The Relationship Between Social Factors and Crime Rates in Toronto: An Analysis

ENVSOCTY4GA3

Submission Date: 22nd April, 2022

TA: Anastasia Soukhov

Group Members:

Yixin Yang 400226742

Xiaosong Xie 400143076

Ling Cen 400181569

This paper reports our analysis of property crimes in Toronto, Canada, and its relationship with household income, unemployment rates and education level. Data were obtained from the Toronto Police Service Portal and the City of Toronto (Social Development, Finance &Administration).

Introduction

Property crimes are one of the main types of crimes happening in Toronto. According to the official government report on Neighborhood Characteristics and the Distribution of Police-reported Crime in the City of Toronto (Charron, 2009), the distribution of property crimes is more even compared to violent crimes and property crime rates are very close to the average for a large part of the city in Toronto. However, there are still some hot spots that represent clusters of property crimes. "Hot spots are small places in which the occurrence of crime is so frequent that it is highly predictable, at least over a one-year period." according to Sherman (1995). To understand why property crimes concentrate in certain regions, this paper will analyze the spatial statistics of property crimes in Toronto. Taking this as the research direction, we put forward the following questions: Is there any relationship between social factors (household income, unemployment rates and education level) and property crime rates in Toronto? To answer this, the property crime data will go through a series of processes under spatial statistical analysis through R studio in relation to household income, unemployment rates and education level. The effects and importance of these factors will be determined through the investigations.

Background

As one of the largest cities in North America, Toronto is also the most populated city in Canada, with about 3 million people. In such a city with a large size and high population, the crime rate in Toronto is relatively low compared to other cities of similar size in North America (Thompson and Gartner, 2007). However, Toronto's shoplifting rates were higher than Montréal's, another central city in Canada (Charron, 2009). Shoplifting is a type of property crime, and property crime is one of the major types of crime in Toronto.

The relationship between social factors and property crime rates has been the subject of much research since decades ago. Becker (1968) developed a theoretical model of criminal behavior, specific to the role played by a deteriorating labor market. Since then, large amounts of literature have examined the substantial relationship between crime and the economy. According to Janko and Popli (2015), they found some evidence of a significant negative short-run relationship between crime and unemployment for property crimes but no long-run relationship between them. This result is different from Andresen's finding that there is evidence of a long-run relationship between crime and unemployment (Andresen, 2013) and such difference might be caused by using different methodologies to identify the long-run relationship (Janko and Popli, 2015). Another common factor in property crime rates would be household income and it can also be considered as a mediating factor of hate crime (Chongatera, 2013).

Based on data from across Canada, income inequality and average income are instead positively correlated and there is an impact of income inequality on property crime rates (Daly, Wilson and Vasdev, 2001). Related to income, educational level provides an indirect measure of an individual's ability to find skilled employment and salary to match the job (Charron, 2009). It's pointed out in Zajac's study (2012) that committing crimes is less attractive for people with higher education levels as they will have a higher opportunity cost to commit crimes. According to Charron's report (2009), locations with high property crime rates correspond to the city's main shopping centers and most small shopping centers are secondary focal points for property crime in Toronto. There is also a high property crime rate in the neighborhoods around the city center with high density and significant commercial activity. On the contrary, industrial areas or green spaces often have low property crime rates. Charron (2009) also mentioned that half of the residents have a university degree in neighborhoods with low property crime rates, while this percentage drops to one-third in neighborhoods with high property crime rates.

Study Area

Property crime rates were studied at the county level in the city of Toronto as shown in Figure 1. There are 140 social planning neighborhoods in total and all of these counties have available data on property crimes for analysis. Toronto Police reported that the crime rate in 2018 is 3,428 incidents per 100,000 population. It decreased by 17% and 38% compared with Province level and Country level respectively. Toronto is a relatively safe city compared to other cities in Canada. For Property crime specifically, there were 2,282 property crimes per 100,000 population(Statistics Canada, 2020). However, the property crime rate for each neighborhood does have a large variation. The area data was created by generating the average property crime from 2014 to 2020 records (Figure 3)

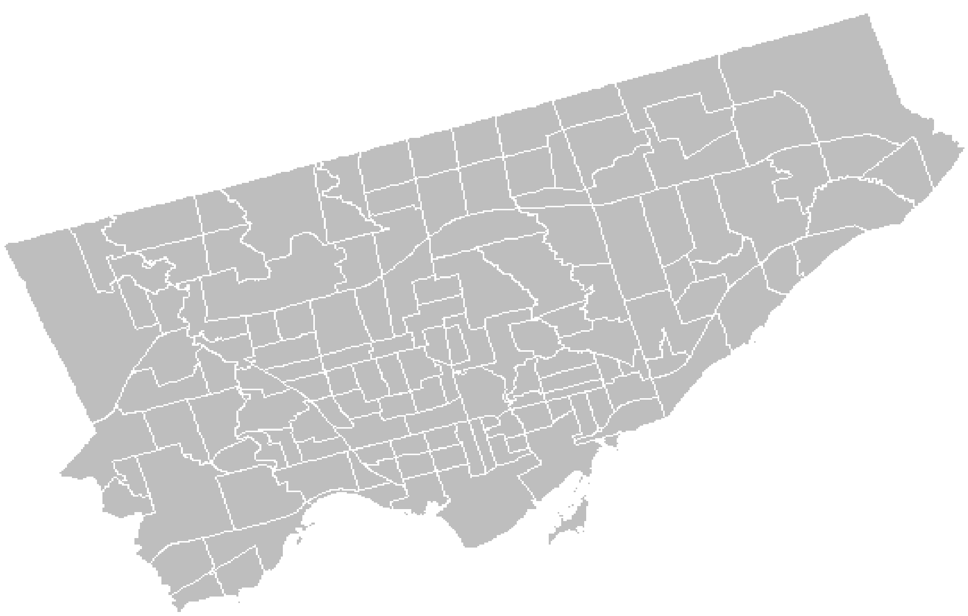

Figure 1: Neighborhood in Toronto, CA

Data

The data used for this project are from the Toronto Police Service Portal and the City of Toronto (Social Development, Finance &Administration). Major Crime Indicators (MCI) from Toronto Police Service Portal show major crime types including Break and Enter, Auto Theft, Robbery and Theft Over from 2014 to 2021. Crime analysis in the later section will use the crime rates from each neighborhood (Toronto Police Service Portal, 2022). In addition, the Wellingbeing Toronto sources the social factors that might contribute to property crimes.

