NYC Crime 2025: Visualization Analysis Report

Authors: Misael Perez, Syed Kauser, Md Omit
Course: CSC 47400 – Information Visualization
Year: 2025

Introduction

This project analyzes crime incidents in New York City during the year 2025 with the goal of a relationship between criminal activity and socioeconomic status. The analysis is based on a merged dataset of 405,514 crime records and median household income data for 177 zip codes.

The work follows the What–Why–How visualization framework:

- What: Crime and income datasets
- Why: Discover trends, correlations, and risk factors
- How: Visual encodings using charts, maps, and interactive dashboards

Datasets

This project uses two primary data files which is `crimedata_with_zip.csv` and `zip_median.csv`

Crime Data with ZIP Codes (`crimedata_with_zip.csv`):

This file contains NYPD-reported crime incidents for New York City in 2025, enriched with ZIP code information to enable spatial analysis. Key characteristics include:

- Individual crime records with date, time, borough, offense description, and location type
- ZIP code associated with each incident
- Categorical attributes (crime type, borough, victim/suspect demographics)
- Quantitative attributes (crime counts, temporal aggregations)

This dataset is used for:

- Computing total crimes per ZIP code

- Identifying the most common crime in each ZIP code
- Time-based, location-based, and demographic visualizations

Median Household Income by ZIP Code (`zip_median.csv`):

This file contains median household income data aggregated by ZIP code, derived from U.S. Census–based sources. Key characteristics include:

- One row per ZIP code
- Numerical attributes representing median household income
- Used as a contextual socioeconomic variable

This dataset is used for:

- Comparing crime frequency against income levels
- Creating scatterplots and choropleth maps relating income to crime patterns
- Supporting conclusions about socioeconomic correlations

Extreme Outliers

- Highest Crime Volume: Zip 11207 (East New York, Brooklyn)
    - Total Crimes: 7,068
    - Median Income: $55,419
    - Top Crime: Harassment 2 (1,318 incidents)
- Lowest Crime Volume: Zip 11697 (Breezy Point, Queens)
    - Total Crimes: 41
    - Median Income: $134,844
- Wealthiest Neighborhood: Zip 10007 (Tribeca, Manhattan)
    - Median Income: $250,001+
    - Total Crimes: 1,356 (Moderate)
    - Top Crime: Petit Larceny
- Poorest Neighborhood: Zip 10454 (Mott Haven, Bronx)
    - Median Income: $27,500
    - Total Crimes: 3,568 (High)
    - Top Crime: Harassment 2

Both datasets are joined using the ZIP code attribute, enabling integrated spatial and socioeconomic analysis.

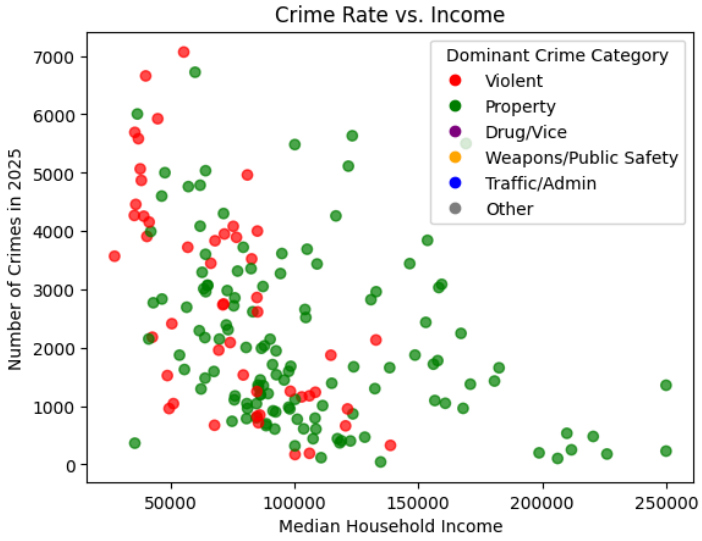
Data Preprocessing

Before visualization, several preprocessing steps were applied:

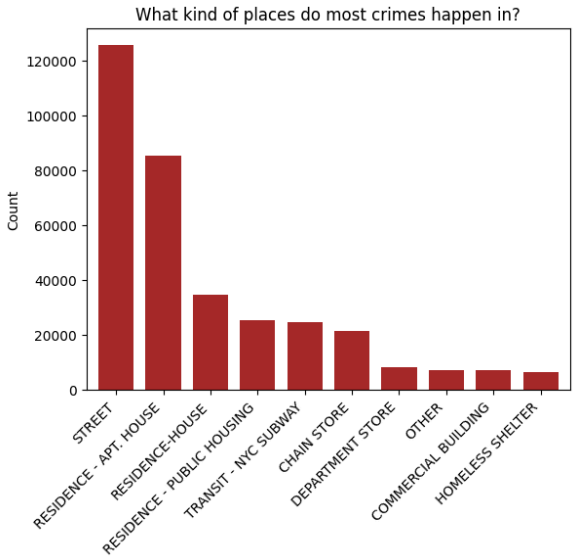
- Added zip codes to crime records
- Removed or handled missing (N/A) values
- Converted date and time fields into usable formats
- Categorized 56 crime types into 6 major categories as seen below

Location-Based Analysis

- A scatter plot visualizing the number of crimes in each area vs. their median household income, grouped by major category of crime, shows crimes are more likely to occur in poorer neighborhoods. This trend is especially clear among the areas where violent crimes dominate.
- A bar chart of top places where crimes occur shows crimes mainly happen in the street, followed by residential areas, the NYC subway, and stores.



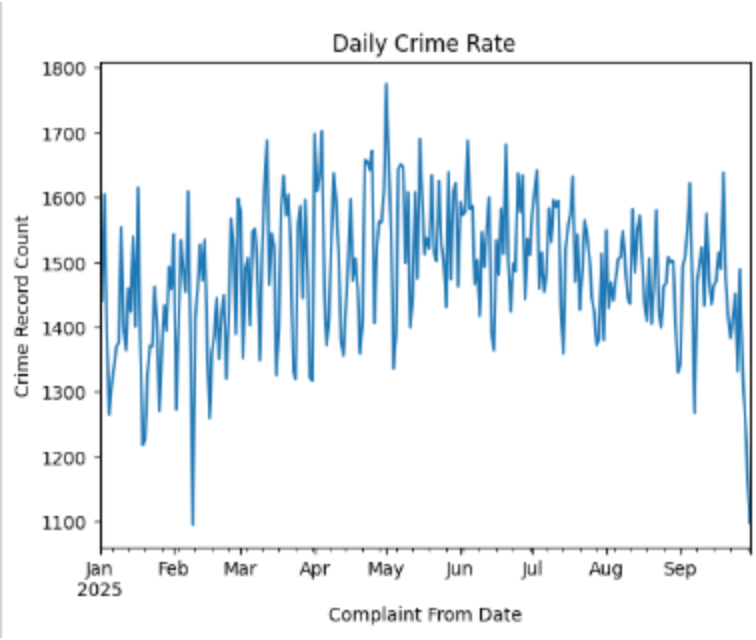
Most recent (2023) median household income from ACS used.
Crime counts were grouped by their location's ZIP code to match against median incomes.



Time-Based Analysis

Daily Crime Rate Over Time

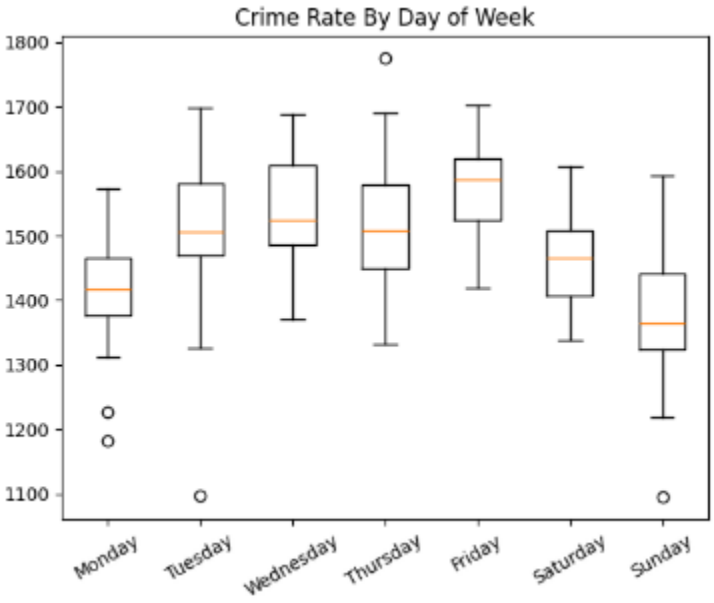
A line chart of daily crime counts across 2025 shows fluctuations with no extreme seasonality but noticeable short-term variations. Line marks are appropriate here because time is an ordered attribute.



