Examining the Potential Relationship Between Tree Canopy Cover and Crime Rates in Toronto, Canada Neighbourhoods

Shakeeb Tahir

April 8th, 2022

GGR463H5S - Geographic Information Analysis and Processing

Prof. Damian Maddalena

1) Introduction / Background

Trees have long been associated for having many benefits such as reducing air pollution, improving wild-life sustainability and reducing surface temperature by providing shade. However, recently there have been an increasing number of studies showing the potential benefits that trees might have on societal factors such as crime rates. For instance, a study in New Haven, US found that greater tree canopy cover was associated with lower violent and property crime rates (Gilstad-Hayden et al., 2015). Another study in Baltimore, US found that a 10% increase in tree canopy was associated with a roughly 12% decrease in crime (Troy et al., 2012). Moreover, when tree cover was lost in Cincinnati, US there was a relative increase in crime that followed (Kondo et al., 2017). Some studies have gone even further and analyzed how different tree types might affect crime rates. A study from Portland, Oregon found that larger trees had a relationship with lower crime rates while areas with smaller trees actually had increased crime rates (Donovan et al., 2010). This trend is not just prevalent in major U.S cities but also in other regions of the worlds. In developing countries such as Colombia, greater tree density and sizes within the city of Bogota were related to lower homicide occurrences (Escobedo et al., 2018).

As such, with all these studies in the United States and developing countries pointing to the possible relationship of greater tree numbers and lower crime rates, I want to know whether this is also the case in Toronto, Canada and to what extent there could be a relationship. All of these studies have been done in major cities within the United States or developing countries such as Colombia, while there has been a lack of research within this area in Canada. By analyzing tree cover within a major Canadian city, it can help to provide evidence as to whether this trend could also exist within Canada. By knowing if such a relationship exists, it can help city officials know where to plant more trees and plan future development projects accordingly.

2) Methods

The study area is the city of Toronto in Ontario, Canada. Toronto is one of the largest and most populous cities within Canada. The research objective is to determine whether the number of trees per square kilometer in Toronto neighbourhoods are associated with various crime rates, specifically assault, auto theft, break and enter, and robbery. The data for the study was obtained from various open-source websites. A tree points shapefile was obtained from the Toronto Open Data Portal (City of Toronto, n.d.) and a crime rates shapefile was obtained from the Toronto Police Services open data portal which included various crime rates such as shooting rates, homicide rates, robbery rates, etc., from the years 2014-2020. In order to stay consistent, only the crime rates from 2014 were used since that is the year the tree data was last updated. The specific crime rates analyzed were assault rate, auto theft rate, break and enter rate, and robbery rate as they were the main crime indicators and also had rates for each neighbourhood while other crime indicators had many 0 or null values which would make it hard to find associations.

The neighbourhood crime rates shape file was in the NAD 1927 UTM Zone 17N projection while the tree points shapefile was in the WGS 1984 projection. After importing the files into QGIS, the projections of the shapefiles was changed to NAD 1983 UTM Zone 17N. Now that the data had been obtained and processed, analysis could be done using a variety of tools and methods.

Firstly, the neighbourhood crime rates shapefile and tree points shapefile were combined into a single layer by using the QGIS "Join Attributes by Location" tool which also counts the number of tree points in each neighbourhood and plots it as a field in the attribute table. Next, the QGIS "field calculator" was used find the size of each neighbourhood in square kilometers, and then to find the number of trees per square kilometer. This is important as the data needs to be standardized to avoid bias as just looking at raw numbers will skew the analysis. For instance, the number of trees in a neighbourhood is likely influenced by the size of that neighbourhood which is why we need to divide the total trees by the area. Next, the maps of the trees per square kilometers, the assault rate, auto theft rate, break and enter rate, and robbery rate were made.

After making all the maps, the attribute table of the combined shapefile layer with all the calculations was exported as a CSV file and then imported in R/RStudio to perform regression analysis in order to analyze the potential associations between trees per square kilometer and the varying types of crime rates. The regression mode was done using the default linear model function within R. The regression models gave p-values to tell whether or not the relationship is significant if the values are less than 0.05. We can also check how well our models fit using R-Squared values. After the regression was done, a Moran's I analysis was conducted for each of the different types of crime rates to make sure there was no spatial clustering, since if spatial autocorrelation exists then we cannot be sure of the association between tree cover and crime rates as the crime rates are likely influenced by clustering. In order to conduct Moran's I, spatial weights needed to be conducted based on the nearby polygons of the neighbourhoods. A spatial weights matrix was made using rgdal and spdep which allowed us to calculate the Moran's I statistic and the significance.

A full workflow diagram of the methods can be found in Appendix A (Figure 1: Workflow Diagram) and the metadata of the data can be found in Appendix B (Table 1: Data Used).

3) Results and Discussion

Initial results from the QGIS maps show trees per square kilometer seem to be located near downtown and the center of Toronto (Appendix D: Figure 5). Assault rates (Appendix D: Figure 1) and auto theft rates (Appendix D: Figure 2) seem to be more towards the west side of the city while break and enter rates (Appendix D: Figure 3) and robbery rates (Appendix D: Figure 4) seem to be fairly dispersed throughout the city. The regression models were all significant with p values less than 0.05 however they all had low R^2 values which indicates the models may not fit well.

