09 April 2019

Analyzing Crime per Capita and Income in Vancouver for 2016 Using Geographic Information Systems

Yassen Atallah, Megan Sheline, Brenna Williams

GEOB 270 Professor Sally Hermansen

The University of British Columbia

ABSTRACT:

Using median income (2015), population (2015), crime (2016), and census tract (2016) data, this report assess the relationship between income and crime per capita within the census tracts of Vancouver, B.C., Canada. The first geographical analysis conducted mapped the different median income levels in the Vancouver census tracts. This map was displayed alongside the second geographical analysis which mapped the number of crimes per capita in the Vancouver census tracts (See Figure A). Two histograms were constructed to represent the frequency and distribution of income and crime per capita respectively (See Figure D and E). The third geographical analysis conducted mapped crime density hotspots in Vancouver (See Figure B). Finally, the fourth geographical analysis was a geographically weighted regression analysis of crime per capita as a function of income to assess the existence and strength of a possible correlation between the two (See Figure C). This report finds that many but not all Vancouver CTs with low income levels have a relatively high rate of crime per capita, that crime is concentrated near downtown Vancouver and follows central roadlines, and that 31.6% of crime per capita in Vancouver can be predicted without error from median income per CT.

TABLE OF CONTENTS:

PROJECT DESCRIPTION:

Vancouver is a coastal seaport city in the lower mainland of British Columbia in western Canada. According to the 2016 census, it has a population of over 600,000, with the region being the third largest metropolitan area in Canada (following Toronto and Montreal) (The Canadian Press 2017), and the city being the fifth most densely populated city in North America (Statistics Canada 2016).

Vancouver is one of the most expensive cities in Canada to afford a house in as housing is severely unaffordable (Connolly 2019). In a city where income makes a big difference in where people can live within Vancouver, one may question in what ways these areas are different. Given that certain areas within Vancouver are more prone to violence than others, could there be a connection between income levels and crime rate in different parts of the city? This project seeks to conduct geographic information system analyses (using ESRI's ArcGIS Desktop 10.6.1) to explore the comparability of median household income and crime per capita across Vancouver, the crime hotspots in Vancouver, and the existence (or lack thereof) of a correlation as well as its strength between the median household income and crime per capita across Vancouver. This information could serve those who seek to move to a particular area of the city to raise a family, or for the police and policymakers to combat crime by focusing their efforts in certain high crime density areas, to name a few. We used population, income, and crime data coupled with Vancouver Census Tracts (CTs) to conduct these analyses.

METHODOLOGY:

A. Acquire:

We acquired our four layers of data from Statistics Canada and City of Vancouver Open Data Catalogue. The 2016 crime point data was obtained from the City of Vancouver in the form of a shapefile (making it easy to import into ArcGIS without any further data preparation). Both the income (per household of 4 people) and population layers were provided by Statistics Canada 2015. The 2016 census tract layer of our map was sourced from Statistics Canada as a shapefile file. This process was done by researching the necessary data within databases, looking for the most current data available. Since the census tract is updated every five years, the 2016 data was selected as it was the most current census taken. For consistency, we looked for data from 2016.

B. Parse Filter:

We began by projecting all of our layers into the UTM 10 projection. Then we modified our CT layer by performing tabular joins with the income and population tabular data. Then we spatially joined the crime points to the new CT layer. We also created a new attribute column defined by the number of crime points in a given CT divided by the population of that CT to obtain crime per capita. As we performed our reclassifications, the out-of-Vancouver CTs disappeared as no

data was available for the analyses to be conducted, effectively resulting in a pseudo-Clip of the new CT layer into the Vancouver boundary shape.

C. Mine:

The first analysis we pursued was reclassifying the income data with a manual breaks classification method to outline CTs that were below the 2016 Canadian poverty line of $44,000 (Statistics Canada 2017). Then we reclassified crime per capita using a manual breaks classification that simplified the classes to facilitate understanding. Afterward, we used the Point Density tool to create a hotspot map of crime density from the crime points, using a manual breaks classification that simplified the classes to facilitate understanding despite being quite close to the default natural break classes. Finally, we conducted a geographically weighted regression analysis using our income data as the independent variable and the crime per capita data as the dependent variable to assess the strength of a correlation between the two variables. Standard residual classification breaks allowed us to observe where the relationship between the two was (relatively) statistically significant, and where the relationship was not statistically significant.

D. Represent:

We decided to create three maps in order to visualise the different analyses we wanted to do with the data that we acquired. This is an important step as data visualisation can have major impacts on how the data is perceived by the audience. For the purposes of our project we created one map displaying the median income per household (of four people) side by side with a map illustrating crime per capita, both using manuals breaks classification (See Figure A). Our data was also represented by a heatmap as our second map (See Figure B), which is a graphical representation of data using a system of colour coding to illustrate the concentration of crimes in particular areas of the city. The third and final map is a geographically weighted regression map using standard deviation classification (See Figure C). This was done as a means to determine whether or not there is a correlation between lower income groups and higher crime rates.

