

Analysis of factors affecting the social safety of communities in the Greater Toronto Area over the past 6 years (2014-2020)*

A study on social security with the City of Toronto's Open Data(Gelfand 2020)

Xiaoyuan Xing

April 21st, 2022

Abstract

This report provides a comprehensive statistical study of safety in different communities in the Greater Toronto Area to provide a valuable reference for people who want to settle in Toronto. The results show that the population flow and density primarily affect that community's security environment. The analysis included logistic regression using the statistical programming language R(R Core Team 2020). The results help local government agencies and the Public Safety Bureau develop better regulatory measures for less secure communities.

Keywords: social security, Toronto area, communities, population, local government and the Public Safety Bureau, toronto citizens, opendatatoronto(Gelfand 2020)

Contents

1	Introduction	3
2	Data	4
2.1	Dataset	4
2.2	Variables	4
2.3	Missing data	5
2.4	Plots	6
3	Model	9
3.1	Methodology	9
3.2	Multiple Linear Regression	9
4	Result	12

*Code and data are available at: <https://github.com/NATHAN0472/FINAL-PAPER.git>.

5	Discussion	13
5.1	What is done in this paper?	13
5.2	What is something that we learn about the world?	14
5.3	What are some weaknesses of what was done?	15
5.4	What is left to learn or how should we proceed in the future?	15
5.5	Limitations of Analysis	16
6	Conclusion	17
	Reference	18

1 Introduction

In recent years, with the urban growth of the Toronto region, cities are getting bigger and bigger, and the gathering of population and wealth makes urban risks gather. As the city's grassroots communities will face new risks and challenges, urban community safety and risk management deserve in-depth study. Cities are composed of one or more large or small communities, which are the basic units of urban operation. A community is a community of people with some interaction, common cultural cohesion, and interconnectedness in a particular field and its activity area. Differences in the types and characteristics of urban communities make the risks they face differently. Urban community security risk management is the foundation of all urban community governance.

Toronto, Canada, is among the safer areas regarding security, but that does not mean that Toronto is a safe city where more severe crimes still occur. While Toronto's homicide rate has been relatively low for several years, there was still a non-negligible increase here in 2016. This increase directly led to Toronto reaching the highest homicide rate of any major Canadian city in 2018(HuffPost 2018). Homicide is a crime that can directly result in death or injury, which is different from burglary and fraud, where homicide is far more violent. This means that although Toronto is one of the safest cities globally, the city's security management cannot be ignored. Therefore, a multifaceted investigation and study of crime in Toronto are essential.

Community safety refers to the state in which community risks are controlled acceptably. That is the state in which community residents live free from threats, hazards, injuries, losses, and the ability to guarantee a continuous state free from hazards and losses. The existence of safety risks in the community poses a direct threat to community safety. It acts directly on the community, affecting the safety and stability of the community, thus reducing the sense of security, happiness, and access of community residents. Community safety requires low risk in the community, the elimination of management deficiencies, the reduction of unsafe behaviors, and a good infrastructure for community services. For risks in the community that are predictable and early warning, control is in place to ensure community safety; for threats in the community that are not early warning, community safety is relative, and community safety requires both measures to eliminate the likelihood of hazards and to deal with the severity of consequences.

When choosing a community to settle for the foreign population, it is essential to consider its safety, as it is about the comfort of living there. Safety factors play an essential role in the personal safety of the settlers. The main reason for studying this topic is to do a comprehensive statistical study on the safety of different communities in the Greater Toronto Area to provide a valuable reference for people who want to settle in Toronto. The most intuitive information about whether an area is safe and suitable for people is whether it is a high-crime area. If the area is prone to burglaries, robberies, and homicides, this is a sure sign that the area is not safe. Conversely, if an area has very few such crimes, then it is somehow a good place to live. These data and studies can provide a good reference for those who want to come to Toronto.

This paper first presents an overview of the original data set and the variables that are relevant to this study. In addition, it aggregates and analyzes the population size of numerous communities in the Toronto area. Next, the ten communities with the highest and lowest crime rates in the Greater Toronto Area are clearly shown. Then, the results are explained separately for the information related to the applied model. Finally, a more detailed discussion of the extensions of this study (implications, limitations, and future work) is presented.

2 Data

2.1 Dataset

To do a comprehensive statistical study of safety in different neighborhoods in the Greater Toronto Area over the last six years, I used a dataset on Neighbourhood Crime Rates (Portal 2018), which includes crime data by neighborhood. Assault, auto theft, burglary, homicide, gunshot, and firearm discharge data are available, and the data also includes crime rates per 100,000 population calculated using population estimates provided by Environics Analytics. The dataset was collected in 2014 and last updated on May 6, 2021. using R (R Core Team 2020), tidyverse (Wickham et al. 2019), tidyr (Wickham 2021), and dplyr (Wickham et al. 2021), I cleaned and extracted the necessary data to complete an exploratory data analysis and modeling.

2.2 Variables

To begin, I created new variables that will be used in this study, the sum of the number of each type of criminal activity for the period 2014-2020, such as “Assault in total,” “AutoTheft in total,” and “BreakAndEnter in total,” among others. These variables represent the sum of each crime that occurred in Toronto-area communities from 2014-to 2020, and these data will play an essential role in this report. The purpose of these new variables is to examine the social security situation in the Toronto area in recent years and filter out the safest and most dangerous neighborhoods (Portal 2018). These criminal activities (assault, auto theft, break, robbery, theft, homicide, and shooting) are equally dangerous to residents’ lives and property and persecution to the community, even though they have different rates and levels of risk. Second, adding up the number of each criminal activity for the six years was to examine the trend of criminal activities occurring in each community over the six years and thus use the resulting model to predict the crime rate. Therefore, the number of criminal activities each year is not as significant. Finally, a new variable called `Crime_in_total` was created, the sum of all criminal activities in each community over the six years. Unlike the previous new variable, this variable contains the sum of all criminal activities in each community over the last six years. It gives a more visual representation of the number of crime events in each community. These newly created variables, all of which can reflect the crime activity in each community, will play a crucial role in the subsequent work of constructing the model.

In addition to that, I removed variables that would not play a role in this study, such as `ids`, where each community has its unique name and is not confused. Therefore there would be no need for additional numbering. Then, after aggregating the number of crimes for each year, the number of each crime for each year can be removed because this paper wants to study the general trend of criminal activity over these six years. For the treatment of crime rates, the crime rates for all crimes over the last six years are still summed to obtain a sum of crime rates for all criminal activities. This variable can visualize the crime rate of each community during these six years, and this variable can be involved in the construction of the model afterward. Moreover, the public security authorities can use the model to manage communities with more safety hazards more strictly, thus improving the community’s safety conditions. Therefore, the crime rate for each specific criminal activity can also be removed. Otherwise, these variables would be duplicated in the total crime rate. The following table contains all important variables (including newly created ones) in this study.

