

# NYPD\_Data

R. Buessing

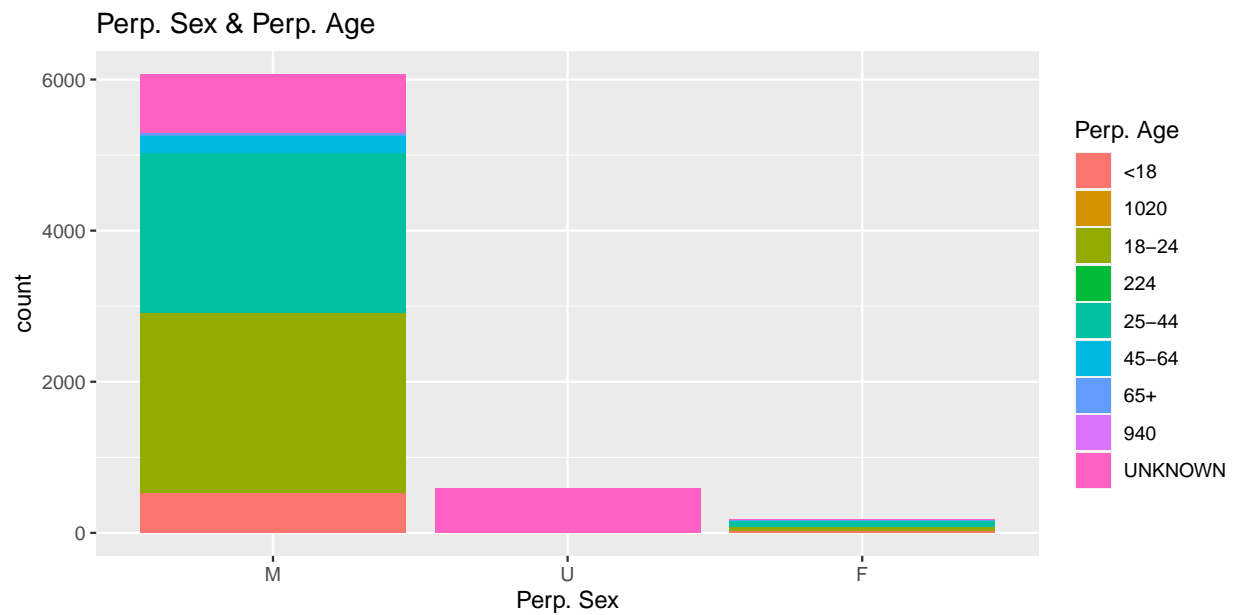
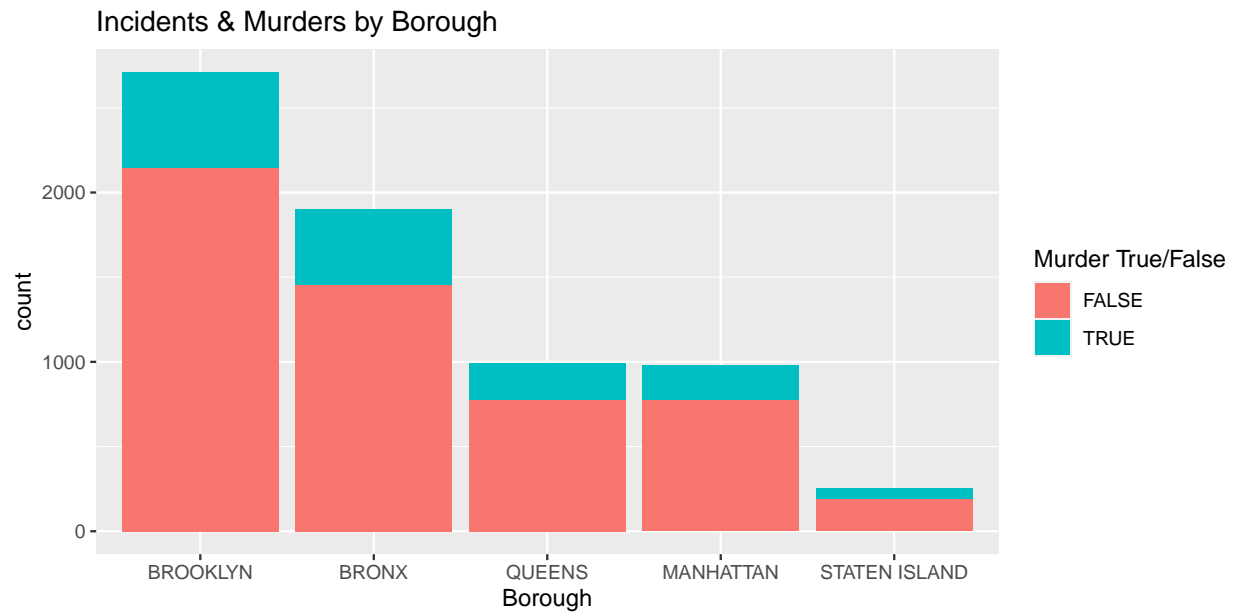
10/11/2021

