

City University of New York (CUNY)

## CUNY Academic Works

---

Dissertations, Theses, and Capstone Projects

CUNY Graduate Center

---

6-2023

### Phantom Shootings

Allan Ambris

*The Graduate Center, City University of New York*

[How does access to this work benefit you? Let us know!](#)

More information about this work at: [https://academicworks.cuny.edu/gc\\_etds/5410](https://academicworks.cuny.edu/gc_etds/5410)

Discover additional works at: <https://academicworks.cuny.edu>

---

This work is made publicly available by the City University of New York (CUNY).

Contact: [AcademicWorks@cuny.edu](mailto:AcademicWorks@cuny.edu)

PHANTOM SHOOTINGS

by

ALLAN ARTHUR AMBRIS

A master's capstone project submitted to the Graduate Faculty in Data Analysis and Visualization  
in partial fulfillment of the requirements for the degree of Master of Science,  
The City University of New York

2023

© 2023

ALLAN ARTHUR AMBRIS

All Rights Reserved

APPROVAL

Phantom Shootings

by

Allan Arthur Ambris

This manuscript has been read and accepted for the Graduate Faculty in Data Analysis and Visualization in satisfaction of the capstone project requirement for the degree of Master of Science.

Approved: April 2023

Michelle A. McSweeney, Advisor

Matthew Gold, Executive Officer

THE CITY UNIVERSITY OF NEW YORK

## ABSTRACT

### Phantom Shootings

#### A Critique of NYC Open Data's NYPD Shooting Incident Data

by

Allan Arthur Ambris

Advisor: Dr. Michelle A. McSweeney, PhD

This capstone is a website designed to critique NYC Open Data reporting with respect to shootings through a series of visualizations and discoveries. The NYPD Shooting Incidents datasets (Historic and Year to Date) introduce themselves to the user by claiming to be a “list of every shooting incident that occurred in NYC.” An obscured paragraph can be revealed if the user clicks “more”. Reading it, they might be intrigued to scroll down the page to download a PDF of footnotes. Then, if the user reads both pages, they eventually learn that the dataset is not every shooting incident. It is a subset of a subset.

After understanding the supporting materials, there are still undisclosed truths. My exploration of the data revealed that a single victim may be represented across multiple entries. Additionally, multiple victims may be represented by a single entry. It is impossible to determine if the relationship between records and individuals is one-to-one, one-to-many, many-to-one, or many-to-many. As such, any analysis of this dataset which counts people should be avoided. These truths are not available in any of the documentation.

This project was inspired by my own inquisitiveness and prior attempts to analyze shooting data. I was interested in seeing if the data reflects or contradicts my assumptions and experiences. This capstone can resonate with those who have experienced the effects of gun violence in any way, plan on joining the conversation, or want to reflect internally. It provides insight into how/why simple questions may have frustratingly convoluted answers.

I also intend for my audience to include those who are affected by the intricacies of data ethics academically, professionally, and socially. This project provides examples which either introduce or reinforce the need to be cautious when using data. It also serves as a reminder for providers to be forthcoming and transparent when making their own data collection available.

## ACKNOWLEDGEMENTS

Thank you, God.

To my family,

You enabled me to turn mistakes into lessons and survival into evolution. You are the foundation that refused to crack when I felt like crumbling. I am still standing because you've held me up. I am standing tall because you've built me up.

To my neighbors,

I cannot think about the future without hearing your words from the past. May this degree serve as a confirmation that echoes your affirmations from long ago.

To my friends,

Thank you for your patience, anticipation, and reassurance. I'm fortunate to have you with me. I am blessed to know that this celebration will be a recap, not a reunion.

To my coworkers,

Thank you for understanding my goals and accepting the impact of this journey. Your energy fueled me before I could ever run on empty.

To the faculty, staff, and my colleagues,

Thank you for this opportunity; for your professional and academic guidance, input, and feedback; for learning with me; for letting me learn from you; for sharing yourselves.

To my advisor, Dr. Michelle A. McSweeney, PhD,

You've been a steady source of encouragement from the beginning. You saw my vision and helped me keep focused. You connected with my story and empowered me to share it with others. Thank you for helping transform this project from words and code into a voice with a message.

