Analysis of Crime Patterns in the Ottawa Nepean Area

Chelsey-Ann Cu

32482135

GEOB 479 L2A

University of British Columbia

Dr. Brian Klinkenberg

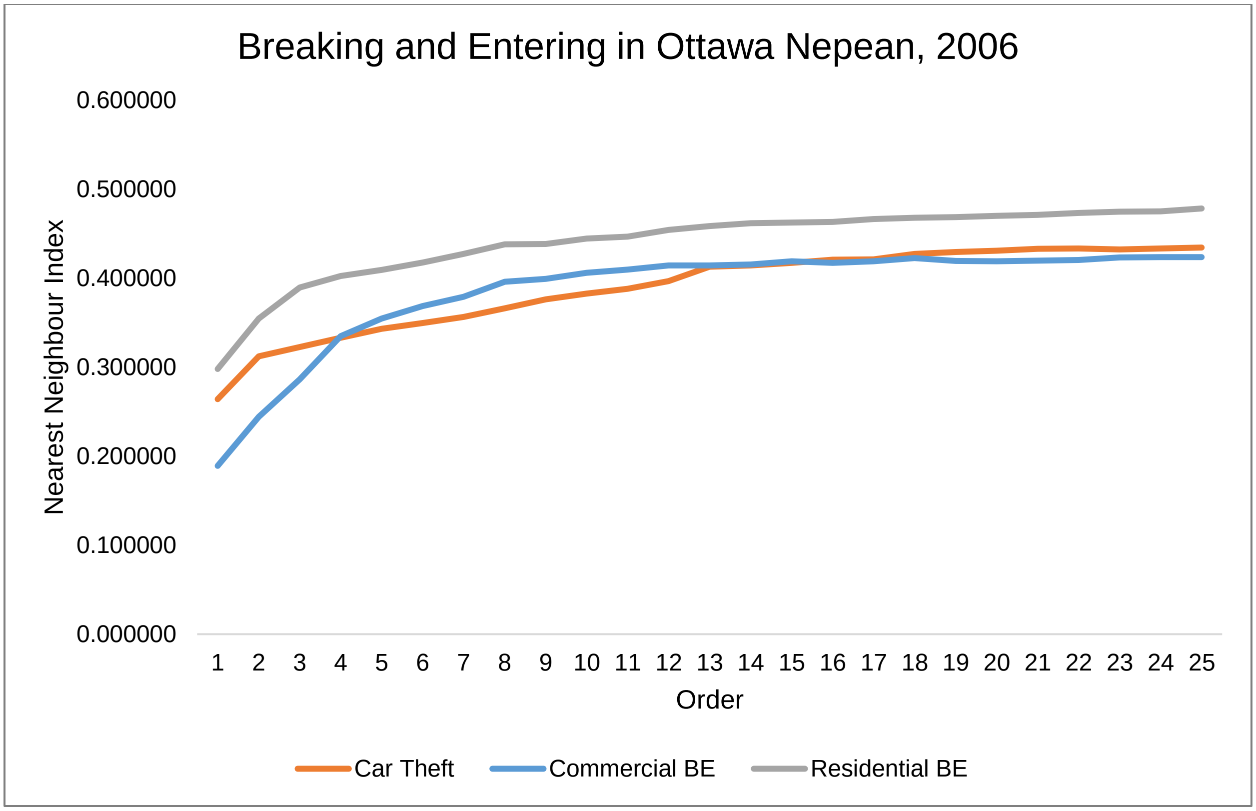
9 March 2018

I. Introduction

Through the use of several different types of spatial distribution analysis, this exercise evaluated the distribution pattern of crimes in the Ottawa Nepean area. The analysis of crime, such as what is seen in this report, can be a crucial asset to aiding law enforcement personnel determine strategies and policies for improving the city. Through different statistical analysis techniques, such as nearest neighbor index, Moran's Index, hot spot analysis, Knox Index, and Kernel density estimation, varying patterns can are revealed. The results of this lab show clusters which can be used to determine policing strategies and policies surrounding areas that should be more closely monitored.

II. Nearest Neighbour Index

The nearest neighbor index provides information on the distance between each crime location and a set number of the closest instances of the same category. This index is determined through the comparison of the distance between a point and any given neighbour (National Institute of Justice, 2010). In this case, it was specified that the comparison be made with 25 of the closest neighbouring points. With this analysis, a value from zero to one is generated, with one being a distance that is expected and zero representing total spatial aggression. The order along the x-axis represented the rank of distance each neighboring point is to any given point being analyzed, making one the closest point, and 25 being the 25^{th} nearest point.



