Research Article GU J Sci, Part B, 9(2): 129-140 (2021)

Gazi University

Journal of Science

PART B: ART, HUMANITIES, DESIGN AND PLANNING

http://dergipark.gov.tr/gujsb

The use of spatial analytical techniques to highlight the geography of burglary in Leeds, UK.

Christopher Uche EZEH[1,*]

[1] 0000-0001-9866-0549, Department of Geography, Faculty of the Social Sciences, University of Nigeria

Article Info

Received: 03/04/2021
Accepted: 25/06/2021

Keywords

Burglary,
K-means,
Socio-demographic,
Geography,
Leeds

Abstract

Socio-demographic classification of attributes of burglary was carried out. The aim was to highlight the geography of burglary across Leeds using housing/socio-economic related data. Three hypotheses were formulated: (i) the denser the residential areas, the higher the likelihood of it being burglars' target. (ii) The denser the population and activities, the higher the burglary rate. (iii) The higher the territory's level of ethnic heterogeneity, the higher the probability of burglary incidence. The K-means technique was used to analyse the data. The hypotheses were upheld. It reveals that the bulk of burglary occur at central south-western part of Leeds which is the city centre with high business activities, population with high diversity and residences. The limitation of the technique is the inability to handle temporal dimension. To have best output, it should be used together with other techniques

1. INTRODUCTION

The geography of burglary in Leeds was studied via socio-demographic classification. It highlighted the nature and geography of burglary. Thereafter, followed the relevant analytical techniques, discussion of findings and conclusion. It did not necessarily seek the causes (why) and temporal (when) dimensions of burglary but the place (where) of burglary. Moreover, to achieve that, these hypotheses were posited: (i) the denser the residential areas, the higher the likelihood that it will be burglars' target. (ii) The denser the population and activities, the higher the burglary rate. (iii) The higher the level of ethnic heterogeneity, the higher the probability of burglary incidence.

Burglary implies all criminal entrances into buildings and vehicles aimed at carrying out criminal act whether successfully or not. It includes theft and other criminal intents like assault or sexual harassment [1]. Burglary event depends on the environment, opportunity, presence or absence of crime reduction programs, motivation amongst others ([2,3]. Similarly, influence of poverty and police activity has been highlighted [4]. This supports the view that deprivation or not of a people/location (dis)encourages incidence of burglary [5] depending on effectiveness of surveillance. The important role of opportunity, to understanding the incidence of burglary has been stressed [6] despite other views like drug and alcohol [7], wages [8]. The physical configuration of a place has been found to significantly influence burglary. For instance, the location of residents at higher altitudes reduces the risk of burglary victimisation [9,10].

The counter views notwithstanding, opportunity is foremost to understanding the geography of burglary [11]. These opportunities can be latent- the innate abilities of the offender over the victims such as smartness, trickiness. The overt opportunities include those offered the offender by the environment which increase as they become more vulnerable to the offender (eq. 1). However, these opportunities are perceived, since the potential burglar is not so certain of success till he ventures. Furthermore, locations that offer most of the opportunities will likely have the highest burglary incidences. This can be represented by this relationship;

* Corresponding author: christopher-uche.ezeh@unn.edu.ng

$$B = f(V) = (O) - (S\ or\ g) \quad 1.$$

where B is burglary, f is function, V is vulnerability, O is opportunity, S is surveillance or g is guard

The knowledge of geography of burglary (the where, why and when of crime event) is critical to understanding how crime can be controlled [12]. This reveals the reasons behind the disparity in distributions of crime among diverse locations as crime is unevenly distributed in space. It has been shown that city/urban centres usually experience higher incidences of crimes than the suburbs/rural [5,13]. There are higher burglary rates in most deprived areas than the least deprived areas [5,14]. Unusual cases of high burglary rates exist around University accommodation [11] including library, student union or hostels [15], ethnic heterogeneity and proximity to offenders' homes [16]. Higher burglary rates also take place on key pathways including major nodes like friend's house, recreation centres or work places where diverse potential offenders converge through routine activities with large numbers of potential victims and targets [16].