# TABLE OF CONTENTS

<b>Abstract</b> .....	iv
<b>Acknowledgements</b> .....	v
<b>Table of Contents</b> .....	vi
<b>Digital Manifest</b> .....	vii
<b>A Note on Technical Specifications</b> .....	viii
<b>Narrative</b> .....	1
Project Description .....	1
Background .....	2
Structure .....	5
Design Choices .....	7
Data Creation .....	9
Visual Design .....	10
Future Iterations .....	12
Evaluation .....	14
<b>Appendices</b> .....	16
List of Variables .....	16
Glossary of Functions .....	16
Geospatial Transformations .....	17
<b>Select Bibliography</b> .....	18

## DIGITAL MANIFEST

- I. Capstone Whitepaper (PDF)
- II. Project Website
  - a. <https://NumberSequence.github.io/PhantomShootings/>
- III. Code and other deliverables
  - a. Zip file containing contents of GitHub repository at time of deposit.
  - b. PDF containing screenshots of Tableau dashboard at time of deposit.



## A NOTE ON TECHNICAL SPECIFICATIONS

This capstone is a webpage which relies on its internal HTML, CSS, JavaScript, images, and mp4 videos of news segments. Tools used include Leaflet.js, SliderControl.js, Cal-Heatmap.js, and Tableau Public. The webpage is intended to be viewed on a full-size landscape display with a movable cursor. The user will be presented with a loading screen while the page's contents are positioned. Once the page loads, this screen is removed.

This page was created and tested using Virtual Studio Code and the Live Server plugin. It is served on GitHub Pages. This resources and supporting references are stored in a GitHub repository located at <https://github.com/NumberSequence/PhantomShootings/>. Any additional changes to this project can be found there. The webpage can be viewed at <https://NumberSequence.github.io/PhantomShootings/>.

## NARRATIVE

### *Project Description*

My capstone is a critique of NYC Open Data's NYPD Shooting Incident Data (Historic and Year to Date) from a Data Feminism framework (D'Ignazio and Klein). I open by arguing that the dataset repeatedly misrepresents and contradicts itself. I conclude by declaring that by overstating its purpose, content, and influence, it can be classified as Big Dick Data (151-152).

It should be noted that although this project is critical of how the dataset presents itself, NYPD Shooting Incident Data does contribute to the movement of making data accessible. It offers itself as a widely accessible resource at no cost to the user. However, this project shows that it shares the same fundamental flaws laid out in Data Feminism regarding open data:

...governments have prioritized the "opening up" part of open data...but lack the capacity to provide any context about the data's provenance, let alone documentation that would allow the data to be made accessible and usable by the general public. (155)

This project attempts to address these concerns.

The dataset's preview declares that it is a "list of every shooting incident that occurred in NYC going back to 2006" ("Dataset Description"). It reasserts this claim before revealing that it is actually a list of shooting victims ("Dataset Footnotes" 1). This stance is softened in the penultimate footnote declares that the dataset is a list of a subset of shooting victims ("Dataset Footnotes" 2). In the end, what's provided is a dataset which may lead users to make inaccurate conclusions even after understanding the documentation's previous claims.

I present evidence showing that dataset is far removed from both reality and its own assertions. This project also introduces an undisclosed caveat which produces a significant miscalculation if left unaccounted. Bringing these findings together shows that being familiar with the supporting documentation's nuances does not prevent users from an unintentional misrepresentation of facts.

These findings are available at <https://NumberSequence.github.io/PhantomShootings/>. There, I walk through the discovery process, providing examples and visualizations of the discrepancies and their consequences. Videos of news segments serve as contextual references while visualizations created using JavaScript plugins/add-ons and Tableau produce the maps, dashboard, and heatmap present throughout the page. External data manipulations were performed in Excel and QGIS.