Methods

This study utilizes RStudio to conduct a series of data processes to formulate and illustrate trends from datasets. The distribution of property crime rates will be mapped out with multiple social factors and spatially analyzed to find any correlation. The data will be visualized to investigate the relationship between crime rates and social factors by using regression analysis and choropleth maps driven by RStudio. The different variables used are low-income population rate, unemployment rate and high education population rate. By the end of the project, the study tends to build a better understanding of how strongly each element is correlated to property crime rates and their cumulative effect.

Results

To begin with, the data was collected from Toronto Portal to show the general trends of property crime rates from 2014 to 2020. The property crime rate is calculated by total cases of auto theft, break and enter, robbery and theft over divided by the total population. To notice that the result is showing in units of 100,000 people. The darker color indicates a higher property crime rate in Figure 2.

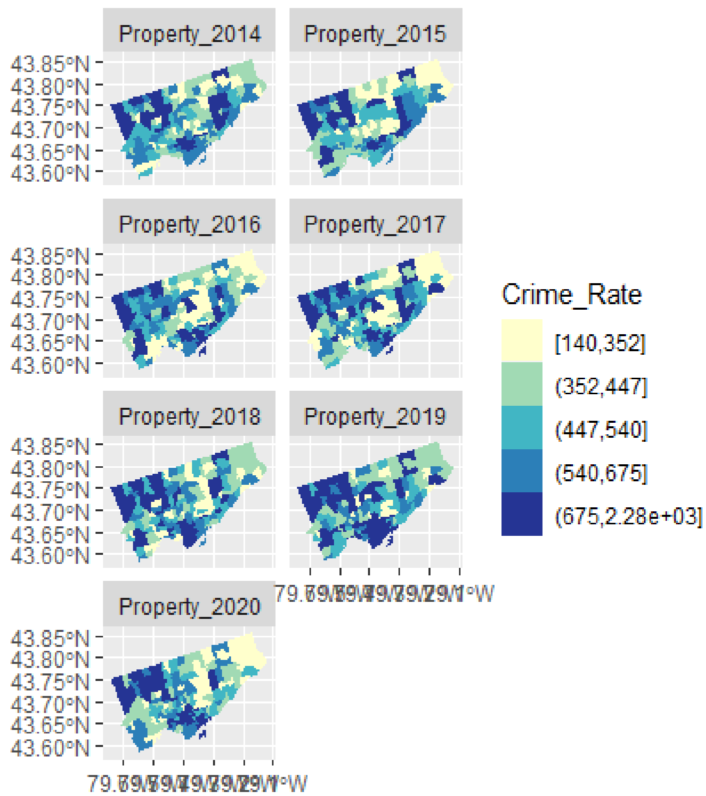

Figure 2: Annual crime rate by neighborhoods in Toronto, CA (2014-2020)

Figure 3 is the map of the average property crime rate from 2014 to 2020 in Toronto. It shows the distribution of crime rates over time.

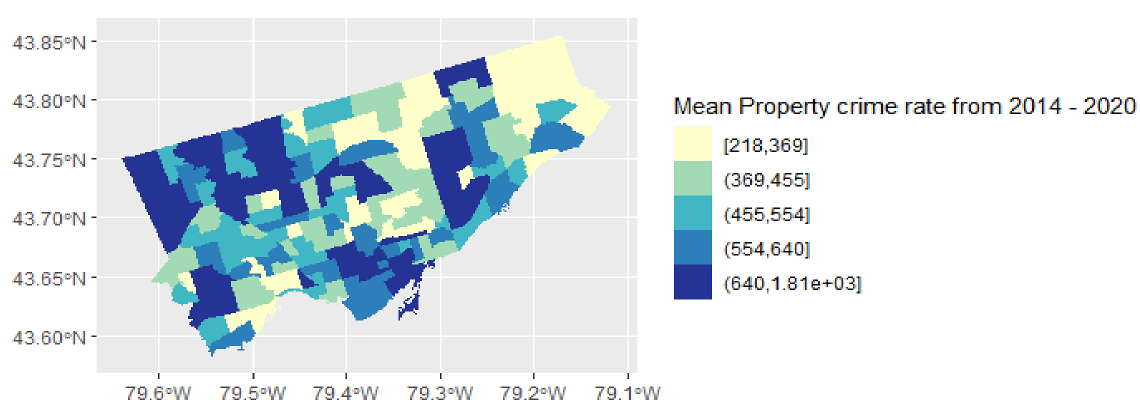

Figure 3: Mean crime rate by neighborhoods in Toronto, CA (7- year average, 2014-2020)

As shown in Figure 4, there is a series of choropleth maps containing the spatial moving average of the target data and six simulated landscapes. These graphs contribute to the further analysis of Moran's scatterplots of spatial moving averages.

