CHARACTERISTIC POINTS AND THEIR SIGNIFICANCE IN CARTOGRAPHIC LINE GENERALIZATION

by

Jill S. Marino
B.A., University of California, Santa Barbara, 1970

Submitted to the Department of Geography and the Faculty of the Graduate School of the University of Kansas in partial fulfillment of the requirements for the degree of Master of Arts.

Professor in Charge

For the Department

ACKNOWLEDGEMENTS

I gratefully acknowledge the guidance and advice from Professor George F. Jenks; and also wish to thank Professor George F. McCleary, Jr. for comments, and Professor Curtis J. Sorenson for assistance regarding geomorphology.

Special thanks are due to the many people who were good-natured enough to stick hundreds of pins into narrow black lines!

I also thank Tony, my husband, for support, encouragement, and advice. A source of inspiration was found in the memory of Henry Bergman.

"...One of the cartographer's greatest problems is the selection and generalization of the data with which he is working. How many contour lines to draw, how much to smooth them, what selection of rivers or towns to make, how to simplify a coastline, and many others, are perennial problems to the cartographer. Considerable importance has been attached to these functions in cartographic writing, and the very frequency of their appearance and concern to the cartographer (especially generalization) has fostered a tendency to think of them as cartographic. All scientific endeavor is constantly faced with the task of evaluation and generalization, and cartography, rather than being an exception, merely follows the rule."

Arthur H. Robinson, The Look of Maps, 1966, page 14.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

Cartography is often described as both an art and a science. In many areas of cartography the subjective and non-quantitative approach has given way to more rigorous quantitative methods. In particular, that part of cartography which deals with line generalization has increasingly abandoned intuitive, *ad hoc* practices and has adopted more systematic procedures. One reason for this shift in methodology is a desire on the part of cartographers to construct a unified theory for their discipline (Morrison, 1974b, 12). But the more important impetus, necessity, arises from increasing use of computers in map-making processes.

Cartographic generalization is a process of evaluation for relative significance, of selection, and of simplification of data in order to transmit a clear and legible message to the map user (Jenks, 1963; Miller and Voskuil, 1964). Although various definitions of cartographic generalization exist, each stresses that it is a reduction of or partial selection from the original information, and a simplification with the retention of the essential or characteristic elements of the data

(Lundquist, 1959, 466; Neumann, 1972, 9; Pannekoek, 1962, 55; Robinson and Sale, 1969, 52).

Cartographic generalization is a function of map scale and map purpose. Clearly, every map is a product of generalization. (An identity mapping would be an exception, but does not produce a map in the sense of the word model.) With a decrease in map scale, it is not physically possible to retain all the initial information in visible, legible form. Small scale thematic maps, derived from larger scale sources, are therefore subject to generalization. In addition, the purpose of some maps is served if base data are generalized. For example, thematic information is often accentuated if base data are simplified. Extreme generalizations, or schematic renderings, are also effective in some cases (Hogben, 1960, 284; Wiedel, 1977).

The instructions for generalization given in 1879 by M. Siegfried--"A general map demands an intelligent reduction, emphasizing the important, restraining the unessential and avoiding the overloading"--are sound today (Neumann, 1972, 6). The procedures, however, are beginning to change.

In an often-quoted paper published in 1942, J.K. Wright discussed the implications of "the fact that maps are drawn by men and not turned out automatically by machines" (1942, 527). With the subsequent development of the computer, its availability and application in various disciplines, including geography and cartography, we now do have maps produced by machines. The computer is not yet a panacea for cartography, for as many have pointed out, automation is disrupting traditional attitudes and is forcing a re-examination of many conventional cartographic procedures (Muehrcke, 1972, 47; Tobler, 1966).

Traditionally, line generalization has been a subjective practice (Lundquist, 1959; Steward, 1974). The "rules of thumb" passed along from teacher to student have stressed experience, personal judgement, skill, and a knowledge of geography. For example, Monkhouse and Wilkinson state "The compiler chooses in the light of his experience the correct line-thicknesses..." (1966, 10). Miller and Voskuil advise one to "combine practical experience with common sense and a flair for the subject" (1964, 19). Pannekoek, Robinson and Sale, and Saunders say that knowing which essential characteristics to stress can only be determined by having a clear geographical understanding of the region to be depicted (1962; 1969, 16; 1958, 11). Raisz instructs with an example: "Intelligent generalization demands a good knowledge of geography and a sense of proportion... . For instance, we would not miss the characteristic sharp point of Cape Hatteras on a map at any small scale" (1962, 38). In the context of generalizing contours, Imhof advises one to rely on knowledge of geomorphology in order to avoid rounding off critical points (1963).

The trend towards automated cartography requires that line generalization techniques be objective. Srnka concludes from generalizations done by cartographers that "certain intrinsic laws are in evidence in the selection of the elements represented." The key problem, he adds, is finding theoretically correct and objective standards, expressed quantitatively, for the selection of represented features (Srnka, 1970, 48). Although Töpfer and Pillewizer devised formulas applying to the number of elements which should be included on a generalized map as scale decreases, they offer no guide as to which elements should be retained (1966). The need to reduce individual bias by establishing impartial, universally applicable criteria for line generalization is evident (Steward, 1974), because efforts for "fully automated cartographic processing have all failed to some degree for lack of objective map generalization guidelines" (Muehrcke, 1972, 47).

Automated cartographic procedures are advantageous for several reasons. Today cartographers are being pressured to decrease mapping time and costs, and yet produce more and better maps (Muehrcke, 1972). Survey data are being rapidly generated by aerial photographs and other remotely sensed images, but traditional mapping methods can not keep pace with this rapidly expanding data supply (Peucker, 1972 a, 61). Maps generated for planning purposes are often out of date before they are even finished. For example, a map of air pollution over a metropolitan area is needed quickly in order to be useful (Muehrcke, 1972, 27). The versatility of digitized map data is also an advantage because the map can be reproduced at different scales, on different projections; be compiled with other geographical base data information; or incorporate updated information (Douglas and Peucker, 1973, 115; Linders, 1973). The advantages of automated map production are speed, accuracy, and flexibility of data manipulation and display. Automated methods should enable cartographers to satisfy the demand for more maps in less time at a lower cost per unit (Tobler, 1959; Muehrcke, 1972, 3).

Two types of line generalization algorithms have been proposed for use in automated cartography: one approach eliminates points, while the other selects points along a line. The majority of algorithms are of the first type. One such suggestion has been to select every n^{th} data point along the line (Bateman, 1974; Tobler, 1964). Another proposal involves simple or weighted averaging of n successive points (Tobler, 1966). Koeman and Van der Weiden recommend averaging points within a specified distance (1970).

Many methods of line generalization incorporate a tolerance limit of some sort. Perkal eliminates indentations less than a radius of 2ϵ, for an ϵ arbitrarily defined (1966). Coles modifies the same idea (1966). Hershey's use of a tolerance limit eliminates points whose distance apart is less than some function of the plotted line width (1963). Boyle interpolates a line between points which are a specified distance apart (1970). Vannicek and Wollnough also rely on a specified tolerance limit, but their method involves a tube of arbitrary width, ϵ, inside of which the generalized segments are constructed (1975). Lang incorporates a tolerance based on the sum of the absolute value of the distances from a proposed segment to all intervening points (1969, 50).

Other suggested methods approximate lines by removing a series of points and replacing them with mathematical functions. Rhind links widely spaced points with mathematical spline functions (1974). Clark approximates a curve using basis functions which fit within the convex hull of the given data points (1974). Brophy smoothes a subset of equally spaced points with polygonal curves (1972). Breward fits polynomials to curves with the constraint that the polynomials approximate segments between turning points (1972).

Although the proposed solutions to systematic cartographic line generalization mentioned above are based on different concepts, the resulting product of each is undesirable because of the indiscriminate shifting of original data points at certain critical locations. Smoothing techniques "tend to depress the effect of extreme points, often the very points which give character to a line" (Douglas and Peucker, 1973, 115). Lang argues that in generalizing a line, one must "keep the salient features but discard excessive detail, to produce what is essentially a caricature of the original" (1969, 50). As Lipkin and Rosenfeld state, in some cases it may only be necessary to display "a visual precis of those points that convey the information" about the object of interest (1970, 34). This point of view suggests a different approach to utilize in developing line generalization algorithms.

Interest in points which characterize a line or line drawing is not restricted to cartographers. Psychologists, engineers, and computer scientists are also concerned with form recognition and search for aspects of a picture which maximize its recognizability. In the widely-quoted experiment conducted by Attneave in 1954, subjects drew a pattern of 10 dots on outline shapes which would resemble the shape as closely as possible. The result was that subjects showed a great deal of agreement "in their abstractions of points best representing the shape, and [that] most of these points were taken from regions where the contour was most different from a straight line" (Attneave, 1954, 185). Noton and Stark conducted a similar experiment and concluded as had Attneave, that the most informative parts of a line drawing are angles and sharp curves (1971). Dent later repeated the same test, using state outlines as the figures to be approximated by subjects. He found a significant correlation between measures of the original and generalized shapes (Dent, 1972). The results of these studies indicated that certain points on a figure appear to be more significant than others.

While Douglas and Peucker are the first geographers to present an algorithm for cartographic line generalization based on locating significant points on a line or figure, they claim that their work was developed independently of previously published studies in the fields of engineering and computer science. Among studies in these other disciplines, Rosenberg, modifying Langridge's earlier work, has established a method for identifying perceptually significant points (dominant points). Dominant points satisfy several conditions: from an initial point, each is most distant and at least $\pi/2$ radians in either direction; each is most distant from a chord whose perimeter-to-chord ratio is greater than the value 1.2. Several iterations may be necessary before stabilization of points selected as dominant occurs (Rosenberg, 1972; Langridge, 1972). Ramer describes an algorithm which approximates curves with straight line segments constrained by a specified tolerance of "fit". If a derived segment does not satisfy the fit threshold tolerance (i.e., the maximum distance between it and the original curve is too great), then the original curve is approximated by two derived segments which intersect at the point of maximum distance between the original curve and the initially derived segment, and checked for fit again. This algorithm applies to both open and closed lines (Ramer, 1972).

Peucker has recently presented to cartographers an algorithm for cartographic line generalization which is very similar to the procedure developed by Ramer. The first step in Peucker's algorithm is to connect the two end points of the original line, thereby defining its "general direction". The perpendicular distance from the "general direction" to every point on the original is then calculated. The point on the original line which is the maximum distance from the "general direction" is retained and becomes the end point for a new "general direction" line. The process is then repeated until the maximum distance between the original line and the derived "general direction" is less than some arbitrary value (Peucker, 1975). This method seems to provide a quantitative expression for the characteristics of the elements of a line, as well as objective rules for their selection. It has not yet been established whether the product of this generalization algorithm is acceptable to cartographers and/or to map users.

The present study examines the issue of the selection of characteristic points along a line by trained and/or experienced map makers, and by cartographically naive individuals. One underlying assumption is that certain points (characteristic points) along a line must be maintained in any generalization in order to communicate the character of the line. That is, the generalized line must pass through these points in order that the character of the line be maintained. "It must be strongly emphasized that we can not merely smooth out coasts and ignore islands. The basic or fundamental character of the shapes and outlines must be retained and emphasized in their simplicity..." (Robinson and Sale, 1969, 55). Secondly, it is assumed that lines symbolizing certain geographical features can and do have a different character from one another.

The working hypothesis for this study is that skilled map makers -- cartographers, geographers, and geologists -- will select the same set of characteristic points along any given line, and that this set will be similar to that selected by cartographically untrained individuals.

Six sample lines were used in this experiment. Classic examples of three types of coastlines and three types of river channel patterns were selected in consultation with a geomorphologist. The two sample groups of respondents were instructed to mark (on the lines) points which should be retained at three increasing levels of generalization.

Therefore, this study differs from others in two respects. First, a comparison was made between responses given by cartographers and those given by non-cartographers. Secondly, three levels of abstraction were required, in each of which respondents decided how many, as well as which, points were necessary to characterize the given line.

Details of the experiment which was performed and of the respondents who participated, follow in chapter two. Chapter three presents the data which were gathered and the results of statistical tests substantiating the similarity of responses given by the two sample groups. The concluding chapter is a discussion of the interpretation and implications of the results of this study.

CHAPTER II

THE EXPERIMENT

For this study an experiment was designed in which a sample group of respondents selected characteristic points on a line. The implicit assumption was that these points could be used in a cartographic generalization of the line. The incentive for conducting this experiment was twofold. If the responses of the sample group were consistent, or showed considerable agreement, this would provide a set of data against which a line generalization model might eventually be tested. In addition, similar responses by the individuals sampled should provide insight into the types of points which are perceptually significant in the generalization of a line. Designing the experiment involved several decisions, including what line(s) to use, how to get a response from an individual, and how to select a sample group of respondents. The decisions, their rationale, and a description of the experiment follow.

Selection of Sample Lines

Two thoughts motivated the selection of lines used in this experiment. Because responses for only one line would have had limited value, and because the possibility of generalizing the results would be greater with several, different types of lines, several lines were used in the experiment. To maintain a high degree of realism, lines representing natural phenomena were chosen rather than a set of constructed lines. Therefore, six lines, representing portions of three rivers and three coastlines, were used in this experiment.

Geomorphologists classify coastlines and river channel patterns by type, providing established categories of line forms. These geomorphological classifications were used to select a group of test lines. According to William Thornbury (1969, 124-125), the noted geomorphologist, river channels may be grouped into one of three forms or patterns: straight, braided, or meandering. The braided stream channel was eliminated from this study because single lines were desired. In consultation with a geomorphologist, the Fall River in Colorado was chosen to represent a straight river channel; the North Fork of the Shenandoah River in Virginia was selected to depict a meandering river channel; and the Mancos River in Colorado and Utah was chosen to represent a stream channel which meanders but is described as youthful, and as a line form is therefore between the relatively straight Fall River and the meandering Shenandoah River (Tuttle, 1970, 60). All three of these rivers are classic examples used in teaching geomorphology, and are shown in Figure 2.1.

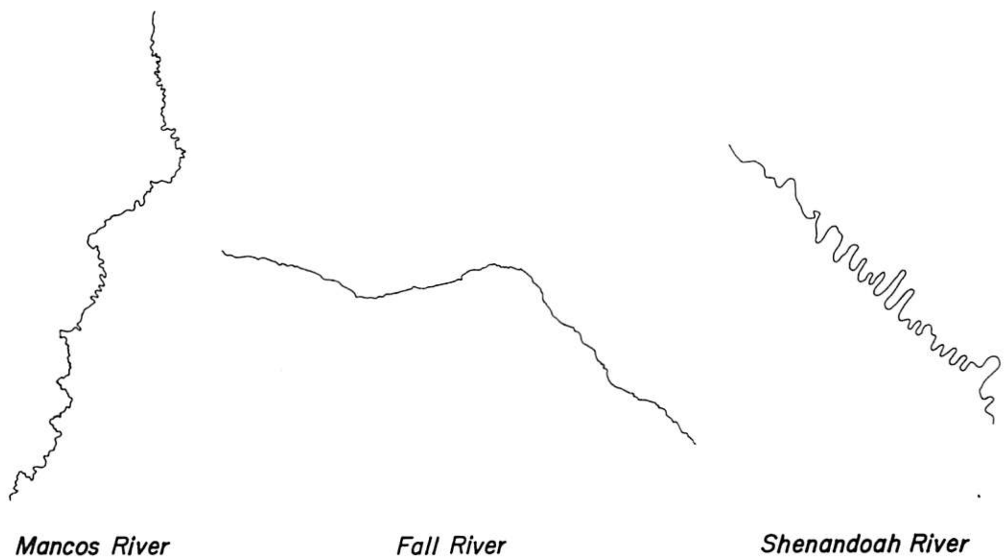


Figure 2.1

Although many classifications of coastlines exist, F.P. Shepard's 1963 genetic classification was chosen for its comprehensiveness (King, 1972, 406-414). Examples of three types of coastlines selected for use in this test include Chesapeake Bay, Maryland and Virginia, a drowned river valley of the primary land erosion type coast; Cape Arago, Oregon, an irregular coastline of the secondary wave erosion type coast; and Cape Hatteras, North Carolina, a special type of barrier coast -- the cuspate foreland -- of the secondary marine deposition type coast (Shepard, in Fairbridge, 1968, 131-133). See Figure 2.2.

Lines representing portions of these six examples of geomorphological features were traced from U.S. Geological Survey topographic maps which are produced according to

Figure 2.2

standard specifications.[1] The largest scale possible was used, subject to the constraint that when reduced to 67%, the feature could be displayed on an 8½ by 11 inch (21.25 by 27.5 cm) format. Multiple copies of these distinctive natural linear features were reproduced by offset printing.

[1]The following maps were used: U.S. Geological Survey (1957), Trail Ridge, Colorado Quadrangle, 1:24,000. U.S. Geological Survey (1961), Estes Park, Colorado Quadrangle, 1:24,000. U.S. Geological Survey (1947), Charlottesville, Virginia, U.S., 1:250,000. U.S. Geological Survey (1961), Cortez, Colorado-Utah, 1:250,000. U.S. Geological Survey (1900), Point Lookout, Maryland-Virginia Quadrangle, 1:62,500. U.S. Geological Survey (1970), Cape Arago, Oregon Quadrangle, 1:24,000. U.S. Geological Survey (1957), Manteo, North Carolina, 1:250,000.

<u>Method Used to Obtain Responses</u>

The objective of this experiment was to have a sample group of individuals select characteristic points on a given line. Pretests revealed that asking individuals to generalize a given line by drawing directly on the line with a pen or pencil, resulted in two major problems with regard to interpreting the results. First, it was often difficult to determine where one vector ended and the next began. Secondly, the extreme variation among the respondents in the number of segments used to generalize the line resulted in rather meaningless comparisons between some responses. It was therefore necessary to indicate to the respondent the degree or level of generalization desired in his or her response. This was accomplished by setting a limit on the number of points which could be selected. It was believed that a task involving a high degree of generalization would be more difficult than one in which little generalization was involved. Therefore it was decided to direct the individual to respond in stages of increasing generalization in order to obtain the desired response in which substantial generalization was present. The response from an individual, then, would consist of three stages, in each of which successive increases in generalization would be required.

The technique employed to obtain a response consisted of using a box of straight, dressmaker, glassheaded pins (approximately $1\frac{1}{4}$ inch (3.125 cm) in length with heads 1/8 inch (0.31 cm) in diameter) which the respondent was instructed to stick directly into the printed line. This method had two advantages. First, it produced easily identifiable points. Secondly, it restricted the number of points which an individual could choose, and therefore suggested the amount of generalization desired in the response. Although this technique set an upper limit on the number of points a respondent could select, the individual was not required to use all the pins. Providing each respondent with a predetermined number of pins to use in selecting characteristic points permitted a response to be obtained at three suggested levels or degrees of generalization.

To enable the respondent to manipulate these pins in his or her selection of characteristic points along the sample line, the first stage response in this experiment began with five pins per inch (2 pins per cm) of printed line. The number of pins provided to the respondent in the second and third stages were 2.5 and 1.25 pins per inch of line (1 and 0.5 pins per cm of line), respectively, based on the assumption that the amount of information which can be shown on a map as the area is reduced, decreases in a geometrical progression (Robinson and Sale, 1969, 53; Töpfer and Pillewizer, 1966, 10. A geometric progression or sequence, with first term a and ratio r is denoted $a_n = ar^{n-1}$ (Saltz, 1974, 141). In this case, a = 5x, and r = 0.5, where x is the length of the sample line.).

The Test Procedure

The test was administered in the office, classroom, or home of the respondent, on an individual basis or in groups of not more than four persons. The instruction sheet read by the respondent asked him or her to indicate, by sticking pins directly into the sample line, those points which he or she considered to be the most important ones in preserving the character of the line. A copy of the instruction sheet used in this experiment is included in Appendix A. After questions had been answered, the respondent was given a copy of one of the six sample lines, mounted on a 12 by 12 inch (30 by 30 cm) piece of soft fiberboard, and a box of pins which were to be stuck into the printed line. The color of the glass heads of the pins was the same in each box, and changed at each stage to implicitly reinforce the idea that each stage involved a different response.

Each sample line was always presented to the respondent with the same orientation (that shown in Figures 2.1 and 2.2) although some individuals rotated the line as they worked. The respondent was free to use as many or as few of the pins as he or she felt was sufficient. No time limit was indicated or enforced, and the response times varied considerably. When the respondent indicated that he or she had completed the first stage of the experiment, materials were removed, and a duplicate of the line was presented with a box of pins containing one half the number originally provided. Upon completion of this second stage, materials were removed, and another duplicate of the line was given to the respondent along with a box of pins containing one fourth the number given in stage one. This completed the response of one individual in selecting characteristic points at three levels of generalization on the same line. The individual was then free to respond to another of the sample lines, if he or she desired. Due to the time, patience, and interest involved in participating in this experiment, most persons responded to only two sets of the six sample lines.

On the individual level, the manner of response by participants was rather evenly divided with respect to the way in which each selected characteristic points along a sample line. The first approach was to begin at one end of the line and to consecutively mark points as the individual progressed to the other end of the test line. The other approach, which appeared to be more efficient, consisted of the selection of a set of primary points along the sample line, with the subsequent inclusion of subordinate points.

It is also notable that most of the individuals sampled responded with three successive increases in generalization. The instructions (complete in Appendix A) merely implied that this type of response was desired.

Selection of the Experimental Population

Because cartographic line generalization has so often been described as a procedure based on skill (Pannekoek, 1962), experience (Miller and Voskuil, 1964), and geographic knowledge (Imhof, 1963), it was decided to select two groups of respondents. One group consisted of trained and/or practicing cartographers, and is subsequently referred to as "cartographers". By definition, these individuals had earned at least a masters' degree in geography with emphasis in cartography; and/or had been or were employed as cartographers by a private, city, state, or federal mapping and/or planning agency; and/or had earned a PhD in geography or geology, having considerable experience working with or making maps. The other group of respondents was untrained cartographically, and is referred to as "non-cartographers". By definition, these individuals had no cartographic training or work experience, and had not completed more than two introductory geography classes.

The experiment was administered to a total of 174 individuals. The responses of 13 persons were later eliminated because the individuals belonged to neither group as defined above. Of the remaining 161 individuals whose responses were studied, 94 were cartographers and 67 were non-cartographers. At the time these persons participated in this experiment, each resided in Kansas, California, or Colorado.

As listed in Table 2.I below, the total sample of respondents was composed of 32% female and 68% male. In terms of years of completed formal education, 40% of the total sample had completed high school but less than four years of college; 19% had earned a BA but not an MA degree; 17% had earned an MA but not a PhD degree; 20% had earned a PhD. The educational status of 4% of the total sample is unknown.

Table 2.I also describes the composition of the group labeled as cartographers as well as that of the group labeled as non-cartographers. Of the cartographers, 66% had earned an MA and/or PhD degree, and 79% were male. Of the non-cartographers, 65% had not completed four years of college, and 60% were male.

The experiment was administered to obtain 30 responses by cartographers for each line, and 30 responses by non-cartographers for each line. Because individuals did not usually respond to all six lines, a description of both groups for each line may be of interest, and is included in Appendix B.

Table 2.I

The Sample Group of Respondents

	Total Sample	Cartographers	Non-Cartographers
Sex			
female	32%	21%	40%
male	68%	79%	60%
Education (y)			
secondary $\leq y <$ college graduate	40%	4%	65%
college graduate $\leq y <$ MA	19%	25%	14%
MA $\leq y <$ PhD	17%	30%	9%
$y \geq$ PhD	20%	36%	10%
unknown	4%	4%	3%

CHAPTER III

THE RESPONSES

The experiment described in the preceding chapter produced two sets of response data. One set consisted of the responses given by the group of cartographers, and the other set included the responses by the group of non-cartographers. In order to compare results between the two groups, it was necessary to cumulate the individual responses given for each sample line, at each stage of generalization by the two groups. These cumulated or summarized group responses were tabulated and graphed. The resulting data sets were then statistically analyzed to determine the degree of correlation.

Data Accumulation

A summary of each group's response was plotted on one copy of that particular sample line. These summary sheets were constructed manually using the following procedure. A copy of a particular sample line was placed over a participant's response line and carefully registered. The points selected by the respondent appeared as red-marked pinholes in the paper, and were visible through the summary sheet when viewed on a light table.

As an individual completed his/her response during the experiment, a red mark was made through the hole which remained as a pin was removed. Not only did this facilitate data accumulation (specifically tabulation), but because repositioning of the pins was permitted, it also ensured that only those points were recorded which were the final choice of the respondent.

Whenever a marked pinhole existed on the response line, a thin pencil line was drawn out from that point on the summary sheet of the line. A dot was placed on this pencil line to signify that this point had been selected by one individual. This process was repeated until all 30 responses by one group (cartographers or non-cartographers) for a particular stage (degree of generalization) of a given line had been summarized onto the summary sheet. See Figure 3.I.

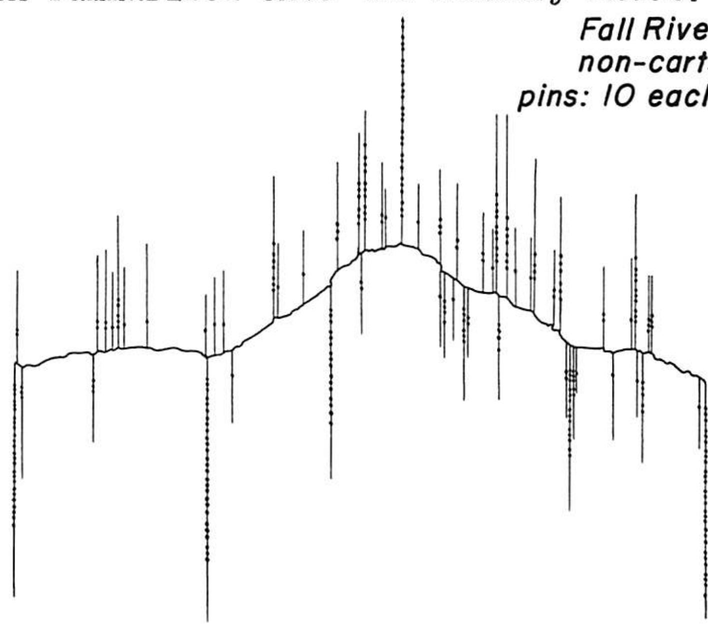


Figure 3.1 *one summary sheet*

Accuracy of Summary Sheets

Several types of errors affected the accuracy of these summary sheets. First, the physical limitations of the size of the pinhole in the response line restricted accuracy to a minimum of 0.046 inch (.115 cm). This resulted from the fact that the mean pinhole diameter, based on a random sample of 65 pinholes from the response data sets, was 0.023 inch (0.0575 cm). The error in recording responses (assigning a given pinhole to a particular pencil line) was judged to be less than or equal to one-half of the average pinhole diameter (0.0115 inch or 0.0286 cm) on either side of a given pinhole. Consequently, the linear distance of assured accuracy was 0.023 inch + (2)(0.0115) inch = 0.046 inch (or 0.0575 cm + (2)(0.0286) cm = 0.1147 cm). Therefore, a conservative estimate of accuracy to two-decimal places yielded 0.05 inch (0.0125 cm). Secondly, slight error could have been introduced when registering the position of a summary sheet over a response line.

