

NYPD Shooting Data

Bryan Craig

5/21/2021

NYPD Historical Shooting Data Analysis

The following is a step-by-step process to conduct a basic analysis of NYPD Historical shooting Data Analysis

1 I am going to read in the data and see what it looks like:

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

summary(shootingdata)

##
## -- Column specification -----
##
## cols(
##   INCIDENT_KEY = col_double(),
##   OCCUR_DATE = col_character(),
##   OCCUR_TIME = col_time(format = ""),
##   BORO = col_character(),
##   PRECINCT = col_double(),
##   JURISDICTION_CODE = col_double(),
##   LOCATION_DESC = col_character(),
##   STATISTICAL_MURDER_FLAG = col_logical(),
##   PERP_AGE_GROUP = col_character(),
##   PERP_SEX = col_character(),
##   PERP_RACE = col_character(),
##   VIC_AGE_GROUP = col_character(),
##   VIC_SEX = col_character(),
##   VIC_RACE = col_character(),
##   X_COORD_CD = col_number(),
##   Y_COORD_CD = col_number(),
##   Latitude = col_double(),
##   Longitude = col_double(),
##   Lon_Lat = col_character()
## )

summary(shootingdata)
```

```

## INCIDENT_KEY OCCUR_DATE OCCUR_TIME BORO
## Min. : 9953245 Length:23568 Length:23568 Length:23568
## 1st Qu.: 55317014 Class :character Class1:hms Class :character
## Median : 83365370 Mode :character Class2:difftime Mode :character
## Mean :102218616 Mode :numeric
## 3rd Qu.:150772442
## Max. :222473262
##
## PRECINCT JURISDICTION_CODE LOCATION_DESC STATISTICAL_MURDER_
FLAG
## Min. : 1.00 Min. :0.0000 Length:23568 Mode :logical
## 1st Qu.: 44.00 1st Qu.:0.0000 Class :character FALSE:19080
## Median : 69.00 Median :0.0000 Mode :character TRUE :4488
## Mean : 66.21 Mean :0.3323
## 3rd Qu.: 81.00 3rd Qu.:0.0000
## Max. :123.00 Max. :2.0000
## NA's :2
## PERP_AGE_GROUP PERP_SEX PERP_RACE VIC_AGE_GROUP
## Length:23568 Length:23568 Length:23568 Length:23568
## Class :character Class :character Class :character Class :character
## Mode :character Mode :character Mode :character Mode :character
##
##
##
## VIC_SEX VIC_RACE X_COORD_CD Y_COORD_CD
## Length:23568 Length:23568 Min. : 914928 Min. :125757
## Class :character Class :character 1st Qu.: 999900 1st Qu.:182565
## Mode :character Mode :character Median :1007645 Median :193482
## Mean :1009363 Mean :207312
## 3rd Qu.:1016807 3rd Qu.:239163
## Max. :1066815 Max. :271128
##
## Latitude Longitude Lon_Lat
## Min. :40.51 Min. : -74.25 Length:23568
## 1st Qu.:40.67 1st Qu.: -73.94 Class :character
## Median :40.70 Median : -73.92 Mode :character
## Mean :40.74 Mean : -73.91
## 3rd Qu.:40.82 3rd Qu.: -73.88
## Max. :40.91 Max. : -73.70
##

```

2 I am now going to start tidying the data. I am not going to be using the Latitude, Longitude, Lon_Lat, X_COORD_CD, or Y_COORD_CD columns. Therefore, I am going to eliminate these columns.

```

shootingdata <- shootingdata %>%
  select(-c(Latitude, Longitude, Lon_Lat))

```

```
shootingdata <- shootingdata %>%
  select(-c(X_COORD_CD,Y_COORD_CD))
```

3 I am going to convert the date field to date formatting.

```
shootingdata <- shootingdata %>%
  mutate(OCCUR_DATE = mdy(OCCUR_DATE))
```

4 I am going to rename the BORO column to 'Neighborhood'.

```
shootingdata <- shootingdata %>%
  rename('Neighborhood' = BORO)
```

5 I am going to remove any data records that have blank, or NA, entries.

