International Journal of Applied Science and Mathematical Theory E- ISSN 2489-009X
P-ISSN 2695-1908, Vol. 8 No. 4 2022 www.iiardjournals.org

Multivariate Analysis of Some Economics Data and Crime Figures in Nigeria

EZE-Emmanuel, Peace
Department of Mathematics(Statistics)
Rivers State University, Port Harcourt

D.O.I: 10.56201/ijasmt.v8.no4.2022.pg26.77

ABSTRACT
Multivariate Analysis of some Gross Domestic Product variables and crime figures in Nigeria was investigated to determine the important economic variables and crime figure variables that have positive effect on Nigerian society. This study examined the performance of these variables using yearly Crime data betwixt 2012 and 2020 and quarterly Nigerian Gross Domestic Product data from 1981Q1-2019Q1.The methods utilised are principal components analysis(PCA), factor analysis(FA) and cluster analysis(CA) multivariate techniques. Using R statistical software, the data were analysed. This research used three Rotation Methods of the Principal Components Factor Analysis to describe the variability in both data sets. Then, find the optimal number of clusters using four Clusters Identification approaches, and group variables into more homogenous groups. This research was able to identify significant differences between (None and Varimax rotation methods) and the Promax rotation methods of the Nigeria Economic variables data considered, while there are no significant differences among the three different rotation method results for crime data (i.e. large and small sample sizes). The communality and uniqueness of the factor analysis for the economic variables showed that principal components lies between 78.2% and 99.5% respectfully, of the total variability; the communality and uniqueness of the factor analysis for the large sample size (crime variables) principal components lies between 17.7% and 99.5% respectfully, of the total variability while the communality and uniqueness of the factor analysis for the small sample size (crime variables) principal components lies between 18.1% and 99.3% respectfully, of the total variability. This study was able to determine seven cluster groups for the economic variables, also seven cluster groups for the large sample size of crime rate variables and five cluster homogenous groups for the small sample size of crime rate variables.

Keywords: *Eigenvalues, Eigenvectors, Rotation, Cluster Analysis (CA), Factor Analysis (FA) And Principal Component Analysis (PCA)*

INTRODUCTION

Crime analysis is an important study to society life and it is of interest to this research because of the consequences and penalties it attracts from fine to death. Thus, this study will estimate an economic model and also show the relationships that exist among the various crime types.
Similarly, it is of great interest for this research to know the important Nigerian economic variables amongst others for the betterment of the country and states. Over the years, researchers have given attention to the subject of classification/factoring-out of variables into pre-determined groups or to reduce the redundancy among the variables by using a smaller number of factors. Complex problems and the results of bad decisions frequently force researchers to look for more objective ways to predict outcomes. That is why this research is interested in comparing results obtained from these three methods (Principal Components Analysis, Factor Analysis and Cluster Analysis) to determine the more suitable one when larger/smaller numbers of variables are involved. Dimensionality reduction is an important part of a pattern recognition system. It is a process in which we represent a system having many degrees of freedom by a smaller number of degrees of freedom. The main aim of the dimensionality reduction algorithms is to obtain a compact, accurate representation of the data that reduces or eliminates statistically redundant components. Principal Component Analysis, Factor Analysis and Cluster Analysis are some of the techniques, which can be used for dimensionality reduction. The aim of this study is to use Principal Components, Factor Analysis and Cluster Analysis to determine the important Nigerian economic variables amongst others and most crime variables that have effect on the Nigeria society. Thus, collected the data sets of Nigerian Gross Domestic Product (GDP) from 1981 to 2019 and Crime figure events in Nigeria from 2012 to 2020 (large and small data sets). The objectives of the study are to; use three Rotation Methods of the Principal Components Factor Analysis to describe the variability in both data sets, find the optimal number of clusters using four Clusters Identification approaches, then group them into more homogenous groups and Compare the results obtained from the two data sets using objectives I and II.

METHODOLOGY

Principal Components Analysis (PCA)

Principal Components Analysis is a useful statistical technique that has many applications in fields such as face recognition and image compression, and is a common technique for finding patterns in data of high dimension. It is concerned with explaining the variance – covariance, standard deviation, eigenvalue and eigenvectors structure through a few linear combinations of the original variables. Principal Components are particular (algebraically) i.e. linear combinations of the p random variables $X_1, X_2, \dots, X_P$. It depends solely on the covariance matrix or correlation matrix of $X_1, X_2, \dots, X_P$.
Consider the linear combinations

$$
\begin{aligned}
Y_1 &= L_1 X = L_{11}X_1 + L_{21}X_2 + \cdots + L_{P1}X_P \\
Y_2 &= L_2 X = L_{12}X_1 + L_{22}X_2 + \cdots + L_{P2}X_P \\
&\vdots \\
Y_p &= L_p X = L_{1p}X_1 + L_{2p}X_2 + \cdots + L_{PP}X_P
\end{aligned} \tag{3.1}
$$

where $L_1, L_2, \ldots, L_P$ are row vector and X's are column vector, such that $L_1 = (L_{11}, L_{21}, \ldots, L_{P1})$, $L_2 = (L_{12}, L_{22}, \ldots, L_{P2}) \cdots, L_P = (L_{1P}, L_{2P}, \ldots, L_{PP})$.

Variance − Covariance Matrix

We recall that covariance is always measured between 2 dimensions. There is usually more than one covariance measurement that may be calculated if we have a data set with more than 2 dimensions. For 3 dimensional data set, (dimension a, b, c), we have: Cov(a, b), Cov(a, c) and Cov(b, c).

$$\underline{\boldsymbol{X}} = \begin{bmatrix} X_1 \\ X_2 \\ . \\ . \\ . \\ X_P \end{bmatrix} \tag{3.2}$$

Then the population variance-covariance matrix (Abdi & Williams, 2010) is

$$\sigma_i^2 = \frac{\sum_{i=1} \left(\mathrm{X}_i - \bar{\mathrm{X}}\right)^2}{n-1} = \frac{\sum X_1^2 - n\bar{X}_1^{\,2}}{n-1} \tag{3.3}$$

$$\mathrm{Cov}(\mathrm{X}_i, \mathrm{X}_j) = \sigma_{ij} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \left(\mathrm{X}_i - \bar{\mathrm{X}}_i\right)\left(\mathrm{X}_j - \bar{\mathrm{X}}_j\right)}{n-1} \tag{3.4}$$

Eigenvalues and Eigenvectors

Many applications of matrices to technological problems involve A.X = λX; where A = $[a_{ij}]$ is a square matrix and λ is a number (scalar). Clearly X = 0 is a solution for any value of λ and is not normally useful. For non-trivial solution, i.e X ≠ 0, the values of λ are called the eigenvalues, character values or latent roots (polynomial) of matrix A and the corresponding solutions of the given equations A.X = λX are called the eigen vectors or characteristic vectors of A.

From the linear combination of Xi. i.e

$$\boldsymbol{\ell_{ij}} = \begin{bmatrix} \ell_{i1} \\ \ell_{i2} \\ . \\ . \\ . \\ \ell_{iP} \end{bmatrix} \quad (3.5)$$

$$AX - \lambda X = 0 , \quad (3.6a)$$

where the x's are the eigenvectors and

$$|A-\lambda I| = 0 \quad (3.6b)$$

λ's are the eigen-values.

Hence the coefficients ℓ_{ij} are collected into the vector.

$$\boldsymbol{\ell_{ij}} = \begin{bmatrix} \ell_{i1} \\ \ell_{i2} \\ . \\ . \\ . \\ \ell_{iP} \end{bmatrix} \text{ and } \lambda_i' = (\lambda_1, \lambda_2, \dots, \lambda_P)$$

Are Eigen-vectors and Eigenvalues.

First Principal Components Analysis (PCA 1)

Y_1, the first principal Components is the linear combination of x-variables that has maximum variance (among all linear combinations), being that it accounts or make up for as much variation in the data as possible. Specifically we will define coefficient $\ell_{11}, \ell_{12}, \dots, \ell_{1p}$ for that Components in such a way that its variance is maximized, subject to the constraint that the sum of the squared coefficients is equal to one. This constraint is required so that a unique value may be obtained.

More formally, select $\ell_{11}, \ell_{12}, \dots, \ell_{1p}$ that maximizes

$$\text{Var}(Y_1) = \sum_{i=1}^{p}\sum_{j=1}^{p} \ell_{1i}\ell_{1j}\sigma_{ij} = \ell_1' \sum \ell_1 \quad (3.7)$$

Subject to the constraint that

$$\ell_1'\ell_1 = \sum_{j=1}^{p} \ell_{1j}^2 = 1$$

Second Principal Components (PCA 2)

Y_2, the second principal Components is the linear combination of x- variables that accounts for as much of the remaining variations as possible, with the constraint that the correlation between the first and second Components is 0. Select $\ell_{21}, \ell_{22}, \ldots, \ell_{2p}$ that maximizes the variance of this new components

$$\mathrm{Var}(Y_2) = \sum_{i=1}^{p}\sum_{j=1}^{p} \ell_{2i}\ell_{2j}\sigma_{ij} = \ell_2' \Sigma \ell_2 \qquad (3.8)$$

Subject to the constraint that the sum of squared coefficients adds up to one

$$\ell_2' \ell_2 = \sum_{j=1}^{p} \ell_{2j}^2 = 1$$

Along with the additional constraint that these two Components will be uncorrelated with one another,

$$\mathrm{Cov}(Y_1, Y_2) = \sum_{i=1}^{p}\sum_{j=1}^{p} \ell_{1i}\ell_{2j}\sigma_{ij} = \ell_1' \sum \ell_2 = 0 \qquad (3.9)$$

All subsequent principal Components have same property, they are linear combinations that account for as much of the remaining variations as possible and they are not correlated with other principal Components. We will do this in the same way with each additional components.

Proportion

Proportion of variance explained by the k^th^ principal component is

$$\mathbf{P.V} = \frac{\lambda_k}{\lambda_1 + \lambda_2 + \ldots + \lambda_p} \times 100\% \qquad (3.10a)$$

where λ_k is the k^{th} eigenvalues.

Cumulative Proportion

Cumulative proportion of variance explained by the first k^{th} principal components is

$$\mathbf{C.P.V} = \frac{\lambda_1 + \lambda_2 + \ldots + \lambda_k}{\lambda_1 + \lambda_2 + \ldots + \lambda_p} \qquad (3.10b)$$

Factor Analysis

In factor analysis we represent the variables $y_1, y_2, \ldots, y_p$ as linear combinations of a few random variables $f_1, f_2, \ldots, f_m$ $(m < p)$ called *factors*. The factors are underlying constructs variables that "generate" the *y*'s. If the random sample $y_1, y_2, \ldots, y_n$ from a homogeneous population with mean vector μ and covariance matrix ∑ is taken.

The factor analysis model expresses each variable as a linear combination of underlying common factors $f_1, f_2, \ldots, f_m$, with an accompanying error term to account for that part of the variable that is unique (not in common with the other variables).

Thus, for $y_1, y_2, \ldots, y_p$ in any observation vector y, the model is given as

$$y_1 - \mu_1 = \lambda_{11} f_1 + \lambda_{12} f_2 + \cdots + \lambda_{1m} f_m + \varepsilon_1$$
$$y_2 - \mu_2 = \lambda_{21} f_1 + \lambda_{22} f_2 + \cdots + \lambda_{2m} f_m + \varepsilon_2$$
$$\vdots \qquad (3.11)$$
$$y_p - \mu_{p1} = \lambda_{p1} f_1 + \lambda_{p2} f_2 + \cdots + \lambda_{pm} f_m + \varepsilon_p$$

where:
Coefficients λ_{ij} are loadings and serve as weights, which shows each y individually depends on the f's (eigenvalues).
Note: *m* should be substantially smaller than *p*; otherwise we have not achieved a parsimonious description of the variables as functions of a few underlying factors.
A simple expression for the variance of y's is

$$\text{var}(y_i) = \lambda_{i1}^2 + \lambda_{i2}^2 + \ldots + \lambda_{im}^2 + \psi_i \qquad (3.12)$$

Thus, the emphasis in factor analysis is on modelling the covariance or correlations among the y's. Model (3.11) can be written in matrix notation as

$$y - \mu = \wedge f + \varepsilon \qquad (3.13)$$

Where
$y = (y_1, y_2, \cdots, y_p)', \mu = (\mu_1, \mu_2, \cdots, \mu_p)', f = (f_1, f_2, \cdots, f_m)', \varepsilon = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_p)'$ and

$$\wedge = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \cdots \lambda_{1m} \\ \lambda_{21} & \lambda_{22} & \cdots \lambda_{2m} \\ & \ddots & \\ \lambda_{p1} & \lambda_{p2} & \cdots \lambda_{pm} \end{bmatrix}$$

The assumptions can be expressed as follow

If $E(fj) = 0$, $j = 1,2,\ldots,m$

(i) $E(f) = 0$ (3.14)

$\text{var}(fj) = 1, \; j = 1, 2, \ldots, m$ and $\text{cov}(f_j, f_k) = 0$, $j \neq k$, therefore

(ii) $\text{cov}(f) = \text{I}$ (3.15)

$E(\varepsilon_i) = 0$, $i = 1,2,\ldots, p$

(iii) $E(\varepsilon) = 0$ (3.16)

$\text{var}(\varepsilon_i) = \psi_i, \; i = 1, 2, \ldots, p$ and $\text{cov}(\varepsilon_i, \varepsilon_k) = 0$, $i \neq k$, therefore

$$(4)\quad \operatorname{cov}(\varepsilon) = \psi = \begin{pmatrix} \psi_1 & 0 & \dots & 0 \\ 0 & \psi_2 & \dots & 0 \\ & & . & \\ & & . & \\ & & . & \\ 0 & 0 & \dots & \psi_p \end{pmatrix} \tag{3.17}$$

and $\operatorname{cov}(\varepsilon_i\ ,\ f_j) = 0\ \ for\ all\ i\ and\ j$

However, Equation (3.11) can be written as

$$\begin{aligned} \Sigma &= \operatorname{cov}(y) = \operatorname{cov}(\wedge\ f\ +\ \varepsilon) \\ &= \wedge \operatorname{cov}(f) \wedge' + \psi \\ &= \wedge \mathrm{I} \wedge' + \psi \\ \Sigma &= \wedge \wedge' + \psi \end{aligned} \tag{3.18}$$

Since μ does not affect variances and covariance of y.

Note: $\wedge$ has only a few columns, say two or three

In general,

$$\operatorname{cov}(y_i\ ,\ f_j) = \lambda_{ij}\ ;\quad i = 1,2,\dots,p\ \ and\ \ j = 1,2,\dots,m \tag{3.19}$$

Since λ_{ij} is the (ij)th element of $\wedge$, Equation (3.19) can be written as

$$\operatorname{cov}(y\ ,\ f) = \wedge \tag{3.20}$$

If standardized variables are used, Equation (3.18) is replaced by

$$\rho = \wedge \wedge' + \psi \tag{3.21}$$

and the loadings become correlations:

$$\operatorname{cov}(y_i\ ,\ f_j) = \lambda_{ij} \tag{3.22}$$

In partitioning variance of y_i into components due to the common factor in Equation (3.12), called the communality, and a components unique to y_i called the specific variance:

$$\begin{aligned} \sigma_{ij} = \operatorname{var}(y_i) &= \left(\lambda_{i1}^2 + \lambda_{i2}^2 + \dots + \lambda_{im}^2\right) + \psi_i \\ &= h_i^2 + \psi_i \\ &= communality + specific\ \ variance \end{aligned} \tag{3.23}$$

$Communality = h_i^2 = \lambda_{i1}^2 + \lambda_{i2}^2 + \dots + \lambda_{im}^2$ is also called common variance

$Specific\ \ variance = \psi_i$ is also called Specificity, unique variance, or residual variance.

The two factor analysis methods considered were Principal Component Approach (PCA) and Maximum Likelihood Method (MLM), while MLM is divided into three parts: (a) None Rotation Method of the factor analysis between the variables using Maximum Likelihood; (b) Varimax Rotation Method of the factor analysis between the variables; and (c) Promax Rotation of the factor analysis between the variables.

Cluster Analysis

Use clustering of observations to classify observations into groups when the groups are initially not known. This procedure uses an agglomerative hierarchical method that begins with all observations being separate, each forming its own cluster. In the first step, the two observations closest together are joined. In the next step, either a third observation joins the first two, or two other observations join together into a different cluster. This process will continue until all clusters are joined into one, however this single cluster is not useful for classification purposes. Therefore, you must decide how many groups are logical for your data and classify accordingly, using the following formulas;

For Euclidean distance,

$$d(i,k)=\sqrt{\sum_j \left(X_{ij}-X_{kj}\right)^2} \tag{3.24}$$

where d(i,k) is the distance between observations i and k.

For Manhattan distance,

$$d(i,k)=\sum_j \left|X_{ij}-X_{kj}\right| \tag{3.25}$$

where d(i,k) in row i and column k is the distance between observations i and k.

For Pearson distance,

$$d(i,k)=\sqrt{\sum_j \left(X_{ij}-X_{kj}\right)^2 \Big/ v_j} \tag{3.26}$$

where d(i,k) in row i and column k is the distance between observations i and k, and v(j) is the variance of variable j.

Note: there are also other methods like the Square of the Euclidean and Pearson methods.

For the correlation distance method,

$$d_{ij} = 1 - r_{ij} \tag{3.27}$$

where r_{ij} is the Pearson product moment correlation between variables i and j.

For the absolute correlation distance method,

$$d_{ij} = 1 - | r_{ij} | \tag{3.28}$$

where r_{ij} is the Pearson product moment correlation between variables i and j.

In average linkage, the distance between two clusters is the average distance between an observation in one cluster and an observation in the other cluster. In terms of the distance matrix,

$$d_{mj} = \frac{N_k d_{kj} + N_1 d_{lj}}{N_m} \tag{3.29}$$

where, Nk, N_l, and N_m are the number of observations in clusters k, l, and m

In centroid linkage, the distance between two clusters is the distance between the cluster centroids or means. In terms of the distance matrix,

$$d_{mj} = \frac{N_k d_{kj} + N_l d_{ij}}{N_m} - \frac{N_k N_l d_{kl}}{N_m^2} \tag{3.30}$$

where, N_k, N_l, and N_m are the number of observations in clusters k, l, and m

In complete linkage, or "furthest neighbor," the distance between two clusters is the maximum distance between an observation in one cluster and an observation in the other cluster. In terms of the distance matrix,

$d_{mj} = \max(d_{kj}, d_{lj})$ (3.31)

In McQuitty's linkage, the formula for the distance matrix =

$$d_{mj} = \frac{d_{kj}+d_{lj}}{2} \quad (3.32)$$

In median linkage, the distance between two clusters is the median distance between an observation in one cluster and an observation in the other cluster. In terms of the distance matrix,

$$d_{mj} = \frac{d_{kj}+d_{lj}}{2} - \frac{d_{kl}}{4} \quad (3.33)$$

In single linkage, or "nearest neighbor," the distance between two clusters is the minimum distance between an observation in one cluster and an observation in the other cluster. When observations lie close together, single linkage tends to identify long chain-like clusters that can have a relatively large distance separating observations at either end of the chain. In terms of the distance matrix,

$d_{mj} = \min(d_{kj}, d_{lj})$ (3.34)

In Ward's linkage, the distance between two clusters is the sum of squared deviations from points to centroids. The objective of Ward's linkage is to minimize the within-cluster sum of squares. In terms of the distance matrix,

$$d_{mj} = \frac{(N_j+N_k)d_{kj}+(N_j+N_l)d_{lj}-N_j d_{kl}}{N_j+N_m} \quad (3.35)$$

where, Nj, N_k, N_l, and N_m are the number of observations in clusters j, k, l, and m

In Ward's linkage, it is possible for the distance between two clusters to be larger than d(max), the maximum value in the original distance matrix, **D**. If this happens, the similarity will be negative.

K-means clustering begins with a grouping of observations into a predefined number of clusters.

i. Evaluates each observation, moving it into the nearest cluster. The nearest cluster is the one which has the smallest Euclidean distance between the observation and the centroid of the cluster.
ii. When a cluster changes, by losing or gaining an observation, recalculate the cluster centroid.
iii. This process is repeated until no more observations can be moved into a different cluster. At this point, all observations are in their nearest cluster according to the criterion listed above.

Unlike hierarchical clustering of observations, it is possible for two observations to be split into separate clusters after they are joined together. K-means procedures work best when you provide good starting points for clusters.

The final grouping of clusters (also called the final partition) is the grouping of clusters which will, hopefully, identify groups whose observations or variables share common characteristics. The decision about final grouping is also called cutting the dendrogram. The complete dendrogram (tree diagram) is a graphical depiction of the amalgamation of observations or variables into one cluster. Cutting the dendrogram is akin to drawing a line across the dendrogram to specify the final grouping.