A third component of error involved the respondent him/herself. When one considers the manual dexterity of a normal individual performing the assigned task, one must consider that slight differences may exist between the actual pinhole in the response line and the exact point which the respondent intended to select[1]. The width of the line on the response map, together with its sinuosity also would have limited the precise location of selected points in some cases.

Aggregating Data for Statistical Analysis

To compare the two group responses quantitatively, the data as accumulated on the summary sheets required aggregation. Statistical tests to measure the degree of correlation between the two group responses required that the two data sets be comparable. In order to compare the responses, the data were aggregated into equal intervals of 0.05 inch (0.125 cm). This interval length was selected as a conservative estimate of accuracy, according to considerations of possible error in pinhole location as discussed in the preceding section. Pinholes which differed in location by less than 0.05 inch (0.125 cm) were judged to be essentially the same.

To segment a line into equal 0.05 inch (0.125 cm) intervals, each sample line was digitized. Points which

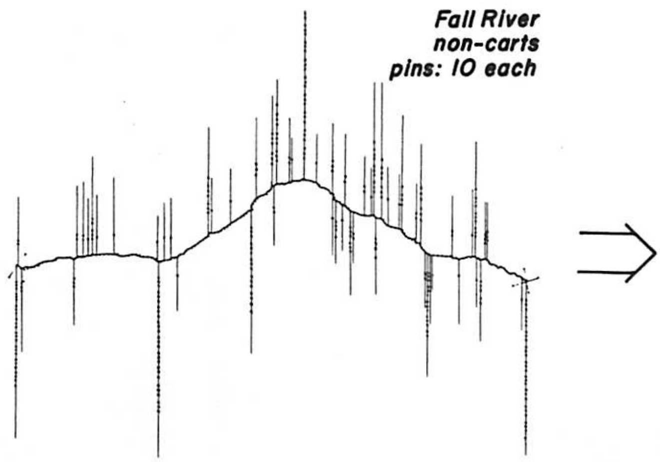
---

[1]Because this experiment was administered on an individual basis, respondents' reactions to the task could be noted. Pertinent reactions usually included a statement by the respondent that he/she felt that the task was rather tedious, and, on the sample line of Chesapeake Bay, especially, that he/she experienced some difficulty in physically being able to place a pin at a particular point which was already surrounded by a cluster of pins.

were 0.05 inch (0.125 cm) apart along the sample line were plotted. These points were reproduced as a film positive. The equally spaced points produced the following number of segments on each sample line: Cape Arago, 299; Cape Hatteras, 463; Chesapeake Bay, 680; Fall River, 164; Mancos River, 233; Shenandoah River, 243. See also, Appendix D.

By placing the film positive over its associated summary sheet of each group's response, it was possible to aggregate the data into a form useable for statistical testing. Figure 3.2 illustrates this aggregation process. The film positive overlay made it possible to count the number of responses within each 0.05 inch (0.125 cm) interval. These numbers were recorded sequentially on a tabulation sheet.

With the data in this new form, it was then possible to transform the summarized group responses for



summary sheet with film positive overlay

segment number	no. dots on first line	no. dots on second line
1	25	2
2	3	0
3	0	0
4	0	0
5	0	0
6	0	0

tabulation sheet

Figure 3.2

a sample line into histogram format. Figure 3.3 illustrates this step which produced a graph of summarized responses.

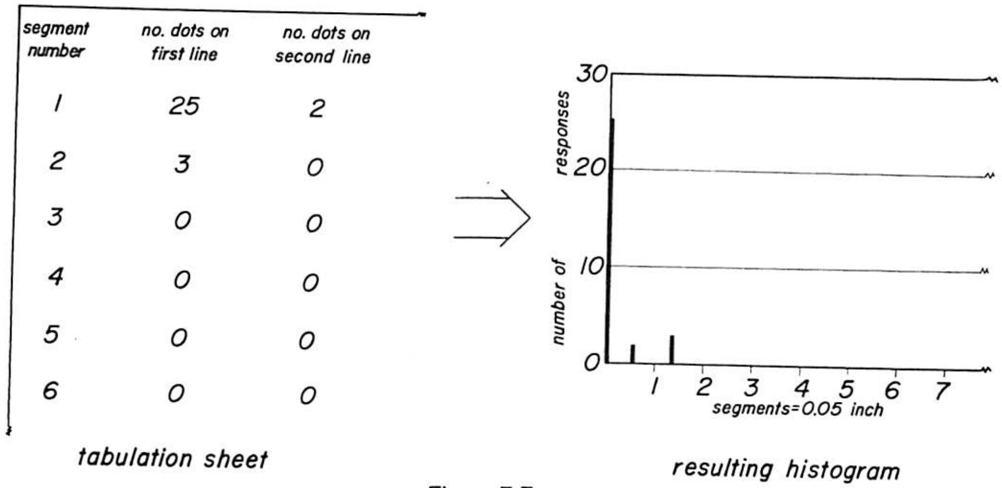


Figure 3.3

Eighteen pairs of corresponding graphs were produced in this manner. Each pair of graphs (for example, the pair shown in Figure 3.4) is comparable because one graph represents the summarized response by the group of cartographers at a given stage of generalization for a given sample line, and the other graph represents the summarized responses by the group of non-cartographers at the same stage of generalization for the same sample line.

Similarity of Aggregated Responses

The similarity between corresponding pairs of graphs is evident. Pearson product-moment correlation coefficients to establish the degree of association present were computed for each corresponding pair.

The null hypothesis set forth in this statistical test for correlation was that the group of cartographers distributed their selection of characteristic points along the sample line in a different manner than did the group of non-cartographers. That is, $H_0 : \rho = 0$; where ρ is the population parameter of which r is the sample estimate. The significance test was derived from use of the Student's t distribution with n-2 degrees of freedom.

The computed values of r for each corresponding pair are included with the graphs in Figures 3.4 through 3.30. In addition, they are listed in comprehensive form in Table 3.I below. (See Appendix E also.) If the true $\rho = 0$, the probability that r values such as those which resulted would occur is 0.001. That is, there is one chance in a thousand that, based on the sample data, these closely associated responses between the group of cartographers and the group of non-cartographers could have occured by chance.

The graphs are presented in their entirety in Figures 3.4 through 3.30 for several reasons. Although the correlation coefficients were statistically significant for every corresponding pair, the graphic presentation of the summarized responses reveals more strikingly the

Table 3.1

Pearson product-moment correlation coefficients

for all corresponding pairs of responses

Sample line and stage of experiment	Pearson product-moment correlation coefficient	significance level
Cape Arago		
stage one	0.8849	0.001
stage two	0.9320	0.001
stage three	0.9063	0.001
Cape Hatteras		
stage one	0.8368	0.001
stage two	0.9096	0.001
stage three	0.8810	0.001
Chesapeake Bay		
stage one	0.9452	0.001
stage two	0.9530	0.001
stage three	0.5482	0.001
Fall River		
stage one	0.8520	0.001
stage two	0.9089	0.001
stage three	0.7950	0.001
Mancos River		
stage one	0.8779	0.001
stage two	0.9301	0.001
stage three	0.9071	0.001
Shenandoah River		
stage one	0.6830	0.001
stage two	0.8861	0.001
stage three	0.8992	0.001

similarity between the responses given by cartographers and those given by non-cartographers. It is evident from these graphs of summarized responses that points were not selected randomly. Also, one may compare the two group responses in detail for any particular point along the sample line, may assess the degree of consensus of opinion by respondents at various points, or may note the points (or types of points) which were selected consistently, regardless of the degree of generalization. Specific comments regarding these response attributes are contained in Chapter Four.

In Figure 3.4 the summarized responses are shown for the sample line of Cape Arago. This is one corresponding pair, in that the graph at the top of the page, labeled "non-cartographers", and the graph directly below that, labeled "cartographers", are comparable. (For a description of these sample groups, see Appendix B.) This figure illustrates the results of the first stage of the experiment, requiring the least amount of generalization, in which each individual was given 73 pins, and therefore could select no more than 73 characteristic points along this line symbolizing Cape Arago. The vertical, black lines on the graphs indicate the number of respondents who selected this particular point on the sample line. For example, 26 cartographers and 28 non-cartographers selected the left-hand end point on the line symbolizing Cape Arago. Cape Arago is shown at the bottom of the page, and is the same size as that used in the actual testing. The two graphs are aligned, and where feasible, a point on the graphs has been tied to its respective location on the sample line itself. For example, the left-hand point on the cartographers' group response graph has been tied to the left-hand end point on the Cape Arago sample line. Because the two sample group response graphs are aligned, this shows that the left-hand point on the non-cartographers' graph is also associated with the left-hand end point on the Cape Arago line. As many graph points were connected with their associated sample line points in this manner as was possible without destroying the legibility for the reader.

Similarly, Figure 3.5 depicts the summarized responses for Cape Arago for stage two of the experiment. In this stage, each respondent could select no more than 36 characteristic points. Figure 3.6 illustrates the greatest amount of generalization, stage three of the test, in which respondents could select not more than 18 characteristic points along the sample line.

To facilitate display of the data, and to retain a consistent scale on each graph for all corresponding pairs, graphs showing the summarized responses for the

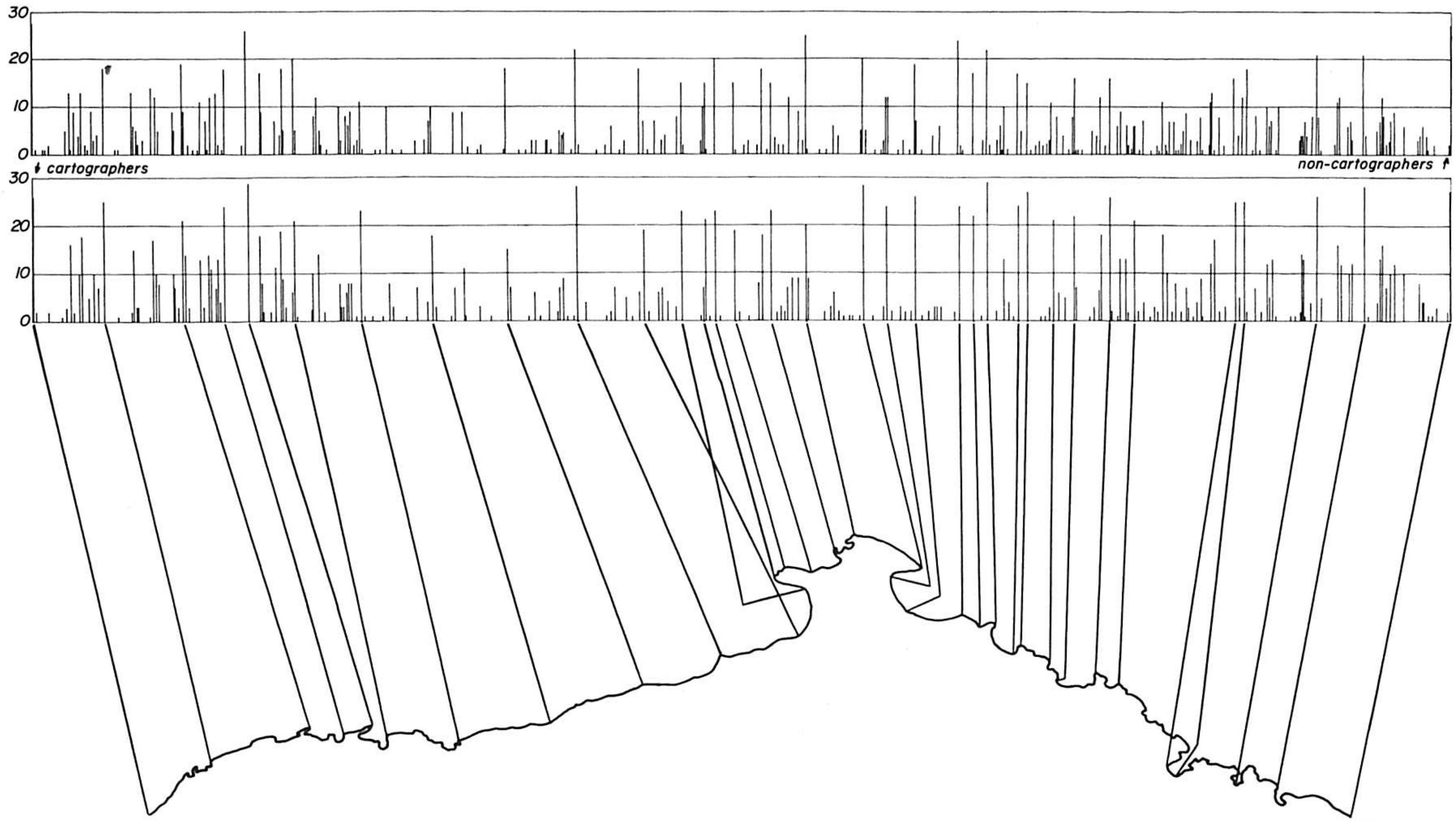


Figure 3.4. *Sample line is Cape Arago. Graphs of summarized responses for stage one (least generalization); 73 possible points could be selected by non-cartographers (top) and cartographers (below); r = 0.8849, 0.001 significance level.*

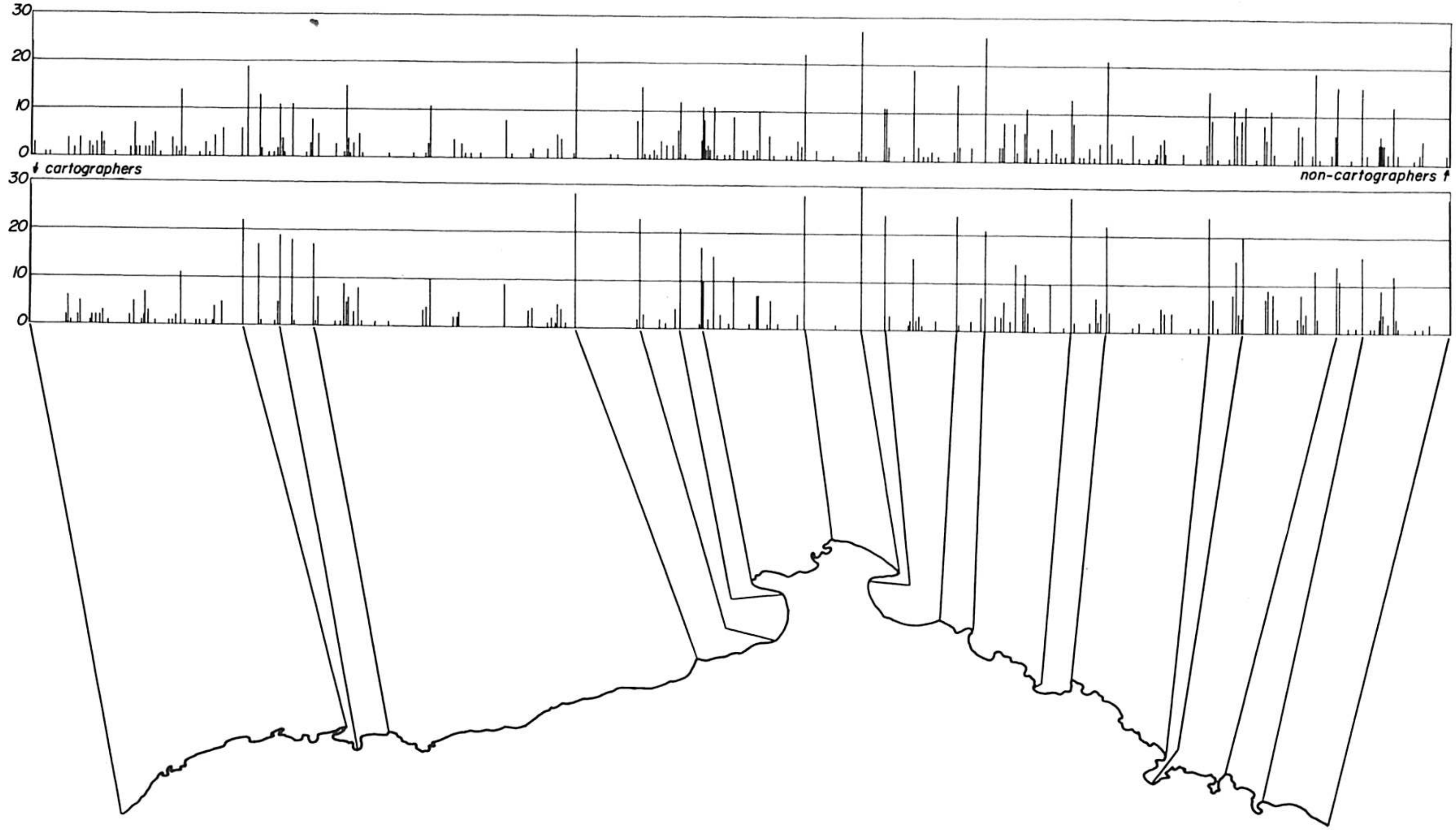


Figure 3.5. *Sample line is Cape Arago. Graphs of summarized responses for stage two; 36 possible points could be selected by non-cartographers (top) and cartographers (below); r = 0.9320, 0.001 significance level.*

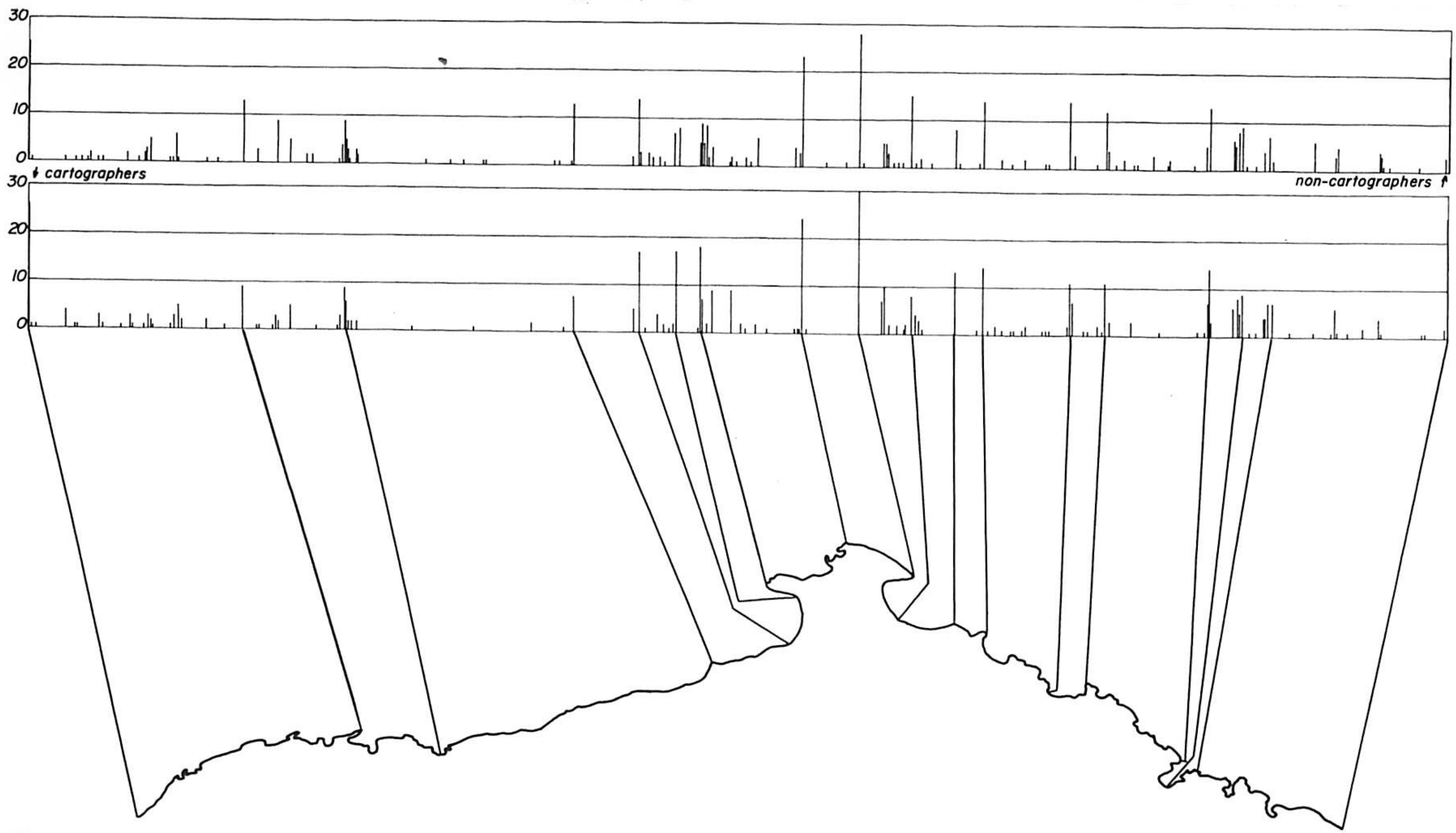


Figure 3.6. *Sample line is Cape Arago. Graphs of summarized responses for stage three (greatest generalization); 18 possible points could be selected by non-cartographers (top) and cartographers (below); r=0.9063, 0.001 significance level.*

sample line of Cape Hatteras have been divided into two parts. The first part coincides with half of the sample line; the second part of the graph matches the remaining half of the Cape Hatteras line.

For example, Figure 3.7 is the graph of first stage responses (116 possible points) for what in this case, is the upper half of the line symbolizing Cape Hatteras. Figure 3.8 is a continuation of the graph of first stage responses for the remaining portion of the Cape Hatteras line. Figures 3.9 and 3.10 should also be examined as a pair, for each shows half of the graphs representing points chosen in the second stage of the experiment, where a maximum of 58 points could be selected by a respondent. Similarly, Figures 3.11 and 3.12 comprise the total summarized responses for the third stage of the experiment, involving the greatest amount of generalization by the respondents who could select not more than 29 points along the sample line.

The graphs of summarized responses for the sample line of Chesapeake Bay have been divided into three portions. Figures 3.13, 3.14, and 3.15 each represent portions of the complete graphs of the summarized responses for stage one of the experiment, when 168 points were the maximum number which a respondent could have selected along the sample line. Likewise, Figures 3.16, 3.17,

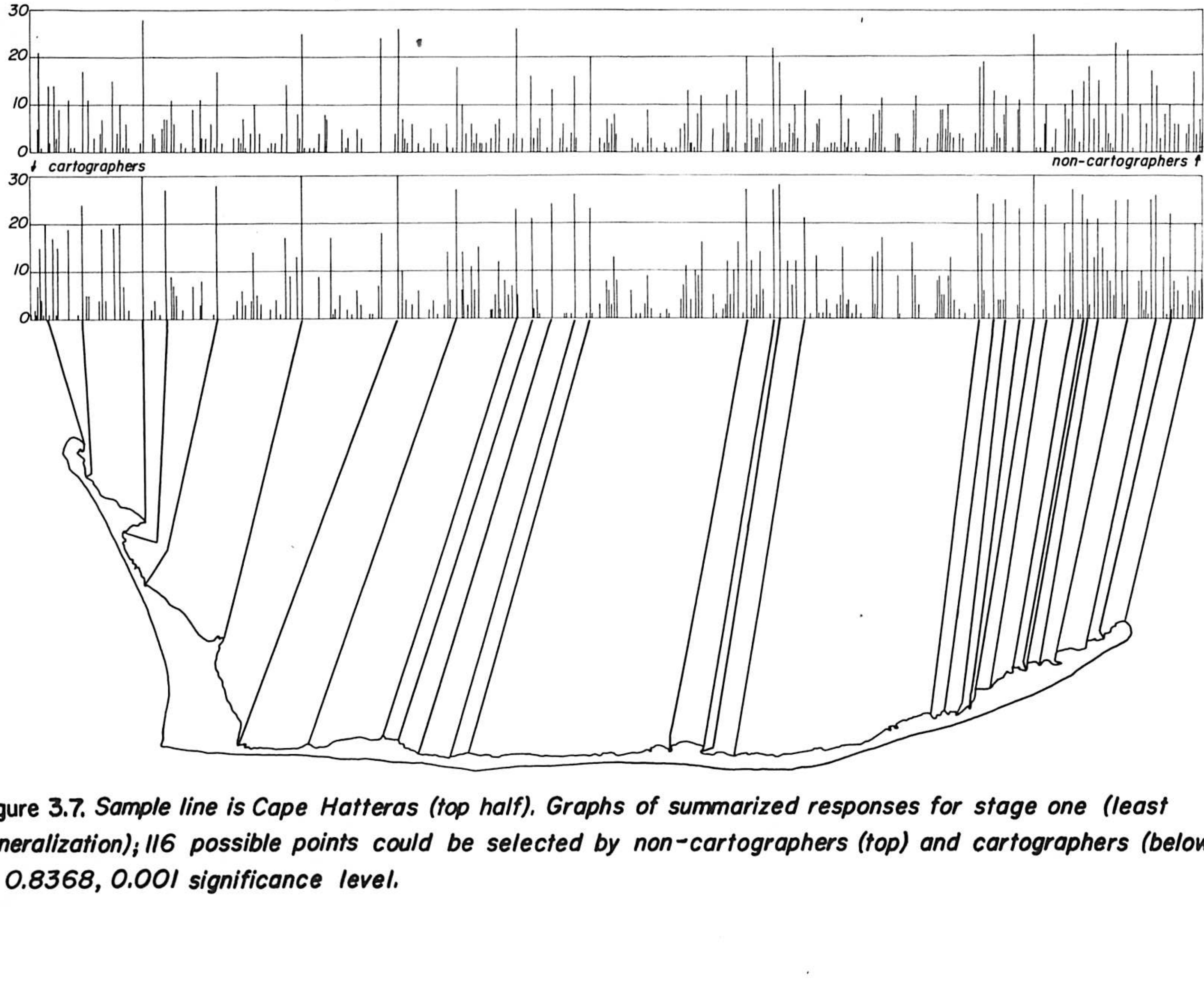


Figure 3.7. *Sample line is Cape Hatteras (top half). Graphs of summarized responses for stage one (least generalization); 116 possible points could be selected by non-cartographers (top) and cartographers (below); r = 0.8368, 0.001 significance level.*