```
shootingdata <- shootingdata %>%
  drop_na(LOCATION_DESC) %>%
  drop_na(PERP_SEX) %>%
  drop_na(PERP_AGE_GROUP) %>%
  drop_na(PERP_SEX) %>%
  drop_na(JURISDICTION_CODE) %>%
  drop_na(LOCATION_DESC) %>%
  drop_na(VIC_AGE_GROUP, VIC_RACE, VIC_SEX)
```

6 Finally, I am going to add a frequency column and add 1 for each record so that I can perform some analysis and visualizations.

```
IncidentCount <- c(1)
shootingdata$new_col <- IncidentCount
```

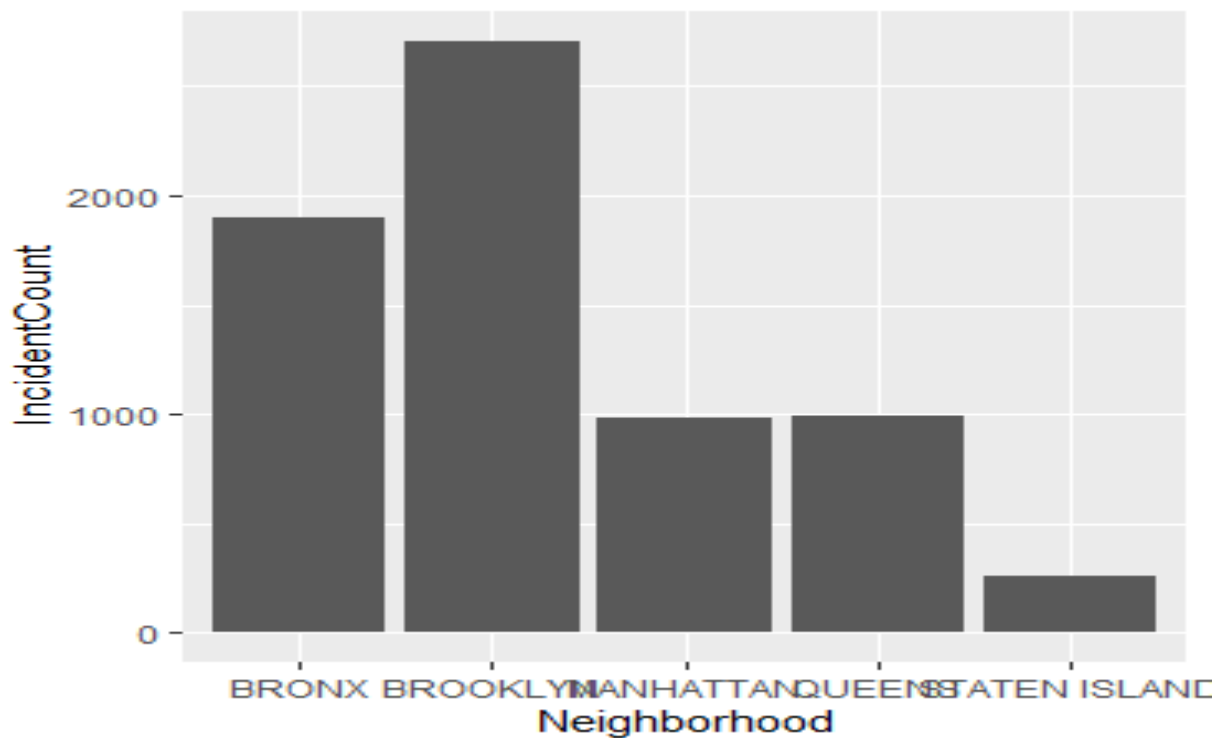
7 Now let's see what the data looks like