Table 1: The description of important variables

Variable	Description	Type
Neighbourhood	Name of City of Toronto neighbourhood	Character
F2020_Population_Projection	2020 Population	Number
Assault_in_total	Total count of this crime	Number
AutoTheft_in_total	Total count of Autotheft	Number
BreakAndEnter_in_total	Total count of Break&Enter	Number
Robbery_in_total	Total count of Robbery	Number
TheftOver_in_total	Total count of Theftover	Number
Homicide_in_total	Total count of Homicide	Number
Shooting_in_total	Total count of Shooting	Number
Crime_in_total	Total count of Crime	Number
total_crime_rate	The sum of all crime rates	Number

Finally, all the newly created variables are formed into a new dataset, and we can finish building the model with the new dataset. The purpose of the final model is to predict the crime rate of each neighborhood in the Toronto area, and the level of this crime rate can directly reflect the safety of this neighborhood. If the crime rate is high, then the community is a safety risk, while if the crime rate is low, then the community is safe and an excellent choice to settle in. Many factors influence crime rates, and our main goal is to explore the primary factor that influences crime rates, which means that this factor can determine the crime rate to some extent. For example, the size of the population and the community’s location; if the population of a community is too large, then it means that the community has a high flow of people, which provides the perfect conditions for criminal activity. Of course, these factors are only conjectures in a general objective sense. We need to determine which factors are the main factors influencing the crime rate through the models and analyses derived subsequently.

2.3 Missing data

Missing data, or missing values, occurs when a set of data sets does not store data for certain variables or participants. Missing data can be due to incomplete data entry, human factors, lost files, and many other reasons(Bhandari 2021). In any data set, there is usually some missing data. It is crucial to discuss data that have been purposely excluded or those left behind because of an oversight in the department’s work in question. In some communities, not filing with the police department because of the small amount of property stolen from residents would result in an artificially low incidence of crime(Chenane 2021). This would imply that the community is not entirely safe but rather that some crimes are not known and investigated, a situation that would interfere with the construction of the model. This type of missing data is not reflected in the model, which means that there will be some errors in the results, so these missing data should be discussed and analyzed if the results are convincing.

In addition, those crimes that are intentionally underreported or not reported seriously affect the outcome. The persuasiveness and authority often depend on the data’s completeness and veracity. In the dataset, some communities have had zero crime incidents in recent years, which by common sense, is unrealistic. No matter how good the security conditions are in an area, there will be no more crime incidents in the area, and obviously, this data is not accurate. So I excluded this outlier because I believe it is due to the sloppy data collection. In addition, the lack of data may be because residents no longer choose to report to the police in the event of a burglary or other crime(Agnew 1999). The main reason for this situation is the low efficiency of the police department and the reluctance of residents to trust the police to catch those criminals, which is another reason why some data is forgotten.

2.4 Plots

To better understand the data, I completed a preliminary analysis using `ggplot2` (Wickham 2016). I sorted and organized the ten neighborhoods with the highest number of crime incidents and populated them with the population of each neighborhood. I then sorted the ten cities with the lowest number of crime incidents and again populated them with the population of each community. The primary purpose of doing this is to analyze the relationship between the number of crime incidents in a community and the community's population, and finding the connection between them is the first task in building the model.

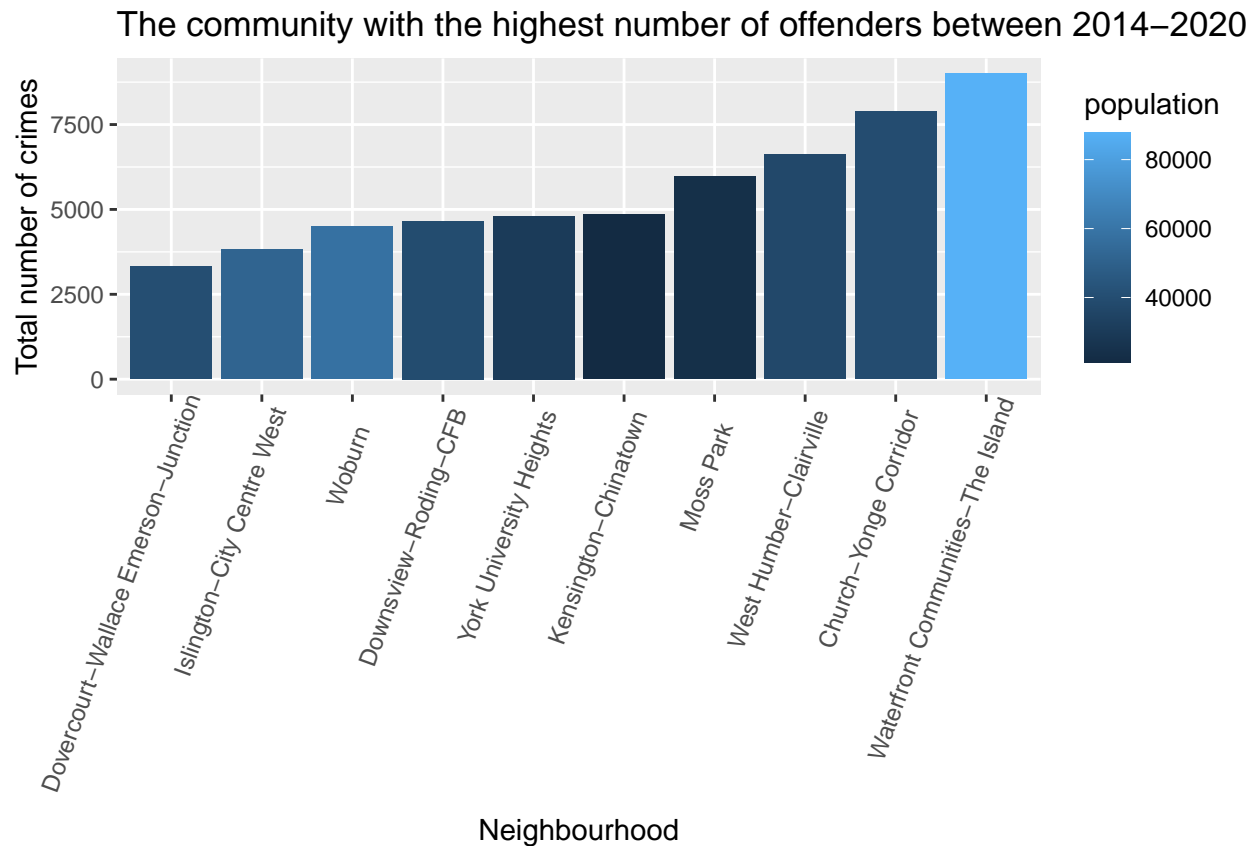


Figure 1: The 10 most crime-ridden neighbourhoods in Toronto and their populations

It is easy to see from Figure 1 that there is some relationship between the population of a community and the number of crime incidents that occur. The waterfront communities, where crime occurs frequently, are also the most populous of the ten communities. However, Kensington, the least populated of the ten communities, has the fourth-highest number of crime incidents, which indicates that the crime rate in this area is very high and that the security management measures of the relevant authorities in this area need to be improved. Residents living in this area should also always be aware of the safety of their living environment and ensure that there are no factors that threaten their safety.

