Impact of different types of environment on the crime rate of public space theft

1. Introduction

Theft of public space is one of the criminal activities that is most closely related to people's daily travel activities (Liu, Song and Xiu, 2014, pp.53-72). It is also one of the highest incidence cases(ibid.). It plays an important role in affecting urban public safety and other issues (Davies and Johnson, 2014, pp.481-507). As the road network becomes denser and the road environment becomes more diverse, the urban spatial structure is also becoming increasingly complex, which has brought about a more complex social environment and neighbourhood environment, and provided more criminal opportunities for criminals(ibid.). Therefore, it is of great significance to clarify the role of the neighbourhood environment of road in influencing crime, and to carry out effective prevention and control.

1.1 Literature review

The impact of road environment on crime rates has been confirmed by some studies (Hillier, 2004, pp.31-45). Different scholars focus on the different aspects of roads. For example, Sohn et al. (2016, pp.86-93) considered the relationship between road density and crime in Seattle and considered all types of roads as a whole. While, Maha (2014, pp.1-13) used the density of main roads and secondary roads as a factor influencing crime when discussing the spatial distribution of crime. There are also related studies focusing on the impact of road access and other attributes on the crime rate and its mechanism (Davies and Johnson, 2014, pp.481-507).

In addition, there is a large difference between the built environment and the social environment in different types of areas, which is the reason for the difference in the impact on the crime rate. Therefore, crime prevention and control against the built environment is important.

1.2 Research Question

Based on the above research, this article takes the environment of the roads as the research object, and establishes diversity by taking into account different types of road attributes. Using linear regression model to explore the impact of different types of neighbourhood environment on the crime rate of theft in public space. Therefore, this article will examine the following two issues:

i. How does the environment of different types of neighbourhoods affect the crime rate of theft in public space?

ii. What neighbourhoods factors are most important to crime in public space?

2. Data Presentation

2.1 Study area

The study area of this article is the city of London, with a total area of 2.9 Km2 and a population of 8072k in 2011.In this article, the theft of public space refers to the theft of other people's property in public space (excluding public transportation).

2.2 Data sources

This article uses the Lower Layer Super Output Area (LSOA, 2011) as the research unit, and summarizes basic geographic information such as alarm data, census data, traffic road network data, and Point of Interest data in Wards. The alarm data mainly comes from the public space theft data [data.police.uk] in the report received by the police from October 2018 to October 2019, and the population and road attribute data come from [data. london.gov.uk], POI data is downloaded from OpenStreetMap. Details can be found in Appendix 2.

There are many reasons for the accumulation of thefts in urban space. Therefore, when discussing the impact of different types of road environment on the crime rate of public space theft, other factors must be considered. Taking into account factors such as related research and data availability, the factors that have strong correlation with the crime rate of public space theft in terms of population, facilities, society, and economy are selected in this article. The dependent variable algorithm in this paper is as follows:

Public space theft crime rate = Theft Pieces / Population

2.3 Seasonal trend

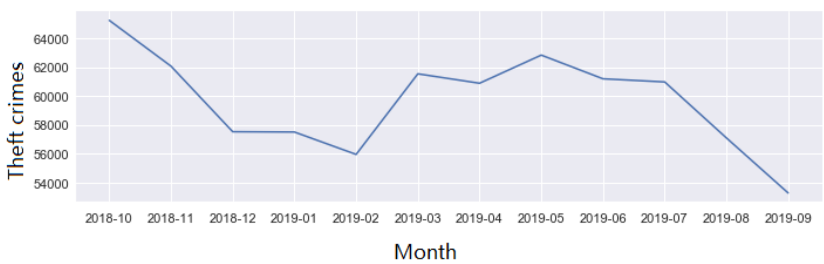

Figure 2. Theft Crimes in London 2018-2019

When investigating the thefts of public space, we can find obvious seasonal changes (seems variation with temperature or light time). The reason may be related to the behavior of tourists and residents. In summer, there are more foreign tourists, and the major attractions are high places of thieves; while citizens are also more active in public space, resulting in more people in public space and a higher probability of theft (Anon, 2018). On the other hand, there is a conjecture that it is related to the clothing of the residents. Perhaps the cool summer dresses make it easier for thieves to get started (ibid.)..

