

# NYPD Shooting Incident Data

2022-09-10

## Description of Data

The NYPD Shooting Incident data set is a comprehensive list of shooting incidents that occurred in NYC starting in 2006 through the end of the previous calendar year. Each incident includes information on the location and time of the event, as well as victim and suspect demographic descriptions.

## Import Data

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

```
## Rows: 25596 Columns: 19
## -- Column specification -----
## Delimiter: ","
## chr  (10): OCCUR_DATE, BORO, LOCATION_DESC, PERP_AGE_GROUP, PERP_SEX, PERP_R...
## dbl  (7): INCIDENT_KEY, PRECINCT, JURISDICTION_CODE, X_COORD_CD, Y_COORD_CD...
## lgl  (1): STATISTICAL_MURDER_FLAG
## time (1): OCCUR_TIME
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

## Initial Summary of Data

```
summary(nypd_data)
```

##	INCIDENT_KEY	OCCUR_DATE	OCCUR_TIME	BORO
##	Min. : 9953245	Length:25596	Length:25596	Length:25596
##	1st Qu.: 61593633	Class :character	Class1:hms	Class :character
##	Median : 86437258	Mode :character	Class2:difftime	Mode :character
##	Mean :112382648		Mode :numeric	
##	3rd Qu.:166660833			
##	Max. :238490103			
##				
##	PRECINCT	JURISDICTION_CODE	LOCATION_DESC	STATISTICAL_MURDER_FLAG
##	Min. : 1.00	Min. :0.0000	Length:25596	Mode :logical
##	1st Qu.: 44.00	1st Qu.:0.0000	Class :character	FALSE:20668
##	Median : 69.00	Median :0.0000	Mode :character	TRUE :4928
##	Mean : 65.87	Mean :0.3316		
##	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:25596	Length:25596	Length:25596	Length:25596
##	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:25596 Length:25596 Min. : 914928 Min. :125757
## Class :character Class :character 1st Qu.:1000011 1st Qu.:182782
## Mode :character Mode :character Median :1007715 Median :194038
## Mean :1009455 Mean :207894
## 3rd Qu.:1016838 3rd Qu.:239429
## Max. :1066815 Max. :271128
##
## Latitude Longitude Lon_Lat
## Min. :40.51 Min. : -74.25 Length:25596
## 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
##
```

## Data Transformation

```
nypd_data <- nypd_data %>% rename(Borough = 'BORO') %>% #change name from BORO to Borough
mutate(OCCUR_DATE = mdy(OCCUR_DATE)) %>% # convert date column entires to a date objects
select(-c(PRECINCT, JURISDICTION_CODE, LOCATION_DESC, STATISTICAL_MURDER_FLAG, X_COORD_CD, Y_COORD_CD)
summary(nypd_data)
```

```
## INCIDENT_KEY OCCUR_DATE OCCUR_TIME Borough
## Min. : 9953245 Min. :2006-01-01 Length:25596 Length:25596
## 1st Qu.: 61593633 1st Qu.:2009-05-10 Class1:hms Class :character
## Median : 86437258 Median :2012-08-26 Class2:difftime Mode :character
## Mean :112382648 Mean :2013-06-13 Mode :numeric
## 3rd Qu.:166660833 3rd Qu.:2017-07-01
## Max. :238490103 Max. :2021-12-31
## PERP_AGE_GROUP PERP_SEX PERP_RACE VIC_AGE_GROUP
## Length:25596 Length:25596 Length:25596 Length:25596
## Class :character Class :character Class :character Class :character
## Mode :character Mode :character Mode :character Mode :character
##
##
##
## VIC_SEX VIC_RACE Latitude Longitude
## Length:25596 Length:25596 Min. :40.51 Min. : -74.25
## Class :character Class :character 1st Qu.:40.67 1st Qu.: -73.94
## Mode :character Mode :character Median :40.70 Median : -73.92
## Mean :40.74 Mean : -73.91
## 3rd Qu.:40.82 3rd Qu.: -73.88
## Max. :40.91 Max. : -73.70
```

