BLACK LIVES MATTER
BL ER
BL ER
BL ER
BL ER
BL ER
BL ER

THE INFRINGEMENT OF NATURAL HUMAN RIGHTS BECAUSE OF POLICE BRUTALITY IN THE UNITED STATES

SSTA05AC PROJECT

BLACK LIVES MATTER

INTRODUCTION

Police brutality is the repression by personnel affiliated with law enforcement when dealing with suspects and civilians. Police violence has become a hot-button problem in the United States in recent years, particularly with Michael Brown's fatal shooting at Ferguson, Missouri, in 2014. The Black Lives Matter Movement founded in 2014 was a vocal aspect of the U.S. Police Brutality Movement through the organization of "die-ins," rallies, and protests against the killing by police of non-white men and women. The data from 2000 to 2016 of the civilian victims killed by police and their racial, gender, age and residence identities have been listed in the following dataset for our Data Analysis using R project.

DESCRIPTION OF THE DATASET

In this Dataset, there is a list of over 12,000 people who were shot dead by the police.
It also includes information about their name, race, gender, age, etc.
The 12 variables in the dataset are listed below:

UID: The unique identification for each individual. (Numerical variable)

Name: Name of the Individual. (Categorical variable)

Age: Age of the Individual. (Numerical variable)

Race: Race of the Individual . (Categorical variable)

Date: The Date of the Incident. (Numerical variable)

City: The City of the Incident. (Categorical variable)

State: The State of the Incident. (Categorical variable)

Manner_of_death: Manner in which the Individual was killed by the Police. (Categorical variable)

Armed: The arms with the Individual at the time of the incident. (Categorical variable)

Mental_illness: The mental state of the Individual Shot. (Categorical variable)

Flee: Information regarding whether the Individual tried to flee or not. (Categorical variable)

Link to the Dataset

https://www.kaggle.com/rishidamarla/individuals-killed-by-the-police

The Dataset at a Glance

> p=Police.Fatalities

The first 6 observations of the Dataset. The head () function gives an idea about the entire dataset appearance.

> head(p)

```
  UID                                    Name  Age  Gender   Race        Date        City
State
1  33                     Mark Edward Lewis  31   Male       02-01-2000 Boca Raton    FL
2  31                     John Mell Camacho  NA   Male       02-01-2000  Southgate    MI
3  32 Jonathon Miller / Brent J. Burchart  NA   Male       02-01-2000  Southgate    MI
4  30            Christopher Todd Frierson  22   Male White 02-01-2000   Picayune    MS
5  65                     Malcolm Ferguson  23   Male Black 03-01-2000      Bronx    NY
6 128                      Richard Stewart  37   Male       05-01-2000   Stockton    CA

  Manner_of_death      Armed Mental_illness  Flee
1            Shot        Gun          FALSE FALSE
2            Shot        Gun          FALSE FALSE
3            Shot        Gun          FALSE FALSE
4            Shot      Knife           TRUE FALSE
5            Shot                     FALSE FALSE
6            Shot Toy weapon           TRUE FALSE
```

The str() function compactly displays the internal structure of the Dataset.

> str(p)

```
'data.frame':     12491 obs. of  12 variables:
 $ UID            : int  33 31 32 30 65 128 167 244 284 286 ...
 $ Name           : chr  "Mark Edward Lewis" "John Mell Camacho" "Jonathon
Miller / Brent J. Burchart" "Christopher Todd Frierson" ...
 $ Age            : int  31 NA NA 22 23 37 19 37 20 25 ...
 $ Gender         : chr  "Male" "Male" "Male" "Male" ...
 $ Race           : chr  "" "" "" "White" …
 $ Date           : chr  "02-01-2000" "02-01-2000" "02-01-2000" "02-01-
2000" ...
 $ City           : chr  "Boca Raton" "Southgate" "Southgate" "Picayune"
...
 $ State          : chr  "FL" "MI" "MI" "MS" ...
 $ Manner_of_death: chr  "Shot" "Shot" "Shot" "Shot" ...
 $ Armed          : chr  "Gun" "Gun" "Gun" "Knife" ...
 $ Mental_illness : logi  FALSE FALSE FALSE TRUE FALSE TRUE ...
 $ Flee           : logi  FALSE FALSE FALSE FALSE FALSE FALSE ...
```

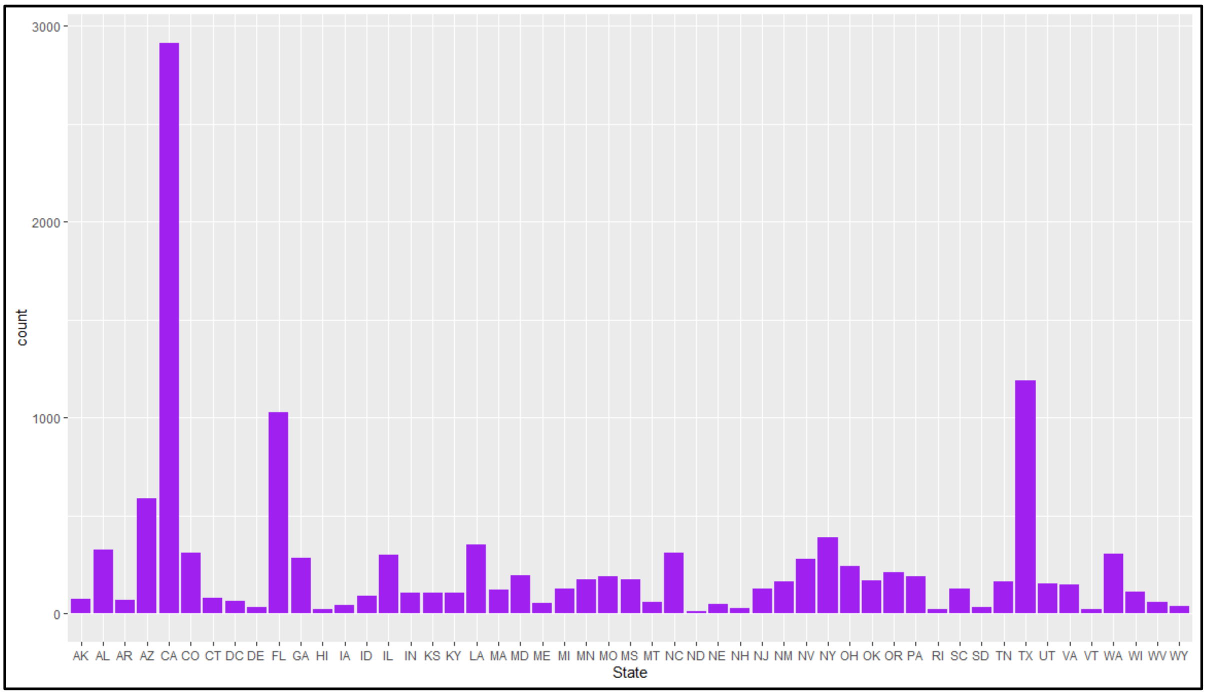
> summary(p)

```
      UID             Name                  Age             Gender         
 Min.   :    2   Length:12491       Min.   :  1.00   Length:12491      
 1st Qu.: 4102   Class :character   1st Qu.: 25.00   Class :character  
 Median : 7782   Mode  :character   Median : 33.00   Mode  :character  
 Mean   : 7760                      Mean   : 35.27                     
 3rd Qu.:11444                      3rd Qu.: 44.00                     
 Max.   :14980                      Max.   :107.00                     
                                    NA's   :233                        

     Race               Date               City              State          
 Length:12491       Length:12491       Length:12491       Length:12491      
 Class :character   Class :character   Class :character   Class :character  
 Mode  :character   Mode  :character   Mode  :character   Mode  :character  


 Manner_of_death       Armed           Mental_illness     Flee        
 Length:12491       Length:12491       Mode :logical   Mode :logical  
 Class :character   Class :character   FALSE:9862      FALSE:11931    
 Mode  :character   Mode  :character   TRUE :2629      TRUE :560      
```

