EDA_emanthony

2024-03-01

Introduction

```
library(readxl)
shootings <- read_excel("C:/Users/ellaa/OneDrive/Desktop/math130/data/fatal-police-shootings-data.xlsx"
```

I will be exploring the fatal police shootings data set which has 3960 observations and 14 different variables pulled from the Washington Post. I will be looking at the manner of death, race, flee variables. My research question is the following: "Is there a correlation between the race and whether of not the victims tried to flee and the manner of death?"

Univariate Exploration

I am going to start by providing a summary of each of the variables being explored.

```
library(ggplot2)
ggplot(shootings, aes(x=manner_of_death, fill=manner_of_death)) +
geom_bar() +
ggtitle("Distribution of Manner of Death") +
xlab("Manner of Death") +
ylab("Count") +
scale_fill_manual(values=c("lightblue","lightblue"),guide="none")
```

Distribution of Manner of Death
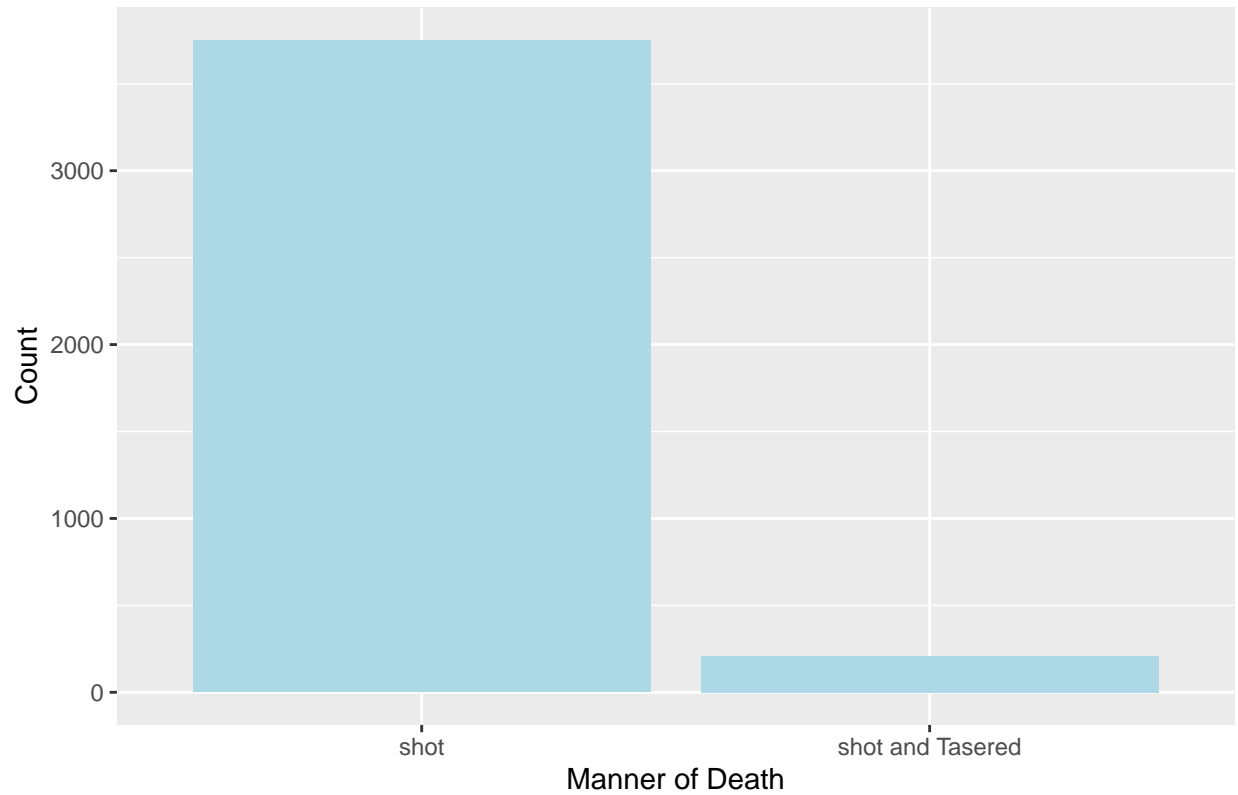


```
table(shootings$manner_of_death)
```

```
##
##          shot shot and Tasered
##          3750              210
```

