Lab 4 – Crime Analysis using CrimeStat

Connor Guilherme

31681144

GEOB479

Mar. 9, 2018

Table of Contents

1.0 Introduction

This report explores the different tools within the CrimeStat software to analyze residential break and enter data, commercial break and enter data and car theft data in the city of Ottawa. CrimeStat offers a variety of techniques to analyze the patterns of crime and outputs files that can be opened in ArcGIS to view spatial results. The combination of CrimeStat and ArcGIS offers valuable data analysis capability that can be applied to land use planning to help project potential theft patterns in urban areas and minimize the environmental conditions that facilitate crime.

2.0 Data and Methods

This report uses numerous techniques to analyze Ottawa crime data. Through the use of the nearest neighbour index and Moran's index, spatial autocorrelation of crime in the city was determined. Hot spot analysis and nearest neighbour hierarchical spatial clustering techniques were used to map residential break and enter crimes. Both non-risk adjusted and risk-adjusted clusters were determined for the residential break and enter crime data to determine the influence of including population in the cluster calculation. The Knox index was used to test the spatial and temporal clustering of car theft based on the location of the crimes and the time of day that they occurred. Finally, single and dual kernel density maps were created to show the interpolated residential break and enter crime intensity for the entire study area.

3.0 Discussion of Results

3.1 Nearest neighbour index results

The nearest neighbour index measures distribution of spatial data and quantifies this measure such that nearest neighbour index values closer to zero indicate higher spatial aggregation or clustering. When the nearest neighbour is greater (closer to one) the spatial distribution is more random. In this report, the number of nearest neighbours to be computed was set as 25. This means that the distances of 25 points was considered for the nearest neighbour index calculation such that point 15 is further from point 5 than point 10.

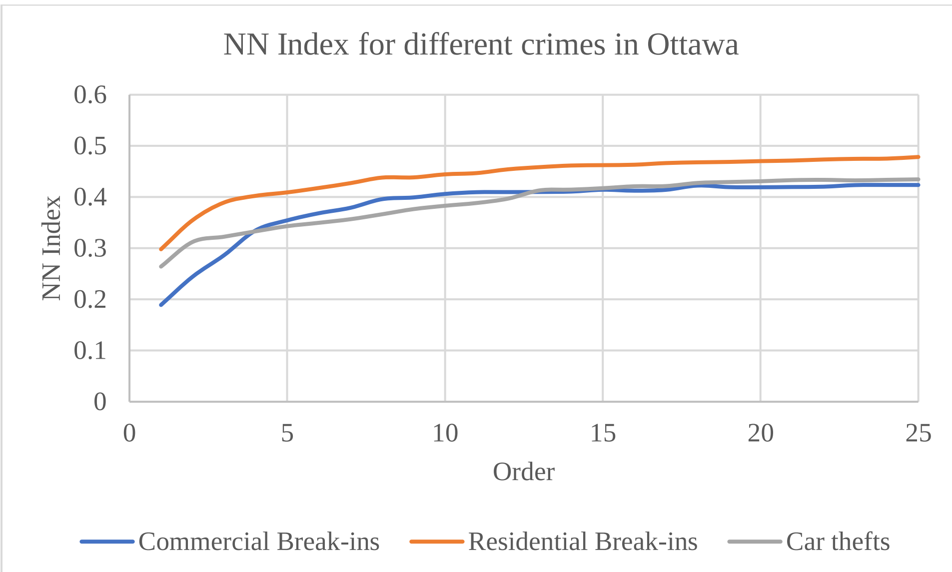

Figure 1: Nearest neighbour (NN) index results for different crime types in Ottawa

As Figure 1 shows above, for all three crime types considered, the nearest neighbour index increases as the order increases. This means that at further distances there is less clustering and more randomness to the crimes. At the first order, the commercial break and enters are more clustered than residential break and enters and car thefts. Therefore, commercial break and enters have the strongest autocorrelation of the three crime types in the analysis (for the first order). In cities, there are large areas zoned for commercial buildings while residential areas are more dispersed throughout the city, therefore there is greater clustering of commercial break and enters than residential break and enters. Figure 1 also shows that as the order increases from order 1 to about order 15 the spatial autocorrelation is strongest and thereafter it stabilizes. This is because closer neighbours have greater spatial aggregation.

