Math 130 Exploratory Data Analysis

Maurice Ledoyen

2/21/2022

```
police <- readxl::read_excel(
  '/Users/mauriceledoyen/Desktop/Math_130/data/fatal-police-shootings-data.xlsx')
```

Description of the Data

I chose the "Police Shootings" (police) dataset which provides the characteristics of individuals killed by police in 2015. The variables I will be exploring in this data analysis will be (1) whether the victim was displaying signs of mental illness (signs_of_mental_illness), (2) the race of the victim (race), (3) the age of the victim (age). I am interested in exploring if race plays a role in police shootings? Does mental illness increase the likelihood of being involved in a police shooting? Is age a factor in police shootings?

Univariate Statistics

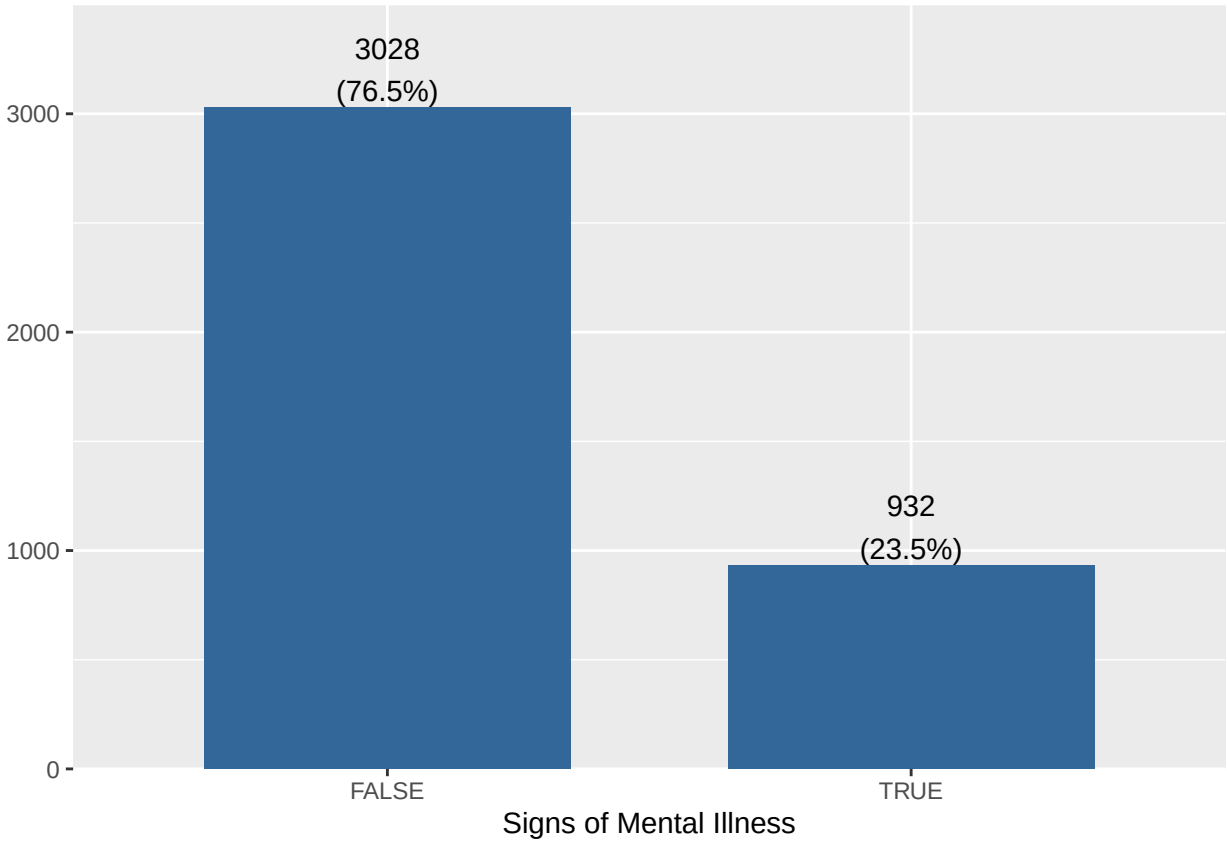
Signs of Mental Illness

The 'signs_of_mental_illness' variable is a nominal variable. 'True' represents signs of mental illness are present. 'False' represents signs of mental illness are not present.

```
table(police$signs_of_mental_illness)
```

```
##
## FALSE  TRUE
##  3028   932
```

