Vehicle Collisions in NYC: A spatial analysis of injuries and risk aversion

Aaron D'Souza, Priyanshi Singh, Nurvirta Monarizqa, Kaylyn Levine, Vishwajeet Shelar, and Patrick Mwangi

Urban Spatial Analytics

December 15th, 2016

Introduction

New York City, like other major urban areas, has a high density of both vehicle and human traffic. Vision Zero, an action plan created by Mayor Bill de Blasio, is a major city project to reduce traffic-related injuries and deaths. The objective of this project, which is to identify areas of New York City that are at high risk for vehicle collisions using a model based on a combination of weighted factors, will provide insights to which areas of the city Vision Zero efforts should be focused on. In this spatial analysis we defined the time period in which collisions are happening as between January 1st, 2015 to January 1st, 2016. Similarly, only vehicle collisions in which personal injuries occurred were considered for this analysis. The vehicle collision data, as well as various internal and external related factors, were analyzed using LISA (local indicators of spatial autocorrelation). The goal is to use our findings to identify at-risk areas of the city based on census tracts and present solutions to the Department of Transportation and the Vision Zero team to increase pedestrian and traffic safety.

The City of New York, under Mayor Bill de Blasio, created the Vision Zero Action Plan to mitigate traffic collision injuries and deaths through increased enforcement, street design changes, public outreach, and legislation changes (Vision Zero Action Plan). Spatial analysis of vehicle collisions in the city can provide the Vision Zero team with high risk areas, as well as hot and cold spot patterns.

Methodology

The vehicle collision data was obtained from NYPD open data and cleaned in python for this spatial analysis. The population data used in this spatial analysis was retrieved from the U.S. Census Bureau and delineated by census tracts. This dataset was combined with MTA Ridership information to incorporate commuters, tourists, and other population influxes into the city. The U.S. Census Bureau shapefile contains population characteristics from the 2010 Census as well as geographic information for mapping (GEOID). The data was clipped to the boroughs of New York City in ArcGIS.

The street assessment rating file was used to understand the impact of street quality on the collisions caused in New York City. This shapefile was retrieved from the NYC Department of Transportation. After reading the metadata file and the table, 3 significant contribution caused by poor street quality were identified viz. Pavement Defective, Lane Marking Improper/Inadequate and Obstruction/Debris. The data was then aggregated to the census blocks using a spatial join for the analysis.

To assess the safety in terms of number of Collisions per census tract, we took the Safety Score (point data) from MIT media lab which has given safety scores for the segment of streets. This score is perceived safety and based on crowdsourcing data. In order to visualise safety per census tract we spatially joined the safety score with census Data. Again, To see the number of Collisions per census

Tract, we spatially joined the two datasets and visualised the data. While Visualising we saw that some census tracts which were given higher safety score had higher number of collisions, which convey that the Census Tract is not safe. So, we concluded that it the number of collisions were not correlated with the safety score data.

To take account of all the external factors, we used Spatial Regression in python with geopandas and pysal package to weigh them as fairly as possible; coefficients from the regression will be the weight for each factor. The dependent variable was people density (number of people exposure divided by area of census tract), while we used following factors as independent variable: average number of driving lane, population density (under 18), average streetscore, and average safety score. We got 0.3034 for the Moran's I score, meaning there are spatial dependencies. Upon regression, we normalized the factors and we got coefficients as follows: average driving lane (0.1406), people density (0.043), under 18 density (0.048), average street score (0.0323), average safety score (0.0321) (Figure 1).

Also from regression, we got predicted value for each observation and we used them as risk value for each census tract. From that, we calculated spatial lag for each census tract and divided them into 10 deciles (External Quantile Map). From here we can start seeing which places are more dangerous than the others. To make it more clear, we calculate the local indicators of spatial correlation (LISA) from all the lag scores with queen as the weights matrix, then divides the data as cold, hot and neutral spots with parameter 1 times spatial lag for cold spots and 3 times spatial lag for hot spots.