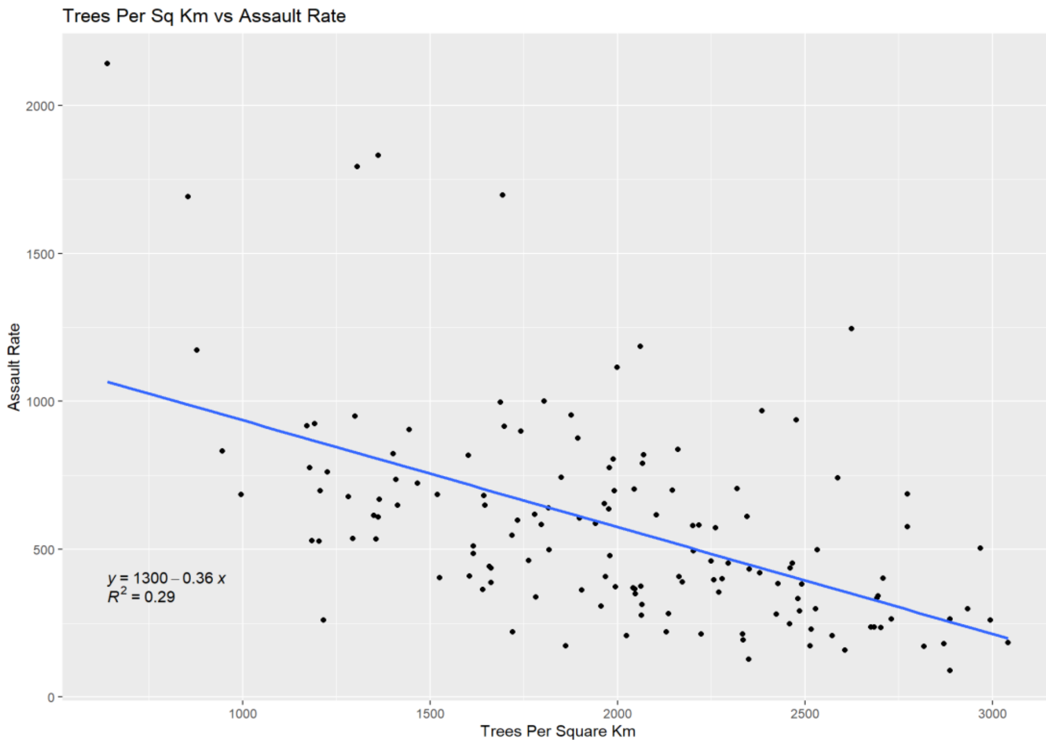


Figure 1: Assault Rate Regression Model

```
Residuals:
    Min      1Q  Median      3Q     Max 
-597.95 -194.18  -58.83  102.12 1075.07 

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)    
(Intercept)           1296.88120   98.89862  13.113  < 2e-16 ***
`Trees Per Square Km`   -0.36101    0.04796  -7.527  6.1e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 295.2 on 138 degrees of freedom
Multiple R-squared:  0.291,	Adjusted R-squared:  0.2859 
F-statistic: 56.65 on 1 and 138 DF,  p-value: 6.096e-12
```

Figure 2: Assault Rate Regression Significance

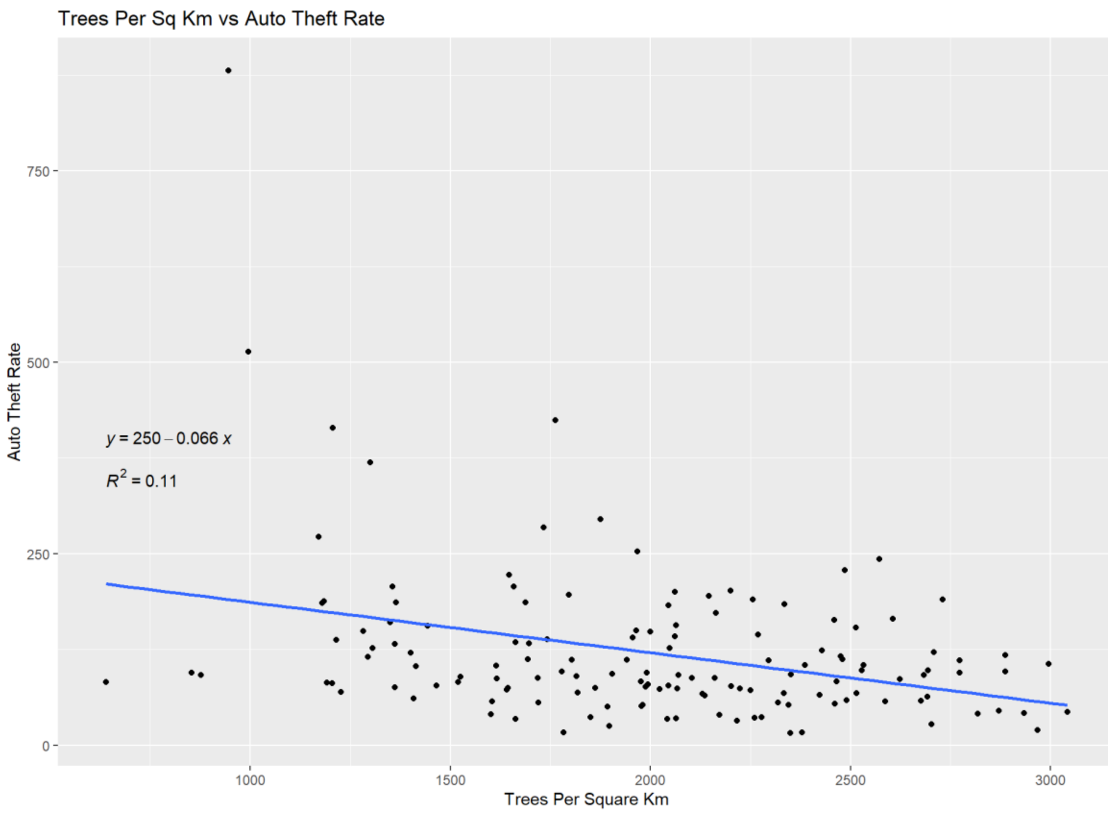


Figure 3: Auto Theft Rate Regression Model

```
Residuals:
    Min      1Q  Median      3Q     Max
-128.43  -52.92  -23.77   28.14  690.43

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)           252.56807   32.65561   7.734 1.96e-12 ***
`Trees Per Square Km`  -0.06586    0.01584  -4.158 5.60e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 97.48 on 138 degrees of freedom
Multiple R-squared:  0.1114,	Adjusted R-squared:  0.1049
F-statistic: 17.29 on 1 and 138 DF,  p-value: 5.6e-05
```

