Lutfulloev Asatillo, Kareem Virani, Jubyaid Uddin

Professor Yunhua Zhao

CSC 474 - Visualization

12/17/2025

Interactive Visualization of NYC Traffic Collisions

Abstract

Traffic collisions are a serious safety problem in large cities like New York, and they are affected by many factors such as time, location, road design, weather, and driver behavior. Looking at only basic statistics or simple charts makes it hard to understand how all of these factors work together. Because of this, planners, researchers, and the general public need better tools to explore and understand collision data. In this project, we created an interactive visual analytics dashboard using New York City motor vehicle collision data to help users examine spatial, temporal, and contextual patterns at the same time.

The dashboard includes features like map-based heatmaps, time filters, and multiple visual views that allow users to find high-risk areas, notice unusual patterns, compare collisions across different boroughs, and explore how factors such as time of day, day of the week, and contributing causes affect crash severity. The design follows common visualization principles like providing an overview first, allowing users to zoom and filter, and then showing more detailed information when needed. Overall, this project shows how interactive visualizations can support better understanding and decision-making for traffic safety and urban planning.

Introduction

Traffic safety is a major concern for city lawmakers, transportation departments, and residents, especially in a dense and complex city like New York. NYC publishes a large and detailed dataset of reported motor vehicle collisions, containing millions of records and dozens of attributes for each incident, such as location, time, contributing factors, and severity. The data is accurate and is updated daily to show new records in real time. While this data is extremely valuable, its size and complexity make it difficult to analyze using simple plots, tables, or summary statistics. Important patterns and problem areas can easily be missed without more interactive and flexible ways to explore the data.

Our group chose this topic because traffic collisions are an issue that directly affects everyday life, and the data is both realistic and relevant to real-world decision making. We were interested in working with a dataset that is large, multivariate, and spatial in nature, since it allowed us to apply many of the visualization techniques discussed in class. The NYC collision dataset also provided an opportunity to move beyond answering a single question and instead design a tool that supports open-ended exploration and discovery.

The motivation for this project is to address the following problem: how can interactive visualization help users better understand multivariate patterns in traffic collisions and identify meaningful risk factors and hotspots? Rather than focusing on one predefined analysis, our goal is to support tasks such as exploration, monitoring trends over time, and detecting anomalies across different locations and conditions. This approach reflects how real analysts often work when they are trying to understand complex systems.

An imagined client for this system is a city traffic safety analyst or urban planner who needs to quickly identify dangerous areas, compare trends across neighborhoods or boroughs, and drill down into specific incidents for more detail. A secondary audience includes journalists and members of the public who want to better understand traffic risks in their own communities and see how factors like time, location, and contributing causes influence crash patterns.

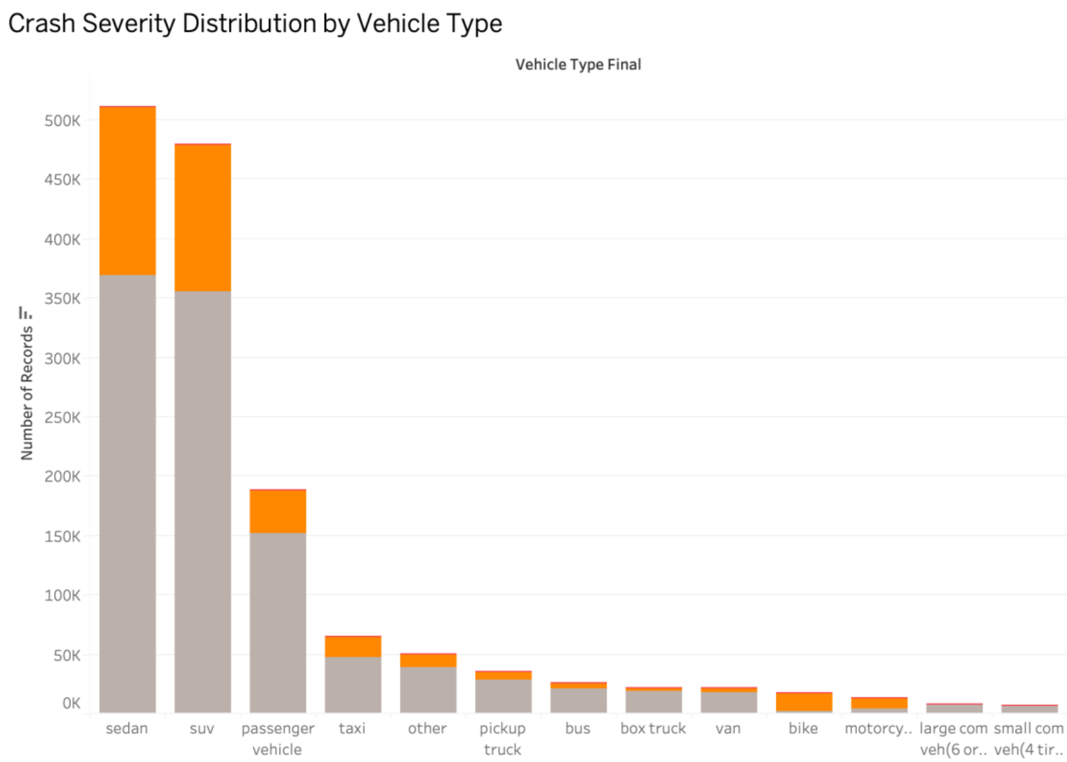
Crash Severity by Vehicle Type and Contributing Factors

