Hate Crimes as recorded and analysed by the FBI, USA

RBGA

December 22nd, 2022

RESEARCH TOPIC This project proposal aims to analyze and annotate the research done by the FBI on their analysis of Hate Crime statistics across the United States of America for the year 2020, which was published on the FBI public database on 30th of Aug 2021. The motive of this research is to capture information about the types of bias that motivate crimes, the nature of the offenses, and some information about the victims and offenders.

On April 23, 1990, Congress passed the Hate Crime Statistics Act, 28 U.S.C. § 534, which required the Attorney General to collect data "about crimes that manifest evidence of prejudice based on race, religion, sexual orientation, or ethnicity." The Attorney General delegated the responsibilities of developing the procedures for implementing, collecting, and managing hate crime data to the Director of the FBI, who, in turn, assigned the tasks to the Uniform Crime Reporting (UCR) Program. Under the direction of the Attorney General and with the cooperation and assistance of many local and state law enforcement agencies, the UCR Program created a hate crime data collection to comply with the congressional mandate.

This R Studio Markdown aims to analyse data from 2020 of FBI's Hate Crime Statistics with various key parameters as a control variable and defend a hypothesis that

1. Most hate crimes with bias 'Race/Ethnicity/Ancestry' happened mostly at the target victim's place of residence the most, followed by Supermarkets to be the next. Minority people are not safe in their own houses and public essential places like supermarkets in the US of A.
2. Most hate crimes are targeted toward Black victims the highest in US of A while the second highest is not even a race, rather a broader classification called Sexual Orientation where the Victims have various Sexual Orientations. This infers that the criminals in America hates the one race - Black more than they hate an entire broader group of populous with various Sexual Orientations.
3. Among juvenile, minors and kids, the hate crime is seriously directed toward Black Juveniles more than any other race present. With the second highest again not being another race but the entire victims of various sexual orientations, White race in USA commits the most range of hate crimes against any other race. The highest crimes are against other people and specifically Intimidation
4. Single Bias Incidents, Race Bias, African Bias are the leading bars showing which type of hate crimes are the highest
5. More Juvenile Hate crime victims lead to more adult hate crime victims
6. White race commits the most number of direct targeted hate crimes like Murder, Rape, Intimidation, followed by Black race being the second
7. Mean value of Known offenders is 600+ and intimidation is one of the biggest repeat offense
8. Known Offenders cause the most number of hate crimes.

In creating the program, the designers recognized hate crimes are not separate, distinct crimes; instead, they are traditional offenses motivated by the offender's bias. (For example, an offender assaults a victim because of a bias against the victim's race.) After much consideration, the developers agreed hate crime data could be derived by capturing the additional element of bias in those offenses already being reported to the UCR Program. Attaching the collection of hate crime statistics to the established UCR data collection procedures, they concluded, would fulfill the directives of the Hate Crime Statistics Act without placing an undue additional reporting burden on law enforcement.

Hate Crime is an active problem in current fast paced Workforce society where the motives of the crime in retrospect with today's ideals are plainly ridiculous. In the age where humanity has surpassed the Nuclear era and is well into the information era, hampering common public workforce from all around the world with baseless hate, racism and discrimination is something massively bizarre and counterproductive that our species are still facing. Coming from the perspective of an author, who is a person of color, this research stresses the importance of analysis of hate crime based on various parameters, to ensure systemic eradication of this barbaric baseless hate. Though the research attempted by FBI is satisfactory, it doesn't exactly produce the right data points with which a potentially better organization or institution can better use to curb this problem. One strong suggestion would be to review those hate crimes on a much deeper scale, including potential factors such as the criminal's mental history, childhood, past traumas, education level and more.

Step -1 Install Necessary Packages and Libraries

```
install.packages("tidyverse")
install.packages("readxl")
install.packages("car")
install.packages("haven")
install.packages("corrr")
install.packages("janitor")
install.packages("rstatix")
install.packages("lmtest")
install.packages ("ppcor")
install.packages("tableone")
install.packages("ggthemes")
install.packages("stargazer")
install.packages("ggrepel")
install.packages("psych")
```

More

```
library(package = "tableone")
library(car) # Load the library car so we can run leveneTest, install if needed
library(tidyverse)
library(readxl) #to import excel files
library(haven) #import files produced in other statistical s oftware
library(corrr) #Includes functions to evaluate correlations within the tidyverse
library(janitor) #To analyze cross-tabulation - chi-squared
library(rstatix) #One of many packages used for t-tests
library(lmtest) #to test for constant variance assumption
library(ppcor) #to obtain partial correlations
library(stargazer) #Produces summary and regression output formatted tables
library(ggrepel) #to add overlapping labels to our plots
library(ggthemes) #additional ggplot themes
library(psych)
```

As Our Data is a spreadsheet XLSX, lets install XLSX

```
install.packages("xlsx")
library("xlsx")
```

Now use read.xlsx to read sheets into R Objects. Use head to view created objects

```
victims <- read.xlsx("Table_7_Victims_Offense_Type_by_Bias_Motivation_2020.xlsx", 1)
```

```
head(victims)
```

```
##                         Bias.motivation Total.victims1
## 1                                 Total          11472
## 2                 Single-Bias Incidents          11126
## 3              Race/Ethnicity/Ancestry:           6880
## 4                             Anti-White           1082
## 5        Anti-Black or African American           3915
## 6 Anti-American Indian or Alaska Native            108
##   Total.number.of.adult.victims2 Total.number.of.juvenile.victims2
## 1                           7729                               791
## 2                           7469                               776
## 3                           5205                               589
## 4                            845                                64
## 5                           2932                               408
## 6                             85                                 6
##   Murder.and.nonnegligent.manslaughter Rape3 Aggravated.assault Simple.assault
## 1                                   22    21               1390           2166
## 2                                   22    21               1370           2131
## 3                                    8    10                976           1455
## 4                                    1     7                145            317
## 5                                    5     2                530            660
## 6                                    0     1                  5             26
##   Intimidation Other4 Robbery Burglary Larceny..theft Motor.vehicle.theft Arson
## 1         4119     32     155      152            383                  36   105
## 2         3940     32     154      150            376                  36   100
## 3         2296     20      90       79            181                  18    41
## 4          231      8      31       18             63                   6     4
## 5         1577      5      23       38             77                   6    27
## 6           14      3       1        5             11                   2     1
##   Destruction..damage..vandalism Other4.1 Crimes.against.society4 NA.
## 1                           2574       85                    232  NA
## 2                           2479       84                    231  NA
## 3                           1515       43                    148  NA
## 4                            168       18                     65  NA
## 5                            902        9                     54  NA
## 6                             25        5                      9  NA
```

More Reading and 'head'ing

```
bias <- read.xlsx("Table_1_Incidents_Offenses_Victims_and_Known_Offenders_by_Bias_Motivation_2020.xlsx"
head(bias)
```

```
##                         Bias.motivation Incidents Offenses Victims1
## 1                                 Total      8263    11129    11472
## 2                 Single-Bias Incidents      8052    10790    11126
## 3              Race/Ethnicity/Ancestry       5227     6677     6880
## 4                             Anti-White       869     1048     1082
## 5        Anti-Black or African American      2871     3819     3915
## 6 Anti-American Indian or Alaska Native        96      103      108
##   Known.Offenders2
```

```
## 1             6780
## 2             6657
## 3             4339
## 4              825
## 5             2302
## 6               74
```

