

# NYPD Shooting Incidents

Konica Patait

January 23, 2023

## NYPD Shooting Incident Analysis

This report is based on an NYPD shooting incident that is publicly available on data.gov site and captures all the incidents reported in New York for the years 2006-2021.

The agenda of this report is to analyze on below:

- Shooting incidents trended over the years.
- Shooting incidents trend by the time during the day.
- Shooting incidents by the age of perpetrator and victim

## Load Data

Below set of lines read the **NYPD Shooting Incident** data for further analysis:

```
url_in <- "https://data.cityofnewyork.us/api/views/833y-fsy8/rows.csv"
data <- read_csv(url_in)
```

```
## 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.
```

```
spec(data)
```

```
## 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_double(),
## Y_COORD_CD = col_double(),
## Latitude = col_double(),
## Longitude = col_double(),
## Lon_Lat = col_character()
## )
```

## Clean Data

The columns that are not relevant to the reports are removed. The row that does not have required values for reporting are dropped from the data.

```
required.columns <- data %>%
  select(
    OCCUR_DATE,
    BORO,
    OCCUR_TIME,
    PERP_AGE_GROUP,
    VIC_AGE_GROUP
  )

clean.data <- required.columns %>%
  rename (
    BOROUGH = `BORO`,
  ) %>%
  mutate(
    OCCUR_DATE = mdy(OCCUR_DATE),
    PERP_AGE_GROUP = as.factor(PERP_AGE_GROUP),
    VIC_AGE_GROUP = as.factor(VIC_AGE_GROUP)
  ) %>%
  drop_na(
    PERP_AGE_GROUP,
    VIC_AGE_GROUP,
    OCCUR_DATE,
    OCCUR_TIME
  )

clean.data
```

```
## # A tibble: 16,252 x 5
##   OCCUR_DATE BOROUGH   OCCUR_TIME PERP_AGE_GROUP VIC_AGE_GROUP
##   <date>      <chr>      <time>      <fct>          <fct>
## 1 2021-07-16 BROOKLYN   22:05      45-64          25-44
## 2 2021-07-11 BROOKLYN   01:09      <18            25-44
## 3 2021-03-07 BROOKLYN   06:15      25-44          25-44
```

```
## 4 2021-07-21 MANHATTAN 00:40      25-44      25-44
## 5 2021-05-09 BRONX      02:50      25-44      25-44
## 6 2021-06-16 BRONX      23:22      25-44      25-44
## 7 2021-01-12 BROOKLYN  22:12      18-24      18-24
## 8 2021-09-04 MANHATTAN 20:18      18-24      18-24
## 9 2021-06-16 BRONX      23:22      18-24      25-44
## 10 2021-09-29 BRONX     12:50      18-24      <18
## # ... with 16,242 more rows
```

