Complementing GIS with Cluster Analysis in Assessing Property Crime in Katsina State, Nigeria

S. A. Yelwa
Department of Geography
Usmanu Danfodio University
PMB 2346, Sokoto, Nigeria.

Y. Bello
Department of Mathematics and Statistics, Hassan Usman Katsina
Polytechnic, Katsina

Abstract

One of the fundamental techniques to combat criminal activities is the better understanding of the dynamics of crime. Techniques are needed to categorise geographical areas according to their similarities in criminal activities that would facilitate the understanding of why and where crimes take place. The detection of crime hot-spot areas would enhance the policing agencies and policy makers to develop appropriate techniques for controlling and prevention of crime. This paper aimed at complementing Geographical Information System (GIS) with multivariate cluster analysis in assessing property crime in the LGAs of Katsina State, Nigeria.

Key words: Property, crime, GIS, Cluster analysis, Criminal Activities

1. Introduction

The relevance of GIS mapping in crime control and prevention is very enormous. It can be used by police personnel to plan effectively for emergency response, determine mitigation priorities, analyse historical events, and predict future events. It can be used as an investigative methodology that uses the locations of a connected series of crimes to determine the most probable area of the residence of the offender (Fajemirokun *et al.*, 2006). The application of GIS to crime mapping and management is being used in developed countries and has been proved to be successful. Nowadays, in Europe and USA, many police departments are mapping crime patterns on a daily basis. The San Bernardino police department in southern California for example, is using GIS technology on an ongoing basis to plan patrols (Radoff, 1993).

In their effort to analyse the crime in urban region, McGuffog *et al,.* (2001) have explored the use of GIS for examining crime occurrence in Brisbane, Australia. Savoie (2008) has also used GIS to determine the spatial distribution of violent crime and robbery incidences in Regina and Winnipeg, Canada respectively. In predicting the geo-temporal variations of crime and disorder and in identifying clusters with relatively high levels of crime (hot spots), the use of GIS is proved to be relevant (Corcoran *et al.*, 2003). In Nigeria the application of GIS started to evolve when Fajemirokun *et al.,* (2006) utilised it in mapping crime in Victoria Island, Lagos, Nigeria.

Most these works cited utilised GIS solely for mapping crime analysis. However, in this work, GIS mapping is complemented with multivariate cluster analysis.

The work of (Rencher, 2002) for example, provides a good relevance of using cluster analysis in crime analysis. He compared the crime rates per 100,000 populations of 1970 for the sixteen US cities and grouped them according to their homogeneous clusters. The result of his work was displayed using the dendrograms of different algorithms for hierarchical method. Hardle and Zdenek (2007) have also analysed the US reported crimes data for 1985 using cluster analysis. Olivares (1996) have applied six different clustering algorithms on the Euclidean distances to classify prison inmates and compared the results with the actual qualitative classification. The result of the Ward method, among others, has satisfactorily matched the actual classification.

The purpose of this paper is to complement the use of GIS with multivariate cluster analysis in assessing property crime in the LGAs of Katsina State.

2. Data and Methods

The crime data for 36 Divisional Police Headquarters of the Local Government Areas (LGAs) of Katsina State was collected from the Statistics (F) department of the Nigeria Police Force, Katsina State command. For easy identification the 36 LGAs were categorized according to the three existing Area Commands (ACs) in the state: Katsina, Funtua and Daura ACs. Each LGA within the three ACs were identified by 1, 2, and 3 respectively.

The data consists of five property crimes reported to the police from the year 2006 to 2008. The crimes are robbery, theft and stealing, unlawful possession, vehicle theft and house and store breakings. The crime Frequencies of crime for each category were averaged over the three years in the study period to control for anomalous years when there may have been an unexplained spike or fall in crime levels prior to the statistical analysis. The value for each crime is converted to crime rate per 100,000 populations of the LGA (Kpedekpo and Arya, 1981), and is calculated as :

$$Crime\ rate = \frac{number\ of\ crime\ committed}{population\ of\ the\ LGA} \times 100{,}000 \qquad (1.1)$$

2.1 Geographic Information Systems (GIS)

ArcView GIS 3.2a, software was utilised for spatial analysis so as to map out the category of crimes enumerated earlier. GIS images of different average property crime rates in Katsina State from 2006 - 2008 are displayed in section three.

