POI DATA VERSUS LAND USE DATA, WHICH IS MORE EFFECTIVE IN MODELING THEFT CRIMES?

JIAJIA FENG[1], YUEBING LIANG[2], QI HAO[3], KE XU[4] and WAISHAN QIU[5]
[1,3,4]Department of Architecture, Tsinghua University, Beijing, China.
[2]Department of Urban Planning and Design, The University of Hong Kong, Hong Kong
[5]Department of City and Regional Planning, Cornell University, Ithaca, NY, USA
[1]jiajiafsz@126.com, 0000-0003-0931-2767
[2]yuebingliang@connect.hku.hk, 0000-0003-2089-4606
[3]hao-q14@tsinghua.org.cn, 0000-0002-6063-0018
[4]xk_thu@163.com, 0000-0003-1286-9777
[5]wq43@cornell.edu, 0000-0001-6461-7243

Abstract. Alleviating crime and improving urban safety is important for sustainable development of society. Prior studies have used either land use data or point-of-interests (POI) data to represent urban functions and investigate their associations with urban crime. However, inconsistent and even contrary results were yielded between land use and POI data. There is no agreement on which is more effective. To fill this gap, we systematically compare land use and POI data regarding their strength as well as the divergence and coherence in profiling urban functions for crime studies. Three categories of urban function features, namely the density, fraction, and diversity, are extracted from POI and land use data, respectively. Their global and local strength are compared using ordinary least square (OLS) regression and geographically weighted regression (GWR), with a case study of Beijing, China. The OLS results indicate that POI data generally outperforms land use data. The GWR models reveal that POI Density is superior to other indicators, especially in areas with concentrated commercial or public service facilities. Additionally, Land Use Fraction performs better for large-scale functional areas like green space and transportation hubs. This study provides important reference for city planners in selecting urban function indicators and modelling crimes.

Keywords. POI; Land Use; Urban Functions; Theft crime; Predictive Power; SDG 16.

1. Introduction

In 2015, the United Nations General Assembly (2015) set up seventeen Sustainable Development Goals, one of which is to “promote peaceful and inclusive societies for

POST-CARBON, Proceedings of the 27th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2022, Volume 1, 425-434.

sustainable development". To achieve this goal, the United Nations aims at reducing all kinds of urban crimes, including violence, illicit financial flows and theft, by 2030. Therefore, investigating how to control urban crime is extremely important for realizing a more sustainable future for global citizens.

Under the tendency of finer-grained planning, neighbourhood-level crime study has become a heated topic. In this urban level, neighbourhood functions have long been revealed to significantly correlate with crime occurrence. Understanding the impact of neighbourhood functions on crime is crucial for urban planners to avoid potential high crime areas. In solving the problem, an increasing number of researches have emerged in the wake of high-resolution urban data. Land use data, which represents urban land types as polygons, is most commonly used to represent neighbourhood functions in urban crime studies. Most crime studies use land use fractions, i.e., the percentage of land use area to the neighbourhood, as the indicator of neighbourhood functions. For instance, Stucky and Ottensmann (2009) found that robberies are more common in neighbourhoods with larger fraction of commercial area. Sohn (2016) discovered that residential areas were negatively related to crime density in the neighbourhood level. Recent studies have also examined how land use mix affects crime patterns. For example, a Herfindahl index of land uses was constructed by Wo (2019) for the Los Angeles neighbourhoods to capture mixed land use and assess the effect of mixed land use on crime. De Nadai et al. (2020) used the average entropy among land uses as the indicator of land use mix and found that its effects on urban crime varies from one city to another. Most recently, big data reflecting city dynamics have become widely available. Within this regard, POI data, which abstracts physical facilities as points, has been proposed to profile neighbourhood functions. To link POI data to urban crime, three indicators have been constructed from POI data, including POI density (i.e., POI count within a certain area), POI fraction (i.e., the proportion of POI count to the total POI count) and POI mix (i.e., the mixture of POI in a neighbourhood). Bendler et al. (2014) first integrated POI density in crime analysis for San Francisco, and found the densities of POIs like nightlife, food and drink to be strong predictors of theft crime. Redfern et al. (2020) calculated the POI kernel densities for crime prediction in ten UK cities, and reported that the additional POI predictors contributed much to the model accuracy. Wang et al. (2016) used both POI fraction and POI raw count to characterize the neighbourhood functions and confirmed the efficiency of POI indicators, especially the raw count, in crime inference problems in Chicago. These studies show the advantage of POI data in reflecting more accurate and diverse urban facilities.

