# Exploratory Data Analysis

Gabriel Loera

2023-09-15

INTRODUCTION

The data set that i will be using for this project will be the Fatal Police Shootings data set. Within this set I will be exploring the relationship of the perpetrators threat level,age , and if body camera footage was recovered after incident. I would like to see which age would most likely result in an altercation with police, my initial hypothesis is that those in their mid to late 20s would be more likely to have interactions. I would also hypothesis that incidence with low threat levels will have less footage recovered than higher threat levels.

UNIVARIATE DISCRIPTION

```
police %>% ggplot(aes(x=threat_level)) + geom_bar(aes(fill = threat_level)) +
 xlab("Threat Level") +
 ylab("Count") +
 ggtitle("Threat Level Displayed by Perpetrator") +
 scale_x_discrete(labels= c("Attack","Other","Undetermined")) +scale_fill_manual(values=c("red","yellow
 theme(legend.position = "none")
```

Threat Level Displayed by Perpetrator
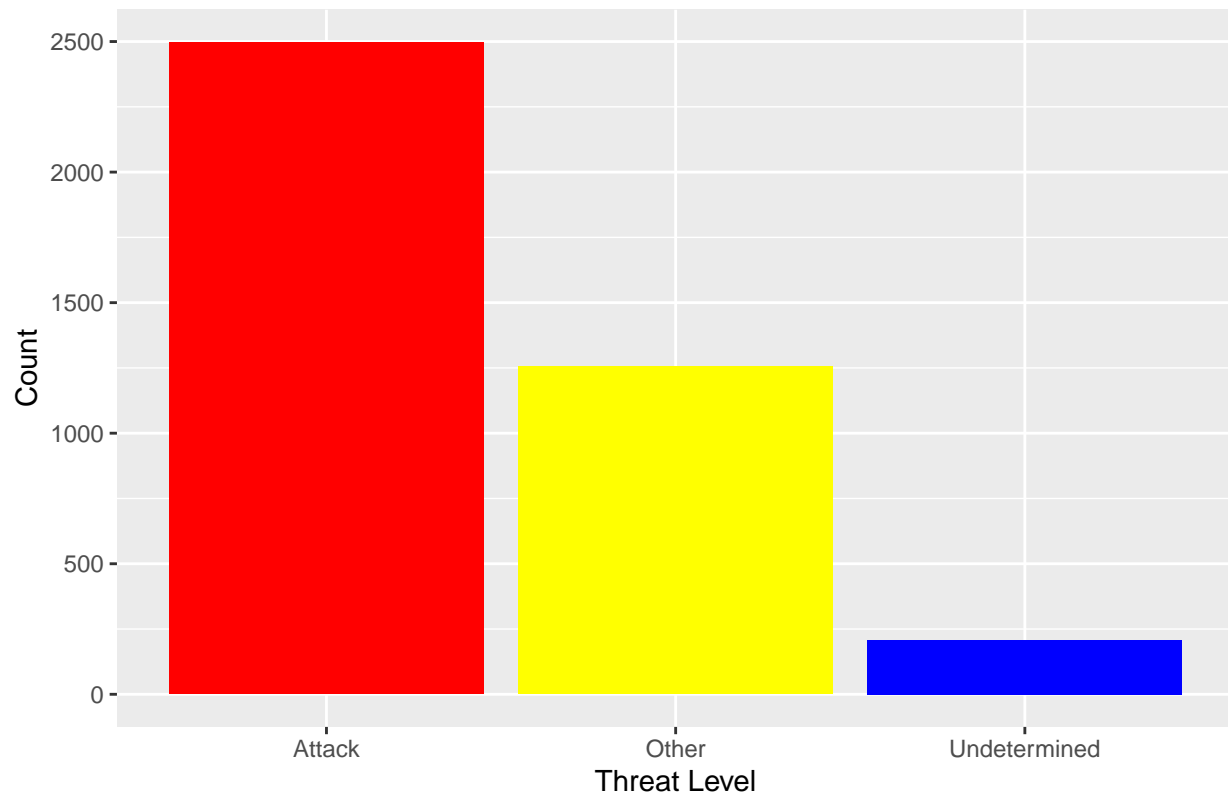
##First we will look at the threat level each perpetrator displayed before lethal force. The levels of threat are a bit vague where only one variable is specific to the threat while the "other" and "undetermined" can be open to speculation.

```
police%>%ggplot(aes(x=age))+
  geom_histogram(colour="red",fill="red")+
 xlab("Age") +
 ylab("Count")+ggtitle("Age at Time of Incident")
```

