Hate Crimes in the US after the 2016 Presidential Election

Mariel Pacada, Jason Cheong

10/11/2019

Introduction

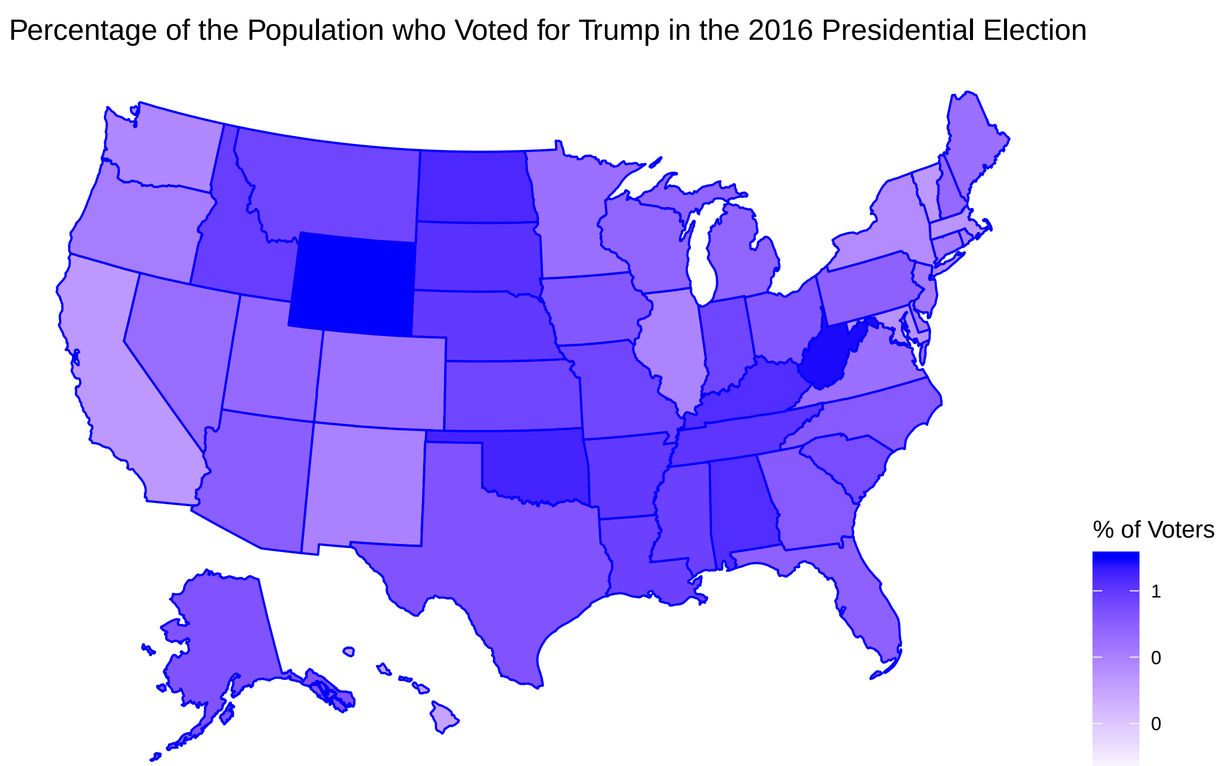
This dataset describes the prevalence of hate crimes in the United States in the context of the most recent presidential election. THe dataset includes several characteristics of each state's population. In this exploratory data analysis, we explore how such characteristics are related with how a state voted for Trump.

Part 1: Macro Analysis

How did the United States vote for Trump?

In this visualization, we take a look at how each state voted for Trump in the 2016 Presidential Election. The darker the purple hue, the bigger proportion of the state's population voted for Trump.

```r
plot_usmap(data = hate_crimes, values = "share_voters_voted_trump", color = "blue") +
  scale_fill_continuous(low = "white", high = "blue", name = "% of Voters", label = scales::comma) +
  theme(legend.position = "right") +
  labs(title = "Percentage of the Population who Voted for Trump in the 2016 Presidential Election")
```