Overall, most existing studies have used either land use or POI data to investigate the correlations between neighbourhood functions and urban crime. In terms of indicators, three types of indicators have been introduced in the literature: Density, Fraction and Mix. Note that Fraction and Mix can be constructed from both POI and land use data while only POI data has Density attributes. Nevertheless, there have not been systematic comparisons among different indicators from land use or POI data on their strength, divergence and coherence to profile neighbourhood functions. There exist disagreements of land use and POI data's impact on crime in previous research. For example, using land use data, Sohn (2016) argued that residential land use was negatively related to crime, while the research of Rumi et al. (2018) showed that there exist no significant correlations between residential POI and various crimes. The use of different types of indicators brings about further contradictions. For instance, Wang et al. (2016) reported that the crime rate was negatively correlated with the raw count of shop POI, while the correlation was statistically insignificant with its fraction. Comparing the performance of land use and POI data, with regard to the construction methods of indicators, can help avoid possible theoretical contradictions regarding the effect of neighbourhood functions on crime, and provide valuable guidance on the selection of indicators for future studies. To achieve the goal, this study provides a comparative study of different indicators based on land use and POI data in association with neighbourhood functions and urban crime in the neighbourhood level.

2. Study Area and Data

This study was conducted for theft crimes in Beijing, China. We focus on the area within the Fifth Ring Road of Beijing, where the majority of urban crimes occurred. The theft crime is chosen as a representative because of its high occurrence, great threats to the citizens' possession, and sensitivity to the urban contexts. The theft crime data used in this study was collected from the China Judgements Online. In this study, we use 4182 theft crimes from 2014 to 2016 that fall inside our study area for research.

The land use data was provided by Beijing Municipal Commission of Planning and Natural Resources. It consists of land parcels covering the study area which are labelled with a land use category. The POI data was crawled from Gaode Map, which provides the geographical coordinates, category and name of each POI. To facilitate comparative analysis with POI data, we reclassified land use data into 10 new categories and POI data into 14 new categories. Each POI category belongs to one land use type and one land use type corresponds to 1-3 POI categories.

Apart from land use and POI data, several supplementary data sources are used to construct control variables: 1) permanent residential population provided by the Sixth National Census of China; 2) road junctions collected from OpenStreetMap; 3) building footprints extracted from Baidu Map; 4) social media data crawled from Microblog; 5) night light intensity data extracted from remote sensing images.

3. Methodology

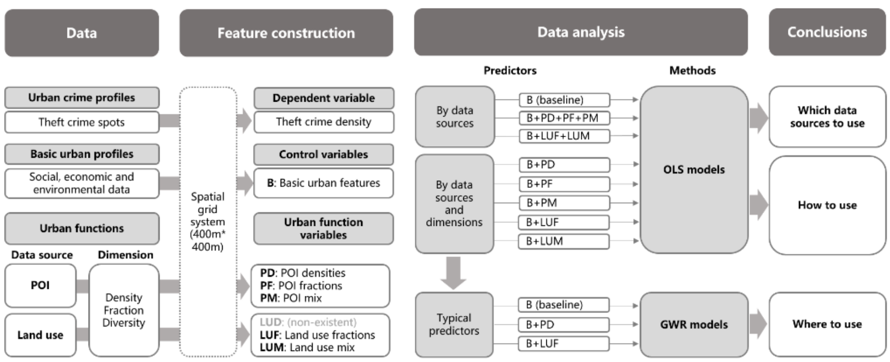


Figure 1. Flowchart for the whole study

In this section, we first introduce the variables and models used in our study, and then elaborate on the experiment framework we developed for the comparative analysis. A flowchart for the whole study is shown in Figure 1.

3.1. VARIABLES

Theft Crime Density: We divide the study area into regular-shaped grid cells and each crime lot is assigned to the grid cell that contains it. The crime intensity of each cell was then measured as the total number of crimes that occurred in the grid. Specifically, a raster grid with a resolution of 400m by 400m was used, which resulted in 4267 grid cells in total. The spatial distribution of theft crime intensity is shown in Figure 2.