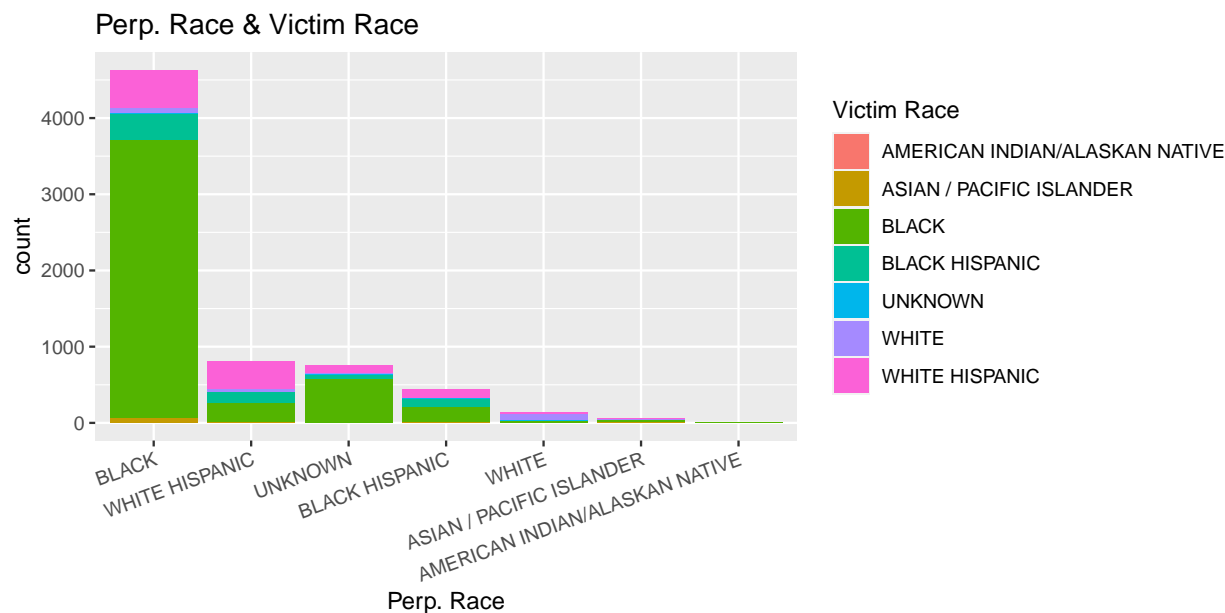
In this report we investigate possible correlation between multiple data points in the NYPD Shooting Incident data. We look at whether or not factors such as age, location, sex, and ethnicity impact the volume and severity of shooting incident in New York City.

The code behind this report can be located at the following public GitHub repository: [https://github.com/rabu4751/dsaf\\_final](https://github.com/rabu4751/dsaf_final)

Below are several charts that detail these potential correlations.

## Data Plots





## Findings

From these charts we get an idea of what factors influence the volume of shooting incidents. We can see that the poorer borough's have a significantly higher number of shooting incidents compared to wealthier borough's. Not only this, the incidents that do occur in poorer regions also tend to have a higher murder rate. We can also see that perpetrators of these crimes are generally male and are of Black or Hispanic ethnicity. It's important to note that there may be bias here in the form of under reporting (similar to sampling bias) for those of White ethnicity and female incidents while on the other hand over reporting could lead to what we see in our data now with a possible misrepresentation of Black and Hispanic incidents. I also purposely left three outliers in the Perp. Age & Perp. Sex graph to show a form of measurement bias in the fact that the ages shown were obviously a result of human error when entering the data.

## Predictions

Below we take several of the data points from the data (displayed below) to predict the probability of surviving a shooting incident in New York City. This model uses logisitic regression to make this prediction.

Table 1: Data to Predict on

BORO	VIC_RACE	VIC_AGE_GROUP	PERP_AGE_GROUP	LOCATION_DESC
BRONX	BLACK	18-24	18-24	MULTI DWELL - PUBLIC HOUSING
QUEENS	WHITE HISPANIC	25-44	25-44	PVT HOUSE
MANHATTAN	WHITE	65+	65+	PVT HOUSE

Table 2: Prediction/Survival Probability

x
0.2060895
0.4114317

$$\frac{x}{0.7195879}$$

## Conclusion

As we can see, the prediction follows the trends we saw earlier with the White, wealthy individual having the highest survival probability compared to the Black and Hispanic individuals located in the poorer regions of the city. The graphs and model used in this report only used complete data while non-complete data was filtered out.