Are Greener Spaces Less Safe Places?*

Exploring the Surprising Relationship Between Urban Green Spaces and Crime Rates in Toronto

Jena Shah

April 19, 2024

This study examines the relationship between green spaces and crime rates in Toronto's neighborhoods using gamma regression models. The analysis reveals a positive association between the amount of green spaces and crime rates, indicating that neighborhoods with more green spaces tend to have higher crime rates. This finding emphasizes the relation between urban green spaces and public safety, suggesting that the presence of green spaces may present challenges for crime prevention and urban planning. Understanding these dynamics can help the government and urban planners in their efforts to create safer and more sustainable cities.

Table of contents

*Code and data are available at: https://github.com/shahjena/greenspace_safeplace.git.

1 Introduction

In urban environments, the presence of green spaces has long been associated with a variety of benefits: from improved mental health to enhanced community cohesion. However, the relationship between green spaces and crime rates is a complex one and has been a subject of debate in urban planning and crime prevention. There are plenty of studies suggesting that green spaces may act as a deterrent to crime through promoting community interaction and physical activity. However, there are also others who raise concerns that these areas may attract criminal activity. Understanding this relationship is crucial for creating safer and more livable cities.

National Geographic has summarized studies indicating that urban areas with more green spaces tend to have lower crime rates. For example, in Flint, Michigan - a city that experienced a decrease in crime rates from 2012 to 2022 - the implementation of green spaces was correlated with this positive change. Similarly, places like Ohio and Philadelphia have seen recommendations to increase green spaces in areas with high crime rates as a potential strategy to reduce crime (Burtka 2023). Not only that, but there is further research backing up the results discussed by National Geographic. A research paper from the Multidisciplinary Digital Publishing Institute (MDPI) conducted a comprehensive literature review to investigate the impact of green spaces on violent crime in urban settings. Their review included over 30000 potential paper titles and they ultimately selected 45 papers for inclusion. After their analysis, their findings and results suggested that green spaces play a significant role in reducing violent crime. The mechanisms proposed for this impact include social interaction, social recreation, community perception, nature-induced stress reduction, climate change, and many more (Shepley et al. 2019).

Unlike the two studies mentioned, this paper will represent the controversial side's contribution to the current understanding of urban dynamics by focusing on the relationship between the prevalence of green spaces and crime rates within Toronto's urban neighbourhoods. By utilizing a rich and expansive dataset along with a sophisticated statistical model, we conduct a thorough analysis to unravel the complexities of this relationship. In doing so, the findings not only focus on the interconnections between green spaces and crime rates, but also offer insights that can significantly inform the formulation of urban planning strategies for crime prevention. The estimand for this study is to quantify the effect of an increase in green spaces on crime rates in urban neighbourhoods - while controlling for relevant factors.

The paper is structured into four main sections. First, I provide a detailed overview of the data utilized in this study, including its sources and any pre-processing/cleaning steps. I also describe the methodology used for the analysis, including the statistical figures made to explore the relationship between green spaces and crime rates. After that, I present the modeling approach used in this study - detailing the statistical techniques applied to analyze the data. This section includes a discussion of the model setup and justification for the chosen modeling approach. Following that, I present the results of my analysis, highlighting the key findings regarding the relationship between green spaces and crime rates in urban neighbourhoods. I use visualizations such as graphs and tables to illustrate these results effectively. Finally, the implications of my findings are discussed for urban planning and crime prevention. I explore how these results can better the policies which are aimed at creating safer and more sustainable urban environments. I also take a look at the weaknesses of my study which could bias the results. Throughout the paper, I utilized the R programming language and its packages which are cited in the references.

