

# Milestone 3: Data Visualization: "NYC: Does Crime Ever Sleep in the City That Never Does?"

Mahdi Fourati 313078,  
Elsa Heitz 316735,  
Benoit Mathey-Doret dit Doret 324620

April 2025

## 1 Process book

When we embarked on the journey of creating a website focused on impactful data visualization, our first goal was to identify a dataset that truly inspired us. During our initial search, we encountered several datasets, but none sparked compelling ideas for engaging visualizations or captivating data narratives. Eventually, we discovered the New York City Crime Complaints dataset, and immediately, each team member envisioned intriguing possibilities for visual exploration and storytelling.

We quickly decided that our website would prominently feature a map showcasing the spatial distribution of crimes throughout NYC. However, defining an engaging and insightful data story remained a challenge. Initially, we considered approaching the data from a feminist perspective, exploring questions like, "Are crimes disproportionately affecting women?" and "What types of crimes are women most vulnerable to?" After conducting an exploratory data analysis, we observed that the crime distributions between women and men were not significantly different enough to yield a compelling narrative.

Next, we shifted our focus toward a socioeconomic perspective, hypothesizing a potential relationship between crime rates and socioeconomic factors such as unemployment and poverty. We envisioned merging our crime dataset with relevant socioeconomic data to examine whether wealthier districts experienced significantly fewer crimes compared to economically disadvantaged areas. After merging the datasets and performing preliminary analyses, we uncovered compelling insights that will be further detailed and discussed in the statistics section of our project.

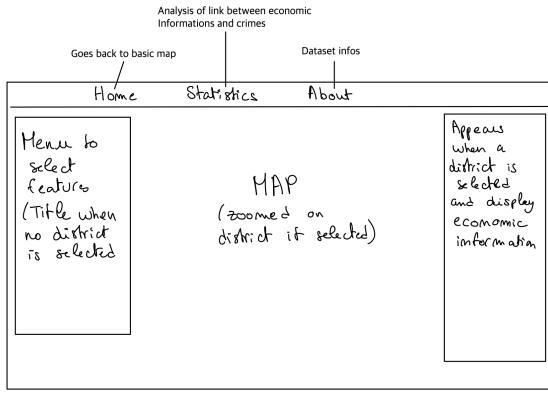


Figure 1: Top 3 most dangerous districts presented in an interactive manner.

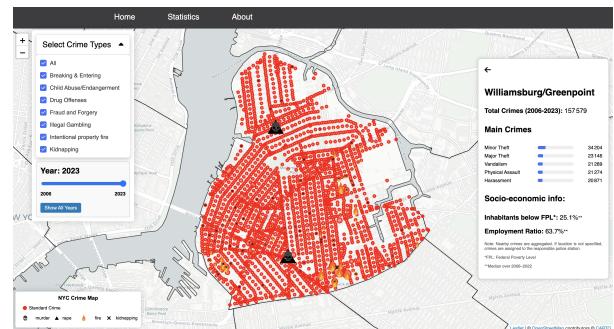


Figure 2: Interactive correlation table.

## 1.1 Home page

For our homepage, we envisioned users encountering an engaging, district-level map of New York City designed to immediately spark their curiosity. Considering our primary audience being individuals planning to visit or relocate to NYC, we prioritized clarity and ease of use.

Initially, we explored presenting a heatmap depicting crime density directly upon entry. However, upon testing this visualization, we found it overwhelming due to significant disparities in crime rates across districts. Consequently, we searched for alternative methods and ultimately selected a choropleth map. This choice provided a simple yet powerful visual message, clearly conveyed through an effective legend positioned at the bottom left.

Following the successful implementation of the crime-based choropleth map, we extended the concept to include socioeconomic data, allowing users to seamlessly toggle between crime and socioeconomic views with the simple click of a button. Initially, both maps shared similar color schemes, but we soon realized that distinct color palettes would significantly enhance user experience. Thus, we switched to a green-based scheme for socioeconomic data to clearly differentiate it from the blue-base representing crime.

A notable adjustment we made from Milestone 2 was the legend used in the choropleth crime map. After internal discussions, we concluded that a legend displaying only percentages was insufficient for clearly conveying the actual magnitude of crime occurrences. Thus, we revised the legend to include specific crime count ranges instead. While the color scheme continues to represent the bottom 20 percent intervals and so on, incorporating numerical crime ranges significantly enhances the clarity and impact of our visualization.

