
Mark S. Monmonier

***Maximum-Difference Barriers: An Alternative
Numerical Regionalization Method****

Regionalization is the primary classification problem in geography, although other typologies are sometimes demanded by specific research endeavors. Groupings of area units can produce either contiguous or fragmented patterns. Discontiguous regionalizations may have the advantage of placing truly alike areal units in the same category and are obviously necessary when the similarity of distant places is sought. The same number of contiguous regions, on the other hand, will most likely produce a more regular map pattern, thereby facilitating the transferral of the printed map into a more coherent and more lasting mental image.¹ Furthermore, many problems, particularly those partitioning space for administrative purposes, demand contiguity. Although some within-group homogeneity is often lost by imposing contiguity, this is a difficulty only for situations with relatively low spatial consistency.² This loss of homogeneity may well be offset by the perceptual advantage of simplicity.

Previous quantitative approaches to regional clustering have largely achieved contiguity by prohibiting linkages from occurring unless the two places abut. This contiguity restriction has been employed most often with hierarchical grouping procedures in which the clustering of N places proceeds through $N - 1$ levels of classification to the ultimate aggregation of all units into a single group.³ At

* The author is grateful to the Research Foundation of the State University of New York for financial support and to Michael Dobson, Peter Gould, and Anthony Williams for helpful suggestions.

¹ For a discussion of the perceptual advantages of regular map patterns, see Monmonier [9].

² Spence [12] discusses this problem and gives an example in which the loss of homogeneity for M regions of N counties decreases as M approaches both 1 and N .

³ Two of the earliest users of this approach were Ray and Berry [11].

Mark S. Monmonier is associate professor of geography at Syracuse University.

each level similarity is measured between all possible pairs of Operational Taxonomic Units (OTUs). The most similar pair is combined into a single OTU before proceeding to the next level. A contiguity constraint permits this test of similarity for adjoining districts only, thus assuring contiguous groups of places at all levels of clustering. Other clustering strategies, such as Q-mode factor analysis and multiple discriminant analysis, which do not group pairs of OTUs at independent stages in a hierarchy, are less amenable to a contiguity restriction.⁴

Alternatives to the use of contiguity as the sole geographic constraint have been suggested by Bunge [1] and Taylor [14]. The former argued that location should be used together with the other discriminant variables in computing similarities. However, if location is measured solely in latitude-longitude or other customary geographic coordinates, this is not sufficient to insure contiguous regions. Taylor clustered enumeration districts in Liverpool around nodes selected for their representativeness of other OTUs within a given distance. Yet even here contiguity played a role both in computing places' scores on his "typicality scale" and in assigning OTUs to specific "typical nodes."

The classification problem has been approached differently by Edwards and Cavalli-Sforza [3], who, instead of forming a grouping hierarchy that starts with N OTUs and ends with a single class, begin with a single region that is progressively split into two and then successively larger numbers of groups. All possible partitions of the population of OTUs into two groups are tested for within-group homogeneity by the analysis of variance and the best one accepted. These two groups are in turn split further into four clusters by testing all possible divisions within each. Spence and Taylor [13] suggest that a contiguity constraint could be applied but are dismayed by the computational effort required. Although the goal of this grouping strategy is minimum within-group variation and maximum between-group variation, these optimum conditions are not achieved with certainty unless all potential divisions of the areas are obtained and tested. Even with today's third-generation computers the cost of this optimal division of a large number of OTUs is considerable. Instead of seeking the global optimum, one is usually better off accepting some satisfactory local optimum in which within-group homogeneity, although not necessarily at its peak value, is nevertheless not sufficiently below the peak to impair seriously the efficiency of the classification.

All of the above procedures are based on measures of the *similarity* between OTUs. This is perhaps understandable in view of the objective of maximum within-group similarity. When similarity-biased regionalizations are mapped, however, the map reader sees not only the area-shading or alphameric symbols used to distinguish the regions; he also is confronted by the boundaries between these regions. How a person interprets a map is far from being clearly understood, but it is believed that the map reader attaches some significance to these boundary lines themselves.⁵

⁴ For discussions of Q-mode factor analysis and discriminant analysis approaches to regionalization, see Hautamaki [4] and Casetti [2].

⁵ Jenks and Caspall [5, p. 218] suggest that map readers attempt to extract from a map

Jenks and Caspall [5, pp. 229–31], in attempting to optimize class-interval selection for choroplethic maps of univariate distributions, recognized that one possible approach was to force as many as possible of the more significant “cliffs” on a three-dimensional data model into the mapped boundaries between classes. They further noted that this objective of minimizing error along map boundaries conflicts with the opposing goal of creating homogeneous classes. In fact, when only the reduction of boundary error is considered, the ranges of the resulting classes tend to overlap. Although this creates an impossible situation for the cartographer attempting to display a single variable, it poses no such conceptual difficulty in multivariate regionalization problems. The procedure outlined in this paper proposes a solution to the regionalization question based on determining the most significant boundary lines through a collection of conterminous areal units. The result is only another locally optimum solution to the minimum within-group variation problems, yet this placing of the primary emphasis on differences across boundaries not only insures contiguous regions, but also provides an interesting alternative against which other contiguity constraint strategies might be compared.⁶ Furthermore, although many of cartography’s psychophysical questions remain unanswered, this study does, as McCleary [7] suggests, propose a solution that might achieve greater significance once the interaction between the map and its user is better understood.

