A Permutation test to assess the geographic patterns of violent crime in downtown Toronto

Minni Xiong

20653732

University of Waterloo

Instructor: Samuel W.K. Wong

Abstract

The purpose of this report is to examine the geographic patterns of violent crimes in Toronto area. More specifically, to verify whether the crime rate is higher in downtown area than in other places. I use a non-parametric testing on crime cases, using traffic accidents as a control group. I will use some plots as visual presentations for easier understandings for non-statistical audiences.

Keywords *Permutation test; spatial data analysis; plots; crime rate; Toronto*

1 Introduction

Toronto is the largest city in Canada, and while violent crimes are considered a serious public health issue, limited research has test on the spatial of the crime rate (Toronto public health, 2019). Criminal activities are usually unevenly distributed across the space (Wang et al.,2019). In 2011, Statistics Canada issued a publication *Neighbourhood Characteristics and the Distribution of Crime in Toronto: Additional Analysis on Youth Crime*. The paper studied the spatial distribution of youth crime in Toronto and tried to identify a relationship between youth crime and neighborhood characteristics. The study finds that crimes are not evenly distributed in a municipality but tend to be concentrated in certain neighbourhoods or 'hot spots'.

In this study, we want to see if downtown Toronto is one of the 'hot spots'. That is, we want to verify if there tends to have increased rate of crime around downtown Toronto.

In section 2, I use a permutation test to carry out the hypothesis testing.

In section 3, I extend the study using another set of dataset to see if the result is the same.

2 Testing

2.1 Motivation

We observe the locations of events within a bounded region A. We suspect the event rates may have an association with a prespecified point x_0 in region A and our goal is to investigate this association.

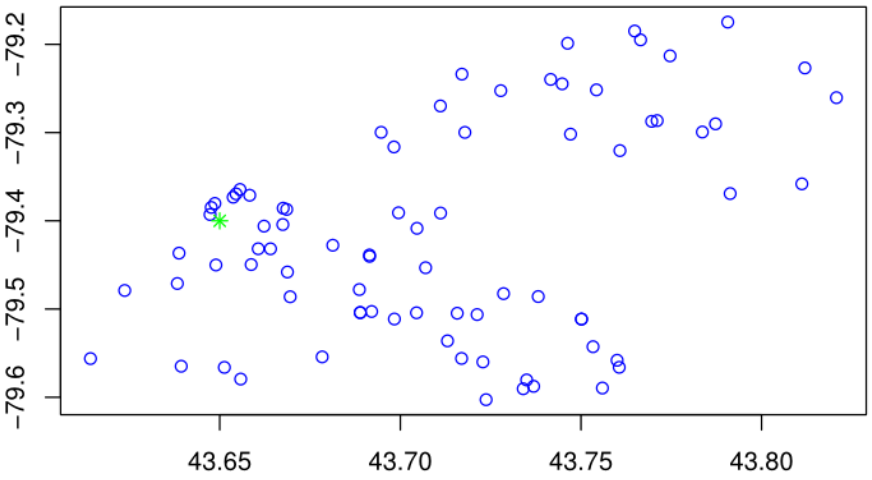

Fig.1 Spatial distribution of 80 homicide cases in the Toronto area in 2019, and (*) is a point in downtown Toronto

In this case, we observe the locations of homicide cases in Toronto. We use the data for homicide because homicide is one of the Major Crime Indicators (MCI). Data is pulled from Toronto Police Service public safety data portal website. The data originally contains all homicide cases from 2004 to 2020. I extracted the data of 2019, to stay relevant for the time. The prespecified point x_0 is chosen by the coordinates from google map when I type in downtown Toronto. We can see there are some cluster cases around the prespecified point x_0. Now, we want to investigate if it is sufficient to establish the association.

Some major methodological issues:

- Homicide cases has a relatively small sample size
- Other geographical or demographic factors might impact the spatial variation of the crime rate.

To solve both issues, we choose to use a non-parametric testing. In this case, I choose the permutation testing. We will use another data set as control group.