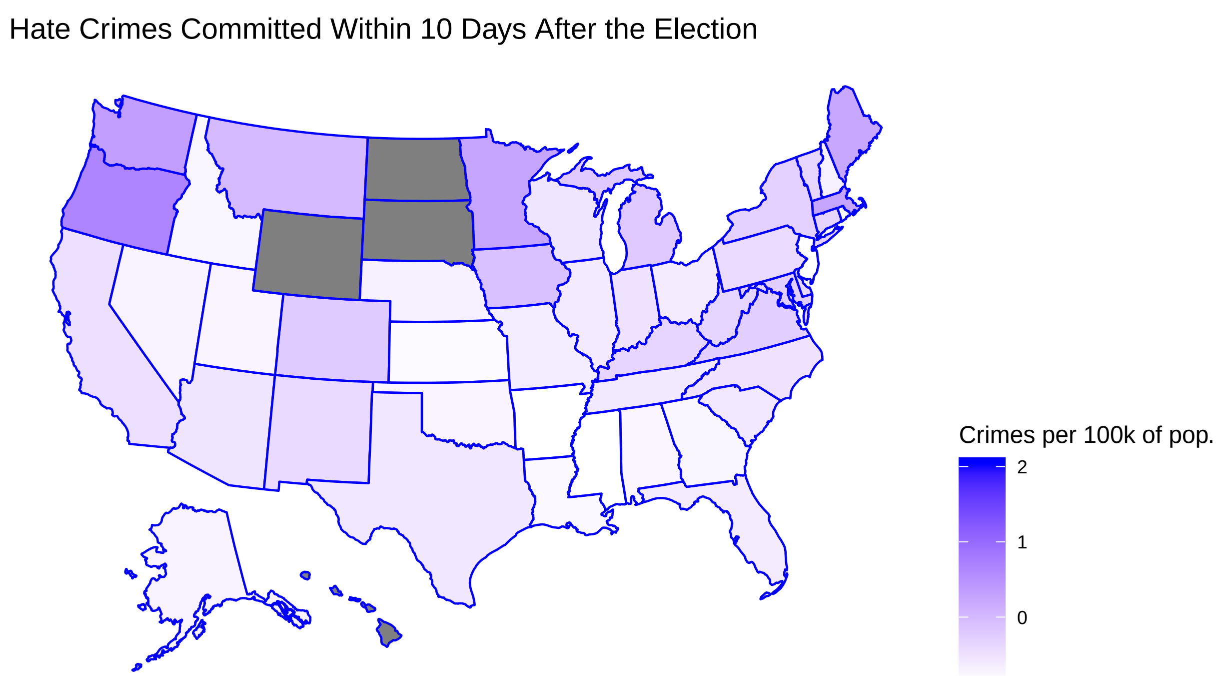


Visually, we can discern that Wyoming and its neighboring states central states voted for Trump at high proportions. Most southern states similarly voted for Trump en masse.

How does Trump voting behavior compare to hate crime behavior?

Here, we look at the hate crime data relative to each state. The hate crime index was collected by the Southern Poverty Law Center, a nonprofit legal advocacy organization specializing in civil rights, based on the hate crimes committed between November 9th and 18th of 2016 (days following the election on November 8th).

```
plot_usmap(data = hate_crimes, values = "hate_crimes_per_100k_splc", color = "blue") +
  scale_fill_continuous(low = "white", high = "blue", name = "Crimes per 100k of pop.", label = scales:
  theme(legend.position = "right") +
  labs(title = "Hate Crimes Committed Within 10 Days After the Election")
```

