COMPARATIVE ADVANTAGES OF GRAPHIC VERSUS NUMERIC REPRESENTATION OF QUANTITATIVE DATA

by

FRED W. LACERDA, JR.

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Environmental Design and Planning

APPROVED:

Robert C. Heterick, Chairman

Shanta P. Kerkar

Harlan B. Miller

Charles W. Steger

Alan W. Steiss

June 1986

Blacksburg, Virginia

10/28/86 MCR

COMPARATIVE ADVANTAGES OF GRAPHIC VERSUS NUMERIC REPRESENTATION OF QUANTITATIVE DATA

by

FRED W. LACERDA JR.

Robert C. Heterick, Chairman

Environmental Design and Planning

(ABSTRACT)

This research proposed to determine, in the context of preliminary data analysis, whether one can generate more—and more complex—"insights" (meaningful or possibly relevant relationships suggested by the data) by looking at a graphic (multiple bar chart) representation—as opposed to numeric table—of a large, multivariate quantitative dataset (twenty variables with twenty four observations each), displayed and manipulated interactively in a personal computer-based system. If the more complex observations made possible by graphic representations can be explored in more detail—with further help from statistical and mathematical techniques—then the probability of achieving truly novel and useful solutions can be increased. The major issue involved is not how to communicate more effectively information to a large audience; it is rather what would stimulate deeper, sharper, and more expeditious analysis of a problem.

An experiment—of a "posttest only control group" design—was conducted, with eighty Subjects. Half of those Subjects were randomly assigned to

a treatment group (graphic representation of a quantitative dataset) and the other half, to a control group (multivariate representation of same dataset). Individual experimental sessions took approximately two hours, with an interactive tutorial—designed to give both groups the same level of basic skills for handling the computer program—followed by sixty minutes (maximum) for problem analysis.

The null hypothesis was there would be no differences between the scores of Subjects looking at a graphic versus a numeric representation of data for each of four classes of "insight" generation:

1 "Insights" ignoring complexity levels

2 Multiple-field "insights", exclusive of single-field "insights"

3 Multiple "field-group" (such as age groups) "insights"

4 Number of different complexity levels

A methodology was developed for objective scoring of the raw data (written notes with requested observations and inferences). Observations were eliminated on the basis of repetition, incompleteness, and lack of validation from underlying dataset.

The differences between "insights" produced by the "graphic" and "numeric" groups were statistically significant. The major differences corresponded to the higher levels of "insight" complexity—those observations relative to a large number of problem variables or to the whole dataset. The "graphic" group produced a significantly larger number of such observations.

ACKNOWLEDGEMENTS

The author is truly grateful to all the members of his Advisory Committee for their support, advice, and critical reviews. Their dedication throughout the whole project served as an inspiration to pursue that "extra mile" in addition to the usual doctoral "marathon".

A special debt is acknowledged to Dr. Harlan Miller, for the incisive and thorough criticisms of the proposal, general framework, and final drafts of this study. Dr. Shanta Kerkar offered particularly valuable assistance during the planning stages and development of the dissertation experiment. Dr. Charles Steger, Dean of the College of Architecture and Urban Studies, was most generous in providing the author with research facilities and financial support for the experiment. Dr. Alan Steiss offered help and guidance not only throughout the course of the project but also during the author's previous years of graduate studies.

The author is particularly indebted to Dr. Robert Heterick, his chairman, who provided the impetus for this study, invaluable guidance, and helped make these years the most intellectually stimulating and "insightful" of the author's life.

A very special thanks is extended to Nora Bentley, Kim O'Rourke, and Janice Huffman, for their assistance during the experimental phase of the research; to the author's colleagues at the EDP program, for the friendly interaction and discussion of ideas so critical for graduate studies; and to Charles Blair, for cheerful and welcome friendship since the beginning of this study.

The author would also like to acknowledge the extended support, at the early stages of his graduate studies, and continuous encouragement provided by CAPES—the Brazilian federal agency sponsoring graduate studies.

The final and most heartfelt thanks must go to the author's parents, Helena and Fred Lacerda. They provided not only encouragement and support—as bona fide parents do—but, foremost, they fostered, since the author's early age, the climate of intellectual curiosity and integrity under which this study was joyfully pursued.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

1 - THE PROBLEM AND ITS SETTING

1.1 - CHARACTERIZATION OF THE PROBLEM

1.1.1 - STATEMENT OF THE PROBLEM

This research proposed to determine, in the context of preliminary stages of data analysis for problem solving, whether one can generate more or better "insights" (meaningful or possibly relevant relationships suggested by the data) by looking at graphic—as opposed to numeric—representation of computer-supported relational quantitative data ("data").

The context of the research, as mentioned above, was on preliminary or exploratory data analysis, but not as a substitute for other kinds of statistical or mathematical manipulations of the data, aimed at further stages of problem solving. It was nonetheless assumed here that there are products, similar to those derived from more complex and formal "data" analysis, which are indispensable if that "data" is to prove useful for problem-solving: compression (summary) of one field (or variable), relationships between two or more fields, observations focusing on "data" subsets or on the whole dataset. There is an implicit ranking in the products just mentioned, where "insights" related to one field only would rank the least complex compared with "insights" abstracting the whole dataset. In this "complexity" scale, the higher the number of data elements (variables or records, or both) correctly abstracted, the more appropriate, or useful, that particular "insight" may prove to be for problem solving.

The operational measurement of "insights" was taken in terms of "complexity". "Complexity" describes the number of fields and "field groups" (a set of closely related fields, such as different age groups) made implicit or explicit for each "insight". Fourteen levels of complexity were distinguished.

1.1.2 - THE PROBLEM

The problem was to determine whether there would be a discernible pattern of responses for "insight" complexity generation for Subjects looking at graphic, as opposed to numeric, interactive computer displays of quantitative relational data.

1.1.3 - THE HYPOTHESIS

The null hypothesis was that there would be no differences between the scores of Subjects looking at a graphic representation of data and the scores of Subjects looking at a numeric representation of the same dataset, for each of four classes of "insight" generation:

Class 1 A count of "insights" ignoring complexity levels

Class 2 A count of multiple-field "insights", exclusive of single-field (level "A") "insights"

Class 3 A count of multiple "field-group" "insights", exclusive of levels "A" through "F"

Class 4 A count of the number of different complexity levels

1.2 - BOUNDARIES OF THE STUDY

1.2.1 - DELIMITATIONS

- The study did not deal either with visual communication issues—involving dissemination of information—or with propaganda or advertising.

- The study did not deal with either cognitive perceptual issues, particularly visual cognition, or human factors—in the ergonomic sense.

- The study was limited to examining one particular type of graphic representation of multivariate data (relational graphic matrix), using monochrome displays.

- The study was limited to undergraduate and graduate students at Virginia Tech.

1.2.2 - DEFINITION OF TERMS

Insights Insights are relationships among elements of a dataset which may prove meaningful or relevant for finding solutions to a problem defined in terms of that dataset.

Observation An observation is a sentence that describes or summarizes a subset of, or the whole of, a dataset; it is derived entirely from the data in question, as a statement of fact, of a state of being; no or minimum value judgement involved, restricted mainly to possible subjective characterization of intensity of state being described.

Inference An inference is a sentence that explains or gives a probable or assumed interpretation of observations made about a dataset or subset; supported mainly by observer's experience and knowledge of data beyond the stated dataset.

Suggestion A suggestion is a sentence that prescribes or suggests possible courses of action to achieve desirable outcome(s); it can become a solution (or one of many) if framed in the context of a defined problem.

Decision Support Systems (DSSs) DSSs are computer-based systems (integrated hardware and software) aimed at facilitating decision-making and problem-solving.

Preliminary Data Analysis (PDA) PDA is data analysis from the very first moment of data gathering until the boundaries of a problem or a decision space are clearly defined.

Multivariate Quantitative Data (MQD) MQD is a set of quantitative data elements representing four or more variables or factors; the cases of one, two, or three variables are considered special instances, extensively treated in the literature, of the general case of n-dimensional quantitative data-sets.

Graphic Matrix A graphic matrix is the graphical analog (one-to-one mapping of individual elements) of a two-dimensional numerical matrix, where each element within the set (or a given subset) is converted to a graphic symbol chosen according to the principle of preservation of scale during data conversion from one representation mode to another.

Tables Tables are two-dimensional numerical matrices with n x m elements.

1.2.3 - ASSUMPTIONS

1. Quantitative skills—among symbolic thinking abilities—are a necessary but not sufficient condition for effective professional performance in the present and in the near future; "numeracy", so to speak, is needed as much as "literacy" if only to enable communication with professionals in other disciplines or to cope with the tremendous amount of quantitative data generated by the design/building process.

2. Relational data models were assumed to be the simplest and most natural way of data organization, from an individual's perspective.

3. There is a substantial proportion of professionals that have highly developed spatial reasoning and visual information processing capabilities; their "visual literacy" skills can be successfully harnessed to handle some of the symbolic thinking abilities traditionally associated only with "numeracy".

4. The two prevailing modes of data representation, particularly for large datasets, are numeric and graphic; they are likely to remain so for the near future.

5. Graphics were assumed to be, in general, better means of presenting whole sets or subsets of data than numeric, tabular formats. Graphics allow and can induce—faster—a holistic view (gestalt) of a dataset or of a (simplified) model of its overall structure, assumed necessary prior to effective mental manipulation of that dataset; tables, on the other hand, are better for retrieving data at the individual element level.

6. Graphics are not the only means of achieving a holistic mental model of a dataset, but allow a common representation, a common language, that can be immediately shared by others.

1.3 - BACKGROUND AND IMPORTANCE OF THE STUDY

There is no need to elaborate on the impact of computers in the way professionals fulfill their responsibilities. However, the full potential of human visual capabilities for computer-based information processing has scarcely been touched, regardless of the increasing ease of use of computer displays for problem solving (as in the case of spreadsheets). Fortunately, the glamour of the introduction of computer graphics is fading away, allowing a more critical view of the relationships between data representation modes and human intellectual abilities. When better understood, those relationships can point the way towards significant enhancement of professional problem-solving, decision-making, and design capabilities. The issue is not how to communicate more effectively information to a large audience; it is rather to discover what would stimulate deeper, sharper, and more expeditious analysis of a problem.

DSSs have been the cornerstone of a large body of empirical research seeking to understand—and to provide for—the information processing needs of decision-makers, particularly business managers, in their search for more efficient and effective problem-solving methods. The research emphasis has gradually shifted from database management to modelling to data representation, as some of the basic issues related to each area have been satisfactorily dealt with, if not in practical, at least in theoretical terms.

The present popularity of spreadsheet programs in the computer software market, for instance, indicates that relational data models—always from the individual user's point of view—are indeed satisfactory for personal data base management systems (DBMS) and DSSs. Modelling capabilities are also apparently making their way into the marketplace, more often than not integrated with database management capabilities and spreadsheet programs. Graphic representations for quantitative data, however, other than univariate (as the ubiquitous pie-chart) or bivariate (scatterplots, histograms, etc.), have been the subject of surprisingly much less research than database management or modelling, the basic DSS capabilities.

The lack of appropriate iconic models to aid conceptualization of multivariate (more than three variables) datasets is at issue here, not the (recognized) importance of uni-, bi-, or trivariate graphic data-representations, which, with the great ease of use afforded by computers, are paving the way for increasing "graphicacy" levels. The trends in the visual literacy arena, however, only make the multivariate graphics problem more urgent, particularly as the computer software industry begins to address the needs of groups with cognitive and intellectual abilities quite different from those of business managers, the traditional target-user of DSS efforts.

There are many ways to achieve creative—novel and useful—solutions. In the context of problem solving, if a deeper understanding of a quantitative dataset is achieved at a fraction of the time by looking at a graphic, as opposed to numeric, representation of that dataset, then the potential time savings involved are enough to warrant a closer look at the representations making those potential savings possible. If, in addition, the more complex observations made possible by graphic representations can be explored in more detail—with further help from formal statistical and mathematical techniques—then the probability of achieving truly novel and useful solutions can be increased. The situation is notably akin to going to the highest point of a region to find one's way around; it does not preclude at all the use of maps or other means of orientation, like a compass; but it can prove tremendously helpful for most people to look at anything that will speed their understanding, of either their surroundings or of abstract data.

2 - REVIEW OF THE RELATED LITERATURE

2.1 - CONTEXT FOR THE RESEARCH

2.1.1 - DEMAND FOR ANALYTIC METHODS

The demand for analytic methods in Architecture and Planning has increased in the past ten to fifteen years (Negroponte, 1975; Mitchell, 1977; Krueckeberg, 1978). That demand is not surprising, since it corresponds to an increasingly more demanding professional environment which, in turn, reflects the increasing complexity of contemporary technological and societal changes. Krueckeberg's definition of "methods", in the context of his research into their changing role as one of the urban planner's skills, is that of "analytic tools and skills, primarily quantitative—or at least analytically rigorous—that are used in the practical work of urban planners at city, regional, and state levels of government." This definition was adopted in the present study, expanded only to include architectural practitioners working as individual consultants, as employees in Architectural and Engineering offices, and at the same levels of government indicated above.

Among the analytic methods examined by Krueckeberg were statistical data analysis, computer usage for analysis, several models of the urban pattern (for analysis and forecasting of such variables as population, land use, and so on), and economic evaluation techniques for impact and cost-benefit analysis. The conclusions of that study supported the thesis of an increase in the importance of analytical methods in planning, as perceived by planners and evidenced by the curricula of major planning schools at the time.

One specific method, statistical data analysis, was the focus of attention of the present study, which assumed that an increase in the depth of analysis could take place concomitantly with a decrease in the time needed for that analysis. This possible increase in the effectiveness of statistical data analysis could be accomplished, in principle, with a more efficient use of the human visual cognitive system, coupled with the support of computers for speed of operations.

The next sub-section sets a frame of reference for computer support in terms not only of statistical data analysis but decision-making and design activities in general.

2.1.2 - COMPUTER SYSTEMS FOR DECISION SUPPORT

2.1.2.a - Characterization of Decision Support Systems

The concept of a "decision support system", or DSS (Gorry and Scott Morton, 1971) has been defined relative to each of its three terms:

Decision Emphasizing "the primary focus on decision making in problem situations rather than ... simple information retrieval, processing, or reporting."

Support Emphazising "the computer's role in aiding rather than replacing the decision maker, thus including those decision situations with sufficient 'structure' to permit computer support, but in which managerial judgement is still an essential element."

System Suggesting the integration in context of "user, machine, and decision environment."

A comparison of DSSs with other computer-based systems, based on task structure differences, indicates significant differences :

- Transaction processing systems (TPS) are directed essentially to the collection, updating, and distribution of information according to predefined procedures, to aid in the solution of essentially structured problems. The primary focus of a TPS is on the process that needs to be facilitated through automation (Chachra and Heterick, 1982).

- Management information systems (MIS) are characterized by "predefined data aggregation and reporting capabilities ... [with] reports usually either printed in batch or queried on demand [with options normally including] only a set of predefined data extraction operators" (Moore and Chang, 1983). The primary focus of a MIS, therefore, is on the presentation of information and statistical data.

- DSSs are characterized in turn as an "extensible system with intrinsic capability to support ad hoc extraction, analysis, consolidation, and reduction [of data], as well as decision-modeling activities" (Moore and Chang, 1983). The nature of the problems and decisions to be supported is usually semistructured or unstructured. The primary focus of a DSS, therefore, becomes the search for insights about the data examined (Chachra and Heterick, 1982).

The published case studies of DSS implementation suggested, according to Keen (1980), that "DSS" might be too broad a concept, which should be further subdivided to differentiate between:

- Personal support system (PSS), "for use by individuals in tasks not involving interdependencies, so that the user can indeed make a decision";
- Group support system (GSS), "for tasks with 'pooled' interdependencies, which thus require substantial face-to-face discussion and communication";
- Organizational support system (OSS), "for tasks that involve 'sequential' interdependencies."

Given the business and managerial focus of much of early implementations of MIS and DSS applications, it is not surprising that Keen also reported a great majority of implemented DSS falling into the organizational category. In Architecture and Planning, there have been very few instances of even theoretical discussion on the subject, possibly because the lack of true computer capabilities to facilitate the intrinsically personal nature of most of the design tasks associated with that discipline.

Personal decision support systems (PSSs) are part of the focus of the present research. Professionals in Architecture and Planning, traditionally trained to employ most of their techniques and methods on an individual basis, were assumed to need more support for their personal decision-making activities than for activities at other levels of their work organization. Furthermore, the approach to such a personalized DSS must avoid capturing "the decision maker in a fixed, sequential process ... [since] no two decision makers can be expected to follow the same route to a decision, nor even to arrive at the <u>same</u> decision. The DSS must facilitate this <u>ad hoc</u> approach to problem solving, it must never inhibit it;" the major difference between PSSs and other computer-based systems, in synthesis, should be that it "must be used not only for number crunching but also for gaining <u>insights</u> [emphasis supplied]" (Chachra and Heterick, 1982).

2.1.2.b - Decision Maker Activities

Carlson (1983) summarized several studies on activities of decision makers making use of installed DSSs, with five salient factors for the analysis or design of those systems:

1. "Decision makers rely on conceptualizations [such as pictures or charts] in making a decision, and a DSS should provide familiar representations (e.g., charts and graphs) to assist in conceptualization.

2. "Decision makers perform Intelligence, Design, and Choice activities [phases of a model of individual human decision making proposed by Simon (1960) and discussed in the next section] while making a decision, so a DSS should provide operations which support these activities.

3. "Decision makers need memory aids [physical, such as scratch paper, memos, or reports; mental, as decision rules; triggers or reminders to execute certain operations, and so on], so a DSS should provide memory aids which help carry out the decision-making process." Examples of suggested DSS memory aids include a data base (from sources internal and external to the decision maker's organization), views (aggregations and subsets) of the available data base, and computer workspaces to display and preserve representations and intermediate data manipulation products.

4. "Decision makers exhibit a variety of skills, styles, and knowledge, so a DSS should help decision makers work in their own idiosyncratic ways.

5. "Decision makers expect to control their decision support, so a DSS should provide control aids which help decision makers exercise direct, personal control [when making semi-structured or unstructured decisions]."

