



International Journal of Environment and Geoinformatics (IJE GEO) is an international, multidisciplinary, peer reviewed, open access journal.

Spatially constrained clustering of Nigerian States: Perspective from Social, Economic and Demographic Attributes

Olanrewaju LAWAL

Chief in Editor

Prof. Dr. Cem Gazioğlu

Co-Editors

Prof. Dr. Dursun Zafer Şeker, Prof. Dr. Şinasi Kaya,

Prof. Dr. Ayşegül Tanık and Assist. Prof. Dr. Volkan Demir

Editorial Committee (April 2020)

Assos. Prof. Dr. Abdullah Aksu (TR), Assit. Prof. Dr. Uğur Algancı (TR), Prof. Dr. Bedri Alpar (TR), Prof. Dr. Lale Balas (TR), Prof. Dr. Levent Bat (TR), Prof. Dr. Paul Bates (UK), İrşad Bayırhan (TR), Prof. Dr. Bülent Bayram (TR), Prof. Dr. Luis M. Botana (ES), Prof. Dr. Nuray Çağlar (TR), Prof. Dr. Sukanta Dash (IN), Dr. Soofia T. Elias (UK), Prof. Dr. A. Evren Erginal (TR), Assoc. Prof. Dr. Cüneyt Erenoğlu (TR), Dr. Dieter Fritsch (DE), Assos. Prof. Dr. Çiğdem Göksel (TR), Prof. Dr. Lena Halounova (CZ), Prof. Dr. Manik Kalubarme (IN), Dr. Hakan Kaya (TR), Assist. Prof. Dr. Serkan Kükrer (TR), Assoc. Prof. Dr. Maged Marghany (MY), Prof. Dr. Michael Meadows (ZA), Prof. Dr. Nebiye Musaoğlu (TR), Prof. Dr. Erhan Mutlu (TR), Prof. Dr. Masafumi Nakagawa (JP), Prof. Dr. Hasan Özdemir (TR), Prof. Dr. Chryssy Potsiou (GR), Prof. Dr. Erol Sarı (TR), Prof. Dr. Maria Paradiso (IT), Prof. Dr. Petros Patias (GR), Prof. Dr. Elif Sertel (TR), Prof. Dr. Nüket Sivri (TR), Prof. Dr. Füsün Balık Şanlı (TR), Prof. Dr. Uğur Şanlı (TR), Duygu Ülker (TR), Assoc. Prof. Dr. Oral Yağcı (TR), Prof. Dr. Seyfettin Taş (TR), Assoc. Prof. Dr. Ömer Suat Taşkın (US), Dr. İnese Varna (LV), Dr. Petra Visser (NL), Prof. Dr. Selma Ünlü (TR), Assoc. Prof. Dr. İ. Noyan Yılmaz (AU), Prof. Dr. Murat Yakar (TR), Assit. Prof. Dr. Sibel Zeki (TR)

Abstracting and Indexing: TR DIZIN, DOAJ, Index Copernicus, OAJI, Scientific Indexing Services, International Scientific Indexing, Journal Factor, Google Scholar, Ulrich's Periodicals Directory, WorldCat, DRJI, ResearchBib, SOBIAD

Spatially constrained clustering of Nigerian States: Perspective from Social, Economic and Demographic Attributes

Olanrewaju Lawal 

Department of Geography and Environmental Management, Faculty of Social Sciences, University of Port Harcourt, Port Harcourt, Rivers State

* Corresponding author:
E-mail: olalaw@hotmail.com

Received 06 June 2019
Accepted 25 March 2020

How to cite: Lawal (2020). Spatially constrained clustering of Nigerian States: Perspective from Social, Economic and Demographic Attributes, *International Journal of Environment and Geoinformatics (IJEgeo)*, 7(1): 68-79. DOI: 10.30897/ijegeo.588032

Abstract

Creation and differentiation of regions are some of the basic tasks in geographic analysis. Regionalisation attempts to create a generalised representation of the processes which is taking place at the level of the amalgamated geographic units. To this end, this study examined the combined use of demographic, economic and poverty characteristics of States across Nigeria to create regions relevant for economic and development planning. The study utilised dependency ratios derived from gridded age structure data, Gross domestic product (GDP), poverty index. K-Means and Max-p algorithm were used for identification of regions. Correlation analysis showed that Youth dependency and total dependency have a strong statistically significant positive relationship ($r=0.998$, $p<0.01$) indicating that dependency in the country is driven by youth. The best K-Mean clustering implementation without considering contiguity identified 12 regions with a ratio of between and total sum of squares (RBTSS) of 0.789. The Max-p algorithm was tested with population constrain, the best result identified 9 regions with RBTSS of 0.611 constrained by a minimum population of 8% and implemented with the greedy local search algorithm, this was the same for the simulated annealing approach (SA). With high dissimilarity still common across a handful of the regions identified, a further test was carried out using a minimum bound of 3 States and the SA local search approach. The best result identified 11 contiguous regions with only one region having a relatively high within region dissimilarity and a RBTSS of 0.626. The results confirmed that there are more than 6 regions as currently defined for the country. The analyses showcased an example of knowledge discovery from a spatial dataset which could support regional development planning. From the results, there is a clear need for re-examination of current regions and designing of better-defined regions to ensure that development is guided by evidence.

Keywords: Constrained Clustering; Geoinformatics; Regionalisation; K-Means; Max-p

Introduction

Creation and differentiation of regions are some of the basic tasks in the study of geography. Creating zones or regions allows geographers to study geographic units – Nations, States, Local Government Areas (LGA), Wards, etc. from the regional point of view. Thus, allowing for the understanding of differences, association and interaction among these units within regions formed or developed. Therefore, regionalisation allows for better understanding of the similarity of attributes and characteristics among members of a region and the differences between regions. Regionalisation attempts to create a generalized representation of the processes which is taking place at the level of the amalgamated geographic units. To this end, this study examined the combined use of demographic, economic and poverty characteristics of States across Nigeria to create regions relevant for economic and development planning. The identification of such regions could serve as a basis for the development of a sponsored growth and balanced regional development measures as proposed by the Institutional Reform Theory postulated by (Myrdal, 1957).

Creation of zones/regions could be based on the understanding of human geography (social, political, economic) while it could also be based on physical geography (soil, topography, geomorphology, etc.). Regionalisation for these, help capture the natural (physical geography) and human (human geography) process. While there are many comprehensive regionalisation efforts based on physical processes, there are very few in the case of human geography, especially for Nigeria. Thus, any effort at creating human geography-based regions, based on empirical data is worthwhile in filling this gap. This effort will also help in support decision making and policymaking for national and regional development strategy formulation. Depending on the level of analysis the outcome of regionalisation could serve various purposes – from an understanding of local needs to development of strategic plans.

One of the oldest attempts at regionalisation of human geography of Nigeria was by Buchanan and Pugh (1955) predating the nation's independence and as such provides a vantage point which is closer to the colonial view of the nation (regions were potentially highlighted by the position of the colonial masters). Another attempt was made by (Udo, 1970), in which the author divided the

country into twenty regions (Table 1) and grouped into three zones namely Guinea coastlands; the middle belt and the Sudanic north.

These zoning efforts combine physical and human characteristics of the countries to identify units with unities and differences in creating the regions. These efforts are laudable in creating thematic zoning which relates to the economies and the societies across the country. However, these need to create new

comprehensive regionalisation of human geography across for the country taking into cognisance recent development and activities across the country. To achieve the development of such regionalisation effort, Geographical Information Systems (GIS), Geocomputation, Big Earth Data (BED) and Remote sensing (RS) are of great value. With increasing accessibility to data from various sources, increasing computer processing power and speed, it is becoming easier to create regions.

