# Polygons, Stars, and Clusters

by

Robert O. Turek

Thesis submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Master of Science

in

Industrial Engineering and Operations Research

APPROVED:

J. S. Greenstein,
Co-Chairman

H. L. Snyder, Co-Chairman

R. D. Dryden

July, 1986

Blacksburg, Virginia

**Polygons, Stars, and Clusters**

by

Robert O. Turek

J. S. Greenstein, Co-Chairman

Industrial Engineering and Operations Research

(ABSTRACT)

One technique for displaying a set of quantitative variables is to represent the set as a polygon. Such displays allow the observer to visualize complex information quickly, as a whole. Polygon displays have been employed to display information for analysis, status, or presentation. An experimental investigation was undertaken to ascertain the effect of variation in certain visual features of the display on the consistency with which people categorize information presented as polygons. Variables included background information of the display, shading, and form. Subjects performed a categorization task on two sets of data; the results are analyzed for consistency between individuals and for consistency with certain standard clustering algorithms. The effects of distinctive portions of the figures on the judgment of similarity, and of the nature of the data and of interactions of combinations of the variables used in the experiment on the consistency of clustering were noted. Implications for the design of polygon displays are discussed.

# Acknowledgements

I would like to thank a number of people who have provided support for this undertaking. My advisor, Dr. Joel Greenstein, has always been helpful and encouraging, as have the other members of my committee, Dr. H. Snyder and Dr. R. D. Dryden. The consultants at the Computer Center and the Statistics Department have been courteous in trying to solve my all-too-numerous problems, especially Mr. Bill Sydor, who helped with the graphics programming. Of course many thanks are due to my wife and family for their support and for being able to get along without me for long periods during my course of study. I would particularly like to thank the General Reference Department of the University Library for making possible the continuing of my studies.

Funding for this project was provided in part by the Human Factors and Speech Technology Branch, Naval Ocean Systems Center, under the technical direction of Dr. G. A. Osga.

Table of Contents


CHAPTER ONE -- GRAPHICS AND PERFORMANCE . . . . . . . . 1

Introduction . . . . . . . . . . . . . . . . . . . . . . 1

Work Relating to Graphics and Human Performance . . . 1

CHAPTER TWO -- THE PRESENT STUDY . . . . . . . . . . . . 14

Background for the Display and the Task . . . . . . . . 14

The Experiment . . . . . . . . . . . . . . . . . . . . . 20

The Display. . . . . . . . . . . . . . . . . . . . 20

Design. . . . . . . . . . . . . . . . . . . . . . 30

Subjects. . . . . . . . . . . . . . . . . . . . . 30

Procedure. . . . . . . . . . . . . . . . . . . . 30

CHAPTER THREE -- EXPERIMENTAL RESULTS AND DISCUSSION . 34

Frequency Investigation . . . . . . . . . . . . . . . . 34

Analysis of Differences . . . . . . . . . . . . . . . . 46

CHAPTER FOUR -- CONCLUSIONS . . . . . . . . . . . . . . 71

Appendix A. Instructions . . . . . . . . . . . . . . . . 75

REFERENCES . . . . . . . . . . . . . . . . . . . . . . . 81

List of Illustrations


Figure 1. Examples of Iconic Representations (from Kleiner . . . . . . . . . . . . . . . . . . 9

Figure 2. Examples of polygons for the experiment. . 21

Figure 3. Examples of levels of additional information. 23

Figure 4. Examples of form variable. . . . . . . . . 25

Figure 5. Experimental design. . . . . . . . . . . . 31

Figure 6. Clusters on the basis of frequency of pairs. 37

Figure 7. Plot of principal components for city data. 40

Figure 8. Polygons for four California cities. . . . 41

Figure 9. Principal components plot for car data. . . 44

Figure 10. Plots of canonical discriminant scores . . 45

Figure 11. Clusters on the basis of K-Means technique. 49

Figure 12. Simplified dendrogram for car data. . . . . 50

Figure 13. Dendrogram from SAS for city data. . . . . 52

Figure 14. Clusters on the basis of hierarchical technique. . . . . . . . . . . . . . . . . . . 53

Figure 15. Analysis of Variance Summary . . . . . . . 55

List of Tables

Table 1. City Data Set. ....................................26

Table 2. Car Data Set. .....................................27

Table 3. Frequency Count for City Data Set. ............35

Table 4. Frequency Count for Car Data Set. .............36

Table 5. Difference Totals with K-Means Algorithm. .....56

Table 6. Anova Results with K-Means Algorithm. .........57

Table 7. Difference Totals with Hierarchical Algorithm. 61

Table 8. Anova Results with Hierarchical Algorithm. ....62

Table 9. Difference Totals from Pair Comparisons. ......66

Table 10. Anova Results from Pair Comparisons. ..........67

# CHAPTER ONE -- GRAPHICS AND PERFORMANCE

## Introduction

The representation of quantity seems to have developed at the same time as written language in human history, but only recently have graphic forms for representing quantitative ideas been developed. The concept of drawing upon the human visual system's capacity for perceiving and comparing patterns, thus allowing the integration of large numbers of individual information items, seems to have flowered in the late eighteenth century, particularly in the work of William Playfair (see examples in Tufte, 1983). While interest in graphical presentation has varied over the years since, in many respects the field has not progressed beyond these early works. The last two decades have seen a renewed enthusiasm for graphical methods of data presentation, drawing in part on an increasing emphasis on visual media and on the growing capabilities of computing and related machinery. One of the areas of new interest is graphic representation in multivariate statistics, a field which itself owes much recent development to applications of the computer.

## Work Relating to Graphics and Human Performance

Work on the graphic representation of quantitative information is found in the literature of a number of different disciplines, each of which propounds its own point of view. In reviewing this work one must be prepared to range over a broad spectrum of fields, from statistics, graphical arts and cartography to education and ergonomics. Within this variety, though, and, indeed, maybe because of it, there has developed no accepted theoretical basis for visual graphics, nor even a consistent body of terminology. At the most basic level, for example, there is the inconsistency in the use of the terms "graph" and "chart," compounded by the word "graphic"; these terms are found interchangeably. MacDonald-Ross (1977) gives a brief glossary and discussion of certain inconsistencies, such as those associated with "nomograph"; various handbooks also provide sets of terms. There is no generally accepted classification of graphic techniques. Furthermore, and what is of interest here, there has only recently evolved the concept of investigating the characteristics and forms of various graphical presentations by observing or measuring the performance of the user (see Kruskal, 1975).

Tufte's recent book (1983) provides a good introduction and background to quantitative graphics, as well as examples of some of the best of the "art." Beniger and Robyn (1976) provide a brief historical overview; Feinberg (1979) has reviewed developments in statistical graphics and noted the paradoxical trend of recent renewed interest in graphics but their generally decreasing use in technical publications. Cleveland (1984b) has also surveyed usage in scientific periodicals; Wainer and Thissen (1981) also provide an overview of recent developments with good examples. The last two decades have been a period of innovation, particularly in exploratory methods of data analysis, in probability plotting procedures, and in multivariate techniques. Izenman (1980) has comprehensively reviewed the contributions during this period. To aid those who employ visual graphics, a number of handbooks or manuals of techniques have appeared, based on intuition and aesthetics; those by Schmid and Schmid (1979), Schmid (1983), and Chambers, Cleveland, Kleiner, and Tukey (1983) are recommended. Furthermore, there have been some efforts expended at various times toward the establishment of standards for graphic presentation (Schmid, 1976). These efforts often appear to have had little impact, though; examples of poor graphics continue to appear regularly, ranging from those that are merely confusing to some that are deceptive (Wainer, 1980; 1984).

Although new methods and forms of the graphical presentation of quantitative data have been developed since the time of Playfair, there has been relatively little empirical evidence established for the preference of particular methods in a given circumstance, or for the use of particular features of a graphic type to best serve the function intended. Several articles have at least partially reviewed the work that has been done (Feinberg, 1979; Kruskal, 1982; MacDonald-Ross, 1977; Wright, 1977). It is interesting to note that Charles Babbage, the forefather of computing machinery, was one of the first to express concern for the presentation of data and its effect on the observer (see Kruskal, 1982). During the 1920's and 1930's some studies were undertaken, primarily by statisticians, to contrast the relative merits of the bar and circle (or, pie) graphs (Croxton, 1927; Croxton and Stein, 1932; Croxton and Stryker, 1927; Eells, 1926; Huhn, 1927; Graham, 1937). There are problems with generalizing from these early experiments, though, due to the limited variety of graphical representation and some methodological considerations; further, their results are at times inconsistent.

This "Bar-Circle Debate," as Kruskal (1982) has termed it, is continuing to the present. Peterson and Shramm (1954) found circles to be more accurately used; Culbertson and Powers (1959) found multiple bars better. Cleveland and McGill (1984), in a sound article attempting to generate a theoretical basis for graphic perception, reported that their subjects could estimate proportion significantly more accurately from multiple or grouped bars than from circle graphs. This result, in part, sustains their hypothesis that judgments of position along a scale are better than those of angle or area. Further, they proposed using dot charts in place of both (Cleveland and McGill, 1984; Cleveland, 1984a). This long-running controversy has spilled over into the cartographic literature, as will be mentioned later.

This debate is indicative of the problems entailed in relating human performance and graphics. Bar and pie charts continue to be the most frequently used graphs in most fields; many software packages will turn them out with ease. And yet, there is still not a solid basis for using or condemning (completely) one form or the other, or for which are the features that will make either best convey what it is supposed to convey. Cleveland and McGill's work is solid and it is hoped that more such will be undertaken to provide a firmer basis for guidance to the designers of graphs. But, we should not overlook the complexity of even these simple graph forms.

Kruskal (1982) has a good discussion of the problems of criteria for judging graphs. In their article, for example, Cleveland and McGill (1984) demonstrate the advantage of showing the difference between five approximately equal portions of a whole, by using dot rather than circle charts, but we should ask whether it is the differences we actually want to convey, or is the approximate similarity more important. If the latter, the circle would seem to work at least as well. Secondly, the impact of portraying the division of a whole of something is more forceful in the circle graph. As they point out, these forms are usually used for data presentation rather than exploration; in such use many factors must be considered. The study of graphical perception and use is only beginning.

Following the Second World War better experimental techniques were applied to the investigation of the relation of graphic characteristics and performance for functional values presented as graphs and as tables (see Carter, 1947a; 1947b). This work was followed more than a decade later by investigations of trend representations in graphs by Schutz (1961a; 1961b).

The recent period of interest in graphics has seen a variety of new techniques and formats being proposed, particularly in the field of statistics. As computer technology has advanced, new graphical forms have been developed to take advantage of the computer's capacity for data manipulation. Research concerning the relation of graphic techniques and performance has also gained interest. Wainer has proposed the development of a body of empirical results which could aid graphic designers in choosing appropriate parameters for specific purposes. He conducted several experiments with this intent, investigating the use of rootograms, a graphical representation of the residuals of the root of nonlinear fit, and a comparison of graph types (1974; Wainer and Reiser, 1976). Examples of a similar nature include investigations of correlation estimation parameters (Cleveland, Harris, and McGill, 1983), the use of bar graph displays for process control (Verhagen, 1981) and further investigations of the relationship between tabular and graphic display (Feliciano, Powers, and Kearl, 1963; Ghani and Lusk, 1982; Powers, Lashley, Sanchez, and Schneiderman, 1984; Remus, 1984). Considering the widespread use of graphics for the representation of data and the availability of computer programs to produce displays of data, though, there has been relatively little work to indicate which types or characteristics of a given type of display are best for fulfilling a particular purpose.

The use of graphic representation has also been studied in the field of education. There has been some indication that graphics are not intrinsically more interpretable than textual material and that their redundant use may be detrimental (Feliciano et al., 1963; Preece, 1983; Roller, 1980; Vernon, 1946, 1950). Other work has investigated certain further aspects of the use of graphics in education (see Eggen, Kauchak, and Kirk, 1978; Kirk, Eggen, and Kauchak, 1978; Washburne, 1927). Cartographers have also shown concern for the relationship between the characteristics of graphics and their capacity to communicate information quickly and accurately (see for review, Phillips, 1979; Potash, 1977). Thematic maps are a special class of quantitative data graphics. The use of symbols, particularly graduated circles and circle graphs, has been extensively investigated (see Chang, 1977; Cleveland, Harris, and McGill, 1982; Cox, 1976; Flannery, 1971; Meihoefer, 1973), as well as various aspects of the use of color for representing area or quantitative values (Cleveland and McGill, 1983; Dobson, 1980; Wainer and Francolini, 1978). Considering the widespread use of data graphics in texts and maps more study of their role in the communication of information is surely needed.

Among the innovative graphic techniques which have been proposed in recent years, those which deal with multivariate data have generated some interest. These techniques derive from the general increase in work on multivariate analysis which has accompanied the growth in computer applications in statistics. Feinberg (1979) and Wainer (Wainer and Thissen, 1981) devote sections of their reviews to these developments, along with providing illustrative examples. Everitt (1978) has a volume on the graphical presentation of multivariate data; Chambers, Cleveland, Kleiner, and Tukey (1983) include a chapter on multivariate methods in their recent work on graphics for data analysis.

One group of techniques for the representation of multivariate quantitative data, termed point representation, uses a particular symbol or icon to represent each point. The different techniques vary from each other primarily in their use of different symbols, ranging from profiles or bar charts to characterized representations of the human face (see Figure 1 for examples, from Kleiner and Hartigan, 1981). The basic procedure is to represent each object of the set of objects to be compared as an individual graphical unit, a symbol whose appearance is determined by the values of the variables measured for the object. The objects of the set can then be compared by observing the set of symbols thus

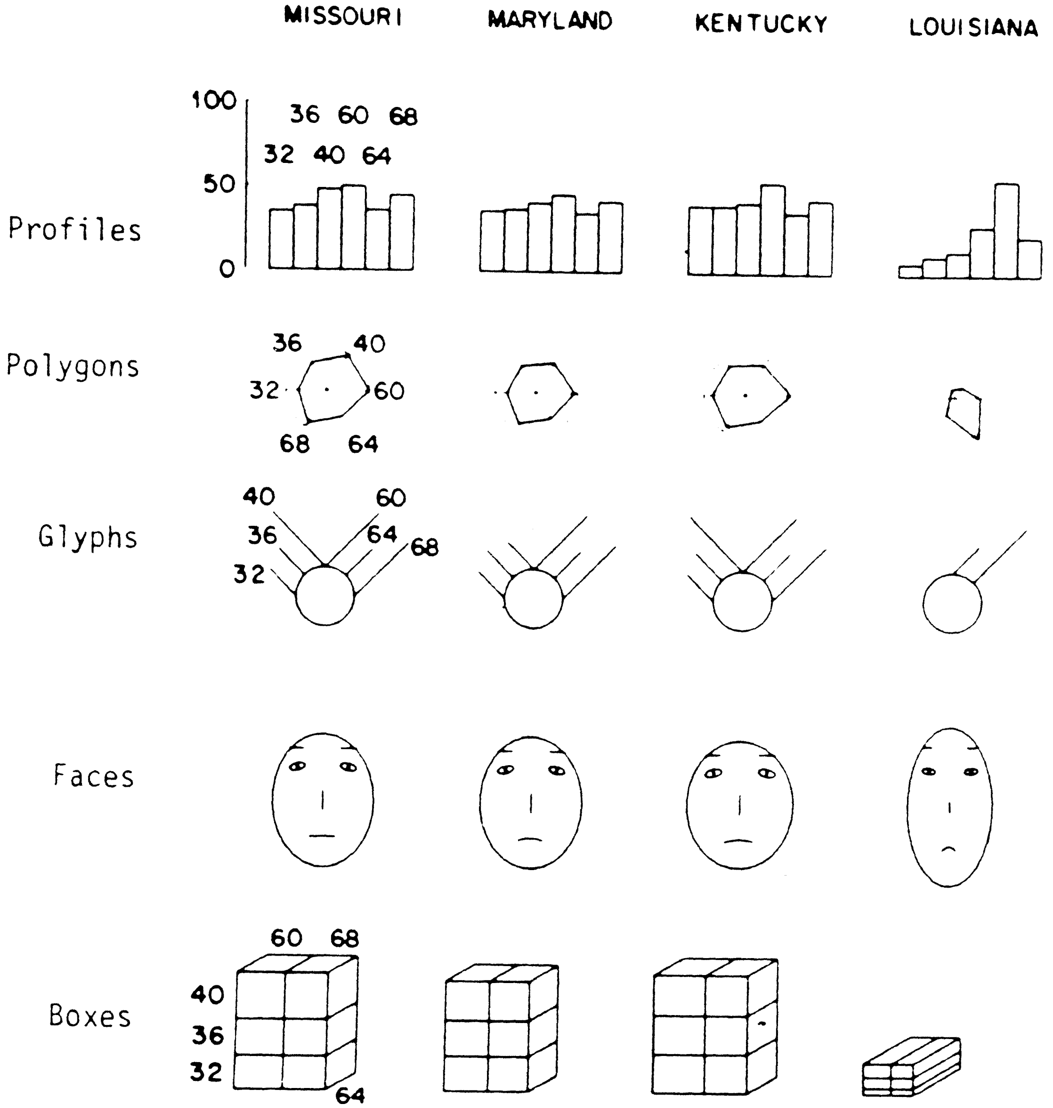



Figure 1. Examples of Iconic Representations (from Kleiner and Hartigan, 1981).

generated. Each symbol provides a single pattern of that data point. Such techniques are usually employed for preliminary cluster identification, detection of unusual data points, and, to a lesser degree, trend identification. Among the advantages often cited for iconic displays is their capacity for each to be perceived in a holistic fashion, allowing comparison of points by the comparison of the single image of each. They allow the user to observe the structure of the data, by eye, directly from the data values and without the scaling and correlation metrics necessary for joint representations. Further, most of these techniques are designed to make use of computer processing and graphic capabilities, allowing the creation of automated systems for producing graphics for initial or exploratory analysis of data.