How do you know where to cut the dendrogram? You might first execute cluster analysis without specifying a final partition. Examine the similarity and distance levels in the Session window results and in the dendrogram. You can view the similarity levels by placing your mouse pointer over a horizontal line in the dendrogram. The similarity level at any step is the percent of the minimum distance at that step relative to the maximum inter-observation distance in the data. The pattern of how similarity or distance values change from step to step can help you to choose the final grouping. The step where the values change abruptly may identify a good point for cutting the dendrogram, if this makes sense for your data. After choosing where you wish to make your partition, rerun the clustering procedure, using either Number of clusters or Similarity level to give you either a set number of groups or a similarly level for cutting the dendrogram. Examine the resulting clusters in the final partition to see if the grouping seems logical. Looking at dendrograms for different final groupings can also help you to decide which one makes the most sense for your data.

Results

Table 4.1: Comparison of Principal Components Factor Analysis of the Economic Variables using different Rotations

Rotation Method	**Variance explained criterion**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
None	SS Loadings	18.04	8.52	0.94	0.71	0.58	0.33	0.25	0.23	0.19	0.16	0.13	0.09	0.08
	Proportion	0.58	0.28	0.03	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.00
	Cumulative	**0.58**	**0.86**	**0.89**	**0.91**	**0.93**	**0.94**	**0.95**	**0.96**	**0.96**	**0.97**	**0.97**	**0.97**	**0.98**
Varimax	SS Loadings	17.94	8.20	1.37	0.86	0.56	0.29	0.28	0.22	0.18	0.10	0.10	0.07	0.07
	Proportion	0.58	0.27	0.04	0.03	0.02	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00
	Cumulative	**0.58**	**0.85**	**0.89**	**0.92**	**0.94**	**0.95**	**0.96**	**0.97**	**0.98**	**0.98**	**0.98**	**0.98**	**0.98**
Promax	SS Loadings	16.32	6.26	1.12	0.86	0.75	0.63	0.53	0.41	0.40	0.31	0.23	0.18	0.13
	Proportion	0.53	0.20	0.04	0.03	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.00
	Cumulative	**0.53**	**0.73**	**0.77**	**0.80**	**0.82**	**0.84**	**0.86**	**0.87**	**0.88**	**0.89**	**0.90**	**0.90**	**0.91**

Table 4.2: Factor Analysis of the Economic Variables using PCA for None, Varimax and Promax Rotation

Test of the hypothesis that 13 factors are sufficient.
The chi square statistic is 1025.64 on 140 degrees of freedom.
The p-value is 1.26e-134 **

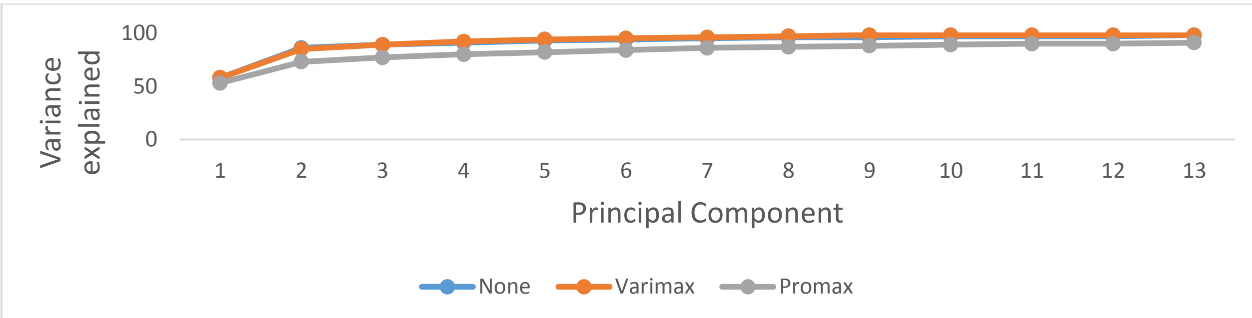

Figure 4.1: Economic Variables Scree Plot of the Cumulative Variance Explained by the Three Rotation Methods Principal Components

Table 4.2 and Figure 4.1 show results of the three different rotation methods used (None, Varimax and Promax rotation methods) and Table 4.7 shows the rotation methods chi-square statistic with its p-value result. This result identified only 13 factors are sufficient for describing the total variance accounted for in the economic variables. However, none and varimax rotation methods show similar variation, while the promax rotation method differ slightly in Figure 4.1.

Table 4.3: Communality and Uniqueness of the Factor Analysis

Variable	Communality	Uniqueness
crop production	0.991348	0.008652
live stock	0.995429	0.004571
Forestry	0.995044	0.004956
Fishing	0.984032	0.015968
cp&ng	0.992328	0.007673
metal ores	9.95E-01	4.96E-03
Quarrying &O M	9.34E-01	6.58E-02
oil refining	9.49E-01	5.09E-02
Cement	9.96E-01	3.68E-03
other manuf.	7.82E-01	2.18E-01
B&C	9.97E-01	2.86E-03
W&RT	9.95E-01	4.94E-03
road trans.	9.97E-01	3.40E-03
R T& Pipeline	8.93E-01	1.07E-01
water trans	9.81E-01	1.87E-02
air trans	9.95E-01	4.75E-03

O T S	9.96E-01	4.32E-03
Telecom.	9.95E-01	4.85E-03
Post	9.62E-01	3.84E-02
Electricity	9.79E-01	0.020612
Water	9.96E-01	4.23E-03
Hotels & Rest.	9.95E-01	4.71E-03
financial inst.	9.88E-01	1.18E-02
Insurance	9.96E-01	3.93E-03
real estate	0.997812	0.002188
business S	8.98E-01	1.02E-01
Pub Admin	9.97E-01	3.30E-03
Education	9.97E-01	3.42E-03
Health	9.99E-01	1.41E-03
Other Services	9.91E-01	8.70E-03
Broadcasting	0.996172	0.003828

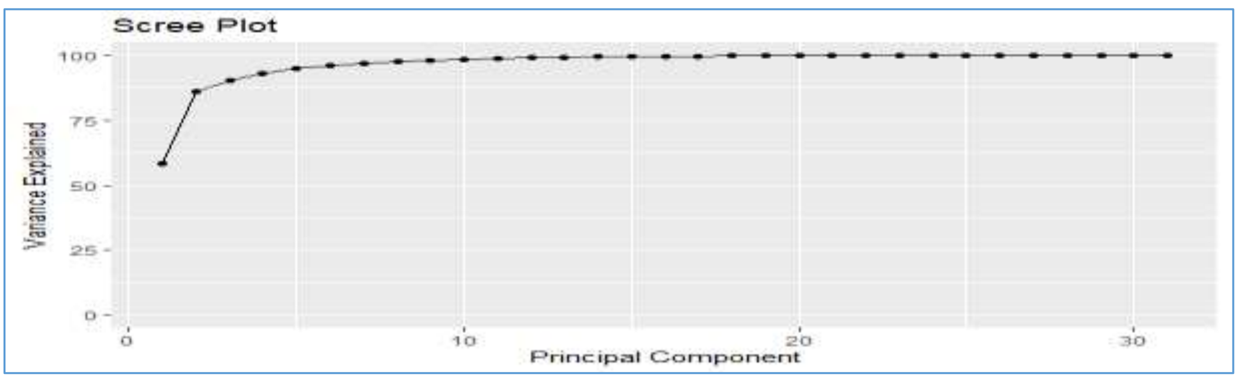

Figure 4.2: Economic Variables Scree Plot of the Cumulative Variance Explained by Principal Components Factor Analysis

Table 4.3 and Figure 4.2 show the communality and uniqueness of the factor analysis together, where all the economic variables principal components lie between 78.2% and 99.7% respectfully, of the total variability. Thus, most of the data structure can be captured in the first-thirteen principal component underlying dimensions. The remaining principal components account for a very small proportion of the variability and are probably unimportant. The Scree plot provides this information visually (Figure 4.2).

Cluster Analysis of the Economic Variables

In order to find the optimal number of clusters for a *k*-means, the following four approaches were considered in this work: 1) Elbow method (which uses the within cluster sums of squares); 2) Average silhouette method; 3) Gap statistic method and 4) NbClust() function (package) to implement the D index method. The above methods are displayed below in Figures 4.3 and its scree plot, 4.4 and 4.5.

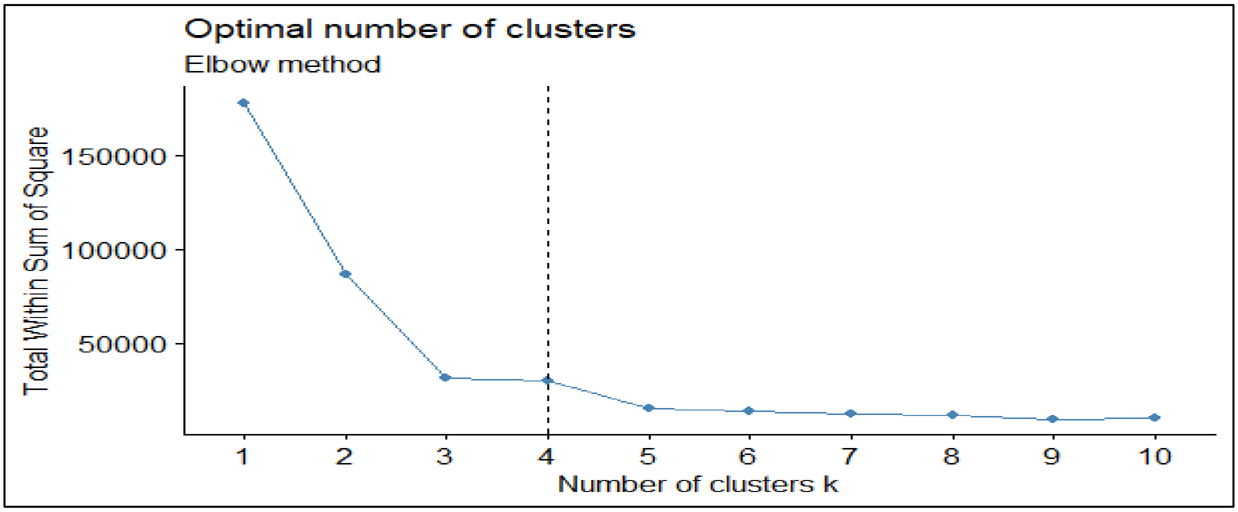

Figure 4.3: Elbow Method for Number of Clusters Identification for the Economic Variables

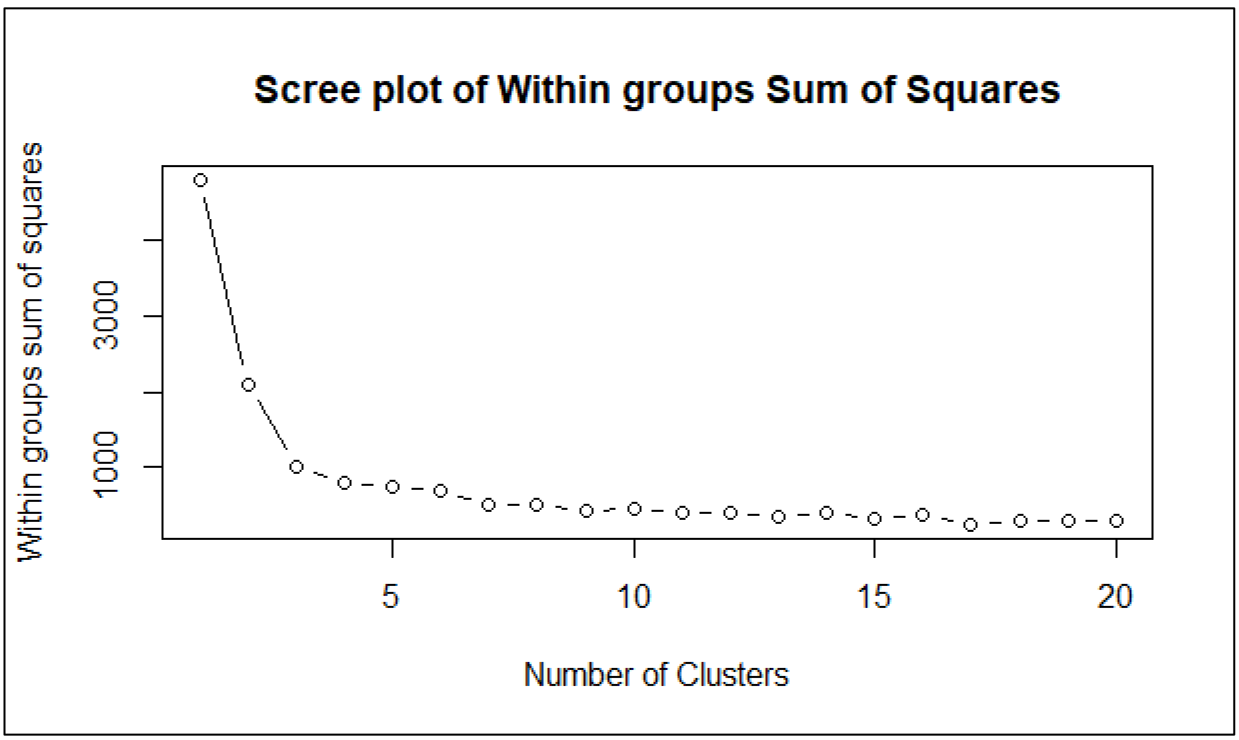

Figure 4.8: Scree Plot of the Within Cluster Sum of Squares for the Economic Variables

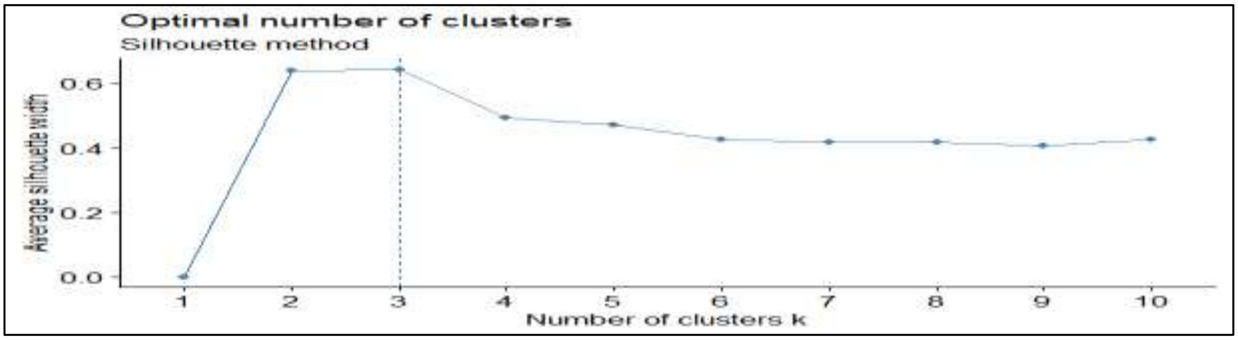

Figure 4.4: Silhouette Method for Number of Clusters Identification for the Economic Variables

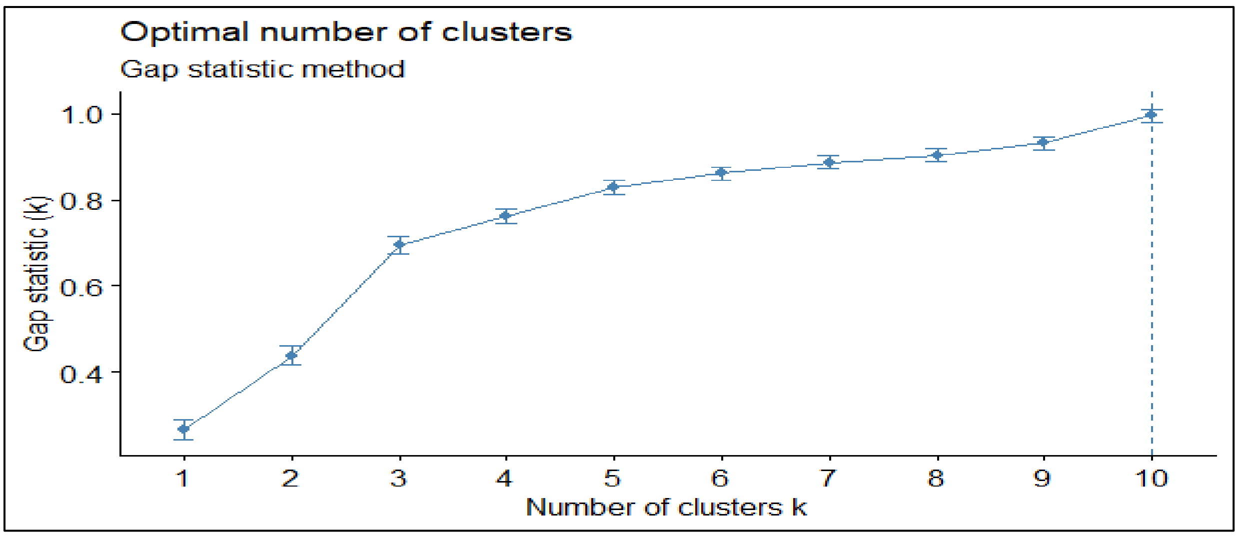

Figure 4.5: Gap Statistic Method for Number of Clusters Identification for the Economic Variables

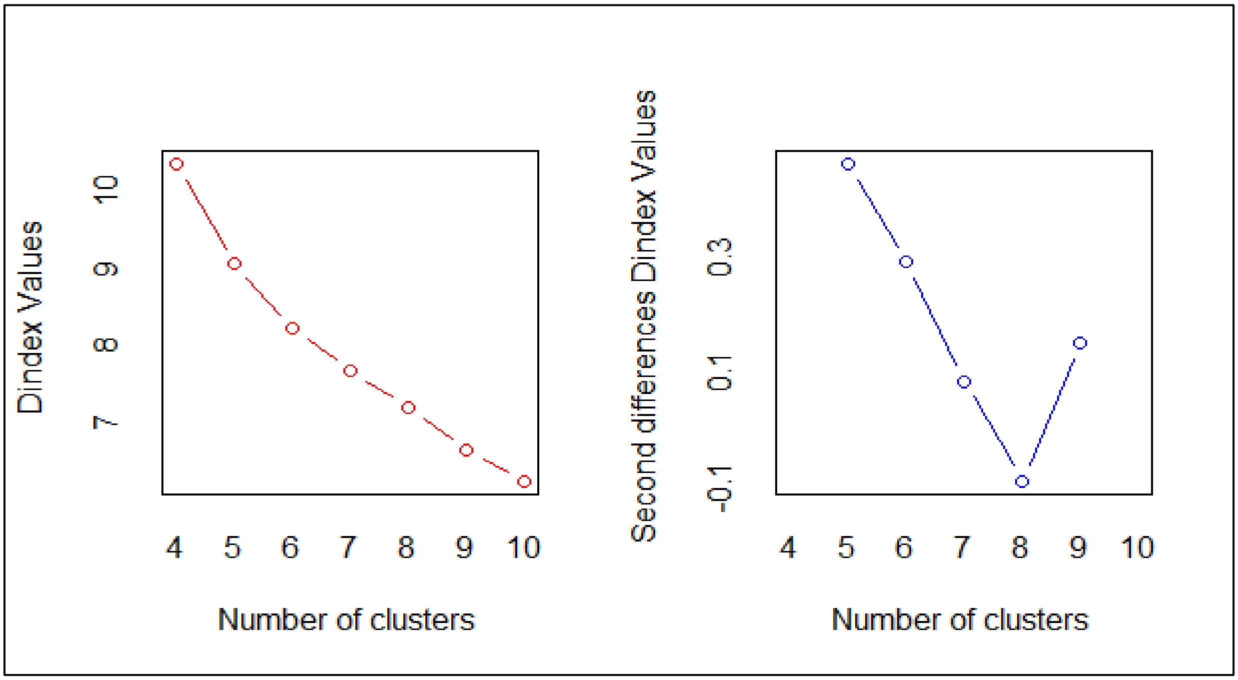

Figure 4.6: D Index Statistic Method for Number of Clusters Identification I for the Economic Variables

The Elbow method seems to suggest 4 clusters. The location of a knee in the plot is usually considered as an indicator of the appropriate number of clusters because it means that adding another cluster does not improve much better the partition; the Silhouette method seems to suggest 3 clusters. Note that the Silhouette method measures the quality of a clustering and determines how well each point lies within its cluster. While the Gap Statistic method suggests 10 clusters. The three methods done so far on the economic variables are suggesting that the optimal number of clusters is between 3 and 10. The D index is a graphical method of determining the number of clusters. In the plot of D index, we seek a significant knee (the significant peak in D index second differences plot) that corresponds to a significant increase of the value of the measure. So, the D index method constructed using the Euclidean distance and permuted optimal number of factors around the interval 3 and 10; for 3 as the minimum and 10 as the maximum number of clusters gave the result displayed in Figure 4.6a above and the corresponding optimal number in Figure 4.6b suggest 8 clusters.