Decision-maker control over DSSs was implicitly assumed here as necessary in practice but, for the purposes of the experimental session program built for manipulation of quantitative data, restricted to the minimum set of commands necessary to handle a limited set of possible dataset views. During the present study, no attempt was made to investigate in more detail decision making styles, briefly mentioned in the next section, with regard to human cognitive processes. This variable—decision making styles—was assumed to be controlled through random sampling of the target population for the study. Some data on these individual styles, however, may emerge from the charts depicting idea generation rates for individual Subjects in the experiment. The data generated, however, would be related only to the rate of production of certain classes of ideas, not to individual intellectual or personality factors.

The paradigm problem used in the experimental sessions is an example of the type and dataset size assumed to be most common in a professional context. The paradigm problem dataset in question happened to be a view of the original dataset from which it was extracted (the original being approximately one hundred times bigger than the derived dataset). However, even as a fraction of the original one, the paradigm problem dataset still was large enough that it may have required, from the Subjects in the experiment, further operations of data decomposition prior to their perception of most of the relationships within the problem dataset. Therefore, the basic set of operations made available to subjects during the experimental session included, primarily, data reorganization and compression facilities.

2.1.2.c - Representations of Information

Decision making phase models are reviewed in the next section. It can be stated at this point, however, that an underlying assumption of this study was that the need for conceptualizations in problem solving is stronger at the initial phases or stages of problem solving.

Conceptualizations, in the context proposed by Carlson (1983), and modelling, in the sense characterized by Chachra and Heterick (1982), can be used interchangeably. Models, classified by the latter as either iconic or analogic, are considered essential to the decision making process. In their definition, the two types can be considered as:

Iconic models	"image-generating, emotional, intuitive and ambiguous forms such as natural language, diagrams and fuzzy sketches.
Analog Models	"measurable, quantifiable, rule-manipulated, rational forms such as mathematical notations, mappings and drawings."

The relative importance of iconic versus analog models was described by Chachra and Heterick in terms of "when the limit of formal analysis is confronted, we turn to some iconic model for 'inspiration' or a fresh perception of a method of attack. The iconic aspect of problem solving is one of the reasons why graphics is so important an aspect of DSS." Iconic models, intrinsically, are physical representations of the particular concepts they embody. These physical representations, as Carlson pointed out, can also be particularly important when communication of a decision, or some aspect of it, becomes necessary. Examples of physical representations of concepts, i.e, iconic models, include histograms, scatter plots, line graphs, maps, tabular reports, spread sheets, architectural drawings, and memoranda.

The problem of representation of quantitative data is a special one. It is mostly characterized by a special symbolic graphic notation, developed over several thousand years of human history, which is so pervasive, particularly in terms of symbolic representation of quantities—Arabic numerals—as to allow a complete fusion of the symbols with the concepts they stand for. However, as known to anybody who has ever used an abacus, there are other ways to represent numbers. A more detailed discussion of this issue is postponed until Section 2.3 ("Representation of Data") of this chapter.

Regardless of the particular mode of representation of quantitative data, the organization of abstract quantitative data elements is somewhat independent of representation mode. The most common methods of data organization are: Relational, hierarchical, and network. Chachra and Heterick (1982) also suggested that "the problem solver tends to think most comfortably about information structured in two-dimensional tables" (a relation) with fields (or variables, in the statistical sense) versus records (data elements, or observations). They proposed that, regardless of the actual storage strategy used by the computer in question, "the user's view of the information should always be relational." Based mainly on the assumption of ease of use for the decision maker, the present research adopted the relational model of data organization as its standard.

Further work and familiarization with problem data may result in internalization, in the decision maker's mind, of most of the salient features or characteristics of the problem. This internalization process may be so complete that a decision maker need perhaps only to rely on mental manipulations of data to search for a solution or to decide which alternative to choose from a set of options. In practice, however, this can be assumed to happen only to a definite minority of problem situations. Representations, even though they may be considered "dead-end snapshots" of the data—since they are end products, not intermediate results capable of further manipulation—can almost certainly be useful for professionals at most of the phases of problem solving and decision making. Particularly, as has been pointed out above, at the early, fuzzy stages of data analysis.

2.2 - THE SEARCH FOR "INSIGHTS"

2.2.1 - DECISION-MAKING AND PROBLEM-SOLVING

2.2.1.a - Basic Issues for the Research

The present discussion is not to be interpreted as extending to all aspects of human problem solving or decision-making. Only those issues bearing more directly on aspects of the research problem are reviewed.

Creativity vs. Problem-Solving: Newell and Simon (1972) presented an extensive treatment of models and explanatory mechanisms for some of the main phenomena in human problem solving behavior. Simon also developed elsewhere (1979, p.139) the argument that "there is no qualitative gulf between ordinary, garden-variety problem solving and that kind of problem solving which, because of its impressiveness, we judge worth of the label 'creative.' The same mechanisms ... are at the root of both kinds of problem solving—creativity is simply problem solving writ large [emphasis supplied]."

Although the dismissal of 'creativity' as a proper label only for "impressive" accomplishments is discussed in the next subsection of this chapter, the interchangeability of the concepts "problem solving" and "creativity" was retained. The "extraordinary" cases of problem solving were not part of the main focus of the present study; rather, the cases in question were exactly those "garden-variety", or better still, "office-variety" problems that professionals face every day. If improvements in the overall effectiveness of professional performance could be associated with different methods of looking at data, one of the main goals of this research would have been fulfilled.

One example of the kind of creative thinking of concern here has been exemplified by Shouksmith (1970, p.210). Writing about the work of early psychological writers concerned with thinking and reasoning, he mentioned William James' discussion of:

> ...inventive genius among reasoners, who extract essential elements from a set of data. Novelty is introduced into reasoning by extracting a minor element and showing it to be of importance. ... This is achieved through extreme instances of association by similarity; creative reasoning occurs where the association of novel sets of data produces a surprising, yet effective, product.

Another interpretation of creativity in data analysis came from Parnes (1976), who noted that a problem solving "challenge may be met by the 'creative' discovery of relevant factors. The creativity in this case consists of devising ideas for uncovering pertinent data."

Information-Rich Domains: As it has already been pointed out, this is not a study on human problem solving per se, rather of the differential impact of data presentation modes on the problem solver. However, the particular choice of paradigm problem structure for the present study was influenced by the emphasis of more recent research (as reported by Simon, 1979) on problem solving in information-rich domains, particularly ill-structured tasks such as Architectural Design. The choice of an ambiguous problem statement, calling on quite a reasonable amount of information about the real world outside the statement wording, and of a relation containing a large number of fields can be considered as an attempt to replicate as much as possible some of the complexities inherent in even the most mundane office or professional activities of interest to the study. An important practical conclusion from some of the studies reported by Simon, in this area, is that even "innocent" changes in problem instructions can have major effects on the way people encode it, i.e., devise solution paths, or models for solution searches. The way of encoding of problem may have, primarily, a major effect on the difficulty of solution searches. No changes in problem instruction were analyzed in the present study; the wording of the problem statement tried exclusively to emphasize an open-ended approach to paradigm problem analysis, open-endedness in the sense that any "insight" search path is as valid as any other one.

Effectiveness and Efficiency: The concepts of effectiveness and efficiency in decision making are quite important for the extension (domain) of the DSS concept itself. Keen and Scott Morton (1978, p.7) defined those terms as:

Effectiveness Identification of "what should be done and ensuring that the chosen criterion [for decision choice or solution certification] is really the relevant one."

Efficiency	Performance of "a given task as well as possible in relation to some predefined performance criterion."

As Bennett (1983) pointed out, in order to achieve more effective decision making:

> ... we need to discover the decision maker's perception of the decision situation in order to increase the decision maker's effectiveness. This is fully as important as identifying the surface 'facts' of the situation. It is often the case that the 'facts' which initially appear important when working within a semistructured or structured decision situation are not the ones that, after they are explored by the decision maker, turn out to be the most influential in affecting decision outcome.

Bennett also suggested that efficiency implies a narrowing of focus, to minimize the "time, cost, or effort to complete a given activity." Effectiveness would imply a broadening of focus, a larger frame of reference, a purposive and more probing search for the right solution to the right problem definition. However, as Moore and Chang (1983) pointed out, "effective decision making by an individual at one level in an organization is often interpreted as only efficient decision making at another, usually higher, level." The concepts of efficiency and effectiveness, therefore, become tied not only to the approach to a decision making "space" but also to the specific level of skills, knowledge, or responsibilities, of the decision maker. Even though Moore and Chang made their observations as a preamble to argue for more attention to efficiency issues in organizational DSS implementation, they were retained here as basis for the assumption that the nature of the paradigm problem for the present experiment was such that all Subjects would have no preconceived notions of how to go about it (the efficiency issue) but would be forced to define their own frame of reference for analysis (the effectiveness issue).

2.2.1.b - Problem Structure Typology

The problem structure concept, as used in DSS research (exemplified by McCosh and Scott Morton, 1978), has its origins in Simon's (1960, p.5-6) decision structure continuum. One extreme of that continuum consists of "programmed" decisions, i.e., repetitive and routine, with standard procedures available for handling them. The other extreme consists of "nonprogrammed decisions", i.e., "novel, unstructured, and consequential ... [with no available] method for handling the problem because it hasn't arisen before, or because its precise nature and structure are elusive or complex, or because it is so important that it deserves a custom-tailored treatment." This last definition is practically another way of defining a creative decision, in terms of the two basic conditions for defining a creative product: novelty and usefulness (expanded upon in the next section of this chapter.) Therefore, the decision programmability continuum could also be restated in terms of a decision creativity continuum.

In the DSS research area, the transformation of that concept into the one of problem structure (Gorry and Scott Morton, 1971) was sought mainly to place more emphasis on the basic character of problems and less on computer dependency. Three basic categories were suggested: structured, semistructured, and unstructured problems or their component tasks. According to Keen (1980), structured "tasks can be automated or routinized, thus replacing judgement, while unstructured ones involve judgement entirely and defy computerization. Semistructured tasks permit a synthesis of human judgement and the computer's capabilities." In the fully unstructured state, all a DSS can provide is "access to the data and useful ways of displaying them."

The problem structure concept, as Keen also pointed out, "leaves unclear whether structure is perceptual or intrinsic to the task." Moore and Chang (1983) suggested that DSSs must be designed or discussed "in terms of user-specific problem structures." In other words, the notion of problem structure should be analyzed against the background of decision maker cognitive styles, skills, and the degree of familiarity of the decision maker with the problem space.

2.2.1.c - Decision Making Phase Models

The problem solving stages first described by Dewey (1910, 1931) became the basic paradigm around which most of decision-making theory evolved, particularly in the work of Simon (1960), who paraphrased Dewey's three factors of judging , or interpretation of facts (1910, pp.101-7) as:

1. What is the problem? [i.e., the "controversy" that requires judgement]
2. What are the alternatives? [i.e., the opposite "claims" that need to be defined and elaborated for a proper decision]
3. Which alternative is best? [i.e., the final decision, or "sentence", closing the particular matter in dispute]

Decision making activities have been classified by several other authors roughly along the same lines, constituting the so-called "phase" models. The most cited model in the DSS literature and of decision making studies, in general, seems to be Simon's (1960, p.1-4) three-phase model, succinctly described as "Intelligence/Design/Choice", for each of its three basic component activities. This model, in turn, can be identified (Kickert, 1980) as an extension of Simon's "bounded rationality" model of decision making (1959, 1964), with one phase for problem identification and another for search and choice of solution(s).

The first phase of Simon's model—"searching the environment for conditions calling for decision"—was called "intelligence" activity, "borrowing the military meaning of 'intelligence'." The "design" phase is concerned with "inventing, developing, and analyzing possible courses of action." The last phase, "choice", consists of "selecting a particular course of action from those available." Needless to say, many extensions have been proposed to this model, several of them discussed in Kickert (1980). One model, in particular (Mintzberg et al., 1976), has as a distinguishing feature not just a more detailed breakdown of Simon's three phases (relabeled "identification", "development", and "selection") but of "dynamic factors", such as interrupts and feedback delays, that seem to account for much of the interactions that take place during decision making.

As for model complexity, Simon stated that his model should be viewed as an iterative hierarchical process, where any component activity could be in turn subdivided in its particular subcomponents for the "intelligence/design/ choice" phases. The Mintzberg model was less "modular" and extensible, but emphasized more strongly a possible arbitrary sequence that activities may follow in a real world setting, with more explicit accounting of environmental factors (organizational, political, informational) that may disturb any "rational" sequence of decision making activities.

Kickert (1980) also reported a study by W. Kirsch to ascertain the descriptive and normative validity of phase models in general, with a general conclusion that phase models constitute only a "very approximate scheme of decision-making processes and only [express] a general tendency". As Sprague (1980) pointed out:

> There is no universally accepted model of the decision-making process, and there is no promise of such a general theory in the foreseeable future. There are just too many variables, too many types of decisions, and too much variety in the characteristics of decision makers.

Although the area of human decision making and problem solving, particularly its basic paradigms, may not have been fully tested and researchers may not have taken all the major parameters of the process into account, the existence of a general tendency of sequential decision making phases was assumed in the present study. This assumption was necessary in order to describe, with the help of categories proposed through phase models, the approximate decision making activities that could most benefit of alternative modes of displaying quantitative data.

The present study attempted to estimate the effect of different modes of data representation for a set of tasks—data manipulation operations—falling within the "design activity", or problem analysis, range of Simon's decision making phase model.

2.2.2 - THE CREATIVITY DIMENSION

2.2.2.a - Creative Behavior and Products

Definitions: The issue of outstanding creative objects (conceptual or physical), in whatever discipline or area of human activity, fell beyond the focus of attention of this study. The interest, here, was much more on the "process" and "products" associated with different types of normal, everyday behavior which nonetheless produce results that can be labeled as 'creative' in particular circumstances. The importance of creativity for data manipulation operations has been expounded by de Bono (1971):

> Ideas are the spectacles through which we look at data in order to see information. Data is useless until we look at it through an idea—only then does it become useful information. Different people looking at the same data will derive different information from it according to the idea which each of them uses to look at the data. Old data looked at through a new idea gives new information. Creativity is concerned with bringing about new ideas and updating old ones. Since data is usually available to everyone, it is the creativity with which an individual can look at the data that makes the big difference.

If creativity becomes a desirable goal, as implied in the above statement, the issue of defining creativity follows suit immediately. A major advance in creativity studies took place, during the first decades of this century, when the emphasis shifted from categorization of degrees of creativity (creative or not, highly or moderately creative, and so on) to necessary conditions for definition of a creative behavior or product, independently of the quality assessment of the behavior or product in question. At this point, two factors need to be taken into account for the definition of creativity. First, the candidate objects—behaviors, products, processes—for consideration as "creative" are context-dependent, since "creativity" implicitly assumes a reference base for comparison. Second, what is truly "creative" in sub-atomic physics cannot be directly compared to a "creative" work of art. The criteria for inclusion of products for evaluation in each of those disciplines are by nature quite different. Second, even within a given discipline or area with acceptable candidate objects, there may be several levels (personal, peer group, and societal) or frames of reference that can be used to determine if conditions for creativity are indeed being met (those frames of reference are analyzed in the next subsection of this Chapter.)

Within the confines of a given discipline or area of activity and of the particular frame of reference chosen, the definition of a creative object becomes possible if fulfillment of a basic set of conditions is met. A current definition of those conditions, in terms of "novelty" and "usefulness" (Guilford, 1976b), was retained here; it is a universally applicable, generalizable but small set. This definition is also analyzed in conjunction with frames of reference.

The measurement of criteria for assessing the creativeness of objects in particular disciplines was not discussed here. There hardly seems to be, at any rate, much agreement in any discipline over the qualitative categories for creative assessment, at least not until the passage of time or pressing concerns force an evaluation (or reevaluation) of an object. Artistic evaluation is a case in point; accepted canons of artistic quality have been subject to extreme fluctuations over time, depending on particular local societal conditions.

At any rate, surrogate measures of creativity were used for analyzing the products generated by Subjects during experimental sessions for the present study. Those measures are presented in more detail in Chapter 3, "Research Framework and Methodology".

Frames of Reference: The three basic frames of reference against which the creativity of an object may be measured (Guilford, 1976a; Glover, 1980) are the personal, peer group, and societal levels. Since consensus of opinion is at the heart of the characterization of a given object as creative, it follows that the criteria for creativity may be more simply stated and less subject to random variations of subjective interpretations in a smaller human group than in a larger one, simply because of a tendency towards homogeneity within groups that are usually defined, to begin with, on the basis of shared interests or traits. The two basic conditions usually accepted as basic for characterization of creativity are discussed below.

Novelty and usefulness can properly be used for determination of the creativity of an object—behavior, process, or product—within a well defined frame of reference, i.e., the human group within whose boundaries and to whose objects the object in question will be compared. Novelty may be a characteristic not of the object per se, but of some of its attributes. A literary novel may not be "novel", i.e., the first of its kind, but its contents, the intellectual and emotional topics it may raise, those would have to be treated in a "novel" manner before the work itself can be so rated. In a similar manner, data manipulation operations in statistics are not "novel" per se, but the relationships uncovered by those operations have to be, obviously within the boundaries of the problem space.

The concept of usefulness has been more closely tied to the third level of reference, the societal level. The social usefulness of an object may be a noble and desirable goal, but its determination will involve, by definition, value judgements which may elude the best efforts at extracting a consensus. It is only a majority consensus, in the long run, that can guarantee fulfillment of the condition. Dadaism, one of the major art movements of the twentieth century, may have been accepted as novel, at the time, but most certainly not useful by the largely bourgeoise European society of post-World War I. It took several decades of profound societal changes to trivialize, i.e., make acceptable, and even part of, most modern industrial societies, the notion that anything—physical and conceptual objects—can be acceptable as valid works of art per se, by the simple manifestation of intent, by an artist, of creating a work of art (the "happenings" of the 1960s, Christo's wrappings of landscapes and physical landmarks, and so on). Marcel Duchamp's famous "bicycle", an assemblage of "objets trouve"—footstool, a bicycle frame, and an umbrella—may not be "liked" or even rated as a "good" work of art. By its sheer novelty and impact, however, it became the Dadaist movement's most well-known and "useful" piece, bearing direct impact on the movement for a major expansion in the set of criteria for definition of works of art.

