CPSC 4030 - NYC Vehicle Collision Visualization

GitHub Pages Website:
https://mareed50.github.io/cpsc4030/

GitHub Repository:
https://github.com/mareed50/cpsc4030/

Group Members:
Michael Reed
mareed@clemson.edu
C19417312

Overview and Motivation:

The project is aimed at offering a comprehensive and visually intuitive representation of motor vehicle collisions within the boroughs of New York City. The motivation behind this visualization project is to create visualizations that create meaning behind the date on the patterns, trends, and contributing factors associated with collisions in the city. By leveraging visualization techniques learned, we are able to provide valuable insights into the contributing factors of collisions and identify high-risk scenarios.

Related Work:

The visualizations used in the project were inspired off of the class assignments in which we learned to create different visualizations in D3. The map visualization was inspired off of the Tableau assignment that used a map to describe the data.

Questions:

The original questions that influenced the visualizations created in this project are:

1. Which boroughs have the highest numbers of collisions and injuries/fatalities related to these collisions.
2. Is there a correlation between collisions and the time of day?
3. What factors contribute the most to vehicle collisions?
4. Which vehicles are most likely to be involved in a collision?

Most of these questions remained the same throughout the creation of the project, though a couple questions were modified or added:

- Which factors contribute to higher deaths compared to injuries?
- What vehicles or factors are most likely to be involved in a collision involving a motorcycle or bike?

Data:

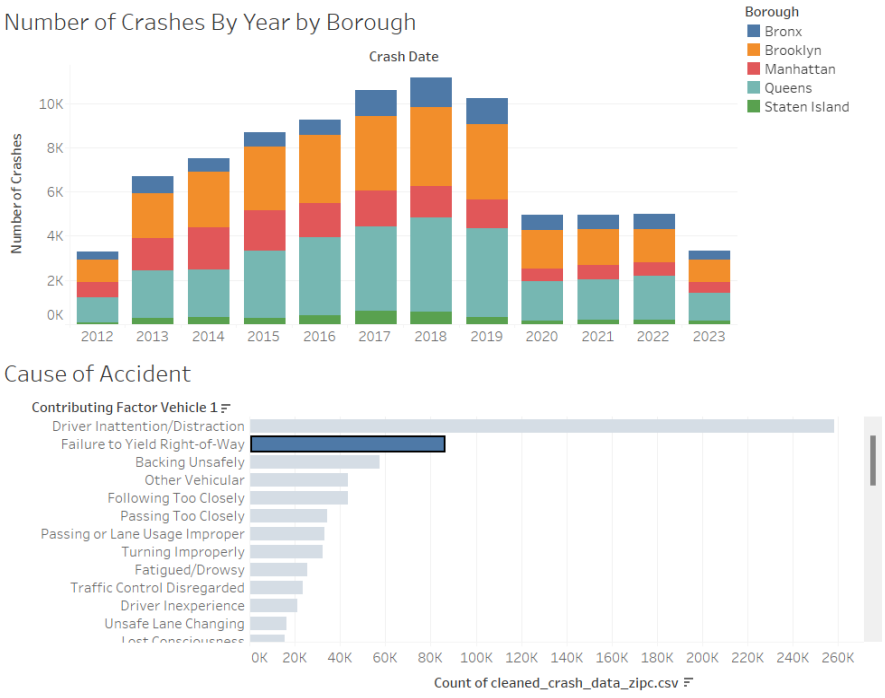
The data used for this project was downloaded directly from the motor vehicle collisions dataset provided by the NYPD.
https://data.cityofnewyork.us/Public-Safety/Motor-Vehicle-Collisions-Crashes/h9gi-nx95

The dataset contains over two million collisions and was filtered and refined to provide more meaningful data.