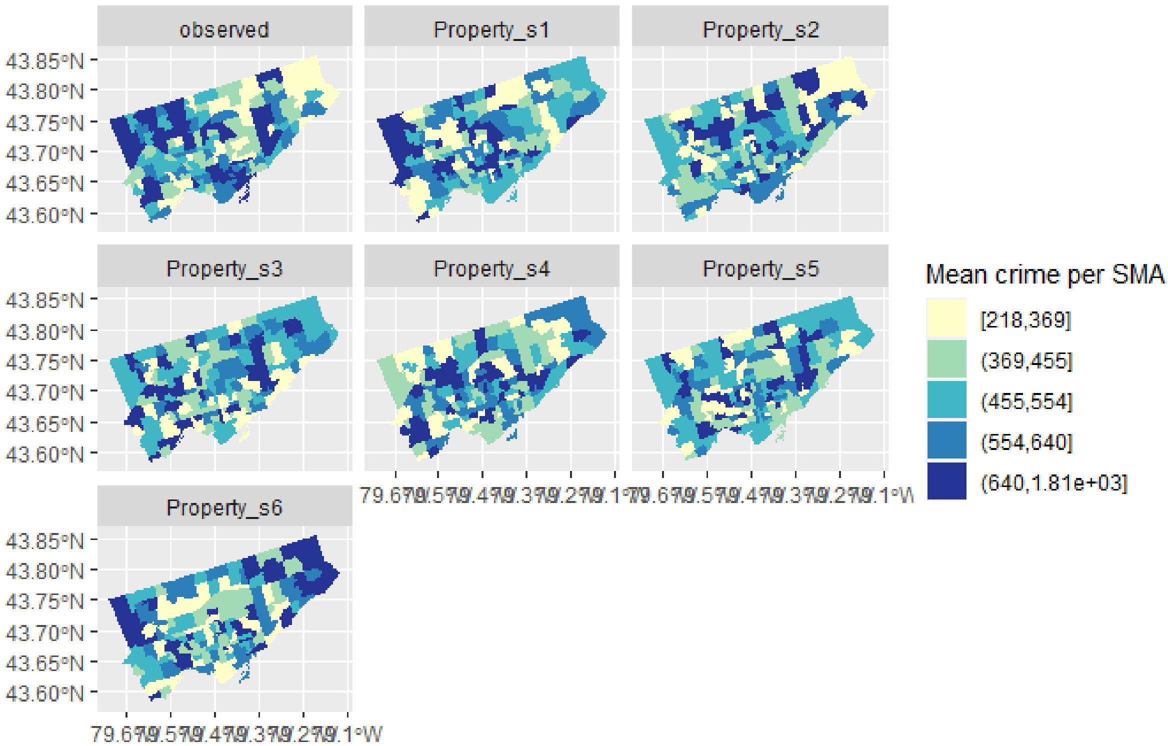

Figure 4: Observed map with 6 simulated landscapes shows the distribution of mean property crime spatial moving average

Figure 5 contains the scatterplots of the empirical average property crime rate and its spatial moving average, as well as the simulated variables and their spatial average. Each scatterplot includes a 45-degree line.

Figure 5: Moran's scatterplot shows the relationship between spatial lag of crime rate and crime rate for observed and 6 simulated landscapes

Figure 6 is Moran's Scatterplot of testing coefficient of spatial autocorrelation for average property crime rate. The Moran I Statistic is 0.243772652 with a p-value equal to 7.052e-08.

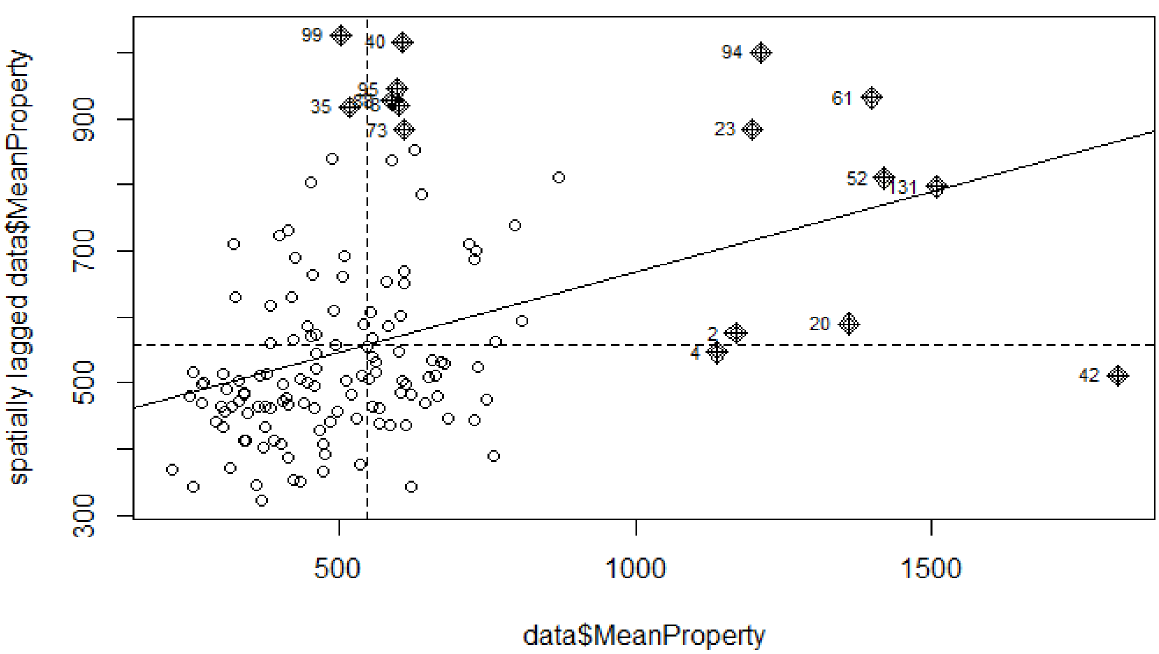

Figure 6:Moran's scatterplot of observed variable

Moran I test under randomization		
data: data$MeanProperty weights: data.w alternative hypothesis: greater		
Moran I statistic standard deviate = 5.2639	p-value = 7.052e-08	
sample estimates:		
Moran I statistic	Expectation	Variance
0.243772652	-0.007194245	0.002273099

Table 1:Moran's I test stats

Figure 7 shows the local statistics of spatial association, which is the local version of Moran' I. According to the previous scatterplot, there are four types of neighborhoods in the plot. One is the neighborhood that has high property crime rates (compared to average property crime rate among all neighborhoods), surrounded by neighborhoods with high property crime rates. This pairing situation is termed "High-High". The other one is that the neighborhood has low property crime rates, surrounded by neighborhoods with low property crime rates. This pairing situation is termed "Low-Low". The other two situations are "Low-High" and "High-Low". Mapping out the local statistics of spatial association helps us to visualize patterns of property crime rates in each neighborhood and its surroundings. It allows us to detect hot spots with high property crime rates, or cold spots with low property crime rates.