## Transform Data

The data is transformed to fetch the number of shooting incidents over the year, time of day, and age group.

```
incidents.by.year <- clean.data %>%
  mutate(YEAR = lubridate::year(OCCUR_DATE)) %>%
  group_by(YEAR) %>%
  summarize(NUM_INCIDENTS = n())

incidents.by.time <- clean.data %>%
  mutate(INTERVAL = lubridate::hour(OCCUR_TIME)) %>%
  group_by(INTERVAL) %>%
  summarize(NUM_INCIDENTS = n())

print(n = 20, x = incidents.by.year)
```

```
## # A tibble: 16 x 2
##   YEAR NUM_INCIDENTS
##   <dbl>         <int>
## 1  2006           1890
## 2  2007           1606
## 3  2008           1737
## 4  2009           1639
## 5  2010           1214
## 6  2011            992
## 7  2012            825
## 8  2013            698
## 9  2014            733
## 10 2015            766
## 11 2016            640
## 12 2017            578
## 13 2018            549
## 14 2019            545
## 15 2020            878
## 16 2021            962
```

```
print(n = 24, x = incidents.by.time)
```

```
## # A tibble: 24 x 2
##   INTERVAL NUM_INCIDENTS
##   <int>         <int>
## 1      0           1176
```

```
## 2      1      1165
## 3      2      1004
## 4      3       940
## 5      4       825
## 6      5       415
## 7      6       214
## 8      7       162
## 9      8       133
## 10     9       155
## 11     10      184
## 12     11      262
## 13     12      319
## 14     13      350
## 15     14      539
## 16     15      631
## 17     16      671
## 18     17      703
## 19     18      786
## 20     19      881
## 21     20      974
## 22     21     1214
## 23     22     1204
## 24     23     1345
```

```
incidents.by.perp.age <- clean.data %>%
  group_by(PERP_AGE_GROUP) %>%
  summarise(NUM_INCIDENTS = n()) %>%
  rename (
    AGE_GROUP = 'PERP_AGE_GROUP'
  )

incidents.by.vic.age <- clean.data %>%
  group_by(VIC_AGE_GROUP) %>%
  summarise(NUM_INCIDENTS = n()) %>%
  rename (
    AGE_GROUP = 'VIC_AGE_GROUP'
  )

incidents.by.age <- incidents.by.vic.age %>%
  right_join(incidents.by.perp.age, by="AGE_GROUP") %>%
  rename (
    NUM_VICTIMS = NUM_INCIDENTS.x,
    NUM_PERPETRATOR = NUM_INCIDENTS.y
  ) %>%
  drop_na()

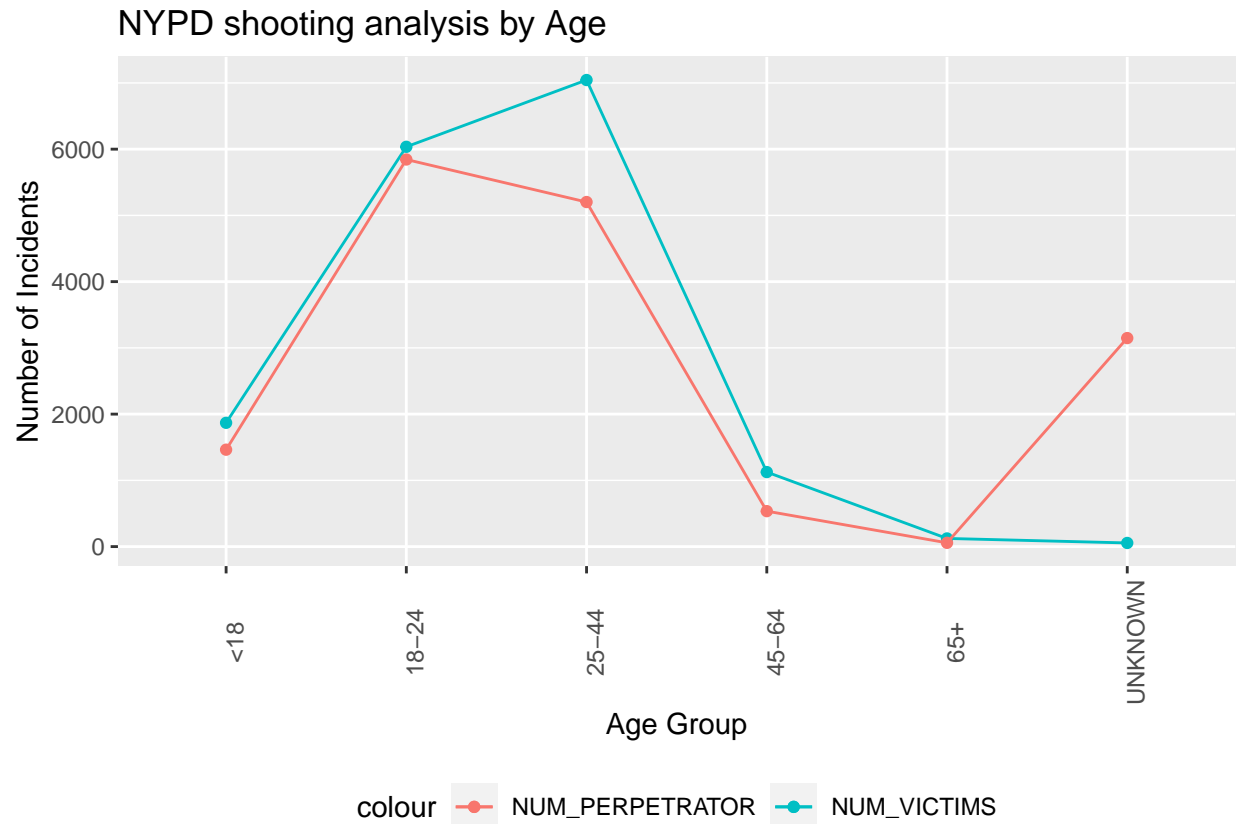
print(x = incidents.by.age)
```

```
## # A tibble: 6 x 3
##   AGE_GROUP NUM_VICTIMS NUM_PERPETRATOR
##   <fct>      <int>      <int>
## 1 <18        1869        1463
## 2 18-24      6036        5844
## 3 25-44      7044        5202
```