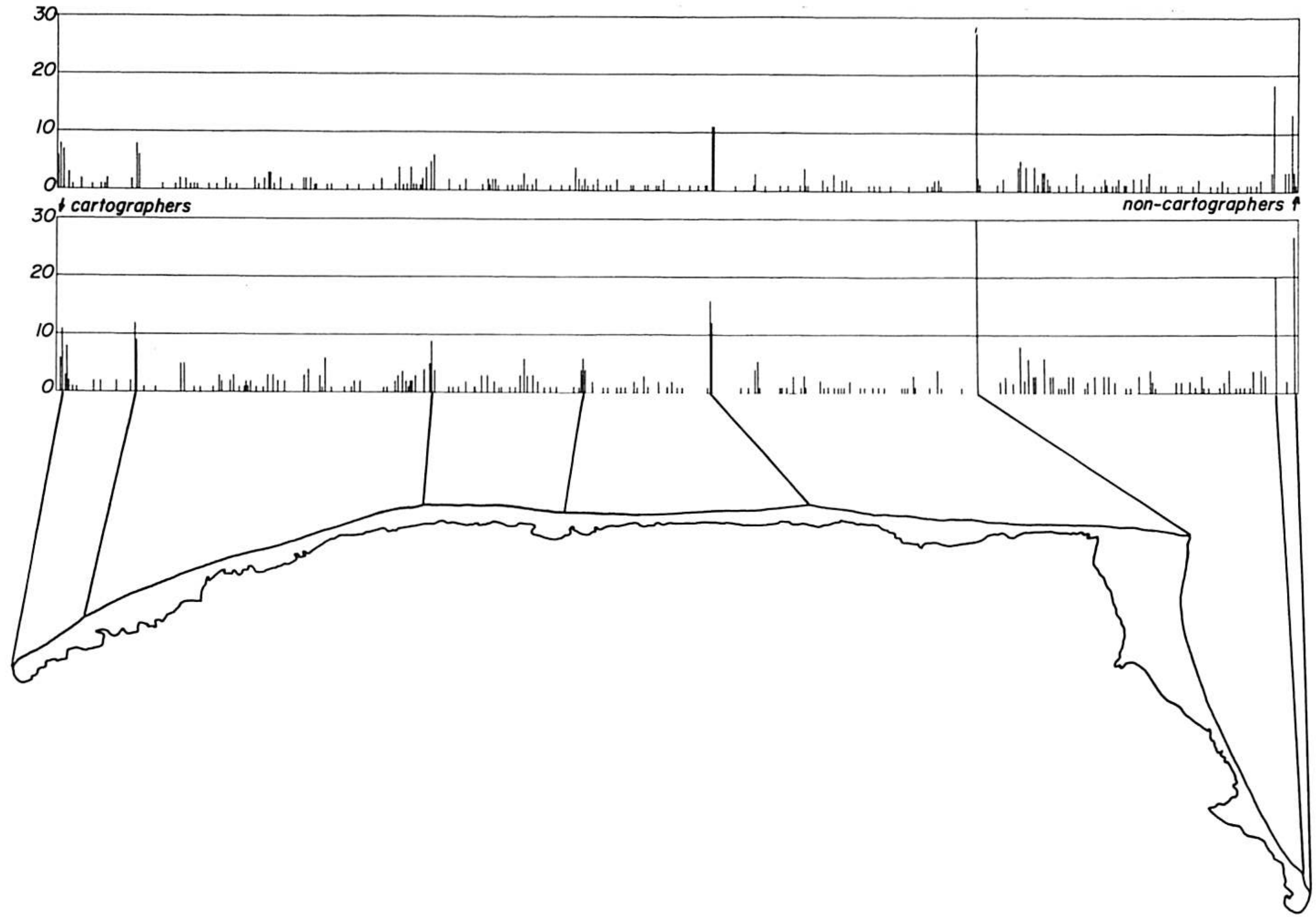


Figure 3.8. *Sample line is Cape Hatteras (bottom half). Graphs of summarized responses for stage one (least generalization); 116 possible points could be selected by non-cartographers (top) and cartographers (below); r = 0.8368, 0.001 significance level.*

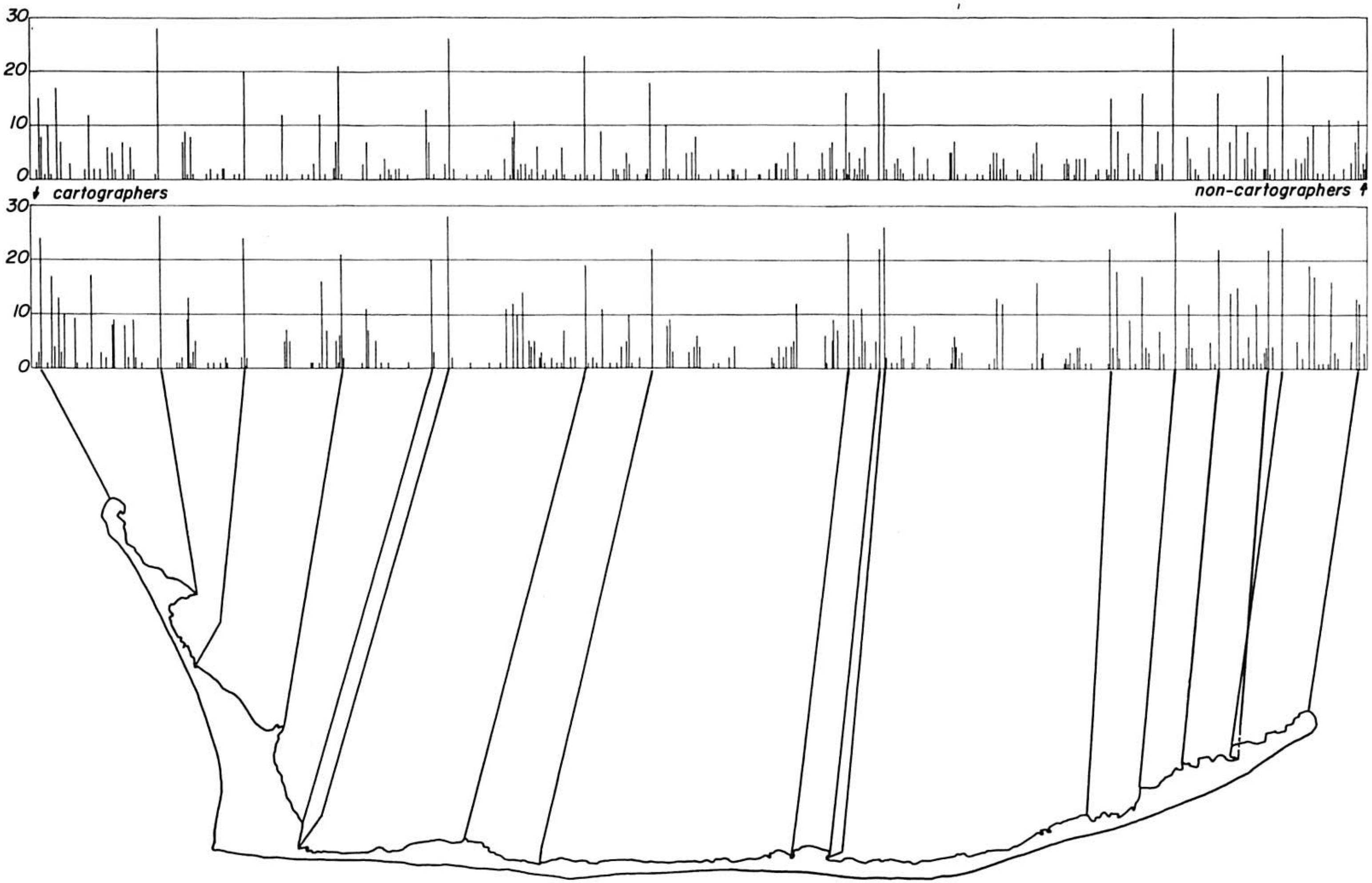


Figure 3.9. *Sample line is Cape Hatteras (top half). Graphs of summarized responses for stage two; 58 possible points could be selected by non-cartographers (top) and cartographers (below); r = 0.9096, 0.001 significance level.*

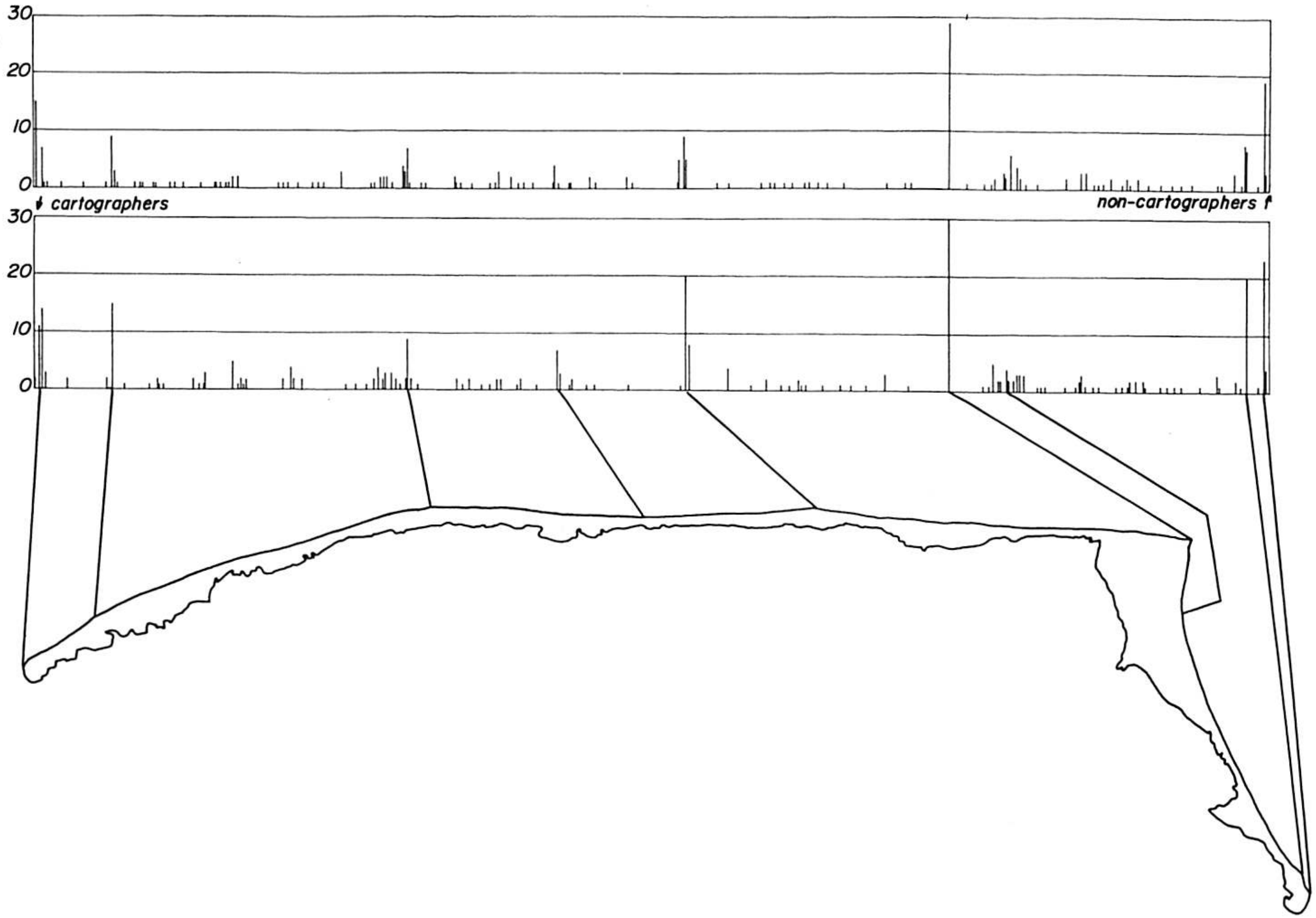


Figure 3.10. *Sample line is Cape Hatteras (bottom half). Graphs of summarized responses for stage two; 58 possible points could be selected by non-cartographers (top) and cartographers (below); r = 0.9096, 0.001 significance level.*

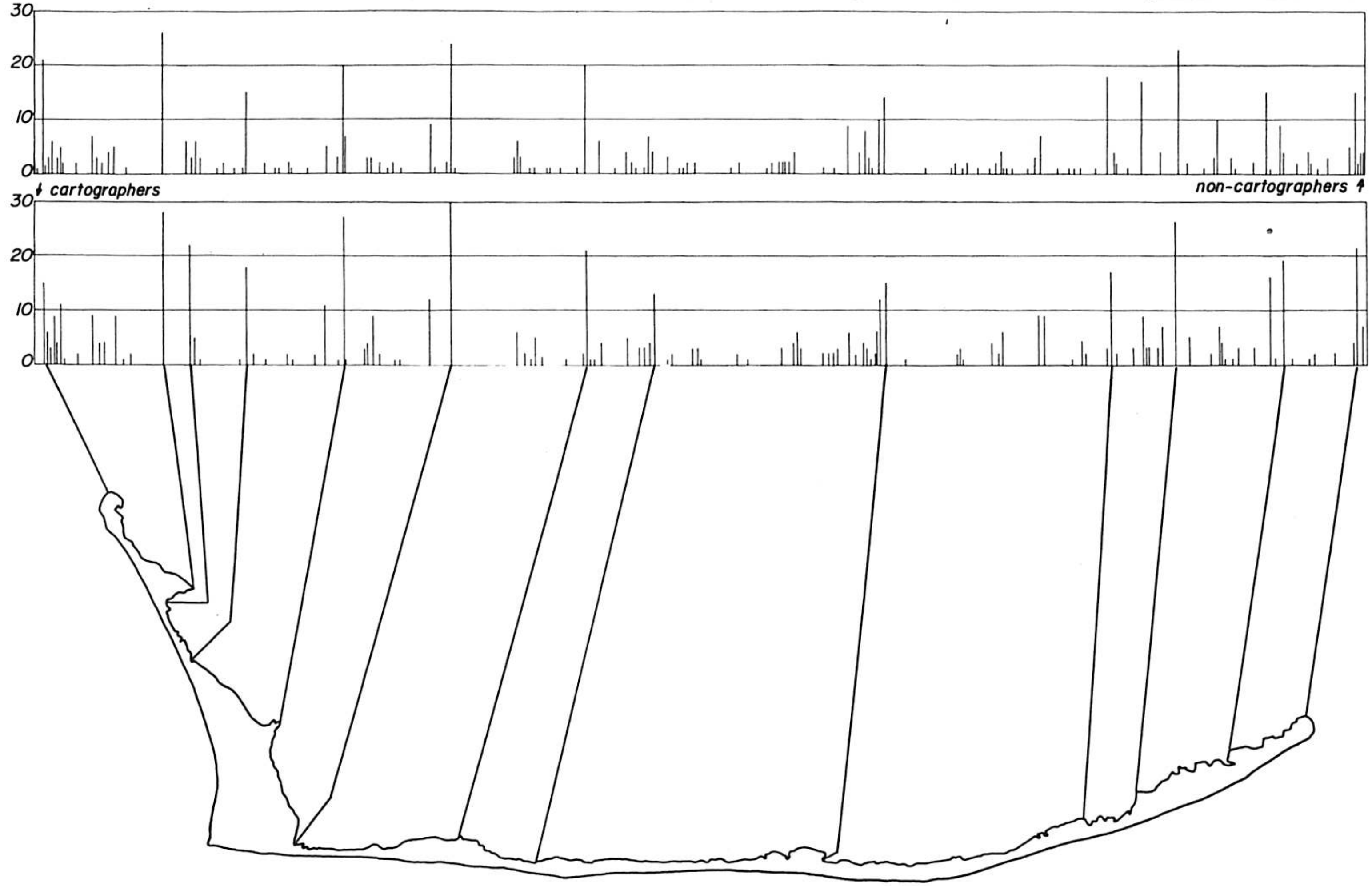


Figure 3.11. *Sample line is Cape Hatteras (top half). Graphs of summarized responses for stage three (greatest generalization); 29 possible points could be selected by non-cartographers (top) and cartographers (below); r = 0.8810, 0.001 significance level.*

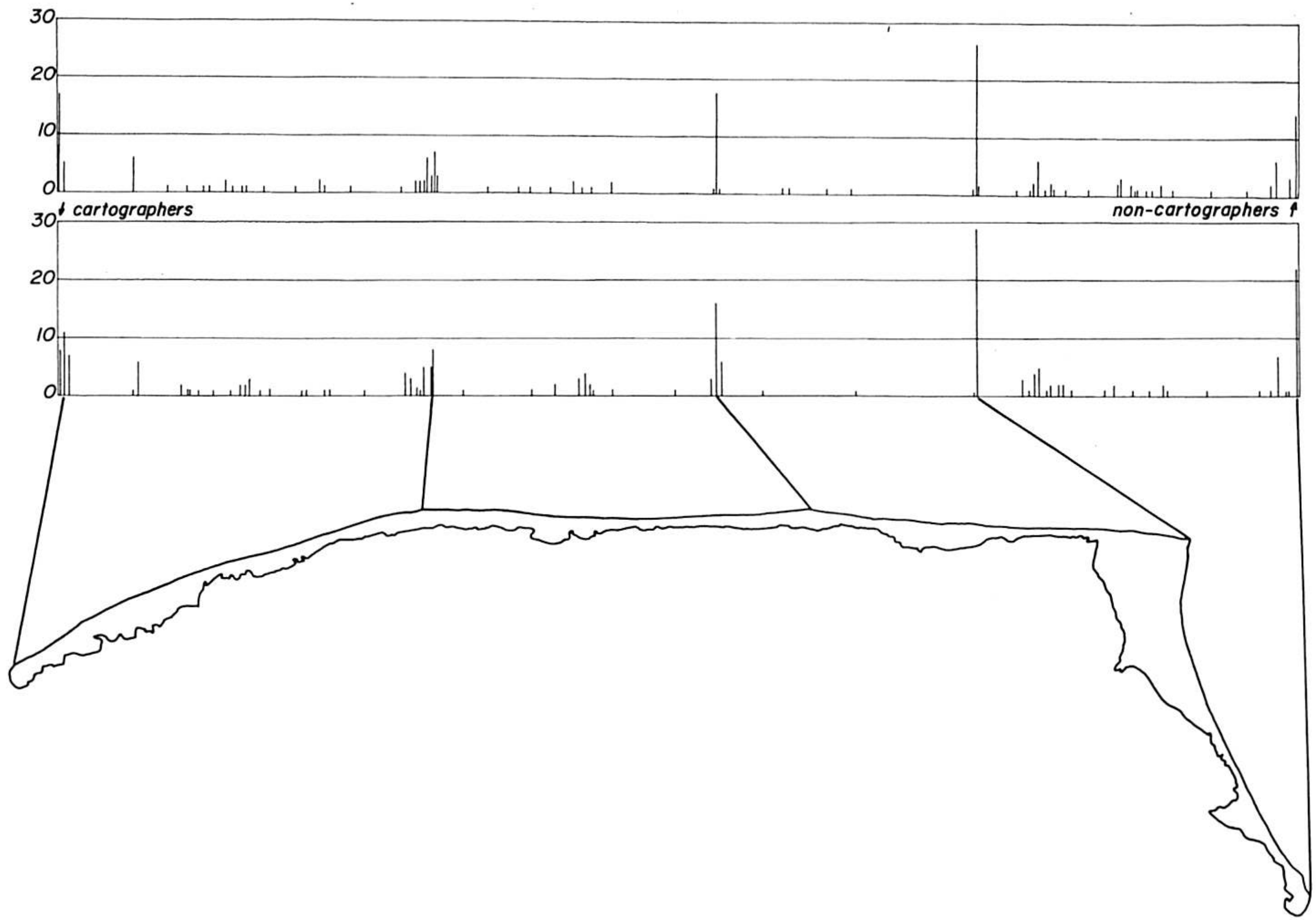


Figure 3.12. *Sample line is Cape Hatteras (bottom half). Graphs of summarized responses for stage three (greatest generalization); 29 possible points could be selected by non-cartographers (top) and cartographers (below); r = 0.8810, 0.001 significance level.*

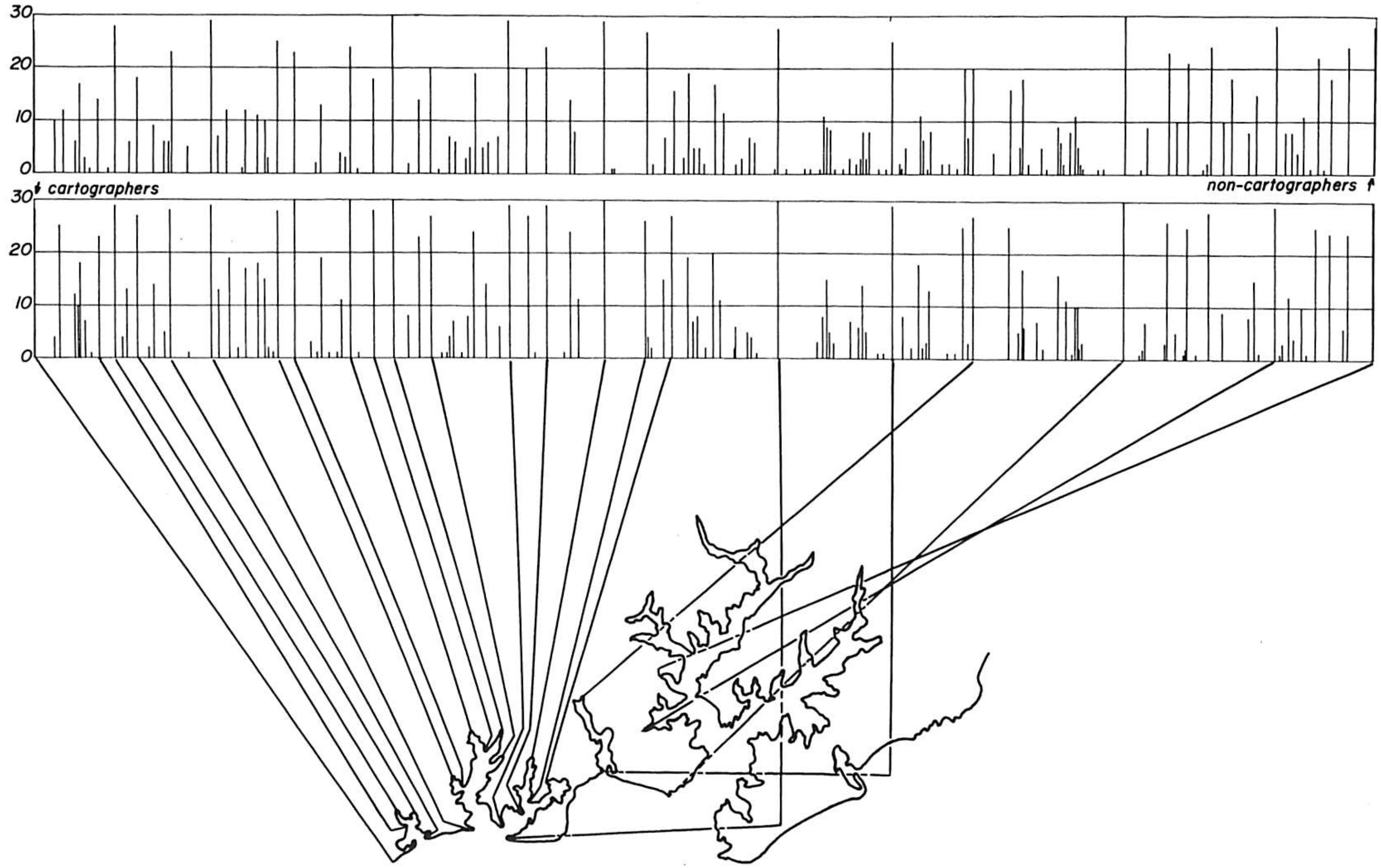


Figure 3.13. *Sample line is Chesapeake Bay (left third). Graphs of summarized responses for stage one (least generalization); 168 possible points could be selected by non-cartographers (top) and cartographers (below); r = 0.9452, 0.001 significance level.*

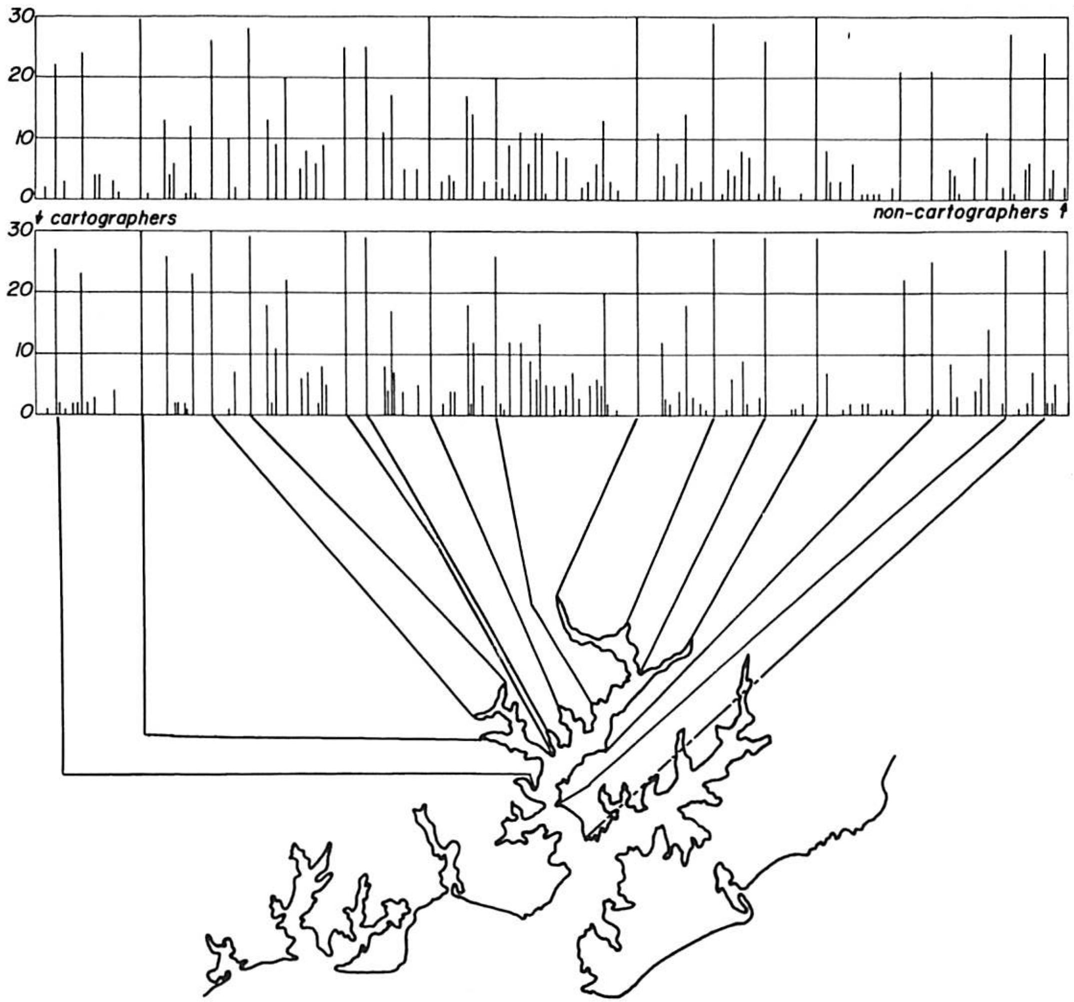


Figure 3.14. *Sample line is Chesapeake Bay (middle third). Graphs of summarized responses for stage one (least generalization); 168 possible points could be selected by non-cartographers (top) and cartographers (below); r = 0.9452, 0.001 significance level.*