These techniques have been employed for data analysis in a number of fields, at least experimentally. Friedman, Farrell, Goldwyn, Miller and Siegel (1972) reported on the use of polygons to classify pathophysiological stages in septic shock; Jacob (1978) reported on the use of cartoon faces, proposed by Chernoff (1972), for classification of personality profiles. Examples of applications in other areas include those reported by Bruckner (1978) at Los Alamos Laboratories and by McDonald and Ayers (1978) at General Motors. An interesting use of faces in presenting basic statistical concepts to beginning students has been proposed by

Pickover (1984). Bertin's recent work (1981) contains a number of examples of the graphical analysis of multivariate data sets in a variety of fields. In his introductory chapter he outlines the use of his matrix procedure to analyze occupancy data for a resort hotel and shows how such analysis can be used for management decisions regarding establishment of rates. The technique first establishes a data matrix, then converts each row to a profile representation (or other visual variable), and clusters the data by permutation of the rows. The display can be used as a method of exploratory analysis, from which further data analysis can proceed, as in this case, by examining the characteristics of the clusters. The display, with some modification, can also be used to communicate the results to others. Thus, this type of display can fulfill two of the three purposes which are commonly cited for the graphic presentation of data -- analysis, communication, and computation (e.g., see Chernoff, 1978; Tufte, 1983).

The relationship between these point or iconic representation techniques and performance has been explored in several studies. Jacob (1978; Jacob, Egeth, and Bevan, 1976) has reported on a series of experiments concerned primarily with the use of Chernoff faces. In a comparison of faces with polygons and digits for clustering data with nine variables, by pattern recognition, faces were found to be significantly more accurate; polygons were as fast, but were less accurate.

In a comparison of forms in a paired-associate learning task, faces tended to be better than the other forms. Mezzich and Worthington (1978) investigated pattern recognition of data by comparing seven forms of representation, including linear profiles, circular profiles (polygons), faces, linear and polar Fourier series, factor analysis in two dimensions, and an ordinal multidimensional scaling. The data consisted of the values of 17 variables for four archetypical psychological patients as assigned by 11 psychiatrists. The factor analysis and multidimensional scaling representations were found to provide the best performance, with the polar Fourier icons next. Chernoff and Rizvi (1975) investigated the effects of the relationship of the features in using faces for data representation by randomly changing the features assigned to the variables. Such changes were found to affect the results of a classification task (dichotomous clustering) by about 25 per cent. A similar problem of the interrelatedness of variable assignment and the perception of form is cited for other forms of iconic representation (see Bruckner, 1978; Egeth, Jacob, Wainer, Kleiner, and Hartigan, 1981; Kleiner and Hartigan, 1981; Naveh-Benjamin and Pachella, 1982).

Wilkinson (1982) compared the performance for four icon types, blobs (polar Fourier series), castles, faces, and polygons, at two levels of dimensionality, by having subjects judge dissimilarity in pairwise comparisons. There were found to be significant differences in reliability and validity among display types, with faces proving better in both. There was also a significant dimensionality effect, but the interaction between dimensionality and type was not significant. Freni-Titulaer and Louv (1984) compared castles, trees (Kleiner and Hartigan, 1981), bar profiles, and bar profiles with the variables ordered according to hierarchical clustering. Subjects sorted the stimuli into two clusters; time and accuracy were measured. Trees were clustered more quickly and more accurately than the other forms. Some effects were also found from differences in the data sets used to generate the graphics.

These studies, for the most part, have been comparative in nature; there has been little effort to identify which of the point representations functions best in particular circumstances, with particular types of data, or for particular tasks. Some forms have not been used in the studies; Bertin's thorough and thoughtful work (1981; 1983), for example, has been mentioned only in passing in most studies. Moreover, there has been little investigation of the relationship between the graphic characteristics of a particular type of representation and the performance with that graphic.

## CHAPTER TWO -- THE PRESENT STUDY

### Background for the Display and the Task

This present study was designed to investigate in more depth the effect of variation in certain characteristics of the iconic representation called Polygons (or variously, Stars or Snowflakes) on the ability of individuals to perform a clustering task. Polygons are formed by representing the value of the variables measured for each object as a point along radii of a circle and connecting these points. Basically each polygon is a profile representation, or line graph, in polar co-ordinates.

The use of polygons for representing multivariate data has been reported in various fields, and several of the comparative studies of multivariate techniques mentioned earlier have included them. Polygons have been used for data exploration and presentation in some studies; Hanson, Kraut, and Farber (1984), for example, used this technique in a study of use patterns for UNIX commands; Zelenka, Cherry, Nir, and Siegal (1984) used polygon displays to present data on the variation in the growth of quail. A single polygon has been employed as a graphical data display in investigations of the role of integration of information (Carswell and Wickens, 1984; Goldsmith and Schvaneveldt, 1982). Goldsmith and Schvaneveldt noted that performance was better for polygons than Chernoff faces in pilot studies. This type of display has been employed in a proposed Safety Parameter Display

System for nuclear power plants (Petersen, Banks, and Gertman, 1982; Woods, Wise, and Hanes, 1981, 1982). Polygon displays are the basis of the decision polar graph, a visual representation of data on multiple criteria intended as an aid to the management decision process; Frazelle (1985) shows an example for alternative material handling systems.

While polygons have not fared as well as some of the other point representation techniques in some comparative studies, they have proved better in others. They are in many respects more straightforward and simpler to understand than some of the other icons, such as blobs (Fourier series in polar co-ordinates) or castles. Faces are burdened with problems of subjective interpretation and correlation of variables. Everitt (1978) has a good example of the dependence of this type of display on the relation between the variables of the data and the facial features used to represent them. Cleveland and McGill (1984) point out the difficulty of interpreting faces due to the complexity of the perceptual judgments which have to be made. The observer must compare such diverse features as length of nose, slant of eye, shape of face, and curvature of mouth, for example, to extract a sense of the relation of the variables in a single data point. Trees and castles depend on the hierarchical clustering of the variables; such clustering will vary with the data sets and may not be desirable in some instances.

The polygons used in comparative studies have varied in their visual features; none of these studies has tried to determine their optimal characteristics. Because of their simplicity, ease of use, and their applications in experimental tasks, process monitoring, and representation for decision making, further investigation of polygons seemed justified. By identifying the characteristics which lead to better performance, polygons can be employed more efficiently and more validly compared to other display types.

Of the tasks for which point representations are generally recommended, it was decided to investigate the performance in a clustering task. This task is one that has various applications, is related to other tasks for which polygons are suited, and has not been investigated in detail with polygons. Given a set of objects for which various attributes are measured, one divides the objects into groups on the basis of their similarity. Such a task may often be an initial step in data analysis. The hotel occupancy example from Bertin, mentioned earlier, illustrates the use of this technique in analyzing data as part of a decision process; plant location, material handling alternatives, or similar problems with multiple variables could be handled in this manner. Clustering is related to categorization, and as such could find application in tasks which require one to categorize a new object into one of several classes on the basis of its similarity to a typical member of the class. Identification tasks or status displays might be considered related to this task. Finally, polygon displays find application in data presentation, to support or explain a clustering or categorization task.

Clustering and categorization have been of interest in both cognitive psychology and statistics. The terms categorization and classification are at times used almost interchangeably in the psychological literature (e.g., Reed, 1972), but here categorization will be used to denote such a task. Classification will be used to denote the process of forming a hierarchy of objects or groups of objects, as a taxonomy. The two processes are not exactly the same, though they are related. At a particular level of a classification hierarchy, one would categorize objects into groups.

Categorization seems to be basic to an organism's information processing. Rosch (1978) points out two principles underlying the process, the need for some efficient or economical way to deal with the complexity of information in the world and the view that the world of stimuli has a correlated structure. These principles, especially the first, are noted by most in discussing categorization. The psychological theories of the process are varied; Anderson (1980) has a good discussion of the various theories and some of the problems with each. At present some form of prototype or schema theory seems to be the most widely supported, depending, in part, on the definition of prototype or schema. According to this theory some concept of a typical member of each category is formed by the observer, whether as a full member or as a set of rules for membership, and new stimuli are then categorized on the basis of these prototypes or schemata.

In categorizing a group of stimuli, it is generally assumed that the categories are formed in such a manner as to retain the maximum similarity between members of the category and maximum difference between contrasting categories (Rosch, 1978; Tversky and Gati, 1978). Similarity judgment has been modelled on the basis of a multidimensional geometric space, similarity being related to some distance measure when stimuli are considered points in this space. Reed (1972), for example, studied a number of different distance metrics as the basis for categorization of visual stimuli, cartoon faces like Chernoff faces. Recently several other models have been proposed. The cue validity model (Reed, 1972; Rosch, 1978) relates similarity to a function of the conditional probabilities of category membership for each aspect of the stimulus. Tversky (1977) has proposed a model, termed the contrast model, in which similarity is a linear function of the common and distinct aspects of the stimuli. These more recent models were developed to try to explain certain problems in the purely geometric models, such as the asymmetry of judgments depending upon the direction of comparison and the role of context. In general, though, the geometric model still provides a reasonable approximation for many tasks, as Tversky (Tversky and Gati, 1978) points out.

In the field of statistics a group of procedures for categorizing objects has been developed, known most commonly as cluster analysis. Introductions to these techniques can be found in various multivariate texts (e.g., Dillon and Goldstein, 1984) and in several single volumes devoted to the subject (Everitt, 1974; Hartigan, 1976). The techniques begin with a distance or similarity matrix of the various objects to be clustered. Two basic methods of finding clusters are used. In hierarchical techniques the objects are progressively merged on the basis of some metric, adding members and combining clusters until one cluster is finally formed; alternatively, some hierarchical procedures begin with one cluster, progressively dividing the clusters. The other group of techniques includes those which partition the objects into a predetermined number of mutually exclusive clusters by minimizing some within cluster metric and maximizing some between cluster metric. There is a variety of clustering methods, and each has its own advantages and problems. The results from different techniques often vary.

It has been suggested that point representations such as polygons might be used for rough clustering or approximating the number of clusters in an exploratory look at the data, as well as for presentation of the patterns in the data (e.g., Everitt, 1978; Dillon and Goldstein, 1984). Such a graphic technique could be used by individuals not well versed in the arcana of matrix algebra and multivariate analysis techniques. For polygons to be effectively used in tasks such as clustering they should be able to be consistently clustered. Visual variables which encourage such consistency should be employed when designing such a display.

The Experiment

The Display. Although some software packages will produce polygons, SAS/GRAPH (Statistical Analysis System) for example, the variations for this investigation necessitated writing programs to produce the displays. Figures were plotted using a Versatec 1200 electrostatic plotter, then separated so that a single polygon appeared on each three by four inch (7.6 x 10.2 cm.) card, with a title and number below. This legend provided a base to orient the figure and allowed keeping track of the cluster membership (see Figure 2).

Certain aspects, actually "dimensions" to use Garner's (1970, 1978) terminology, of the graphic were varied to determine whether changes in them would affect the consistency of clustering. The first visual aspect to be varied was the shading of the figure. It was hypothesized that a shaded figure would be perceived more readily as a whole than an outline and be more consistently clustered. The reported studies have all used outline figures.

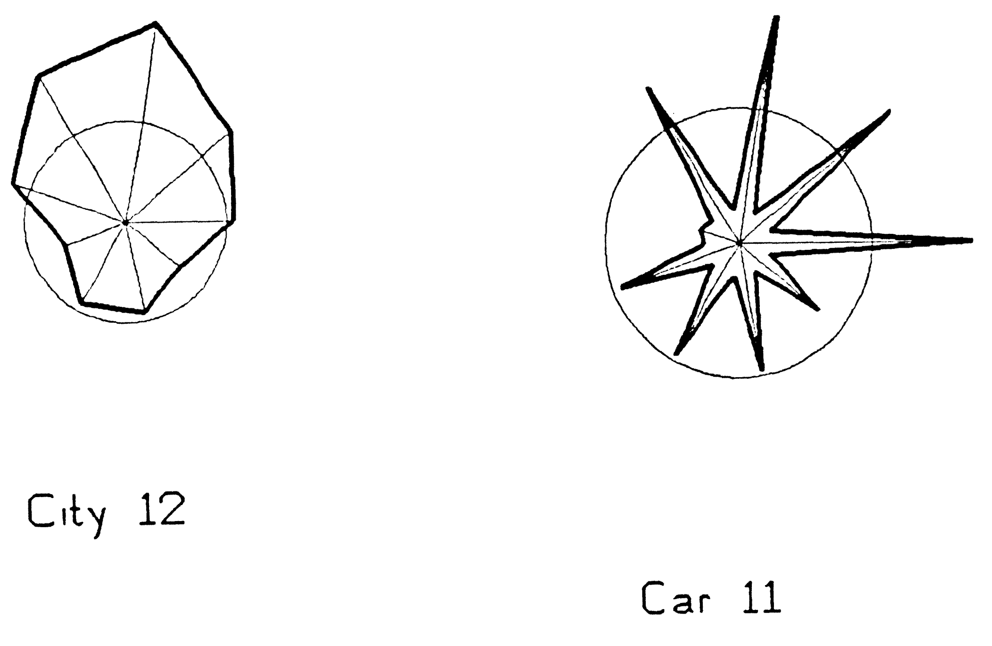



Figure 2. Examples of polygons for the experiment.

The amount of additional information displayed with each polygon was thought to interact with the tasks for which it is used. Goldsmith and Schvaneveldt (1982) noted an increase in performance on an information integration task with a polygon display which included internal radii. For the task of clustering, though, one might expect that excess information would detract from the comparisons by shape. Three levels of additional information were used in this study. The first was no additional information; the polygon alone was presented. The second level included a background circle through the means of each variable and internal radii. The final level included a circular, polar grid, against which the polygon was presented, including internal radii (see Figure 3).

The task of clustering is one of grouping the figures on the bases of their perceived similarity or dissimilarity. The variation between polygons, of course, is created by the differences in vector length of the radii; judgments of length may play a role in the comparisons. It might be expected, then, that accentuating this length would be beneficial to performing the clustering task. On the other hand, the grouping task may be performed more on the perception of the overall shape or pattern, in which case an accentuated figure would prove no easier to group than the standard polygon. To investigate the relation between form and performance, polygons were presented in two different forms, the

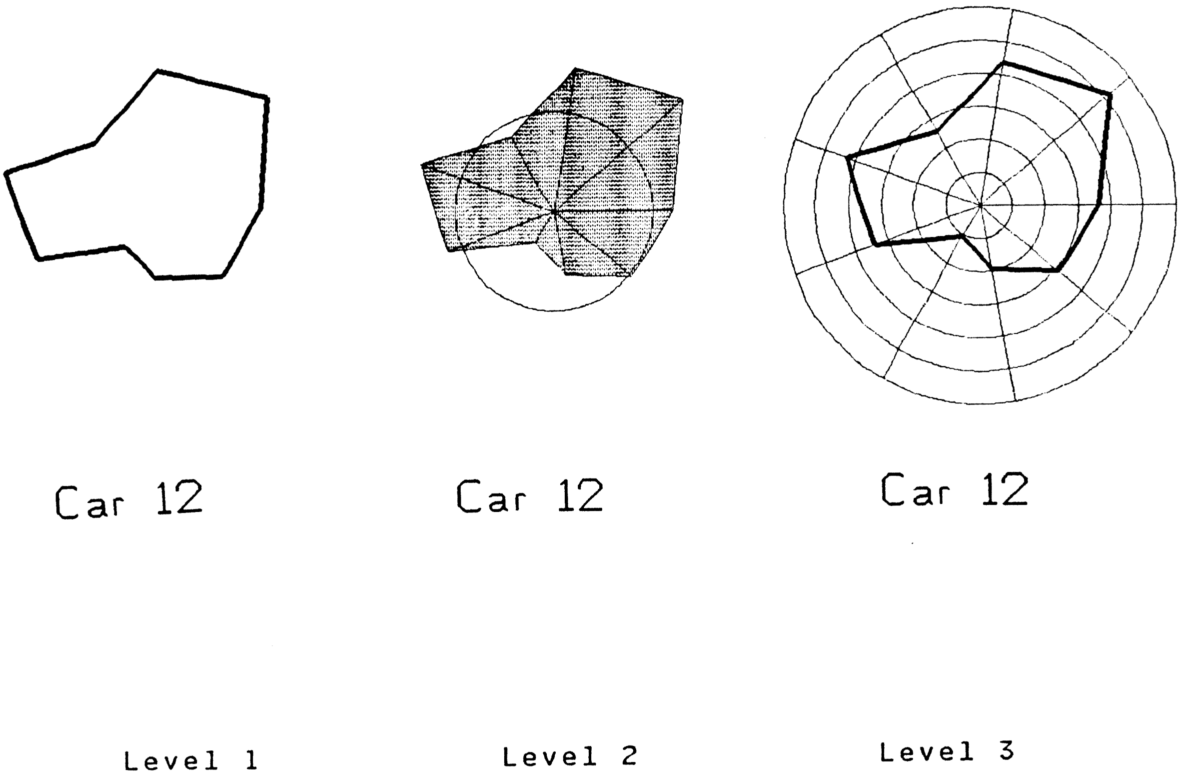



Figure 3. Examples of levels of additional information.

standard polygon and an accentuated form, the star. By connecting each value point on the radius with a small base circle mid-way between each pair of radii, the length along the radius is accentuated. The rays of the star were approximately the same length as the radii of the corresponding standard polygon, but the attribute of length will be visually emphasized (see Figure 4).