2.2 Cluster Analysis

Given a data matrix containing multivariate measurements on a large number of individuals (or objects), the objective is to build subgroups for which the observations or objects within each cluster are similar, but the clusters are dissimilar to each other according to some appropriate criterion.

Cluster analysis can be divided into two fundamental steps:

1. Choice of a proximity measure:
   One checks each pair of observations (objects) for the similarity of their values. A similarity (proximity) measure is defined to measure the *closeness* of the objects. The "closer" they are, the more homogenous they are.
2. Choice of group-building algorithm:

On the basis of the proximity measures the objects assigned to groups so that differences between groups become large and observations in a group become as close as possible (Hardle and Simar, 2003).

The stating point of a cluster analysis is a data matrix **χ**(*n*p*) with *n* measurements (objects) of *p* variables (Hardle and Simar, 2003). The proximity (similarity) among objects is described by a matrix *D* (*n*n*)

$$D = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & \cdots & d_{nn} \end{bmatrix} \qquad (1.2)$$

The matrix *D* contains measures of similarity among the *n* objects. If the values d_{ij} are distances, they measure similarity. The greater the distance, the less similar are the objects. The Euclidean distance d_{ij} between two cases, *i* and *j* with variable value $x_i = (x_{i1}, x_{i2}, ..., x_{ip})$ and $x_j = (x_{j1}, x_{j2}, ..., x_{jp})$ is define by

$$
\begin{aligned}
d_{ij} &= \sqrt{\left(x_i - x_j\right)'\left(x_i - x_j\right)} \\
&= \sqrt{\left(x_{i1} - x_{j1}\right)^2 + \left(x_{i2} - x_{j2}\right)^2 + - - - - + \left(x_{ip} - x_{jp}\right)^2} \\
&= \sqrt{\sum_{k=1}^{p} \left(x_{ik} - x_{jk} \right)^2}
\end{aligned}
$$

Clusters formed in this method are compared using the averages of the standardized data set within the clusters obtained to determine the clusters of higher and lower crime concentrations (Hardle and Zdenek, 2007).

There are two types of clustering algorithm: hierarchical and non-hierarchical (partitioning) algorithm. The hierarchical algorithm can be divided into agglomerative and divisive procedures. The agglomerative hierarchical methods start with the individual objects. Thus, there are initially as many clusters as objects. The most similar objects are first grouped, and these initial groups are merged according to their similarities. Eventually, as the similarity decreases, all subgroups are fused into a single cluster (Richard and Dean, 2001).

The Ward algorithm does not put together groups with smallest distance. Instead, it joins groups that do not increase a given measure of heterogeneity "too much". The aim of the Ward procedure is to unify groups such that the variation inside these groups does not increase too drastically: the resulting groups are as homogenous as possible.

The heterogeneity of group *R* is measured by the inertia inside the group. This inertia is defined as:

$$I_R = \sum_{i=1}^{n_R} d^2\left(x_i, \overline{x_R}\right), \qquad (1.3)$$

where $\overline{x_R}$ is the arithmetic average and the n_R the number of observation within group *R* (Ward, 1963). If the usual Euclidean distance is used, then I_R represents the sum of variances of the component of x_i inside group *R*.

3. Results and Discussion

The five crimes against property in the analysis are robbery, theft and stealing, unlawful possession, vehicle theft and house and store breakings. Table 1 and Figure 1 show the distributions of the crimes.

Theft and stealing appear to have the highest average rate with the highest standard deviation among all the property crimes in Katsina State; robbery has the second average rate, while vehicle theft has the lowest average among property crimes in the state. The figure has shown a very high variability of the crimes among the LGAs of the state to a situation that the standard deviation of each crime is relatively than the average value. This is an indication of different cluster formation among the LGAs of the state for each of the crime.