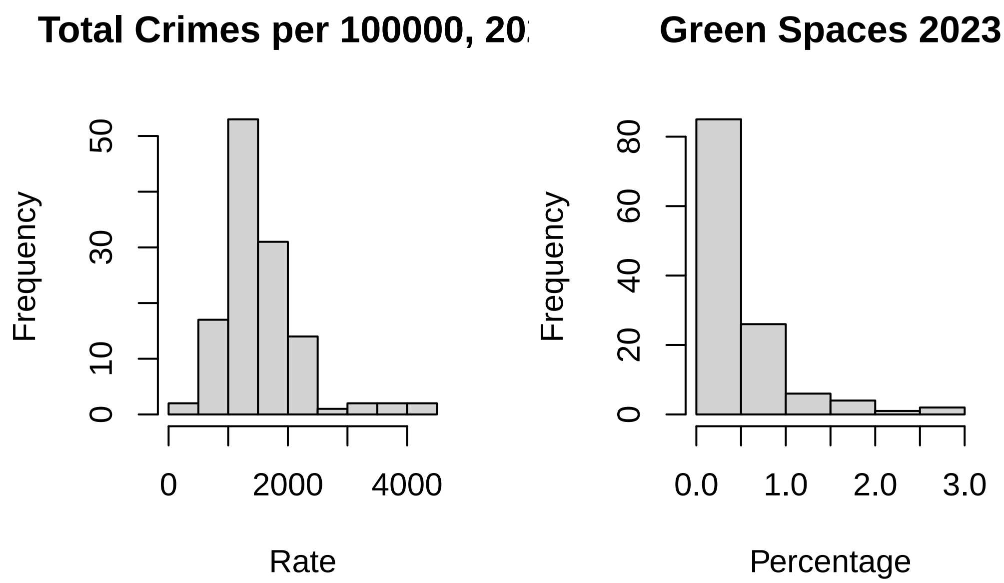


Figure 1: Distribution of Green.Spaces and TOTAL_CRIMES.

2 Data

2.1 Data Collection and Cleaning

The datasets for this project were obtained from the City of Toronto's open data portal. There were many other sources that could provide similar data but they were much older, hence it was more appropriate to use the data from Open Data Toronto's portal. The first dataset, "Neighbourhood Crime Rates," contains all neighbourhoods in Toronto and gives the rates (out of 100000) of nine different types of crime from the years 2014 to 2023. Example of crimes include, Assault, Auto Theft, Break and Enter, Robbery, Theft Over $5000, and more. The second dataset, "Wellbeing Toronto Environment," contains information on green spaces and environmental factors. This dataset was provided by Toronto Public Health, Toronto Parks, Forestry and Recreation, the Federal Ministry of the Environment, and Toronto Water.

It includes data on Green Rebate Programs, Pollutant Carcinogenic TEP Score, Pollutant Non-Carcinogenic TEP Score, and Pollutants Released to Air, among other variables.

The datasets were combined using the neighbourhood ID (HOOD_ID) and filtered for the variables of interest, which include the neighbourhood ID, neighbourhood name (AREA_NAME), and crime rates for Assault, Auto Theft, Break and Enter, Robbery, and Theft Over for the year 2023. Additionally, the dataset includes information on Green Spaces for each neighbourhood. Since the data of crime across Toronto did not provide a total amount of crimes per neighbourhood and only focused on the different types of crime and their rates, I created a new variable which combined all the different crimes and merged it into a variable called TOTAL_CRIMES which provides the total amount of crimes per 100000 people in a neighbourhood.

The decision to focus on the year 2023 out of all the years from 2014 to 2023 was made to use the most recent data available. In the crimes data, there were a total of 9 types of crimes used but this paper focuses on using the top 5 crimes in Toronto in the year of 2023 according to Vilkhov Law. They were selected for analysis and include Assault, Auto Theft, Break and Enter, Robbery, and Theft Over (Unknown 2024). Other years and additional crimes were available in the raw crime data, but for the purposes of this study, only the year 2023 and the selected top 5 crimes were considered.

Overall, the data collection and pre-processing steps involved obtaining, combining, and filtering the datasets to create a consolidated dataset for analysis. These steps ensured that the data used in the study were relevant, accurate, and suitable for investigating the relationship between green spaces and crime rates in Toronto's neighbourhoods.

