An Analysis of Criminal Activity Distribution in the Ottawa – Nepean Area

Based on data from Statistics Canada and the Ottawa Police Department

By: Grihalakshmi Soundarapandian, 37082138
GEOB 479 L2A
March 15, 2018

Nearest Neighbour Index

User Input Z – Score for Commerical B & E's = -67.8104
User Input Z – Score for Residential B & E's = -77.4498
User Input Z – Score for Stolen Vehicles = -65.3034

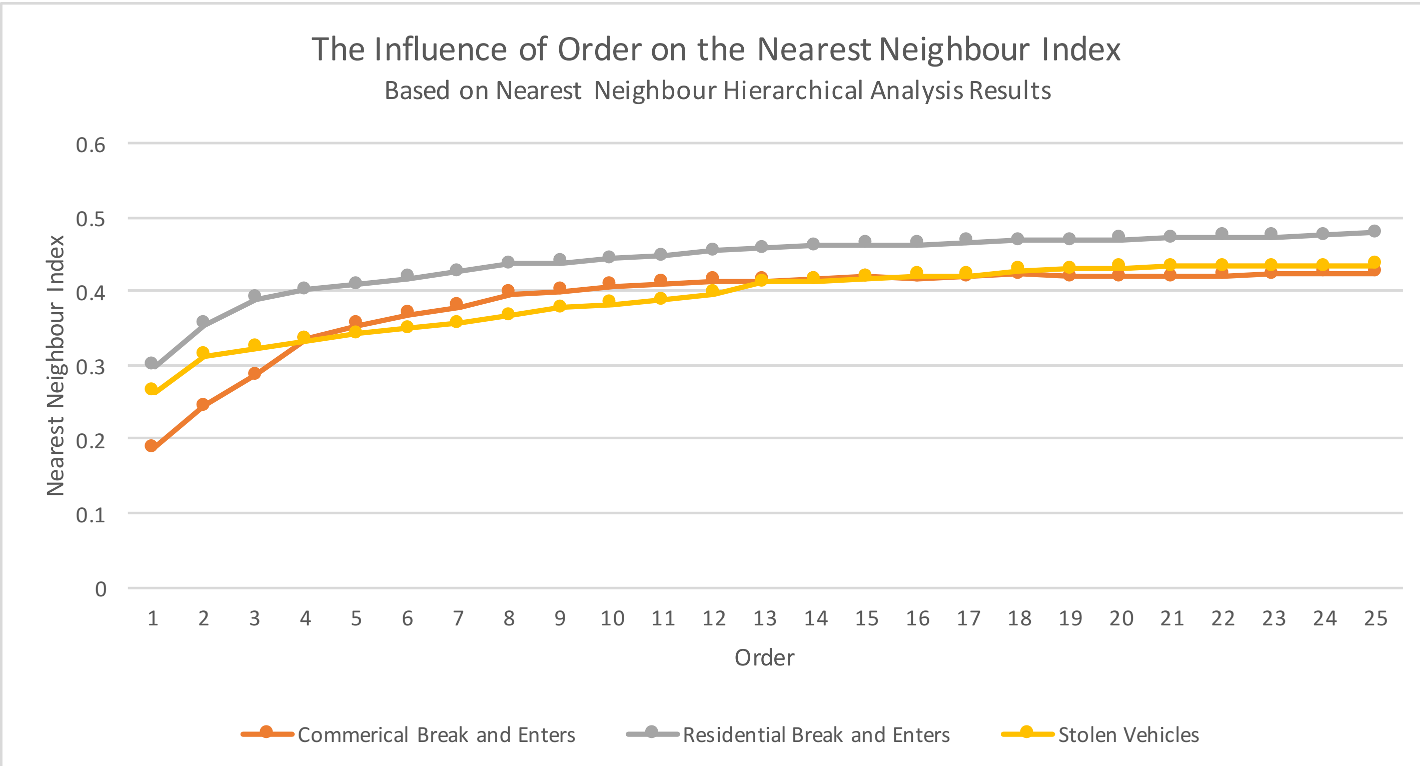

Figure 1. The Influence of Order on the Nearest Neighbour Index

Based on (Levine, Chapter 6), if the nearest neighbour index (NNI) is less than 1.0, then the data is clustered, meaning that the points are closer together than expected. Because all our first-order NNI values for all 3 variables fall under 1.0, they are more spatially aggregated, or have a stronger spatial autocorrelation, than expected. This is seen in Figure 1 above, where the first-order NNI values for Commercial B & E's, Residential B & E's, and Stolen Vehicles are approximately 0.19, 0.3, and 0.28 respectively, which all fall below 1.0.

