Ahsanul_Choudhury_DATA_609_Project

Ahsanul Choudhury

May 15, 2018

Contents

Introduction

According to a report titled "The number of hate crimes rose in 2016" published on CNN.com (https://www.cnn.com/2017/11/13/politics/hate-crimes-fbi-2016-rise/index.html) on November 13, 2017, hate crimes in the United States have increased to a point not seen in recent history. They showed some statistics from Federal Bureau of Investigation and used graphics to back their claim. On a graph the visualized how hate crime spiked around the time of November 2016 election. One of the reason put forward for this is share of proverty in the majority race member of the population and which was blamed on influx of immigrant by some political leaders during campaigning.

My Goal for this project is to collect data form various sources and construct a mathematical model to see the relationship between the share of poverty in white population and reported hate crimes in each state.

Data Collection and Preparation

To collect my data I have used the following three sources:

https://ucr.fbi.gov/hate-crime/2016/tables/table-11

https://www.census.gov/data/datasets/2016/demo/popest/state-total.html

https://github.com/fivethirtyeight/data/blob/master/hate-crimes/hate_crimes.csv

Data from these source downloaded cleaned and combined to create my final dataset. All the information for the state of Hawaii was not available so it was left out for this project. The following table shows the final dataset; the first column represent the states, second column represents the Share of white residents who are living in poverty and the third column represents the number of hate crimes reported by FBI for the year of 2016. I chose to convert the reported hate crime for every 100,000 of population due large variance in total population in each state.

```
#Download and read data
url <- "https://github.com/choudhury1023/DATA-609/raw/master/Project/table_11_offenses_offense_type_by_
GET(url, write_disk("2016_data.xls", overwrite=TRUE))
```

```
## Response [https://raw.githubusercontent.com/choudhury1023/DATA-609/master/Project/table_11_offenses_
##   Date: 2018-05-16 23:37
##   Status: 200
##   Content-Type: application/octet-stream
##   Size: 41 kB
## <ON DISK>  C:\Users\ahsan\Documents\DATA 609\Project\2016_data.xls
```

```
df <- read_excel("2016_data.xls")

#Data Cleanup and manipulation

df <- df[-c(1:6, 57:60), ]

df1 <- df[ ,-c(3:17)]

colnames(df1) <- c("state","hate.crimes")

write.csv(df1, "df1.csv")
```

```
url <- c("https://github.com/choudhury1023/DATA-609/raw/master/Project/nst-est2016-01.xlsx")
GET(url, write_disk("population.xlsx", overwrite=TRUE))
```

```
## Response [https://raw.githubusercontent.com/choudhury1023/DATA-609/master/Project/nst-est2016-01.xlsx
##   Date: 2018-05-16 23:37
##   Status: 200
##   Content-Type: application/octet-stream
##   Size: 18.3 kB
## <ON DISK>  C:\Users\ahsan\Documents\DATA 609\Project\population.xlsx
```

```
df6 <- read_excel("population.xlsx")

df6 <- df6[-c(1:8, 20, 60:67), ]
df6 <- df6[-c(2:9)]
colnames(df6) <- c("state","population")

write.csv(df6, "pop.csv")
```

```
raw <- getURL("https://raw.githubusercontent.com/choudhury1023/DATA-609/master/Project/hate_crimes.csv")
df7 <- read.csv(text = raw)
df7 <- df7[-c(2:6, 8:12)]
df7 <- df7[-c(12), ]



df8 <- bind_cols(df1, df6, df7)
df8 <- transform(df8, hate.crimes = as.numeric(hate.crimes))
df8$avg_hatecrimes_per_100k_fbi_2016 <- (df8$hate.crimes/df8$population)*100000
df8 <- transform(df8, share_white_poverty = as.numeric(share_white_poverty))
df8 <- transform(df8, avg_hatecrimes_per_100k_fbi_2016 = as.numeric(avg_hatecrimes_per_100k_fbi_2016))
final_df <- df8[-c(2:5)]
final_df1 <- df8[-c(2:5)]
final_df2 <- df8[-c(2:5)]

knitr::kable(final_df)
```