Figure 4: Auto Theft Regression Significance

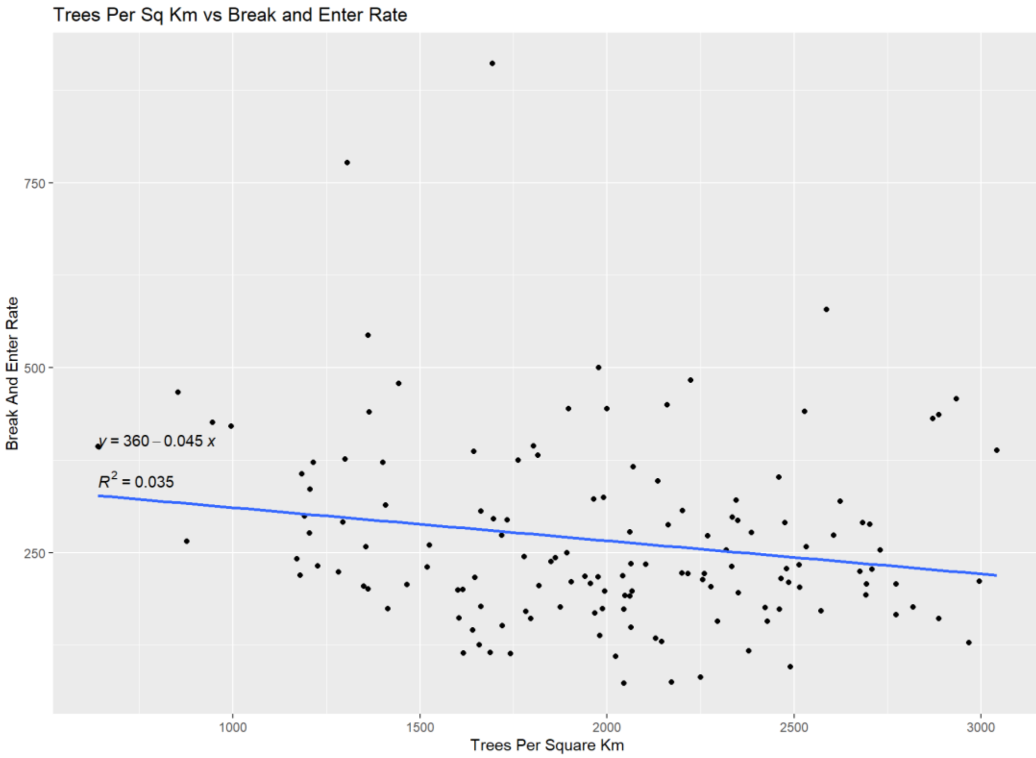


Figure 5: Break & Enter Rate Regression Model

```
Residuals:
    Min      1Q  Median      3Q     Max
-190.56  -72.63  -28.76   53.98  631.02

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)           355.73828   41.47043   8.578 1.77e-14 ***
`Trees Per Square Km`  -0.04479    0.02011  -2.227   0.0276 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 123.8 on 138 degrees of freedom
Multiple R-squared:  0.0347,    Adjusted R-squared:  0.0277
F-statistic:  4.96 on 1 and 138 DF,  p-value: 0.02756
```

Figure 6: Break & Enter Rate Regression Significance

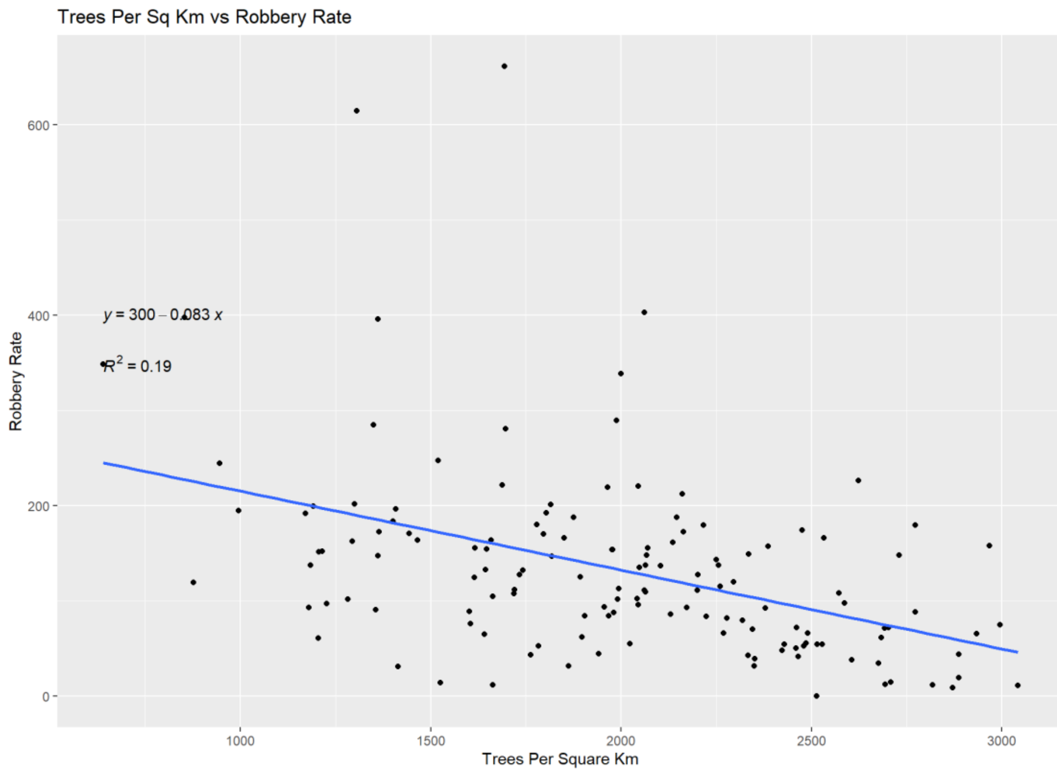


Figure 7: Robbery Rate Regression Model

```
Residuals:
    Min      1Q  Median      3Q     Max
-157.98  -45.87  -16.10   24.75  503.56

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)          298.14694   30.35504   9.822  < 2e-16 ***
`Trees Per Square Km`  -0.08289    0.01472  -5.630 9.69e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 90.62 on 138 degrees of freedom
Multiple R-squared:  0.1868,	Adjusted R-squared:  0.1809
F-statistic:  31.7 on 1 and 138 DF,  p-value: 9.69e-08
```

Figure 8: Robbery Rate Regression Significance

Since all of the regression models were significant, Moran's I needed to be tested to make sure these results were not due to the clustering of the crime rates. A spatial weights matrix was created in R using rgdal and spdep in order to test Moran's I.

