Trans. Inst. Indian Geographers
Vol. 19. No.2, July 1997

ON THE PRINCIPLES, METHODS AND PROBLEMS OF CLASSIFICATION AND REGIONALIZATION IN GEOGRAPHY.

TEMI. E. OLOGUNORISA, Nigeria

ABSTRACT: Classification is a scientific tool of analysis without which geographers will be unable to generalize or see underlying regularities, and by implication formulation of laws and theories. The paper argues that since geographers study spatial analysis of phenomena on the earth's surface, they are often faced with a whole problem of collecting data on a large number of variables, and also on the problem of classifying data using multivariate analysis. A consideration of the principles, methods and problems of classification and regionalization in geography which could further our understanding form the major thrust of this paper. The paper concludes that classification and regionalization are part of the basic tools we use in dealing with the world around us, and should therefore be seen and understood as means to an end rather than an end in itself.

INTRODUCTION

Since the beginning of geography, geographers have been doing some sort of classification and regionalisation. This is because geographers like other scientists have continued to be interested in how things happen in reality. For instance, the geographic zonation of the world according to the Greeks into torrid, temperate and frigid zones is a form of classification. Also Koppen's, Thorthwaite and Miller's climatic classification schemes are based on the methods of classification; or even the division of soil (in soil geography) into loamy, clay and sandy soils based on the soil properties-soil texture, colour, structure and nutrient status; or the Davisian cycle of erosion and stream ordering in geomorphology are all forms of classifiction. Even in human geography, the central place theory, von Thunnen and gravity models are all forms of classification. In short, there is no branch of geography that is not engaged in some form of classification or regionalisation.

Classification is a scientific tool of analysis without which geographers will be unable to generalise or see underlying regularities. And so making of law-like statements of universal validity will be impossible. With this background, we can argue that since geographers study spatial analysis of phenomena on the earth's surface, they are often faced with a whole problem of collecting data on a large number of variables. And these data have to be classified or subjected to a classification system in order to obtain an efficient arrangement of information in a simplified and generalised form (Abumere, 1988). Geographers began to run into problems when they started using multivariate factors in classification. It is in recognition of their roles

in geographic studies that this paper has the primary objective of examining the principles, methods or procedures and problems of classification and regionalisation in geography.

THE DIFFERENCE AND SIMILARITY BETEEN CLASSIFICATION AND REGIONALISATION

The question has always been asked if classification is the same thing as regionalisation. The term "Classification" according to Master (1976) has two connotations. First, it can be used as a verb, in this case, as a process of identifying the number, nature and composition of relatively homogeneous groups which together make up the data set under scrutiny. And secondly, as a noun in the sense of "arrangement" or "Scheme". Harvey (1969) on the other hand, defined classification as a set of rules for assigning data to their apporpriate classificatory boxes. The application of the rule according to him depends on the purpose or objective of classification. In other words, classification is a rational and scientific device or procedure by which we impose some sort of order and coherence on geographic reality or upon the vast in-flow of information from the real world.

A better clarification on the meaning of classification was however provided by Grigg (1965) as the grouping of objects into classes on the basis of some similarity in either properties, or in the relationships between the objects. The objects to be classified according to him are called the individuals, and all individuals (are presumed to) have a number of properties. And the total number of individuals classified is called the universe.

Regionalisation on the other hand, is a form of classification which brings in the criteria of contiguity constraints. Geographers do classification not only on the basis of similarity but also contiguity. Grouping on the basis of similarity alone is classification and not regionalisation. From this therefore, it could be deduced that regionalisation is a special form of classification. It deals with regional delimitation and grouping. And the delimitation of a region is basiclly a problem in classification. It can be argued that regionalisation is step ahead of classification. Classsification is very popular in biological sciences especially in taxonomy, ecology and zoology, whereas geographers are interested in areal form of classification called regionalisation. In delimiting regions, geographers are concerned with identifying areas or regional systems in which the variation betweem areas is distinctly evident and in which internal variation is minimal. Those regional systems are areal classification lying somewhere between uniqueness and maximum generality.

Debate within the academic discipline of geography rumbles on the definition of a region. Regions within the scope of this paper can be conceptualised as points, lines, areas, depending on the scale of observation. For instance, in regional scale, towns can be regarded as points and highways or roads as lines. But at a local scale, towns can be regarded as areas, and roads as lines, Site and situation characteristics are also useful in explaining the term "region" It is not the task of this paper to look at the types of regions but for more details, see Abler et al (1971), Haggett et. al (1971), Friedman (1966), Richardson (1969), Bourdeville (1966) and Berry (1966).