state	share_white_poverty	avg_hatecrimes_per_100k_fbi_2016
Alabama	0.12	0.3495569
Alaska	0.06	1.7522719
Arizona	0.09	4.1984853
Arkansas	0.12	0.5019664
California	0.09	2.9095529
Colorado	0.07	2.2741445
Connecticut	0.06	3.4671233
Delaware	0.08	1.9956621
District of Columbia	0.04	18.6443913
Florida	0.11	0.5336583
Georgia	0.09	0.4461527
Idaho	0.11	2.0794467
Illinois	0.07	0.9530104
Indiana	0.12	2.4573903
Iowa	0.09	0.6699221
Kansas	0.11	2.4077414
Kentucky	0.17	5.4316298
Louisiana	0.12	0.7048773
Maine	0.12	3.3797003
Maryland	0.06	0.6648442
Massachusetts	0.08	6.5181210

state	share_white_poverty	avg_hatecrimes_per_100k_fbi_2016
Michigan	0.09	4.6231480
Minnesota	0.05	2.9891564
Mississippi	0.14	0.3345907
Missouri	0.07	1.8709995
Montana	0.10	1.9184284
Nebraska	0.07	1.8352318
Nevada	0.08	1.6666338
New Hampshire	0.06	2.9967149
New Jersey	0.07	3.4434688
New Mexico	0.10	1.7299251
New York	0.10	3.0285705
North Carolina	0.10	2.0499098
North Dakota	0.09	1.0554758
Ohio	0.10	4.3738909
Oklahoma	0.10	0.8920468
Oregon	0.10	2.9803602
Pennsylvania	0.09	0.5240833
Rhode Island	0.08	1.2305642
South Carolina	0.09	0.4636051
South Dakota	0.08	2.4264721
Tennessee	0.13	2.8566300
Texas	0.08	0.8039452
Utah	0.08	2.3597142
Vermont	0.10	4.1627041
Virginia	0.07	1.7832076
Washington	0.09	6.6136114
West Virginia	0.14	2.6213723
Wisconsin	0.09	0.7268061
Wyoming	0.09	0.6831756

Model Construction

Geometric Similarity Model

Assumptions:

We assume that the share of white poverty is proportional to any characteristic dimension cubed.

$$V \alpha l^3$$

If we assume a constant share of white poverty, then the share state's number of hate crimes reported is proportional to it's share of white poverty,

$$V \alpha W$$

$$V \alpha W \alpha l^3$$

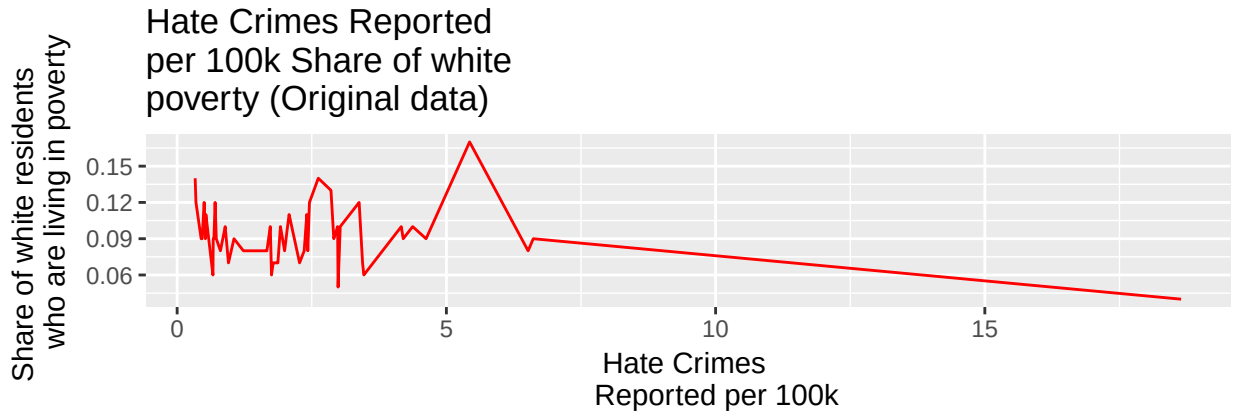
Let the characteristic dimension of l be the number of hate crime reported, which is chosen because it is reported as an indicator of share of white poverty.