However, not all burglaries occur within the offenders' neighbourhood [17]. Several factors influence the distance-to-crime covered, like to avoid being recognised to mobility means and accessibility. Another dimension to this is the rational choice perspective [18] where the offenders tend to weigh the potential benefits of a successful hit with potential risks of being apprehended. However, it has been argued contrariwise that offenders tend to be motivated more by the benefits than the risks involved [19]. In this regard, the offender is likely to strike if the venture is perceived highly profitable to the neglect of the risks of been apprehended.

In fact, the geography of crime is complex especially with respect to its location [19]. It has been described via the concept of 'environmental backcloth' that is multifaceted including physical and social elements [19]. The next question is; which technique is needed to establish the spatial pattern of burglary?

2. METHOD

There are several classification techniques for grouping or regionalising events/areas. They include the hierarchical, PCA, K-means, K-medoid (for more, see [21]. The K-means involves classification of items based on similarity or common characteristics. This implies that areas in the same class may not necessarily be contiguous. K-means classification technique has numerous merits though not without demerits (Table 1).

Table 1. *Pros and Cons of K-means classification in geography*

Pros	**Cons**
It harmonises diversity through standardisation and grouping	Often involves daunting statistics
It reveals areal similarities and dissimilarities	No hard and fast rule to what constitute a group
Brings at a glance areal pattern	Often judgement is subjective and can be influenced by sentiment or prejudice
It is easy to interpret result	It is very sensitive to outliers
It allows subjects to move from one cluster to another	Ascertaining number of clusters is difficult, so multiple analysis may be done
It facilitates the assessment of properties of an area	Dominant members of the group mask the characteristics of the few other members.
It makes for easy comparisons	Very sensitive to the initial choice of cluster centres
It works well even when all the assumptions are not satisfied	It may not work well in a set with different densities

Source: [22-24].

Non-hierarchical methods like k-means entail total enumeration of all possible items, which for large datasets, may be unfeasible [21]. As a result, the technique is iterative and performs best for limited amount of data.

However, these weaknesses can be overcome with hybrid k-means. These include the fuzzy k-means [23]

[25], fast k-means [26], global k-means [27] and cluster ensembles [28]. These hybrid procedures are computationally faster [26]. It has been indicated that ensemble of clustering solutions is the best of the hybrid clustering techniques [28]. They argue that given large datasets, the clustering may be done on amenable size disjoint subsets that merges the partitions in a natural fit. Conversely, the weakness of these approaches is that given a small dataset, it is computationally 'time wasting' to use. Again, it is computationally complex and costly, involving advanced skill to use.

Therefore, the justification of simple k-means is its added advantage of ease of application, simplicity, efficiency and empirical successes over the years (Table 1) [29]. He reiterates, there is no best clustering algorithm as each imposes a structure on the data and a good partition is obtained where a good match exists between the model and the data. Hence, "clustering is in the eye of the beholder" ([29], p663) and none outperforms others across all applications.

2.1. Data

The variables were selected from 2001/2000-03 burglary/socio-economic data respectively (Table 2). No hard and fast method exist for choosing the most suitable number of clusters or variables [22,23].

Table 2. *Data and selected variables*

Data type	source	Variable Domain	Relation	Selected variables in percentage
Tabulated Sav	ONS	Housing	Housing	Terraced, detached, flats, Pub. Rented, owner
		H.hold composition	Housing	Households
		Demographic	Economic	Pop. Density, single
		Ethnicity	Social	sthasian, white
		Employment	Economic	Unemployed
LSOA shp file	ONS			

Variable selection based on PCA often offers no assurance that the component loadings has pointers to what one wants to detect via clustering [30] (Table3). Moreover, the use of correlation is fraught with weaknesses too. First, is the case of non-linearity and outliers [31], collinearity [32], overestimation of the relationship between large variables [33].

To overcome these weaknesses, several techniques have been proposed such as the sparse clustering framework [30], Feature Annealed Independence Rule [34], extended LARS technique [35], Cluster indicator matrix [36]. However, the limitation to their usage is their computational complex nature. Moreover, none has been a provably acceptable feature selection for k-means [36].

