Collision Hotspots Project Proposal

Bartosz Jerzy Bonczak (bjb417), Awais Malik (am5801) & Radu Stancut (rs1933)

Introduction

Last year, 268 fatal traffic crashes occurred in New York City.[1] In order to achieve "Vision Zero" (the ideal scenario where NO fatal traffic accidents take place), the New York City Police Department (NYPD) will need to better understand city traffic patterns in order to prioritize allocation of resources to any collision "hotspots." The objective of this project is to provide the NYPD with the desired enhanced cognition of traffic collision patterns in the city by identifying these hotspots and correlating them with specific enforcement-related causes (such as DUIs and cell phone use). We hope our visualization tool will allow the NYPD to aggregate and visualize the available data in a way that will be easier to understand and explore than before.

Related Work

Much of the focus and work on car crashes/collisions have dealt with the technical aspects of making driving safer, focused on highway safety, or looked at the issue at an inter/national level. The exciting aspect of this project is the leveraging of data to help better understand the phenomenon of vehicle collisions specific to New York City. Despite the New York-centric perspective inherent in the forthcoming work the available data from near and far is informative as it relates to the seriousness of the problem and ways of tackling the issue.

Injuries and deaths related to motor vehicle accidents take a heavy toll on communities worldwide; according to the World Health Organization 3,400 people die each day around the globe.[2] The United States has seen an yearly death toll anywhere between 30,000 to 40,000 over the past decade, with another two and a half million injured.[3] Additionally, road traffic injuries are a leading cause of deaths around the world for children and young adults.[4]

---

[1] http://project.wnyc.org/traffic-deaths/
[2] http://www.who.int/violence_injury_prevention/road_traffic/en/
[3] http://www-nrd.nhtsa.dot.gov/Pubs/811856.pdf
[4] http://www.who.int/violence_injury_prevention/publications/road_traffic/world_report/chapter1.pdf?ua=1

The crisis surrounding car accident deaths and injuries is a serious public health issue, one that all countries/cities must contend with (in fact, the name "Vision Zero" comes from a Swedish initiative started in the 1990's to deal with these issues). We view this project as a practical step towards better understanding the nature of NYC collisions by aiming to create an interactive visual platform that is meant to extract meaning and information not inherently accessible in its current state.

Data

The main dataset used for this project is NYPD Motor Vehicle Collisions available on the NYC Open Data platform[5] and can be downloaded directly from the NYC Open Data website in either CSV or JSON, among other types. The dataset comes from processed TrafficStat reports made by NYPD and was used inter alia as a basis for developing the Vision Zero View web application. It represents each collision reported to NYPD from 7/1/2012 onward. The dataset is updated daily with an one or two day delay, making it a dynamic dataset with more than 543,000 observations, as of 9th of March 2015. The data type is a flat table consisting of 29 attributes describing each collision, which can be grouped into 6 categories. Both categories and described attributes are listed in a table in the appendix (Table 1).

For the purpose of our project, the dataset will be treated as a static table limited to collisions from 07/01/2012 to 02/28/2015. The size of the generated file (CSV) is ~100MB. The dimensions of the filtered table is 541,764 rows by 29 columns. Because the project will be presenting spatial distribution, location information is crucial. Therefore we decided to focus only on the collisions that provided latitude and longitude. There are 455,575 such items which account for 84% of the total reported collisions over the selected time period.

Though location will be our primary spatial attribute, we anticipate using nearly all attributes in order to provide in-depth descriptive information and/or for aggregation purposes. Temporal features (DATE, TIME) will be used to show collisions in different time frames selected by users, while victim, contributing factor, and vehicle information will create additional layers to choose from.

---

[5] https://data.cityofnewyork.us/Public-Safety/NYPD-Motor-Vehicle-Collisions/h9gi-nx95

In addition to the NYPD Motor Vehicle Collisions Data we will be using geometry type datasets like New York City's shapefiles including borough boundaries, zip-codes, police precincts, street network etc. available through BYTES of the BIG APPLE database.[6] They will be used mostly to create a base map for the visualization.

Visualization Questions

The following is a list of preliminary questions we hope to answer with the support of our visualization:

1. Are there areas of the city seeing new collision hotspots?
2. Are there local hotspots that correspond to increases in enforcement-related causes (such as cellphone use, illegal drug use, or alcohol involvement)?
3. Which areas should the NYPD prioritize its resources toward?
4. Which enforcement-related causes should the NYPD pay particular attention toward?

Visualization Sketch

Figure 1 shows the sketch of our desired visual solution: an interactive heat-map of the number of collisions in each locality in New York City (spatial options include police precincts, zip codes and census tracts). Temporal options include selecting figures for a particular day of the week, week of the month, month of the year, a particular year in the dataset, as well as by day and night. The timeline scroll bar will assist in any temporal analysis (by day, month or year). Once a particular region is clicked (or the mouse is hovered over it), that region's ID (borough name and zip code/police precinct/census tract) will be displayed as well as a bar chart of the major causes of collisions.