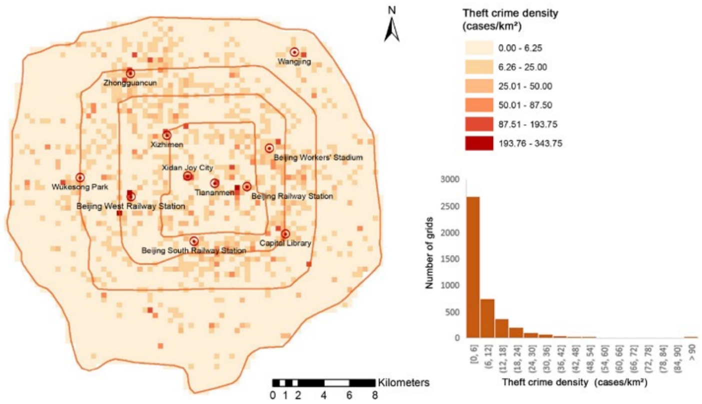


Figure 2. Spatial and statistical distribution of theft crime intensity

Neighbourhood Function Variables: Two types of indicators were constructed from land use data to profile neighbourhood functions: 1) Land Use Fraction (LUF): the proportion of land use area per category to the area of each grid; 2) Land Use Mix (LUM): a Herfindahl index of the ten specific land use categories for each grid following Wo (2019):

$$LUM = 1 - \sum_{j=1}^{N} LUF_j^2 \tag{1}$$

where LUF_j represents the area fraction of land use category j out of N categories. In terms of POI data, three types of indicators are constructed: 1) POI Density (PD): the normalized count of POIs per category in each grid; 2) POI Fraction (PF): the proportion of POI count per category to the total POI count in each grid (Chi et al. 2016); 3) POI Mix (PM): a Herfindahl index of the POI categories for each grid.

Control Variables: Apart from neighbourhood function variables, we also include five types of control variables, including population density, microblog count, night light intensity, road junction count and floor area ratio of each grid.

3.2. MODELS

OLS regression: OLS regression is one of the most widely used regression approaches in urban studies. In this study, the OLS regression takes theft crime density as the dependent variable, and the independent variables are our proposed indicators of neighbourhood functions using land use or POI data, along with other basic urban variables as control variables.

Geographically Weighted Regression (GWR): One possible limitation of OLS regression is their poor consideration of spatial non-stationarity. To tackle this issue, this study adopts a GWR model which allows the estimated parameters to vary over the spatial domain. It also takes spatial autocorrelation into account and thus can mitigate the bias caused by spatial autocorrelation.

3.3. EXPERIMENT FRAMEWORK

To compare the performance of POI and land use data as indicators of urban functions in urban crime studies, we develop a two-step experiment framework. First, we use the OLS regression model to compare the global prediction performance of different indicators constructed from POI or land use data. A baseline model is constructed using only control variables. Besides, we also build two models incorporating all the land use features and all the POI features respectively, and five more specific models with LUF, LUM, PD, PUF and PM as predictors respectively. Second, we devise a GWR model to uncover the local prediction performance for urban crime using POI or land use data. Specifically, GWR models with variables from POI and land use data are conducted, and the spatial variance of the model performance can be illustrated with local R^2. We further distinguish the applicability of land use and POI features in different neighbourhoods using correlation analysis and case studies.

4. Results

4.1. RESULTS FROM LR MODELS

While OLS Model 1 acts as a baseline, land use variables and POI variables are applied in OLS Model 2 and OLS Model 3 respectively. As shown in Table 1, OLS Model 3 has the best accuracy of all, with an adjusted R^2 of 0.234. Therefore, POI data generally provide more powerful predictors for theft crime prediction than land use data do.

Table 1. Performance of the OLS models

	OLS Model							
	M1	M2	M3	M2-2	M2-3	M3-1	M3-2	M3-3
R^2	0.120	0.139	0.240	0.139	0.121	0.236	0.139	0.122
Adjusted R^2	0.119	0.136	0.234	0.136	0.120	0.232	0.135	0.121
AIC	3.47e+4	3.46e+4	3.41e+4	3.46e+4	3.47e+4	3.41e+4	3.46e+4	3.47e+4

Predictors for the OLS models: M1: B; M2: B+LUF+LUM; M3: B+PD+PF+PM; M2-2: B+LUF; M2-3: B+LUM; M3-1: B+PD; M3-2: B+PF; M3-3: M+PM.

Regarding the p-values for urban functional variables, 9 out of 14 categories among POI density variables show significant impact on theft crime density, while only 4 of the 10 land use fraction categories are significant predictors, which reveals that POI densities might be generally more effective than land use fractions in theft crime prediction. POI mix proves to be significant as well.