Therefore, a simple approach was adopted, employing the advantages of correlation and standard deviation. Correlation is used to obtain validity and reliability of evidence [31] and increases with higher variability. Therefore, a correlation was used to eliminate redundancy among data while the standard deviation was used to include all variables that would likely impact the event understudy by considering their variation across the LSOAs [22] (Tables 4 and 5). Hence, of the 25 variables, 11 were selected (Table 2), though the emphasis was on housing-related. Highly correlated variables had one discarded. For instance, full-time students had a strong positive correlation with aged 15 – 24 and single, so single was selected based on the extent of variation (Table 4, 5). Owner had higher variation than married; married was discarded (Table 5). Others with low standard deviations like unemployed was selected at the expense of routine occupation since they fall within the same class of deprived but unemployed had higher PCA (Table 3). The high professional was dropped on the basis that they are the least target to burglars due to their position in the society and affluent residential quarters [11].

***Table 3**. PCA loadings on the variables*

Component Matrix[a]

	Component					
	1	2	3	4	5	6
Persons per hectare	.65	-.39	-.31	-.04	-.04	-.27
% Age 45-64	-.77	.38	.19	-.16	.10	-.13
% Married	-.89	.29	-.08	-.25	.09	-.06
% Single	.70	-.66	.08	.02	-.18	.04
% Employed Full Time	-.69	.18	-.38	.50	-.12	.03
% Unemployed	.78	.38	-.05	.03	.27	.19
% Lower professionals	-.85	-.34	-.08	.26	.16	.12
% Detached	-.76	.16	.09	-.44	-.003	.32
% Public rented	.65	.60	.25	.04	-.05	.26
% Owner occupiers	-.94	-.07	-.25	-.07	.05	-.12
% Lone parents	.65	.55	-.14	-.08	.14	.34

Extraction Method: Principal Component Analysis.

a. 6 components extracted.

***Table 4.** Highly correlated variables positive/negative*

Variable 1	**Variable 2**	**Corr. Co.**	**Rd**	**Variable 1**	**Variable 2**	**Corr. Co.**	**Rd**
15 - 24	single	0.93	0.87	sthasian	white	-0.91	0.83
45 - 64	married	0.85	0.72	45 64	single.	-0.79	0.62
15 - 24	student	0.95	0.90	single	married	-0.86	0.74
Student	single	0.87	0.76				
Single	detached	0.80	0.64				
Married	Owner	0.87	0.76				

***Rd** = Redundancy, ***Corr. Co** = Correlation coefficient

Table 5. *Standard deviation*

	Std. Deviation
Count of households in each LSOA	92.52
Persons per hectare	36.06
% Age 15-24	11.08
% Age 25-44	5.20
% Age 45-64	5.67
% Age 65+	5.87
% Married	14.27
% Single	13.13
% Indian, Pakistani & Bangladeshi ethnicity	7.45
% White ethnicity	11.55
% Employed Full Time	8.56
% Unemployed	2.14
% Full-time students	3.28
% No qualifications	13.26
% Qualification level 4 or 5 (degree, professional)	11.54
% Higher professionals	3.69
% Lower managerial and professionals	6.42
% Routine occupations	4.93
% Detached / semi-detached housing	26.90
% Terraced house	22.00
% Flats	12.98
% Public rented	19.84
% Private rented	6.24
% Owner-occupiers	24.86
% Lone parents	5.46

2.2. Analysis

To remove bias in the dataset, they were standardised. Diverse approaches abound for doing this, some of which were outlined by Vickers et al. (2003) but this study employed the standard normal variate (eq. 2).

$$Zi = \frac{\sqrt{\text{xi}-\bar{\text{x}}}}{6\text{x}} \qquad 2$$

where Zi is the Z-score, Xi is the individual variable, $\bar{x}$ is the mean of x, бx is the standard deviation of x.

The procedure in K-means classification involves partitioning n data points with m variables into k clusters which results in a matrix of cluster centres m (k, j) [22]. This minimises the Euclidean sum of squares as shown in equation 3 [22] to obtain spherical or ball-shaped clusters in the data.

$$J(k,m) = \sum_{i=1}^{n} \quad \sum_{l=1}^{m}(Zij - Zcj \quad)\char`^2 \qquad 3$$

where Zcj is the value of the cluster c and variable j, Zij is the value of object i and variable j.

