Lab Assignment 2

Student

Geography 418: Assignment 2 Spatial Distribution Analysis of Mischief and Breaking and Entering Occurences in the Vancouver Area

Introduction

Understanding patterns in criminal activity provides the opportunity to interpret the reasoning or sources of the crimes studied. Being aware of the spatial correlation of specific crimes can also allow for mitigation techniques to be employed reducing the frequency and potentially in a perfect world treat the cause instead of the symptoms of the activity. Through the course of this analysis the spatial attributes of criminal mischief and residential breaking and entering of residential and other (BNE) will be analyzed by a variety of point pattern analysis techniques during 2019. The spatial statistical tests that were used are: quadrat analysis, nearest neighbour analysis, k-function analysis, and kernel density estimation. With the results from each test determining potential influencing factors but the main focus will be characterizing the spatial correlation of the crimes mentioned will conclude this analysis.

Study Area

The region this analysis is performed on is Vancouver area. The greater Vancouver area itself has a population of nearly two and a half million (Statistics Canada, 2016) people making it the largest population center in BC as well as the Canadian Coastal Pacific.

Data

The data collect for this study was taken from the Vancouver Police Department's open-source data catalog. The data collected is also known to be offset and to have reduced accuracy to protect people and locations nearby and associated to the crime reports.

Data cleaning was performed prior to analysis for this report, the focus of the cleaning was to remove years, crimes and locations that were not associated with this specific analysis. Removing locations which were left blank for the protection of identity or for crimes which we committed at an unknown location, wherein x,y UTM coordinates were set to 0.

Methods

Nearest Neighbour Distance

The nearest neighbour analysis is a statistical method developed by David Waugh for comparing point patterns or spatial correlation based upon their randomness, clustering, or dispersion. The process starts by summing the Euclidean distance of each point in a dataset then dividing by the total number of point / observations in the dataset which is the mean nearest neighbour distance [1]. Following that the point density over the total area is calculated which will be used to determine what an ideal or random nearest
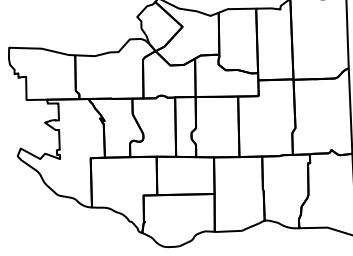
Figure 1: Vancouver British Columbia

neighbour distribution [2]. Additionally, a dispersed iteration of the equation [3] is also produced to aid in comparison. The equation is then standardized [4] for local comparison of different areas. Finally, onto determining significance of the nearest neighbour analysis we can use a Z-test using formula [6] to determine whether the spatial distribution of the pattern is random or not.

$$\overline{NND} = \frac{\Sigma NND}{n}$$

(1)

$$\overline{NND_R} = \frac{1}{2\sqrt{Density}}$$

(2)

$$\overline{NND_D} = \frac{1.07453}{\sqrt{Density}}$$

(3)

$$R = \frac{\overline{NND}}{\overline{NND_R}}$$

(4)

$$\sigma_{\overline{NND}} = \frac{0.26136}{\sqrt{Density}}$$

(5)

$$Z_n = \frac{\overline{NND} - \overline{NND_R}}{simga_{\overline{NND}}}$$

(6)

Quadrat Analysis

Quadrat analysis is based upon counting the number of points within a cell, of a specified size and taking note of how many cells feature that many number of points. Together these variables are used to

produce the mean [7] within the quadrat analysis enabling standardized comparison. This mean divided by the variance [8] produces variance mean ratio (VMR) which can then use a chi square test [9] to determine whether there is clustering to be found within the data set.