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

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

Age at Time of Incident

##Within the age variable we will observe which ages are involved most with deadly police shootings.

```
police %>% ggplot(aes(x=body_camera)) + geom_bar(aes(fill = body_camera)) +
 xlab("Body camera status") +
 ylab("Count") +
 ggtitle("Body Camera Footage Recovered") +
 scale_x_discrete(labels= c("Not Recorded", "Recorded")) +scale_fill_manual(values=c("red","blue"))+
 theme(legend.position = "none")
```

Body Camera Footage Recovered
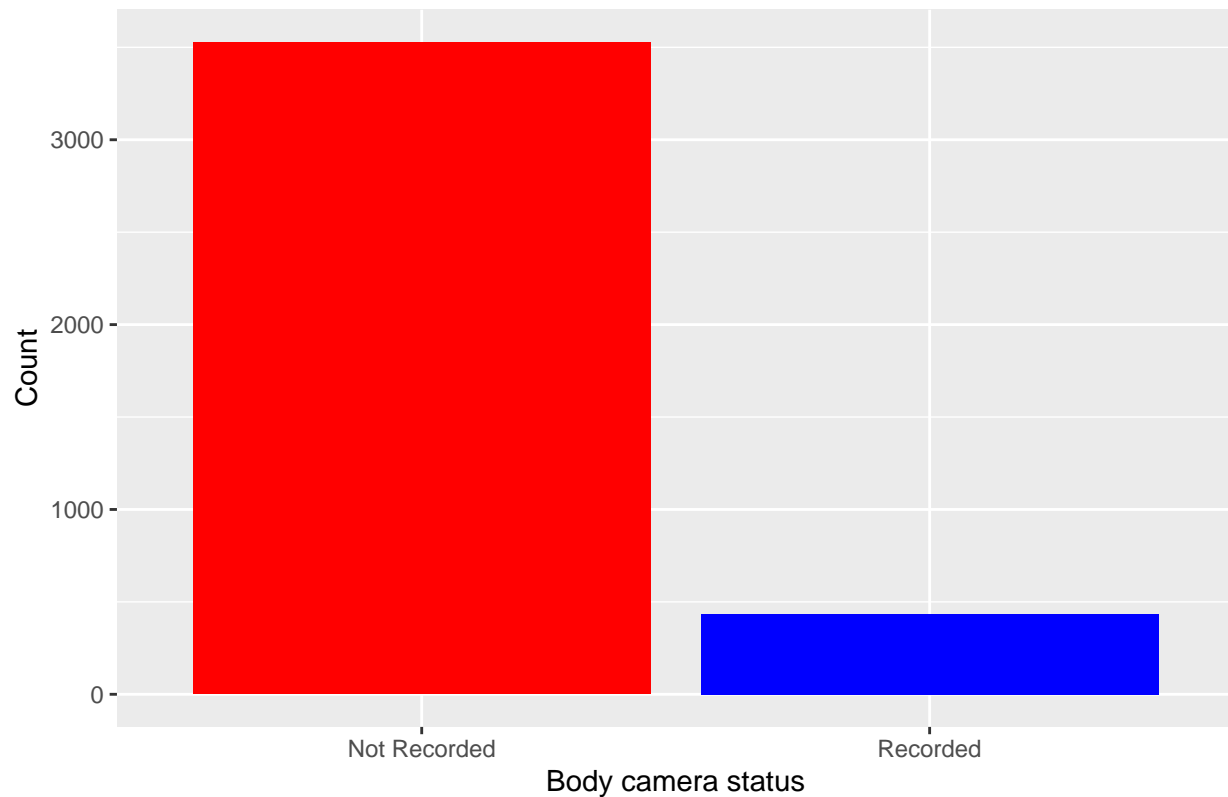
##Because the variable of body_camera does not determine whether a body camera was available at the time of incident we will assume all situations had a camera And that the footage was either recorded or not recorded.

SUMMARY STATS

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

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

```
2497/3960
```

```
## [1] 0.6305556
```

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

```
## [1] 36.849
```

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

```
## [1] 35
```

```
summary(police$body_camera)
```

```
##    Mode   FALSE    TRUE
## logical   3527     433
```

```
433/3960
```

```
## [1] 0.1093434
```

###Of the 3960 observations 63% of the use of lethal force were warranted under direct attack while the other 47% may or may not have been depending on the unrecorded situations. the average age of a person to be involved in these incidents were about 37 years old with a median age of 35. The age group with the most incidents were those between 25 and 37. The majority of body camera footage was not recorded with only 10% of incident being recorded.

