Spatial Patterns in Crime across Chicago Community Areas from 2018 to 2022

Amy Hilla
Computer Science
Virginia Tech
Falls Church, VA, USA
ahilla@vt.edu

Harditya Sarvaiya
Computer Science
Virginia Tech
Falls Church, VA, USA
harditya@vt.edu

ABSTRACT

In this research, our objective is to understand the spatial correlation and effects of economic and social factors linked to crime within community areas in Chicago over a five-year study period spanning from 2018 to 2022. Previous work suggests numerous factors that can influence and be influence by crime in different locations, but we examine nine specifically: total crime, crime density, per capita income, percent of population without a high school diploma, percent of households in poverty, percent unemployed, percent of households with overcrowding, percent of population under 18 or over 65, and hardship index. We use Ordinary Least Squares, Random Forest, and Moran's I to examine the data, as well as Getis-Ord Gi* Hot Spot Analysis for exploratory data analysis. We include seven different subsets of the data, to examine how the models performed on different sub-groups of crime incidents. We find that narcotics crimes and theft crimes may be more spatially autocorrelated, and that percent of community area population under 18 or over 65 is spatially autocorrelated and associated with increased crime density. These results could be used by policy makers and law enforcement decision makers in Chicago to enact more effective policing strategies and more effective social outreach resources.

KEYWORDS

Crime, Chicago, socioeconomic, hardship, Moran's I, geospatial

ACM Reference format:

Amy Hilla and Harditya Sarvaiya. 2023. Spatial Patterns in Crime across Chicago Community Areas from 2018 to 2022.

1 Introduction and Literature Review

Understanding what variables correlate with crime incidence is a critical effort for cities worldwide. By knowing trends and associations between crime and often seemingly unrelated features, city officials can make better informed choices about where to direct resources. It is easy to think that knowing what factors correlate with crime is intuitive: we think that crime is associated with nighttime, with "bad" neighborhoods, and with shady alleyways. But a more data driven approach to understanding these correlations is necessary to derive truly useful conclusions. In this paper, we do not seek to predict the causes of crime, but to highlight some trends: what economic, environmental, or social factors tend to be co-located with places where crimes occurred?

Previous work on this issue has focused on everything from graffiti to the weather (Bloch 2020, Towers et al 2018), suggesting that there could potentially be infinite factors that correlate with crime, depending on time and place. But in Chicago specifically, some narrower patterns emerge. Homicides tend to cluster in the south and west parts of the city, which also tend to be lower income (Bolter 2022). Research comparing crime in downtown versus in suburbs of Chicago suggests that the location of retail and manufacturing activities may play a role in where violent crime versus property crime occurs (Brown 2016). Another paper finds spatial correlation between rates of crime in Chicago's community areas and rates of gonorrhea and chlamydia (Marotta 2017). Clearly, the elements that influence crime rates and placement are complex and cannot be easily untangled through a single analysis. However, despite this complexity, we chose seven variables to look at more closely in the hopes of shedding light on some of the puzzle pieces that may help decision makers understand the question of where crime occurs, slightly better.

We chose to examine the associations between level of income, level of education, percent of households below the poverty line, rate of unemployment, rate of housing overcrowding, time of day, and presence of a nearby police station to incidence of crime across Chicago community areas. Our hypothesis that associations exist between the socioeconomic variables (income, education, poverty, unemployment, and overcrowding), and crime is supported by the literature, which suggests that crime clusters with variables that connote disadvantage and hardship, though some research also suggests that these associations may not be due to direct relationships and instead due to interaction effects (Stretesky, Shuck, & Hogan 2006, Hooghe et al. 2010).

We chose to study these associations at the community area level. The city of Chicago is divided into 77 community areas. These community areas are formed based on close neighborhoods, although some community areas encompass multiple neighborhoods. The community areas were first designed and published in 1920, by sociologists Robert Park and Ernest Burgess, and have remained relatively stable since then, making them an ideal unit of measurement for analysis over time (Burgess & Newcomb 1920). Using community areas was also ideal because of processing limitations. Trying to analyze the spatial distribution of individual crime incidents, or crime incidents across smaller geographic units such as census tract, led to extremely slow processing speeds due to the size of the data. For example, when trying to geo-reference each crime incident to its corresponding census tract, using the latitude and longitude reported for each crime incident and utilizing the censusgeocode package in Python (PyPi), the code was only able to process the census tracts for a 5% sample of the dataset, which took over four hours. At that rate, getting the census tract for each crime incident in the full 5-year study period would have taken approximately 50 hours or more. Thus, in the interest of practicality, we summed the crime incidents to the community area level before running our models.