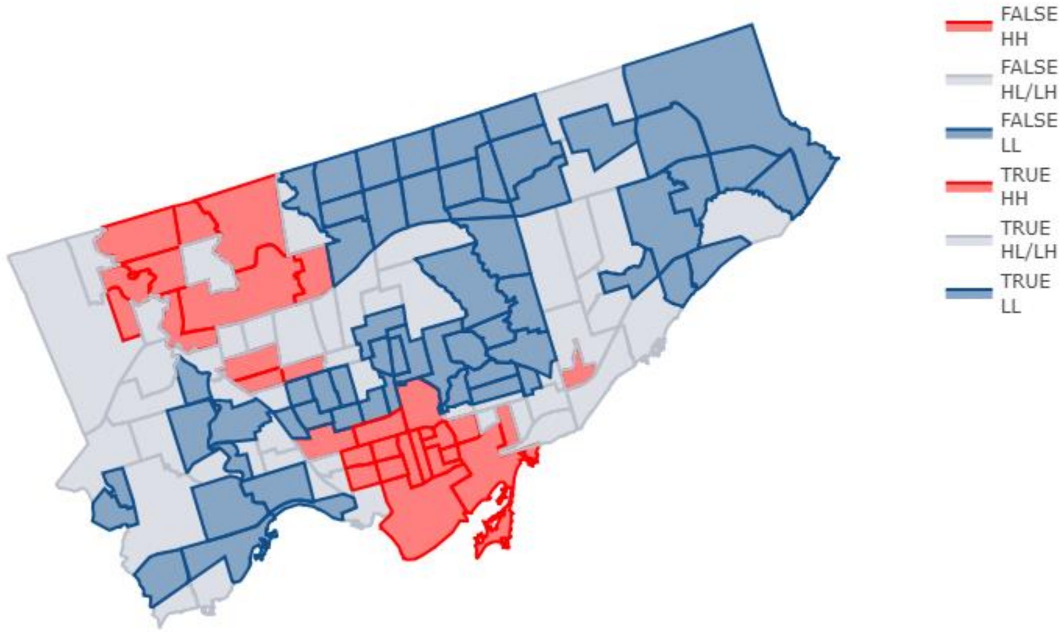

Figure 7: Map of local statistics of spatial association, by local Moran coefficients

Figure 8 represents the low-income population rate in 2016 from 2016 census data which are obtained from Toronto wellbeing (note: the low-income rate is multiplied by 100,000 for comparison). According to Statistics Canada's Low-income measures(LIMs), the low-income population is set at 50% of adjusted median household income (Statistic Canada,2015). The Low-income measure with household size refers to table 11-10-0232-01 (Statistic Canada, 2015). It is calculated by the low-income population divided by the total population for each neighborhood. The darker the area is, the more population with low income are presented. The Figure shows household income trends which are the main factors that contributed to the crime.

Education level is the second factor that is considered the main factor contributing to the crime rate in Toronto. Figure 10 displays each neighborhood rate with College Certificate/Diploma(note: college rate is multiplied by 100,000 for comparison).

The third factor is the unemployment rate. Figure 12 shows the unemployment rate stats for 140 neighborhoods (note: the unemployment rate is multiplied by 100,000 for comparison). Those three maps will be used for comparing the average crime rates.

All of the main factors need to be compared with the average crime rates. The linear regression models are made for the Low-income population rate, college rate and unemployment rate in

Figure 9, Figure 11 and Figure 13 respectively. The following tables are the result of the spatial error model.

