Third Party Signs and Traffic Accidents in Toronto – Final Report

Shawn Mills – 304420
York University – CSDA1050
August 23, 2019

Introduction and Literature Review/Background

The concept of distracted driving is certainly far from a new phenomenon. For as long as vehicles have been on roads, there have been distractions – audible and visual - for those behind the wheel. These distractions can come in many forms, including other vehicles, pedestrians, animals, advertisements, music, and most recently cell phones. While there have been an abundance of recent news stories and awareness campaigns towards the dangers of cell phones and texting, discussions surrounding the effects of excess signage are currently minimal by comparison.

The city of Toronto enacted its Sign By-law in 2010. Under these by-laws, signage is divided into two categories:

- First-party signs – those signs that are used to identify the business or service at the location
- Third-party signs – those signs that are used to advertise goods or services that are not related to the businesses on premises

Permits for third-party signs take about 10 days to be issued once the application is received and must be renewed every 5 years. Failure to acquire a permit before installing these signs results in a minimum $305 fine, court appearances, and all related costs of removing the signs. Larger signs and signs with digital displays are restricted to Commercial Sign Districts, whereas other signs may be allowed in Residential or Open Space Sign Districts.

Additional details on the by-laws can be found on this site:

https://www.toronto.ca/services-payments/building-construction/sign-permits-information/general-sign-inquiries/

Some examples of articles addressing signs as a driving distraction include:

The role of roadside advertising signs in distracting drivers - https://www.researchgate.net/publication/222691471_The_role_of_roadside_advertising_signs_in_distracting_drivers

The above article discusses an experiment with twelve participants in a driving simulator who were judged on their driving behaviours when presented with a street with no advertising signs and when presented with that same street with advertising signs added. The results showed more erratic driving behaviours during the simulation with advertising signs added.

Digital signage is a distracted driving hazard - https://www.myparkingsign.com/blog/digital-signage-distracted-driving/

The above article discusses an experiment that monitored driver eye motion (glances) as the drivers passed by a digital billboard on a highway. The results of this were compared to the driver eye motion when there were no signs present along that same highway. The results concluded that an increase in the number of glances were noted as the test subjects were passing the advertising sign regardless of the direction the vehicles were driven.

Research Question

Does the presence of third-party signs affect the volume of traffic accidents in Toronto?

Data and Description

The first step in performing the analysis was to gather relevant datasets. After some review, the following datasets were selected:

Third Party Signs Inventory: https://www.toronto.ca/city-government/data-research-maps/open-data/open-data-catalogue/business/#9b6f952e-52d7-1fc4-51f6-4ad6bc913218

- This dataset provided the locations of all licensed third-party signs in Toronto.

Toronto Police Service Public Safety Data Portal:
http://data.torontopolice.on.ca/datasets/ksi/data

- This dataset provided details on the traffic accidents throughout Toronto.

SimplyAnalytics – Census Tracts and Demographic Data:

https://simplyanalytics.com/

- This dataset provided the appropriate level of geographic detail for comparing the sign and traffic accident data. As well, it provided helpful demographic data in each census tract.

Methodology

The analysis made extensive use of Geospatial analytics due to the requirements of the research question. In order to answer the question, the location of the signs relative to the location of the traffic accidents must be known. Once this was in place, a technique known as Geospatial Autocorrelation was used to uncover the geographical relationships between the signs and accidents. Through this methodology, answers to the following would be uncovered:

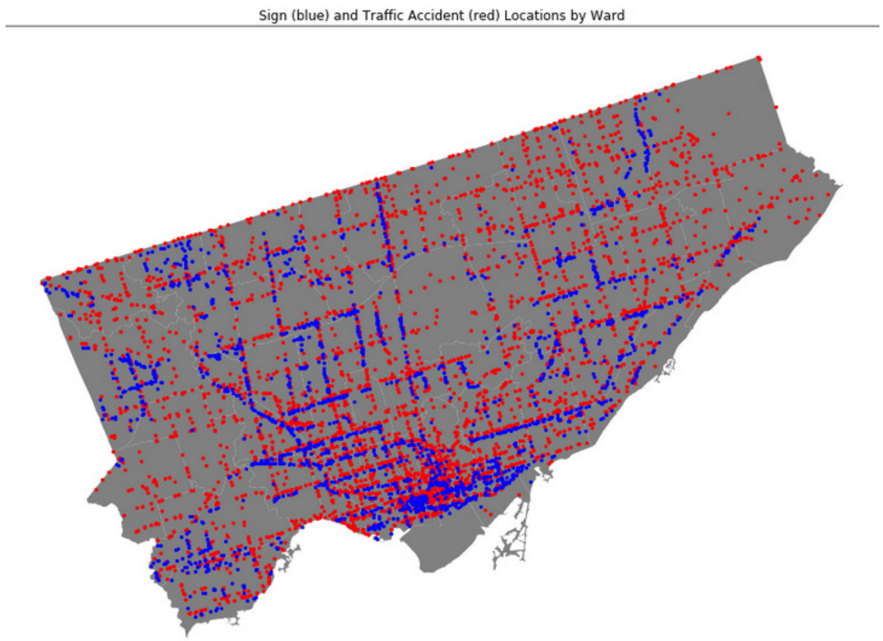
1. Is there a relationship amongst the placement of the advertising signs themselves?
2. Is there a noticeable relationship between the number of traffic accidents surrounding the locations of the advertising signs?
3. Is there a demographic influence on the sign placements and, by extension, the number of related traffic accidents?

The following python (version 3.6.8) libraries were used for this analysis:

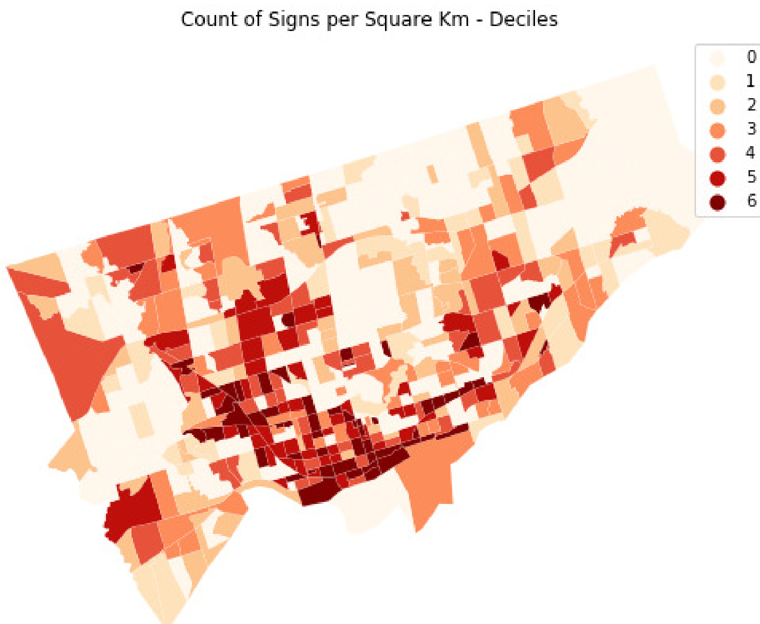
- Pandas 0.25.0
- Numpy 1.16.4
- Xlrd 1.2.0
- Matplotlib 3.1.1
- Seaborn 0.9.0
- Geopandas 0.5.1
- Shapely 1.6.4.post1
- Fiona 1.8.4
- Pycrs 1.0.1
- Pysal 2.0.0

Results

Once the data was cleansed, the geography columns in each dataset needed to be standardized to the same projection or CRS (Coordinate Reference System). MTM Zone 10 (epsg=32190), a standard used in Canadian projections, was selected for this analysis. From this point a visualization of the advertising signs and traffic accident locations could be produced as follows:

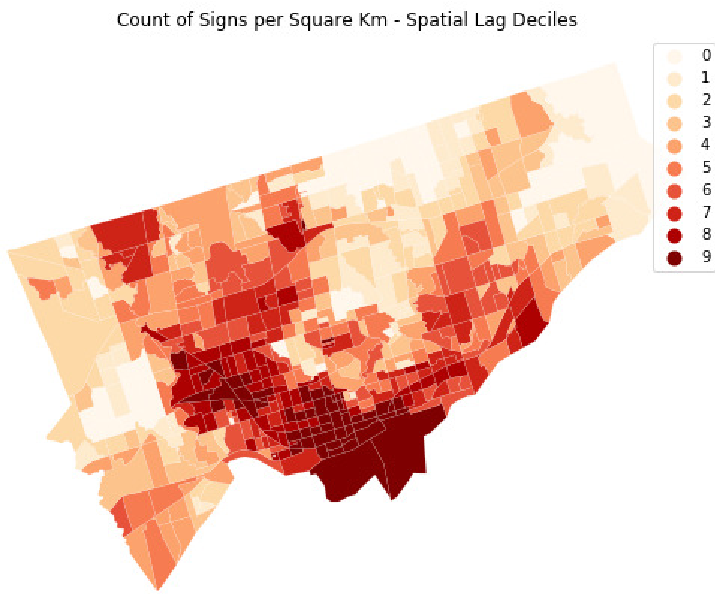


In this image it was observed that there were many areas where both traffic accidents (red) occurred in areas close to clusters of advertising signs (blue). However, given the number of areas where traffic accidents occurred without the presence of the advertising signs, we can conclude that not all traffic accidents were the result of these signs. Further investigation into the areas where both accidents and signs were present was necessary.

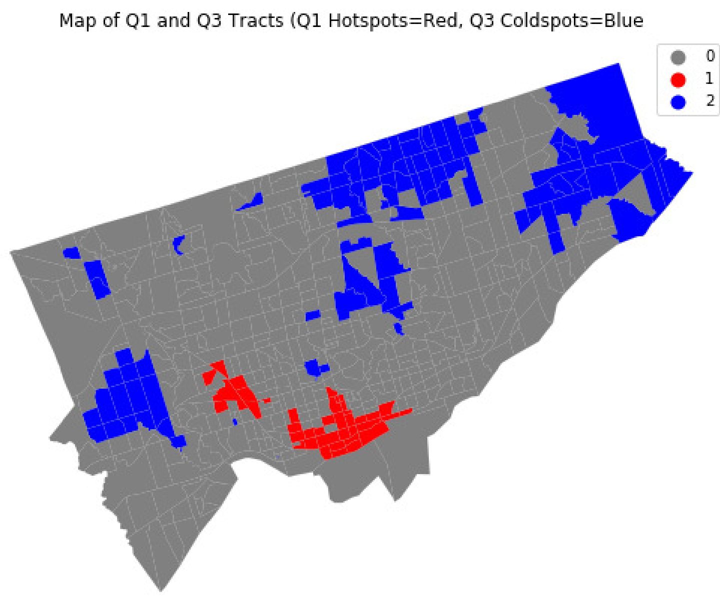


The above map illustrates which census tracts had the greatest number of advertising signs per square kilometre. The darker shaded regions indicate tracts containing greater numbers of signs.

