Exploring Police Shootings

Viren Halaharivi

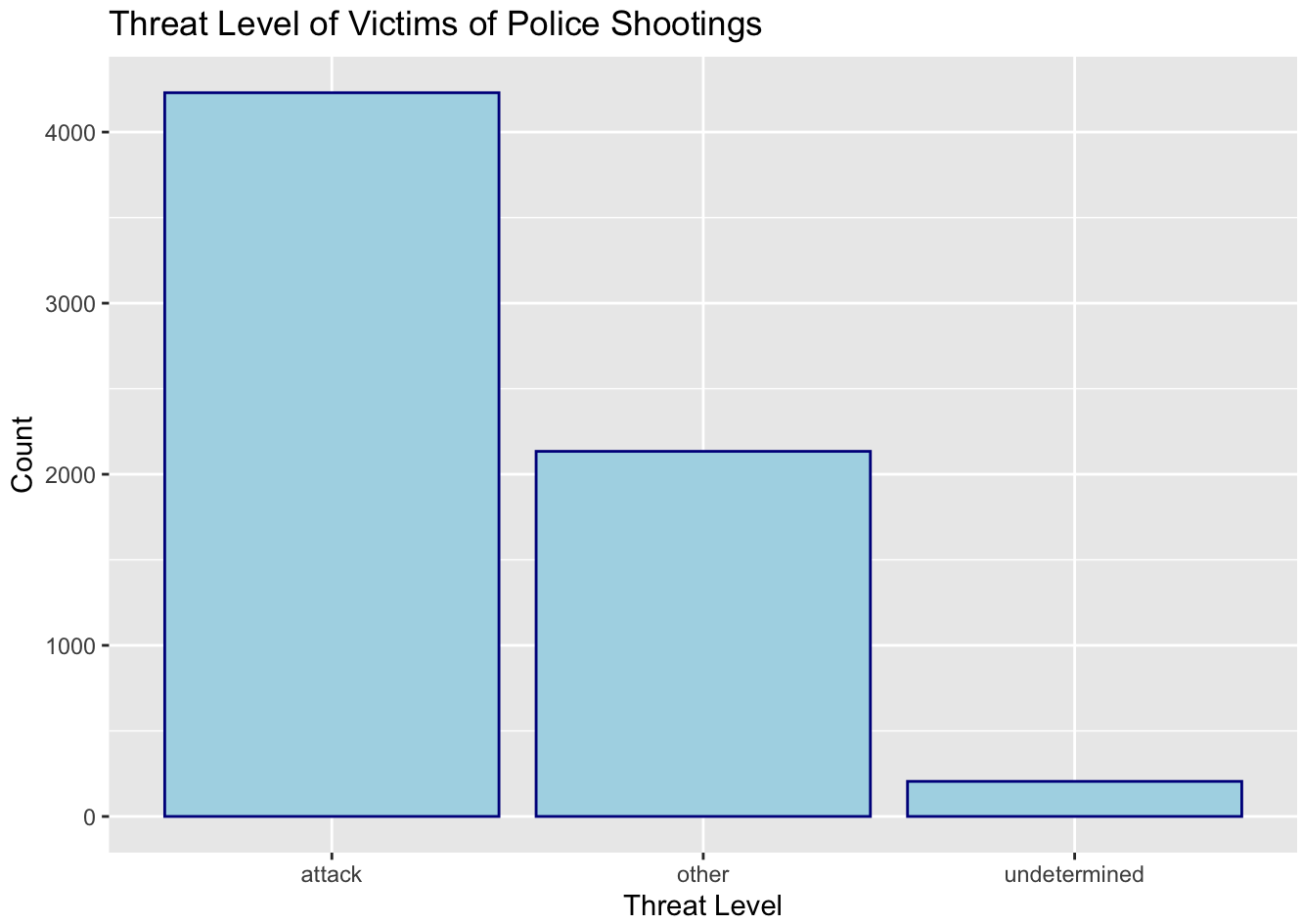
2022-09

Introduction

The dataset I will investigate for this project is called **Project1_police_shootings.csv**. This police shootings dataset contains a plethora of variables which includes the Demographics of victims(Person.name, Person.age, Percent.gender, Person.race), Incident detail(Incident.Date.Month, Incident.Date.Day, Incident.Date.Year, Incident.Date.Full, Incident.Location.City, and Incident.Location.State), Factors of the incident(Factors.Mental.Illness, Factors.Fleeing, Factors.Threat.Level), the result of the incident(Shooting.Manner), and whether or not a body camera was used(Shooting.Body-Camera). The outcome variable is **Factors.Threat.level**, specifically attack level threats, and the predictors I will use to investigate my outcome are **Person.gender**, **Person.race**, and **Incident.Location.State**.