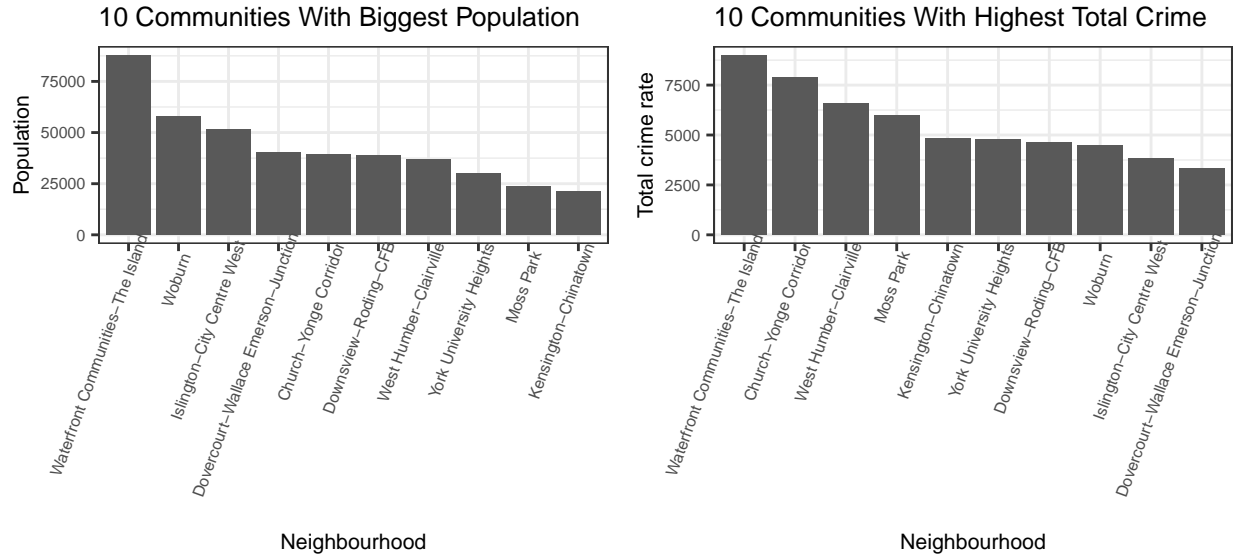


Figure 2: Number of crimes in the 10 most populated neighborhoods

As can be learned from Figure 2, there is a correlation between the population of a community and the probability of having a crime incident. For example, the population of waterfront communities is the largest of all communities, with nearly 80,000 people, so the city has the highest number of crimes in the city, with nearly 8,000 crimes over the past six years. We can use this data to estimate that this community has experienced three to four crimes per day over the past six years. This data reflects a very frightening situation, and we cannot imagine what kind of security situation the residents in this community face every day. The relationship between the community's population and the number of crime incidents has been further illustrated in this chart.

Figure 3 shows the top 10 communities with the lowest number of crime incidents between 2014-2020, based on the population of residents in the community. The graph visually shows that the least populated neighborhoods also have the lowest number of crimes, particularly Lambton baby point, which has a low crime count of less than 400 over the six-year period, which is a small number compared to other neighborhoods in Toronto. In addition, the most populous of the ten neighborhoods, Mount Pleasant East, had just over 600 crimes over six years, reflecting the relatively low incidence of crime in some of the less populous neighborhoods. From these two graphs, it can be inferred that there is a relationship between the population of a community and the number of crime incidents. We will then analyze and infer whether a specific model can describe this relationship.

It is clear from Figure 4 that Lambton Baby Point is the second smallest community in terms of population, with almost the same population as Woodbine, the smallest community in terms of population, and that Lambton Baby Point also has the lowest number of crime incidents over the past six years, at just under 400. The Lambton Baby Point community has had an average of one crime per week over the past six years. This is in stark contrast to the average of four crimes per day in Waterfront communities. The probability of being troubled by a crime in Waterfront communities is approximately 28 times greater than in Lambton Baby Point. Living in a less populated community appears to be relatively safe.

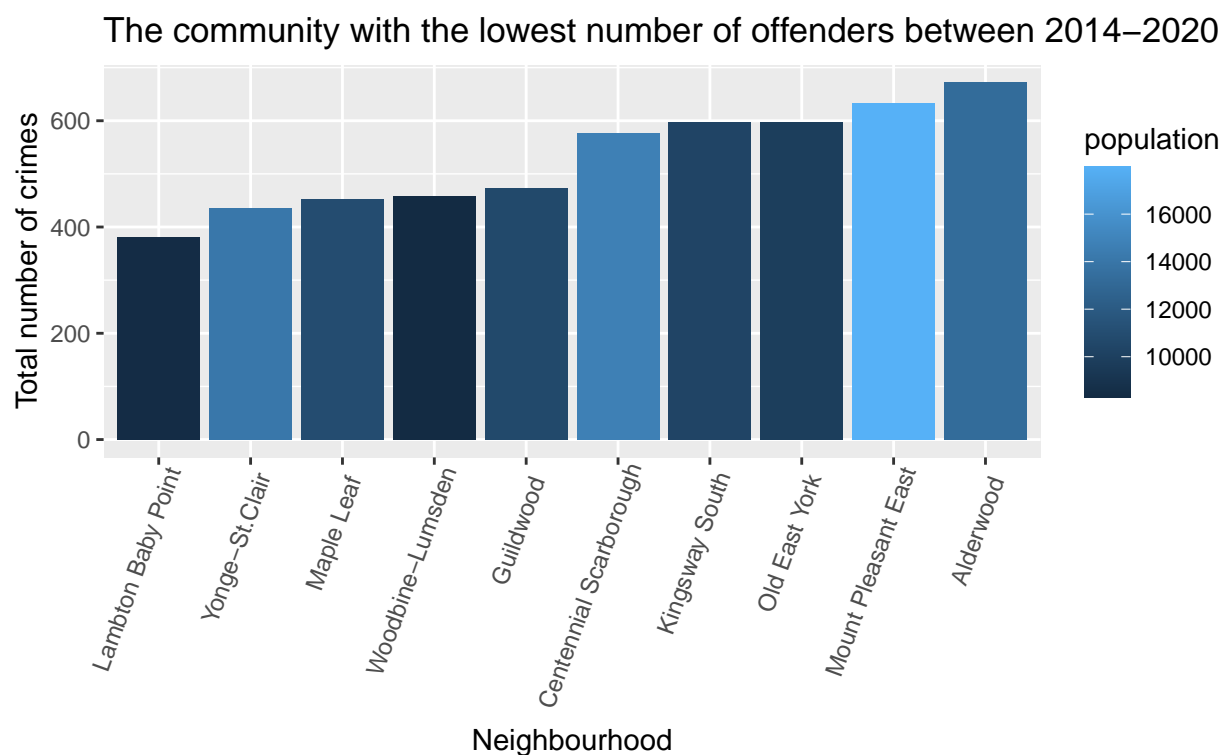


Figure 3: The 10 least crime-ridden neighbourhoods in Toronto and their populations