Scientific works are usually not subject to the same frame of reference as works of art in general. Peer group evaluation is normally the frame of reference used, simply due to complexity of subject matters. Work considered useful for advancement of science, by review of a scientist's peers, may not be viewed as such by society at large, as might possibly be the case with atomic bombs. In the case of professional and managerial settings, peer review can range from one's co-workers to the larger component organization or to a whole multinational industrial sector. Corporate "raiding", for instance, the practice of (mostly unfriendly) acquisition of a company's publicly traded stock for the purpose of assuming control of that company's operations, would certainly not be voted as a "useful" practice—novelty issue aside—by many corporate workers at all organizational levels. But it has been argued that it may be a useful practice from a societal standpoint, with a streamlining of larger scale operations, increased efficiency (albeit not necessarily effectiveness) in the production of goods, and so on. The value judgments inherent in this particular issue and many others, however, cloud any attempts at a more "rational", or measurable, definition of usefulness.

To be useable at all, "usefulness" needs to be measured against objective criteria, quantifiable or not, but at least verifiable. At the peer group or the personal level, usefulness may depend on particular circumstances—goals, objectives, time pressure, resource constraints—in a constant state of flux. Data manipulation operations, as only one group within a much larger set, have ultimately to be measured against the problem solving context and the choice of alternative outcomes. Therefore, it may be very difficult, if not impossible, to assess a priori the ultimate usefulness of data manipulation per se or of its products.

The usefulness of an alternative mode of data representation can only be fully tested under a battery of real, or realistic, problem solving situations in information-rich domains. More simply, such a test can sometimes be met by the introduction of the data representation system in the community of potential users, either by commercial marketing or through other means of diffusion. An alternative approach, adopted here, involves the examination of the chain of elements contributing to successful problem solving, particularly the stages of individual creative thought.

Stages of Creative Thought: Among experimental studies of stages of the creative process reviewed in the literature (Guilford, 1976a, b; and Freeman et al., 1971), the pioneering work of Patrick (1938, 1941) remains a major reference point. Patrick investigated creative thinking in terms of its division "into four stages of thought, namely: preparation, incubation, illumination, and verification," as proposed by Wallas (1921) and Poincare (1946). Preparation refers simply to familiarization with the problem space. Incubation, which follows preparation (or may accompany it), refers to the intermittent recurrence of idea(s), with more or less modification, while the individual concentrates on other topics. Illumination, which normally follows incubation, corresponds to the first written draft, or sketch, or physical (tangible) manifestation of the general idea. These first three stages are akin to Poincare's three conditions for creativity: conscious problem analysis, unconscious work, and careful nurturing and analysis of ideas unconsciously developed. The fourth stage, verification, or revision, is an evaluation of the adequacy of the tangible "product" to represent the particular thought or idea. Guilford (1976a) referred to this stage as "elaboration".

Patrick found that the creative steps above, which represent ultimately an alternative way of describing problem-solving phases, did effectively occur in a study of poets and one of artists, although departing from the exact sequence described (preparation, incubation, illumination, and verification) and even overlapping. Mintzberg's model (1976) could be adapted to Wallas's creativity stages as a way to account for potentially significant shifts from the sequence originally postulated. The present study (as already emphasized in the subsection on problem-solving and decision-making) focused on inducing more favorable conditions for the first two stages of creativity according to Wallas. These favorable conditions are, in essence, availability of methods for holistic data analysis in a much shorter time frame.

2.2.2.b - Conditions Affecting Creative Thinking

Cognitive Styles: Strategies (regularities in decision making patterns) "which recur consistently within the same individual make up what may be referred to as the cognitive style of that individual" (Shouksmith, 1970; Witkin and Goodenough, 1981). Several polarities, representing selected aspects of thinking processes, have been investigated by creativity and decision-making researchers. Among the cognitive styles identified in that manner are the "reactive/self-paced", "analytic/heuristic" and "divergent/convergent production" styles. A "reactive" style is more characteristic of individuals who perform better under pressure and quickly shifting conditions; a "self-paced" style is more characteristic of those who prefer to manage their own time and "pace" themselves toward conclusion of their tasks (Glover, 1980). In terms of the second set of styles, an "analytic" decision-maker would be one more inclined to work with quantitative data; a "heuristic" one would tend to examine the entire problem. In terms of production of ideas, "divergent production" is a style requiring a flexible, free-associative method of thinking, while "convergent production" is a much more focused, abstract and persevering mode (Shouksmith, 1970, p. 210-4; Guilford, 1976a). One general conclusion mentioned seems to be that there is no single cognitive style, within a given polarity, associated with a creative individual; it is rather the ability to switch to the appropriate style as conditions demand that would characterize a creative thinker. For example, a heuristic, divergent-production "style" would be appropriate, say, for early creative stages, but for "verification" or "elaboration" of a final solution, an abstract, convergent-production style would be more desirable, at least in principle.

The present study was more concerned, in the context of cognitive styles, with investigating the appropriateness of graphic and numeric representations of quantitative data to foster a more holistic, flexible examination of the problem space at the early stages of problem solving. For this reason, a more detailed review of conditions affecting creative, "divergent" thinking is presented below.

Deferred Judgment: Evaluation of ideas is one of the conditions affecting "creative-thinking processes, by way of facilitation or inhibition. ... Absence of self-evaluation while generating ideas [is] known as 'deferred judgement'" (Guilford, 1976a). The concept was first introduced by Osborn, who also labeled it "brainstorming", the name under which it is better known at the present. Osborn (1957, p.228; 1962), however, placed equal emphasis on exclusion of criticism and group evaluation of ideas, which were not reviewed during this study. The major thrust of deferred judgement is on establishing a clear distinction between generation and evaluation of ideas, to give a better chance to potentially good ideas that might otherwise be prematurely rejected. A criticism leveled against early studies investigating the effectiveness of "brainstorming" techniques was the lack of adequate controls for the "relative effects of prior training in the actual subject-matter of the problems themselves besides the effects of prior training in 'brain-storming', ... [with that prior training corresponding to] the stage of creative thought described by Patrick ... as 'preparation'" (Freeman et al., 1971, p.57).

The present study included controls not only for familiarity with subject-matter (practically non-existent, for all Subjects) but also for the minimum level of skills necessary for manipulation of the interactive computer system designed for the present experiment (see Chap. 4, "Data Collection").

Divergent versus Convergent Thinking: The "divergent/convergent" cognitive styles imply, by their very labels, that the individual problem-solver starts from given information towards a desired solution or creative product. Divergent thinking would be "a matter of scanning one' [sic] stored information to find answers to satisfy a special search model" (Guilford, 1976a); convergent thinking would be, accordingly, a matter of narrowing the search scope and focusing on practical ways to cope with the limitations imposed on the problem solution space. As Bennett (1983) pointed out, discovering "how to be more effective requires 'divergent' thinking as the decision maker is stimulated to expand the set of open decision possibilities [whereas discovering] how to be more efficient requires 'convergent' thinking as the decision maker uses tools to achieve results in reduced time or at reduced cost."

Evaluation would be, relative to divergent and convergent thinking, a matter of determining the particular fit of a given answer (or set of answers) to the problem situation. According to Guilford (1976b):

> Relaxed evaluation [deferred judgement] would permit a broadening of the base of the search, whereas an evaluative attitude with some degree of strictness would narrow the search. In doing so, however, it may lead more efficiently to good answers. This should depend upon the clarity and accuracy of the search model. If the thinker has a good search model, the application of evaluation while he thinks should be helpful.

Personality and Motivational Characteristics: There are several factors, besides intellectual abilities as divergent or convergent thinking, that have been found or postulated significant for the understanding of creativity. Among them are intelligence, perseverance, tolerance for ambiguity, risk-taking ability, innovativeness, and even sense of humor. Of particular interest here is the case of "intelligence", as measured by Intelligence Quotient (IQ) tests. The "threshold", or "triangular scatterplot" theory (Torrance, 1962, p.63), hypothesizes that IQ sets an upper limit to creative potential, but that the relationship is not symmetrical. In other words, there seems to be a cut-off point in IQ levels, experimentally situated in the 100-120 range, below which only low levels of creativity have been identified and above which creativity does not seem to be correlated with IQ levels. This might be one of the most relevant control factors in future studies expanding the scope of the present research.

The Structure-of-Intellect Model: Competing views of the concept of "intelligence" have been presented in two basic ways: as a broad unitary factor or as a combination of distinct intellectual abilities (Freeman et al., 1971, p.10). While the fragmentation of the unitary "intelligence" concept may seem unwarranted to many psychologists, as reported by them, the postulation of discrete intellectual abilities—which may or may not constitute facets of "intelligence"—has been reasonably successful. For the present research, the issue of a superordinate versus an aggregate concept of "intelligence" was less relevant than the insights and powerful synthesis of human intellectual abilities brought by the "fragmentary" view of intelligence.

Guilford's work (1959, 1967, 1976a, 1983, 1984), has been the pivotal element in the polarization of the views on intelligence. His "Structure-of-Intellect" (SOI) theory of intelligence and its components is based on an extended series of studies, carried out at the Aptitudes Research Project at the University of Southern California, applying multivariate methods of factor analysis. According to Guilford (1976a, p.9):

> Rejecting the prevailing doctrine that intelligence is a single monolithic ability, and also the view that creative talents are something outside the realm of intelligence, the studies began with the assumption that there are several, perhaps many, distinguishable abilities involved. It was also assumed that creative talents are not confined to a few favored individuals, but are probably distributed to different degrees throughout the population. Creative talents could therefore be investigated without being restricted to observation of the gifted few.

Guilford's SOI model, presented in Figure 1 on page 40, is conceptually quite simple. It articulates the range of intellectual operations (or processes) that act upon certain "objects" to derive certain "products" (mental structures). Among the operations postulated, divergent-production and convergent-production of ideas have been suggested by Guilford as quite important for creative thinking.

The development of the SOI theory has suggested that the intellectual abilities encompassed by the model act usually in combination with several others. Although there may be, for instance, predominance of materials of figural (visual) or semantic (verbal) content associated with cognitive processes, information may be derived from materials in other "languages", i.e., symbolic (numbers and letters), auditory, or even behavioral. The optimum combination of abilities to absorb information or to perform creative problem-solving, and so on, would vary from individual to individual and domain to domain, with a redundancy of "languages" for transmission of information being probably more effective (Freeman et al., 1971, p.24).

An important concept associated with the model is that the abilities are to be recognized mainly "as distinct but generalized thinking skills" (Guilford, 1976b), which imply a rather dynamic, almost constantly "shifting" view of intelligence, a view that recognizes possible limitational factors (as the IQ "threshold" theory) but that accepts the possibility of increasing existing levels of intellectual skills.

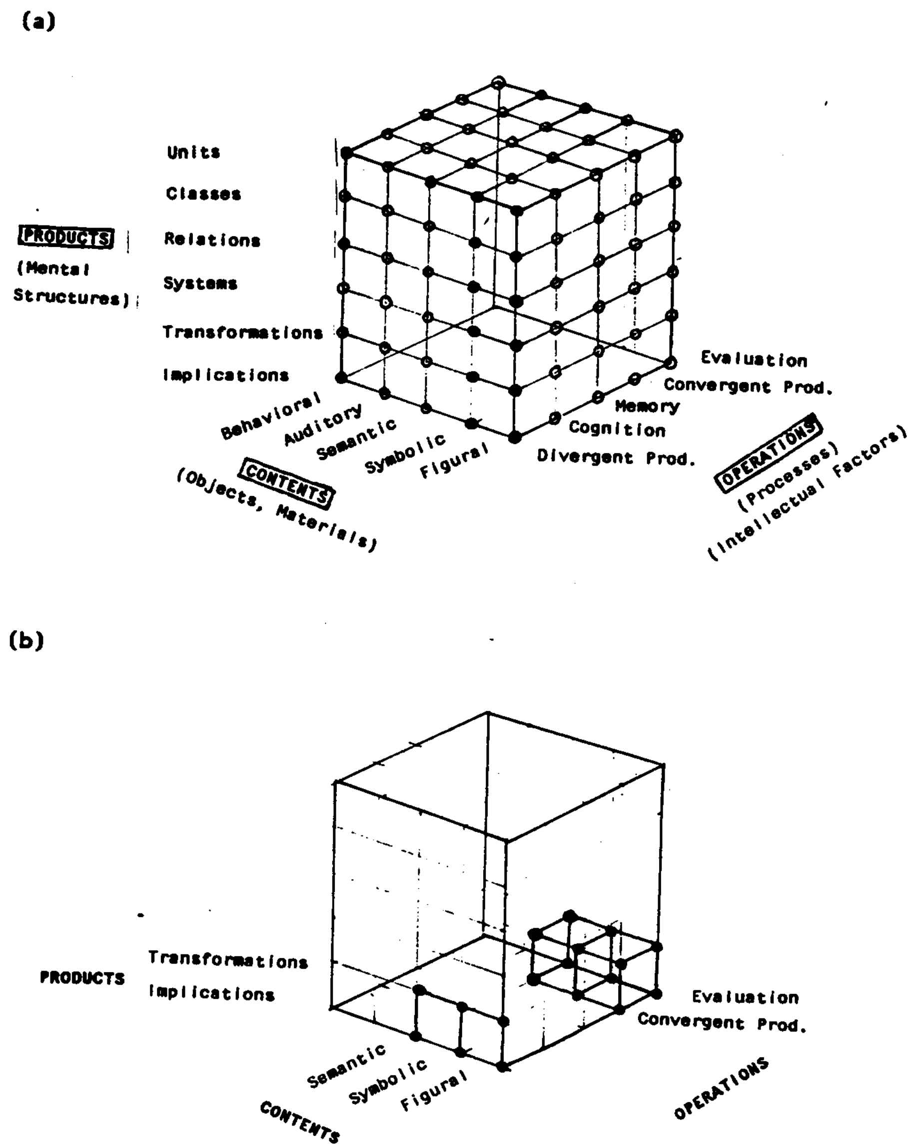


Figure 1. Guilford's Structure-of-Intellect Model: (a) Diagrammatic representation and (b) selected abilities relevant for data analysis.

In view of particular demands from quantitative data analysis activities, there are a few intellectual abilities which might prove relevant for predicting the probability that certain individuals are more likely to benefit from one "language" of presentation of information than another (see Figure 1 on page 40). The subset of abilities in question is defined, in terms of content of original information, by "figural", "symbolic", and "semantic" objects, processed by "divergent-production", "convergent-production", and "evaluation" operations, with particular attention to the products "transformations" (changes to the original information content) and "implications" (ability to reorganize associated items).

While the investigation of the intellectual profiles of problem solvers was considered premature in the context of the present research, it may be a natural direction for future research in the human information processing area. It should be noted that there are already a number of studies, particularly in Management Science, devoted to this topic, albeit limited to one or two cognitive styles or intellectual abilities as control variables, and not to a large subset of those abilities.

2.2.2.c - Measurement of Creativity

In qualitative terms, the two basic criteria for evaluation of creativity of an object (a physical or conceptual entity, or a behavior, for instance) are novelty and usefulness. However, albeit important, they are hardly quantifiable, even at the personal level of reference, which would in principle should be the simplest and easiest to ascertain. There are four other criteria, however, that capture some of the facets of the creativity phenomenon, and can serve as useful surrogates for the qualitative ones. Those four, addressed in more detail below, purport to measure specific, tangible products (ideas, in written or oral form; drawings, etc.) created by an individual in any kind of creative process. The four criteria are: fluency (number of different objects generated), flexibility (number of different kinds or classes of objects), elaboration (the level of detail in the development of an object), and originality (in terms of statistical infrequency of an object).

Fluency: This measure corresponds to the quantity of different ideas, or objects, that an individual generates in response to a problem situation or creative impulse. Implicit in this concept is that the number of ideas that an individual can generate is directly proportional to the number of possible problem solutions he can achieve. It has been pointed out that quantity of ideas generated may not be directly related to their quality, but also that this situation may be context (problem) dependent. For problem solving, however, particularly in the first phase of "intelligence", the discovery of relevant factors may be strongly correlated with an ability to uncover observations, or facts, necessary to stimulate the problem-solver to search for possible solutions (Parnes, 1976). Fluency—in terms of observations or ideas—is, for practical purposes, what the expression "food for thought" is all about; unfortunately, no "recommended daily allowances" are known for problem-solution intellectual "diets".

Fluency can be measured, in practice, in several different ways, once the object-unit to be measured is defined, within any cognitive representation mode—written, oral, numeric, graphic, tactile, gustatory, or olfactory. In most creativity studies, written or oral ideas have been the predominant unit of measurement, ranging from single words to complex simple sentences, or even whole poems, as in Patrick's case (1938, 1941).

A particularly common way of characterizing fluency is through the measurement of ideational fluency, or rate of generation of ideas (belonging to the same class). In practice, there is less emphasis on the rate of generation of ideas than on the absolute scores achieved within a strict time limit for experimental task performance. Guilford (1976b) also related two other kinds of fluency, associational and expressional, the latter only indirectly relevant for the present study. Associational fluency refers to the ability to complete relationships between a given object and any other objects or ideas that an individual can generate. Thinking by analogy, in problem solving, would be an example of this particular ability. Expressional fluency refers to ease of generation of sentences.

Flexibility: According to Guilford (1976b):

> Flexibility in thinking means a <u>change</u> of some kind—a change in meaning, interpretation, or use of something, a change in understanding of the task, a change in the strategy of doing the task,

or a change in direction of thinking, which may mean a new interpretation of the goal.

Two measures of flexibility, related by Guilford, are: spontaneous and adaptive. The first one corresponds simply to the variation in the kind of solutions, or ideas, generated in response to a problem situation. Spontaneous flexibility can be, in turn, measured in terms of the total number of categories of ideas generated, or in terms of the rate of change in the categories of ideas. The second general measure, "adaptive flexibility", involves problem situations which specifically call for changes "in interpretation of the task, in approach or strategy, or in possible solutions"; this measure, however, is considered to be actually an operationalization of "originality" (see below).

