

NYC Motor Vehicle Collision Shiny Project

...

Hua Yang
Jan 31, 2018



Why motor vehicle collision?

Critical issue:

Life-or-death situation

Relevant to everybody's daily life

Inevitable



Core interest:

“What can we learn from these info to better prevent and/or avoid collisions?”

Personal interests

How is the data look like on a map? Can we find particularly more dangerous/risky regions of concern? What their collision history data look like, can we get any useful insight?...

Esp for cases with walker injured/killed and the cases with people killed in general, are their distribution over location and/or time shows any significant different pattern/feature than the overall cases? Can we identify particular dangerous spots/areas for pedestrians, cyclists, or the lethal collisions, etc. ? ...

What are the top, say 20, most often seen collision causes and involved types of vehicles? What can learn from it? ...

General questions

What is the trend of total number of collisions from year to year? What can we predict for 2018?

What is the composition ratio of different types of victims (pedestrian, cyclist, motorist) and different levels of severities (no hur, injured, lethal)?

Was the situation different for different boroughs (5 boroughs of NYC)? What about different days in a week, hours in a day, month in a year? ...

What data?

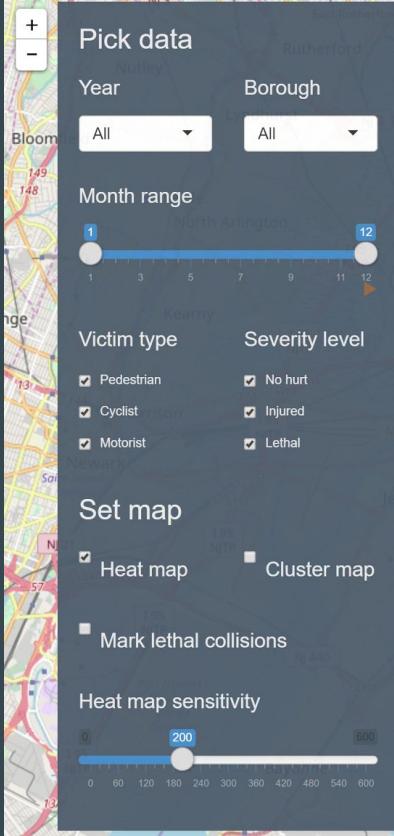
“NYC OpenData” Motor Vehicle Collision

- NYPD collision records since Jul 1, 2012
 - Take only full year data 2013-2017, random sampled by 5 times for demo.
 - Total ~150k collision records
-
- Date, time, location, borough
 - Number of injured, killed: pedestrian, cyclist, motorist
 - Causes
 - Vehicle info

Project objective:

1. Develop an interactive map tool to easily check and explore collision case info on a real city map.
2. Do some preliminary exploratory data analysis to get overview results on interested questions.