BIVARIATE DISCRIPTION

```
police%>%ggplot(aes(x=age)) + geom_bar(aes(fill = threat_level)) +
 xlab("Age") +
 ylab("Count") +
 ggtitle("Threat Level at Age")+scale_fill_manual(values=c("red","green","blue"),(name="Threat Level"))
```

```
## Warning: Removed 152 rows containing non-finite values ('stat_count()').
```
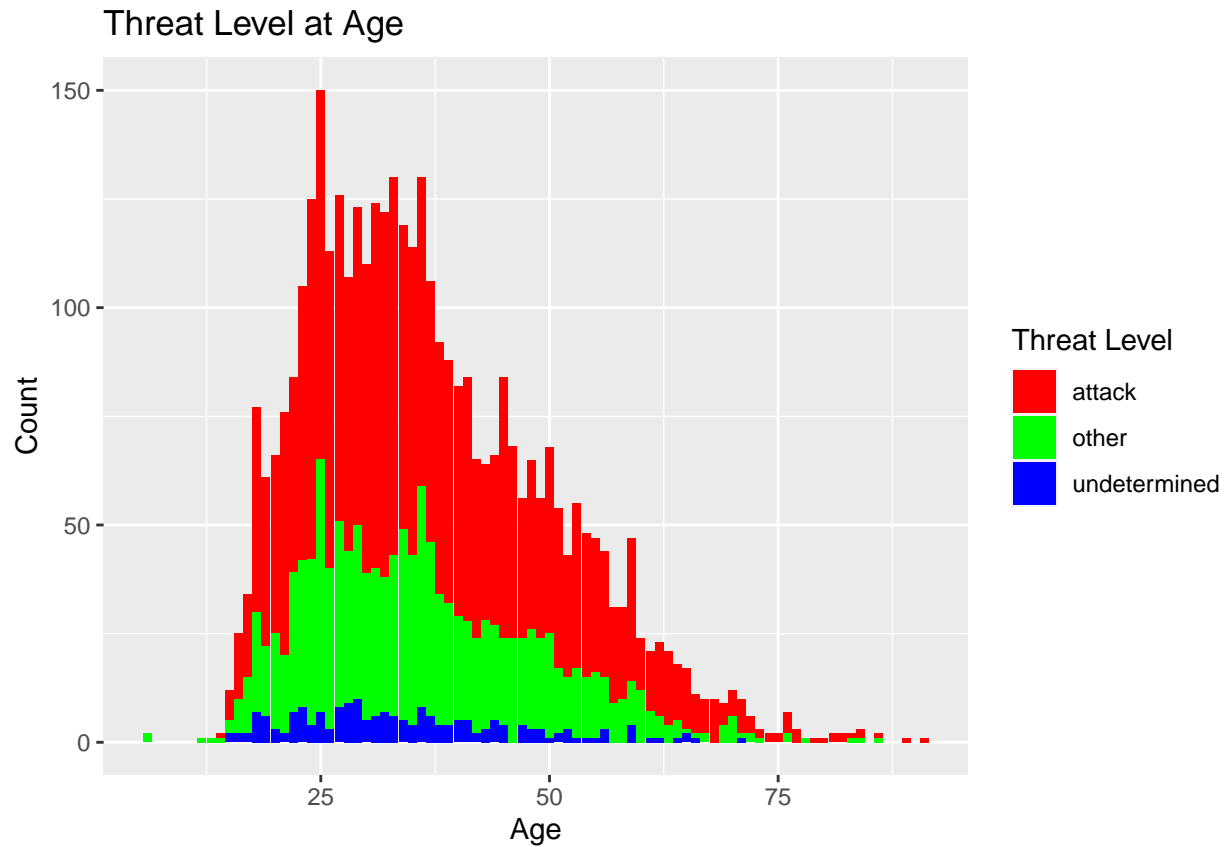
Threat Level at Age

##In this graph there is a comparison between threat level and age and we see ther is a higher threat level between the ages of twenty five and thirty five much like was hypothesised in the introduction. There is also and increase in threat level "other" and somewhat of in increase in threat level "undetermined".

```
ggplot(police, aes(x=body_camera, fill=threat_level)) + geom_bar()+xlab("Body camera") +
 ylab("Count")  +ggtitle("Body Camera Footage Evidence of Threat Level")+
scale_fill_manual(values=c("red", "green","blue"),name="Threat Level")
```