Car Theft z-score: -65.3034
Commercial BE z-score: -67.8104
Residential BE z-score: -77.4486

Figure 1. Nearest Neighbour Analysis Results

As Figure 1 shows, the crime data is more spatially aggregated than expected since all index values are less than one. Therefore, the distance to the nearest crime is less than what would be expected in a random pattern. The index shifts up and down depending on the type of crime. For instance, residential break-ins have a slightly higher index value, making them more spatially aggregated compared to car thefts and commercial break-ins. This is could be attributed to residential areas being naturally more dispersed in comparison to commercial regions and cars. Moreover, certain neighbourhoods could be targeted based on wealth dispersion and amount of security in varying residential areas. Therefore, this would result in less spatial clustering of the other two categories.

III. Moran's Index

The Moran's Correlogram provides information on the diffuse or concentrated nature of the data in terms of spatial autocorrelation. From a correlogram, it can be determined whether hot spots are isolated concentrations or by-products of spatial clustering over a large area (National Institute of Justice, 2010). Also, it shows the degree of decline in spatial autocorrelation with distance. Here, a value of one would mean perfect spatial autocorrelation, while a value of zero would represent total random distribution. The x-axis shows distances (in metres), with the lowest distance representing the closest dissemination areas.

In this case, the crimes are aggregated to dissemination areas in Ottawa. From Figure 2, the correlogram of the different intensities, one can note that residential break-ins have the highest spatial autocorrelation followed by car theft and commercial break-ins, with population over the age of 15 representing the default, null situation. The null value is indicative of a situation in which crime is distributed based on population distribution, meaning that the Moran's I values for all types of crimes should equal that of the population values. Since the values for car theft, commercial break-ins, and residential break-ins rank higher than the null values on the Moran's Index, it reaffirms their spatial autocorrelation links to population distribution coupled with other underlying trends. Based on the graph, the dissemination areas that experience high crime rates, are located in close proximity other dissemination areas that experience the same phenomenon.

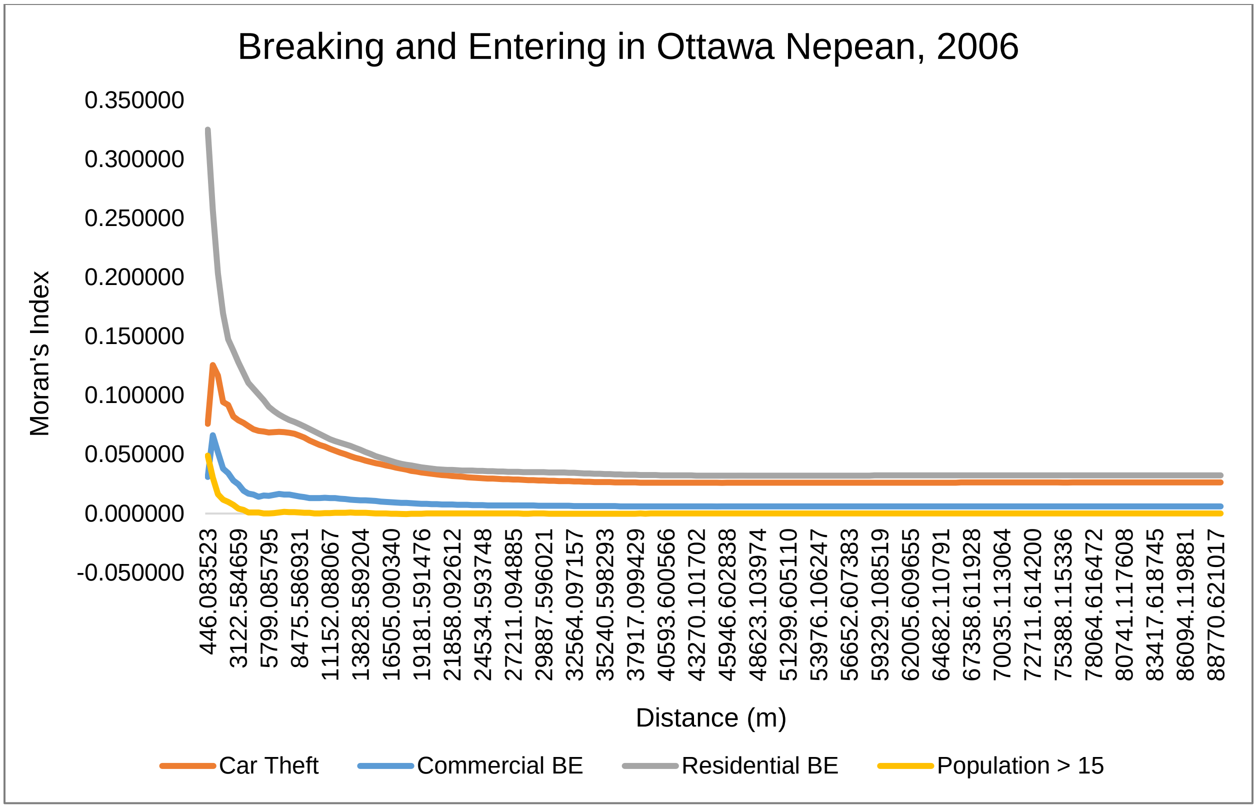


Figure 2. Results of Moran's I of different crime types

Looking at the results from the Moran's Index in comparison to the results from the nearest neighbour analysis, there are different patterns that emerge. Residential crimes, which were previously analyzed to be the least spatially aggregated in the nearest neighbour analysis, are now seen as the most spatially autocorrelated crime type according to Moran's Index. Again, the same initial reasoning put forth for the nearest neighbour analysis can be applied here. Residential areas are more widespread as opposed to the other categories of crime types. Thus, while individual residential break-ins are not significantly spatially aggregated, as a whole residential break-ins are clustered within areas that are classified as residential land use, creating a correlation.