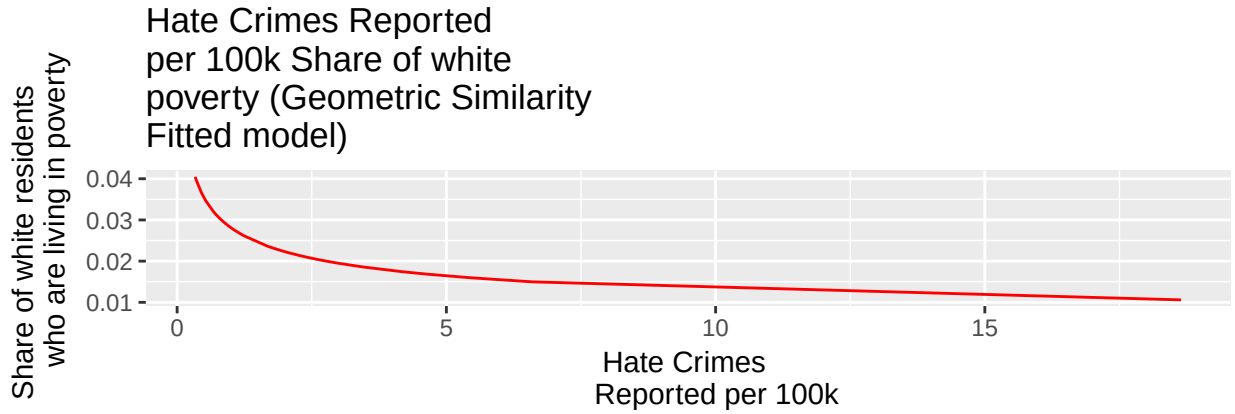
$$W = kl^3, k > 0$$

```
final_df$z1 = c(NA,tail(final_df$share_white_poverty,-1)-head(final_df$share_white_poverty,-1))
final_df$z2= c(NA,tail(final_df$avg_hatecrimes_per_100k_fbi_2016,-1)-head(final_df$avg_hatecrimes_per_1

# finding the slope
final_df$slope<- final_df$z1 / final_df$z2

slope<- mean(final_df$slope, na.rm= TRUE)
slope
```

```
## [1] 0.02809108
```

```
# calculating f(x) using the slope
final_df$fx<- slope*(final_df$avg_hatecrimes_per_100k_fbi_2016)^(-1/3)


p1<- ggplot(data = df, aes(x = final_df$avg_hatecrimes_per_100k_fbi_2016, y = final_df$share_white_pove
  labs(x = "Hate Crimes
       Reported per 100k", y = "Share of white residents
       who are living in poverty") +
  geom_line(color = "red")+
  ggtitle("Hate Crimes Reported
per 100k Share of white
poverty (Original data)")

p2<- ggplot(data = df, aes(x = final_df$avg_hatecrimes_per_100k_fbi_2016, y = final_df$fx)) +
  labs(x = "Hate Crimes
       Reported per 100k", y = "Share of white residents
       who are living in poverty") +
  geom_line(color = "red")+
  ggtitle("Hate Crimes Reported
per 100k Share of white
poverty (Geometric Similarity
Fitted model)")

grid.arrange(p1, p2, nrow=2, ncol=1)
```