More

```
location <- read.xlsx("Table_10_Incidents_Bias_Motivation_by_Location_2020.xlsx", 1)
head(location)
```

```
##                            Location Total.incidents Race..ethnicity..ancestry
## 1                             Total            8263                      5227
## 2      Abandoned/condemned structure              10                         6
## 3             Air/bus/train terminal             118                        77
## 4                      Amusement park               4                         3
## 5 Arena/stadium/fairgrounds/coliseum               9                         6
## 6           Auto dealership new/used              11                         6
##   Religion Sexual.orientation Disability Gender Gender.identity
## 1     1244               1110        130     75              266
## 2        0                  0          0      0                2
## 3        9                 20          2      0                6
## 4        0                  1          0      0                0
## 5        3                  0          0      0                0
## 6        2                  0          0      0                3
##   Multiple..bias.incidents1
## 1                       211
## 2                         2
## 3                         4
## 4                         0
## 5                         0
## 6                         0
```

More

```
race <- read.xlsx("Table_3_Offenses_Known_Offenders_Race_and_Ethnicity_by_Offense_Type_2020.xlsx", 1)
head(race)
```

```
##                          Offense.type   NA. White Black.or.African.American
## 1                               Total 11129  4958                      1390
## 2             Crimes against persons:  7750  4194                      1148
## 3 Murder and nonnegligent manslaughter    22     8                         3
## 4                               Rape2    21     9                         5
## 5                  Aggravated assault  1390   746                       311
## 6                      Simple assault  2166  1136                       488
##   American.Indian.or.Alaska.Native Asian
## 1                               73    75
## 2                               64    60
## 3                                0     0
## 4                                0     0
## 5                               12    15
## 6                               37    27
##   Native.Hawaiian.or.Other.Pacific.Islander Multiple Unknown.Race
```

```
## 1                                    29      223          947
## 2                                    24      187          583
## 3                                     0        1            5
## 4                                     1        1            3
## 5                                     3       48          119
## 6                                     6       92          155
##   Hispanic.or.Latino Not.Hispanic.or.Latino Group.of.multiple.ethnicities
## 1                608                  2625                            69
## 2                509                  2150                            52
## 3                  3                     6                             0
## 4                  5                     6                             0
## 5                162                   530                            19
## 6                193                   742                            26
##   Unknown.ethnicity NA..1 NA..2
## 1              2915  3434    NA
## 2              2342  1490    NA
## 3                 7     5    NA
## 4                 8     2    NA
## 5               472   136    NA
## 6               791   225    NA
```

More

```
vicincidents <- read.xlsx("Table_8_Incidents_Victim_Type_by_Bias_Motivation_2020.xlsx", 1)
head(vicincidents)
```

```
##           Bias.motivation Total.incidents Individual
## 1                   Total            8263       6546
## 2  Single-Bias Incidents:            8052       6406
## 3 Race/Ethnicity/Ancestry            5227       4345
## 4                Religion            1244        624
## 5      Sexual Orientation            1110       1025
## 6             Disability             130        106
##   Business..financial.institution Government Religious.organization
## 1                             551        343                    200
## 2                             535        321                    195
## 3                             328        212                     20
## 4                             162         91                    169
## 5                              25         14                      6
## 6                              11          1                      0
##   Society..public1 Other..unknown..multiple
## 1              185                     438
## 2              185                     410
## 3              107                     215
## 4               41                     157
## 5                9                      31
## 6               10                       2
```

A Whole Lot More Reading and Heading

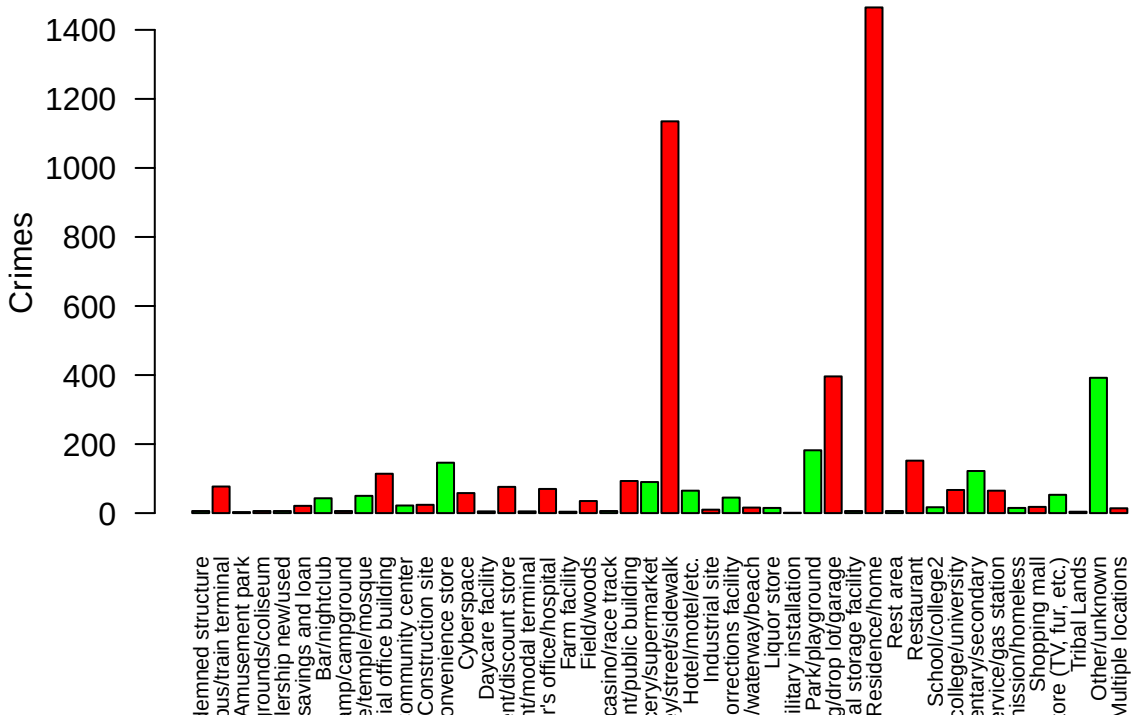
```
table1 <- read.xlsx("Table_1_Incidents_Offenses_Victims_and_Known_Offenders_by_Bias_Motivation_2020.xlsx

table2<-read.xlsx("Table_2_Incidents_Offenses_Victims_and_Known_Offenders_by_Offense_Type_2020.xlsx", 1
```

```
table3<-read.xlsx("Table_3_Offenses_Known_Offenders_Race_and_Ethnicity_by_Offense_Type_2020.xlsx", 1)

table4<-read.xlsx("Table_4_Offenses_Offense_Type_by_Bias_Motivation_2020.xlsx", 1)

table7<-read.xlsx("Table_7_Victims_Offense_Type_by_Bias_Motivation_2020.xlsx", 1)

table8<-read.xlsx("Table_8_Incidents_Victim_Type_by_Bias_Motivation_2020.xlsx", 1)

table9<-read.xlsx("Table_9_Known_Offenders_Known_Offenders_Race_Ethnicity_and_Age_2020.xlsx", 1)

tablex<-read.xlsx("Table_10_Incidents_Bias_Motivation_by_Location_2020.xlsx", 1)

table5<-read.xlsx("Table_5_Offenses_Known_Offenders_Race_and_Ethnicity_by_Bias_Motivation_2020.xlsx", 1
```

Now we use Barplot function to create a barplot with 'Race/Ethnicity/Ancestry' as control factor for various locations. This is a formatted bar graph