Interactive Map Tool



For data, can select

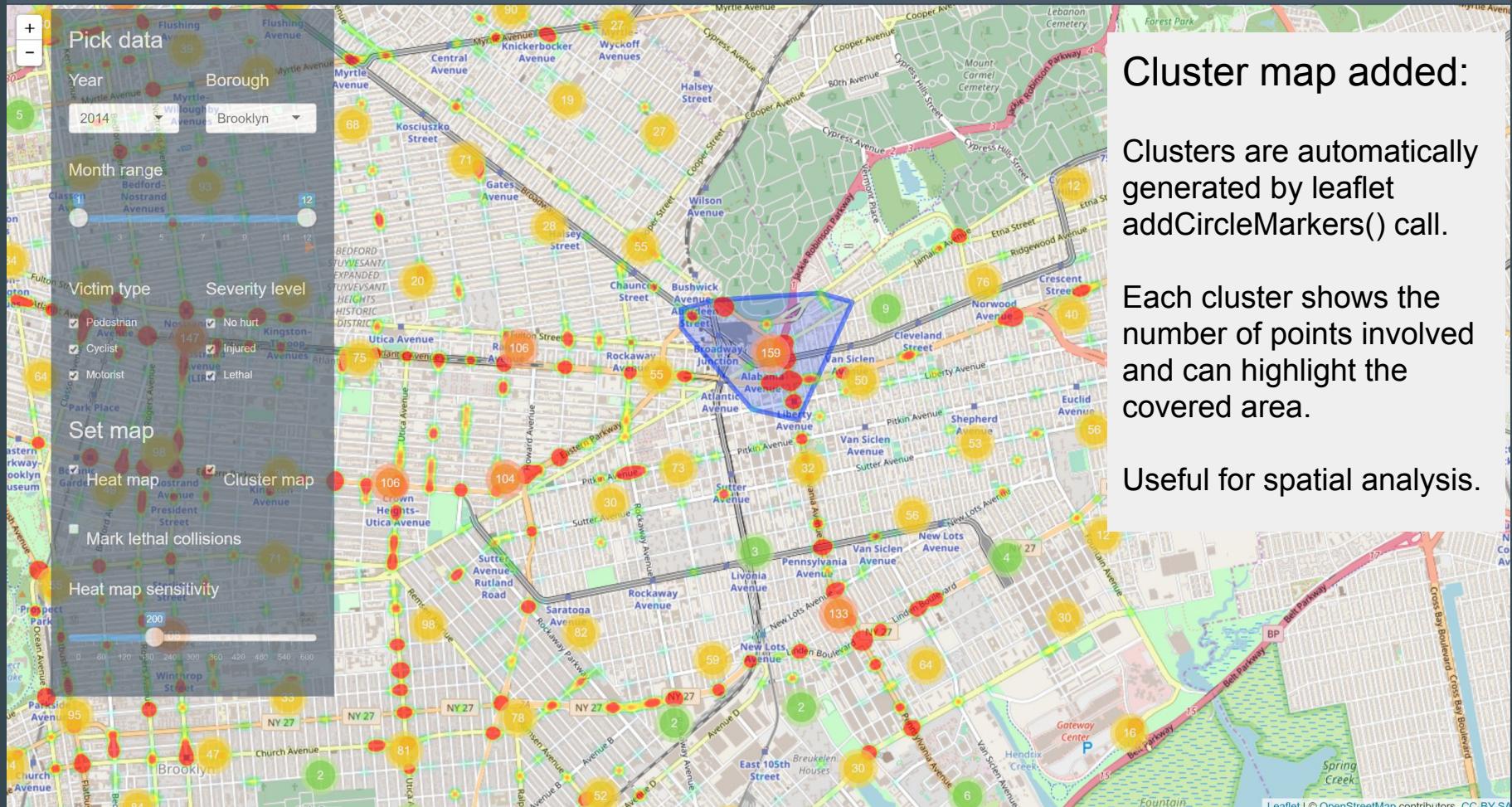
- Year, borough, month range, victim type (ped, cyc, mot), severity level (no hurt, injured, lethal).