2 Methods

Subsets of the data considered (for each year)
Only crime incidents occurring during the day
Only crime incidents occurring at night
Only crime incidents concerning narcotics
Only violent crime incidents
Only non-violent crime incidents
Only theft incidents
Only burglary incidents

Analytic methods used
Getis-Ord Gi* Hotspot Analysis
Moran's I
OLS with Rook weights
Random Forest

Variables of interest
Total count of crime occurring each year
Crime density (total crime / population) in the community area each year
Average income per capita in the community area*[1]
Percent of population 25+ without high school diploma in the community area*
Percent of households below the poverty line in the community area*
Percent of population unemployed in the community area*
Percent of households experiencing overcrowding in the community area *
Percent of population under 18 or over 65 in the community area
Hardship index**[2]

[1] * As of 2020 census

[2] ** The Hardship Index is a composite score reflecting hardship in the community (higher values indicate greater hardship). It incorporates unemployment, age dependency, education, per capita income, crowded housing, and poverty into a single score that allows comparison between geographies (Chicago Health Atlas, Nathan & Adams 1976).

Much of the prior research in this space has used Moran's I tests to determine spatial autocorrelation (Moran 1950, Marotta 2017, Anselin et al. 2000). Moran's I measures the spatial autocorrelation of values, characterized by a correlation in some kind of other value spread across nearby locations. Spatial correlation allows for multiple dimensions and directions of space. Moran's I takes a set of features (in our case, the Chicago community areas) and evaluates whether their values (in our case, each community area-level variable) are distributed in a pattern that is clustered, dispersed, or random. The Global Moran's I statistic indicates how spatially autocorrelated each variable is across the input features, but it is an inferential statistic, so we must also calculate statistical significance to interpret how meaningful it may be, using the z-score and p-value. Mathematically, it is:

$$I = \frac{N}{W} \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{N} (x_i - \bar{x})^2}$$

Where:

N is the number of spatial units indexed by i and j
x is the variable of interest
x is the mean of x
wij are the elements of a matrix of spatial weights with zeroes on the diagonal
W is the sum of all wij

The null hypothesis when evaluating Moran's I is that the data is random, and thus there is no spatial autocorrelation, so the expected value of I would be E(I) = -1 / N – 1.

Rook weights and queen weights are used by Moran'I to construct spatial weight matrix. Spatial weights are applied to observations in the simplest form to simply express whether two features are neighbors, or not. The rook finds neighbors based on the definition that neighbors share a common edge. The queen finds neighbors based on the definition that neighbors share a common edge or a common vertex (Anselin 2020). We applied both these methods to our data to determine which approach improved model performance the most.

Random Forest is a method of averaging the results of many decision trees to reduce variance and gain an estimate of the importance of individual features within the model (Ho 1995). In regression tasks, such as in our case, the random forest model will return the mean predicted value for the target variable (in our case, level of crime). The Random Forest randomly selects variables to include in various decision trees, and fits trees to the different selections, then averages all the predictions across all the models . Mathematically, it is:

For *b* = 1, ..., *B*:

1. Sample, with replacement, n training examples from X, Y; call these Xb, Yb.
2. Train a classification or regression tree fb on Xb, Yb.
3. Then average the predictions on x':

$$\hat{f} = \frac{1}{B} \sum_{b=1}^{B} f_b(x')$$

Getis-Ord Gi* Hot Spot analysis allows us to see where variables with high or low values cluster spatially, by looking at each feature in the context of neighboring features (Getis and Ord 1992). These values, like Moran's I, require z-scores and p-values to interpret. The Gi* statistic for each feature is a z-score, and for statistically significant positive scores, higher values denote higher levels of clustering, and vice versa for statistically significant negative scores. We used the Hot Spot Analysis Tool in ArcGIS Pro to calculate these statistics for our features using the input variables (Esri 2023).