3.2 Moran’s index spatial autocorrelation results

The Moran’s index is a measure of spatial autocorrelation and in this analysis it is determined by the aggregating crimes based on dissemination areas. A positive Moran’s I indicates clustering (spatial autocorrelation) while a negative Moran’s I indicates dispersion and a Moran’s I of zero indicates no spatial autocorrelation. Since the Moran’s I in this analysis is determined by aggregating the crimes based on dissemination areas, the ranking of spatial autocorrelation for each crime differs than in the above nearest neighbour analysis. Above, commercial break and enters are more clustered than car thefts which are more clustered than residential break and enters.

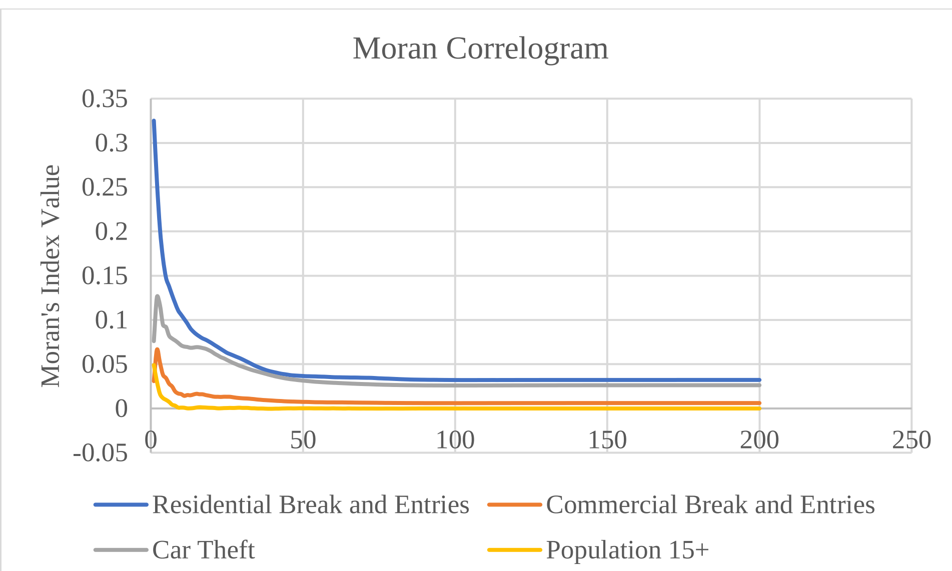

Figure 2: Moran’s index for different crime types in Ottawa

But, the Moran’s index values indicated in Figure 2 above, show that the residential break and enters are more clustered than car thefts which are more clustered than commercial break and enters. Since residential areas are dispersed around the city, the aggregation of residential break and enter crimes to dissemination areas results in greater spatial autocorrelation for this crime type in the Moran's I analysis. It should also be noted that all of the crime data lines have higher Moran's I values than the baseline or 'null value' population over 15 years of age data. This means that the spatial autocorrelation for each crime type is not due population distribution.

3.3 Fuzzy hot spot analysis and nearest neighbor hierarchical spatial clustering

The hot spot analysis identifies points in Ottawa with varying frequency of residential break and enter crimes. As Figure 3 shows, the downtown core is mainly covered in red points that represent 187 to 247 residential break and enters. Areas, such as the downtown core, where points are clustered together represent higher spatial autocorrelation. As the distance from the core increases the frequency of residential break and enters and the clustering of points decreases. Thus, as the distance from the core increases the spatial autocorrelation generally decreases, but pockets of clustering can be found, likely in areas where the large housing developments outside of the downtown core.

