DOCUMENT RESUME

ED 077 799 SO 005 659

AUTHOR	Margulis, Harry L.
TITLE	Geographical Data Conceptualization: An Approach to Student Understanding.
PUB DATE	26 Oct 72
NOTE	12p.; Paper presented at the Annual Convention, National Council for Geographic Education, Milwaukee, October 26-28, 1972
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Concept Teaching; *Geographic Concepts; *Geography Instruction; Map Skills; Speeches; Thought Processes
IDENTIFIERS	*National Council for the Social Studies

ABSTRACT

Structuring geographical concepts so that they are intelligible to an uninitiated student is a recurrent problem in geography teaching. The subtle interrelationships between map distributions and analytical procedures are not intuitively clear, nor is it plainly evident that concise map analyses depend upon specific measures of distributions. This paper is concerned with the development of a thought process which introduces and leads students to a more thorough understanding of geographical phenomena, conceptualization, and generalization. Specifically, it is concerned with data presentation, analyses, and the generalization of the phenomenal content of earth space through the use of traditional geographical modes of thought. Thirty five millimeter slides of the graphics which accompany this paper are available to interested parties from The Department of Geography, Rutgers University, Newark, New Jersey 07102. (Author)

FILMED FROM BEST AVAILABLE COPY

ED 077799

NATIONAL COUNCIL FOR GEOGRAPHIC EDUCATION, ANNUAL CONVENTION, HOTEL SHERATON SCHROEDER, MILWAUKEE, WISCONSIN, OCTOBER 26-28, 1972

Submitted 4/12/72

TO: John O. Towler
Associate Professor of Education
School of Humanities, Social Science and Education
Purdue University, Lafayette, Indiana 47907

Dr. Harry L. Margulis
Department of Geography
Rutgers University
Newark, N. J. 07102

TITLE OF PAPER: GEOGRAPHICAL DATA CONCEPTUALIZATION: AN APPROACH TO STUDENT UNDERSTANDING

CURRICULUM LEVEL: Undergraduate freshman-sophmore
ADDITIONAL PERSONS INVOLVED: Dr. Raphael J. Caprio
TIME REQUIRED: 20 minutes

EQUIPMENT NEEDED: 35 mm slide projector

ABSTRACT. Structuring geographical concepts so that they are intelligible to an uninitiated student is a recurrent problem in geography teaching. The subtle interrelationships between map distributions and analytical procedures are not intuitively clear, nor is it plainly evident that concise map analyses depend upon specific measures of distributions. This paper is concerned with the development of a thought process which introduces and leads students to a more thorough understanding of geographical phenomena, conceptualization, and generalization. Specifically, it is concerned with data presentation, analyses, and the generalization of the phenomenal content of earth space through the use of traditional geographical modes of thought.

The character of geography derives not so much from its specific phenomena of study, but from its distinctive processes and problem-solving approaches. Embodied in this statement is a traditional way of looking at geographic reality which strives to answer the question: "What is where?" (Abler, Adams, and Gould, 1971). Precisely, given all locations what kinds of phenomena are found, what kinds of interrelationships exist, and how do phenomena covary spatially? Berry (Berry, 1964) presents a simple technique which partially structures these geographical questions. His description commences with the construction of a geographical matrix. The matrix, referred to as a geographic data file, consists of an orderly, intersecting arrangement of columns (places) and rows (characteristics). Each cell contains a geographic fact or a minute piece of reality. The manner in which these columns and rows are viewed provides a structure for geographical perspectives. A single row of cells recording a characteristic at a series of places leads to a study of spatial variations and their associative map distributions. Comparison of two or more rows conveys the study of spatial covariations and associations. Column arrangements are also of two types. When a single place (column) is compared for a series of characteristics, the arrangement is known as a locational inventory. When two or more columns are compared, this leads to a discussion of areal differentiation.

Berry's intention was not to create a pedogogical tool but rather to illustrate the types of "spatial questions" asked by the various schools of geography. His conceptual model, nonetheless, is an excellent pedogogical tool. It illustrates well what a geographer means by a "spatial perspective" and is a fine example of a conceptual model. Owning to their heuristic potential, conceptual models form a basic framework for teaching introductory geography courses. In the case of the geographic matrix, the framework provided is very detailed since not only is the disciplo defined, but also the links and aims of the diverse parts of the field are exposed.

