NYPD Analisys

RT

2023-04-02

Importing Data

I started importing the NYPD Shooting data from https://catalog.data.gov/dataset/nypd-shooting-incident-data-historic. So, I created a "url" variable to read the data which the website has.

```
url_data<-"https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD"
```

Readind Data

Let's read in the data and see what I have.

```
data_shopting <- read.csv(url_data)
```

Tidying Data

After looking at the data. I would like to tidy data set and remove the columns that I am not interested and rename other. Also, I would like see the range of dates in my data set, so I summarized my data, to have an idea what I would like to visualize and create my model. Other modification that I did was mutating the data from a character to actually a date.

```
data_shopting<-data_shopting %>% select(-c(INCIDENT_KEY,OCCUR_TIME,JURISDICTION_CODE,
LOCATION_DESC,STATISTICAL_MURDER_FLAG,PERP_AGE_GROUP,PERP_SEX,PERP_RACE,X_COORD_CD,
Y_COORD_CD,Latitude,Longitude,Lon_Lat))
data_shopting<-data_shopting %>% rename(CITY = BORO)
data_shopting <- data_shopting %>% mutate(OCCUR_DATE = mdy(OCCUR_DATE))
data_shopting$year <- format(as.Date(data_shopting$OCCUR_DATE, format="%m/%d/%Y"),"%Y")
summary(data_shopting)
```

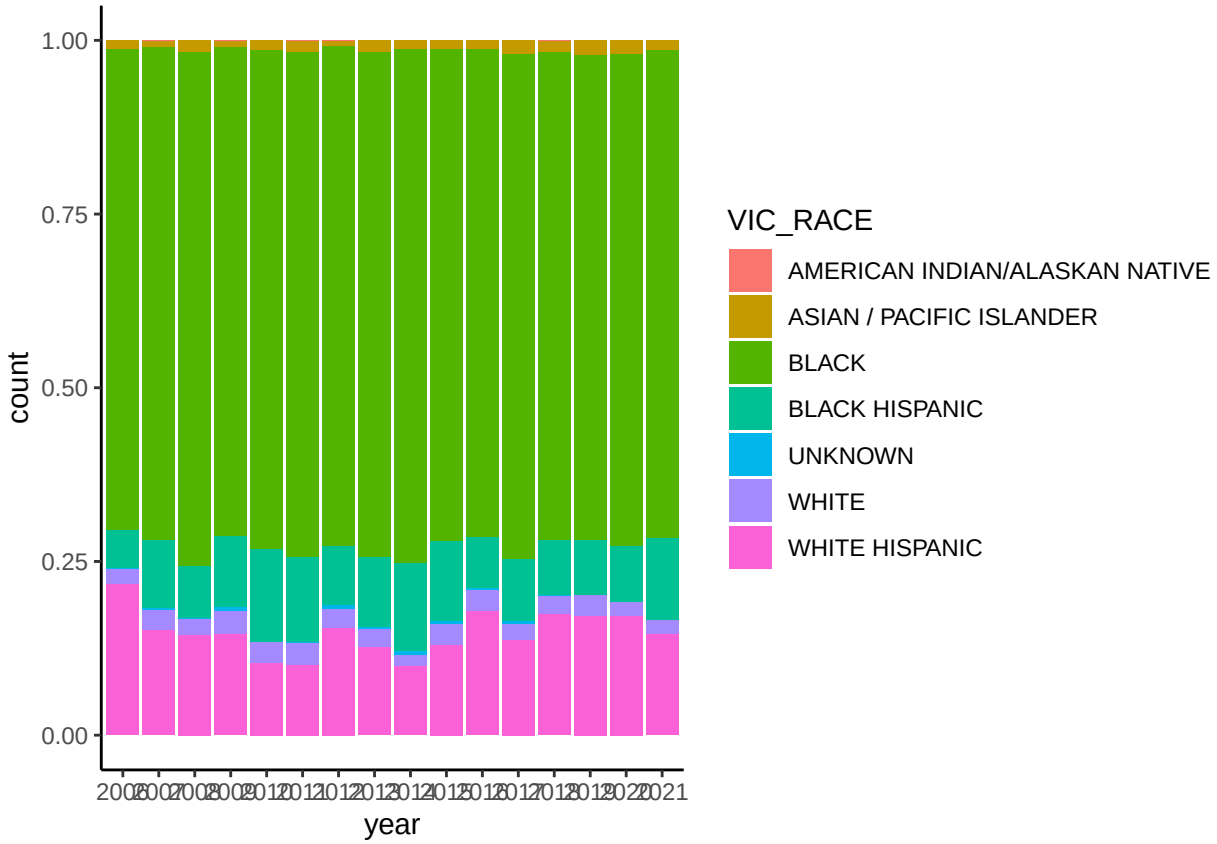
```
##    OCCUR_DATE            CITY              PRECINCT      VIC_AGE_GROUP
##  Min.   :2006-01-01   Length:25596       Min.   :  1.00   Length:25596
##  1st Qu.:2009-05-10   Class :character   1st Qu.: 44.00   Class :character
##  Median :2012-08-26   Mode  :character   Median : 69.00   Mode  :character
##  Mean   :2013-06-13                      Mean   : 65.87
##  3rd Qu.:2017-07-01                      3rd Qu.: 81.00
##  Max.   :2021-12-31                      Max.   :123.00
##    VIC_SEX            VIC_RACE             year
##  Length:25596       Length:25596       Length:25596
##  Class :character   Class :character   Class :character
```

```
##  Mode   :character   Mode   :character   Mode   :character
##
##
##
```