Graphical Summary of the Dataset

1. State-wise Comparison of Fatalities

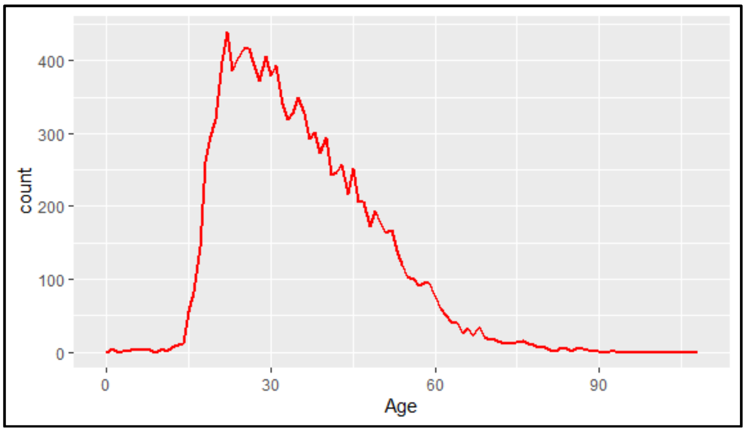
```
> p2=p%>%filter(State!='')
> a3<-ggplot(p2,aes(State))
> a3+geom_bar(fill='purple')
```



The State Wise comparison of the fatalities displays the distribution of fatalities among the different States of the US. As seen from the graph above, States like California have a huge number of fatalities followed by Texas and Florida.

2. Age-wise Comparison of Fatalities

```
> #Age-wise Comparison
> p3=p%>%filter(p$Age!='')
> a4<-ggplot(p3,aes(Age))
> a4+geom_freqpoly(binwidth=1,color='red',size=1)
```



The age wise graph displays the trend of age v/s the count of fatalities. As seen from the graph above, Individuals in the age group 20-30 are the main victims of the fatalities.

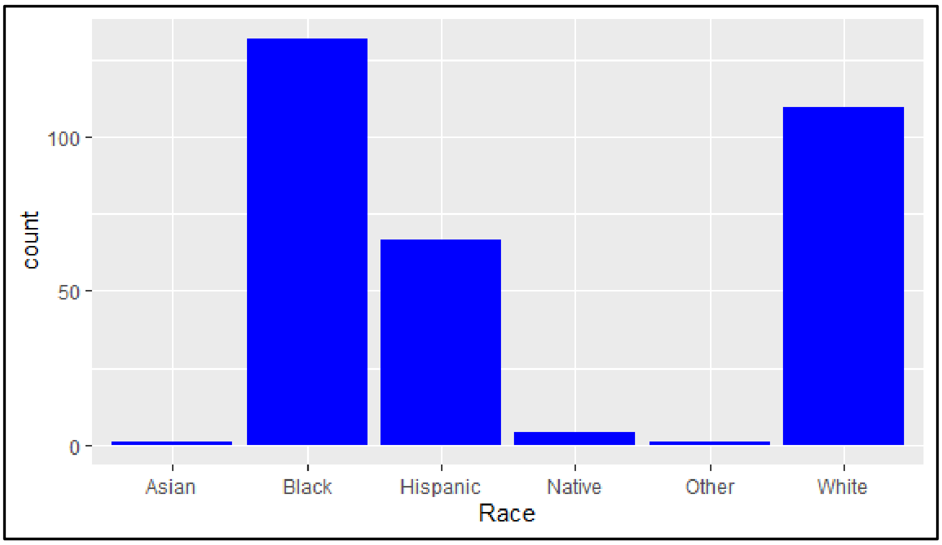
3. Arms & Race Comparison

```
> p5=p%>%filter(p$Armed!='',p$Race!='')
> a13=p5%>%select(Armed,Race)
```

3.1 Unarmed and Race

The following code and graph display the count of the individuals who were unarmed and their Race composition. The graph indicates that Black people were most likely to be killed even though unarmed followed by the White people and Hispanic people.

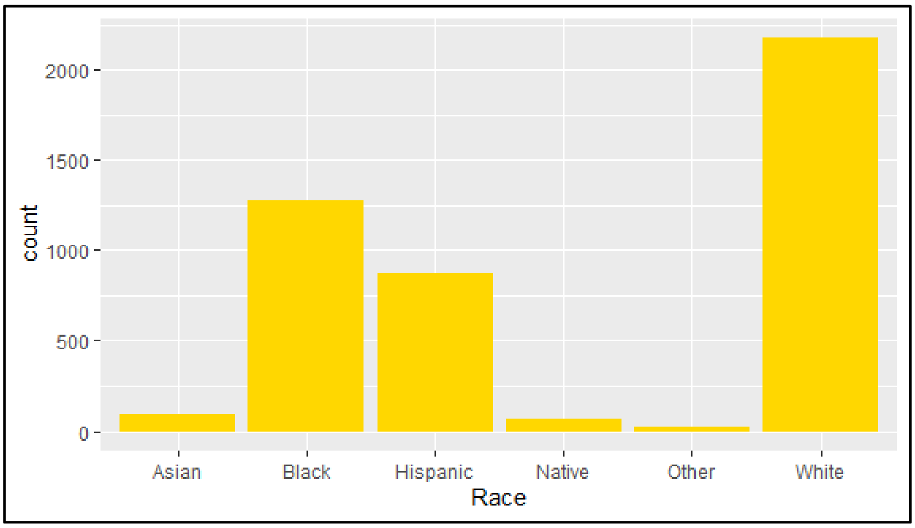
```
> a14=a13%>%filter(Armed=='Unarmed')
> a16<-ggplot(a14,aes(Race))
> a16+geom_bar(fill='blue')
```



