Understanding unsolved crimes hotspots: a spatial approach

Cabral Juan Andrés

Work in progress

Abstract

This paper aims to investigate unsolved crime hotspots through a spatial approach. We use data from more than 52,000 homicides that took place over the past decade in America's largest cities. Common tests are carried out to detect the presence of spatial dependence, and spatial mapping and modeling tools are used. The analysis carried out is intended to be useful to understand and prevent those crimes that are not solved.

Introduction

Over the past few years, numerous studies have explored the determinants and effects of crime hotspots. Many studies suggest that crime is not evenly distributed throughout space, but rather is concentrated in certain areas called "hotspots" (L. W. Sherman, Gartin, and M. E. Buerger 1989, D. Weisburd et al. 1992 and Eck and D. L. Weisburd 2015).[1] As defined by L. W. Sherman 1995, hotspots are small places in which the occurrence of crime is so frequent that it is highly predictable (at least over one year). Given this definition, undoubtedly knowing which areas are hotspots can be used to fight crime; there is abundant evidence on the effectiveness of police interventions in hotspots (see for example Kuo, Lord, and Walden 2013 and for a summary of the available evidence see Braga, Papachristos, and Hureau 2014). Therefore, the analysis of these areas is very useful to reduce crime.

An important part of the literature on hotspots stick to the following structure: mapping crimes, identifying hotspots through a particular method, using some type of spatial dependency test and later using a particular approach, such as modeling spatial dependency (see Baller et al. 2001). Another part uses OLS models to see the impact of hotspots on some variable (see, for example, Ceccato and Wilhelmsson 2020 where they see the impact of hotspots on property prices). Finally, there exists a third group of studies that try to assess which variables can explain crime without explicitly including spatial dependence. An example is Favarin 2018, where it is shown that in Milan Italy there are hotspots which are stable. The author then tries to explain them through different factors using negative binomial regressions models.

Since hotspot phenomena are essentially spatial phenomena, it is difficult to find a study that omits tools of spatial econometrics entirely. Furthermore, all the studies cited here that have performed a Moran's I or similar test obtained results

[1] Some authors write "hotspots" as "hot spots".

that suggest spatial dependence, so ignoring spatial effect in this type of analysis is not desirable.

In this work, the hotspots of unsolved crimes are investigated through spatial analysis. As said before, hundreds of papers have been devoted to analyzing the hotspots of different types of crimes and their effects on other variables, but to date, no study has looked specifically at hotspots of those crimes that are not solved.

This paper contributes mainly to two pieces of literature. On one hand, it analyzes crime hotspots like many other works. It seeks to replicate some stylized facts, such as showing that crimes are not distributed randomly but that there are crime clusters.[2] On the other hand, this work contributes to crime prevention literature. Much of the criminology work focuses on crimes without discriminating whether they were solved or not. Focusing on unsolved crimes can be useful to dedicate resources precisely to avoid those crimes that for various reasons cannot be solved.

Spatial Econometrics background

As Herrera Gómez et al. 2017, p. 2 comments, the birth of spatial econometrics is located around 1979. In its early years, the models that explicitly incorporated space were focused on issues of geographic or urban economy, however, spatial econometric models began to be applied in other areas over the years. Among the areas that were influenced by spatial econometrics is the study of crime.

Given that the phenomenon of hotspots is related to non-random distributions in space, the use of spatial tools to identify clusters is commonplace (for example, many of the examples used in spatial econometrics textbooks are related to criminology, see Sarrias 2020, LeSage 2015 and Elhorst 2014). One of the first steps in studies that analyze hotspots is to map these hotspots. For this, there are various methods as dot maps, line maps, ellipse maps, etc. The method to be used depends on the type of data and what you want to show. Line maps, for example, are useful when hotspots are along streets whereas Dot Maps, are used when hotspots are at specific addresses, corners, and other places. Ellipse and choropleth maps are used when the designated hotspots share the same risk level, so a specific location inside that area is irrelevant (see Table 1 for more methods and examples).

As previously stated, hotspot analyzes begin through the use of some spatial dependency test, whether local or global. Some of these tests are Moran's I and Geary's C. These tests are generally computed through particular softwares; for example, the CrimeStat spatial software includes these tests (Levine 2006, p. 44). Local Moran's I tests are common in the literature of crime hotspots; Kerry et

[2] As it is noted in figure 7 of Kedia 2016, there are different types of hotspots, in this work, we focus on moderately concentrated clusters but not crimes exactly positioned at one point.

Table 1: Tools used in crime concentration investigation

Methods	Example
Mapping	
Dot maps	Ceccato and Wilhelmsson 2020
Line maps	Telep and Hibdon 2019
Ellipse maps	Zhang et al. 2010
Grid thematic maps	Chainey, Tompson, and Uhlig 2008
Choropleth/thematic maps	Chainey, Tompson, and Uhlig 2008
Isoline maps	Telep and Hibdon 2019
Interpolation and continuous surface smoothing methods	Chainey, Tompson, and Uhlig 2008
Test for spatial auto correlation	
Moran's I	Baller et al. 2001
Geary's C	Eck, Chainey, et al. 2005
Local Indicators of Spatial Association (LISA) statistics	
Gi and Gi*	J. H. Ratcliffe and McCullagh 1999
Local Moran's I	Rogerson and Kedron 2012
Local Geary's C	Anselin 2019
Commonly used software	
CrimeStat	Levine 2006
GeoDa	Leitner and Brecht 2007
ArcView	Brunsdon, Corcoran, and Higgs 2007
Regression models	
Spatial lag model (SLM)	Kubrin and Weitzer 2003
Spatial error model (SEM)	Miles-Doan 1998

Table 1 shows a summary of different tools used to detect and analyze hotspots and spatial interactions in crime related research. The table organization is mainly based on the works of Chainey, Tompson, and Uhlig 2008 and Chainey and J. Ratcliffe 2013.

al. 2010 investigates car theft, Murray et al. 2001 investigates property crimes, Baller et al. 2001 turned their attention to homicides and Cohen and Tita 1999 attempted to identify clusters in violent crimes.

When looking for jobs that include spatial analysis, it is convenient to review the most recent jobs because these tend to use particular tools of spatial econometrics. Ahmed and Salihu 2013 investigate hotspots for various categories of crime in Nigeria through interpolation methods (particularly using the Inverse Distance Weighted method). In most papers crimes are mapped, however, not many of them aim to analyze the predictive capacity of the method used. Chainey, Tompson, and Uhlig 2008 show that it is not only relevant to be able to map hotspots but also to evaluate the predictive capacity of crime, since it is useless to find hotspots if they will disappear in the future. By comparing various methods, it shows how effective each one is in predicting crime.

