Does geography have many solutions for one research problem? A case of distribution pattern

Aung Kyaw

Abstract

Although there are many definitions of geography developed through out its history, geographical studies constantly emphasize on the space. Density, dispersion and pattern are three major methods use in analyzing the distribution of geographical phenomena on the space. Geographers trace the spatial *pattern* that occurring on the earth surface by using above methods and find out the major *processes* that create the observing pattern. By using these major processes researcher could *predict* the future distribution patterns of geographical feature. Therefore, distribution analysis is a vital tool in geographical studies. The methods traditionally used in geography for analyzing the distributions were not properly considered the spatial nature. This paper tried to put forward the weakness of these traditional distribution methods and highlighted some methods those are spatially more relevant. The results reveal that there are many possible solutions that could be resulted from the different distribution analysis method of same problem and area. Each result, however, has its own proper explanations and correct to the certain extent. The most relevant answer, however, could be derived only if spatially considered method is employed systematically.

Key words: Geography, distribution pattern, controlling process, spatial consideration

Introduction

The origin of geographic discipline started from the curiosity of people about their environment. People recognized the variations of their environment on the earth surface. Therefore, early geographical studies were made by travelers. Then, scholars tried to invent the system that could be more precisely represented the variations on the earth surface. Precise latitudes and longitudes system was invented with the development of chronometer by John Harrison in 1740 (Claval, 1998). With the accurate determination of graticules and scientific mapping of the earth, place information became more reliable.

In the early time, geographers discussed about the pattern and process of climate, vegetation and landforms of the earth. Later, they considered variation of human and its activities on the earth surface. Description of space became easier with the development of formal administration unit in the 18th Century. Introduction of statistics provided geographers with new quantitative information of demographic and economic data. National census, trade statistics and ethnographic studies made human geographic studies more precise in 19th Century (Claval, 1998).

There are many definitions of geography. Fellmann, et al. (1990, P2), for example, defined geography as "the study of spatial variation of how— and why– things differ from place to place on the surface of the earth". According to Rubenstein (2003, P3), "Geography is the study of where things are found on Earth surface and the reason for the location." He added two major questions related to geographic interest: where and why. Where are the studied activities found? Why are they found there?

Associate Professor, Dr, Department of Geography, Taungoo University

Whatever the definition throughout its history, geographical studies always emphasize on distribution of natural feature and human activities on the earth surface before explaining the causes that make the distribution. However, some traditional methods using in the analysis of distribution pattern did not match with spatial characteristics. Therefore, this paper tried to discuss the major approaches of geography by analyzing previous studies to confirm the analysis of spatial distribution as a major way of scientific geographical methods to gain applicable results. Then, nature of geographical distribution analysis mainly used methods are reviewed their validity.

Geography as Pattern and Process

Although the terminologies of 'geography' is a little bit differ among scholars, emphasize is similar. By reviewing the way of geographical studies we could clarify the approaches of discipline. There are three major sequential steps in geography. First, geographers study the spatial variation ***(pattern)*** of geographical features (specific problem) that are under consideration. These features could be urban expansion, development of retail shops, cropping pattern changes, landform variation, natural vegetation cover, soil types, climate types or any geographic investigation occurring on the earth surface.

After confirming these variations, geographer tried to find out the underlying forces that caused the above patterns in the second stage. First stage and second stage are not reversible, because first step could be confirmed by means of precise measurements on the earth surface. Only after confirming the spatial pattern one could continue with the ***processes*** that generate above spatial pattern by relating other geographic information. The processes (reasons) that caused the spatial pattern generally find out from the many pre-assumed variables which are sometimes difficult to measure precisely. These pre-assumed variables are referred as 'hypotheses' in scientific term. Therefore, geographers should start with precise patterns to be able to find out accurate answer (processes). Statistical analysis usually considers pattern as a dependent variable and processes as independent variables.

Why do geographers study pattern and process?

It is an important question. All scientific discipline should have its implication to the benefit of human being directly or indirectly. This implication is based on the predictability of the future. Geographer needs to conduct above two steps to find out the underlying factors (processes) that creating the present pattern. If one could find out these underlying factors precisely, she/he could control the future patterns of investigating features or events by means of favouring or oppressing the controlling processes. In the study of retail development, for example, one could find out the spatial variation and changes of retail shop locations as pattern. Then, the major processes generated current distribution pattern of retail shops are find out related to other factors such as urban structure, land use policy, nature of economy and competition, etc. By analyzing these variables, the major controlling factors of newly open and closed retail shops could be found out. If proper statistical techniques are applied the extent of each major controlling factor on the development of retail shop could also find out. Then, these finding could be use in urban planning to control the development of retail shop in preferring place and direction by means of favouring and restricting underlying variables. This is the third stage that geographers should do. Although actual application of finding to the policy and planning is beyond the scholar, geographer should carried out up to the stage of giving suggestion or prediction for future prospect of investigated topic based on their scientific research findings.