Regions	Regions
1. Lagos Metropolitan District	11. Afenmai Uplands, Ekiti and Western Kabba
2. Creeks and Lagoons of South-Western Nigeria	12. Middle Niger Valley
3. Cocoa Belt and Yorubaland	13. The Jos Plateau
4. Benin Lowlands	14. Benue Valley
5. Niger Delta	15. North Central Highlands/Plains
6. Palm Belt of South-Eastern Nigeria	16. Rima Basin
7. The Scarplands	17. Kano Region
8. Cross River Basin	18. Chad Depression
9. South-Western Grass Plains	19. Eastern Borderlands
	20. Gongola Basin

The current endeavour is targeted at creating regions from the 36 States and the Federal Capital Territory using selected population and socio-economic factors. Thus, assessing various techniques currently used in the creation of spatially contiguous regions. Population and socio-economic factors were selected based on their ability to indicate the manifestation of history, culture, tradition as well as the economic realities of the society. Clustering of economic, social, cultural and political attributes and activities could provide an understanding of the potential impact of location on these, thereby contributing to informed decision making for better social and economic planning. Therefore, the zoning of geographic units creates new regions that can highlight similarities and differences which could support the identification of needs and interventions. Consequently, by identifying units with similar characteristics, policy tools and measure could become clearer and purposeful in harnessing the economic potential and policy targeting for each region. The study aims that creating spatially contiguous regions with similarities in age structure and selected economic indicators, thereby contributing to the development of locally relevant economic policies for stimulating economic development in Nigeria.

Observation in space often suffers from spatial autocorrelation - the exhibition of correlation between the values of a variable attributable to their location. This thus nullifies the independent observation assumption, of conventional statistics as alluded to by Tobler first law of Geography (Griffith, 1987). Therefore, analysis of spatial autocorrelation presents an opportunity for the creation of regions across an area of interest. Spatial clustering is fundamentally based on our understanding of spatial autocorrelation. Spatial clustering could be based on attribute or location similarities. Spatially constrained clustering combined these two to create contiguous regions (with locational and attribute similarities). Thus, allowing the problem of regionalisation to be solved using computational techniques.

Human activities and habitation are directly impacted by location. This interrelationship and dependencies often lead to the increasing agglomeration of businesses and industries at specific locations (e.g., established urban centres). Similarly, there is a tendency for segregation of people with a similar culture, tradition, behaviour, social class, etc. across different region and locations. Spatial clustering impact economic activity, leading to spatial unities and differences. Therefore, regionalisation using multivariate clustering methods could provide new insights into the interaction among location, demographic and socio-economic factors.

There is a huge number of works examining spatial clustering across a wide range of topics e.g. Kim, Phipps, and Anselin (2003) - air quality and house prices; Grineski and Collins (2010) - environmental inequality/injustice; Longley and Tobón (2004) - urban deprivation; Ceccato and Uittenbogaard (2014) - real state (Kaya & Gazioglu, 2015- crime pattern; Yu, Wei, and Wu (2007) - house prices; Zou (2014) - higher-priced mortgages; (Fan & Myint, 2014) - landscape fragmentation; (Lawal, 2015) - migrations (Ülker et al., 2018) - age dependencies, Olanrewaju Lawal (2017) - dependency and socio-economic factors.

Clustering with constraints can be defined as a group of clustering methods which is aimed at delimiting homogenous regions based on univariate or multivariate characteristics while forming blocks/regions that are adjacent over space and/or time (Legendre, 1987). Essentially, constrained clustering could be carried out temporally, spatially or for a combination of both. Adams, et al., (2016) created and Ülker et al., 2018 suggested four subregions within a large basin (St. Lawrence Basin, Canada) using 21 attributes (for defined ecological units) by employing spatially constrained clustering. This was carried out to aid the delineation of homogenous units which could be used for optimising integrated water monitoring network for the basin. Thus, showcasing the use and benefit of constrained spatial clustering the water resource management. This was

used to address the computational resource demand in the optimisation of an integrated network of monitoring stations for a basin.

Our population and activities represent one of the major sources of change on the planet, thus understanding of these and its manifestations in the social, economic and political sphere is pertinent. Demography studies the dynamics of the population and various demographic processes to derive insights into the population pattern and structure. These have wide implications for the environment, culture, economy, politics and other social issues. Insights from demography could help in capturing an aspect of human geography which could be used in the differentiation and creation of regions.

Dependency ratios (DR) gives an indication of the working population and the potential withdrawals by the economically inactive population. The Total DR (TDR)

is the proportion of the people younger than 15 years and those older than 64 compared to that of the working age population of 15-64 years old (The World Bank Group, 2017). Youth DR (YDR) is the ratio of the total population below 15 years and working population age of 15-64-year-olds (Pirani, 2014) while Aged DR (ADR) is that of ages above 64 compared to the working age population.

Recent estimates by the World Bank Group (2017), shows that globally, TDR has been in decline across the world and stood at an average of 54.2 in 2016. However, there are significant regional variations across the world as shown in Figure 1. Values for Nigeria in comparison with Ghana shows a recent slight increase where that of Ghana show a continuous decline since 1983, with Nigerian figures slightly above the Sub-Saharan African figures.

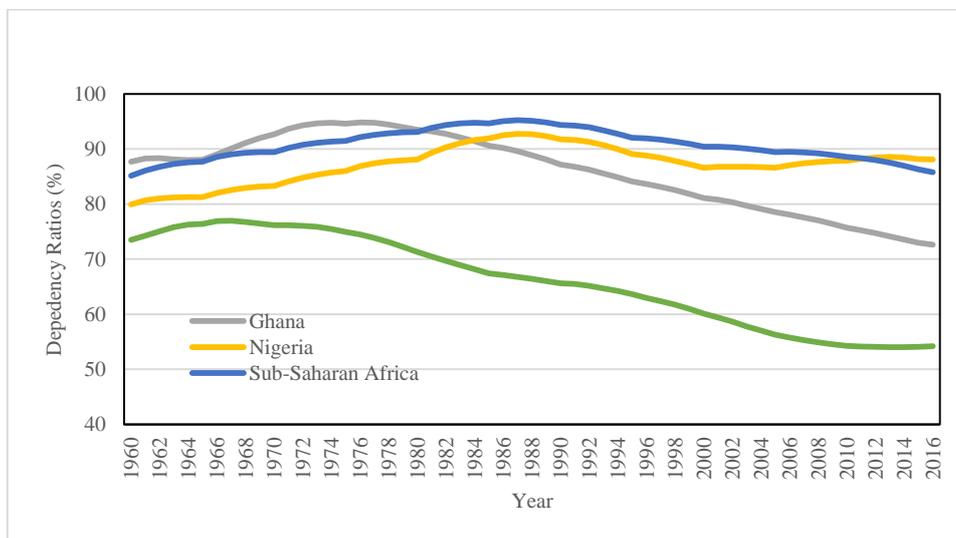


Fig. 1: Trend in age dependency ratio across the selected region and countries Source: Data from The World Bank Group (2017).

According to the Central Intelligence Agency (2016), the population pyramid for Nigeria indicates a very youthful population with more than half of the population (62.27%) between the ages of 0 and 24 years, 30.65% are between the ages of 25 and 54 years, around 4% are between 55 and 64 while the remaining are 65 years old and over. Clearly, more than 60% of the population are at the stage in their life where they are not yet ready for school, in school or just out of school, thus putting a lot of strain on the working age group and government to support the dependent population. The most recent World Population Prospect (United Nations Department of Economic and Social Affairs Population Division, 2019) indicated some positive development, with Sub-Saharan Africa experiencing a decline in fertility rate – providing a basis for rapid economic development and reaping of demographic dividend.