For this analysis the quadrats selected were based larger than the typical Canadian city (~100m x ~200m) so quadrat sizes were roughly 300x 300 meters. This totaled with 1600 quadrats or 40 x 40 quadrats were developed for both portions of the study.

$$MEAN = \frac{Number of Points in the Dataset(n)}{Number of Quadrats(m)}$$

(7)

$$VAR = \frac{\sum f_i x_i^2 - [\frac{(\sum f_i x_i)^2}{m}]}{m-1}$$

(8)

$$VMR = \frac{VAR}{MEAN}$$

(9)

$$X^2 = VMR(m-1)$$

(10)

K-Function [11]

This method produces a graphically representation for determining spatial characteristics. Similarly, to Quadrat analysis it utilizes point density, E represents the estimated number of points within the specific distance d, and N is the number of points with the given distance d. This allows us the determine whether a data set is clustered at a specific distance. The values from the k- function can be compared on a graph with [12] or the complete spatial random output to determine the spatial pattern being analyzed.

$$K(d) = \lambda^{-1} E(N_d)$$

(11)

$$K_{CSR}(d) = \pi d^2$$

(12)

Kernel Density Estimation [13]

The kernel density estimation provides a method which presents data in a spatial format for visualization. To do this a specific cell size is determined and the density of points within a circle at each point creating a 'kernel'. Looking at the formula we see that C is the circle with radius r is placed at each p or point. Then to be divided by the area of the circle C.

For this portion of the analysis multiple circle/ kernel sizes were produced at scales of 50, 100, 200, and 500 meters respectively.

$$\hat{\lambda}_p = \frac{no.[S \subset C(p,r)]}{\pi r^2}$$

(13)

3

Analysis Ouputs

Table 1: Nearest Neighbour Analysis Results for Mischief in 2019

NNDd	NNDr	NND	Zscore
200.799	93.4357	57.8053	-0.7295

Table 2: Nearest Neighbour Analysis Results for Break and Enter Other in 2019

NNDd	NNDr	NND	Zscore
280.0543	130.3148	98.9765	-0.4601

Mischief in 2019

Figure 2: Mischief Mapped

Table 3: Quadrat Analysis Results table for Mischief in 2019

Mean	Variance	VMR
2.558	30.059	11.753

Break and Enter of Residential and Other in 2019

Figure 3: Break and Enter of Residential and Other Map

Table 4: Quadrat Analysis Results table for BNE in 2019

Mean	Variance	VMR
1.304	3.243	2.486

Results

Nearest Neighbour Analysis

Based upon the results (tables 1-2) of the nearest neighbour analysis in conjunction with its related Z scores we can reject the null hypothesis that the Mischief and the BNE datasets are spatially random. Rather that both data sets exhibit a high and moderate level of clustering respectively.

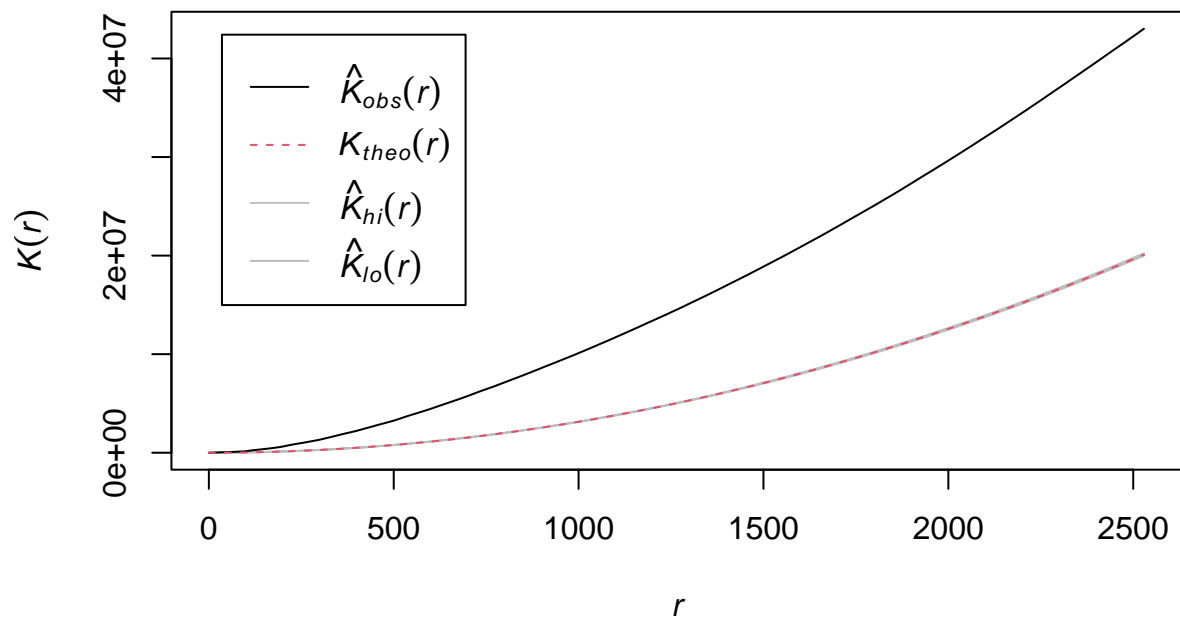# K Function Results for Mischief in 2019