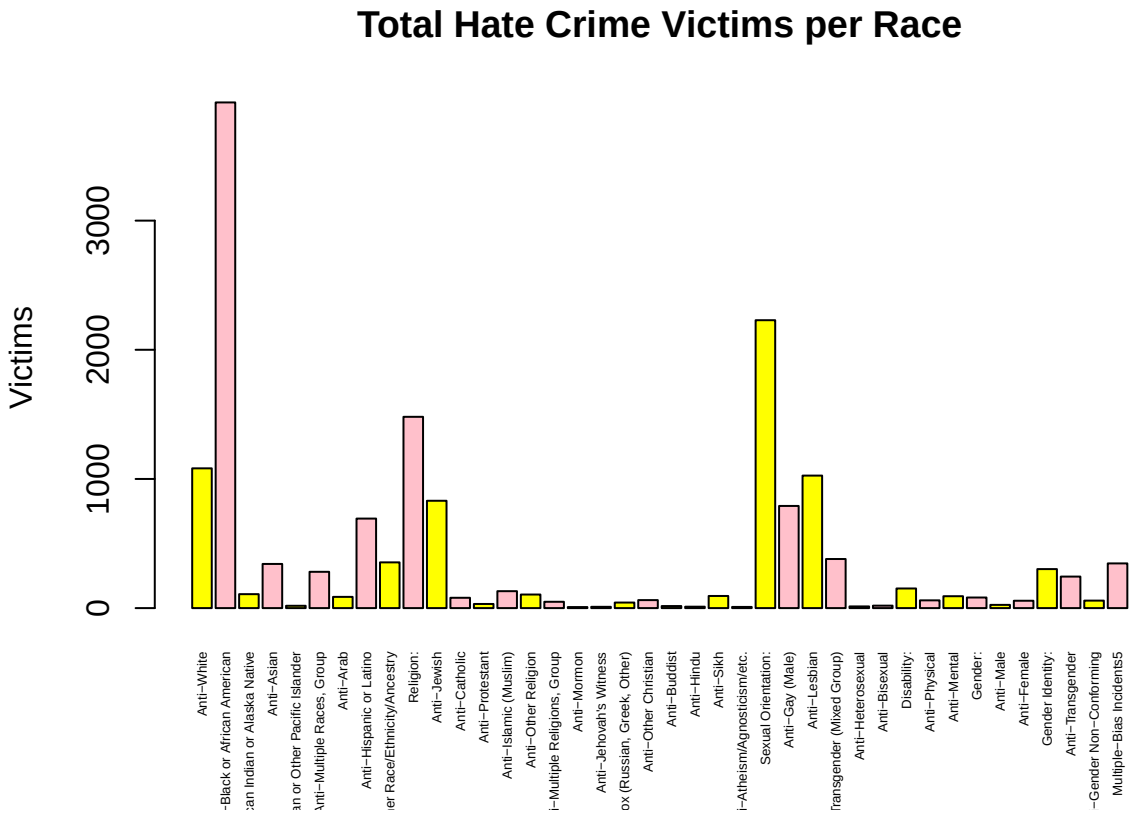
From the graph we clearly see that most hate crimes with bias 'Race/Ethnicity/Ancestry' happened mostly at the target victim's place of residence the most, followed by Supermarkets to be the next. Minority people are not safe in their own houses and public essential places like supermarkets in the US of A. This is a demonstration of R's base barplot() function. This is from the location vs various biases data from the FBI in table 'location'.

```
barplot(location$Race..ethnicity..ancestry[2:47],
        names=location$Location[2:47],
        horiz = FALSE,
        col=c("green","red"),
        cex.names=0.7,
        las=2,
        main="Race Bias Crimes vs Locations",
        ylab = "Crimes")
```



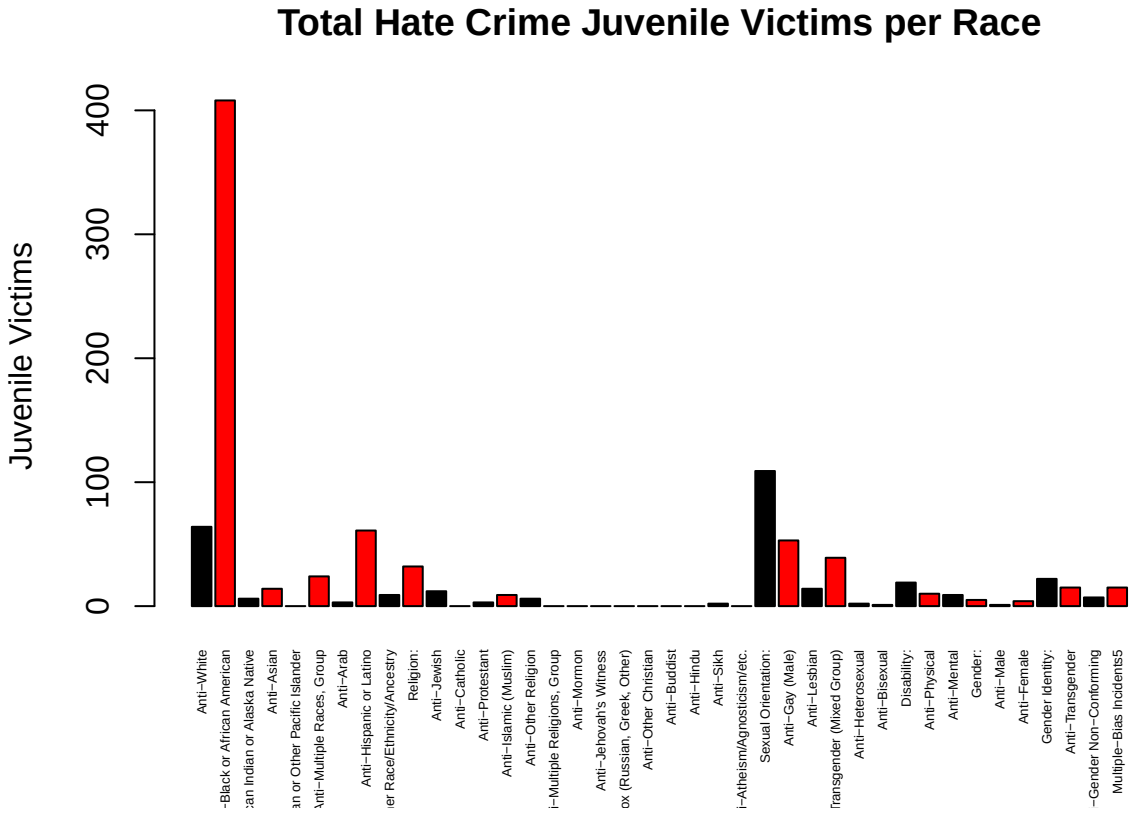
Another barplot from object victims which has victims of different races based on different biases. From this plot it is evident that most hate crimes are targeted toward Black victims the highest in US of A while the second highest is not even a race, rather a broader classification called Sexual Orientation where the Victims have various Sexual Orientations. This infers that the criminals in Amercia hates the one race - Black more than they hate an entire broader group of populous with various Sexual Orientations.

```
barplot(victims$Total.victims1[4:43],
        names = victims$Bias.motivation[4:43], horiz=FALSE,
        cex.names=0.4,
        col=c("yellow","pink"),
        las=3,
        main="Total Hate Crime Victims per Race",
        ylab="Victims")
```



From the object 'Victims', a barplot with 'Total.number.of.juvenile.victims2' as the main variable is plotted against bias motivation. From the graph, the data shows that even amongst juvenile, minors and kids, the hate crime is seriously directed toward Black Juveniles more than any other race present. With the second highest again not being another race but the entire victims of various sexual orientations.

```
barplot(victims$Total.number.of.juvenile.victims2[4:43],
        names = victims$Bias.motivation[4:43],
        horiz=FALSE,
        cex.names=0.4,
        col=c("black","red"),
        las=3,
        main="Total Hate Crime Juvenile Victims per Race",
        ylab="Juvenile Victims")
```



Now let's extract data from object 'race' into Data Frame 'racetable' which will have the columns as shown below. Thereby removing unwanted random rows of gibberish data.

```
racetable <- data.frame(race$White[2:17],
                        race$Asian[2:17],
                        race$Black.or.African.American[2:17],
                        race$American.Indian.or.Alaska.Native[2:17])

colnames(racetable) <- c("White", "Asian", "Black", "Native Indian")
head(racetable)
```