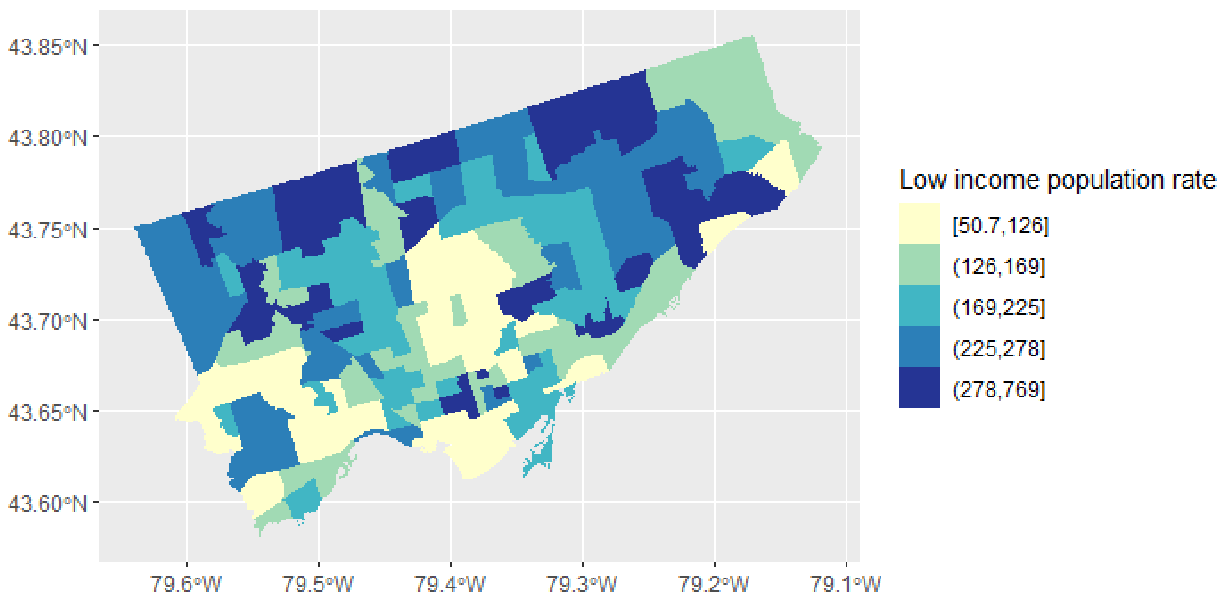

Figure 8: Low-income population rate in 2016

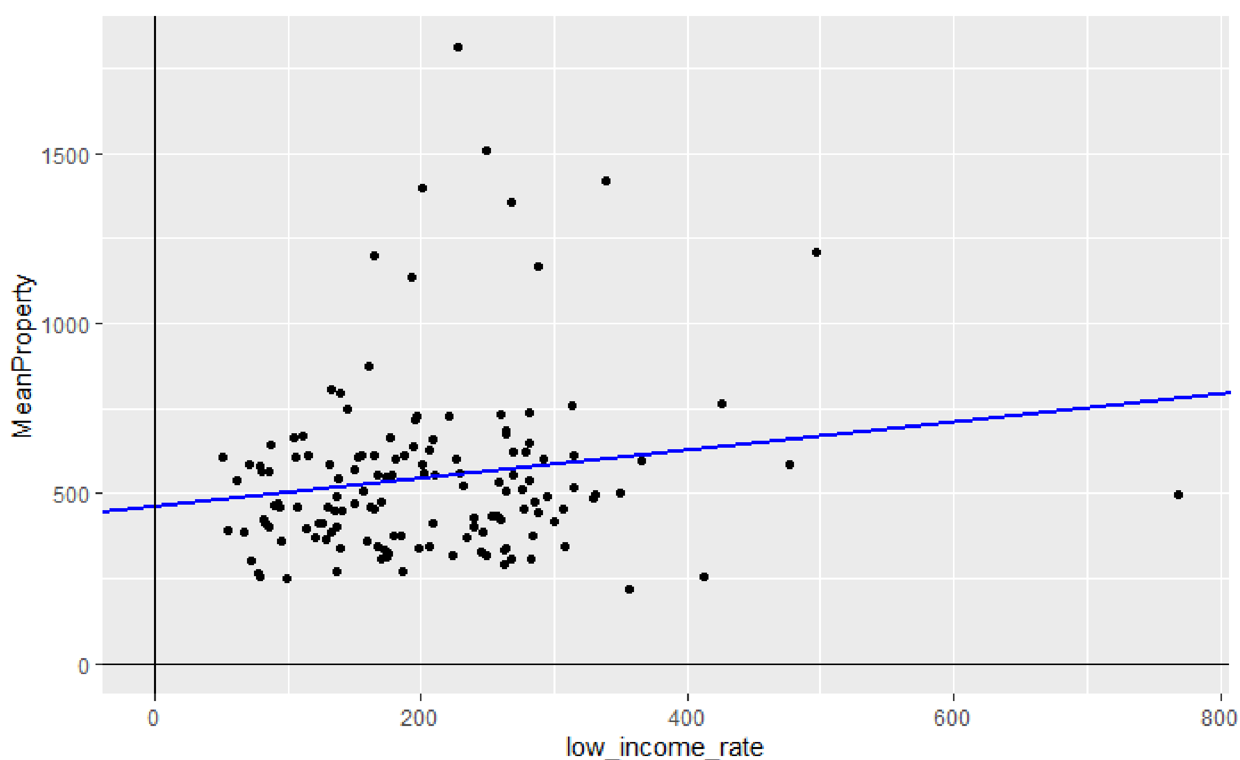

Figure 9: Regression analysis of low income rate and mean property crime

<table>
<tr><th colspan="5">Residuals</th></tr>
<tr><td>Min</td><td>1Q</td><td>Median</td><td>3Q</td><td>Max</td></tr>
<tr><td>-294.487</td><td>-151.027</td><td>-56.385</td><td>92.576</td><td>1281.272</td></tr>
<tr><td colspan="5">Call:errorsarlm(formula = MeanProperty ~ low_income_rate, data = data3, listw = data3.w)

Type: error

Coefficients: (asymptotic standard errors)</td></tr>
<tr><th colspan="5">Estimate Std. Error z value Pr(>|z|)</th></tr>
<tr><td>(Intercept)</td><td>480.43484</td><td>55.81006</td><td>8.6084</td><td><2e-16</td></tr>
<tr><td>low_income_rate</td><td>0.29368</td><td>0.21213</td><td>1.3844</td><td>0.1662</td></tr>
<tr><td>Lambda: 0.41937</td><td colspan="3">LR test value: 14.482</td><td>p-value: 0.00014152</td></tr>
<tr><td colspan="5">Asymptotic standard error: 0.11057</td></tr>
<tr><td colspan="2">z-value: 3.7929</td><td colspan="3">p-value: 0.00014889</td></tr>
<tr><td colspan="2">Wald statistic: 14.386</td><td colspan="3">p-value: 0.00014889</td></tr>
</table>

Log likelihood: -966.3502 for error model ML residual variance (sigma squared): 55988, (sigma: 236.62) Number of observations: 140 Number of parameters estimated: 4 AIC: 1940.7, (AIC for lm: 1953.2)

Table 2: Spatial Error Model of low-income rate and mean property crime rate stats

Moran I test under randomization		
data: model.sem1$residuals weights: data3.w alternative hypothesis: less		
Moran I statistic standard deviate = -0.50056	p-value = 0.3083	
sample estimates:		
Moran I statistic	Expectation	Variance
-0.030960610	-0.007194245	0.002254275

Table 3: Moran’s I test stats for Spatial Error Model for low-income rate analysis