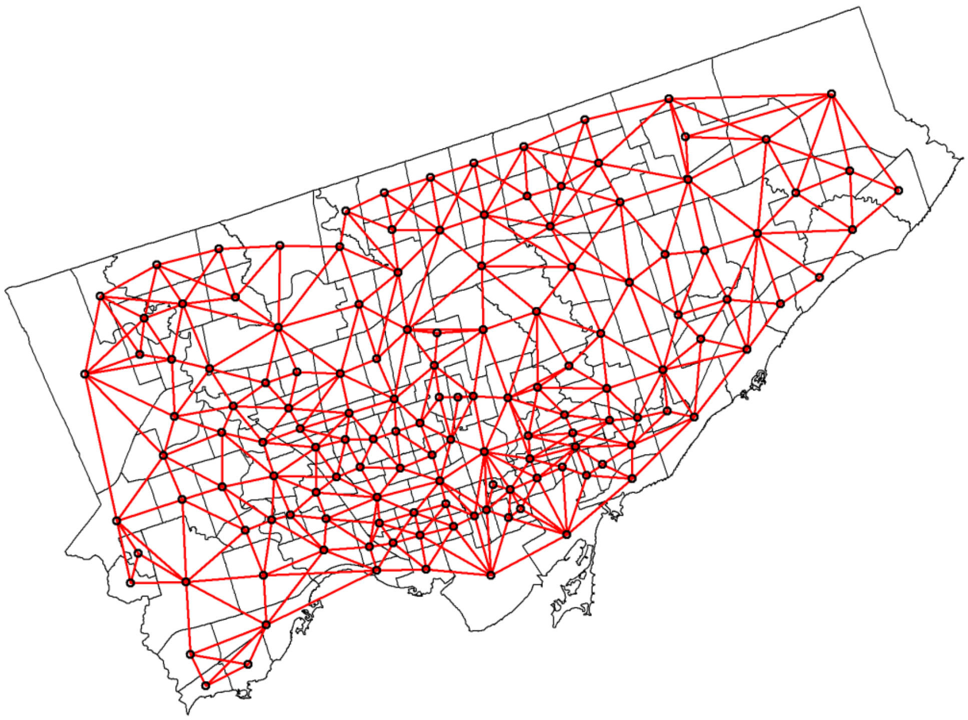


Figure 9: Polygon neighbour connections

After conducting Moran's I, it was shown that the p-values for each of the crime rates was significant which indicates that spatial clustering exists and we cannot be conclusive of the results from the regression models. The Moran's I statistic was also positive for every rate which indicatives positive spatial autocorrelation.

```
        Moran I test under randomisation

data:  neighbourhoods_spdf$Assault_Ra
weights: Toronto_listw

Moran I statistic standard deviate = 10.486, p-value < 2.2e-16
alternative hypothesis: greater
sample estimates:
Moran I statistic       Expectation          Variance
      0.496659982      -0.007194245       0.002308988
```

Figure 10: Assault Rate Moran's I

```
	Moran I test under randomisation

data:  neighbourhoods_spdf$AutoTheft_
weights: Toronto_listw

Moran I statistic standard deviate = 6.3178, p-value = 1.326e-10
alternative hypothesis: greater
sample estimates:
Moran I statistic       Expectation          Variance
      0.275823953      -0.007194245       0.002006745
```

Figure 11: Auto Theft Rate Moran’s I

```
	Moran I test under randomisation

data:  neighbourhoods_spdf$BreakAndEn
weights: Toronto_listw

Moran I statistic standard deviate = 6.0771, p-value = 6.118e-10
alternative hypothesis: greater
sample estimates:
Moran I statistic       Expectation          Variance
      0.283684974      -0.007194245       0.002291017
```

Figure 12: Break & Enter Rate Moran’s I

```
	Moran I test under randomisation

data:  neighbourhoods_spdf$Robbery_Ra
weights: Toronto_listw

Moran I statistic standard deviate = 6.928, p-value = 2.134e-12
alternative hypothesis: greater
sample estimates:
Moran I statistic       Expectation          Variance
      0.320658294      -0.007194245       0.002239462
```

Figure 13: Robbery Rate Moran’s I

The results show that there does seem do be some slight correlation between the trees per square kilometer and the 4 types of crime rates however it is not completely reliable as spatial autocorrelation exists within the crime data. Due to the existence of these clusters, crime rates might not be related to tree cover. These results however can somewhat confirm previously observed phenomena from other studies such as from Portland (Donovan et al., 2010) or the greater Baltimore regions (Troy et al., 2012) which found that increased tree cover was associated with decreased crime rates. We cannot be fully sure that the results are due to clustering but there is still a chance and as such would involve further inspection and analysis into these clusters.

4) Conclusion

In conclusion, the study found that some slight relation seems to exist between the number of trees per square kilometer and crime rates (assault, auto theft, break and enter, robbery) within Toronto, Canada, however the evidence is not fully conclusive and could be the result of clusters caused by spatial autocorrelation of the crime rates data. The maps generated also showed where tree cover is high and low in density along with which neighbourhoods in Toronto have high and low rates of the different types of crimes. In order to expand on the analysis in the future, more data or information could be collected about what is causing the crime rate clusters such as societal factors as it would help to better understand the data. Furthermore, perhaps analysis could be done within the non clustered regions or split neighbourhood regions such as the east, center, and west side of the city in order to see how the relationships differ.

References

Donovan, G. H., & Prestemon, J. P. (2010). The effect of trees on crime in Portland, Oregon. *Environment and Behavior*, *44*(1), 3–30. https://doi.org/10.1177/0013916510383238

Escobedo, F. J., Clerici, N., Staudhammer, C. L., Feged-Rivadeneira, A., Bohorquez, J. C., & Tovar, G. (2018). Trees and crime in Bogota, Colombia: Is the link an ecosystem disservice or service? *Land Use Policy*, *78*, 583–592. https://doi.org/10.1016/j.landusepol.2018.07.029

Gilstad-Hayden, K., Wallace, L. R., Carroll-Scott, A., Meyer, S. R., Barbo, S., Murphy-Dunning, C., & Ickovics, J. R. (2015). Research note: Greater Tree Canopy cover is associated with lower rates of both violent and property crime in New Haven, CT. *Landscape and Urban Planning*, *143*, 248–253. https://doi.org/10.1016/j.landurbplan.2015.08.005

Kondo, M. C., Han, S. H., Donovan, G. H., & MacDonald, J. M. (2017). The association between urban trees and crime: Evidence from the spread of the emerald ash borer in Cincinnati. *Landscape and Urban Planning*, *157*, 193–199. https://doi.org/10.1016/j.landurbplan.2016.07.003

Toronto Police Services. *Neighbourhood crime rates 2020*. Public Safety Data Portal. (2021). https://data.torontopolice.on.ca/datasets/TorontoPS::neighbourhood-crime-rates-2020-1/about

City of Toronto. *Topographic Mapping – Physical location of trees*. Open Data Portal. (n.d.). https://open.toronto.ca/dataset/topographic-mapping-physical-location-of-trees/

Troy, A., Morgan Grove, J., & O'Neil-Dunne, J. (2012). The relationship between tree canopy and crime rates across an urban–rural gradient in the Greater Baltimore region. *Landscape and Urban Planning*, *106*(3), 262–270. https://doi.org/10.1016/j.landurbplan.2012.03.010