Since GIS can be used to analyse a particular crime at a time it is complimented here in multivariate cluster analysis so as to identify different cluster classifications of property crimes in all the LGAs of Katsina state.

Figure 2a and Table 2 show the LGAs with the very higher levels of robbery rates. These are Faskari, Funtua and Malumfashi from AC2, Danmusa and Katsina from AC1. The crime has median rates in Mani, Sandamu and Daura from AC3, Jibia and Batsari from AC1 and Dandume from AC2, whereas Ingawa, Bindawa, Dutsi, Mashi, Maiadua and Zango from AC3, Musawa and Rimi LGAs from AC2 and 1 respectively have very lower rates of the crime are. The LGA with the highest rate in the state is Danmusa LGA. The southerly AC2 has the highest level while north-east AC3 has the lowest level of the crime. Therefore, the robbery prevalence in the state increases from north to south.

From Figure 2b, the LGAs with very higher rates of theft cases are Katsina and Funtua LGAs, higher rates at Daura and Baure from AC3 and Faskari from AC2. The LGAs with median rates of the crime are Mashi and Maiadua from AC3, Malumfashi, Bakori, Danja and Dandume from AC2, Danmusa, Dutsinma and Charanchi from AC1. The rest of the LGAs have either lower or very lower rates of the crime. AC2 has highest relative rates of the crime in the state, while Katsina LGA has the highest rate in the state.

There are no much vehicle theft prevalence cases in the state as we can see from Figure 3a; the only recorded cases are from five LGAs, with rate being low in Bakori, median in Mashi, higher in Dutsinma and Funtua and very higher in Katsina in LGAs. The rest of the LGAs have very lower incidences of the crime.

The LGAs from figure 3b with very higher rates of house and store breakings are Katsina and Kusada LGAs, and higher in Daura from AC3, Funtua and Sabuwa from AC2. The LGAs with median rates are Safana and Charanchi from AC1 and Zango from AC3, while the rest of the LGAs have either lower or very lower rates. Katsina LGA has the highest rate. All the extreme northern LGAs of AC1 and 3 that shared boarder with Niger republic, excepting Zango and Baure have very lower rate of the crime.

From Figure 4, Katsina and Rimi LGAs have the very higher rate of the unlawful possessions from AC1, followed by Safana and Mani LGAs. Only Funtua LGA from southern part of the state has at most the median rate, and therefore the crime is prevalence toward the northern part of the state.

The multivariate cluster analyses consider multiple crimes at a time for analysis, however. This could enhance exploring the overall distribution of the pattern of the crimes in a given geographical area.
The results of the Ward algorithm performed on the Euclidean distances are plotted and shown in Figure 5 and also summarised in Table 2. The resulting dendrogram suggested to splits the data set into five clusters.

The LGAs were later classified into higher and lower crime clusters using the averages of the standardized number of crimes which are summarized on Table 3.

Cluster 1 contains the LGAs with very low crime rates since the averages of the standardized number of all crimes are negative. Cluster 5 consisting of Katsina and Funtua LGAs has the higher property crime rate in the state. All the crimes have higher prevalence in cluster 5, except for robbery that is relatively more prevalence in cluster 3 that consist of Bakori, Malumfashi and Danmusa LGAs.

Cluster 2 corresponds to the LGAs with tendency towards unlawful possession, while cluster 4 corresponds to LGAs that have a considerable rate of theft and stealing. Base on the clusters averages, the clusters 1 to 5 can be level as very low, low, moderate, high and very high property crime rate respectively.

4. Conclusions

The property crime data set in Katsina State has shown that theft and stealing has the highest mean rate in the state followed by robbery, while vehicle theft has the lowest mean rate. The data set has high dispersion for all the crimes, which is an indication of high variability amongst the crime values of the LGAs.