At this point, the greatest anticipated challenge is allowing the visualization to surface hotspots without undue burden on user's time and effort. Supplemental charts to the map under consideration include, but are not limited to, tree maps, parallel coordinates, and time series displays. Options will be considered and tested with increased understanding of the data and feedback from the client on proposed visual idioms moving forward.
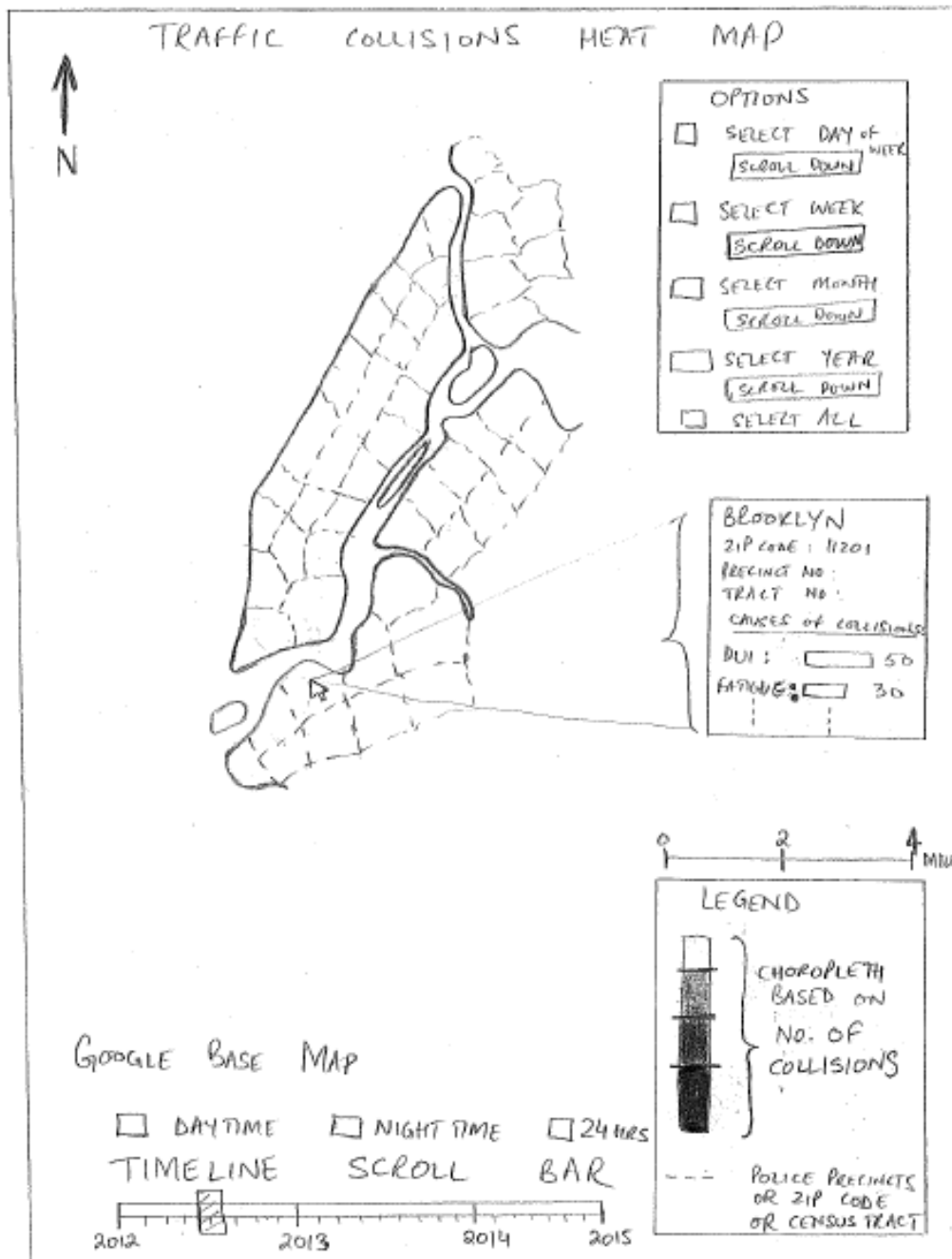
---

[6] http://www.nyc.gov/html/dcp/html/bytes/applbyte.shtml

Figure 1: Visualization Sketch

Work Plan

The first priority identified by our team is to better understand the data and prepare it for use moving forward. As referenced above we will be focusing exclusively on collisions that provide latitude/longitude information. The task of isolating these records, as well as identifying and dealing with additional anomalies, will fall under data cleaning. Two other exercises will be going on concurrently as the data is being readied: spatial joining and D3 structural preparation. In anticipation of the finalized collision table, team members will work with representative data subsets to work through the process of assigning each collision to a police precinct via lat/long coordinates, which will be critical in aggregating collision data at the client's desired spatial level, and testing the importing of data into a first pass skeletal map using D3.