We did the same thing for the top 6 contributing factors (driver inattention/distraction, failure to yield right-of-way, fatigue/drowsy, other vehicular, traffic control disregard, and physical disability). In the data exploration, we found the most of injuries are caused by driver inattention/distraction (Figure 2). However, from the spatial regression, we found that traffic control disregarded has the highest coefficient (1.44), so it affects the most on the number of injuries than driver inattention (0.953) (Figure 3).

Analysis/Results

Upon initial exploration of the vehicle collision dataset, the most prevalent contributing internal factor for injury collision was driver inattention. However, our spatial analysis of the inside factors resulted in traffic control disregard being the factor that impacted collision risk the most. Traffic control disregard, which can be defined as ignoring signals and signs or failing to stop, is therefore a significant risk factor for vehicle collision injuries in New York City. Upon further analysis of traffic control disregard, there were more in in Brooklyn and Bronx than Manhattan, which is dissimilar from the analysis including all collisions (with and without injuries).

After analysis of the external factors (population, number of traffic lanes, street score, and safety score), it was determined that the number of traffic lanes impacted the injury collision risk most prominently. Each external factor analyzed independently resulted in uncorrelated areas with high risk for injury.

Conclusions

This spatial analysis of vehicle collisions revealed that injuries are distributed unequally throughout the city, and highlighted which internal and external contributing factors impact risk the most. From our analysis we have found that for internal factors, traffic disregard affects the risk of injuries the most. Whereas for external factors, the number of lanes on the street affects the risk of injury the most. Considering the two highest contributing factors, traffic disregard and number of lanes per street, we came up with up with the following recommendations:

- Higher enforcement by the NYPD on jaywalking would act as a deterrent for pedestrians to cross streets without paying attention to signs
- The addition of road treatments such as barriers or pedestrian refuges on high lane-count streets would reduce pedestrian injuries as pedestrians would have a place to stop in the event that they encounter oncoming traffic
- The removal of parking adjacent to corners at intersections would increase visibility for both pedestrians and drivers, giving them enough time to avoid collisions
- Addition of pedestrian detection measures such as video detection at pedestrian crossings that could help extend crossing time for late crossing pedestrians
- Since traffic disregard was a high contributor towards collision counts, higher fines for drivers who disregard signage or other measures in place would reduce the number of collisions
- Further reductions in speed limits. With increased visibility and better pedestrian protections, a lower speed limit would give drivers more time to react to situations, thus avoiding collisions.

Works Cited

Branch, Geographic Products. "TIGER/Line® with Data." *US Census Bureau Geography*. N.p., n.d. Web. 5 Dec. 2016. <https://www.census.gov/geo/maps-data/data/tiger-data.html>.

Campbell, Jon. "The NYPD Enforces Jaywalking When It Damn Well Feels Like It." *Village Voice*. N.p., 28 July 2016. Web. 15 Dec. 2016. http://www.villagevoice.com/news/the-nypd-enforces-jaywalking-when-it-damn-well-feels-like-it-7650140>.

Curtin-Monash Accident Research Center. "An investigation of pedestrian crashes at traffic intersections in the Perth Central Business District". December 2013

Lab, MIT Media. *Street Safety Score Data*. MIT, n.d. Web. 5 Dec. 2016. <http://streetscore.media.mit.edu/data.html>.

MTA New York. "Annual Subway Ridership." Web. 5 Dec 2016. <http://web.mta.info/nyct/facts/ridership/ridership_sub_annual.htm>

"NYPD Motor Vehicle Collisions | NYC OpenData." *NYPD Motor Vehicle Collisions | NYC OpenData*. NYC Open Data, n.d. Web. 30 Nov. 2016. https://data.cityofnewyork.us/Public-Safety/NYPD-Motor-Vehicle-Collisions/h9gi-nx95>.

Street Score Data. NYC DOT - Data Feeds, n.d. Web. 5 Dec. 2016. <http://www.nyc.gov/html/dot/html/about/datafeeds.shtml>.

Vision Zero Action Plan. New York City: n.p., 2012. City of New York. Web. 14 Dec. 2016. <http://www.nyc.gov/html/visionzero/pdf/nyc-vision-zero-action-plan.pdf>.