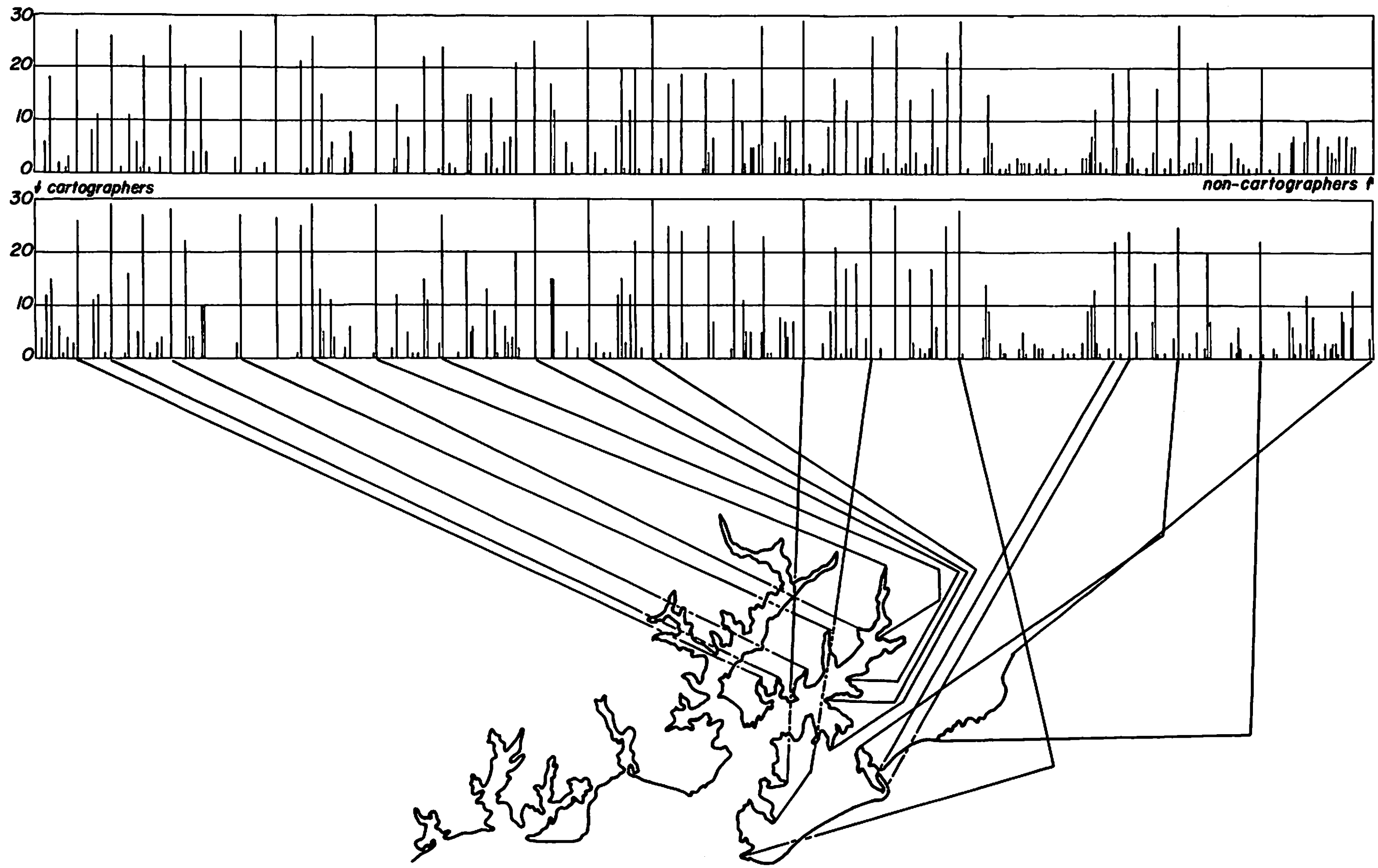


Figure 3.15. *Sample line is Chesapeake Bay (right third). Graphs of summarized responses for stage one (least generalization); 168 possible points could be selected by non-cartographers (top) and cartographers (below); r = 0.9452, 0.001 significance level.*

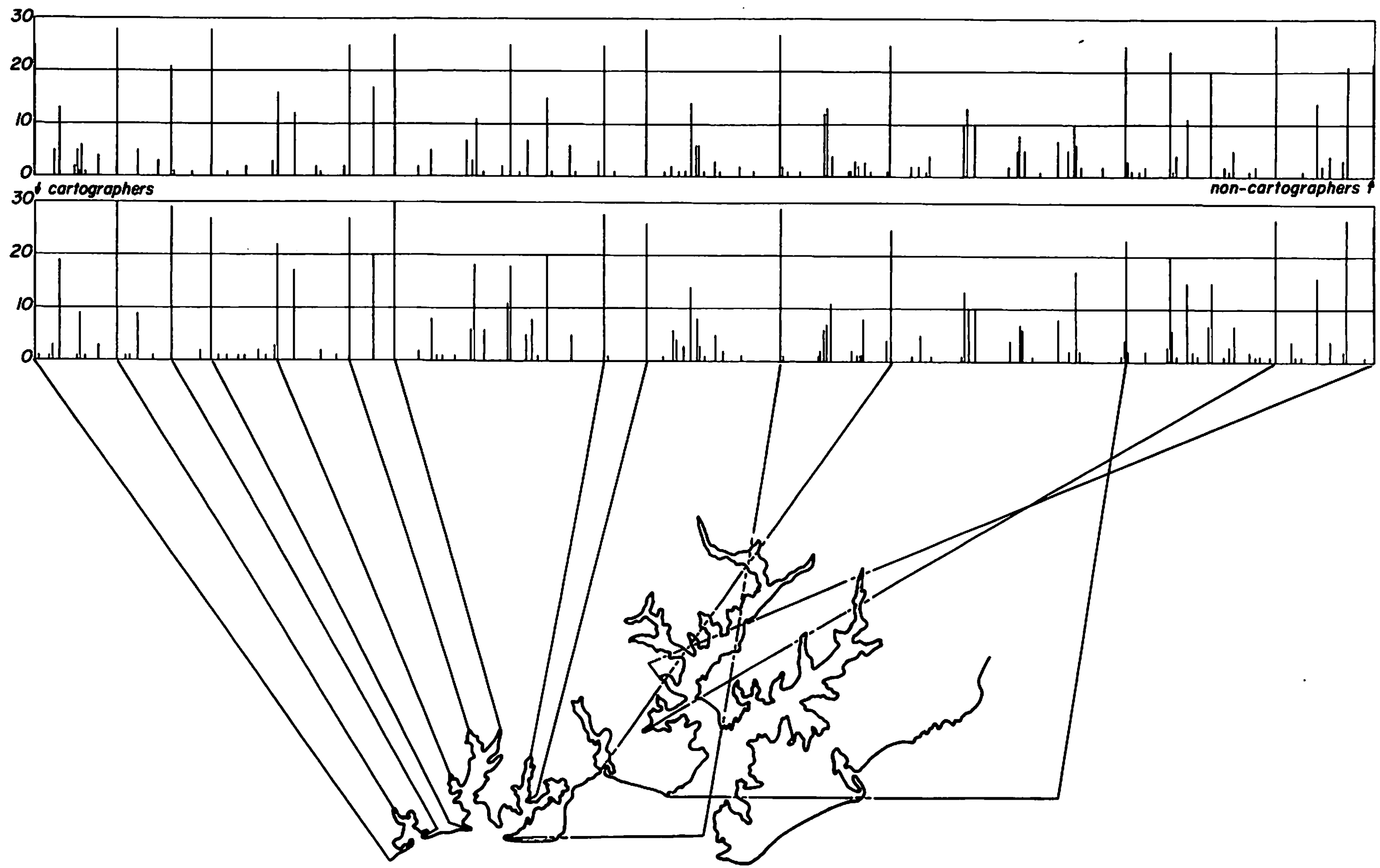


Figure 3.16. *Sample line is Chesapeake Bay (left third). Graphs of summarized responses for stage two; 84 possible points could be selected by non-cartographers (top) and cartographers (below); r=0.9530, 0.001 significance level.*

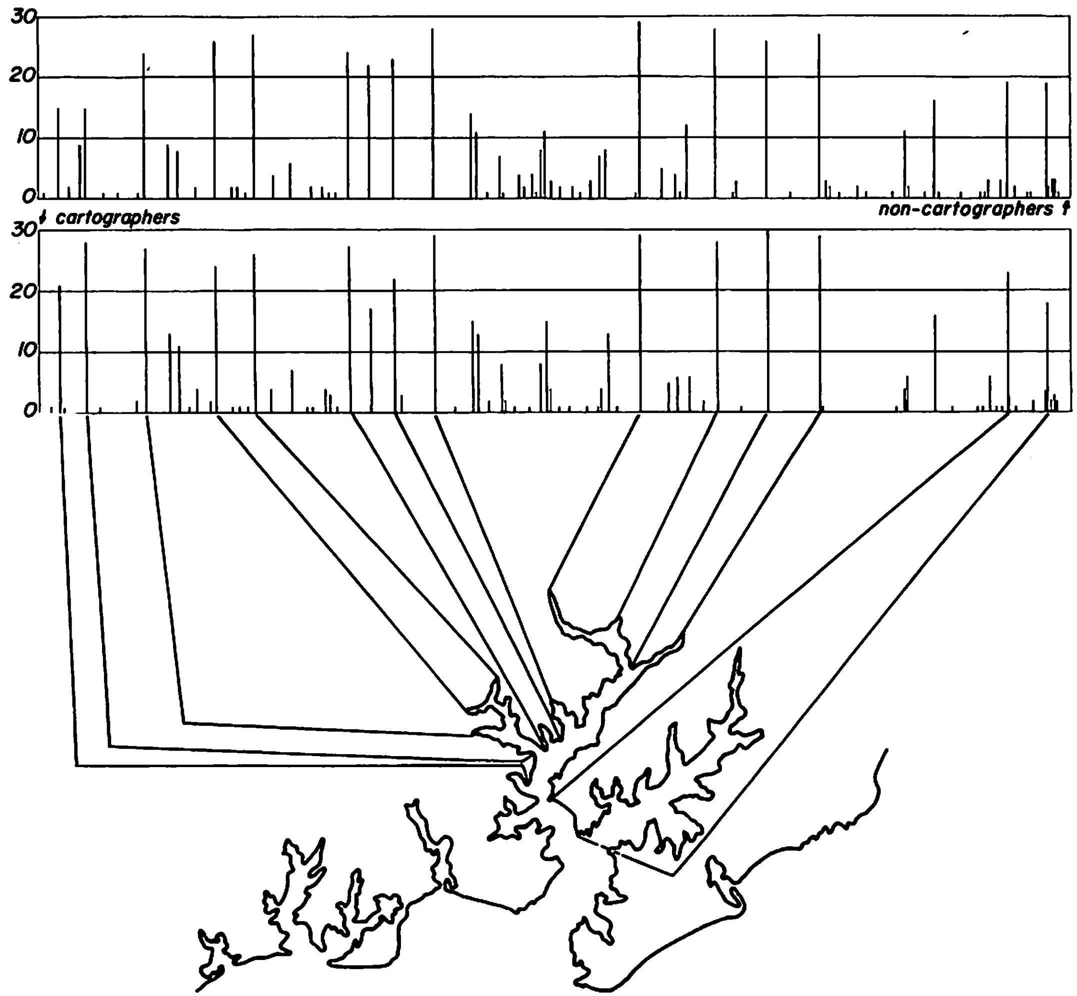


Figure 3.17. *Sample line is Chesapeake Bay (middle third). Graphs of summarized responses for stage two; 84 possible points could be selected by non-cartographers (top) and cartographers (below); r=0.9530, 0.001 significance level.*

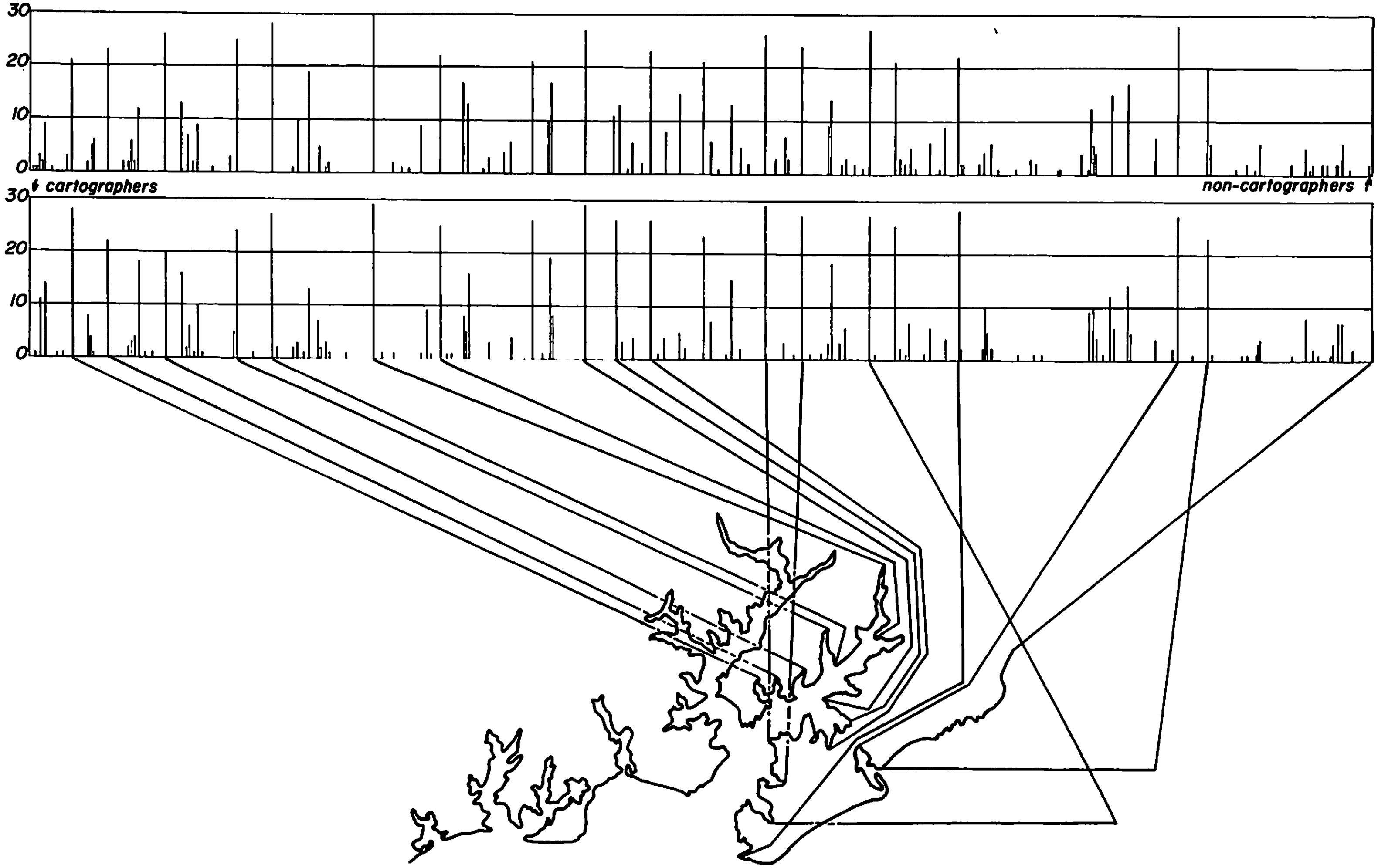


Figure 3.18. *Sample line is Chesapeake Bay (right third). Graphs of summarized responses for stage two; 84 possible points could be selected by non-cartographers (top) and cartographers (below); r=0.9530, 0.001 significance level.*

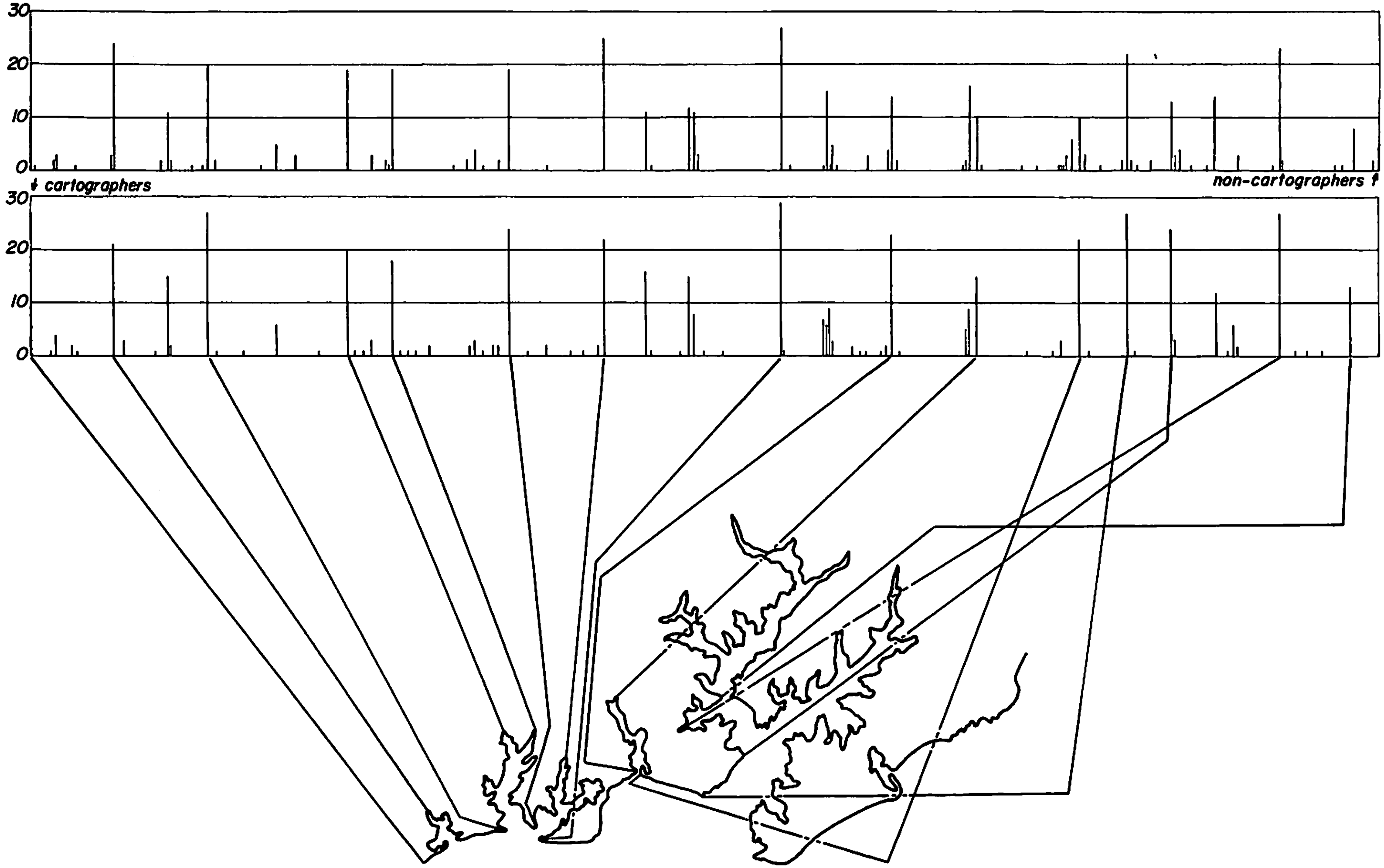


Figure 3.19. *Sample line is Chesapeake Bay (left third). Graphs of summarized responses for stage three (greatest generalization); 42 possible points could be selected by non-cartographers (top) and cartographers (below); r=0.5482, 0.001 significance level.*

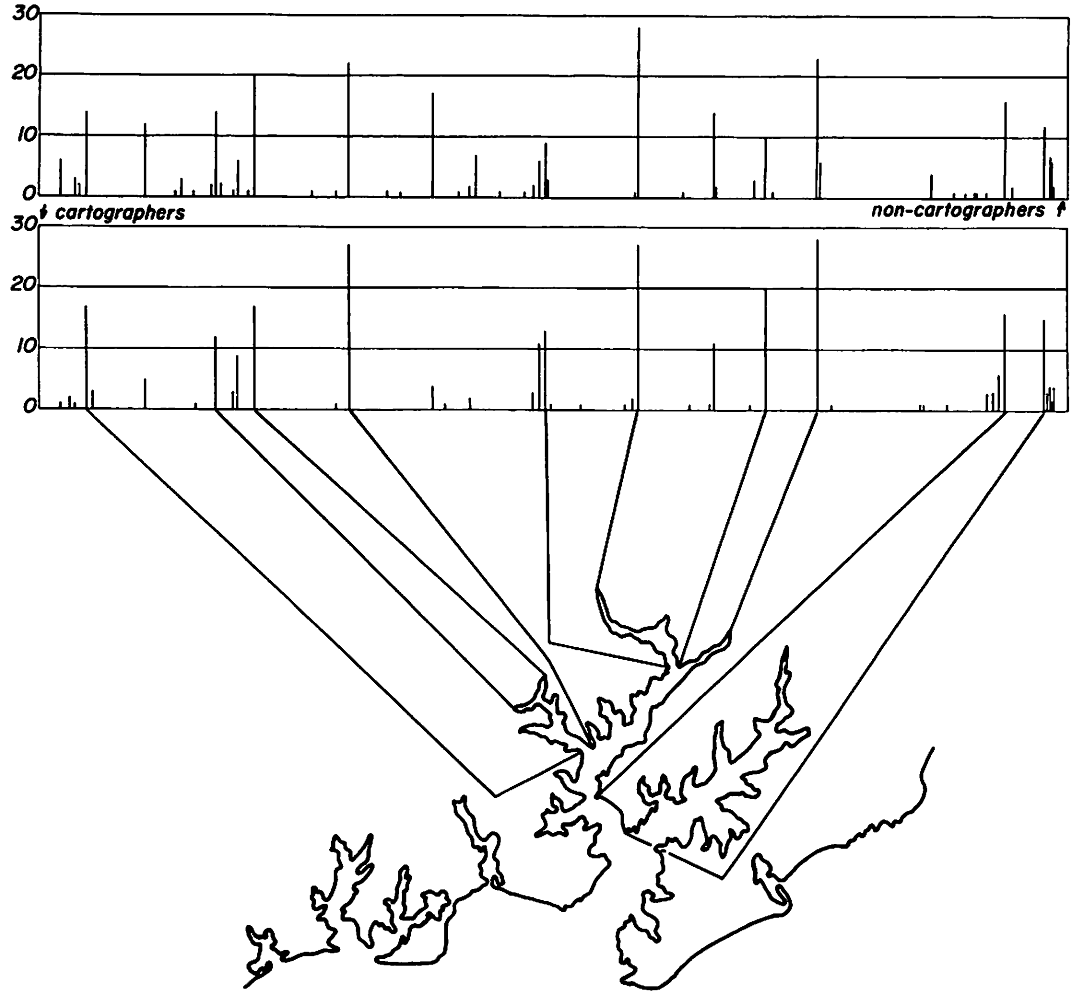


Figure 3.20. *Sample line is Chesapeake Bay (middle third). Graphs of summarized responses for stage three (greatest generalization); 42 possible points could be selected by non-cartographers (top) and cartographers (below); r = 0.5482, 0.001 significance level.*

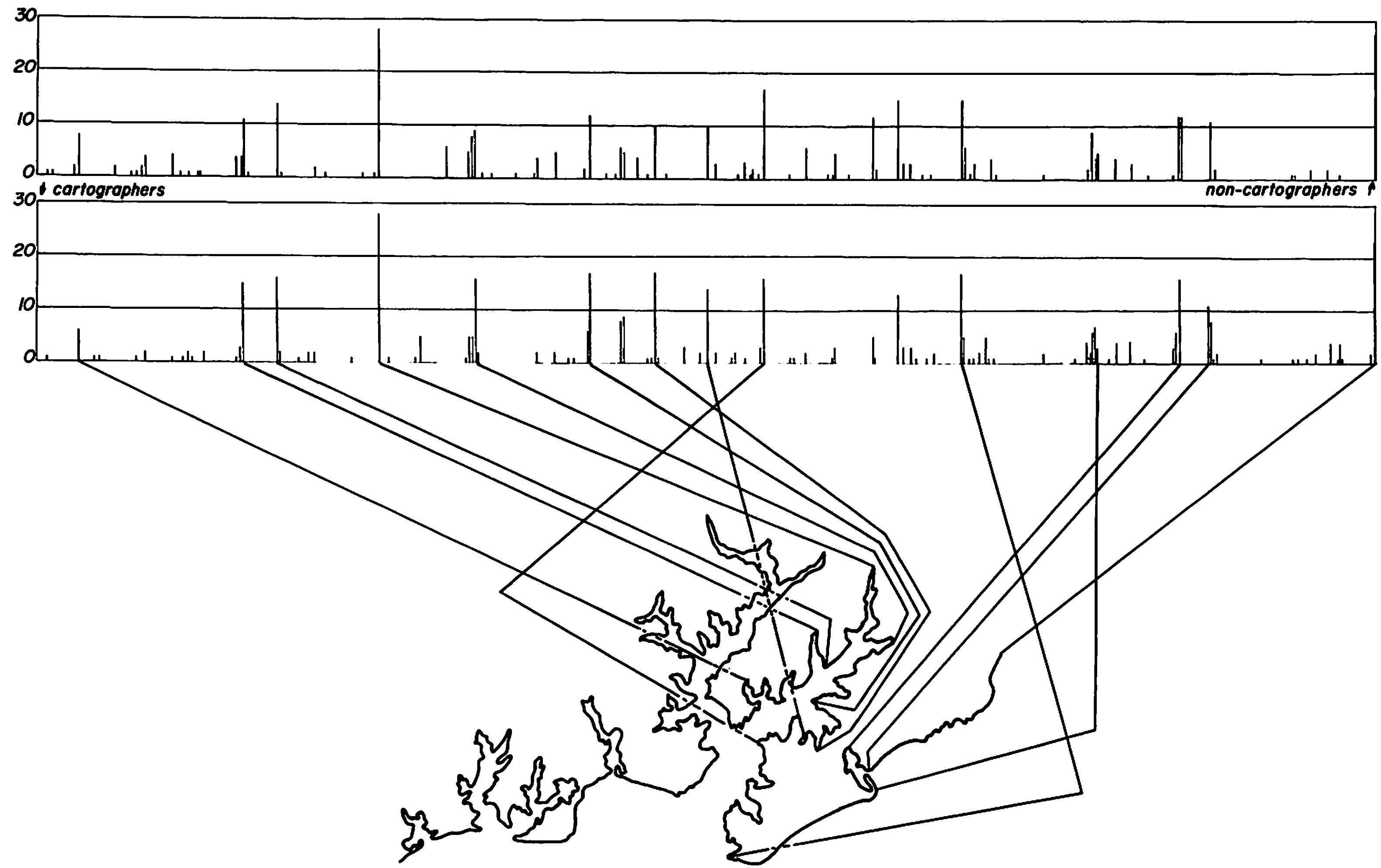


Figure 3.21. *Sample line is Chesapeake Bay (right third). Graphs of summarized responses for stage three (greatest generalization); 42 possible points could be selected by non-cartographers (top) and cartographers (below); r=0.5482, 0.001 significance level.*