IV. Fuzzy Mode Hot Spot Clusters and Nearest Neighbour Hierarchical Spatial Clustering

The fuzzy mode cluster results are derived from the frequency of residential break-ins within 1000 metres of a point. As seen in Figure 3, most of the crimes occur in the downtown core of the city where there is a cluster of red points indicating 99-247 reported crimes within a 1000 metre area. As one moves further away from the city's core, the dominant clustering classes are green (1-20 reported crimes), light green (21-26 reported crimes), and yellow (27-45 reported crimes) as opposed to red. This shows a significant decrease in residential break-ins in more suburban areas and also the high spatial aggregation and autocorrelation of the data.

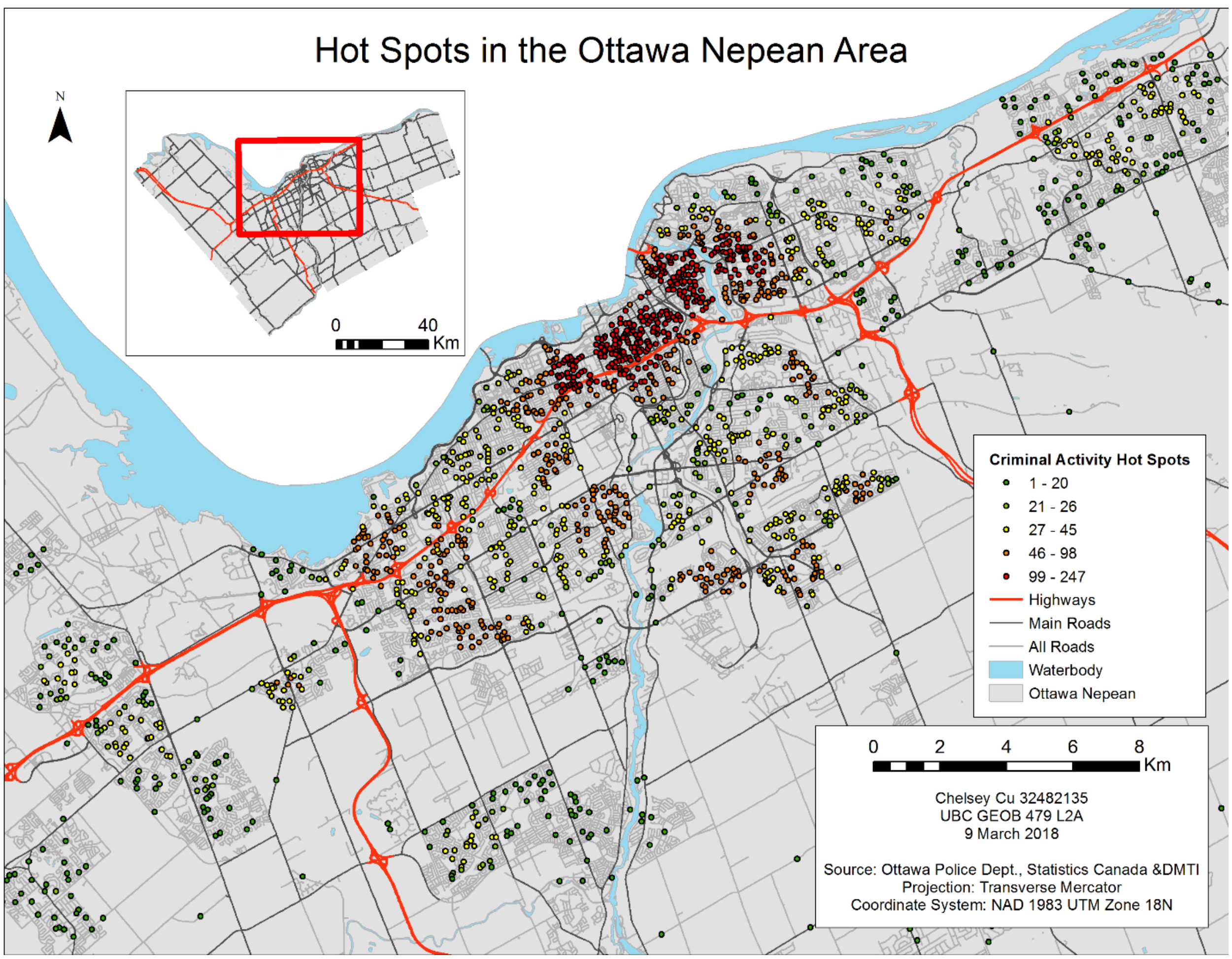


Figure 3. Results of fuzzy mode hot spot cluster analysis

The nearest neighbour hierarchical spatial clustering analysis is used to identify hot spot areas rather than individual points that cluster. This is done by repeatedly grouping reported break-in incidents that were spatially close to one another into single clusters. The resulting ellipses (shown in dark blue on Figure 4), summarizes