The GIS classification has revealed that, Katsina, Danmusa, Malumfashi, Funtua and Faskari LGAs have the very highest rate of robbery in the state. The dispersion of the crime has covered the state especially in the LGAs of AC1 and 2. Theft cases have high prevalence in Katsina, Funtua, Daura, Baure and Faskari LGAs. The dispersion of the crime has al most equally cover the three ACs. There is low prevalence of vehicle theft in the state, and the significant reports are from Katsina, Funtua and Dutsinma LGAs. The LGAs with high prevalence of house and store breakings are Katsina, Funtua, Daura, Kusada and Sabuwa LGAs. The LGAs at the northern border with Niger republic, excepting ZNG and BRE have very lower rate of the crime, while all the LGAs at the southern border with Kaduna and Zamfara States have at least median rate of the crime. Unlawful possessions have shown high prevalence in the northern part of the state, and are higher in Katsina, Rimi, Mani and Safana LGAs.

The cluster analysis has combined all the crimes at a time for analysis, instead. Katsina and Funtua LGAs have the higher concentrations of all the property crimes in the state, while the LGAs with the lower concentrations of the property crimes are Ingawa, Musawa, Kaita, Dutsi, Kurfi, Kafur, Kankara, Sandamu, Jibia, Mani, Safana, Sabuwa, Kankia and Matazu LGAs. Robbery is relatively higher in Bakori, Mslumfashi and Danmusa LGA. Theft and stealing are prevalent in Dutsinma, Faskari, Baure and Daura LGAs.

Although, the techniques of GIS and cluster analysis are different, the latter has complemented the former in analysing property crime in Katsina State.

Table 1 Descriptive Statistics of the crimes

Variable	Mean	Median	StDev
Robbery	0.709	0.425	0.741
Theft/Stealing	2.786	2.450	2.861
Unlawful Possession	0.243	0.000	0.366
Vehicle Theft	0.107	0.000	0.305
H/S Breaking	0.463	0.350	0.519