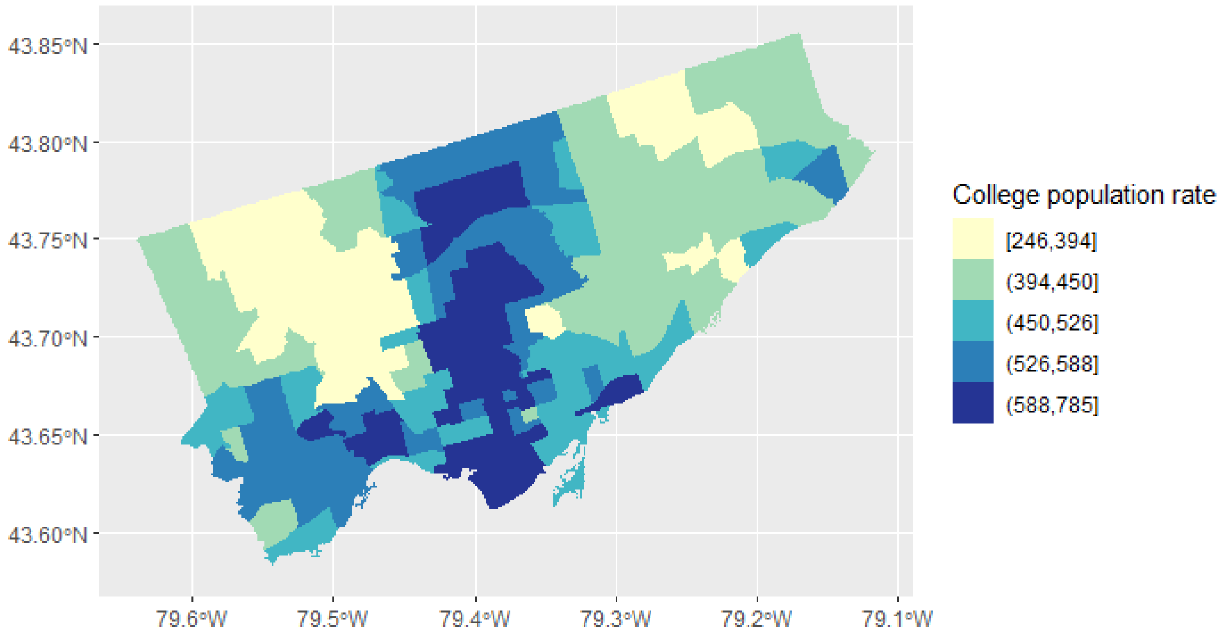

Figure 10: College population rate in 2016

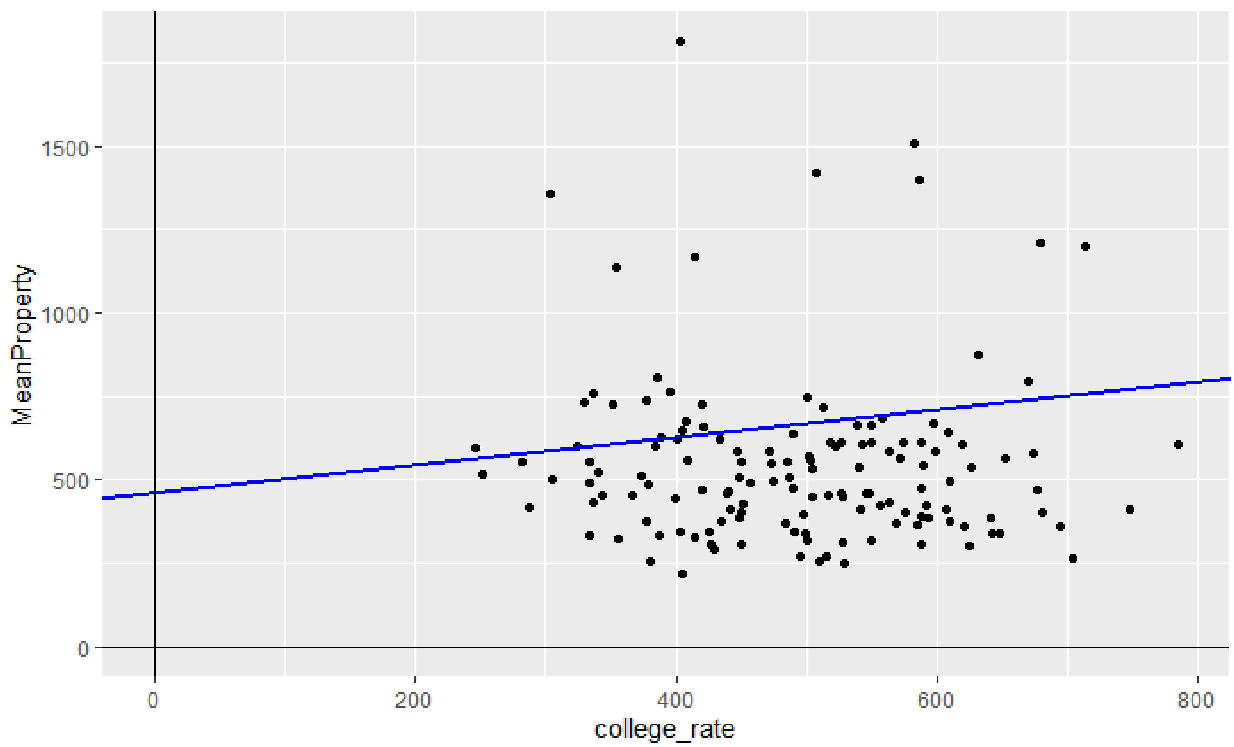

Figure 11: Regression analysis of college rate and mean property crime

<table>
<tr><th colspan="5">Residuals</th></tr>
<tr><td>Min</td><td>1Q</td><td>Median</td><td>3Q</td><td>Max</td></tr>
<tr><td>-299.542</td><td>-145.814</td><td>-60.078</td><td>82.011</td><td>1280.454</td></tr>
<tr><td colspan="5">Call:errorsarlm(formula = MeanProperty ~ college_rate, data = data3, listw = data3.w)

Type: error

Coefficients: (asymptotic standard errors)</td></tr>
<tr><td colspan="5">Estimate Std. Error z value Pr(>|z|)</td></tr>
<tr><td>(Intercept)</td><td>587.420907</td><td>130.177311</td><td>4.5125</td><td>6.408e-06</td></tr>
<tr><td>college_rate</td><td>-0.095036</td><td>0.254384</td><td>-0.3736</td><td>0.7087</td></tr>
<tr><td>Lambda: 0.43894</td><td colspan="3">LR test value: 16.278</td><td>p-value: 5.4708e-05</td></tr>
<tr><td colspan="5">Asymptotic standard error: 0.10849</td></tr>
<tr><td colspan="2">z-value: 4.046</td><td colspan="3">p-value: 5.2102e-05</td></tr>
</table>

Wald statistic: 16.37	p-value: 5.2102e-05
Log likelihood: -967.214 for error model ML residual variance (sigma squared): 56478, (sigma: 237.65) Number of observations: 140 Number of parameters estimated: 4 AIC: 1942.4, (AIC for lm: 1956.7)	

Table 4:Spatial Error Model of college rate and mean property crime rate stats

Moran I test under randomization		
data: model.sem2$residuals weights: data3.w alternative hypothesis: less		
Moran I statistic standard deviate = -0. 51247	p-value = 0. 3042	
sample estimates:		
Moran I statistic	Expectation	Variance
-0.031529227	-0.007194245	0.002254902