Additionally, the index values indeed increase as a function of the type of crime and the order. This indicates that as the order of clusters increase (ie. as there are more clusters being grouped together to form more concise clusters), the spatial aggregation also increases. Hence, this shows that spatially, the crimes are distributed much like land use across the city. Land use does not change abruptly or randomly. It changes gradually, and has concentrations of certain land uses in particular areas. Likewise, the graph above indicates that the distribution of the crimes across the area is such that the crime does not happen here and there or randomly, but there are specific areas and clusters within which it happens. And the change from high crime to low crime areas is gradual. This is exactly what is seen with Maps 1, 2 and 3. We see areas with high concentration or frequency of crimes (central study area), and then the concentration decreases outward until there are only outliers left. Hence, the nearest neighbour analysis is a great way of both statistically and visually analyzing the distributions and significance of variables. Additionally, we can see that the NNI for Commercial B & E's, at a value of 0.19, is less than that of Residential B & E's and Stolen Vehicles . This indicates that the Commerical B & E's are the most spatially aggregated and autocorrelated. This is because, commercial landuse is not as wide spread and is more clustered than residential land use. Additionally, residential areas are usually not located immediately beside commercial or industrial areas . As well, the index for stolen vehicles falls between that of Commerical and Residential B&E's. This means that car theft crime is influenced by both commercial and residential land use. Hence, all of the above is evidence that the occurrence of these crimes follows that of land use distribution, and are not random, as explained in .

Moran's I

Moran's I, similar to the nearest neighbour analysis, is a tool which is used to test the spatially autocorrelation of variables. But, nearest neighbourhood analyses the patterns of the variables themselves, whereas Moran's I can be used to analyze spatial autocorrelation within DA's, and the intensity of variables . Moran's I essentially calculates a mean for a variable, then a deviation from this mean. Afterwards, it compares the value of this variable to the value of the variable at other locations or deviations (Levine, Chapter 5). The Moran's results for each variable used as listed below:

Table 1. Moran's I Results

Variable	**Moran's I Value**
Stolen Vehicles	0.026145
Pop15	-0.000028
Commerical B & E's	0.006124
Residential B & E's	0.032224

All the Moran's results for the variables are greater than zero, with the exception of the pop15 result. states that Moran's I correlogram analysis, an I value of less than zero indicates no spatial autocorrelation, and anything greater than zero indicates stronger spatial autocorrelation. Therefore, based on this criteria, is it clear that the population is not spatially autocorrelated, but the other variables are more so correlated, but this correlation may not be very strong because the Moran's I values are still close to zero. These values are depicted in figure 2 below, which is a graph showing the influence of distance on the Moran's I values. It can be seen that the shorter the distance used, the greater the Moran's I value is, meaning that the stronger the spatial autocorrelation. Of the three positive Moran's I values (indicated in Table 1 and Figure 2), Residential B & E's is the greatest, followed by Stolen Vehicles, and then by Commerical B & E's. This means that the spatial autocorrelation also decreases amongst these variables respectively. This is surprising because the nearest neighbour results showed stronger spatial autocorrelation than expected. Therefore, both analyses do not agree much.

Additionally, the Moran’s I values for the types of crimes certainly do different from the value for population. This means that the occurrence of crimes does not follow population spread or distribution, which is why it is important to investigate hotspots to see if they are a hotspot because of population or truly because of crime risk (this concept is explained later). This further raises the concept that population does not influence the distribution of these crimes as per Moran’s I results. This could be because offenders of these crimes commit these crimes regardless of high or low population areas.

