NYC Motor Vehicle Collision Data Analysis: Where is New York Safe?

Nick Herman

R Data Analysis Project

11/03/2023

Dataset(s)

The city of New York provides open data on many topics. The Motor Vehicle Collisions data table has information from all incidents where there is an injury, death or $1000+ of damage, requiring the NYPD to submit form MV104-AN.

Each row in the datatable is references specific crash event and provides location and time information, descriptions of the vehicle(s) involved, potential contributing factors for each vehicle, and the number of injuries/deaths which occurred.

https://data.cityofnewyork.us/Public-Safety/Motor-Vehicle-Collisions-Crashes/h9gi-nx95

Additionally from NYC Open Data, a table containing census information on population was merged with the above table in order to determine the density of accidents in each borough.

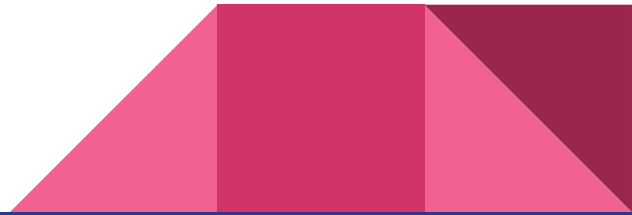
https://data.cityofnewyork.us/City-Government/New-York-City-Population-by-Borough-1950-2040/xywu-7bv9

Data Cleaning

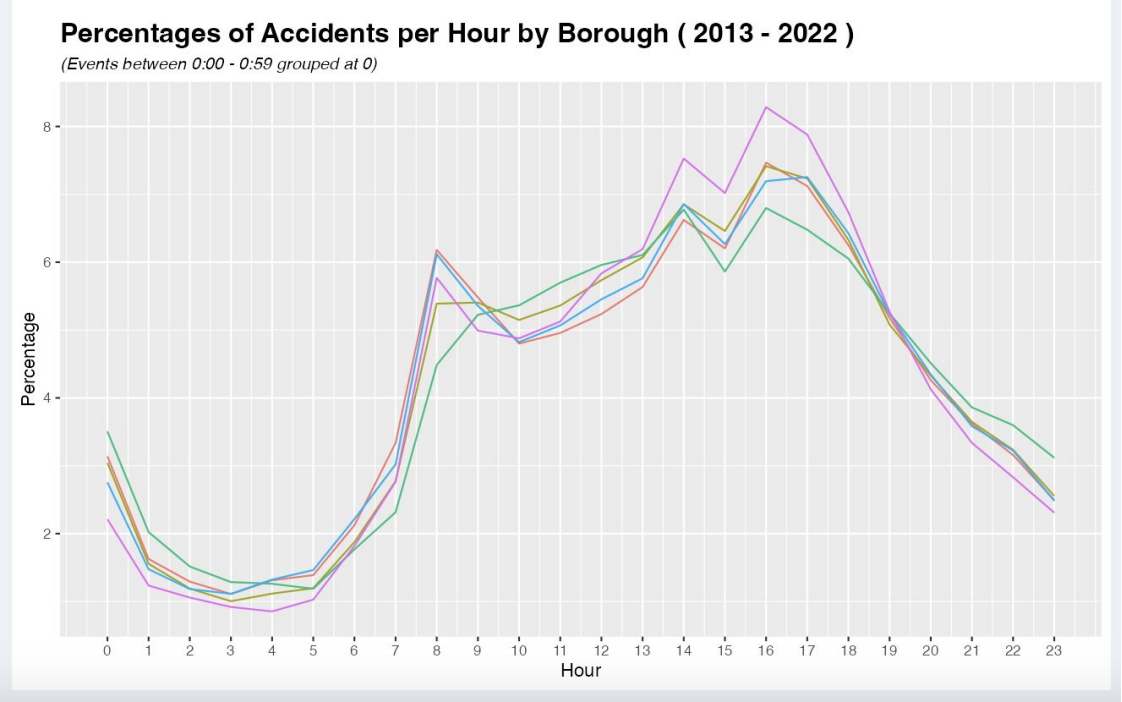
There were lots of rows with missing location, borough, or zip code information. The incidents with missing locations were removed. The incidents with missing boroughs were relabeled as unlisted. This allows for the data to still be included in the maps.

There were descriptive columns for the vehicles involved in the crashes with over 2000 different vehicle types listed. I took all the entries which appeared over 50 times and sorted them into the following categories: Bicycle/Scooter, E-Bicycle/Scooter, Motor Bike, Other/Unknown, Passenger Car, Truck. The large majority of the data is in the Passenger Car category.