One part of our dashboard focuses on understanding how crash severity differs depending on the type of vehicle involved and the contributing factors reported for each collision. While the dataset contains detailed numeric information about injuries and fatalities, these values are difficult to interpret without grouping and visual comparison. To make these patterns easier to understand, we created a set of visualizations that summarize crash severity across vehicle types and contributing factors.

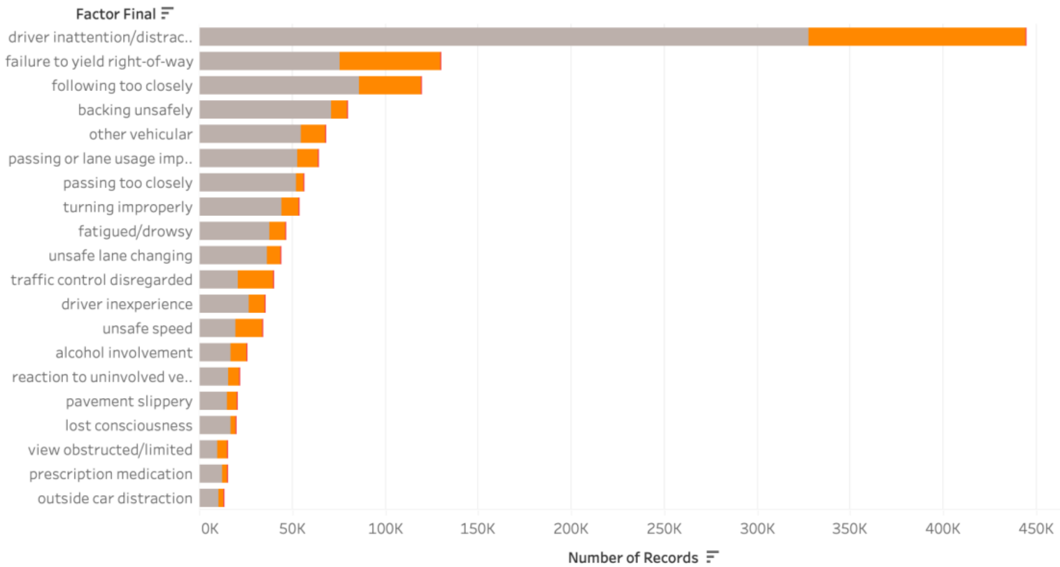
This section includes stacked bar charts that show the number of fatal, injury, and property-damage-only crashes for each major vehicle category and contributing cause. These charts make it clear that passenger vehicles such as sedans and SUVs are involved in the largest number of severe crashes overall, which is expected given how common these vehicles are on New York City roads. However, the charts also show that motorcycles, while involved in fewer crashes overall, have a much higher proportion of injury and fatal outcomes compared to other vehicle types.

The contributing factor visualizations show similar contrasts. Driver inattention or distraction appears most frequently in severe crashes, making it the most common contributing factor associated with injuries and fatalities. Other factors, such as failure to yield right-of-way and unsafe speed, appear less often overall but are associated with a higher share of fatal crashes. This suggests that some behaviors may be less common but lead to more serious outcomes when they do occur.

In addition to the bar charts, we included matrix-style visualizations that display severity categories across vehicle types and contributing factors in a grid format. These views allow users to compare severity distributions across multiple categories at once and confirm patterns seen in the bar charts. Together, these visualizations help users move beyond simple crash counts and better understand how different vehicles and behaviors relate to more severe traffic collisions.