Table 2. The coefficients of significant urban functional features with the top-3 positive or negative impacts in OLS M2 and OLS M3

	Feature	Coefficient (top-3 positive)	Feature	Coefficient (top-3 negative)
OLS M2	LUF(Medical)	15.086	LUF(Residential)	-2.080
	LUF(Commercial)	9.003		
	LUF(Financial)	6.166		
OLS M3	PD(Shop and catering)	84.795	PD(Government)	-17.885
	PD(Nightlife)	27.024	PD(Enterprise)	-13.518
	PD(Station)	20.996	PF(Station)	-4.200

In OLS M2, less than three of the significant urban function features ($p<0.1$) have negative coefficients.

Table 2 shows the coefficients of significant urban functional features with the top-3 positive or negative impacts in OLS Model 2 and OLS Model 3. Among all the significant predictors, the POI densities of governments and enterprises, the POI fractions of stations and the land use fraction of residence are negatively related to crime density, while the rest predictors have positive effects.

Some influencing patterns of urban functions found from OLS Model 3 are coherent to those from OLS Model 2. For instance, POI densities of typical commercial facilities are all significant predictors with positive coefficients, which is in accordance with the effects of commercial land in OLS Model 2.

OLS Model 3 also reveals more detailed impacts from other urban functions that are not shown in OLS Model 2. Enterprise and bank are two typical POI categories in financial land, but their densities have adverse impacts on theft crime. Moreover, for the same urban function, its density feature and fractional feature might have opposite impacts. For example, the POI density of stations has a positive coefficient, while the POI fraction of stations has a negative coefficient. As for diversity features, OLS Model 3 shows that POI mix has a significant promoting effect on theft crime, while land use mix in OLS Model 2 is insignificant. Compared with land use data, POI data might provide a higher-resolution depiction of urban functions for a land parcel and thus contribute to theft crime prediction.

To construct a well-performed and efficient model with the least independent variables necessary, systematic experiments are carried out to explore which dimensions of information from land use or POI data are most effective. According to their results shown in Table 1, POI densities have the strongest predictive power; POI fractions and land use fractions have similar and moderate predictive power; POI mix and land use mix are almost ineffective.

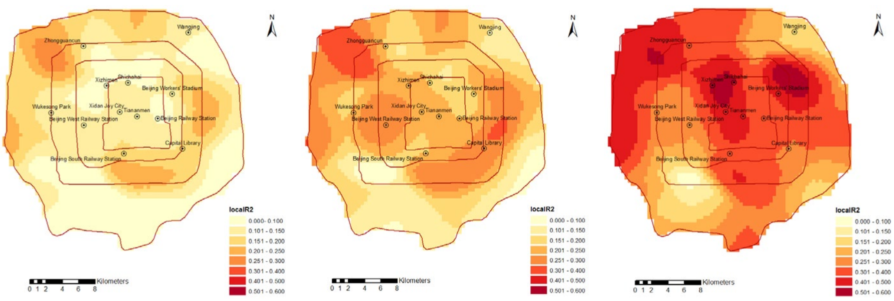
4.2. RESULTS FROM GWR MODELS

Based on the results of the OLS models, the GWR experiments only focus on the former three kinds of effective theft crime predictors: land use fractions, POI fractions and POI densities. We construct 4 GWR models, namely GWR Model 1 (B), GWR Model 2-2 (B+LUF), GWR Model 3-1 (B+PD), and GWR Model 3-2 (B+PF). The adjusted R^2 and AIC values testifies that in most cases, GWR model has stronger explanatory power over OLS model due to its consideration of spatial heterogeneity (Table 3).

Table 3. Results of geographically weighted model for thieve crimes

	GWR Model 1 (B)	GWR Model 2-2 (B + LUF)	GWR Model 3-1 (B + PD)	GWR Model 3-2 (B + PF)
R^2	0.178	0.267	0.385	0.284
Adjusted R^2	0.150	0.194	0.314	0.186
AIC	34618.96	34498.87	33835.37	34587.64

On the local scale, by contrast, POI densities do not always have an advantage. Figure 3 shows the spatial distributions of local R^2 of the GWR models, each of which has distinctive spatial patterns. (For POI data, only GWR Model 3-1 is retained as the better-performing model as compared with GWR Model 3-2). Compared with the baseline (GWR Model 1), GWR Model 2-2 obtains prominent increases in local R^2 in the nearby area of the western 3rd Ring Road, which implies the strong predictive power of land use fractions. Correspondingly, the result of GWR Model 3-1 indicates that POI density features have notable predictive power near central Beijing, especially in the northern areas of the 2nd Ring Road.



