New York City Shooting Incidents

Seth Porter

2023-03-20

What are we looking at, and why should I care?

We will examine "shooting incident" data from the New York City Police Department.

Shooting incidents are extreme events which can end or radically change the lives of both the victim and the shooter in a single action.

Understanding their patterns is crucial to policy and intervention decisions from law enforcement, other city agencies, and non-governmental actors.

Where is the data from?

- Dataset sourced from https://catalog.data.gov/dataset/nypd-shooting-incident-data-historic, described in Footnotes (https://bit.ly/3KSLRjA)
- "List of every shooting incident that occurred in NYC going back to 2006 through the end of the previous calendar year."
- "Only valid shooting incidents resulting in an injured victim are included in this release."

Secondary data sources

In addition to the primary data, some secondary sources are used to contextualize or suggest possible causal variables:

- Temperature data from https://www.weather.gov/media/okx/Climate/CentralPark/DailyAvgTNormals.pdf
- Population data from https://www.census.gov/quickfacts/newyorkcitynewyork

Preliminary Cautions

Be cautious; this data is from a agency reporting on its own jurisdiction, and may have resulting biases or gaps in the collection, characterization, and reporting process.

Self Links

Github repo:
https://github.com/symmatree/data-science/tree/main/r

- These slides: Source (https://github.com/symmatree/data-science/raw/main/r/NYPDShootingDataSlides.Rmd) / pdf / pptx
- The full analysis document has more discussion of data tidying and methods: https://github.com/symmatree/data-science/raw/main/r/NYPDShootingData.Rmd (source) / pdf

Two Major Approaches

We will consider

- patterns-over-time and seasonality
- demographics of victims and perpetrators

in hopes of understanding the patterns and impacts of these incidents.

Data Analysis - Temporal trends

Our first layer of analysis will focus on **when** shootings occur, considering both the overall trend and various scales of periodic behavior.

Long-term trends

Reported shooting incidents dropped steadily, then rose sharply in 2020 (presumably pandemic-related, though possibly a reporting change). This is such a sharp reversal that it dominates other trends. The rest of this analysis will normalize by annual incident count to reveal patterns masked by this macro-trend.
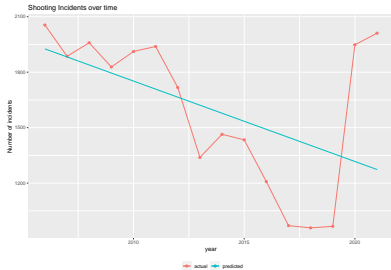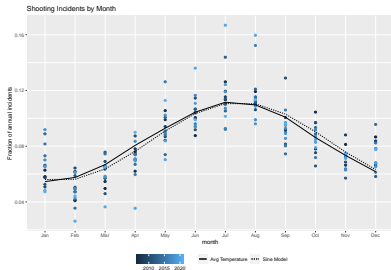
Seasonality: Month of year

A shifted and vertically-scaled sine wave fits the monthly data quite nicely.

Average NYC temperatures correlate strongly with both this model and the observed incident counts.

This suggests (though does not prove!) that temperature-driven behavior patterns are a major driver of the overall level of shootings.

Seasonality: Day of week

Weekly rhythms have a strong influence on shooting incidents; a scaled and phase-shifted sine wave fits quite nicely (RMSE ~2 percentage points of annual incidents).