Among the many characteristics of conceptual models, two are of particular importance - selectivity and distortion. All models are selective in respect to the components of reality stressed. Models are prototypes of reality which selectively ignore irrelevant facts so that reality becomes simplistic and more meaningful.

In order to appreciate the selective nature of a model, a student must come to recognize that selectivity distorts reality. The geographic matrix subsequently furnishes a relative devise for illustrating these principles. By manipulating one of two conceptual dimensions, we can change the emphasis of a problem by altering the questions asked, i.e., what and where. One may argue that in its original form the data matrix does not distort "real data": however, if the matrix is to be operationalized scale adjustments are required. Scale adjustments are obtainable by consideration of a submatrix possessing a finite number of rows and columns. The reduced scale of the data unit, whether at the scale of a household, census tract, county, or country, assumes some form of distortion.

Students readily recognize the need of a submatrix to satisfactorily constrain the scope of geographic reality and they appreciate the infinite variety of row and column arrays. While scale adjustments change the perspective of a problem by transforming corresponding areal associations, it is a manditory step for mapping spatial distributions at reduced topographic scales and for measuring spatial properties. We have found that from this point on, with proper presentation, a submatrix can be used with surprising success to introduce beginning students to the subleties of geographical conceptualization. The manner in which this is approached is illustrated as a flow diagram in Figure 1.

Location

In general terms, location implies a place or position in "space." The location of points in a scatter diagram is conceptually similar to a dot distribution of individuals in two-dimensional "geographic space." The particular difference in the former case is that each data unit is located with respect to two social variables (such as education and income), while in the latter case, each data unit is located with respect to two orthogonal compass directions (such as north-south and east-west). Thus arises the distinction between location in "social space" as compared to location in "geographic space."

Most beginning geography students can readily comprehend this difference and have seen or used scatter diagrams in other disciplines. What they are usually unaware of is that the location of 29 data units, for example, census tracts, in a two-dimensional social space can be viewed as a 29 x 2 sub-matrix. The scatter diagram also shows what a geographer means when he discusses apparent spatial patterns. Consider Figures 2A and 2B. Figure 2A shows a scatter diagram of 29 census tracts (points) and their location relative to two characteristics, income and median education. Most students, when asked to explain the scatter diagram pattern do so by saying that there is apparently a <u>relationship</u> between income and medium education. The causal factor, improved ability to compete in the labor market as approximately measured by years of formal education, is easily perceived. When presented with Figure 2B and told that the data units are houses located in geographic space, they again respond that some form of relationship is evident between the independent (X) and dependent (Y) variables. When asked to hypothesize about the causal relationships, it does not take long to elicit several possibilities.

Mapping

Most spatial phenomenon are reduceable to one of four abstract dimensions: points (non-dimensional), lines (one-dimensional), areas (two-dimensional), and volumes (three-dimensional). When these qualities are mapped, they become iconic or representational models. A map is an analytical tool which communicates information to the perceiver by abstracting significant portions of reality. Reality is scaled and simplified to achieve greater comprehensibility. Cartographic processes, the encoding and transmitting of messages through mapping, is a traditional extension of geographical conceptualization.

Two types of maps are particularly useful in furthering our understanding of geographical conceptualization. These are the choropleth and isoline maps, both classified as quantitative areal maps. Once location in geographical space is defined by a finite data matrix, the data can be mapped using one or the other of these techniques. However, an indispensable assumption of choropleth mapping is that the phenomena grouped by class intervals coincides with the areal unit. The isoline map, on the other hand, displays distributional phenomena by connecting points of equal value. Class intervals are independent of the areal units and it is assumed that an interpolative continuum exists between datum points. While it is desirable at times to reshape our data from areal to point, this procedure does violence to our data because of the absence of standardized collecting units. Since most data is gathered for areas of unequal size and shape, a data bias is built into any given geographical problem utilizing areal data. We compound this data bias by arbitarily moving from areal to point data. While crude adjustments can be made in converting data by weighting or regularizing collecting units, we find it more expedient to concentrate on the common conceptual bonds existing between quantitative maps and the geographic data matrix than on problems of incongruous scales. Being able to selectively change the nature of our scale perspective makes it possible to whow the interchangeability of the idiographic and integrative modes of geographical thought.