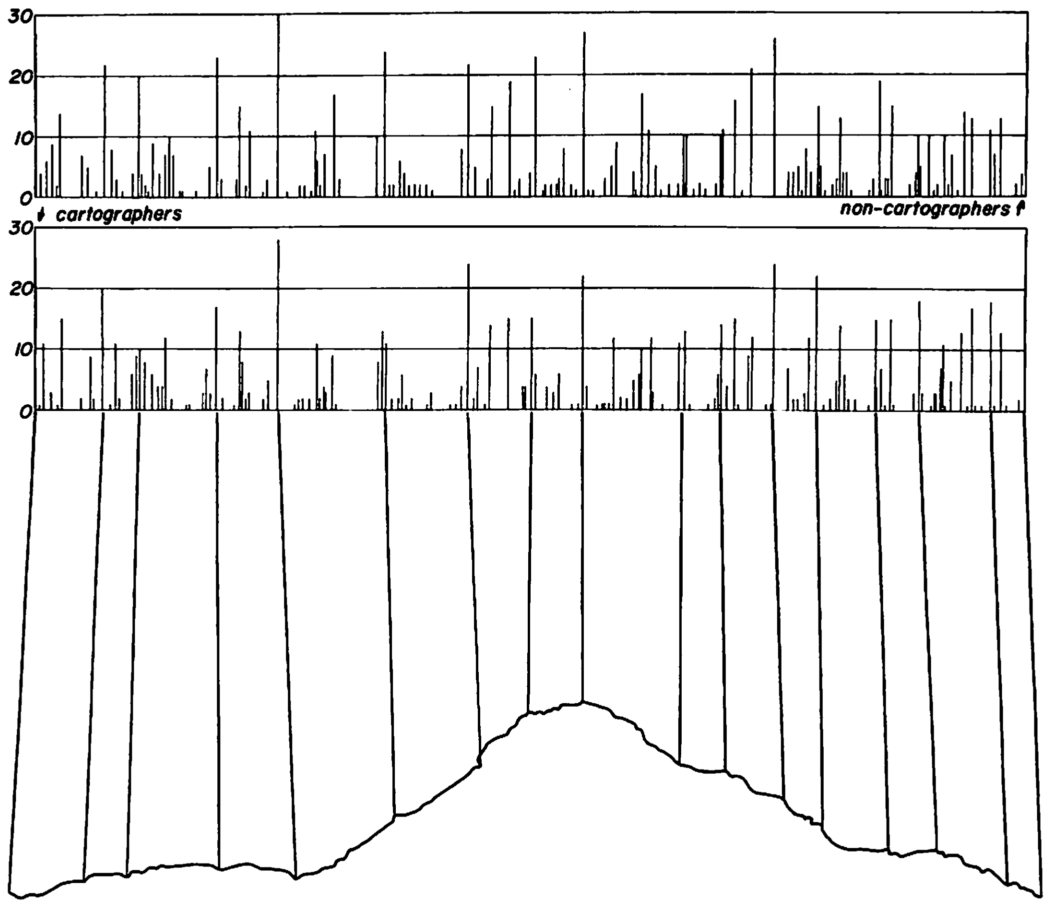


Figure 3.22. *Sample line is Fall River. Graphs of summarized responses for stage one (least generalization); 42 possible points could be selected by non-cartographers (top) and cartographers (below); r = 0.852, 0.001 significance level.*

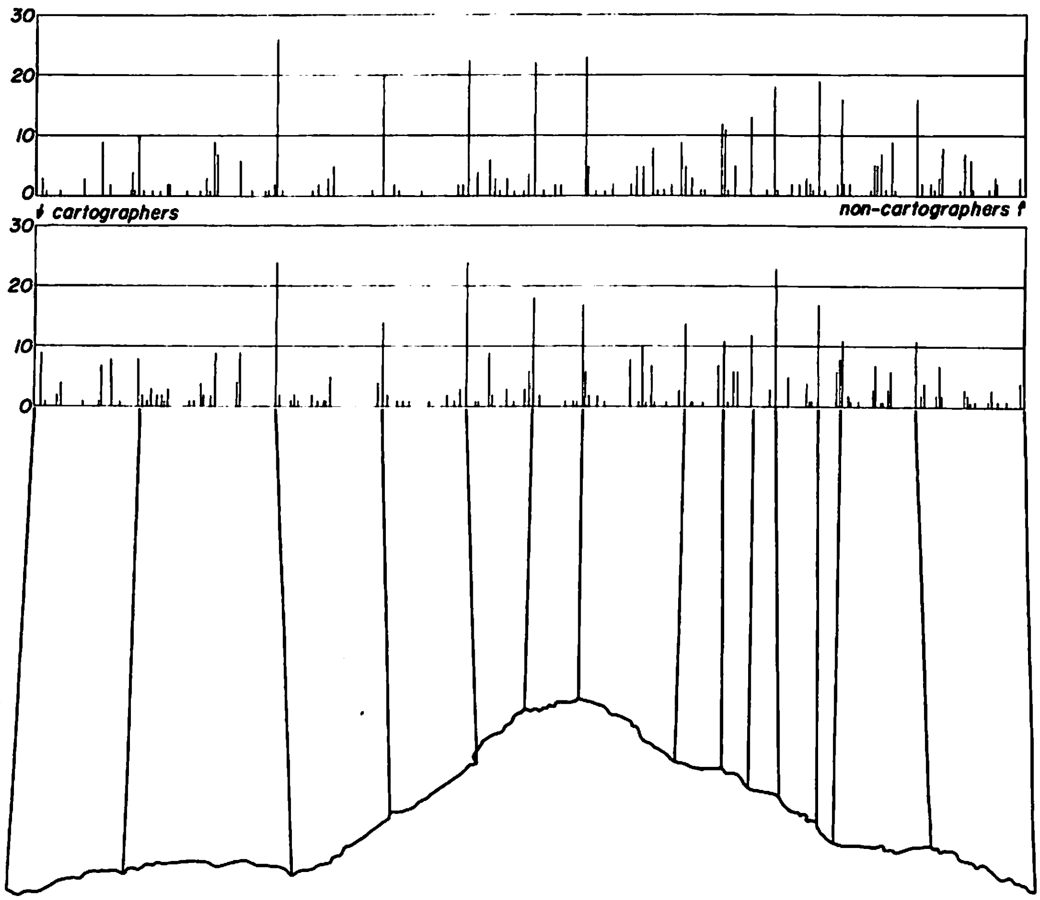


Figure 3.23. *Sample line is Fall River. Graphs of summarized responses for stage two; 20 possible points could be selected by non-cartographers (top) and cartographers (below); r=0.9089, 0.001 significance level.*

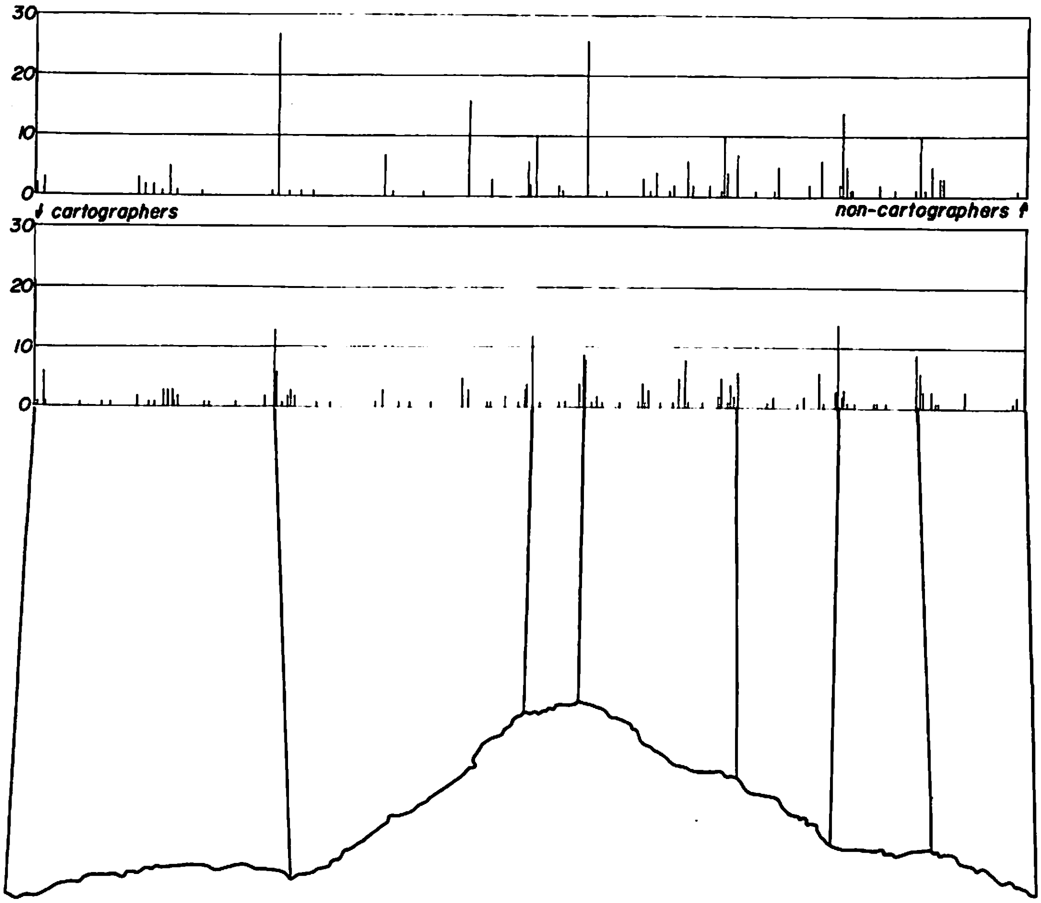


Figure 3.24. *Sample line is Fall River. Graphs of summarized responses for stage three (greatest generalization); 10 possible points could be selected by non-cartographers (top) and cartographers (below); r=0.795, 0.001 significance level.*

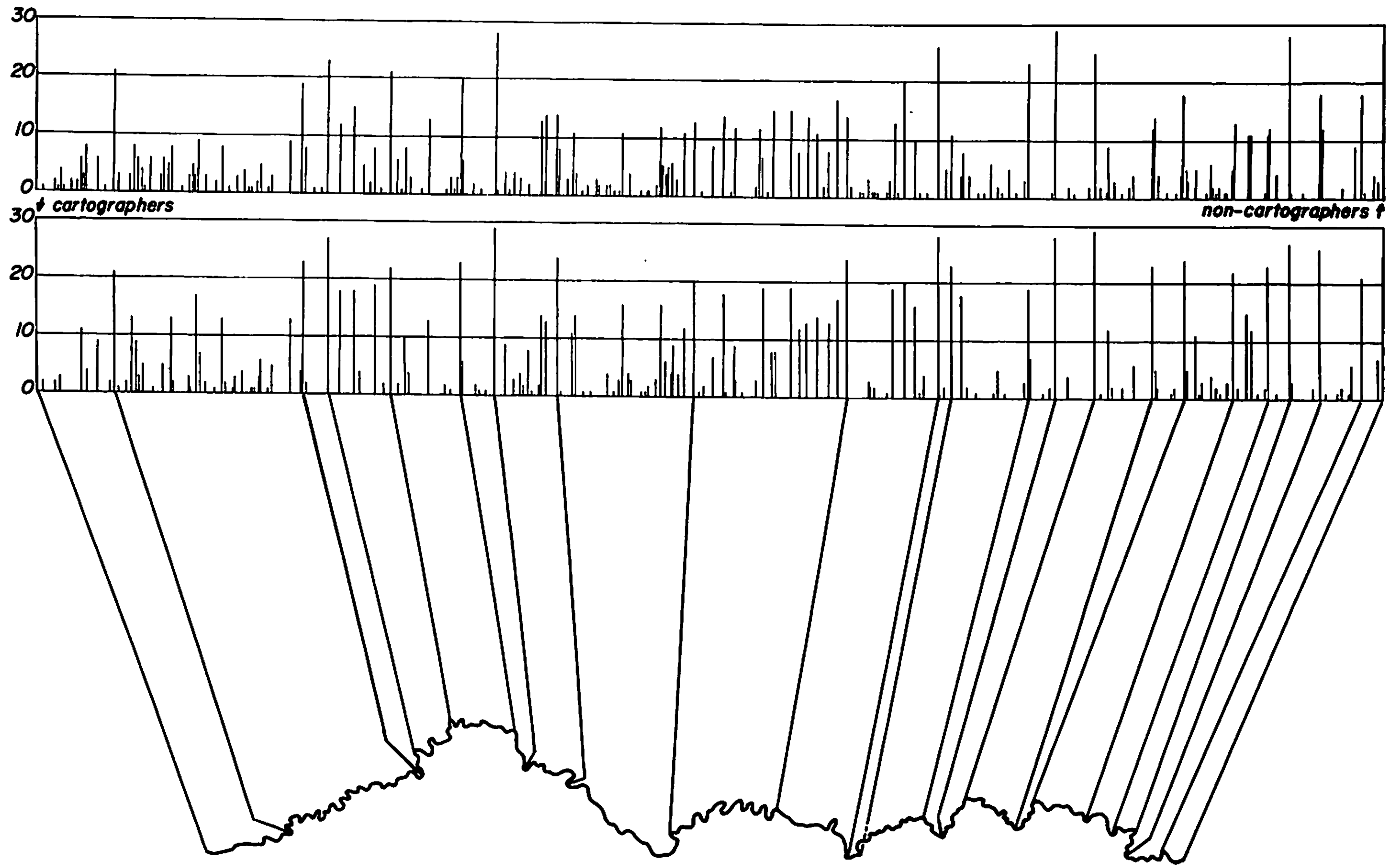


Figure 3.25. *Sample line is Mancos River. Graphs of summarized responses for stage one (least generalization); 56 possible points could be selected by non-cartographers (top) and cartographers (below); r=0.8779, 0.001 significance level.*

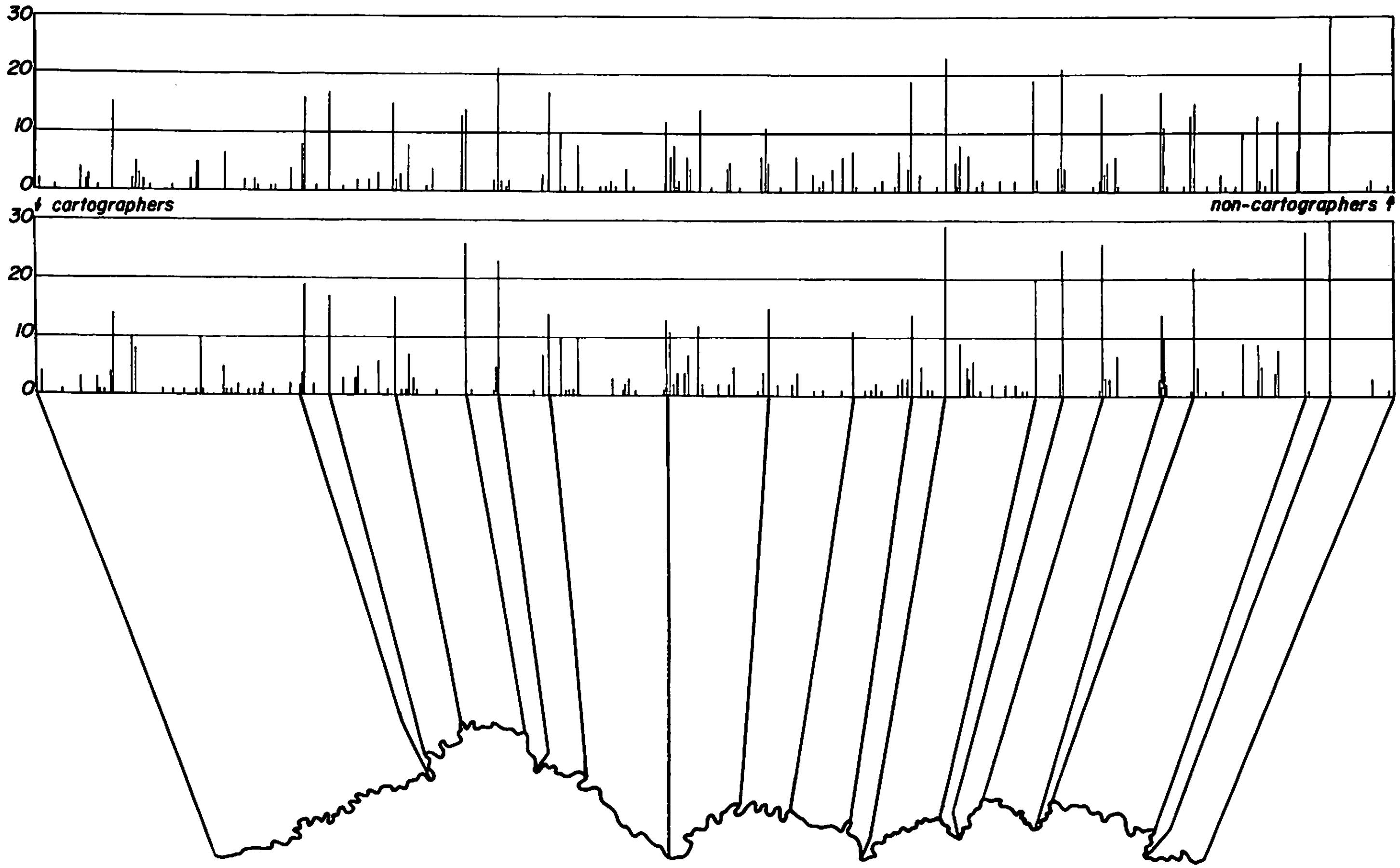


Figure 3.26. *Sample line is Mancos River. Graphs of summarized responses for stage two; 28 possible points could be selected by non-cartographers (top) and cartographers (below); r=0.9301, 0.001 significance level.*

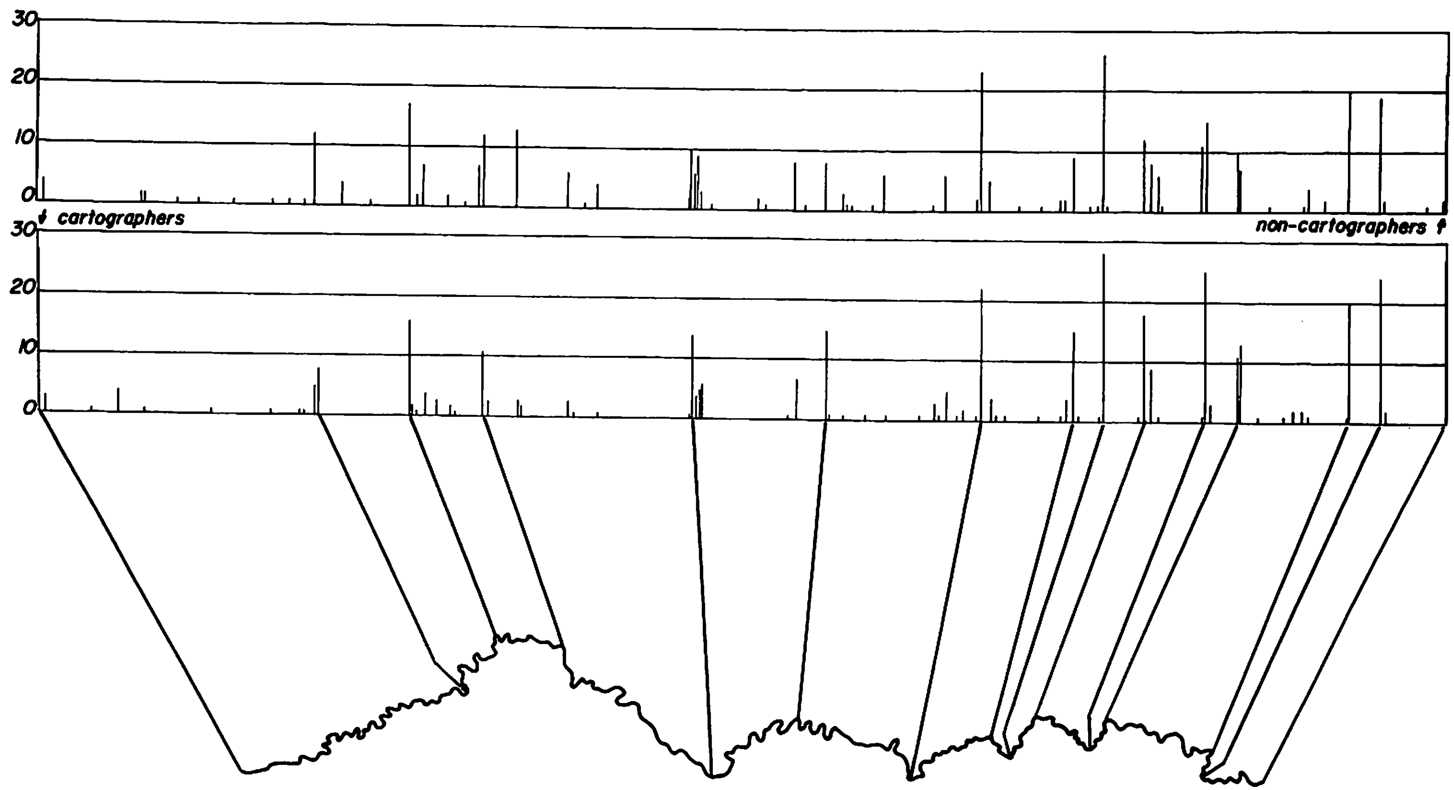


Figure 3.27. *Sample line is Mancos River. Graphs of summarized responses for stage three (greatest generalization); 14 possible points could be selected by non-cartographers (top) and cartographers (below); r=0.9071, 0.001 significance level.*

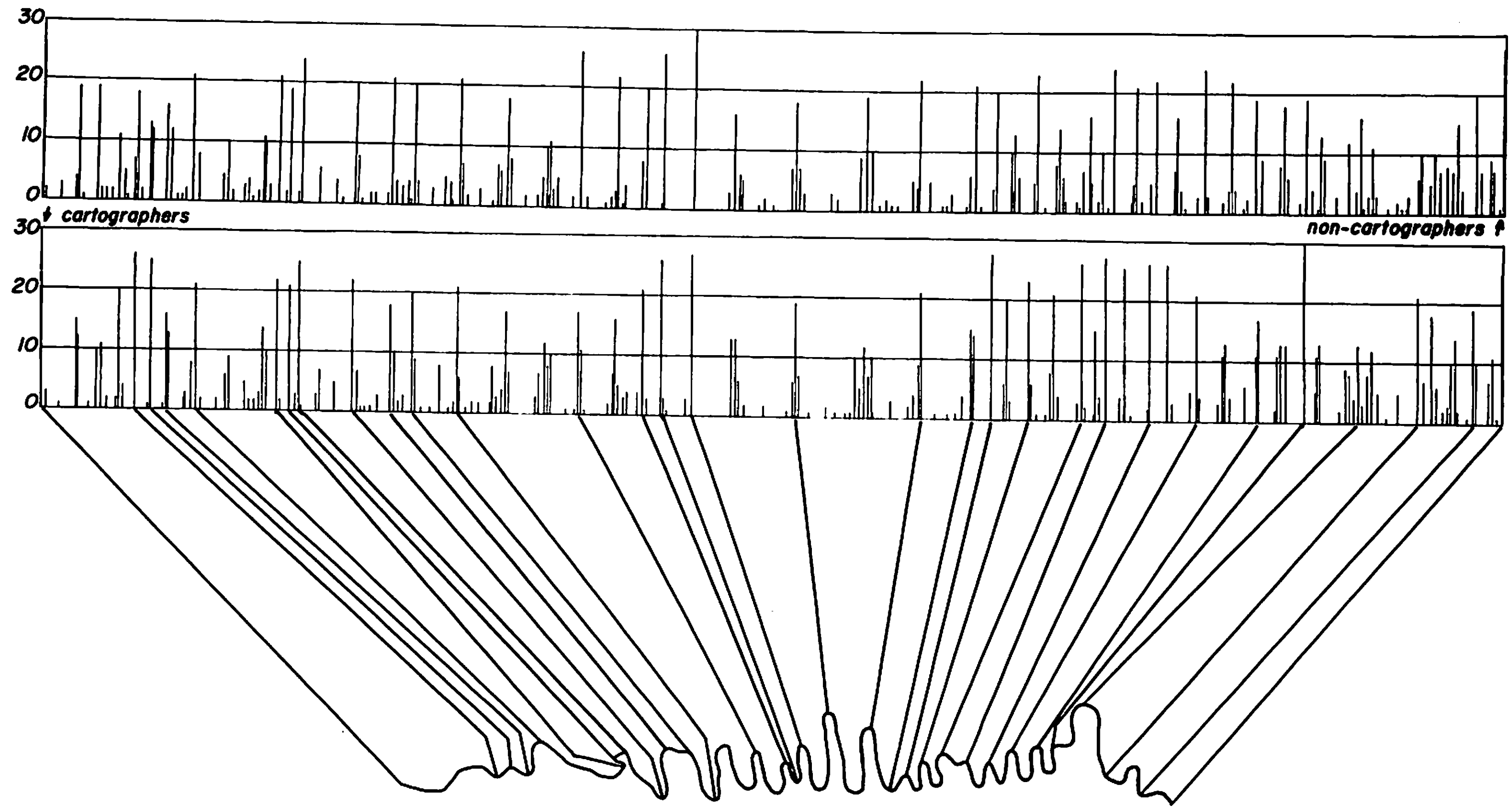


Figure 3.28. *Sample line is Shenandoah River. Graphs of summarized responses for stage one (least generalization); 66 possible points could be selected by non-cartographers (top) and cartographers (below); r=0.6830, 0.001 significance level.*