Two data sets were selected for use in the experiment (see Tables 1 and 2). The first is a subset of data used in examples in Chambers , Cleveland, Kleiner, and Tukey (1983); the second came from McDonald and Ayers (1978). Subsets of the variables were chosen, nine in each case, to give reasonably complex polygons without being too complex. I have seen no reported studies of the relation between the number of variables portrayed with polygons and performance, but it would seem intuitive that there is an upper limit. Representations with very few variables also seem more difficult to cluster. The number of variables, of course, would be related to the perceived complexity of the figures; Attneave (1957) found that degree of perceived complexity was primarily determined by the number of turns in a figure, but also by the angular variability, which would be a function of the juxtaposition of variables and their value in each individual case. The variables were selected to try to preserve some of the relationships in the full data sets, though

Polygon

Star

Figure 4.  Examples of form variable.

TABLE 1

City Data Set (from McDonald and Ayers, 1978)

DATA FOR CITIES

| Precip. in. | Jan. Temp. | July Temp. | Years Educ. | Pop. Density sq. mi. | HC | NOx | $SO_2$ | Mortality Rate | Number | City |
|---|---|---|---|---|---|---|---|---|---|---|
| 36 | 27 | 71 | 11.4 | 3243 | 21 | 15 | 59 | 921.870 | 1 | Akron |
| 35 | 23 | 72 | 11 | 4281 | 8 | 10 | 39 | 997.875 | 2 | Albany |
| 47 | 45 | 79 | 11.1 | 3125 | 18 | 8 | 24 | 982.291 | 3 | Atlanta |
| 53 | 45 | 80 | 10.2 | 3325 | 30 | 32 | 72 | 1030.380 | 4 | Birmingham |
| 36 | 27 | 72 | 10.7 | 4213 | 12 | 7 | 20 | 912.347 | 5 | Canton |
| 52 | 42 | 79 | 9.6 | 2302 | 18 | 8 | 27 | 1017.613 | 6 | Chattanooga |
| 35 | 28 | 71 | 11.1 | 3042 | 31 | 21 | 64 | 985.950 | 7 | Cleveland |
| 35 | 46 | 85 | 11.8 | 1441 | 1 | 1 | 1 | 860.101 | 8 | Dallas |
| 31 | 24 | 72 | 10.9 | 3226 | 5 | 3 | 10 | 871.338 | 9 | Grand Rapids |
| 31 | 45 | 85 | 11.4 | 1844 | 1 | 1 | 1 | 891.708 | 10 | Ft. Worth |
| 42 | 40 | 77 | 10.4 | 2269 | 8 | 3 | 5 | 971.122 | 11 | Greensboro |
| 46 | 55 | 84 | 11.4 | 2647 | 6 | 5 | 1 | 952.529 | 12 | Houston |
| 35 | 31 | 81 | 12 | 3262 | 7 | 4 | 4 | 919.729 | 13 | Kansas City |
| 11 | 53 | 68 | 12.1 | 4700 | 648 | 319 | 130 | 861.833 | 14 | Los Angeles |
| 37 | 31 | 75 | 11.9 | 4259 | 23 | 9 | 15 | 958.839 | 15 | Columbus |
| 50 | 42 | 82 | 10.4 | 3497 | 15 | 18 | 34 | 1006.490 | 16 | Memphis |
| 45 | 40 | 80 | 10.1 | 2682 | 17 | 14 | 78 | 961.009 | 17 | Nashville |
| 10 | 55 | 70 | 12.1 | 3033 | 144 | 66 | 20 | 839.709 | 18 | San Diego |
| 54 | 54 | 81 | 9.7 | 3172 | 20 | 17 | 1 | 1113.156 | 19 | New Orleans |
| 36 | 30 | 75 | 11.4 | 4029 | 6 | 4 | 16 | 936.234 | 20 | Dayton |
| 13 | 49 | 68 | 12.2 | 2702 | 105 | 32 | 3 | 790.732 | 21 | San Jose |
| 44 | 39 | 78 | 11 | 3768 | 12 | 9 | 48 | 1025.502 | 22 | Richmond |
| 30 | 24 | 72 | 10.8 | 3694 | 11 | 4 | 11 | 941.181 | 23 | Flint |
| 32 | 25 | 72 | 11.1 | 4335 | 7 | 4 | 18 | 874.281 | 24 | Rochester |
| 38 | 28 | 72 | 10.7 | 3451 | 14 | 13 | 39 | 954.442 | 25 | Youngstown |
| 18 | 48 | 63 | 12.2 | 4253 | 311 | 171 | 86 | 911.701 | 26 | San Francisco |
| 39 | 29 | 75 | 11.4 | 4412 | 13 | 7 | 33 | 968.665 | 27 | Indianapolis |
| 46 | 30 | 72 | 11.3 | 3327 | 4 | 3 | 8 | 923.234 | 28 | New Haven |

TABLE 2

Car Data Set (from Chambers, et al., 1983)

CAR DATA SET

| Price | Displac. cu. in. | Rear Seat Room | Weight lbs. | Head Room in. | MPG | Repair Record 1 - 5 | Trunk Size cu. ft. | Length in. | Number | Model |
|---|---|---|---|---|---|---|---|---|---|---|
| 14500 | 350 | 30 | 3900. | 3.5 | 14 | 2 | 16 | 204 | 1 | Cad. Eldorado |
| 4504 | 200 | 28.5 | 3180. | 3.5 | 22 | 3 | 17 | 193 | 2 | Chev. Malibu |
| 4589 | 85 | 23.5 | 2020. | 3.5 | 35 | 5 | 8 | 165 | 3 | Darsun 210 |
| 5010 | 318 | 29 | 3600. | 4 | 18 | 2 | 17 | 206 | 4 | Dodge Diplomat |
| 13594 | 400 | 28.5 | 4720. | 2.5 | 12 | 3 | 18 | 230 | 5 | Lincoln Cont. M.V |
| 3748 | 97 | 24.5 | 2200. | 3 | 31 | 5 | 9 | 165 | 6 | Toyota Corolla |
| 4816 | 196 | 29 | 3250 | 4.5 | 20 | 3 | 16 | 196 | 7 | Buick Century |
| 15906 | 350 | 30 | 4290 | 3 | 21 | 3 | 13 | 204 | 8 | Cad. Seville |
| 3984 | 98 | 24 | 2120 | 2 | 30 | 5 | 8 | 163 | 9 | Dodge Colt |
| 6342 | 225 | 28 | 3740 | 4.5 | 17 | 2 | 21 | 220 | 10 | Dodge St.Regis |
| 4499 | 91 | 23.5 | 1760 | 2.5 | 28 | 4 | 5 | 149 | 11 | Honda Civic |
| 13466 | 302 | 27 | 3830 | 3.5 | 14 | 3 | 15 | 201 | 12 | Lincoln Versailles |
| 3995 | 86 | 25.5 | 1980 | 3.5 | 30 | 4 | 11 | 154 | 13 | Mazada GLC |
| 3291 | 140 | 29 | 2830 | 3.5 | 20 | 3 | 17 | 195 | 14 | Merc. Zephyr |
| 4733 | 231 | 28 | 3300 | 4.5 | 19 | 3 | 16 | 198 | 15 | Olds. Cutlass |
| 10371 | 350 | 30 | 4030 | 3.5 | 16 | 3 | 17 | 206 | 16 | Olds. Toronado |
| 3798 | 97 | 25.5 | 2050 | 2.5 | 35 | 5 | 11 | 164 | 17 | Subaru |
| 4389 | 98 | 26 | 1800 | 2 | 28 | 4 | 9 | 147 | 18 | Ford Fiesta |
| 4425 | 86 | 23 | 1800 | 2.5 | 34 | 5 | 11 | 157 | 19 | Plym. Champ |
| 6165 | 302 | 30.5 | 3720. | 3.5 | 15. | 3 | 23 | 212 | 20 | Merc. Marquis |
| 8814 | 350 | 31.5 | 4060 | 4 | 21 | 4 | 20 | 220 | 21 | Olds. 98 |
| 7827 | 350 | 31.5 | 4080. | 4 | 15. | 4 | 20 | 222 | 22 | Buick Electra |
| 5222 | 231 | 28.5 | 3210 | 2 | 19 | 3 | 16 | 201 | 23 | Pont. Grand Prix |
| 5798 | 231 | 29 | 3700 | 4 | 18 | 4 | 20 | 214 | 24 | Pont. Catalina |

the decision was, at times, arbitrary. Basically, a workable subset of the full data sets was sought.

A group of between 20 and 30 objects was chosen which seemed to divide reasonably well into a small number of clusters, as determined by pretesting. Enough objects were included to make the task non-trivial and still allow subjects to complete four clustering tasks in about an hour. All manipulations and decisions on the data sets were made using polygon displays generated in the same manner as those for the experimental task.

Earlier studies using icons in categorization tasks have used artificially generated data. Jacob (1978; Jacob et al., 1976) had subjects stimuli categorize with various icon representations from randomly generated data. Chernoff and Rizvi (1975) and Wilkinson (1982) also used random data in categorization and similarity judgment tasks; Freni-Titulaer and Louv (1984) generated special data sets, varying specific parameters. It is certainly easier to study the effects of various types of data on the task with generated sets, but in an exploratory task, for which icons are often recommended, one probably won't know what parameters characterize the data. Thus it was felt that real data would at least have face validity.

A second factor with randomly generated data is related to its use as the criterion with accuracy as the measure. This issue is at the basis of judging categorization. If we generate two data sets with a pseudo-random number generator, about two distinct points in multidimensional space, such that there is no overlap, can we rightly say that the subjects "miscategorized" the points if they don't reproduce the geometric clusters? Geometry may not be the appropriate measure for pattern similarity of figures, nor for similarities in data. This problem is, to some extent, a part of the difficulties that cluster analysis faces.

Two aspects of the subjects' clusterings were used as measures of performance. The first was the agreement of the clusterings with some standard clustering of the data. With generated data the standard would be provided from the algorithm for generating the data. Since "real" data were being used as the basis of the displays, these data sets were subjected to cluster analysis techniques to derive the standards. Differences between the standard and the subjects' results became the metric.

There are numerous cluster analysis techniques available and the results of different techniques are not always consistent. For any graphic clustering method to be reliable, though, it should be reasonably consistent across individuals, as well as with the same individual across time. Visual aspects of the display which show more consistent performance should be identified. The second measure of performance was consistency across the subjects at each level of the variables. Differences between the subjects' clusterings were used as a metric of this consistency.

Design. The three visual variables of the display and data set were combined in a 2 x 3 x 2 x 2 full factorial experimental design (see Figure 5). The variables of additional information and shading were between-subjects variables. In order to make best use of subjects and control for some of the expected variation, the variable of form was treated as a within-subject variable. Pretesting indicated that this was feasible. The titles of the data sets, printed at the bottom of each polygon, were interchanged between forms, and the order of the polygons varied. Since both data sets were seen by each subject, data set was also a within-subject variable. Each subject thus saw four sets of figures, both data sets at both levels of form.

Subjects. A broad spectrum of subjects participated in the experiment, 36 in all. They came from the academic community and had at least some post-secondary education. They ranged in age and occupation, from students to middle-aged professionals. It was felt that such an exploratory experiment as this should include a range of subjects and not be limited to a specific expertise. Through informal inquiry it was learned that none of the subjects had seen this type of graphic display.

Procedure. The experiment was run over a two and a half week period, all sessions being in the afternoon or early

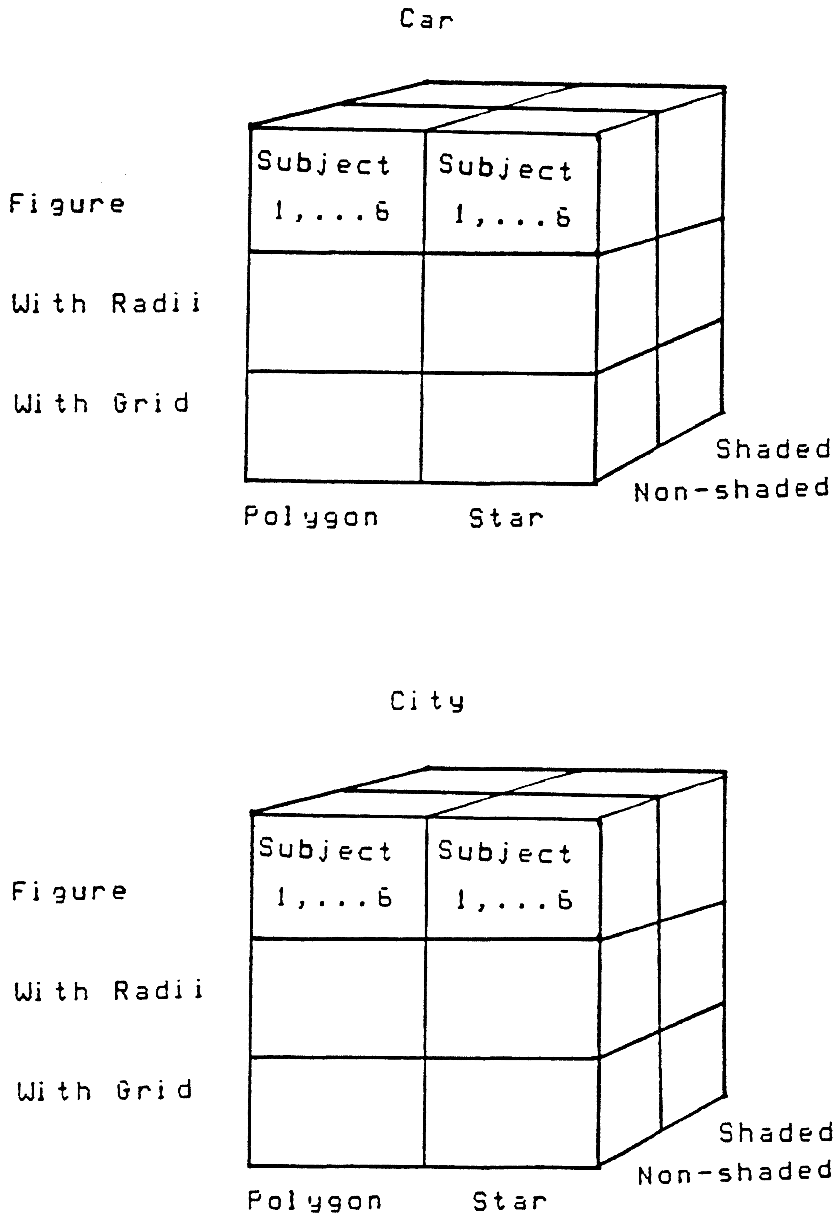



Figure 5. Experimental design.

evening. A classroom at Virginia Polytechnic Institute and State University was used for all sessions, providing a quiet, well-lighted, and comfortable place to work. If more than one subject participated at a time, sessions were staggered.

Each session began with the subject reading the consent form. A set of instructions was then given which provided a brief overview and background of the display and its use, an illustrative example, and the actual instructions for the task (see Appendix). The shading of the figures in the instructions corresponded to that which the subjects would be given; the form, with that which the subjects would see first. After it was determined that the subjects had read and understood the instructions, the four sets of figures were presented, one set at a time. Subjects were requested to initially spread all the figures out before beginning the clustering; the figures were presented in numerical order in a pack. On top of the pack was a card indicating the number of groups into which to divide the figures and explaining the background (level of additional information). The order of presentation of the levels of additional information and of shading was balanced throughout the experiment by rotating through the six combinations. The initial order was determined randomly. The order of form and data set was also balanced.

Training of the subjects was considered, but training must have a well-defined criterion from which feedback can be provided. Several of the studies of graphics in education, mentioned earlier, have commented on the fact that the graphic perception of ideas is not necessarily intuitive, and Kruskal (1982) comments on the need for more work relating to graphic interpretation. This study was intended as exploratory, to investigate several aspects of polygon displays in relation to clustering with a fairly broad range of subject backgrounds. It was intended to look at the subjects' existing ability to perceive pattern and judge similarity. In addition, a specific criterion to which subjects could be trained was lacking. There was no training of subjects undertaken beyond the explanation and example given in the instructions.

An effort was made to keep the sessions informal, though serious, and to answer any questions which might arise. At the end of each session, any further questions were answered, along with discussion of possible applications or problems, if the subject was interested. Many of the subjects expressed interest in the display and some, ways it might be used. Sessions ranged in length from about 45 to 80 minutes.

# CHAPTER THREE -- EXPERIMENTAL RESULTS AND DISCUSSION

## Frequency Investigation

As a preliminary step to the analysis of the effects of the visual variables used in this study, the full results of the subjects' clusterings were reviewed by observing the frequency of the data points being clustered together. This analysis would allow a look at the overall pattern of clustering. A frequency count indicating how often each pair of objects appeared in the same cluster was made (see Tables 3 and 4). The range of values was very broad, as can be seen from the tables. Of the 72 times each pair of data points appeared, for the car data the frequencies ranged from 71 to zero; the city data had a similar range. From these frequency matrices the clusters of points found most frequently together were determined by a heuristic method. The pairs were ordered by frequency; then by descending through the order, members were added to form groups on the basis of their high frequency with other group members and low frequency with members of the other groups. These groups are shown in Figure 6.