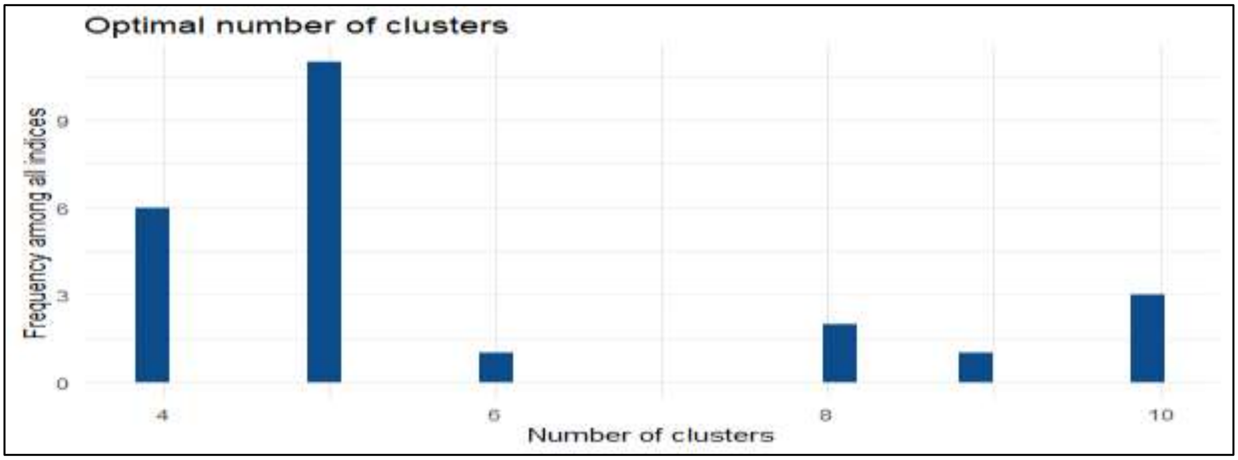

Figure 4.7: D Index Statistic Method for Number of Clusters Identification II for the Economic Variables

In Figure 4.7 among all indices, 6 proposed 4 as the best number of clusters, 11 proposed 5 as the best number of cluster, 1 proposed 6 as the best number of clusters, 0 proposed 7 as the best number of cluster, 2 proposed 8 as the best number of clusters, 1 proposed 9 as the best number of clusters and 3 proposed 10 as the best number of clusters. According to the majority rule, the best number of clusters is 10. Finally, the four methods have given an idea of the number of clusters that is optimal in grouping the economic variables' data. Hence, the highest number of clusters is chosen and this number is 10. To confirm that this chosen number of classes is indeed optimal, there is a way to evaluate the quality of your clustering via the silhouette plot (which shows the silhouette coefficient on the *y* axis). We draw the silhouette plot for 10 clusters, as chosen. The silhouette plot for economic variables shows that the best number of clusters is 7 in Figure 4.9. More so, the cluster mapping output for economic variables and the grouping of the economic variables are shown in Figures 4.10, 4.11 and Table 4.9.

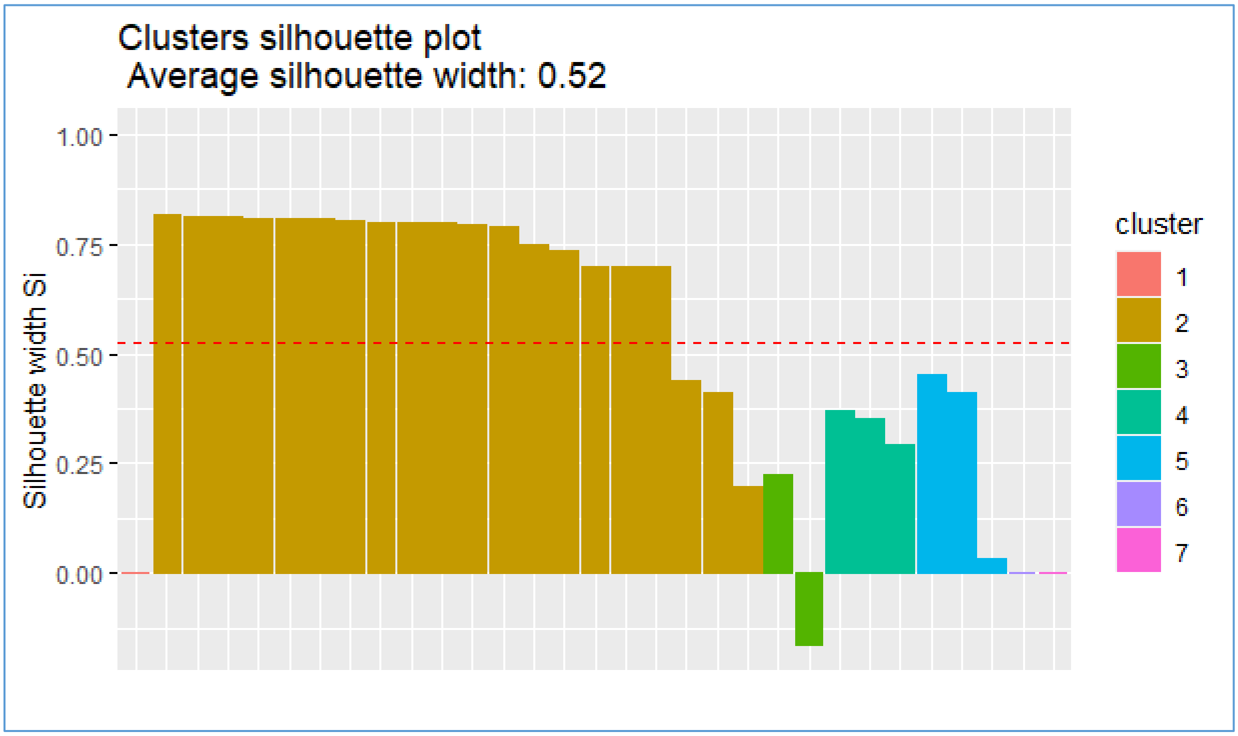

Figure 4.9: Silhouette Plot for Economic Variables

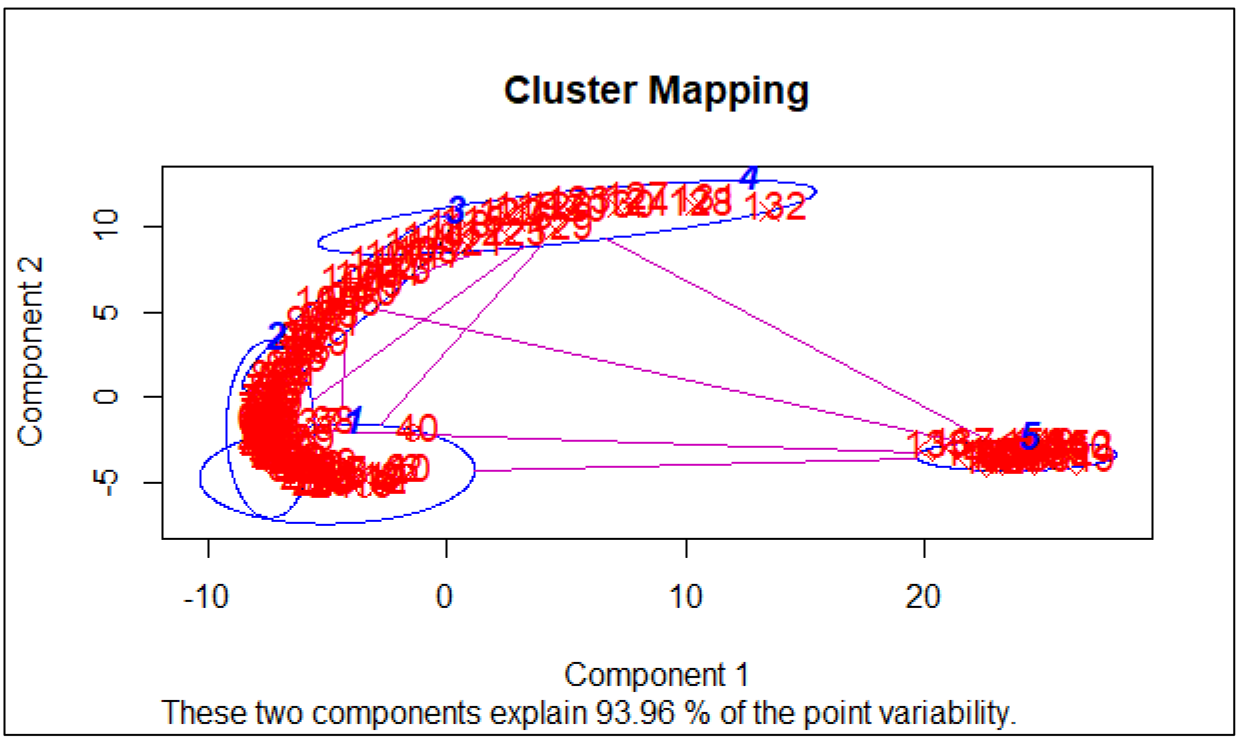

Figure 4.10: Cluster Mapping Output for Economic Variables

A principal component analysis is performed to represent the variables in a 2 dimensions' plane as shown in Figure 4.10. The two components from the principal component analysis explain about 93.96% of the point variability for the cluster model. The cluster plot in Figure 4.11 gives further clarification on the percentage contribution of each of these principal components.

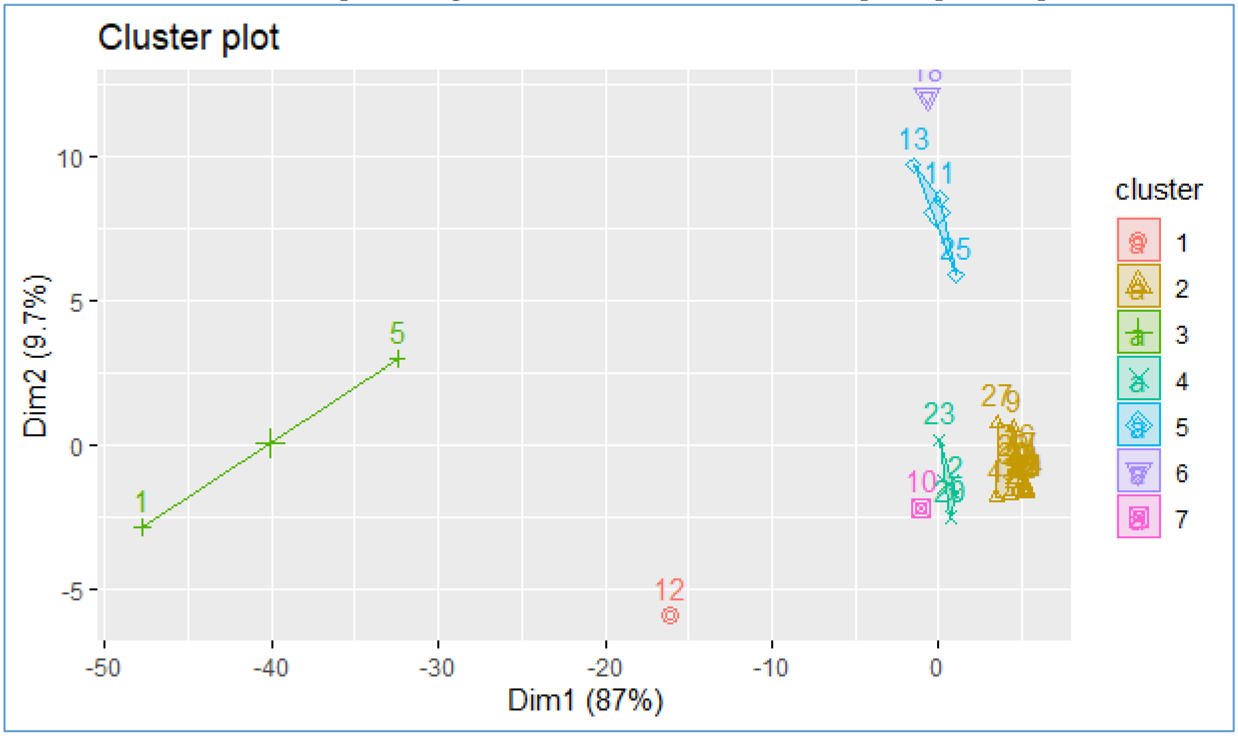

Figure 4.11: Cluster Plot Output for Economic Variables

Having determined the optimal number of clusters, the next step is to normalize the data set. The data set is normalized and k-means cluster analysis was conducted. Hence, the economic variables data chooses best number of clusters which is 7, is as follow in Table 4.9. Clusters 1, 6 and 7 have one economic variable each; while Cluster 3 has two economic variables; Clusters 4 and 5 have three economic variables each and Cluster 2 has the highest number of economic variables with 20 variables.

Crime Variables Analysis

Normal distribution in Minitab 21 was used to generate small and large sample. 29 random numbers were generated from a normal distribution with their means and standard deviations for small sample. 40 random numbers were simulated plus the actual 9 observations to give the large sample (that is 49 observations for large sample).

Table 4.4: Comparison of Principal Component Factor Analysis using Different Rotation Methods for Large Sample Size of Crime Variables

Rotation Method	Variance explained criterion	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
None	SS Loadings	1.66	1.642	1.453	1.067	1.03	0.838	0.793
	Proportion	0.098	0.097	0.085	0.063	0.061	0.049	0.047
	Cumulative	**0.098**	**0.194**	**0.28**	**0.342**	**0.403**	**0.452**	**0.499**
Varimax	SS Loadings	1.541	1.391	1.254	1.213	1.184	1.021	0.878
	Proportion	0.091	0.082	0.074	0.071	0.07	0.06	0.052
	Cumulative	**0.091**	**0.172**	**0.246**	**0.318**	**0.387**	**0.447**	**0.499**
Promax	SS Loadings	1.448	1.412	1.352	1.343	1.228	1.189	1.04
	Proportion	0.085	0.083	0.08	0.079	0.072	0.07	0.061
	Cumulative	**0.085**	**0.168**	**0.248**	**0.327**	**0.399**	**0.469**	**0.53**

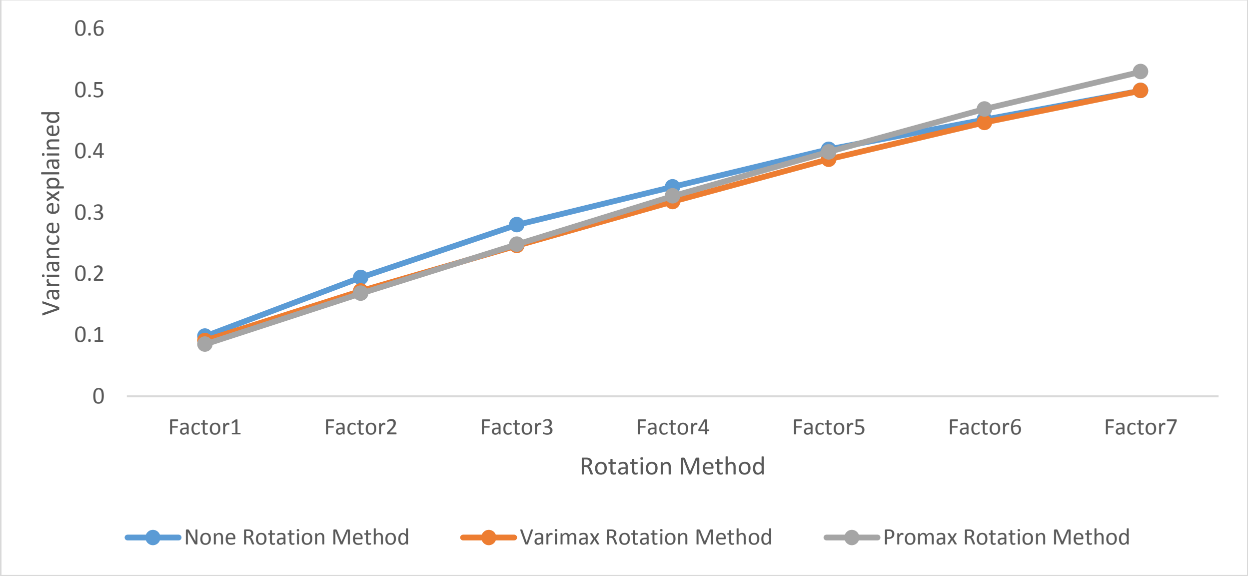

Figure 4.12: Large Sample Size for Crime Variables Scree Plot of the Cumulative Variance Explained by the Rotation Methods Principal Component

Table 4.5: Factor Analysis using PCA for Promax Rotation of the Large Sample Size for Crime Variables

Call:

```
factanal(x = mydata, factors = 7, rotation = "promax")

Uniquenesses:
   X1     X2     X3     X4     X5     X6     X7     X8
0.681  0.356  0.097  0.756  0.46   0.238  0.598  0.723
   X9    X10    X11    X12    X13    X14
0.629  0.823  0.789  0.738  0.005  0.203
  X15    X16    X17
 0.41  0.464   0.55

Loadings:
     Factor1 Factor2 Factor3 Factor4 Factor5 Factor6 Factor7
X1    0.354  -0.116  -0.331  -0.212
X2   -0.179   0.113   0.121  -0.106   0.119   0.817  -0.174
X3     0.99   -0.12  -0.142   0.125
X4    0.198  -0.102  -0.131  -0.464
X5    0.158   0.164  -0.179   0.735
X6    0.933  -0.214  -0.134   0.126   0.122
X7    0.328  -0.153     0.5    0.16
X8    0.341   -0.12   0.218   0.277
X9    0.329   0.305  -0.155  -0.187
X10    0.27  -0.146   0.157
X11  -0.131   0.388
X12   0.105   0.254   0.368    0.11
X13  -0.146   1.025  -0.101   0.192
X14  -0.125  -0.215   0.928
X15   -0.24   0.759
X16   0.241  -0.103   0.335   0.497
X17   0.228   0.394   0.127  -0.195   0.221  -0.139

Test of the hypothesis that 7 factors are sufficient.
The chi square statistic is 26.03 on 38 degrees of freedom.
The p-value is 0.93
```

Table 4.5 and Figure 4.12 show the results of the three different rotation methods used (None, Varimax and Promax rotation methods) and Table 4.15 shows the rotation methods chi-square statistic with its p-value result. This result identified only 7 factors are sufficient for describing the total variance accounted for in the large sample size for crime variables, since the p-value

(0.93) is not significant. However, the three rotation methods display similar variation in Figure 4.12.

Table 4.6: Communality and Uniqueness of the Factor Analysis of the Large Sample Size for Crime Variables

Variable	Full Name	Communality	Uniqueness
X_1	Homicide	0.319152	0.680848
X_2	Offences against morality	0.644443	0.355557
X_3	Other offences against persons	0.903217	0.096783
X_4	Robbery	0.243693	0.756307
X_5	Breakings	0.539893	0.460107
X_6	Theft of stock	0.762216	0.237784
X_7	Stealing	0.401838	0.598162
X_8	Theft by servant	0.277178	0.722822
X_9	Vehicle and other thefts	0.37109	0.62891
X_{10}	Dangerous drugs	0.17691	0.82309
X_{11}	Serious Traffic offences	0.211304	0.788696
X_{12}	Criminal damage	0.262148	0.737852
X_{13}	Economic crimes	0.995004	0.004996
X_{14}	Corruption	0.797176	0.202824
X_{15}	Offences involving police officers	0.59023	0.40977
X_{16}	Offences involving tourist	0.536142	0.463858
X_{17}	Other penal code offences	0.450049	0.549951

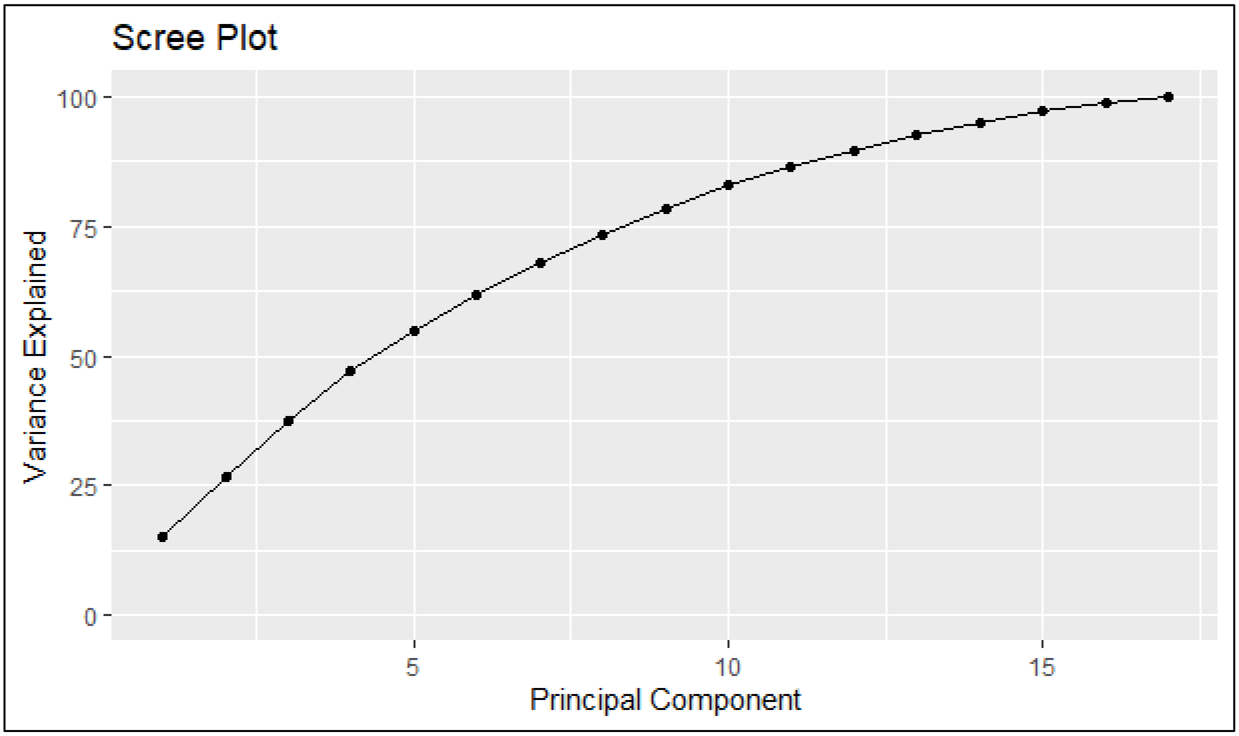

Figure 4.13: Large Sample Size for Crime Variables Scree plot of the Cumulative Variance Explained by Principal Components

Table 4.6 and Figure 4.13 show the communality and uniqueness of the factor analysis together, where all the large sample size for crime variables principal components lie between 17.7% and 99.5% respectfully, of the total variability. Thus, most of the data structure can be captured in the seven principal components underlying dimensions from Figure 4.13, since it accounts for almost 75% proportion of the variability. The remaining principal components account for 25% proportion of the variability and are probably unimportant. The Scree plot provides this information visually (Figure 4.13).