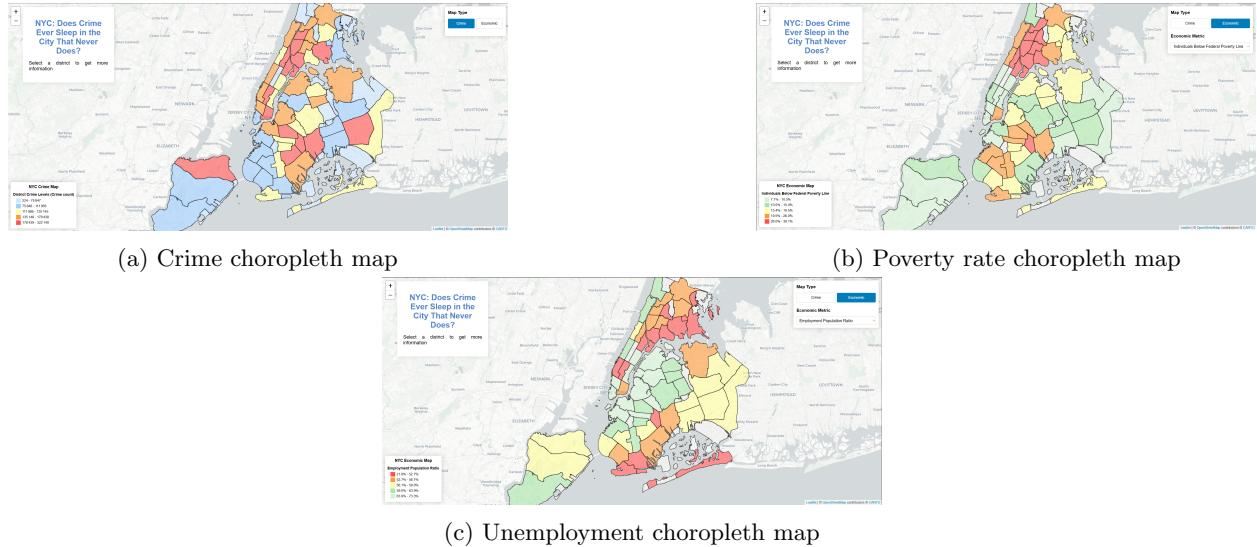


Figure 3: The 3 different choropleth maps of our website

Recognizing that our users, especially prospective residents, would be interested in detailed neighborhood-level information, we implemented an interactive feature allowing users to select a specific district and view precise crime locations within it. Our first approach employed simple red circles to indicate crimes; however, given NYC's high crime volume, this representation quickly became overly dense and visually aggressive.

To address this, we initially considered designing unique icons for different crime types, not all of them thought since it did the same result of overcharging the view. We further enhanced user interaction by introducing a feature that enables selecting specific crime categories. For example, individuals particularly concerned about home burglaries could easily filter and visualize only these occurrences like shown in figure 5. This significantly improved the clarity and effectiveness of our visualization.

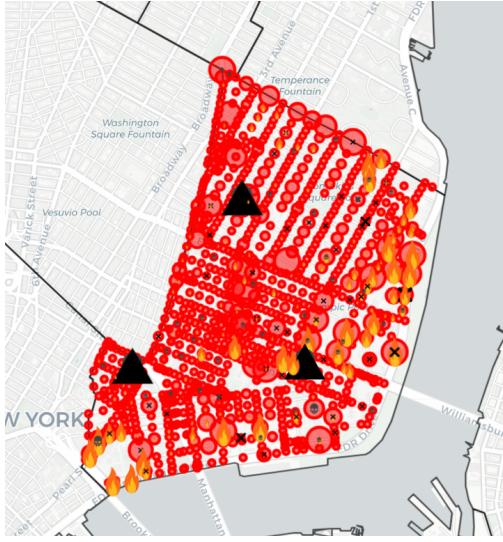


Figure 4: Crime display

Lastly, we incorporated a dynamic 'year slider,' empowering users to explore temporal changes in crime patterns effortlessly. This feature allows users to track crime evolution or focus on specific crime trends in their desired neighborhood or district, adding substantial value to their browsing experience.