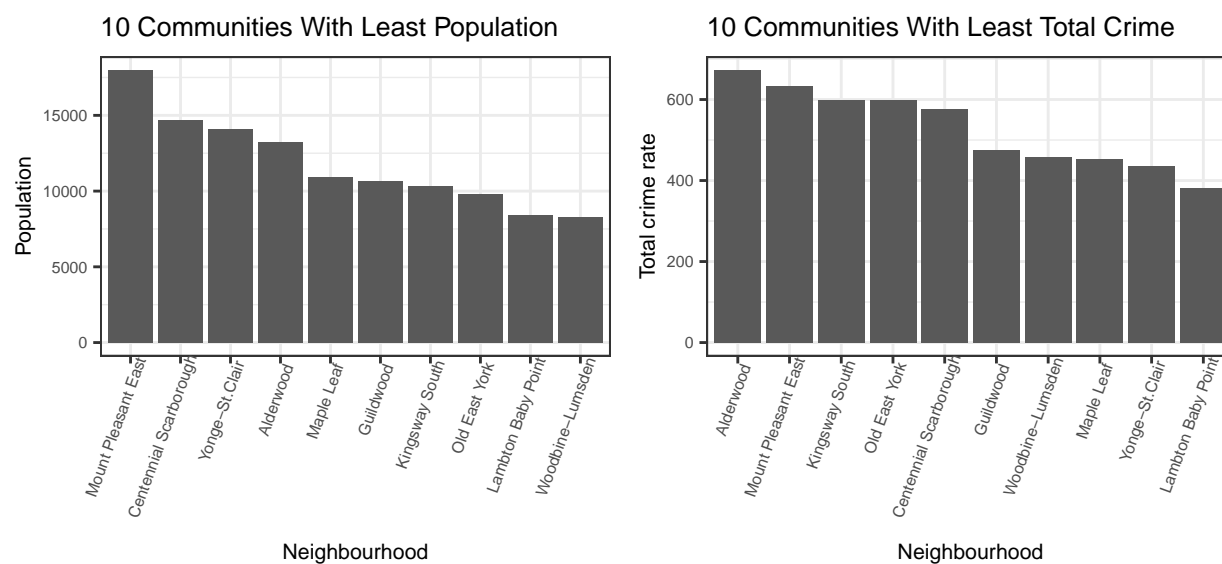


Figure 4: Number of crimes in the 10 least populated neighborhoods

3 Model

3.1 Methodology

First, I created new variables that will be used in this study, namely the sum of the number of each type of criminal activity, such as “Total Assault Crimes,” “Total Auto Theft Crimes,” and “Total Breaking Crimes,” for the period 2014-2020. Second, the number of each crime activity over the six years is summed to examine trends in the criminal activity that occurred in each community over the six years. The resulting model can be used to predict crime rates. Therefore, the number of criminal activities per year was not as significant. Secondly, a new variable called `Crime_in_total` was created, the sum of all criminal activity in each community over the six years. Unlike the previous new variable, this one contains the sum of all criminal activity for each community over the last six years. It gives a more intuitive representation of the number of crime incidents in each community. In addition to this, I removed variables that would not play a role in this study, such as `id`, where each community has its unique name and is not confused. Therefore there would be no need for additional numbering. Then, after aggregating the number of crimes for each year, the number of each crime for each year can be removed because this paper wants to examine the overall trend of criminal activity during these six years. For the treatment of crime rates, the crime rates for all crimes in the last six years are still added together to obtain the sum of crime rates for all criminal activities. In addition, public security authorities can use this model to improve the safety of communities with more security risks by imposing stricter management on them. Therefore, the crime rate of each specific criminal activity can also be removed. Otherwise, these variables would be repeated in the total crime rate. The table below contains all the essential variables in this study (including the newly created variables). To see if there is a relationship between the number of crime events in each community and the population of the community we previously predicted, I performed a multiple linear regression using the R language (R Core Team 2020). This model is well suited for exploratory data analysis and allows for further investigation of this relationship and how it might be conducted in the future.

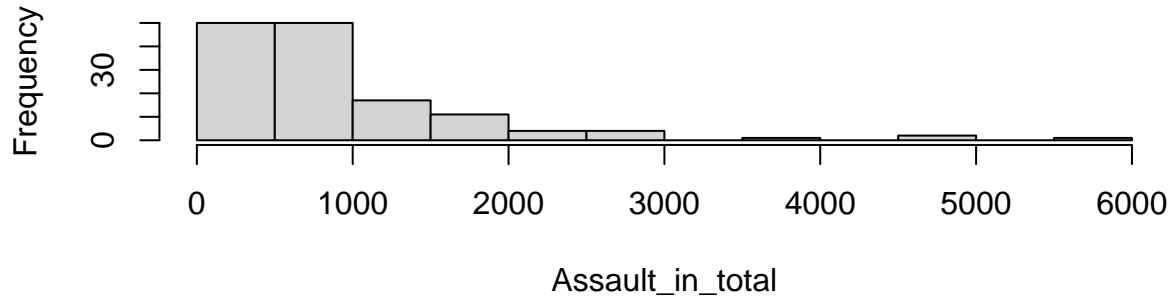
A simple way to check the first condition is to draw a curve of the response versus the fitted values. The fitted value is satisfied if all data points are randomly scattered on a straight line. We can draw a scatter plot of all predictors for the second condition. If all the paired relationships appear linear, this condition is satisfied. If all the additional conditions hold, I will draw two residual plots to test the model’s assumptions. residuals versus fitted values, and the normal Quantile-Quantile (QQ) plot. These steps will ensure that we end up with a statistically sound model and can answer our research question. Finally, I used a test set to evaluate the validation of the model. Then, to obtain the final model, I performed a partial F-test to find the better model between the first model (complete) and the second model (reduced). The null hypothesis is that all coefficients removed from the complete model are zero. The alternative hypothesis is that at least one coefficient removed from the complete model is not zero. If the p-value is more significant than 0.05, then we cannot reject the null hypothesis and should choose the reduced model. Otherwise, we should choose the full model. Using data from the test set, I fit a new model with the same formula as the final model formula (picking some essential variables). The next essential step was to compare their coefficients and determine if the model was reasonable based on the similarity of the coefficients of the variables in the two models. This is because the similarity of the coefficients of the variables in the two models can be used to determine whether the model is meaningful or not.