As mentioned above, many works perform spatial dependency tests. For example, Chakravorty 1995, p. 57 carries out an analysis of Philadelphia and recommends that to locate hotspots it is advisable to perform LISA tests instead of global tests because LISA tests identify small local clusters when large heterogeneous areas are used.[3] Additionally, Rogerson and Kedron 2012 suggests that due to the problem of using different weighting matrices it is necessary to adjust for the multiple tests performed. They develop a local Moran's I test that adjusts by testing with multiple weighting matrices.

As mentioned before, there are works that explicitly model spatial effects. What is interesting about the work of Baller et al. 2001 is that they use models that explicitly take spatial effects into account and compare them. The authors find that there is residual spatial autocorrelation in all the investigated periods, in addition to the fact that there are diffusion effects in which certain crimes in some counties of United States have an influence on crimes in others.

Law, Quick, and Chan 2015 explored hotspots for violent crime in Toronto (2006-2007) through a Bayesian approach. In their paper, they identify hotspots based on the trend from 2006 to 2007. An interesting aspect of this work is that it adopts a non-frequentist approach that according to the authors is convenient from a law enforcement perspective. This type of study is not very common.

It is also relevant to highlight that the works that analyze hotspots do so by selecting a certain time window (commonly months or years). These works also use panel data so it is possible to perform spatio-temporal analyzes. Among those works that study crime hotspots with a spatio-temporal approach is He, Páez, and Liu 2017 which explores the temporal persistence of hotspots. The study shows that certain socio-economic factors manage to explain the presence

[3] An interesting point that Quick and Law 2013 makes is that as the Local Moran's I test detect small and concentrated clusters, then they are very useful to design prevention policies because they allow the police forces to be directed towards areas that do not need a large amount of police personnel to be covered. This is why if you want to focus on prevention, it is relevant to carry out Local Moran's I tests.

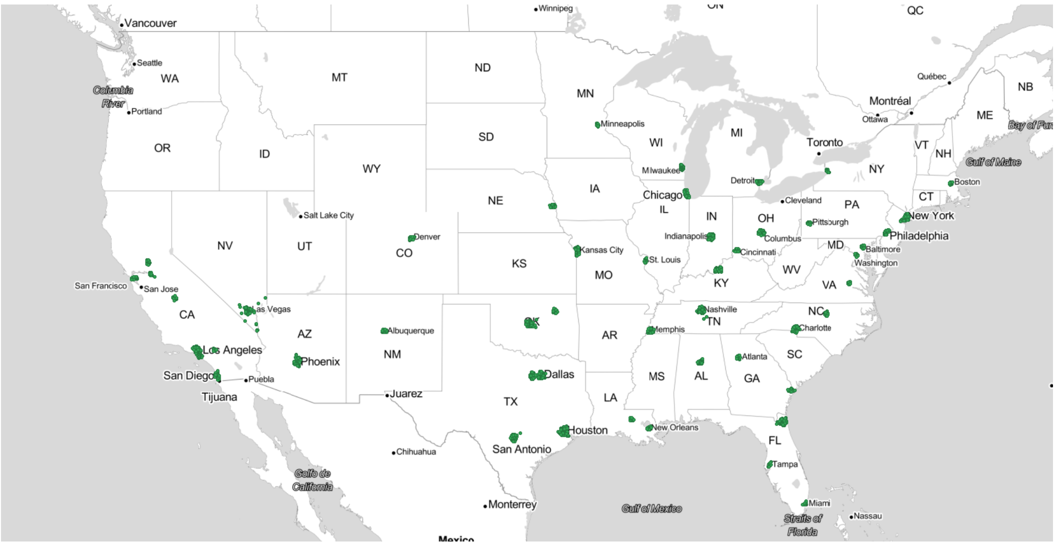

Figure 1: Map of the United States with the database murders.

of violent crimes in hot zones.

Data

The data comes from a Washington Post work. The database contains more than 52,000 homicides that took place in the past decade in the 50 largest cities in the United States. The database includes the location (latitude and longitude), if the case is closed and if there were any arrests among other basic data of the victim (race, sex, age). This database was built through data that comes from different sources. As the repository from Washington Post said, reporters worked for months to clean up the data and standardize it. In comparisons with FBI data, it was verified that the database is consistent with other sources. Some analysis of the database can be found at Washington Post. More information on the database can be found in a Github Repository.

A homicide is considered to be closed by arrest when police reported that to be the case. Cases were counted as closed without arrest if they were reported to be "exceptionally cleared". This means that there is enough evidence but an arrest is not possible; an example of this is when the suspect has died. All other cases were classified as having no arrest.

In Figure 1 it can be seen the total number of murders on a map of the United States. If the image is enlarged, it is possible to see the distribution of crimes within a particular state, for example in LA (Figure 2).

Other data was obtained from Chicago data portal; for example, Chicago communities boundaries was used for the Chicago communities shapefile. To carry out the spatial modeling, data on education, poverty, and income from selected socioeconomic indicators was also consulted. Different programs were used to