The Getis-Ord local statistic is given as:

$$G_i^* = \frac{\sum_{j=1}^{n} w_{i,j}x_j - \bar{X}\sum_{j=1}^{n} w_{i,j}}{S\sqrt{\frac{\left[n\sum_{j=1}^{n} w_{i,j}^2 - \left(\sum_{j=1}^{n} w_{i,j}\right)^2\right]}{n-1}}} \tag{1}$$

where x_j is the attribute value for feature j, $w_{i,j}$ is the spatial weight between feature i and j, n is equal to the total number of features and:

$$\bar{X} = \frac{\sum_{j=1}^{n} x_j}{n} \tag{2}$$

$$S = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - (\bar{X})^2} \tag{3}$$

The G_i^* statistic is a z-score so no further calculations are required.

via Esri 2023

Spatial dependence in spatial data can result in the spatial autocorrelation of regression residuals. Spatial autocorrelation occurs when the values of variables sampled at nearby locations are not independent from each other. Spatial autocorrelation may be either positive or negative. Positive spatial autocorrelation occurs when similar values appear together in space, while negative spatial autocorrelation occurs when dissimilar values appear together.

We employ the spatial lag model to conduct spatial autoregression. The spatial lag model incorporates an additional covariate into the conventional terms for independent variables and errors used in ordinary least squares (OLS) regression. The extra variable, referred to as the "spatial lag" variable, represents the weighted average of values for the dependent variable in areas identified as "neighbors."

Following the standard OLS regression model and adhering to the approach outlined by Anselin (1988), the spatial lag model can be expressed as follows:

$$Y = \rho W y + X\beta + \varepsilon,$$

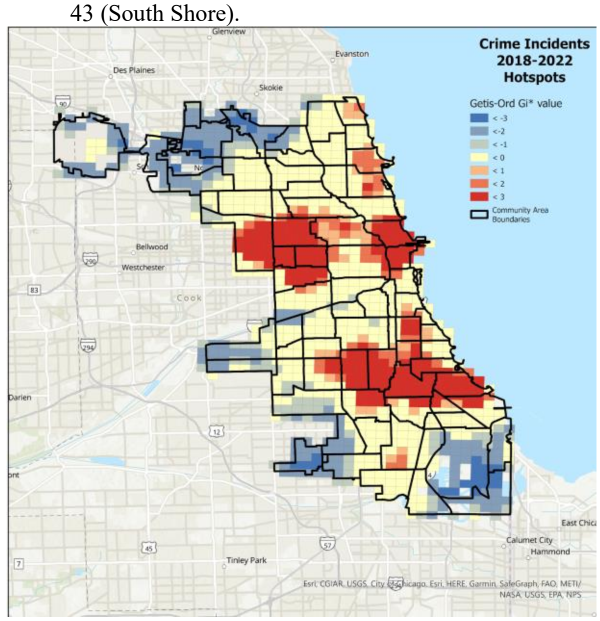
Where Y is the dependent variable of interest, ρ is the autoregression parameter, W is the spatial weights matrix, X is the independent variable, and ε is the error term.

3 Experiments and Results

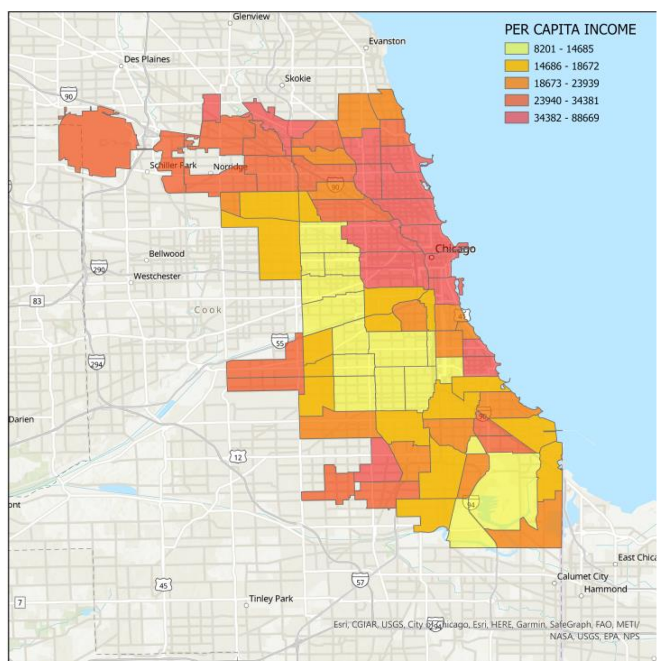
For exploratory data analysis, to get a sense of the general spatial pattern of the crime data across the entire study period, we used the Getis-Ord Gi* Hot Spot analysis tool in the ArcGIS Pro software with the crime incidents dataset.