3.2 Multiple Linear Regression

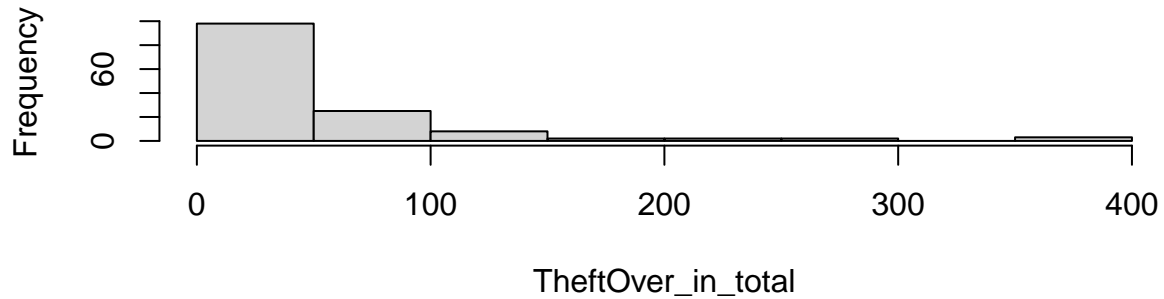
To see if there is a relationship between the number of crime incidents in each community and the population of the community we previously predicted, I performed a multiple linear regression using the R language (R Core Team 2020). This model is well suited for exploratory analysis of data and allows for further investigation of this relationship and the way it might be conducted in the future. I believe that multiple linear regression is a very intuitive statistical method that allows us to summarize and investigate the relationship between multiple continuous () variables. Some variables, denoted as x , are considered as

predictors, while another variable, considered as a response or outcome, is denoted as y . With multiple linear regression, we cannot yet draw decisive conclusions about causality, because it is very hasty. This model allows to explore any relationships that exist and their potential strength.

Histogram of Assault_in_total

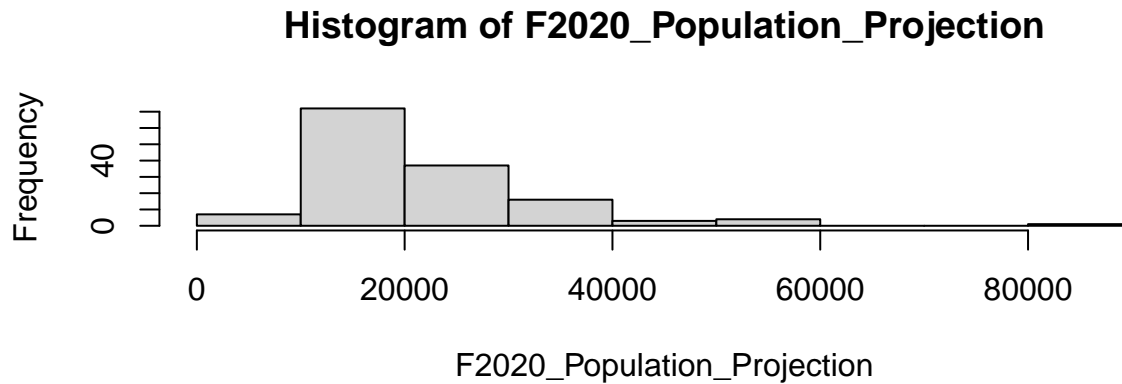
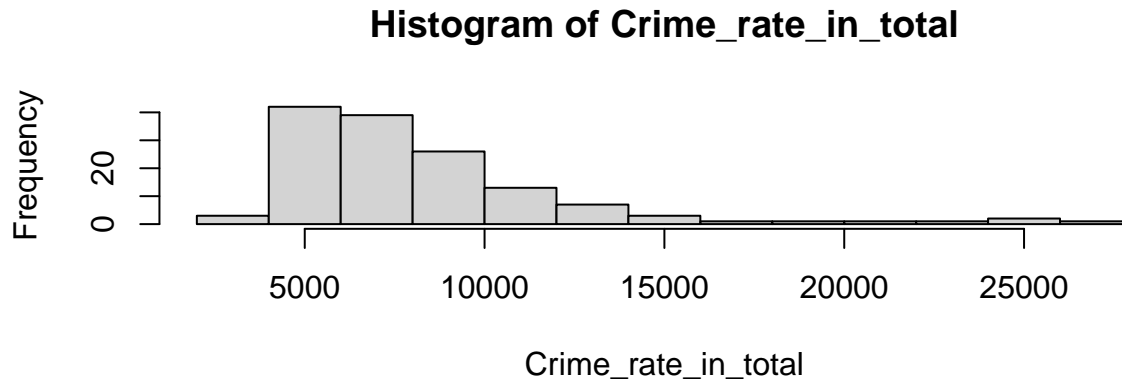
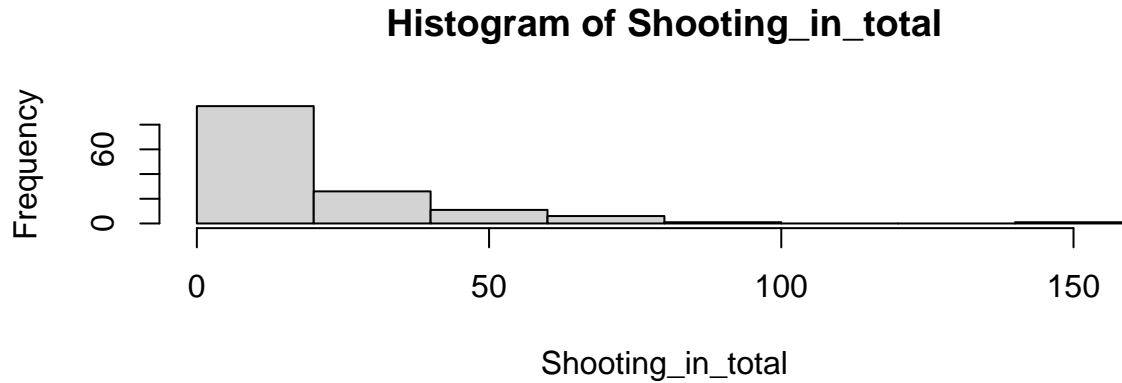


Histogram of TheftOver_in_total



Histogram of Homicide_in_total





For the selection of model variables, I included in the first model all variables that could affect the response (i.e., crime rates in the 140 communities) based on relevant data information.

$$Y_1 = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + \beta_9 X_9 + \epsilon$$

In this model, Y_1 is the crime rate for each community for the last 6 years (2014-2020) and X_1 - X_9 represent 9 different variables: Assault_in_total, AutoTheft_in_total, BreakAndEnter_in_total, Robbery_in_total, TheftOver_in_total, Homicide_in_total, Shooting_in_total, Crime_rate_in_total, and

F2020_Population_Projection which is temporary populations (such as commuters and business customers, are not included). Each of these nine different variables plays an important role in studying the amount of criminal activity that occurs in a community, so in the first model we want to include all the variables that would affect response value. β_0 denotes the projected value of Y when X is 0, and β is the expected change in Y when X increases.