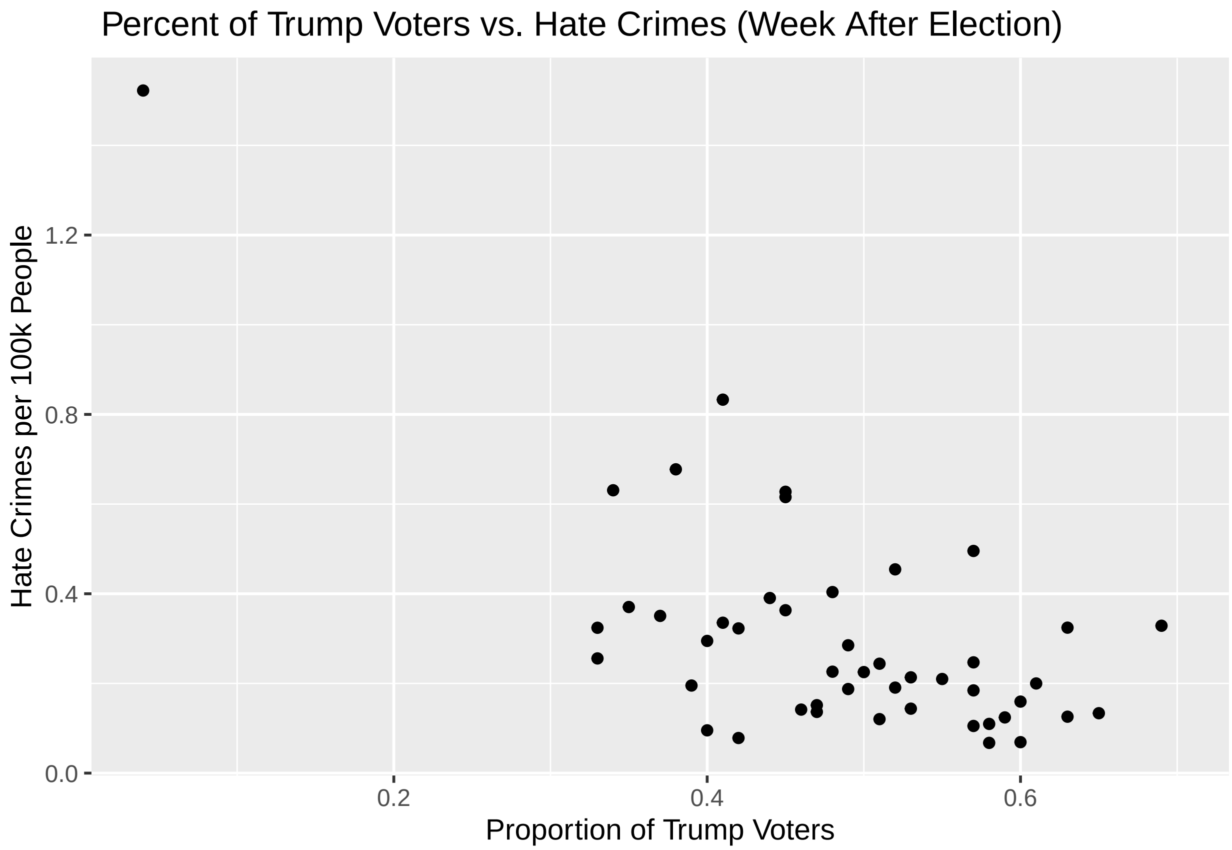


(Note: The states with a gray fill – Hawaii, Wyoming, North Dakota, and South Dakota – do not have data for hate crimes.)

Interestingly, if we compare the hues of the first map with that of the second, we can see that the states who voted for Trump at the highest percentages experienced the least hate crime after he was elected. Below, we confirm this with finding the correlation coefficient.

```r
ggplot(data = hate_crimes, aes(x = share_voters_voted_trump, y = hate_crimes_per_100k_splc)) + geom_poin
  labs(title = " Percent of Trump Voters vs. Hate Crimes (Week After Election)",
  x = "Proportion of Trump Voters", y = "Hate Crimes per 100k People")
```

```
## Warning: Removed 4 rows containing missing values (geom_point).
```




```r
cor(hate_crimes$share_voters_voted_trump, hate_crimes$hate_crimes_per_100k_splc, use = "complete.obs")
```