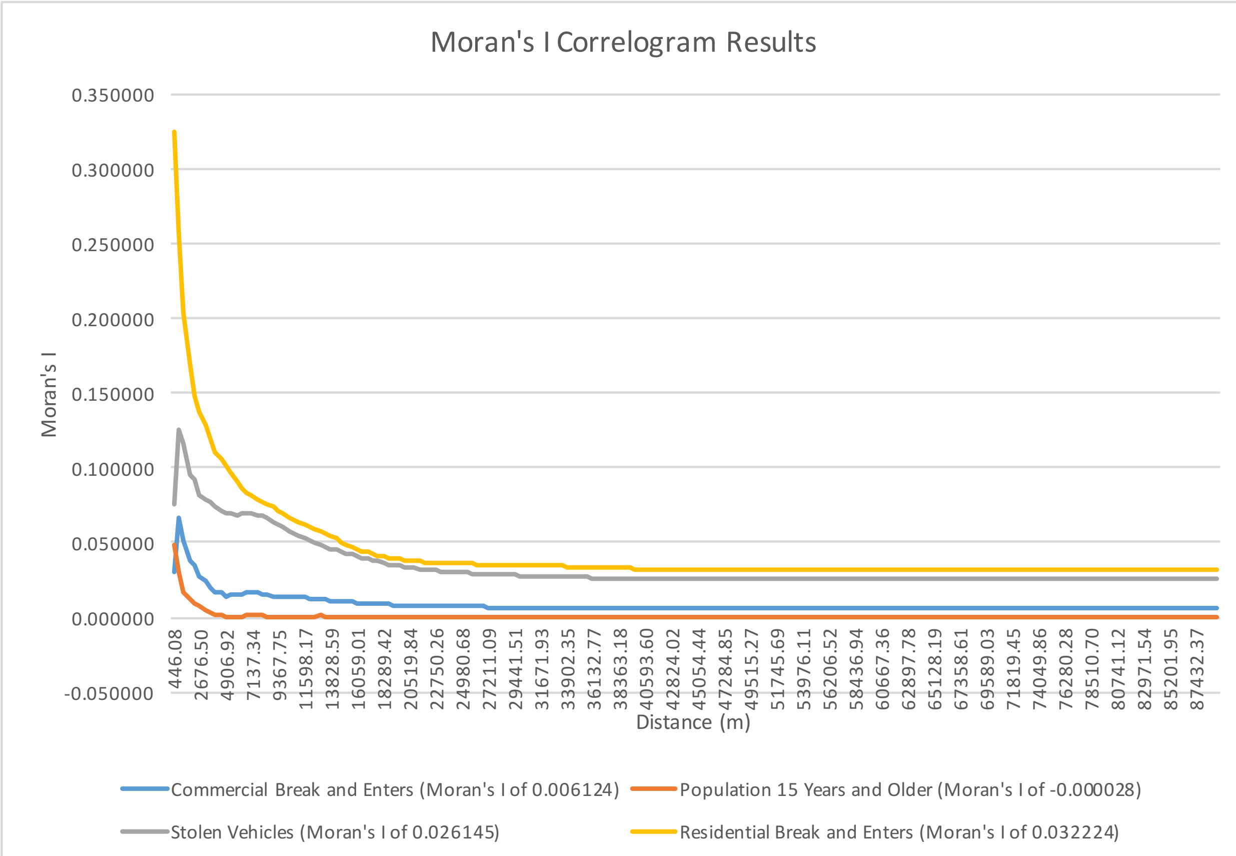

Figure 2. Moran's I Correlogram Results for all Four Intensity Variables

Fuzzy Mode vs. Nearest Neighbour Hierarchical Spatial Clustering Results

Spatial clusters, or hotspots, are areas where there is a higher occurrence of a phenomena than surrounding areas. Hotspots can vary over time and space (Levine, Chapter 7), thereby creating important visual and analytical results to better understand various phenomena. This concept has been widely used in police departments to identify neighbourhoods (or clusters) with higher crime rates, and target their work accordingly. In order to generate the hotspots in this case, two techniques were used: fuzzy more and hierarchical spatial clustering. Fuzzy mode uses location point data to indicate its hotspots and does so by looking for incidents of the phenomena around the location and within a defined search radius (in this case, the radius was 1000m) (Levine, Chapter 7).