$$K(r)$$

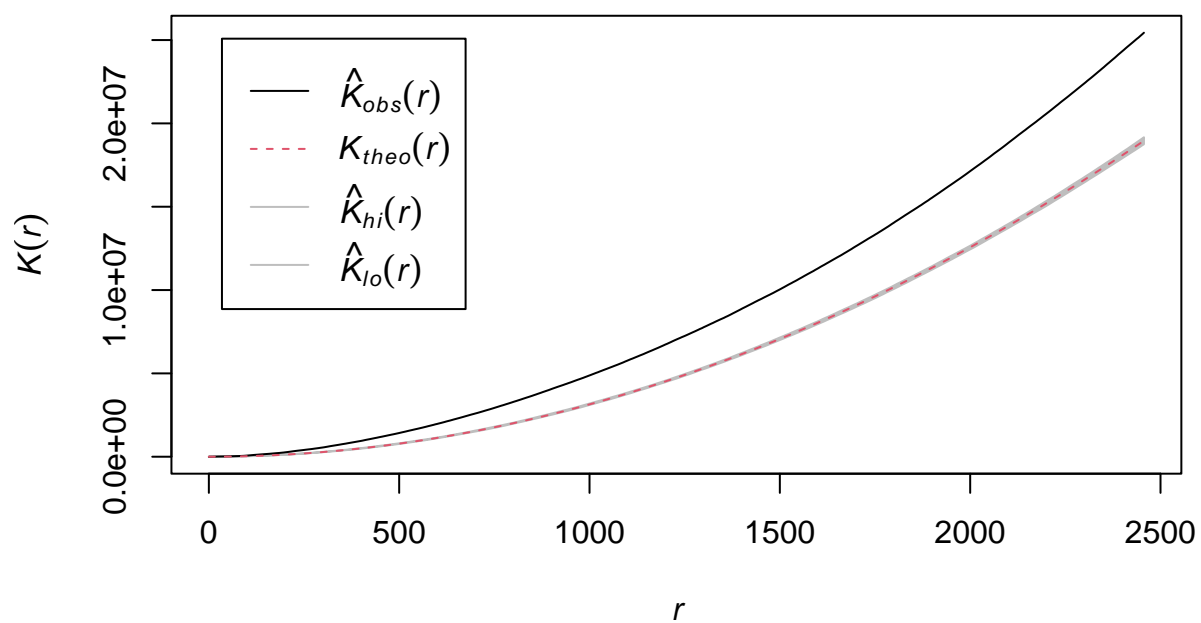Figure 4: K-Function Results for Mischief