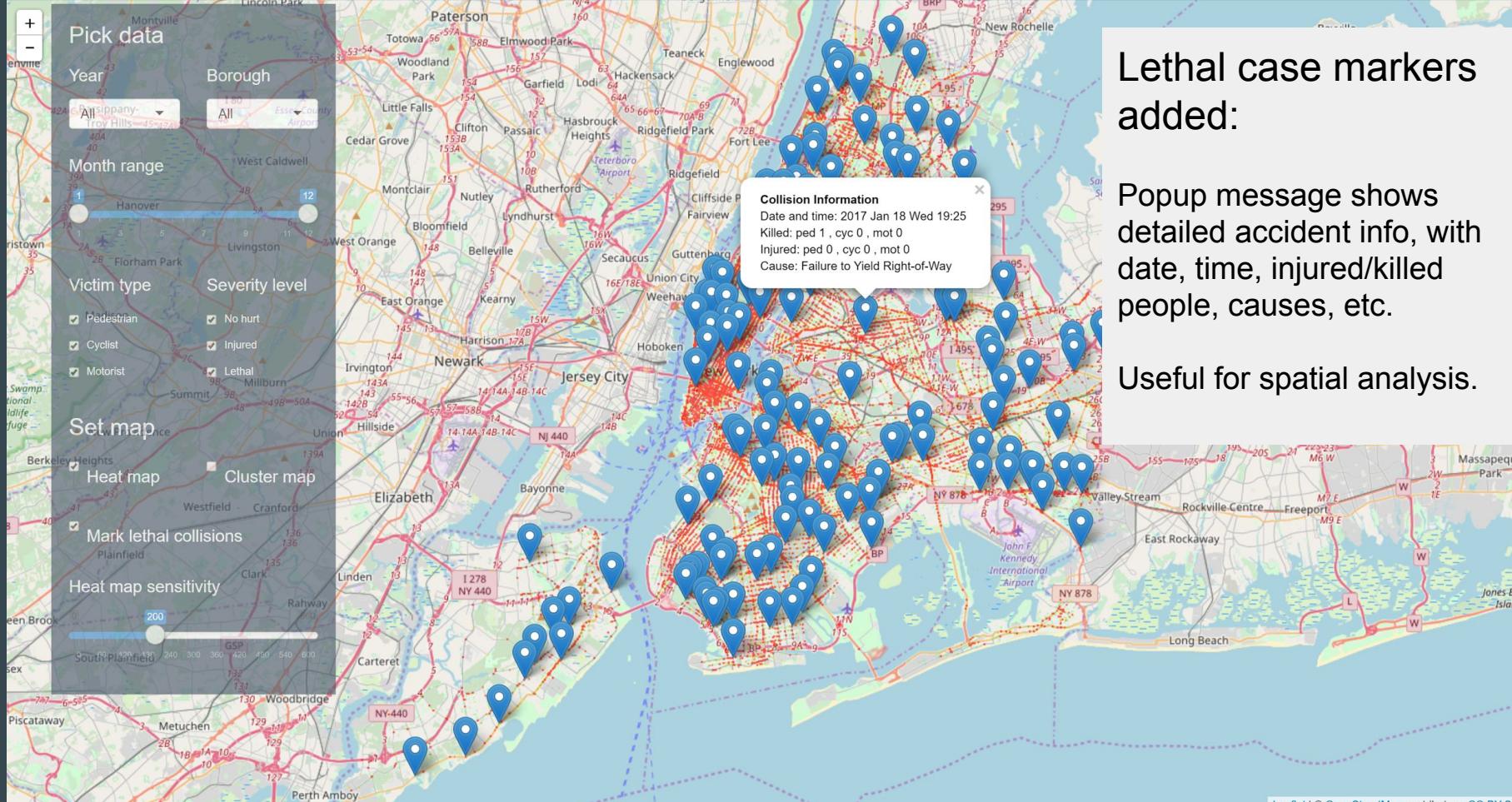
For map, can see

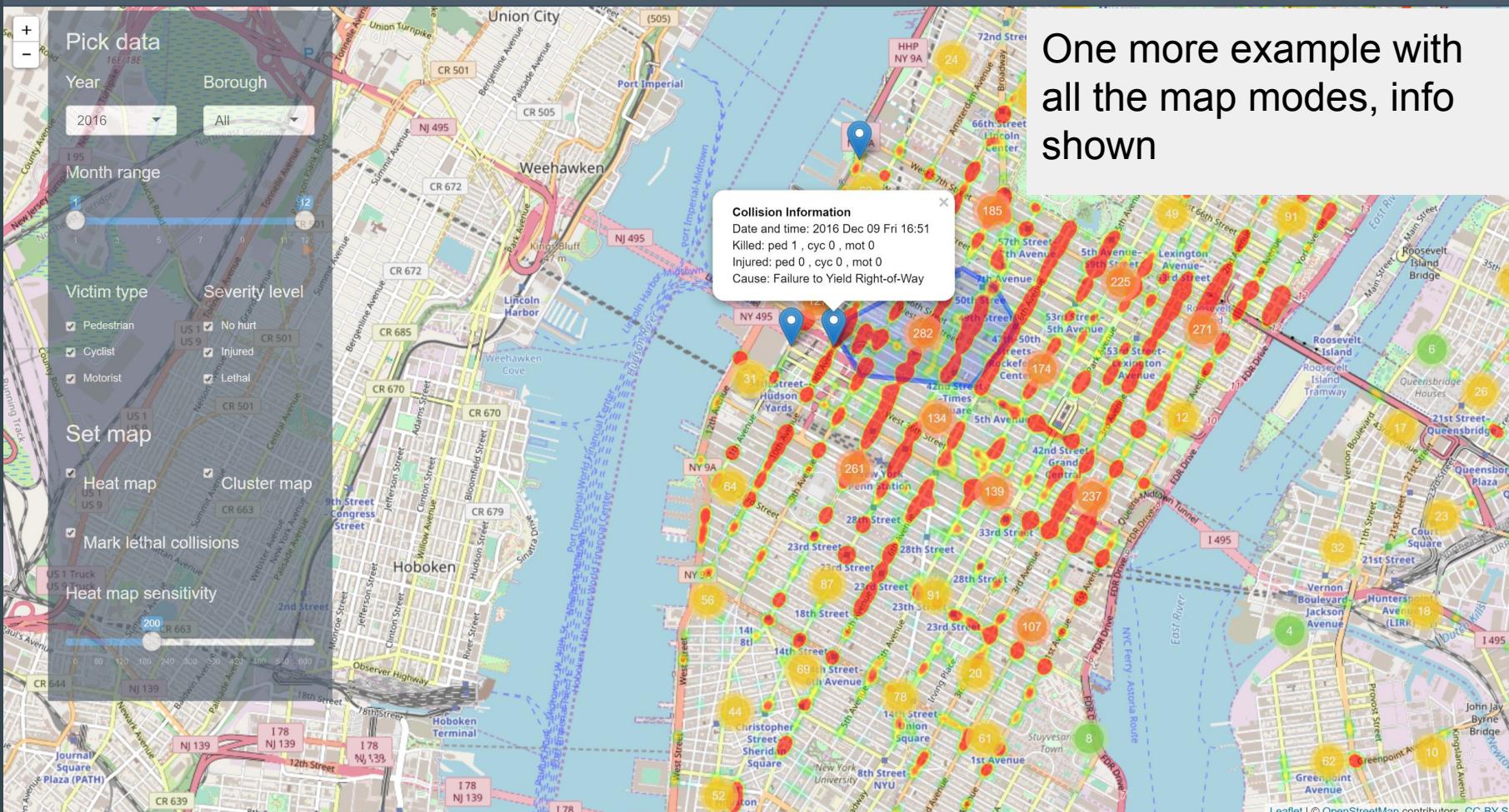
- Heat map (default)
- Cluster map
- Mark out lethal collisions

Can be all shown at the same time.