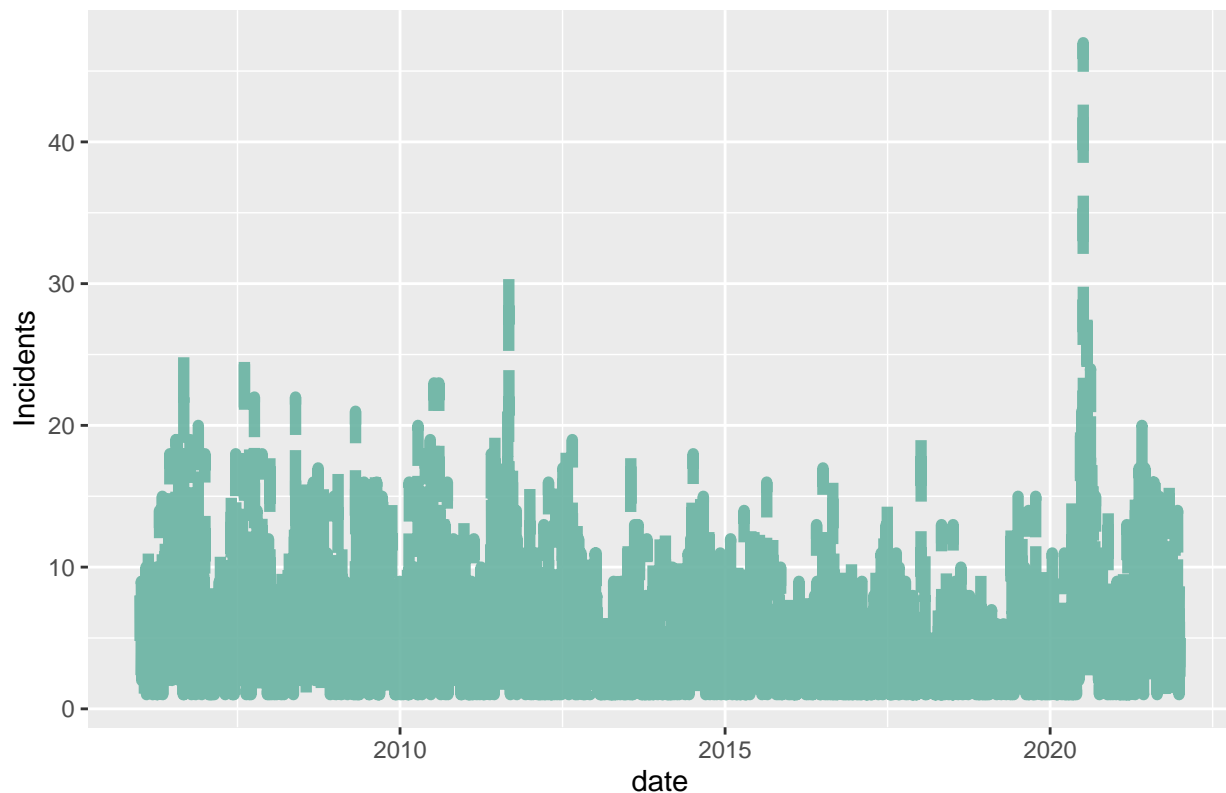
There is missing demographic data for the perpetrator, which makes sense because they may not have been caught or identified. This missing data could be omitted with the following code if that information was needed in the analysis.

```
#sum(is.na(nypd_data))  
#nypd_data <- na.omit(nypd_data)
```

This code is not going to be used because this data exploration and analysis will focus on the specific boroughs and associated incident occurrences.