areas where similar number of residential crimes have occurred within a fixed distance of 1000 metres. This provides a clearer indication of where the hot spot clusters are located.

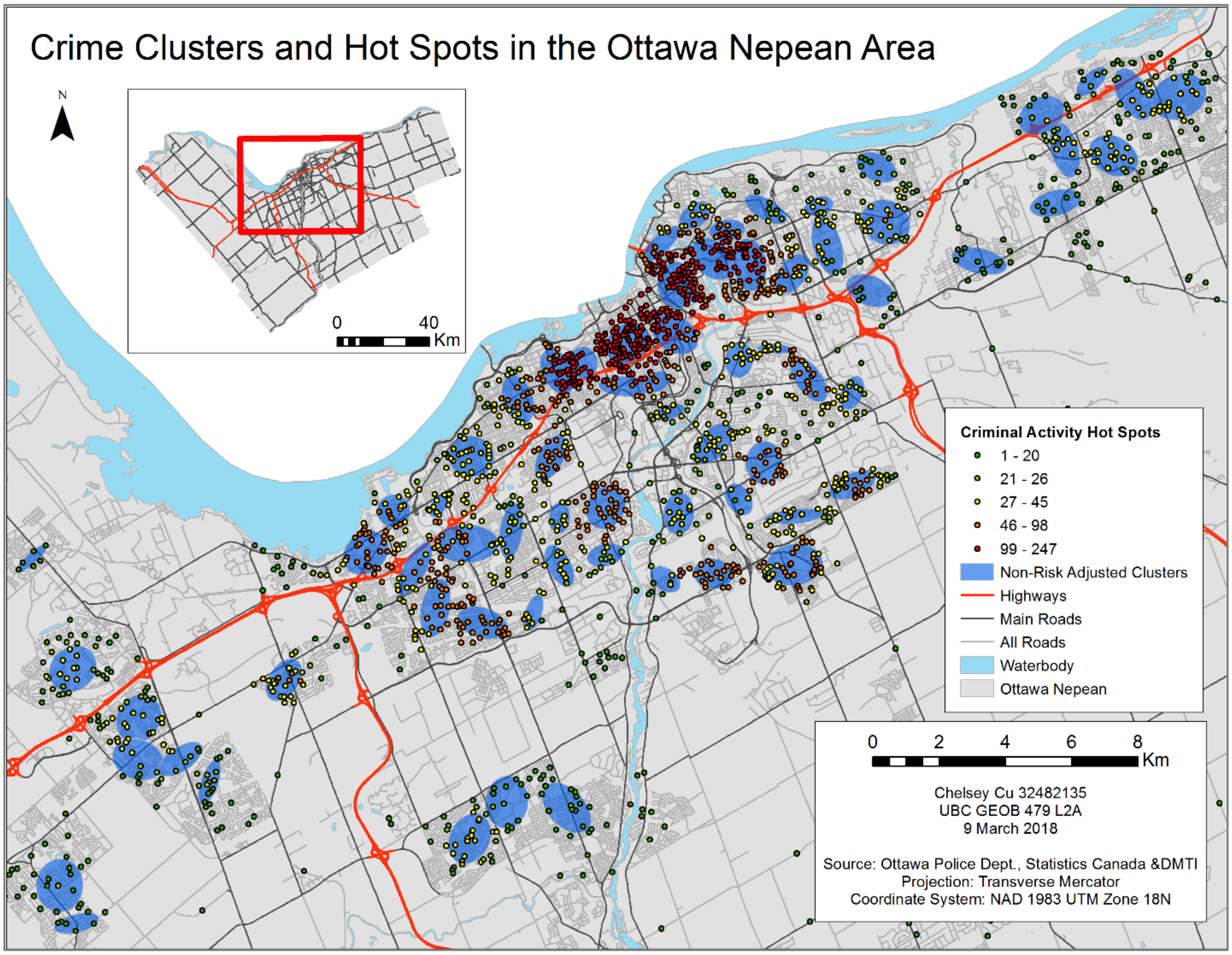


Figure 4. Results of fuzzy mode hot spot cluster analysis and nearest neighbour hierarchical spatial clustering analysis

Results from both analysis, as seen in Figure 4, are quite similar. In areas where large clusters of criminal activity hot spots are located, non-risk adjusted clusters, which are determined by the nearest neighbour analysis, are also present. This can be expected because the nearest neighbour analysis determines if there are any clusters in the data, in other words, areas where similar frequency of crimes are being reported. Notably, all the red hotspots (99-247 crime sites within 1000 metres) are identified as clusters. From this it can be determined that the clusters are partially a function of land use since the largest clusters occur where the land is mostly dedicated for residential uses. These clusters can help to determine the linkage of crime to underlying geographical social, political and economic conditions.