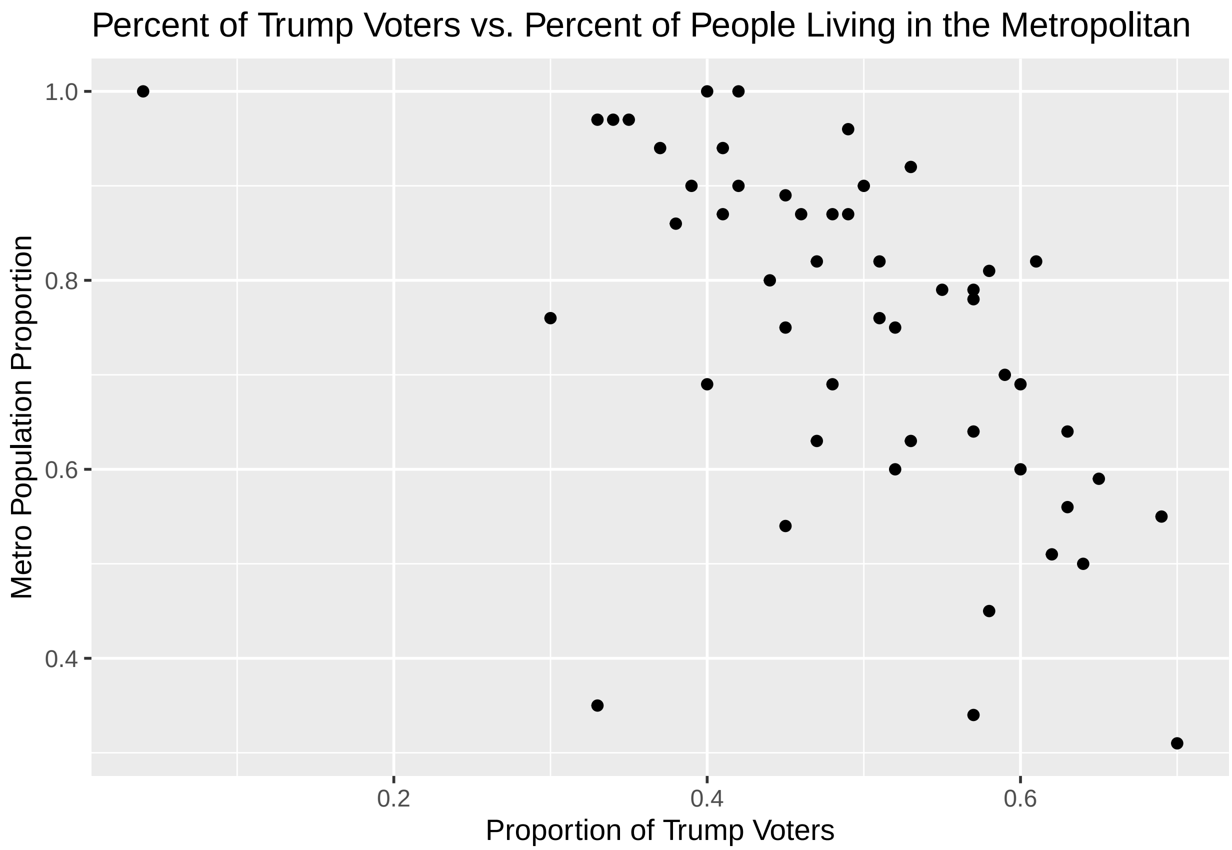
```
## [1] -0.6570672
```

Using the correlation function, we can see that the the percentage of Trump voters in a state and the number of hate crimes per 100,000 people has a slightly strong negative correlation coefficient of **-0.657**. Along with our visual evidence with the US maps above, this numerical evidence suggests that states with greater percentages of Trump voters is less likely to commit hate crime.

Part 2: Micro Analysis

This section will explore two more variables that have a considerable correlation with the percentage of Trump voters per state: **metropolitan population proportion** and **proportion of the population that is white and living under poverty**.

```r
ggplot(hate_crimes, aes(x = share_voters_voted_trump, y = share_population_in_metro_areas)) + geom_poin
  labs(title = "Percent of Trump Voters vs. Percent of People Living in the Metropolitan",
  x = "Proportion of Trump Voters", y = "Metro Population Proportion")
```

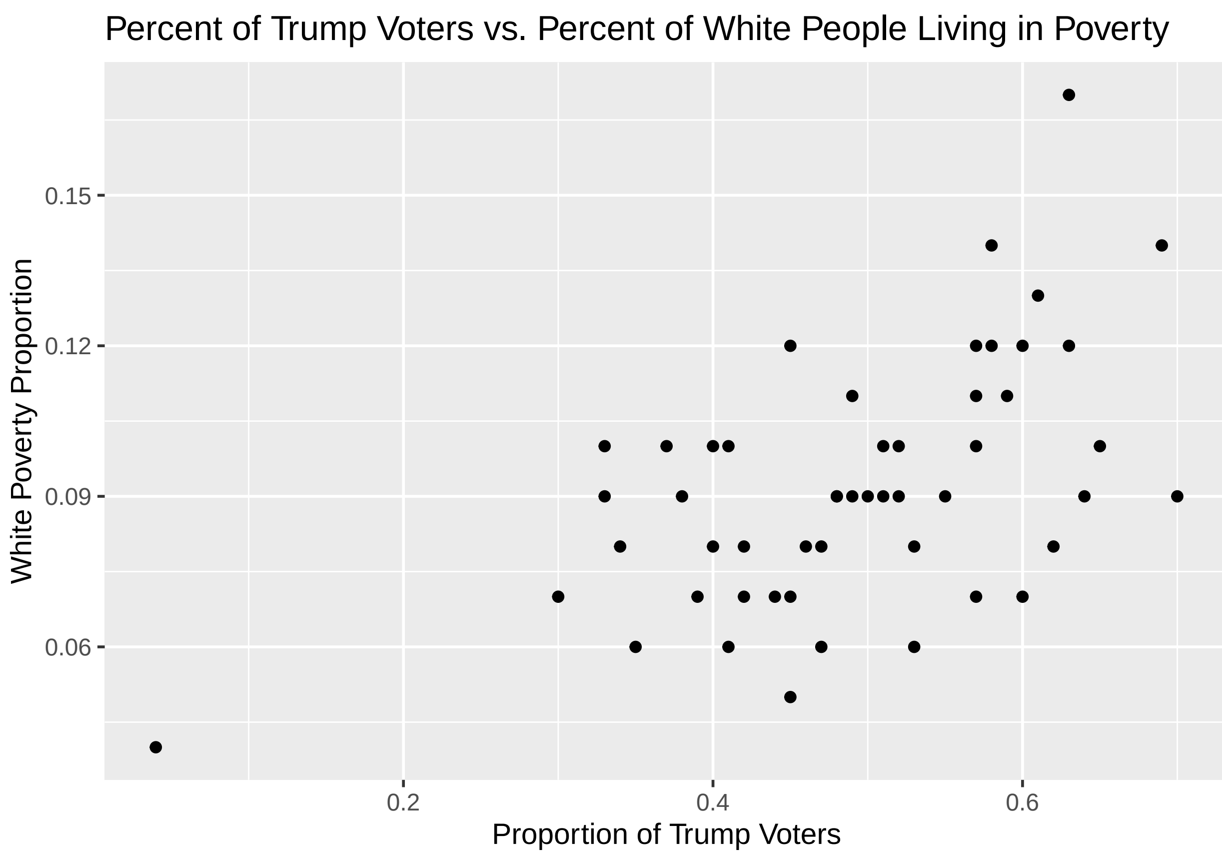



```r
cor(hate_crimes$share_voters_voted_trump, hate_crimes$share_population_in_metro_areas)
```

```
## [1] -0.5706947
```