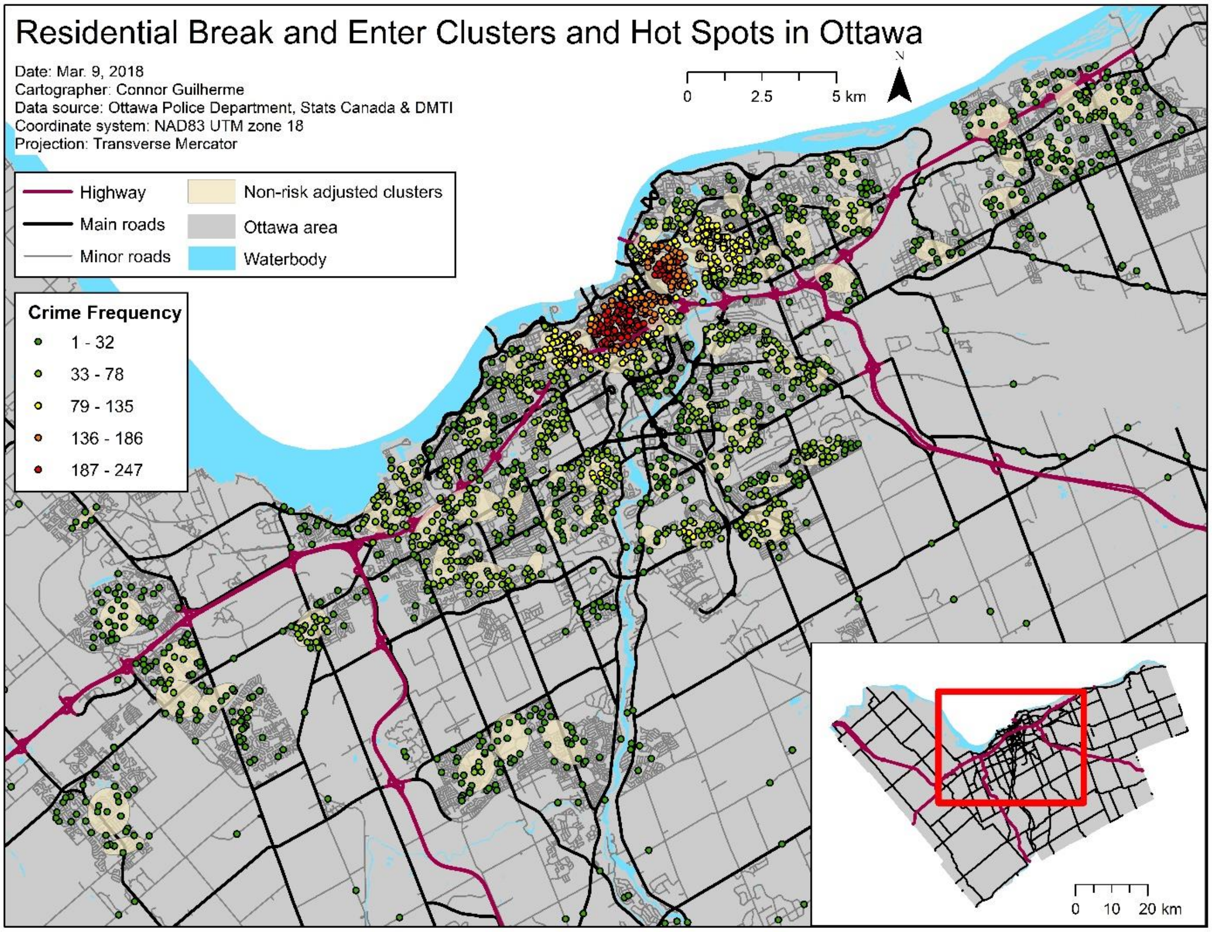

Figure 3: Hot spot analysis crime frequency points overlaid on residential break and enter clusters

The beige clusters in Figure 3 highlight the areas where there is greater spatial autocorrelation and were set to have 1000m radii. As is shown above there is clear overlap of the crime frequency points and the clusters. The overlap of the nearest neighbour clusters and the frequency points suggests that these areas have frequently occurring residential break and enter crimes due to certain environmental conditions (e.g. house density). Another point to note is that the clusters and points typically overlap in areas with lots of minor roads which are likely home to subdivision developments and thus higher house density.

3.4 Clustering crimes by risk

The hot spot frequency points and the clusters in Figure 3 do not represent risk-adjusted data. Since population distribution varies across cities, the risk of residential break and enter varies for each location. For example, areas with more houses (i.e. more people) might be greater targets for break and enter and thus the risk to a person in one area may be different than the risk in another area. The first order non-risk adjusted clusters are shown in Figure 4 below.