4 Result

Table 2: Summary table of the first model

	Estimate	Std.Error	T value	Pr	Significance
(Intercept)	7874.21408	382.017	20.612	< 2e-16	***
F2020_Population	-0.30368	0.0222	-13.654	< 2e-16	***
Assault_in_total	2.46888	0.8068	3.060	0.002926	**
AutoTheft_in_total	1.59705	1.4223	1.123	0.264533	
BreakAndEnter_in_total	6.96331	1.7537	3.971	0.000145	***
Robbery_in_total	7.75688	2.8612	2.711	0.008048	**
TheftOver_in_total	0.67063	8.3129	0.081	0.935883	
Homicide_in_total	24.73884	74.3394	0.333	0.740081	
Shooting_in_total	21.73055	11.8951	1.827	0.071075	.
Crime_in_total	NA	NA	NA	NA	

From this table, we can know that there are four significant variables: F2020_Population_Projection, Assault_in_total, BreakAndEnter_in_total, Robbery_in_total. Then we can get the reduced model which includes all significant variables.

$$Y_{Reduced} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon$$

In the reduced model, Y_1 is the crime rate for each community for the last 6 years (2014-2020) and X_1 - X_4 represent 4 different variables: F2020_Population_Projection, Assault_in_total, BreakAndEnter_in_total, and Robbery_in_total. This model tells us that crime rates in some Toronto area neighbourhoods over the past six years have a strong relationship with the neighbourhood's population and the three crime events of assault, break and robbery. In other words, in areas where these three crimes are prevalent, then the crime rate of the community will be increased. This result is in line with the popular consciousness that in some areas where robbery and theft occur frequently, the probability of other more serious crimes occurring in this community will also be high. The results obtained from this model are also the primary purpose of our study.

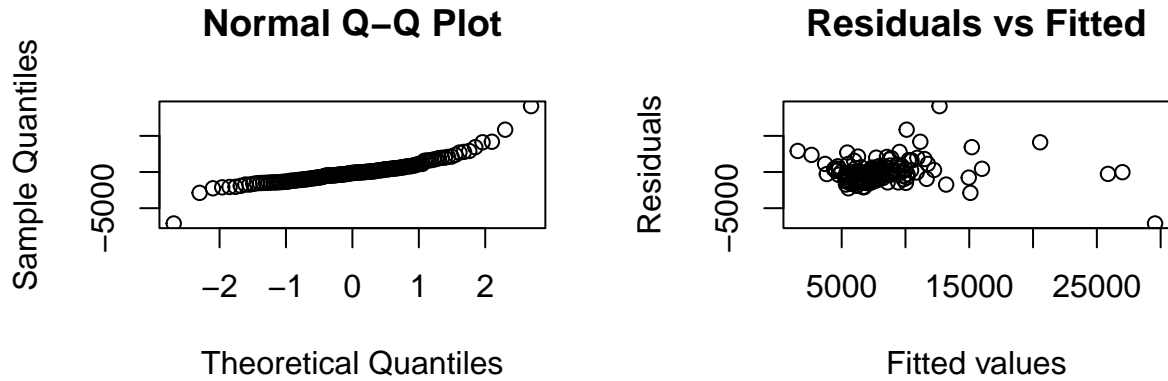


Figure 5: Residual scatter plots

As can be seen in the plot of residuals versus fitted values on the left, we can see that there is no systematic or fan pattern present, which is very good for the model. The points in the QQ plot lie on the diagonal of a straight line and, most importantly, have very little deviation from the line. Therefore, we can conclude that the model satisfies all four assumptions. Then, we can find problematic observations of the second model. There are three observations (1st, 63rd, 67th) whose standardized residual is not between -2 and 2, so three outliers have been found. Then I calculated the Cook's distance for each observation and compared it with the cutoff value. It turns out that all the Cook's distances are smaller than the cutoff value, which is a very good sign because it means that no influential points were found. The p-value is 0.8454, which is greater than 0.05, so we should choose the reduced model (Total crime rate = $\text{lm}(\text{Crime_rate_in_total} \sim \text{F2020_Population_Projection} + \text{Assault_in_total} + \text{BreakAndEnter_in_total} + \text{Robbery_in_total})$) for final model.

5 Discussion

5.1 What is done in this paper?

This paper analyzes the number of crime incidents in a community located in the Toronto area over the last six years from different perspectives about the population and other factors related to this community and its potential impact on social development, which is of importance to both individuals and the Toronto area policing authorities. The original survey categorized different crime incident types (e.g., robbery, theft, and murder). The original dataset recorded the number of different crime incidents and the probability of their occurrence over these six years. In other words, the subject of our study is the impact of these specific crime events and community demographics on a community's policing environment. First, we ranked the 10 communities with the highest and lowest number of crime events in the last six years (2014-2020) in 140 communities. Next, we found a clear relationship between the number of crime incidents in a community and the population size of that community, meaning that the population size of a community directly affects the probability of crime incidents in the community.

In addition, we used some statistical reference values (e.g., mean, median, etc.) to investigate the frequency and specifics of different crime events over the last six years. The significance of this step is that it can help social security staff target poorly policed communities for better management to develop more reasonable regulations and visualize the security situation and the living standard of residents in the Toronto area. This is because, in some communities that do not perform well in terms of social security, residents' standard of living in this community is also lower. The data in this paper show that the frequency of crime incidents

in most areas is still at an average level in recent years. Second, a linear regression model analyzes the relationship between the number of crime incidents and the community's population. This view is consistent with the public perception that communities with large populations will also have more crime incidents. However, the final model shows that factors affect the number of crime incidents in a community beyond the population. There are specific crime incidents (e.g., assault, break, and robbery). Overall, this paper presents a different perspective on the potential relationship between the number of crime incidents in a community and the community's population and some other factors related to the security situation in the community.

5.2 What is something that we learn about the world?

After completing this report, I have the following thoughts about the current social security and community safety conditions. In the current social context of great integration and increasing openness, although social civilization has achieved leaps and bounds, at the same time the difficulty and complexity of managing all aspects of social security is increasing, and security control is becoming more and more indispensable (Foster 2008). As government officials and social residents, we need to understand the importance of security control, analyze the current weaknesses of security control, and propose some corresponding improvement countermeasures to strengthen the control function of security personnel and maintain social security stability.