Consider another set of data, that contains the location of a more common event. This data should adjust for the demographic factors in the phenomenon of interest, but should not be assumed to have an association of the prespecified point x_0. This dataset is considered the control group of the testing.

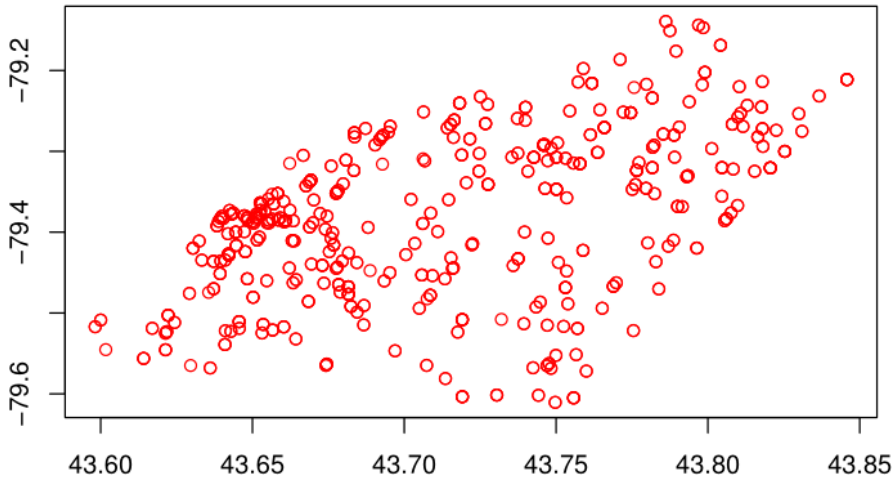

Fig.2 Spatial distribution of 808 traffic collisions (KSI) in the Toronto area in 2019

Fig. 2 shows the location of all 808 traffic collisions where a person was either Killed or Seriously Injured (KSI) in the Toronto area in 2019. This map should correctly reflect the spatial variation adjusted for geographic and demographic factors. For the purpose of our study, I would assume there is no association between fatal traffic collisions and downtown area. Data is pulled from Toronto Police Service public safety data portal website. The data includes all traffic

collisions where a person was either Killed or Seriously Injured (KSI) from 2006 to 2019. Again, I extracted the 2019 data.

Note that, since we are using observational data, we can only examine association but not causation.

2.2 Visualization of data

I have provided two figures to show the location of events on the map of Toronto.