Geography as a science has its systematic investigation methods based on cause and effect. In medical science, physicians enquire the feeling of a patient and examine the external appearance of patient (body temperature, colour of skin and eye, etc.) at first step. In this step physician could use thermometer, stethoscope, and sphygmomanometer as the tools for examination. They will also use internal measurement machines (CT (computerized tomography) scans, X-Ray, etc.) to able to predict the exact situation of the problem. All these investigations are examining and measuring the appearances (effects or patterns) of problem (disease). By using results of above observations and professional knowledge of physician (learn personally and from literature) one will predict the causes (process) of disease (heart or lung or lever or . . . etc.). After that physician would try to cure the disease by giving proper treatments. In that case he/she try to stop the causes (process) of disease that generate the external and internal appearances (pattern) of the patient. Only when physician could predict the source of disease by using patterns he/she could effectively and quickly cure the patient.

A geographer investigates earth surface (instead of patient in case of physician) based on the particular problem (urban expansion, flood hazard, environmental pollution, etc.) by using cartography, statistics, Geographic Information Systems, field methods, etc. as the tools. Direct observation data and census data are also used to verify the pattern. Then, they tried to find out the process (causes) that could be the main reasons for the generation of current pattern (effects) by using scholar's existing knowledge. Although the resulted outcome do not use for the curing of a particular patient, geographical finding are applicable to solve the social problems and could be used for planning (applied research) or for knowledge building (basic research) of future generation.

Does geography start explain from process and proceed to pattern?

If we have many theories and hypotheses that are necessary to test for their verification or consistency, it is necessary to find out the places that have similar process under debating. Then, observation is carried out based on the problem that "whether similar processes produce same outcome or not". In that case we consider process first then, examine the outcome. The result derived from the above research rather lead to theory and model building. In addition, it is necessary to examine the wider area of field and literature before building new hypothesis or finding weaknesses of existing literatures of specific field. These kinds of researches (process to pattern) are known as basic research while formerly explained pattern to process researches are considered as applied research. In developing countries like Myanmar, many geographic questions are related to social research nature and literature resources are limited to the certain extent. Thus, it is more appropriate to start from the pattern and find out the process for further applications.

In general, applied geographer should start with pattern and continue to the process to be able to predict the future precisely. Pattern could be observed by asking some questions: Where is it located? How is it located? Process on the other hand, was generally asked by 'why' (why is it located there?). It is also possible to ask the process with 'what'. For example, "what are the major forces that caused observed patterns?" Thus, the starting word of a question is not important but the meaning of question.

Spatial Distribution: Many Methods and Many Patterns

Rubenstein (2003) pointed out five distinguished ways of geographic thinking about the world: space, place, region, scale, connection. All these ways of geographical looking on the observing features are important to understand the pattern and processes. Space could be

considered as an arrangement of natural features and human activities over the earth surface. We could measure spatial variations of a particular feature by means of distribution, density, concentration and pattern. Place is used to distinguish the uniqueness of observed items by means of location, place name, site and situation. Region is basic classification system for spatial patterning of observed item by means of formal, functional, and vernacular regions. Scale is also important as a basic unit of analysis because analysis conducted at different scale could result varied out put. In addition, geographer should also consider the connection that could occur across the space.

In all five ways of geographical looking, it is necessary to define a study area before actual observation is conducted. Then, it is necessary to make classification in all measurement of space. Distribution, for example, has to examine by drawing regions that based on the designated class. Density, concentration and pattern all have to base on the classes. In that case, there are many ways (will discuss detail in late section) to classify derived data. Place also has to be generalized based on their location (e.g. between latitude 10 degree north and south) and place names of same categories (for example, city names derived from the natural features such as Taunggyi, Inlay, etc.). Region also formed based on the criteria of common features (cotton region). In both scale and connection, it is also necessary to consider "which level should we consider for the specified problem?" Although it depends on the nature of problem, there are variations in defining the area. The trade connection of a city, for example, could be reach up to international level. But it is also difficult to study the whole trade connection network. Then, where should we limit the boundary? This limitation should be logical rather than individual preference.