(a) GWR Model 1(B) (b) GWR Model 2-2(B+LUF) (c) GWR Model 3-1(B+PD)

Figure 3. Spatial distribution of local R^2 from the GWR models

Since the predictive power of land use fractions and that of POI density features show distinctive patterns, it is necessary to make a direct comparison. Therefore, a Power ratio indicator is calculated for each grid via the following equation:

$$Power\ ratio = \frac{local\ R^2\ in\ GWR\ Model\ 3-1}{local\ R^2\ in\ GWR\ Model\ 2-2} \tag{2}$$

The higher Power ratio is, the more advantage POI density features have over land use features. As shown in Figure 4, most grids have Power ratio larger than 1. The patterns of the power ratio agree with the intuitive judgement from local R^2 as discussed above. In this study, grids with power ratios larger than 2 are defined as prominent pro-

POI grids, and those with power ratios smaller than 1 are defined as prominent pro-land-use grids.

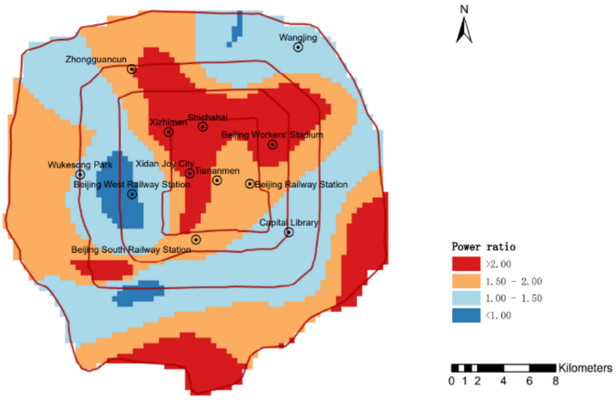


Figure 4. Spatial distribution of the power ratio

In total, 4 clusters of pro-POI grids and 2 clusters of pro-land-use grids are recognized. The northern area of the 2nd Ring Road has remarkably large power ratios, where POI density features have prominent advantages over land use fractions. By contrast, a cluster of grids near the western 3rd Ring Road have power ratios smaller than 1, indicating that GWR Model 2-2 obtains better local R^2 than GWR Model 3-1 does with less independent variables.

Table 4. Pearson Correlation between urban functional variables and the power ratio

Land use feature	Correlation	POI density feature	Correlation
LUF(Residential)	.015	PD(Residence)	.019
LUF(Administration)	.022	PD(Government)	.154**
		PD(Police station)	.083**
LUF(Culture)	.043**	PD(Culture)	.053**
LUF(Education)	.097**	PD(Education)	.095**
LUF(Medical)	.021	PD(Medical)	.089**
LUF(Commercial)	.064**	PD(Hotel)	.076**
		PD(Shop and catering)	.063**
		PD(Nightlife)	.128**
LUF(Financial)	.056**	PD(Enterprise)	.054**
		PD(Bank)	.056**
LUF(Municipal)	-.050**	PD(Public facility)	.151**
LUF(Green space)	-.137**	PD(Park)	.064**
LUF(Transportation)	-.094**	PD(Station)	.015

** P < .01. * P < .05.

To further explore factors that decide whether land use features or POI density features perform better locally, the Pearson correlation coefficients are calculated between urban functional variables and the power ratio (Table 4). A significant positive coefficient indicates that the more intense the urban function is, the more superior POI density features are to land use fractions in theft crime prediction, and vice versa.

For all the POI categories except residence and station, the density features are positively correlated to the power ratio at the $p<0.01$ level. The top five most influential POI categories are government, public facility, nightlife, education, and medical. The results indicate that in places with high POI prosperity, POI density features generally have stronger predictive power over land use features.

The Pearson correlations between land use features and the power ratio are more complicated. For cultural land, educational land, commercial land and financial land, their land use fractions are positively correlated to the power ratio at the $p<0.01$ level, where POI density features tend to have stronger predictive power. Meanwhile, the fractions of municipal land, green space and transportation land are negatively correlated to the effect ratio at the $p<0.01$ level. In these lands, POI density features tend to be inferior to land use fractions.

5. Conclusion

This study has three major findings:

(1) Among two data sources of urban functions, land use and POI, the latter one could provide generally more effective predictors in urban crime prediction.