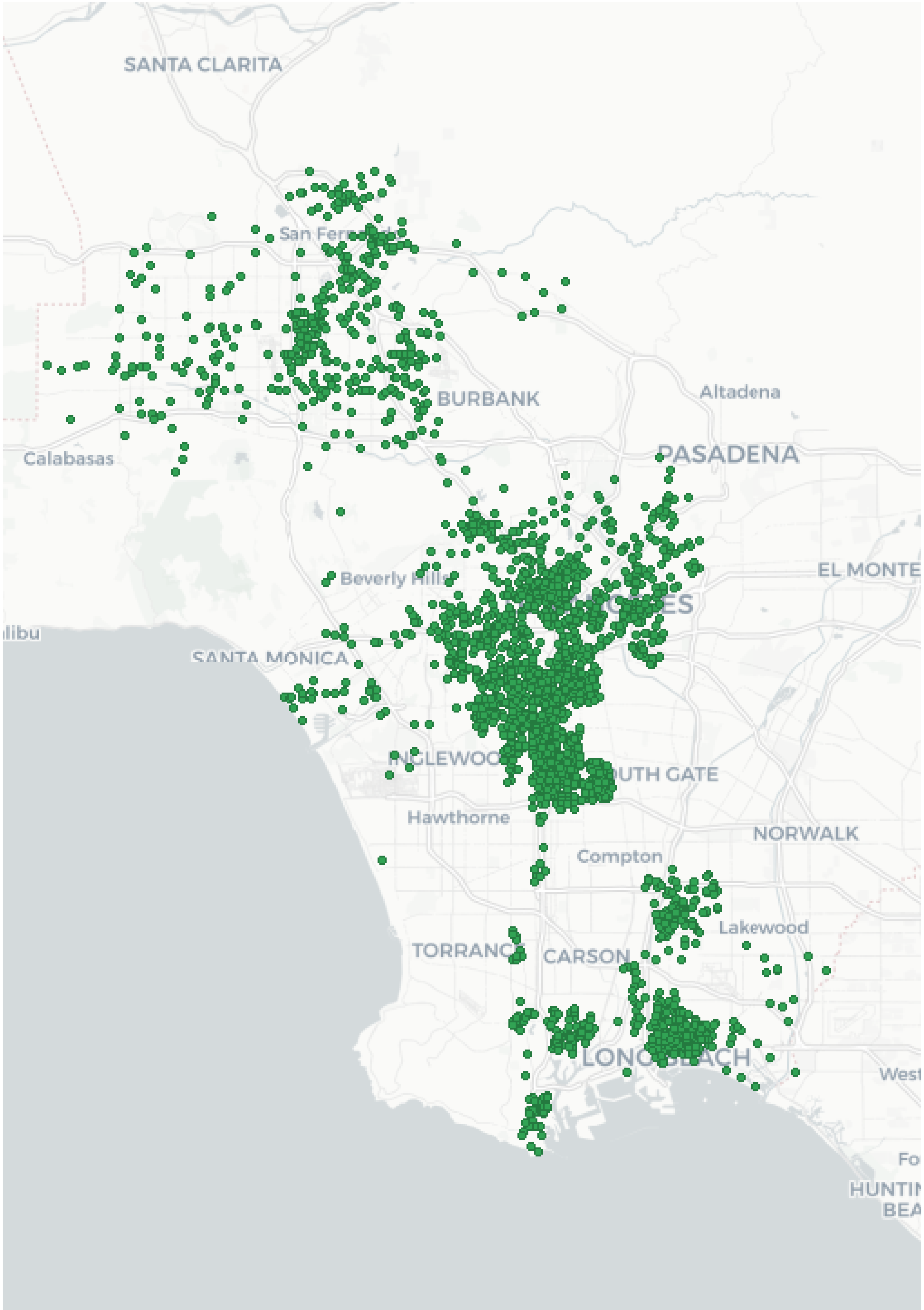

Figure 2: Map of Los Angeles with the database murders.

view the data and carry out different tests. R was used for handling the .csv files and spatial modeling, QGIS was used to perform spatial joins, and GEODA for the creation of maps and spatial dependency tests. A summary of the path taken can be seen in Figure 3.

Descriptive statistics

Descriptive statistics can be seen in Table 2. Most of the variables are dummies. It is possible to see that the victims tend to be relatively young (around 30 years old) and that the cases that are closed and not resolved are only 5%.

One of the relationships that can be observed through this dataset is that the proportion of unsolved cases (open or closed without arrest) is approximately 51%. Further, through the data of the victims we can see that in the case that the victim is a black person, it is more likely that the case is not resolved (54% for black people, compared to 37% for people white). Something similar happens with men; men's cases tend to be resolved less (53% compared to 42% in the case of women).

Table 2: Descriptive statistics

Statistic	N	Mean	Min	Max
open_or_no_arrest	52,179	0.508	0	1
closed_without_arrest	52,179	0.056	0	1
arrest	52,179	0.492	0	1
black	52,179	0.639	0	1
white	52,179	0.121	0	1
male	52,179	0.781	0	1
age	49,180	29.120	1.000	101.000

A temporal analysis shows us that over time, there was a greater increase in unsolved crimes than in solved crimes (Figure 4).

To perform the spatial analysis, the 2 cities with the highest number of observations were selected; Table 3 shows that these are Chicago and Philadelphia.

Methodology

Both Chicago and Philadelphia crimes are analyzed with disaggregated data. First, a matrix of spatial weights is computed based on Almeida, Haddad, Hewings, et al. 2005 who use the method of inverse distance. Some tests are carried out to detect spatial dependence. Second, once the Chicago and Philadelphia results have been obtained, the Chicago data is added by community and a more conventional analysis is carried out through a matrix of spatial weights based on

Table 3: Descriptive statistics for cities

	City	Observations	% unsolved cases	% black victim
10	Omaha	409	0.413	0.606
11	Baton Rouge	424	0.462	0.884
12	Richmond	429	0.263	0.876
13	Stockton	444	0.599	0.327
14	San Diego	461	0.380	0.321
15	Fresno	487	0.347	0.300
16	Buffalo	521	0.612	0.814
17	Fort Worth	549	0.464	0.454
18	Louisville	576	0.453	0.665
19	Tulsa	584	0.330	0.488
20	Boston	614	0.505	0.708
21	New York	627	0.388	0.560
22	Pittsburgh	631	0.534	0.865
23	San Francisco	663	0.507	0.522
24	Oklahoma City	672	0.485	0.458
25	Charlotte	687	0.300	0.707
26	Cincinnati	694	0.445	0.849
27	Miami	744	0.605	0.409
28	Nashville	767	0.362	0.636
29	Birmingham	800	0.434	0.882
30	San Antonio	833	0.429	0.218
31	Phoenix	914	0.551	0
32	Oakland	947	0.536	0.725
33	Atlanta	973	0.383	0.909
34	Columbus	1,084	0.530	0.708
35	Milwaukee	1,115	0.361	0.795
36	Jacksonville	1,168	0.511	0.681
37	Kansas City	1,190	0.408	0
38	Indianapolis	1,322	0.449	0.688
39	Washington	1,345	0.438	0.905
40	Las Vegas	1,381	0.414	0.327
41	New Orleans	1,434	0.649	0.882
42	Memphis	1,514	0.319	0.849
43	Dallas	1,567	0.481	0
44	St. Louis	1,677	0.540	0.894
45	Los Angeles	2,257	0.490	0.393
46	Detroit	2,519	0.588	0.892
47	Baltimore	2,827	0.646	0.918
48	Houston	2,942	0.507	0.512
49	**Philadelphia**	3,037	0.448	0.777
50	**Chicago**	5,535	0.736	0.768

queen contiguity. Later, this same aggregate database is used to explicitly add the spatial effects.

Spatial dependence

As the outcome variable in the case of disaggregated data is binary (1 if the crime was unsolved and 0 if it was solved), the appropriate (Anselin and Li 2019) test to detect spatial autocorrelation is the Join-count test. However, local and global Moran's I tests also show evidence of spatial autocorrelation. All tests are computed through Geoda, using 999 permutations.

Results

Chicago

Data points

For the Chicago crime database Moran's I spatial dependence test is computed, and the result suggest spatial dependence (p-value=0.001). Through this test it is verified that there is spatial dependence in unsolved crimes. This indicates that at least in this regard there are no differences between these types of crimes (because most of the literature finds spatial dependence in all types of crimes).

In Figure 8 we can see the different clusters that are identified through a join-count test. This is done for unsolved crimes in Chicago. The logic behind the clusters identified by the join-count test is that those points that, unlike the maps generated by local Moran's I tests, there are not 4 classifications but only one. What these maps show are those points that are surrounded by other points that take a value equal to 1 in the binary variable of interest (in this case being an unsolved homicide). Therefore, all dots being homicides, the green dots are unsolved homicides surrounded by other unsolved homicides.