3. Methodology

3.1 Variable transformation

Firstly, the dependent variable does not meet the normality assumption of the linear model. Therefore, as shown in figure 1, taking the logarithm of the crime rate and the dependent variable after taking the logarithm are more similar to the normal distribution.

In addition, the extremely high value of the theft rate of public space in cities such as London will cause strong leverage to the model, which seriously affects the fitting effect of the model and the robustness of the model. Therefore, in this paper, the block method (Chen, Xu and Tian, 2017, pp.26) is used to replace the value above 99% of the model with the value of the 99% quantile, and replace the value below 1% with the value of the 1% quantile.

3.2 Multiple imputation

Multiple imputation is a method for processing missing values based on repeated simulations. It will generate a complete set of data, and the missing data will be filled by Monte Carlo methods. The interpolation of missing values is done by Gibbs sampling. Each variable containing missing values can be predicted by other variables in the data set by default, than these prediction equations could be used to predict the missing values (Kabacoff, R., 2011, pp.394).

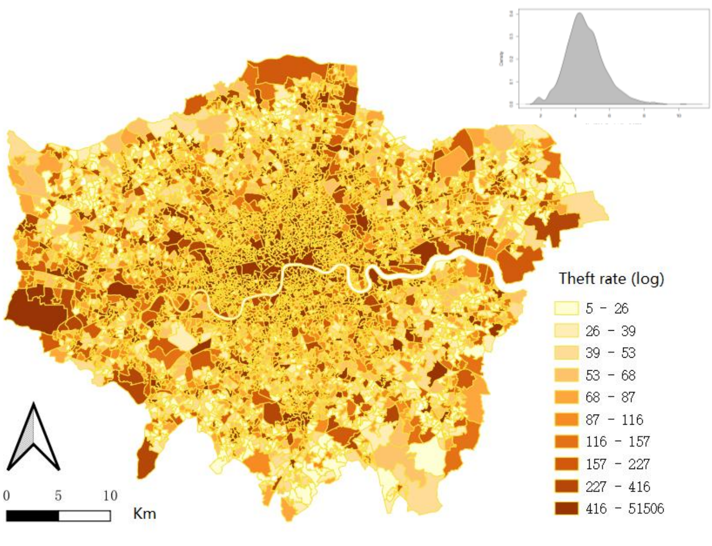

Figure 1. Theft Rate in London 2018-2019

3.3 Variables clustering

Clustering of variables is a method of arranging variables into homogeneous clusters based on their correlation (Chavent, M. et al., 2012, pp.1–16.). Then according to the divided clusters, select the variable that is the most representative for each cluster in this cluster. The algorithm in this article uses 1-R2, the formula is as follows:

$$1 - R^2\ \text{Ratio} = \frac{1 - R^2_{\text{own-cluster}}}{1 - R^2_{\text{next-closest}}}$$

3.4 Relative weight

Relative weight is a more common method for measuring the importance of variables this year. It is an approximation of the average increase in R-squared by adding a predictor to all possible sub-models (Kabacoff, R., 2011, pp.196).