Table 5: Moran’s I test stats for Spatial Error Model for college rate analysis

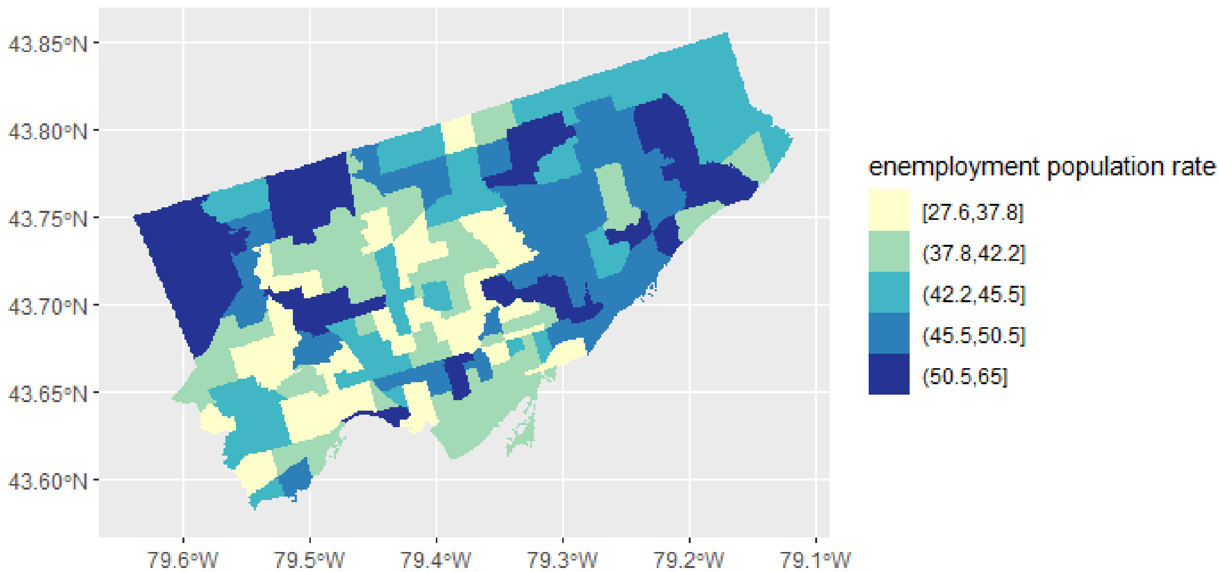

Figure 12: unemployment population rate in 2016

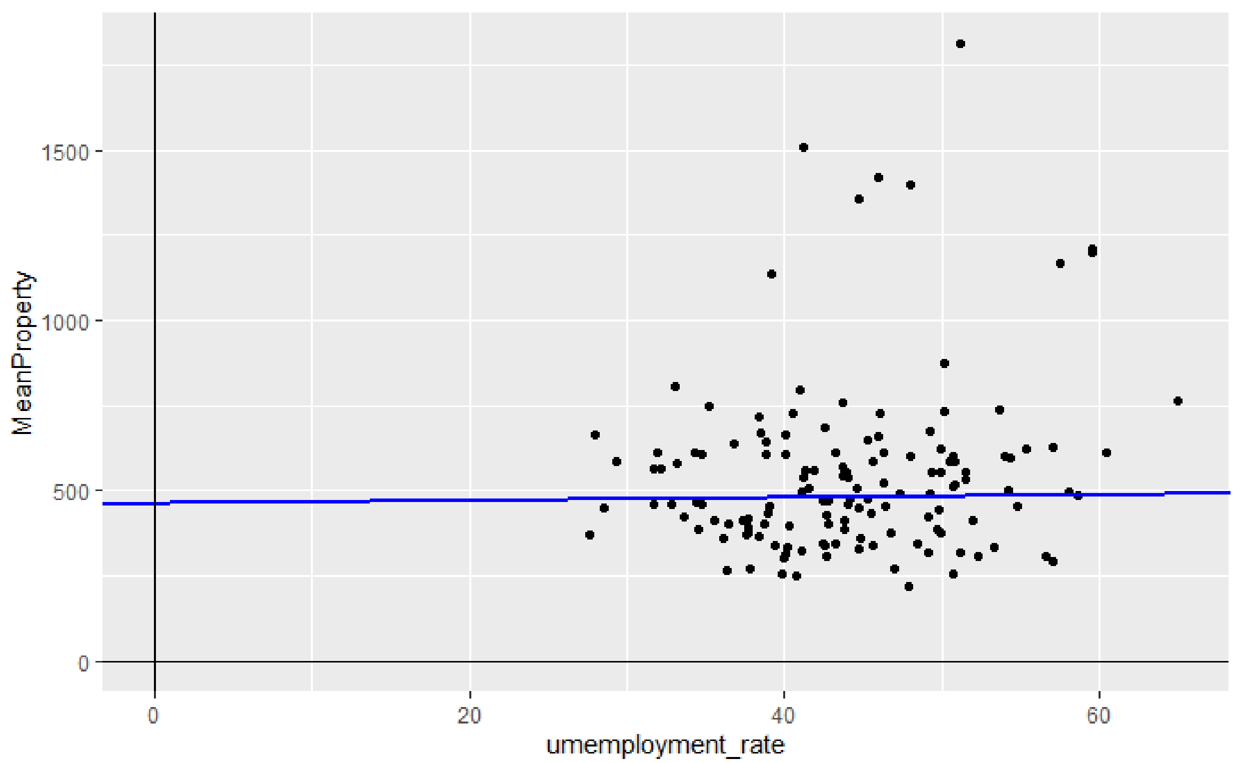

Figure 13: Regression analysis of unemployment rate and mean property crime

Residuals				
Min	1Q	Median	3Q	Max
-280.388	-141.616	-50.228	84.975	1264.815
Call:errorsarlm(formula = MeanProperty ~ umemployment_rate, data = data3, listw = data3.w) Type: error Coefficients: (asymptotic standard errors)				
Estimate Std. Error z value Pr(>|z|)				
(Intercept)	384.3193	134.8086	2.8509	0.00436
umemployment_rate	3.5517	2.9480	1.2048	0.22829
Lambda: 0.41017	LR test value: 13.375	p-value: 0.00025506		
Asymptotic standard error: 0.11152				
z-value: 3.6779	p-value: 0.00023518			
Wald statistic: 13.527	p-value: 0.00023518			

Log likelihood: -966.6016 for error model ML residual variance (sigma squared): 56281, (sigma: 237.24) Number of observations: 140 Number of parameters estimated: 4 AIC: 1941.2, (AIC for lm: 1952.6)

Table 6: Spatial Error Model of employment rate and mean property crime rate stats

Moran I test under randomization		
data: model.sem3$residuals weights: data3.w alternative hypothesis: less		
Moran I statistic standard deviate = -0.51642	p-value = 0.3028	
sample estimates:		
Moran I statistic	Expectation	Variance
-0.031727724	-0.007194245	0.002256915

Table 7: Moran's I test stats for Spatial Error Model for employment rate analysis

Analysis
In Figure 2, we can see that the property crime rates are changing in the counties from 2014 to 2020. Darker colored areas correspond to higher property crime rates. However, the regions with high property crime rates are still concentrated in the northwest and the south of the city over the years, and the darker colors cover more of those areas in 2020 compared to 2014. Another thing to notice is that the city's crime rate in the northeast dropped in 2020 compared to 2014. Figure 3 shows the average crime rates by neighborhoods in Toronto. It also illustrates the areas with the highest average property crime rates over seven years, such as the abovementioned regions.

To identify if the observed map is random, we generate six simulated maps to compare them with the empirical variable. Figure 4 indicates that the empirical variable is not spatially autocorrelated. The high values of spatial moving average tend to be clustered at the northwest corner and lower middle, whereas the low values shown in yellow are clustered in the northeast region. The trend of the empirical variable is decreasing from west to east, and the values in 6 simulated maps are normally distributed.

In Figure 5, we create spatial moving average scatterplots for observed and simulated maps. It is obvious to see significant differences between empirical and simulated variables. First, the average crime rate's spatial lag has many plots above 900 for empirical variables. It rarely occurs in simulated maps. In addition, the slope of the line tends to be flat for simulated variables. However, the slope of the line for the observed variable is close to the 45-degree line. The k-value for the observed scatterplot is higher than the simulated scatterplot. Those differences suggest that the values of the observed variables are not independent of the values of their local means. Therefore, the probability of the observed variables being random is very low.