E. Table of dataset:

Table 1: Datasets

Layer / Datafile Name	Source	Uses	Entity/Data Model	Attributes	Modifications
Original: crime_shp_2016 Renamed: Crime	City of Vancouver 2016, .shp	Used to show crime per capita for a comparison with income	Points	ObjectID, Shape, Type, Year, Month, Day, Hour, Minute, H_Block, N_Hood	Point density
Original: mAHTiPb56Ph_data Renamed: Income	Statistics Canada 2015, .dbf	Used to map income per household of 4 people for a comparison with crime	Tabular Data	Income, CTUID	Tabular join
Original: E7lke7gSnxF4_data Renamed: Population	Statistics Canada 2015, .dbf	Used to determine crime per capita	Tabular Data	Population, CTUID	Tabular join
Original: Lpr_000b16a_e Renamed: MVanCT	Statistics Canada 2016, .shp	Used to identify CT boundaries for analyses	Vector Polygon	ObjectID, Shape, CTUID, CTNAME, PRUID, PRNAME, CMAUID, CMANAME, CMATYPE, Shape_length, Shape_area	Tabular join, Spatial join, Clip

DISCUSSION & RESULTS:

Figure A consists of two different maps, one displaying median income while the other displaying crime per capita, both of which within the city of Vancouver. The data we acquired and used for income is for a household of four people. We decided that using median household income data as opposed to using average household income was preferred to realize our objectives as the data would not be skewed by any outliers. The median income map uses a green colour scheme to represent the data. As can be seen in the legend, lower income households are illustrated by the lightest shade of green while the darker green shaded areas are the highest income bracket. The legend for the income map starts at the $44,000 poverty line for Vancouver in 2016 in order to provide a baseline for the general cost of living in the city. With this is mind, the manual breaks classification method presented itself as the best method to visually display the data while also enabling an ease of map legibility and understanding for a wider audience. Using the poverty line value and the highest income value we calculated equally distanced classes in order to formulate our manual breaks. To maintain consistency, the crime per capita map also contains five near-equal interval classes in order to make the data easy to understand. The numbers indicated in the legend represent the number of crimes occur per person in a given CT. For instance, the first bracket suggests that 0.01-0.03 crimes occurred per person.

Comparing these two maps, it is apparent that the areas of lowest median income seem to have the highest number of crime. This is especially noticeable around Vancouver's downtown eastside. The downtown eastside is known to have a high homeless population, meaning that many people living in those CTs live within or well below the poverty line. Not surprisingly, areas of higher income (within the highest two brackets) have the lowest amounts of crime overall. This is not true, however, in the area of Coal Harbour which experienced high levels of crime. Coal harbour has a median income between 92,000 and 116,000 (the second highest bracket). It also sits within the second highest bracket of crime, experiencing 0.10 to 0.21 crimes per person. One potential reason for this is that the population for the Coal Harbour region could be quite small thereby giving the perception of a higher crime rate even if the overall amount of crime is low. These maps, however, do not take into considerations the different types of crime being committed, such as theft, assault, arson, or homicide. Even though certain areas may experience high crime rates, the crimes could be mostly petty crimes such as stealing for sustenance as opposed to more serious crimes such as homicide. Given that crimes are not all the same, focusing on crime frequency may not provide an accurate representation of safety in a given CT.

Two histograms were created to support the maps illustrating income and crime per capita. Figure D displays a graph of the of median income in Vancouver. As expected, the curve is relatively normally distributed, being positively skewed, with either extremes not having a significantly high frequency in terms of income earned. The most common median income per CT earned sits between $64,340 and $80,103. Figure E represents the frequency of crime per capita in Vancouver. This histogram shows how smaller rates of crime are much more frequent among a higher number of census tracts as opposed to higher rates of crime. For instance, the

lowest class in terms of rate of crime per capita (0.012 - 0.033 crimes per person) makes up a little over 50 of the census tracts (about 43% of all Vancouver CTs) , whereas the highest class in terms of rate of crime per capita (0.513 - 0.535 crimes per person) makes up only one census tract (less than 1% of all Vancouver CTs).

Figure B is a heatmap showing the density of crime within Vancouver. While this map uses much of the same data found in the lower half map of Figure A, this map provides a gradient of density not bound by the CT borders nor the unique population values associated with each CT. The use of a heat map for this situation allows for the identification of patterns in the location and range of crime data. Density was determined by the number of crimes in a given area (number of crimes divided by area), with the red colour gradient representing areas of higher crime rates per capita, and the yellow colour gradients representing areas of lower crime rates per capita. White represents the areas lacking data, including lakes, rivers, out-of-boundary zones, and the Pacific Ocean.

As can be seen in Figure B, there appears to be the highest density of crime centered around downtown Vancouver, mainly focused in the eastern side, including Central Vancouver, Gastown, and the Eastside. The eastern side of downtown Vancouver has a lower average income (seen in Figure A), which most likely is a contributing factor toward a higher crime rate. Additionally, City Centre - including areas such as the Pacific Centre shopping mall and an overall plethora of upscale shops - has a high population density and is a hotspot for tourists, which could also result in a higher crime rate. However, as mentioned earlier, the crime data does not distinguish between different types of crimes, therefore making it difficult to pinpoint specific causes as to why these downtown areas have higher crime rates. Generally, the farther an area is located from the city centre, the lower the crime rate. This region may suggest that there is some level of correlation between urbanization level and crime rates. An additional pattern that is predominant throughout the map is that higher crime rates seemingly follow central roads. This is prevalent along Broadway, Hastings, and Kingsway. This reinforces the hypothesis that there may be a correlation between high urbanization level and high crime rates in a given area, as these roads are urban centres teeming with people and stores. In the southern portion of the map mainly dominated by lower crime rate, a few “dots” of medium crime rates can be seen. Compared with a topographical map, many of these spots align with public parks.