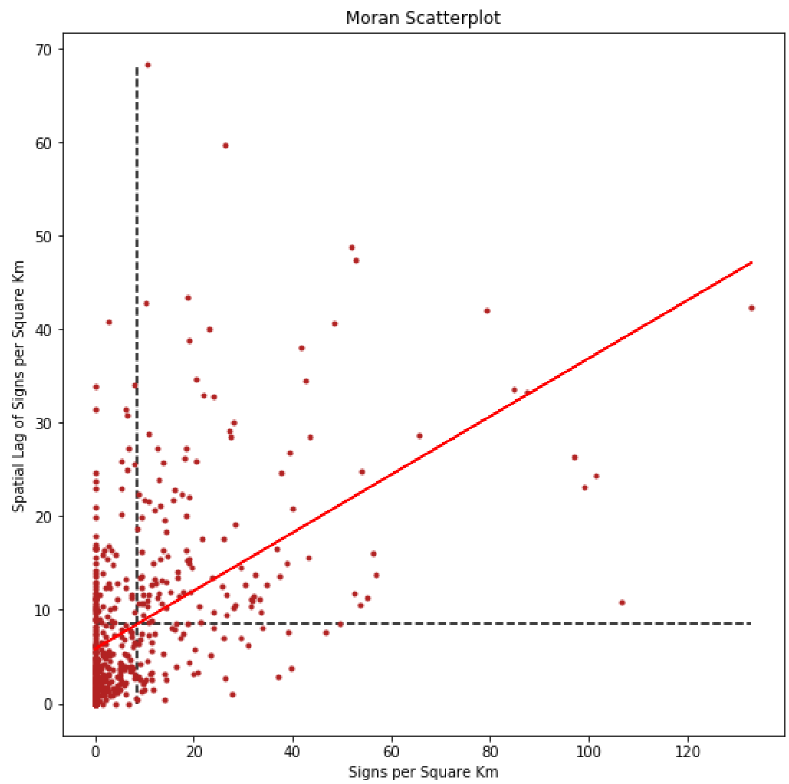
Spatial Autocorrelation is a technique that provides a means of visually clustering similar regions based on a given feature. The PySAL library provides the tools necessary for this type of analysis. Spatial similarity between regions is evaluated using spatial weights representing which regions are neighbours. Queen weights were used in this analysis so that each census tract that shared the same edge or vertex with another tract would be classed as each other's neighbour. Attribute similarity compares how similar a given feature is between neighbouring regions and is measured as the spatial lag. Using "Signs per Square Km" as the feature, the following map illustrating the spatial autocorrelation of the regions was produced:



The above map shows greater clustering of similar census tracts based on signs per square kilometer in the south end of the city and running up the central-west regions. Visually, there appears to be some spatial correlation amongst the census tracts.