This scatterplot plots the proportion of Trump voters against the proportion of the state's population that resides in the metropolitan. Visually, we can see a slightly strong negative correlation, and we confirm with the cor() function that the correlation coefficient between the two is **-0.57**. Note that this is for proportion of the population living in metropolitan areas. In turn, we can assume that the correlation between the share of Trump voters and a state's population that resides in **rural** areas is a slightly strong positive correlation. Therefore, we have evidence to suggest that people living in rural areas are more likely to vote for Trump than those living in metropolitan areas.

```r
ggplot(hate_crimes, aes(x = share_voters_voted_trump, y = share_white_poverty)) + geom_point() +
  labs(title = "Percent of Trump Voters vs. Percent of White People Living in Poverty", x = "Proportion
  y = "White Poverty Proportion")
```




```
cor(hate_crimes$share_voters_voted_trump, hate_crimes$share_white_poverty)
```

```
## [1] 0.5528492
```

This plot shows the proportion of white people in the state living in poverty as a function of the state's proportion of Trump voters. Similarly with the previous scatterplot, we see a slightly strong correlation – this time, positive. We use the cor() function and gather that these two variables have a correlation coefficient of **0.553**. Thus, we can say that states with more people living in white poverty are more likely to vote for Trump at high proportions.

Part 3: Case Studies

Looking at the two scatterplots above, we can easily determine one obvious outlier in each. These two dots actually correspond to the same region: District of Columbia. Below, we examine the characteristics of Washington, D.C. across all given variables.

Case Study 1: Washington, D.C.

```
dc_ranked_trump <- hate_crimes %>%
                      select(state, share_voters_voted_trump) %>%
                      arrange(desc(share_voters_voted_trump))

dc_ranked_hate <- hate_crimes %>%
```

```r
                    select(state, hate_crimes_per_100k_splc) %>%
                    arrange(desc(hate_crimes_per_100k_splc))
```

D.C. voted for Trump in the least proportion.

```r
tail(dc_ranked_trump)
```

```
##                   state share_voters_voted_trump
## 46             Maryland                     0.35
## 47        Massachusetts                     0.34
## 48           California                     0.33
## 49              Vermont                     0.33
## 50               Hawaii                     0.30
## 51 District of Columbia                     0.04
```

D.C. experienced the most hate crimes.

```r
head(dc_ranked_hate)
```

```
##                  state hate_crimes_per_100k_splc
## 1 District of Columbia                 1.5223017
## 2               Oregon                 0.8328496
## 3           Washington                 0.6774876
## 4        Massachusetts                 0.6308106
## 5            Minnesota                 0.6274799
## 6                Maine                 0.6155740
```

Despite having the **lowest share of Trump Voters**, Washington D.C. has the **highest number of hate crimes**, during the week after Election 2016. Why might this be the case?

Let us examine what other variables (not mentioned in Part 2) may be significant in explaining why D.C. has the highest number of hate crimes.

```r
dc_ranked_income <- hate_crimes %>%
                      select(state, median_household_income) %>%
                      arrange(desc(median_household_income))

dc_ranked_employ <- hate_crimes %>%
                      select(state, share_unemployed_seasonal) %>%
                      arrange(desc(share_unemployed_seasonal))

dc_ranked_citizen <- hate_crimes %>%
                       select(state, share_non_citizen) %>%
                       arrange(desc(share_non_citizen))

dc_ranked_white <- hate_crimes %>%
                     select(state, share_non_white) %>%
                     arrange(desc(share_non_white))

dc_ranked_poverty <- hate_crimes %>%
                       select(state, share_white_poverty) %>%
                       arrange(desc(share_white_poverty))
```

```r
dc_ranked_gini <- hate_crimes %>%
                    select(state, gini_index) %>%
                    arrange(desc(gini_index))
```

D.C. is the 5th richest region.

```r
head(dc_ranked_income)
```

```
##                  state median_household_income
## 1             Maryland                   76165
## 2        New Hampshire                   73397
## 3               Hawaii                   71223
## 4          Connecticut                   70161
## 5 District of Columbia                   68277
## 6               Alaska                   67629
```