## 4	45-64	1125	535
## 5	65+	123	57
## 6	UNKNOWN	55	3148

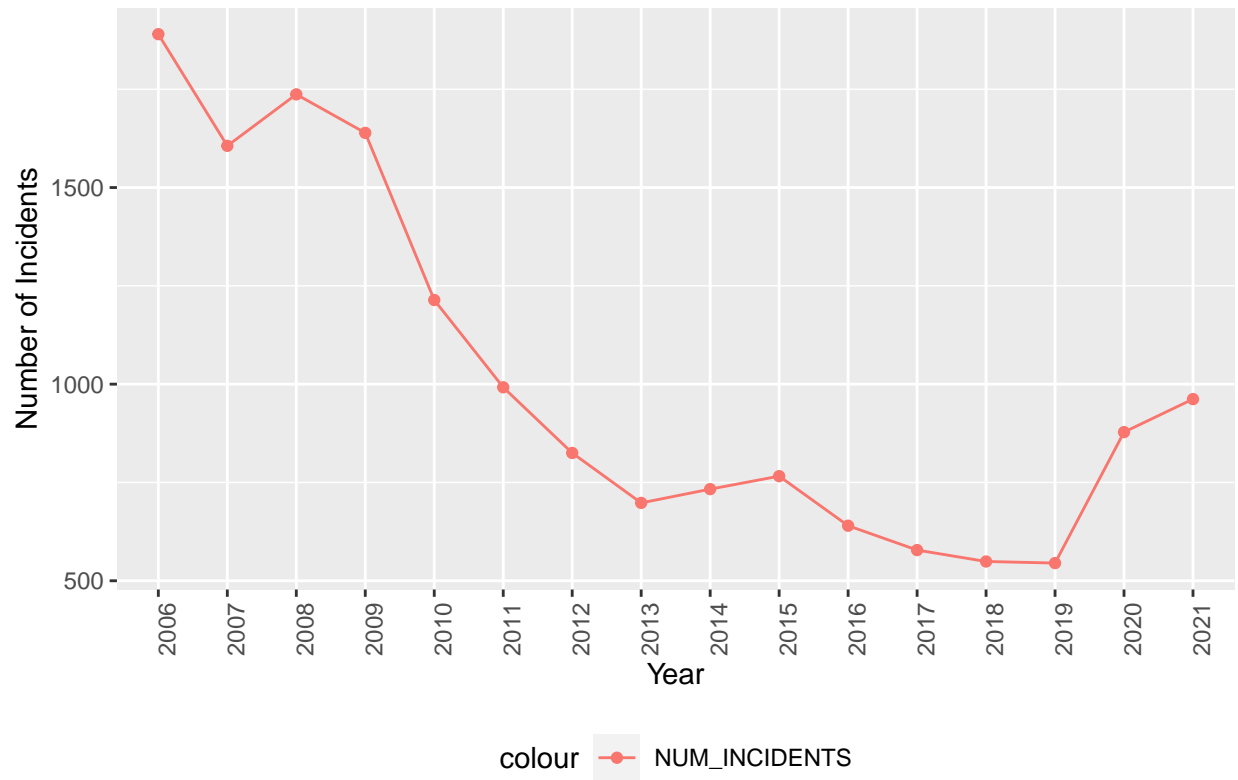
## Visualization

The below plot represents the number of shooting incidents reported by the New York Department of Police(NYPD) by the perpetrator's age and victim's age. It can be noticed that young adults commit most shooting incidents, i.e., 18-24 years of age. Based on the data, a similar trend can also be seen for victims.