4. Results

4.1 Result of variables clustering

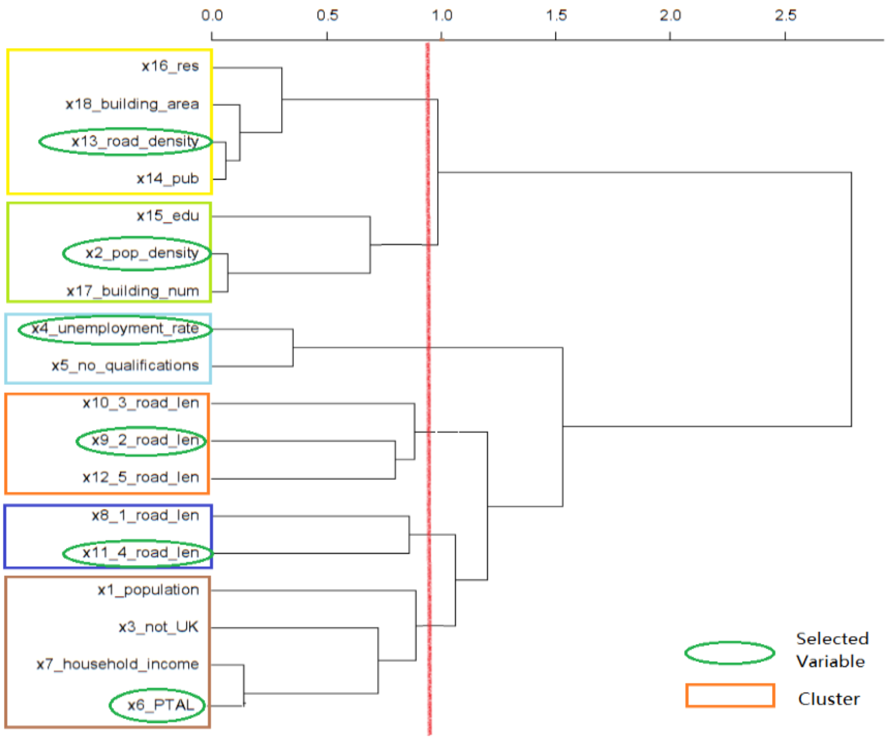

Figure 2. Cluster Dendrogram

From the results of variable clustering in the figure above, it can be seen that the 18 independent variables selected could be divided into 6 clusters based on their correlation coefficients. Among them, the optimal variables in each cluster are selected according to the information representation degree of the cluster data, and the remaining variables selected are road density (X13), population density (X2), unemployment rate (X4), and main road The proportion of the road (X9), the proportion of the road (X11), and the public transport reachability (X6). In addition, public transport accessibility is positively related to household income and the proportion of migrants, while the unemployment rate and education level are also positively related. After testing, the variance expansion coefficient (VIF) values between the six selected variables are all lower than 2, which means that the influence of multivariate collinearity is excluded by variable clustering.

4.2 Result of regression model

Table 1. Results of Regression Model

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.042e+00	4.361e-02	69.754	< 2e-16 ***
X4 Unemployment Rate	1.438e-02	3.652e-03	3.937	8.36e-05 ***
X6 PTAL	3.803e-01	8.163e-03	46.586	< 2e-16 ***
X11 Tertiary Road	-2.018e-01	5.098e-02	-3.958	7.66e-05 ***
X9 Primary road	1.591e-01	5.107e-02	3.116	0.00185 **
X13 Road density	1.445e-04	1.845e-05	7.834	5.80e-15 ***
Residual standard error:	0.8507 on 4711 degrees of freedom			
Multiple R-squared:	0.3432			
F-statistic:	492.4 on 5 and 4711			
p-value	< 2.2e-16			

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

The model results show that the goodness of fit of the model is 0.3432, which means 34% of the dependent variable can be explained by the independent variables. In addition, the independent variables are all significant, among which unemployment rate, public transport accessibility, trunk road density, road network density and public space are positively correlated; and highway density and crime rate are negatively correlated. The impact of population density on crime rate is not significant, so it is eliminated in the process of model construction.