Crime by Day of Week

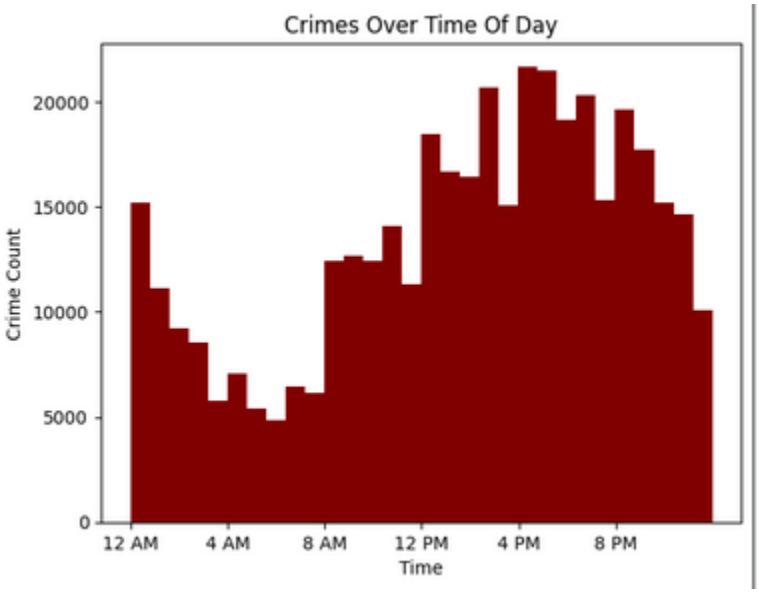
Boxplots comparing crime rates by weekday show:

- Higher median crime rates on weekdays, especially mid-week
- Slightly lower crime rates on weekends



Crime by Time of Day

A histogram of crimes over the 24-hour day shows peaks during the afternoon and evening hours, indicating higher risk periods.

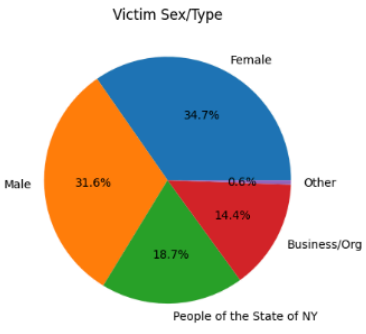


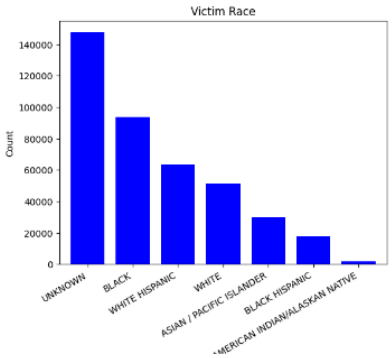
Demographic Analysis

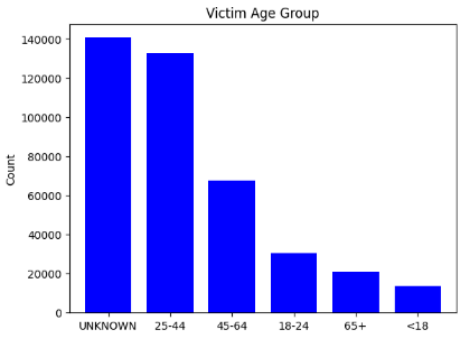
Victim Demographics

Visualizations of victim data show:

- Victims are slightly more likely to be female than male
- The most common reported victim age group is 25–44