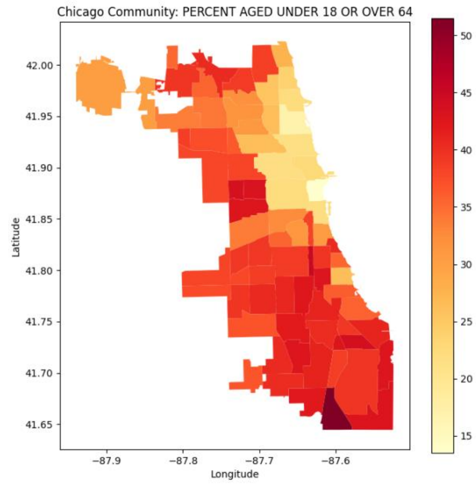
We found that over our 5 year study period, three major crime hotspots were identified. After overlaying with the community area boundaries, it can be seen that these hotspots are mainly located in the following community areas:

- Hotspot 1: 25 (Austin), 23 (Humboldt Park), 26 (West Garfield Park), 27 (East Garfield Park), 29 (North Lawndale)
- Hotspot 2: 36 (Near North Side), 37 (Loop)
- Hotspot 3: 64 (West Englewood), 65 (Englewood), 66 (Greater Grand Crossing), 42 (Woodlawn), and 43 (South Shore).



As part of our exploratory data analysis, we also plotted the variables across community areas to look for patterns. In the interest on concision we will not include all the maps here, but some of the interesting ones include:





Next, we ran three other types of models on our seven different subsets of the data, running once for each of the five years, and including each of our nine variables of interest. Obviously, this created many more results that we can concisely report here. Below, we report on the performance of the different models, reporting R-squared and p-values where applicable. We do not report the coefficient on every explanatory variable for every year for every subset from the OLS model, but instead report the p-value on the spatial lag variable (lag_count_per_population_rook) for the variations of the OLS models (Table 4). We report on this variable specifically because it provides information about the significance of spatial autocorrelation. We do also report the p-value on the other variables for the main dataset including all crime incidents (Table 5).

Likewise, for the Random Forest model, we report the result of feature importance value for the spatial lag variable (Table 1). For Moran's I, we report the Moran's I values (which are already z-scores) for the total crime per year variable for each subset of the data (Table 2), and the Moran's I value for all variables for the main dataset including all crime incidents (Table 3).

Models:

- Global Moran's
- Basic OLS with Rook Weights
- Random Forest

We ran each of the above models on different subsets of the data for each year:

- Only crime incidents occurring during the day
- Only crime incidents occurring at night
- Only crime incidents concerning narcotics
- Only violent crime incidents
- Only non-violent crime incidents
- Only theft incidents (excluded from RF results due to very low R-squared)
- Only burglary incidents (excluded from RF results due to very low R-squared)

Table 1
Random Forest - Spatial Lag only
Feature Importance values for spatial lag variable within each subset model. Each cell represents a different model, run on only the data for the given year and given data subset. The top value represents the feature importance value and the bottom value represents the R-squared value. R-squared values over or equal to 0.6 are in bold.

	2018	2019	2020	2021	2022
Day	0.56 0.47	0.57 0.52	0.58 0.59	0.57 0.57	0.52 0.45
Night	0.60 0.56	0.59 0.53	**0.58** **0.60**	0.57 0.56	0.54 0.45
Violent	**0.43** **0.64**	**0.43** **0.67**	**0.33** **0.62**	**0.41** **0.64**	0.42 0.55
Non-violent	0.51 0.43	0.52 0.43	0.53 0.58	0.48 0.45	0.43 0.39
Narcotics	0.19 0.21	0.16 0.26	Data not available	Data not available	0.18 0.08

Table 2
Moran's I values for total crime per year variable. Values over 0.4 are in bold.