```
library(sjPlot)
plot_frq(police$signs_of_mental_illness, axis.title='Signs of Mental Illness')
```



More than 76% of police shooting victims were not showing signs of mental illness at the time of the incident.

Race

The 'race' variable is nominal variable. 'A' represents Asian, 'B' represents black, 'H' represents Hispanic, 'N' represents Native American, 'O' represents other, 'W' represents white.

```
table(police$race)
```

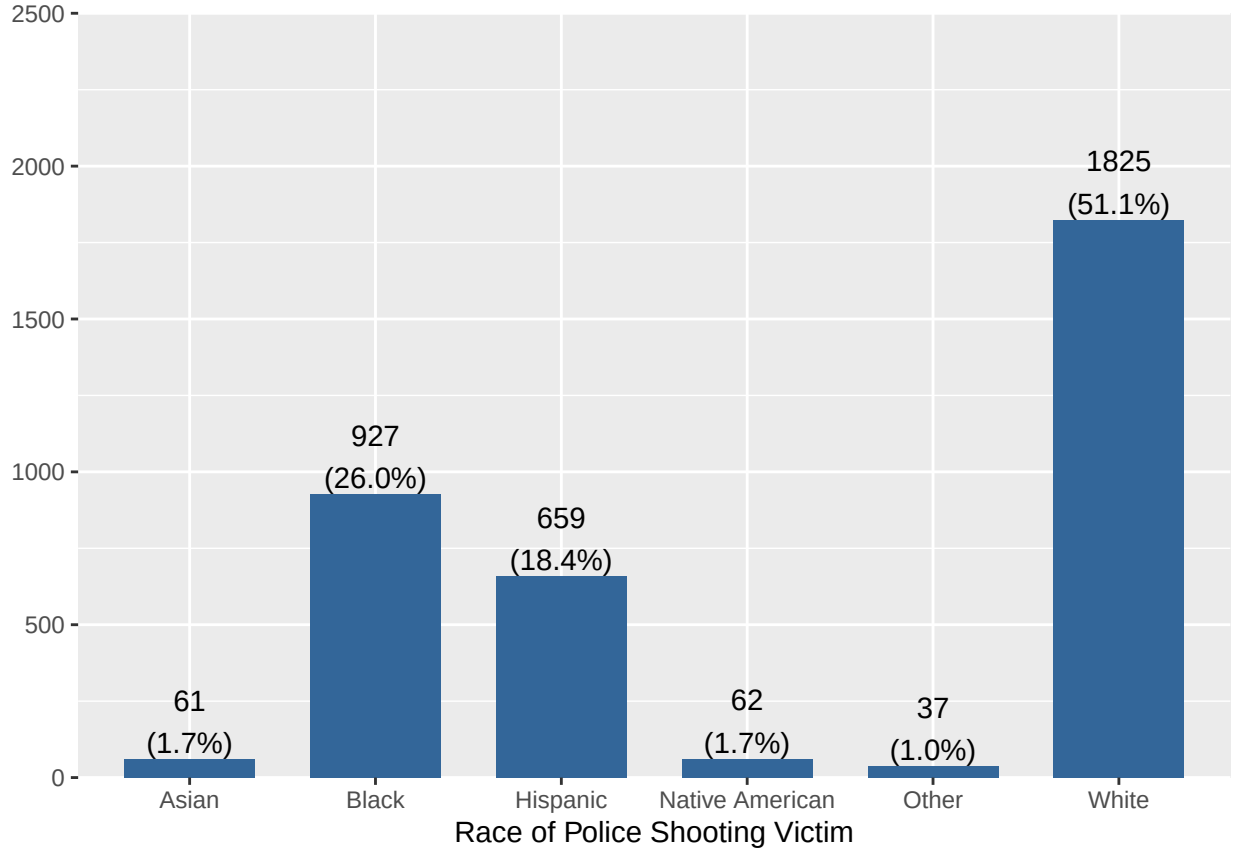
```
##
##    A    B    H    N    O    W
##   61  927  659   62   37 1825
```

```
library(forcats)
police$fctrace <- fct_recode(police$race, "Asian"="A", "Black"="B",
                             "Hispanic"="H", "Native American"="N", "Other"="O", "White"="W")
```

```
table(police$fctrace)
```

```
##
##            Asian            Black         Hispanic Native American
##               61              927              659               62
##            Other            White
##               37             1825
```

```
plot_frq(police$fctrace, axis.title = "Race of Police Shooting Victim")
```