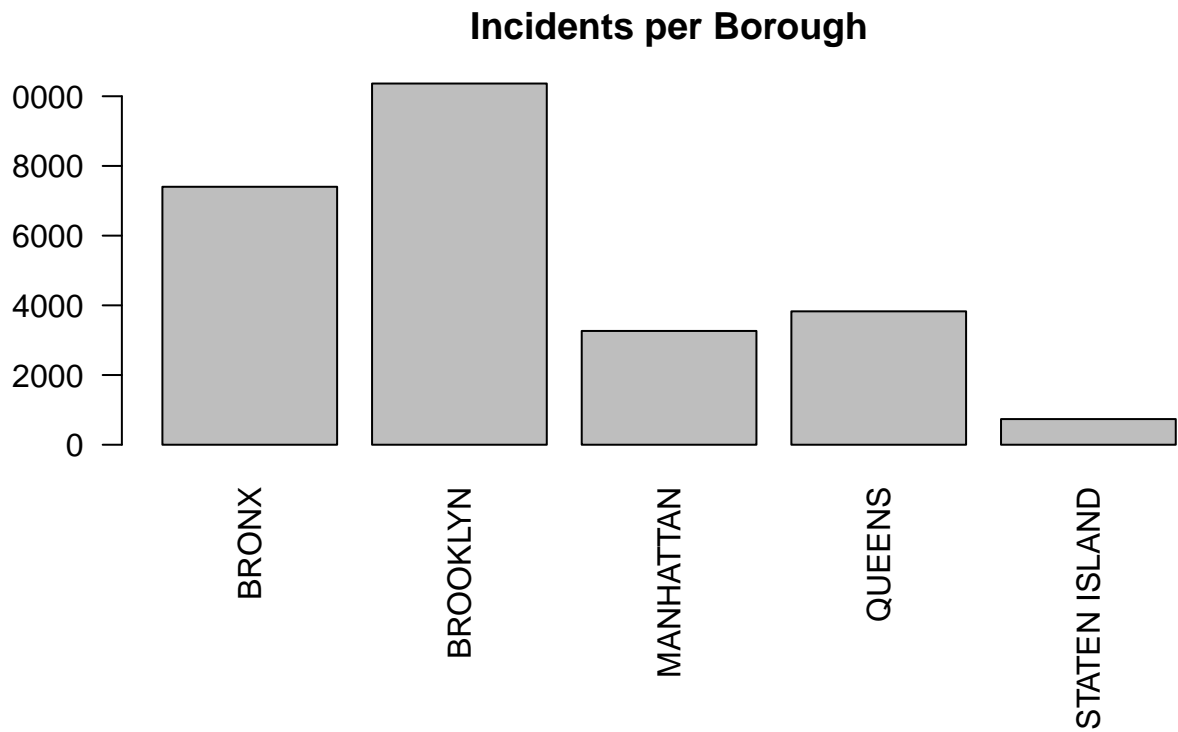
## Visualize the Data

### NYPD Shooting Incident Per Day

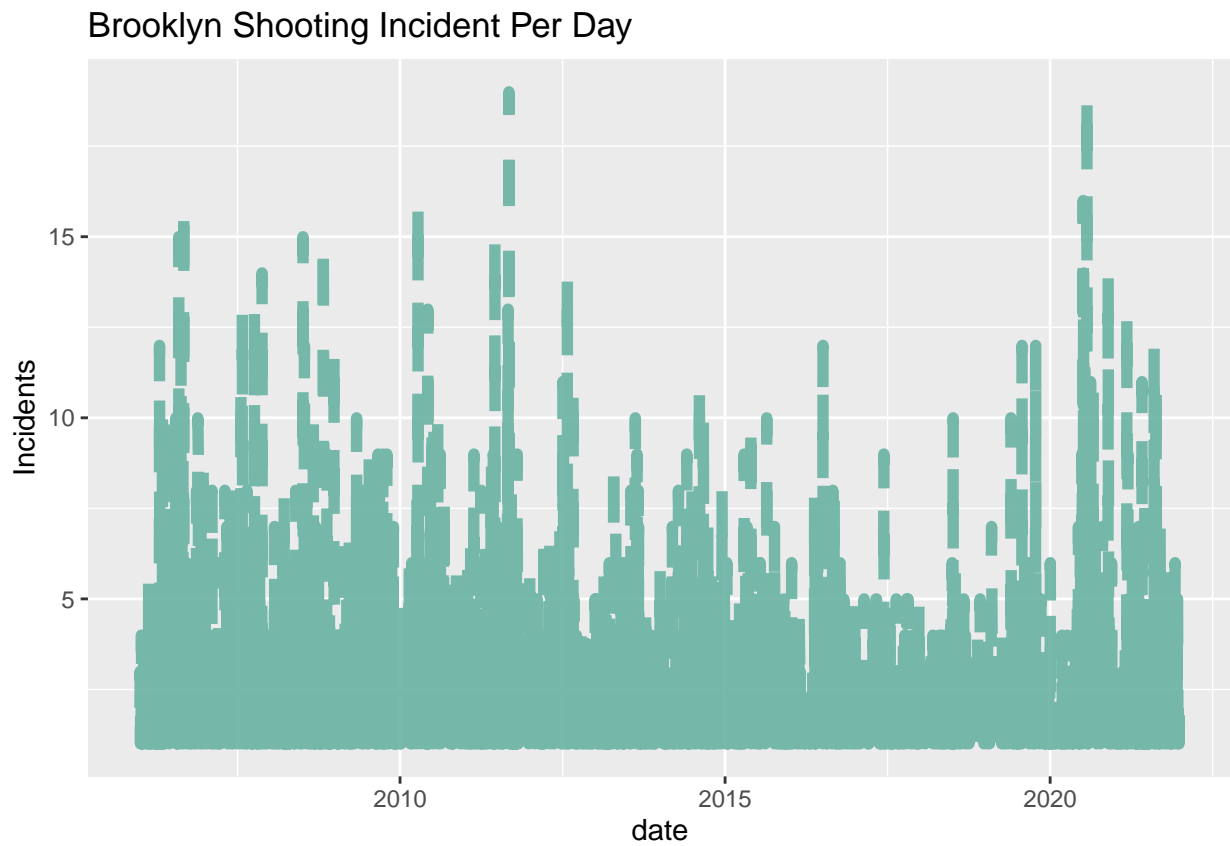


Here we can see that the daily number of incidents for NYC has not changed dramatically over time. There seems to be an overall spike in incidents sometime around the beginning of 2020.

This plot does not provide insight into where the incidents are happening, so the following box plot will be used to explore that distribution



It looks like Brooklyn has the most reported incidents, and the following chart shows how many incidents per day are occurring there.



Like the graph of the entire city, there is not a drastic increase or decrease in the number of daily occurrences

through the time period capture in this data set.

## Analyze the Data

While there is not a drastic increase or decrease in the number of daily incidents over a long period of time, there do seem to be wide variations during smaller periods of time. Because of that variation it would be interesting to see the daily average incidents occurring in each borough.

```
##   Borough      OCCUR_DATE      n
## Length:12239   Min.      :2006-01-01   Min.      : 1.000
## Class :character 1st Qu.:2009-08-04   1st Qu.: 1.000
## Mode  :character Median :2013-05-06   Median : 1.000
##                Mean  :2013-09-14   Mean   : 2.091
##                3rd Qu.:2017-09-26   3rd Qu.: 3.000
##                Max.   :2021-12-31   Max.    :19.000

## # A tibble: 5 x 2
##   Borough      daily_average
##   <chr>          <dbl>
## 1 BRONX          2.20
## 2 BROOKLYN       2.53
## 3 MANHATTAN       1.66
## 4 QUEENS          1.70
## 5 STATEN ISLAND   1.32
```

These daily averages per borough are all relatively similar. However these numbers do not take into account the size of the populations of these boroughs. It would be interesting to see where the number of incidents increase as the population density between boroughs increases. In order to successfully explore that, data on the population sizes of these boroughs at the time of the reported incidents would need to be collected. It would also be interesting to see if demographic features of the boroughs play a role in the number of reported incidents.

## Conclusions and Biases

In conclusion, the number of incidents reported in NYC has not dramatically increased or decreased since 2006. That same state holds for the specific boroughs as well, although the different boroughs do not have the same level of reported incidents. For example Brooklyn has the highest average reported incidents compared to the other four boroughs.

The analysis presented here does not include any of the demographic data, which would be the most likely aspect of this data set to introduce bias. However personal knowledge of different boroughs and stereotypes associated with them could lead to a lack of further exploration of the data. As stated above Brooklyn had the highest average number of reported incidents, and if someone had heard that Brooklyn wasn't safe because of a certain group of people who live there or any other reason, that prior knowledge might stop further analysis because the current analysis confirmed a preconceived notion. This would be inappropriate because all reasons for the higher reported incidents would not be explored, therefore making the results inaccurate.