Spatiotemporal pattern of drug activities in Chicago

GEOG 788P course project report, Dec 2019

Zhiyue Xia

1. Introduction

Substance use disorder, referring to clinical impairments such as mental health problems and disabilities caused by alcohol or drugs (Center for Disease Control and Prevention, n.d.) affects millions of adolescents and adults in the United States and is an urban public health concern (Center for Behavioral Health Statistics and Quality, 2015). Research on spatial and temporal patterns drug activities and finding drug activities clusters helps understanding the hot zones of drug activities and assessing the availability of drug treatment facilities provides an opportunity to investigate the degree to which locations where drug-related activities are occurring in a city coincide with locations of substance use treatment services in order to ensure that treatment is available as close to locations of drug use as possible.

Substance disorder is an important public health issue and has been noted globally. Mounting previous geographical research on substance disorder have been conducted from various aspects. For example, previous research studied on the dynamics of drug overdose epidemic in the United States between 1979 and 2016 (Jalal et al., 2018). Another research has examine the spatial pattern of deaths involving heroin across the United States from 2000 through 2014 (Stewart et al., 2017). In a study investigating the spatial patterns of adolescent drug use in Cincinnati metropolitan region, Ohio, researchers analyzed student drug use survey data and found significant clustering pattern of high risk for drug use (Chaney and Rojas-Guyler, 2015). A study conducted in New York City found that built environment and neighborhood socioeconomic were associated with analgesic overdose fatalities (Cerdá et al., 2013). National Drug and Alcohol Research Centre at University of New South Wales (Australia) researched on the geographical injecting locations in Sydney for drug users and reported that nearly all drug users, to be specific 96% had injected in public spaces like streets, vehicles and public toilets (Darke et al., 2001); In Mexico, a study collected drug arrest data from police sectors and found college education, housing conditions and female-headed households are positively correlated with arrests for marijuana, but no sociodemographic correlates are significantly established for cocaine arrests (Vilalta, 2010).

In this project, I will take Chicago as a case study area. Chicago is located in Cook County, Illinoi, which has been identified as a crucial county by High Intensity Drug Traffic Areas (HIDTA) program. Chicago is a typical U.S. city that is an important place where drug activities are likely to occur. I will investigate the spatial distribution and temporal patterns of drug activities, mainly focus on heroin related death and arrests in Chicago between 2016 and 2019.

Three hypothesis were put forward in this project. The first one is that the spatial patterns of drug activities in city areas are significantly clustered; the second one is that drug arrests is an effective proxy of drug activities; the last one is that the spatial patterns of drug activities are correlated with built environment, demographic and socioeconomic factors.

2. Methods

2.1 Data

Drug activity data: drug arrests and opioid related death data

For this project, I used a public safety dataset provided by the Chicago data portal[1] that contains details of reported incidents of crime from 2001 to present for the City of Chicago, Illinois. The dataset includes 35 different types of crime such as arson, burglary, narcotics, weapons violations, and so on. Drug-related arrests during 2016, 2017, and 2018 were recorded based on two types of crime, 'narcotics' and 'other narcotic violations', and were extracted from the portal. For the reported incidents, the police recorded the specific reasons for the arrests such as possession, delivery and manufacture of drugs and also described the drug types. I summarized the drug arrests for all reasons and categorized the raw data by drug types.

Another drug activity data is opioid related death data that accessed from Medical Examiner Case Archive dataset provided by Cook County open data[2]. This dataset provides information about deaths such as time, location, reason from 2014 to present. The amount of death related data is much less than drug arrest data. Drug arrest data is more suitable for time series analysis because it includes sufficient data and is able to generate a more reliable spatial patterns in a fine time granularity. Therefore, drug related death data is used as a comparing dataset here to implement bivariate analysis to examine whether its pattern is similar to drug arrests, then test my hypothesis two: drug arrests is an effective proxy of drug activities.

---

[1] Chicago data portal https://data.cityofchicago.org/
[2] Cook County open data https://hub-cookcountyil.opendata.arcgis.com/

Previous research (Beckett et al., 2005; Cerdá et al., 2013; Darke et al., 2001; Lipton et al., 2013; Vilalta, 2010) has highlighted the role of socioeconomic and demographic factors for drug activity and in this study, median household income. EPA Smart Location Mapping[3] dataset provides some socioeconomic information such as number of low, median and high wage workers by block group, but some other potential factors that are found highly correlated with drug activities are not included in EPA dataset such as education and occupancy. To investigate the linkage between the spatial patterns of drug activities and the integrated neighborhood characteristics, we also collected and calculated demographic (e.g., percentage of population with college level education or higher), economic (e.g., median household income) and social factors (e.g., percentage of full time employ) from U.S. census. American Community Survey[4] (ACS) 5-year estimate data for 2017 at the granularity of block group was used for this research. Selected variables used in this project are shown in table 1.