K Function Results for BNE in 2019

Figure 5: K-Function Results for Break and Enter of Residential and Other
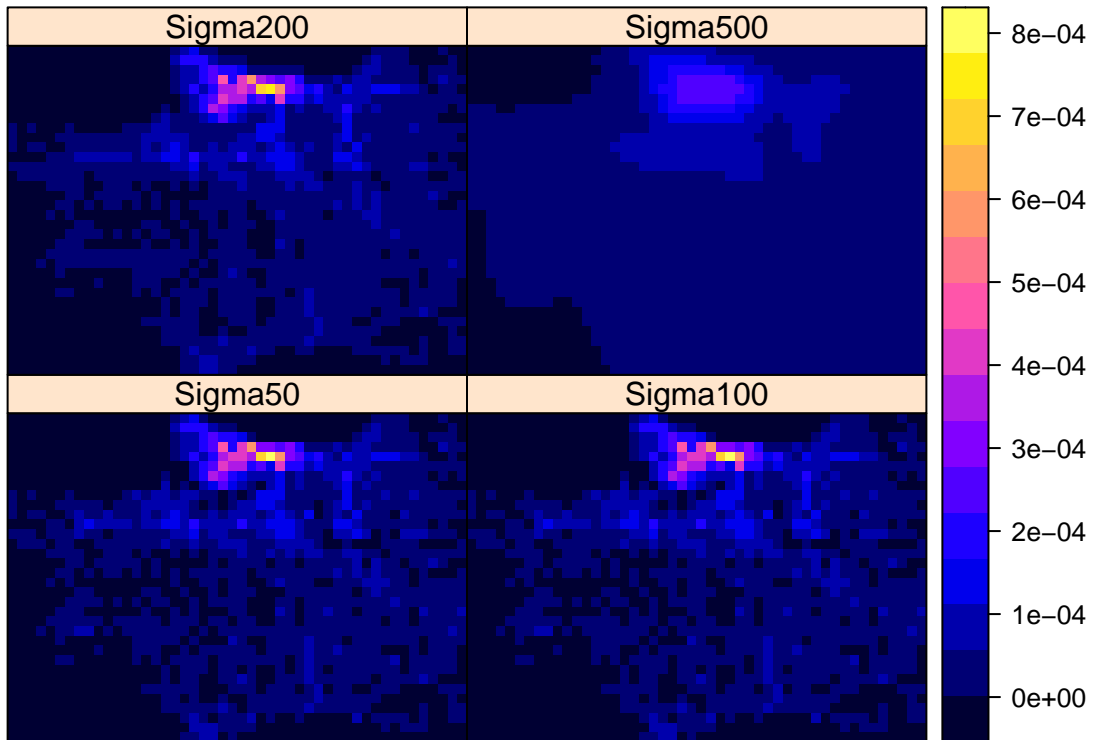
Figure 6: Kernel Density of Mischief at scales: 50, 100, 200, 500 meters
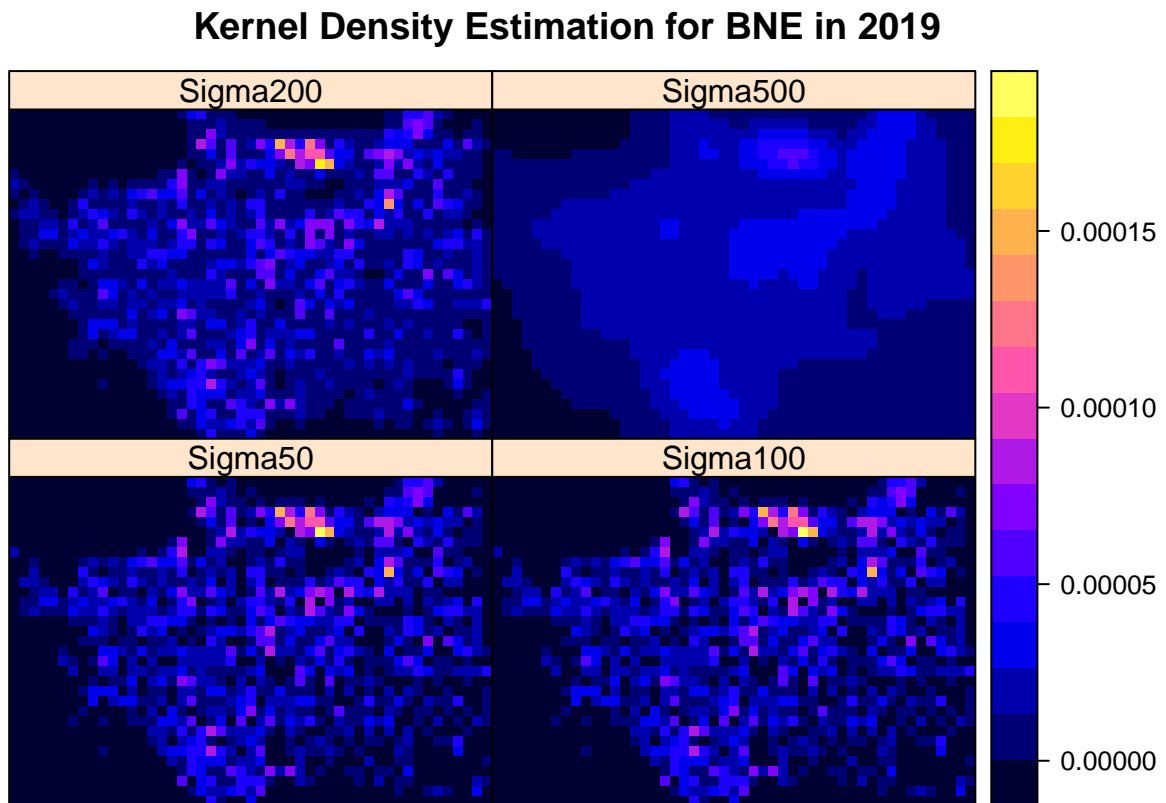
Kernel Density Estimation for BNE in 2019

