Impact Of Reckless Driving On Low-Income New Yorkers

Jason Chang
Information Science, Courant Institute
New York, NY
jasonchang@nyu.edu

Scott Wyman Neagle
Computer Science, Courant Institute
New York, NY
scott.neagle@cs.nyu.edu

Raghav Singhal
Mathematics, Courant Institute
New York, NY
rs4070@nyu.edu

***Abstract*—**

We will analyze NYPD crash data, 311 street complaints, and American Community Survey commuting and income data to determine if low-income New Yorkers are disproportionately impacted by reckless driving and dangerous street conditions.

analytics, commuting, transportation, cycling, walking, income

I. Introduction

When commuting and transportation policy are discussed, issues of class are often raised. Although there isn't a one-to-one relationship between income and class, we can gain some insight into this issue by considering the income levels of people using different means of transportation. It's well established that bike commuters have the lowest median income of any commuting group. This analytic will map crashes and street infrastructure complaints in New York City. We will group these by US Census Public Use Metadata Areas and see if PUMAs with a high number of low-income people using non-motorized commuting modes also have high numbers of street complaints and crashes that injure low-income pedestrians and cyclists. We will attempt to determine if low-income New Yorkers are disproportionally impacted by reckless driving, and to identify areas where they could benefit from targeted traffic enforcement and improved infrastructure.

Our application can be used by policymakers and urban planners to target infrastructure improvements to the most vulnerable populations. We will attempt to gain insight on New York neighborhoods that are underserved by transportation infrastructure.

II. Motivation

As New York City's transportation infrastructure continues to crumble, public dollars must be spent in more efficient ways. Cycling and walking are the two lowest impact and cheapest ways to move people. However, they are underutilized. People who could bike or walk cite danger as the most common reason that they don't. [3]

Our work identifies community districts in NYC where cycling and walking are most critical transportation options. We focus on low-income commuters because these workers often don't have access to more expensive transport options. We then focus on districts with a high number of crashes injuring pedestrians and cyclists and also a high number of street condition complaints.

III. Related Work

McKenzie, writing for the US Census, found that commuting by non-motorized means was much more common among very low-income people. The author was challenged by the fact that the sample size of people who use non-motorized means of commuting to work is relatively small compared to those who drive or use public transit. For many groups, less than one percent use non- motorized commuting modes. However, he was able to identify certain trends among these commuters. There is a strong correlation between low-income households and non-motorized commuting. with over five percent of commuters from households earning less than ten thousand dollars per year walking or biking to work. [1]

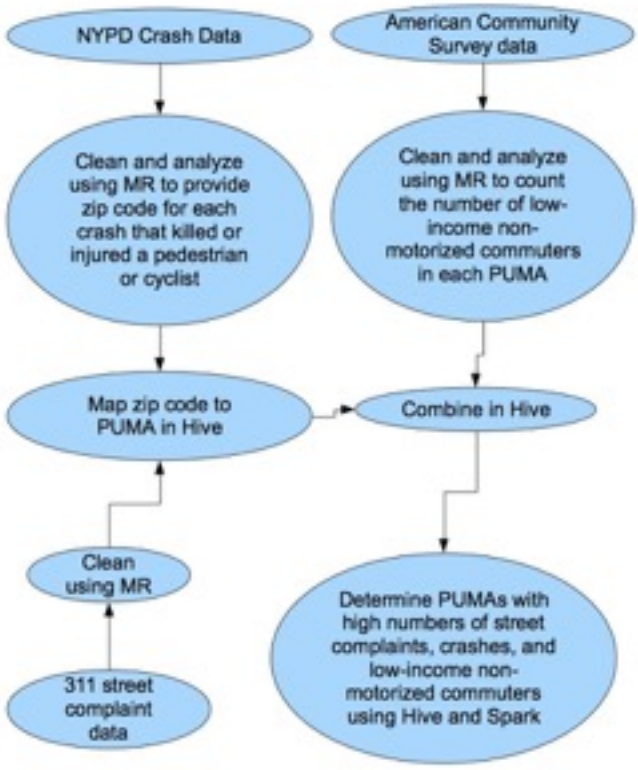
Researchers have always been finding ways to better understand the factors of crashes and hope to predict the likelihood of crashes and provide direction for policy makers to reduce the number of crashes. But some crash-frequency data can pose problems in terms of data characteristics, such as dispersion issues, sampling size issue, omitted-variables bias, fixed parameters, etc. To deal with these data and methodological issues, several innovative modeling methods have been introduced for analyzing crash-frequency data. These modelings have their strengths and weaknesses and have their own statistical fit/or predictive capabilities. The two most common methods used for crash-frequency models are maximum likelihood estimation and Bayesian methods. Some exciting statistical approaches like random-parameter models, finite mixture models, Markov switching models all hold great promise in improving our understanding of the factors that affect the frequency of crashes. However, it is important to know that the work has been inherently limited by the available data which has been restrictive like detailed driving data (acceleration, braking and steering information, driver response to stimuli, etc.) and crash data (from vehicle black-boxes). When these data become available, it will open up an entire