Elaboration: In creativity terms, elaboration is the process of detailing ideas generated during a problem solving or creative situation. Elaboration has been reported (Glover, 1980) to be negatively correlated with fluency. In a similar manner to the situation with divergent and convergent thinking, there are probably different needs for either fluency or elaboration at different stages of problem solving, and the more creative individual has to have the capacity to switch from one "style" to another. Simon's "design" phase of problem-solving, or Wallas's "illumination" and "verification" creative stages, represent situations where the ability to fill out ideas or outlined solutions can be highly useful. Guilford referred to this particular characteristic as the ability to generate a "variety of implications".

Originality: Originality, one could say, is in the eyes (or the mind) of the beholder. Of the two major criteria for creativity in general—novelty and usefulness—only the former can be approximately measured, in terms of originality. Objective measurements of originality can only be made in a very rudimentary way, in terms of statistical infrequency of a particular object, or idea, within a restricted group or population, since there is no possible way to determine, even relative to a single individual, whether his ideas are truly novel, or original.

2.2.3 - MEANINGFUL DATA RELATIONSHIPS

If, as R. W. Hamming said, "the purpose of computing is insight, not numbers," then it might be useful to examine what is generally referred to as "insight". As implied by Poincare's stages of creativity, "insight" would correspond to a sudden "illumination", primarily driven by the unconscious side of the intellect. Freeman et al. (1971, p.52) listed several authors and their descriptions of what characterizes "insight":

> [W. Kohler] ... defines "the appearance of a complete solution with reference to the whole layout of the field". [C. E. Osgood]: "the solution of the problem precedes the actual execution of it". [R. M. Yerkes]: "the solution is suddenly, directly and definitely carried out". In most of the accounts of 'insight' the successful response appears suddenly, it is easily repeated in future trials and is not easily forgotten or eliminated. ... this kind of situation is characteristic of the 'eureka' sensation in creative discovery. In view of [H. F.] Harlow's findings it would appear that 'insight' might be, demonstrably, a result of extensive practice on related problems.

Most of the definitions above may have to do with "Insight", in the sense of a major solution or definitive response to a creative impulse. An interpretation of "insight" as an ability, parallel to Harlow's, had already been expounded by Dewey (1910, p.104), who defined "judgements" as the basis for the whole process of thinking, culminating in a final judgement, "the conclusion", i.e., the solution to a problem or final decision. In his words:

> To be a good judge is to have a sense of the relative indicative or significative values of the various features of the perplexing situation; to know what to let go as of no account; what to eliminate as irrelevant; what to retain as conducive to outcome; what to emphasize as a clue to the difficulty. This power in ordinary matters we call knack, tact, cleverness; in more important affairs, insight, discernment.

However, a more pedestrian notion of "insight" is necessary to indicate the discovery of those factors, in problem solving, that are no more than possible clues to the solution of the problem at hand, that may never lead to anything, but that just might trigger the right conditions for the major "Insight". For the purposes of the present study, "insight" was defined as those observations—facts, inferences—that are generated by an individual in direct response to a problem situation, as part of "preparation" or "intelligence" activities during the course of problem solving, or even during "incubation", "illumination" or "design" activities.

The quantity or the quality of "insights" (or both) generated by an individual can safely be said to be dependent on many factors, among them the individual's level of thinking skills, experience in the problem area or similar ones, personality and environmental factors, and so on. The particular cognitive mode of presentation of the data associated with the problem can also be assumed to determine, to a large extent, individual performance in problem solving, even controlling for level of skills in information processing in the cognitive mode(s) in question. The next section of this chapter deals with cognitive aspects of the human intellect, and the particular modes of data representation that can be, or have been, used to increase human information processing capacity during problem solving, to generate not only more "insights" but better "Insights" as well.

2.3 - REPRESENTATION OF DATA

2.3.1 - HUMAN COGNITION AND INFORMATION PROCESSING

2.3.1.a - Information Processing

The study of communication of information, regardless of the particular sensorial mode involved, was enormously expanded in the past few decades. Particularly noteworthy is the work of Miller (1956), who, in a seminal paper, presented several important new concepts for that field, several of them relevant for the present research. The minimum unit of information, the "bit", was defined as "the amount of information that we need to make a decision between two equally likely alternatives", with a derived general rule stating that "every time the number of alternatives is increased by a factor of two, one bit of information is added." He further distinguished between "bits" and "chunks" of information, with the latter constituting a not always very clear amalgam of bits of information, reorganized or grouped into familiar units (chunks). This reorganization, or recoding, of information would be a crucial factor to explain efficiency of information stored in memory.

The evidence presented by Miller suggested that the "span of absolute judgement" (the "channel capacity", or the number of bits of information that are effectively transmitted to the observer) is limited by the amount of bits of information (approximately seven). The "span of immediate memory", a different but related kind of limitation, is measured not by bits but by chunks of information, also limited to approximately seven (Miller's "magic number", again), regardless of the number of bits of information stored in each chunk. In other words, "we can increase the number of bits of information that [each chunk] contains simply by building larger and larger chunks, each chunk containing more information than before." The recoding process, therefore, becomes an "extremely powerful weapon for increasing the amount of information that we can deal with." The power of graphics for information processing, it is contended here, derives precisely from its data recoding capabilities.

2.3.1.b - Modes of Perceiving

Most of the sense modalities have been extensively dealt with in the literature: speech, auditory, visual, and, to a lesser extent, tactile (e.g., pseudo "visual" representations for the blind), olfactory and gustatory (essence making, wine and tea tasting, and so on). More refined modal distinctions have been suggested by Pick and Saltzman (1978, pp.2-4), based not on the receptor per se but on the nature of the information desired:

> ... different modes of perceiving may be implied: (1) when one type of information rather than another is extracted from a given pattern of stimulation; and (2) when a specific type of information is extracted from very different patterns of stimulation. In the first case, the functional distinctiveness of information is critical. The stimulation itself is ambiguous, that is, it can be read for one meaning or another. ... In the second case, the generality is critical.

It seems that quantitative data analysis would fall under the second category mentioned above, inasmuch as there is an abstract, "invariant set" of data element relationships that are theoretically attainable by whatever sensorial modality used to extract them. On an individual basis, that might certainly be possible. However, there are limitations in the skills possessed by individuals to analyze information efficiently through a given perceptual mode. Different individuals would, therefore, extract different sets (or rather, subsets) of information from the same pattern of stimulation, subject to their relative skills in interpretation of that particular sensorial modality.

The question becomes not so much whether quantitative data analysis is effectively a general type of information retrievable through any type of stimulation, but rather what particular sense modality can effectively help individuals to uncover the broadest possible range of data relationships from the "original", "invariant" dataset.

2.3.1.c - Visual Cognition

Seymour (1979) defined visual cognition as encompassing two areas of human competence, the first "competence to deal intelligently with the configuration of objects which he encounters moment by moment as he moves about his world", the other "his competence in the interpretation of visual symbols", encompassing literacy and numeracy, "and refers most directly to the ability to read and use symbols in mathematical and scientific problem solving." Seymour cited, in support of such distinction, Guilford's (1967) studies of human intellect, particularly the identification of abilities for solving "figural" and "symbolic" types of test item.

The present research focused not only on the second area mentioned above—visual symbol interpretation—but on specifically two of its submodalities: numeracy and "graphicacy" (akin to speech versus non-speech modal distinctions).

2.3.1.d - Visual Languages

The thrust for further definition of key concepts and descriptive models for symbolic visual cognition is more apparent in the work of individuals more actively involved in its practical applications, ranging from graphic designers (Twyman, 1979) to geographers (Bertin, 1981, 1983; his ideas are discussed in more detail in the next subsection of this chapter).

Twyman presented a schema for the study of graphic languages which, although tentative in places, offer an useful diagrammatic model embracing all graphic languages (see Figure 2 on page 52.) The model, presented in a two dimensional matrix structure, relates modes of symbolization and method of configuration (i.e., the structure of a graphic message influencing or even determining the perceptual behavior of the user). The particular data representation displays used in the present study (both matrices) occupy cells '5' (verbal/numerical matrix) and '26' (schematic matrix) of the model. Twyman discussed the expansion of the model to three or four dimensions, including temporal factors and the graphic variables isolated by Bertin (discussed in the next subsection). Although important in themselves, the lack of those possible additions does not detract from the empirical usefulness of the basic model.

For the specific type of visual language of interest here—graphics—the most precise and exacting model, however, comes from Bertin. His "synoptic" of graphic constructions (Figure 3 on page 53) contrasts the number of characteristics of the statistical object (i.e., the independent variables) with three basic traits of a graphic object, namely, whether it is ordered, reorderable, or topographic. The particular graphic constructions used in this study fall in the upper-left position of the synoptic, constituting the so-called "reorderable matrix" (equivalent to the basic matrix from a relational data base model).

METHOD OF CONFIGURATION

MODE OF SIMBOLIZATION	Pure linear	Linear interrupted	List	Linear branching	Matrix	Non-Linear directed viewing	Non-Linear most options open
Verbal/ Numerical	1	2	3	4	5	6	7
Pictorial & verbal/ numerical	8	9	10	11	12	13	14
Pictorial	15	16	17	18	19	20	21
Schematic	22	23	24	25	26	27	28

Figure 2. Twyman's Graphic Languages Model: Diagramatic representation.

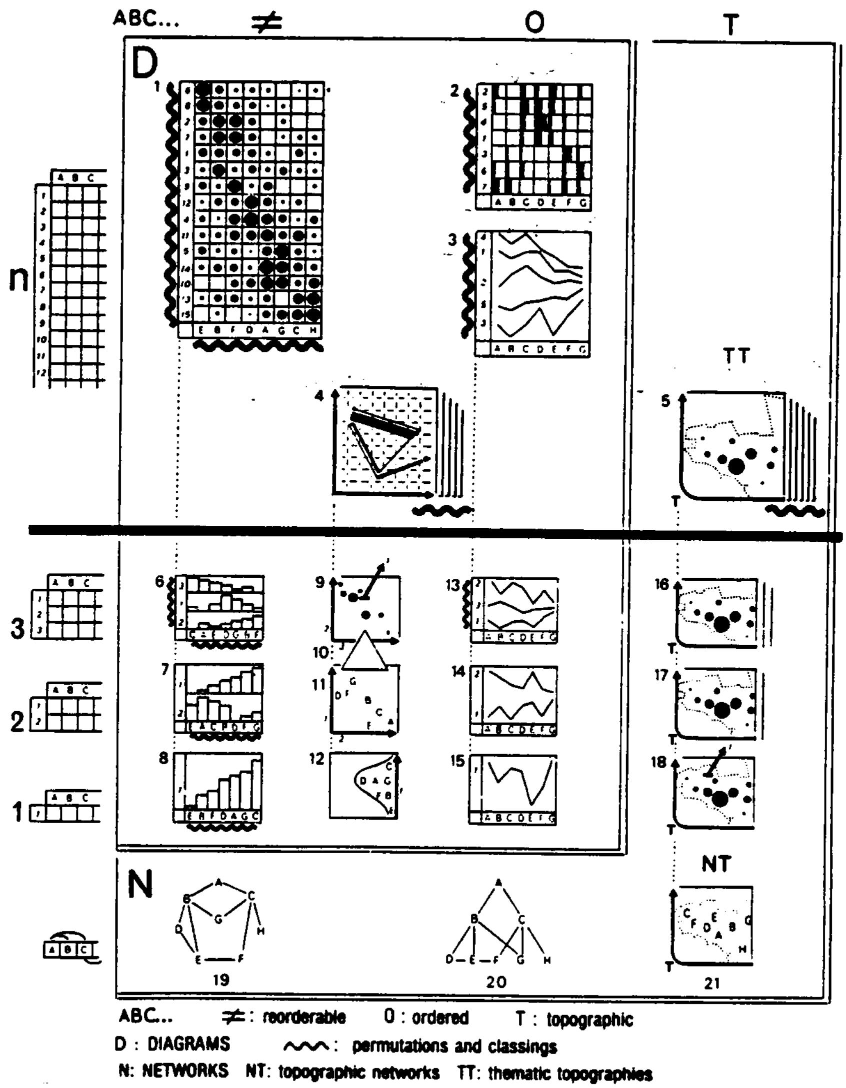


Figure 3. Bertin's Synoptic of Graphic Constructions: Diagramatic representation.

The following subsection deals in more detail with specific issues of graphics as a very specific type of visual language, particularly developments in statistical graphics.

2.3.2 - GRAPHIC MODE OF DATA REPRESENTATION

2.3.2.a - Statistical Graphics

No detailed review of statistical graphics has been attempted here. For such reviews, see Beniger and Robyn (1976), Izenman (1978), and Fienberg (1979). Of the types of graphics analyzed in those reviews, only graphics involving display of multidimensional data are of interest here.

An important semantic distinction must be established here between the dimensions of the statistical object (the quantitative data set with "n" variables), the dimensions of a conceptual representation of the object, and the dimensions of a physical representation that portrays one "facet" (one conceptual representation) of that statistical object. The overwhelming majority of physical representations are, not surprisingly, two-dimensional, like those displayed in a flat paper surface or on a computer screen, although three or even four-dimensional representations (interactive versions of existing holographic displays) offer tantalizing possibilities for the not so distant future. Bertin, however, in his synopsis of graphic constructions, presented earlier, addressed exclusively this type of physical representation, on the sensible grounds that there is where the simplest and widest means of diffusion of information really are.

Two-dimensional physical representations of quantitative data were chosen, accordingly, as the appropriate representation type for the purposes of this study and of DSSs in general. From now on, the discussion of multidimensional representations refer exclusively to conceptual representations of data.

Three particular multidimensional data representations should be mentioned here. The first one is actually a class of representations, labeled "hypergraphics" by Brisson (1978), consisting of an expansion of tri-dimensional, Cartesian coordinates to n-dimensions, with all data plotted in the same construct, where the hierarchy of coordinates (dimensions) determines what data subset will be plotted along the immediately higher dimension. In spite of the originality of the work, including its mathematical treatment, it is not intuitively intelligible above five or even four dimensions. Although there have been apparently no studies investigating the issue, it is possible that changing the order of dimensions to be plotted might result in quite different perceptions of the data. In spite of these real (and possible) shortcomings, it represents one of the most intriguing approaches to multidimensional data representations developed in the past twenty years.

The second graphic representation corresponds to Anderson's (1957) "glyphs" and "metroglyphs", consisting of a dot representing each data element and rays of various lengths, emanating from the dot, visually displaying the values for each variable. Best display results were associated with a maximum of seven variables. The third representation corresponds to Chernoff's (1972) "faces", consisting of a cartoon-like human face for each data element, with different facial features associated with different variables.

A major problem with this type of display, as pointed out by Izenman (1980), is that:

> ... the clusters produced by visual inspection of the above types of displays are not independent of the particular permutation of coordinates used to produce those displays. This point was demonstrated for the case of the faces display in an experimental setting by Chernoff and Rizvi ... It would certainly be desirable to have some type of permutation-invariant representation for clustering multivariate data, but this is probably very difficult to realize ...

In defense of the "subjectivity of polygons [a variation on Anderson's displays] or [Chernoff's] faces", Newton (1976) argued that:

> One could agree if hypothesis validation were the intent of their use. But discovery would seem to lever upon the risks inherent in the creativity and idiosyncrasies of our knowing minds. In the birth of ideas, one again might wish to affirm "vive la difference", and to assist this difference unbegrudgingly.

Most of the examples of multivariate graphics examined by Bertin (1981, 1983), since the late sixties, however, can be associated precisely with data permutation invariance. Referring to such graphics as "multiples" or "collections" (of similar constructs), Bertin essentially proposed the use of univariate or bivariate representations of data elements displayed collectively for the whole dataset. Chambers et al. (1983) proposed one particular type of such multiple representations (bivariate scatterplots with normalized ranges, ambiguously labeled "draftsman's" displays). By reducing the complexity of the graphic image ,for representing each data element, and combining those similar representations in a single, permutable display, both Bertin and Chambers et al. managed to achieve displays that are indeed permutation invariant and quite effective for data clustering analysis. Such graphic "multiples", it is assumed here, are the most appropriate, at least for the near future, for DSSs, where an ideal combination of data complexity and interactivity can be presently achieved.

2.3.2.b - Graphics for Communication

Most works dealing with graphics are, in fact, concerned more with communication and less with data-analytical graphics. A few authors well exemplify this category: Brinton (1914), in a comprehensive, pioneering handbook of graphics; Schmid and Schmid (1979) and Schmid (1983), more recent, comprehensive handbooks; and Tufte (1983), whose work is perhaps the most elegant and eloquent defense of economy of means to achieve maximum graphic communication. For a thorough, detailed analysis of graphics as an information processing, rather than communication, tool, one must return once more to Bertin.

2.3.2.c - Graphic Information Processing

Bertin (1980, 1981) posed two questions as "the basic test of the graph", corresponding to "two stages of perception: 1st: [sic] What are the [data] elements in question? 2nd: What are the relationships among those elements [?]" In the first stage of graphic perception, "external identification", all the verbal conventions and identifying information are interpreted to isolate unambiguously the abstract data set to be represented by the graphic image. This is precisely the point made by William Playfair, in 1786, in the Commercial and Political Atlas (as cited in Tufte, 1983), in which he introduced the first acknowledged statistical graphic, the bar chart.

The emphasis, from both Playfair and Bertin, is on legibility and clarity of explanations to help the observer with the decoding process; the need may be more acute for communication graphics (from author to other individuals) but remains important even in the case of one person (the data analyst). Economy of description does not mean eliminating essential information, it does imply eliminating any extraneous materials—text, symbols, even color—that might prove redundant, distracting, or misleading. This is the province of most works on communication graphics, such as Tufte's.

Bertin's second stage of graphic perception, "internal identification", is characterized by discovery of relationships, as the "true domain of graphics ... [that] utilizes only the relationships between signs [or] visual variations." Visual variables are discussed first, in this section.