3.2 Armed and Race

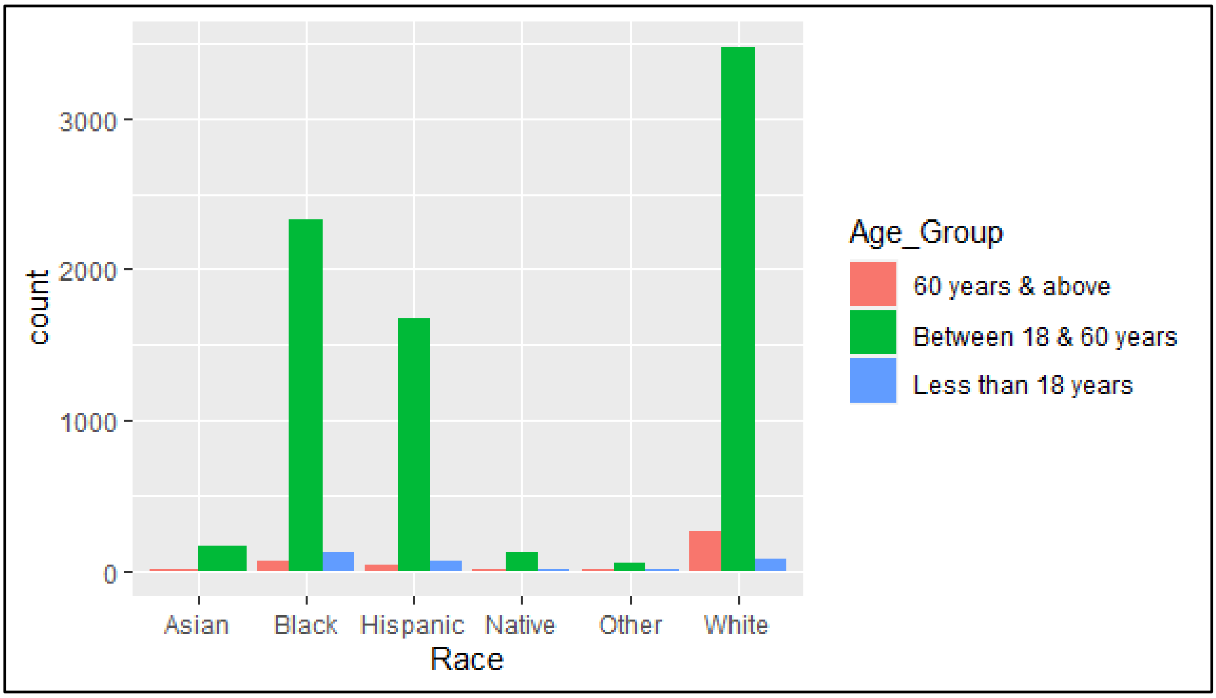
The following code and graph displays the count of the individuals who had arms during the killing and their Race composition. The graph indicates that White people were most likely to be armed followed by the Blacks and Hispanics.

```
> a15=a13%>%filter(Armed!='Unarmed')
> a17<-ggplot(a15,aes(Race))
> a17+geom_bar(fill='gold')
```



4. Age & Race Comparison

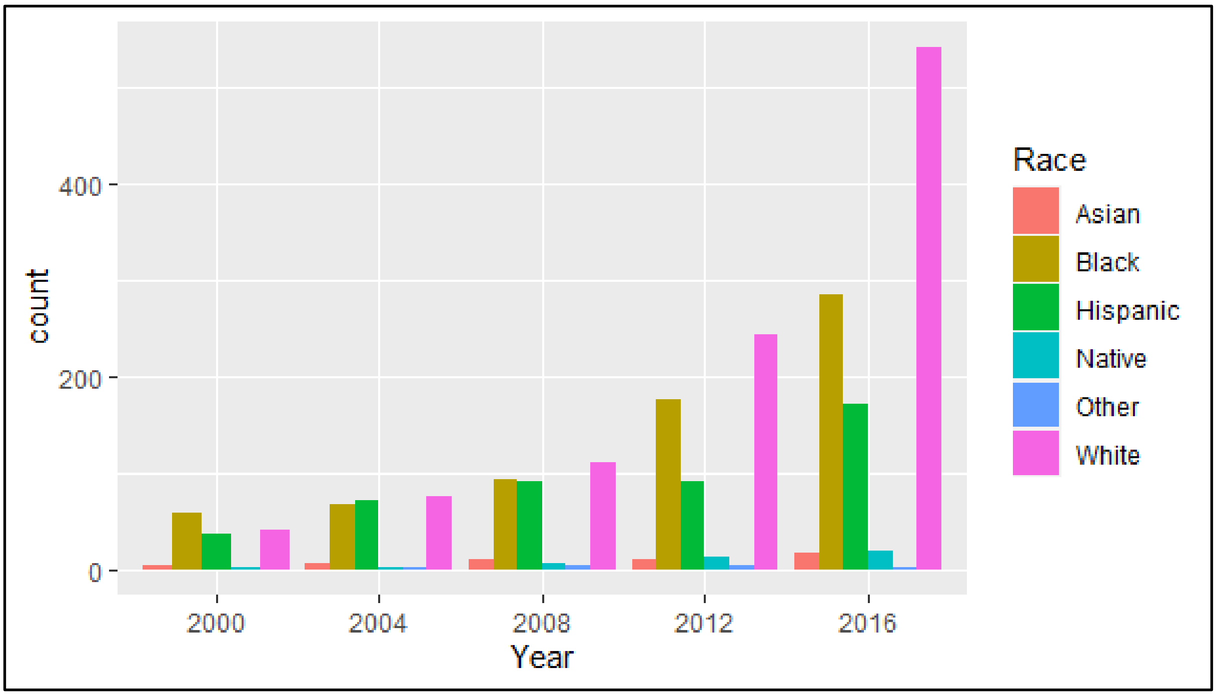
```
> a6=p4%>%select(Age,Race)
> a7=a6%>%filter(Age<18)
> bind2=cbind(a7,Age_Group='Less than 18 years')
> a9=a6%>%filter(Age>=18&Age<59)
> bind3=cbind(a9,Age_Group='Between 18 & 60 years')
> bind4=cbind(a11,Age_Group='60 years & above')
> a20=rbind(bind2,bind3,bind4)
> a21<-ggplot(a20,aes(Race,fill=Age_Group))
> a21+geom_bar(position='dodge')
```



The above graph displays the race wise distribution of fatalities among the different age groups. In the age group of less than 18 years, victims from the Black community were the highest. In the age group between 18 & 60 years, victims from the White community were the highest and the same trend was seen in the age group of 60 years and above.

