NOTProject1

Jared Brotamonte

1/26/2024

Introduction

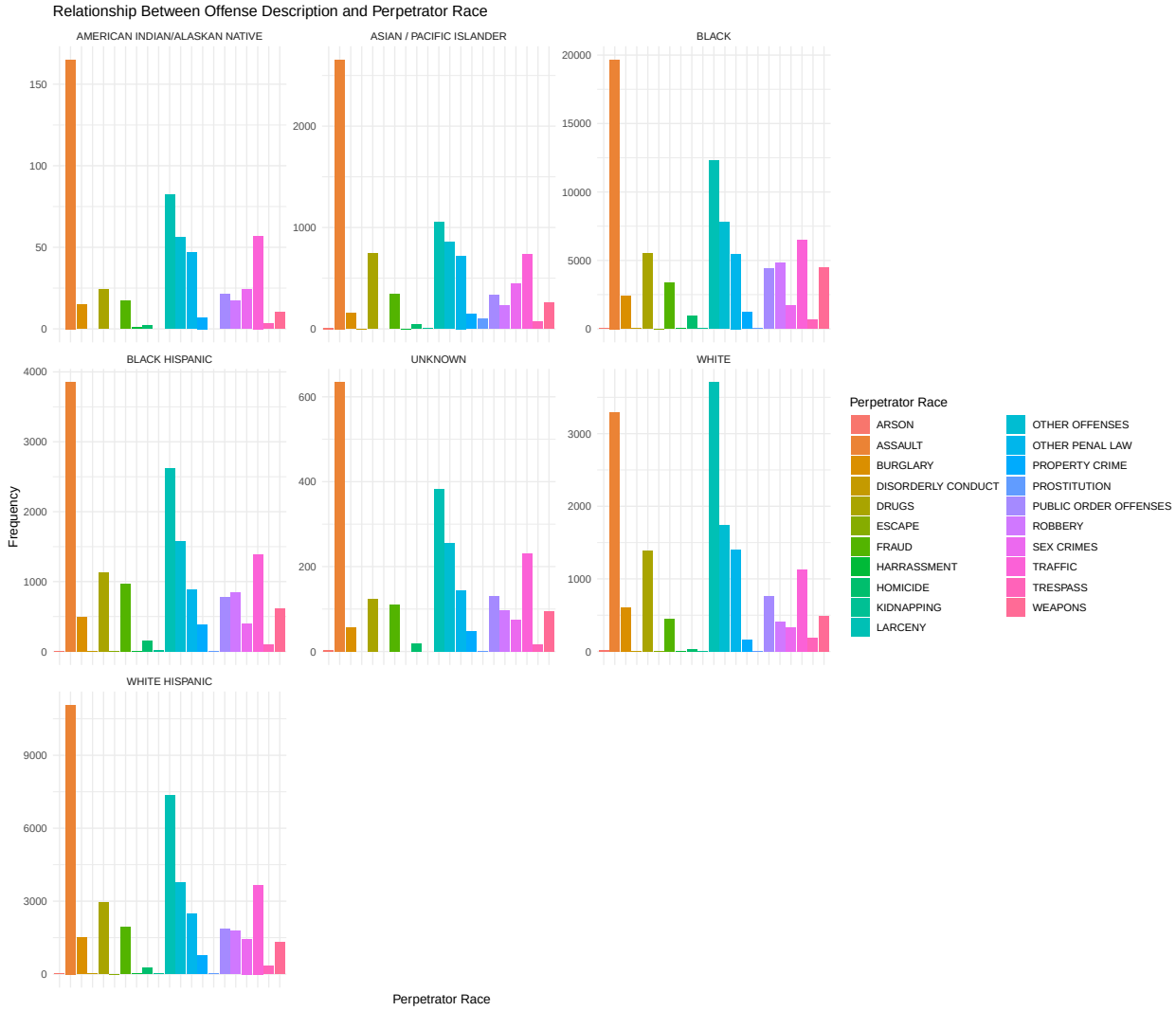
For this project, I decided to look into the topic of law enforcement and whether or not law enforecment truly targets minorities and people of color.
The difference in power when comparing law enforcement to the average citizen has always been hugely drastic. This imbalance of power has recently led to many conflicts arising between law enforcement and US citizens with the bases of these conflicts being that law enforcement has too much ability to abuse their power without enough consequences. In particular, I wanted to focus on claim that law enforcement abuses their power towards minorities or more specifically people of color

Data Cleaning

In order to clean the data and make it more usable, I had to discern what variables where actually neccessary for analyzing the data. Many variables held similar information as other variables like how OFNS_DESC and PD_DESC both describe the type of offense that was commited, but one is just more specific than the other. There are also variables that are just hard to interpret, such as the variable LAW_CODE which is just a number which corresponds to the law broken, but because I was given just the law code number, it makes it hard to use this number and generate some type of analysis. Much of the data was cleaned out due to the reasons prior which in turn left me with 10 variables left over.
To clean the left over variables I had to deal with a couple different problems. Firstly I made sure to properly format the variable ARREST_DATE as dates. I then made sure to only keep valid data points, thus I got rid of data points containing null values as well as data points that did not contain valid information. Then lastly I had to make sure to turn the variables that needed to be factored into factors. This also included cleaning up factors. For the variable OFNS_DESC, since there where too many factors in for this variables, I had to go through and change the factors to be more generalized. As for the other factors, I had to make sure to check that those variables didn't contain any factors that shouldn't exist.

```r
library(NOTProject1)
library(ggplot2)
```

```r
ggplot(NYPD_Arrest, aes(x = OFNS_DESC, fill = OFNS_DESC)) +
  geom_bar(position = position_dodge()) +
  facet_wrap(~PERP_RACE, scales = "free_y") +  # scales = "free_y" for better visualization
  labs(
    x = "Perpetrator Race",
    y = "Frequency",
    title = "Relationship Between Offense Description and Perpetrator Race",
    fill = "Perpetrator Race"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_blank())  # Remove x-axis labels due to clutter
```



```r
ggplot(NYPD_Arrest, aes(x = ARREST_BORO, fill = ARREST_BORO)) +
  geom_bar(position = position_dodge()) +
  facet_wrap(~PERP_RACE, scales = "free_y") +  # scales = "free_y" for better visualization
  labs(
    x = "Arrest Borough",
    y = "Frequency",
    title = "Relationship Between Arrest Borough and Perpetrator Race",
    fill = "Perpetrator Race"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_blank())  # Remove x-axis labels due to clutter
```