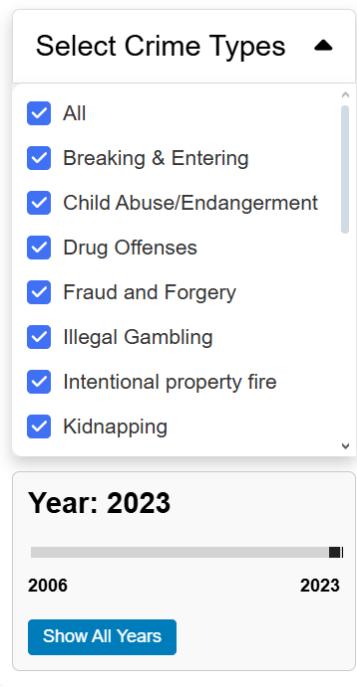


Figure 5: Year slider + Crime type selection

With our engaging visualizations in place, our next step was to incorporate essential information to enhance user interaction upon selecting a district. We determined that displaying the district name prominently was fundamental. Additionally, we included key statistical data such as the total crime count from 2006 to 2023 and identified the five most frequent types of crimes in each district. To provide further insight, we integrated socioeconomic metrics, specifically the percentage of inhabitants below the Federal Poverty Level (FPL) and the employment ratio.

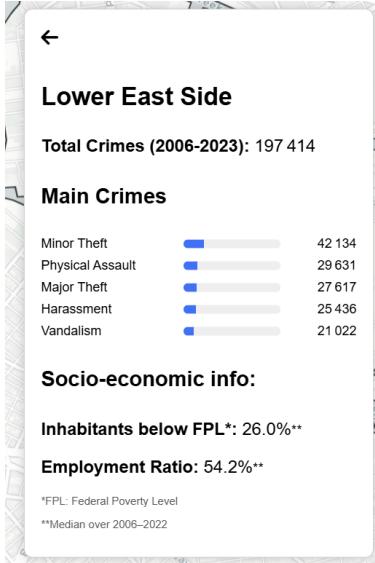


Figure 6: Useful information box

With these final additions, our visualization became comprehensive and user-friendly. Visitors to our website can now effortlessly access valuable insights tailored to their specific interests, making the experience both informative and intuitive.

## 1.2 Statistics

While our front page provides a rich and interactive exploration of crime and socioeconomic conditions in New York City, the sheer volume of information presented can quickly overwhelm users. To address this, we developed a Statistics page aimed at guiding users through key insights and crafting a coherent narrative from the data.

Our statistical analysis is structured into three main parts:

- **Motivation:** In this section, we explain the relevance of our work, highlighting the challenges of accessing clear, localized crime data in New York City. Our platform is designed for a broad audience, including residents, tourists, and policymakers, who seek to better understand crime patterns in their neighborhoods. While our interactive map is a powerful tool, it can be overwhelming due to the volume of information it presents. To guide users more effectively, we introduce this statistical analysis section, which provides structured insights and a clearer narrative.
- **Crime Situation in New York City:** Here, we take a deeper look into the NYC Crime Complaint Dataset, offering an overview of crime trends from 2006 to 2023. This includes insights such as how crimes are distributed across boroughs, which districts report the highest numbers, and what times of day are most dangerous. We also highlight the most common types of crime and examine demographic trends among victims and suspects, helping users contextualize the data beyond simple counts.
- **Relationship with Socio-Economic Factors:** In the final section, we investigate how crime correlates with socioeconomic indicators such as poverty and unemployment. Using exploratory data analysis (EDA), we found strong correlations between certain types of crime, particularly drug offenses, physical assaults, and intentional property fires. These findings are presented using clear visualizations, including scatter plots and summary tables, allowing users to explore how social conditions may influence public safety in different districts.

To integrate these findings into a seamless user experience, we adapted our plots for web display. As shown in Figure 5, initial EDA outputs were redesigned to align with the site's artistic direction, using consistent color schemes and interactivity features.