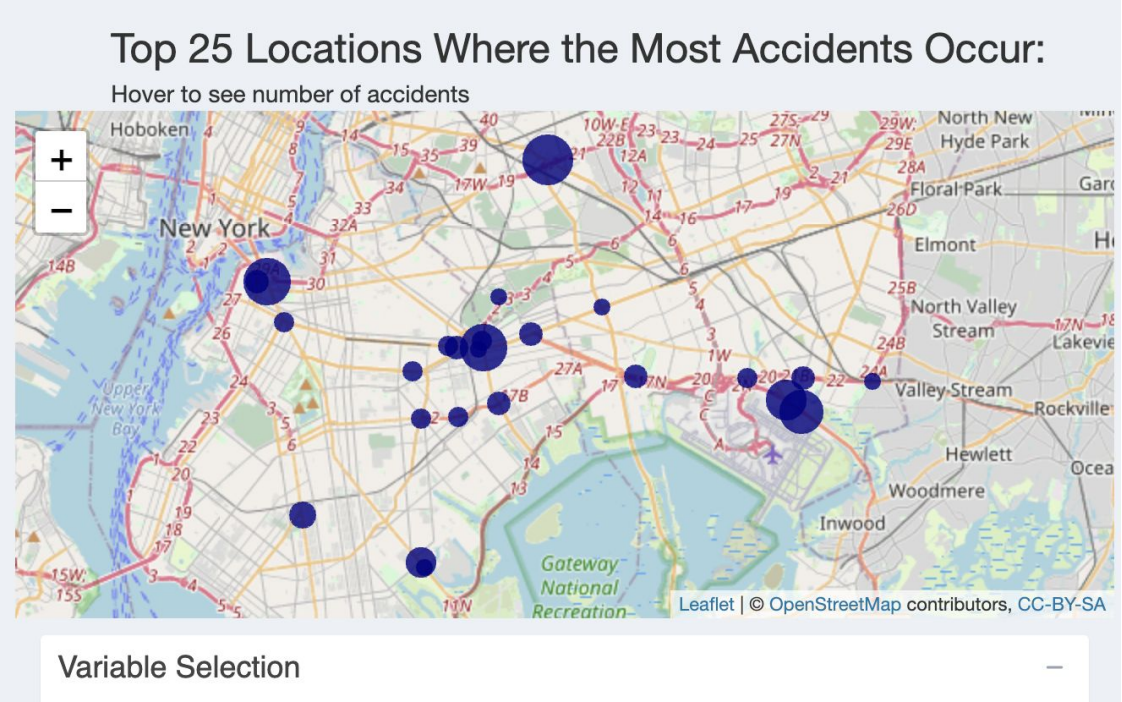
A few miscellaneous points with clearly incorrect locations were removed, but there are many more I only discover accidentally as I create new maps.



Shiny App

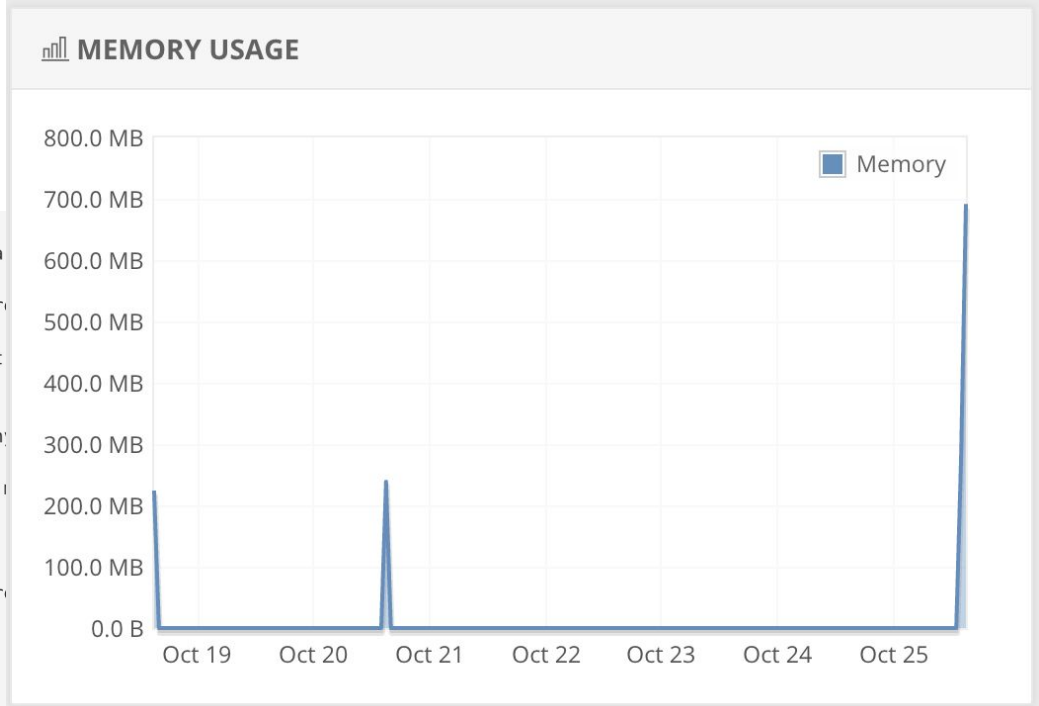

Percentages of Accidents per Hour by Borough (2013 - 2022)
(Events between 0:00 - 0:59 grouped at 0)
Percentage
8
6
4
2
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Hour