Figure C represents the geographically weighted regression map of Vancouver in which crime per capita is a function if median income, using the standard deviation classification method to represent standard residual deviation. Following the analysis, the coefficient of determination obtained was an R-squared of 31.6%. Thus, 31.6% of crime per capita in Vancouver can be predicted without error from median income per CT. The acquired Residual Squares is 0.302 and the acquired AICc is -345, with the Sigma being 0.05. Interestingly, most of Vancouver fits within half a standard residual deviation of the mean of the regression, resulting in most of Vancouver’s CT’s crime per capita rates being explained by the median income values in those CTs. Areas demarcated in dark red – including northern downtown and Gastown – have a relatively high residual greater than 0.19, with areas demarcated in tomato

red having a residual near 0.05, and areas demarcated in salmon red having a residual slightly greater than 0.02. Similarly, areas demarcated in cesious blue – including much of the West End, Coal Harbor, Stanley Park, South Granville, northern Hastings, southern Sunrise, and southern Strathcona – have a relatively low residual lower than -0.025. It is worth noting that no CTs in Vancouver fall within the dark blue to deep blue range in terms of the residual distribution curve, meaning that the distribution curve is skewed to the right. Areas demarcated in a red or blue hue are not very well represented by the crime per capita as a function of income regression model. Therefore, there likely is at least one key explanatory variable missing from the model to best explain the correlation and even the causality of the crime per capita rates in Vancouver, perhaps including populations' education levels or employment rates. In this regard, greater analysis could be conducted to include these and other variables to assess the strength of the correlations.

ERROR & UNCERTAINTY:

Sources of error include the maps' relatively coarse resolution, errors in the classification and generalisation steps, and the age of the data as it may be too old to use to infer knowledge about crime density and income in Vancouver today (although it is not likely that there were significant changes in trends from 2015/2016 compared to present). Great uncertainty results from combining 2015 income and population data with 2016 census tract boundaries and crime data, resulting in inaccurate analyses of maps that do not exactly reflect the real world for a particular year. Nonetheless, these temporally close data sets were used as we encountered difficulty with finding a complete dataset for the 2015 or 2016 years. Also, there is a lack of fine accuracy with regard to the crime points data, which could be a result of geocoding errors. As we treated all crimes the same without taking the different types of crime into calculation for this project, the analyses assumed that all crimes were of the same severity, such as a bike theft compared to a homicide. Furthermore, there is an important discrepancy with using the 2016 Canada-wide poverty line for Vancouver specifically considering that Vancouver is especially unaffordable compared to the rest of Canada. This generalised poverty line statistic was employed as we encountered difficulties with finding vancouver-specific poverty line statistics for 2016.

As with many GIS analyses using data sourced from Statistics Canada, the income data we sourced from Statistics Canada was incomplete due to Statistics Canada's following data suppression rules:

- No characteristic or tabular data can be released for standard areas with a population size under 40 people, and for nonstandard areas with a population of under 100.
- Clients cannot be provided a list of postal codes except postal codes with specified names in their request.
- Quantitative data will be suppressed if below a certain specified threshold (such as income) or if the number of data points is less than 4. Statistics will also be suppressed if an outlier exists that surpasses a given threshold. (Statistics Canada 2015)

Likewise, the crime points data is likely to also be incomplete since it is highly unlikely that all crimes were discovered in a timely manner or were even reported to the police.

Moreover, we recognize the consequences of using certain data classification methods and colour gradients over other ones as a result of personal biases, and how our decisions impact the perceptions of our map viewers. No conflict of interest was present during the execution of this project. Also, any errors made at the initial stages of the analysis would have resulted in propagating errors affecting our later analyses. Lastly, as with all human-based projects and experiments, human error was potentially involved in our analyses.

FURTHER RESEARCH & RECOMMENDATIONS:

Further research into this topic could take on a larger scope and look into the crime rates in Greater Vancouver. This research would take into account areas outside of the city of Vancouver, mapped in this project. For this research, the data from other municipalities (such as Burnaby and Surrey) would need to be obtained from separate sources due to the data not being compiled into a single database and released as open-source data. It should be taken into consideration that much of the land in North Vancouver as well as many areas in Langley and Surrey are made up of mainly forest and agricultural land respectively. This could skew the data because it is likely that crimes per capita would be higher, not necessarily due to an association with low income but rather as a result of low population density due to the isolated nature of these areas.

The specific types of crime committed across the city of Vancouver could be further researched. Such a project would delve deeper into the relationship between crime and income, as it would analyze the patterns between the different types of crime and the nature of the given locations. It would be recommended to sort the types of crimes into broader groupings such as violent crimes, property crimes, and statutory crimes by associating different severity values to each type of crime to make the results much more insightful about safety in a given location as well as to avoid overcrowding the map with excess information. Additionally, it may prove difficult to obtain information regarding the specific crimes in a given location. Therefore, it may even be necessary to provide broad groups of crime types based on severity.