5. Year & Race Comparison

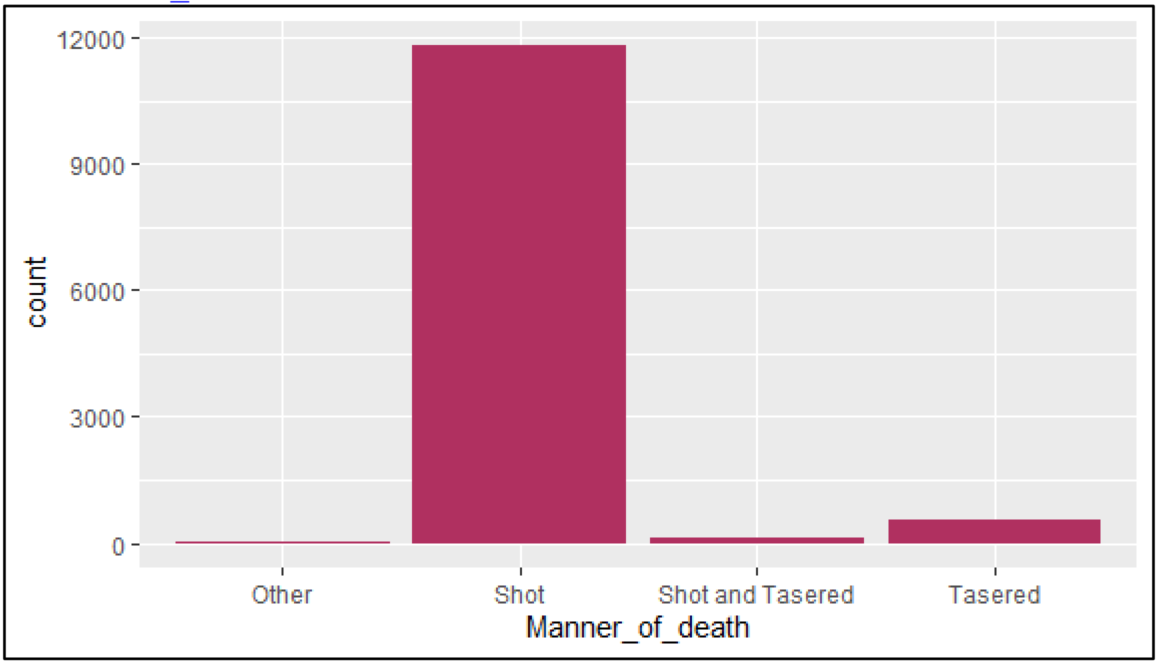
```
> p8=p%>%filter(p$Date!='',p$Race!='')
> a27=p8%>%select(Date,Race)
> Y_2000=a27%>%filter(grepl('2000',Date))
> bind_2000=cbind(Y_2000,Year='2000')
> Y_2004=a27%>%filter(grepl('2004',Date))
> bind_2004=cbind(Y_2004,Year='2004')
> Y_2008=a27%>%filter(grepl('2008',Date))
> bind_2008=cbind(Y_2008,Year='2008')
> Y_2012=a27%>%filter(grepl('2012',Date))
> bind_2012=cbind(Y_2012,Year='2012')
> Y_2016=a27%>%filter(grepl('2016',Date))
> bind_2016=cbind(Y_2016,Year='2016')
>
Y_bind=rbind(bind_2000,bind_2004,bind_2008,bind_2012,bind_2016
)
>
ggplot(Y_bind,aes(Year,fill=Race))+geom_bar(position='dodge')
```



The above graph displays the yearly trend of the number of fatalities and the Race wise composition of each year. The Asian and the Native community shows a slight increase over the years while there is a steep increase in the number of fatalities of the White community. The Black community also shows and increasing pattern along with the Hispanics over the years.

6. Graph on Manner of Death

```
> p14=p%>%filter(p$Manner_of_death!='')
> a41=p14%>%select(Manner_of_death)
> a42<-ggplot(a41,aes(Manner_of_death))
> a42+geom_bar(fill='maroon')
```



This graph displays the count of the manner in which the killings occurred. Most of the victims were shot by the Police followed by being Tasered. Some individuals were shot as well as tasered.

Descriptive Statistics

1. Maximum and Minimum State of Fatalities

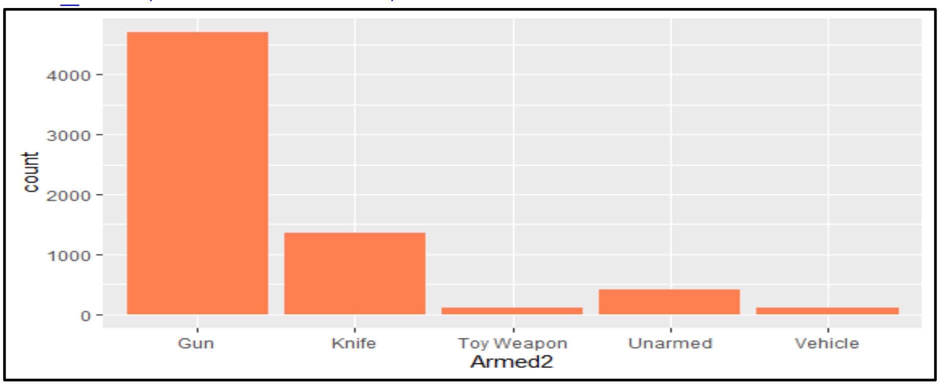
```
> p2=p%>%filter(State!='')
> sh=function(x) {
+   u=unique(x)
+   u[which.max(tabulate(match(x,u)))]
+   }
> u1=unique(p2$State)
> u1[which.max(tabulate(match(p2$State,u1)))]
[1] "CA"
> sl=function(x) {
+   u=unique(x)
+   u[which.min(tabulate(match(x,u)))]
+ }
> u2=unique(p2$State)
> u2[which.min(tabulate(match(p2$State,u2)))]
[1] "ND"
```

Thus, California was the state with the highest number of fatalities and North Dakota was the state with the minimum number of fatalities.