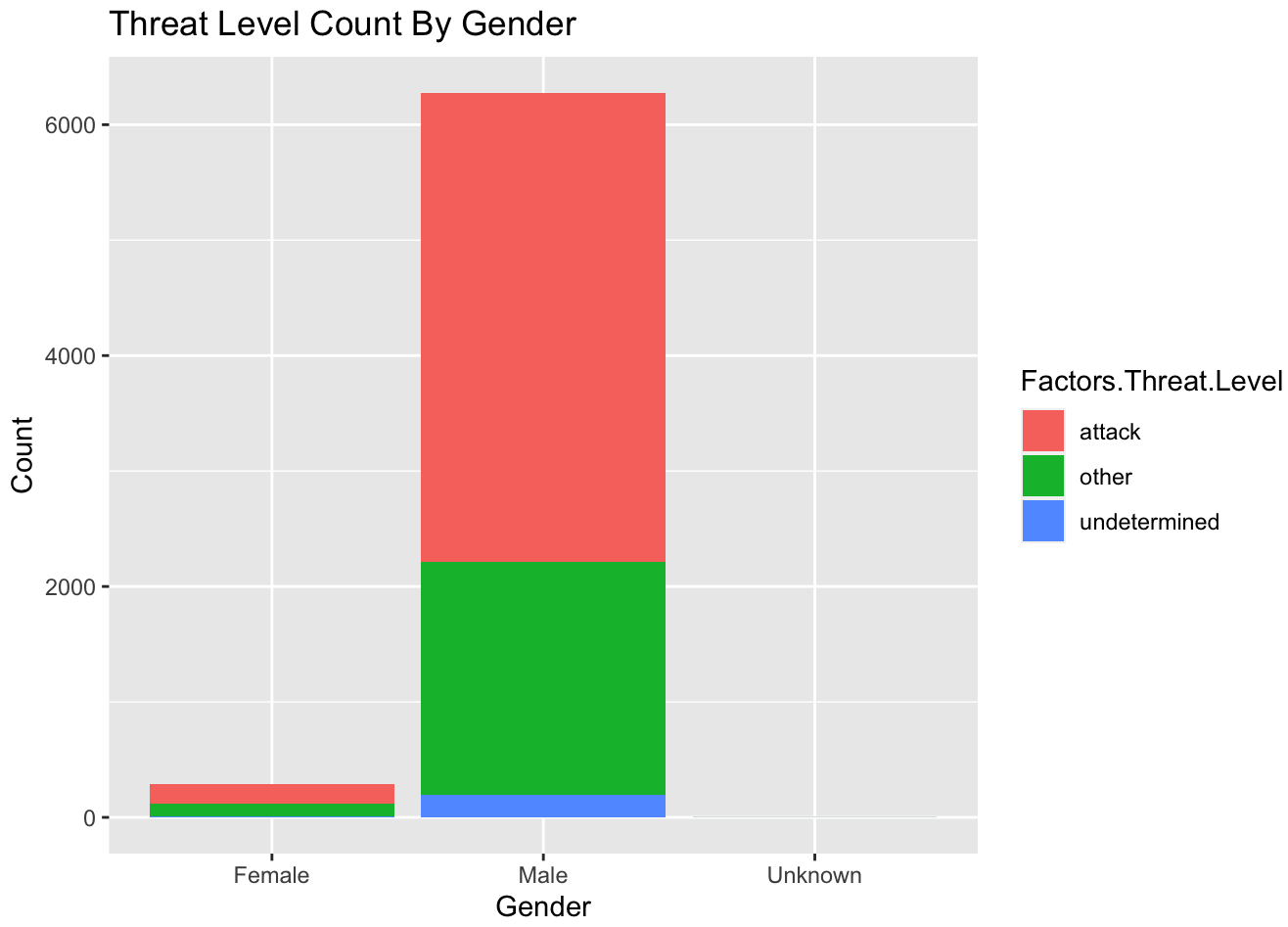
Are there discrepancies in Police shooting victims?