V. Non-Risk and Risk-Adjusted Clustering

The nearest neighbour hierarchical spatial clustering analysis (also known as the non-risk-adjusted spatial clustering) mentioned in the previous section looks solely at the frequency of crimes and lacks spatial population density. Figure 5 shows the same results as Figure 4 but with the hot spot points removed to clearly show the formed clusters. The issue with using the map in Figure 5 is that without population density as a factor, it inaccurately depicts the risk to individuals that reside in these areas.

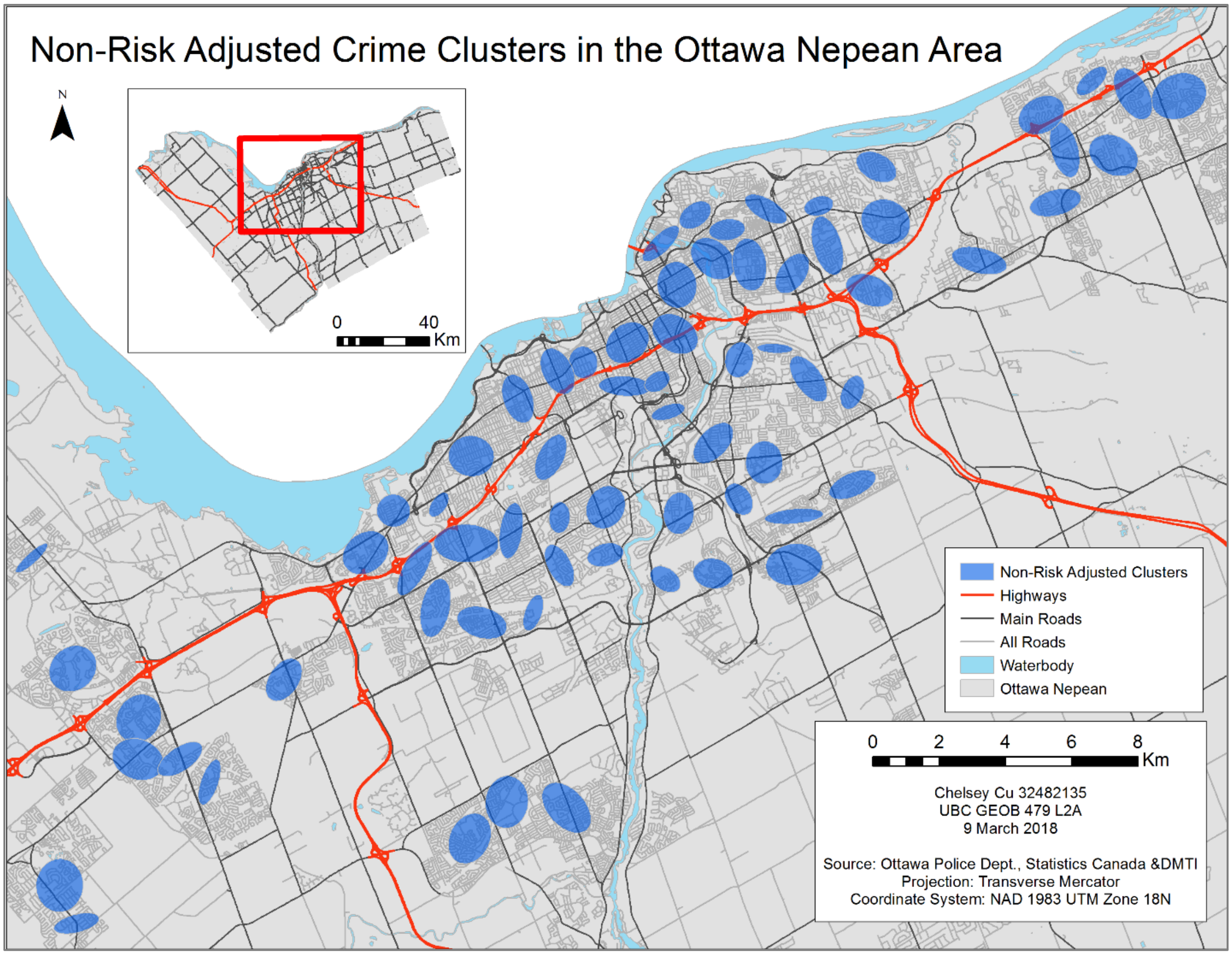


Figure 5. Results of the nearest neighbour hierarchical spatial clustering analysis