2. Most used arms by the victims

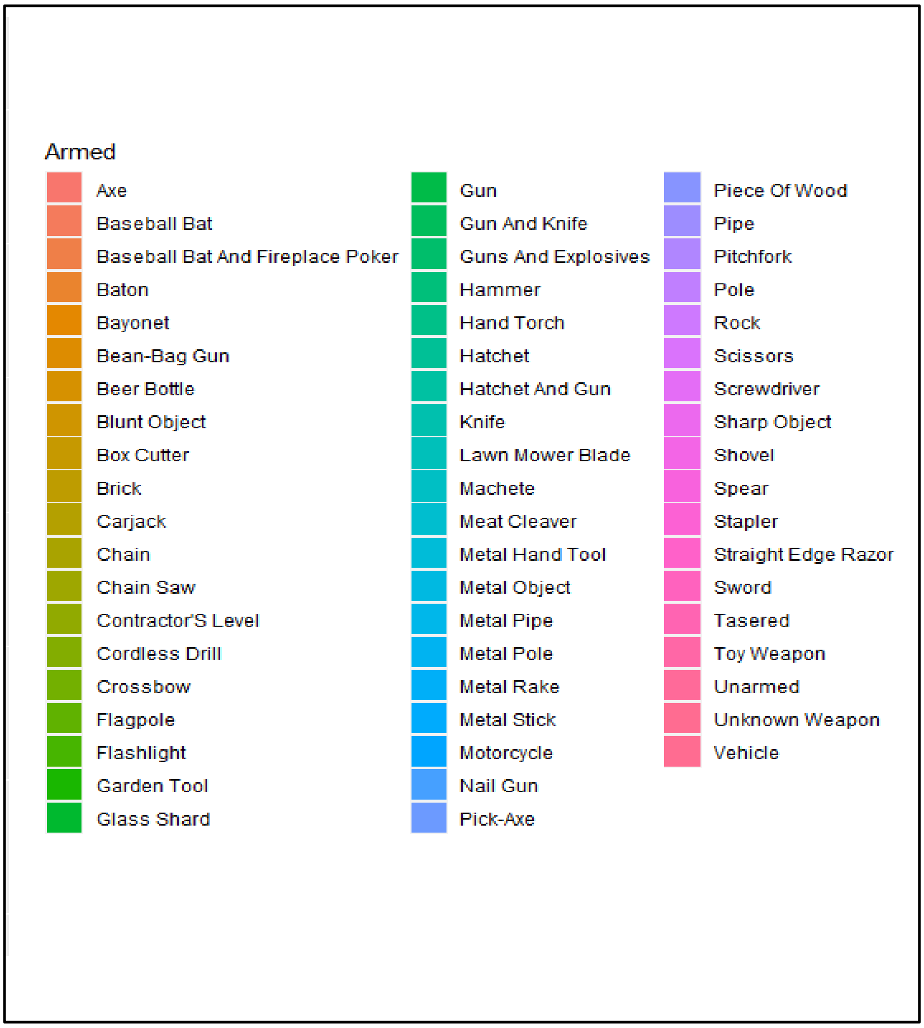
The graph displays the top 5 most used arms by the victims. As seen from the graph, the majority of the victims had a gun while the second most common arm was a knife. The graph also displays how many of the number of victims were unarmed.

```
>p10=p%>%filter(p$Armed=='Gun'|p$Armed=='Knife'|p$Armed=='Vehi
cle'|p
$Armed=='Toy Weapon'|p$Armed=='Toy weapon'|p$Armed=='Unarmed')
> a45=p10%>%select(Armed)
> a46=a45%>%filter(Armed=='Toy Weapon'|Armed=='Toy weapon')
> bind13=cbind(a46,Armed2='Toy Weapon')
> a47=a45%>%filter(Armed!='Toy Weapon',Armed!='Toy weapon')
> bind14=cbind(a47,Armed2=a47[,1])
> a48=rbind(bind13,bind14)
> a30=ggplot(a48,aes(Armed2))
> a30+geom_bar(fill='coral')
```



3. Details of how the victims were armed

```
> p9=p%>%filter(p$Armed!='Toy weapon',p$Armed!='')
> a28=p9%>%select(Race,Armed)
> a29=ggplot(a28,aes(Race,fill=Armed))
> a29+geom_bar(position='dodge')
```



This displays the arms that were mentioned in the dataset. From toy weapons to Baseball bats, the different kinds of arms mentioned give an interesting insight. Even objects including Stapler, Piece of wood, etc. were included as arms.

Statistical Analysis

The Chi- Square test of independence is used to detect the independence or dependency of two variables.

1. Test of Independence between the manner of death of the victim & how they were armed

H0: The Manner of death of the victim & how they were armed is independent.
H1: Not H0

```
> p1=p%>%filter(p$Manner_of_death!='',p$Armed!='')
> a1=p1%>%select(Manner_of_death,Armed)
> a2_1=a1%>%filter(Armed=='Gun')
> bind10=cbind(a2_1,Armed2='Gun')
> a2_2=a1%>%filter(Armed=='Knife')
> bind11=cbind(a2_2,Armed2='Knife')
> a2_3=a1%>%filter(Armed!='Gun',Armed!='Knife')
> bind12=cbind(a2_3,Armed2='Other')
>
a2_bind=rbind(bind10,bind11,bind12)%>%select(Armed2,Manner_of_
death)
> y1=table(a2_bind)
> chi_y1=chisq.test(y1); chi_y1

Warning message:
In chisq.test(y1) : Chi-squared approximation may be incorrect
```

Pearson's Chi-squared test

```
data:  y1
X-squared = 246.37, df = 6, p-value < 2.2e-16

> chi_y1$expected

       Manner_of_death
Armed2       Other      Shot  Shot and Tasered   Tasered
  Gun    4.8313766 4560.8195         79.37262 57.976519
  Knife 1.3981509 1319.8544         22.96962 16.777810
  Other 0.7704726  727.3261         12.65776  9.245671
```

Since some of the expected frequencies are less than 5, we pool some of the frequencies.

```
> z1=data.frame(y1[,1]+y1[,4],y1[,2],y1[,3])
> chi_z1=chisq.test(z1); chi_z1
```

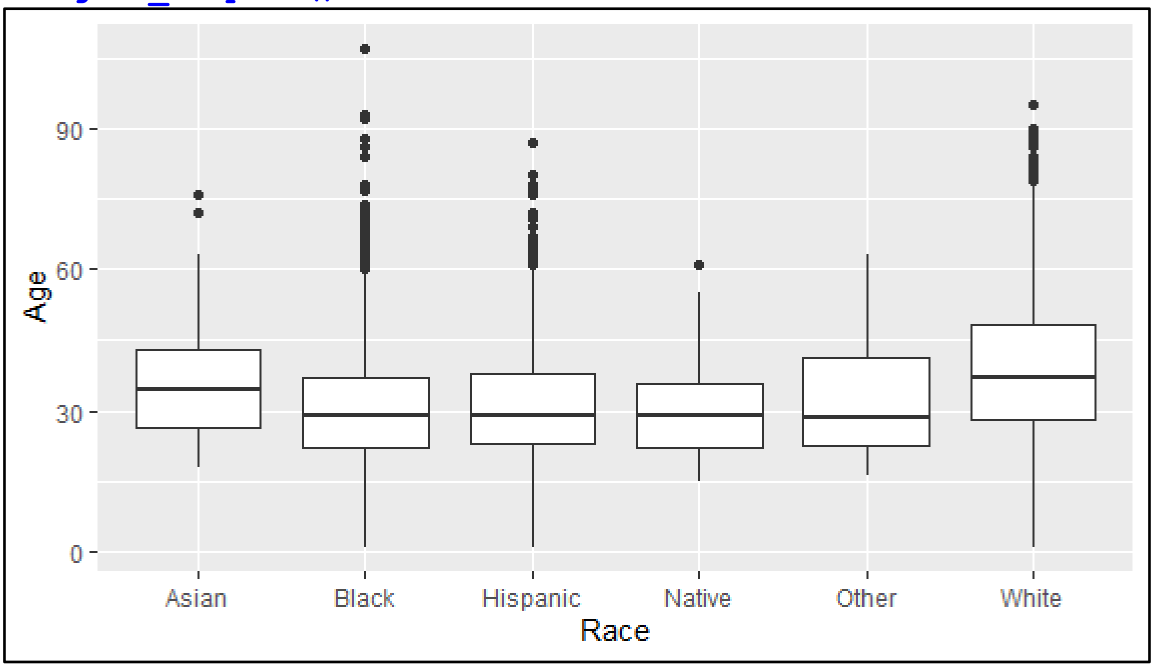
```
Pearson's Chi-squared test

data:  z1
X-squared = 240.13, df = 4, p-value < 2.2e-16
```