Figures

```
REGRESSION
----------
SUMMARY OF OUTPUT: MAXIMUM LIKELIHOOD SPATIAL LAG (METHOD = FULL)
-----------------------------------------------------------------
Data set            :nyct2010_16d/ct_filtered.dbf
Weights matrix      :       queen
Dependent Variable  :  inj_per_ar              Number of Observations:       2120
Mean dependent var  :      0.8308              Number of Variables   :          7
S.D. dependent var  :      0.7306              Degrees of Freedom    :       2113
Pseudo R-squared    :      0.3142
Spatial Pseudo R-squared:  0.1736
Sigma-square ML     :       0.368              Log likelihood        :  -1993.474
S.E of regression   :       0.607              Akaike info criterion :   4000.948
                                               Schwarz criterion     :   4040.562

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT       0.4309337       0.0238077      18.1005801       0.0000000
           Avg_TLane       0.1406981       0.0143758       9.7871453       0.0000000
             PEO_DEN       0.0438682       0.0136493       3.2139515       0.0013092
           UN_18_DEN       0.0481207       0.0139590       3.4473014       0.0005662
          average_st       0.0323282       0.0133329       2.4246916       0.0153214
          Qscore_avg       0.0321207       0.0142503       2.2540383       0.0241938
         W_inj_per_ar      0.0770712       0.0038098      20.2296155       0.0000000
------------------------------------------------------------------------------------
=============================== END OF REPORT ======================================
```

Figure 1. Spatial Regression of External Factors

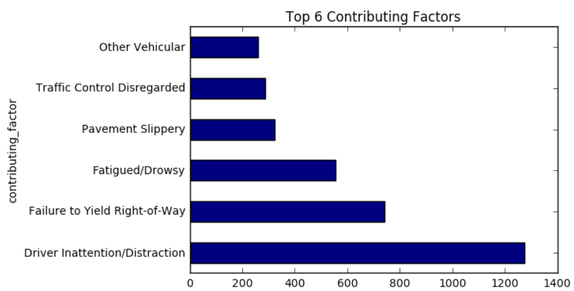


Figure 2. Top 6 Contributing Factors

```
REGRESSION
----------
SUMMARY OF OUTPUT: MAXIMUM LIKELIHOOD SPATIAL LAG (METHOD = FULL)
-----------------------------------------------------------------
Data set            :nyct2010_16d/ct_filtered.dbf
Weights matrix      :       queen
Dependent Variable  :  inj_per_ar              Number of Observations:       1859
Mean dependent var  :     21.2690              Number of Variables   :          8
S.D. dependent var  :     17.7777              Degrees of Freedom    :       1851
Pseudo R-squared    :      0.7160
Spatial Pseudo R-squared:  0.6732
Sigma-square ML     :      89.712              Log likelihood        :  -6824.784
S.E of regression   :       9.472              Akaike info criterion :  13665.568
                                               Schwarz criterion     :  13709.790

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT       2.3152913       0.4298448       5.3863423       0.0000001
Driver Inattention/Distraction       0.9537904       0.0720780      13.2327592       0.0000000
Failure to Yield Right-of-Way       0.4969428       0.0973546       5.1044608       0.0000003
     Fatigued/Drowsy       0.8627573       0.0837037      10.3072826       0.0000000
     Other Vehicular       0.3789080       0.1240211       3.0551888       0.0022492
Traffic Control Disregarded       1.4401197       0.1335451      10.7837739       0.0000000
 Physical Disability       0.6103541       0.1351213       4.5170824       0.0000063
        W_inj_per_ar       0.0366089       0.0030576      11.9731368       0.0000000
------------------------------------------------------------------------------------
=============================== END OF REPORT ======================================
```