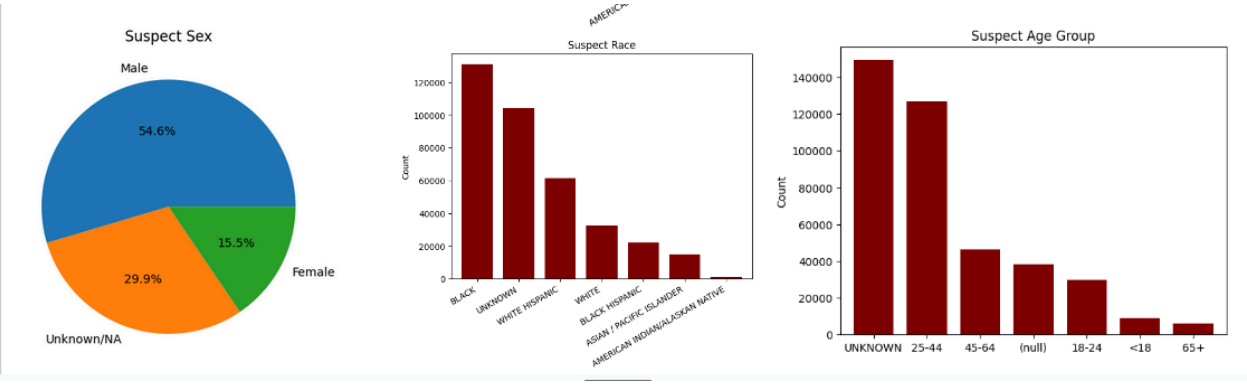




Suspect Demographics

Suspect-based charts reveal:

- A higher proportion of male suspects
- Most suspects fall within the 25–44 age group

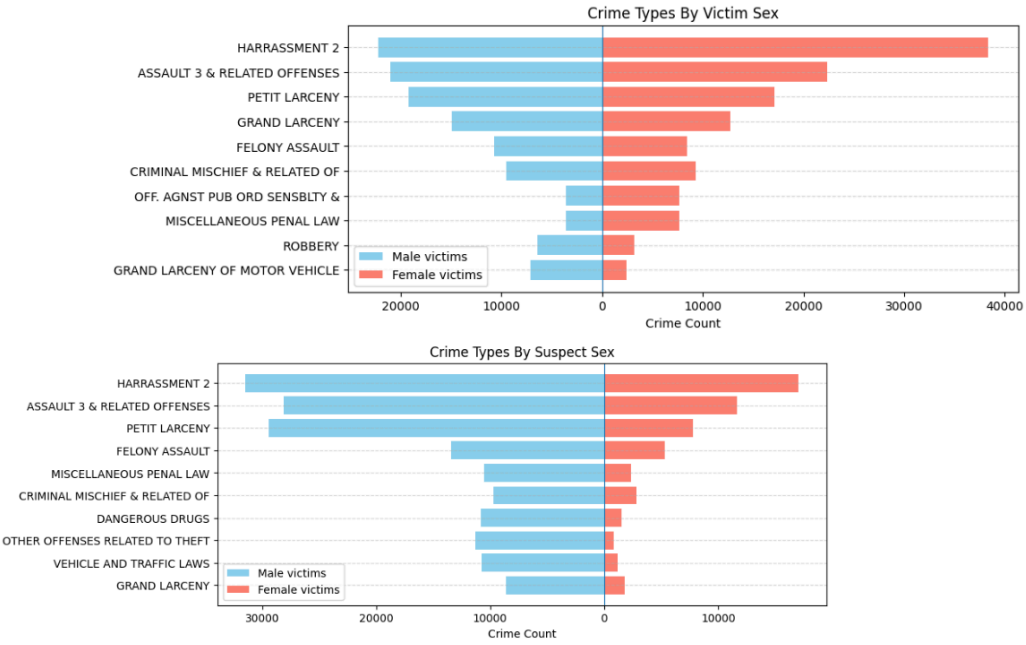


Crime Type Analysis

Diverging bar charts compare crime types by:

- Victim sex
- Suspect sex

Crimes such as harassment, assault, and larceny dominate across categories, while severe crimes appear less frequently. The diverging layout allows direct comparison between groups.



Interactive Map

An interactive choropleth map was created using:

- GeoPandas for spatial processing
- NYC MODZCTA zip code boundaries
- Folium / Leaflet.js for interactivity

Features:

- Displays total crime count per zip code
- Shows most common crime per zip code
- Allows filtering by crime type