Appendix A: Workflow Documentation

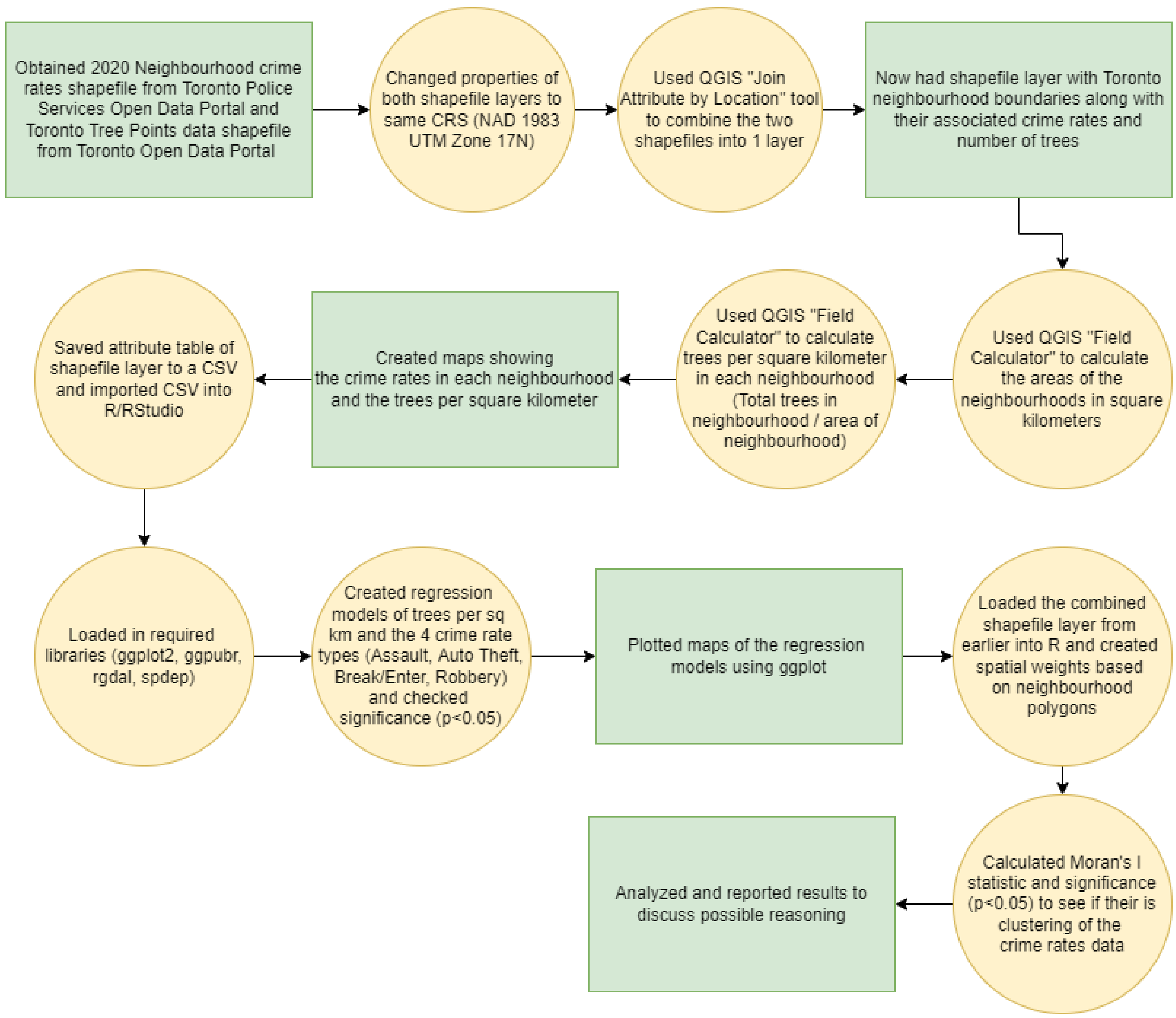


Figure 1: Workflow Diagram

Appendix B: Metadata

File Name	Description	Type	Projection	Extent	Source
Neighbourhood_Crime_Rates_2020.shp	Shape file containing Toronto neighbourhood boundaries along with their associated major crime indicators from 2014-2020	Polygons	NAD 1927 UTM Zone 17N	Meters	Toronto Police Service Open Data Portal
TOPO_TREE_WGS84.shp	Shape file representing the physical location of all trees in Toronto	Points	WGS 1984	Degrees	Toronto Open Data Portal

Table 1: Data Used

Appendix C: Evidence of Open-Source Software

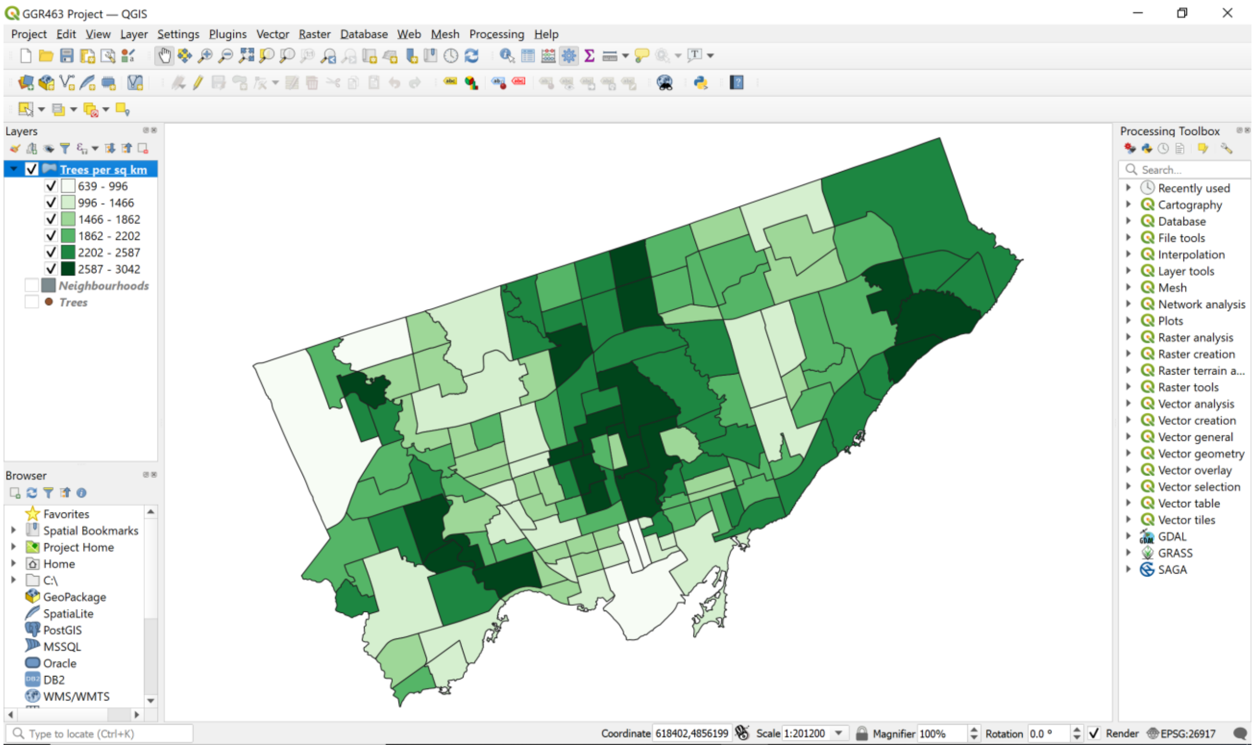
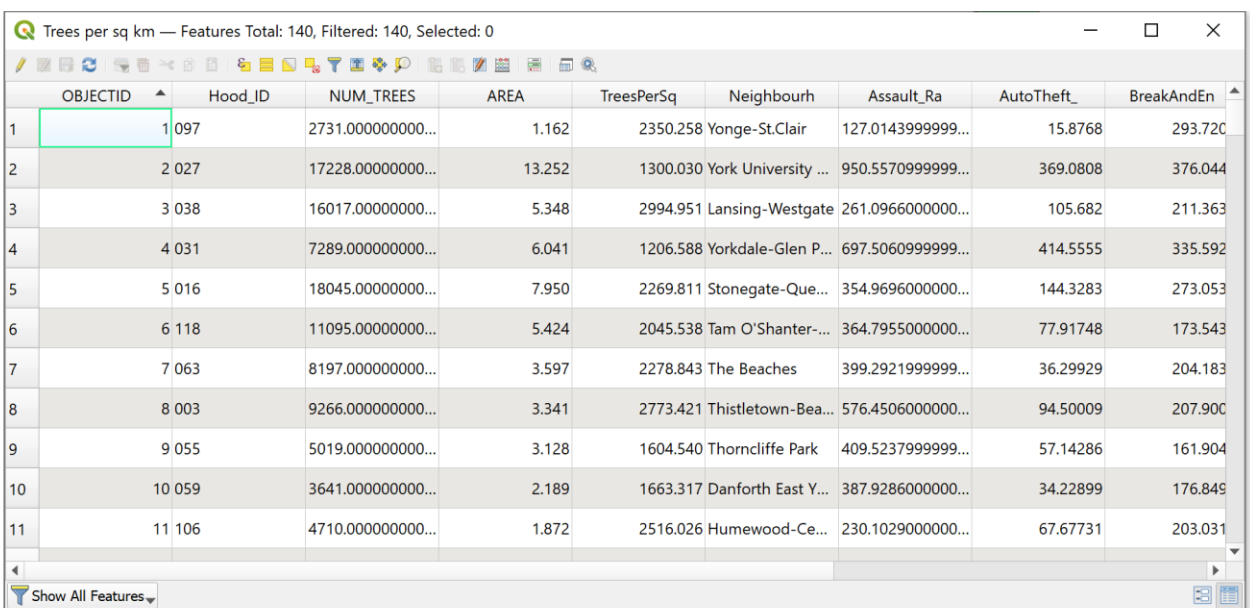