Visualization

As I have the data from 2006 to 2021, I would like to see the data by year compared by race and see what it would bring me. So I created a percent stack chart by year.

```
ggplot(data_shopting, aes(x = year, fill = VIC_RACE)) +
geom_bar(position = "fill") +
theme_classic()
```

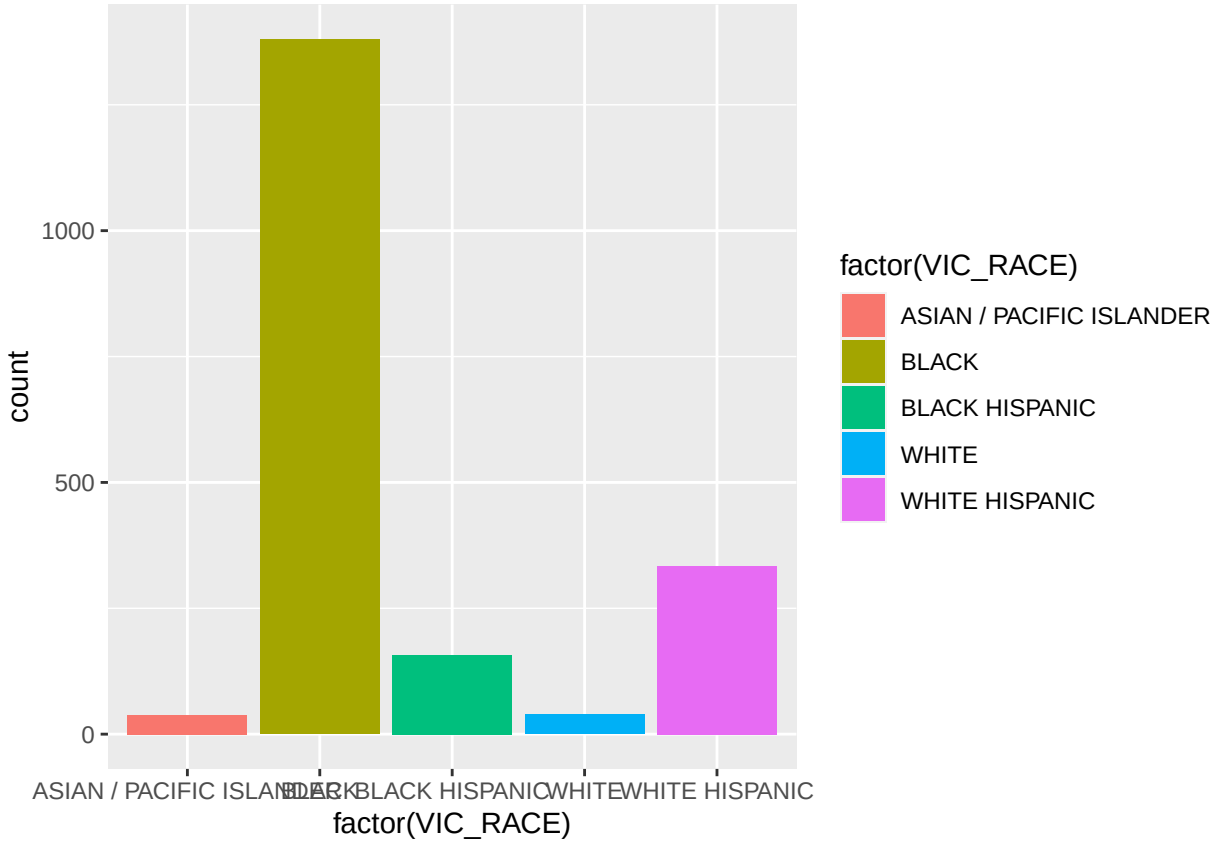


Second, I filtered only the 2020 year to start thinking about a model that I would like to check. The last Census of polulatuon New York had was 2020.

```
data_2020 <- data_shopting%>% filter(year==2020)
```

So, I created a bar plot for only 2020 quantity of shoots by race, to see what would come.

```
ggplot(data_2020, aes(factor(VIC_RACE),
fill = factor(VIC_RACE))) +
geom_bar()
```