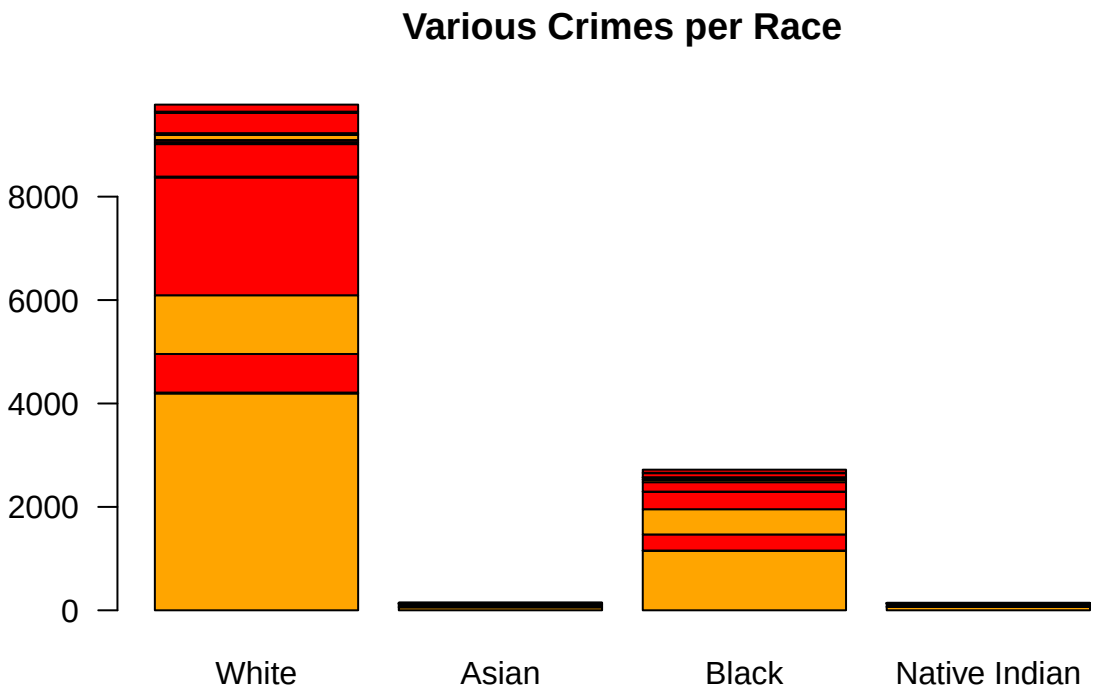
```
##   White Asian Black Native Indian
## 1  4194    60  1148            64
## 2     8     0     3             0
## 3     9     0     5             0
## 4   746    15   311            12
## 5  1136    27   488            37
## 6  2277    17   335            14
```

We now plot a Stacked bar plot for various races for various hate crimes they have commited.

This stacked bar plot shows that the white race in USA commits the most range of hate crimes against any other race. The highest crimes are against other people and specifically Intimidation. Oppression begins as a form of intimidation or bullying.

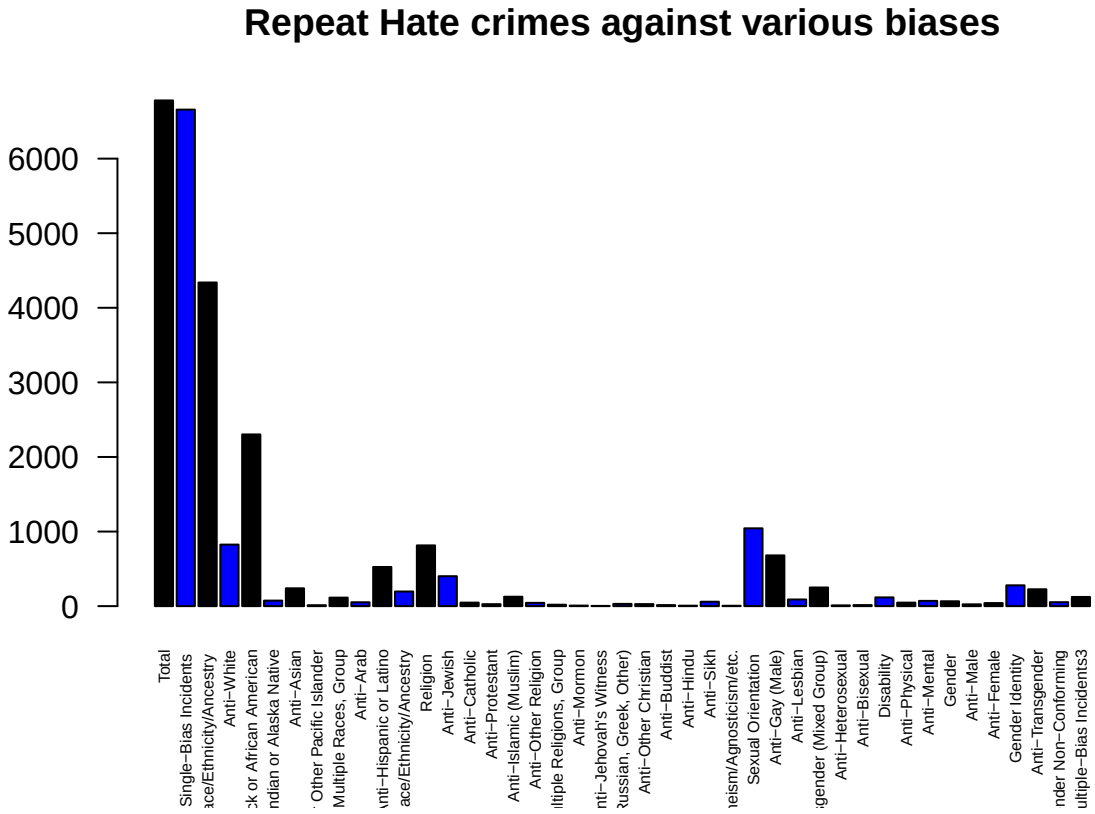
```
barplot(as.matrix(racetable), main = "Various Crimes per Race",
        las = 1,
        cex.names=1,
        col = c("orange","red"),
```

```
        beside = FALSE
        )
```



We not plot a barplot for Known HAte Crime Offenders or Repeat Offenders for various hate crime offenses and the result shows that Single Bias Incidents, Race Bias, African Bias are the leading bars showing which type of hate crimes are the highest.

```
barplot(bias$Known.Offenders2,
        names = bias$Bias.motivation,
        col=c("black","blue"),
        cex.names=0.5,
        las=2,
        main="Repeat Hate crimes against various biases", )
```