Exploratory Data Analysis:

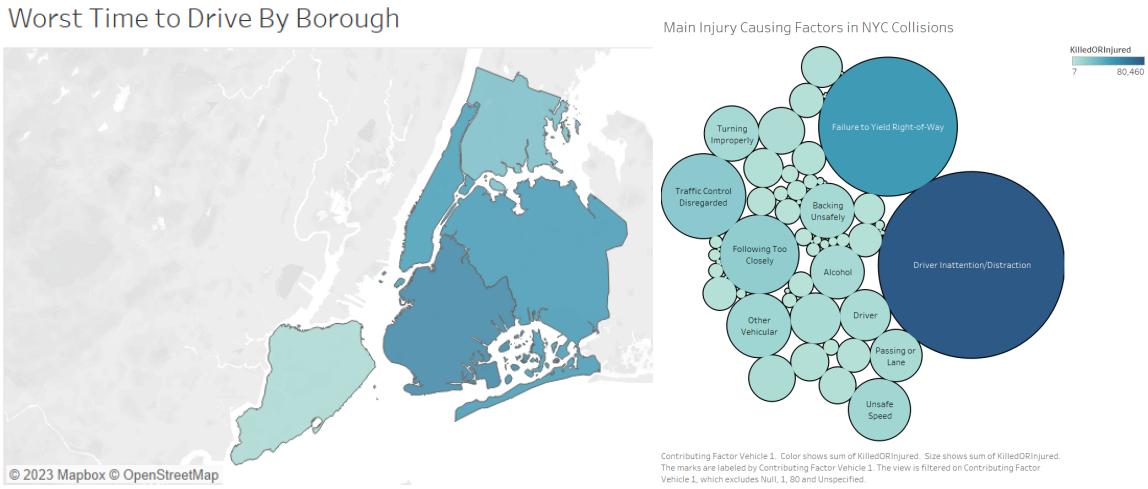
In order to get an initial idea of what visualization tools might be useful for this project, we started by creating a few visualizations on Tableau, a data visualization application that makes it quick and easy to make unique visualizations. We tested our data with bar charts, scatter plots, maps, line graphs, and other tools to visually represent the distribution of collisions. These visualizations revealed patterns, such as higher amounts of collisions at specific times. The designs played a critical role in creating the final visualizations of our project. By gaining a deeper understanding of the data through these visualizations, we were able to adjust the visualizations used to address specific questions that we wanted to answer as well as provide a more informed view of the collision data in New York.

Design Evolution:

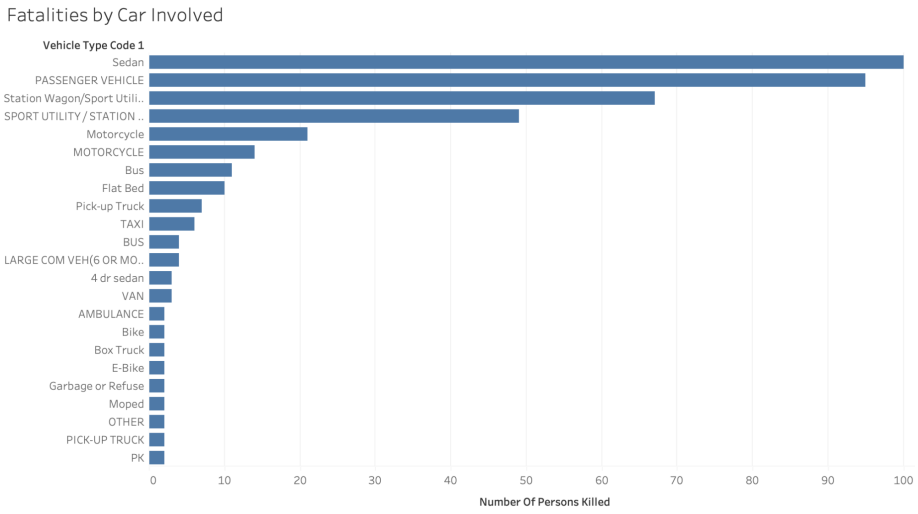


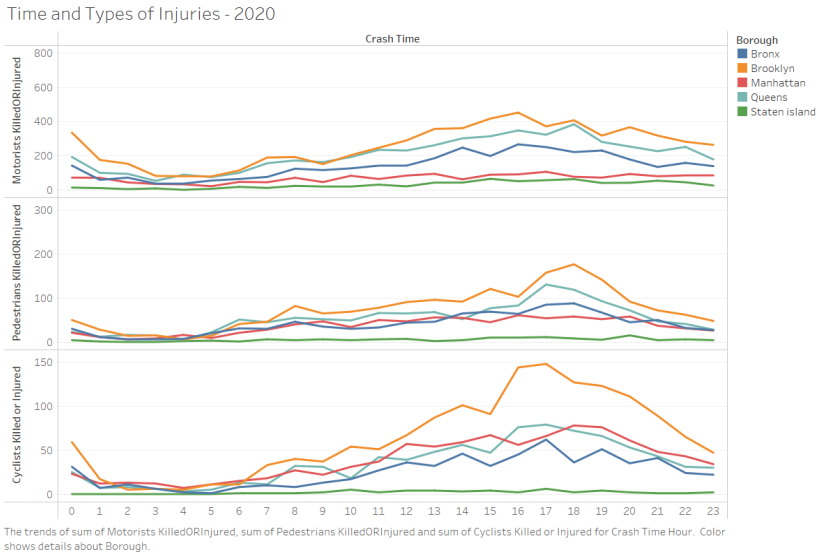
This image is one of the original visualizations considered for the project. It was designed to show the number of collisions in relation to year separated by borough as well as the contributing factors towards collisions. This quickly evolved into visualizations that were designed to have less cluttered information.

The charts shown below were the next iteration. They better described the relationships between the data without cluttering the visualization with unnecessary information.



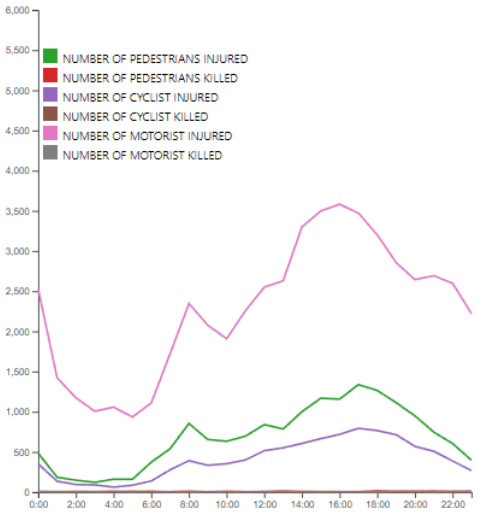
Two more visualizations were added to create the outline for the final product, a bar chart describing the relationship between vehicle and number of collisions and a multiple line graph that relates collisions with time.

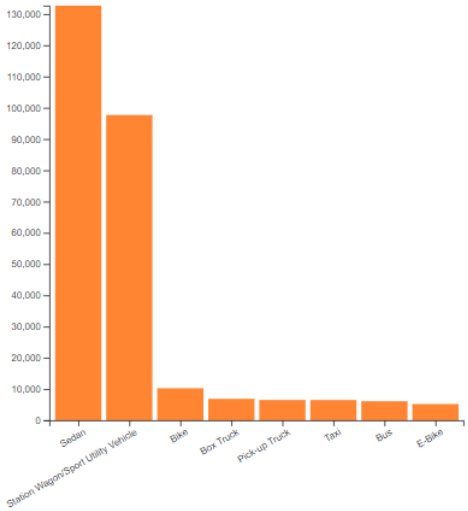




After finalizing the visualization tools that would be used in this project, it was decided to simplify them in order to not overwhelm or overcomplicate the main goals of the project.

To do this, we removed the unnecessary color scale from all of the plots, as the information given by it was already shown by another means, we removed a few of the lines from the time graph as they were unnecessary and could instead be combined in some places to create a more readable visualization, and we reorganized the bar chart to be easier to understand.





The above visualizations are the final charts created to illustrate the dataset.