Top Contributing Factors by Crash Severity



Crash Severity by Vehicle Type

	Severity Category		
Vehicle Type Fi..	Fatal	Injury	PDO
sedan	552	141,658	369,193
suv	590	123,207	355,812
passenger vehicle	213	36,620	151,891
taxi	49	17,295	47,742
bike	64	14,935	2,554
other	143	11,292	39,448
motorcycle	304	8,881	4,026
pickup truck	64	7,390	28,285
bus	59	5,226	20,735
van	36	3,853	17,964
box truck	46	3,184	19,021
large com veh(6 or more tires)	24	787	7,779
small com veh(4 tires)	3	725	6,376

Crash Severity by Contributing Factor

Factor Final	Severity Category: Fatal	Injury	PDO
driver inattention/distraction	390	117,022	327,314
following too closely	23	32,939	86,293
failure to yield right-of-way	289	54,112	75,285
backing unsafely	41	8,540	70,888
other / rare	329	35,994	67,905
other vehicular	26	13,113	54,914
passing or lane usage improp..	46	10,900	52,757
passing too closely	3	4,272	52,048
turning improperly	20	9,350	44,333
fatigued/drowsy	3	8,165	37,899
unsafe lane changing	16	7,296	36,453
driver inexperience	66	8,248	26,981
traffic control disregarded	270	18,881	20,986
unsafe speed	402	14,320	19,229
alcohol involvement	108	8,101	17,138
lost consciousness	56	2,934	16,788
reaction to uninvolved vehicle	11	5,872	15,821
pavement slippery	11	5,211	15,092
prescription medication	6	2,050	12,615
outside car distraction	7	2,602	10,160
view obstructed/limited	24	5,131	9,927

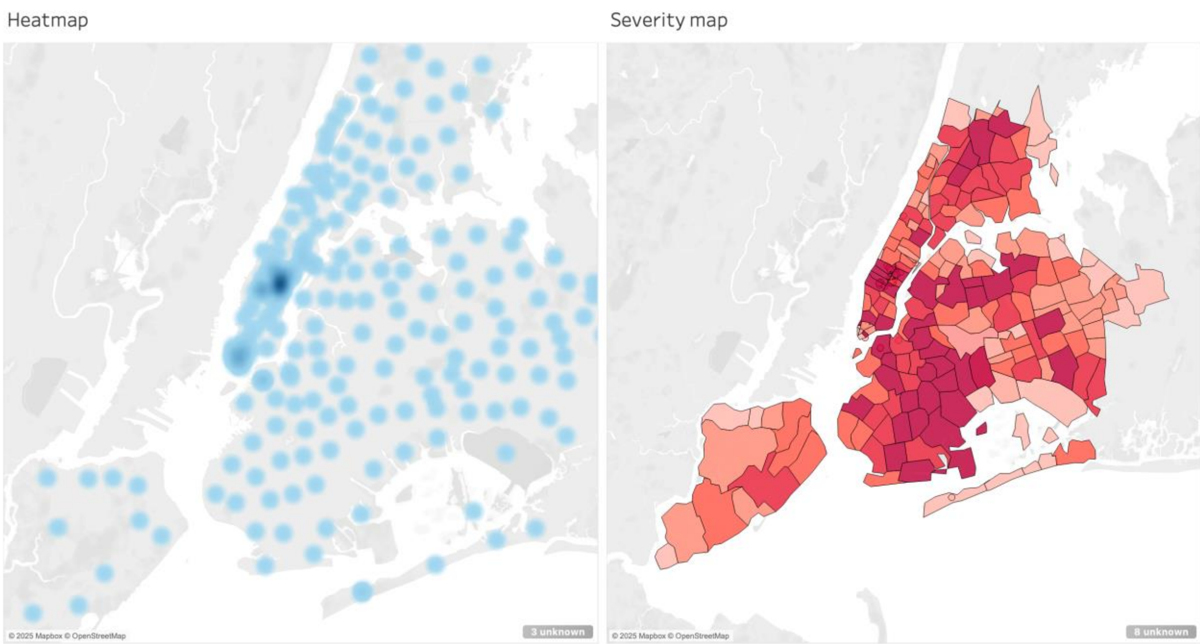
Spatial Analysis

The main goal of spatial analysis is to gain insights that wouldn't be obvious just by looking at the data in a table. By analyzing location-based information, spatial analysis helps identify hotspots, clusters, and differences between areas, as well as how physical space influences behavior, activity, or outcomes. It is often used to support decision-making, planning, and problem-solving in fields like urban planning, environmental studies, public health, transportation, and business. If looks at where things are located and why they are there, I use maps, data layers, and spatial tools to make sense of real-world locations and how they interact with each other.