To correct for this, first, second and third order risk-adjusted clusters are determined based on the population over the age of 15 living in the area (see Figure 6). The first order clusters, similar to the non-risk-adjusted clusters, are determined based on number of reported crimes and distance. Therefore, the first order clusters (shown in red) are the most similar to the non-risk-adjusted clusters that were shown in Figure 4 and 5. The difference between nearest neighbour clustering and first order clusters is that, by taking into account population densities, the latter results show the areas that are pose the highest risks rather than merely the highest volume of crimes. The second order clusters are formed through identifying clusters of groups of criminal activities based around the centers of the first order clusters. Likewise, the third order clusters are determined in a similar grouping fashion of criminal activities, but representing a convergence of all the crime sites. The latter order is the final cluster possible for the analysis as there is no larger cluster possible past this point. Through the different order clusters, becomes easier to generalize areas where a higher number of crimes are reported.

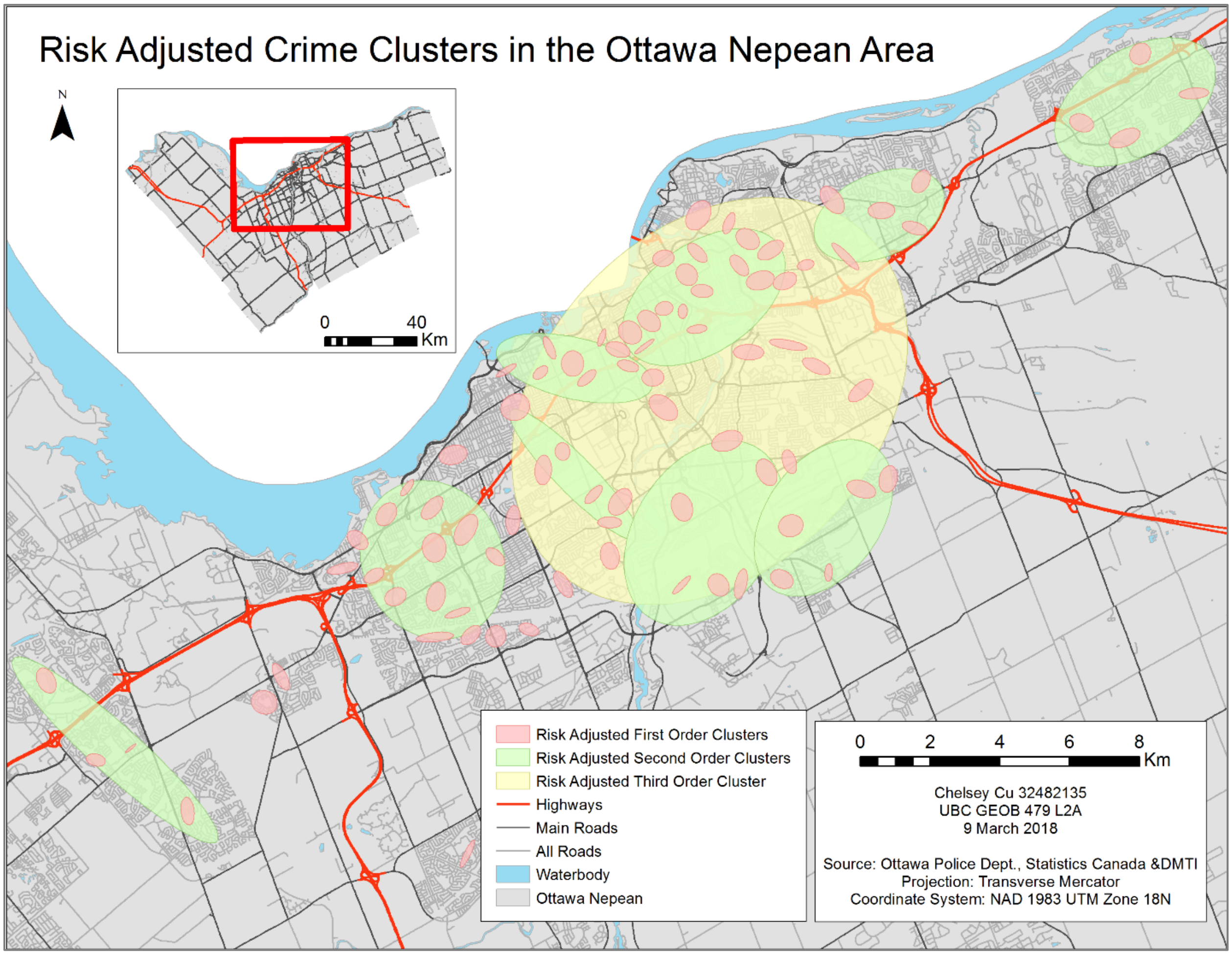


Figure 6. Results of the risk-adjusted crime cluster analysis

VI. Knox Index

The Knox Index compares the relationship between incidents in terms of space and time (National Institute of Justice, 2010). In the analysis, the distance between car theft points is divided into those that are close in time and those that are not. The analysis was run 19 times and looking at the 2 x 2 table produced in Appendix A, one can see that values have been predicted given that "closeness in time" was set to six hours and "closeness in distance" was set to five kilometres. The results are as followed: close in time and space (325473), not close in time and space (759110), close in time and not close in space (986924), and not close in time and close in space (242964). These values can be compared to the expected frequencies to determine the significance of the model. In this case, our results showed that the chi-square statistic is 94.02. The high value shows that car theft distribution is not random. To reinforce the findings, one can also look at the p-value (0.00010) which indicates that this model would only arise by chance 0.01 percent of the time. As shown, the Knox Index helps to determine the spatial and temporal distribution patterns of car thefts.

The problem with this particular analysis is that the software does not take into account crimes that are committed within a six-hour period overnight (ie. one point at 23:00 and another point at 05:00) because of the linearity of the analysis. Furthermore, depending on the cut-off point of the analysis, the chi-square statistic that is produced changes.

VII. Kernel Density Estimation

Kernel density estimation "is a technique for generalizing incident locations to an entire area" (National Institute of Justice, 2010, p. 10.1). This differs from spatial distribution methods in that it takes data incidents which were statistically determined and generalizes it to an entire region by placing a symmetrical surface over each point. The result is an estimated intensity variable (Z-value) at a particular location.

A single and duel Kernel density analysis were conducted. For this, a similar method for producing the risk –adjusted cluster analysis was used with the number of crimes set to 250 per square metre area. The single Kernel density results were of absolute density crimes. Here, as seen in Figure 7, a high number of crimes are clustered in the core of the city, giving the same results as the previous analyses (ie. hot spot clusters, non-risk adjusted clusters and risk-adjusted clusters). In the duel Kernel density analysis looks at the relative density of residential break-ins with regards to the density of the population over 15 years old. The results show areas with high rates of reported crime in comparison to the number of residents in the area. In the duel Kernel density map, one can see that the core of the city is no longer a hot spot for crimes (see areas in red). This is because with an increase in population, there would be an increased volume in crime rates. Therefore, the ratio of population to reported crime in downtown area is lower than the same ration in some suburban residential areas.

