

MATH130- Exploratory Data Project

Julianna Messina

2024-03-01

Introduction

The fatal police shooting data set explored the characteristics of victims of police shootings in 2015. There are 3960 observations with 14 variables.

I am going to explore the correlation between a victim showing signs of mental illness and threat level they posed. The two variables that I will be looking at are `signs_of_mental_illness` and `threat_level`.

My research question is: Is there a higher threat level in people with signs of mental illness or in people without signs of mental illness.

I first imported the data set through my library.

```
library(ggplot2); library(dplyr)

##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
knitr::opts_chunk$set(warning=FALSE, message=FALSE)

library(readxl)
police <- read_excel(path = "/cloud/project/police.xlsx")
```

Univariate

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

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

This table shows that less individuals showed signs of mental illness.

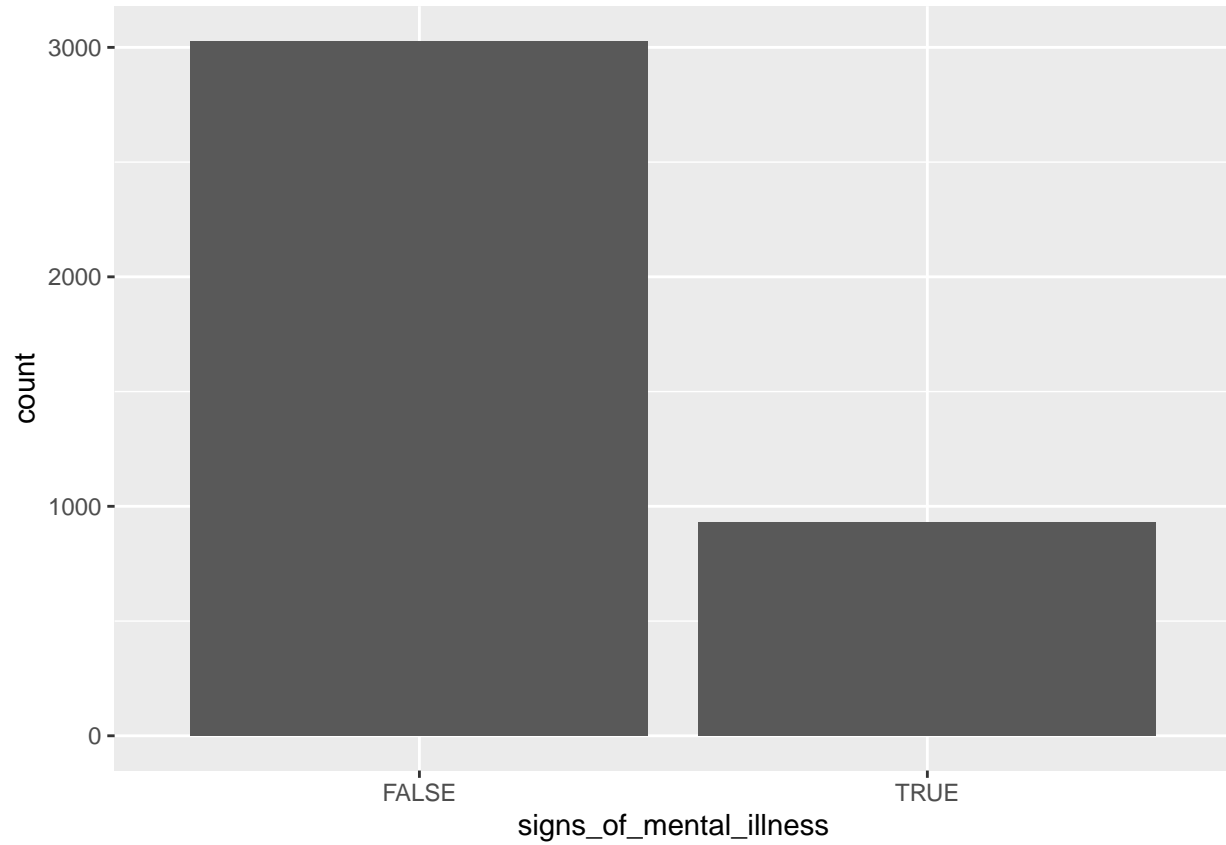
```
table(police$threat_level)
```

```
##
##      attack      other undetermined
##      2497      1255      208
```

This table shows that majority of individuals had a threat level of attack.