Cluster Analysis of the Large Sample Size for Crime Rate Variables

Similarly, in order to find the optimal number of clusters for a *k*-means, the following four approaches were considered in this work: (1) Elbow method (which uses the within cluster sums of squares); (2) Average silhouette method; (3) Gap statistic method and (4) NbClust() function (package) to implement the D index method. The Figures of the above methods are displayed in Figures 4.14, 4.15 and 4.16.

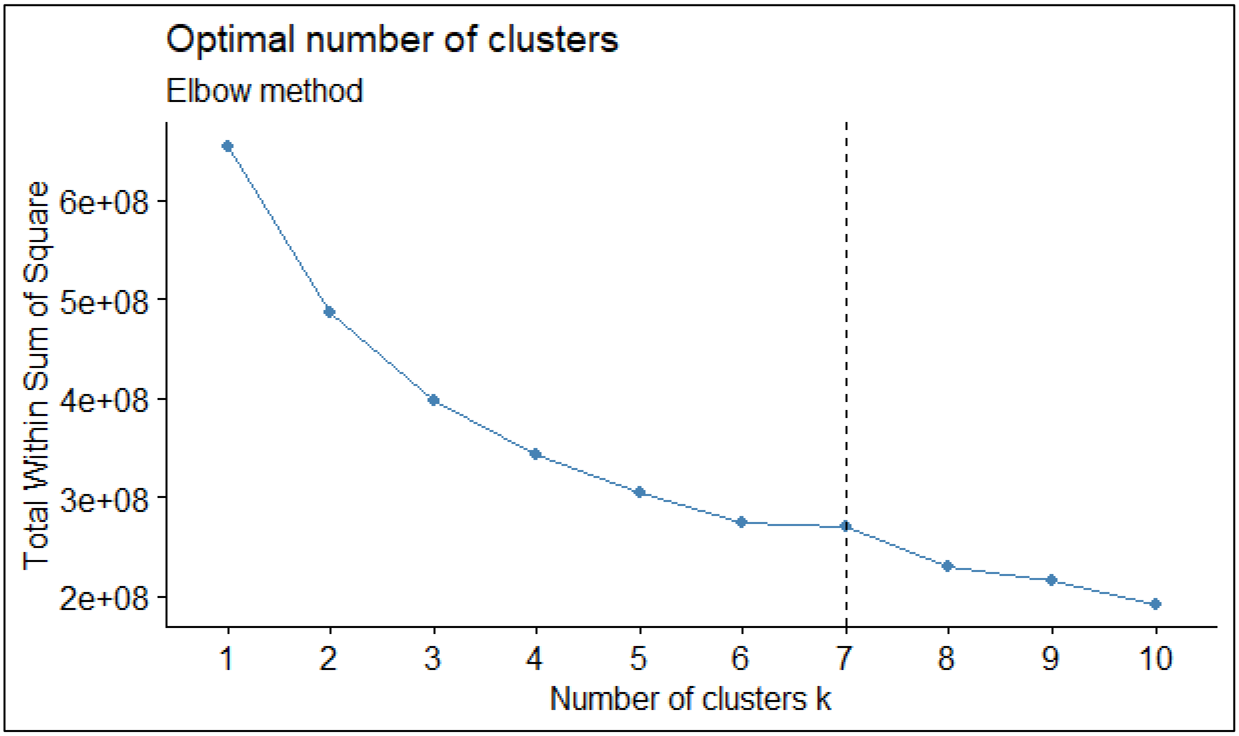

Figure 4.14: Elbow Method for Number of Clusters Identification for Large Sample Size of the Crime Variables

Figure 4.14 shows that the Elbow method seems to suggest 7 clusters. The location of a knee in the plot is usually considered as an indicator of the appropriate number of cluster because it means that adding another cluster does not improve much better the partition.

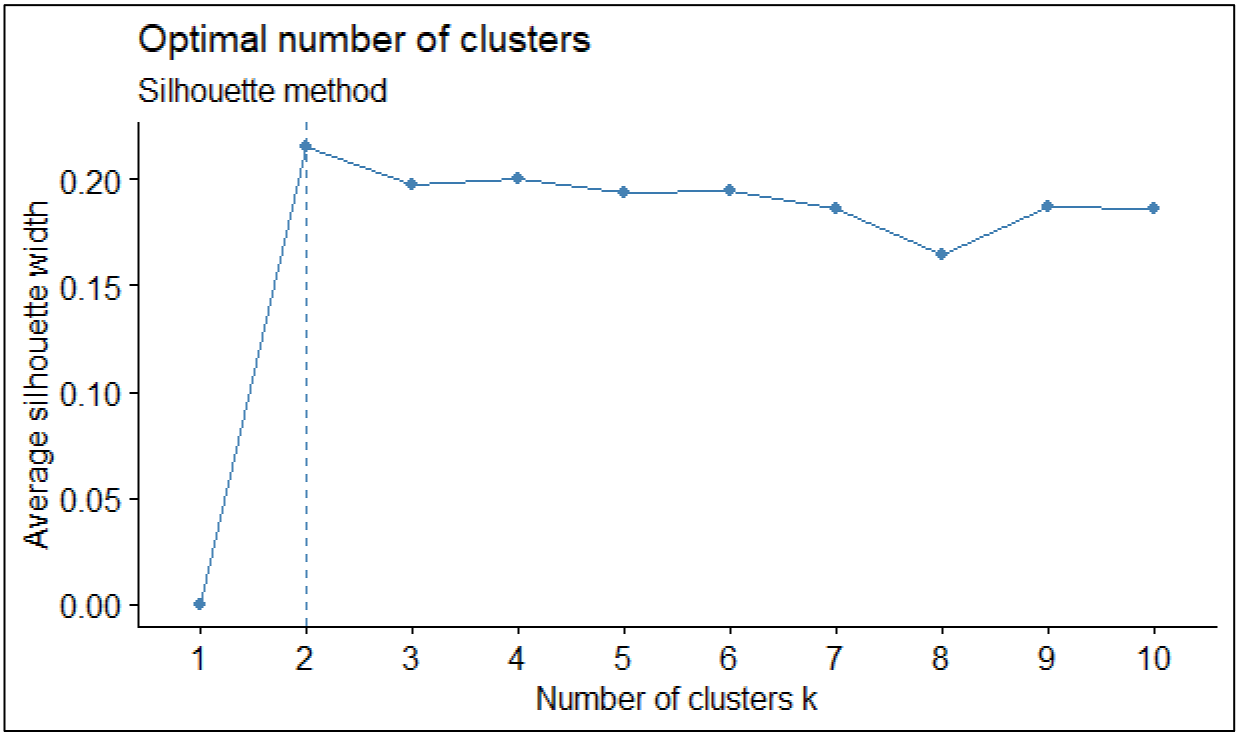

Figure 4.15: Silhouette Method for Number of Clusters Identification for Large Sample Size of the Crime Variables

Figure 4.15 Depicts that the Silhouette method seems to suggest 2 clusters. Note that the Silhouette method measures the quality of a clustering and determines how well each point lies within its cluster.

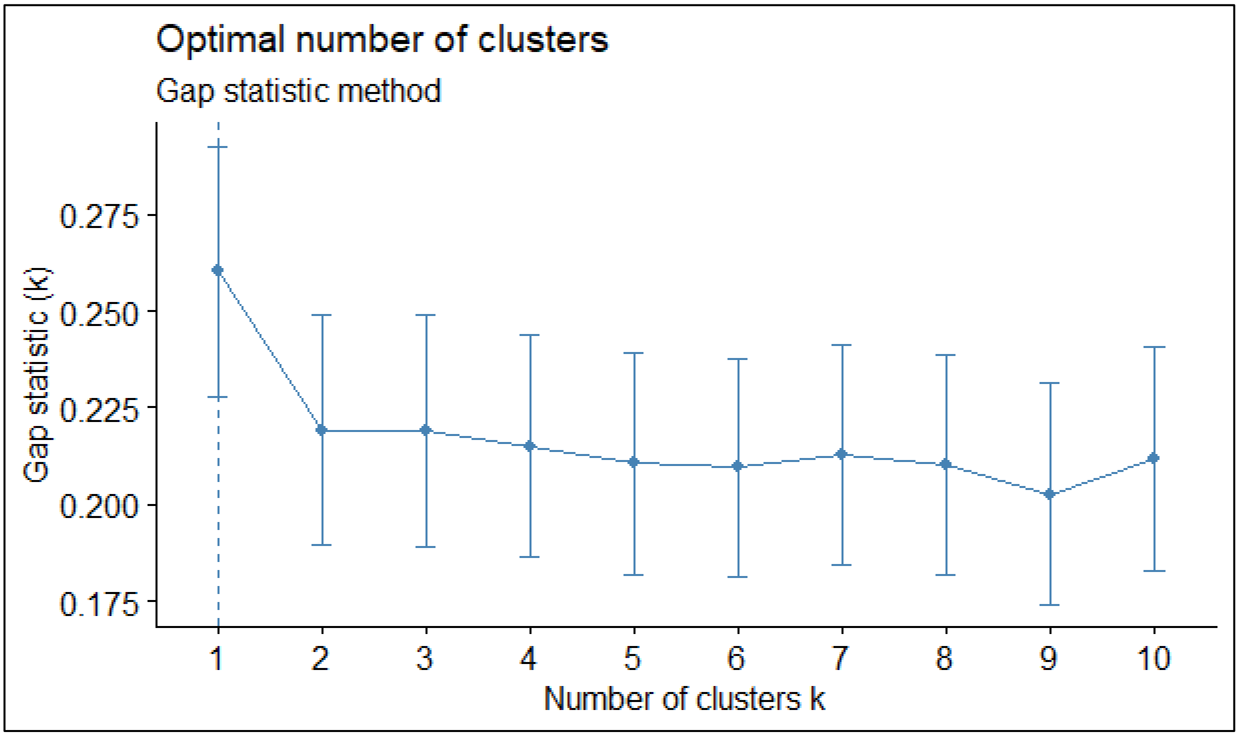

Figure 4.16: Gap Statistic Method for Number of Clusters Identification for Large Sample Size of the Crime Variables

Figure 4.16 shows that the Gap Statistic method suggests 10 clusters. The three methods done on large sample size for crime variables so far are suggesting that the optimal number of clusters is between 2 and 10. In addition to these methods above, the D index method constructed below using the Euclidean distance and permuted optimal number of factors around the interval 4 and 5; for 4 as the minimum and 5 as the maximum number of clusters gave the result displayed in Figure 4.17a and the corresponding optimal number in Figure 4.17b suggests 5 clusters.

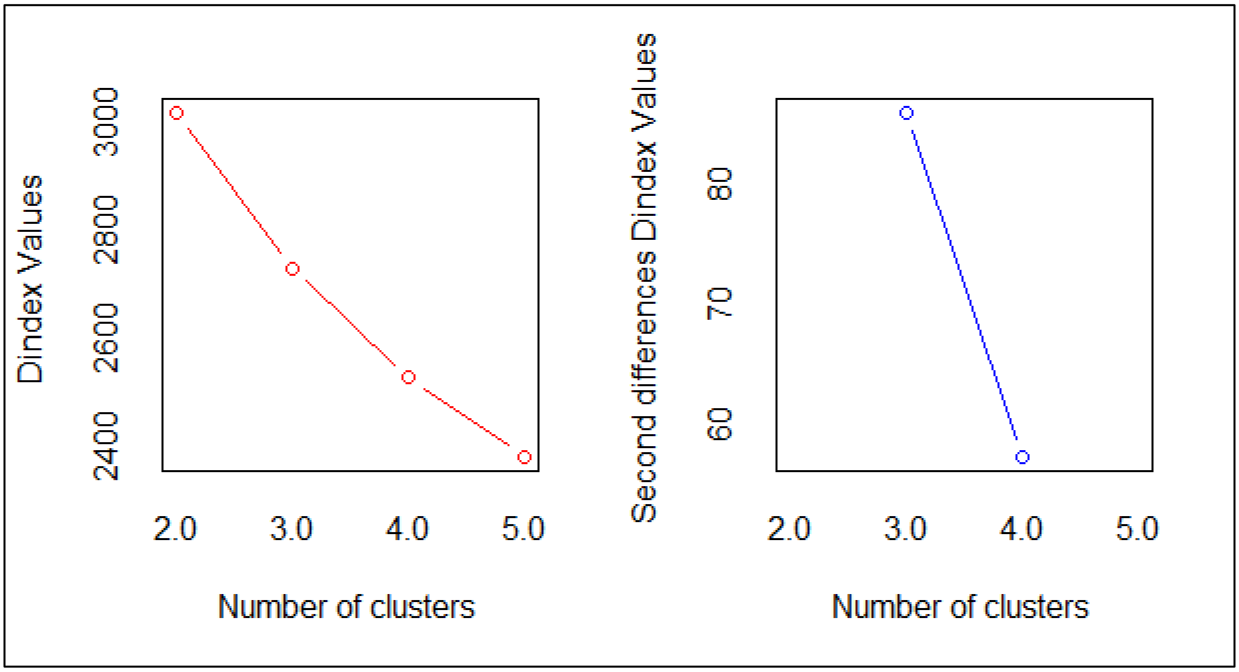

Figure 4.17: D Index Statistic Method for Number of Clusters Identification I for Large Sample Size of the Crime Variables

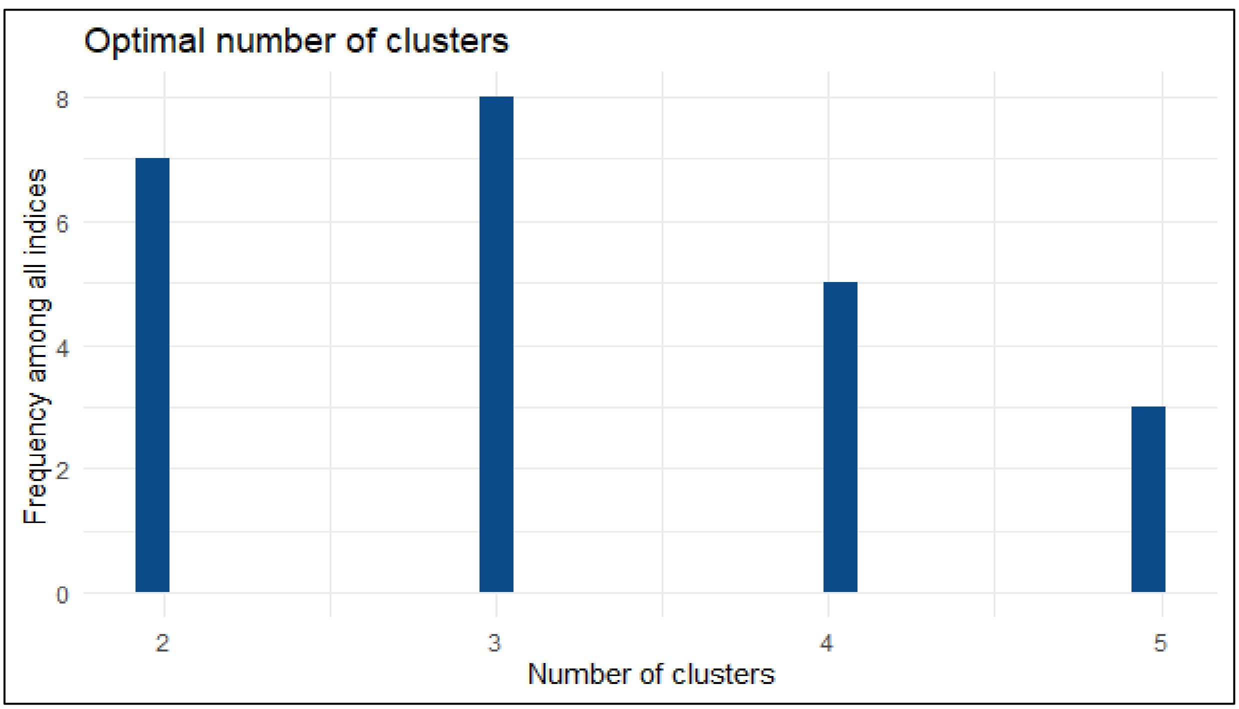

Figure 4.18: D Index Statistic Method for Number of Clusters Identification II for Large Sample Size of the Crime Variables

In Figure 4.18, 7 proposed 2 as the best number of clusters, 8 proposed 3 as the best number of clusters, 5 proposed 4 as the best number of cluster and 3 proposed 5 as the best number of clusters. According to the majority rule, the best number of cluster is 5, however Figure 4.19 proposed 7 as the best number of clusters for large sample size of the crime variables.