Visual Variables: Bertin proposed a set of eight visual variables to transcribe the "relationships of resemblance, order, and proportion" at the core of graphic analysis. The first four were referred to as "variables of the image": the X and Y dimensions of the plane (i.e., a Cartesian grid of the two-dimensional physical representation of the data), the "size" (length or area) and the "value" (i.e., tonal gradation from white to black) of the graphic element. The other four, "differential variables", correspond to texture, color, orientation, and shape of graphic elements. Bertin's argument for so labeling them was that only nominal data (as opposed to ratio, interval, or even categorical scale data) can be adequately represented with the aid of those last four variables. Some experimental studies confirmed this dichotomy (McCleary, 1983; Cleveland and McGill, 1985).

Morse (1981) cogently encapsulated the issue of graphic variables:

> If a display is to be easy to read, the graphical encoding scheme ought to reflect the significant data relationships. The encoding of a scale of data values should result in a scale of perceived values that parallels the data scale. This is the principle of preservation of scale [emphasis supplied]. Stated more or less formally: the psychophysical effect generated by the display of data should preserve the relationships among the data.

According to the principle above, Bertin's argument could be restated as: quantitative data can be adequately represented only by the "size" and "value" variables of the image, as the only two that can effectively preserve the original data scale. Cleveland and McGill (1985) ranked "size", in particular, as the most efficient in scale preservation terms (positioning of elements along aligned or non-aligned identical scales). The particular graphic representation used in the present experiment—multiple bar charts—utilizes "size" to represent variations in data elements.

Relationships between Signs: According to Bertin, "information is a relationship" (1981, p.12) and to "understand is to reduce the complete set of relevant data to the groupings that the relations generate" (1980). Useful information would correspond not "to the general categories ... which enabled us to conceive the data table but, on the contrary, to the new groupings defined by the set of relationships constructed by the interplay of data" (1981, p.20). As Bertin also pointed out, "relationships can exist among [data] elements, subsets of [whole data] sets" (1981, p.12). The distinctions between the three levels are:

Elementary It corresponds, in any given data matrix, to the individual cell, containing a single datum. He argued that "information processing 'must struggle' against it [elementary level] ... [because] our memory cannot retain this multiplicity of elementary data. We must reduce this multiplicity, discover similar elements, group them, and class them. These are the prerequisites to understanding and deciding."

Intermediate comprises all relationships between data subsets.

Overall level The relationships between all dataset elements would be "the true objective" of information processing, "the level necessary for decision-making".

The three levels above, and their implicit hierarchy, served as the basis for categorization of "insights" produced by Subjects of the experiment in this study.

2.3.3 - COMPARISON BETWEEN THE GRAPHIC AND NUMERIC MODES

The remainder of this section is a limited review of experimental studies comparing graphic types of data representation to alternative representations (numeric, even verbal) of that data. Eight studies are reviewed here, three of them published before 1950, all the others after 1975. Only one of those studies dealt with a relatively large dataset (seventeen variables with forty four observations each). None of them are directly comparable to the present study because of different dependent and independent variables, particularly their emphasis on objective measurement of observer performance without distinction between decision-making or problem solving stages. None of those studies provided interactive modification of displays, although, in two cases, limited interactive simulation capabilities were provided.

In spite of the differences pointed out, all of those studies present at least one or two relevant factors that were useful for preparation of the experiment presented in this study. It is in this spirit—not that of a comprehensive review of the literature in the field—that the nine following studies are discussed.

The Washburne Study: Washburne (1927), in a pioneering study, investigated the effectiveness of various types of data presentation—in graphic, tabular, and textual form—on comprehension of quantitative datasets (very small, with three to four variables, with five observations each). The Subjects consisted of several thousand junior high school children, divided in eight groups. Several combinations of questions were asked from each group, involving recall of specific data and comparisons between data subsets. Summarizing the results of the experiments, Washburne reported that:

- "The simpler the visual pattern, and the fewer the data, the more specific the recall [this corresponds to Miller's (1956) 'span of absolute judgement', as one of the limitations of short-term memory]
- "The more complex the visual pattern, and the more numerous the data, the more general the recall (and apparently the more effective the making of static and dynamic comparisons) [i.e., relative amounts and fluctuations, or relative rates of change]
- "Increase in the number of data presented in a graph does not effect unfavorably the recall of specific amounts ...
- "The bar-graph is the form the most favorable to the recall of relative amounts (static comparisons) when the comparisons called for involve a fair degree of difficulty ... [partly validated by Cleveland and McGill, 1985]
- "The line graph is the form most favorable to the recall of relative increase, decrease, and fluctuation (dynamic comparisons).

- "The statistical table is the form most favorable to the recall of specific amounts.

The Carter Studies: The purpose of the first study (1947a) "was to compare the relative efficiency with which data presented in tables and graphs can be used." The concern of the researcher was directed at the needs of aircrew members (of that time) to "refer to different mathematical functions in performing their flight duties." Speed and accuracy of information retrieval were the dependent variables. The measure of speed was the number of problems completed in a given amount of time. Accuracy was measured in terms of the number of errors made, average magnitude of those errors, and average magnitude of those errors in terms of the number of problems attempted.

A table and a graph for each of four mathematical functions were used. Each table contained approximately 100 to 150 elements, arranged in a matrix with five columns. There were time limits ranging from 1.5 minutes (for questions involving recall of specific values) to 3 to 4 minutes (for questions involving simple and double interpolation questions). Carter concluded that:

> It [was] apparent that when the tables are entered with tabulated values we find the table superior to the graph in both speed and accuracy. On the other hand, whenever the table cannot be entered directly and some type of interpolation must be used, the data gathered from the graph is just as accurate and perhaps more accurate than that gathered from the table. With the graph a large number of errors are made but they are usually very small errors.
>
> It seems apparent that when relatively accurate answers are required and it is not possible to tabulate all the values that will be used, the material should be presented in graphic form. On the

> other hand, if speed is essential and some accuracy can be sacrificed, it seems better to have the material presented in tabular form and not to require any interpolation by allowing all entries to be made with the nearest tabulated argument.

The purpose of the second study (1947b) "was [also] to determine the relative effectiveness of different techniques for presenting numerical functions," restricted as well to tables and graphs, with speed and accuracy of information retrieval as the major concern. Among the four problems investigated, only the fourth one is of interest here: whether "it was better to use the best designed table [as determined by a previous part of the experiment] as compared to the best designed graph [also previously determined]."

Carter concluded that "within the limits of the material investigated, a table in which every point is given is as rapid, and more accurate to use, than [the graphic] method of presenting data." It is important to reemphasize that speed and accuracy—direct information retrieval—were the determinants of both studies, not preliminary data analysis of unfamiliar datasets.

The Minnesota Experiments: This series of experiments, reported by Dickson et al. (1977), "were conducted to examine the significance of various [management] information systems characteristics on decision activity ... By varying the manner in which information was provided to participants in each experiment, the impact of various [management] information systems characteristics and individual differences on decision effectiveness were investigated." The Minnesota researchers used one particular type of laboratory study—"man-machine" experimental gaming—characterized by interactive computer applications, imbedded mathematical simulation capabilities, and full experimental control over the decision making environment.

Nine experiments were conducted, based upon five different simulated decision-making environments (labeled "simulators"). The independent variables consisted of two basic types:

> (1) subjects' characteristics or attributes (e.g., psychological or experience measures); and (2) characteristics of the information system provided to the subjects (e.g., monitor display versus Batch [sic] paper output, various forms of output format [such as tabular versus graphic.]
>
> The dependent variables used to measure the decision making performance varied from ... measures of decision quality (e.g., cost of production over the range of the experiment) ... [to] attributes of decision performance [such as]: (1) the time taken to make the decisions; (2) the confidence placed in the decisions made; (3) the data selected (or reports used) to make decisions; and (4) the kind of decisions made (decision outcomes).

The present study, by contrast, utilized only the mode of data representation (or "output format", in their terminology) as its independent variable, and "insight" (idea) generation as its dependent variable.

From the nine experiments, of specific interest here was the one conducted by I. Benbasat. In the experiment—conducted in an inventory management setting—the Subjects were assigned five different treatments within a factorial study design framework. In terms of one of those treatments—tabular versus graphic output formats—the major findings were that "subjects receiving graphical output and decision aids had the lowest costs" and that "subjects receiving graphic output used the least reports." A general conclusion, related to all the nine experiments, stated that "CRT [cathode ray tube] system output [as opposed to various other forms of paper output] can lead to faster decisions and use substantially less data ... Graphic [paper] output format may have results similar to that of CRT systems and may even lead to 'better' decision making."

The Mezzich and Worthington Study: The experiment conducted by Mezzich and Worthington (1978) compared seven different representations of multivariate quantitative data (relative to psychiatric diagnostic classification). The dataset in question contained seventeen variables and forty four observations, corresponding to four major categories of psychiatric patients. This is the largest of all datasets examined by the studies related in this section, and the only one comparable to the dataset used for the present research.

The two basic objectives of the study were: "1. Assessment of the comparative difficulty experienced by human judges in using the graphical methods for grouping the archetypal patients into [the four original groups; and] 2. Assessment of the actual performance of these judges in using the graphical methods to recover the four diagnostic groups built into the data." The seven graphic representations used were:

- Linear profile (multiple line graphs)

- Circular profile (a variation of Anderson's "metroglyphs", with equally spaced rays)
- Chernoff's faces
- Linear Fourier representations
- Polar Fourier representations
- Factor scores in two dimensions, directly plotted
- Ordinal multidimensional scaling (reduced to a two dimensional solution space).

Thirteen judges (experienced psychiatrists) attempted to sort out the information into the four original groups underlying the data. Dimensional reduction to representations as a point in the Cartesian plan (the factorial and the ordinal multidimensional techniques) were generally the most effective, although for the best performers the best approach consisted of polar Fourier representations, and the low scorers had better results with Chernoff faces.

The Lusk and Kersnick Study: This study (1979) sought, among other objectives, information on two Management Information Systems design questions, namely, how psychological character of individuals and report formats (the two independent variables) affect task performance (the dependent variable). Also analyzed, as a secondary dependent variable, was the relative "desirability" of reports, defined in terms of perceived complexity ("the level of difficulty anticipated in using a report to provide specific information.") No learning instructions of any kind were given to users of the different report types.

Task performance was measured in terms of correct answers to twenty questions involving development of only discrete values, "rather than to perceive a trend or recognize a relationship." (In other words, efficiency, not effectiveness in information retrieval, was the major concern.) There were 219 Subjects in the experiment. A time limit was imposed, to allow the "average" Subject to complete about 75 percent of the questions. The problem dataset consisted of four variables with nine observations each.

Psychological individual type was characterized in terms of a cognitive style dichotomy, "low/high analytical", involving the ability to isolate relationships from a given context (cognitive differentiation). Report formats consisted of five types:

- Tabular raw data
- Tabular percentage relationship of the raw data
- Frequency histogram of raw data
- Cumulative frequencies graph of raw data
- Cumulative frequencies graph of percentage data

The null hypothesis that the "high analytic" group would not have higher task performance scores that the "low analytic" group, controlling for report type, was rejected for both the "tabular" and "graphic" groups (at the 0.01 level) but not for representations of transformed (percentages) data. The tabular reports were considered less complex than the graphic ones, with task performance scores inversely related to perceived report complexity. These results are consistent with those from Carter

(1947a, b), inasmuch as speed and accuracy of information retrieval are considered the main determinants of performance.

The Lucas Study: In this study, Lucas (1981) investigated the impact of computer-based graphics on decision-making. The independent variables used were type of terminal output (monitor displays versus hardcopy terminals, graphic versus numeric output) and decision style (a cognitive style dichotomy, "analytic/heuristic", characterizing the individual predominant approach to problem solving.) The dependent variables included task performance on a simulation exercise, self reports of information usefulness, and objective tests of problem understanding (only discrete values were asked for). The outputs for the simulation problem included one or two variables (in the case of time series) with twenty observations. There were 119 Subjects in all.

Three of the four hypotheses tested are of interest here. The first one predicted that groups receiving graphic output would score higher (in terms of all three dependent variables) than the "tabular" groups (both using monitor displays). According to Lucas:

> ... the hypothesis [was] not confirmed for performance or the usefulness of information. However, the evidence [suggested] that the graphics groups developed a better understanding of the problem. Some of the evidence, though, is for two groups receiving probability graphs on a test designed to measure the understanding of inventory, not probabilities. One group receiving graphical output on the simulation also did better in the probability test. Thus, these results must be treated with caution.

The next hypothesis predicted that "graphic subjects receiving both graphical and tabular output will have higher scores than graphics output receiving only graphical output." There was limited support for this hypothesis; scores were higher for the graphical/tabular output group only in the inventory test (specific information retrieval).

The third hypothesis was that the "heuristic decisionmakers [would] have higher scores for graphical treatments and analytic decisionmakers [would] have higher scores for tabular treatments" (both groups only with monitor display output). The first group "had the best (lowest cost) simulation results ... [whereas there was] no difference between treatments for the analytics." For inventory understanding, the "graphics" groups had the highest scores, the differences being greatest for "analytics", who, contrary to the hypothesis, had the highest scores.

Lucas speculated that the different response of the "heuristic" and "analytic" groups to graphic treatments might be due to "analytics" having:

> ... a model in mind when beginning the exercise. If this is true, the mode of presentation would be relatively less important to them because they used output data to confirm or disprove their prior model.
>
> Heuristics, on the other hand, performed better under graphics treatments where they could see a picture of the data. Heuristics may have relied on the displays more due to their lack of an <u>a priori</u> model.

The Powers, Lashley, Sanchez and Schneiderman Study: This experiment (1984) was "designed to test the hypothesis that more usable information can be conveyed using a combination of graphical and tabular data than by using either form alone." It was expected that, in a sense, one form of data representation might verify the other with which the user is most familiar. Two independent variables were used: memory (recall and nonrecall situations) and presentation format (tabular, graphic, or combination).

The dependent variable was comprehension performance, measured with a multiple choice test with three types of questions: (1) simple recall or "look-up" of specific data observations; (2) simple recall or "look-up" of several observations, with comparison of those elements; and (3) extensive recall or "look-up" of observations "with a comparison and/or arithmetic manipulation of the data recalled." The measurement of the dependent variable was operationalized in seven different ways, ranging from the number of questions correctly answered (total and by question type) to percentage of questions answered correctly of all those attempted. Five minutes were allocated for familiarization with the data, and ten for answering the questions. The problem dataset contained five variables, one with twenty observations, the others with only five each. There were seventy four Subjects in the experiment.

Subjects in the "tabular" group performed better, overall, than their counterparts in the "graphical" and "combination" groups. Subjects in the "nonrecall" group (who could refer to the original data while answering the questionnaire) also performed better than their counterparts in the "recall" group. Controlling for memory, there was a slight advantage to the "graphical/tabular" group compared to the "graphical" group, more strongly in terms of accuracy, and less so in terms of speed of performance (best handled, as other experiments have also indicated, by numeric tabular representations alone). Powers et al. attributed the differences in presentation format group performances to several factors:

- The "tabular format was probably a more familiar and natural form of data presentation for our subjects [than the] pie chart [and bar chart forms]."
- Subjects in the "graphics/tabular combination group [may have been] overwhelmed by the amount of data presented by them."
- The "design and format of the experiment was geared towards the Subjects in the tabular group [which] received data in much the same format as that in which they were required to recall it ... The graphic treatment gave the subjects a good overall 'view' of the data, but did not easily supply facts to the Subjects. The questions demanded recall of specific facts provided by the materials presented, therefore the questions were geared towards the tabular treatment."

2.3.4 - THE PRESENT STUDY

One of the basic premises of this study, which found empirical support in some of the studies reviewed above, is that graphics can induce a more holistic view of quantitative data than numeric representations, which are more appropriate for specific data element retrieval. The focus of the study is on exploratory statistical analysis of unfamiliar datasets, not as a substitute for more complex mathematical analyses but as a tool to increase the effectiveness of the whole data analysis process, by creative exploration of new and—it is expected—more cogent lines of inquiry. In that context, computer-based personal support systems, facilitating an ad hoc approach to problem solving, are assumed to be the best tool presently available for gaining "insights" into the data.

A major goal of this research is to contribute to the search for methods to increase the overall effectiveness of professional performance. For this reason, none of potential dependent variables dealing with task efficiency (such as speed and accuracy of information retrieval) were selected for this study. By emphasizing an open-ended approach to an information-rich, purposefully novel, unstructured, and ambiguous problem space (all from a Subject's point of view), it was simply natural to focus on the ideas ("insights") generated as the appropriate dependent variable to study effectiveness—not efficiency—issues in quantitative data analysis.

Another restriction associated with the dependent variable chosen ("insight" generation) is that only the early stages of problem analysis were considered in the research. Those stages include familiarization with the problem, incubation of the first ideas associated with the analysis, and first concrete manifestations ("distillation") of the understanding of the problem space.

The independent variable chosen, "data representation mode", consists, in cognitive terms, of two visual symbolic submodalities—numeric and graphic. For information processing purposes, the graphic mode was assumed to be potentially more powerful than the numeric mode, in virtue of its data recoding capabilities—an essential factor in the increase of amount of information an individual can process. No mixed graphic-numeric format was included due to little empirical support for differentiation between purely graphic and combination formats (Powers et al., 1984).

Only one type of graphic representation—multiple bar chart, a relational graphic matrix—was used, given the exploratory nature of the study. Although other types of multivariate representation of quantitative data are available, bar charts were chosen primarily because its identifying visual variable "size" (more specifically, "length") is arguably the best for representation of quantitative data, i.e., for preservation of a close relationship between the perceived scale and the underlying scale of the original, abstract set of numbers (Bertin, 1981).

Both data representation modes were displayed in computer monitors, with no printed output of any kind other than supporting, textual information. The interactiveness of a DSS fully implies a higher degree of control over data representations—and much faster response time—than any kind of printed materials can offer for exploratory analysis of large quantitative datasets. It is quite possible, however, that instantaneous "snapshots" of DSS data manipulations, created by the user, might be useful for many users as an extended documentation of their "exploratory" efforts (Carlson, 1983). For several reasons, that option could not be explored during the present study; however, it is assumed here that it is a better alternative format to monitor displays that the fixed, printed output format employed in several of the studies reviewed above.