Lastly, before moving forward and testing different visual idioms with the client, each team member will be responsible for data exploration in a tool of their choice, as well as Tableau. This exercise is meant to mirror the workflow of the sketches, where each team member worked independently and later came together to share ideas and insights.

All of the work tasks described are referenced in the table below, with the responsible team member identified. We stress that this is the preliminary work plan needed to better understand our data while we continue communicating with our client. It is believed that once the work above has been accomplished we will be in a stronger position to do the necessary data and task abstractions as well as any additional data transformations.

Task	Awais	Bartosz	Radu
Data Cleaning		x	
Data/Spatial Joining			x
D3 (map/plots)	x		
Data Exploration (Tableau/R)	x	x	x

Appendix 1

Table 1 Attributes and their description

Data category	Name	Format	Type	Description
Temporal identification	DATE	date format (MM/DD/YYYY)	sequential	day when collision happened
	TIME	plain text in 24 hours format (HH:MM)	cyclic	time at what collision happened or was reported
Geographical reference	BOROUGH	plain text	categorical	name of the NYC borough for collision location
	ZIP CODE	plain text	categorical	zip-code for collision location
	LATITUDE	number	diverging	geographical latitude in decimal format of collision location
	LONGITUDE	number	diverging	geographical longitude in decimal format of collision location
	LOCATION	geographical location (degrees)	diverging	georeferenced attribute locating accident
	ON STREET NAME	plain text	categorical	name of the street where collision happened
	CROSS STREET NAME	plain text	categorical	name of the nearest cross street where collision happened
	OFF STREET NAME	plain text	categorical	address where collision happened

Victims	NUMBER OF PERSONS INJURED	number	quantitative	self-explanatory
	NUMBER OF PERSONS KILLED	number	quantitative	self-explanatory
	NUMBER OF PEDESTRIANS INJURED	number	quantitative	self-explanatory
	NUMBER OF PEDESTRIANS KILLED	number	quantitative	self-explanatory
	NUMBER OF CYCLIST INJURED	number	quantitative	self-explanatory
	NUMBER OF CYCLIST KILLED	number	quantitative	self-explanatory
	NUMBER OF MOTORIST INJURED	number	quantitative	self-explanatory
	NUMBER OF MOTORIST KILLED	number	quantitative	self-explanatory
Contributing factors	CONTRIBUTING FACTOR VEHICLE (1 to 5)	plain text	categorical	factor most contributing to collision of vehicle number 1 to 5 involved in an accident grouped into 46 different categories (appendix 1)
Identification	UNIQUE KEY	number	ordinal	unique key number given to each collision
Vehicle type	VEHICLE TYPE CODE (1 to 5)	plain text	categorical	vehicle type involved in a collision grouped into 17

				different categories (appendix 2)

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

Appendix 2

List of categories for contributing factor

1.	Lost Consciousness
2.	Fatigued/Drowsy
3.	Pavement Slippery
4.	Driver Inattention/Distraction
5.	Unspecified
6.	Brakes Defective
7.	Glare
8.	Backing Unsafely
9.	Traffic Control Disregarded
10.	Physical Disability
11.	Failure to Keep Right
12.	Prescription Medication
13.	Failure to Yield Right-of-Way
14.	Other Vehicular
15.	Driver Inexperience
16.	Turning Improperly
17.	Outside Car Distraction
18.	Oversized Vehicle
19.	Illness

20.	View Obstructed/Limited
21.	Alcohol Involvement
22.	Passenger Distraction
23.	Reaction to Other Uninvolved Vehicle
24.	Steering Failure
25.	Other Electronic Device
26.	Unsafe Speed
27.	Fell Asleep
28.	Aggressive Driving/Road Rage
29.	Obstruction/Debris
30.	Unsafe Lane Changing
31.	Traffic Control Device Improper/Non-Working
32.	Windshield Inadequate
33.	Shoulders Defective/Improper
34.	Accelerator Defective
35.	Pavement Defective
36.	Passing or Lane Usage Improper
37.	Lane Marking Improper/Inadequate
38.	Cell Phone (hand-held)
39.	Tire Failure/Inadequate
40.	Drugs (Illegal)
41.	Other Lighting Defects

42.	Following Too Closely
43.	Animals Action
44.	Pedestrian/Bicyclist/Other Pedestrian Error/Confusion
45.	Headlights Defective
46.	Cell Phone (hands-free)

Appendix 3

List of categories for involved vehicles

1.	BUS
2.	PASSENGER VEHICLE
3.	LARGE COM VEH(6 OR MORE TIRES)
4.	SPORT UTILITY / STATION WAGON
5.	TAXI
6.	UNKNOWN
7.	VAN
8.	OTHER
9.	SMALL COM VEH(4 TIRES)
10.	PICK-UP TRUCK
11.	FIRE TRUCK
12.	LIVERY VEHICLE
13.	AMBULANCE
14.	MOTORCYCLE
15.	BICYCLE
16.	SCOOTER
17.	PEDICAB