

Assignment 2

Alex Zhang

11/14/2021

Section 1

The dataset contains the hate crimes per 100000 population in different states, as well as other datapoints such as household income, unemployment rate, urban population share, high school graduate share, race and etc. The research question that I try to answer is that what are the major factors that contribute to higher hate crimes. Data file type is CSV. It's delimited and the delimiter is comma.

Section 2

Will use read_csv function to read in the comma delimited data. It's from the Tidyverse package.

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.0.5
```

```
## -- Attaching packages ----- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5      v purrr   0.3.4
## v tibble  3.1.0      v dplyr  1.0.7
## v tidyr   1.1.3      v stringr 1.4.0
## v readr   1.4.0      v forcats 0.5.1
```

```
## Warning: package 'ggplot2' was built under R version 4.0.5
```

```
## Warning: package 'dplyr' was built under R version 4.0.5
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

```
hc <- read_csv("https://raw.githubusercontent.com/fivethirtyeight/data/master/hate-crimes/hate_crimes.csv")
```

```
##
## -- Column specification -----
## cols(
##   state = col_character(),
##   median_household_income = col_double(),
##   share_unemployed_seasonal = col_double(),
```



```
## -- Column specification -----
## cols(
##   '# Hate Crimes' = col_character()
## )

## Warning: 14 parsing failures.
## row col expected actual
## 2 -- 1 columns 2 columns 'https://raw.githubusercontent.com/fivethirtyeight/data/master/hate-crime'
## 3 -- 1 columns 2 columns 'https://raw.githubusercontent.com/fivethirtyeight/data/master/hate-crime'
## 4 -- 1 columns 2 columns 'https://raw.githubusercontent.com/fivethirtyeight/data/master/hate-crime'
## 5 -- 1 columns 2 columns 'https://raw.githubusercontent.com/fivethirtyeight/data/master/hate-crime'
## 6 -- 1 columns 2 columns 'https://raw.githubusercontent.com/fivethirtyeight/data/master/hate-crime'
## ... ..
## See problems(...) for more details.
```

```
kable(readme[4:15,], caption = "Data Dictionary")
```

Table 1: Data Dictionary

Hate Crimes
state
median_household_income
share_unemployed_seasonal
share_population_in_metro_areas
share_population_with_high_school_degree
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

```
##cannot get the second column to show
```

```
##Section 5
```

```
keep2 <- hc %>%
  select(gini_index,median_household_income,share_unemployed_seasonal)

stats <- summary(keep2)

print(stats)
```

```
##      gini_index      median_household_income      share_unemployed_seasonal
##  Min.      :0.4190      Min.      :35521      Min.      :0.02800
##  1st Qu.:0.4400      1st Qu.:48657      1st Qu.:0.04200
##  Median :0.4540      Median :54916      Median :0.05100
##  Mean   :0.4538      Mean   :55224      Mean   :0.04957
##  3rd Qu.:0.4665      3rd Qu.:60719      3rd Qu.:0.05750
##  Max.   :0.5320      Max.   :76165      Max.   :0.07300
```