Shiny App


Top 25 Locations Where the Most Accidents Occur:
Hover to see number of accidents
Hoboken
New York
North New Hyde Park
Floral Park
Elmont
North Valley Stream
Valley Stream
Rockville
Hewlett
Woodmere
Inwood
Gateway National Recreation
Leaflet | © OpenStreetMap contributors, CC-BY-SA
Variable Selection

Out of Memory

```
2023-10-25T14:59:54.533700+00:00 shinyapps[10223466]:
2023-10-25T14:59:54.538853+00:00 shinyapps[10223466]: Attaching package: ‘data
2023-10-25T14:59:54.544368+00:00 shinyapps[10223466]:
2023-10-25T14:59:54.549968+00:00 shinyapps[10223466]: The following objects ar
2023-10-25T14:59:54.555780+00:00 shinyapps[10223466]:
2023-10-25T14:59:54.561562+00:00 shinyapps[10223466]:     between, first, last
2023-10-25T14:59:54.567197+00:00 shinyapps[10223466]:
2023-10-25T14:59:54.572820+00:00 shinyapps[10223466]:
2023-10-25T14:59:54.578356+00:00 shinyapps[10223466]: Attaching package: ‘shin
2023-10-25T14:59:54.584061+00:00 shinyapps[10223466]:
2023-10-25T14:59:54.589474+00:00 shinyapps[10223466]: The following object is
2023-10-25T14:59:54.594836+00:00 shinyapps[10223466]:
2023-10-25T14:59:54.600012+00:00 shinyapps[10223466]:     runExample
2023-10-25T14:59:54.605443+00:00 shinyapps[10223466]:
2023-10-25T14:59:54.610983+00:00 shinyapps[10223466]: The following objects ar
2023-10-25T14:59:54.617027+00:00 shinyapps[10223466]:
2023-10-25T14:59:54.622455+00:00 shinyapps[10223466]:     removeClass, show
2023-10-25T14:59:54.627957+00:00 shinyapps[10223466]:
2023-10-25T14:59:54.633541+00:00 shinyapps[10223466]:
2023-10-25T14:59:54.638836+00:00 shinyapps[10223466]: Attaching package: ‘shinywidgets
2023-10-25T14:59:54.644694+00:00 shinyapps[10223466]:
2023-10-25T14:59:54.650534+00:00 shinyapps[10223466]: The following object is masked from ‘package:shinyjs’:
2023-10-25T14:59:54.655945+00:00 shinyapps[10223466]:
2023-10-25T14:59:54.661570+00:00 shinyapps[10223466]:     alert
2023-10-25T14:59:54.666974+00:00 shinyapps[10223466]:
2023-10-25T14:59:54.672239+00:00 shinyapps[10223466]: The following object is masked from ‘package:shinydashboardPlus’:
2023-10-25T14:59:54.677668+00:00 shinyapps[10223466]:
2023-10-25T14:59:54.683063+00:00 shinyapps[10223466]:     progressBar
2023-10-25T14:59:54.688555+00:00 shinyapps[10223466]:
2023-10-25T15:00:18.497125+00:00 shinyapps[10223466]: Container event from container-8776991: oom (out of memory)
```



Future Work

Data cleaning for clear analysis removed a subset of the data that could lead to important insights. There were many rows with missing/inaccurate information. Some with gps coordinates that didn't match the borough or zip code they were listed under. With proper gps info the zip code and borough could be ascertained and the dataframe updated.

> Allowing the user to either exclude the data with missing information or replace it with approximate/averaged data in missing columns might also help learn about where/when/why the officers filing the reports leave certain fields blank.

Adding additional data might lead to now and interesting insights. **Weather** data could help determine if rainy days lead to more accidents. Conditions might be worse for cyclists but there would also be less on the road. **Demographic** data for the areas as well as the people involved in the accidents would also be interesting to study to see who is impacted the most be the events recorded in the dataframe.

I think the app would also benefit from the ability to play through images or export gifs of user selected date/time ranges and areas to show how the most dangerous spots change over time.

Incorporation of the population data into the maps would help better visualize the levels of danger.