Mental Illness

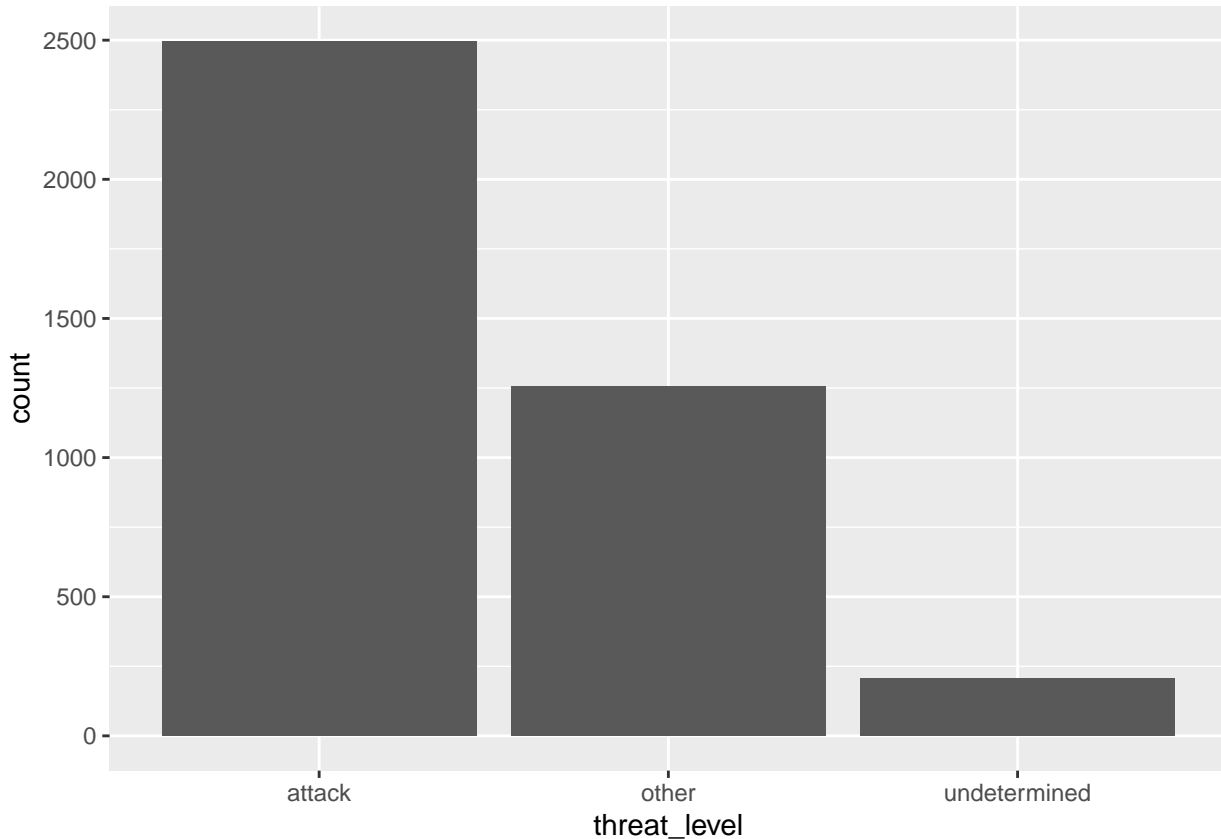
```
ggplot(police, aes(x=signs_of_mental_illness)) + geom_bar()
```



This bar chart shows that majority of individuals showed signs of mental illness.

Threat Level

```
ggplot(police, aes(x=threat_level)) + geom_bar()
```



This bar graph shows that majority individuals showed signs of attack.

Bivariate Exploration

I am going to compare both variables in a table.