Interactive Map

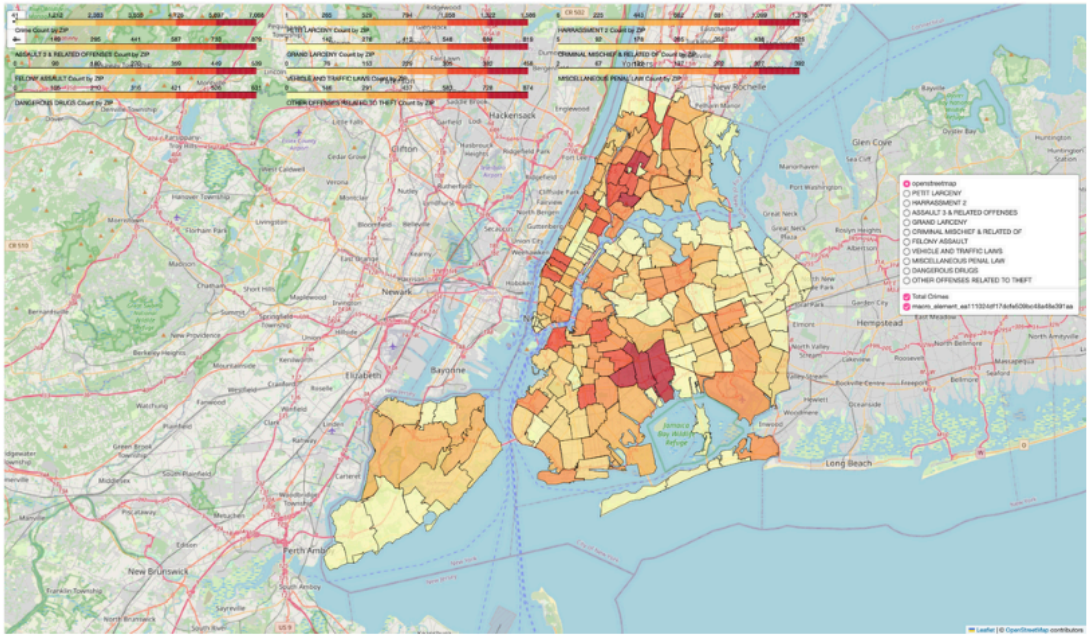
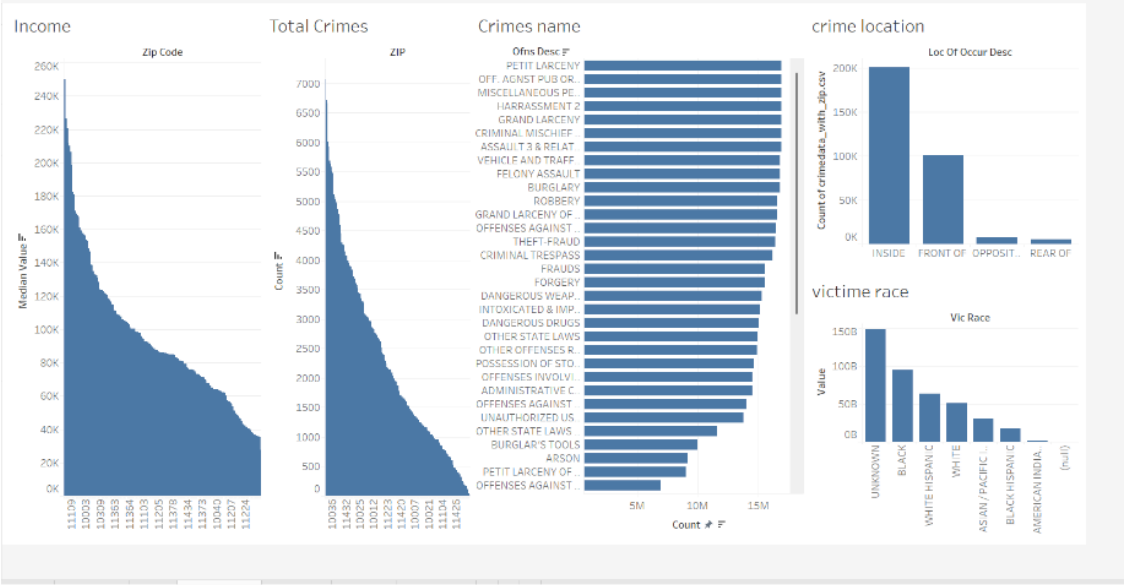


Tableau Dashboard

A Tableau dashboard integrates multiple coordinated views:

- Income vs. zip code
- Total crimes by zip
- Crime type frequency
- Crime location and victim race

Tableau



Conclusion

This project demonstrates how visualization can reveal meaningful patterns in large, complex datasets. Key findings include:

- Strong association between lower income and higher crime
- Clear spatial and temporal crime patterns
- Demographic trends among victims and suspects

By applying visualization principles such as expressiveness, effectiveness, and the What–Why–How framework, the project successfully transforms raw crime data into actionable insights.