Shoplifting Trends in Time and Space: A Study of Two Major American Cities

Methodology

The analysis, entitled Shoplifting Trends in Time and Space: A Study of Two Major American Cities, examined recent trends in reported shoplifting in Chicago, IL, and Los Angeles, CA. The reported shoplifting data as well as data for other offenses in this report were drawn from incident-level data from January 2018 to December 2023 taken directly from law enforcement agency or city websites. Study cities were selected based on the availability of incident-level, geocoded data on their online portals. Finally, the choice of cities was also intended to represent different geographic contexts. The values may differ from data published by individual police departments due to data updating over time and from official counts released by the FBI. For the most up-to-date information for a specific city, please visit its website. Shapefiles for city block groups came from the United States Census Bureau[1] and neighborhood boundary shapefiles were from city websites.[2] Population measures were also from the United States Census Bureau.[3]

The retail outlet data in this report came from city datasets for business licenses. Both datasets were available from city data portal websites.[4] The data provided general descriptions of business activity but did not provide a direct way to differentiate businesses in terms of their size or potential for reported shoplifting or other activity. We included all retail business locations that could experience reported shoplifting, regardless of the actual likelihood. To meet this criteria, businesses required a physical location and a retail license. Finally, we established a retail location as operating in a given year if that location had an active business license for at least six months of the year.

The maps provided in this report use a method known as Kernel Density Estimation (KDE) to produce visualizations of the level of crime across a geographic study unit. KDE was performed and all maps were generated using ArcPro 3.2. KDE generates a grid

---

[1] United States Census Bureau. (n.d). *TIGER/line shapefiles*. https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html

[2] Chicago Data Portal. (n.d.). *City of Chicago*. https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-Neighborhoods/bbvz-uum9; City of Los Angeles. (n.d.). https://geohub.lacity.org/search?collection=Dataset&q=neighborhoods

[3] United States Census Bureau. (n.d.). *Explore census data*. https://data.census.gov/

[4] Chicago Data Portal. (2024). *Business licenses*. https://data.cityofchicago.org/Community-Economic-Development/Business-Licenses/r5kz-chrr/about_data; City of Los Angeles. (2024). *Listing of all businesses*. https://data.lacity.org/Administration-Finance/Listing-of-All-Businesses/r4uk-afju/about_data

across the study site and calculates the density of incidents occurring around each grid cell. KDE fits a symmetrical, smoothed curved surface—a kernel—over each incident point location, where the value of the surface is greatest directly at the location of the point incident and decreases as the distance from the point increases. For each cell in the grid, the density of incidents is calculated by adding together the values of all surfaces in the center point of the cell.

When KDE is used, there are several parameters that must be set: the desired kernel function, radius/bandwidth, and cell size. Multiple methods can be used to generate the kernel function. This analysis used the quartic kernel function, which specifies that the intensity of the kernel surface diminishes gradually as the distance from the point incident increases. The bandwidth in our analysis was selected based on Silverman's Rule-of-thumb formula,[5] We generated a cell size using the ArcPro 3.2 default setting, which calculates cell size by dividing the shorter side of the study area by 250.[6] Additionally, KDE results are dependent upon how they are presented. In this report, we used a method suggested by Chainey and colleagues which uses incremental multiples of the grid cells' mean value to visually present results.[7] The average density for all non-zero grid cells is calculated, and maps are altered to represent multiples of that mean value from one to greater than five times the mean density value. This method has the advantage of presenting visually appealing results and being intuitive and easy to understand and read.[8]

However, this method also requires somewhat arbitrary, or at least heuristic decisions. As a result, the same analysis could be performed using different but still reasonable parameters, which could result in slightly different conclusions or numbers. Thus, the count of incidents and location of "clusters" in this report are not precise values but aggregations across space using the chosen parameters of the method. They are better suited to speak to larger patterns across space and time rather than exact values or demarcation of places at a small scale.

The analysis of clusters is limited to areas with at least twice the mean level of reported shoplifting. This is done two reasons. First, examining areas with the highest concentration of reported shoplifting, and areas with average to twice the average amount of reported shoplifting may constitute "concentrated" reported shoplifting rates. Second, for many years (for both cities) there were a multitude of areas with an average to twice the average amount of reported shoplifting. By limiting the definition of a reported shoplifting cluster to areas with at least twice the average amount of reported