### *Background*

I began interacting with this dataset in my first course at The Graduate Center, Introduction to GIS. Seeing an opportunity to combine two areas of focus into one cohesive statement, I transformed a static poster created in QGIS to an interactive map using techniques learned in Interactive Data Visualization. During a functionality check of my final projects, I noticed a disproportionate number of even values. After troubleshooting my code, I realized that the fault was with the dataset. I didn't have time to hypothesize and test what could have caused the issue, so I opted to limit my findings and present a visualization of locations, dates, and times. I felt lucky to make this discovery, but also frustrated that I didn't notice it sooner.

I attempted to use this dataset again in *Working with Data: Fundamentals*, but with a more critical eye. I quickly realized that the dataset didn't account for many of the incidents I was interested in and therefore wouldn't be able to create a truthful visualization to express my position. I began looking at my own techniques and wondered whether the assumptions I made were due to excitement, inexperience, or if I had truly been misguided by trusting the data. I read *Data Feminism* (D'Ignazio and Klein, 2020) that semester and felt relieved that there was already a movement which put words behind the thoughts that I've been having about responsible data practices.

In preparation for this capstone, I combed through the dataset's documentation and descriptions. I validated my theory on why there were so many even values in my initial project by isolating incidents that had more than one murder classification associated with it. For each incident, the count of each `STATISTICAL_MURDER_FLAG` (a binary value) was the same. I then checked if the demographics provided for each true value had a matching false value by separating and sorting each category. After eliminating the murder classification, they matched each other perfectly.

Working backwards, I checked the footnotes again for any mention of this phenomenon with no success. I then went through the dataset's description which reminded me of my first experience with it. Specifically, I remembered the panic of a last-minute pivot, creating a new summary, and re-coding my project to ensure visualizations reflected the findings. I kept focusing on the introduction that inspired my confidence and decided to share my journey, using the initial close call as a starting point to my culmination project.

With *Data Feminism* in mind, I took issue with how much the dataset's introduction varied from what the data is. The absence of cautious keywords such as "reported", "available",

or “documented” distanced the dataset from reality. The footnotes were alluded to as “additional” (not required) information. I envisioned showing the potential impact of taking the dataset at face value. What I found was that even with due diligence, pitfalls remain.

Like my first attempt to use this dataset in a project, I made another late discovery. After noticing a post-analysis mass shooting location seemed out of place (the implications of such a statement are not lost on me), I searched for stories regarding the incident. An article showed that there were 2 victims (Fitzsimmons). It did not truly qualify as a mass shooting, leading me to explore the data further.

I investigated whether my previous theory (record changes may be appended instead of updated) applies to fields other than murder classifications. This resulted in a series of discoveries, each more damaging than the last. Ultimately, I found that (according to the data) there was never an incident where a perpetrator shot more than 1 person of a specific demographic. In essence, victims and perpetrators were not considered to be individuals, they were a combination of age, sex, and race. Each only appears once per incident. With this, I was able to conclude that NYPD Shooting Incident Data should not be used to perform any analysis that relies on / reports the number of people involved in a shooting. Any attempt to do so is subject to failure.

The findings took this capstone beyond my anticipations. Since these discoveries all stemmed from interactions with visualizations, it also validates my methodology. I likely would not have observed nor been inclined to investigate the data at a deeper level without observing it through Data Visualization.

### *Structure*

I began by rebuilding the map that brought the first discrepancy to my attention. This time, I chose to show mass shooting locations both before and after analysis. I decided against letting the audience control which view they received for several reasons. Primarily, this interface could be bypassed without any interaction, making the analysis and findings irrelevant to the user. Alternatively, a “breathing” map can deliver the message and tell a story independent of the user actions. I considered making this an animated gif, but doing so would not provide inquisitive members of the audience the opportunity to explore the data early on. It would also remove tooltips and in my opinion reduce the legitimacy of my argument by not presenting the findings/reasons why a particular incident disappears or stays to those who felt the need to explore.

Additionally, since the audience is still near the top of the page, I didn’t want them to work for my evidence. To make it as effortless as possible, I included videos to support the validity and importance of my argument as opposed to posting links to their articles. I debated describing the first video to add depth by pointing out people other than the shooter in the video. These were the person seen entering the adjacent building before the shooting occurred, the people on the near sidewalk whom the shooter appears to aim at in his final shot(s), the person across the street sitting on the stairs, and their neighbor who appears to drop a white bag.