Figure 7: Kernel Density at scales: 50, 100, 200, 500 meters of Break and Enter of Residential

Mischief Kernel Density Graph

Figure 8: Mischief Kernel Density Graphed

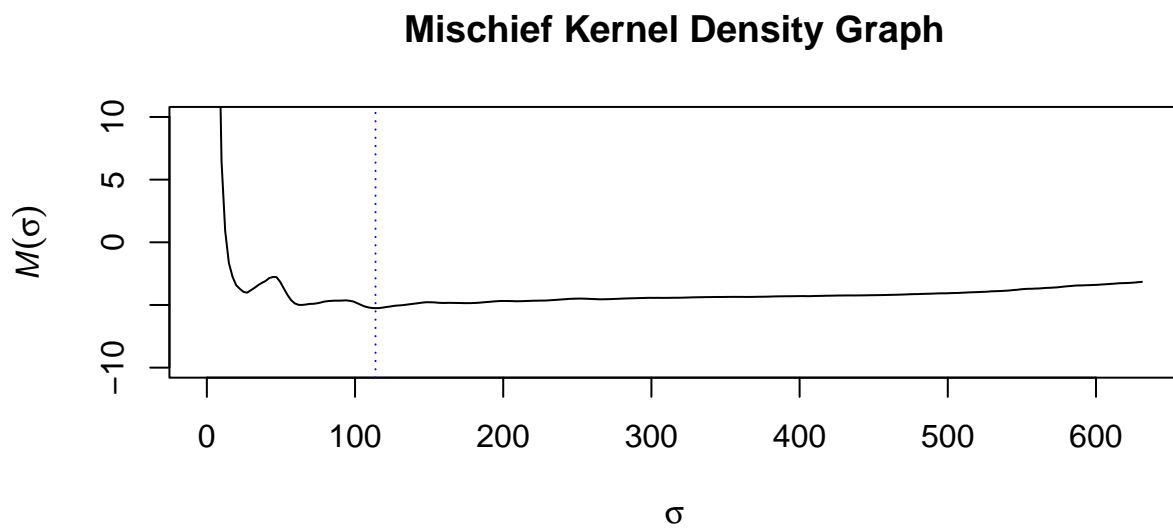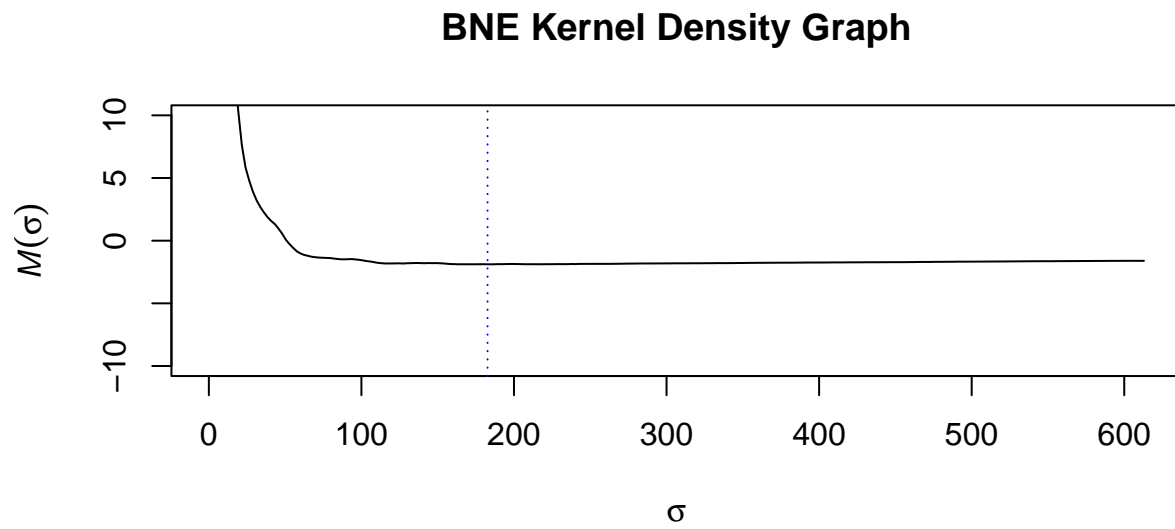## BNE Kernel Density Graph