2.2 Data Variables

The dataset for this study comprises four key variables that are central to the analysis. These variables are properly descibed below.

1. HOOD_ID (Neighbourhood ID): This variable serves as a unique identifier for each neighborhood in Toronto, allowing us to differentiate between different areas in our analysis.
2. AREA_NAME (Neighbourhood Name): This variable provides the name of each neighborhood, providing context and a more intuitive understanding of the data.
3. Green.Spaces (Amount of Green Spaces): This variable quantifies the amount of green spaces in each neighborhood, measured in square kilometers. Green spaces include parks, forests, and other areas covered by vegetation.
4. TOTAL_CRIMES (Total Crimes per Neighbourhood per 100000 People): This variable represents the total number of reported crimes per neighborhood, standardized to a

rate per 100000 people. It provides a measure of the overall crime burden in each neighborhood, accounting for differences in population size.

These variables were selected based on their relevance to the study's objective of examining the relationship between green spaces and crime rates. The inclusion of the neighborhood ID and name allows us to identify and differentiate between neighborhoods, while the Green.Spaces variable provides a measure of the presence and extent of green spaces in each area. The TOTAL_CRIMES variable provides a measure of the overall crime rates, allowing us to assess the impact of green spaces on crime in different neighborhoods. Figure 1 shows the different variables and we can see the right-skew in the crimes total, which can help us decide what regression model to use in the next step.

3 Model

3.1 Model Set-up

This study was done to investigate the relationship between the abundance of green spaces and crime rates in urban neighborhoods in Toronto by using a gamma regression model. This approach is suitable for analyzing count data with a skewed distribution, such as crime rates, where the response variable is non-negative and continuous.

3.1.1 Variables

The key variables used in the model are:

- Dependent Variable: TOTAL_CRIMES (total crimes per neighborhood per 100000 people).
- Independent Variable: Green.Spaces (amount of green spaces in square kilometers).

3.1.2 Model Details

The gamma regression model is specified as follows:

$log(E(Y)) = \beta_0 + \beta_1 \times \text{Green.Spaces}$

Where:

- (E(Y)) is the expected value of the dependent variable (TOTAL_CRIMES).
- β_0 represents the intercept term.
- β_1 is the coefficient associated with the Green.Spaces variable, indicating the effect of green spaces on crime rates.

3.1.3 Assumptions

The gamma regression model assumes that the dependent variable follows a gamma distribution. We also assume that the relationship between green spaces and crime rates is linear, meaning that a one-unit increase in green spaces is associated with a constant percentage change in crime rates.

3.2 Model Justification

The choice of a gamma regression model is appropriate for this study because it is specifically designed for analyzing skewed count data. We noticed the right-skew in the histograms of green spaces and total crimes in Figure 1. By using a gamma regression approach, we can account for the skewed nature of crime rate data and obtain more reliable estimates of the relationship between green spaces and crime rates.

Overall, the gamma regression model is well-suited for analyzing the relationship between green spaces and crime rates in Toronto's neighborhoods, providing a statistically interpretable approach to examining this important issue.

4 Results

In the results, I found that green spaces have a statistically significant relationship with crime rates in Toronto's neighborhoods. The gamma regression model showed that as the amount of green spaces in a neighborhood increases, the total crimes per 100000 people also tend to increase. Specifically, for every unit increase in green spaces (measured in square kilometers), there was a corresponding increase in total crimes.

(**tab-one?**) presents the results of the explanatory model for crime rates based on green spaces. The intercept and coefficient for green spaces are both statistically significant, indicating that green spaces have a significant impact on crime rates. The model had 122 observations, and the log-likelihood was -943.227. The model's predictive performance was assessed using ELPD, LOOIC, WAIC, and RMSE, indicating a good fit of the model to the data.

Figure 1 illustrates this relationship with a scatterplot. It shows a concentration of data points at low to medium crime rates and low levels of green spaces. As the amount of green spaces increases, there are much fewer data points and they are more scattered, indicating higher crime rates in neighborhoods with more green spaces. The line of best fit in the scatterplot shows a nearly straight line, yet still highlighting the positive relationship between green spaces and crime rates.