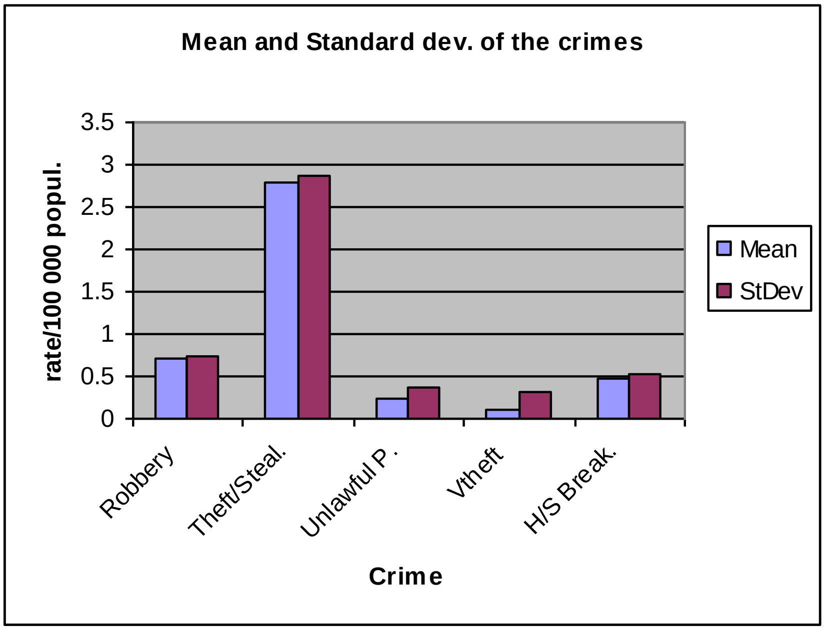
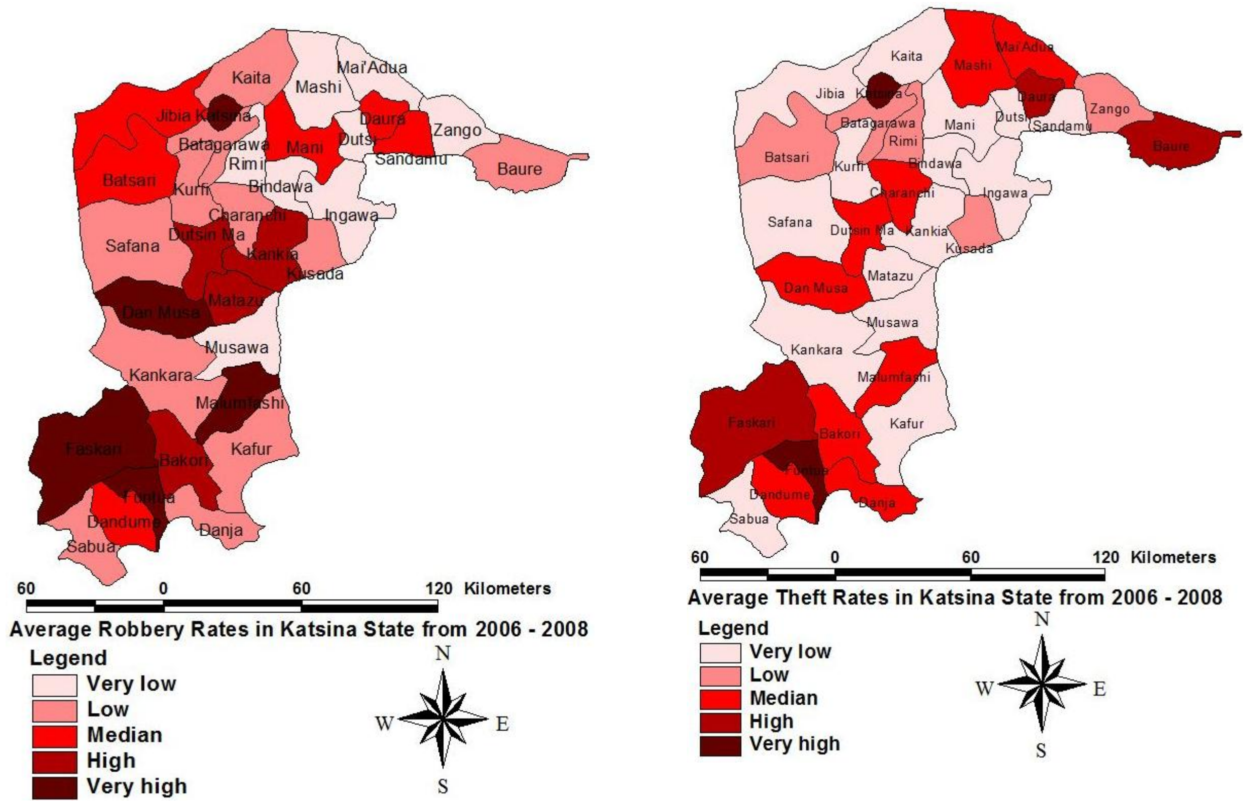