Figure 9: Break and Enter of Residential and Other Kernel Density Graphed

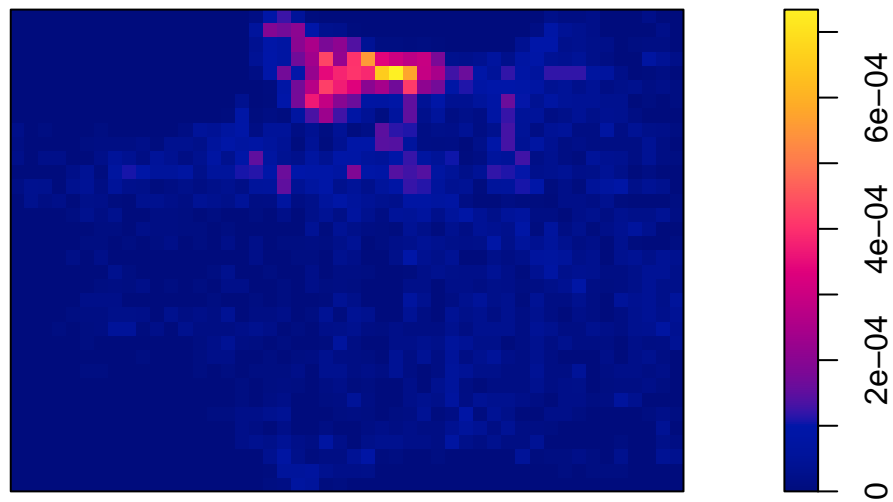## Validated Band for Mischief Kernel Density

Figure 10: Cross Validated Bandwidth Selection of Mischief Kernel Density Graphed

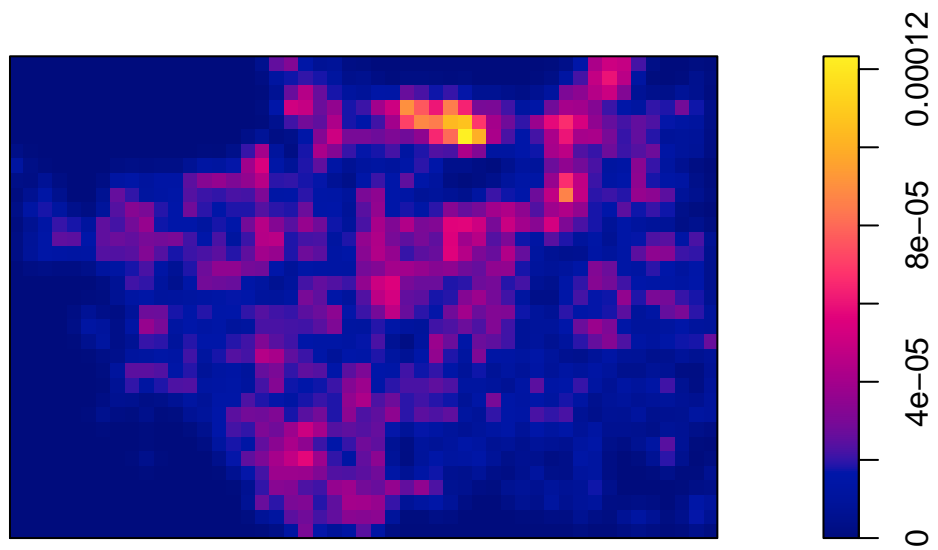**Validated Band Selection for BNE Kernel Density**

Figure 11: Cross Validated Bandwidth Selection for Break and Enter Kernel Density

Regarding both datasets, there are likely multiple factors impacting their spatial correlation beyond just their zero-order process. Contributing factors which could be driving mischief for example could be correlated to proximity of schools, given that underage males are statistically most likely to be charged for criminal mischief (Mahony, 2017). Factors effecting break and enter statistics similarly could be correlated to medium median income neighbourhoods (Bawagan, 2017).

