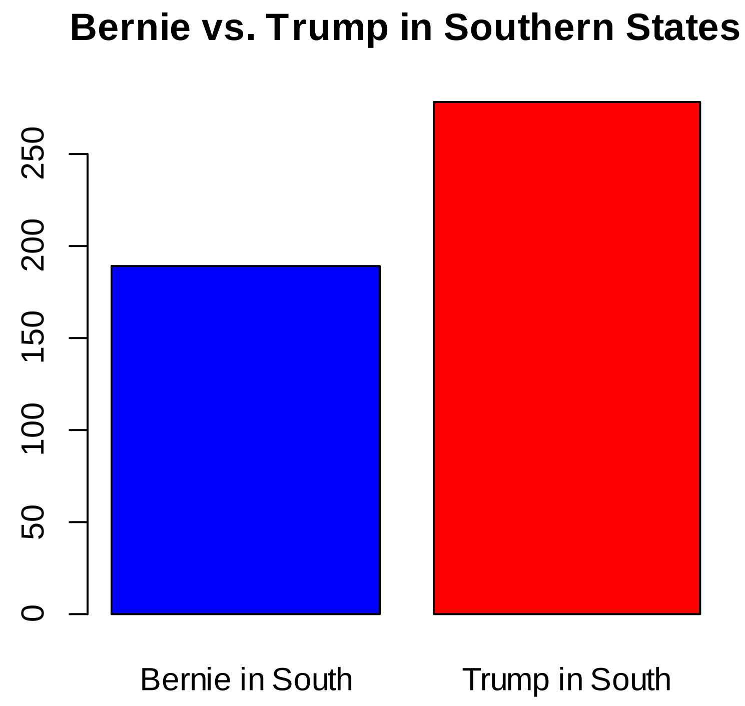


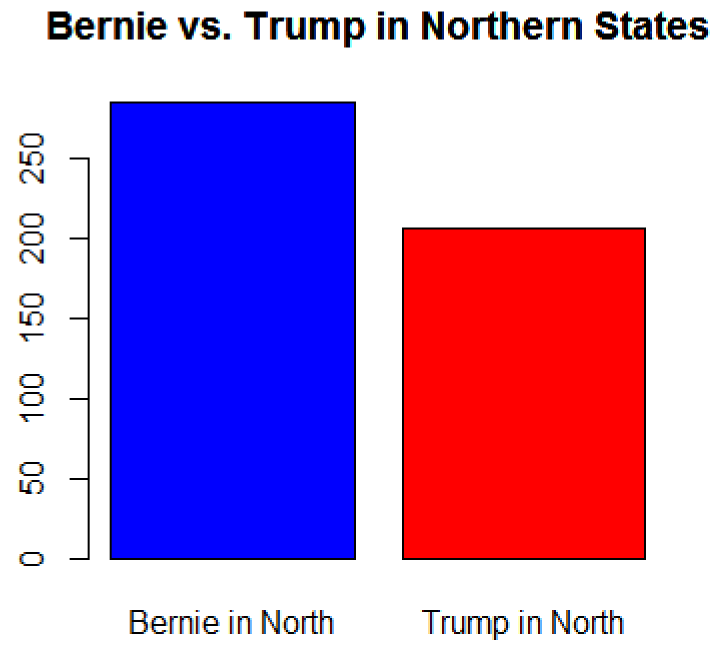
Bernie, 2016



Bernie vs. Trump, 2016




Bernie vs. Trump in Southern States
0
50
100
150
200
250
Bernie in South
Trump in South


Bernie vs. Trump in Northern States
0
50
100
150
200
250
Bernie in North
Trump in North

```
# Souther States on count of postive tweets

>  sum_below_line_trump_south = sum(mydata_trump$lat < 38)
>  sum_below_line_trump_south
[1] 278

> sum_below_line_bernie_south = sum(mydata_bernie$lat < 38)
> sum_below_line_bernie_south
[1] 189

# Northern States on count of postive tweets

> sum_above_line_trump_north = sum(mydata_trump$lat > 38)
> sum_above_line_trump_north
[1] 206

> sum_above_line_bernie_north = sum(mydata_bernie$lat > 38)
> sum_above_line_bernie_north
[1] 284
```

```
prop.test(sum_below_line_trump_south, n = trump_tweets_total, p = .5, alternative
="greater"
#one-sample test
1-sample proportions test with continuity correction

data:  sum_below_line_trump_south out of 484, null probability 0.5
X-squared = 10.4153, df = 1, p-value = 0.0006249
alternative hypothesis: true p is greater than 0.5
95 percent confidence interval:
 0.5360614 1.0000000
sample estimates:
        p
0.5743802
```

```
>Because the p-value is less than .000629 the significance level of .01 ,there is enough
evidence to claim that in the southern states Trump is more popular which was divided by
divided by latitude 38 and below(southern states)
```

```
> prop.test(sum_above_line_bernie_north, n = bernie_tweets_total, p = .5, alternative =
"greater")

        1-sample proportions test with continuity correction

data:  sum_above_line_bernie_north out of bernie_tweets_total, null probability 0.5
X-squared = 18.6808, df = 1, p-value = 7.727e-06
alternative hypothesis: true p is greater than 0.5
95 percent confidence interval:
 0.5618409 1.0000000
sample estimates:
        p
0.6004228
```

```
Because the p-value is less than ( 7.727e-06) the significance level of .01 ,there is
enough evidence to claim that in the Northern  states Bernie is more popular which was
divided by divided by latitude 38 and above(Northern states)
```