Table 1. Selected variables used this project

	Variables	**Descriptive variables**
Economic	Income	Median household income, Number of household with income below poverty level...
Demographic	Employment	Total employment, Full time employment, Percent of population that is working aged…
	Education	Population of Bachelor's degree or higher…
	Race	White population, black population, Asian population, Hispanic or Latino population…
Built environment	Drug activity key location	Vacant building, vacant lot, gas station, alley, parking lot
	Urban design	Road network density, Street intersection density…
Transit	Transit distance	Distance from population weighted centroid to neatest transit stop, Proportion of employment within ¼ mile to transit stop….
	Access to workers	Jobs within 45 minutes auto travel time, Working age population within 45 minutes auto travel time…

[3] EPA Smart Location https://www.epa.gov/smartgrowth/smart-location-mapping
[4] U.S. census American Community Survey https://www.census.gov/programs-surveys/acs

2.2 Explore Spatial Data Analysis

The geospatial analysis in this study includes understanding the spatial patterns of opioid related deaths and heroin related arrests. The granularity of spatial analysis for this study is U.S. census block group. The frequency of reported incidents for drug activities during one year was aggregated by block group units and normalized by the population. To investigate whether the spatial pattern of drug activities is dispersed, clustered or random, Global Moran's I statistic was used to measure the spatial autocorrelation of drug activities (Li et al., 2007; Moran, 1950). Moreover, to find the hot zones for drug activities, Anselin Local Moran's I statistic was performed to identify hot spots that have high frequencies of drug arrests and were surrounded by neighborhoods that also have high frequencies (Anselin, 1995). The block group that was in a significant high-high value cluster at 95% confidence interval level then identified as a hotspot block group.

2.3 Bivariate analysis

In order to examine the similarity of spatial patterns of opioid related death and heroin related drug arrests, bivariate Moran's statistic (Anselin et al., 2006) is implemented. Bivariate analysis for these two type of drug related data assist in determining whether there is spatial autocorrelation between drug deaths and drug arrests.

2.4 Geographically weighted regression and multiscale geographically weighted regression

To investigate the quantitative relationship and interpret the correlation between independent factors (built environment, sociodemographic and economic variables) and drug activities. Geographically weighted regression (Brunsdon et al., 1998) and multiple scale geographically weighted regression (Oshan et al., 2019) were implement in Python. Five factors from different categories that are identified as the most important variables by preliminary Random Forest variable importance were input into GWR and MGWR. The results of GWR and MGWR are analyzed and compared in the result session. The dependent variable is frequency of drug arrest per 1000 population. The independent variables are VacantLot (number of vacant lots in the block group), ParkingLot (number of parking lots in the block group), PBachelorHigher (percentage of population with bachelor's or higher degree), Pct_LoMeWg (percent of low-medium wage workers have accessibility to public transit) and D5ar (jobs accessible by 45 min transit).

2.5 Random Forest: time series analysis

To capture the changing patterns of heroin, I aggregated the drug arrests to a finer time granularity of six months instead of one year. The time granularity was chosen by taking into account the frequency of drug arrests happened during each time period. I calculated 4 types of variables related to spatial and temporal autocorrelations, including temporal lagged variables, spatiotemporal lagged variables, trend of self-change and trend of neighborhood change for time lag = 1, 2, and 3.