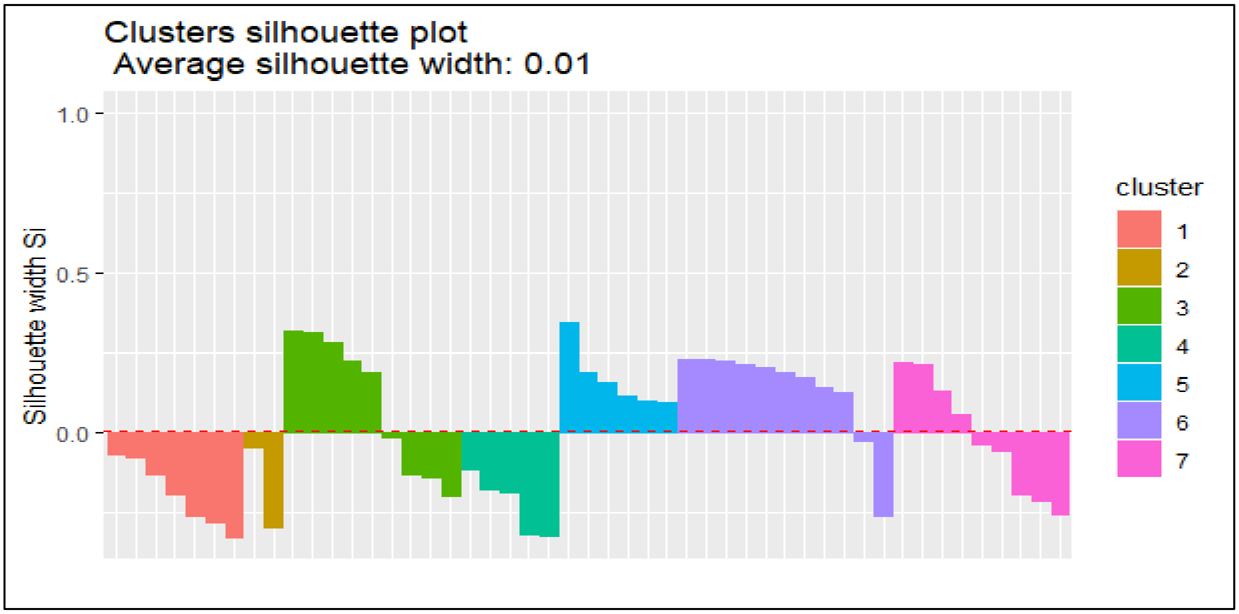

Figure 4.19: Silhouette Plot for the Years II for Large Sample Size of the Crime Variables

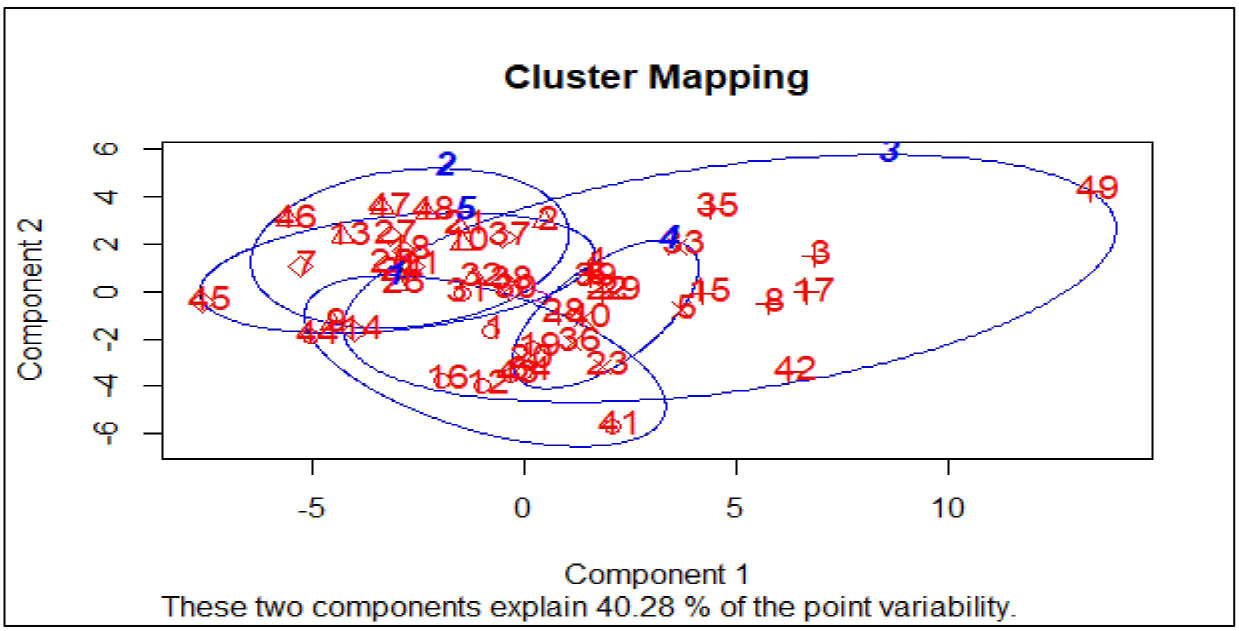

Figure 4.20: Large Sample Size of the Crime Variables Cluster Mapping Output for the Years

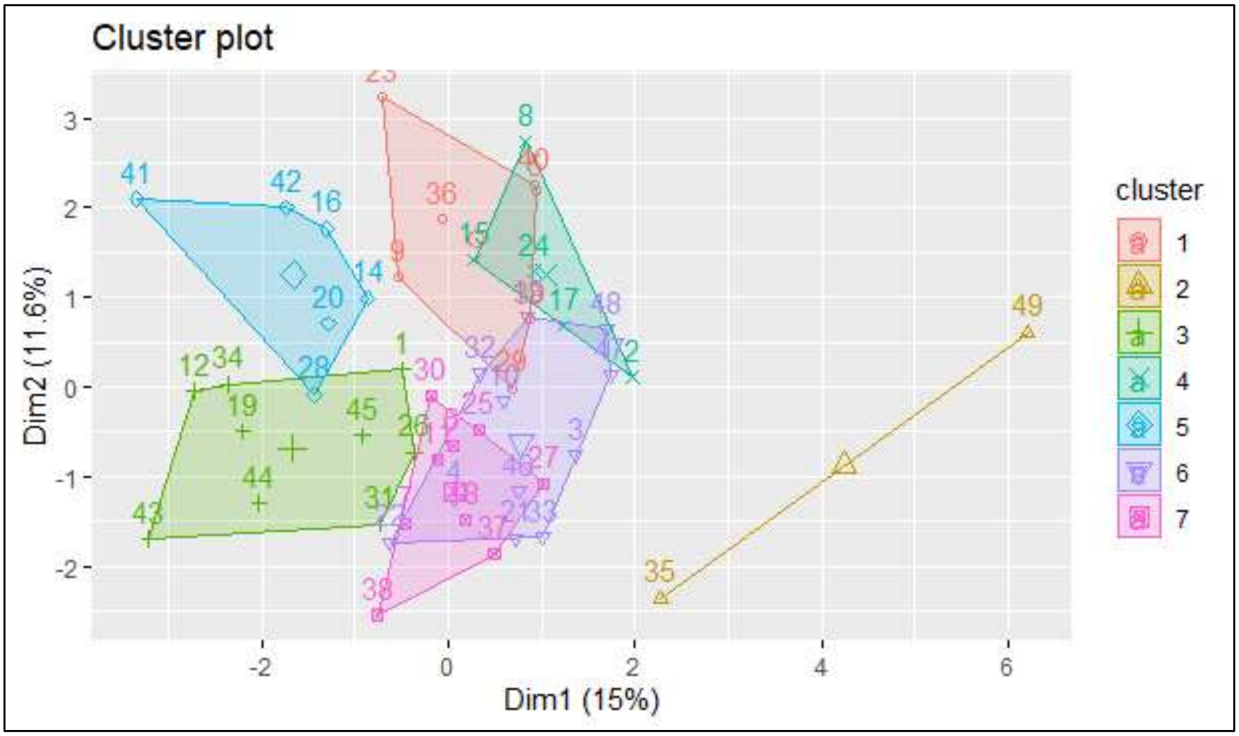

Figure 4.21: Large Sample Size of the Crime Variables Cluster Plot output for the Years

A principal component analysis is performed to represent the variables in a 5 dimensions' plane as shown in Figure 4.20. The two components from the principal component analysis explain about 40.28% of the point variability for the cluster model. The cluster plot in Figure 4.21 gives further clarification on the percentage contribution of each of these principal components, which is 11.6% and 15% respectively for the two components.

The simulated data for large sample size of the crime variables have 49 years, starting from 1972 to 2020. Hence, the yearly clustering for the large sample size of crime variables was done in Table 4.17.

Table 4.7: The Yearly Clustering for the Large Sample Size of Crime Variables

Cluster	cluster 1	cluster 2	cluster 3	cluster 4	cluster 5	cluster 6	cluster 7
Years	1976	2006	1972	1973	1985	1974	1977
	1980	2020	1983	1979	1987	1975	1978
	1994		1990	1986	1991	1981	1982
	2000		1997	1988	1999	1984	1989
	2007		2002	1995	2012	1992	1996
	2010		2005		2013	1993	1998
	2011		2014			2003	2001
			2015			2004	2008
			2016			2017	2009
						2018	

						2019	
Number of Year	**7**	**2**	**9**	**5**	**6**	**11**	**9**

The yearly clustering for the large sample size of the crime variables data chooses best number of clusters which is 7, is as follow in Table 4.17. Cluster 1 has seven years' crimes; Cluster 2 has only two years' crimes; while Clusters 3 and 7 have nine years of crimes each; Clusters 4 and 5 have five and six years of crimes respectively and Cluster 6 have the highest number of years with 11 crimes.

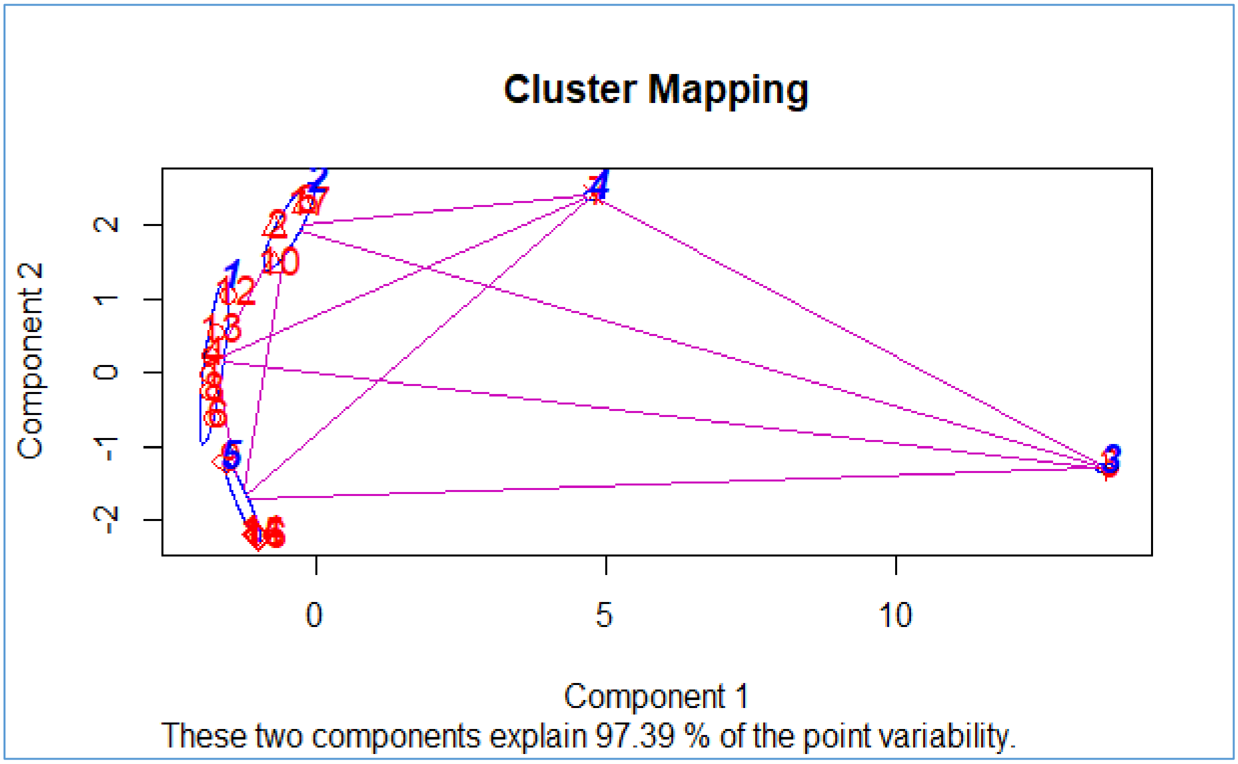

Figure 4.22: Large Sample Size of the Crime Variables Cluster Mapping for Crime Rates

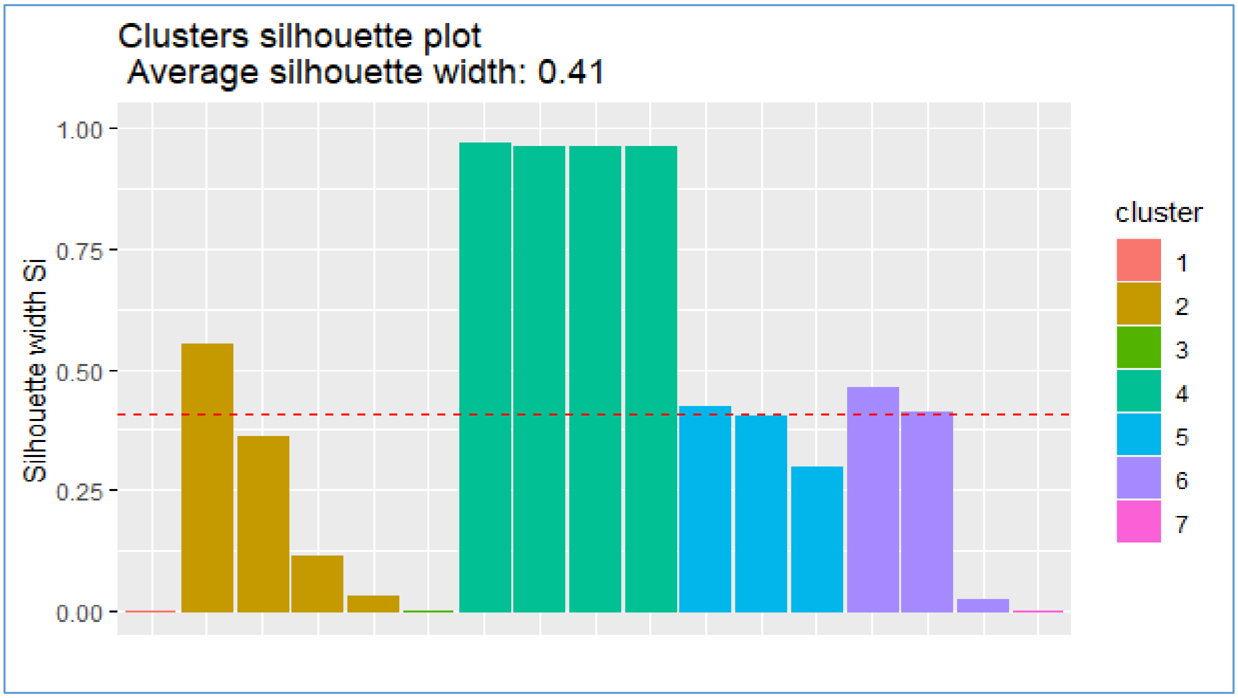

Figure 4.23: Silhouette Plot for the Crime Rates (Large Sample Size of the Crime Variables)

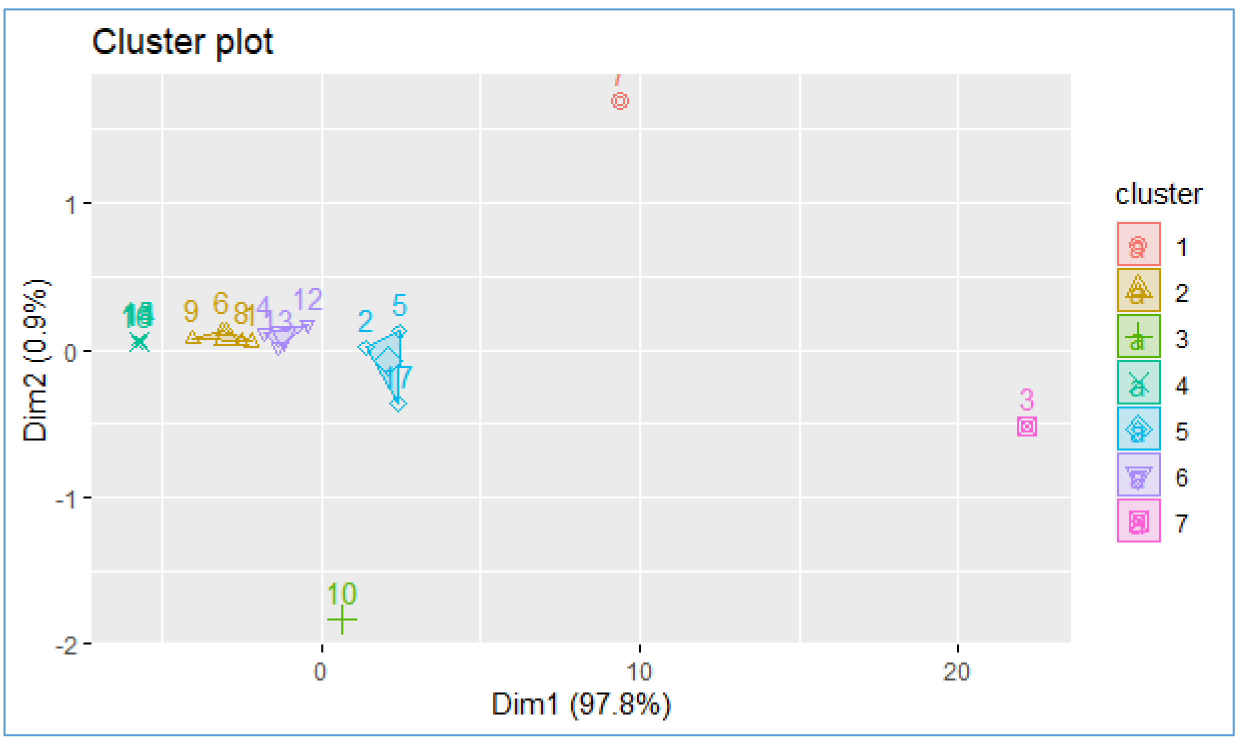

Figure 4.24: Large Sample Size of the Crime Variables Cluster Plot Output

Figure 4.22 is a principal component analysis performed where two components from the principal component analysis explain about 97.39% of the point variability for the cluster model. Figure 4.23 shows majority of the silhouette coefficients are positive and also suggests 7 as the best number of cluster. The cluster plot in Figure 4.24 gives further clarification on the percentage contribution of each of these principal components, which is 97.8% and 0.9% respectively for the two components. Next, the simulated data for large sample size of the crime rate classification was done below.

Table 4.8: The Clustering for the Large Sample Size of Crime Rate Variables

cluster	**cluster 1**	**cluster 2**	**cluster 3**	**cluster 4**	**cluster 5**	**cluster 6**	**cluster 7**
crime rate variables	Stealing	Homicide	Dangerous drugs	Serious Traffic offences	Offences against morality	Robbery	Other offences against persons
		Theft of stock		Corruption	Breakings	Criminal damage	
		Theft by servant		Offences involving police officers	Other penal code offences	Economic crimes	
		Vehicle and other thefts		Offences involving tourist			
Number of Variable	**1**	**4**	**1**	**4**	**3**	**3**	**1**

Table 4.18 shows the clustering for the large sample size of the crime rate variables where the chosen best number of clusters is 7. Then, Clusters 1, 3 and 7 have only one crime rate variable each; Clusters 5 and 6 have three crime rate variables each; while Clusters 2 and 4 have four crime rate variables each, which are the highest number of crime rate variables.

Small Sample size for Crimes in Nigeria

The factor analysis by the principal component analysis approach with three different rotations were done and presented in Table 4.23. The rotation with the highest cumulative variance explained by the four factors extracted is chosen as the best rotation to be used for the study.

Table 4.9: Comparison of Principal Component Factor Analysis using different Rotations (Small Sample size for Crime Rate variables)

Rotation Method	Variance explained criterion	Factor1	Factor2	Factor3	Factor4
	SS Loadings	1.855	1.702	1.059	1.047
None	Proportion Var	0.185	0.17	0.106	0.105
	Cumulative Var	**0.185**	**0.356**	**0.462**	**0.566**
	SS Loadings	1.761	1.453	1.293	1.156
Varimax	Proportion Var	0.176	0.145	0.129	0.116
	Cumulative Var	**0.176**	**0.321**	**0.451**	**0.566**
	SS Loadings	1.803	1.441	1.336	1.144
Promax	Proportion Var	0.18	0.144	0.134	0.114
	Cumulative Var	**0.18**	**0.324**	**0.458**	**0.572**

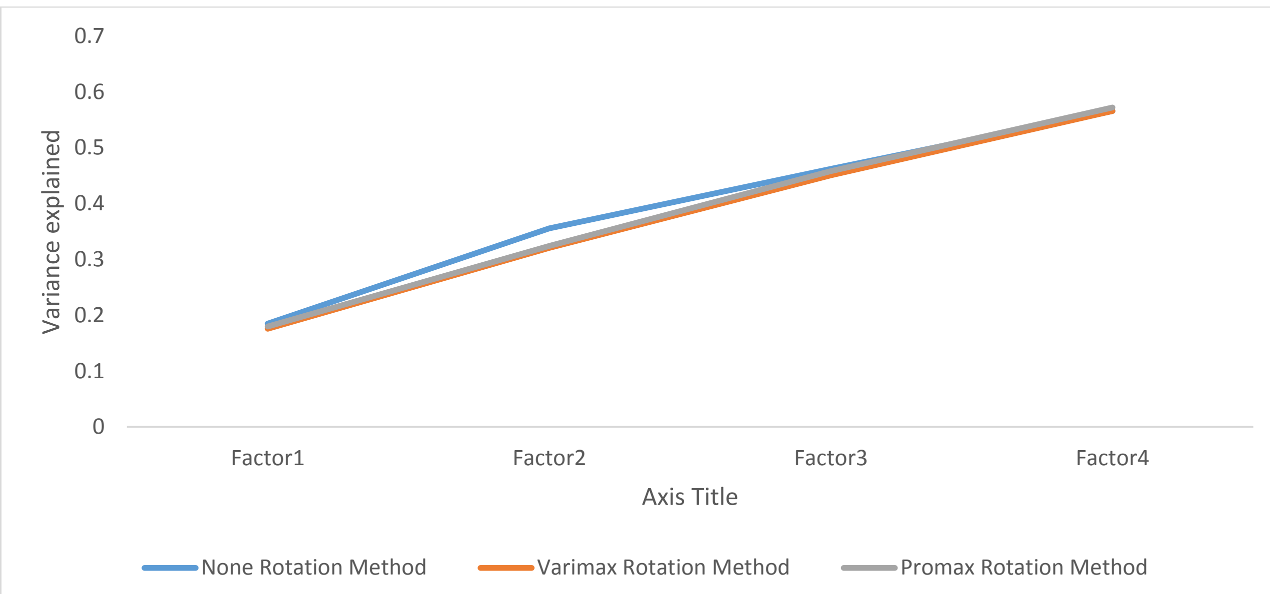

Figure 4.25: Small Sample Size for Crime Variables Scree Plot of the Cumulative Variance Explained by the Rotation Methods Principal Components

In Figure 4.25 and Table 4.23, the row Cumulative Var gives the cumulative proportion of variance explained by each factor. These numbers range from 0 to 1. The row Proportion Var gives the proportion of variance explained by each factor, and the row SS loadings gives the sum of squared loadings. This is sometimes used to determine the value of a particular factor. A factor is worth keeping if the SS loading is greater than 1 (Kaiser's rule). Based on the cumulative var for the three rotations, the Promax rotation has the highest cumulative proportion of variance in the ten variables explained by the four factors extracted and therefore is chosen as the best rotation for the factor analysis. The principal component analysis using the Promax rotation method output is given in Table 4.24.