	2018	2019	2020	2021	2022
Day	0.4334	0.4418	0.3804	0.3714	0.3939
Night	0.4053	0.4097	0.3738	0.3738	0.3686
Violent	0.3658	0.3649	0.3480	0.3482	0.3347
Non-violent	**0.4597**	**0.4649**	0.4021	0.3973	0.4115
Narcotics	**0.47**	**0.47**	**0.47**	**0.46**	**0.47**
Theft	**0.53**	**0.51**	**0.49**	**0.51**	**0.54**
Burglary	0.27	0.29	0.31	0.26	0.27

Table 3A
Moran's I for all data - Moran's values over 0.6 are in bold.

Features	Moran's I
% in poverty	0.49
Per capita income	**0.61**
% w/o high school diploma	0.55
% of households overcrowded	0.47
% unemployed	0.57
% under 18 or over 65	**0.63**
Hardship Idx	0.56

Table 3B
Moran's I for total crime during the study period compared to average crime per year during the study period

	2018	2019	2020	2021	2022
Crime	0.42	0.43	0.37	0.37	0.38
Crime density	0.50	0.49	0.55	0.58	0.56

Table 4
OLS with Rook - Each cell represents a separate model - the top value is the p-value for spatial la), and the bottom is the R-squared value for the model, with values over 0.6 in bold.

	2018	2019	2020	2021	2022
Day	0.00002 0.55	0.00002 0.56	0.00004 0.50	0.00006 0.49	0.00013 0.51
Night	0.00001 0.52	0.00001 0.53	0.00001 0.49	0.00001 0.49	0.00004 0.49
Violent	0.00002 0.50	0.00002 0.50	0.00004 0.52	0.00006 0.51	0.00013 0.53
Non-violent	0.00001 0.57	0.00000 0.58	0.00000 0.47	0.00001 0.47	0.00003 0.46
Narcotics	0.00000 0.58	**0.00000** **0.66**	**0.00000** **0.72**	**0.00000** **0.74**	**0.00000** **0.67**
Theft	**0.00074** **0.67**	**0.00182** **0.65**	**0.00007** **0.63**	**0.00021** **0.64**	**0.00016** **0.68**
Burglary	0.00001 0.38	0.00001 0.39	0.00034 0.39	0.00001 0.37	0.00002 0.38

Table 5
OLS p-values for all data. Each column represents a separate model, each row represents a variable within the models. Statistically significant p-values (less than 0.05) are in bold.

Features	2018	2019	2020	2021	2022
% Housing Overcrowded	0.34008	0.28165	0.34132	0.33352	0.31090
% in Poverty	0.29962	0.28421	0.32520	0.37487	0.46700
% Unemployed	0.56389	0.54751	0.61389	0.58217	0.47939
% w/o high school diploma	0.05866	**0.04611**	0.06021	0.07292	0.11023
% under 18 or over 65	**0.01152**	**0.01005**	**0.01489**	**0.01394**	**0.00979**
Per capita income	0.27643	0.33006	0.96180	0.75315	0.62186
Hardship index	0.13331	0.13122	0.20994	0.22204	0.30444
Spatial Lag	**0.00001**	**0.00000**	**0.00001**	**0.00002**	**0.00004**
R-squared	0.5458	0.5566	0.5019	0.4961	0.5043