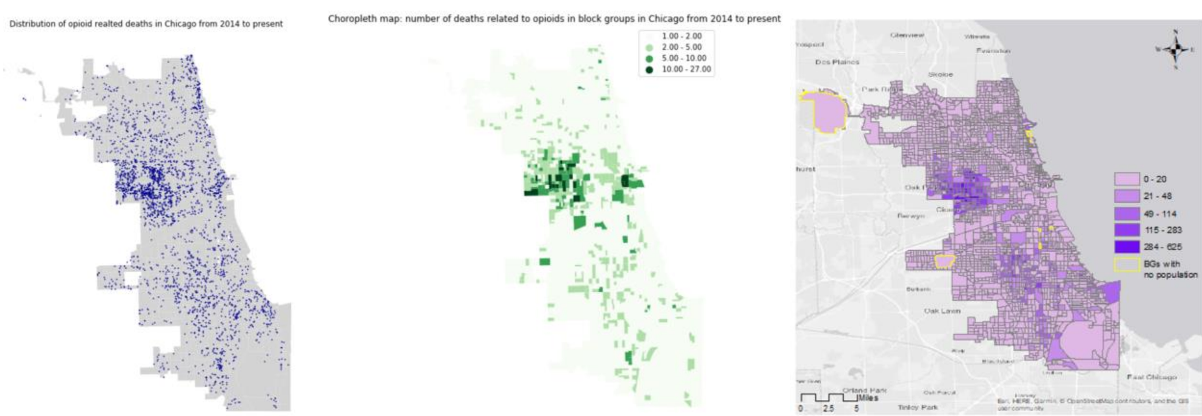
I used the Random Forests model (Breiman, 2001) to learn the geospatial patterns of heroin related activities. The model was trained drug arrest data between 2016 and 2018 was serve as the training dataset for spatial patterns of illicit drug activities in Chicago. The drug arrests for Jan-June 2019 was used as independent test dataset. Potential predictor variables were extracted from multisource datasets and were inputted into a Random Forest classifier to predict whether a block group is a hotspot of illicit drug activities or not.

3. Results

3.1 Spatial patterns of drug arrests and opioid related deaths

Between 2016 and 2018, the west side heroin hotspots involved neighborhoods mainly including Austin, Humboldt Park, West Garfield Park, East Garfield Park, North Lawndale, and a little spread to the north edge of South Lawndale and west edge of Near West Side neighborhood. The high clustering areas of opioid deaths (Figure 1 a, b) are similar to high drug arrest frequency clustering areas (Figure 1 c).

Two major heroin hotspots were identified in 2016 (Figure 2). The lower heroin hotspot around southwest side gradually disappeared during the past three years and only impacted two block groups in 2018 (Figure 2). The clustering pattern shifted from southwest side and west side to only west side indicates that heroin activities was changing to be centered in the west side during the three years.



(a) (b) (c)

Figure 1. (a) Distribution of opioid related deaths in Chicago between 2014 and 2019; (b) category map of opioid related deaths in Chicago between 2014 and 2019; (b) category map of heroin related arrests in Chicago between 2016 and 2018

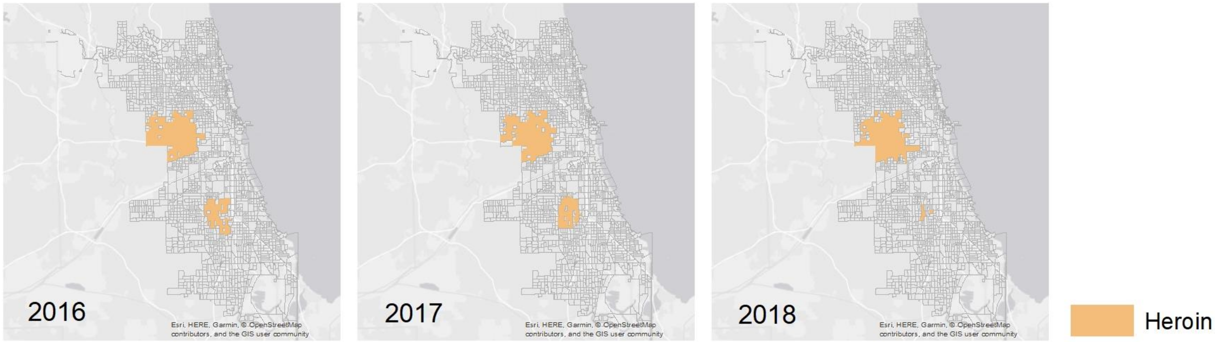
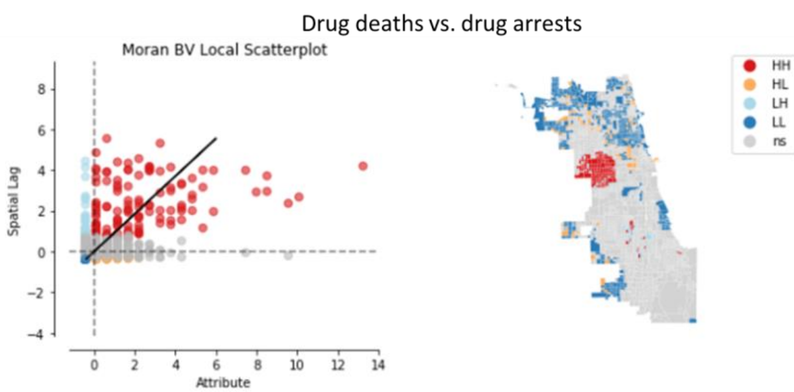