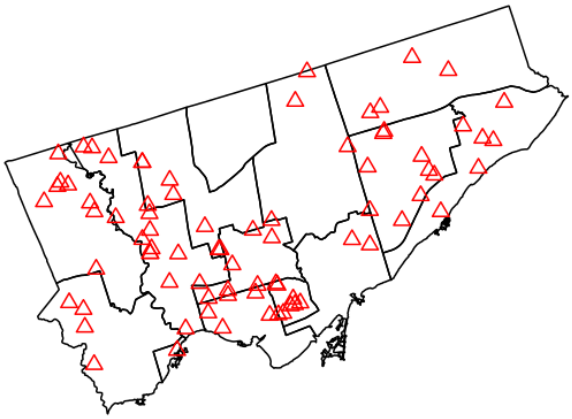

Fig.3 Location of homicide cases in 2019 in Toronto

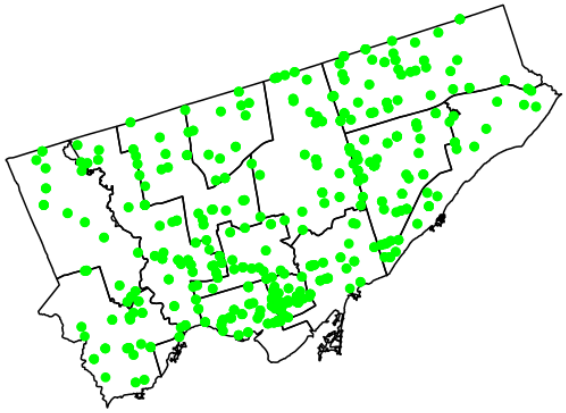

Fig.4 Location of KSI in 2019 in Toronto

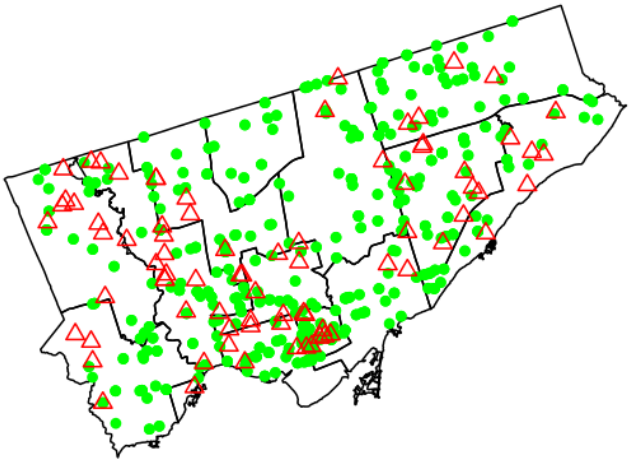

Fig.5 Location of homicide cases and KSI

We can see that KSI cases are distributed more evenly throughout the city, while homicide cases are more concentrated in certain area.

2.2 Permutation test

We start with two sets of data and want to test out if these two sets of data share the same spatial distribution. The idea is that, if they share the same distribution, we can "shuffle" their labels. That is, we can randomly assign labels to the collection of points. We then choose a random test statistic and run the test, and see if the results comply with a random allocation. If the null hypothesis is true, then no matter what test statistic we choose, the result should not change.

In our case, we use the two data sets: homicide cases $\{x_i, i = 1, \dots n\}$ and traffic collisions $\{y_j, j = 1, \dots m\}$. Our null hypothesis is these two data come from the same spatial distribution. Under the null hypothesis, we should be able to assign "homicide" and "traffic accidents" randomly to the collection of points. We choose the test statistic to be the sum of the squared distances from the i^{th} homicide cases to the prespecified point x_0. This test statistic allows us to check if there is increased incidence rate around x_0. Then, we use the randomization test to see if the test statistic is extreme compared to a random allocation. Since we assume there is no association between fatal traffic collisions and downtown area, if the two data come from the same distribution, that means there is no association between homicide cases and downtown area.

I will present the result in histogram for visual presentation.

2.3 Result

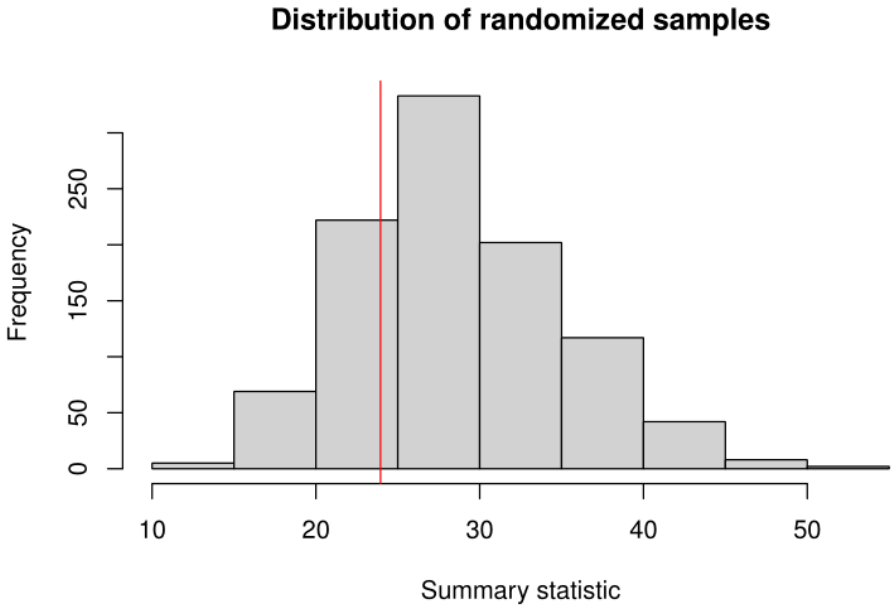

Fig.6 Distribution of randomized samples and observed test statistic.