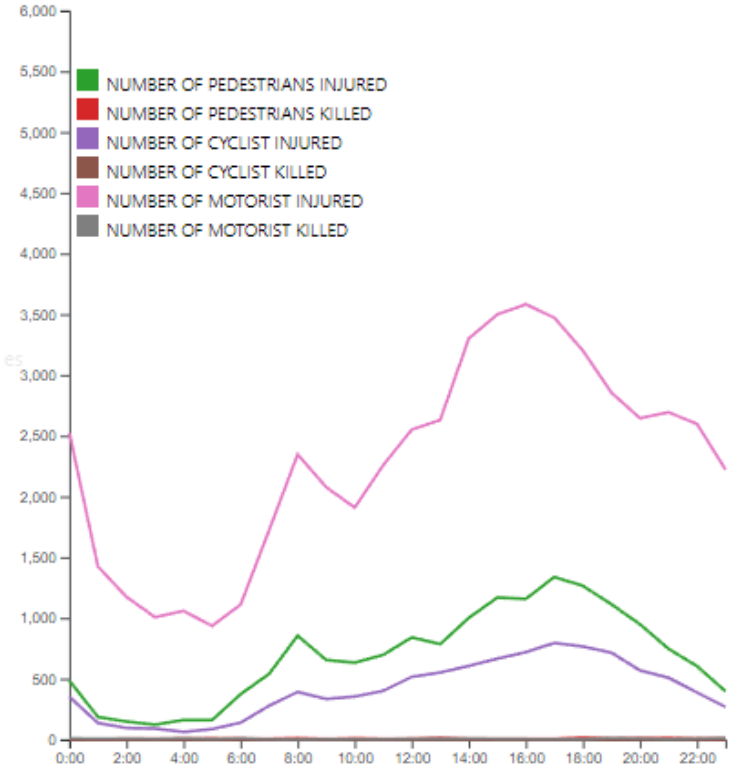
Project Implementation:

Borough Map

The map was designed with the simple idea of showing the relationship between borough and number of collisions. When hovered, a tooltip will appear with the name of the borough and number of collisions in the borough.

When a borough is selected, the other visualizations adjust to show data from that borough alone.

Time Plot Graph

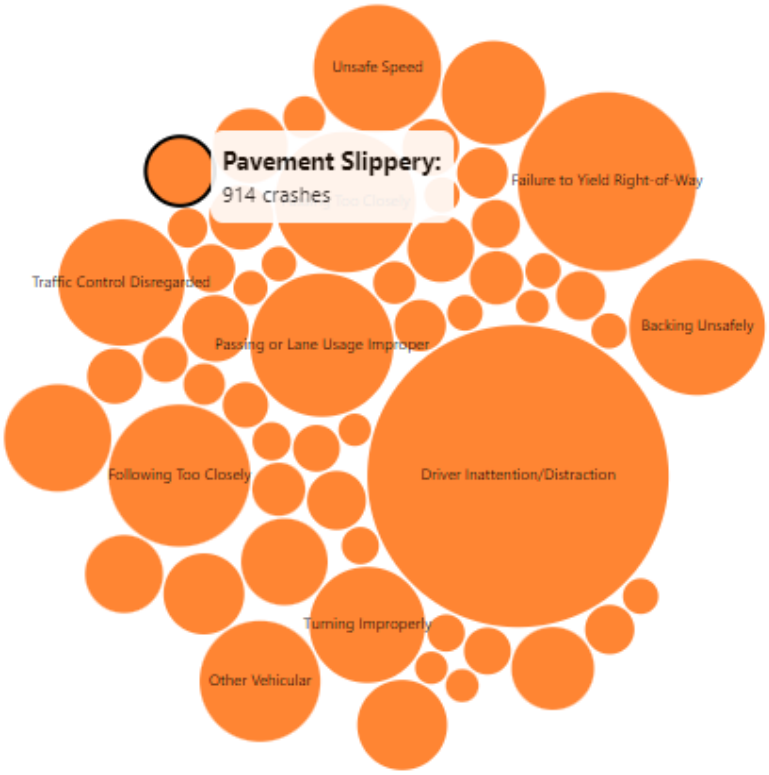


The multiple line time chart was intended to find out if there is any sort of correlation between collisions and time of day.