Figure 2 enhances this by presenting a line graph that shows the uncertainty shadow of the line of best fit. It demonstrates that the shadow is small at low to medium crime rates and low levels of green spaces, indicating higher certainty in the relationship. However, as the amount of green spaces increases, the shadow expands a lot, indicating more uncertainty in the relationship.

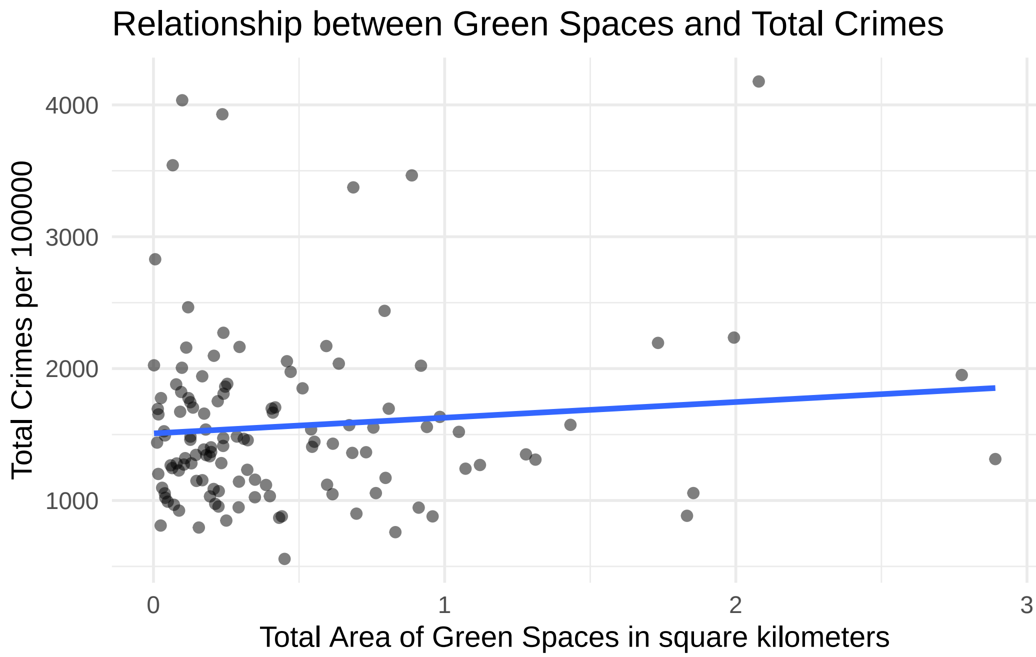


Figure 2: Crime rate (out of 100000) frequency depending on the area of green spaces (square kilometers)

5 Discussion

In this study, I analyzed the relationship between green spaces and crime rates in Toronto's neighborhoods using gamma regression models. The results suggest a positive relationship between the amount of green spaces and crime rates, indicating that areas with more green spaces tend to have higher crime rates. However, it is important to note that this relationship does not imply causation.

One of the key insights gained from this study is the potential impact of green spaces on crime rates in urban areas. The findings suggest that neighborhoods with more green spaces may face unique challenges in terms of crime prevention and urban planning. This knowledge can inform policymakers and urban planners in their efforts to create safer and more sustainable cities.