The highest percentage of police shooting victims were White at 51.1%, followed by Black at 26%, and Hispanic at 18.4% of the total.

Age

The 'age' variable is a continuous variable. The age variable represents the age of the victim at the time of the shooting.

```
summary(police$age)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    6.00   27.00   35.00   36.85   45.00   91.00     152
```

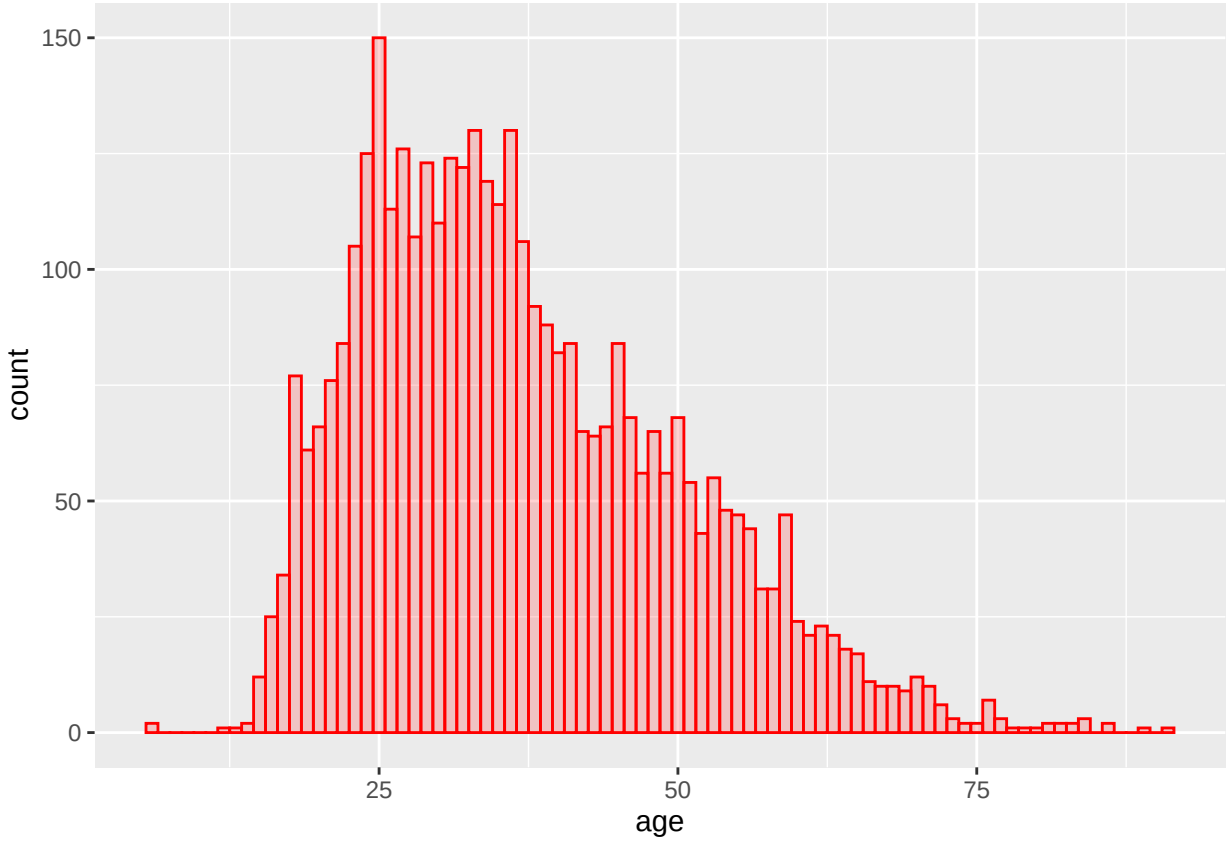
```
IQR(police$age, na.rm=TRUE)
```

```
## [1] 18
```

```
library(ggplot2)

ggplot(police, aes(x=age)) + geom_histogram(binwidth=1,
                                            color='red', alpha=.3 , fill="#FF6666")
```

```
## Warning: Removed 152 rows containing non-finite values (stat_bin).
```