Essentially, it allows the area around the location of a phenomena to be slightly modified, allowing for less precision but greater analysis by discovering small hotspots than exact locations (Levine, Chapter 7). Fuzzy mode produces 4 outputs: ranking of locations based on the number of incidents within each location, the frequency of incidents, and the X and Y coordinates of the location (Levine, Chapter 7). For Map 1, the frequency was used to characterize the fuzzy mode results. On the other hand, Nearest Neighbour Hierarchical techniques (Nnh) are those that first group individual incidents together based on a certain set of criteria that the individual instances must be closer than a set threshold distance and/or belong to a group meeting the minimum number of points needed (1000m and 10 points respectively in this case), in order to be first-order clustered (Levine, Chapter 7). These first-order groups are then re-grouped into bigger clusters, called second-order clusters, which then are further combined and re-grouped, forming higher order clusters, until either all the incidents belong to a single cluster, or the clustering fails (Levine, Chapter 7). Figure 1 depicts this process, which is known to replicate a hierarchical tree diagram (Levine, Chapter 7).

Looking at Maps 1, 2, and 3, it is clear that they generally agree with each other. The high frequency points of map one are in the same area (central study area) as the clusters of the Nnh and RNnh (Risk Adjusted Nnh) results. This shows that though both the fuzzy mode and Nnh/RNnh methods operate differently and require different criteria, their results can be comparable in terms of crime occurance. Both methods definitely indicate that there are crime hotspots within the study area. However, the fuzzy mode results are more suitable when looking at the relative frequency of the incidents. For example, in Map 1, it is clear that though there are many clusters or occurrences of the dark blue hotspots, they are much lower in crime frequency than the red or green hotspots. Therefore, if a police department would like to target their work to neighbourhoods that have a higher frequency of crime, the results from the fuzzy mode would be more useful than that of Nnh/RNnh. For example, the Nnh clusters in Map 2 can mislead a reader into thinking that all clusters are also equal crime frequency areas. But, Map 3, which depicts RNnh, concludes to provide a better medium between the Nnh and fuzzy mode results. Because RNnh takes the relative risk into consideration (explained more in detail below), a reader is more informed about the level of crime risk. Though they are still not given the frequency, they can see that certain areas may have a higher risk than others, as the third-order cluster of RNnh aligns with the higher frequency hotspots of Map 1.

Standard vs. Risk-Adjusted Nearest Neighbour Hierarchical Spatial Clustering

As mentioned earlier, hierarchical spatial clustering techniques can be quite useful in determining hotspots of phenomena. Two such techniques are the Standard Nearest Neighbour Hierarchical Spatial Clustering (Nnh), and the other is Risk-Adjusted Nearest Neighbour Hierarchical Spatial Clustering (RNnh), the results of which are seen in Map 2 and Map 3 respectively. The main concept that sets these Nearest Neighbour methods apart from others is that the clusters are generated based on the spatial closeness of the points (Levine, Chapter 7). Though it was mentioned earlier that hierarchical techniques produce several order clusters, Map 2 shows a difference between Nnh and RNnh results: that only a first order cluster is produced via the Nnh method here, due to the spatial distribution of the data . Therefore, Map 2 shows the results of Nnh, which are several orange clusters that give an idea of the crime hotspot distribution across the study area. Now, comparing these results to that of Map 3, obvious visual differences can be noted. Using RNnh, Map 3 contains higher order clusters, which are indicated by the yellow, orange, and blue clusters. The yellow clusters are the **first-order** clusters; hence they are the greatest in number. Then, as we move onto the second-order clusters, the clusters become less in number but greater in size. Finally, the third-order cluster is singular and greatest in size.