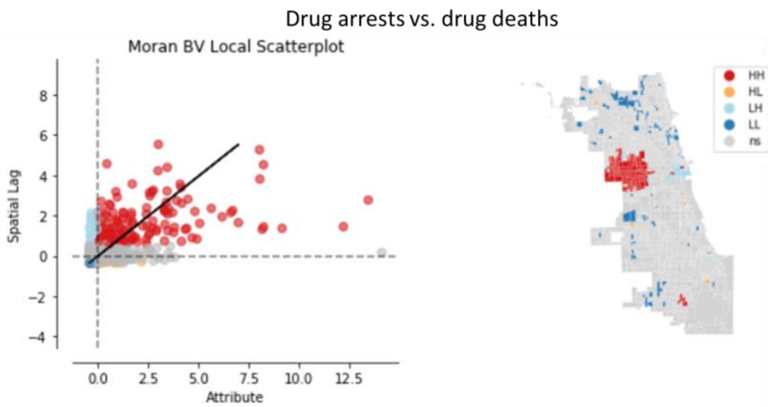
Figure 2. Hotspots of heroin related arrests in Chicago for 2016, 2017 and 2018

3.2 Bivariate analysis

Global scale bivariate analysis was implemented, the global bivariate Moran's I is 0.42 with p-value smaller than 0.001 that indicates positive correlation between death and arrest in nearby areas. To investigate the local patterns, local bivariate Moran's statistics are also implemented. Figure 3 a shows the local statistics of death surrounded by arrest. HH cluster indicate high death rate surrounded by high drug arrest. Figure 3 b shows arrest surrounded by death. HH cluster indicate high frequency of arrest surrounded by high death rate.



(a)



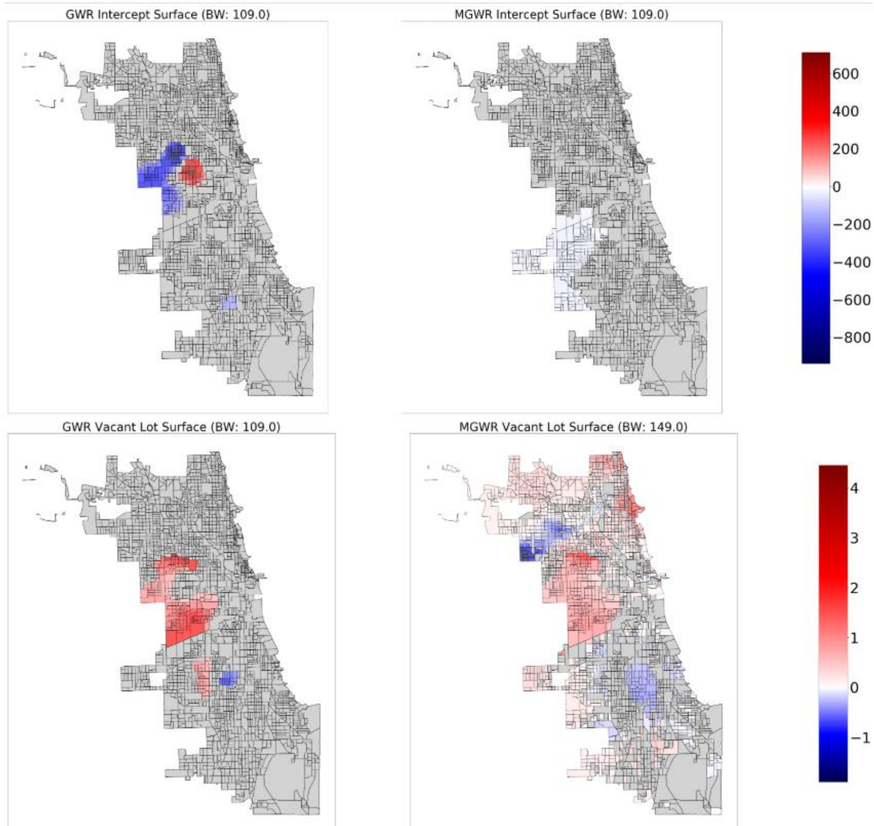
(b)