Figure 6 shows a more detailed Moran's scatterplot of the average property rates per 100,000 population that appeared in Figure 5. Moran's test indicates an extremely low p-value and positive Moran's I coefficient. The statistic's p-value is interpreted as the probability of making a mistake by rejecting the null hypothesis. Thus, we can reject the null hypothesis of randomization with high confidence. There is likely to be a spatial pattern of property crimes in the Toronto area. Some observations are highlighted in the scatterplot, which means they are "impactful" because they contribute a lot to the calculation of Moran's I. These highlighted observations are concentrated in the first quadrant from the scatterplot, meaning they are positive on both axes. This matches our result of the positive coefficient from Moran's test. There are also many observations concentrated in the third quadrant. We next map out the local statistics to understand the spatial pattern in the region.

Figure 7 indicates whether average property crime rates in a neighborhood are high, surrounded by other neighborhoods with average high property crime rates (HH), or low, surrounded by other neighborhoods with average low property crime rates (LL). Other neighborhoods have either low average property crime rates and are surrounded by neighborhoods with high average property crime rates, or vice-versa (HL/LH). According to the map, the "HH" neighborhoods are distributed in the northwest and south of Toronto. The "LL" neighborhoods are in the center and northeast of the city.

Figure 8 shows that the low-income population concentrates in the city's north. A tiny area with a high low-income population is also represented in the downtown area. Only a very small proportion of low-income people live in the Toronto downtown area. In this figure, darker colors indicate areas with higher rates of the low-income population. Compared to the average property crime rate figure (figure 4), the overall distribution of dark color areas is generally the same except for the downtown area.

Figure 9 shows a regression analysis between the low-income rate and the average property crime rate with a linear model. Some noises are observed in the model, represented by the dots far apart from the trend line. The slope of the trend line is positive with many dots above it. According to the analysis, the crime rate increases when the low-income population rate increases.

To justify if the model is appropriate, we use the spatial error model shown in Table 2 and Table 3 to take direct remedial action for residual spatial autocorrelation. Since the covariate is missing, the logarithm of the average property crime rate is used for this model. The result shows that 41% of the moving average contributes to residuals with a p-value of 0.00014152. The Moran's I coefficient is negative in Table 3 proves there are fewer alternatives. All in all, the residuals in this situation are spatially uncorrelated.

From Figure 10, we can see that the distribution of college population rates has a very clear pattern that counties with high college population rates are concentrated in the middle and southwest part of the city, whereas the whole west and east region of Toronto has a generally low college population rate. Additionally, the counties with high college population rates are especially clustered in the downtown area.

Similar to Figure 9, there are many noises in the regression model in Figure 11. Most of the dots are clustering below the trend line. The regression analysis demonstrates a positive relationship between the college rate and the average property crime rate. However, we expect a negative correlation between a highly educated population and property crime. Toronto's high average education level may be one reason that leads to this weak and unexpected correlation. We use the same method for The Spatial Error Model. The model shows that 43% of the moving average contributes to the residual with a p-value of 5.4708e-05 in Table 4. The negative sign of Moran's I coefficient suggests the alternative is less in Table 5. In conclusion, the residuals are spatially uncorrelated.

Figure 12 demonstrates the distribution of the unemployment population rate. Most counties on the north and east sides of Toronto have a high unemployment population rate, whereas the general unemployment rate is low in the downtown area and the southwest part of the city. Only very few counties with high unemployment rates are located in the city's middle and the south part.

The dots distribute evenly around the trend line in the regression model based on Figure 13 with many noises. The flat trend line indicates that the relationship between the unemployment rate and property crime is too weak. The weighted average percentage contributes to residuals is 41% in Table 6 with a p-value of 0.00025506. The negative Moran's I coefficient proves that there are fewer alternatives. To sum up, the residuals are spatially uncorrelated again.

Conclusions

According to our study, the distribution of crime rate is not random, which was revealed by the low p-value in Moran's I test. Furthermore, we find out that neighborhoods with more low-income populations tend to have higher property crime rates. Similarly, the neighborhoods with a highly educated population rate tend to correlate with property crime rates but are much weaker positively than the correlation of the low-income population. In contrast, the correlation between the unemployment rate and property crime rate is too weak to be detected by the trend line from our regression model. This suggests that there is not much of a relationship between these two factors.

The conclusion can be drawn that there is a causal relationship between the low-income population and the crime rate after conducting this series of analyses. Yet, the relationship between education level, unemployment factors and property crime remains unclear. Therefore, we recommend more accurate data within different regions be carried out in future studies to validate the relationships. In addition, this study could be used for comparing the crime rate with other regions in Canada.

Reference

Andresen, M. A. (2013). Unemployment, business cycles, crime, and the Canadian provinces. Journal of Criminal Justice, 41(4), 220-227.

Becker, G. S. (1968). Crime and punishment: An economic approach. In The economic dimensions of crime (pp. 13-68). Palgrave Macmillan, London.

Charron, M. (2009). Neighbourhood Characteristics and the Distribution of Police-reported Crime in the City of Toronto. Ottawa: Statistics Canada.

Chongatera, G. (2013). Hate-crime victimization and fear of hate crime among racially visible people in Canada: The role of income as a mediating factor. Journal of Immigrant & Refugee Studies, 11(1), 44-64.

Daly, M., Wilson, M., & Vasdev, S. (2001). Income inequality and homicide rates in Canada and the United States. Canadian Journal of Criminology, 43(2), 219-236.

Janko, Z., & Popli, G. (2015). Examining the link between crime and unemployment: a time-series analysis for Canada. Applied Economics, 47(37), 4007-4019. doi: 10.1080/00036846.2015.1023942

Major Crime Indicators - Toronto Police Service Portal (2022). Retrieved 12 February 2022, from https://data.torontopolice.on.ca/datasets/major-crime-indicators-1/explore?location=20.413834%2C-40.019598%2C4.62

Sherman, L. W. (1995). Hot spots of crime and criminal careers of places. Crime and place, 4, 35-52.

Statistics Canada. Table 11-10-0232-01 Low income measure (LIM) thresholds by income source and household size DOI: https://doi.org/10.25318/1110023201-eng

Statistics Canada. 2020. Safe Cities profile series: Key indicators by census metropolitan area - Toronto, Ontario.

Thompson, S. K., & Gartner, R. (2007). Effective crime prevention in Toronto, Canada. Enhancing urban safety and security: global report on human settlements 2007.

Zajac, R. (2012). Crime Rates and Economic Factors in Canada (Doctoral dissertation, Acadia University).