Poverty is a multidimensional problem which could be examined and categorised from a different perspective, e.g. income level, lack of access to facilities/support or basic needs, employment status or level of deprivation (Anyanwu, 2014). Its analysis is pertinent for economic

planning and development. Attributes of households can predispose them either rising above or falling below the poverty line. The World Bank Group (2016b) estimated that as of 2009 about 53.5% of the Nigerian population live in extreme poverty and made a projection for a drop to 46.8% in 2015. With the recent economic turmoil, the number is projected to edge upwards to around 48.4% in 2016. The report also highlighted poor translation of economic growth to poverty reduction in Nigeria with growth elasticity of poverty of around 0.6 % (i.e. for every 1% increase in GDP per capita there is only 0.6% reduction in poverty).

Lucas (1988) argued that human capital has the potential to contribute to faster economic growth, this would only be possible if the appropriate investments (i.e., education, training, health facilities, public health protection, etc.) were made to improve the productivity of the working population. Evidently, for the huge human capital of Nigeria, is a commensurate need for investment for such human capital to be able to contribute to economic growth. Increasing population on dependants demand resources from the working

population, thus giving rise to a negative demographic dividend (slow economic growth and a lesser amount of monies available for investment and research to spur more growth in the economy). Low TDR will usually bring primary demographic dividend – to higher economic output and productivity due to a large population of working age group. Furthermore, the lower TDR will afford the working population the opportunity to save which can support investment in businesses, research and consequently faster economic growth (secondary demographic dividend). These have been confirmed across the world, therefore, it is possible to link age dependency and demographic change to economic growth (Miles, 1999; Prskawetz et al., 2007). These authors have also reported a positive relationship between the growth rate of economically active population and productivity growth rate per worker, and this consequently manifests as a negative correlation between dependency ratio and economic growth rate. However, decreasing population is also not desirable it could induce dissaving with a negative impact on the growth of the economy (Fukuda & Morozumi, 2004). Therefore, population growth and demographic structure

must be at a level which can support the sustained growth of the economy. Evidently, different age groups within the population contribute differently to the economy, and understanding of the spatial distribution of such groups is important for economic planning and development. Kögel (2005) demonstrated that age structure can account for the difference in global output per worker. Feyrer (2007) concluded that workforce demography is strongly related to productivity and output and, the difference in productivity between OECD and low-income countries is related to the difference in the demographic structures. In addition, the author showed that individuals between the ages of 40 and 49 contributed the most to higher output per worker. Furthermore, Prskawetz et al. (2007) showed that age groups 30-49 and 50-64 have positive and significant absorption parameters, while the age group 40-49 contribute most to economic growth. They thus, concluded that it is highly plausible that highly educated young people drive the absorption process while matured adults ensure the running of the production process.

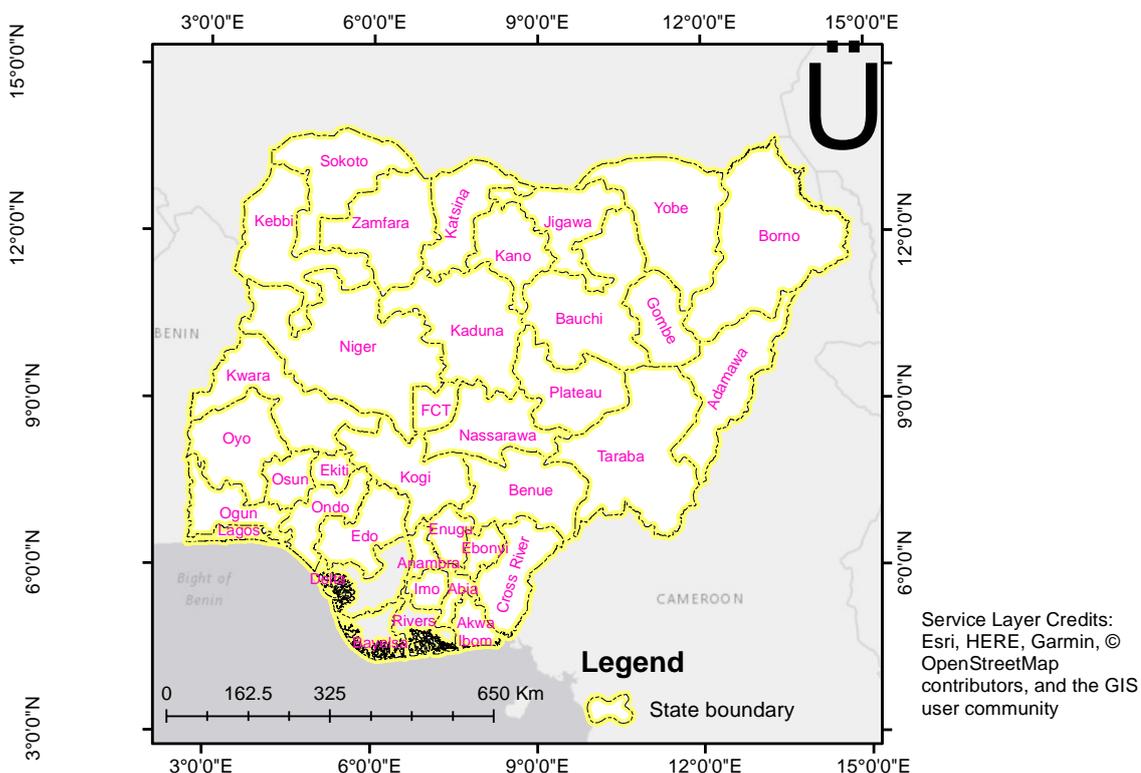


Fig. 2: the Nigerian States and the neighbouring countries

From the foregoing, it is evident that dependency within an area can influence economic growth and development. Higher dependency ratios could hinder economic development and pull more people into poverty. Recent work at the household level shows that there is a relationship between poverty, age, gender, educational attainment and regional location of the household (Anyanwu, 2014). Poverty and dependency ratios are likely to exhibit spatial autocorrelation due to neighbourhood effect which often results in places and localities sharing similar social and cultural attributes. Furthermore, access to facilities, socio-economic, level of awareness, perception, and attitude, as well as social

and political institutions, are also likely to be similar. Therefore, in creating regions relevant for economic planning and development it makes sense to employ these metrics.

From the foregoing, it is evident that age structure influence economic development and spatial pattern plays an important role in the distribution of human activities. Therefore, the creation of regions based on this understanding would serve better in the design of regional development plans which are based on similarities within regions and disparities across regions. Thereby, supporting the generation of a new understanding of economics and development with sustainable development at its core.