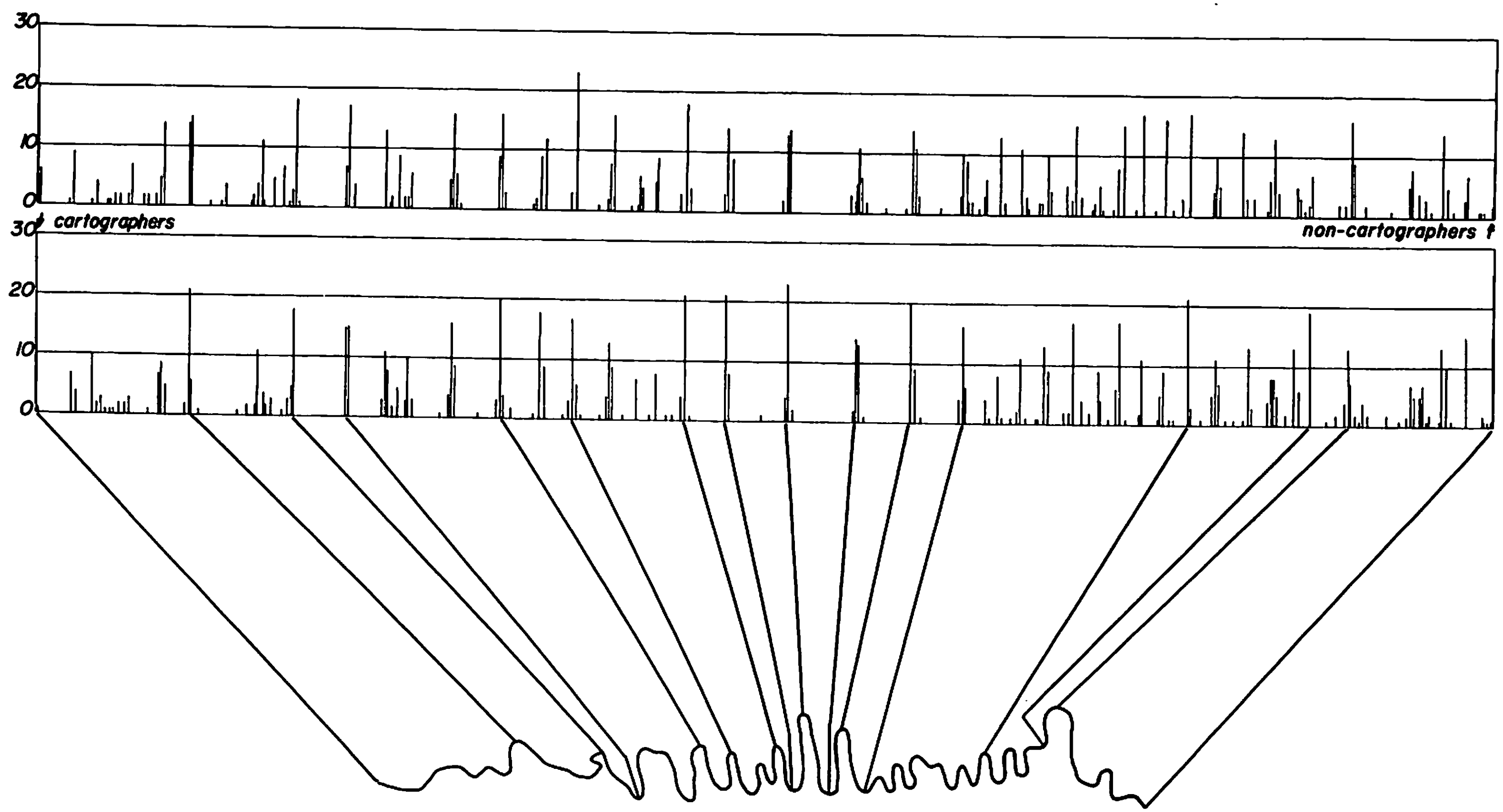


Figure 3.29. *Sample line is Shenandoah River. Graphs of summarized responses for stage two; 34 possible points could be selected by non-cartographers (top) and cartographers (below); r=0.8861, 0.001 significance level.*

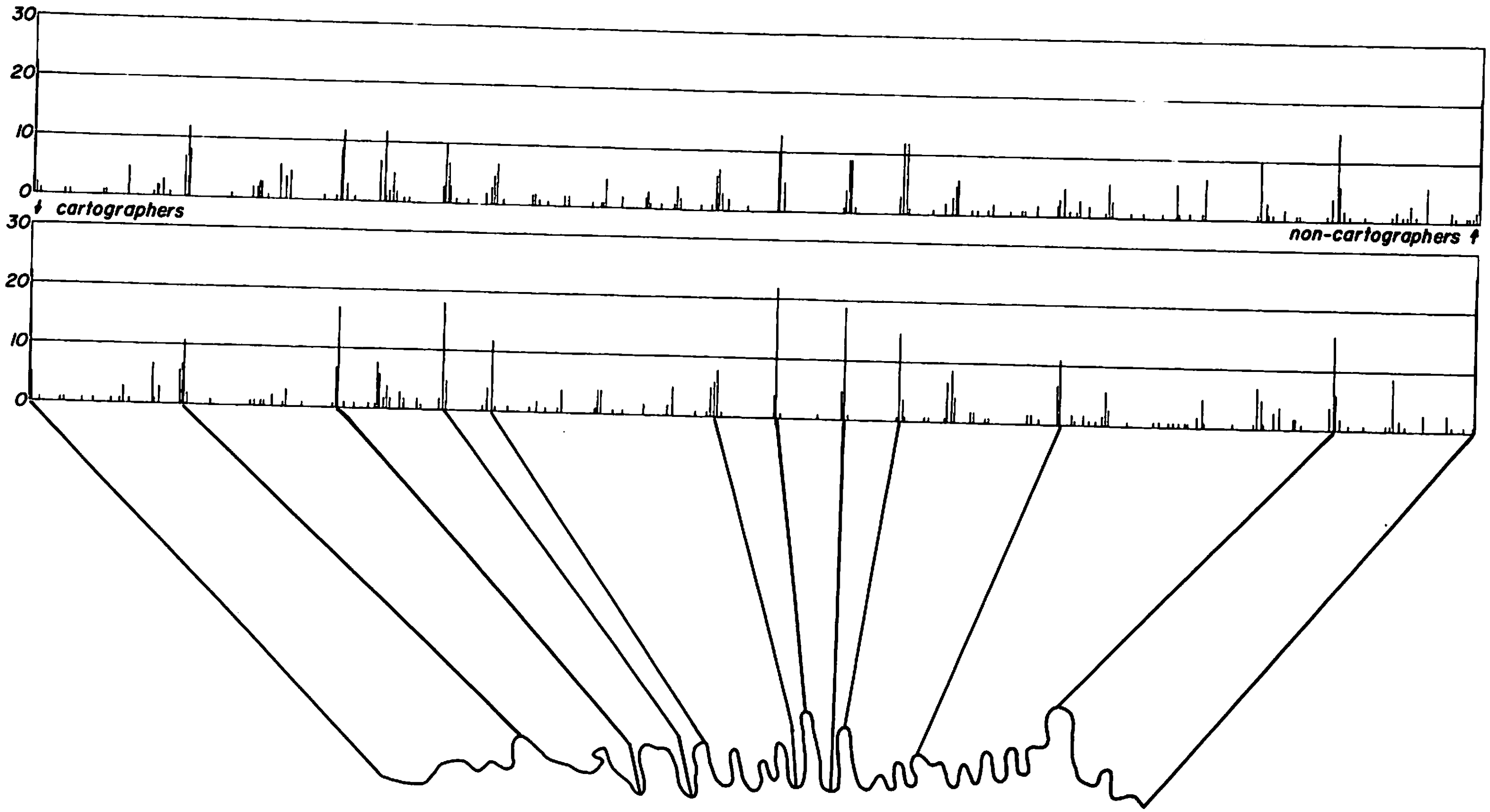


Figure 3.30. *Sample line is Shenandoah River. Graphs of summarized responses for stage three (greatest generalization); 17 possible points could be selected by non-cartographers (top) and cartographers (below); r=0.8992, 0.001 significance level.*

and 3.18 portray the responses given to stage two of the experiment (a maximum of 84 possible points). This corresponding pair showed the highest degree of correlation, with a computed r value of 0.953. Figures 3.19, 3.20, and 3.21 summarize the responses given in the third stage of the experiment, when a maximum number of 42 points could have been selected by each individual (the highest degree of generalization). This corresponding pair showed the lowest degree of correlation, with a computed r value of 0.5482.

The graphs of responses for the sample line of the Fall River are shown in Figures 3.22, 3.23, and 3.24; those for the Mancos River sample line, in Figures 3.25, 3.26, and 3.27; response graphs for the sample line of the Shenandoah River are shown in Figures 3.28, 3.29, and 3.30. These graphs are complete on each page. For example, Figure 3.22 illustrates the responses for stage one of the experiment; Figure 3.23 portrays the results of stage two; and Figure 3.24 shows the points selected in stage three of the experiment on the sample line of the Fall River.

Points Selected by a Majority of Each Group

The extent of agreement existing within each group's response was evaluated using the majority opinion of each group. The majority opinion refers to the circumstance in which more than half of the sample group selected the same point.[2] As the number of points selected by a majority increases, so does the amount of clustering of selected points. On the other hand, few or no points selected by a majority implies less clustering, and a tendency towards a uniformly even distribution of selected points along the sample line.

From the graphs of corresponding pairs, one can isolate and compare points selected by a majority of either group. With three exceptions, the cartographers, as a majority, selected a greater number of points than did the non-cartographers. The exceptions occur for Cape Arago, stage three, and for the Fall River, stages two and three. However, a difference of proportions test statistic computed for each matching pair of majority responses reveals that differences between the two groups of respondents are not significant, except in the stage one responses for Cape Hatteras and Chesapeake Bay, where the proportion of segments selected by a majority of cartographers is significantly different than that selected by a majority of non-cartographers. One will notice that these sample lines are the second, and the

---

[2] Note that the word point used in this context refers to segments less than 0.05 inch (0.125 cm) of sample line, the minimum interval on which location of selected points can be considered accurate.

greatest in length, respectively. (See Appendix D for a complete description.)

Figures 3.31 through 3.38 reveal points (segments < 0.05 inch or < 0.125 cm) which were selected by a majority of each group. For example, Figure 3.31 shows, on the far left, the segments selected by at least 16 of the 30 cartographers who responded to this sample line of Cape Arago, stage one. Second to the left, those segments selected by more than 15 of the 30 non-cartographers responding on Cape Arago, stage one, are marked. The two lines in the middle illustrate the segments selected in stage two (again, cartographers' choices are to the left). The two lines at the right show which segments were selected in stage three for this sample line of Cape Arago (non-cartographers' choices are on the far right).

Although the number of segments selected by a majority of cartographers is greater than for the other group, the proportion of segments selected by each group is not significantly different, except in two cases.

The two cases in which a significant difference exists between the proportion of segments selected by a majority of both groups, are found in Figures 3.32, the stage one response for Cape Hatteras; and in Figure 3.35, the stage one response for Chesapeake Bay.

An inspection of Figures 3.31 through 3.38 will reveal that points selected in stage three of the responses

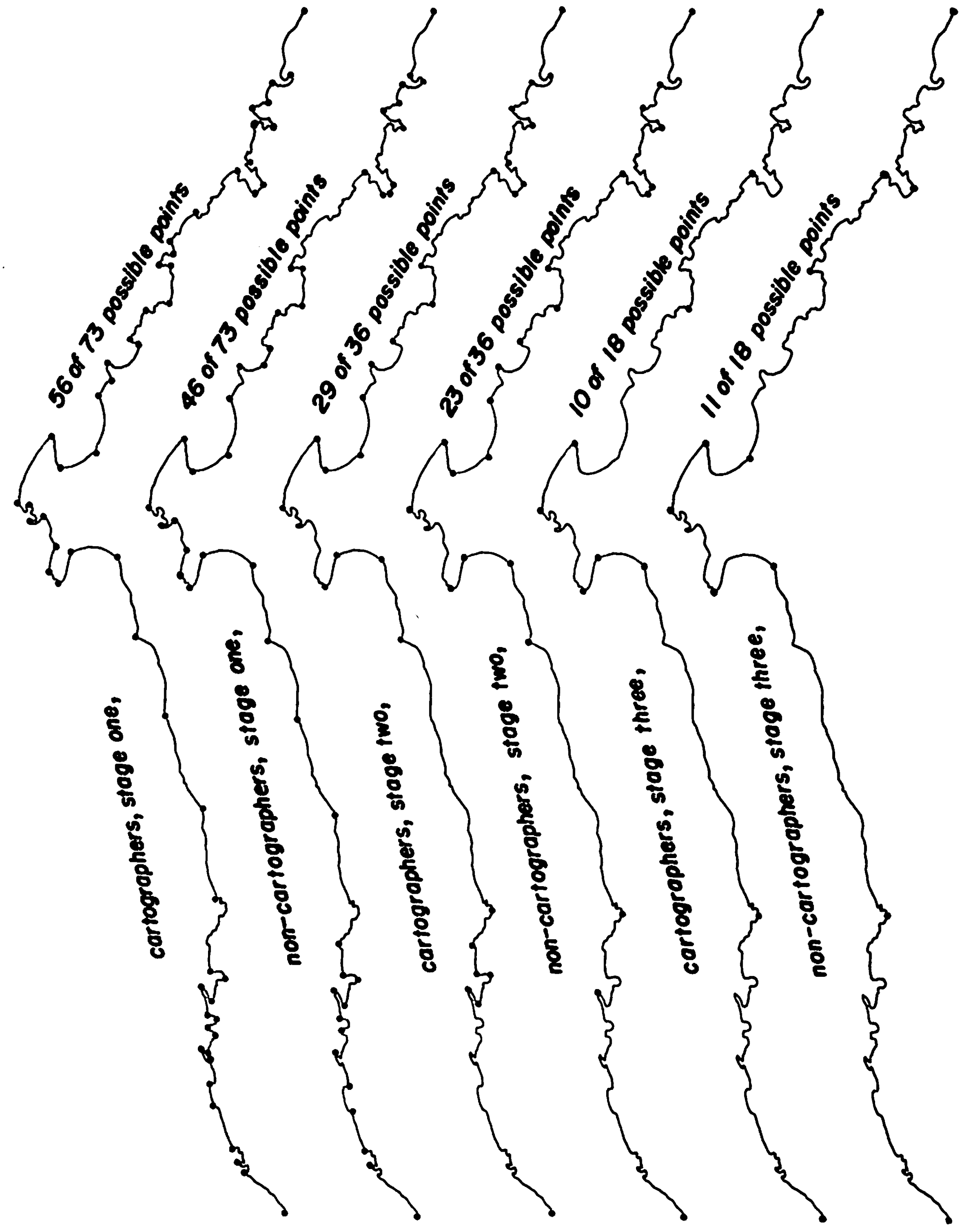


Figure 3.31. *Points selected by a majority of cartographers and non-cartographers at three stages. Generalization increases on each pair from left to right.*

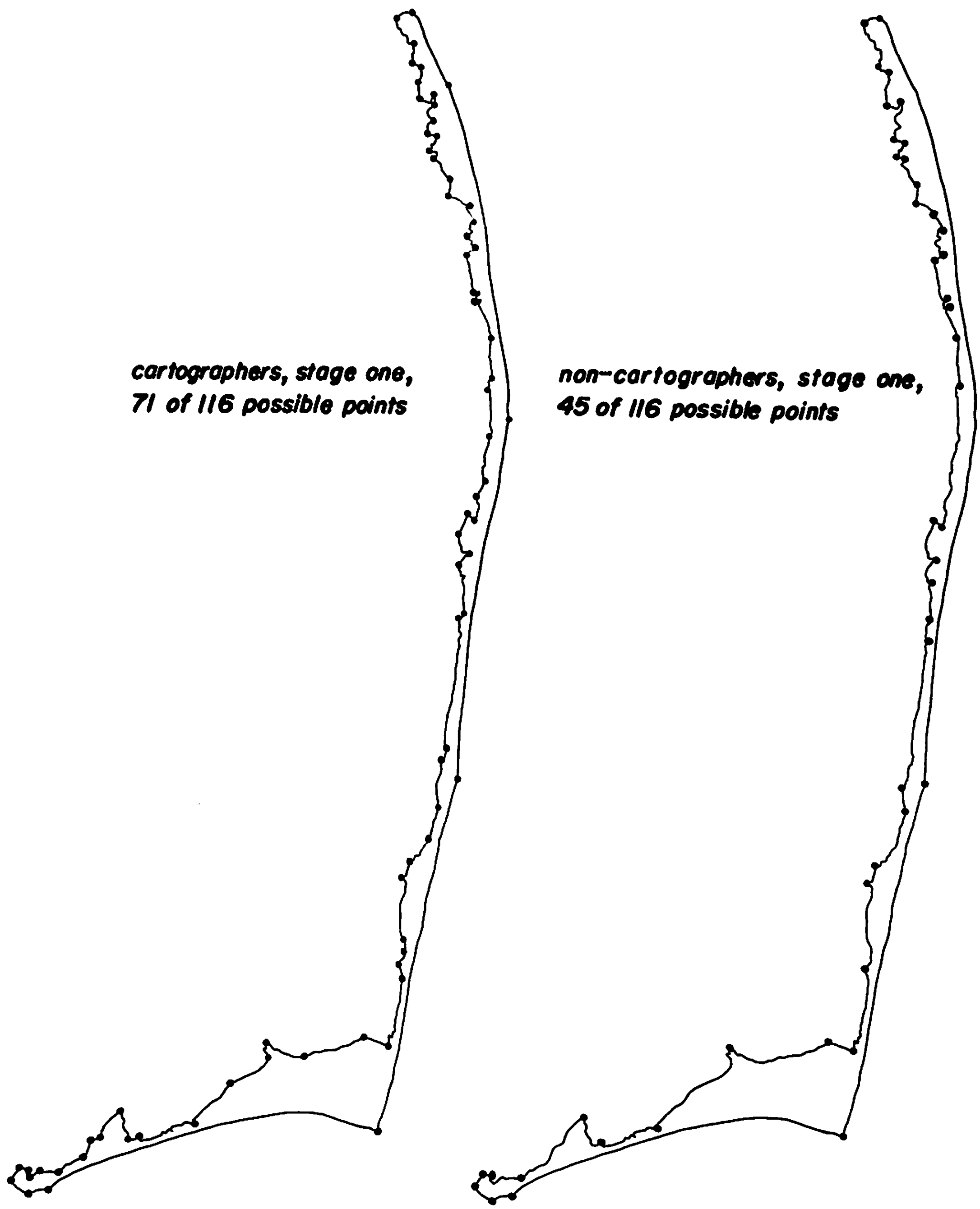


Figure 3.32. *Points selected by a majority of cartographers (left) and non-cartographers (right) for stage one (least generalization).*

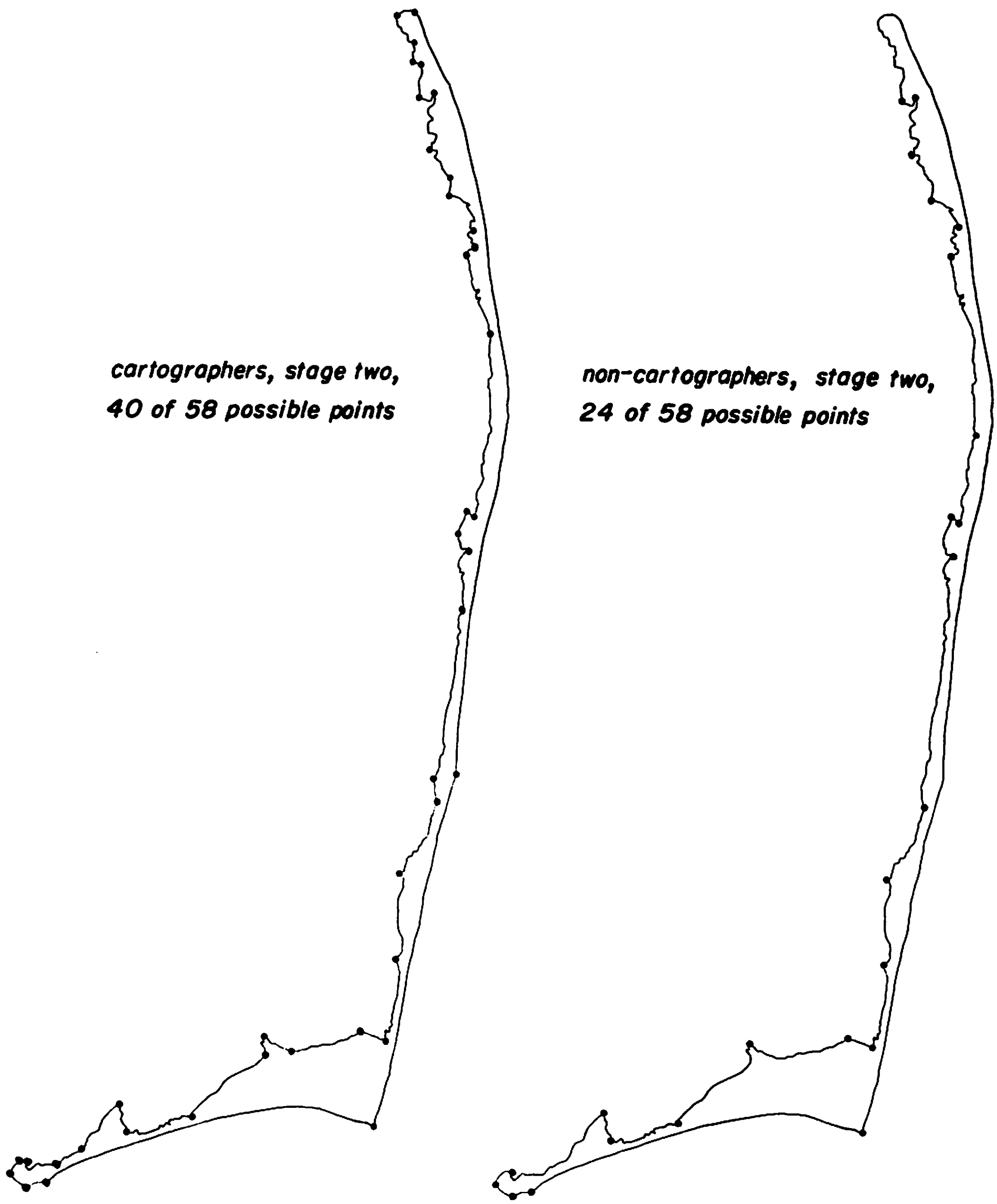


Figure 3.33. *Points selected by a majority of cartographers (left) and non-cartographers (right) for stage two.*

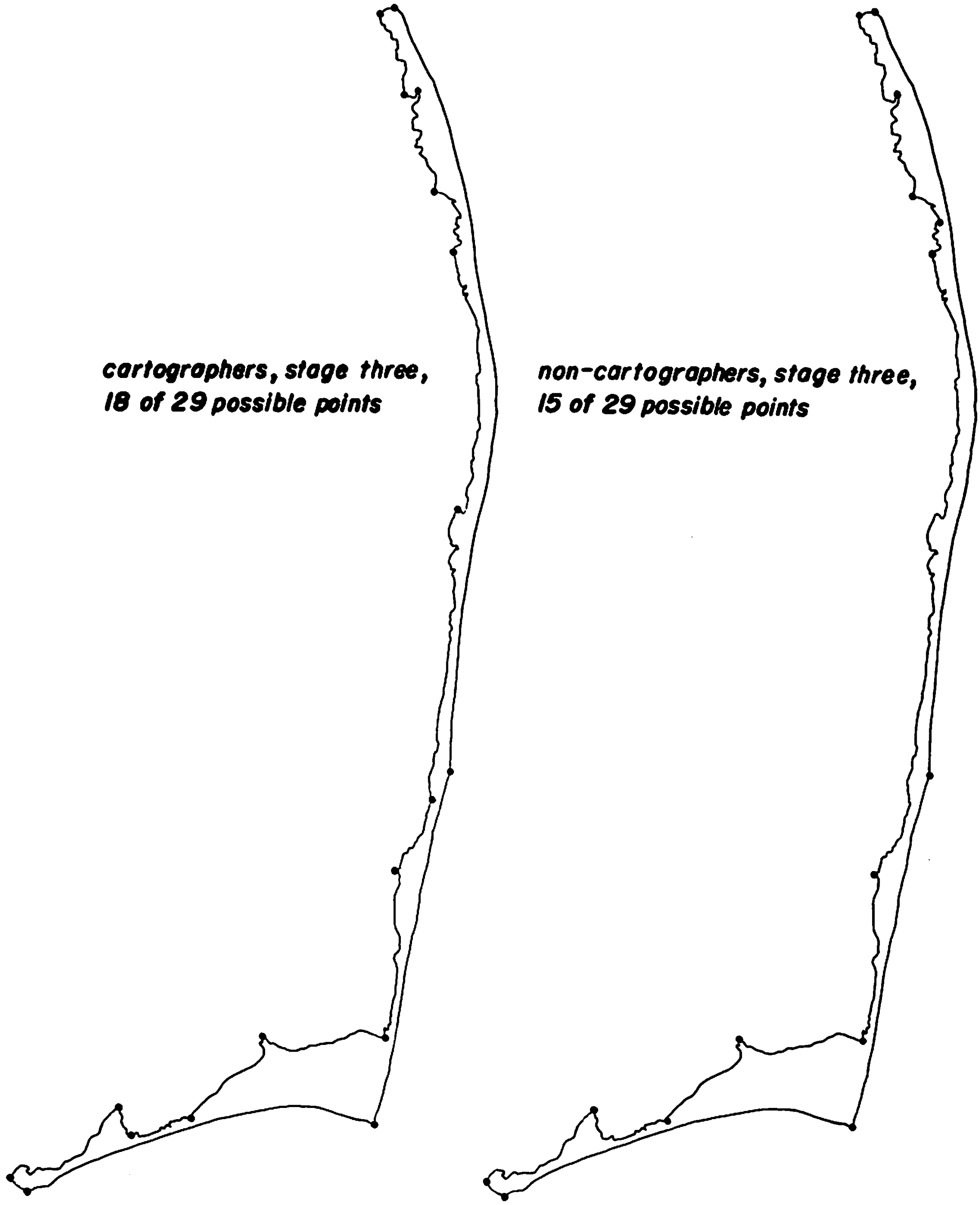


Figure 3.34. ***Points selected by a majority of cartographers (left) and non-cartographers (right) for stage three (greatest generalization).***