The variable manner of death is referring to how each victim was killed; either shot or shot and tasered. In this data set the majority of the victims were shot.

```
ggplot(shootings, aes( x=race, fill=race)) +
geom_bar() +
ggtitle("Distribution of Race") +
xlab("Race") +
ylab("Count") +
scale_fill_manual(values=c("lightblue","lightblue","lightblue","lightblue","lightblue","lightblue","ligh
```
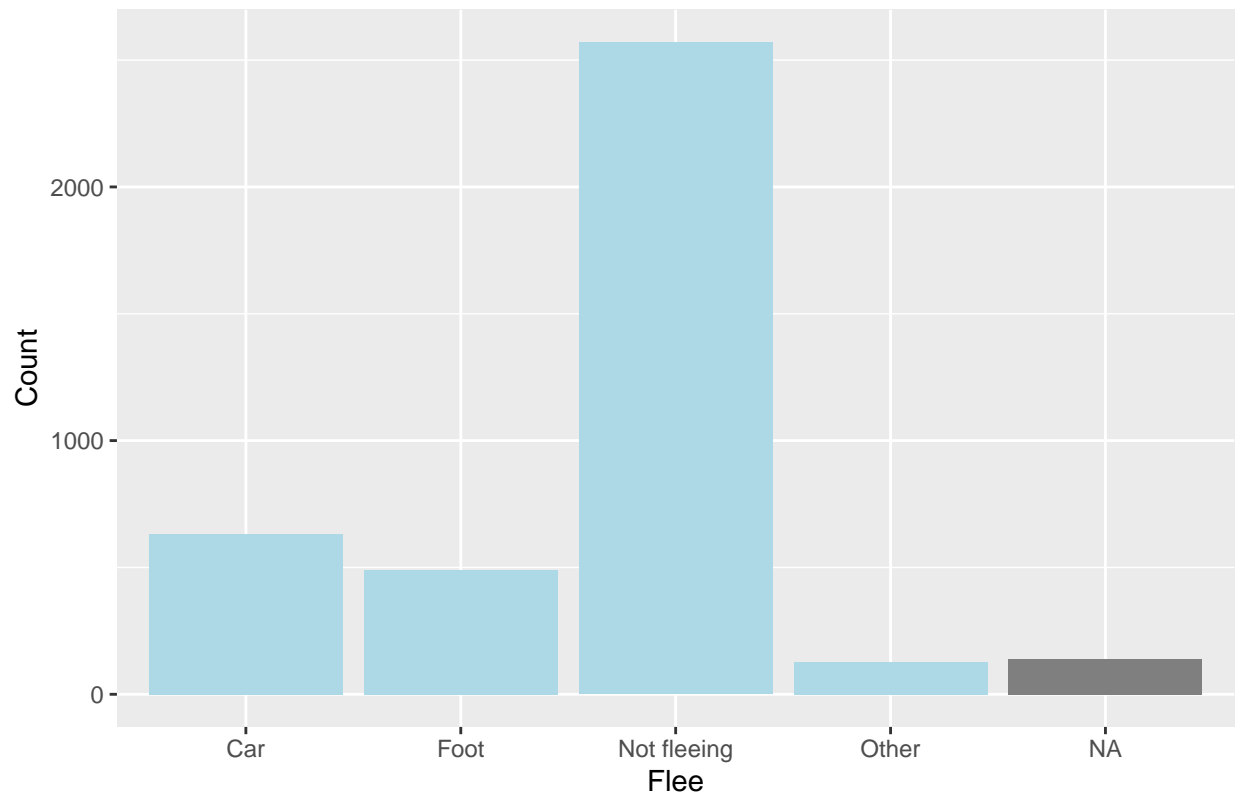
Distribution of Race
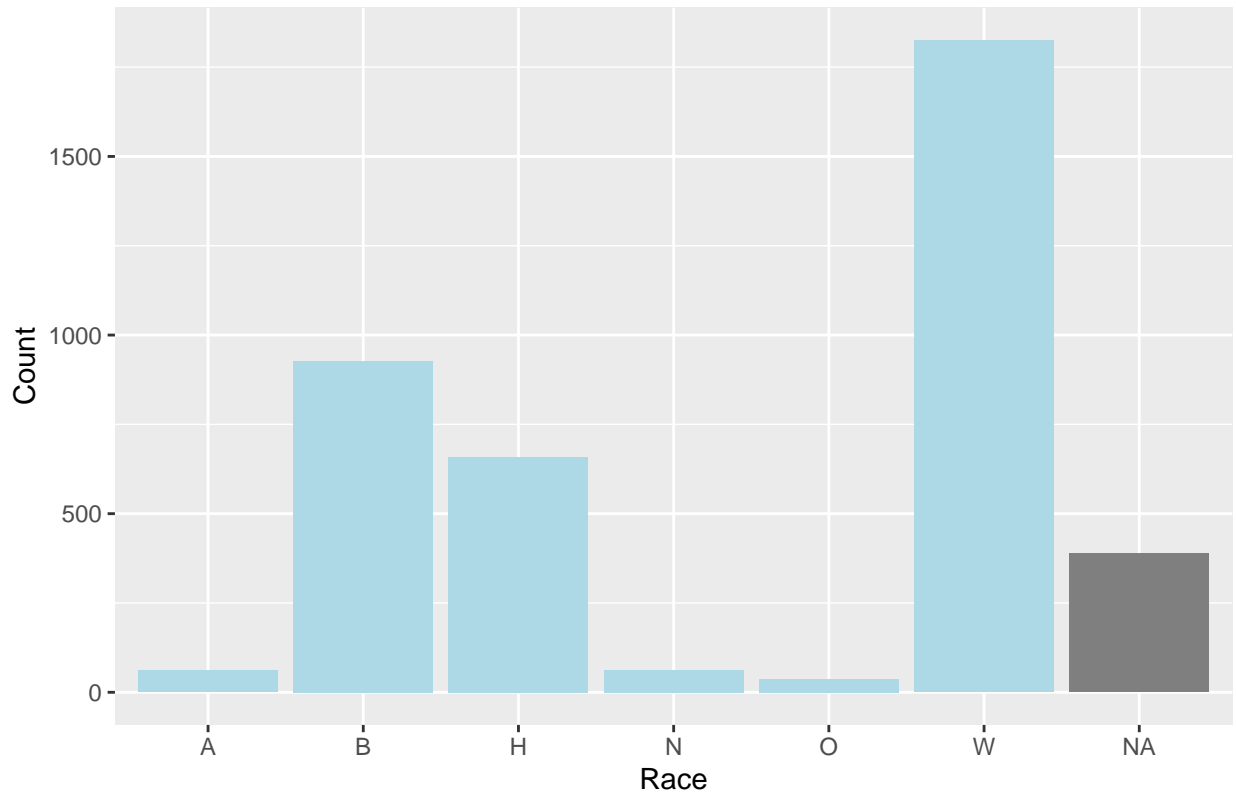

```
table(shootings$race)
```

