Analysis Report

Jiyeon Song

Introduction

The goal of this study is to investigate the impact of greening sites on the incidence of assault crimes involving firearms by employing a spatial-temporal modeling approach. By analyzing paired greening and control sites, we aim to assess the association between the presence of greening sites and the occurrence of such crimes, contributing to our understanding of the potential role of greening initiatives in crime prevention strategies.

Data

This study utilizes three distinct datasets. Firstly, crime incident data for Indianapolis, IN was obtained from the Indianapolis Metropolitan Police Department. This dataset includes details on the type, location, and time of crimes committed in the city and covers the period from 2010 to 2021. We focus on crimes related to firearms, such as Part I violent crime, assault with a firearm, and homicide. Secondly, we have compiled a list of 58 greenspaces and 110 control spaces, each with recorded latitude and longitude coordinates. Additionally, we have collected data on neighborhood characteristics, including poverty rate, percentage of non-white population, and renters. Lastly, our study incorporates several social vulnerability-related variables that may impact crime rates, such as the percentage of households living below the poverty line, households with no vehicles, and households eligible for food stamps, to account for the potential impact of social vulnerability on crime rates, and to assess the equity implications of greening projects.

Preliminary Analysis

Our preliminary analysis indicates fluctuations in the number and types of crimes over the years. Notably, there has been a significant surge in gun-related crimes between 2019 and 2021 (Table 1). In addition, since the accepted greenspaces and control spaces are in close proximity and on irregular locations (Figure 1), failing to account for their spatial locations may lead to misleading conclusions about the impact of greening initiatives. Furthermore, our comparison of social vulnerability conditions between greenspaces and rejected spaces highlights significant disparities such as the proportion of the population over 65 years old and the non-white population (Table 2), emphasizing the need to fully comprehend the relationships between greening initiatives, crime rates, and social vulnerability. This understanding is crucial for the design and implementation of effective public health interventions.

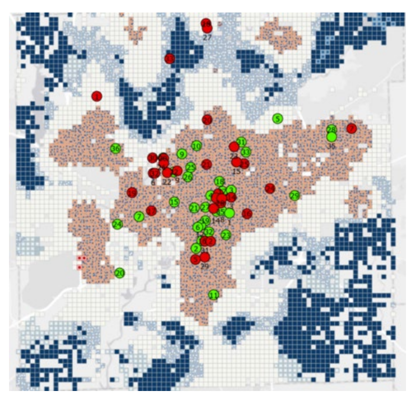

Figure 1. Distribution of Greenspaces (Green) and Rejected Sites (Red)

Table 1. Summary by Crime Type

Year	P1 ALL	P1 Violent	P1 Property	Assaults	Assaults w/ Firearm	Homicides
2010	50,651	8,830	41,821	5,300	880	86
2011	51,653	8,321	43,332	4,810	981	89
2012	52,825	8,066	44,759	4,253	1,367	92
2013	52,617	10,016	42,600	5,667	1,377	121
2014	50,874	10,470	40,404	6,103	1,620	129
2015	51,664	11,131	40,532	6,663	2,027	141
2016	49,887	11,266	38,618	6,704	2,080	146
2017	46,448	10,797	35,645	6,678	2,043	149
2018	44,651	10,550	34,101	6,819	2,232	156
2019	38,596	8,546	30,042	5,305	2,205	137
2020	39,533	9,545	29,977	6,611	3,718	209
2021	38,639	9,225	29,410	6,630	4,112	245

Table 2. comparisons of social vulnerability conditions between the control and treatment groups

Variable: Mean (SD)	Control	Greening
Percent of HH below 200% poverty level	0.52 (0.20)	0.55 (0.20)
Percent of HH with no vehicle	0.18 (0.10)	0.18 (0.09)
Percent of population underage 18	21.08 (8.08)	22.69 (8.99)
Percent of population over age 65	10.34 (3.75)	10.30 (5.28)
Percent of population with a disability	16.88 (4.96)	16.90 (4.68)
Percent of HH with limited English	2.17 (2.73)	2.67 (3.38)
Percent of non-white population	42.84 (26.70)	39.63 (23.85)
Percent of HH eligible for food stamps	23.71 (12.70)	27.07 (12.60)
Percent of population that are renters	55.90 (14.45)	55.27 (14.92)
Homeowners who pay more than 30% of their income on mortgage payments	29.83 (10.15)	31.22 (10.04)
Population with no high school diplomas	18.86 (10.60)	22.55 (11.13)

Method

We initially formed 36 pairs of greening sites and control sites that exhibited similar background characteristics in terms of population. This pairing was essential to ensure comparability and minimize confounding factors, thus allowing us to observe the specific impact of greening efforts. For the spatial-temporal analysis, the Integrated Nested Laplace Approximation (INLA) method with the stochastic partial differential equation approach (SPDE) were employed. INLA is a powerful Bayesian approach that efficiently models spatial and temporal dependencies in the data.

In this study, we used the INLA model to analyze the paired greening and control sites with the Poisson regression framework. The outcome variable in the Poisson regression model was the count of crimes, allowing us to assess the relationship between the presence of greening sites and crime incidence. To address spatial autocorrelation, the INLA model incorporated an autoregressive term of order 1 (AR(1)). Furthermore, the population density was added to the model as an offset term, effectively adjusting for the underlying population at risk. For evaluation, we measured the Deviance Information Criterion (DIC), a Bayesian measure analogous to the Akaike Information Criterion (AIC). However, it is important to note that INLA does not provide p-values for assessing the significance of individual predictors. Instead, one can infer significance by examining the overlap between the 2.5% and 97.5% posterior estimates of the predictors with zero.

