

Final Project

Faith Whittaker

9/27/2021

Introduction

The data set I am analyzing is the Fatal Police Shootings. A little background on this data set, it is from the “Washington Posts” database. They collected records of every fatal shootings in the U.S. by a police officer on duty since January 1, 2015. It is only including reports of on-duty/in the line of duty officers, where they have killed civilians by police shootings.

```
"shootings"<- read_excel("C:/Users/Fcwhittaker/Desktop/MATH130/Data/fatal-police-shootings-data.xlsx",  
sheet=1,col_names=TRUE)
```

Three Variables I will be Exploring

1. threat levels
2. Race of the Victims
3. Signs of mental illness

```
shootings %>% select(threat_level,signs_of_mental_illness,race)
```

```
## # A tibble: 3,960 x 3  
##   threat_level signs_of_mental_illness race  
##   <chr>         <lgl>                <chr>  
## 1 attack       TRUE                 A  
## 2 attack       FALSE                W  
## 3 other        FALSE                H  
## 4 attack       TRUE                 W  
## 5 attack       FALSE                H  
## 6 attack       FALSE                W  
## 7 attack       FALSE                H  
## 8 attack       FALSE                W  
## 9 other        FALSE                W  
## 10 attack      FALSE                B  
## # ... with 3,950 more rows
```

Univariate Analysis of Variables

Threat Level of Victims

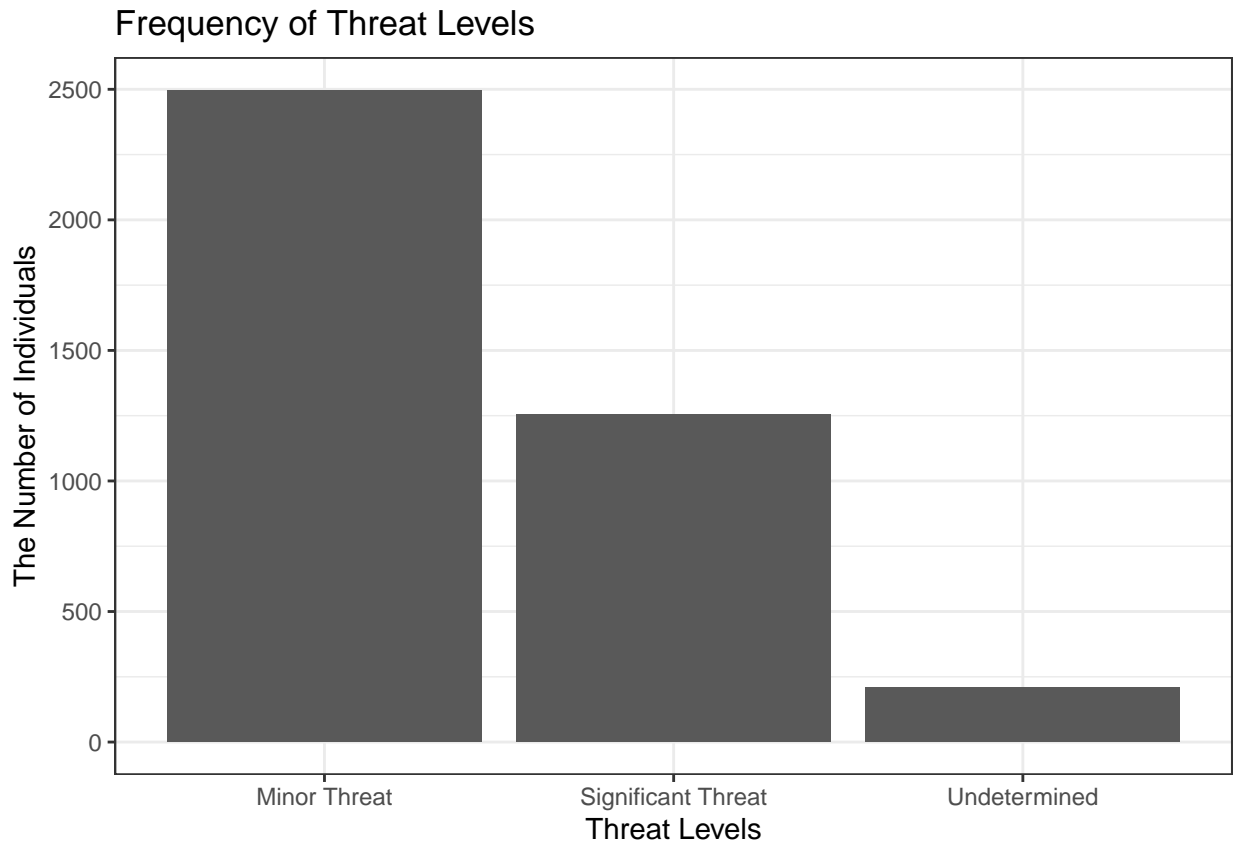
```
shootings$threat_level_new<-fct_recode(shootings$threat_level,"Undetermined"="undetermined",
"Minor Threat"="attack","Significant Threat"="other")

table(shootings$threat_level_new)
```

```
##
##      Minor Threat Significant Threat      Undetermined
##              2497              1255              208
```