Since $p_{cal} < p_{tab}$, we reject H0 at 5% L.O.S.
Hence, we conclude that the manner of death of the victim is dependent on how they were armed at 5% L.O.S.

2. Test of Independence between the Age & Race of the victim

H0: Age & Race of the victim is independent of each other.
H1: Not H0

```
> p4=p%>%filter(p$Age!='',p$Race!='')
> a5<-ggplot(p4,aes(Race,Age))
> a5+geom_boxplot()
```




```
> a6=p4%>%select(Age,Race)
> a7=a6%>%filter(Age<18)
> a8=a7%>%count(Race)
> bind1=rbind(list('Asian',0),a8)
> a9=a6%>%filter(Age>=18&Age<59)
> a10=a9%>%count(Race)
> a11=a6%>%filter(Age>=60)
> a12=a11%>%count(Race)
> y2=data.frame(bind1$n,a10$n,a12$n)
> chi_y2=chisq.test(y2); chi_y2
Warning message:
In chisq.test(y2) : Chi-squared approximation may be incorrect
```

```
	Pearson's Chi-squared test

data:  y2
X-squared = 144.04, df = 10, p-value < 2.2e-16

> chi_y2$expected
        bind1.n      a10.n      a12.n
[1,]   5.035205  155.51189   7.452902
[2,]  74.988582 2316.01641 110.995005
[3,]  52.629876 1625.46955  77.900571
[4,]   3.896289  120.33658   5.767127
[5,]   1.438630   44.43197   2.129401
[6,] 114.011418 3521.23359 168.754995
```

Since some of the expected frequencies are less than 5, we pool some of the frequencies.

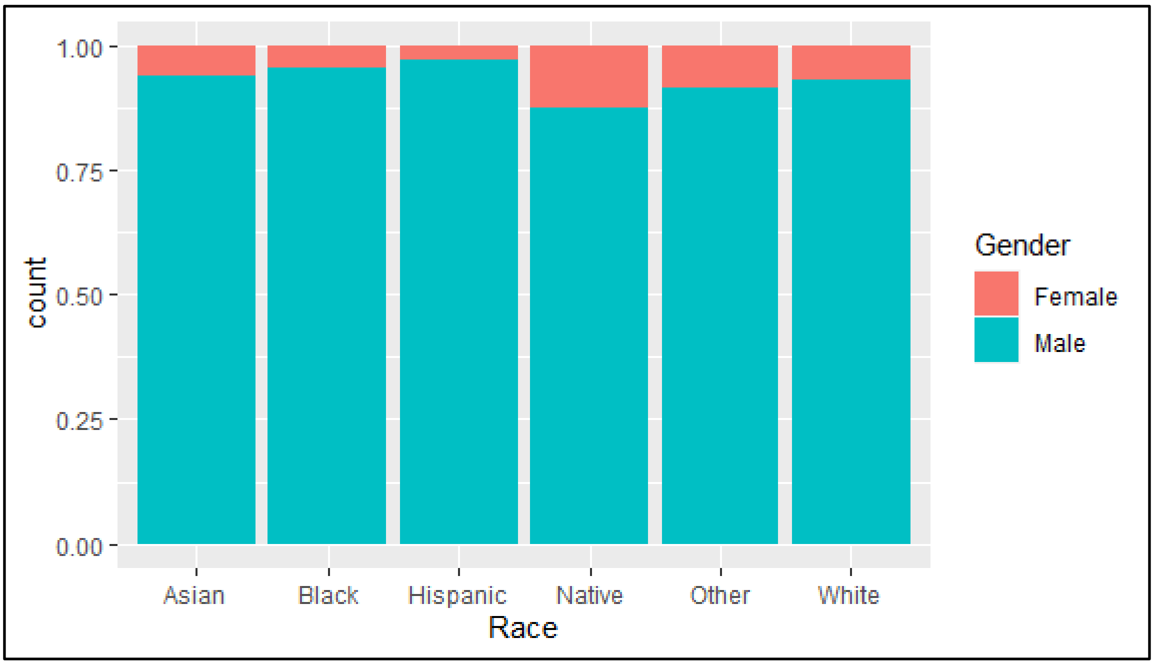
```
> t2=t(y2); t2
        [,1] [,2] [,3] [,4] [,5] [,6]
bind1.n    0  113   61    5    1   72
a10.n    161 2320 1665  123   45 3469
a12.n      7   69   30    2    2  263
> z2=data.frame(t2[,1]+t2[,2],t2[,3]+t2[,4]+t2[,5],t2[,6])
> chi_z2=chisq.test(z2); chi_z2

	Pearson's Chi-squared test

data:  z2
X-squared = 131.43, df = 4, p-value < 2.2e-16
```

Since $p_{cal} < p_{tab}$, we reject H0 at 5% L.O.S.
Hence, we conclude that the age and race of the victim are dependent at 5% L.O.S.

3. Test of Independence between the Gender & Race of the victim.

H0: Gender & Race of the victim are independent of each other.
H1: Not H0

```
> p6=p%>%filter(p$Gender!='',p$Race!='')
> a18=p6%>%select(Gender,Race)
> a19<-ggplot(a18,aes(Race,fill=Gender))
> a19+geom_bar(position='fill')
```




```
> y3=table(p6$Gender,p6$Race)
> chi_y3=chisq.test(y3); chi_y3
Warning message:
In chisq.test(y3) : Chi-squared approximation may be incorrect

	Pearson's Chi-squared test

data:  y3
X-squared = 57.742, df = 5, p-value = 3.555e-11

> chi_y3$expected

               Asian     Black   Hispanic     Native     Other     White
  Female    9.093512  133.2489   93.72099   6.833275  2.523055  202.5803
  Male    163.906488 2401.7511 1689.27901 123.166725 45.476945 3651.4197
```

Since some of the expected frequencies are less than 5, we pool some of the frequencies.