Table 1: Explanatory models of the crime rates based on the green spaces

	Model
(Intercept)	0.00
	(0.00)
Green.Spaces	0.00
	(0.00)
Num.Obs.	122
Log.Lik.	−943.227
ELPD	−946.4
ELPD s.e.	11.4
LOOIC	1892.8
LOOIC s.e.	22.9
WAIC	1892.7
RMSE	647.58

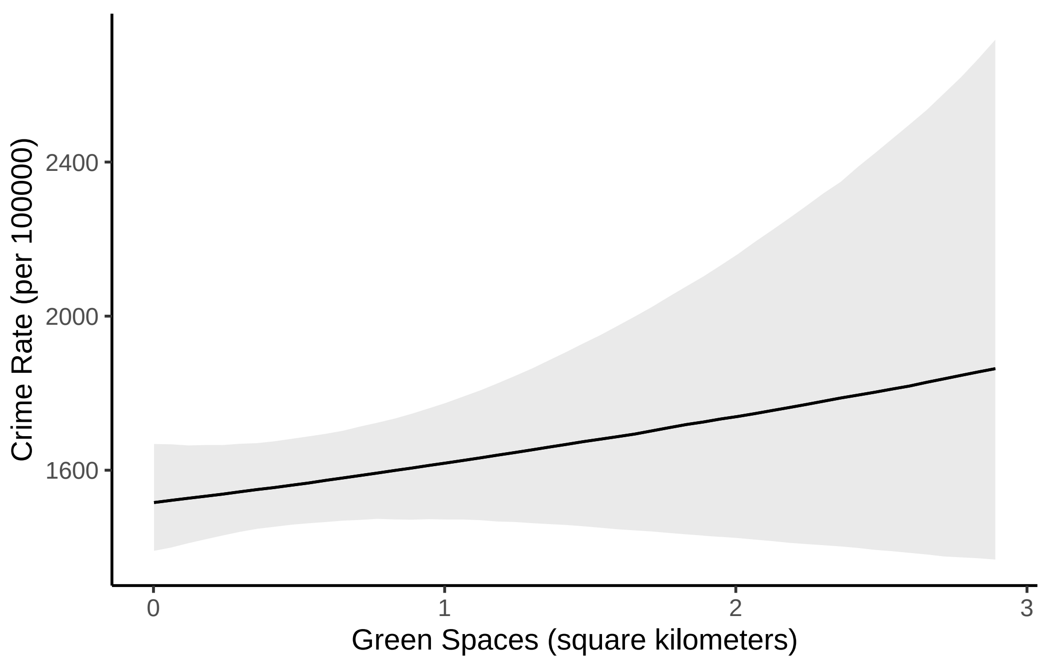


Figure 3: The number of crimes predicted by the model for the area of green spaces per neighbourhood

5.1 Weaknesses and Next Steps

Several limitations in this study should be acknowledged. Firstly, the green space data used in this analysis is from 2011, which may not accurately reflect the current green space distribution in Toronto. Additionally, the process of reducing the number of neighborhoods from 158 to 138 due to the removal the NA and 0 values could introduce bias into the analysis (so a total of 20 nighbourhoods were not considered in this study).

To deepen the understanding of relationships between green spaces and crime rates, future research should focus on using more recent green space data and exploring other factors that may influence crime rates in urban areas. For instance, economic status, social factor, and other demographics relating to the different neighbourhoods. Additionally, conducting similar studies in other cities instead of Toronto could help to determine the general idea behind these findings.

5.2 Conclusion

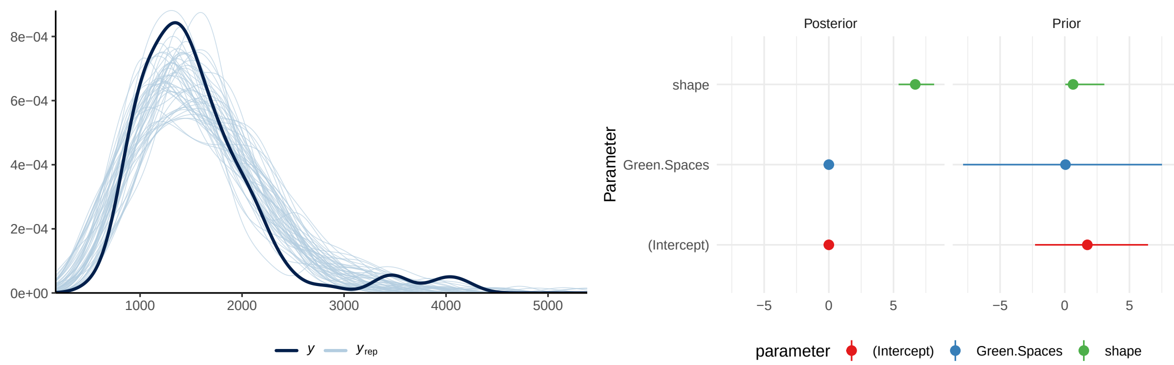
In conclusion, this study provides details about the relationship between green spaces and crime rates in urban areas. By understanding the impact of green spaces on crime rates, the policymakers, government and urban planners can work towards creating even more safe and sustainable cities for their residents.