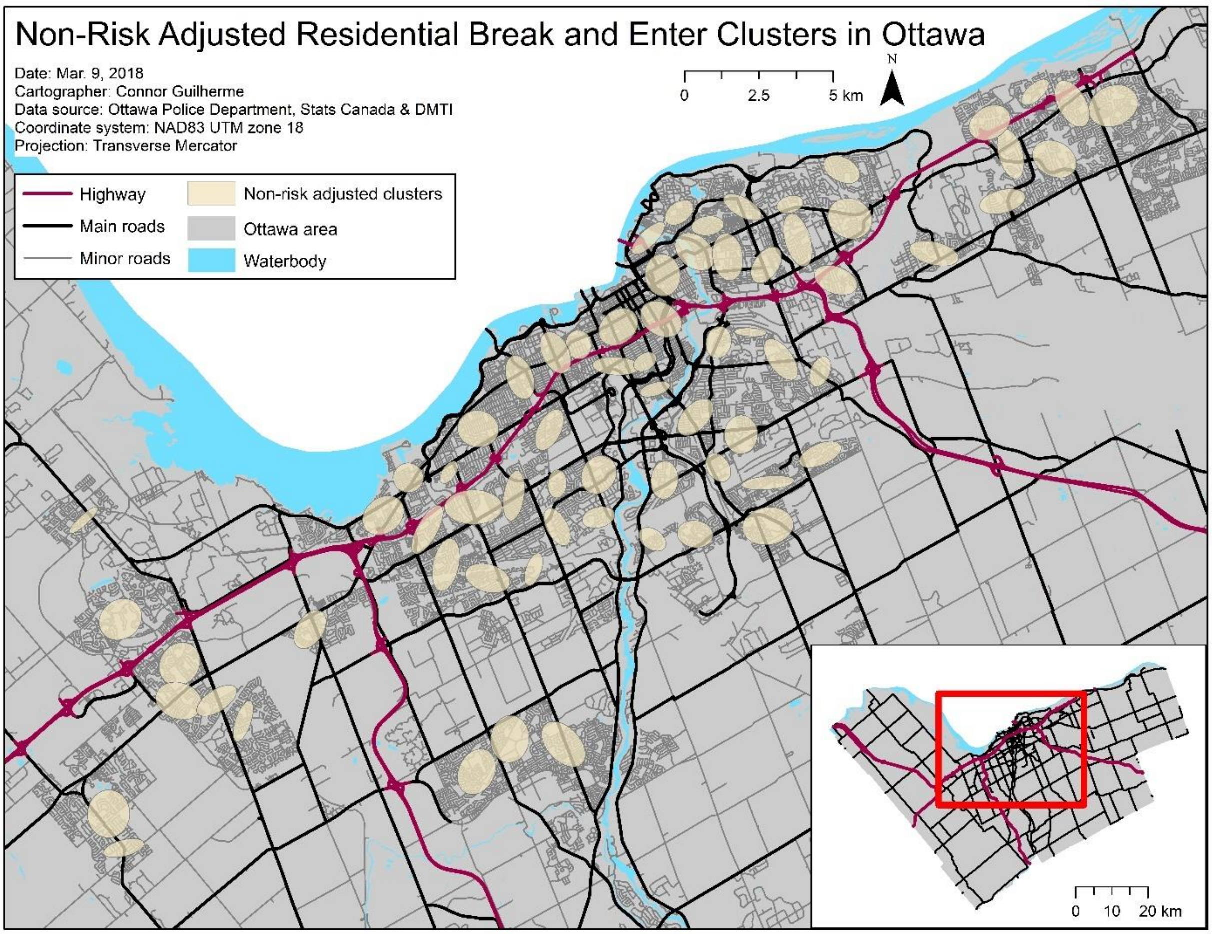

Figure 4: Non-risk adjusted residential break and enter clusters in Ottawa area

It is necessary to account for population in order to identify risk-adjusted residential break and enter clusters and thus, better represent the crime risk. Figure 5 below shows the first, second and third order risk adjusted clusters for the same area in Ottawa.