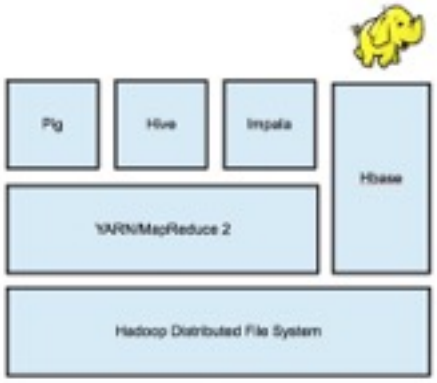
new direction of research and provide valuable insights into the factors of crash-frequencies. [2]

Jain, Ahuja, Anuranjana, and Mehrotra used data mining methods to figure out the main causes of road accidents in several different regions. This study uses the educational background of the drivers along with other factors to give region-based causes which are vital for the facilitating road safety in those areas. The main thrust of the paper is that it combines Data Mining techniques to first identify accident prone states and regions in India by using K-Means Cluster Analysis and then builds a decision tree model to predict the cause of that accident. The features of the dataset included number of deaths in an accident, whether motorbikes, cars or pedestrians were involved, blood-alcohol level of the driver, educational qualifications, etc. K-Means provided insight into which states have similar causes. Similar policies can be implemented there. The states in each cluster had similar geographic features, for instance Rajasthan, Gujarat and Madhya Pradesh were in a similar cluster (All three states are mostly desert states). The decision tree model is a classification model that classifies the regions into how accident-prone the regions are and these methods can be further generalized into determining which season of the year is more accident-prone and which hour of the day is more accident prone. [4]

Cycling as a means of transportation continues to see strong growth in New York City as the city population sets new records and subway transit - historically the primary means of transport for most New Yorkers - is increasingly unreliable. Century old switch systems deteriorate and work continues to repair the damage done by Hurricane Sandy. Recent studies have shown that New York City now has more bike commuters than any other American city. [5] The increase in cyclists has been accompanied by a significant drop in cyclist deaths per bike trip. This safety increase is largely confined to areas where there have been improvements in bike infrastructure. Nearly 90% of cyclist deaths have occurred on streets with no bike infrastructure improvements. Last year, 0 cyclists were killed in “protected” bike lanes, meaning those lanes separated from motor vehicle traffic by Jersey barriers, a parking lane, or other permanent fixtures like fencing or planters.[7] The outer borough neighborhoods of East Elmhurst, Jackson Heights, Glendale, Ridgewood, Corona, Bed-Stuy, East New York, Borough Park, East Flatbush, Flatbush, Midwood, Brownsville, Gravesend and Manhattan Beach account for a quarter of “cyclist fatalities and serious injuries”. These neighborhoods also are poorly served by bike infrastructure, accounting for only 17% of bike lanes. They are also some of the poorest neighborhoods in New York and indeed the country. East New York has the lowest median income of any neighborhood in Brooklyn and it is among the lowest in New York City and nationally.[6]

IV. Design





V. Results

We had many challenges working with location data. The American Community Survey data has a Public Use Metadata Area column (which in New York City is equivalent to community district) but no other location column. The NYPD crash data and the 311 complaint data have many location columns, including latitude, longitude, and zip code, but they don't include PUMA or community district. We attempted to use Hive UDFs provided by ESRI to put each crash and complaint into a PUMA using latitude and longitude. We were able to process the U.S. Census PUMA shapefile that defines the PUMAs into a plain text file using pyshp. It was then trivial to convert each shape into GeoJSON format and pull the shapes into Hive. However, we ran into issues using ESRI's Hadoop spatial framework. The framework contains a user defined type called ST_Geometry, and the first step in putting coordinates inside shapes is converting strings in whatever format you have (in our case, GeoJSON) into the Hive type ST_Geometry. This was an issue for us due to the fact that the provided UDFs were designed for an earlier version of Hive than what is currently running on Dumbo, causing version conflicts. We were not able to create ST_Geometry objects and were blocked there.