6 Appendix

6.1 Datasheet

6.2 Posterior predictive check

This check is a way to assess the goodness-of-fit of a model by comparing its predictions to actual observations. (a) Peaks in Figure 4 check around 1000-1500 indicate areas where the model may not accurately capture the data's variability or patterns. (b) The posterior estimates are more in the middle, indicating greater certainty after adding in the data, whereas prior estimates are more dispersed, reflecting greater uncertainty before observing the data.



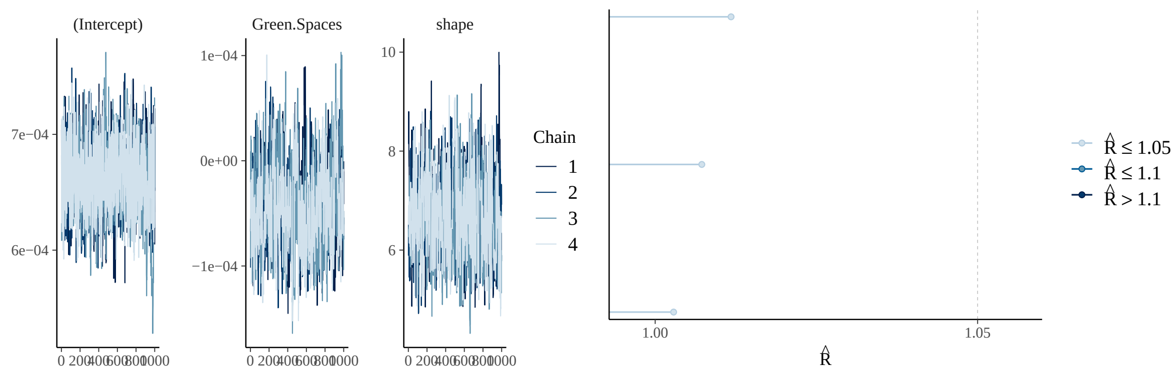
(a) Posterior prediction check

(b) Comparing posterior with prior

Figure 4: Examining how the model fits and is affected by the data

6.3 Diagnostics

The diagnostics help assess whether the model has sampled from the posterior distribution effectively. In Figure 5 (a) The trace plot shows that chain 4 appears to have higher variability than the others. (b) The Rhat plot displays values around 1 indicating convergence of the chains.



(a) Trace plot (b) Rhat plot

Figure 5: Checking the convergence of the MCMC algorithm

7 References

R Core Team (2023) Gelfand (2022) Burtka (2023) Shepley et al. (2019) Unknown (2024)

Burtka, Allison Torres. 2023. "One Unexpected Way to Reduce Violent Crime? Create Green Spaces." *National Geographic*. https://www.nationalgeographic.com/environment/article/urban-greening-violent-crime-mental-health-nature.

Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal*. https://CRAN.R-project.org/package=opendatatoronto.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Shepley, Mardelle, Naomi Sachs, Hessam Sadatsafavi, Christine Fournier, and Kati Peditto. 2019. "The Impact of Green Space on Violent Crime in Urban Environments: An Evidence Synthesis." *National Library of Medicine*. https://doi.org/10.3390/ijerph16245119.

Unknown. 2024. *Toronto Crime Rates in 2023*. Vilkhov Law. https://vilkhovlaw.ca/toronto-crime-rates/.