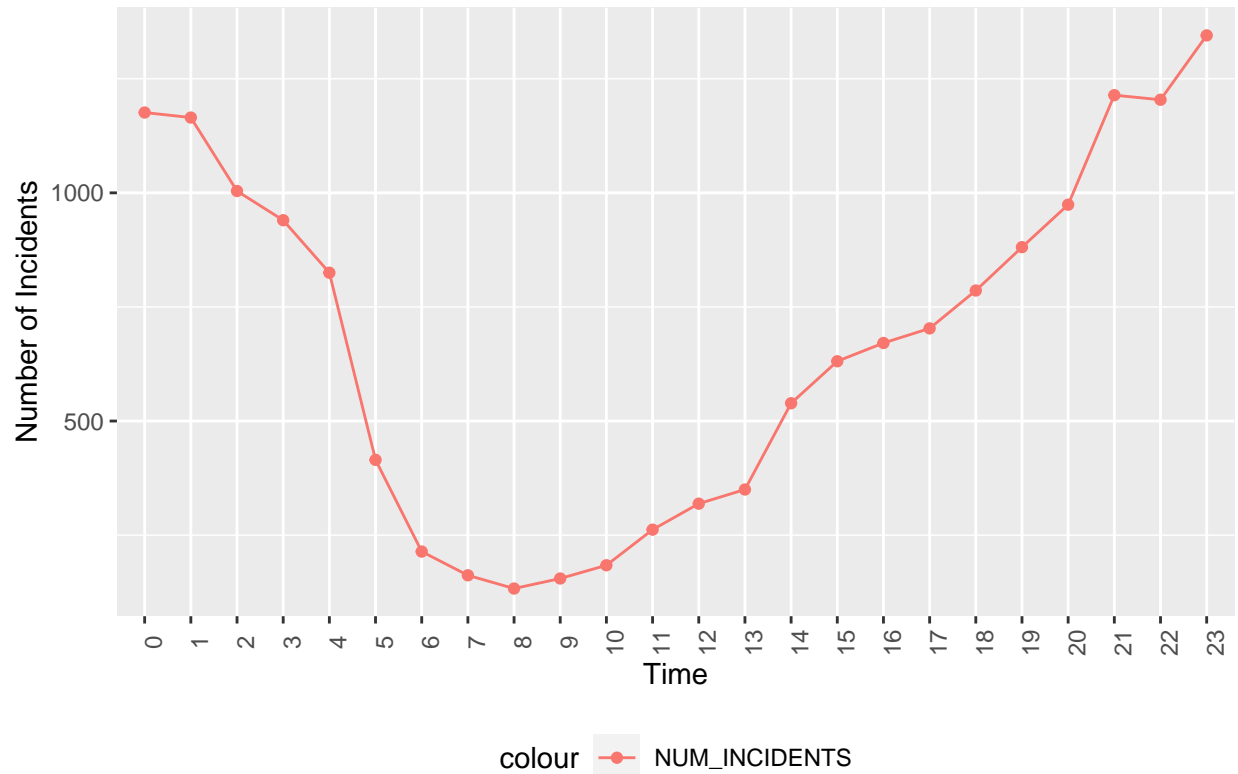


The below plots show the shooting incident trend over the years and the time during the day. It can be noticed that shooting incidents dropped from 2009 till 2019, but suddenly the number of shooting incidents rose in 2020.

NYPD shooting analysis by Years



## NYPD shooting analysis by Time



## Bias Identification

Following are the possible sources of bias in the above analysis:

- Assumption is the data captured without personal bias along with the reported date and time.
- Data contained incidents where perpetrator or victims data was missing or incorrect. This could lead to bias in the analysis comparing the incidents by the age.
- The analysis based on the time assumes that the cases are reported with minimal delay.

## Conclusion

Based on the available data, the shooting cases reported by New York Police department(NYPD) confirm that the number of shooting incidents has decreased over the years since 2009. Although, the number of shooting incidents started rising in 2020, but we don't have enough information if these are somehow related to COVID-19 or not. It can be noted that during the day the majority of the shooting incidents occurs post 4 PM and least number of incidents happen during mornings i.e., 5-11 AM. Data confirms that young adults and victims who commit the majority of shooting incidents are usually between 18-44 years old.