I thought it would be better to be more specific in the different threat levels. I felt like the other names weren't that's specific and can be a little confusing. Based on the table, "Minor Threat" signifies if the the suspect or individual did a non life threatening attack on the police officer. The "Significant Threat" is when the suspect or individual did damage or a significant threat to the officers and "undetermined" is when the people filing the report are not sure what exactly took place at the scene or if there was any attacks initially made to the officer. As we can see, there were more minor threats made, 2497, than Significant and Undetermined threats. There is still good amount of "Undetermined" threats, 208, which can be a problem when it comes to making sure we have all the correct data information to properly conduct our analysis.

```
ggplot(shootings,aes(x=threat_level_new))+geom_bar()+theme_bw()+xlab("Threat Levels")+
ylab("The Number of Individuals")+ggtitle("Frequency of Threat Levels")
```



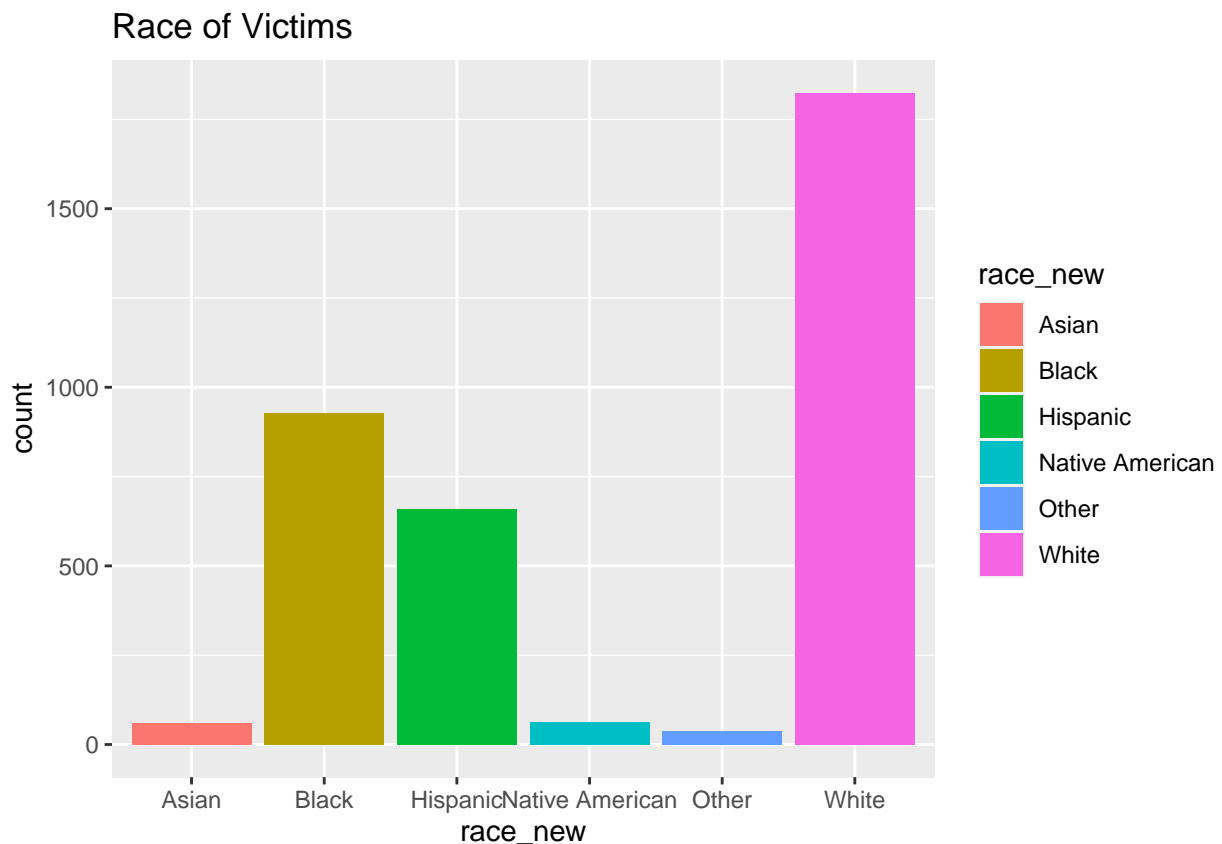
This bar graph just gives further visual information on how there are more Minor threat levels than any other threat levels, which one can say that the victims were never really a huge threat to the officers to begin with. The height of the "Minor Threat" bar is about twice the size of the "Significant Threat" level and triple the size or more of the "undetermined" threat level.