Aggregated data

Aggregating the data in Chicago and calculating per capita unsolved crimes per community allow us to obtain Figure 6. With regard to aggregated data, the results of the different tests suggest spatial dependence, for example Moran's I can be seen in Figure 5 (0.546, and p-value=0.001)

Figure 7 combines information of the Moran scatterplot and the LISA statistic. Four categories are created for classification (all categories are statistically significant in terms of the LISA concept). These categories as usual represent combinations of high values with high values, low values with low values, and low values with highs and high with lows. It should be noted that the Local Moran statistics suffer from sensitivity to outliers because it is built from neighbor averages.

Philadelphia

The results for the Philadelphia dataset suggests that there is spatial dependence (Moran's I test gives a p-value of 0.006). As before, a join-count test is also performed. The resulting map can be seen in Figure 9. These results suggest clusters as in the previous case.

Spatial modelling

When deciding on the specification of models, there are various strategies both theory-driven and data-driven. In this case, the specification will be chosen simply by the results that the data gives. One strategy will be carried out: from the general to the particular (with LR tests). The LR tests are based on maximum likelihood. The LR tests involve estimating the model under alternative hypothesis and null hypothesis and comparing the restricted and unrestricted objective functions. For the choice of models, we will resort to exploring the models detailed in the Figure 10.

The strategy from the general to the particular (á la Hendry) implies starting from a model that includes many terms and captures spatial effects. Then, through tests, reduces it to a simpler model. One of the advantages of this strategy is that it tends to be more robust to anomalies in the data generating process, but in general the 2 strategies (from particular to general and from the general to the particular) tend to achieve the same results (Mur and Angulo 2009). Instead of starting with the GNS model (Figure 10) that cannot be estimated (because it is weakly identified), we start by estimating an SDM model since from it one can arrive at an SLM, SLX, or SEM.

In case of arriving at the simplest model, a result like the one in Table 4 would be obtained.

Once SDM is estimated, it is tested if $\rho = 0$, if $\theta = 0$, and if $rho\beta + \theta = 0$. The results of the common factors test show us that SDM cannot be reduced to SEM, but it can be reduced to SLX. On the other hand, we cannot reduce to SLM either. Table 5 shows the results of the estimated SLX model. The direct effects of the SLX model are equal to the estimated coefficient of each variable. While the indirect effect is equal to the estimated coefficient of the spatial lagged value. This is an advantage of models like SLX since they allow an interpretation without carrying out major calculations. For example, the percentage of households below the poverty line has a positive (and significant) effect on increasing the number of unsolved homicides per capita, but has a negative indirect effect. On the other hand, the percentage of crowded houses has an indirect effect on unsolved crimes per capita. The indirect effect suggests that an increase in the percentage of crowded houses in nearby neighborhoods increases unsolved crimes per capita. Finally, if one wants to obtain the total effect, it is only a matter of adding the direct effect and the indirect effect.

	Dependent variable:
	crimepc
Per capita income	0.000
	(0.00000)
% aged under 18 or over 64	0.00004
	(0.00003)
% aged 16+ unemployed	0.0001***
	(0.00003)
% households below poverty	0.0001***
	(0.00002)
% of housing crowded	0.0002***
	(0.0001)
% aged 25+ without high school diploma	−0.0001***
	(0.00002)
Constant	−0.002*
	(0.001)
Observations	77
R^2	0.748
Adjusted R^2	0.726
Residual Std. Error	0.001 (df = 70)
F Statistic	34.608*** (df = 6; 70)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 4: OLS regression

	Dependent variable:
	crimepc
Per capita income	0.000
	(0.00000)
% aged under 18 or over 64	0.00004
	(0.00003)
% aged 16+ unemployed	0.0001**
	(0.00003)
% households below poverty	0.0001***
	(0.00002)
% of housing crowded	0.0001*
	(0.0001)
% aged 25+ without high school diploma	−0.0001**
	(0.00003)
W: Per capita income	0.00000*
	(0.00000)
W: % aged under 18 or over 64	0.0001
	(0.0001)
W: % aged 16+ unemployed	0.0001**
	(0.0001)
W: % households below poverty	−0.00003
	(0.00004)
W: % of housing crowded	0.0005***
	(0.0002)
W: % aged 25+ without high school diploma	−0.0001
	(0.0001)
Constant	−0.009***
	(0.003)
Observations	77
R^2	0.795
Adjusted R^2	0.756
Residual Std. Error	0.001 (df = 64)
F Statistic	20.649*** (df = 12; 64)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 5: SLX

Conclusions

Through a spatial analysis of unsolved crimes in the largest cities of the United States, different tests are carried out to detect spatial dependence, reaching results that suggest spatial effects. This result is similar to various studies that examine crime hotspots without discriminating on whether they were solved or not.

After an examination of the available literature, we reach the conclusion that this is the first work to analyze crime hotspots that discriminate by the fact that the case was solved or not.

Once the spatial dependence had been detected, said spatial effects were modeled through models that explicitly incorporate the importance of space. This first analysis suggests opening the way to other later studies that achieve a better understanding of the characteristics of unsolved crimes. The final result suggests that the correct model to use is SLX, at least for the chosen city (Chicago) and for the available explanatory variables. However a warning must be made carrying out robustness checks, in the case of using a matrix of spatial weights rook contiguity, the results remain practically the same. Although, when the checks are carried out with a matrix of spatial weights generated by the inverse distance, the exact same results are not achieved.

Finally, the results obtained are similar to those found in studies that do not discriminate based on whether the crime was solved or not.

References

Ahmed, M and RS Salihu (2013). “Spatiotemporal pattern of crime using geographic information system (GIS) approach in Dala LGA of Kano State, Nigeria”. In: *American Journal of Engineering Research (AJER)* 2.3, pp. 51–58.

Almeida, Eduardo Simões de, Eduardo Amaral Haddad, Geoffrey JD Hewings, et al. (2005). “The spatial pattern of crime in Minas Gerais: an exploratory analysis”. In: *Economia Aplicada* 9.1, pp. 39–55.

Andresen, Martin A (2011). “The ambient population and crime analysis”. In: *The Professional Geographer* 63.2, pp. 193–212.

Anselin, Luc (2019). “A local indicator of multivariate spatial association: extending Geary’s C”. In: *Geographical Analysis* 51.2, pp. 133–150.

Anselin, Luc and Xun Li (2019). “Operational local join count statistics for cluster detection”. In: *Journal of geographical systems* 21.2, pp. 189–210.

Baller, Robert D et al. (2001). “Structural covariates of US county homicide rates: Incorporating spatial effects”. In: *Criminology* 39.3, pp. 561–588.

Braga, Anthony A, Andrew V Papachristos, and David M Hureau (2014). “The effects of hot spots policing on crime: An updated systematic review and meta-analysis”. In: *Justice quarterly* 31.4, pp. 633–663.