The selected variables were then analysed via the SPSS. Iteration of 100 was used. Several runs were done starting from 9 down to 5 of which 5 was acceptable (Figure 1).

3. RESULT AND DISCUSSION

The resulting groups of the cluster are displayed in Figures 1 and 2.

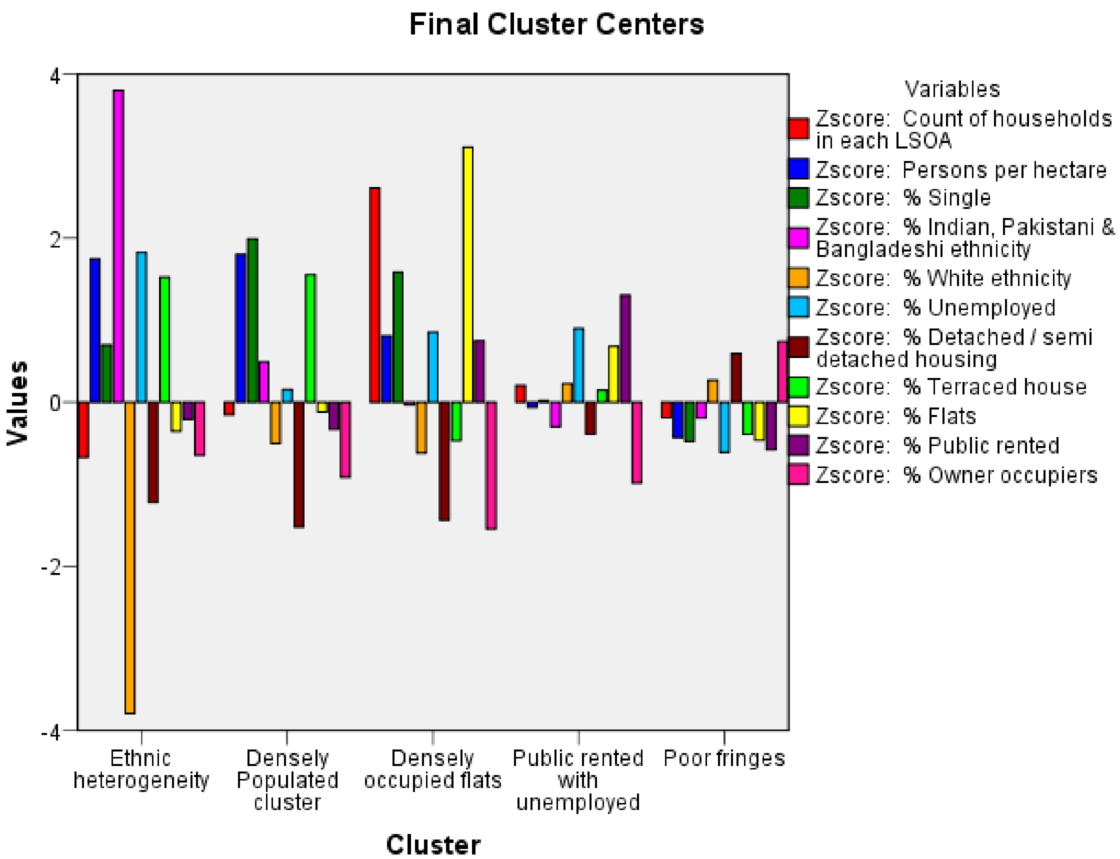

Figure 1. *Number of clusters*

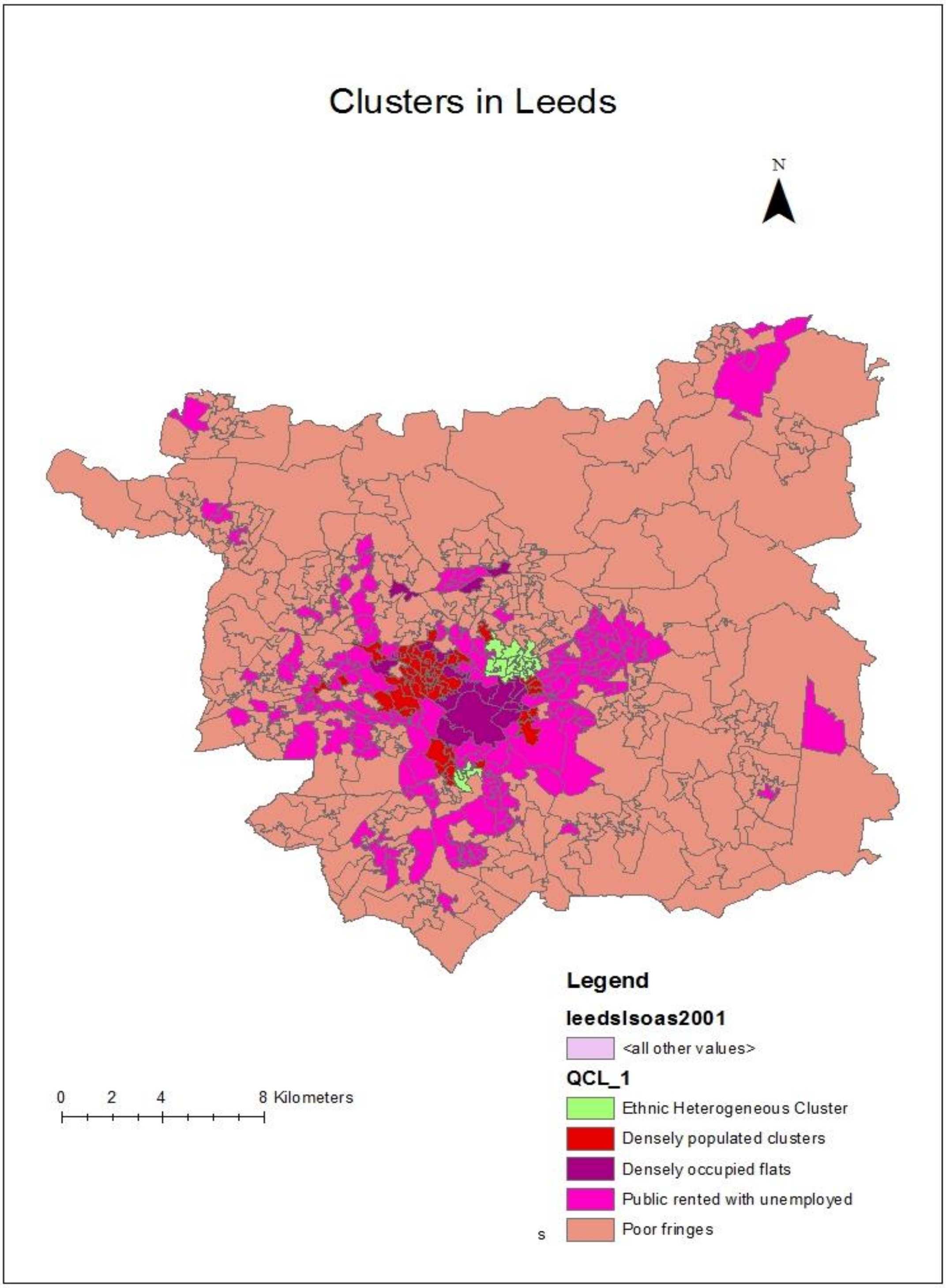

***Figure 2**. Socio-demographic classification of Leeds*

The findings reveal the socio-economic variables that have the greatest synergies with the burglary at certain locations (Figures1 and 2). It shows that the high clustering of burglary cases in about 43% of the LSOAs was due to ethnic heterogeneity, dense population, densely occupied flats and unemployment (clusters 1-4). It reveals also that higher crime intensity concentrates around the city periphery (26.3%, clusters 4) which is public rented flats with unemployed. It informs that the central south-western part (city centre) demands more surveillance from the police especially the periphery where unemployment is higher.