A BASIS FOR BOUNDARY DELIMITATION

If boundaries are to be the basis for regional division, the first requirement is an index of boundary significance. One obvious criterion is the difference in value between the two adjacent areas divided by the boundary. When the regionalization is based on a single variable X , the difference D_{ij} between adjoining OTUs i and j may be expressed as

$$D_{ij} = |X_i - X_j|.$$

The use of absolute values recognizes that a boundary’s relevance is not affected by the sign of the difference. For the digital computer the computation time can be reduced slightly by taking the square of the difference.

$$D_{ij} = (X_i - X_j)^2$$

one or a combination of three facets of the distribution: the overview of general trends, the data values for specific places, and the boundary lines between patterns. These authors hold that these boundaries are compared with mental images of other aspects of the area mapped.

⁶ Another project concerned with the significance of class boundaries is reported on p. 65 in Vol. 4, No. 1 of *Area*. In the Symposium on Cartographic Automation and Geographic Analysis, Richard Webster presented a paper on a method for recognizing the boundaries with the steepest gradients of variation between soil classes. Unlike that presented here, Webster’s approach is concerned with scanning soil property data along a one-dimensional transect.

will result in different values of D , but the rank order of the boundaries compared is not destroyed. This formulation can be recognized as a special case of the formula for m -variate taxonomic distance squared

$$D_{ij}^2 = \sum_{k=1}^m (X_{ik} - X_{jk})^2,$$

where m is the number of and k is the subscript for the dimensions on which dissimilarity is based. These computations assume, of course, that the values of X are uniform within the OTU. Yet, the possible absence of complete homogeneity throughout each of the fundamental areal units being classified cannot be labeled a deficiency unique to the technique proposed here since OTU uniformity influences the result of *any* classification method. Nevertheless, the individual researcher must evaluate the appropriateness of the level of aggregation of his OTUs for his particular research objective.

Given the strengths of all boundaries between pairs of adjacent OTUs, a technique is needed for consistently selecting a string of boundaries to form a barrier between higher-order regions. Two considerations are important. First, all boundaries forming the barrier must be linked together and must either form a closed loop or have both ends terminate at the edge of the study area or against another barrier. Second, the most important barrier should contain the boundary with the steepest gradient. The latter is not an absolute requirement since the aggregates of all boundary distances across the barrier could also provide the yardstick for barrier demarkation. Yet, to avoid the necessity of examining all possible barriers, the heuristic search procedure suggested here uses the boundary with maximum dissimilarity as a logical starting point and begins to extend the barrier by adding at each end the boundary across which the dissimilarity between adjacent areal units is greatest. Search proceeds one step at a time away from each end of the initial boundary with the steepest gradient and concludes when both ends join the outside boundary or the barrier itself. A second barrier can then be erected to partition one of the first two regions into two sub-divisions. The process can be repeated until a set number of divisions is reached or the ratio of within-group variation to between-group rises to an unacceptable level. This barrier extension approach is illustrated for the North Central States (Figure 1). A single variable, percentage of population increase between 1960 and 1970, is used in this example. The differences between all contiguous states are shown in Table 1. The boundary with the steepest gradient (13.8) is between Minnesota and North Dakota. The barrier it initiates is not extended farther north since its northern node is on the edge of the study area. Its southern node provides two choices, the North Dakota-South Dakota boundary with only a 0.2 difference and the Minnesota-South Dakota border whose gradient of 13.6 makes it the next addition. At the next node, because Iowa is more different from Minnesota (9.1) than from South Dakota (4.5), the regional boundary turns east. The barrier finally meets the edge between Kansas and Missouri and terminates. It is interesting to note that the five greatest across-boundary dis-

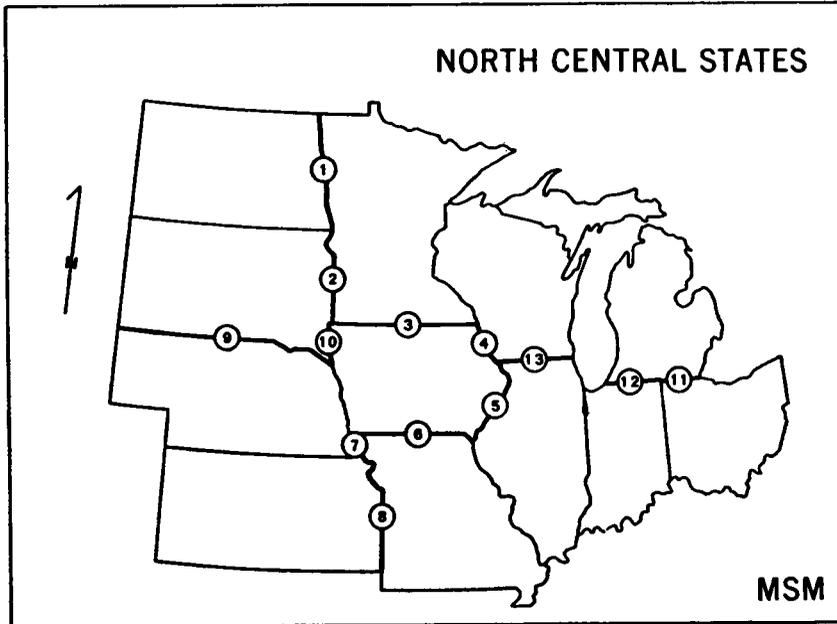


FIG. 1. North Central States Barriers Based on Percentage Population Increase, 1960-70. Numbers indicate sequence in which boundaries joined barriers.