Body Camera Footage Evidence of Threat Level
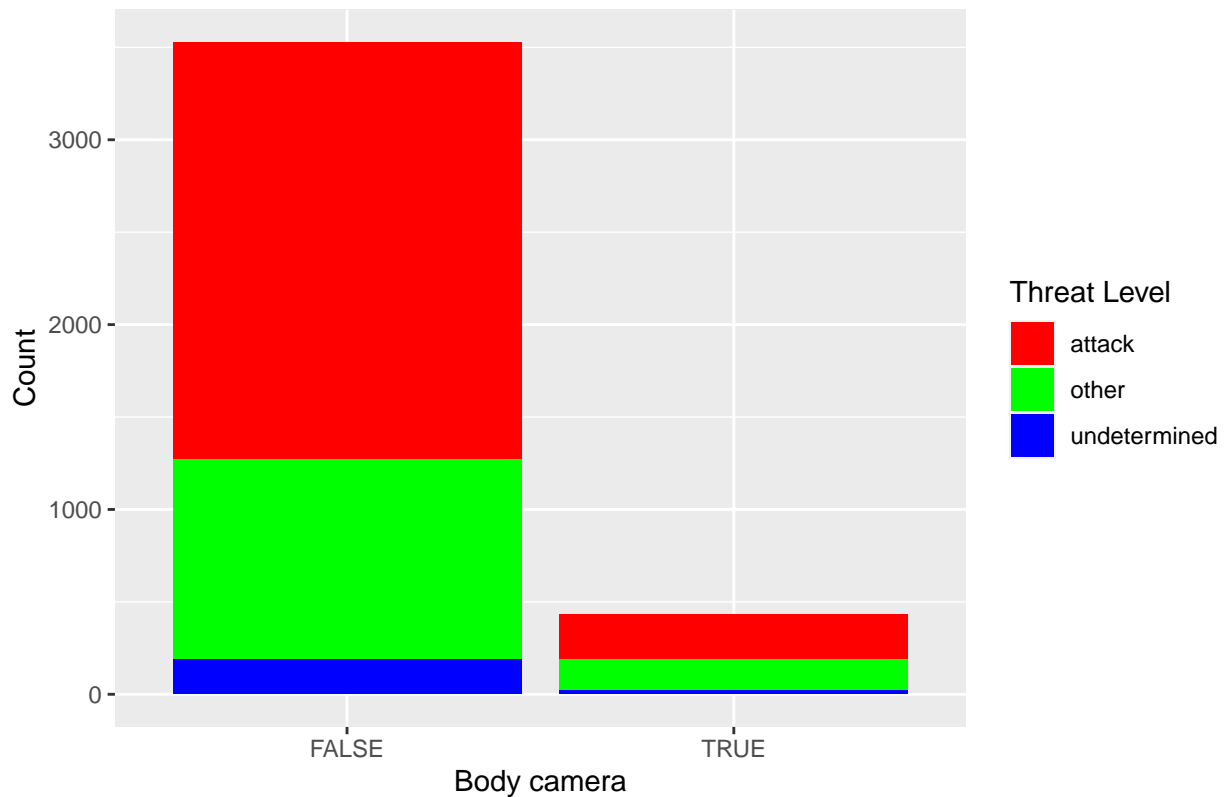
##When comparing threat level with body camera we see that the majority of incidence were not recorded with a body camera. In such stressful sittuations it is possible an officer may forget to turn on his camera, however if such a tool is available it should be used in every cas possible and should be turned on before every interaction with the public both for officer and civilian safety.

CONCLUSION

#With all the data observed we can say that interactions with people ages 25 through 37 there is a higher incident rate that those in other age groups especially with an average age of incident at 37. Also the incidents that do occur wil not have video evidence which puts and extra enphasis on collecting witness and officer reports. To rule out illegal use of force video evidence is a highly valuble variable and the rate at which it is recorded or the rate the cameras are being used is not enough, especially withthe number of deaths as a result of threat level "other". Although we cannot detemine if video evidence has been tampered with we can see a trend in the use of body cameras or possibly even the recovery of body cam footage. Either way it seems that more officers need to use their body cameras at a higher rate and new policies may be required to increas use of body cameras.