Result

The INLA mesh (See Figure 2) provides a visual representation of the spatial structure and discretization of the model. The mesh divides the study area into a set of small polygons or cells, which are used to represent the spatial domain. By examining the mesh, one can identify spatial patterns and trends, such as consistent increases or decreases in the variable across neighboring mesh cells.

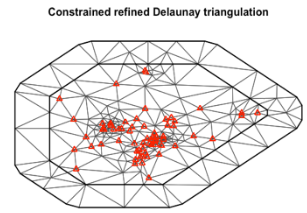

Figure 2. INLA mesh *Distribution of Sties*

Table 3. Coefficient and Confidence Intervals of the Covariates in the Poisson Regression: Assaults with Firearm Count

Variable	Mean	Sd	ExpBeta	0.025quant (Lower)	0.975quant (Upper)	DIC
Intercept	-2.473	0.393	0.084	-3.247	-1.702	12056.08
Group_greenspace	**-0.129**	**0.035**	**0.879**	**-0.197**	**-0.061**	
Percent of HH below 200% poverty level	1.766	0.408	5.850	0.967	2.566	
Percent of HH with no vehicle	4.609	0.491	100.429	3.646	5.573	
Percent of population underage 18	-0.010	0.006	0.990	-0.023	0.002	
Percent of population over age 65	0.017	0.009	1.018	-0.001	0.036	
Percent of population with a disability	-0.029	0.009	0.971	-0.046	-0.013	
Percent of HH with limited English	0.043	0.013	1.044	0.017	0.069	
Percent of non-white population	0.001	0.002	1.001	-0.003	0.004	
Percent of HH eligible for food stamps	0.021	0.005	1.021	0.010	0.032	
Percent of population that are renters	-0.036	0.004	0.965	-0.044	-0.028	
Homeowners who pay more than 30% of their income on mortgage payments	0.015	0.003	1.015	0.010	0.021	
Population with no high school diplomas	-0.014	0.006	0.986	-0.026	-0.002	

Table 3 presents the estimated coefficient of the Poisson regression model, where the count of assault crimes involving firearms serves as the outcome variable. The coefficient for the "Group_greenspace" variable was estimated to be -0.129. This indicates that, when controlling for the effects of other variables, greenspace sites are associated with a statistically significant decrease of approximately **12.1%** (1 - 0.879 = 0.121) in the expected assault gun count. The negative coefficient suggests that the greenspace within the studied areas is linked to a reduced likelihood of crime occurrence.

However, if we fit the model for the total crime count by summing all types of crime, we see the greenspace effect is insignificant. Table 4 shows the estimated coefficient of the Poisson regression when the total crime count is used as an outcome variable.

Table 4. Coefficient and Confidence Intervals of the Covariates in the Poisson Regression: Total Crime

Variable	Mean	Sd	ExpBeta	0.025quant (Lower)	0.975quant (Upper)	DIC
Intercept	0.519	0.896	1.680	-1.280	2.285	21290.82
Group_greenspace	-0.042	0.054	0.959	-0.148	0.065	
Percent of HH below 200% poverty level	-0.993	0.590	0.370	-2.157	0.166	
Percent of HH with no vehicle	1.466	0.873	4.332	-0.255	3.182	
Percent of population underage 18	0.001	0.009	1.001	-0.016	0.019	
Percent of population over age 65	0.006	0.014	1.006	-0.023	0.034	
Percent of population with a disability	-0.010	0.012	0.990	-0.035	0.014	
Percent of HH with limited English	0.094	0.026	1.099	0.043	0.144	
Percent of non-white population	-0.006	0.003	0.994	-0.012	0.001	
Percent of HH eligible for food stamps	0.037	0.009	1.038	0.020	0.054	
Percent of population that are renters	-0.003	0.007	0.997	-0.018	0.012	
Homeowners who pay more than 30% of their income on mortgage payments	0.002	0.004	1.002	-0.007	0.010	
Population with no high school diplomas	-0.035	0.011	0.966	-0.056	-0.014	

Discussion

Reference

Appendix

1. Modeling

Spatial or spatio-temporal models are commonly employed in various domains, such as crime or disease mapping analysis. The principal objective of analyzing such data is to effectively smooth and forecast the temporal progression of specific response variables within a designated spatial domain. Assessing the security of a neighborhood often hinges on comprehending the geographical (or spatial) fluctuations of crime incidence. Spatial data can be defined as realizations of a stochastic process that is indexed by space.

$$Y(s) \equiv \{y(s), s \in \mathcal{D} \in \mathbb{R}^2\}$$

The actual data can be then represented by a collection of observations $\mathbf{y} = \{y(s_1), \dots, y(s_n)\}$, where the set $(s_1, \dots, s_n)$ indicates the spatial units at which the measurements are taken. The concept of spatial process can be extended to the spatio-temporal case, including a time dimension. The data are then defined by a process

$$Y(s,t) \equiv \{y(s,t), (s,t) \in \mathcal{D} \in \mathbb{R}^2 \times \mathbb{R}\}$$

and are observed at n spatial locations or areas and at T time points. Consequently, a comprehensive approach to addressing this issue involves modeling the mean for each unit (i-th) through an additive linear predictor, which is defined on a suitable scale.

$$\eta_i = \alpha + \sum_{m=1}^{M} \beta_m \, x_{mi} + \sum_{l=1}^{L} f_l\,(z_{li}), \quad i = 1, \cdots, n.$$

By varying upon the form of the functions $f_l(\cdot)$, this formulation can accommodate a wide range of models, from standard and hierarchical regression to spatial and spatio-temporal models.