tances have been incorporated in this first regional partition. The greatest remaining boundary gradient is between Nebraska and South Dakota. Adding their most important diverging boundary partitions, the Dakotas form the southern members of the western region. Similarly, a division initiated by the Michigan-Ohio difference (3.7) extends westward to form northern and southern subdivisions in the eastern part. This barrier, like the previous one, ends when it encounters the main north-south barrier.

TABLE 1
VALUE DIFFERENCES BETWEEN CONTIGUOUS STATES

Iowa	—												
Illinois	7.8	—											
Indiana	—	1.2	—										
Kansas	—	—	—	—									
Michigan	—	—	3.2	—	—								
Minnesota	9.1	—	—	—	—	—							
Missouri	5.9	1.9	—	5.1	—	—	—						
Nebraska	2.7	—	—	1.9	—	—	3.2	—					
North Dakota	—	—	—	—	—	13.8	—	—	—				
Ohio	—	—	1.7	—	3.7	—	—	—	—	—			
South Dakota	4.5	—	—	—	—	13.6	—	7.2	0.2	—	—		
Wisconsin	9.4	1.6	—	—	1.6	0.3	—	—	—	—	—	—	—
	IO	IL	IN	KS	MI	MN	MO	NB	ND	OH	SD	WS	

COMPUTATIONAL PROCEDURES

In order to be practical for large sets of data, the maximum-distance boundary extension technique, like most heuristic methods, must be defined in a series of logical steps that can be performed by a digital computer. Although the preceding example appears simple in concept, the general algorithm is necessarily more complicated to allow for a variety of complications.⁷ The discussion that follows requires as input a data matrix X , from which taxonomic distances are computed, and a contiguity matrix C that specifies the linkages among areal units and boundaries. In addition to the usual values of C_{ij} , which are ones when OTUs i and j share a common boundary and zeros otherwise, C_{ij} is set to two when i and j touch only at a point rather than along a line. Point and linear linkages must be differentiated and, as will be evident later, both types of contiguity must be accessible to enable detection of all boundaries meeting at a node. Contiguity to the edge of the study area must also be specified in an additional row and column of C to permit barrier penetration to and termination at the outside perimeter. The procedure outlined here assumes that no OTU has an outlier and that the entire region has no enclaves that do not participate in the clustering. These are not severe limitations, since outliers can be designated separate OTUs receiving the same data values as their main part, and interior holes can be partitioned among all neighbors. This latter division merely involves tying together at a point all boundaries that encounter the enclave so that any barrier that enters the region will be able to leave along a different diverging boundary. The enclave or lake can be shown on the finished map but the contiguity of the regional divisions will remain unaffected.⁸

Four general steps can be outlined.

1. *Compute the taxonomic distances between all OTUs contiguous along a line. Choose the two OTUs with the greatest difference to specify the starting boundary for the barrier.*
2. *Follow the barrier to the left until either the edge or a barrier is encountered.*
3. *Follow the barrier to the right until either the edge or a barrier is encountered.*
4. *Return to Step 1 unless some criterion (for instance, the within-group sum of squares) indicates that regional division has progressed far enough. An additional step must, however, be inserted to allow for the possibility that, of the remaining boundaries diverging from a node a barrier has entered, more than one might be tied for the greatest difference. When this condition is sensed only one of the tied boundaries is followed to the barrier's conclusion and the others are placed in a queue that is processed at the completion of step 3. Thus,*

⁷ A listing of the computer program MAXDIST is available from the author.

⁸ In the North Central States example, Lake Michigan can be divided between Illinois, Indiana, Michigan, and Wisconsin. These states would then meet at a node somewhere near the southern end of the lake. In this event, after the barrier enters the node along the Indiana-Michigan border, both the Illinois-Wisconsin and the Michigan-Wisconsin boundaries are tied for the maximum difference. The computer program would thus split the barrier to include the Michigan-Wisconsin segment and form five regions instead of four.

3a. *Follow all diverging boundary ties in the queue until they encounter either the edge or another barrier.* Another queue is needed since more than one boundary might be tied for the maximum distance computed in Step 1. Whereas it might be argued that this tied starting boundary would be picked up on a subsequent reiteration of Step 1, this second queue is needed to prevent possible termination of the division process in Step 4. The queue guards against the possibility of not forming a barrier when others with an equal starting dissimilarity are included. Thus,

3b. *Repeat Steps 1 to 3a for all tied initial boundaries.*

As observed in the example for the North Central states, the boundary with the second greatest difference may have been included in the barrier initiated by the boundary with the maximum gradient. To prevent redundancies a second contiguity matrix K (which in the computer program can be the lower half of C) is required for the storage of *current* contiguities. Whenever a boundary $i-j$ is added to a barrier, K_{ij} is set to zero. K can then be inspected so that existing barriers are not retraced.