Another area of further research could be to look into the reverse relationships between income and crime which appeared in certain areas in Vancouver. This project would specifically look into the areas where there is high income accompanied by a high rate of crime per capita. If this project were to be executed, it is recommended to find geographical area divisions with a scale smaller than that of the census tracts in order to better evaluate if the crime follows any unique pattern in respect to gathering around central roads, shopping centres, sports arena, or parks for example, and which may support an explanation for this reverse relationship. Additionally, it would be recommended to take the density of the populations within these areas into consideration to assess if a correlation exists.

WORKS CITED:

The Canadian Press. (2017). Population of Metro Vancouver Outpaced National Growth Rate. Retrieved from vancouversun.com/news/local-news/new-census-data-population-of-metropolitan-area-of-vancouver-outpaced-national-growth-rate

Connolly, Joannah. (2019). Vancouver Now Ranked 'Second-least Affordable' Global Housing Market. Retrieved from www.vancourier.com/real-estate/vancouver-now-ranked-second-least-affordable-global-housing-market-1.23606899

Statistics Canada. (2015). Chapter 2 – Confidentiality (Non-disclosure) Rules. Retrieved from www12.statcan.gc.ca/nhs-enm/2011/ref/DQ-QD/guide_2-eng.cfm

Statistics Canada. (2016). Population and Dwelling Count Highlight Tables, 2011 Census. Retrieved from www12.statcan.gc.ca/census-recensement/2011/dp-pd/hlt-fst/pd-pl/Table-Tableau.cfm?LANG=Eng&T=307&SR=1&S=10&O=D

Statistics Canada. (2017). Table 4.2: Low-income Measures Thresholds (LIM-AT and LIM-BT) for Private Households of Canada, 2015. Retrieved from www12.statcan.gc.ca/census-recensement/2016/ref/dict/tab/t4_2-eng.cfm

APPENDICES - Maps and Figures:

Figure A

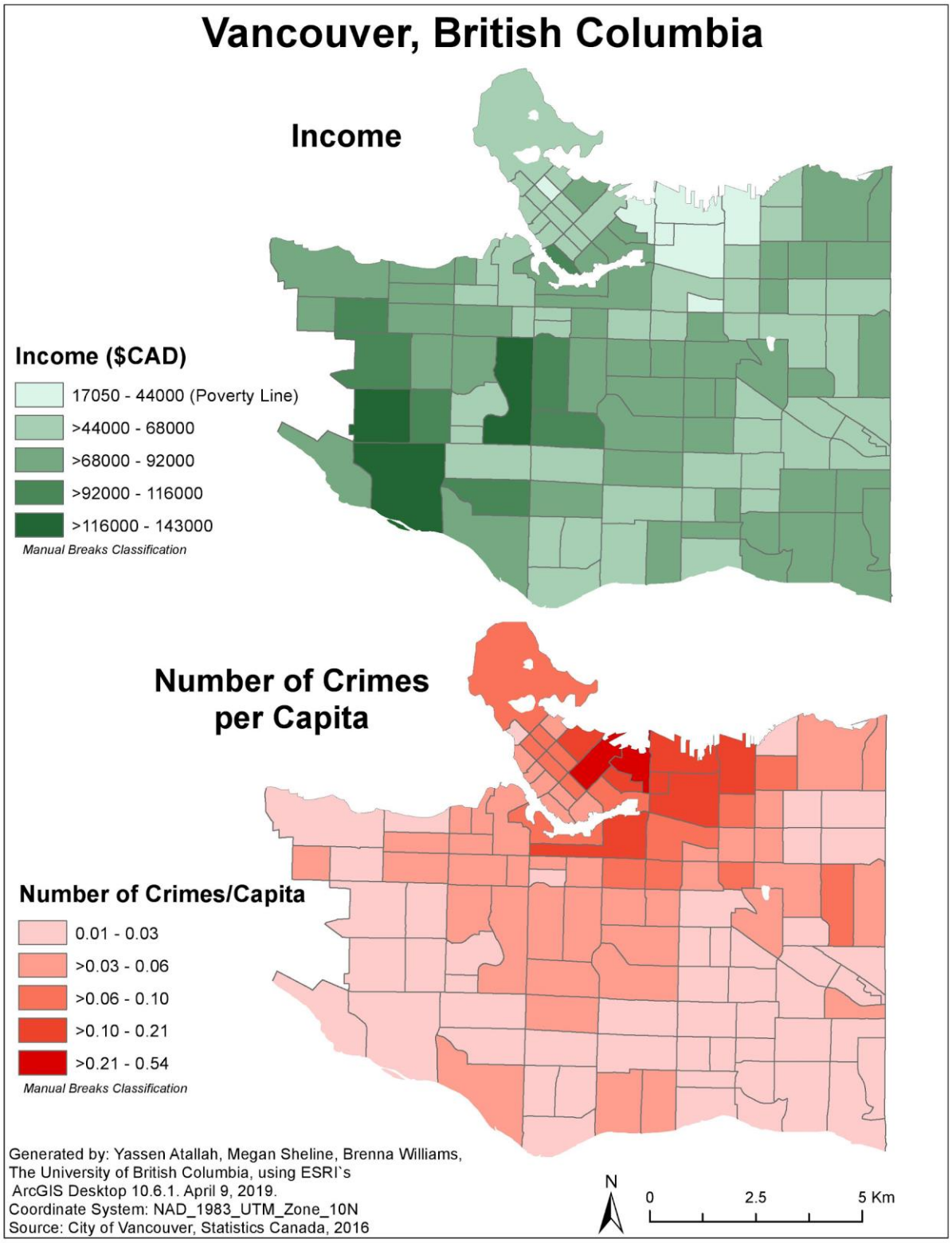

Vancouver, British Columbia
Income
Income ($CAD)
17050 - 44000 (Poverty Line)
>44000 - 68000
>68000 - 92000
>92000 - 116000
>116000 - 143000
Manual Breaks Classification
Number of Crimes per Capita
Number of Crimes/Capita
0.01 - 0.03
>0.03 - 0.06
>0.06 - 0.10
>0.10 - 0.21
>0.21 - 0.54
Manual Breaks Classification
Generated by: Yassen Atallah, Megan Sheline, Brenna Williams,
The University of British Columbia, using ESRI`s
ArcGIS Desktop 10.6.1. April 9, 2019.
Coordinate System: NAD_1983_UTM_Zone_10N
Source: City of Vancouver, Statistics Canada, 2016
N
0
2.5
5 Km

Figure B

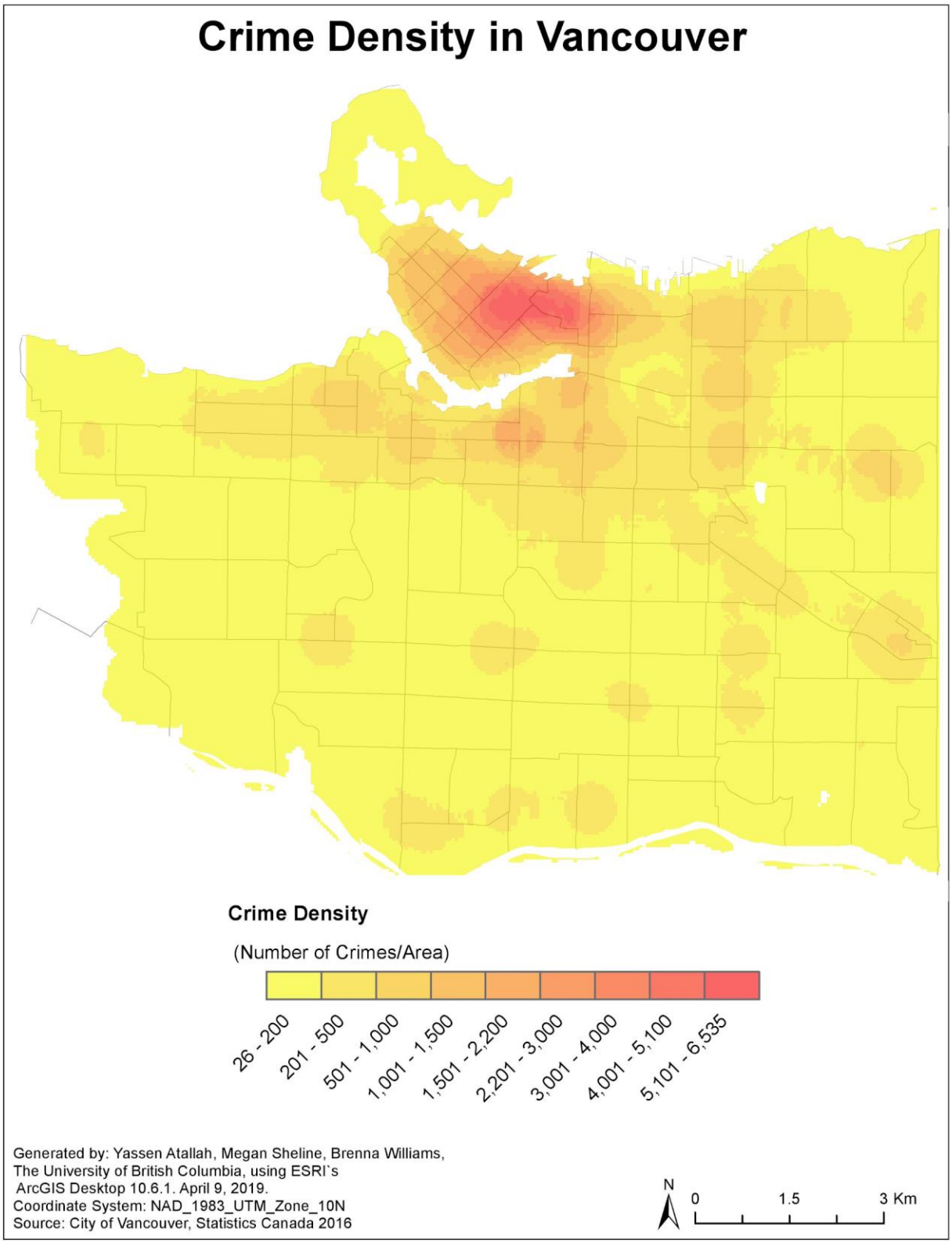

Crime Density in Vancouver
Crime Density
(Number of Crimes/Area)
26 - 200
201 - 500
501 - 1,000
1,001 - 1,500
1,501 - 2,200
2,201 - 3,000
3,001 - 4,000
4,001 - 5,100
5,101 - 6,535
Generated by: Yassen Atallah, Megan Sheline, Brenna Williams,
The University of British Columbia, using ESRI's
ArcGIS Desktop 10.6.1. April 9, 2019.
Coordinate System: NAD_1983_UTM_Zone_10N
Source: City of Vancouver, Statistics Canada 2016
N
0
1.5
3 Km