Space or arrangement of the objects is represented by distribution. Distribution of an observed pattern could measure by means of density, concentration and patterns. There are many ways to measure density based on the purpose of measurements. Some use arithmetic density (the total number of object in an area unit) while others used physiological density (the number of persons per unit area suitable for agriculture) or agriculture density (number of farmers per unit area of farmland) (Rubenstein, 2003).

Concentration differed from density and measures the clustering or dispersing of objects in aerial term. Pattern represents the geographic arrangement of objects in space. In all three descriptions, there should have standardized classification based on the studying theme(s). All geographic data are represented by means of area, line, and points. All geographical thinking includes spatial grouping or classification of derived data one way or another since it is impossible to trace the underlying processes without generalizing the observed pattern. Each of above measurement could be represented by the point, line and area (polygon) in cartographic techniques.

Nature of geographical data

From the nature of geographical research both pattern and processes are derived from space of the earth. There are three main forms in deriving data: point, line, and area.

Point and area data are interchangeable based on the scale. A city could be area data for the city level analysis while it is a point at the national or regional level analysis. Point data of observing item could be derived from census (such as population of a town) or field observation (soil type of study sample point or number of cars passing through the particular node daily).

Line data are categorized based on their designated level (e.g. 1st order stream) or field observation based data (e.g. average amount of commodity flow in particular segment of road network).

In the case of area data generalization problem could be encountered. Census data, for example, is defined to the location where settlement is located and sometime it is not coincide with the available geographic information. Census data are collected for general purpose and it is important to consider the compatibility of it with observing theme. In above situation, field observation becomes important. Again, validity of field data is relied on the using field techniques and skillfulness of the researcher. It is important to get correct and accurate data because the whole geographic analysis process relied on the accuracy of data.

Methods of distribution analysis

Although derived data are accurate and reasonable, there is still a problem in the ways of analyzing their distribution pattern. Both statistical model and simple qualitative methods could be used in the analysis. Whatever the analysis is it is necessary to categorize the pattern or/and processes to get required findings. With modern techniques of statistics and computer calculation, it is possible to correlate each case of dependent variable with other independent variables. In multiple-regression, for example, each case that could be considered as pattern is considered as dependent variable by using each or relevant independent variables. This complex statistical analysis is beyond the scope of this research and it will only focused on the simple spatial based quantitative techniques.

There are three major methods of classification derived from different sources. First method is traditionally used by many geographers. It classifies the data into *equal interval.* Second method derived from some statistical techniques. *Standard deviation, natural break* are involved in this method. The third source is related to GIS applications. *Equal area* classification method, for example, is derived from this method. All these classification methods are readily available in modern Geographic Information Systems software. It is worthwhile to review a brief account of these classification techniques before analyzing advantages and disadvantages of each classification.

The equal interval method divides the range of attribute values into equal sized sub-ranges. Then the features are classified based on those sub-ranges. This method is mainly used by traditional geographer. If frequency data values are not evenly distributed 'over generalization' problem encounter in this method. It means that designated ranks do not effectively represented included data.

Standard deviations method finds the mean value and then places class breaks above and below the mean at intervals of either 1/4, 1/2, or 1 standard deviation until all the data values are contained within the classes. This method of classification is based on the deviation of individual values from the average contain in the data set.

In the quartile classification method, each class contains the same number of features. Quartile classes are perhaps the easiest to understand, but they can be misleading. Population counts (as opposed to density or percentage), for example, are usually not suitable for quartile classification because only a few places are highly populated. One can overcome this distortion by increasing the number of classes. Imagine the difference, for example, if five classes are used in the chart instead of three. Quartiles are best suited for data that is linearly distributed; in other words, data that does not have disproportionate numbers of features with similar values. Nevertheless, this method also bases on the distribution of individual data contain in the data set.

Natural breaks classification identifies breakpoints between classes using a statistical formula (Jenk's optimization). This method is rather complex, but basically the Jenk's method minimizes the sum of the variance within each of the classes. Natural breaks find groupings and patterns inherent in your data (Andy, 1999).

The equal area method classifies polygon features by finding breakpoints so that the total area of the polygons in each class is the approximately the same. In GIS software the area of each aerial unit was automatically extracted and used in the calculation. Classes determined with the equal area method are typically very similar to quartile classes when the sizes of all the features are roughly the same. Equal area will differ from quartile if the features are of vastly different areas.