Identifying Nodes

To keep input simple this technique requires only contiguity and data matrices. If the nodes from which boundaries diverge had to be detected manually, the method would be made excessively laborious for the average user. Yet these nodes must be referenced in terms of their diverging boundaries so that all alternate exits for an advancing barrier can be tested. On the assumption that most users will not require the incorporation of all boundaries into barriers and thus make every OTU a region, the present approach does not first attempt to identify all nodes in the boundary network. Instead, only those nodes specified by a single boundary are accessed as needed.

If every pair of adjoining areas is separated by a single continuous line, the problem is greatly simplified. The possibility remains, however, that two adjacent places may meet along more than one boundary segment (see places i and j in Figure 2). Yet, if they are to be partitioned, the entire collection of separated boundary segments must be added to the barrier. Hence, they can still be represented by a single pair of cells in a symmetric contiguity matrix.⁹ Moreover, the barrier must now be extended between more than one pair of nodes in order to remain a continuous line. In the case illustrated by Figure 2, a barrier between i and j and entering at node A must leave not only at node B but must also select a path through nodes C, C', D, E, F, and G. Of the eight nodes labeled in the figure, however, only seven are independent. Node C' is not independent of node C since the path chosen at C will automatically specify the path selected at C'. Nodes D and E, on the other hand, are independent since the diverging route chosen from one does not provide a direct path immediately converging on the other. Similarly, F and G are independent because the barrier can choose a diverging route from F and not be linked immediately with G. Thus, it is

⁹ It is assumed that no pair of places will be contiguous at both a line and a point.

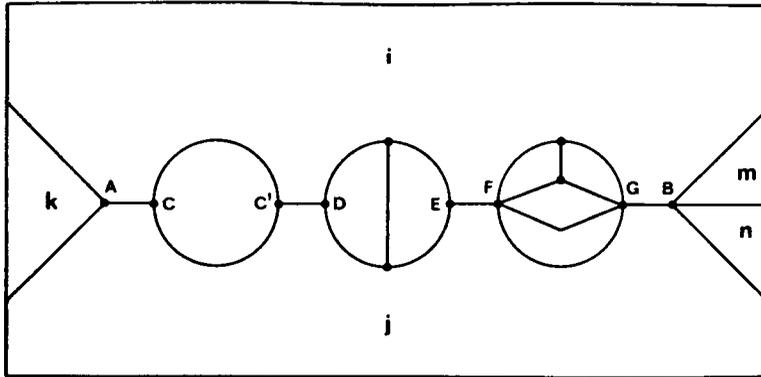


FIG. 2. Example of More Than Two Independent Nodes Along a Boundary.

necessary to identify all independent nodes along boundary $i-j$ and add each of these nodes to the barrier until the partition between i and j is complete.

A node can be represented as a list of all areas that are mutually contiguous to each other either across a boundary entering the node or merely across the nodal point itself. Those boundaries separating the linearly contiguous areas are the only boundaries diverging from the node. For example, node A in Figure 2, the simplest kind of node possible, can be recognized because i is contiguous to both j and k , j is contiguous to both i and k , and k is contiguous to both i and j . Node B, on the other hand, involves point contiguities (i and n ; j and m) but, here again, i , j , m , and n are all mutually contiguous.

An automated procedure for detecting nodes is surprisingly simple. Given the boundary $i-j$, find first all other areas contiguous to both i and j either at a point or along a line. *Restrict further checks to just these areas.* Then, considering j first, choose an area contiguous to j along a line. If this area is also contiguous along a line to i , the node has been identified. If the area is not contiguous along a line to i , as in the case of area n at node B, select an area other than j to which it is linearly contiguous. In Figure 2, this would be area m . Since m is linearly contiguous to i , the node is defined as $j-n-m-i$. This search procedure is equivalent to leaving area j by crossing boundary $j-n$ and ultimately crossing $m-i$ into area i . It can, therefore, be readily generalized to accommodate more than four areas that are mutually contiguous at a single node.

The process is repeated by entering *all* areas contiguous to either i or j . In this way, for instance, F and G in Figure 2 can be identified as separate nodes. Obviously some redundancies will occur and it is necessary to examine each node's list of area-identifying numbers in order to eliminate duplication. The node through which the barrier originally entered the section can be eliminated so that only new independent nodes remain. Every pair of areas linearly contiguous at a node represents a boundary whose taxonomic distance must be examined before choosing the next link in the barrier. Extra independent nodes can be

placed in a queue to be examined after the boundary encounters the edge or another boundary.

In a general sense the topology of the unit areas may be so complex that the above procedure will, if not modified, falsely identify a nonexistent node. Obviously, in Figure 3a there is but one node $i-k-j$. However, the configuration in Figure 3b will yield an additional false node $i-l-j$ since a path exists from i through l to j . Examination of just the linkages between mutually contiguous areas is not sufficient because the cases shown in Figures 3c and 3d have the same contiguities among i, j, k , and l as that in Figure 3b, yet in 3c and 3d there are, in fact, two independent nodes. The general solution is to see if a path through the areal units exists from node $i-l-j$ to the edge of the study region without passing through either area i or area j . In this way the conditions in Figures 3c and 3d, where no such path is possible, can be differentiated from that in Figure 3b. Many additional computations will be involved, and manually separating l into l and p , as in Figure 3e, will substantially reduce computer usage costs. It should be noted, however, that this problem is likely to occur primarily where l is an area surrounding the study region. This case can be dealt with easily since the false node would not produce a false additional branch of the barrier. Furthermore, multiple independent nodes (Fig. 2) are also extremely rare. Most boundaries have but two nodes and, if this is known, a less elaborate and less expensive algorithm will suffice.