D.C. is the 3rd most unemployed region.

```r
head(dc_ranked_employ)
```

```
##                  state share_unemployed_seasonal
## 1        West Virginia                     0.073
## 2           New Mexico                     0.068
## 3 District of Columbia                     0.067
## 4               Nevada                     0.067
## 5               Alaska                     0.064
## 6              Arizona                     0.063
```

D.C. has the 2nd highest non-citizen population percentage.

```r
head(dc_ranked_citizen)
```

```
##                  state share_non_citizen
## 1           California              0.13
## 2 District of Columbia              0.11
## 3           New Jersey              0.11
## 4                Texas              0.11
## 5              Arizona              0.10
## 6               Nevada              0.10
```

D.C. has the 2nd highest non-white population percentage.

```r
head(dc_ranked_white)
```

```
##                  state share_non_white
## 1               Hawaii            0.81
## 2 District of Columbia            0.63
## 3           New Mexico            0.62
## 4           California            0.61
## 5                Texas            0.56
## 6             Maryland            0.50
```

D.C. has the smallest percentage of white people living in poverty.

```
tail(dc_ranked_poverty)
```

```
##                   state share_white_poverty
## 46               Alaska                0.06
## 47          Connecticut                0.06
## 48             Maryland                0.06
## 49        New Hampshire                0.06
## 50            Minnesota                0.05
## 51 District of Columbia                0.04
```

D.C. experiences the most income inequality.

```
head(dc_ranked_gini)
```

```
##                  state gini_index
## 1 District of Columbia      0.532
## 2             New York      0.499
## 3          Connecticut      0.486
## 4            Louisiana      0.475
## 5        Massachusetts      0.475
## 6              Florida      0.474
```

Based on the above, we can see that Washington, D.C. has significant rankings in relation to US state in the variables of **median household income**, **unemployment**, **metropolitan population**, **non-citizen population**, **population of white people living under poverty**, and **Gini index**. These chracteristics of Washington, D.C. provide good exploratory evidence to potentially proceed to study how several of these variables *work together* to create conditions that ultimately affect how a state votes during a big election.

Case Study 2: West Virginia

While D.C. was the region to vote for Trump in the least proportion, Wyoming was the one who voted in the highest proportion. However, Wyoming did not have complete data in this dataset, so we take a look at the state who voted for Trump in the second highest percentage: West Virginia.

```
wv_ranked_trump <- hate_crimes %>%
                     select(state, share_voters_voted_trump) %>%
                     arrange(desc(share_voters_voted_trump))
```

WV voted for Trump in the second highest proportion.

```
head(wv_ranked_trump)
```

```
##           state share_voters_voted_trump
## 1       Wyoming                     0.70
## 2 West Virginia                     0.69
## 3      Oklahoma                     0.65
## 4  North Dakota                     0.64
## 5       Alabama                     0.63
## 6      Kentucky                     0.63
```

Similarly with our exploratory data analysis on D.C., we take a look at how West Virgina compares with the other state in regards to the dataset's variables.

```
wv_ranked_income <- hate_crimes %>%
                     select(state, median_household_income) %>%
                     arrange(desc(median_household_income))

wv_ranked_employ <- hate_crimes %>%
                     select(state, share_unemployed_seasonal) %>%
                     arrange(desc(share_unemployed_seasonal))

wv_ranked_citizen <- hate_crimes %>%
                      select(state, share_non_citizen) %>%
                      arrange(desc(share_non_citizen))

wv_ranked_white <- hate_crimes %>%
                    select(state, share_non_white) %>%
                    arrange(share_non_white)

wv_ranked_poverty <- hate_crimes %>%
                      select(state, share_white_poverty) %>%
                      arrange(desc(share_white_poverty))
```

WV is the 2nd poorest state.

```
tail(wv_ranked_income)
```

```
##            state median_household_income
## 46     Tennessee                   43716
## 47      Kentucky                   42786
## 48     Louisiana                   42406
## 49       Alabama                   42278
## 50 West Virginia                   39552
## 51   Mississippi                   35521
```

WV is the most unemployed state.

```
head(wv_ranked_employ)
```

```
##                  state share_unemployed_seasonal
## 1        West Virginia                     0.073
## 2           New Mexico                     0.068
## 3 District of Columbia                     0.067
## 4               Nevada                     0.067
## 5               Alaska                     0.064
## 6              Arizona                     0.063
```

WV has the lowest non-citizen population percentage (tied with VT and MT).

```
tail(wv_ranked_citizen)
```

```
##            state share_non_citizen
## 46       Montana              0.01
## 47       Vermont              0.01
## 48 West Virginia              0.01
## 49         Maine                NA
## 50   Mississippi                NA
## 51  South Dakota                NA
```

WV has the 2nd highest white population.

```r
head(wv_ranked_white) #2nd most white
```

```
##           state share_non_white
## 1       Vermont            0.06
## 2 West Virginia            0.07
## 3         Maine            0.09
## 4 New Hampshire            0.09
## 5       Montana            0.10
## 6          Iowa            0.15
```