Can be animated with a looped slideshow with month range iterated.

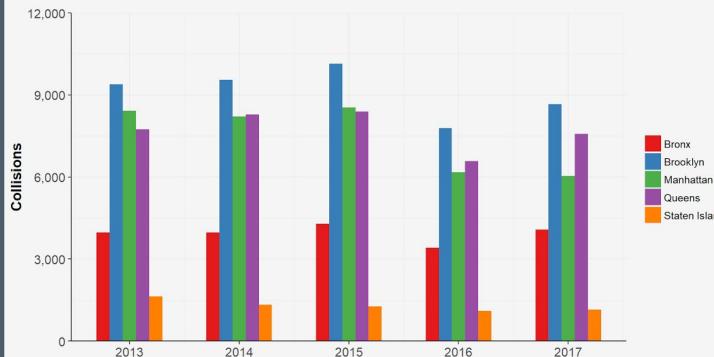
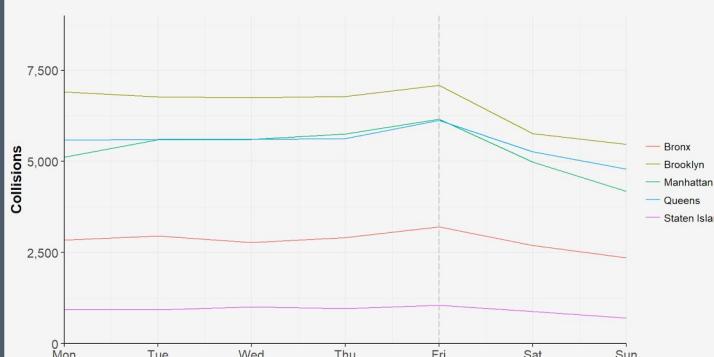
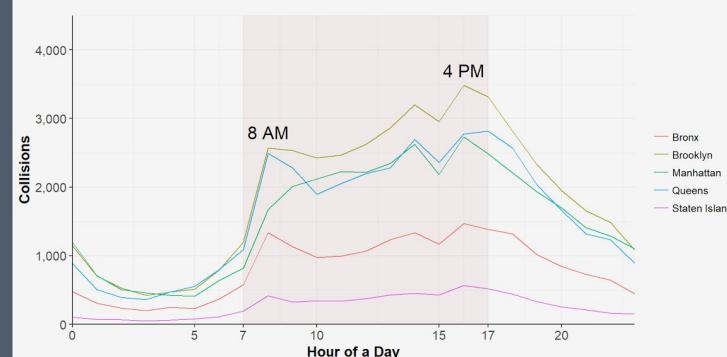