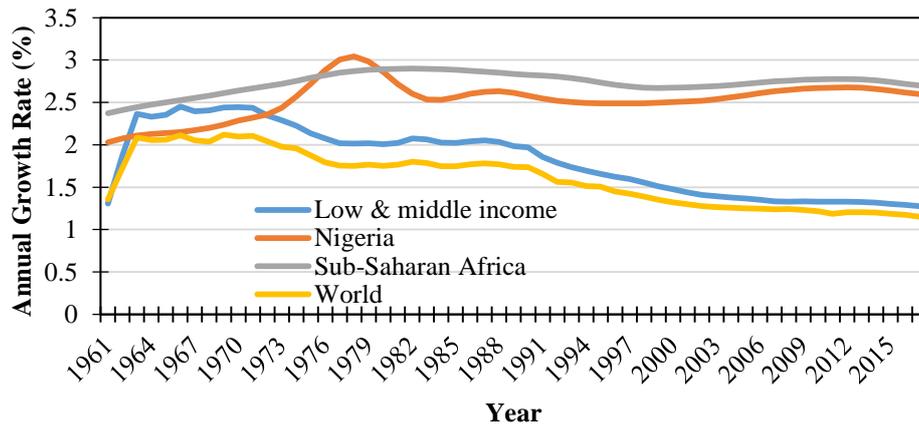


Fig. 3: Population growth rate trend among selected regions and Nigeria (The World Bank Group, 2018)

Data and Method

Study area

Nigeria is the most populous in Africa, with a population of about 182.2 million, which accounts for around 47% of the total population of West Africa (The World Bank Group, 2015, 2016a). Nigeria as a federation has 36 states and Abuja (Figure 2) as the Federal Capital Territory (FCT) and these are further divided into 774 Local Government Areas (LGA). The country is also divided into 6 geopolitical zones or regions (South-South, South-West, South-East, North-East, North Central, North-West) and this formed the basis for the division of resources and planning for the country. The population growth rate peaked in the 1970s and has been stable since the 1990s (Figure 2) when compared to the regional average (Sub-Saharan Africa) Nigeria’s population growth rate has been below the average since 1980 and currently stood at 2.6% per annum (2017). This is considerably higher than what is obtained among the lower- and middle-income countries (Figure 2).

According to the 2011 estimates, 37% of the Nigerian landmass is under arable crop production, 33% under permanent pasture while other use (urban, transport and industrial use) occupies around 13%, the remaining areas covered by forest and permanent crops (10% and 7% respectively) (Central Intelligence Agency, 2015). The Gross domestic product (GDP) in 2014 was estimated at 1.049 trillion dollars with a growth rate of about 6.5%. Furthermore, about 54% of the GDP comes from the service sector, 20% and 26% comes from the agricultural and industrial/manufacturing sectors respectively (Central Intelligence Agency, 2015). There are four climatic zones within the country (Kottek, Grieser, Beck, Rudolf, & Rubel, 2006) namely Equatorial monsoon (Am), Equatorial savannah with dry winter (Aw), Arid desert with hot arid temperature (BWh) and Arid, Steppe with hot arid temperature (BSh).

Data

Age structure gridded population data was sourced from the Geodata Institute (GeoData Institute, nd). Methods used in the production of this dataset is described by

Linard, Gilbert, Snow, Noor, and Tatem (2012). The age group is broken down into 14 groups (five years interval) starting from 0-5years and ending at above 65years old. The total population was also derived from this dataset. Socio-economic data used for this study were collated from the National Bureau of Statistics (2012) socio-economic indicators dataset and enumerated in Table 2.

Table 1: Selected social and economic variables for the study.

Variables	Acronyms
GDP (PPP billion \$) (2007)	GDPBDL07
Dollar-based on dollar/day based on adjusted PPP (as at 2010)	PPPPR10

Method of Analysis

Zonal summation was carried out within ArcGIS (ESRI, 2015), to arrive at the State total for all the 14 age groups within the population dataset. This was subsequently used in the computation of the TDR and this was computed as follow:

$$TDR = \frac{([0 - 5] + [5 - 10] + [10-15] + [above 65])}{[15 - 65]} \times 100$$

$$YDR = \frac{([0 - 5] + [5 - 10] + [10-15])}{[15 - 65]} \times 100$$

$$ADR = \frac{[above 65]}{[15 - 65]} \times 100$$

Where [XXX] represent the population of specified age group.

A correlation analysis (Pearson Product Moment Correlation) was carried out to examine the linear relationship between the demographic and the socio-economic variables selected, this was carried within SPSS version 23 (IBM, 2015). This step was carried out to eliminate any redundant variables among those selected. Constrained and non-constrained cluster analyses were carried out within the GeoDa software (Anselin, Syabri, & Kho, 2006). K-means (constrained and unconstrained) and Max-P (Duque, Anselin, & Rey, 2012) were tested to find the optimal solution. Max-p algorithm was selected for two reasons (a) there is no need to have prior knowledge of the number of regions

and (b) minimum bound could be defined for membership. Ultimately, it is also able to create contiguous regions compared to the K-Means approach.

Results

The relationship among selected variables

Correlation analysis result (Table 3) shows that YDR and TDR have a strong statistically significant positive relationship. This is a clear indication that the total dependency in the country is among young people. Evidently showing that the country has a very youthful population an attribute that could be a risk and an opportunity depending on how it is managed (Lawal, 2017).

Table 2: Relationship among selected demographic and socioeconomic variables

	TDR	YDR	ADR	GDPBDL07
YDR	0.998**			
ADR	-0.046	-0.106		
GDPBDL07	-0.615**	-0.609**	-0.067	
PPPPR10	0.249	0.259	-0.185	-0.071

** . Correlation is significant at the 0.01 level (2-tailed).

The State Level GDP showed a medium but statistically significant negative correlation with both TDR and YDR. Thus, indicating that economic output is negatively associated with a high level of dependency. This could be expected since the dependent population is not actively demanding from the working age population thus reducing resources available for investments, research, and development. Evidently, as it has been previously reported by other authors (Sinding, 2009; Vijayakumar, 2013) high fertility rates lead to an increase in dependency which reduces per capita expenditure and ultimately economic growth.

The poverty index used (PPPPR10) shows a positive association with TDR and YDR and a negative association with GDP and ADR. However, none of these associations is statistically significant. While it is evident that there are relationships however, the data at hand indicate the relationship is likely only to be due to chance.

From this result, the decision was made to drop the YDR in the regionalisation operation since it shows a very strong positive relationship with TDR. It was deemed redundant based on its a strong relationship with TDR.

Regions from K-mean Clustering Approach

K-mean clustering benchmark

In order to have a basis for comparison, the K-mean Algorithm was tested initially to examine what is possible with the selected variables for creating geographical units. In doing this, the number of regions to create was set to 6. The result (Figure 4) shows that in 12 States across the Northern parts and the Middle Belt of Nigeria are similar in terms of the selected variables (Class 1).

Moreso, the FCT displayed a uniqueness which made it standalone (Class 6) while Lagos and Rivers (Class 5) also exhibited a level of similarity which allowed for grouping them as a Class. Class 1 and 3 are the only classes which showed contiguity across their members.

With a total sum of squares of 144 and a between-cluster sum of squares of 113.584, the RBTSS (0.789) gave an indication that the clustering is quite efficient. A look at the within cluster sum of squares for each of the classes that this ranges between 0 (for Class 6) and 7.749 (Class 4). This could be expected since the clustering is not spatially constrained.

K-mean clustering with weighted optimisation

To examine the possibility of getting contiguous regions using the K-means algorithm, weight was assigned to the geometric centroid (GC) of each polygon (State). Essentially, this allows for consideration of location in grouping the States into regions. The test started with full weight (1) given to the geometric centroid and similarity (or differences) indices were computed. For optimisation a bisection search approach was adopted to find the best weight for which spatial contiguity can be achieved. The results of the search are presented in table 4.