The problem dataset used for the present experiment was assumed to be representative of the type and size of datasets most common in a professional context. The quantitative dataset contains 20 variables with 24 observations each (a total of 480 elements). Very few comparative studies of graphic versus numeric mode differences seem to have employed datasets with more than 5 variables or a total of 100 elements, or both.

In summary, the present study proposed to determine, in the context of preliminary data analysis, whether one can generate more, and more complex, "insights" by looking at a graphic (multiple bar chart) representation—as opposed to a numeric table—of a large, multivariate quantitative dataset, displayed and manipulated interactively in a personal computer-based system.

The next chapter presents a discussion of the research framework and study design used, as well as of possible threats to study validity.

3 - RESEARCH FRAMEWORK AND METHODOLOGY

3.1 - RESEARCH FRAMEWORK

3.1.1 - DATA CHARACTERISTICS

3.1.1.a - Data Type

The data of this research are of the primary kind—directly associated with the research—and generated experimentally during the study. The nature of the research data is uniform, consisting exclusively of notes written by Subjects during experimental sessions.

3.1.1.b - Data Needed

The research problem was to determine what are the effects, if any, of having Subjects look at graphic, as opposed to numeric, interactive computer displays of quantitative relational data. Those effects were measured in terms of a discrete pattern of responses for "insight" complexity scores.

The data needed for solving the research problem are:

- The notes written by Subjects with lists of "insights" (the dependent variable) reached while examining the paradigm problem;

- the experimental group (graphic or numeric mode of data presentation) to which each Subject was randomly assigned (the independent variable);

- the time categories (six consecutive intervals of 10 minutes) under which "insights" were produced, to examine rates of generation of ideas ("ideational fluency").

3.1.1.c - Means of Obtaining the Data

All the data for each Subject was gathered during a single session. Detailed information is presented in the next chapter about data collection, including the session protocol, the experimental paradigm problem, and the interactive computer program for manipulation of the experimental paradigm problem. The procedures for data collection conform to Virginia Tech norms for research involving human subjects and were exempted from review by the University's Institutional Review Board. Appendix A.1 ("Review Exemption") exhibits the Form submitted for certification of exemption and the statement on specific procedures to ensure Subject data anonymity.

3.1.2 - CLASSIFICATION OF THE STUDY EFFORT

The classification scheme suggested by Campbell (1957) was used here for characterization of the study effort. Within that framework, the present study used the experimental method for data gathering called "Posttest-Only Control Group Design".

(a)	A	X	O_1	or	(b)	A	X_t	O_1
	A		O_2			A	X_c	O_2

Figure 4. Experimental Research Method: Format of experimental design.

The format of the experimental method used is shown in Figure 4. The two diagrams in that figure are equivalent. Symbol A indicates that groups were made equivalent by random sampling assignment at a specific point in time prior to the experimental treatment. The treatment is represented by X, in diagram (a), or by Xt in diagram (b). The lack of treatment (or "control" treatment) is indicated by a blank, in diagram (a), or by Xc, in diagram (b). The two groups from which measurements were taken are O1 and O2, standing respectively for the graphic (treatment) and numeric (control) groups.

The present study was set to investigate the quantitative and qualitative differences, if any, in the written responses of two groups requested to generate "insights" about the same quantitative dataset. One group was assigned to an active treatment (use of a graphic display of the experimental paradigm problem); the other group was assigned to a control treatment (use of a numeric display of the paradigm problem). The experimental units (the "Subjects") consisted of students (at the undergraduate and graduate levels) at Virginia Tech.

The main interest of this study was to find the statistical significance and the strength of the relationship between "insight generation" (the dependent variable) and "data display mode" (the independent variable), as well as the magnitude of that relationship. It was expected, based on preliminary results from a pilot study conducted in July 1985, that there would be primarily a qualitative, rather than quantitative, difference in "insight generation" between the two groups, indicated mainly by "complexity" levels.

3.1.3 - MEASUREMENT OF THE DEPENDENT VARIABLE

The dependent variable "insight production" was measured in four different ways, reflecting different classes of "insight" "complexity" (see Figure 5 on page 80).

I_1 Raw count of "insights" ignoring complexity levels

I_2 Count of multiple-field "insights", exclusive of single-field "insights"

I_3 Count of multiple "field-group" "insights", exclusive of levels "A" through "F"

I_4 Count of the number of different complexity levels

Figure 5. "Insight" Production: Operationalization of the dependent variable for the research question.

The classification of ideas, or "insights", in the present study, is dependent on the number of fields (i.e., variables of the paradigm problem dataset) and number of field groups (set of fields conceptually related)[1] mentioned in any one given "insight". The major distinctions that could safely be established, given the exploratory nature of the overall study, are dichotomies in the number of fields (F) and field groups (G) mentioned in any given "insight". The first dichotomy is characterized by the categories "Single Field" (sum of $F=1$) and "Multiple Fields" (sum of $F>1$). The second dichotomy consists of "Single Field-Group" (sum of $G=1$) and "Multiple Field-Groups" (sum of $G>1$). A more detailed explanation of the complexity levels used is presented in Chapter 5 ("The Treatment of the Raw Data"). Other schema may be possible, but for an exploratory

[1] The field-group "Age", for instance, is a set that encompasses all the pre-determined age categories presented in the paradigm problem.

classification scheme, only references to the number of fields of the dataset under study were deemed unambiguous, objective, and measurable.

3.1.4 - RANDOMIZATION OF THE INDEPENDENT VARIABLE

The Subjects who volunteered for participation in the study (a total, arbitrarily pre-determined, of eighty individuals) were already assigned, by simple random sampling, to either the treatment group (Xt), which used the "graphic" mode of data display, or the non-treatment, or control, group (Xc), which used the "numeric" mode of data display (see also Section 3.1.6, "The Population Under Study", for a discussion on population sampling).

A table of random numbers (Blalock, 1972, Table B, p.554) has been used to assure equal probabilities of Subject participation in the treatment group. Appendix A.2 ("Random Sampling Procedures") presents in detail the steps taken for determination of treatment-group membership.

3.1.5 - OPERATIONAL HYPOTHESES

Based on the measurement of the dependent variable, presented above, the study hypothesis was subdivided into four related but independent operational hypotheses, to test each of the four classes of complexity of "insight" generation. Each operational hypothesis is first presented in terms of a directional hypothesis; its corresponding null hypothesis, presented immediately following it, was used to test the statistical significance of the relationship between the dependent variable (as measured for the particular hypothesis) and the independent variable (common to all the four operational hypotheses).

3.1.5.a - The First Hypothesis (Overall "Insight" Scores)

H1 The sum of the ranks of "insight" scores generated by Subjects looking at a graphic display of the "paradigm problem" data would be significantly higher than the sum of ranks of "insight" scores generated by Subjects looking at a numeric display of the same data, i.e., there would be a positive treatment effect ($\Delta > 0$).

$H1_0$ There would be no significant differences between the sum of the ranks of "insight" scores generated by Subjects looking at the graphic display of the "paradigm problem" data and the sum of the ranks of "insight" scores generated by Subjects looking at the numeric display of the same data, i.e., there would be no treatment effect ($\Delta = 0$).

3.1.5.b - The Second Hypothesis ("Multiple Fields" Scores)

H2 The sum of the ranks of "Multiple-Fields" "insight" scores generated by Subjects looking at a graphic display of the "paradigm problem" data would be significantly higher than the sum of ranks of "Multiple-Fields" "insight" scores generated by Subjects looking at a numeric display of the same data, i.e., there would be a positive treatment effect ($\Delta > 0$).

$H2_0$ There would be no significant differences between the sum of the ranks of "Multiple-Fields" "insight" scores generated by Subjects looking at the graphic display of the "paradigm problem" data and the sum of the ranks of "Multiple-Fields" "insight" scores generated by Subjects looking at the numeric display of the same data, i.e., there would be no treatment effect ($\Delta = 0$).

3.1.5.c - The Third Hypothesis ("Multiple Field-Groups" Scores)

H3 The sum of the ranks of "Multiple Field-Groups" "insight" scores generated by Subjects looking at a graphic display of the "paradigm problem" data would be significantly higher than the sum of ranks of "Multiple Field-Groups" "insight" scores generated by Subjects looking at a numeric display of the same data, i.e., there would be a positive treatment effect ($\Delta > 0$).

$H3_0$ There would be no significant differences between the sum of the ranks of "Multiple Field-Groups" "insight" scores generated by Subjects looking at the graphic display of the "paradigm problem" data and the sum of the ranks of "Multiple Field-Groups" "insight" scores generated by Subjects looking at the numeric display of the same data, i.e., there would be no treatment effect ($\Delta = 0$).

3.1.5.d - The Fourth Hypothesis (Different Complexity Levels)

H4 The sum of the ranks of the total number of "insight" levels generated by Subjects looking at a graphic display of the "paradigm problem" data would be significantly higher than the sum of ranks of the total number of "insight" levels generated by Subjects looking at a numeric display of the same data, i.e., there would be a positive treatment effect ($\Delta > 0$).

$H4_0$ There would be no significant differences between the sum of the ranks of the total number of "insight" levels generated by Subjects looking at the graphic display of the "paradigm problem" data and the sum of the ranks of the total number of "insight" levels generated by Subjects looking at the numeric display of the same data, i.e., there would be no treatment effect ($\Delta = 0$).

3.1.6 - THE POPULATION UNDER STUDY

3.1.6.a - The Reference Population

The reference population, or the group at large whose needs represent the focus of attention of the present study, consists of professionals, managers, undergraduate and graduate students, and all those using (or potentially using) computers as a means of processing quantitative data, in the form of spreadsheets and relational databases, and particularly taking advantage of the graphics capabilities of machine-generated displays.

3.1.6.b - The Target Population

It was obviously impractical to sample the reference population mentioned above, not just because of sheer numbers but also due to its recognizably imprecise boundaries. The population used, by assumption, as a representative subset of the reference population consists of students (undergraduate and graduate) at Virginia Tech. Two issues need to be raised, in conjunction with sampling of this reference population. First, the total population of University students (above 22,000) could not be adequately sampled, given limitations in accessibility to unbiased, updated student records. Second, the overall research effort was basically exploratory, where questions about the existence (or not) of the phenomenon under examination took precedence over questions about its extent.

To determine the total number of individuals for participation in the study, two basic factors were taken into account: sensitivity of test statistics to sample size, and resource constraints (particularly time) for research. A total of eighty Subjects was considered adequate for the overall target population for this study, evenly divided between the treatment and non-treatment (control) groups.

University students were invited to participate voluntarily in the study—and be remunerated for it—with the first eighty volunteers to complete the experimental sessions constituting the effective population of the study. The selection of Subjects for participation was made on a "first-come" basis, until the pre-determined total of Subjects completing the sessions was attained.

3.1.6.d - Subject Recruitment Process

A special effort was made to ensure that all student groups at Virginia Tech were equally approached or informed of the research, thorough notices posted at each department or program at the University as well as at major concentration points of students on campus: McBryde Hall, Squires Student Center, and undergraduate dining halls.

Each prospective Subject was given a "Subject Participation" handout describing the study in very brief terms and the requirements for the Subject's participation. Upon acceptance of the conditions and limitations described in the handout, a "Scheduling Participation Sheet" and a release form were respectively filled in and signed by the Subject.

Appendix A.3 ("Recruitment Materials") exhibits a copy of the "Subject Participation" handout, "Scheduling Participation Sheet", "Release Form", recruitment notice, as well as a list and map of locations where recruitment notices were placed.

3.2 - STUDY DESIGN

3.2.1 - MODEL FOR NULL HYPOTHESIS TESTING

The model used for studying the differences between two independent random samples (Hollander and Wolfe, 1973, p.67,) with an ordinal scale of measurement, is presented in Figure 6 on page 87.

It was also assumed that, with regard to the model, each of the N e's comes from the same continuous population, and that they are mutually independent.

$$X_i = e_i, \qquad i = 1, \ldots, m$$

and

$$Y_j = e_{m+j} + \Delta, \qquad j = 1, \ldots, n$$

Where:

Xi and Yj	=	N = m + n observations
$e_{m+1}, \ldots, e_{m+n}$	=	unobservable random variables
Δ	=	(unknown) shift in location due to treatment

Figure 6. Null Hypothesis Test Model

3.2.2 - TEST STATISTICS

The statistics used for analysis of the relationship between the dependent and the independent variables are:

- Measure of statistical significance of the relationship:

  Distribution-free rank sum test (Wilcoxon, or Mann-Whitney, test)

- Measure of strength of association of the two variables:

  Cramer's "V" ("V-squared")

- Measure of the magnitude of the relationship:

  Difference of means

Given the data types of the dependent and independent variables (ordinal and nominal scales, respectively) and the nature of the relationship sought (the differences between two samples), the Wilcoxon rank-sum test is not just an appropriate test but one of the most researched and well-established nonparametric tests for measuring statistical significance involving ordinal data and even interval data. The major reasons for choosing a nonparametric test were the conditions pointed out by Blalock (1972, p.244): (1) "we cannot legitimately use an interval scale but ordering of scores is justified, or (2) the sample is small and normality cannot be assumed."

The results of the statistical significance test are important for this study because of the small sample size for the two groups ($N1 = N2$; $N1 \leq 40$). However, an additional statistic to measure the strength of the relationship between the two variables was in order, since statistical significance is normally a necessary but not sufficient condition to indicate the overall significance, i.e., the true meaning of the relationship between variables.

The measure known as Cramer's "V" (or "V-squared") was used to measure the strength of the association between the two variables. It was selected, from several other equivalent measures in the literature, since it can attain unity ($0 \leq V^2 \leq 1$) and therefore its result is intuitively easier to interpret than other measures (Blalock, 1972, p. 297). For purposes of the "V-squared" test, the dependent variable "insight generation" was reduced to a nominal, or categorical, scale, to enable its use in the contingency tables for the test. To measure the impact, or magnitude, of the differences between the two samples, the measure used was the difference in their mean scores for "insight" generation.

3.2.3 - SIGNIFICANCE LEVEL AND CRITICAL REGION

Since there was some experimental evidence that a positive difference between "insight" scores for the treatment and control groups could be expected, a directional (one-tail) test was used for all hypotheses. Among major criteria for selecting the significance level for hypothesis testing (Labovitz, 1970), the following were directly relevant for testing of the operational hypotheses:

Power of Test - Sample Size (N): "The power of a test varies directly with sample size"; since with a small N "even large differences may not reach the predetermined level", large error rates (.10 or .05) should be used.

Degree of Control in Design: The larger the amount of control to reduce alternative interpretations of an experiment, the larger the level of significance that can be tolerated.

One-tail vs. Two-tail Tests: Since there is a probability "that we have some idea of the direction of [a] hypothesis, but [with] a small to large amount of uncertainty in our reasoning, ... we should neither accept the z-score equivalent of the one-tail or two-tail test ... but an intermediate score between the two values. ... This is the equivalent of saying we should choose a larger or smaller error rate depending upon our degree of confidence in the direction of our hypothesis."

Testing vs. Developing Hypotheses: In the case of exploration of "a set of interrelations for the purpose of developing hypotheses to be tested in another study, a larger error rate will tend to yield more hypotheses" for subsequent validation; at the "exploration stage, the .10 or .20 level would be sufficient."

The following significance levels were used for testing the statistical significance of the operational hypotheses, taking in consideration the criteria stated above and the different complexity levels associated with "insight" generation for each hypothesis:

$H1_0$	.10
$H2_0$	.05
$H3_0$	.025
$H4_0$	.025

3.2.4 - COMPUTATION OF TEST STATISTICS

All the statistical formulas in this section are presented in two ways: (a) conventional statistical notation and (b) Iverson's mathematical notation (1976). The actual computation of all the measures in question was made through an implementation of Iverson's notation in a computing environment, APL2 from IBM (IBM, 1982; Gilman and Rose, 1984).

3.2.4.a - Significance of Relationship

The measures presented in this sub-section are discussed in Hollander and Wolfe (1973, Chap. 4).

Wilcoxon Rank-Sum Test: Figure 7 presents the formula for computation of the Wilcoxon rank-sum statistic, applicable mainly to small samples, and used as part of a corrected Wilcoxon rank-sum statistic for large samples.

(a)

$$W = \sum_{j=1}^{n} R_j$$

Where:

W = Wilcoxon rank-sum test
R_j = Ranks assigned to the Y's (treatment group observations)

(b)

W: +/R

Where:

W ↔ Wilcoxon rank-sum test
R ↔ Ranks assigned to elements of data vector Y

Figure 7. Wilcoxon Rank-Sum Test: Formula for small samples.

Corrected Wilcoxon Rank-Sum Test: Figure 8 on page 93 presents the formula for computation of the Wilcoxon rank-sum statistic, corrected for large sample sizes ($N > 20$).

Wilcoxon Test Compensation for Ties: Figure 9 on page 94 presents the formula for the variance of the Wilcoxon rank-sum test, compensating for ties. In case of ties, the regular **W** statistic was computed with average ranks, i.e., the arithmetic mean of the ranks of groups of tied scores.

Decision on Statistical Significance: The decision to reject or accept the null hypotheses at the stated levels of significance, using the corrected Wilcoxon rank-sum test (with a normal approximation) was made as follows:

Reject H_0	if	$W^* \geq z(\alpha)$
Accept H_0	if	$W^* < z(\alpha)$

The table used for determination of correspondence between significance levels (area under the normal curve) and z standard-deviation units from the mean is in Blalock (1972, p. 558).