The main reason for these differences in results is because RNnh takes population concentration into account (Levine, Chapter 7), where as Nnh produces it's results regardless of the effects of population. Therefore, the clusters that Nnh produces could simply be because there is a higher concentration of people within that area, hence the chance of crime is greater, creating misleading results . But, RNnh creates clusters based on a secondary surface of population data (pop15 in this case), and uses a dynamic threshold distance that increases or decreases depending on the population distribution across the study area (Levine, Chapter 7). Therefore, RNnh produces results that create a relative risk surface, focusing on the risk of the phenomena occurring, and offer more details and meaning to the data, instead of using a constant threshold distance and producing results that focus on the volume or absolute number of incidents, like the Nnh does. Though both the results in Map 2 and Map 3 show approximately the same spatial variability and distribution, In Map 2, it cannot be determined which cluster is due to population concentration, and which is due to actual risk. Hence, some clusters in Map 2 may not actually be of much importance. But, Map 3 provides a clear indication of the relative risk. Consequently, the clusters formed in Map 3 have eliminated some clusters of Map 2, and provide a more refined and focused result showing clusters of genuine importance.

Knox Index

Because the incidents of crime were reported, meaning that they have both locations and time periods associated with them, the interactions between space and time can be further explored to see if there is clustering both spatially and temporally . The Knox Index allows one to do just that, as it fosters discovery surrounding location and time based relationships between incidents of a phenomenon (Levine, Chapter 12). This Index compares an individual incident to another in terms of distance and time, in which each incident categorized in “close in distance or not close in distance”, and “close in time and not close in time” (Levine, Chapter 12). Here, “close in distance” was defined as 5km and “close in time” was indicated as 6 hours. If the two incidences being compared were close in distance, time, or both, then there lies some evidence of clustering. Appendix 1 shows the generated results of this analysis. There were more incidents that feel into the category of “close in space and time” and “not close in space and time”. This further reflects what was mentioned earlier in the nearest neighbour analysis, that the occurrence of car theft is influenced by commercial and residential landuse. For example, more car theft can occur at a shopping mall during mall operation hours (ie. more evidence for the “close in space and time” incidents shown in Appendix 1). On the other hand, car thefts can occur at any given space or time in residential areas, which provides more evidence for the “not close in space in time” incidents in Appendix 1.

Also, knowing that the greater the chi-squared value, the greater the significance of the data is also important here. Based on this criteria, it can be stated that our chi-squared value of 94 is statistically significant. Within the Monte Carlo simulations, the maximum chi-squared value is 1.71090, which is greatly different from the observed value of 94. Additionally, the Monte Carlo simulations were conducted under the null hypothesis that the distribution of the crimes were random. Because the observed chi-square value is greatly different from the Monte Carlo value, the null hypothesis can be rejected, stating that the crimes are not random and that there is a clear indication of spatial aggregation and autocorrelation. This idea is further supported by the fact that 19 “simulation runs” were conducted, and all 19 times, the generated chi-squared value was vastly different from the observed chi-square value. Hence, it can be stated that the p-value of 0.00010 holds true, and that the chance of the observed chi-squared value being wrong is quite low. Hence, the Knox index also agrees with the nearest neighbour analysis, and further supports the idea that the distribution of crime within the area is not random.

Kernel Density Estimation vs. Fuzzy Mode and Nearest Neighbour Hierarchical Spatial Clustering

Previously, the creation and existence of hotspots and more precise locations of crime intensity were identified through various methods, and the results were compared. But what method is used to get a generalized idea of the spatial variability of a phenomenon across a whole study area? This is where kernel density interpolation estimation (KD) takes place. KD is a method that interpolates the point data of phenomenon incidents to create a surface which shows the spatial distribution of that phenomenon, transforming locations into larger areas (Levine, Chapter 10). Specifically, KD allows the visualization of the density of a phenomenon across the study area, using the Z or intensity value (Levine, Chapter 10). Results produced by such an analysis can be useful when, for example, a police department would like to see how crime distribution has changed over time, or where the generalized areas of high and low crime are. With this data, they can make better assumptions as to why distributions have or have not changed over a time period, and what makes one area more susceptible to crime than another. Then, similar to the previously mentioned techniques, targeting work or intervention can take place. These concepts can be seen in Maps 4 and 5, where single and dual KD analysis was used.