The results show that burglary rate is higher in the city centre (Figure 2). This is because the city centre is where there is higher population, ethnic heterogeneity and other opportunities. Hence, it supports the theory that burglary is higher in the city than in the rural districts [14] (Fig 2). It shows that burglary decreases as opportunity decreases [6,11,2]. From Figure 2, burglary rates reduce centrifugally from the periphery. This shows that accessibility could be a factor considered by burglars for which its absence, limits opportunity [5,13]. The findings strongly support the formulated hypotheses that:

- The denser the residential areas, the higher the likelihood that it will be a target to burglars (Cluster 3)
- The denser the population and activities, the higher the burglary rate will be (Cluster 2)
- The higher the territory's level of ethnic heterogeneity, the higher the probability of burglary incidence (cluster 1). Therefore, burglary is strongly tied with where opportunities exist [3], and with an absence of surveillance, the burglars will burgle (eq.1). These perceived opportunities are more in the central south-western part of the study area (Figure 2).

The findings in figure 2 were validated in figure 3. It shows that the burglary rate is highly concentrated around the central south-western part of Leeds. This entails that the simple k-means is a good technique for the spatial study of burglary of a given geographical location.

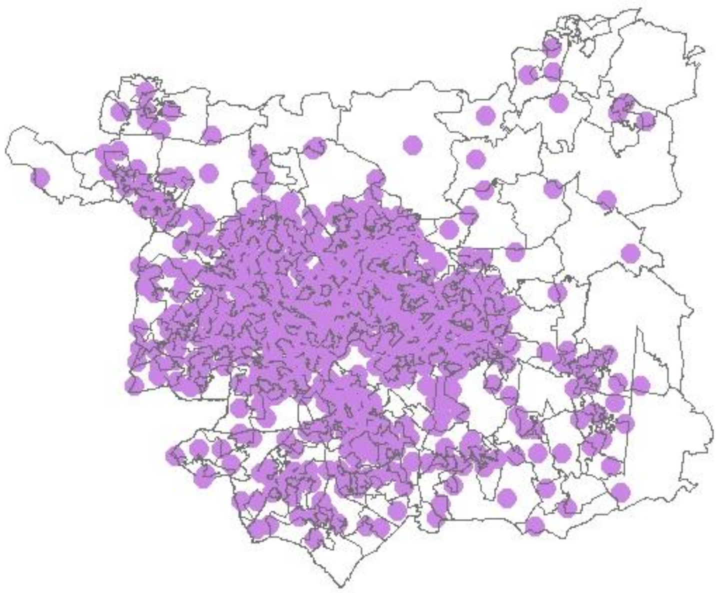

Figure 3. *Burglary rate*

4. LIMITATIONS TO THE STUDY

Confidentiality requirements pose problems, as multiple crimes are assigned to an areal location [13], so certain details would be concealed [11]. The visual display gives the erroneous impression that burglary rates are evenly distributed among households across a given output area (figures 1 and 2). Time dimension and offenders' locations were not part of the data, so the effect of season or day/night relations with burglary could not be considered as well as burglars' travel distance to the target.

5. CONCLUSION

The posited hypotheses were upheld implying that burglary depends on opportunities. The k-means produces good result visually. However, to overcome the outlined weaknesses, it should be used together with other techniques like kernel density, Ward's hierarchical and others that can handle temporal dimensions. Tools like Agent-based modelling could be employed to assess burglars' travel time [2] and given its predictive power can enhance the effectiveness of police in combating burglary. Finally, updated data having the noted missing attributes is necessary. This will improve research in this field like burglary trend [13], temporal dimension [14].

REFERENCES

[1] Nee, C. 2010. Residential Burglary: methodological and theoretical underpinnings in Brown and Campbell (eds), Cambridge Handbook of forensic Psychology. Cambridge, UK. Cambridge University Press. Retrieved 20th December, 2016 from http://eprints.port.ac.uk/11264/1/filetodownload,102435,en.pdf

[2] Malleson, N. and Evans, A. 2013. Agent-based Models to predict Crime at Places. Retrieved 28th November, 2016 from http://nickmalleson.co.uk/wp-content/uploads/2012/01/2013-crim_enc-abm.pdf

[3] Malleson, N., Heppenstall, A., See, L. and Evans, A., 2013. Using an agent-based crime simulation to predict the effects of urban regeneration on individual household burglary risk. *Environment and Planning B: Planning and Design*, *40*(3), pp.405-426.