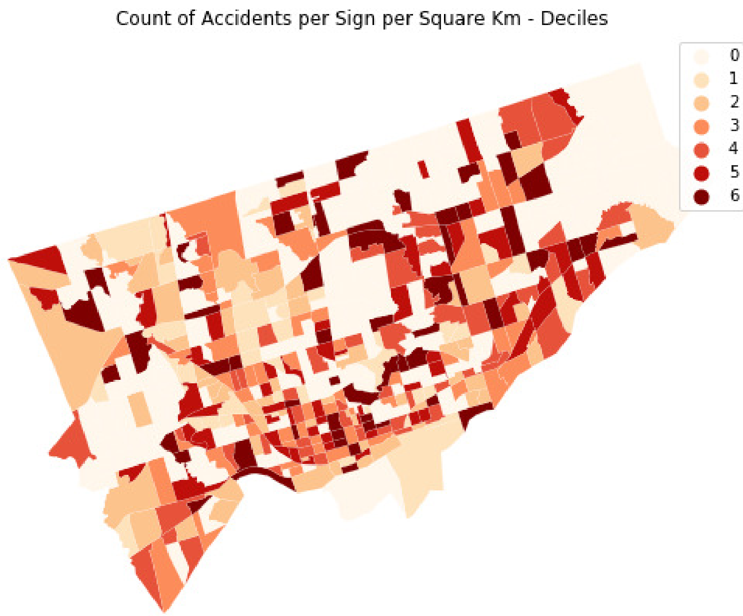


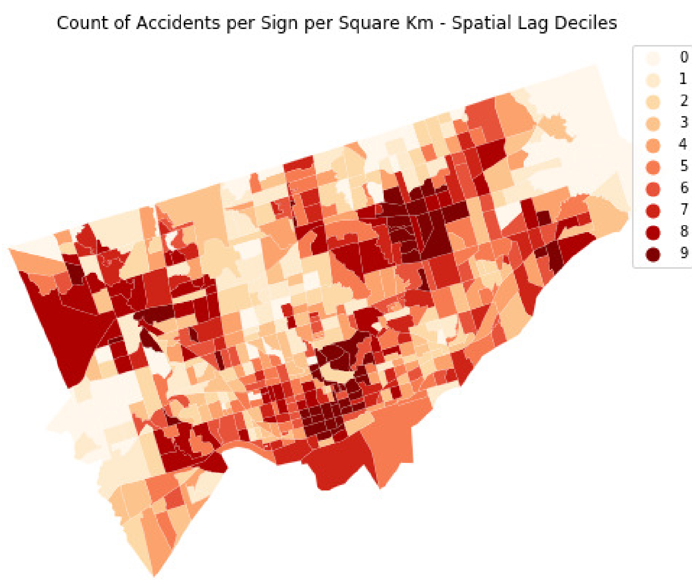
The above uses Local Indicators of Spatial Association (LISA) to highlight statistically significant clusters. The cluster of tracts with significantly higher number of signs per square kilometer appear in red, and significantly lower number of signs per square kilometer appear in blue. Higher numbers tend to be clustered in the heart of the city while lower values tend to be found towards the outer edges coinciding with the global autocorrelation results in the previous map.



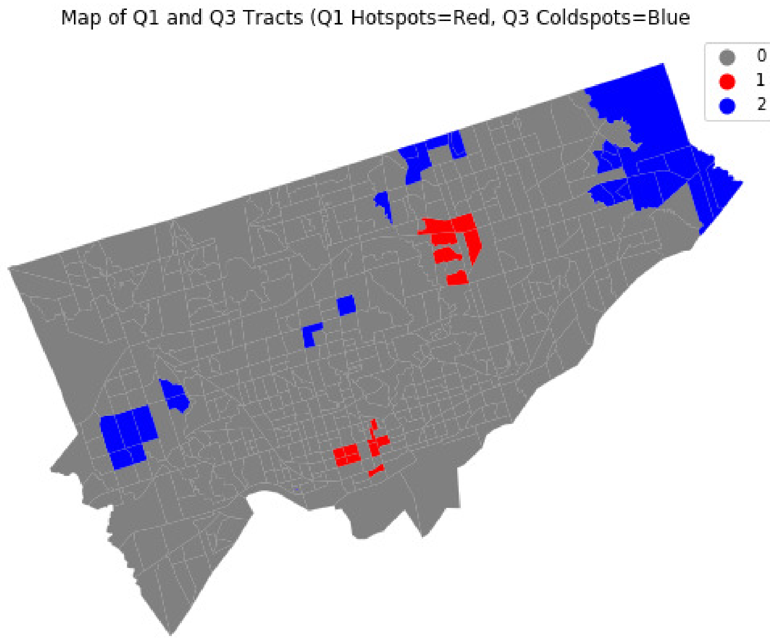
The results can also be statistically verified using Moran's I. Based on the Moran's I value of 0.31, we can conclude that there is some clustering of similar values (ie. I is a positive value), but the values are somewhat close to being random in nature (ie. I is close to 0). The p -value of 0.001 indicates that this I value was not generated by chance. The above Moran scatterplot of signs per square kilometer versus the spatial lag illustrates this relationship. Moran's I value is represented by the solid red line (slope = 0.31). The dashed lines mark the mean of the plots for both axes.

Seeing that there is a spatial relationship with the number of signs per square kilometre amongst the census tracts, does the same hold true when the traffic accident data is introduced into the analysis?