Brunsdon, Chris, Jonathan Corcoran, and Gary Higgs (2007). “Visualising space and time in crime patterns: A comparison of methods”. In: *Computers, environment and urban systems* 31.1, pp. 52–75.

Ceccato, Vania and Mats Wilhelmsson (2020). “Do crime hot spots affect housing prices?” In: *Nordic Journal of Criminology* 21.1, pp. 84–102.

Chainey, Spencer and Jerry Ratcliffe (2013). *GIS and crime mapping.* John Wiley & Sons.

Chainey, Spencer, Lisa Tompson, and Sebastian Uhlig (2008). “The utility of hotspot mapping for predicting spatial patterns of crime”. In: *Security journal* 21.1-2, pp. 4–28.

Chakravorty, Sanjoy (1995). “Identifying crime clusters: The spatial principles”. In: *Middle States Geographer* 28, pp. 53–58.

Cohen, Jacqueline and George Tita (1999). “Diffusion in homicide: Exploring a general method for detecting spatial diffusion processes”. In: *Journal of quantitative criminology* 15.4, pp. 451–493.

Eck, John, Spencer Chainey, et al. (2005). “Mapping crime: Understanding hotspots”. In:

Eck, John and David L Weisburd (2015). “Crime places in crime theory”. In: *Crime and place: Crime prevention studies* 4.

Elhorst, J Paul (2014). *Spatial econometrics: from cross-sectional data to spatial panels.* Vol. 479. Springer.

Favarin, Serena (2018). “This must be the place (to commit a crime). Testing the law of crime concentration in Milan, Italy”. In: *European Journal of Criminology* 15.6, pp. 702–729.

He, Li, Antonio Páez, and Desheng Liu (2017). "Persistence of crime hot spots: An ordered probit analysis". In: *Geographical analysis* 49.1, pp. 3–22.

Herrera Gómez, Marcos et al. (2017). *Fundamentos de Econometrıa Espacial Aplicada [Fundamentals of Applied Spatial Econometrics]*. Tech. rep. University Library of Munich, Germany.

Kedia, Pranav (2016). "Crime Mapping and Analysis using GIS". In: *International Institute of Information Technology* 1.1, pp. 1–15.

Kerry, Ruth et al. (2010). "Applying Geostatistical Analysis to Crime Data: Car-Related Thefts in the Baltic States." In: *Geographical analysis* 42.1, pp. 53–77.

Kubrin, Charis E and Ronald Weitzer (2003). "Retaliatory homicide: Concentrated disadvantage and neighborhood culture". In: *Social problems* 50.2, pp. 157–180.

Kuo, Pei-Fen, Dominique Lord, and Troy Duane Walden (2013). "Using geographical information systems to organize police patrol routes effectively by grouping hotspots of crash and crime data". In: *Journal of Transport Geography* 30, pp. 138–148.

Law, Jane, Matthew Quick, and Ping W Chan (2015). "Analyzing Hotspots of Crime Using a B ayesian Spatiotemporal Modeling Approach: A Case Study of Violent Crime in the Greater Toronto Area". In: *Geographical Analysis* 47.1, pp. 1–19.

Leitner, Michael and Henrike Brecht (2007). "Software review: Crime analysis and mapping with GeoDa 0.9. 5-i". In: *Social Science Computer Review* 25.2, pp. 265–271.

LeSage, James (2015). "Spatial econometrics". In: *Handbook of research methods and applications in economic geography*. Edward Elgar Publishing.

Levine, Ned (2006). "Crime mapping and the Crimestat program". In: *Geographical analysis* 38.1, pp. 41–56.

Miles-Doan, Rebecca (1998). "Violence between spouses and intimates: Does neighborhood context matter?" In: *Social Forces* 77.2, pp. 623–645.

Mur, Jesus and Ana Angulo (2009). "Model sQelection strategies in a spatial setting: Some additional results". In: *Regional Science and Urban Economics* 39.2, pp. 200–213.

Murray, Alan T et al. (2001). "Exploratory spatial data analysis techniques for examining urban crime: Implications for evaluating treatment". In: *British Journal of criminology* 41.2, pp. 309–329.

Quick, Matthew and Jane Law (2013). "Exploring hotspots of drug offences in Toronto: A comparison of four local spatial cluster detection methods". In: *Canadian journal of criminology and criminal justice* 55.2, pp. 215–238.

Ratcliffe, Jerry H and Michael J McCullagh (1999). "Hotbeds of crime and the search for spatial accuracy". In: *Journal of geographical systems* 1.4, pp. 385–398.

Rogerson, Peter A and Peter Kedron (2012). "Optimal weights for focused tests of clustering using the Local Moran statistic". In: *Geographical Analysis* 44.2, pp. 121–133.

Sarrias, Mauricio (2020). "Notes on Spatial Econometrics". In:
Sherman, Lawrence W (1995). "Hot spots of crime and criminal careers of places". In: *Crime and place* 4, pp. 35–52.
Sherman, Lawrence W, Patrick R Gartin, and Michael E Buerger (1989). "Hot spots of predatory crime: Routine activities and the criminology of place". In: *Criminology* 27.1, pp. 27–56.
Telep, Cody W and Julie Hibdon (2019). *Understanding and Responding to Crime and Disorder Hot Spots*. Department of Justice.
Vega, Solmaria Halleck and J Paul Elhorst (2013). "On spatial econometric models, spillover effects, and W". In: *53rd ERSA Congress, Palermo, Italy*.
Weisburd, David et al. (1992). "Contrasting crime general and crime specific theory: The case of hot spots of crime". In: *Advances in criminological theory* 4.1, pp. 45–69.
Zhang, Xiang et al. (2010). "Detecting and mapping crime hot spots based on improved attribute oriented induce clustering". In: *2010 18th International Conference on Geoinformatics*. IEEE, pp. 1–5.

Appendix

In this appendix, a robustness analysis (with Chicago database) is carried out through another matrix of spatial weights. Following Almeida, Haddad, Hewings, et al. 2005 and Andresen 2011 rook contiguity will be used. The result for Moran's I test is maintained (p-value = 0.001). If a cluster map is generated with the rook contiguity weight, the result is practically identical to the one obtained previously. The same happens with Local Geary cluster maps or Gi cluster maps.

When selecting spatial models, an SDM is first estimated as was done in the previous case, the tests suggest that this model can be reduced to SLX but not to SEM or SLM, as in the previous case. By last, The SLX model cannot be reduced to OLS. By way of illustration, the SLX estimates are presented in Table 6. The results in terms of estimates are practically identical.

