**PUI2015 Extra Credit Project**

**Vision Zero Crash Data Analysis**

**<Alexey Kalinin, GitHub – ak6129, NYU ID - ak6129>**

**Abstract**: The project was focused on exploring fatalities occurred in New York City (NYC) among 3 major groups involved in traffic accidents between 2009 and 2016: pedestrians, bicyclists, and motor vehicle occupants(MVO). The project’s results show that overall trend in fatalities is declining, while trend analysis for each group shows that fatalities among bicyclists is increasing. Also, further analysis revealed that the highest number of fatalities occurred: for pedestrians in lower east side downtown Manhattan, zip code 10002; for bicyclists in east Harlem uptown Manhattan zip code 10029; for MVO in East Flatbush, Brooklyn, zip code 11203.

**Introduction:** In New York City (NYC), nearly 4,000 residents are injured and more than 250 residents are killed due to traffic collisions each year1.  According to NYC government, collisions injure or kill some City residents every two hours2.  In response to the high injury rate, the City has taken a number of initiatives to mitigate the volume of injuries, including, expanding enforcement against speeding and yielding to pedestrians, harsher penalties for dangerous drivers, and new street designs that improve safety. The questions for this project are to find out tendency for fatalities due to road accidents in NYC for period from 2009 to 2016, localize some high-risk locations for 5 boroughs for 3 major groups – pedestrians, bikers, and motor vehicle occupants, and explore what are the changes for these locations. To answer the stated questions, I’m going to analyze a data provided by Department of Transportation (DOT) of NYC.

**Data:** The main resource of the data for this project was DOT specifically Vision Zero Data Feeds3. The project will include two sets of data: fatalities distributed by year and fatalities distributed by month. Both data sets consist of total number of fatalities for period from 2009 to 2016, distribution of fatalities among 3 major groups, year or month for registered fatalities, and location within 5 boroughs in NYC. To prepare the data for analysis I formatted and merged originally separated columns - ‘YR’ for year, and ‘MN’ for month – to yyyy/mm/dd format and added a new column “Date”. Then I used groupby function to group and sum up data in every column and new data frame:

| Date | Fatalities | PedFatalit | BikeFatali | MVOFatalit |
| --- | --- | --- | --- | --- |
| 2009-01-01 | 733 | 649 | 0 | 84 |
| 2009-02-01 | 830 | 646 | 0 | 184 |
| 2009-03-01 | 253 | 226 | 6 | 21 |
| 2009-04-01 | 679 | 263 | 40 | 376 |

The same principle I used to format geospatial data set. Also for datasets for heatmaps I converted ‘0’ to NaN and drop NaN’s with function dropna. It was done to make sure that ‘0’ are not counted as points.

For example, dataset of Pedestrians fatalities prepared for heatmap:

PedFatalit Date geometry

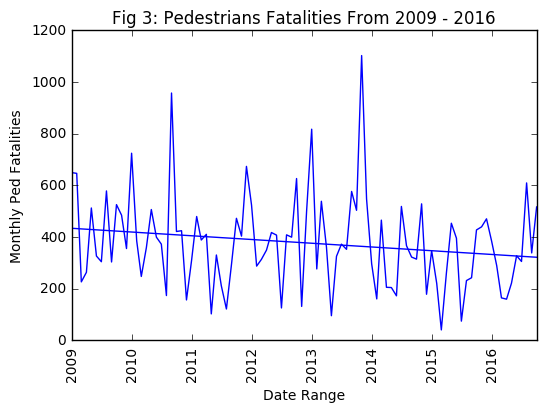
1.0 2009-10-01 POINT (-73.89052628583379 40.81151194094214)

1.0 2009-10-01 POINT (-73.9002758732755 40.86378763472604)

1.0 2009-10-01 POINT (-73.96759867796442 40.58037846976183)

1.0 2009-10-01 POINT (-73.86272633841989 40.74986793616723)