Figure 3 summarizes this step-wise transition from geographic matrix to quantitative mapping. This process implicitly shows how spatial relationships are linked to various scales of data abstraction. Every student is required to develop a working knowledge of this procedure and their understanding is enhanced through several exercises.

To provide a data base for these exercises, students are referred to the Census of Population and Housing (PHC-I) and instructed to obtain data for a number of selected census tracts. In our course, the cities of Orange and East Orange, New Jersey are used as case study areas because of relatively large social variations between their 29 census tracts, a total number of observations not impractical to work with at this introductory level. Five variables are chosen thereby yielding a 29 x 5 data matrix. Figure 3A illustrates a typical data listing. From this listing students are instructed to prepare a number of choropleth maps (See Figure 3B). To introduce the ideas of meaningful map intervals, standard scores and central tendency measures, mean and median, are discussed. These descriptive statistics assist us in clarifying three assumptions of choropleth mapping. First, local variations within census tracts are cancelled out by assuming that data distributions are equal. Second, distributions in choropleth mapping are considered discontinuous. Few distributions in reality terminate abruptly at specific boundaries; rather at the interface between distributions there is a gradual submergence of one variable and the emergence of a more dominant variable feature. Third, choropleth class intervals are not standardized, but they are specific to the geographical problem. Changing the intervals of the classes automatically alters the patterns displayed.

The next step in the exercise sequence is to superimpose an arbitrary cartesian coordinate system over the base map. Each census tract is located by an X and Y coordinate thus yielding the modified 29 by 7 data matrix illustrated in Figure 3D. Isoline maps, because they require a three-dimensional input, are constructed from 29 x 3 submatrices of the larger 29 x 7 modified matrix (See Figure 3E). In choropleth mapping, we can proceed from map to data matrix, but it is impossible to reverse this order because we fail to preserve the shape of the areal unit. However, without undue representational difficulties, we can alternate back and forth between the isoline map and the data matrix. Though substantial information concerning areal units is lost in isoline-data matrix transference, this sacrifice is more than compensated for by our ability to choose other bases for conceptual demonstration.

Because of idealization, dimensionality, and representation problems, maps are more limited in scope than are geographic matrices. Geographic matrices, unlike maps, can easily be changed to achieve different levels of abstraction. This same process carried out with mapping is much slower and added detail merely increases the difficulty of concise, quick interpretation. Maps are effective tools, nevertheless, for assembling a cumulative picture of an area of interest. In combination with applied statistics, greater freedom exists for more stringent problem testing and greater clarity of exposition.

Covariation, Generic Classification, and Areal Differentiation

Covariation is defined by both a nominal and operational definition. As nominally defined, covariation implies interrelatedness or intervariation among classes of geographical phenomena. Since phenomenal combinations rather than singular categories reveal the character of the earth surface, the manner in which these combinations change is the study of covariation. Geography, as a chorological science, considers areal content by describing covariations among phenomena, but this leads to inexactness in determining the strength of association between one or more classes of commonly occurring variables. An operational definition is therefore required which gives a more exact rule of association. This rule is derived from analytical statistics which refers to the expression $\sum_{i=1}^{N} (X_i - \bar{X})(Y_i - \bar{Y})$ as the covariation of X and Y. This expression provides a measure of how X and Y vary together and, if divided by N, the number of observations, the covariance between X and Y is determined. The sum of the products is positive, X is related to Y, when $X > \bar{X}$, and $Y > \bar{Y}$, or when $X < \bar{X}$ and $Y < \bar{Y}$. If $X > \bar{X}$ and $Y < \bar{Y}$, then the resulting relationship between the sum of the products is negative. The variables are independent of each other when the sum of the products is zero. Suspected covariant relationships can thus be submitted to analyses using several types of statistical tests. By simple shifts in data scales from nominal to interval and by variations in measurement techniques, the covariation between variables can be explored for amount and direction of change.

Once covariant relationships are determined, the desire to reduce the complexity of phenomena is of overriding importance. This development takes the form of a search for generic relationships. It involves categorization, classification, and regionalization of specific phenomena. Again, statistical tools are employed, but solely as a means of grouping data, a procedure which is analogous to scale reduction in mapping. In the past, grouping was accomplished through the use of subjective empirical rules of procedure, but this method is open to criticism and methodological debate. Statistical grouping has stricter requirements and, while the choice of techniqueapplied is open to debate, the rules of manipulation are prescribed and less alterable.