Figure 3. Spatial Regression of Internal Factors

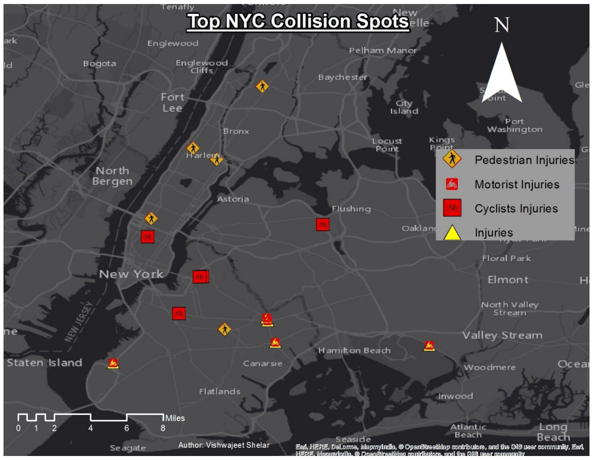


Figure 4. Top NYC Collision Spots

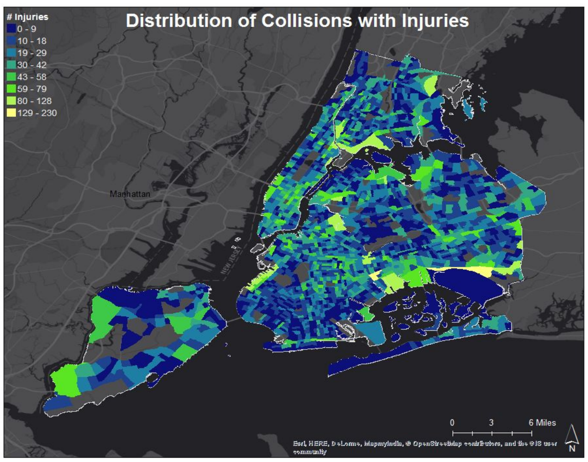


Figure 5: Distributions of Collisions that resulted in injuries

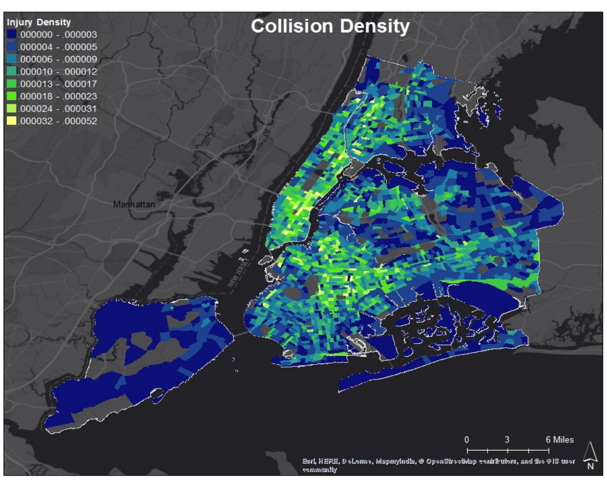


Figure 6: Collision density

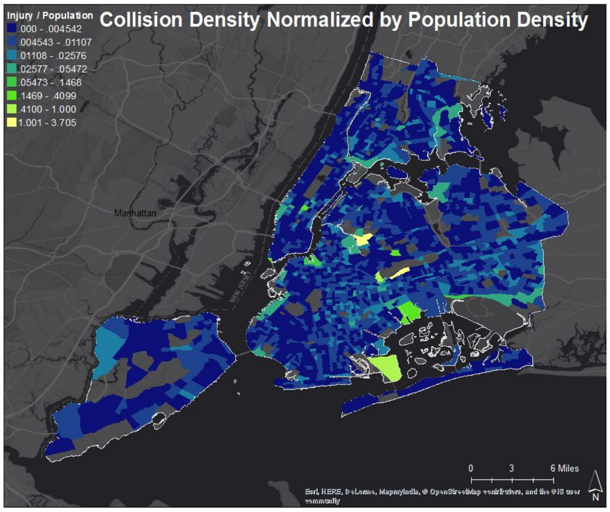