Analyzed from the social level, the implementation of security prevention and control work is conducive to ensuring the stability of social order, reducing undesirable behaviours such as riots and crimes, and reducing the negative impact of undesirable behaviours on society (Skogan 1986). This is important for protecting national interests and maintaining social stability. And it is conducive to maintaining the market order, which has important significance and influence on maintaining and promoting society's stable and orderly development. Analyzed from the people's point of view, the implementation of security and control has an important role in protecting and improving the legitimate rights and interests of the people, providing convenient and effective means of protection for the people's daily travel, buying and selling transactions, etc., so that conflicts and disputes among the people, etc., can be effectively handled to avoid more serious consequences.

As mentioned above, the current society is dynamic and open. Since public security control develops and changes with the development of society, it presents dynamic characteristics and begins to develop in a dynamic direction. Under such a new situation, the personnel of public security organs should strengthen the concept of public security control and learn to analyze specific problems with scientific methods of thought. Focus on the root cause of the problem and eradicate it thoroughly, and learn to solve the problem at the source. For the use of police force, public security authorities should follow the principle of economy (Bursik Jr 1993), learn to make good use of police force, cherish the use of police force, forbid to invest large police force in small problems, adhere to the realistic and pragmatic, focus on the effectiveness of public security control work, and be good at grasping the key points of control. Primarily for communities where crimes always occur, a more stringent management system should be carried out to ensure the safety of community residents.

For a community with a high incidence of crime in the last six years, there is a gap in the security of this community. In terms of the number of existing permanent civilian police allocations, a large number of police forces are occupied by non-policing, and the regular allocation of resources has caused a strain on police resources, increased the funding of public security organs, and resulted in a disproportionate investment of funds and police forces to the effectiveness of security control (Messer 2006). At the same time, the typical allocation of police resources does not match the current dynamic security control (Diamond 2004). It is challenging to meet the actual needs of dynamic security control, so the phenomenon of frequent crime incidents in specific communities has emerged. Although Canada's laws and regulations are constantly

being amended and improved, the amendment of the law is a phase of work that can not be sustained, which determines its time has a certain lag. This lag makes the laws and regulations can not be on the security control process of the new problem of immediate and effective management, security control lack of vital disposal means, police officers in the active response to the case on the initiative is not enough(Feldstein 2002). This has led to the residents of some communities not being able to get timely protection of their lives and property when they are robbed or infringed upon.

5.3 What are some weaknesses of what was done?

- In this paper, we analyze the variables in the database. We roughly assume that all crimes in each community in the last six years are recorded in the dataset. However, a persuasive paper should analyze the data in conjunction with those not included in the dataset. Therefore, the first shortcoming of this paper is that it does not consider data that are hidden or forgotten. Secondly, this study does not analyze the deeper causes that contribute to the high incidence of crime. From the data, we can see that the number of crime incidents in the Toronto area increased significantly in 2020, and we cannot ignore the reasons behind it. The outbreak of Covid-19, for example, has caused harm to any area, causing many residents to lose their jobs and become homeless(Gaetz 2014), unable to secure their basic livelihood and living conditions have become so bad that they do not even have enough money to treat the physical damage caused by the virus. The epidemic has had a considerable impact on the world economy, severely affecting people's wages and incomes and further increasing inequality in income distribution. This has directly led to an elevated crime rate in the community, and people have had to resort to robbery to get money to secure their livelihoods. In completing this paper, we should analyze it in the social context of the year.
- The statistics of the original database are not complete, resulting in less than convincing conclusions. This data only classifies the different crime incidents and does not provide statistics on the number of residents who die in each community each year because of the crime incidents. In other words, the number of community residents who died as a result of assaults is the best indicator of the safety of their community. The incompleteness of the data statistics directly leads to a certain lack of rigor in the results of the study, because some communities with a high rate of fatalities (e.g., murders, shootings, and rapes) even though the number of crime incidents is not very high, will have a much higher risk factor than those communities where theft is a frequent occurrence. However, this paper only analyzes the impact of different crime events and population sizes on community safety. It does not consider the number of deaths and injuries caused by crime events each year. However, we cannot deny that it is crucial to focus on areas where incidents frequently lead to casualties, regardless of the neighbourhood and city.
- The dataset used in this paper is as of 2020, and the social environment and policing in the Toronto area has changed significantly over the past seven years. The standard of living and educational attainment of residents is different from the past, and it is essential to gather some recent data to support the ideas in this paper.

5.4 What is left to learn or how should we proceed in the future?

First of all, through this study, we learned that multiple factors shape the security situation of a person's community. Community risk is a combination of the likelihood and consequences of a hazardous situation occurring in a community(Flake 2005). It is a collective term for all types of risks that threaten community safety, which act directly on the community and affect its safety and stability, thereby reducing the sense of security, well-being, and access of community residents. For the predictable risks in the community, control

is put in place to ensure the community's safety. For threats that are not foreseeable in the community, residents need to be aware of self-protection. Community safety requires both measures to reduce the likelihood of danger and the ability to respond to it. We need to be alert to the dangers our surroundings may pose and develop our safety awareness in our daily lives(Kim 2015).

Secondly, academically, we learned how to analyze problems with different perspectives through this study. As well as to relate the data provided to the natural environment rather than just analyzing the data mechanically, so that the reports made are convincing. In subsequent academic research, the analysis should be done in the context of the social environment of the time. In addition, when collecting data, it is essential to separate the data into categories and not to use a single characteristic to explain all individuals, which means that we should respect differences. In addition, academic research should classify data as carefully as possible.