Table (1) Advantages and disadvantages of major classification methods

Method	Advantages	Disadvantages
Natural Break	1. It is good for mapping features that are not evenly distributed, since it place clustered values in the same class.	1. Since the class ranges are specified to the individual dataset, it is difficult to compare the map to other map 2. Choosing the optimum number of classes is difficult, especially if the data is evenly distributed
Quartile	1. Comparing area that are roughly the same size 2. Mapping data in which the values are evenly distributed 3. Emphasizing the relative position of a feature among other features.	1. Features with close values may end up in different classes, especially if values cluster. This may exaggerate the differences between features. 2. If the areas vary greatly in size, a quartile classification can skew the patterns on the map.
Equal Interval	1. Good for presenting information to a non-technical audience 2. Mapping continuous data such as precipitation and temperature	1. If the data values are clustered rather than evenly distributed, there may be many features in one or two classes and some classes with no features.
Standard Deviation	1. Good for seeing which feature are above or below an average value 2. Displaying data that has many values around the mean, and few further from the mean (normal distribution)	1. The map doesn't show the actual values of the features, only how far their value is from the mean 2. Very high or low values (outliers) can skew the means so that most features will fall in the same class.

Source: Extract from Andy (1999).

Advantages and disadvantages of major classification methods are shown in table (1). Usefulness of all methods except from Natural Break has relied on the conditions that data included in the data set are evenly distributed or normal distribution pattern. These pre-requirements are somehow contradict with the spatial distribution nature of patterns. If it is assumed that soil type changes with underlying rock, there should be distinguished changes in characteristics of soil when underlying rock changes from one type to another. Minor changes in soil characteristics could be considered as a result of other minor factor changes rather than the underlying rock. As disadvantages, Natural Break method has weakness in comparing among the classifications of same area for different periods or same period for different area.

On the other hand, other methods have weaknesses in possibility of grouping not too vary values in different group or concentration of many values in one or two group.

Geography's first step analysis is to classify the observed patterns based on the space. Observed patterns are occurred as results of underlying processes. In other words, current pattern of observed objects are related to the one or more similar spatial processes. Geographers conduct spatial analysis based on above assumption. For example, yield per acre of paddy is related to many variables as mentioned before. It means that spatial distribution pattern of yield per acre is to certain extent similar to the one or more distribution pattern of soil fertility, slope, rainfall, etc. The generalization (classification) natures of data are not similar in all above variables.

If current observed pattern is generated by some spatially distributed underlying processes, there should have distinguished break between classified groups. In that spatial context, it is not necessary to consider the equal distribution of cases in classes. However, all above mentioned classification methods are based on the data set itself and classified based on the consistence of data set itself. They are not particularly reference on the spatial variation of space.

In above classification methods, equal interval, standard deviation, and quartile classification methods are considered only on the distribution of individual data within data set instead of spatial nature they are representing. In that case 'over generalization' and 'spatial negligence' error could be encountered. In case of natural break classification, it is based on the point variance of each group has least after classification. It means that differences within each category are very few and each class fully represents the data included in the data set. Thus, if we accept the assumption that "patterns of geographical features are generated by other geographical features those are spatially distributed and there should have a sharp break between the classes of observed geographical feature" natural break is the most appropriate method to classify the distribution patterns of geographical features.

Conclusion: Does geography have many answers?

Whatever the classification method is used, one spatial pattern will find out as a result. The resulted pattern, however, will be differed based on the method employed in study. Since one spatial pattern is result of many spatial processes, all resulted patterns from various classification methods should have their own explanations. If so the question 'Does geography have many solutions for one research problem?' will arise. From the above discussions, it is clear that 'Geography could have many answers'. But the level of correctness and accuracy of each answer is differed based on the way of analyzing spatial data and experience of researcher.

References

Andy, M., (1999), *The ESRI guide to GIS analysis: geographic pattern and relationship*, ESRI, Inc., Redlands, California.

Claval, P., (1998), *An introduction to regional geography*, Blackwell Publisher, Oxford.

Fellmann, J., Getis, A. and Getis, J., (1990), *Human geography: landscape of human activities*, Wm C. Brown Publishers, Dubuque, USA.

Rubenstein, J. M., (2003), *The cultural landscape: an introduction to human geography*, Pearson Education Inc., USA.