One more example with all the map modes, info shown

Time Factors

Collisions by Borough over Years**Collisions by Month****Collisions by Day in a Week****Collisions by Hour of a Day**

Observations

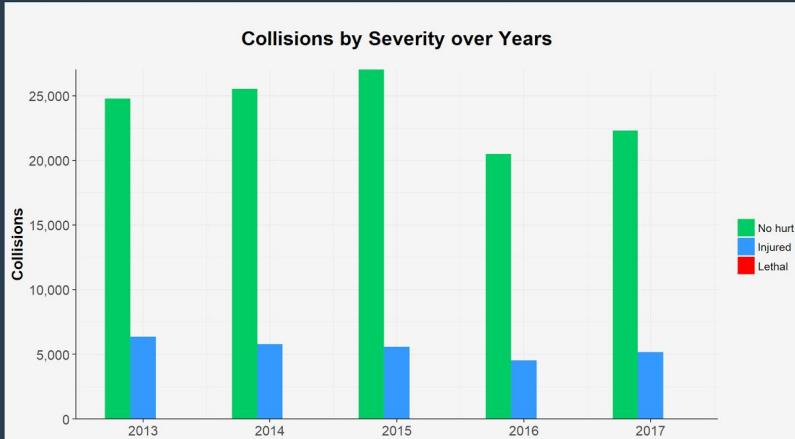
- Yearly-wise, gradual increase since 2013, drop at 2016, but then increase again at 2017. So, what to guess for 2018? ...
 - Month, weekday, hour factors show results generally well aligned w/ common sense.
 - Interestingly, Friday looks a little more like a peak day of a week in terms of number of collisions.
-