Closer examination of the matrix for car data shows one group of cars which was found together very often; these are the subcompacts. The average frequency for pairs in this group was 67.54 ($\underline{s}$ = 2.08); the other two groups showed lower averages and higher deviations (for the group labeled

TABLE 3

Frequency Count for City Data Set

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | City Number |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| - | 50 | 12 | 4 | 54 | 4 | 59 | 0 | 54 | 1 | 16 | 5 | 35 | 1 | 44 | 8 | 8 | 2 | 1 | 58 | 0 | 13 | 55 | 49 | 66 | 0 | 50 | 52 | 1 |
| | - | 3 | 4 | 50 | 7 | 48 | 9 | 56 | 14 | 9 | 1 | 35 | 4 | 42 | 5 | 5 | 3 | 9 | 43 | 5 | 6 | 60 | 58 | 50 | 4 | 49 | 44 | 2 |
| | | - | 20 | 5 | 19 | 16 | 7 | 8 | 7 | 33 | 33 | 16 | 1 | 18 | 25 | 18 | 1 | 11 | 21 | 1 | 65 | 6 | 6 | 12 | 2 | 13 | 18 | 3 |
| | | | - | 5 | 49 | 8 | 35 | 6 | 37 | 33 | 27 | 13 | 7 | 5 | 55 | 50 | 2 | 45 | 3 | 5 | 20 | 6 | 5 | 3 | 7 | 7 | 4 | 4 |
| | | | | - | 1 | 49 | 4 | 49 | 7 | 12 | 7 | 40 | 3 | 49 | 5 | 6 | 3 | 2 | 45 | 4 | 6 | 52 | 54 | 57 | 2 | 48 | 51 | 5 |
| | | | | | - | 5 | 41 | 7 | 38 | 35 | 23 | 8 | 6 | 3 | 49 | 55 | 8 | 49 | 4 | 5 | 18 | 7 | 7 | 2 | 8 | 7 | 3 | 6 |
| | | | | | | - | 1 | 46 | 2 | 17 | 9 | 35 | 2 | 43 | 10 | 10 | 3 | 2 | 53 | 3 | 19 | 47 | 46 | 55 | 1 | 51 | 51 | 7 |
| | | | | | | | - | 7 | 59 | 19 | 29 | 18 | 18 | 6 | 31 | 34 | 18 | 43 | 2 | 20 | 8 | 7 | 7 | 0 | 18 | 7 | 3 | 8 |
| | | | | | | | | - | 11 | 12 | 1 | 29 | 1 | 36 | 4 | 6 | 1 | 7 | 49 | 1 | 9 | 63 | 57 | 55 | 1 | 44 | 49 | 9 |
| | | | | | | | | | - | 19 | 28 | 20 | 13 | 4 | 30 | 32 | 15 | 42 | 2 | 18 | 7 | 12 | 13 | 0 | 14 | 6 | 3 | 10 |
| | | | | | | | | | | - | 32 | 15 | 0 | 13 | 38 | 37 | 3 | 23 | 17 | 4 | 34 | 13 | 12 | 16 | 0 | 14 | 21 | 11 |
| | | | | | | | | | | | - | 22 | 10 | 11 | 33 | 23 | 7 | 30 | 8 | 9 | 34 | 0 | 2 | 3 | 7 | 9 | 14 | 12 |
| | | | | | | | | | | | | - | 5 | 45 | 14 | 12 | 8 | 9 | 35 | 7 | 18 | 30 | 35 | 31 | 4 | 42 | 35 | 13 |
| | | | | | | | | | | | | | - | 3 | 4 | 5 | 57 | 18 | 1 | 58 | 0 | 1 | 1 | 1 | 67 | 1 | 2 | 14 |
| | | | | | | | | | | | | | | - | 6 | 7 | 2 | 1 | 54 | 3 | 20 | 40 | 42 | 44 | 2 | 49 | 47 | 15 |
| | | | | | | | | | | | | | | | - | 52 | 4 | 40 | 5 | 6 | 25 | 4 | 3 | 6 | 6 | 12 | 6 | 16 |
| | | | | | | | | | | | | | | | | - | 9 | 43 | 7 | 8 | 20 | 6 | 5 | 6 | 7 | 6 | 6 | 17 |
| | | | | | | | | | | | | | | | | | - | 16 | 1 | 64 | 3 | 1 | 4 | 1 | 55 | 2 | 2 | 18 |
| | | | | | | | | | | | | | | | | | | - | 1 | 12 | 10 | 6 | 7 | 0 | 15 | 4 | 1 | 19 |
| | | | | | | | | | | | | | | | | | | | - | 0 | 22 | 48 | 45 | 56 | 0 | 50 | 55 | 20 |
| | | | | | | | | | | | | | | | | | | | | - | 3 | 1 | 4 | 0 | 60 | 0 | 1 | 21 |
| | | | | | | | | | | | | | | | | | | | | | - | 7 | 8 | 12 | 1 | 12 | 21 | 22 |
| | | | | | | | | | | | | | | | | | | | | | | - | 60 | 57 | 1 | 46 | 45 | 23 |
| | | | | | | | | | | | | | | | | | | | | | | | - | 50 | 3 | 44 | 46 | 24 |
| | | | | | | | | | | | | | | | | | | | | | | | | - | 0 | 52 | 52 | 25 |
| | | | | | | | | | | | | | | | | | | | | | | | | | - | 0 | 1 | 26 |
| | | | | | | | | | | | | | | | | | | | | | | | | | | - | 49 | 27 |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | - | 28 |

TABLE 4

Frequency Count for Car Data Set

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | Car Number |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| - | 15 | 2 | 56 | 50 | 2 | 10 | 34 | 3 | 50 | 2 | 40 | - | 11 | 14 | 45 | 2 | 3 | 3 | 36 | 25 | 40 | 35 | 25 | 1 |
| | - | 0 | 21 | 14 | 2 | 61 | 39 | 0 | 22 | 0 | 29 | 0 | 59 | 57 | 29 | 1 | 0 | 0 | 32 | 50 | 34 | 35 | 52 | 2 |
| | | - | 3 | 3 | 65 | 0 | 1 | 67 | 2 | 67 | 2 | 67 | 2 | 0 | 1 | 68 | 66 | 69 | 2 | 0 | 3 | 3 | 0 | 2 |
| | | | - | 46 | 3 | 24 | 22 | 3 | 58 | 2 | 40 | 0 | 23 | 24 | 31 | 3 | 3 | 3 | 38 | 19 | 34 | 49 | 27 | 4 |
| | | | | - | 4 | 21 | 37 | 1 | 44 | 0 | 48 | 1 | 20 | 25 | 47 | 5 | 1 | 1 | 42 | 26 | 40 | 35 | 26 | 5 |
| | | | | | - | 2 | 1 | 64 | 2 | 66 | 3 | 68 | 5 | 2 | 2 | 69 | 65 | 66 | 1 | 0 | 2 | 2 | 0 | 6 |
| | | | | | | - | 36 | 0 | 27 | 0 | 30 | 0 | 66 | 66 | 28 | 1 | 0 | 0 | 35 | 43 | 27 | 32 | 45 | 7 |
| | | | | | | | - | 1 | 15 | 0 | 38 | 0 | 37 | 42 | 58 | 1 | 1 | 0 | 43 | 53 | 41 | 12 | 43 | 8 |
| | | | | | | | | - | 4 | 70 | 2 | 68 | 0 | 1 | 1 | 65 | 71 | 70 | 1 | 0 | 1 | 5 | 0 | 9 |
| | | | | | | | | | - | 3 | 42 | 1 | 24 | 29 | 26 | 2 | 4 | 2 | 37 | 16 | 32 | 53 | 28 | 10 |
| | | | | | | | | | | - | 1 | 70 | 1 | 0 | 0 | 65 | 71 | 70 | 0 | 0 | 0 | 3 | 0 | 11 |
| | | | | | | | | | | | - | 2 | 31 | 32 | 44 | 3 | 2 | 0 | 45 | 27 | 40 | 37 | 31 | 12 |
| | | | | | | | | | | | | - | 3 | 0 | 1 | 67 | 69 | 68 | 1 | 0 | 1 | 1 | 0 | 13 |
| | | | | | | | | | | | | | - | 63 | 27 | 4 | 0 | 1 | 32 | 41 | 26 | 33 | 43 | 14 |
| | | | | | | | | | | | | | | - | 32 | 1 | 1 | 0 | 37 | 45 | 31 | 30 | 43 | 15 |
| | | | | | | | | | | | | | | | - | 1 | 1 | 0 | 55 | 45 | 45 | 15 | 35 | 16 |
| | | | | | | | | | | | | | | | | - | 64 | 67 | 1 | 0 | 2 | 3 | 0 | 17 |
| | | | | | | | | | | | | | | | | | - | 69 | 1 | 0 | 1 | 4 | 0 | 18 |
| | | | | | | | | | | | | | | | | | | - | 0 | 0 | 1 | 3 | 0 | 19 |
| | | | | | | | | | | | | | | | | | | | - | 48 | 50 | 30 | 46 | 20 |
| | | | | | | | | | | | | | | | | | | | | - | 52 | 25 | 58 | 21 |
| | | | | | | | | | | | | | | | | | | | | | - | 31 | 52 | 22 |
| | | | | | | | | | | | | | | | | | | | | | | - | 35 | 23 |
| | | | | | | | | | | | | | | | | | | | | | | | 0 | 24 |

For City data --

A Akron, Albany, Canton, Cleveland, Grand Rapids, Kansas City, Columbus, Dayton, Flint, Rochester, Youngstown, Indianapolis, New Haven

B Atlanta, Greensboro, Houston, Richmond

C Birmingham, Chattanooga, Dallas, Ft. Worth, Memphis, Nashville, New Orleans

D Los Angeles, San Diego, San Jose, San Francisco

For Car Data --

A Cad. Eldorado, Dodge Diplomat, Lincoln Cont., Cad. Seville, Dodge St. Regis, Lincoln Versailles, Olds. Toronado, Pont. Grand Prix

B Chev. Malibu, Buick Century, Merc. Zephyr, Olds. Cutlass, Merc. Marquis, Olds 98, Buick Electra, Pont. Catalina

C Datsun 210, Toyota Corolla, Dodge Colt, Honda Civic, Mazda GLC, Subaru, Ford Fiesta, Plym. Champ

Figure 6. Clusters on the basis of frequency of pairs.

A, average = 39.14, $\underline{s}$ = 12.59; for B, average = 46.21, $\underline{s}$ = 11.6).

For the city data one group was most frequently clustered; labeled D, it had an average frequency for pairs of 60.17 ($\underline{s}$ = 4.54). These were all California cities. The other groups had lower averages and more variation. For the Northern and Mid-West cities, A, the average was 48.05 (s = 7.70); the averages for the two groups of Southern cities were, for B, 38.5 ($\underline{s}$ = 13.00) and C, 43.29 ($\underline{s}$ = 8.48).

Since all the pre-experimental work with the data sets had been done using graphic representations, the data sets used and the clusters found based on the frequency analysis were analyzed by several more traditional statistical techniques to observe the structure of the data sets.

A principal components analysis of both data sets was conducted to reduce the dimensionality of the data. Principal components analysis finds the orthogonal linear transformations of the variables which account for most of the variance in the original data. The eigenvectors of the covariance matrix are used for this transformation. The technique is often used to reduce the dimensionality of the data by finding certain principal factors or components which account for most of the variance in the data, though interpretation of the derived components is not always clear. This reduction of dimensionality can sometimes also be used to graphically portray the overall structure of the data. By plotting the scores of the major principal components against each other one may be able to get a two dimensional representation of the distribution of the data. Clusters appear as groups of points when the data are thus projected onto a plane. Using the clusters from the frequency analysis as symbols, the principal components were plotted for both sets of data. For the city data the plots of the first and second principal components and of the first and the third are shown (Figure 7). That for the second and third is similar, but more uniformly distributed.

These plots show that there is not a simple structure in this data set, that the clusters aren't nicely separated or "natural." The group designated with A, the Northern and Mid-West cities, does seem to form a grouping, and is somewhat separated, especially in the second plot. The two groups of Southern cities, B and C, are not well separated. The four cities identified by D, the California cities, appear spread out, apart from the other points. These four points are the ones with which the mathematical clustering methods, discussed later, had trouble; such techniques are sensitive to outliers (Dillon and Goldstein, 1984).

Considering these four points more closely leads to an interesting aspect of performance with polygon type displays (see Figure 8). The frequency analysis indicated that pairs of these four points were most commonly grouped together. Comments made by some of the subjects indicated that they had

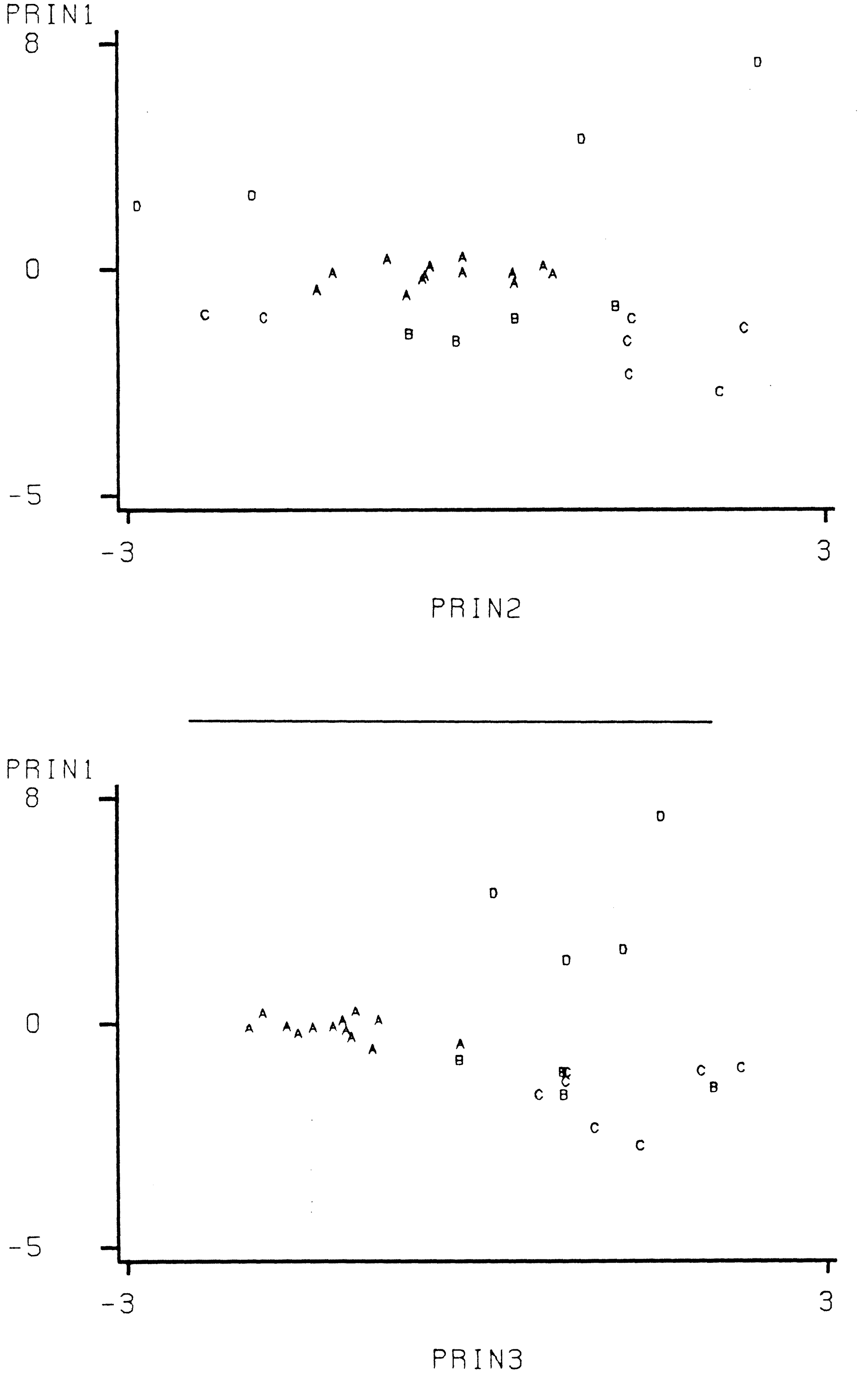



Figure 7. Plot of principal components for city data.