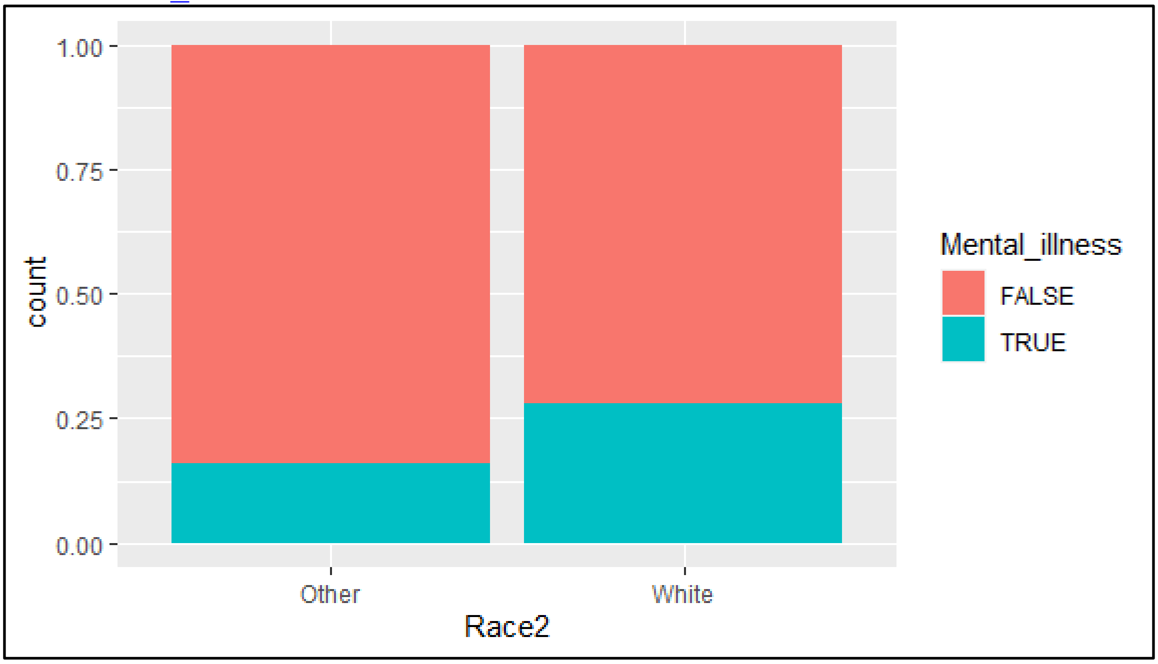
```
> z3=data.frame(y3[,1],y3[,2],y3[,3],y3[,4]+y3[,5],y3[,6])
> chi_z3=chisq.test(z3); chi_z3

	Pearson's Chi-squared test

data:  z3
X-squared = 56.631, df = 4, p-value = 1.479e-11
```

Since $p_{cal} < p_{tab}$, we reject H0 at 5% L.O.S.
Hence, we conclude that the gender and race of the victim are dependent at 5% L.O.S.

4. Test of Independence between the Mental Illness & Race of the victim.

H0: Mental Illness & Race of the victim are independent of each other.
H1: Not H0

```
> p7=p%>%filter(p$Mental_illness!='',p$Race!='')
> a22=p7%>%select(Mental_illness,Race)
> a23=a22%>%filter(Race=='White')
> bind5=cbind(a23,Race2='White')
> a24=a22%>%filter(Race!='White')
> bind6=cbind(a24,Race2='Other')
> a25=rbind(bind5,bind6)
> a26<-ggplot(a25,aes(Race2,fill=Mental_illness))
> a26+geom_bar(position='fill')
```




```
> y4=table(a24)
> chi_y4=chisq.test(y4); chi_y4

	Pearson's Chi-squared test

data:  y4
X-squared = 36.702, df = 4, p-value = 2.075e-07
```

Since $p_{cal} < p\_tab$, we reject H0 at 5% L.O.S.
Hence, we conclude that the mental illness of the victims depends on their race at 5% L.O.S.

5. Test of Independence between the Gender & Mental Illness of the victim.

H0: Mental Illness & Gender of the victim are independent of each other.
H1: Not H0

```
> p15=p%>%filter(p$Gender!='',p$Mental_illness!='')
> a43=p15%>%select(Gender,Mental_illness)
> y5=table(a43)
> chi_y5=chisq.test(y5); chi_y5

      Pearson's Chi-squared test with Yates' continuity correction

data:  y5
X-squared = 20.392, df = 1, p-value = 6.31e-06
```

Since $p_{cal} < p_{tab}$, we reject H0 at 5% L.O.S.
Hence, we conclude that the mental illness of the victim depends on their gender at 5% L.O.S.

6. Test of Independence between whether the victim tried to flee & whether they had any mental illness

H0: Mental Illness & Fleeing of the victim are independent of each other.
H1: Not H0

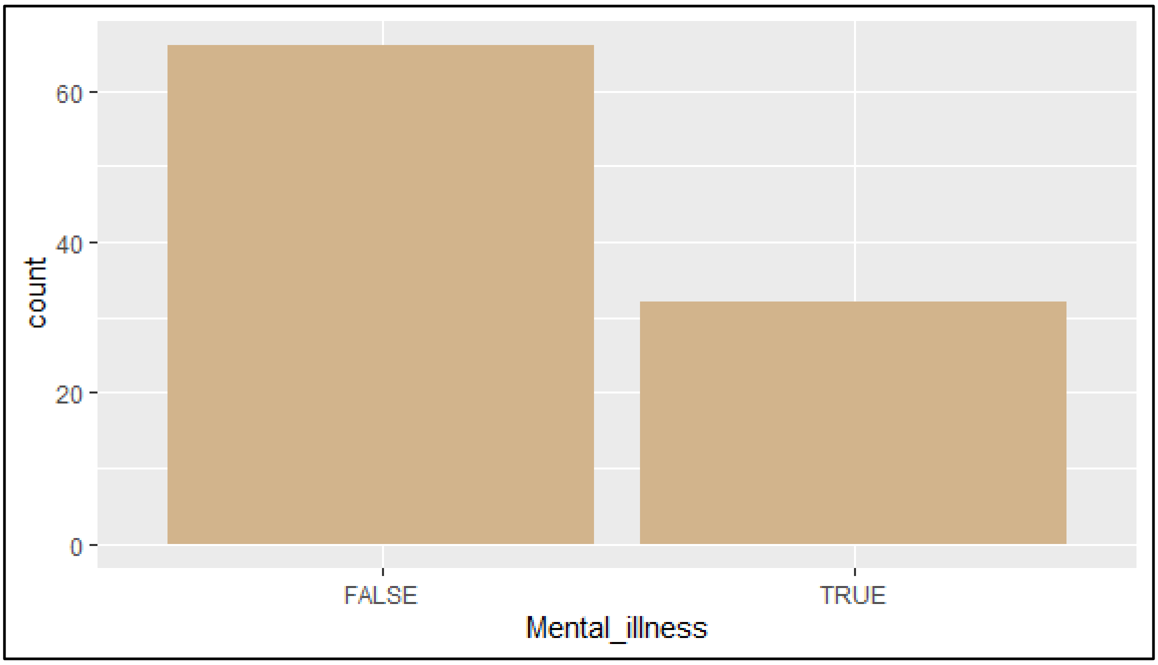
```
> p16=p%>%filter(p$Flee!='',p$Mental_illness!='')
> a44=p16%>%select(Flee,Mental_illness)
> y6=table(a44)
> chi_y6=chisq.test(y6); chi_y6

      Pearson's Chi-squared test with Yates' continuity correction

data:  y6
X-squared = 34.485, df = 1, p-value = 4.296e-09
```