The average age of police shooting victims was 36.85, a median age of 35, and an interquartile range of 18 (1st quartile = 27, 3rd quartile = 45).

Bivariate Exploration

Comparing age and race.

```
police$age_group <- cut(police$age, seq(0,100, 20))
table(police$fctrace, police$age_group)
```

```
##
##                   (0,20] (20,40] (40,60] (60,80] (80,100]
##   Asian                6      34      19       1        0
##   Black              109     619     159      24        0
##   Hispanic            57     437     136      12        0
##   Native American      6      47       9       0        0
##   Other                4      23       9       0        0
##   White               86     946     652     121        8
```

The first table shows that regardless of race, the majority of police shooting victims are between the ages of 20 to 40, followed by the next highest age group of 40 to 60 years old.

Comparing race and signs of mental illness.

```
table(police$fctrace, police$signs_of_mental_illness)
```

```
##
##                   FALSE TRUE
##   Asian              43   18
##   Black             785  142
##   Hispanic          535  124
##   Native American    53    9
##   Other              28    9
##   White            1282  543
```

The second table show signs of mental illness by race of victims at the time of the shooting incident.

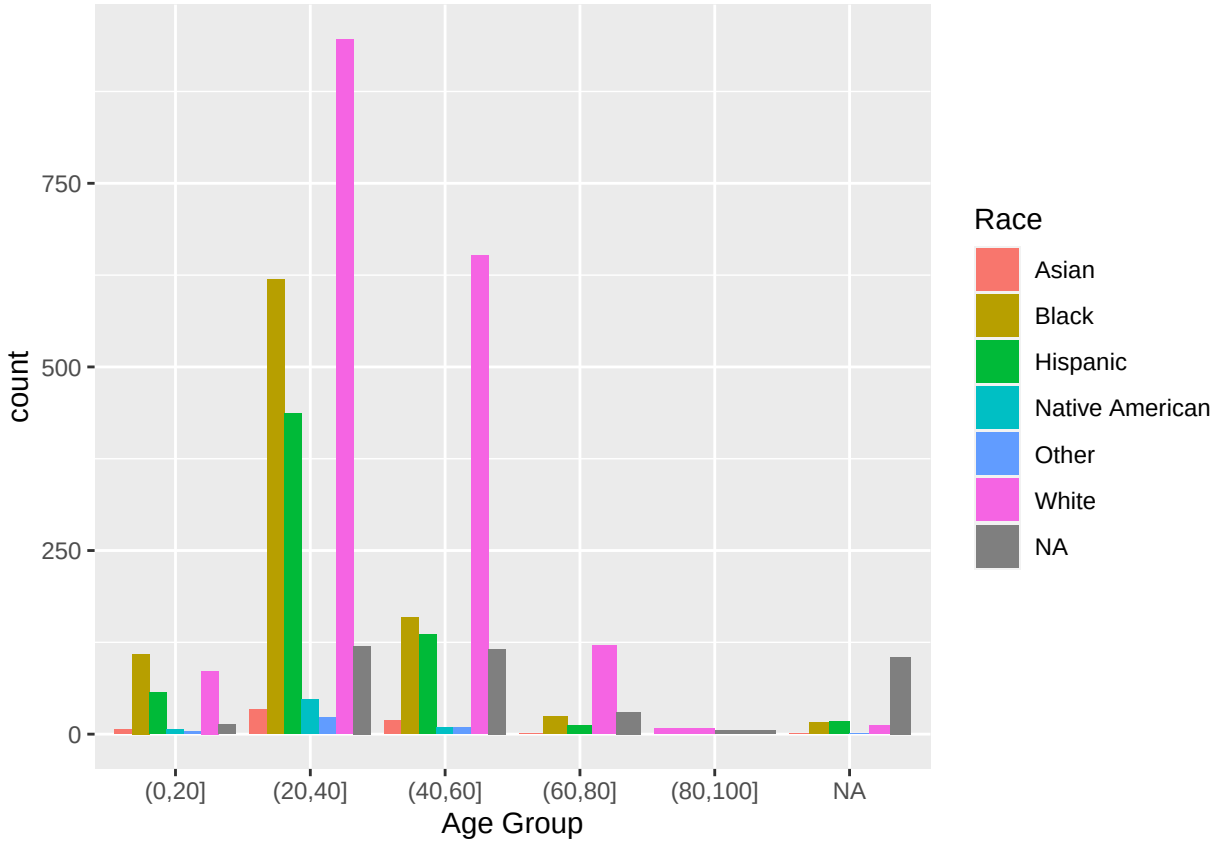
Comparing age group and signs of mental illness.

```
table(police$age_group, police$signs_of_mental_illness)
```

```
##
##            FALSE TRUE
##   (0,20]     234   47
##   (20,40]   1732  494
##   (40,60]    806  294
##   (60,80]    125   63
##   (80,100]     7    6
```