Trying a different approach worked much better, although the resulting data is less accurate. Baruch College provides a crosswalk from zip code to PUMA for New York City (https://www.baruch.cuny.edu/confluence/display/geoportal/NYC+Geographies). After pulling the CSV file into Hive, it was trivial to calculate reasonably accurate totals for complaints and crashes in each PUMA. We were able to use the resulting tables to do analysis.

Top 5 areas in descending order of low-income non-motorized commuters
Brooklyn community district 12 - Borough Park, Kensington, & Ocean Parkway
Brooklyn community district 1 - Greenpoint & Williamsburg
Brooklyn community district 7 - Sunset Park & Windsor Terrace
Manhattan community district 3 - Chinatown & Lower East Side
Bronx community districts 3 & 6 - Belmont, Crotona Park East & East Tremont

Top 5 areas in descending order of crashes that killed or injured a pedestrian or cyclist
Manhattan community districts 4 & 5 - Chelsea, Clinton & Midtown Business District
Manhattan community district 6 - Murray Hill, Gramercy & Stuyvesant Town
Manhattan community districts 1 & 2 - Battery Park City, Greenwich Village & SoHo
Manhattan community district 3 - Chinatown & Lower East Side
Brooklyn community district 2 - Brooklyn Heights & Fort Greene

Top 5 areas in descending order of 311 street condition complaints
Staten Island community district 3 - Tottenville, Great Kills & Annadale
Staten Island community district 1 - Port Richmond, Stapleton & Mariners Harbor
Queens community district 7 - Flushing, Murray Hill & Whitestone
Staten Island community district 2 - New Springville & South Beach
Manhattan community districts 4 & 5 - Chelsea, Clinton & Midtown Business District

Notice that there is not one neighborhood that appears in all of the above tables. Most of the neighborhoods with high numbers of low-income non-motorized commuters are medium- to high-density neighborhoods with traditionally significant ethnic immigrant populations. These neighborhoods are not the poorest in the city, however. The neighborhoods with the lowest incomes in the city may be too far from significant commercial activity to have high numbers of people biking or walking to work. Neighborhoods with very low incomes also have high numbers of unemployed people, driving down the amount of commuters of all types. The neighborhoods with high numbers of crashes are very dense neighborhoods in lower Manhattan and "close-in" Brooklyn, which seems reasonable: Lots of people and cars existing together in a small space causes crashes. The neighborhoods where we see high numbers of 311 street condition complaints tend to be low-density suburban neighborhoods, which is somewhat surprising, although maybe these neighborhoods'

distance from the core commercial activity of the city makes them a low priority for the Department Of Transportation. Chinatown and the Lower East Side do have high numbers of non-motorized low-income commuters and crashes, but rank 25th in street condition complaints - in fact, this area has seen significant improvements in street infrastructure, most notably the "complete street" projects on Grand Street and Chrystie Street. Midtown is a stronger candidate for our hypothesis with very high numbers of crashes and street complaints and a relatively high number of low-income non-motorized commuters, ranking 13th. This area does have the West Side greenway, pedestrian plazas on Broadway, and protected bike lanes on 8th and 9th avenues, but this infrastructure is clearly inadequate for moving people around the neighborhood - a protected crosstown bike route is conspicuously missing.

Failing to find a strong link between these indicators, we considered looking at total non-motorized commuters rather than low-income non-motorized commuters.

Top 4 areas in descending order of non-motorized commuters
Manhattan community districts 4 & 5 - Chelsea, Clinton & Midtown Business District
Manhattan community district 6 - Murray Hill, Gramercy & Stuyvesant Town
Manhattan community districts 1 & 2 - Battery Park City, Greenwich Village & SoHo
Manhattan community district 3 - Chinatown & Lower East Side

These are the same areas that have a high number of crashes. We can see a strong case emerging for improving street infrastructure and traffic enforcement in Midtown. Interestingly, crashes correlate more strongly with non-motorized commuters with incomes over $25k/year. This correlation has a Pearson coefficient of 0.81895484, compared to the much weaker Pearson of 0.32650983 for the correlation between crashes and low-income non-motorized commuters. Also, these areas have a very high ratio of high-income to low-income commuters, about 4:1.