The two maps show how activity is spread across New York City and nearby areas. The first map breaks the region into three color-coded zones- red, orange, and yellow. These zones likely represent different levels of importance, population density, or some other geographic factor. The red zone, which includes parts of Brooklyn and Staten Island, stands out as the most significant, while the orange and yellow zones extend into the outer boroughs and parts of New Jersey.

The second map adds a heatmap on top of this layout to show where activity is most concentrated. Warmer colors like yellow and green represent higher levels of activity, which are mostly found in central New York, especially Manhattan and parts of Brooklyn. These high-activity areas line up closely with the red zones from the first map, suggesting that the most important areas are also the most active. In contrast, places like southern Queens and the area around Jamaica Bay show cooler blue colors, meaning lower activity levels.

Overall, the maps together show a clear connection between key urban areas and higher levels of activity, with the city's core standing out as both the most important and the most active region.

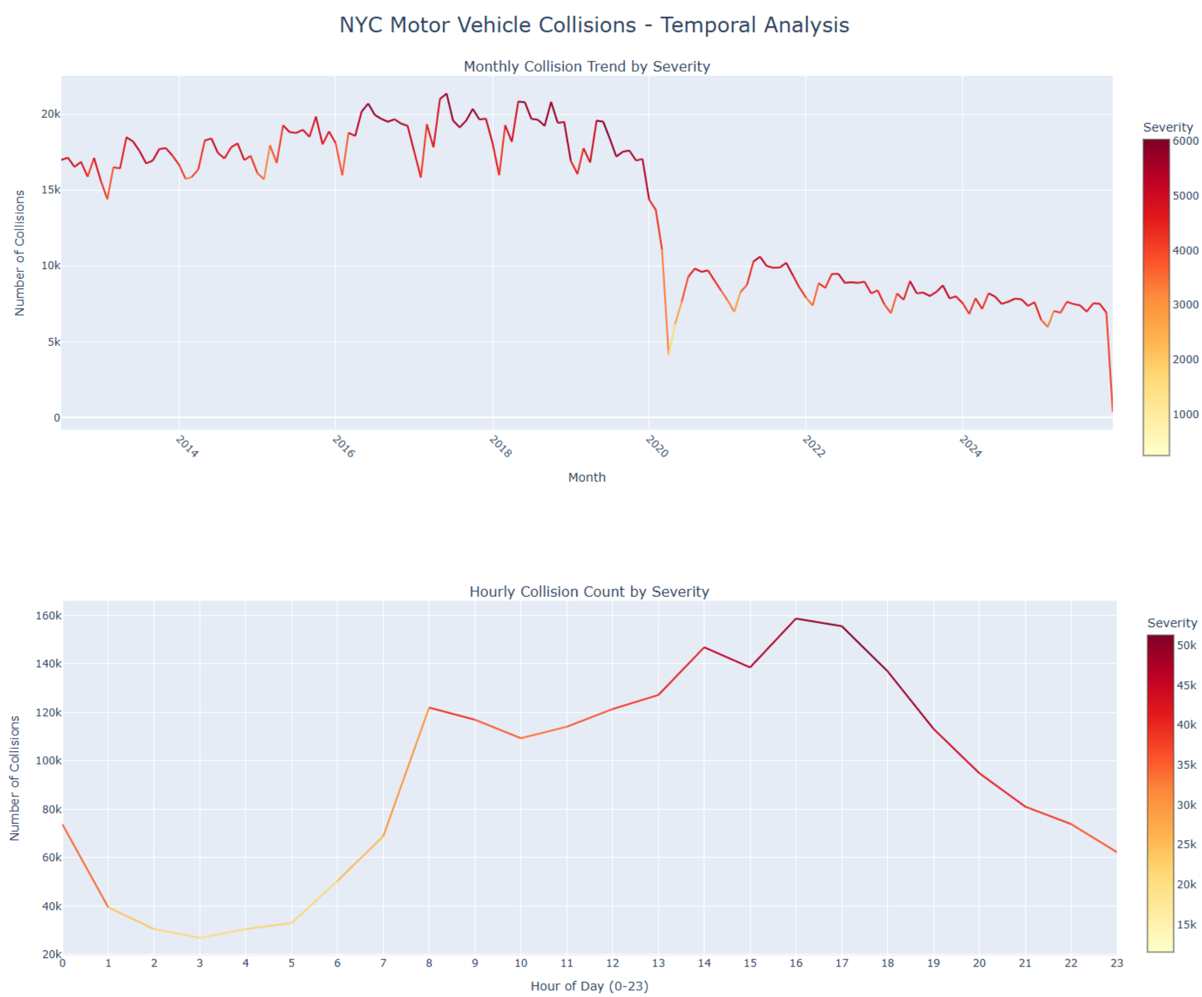


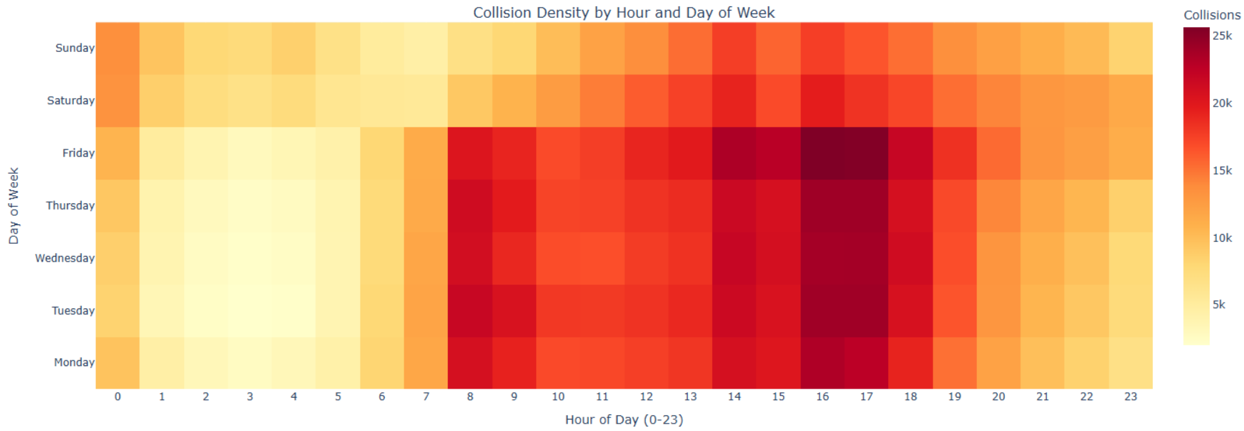
Temporal Analysis

The temporal analysis focuses on understanding how motor vehicle collisions in New York City vary over time, both in terms of crash volume and severity. In the respective dashboard, temporal patterns were examined across multiple time scales, including monthly trends, hourly distributions, and day-of-week variations. Severity is encoded using a derived severity metric and visualized alongside collision counts to determine correlation between high-frequency events and high-impact events.