In our flow diagram, Figure I, areal differentiation follows from generic classification. The placement of this concept is more a convenience than an absolute. Areal differentiation and covariation are merely different ways of viewing column and row structures in a geographic matrix. The techniques used to explain variable co-variations are just as applicable to a study of areal differentiation. Only the manner in which the data is handled and the ways in which the results are interpreted are substantially different. For example, in normal computer card listing, geographical variables are key punched in the columns of cards. The cards are usually coded to identify the places where the variable attributes occur. It is equally feasible for each card to represent a variable and for columns to represent places.

Since each student has manipulated a relatively small data matrix (29 x 5) in several map exercises, students can be introduced to simple computer procedures with little difficulty. Key punching and machine control cards are discussed in order to facilitate a student's use of a number of "canned" library programs. Specifically, two programs, Stat-I and Numerical Taxonomy, are utilized to deal with the problem of spatial variation, co-variation, and generic classification. Using 29 x 5 data matrix input, the Stat-I program computes the means, standard deviations, and correlation matrix for the five variables. Using z-scores to define map intervals, students are instructed to map both related and unrelated variables. The spatial distribution of a single variable, as well as the spatial co-variation of related variables, are noted. The correlation matrix is analyzed with an eye toward the elimination of redundancies in the data. Since highly related variables often "explain" a similar aspect of reality, students are introduced to the notion of a parsimonious choice. By examining the correlation matrix, students are required to reduce the number of variables by eliminating any which are highly correlated with another. Having reduced the matrix from 29 x 5 to 29 x 3 or 29 x 4, the remaining unrelated variables are used as input to a numerical taxonomy. The basis of this generic classification program, is that those cases most similar should be grouped first. Most similar is defined by the distance separating any two cases in "taxonomic space". Based on the pythagorean theorem, the technique is less sophisticated than factor analysis or other clustering techniques. This not only makes it easier to work with, but is also easier to present from a pedigogical viewpoint. In constructing the "taxonomic tree", a student can appreciate the increased generality of grouping, as well as the concomitant problems of empirically derived "regions". In this series of exercises emphasis is placed on understanding the input and output of the computer. For all intents and purposes, the computer is treated as a "black box", but some elementary explanations of computer programming are discussed. This, as we have found, does not distract from the purpose of explaining geographical conceptualization. In fact, it adds appreciably to the general tone of presentation by reinforcing the ideas of logic and flow.

The placement of areal differentiation following generic classification is significant in that grouped categories of data can be mapped. Maps are ideally suited for this purpose because of the following reasons (Thomas, 1968). First, maps are traditional devices for finding typical and atypical cases among geographical distributions. Second, patterning provides a visual means for modifying hypotheses. Third, abstracting regional boundaries organizes space in a meaningful manner so that data may be regrouped for continued analysis. Reverting back to maps therefore enables us to derive new and possibly original insights. Only then is it possible to speculate about genetic processes and causation.

Scatter Diagrams - Further Uses

Geographic distributions are never immutable, but forever changing. The search for new methods of geographical inquiry are therefore continuous and unending. In developing other generalizing techniques for concise measurement of distributions, attention is once again focused upon scatter diagrams. Scatter diagrams, besides permitting us to examine the clustering of variables, also have distributional properties which offer additional prospects for geographical conceptualization. Among these properties are the shape of the distribution, the spacing - regular or irregular - of the points in the distribution, the mean center, and the dispersion of the distribution around the mean center.

In a recent paper prepared by the authors,[3] a computer program for measuring mean center and dispersion properties is discussed. The procedures, known as the ellipse and modified sectogram methods, are used to point out the types of questions geographers ask relative to summarizing spatial distributions. Other summary distribution measures such as nearest neighbor map analysis and some recent work on shapes of distributions are presently under review for incorporation into the course.

REFERENCES CITED

Abler, Ronald, Adams, John S., and Gould, Peter, Spatial Organization (Englewood Cliffs, N.J.: Pr[illegible]-Hall, Inc., 1971).

Berry, Brian J. L., "Approaches To Regional Analysis: A Synthesis," Annals of the Association of American Geographers, 54 (1964) 2-11.

Thomas, Edwin N., "Maps of Residuals From Regression," in Brian J. L. Berry and Duane F. Marble, Spatial Analysis (Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1968) 326-352.