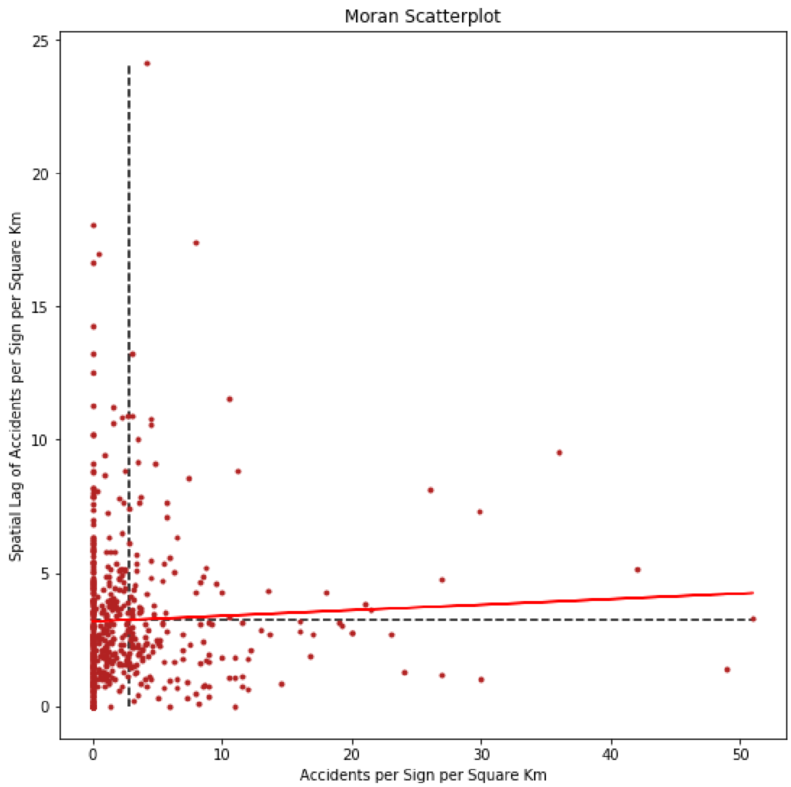




The above maps show the spatial similarities when comparing traffic accidents per sign per square kilometer. While there does appear to be some clustering of colours, the clusters appear to be more randomly placed when compared to the maps produced with only the signs per square kilometer.

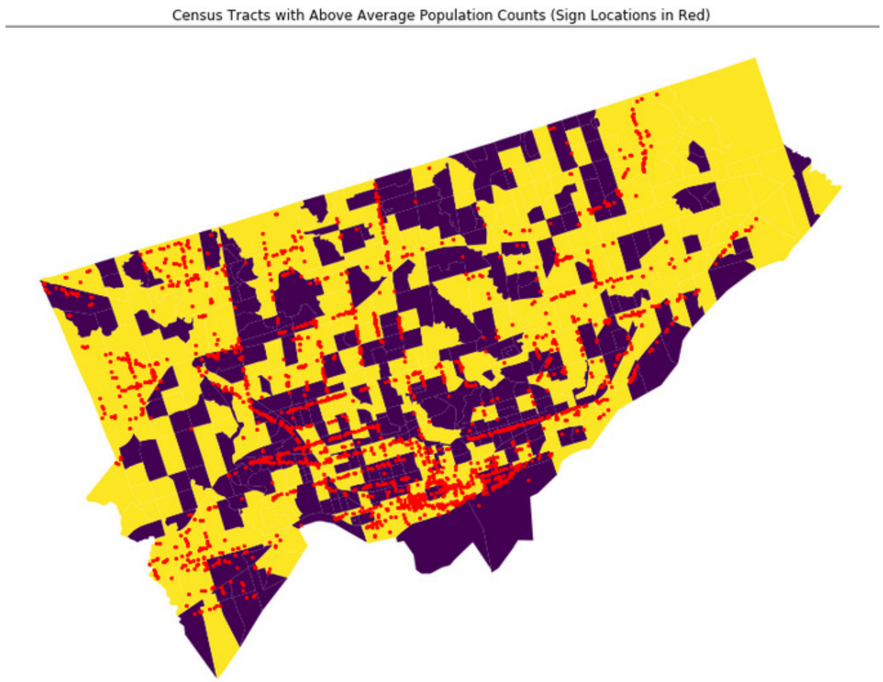


The above map of the clusters produced by the LISA statistics further illustrates the difference to the signs per square kilometer results, especially when comparing the mapped red areas. In this case, the red areas represent higher numbers of traffic accidents per sign per square kilometer.



Based on a Moran's I value of 0.02, we can conclude that there is no autocorrelation, indicating close to perfect randomness (ie. I is almost a zero value). The somewhat high p -value of 0.181 indicates that this I value could have been generated by chance. We cannot reject the null hypothesis here.