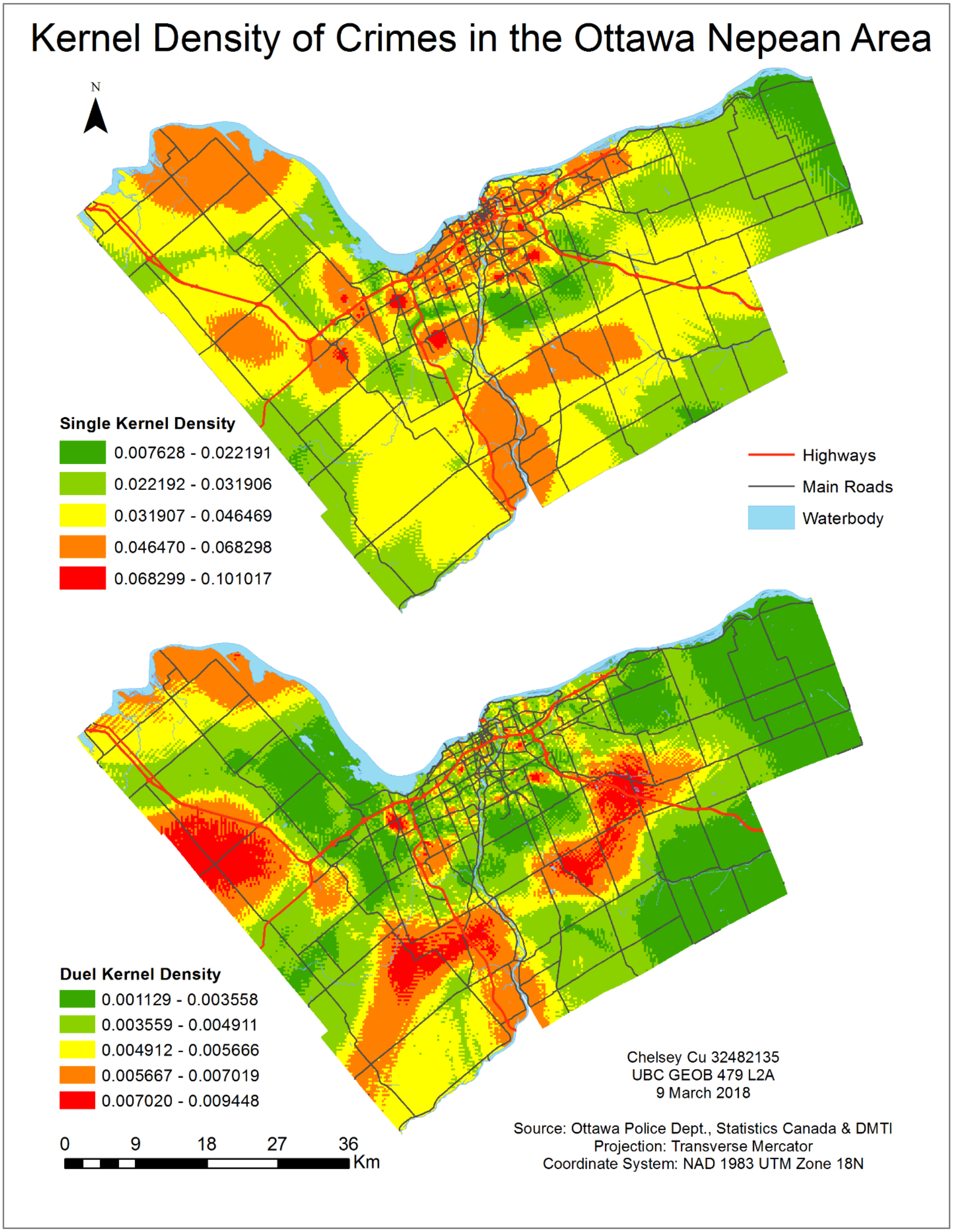


Figure 7. Results of the single and dual Kernel density estimation

References

National Institute of Justice. (2010). Houston, Texas.

<https://www.nij.gov/topics/technology/maps/pages/crimestat-downloads.aspx>

Appendix A

These are the results from the Knox analysis, which analyzed the space-time patterns in car theft data.

```
Knox Index: Interaction of Space and Time
-----------

    Sample size ...........:  2152
    Measurement type ......:  Direct
    Input units .... ......:  Meters
    Time units ............:  Hours
    Simulation runs .......:  19
    Start time ............:  02:56:50 PM, 03/01/2018

    "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 ..............:  02:56:50 PM, 03/01/2018

    Distribution of simulated index (percentile):
    Percentile      Chi-square
    ---------- ---------------
           min         0.00115
           0.5         0.00115
           1.0         0.00115
           2.5         0.00115
           5.0         0.00115
          10.0         0.01652
          90.0         3.62385
          95.0         5.16675
          97.5         5.16675
          99.0         5.16675
          99.5         5.16675
           max         5.16675

    Simulation ended ......:  02:56:53 PM, 03/01/2018
```

Appendix B

These are the results from the Moran's Index for each of the Intensity variables.

I. Car Theft:

```
Spatial Autocorrelation for Point Data:
---------------------------------------

    Sample size ......................: 1328
    Measurement type .................: Direct
    Start time........................: 11:06:09 AM, 03/02/2018

    Moran's "I" ......................:     0.026264
    Spatially random (expected) "I" ...:   -0.000754
    Standard error of "I"......... ...:    0.001080
    Normality significance (Z) ........:   25.016781
    p-value (one tail) ................:     0.0001