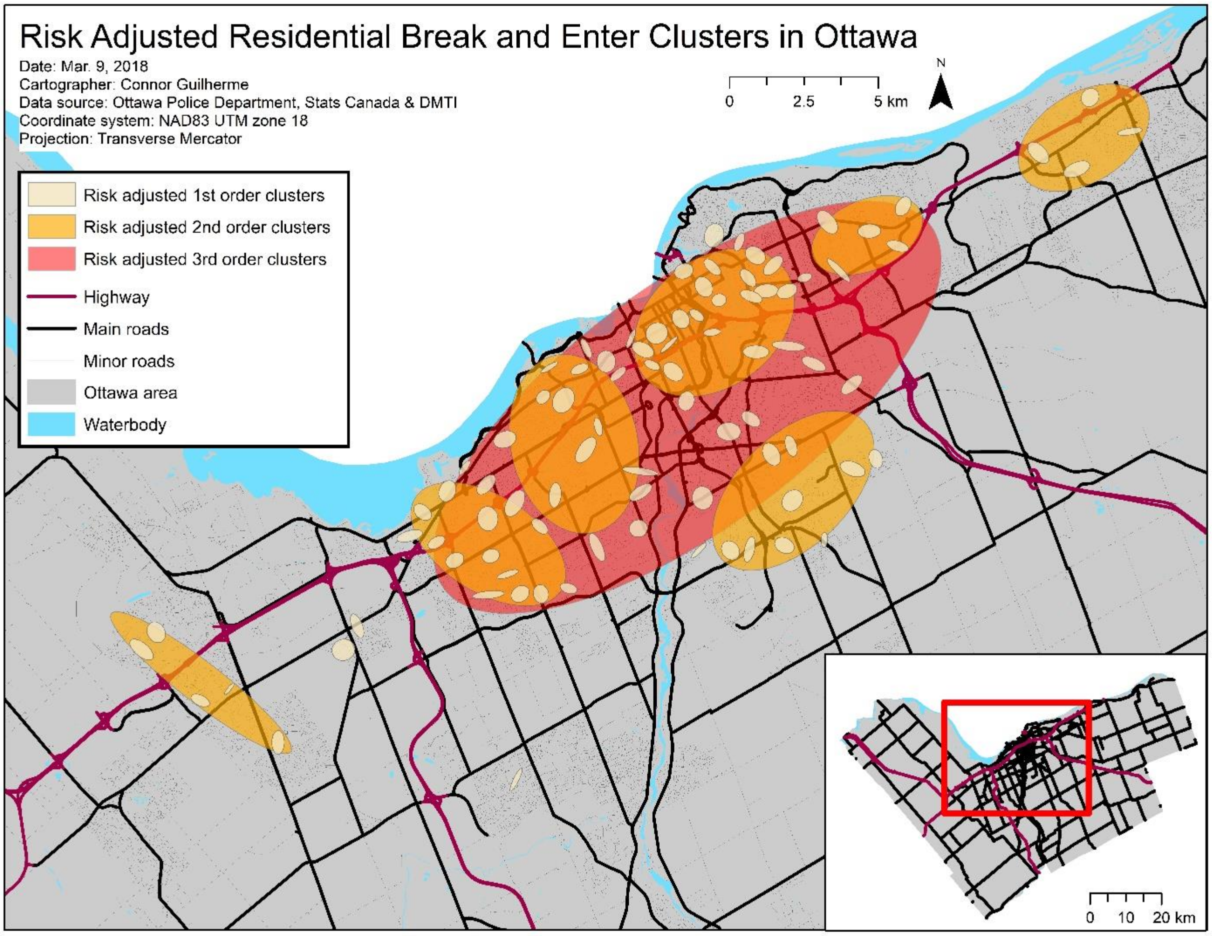

Figure 5: Risk adjusted residential break and enter clusters in Ottawa area

It can be seen that the non-risk adjusted clusters in Figure 4 are most similar to the first risk adjusted clusters in Figure 5, but they are not identical. This is because the risk adjusted clusters account for the population data in order to determine areas of strong spatial autocorrelation for the residential break and enter data. Therefore, the adjusted cluster identify areas of higher risk rather than just greatest frequency of crime. The construction of three orders is a function of the risk adjustment by the CrimeStat software. As the first order clusters are based on distances of residential break and enter frequency points, the second order clusters are based on the distances of the first order clusters and the third order cluster is based on the distances of the second order clusters. As the clusters get larger from first order to second order and second order to third order the scope of the residential break and enter data becomes broader so a more generalized spatial trend can be seen. With this generalization, some of the clusters from the previous order are not encompassed. For example, the narrow second order cluster in the West is not encompassed by the single large third order cluster.