Table 3: Weight optimisation result for K-means clustering

Weight for GC	Between-cluster sum of squares	Ratio of between to total sum of squares (RBTSS)
1	48.585	0.337
0.5	97.287	0.676
0.25	113.584	0.789
0.375	97.287	0.676
0.75	76.232	0.529
0.875	53.166	0.369

NB. Total sum of squares = 144

With a weighting of 1 for the GC (i.e. no consideration for the selected variables only the location was a consideration), expectedly contiguity was achieved (Figure 5a). However, there was a high level of dissimilarities among the members as reflected in the RBTSS value (0.337). Reduce the dissimilarity (within group sums of squares) weight of 0.5 was considered. The regions illustrated in Figure 5b and RBTSS (0.676) for this weighting (0.5) shows that some the regionalisation is not spatially contiguous. In addition, Classes 1 and 2 also have high dissimilarities (within-cluster sum of squares 20.92 and 13.06 respectively) among members. With complete contiguity not achieved further test were conducted using different weighting as presented in Table 4. The last weight of 0.875 achieved contiguity (Figure 5f). However, the RTBSS (0.369) shows that there is a high level of dissimilarity within each region. For the six clusters (classes) the within-cluster sum of squares ranges between 34.226 (Class2) and 3.486 (Class 6), with Classes 1 – 3 having a comparatively higher level of dissimilarity. With contiguity achieved, there is another benchmark from K-means for which results from Max-p algorithm can be compared.

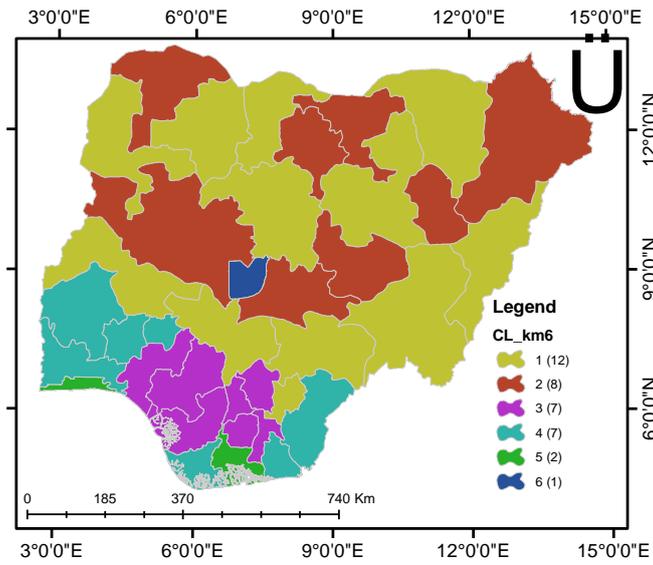


Fig. 4: K-means regions for States using selected variables

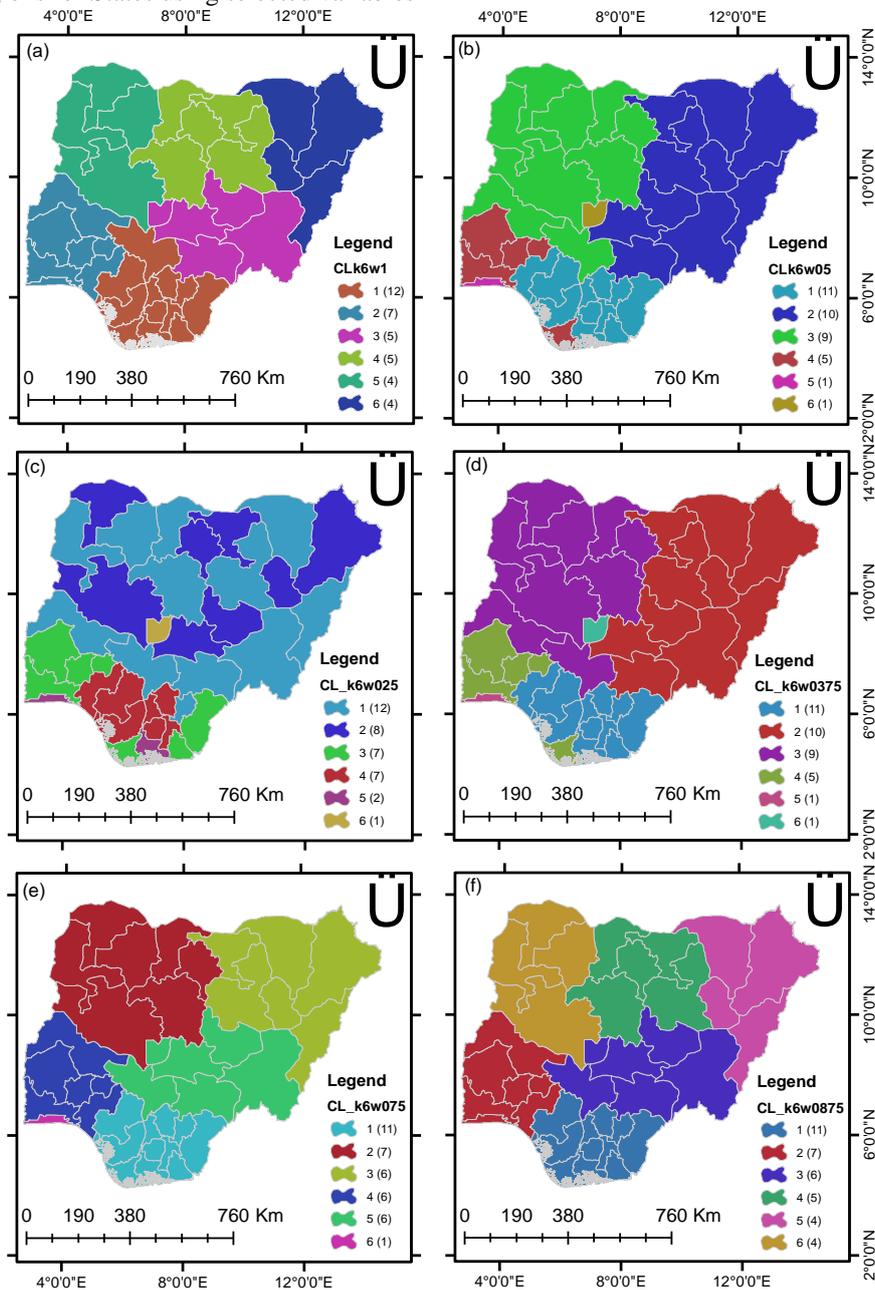


Fig. 5: Weight optimisation cluster with K-means.

Table 4: Results of Max-P algorithm region optimisation with greedy local search.

Minimum Size Constrain	Between-cluster sum of squares	Ratio of between to total sum of squares (RBTSS)
Greedy Algorithm Local Search, 1000 iteration		
TPOP = 13%	75.050	0.521
TPOP = 10%	81.229	0.564
TPOP = 5%	110.199	0.765
TPOP = 8%	87.922	0.611