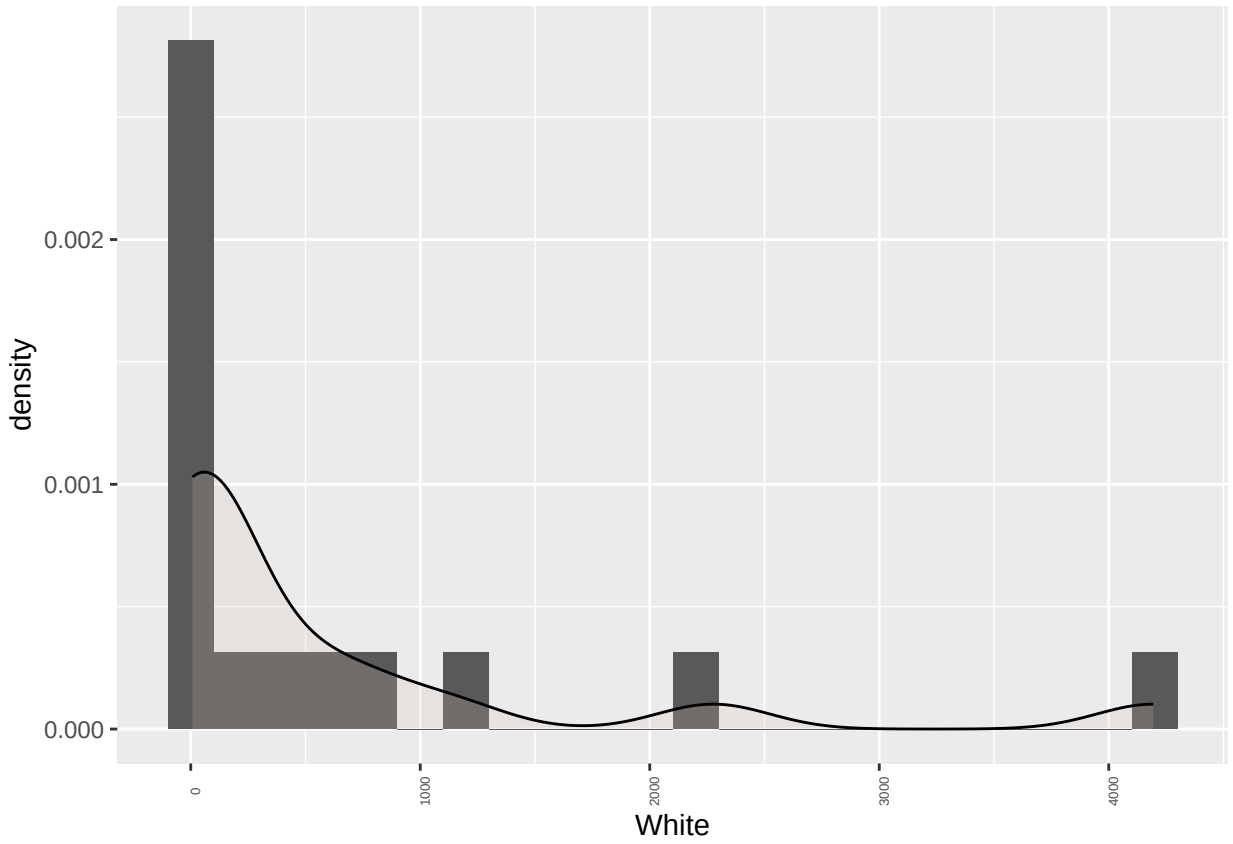


Lets load some more libraries

```
library(xlsx)            # for reading in Excel data
library(dplyr)           # for data manipulation
library(tidyr)           # for data manipulation
library(magrittr)        # for easier syntax in one or two areas
library(gridExtra)       # for generating the bin width comparison plot
library(ggplot2)         # for generating the visualization
```

We now use ggplot() to plot a histogram and a geom_density() for the race data with White being the key variable.

```
ggplot(race[2:17, ], aes(x = White)) +
        theme(axis.text.x = element_text(angle = 90, size = 5))+
        geom_histogram(aes(y = after_stat(density)),binwidth = 200) +
        geom_density(alpha = .2, fill = "antiquewhite3")
```



We try to find the Linear Regression Model for the victims data between Adult Victims and Juvenile Victims.

The summary functions returns the calculated linear regression for the above proposed model. The output returns back with the LM test results. We see -

Residuals Median is -17.87, meaning the predictions are not good for lower numbers

Coefficients Y = MX + B Y = 9.26X + 31.8

This shows that with increase in Juvenile Victims, Adult Victims are considerably increasing too.

The Pr value is <0.05 meaning the model is significant.

```
juvymod <- lm(Total.number.of.adult.victims2 ~ Total.number.of.juvenile.victims2, data = victims)
summary(juvymod)
```

```
##
## Call:
## lm(formula = Total.number.of.adult.victims2 ~ Total.number.of.juvenile.victims2,
##     data = victims)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -878.56  -36.05  -17.87   29.41  411.76
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       31.8709    30.5430   1.043    0.303
## Total.number.of.juvenile.victims2   9.2615     0.1506  61.508   <2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 186.9 on 41 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.9893, Adjusted R-squared:  0.989
## F-statistic:  3783 on 1 and 41 DF,  p-value: < 2.2e-16
```

Lets load more packages

```
library(stargazer)
library(ggthemes)
```

A Stargazer result of our Linear Model

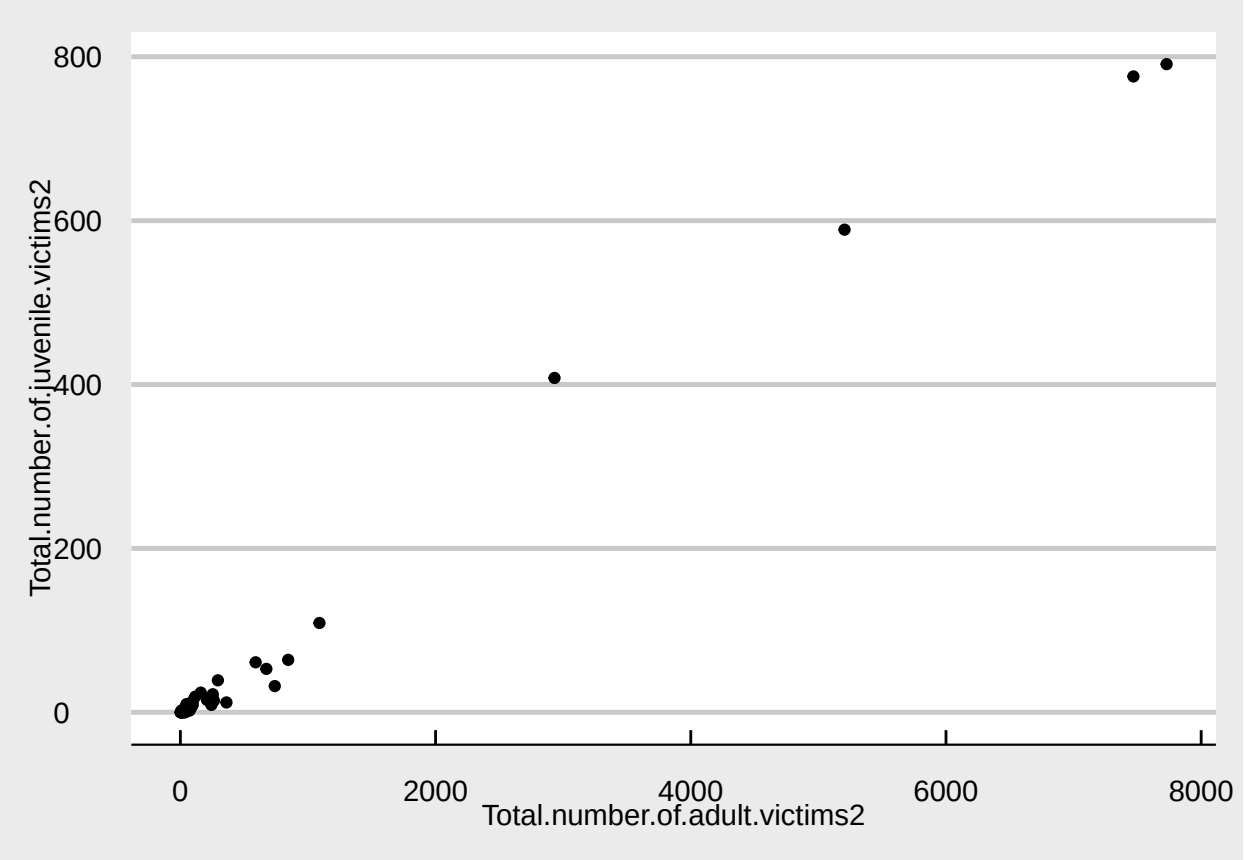
F statistic is high indicating a relation between the two observed data points

```
stargazer(juvymod, type = "text", title = "Higher Juvenile Criminals Produce Higher Adult Criminals")
```

```
##
## Higher Juvenile Criminals Produce Higher Adult Criminals
## ================================================================
##                                       Dependent variable:
##                                   -----------------------------
##                                   Total.number.of.adult.victims2
## ----------------------------------------------------------------
## Total.number.of.juvenile.victims2            9.262***
##                                              (0.151)
##
## Constant                                      31.871
##                                              (30.543)
##
## ----------------------------------------------------------------
## Observations                                    43
## R2                                             0.989
## Adjusted R2                                    0.989
## Residual Std. Error                      186.908 (df = 41)
## F Statistic                          3,783.182*** (df = 1; 41)
## ================================================================
## Note:                                  *p<0.1; **p<0.05; ***p<0.01
```