3.5 Knox index results

The Knox index is used to determine if there is both spatial and temporal clustering of car thefts. Therefore, both the geographic location of the theft and the time that it occurred are inputs for the Knox index analysis. To determine whether there was clustering the Knox index has a "closeness" setting which was set as car thefts within 6 hours of each other (to determine temporal clustering) and car thefts within 5km (to determine spatial clustering). 19 simulations were run in order to classify closeness in terms of 4 categories: close in time and space (322,327), close in time but not close in space (990,075), not close in time but close in space (246,110), and not close in time or space (755,964). Therefore, the Knox index suggests that car thefts are most likely to be close in time but not close in space. Refer to the appendix for the txt file with the complete results of the Knox index.

The chi-square value for the Knox index quantifies the significance of the classification that it determines. Since the chi-square value is a measure of the difference between the classification based on the data compared to the values that the simulation produces, a high chi-square value means that the classification is significant. The chi-square value determined by the Knox index for this analysis was 94.016, thus the classification described above is significant. The p-value of the chi-square is very low (0.0001), this also suggests that the classification is significant.

3.6 Kernel density results

The hot spot and cluster analysis presented in Figures 3, 4 and 5 represent residential break and enter crimes as either points or areas (clusters). The kernel density interpolates the data to create an estimate of residential break and enter crime intensity for the entire study area. For this analysis the study region was divided into 250 square metre grids and a value for crime intensity was assigned to each grid square. The benefit to this analysis is that for every point in the study area there is a quantitative value for residential break and enter crimes. In the hot spot, cluster and nearest neighbour techniques a large portion of the study area is not quantified and thus there is only information for areas that show high crime frequency and/or high spatial autocorrelation.

Figure 6 includes both a single kernel density map and a dual kernel density map. The single kernel density map is similar to the non-risk adjusted cluster map (Figure 4) because it does not consider population. The dual kernel density map considers a secondary surface – population over 15 years of age. Thus, the dual kernel density map is similar to Figure 5, the risk adjusted cluster map. Without the secondary surface it appears that the downtown area has the greatest residential break and enter crime intensity (single kernel density map), but when the population is accounted for as a secondary surface, the residential break and enter crime intensity is lower in the downtown area (dual kernel density map).