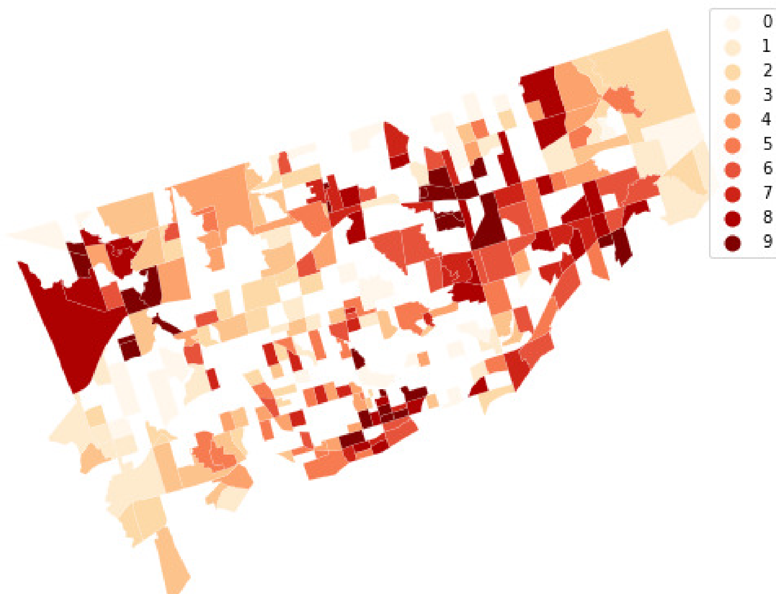
Is there a demographic reason for this? During the data discovery phase, 61.4% of signs were found in areas where the population counts in each tract were higher than the average population per tract. Population count represented one of the stronger factors in determining sign placement, so it was selected for this final analysis.



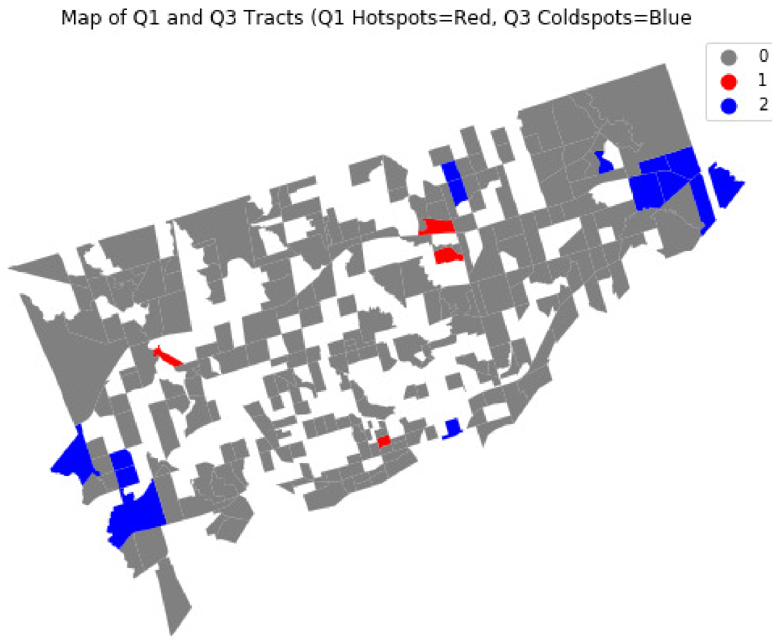
The above map shows the placement of signs in the city marked in red. The yellow shaded regions represent the tracts with above average population counts.

Count of Accidents per Sign per Square Km - Deciles
Above Average Population Count Tracts Only