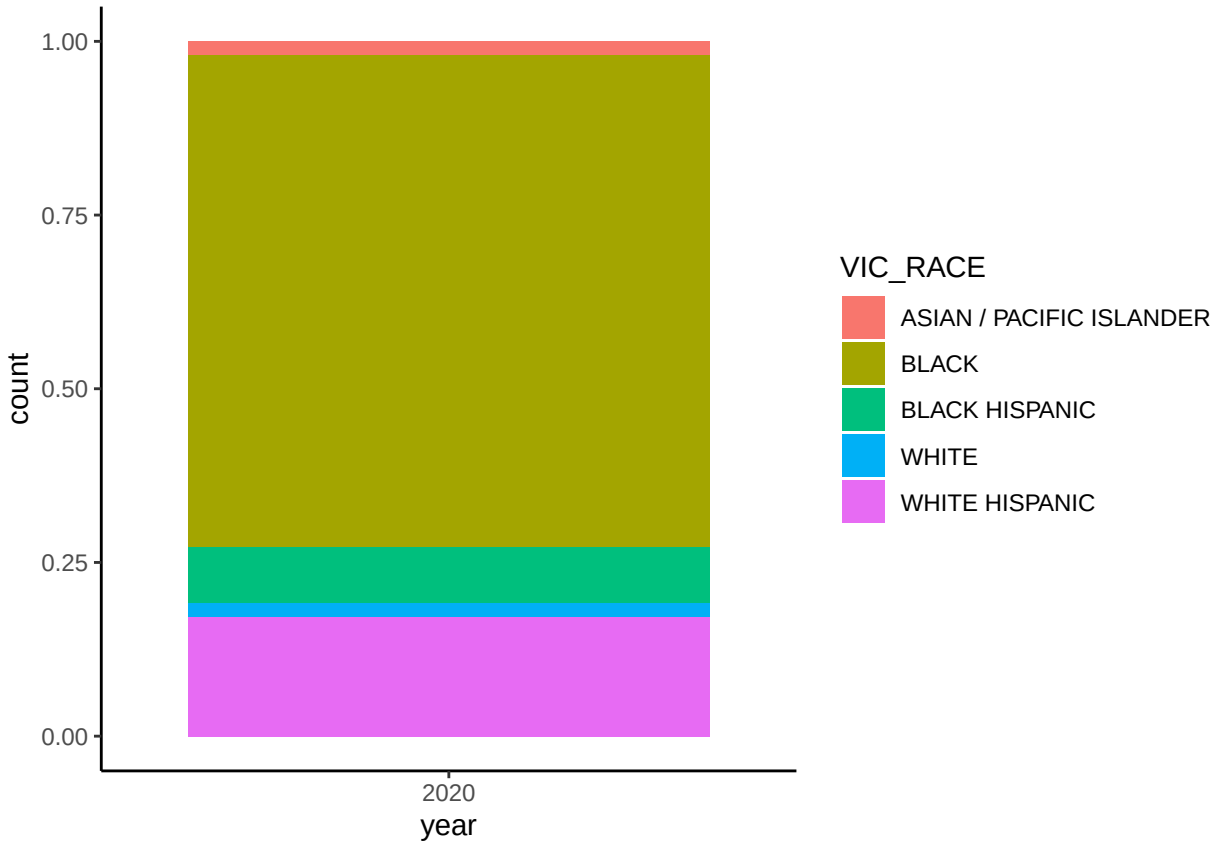
Model and Analysis

Once I see the data and the visualizations I decided to compare the population in NY state with the shot by race in 2020. The Census 2020 is found here https://www.census.gov/library/stories/state-by-state/new-york-population-change-between-census-decade.html.

It says that 12.4% of the population is Black and 61.6% is White, 18.7% is hispanic (black and white), 6.2% is Asian/Pacific Islander and 6.1% others. when you compare with the shot distribution by race you see that even though the population of white people is almost 5x bigger than black, the black people (70.8% of the shots) get shot almost 25x more than white people (only 2% of the shots).

This is a huge indicative of the social and racism problems NY and most likely the USA have in your society. Asian and white people are least likely to get shot than Hispanic and mainly black people in New York state.

```
ggplot(data_2020, aes(x = year, fill = VIC_RACE)) +
geom_bar(position = "fill") +
theme_classic()
```



Conclusion

When I saw this I assignment I had already the decision what to do with it, it was to show that black and Hispanic people suffers more with violence than white in NY state. the was my bias because even before have any visualization or anything I would like to show that, the results show that, but maybe because I am Brazilian and an immigrant my personal bias was already taking the lead on my analysis to show what it is true that Hispanic and black people suffers more with violence than white even though black and Hispanic population are much smaller than white. we can discuss the reasons and for future analysis include other facts to go deeper in the problem, but the social problems in this data show that something needs to change so we can have the same opportunities to all people in the world regardless their race or origin.