Table 4.10: Factor Analysis using PCA for Promax Rotation (Small Sample Size for Crime Variables)

Call: factanal(x = mydata, factors = 4, rotation = "promax")

Uniquenesses:

X_1	X_3	X_5	X_7	X_9	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}
0.005	0.795	0.502	0.630	0.262	0.819	0.005	0.754	0.005	0.560

Loadings:

	Factor1	Factor2	Factor3	Factor4
X_1			1.027	
X_3	0.211	-0.270	0.194	
X_5	0.637	-0.107	-0.153	0.252
X_7	0.242	0.487	-0.227	-0.173
X_9	-0.824	-0.204	-0.205	
X_{12}	-0.296	0.214		0.161
X_{13}	0.156	0.100		0.984
X_{14}	-0.314		-0.271	0.197
X_{15}		1.002		0.108
X_{16}	0.636	-0.121	-0.207	

Factor Correlations:

	Factor1	Factor2	Factor3	Factor4
Factor1	1.0000	-0.3032	-0.13622	-0.09366
Factor2	-0.3032	1.0000	0.24596	0.09333
Factor3	-0.1362	0.2460	1.00000	0.00638
Factor4	-0.0937	0.0933	0.00638	1.00000

Test of the hypothesis that 4 factors are sufficient.
The chi square statistic is 9.35 on 11 degrees of freedom.
The p-value is 0.589

The first chunk in Table 4.24 provides the uniqueness, which range from 0 to 1. The uniqueness corresponds to the proportion of variability, which cannot be explained by a linear combination of the factors. A high uniqueness for a variable indicates that the factors do not account well for its variance. The next section is the loadings, which range from −1 to 1. The loadings are the contribution of each original variable to the factors. Variables with a high loading are well explained by the factor. Notice there is no entry for certain variables. That is because R-software does not print loadings less than 0.1. This is meant to help us spot groups of variables. The chi – squared test of hypothesis that test the hypothesis that 4 factors are sufficient for this model is not significant at 5% level of significance (p-value 0.589 < alpha 0.05). This means that the optimal number of factors is not significantly different from four). The proportion of the variability is denoted as communality. An appropriate factor model results in low values for uniqueness and high values for communality. The communality and uniqueness are displayed in Table 4.25.

Table 4.11: Communality and Uniqueness of the factor analysis of the Small Sample Size for Crime Rate Variables

Variable	Full Name	Communality	Uniqueness
X_1	Homicide	0.995004	0.004996
X_3	Other offences against persons	0.205171	0.794829
X_5	Breakings	0.497876	0.502124
X_7	Stealing	0.369719	0.630281
X_9	Vehicle and other thefts	0.737806	0.262194
X_{12}	Criminal damage	0.181195	0.818805
X_{13}	Economic crimes	0.99500	0.00500
X_{14}	Corruption	0.245591	0.754409
X_{15}	Offences involving police officers	0.995003	0.004997
X_{16}	Offences involving tourist	0.440098	0.559902

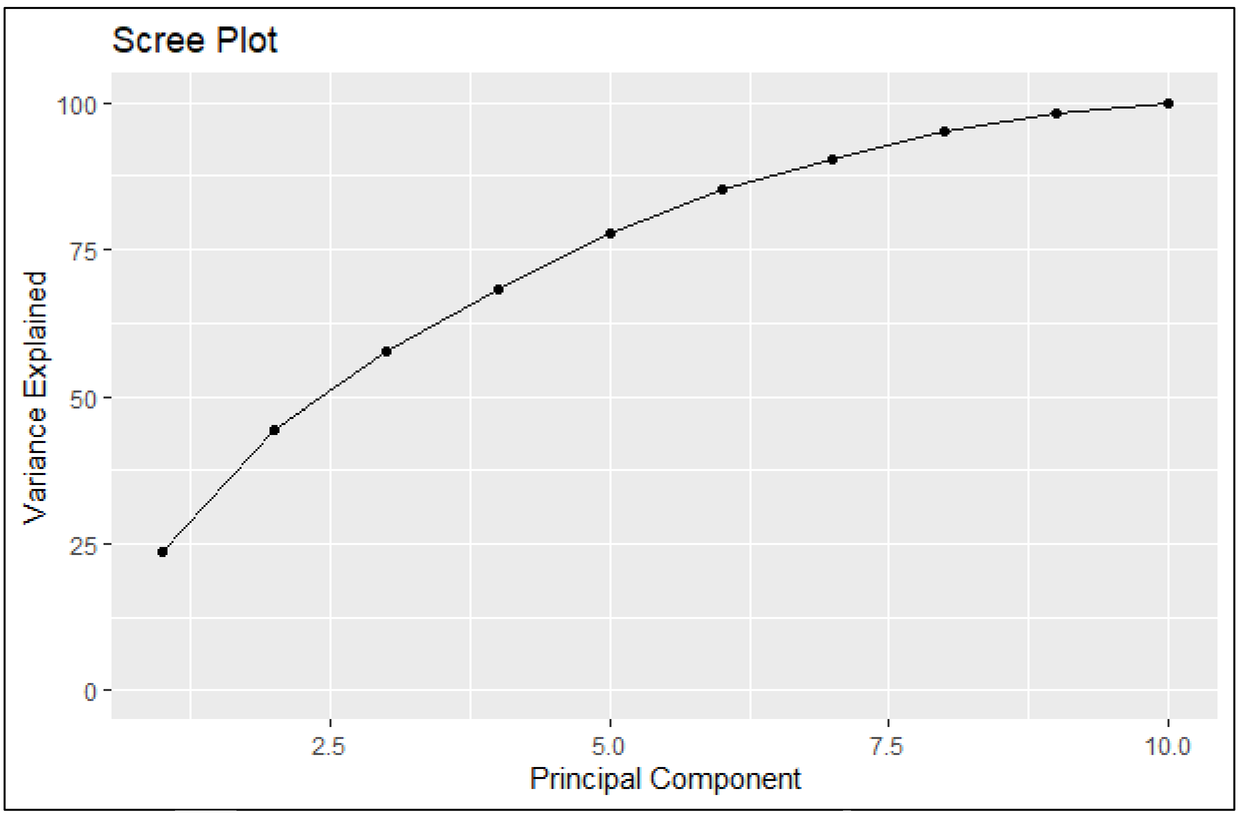

Figure 4.26: The Small Sample Size for Crime Rate Variables Scree plot of the Cumulative Variance Explained by the Principal Components

The scree plot in Figure 4.26 shows that four factors (principal components) explain about 68.19% of the common variance in the ten variables. The residual matrix for the factor analysis model is computed and displayed in Table 4.26.

Table 4.12: The Small Sample Size for Crime Rate Variables Residual Matrix

	X_1	X_3	X_5	X_7	X_9	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}
X_1	-4E-06	0.000592	-0.00016	-0.00019	-0.00023	0.000259	0.000001	-0.00037	0	-0.00063
X_3	0.000592	-2E-06	0.096114	0.049507	0.091502	0.169056	-0.00028	-0.13916	0.000167	0.141701
X_5	-0.00016	0.096114	0	0.102194	-0.00567	0.043415	0.000239	-0.04193	-0.00042	-0.07969
X_7	-0.00019	0.049507	0.102194	0.000001	0.040401	-0.07244	0	-0.14031	0.000326	-0.09064
X_9	-0.00023	0.091502	-0.00567	0.040401	0	0.04362	0.000152	-0.06185	-0.00013	-0.04211
X_{12}	0.000259	0.169056	0.043415	-0.07244	0.04362	-1E-06	-0.00053	0.114534	0.000964	0.110627
X_{13}	0.000001	-0.00028	0.000239	0	0.000152	-0.00053	0	-0.00016	0.000001	-6E-06
X_{14}	-0.00037	-0.13916	-0.04193	-0.14031	-0.06185	0.114534	-0.00016	0	-0.0001	-0.00461
X_{15}	0	0.000167	-0.00042	0.000326	-0.00013	0.000964	0.000001	-0.0001	-3E-06	0.000229
X_{16}	-0.00063	0.141701	-0.07969	-0.09064	-0.04211	0.110627	-6E-06	-0.00461	0.000229	-1E-06

The matrix in Table 4.26 is called the residual matrix. Numbers close to 0 indicate that our factor model is a good representation of the underlying concept. It is evident that the factor model with Promax Rotation is a good representation of the underlying concept.

Cluster Analysis of the Small Sample Size for Crime Rate Variables

Likewise, in order to find the optimal number of clusters for a *k*-means, the following four approaches are considered in this work: The above methods are displayed in Figures 4.27, 4.28 and 4.29.

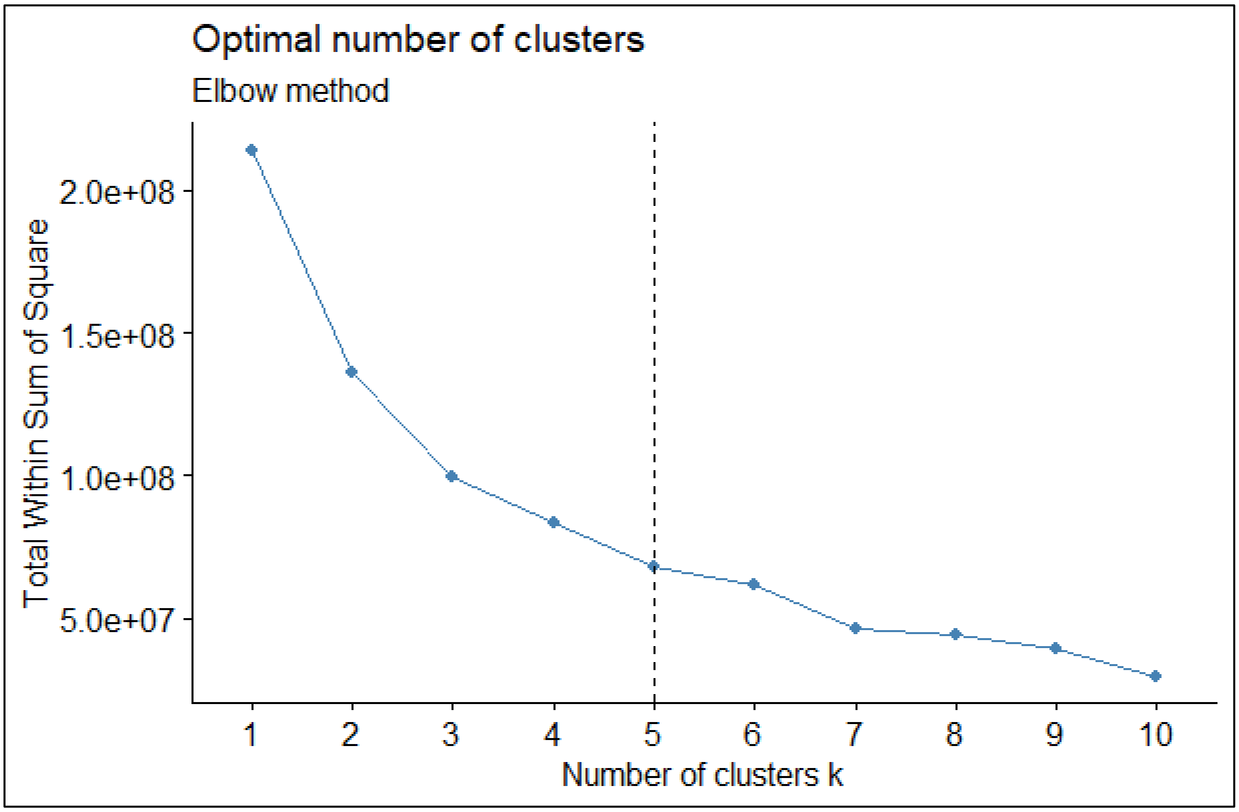

Figure 4.27: Elbow Method for Number of Clusters Identification for the Small Sample Size of the Crime Rate Variables

The location of a knee in the plot is usually considered as an indicator of the appropriate number of clusters because it means that adding another cluster does not improve much better the partition. This method seems to suggest 5 clusters in Figure 4.27.

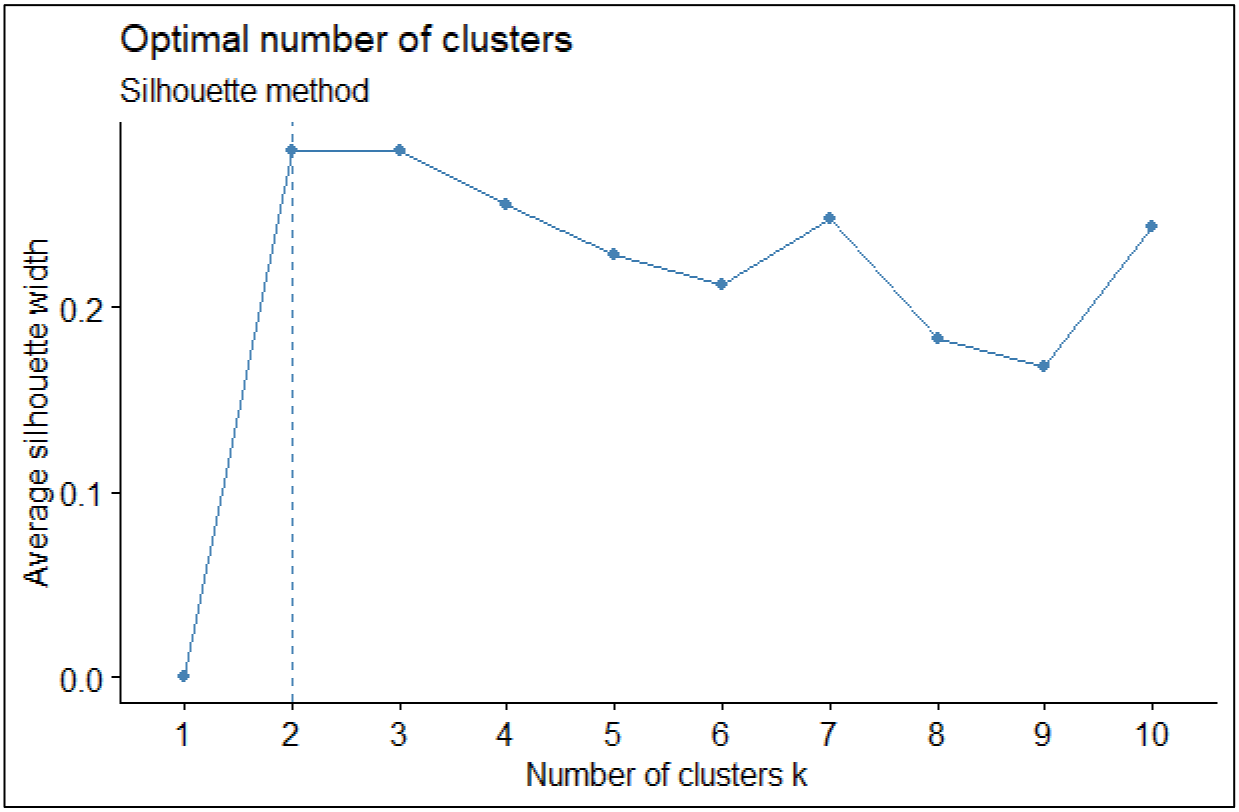

Figure 4.28: Silhouette Method for Number of Clusters Identification for the Small Sample Size of the Crime Rate Variables

The Silhouette method measures the quality of a clustering and determines how well each point lies within its cluster. The Silhouette method suggests 2 clusters in Figure 4.28.

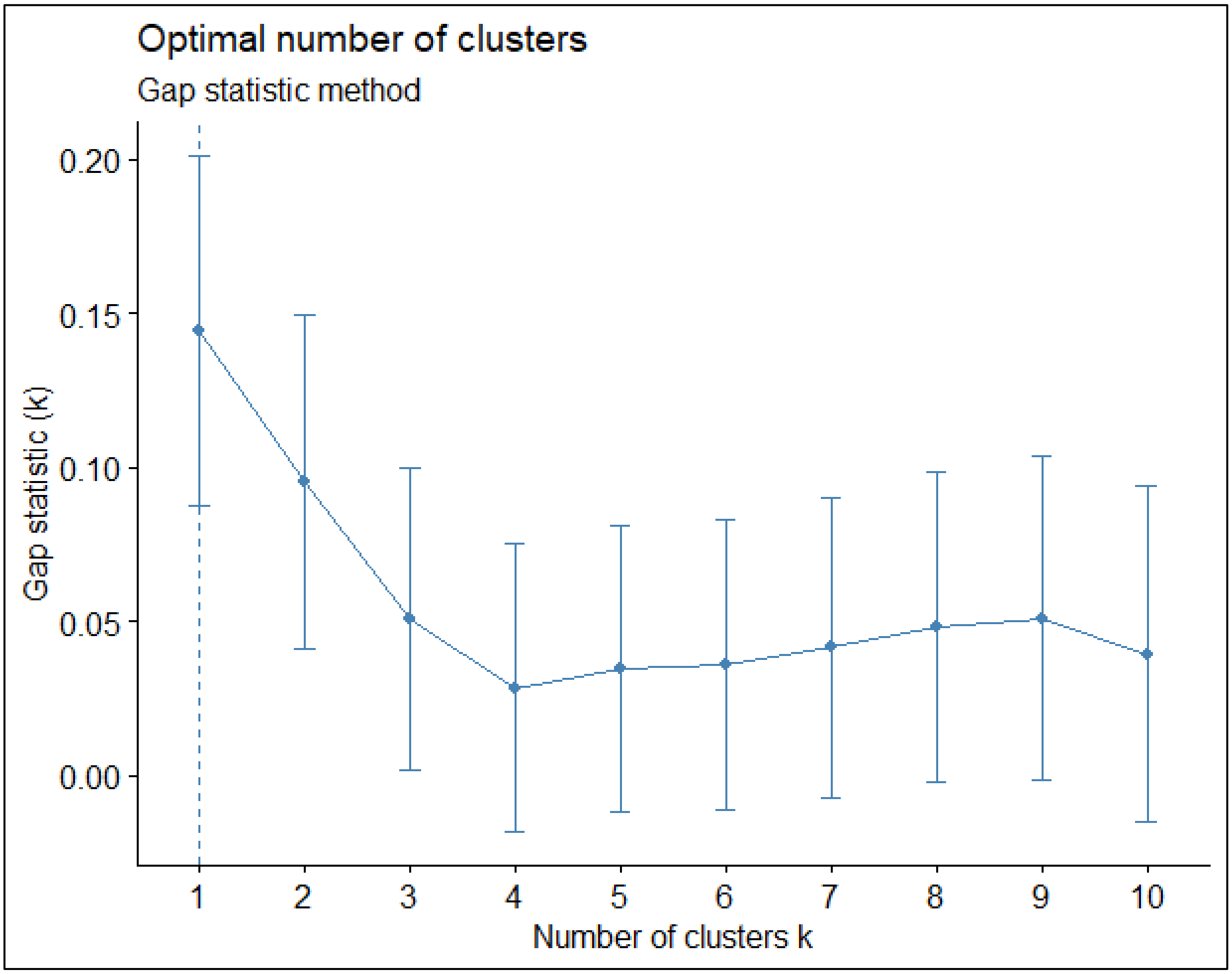

Figure 4.29: Gap Statistic Method for Number of Clusters Identification for the Small Sample Size of the Crime Rate Variables

The optimal number of clusters is the one that maximizes the gap statistic. After 500 bootstraps with a starting point of 25, this method suggests only 1 cluster (which is therefore a useless clustering). The three methods done so far are suggesting that the optimal number of clusters is between 1 and 5. So, the D index method constructed using the Euclidean distance and permuted optimal number of factors around the interval 1 and 5 for 1 as the minimum and 5 as the maximum number of cluster gave the result displayed in Figure 4.30 and the corresponding optimal number in Figure 4.30.

