Assignment 2

Mayuree

2022-06-15

Section 1: Data Description The Hatecrime Dataset entails the details of the people who died due to Hate crime. It represents Median_household_Income which represents Median household Income,2021.It also represents Share of the population that is unemployed (seasonally adjusted) & metropolitan areas.Also, Share of adults 25 and older with a high-school degree & Share of the population that are not U.S. citizens Research Question is- Whether High Rates of Crimes is dependent to Income Inequality. Is there a correlation between median_household_income & hate_crimes_per_100k_splc Format of the File.This data is saved in CSV file format in github. This is a flat file. This is a delimited file & comma is the delimiter used in the file.

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

Section 2: Reading the data into R

Section 3: Clean the data

```
library(tidyverse)
```

Renaming column names with upper case.

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

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.8
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1

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

```
data <- data %>%
  rename_with(toupper)
```

```
data %>%
  rename(
    Hate_Crime = HATE_CRIMES_PER_100K_SPLC,
    Avg_HateCrimes = AVG_HATECRIMES_PER_100K_FBI
  )
```

Renaming columns to Hate_Crime and AVg_HateCrimes

```
##                   STATE MEDIAN_HOUSEHOLD_INCOME SHARE_UNEMPLOYED_SEASONAL
## 1               Alabama                  42278                       0.060
## 2                Alaska                  67629                       0.064
## 3               Arizona                  49254                       0.063
## 4              Arkansas                  44922                       0.052
## 5            California                  60487                       0.059
## 6              Colorado                  60940                       0.040
## 7           Connecticut                  70161                       0.052
## 8              Delaware                  57522                       0.049
## 9  District of Columbia                  68277                       0.067
## 10              Florida                  46140                       0.052
## 11              Georgia                  49555                       0.058
## 12               Hawaii                  71223                       0.034
## 13                Idaho                  53438                       0.042
## 14             Illinois                  54916                       0.054
## 15              Indiana                  48060                       0.044
## 16                 Iowa                  57810                       0.036
## 17               Kansas                  53444                       0.044
## 18             Kentucky                  42786                       0.050
## 19            Louisiana                  42406                       0.060
## 20                Maine                  51710                       0.044
## 21             Maryland                  76165                       0.051
## 22        Massachusetts                  63151                       0.046
## 23             Michigan                  52005                       0.050
## 24            Minnesota                  67244                       0.038
## 25          Mississippi                  35521                       0.061
## 26             Missouri                  56630                       0.053
## 27              Montana                  51102                       0.041
## 28             Nebraska                  56870                       0.029
## 29               Nevada                  49875                       0.067
## 30        New Hampshire                  73397                       0.034
## 31           New Jersey                  65243                       0.056
## 32           New Mexico                  46686                       0.068
## 33             New York                  54310                       0.051
## 34       North Carolina                  46784                       0.058
## 35         North Dakota                  60730                       0.028
## 36                 Ohio                  49644                       0.045
## 37             Oklahoma                  47199                       0.044
## 38               Oregon                  58875                       0.062
## 39         Pennsylvania                  55173                       0.053
## 40         Rhode Island                  58633                       0.054
## 41       South Carolina                  44929                       0.057
## 42         South Dakota                  53053                       0.035
## 43            Tennessee                  43716                       0.057
```

```
## 44                Texas                  53875                      0.042
## 45                 Utah                  63383                      0.036
## 46              Vermont                  60708                      0.037
## 47             Virginia                  66155                      0.043
## 48           Washington                  59068                      0.052
## 49        West Virginia                  39552                      0.073
## 50            Wisconsin                  58080                      0.043
## 51              Wyoming                  55690                      0.040
##    SHARE_POPULATION_IN_METRO_AREAS SHARE_POPULATION_WITH_HIGH_SCHOOL_DEGREE
## 1                             0.64                                    0.821
## 2                             0.63                                    0.914
## 3                             0.90                                    0.842
## 4                             0.69                                    0.824
## 5                             0.97                                    0.806
## 6                             0.80                                    0.893
## 7                             0.94                                    0.886
## 8                             0.90                                    0.874
## 9                             1.00                                    0.871
## 10                            0.96                                    0.853
## 11                            0.82                                    0.839
## 12                            0.76                                    0.904
## 13                            0.70                                    0.884
## 14                            0.90                                    0.864
## 15                            0.79                                    0.866
## 16                            0.60                                    0.914
## 17                            0.64                                    0.897
## 18                            0.56                                    0.817
## 19                            0.81                                    0.822
## 20                            0.54                                    0.902
## 21                            0.97                                    0.890
## 22                            0.97                                    0.890
## 23                            0.87                                    0.879
## 24                            0.75                                    0.915
## 25                            0.45                                    0.804
## 26                            0.78                                    0.868
## 27                            0.34                                    0.908
## 28                            0.60                                    0.898
## 29                            0.87                                    0.839
## 30                            0.63                                    0.913
## 31                            1.00                                    0.874
## 32                            0.69                                    0.828
## 33                            0.94                                    0.847
## 34                            0.76                                    0.843
## 35                            0.50                                    0.901
## 36                            0.75                                    0.876
## 37                            0.59                                    0.856
## 38                            0.87                                    0.891
## 39                            0.87                                    0.879
## 40                            1.00                                    0.847
## 41                            0.79                                    0.836
## 42                            0.51                                    0.899
## 43                            0.82                                    0.831
## 44                            0.92                                    0.799
## 45                            0.82                                    0.904
```