The Bronx is the poorest borough, and the second smallest (after Staten Island). We found that the Boogie Down has the highest rate of crashes per non-motorized commuter. In this sense, it is a more dangerous place to bike or walk to work, even though it doesn't have areas with high concentrations of crashes. The "Real Little Italy" Belmont area, which has welcomed large numbers of immigrants for over a century and continues to do so, may be another area that the Department Of Transportation and NYPD should improve safety in.

VI. Future Work

Our analytic could be improved with more accurate numbers provided by definitively mapping crashes and complaints into PUMAs. The method we used cross walking zip codes was a good-enough solution, but it assumes that crashes and complaints are evenly distributed across a zip code. The possibility exists that a large majority of crashes in a zip code are concentrated at one intersection or on one street and this could have an impact on results.

Street infrastructure and traffic enforcement are highly dynamic in New York City, particularly under the current administration's Vision Zero initiative. Our data covers a large time range which includes data from the Bloomberg era. It would be good to see if crashes in Midtown, for example, dropped significantly after the Broadway pedestrian plazas were installed. If we only looked at more recent data, we may have reached a conclusion similar to the one found by the Department Of Transportation in their report.

VII. Conclusion

The City's open data system provides heat maps generated using similar data to the data we used in our analytic. The similarity of the results suggest our work is as error-free as possible given the limitations of geographic keys noted above. The Department Of Transportation released a report this summer that attempts to do the same thing we did, identify neighborhoods that are underserved by street infrastructure. They used different methodology and data and we believe our analytic provides an important alternate perspective.

The need for crosstown protected bike lanes in Manhattan has been noted by a number of advocates and researchers. [7] [8] Our analytic suggests a similar need. Despite the significant improvements in infrastructure in Midtown over the past 10 years, the commuters in this area are still threatened daily by reckless drivers and inadequate street design. There are certainly neighborhoods with worse streets. East New York and Maspeth have virtually no bike infrastructure whatsoever, as does nearly all of Staten Island, as we saw in our complaint data. The case exists for improving infrastructure in these areas to encourage more walking and cycling, but as it stands, these areas don't have high numbers of people walking or biking to work, and people are not being victimized by reckless driving at high rates. We believe that improving infrastructure in the Bronx will have a stronger impact than in the wealthier outer boroughs. This may be due to the fact that Brooklyn and Queens have much better transit options and Staten Island is more like a car-centric, low-density suburb than a dense urban borough. Researcher S. Neagle spoke to traffic enforcement agents in Midtown, and they reported that they were ordered to ignore certain traffic violations at roll call. This area clearly needs greater improvements. This has been further highlighted by high-profile traffic deaths occurring in the area this summer.

Of the neighborhoods identified for targeted improvement by the NYCDOE, our analytic suggests that Borough Park may be the neediest, as it has the highest number of low-income non-motorized in the city. This area has been the subject of a high-profile political debate around speed cameras lately. The speed cameras on Ocean Parkway have been very effective in reducing deaths and injuries in the neighborhood. We believe they should stay.

REFERENCES

1. B. McKenzie. Modes Less Traveled—Bicycling and Walking to Work in the United States: 2008-2012. *American Community Survey Reports*, ACS-26, U.S. Census Bureau, Washington, DC, 2014.
2. D. Lord and F. Mannering. The statistical analysis of crash frequency data: A review and assessment of methodological alternatives. *Transportation Research*, A-44, pp. 291-305, February 2010.
3. M. Andersen. The First National Survey Of People 'Interested But Concerned' About Biking. *Streetsblog*, 2015.
4. A. Jain, G. Ahuja, Anuranjana, D. Mehrota. Data Mining Approach to Analyse the Road Accidents in India. 2016 5th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), Sep. 7-9, 2016, AIIT, Amity University Uttar Pradesh, Noida, India
5. Safer Cycling: Bicycle Ridership and Safety In New York City. New York City Department Of Transportation, New York, NY 2017.
6. E. Whitford. Is NYC Doing Enough To Protect Its Booming Cyclist Population? *Gothamist*, 2017.
7. E. Whitford. Cyclists Demand Protected Crosstown Bike Lanes In Wake Of Fatal Citi Bike Crash. *Gothamist*, 2017
8. D. Meyer. TransAlt Calls for Protected Crosstown Bike Routes After Bus Driver Kills Dan Hanegby. *Streetsblog,* 2017.