```
## INCIDENT_KEY      OCCUR_DATE      OCCUR_TIME      BORO
## Min.   : 9953245    Length:23568    Length:23568    Length:23568
## 1st Qu.: 55317014   Class :date      Class1:hms      Class :character
## Median : 83365370   Mode  :date      Class2:difftime Mode  :character
## Mean   :102218616                    Mode  :numeric
## 3rd Qu.:150772442
## Max.   :222473262
##
## PRECINCT      JURISDICTION_CODE LOCATION_DESC      STATISTICAL_MURDER_
FLAG
## Min.   : 1.00    Min.   :0.0000    Length:23568    Mode :logical
## 1st Qu.: 44.00    1st Qu.:0.0000    Class :character FALSE:19080
## Median : 69.00    Median :0.0000    Mode  :character TRUE :4488
## Mean   : 66.21    Mean   :0.3323
## 3rd Qu.: 81.00    3rd Qu.:0.0000
```

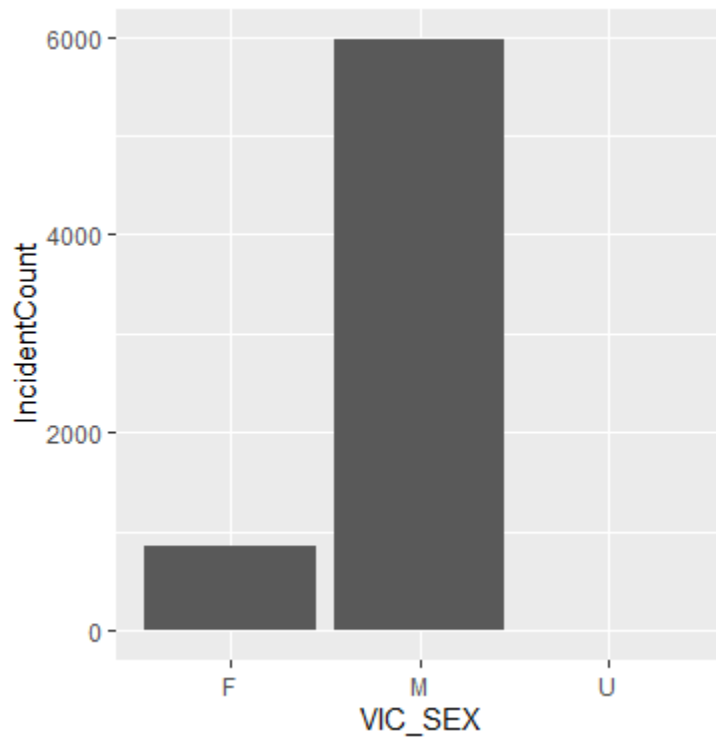
```