Figure 7: Collision Density Normalized by Population Density

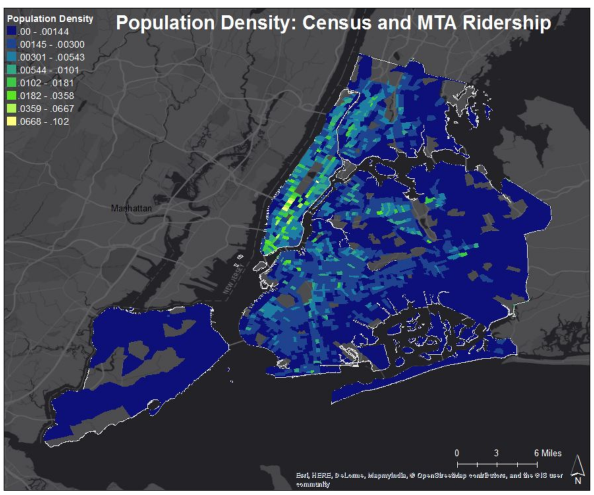


Figure 8: Population Density -: Census and MTA Ridership

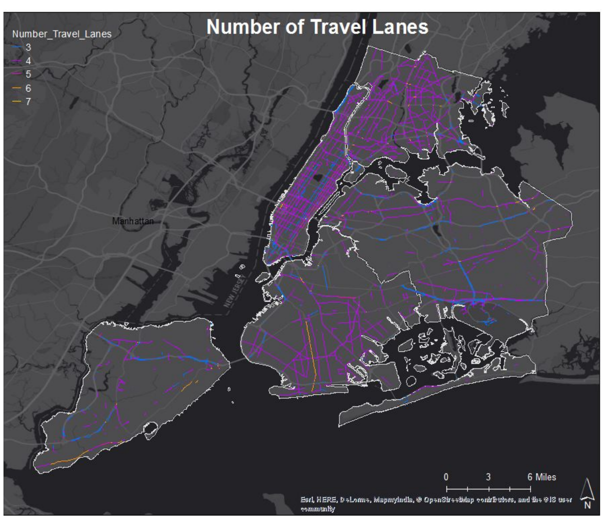


Figure 9: Number of Travel Lanes

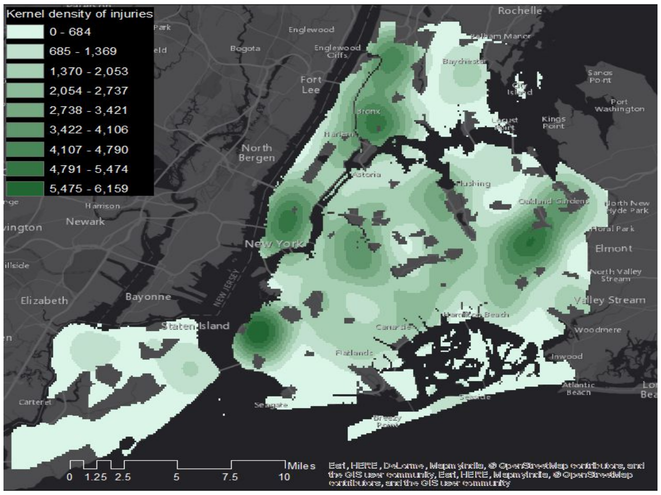
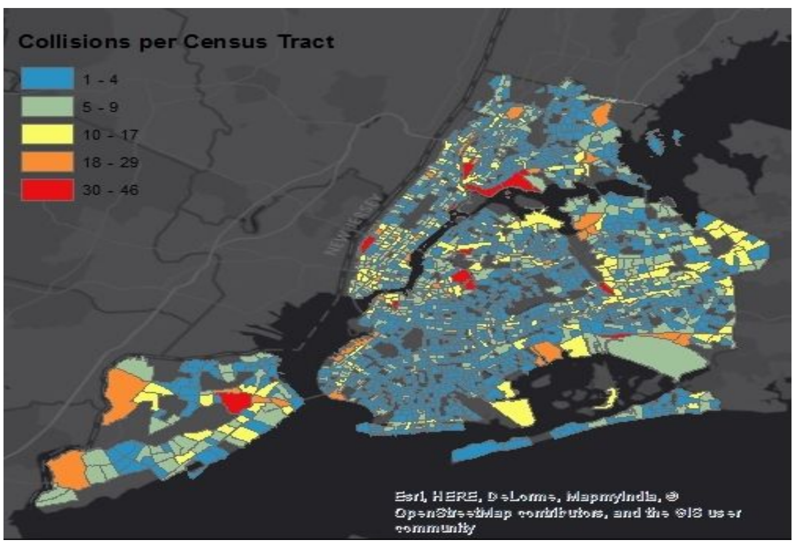