Race

```
shootings$race_new<-fct_recode(shootings$race, "Black"="B", "Native American"="N", "Hispanic"="H", "Other"="O", "White"="W")
table(shootings$race_new)%>%prop.table*100
```

```
##
##      Asian      Black  Hispanic Native American      Other
##  1.708205  25.959115  18.454215  1.736208  1.036124
##      White
##  51.106133
```

```
shootings%>%select(race_new)%>%filter(!is.na(race_new))%>%
ggplot(aes(x=race_new,fill=race_new))+
geom_bar(position="dodge")+ggtitle("Race of Victims")
```



Based on both the table and bar graph, I can see that there are more White, Black, and Hispanic victims than any other race. I made the table into a proportion table where I can see the percentage of each race and there is a clear percentage of more White individuals than the other races. Out of 100% there are 51.1% of white individuals that make up the data records and only 25.95% of Black Individuals that make up the data set and the rest are Asian, Hispanic, and Native American.

Signs of Mental Illness

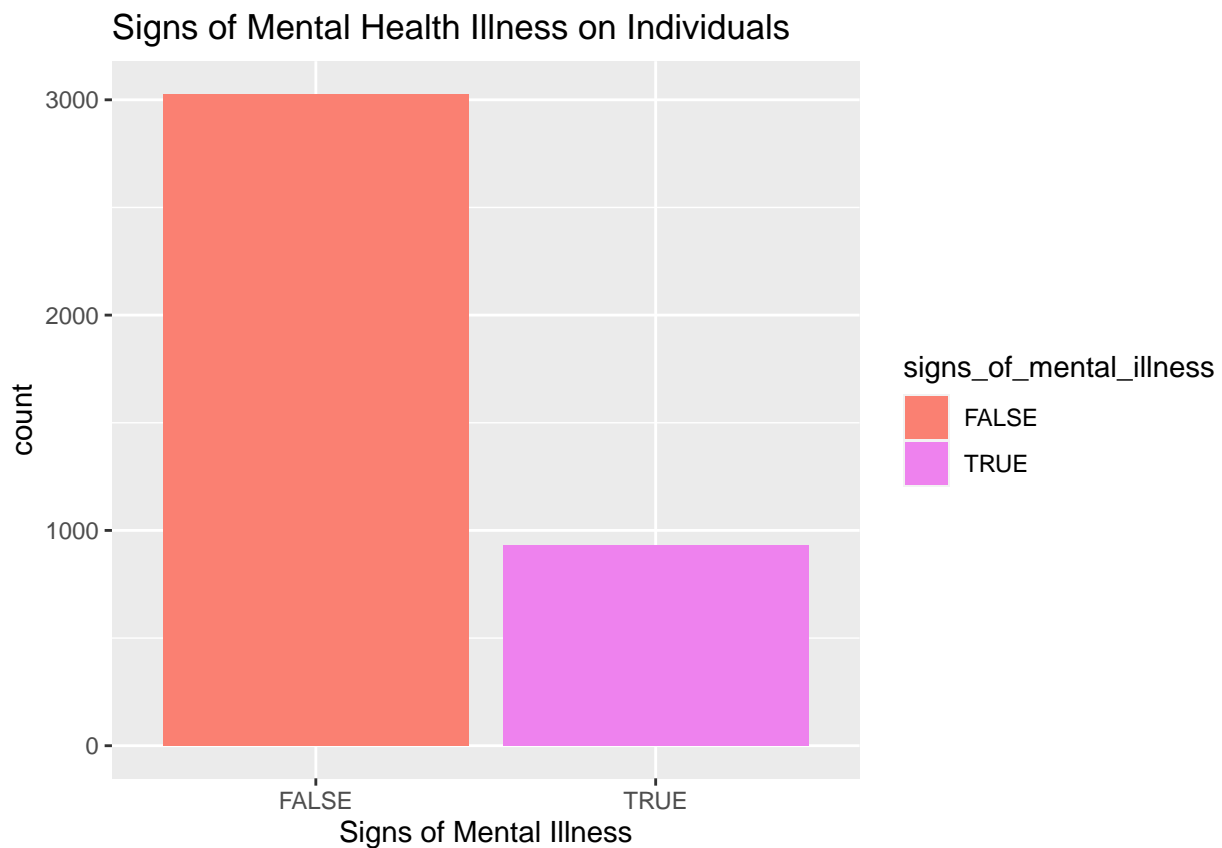
```
mean(shootings$signs_of_mental_illness)
```

```
## [1] 0.2353535
```

```
summary(shootings$signs_of_mental_illness)
```

```
##      Mode   FALSE    TRUE  
## logical  3028    932
```

```
ggplot(shootings,aes(x=signs_of_mental_illness,fill=signs_of_mental_illness))+  
geom_bar()+xlab("Signs of Mental Illness")+  
ggtitle("Signs of Mental Health Illness on Individuals")+  
scale_fill_manual(values=c("salmon","violet"))
```