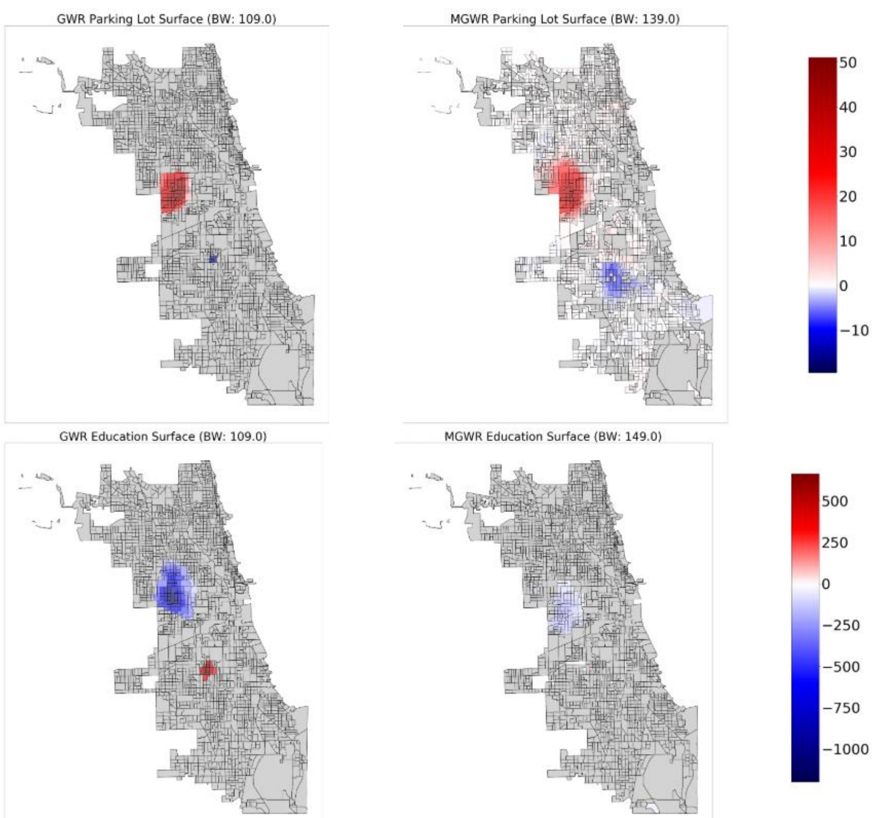
Figure 3. (a) Bivariate Moran's statistic between drug deaths and drug arrests; (b) Bivariate Moran's statistic between drug arrests and drug deaths

3.3 Results of GWR and MGWR

The variables input into GWR and MGWR include ehe dependent variable -- frequency of drug arrest per 1000 population, and the independent variables -- VacantLot (number of vacant lots in the block group), ParkingLot (number of parking lots in the block group), PBachelorHigher (percentage of population with bachelor's or higher degree), Pct_LoMeWg (percent of low-medium wage workers have accessibility to public transit) and D5ar (jobs accessible by 45 min transit).

Frequency of drug arrests is significantly correlated with some factors in some areas (Figure 4). In some block groups, especially near the upper hot spot area, the number of vacant lot is significantly positive correlated with drug arrests (Figure 4 b). While in some block groups, the percentage of population with Bachelor's degree or higher is significantly negative correlated with drug arrests (Figure 4 d). Many block groups are identified not significant correlated since a lot of block groups never occur drug arrests. Drug activities are highly crusted and concentrated in the hotspots. Logistic regression and classification may perform better in this problem.


GWR Intercept Surface (BW: 109.0)
MGWR Intercept Surface (BW: 109.0)
600
400
200
0
−200
−400
−600
−800
GWR Vacant Lot Surface (BW: 109.0)
MGWR Vacant Lot Surface (BW: 149.0)
4
3
2
1
0
−1


GWR Parking Lot Surface (BW: 109.0)
MGWR Parking Lot Surface (BW: 139.0)
50
40
30
20
10
0
−10
GWR Education Surface (BW: 109.0)
MGWR Education Surface (BW: 149.0)
500
250
0
−250
−500
−750
−1000