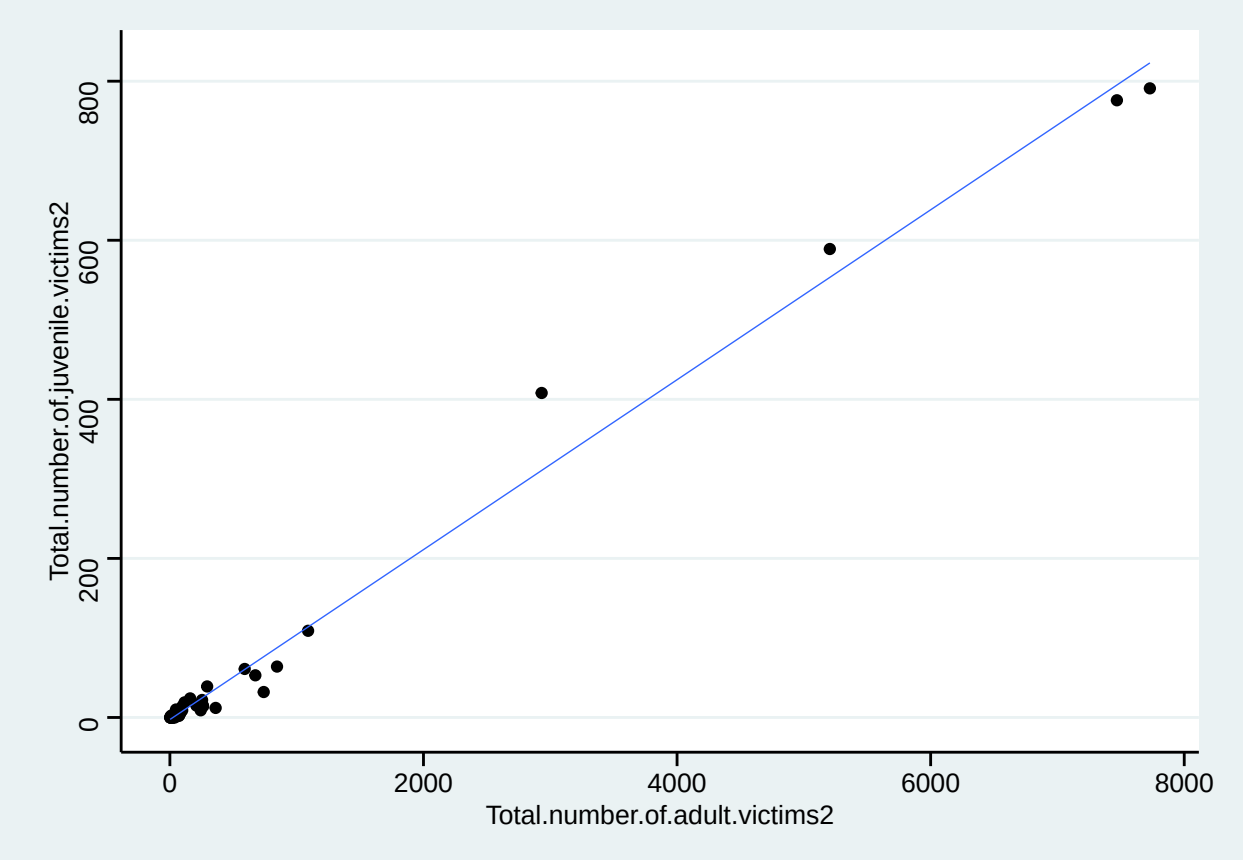
We will use GGPLOT() to plot a Scatter Plot of Adult and Juvenile Victims

```
ggplot(data=victims, aes(Total.number.of.adult.victims2, Total.number.of.juvenile.victims2)) +
  geom_point() +
  theme_economist_white()
```



We also add a regression line which is an indication and prediction of a direct correlation between the two data sets indicationg theat more Juvenile victims leads to more adult victims

```
ggplot(data=victims, aes(Total.number.of.adult.victims2, Total.number.of.juvenile.victims2)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, size = 0.25)+
  theme_stata()
```



Lets now create a Linear Model for various crimes race wise for White and Black

```
wbmod <- lm(table3$White[c(2:7)] ~ table3$Black.or.African.American[c(2:7)])
summary(wbmod)
```

```
##
## Call:
## lm(formula = table3$White[c(2:7)] ~ table3$Black.or.African.American[c(2:7)])
##
## Residuals:
##       1       2       3       4       5       6
##   73.09  -40.05  -46.17 -397.63 -637.24 1048.00
##
## Coefficients:
##                                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)                               37.3805   371.4405   0.101  0.92468
## table3$Black.or.African.American[c(2:7)]   3.5571     0.6848   5.194  0.00654
##
## (Intercept)
## table3$Black.or.African.American[c(2:7)] **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 646.4 on 4 degrees of freedom
## Multiple R-squared:  0.8709, Adjusted R-squared:  0.8386
## F-statistic: 26.98 on 1 and 4 DF,  p-value: 0.006543
```

Another Model with Asians as an added control variable

```
wbamod <- lm(table3$White[c(2:7)] ~ table3$Black.or.African.American[c(2:7)] + table3$Asian[c(2:7)])
summary(wbamod)
```

```
##
## Call:
## lm(formula = table3$White[c(2:7)] ~ table3$Black.or.African.American[c(2:7)] +
##     table3$Asian[c(2:7)])
##
## Residuals:
##        1        2        3        4        5        6
##    9.421   -5.057  -37.544 -674.107 -247.753  955.041
##
## Coefficients:
##                                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)                              -37.17     410.18  -0.091    0.934
## table3$Black.or.African.American[c(2:7)]    16.74      18.51   0.904    0.432
## table3$Asian[c(2:7)]                     -250.00     350.69  -0.713    0.527
##
## Residual standard error: 690.3 on 3 degrees of freedom
## Multiple R-squared:  0.8896, Adjusted R-squared:  0.816
## F-statistic: 12.08 on 2 and 3 DF,  p-value: 0.03669
```