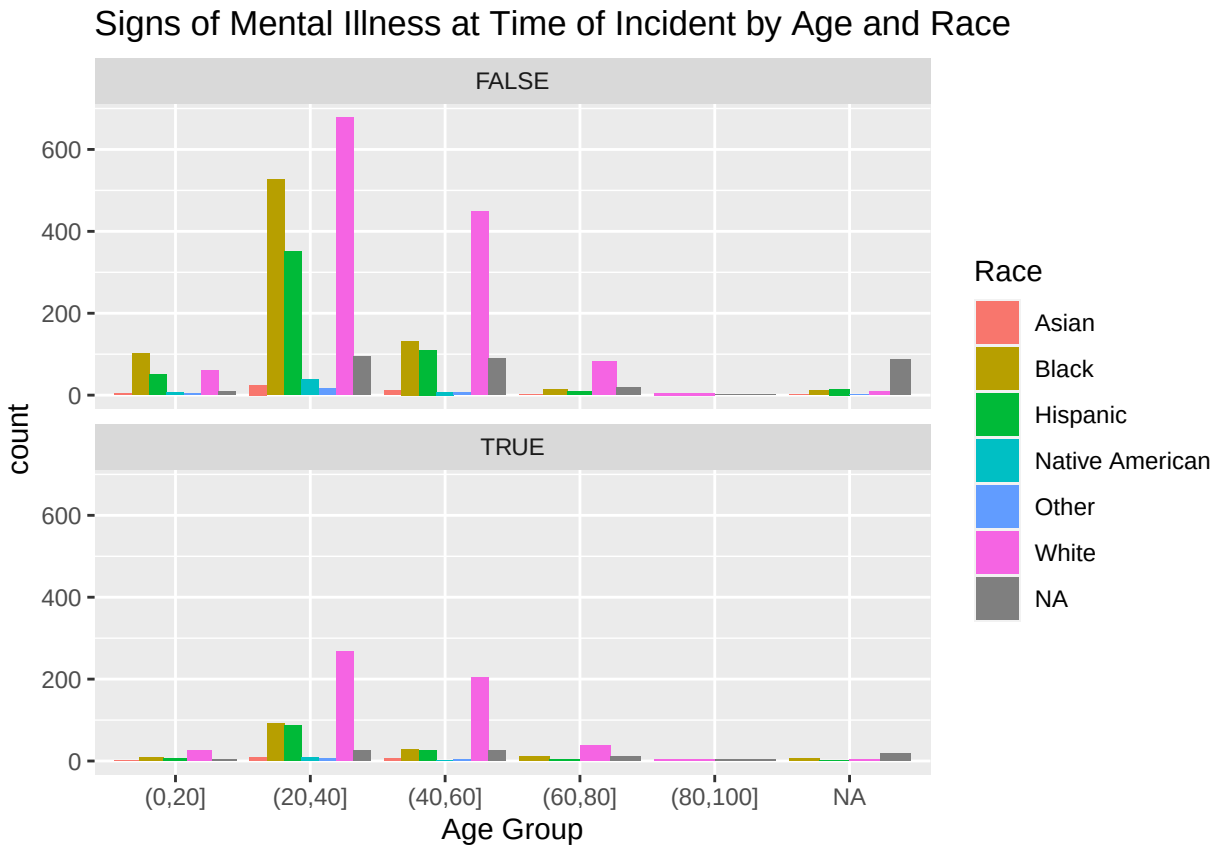
The third table shows signs of mental illness by age group.

```
library(RColorBrewer)
ggplot(police, aes(x=age_group, fill=fctrace)) +
  geom_bar(position='dodge') +
  scale_x_discrete('Age Group') +
  scale_fill_discrete(name="Race")
```



The bar chart shows the relationship between age and race of police shooting victim. Most victims regardless of race are between the ages of 20 to 40 with white, black and Hispanics comprising the majority. The bar chart also shows that there are more <20 year old black police shooting victims than any other race, and more white 60 to 80 year old police shooting victims than all other races combined for this age group.

```
ggplot(police, aes(x=age_group, fill=fctrace)) +
  geom_bar(position='dodge') + scale_fill_discrete(name="Race") +
  scale_x_discrete('Age Group') + facet_wrap(~signs_of_mental_illness, ncol = 1) +
  ggtitle('Signs of Mental Illness at Time of Incident by Age and Race')
```



This bar chart shows the number of people killed by police shootings broken down by age group and race and paneled by signs of mental illness at the time of the incident. The top chart shows age group and race with no sign of mental illness at the time of the shooting incident. The bottom chart shows age group and race with signs of mental illness. The trends between age group and race, with or without mental illness present, are relatively consistent. However, the chart does show a higher frequency of black and Hispanic victims, ages 20-60, with no mental illness present are killed by police when compared with the mental illness present group.

Conclusion

This exploratory data analysis showed several interesting characteristics of individuals killed by police in 2015. By far the highest frequency of victims killed by police were ages 20-40, with the average victim age of 36.85. The race statistics show that white people make up the majority of victims killed, followed by black and Hispanic victims respectively. This is raw data however and does not take into account the proportion of each race within the greater population which would be interesting to look at in future analysis of this data. Most victims did not show signs of mental illness at the time of the incident. White had the highest frequency of mental illness present at time of the incident, with most being between ages 20 to 60.