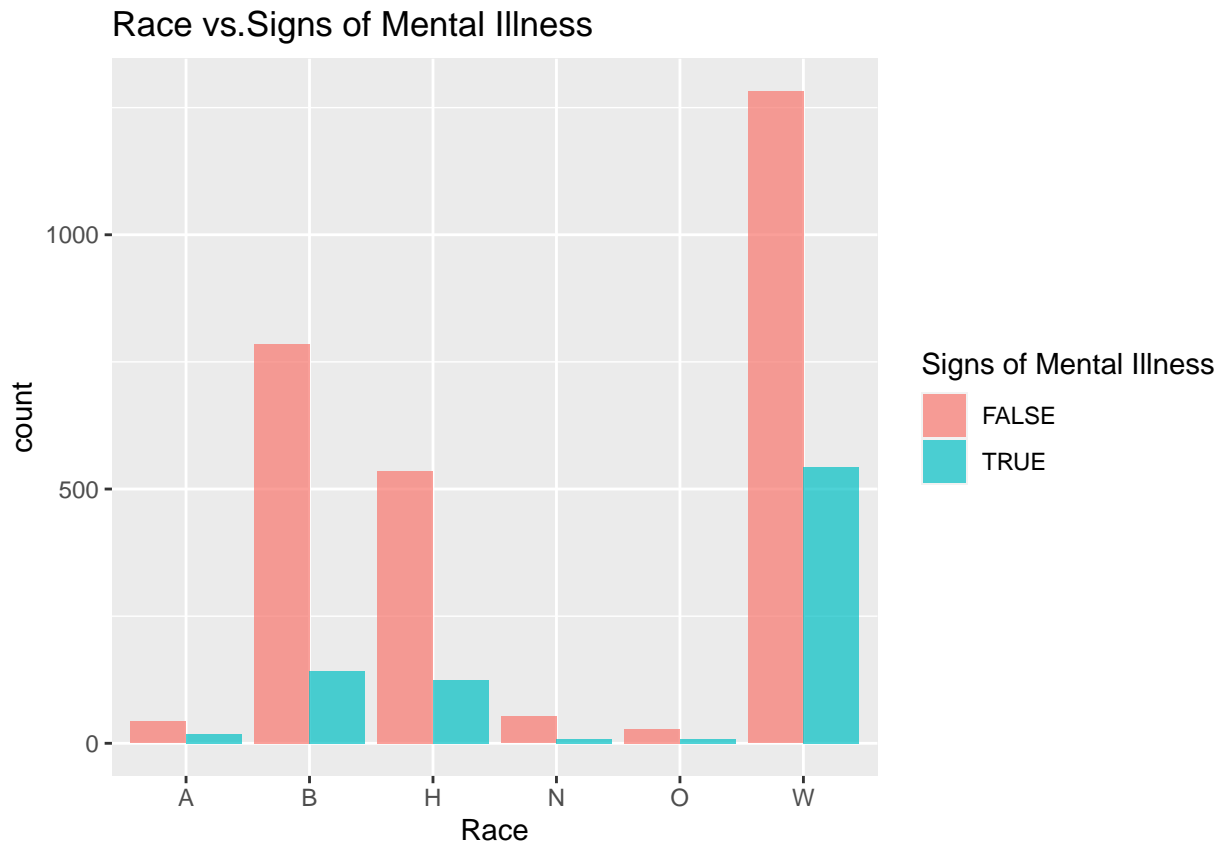


This bar graph shows that there wasn't a significant amount of individuals who were shot, who had any signs of mental illness. There are more "false" of signs of mental illness than "true". There are 3000 people who showed no signs of mental illness, which is 2000 more than the amount of people who did show signs of mental health issues, 1000.

Bivariate Comparisons between two variables

Race vs. Signs of Mental Illness

```
shootings%>%select(race,signs_of_mental_illness)%>%filter(!is.na(race))%>%  
ggplot(aes(x=race,fill=signs_of_mental_illness))+geom_bar(position="dodge",alpha=0.7)+  
scale_fill_discrete(name="Signs of Mental Illness")+  
xlab("Race")+ggtitle("Race vs.Signs of Mental Illness")
```



Based on this bar graph, there is a high count for individuals who were White, Black, and Hispanic that showed no signs of having a mental illness when they were shot. They also have a higher count when it came to showing signs of mental illness, but it is not a high count to out way those who didn't show any signs of having a mental illness. White individuals has a count of over 1000 of victims having o signs of a mental illness, while Black and Hispanic individuals had a count under 1000. Asians, Native Americans, and Other races didn't make or meet the spectrum. For those who did shows signs of a mental illness, the highest was White people with a count of a little over 500, while the other races were barely more than 10 counts.

Threat Level vs. Race

```
table(shootings$race_new,shootings$threat_level_new)%>%prop.table(margin=1)*100%>%round(2)
```

##

##		Minor Threat	Significant Threat	Undetermined
##	Asian	52.459016	44.262295	3.278689
##	Black	65.695793	29.234088	5.070119
##	Hispanic	55.993930	37.481032	6.525038
##	Native American	56.451613	41.935484	1.612903
##	Other	64.864865	35.135135	0.000000
##	White	65.095890	30.684932	4.219178

This table give information on the correlation between Threat level and Race. Based on the table, it looks like there is a close correlation between all of the races when it comes to Minor Threats attacks. All of the races has between 52%-65% minor Threat when it comes to the police officers. Meanwhile, in the "Significant Threat Level", Asians has a 44.3% attack percentage compare to Black people who has 29.2% percentage of being a significant threat towards police officers. And in the "undetermined" category, it has a very low percentage for all races. But what I can draw from this is that all of the races has a chance of being shot at and killed because of the correlation they have with their race and threat level, but White, Black, and Asian individuals has a higher likelihood of this happening.