Figure 1: QGIS used to conduct spatial analysis and create maps



Trees per sq km — Features Total: 140, Filtered: 140, Selected: 0

	OBJECTID	Hood_ID	NUM_TREES	AREA	TreesPerSq	Neighbourh	Assault_Ra	AutoTheft_	BreakAndEn
1	1	097	2731.000000000...	1.162	2350.258	Yonge-St.Clair	127.0143999999...	15.8768	293.720
2	2	027	17228.00000000...	13.252	1300.030	York University ...	950.5570999999...	369.0808	376.044
3	3	038	16017.00000000...	5.348	2994.951	Lansing-Westgate	261.0966000000...	105.682	211.363
4	4	031	7289.000000000...	6.041	1206.588	Yorkdale-Glen P...	697.5060999999...	414.5555	335.592
5	5	016	18045.00000000...	7.950	2269.811	Stonegate-Que...	354.9696000000...	144.3283	273.053
6	6	118	11095.00000000...	5.424	2045.538	Tam O'Shanter-...	364.7955000000...	77.91748	173.543
7	7	063	8197.000000000...	3.597	2278.843	The Beaches	399.2921999999...	36.29929	204.183
8	8	003	9266.000000000...	3.341	2773.421	Thistletown-Bea...	576.4506000000...	94.50009	207.900
9	9	055	5019.000000000...	3.128	1604.540	Thorncliffe Park	409.5237999999...	57.14286	161.904
10	10	059	3641.000000000...	2.189	1663.317	Danforth East Y...	387.9286000000...	34.22899	176.849
11	11	106	4710.000000000...	1.872	2516.026	Humewood-Ce...	230.1029000000...	67.67731	203.031

Show All Features

Figure 2: QGIS attribute table after conducting calculations with field calculator

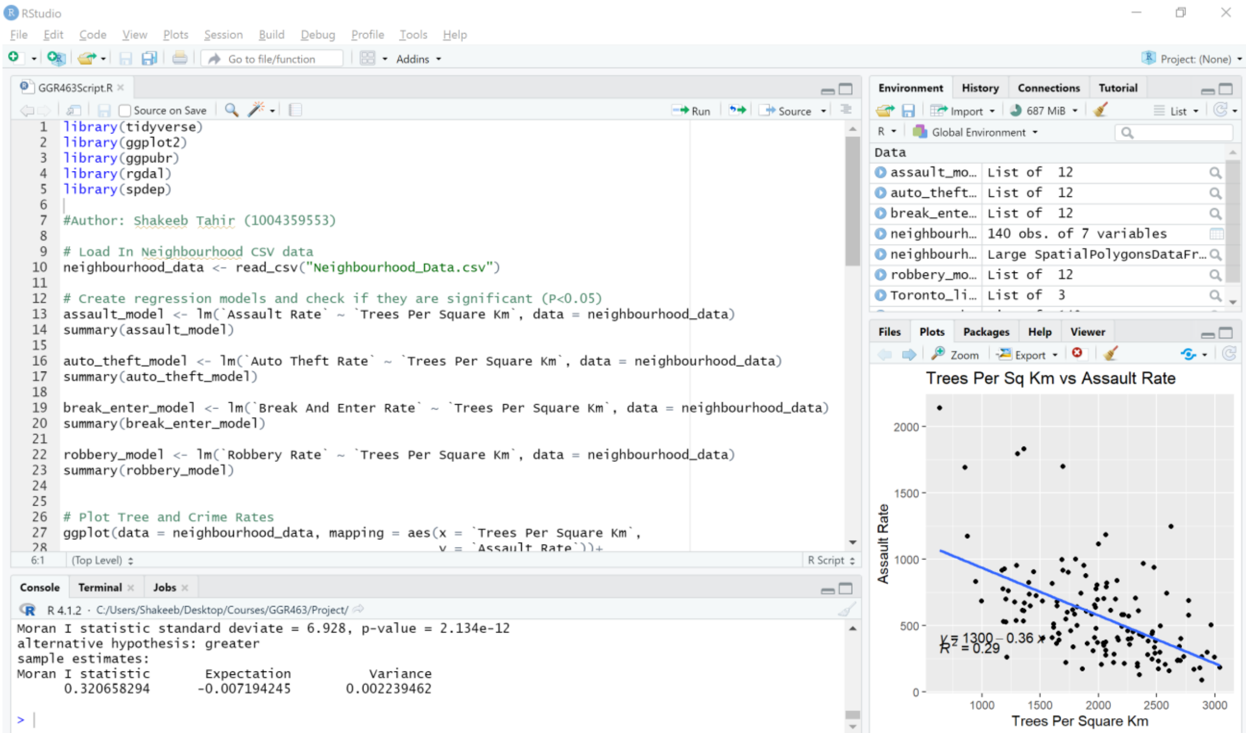


Figure 3: R/RStudio used to conduct statistical analysis (regression / Moran's I)