Figure 10: Kernel Density of Injuries



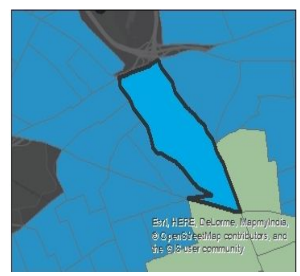
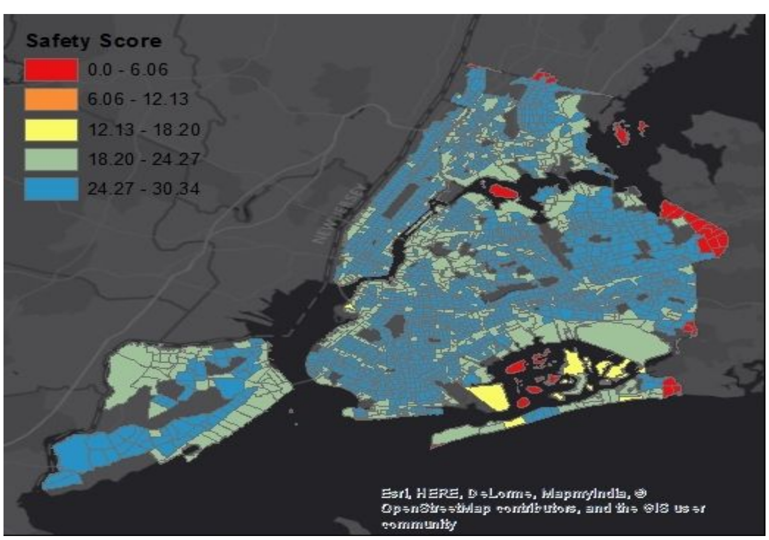


Figure 11: Safety Score(Perceived) by MIT Media Lab



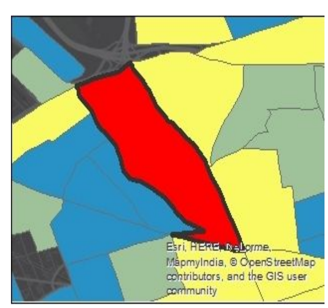


Figure 12: Number of Collisions per Census Tracts **(Actual Safety is not same as perceived safety)**