Single KD analysis is used when analyzing the spread of a single variable, whereas dual KD investigates the relationship between two variables (or surfaces) and how both influence each other. Therefore, Map 4 looks solely at residential break and enters, and Map 5 uses pop15 as the secondary surface and shows the interaction between residential break and enters and population concentration. Essentially, single KD is similar to the standard Nnh and fuzzy mode, whereas dual KD is similar to the RNnh. These comparisons are also visible on their respective maps as well. Map 4, or single KD, shows higher crime density across the study area compared to the dual KD results of Map 5. This indicates that some of the high crime areas only appear so due to higher population, which is the idea that was analyzed earlier when comparing Nnh to

RNnh. Additionally, the single KD results coincided better with the fuzzy mode, Nnh, and the RNnh results. The areas of high crime density (central study area) detected by the single KD analysis were similar to the areas that were detected by the other techniques as well. Dual KD, on the other hand, does not agree much with the results of the other techniques as it shows a very low crime density in areas that the other techniques showed as high in crime, and vice versa. This could be because once population was taken into account, it provided another layer of criteria for the data to consider, and measured the risk of crime density based on population. However, it is interesting to see that though this was the case, and though a secondary surface was used here as with the RNnh, the results of both the dual KD and RNnh do not agree as much as expected. This could simply be a result of the interpolation within the dual KD, which therefore leads to the generalization. Additionally, because triangular interpolation was used, the “peaks and valleys” of the data are highlighted (Levine, Chapter 10), and this may have contributed to the visual uniqueness of these dual KD results.

Appendix 1

Knox Index: Interaction of Space and Time

Sample size ....: 2152

Measurement type ......: Direct

Input units ....: Meters

Time units ....: Hours

Simulation runs ....: 19

Start time ....: 01:38:42 PM, 03/08/2018

"Close" time ....: 6.00000 hours

"Close" distance ....: 5000.00000 m

	Close in space(1)	Not close in space(0)	
Close in time(1)	325473	986929	1312402
Not close in time(0)	242964	759110	1002074
	568437	1746039	2314476

Expected:

	Close in space(1)	Not close in space(0)	
Close in time(1)	322326.89199	990075.10801	1312402.00000
Not close in time(0)	246110.10801	755963.89199	1002074.00000
	568437.00000	1746039.00000	2314476.00000

Chi-square ....: 94.01612

P value of Chi-square: 0.00010

End time ....: 01:38:42 PM, 03/08/2018

Distribution of simulated index (percentile):

Percentile	Chi-square
min	0.00016
0.5	0.00016
1.0	0.00016
2.5	0.00016
5.0	0.00016

10.0	0.00292
90.0	1.68066
95.0	1.71090
97.5	1.71090
99.0	1.71090
99.5	1.71090
max	1.71090

Simulation ended ......: 01:38:45 PM, 03/08/2018

Works Cited

Levine, Ned. (n.d.) *CrimeStat IV Chapter 5: Spatial Autocorrelation Statistics.* <http://ibis.geog.ubc.ca/courses/geob479/notes/crime/CrimeStat%20IV%20Chapter%205.pdf>

Levine, Ned. (n.d.) *CrimeStat IV Chapter 6: Distance Analysis I and II.* <http://ibis.geog.ubc.ca/courses/geob479/notes/crime/CrimeStat%20IV%20Chapter%206.pdf>

Levine, Ned. (n.d.) *CrimeStat IV Chapter 7: Hotspot Analysis of Points: I*. <http://ibis.geog.ubc.ca/courses/geob479/notes/crime/CrimeStat%20IV%20Chapter%207.pdf>

Levine, Ned. (n.d.) *CrimeStat IV Chapter 10: Part IV: Spatial Modeling I.* <http://ibis.geog.ubc.ca/courses/geob479/notes/crime/CrimeStat%20IV%20Chapter%2010.pdf>

Levine, Ned. (n.d.) *CrimeStat IV Chapter 12: Space-Time Analysis.* < http://ibis.geog.ubc.ca/courses/geob479/notes/crime/CrimeStat%20IV%20Chapter%2012.pdf>