Regional Membership

Barrier significance can be measured by summing the across-boundary gradients of all boundaries linked together to form the barrier and taking their average.

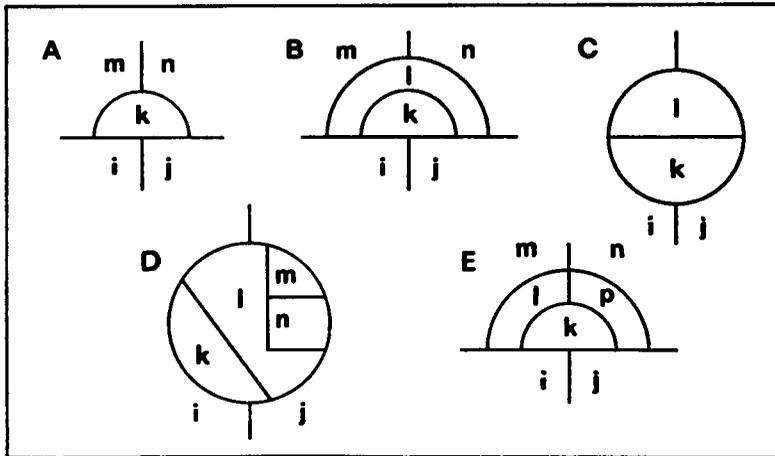


FIG. 3. Cases Illustrating False Node Difficulties. See text for explanation.

Although this average barrier difference is a useful indicator of barrier strength, it is not a measure of regional homogeneity. The standard gauge of a regionalization's significance, the ratio F of between-group variance to within-group variance, involves all areas in a division, interior places as well as those on the border. Thus, if the F -ratio is to form the basis for stopping the division process, an algorithm is needed to identify the groups into which the areas have been partitioned.

The method developed here for establishing regional membership recognizes that a barrier with no tied branches divides a previously unpartitioned region into, at most, three subdivisions. The number of groups, when it increases, is always incremented by one except when the two limbs of a barrier diverging in opposite directions bend around and encounter themselves; this kind of barrier can be pictured as two closed loops joined by a chain of boundaries, one of which has the greatest current across-boundary distance. Each tied branch adds one group regardless of whether it terminates at the outside border of the study area, at another part of the same boundary, or on itself. The important point is that each time a barrier limb stops it might or might not divide the area it penetrates into two parts but never more than two. No new division occurs when (1) a first limb followed away from an initiating boundary ends at the outside edge or at another boundary or (2) a second limb meets the edge or another boundary after the first limb has closed on itself. Thus, regional membership needs to be updated only when the second limb or a tied branch terminates or when the first limb closes on itself. This latter condition is recognized when the three or more areas meeting at the end node all belong to the same region.

Regional membership can be determined by treating all places as follows. First, if the place has not yet been assigned to a region it is given a unique region membership number. (Area One is thus always placed in Region One.) It is then compared with all other places. If a pair of places is not contiguous or has been separated by a barrier, the next sequential pair is considered. If they are still joined, however, a check is made to see if they have been assigned to the same region. If so, the next pair is considered. If not, a check is made to see if one has been assigned to any region at all. If it has not, it is assigned to the region of the other. If both have been assigned to different regions, one area is assigned to the region of the other and all places in the deleted region are also placed in the surviving region. At the conclusion of these $N^2 - N$ comparisons all areas in the same region have the same region membership number, yet some numbers between 1 and N will specify now empty regions. For convenience the regional identifying numbers are compressed so that with M regions the membership numbers will range from only 1 to M . At this point appropriate measure of internal homogeneity can be calculated.

AN EXAMPLE

To illustrate the maximum-difference criterion for barrier erection, this procedure was applied to 60 Km^2 -cells in the Albany, New York area (see Figure