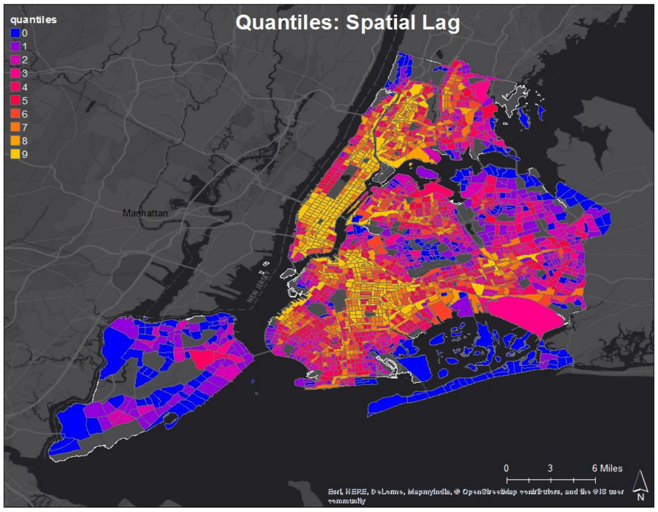


Figure 13: Spatial Lag

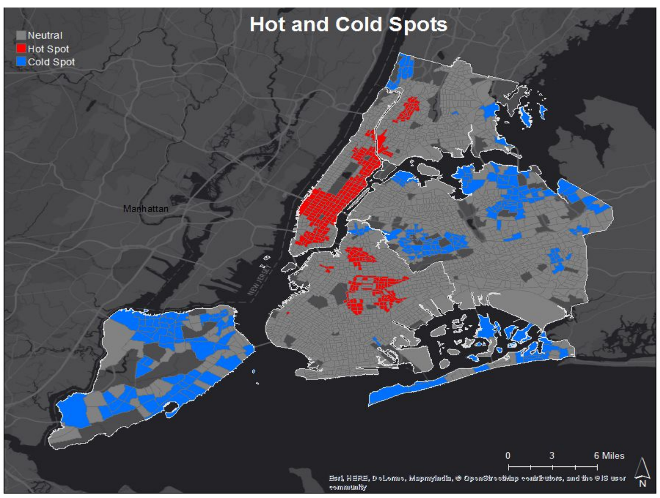


Figure 14: Hot and Cold Spots

Figure 15: Hot Spot in Bay Ridge

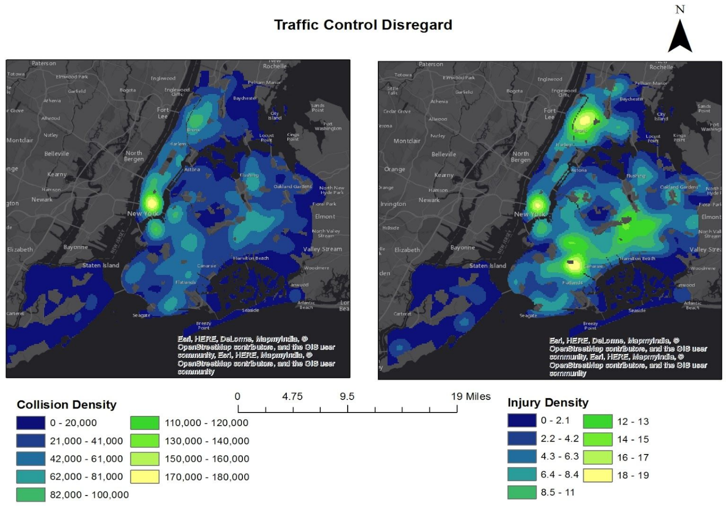


Figure 16: Comparison of Collision density and Injury Density due to Traffic Control Disregard

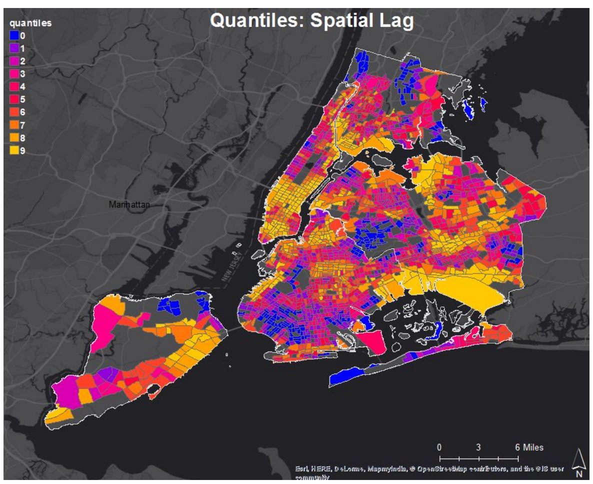


Figure 17: Internal Factor LISA Quantiles

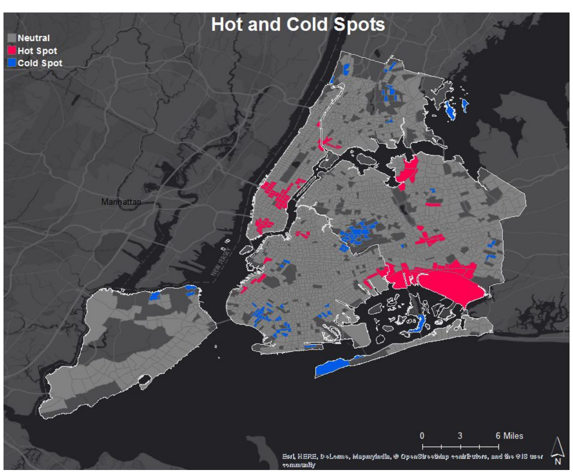


Figure 18: Hot and Cold Spots for Internal Factors