WV is the 2nd poorest state (tied with MS).

```r
head(wv_ranked_poverty) #2nd poorest (tie with ms)
```

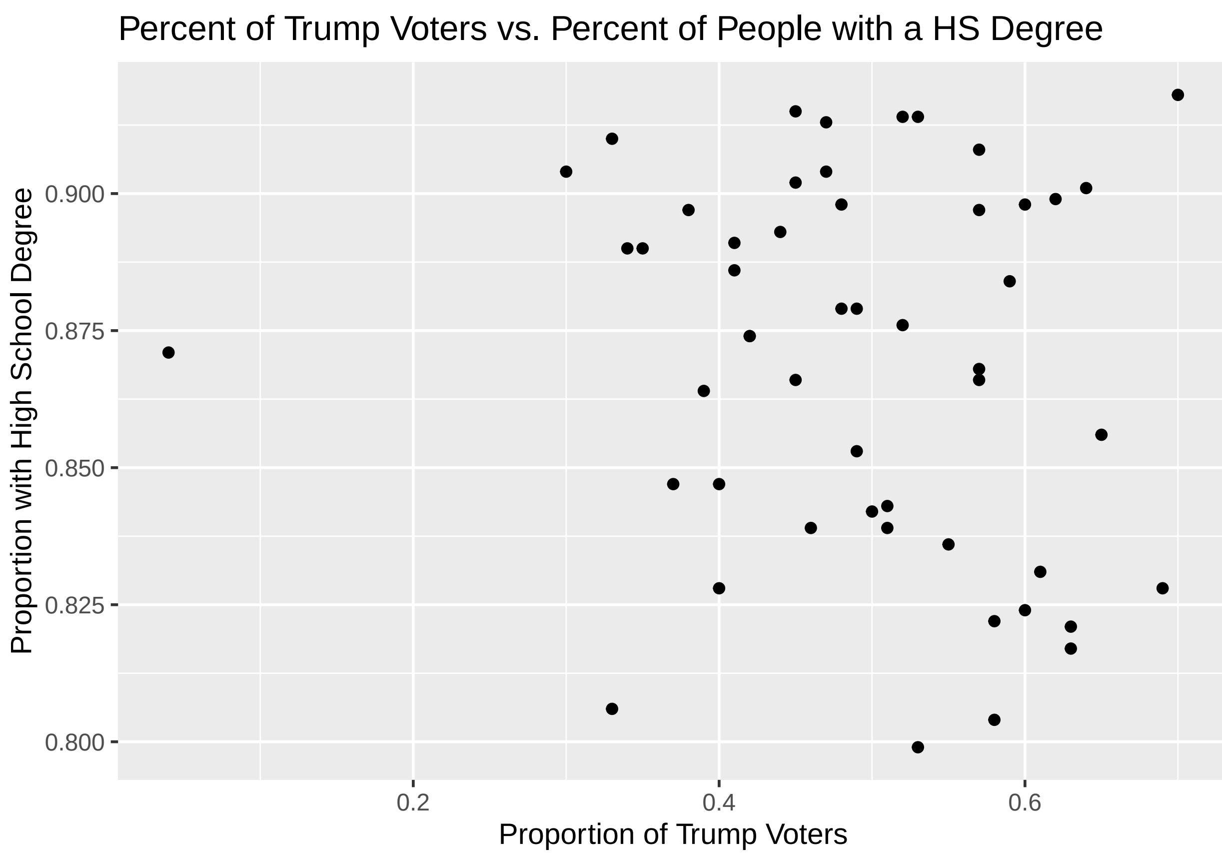
```
##           state share_white_poverty
## 1      Kentucky                0.17
## 2   Mississippi                0.14
## 3 West Virginia                0.14
## 4     Tennessee                0.13
## 5       Alabama                0.12
## 6      Arkansas                0.12
```

Like D.C., West Virginia had significant rankings in the variables **median household income**, **unemployment**, **non-citizen population**, **white population**, and **population of white people living under poverty**. Looking at these two case studies may provide insight on further analysis on how the variables interact with each other.

Concluding Thoughts

While our exploration of the data above may suggest relationships among several variables, we must prevent ourselves from forming biases based on these alone. As we conclude, let us quickly examine the variables **proportion of the population with a high school degree** and **proportion of the population that is unemployed**.