## Max.      :123.00    Max.      :2.0000
##                                     NA's      :2
## PERP_AGE_GROUP      PERP_SEX      PERP_RACE      VIC_AGE_GROUP
## Length:23568        Length:23568    Length:23568    Length:23568
## Class :character    Class :character Class :character Class :character
## Mode  :character    Mode  :character Mode  :character Mode  :character
##
##
##
## VIC_SEX      VIC_RACE
## Length:23568  Length:23568
## Class :character Class :character
## Mode  :character Mode  :character
##
##
##
##
```

8 Now we are going to do some data visualizations and analysis to see what the shooting data tells us.

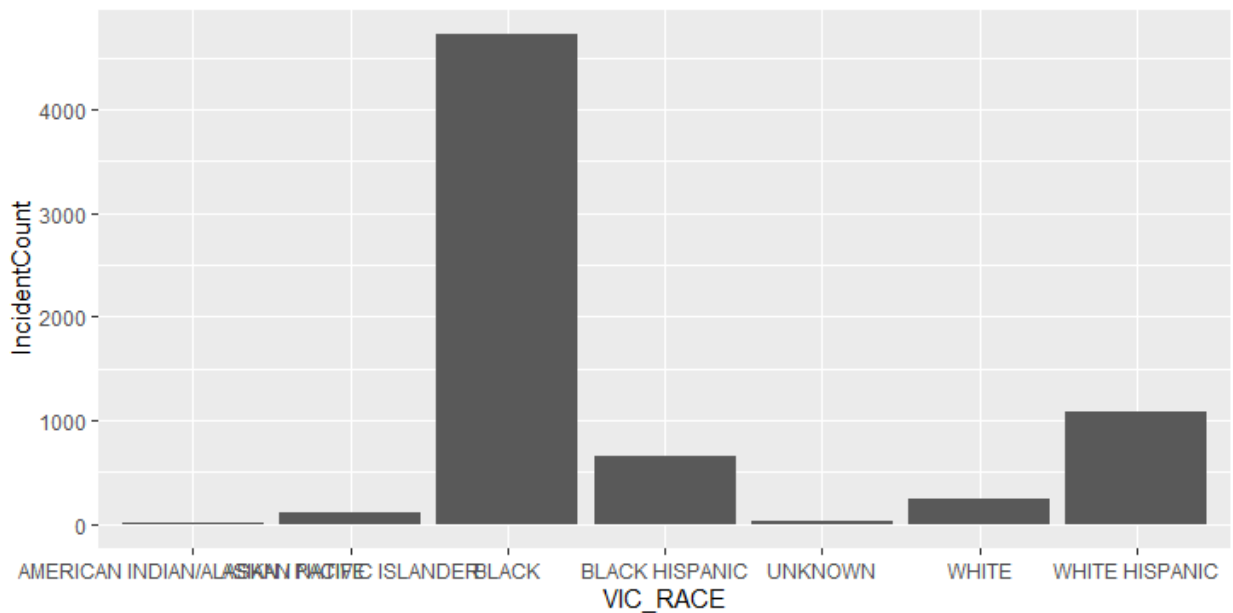
8a. Analysis to see which neighborhoods had shooting incidents. It looks like Brooklyn.



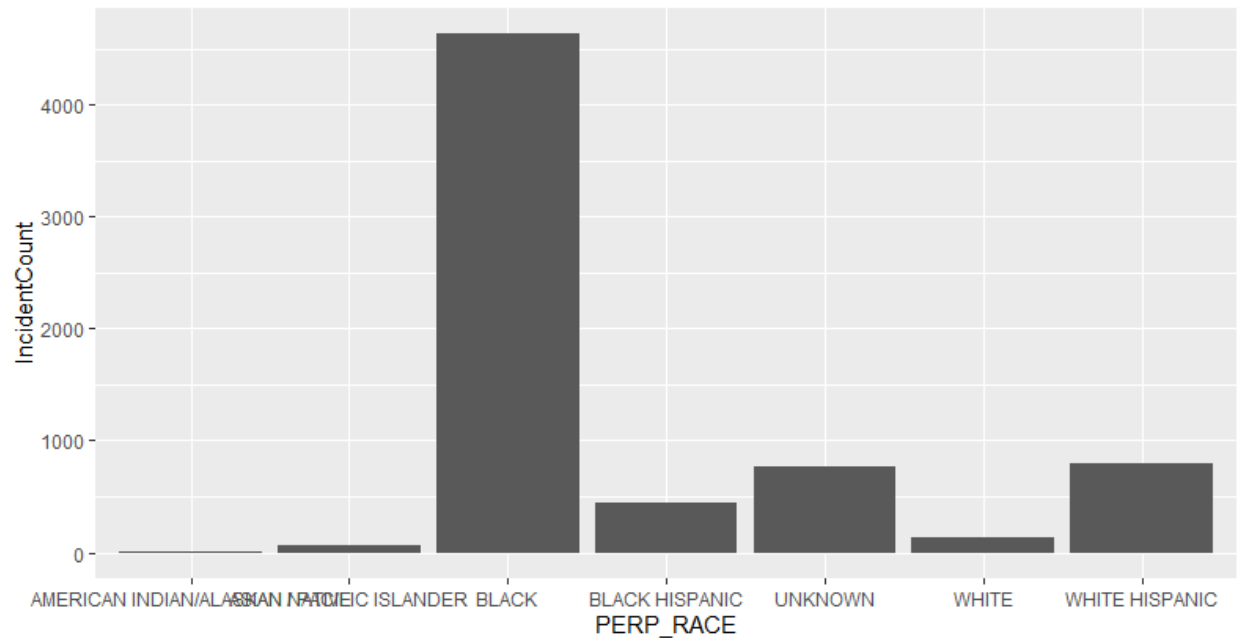
8b. Analysis to see which victim gender had shooting incidents. It looks like Males are predominant victims.



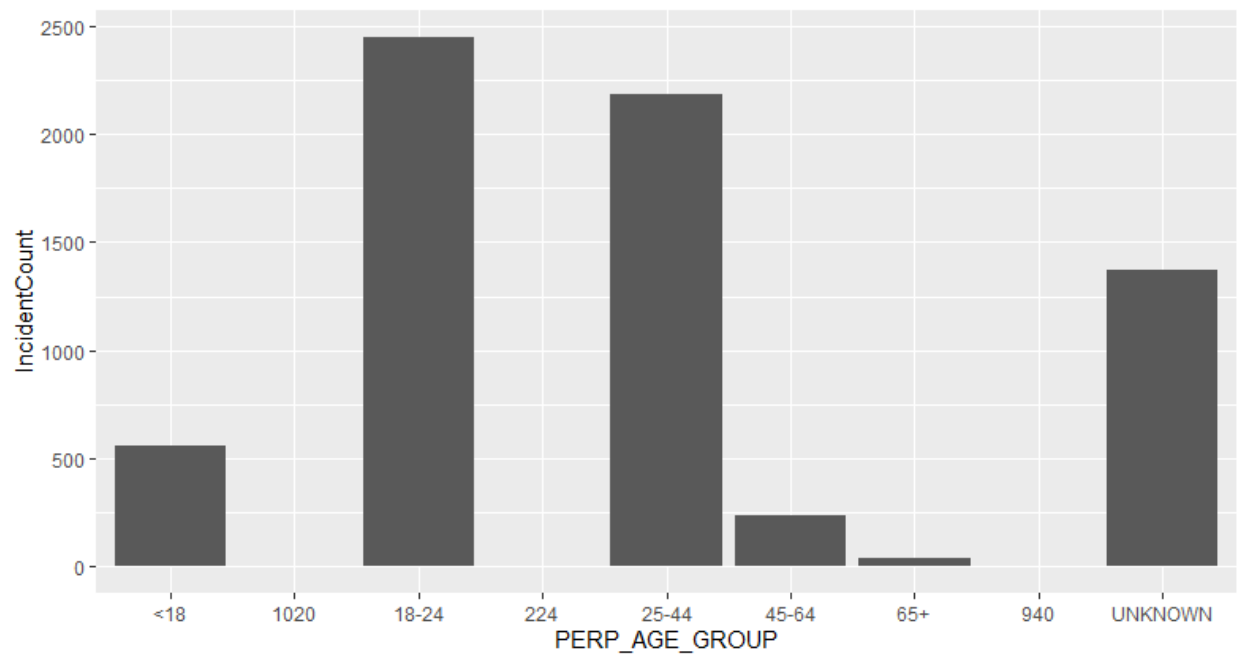
