NYPD Copaganda - Unveiling Chronological Periodicity

Jansen Smith

2023-08-14

Purpose

This is an analysis of NYPD shooting incidents occurring from the start of 2006 to the end of 2022. The data and analysis presented herein is an effort to understand and analyze the provided data and to draw some meaningful conclusions as part of the "Data Science As A Field Course" at UC Boulder in the Summer 2 session in 2023. Due to the source of the data provided, all analysis should be considered purely conjectural. In particular, this report seeks to establish that the NYPD's reported shooting data demonstrates periodicity by the time of year.

Data Source and Description

The data was obtained from NYC's open data API and contains information on NYPD shooting incidents. The dataset includes columns like OCCUR_DATE (date of incident), to which columns such as Week_of_Year were added to facilitate analysis.

Importing required libraries

The only libraries required to run the following analysis are tidyverse and lubridate, and this code block is designed to automatically install tidyverse if the user has not already done so.

```
# Install the tidyverse package if not already installed
if (!requireNamespace("tidyverse", quietly = TRUE)) {
  install.packages("tidyverse")
}

# Load the tidyverse package
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.8
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
# Load the lubridate package for parsing dates
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

NYPD Shooting data import

The wisdom of using the NYPD as a source for crime statistics is a question best left to the reader. For more on this topic, please refer to the many historical overturned convictions due to demonstrated systemic NYPD illegality: https://abc7ny.com/convictions-overturned-brooklyn-da-eric-gonzalez-nypd-misconduct-corruption/12212052/

To disassociate this report from the worst excesses of the NYPD's misconduct, the asserted racial identities of both victims and alleged perpetrators are immediately removed from the dataset.

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

```
## Rows: 27312 Columns: 21
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (12): OCCUR_DATE, BORO, LOC_OF_OCCUR_DESC, LOC_CLASSFCTN_DESC, LOCATION...
## 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.
```

```
data = data %>%
  mutate(OCCUR_DATE = mdy(OCCUR_DATE), Week_of_Year = lubridate::week(OCCUR_DATE)) %>%
  select(-X_COORD_CD, -Y_COORD_CD, -Latitude, -Longitude, -Lon_Lat, -BORO, -LOC_OF_OCCUR_DESC, -PRECINC
spec(data)
```