Finally, using a matrix of spatial weights generated by the inverse distance, the results begin to vary. In principle, there is spatial dependence but Moran's I is lower (0.215). The clusters generated by the Local Moran's I test also differ in part from those obtained previously. The same is true for the Local Geary and G cluster maps. If a spatial model selection analysis is performed starting again with SDM, the result leads to SLX, but unlike the previous cases now SLX can be reduced to OLS (Table 4). In this way, it can be concluded that this last result is not robust to the specification of the spatial weight matrix. In fact, when a Moran I test is performed on the residuals of the linear model with the explanatory variables used previously, H0 cannot be rejected, that is, the result suggests that there is no spatial dependence.

Other robustness checks may include testing other spatial weight matrices or even following a strategy from particular to general and expecting the same result.

	Dependent variable:
	crimepc
Per capita income	−0.000
	(0.00000)
% aged under 18 or over 64	0.00005
	(0.00003)
% aged 16+ unemployed	0.0001**
	(0.00003)
% households below poverty	0.0001**
	(0.00002)
% of housing crowded	0.0001**
	(0.0001)
% aged 25+ without high school diploma	−0.0001**
	(0.00003)
W: Per capita income	0.00000
	(0.00000)
W: % aged under 18 or over 64	0.0001
	(0.0001)
W: % aged 16+ unemployed	0.0001*
	(0.0001)
W: % households below poverty	−0.00001
	(0.00004)
W: % of housing crowded	0.0004***
	(0.0001)
W: % aged 25+ without high school diploma	−0.0001
	(0.0001)
Constant	−0.008**
	(0.003)
Observations	77
R^2	0.790
Adjusted R^2	0.751
Residual Std. Error	0.001 (df = 64)
F Statistic	20.113*** (df = 12; 64)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 6: SLX