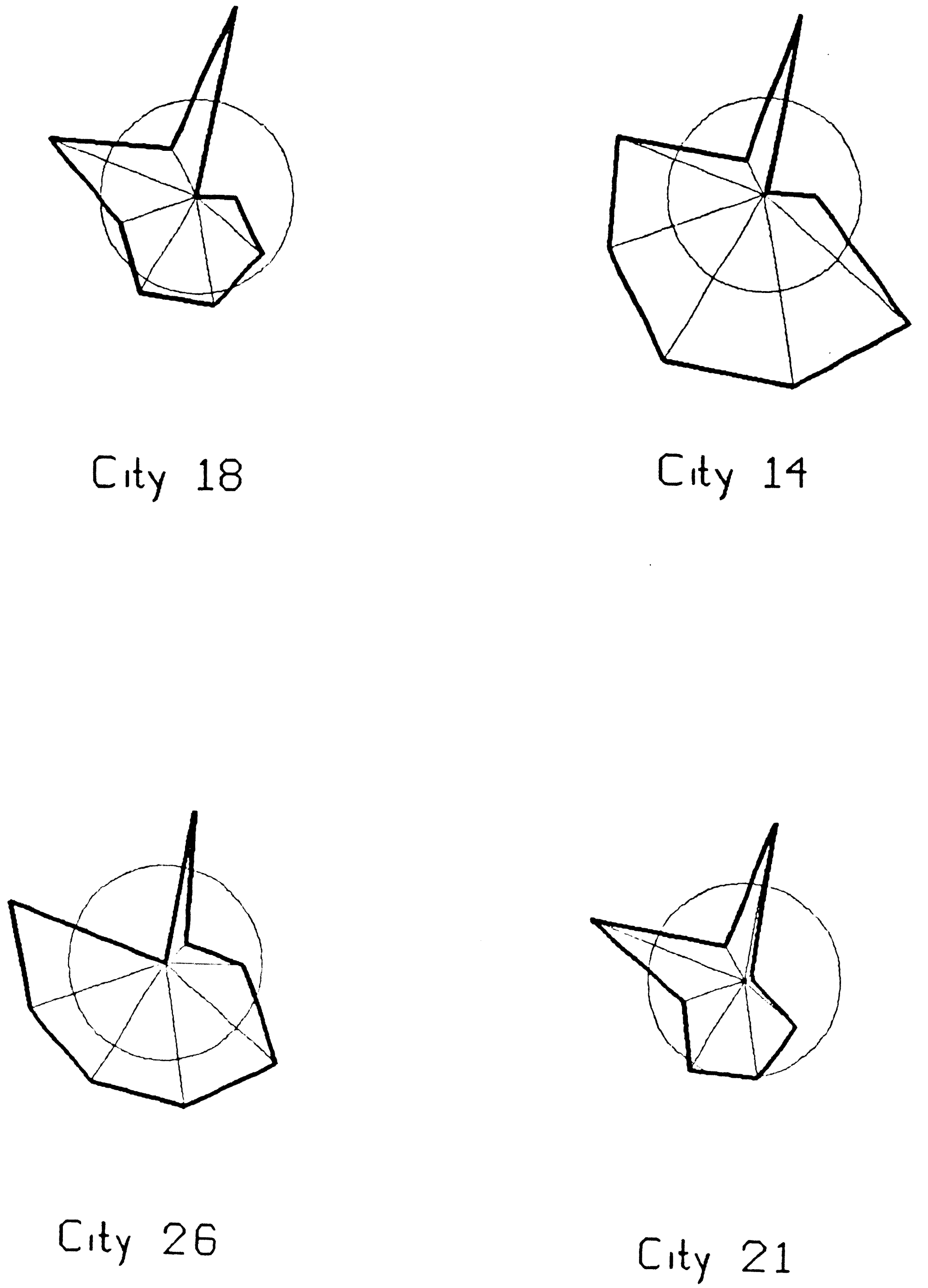



Figure 8. Polygons for four California cities.

given special consideration to the spiked shape at the top of these figures; observation of subjects revealed that these four were often the first grouped. The fact that these cities were grouped together so frequently may be accounted for by considering Tversky's hypothesis of asymmetry in similarity judgments (1977; Tversky and Gati, 1978). People tend to overemphasize similar aspects of stimuli when making similarity judgments, as subjects were here requested to perform. Some subjects indicated in remarks that they had used distinctive aspects to form other groups.

There are two implications of this observation for the use of polygon displays for clustering or for categorization. In effect the subjects weighted certain variables in their clustering. This weighting would be related to the order of presentation of the variables, which produces the distinctive pattern. Reordering the variables might bring different patterns to the fore, and, of course, in exploratory analysis, one would probably not know which variables are likely to produce such patterns. For categorization tasks where the weighting would hinder the accuracy, the display design and ordering of variable presentation should take this phenomenon into account.

On the other hand, observing such patterns in the data is one of the purposes of graphic analysis. In this experiment most subjects picked out the only California cities from the set, something the mathematical algorithms had problems with.

For the car data the principal components plot shows one fairly distinct group (indicated with C, Figure 9). The other data points are fairly spread out and there is overlap in the groups identified by the subjects.

Everitt (1978) proposes plotting the canonical discriminant scores from graphic clustering to observe separation. Canonical discriminant (variate) analysis seeks to find orthogonal linear combinations of the variables of the data for which category membership is already known. The weights for these functions are based on the ratio of the between sums of squares matrix and the within; the eigenvectors of the product of the two matrices are used. The canonical variate scores can then be plotted against each other to produce a two-dimensional representation of the discrimination between the groups. Since mathematical clustering techniques find clusters on the basis of maximizing separation by manipulating these matrices, canonical discriminant analysis can not be used to evaluate such results. With graphic clusterings, though, there may be some rough validity in performing such an evaluation. If subjects were actually grouping random points, the analysis could not find a good transformation and the plot would not show any separation. The plots show some separation, as did the principal components plots (Figure 10).

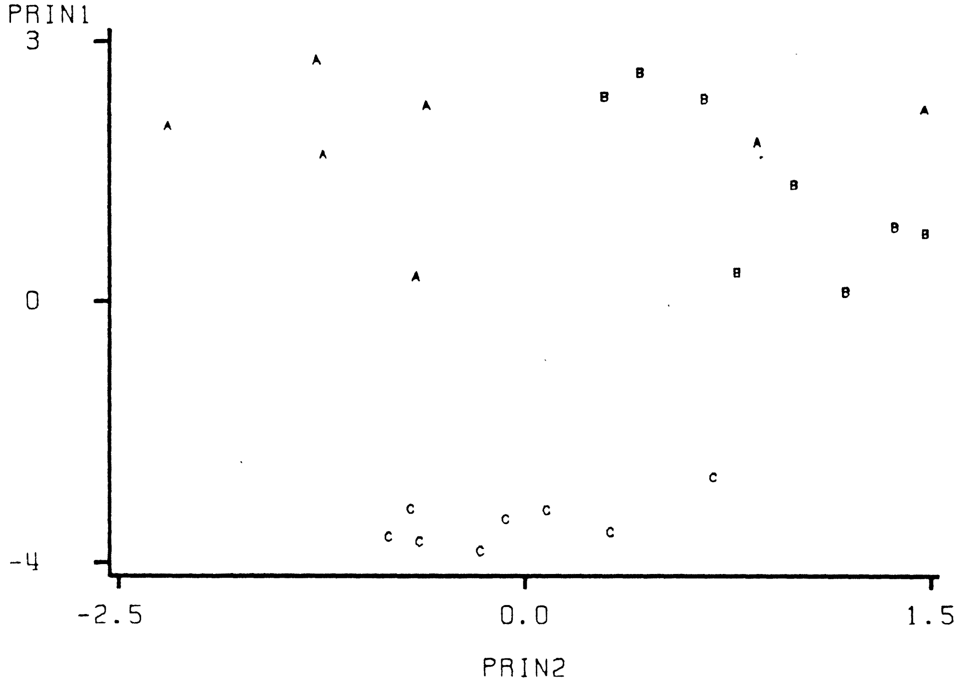



Figure 9. Principal components plot for car data.

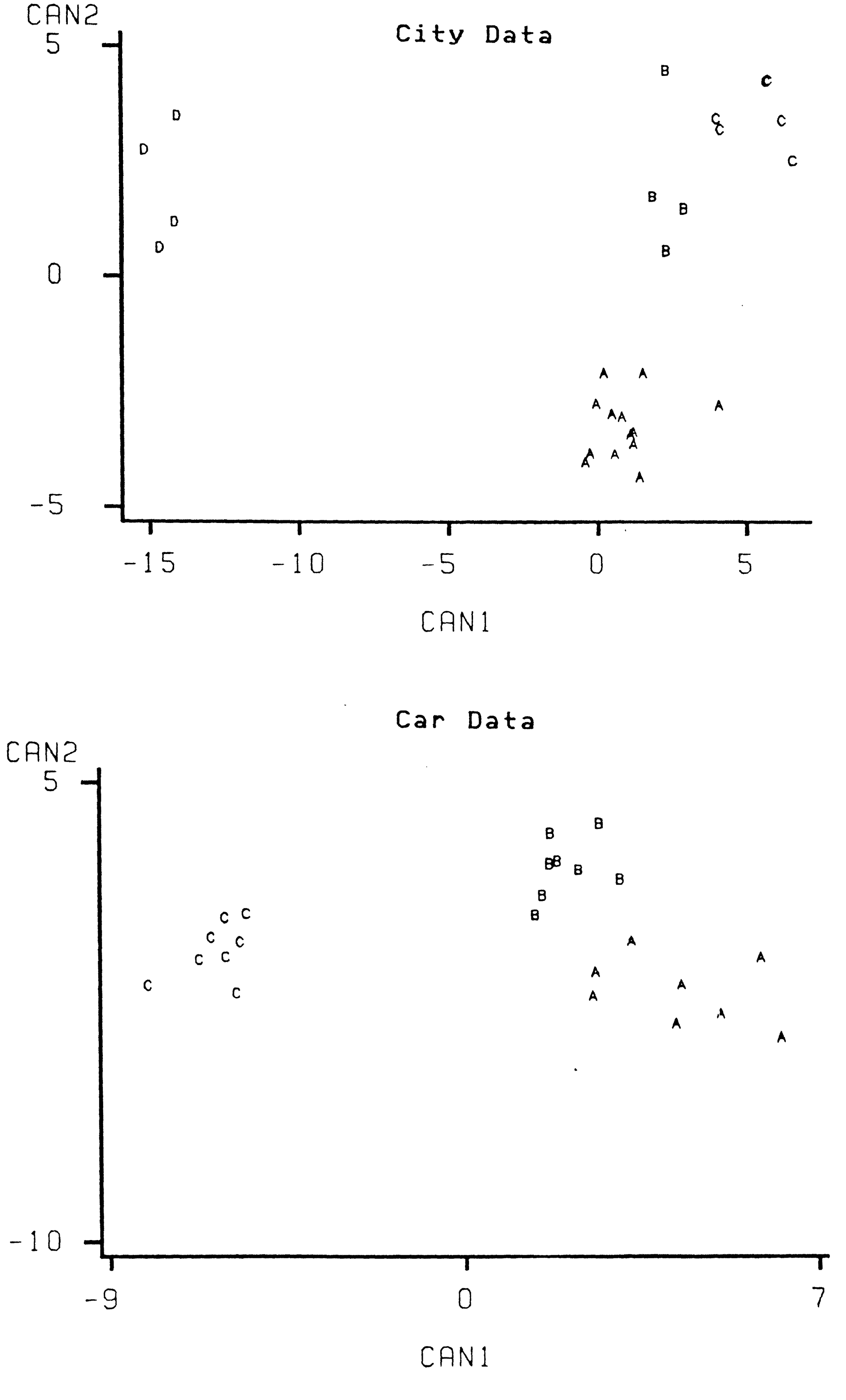



Figure 10. Plots of canonical discriminant scores

The preliminary investigation of frequency confirms that subjects using polygon type displays can generally find groups of similar objects in a data set, clusters if they exist, as some earlier studies which used polygons to study similarity judgments have assumed (Rosch, 1978; Tversky and Gati, 1978). In this study subjects seem to have done well finding fairly distinct groups in the data sets used; as the data became more uniform, the membership of the groups found varies more. Further, subjects showed a fairly good ability to pick out points which were unusual in the data; outlier identification is one of the recommended uses of point representation graphics. One potential problem with such displays was noted, though; subjects did seem to weight their judgments depending on the distinctive characteristics of the representation. This tendency may affect the clusterings found, especially when compared with statistical techniques.

Analysis of Differences

The primary analysis of the results of the experiment was designed to investigate the consistency of the results of the clustering task by using a measure of differences -- first, those from standard cluster analysis techniques; then those between subjects at every level of the variables. The results of each subject's clustering were tabulated and the difference scores were figured by comparing each cluster from a subject's results with the corresponding cluster of the standard or another subject's results. Since the identification of an individual cluster is determined by its elements and not some external attribute, the corresponding clusters were considered to be those with the least number of differences and the total of the cluster differences was used as the difference score for that pair. The difference scores were adjusted to take into account the difference in the size of the two data sets used in the experiment. Difference scores for the car data set were divided by 32; for the city data, by 42. These adjustment factors constitute the maximum possible difference scores determined by comparisons of randomly generated clusters.

For investigating consistency with standard algorithms, it was felt that a comparison should be made with each of the two major types of clustering techniques, hierarchical and partition. The choice of specific techniques was based primarily on their availability and common use. Those in standard statistical packages are the ones most likely to be employed and the ones considered. For a partitioning method, the K-Means technique was selected; it is the basis of the partitioning procedure in BMDP (1983) and the SAS (1985) partitioning procedure, Fastclus, is partially modelled on it. This technique finds clusters by rearranging the membership until the error component, defined as the distance between cluster members and the cluster mean, is minimized (see Dillon and Goldstein, 1984; Hartigan, 1975). As with other partition methods, membership of the clusters is reevaluated when members are shifted. In hierarchical techniques membership is fixed when a point is joined to another or to a cluster. Subjects were observed shifting polygons from one group to another, re-evaluating cluster membership, as a partition technique would do. The partition procedure in BMDP was used since that in SAS is designed for large data sets and is sensitive to the order of presentation with smaller sets. A list of the members of the clusters from this procedure is given in Figure 11.

For the hierarchical clustering technique both SAS and BMDP were used. For the car data, all three of the distance metrics available on SAS gave the same clustering at the three cluster level, and BMDP results were similar (Figure 12). The dendrogram derived from that produced by BMDP shows the clusterings found at each level of joining members or groups; the membership at a given number of clusters can be determined from the branches by running a line across the dendrogram at the appropriate place.

For the city data set Ward's method was selected for use to determine distances (see Figure 13). This data set contained four points which were somewhat unique. Ward's method joins objects on the basis of the least increase in the error sums of squares (see Dillon and Goldstein, 1984; Everitt, 1974). The choice of an appropriate clustering technique is always a matter of some concern (see Everitt, 1974, and brief discussion and examples with different types of data in SAS,

For City data --

A Akron, Albany, Canton, Cleveland, Grand Rapids, Kansas City, Columbus, Dayton, Flint, Rochester, Youngstown, Indianapolis, New Haven

B Atlanta, Birmingham, Chattanooga, Greensboro, Houston, Memphis, Nashville, New Orleans, Richmond

C Dallas, Ft. Worth, San Diego, San Jose

D Los Angeles, San Francisco

For Car Data --

A Cad. Eldorado, Dodge Diplomat, Cad. Seville, Lincoln Versailles, Olds. Toronado, Olds. 98, Buick Electra

B Chev. Malibu, Lincoln Cont., Buick Century, Dodge St. Regis, Merc. Zephyr, Olds. Cutlass, Merc. Marquis, Pont. Catalina, Pont. Grand Prix

C Datsun 210, Toyota Corolla, Dodge Colt, Honda Civic, Mazda GLC, Subaru, Ford Fiesta, Plym. Champ

Figure 11. Clusters on the basis of K-Means technique.

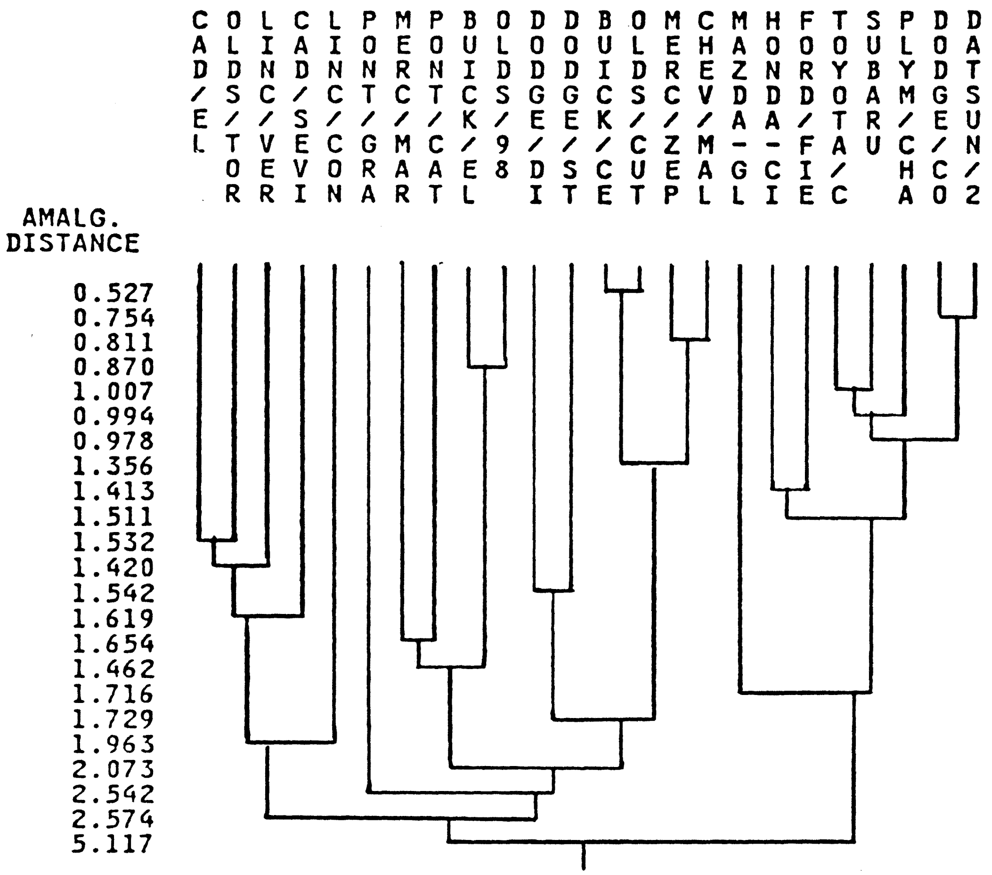



Figure 12. Simplified dendrogram for car data.