The red line indicates the observed squared distances from the homicide cases to downtown area. From the graph, we can see that the homicide cases do not belong to the same distribution as traffic accidents. However, the difference is not very obvious. To further investigate for the question, I want to test on another set of data and see if the difference will be more obvious.

3 Extension

Since the result from the previous testing did not show a clear difference in spatial distribution between two groups of data, I pull up another dataset, shooting cases, to do the test and see if we can reach the same result. I choose shooting cases since shooting is also one of the Major Crime Indicators (MCI) according to Toronto Police Service.

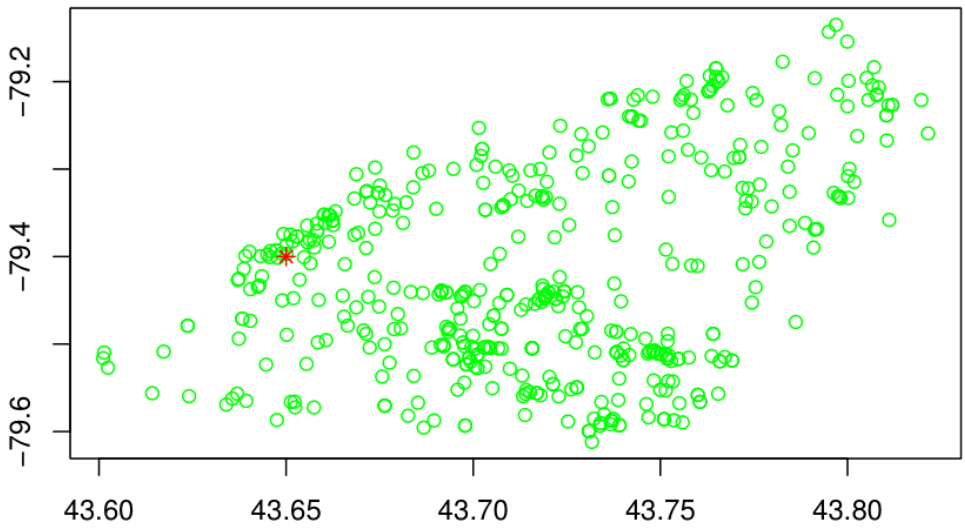

Fig.7 Spatial distribution of 493 shootings and firearm discharge in the Toronto area in 2019, and (*) is a point in downtown Toronto

Fig. 7 shows the location of all 493 shootings and firearm discharge in the Toronto area in 2019. Again, we see some cluster cases around the predetermined point x_0. We want to know if there is sufficient evidence to establish association of increased shooting accidents around downtown Toronto. Data is pulled from Toronto Police Service public safety data portal website. The data contain all shooting-related occurrences reported to the Toronto Police Service and I extracted 2019 data for consistency.

The setup is similar as above. We construct a permutation test, using the sum of the squared distances from the i^{th} shooting events to the prespecified point x_0 as test statistics and construct a histogram.

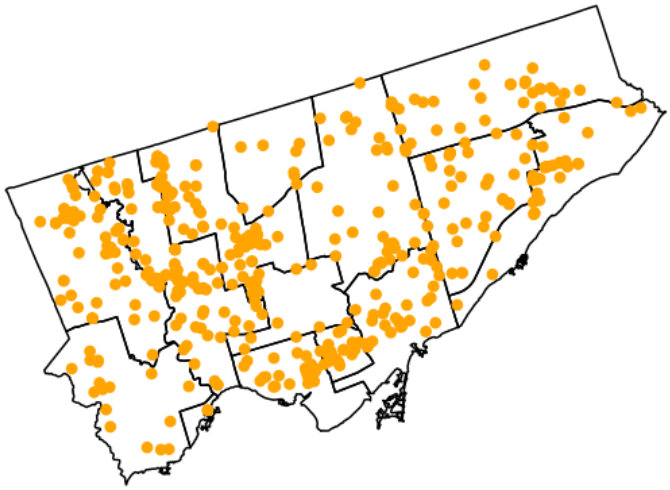

Fig.8 Location of shooting cases in 2019 in Toronto

Again, I plot out the location of shooting cases in 2019 on the map of Toronto. We can see the cases are not evenly distributed, but more concentrated on the left side.