8c. Analysis to see which victim race had shooting incidents. It looks like predominant race as victims were blacks.



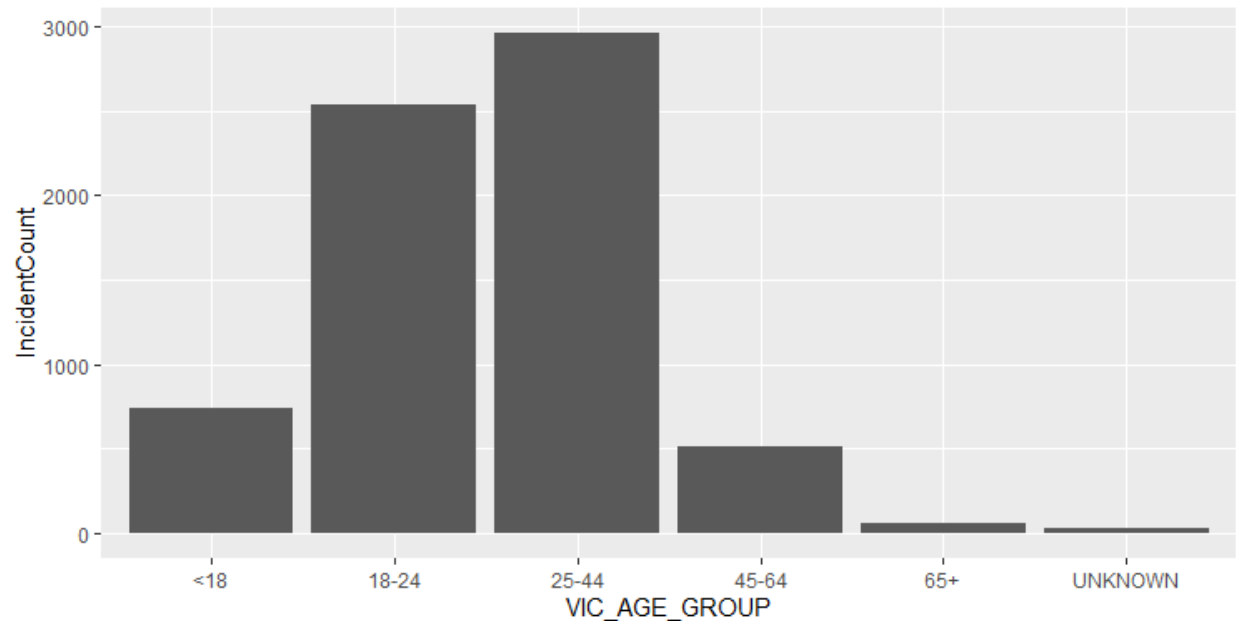
8d. Analysis to see which perpetrator race had shooting incidents. It looks like predominant race as perpetrator were blacks.



8e. Analysis to see which perpetrator age group had shooting incidents. It looks like predominant age group as perpetrator were 18-24 year olds.



8f. Analysis to see which victim age group had shooting incidents. It looks like predominant age group as victims were 25-44 year olds.



8g. Now I am going to do a basic model to see if there is an association between murder rates and victims who are female. I will have to convert the Murder Rate Flag from TRUE/FALSE to binary, as well as convert Victim Sex of M/F to binary.

Convert Murder to integer for modelling