At a monthly level, collision volume shows a relatively stable pattern from 2012 through early 2020, followed by a sharp decline coinciding with the onset of the COVID-19 pandemic. Although overall crash counts decrease substantially after 2020, severity remains elevated in certain periods, indicating that lower traffic volume did not proportionally reduce crash impact. This highlights a disparity between frequency and severity that would not be visible using volume alone. Hourly analysis reveals strong daytime patterns. Collision counts are lowest during early morning hours and increase steadily throughout the day, peaking during afternoon and early evening hours. These peak periods also correspond to higher average severity, suggesting that rush-hour traffic conditions contribute not only to more frequent crashes but also to more severe outcomes. Finally, a heatmap of collision density by hour and day of week exposes distinct weekday and weekend behaviors. Weekdays exhibit pronounced peaks during morning and evening commuting hours, while weekends show a flatter distribution with elevated activity later in the day. Severity patterns persist across these temporal slices, reinforcing that time-of-day and day-of-week are critical factors in understanding crash risk.

Overall, the temporal analysis demonstrates that collision behavior is highly structured in time, with clear interactions between crash frequency, severity, and human activity patterns. These findings motivate deeper exploration when combined with spatial and vehicle-level analyses dashboards.





Collision Density by Hour and Day of Week
Collisions
25k
20k
15k
10k
5k
Sunday
Saturday
Friday
Thursday
Wednesday
Tuesday
Monday
Day of Week
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Hour of Day (0-23)