Quadrat Analysis

The results of both quadrat analysis' (tables 3-4) resulted in a p value small enough that R rounded the value to zero, more specifically the pchisq function utilized to produces inaccuracies at extreme values +/-1e^5. Given the negligible P value we can reject the null hypothesis and confirm that the datasets do not have a random distribution.

K Function Analysis

Using the graphs (figure 4-5) from the k-function compared to the K-complete spatial random data we can affirm that both are noticeably clustered due to their curve over the CSR. Similarly, BNE is also clustered but to a lesser extent.

Kernel Density Analysis

Now looking at the KDE results we can see that there are a few locations which feature a greater density of mischief and BNE in the downtown Vancouver, East Vancouver and Grandview-Woodland locations as seen in figures 6 and 7. Looking further into figure 6's 500-meter kernel estimation it appears as though the aforementioned downtown / East Vancouver regions are the only locations exhibiting spatial clustering. BNE however shows further clustering along the Northern side of the Fraser River.

That said taking into consideration the 'optimal' sigma size for the kernel density appears to lie between 100 and 200 meters. Nevertheless, the map outputs all produce similar maps just ranging in intensity at the specified locations.

Further along, other order causes for there to be clustering for mischief and BNE are likely correlated to the clustered regions having a lower socioeconomic demographic at those locations. Additionally, a BNE is also dependent on the dominant location where residential houses and "other" are located, since most neighbourhoods are organized, it is unlikely to see a break and enter of a residence in an industrial or commercial district. Additionally, both locations could feature a higher amount of incidences in their allotted regions due to greater ticketing occurrence or greater police activity in that specific area due to specified patrol routes or law enforcement placement.

Conclusion

To conclude clustering was found for both datasets based on the results of all the statistical tests, as well they were found be to be predominantly clustered in similar downtown / East Vancouver locations, with BNE featuring a slightly greater density along the Northern side of the Fraser River.

The exact causes for these locations and crimes to be clustered cannot be said for certain as multiple factors could be contributing to their placement, such as: socioeconomic demographics at those locations, the overall locations of residences, and the overall law enforcement presence within the area.

Some limitations of this study can be found in production of the initial data, specifically the reduction in accuracy for the preservation of privacy of parties involved in the criminal activity, possible edge effects from the quadrat analysis as well as optimizing cell size/ quadrat count for that analysis could also be considered for error. Finally selecting the optimal kernel size for the KDE analysis could also be considered as a source of error, as the estimations produces both featured deviations between 15- 50 meters in their size and the proposed ideal kernel size.

Further adaptions of this concept could include a variety of other data to delineate what factors may be causing these crimes to occur in the areas that they do. Examples of this data could be census profile average income vectors, school locations, proximity to liquor stores, near by institutions etc....

References

Bawagan, L. (2017). REGRESSION & CORRELATION ANALYSIS OF CRIME IN THE CITY. In Geographical Biogeosciences. http://blogs.ubc.ca/luciampbawagan/files/2017/12/lab3_2017_answersheet.pdf

Mahony, H. (2017). Women and the Criminal Justice System. Statcan.gc.ca. https://www150.statcan.gc.ca/n1/pub/89-503-x/2015001/article/14785-eng.htmStatistics Canada. (2017, February 8).

Census Profile, 2016 Census - Greater Vancouver, Regional district [Census division], British Columbia and British Columbia [Province]. Www12.Statcan.gc.ca. https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page.cfm?Lang=E&Geo1=CD&Code1=5915&Geo2=PR&Code2=59&Data=Count&SearchType=Begins&SearchPR=01&B1=All

Vancouver Police Department. (2020). VPD OPEN DATA. Geodash.vpd.ca. https://geodash.vpd.ca/opendata/