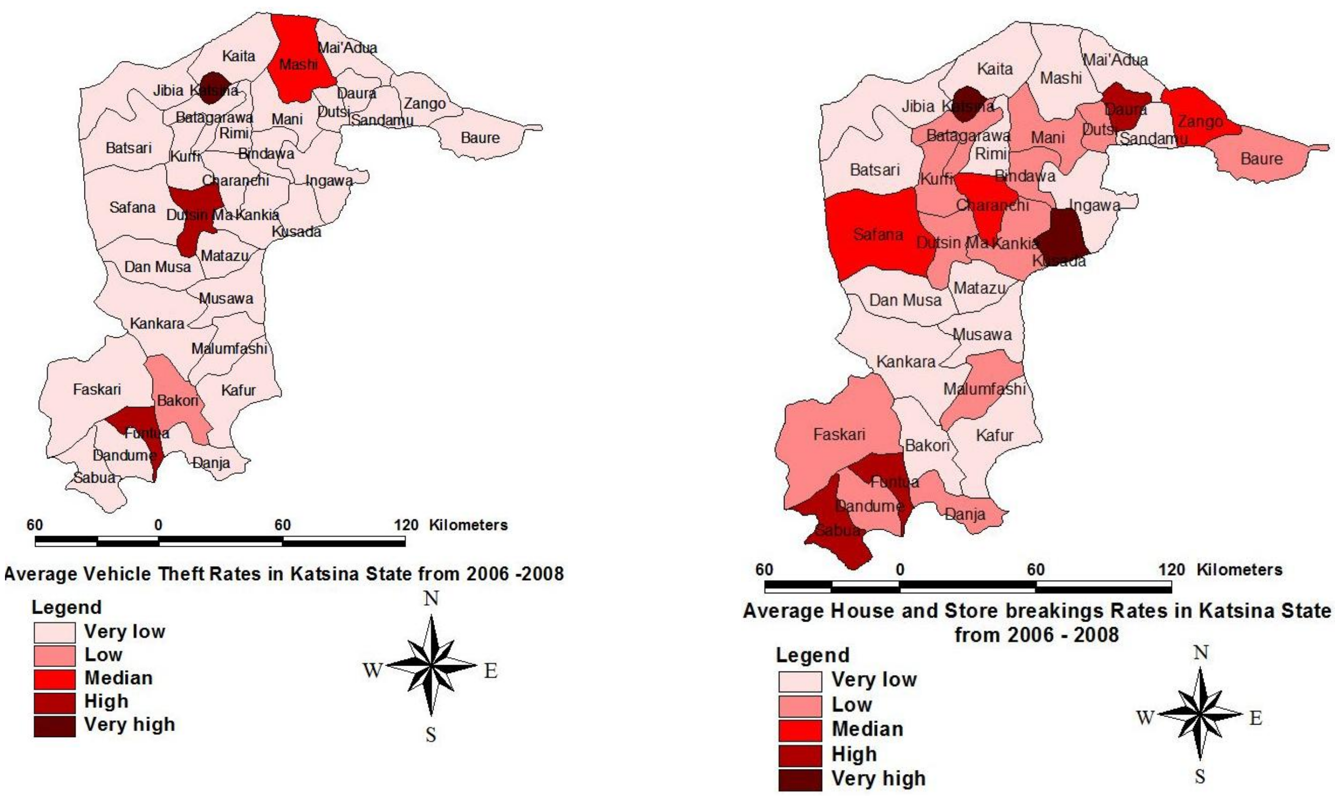