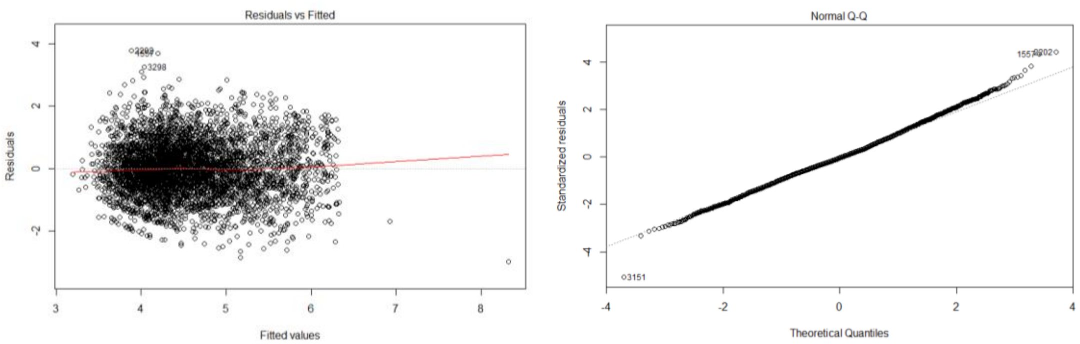

Figure 3. Residual Plot and Q-Q Plot

From the model's residual plot and Q-Q plot, we can see that the model's residuals shows a random distribution and satisfy the assumption of linear regression.

4.3 Relative importance of predictor variables

The results given by the relative weight algorithm can indicate the degree to which each predictor variable explains the model variance, that is, how important the variable is. It can be seen that the most important variable is public transport accessibility level (91.86%), which has the highest importance, followed by the density of branch roads (4.12%). The other three are not significantly different.

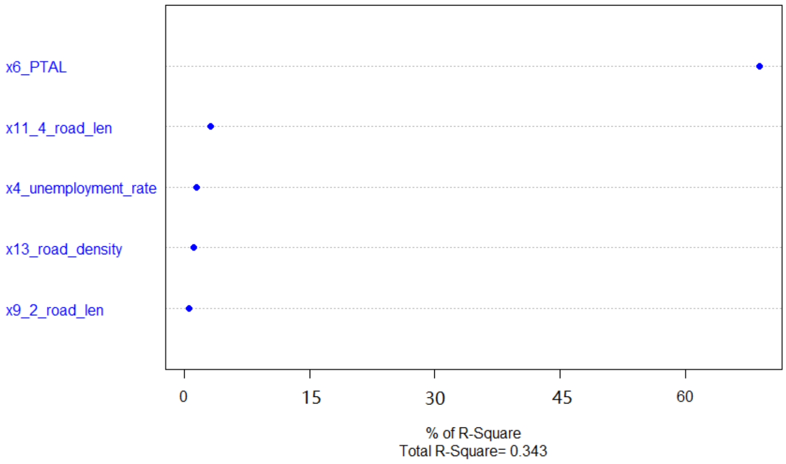

Figure 4. Relative Importance of Predictor Variables

5. Discussion

The good accessibility of public transportation usually means attracting the distribution of various facilities and the wide range of services provided by various facilities. It is easy to attract crowds of people and generate more victims and potential offenders. In addition, the population of developed public transportation areas Is highly mobile, anonymous, and has a poor recognition of potential offenders, which will result in a poor "watch effect" (Hakluyt and Burrage, 1959, pp.26-46), which means a lack of effective regulation; therefore victims, potential offenders, and lack of effective supervision gather in public space has created a criminal opportunity for theft.

Moreover, the accessibility of the city's secondary arterial roads is good, people are active, and the potential perpetrators have a wide range of spatial perceptions (Wortley and McFarlane, 2009, pp.149-156). Based on this, the crime rate of public space theft is high in densely populated areas of secondary arterial roads. Urban secondary arterial roads are regional arterial roads in the urban road network. Roads that undertake the function of traffic distribution between the main roads and the districts have both transportation and service functions.

Finally, an increase in the unemployment rate will lead to the concentration of criminals in the area, which will lead to more theft.

6. Conclusion

In short, by constructing a regression model, this article finds that different road environments have a significantly impact on theft rate. And the model results show that the built environment of the neighbourhood has a significant impact on theft crimes, especially factors such as public transport accessibility, road density, road grade, and unemployment rate.