Severity & Victim Analysis

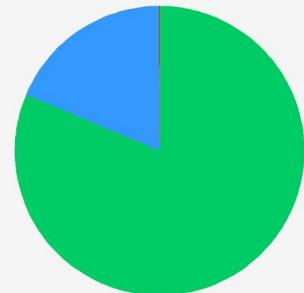
Severity data

Year

Borough



Collision Severity Composition

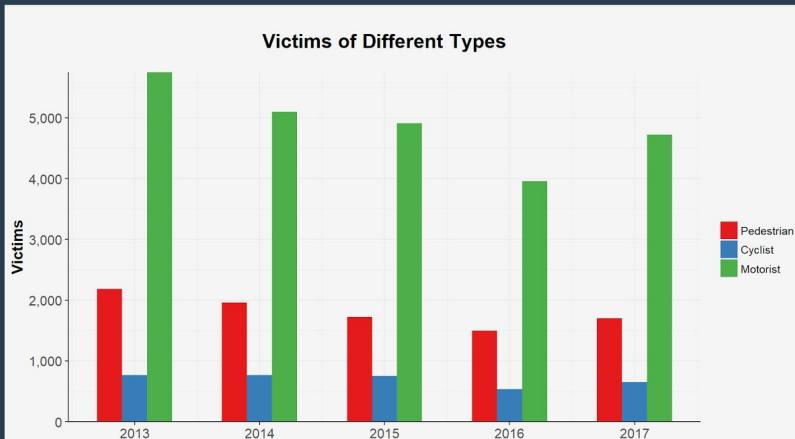


- No hurt
- Injured
- Lethal

Victim data

Year

Borough



Collision Victims Composition



- Pedestrian
- Cyclist
- Motorist

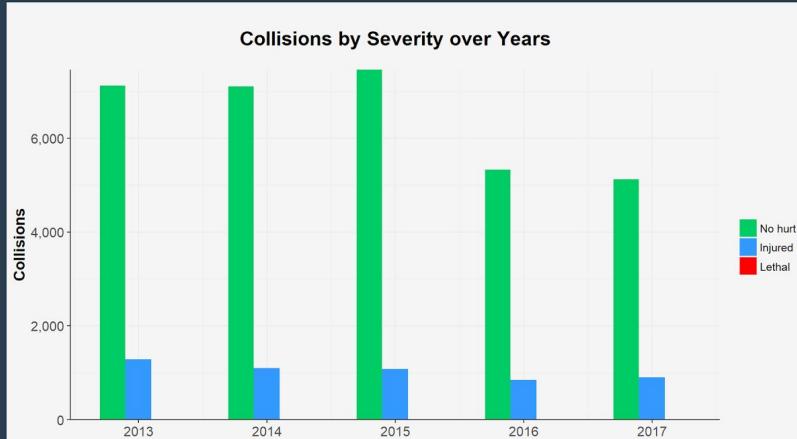
Observations

- Most cases are “No-hurt”, much higher ratio than “Injured” cases, while “Lethal” collisions are very rarely seen.
 - In terms of victims, pedestrians consists of a significant portion of the total collisions, and is significantly higher than that of cyclists.
-