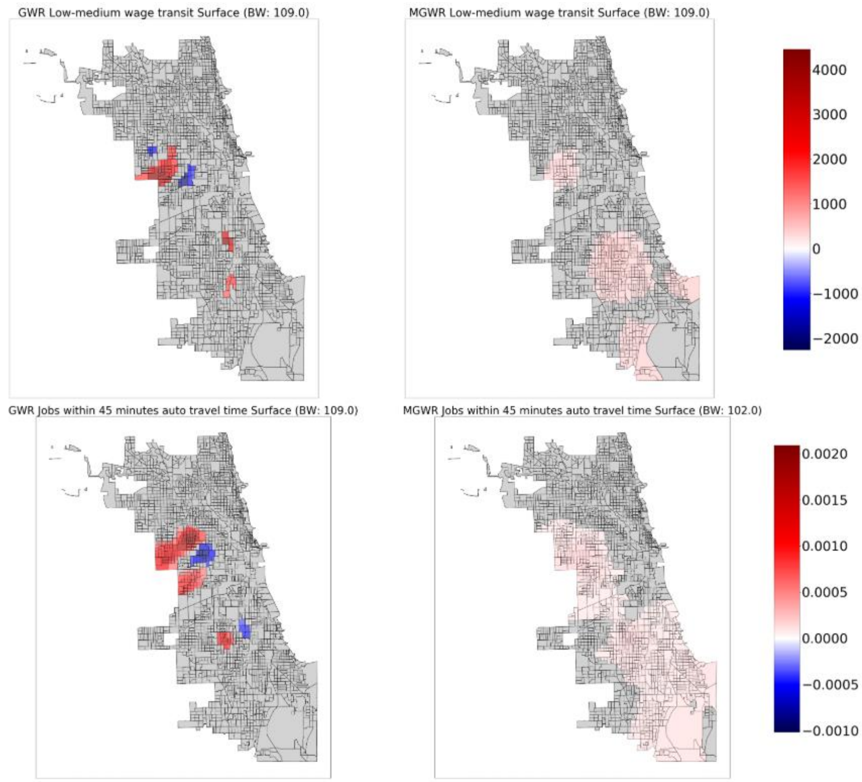


Figure 4. Coefficient surfaces of GWR and MGWR: (a) intercept (b) vacant lot (c) parking lot (d) percentage of Bachelor's degree or higher (e) percentage of low-medium wage workers have accessibility to public transit (f) jobs accessible by 45 min transit

3.4 Changing patterns of drug activities

Time series analysis were implemented in R since I chose use Random Forest model for this project and R has more derivative Random Forest package such as 'ranger','pdp' as well as geographically Random Forest in 'spatialML' that includes more functions than python 'scikit-learn'.

To capture the changing pattern of heroin hotspots over time. I trained a predictive Random Forest model by inputting built environment, sociodemographic, and economic variables and using different time lag (time 1, 2, and 3) and fitting the data between 2016 and

2018. Then predict heroin hotspots in 2019 and compare it with the real calculated hotspots in 2019 (Figure 5). The model using time lag 3 perform best among all models.