However, there are many limitations in this study. Firstly, the R-squared value of the model is not well, mainly because the scales is not suitable, because theft of public space usually occurs on main streets, but the division of LSOA areas is also based on streets. So, theft of the same street may be divided into two different partitions. Secondly, the theft data has been obscured, so there is a certain deviation in the geographical location, and the quality of the original data is poor, about one-third of the records have no geographical location information.

Therefore, the future research research could focus on the street scale. In addition, the accessibility of public transportation seems to be an interesting direction. In the future, in-depth research on public transportation and theft can be conducted.

(Word count: 1798 words in 7 pages)

Appendix 1: Important Link

Github: https://github.com/yangznufe/QM-Project-2019----Theft-in-London

Appendix 2:

Table 2. Description of Independent Variables

Variable Name	Description	Min	Max	Mean
Social environment				
X1 Population	Number of Population in each LSOA	985	4933	1691
X2 Population density	Number of Population in each Km^2	1.8	82.30	36.11
X3 Foreign Rate	Proportion of the population of foreign	3.6	82.3	37.2
X4 Unemployment Rate	Proportion of the population of Unemployment	1.2	27.6	7.435
X5 Noqualification Rate	Proportion of the population of no qualications	1	43.3	17.84
X6 PTAL	Public transportation accessible level	0.3	8.0	3.749
X7 Household Income	Household income level	0.3	8.0	3.749
Road properties				
X8 Trunk road	Proportion of trunk road	0	1	0.06
X9 Primary road	Proportion of primary road	0	1	0.16
X10 Secondary road	Proportion of secondary road	0	1	0.06
X11 Tertiary Road	Proportion of tertiary road	0	1	0.15
X12 Residential road	Proportion of residential road	0	1	0.21
X13 Road density	Length of Road in each Km^2	0.0	70.5	0.10
Block environment				
X14 Pub density	Number of pub in each Km^2	0	6.32	0.01
X15 Education density	Number of education agents in each Km^2	0	1.53	0.003
X16 Food Poi density	Number of Food provider in each Km^2	0	23.27	0.04
X17 Buildings density	Number of buildings in each Km^2	0.002	709.2	4.81
X18 Built area rate	Proportion of built area	21.8	107.7	1591

Appendix 3: References

Anon, 2018. Nearly Three-Quarters of Americans Concerned About Identity Theft During Holiday Shopping Season. PR Newswire, pp.PR Newswire, Nov 28, 2018.

Chavent, M. et al., 2012. ClustOfVar: An R Package for the Clustering of Variables. Journal of Statistical Software, 50(13), pp.1–16.

Davies, T. and Johnson, S. (2014). Examining the Relationship Between Road Structure and Burglary Risk Via Quantitative Network Analysis. Journal of Quantitative Criminology, 31(3), pp.481-507.

Hakluyt, R. and Burrage, H. (1959). Early English and French voyages. New York: Barnes & Noble.

Hillier, B. (2004). Can streets be made safe?. URBAN DESIGN International, 9(1), pp.31-45.

Kabacoff, R., 2011. R in action : data analysis and graphics with R / Robert I. Kabacoff., Shelter Island, N.Y.: Manning.

Liu, D., Song, W. and Xiu, C. (2014). Spatial patterns of violent crimes and neighborhood characteristics in Changchun, China. Australian & New Zealand Journal of Criminology, 49(1), pp.53-72.

Maha J, Malaika, Hamid H. (2014). Achieving the efficient distribution of police stations and rescue police points in Duhok city/ Iraq by using (GIS), Journal of Information Engineering & Applications, 4(4), pp.1-13.

Szmadzinski, T. (1997). Probability, stastics & graphing. Grand Rapids, MI: Instructional Fair.

Wortley, R. and McFarlane, M. (2009). The role of territoriality in crime prevention: A field experiment. Security Journal, 24(2), pp.149-156.