Figure C

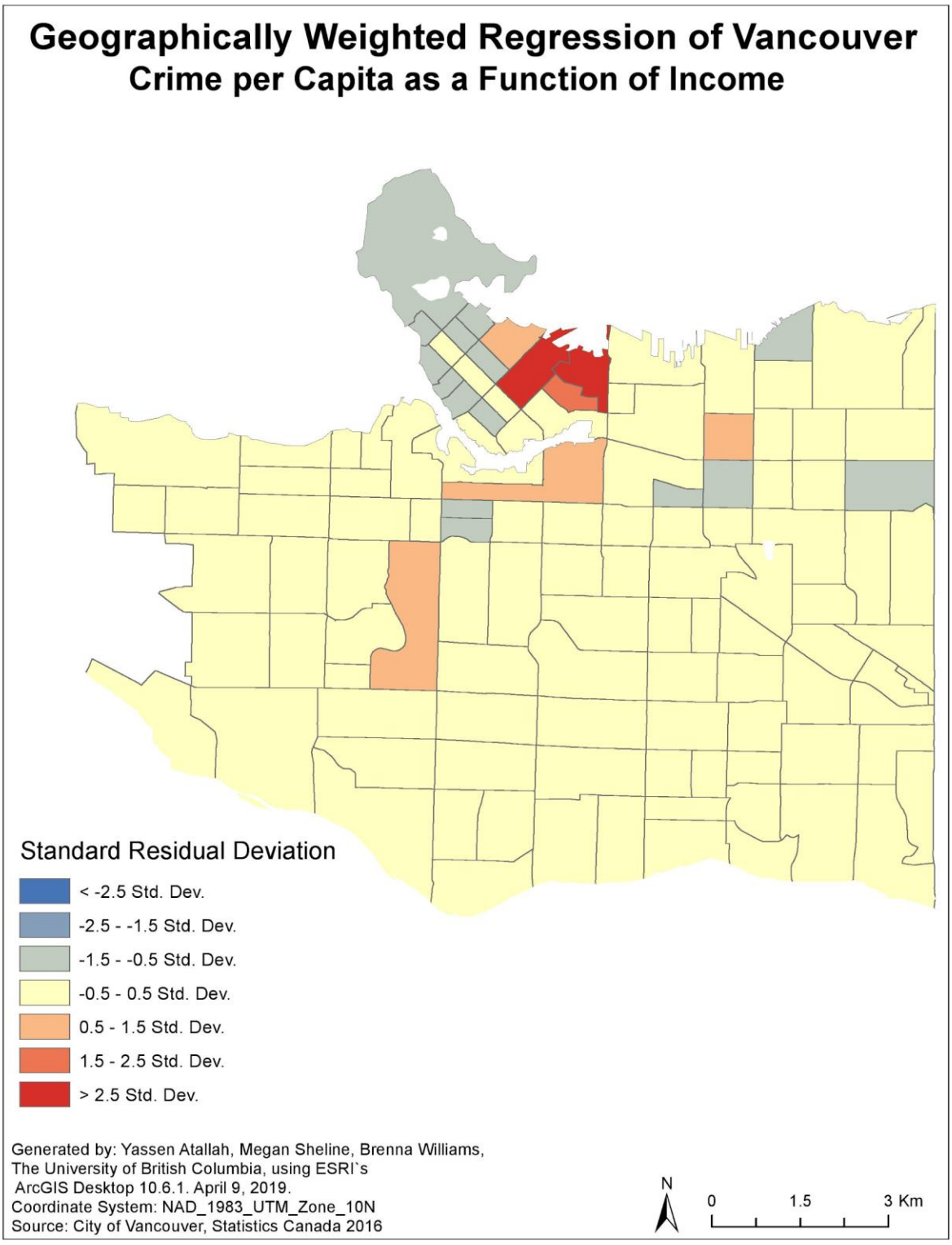

Geographically Weighted Regression of Vancouver
Crime per Capita as a Function of Income
Standard Residual Deviation
< -2.5 Std. Dev.
-2.5 - -1.5 Std. Dev.
-1.5 - -0.5 Std. Dev.
-0.5 - 0.5 Std. Dev.
0.5 - 1.5 Std. Dev.
1.5 - 2.5 Std. Dev.
> 2.5 Std. Dev.
Generated by: Yassen Atallah, Megan Sheline, Brenna Williams,
The University of British Columbia, using ESRI`s
ArcGIS Desktop 10.6.1. April 9, 2019.
Coordinate System: NAD_1983_UTM_Zone_10N
Source: City of Vancouver, Statistics Canada 2016
N
0
1.5
3 Km