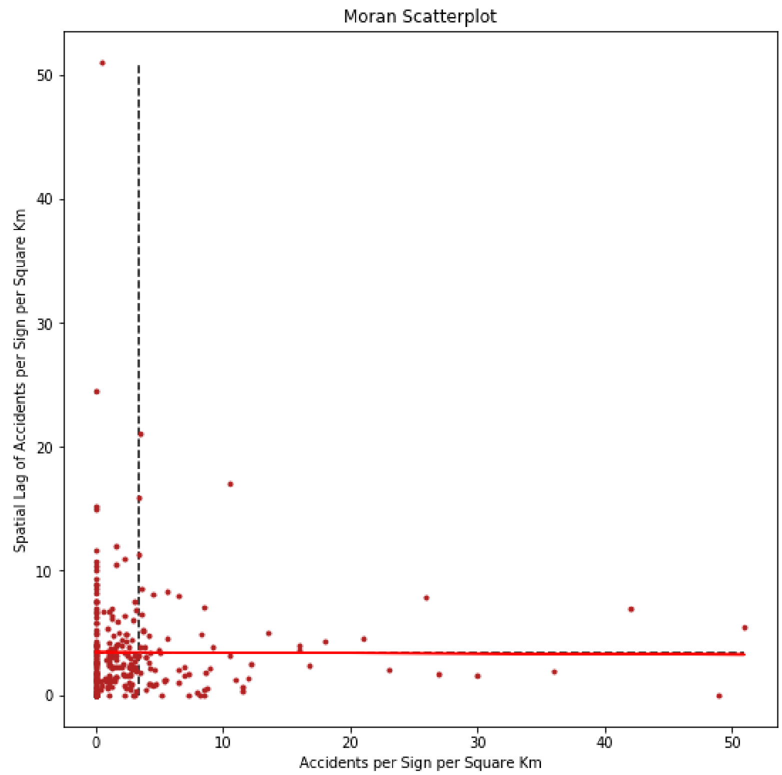
Count of Accidents per Sign per Square Km - Spatial Lag Deciles
Above Average Population Count Tracts Only



The above maps show the spatial similarities when comparing traffic accidents per sign per square kilometer in census tracts with above average population counts. As with what was shown in the accidents per sign per square kilometer maps, there appears to be some randomness to the areas with similar counts.



The hotspots highlighted in the above using the LISA statistics confirms the randomness of areas with both high (red) and low (blue) occurrences of accidents per sign per square kilometer in census tracts with above average population counts.



Based on a Moran's I value of -0.004, we can conclude that there is no autocorrelation, indicating close to perfect randomness (ie. I is almost a zero value). The p-value of 0.443 indicates that this I value very likely could have been generated by chance. We cannot reject the null hypothesis here.

One issue with removing the census tracts with low populations from this analysis was the loss of much of the spatial similarities. A tract with a high population that was surrounded by tracts of low population would no longer have adjoining neighbours resulting in a different set of spatial weights containing mostly zero values.

Conclusion

From a geospatial perspective, there appears to be no correlation between the number of traffic accidents and the presence of advertising signs. Accidents frequently occur in areas with no advertising signs, so the signage was obviously not going to be the only factor in these occurrences. The analysis also revealed small clusters of census tracts with both high and low spatial and attribute similarities. However, the volume of the accidents did not necessarily increase with the presence of more signs in all tracts and no statistically significant relationship was found.

Greater accuracy of this analysis could have been possible if the installation and removal dates of the signs were included in the data set.

Sources and Citations

Introduction to Geospatial Data in Python: https://www.datacamp.com/community/tutorials/geospatial-data-python

Introduction to Python GIS (full tutorial): https://automating-gis-processes.github.io/CSC18/lessons/L1/Intro-Python-GIS.html

GeoPython – AutoGIS – Geometric Operations: https://automating-gis-processes.github.io/2016/Lesson4-geometric-operations.html

Geo-Python: https://geo-python.github.io/2017/index.html

GeoPandas: http://geopandas.org/index.html

MTM Notes: http://leware.net/geo/mtmNotes.htm

Geographic Data Science with PySAL and the pydata stack (full tutorial): http://darribas.org/gds_scipy16/

***techniques from these pages were used extensively -- http://darribas.org/gds_scipy16/ipynb_md/03_spatial_weights.html and http://darribas.org/gds_scipy16/ipynb_md/04_esda.html

pysal Documentation Release 2.0.0: https://buildmedia.readthedocs.org/media/pdf/pysal/latest/pysal.pdf

Spatial Autocorrelation Functions: http://ljwolf.org/post/spatial_acf/

Moran's I: Definition, Examples: https://www.statisticshowto.datasciencecentral.com/morans-i/