    p-value (two tail) ................:     0.0001
    Randomization significance (Z) ....:   25.113107
    p-value (one tail) ................:     0.0001
    p-value (two tail) ................:     0.0001

    End time..........................: 11:06:09 AM, 03/02/2018
```

II. Commercial Breaking and Entering:

```
Spatial Autocorrelation for Point Data:
---------------------------------------

    Sample size ......................: 1328
    Measurement type .................: Direct
    Start time........................: 11:10:55 AM, 03/02/2018

    Moran's "I" ......................:     0.006119
    Spatially random (expected) "I" ...:   -0.000754
    Standard error of "I"......... ...:    0.001080
    Normality significance (Z) ........:    6.363790
    p-value (one tail) ................:     0.0001
    p-value (two tail) ................:     0.0001
    Randomization significance (Z) ....:    6.462475
    p-value (one tail) ................:     0.0001
    p-value (two tail) ................:     0.0001

    End time..........................: 11:10:55 AM, 03/02/2018
```

III. Residential Breaking and Entering:

```
Spatial Autocorrelation for Point Data:
---------------------------------------

    Sample size .......................: 1328
    Measurement type ..................: Direct
    Start time.........................: 11:09:35 AM, 03/02/2018

    Moran's "I" .......................:     0.032215
    Spatially random (expected) "I" ...:    -0.000754
    Standard error of "I"......... ...:     0.001080
    Normality significance (Z) ........:    30.531184
    p-value (one tail) ................:      0.0001
    p-value (two tail) ................:      0.0001
    Randomization significance (Z) ....:    30.561449
    p-value (one tail) ................:      0.0001
    p-value (two tail) ................:      0.0001

    End time...........................: 11:09:35 AM, 03/02/2018
```