```
## 46                          0.35                                  0.910
## 47                          0.89                                  0.866
## 48                          0.86                                  0.897
## 49                          0.55                                  0.828
## 50                          0.69                                  0.898
## 51                          0.31                                  0.918
##    SHARE_NON_CITIZEN SHARE_WHITE_POVERTY GINI_INDEX SHARE_NON_WHITE
## 1               0.02                0.12      0.472            0.35
## 2               0.04                0.06      0.422            0.42
## 3               0.10                0.09      0.455            0.49
## 4               0.04                0.12      0.458            0.26
## 5               0.13                0.09      0.471            0.61
## 6               0.06                0.07      0.457            0.31
## 7               0.06                0.06      0.486            0.30
## 8               0.05                0.08      0.440            0.37
## 9               0.11                0.04      0.532            0.63
## 10              0.09                0.11      0.474            0.46
## 11              0.08                0.09      0.468            0.48
## 12              0.08                0.07      0.433            0.81
## 13              0.04                0.11      0.433            0.16
## 14              0.07                0.07      0.465            0.37
## 15              0.03                0.12      0.440            0.20
## 16              0.03                0.09      0.427            0.15
## 17              0.04                0.11      0.445            0.25
## 18              0.03                0.17      0.466            0.15
## 19              0.02                0.12      0.475            0.42
## 20                NA                0.12      0.437            0.09
## 21              0.08                0.06      0.443            0.50
## 22              0.09                0.08      0.475            0.27
## 23              0.04                0.09      0.451            0.24
## 24              0.05                0.05      0.440            0.18
## 25                NA                0.14      0.468            0.44
## 26              0.02                0.07      0.455            0.20
## 27              0.01                0.10      0.435            0.10
## 28              0.05                0.07      0.432            0.21
## 29              0.10                0.08      0.448            0.50
## 30              0.03                0.06      0.425            0.09
## 31              0.11                0.07      0.464            0.44
## 32              0.06                0.10      0.464            0.62
## 33              0.10                0.10      0.499            0.42
## 34              0.05                0.10      0.464            0.38
## 35              0.03                0.09      0.433            0.15
## 36              0.03                0.10      0.452            0.21
## 37              0.04                0.10      0.454            0.35
## 38              0.07                0.10      0.449            0.26
## 39              0.03                0.09      0.461            0.24
## 40              0.08                0.08      0.467            0.28
## 41              0.03                0.09      0.461            0.36
## 42                NA                0.08      0.442            0.17
## 43              0.04                0.13      0.468            0.27
## 44              0.11                0.08      0.469            0.56
## 45              0.04                0.08      0.419            0.19
## 46              0.01                0.10      0.444            0.06
## 47              0.06                0.07      0.459            0.38
```

```
## 48              0.08               0.09      0.441            0.31
## 49              0.01               0.14      0.451            0.07
## 50              0.03               0.09      0.430            0.22
## 51              0.02               0.09      0.423            0.15
##    SHARE_VOTERS_VOTED_TRUMP Hate_Crime Avg_HateCrimes
## 1                      0.63 0.12583893      1.8064105
## 2                      0.53 0.14374012      1.6567001
## 3                      0.50 0.22531995      3.4139280
## 4                      0.60 0.06906077      0.8692089
## 5                      0.33 0.25580536      2.3979859
## 6                      0.44 0.39052330      2.8046888
## 7                      0.41 0.33539227      3.7727015
## 8                      0.42 0.32275417      1.4699796
## 9                      0.04 1.52230172     10.9534797
## 10                     0.49 0.18752122      0.6980703
## 11                     0.51 0.12042027      0.4120118
## 12                     0.30         NA             NA
## 13                     0.59 0.12420817      1.8913305
## 14                     0.39 0.19534455      1.0440158
## 15                     0.57 0.24700888      1.7573566
## 16                     0.52 0.45442742      0.5613956
## 17                     0.57 0.10515247      2.1439867
## 18                     0.63 0.32439697      4.2078896
## 19                     0.58 0.10973335      1.3411696
## 20                     0.45 0.61557402      2.6266109
## 21                     0.35 0.37043897      1.3248395
## 22                     0.34 0.63081059      4.8018993
## 23                     0.48 0.40377937      3.2004423
## 24                     0.45 0.62747993      3.6124118
## 25                     0.58 0.06744680      0.6227460
## 26                     0.57 0.18452351      1.9089550
## 27                     0.57 0.49549103      2.9549594
## 28                     0.60 0.15948963      2.6862484
## 29                     0.46 0.14167316      2.1139902
## 30                     0.47 0.15154960      2.1059886
## 31                     0.42 0.07830591      4.4132026
## 32                     0.40 0.29481132      1.8864352
## 33                     0.37 0.35062045      3.1021643
## 34                     0.51 0.24400659      1.2626798
## 35                     0.64         NA      4.7410699
## 36                     0.52 0.19071396      3.2404204
## 37                     0.65 0.13362910      1.0816721
## 38                     0.41 0.83284961      3.3948861
## 39                     0.49 0.28510109      0.4309276
## 40                     0.40 0.09540164      1.2825718
## 41                     0.55 0.20989442      1.9370828
## 42                     0.62         NA      3.3017371
## 43                     0.61 0.19993848      3.1360512
## 44                     0.53 0.21358394      0.7527683
## 45                     0.47 0.13654673      2.3840650
## 46                     0.33 0.32414911      1.9030814
## 47                     0.45 0.36324890      1.7247546
## 48                     0.38 0.67748765      3.8177403
## 49                     0.69 0.32867707      2.0370536
```

```
## 50                      0.48 0.22619711      1.1219447
## 51                      0.70         NA      0.2669408
```