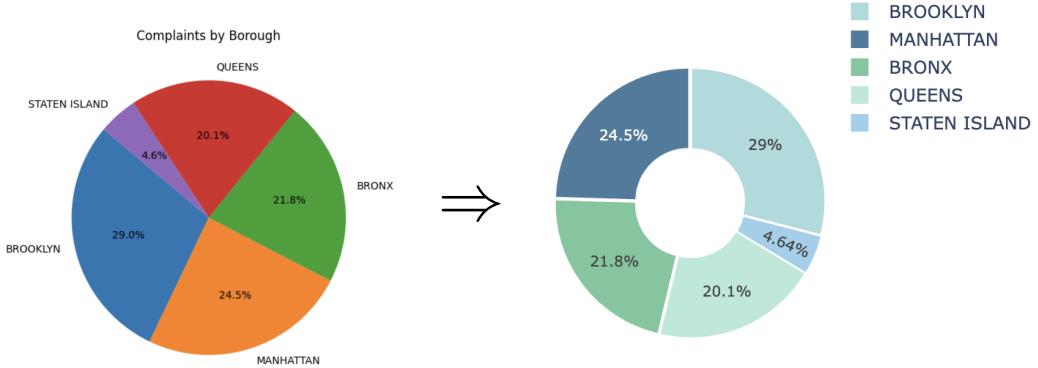


Figure 7: Evolution of a plot to be integrated in the website. The final plot respect the color artistic direction and is made interactif.

Figures 6 and 7 shown somes nice interactive visualization and also illustrate how users can engage with these insights directly. From filtering data via a search bar to toggling between crime types and socioeconomic indicators, users are empowered to explore the data independently while following a guided analytical path.



Figure 8: Top 3 most dangerous districts presented in an interactive manner.

Crime & Economic Analysis				
Correlation between crime types and economic factors				
Search crime types...				
Crime Type	Below Poverty Line Correlation	Below Poverty Line P-value	Employment Ratio Correlation	Employment Ratio P-value
Drug Offenses	+0.724	9.8e-11	-0.494	7.1e-5
Intentional Property Fire	+0.654	1.9e-8	-0.657	1.6e-8
Physical Assault	+0.611	2.7e-7	-0.448	4.8e-4
Weapon Possession	+0.601	4.8e-7	-0.469	1.8e-4
Murder	+0.570	2.4e-6	-0.478	1.3e-4
Prostitution Offenses	+0.531	1.5e-5	-0.433	6.2e-4
Illegal Gambling	+0.510	6.8e-5	-0.294	0.03
Kidnapping	+0.473	1.5e-4	-0.379	0.003
Public Disturbance	+0.471	3.7e-4	-0.486	0.001

Figure 9: Interactive correlation table.

Our statistical analysis complements the interactive features of the website by offering a deeper understanding of crime patterns in New York City. By combining historical crime data with socioeconomic indicators, we not only highlight where and when crimes occur, but also explore potential underlying causes. This layered approach encourages users to move beyond surface-level observations and consider broader structural factors that impact public safety. Ultimately, our goal is to make complex data accessible, meaningful, and actionable for anyone seeking to better understand and engage with their city.

### 1.3 Conclusion

This project has been a comprehensive and rewarding journey through the world of data visualization, web development, and collaborative storytelling. Starting from a raw and complex dataset, we learned how to extract meaningful insights and present them in a way that is both engaging and accessible to a broad audience.

One of the key challenges we faced was structuring a real-scale web application using HTML, CSS, and JavaScript. We gained valuable experience organizing a multi-file project, ensuring consistency in style and functionality across pages, and maintaining clean, modular code. We also learned how to properly manage and process data to optimize performance. By carefully filtering, aggregating, and formatting our datasets, we ensured that our website remained lightweight and responsive.

In parallel, we deepened our understanding of how to present data effectively. It's not enough to visualize data, one must also tell a story. From choosing appropriate color schemes to designing interactive filters and ensuring the responsiveness of our plots, we learned how to prioritize clarity and usability. Our iterative design process helped us make complex information digestible without oversimplifying the story behind the numbers.

Beyond technical skills, this project reinforced the importance of considering the societal context of our work. By exploring the links between crime and socioeconomic factors, we aimed not only to inform, but also to provoke thought and empower users to engage with their environment more critically.

## 2 Peer assessment

**Together:** We had regular meetings to think about the design all together and put all our ideas in common. We also all dived into the data to find insights that could be interesting to present through an exploratory data analysis.

**Mahdi Fourati:** He cleaned and processed the dataset with economic information for the choropleth maps, created the map for crimes and economic data. He also implemented the slider to chose which year we want to display.

**Elsa Heitz:** She cleaned and processed the crime data. She used it to create the crime type filter with the dropdown menu. She also implemented the information menu on the district on Figure ???. She created figures for the statistics page, participated to design and write the information on the page.

**Benoit Mathey-Doret dit Doret:** He created ad processed the boarder dataset. He found how to integrate the map and make it interactive. He incorporated the boarders of the districts. He created the About and stats pages for which he conducted further analysis, created figures and participated in writing it.

Overall we are very satisfied on the work we performed through this project and in the way the workload was share among the team.