4 Discussion

Table 1 shows that the variable with the highest R-squared values in the Random Forest model is the version of the model run on only violent crime incidents, suggesting that the model performs better on this subset of the data. However, it does not predict very high feature importance values for spatial lag in this model, compared to the other subset models, suggesting that while violent crime exhibits some degree of spatial autocorrelation, its occurrence may also be influenced by other variables. Table 2 shows that the only-narcotics and only-theft subsets of the data show greater spatial autocorrelation than the other subset variations, suggesting that narcotics crimes and theft crimes may be more correlated with neighborhoods than crimes like burglary, which had lower Moran's values. Table 3A shows the Moran's values for individual variables, for the main dataset with no subsetting. The highest Moran's I is for per capita income, which is an expected result, given that income can be very spatially clustered and vary greatly depending on neighborhoods. The next highest Moran's I is for the percent of population in each community area that is under the age of 18, or over the age of 65. This is interesting because it suggests that age is spatially correlated. This might suggest that working age adults are more likely to be clustered in certain neighborhoods, and likewise non-working age individuals are clustered spatially as well. Table 3B shows the Moran's I values for total crime as compared to crime density (total crime / population as of the 2020 census) for each year. The crime density Moran's values are all higher than the total crime values, implying that crime density may provide a more effective representation of spatial patterns than raw count alone. This makes sense because total crime by community area does not account for the fact that some community areas have a higher population and thus may have inherently higher crime, but crime density may allow for a clearer picture of how crime correlates with location. Another interesting result in Table 3B is the jump in Moran’s I values for crime density in 2020 and after. During the Covid-19 pandemic, some kinds of crime increased (such as domestic violence), while others decreased (such as robbery) (Meyer et al. 2022). It is interesting to see that Moran's I scores for crime density increased during this time, suggesting that perhaps certain kinds of crime became more spatially correlated. One possible explanation might be that the Covid-19 pandemic affected and limited people’s movements around the city, but more granular data and analysis would be needed to say definitively. Table 4 shows that the Theft and Narcotics data subset variations of the OLS model had the highest R-squared values, suggesting that more of the variation in the data could be explained by the model when run on these subsets. However, many of the subsets had statistically significant p-values on the spatial lag variable. This suggests that in the OLS model, these subsets of crime incidents could experience a possible diffusion process, i.e, the sub-groups of crime data points have associations across neighboring community areas. Table 5 shows the p-values for each variable in the OLS model run on all data without subsetting. The variables with consistently low p-values across all years are the spatial lag variable, which can be expected given the results in Table 4, and the percent of population in each community area that is under the age of 18, or over the age of 65, which is consistent with the results from Table 3B and further suggests that age distribution may be clustered across neighborhoods and correlated with crime levels in community areas.

Some ethical considerations with regard to these results are the concern that these associations, such as the low p-value on percent under 18 or over 65, might be interpreted to mean that certain groups, such as young people under 18, are more likely to commit crimes. It is important to understand that these results should not be used to draw these sorts of conclusions, or to suggest that crime in certain locations might be “inevitable” because of the high spatial autocorrelation of crime in those areas. It is important to note that our research does not attempt to address causality or speak to the root causes of crime. Instead, we focus on variables that may go hand-in-hand with crime incidents across neighborhoods, to better understand what factors tend to co-locate.

These results could be used by policy makers and law enforcement decision makers in Chicago to enact more effective policing strategies. By knowing what variables are correlated with crime and are spatially dependent, policing resources can be deployed more strategically to increase the chances that officers are nearby when crimes occur. For example, the finding that percent of community area population under 18 or over 65 may be associated with higher crime density in community areas could influence policing resources to focus on areas with higher percentages of young people (based on the assumption that this relationship is mostly driven by the younger population, although more research would be needed). Additionally, city resources besides law enforcement, such as social workers, drug counselors, and more, might be able to use the findings that Narcotics crimes are highly spatially correlated to more effectively target drug-addiction and anti-distribution resources to neighborhoods that need them.

5 Conclusion

This research aims to shed light on the socioeconomic variables that correlate with crime, and to understand what variables related to crime are spatially correlated. We examined nine socioeconomic variables and compared three different models and seven different subsets of the data. We found that narcotics crimes and theft crimes may be more correlated with certain neighborhoods, as well as percent of community area population under 18 or over 65. We also found widespread statistical significance of spatial lag across many of the data subsets, suggesting that the various sub-groups of crime data points have associations across neighboring community areas. These results are not causal in nature and more research is needed to determine what, if any, relationship these factors may have to the root causes of crime. Additionally, more research is needed to see how these different factors might vary within individual community areas, because it is highly possible that the different clustering effects among the variables are more localized than the community area boundaries can reflect.

ACKNOWLEDGMENTS

We would like to acknowledge the City of Chicago Department of Innovation and Technology which maintains their open data portal that we used to gain access to the data for this research. We would also like to acknowledge Professor Naren Ramakrishnan, who provided feedback on our initial research proposal, suggested various methods of analysis along the way, and provided valuable advice on starting early, which we ignored.