Having examined the difference and similarity between classification and regionalisation, it is

important to note Hannagan's (1966) caution that classification (by implication regionalisation too) is based on the basic assumption that every piece of data has characteristics. Some of which are measurable attributes or variables (such as weight) and some are non-measurable attributes (such as beauty). Measurable variables can be divided into:

(a) discrete variables measured in single events (such as people, towns, houses, cars etc);

(b) continuous variables are in units of measurement which can be broken down into definite graduations. Examples include: temperature in decimals of a degree, height or length in decimals of centimetre or fraction of an inch.

It should be noted however that every classification scheme must have a purpose. And the purpose depends on what the researcher seeks to achieve. But one thing that is clear is that, the aim of classification or regionalisation or any grouping generally is that it tries to minimize intra-group variability and maximize inter-group variability. Any grouping or classificatory technique that does not achieve this is not grouping or classification. For more detailed discussion on the philosophical basis of classification see Harvey (1969); Sneath and Sokal (1973); Cormack (1971) and Johnson, (1968).

PRINCIPLES OF CLASSIFICATION/ REGIONALISATION

All classification systems must necessarily possess the following important qualities.

First is the property of exhaustivity. This means that no object should be left out or omitted during the process of classification. Secondly, classification must be mutually exclussive.This property of mutual exclusivity implies that there must be no cases of objects standing on the fence or belonging to two groups at the same time.

The thrid property is that there should be internal consistency and coherence. This means that there should be consistency and coherence in the choice of property or criteria used at every stage in the division of objects into classes. In other words, the principle of division at every stage should be uniform throughout.

In addition to this, all classification of whatever form should have a specific purpose. In other words, any classification scheme must aim at something. Any classification designed without a specific purpose will be useless and meaningless.

Moreover, the classification of objects or individuals into groups should be based on the properties of the objects to be classed. Furthermore, classification should be dyamic. It shouldd be a continuous affair in order to accommodate changes as more knowledge is gained about the objects to be classed. In other words, classifications are not absolute, or should not be too rigid, they should be flexible.

Apart from this, objects or what are called geographic individuals are the variables classed or grouped during classification. And there should be differences in kind between objects, as objects which differ in kind will not easily fit into the same classification.

Finally, during classification, the principle adopted must be applicable to higher hierarchy than lower hierarchy. In other words the properties that are used to group or classify in the higher categories must be more important for the purpose of division than those used in the lower categories. (Chorley and Haggett, 1967).

METHODS OF CLASSIFICATION AND REGIONALISATION

Generally speaking, there are two methods or procedures of classificatin. These according to Harvey (1969) include:

(1) Classification from above. This is also called "logical division" or deductive "classification" or monothetic classification.

This type of classification is based on the principls that a universal set, U, is partitioned according to a particular property. For example, the classification of a Nation into States, Local Goverment Areas, Communitis and settlements. The division of the universal set according to Stebbing (1961) takes place in a series of steps, such that at each step one property or set of properties is used to differentiate between classes. The criteria or properties used in classification need to be placed in order of priority. Another good example of logical division is provided in Figure 1.

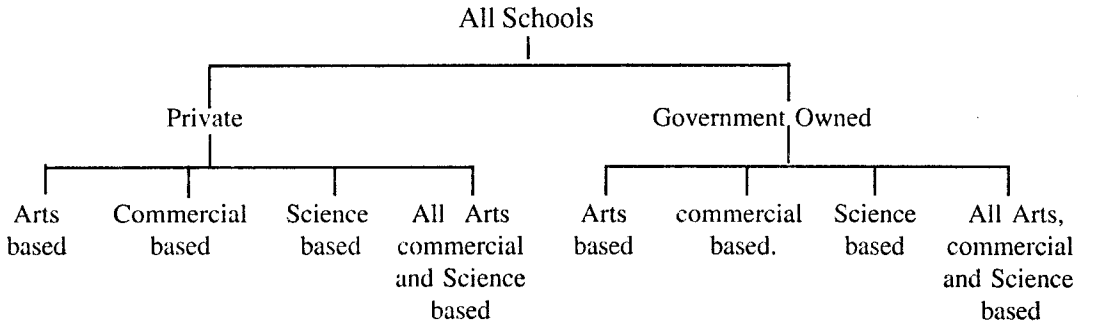

Fig. 1

Source : Author's classification

(2) The second form of classification is the classification from below. This according to Harvey (1969) is also called "Grouping" or "Inductive classification" or polythetic classification. This form of classification is based on multiple attributes, and this is also called cluster analysis. Cluster analysis which is a method of classificaton based on multiple attributes has various methods: (see fig.2).

(a) Hierarchical grouping technique looks like this Fig 2.