1984). Both the centroid and average link measures led to chaining, a predilection to join single points to already existing clusters, and several points remained separate until late in the clustering procedure; at the level of four clusters there were two large clusters and two clusters with one each, California cities. Ward's method tended to give more distinct clusters with these data. Figure 14 contains the clusterings from the hierarchical procedure for both data sets.

As is not uncommon in using cluster analysis techniques in their present state of development, there were different clusterings from the different methods. This inherent inconsistency in the results from these techniques and the problems in assessing the relative appropriateness of the various results would make the use of a particular technique questionable as a criterion for training.

The analysis of differences in this study was not intended to be used as an investigation of which clustering algorithm more closely approximates the way people categorize geometric patterns or the converse, but rather to use commonly available statistical procedures as a standard against which the effect of the visual variables might be compared. Reed (1972) has already investigated the relationship between categorization and certain mathematical procedures, pointing out the general ability of people to abstract a prototype based on some central tendency. There are problems in such

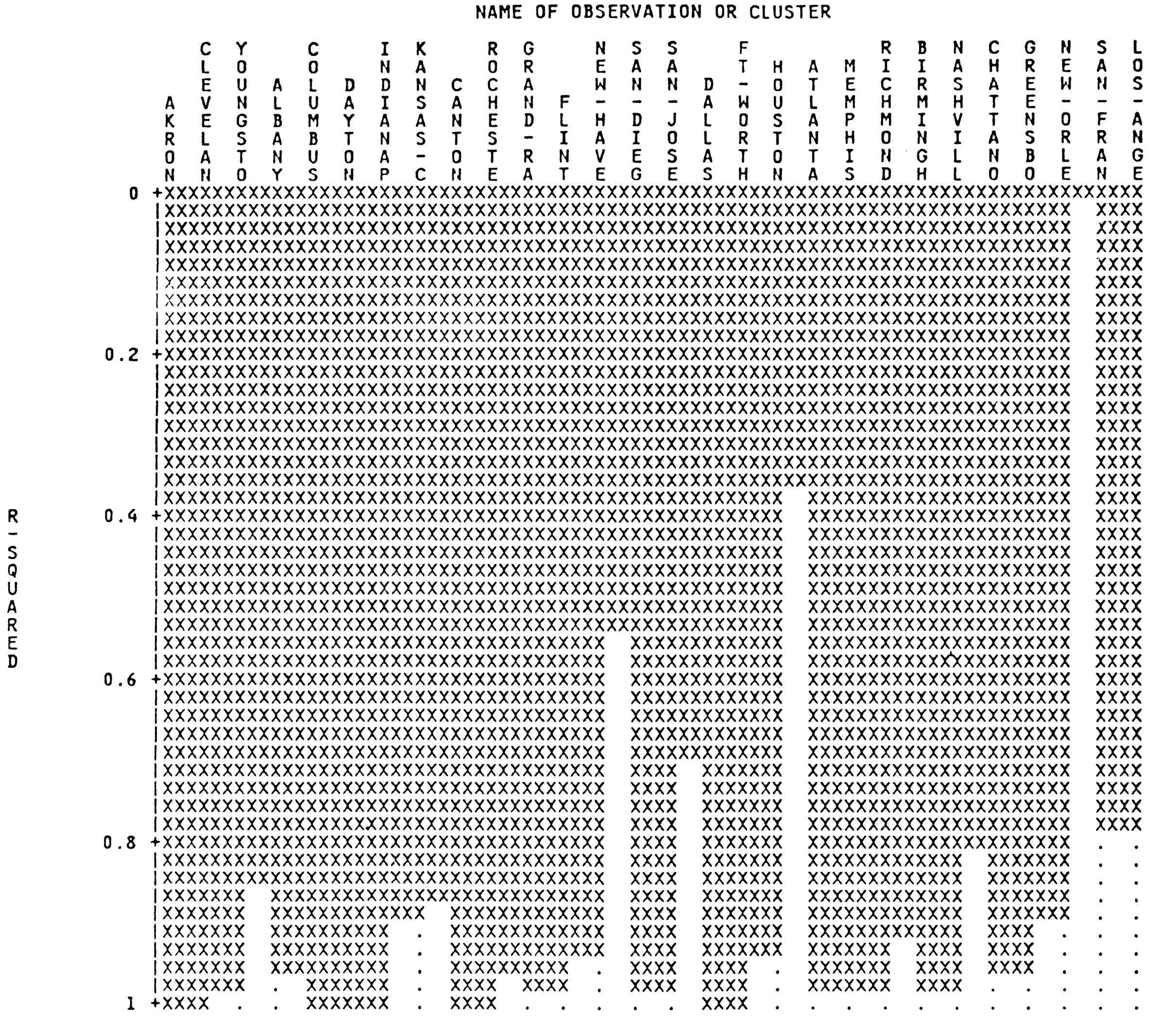



Figure 13. Dendrogram from SAS for city data.

For City data --

A Akron, Albany, Canton, Cleveland, Grand Rapids, Kansas City, Columbus, Dayton, Flint, Rochester, Youngstown, Indianapolis, New Haven

B Atlanta, Birmingham, Chattanooga, Greensboro, Memphis, Nashville, New Orleans, Richmond

C Dallas, Ft. Worth, Houston, San Diego, San Jose

D Los Angeles, San Francisco

For Car Data --

A Cad. Eldorado, Lincoln Continental, Cad. Seville, Lincoln Versailles, Olds. Toronado

B Chev. Malibu, Dodge Diplomat, Buick Century, Dodge St. Regis, Merc. Zephyr, Olds. Cutlass, Merc. Marquis, Olds 98, Buick Electra, Pont. Catalina, Pont. Grand Prix

C Datsun 210, Toyota Corolla, Dodge Colt, Honda Civic, Mazda GLC, Subaru, Ford Fiesta, Plym. Champ

Figure 14. Clusters on the basis of hierarchical technique.

investigations, though; as has recently been pointed out (Rosch, 1978; Tversky, 1977; Tversky and Gati, 1978), judgments of similarity and categorization may not be best modelled by relationships based on geometric distances. Were one to look for a theoretical basis in an experiment such as this, some aspects of the visual patterns themselves probably should be used, as well as purely geometric distance techniques. For example, a measure might include the ratio of overlap to non-overlap area for each pair of figures (Rosch, 1978), along with some overall measure of the angular variability of the figure (Attneave, 1954, 1957).

To investigate the effects of the visual variables, an analysis of variance technique was employed (see Figure 15 for the design of the ANOVA table and error terms). Difference scores were used as the metric. Three analyses were completed, one for the K-Means comparison, one for the hierarchical comparison, and one for the comparison between subjects.

The first analysis was the comparison with the results of the K-Means clusterings. The adjusted difference scores between the subjects' clusterings and those of the K-Means technique for each cell are given in Table 5. These entries are the total in each cell of the difference scores adjusted as described earlier. A summary of the analysis of variance is given in Table 6.

| Source | Deg. of Freedom (n = 6) | F Ratio |
|---|---|---|
| A (add. info.) | 2 | MS(A)/MS(Subj./Gr.) |
| S (shading) | 1 | MS(S)/MS(Subj./Gr.) |
| AS | 2 | MS(AS)/MS(Subj./Gr.) |
| Subj./Groups | 30 | |
| F (form) | 1 | MS(F)/MS(F x Subj./Gr.) |
| D (data) | 1 | MS(D) / MS(D x Subj./Gr.) |
| FD | 1 | MS(FD) / MS(fd x Subj./Gr.) |
| AF | 2 | MS(AF)/MS(F x Subj./Gr.) |
| AD | 2 | MS(AD)/MS(D x Subj./Gr.) |
| SF | 1 | MS(SF)/MS(F x Subj./Gr.) |
| SD | 1 | MS(SD)/MS(D x Subj./Gr.) |
| ASF | 2 | MS(ASF) / MS(F x Subj./Gr.) |
| ASD | 2 | MS(ASD)/MS(D x Subj./Gr.) |
| AFD | 2 | MS(AFD)/MS(FD x Subj./Gr.) |
| SFD | 1 | MS(SFD) / MS(FD x Subj./Gr.) |
| ASFD | 2 | MS(ASFD)/MS(FD x Subj./Gr.) |
| F x Subj./Gr. | 30 | |
| D x Subj./Gr. | 30 | |
| FD x Subj./Gr. | 30 | |

Figure 15. Analysis of Variance Summary

TABLE 5

Adjusted Difference Totals from Comparison with K-Means Algorithm

Additional Information

| | Level 1 (figure alone) | | Level 2 (with radii) | | Level 3 (with grid) | |
|---|---|---|---|---|---|---|
| | Polygon | Star | Polygon | Star | Polygon | Star |
| City Data | | | | | | |
| Shaded | 3.00 | 2.57 | 2.14 | 2.05 | 2.38 | 3.19 |
| Non-shd. | 2.14 | 2.43 | 2.67 | 3.10 | 3.29 | 2.86 |
| Car Data | | | | | | |
| Shaded | 2.44 | 1.94 | 1.44 | 1.63 | 2.75 | 2.25 |
| Non-shd. | 2.38 | 2.13 | 2.25 | 2.44 | 3.25 | 1.94 |

TABLE 6 (Part a)

ANOVA Results from Comparison with K-Means Algorithm

| Source | df | MS | F | p |
|---|---|---|---|---|
| A (add.info.) | 2 | 0.096 | 2.05 | 0.146 |
| S (shading) | 1 | 0.066 | 1.41 | 0.245 |
| AS | 2 | 0.088 | 1.86 | 0.172 |
| Subj./Groups | 30 | 0.047 | | |
| F (form) | 1 | 0.018 | 1.29 | 0.265 |
| D (data) | 1 | 0.173 | 5.60 | 0.025 |
| FD | 1 | 0.053 | 2.82 | 0.104 |
| AF | 2 | 0.026 | 1.84 | 0.177 |
| AD | 2 | 0.005 | 0.16 | 0.856 |
| SF | 1 | 0.002 | 0.16 | 0.695 |
| SD | 1 | 0.004 | 0.14 | 0.710 |
| ASF | 2 | 0.055 | 3.93 | 0.030 |
| ASD | 2 | 0.013 | 0.41 | 0.671 |
| AFD | 2 | 0.028 | 1.47 | 0.246 |
| SFD | 1 | 0.002 | 0.12 | 0.734 |
| ASFD | 2 | 0.006 | 0.31 | 0.733 |
| F x Subj./Gr. | 30 | 0.014 | | |
| D x Subj./Gr. | 30 | 0.031 | | |
| FD x Subj./Gr. | 30 | 0.019 | | |

TABLE 6 (Part b)

Means from Comparison with K-Means Algorithm

| Added Information | |
|---|---|
| Figure alone | 0.40 |
| With radii | 0.37 |
| With grid | 0.46 |

| Shading | |
|---|---|
| Non-shd. | 0.43 |
| Shd. | 0.39 |

| Form | |
|---|---|
| Polygon | 0.42 |
| Star | 0.40 |

| Data Set | |
|---|---|
| Car | 0.37 |
| City | 0.44 |

| Added Infor. | Shading | Form | |
|---|---|---|---|
| Figure alone | Non | Polygon | 0.38 |
| | Non | Star | 0.38 |
| | Shd. | Polygon | 0.45 |
| | Shd. | Star | 0.38 |
| With radii | Non | Polygon | 0.41 |
| | Non | Star | 0.46 |
| | Shd. | Polygon | 0.30 |
| | Shd. | Star | 0.31 |
| With grid | Non | Polygon | 0.54 |
| | Non | Star | 0.40 |
| | Shd. | Polygon | 0.43 |
| | Shd. | Star | 0.45 |

As can be seen from the results only the interaction among the level of background (additional information), shading, and form was significant ($p = 0.030$). Looking at the means we see that shaded stars or polygons were clustered more consistently with the clustering of the K-Means technique at level two of background than any of the other combinations. The two forms were clustered with about the same consistency at this level; simple effects test shows no form effect ($p = 0.49$). At this level shading shows lower means, though not significantly lower ($p = 0.06$). Level two consists of a circle indicating the means and internal radii; polygons with internal radii were found to give better performance in a display integration experiment (Goldsmith and Schvaneveldt, 1982). At level three (grid pattern) there is an interaction between form and shading ($p = 0.02$), with shading appearing to help polygons, but hurt stars.

There is a difference between the data sets ($p = 0.025$). The car data set was clustered more consistently with the results for the partition technique. A simple effects test by form shows that the data set did not effect the clustering of polygons ($p = 0.36$), but did effect stars ($p = 0.01$). Polygons are the more commonly used form of this graphic display.

The adjusted difference results from the comparison with the hierarchical clustering technique are given in Table 7; ANOVA results are in Table 8. Here again the data set shows an effect ($p$ = 0.008); there is also a marginally significant ($p$ = 0.05) interaction between form and data set. The car data set was more consistently clustered with the hierarchical standard, as it was with the K-Means. Simple effects tests are also similar to the K-Means results, showing that the data set effect is not significant for polygons ($p$ = 0.34), but is for stars ($p$ = 0.002).

The interaction of background and form also shows a significant effect ($p$ = 0.015), with the second level (the means and internal radii indicated) again having the least mean differences, here with polygons. The simple effects test shows a background effect ($p$ = 0.030) for polygons, with all levels different on the Student-Newman-Keuls test. The background is not significant for stars ($p$ = 0.66). The interaction of background, shading, and form, though, does not show a significant effect in this analysis.

The results of the comparisons with the clusters found by the two cluster analysis techniques are similar. When one notes that the cluster membership found by the techniques was similar, the similarity of the ANOVA results is understandable. These results show a fairly high number of differences between the subjects' clusterings and those of the statistical techniques; on the car data set the mean difference score was 12 (of 32), and for the city set, 18 (of 42), for the K-Means comparison. Part of these high difference scores may be accounted for by the California cities, mentioned

TABLE 7

Adjusted Difference Totals from Comparison with the Hierarchical Algorithm.

| | Additional Information | | | | | |
|---|---|---|---|---|---|---|
| | Level 1 (figure alone) | | Level 2 (with radii) | | Level 3 (with grid) | |
| | Polygon | Star | Polygon | Star | Polygon | Star |
| <u>City Data</u> | | | | | | |
| Shaded | 2.90 | 2.43 | 2.24 | 2.14 | 2.24 | 3.19 |
| Non-shd. | 2.19 | 2.43 | 2.81 | 3.33 | 3.43 | 2.86 |
| <u>Car Data</u> | | | | | | |
| Shaded | 2.81 | 2.19 | 1.19 | 2.00 | 2.88 | 2.00 |
| Non-shd. | 2.63 | 1.75 | 2.13 | 2.31 | 3.00 | 1.69 |

TABLE 8 (Part a)

ANOVA Results from Comparison with Hierarchical Algorithm

| Source | df | MS | F | p |
| --- | --- | --- | --- | --- |
| A (add.info.) | 2 | 0.052 | 1.17 | 0.324 |
| S (shading) | 1 | 0.038 | 0.86 | 0.362 |
| AS | 2 | 0.099 | 2.23 | 0.126 |
| Subj./Groups | 30 | 0.044 | | |
| F (form) | 1 | 0.031 | 2.12 | 0.156 |
| D (data) | 1 | 0.220 | 8.15 | 0.008 |
| FD | 1 | 0.074 | 4.27 | 0.048 |
| AF | 2 | 0.071 | 4.85 | 0.015 |
| AD | 2 | 0.030 | 1.08 | 0.351 |
| SF | 1 | 0.016 | 1.07 | 0.309 |
| SD | 1 | 0.015 | 0.55 | 0.462 |
| ASF | 2 | 0.035 | 2.35 | 0.113 |
| ASD | 2 | 0.007 | 0.25 | 0.782 |
| AFD | 2 | 0.052 | 3.00 | 0.065 |
| SFD | 1 | 0.009 | 0.51 | 0.483 |
| ASFD | 2 | 0.039 | 1.95 | 0.160 |
| F x Subj./Gr. | 30 | 0.015 | | |
| D x Subj./Gr. | 30 | 0.027 | | |
| FD x Subj./Gr. | 30 | 0.017 | | |

TABLE 8 (Part b)

Means from Comparison with Hierarchical Algorithm

| Added Information | |
|---|---|
| Figure alone | 0.40 |
| With radii | 0.38 |
| With grid | 0.44 |

| Shading | |
|---|---|
| Non-shd. | 0.42 |
| Shd. | 0.39 |

| Form | |
|---|---|
| Polygon | 0.42 |
| Star | 0.39 |

| Data Set | |
|---|---|
| Car | 0.37 |
| City | 0.45 |

| Added Information | Form | |
|---|---|---|
| Figure alone | Polygon | 0.44 |
| | Star | 0.37 |
| With radii | Polygon | 0.35 |
| | Star | 0.41 |
| With grid | Polygon | 0.48 |
| | Star | 0.41 |

| Form | Data Set | |
|---|---|---|
| Polygon | Car | 0.41 |
| | City | 0.44 |
| Star | Car | 0.33 |
| | City | 0.46 |

earlier. Overall, though, the subjects were not consistent with the statistical techniques.

From these two analyses it appears that the variability in performance noted in the frequency analysis overshadows the visual variables effects. It would appear that the subjects' inconsistency in making judgments of similarity between figures combined with differences between subjects' judgments is stronger than the changes in the visual variables used in this study. The interaction effects of the varibles are also inconsistent. The interaction of form and shading when against a brackground grid, but the absence of the interaction when figures have the radii and means indicated, for example, is difficult to explain. The effect of data sets is troublesome when considering the display for exploratory analysis.