NB. Total sum of squares = 144

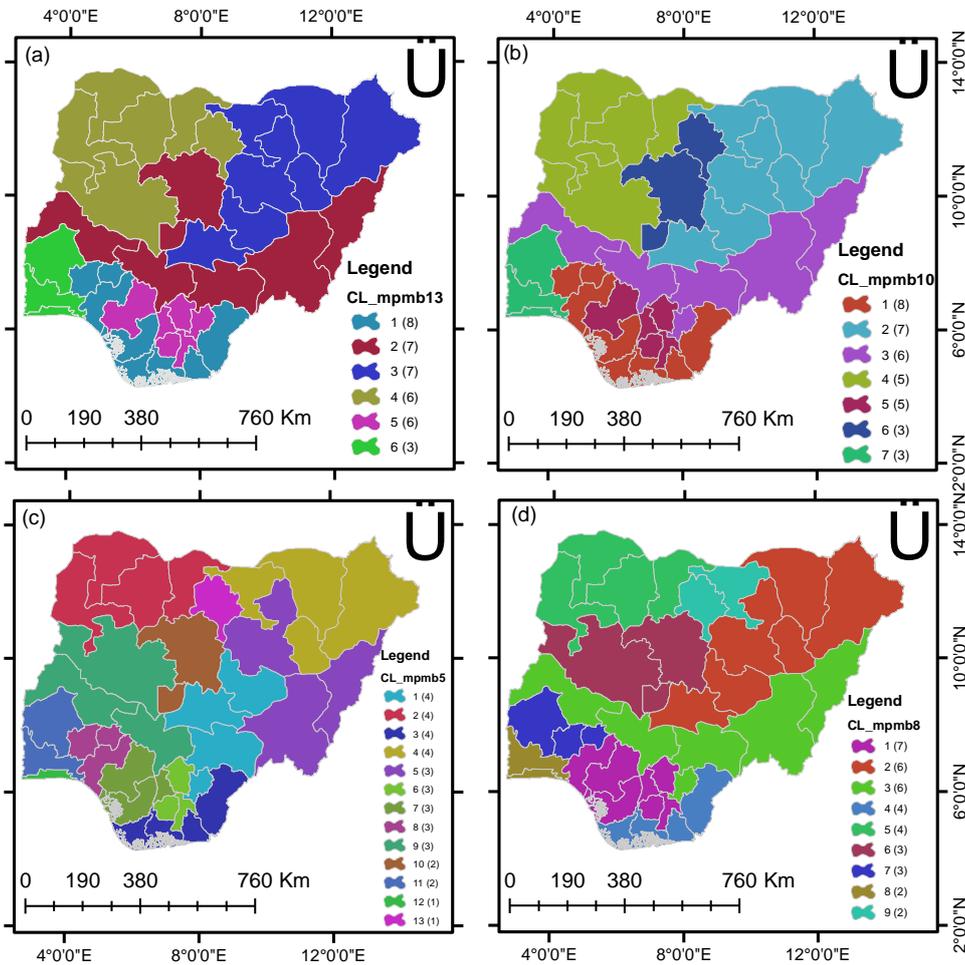


Fig. 6: Max-p algorithm with greedy local search and population constrain.

Spatially constrained clustering with Max-P Algorithm

Max-P allows for the initiation of clustering without prior knowledge of a number of regions or clusters to be created. Therefore, the algorithm can search for the appropriate number of clusters to create. However, there is a need to specify the minimum size constrain on the solution finding process. In this case, the total population (of the entire country was used), and a bound of 13% (12.5% was approximated to 13%) was used. This was selected because the previous test gave an indication that 6 regions were not adequate (high within-cluster sum of squares) and using 13% could divide the regions into 8 regions of similar population size. This thus allows for the separation of some of the highly dissimilar members of the previously identified regions (Figure 5).

Running 1000 iterations using the Greedy Algorithm for the local search, several tests were conducted with the aim of achieving contiguity and minimising dissimilarity within regions. The results are presented in table 4 and figure 6.

Initial test with 13% total population constrain (Table 5 and Figure 6a) showed that 6 contiguity regions could be achieved. However, there is a high dissimilarity within three classes (within-cluster sum of squares of 14.229, 16.764, 18.825 for Classes 1, 2, 6 respectively). With this level of within-cluster dissimilarity, another test was carried out using a lower population of constrain (10%). The results showed a slight improvement in the RTBSS value (0.564 compared to 0.521) however three class still exhibited high within cluster disparity even though there are 7 contiguous regions created (Figure 6b).

An additional search was conducted, using a reduced population constrain of 5%. This created 13 regions and improve the RTBSS (0.765) value. But Classes 12 and 13 (Lagos and Kano State - Figure 6c) while all the other regions have a lower within-cluster sum of squares. Essentially, this reduced similarity at the expense of contiguity. A new test was conducted using a population constrain of 8% and the result in Table 5 and Figure 6d showed that 9 contiguous regions could be created. However, two regions showed a high level of dissimilarity (within sum of squares > 10) – Class 6 and 8. It is important to note that these regions have Abuja and Lagos States respectively as members.

From the foregoing, it is obvious that there more than 6 regions are required to effectively regionalise the States and Max-p can be adopted in ensuring contiguity. To further ensure the robustness of the solution identified another local search algorithm was tested with the Max-p algorithm. The simulated annealing (SA) was tested and the results are presented in Table 6 and Figure 7.

The 13% population constrain showed a very slight improvement in RTBSS (0.522 compared to 0.521) and created 7 contiguous regions. But the level of dissimilarity with the regions is still high for Classes 2, 5 and 7. This high dissimilarity could be attributed to the presence of FCT, Rivers and Lagos State (respectively) within these regions. SA also performed slightly better when 10% of population constrain was implemented, creating 8 contiguous regions with RTBSS of 0.574 (compared to 0.564) with high dissimilarity for Class 7 and 8. This could be attributed to the presence of Lagos and FCT (respectively for each Class).

Using Max-p the regionalisation was also carried out with the minimum number of members as the constrain and in this case a minimum of three was adopted. This test was carried out find out if there is a clustering pattern that would ensure a further reduction in the internal inconsistencies for some of the regions derived from previous simulations. The test was carried out using a cooling rate of 0.85, simulated annealing for local search algorithm and 1000 iterations as well as the selected variables. This produced 11 contiguous regions (Figure 8) with a RTBSS of 0.626. All these regions have dissimilarity measure (i.e. within-cluster sum of squares) less than 10 except Class 10. Indicating that which there are some dissimilarities within the regions (based on the selected socio-economic variables), such are differences are relatively minimal. Class 10 show a within-cluster sum of squares of 18.82. This could be attributed to the fact that Lagos is mostly a national outlier for many factors (social and economic) in the country.

This result (Figure 8) represents a better trade compared to all the other schemes that have been presented (Figure 4 -7), with contiguity achieved and the dissimilarities minimised. To ensure that the solution identified is not trapped in a local optimal a sensitivity analysis was carried out by increasing the number of iterations for the best solutions obtained. For this, 5000 and 10000 iterations were tested to check the consistency of the results. The result for the two tests (Table 7), gave an indication that the solution remained consistent irrespective of the number of iterations.

Table 5: Results of Max-P algorithm region optimisation with Simulated Annealing local search

Minimum Size Constrain	Between-cluster sum of squares	Ratio of between to total sum of squares (RBTSS)
Simulated Annealing local search algorithm, 1000 iterations, cooling rate of 0.85		
TPOP = 13%	75.147	0.522
TPOP = 10%	82.682	0.574
TPOP = 8%	Same as in Table 5 for Greedy algorithm	

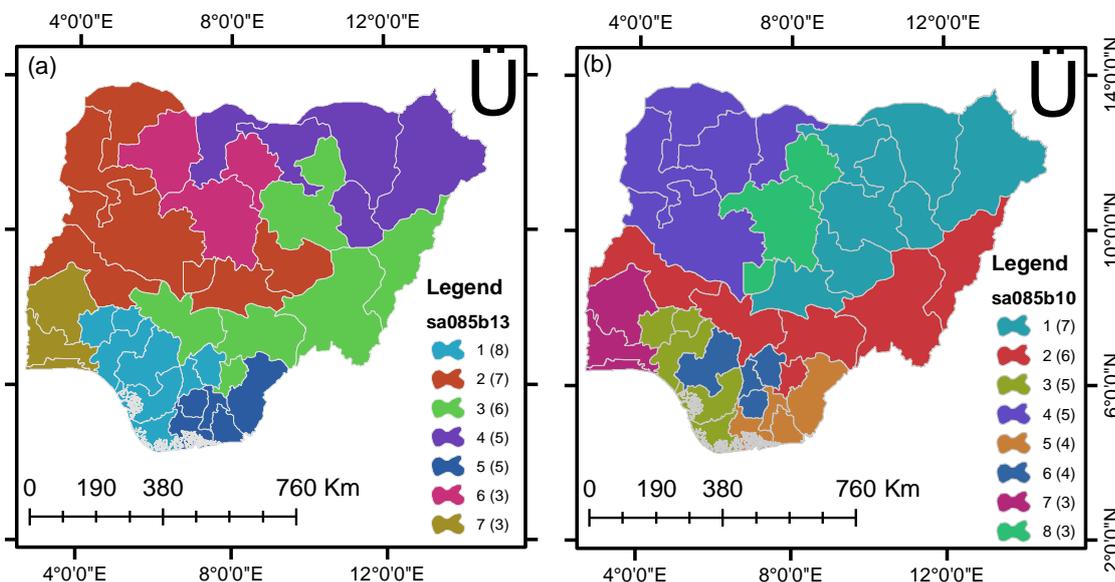


Fig. 7: Max-p with simulated annealing and population constrain.