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

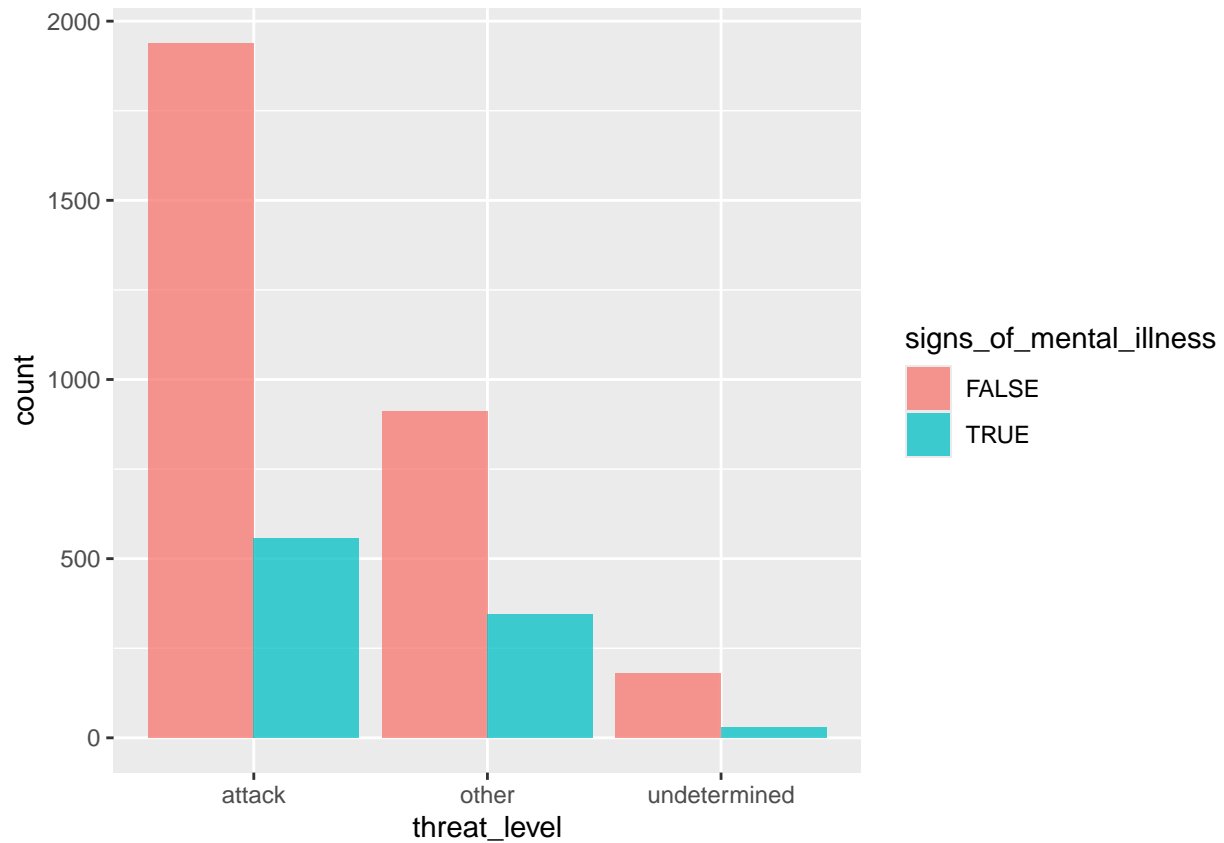
```
##
##      attack other undetermined
## FALSE  1939   910           179
## TRUE   558   345            29
```

```
table(police$signs_of_mental_illness, police$threat_level) %>% prop.table
```

```
##
##      attack      other undetermined
## FALSE 0.489646465 0.229797980 0.045202020
## TRUE  0.140909091 0.087121212 0.007323232
```

According to these statistics, roughly 48% of victims who died from police shootings showed signs of attacking along without mental illness. About 14% of victims who showed signs of attacking did have signs of mental illness. About 23% of victims with mental illness signs did not attack and had a different threat level. About 8.7% of victims with signs of mental illness did not attack. A total of 5.2% of victims threat level was undetermined with 4.5% of those who did not show signs of mental illness.

```
ggplot(police, aes(x=threat_level, fill=signs_of_mental_illness)) + geom_bar(alpha=0.75, position="dodge")
```



This comparison graph show that individuals with signs of mental illness were less likely to attack.

Conclusion

To best show my variables, I had decided to use bar charts since they are categorical. Through my univariate exploration you can clearly see that signs of mental illness were not in majority and individuals with an attack threat level were. To better compare the variables together, I explored them bivariately. By doing this, it clarified my earlier statement to examine if those with an attack threat level showed signs of mental illness. OI was surprised to see that individuals who attacked did not show signs of mental illness. Although this leads me to another question if individuals with mental illness signs were scared rather than wanting to fight.