Figure 1: Mean and Standard Deviation of Crimes



(a) (b)

Figure 2: Average Robbery Rates in Katsina State



(a) (b)

Figure 3: Average of Theft Rates in Katsina State

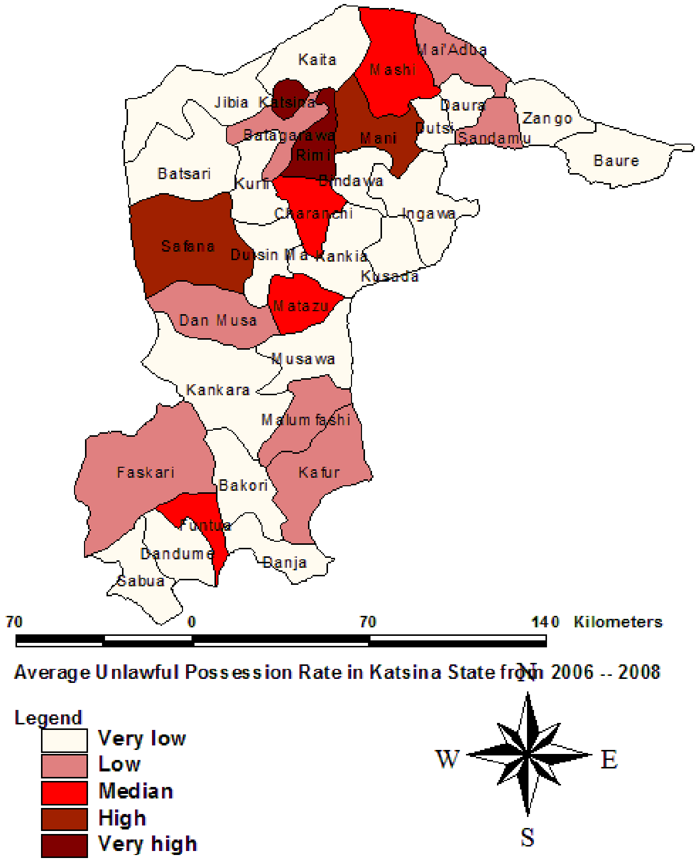


Figure 4: Average Unlawful Possession Rates in Katsina State

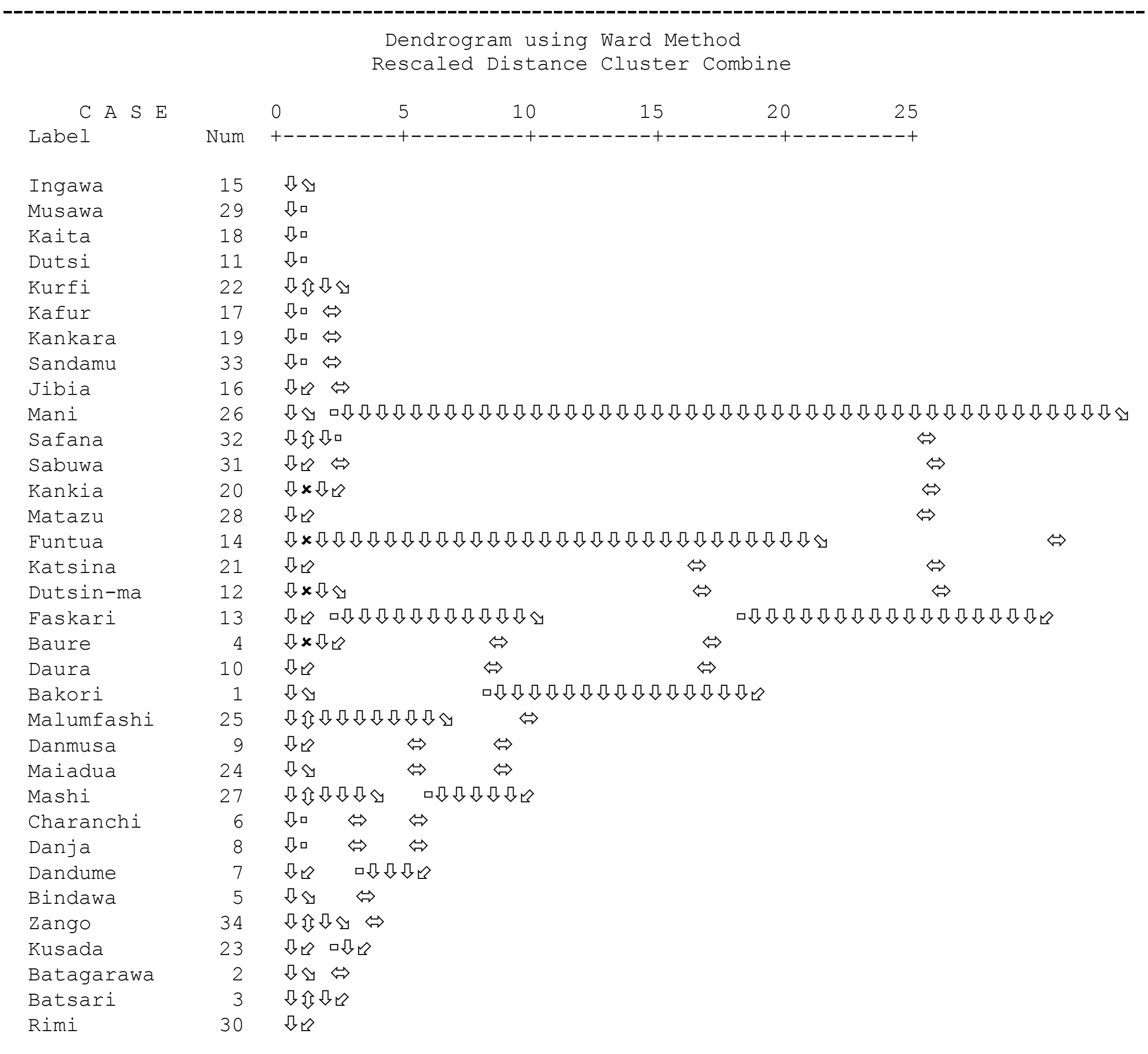


Figure 5: Dendrogram showing analysis of crimes in Katsina State

Table 2. Result of Ward method on Euclidean distances

CLUSTER 1	CLUSTER 2	CLUSTER 3	CLUSTER 4	CLUSTER 5
3Ingawa 3Sandamu	3Maiadua 3Kusada	2Bakori	1Dutsinma	2Funtua
2Musawa 1Jibia	3Mashi 1Batagarawa	2Malumfashi	2Faskari	1Katsina
1Kaita 2Mani	1Charanchi 1Batsari	1Danmusa	3Baure	
3Dutsi 1Safana	2Danja 1Rimi		3Daura	
1Kurfi 2Sabuwa	2Dandume			
2Kafur 2Kankia	3Bindawa			
2Kankara 2Matazu	3Zango			

Table 3. Average of the Standardised Crimes within the five clusters

	Robbery	Theft	Unlawful. P	V.Theft	H/Sbreak	Average
1	-0.241	-0.829	-0.033	-0.352	-0.359	-0.36
2	-0.646	0.039	0.167	-0.237	0.093	-0.12
3	1.836	0.298	-0.113	-0.108	-0.658	0.25
4	0.386	1.139	-0.314	0.292	0.364	0.37
5	1.712	2.861	2.116	3.330	2.247	2.45

References

Corcoran, J. J., Wilson, I. D. and Ware, J. A. (2003). Predicting the geo-temporal variations of Crime and disorder. *International Journal of Forecasting.* 19,. 623 – 634.

Fajemirokun, F., Adewale, O., Idowu,T., Oyewusi, A., and Maiyegun, B. (2006). A GIS Approach to Crime Mapping and Management in Nigeria: A Case Study of Victoria Island Lagos. www.oicrf.org/searchdocument.Asp?p=15&e=40 - 12k (Nov. 23, 2011)