It is also interesting to note that the visually more complex form, the stars, with its higher angular variability, is not significantly different from polygons in comparison to the standard algorithm clustering. While form shows no main effect, stars do seem to be affected by differences in data sets more than polygons.

The third analysis involves making a pairwise comparison of the cluster membership in each cell of the design to assess the consistency with which subjects clustered the data at each level of the variables. This analysis allows a comparison of the effects of the variables without resort to some outside standard.

Difference scores for each pair of clusterings in each cell of the design were computed. The adjusted totals for the cells are given in Table 9 and the ANOVA results in Table 10.

In this analysis the data set again shows a significant effect ($p$ = 0.0001), with the car data being more consistently clustered. Again the means show a high degree of variability on both data sets. A simple effects test indicates that the data set effect is significant for both forms, though more so for stars ($p$ = 0.0001; for polygons, $p$ = 0.04). Form and data set show an interaction effect, with stars having the highest and lowest means on different data sets. The interaction of form, data set, and background is also significant. Data sets show an effect at both the second and third levels of background in simple effects tests ($p$ = 0.0004 and $p$ = 0.0001, respectively).

The background shows a significant effect in this analysis, the lowest differences at the simplest level. The level of background is significant for stars and for non-shaded figures in test for effects at those levels of the variables ($p$ = 0.01 for both). The interaction of background and the other visual variables and data sets is also significant. From the means it can be seen that, for stars, the performance with the two simpler background levels are closer

TABLE 9

Adjusted Difference Totals from Pair Comparisons

| | Additional Information | | | | | |
|---|---|---|---|---|---|---|
| | Level 1 (figure alone) | | Level 2 (with radii) | | level 3 (with grid) | |
| | Polygon | Star | Polygon | Star | Polygon | Star |
| City Data | | | | | | |
| Shaded | 6.52 | 6.95 | 6.86 | 6.29 | 5.38 | 9.10 |
| Non-shd. | 4.86 | 7.62 | 8.05 | 8.67 | 8.14 | 9.05 |
| Car Data | | | | | | |
| Shaded | 6.06 | 4.69 | 5.19 | 5.38 | 6.81 | 4.50 |
| Non-shd. | 5.63 | 4.63 | 6.00 | 4.69 | 5.94 | 5.81 |

TABLE 10 (Part a)

ANOVA Results from Pair Comparisons

| Source | df | MS | F | p |
|---|---|---|---|---|
| A (add.info.) | 2 | 0.144 | 3.46 | 0.036 |
| S (shading) | 1 | 0.104 | 2.50 | 0.118 |
| AS | 2 | 0.061 | 1.47 | 0.237 |
| Comb./Groups | 84 | 0.042 | | |
| F (form) | 1 | 0.007 | 0.33 | 0.568 |
| D (data) | 1 | 1.290 | 52.80 | 0.0001 |
| FD | 1 | 0.492 | 25.92 | 0.0001 |
| AF | 2 | 0.026 | 1.27 | 0.286 |
| AD | 2 | 0.028 | 1.14 | 0.325 |
| SF | 1 | 0.007 | 0.35 | 0.556 |
| SD | 1 | 0.100 | 4.07 | 0.047 |
| ASF | 2 | 0.038 | 1.60 | 0.208 |
| ASD | 2 | 0.059 | 2.19 | 0.118 |
| AFD | 2 | 0.093 | 4.92 | 0.010 |
| SFD | 1 | 0.001 | 0.06 | 0.805 |
| ASFD | 2 | 0.144 | 7.56 | 0.001 |
| F x Comb./Gr. | 84 | 0.020 | | |
| D x Comb./Gr. | 84 | 0.024 | | |
| FD x Comb./Gr. | 84 | 0.019 | | |

TABLE 10 (Part b)

Means from Pair Comparisons

| Added Information | | Shading | |
|---|---|---|---|
| Figure alone | 0.39 | Non-shd. | 0.44 |
| With radii | 0.41 | Shd. | 0.40 |
| With Grid | 0.46 | | |

| Form | | Data Set | |
|---|---|---|---|
| Polygon | 0.41 | Car | 0.36 |
| Star | 0.42 | City | 0.48 |

| Added Inform. | Shading | Form | Data | |
|---|---|---|---|---|
| Figure alone | Non. | Poly. | Car | 0.36 |
| | Non. | Poly. | City | 0.33 |
| | Non. | Star | Car | 0.31 |
| | Non. | Star | City | 0.51 |
| | Shd. | Poly. | Car | 0.40 |
| | Shd. | Poly. | City | 0.43 |
| | Shd. | Star | Car | 0.30 |
| | Shd. | Star | City | 0.46 |
| With radii | Non. | Poly. | Car | 0.39 |
| | Non. | Poly. | City | 0.54 |
| | Non. | Star | Car | 0.30 |
| | Non. | Star | City | 0.56 |
| | Shd. | Poly. | Car | 0.34 |
| | Shd. | Poly. | City | 0.43 |
| | Shd. | Star | Car | 0.36 |
| | Shd. | Star | City | 0.39 |
| With grid | Non. | Poly. | Car | 0.40 |
| | Non. | Poly. | City | 0.54 |
| | Non. | Star | Car | 0.39 |
| | Non. | Star | City | 0.60 |
| | Shd. | Poly. | Car | 0.45 |
| | Shd. | Poly. | City | 0.36 |
| | Shd. | Star | Car | 0.30 |
| | Shd. | Star | City | 0.61 |

TABLE 10 (Part c)

Means from Pair Comparisons

| Added Infor. | Data Set | Form | |
|---|---|---|---|
| Figure alone | Car | Polygon | 0.39 |
| | Car | Star | 0.30 |
| | City | Polygon | 0.38 |
| | City | Star | 0.48 |
| With radii | Car | Polygon | 0.37 |
| | Car | Star | 0.33 |
| | City | Polygon | 0.48 |
| | City | Star | 0.47 |
| With grid | Car | Polygon | 0.43 |
| | Car | Star | 0.34 |
| | City | Polygon | 0.45 |
| | City | Star | 0.60 |

| Form | Data Set | |
|---|---|---|
| Polygon | Car | 0.39 |
| | City | 0.44 |
| Star | Car | 0.33 |
| | City | 0.52 |

| Shading | Data Set | |
|---|---|---|
| Non-shd. | Car | 0.36 |
| | City | 0.51 |
| Shaded | Car | 0.36 |
| | City | 0.44 |

to each other than to that of the grid pattern background. The fact that the internal radii and the circle through the means provide a center for polygons, but that stars are visually centered by the nature of their form, may explain this similarity of performance. As in the previous analyses, the grid pattern generally shows higher difference scores.

At the intermediate level of background information shading shows lower mean differences on both forms, as it does in the analyses of the comparison with the clustering algoritms. Here, it also shows lower differences at the most complex level. The effect of shading is not statistically significant at either level, though ($\underline{p}$ = 0.15 and $\underline{p}$ = 0.14, respectively). While showing generally lower means, shading does not show a significant overall effect.

In general, the analysis of consistency between subjects shows again the variance in the clusterings which were done. The mean difference scores are again high. Clustering consistency seems to be again related to the data set. Use of a lower level of added background information and of shading may be indicated, though the interaction effects make generalization difficult. The subjects' ability to perceive patterns and to compare one to others and the variation between subjects' judgments seem to override most of the effects of the variables.

# CHAPTER FOUR -- CONCLUSIONS

The results of this investigation have several implications for polygon displays when used for tasks involving clustering or categorization. Although the results of the analyses are somewhat inconsistent, there are a number of points which came to light in this exploratory experiment.

In designing or using polygon type graphic displays, we should keep in mind that judgments of similarity may be weighted toward distinctive, similar features of the figures. Because of this tendency, decisions of membership in a category may be made on the basis of a subset of the variables rather than the overall figure. In situations where the task might be adversely affected, such as in status displays or identification tasks, the display should be so designed as to minimize this effect. If the possible range and relation of values are known, the variables could be so ordered as to lessen the appearance of distinctive patterns, which might distract from the overall shape. While training of those using the display may be able to compensate, if this tendency is based in a perceptual process, it may not be amenable or even desirable to try to train individuals to avoid it.

In exploratory analysis this tendency may prove both helpful and detrimental. It will allow perceptions of patterns in the data which other methods of presentation may overlook, or find with more difficulty. On the other hand, such weighting of variables may hinder the observer from perceiving less distinctive but important patterns in the other variables. It would be interesting to ascertain the interplay between the distinctiveness of the graphic pattern and the relationships between the data points which these patterns reflect. In the city data set used in this study, for example, it was the distinctive shape of the polygons which caused the grouping of four cities generally separated by the standard clustering methods. There was some natural reason for these cities to be grouped. The graphic presentation allowed one to observe a relationship in the data which did exist and which might have been overlooked using other clustering methods. Further study could indicate with which types of data or under what circumstances graphic presentation might yield insights and where it might be misleading.

When using polygons for cluster identification, the data have an important effect. Although the form variable in this experiment did not show an overall strong effect, there was some indication that the polygon form was less affected by differences in the data sets. From the frequency analysis it was noted that subjects seem to group objects which are members of distinct clusters with fair regularity; the groups that overlap are less consistently clustered. Even if one considers that these are not actually natural clusters, not coming from different populations, subjects tended to group those data points which were similar in the distribution.

The question of how well such graphic displays work for initial determinations of the numbers of clusters needs further investigation. At any rate, the graphic seems to work best where the clusters are fairly well separated.

The variability of subjects from a general population sample seems to have been a problem in this study. To all of the subjects this type of display was new. While they seemed to understand the task and the explanation of how the display could be used, none had actually done such a task before. Many of the other studies have used subjects who have probably had more experience in data analysis than did the present subjects (e.g., Freni-Titulaer and Louv, 1984; Wilkinson, 1982); Woods, Wise, and Hanes (1982) commented on the unfamiliarity of their subjects with the polygon display for safety parameters and its effect. Some exposure to the display and the task would probably bring the variability down. Training for consistency is feasible, involving repeated clusterings of similar data sets, with feedback on differences from earlier groupings.

For display design there is some indication, at least in relation to the consistency with clustering algorithms results, that internal radii and an indication of means are helpful for displays using the usual polygons, and that shading may be beneficial. That the added information should prove helpful has an intuitive explanation as well. The internal radii give the figure a center, providing a basis to judge relative lengths or proportions of area from figure to figure. This basis may help to explain the relatively good performance of the star form, especially in the pair comparison. Cleveland and McGill (1984) hypothesized that judgments of length are higher in level of graphical perceptual skill than of area, and maybe of overall pattern as well.

Finally one should recognize the ability of subjects to discern the pattern of the figure. More study is needed on the intra-subject consistency in the use of polygon displays, but it would appear that to a certain extent the perception of patterns overshadowed the effects of the visual variables in this experiment. It would be worthwhile to investigate whether these variables affect tasks other than categorization, such as information integration.

Polygons displays do have their advantages. They are relatively simple to produce, are abstract in form, thus unburdened by subjective perceptions, and seem to be understandable to a wide population.

Appendix A. Instructions

## INSTRUCTIONS

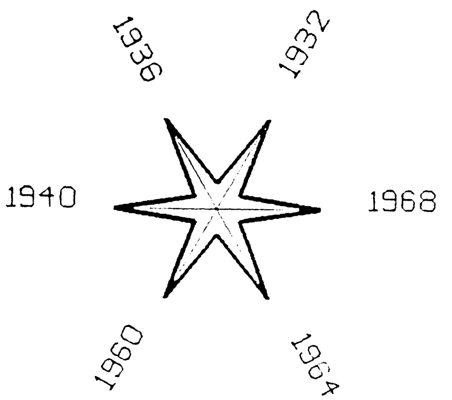
The experiment in which you are participating is intended to investigate the effect of certain visual characteristics of a graphic display on the ability of people to use that display. It is hoped that by identifying the optimal characteristics for this display, designers can create more efficient and effective displays.

This type of graphic display is used to portray complex data in a simplified form. Suppose, for example, we take the percentage of Republican votes in six Presidential elections for nine States. Such data might look like this:

Per Cent Republican Voting in Presidential Elections

| *1932* | *1936* | *1940* | *1960* | *1964* | *1968* | | *State* |
|---|---|---|---|---|---|---|---|
| *14* | *13* | *14* | *42* | *69* | *19* | *1* | *Alabama* |
| *25* | *24* | *26* | *51* | *49* | *41* | *2* | *Florida* |
| *8* | *13* | *8* | *37* | *54* | *30* | *3* | *Georgia* |
| *40* | *40* | *42* | *54* | *36* | *44* | *4* | *Kentucky* |
| *7* | *11* | *14* | *29* | *57* | *23* | *5* | *Louisiana* |
| *36* | *37* | *41* | *46* | *35* | *42* | *6* | *Maryland* |
| *35* | *38* | *48* | *50* | *36* | *45* | *7* | *Missouri* |
| *29* | *27* | *26* | *48* | *44* | *40* | *8* | *North Carolina* |
| *2* | *1* | *4* | *49* | *59* | *39* | *9* | *South Carolina* |
| *33* | *31* | *32* | *53* | *44* | *38* | *10* | *Tennessee* |
| *30* | *29* | *31* | *53* | *46* | *43* | *11* | *Virginia* |

Each data value for an individual State may be plotted as a point along one of six lines radiating from a center, one line for each year. These points can then be connected to form a multi-sided figure, or star. A high per cent in a particular year for one State would be farther from the center than a lower per cent for another State. The shape of the figure for each State is determined by the values in the data. On the next page is a sample star, drawn for the <u>average</u> values for the election data.

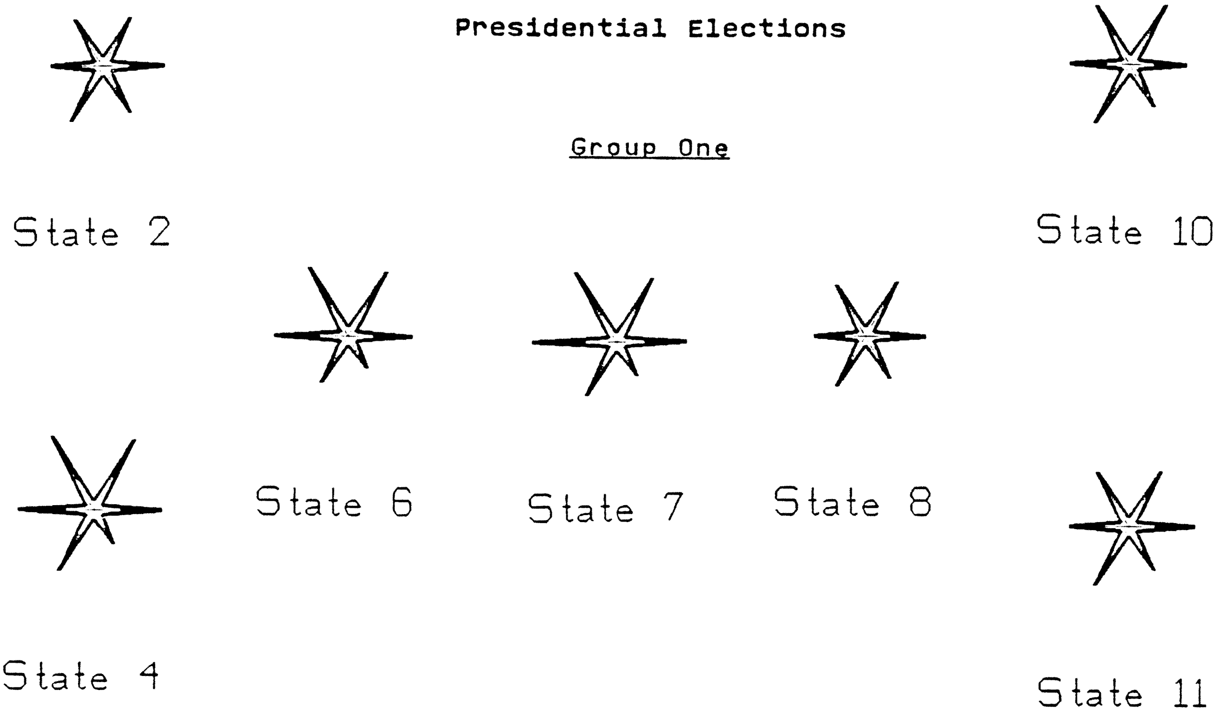



State Number

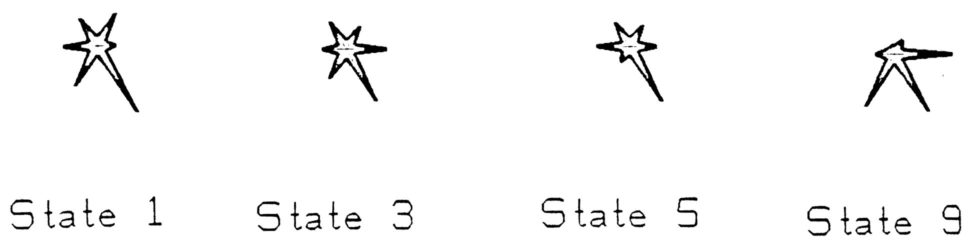
When figures are plotted for each of the States in the preceding example, these figures can be used to divide the States into groups or clusters of States with similar voting records, based on the shape of the stars. The States appear to fall into two groups, as can be seen on the next page.