Univariate Distribution of Outcome Variable



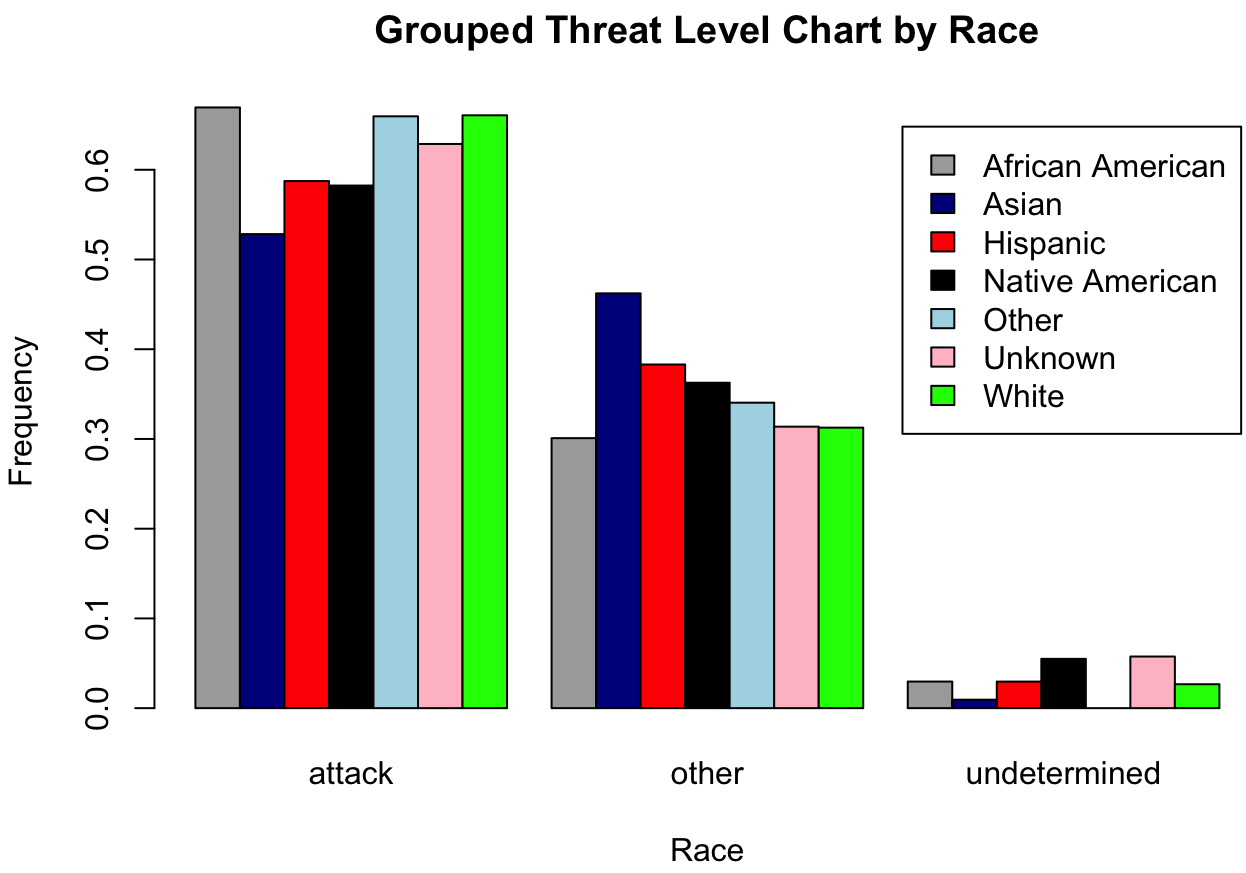
To display my outcome variable I used a barplot to distinguish the count for each Threat Level. **4230** police shooting victims had a Threat level of *attack*, **2134** victims had a Threat level of *other*, and **205** victims had an *undetermined* Threat level.