Another model with Native Indians as added control variable for the Linear Model

```
wbanmod <- lm(table3$White[c(2:7)] ~ table3$Black.or.African.American[c(2:7)] + table3$Asian[c(2:7)] + t
summary(wbanmod)
```

```
##
## Call:
## lm(formula = table3$White[c(2:7)] ~ table3$Black.or.African.American[c(2:7)] +
##     table3$Asian[c(2:7)] + table3$American.Indian.or.Alaska.Native[c(2:7)])
##
## Residuals:
##      1      2      3      4      5      6
##  35.76 -54.40 129.94 -56.74 -36.69 -17.87
##
## Coefficients:
##                                                 Estimate Std. Error t value
## (Intercept)                                       337.39      76.75   4.396
## table3$Black.or.African.American[c(2:7)]          -91.67      10.87  -8.430
## table3$Asian[c(2:7)]                             2273.58     249.72   9.105
## table3$American.Indian.or.Alaska.Native[c(2:7)]  -427.53      41.15 -10.389
##                                                 Pr(>|t|)
## (Intercept)                                      0.04805 *
## table3$Black.or.African.American[c(2:7)]         0.01378 *
## table3$Asian[c(2:7)]                             0.01185 *
## table3$American.Indian.or.Alaska.Native[c(2:7)]  0.00914 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 114 on 2 degrees of freedom
## Multiple R-squared:  0.998,  Adjusted R-squared:  0.995
## F-statistic: 331.2 on 3 and 2 DF,  p-value: 0.003012
```

Now lets do a Stargazer model for all these linear models above

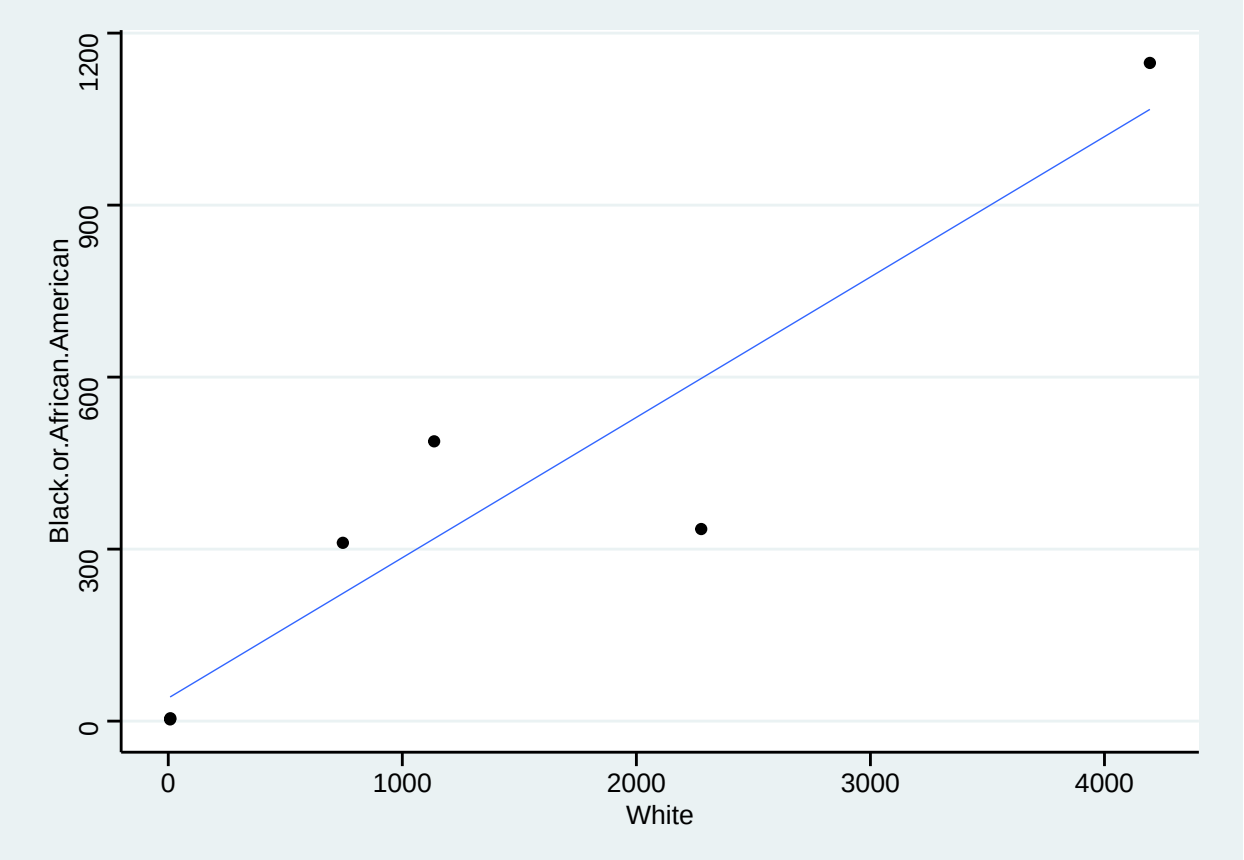
The f statistic is pretty far from 1 which indicates the alternative hypothesis is true, meaning there is a higher effect of hate crimes by white more than other races.

```
stargazer(wbanmod, type = "text", title = "Direct Crimes against victim of hate, race wise")
```

```
##
## Direct Crimes against victim of hate, race wise
## ================================================================
##                                            Dependent variable:
##                                        ---------------------------
##                                               White[c(2:7)]
## ----------------------------------------------------------------
## Black.or.African.American[c(2:7)]                -91.666**
##                                                  (10.874)
##
## Asian[c(2:7)]                                   2,273.581**
##                                                  (249.719)
##
## American.Indian.or.Alaska.Native[c(2:7)]        -427.528***
##                                                  (41.152)
##
## Constant                                         337.393**
##                                                  (76.754)
##
## ----------------------------------------------------------------
## Observations                                          6
## R2                                                  0.998
## Adjusted R2                                         0.995
## Residual Std. Error                          114.029 (df = 2)
## F Statistic                                331.198*** (df = 3; 2)
## ================================================================
## Note:                                *p<0.1; **p<0.05; ***p<0.01
```

Create a table from obejct with row and column range values and do a GGPLOT() for various races. This is the data which contains direct hate crimes against individuals like Murder, Rape and assault race wise.

```
t3d <- table3[2:7, 3:6]
ggplot(data=t3d, aes(White, Black.or.African.American, American.Indian.or.Alaska.Native, Asian)) +

  geom_point() +

  geom_smooth(method = "lm", se = FALSE, size = 0.25)+

  theme_stata()
```



Then do the ANOVA test

From the above graph, it is clear that Black race has the second highest hate crimes committed after white race. The numbers are exponentially staggering with White being the highest, then nearly halved is black and so on. The ANOVA test comes back with results of the test which also indicates, that African American come up the second highest at perpetrating direct targetted hate crimes after White, which has the highest, as inferred from the F values

```
anova <- aov(t3d$White ~ t3d$Black.or.African.American + t3d$American.Indian.or.Alaska.Native + t3d$Asia
summary(anova)
```

```
##                                    Df   Sum Sq  Mean Sq F value  Pr(>F)
## t3d$Black.or.African.American       1 11273814 11273814  867.04 0.00115 **
## t3d$American.Indian.or.Alaska.Native  1   567722   567722   43.66 0.02215 *
## t3d$Asian                           1  1077830  1077830   82.89 0.01185 *
## Residuals                           2    26005    13003
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Lets Install more packages

```
install.packages("rstatix")
library(rstatix)
```

From TAble1 data of offenses, find the mean of Known Offenders

```
table1 %>%
  summarize(mean_koff = mean(Known.Offenders2, na.rm = T))
```

```
##   mean_koff
## 1  624.9767
```

Lets run a T - Test against the mean value 624 for all mu 500 and above. From the Statistic and p-value which is greater than alpha 0.05, meaning the null hypothesis is valid and the mean is true.