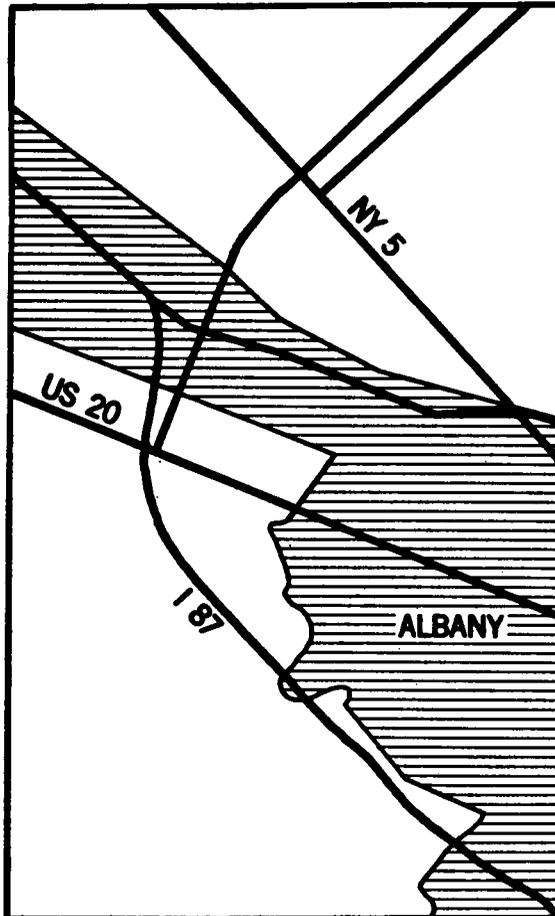


FIG. 4. Major Roads (heavy lines) and Incorporated Portion (shaded) of Test Study Region around Albany, New York. This 600-km² region is bounded on the east and west by the 599 and 593 1000-meter lines of zone 18 of the Universal Transverse Mercator Grid. The northern and southern limits are the 731 and 721 1000-meter lines. Not all of the city of Albany is included.

4). The raw data were accessed from the New York State Land Use and Natural Resource Inventory, a computerized data bank developed from areal photography in the late 1960's to provide rapid information retrieval for comprehensive regional planning [8, 10]. Nine percentage variables were computed from more specific categories and reduced to the four rotated principal components shown in Table 2. Although the relatively small size of the data cells might be considered adequate to insure internal homogeneity, the combining of, for instance, all commercial land use categories (shopping centers, commercial strips, and central business district) into a single percentage variable was deemed desirable.

TABLE 2
ROTATED PRINCIPAL COMPONENT LOADINGS FOR LAND USE DATA

Variable (Percentage of Land Use)	Communality	Component			
		I	II	III	IV
Agricultural	63	-53	-47	-02	-36
Commercial	49	62	25	-11	-17
Extractive	75	-39	68	-36	03
Forested	77	-82	02	-16	-29
Industrial	64	20	74	22	-12
Outdoor Recreation	85	-16	06	91	-03
Public and Semi-Public Areas	92	04	-06	-06	95
Medium- and Low-Density Residential	64	76	-08	-23	-07
Inactive Lands (not under construction)	57	75	00	00	10
Percentage of Total Variation	69.7	29.9	14.5	12.2	13.1

NOTE: Loadings and communalities are in 1/100ths for convenience.

The original land use categories frequently measure the results of growth processes that operate on a more detailed scale than can be portrayed effectively by a Km^2 -grid. The aggregation of land use types and the elimination of less significant variation by using only those principal components with an eigenvalue greater than 1.0 attempt to make these areal units and their values compatible.

These four orthogonal land use dimensions also provided the basis for a land use regionalization according to Ward's hierarchical group algorithm [15, 16]. So that the Ward procedure would serve as an effective control comparison, a contiguity constraint was employed to prevent disconnected regions, and the linkages were followed backwards manually so that the barrier partitions could be mapped and contrasted with those derived from the maximum-difference method. Although other algorithms were available for this comparison, Ward's method is more efficient than the centroid method, the only other widely used technique for regional taxonomy [6, pp. 198-99].

The principal difference between the two approaches is that Ward's algorithm is a fusion process whereas the maximum-difference method is a fission process. In Ward's algorithm each observation is initially a separate group. These groups are combined so that at each successive stage the pairing that is accepted (subject to the contiguity constraint) increases the within-group sum of squares by the smallest possible amount. Thus, Ward's and other linkage methods proceed from N groups to a single group. The maximum-difference technique, on the other hand, operates "backwards" starting with a single region and terminating when a specified number of groups has been delimited.

Since the number of observations is relatively large and since more than two dimensions are used to compute dissimilarity, the two methods are compared here by examining their map patterns rather than by referencing linkage tree diagrams or point plots of clusters. The most obvious effect of the maximum-difference method is the early isolation of relatively unique cells. For example, the first barrier completely surrounds a cell dominated by a complex of state government office buildings (see Figure 5). Also, as indicated by the second

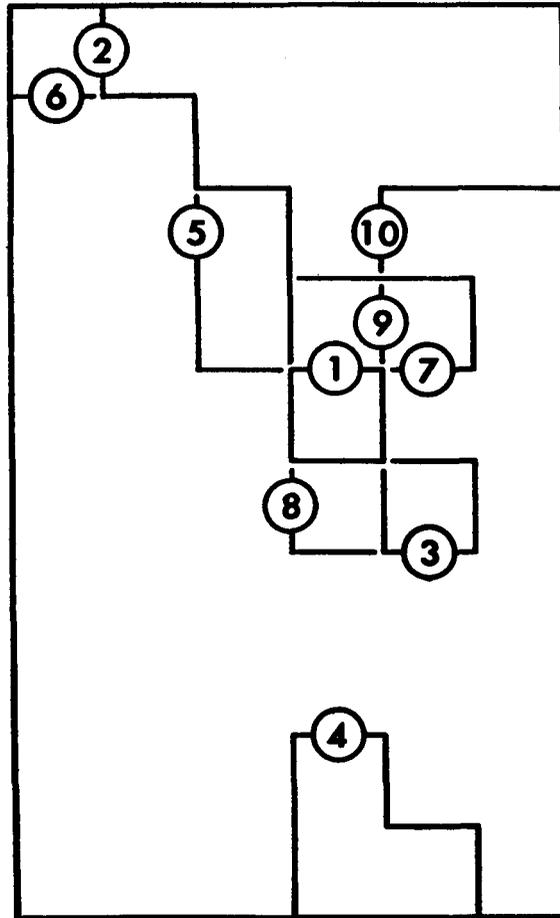


FIG. 5. Barriers Formed by the Maximum-Distance Method for Four Land Use Components. Numbers indicate sequence in which barriers were erected and are positioned at the initiating boundary in the barrier. Note that Barrier 2 did not produce a partitioning from two into three regions.