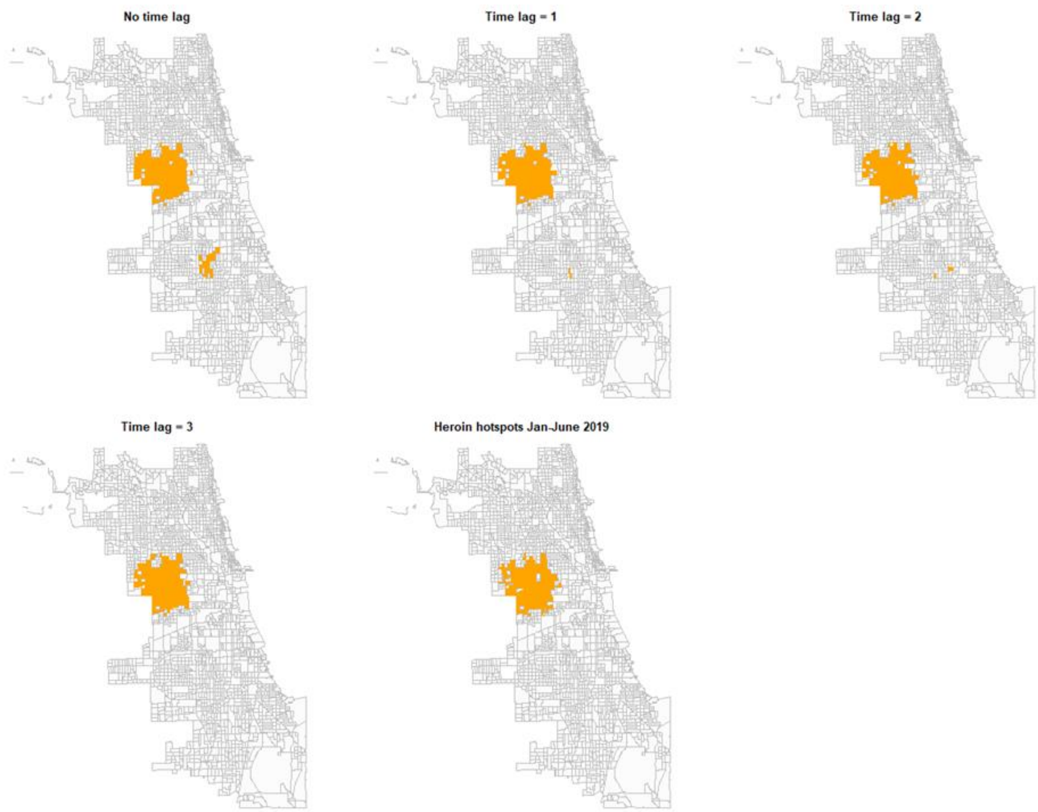


Figure 5. Rnaom Forest model prediction maps of heroin hotspots for Jan-June 2019 not using time lag variables, using time lag 1, 2 and 3, and real heroin hotspots map for Jan-June 2019.

The top 20 most important variables (Figure 6) include 6 spatiotemporal variables for predicting heroin hotspots, such as number of neighboring block groups belonged to a hotspot during t-1, t-2, t-3 period and binary variable whether the given block group belonged to a hotspot during the past periods, as well as that changing trend of this block groups and changing trend in the neighborhood. Some transit variables, for examples, working age population within

45 minutes to works and percentage of medium-wage workers accessible by transit ranked high among all variables.

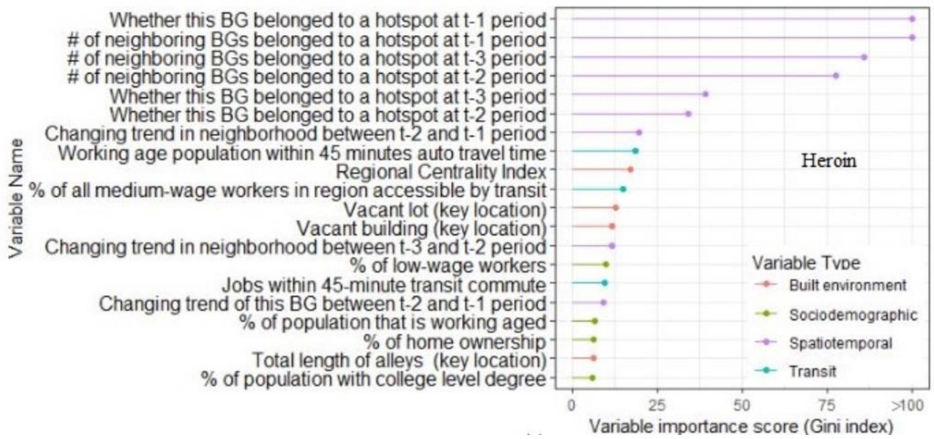


Figure 6. Top 20 most important variables for predicting heroin hotspots

4. Discussion, conclusion and future works

At the beginning of the project, I imagined two challenges in this project as I wrote in the project proposal. 1) Drug activates are episodic and difficult to capture the temporal patterns. 2) The spatial patterns of different drug types are varied so it is hard to explain why some built environment factors may have different impacts on different drug type activities. To solve problem 1, I use a finer time granularity of six months instead of one year to do time series analysis. A finer time granularity provides subtle changing trends and more training sample in a given time period. To solve problem 2, I fellow Dr. Oshan's suggestion in my first project update, I limited my research object to heroin drug instead of all drug arrest which may include cocaine, marijuana etc. Therefore, I can find the driving factors for changing patterns of heroin.

Three hypotheses: spatial patterns of drug activities in city areas are significantly clustered; drug arrests is an effective proxy of drug activities; and the spatial patterns of drug activities are correlated with built environment, demographic and socioeconomic factors all have been tested and the null hypotheses were rejected.

Research on spatial and temporal patterns drug activities and finding drug activities clusters helps understanding the hot zones of drug activities and assessing the availability of drug treatment facilities provides an opportunity to investigate the degree to which locations where drug-related activities are occurring in a city coincide with locations of substance use treatment services in order to ensure that treatment is available as close to locations of drug use as possible. I will continue with this project. Future works will include spatial accessibility of opioid treatment centers to heroin hotspots.