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

The race distribution refers to the victim's races.The distribution also includes people's races that were not identified.

```
ggplot(shootings, aes(x=flee, fill=flee)) +
geom_bar() +
ggtitle("Distribution of Flee") +
xlab("Flee") +
ylab("Count") +
scale_fill_manual(values=c("lightblue","lightblue","lightblue","lightblue","lightblue"), guide="none")
```

Distribution of Flee


```r
table(shootings$flee)
```

```
##
##         Car        Foot Not fleeing       Other
##         631         491        2570         128
```

The flee distribution refers to how the victims fled. This also includes victims that were not identified to have feld.

Bivariate Exploration

Now I am going to provide a comparison of the variables manner of death and race, along with manner of death and flee.

```r
ggplot(shootings, aes(x=manner_of_death, fill=race )) +
geom_bar(position="dodge") +
ggtitle("Manner of Death vs. Race") +
xlab("Manner of Death") +
ylab("Count")
```

Manner of Death vs. Race
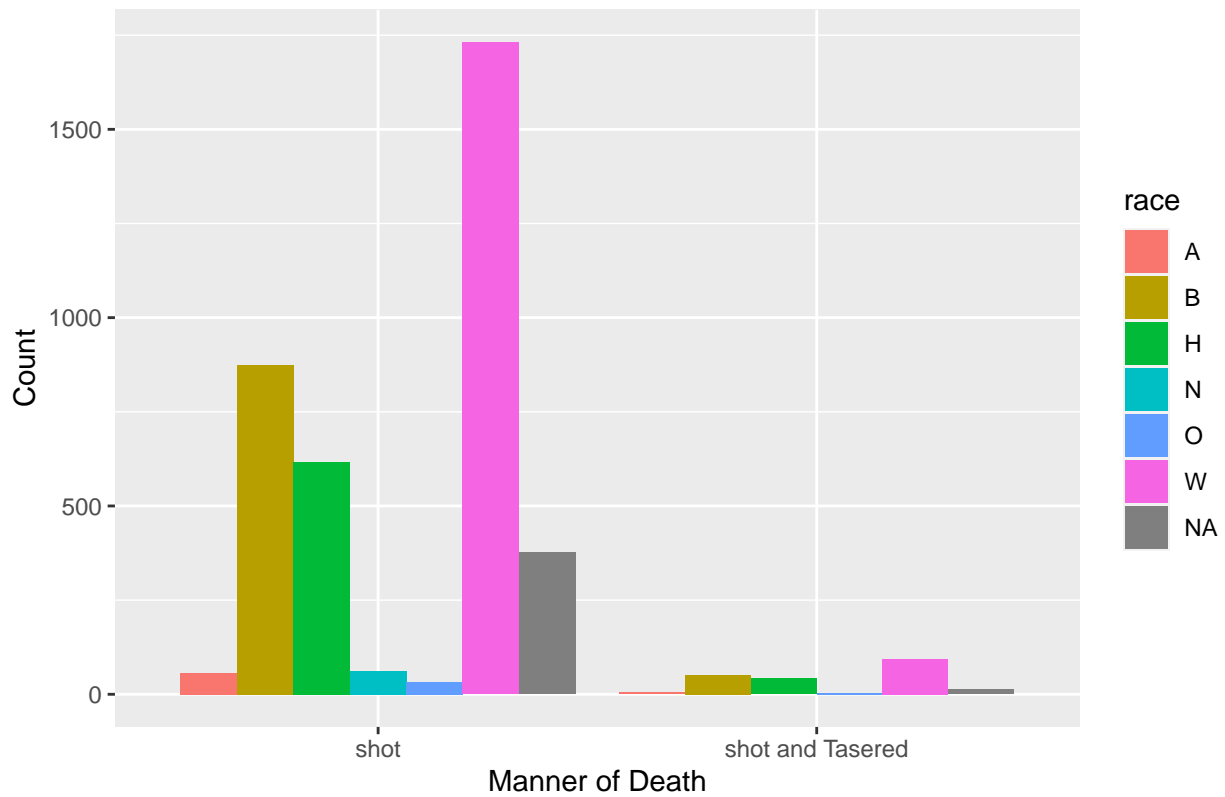