maximum-difference barrier, a new barrier need not produce a total partition of the region it penetrates. Although Ward's algorithm also segregates some of these more singular cells until the later stages of pairing, the within-group similarity criterion tends to produce groups of relatively more uniform size (see Figure 6). For instance, the third partition of the Ward algorithm divides the eastern and western portions of the test area into two distinct regions. These areas are primarily residential but the western section contains more agricultural and inactive land. The maximum-difference method does not yield this separation until the 21st partition since few of the adjacent cells here are notably dissimilar. Yet,

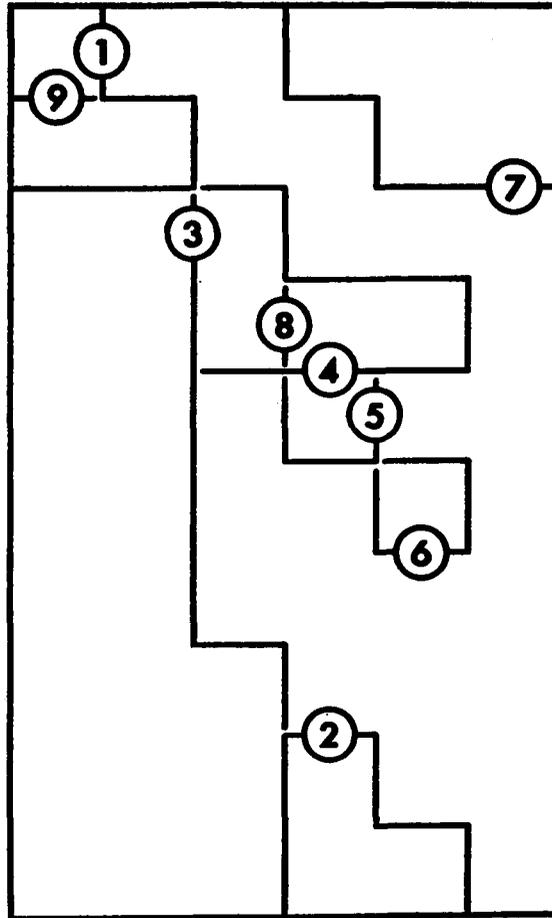


FIG. 6. Barriers Derived from Hierarchical Grouping Algorithm for Four Land Use Components. Numbers indicate the sequence in which the barriers were erected and are positioned at the boundary with the greatest cross-boundary distance.

for the ten most basic regions shown in Figures 5 and 6, five are identical in outline for both methods. These include an extensive tract of public land in the south which is delimited by the second hierarchical grouping barrier and the fourth maximum-difference barrier.

As the within-group sums of squares in Table 3 indicate, hierarchical grouping, as expected, yields more internally homogeneous regions than does the maximum-difference method. This is particularly true when the number of regions is small, as is the latter technique's advantage in greater average barrier distance. But, whereas the within-group sum of squares advantage always holds, the average barrier distance is not necessarily greater for the maximum-difference method at

TABLE 3
COMPARATIVE RESULTS FOR SELECTED REGIONAL PARTITIONS

NUMBER OF REGIONS	WITHIN-GROUP SUM OF SQUARES		NEW BARRIER DISTANCE			
	MAXDIF*	HGROUP*	Average MAXDIF	Average HGROUP	Maximum MAXDIF	Maximum HGROUP
2	213.5	195.7	4.37	3.44	6.65	5.70
2†	213.5	195.7	3.83	3.44	5.70	5.70
3	193.1	159.7	3.92	3.34	4.72	3.83
4	156.0	130.8	3.43	1.93	3.83	3.20
5	142.2	106.4	3.78	3.27	3.20	6.65
6	114.9	85.0	2.72	3.61	2.72	3.95
7	95.4	65.1	2.50	3.93	2.60	4.72
8	90.5	54.4	2.46	1.51	2.52	2.12
9	87.4	47.5	2.49	3.26	2.49	3.26
10	83.7	42.4	2.03	2.72	2.35	2.72
15	40.8	24.8	1.63	1.64	1.68	1.89
20	23.2	15.4	1.42	1.63	1.42	1.63
30	6.0	5.5	.90	.76	.90	.87
40	2.7	1.9	.59	.61	.59	.61
50	.5	.5	.41	.41	.41	.41
59	.0	.0	.24	.29	.24	.29

*MAXDIF and HGROUP refer to the maximum-difference and hierarchical grouping partitioning algorithms.
†The maximum-difference method's second barrier did not result in a division into three regions.

every level of partitioning. Since these barriers are initiated by the currently most significant boundary and since they are extended with the maximum-difference criterion employed at every junction until they terminate at the edge or at a barrier, a barrier can wander into an area of low contrast between adjacent cells. In addition, the more significant boundaries are likely to be added to a partition at an early stage of fission. For these reasons the average barrier distances, as shown in Table 3, also need not decrease monotonically. Nevertheless, this technique generally produces more significant barriers than does hierarchical grouping when the number of regions is small.