Hardle, W. and Simar, L. (2003). Applied Multivariate Statistical Analysis, 2nd edition. New York: Springer-Verlag.

Hardle, W. and Zdenek, H. (2007). Multivariate Statistics: Exercises and Solutions. New York: Springer-Verlag.

Kpedekpo, G.M.C. and Arya, P.L. (1981). Social and Economics Statistics for Africa. London: George Allen and Unwin.

McGuffog, I., Western, J.S. and Mullins, P. (2001). Exploratory Spatial Data Analysis Techniques for Examining Urban Crime. *British Journal of criminology.* 41, 309 – 329.

Olivares, A. M. (1996). Classification of Prison Inmates based on Hierarchical Cluster Analysis. *Psicothema*. 8, 709 – 715.

Radoff, D. (1993), GIS Responds to Emergency Management. International GIS Sourcebook, (pp. 209-210).

Rencher, A. C. (2002), Methods of Multivariate Analysis, 2nd edn. New York: Wiley.

Richard, A. J. and Dean, W. W. (2001). Applied Multivariate Statistical Analysis, 3rd edn. New Delhi: Prentice-Hall.

Savoie, J. (2008). Analysis of the Spatial Distribution of Crime in Canada: Summary of Major Trends. *Crime and Justice Research Paper Series. Statistics Canada Catalogue no. 85-561-M-no.015.* (Nov. 23, 2011)

Ward, J. H. (1963). Hierarchical grouping methods to optimize an objective function. *Journal of the American Statistical Association.* 58, 236-244.