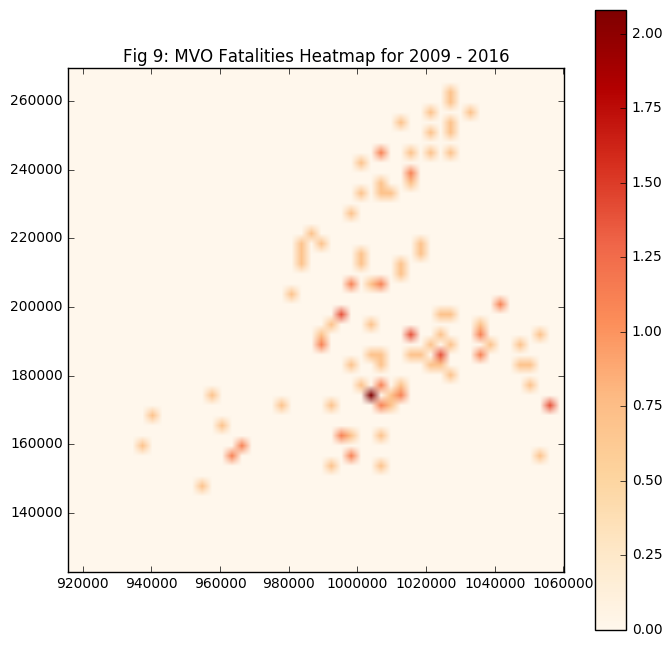
1.0 2009-10-01 POINT (-73.98134068728926 40.74707866223923)

**Methodology:** To show a trend of fatalities for every major group I created a statistical model using ordinary least square. This model clearly visualized data within a date range (2009-2016) and showed a trend.

This model can be used in similar contexts:

x = [i for i in range(0,len(fatalities\_data\_monthly\_v2))]

model = sm.OLS(fatalities\_data\_monthly\_v2['Fatalities'], sm.add\_constant(x)).fit()

For geospatial analysis, I used a heatmap plots from geopandas point geodataframes. Defined the heatmap function which takes a GeoDataFrame with point geometries and shows a matplotlib plot of heatmap density. This is done using numpy's 2D histogram binning with smoothing from scipy.7

**Conclusions:** The project’s results show that overall trend in fatalities is declining, while trend analysis for each group shows that fatalities among bicyclists is increasing. The fact of increasing fatalities among bicyclists could be resulted from significant increasing in number of bicycles from Citi Bike project that was launched in 2013. Citi Bike is the nation's largest bike share program, with 10,000 bikes and 600 stations across Manhattan, Brooklyn, Queens and Jersey City.4 Also, further analysis revealed that the highest number of fatalities occurred: for pedestrians in lower east side downtown Manhattan, zip code 10002; for bicyclists in east Harlem uptown Manhattan zip code 10029; for MVO in East Flatbush, Brooklyn. The main factor that could lead to high fatality rate among pedestrians in zip code 10002 is an extremely large population density: 92,573 people per sq mi.5 The people living in ZIP code 10002 are primarily Asians. For bicyclist fatalities, the factor of extremely large population density is also significant, as well as close localization to the Central Park. Population density for zip code 10029 is 92,116 people per sq mi.5 MVO high risk fatalities spot is in zip 11203, Brooklyn. As per NYPD Vehicle Collision Data6 the highest rate of vehicle collisions happened in Brooklyn, the most populated borough of New York. Distinctive features of zip 11203 are: annual average household income is less than 25,000 USD, 40% of population have no earnings5. Though it may not directly relate to the high MVO fatalities numbers in this location. While it is hard to find out what exact changes were done for these specific high risk zones, Vision Zero is making major strides in safer street design. DOT is adding safety engineering improvements at 50 intersections and corridors, creating 25 new arterial slow zones, and implementing eight new neighborhood slow zones, among many other improvements.1

**Future work:** Further steps could include analysis on streets intersections scale and people age. To do this analysis focus should be zoomed to streets level and intersections where the most fatalities occurred. To make analysis that will reveal age of people involved in road accidents with fatalities dataset should have personal data of pedestrians, bicyclists, or mvo.

**Bibliography**

1. Vision Zero Program web site <http://www1.nyc.gov/site/visionzero/index.page>
2. Vision Zero Investment article <https://www.transalt.org/getinvolved/vision-zero-investment>
3. Vision Zero Data Feeds <http://www.nyc.gov/html/dot/html/about/vz_datafeeds.shtml#crash>
4. Citi Bike program official web site <https://www.citibikenyc.com/>
5. Statistical data by the U.S. zip code <http://www.unitedstateszipcodes.org/>
6. NYC Open Data, NYPD motor vehicle collision data <https://data.cityofnewyork.us/Public-Safety/NYPD-Motor-Vehicle-Collisions/h9gi-nx95>
7. Geopandas heatmap instruction <http://nbviewer.jupyter.org/gist/perrygeo/c426355e40037c452434>