Section 4: Characteristics of the data This data frame has rows 51 and columns 12. The names of the columns and a brief description of each are in the table below:

```
Fields <- data.frame(
                                `Column Name` = c('state',   'median_household_income',  'share_unemploye
                                                  'share_population_with_high_school_degree',    'share_n
                                                  'share_non_white','share_voters_voted_trump','hate_crim

                                 Description =  c('State name',
                                                  'Median household income, 2016',
                                                  'Share of the population that is unemployed (seasonally
                                                  'Share of the population that lives in metropolitan are
                                                  'Share of adults 25 and older with a high-school degree
                                                  'Share of the population that are not U.S. citizens, 20
                                                  'Share of white residents who are living in poverty, 20
                                                  'Gini Index, 2015','Share of the population that is not
                                                  'Share of 2016 U.S. presidential voters who voted for D
                                                  'Hate crimes per 100,000 population, Southern Poverty L
                                                  'Average annual hate crimes per 100,000 population, FBI
)
)

library(knitr)
kable(Fields)
```

Column.Name	Description
state	State name
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 adults 25 and older with a high-school degree, 2009
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: Summary Statistics

```
# picking 3 columns from the data set
data_filtered <- data[, c('STATE', 'MEDIAN_HOUSEHOLD_INCOME', 'SHARE_UNEMPLOYED_SEASONAL')]
```

```r
missing <- as.data.frame(lapply(data_filtered, function(x) sum(is.na(x))))

missing <- rename(missing, state_missing_val = STATE, median_household_income_missing_val = MEDIAN_HOUSE

missing
```

missing values

```
##   state_missing_val median_household_income_missing_val
## 1                 0                                   0
##   share_unemployed_seasonal_missing_val
## 1                                     0
```

```r
# how to create an object
summary_statistics <- summary(data_filtered)
summary_statistics
```

```
##     STATE           MEDIAN_HOUSEHOLD_INCOME SHARE_UNEMPLOYED_SEASONAL
##  Length:51          Min.   :35521           Min.   :0.02800
##  Class :character   1st Qu.:48657           1st Qu.:0.04200
##  Mode  :character   Median :54916           Median :0.05100
##                     Mean   :55224           Mean   :0.04957
##                     3rd Qu.:60719           3rd Qu.:0.05750
##                     Max.   :76165           Max.   :0.07300
```