Least-Squares Criterion Model

Our formula for Least-Squares Criterion Model:

$$a = \frac{\Sigma x_i^n y_i}{\Sigma x_i^{2n}}$$

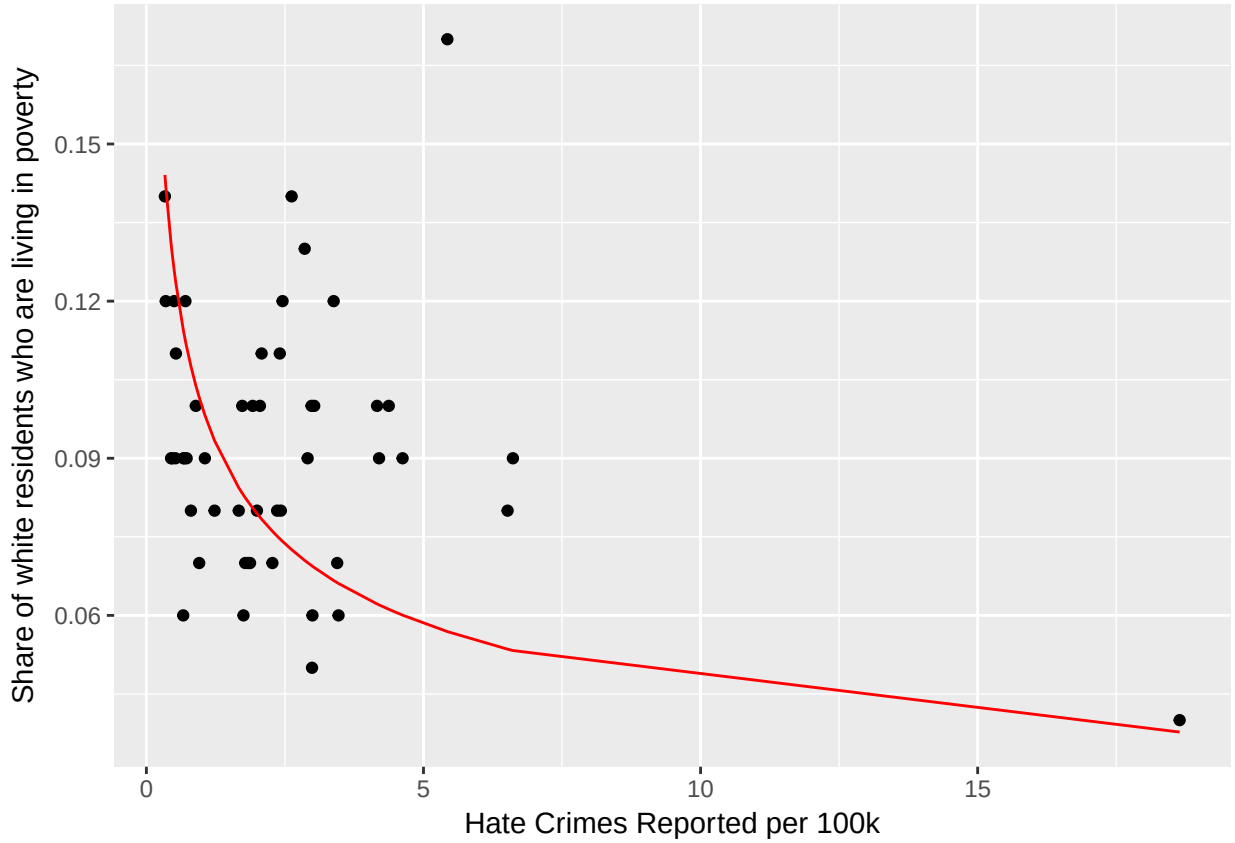
Where x = avg_hatecrimes_per_100k_fbi_2016 and y = share_white_poverty

n is fixed to given collection of data points

```
a <- sum(final_df1$avg_hatecrimes_per_100k_fbi_2016^(-1/3)*final_df1$share_white_poverty)/sum(final_df1$
a
```