Figure D

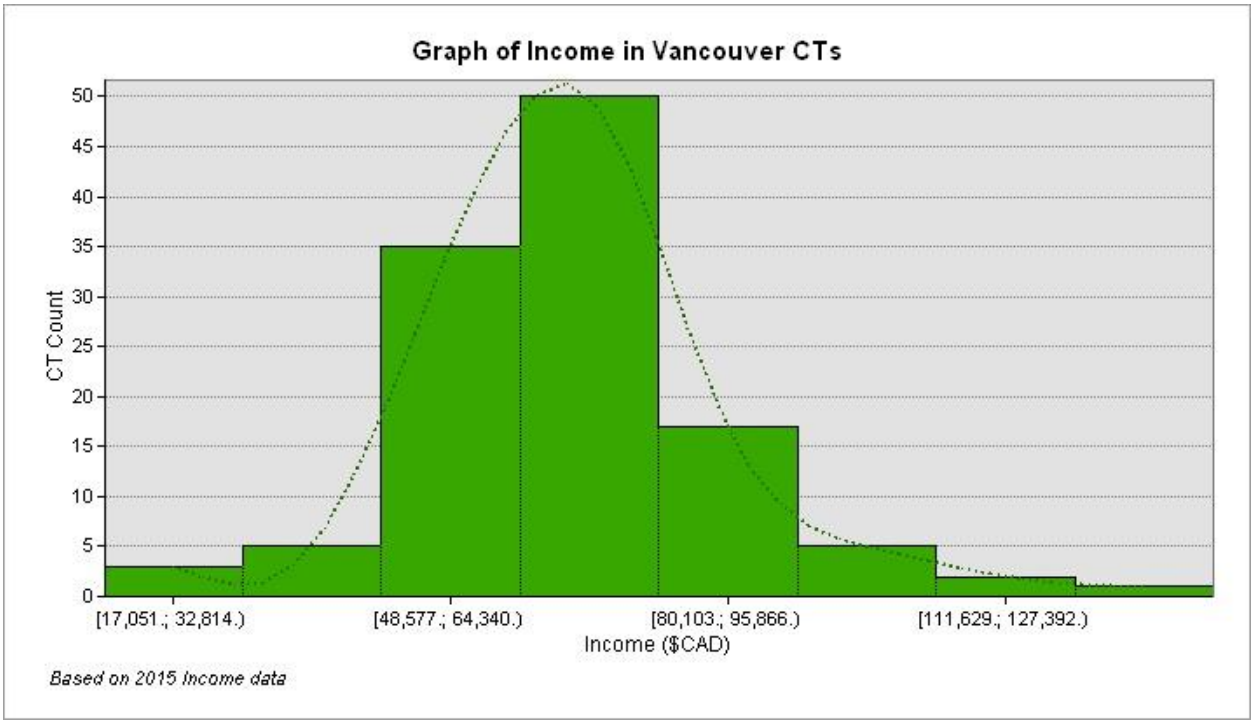
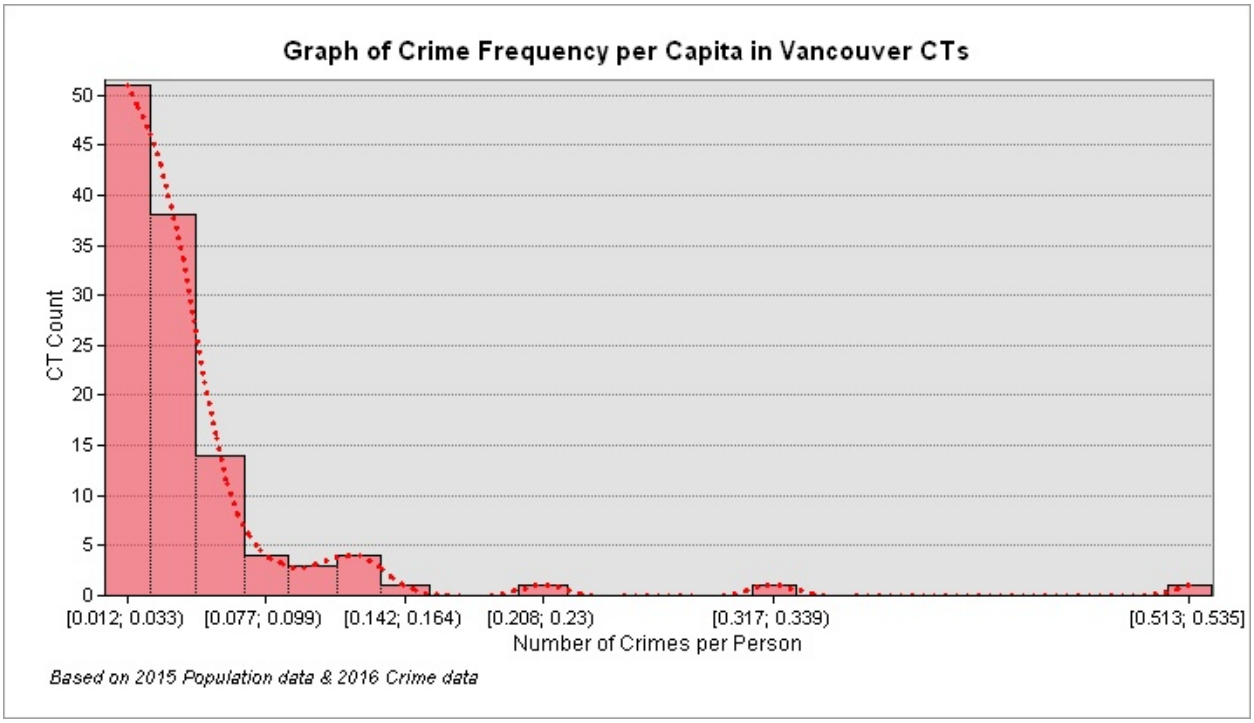