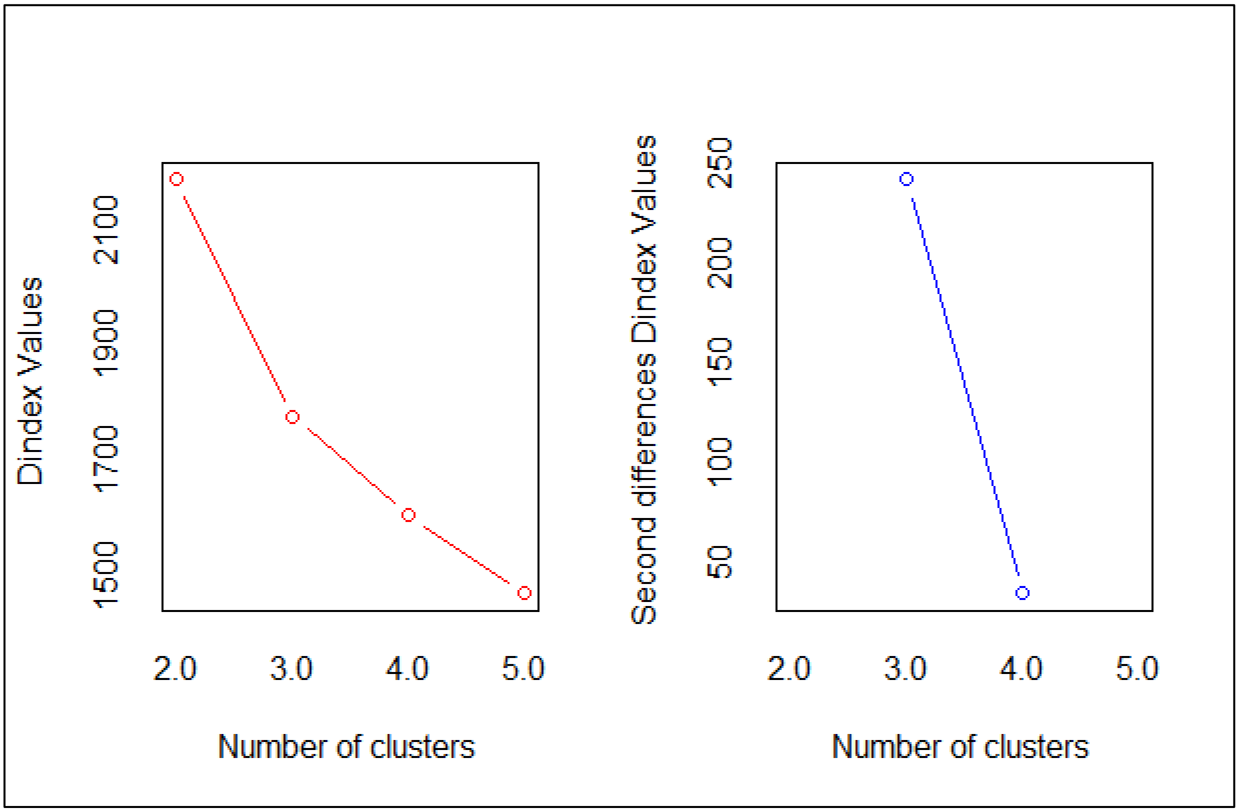

Figure 4.30: D Index Statistic Method for Number of Clusters Identification I for the Small Sample Size of the Crime Rate Variables

The D index is a graphical method of determining the number of clusters. In the plot of D index in Figure 4.30, we seek a significant knee (the significant peak in D index second differences plot) that corresponds to a significant increase of the value of the measure.

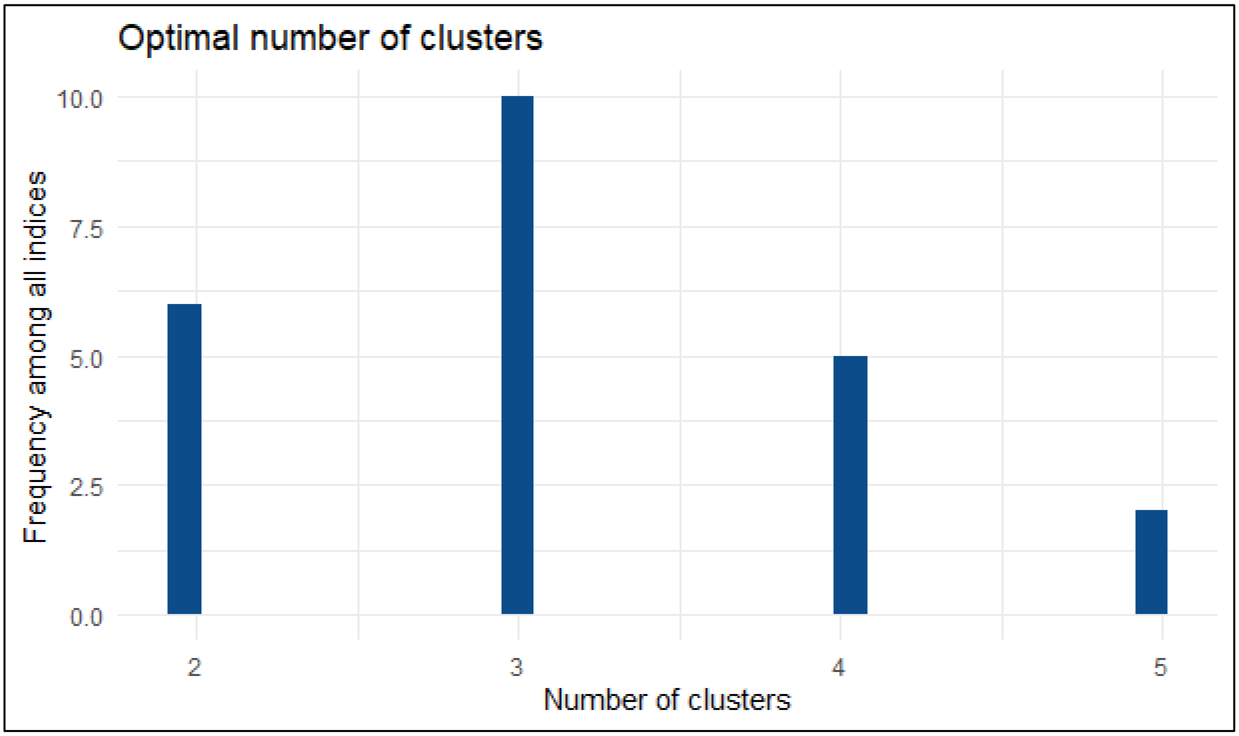

Figure 4.31: D Index Statistic Method for Number of Cluster Identification II for the Small Sample Size of the Crime Rate Variables

Among all indices in Figure 4.31; 6 proposed 2 as the best number of clusters, 10 proposed 3 as the best number of clusters, 5 proposed 4 as the best number of clusters and 2 proposed 5 as the best number of clusters. According to the majority rule, the best number of clusters is 3.
Finally, the four methods have given an idea of the number of clusters that is optimal in grouping the data. We therefore choose the highest number of cluster and this number is 5. To confirm that this chosen number of classes is indeed optimal, there is a way to evaluate the quality of your clustering via the silhouette plot (which shows the silhouette coefficient on the *y* axis). We draw the silhouette plot for 5 clusters, as chosen. Figure 4.32 shows the output of the plot.

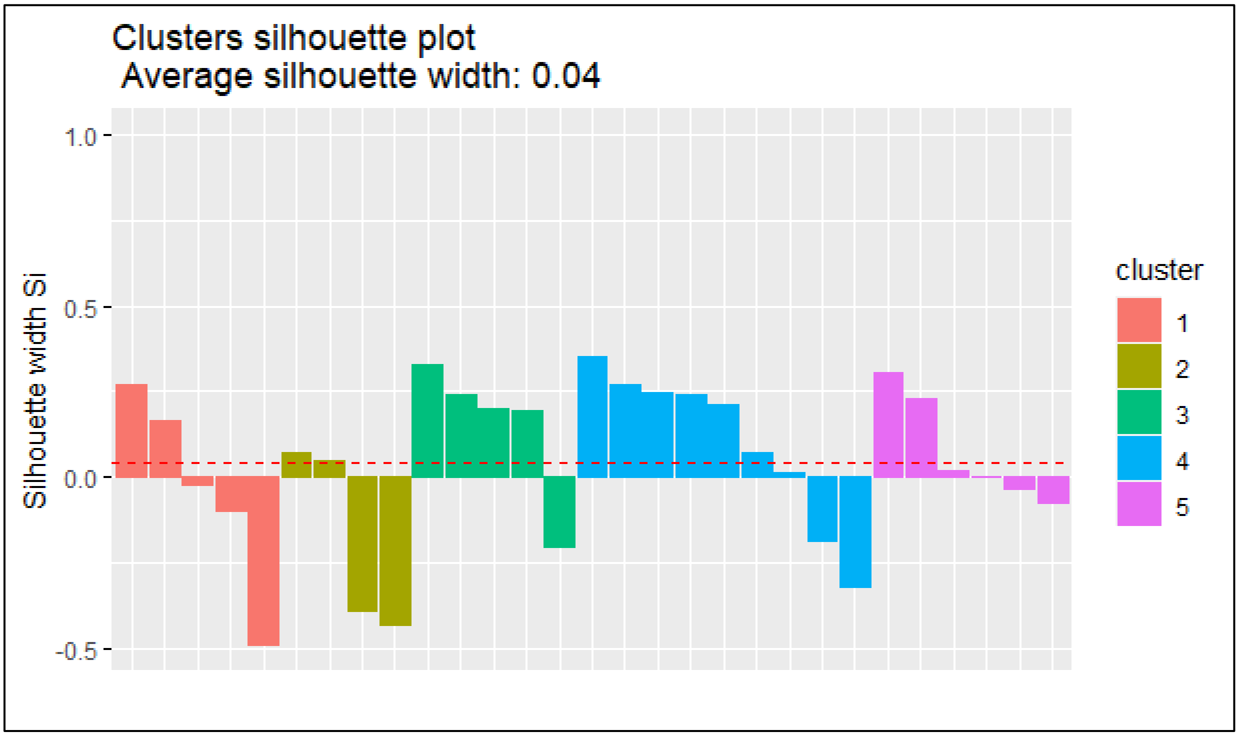

Figure 4.32: Silhouette Plot for the Small Sample Size of the Crime Rate Variables

As a reminder, the interpretation of the silhouette coefficient is as follows:

i. Greater than zero means that the observation is well grouped. The closer the coefficient is to 1, the better the observation is grouped.

ii. Less than zero means that the observation has been placed in the wrong cluster.

iii. Equal to zero means that the observation is between two clusters.

Since a large majority of the coefficients are positive, it indicates that the choice of five as the optimal number of clusters is okay. Having determined the optimal number of clusters, the next step is to normalize the data set. The data set is normalized and k-means cluster analysis was conducted. The cluster mapping is displayed in Figure 4.33

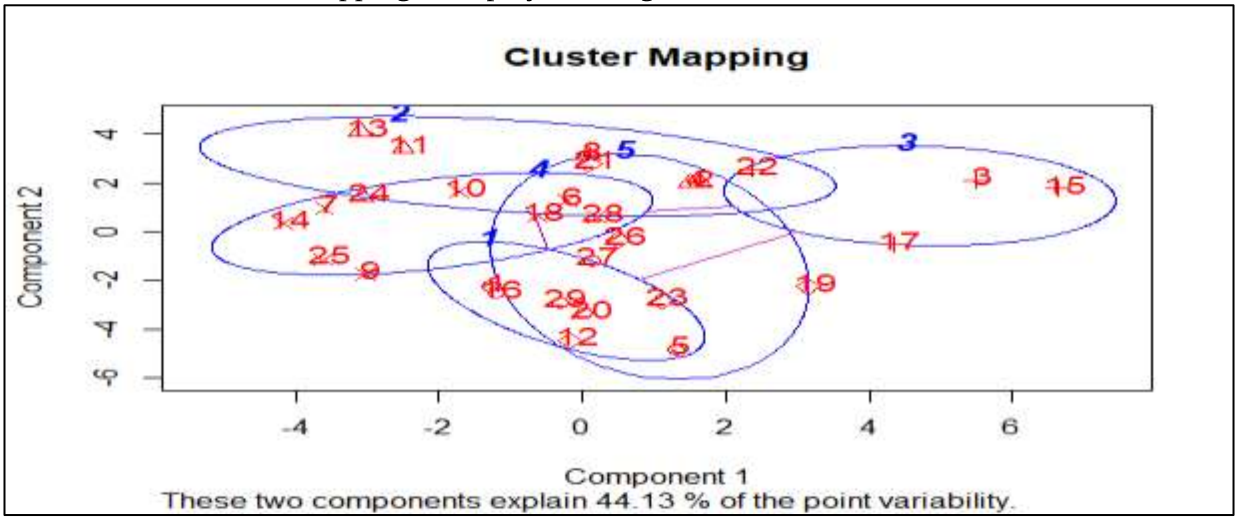

Figure 4.33: Cluster Mapping Output for the Small Sample Size of the Crime Rate Variables

A principal component analysis is performed to represent the variables in a 2 dimensions' plane as shown in Figure 4.33. The two components from the principal component analysis explain about 44.13% of the point variability for the cluster model. The cluster plot in Figure 4.34 gives further clarification on the percentage contribution of each of these principal components.

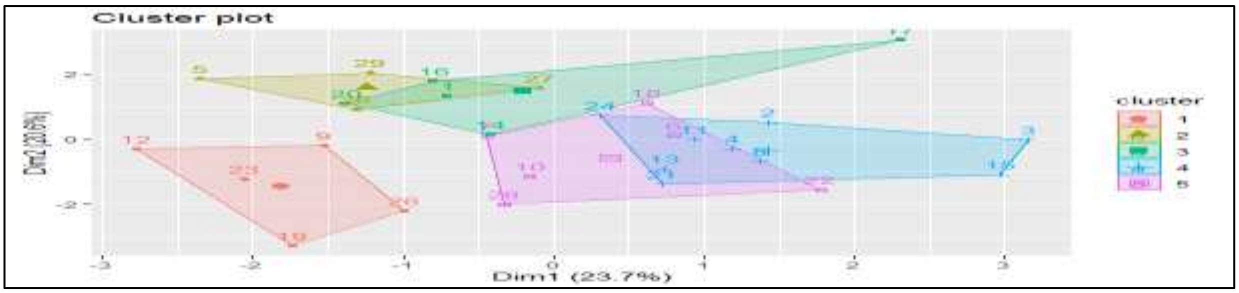

Figure 4.34: Cluster Plot Output for the Small Sample Size of the Crime Rate Variables

Table 4.13: The Yearly Clustering for the Small Sample Size of Crime Rate Variables

S/N	X_1	X_3	X_5	X_7	X_9	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}	Cluster
1	2498	16586	5616	12359	1232	4071	2832	111	86	58	**3**
2	2641	23174	6244	9054	1192	4044	4347	118	59	67	**4**
3	2562	20938	6601	8954	896	2710	3422	88	34	65	**4**
4	2749	20834	7456	9306	1207	3418	3893	109	36	6	**4**
5	2482	18542	6291	9758	1510	4526	4710	155	95	20	**2**
6	2698	22948	5898	11856	1054	4343	2352	113	97	46	**5**
7	2653	21579	5820	12272	1264	3755	3064	107	97	7	**5**
8	2696	21690	6472	11315	1063	3404	2950	136	2	34	**4**
9	2526	17900	6275	9347	1358	4137	3040	130	49	9	**1**
10	2799	20332	6136	10642	1432	4291	3296	70	46	10	**5**
11	2644	23246	6402	11362	1411	3597	3432	107	64	46	**4**
12	2529	16193	5101	10021	1319	3815	2968	150	63	3	**1**
13	2737	22468	6141	11149	1367	3776	3319	103	29	34	**4**
14	2591	18015	6198	11594	1418	3722	2787	104	62	39	**3**
15	2651	23786	7493	9969	1132	4381	2632	92	-35	61	**4**
16	2342	18535	6474	12293	1371	3376	3269	143	79	38	**3**
17	2476	18991	7142	13321	980	3372	4132	103	71	74	**3**
18	2646	20534	6149	13994	1144	3568	3496	74	98	13	**5**
19	2745	19147	5128	7300	1614	3389	2497	133	13	15	**1**
20	2222	20178	6271	12308	1561	4376	3027	102	57	27	**3**
21	2818	21403	5643	8709	1344	4014	3924	89	28	47	**4**

22	2902	19663	6354	10941	1136	2840	2378	89	41	33	**5**
23	2531	21942	5575	10765	1675	4200	2882	167	8	25	**1**
24	2548	19327	6494	9930	1193	4033	3738	123	43	45	**4**
25	2593	21074	5859	11309	1412	4314	3368	128	124	27	**2**
26	2793	19960	4807	8271	1271	4296	2999	125	23	29	**1**
27	2781	19066	6594	10650	1087	4802	3442	118	134	38	**2**
28	2717	20324	5194	11586	1484	3499	2614	61	39	13	**5**
29	2722	17470	6316	13397	1249	4156	4250	152	84	20	**2**

Table 4.27 shows the yearly clustering for the small sample size of crime rate variables suggests 5 as the best number of clusters with cluster 4 consisting of the highest number of years.

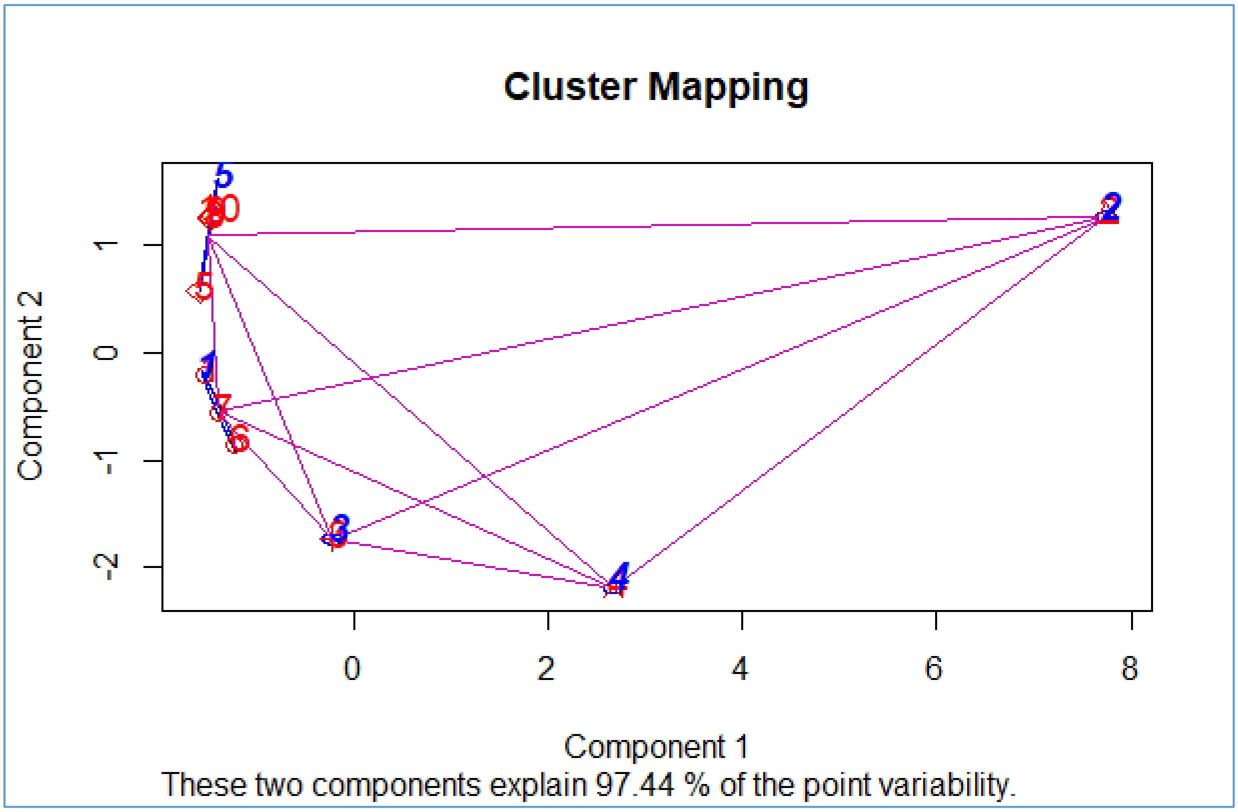

Figure 4.35: Cluster Mapping for the Small Sample Size of the Crime Rate Variables

