Data Analysis Project

Kaitlyn Walsh

2023-09-18

Introduction:

For my final project, I will be looking at the Police Shootings data set. The two variables I will be looking at are signs of mental illness and age. I want to find out if there is any relationship between the age of victims and signs of mental illness in police shootings.

This data set has 3,960 obervations and 14 variables.

```
library(readxl); library(ggplot2); library(RColorBrewer)
policeshooting <- read_excel("/Users/kaitlynwalsh/Desktop/math130/fatal-police-shootings-data.xlsx")
```

```
str(policeshooting)
```

```
## tibble [3,960 x 14] (S3: tbl_df/tbl/data.frame)
##  $ id                     : num [1:3960] 3 4 5 8 9 11 13 15 16 17 ...
##  $ name                   : chr [1:3960] "Tim Elliot" "Lewis Lee Lembke" "John Paul Quintero" "Matth
##  $ date                   : POSIXct[1:3960], format: "2015-01-02" "2015-01-02" ...
##  $ manner_of_death        : chr [1:3960] "shot" "shot" "shot and Tasered" "shot" ...
##  $ armed                  : chr [1:3960] "gun" "gun" "unarmed" "toy weapon" ...
##  $ age                    : num [1:3960] 53 47 23 32 39 18 22 35 34 47 ...
##  $ gender                 : chr [1:3960] "M" "M" "M" "M" ...
##  $ race                   : chr [1:3960] "A" "W" "H" "W" ...
##  $ city                   : chr [1:3960] "Shelton" "Aloha" "Wichita" "San Francisco" ...
##  $ state                  : chr [1:3960] "WA" "OR" "KS" "CA" ...
##  $ signs_of_mental_illness: logi [1:3960] TRUE FALSE FALSE TRUE FALSE FALSE ...
##  $ threat_level           : chr [1:3960] "attack" "attack" "other" "attack" ...
##  $ flee                   : chr [1:3960] "Not fleeing" "Not fleeing" "Not fleeing" "Not fleeing" ...
##  $ body_camera            : logi [1:3960] FALSE FALSE FALSE FALSE FALSE FALSE ...
```

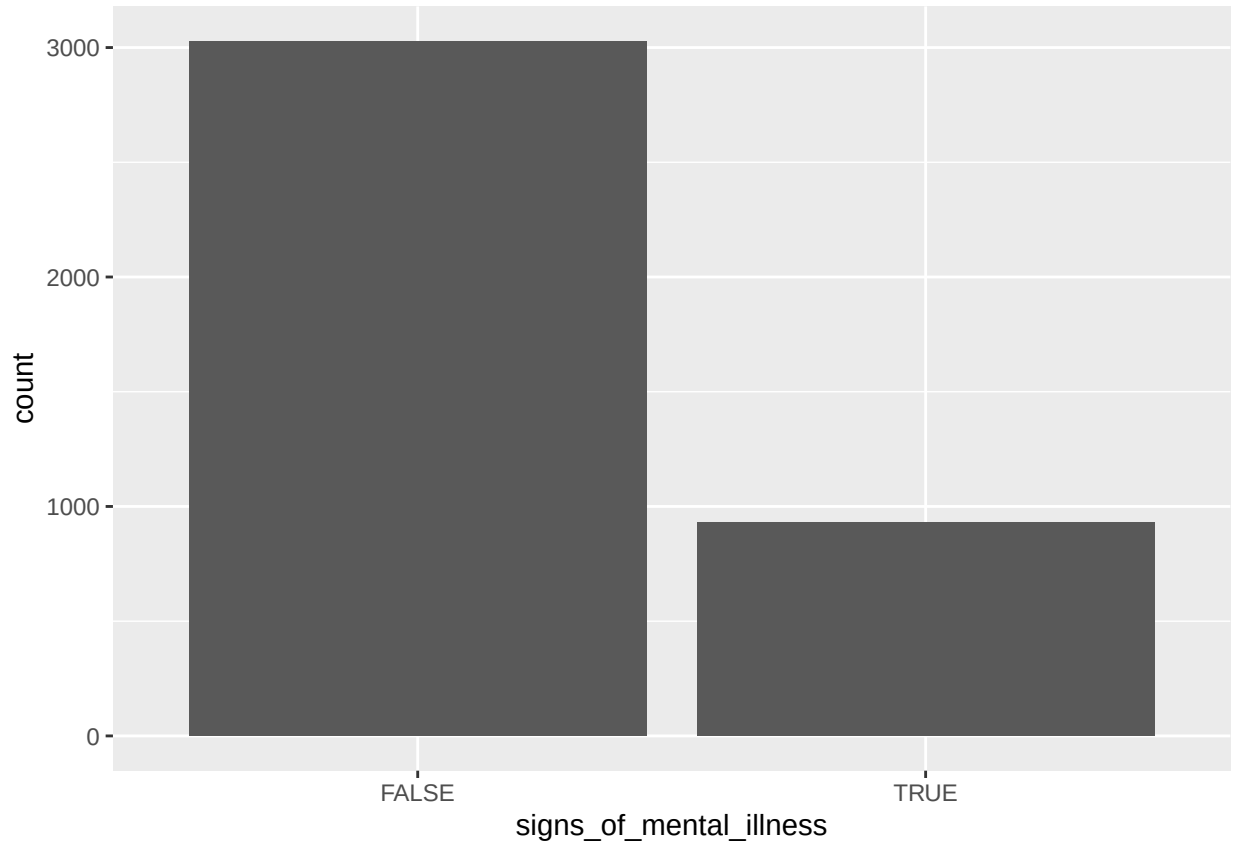
Univariate Exploration:

Race

```
table(policeshooting$signs_of_mental_illness, useNA="always")
```

```
##
## FALSE  TRUE  <NA>
##  3028   932     0
```

```
ggplot(policeshooting, aes(x=signs_of_mental_illness)) +geom_bar()
```



From this graph you can see that there is rarely any victims with signs of mental illness.

Age

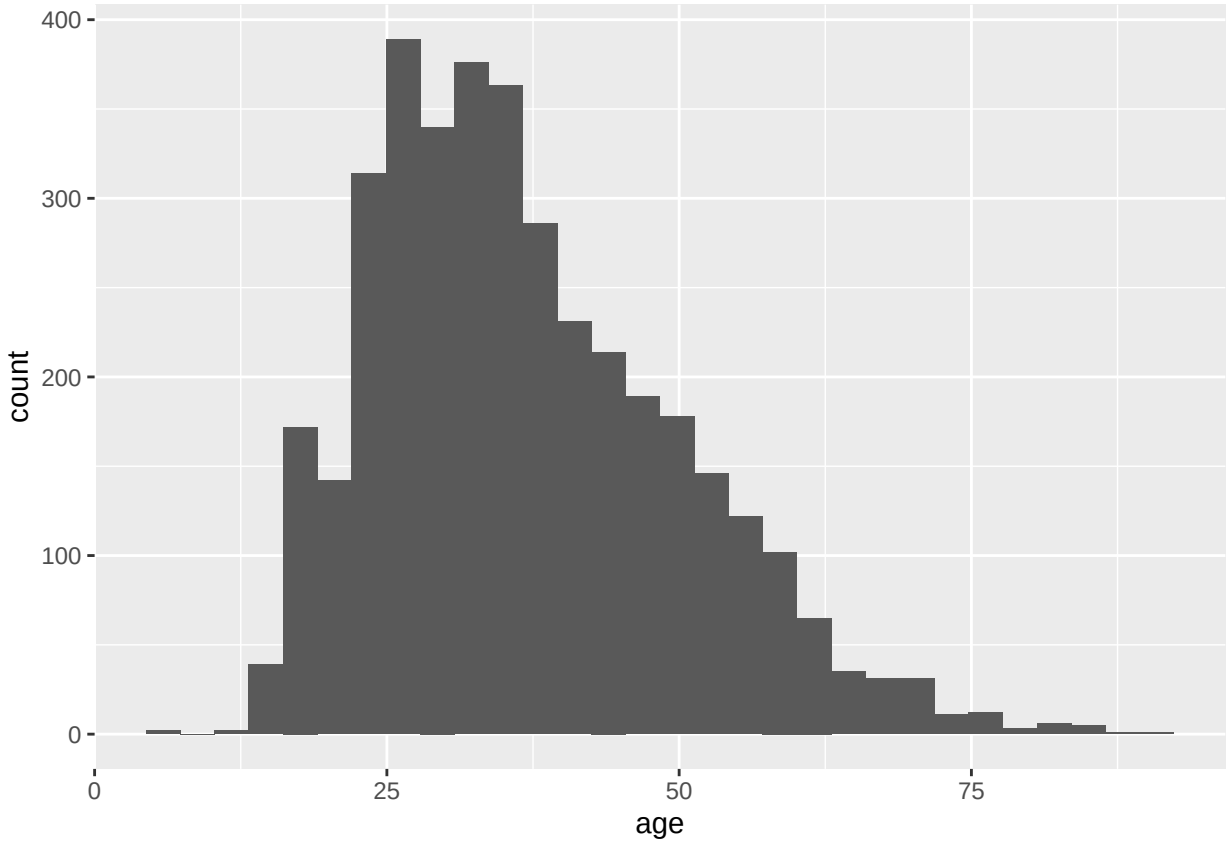
```
summary(policeshooting$age, useNA="always")
```

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

```
ggplot(policeshooting, aes(x=age)) + geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

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



By looking at both the summary table and the graph, we can see that the oldest police shooting victim was 91 years old and the youngest was 6 years old. The mean was about 36 years old and the median was about 35 years old.

Something we also notice about the graph is that its skewd to the right, which signifies that most of the victims are younger in age.

Bivariate Exploration:

```
table(policeshooting$signs_of_mental_illness, policeshooting$age)
```

```
##
##           6  12  13  14  15  16  17  18  19  20  21  22  23  24  25  26  27  28
##   FALSE   2   1   1   2  10  19  26  70  51  52  61  68  84 100 120  91 104  83
##   TRUE    0   0   0   0   2   6   8   7  10  14  15  16  21  25  30  22  22  24
##
##          29  30  31  32  33  34  35  36  37  38  39  40  41  42  43  44  45  46
##   FALSE  90  84  88  94 105  95  85  99  83  73  62  63  67  50  49  50  66  51
##   TRUE   33  26  36  28  25  24  29  31  23  19  26  19  17  15  15  16  18  17
##
##          47  48  49  50  51  52  53  54  55  56  57  58  59  60  61  62  63  64
##   FALSE  43  51  40  43  36  33  40  33  34  26  25  19  34  16  16  16  10  10
##   TRUE   13  14  16  25  18  10  15  15  13  18   6  12  13   8   5   7  11   8
##
##          65  66  67  68  69  70  71  72  73  74  75  76  77  78  79  80  81  82
##   FALSE  13   8   6   8   4  11   7   2   2   2   1   6   1   1   0   1   2   1
##   TRUE    4   3   4   2   5   1   3   4   1   0   1   1   2   0   1   0   0   1
```

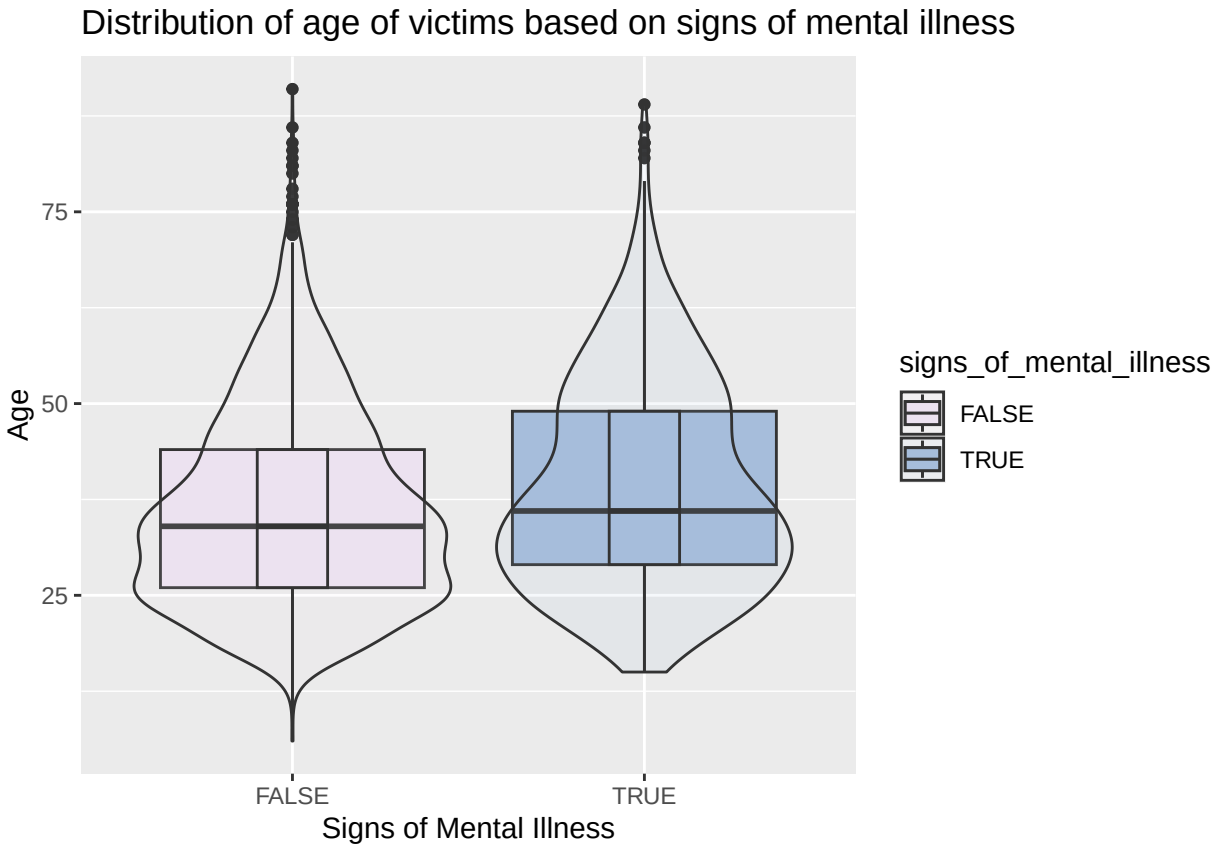
```
##
##          83  84  86  89  91
##   FALSE   1   1   1   0   1
##   TRUE    1   2   1   1   0
```

```
ggplot(policeshooting, aes(x=signs_of_mental_illness, y=age, fill=signs_of_mental_illness)) + geom_boxpl
  geom_violin(alpha=.1) +
  geom_boxplot(alpha=.5, width=.2) +
  ggtitle("Distribution of age of victims based on signs of mental illness") +
  ylab("Age") + xlab("Signs of Mental Illness") + scale_fill_brewer(palette="PuBuGn")
```

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

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

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



Conclusion

We can conclude that the victims of police shootings who showed signs of mental illness tended to be other than those who didn't show signs.