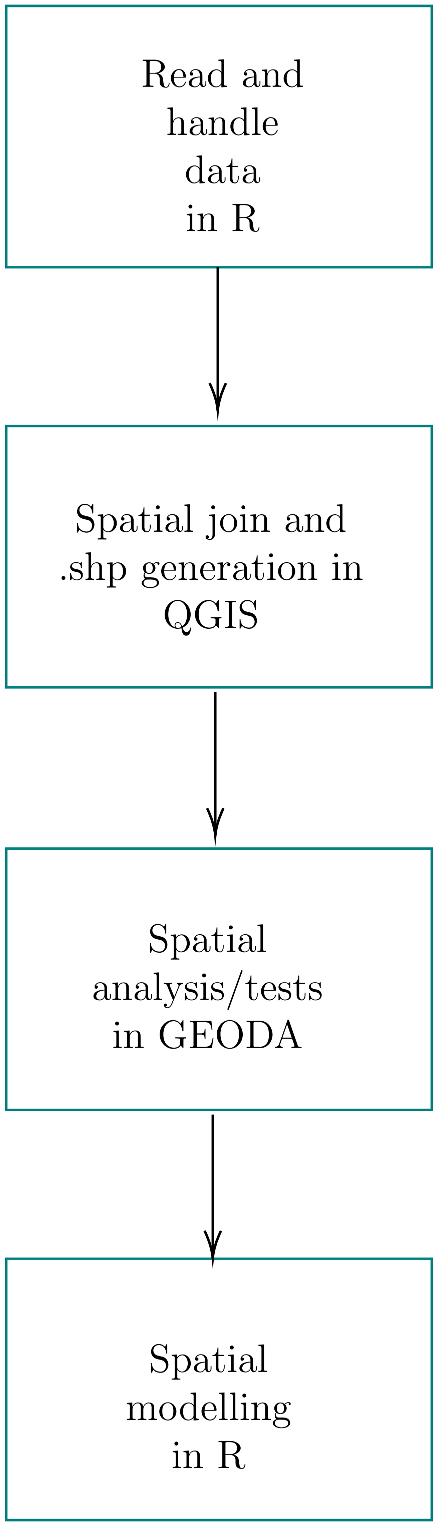

Figure 3: Summary of data handling

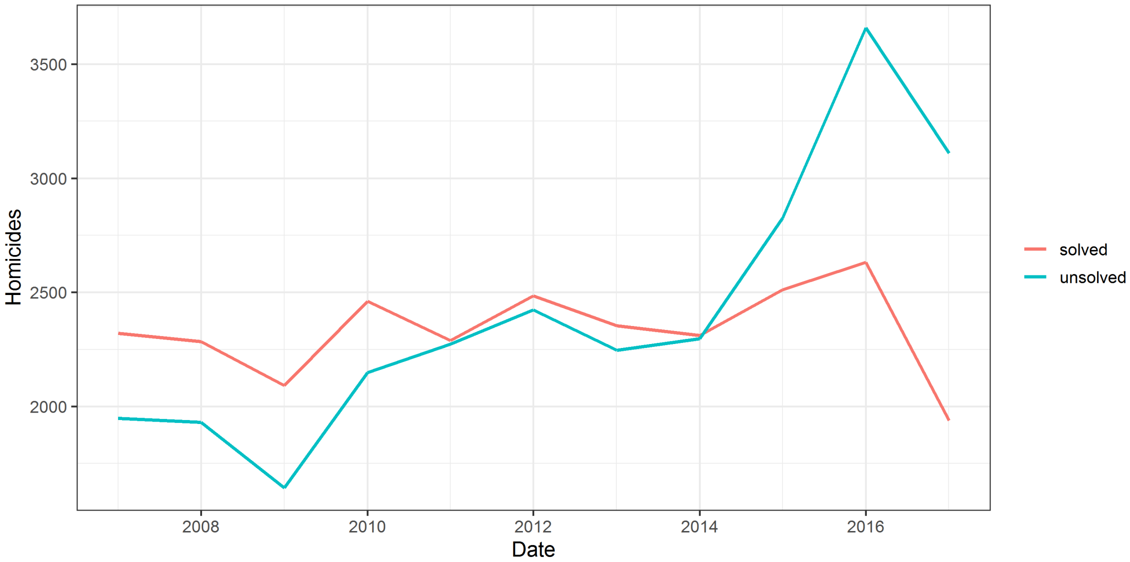

Figure 4: Evolution of solved and unsolved homicides

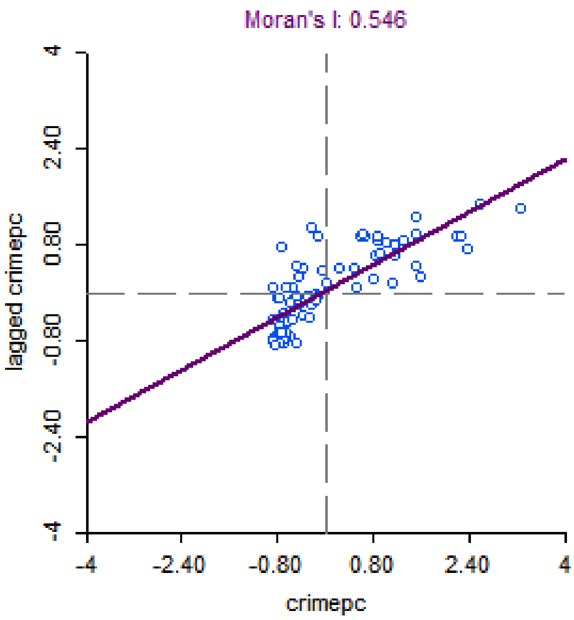

Figure 5: Moran's I scatterplot for Chicago aggregate data

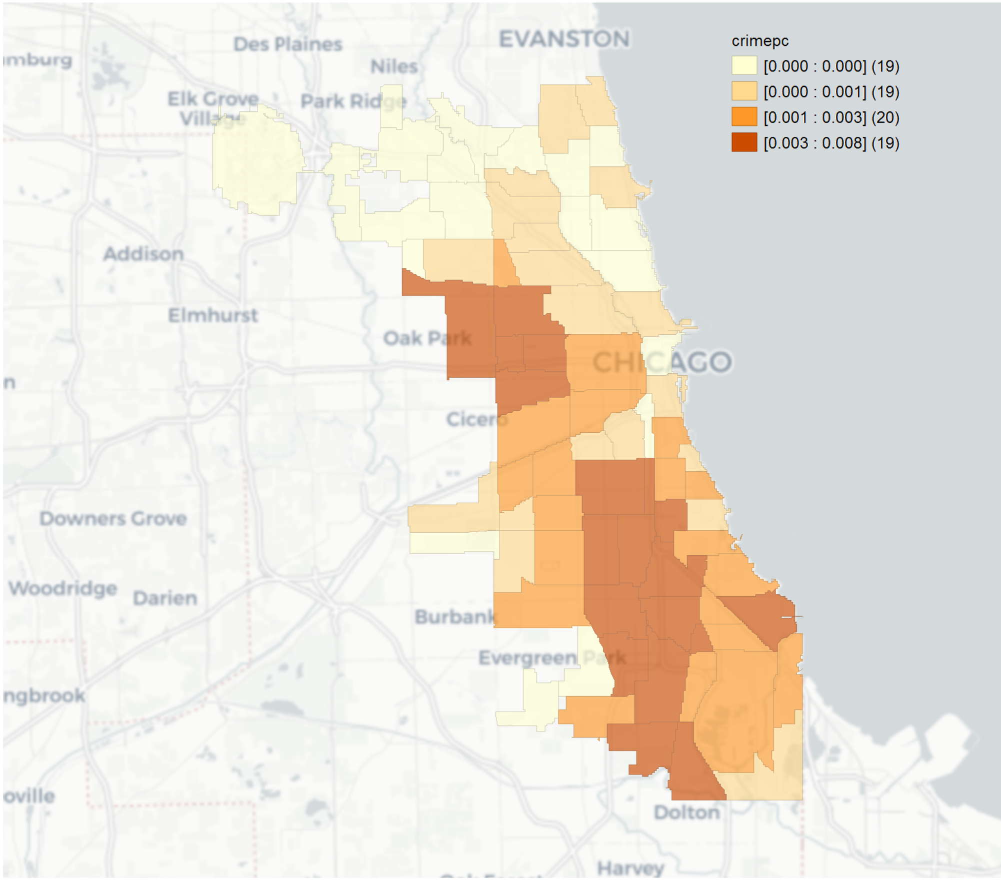

Figure 6: Unsolved crimes per capita in Chicago

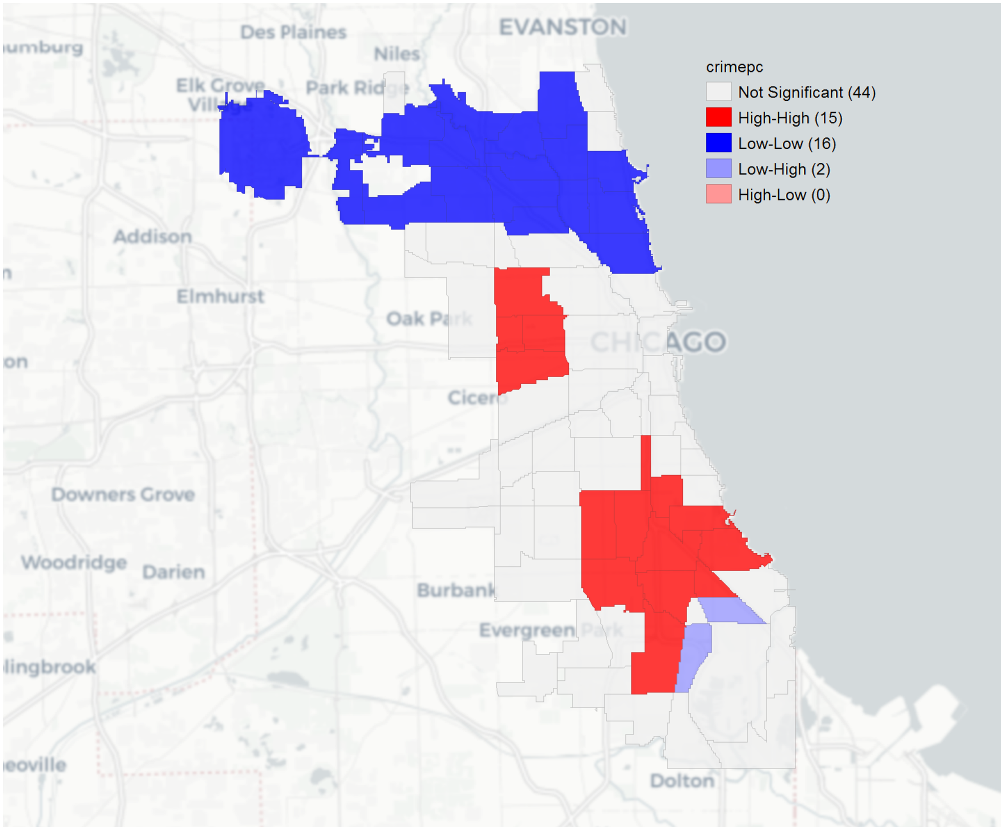