```
## NULL
```

```
head(data, 20)
```

```
## # A tibble: 20 x 4
##    INCIDENT_KEY OCCUR_DATE OCCUR_TIME Week_of_Year
##           <dbl> <date>     <time>            <dbl>
##  1    228798151 2021-05-27 21:30                21
##  2    137471050 2014-06-27 17:40                26
##  3    147998800 2015-11-21 03:56                47
##  4    146837977 2015-10-09 18:30                41
```

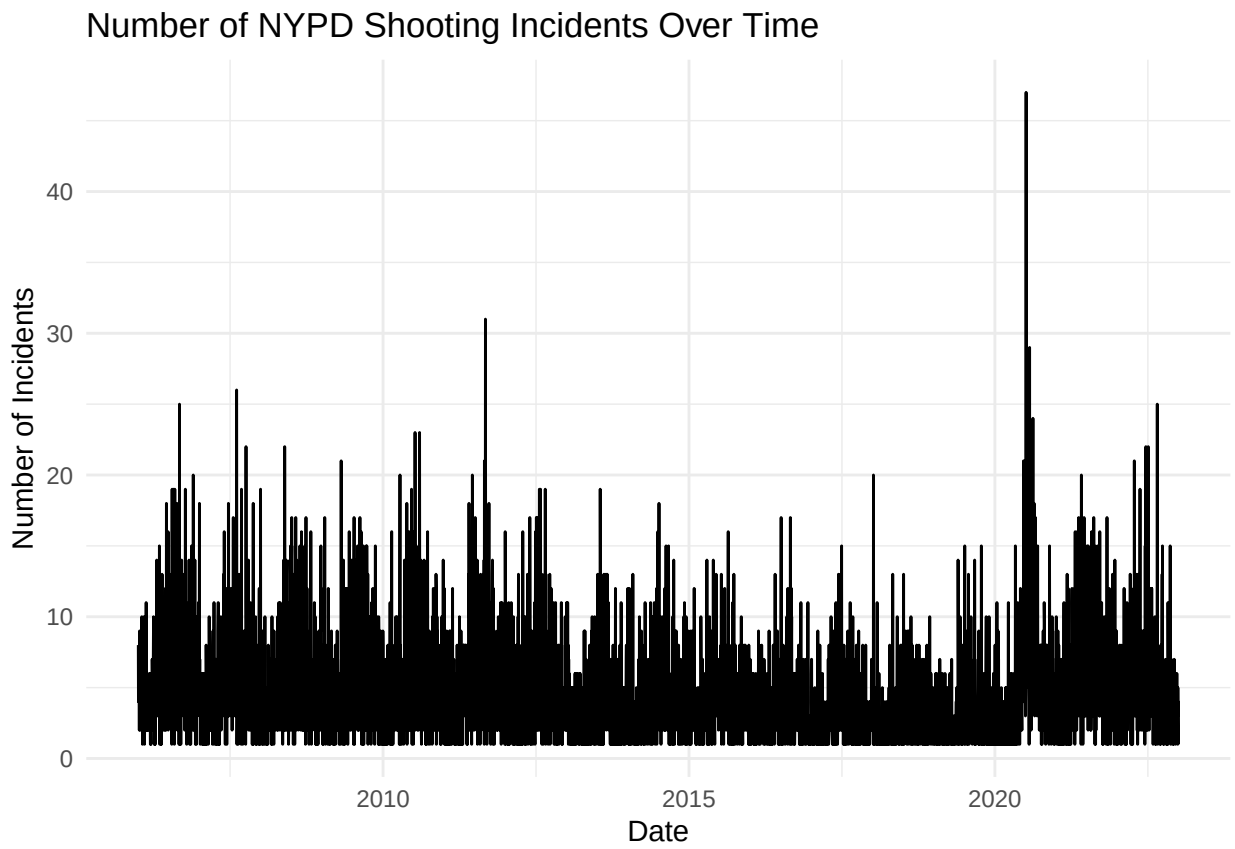
```
##  5     58921844 2009-02-19 22:58                8
##  6    219559682 2020-10-21 21:36               43
##  7     85295722 2012-06-17 22:47               25
##  8     71662474 2010-03-08 19:41               10
##  9     83002139 2012-02-05 05:45                6
## 10     86437261 2012-08-26 01:10               35
## 11     74998200 2010-10-10 03:21               41
## 12     74363835 2010-08-29 01:27               35
## 13    234756217 2021-10-09 20:17               41
## 14     78961869 2011-05-25 21:58               21
## 15     53891791 2008-11-09 20:13               45
## 16     32914993 2007-07-05 01:27               27
## 17     73879554 2010-07-27 02:22               30
## 18     87694905 2012-11-14 21:07               46
## 19    230311079 2021-07-01 02:44               26
## 20    225297069 2021-03-07 21:17               10
```