Figure E



Flowchart:

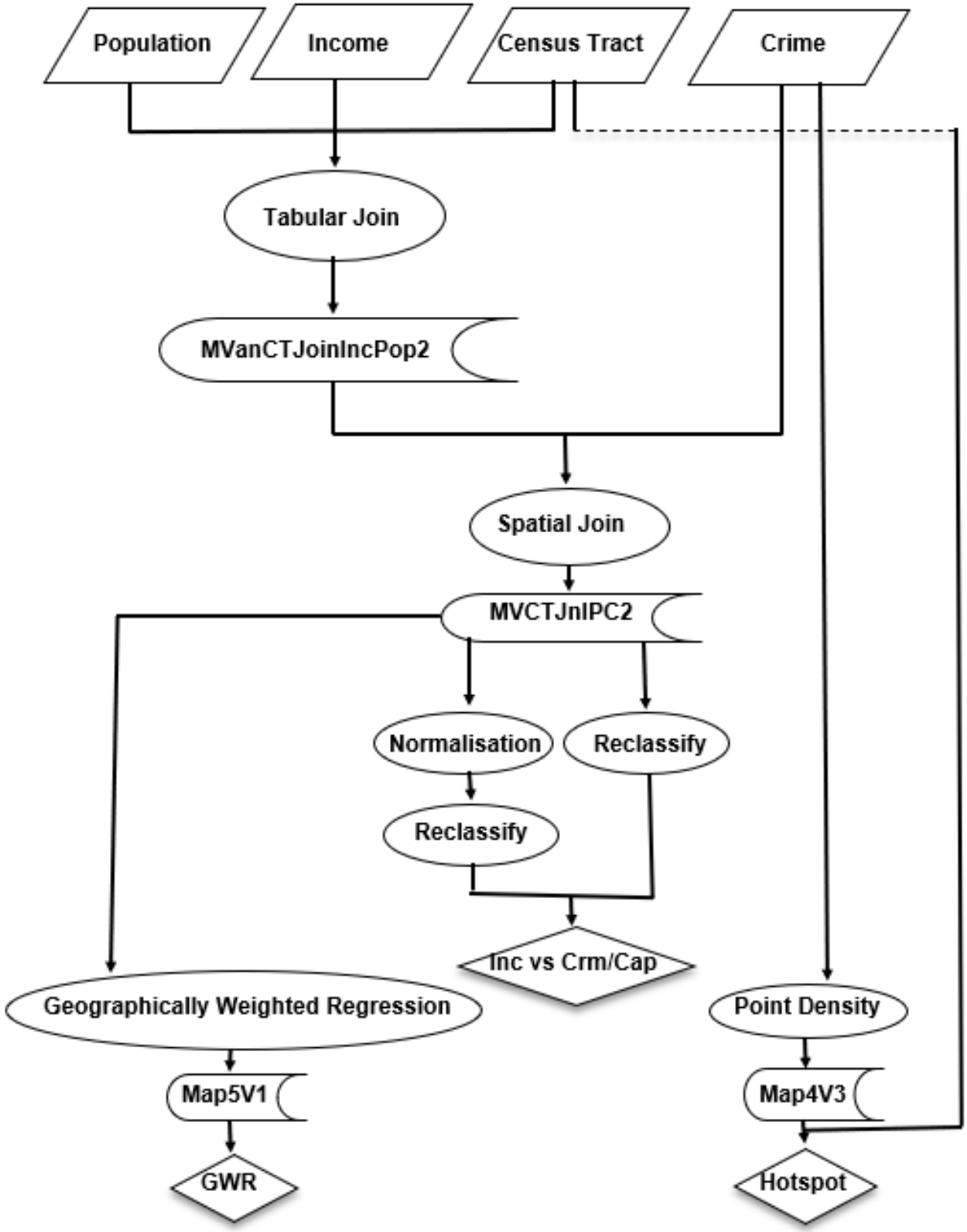

Population
Income
Census Tract
Crime
Tabular Join
MVanCTJoinIncPop2
Spatial Join
MVCTJnlPC2
Normalisation
Reclassify
Reclassify
Inc vs Crm/Cap
Geographically Weighted Regression
Point Density
Map5V1
Map4V3
GWR
Hotspot

Table 1: Datasets

Layer / Datafile Name	Source	Uses	Entity/Data Model	Attributes	Modifications
Original: crime_shp_2016 Renamed: Crime	City of Vancouver 2016, .shp	Used to show crime per capita for a comparison with income	Points	ObjectID, Shape, Type, Year, Month, Day, Hour, Minute, H_Block, N_Hood	Point density
Original: mAHTiPb56Ph_data Renamed: Income	Statistics Canada 2015, .dbf	Used to map income per household of 4 people for a comparison with crime	Tabular Data	Income, CTUID	Tabular join
Original: E7lke7gSnxF4_data Renamed: Population	Statistics Canada 2015, .dbf	Used to determine crime per capita	Tabular Data	Population, CTUID	Tabular join
Original: Lpr_000b16a_e Renamed: MVanCT	Statistics Canada 2016, .shp	Used to identify CT boundaries for analyses	Vector Polygon	ObjectID, Shape, CTUID, CTNAME, PRUID, PRNAME, CMAUID, CMANAME, CMATYPE, Shape_length, Shape_area	Tabular join, Spatial join, Clip