While additional commentary may have added some emotion, it was my emotion and not that of the audience. Although it’s my hope that the people in the video do not go unnoticed by some of the audience, I decided not to inject my observations and left the user to create their own emotional connection and response. The 2<sup>nd</sup> video was a powerful example which I believe achieved that effect.

Although the map displays the discrepancies between the pre/post analysis data, the Tableau dashboard gives the audience more detailed visualizations of the findings. My thought was that seeing individual bars decrease by such a large percentage is telling. Unlike the pulsing map markers, I chose not to animate these transitions. Their sudden change is an attempt to grab attention. As if to say “wrong!” and not “well... not quite”. I included changing titles for the visualizations to suggest that although having a longer title may be wordy, it also makes the dashboard more accurate. The goal was to create a connection between these changes and the word choices made in the dataset’s description.

I placed the final discovery just before my conclusion utilizing a third news segment to provide background. The video was purposefully trimmed so it opened with the words “two teenaged boys are shot” and concluded with “injuring two teenagers”. The intent was to create an expectation for the incident to have two entries.

The final statement was made to contrast the dataset description’s final sentence which did not provide a strong warning about its limitations. With this, my capstone’s final words also serve as the title of the heatmap while the background comes into view as an interactive visualization. I hope users who interact with this heatmap remember that the data is not complete while continuing to follow their curiosity. Again, I did not want to coax the audience, so I ensured the heatmap could be viewed in its entirety with nothing else on the screen; leaving the user alone with the data and hopefully the knowledge that it’s not what it claims to be.

## *Design Choices*

The size and position of page elements are mostly a function of the viewport's width and height. A blue color theme was chosen to avoid creating an off-putting, intimidating, scary, or insensitive space. The horizontal gradient borders framed the website and softened the tones in the heatmap. Additionally, they obscure the y-axis text and the beginning/end of year rugged edges, helping hide the fact that the background is actually a relevant data visualization. Since the shades in the heatmap can vary drastically, a background was applied to paragraph elements. These stylings were performed in CSS.

To avoid revealing the heatmap/background before the concealing elements were in place, a loading screen was used. It foreshadows the "Description, Details, and Data" cycle that would be the foundation of the upcoming argument.

I wanted to make this website easy to read, particularly in the beginning. I avoided placing any material towards the edges of the screen by keeping most content contained to the center 60% of the page. Large text in short segments served as reminder for me to remain concise and assure the audience that I'll get to the point quickly. I took this a step further by ensuring there was always a break in the text no matter where the user happened to be on the page. Depending on the user's screen resolution, the "Description Contradicted by the Details" section may create the wall of text that I hoped to avoid. Placing portions of the recycling image here helped avoid that outcome. These arrows were modified in CSS to change their appearance from green to blue.

I added rotating text around the final section's arrows since the symbolism of these images may not have been clear to the audience. I felt that including it did not take away from



the message nor did it appear overly repetitive. Its presence ensured the audience was aware that the individual discoveries/arrows combine to create a cycle of misinterpretation. Knowing that the text likely wouldn't be modified after implementation, I avoided creating this with HTML. An animated gif proved to be simpler and just as effective but wouldn't allow me to readily modify the rotational speed. The solution was to create a transparent image whose rotation would be controlled using CSS.

Since the words appear in a circle centered around a triangular shape, to create a sense of symmetry, the length of the phrases needed to be as close to one another as possible. I opted to phrase this as "Data misrepresented by the Description" as opposed to "Description misrepresents Data." The latter created an unappealing set of gaps between words and a broken reading cadence.

Hoverable references were chosen using words that don't appear to be bolded for emphasis alone, hopefully suggesting that the user can interact with it. I debated including more hoverable references to shootings that don't appear in the dataset. These would be attached to the word "nothing" but ran the risk of losing viewers with redundancy. I decided against it as the section was already strongly worded and the prior examples were powerful enough to be remembered. Originally located near the end of the page, this section was repurposed as an introduction. It is purposefully the only time I make any mention of myself on the page. Relocating it allows me to own my position from the outset.