5.5 Limitations of Analysis

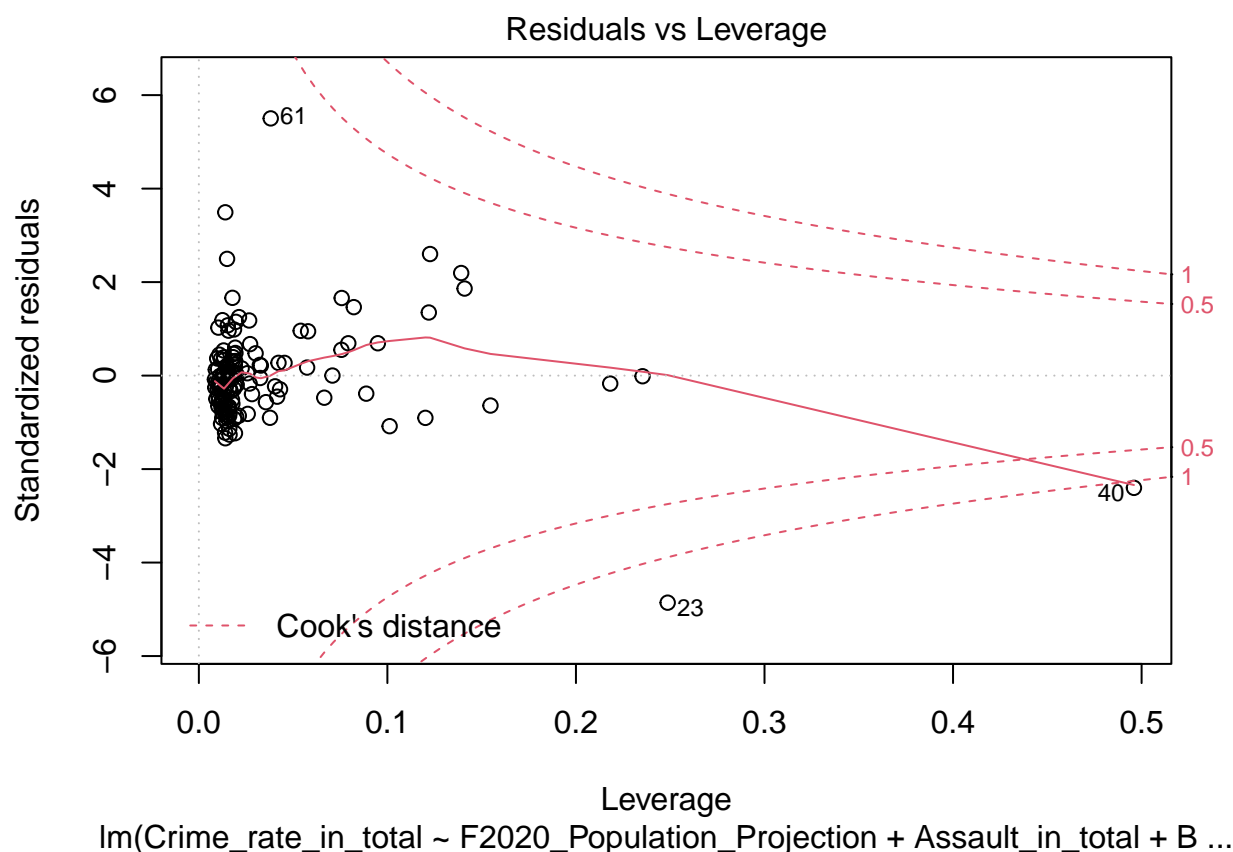


Figure 6: Residuals & leverage plot

Our final model may not fit the data well because, inevitably, there are problematic observations (outliers and leverage points such as the 63rd, 67th, and 40th positions). However, we cannot change or remove data to make the model fit the data better (this is statistically meaningless, and the results obtained in this way are not credible and informative), so it is difficult to correct. In future professional studies, we are also

not allowed to modify experimental data to get the established results in our minds, which is against the authenticity of academic reports. Also, since this data is from the last six years in Toronto, Canada, this model may only apply to Canada.

6 Conclusion

This report focuses on a comprehensive statistical study of the security situation in different communities in the Greater Toronto Area to provide a valuable reference for people who want to settle in Toronto and guide government workers responsible for social security. In general, communities where crime is frequent, tend to have poor social security systems that allow criminals to commit crimes, and communities where fewer crimes occur mean that local police supervision is reasonable and strict (Taylor 1995). Using statistical methods, we obtained a model of the security situation in Toronto communities, a model that explains well the factors that contribute to crime incidents in Toronto. This model tells us that crime rates in some Toronto-area communities over the past six years are strongly related to the population of that community and three types of crime incidents: assault, break-ins, and robberies. In other words, in areas where these three crimes are prevalent, the crime rate in that community is higher. This result is consistent with the widespread perception that in some areas where robbery and burglary are frequent, the probability of other more serious crimes in that community will also be high. The results obtained from this model are also the primary purpose of our study.

Reference

- Agnew, R. 1999. *A General Strain Theory of Community Differences in Crime Rates*. *Journal of Research in Crime and Delinquency*. <https://journals.sagepub.com/doi/abs/10.1177/0022427899036002001>.
- Bhandari, Pritha. 2021. *Missing Data | Types, Explanation, & Imputation*. <https://www.scribbr.com/statistics/missing-data/>.
- Bursik Jr, & Grasmick, R. J. 1993. *Economic Deprivation and Neighborhood Crime Rates*. <https://heinonline.org/HOL/LandingPage?handle=hein.journals/lwsocrw27&div=26&id=&page=>.
- Chenane, & Wright, J. L. 2021. *The Role of Police Officer Race/Ethnicity on Crime Rates in Immigrant Communities*. <https://journals.sagepub.com/doi/full/10.1177/2153368718777278>.
- Diamond, P. 2004. *Social Security*. *American Economic Review*. <https://pubs.aeaweb.org/doi/abs/10.1257/000282804322970670>.
- Feldstein, & Liebman, M. 2002. *Social Security*. *Handbook of Public Economics*. <https://www.sciencedirect.com/science/article/abs/pii/S1573442002800118>.
- Flake, D. F. 2005. *Individual, Family, and Community Risk Markers for Domestic Violence in Peru*. *Violence Against Women*. <https://journals.sagepub.com/doi/abs/10.1177/10778012042721298>.
- Foster, & Giles-Corti, S. 2008. *The Built Environment, Neighborhood Crime and Constrained Physical Activity: An Exploration of Inconsistent Findings*. <https://www.sciencedirect.com/science/article/abs/pii/S0091743508001588>.
- Gaetz, Gulliver, S. 2014. *The State of Homelessness in Canada 2014*. *Canadian Homelessness Research Network*. <https://yorkspace.library.yorku.ca/xmlui/handle/10315/29368>.
- Gelfand, Sharla. 2020. *Opendatatoronto: Access the City of Toronto Open Data Portal*. <https://CRAN.R-project.org/package=opendatatoronto>.
- HuffPost. 2018. *Toronto Blows Past Winnipeg for Highest Homicide Rate in Canada*. Toronto, Canada. https://www.huffpost.com/archive/ca/entry/toronto-homicide-rate-murder_a_23594924.
- Kim, S. Y. 2015. *Safety Awareness and Safety Practice Behavior of College Students*. *Journal of Digital Convergence*. <https://www.koreascience.or.kr/article/JAKO201509057414174.page>.
- Messer, Kaufman, L. C. 2006. *Neighborhood Crime, Deprivation, and Preterm Birth*. *Annals of Epidemiology*. <https://www.sciencedirect.com/science/article/abs/pii/S1047279705003182>.
- Portal, Public Safety Data. 2018. *Neighbourhood Crime Rates (Boundary File)*. <https://data.torontopolice.on.ca/datasets/neighbourhood-crime-rates-boundary-file-/explore>.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Skogan, W. 1986. *Fear of Crime and Neighborhood Change*. <https://www.journals.uchicago.edu/doi/abs/10.1086/449123>.
- Taylor, R. B. 1995. *The Impact of Crime on Communities*. *The Annals of the American Academy of Political and Social Science*. <https://journals.sagepub.com/doi/abs/10.1177/0002716295539001003>.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- . 2021. *Tidyr: Tidy Messy Data*. <https://CRAN.R-project.org/package=tidyr>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemond, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2021. *Dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>.