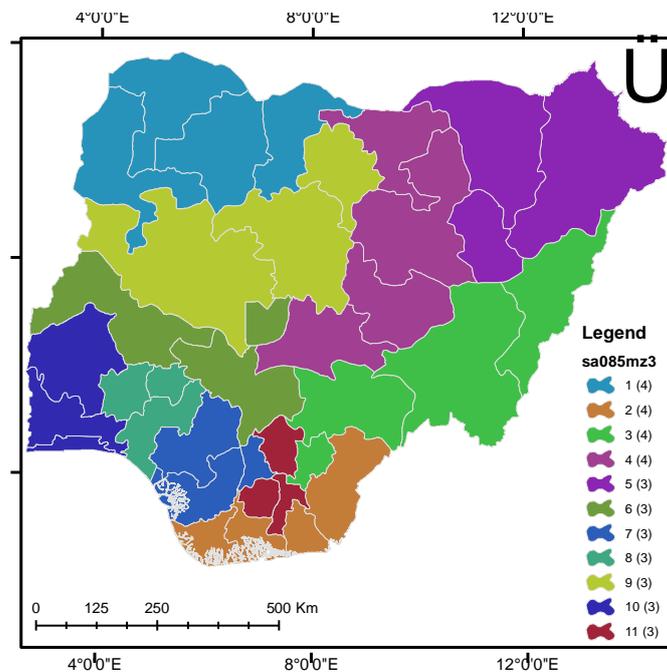


Fig. 8: Max-p with simulated annealing and minimum member constrain.

Table 6: Dissimilarity and Similarity within regions (1000, 5000, 10000 iterations)

Class	Within Cluster Sum of Squares
1	3.112
2	7.087
3	2.643
4	4.405
5	1.175
6	8.958
7	0.873
8	1.558
9	1.607
10	18.825
11	3.605

Subsequent analyses were based on this result to explore the regions created. Class 10 made up of Lagos, Ogun and Oyo States has the highest level of internal differences. This reason for this becomes apparent looking at the descriptive statistics for this region. For example, the GDP (PPP billion \$) ranges between 10.500 (Ogun State) and 33.700 (Lagos State), with a mean of 20.100 and standard deviation of 12.106. The same high level of variation was recorded for TDR with mean of 65.397, (SD = 10.663); ADR with mean of 5.621 (SD = 1.844); and PPPPR10 (dollar/day) with mean of 0.447 (SD = 0.152).

Class 6 (FCT, Kogi, Kwara) and Class 2 (Akwa Ibom, Bayelsa, Cross Rivers, and Rivers) showed a moderately high internal disparity. However, for the sake of ensuring contiguity, these level of disparity is an acceptable trade-off. The disparity for this group could be attributed to the values of the selected variables for FCT (lowest TDR and ADR in the region) and Rivers (high GDP and

PPPPR10) for Class 6 and 2 respectively. With the remaining 8 classes showing low internal disparity and signifies where similar States are based on the selected data.

Discussion and Conclusion

The analyses showcased an example of knowledge discovery from data. Thus, highlighting the relevance of data and evidence-driven planning and policymaking to support the achievement of the Sustainable development goal and providing new understanding and knowledge.

In addition to States and LGAs, Nigeria is also divided into 6 geopolitical zones or regions - South-South (SS), South-West (SW), South-East (SE), North-East (NE), North Central (NC) and North-West (NW) and has often been used in the division of resources and planning for the country. The work of Udo (1970) identified 20 regions based on human and biophysical characteristics of the country as at then. However, in the light of dichotomous growth and development across the country (advantaged and disadvantaged States), it has become obvious that for economic planning and development there is a need for regionalisation to be based on data and evidence. Simulations with different criteria and algorithms show that based on selected variables there are more than six regions across the country. Eleven regions could be identified using the Max-p algorithm with a considerably high level of internal consistency across most of the regions and contiguity - which could support regional policy and development planning.

Given the possibility of starting with an unknown number of regions, Max-p Algorithm displayed extensive flexibility which could accommodate different criteria for the initiation of the regionalisation exercise. K-Means clustering outperformed Max-P clustering in ensuring lower internal dissimilarity (i.e. within-group differences). However, the aim for regionalisation is to

find similar and contiguous regions, as such Max-P outperformed the K-Mean algorithm in ensuring a contiguous and minimising dissimilarity.

Given that States are the next level of administration beyond the national boundaries, creating regions based on State attributes is vital in ensuring that regionally relevant policies and plans can be identified and implemented to jumpstart development across the country. The identification of these regions using a combination of demographic (dependencies), social (poverty index) and economic (GDP) captured some of the essential aspects necessary for development planning. The need for the data-driven policy is even more important now than ever as Nigeria is predicted to be the Third most populous country in the world according to the World Population Prospect (United Nations Department of Economic and Social Affairs Population Division, 2013). Thus, there is a clear need for re-examination of current regions and designing of better-defined regions to ensure that development is guided by evidence and not political reasons.

The characteristics examined in this study are relevant to economic and social development. For example, the dependency attributes could help in understanding how the demographic dividend could be harnessed in each region and in identifying which economic functions and resources should be allocated to each region. The identified regions and their structure could support the decision on how to ensure the spread of economic development across the country using the regions as the basis for targeted policy formulation. The number of regions could also help in curtailing divergent economic growth (an unsustainable development) as currently evident across the country. Consider, the entire country as a landscape (socio-eco-political landscape) the interaction among the properties such as function (interaction between spatial elements) of each region and structure (the spatial relationship between distinctive patches) of the regions will bring about change which will iteratively modify the function and the structure of the entire landscape (Lawal, 2009). Thus, the modification of the current 6 zones, if backed up with proper policy and implementation is bound to change the economic development of the country.

According to the Multiplier Effect (Myrdal, 1957), if the dualism persists backwash effect will continue to drain the less developed area of human capital and other resources and concentrate the positive effect only in the developed areas unless there is an intervention. Therefore, the intervention could be in the form of redefining the regions to better identify relevant policies for each region. According to Myrdal (1968) - The Institutional Reform Theory, effort must be made to curtail dualism (divergence) in development as such balanced development should be the focus of any nation. To achieve this balanced growth, thus regions must be defined and policy-relevant for their growth and development must be implemented accordingly. Clearly, the first step is to define those regions.

In conclusion, the study has revealed that 11 regions could be identified using the demographic, social and economic characteristics of the 36 States and FCT in Nigeria. Thus, showing that the current regions as defined in the country are not adequate to support balanced economic development across the country (region creation must be based on data and not bias for balanced growth is going to be achieved).

The structure of the regions as identified in this study create a basis for further enhancing of spatial interaction among the entities within each region and across the regions. By maximising similarities and contiguity, the result can support effort for the reduction of spatial inequality and promote sustainable development in the country.