The grayscale basemap was chosen to align with the page's color theme. Marker opacity was set to 70% since some incident locations overlap. This created a gradient effect as clusters of shootings appeared, adding depth to the visualization. At first, they were sized relative to the map. This works initially but at higher zoom levels, markers become tiny dots. This made them

harder to identify and interact with. As a solution, markers were given an absolute radius.

Shootings whose post-analysis victim count was less than 4 were assigned to one CSS class. All markers with this classification were styled in CSS to create the pulsing effect.

This decision came after struggling to create an interface that would allow the user to remove these invalid incidents manually. After realizing that I could assign a class to markers as they were created, I styled it to pulse. I chose this approach rather than have the user change the styling since it could represent the contrast between pre and post discovery mass shootings with zero user action. JavaScript was used to prevent inadvertent zooming while users scroll past the map. The map was purposefully (and poetically) centered at the midpoint of the Brooklyn Bridge. It's a nod to my connection to The Graduate Center, but also the literal bridge that separates the New York that's seen by some and the New York that's lived by others, including myself.

### *Data Creation*

I created the map's data source by merging the historical and year to date datasets, keeping the INCIDENT\_KEY, OCCUR\_DATE, STATISTICAL\_MURDER\_FLAG, Latitude, and Longitude. In Excel, I created columns (Mtrue / Mfalse) which counted each murder classification by incident key. Incidents where  $M_{true} + M_{false} < 4$  were removed. Any duplicate entries were then deleted.

To ensure overlapping incidents were individually identifiable, I modified locations that appeared more than once. These shifts moved the marker approximately 100ft from the previously rounded location by adding/subtracting .0004 to both the latitude and longitude. This

allows up to 5 unique incidents to be observed in one relatively small location (one original location followed by SW, SE, NW, and NE shifts). A detailed explanation of this process is provided in the appendices. There are no concerns with performing this transformation as the dataset acknowledges that coordinates are not intended to provide the true locations of incidents (“Data Footnotes” 1). This data was converted to GeoJSON using QGIS and used as the data source for the Leaflet map and slider.

The dashboard’s data source is a merge of the Historic and Year to Data datasets. For consistency between the datasets, Y/N values for STATISICAL\_MURDER\_FLAG were converted to text values of TRUE/FALSE. All text values of “(null)” removed so they could be read as true null values. Four parameters were created to change the titles and time ranges of the graph. The date slider is a master data source filter created from the month of the OCCUR\_DATE.

The heatmap’s data source is a merge of the datasets. It contains INCIDENT\_KEY and OCCUR\_DATE without duplicates.

### *Visual Design*

The Tableau dashboard is a collection of 3 sets of stacked dual axis bar graphs. Orange and blue shading represent whether the incident was / was not classified as murder respectively. Each has a synchronized y-axis and an x-axis determined by the user’s date and time selections. They can switch between “Claims” and “Reality” to see the values change. If they select “Reality”, a gray bar is visible behind the blue and orange stacked bars. This provides a reference to the pre-analysis values. All bars have tooltips which provide the user with the specific

category's count. From top to bottom, the bar charts represent the count of victims, mass shooting victims, and mass shooting incidents.

I included this dashboard to give the audience an opportunity to explore the data after the discoveries are made. The goal behind this design choice was to provide an extra level of detail if they felt compelled to explore further after learning more about the dataset's discrepancies. Ideally, the user now understands the data and the concerns raised after minimal time and effort. This is an opportunity to transform users from a member of the audience to a stakeholder. I chose to introduce the dashboard instead of instructing the user how to use it to avoid applying any type of pressure or coaxing to the audience. Should they choose to interact with the dashboard, it should be out of organic curiosity and interest.

The default locations of the worksheets' titles were unappealing and ended up being an inefficient use of limited vertical space. Three additional worksheets were created with the sole purpose of displaying titles and placed alongside their corresponding charts. This solution made use of the ample horizontal space and alleviated vertical congestion.