[4] Kelly, M. 2000. Inequality and Crime. *Review of Economics and Statistics. Vol.82, No.4, pp530 – 539.*

[5] Higgins, N, Robb, P. and Britton, A. 2010. Geographic Patterns of Crime, in Flatley et al.(eds) Crime in England and Wales 2009/10, London, Home Office. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/116360/hosb1210-chap7.pdf

[6] Felson, M. and Clarke, R.V. 1998. Opportunity makes the thief: Practical theory for crime prevention. Police Research Series, Paper 98. Home office, Policing and reducing Crimes Unit. Research, Development and Statistics Directorate London (Webb, B. ed.) London.

[7] Kristie, R. 2012. Understanding Decisions to Burglarise from the offender's Perspective. UNC Charlotte. Retrieved 28th December from https://scholar.google.co.uk/scholar?cluster=7718305242168458044&hl=en&as_sdt=0,5 and http://s3.amazonaws.com/academia.edu.documents/44000995/Understanding_Decisions_to_Burglarize_fr20160322-21398-1jcdk0u.pdf?AWSAccessKeyId=AKIAJ56TQJRTWSMTNPEA&Expires=1482983482&Signature=igAcyh%2BigJovLeoeOHrxFUiWpLs%3D&response-content-disposition=inline%3B%20filename%3DUnderstanding_Decisions_to_Burglarize_fr.pdf

[8] Gould, E.D., Weinberg, B.A. and Mustard, D.B. 2002. Crime Rates and Local Labor Market Opportunities in the United States: 1979 – 1997. *Review of Economics and Statistics. Vol.84, No.1, Vol.14, No.21, pp45 – 61.*

[9] Felson M, 2002 Crime and Everyday Life 3rd edition (Sage, Thousand Oaks, CA)

[10] Breetzke, G.D., 2012. The effect of altitude and slope on the spatial patterning of burglary. *Applied Geography*, *34*, pp.66-75.

[11] Malczewski, J and Poetz, A. 2005. Residential Burglaries and Neighbourhood Socio-economic Context in London, Ontario: Global and Local Regression Analysis. *The Professional Geographer*, *Vol.57, No.4, pp516 – 529.*

[12] Cahill, M.E. 2005. Geographies of Urban Crime: An Intraurban Study of Crime in Nashville, T.N; Portland, OR; and Tucson, AZ. A Dissertation submitted the Faculty of the Department of Geography and Regional Development , University of Arizona. Retrieved 1st January, 2017 from https://www.ncjrs.gov/pdffiles1/nij/grants/209263.pdf

[13] Gale, C.G., Singleton, A.D. and Longley, P.A. 2015. Profiling Burglary in London using Geodemographics. Retrieved 1st January, 2017 from http://leeds.gisruk.org/abstracts/GISRUK2015_submission_123.pdf

[14] Thornley, A.M. 2004. Crime in Space and Time: A Spatial and Temporal Analysis of Burglary in Christchurch. A thesis submitted to the Department of Geography, University of Canterbury.

[15] Brantingham, P.L., Brantingham, P.J. and Seagrave, J. 1995. Crime and fear at a Canadian University, in Fisher and Sloan (eds) Campus Crime: Legal, Social and Policy Perspectives, Springerfield, pp123 – 155.

[16] Bernasco, W. and Nieuwbeerta, P. 2005. How do Residential Burglars select Target Areas? *Brit. J. Criminol*., *Vol.44, pp296 – 315*.

[17] Blevins, K.R., Kuhns, J.R. and Lee, S.Z. 2012. Understanding Decisions to Burglarize from the Offender's Perspective. UNC Charlotte. The University of North Carolina at Charlotte, Department of Criminal Justice and Criminology.

[18] Clarke R V, Cornish D B, 1985, "Modeling offenders' decisions: a framework for research and policy" Crime and Justice 6 147–185

[19] Vito G F, Maahs J A, Holmes R M, 2007 Criminology: Theory, Research and Policy (Jones and Bartlett, Sudbury, MA).