Figure 7: LISA cluster map with queen contiguity weight

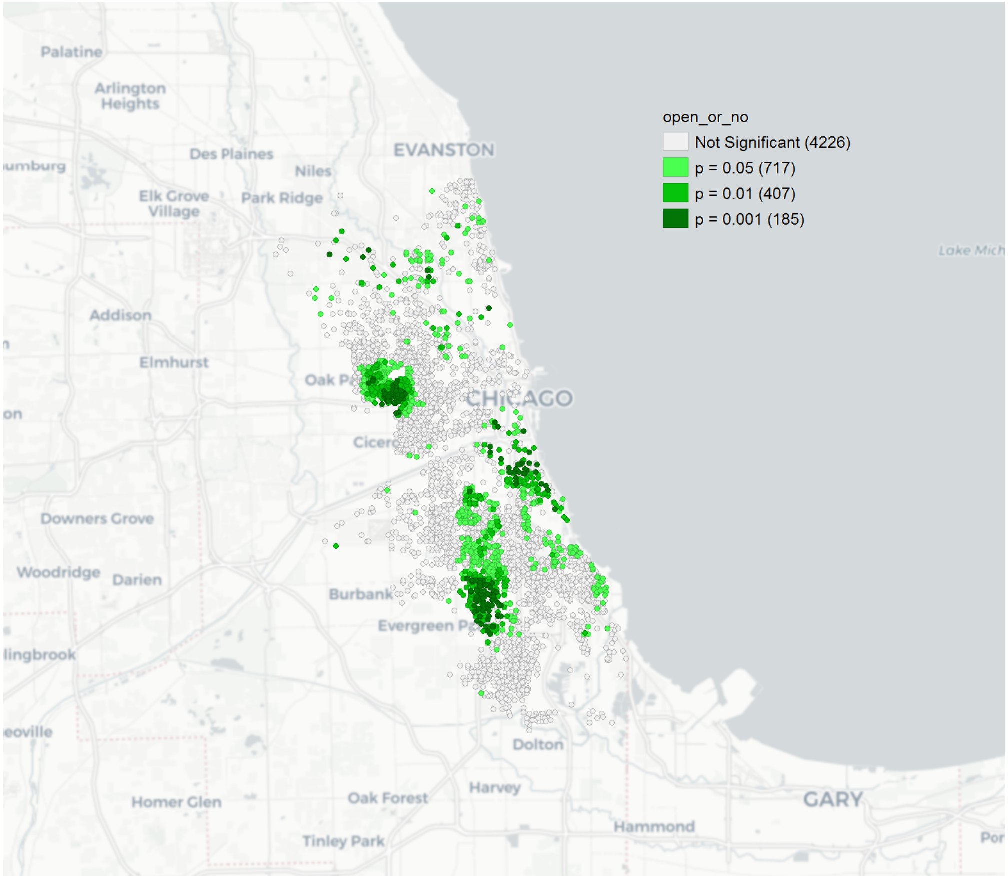

Figure 8: Join-count significance map for unsolved crimes in Chicago

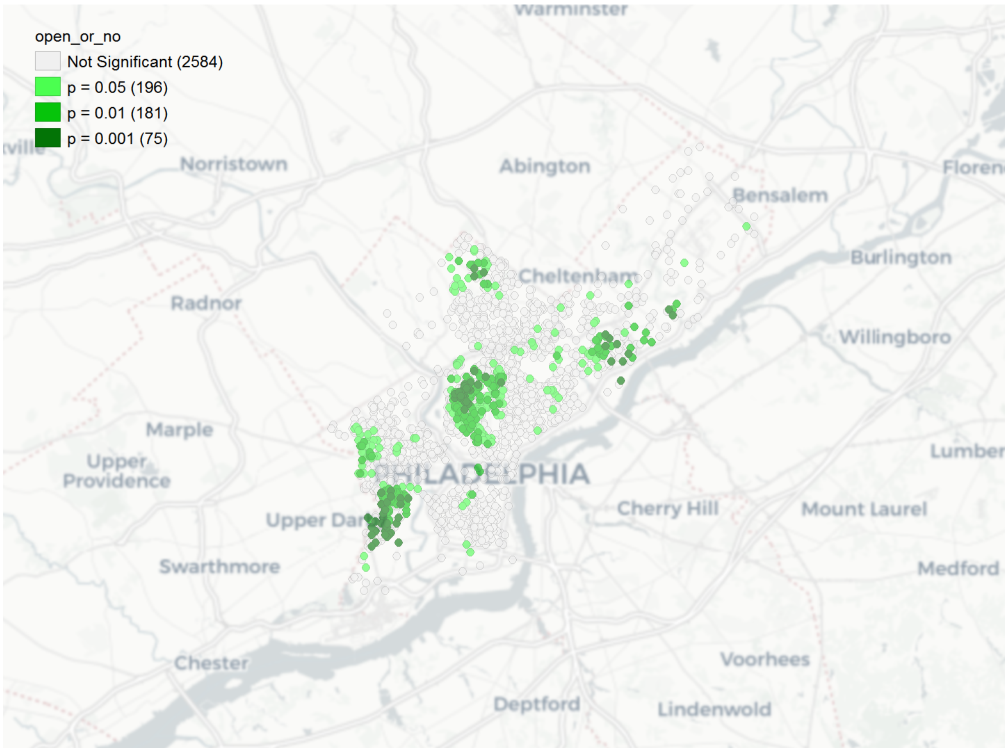

Figure 9: Join-count significance map for unsolved crimes in Chicago

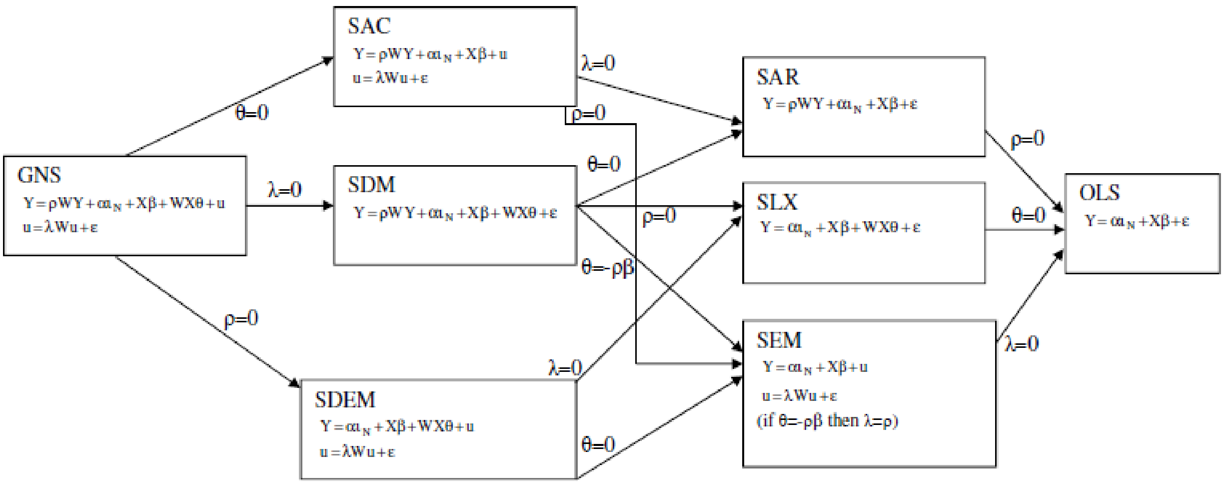

Figure 10: Relationship between the different models, source: Vega and Elhorst 2013