Acknowledgement

Research reported in this paper is the final project for course GEOG 788P Models & Methods for Spatial Data Science, FALL 2019. The author would like to thank the course instructor, Dr. Taylor Oshan, and great peers, Jeffery Sauer, Zheng Liu, Jamis Bruening, Brian Slobotsky, Kaleb Tsegaye and her boyfriend Mofeng Yang for their insights and suggestions for improving this work. The content is solely the responsibility of the author and does not necessarily represent the official views of this course.

Reference

Anselin, L., 1995. Local Indicators of Spatial Association—LISA. Geogr. Anal. 27, 93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x

Anselin, L., Syabri, I., Kho, Y., 2006. GeoDa : An Introduction to Spatial Data Analysis. Geogr. Anal. 38, 5–22. https://doi.org/10.1111/j.0016-7363.2005.00671.x

Beckett, K., Nyrop, K., Pfingst, L., Bowen, M., 2005. Drug Use, Drug Possession Arrests, and the Question of Race: Lessons from Seattle. Soc. Probl. 52, 419–441. https://doi.org/10.1525/sp.2005.52.3.419

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324

Brunsdon, C., Fotheringham, S., Charlton, M., 1998. Geographically Weighted Regression. J. R. Stat. Soc. Ser. Stat. 47, 431–443. https://doi.org/10.1111/1467-9884.00145

Center for Behavioral Health Statistics and Quality, 2015. Behavioral Health Trends in the United States: Results from the 2014 National Survey on Drug Use and Health 64.

Center for Disease Control and Prevention, n.d. Substance Use Disorder Treatment in Persons Who Inject Drugs | CDC [WWW Document]. URL https://www.cdc.gov/pwid/substance-treatment.html (accessed 8.9.19).

Cerdá, M., Ransome, Y., Keyes, K.M., Koenen, K.C., Tardiff, K., Vlahov, D., Galea, S., 2013. Revisiting the Role of the Urban Environment in Substance Use: The Case of Analgesic Overdose Fatalities. Am. J. Public Health 103, 2252–2260. https://doi.org/10.2105/AJPH.2013.301347

Chaney, R.A., Rojas-Guyler, L., 2015. Spatial patterns of adolescent drug use. Appl. Geogr. 56, 71–82. https://doi.org/10.1016/j.apgeog.2014.11.002

Darke, S., Kaye, S., Ross, J., 2001. Geographical injecting locations among injecting drug users in Sydney, Australia. Addiction 96, 241–246. https://doi.org/10.1046/j.1360-0443.2001.9622416.x

Jalal, H., Buchanich, J.M., Roberts, M.S., Balmert, L.C., Zhang, K., Burke, D.S., 2018. Changing dynamics of the drug overdose epidemic in the United States from 1979 through 2016. Science 361. https://doi.org/10.1126/science.aau1184

Li, H., Calder, C.A., Cressie, N., 2007. Beyond Moran's I: Testing for Spatial Dependence Based on the Spatial Autoregressive Model. Geogr. Anal. 39, 357–375. https://doi.org/10.1111/j.1538-4632.2007.00708.x

Lipton, R., Yang, X., A. Braga, A., Goldstick, J., Newton, M., Rura, M., 2013. The Geography of Violence, Alcohol Outlets, and Drug Arrests in Boston. Am. J. Public Health 103, 657–664. https://doi.org/10.2105/AJPH.2012.300927

Moran, P.A.P., 1950. Notes on Continuous Stochastic Phenomena. Biometrika 37, 17–23. https://doi.org/10.2307/2332142

Oshan, T.M., Li, Z., Kang, W., Wolf, L.J., Fotheringham, A.S., 2019. mgwr: A Python Implementation of Multiscale Geographically Weighted Regression for Investigating Process Spatial Heterogeneity and Scale. ISPRS Int. J. Geo-Inf. 8, 269. https://doi.org/10.3390/ijgi8060269

Stewart, K., Cao, Y., Hsu, M.H., Artigiani, E., Wish, E., 2017. Geospatial Analysis of Drug Poisoning Deaths Involving Heroin in the USA, 2000–2014. J. Urban Health 94, 572–586. https://doi.org/10.1007/s11524-017-0177-7

Vilalta, C.J., 2010. The spatial dynamics and socioeconomic correlates of drug arrests in Mexico city. Appl. Geogr. 30, 263–270. https://doi.org/10.1016/j.apgeog.2009.06.001