An obvious concern when comparing any numerical classification procedure is computation time, which tends to increase drastically as the number of OTUs rises. Table 4 shows the execution times required for four different computer runs. The most notable fact is that Ward's original algorithm with the contiguity constraint imposed did not yield a single pairing after 200 seconds of computation. This was initially thought to result from the program's first selecting a

TABLE 4
COMPARATIVE EXECUTION TIMES FOR PARTITIONING METHODS

Algorithm	Time in Seconds*
Hierarchical Grouping (List Structure)	4.125
Maximum-Difference (20 Regions)	10.317
Maximum-Difference (60 Regions)	35.600
Hierarchical Grouping (Contiguity Constraint)	200.000+†

*Central Processing Unit time only on UNIVAC 1108; does not include program compilation or link-editing.

†This program, which operated successfully with smaller numbers of observations, produced no groupings at all after exceeding the maximum execution time limitation.

potential pairing on the basis of the lowest increment in the within-group sum of squares and then checking for contiguity. However, when the contiguity constraint was examined first, the result was the same. Ward's algorithm was made efficient only when a self-updating list of contiguous regions was incorporated into the program; in this way, inspection, at every stage of pairing, of the large number of zero entries in the contiguity matrix was eliminated.

Similar economies could be achieved for the maximum-difference method if inspections for more than two independent nodes and for tied barrier branches were eliminated from the program. If the determination of regional memberships (included so that the within-group of squares could be computed) was also eliminated, even greater efficiencies would result. Thus, the maximum-difference method, although conceptually distinct from hierarchical grouping, can be made competitive in terms of computation cost if the study area does not have complex boundary relationships. Furthermore, since this is a fission rather than a fusion process, additional savings can result from requesting no more regions than could be shown effectively on a map.

CONCLUDING REMARKS

The significance of this research lies primarily in developing a programmable alternative to regionalization procedures based solely on *total-group* similarity. For problems in which maximum internal homogeneity is the clustering objective, hierarchical grouping procedures, centroid grouping, Neely's neighborhood limited algorithm [6], or discriminant analysis will be preferred. There are, however, instances in which the regions formed must take second place to their boundaries. Possible applications include problems where the significance of the boundary lines themselves is paramount or whenever it is necessary to segregate incompatible land uses or social groups. For these studies the maximum-difference barrier method may prove more relevant than present techniques.

LITERATURE CITED

1. BUNGE, W. "Gerrymandering, Geography and Grouping," *Geographical Review*, 56 (1966), 256-63.
2. CASETTI, E. *Classificatory and Regional Analysis by Discriminant Iterations*. Technical Report No. 12. Evanston, Ill.: Northwestern University, Geography Department, 1964.
3. EDWARDS, A. W. F. and L. L. CAVALLI-SFORZA. "A Method for Cluster Analysis." *Biometrics*, 21 (1963), 362-75.
4. HAUTAMAKI, L. "Some Classification Methods in Regional Geography," *Fennia*, 103 (1971).
5. JENKS, G. F. and F. CASPALL. "Error on Choroplethic Maps: Definition, Measurement, Reduction," *Annals, Association of American Geographers*, 61 (1971), 217-44.
6. LANFORD, P. M. "Regionalization: Theory and Alternative Algorithms," *Geographical Analysis*, 1 (1969), 196-212.
7. MCCLEARY, G. F. "Beyond Simple Psychophysics: Approaches to the Understanding of Map Perception," *Technical Papers, 30th Annual Meeting of the American Congress on Surveying and Mapping, 1970*. Pp. 189-209.

8. MONMONIER, M. S. "New York's Land Use Data Bank," *Geographical Review*, 61 (1971), 597-98.
9. ———. "Contiguity-Biased Class-Interval Selection: A Method for Simplifying Patterns on Statistical Maps," *Geographical Review*, 62 (1972), 203-28.
10. New York State Office of Planning Services. *LUNR: What It Is and How to Use It*. Albany: New York State Office of Planning Services, 1972.
11. RAY, D. M. and B. J. L. BERRY. "Multivariate Socioeconomic Regionalization: a Pilot Study in Central Canada," in S. Ostry and T. Rymes, eds., *Regional Statistical Studies*. Toronto: University of Toronto Press, 1966. Pp. 75-130.
12. SPENCE, N. A. "A Multifactor Regionalization of British Counties on the Basis of Employment Data for 1961," *Regional Studies*, 2 (1968), 87-104.
13. SPENCE, N. A. and P. J. TAYLOR. "Quantitative Methods in Regional Taxonomy," *Progress in Geography*, 2 (1970), 1-64.
14. TAYLOR, P. J. "The Location Variable in Taxonomy," *Geographical Analysis*, 1 (1969), 181-95.
15. VELDMAN, D. J. *FORTTRAN Programming for the Behavioral Sciences*. New York: Holt, Rinehart and Winston, 1967. Pp. 308-17.
16. WARD, J. H., JR. "Hierarchical Grouping to Optimize an Objective Function," *American Statistical Association Journal*, 58 (1963), 236-44.