```
shootingdata$MurderInt <- as.integer(shootingdata$STATISTICAL_MURDER_FLAG)
```

Convert Victim Sex to integer for modelling

```
shootingdata$VictimSexInt <- as.integer(shootingdata$VIC_SEX)
```

Conduct linear model and graph

```
install.packages("modelr")
```

```
library(modelr)
```

```
library(ggplot2)
```

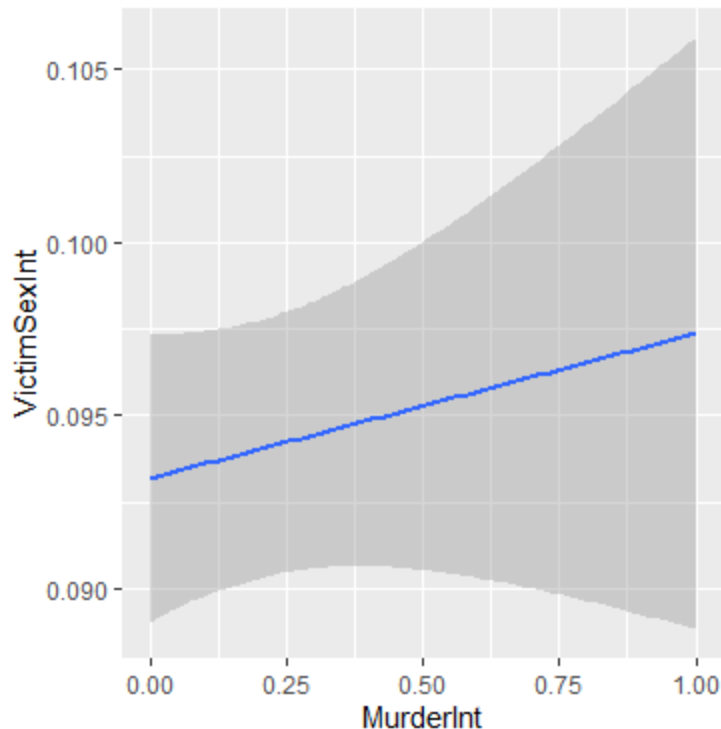
```
model<-lm(VictimSexInt ~ MurderInt, data = shootingdata)
```

```
model
```

```
ggplot(shootingdata, aes(MurderInt, VictimSexInt)) +  
  stat_smooth(method=lm)
```

Coefficients:

(Intercept)	MurderInt
0.093187	0.004184



It appears that there is a strong association that murder victims were women.

Conclusion

Basic analysis indicates majority of shooting incidents are in Brooklyn. It appears that the predominant victims are black males between the ages of 25-44. It also appears that the predominant perpetrators of shooting incidents are black males between the ages of 18-24. Finally, basic modelling and analysis shows that there is a high correlation of shooting murders involving women victims.

Bias

Possible sources of bias could include perceiving that New York City, or any larger urban areas, are dangerous and have high numbers of shootings. Also, another source of bias is the perception that shootings are specific to only a few socio-economic and race groups (i.e. crimes are only committed by one type of person).

To mitigate against such bias is to run additional analysis against other cities and small-/mid-size towns, and conduct the incidents as a per capita percentage. Also, conduct analysis by looking at data across population data to gain better insights about the socio-economic and racial groups living across a broader swath of neighborhoods.

Citation

1. City of New York, Open Data (November 10, 2020). Retrieved from <https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv?accessType=DOWNLOAD>