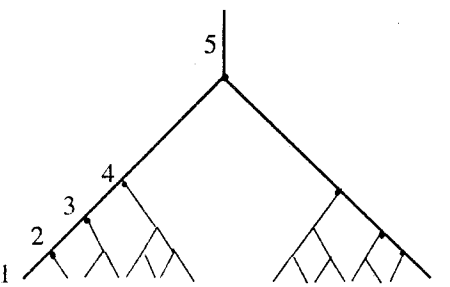

Fig. 2

Source : Abumere, 1988.

One can only belong to a group at a time and not two at a time because the two are exclusive of each other.

(b) The second is non-hierarchical grouping technique. This type of cluster analysis takes place on the basis of similarity and dissimilarity.

Similarity is measured by the correlation coefficient, while dissimilarity is not just the opposite but takes values of between 0 - 1. It measures the distance which two variables are from each other. This can be defined in many ways depending on the type of data (either interval, ratio etc.)

The dissimilarity analysis has the following attributes :

(a) Metric dependent: That is based on ordinal, nominal, interval or ratio scale;

(b) Variables are in the same units of measurements e.g. feet, metres, kilometres etc.,

It should be noted that all classification or grouping techniques have to achieve two outstanding qualities. These are the principle of exhaustivity and exclusivity in addition to those already mentioned earlier under principles of classification. These two properties can only be achieved in two ways.

(1) Through Q- mode factor analysis; and

(2) Through hierarchical and non-hierarchical grouping techniques.

According to Abumers (1988), experienced scholars make use of factor analysis to get grouping but argues that the results of such grouping are often sub - optimal. As an illustrative example, assuming we want to classify on the basis of similarity, four local government areas having different performances.

Let the four Local Government Areas be represented by capital Letters A, B, C, D.

	CASES/PERFORMANCES		
	1	2	3
A	6	5	6
B	5	6	6
C	3	2	3
D	2	3	2

If we use factor analysis, two groups will emerge, and the two groups that will be formed are AC, and BD. This is because these two have the same "profile". That is correlation between the two will be the highest for any combination of two local governments.

Other grouping techniques that group on the basis of similarity will computer AB, CD. And there are two ways of computing AB and CD. First, is through Distance matrix, and secondly, by Mahalanobis D^2 statistic.

In the case of Distance Matrix, it is based on Pythagoras theorem

$$\sqrt{(fi - fj)^2}$$

Note that fi and fj stand for any pair e.g. fi might stand for A, and fj for B as used in table 1.

On the basis of this formula, similarity matrix could be computed the smaller the value, the more similar the values.

The second way of computing AB is by Mahalanabis D^2 statistic. It is not the task of this paper to go into details about its calcuation but to give the usefulness of the method. This statistics was developed by Mahalanobis (1927, 1936) and is one of the earlier measures to define the degree of affinity between different classess. It should be noted that the statistic applies to already established classes and states

how similar they are to one another. This statistic was later developed by Rao (1952). Berry (1960,1967) and king (1960) provide examples of its use in geography. The statistic is very useful for discriminating within a relatively homogeneous population.

The other method of achieving the property of exhaustivity and mutual exclusivity is by hierarchical and non-hierarchical grouping techniques.

The hierarchical grouping technique has already been discussed. We would only give two reservations about its use. First, there is the problem which is refered to in the geographic literature as "migration problem". This means that if there is any mistake in any of the groupings, the problem will be carried on throughout.

The second is the problem of chaining, which is very common. That is, the problem of grouping the existing groups. In other words, one of the cases may not properly belong to the larger group. For these two problems to be corrected, Abumere (1988) argues that linear discriminant analysis is used. Factor analysis is normally carried out before using hierarchical techniques.

While the non-hierarchical grouping technique does not have the two problems mentioned above. This is because it examines each individual group at every step and it is on the basis of this that the non-hierarchical technique is often considered better.

However, after grouping if we discover that we still have an existing group that is not efficient, discriminant analysis is used. There are two types of discriminabt analysis:

(1) Linear discriminat analysis and

(2) Multiple discriminat analysis.

The distinction is that simple discriminant analysis is used when two groups are involved, while multiple discriminant analysis is used when more than two groups are involved.

Discriminant analysis performs two functions: First, it is used for assessing the optimality of any grouping or existing grouping; and secondly, used as a groupimg technique itself.

Apart from this, it also provides two additional measures of whether they (i.e groupings) are significant or not by:

(a) computing the chi-square to test it, and

(b) by using Mahalanobis statistic to test the grouping.

Other statistical technique that is used for grouping or classification is principal - component analysis. This statistic has a lot of similarity with factor analysis. They differ only in terms of their origin. Principal component analysis has a mathematical and statistical origin while factor analysis has its origin in the area of psychology. The similarity between the two are as follows:

(a) Both will collapse many variables into fewer variables. This ability to collapse many variables into fewer factors is called parsimony.