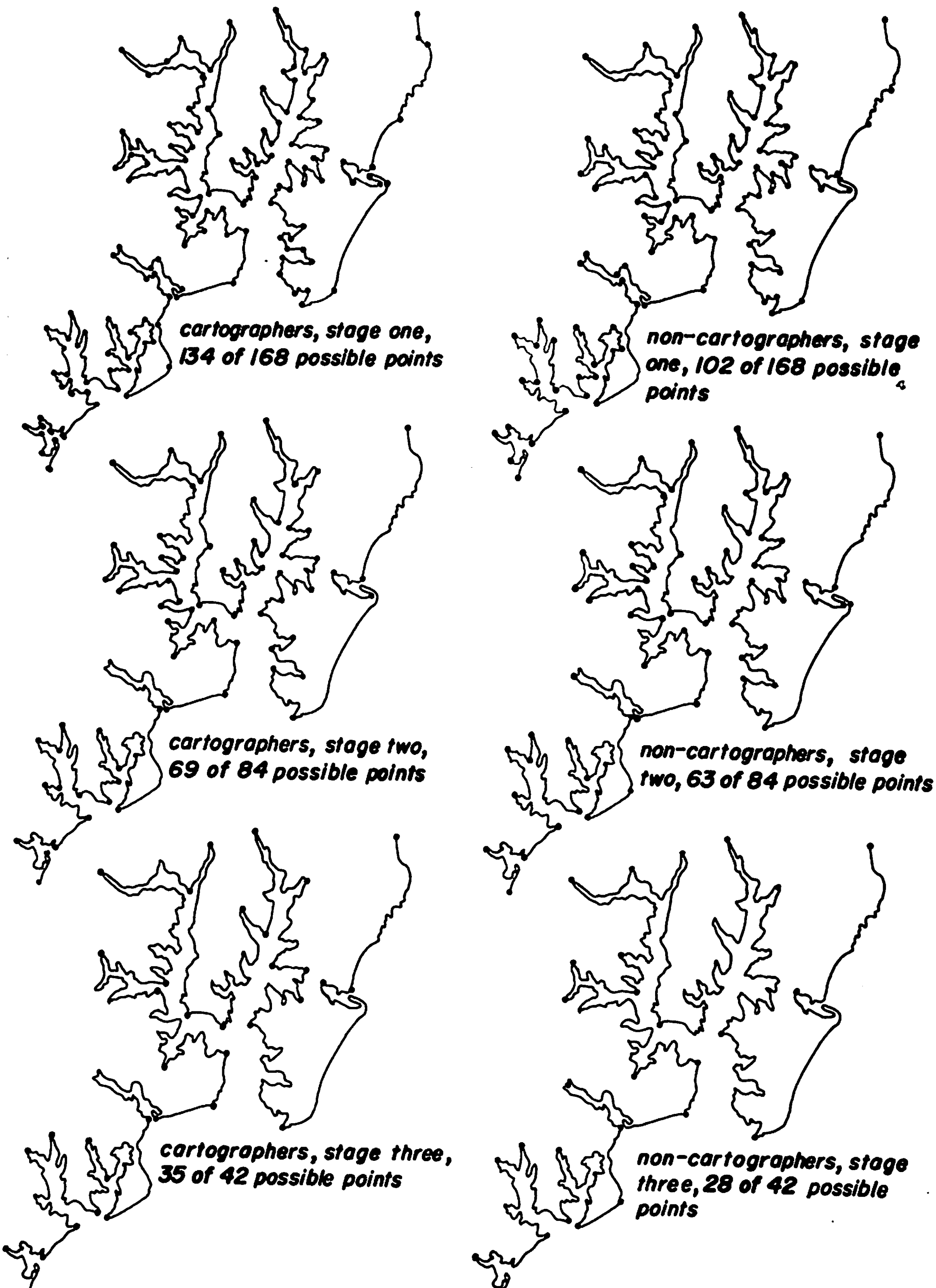


Figure 3.35. *Points selected by a majority of cartographers (left) and non-cartographers (right) at three stages. Generalization increases on each pair from top to bottom.*

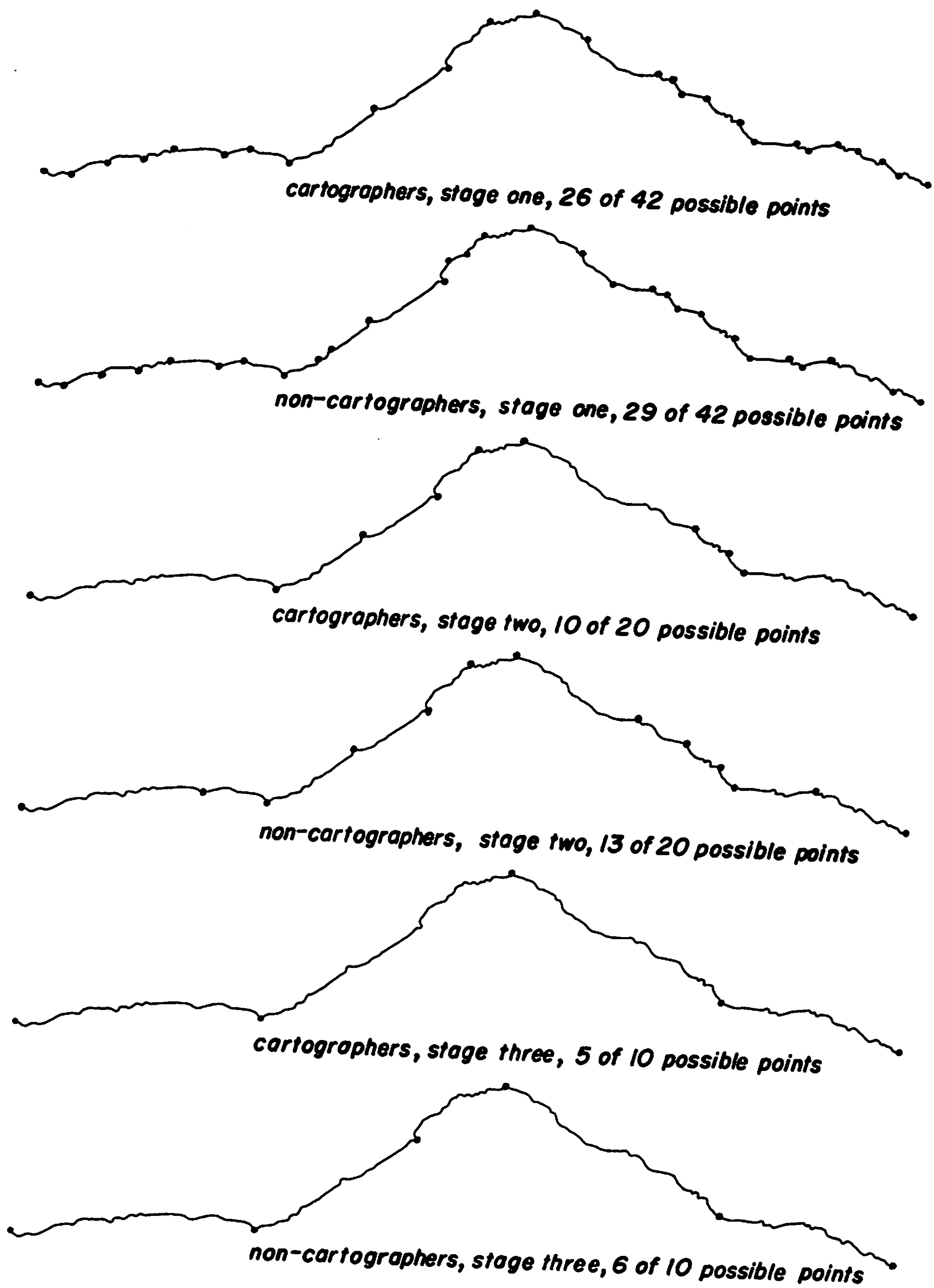


Figure 3.36. *Points selected by a majority of cartographers and non-cartographers at three stages. Generalization increases on each pair from top to bottom.*

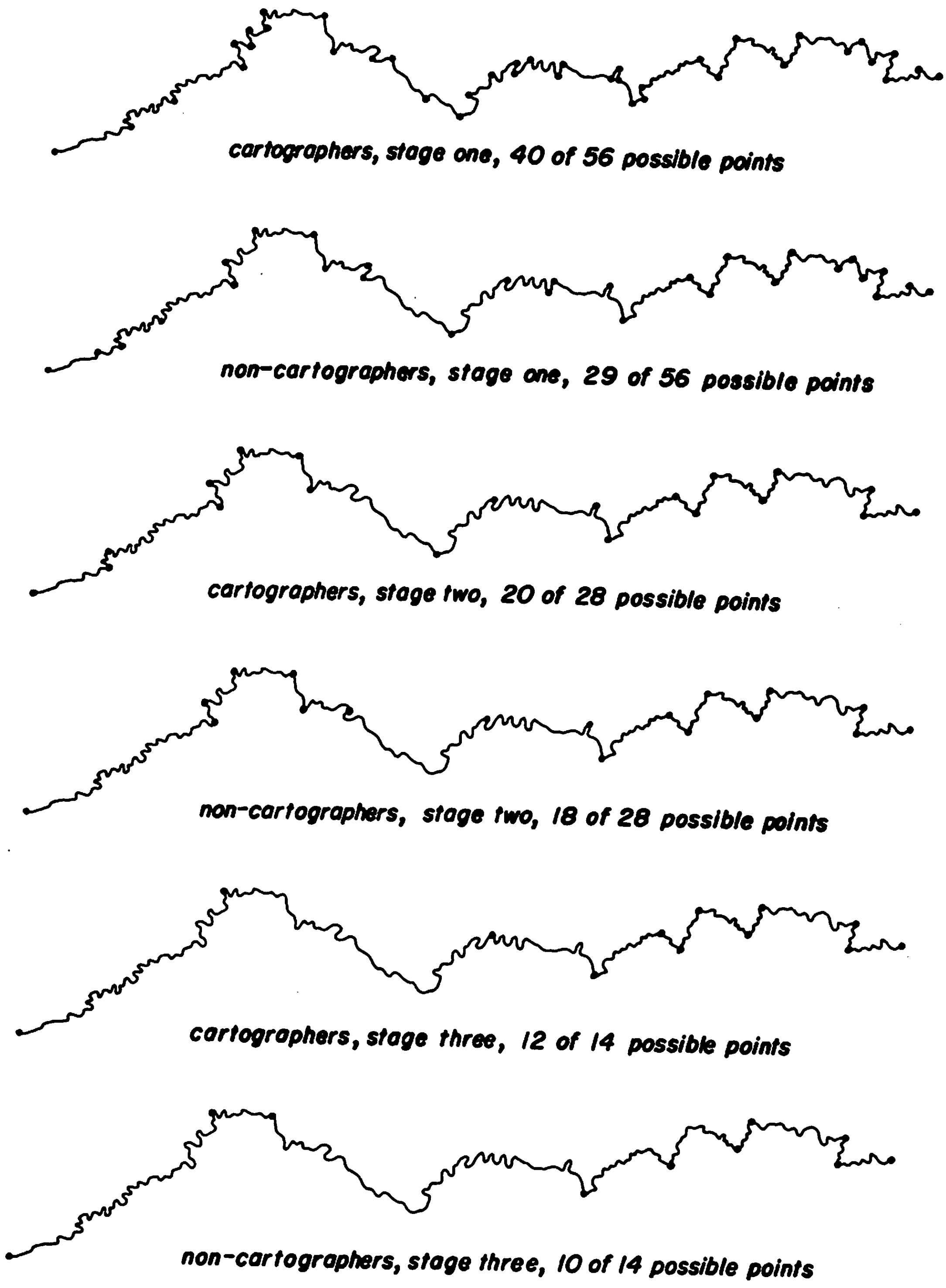


Figure 3.37. *Points selected by a majority of cartographers and non-cartographers at three stages. Generalization increases on each pair from top to bottom.*

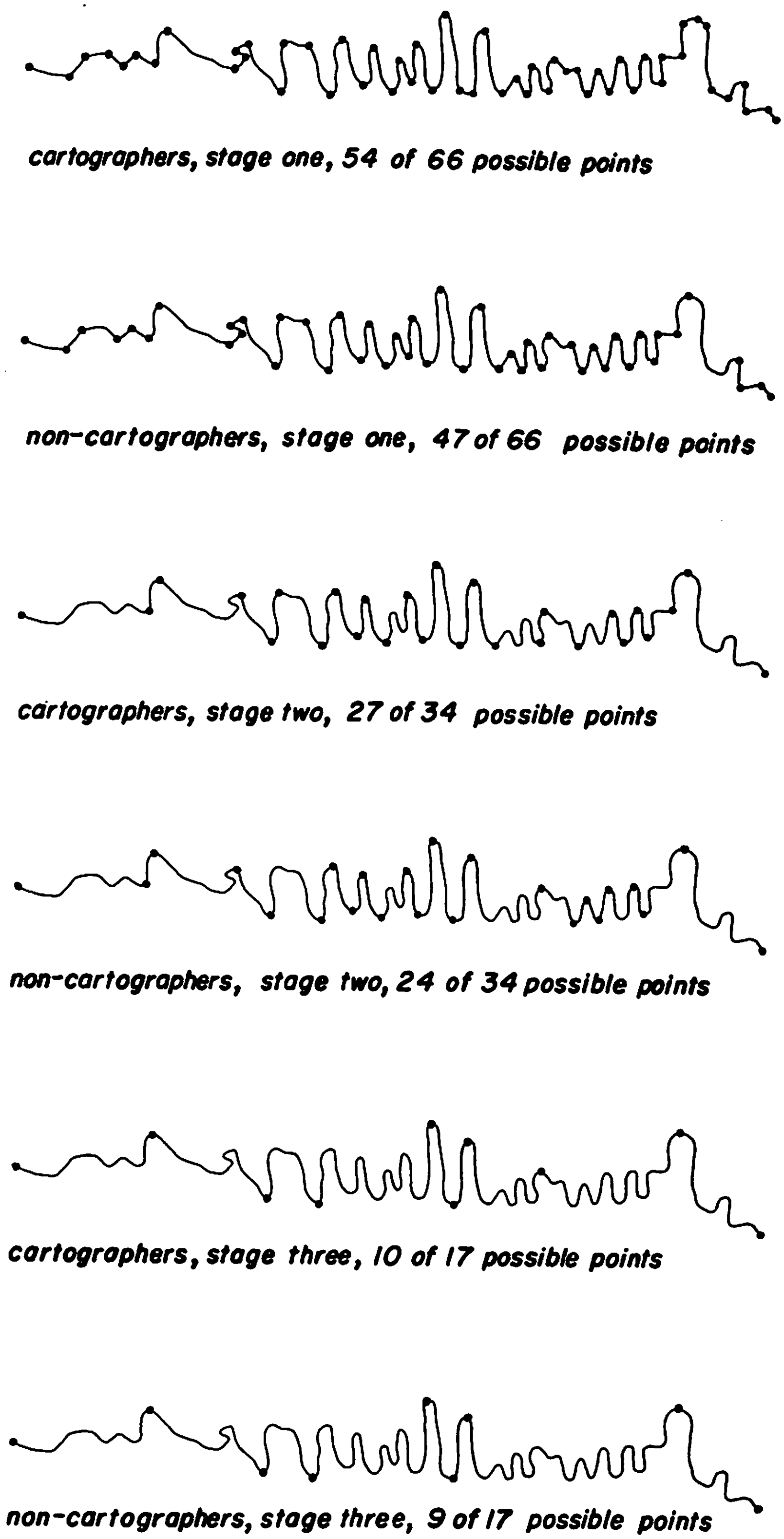


Figure 3.38. ***Points selected by a majority of cartographers and non-cartographers at three stages. Generalization increases on each pair from top to bottom.***

appear again in stages two and one (in which less generalization was required). This suggests that the respondents may have viewed the stage three points as those with highest priority, which were supplemented in stages two and one. Further discussion of this phenomenon follows in the concluding chapter.

Average Number of Points Selected by Each Group

On a group basis, cartographers selected a greater number of points than did the non-cartographers. For all but one of the 18 corresponding pairs, the average number of points selected per segment by the group of cartographers was greater than that selected by the group of non-cartographers. The exception occurred in the responses for the Fall River, stage one. (Figure 3.22. For a complete list of means, see Appendix C.) However, a difference of means test computed for each corresponding pair of responses indicates no significant difference between the means of the two groups, except in one case. This exception occurs in the responses for Cape Hatteras, stage one, in which the computed test statistic indicates that the means of the two groups (cartographers and non-cartographers) are significantly different.

The responses between corresponding pairs reveal the amount of agreement existing among all individuals tested (as well as between the two groups) concerning which

CHAPTER IV

CONCLUSION

The results of this study, an experiment in which two groups of individuals selected characteristic points at three levels of generalization along six sample lines, have been displayed in the preceding chapter. The similarity between the responses given by cartographers and those given by non-cartographers, which is visually evident in the graphs of corresponding pairs, is substantiated statistically with significant Pearson correlation coefficients.

Although statistical tests show the degree of similarity existing between the two group responses, in light of insights gained from administering the experiment, slight differences were noted in the manner in which each group (taken as a whole) approached the task. These slight differences are now discussed. They seem to indicate that the group of cartographers were fussier than the group of non-cartographers.

Slight Differences Between the Two Group Responses

Further statistical tests reveal subtle differences between the responses given by the two sample groups. On the average, the group of cartographers tended to select a greater number of points than did the non-cartographers (see Appendix C). However, the mean number of points selected (per segment, 0.05 inch (0.125 cm) of sample line) by each group on all corresponding pairs is not significantly different at the 5 per cent level, except in one case. This exception occurs in the stage one responses to the Cape Hatteras sample line. This finding suggests that the group of cartographers may be predisposed or accustomed to work involving great detail, because the sample line of Cape Hatteras is the second longest in length, and therefore involved the second largest number of pins. One might expect that a significant difference would also have been found for the sample line of Chesapeake Bay, which was the longest sample line, and therefore involved the greatest number of pins in the stage one response. The only significant difference occurred on the Cape Hatteras line probably because the fine detail existing there was either more apparent to the group of cartographers, or seemed more important to them. Although the Chesapeake Bay line is intricate, it does not possess the fine detail found on the Cape Hatteras line.

Majority opinion was used to assess the amount of agreement existing among the individuals of each group. Using a difference of proportions test, the difference in the proportion of total points (segments < 0.05 inch (0.125 cm) length) which had been selected by a majority of each group, at comparable response stages on a sample line, was not found to be significantly different at the 5 per cent significance level except in two cases. The two exceptions occur in the stage one responses for the sample lines of Cape Hatteras and Chesapeake Bay, where the majority of cartographers selected a greater proportion of points in each case. (See Appendix D.) Again, this finding seems to suggest more diligence on the part of the group of cartographers, perhaps because of their work-related background, or possibly because they may have had a more serious attitude towards the experiment, appreciating the implications of the study. Cape Hatteras and Chesapeake Bay were the sample lines involving, respectively, a possible 116 pins for 23 inches (57.5 cm) of line, and 168 pins for 34 inches (85 cm) of line. It is easy for one to imagine, as many participants declared, the tedium involved in this task. The slight tendency by non-cartographers, relative to the cartographers, towards a more uniformly even distribution of selected points as evaluated by majority opinion, suggests a less deliberate approach taken by the non-cartographers. Because this occurred at a significant level on the two most lengthy and intricate sample lines only, it is conjectured that this resulted primarily from a difference in attitude between the two sample groups.

Aside from an apparent predisposition towards detailed work among the cartographers with respect to the two sample lines just noted, the sample data indicate that there is no appreciable difference between the sets of points which were selected by the two groups of individuals who were tested. As a suggestion for an explanation of the outcome of like points being selected by cartographers and non-cartographers, note that cartography, despite its long existence, has yet to establish an accepted, systematic method for cartographic line generalization (Thrower, 1972; Morrison, 1974).

Types of Points Selected

The selection by the respondents of several types of points reoccurs across the six sample lines as well as between the three stages of generalization. Points (segments < 0.05 inch (0.125 cm) length) selected by a majority of either group possess one or more of the following characteristics: end points of lines; acute or near acute angular changes, especially cusps; midpoints of gentle curves; "corners" of squarish curves; sharp angles or midpoints of curves which separate segments with visibly different degrees of indentation or undulation.

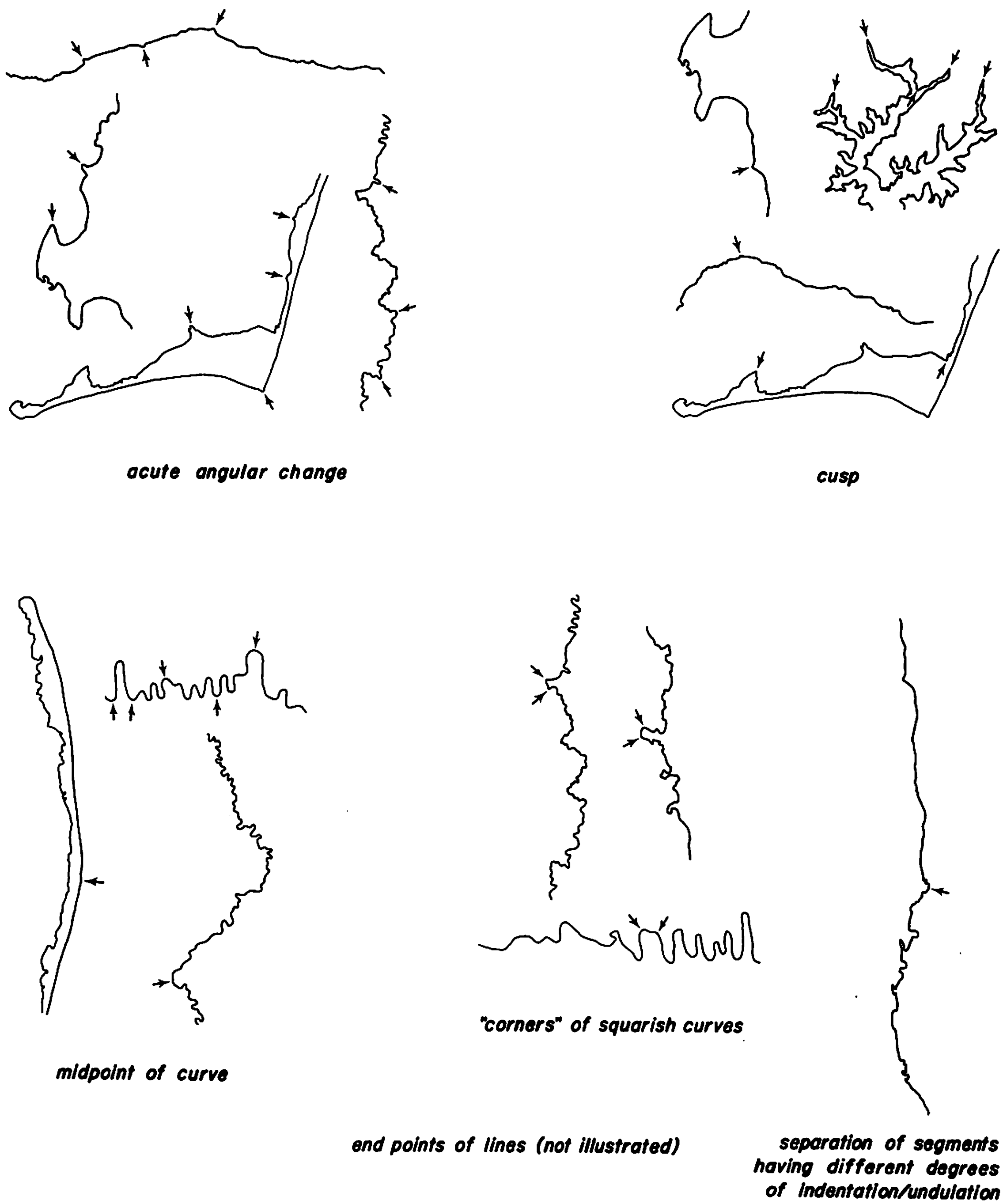


Figure 4.1. *Examples of the types of points selected by a majority of both groups.*

Figure 4.1 illustrates each type of these points with examples from the data gathered in this study. One can refer to Figures 3.31-3.38 in chapter three to compare this description of characteristic points with the complete set of points which resulted in this experiment.

Excluding the cusp point in some cases, it is possible to generalize the description of all other points selected by a majority of either group. Points were selected which mark, with varying degree, major deviations from the "general direction" of the sample line. This concept is depicted geometrically in Figure 4.2.

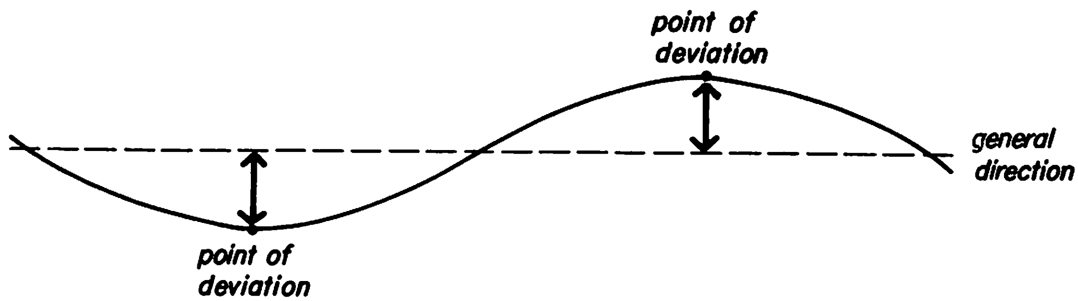


Figure 4.2.

Relationship of Selected Points to Generalizing Schemes

One will recognize in Figure 4.2 a method for cartographic line generalization which has been suggested in previous works by others. The idea of describing a line by selecting key points rather than by eliminating points has been presented in the algorithms for locating dominant points by Langridge (1972) and Rosenberg (1972);

in algorithms for curve approximation with straight lines by Ramer (1972); in the point reduction algorithm by Douglas and Peucker (1973); and in the "line as band width theory" by Peucker (1975). (Peucker claims that his generalization scheme (1975) based on a previous study he and Douglas had published (1973), was developed independently of the work published earlier by Langridge (1972), Rosenberg (1972), and Ramer (1972).)

Because Peucker's algorithm was presented specifically in the context of cartographic line generalization, it is applied to one of the sample lines used in this experiment. Peucker's method consists essentially of two steps, which are repeated until the desired level of abstraction is attained. First, one defines the "general direction" of the line being generalized by connecting the two end points with a straight line (Peucker himself notes that defining the general direction by calculating the principal axis of the point set would likely yield better results (1975, 512)). The general direction is the locus equidistant from the two sides of a minimum bounding rectangle, and whose perpendicular bisector is the locus equidistant from the two ends of the bounding rectangle. The bounding rectangle is called the band width of the line. The second step in this algorithm is to select those points on the original line which touch the sides of the band width. Step one is then repeated by treating the newly derived points as end points of a straight line defining the general direction of a subset of band widths.

A comparison of the points obtained by applying Peucker's algorithm to the sample line of Cape Arago used in this experiment with those points selected by the group of cartographers sampled in this experiment is shown in Figure 4.3. In Figure 4.3 the points selected by the group of cartographers are those obtained in the third stage of the experiment, where the greatest amount of generalization was required. Similarly in Figure 4.4 the points obtained by applying Peucker's algorithm have been compared with the points selected by the group of non-cartographers sampled in stage three of this experiment.