```r
table(shootings$manner_of_death, shootings$race)
```
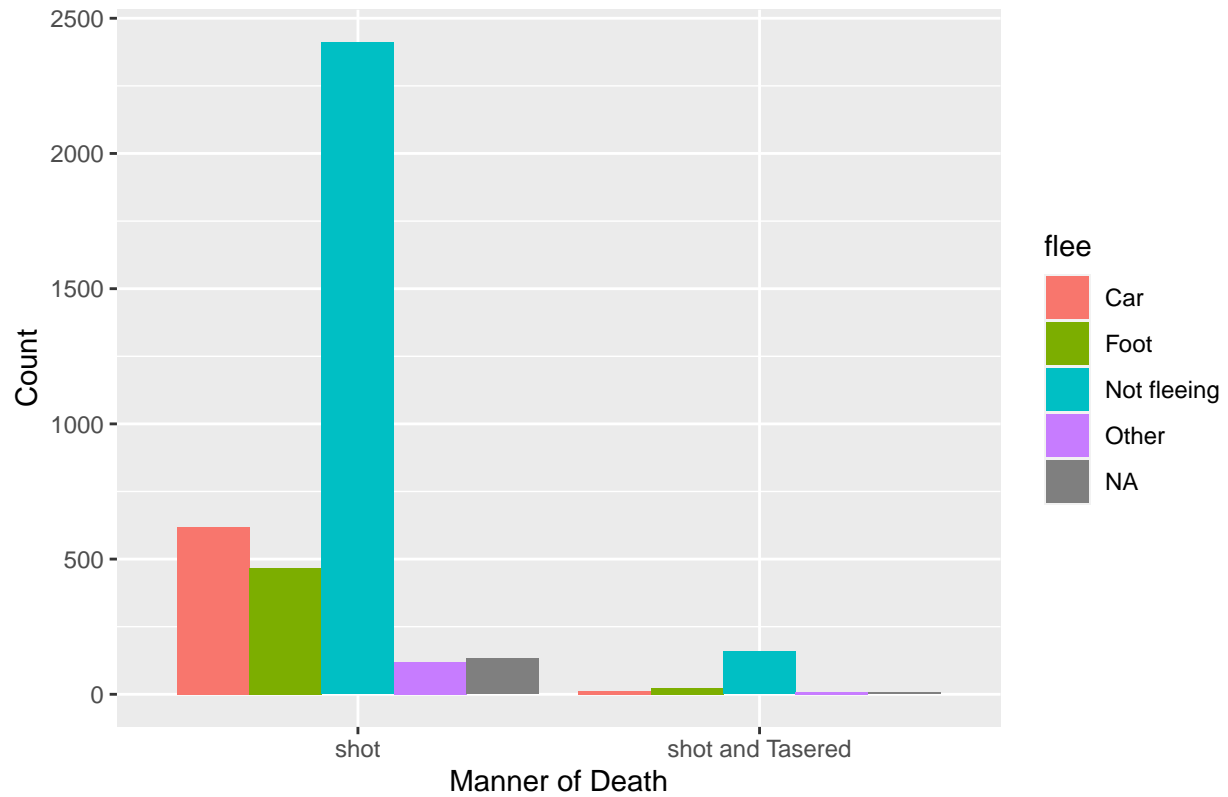
```
## 
##                     A    B    H    N    O    W
##   shot             56  875  617   62   33 1731
##   shot and Tasered  5   52   42    0    4   94
```

```r
ggplot(shootings, aes(x=manner_of_death, fill=flee )) +
geom_bar(position="dodge") +
ggtitle("Manner of Death vs. Flee") +
xlab("Manner of Death") +
ylab("Count")
```

Manner of Death vs. Flee


```r
table(shootings$manner_of_death, shootings$flee)
```

```
## 
##                      Car Foot Not fleeing Other
##    shot              618  468        2411   120
##    shot and Tasered   13   23         159     8
```

Conclusion

To conclude, the distribution comparing race and the manner of death, we can tell that there are more victims for certain races, and the distribution comparing the manner of death and how victims tried to flee, the majority of the victims were shot when they did not flee. There seems to be some level of correlation between both of the distributions.