```
table1 %>%
t_test(Known.Offenders2 ~ 1, mu = 500, detailed = TRUE)
```

```
## # A tibble: 1 x 12
##   estimate .y.    group1 group2     n stati~1     p    df conf.~2 conf.~3 method
## *    <dbl> <chr>  <chr>  <chr>  <int>   <dbl> <dbl> <dbl>   <dbl>   <dbl> <chr>
## 1     625. Known~ 1      null ~    43   0.528 0.601    42    147.   1103. T-test
## # ... with 1 more variable: alternative <chr>, and abbreviated variable names
## #   1: statistic, 2: conf.low, 3: conf.high
```

CONCLUSION OF THE TEST The T Test for Known offenders and mean The P value is 0.61 > alpha - 0.05 meaning that the two variables are dependent on each other meaning that known offenders have the highest number of incidents inferring that most hate crimes are repeat offenses, meaning the null hypothesis is valid and the mean is true.

More libraries

```
install.packages("janitor")
library(janitor)
```

Let us now do a Chi Square Test for Known Offenders and incidents but as p value is greater than alpha, both data are independent

```
table1<-table1 %>%  drop_na()
table1 %>%
 tabyl(Incidents, Known.Offenders2)%>%
chisq.test()
```

```
##
##  Pearson's Chi-squared test
##
## data:  .
## X-squared = 1526.5, df = 1480, p-value = 0.1954
```

CONCLUSION OF THE TEST The Chi Square Test Test for Known offenders and Incidents. The P value is 0.19 > alpha - 0.05 meaning that the two variables are dependent on each other meaning that known offenders have the highest number of incidents inferring that most hate crimes are repeat offenses.

FInd the Correlation between known offenders and incidents. As the value is +ve with close to 1, we can say both data change in same direction proportionally.

```
cor(table1$Known.Offenders2, table1$Incidents)
```

```
## [1] 0.9992666
```

A T test for the two data sets returns back p value greater than alpha meaning the data are dependent.

```
t.test(table1$Known.Offenders2, table1$Incidents, var.equal=TRUE)
```

```
##
##  Two Sample t-test
##
## data:  table1$Known.Offenders2 and table1$Incidents
## t = -0.3594, df = 84, p-value = 0.7202
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -874.5321  606.8112
## sample estimates:
## mean of x mean of y
##  624.9767  758.8372
```

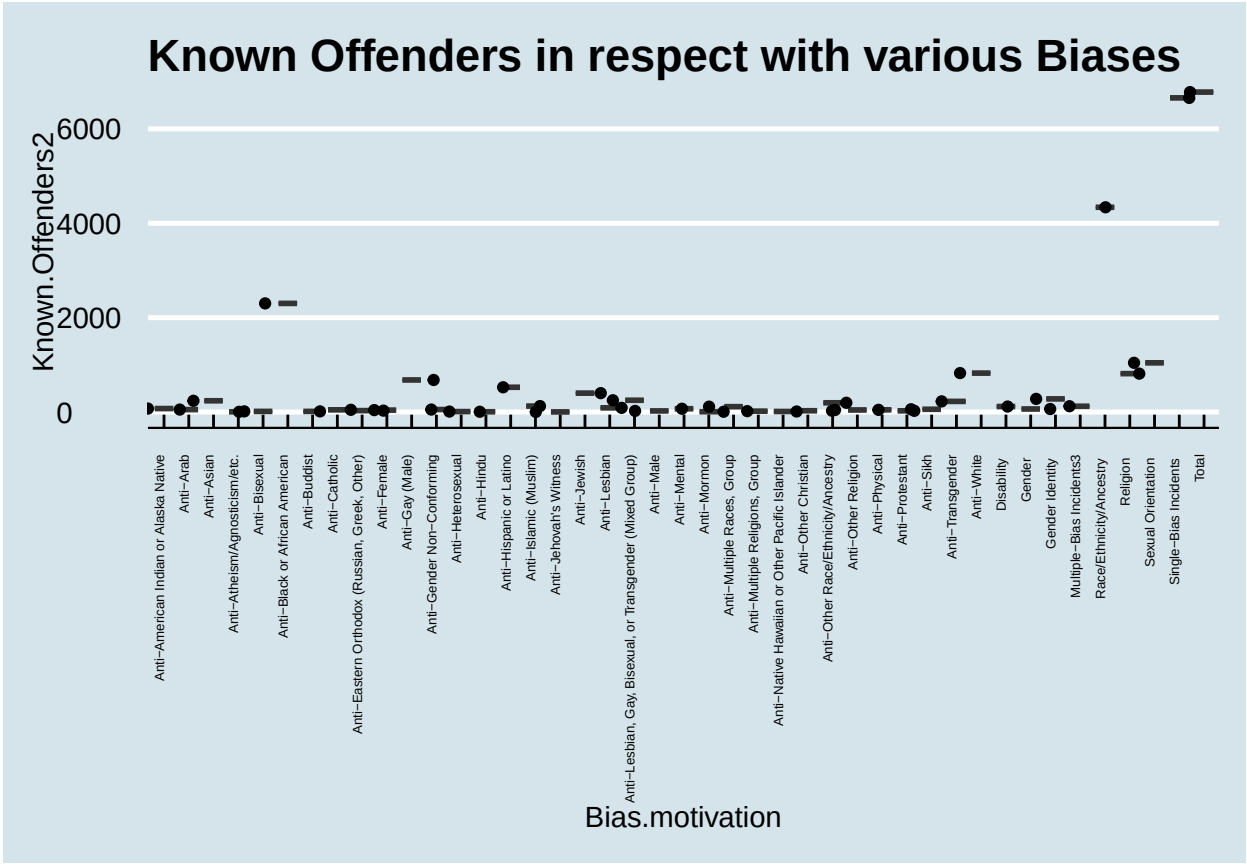
CONCLUSION OF THE TEST The T Test for Known offenders and Incidents. The P value is 0.72 > alpha - 0.05 meaning that the two variables are dependent on each other meaning that known offenders have the hoghest number of incidents inferring that most hate crimes are repeat offenses.

More libraries

```
library(tidyverse)
theme_set(theme_bw(16))
```

Box plot for known offenders and various biases

```
table1 %>%
ggplot(aes(Bias.motivation, Known.Offenders2 )) +
  geom_boxplot(outlier.shape = NA) +
  geom_jitter(width = 1) +
  theme_economist() +
  theme(axis.text.x = element_text(size=5, angle=90, hjust = 1)) +
  labs( title= "Known Offenders in respect with various Biases")
```



CONCLUSION OF THE TEST This boxplot graph gives a clear vivid picture on the regular biases for various Knwon offenders. We see that aside from the total, Single Bias incidents are the highest followed by Race/Ethnicity biased hate crimes which is the second highest.

FINAL CONCLUSIONS

This project aims to demonstrate the uses of R Studio in creating, developing and analysis of a structured data study with many various features, tests and more. The data set chosen is the 2020's FBI Hate Crime Statistics data set of USA with various races, Biases, locations and more. Through various analysis and tests, a defense on hypothesis that 1. Most hate crimes with bias 'Race/Ethnicity/Ancestry' happened mostly at the target victim's place of residence the most, followed by Supermarkets to be the next. Minority people are not safe in their own houses and public essential places like supermarkets in the US of A. 2. Most hate crimes are targeted toward Black victims the highest in US of A while the second highest is not even a race, rather a broader classification called Sexual Orientation where the Victims have various Sexual Orientations. This infers that the criminals in America hates the one race - Black more than they hate an entire broader group of populous with various Sexual Orientations. 3. Among juvenile, minors and kids, the hate crime is seriously directed toward Black Juveniles more than any other race present. With the second highest again not being another race but the entire victims of various sexual orientations, White race in USA commits the most range of hate crimes against any other race. The highest crimes are against other people and specifically Intimidation 4. Single Bias Incidents, Race Bias, African Bias are the leading bars showing which type of hate crimes are the highest 5. More Juvenile Hate crime victims lead to more adult hate crime victims 6. White race commits the most number of direct targeted hate crimes like Murder, Rape, Intimidation, followed by Black race being the second 7. Mean value of Known offenders is 600+ and intimidation is one of the biggest repeat offense

8. Known Offenders cause the most number of hate crimes.