## Presidential Elections

### Group One




### Group Two



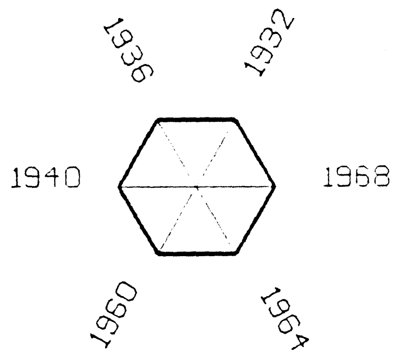
We can examine the groups to see the similarity between the members. Notice that one group of States shows a relatively consistent voting record over the years in this example. The other group of States shows a distinct change in voting pattern, especially in the 1964 election; these are generally deep South States.

In a similar manner figures can be drawn from the measurements of various attributes for any set of objects. Such graphic analysis can be useful when dealing with large amounts of data, to cluster or group the objects on the basis of their similarity. One can then examine a smaller number of groups more easily than the numerical data.

If you have any questions concerning this graphic display, please feel free to ask the experimenter.

## Procedure

You will be presented with four different sets of figures, one set at a time. The data for some of these sets consist of measurements on nine different attributes of various cities. These variables include levels of pollutants (hydrocarbons, oxides of nitrogen, and sulfur dioxide), mean temperature for January and July, precipitation, mean level of education, population density, and mortality rate, though not in this order. The other sets consist of measurements on nine attributes of various car models, including mileage, trunk space, length, repair record, rear seat room, price, weight, and displacement. Two of the sets will have stars drawn as in the preceding example; the figures in the other sets will be drawn by connecting the data points directly, forming a polygon. As with the stars, the data is represented by the distance from the center. Below is a sample of the polygon for the average values in the election example.




State Number
Figure indicates average value

There may be a different background on the sets; a legend with each set will explain the background.

Your task will be to divide each set into a specific number of groups; the legend will tell you how many groups. Please fit all of the figures into one or another of the groups.

- Divide the figures on the basis of the similarity of their overall shape, considering the whole figure.
- There may be different numbers of figures in the different groups.
- Note that the figures are in a fixed position; don't rotate them. A figure which is elongated to the right is different from one that is elongated to the left.
- Begin by laying all the figures out in order in three rows; then proceed to divide them into groups of similar figures.
- Please work as conscientiously and as accurately as possible, but don't take excessive time. Fifteen or twenty minutes should be sufficient to complete each set.
- When you have finished with a set, leave the piles of cards on the table; the experimenter will collect them and give you the next set.
- This is not a test of your skill. There is no one correct set of groupings. We are interested in the consistency with which different people group the graphs.

Please feel free to ask any questions you may have about the task. The experimenter will try to answer them satisfactorily. The decisions as to the groupings, or similarity of particular figures, will have to be your own.

Once again thank you for your participation.

# REFERENCES

Anderson, J. R. (1980). Cognitive psychology and its implications. San Francisco: Freeman.

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

Attneave, F. (1957). Physical determinants of the judged complexity of shapes. Journal of Experimental Psychology, 53, 221-227.

Beniger, J. R. and Robyn, D. L. (1976). The history and future of graphics in statistics. Proceedings of the Social Statistics Section, American Statistical Association.

Bertin, J. (1981). Graphics and graphic information-processing. Translated by W.J.Berg and P.Scott. Berlin: DeGruyter.

Bertin, J. (1983). Semiology of graphics; diagrams, networks, maps. Translated by W.J.Berg. Madison, Wisc.: University of Wisconsin.

BMDP statistical software. (1983). Berkeley: University of California.

Bruckner, L. A. (1978). On Chernoff faces. In P. C. C. Wang (Ed.), Graphical representation of multivariate data (pp.93-121). New York: Aacademic.

Carswell, C. M. and Wickens, C. D. (1984). Stimulus integrality in displays of system input-output relationships: a failure detection study. In Proceedings of the Human Factors Society 28th Annual Meeting (pp. 534-537). Santa Monica, Calif.: Human Factors Society.

Carter, L. F. (1947a). An experiment on the design of tables and graphs used for presenting numerical data. Journal of Applied Psychology, 31, 640-650.

Carter, L. F. (1947b). The relative effectiveness of presenting numerical data by the use of tables and graphs. In P. M. Fitts (Ed.), Psychological research on equipment design, Army Air Forces, Aviation Psychology Program (Research Report 19) (pp. 65-72). Washington, D.C.: GPO.

Chambers, J. M., Cleveland, W. S., Kleiner, B., and Tukey, P. A. (1983). Graphical methods for data analysis . Belmont, Calif.: Wadsworth International.

Chang, K. (1977). Visual estimation of graduated circles. Canadian Cartographer, 14, 130-138.

Chernoff, H. (1972). The use of faces to represent points in n-dimensional space graphically. In Computer Science and Statistics. Sixth Annual Symposium on the Interface, October 16-17, 1972 (pp. 43-44).

Chernoff, H. (1978). Graphical representation as a discipline. In P. C. C. Wang (Ed.), Graphical representation of multivariate data (pp. 1-11). New York: Academic.

Chernoff, H. and Rizvi, M. H. (1975). Effect on classification error of random permutations of features in representing multivariate data by faces. Journal of the American Statistical Association, 70, 548-554.

Cleveland, W. S. (1984a). Graphical methods for data presentation: full scale breaks, dot charts, and multibased logging. The American Statistician, 38, 270-280.

Cleveland, W. S. (1984b). Graphs in scientific publications. The American Statistician, 38, 261-269.

Cleveland, W. S., Harris, C. S., and McGill, R. (1982). Judgments of circle sizes on statistical maps. Journal of the American Statistical Association, 77, 541-547.

Cleveland, W. S., Harris, C. S., and McGill, R. (1983). Experiments on quantitative judgments of graphs and maps. The Bell System Technical Journal, 62, 1659-1674.

Cleveland, W. S. and McGill, R. (1983). A color-caused optical illusion on a statistical graph. The American Statistician, 37, 101-105.

Cleveland, W. S. and McGill, R. (1984). Graphical perception: theory, experimentation, and application to the development of graphical methods. Journal of the American Statistical Association, 79, 531-554.

Cox, C. W. (1976). Anchor effects and the estimation of graduated circles and squares. American Cartographer, 3, 65-74.

Croxton, F. E. (1927). Further studies in the graphic use of use of circles and bars. II. Some additional data. Journal of the American Statistical Association, 22, 36-39.

Croxton, F. E. and Stein, H. (1932). Graphic comparisons by bars, squares, circles, and cubes. Journal of the American Statistical Association, 27, 54-60.

Croxton, F. E. and Stryker, R. E. (1927). Bar charts versus circle diagrams. Journal of the American Statistical Association, 22, 473-482.

Culbertson, H. M. and Powers, R. D. (1959). A study of graph comprehension difficulties. A V Communication Review, 7, 97-110.

Dillon, W. R. and Goldstein, M. (1984). Multivariate analysis; methods and applications, New York: Wiley.

Dobson, M. W. (1980). The acquisition and processing of cartographic information: some preliminary experimentation. In P. A. Kolers, M. E. Wrolstad, and H. Bouma (Eds.), Processing of visible language 2 (pp. 291-304). New York: Plenum.

Eells, W. C. (1926). The relative merits of circles and bars for representing component parts. Journal of the American Statistical Association, 21, 119-132.

Egeth, H., Jacob, R. J. K., Wainer, H., Kleiner, B., and Hartigan, J. A. (1981). Comments and rejoinder. Journal of the American Statistical Association, 76, 269-276.

Eggen, P., Kauchak, D., and Kirk, S. (1978). The effects of generalizations as cues on the learning of information from graphs. Journal of Educational Research, 71, 211-213.

Everitt, B. (1974). Cluster analysis. London: Heinemann.

Everitt, B. S. (1978). Graphical techniques for multivariate data. New York: North-Holland.

Feinberg, S. E. (1979). Graphical methods in statistics. The American Statistician, 33, 165-178.

Feliciano, G. D., Powers, R. D., and Kearl, B. E. (1963). The presentation of statistical information. A V Communication Review, 11, 32-39.

Flannery, J. J. (1971). The relative effectiveness of some common graduated point symbols in the presentation of quantitative data. Canadian Cartographer, 8, 96-109.

Frazelle, E. (1985). Suggested techniques enable multi-criteria evaluation of material handling alternatives. Industrial Engineering, 17 (2), 42-48.

Freni-Titulaer, L. W. J. and Louv, W. C. (1984). Comparisons of some graphical methods for exploratory multivariate data analysis. The American Statistican, 38, 184-188.

Friedman, H. P., Farrell, E. J., Goldwyn, R. M., Miller, M., and Siegel, J. H. (1972). A graphic way of describing changing multivariate patterns. In Computer Science and Statistics. Sixth Annual Symposium on the Interface, October 16-17, 1972 (pp. 56-59).

Garner, W. R. (1970). The stimulus in information processing. American Psychologist, 25, 350-358.

Garner, W. R. (1978). Aspects of a stimulus: features, dimensions, and configurations. In E. Rosch and B. Lloyd (Eds.), Cognition and categorization (pp. 99-133). Hillsdale, N.J.: Erlbaum.

Ghani, J. and Lusk, E. J. (1982). The impact of a change in information representation and in the amount of information on decision performances. Human Systems Management, 3, 270-278.

Goldsmith, T. E. and Schvaneveldt, R. W. (1982). The role of integral displays in decision making. In _Proceedings. Human Factors in Computer Systems. March 15-17, 1982. Gaithersburg, Maryland_ (pp. 197-201).

Graham, J. L. (1937). Illusory trends in the observation of bar graphs. _Journal of Experimental Psychology_, _20_, 597-608.

Hanson, S. J., Kraut, R. E., and Farber, J. M. (1984). Interface design and multivariate analysis of UNIX command use. _ACM Transactions on Office Information Systems_, _2_, 42-57.

Hartigan, J. A. (1975). _Clustering algorithms_. New York: Wiley.

Huhn, R. von (1927). Further studies in the graphic use of circles and bars I. A discussion of the Eells' experiment. _Journal of the American Statistical Association_, _22_, 31-36.

Izenman, A. J. (1980). Developments in statistical graphics, 1960-1980. In _Proceedings of the first general conference on social graphics, Leesburg, Virginia,_ _October 22-24, 1978_ (Tech. paper 49)(pp. 51-79). Washington, D.C.: Bureau of the Census.

Jacob, R. J. K. (1978). Facial representation of multi-variate data. In P. C. C. Wang (Ed.), _Graphical representation of multivariate data_ (pp. 143-168). New York: Academic.

Jacob, R. J. K., Egeth, H. E., and Bevan, W. (1976). The face as a data display. _Human Factors_, _18_, 189-199.

Kirk, S., Eggen, P., and Kauchak, D. (1978). The effect of cue specificity and locus on learning from graphical material. _Journal of Educational Research_, _72_, 39-44.

Kleiner, B. and Hartigan, J. A. (1981). Representing points in many dimensions by trees and castles. _Journal of the American Statistical Association_,, _76_, 260-269.

Kruskal, W. (1975). Visions of maps and graphs. In _Auto-carto II; Proceedings of the International Symposium on Computer-assisted Cartography, September 21-25, 1975_ (pp. 27-36). Washington, D.C.: Bureau of the Census.

Kruskal, W. (1982). Criteria for judging statistical graphics. Utilitas Mathematica, 21B, 283-310.

MacDonald-Ross, M. (1977). How numbers are shown; a review of research on the presentation of quantitative data in texts. A V Communication Review, 25, 359-409.

McDonald, G. C. and Ayers, J. A. (1978). Some applications of the "Chernoff faces": a technique for graphically representing multivariate data. In P. C. C. Wang (Ed.), Graphical representation of multivariate data (pp. 183-197). New York: Academic.

Meihoefer, H. (1973). The visual perception of the circle in thematic maps. Canadian Cartographer, 10, 63-84.

Mezzich, J. E. and Worthington, D. R. L. (1978). A comparison of graphical representations of multidimensional psychiatric diagnostic data. In P. C. C. Wang (Ed.) Graphical representation of multivariate data (pp. 123-139). New York: Academic.

Naveh-Benjamin, M. and Pachella, R. G. (1982). The effect of complexity on interpreting "Chernoff" faces. Human Factors, 24, 11-18.

Petersen, R. J., Banks, W. W., and Gertman, D. I. (1982). Performance-based evaluation of graphic displays for nuclear power plant control rooms. In Proceedings. Human Factors in Computer Systems. March 15-17, 1982, Gaithersburg, Maryland (pp.182-189).

Peterson, L. V. and Schramm, W. (1954). How accurately are different kinds of graphs read? A V Communication Review, 2, 178-189.

Phillips, R. J. (1979). Making maps easy to read -- a summary of research. In P. A. Kolers, M. E. Wrolstad, and H. Bouma (Eds.), Processing of visible language 1 (pp. 165-174). New York: Plenum.

Pickover, C. A. (1984). The use of computer-drawn faces as an educational aid for the presentation of statistical concepts. Computers & Graphics, 8, 163-166.

Potash, L. M. (1977). Design of maps and map-related research. Human Factors, 19, 139-150.

Powers, M., Lashley, C., Sanchez, P., and Shneiderman, B. (1984). An experimental comparison of tabular and graphic data presentation. International Journal of Man-Machine Studies, 20, 545-566.

Preece, J. (1983). Graphics are not straight forward. In T. Green and S.J. Payne (Eds.), Psychology of computer use (pp. 41-56). New York: Academic.

Reed, S. K. (1972). Pattern recognition and categorization. Cognitive Psychology, 3, 382-407.

Remus, W. (1984). An empirical investigation of the impact of graphical and tabular data presentation on decision making. Management Science, 30, 533-542.

Roller, B. V. (1980). Graph reading abilities of thirteen-year-olds. In P. A. Kolers, M. E. Wrolstad, and H. Bouma (Eds.), Processing of visible language 2 (pp. 305-314). New York: Plenum.

Rosch, E. (1978). Principles of categorization. In E. Rosch and B. Lloyd (Eds.), Cognition and categorization (pp. 27-48). Hillsdale, N.J.: Erlbaum.

SAS Institute (1985). SAS user's guide: statistics. 5 ed. Cary, N.C.: SAS.

Schmid, C. F. (1976). The role of standards in graphic presentation. Proceedings of the Social Statistics Section, American Statistical Association (pp. 74-81).

Schmid, C. F. (1983). Statistical graphics: design principles and practices. New York: Wiley.

Schmid, C. F. and Schmid, S. E. (1979). Handbook of graphic presentation (2nd ed.). New York: Wiley.

Schutz, H. G. (1961a). An evaluation of formats for graphic trend displays -- Experiment II. Human Factors, 3, 99-107.

Schutz, H. G. (1961b). An evaluation of methods for presentation of graphic multiple trends -- Experiment III. Human Factors, , 3, 108-119.

Tufte, E. R. (1983). The visual display of quantitative information. Cheshire, Conn.: Graphics.

Tversky, A. (1977). Features of similarity. Psychological Review, 84, 327-352.

Tversky, A. and Gati, I. (1978). Studies of similarity. In E. Rosch and B. Lloyd (Eds.), Cognition and categorization (pp. 79-98). Hillsdale, N.J.: Erlbaum.

Verhagen, L. H. J. M. (1981). Experiments with bar graph supervision displays on VDU's. _Applied Ergonomics_, _12_, 39-45.

Vernon, M. D. (1946). Learning from graphical material. _British Journal of Psychology_, _36_, 145-158.

Vernon, M. D. (1950). The visual presentation of factual data. _British Journal of Educational Psychology_, _20_, 174-185.

Wainer, H. (1974). The suspended rootogram and other visual displays: an empirical validation. _The American Statistician_, _28_, 143-46.

Wainer, H. (1980). Making newspaper graphs fit to print. In P. A. Kolers, M. E. Wrolstad, and H. Bouma (Eds.), _Processing of visible language 2_ (pp. 123-142). New York: Plenum.

Wainer, H. (1984). How to display data badly. _The American Statistician_, _38_, 137-147.

Wainer, H. and Francolini, C. M. (1980). An empirical inquiry inquiry concerning human understanding of "two variable color maps". In _Proceedings of the first general_ conference on social graphics, Leesburg, Virginia, October 22-24, 1978 (Tech. Paper 49)(pp.80-113). Washington, D.C.: Bureau of the Census.

Wainer, H. and Reiser, M. (1976). Assessing the efficacy of visual displays. Proceedings of the Social Statistics Section, American Statistical Association (pp. 89-92).

Wainer, H. and Thissen, D. (1981). Graphical data analysis. Annual Review of Psychology, 32, 191-241.

Washburne, J. N. (1927). An experimental study of various graphic, tabular and textual methods of presenting quantitative material. Journal of Educational Psychology, 18, 361-376, 465-476.

Wilkinson, L. (1982). An experimental evaluation of multivariate graphical point representations. In Proceedings. Human Factors in Computer Systems. March 15-17, 1982. Gaithersburg, Maryland (pp. 202-209).

Woods, D. D., Wise, J. A., and Hanes, L. F. (1981). An evaluation of nuclear power plant safety parameter display systems. In Proceedings of the Human Factors Society 25th Annual Meeting (pp. 110-114). Santa Monica, Calif.: Human Factors Society.

Woods, D. D., Wise, J. A., and Hanes, L. F. (1982). Evaluation of safety-parameter display concepts. (Final Report, EPRI-- NP--2239). Palo Alto, Calif.: Electric Power Research Institute.

Wright, P. (1977). Presenting technical information: a survey of research findings. Instructional Science, 6, 93-134.

Zelenka, D. J., Cherry, J. A., Nir, I., and Siegel, P. B. (1984). Body weight and composition of Japanese quail (coturnix coturnix japonica) at sexual maturity. Growth, 18, 16-28.

The vita has been removed from the scanned document