Assignment 2

Jai Chi Cham

2022-09-05

## Section 1: Description of the data

The source of this dataset is from Southern Poverty Law Centerand FBI. The variables includes median household income, unemployed rate, population and etc . The hypothesis of this analysis is Higher rates of hate crimes are tied to income inequality. The research questions includes: 1. what is the key factor causing the hate crime. 2.Does the political ideology lead to hate crimes in a specific state? 3. What is the crime rate before and after the election, does it change? The file format is csv. It is a delimited file and it uses comma to separate the values. The databases are not fixed width since they are saved as CSV, and CSV files can’t have a fixed length of their columns.

## Section 2 : Reading the data into R

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

#using read.csv to read data from csv file from a URL  
hate\_crimes <- read.csv("https://raw.githubusercontent.com/fivethirtyeight/data/master/hate-crimes/hate\_crimes.csv")

## Section 3 : clean the data

#Remove gini\_index since it is there is share\_of\_poverty, so it is more or less duplicate fields.  
subset1 = select(hate\_crimes,-gini\_index)  
  
#Sorts the states by Median household income  
sort\_subset1 = arrange(subset1,median\_household\_income)  
hate\_crimes = sort\_subset1

## Section 4 : Characteristics of data

This dataframe has 51 rows and 11 columns. The names of the columns and a brief description of each are in the table below:

#this makes a new data.frame called text\_tbl with two columns, Names and Description  
library(knitr)  
columns\_summary <- data.frame(  
Columns = c(colnames(hate\_crimes)),  
Description = c("State","Median household income, 2016", "Share of the population that is unemployed (seasonally adjusted), Sept. 2016","Share of the population that lives in metropolitan areas, 2015","Share of the population that are not U.S. citizens, 2015","Share of white residents who are living in poverty, 2015","Gini Index, 2015","Share of the population that is not white, 2015","Share of 2016 U.S. presidential voters who voted for Donald Trump","Hate crimes per 100,000 population, Southern Poverty Law Center, Nov. 9-18, 2016","Average annual hate crimes per 100,000 population, FBI, 2010-2015")  
 )  
kable(columns\_summary, caption = "Columns Description")

Columns Description

| Columns | Description |
| --- | --- |
| state | State |
| median\_household\_income | Median household income, 2016 |
| share\_unemployed\_seasonal | Share of the population that is unemployed (seasonally adjusted), Sept. 2016 |
| share\_population\_in\_metro\_areas | Share of the population that lives in metropolitan areas, 2015 |
| share\_population\_with\_high\_school\_degree | Share of the population that are not U.S. citizens, 2015 |
| share\_non\_citizen | Share of white residents who are living in poverty, 2015 |
| share\_white\_poverty | Gini Index, 2015 |
| share\_non\_white | Share of the population that is not white, 2015 |
| share\_voters\_voted\_trump | Share of 2016 U.S. presidential voters who voted for Donald Trump |
| hate\_crimes\_per\_100k\_splc | Hate crimes per 100,000 population, Southern Poverty Law Center, Nov. 9-18, 2016 |
| avg\_hatecrimes\_per\_100k\_fbi | Average annual hate crimes per 100,000 population, FBI, 2010-2015 |

## Section 5 : Summary statistics

hc\_subset = select(hate\_crimes,"median\_household\_income","share\_unemployed\_seasonal","share\_population\_in\_metro\_areas")

subset\_Summary <-summary(hc\_subset)   
subset\_Summary

## median\_household\_income share\_unemployed\_seasonal  
## Min. :35521 Min. :0.02800   
## 1st Qu.:48657 1st Qu.:0.04200   
## Median :54916 Median :0.05100   
## Mean :55224 Mean :0.04957   
## 3rd Qu.:60719 3rd Qu.:0.05750   
## Max. :76165 Max. :0.07300   
## share\_population\_in\_metro\_areas  
## Min. :0.3100   
## 1st Qu.:0.6300   
## Median :0.7900   
## Mean :0.7502   
## 3rd Qu.:0.8950   
## Max. :1.0000