Caution is necessary before drawing any conclusions from Figures 4.3 and 4.4. First, the points resulting from applying Peucker's method, and those selected by a majority of the two groups sampled in this experiment are not entirely comparable. The majority of cartographers in the stage three response for Cape Arago selected 10 points (segments $\leq$ 0.05 inch (0.125 cm)), while the majority of non-cartographers in this stage selected 11 points (segments $\leq$ 0.05 inch (0.125 cm)). Peucker's method does not usually yield a number of points which is incremented by one from one iteration to the next. Therefore, it is

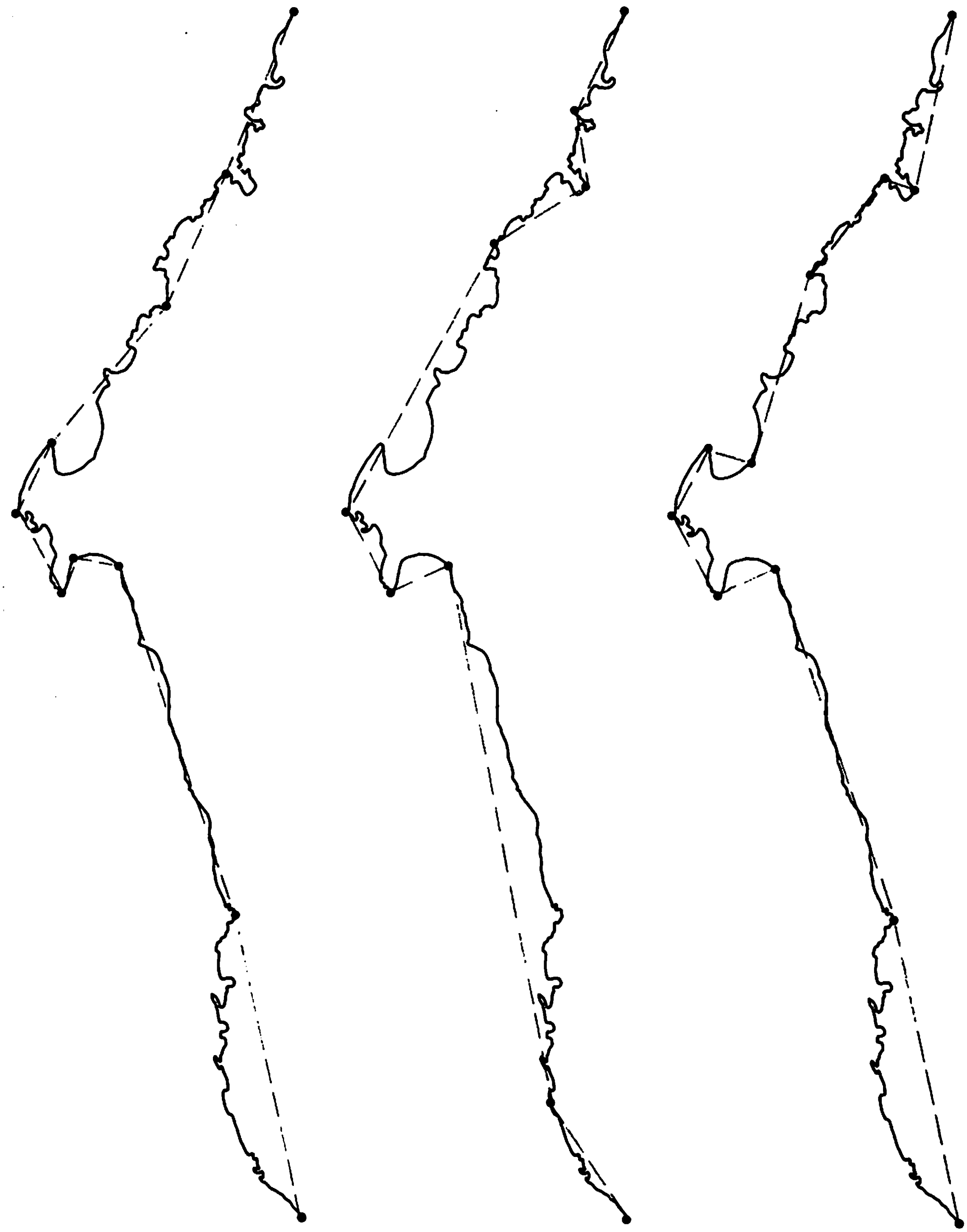


Figure 4.3. Figure 4.4.

Left, points selected by majority of cartographers for stage three. Middle, points derived from Peucker's algorithm. Right, points selected by majority of non-cartographers for stage three. Note, number of points is unequal: see page 44 in text for explanation.

not usually possible to derive an arbitrary number of points using Peucker's algorithm, in order to be able to compare those resulting points with a specified number of points, such as the number selected by a majority of one group in this experiment. Peucker's method will always produce at least one and usually at least two new points per band in each iteration, but may produce more than two points per band, depending on the configuration of the original line segments. Therefore, the increment in the number of points defining the original line does not necessarily increase systematically.

One must also be cautious of evaluating the lines which result by connecting the points obtained by the sample groups in this experiment. Although the points resulting from this study were not selected according to instructions that they be connected by straight line segments in order to represent the sample line, it is likely that some, and perhaps many, of the respondents proceeded according to that assumption.[1] However, the points resulting from this study should be viewed in their

---

[1]Approximately 10% of the respondents said that they were selecting points which they viewed as being connected by straight line segments. Because discussion of the experiment was not encouraged while an individual was performing the task, it is not known exactly how many participants proceeded with this thought in mind. The experiment was not designed with the intention that the selected points would later be connected with straight line segments. The selected points were to be viewed only as points through which a generalization of the line should pass.

proper context, that is, they represent points through which any generalization of the sample line should pass. In this sense, for a cartographic generalization of the sample lines, one set of points may not be all inclusive because, for example, curves may be interpolated between them.

The value of the sample data presented here lies in the insight it can provide about the kind or type of point(s) which this sample group believed should be included in a generalization of these sample lines. In addition, responses to the three stages suggest an hierarchy of importance for the points themselves. One will note that points selected in stage three are usually supplemented in stages two and one by points which mark segments adding more detail to the line. It is as if stage three represents a skeleton, to which finer detail is added in stages two and one.

A combination of the characteristic of the selected points along with the apparent hierarchy of their importance may provide a basis for a systematic procedure to follow in generalizing these sample lines. Because the character of the sample lines varies considerably, an extension to other lines seems feasible. The sample data from this study appear to support the concept of segmenting a line into band widths containing various frequencies, such as that proposed by Peucker (1975).

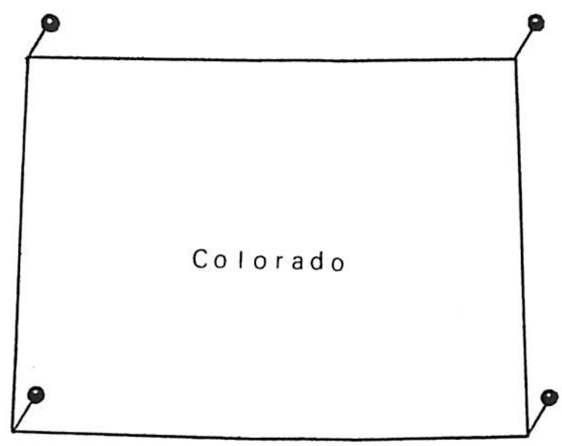
In conclusion, the sample data presented here show that there is substantial agreement among all individuals tested as to which points along a given line are considered to be essential to the characterization of that line. These characteristic points represent major deviations from the general direction of the sample line, at varying scales. This finding suggests that a promising approach to use in developing a cartographic line generalization model would incorporate procedures which identify points of maximum deviation from straight lines defining the general direction of sequentially smaller segments of the original line.

APPENDIX A

Instructions

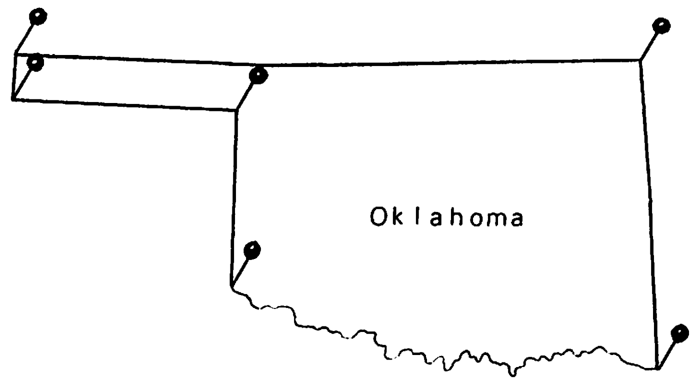
When cartographers make maps, they must often generalize from the lines on the original map. In generalizing, they select some points or parts of the original line to be retained, but they eliminate some of the other original detail.

We believe that some points on a line are critical; that is, that these points must be retained and must remain in the same position in order to preserve the character of the line. Using the outline of the state of Colorado as an example, we believe that there are four critical points which characterize this line. This is illustrated below, where pins (represented by the symbol, ) have been stuck into the line to mark the critical points.



Now consider the outline of the state of Oklahoma. We believe that the six pins stuck into this outline, as shown below, represent some of the critical points which must be retained in order to characterize this line.

However, we do not know which of the other points on the line are critical, particularly those on the southern-most boundary, which is shown below as a gray line. We need to know which points on a line are critical; therefore we are asking you to indicate for us which points on a given line are the most important ones to you in preserving the character of the line.



In this experiment, you will be given sample lines which are not straight as in the example of Colorado above. First you will be given a sample line and a box containing about 100 pins. The exact number of pins varies with each sample line. Please stick the pins into the line itself, to mark those points which you feel are critical. You may use as many of the provided pins as you feel are necessary to indicate the critical points, and you may change the position of the pins as often as you like.

When you have completed this, you will be given another copy of the same sample line, and half as many pins as in the first box. Again, you may use as many of the pins as necessary, and you may change their position. After completing this, you will be given another copy of the same sample line for a third and final time, and a new box of pins containing 1/4 the number used originally. Use as many of the pins as you wish, and reposition them as you desire. The only restriction is that you stick the pins into the line itself.

By successively reducing the number of pins available, we are simulating three different levels or degrees of generalization.

If anything is unclear, please ask questions before we begin. Thank you for participating.

APPENDIX B

Description of Sample Groups Responding

to Each Line

	Cape Arago	Cape Hatteras	Chesapeake Bay	Fall River	Mancos River	Shenandoah River
CARTOGRAPHERS						
female	13%	30%	27%	20%	13%	20%
male	87%	70%	73%	80%	87%	80%
student, MA in cartography	27%	34%	40%	27%	30%	20%
employed cartographer	17%	13%	26%	37%	20%	24%
Ph.D., cartography	17%	13%	17%	10%	17%	20%
mapping with remote sensing	17%	23%	10%	17%	13%	27%
Ph.D., geography and/or geology	23%	17%	7%	10%	20%	10%
NON-CARTOGRAPHERS						
female	47%	43%	50%	50%	47%	43%
male	53%	57%	50%	50%	53%	57%
student, non-geography major	37%	57%	53%	43%	43%	47%
teacher, nurse, real estate	13%	3%	10%	3%	6%	10%
clerk, secretary, beautician	13%	13%	10%	10%	17%	13%
factory, assembly line	3%	3%	3%	3%	7%	7%
contractor, artist, landscaper	13%	10%	9%	14%	13%	14%
professor, non-geography/ geology	10%	13%	13%	20%	7%	10%
other and/or unknown	10%	0%	0%	3%	6%	0%

APPENDIX C

Number of Points Selected by Respondents, Listed by Sample Line

Sample line and stage of experiment	maximum number of pins available	mean number of points selected per 0.05 inch line segment: C*	N-C**	mean number of points selected per respondent: C*	N-C**
Stage One (least generalization)					
Cape Arago	73	6.35	5.63	63.33	56.13
Cape Hatteras	116	5.56	4.31	85.83	66.53
Chesapeake Bay	168	6.41	5.63	145.37	127.57
Fall River	42	5.70	5.98	31.13	32.67
Mancos River	56	6.07	5.45	47.17	42.30
Shenandoah River	66	6.77	6.33	54.80	51.30
Stage Two					
Cape Arago	36	3.50	3.35	34.90	33.40
Cape Hatteras	58	3.48	2.87	53.73	44.27
Chesapeake Bay	84	3.69	3.51	83.73	79.53
Fall River	20	3.32	3.30	18.17	18.07
Mancos River	28	3.35	3.32	26.00	25.77
Shenandoah River	34	4.14	3.85	33.57	31.17
Stage Three (greatest generalization)					
Cape Arago	18	1.78	1.72	17.73	17.13
Cape Hatteras	29	1.87	1.62	28.87	25.00
Chesapeake Bay	42	1.82	1.12	41.20	25.33
Fall River	10	1.77	1.76	9.70	9.60
Mancos River	14	1.80	1.79	13.97	13.87
Shenandoah River	17	2.09	2.05	16.90	16.60

C* denotes cartographers
N-C** denotes non-cartographers

APPENDIX D

MAJORITY OPINION

sample line and stage of experiment	number of 0.05 inch segments along sample line	number of segments*** selected by > 15 of each group: C*	N-C**	significant difference in proportion of majority opinion between groups
Cape Arago				
stage one	299	56	46	no
stage two	299	29	23	no
stage three	299	10	11	no
Cape Hatteras				
stage one	463	71	45	yes
stage two	463	40	24	no
stage three	463	18	15	no
Chesapeake Bay				
stage one	680	135	102	yes
stage two	680	69	63	no
stage three	680	35	28	no
Fall River				
stage one	164	26	29	no
stage two	164	10	13	no
stage three	164	5	6	no
Mancos River				
stage one	233	40	29	no
stage two	233	20	18	no
stage three	233	12	10	no
Shenandoah River				
stage one	243	54	47	no
stage two	243	27	24	no
stage three	243	10	9	no

*** segments < 0.05 inch in length
C* denotes cartographers
N-C** denotes non-cartographers

APPENDIX E

Correlation Coefficients

Summarized Group Response for Cartographers with Non-Cartographers

sample line and stage of experiment	PCC**	SL*	KCC***	SL*	SCC****	SL*
Cape Arago						
stage one	0.8849	0.001	0.6349	0.001	0.7562	0.001
stage two	0.9320	0.001	0.6222	0.001	0.7062	0.001
stage three	0.9063	0.001	0.6047	0.001	0.6587	0.001
Cape Hatteras						
stage one	0.8368	0.001	0.5842	0.001	0.7060	0.001
stage two	0.9096	0.001	0.5777	0.001	0.6646	0.001
stage three	0.8810	0.001	0.5044	0.001	0.5551	0.001
Chesapeake Bay						
stage one	0.9452	0.001	0.7224	0.001	0.8138	0.001
stage two	0.9530	0.001	0.6424	0.001	0.7076	0.001
stage three	0.5482	0.001	0.5507	0.001	0.5912	0.001
Fall River						
stage one	0.8520	0.001	0.6099	0.001	0.7575	0.001
stage two	0.9089	0.001	0.5026	0.001	0.5892	0.001
stage three	0.7950	0.001	0.5304	0.001	0.5935	0.001
Mancos River						
stage one	0.8779	0.001	0.6169	0.001	0.7360	0.001
stage two	0.9301	0.001	0.5402	0.001	0.6153	0.001
stage three	0.9071	0.001	0.5378	0.001	0.5724	0.001
Shenandoah River						
stage one	0.6830	0.001	0.4791	0.001	0.6019	0.001
stage two	0.8861	0.001	0.6523	0.001	0.7474	0.001
stage three	0.8992	0.001	0.5654	0.001	0.6270	0.001

PCC** denotes Pearson Correlation Coefficient
KCC*** denotes Kendall Correlation Coefficient
SCC**** denotes Spearman Correlation Coefficient
SL* denotes significance level

BIBLIOGRAPHY

Attneave, Fred (1954). "Some informational aspects of visual perception." Psychological Review, 61: 183-193.

________, and Arnoult, M.D. (1956). "The quantitative study of shape and pattern perception." Psychological Bulletin, 53: 452-471.

Balchin, W.G.V. and Coleman, A.M. (1967). "Cartography and Computers." The Cartographer, IV (2): 120-127.

Bickmore, D.P. (1973). "New Developments at the Experimental Cartography Unit -- Cartographic Data Banking." International Yearbook of Cartography, XIII: 90-96.

Blalock, Hubert M. (1972). Social Statistics, 2nd Edition. New York: McGraw-Hill.

Boyle, A.R. (1970). "The Quantised Line." The Cartographic Journal, 7 (2): 91-94.

Breward, R.W. (1972). "A Mathematical Approach to the Storage of Digitized Contours." The Cartographic Journal, 9 (2): 82-92.

British Cartographic Society (1974). Automation in Cartography. London: British Cartographic Society.

Brophy, David M. (1972). "Automated Linear Generalization in Thematic Cartography." Unpublished M.S. thesis, University of Wisconsin.

________ (1973). "An Automated Methodology for Linear Generalization in Thematic Cartography." Proceedings of ACSM, 33rd Annual Meeting. Washington, D.C.

Clark, James H. (1974). Three Dimensional Design of Free-Form B-spline Surfaces. Report Number: UTEC-CSc-74-120, University of Utah, Salt Lake City, Utah.

Coles, Melvin W. (1966). "Applications of the Electronic Digital Computer in Nautical Cartography." Proceedings of ASP-ACSM Convention, March 1966. Washington, D.C.

Dent, Borden D. (1972). "A Note on the Importance of Shape in Cartogram Communication." Journal of Geography, LXXI (7): 393-401.

Douglas, David H. and Peucker, Thomas K. (1973). "Algorithms for the reduction of the number of points required to represent a digitized line or its caricature." Canadian Cartographer, 10 (2): 112-122.

Fairbridge, R.W., ed. (1968). The Encyclopedia of Geomorphology. New York: Reinhold Book Co.

Freeman, Herbert (1961). "On the encoding of arbitrary geometric configurations." IRE Transactions on Electronic Computers, 10: 260-268.

________. (1961). "Techniques for the Digital Computer Analysis of Chain-Encoded Arbitrary Plane Curves." Proceedings of the National Electronics Conference, XVII: 421-432. Chicago, Ill.

________. (1967). "On the Classification of Line-Drawing Data." Models for the Perception of Speech and Visual Form. Edited by Wathen-Dunn. Cambridge, Mass.: MIT Press.

Green, R.T. and Courtis, M.C. (1966). "Information Theory and Figure Perception: The Metaphor that Failed." Acta Psychologica, XXV: 12-36.

Hájek, Milan (1972). Automation and the System of Cartographic Generalization: Problems. Slovak Technical University, Bratislava, Czechoslovakia.

Hershey, A.V. (1963). The Plotting of Maps on CRT Printer. Naval Weapons Laboratory Report 1844, U.S. Navy. Dahlgren, VA.

Hoel, Paul G. (1966). Elementary Statistics, 2nd Edition. New York: John Wiley and Sons, Inc.

Hogben, Lancelot (1960). Mathematics in the Making. Mladinska Knjiga, Ljubljana, Yugoslavia: Crescent Books, Inc.

Imhof, Eduard (1957). "Generalisierung der Höhenkurven." Kartographische Studien, Erganzungband Nr. 264 zu Petermanns Geographische Mitteilungen, 89-99. Translation, "Generalization of Contour Lines."

Jenks, George F. (1963). "Generalization in Statistical Mapping." Annals of the AAG, 53 (1): 15-26.

Kao, R.C. (1963). "The Use of Computers in the Processing and Analysis of Geographical Information." Geographical Review, LIII (4): 530-547.

King, Cuchlaine A.M. (1972). Beaches and Coasts, 2nd Edition. New York: St. Martin's Press.

Koeman, C. and Van der Weiden, F.L. (1970). "The Application of Computation and Automatic Drawing Instruments to Structural Generalization." The Cartographic Journal, 7: 47-49.

Lang, T. (1969). "Rules for Robot Draughtsmen." Geographical Magazine, XLII (1): 50-51.

Langridge, D.J. (1972). "On the Computation of Shape." Frontiers of Pattern Recognition. Edited by S. Watanabe. New York: Academic Press.

Leopold, L.B.; Wolman, W.G.; and Miller, J.P. (1964). Fluvial Processes in Geomorphology. San Francisco: W.H. Freeman and Co.

Linders, J.G. (1973). "Computer Technology in Cartography." International Yearbook of Cartography, XIII: 69-80.

Lipkin, B.S. and Rosenfeld, A. (1970). Picture Processing and Psychopictorics. New York: Academic Press.

Lobeck, A.K. (1939). Geomorphology. New York: McGraw-Hill.

Lundquist, Gösta (1959). "Generalization -- A Preliminary Survey of an Important Subject." The Canadian Surveyor, XIV: 466-470.

Miller, O.M., and Voskuil, R.J. (1964). "Thematic Map Generalization." Geographical Review, LIV: 14-19.

Monkhouse, F.J., and Wilkinson, H.R. (1966). Maps and Diagrams, 2nd Edition. London: Methuen and Co., Ltd.

Monmonier, Mark S. (1969). GIPSY: A Geographic Incremental Plotting System, Papers in Geography, No. 4. University Park, PA: Pennsylvania State University.

Morrison, Joel L. (1974a). "A Theoretical Framework for Cartographic Generalization with Emphasis on the Process of Symbolization." International Yearbook of Cartography, 14: 113-127.

________ (1974b). "Changing Philosophical-Technical Aspects of Thematic Cartography." The American Cartographer, 1 (1): 5-14.

Muehrcke, Phillip C. (1970). "Trends in Cartography." Focus on Geography. 40th Yearbook of National Council for the Social Studies. Edited by Philip Bacon.

________ (1972). Thematic Cartography. Resource Paper No. 19. Washington, D.C., Association of American Geographers.

Neumann, J. (1972). "Genesis and definition of the term 'cartographic generalization'." 5th Technical Conference of the International Cartographic Association, Commission II. Montreal/Ottawa.

Noton, D. and Stark, L. (1971). "Eye Movements and Visual Perception," Scientific American, 224 (6): 35-43.

Pannekoek, A.J. (1962). "Generalization of Coastlines and Contours." International Yearbook of Cartography, II: 55-75.

Perkal, J. (1966). "An Attempt at Objective Generalization," trans. by J. Jackowski. Michigan Inter-University Community of Mathematical Geographers. Discussion Paper No. 10. Ann Arbor: University of Michigan.

Peucker, Thomas K. (1972a). Computer Cartography. Resource Paper No. 17. Association of American Geographers.

________ (1972b). Computer Cartography -- A Working Bibliography. University of Toronto, Department of Geography, Discussion Paper No. 12. Toronto: University of Toronto.

________ (1975). "A Theory of the Cartographic Line." Proceedings of the International Symposium on Computer-assisted Cartography, Auto-Carto II. U.S. Department of Commerce, Bureau of the Census.

Raisz, Erwin (1962). Principles of Cartography. New York: McGraw-Hill.

Ramer, Urs (1972). "An Iterative Procedure for the Polygonal Approximation of Plane Curves." Computer Graphics and Image Processing, 1 (3): 244-256.

Ratajski, Lech (1967). "The Phenomenon of Generalization Points." International Yearbook of Cartography, VII: 143-152.

Rhind, D.W. (1973). "Generalization and Realism within Automated Cartographic Systems." The Canadian Cartographer, 10 (1): 51-62.

________ (1974). "An Introduction to the Digitizing and Editing of Mapped Data." Automated Cartography. Edited by British Cartographic Society. London: British Cartographic Society.

Robinson, Arthur H. (1966). The Look of Maps. Madison, Wisconsin: University of Wisconsin Press.

________ (1967). "Cartography -- Which Way?" Journal of Geography, LXVI (1): 4-5.

________, and Sale, R.D. (1969). Elements of Cartography, 3rd Edition. New York: John Wiley and Sons, Inc.

Rosenberg, B. (1972). "The Analysis of Convex Blobs." Computer Graphics and Image Analysis, 1 (2): 183-192.

Salichtchev, K.A. (1973). "Some Reflections on the Subject and Method of Cartography After the 6th International Cartographic Conference." The Canadian Cartographer, 10 (2): 106-111.

Saltz, Daniel (1974). A Short Calculus, An Applied Approach, Revised Edition. Pacific Palisades, CA: Goodyear Publishing Co., Inc.

Saunders, B.G.R. (1958). "Design and Emphasis in Special Purpose Maps." Cartography, 2: 108-113.

Shelton, J.S. (1966). Geology Illustrated. San Francisco: W.H. Freeman and Co.

Sherman, John C. (1961). "New Horizons in Cartography: Functions, Automation, and Presentation." International Yearbook of Cartography, I: 13-19.

Srnka, E. (1970). "The Analytical Solution of Regular Generalization in Cartography." International Yearbook of Cartography, X: 48-63.

Steward, H.J. (1974). Cartographic Generalization: Some Concepts and Explanation. Cartographica Monograph No. 10. Toronto: Gutsell, York University, Toronto, Canada.

Sukhov, V.I. (1970). "Application of Information Theory in Generalization of Map Contents." International Yearbook of Cartography, X: 41-47.

Thornbury, W.D. (1969). Principles of Geomorphology. New York: John Wiley and Sons, Inc.

Thrower, Norman J.W. (1972). Maps and Man. Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Tobler, W.R. (1959). "Automation and Cartography." Geographical Review, XLIX (4): 526-534.

________ (1964). "An Experiment in the Computer Generalization of Maps." Technical Report No. 1. Ann Arbor, Michigan: Geography Department, University of Michigan.

________ (1965). "Automation in the Preparation of Thematic Maps." The Cartographic Journal, 2: 32-38.

________ (1966). "Numerical Map Generalization." Michigan Inter-University Community of Mathematical Geographers. Discussion Paper No. 8. Ann Arbor, Michigan: Department of Geography, University of Michigan.

Töpfer, F., and Pillewizer, W. (1966). "The Principles of Selection." The Cartographic Journal, III: 10-16.

Tuttle, S.D. (1970). Landforms and Landscapes. Dubuque, Iowa: Wm. C. Brown, Co., Publishers.

Vanicek, P., and Woolnough, D.F. (1975). "Reduction of Linear Cartographic Data Based on Generation of Pseudo-Hyperbolae." The Cartographic Journal, 12 (2): 112-119.

Wiedel, Joseph (1977). "Maps for the Blind." Presentation at University of Kansas, Geography Department Colloquium, April 1977.

Wood, M. (1972). "Human Factors in Cartographic Communication." The Cartographic Journal, 9 (2): 123-132.

Wright, John K. (1942). "Map Makers are Human, Comments on the Subjective in Maps." Geographical Review, 32 (4): 527-544.

Zusne, Leonard (1970). Visual Perception of Form. New York: Academic Press.