(a)

$$W^* = \frac{W - E_0(W)}{[var_0(W)]^{1/2}} = \frac{W - [n(m+n+1)/2]}{[mn(m+n+1)/12]^{1/2}}$$

Where:

W^*	=	Wilcoxon rank-sum test (large sample approximation)
W	=	Wilcoxon rank-sum test
$E_0(W)$	=	Mean value of Wilcoxon rank-sum test
$var_0(W)$	=	Variance of Wilcoxon rank-sum test
n	=	Number of observations for treatment group (Y)
m	=	Number of observations for control group (X)

(b)

```
WLSA:  (W - WE)÷WVAR*.5
  WE:   N×.5×M+N+1
WVAR:  (M+N+1)×M×N÷12
   M ←  ρY
   N ←  ρX
```

Where:

WLSA	↔	Wilcoxon rank-sum test (large sample approximation)
W	↔	Wilcoxon rank-sum test
WE	↔	Mean value of Wilcoxon rank-sum test
WVAR	↔	Variance of Wilcoxon rank-sum test
N	↔	Number of elements of vector Y
M	↔	Number of elements of vector X
Y	↔	Data vector with observations from treatment group
X	↔	Data vector with observations from control group

Figure 8. Corrected Wilcoxon Rank-Sum Test: Large-sample formula

(a)

$$\mathrm{var}_e(W) = \frac{mn}{12}\left[m+n+1 - \frac{\sum_{j=1}^{g} t_j(t_j^2-1)}{(m+n)(m+n-1)}\right]$$

Where:

$\mathrm{var}_e(W)$	=	Variance of Wilcoxon rank-sum test (with ties)
g	=	Number of tied groups
t_j	=	Size of tied group j
n	=	Number of observations for treatment group (Y)
m	=	Number of observations for control group (X)

(b)

```
WVART: (M×N÷12)×(M+N+1)-(+/T×((T*2)-1)÷(M+N)×(M+N-1)
T   ←   +/((0=Y=¯1⌽Y)/Y)∘.=Y[⍋Y]
```

Where:

WVART	↔	Variance of Wilcoxon rank-sum test (with ties)
T	↔	Vector with sizes of tied groups
N	↔	Number of elements of vector Y
M	↔	Number of elements of vector X
Y	↔	Data vector (treatment group)
X	↔	Data vector (control group)

Figure 9. Variance of Wilcoxon Test: Compensation for tied scores.

3.2.4.b - Strength of Association

The measures presented in this sub-section are discussed in Blalock (1972, Chap. 15).

Contingency Tables: Figure 10 presents the contingency tables for computation of the chi-square statistic, used for calculation of the measure of strength of association. The categories in those tables were revised, as expected, since the range of the number of "insights" produced (and their frequency, within each category) proved to be substantially different from pilot-test data. The need for revisions arose from efficiency problems associated with the chi-square statistic, which needs at least five observations in each contingency table cell.

(H1)

I	Y	X
1- 5		
6-10		
11-15		
16-20		
21-25		

(H2)

I	Y	X
1- 4		
5- 8		
9-12		
13-16		
17-20		

(H3)

I	Y	X
1-3		
4-6		
7-9		
10-12		

(H4)

I	Y	X
1-2		
3-4		
5-6		
7-8		

Where: H1, ..., H4 = Operational hypotheses
I = "Insight"-generation categories (unit: "insight")
Y = Treatment group (graphic mode)
X = Control group (numeric mode)

Figure 10. Contingency Tables

Chi-Square: Figure 11 presents the formula for computation of chi-square, needed for calculation of the measure of strength of association.

(a)

$$\chi^2 = \Sigma \frac{(fo - fe)^2}{fe}$$

Where:

χ^2	=	Chi-square test
fo	=	Observed frequency for each cell
fe	=	Expected frequency for each cell

(b)

```
CHI:  +/,((FO - FE)*2)÷FE
FE    ← ((+/FO)∘.×(+⌿FO))÷(+/,FO)
```

Where:

CHI	↔	Chi-square test
FO	↔	Contingency table with observed frequencies
FE	↔	Contingency table with expected frequencies

Figure 11. Chi-Square Formula

Cramer's "V": Figure 12 presents the formula for computation of Cramer's "V", the measure of strength of association between the variables "insight" generation and data presentation mode.

(a)

$$V^2 = \frac{x2}{N.Min(r-1,c-1)}$$

Where:

V^2	=	Cramer's "V" test
χ^2	=	Chi-square test
N	=	Pooled frequency of samples
r	=	Number of rows of contingency table
c	=	Number of columns of contingency table
Min(r-1,c-1)	=	Either r-1 or c-1, whichever is smaller

(b)

```
VCRAMER:  CHI÷N×R|C
R   ← 1↑ρFO
C   ← 1↓ρFO
```

Where:

VCRAMER	↔	Cramer's "V" test
CHI	↔	Chi-square test
N	↔	Total frequency of contingency table FO
R	↔	Number of rows of contingency table FO
C	↔	Number of columns of contingency table FO
FO	↔	Contingency table with observed frequencies

Figure 12. Cramer's "V" Formula

3.2.4.c - Magnitude of Relationship

Difference of Means: Figure 13 on page 97 presents the formula for computation of the difference between the means of the treatment and the control groups for the variable "insight" generation.

(a)

$$\Delta\bar{X} = \bar{X}_1 - \bar{X}_2$$

Where:

ΔX = Difference of sample means
$\bar{X}_1$ = Arithmetic mean of treatment group scores
$\bar{X}_2$ = Arithmetic mean of control group scores

(b)

ΔX: XM1 - XM2

Where:

ΔX ↔ Difference of sample means
XM1 ↔ Arithmetic mean of data vector Y (treatment group)
XM2 ↔ Arithmetic mean of data vector X (control group)

Figure 13. Difference-of-Means Formula

Proportion of Difference: Figure 14 presents the formula for computation of the proportion between the difference of means and the sample mean for the control group, as another measure to indicate the magnitude of the relationship.

(a)

$$PX = \frac{\Delta\bar{X}}{\bar{X}_2}$$

Where:

$P\bar{X}$ = Proportion of difference-of-means to control group mean
$\Delta\bar{X}$ = Difference of sample means
$\bar{X}_2$ = Arithmetic mean of control group scores

(b)

PX: ΔX÷XM2

Where:

PX ↔ Proportion of difference-of-means to control group mean
ΔX ↔ Difference of sample means
XM2 ↔ Arithmetic mean of data vector X (control group)

Figure 14. Proportion Difference Formula

3.2.5 - DECISIONS AND REPORTING OF RESULTS

For the statistical significance test, a decision to reject or not to reject each of the null hypotheses is stated in the summary of results presented in Figure 18 on page 134 from Chapter 6 ("Presentation of the Results"). The measurement of the strength of association and the measurement of the magnitude of the relationship between the dependent and the independent variables is presented in Figure 19 on page 134, in the same chapter.

3.3 - THREATS TO STUDY VALIDITY

Possible threats to validity of experiments have been extensively investigated in the literature (Manheim, 1977; Drew, 1980; Holt, 1982). Some of the threats that might specifically affect the present study, as well as measures taken against them, are listed in this section.

3.3.1 - INTERNAL VALIDITY

3.3.1.a - Maturation

Maturation "refers to factors that influence Subjects' performance because of time passing rather than specific incidents." For the present study, the time frame in question was the duration of each experimental session, scheduled to take one and a half hours, and the specific factors that might have been at work are hunger and fatigue. The duration of the experiment was relatively too short for consideration of hunger as a valid issue.

In terms of the fatigue factor, only Subjects obviously drowsy or sleepy were to be screened out at the beginning of experimental sessions. None were screened out; all Subjects were assumed to be moderately rested. Since all the sessions took place at times ranging from 10:00 to 22:00 hours, it is possible that some Subjects were not at the best of their possible concentration at their specific session time; however, this factor was assumed to be controlled for both groups under the random Subject assignment procedures.

3.3.1.b - Testing or Test Practice

This was not a specific threat to the present study, since there was no pretest per se (i.e., measurement of the prior condition of the dependent variable). Rather, "test practice", interpreted here as the accumulation of skills in handling one particular type of problem-solving situation, was exactly what the first part (tutorial) of the experimental sessions tried to ensure. The use of spreadsheets and database manipulation systems involve a minimum amount of skills before the user can actually start concentrating in getting results rather than in learning procedural rules. The "internalization" of such rules, ensuring for all Subjects of this study a minimum level of proficiency in the use of a data manipulation system, was the objective of the "tutorial" section designed for the experimental session (presented in more detail in Chapter 4, "The Data Collection").

3.3.1.c - Instrumentation

Instrumentation problems, in this particular case, have to do essentially with the treatment of raw data (particularly the scoring of "insights"). Special attention was paid to the procedures for treatment of raw data, in order to minimize any subjective interpretations or manipulations of the data. The different stages of raw data manipulation are presented at the end of Chapter 5 ("The Treatment of the Raw Data"). A waiting period of one week was observed before the beginning of the verbatim transcription of the notes written by a group of Subjects within a given week. Another week passed before the two next stages of raw data analysis (parsing and compression of sentences, and "insight" reordering) took place, immediately followed by the scoring and categorization of "insights". It was assumed that these two weeks would be enough to eliminate any potential interpretation bias towards any subset of the two groups, particularly since no group (treatment vs. non-treatment) identification was present in the worksheets filled out by Subjects and derived material. The final results for both samples were aggregated only after final compilation of individual Subject results.

3.3.1.d - Hawthorne Effect

Any changes "in the sensitivity, performance, or both by the Subjects that may occur merely as a function of being in an investigation" (Drew, 1980) were assumed to be operative to a normal degree in both samples, on both the treatment and the experimental groups, and were also assumed to be controlled through random assignment of Subjects to the experimental groups.

3.3.1.e - Bias in Group Composition

Any systematic differences between the treatment and the control groups in the present study, in addition to the treatment under study, were assumed to be controlled through random assignment of Subjects to the experimental groups.

3.3.1.f - Experimental Mortality

Experimental mortality can became a threat to the internal validity of the present study if a differential loss of Subjects takes place between the comparison groups. A Subject loss, in terms of this study, was interpreted as the total invalidation of the notes taken by a given Subject during the experimental session. The invalidation could be caused by a combination of the following factors:

- Refusal to continue with the experimental session or rendering the worksheets useless, or both (assumed unlikely; no cases occurred);

- Inability to complete any sentence describing an "insight" during the whole time period allocated for paradigm problem analysis (moderate probability, associated with failure to follow problem instructions; one case occurred);

- Invalidation, during raw data transcription and scoring, of all "insights" written by Subject (assumed unlikely or quite remote);

The only rule that was set a priori to deal with the last factor mentioned was: If in the final count the two comparison groups lose Subjects in different proportions, and the total difference becomes more than ten percent of the total for the smaller group (an arbitrary proportion), then a number of cases of the larger group could be randomly chosen to be excluded from the final data analysis, in order to bring the total difference to within the ten percent caution point. One possible invalidating factor—which did occur—was not included in that list: problems with computer hardware or software, or both. There was one instance of software malfunctioning, which resulted in immediate invalidation of session results, although the session was resumed, after a few minutes, without further incidents.

3.3.2 - EXTERNAL VALIDITY

3.3.2.a - Population-Sample Differences

The "degree to which ... Subjects in [a given] study are representative of the population to which generalization is desired" (Drew, 1980) is one of the most important factors determining the external validity of a study. As mentioned earlier, in Section 3.1.6 ("The Population Under Study",) the major concern of the present study was to determine the nature of possible differences between different modes of data representation. It was considered somewhat premature, given the exploratory nature of the research and limitations of the sampling process, to generalize results beyond the limits of the population studied, although there are possible analogies to results of other studies in the visual cognition area (as in Carter, 1947a). One must bear in mind, however, that most of those studies present different choices of dependent and even independent variables, making cross comparisons for validation purposes cumbersome, if not outright improper.

Only several replicative studies, with samples drawn from a much wider variety of groups within the reference population, could fully address the issue of external validity of results from the present experiment, assuming that other possible threats to external validity had been properly neutralized.

3.3.2.b - Artificial Research Arrangements

Drew also pointed out that the reactive effects of experimental arrangements can threaten the external validity of an experiment "to the degree that the Subject's responsivity is altered". The choice of an office setting (versus a computer laboratory), a relatively common type of space with which most students are familiar, was made considering this particular problem. The specific characteristics of the experimental session workstation setting were also chosen to maximize privacy and opportunity to concentrate, with a minimum of distractions, on paradigm problem analysis.

3.3.2.c - Pretest Influence

As mentioned before (in terms of the testing practice threat to internal study validity), the research task tried to emulate as much as possible the "real world" setting represented by the way professionals, managers, and college students learn minimum skills about a new piece of software and try to utilize that new "tool" in their work or study environment. Although the tutorial could have been considered either as a warm-up or "pretest" task, its main objective was to provide all Subjects with a comparable, minimum set of acquired skills necessary prior to the problem analysis phase of the experimental session. Any results of the study, therefore, can be generalized only towards individuals that eventually acquire a basic and equivalent set of skills to handle quantitative data in either the graphic or numeric mode. The learning in question, however, is expected to occur gradually among the "reference" population for this study, due to the impact of personal computing in the professional arena.

4 - DATA COLLECTION

4.1 - DATA COLLECTION FRAMEWORK

4.1.1 - FORMAT OF DATA COLLECTION

The experimental problem for this study was to examine the relationship between the "insight" production scores of two groups looking at data displayed through two different modes on a computer screen. The format of data collection used in this study can be described as "systematic stimuli/systematic responses", using the scheme suggested by Galtung (1969). By "systematic stimuli", Galtung meant "stimuli that are kept constant when [Subjects] are changed, in the sense that all [Subjects] are exposed to the same stimuli" (except for only one factor, the experimental treatment, if both Subject groups are considered); by "systematic responses", he meant that "what is kept constant are the response categories."

The study is assumed to have taken place at a single moment in time, both in terms of sampling and of experimental session scheduling. In practice, given the use of only one personal workstation, only one Subject did participate in an experimental session at a time.

The administration of all experimental sessions was undertaken by a single person. Interaction (questions during the course of the session) between the session administrator and Subject was kept to a minimum. The Subject was expected to derive all the information necessary for the session activities from the written instructions and the "Help" information available through the computer system. The Subject was expected to spend the whole session (one to one and a half hours) at the experimental session workstation. The duration of sessions actually ranged from one half to two and a half hours. No interaction was allowed between Subjects and any person entering the office area while an experimental session took place; the workstation itself was well protected, visually, by a partition six feet high. The session administrator also ensured that, even when other persons entered the office area in question during an experimental session, Subjects remained undisturbed.

Before each session, a package with all the necessary worksheets and instruction sheets was ready for use by the scheduled Subject. The session control sheet and the protocol of verbal instructions were also ready for the session administrator, prior to the beginning of the session, in a separate desk, apart from the Subject's workstation.

4.1.2 - PILOT EXPERIMENTAL SESSIONS

In preparation for data collection and as a "feasibility" study for the research endeavor, a series of experimental sessions took place in July 1985. Alternative versions of the tutorial were examined, as well as of the several forms created for data gathering, session control, and treatment of raw data. In that pilot study, ten Subjects participated, evenly divided between the treatment and the control groups.

Several format modifications were made to the parts of the experimental sessions, as a result of recommendations and complaints submitted by the Subjects. The session component mostly affected by those recommendations was the tutorial, which was considerably shortened and simplified. The computer system used was also changed as a result; due to screen resolution for the graphics mode (considered too low), a move to a personal computer offering a better resolution was deemed necessary. A draft summary of pilot study results is presented in Appendix C.6.

4.2 - EXPERIMENTAL SESSION DESIGN

The experimental session consisted of the following sequence:

1. Questionnaire requesting information on previous experience with computers, statistics, and selected demographic variables (age, sex, major curriculum, or program of study, years of higher education); scheduled to take 2-3 minutes.

2. Instructions tutorial on the commands of the data manipulation system, based on a sample problem, comparable to the paradigm problem; scheduled to take from 20 to 30 (maximum) minutes (the actual range was from 10 to 70 minutes);

3. Analysis of the paradigm problem data; scheduled to last a maximum of 60 minutes (hidden time limit, not divulged prior to the end of the session in order to minimize performance anxiety).

4. Questionnaire on enjoyment of experimental session (tutorial and problem analysis) and specific comments on problems and difficulties encountered during those phases; scheduled to take 2-5 minutes.

4.2.1 - QUESTIONNAIRES

The first questionnaire contains questions relative to skills which, even if not possible to control prior to the beginning of the study, could at least be used to portray a composite image of the two samples, to pinpoint any extreme conditions that might explain possible anomalies in the results. The "previous experience" questions were scheduled for completion prior to paradigm problem analysis in order to avoid any interference between each Subject's perceived session "performance" and the answers to those questions. The answers to the experience questions were given in terms of a self-reported scoring along a visual scale ranging from '1' (minimum experience) to '7' (maximum experience).

The objective of the second questionnaire was to uncover major unforeseen difficulties experienced by Subjects regarding the format and commands of the experimental system, difficulties that might possibly have influenced their perception of the system's capabilities, at the time of their experimental session. Appendix B.1 ("Questionnaires for Experimental Session") exhibits the two questionnaires.

4.2.2 - TUTORIAL

4.2.2.a - Tutorial Objectives

The tutorial had two objectives:

1. To instruct Subjects in the data manipulation capabilities of the system, particularly on the display format for the specific data display mode for each Subject.

2. To provide a "hands-on" experience of data manipulation of a small problem dataset, to ensure that all Subjects had passed at least once through the same process of acquiring the specific skills necessary to operate the system, and were therefore able to start the problem analysis on an equal footing with other Subjects, regardless of their level of expertise or familiarity with computer programs of any kind.

4.2.2.b - Tutorial Time Frame

The issue of skill acquisition was very important for the present study as a possible source of bias in the rate of generation of "insights". It can be safely assumed that graphics, particularly the type of multivariate data representation used in the experimental system, are not familiar to the majority of the reference and target populations of this study. Based on the "insight" generation rates evidenced by Subjects of the pilot experimental sessions, there might have been a time-lag of approximately 10 to 20 minutes before Subjects in the treatment group (graphic mode of data display) started to significantly generate "insights", in the absence of a "hands-on" tutorial. It was assumed here that there is a small but significant time differential between the learning curves for the two groups, treatment and non-treatment (numeric mode of data display), simply because "numeracy" is a more widespread skill in our society than "graphicacy".