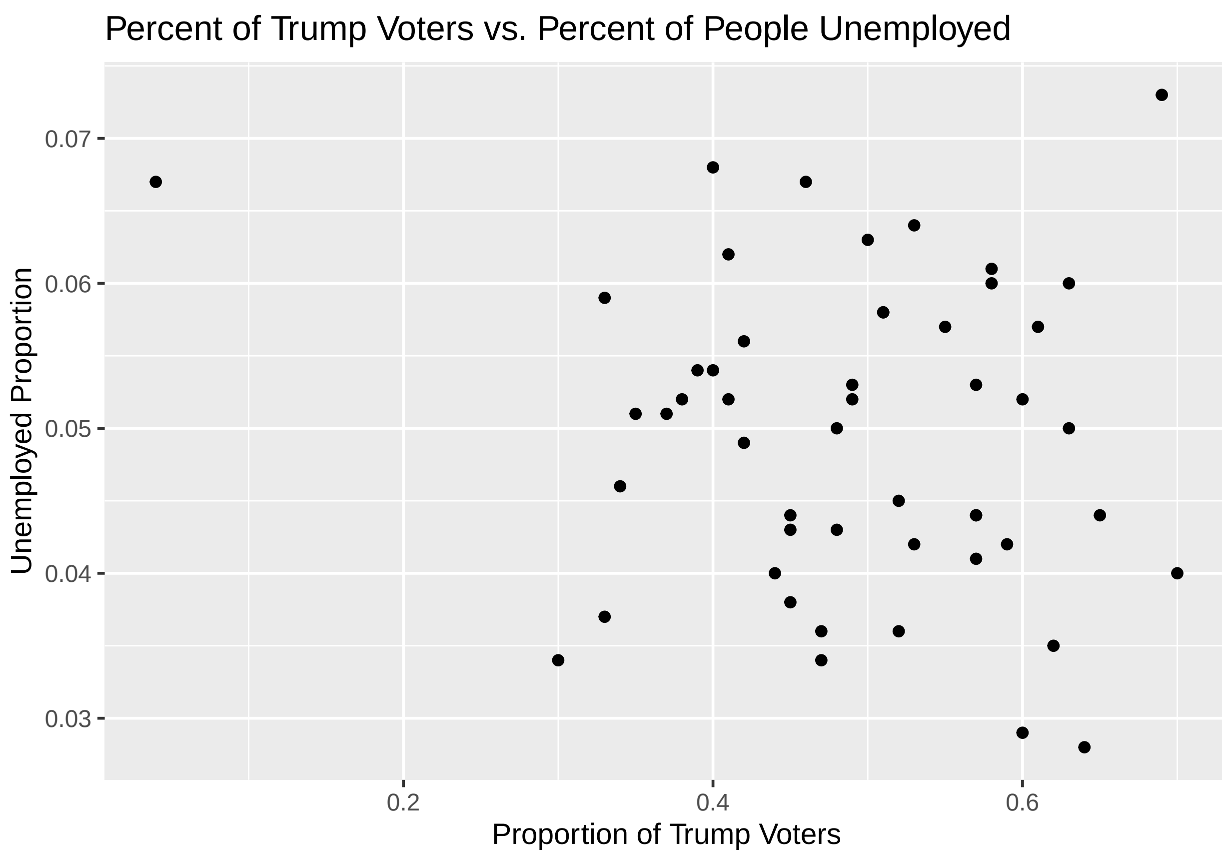
```r
ggplot(hate_crimes, aes(x = share_voters_voted_trump, y = share_population_with_high_school_degree)) +
  geom_point() + labs(title = "Percent of Trump Voters vs. Percent of People with a HS Degree",
  x = "Proportion of Trump Voters", y = "Proportion with High School Degree")
```




```r
cor(hate_crimes$share_voters_voted_trump, hate_crimes$share_population_with_high_school_degree)
```

```
## [1] -0.1556557
```

```r
ggplot(hate_crimes, aes(x = share_voters_voted_trump, y = share_unemployed_seasonal)) + geom_point() +
  labs(title = "Percent of Trump Voters vs. Percent of People Unemployed",
  x = "Proportion of Trump Voters", y = "Unemployed Proportion")
```




```r
cor(hate_crimes$share_voters_voted_trump, hate_crimes$share_unemployed_seasonal)
```

```
## [1] -0.1481921
```

Any discourse surrounding Donald Trump and his unanticipated rise to presidency is controversial, to say the least. Regardless of position on the political compass, communities can easily become echo chambers for biases against their respective opposing sides. For example, some liberal-leaning communities might stereotype a Trump voter as too uneducated to vote for another candidate or too desperate to find a job and in turn trusts a candidate who promises him or her one. However, from above we see that the correlation coefficient between the percentage of Trump voters and *not only* **percentage of the population with a high school degree** *but also* **percentage of the population that is unemployed** is very low. This revelation trumps (pun definitely intended) the idea that the typical Trump voter is of one particular identity.

Exploratory data analysis is a powerful approach to telling stories and sharing information with the general public. The creator of a data visualization has full control over how to present a insight from a dataset. As data scientists, we must remember to practice ethical exploration of data and strive to paint a full picture always.