ANA 515 Assignment 2

Abhijeet

18 June 2022

Section 1

The dataset is a combination of hate crimes data from FBI and Southern Poverty Lawa Center. The data was collected on the basis of key socioeconomic factors for each state, including education, diversity, geographic heterogeneity, economic health, political ideology. The data is measuring the gate crimes per 100k population, and the other columns take into consideration the socioeconomic factors affecting this target variables.

The research questions that can be asked with the data are 1) Why are the hate crimes not uniformly distributed across the United States? 2) What factors influnec the hate crime numbers to rise? 3) Does the poiltical ideology affect the amount of hate crimes in a specific state? 4) Do results vary based on the data collected before and after elections?

The file format is csv. Yes, it is a delimited file which uses comma to seperate the values.

Section 2

```
#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

```
#Removing share_population_with_high_school_degree since the it is amost same for all the rows and the column data is from 2009 which does not match the year rest of data was collected
subset1 = select(hate_crimes,-share_population_with_high_school_degree)

#Sorts the states by Median household income
sort_subset1 = arrange(subset1,median_household_income)

hate_crimes  = sort_subset1
```

Section 4

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
df_table <- data.frame (Name  = c("state",
```

```
"median_household_income","share_unemployed_seasonal","share_population_in_metro_areas","share_non_citizen","share_white_poverty","gini_index","share_non_white","share_voters_voted_trump","hate_crimes_per_100k_splc","avg_hatecrimes_per_100k_fbi"),
                         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(df_table, caption = "Columns Description")  #prints the table
```

Columns Description

Name	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_non_citizen	Share of the population that are not U.S. citizens, 2015
share_white_poverty	Share of white residents who are living in poverty, 2015
gini_index	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

```
hc_subset =
select(hate_crimes,"median_household_income","share_unemployed_seasonal","share_population_in_metro_areas")

subset_Summary <-summary(hc_subset) #creates the summary
subset_Summary #prints the summary in your output
```

```
##  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
```