(b) They also help in grouping, there is what is called a Q-mode factor analysis and R-mode factor analysis. When we collapse the columns variables, we say we are doing R-mode factor analysis while Q-mode factor analysis is when we collapse the cases. Finally, the distinction between principal component and Factor analysis from the standpoint of classification is that we classify without any theory in the case of principal component while we classify with theory in the case of Factor analysis (Kendall, 1957).

PROBLEMS OF CLASSIGICATION AND REGIONALISATION

First, there is the problem of inadequacy of data on most geographic phenomena in terms of coverage, duration and reliability. There is also the problem of appropriate data which are suitable to the application of various high powered techniques used in classification and regionalisation. It has been noted for example that there are problems inherent in the factor analytic method especially in the estimation of communalities and the stability of the factors under sampling variation. There are also problems too in all the methods of clustering, linking and grouping.

Harvey (1969, P. 347) has noted that the limitations of the choice of the product-moment correlation co-efficient for regionalisation creates problem. This is because one of the technical requirements of this statistic is indepedence in the observations. Since the aim of regionalisation is to produce contiguous regions which are internally relatively homogeneous, it appears that this condition of independence in the observations wil be violated.

Apart from this, there is the problem of indentifying the crucial parameter or attributes to consider as geography deals with multivariate phenomena. For example, in the classification of climate, reinfall and temperature are often used, and this may be wrong because other climatic elements like wind. humidity, pressure and radiation also contribure to the climate of any area or region.

We would also like to recognise the fact that all classifications are artificial to the extent that we impose order (boundary) on a complexity or continuum. Consequently, many classifications are subjective (Ayoade, 1988).

In addition to this, there is the need for adequate knowledge of theory of the phenomena being grouped or classified. Otherwise, this may lead to methodological problem.

Finally, there is the problem of demarcating boundaries as also the problem of recognising the regional hierarchies. And we should note that grouping depends on the degree of affinity of the phenomena being grouped.

REFERENCES

Abler, R.J.S.; Adams and R.R. Gould (1971) Spatial *Organisation: A geographer's view of the world.* New Jersey: Engle Wood Cliffs.

Abumere, S.I. (1988) 'A note on techniques of Geographical Investigation for undergraduate students", Unpnlished Lecture Note. Ibadan: University of Ibadan.

Ayoade J.O. (1988) *Introduction to climatology for the tropics.* Ibadan: Spectrum.

Berry B.J.C.(1958) "A note concerning methods of Classification". *Annal. Ass. Amer. Geograph.48* pp. 300-3.

Berry B.J.C. (1966) Essays in commodity flows and the spatial structure of the Indian Economy. *Res. Pap. No. 111.* Department of Geography, University of Chicago.

Berry B.J.C. (1967) "Grouping and Regionalising" in Garrision, W.C. and Marble. D. (eds) *Quantitative Geography, Part I&II,* Nort West University studies in Geography No. 13 14

Chorley. R.J. and Haggett, P. (1967) *Integrated Models in Geography.* London: Methuen and CO. Ltd.

Frey, A.D. Cliff and P. Haggett (1971) *Locational Methods* Lodon: Edward Arnold.

Grigg,D.B. (1965), *The Logic of Regional System Ann. Ass.* Amer. Geogr. 55 pp 465 - 91.

Grigg, D.B. (1967) "Regions, Modes, and Classes" in chorley, R.J. and Haggett, P. (eds) *Integrated models in Geography*. London: Methuen & Co. Ltd.

Harvey. D| (1969) *Explanation in Geography,* London: Edward Arnold.

Kendall, M.G. (1957) *A Course in Multivariate Analysis* (London).

Mahalanobis, P.C. (1927) Analysis of mixture in Bangal. *Fan. Ass. Bengal* 23 PP 301 - 333

Mahalanobis, P.C (1936) "On the generalised distance in Statistics". *Proc. Nat. Inst.Sc. (India)* 12. P. 49.

Mather, (1976) *Computation Methods in Physical Geography*. London: Methuen & Co Ltd.

Rao, C.R. (1948) "The Utilization of multiple Measurement in problems of biological classification." *FI.R. Statistics, Soc; Ser. P.P.* 51-203.

Rao, C. R. (1952) *Advance Sttistical Methods in Biometric Research* (New York).

Stebbling, L.S. (1961) *A modern elementary logic.* (London).

ADDRESS OF THE AUTHOR

Temi. E. Ologunorisa
Department of Geography,
Faculty of Social Sciences,
University of Port Harcourt,
Port Harcourt, Rivers State,
Nigeria.