JavaScript libraries were crucial to creating this website. Cal-Heatmap.js was used to create the calendar heatmap but required customization in its implementation/creation as well as additional CSS styling. Options and parameters were set after calculating the number of columns and rows needed to have the element fill the entire viewport height and appear horizontally centered. Gutters were set to 0 to create a visualization that more closely resembled a mosaic for the background. The heatmap's x-values were the date and the y-values were grouped to output the count of incidents. While the same visualization could be created using only the dates (counting the number of times a date appears as the number of incidents), it would have complicated the ability to troubleshoot/audit the heatmap's results.

The Leaflet.js library was used to create the map. Styling options were set to determine marker size, color, and opacity. Options were set with the user experience in mind as well as the geospatial range of the data. Selections defined zoom increments, set a minimum zoom limit, and reduced the likelihood of unintended interactions caused by scrolling over the map. Details regarding this visualization's options, variables, and dependencies are provided in the appendices.

SliderControl.js was used to create the date filter used in the Leaflet map. The slider and accompanying labels were styled in CSS but needed to be moved so they did not obscure the map. Attempts to do this using CSS had varying degrees of success but also unintended consequences. Positioning elements to account for the numerous shifts became inefficient and this idea was eventually scrapped. An element created in HTML and styled in CSS served as a placeholder for the slider elements. Once the slider is created, JavaScript relocates it into this placeholder. The zoom controls located in the top-left corner were not removed. While they are not required for navigation, they occupied minimal space and serve as another cue to the audience that the map is interactive.

JavaScript was also used to resize videos, hide/show the loading screen and main page, and adjust page contents if the window is resized. It also resets the user to the top of page if the page is refreshed. A list of functions and their purpose are provided in the appendices.

### *Future Iterations*

This project would not be possible without all three areas of study; Data Analysis, Data Visualization, and Data Studies. Data Analysis provided the tools to not only manipulate the data

into workable formats, but to test multiple theories and reach the conclusions laid out in this project. Data visualization served as the medium through which these findings were presented, creating a project capable of reaching audiences of various interests. Data Studies provided the desire and reassurance to contribute to the evolution of the field's scientific and societal responsibilities.

This capstone is the result of my journey through coursework, conversations, and interactions with my professors and colleagues. Their feedback, critiques, motivation, and inspiration have all played a role in this project reaching its current state. Nonetheless, future iterations using their influence are possible.

The heatmap relies on an externally hosted library. Attempts to run the script from an internal source were unsuccessful. As such, viewing the page does not guarantee that the heatmap will appear. Future iterations should adjust the page's formatting (particularly the space at the end of the page) for those instances.

The dashboard also relies on external hosting. I aim to substitute the Tableau dashboard with visualizations created using d3.js or a similar library. It's also possible to have one set of controls affect multiple visualizations. This may require using frameworks, rethinking how the visualizations are created, and selecting different filetypes as data sources. With these changes, the user will see the effects of their selections across the entire project, hopefully increasing engagement and personalized investment in the findings.

Accessibility is also a concern. The page has not yet been optimized for those with visual impairments, namely color-blindness. Additionally, it is intended for full displays in landscape

mode, not smaller mobile devices. The page's design may need to evolve to reach a wider audience.

There may be other issues that have not been addressed in the capstone nor this whitepaper. Future work may continue finding flaws with how the dataset is created, managed, and represented. The purpose was to show that the dataset as a whole is misleading and inaccurate, which has been demonstrated. Iterations of this type may become moot if the dataset is overhauled.

While the above topics are specific to this project, the approach of critiquing the description, details, and data can—and should—be applied to all datasets. Iterations of this nature would be an evolution of this critique's methodology while still providing contrasting visualizations of stated truths and the reality of the discoveries.

### *Evaluation*

This project met the goals set out for it in the initial application. It adds to the narrative that missing data (particularly shooting data) should not be ignored. It simultaneously presents strong evidence that better data management and reporting practices are needed for existing data. An argument reflecting the fact that it's not possible to capture every shooting incident was not included in this project. As important as it is conceptually, it did not add to the overall tone that developed. The more this project progressed, the more out of place it felt.

The overarching goal was to have a project that will provide examples which introduce/reinforce the need to be cautious when using data as a consumer and ensuring that one is forthcoming and transparent when making their own data collection available. As such, I

intend to contact NYC Open Data to express my concerns with this dataset. It is my hope that the description, details, and data align with one another in the near future.