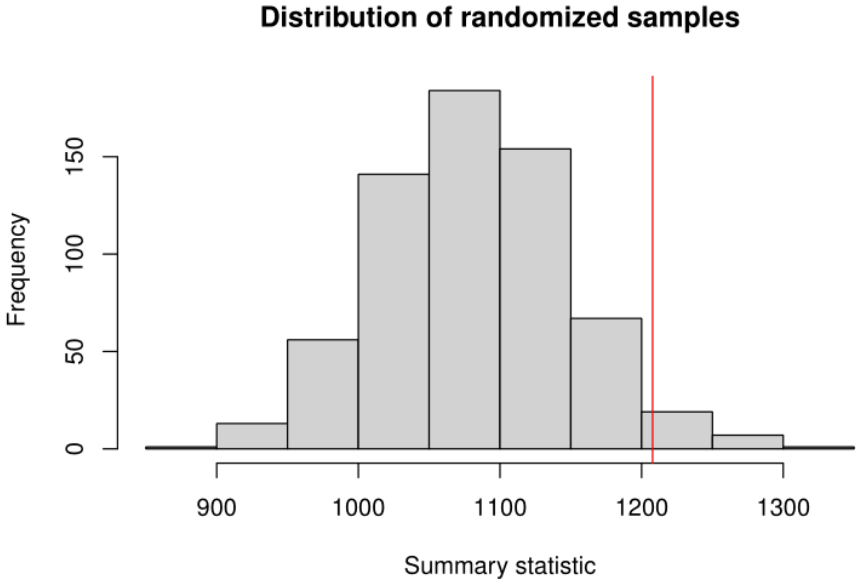

Fig.9 Distribution of randomized samples and observed test statistic.

From Fig.5, we can see that test statistics are extreme relative to the random allocation. Therefore, we can reject the null hypothesis of the two dataset coming from the same distribution. We can see that the test statistics are more extreme in this test than the previous test.

4 Interpretation

From the 2 permutation tests we did, we both reject the null hypothesis that the case data and the control data come from the same spatial distribution. Therefore, we can establish an association between increased crime rates and downtown Toronto. Notice that the significance or the extremity from the two tests are different. One possible reason is that shooting exhibits a stronger association compared to homicide cases. Therefore, the types of the crimes might affect the level of association.

5 Discussion

In this study, we use two dataset and a control set to test out the association between increased crime rates and a position in downtown Toronto. The tests show evidence of association. We found that different crime activities show different level of deviations from the control group distribution. This study can be extended by using other datasets from Major Crime Indicators (MCI) such as assault, auto theft, break and enter, to do the tests and see the disparity between different datasets. We can also use other dataset as control group and do the testing.

This is only a preliminary test to establish an association. Later, we can use modelling to find the predictors for the increased crime rate.

Another important thing to note is that, the data we use is from 2019. Consider the current situation (Covid-19 and lockdown), the spatial distribution of crime rates might be different now.

References:

1. Wang, L., Lee, G., & Williams, I. (2019, January 21). *The Spatial and Social Patterning of Property and Violent Crime in Toronto Neighbourhoods: A Spatial-Quantitative Approach*. MDPI. https://www.mdpi.com/2220-9964/8/1/51.

2. Public Safety Canada / Sécurité publique Canada. (2015, December 3). *The spatial distribution of police-reported youth crime in Toronto*. https://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/dstrbtn-yth-crm/index-en.aspx.

3. Statistics Canada (2015, November 27). *Neighbourhood Characteristics and the Distribution of Crime in Toronto: Additional Analysis on Youth Crime.* https://www150.statcan.gc.ca/n1/pub/85-561-m/2011022/part-partie1-eng.htm.