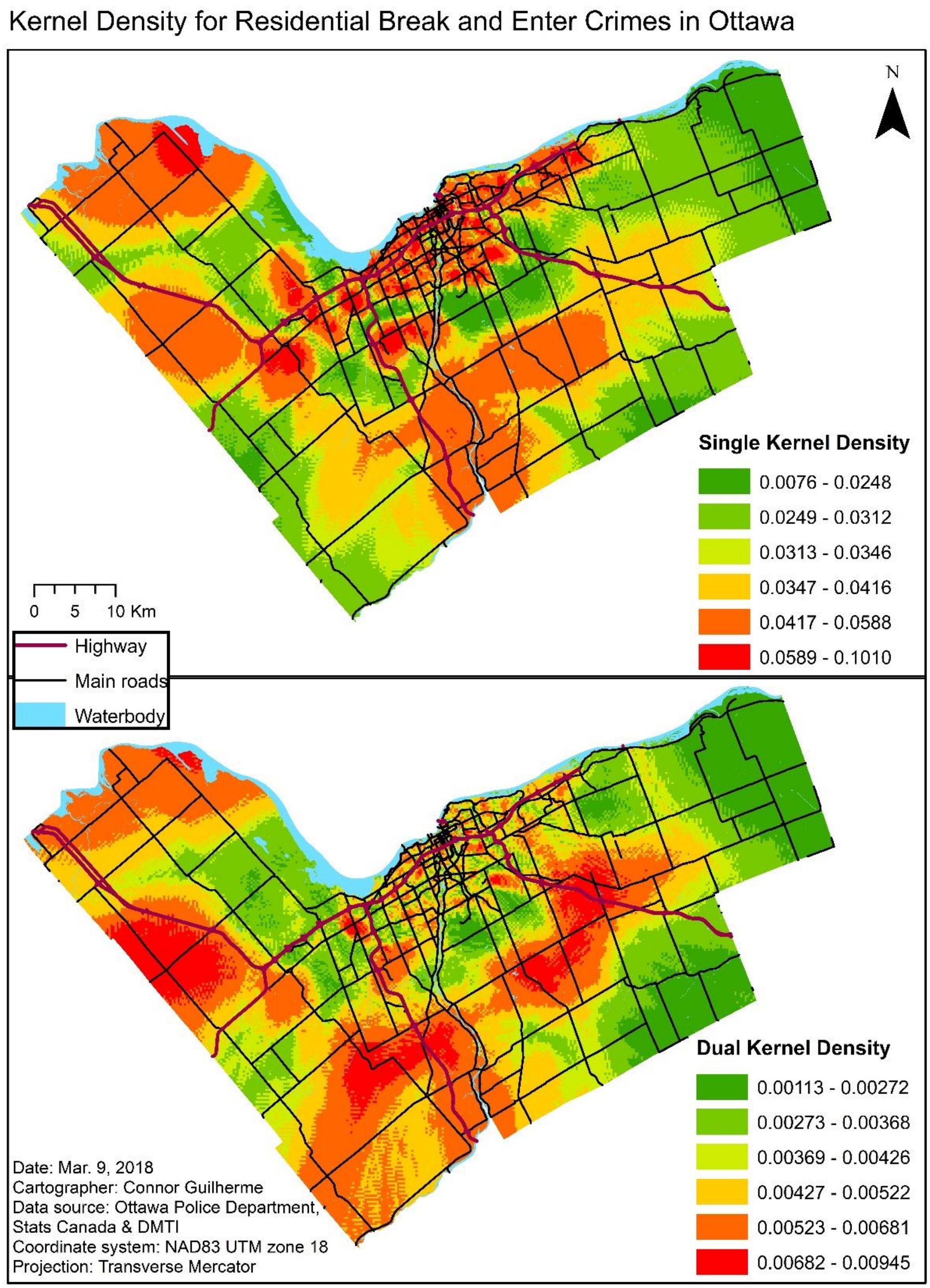

Figure 6: Kernel density of residential break and enters in Ottawa area (top map shows single kernel density and bottom map shows dual kernel density)

4.0 Conclusion

The crime analysis presented in this report uses a variety of techniques based on the CrimeStat software. As was shown through the numerous maps and techniques described above, there are limitations with each of the techniques. For example, the kernel density technique was the only technique that determined the crime intensity for the entire study area while the Knox index was the only technique that accounted for the time of the crime. Thus, it can be seen that depending on the crime that is being analyzed and the purpose of the analysis, the suitable technique may be different or a combination of the use of multiple techniques (such as this report includes) may be most suitable. This report shows that there is significant influence of environmental factors, such as land use and population, on crime in the city of Ottawa.

Works Cited

Klinkenberg, Brian, Lab 4: Crime Analysis Using CrimeStat. Geob 479. 2018. <http://ibis.geog.ubc.ca/courses/geob479/labs/lab4.htm>

Ned Levine (2015). CrimeStat: A Spatial Statistics Program for the Analysis of Crime Incident Locations (v 4.02). Ned Levine & Associates, Houston, Texas, and the National Institute of Justice, Washington, D.C. August.

Appendix:

Below are the results for the Knox identity discussed in section 3.5:

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

    Sample size ...........:  2152
    Measurement type ......:  Direct
    Input units .... ......:  Meters
    Time units ............:  Hours
    Simulation runs .......:  19
    Start time ............:  11:26:20 AM, 02/16/2018

    "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 ............:  11:26:20 AM, 02/16/2018

    Distribution of simulated index (percentile):
    Percentile      Chi-square
    ---------- ---------------
           min         0.00004
           0.5         0.00004
           1.0         0.00004
           2.5         0.00004
           5.0         0.00004
          10.0         0.06012
          90.0         3.52535
          95.0         4.84596
          97.5         4.84596
          99.0         4.84596
          99.5         4.84596
           max         4.84596

    Simulation ended ......:  11:26:23 AM, 02/16/2018
```