## APPENDICES

*List of Variables*

## JavaScript

classSet	Class indicating if incident does not meet post-analysis criteria
truecount	Post-analysis count of victims
sliderMove	Selects the Slider element
sliderPut	Selects the future location of the Slider element
existing	Binary value indicating if slider was created
happened	Binary value indicating if slider was moved to new div

## GeoJSON

origLat; origLong	Latitude / Longitude originally assigned to the incident
roundLat; roundLong	Latitude / Longitude rounded to .0001
locCount*	Count of prior incidents at the “rounded” location
Latitude; Longitude	Latitude / Longitude assigned after geospatial transformations
Mtrue; Mfalse*	Count of entries with / without an affirmative murder classification
locCountOrig*	Count of all incidents at the “rounded” location

*\*Considers incidents having a pre-analysis victim/row count greater than or equal to 4.*

*Glossary of Functions*

## Javascript

moveit	Relocates the slider after it is loaded
observer	Observes mutations until slider has been loaded and relocated

### *Geospatial Transformations*

A list of INCIDENT\_ID, OCCUR\_DATE, OCCUR\_TIME, STATISTICAL\_MURDER\_FLAG, Latitude, and Longitude were preserved.

Mtrue and Mfalse were created to count the number of each STATISTICAL\_MURDER\_FLAG by INCIDENT\_ID.

This calculation was converted to a static number by copying and pasting as values.

STATISTICAL\_MURDER\_FLAG was then deleted.

Duplicate values were removed from the dataset.

The provided coordinates of all records were rounded to the nearest ten-thousandth.

Incidents were sorted chronologically.

A column was created to count the number of prior occurrences of a record's set of coordinates.

Incidents with a prior count of 0 were assigned their rounded coordinates, otherwise:

- .0004 was added to the longitude value if the prior count was even.

- .0004 was subtracted from the longitude value if the prior count was odd.

- .0004 was added to the latitude value if the prior count was greater than 2.

- .0004 was subtracted from the latitude value if the prior count was less than 2.

This manipulation allowed up to 5 pre/post analysis mass shooting incidents of identical locations to be individually identifiable on the map. No location had more than 5 incidents associated with it.

## SELECT BIBLIOGRAPHY

Carvajal, Kathy “Video: Man randomly opens fire in Brooklyn residential area” *FOX 5 NY*, 10 June 2022 <https://www.fox5ny.com/news/video-man-randomly-opens-fire-in-brooklyn-residential-area>.

D’Ignazio, Catherine and Klein, Lauren F., *Data Feminism*, The MIT Press, 2020

“Dataset Description”, <https://data.cityofnewyork.us/Public-Safety/NYPD-Shooting-Incident-Data-Historic-/833y-fsy8>.

“Dataset Footnotes”, [https://data.cityofnewyork.us/api/views/833y-fsy8/files/e4e3d86c-348f-4a16-a17f-19480c089429?download=true&filename=NYPD\\_Shootings\\_Incident\\_Level\\_Data\\_Footnotes.pdf](https://data.cityofnewyork.us/api/views/833y-fsy8/files/e4e3d86c-348f-4a16-a17f-19480c089429?download=true&filename=NYPD_Shootings_Incident_Level_Data_Footnotes.pdf).

Fitzsimmons, Emma. “Brooklyn Man Charged With Killing 2 and Dumping Guns.” *The New York Times*, 13 March 2013, <https://www.nytimes.com/2013/03/13/nyregion/brooklyn-man-held-in-killing-of-2-and-dumping-of-guns.html>.

“Police searching for man who opened fire on street in the Bronx” *Eyewitness News WABC*, 28 February 2019 <https://abc7ny.com/shooting-bronx-man-opens-fire-shots-fired/5160404/>.

“Sweet 16 party turns violent: 2 teens shot and wounded on Staten Island” *Eyewitness News WABC*, 31 July 2021 <https://abc7ny.com/staten-island-shooting-teens-shot-at-sweet-16-party-teenage-boys-and-wounded/10921118/>.