[20] Brantingham P L, Brantingham P J, 1998, "Mapping crime for analytic purposes: location quotients, counts, and rates", in Crime Mapping and Crime Prevention Eds D Weisburd, T McEwen (Criminal Justice Press, Monsey, NY) pp 263–288

[21] Izenman, A.J. 2008. Modern Multivariate Statistical Techniques: Regression, Classification and Manifold Learning. Springer.

[22] Vickers, D., Rees, P. and Birkin, M. 2003. A new classification of UK Local Authorities using 2001 Census Key Statistics. Working Paper 03/03. Retrieved December 1st, 2016 from https://vlebb.leeds.ac.uk/bbcswebdav/pid-4001155-dt-content-rid-7293652_2/courses/201617_26163_GEOG5520M/WP-03-3.pdf

[23] Cornish, R. 2007. Statistics: Cluster Analysis. Mathematics Learning Support Centre. Retrieved 5th January, 2017 from http://www.statstutor.ac.uk/resources/uploaded/clusteranalysis.pdf

[24] Iyigun, C., Turkes, M., Batmaz, I., Yozgatligil, C., Purutcuoglu, V., Koc, E.K. and Ozturk, M.Z. 2013. Clustering Current Climate regioins of Turkey by using a multivariate statistical method. *Theoret. And Applied Climatol., Vol.114, Issue 1, pp95 – 106*.

[25] Borough, P., Wilson, J.P., van Gaans, P.F.M. and Hansen, A.J. 2001. Fuzzy K-means Classification of Topo-climatic data as an aid to forest mapping in the Greater Yellowstone Area USA. *Landscape Ecology Vol.16, pp523 – 546*.

[26] Lee, S.S. and Lin, J.C. 2013. Fast k-means clustering using Deletion by Center Displacement and Norms Product (CDNP). *Pattern Recognition and Image Analysis. Vol.23, Number 2, pp199 – 206.*

[27] Likas, A., Vlassis, N. and Verbeek, J.J. 2003. The Global K-means Clustering Algorithm. *The Journal of the Pattern Recognition Society, Vol.36, pp451 – 461*.

[28] Hore, P., Hall, L.O. and Goldgof, D.B. 2009. A Scalable framework for cluster Essembles. *Pattern Recognition, Vol.42, pp676 – 688*.

[29] Jain, A.K. 2010. Data Clustering: 50 years beyond K-means. *Pattern Recognition Letters, Vol.31, pp651 – 666.*

[30] Witten, D.M and Tibshirani, R. 2010. A Framework for Feature Selection in Clustering. Journal of American Statistical Association, Vol.105, No.490, pp713 – 726.

[31] Goodwin, L.D. and Leech, N.L. 2006. Understanding Correlation: Factor that Affect the Size of r. The Journal of Experimental Educatiion, Vol.73, No.3, pp251 266.

[32] Fan, J. and Lv, J. 2010. A Selective Overview of Variable Selection in High Dimensional Feature Space. Statistica Sinica, Vol.20, pp101 – 148.

[33] Hole, G. 2015. Eight things you need to know about interpreting correlation. Research Skills One, Correlation Interpretation. Retrieved 10th January, 2017 from http://users.sussex.ac.uk/~grahamh/RM1web/Eight%20things%20you%20need%20to%20know%20about%20interpreting%20correlations.pdf

[34] Fan, J. and Fan, Y. 2008. High-Dimensional Classification Using Features Annealed Independence Rules. Annals of Statistics, Vol.36, No6, pp2605 – 2637.

[35] Yuan, M, Joseph, V.R. and Lin, Y. 2007. An Efficient Variable Selection Approach for Analysing Design Experiments. American Statistical Association and the American Society for Quality Technometrics Vol.49, No.4, pp430 – 439.

[36] Boutsidis, C, Mahoney, M. and Drineas, P. 2009. Unsupervised Feature Selectiion for the K-means Clustering Problem. NIPS proceedings , retrieved 10th January, 2017 from http://www.stat.berkeley.edu/~mmahoney/pubs/NIPS09.pdf