Appendix D: Maps and R Code

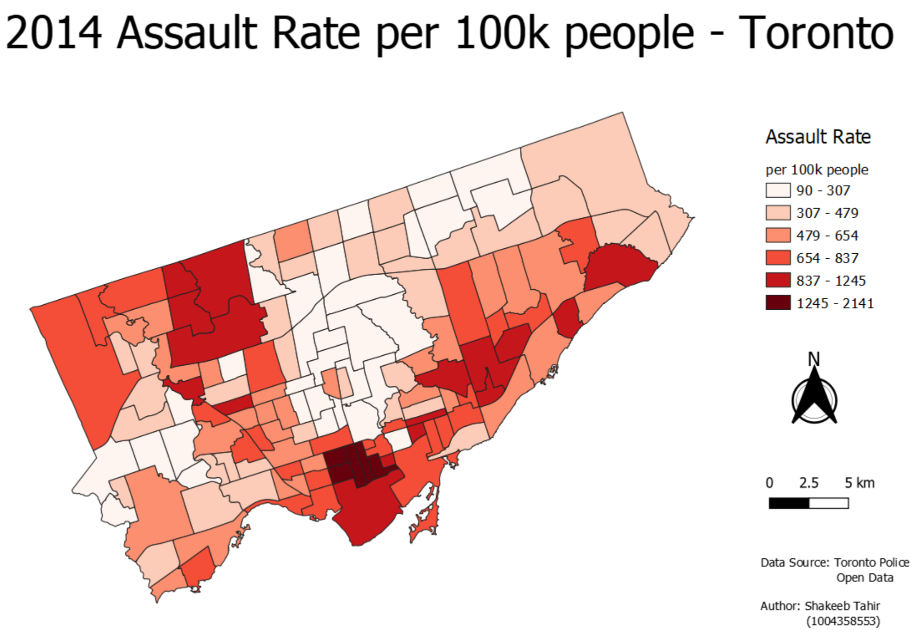


Figure 1: Assault Rate

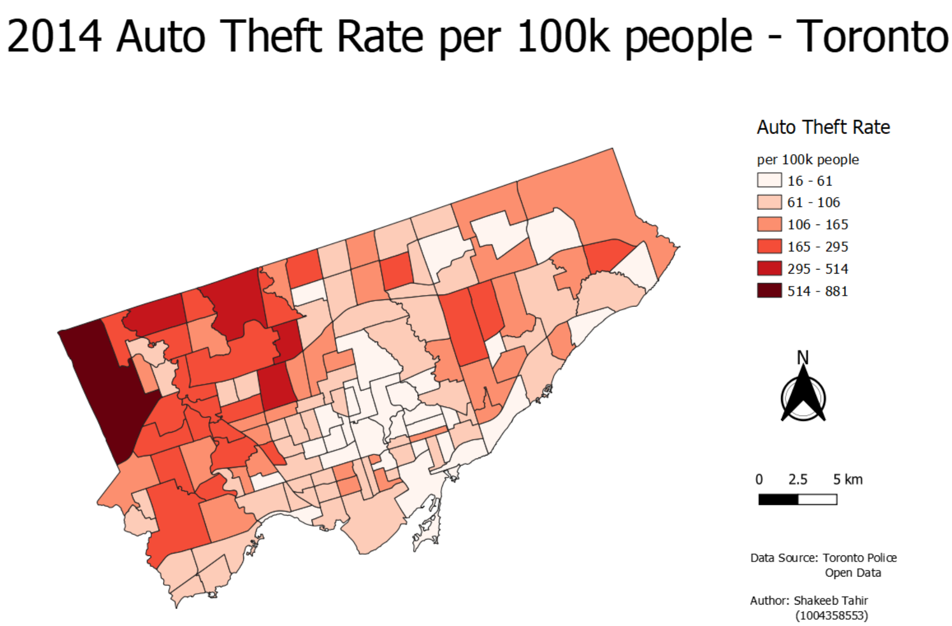


Figure 2: Auto Theft Rate

2014 Break & Enter Rate per 100k people - Toronto

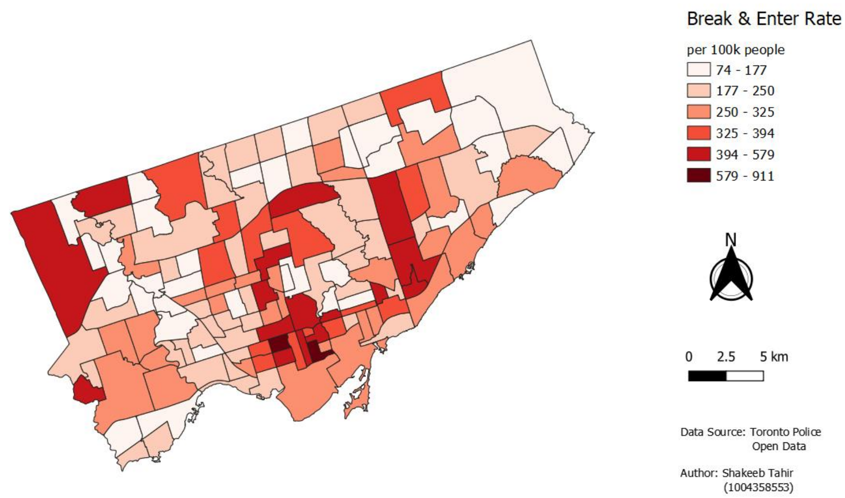


Figure 3: Break & Enter Rate

2014 Robbery Rate per 100k people - Toronto

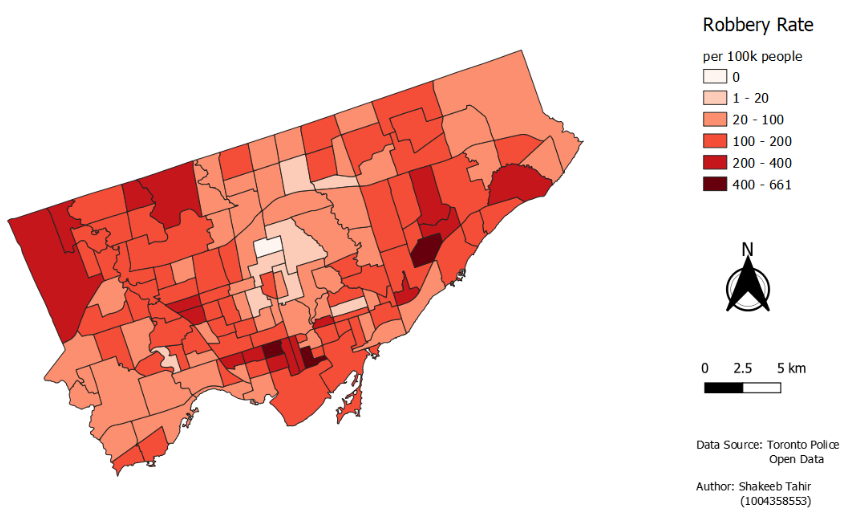


Figure 4: Robbery Rate

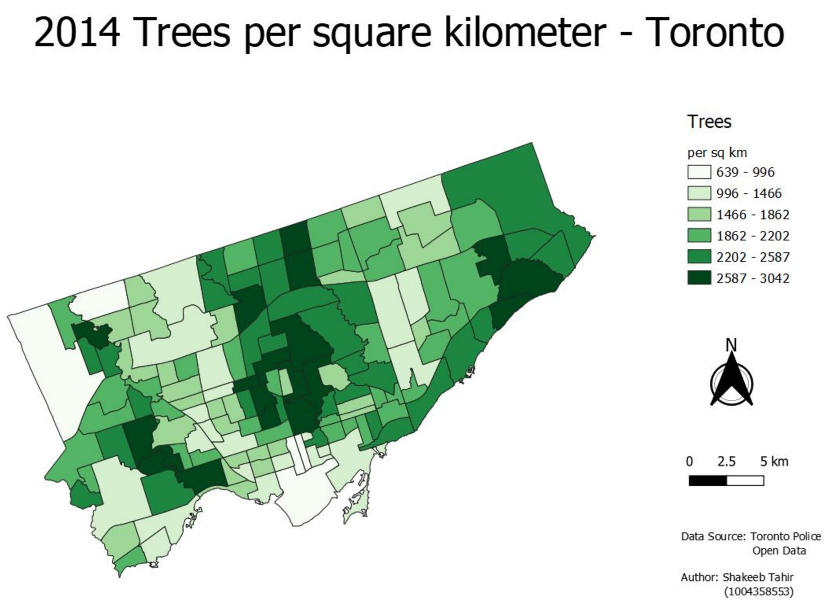


Figure 5: Trees per sq km

GGR463Script.R

Source on Save

```r
library(tidyverse)
library(ggplot2)
library(ggpubr)
library(rgdal)
library(spdep)

#Author: Shakeeb Tahir (1004359553)

# Load In Neighbourhood CSV data
neighbourhood_data <- read_csv("Neighbourhood_Data.csv")

# Create regression models and check if they are significant (P<0.05)
assault_model <- lm(`Assault Rate` ~ `Trees Per Square Km`, data = neighbourhood_data)
summary(assault_model)

auto_theft_model <- lm(`Auto Theft Rate` ~ `Trees Per Square Km`, data = neighbourhood_data)
summary(auto_theft_model)

break_enter_model <- lm(`Break And Enter Rate` ~ `Trees Per Square Km`, data = neighbourhood_data)
summary(break_enter_model)

robbery_model <- lm(`Robbery Rate` ~ `Trees Per Square Km`, data = neighbourhood_data)
summary(robbery_model)


# Plot Tree and Crime Rates
ggplot(data = neighbourhood_data, mapping = aes(x = `Trees Per Square Km`,
                                                y = `Assault Rate`))+
  geom_point() +
  geom_smooth(method = "lm", se = F) + ggtitle("Trees Per Sq Km vs Assault Rate") +
  stat_regline_equation(label.y = 400, aes(label = ..eq.label..)) +
  stat_regline_equation(label.y = 350, aes(label = ..rr.label..))


ggplot(data = neighbourhood_data, mapping = aes(x = `Trees Per Square Km`,
                                                y = `Auto Theft Rate`))+