Author contributions

Harditya Sarvaiya: ran the OLS, Random Forest, and Moran’s I calculations

Amy Hilla: located datasets, ran the Hotspot Analysis, conducted literature review

Dataset and code

Our code is available to download at https://github.com/darthwaydr007/CS-5834

Data sources:

- Population data: Chicago Metropolitan Agency for Planning
- Income data: Per Capita Income | City of Chicago | Data Portal
- Crime data: Crimes - 2001 to Present | City of Chicago | Data Portal
    - Download only years 2018-2022
- Community areas: Boundaries - Community Areas (current) | City of Chicago | Data Portal

REFERENCES

Anselin, Luc (1995). "Local Indicators of Spatial Association—LISA". Geographical Analysis. 27 (2): 93–115. doi:10.1111/j.1538-4632.1995.tb00338.x.

Anselin, Luc & Cohen, Jacqueline & Cook, David & Gorr, Wilpen & Tita, George. (2000). Spatial Analyses of Crime. https://www.publicsafety.gc.ca/lbrr/archives/cnmcs-plcng/cn34984-v4-213-262-eng.pdf

Anselin, Luc. (2020). Contiguity-Based Spatial Weights. https://geodacenter.github.io/workbook/4a_contig_weights/lab4a.html

Bloch, Stefano. (2020). Broken Windows Ideology and the (Mis)Reading of Graffiti. Critical Criminology. 28. doi:10.1007/s10612-019-09444-w.

Bolter, M. (2022, March 21). Urban patterns of murder in Chicago. ArcGIS StoryMaps. https://storymaps.arcgis.com/stories/525fe80df30f492e8cf29b433917636b

Brown, Marilyn A. (1982) Modelling the Spatial Distribution of Suburban Crime, Economic Geography, 58:3, 247-261, DOI: 10.2307/143513

Burgess, Ernest W., and Charles Newcomb, eds. Census Data of the City of Chicago, 1920. http://www.encyclopedia.chicagohistory.org/pages/319.html

Esri. (2023). How Hot Spot Analysis (Getis-Ord Gi*) works. ArcGIS Pro. https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/h-how-hot-spot-analysis-getis-ord-gi-spatial-stati.htm#:~:text=The%20Hot%20Spot%20Analysis%20tool,the%20context%20of%20neighboring%20features.

Getis, A. and J.K. Ord. (1992) The Analysis of Spatial Association by Use of Distance Statistics. Geographical Analysis 24(3).

Hardship Index - Chicago Health Atlas. (n.d.). https://chicagohealthatlas.org/indicators/HDX?topic=hardship-index

Ho, Tin Kam (1995). Random Decision Forests. Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, 14–16 August 1995. pp. 278–282. https://web.archive.org/web/20160417030218/http://ect.bell-labs.com/who/tkh/publications/papers/odt.pdf

Hooghe, Marc & Vanhoutte, Bram & Hardyns, Wim & Bircan, Tuba. (2010). Unemployment, Inequality, Poverty and Crime: Spatial Distribution Patterns of Criminal Acts in Belgium, 2001–06. British Journal of Criminology. 51. 1-20. 10.1093/bjc/azq067.

Marotta, P. Assessing Spatial Relationships Between Rates of Crime and Rates of Gonorrhea and Chlamydia in Chicago, 2012. J Urban Health 94, 276–288 (2017). https://doi.org/10.1007/s11524-016-0080-7.

Mikaela Meyer, Ahmed Hassafy, Gina Lewis, Prasun Shrestha, Amelia M. Haviland & Daniel S. Nagin (2022) Changes in Crime Rates during the COVID-19 Pandemic, Statistics and Public Policy, 9:1, 97-109, DOI: 10.1080/2330443X.2022.2071369

Moran, P. A. P. (1950). "Notes on Continuous Stochastic Phenomena". Biometrika. 37 (1): 17–23. doi:10.2307/2332142. JSTOR 2332142. PMID 15420245.

Nathan, R. P., & Adams, C. (1976). Understanding Central City Hardship. Political Science Quarterly, 91(1), 47–62. https://doi.org/10.2307/2149158

Stretesky, P. B., Schuck, A. M., & Hogan, M. J. (2004). Space matters: an analysis of poverty, poverty clustering, and violent crime. Justice Quarterly, 21(4), 817-842.

Towers S, Chen S, Malik A, Ebert D. Factors influencing temporal patterns in crime in a large American city: A predictive analytics perspective. PLoS One. 2018 Oct 24;13(10):e0205151. doi: 10.1371/journal.pone.0205151. PMID: 30356321; PMCID: PMC6200217.