---

[5] Silverman, B. (1986). *Density estimation for statistics and data analysis*. Chapman and Hall.
[6] ESRI. (n.d.). *How kernel density works*. https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/how-kernel-density-works.htm
[7] Chainey, S., Reid, S., & Stuart, N. (2002). When Is a hotspot a hotspot? A procedure for creating statistically robust Hotspot maps of crime. In D. Kidner, G. Higgs, & S. White (Eds.), *Socio-economic applications of geographic information science*, pp. 20-35. CRC Press. https://doi.org/10.1201/b12606
[8] Eck, J., Chainey, S., Cameron, J., Leitner, M, & Wilson, R. (2005). *Mapping crimes: Understanding hot spots. National Institute of Justice special report*. https://www.ojp.gov/pdffiles1/nij/209393.pdf

shoplifting, a smaller number of areas emerged as clear concentrated clusters of reported shoplifting. The same logic applies to the analysis of retail outlets.
The mean level of reported shoplifting (and retail outlets) was calculated for each city separately. It is also calculated across the entire study period (2018-2023). This was done to examine how reported shoplifting was concentrated in the cities across space and time. This is a more holistic and pragmatic way to see reported shoplifting trends, rather than normalizing reported shoplifting to each year of occurrence.

Regression Analyses

METHODOLOGY

In order to examine larger-scale relationships between reported shoplifting and census tracts, regression analyses employed multilevel linear and negative binomial regression models. These models allow us to better test whether the relationships seen in the spatial analyses are in fact statistically significant, though at a different level of geography. Two main dependent variables were analyzed: *total reported shoplifting* and *reported shoplifting proportion*. *Total reported shoplifting* is the number of reported shoplifting incidents that occurred in the tract. *Reported shoplifting proportion* is a measure of the proportion of the city's reported shoplifting incidents for a particular year that occur in a particular Census tract. To construct this variable, we take the total number of reported shoplifting incidents that occur within a particular tract and divide it by the total number of reported shoplifting incidents that occurred in the entire city. This proportion is calculated for each year included in the analysis. This variable is designed to estimate the share of a city's reported shoplifting incidents that occur in one tract. Stated differently, it tells us whether the tract experienced a larger or smaller share of the city's reported shoplifting incidents relative to other tracts. This mirrors the KDE estimates of the concentration of reported shoplifting events, but places them within set geographical boundaries to be able to use in a regression analysis capable of estimating statistical significance.

Models also included a measure of the proportion of *retail business licenses* in a particular tract relative to the total number of business licenses in the city. Depending on the question being examined, this variable was at times swapped for a total number of licenses. These corresponded to whether the dependent variable was a proportion or a total. Additionally, a lagged variable for tract reported shoplifting offenses, where prior year's other offense numbers predicted the current year's reported shoplifting level for *violent crime*, *burglary*, *motor vehicle theft*, and *non-reported shoplifting larceny*. Lastly, a time measure representing *year* was also included.

Two kinds of random intercept multilevel models were employed, based on the dependent variable utilized in the analysis. The models predicting the reported shoplifting proportion variable employed linear multilevel regression models. The models predicting the total number of reported shoplifting incidents used negative binomial multilevel regression models, as the outcome variable was heavily skewed to lower values. In each of these multilevel models, year (level 1) is nested within Census tract (level 2). This means that in these models, tracts are compared to themselves over time. This controls for time stable differences between tracts that may otherwise influence the within-tract estimates, such as differences across tracts in population, poverty, racial heterogeneity, and other traditionally examined neighborhood-level variables. The portion of the models of importance for this particular analysis are the within-tract estimates, which are similar to fixed effects models, but have been shown to instead be a more efficient modeling strategy.[9] Models were estimated separately for Chicago and Los Angeles.

RESULTS

Table S1 presents the regression coefficients for the models predicting the total reported shoplifting in a tract. For Los Angeles, when a tract has more retail business licenses (compared to itself at times when there are fewer) the tract experiences significantly higher counts of reported shoplifting. With the exception of motor vehicle theft, at times when tracts have higher counts of other types of crime, they also experience significantly higher counts of reported shoplifting. However, the overall magnitude of these relationships is small. Within Los Angeles tracts, shoplifting counts were not significantly different in 2019 compared to 2018 after controlling for the other factors in the model. Tracts in 2020-2023 experienced significantly lower rates of shoplifting compared to the amount of shoplifting within each tract in 2018.

In Chicago, we do not find a similar significant effect of the number of business licenses in a tract at a given time point. Here, when motor vehicle theft and violent crime are higher, there are significantly more shoplifting incidents in a given tract. The measure of the Level 2 variance indicates that in both cities, there is significant variation between tracts for both cities in the total amount of reported shoplifting incidents, which is unsurprising given the results of the KDE analysis. When looking across time, compared to an individual tract's shoplifting levels in 2018, we find that counts in 2019 and 2020 were significantly higher. However, from 2021-2023 the amount of shoplifting was significantly lower to levels within the same tract in 2018.