```
summary(data)
```

```
##   INCIDENT_KEY         OCCUR_DATE          OCCUR_TIME        Week_of_Year
##  Min.   :  9953245   Min.   :2006-01-01   Length:27312      Min.   : 1.00
##  1st Qu.: 63860880   1st Qu.:2009-07-18   Class1:hms        1st Qu.:18.00
##  Median : 90372218   Median :2013-04-29   Class2:difftime   Median :28.00
##  Mean   :120860536   Mean   :2014-01-06   Mode  :numeric    Mean   :27.93
##  3rd Qu.:188810230   3rd Qu.:2018-10-15                     3rd Qu.:39.00
##  Max.   :261190187   Max.   :2022-12-31                     Max.   :53.00
```

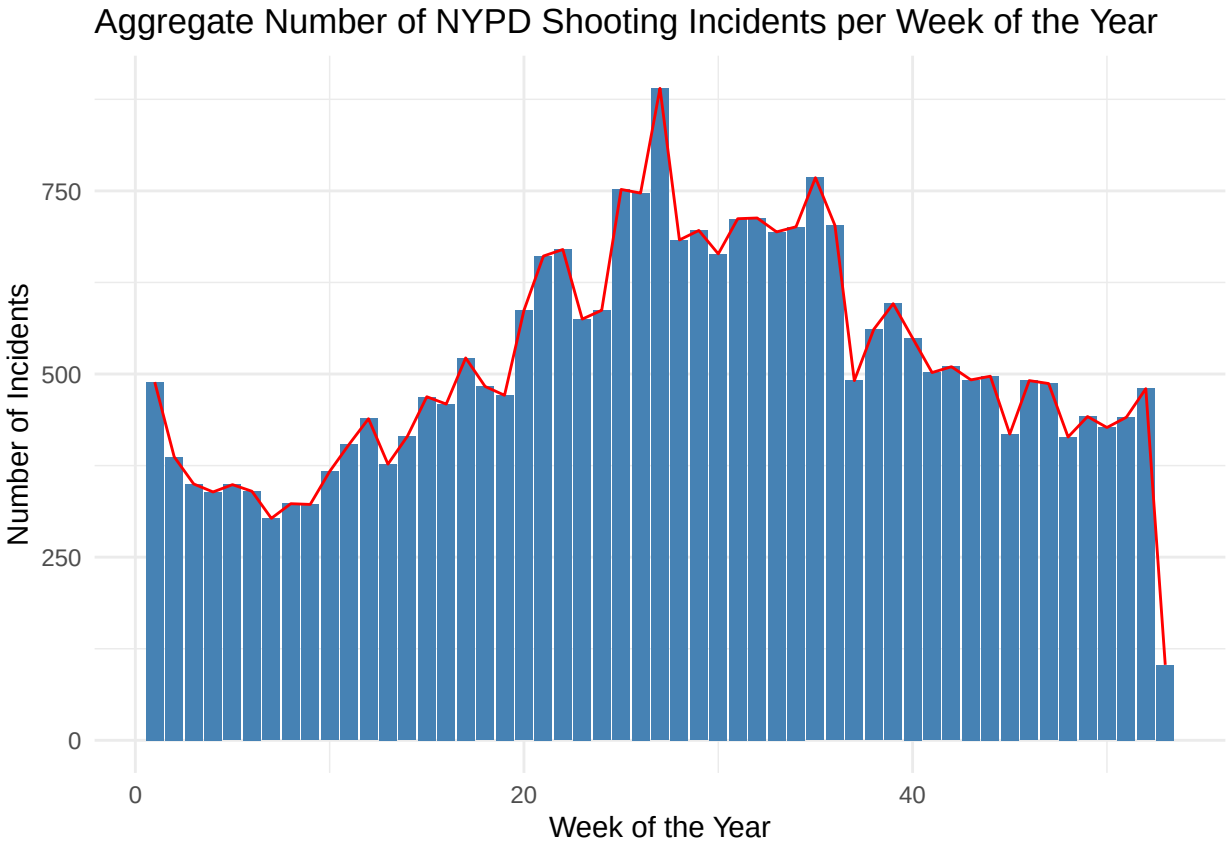
NYPD Shooting data visualization

```
data %>%
  ggplot(aes(x = OCCUR_DATE)) +
  geom_line(stat = "count") +
  labs(title = "Number of NYPD Shooting Incidents Over Time",
       x = "Date",
       y = "Number of Incidents") +
  theme_minimal()
```




```
# Plot the number of incidents per week of the year, aggregating all years
data %>%
  group_by(Week_of_Year) %>%
  summarize(Num_Incidents = n(), Median_Incidents = median(Num_Incidents)) %>%
  ggplot(aes(x = Week_of_Year, y = Num_Incidents)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  geom_line(aes(y = Median_Incidents), color = "red", linewidth = 1.5) +  # Add a red line for the medi
  labs(title = "Aggregate Number of NYPD Shooting Incidents per Week of the Year",
       x = "Week of the Year",
       y = "Number of Incidents") +
  theme_minimal()
```

```
## Warning: Ignoring unknown parameters: linewidth
```



Analysis

- As we can see from the shooting incidents over time, the data shows yearly periodicity.
- When aggregated by week, this periodicity can be seen more clearly, with significantly more incidents occurring during the summer months.
- Based on these findings, there is a potential argument for the NYPD to reevaluate its approach towards public safety and community support during the winter months, considering the demonstrated slowdown in shooting activity during that time.

Bias Assessment

- The format of the analysis was specifically chosen to avoid known sources of bias, with the notable exception of inherent bias due to trusting NYPD incident reporting sources.
- Relying on the NYPD's self-reported statistics to establish groundtruth crime data is itself inherently flawed. For example, the NYPD is likely to diminish the prelevance of crimes by the NYPD itself.
- My hope is that chronological periodicity is a topic that would be cross-cultural and thus mitigate the impacts of my own cultural bias.
- Political bias, however, should not be discounted. It is possible that the data suggests patterns that the author's mistrust of the NYPD may leave uncovered.