(2) Among three dimensions of urban functional features, i.e., density, fraction and diversity, density features have the strongest correlations with urban crime occurrences, and these features could only be constructed from POI data.

(3) On most occasions, POI densities are optimal urban crime predictors in functional areas with small or medium-sized facilities, especially administration and public services, and commercial services. However, land use fractions provide better depictions for large-scale infrastructures like transportation hubs, green space and municipal services, as well as function zones with buildings of divergent scales like residential.

As alleviating crime becomes an important aspect of sustainable development in the social dimension, our findings provide several notable implications in the neighbourhood level in terms of urban crime prediction and planning policy making towards safer cities.

First, our research suggests that POI densities are effective predictors in theft crime prediction. While urban functional organizations have been proven to be related to crime occurrences, it might be difficult to acquire up-to-date and high-resolution land use data for many cities. Fortunately, POI data are now highly accessible, which could become a potential substitute to land use data.

Second, the superiority of POI densities over all fractional and diversity features indicates that the intensity of urban functions has notable impacts on theft crime. Previous planning policies are usually coarse organizations of different land use types, while our study suggests that the facility densities in a function zone should be also taken into account. For example, among commercial lands, places where nightlife spots gather tend to trigger more theft crimes, so limiting the density and business hours of nightlife spots might be of great help to reduce theft crime rate.

Third, our study also points out that POI densities and land use fractions are fit for different urban functions when studying their associations with theft crimes. To promote the security of cities through crime prediction or function zone adjustment, the best solution might be to focus on the POI densities of those functions undertaken by small or medium-sized facilities, and the land use fractions of large-scale infrastructures or function zones with heterogeneous building forms.

Last but not least, our findings regarding the power of POI data and land use data in urban modelling might be applied in other topics. The principles we discuss on how to choose between the two data sources and further construct proper features to depict different urban functions might be also useful for modelling other phenomena associated with urban functional organizations, such as urban vitality.

This study also has some limitations that need to be addressed in future research. Land use data and POI data were generated in different years (2014 and 2016). Although we use theft crime data from 2014 to 2016 accordingly, the inconsistency of data collection time might still influence the results. Another problem is that the predictive powers of POI data and land use data may vary at different spatial granularity (400m by 400 m in our case). Further research should make clear how to choose urban functional indicators when the spatial granularity changes. Besides, this study is conducted specifically for theft crime. More researches are needed to reveal if the superior predictive power of POI data remains true for other types of urban crimes.

References

Bendler, J., Ratku, A., & Neumann, D. (2014). Crime Mapping through Geo-Spatial Social Media Activity. *ICIS 2014* Proceedings. https://aisel.aisnet.org/icis2014/proceedings/ConferenceTheme/12

De Nadai, M., Xu, Y., Letouzé, E., González, M. C., & Lepri, B. (2020). Socio-economic, built environment, and mobility conditions associated with crime: A study of multiple cities. *Scientific Reports,* 10(1), 13871. https://doi.org/10.1038/s41598-020-70808-2

Redfern, J., Sidorov, K., Rosin, P. L., Corcoran, P., Moore, S. C., & Marshall, D. (2020). Association of violence with urban points of interest. *PLOS ONE*, 15(9), e0239840. https://doi.org/10.1371/journal.pone.0239840

Rumi, S. K., Deng, K., & Salim, F. D. (2018). Crime event prediction with dynamic features. *EPJ Data Science,* 7(1), 43. https://doi.org/10.1140/epjds/s13688-018-0171-7.

Sohn, D. W. (2016). Residential crimes and neighbourhood-built environment: Assessing the effectiveness of crime prevention through environmental design (CPTED). *Cities,* 52, 86-93. https://doi.org/10.1016/j.cities.2015.11.023.

Stucky, T. D., & Ottensmann, J. R. (2009). Land use and violent crime. *Criminology,* 47(4), 1223-1264. https://doi.org/10.1111/j.1745-9125.2009.00174.x

United Nations General Assembly (2015, October 21). *Transforming our world: the 2030 Agenda for Sustainable Development.* Retrieved December 2, 2021, from https://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1

Wo, J. C. (2019). Mixed land use and neighborhood crime. *Social science research,* 78, 170-186. https://doi.org/10.1016/j.ssresearch.2018.12.010

Wang, H., Kifer, D., Graif, C., & Li, Z. (2016, August). Crime rate inference with big data. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 635-644).