The coefficients comparing shoplifting rates within tracts over time in both cities further highlight the findings of the KDE analysis that shoplifting, and therefore the increases and decreases in the rates over time, are highly concentrated. The models above estimate within-tract change, ultimately showing the average within-tract change across all tracts within each city. The differences in these relationships compared to the descriptive changes in the counts of monthly shoplifting incidents shown in Figure 1 in

[9] Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods.* Sage Publications.

the main report, highlight that these changes over time occurred in a small number of the total number of tracts in both cities.

Table S1. Multilevel Negative Binomial Regression of Census Tract Reported Shoplifting Counts

	Chicago			Los Angeles		
	Coef	SE		Coef	SE	
Business license proportion	-0.190	0.016	***	-0.030	0.056	
Burglary rate	-0.000	0.000	***	0.000	0.000	***
MVT rate	0.000	0.000	***	-0.000	0.000	
Non-retail theft larceny	0.000	0.000	***	0.000	0.000	***
Violent crime	0.000	0.000		0.000	0.000	***
Year						
2019	-0.000	0.000		0.000	0.000	
2020	0.000	0.000		0.000	0.000	***
2021	0.000	0.000		0.000	0.000	**
2022	-0.000	0.000		0.000	0.000	*
2023	0.000	0.000	**	0.000	0.000	**
Constant	0.000	0.000	**	-0.000	0.000	*
Level 2 variance	0.000	0.000		0.000	0.000	

** p<.05; ** p<.01; *** p<.001*

Supplemental Tables

The following are the offense tables that support the findings on shoplifting areas and other crimes.

Table S2. Reported Shoplifting and Other Offenses in Chicago, 2018

Cluster	Reported shoplifting Incidents per square mile	Violent Offenses per Square Mile	Robbery Offenses per Square Mile	Motor Vehicle Theft per Square Mile	Theft Offenses per Square Mile	Burglary Offenses per Square Mile
A	1,202	1,404	242	138	3,171	71
B	1,357	735	109	78	1,326	44
C	12	466	71	80	289	104
D	96.6	341	63	67	578	133
Chicago	47	354	43	45	234	52

Table S3. Reported Shoplifting and Other Offenses in Chicago, 2020

Cluster	Reported shoplifting Incidents per square mile	Violent Offenses per Square Mile	Robbery Offenses per Square Mile	Motor Vehicle Theft per Square Mile	Theft Offenses per Square Mile	Burglary Offenses per Square Mile
A	941	1,353	217	213	1,494	404
B	-	-	-	-	-	-
C	17	391	43	49	275	77
D	78	253	58	69	347	83
Chicago	28	300	35	44	154	39

Table S4. Reported Shoplifting and Other Offenses in Chicago, 2023

Cluster	Reported shoplifting Incidents per square mile	Violent Offenses per Square Mile	Robbery Offenses per Square Mile	Motor Vehicle Theft per Square Mile	Theft Offenses per Square Mile	Burglary Offenses per Square Mile
A	1,347	78	115	141	750	42
B	1,100	190	37	63	277	27
C	21	220	20	67	173	46
D	85	177	41	139	225	44
Chicago	42	160	20	64	96	16

Table S5. Reported Shoplifting and Other Offenses in Los Angeles, 2018

Cluster	Reported shoplifting Incidents per square mile	Violent Offenses per Square Mile	Robbery Offenses per Square Mile	Motor Vehicle Theft per Square Mile	Theft Offenses per Square Mile	Burglary Offenses per Square Mile
E	132	67	32	55	262	47
F	126	496	226	126	1,047	106
G	152	28	15	20	186	46
H	128	73	46	32	391	99
L.A.	14	52	22	37	136	34

Table S6. Reported Shoplifting and Other Offenses in Los Angeles, 2020

Cluster	Reported shoplifting Incidents per square mile	Violent Offenses per Square Mile	Robbery Offenses per Square Mile	Motor Vehicle Theft per Square Mile	Theft Offenses per Square Mile	Burglary Offenses per Square Mile
E	99	36	23	41	170	30
F	417	667	276	276	1,731	305
G	840	150	90	45	340	125
H	-	-	-	-	-	-
L.A.	14	49	17	44	104	29

Table S7. Reported Shoplifting and Other Offenses in Los Angeles, 2023

Cluster	Reported shoplifting Incidents per square mile	Violent Offenses per Square Mile	Robbery Offenses per Square Mile	Motor Vehicle Theft per Square Mile	Theft Offenses per Square Mile	Burglary Offenses per Square Mile
E	188	72	29	46	199	46
F	203	419	156	299	822	134
G	218	46	29	32	155	44
H	326	72	36	76	280	83
L.A.	26	53	19	57	121	32