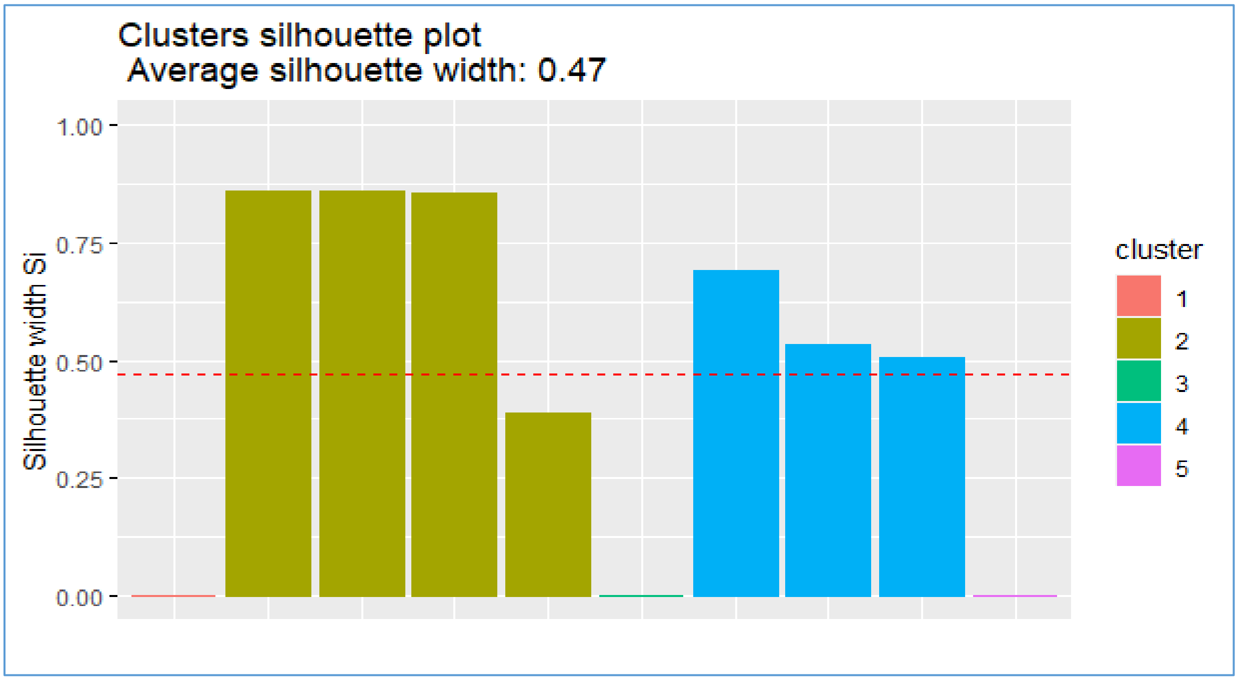

Figure 4.36: Silhouette Plot for the Small Sample Size of the Crime Rate Variables

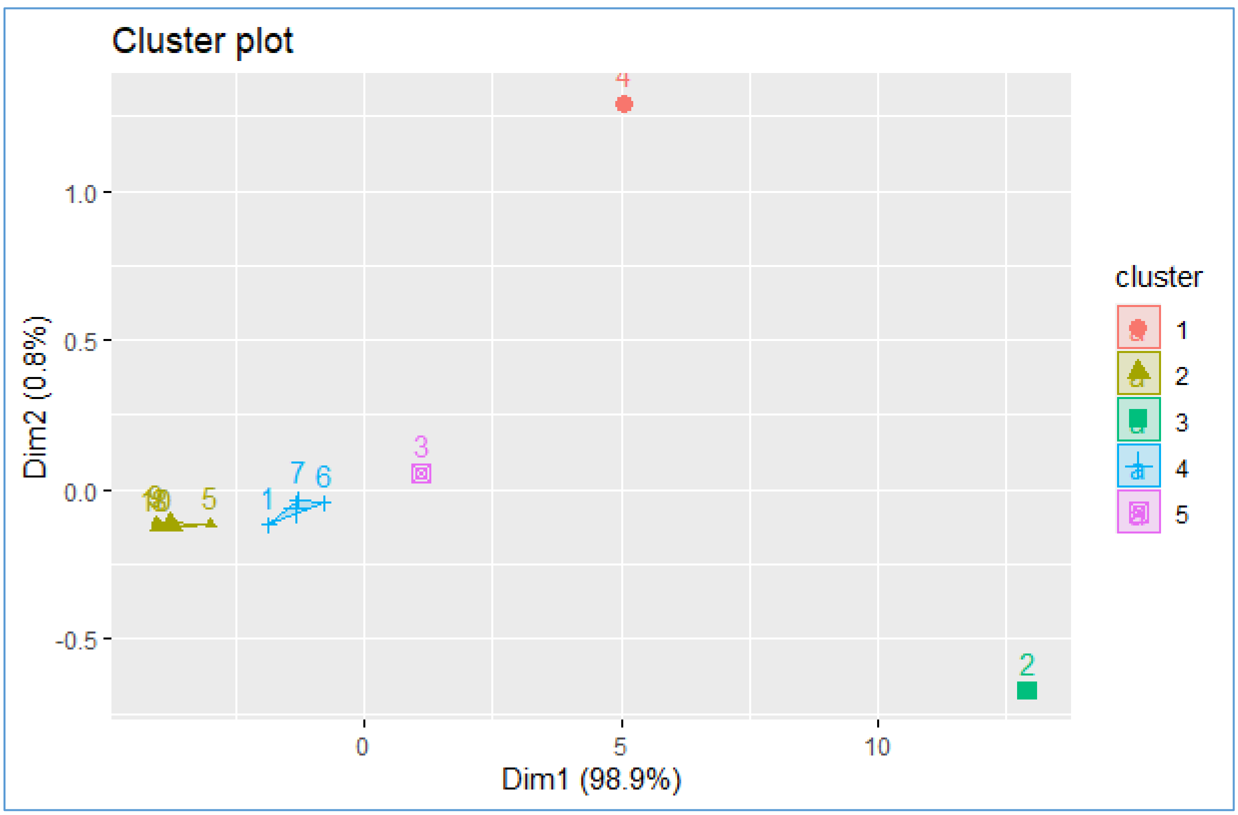

Figure 4.37: Cluster plot Output for the Small Sample Size of the Crime Rate Variables

The two components from the principal component analysis explain about 97.44% of the point variability for the cluster model in Figure 4.35. The cluster plots in Figure 4.36 and 4.37 give further clarification on the percentage contribution of each of these principal components.

Table 4.28a: The Clustering for the Small Sample Size of Crime Rate Variables

Crimes	**Cluster**
Stealing	1
Vehicle and other thefts	2
Corruption	2
Offences involving police officers	2
Offences involving tourist	2
Other offences against persons	3
Homicide	4
Criminal damage	4
Economic crimes	4
Breakings	5

Table 4.28b: The Clustering for the small Sample Size of Crime Rate Variables

Cluster	**cluster 1**	**cluster 2**	**cluster 3**	**cluster 4**	**cluster 5**
Crime Rate Variables	Stealing	Vehicle and other thefts	Other offences against persons	Homicide	Breakings
		Corruption		Criminal damage	
		Offences involving police officers		Economic crimes	
		Offences involving tourist			
Number of Variable	**1**	**4**	**1**	**3**	**1**

Table 4.28b shows the clustering for the small sample size of the crime rate variables where the chosen best number of cluster is 5. Then, Clusters 1, 3 and 5 have only one crime rate variable; Cluster 4 has three crime rate variables while Clusters 2 has four crime rate variables, which is the highest number of crime rate variables.

CONCLUSION

The three different rotation methods (None, Varimax and Promax rotation methods) result indicated that None and Varimax rotation methods exhibit similar variation, while the Promax rotation method differ slightly in terms of the economic variables data set, while the three different rotation methods results for large and small sample sizes (crime data) showed similar variation among the variables and identified some factors are sufficient for describing the total variance accounted in the crime variables.

The communality and uniqueness of the factor analysis for the economic variables shows that principal components lie between 78.2% and 99.5% respectfully, of the total variability. Thus, most of the data structure can be captured in the first-thirteen principal component underlying dimensions. The remaining principal components account for a very small proportion of the variability and are probably unimportant. The communality and uniqueness of the factor analysis for the large sample size for crime variables principal components lies between 17.7% and 99.5% respectfully, of the total variability. Thus, most of the data structure can be captured in the first-seven principal component underlying dimensions from Figure 4.13, since it accounts for almost 75% proportion of the variability. The remaining principal components account for 25% proportion of the variability and are probably unimportant. While the communality and uniqueness of the factor analysis for the small sample size for crime variables principal components lies between 18.1% and 99.3% respectfully, of the total variability.

In order to find the optimal number of clusters in both the economic and crime figures data, we use four cluster analysis approaches; Elbow method, Average silhouette method, Gap statistic method and NbClust() function (package) to implement the D index method. These results show seven cluster groups for the economic variables, also seven cluster groups for the large sample size of crime rate variables and five cluster homogenous groups for the small sample size of crime rate variables.

CONTRIBUTION

i. This research was able to identify significant differences between (None and Varimax rotation methods) and the Promax rotation method of the Nigeria Economic variables data considered, while there are no significant differences among the three different rotation methods results for crime data (i.e. large and small sample sizes).
ii. Based on the results of the study, it was identified that carrying out the cluster analysis (classification) process, it requires consideration of factor analysis that would show the effect/expression of the data structure.
iii. This study shows that variables must be controlled when there are many covariates in the data set.
iv. It was able to demonstrate the role of covariates in each data sets [Nigeria Economic variables: crop production, livestock and Forestry; Nigeria Crime Figures data: Homicide, Offences against morality, Other offences against persons, Robbery, Breakings, Theft of stock and Stealing].
v. This study was able to determine seven cluster groups for the economic variables, also seven cluster groups for the large sample size of crime rate variables and five cluster homogenous groups for the small sample size of crime rate variables.

REFERENCES

Abdi, H. & Williams, L. J. (2010). Principal Components Analysis. *Wiley Interdisciplinary Journal of Reviews in Computational Statistics, 2*(4), 433–459.

Abhishek, B. (2012). Impact of principal components analysis in the application of image processing. *International Journal of Advanced Research in Computer Science and Software Engineering,* 2(1).

Aebi, M.F., Aubusson de Cavarlay, B., Barclay, G., Gruszezynska, B., Harrendorf, B. & Heiskanen, M. (2010). European source book of crime and criminal justice statistics (4th edition). Den Haag: Boom Juridische Uitgevers. Available online at: *http://english.wode.nl/onderzocksdatabase/europeansourcebook-4c-editie.aspx?cp=45&sc=6796.*

Andresen, M. A. (2006). A spatial analysis of crime in Vancouver, British Columbia: A synthesis of social disorganization and routine activity theory. *Can. Geogr*., 50(4), 487–502.

Arash A., (2005). Color image processing using principal component analysis. *Unpublished Master's Thesis*, Mathematics Science Department, Sharif University of Technology, Tehran, Iran.

Arash, A. & Shohreh, K. (2008). Colour PCA Eigen Images and their Application to Compression and watermarking, 26(7), 0260-8858

Baumer, E.P. & Lauritsen, J.L. (2010). Reporting crime to the police: A multivariate analysis of long-term trends in the national crime survey (NCS) and national crime victimization survey (NCVS). *Criminology*, 48, 131-186.

Bocuvxa L., Vacek O. & Jehlicka J. (2005). Principal Component Analysis as a Tool to Indicative the Origin of Potentially Toxic Elements in Soils. *Geoderma*, 128, 289-300.

Bosick, S., Rennison, C., Gover, A. & Dodge, M. (2012). Reporting violence to the police: Predictos through the life course. *Journal of Crime Justice,* 441-451. https://doi.org/10.1016/j.jcrimjus.2012.05.001.

Brook, J. S., Whiteman, M. & Nomura, C. (2008). Personality, family, and ecological influences on adolescent drug use: A developmental analysis. *Journal of Chemical Dependency Treatment*, 1, 123-161.

Carreira-Perpinan, M. A. (2001). Continuous latent variable models for dimensionality reduction and sequential data reconstruction, *Unpublished PhD Thesis*.

Chang, Y., Cesarman, E., Pessin, M. S., Lee, F., Culpepper, J., Knowles, D. M. & Moore, P. S. (2013). Identification of herpes virus-like DNA sequences in AIDS-associated Kaposi's sarcoma. *Science;* 266, 1865–1869.

Cheng, S.C., & Hsia, S.C. (2003). Fast Algorithm's for Color Image Processing by Principal Component Analysis. *Journal of Visual Communication and Image Representation*, 14, 184-203.

Christens, B. & Speer, P. W. (2005). Predicting violent crime using urban and sub-urban densities. *Behaviour and Social Issues*, 14, 113 –127.

Dragovic, S. & Onjia, A. (2006). Classification of Soil Samples According to their Geographic Origin Using Gamma-ray Spectrometry and Principal Component Analysis. *Journal of Environmental Radioactivity*, 84, 150-158.

Duda, R. O., Hart, P. E. & Stork, D. G. (2001*). Pattern classification* (Second Edition). John Wiley and Sons.

Ekpo, A. H. & Umoh, O.J. (2012). *Overview of the Nigerian Economic Growth and Development*. Eghosa Osagie (1992) Edited. Structural Adjustment Programme in the Nigeria Economy, National Institute for Policy and Strategic Studies, Kuru, Jos Nigeria. 71-104.

Eric, B., Trevor, H., Debashis, P. & Robert, T. (2006). Prediction by Supervised Principal Components. *Journal of the American Statistical Association*, 101 (473), 119–137.

Eze-Emmanuel, P. & Etuk, E. H. (2018). Principal Component Analysis of Nigerian Economic Variables. *International Journal of Science and Advanced Innovative Research*, 3(2), 1-28.

Fukunaga, K. (1990). *Introduction to Statistical Pattern Recognition*, (Second Edition), Academic Press,

Gerns, S. H. L., Richardson, B. A., Singa, B., Naulikha, J., Prindle, V. C., Diaz-Ochoa, V. E., Felgner, P. L., Camerini, D., Horton, H., John-Stewart G. & Judd, L. W. (2014). Use of principal components analysis and protein microarray to explore the association of HIV-1-Specific IgG responses with disease progression, AIDS. *Research and Human Retroviruses*, 30(1), 39-44.

Golub, G. H. & Van Loan, C. F. (2014). *Matrix Computations*. Baltimore, Maryland: Johns Hopkins University Press.

Harman, H. H. (2013). Applied Factor Analysis. *Journal of Education Psychology*, 312-320.

Hart, T.C. & Rennison, C.M. (2003). *Reporting crime to the police*. U.S Government Printing Press, Washington DC.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24, 417–441, and 498–520.

Hough, M. & Sato, M. (2011). Trust in justice: Why it is important for criminal policy and how it can be measured. *Final report of the Euro-Justis Project*. Helsinki: Heuni-Unodc.

Huey, L. & Quirouette, M. (2010). Access to justice as a component of citizenship: Reconsidering policing services for Canada's homeless. Available on *http://ir.lib.nwo.ca/sociologypub/10.* Retrieved on 2/17/2021.

Jaisheel, M., Fulufhelo, V. N. & Tshilidzi, M. (2009). Missing data estimation using principle components analysis and auto associative neural networks. *Systemic, Cybernetics and Informatics*, 7(3), 1690-4534.

Johnson, R. A. & Wichern, D. W. (1998). *Applied multivariate statistical analysis*, (Fourth Edition). Prentice- Hall.

Jolliffe, I. T. (2002). *Principal Components Analysis*, (2nd edition) Springer-Verlag. ISBN 978-0-387-95442-4.

Kääriäninen, J. & Siren, R. (2011). Trust in the police, generalized trust and reporting crime. *European Journal of Criminology,* 8, 65-81. http://doi.org/10.1177/14773708103 576562.

Kaiser, H.F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20, 141-151. http://doi.org/10.1177/0013164460020 0116.

Kambhatla, N. & Leen, T. (2014). Dimensionality Reduction by Local PCA, Neural Computation, 9, 1493.

Kamolchanok, P., Kanokporn, S., Natdhera, S. & Daoroong, S. (2012). Principal Components Analysis for the characterization in the application of some soil properties; *International journal of Environment and Ecological Engineering*. 6(5).

Kepple, N. J. & Freisthler, B. (2012). Exploring the ecological association between crime and medical marijuana dispensaries, *J. Stud. Alcohol Drugs*, 73 (4), 523–530.

Khaled, L., Rao, V. & Vemuri, I. (2004). An Application of Principal Components Analysis to the DETECTION and Visualization of Computer Network Attack. *Annals of Telecommunication*, 61(1), 218-234.

Kilishi, A. A., Mobolaji, H. I. Usman, A. Yakibu, A. T. & Yaru, M. A. (2014). The effect of unemployment on crime in Nigeria: a panel data analysis. *British Journal Economic Manag*ement Trade, 4(6), 880–895.

Kim, J. & Chung, J. (2003). *Reduction of Dimension of HMM parameters using ICA and PCA in MLLR Framework for Speaker Adaptation*, Eurospeech, Geneva.

Kunnuji, M. O. N. (2016). Population density and armed robbery in Nigeria: an analysis of variation across states, *African Journal of Criminology*, Justice Studies, 9(1), 62–73.

Landis, J. R. & Koch, G. G. (2003). The Measurement of Observer Agreement for categorical data. *Biometrics*. 33, 159-174

Libin, Y. (2015). An application of Principal Components Analysis to Stock Portfolio Management. Department of Economics and Financial University of Canterbury.

Linde, Y., Buzo, A. & Gray, R. M. (2014). An Algorithm for Vector Quantizer Design. *IEEE Transactions on Communications*, 28(1).

Lindsay, I. S. (2002). A Tutorial on Principal Components Analysis. *Journal of Business and Management Science*, 2(1), 10-20.

Lldiko E. F. & Jerome H. F. (2012). A Statistical View of Some Chemometrics Regression Tools. *Technometrics*, 35(2), 109–135.

Maike, R. (2016). Factor Analysis: A Short Introduction. *http://www.theanalysisfactor.com/factor-analysis-1-introduction/*. Retrieved on 10/23/2016.

Malby, S. (2010). Data collection on [New] forms and manifestations of crime. In M. Joutsen (Ed.), *New types of crime*. Helsinki: Heuni-Unodc.

Marshal N., Faust M. & Hendler T., (2005) The Role of the Right Hemisphere in Processing Nonsalient Metaphorical Meaning: Application of Principal Components Analysis to FMRI data. *Neuropsychologia*, 43(14), 2084-2100.

Nigeria's National Bureau of Statistics (2016). Crime Statistics: Reported Offences by Type and States, *www.nigerianstat.gov.ng*.

Noko, E. J. (2016). Economic recession in Nigeria: Causes and Solution. *http://educacinfo.com/economic-recession- Nigeria*. Retrieved on 11/4/2020.

Odumosu, O. F. (2009). Social costs of poverty: the case of crime in Nigeria. *Journal of Social Development in Africa*, 14(2), 71–85.

Ogoke, U. P., Nduka, E. C., Biu, E. O. & Ibeachu, C. (2003). A Comparative Study of Foot Measurements Using Receiver Operating Characteristics (ROC) Approach. *Africana Journal of Pure and Applied Sciences*. 12(1), 76-88.

Omisakin, I.S. (1998). *Crime trends and prevention strategies in Nigeria*. *A study of Old Oyo State Monograph Series*. 9, NISER, Ibadan.

Omotor, D. G. (2010). Demographic and Socio-economic determinants of crimes in Nigeria (A panel data analysis). *Journal of Applied Business Economics*, 11(1), 181–195.

Paliwal, K. K. (1992). Dimensionality Reduction of the Enhanced Feature Set for the HMM-based Speech Recognizer. *Digital Signal Processing*, 2, 157-173.

Pearson, K. (2010). On Lines and Planes of Closest Fit to Systems of Points in Space (PDF). *Philosophical Magazine*, 2(11), 559–572.

Rafael Do, E. S. (2012). Principal Components Analysis Applied to digital image compression; 10(2), 135-139.

Rennison, C.M., Gover, A.R., Bosick, S.J. & Dodge, M. (2011). Reporting violent victimization of the police: A focus on black, white, Asian and Hispanic adolescent victims. *The Open Family Studies Journal*, 4, 54-67.

Robert, P. & Zauberman, R. (2011). *Mesurer la delinquance*. Paris: Presses de Sciences Po.

Romesburg, C. (2004). *Cluster analysis for researcher*. Lulu Press.

Rosti, A. & Gales, M. (2003). Factor Analysis Hidden Markov Models for Speech Recognition, CUED/F-INFENG/TR.453, Cambridge University.

Roweis, S. & Ghahramani, Z. (2014). A Unifying Review of Linear Gaussian Models, *Neural Computation*, 11(2).

Santanu, P. & Jaya, S. (2008). Rice Disease Identification using Pattern Recognition Techniques. *Proceedings of the 11th International Conference - ICCTT*, 420-423.

Tarling, R. & Morris, K. (2010). Reporting crime to the police. *The British Journal of Criminology*, 50(3), 474-490.

Vukosi, N. M. (2015). Autoennodes, Principal Components Analysis and Support Vector Regression for Data Imputation. *Journal of Research Gate*; School of Electrical and Information Engineering, University of the Witwatersrand. Johannesbury, South Africa.

Ye, N., Emran, S., Chen, Q. & Vilbert, S. (2002). Multivariate Statistical Analysis of Adult Trails for Host Based Institution Detection. *IEEE Transaction on Computers*. 51(7).