Severity data

Year

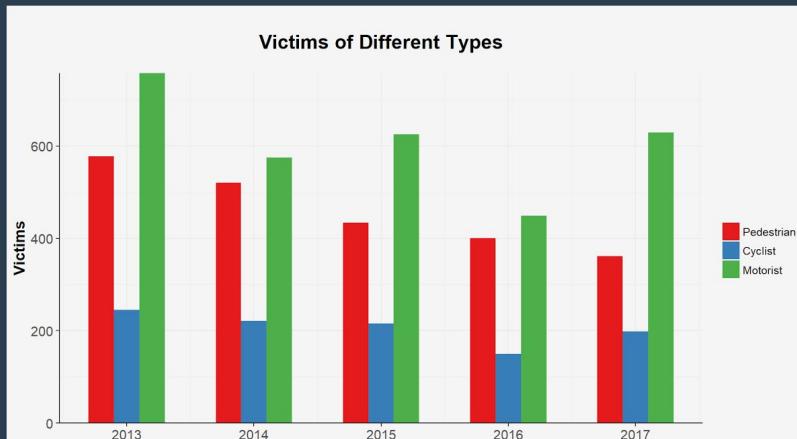
Borough



Victim data

Year

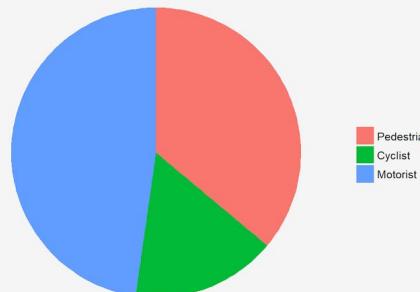
Borough



Collision Severity Composition



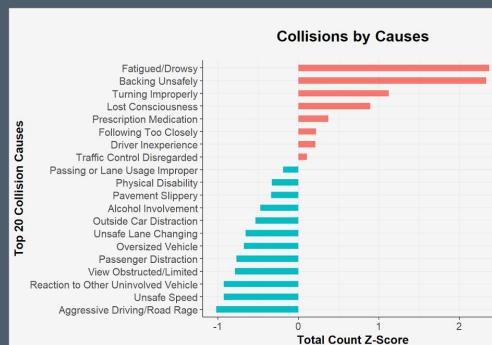
Collision Victims Composition



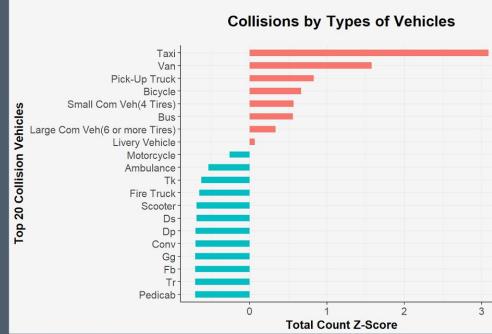
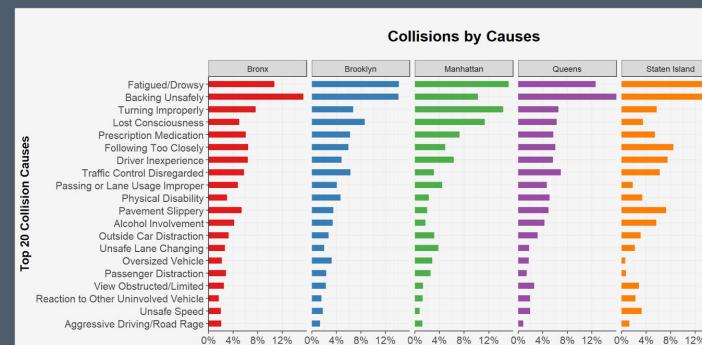
Observations

- Interestingly, Manhattan has a much higher ratio of pedestrian victims than those of the other 4 NYC boroughs, due to its special situation of highly crowded skyscraper buildings.
 - And because of this, its “no-hurt” ratio is also more higher than those of the other boroughs.
-

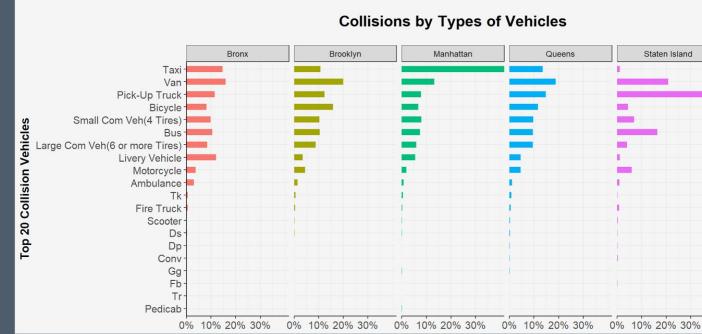
Cause & vehicle analysis



Note: To be more informative,
excluded two most common top items:
"Driver Inattention/Distraction"
"Failure to Yield Right-of-Way".



Note: To be more informative,
excluded two most common top items:
"Passenger Vehicle".
"Sport Utility / Station Wagon".



Top recorded causes

- Driver inattention/distraction
- Failure to yield right-of-way
- Fatigue/Drowsy
- Backing unsafely
- Turning improperly
- Lost consciousness
- Prescription medication
- Following too closely
- Driver inexperience
- Traffic control disregarded

Top involved types of vehicles

- Passenger vehicle
- Sport utility/station wagon
- Taxi (esp for Manhattan)
- Van
- Pick-up truck
- Bicycle
- Small com veh (4 tires)
- Bus
- Large com veh (6 or more tires)
- Livery vehicle
- Motorcycle
- Ambulance (even see firetruck)

Observations

- Besides the commonly known reasons of bad driving habit/skills, a big portion of top causes are related to mental unconsciousness/fatigue/drowsiness, etc.
 - Look at top involved vehicles, can find most of them are commercial vehicles.
 - Looks like reasonably high correlation between these two factors. Deserve further study...
-

What next?

Further/deeper analysis

- Time series analysis to predict future trend
- Correlation analysis between drowsy and commercial drivers
- Correlation analysis w/ other data sets e.g. bad/extreme weather data, special event/celebration data, etc. w/ collision dataset ...

Takeaways

Shiny app

- Interactive map tool
- Preliminary data analysis

Road safety advice

- Cautiously assume/take right-of-way
- Watch out drowsy/commercial drivers

Suggestions for city government

- New ways and/or stronger regulations to better prevent drowsy driving, unsafe backing, improper turning, etc. ?



Thank you!