It is clearly seen that the number of collisions spikes at the times of 8:00am and 4:00 to 6:00pm. It can be inferred that these collisions are likely correlated with the traffic in the city from people going to and leaving their jobs each day.

This visualization was not designed to have any sort of filter directly on it, though when any of the other visualizations are filtered down, this chart adjusts with them.

Contributing Factor Bubble Chart

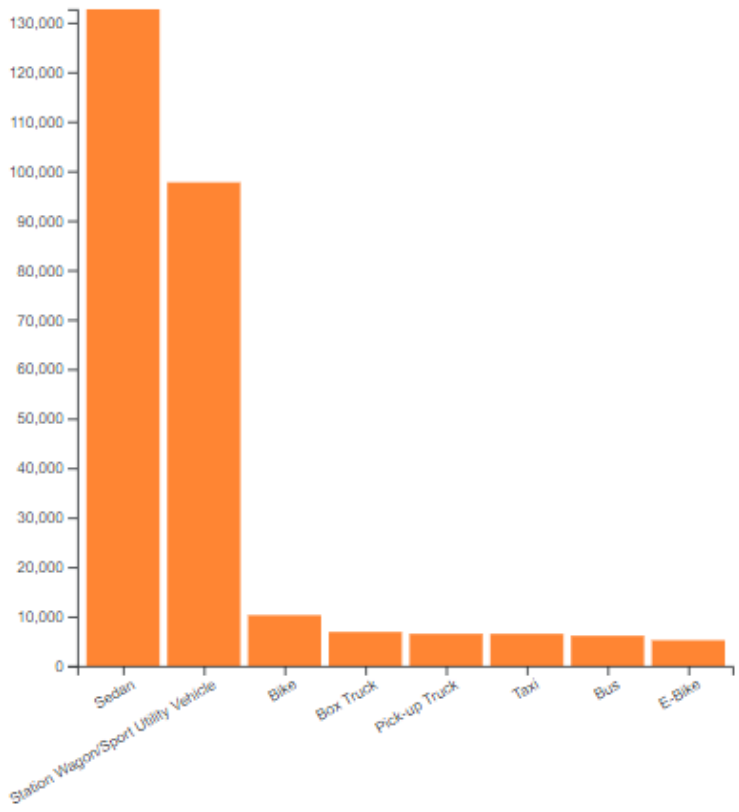


The bubble chart details the factors that most contribute to collisions in the city. The larger the bubble, the greater the number of collisions that have occurred because of that factor.

Only the top ten factors that contribute to collisions have their names listed in order to keep the visualization clean.

Like the borough map, these bubbles can be selected to drill down the visualizations to show only data linked to the contributing factor selected.

Vehicle Bar Chart



The bar chart visualizes the vehicles most likely to be involved in a collision based on historical data.

If a bar is selected, each graph will drill down to only show data for collisions where that vehicle is involved.

For example, if ‘bike’ is selected, the time graph will adjust and show that most injuries and deaths are bikers, not those in other vehicles.

Evaluation:

The use of visualizations provides key insights into the collisions within New York City and helps inform users on where, why, and how collisions occur. The visualizations help to answer key questions related to the distribution of collisions and identify high risk scenarios that could lead to collisions.

Though the visualization effectively answers the original questions that drove the project, it can definitely be improved upon to answer further questions. For example, an advanced map of the boroughs with street views showing hotspots for collisions could play a key role in further enhancing this visualization. It could give users access to in depth details in order to help avoid collisions in the future and allow us to analyze patterns and hotspots in the map to determine further factors that contribute to collisions such as road conditions and weather.

Overall, the visualization effectively achieved the goals of the project in analyzing the collision data in New York City. This project could be a stepping stone into a greater project that analyzes not just New York City, but possibly the entire state or country in order to contribute to creating safer and more efficient roads.