Acknowledgments

The authors declare that they have no financial or personal relationships which may have inappropriately influenced them in writing this article.

References

- Adams, M. D., Kanaroglou, P. S., Coulibaly, P. (2016). Spatially constrained clustering of ecological units to facilitate the design of integrated water monitoring networks in the St. Lawrence Basin. *International Journal of Geographical Information Science*, 30(2), 390-404.
- Anselin, L., Syabri, I., Kho, Y. (2006). GeoDa: an introduction to spatial data analysis. *Geographical analysis*, 38(1), 5-22.
- Anyanwu, J. C. (2014). Marital Status, Household Size and Poverty in Nigeria: Evidence from the 2009/2010 Survey Data. *African Development Review*, 26(1), 118-137.
- Ashford, L. S. (2007). Africa's youthful population: Risk or opportunity, (pp. 4). Washington DC: Population Reference Bureau.
- Buchanan, KM., Pugh, JC. (1955). *Land and people in Nigeria: The human geography of Nigeria and its environmental background*: University of London Press.
- Ceccato, V., Uittenbogaard, A. C. (2014). Space-Time Dynamics of Crime in Transport Nodes. *Annals of the Association of American Geographers*, 104(1), 131-150.
- Central Intelligence Agency. (2015). Nigeria: The World Factbook. Retrieved November 8, 2015, 2015.
- Central Intelligence Agency. (2016). Age Structure: The World Factbook.
- Duque, J. C., Anselin, L., Rey, S. J. (2012). The Max-P-Regions Problem. *Journal of Regional Science*, 52(3), 397-419.
- ESRI. (2015). ArcGIS Desktop (Version 10.4). Redlands, CA: Environmental Systems Research Institute.
- Fan, C., Myint, S. (2014). A comparison of spatial autocorrelation indices and landscape metrics in measuring urban landscape fragmentation. *Landscape and Urban Planning*, 121(0), 117-128.

- Feyrer, J. (2007). Demographics and productivity. *The Review of Economics and Statistics*, 89(1), 100-109.
- Fukuda, S.-i., Morozumi, R. (2004). Economic growth under the demographic transition: a theory and some international evidence* *The Economics of an Ageing Population Macroeconomic Issues*. Cheltenham, UK: 'Edward Elgar Publishing, Inc.'
- GeoData Insitute. (nd). Alpha version 2010 and 2014 estimates of numbers of people per grid square, with national totals adjusted to match UN population division estimates. Nigeria: Population per 100 meters grid Cell Retrieved 20th January, 2015.
- Griffith, D. A. (1987). *Spatial autocorrelation. A Primer*. Washington DC: Association of American Geographers.
- Grineski, S. E., Collins, T. W. (2010). Environmental injustices in transnational context: urbanization and industrial hazards in El Paso/Ciudad Jurez. *Environment and Planning A*, 42(6), 1308-1327.
- IBM. (2015). IBM SPSS Statistics (Version 23). Armonk, New York: IBM Corporation.
- Kaya, H., Gazioğlu, C. (2015). Real Estate Development at Landslides. *International Journal of Environment and Geoinformatics*, 2(1), 62-71.
- Kim, C. W., Phipps, T. T., Anselin, L. (2003). Measuring the benefits of air quality improvement: a spatial hedonic approach. *Journal of environmental economics and management*, 45(1), 24-39.
- Kögel, T. (2005). Youth dependency and total factor productivity. *Journal of Development Economics*, 76(1), 147-173.
- Kottek, M., Grieser, J., Beck, C., Rudolf, B., Rubel, F. (2006). World Map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift*, 15(3), 259-263.
- Lawal, O. (2009). *Analysis of land use changes in the Thames Gateway: the change-pattern approach*. (PhD), University of East London, London.
- Lawal, O. (2015). Geodemographic analysis of age dependencies in Nigeria. *Port Harcourt Journal of Social Sciences*, 6(1 2), 115 – 132.
- Lawal, O. (2017). Mapping Economic Potential Using Spatial Structure of Age Dependency and Socio-economic factors. *African Journal of Applied and Theoretical Economics, Special Edition*(November), 32-49.
- Legendre, P. (1987). Constrained clustering *Developments in Numerical Ecology* (pp. 289-307): Springer.
- Linard, C., Gilbert, M., Snow, R. W., Noor, A. M., Tatem, A. J. (2012). Population Distribution, Settlement Patterns and Accessibility across Africa in 2010. *PLoS ONE*, 7(2), e31743.
- Longley, P. A., Tobón, C. (2004). Spatial Dependence and Heterogeneity in Patterns of Hardship: An Intra-Urban Analysis. *Annals of the Association of American Geographers*, 94(3), 503-519.
- Lucas, RE. (1988). On the mechanics of economic development. *Journal of Monetary Economics*, 22(1), 3-42.
- Miles, D. (1999). Modelling the Impact of Demographic Change Upon the Economy. *The Economic Journal*, 109(452), 1-36.
- Myrdal, G. (1957). *Economic Theory and Under-Development Regions*: Gerarld Duckworth.
- Myrdal, G. (1968). Corruption: Its causes and effects *Asian drama: An inquiry into the poverty of nations* (Vol. 2, pp. 953-961). New Yor: Pantheon.
- National Bureau of Statistics. (2012). *Nigeria Socio-economic Indicators*.
- Pirani, E. (2014). Youth Dependency Ratio *Encyclopedia of Quality of Life and Well-Being Research* (pp. 7287-7288): Springer.
- Prskawetz, A., Fent, T., Barthel, W., Crespo-Cuaresma, J., Lindh, T., Malmberg, B., Halvarsson, M. (2007). The Relationship Between Demographic Change and Economic Growth in the EU (pp. 1-112). Wien: Institut Für Demographie Österreichische Akademie Der Wissenschaften.
- Sinding, SW. (2009). Population, poverty and economic development. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1532), 3023-3030.
- The World Bank Group. (2015). Nigeria. Retrieved 9th December, 2016.
- The World Bank Group. (2016a). Annual Population Growth Rate. Retrieved December 23, 2016, from The World Bank Group.
- The World Bank Group. (2016b). Poverty and Equity Data Portal: Nigeria. Retrieved November 11, 2017, from The World Bank Group.
- The World Bank Group. (2017). World Bank Development Indicators: Age dependency ratio. Retrieved September 12, 2017, from The World Bank Group <http://data.worldbank.org/indicator/>
- The World Bank Group. (2018). Annual Population Growth Rate. Retrieved April 24, 2019, from The World Bank Group.
- Udo, R. K. (1970). *Geographical Regions of Nigeria*: University of California Press.
- Ülker, D., Ergüven, O. Gazioğlu, C. (2018). Socioeconomic impacts in a Changing Climate: Case Study Syria. *International Journal of Environment and Geoinformatics*, 5(1), 84-93.
- United Nations Department of Economic and Social Affairs Population Division. (2013). *World Population Prospects: The 2012 Revision, DVD Edition*.
- United Nations Department of Economic and Social Affairs Population Division. (2019). *World Population Prospects 2019: Highlights*. New York: United Nations.
- Vijayakumar, S. (2013). An Empirical Study on the Nexus of Poverty, GDP Growth, Dependency Ratio and Employment in Developing Countries. *Journal of Competitiveness*, 5(2), 67-82.
- Yu, D., Wei, Y. D., Wu, C. (2007). Modeling spatial dimensions of housing prices in Milwaukee, WI. *Environment and Planning B: Planning and Design*, 34(6), 1085-1102.
- Zou, Y. (2014). Analysis of spatial autocorrelation in higher-priced mortgages: Evidence from Philadelphia and Chicago. *Cities*, 40, Part A(0), 1-10.