  geom_point() +
  geom_smooth(method = "lm", se = F) + ggtitle("Trees Per Sq Km vs Auto Theft Rate") +
  stat_regline_equation(label.y = 400, aes(label = ..eq.label..)) +
  stat_regline_equation(label.y = 350, aes(label = ..rr.label..))


ggplot(data = neighbourhood_data, mapping = aes(x = `Trees Per Square Km`,
                                                y = `Break And Enter Rate`))+
  geom_point() +
  geom_smooth(method = "lm", se = F) + ggtitle("Trees Per Sq Km vs Break and Enter Rate") +
  stat_regline_equation(label.y = 400, aes(label = ..eq.label..)) +
  stat_regline_equation(label.y = 350, aes(label = ..rr.label..))


ggplot(data = neighbourhood_data, mapping = aes(x = `Trees Per Square Km`,
                                                y = `Robbery Rate`))+
  geom_point() +
  geom_smooth(method = "lm", se = F) + ggtitle("Trees Per Sq Km vs Robbery Rate") +
  stat_regline_equation(label.y = 400, aes(label = ..eq.label..)) +
  stat_regline_equation(label.y = 350, aes(label = ..rr.label..))

# Load in shapefile layer with data
neighbourhoods_spdf <- readOGR("C:/Users/Shakeeb/Desktop/Courses/GGR463/Project/Rates_w_Trees",
                               layer = "FinalLayer")

# Create a list of polygons that share borders
Toronto_nb <- poly2nb(neighbourhoods_spdf)

# Plots neighbour connections
par(mai=c(0,0,0,0))
plot(neighbourhoods_spdf)
plot(Toronto_nb, coordinates(neighbourhoods_spdf), col='red', lwd=2, add=TRUE)

# Convert Neigbours List to Spatial Weights
Toronto_listw <- nb2listw(Toronto_nb)

# Run Moran's I tests
moran.test(neighbourhoods_spdf$Assault_Ra, Toronto_listw)
moran.test(neighbourhoods_spdf$AutoTheft_, Toronto_listw)
moran.test(neighbourhoods_spdf$BreakAndEn, Toronto_listw)
moran.test(neighbourhoods_spdf$Robbery_Ra, Toronto_listw)
```

Figure 6: R Code