Graphs of Predictor Variables



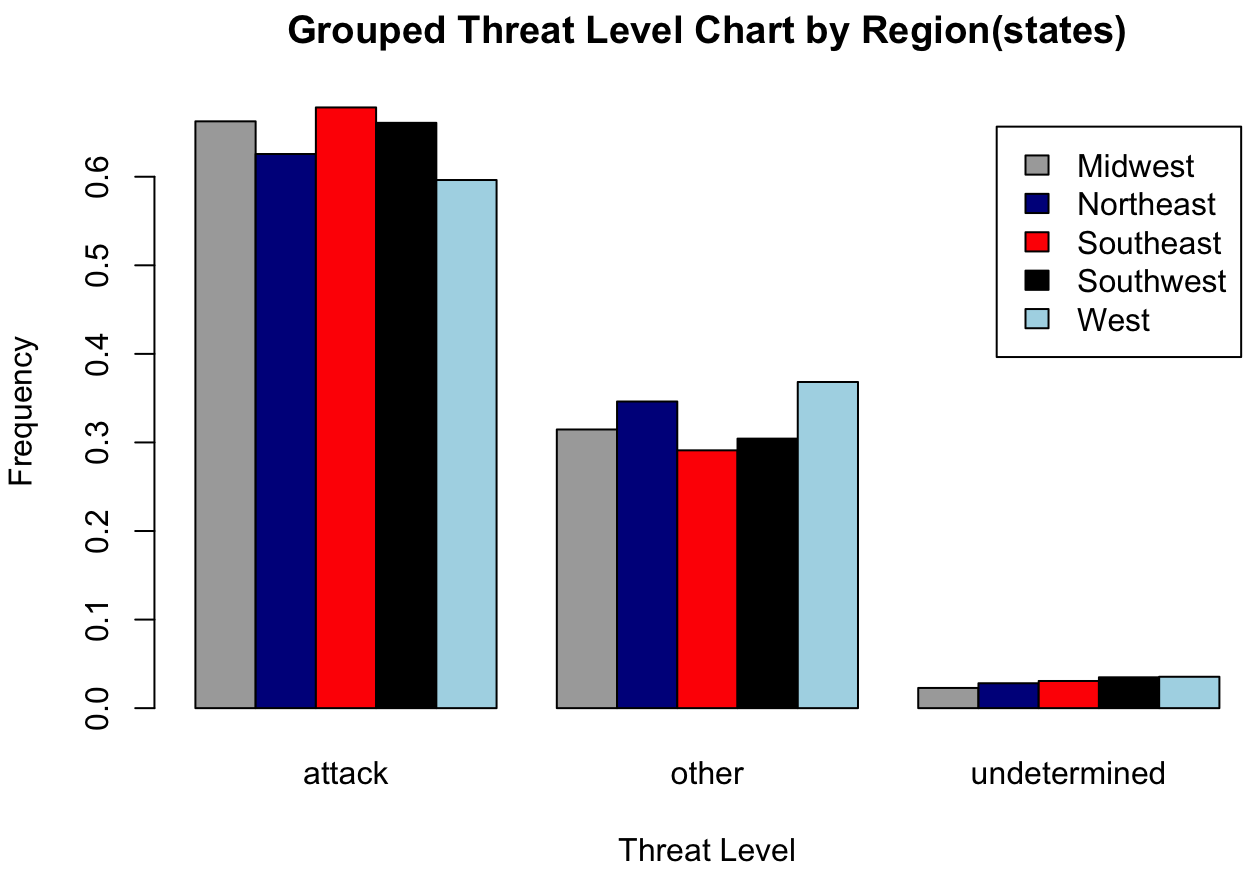

```
##
##           attack other undetermined
##   Female     170   114            6
##   Male      4059  2019          198
##   Unknown      1     1            1
```

My bivariate graph of my gender predictor shows the count of each threat level separated by gender. The graph shows that even though there is a majority of male police shooting victims there is a difference between the proportion of males, females, and unknowns that are considered “attack” level threats. The graph shows that males are largely considered “attack” level threats at about 64.6% while females even with less representation are at 58.7%.




```
##
##                   attack other undetermined
##   African American  0.669 0.301        0.030
##   Asian             0.528 0.462        0.009
##   Hispanic          0.587 0.383        0.030
##   Native American   0.582 0.363        0.055
##   Other             0.660 0.340        0.000
##   Unknown           0.629 0.314        0.058
##   White             0.661 0.313        0.027
```

My second graph is a grouped chart of the frequencies of Threat Levels by race. In this case my predictor is Race and my outcome variable remains the Threat Level. The chart shows that African Americans are considered attack level threats at the **greatest frequency** however the disparity between other demographics is not statistically significant. What should be considered is that Asians were considered an “attack” level threat at the **lowest frequency**, but were considered an “other” level threat at the **highest frequency** by a fair margin. Overall the chart shows no general disparity between races and we can conclude that there is no significant relationship between Race and Threat Level.




```
##
##             attack other undetermined
##   Midwest    0.663 0.315        0.023
##   Northeast  0.626 0.346        0.028
##   Southeast  0.678 0.291        0.031
##   Southwest  0.661 0.304        0.035
##   West       0.596 0.368        0.035
```

My final predictor graph shows the frequencies of attack, other, and undetermined threat levels by state region. The west region has the **lowest** proportion of “attack” level threats and the **highest** proportion of “other” level threats. The Southeast region of the US has the highest proportion of “attack” level threats and the lowest number of “other” level threats. It’s safe to conclude that there is no visible discrepancy between each region however there seems to be a relationship between a **high** proportion of attack level threats and a **low** proportion of other level threats. Undetermined threats are low for all regions and all three charts showed that.

Conclusion

I created a series of bar plots to describe the relationship between Threat Level and (Race, Gender, and State Region). I found that in the data set most police shooting victims were considered attack level threats, which I think was a slight limitation of the data in general. If we had more data which included a variety of threat levels we could clearly distinguish if there is any disparity between race, gender, and even state. In 2020 research was done about Police shooting victims and it was found that people of color were getting disproportionately shot

compared to other demographics. In this data set my bar plots could not conclude that because of the limitations of threat level and race. For more information on Police shootings and racial disparities here is an article (https://www.manhattan-institute.org/verbruggen-fatal-police-shootings).