There is no immediately-obvious explanation of this pattern, other than the incident rates being higher on Fridays and weekends when more social interaction is likely to occur.
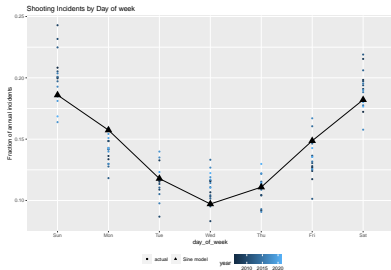
Seasonality: Hour of day

While not a simple sinusoidal, there is a repeated cycle throughout the day, with incidents rarest in the morning hours and most common at night until 4am (which is when bars close in NYC).
This aligns with the weekly pattern: shootings happen less during typical working hours and days.
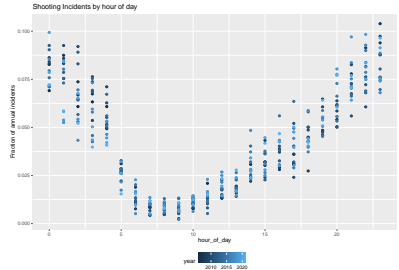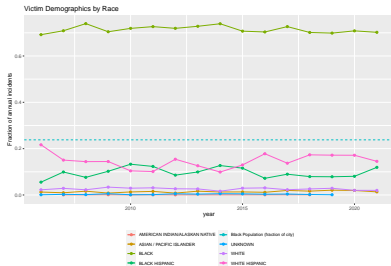
Conclusions - Temporal Conclusions

Despite sharp changes in overall incident counts, there are strong seasonal patterns at multiple scales: over the course of a day, a week, and a year
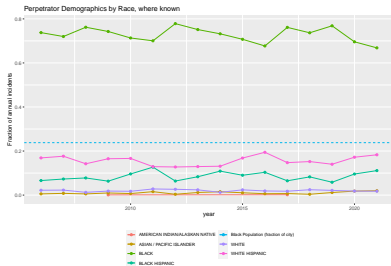
Demographics: Victim, By Race

The victim's race is overwhelmingly coded as Black, disproportionate to the Black share of the overall city population.

Demographics: Perpetrator, by Race

The same holds true, in almost the same proportions, for the perpetrator's race (where known), again disproportionately so.

Perpetrator Demographics by Race, where known

Younger Black males

We can narrow the data to consider only the combined fractions for victims and perpetrators coded as Black males in the 18-24 or 25-44 age groups.

More than half of all shooting incidents involve someone in this demographic as victim, and roughly the same proportion as perpetrator (where known).

The *entire* Black population of NYC is just under a quarter of the city, so the corresponding line is an overestimate of the fraction of Black males in this age group.

Shooting Incidents, Black males age 18–44
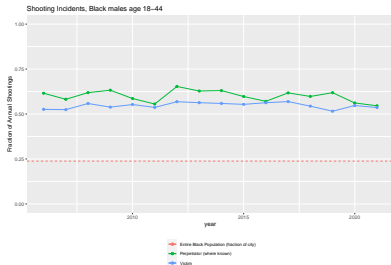
Conclusions - Demographic Conclusions

The impact of shootings (both as victim and perpetrator) falls wildly disproportionately on individuals reported as Black males aged 18-44.

Any strategy or intervention must consider both these temporal and demographic factors or it will be mis-targeted or inefficient.

A Caution

The demographic concentration of shootings should not be read as *causal* but rather as descriptive and correlated. Demographics are correlated with many other factors including socio-economic status, education, job opportunities, geographic location, historical deprivation etc., which are more likely to be actual causation.

Error and Bias

Finally, we consider the many ways this analysis might be wrong or misleading.

Error and Bias: Data Source Motivations

The underlying data comes from the NYPD. They are not a neutral party: this data may be used to demonstrate their effectiveness, to argue for more resources in certain areas, or to support theories of crime and victimhood.

Any of these would be a reason to bias the data collection and reporting:

- Skewing demographic data toward desirable patterns
- More subtly, controlling what is reported as a shooting incident and what is absent altogether

Error and Bias: Analysis

Beyond the data, the analysis may be flawed, biased or misleading.

- The focus on proportional analysis masks the changing absolute numbers of incidents, which may be a better measure of "cost" in many dimensions
- There may be important subtleties in who reports data and how, unknown to the analyst. Demographic fields are subjective and vulnerable to misinterpretation.

That said, the patterns reported *seem* to be so strong as to resist most small-scale biases or misinterpretations!