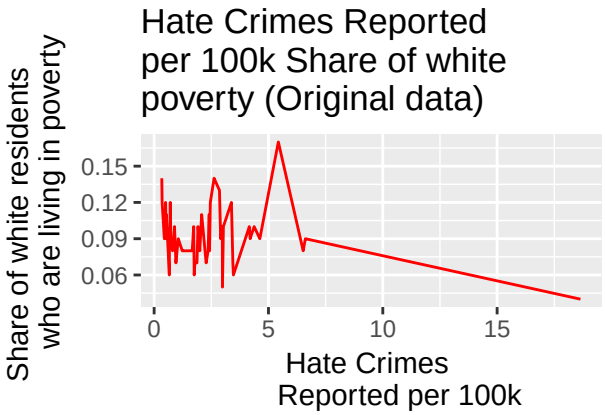
```
## [1] 0.1000343
```

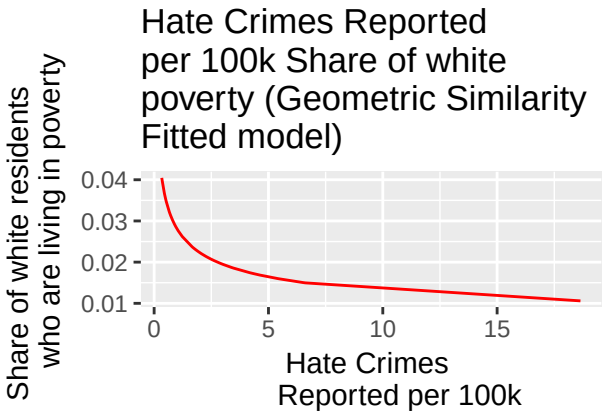
```
final_df1$share_white_poverty_pred <- a * final_df1$avg_hatecrimes_per_100k_fbi_2016^(-1/3)


ggplot(final_df1, aes(x = avg_hatecrimes_per_100k_fbi_2016, y = share_white_poverty)) + geom_point() +
  geom_line(aes(x = avg_hatecrimes_per_100k_fbi_2016, y = share_white_poverty_pred), color = "red") +
  labs(x = "Hate Crimes Reported per 100k", y = "Share of white residents who are living in poverty")
```

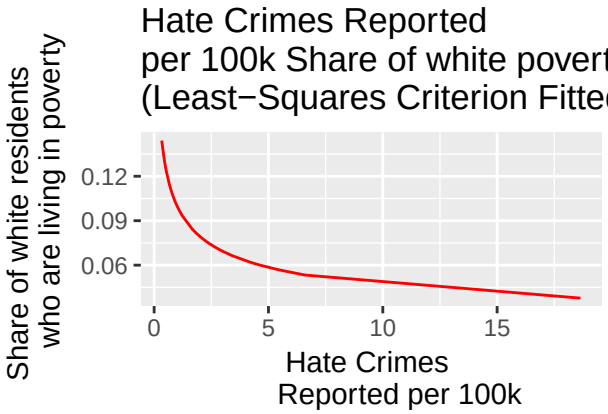


Conclusion

```
p3<- ggplot(data = final_df1, aes(x = final_df1$avg_hatecrimes_per_100k_fbi_2016, y = final_df1$share_wh
  labs(x = "Hate Crimes
       Reported per 100k", y = "Share of white residents
       who are living in poverty") +
  geom_line(color = "red")+
  ggtitle("Hate Crimes Reported
per 100k Share of white poverty
(Least-Squares Criterion Fitted model)")

grid.arrange(p1, p2, p3, nrow=2, ncol=2)
```







The models do not fit tightly with our observed data. Hate crimes alone can not be a predictor of poverty in a portion of a state's population and poverty by itself can not be the sole contributing factor to hate crime. Also, our data have some outliers, for example District of Columbia has a large number of hate crimes reported in 2016 and has a relatively low share of poverty compared to other states. There are other socioeconomic factors which may be a better indicator of poverty in a segment of population like level of education or combination of various other factor. We started the project to see the relationship between hate crimes reported in a state and share of poverty in a segment of population and we are concluding this project by saying the relationship is complex and not truly linear.

Reference

Giordano, F. R., Fox, W. P., & Horton, S. B. (2014). A first course in mathematical modeling. Australia: Brooks/Cole, Cengage Learning.

Table 11. (2017, November 03). Retrieved from https://ucr.fbi.gov/hate-crime/2016/tables/table-11

https://www.census.gov/data/datasets/2016/demo/popest/state-total.html

https://github.com/fivethirtyeight/data/blob/master/hate-crimes/hate_crimes.csv