Since $p_{cal} < p_{tab}$, we reject H0 at 5% L.O.S.
Hence, we conclude that whether the victim tried to flee depends on whether they had a mental illness at 5% L.O.S.

Some interesting analysis of the dataset

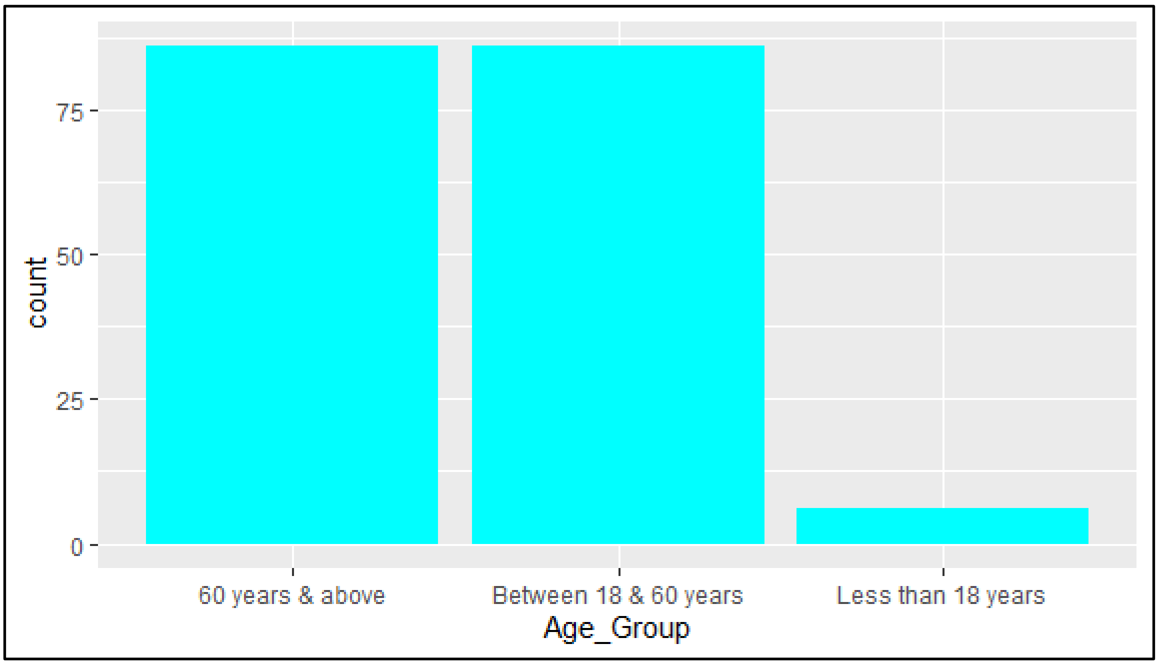
Toy Weapon & Mental Illness

```
> p12=p%>%filter(p$Armed=='Toy Weapon'|p$Armed=='Toy weapon',
p$Mental_illness!='')
> a33=p12%>%select(Armed,Mental_illness)
> a34<-ggplot(a33,aes(Mental_illness))
> a34+geom_bar(fill='tan')
```



Toy Weapon & Age

```
>      p13=p%>%filter(p$Armed=='Toy      Weapon'|p$Armed=='Toy
weapon',p$Age!='')
> a35=p13%>%select(Armed,Age)
> a36=a35%>%filter(Age<18)
> bind7=cbind(a36,Age_Group='Less than 18 years')
> a37=a35%>%filter(Age>=18&Age<59)
> bind8=cbind(a37,Age_Group='Between 18 & 60 years')
> a38=a35%>%filter(Age>=60)
> bind9=cbind(a37,Age_Group='60 years & above')
> a39=rbind(bind7,bind8,bind9)
> a40<-ggplot(a39,aes(Age_Group))
> a40+geom_bar(fill='cyan')
```



According to research, from The Washington Post, in the previous two years, police around the country have murdered dozens of people carrying realistic toy weapons and replicas. Thirty-eight of those killed suffered from mental illness, according to family accounts and police reports, and 53 were carrying pneumatic BB or pellet guns. The latest was the 16-year-old boy shot dead by a Maryland State Police squad, after the teen targeted a toy weapon at the officer, in April 2021 which sparked protest across the United Nations against the ongoing police brutality.

Conclusion and Highlights of the Dataset

- The dataset helps identify the victims, the race and the age of the victims.
- There were more than 12000 police fatalities over the years.
- California was the state with the highest no of fatalities and North Dakota was the state with the lowest no of fatalities.
- White people were most likely to be armed during the incident while the Black people were most likely to be killed even without an arm.
- The manner of death shows how the majority of the victims were shot followed by being tasered. Some of the victims were shot and tasered as well.
- Many of the victims were unarmed.
- Under the armed section, arms such as scissors, toy weapons, vehicles, baseball bats were included. Even objects including Stapler, Piece of wood, etc. were included as arms.

Police violence is a major cause of death in the United States among young males. Over the course of life, around 1 out of 1,000 black men may expect the cops to kill them. The risk of murder of men and women and all racial and ethnic groups at police peaks between 20 and 35 years. Black women and men and Americans and Alaskans Indigenous women and men are much more likely to be killed by the police than white women and men. Latino men are more likely than white men to be slain by the police.

The end of Police brutality is extremely important.

Link for the R Script:

https://drive.google.com/file/d/1VgP0D-ggMWndUuNGS_ly8c09vTjUGB9U/view?usp=drivesdk