The 20 to 30 minutes (maximum) time frame for the tutorial was considered sufficient to enable all Subjects to start the next phase on an equal basis, in terms of the skills necessary to manipulate all command systems. It was heavily stressed in the tutorial instructions that the use of all the commands was not imperative. It is possible, however, that Subjects may have interpreted the simple availability of those commands as an indication that they were expected to use them. At any rate, at the end of the tutorial every Subject was expected to be able to successfully specify and complete any given operation in the system, assuming that the Subject might eventually decide to specify that operation.

Appendix B.2 ("Tutorial Section of Experimental Sessions") exhibits the tutorial instructions (to be handed to Subjects at the beginning of the tutorial) and a sample notesheet. Appendix D ("Information Tutorial

Screens") presents the basic screens displayed to the two groups prior to the "hands-on" tutorial.

4.2.2.c - Supporting Information

Two information sheets were made available to Subjects, to help maintain task concentration and avoid distractions or unnecessary screen movements (change in displays). The sheets contained, respectively, a summary of the "Help" information for all system commands, and a list of all field names and descriptions. Those sheets were intended for use during both the tutorial and problem analysis sections of the experimental sessions. Appendix B.2 also exhibits a copy of both information sheets.

4.2.3 - ANALYSIS OF THE PARADIGM PROBLEM

The problem analysis component of the experiment had one objective: to provide Subjects with the opportunity to look at, and manipulate at will (with a limited but sufficient set of operations) a multivariate quantitative dataset.

4.2.3.a - Problem Instructions

The main goal of the problem instructions was to provide a "loose" problem setting, a context against which a quantitative dataset could be interpreted. It is recognized here that "observations" and "inferences" can be made about any isolated quantitative dataset (a data matrix without any description of what type of data is in storage). However, it seemed common-sensical to embed a quantitative dataset in a less "abstract" frame of reference, to minimize stresses in Subjects who may vary widely in their familiarity with data manipulation and statistics. The paradigm problem setting is discussed in more detail in the next section of this chapter.

The problem instructions emphasized the need for deferred judgement (a conscious effort to avoid evaluation of the validity or worth of observations.) The validity of the observations (or "insights") were judged a posteriori, based exclusively on plausibility and common sense; obviously nonsensical remarks were excluded and invalidated; all others were considered valid, if supported by the quantitative dataset. The part of the instructions emphasizing deferred judgement amounted, for all practical purposes, to a form of instruction in divergent thinking. Although this last factor could be considered a significant bias, it was assumed controlled in the sense that it was equally shared by all Subjects. The instructions on deferred judgement were deemed a better alternative than allowing an unrestricted and unknown degree of mental "censorship" of ideas considered by Subjects.

Subjects were requested to confine all their written notes to the worksheets provided and to write "observations" on the left side of the worksheet, "inferences" on the right. Appendix B.3 ("Paradigm Problem Analysis") exhibits a copy of the problem instructions and of a sample worksheet.

4.2.3.b - Time Frame

A maximum of 60 minutes was allowed for the problem analysis phase. Every 10 minutes after the beginning of this part of the experimental session, the session administrator marked the worksheets with red pen to indicate the time intervals during which "insights" were generated. Subjects were not told in advance of the 60 minute limit, in order to minimize possible performance anxiety; the problem instructions emphasized that they could take as much time as needed to examine the dataset. Subjects remained free, if they wished, to terminate the paradigm problem analysis at any time prior to the 60 minute limit (one-fourth of the eighty Subjects did that).

4.2.4 - CONTROL OF ACTIVITIES

A session control form was used to keep time records of each session, including the six 10 minute intervals. A short sequence of verbal instructions was also given, in addition to the written instructions. The written instructions were read aloud, verbatim, for each Subject; this reinforcement was assumed to ensure a greater uniformity of instruction comprehension than what would have taken place with no reinforcement. Appendix B.4 ("Time Control Sheet and Verbal Instructions") exhibits a copy of a sample session control sheet, and a list of the verbal instructions.

4.2.5 - ENVIRONMENTAL CONDITIONS

An actual office work environment (201 Cowgill Hall, the administrative offices of the College of Architecture and Urban Studies at Virginia Tech) was used for conducting the experimental sessions. Although it would have been impossible to remove most test-related stresses for all Subjects, the avoidance of a "laboratory"-type of environment was assumed important as a way to remove one major possible aspect of those stresses. Office environments, when presenting a certain amount of "normal" clutter on desks and work areas, and particularly when sheltered from excessive noise and distractions, can be considered more congenial to concentration than a computer terminal room or laboratory, particularly for novice computer users. This premise was validated by the comments of several self-reported non-experienced computer users.

All sessions were conducted off-hours, relative to the normal working hours (Monday through Friday, 08:00 to 17:00 hours) of the offices in question. This schedule ensured a minimum of distractions and disrupting background noises. The offices were air-conditioned, with an average room temperature of 69 ± 2 °F which actually ranged from 65 to 74 °F, with an electric heater made available for Subjects who felt uncomfortable, on occasion, with the lower temperatures. The air-conditioning system provided a constant background noise; a large volume laser printer terminal in an enclosed room adjacent to the premises contributes a certain amount of sporadic noise, even off-hours. General lighting was provided by fluorescent ceiling fixtures, with an average illuminance level of 30 foot-candles around the workstation area for the experimental problem. There was a certain amount of glare produced by the rows of ceiling lights on the computer screen of the workstation used.

Appendix B.5 ("Workstation Area for Experimental Sessions") exhibits photographs and floor plan of the workstation area, as well as of a simulated experimental session.

4.3 - THE PARADIGM PROBLEM

4.3.1 - DESCRIPTION OF THE PROBLEM

The paradigm problem consisted of time-series data on recent past performance (for two years) of a small-scale hotel (based on Bertin, 1983). The exploration of alternatives (solutions) for improving performance of that hotel was deliberately excluded from the present study, as explained before, in order to examine only the early stages of problem-solving. The inclusion of the final stages of problem-solving in future studies is obviously a natural and necessary extension of the present study.

The paradigm problem had both an objective and a subjective component, the latter consisting of each individual Subject's perceptions and interpretations of their real-life experiences with hotels of any kind. The objective component of the problem consists of quantitative data summarizing the available data on the Subject.

Appendix B.3 ("Paradigm Problem Analysis") exhibits a complete description of the problem's quantitative dataset, as presented to the Subjects and as available in the computer-based system.

4.3.2 - SELECTION OF THE PROBLEM

The paradigm problem in question was chosen by taking into consideration:

Datataset Size Approximately 500 elements, the upper side of the range of 100-500 dataset elements assumed to be most representative of the problems professionals face in their daily work.

Dimensions Approximately 20 variables, or dimensions, assumed also to fall within the span of perceptual dimensionality associated with the cognitive systems of human beings (Miller, 1956); this number of dimensions, however, falls considerably outside the range of dimensions of graphic representations in present use by professionals.

. **Familiarity** Conceptualization of the "reality" of a small hotel can safely be assumed to be relatively easy for the study Subjects. The dynamics of the problem involve variables within the experiential frame of reference of the target population, sufficiently familiar as to avoid perceptions of excessive problem complexity, but deliberately vague to avoid being a case study on hotel management.

4.3.3 - ASSUMPTIONS ABOUT THE PROBLEM

4.3.3.a - Problem Dataset

All the data elements refer to the same problem. The dataset was a representative summary of the entire original dataset available at the time of problem formulation.

4.3.3.b - Problem Structure

The problem can be considered unstructured, from the Subjects' point of view, in terms of having enough novel aspects—in spite of a sense of familiarity purposefully cultivated—to demand that the Subjects "must first develop what may be an ad hoc procedure for reaching a decision" (Moore and Chang, 1983). The size of the dataset, particularly the number of variables in question, assured the almost practical impossibility of any given Subject to assume full familiarity with the problem domain in the time frame alloted.

4.3.3.c - Other Observations about the Problem

It was assumed that the paradigm data represents a small scale database, obtained and digitized through any available means—the specific means being irrelevant to the research question. No analysis of the impact of database availability or access difficulty on decision performance was made in this study; those factors, however, can present more serious problems, in real working conditions, for much larger datasets.

4.4 - THE COMPUTER SYSTEM USED

4.4.1 - COMPUTER EQUIPMENT

For the present study, the hardware used was assumed to be representative—in terms of relevant characteristics, as display resolution and memory size—of the present and near future (2-5 years) base of installed personal computers with DSS emphasis. The application programming language was made transparent to Subjects, and the overall structure of the application procedural language ensured that, novice and experienced computer users alike, every Subject started the experimental session at the same level (i.e., lack of knowledge) about the system. A variable time-length tutorial allowed for more experienced users to proceed somewhat faster than novice users towards the beginning of paradigm problem analysis.

The central processing unit used was an IBM PC-AT, with a 20 M-byte fixed disk-drive, floating-point co-processor, and 512 K-bytes of installed random-access memory.

The monitor used was an Amdek Video-310A monochrome display, 12 inches diagonally wide, with amber characters displayed on a dark-gray background. The display has a resolution of 720 (horizontal) versus 348 (vertical) pixels, made available through use of a Hercules Graphics Card installed in the processing unit.

For the present study, the use of color in displays was deliberately avoided. Colors are simply not efficient as an aid for discrimination among dataset elements measured along interval or ratio scales (Bertin, 1971, and 1983; Cleveland and McGill, 1985), although they can be quite useful for differentiation among elements of datasets measured along nominal and categorical scales.

4.4.2 - COMPUTER PROGRAMMING

4.4.2.a - Programming Language

The application programming language was STSC's APL*PLUS/PC, an extended version of the APL language developed for IBM PC computers and compatibles. The major advantage of APL for the present study, from a programmer's point of view, lies with its data array processing capabilities.

4.4.2.b - Application Program

The application program consisted of approximately 40 separate APL functions (routines, or sub-programs), handling user requests, error-checking, data formatting and displaying, and session monitoring. The application program interface consisted of a limited set of commands which performed operations deemed necessary for manipulation of quantitative relational data. All the APL functions and commands were the same for both experimental groups, with only the screen output subject to the experimental differences between the active treatment (graphic display of data) and control treatment (numeric display of data).

All the commands could be directly requested by pressing only two or three keys. Therefore, only the argument for each specified operation (eg., a user-selected table name) had to be typed in full by the Subject; this minimized the time and stress (for novice users) associated with computer command specification. The list of the available commands is presented in Appendix B.2.3 (" 'Help' Information Sheet").

The "Help" information sheet, containing the synopsis of available commands, was made available to each Subject in order to minimize on-screen display of "Help" information, which, during pilot experimental sessions, apparently proved to be quite distracting for some Subjects, in terms of their concentration on the analytical task. Windows on the screen display were one alternative originally rejected, in order to keep constant the display area reserved for current view of the paradigm problem data.

"Help" windows were finally retained as a complement to the written sheet but occupying, however, only a small part of the screen display, superimposed on a current data table display.

4.4.2.c - Session Monitoring

All commands and arguments requested by a Subject during the course of an experimental session were recorded and stored in the processing environment, in the form of an APL component file (data storage matrix), with one file for each Subject. Each file (a relation) contains five basic fields: time stamp, command description, command-request error code (if any), argument specification (if any), and argument-request error code (if any). Those files allowed the reconstruction of all the steps taken by Subjects during the session, in case of interpretation questions regarding the worksheet notes with "observations" and "inferences".

5 - TREATMENT OF THE RAW DATA

5.1 - PREPARATION OF THE RAW DATA

5.1.1 - VERBATIM TRANSCRIPTION OF RAW DATA

The first stage in the preparation of the raw data was the verbatim transcription of Subject notes taken during the experimental problem analysis session. The data were transcribed into machine-readable form by means of the text-processing system SCRIPT/VS Release 3 from IBM, operating under VM/SP Release 3 with CMS Release 3. Items from the original worksheets were duplicated word by word, line by line, with the same abbreviations, upper and lower case letters, and punctuation used by Subjects. Comments about orthography or illegible words or symbols were included as necessary. Numbering of "observations" and "inferences" (in two separate lists, according to worksheet page layout) conformed to numbering scheme used by the Subject, if that was the case; otherwise, the existing logical separation of sentences—by spaces, bullets, arrows, or other means—was used for the numbering of "observations" and "inferences". Appendix C.1 ("Sample Verbatim Transcription of 'Insights'") exhibits a sample set of worksheets filled with handwritten notes, and their verbatim transcription.

5.1.2 - PARSING AND COMPRESSION OF RAW DATA

The second stage in the preparation of the raw data was the decomposition and reorganization of the essential parts of each sentence into a standard format for data produced by all the Subjects. At this stage, a new set of SCRIPT/VS files was created, for the manipulation and storage of research data.

5.1.2.a - Parsing of Sentence Elements

Subjects were requested to present individual "insights" in separate, simple sentences. As this request was primarily a guideline, it was expected that other formats for presentation of "insights" would take place, as proved to be the case. "Observations" and "inferences" were broken into simple sentences, whenever possible, following the same order of appearance found at the verbatim transcription. Sentences, classified according to the typology suggested by Baker (1976), were subject to the changes outlined in Figure 15 on page 125.

Type:	Changes:
Simple	None.
Compound Simple	Separate the sentences only if they refer to different relational fields of the experimental problem dataset.
Complex	None.
Compound Complex	Separate into individual complex sentences.

Figure 15. "Insight" Sentence Modification: Structural changes by sentence type.

5.1.2.b - Substitution and Elision of Sentence Elements

The structure of each "observation" was modified, if necessary, to reflect the following sentence structure:

Subject - Verb - Description of State - Specifier

Sentence elements were elided, or had abbreviations or more compact expressions substituted for them. These operations took place to reduce the notation of "insights" to the shortest intelligible and unambiguous representation for each "insight". A terse notation, of a format common to data produced by all Subjects, was assumed to facilitate overall readability and comparison among "insights" produced by different Subjects.

Sentence subjects should correspond, without exception, to the field names of the experimental problem data. Those names were presented in abbreviated form. Verbs, since they belong by definition to the same type (linking a sentence subject to a state of being, and commonly represented by the verb "to be"), were elided from all sentences.

"Description of State" elements were translated into a basic set of 10 to 15 adjectives or adverbs, either in normal, comparative, or superlative form, according to the original verbatim sentence. A "specifier", in the context of the research data, is a sentence element with more detailed information to complement or narrow down the description of a given state. Specifiers were used primarily to indicate time periods, taken from the "Months" field, in the experimental problem dataset.

In principle, all propositions, conjunctions, and interjections were removed, with the exception of a very few whose removal might cause extreme ambiguity in sentence interpretation. Missing or unreadable sentence elements, if deemed relevant (as indicated by position within the sentence structure), were indicated by an ellipsis placed between brackets ('[...]'). Unreadable or misinterpreted information, if and when its true expression could be ascertained, were replaced by the correct information between brackets.

Appendix C.2 ("Sample Parsing and Compression of 'Insights'") exhibits an example of a verbatim transcription of notes and of the parsing and compression of those notes.

5.1.3 - REORDERING OF RAW DATA

The third stage in the preparation of the raw data was the reorganization of "insights" in three distinct blocks, corresponding to the final classification of "insights": certified "observations", "inferences", and "suggestions". A new set of SCRIPT/VS files was created to manipulate and store reordered research data. Appendix C.3 ("Sample Reordering of 'Insights'") exhibits an example of reordered, parsed "insights".

5.2 - SCORING OF "INSIGHTS"

Only "observations" and "inferences" were retained for final analysis. Suggestions, in spite of experimental problem instructions which specifically excluded them, were considered nonetheless likely to occur, due to either failure to follow instructions properly or, as one Subject commented, because "... it was hard not to do [it]." However, "suggestions" were excluded mainly because of the research problem emphasis on generation of ideas possibly relevant for problem solving, not on problem solutions per se, which belong to another, quite distinct realm of problems.

Format Rules

P Partial "insights" (non-overlapping) refer only to different subsets of one single field. For each subject, only the first occurrence of such sentences should be kept. The other occurrences should be subsumed under the first one.

R For each subject, repeated "insights" should be eliminated, with the exception of the first one in the series.

I Incomplete "insights", for which no specifier has been or can be identified, should be eliminated.

Content Rules

S "Spot" "insights", i.e., written descriptions of one single record of one single field of the experimental problem dataset, should be eliminated on the grounds of data repetition.

W Wrong "insights", i.e., those totally unsupported by examination of the experimental problem dataset , or simply nonsensical, should be eliminated.

T Tautological sentences should be eliminated.

Figure 16. Scoring "insights": Guidelines for rejection of ill-formed or incorrect "observations" and "inferences".

At this phase, "insights" were screened according to the set of rules presented in Figure 16 on page 128. Those "insights" falling in at least one of those categories were eliminated, with the exceptions indicated. The rules are divided in two types: formation and content. The objective of formation rules was to make sure that only well formed and unique "insights" would be present for final analysis. The objective of content rules was to eliminate sentences obviously nonsensical, inappropriate, or incorrect (in terms of validation by experimental problem data). The scoring procedures were repeated once, in order to ensure a reasonable level of accuracy in "insight" scoring.

Appendix C.4 ("Sample Scoring of 'Insights'") exhibits an example of a scoring sheet for "insights".

5.3 - CATEGORIZATION OF "INSIGHTS"

This phase of data preparation involves the determination of proper "insight" category. An open-ended schema for classification of "observations" and "inferences" was used; it is based on the number of fields and field groups stated or implied in each "insight". An increasing number of fields or field groups mentioned in one sentence was assumed to be a rough, but reliable, indicator of "insight" complexity; the schema presented in Figure 17 on page 130 was be referred to, for that reason, as complexity categories for "insight" classification.

Field Category	No. of Fields	Levels	No. Fld. Groups	Field-Group Category
Single-field	1	A	1	Single-group
Multiple-flds.	2	B	1	
	3	C	1	
	4	D	1	
	5	E	1	
	6	F	1	
	2	G	2	Multiple-group
	3	H	2-3	
	4	I	2-4	
	5	J	2-5	
	6	K	2-6	
	7	L	2-7	
	8	M	2-8	
	16-20	N	5-9	

Figure 17. Complexity Levels: Framework for classification of "insights" generated by Subjects during experimental sessions.

For each Subject, the individual categorization sheets were scored to obtain the algebraic sum of well-formed, meaningful "observations" and "inferences" in each complexity category (of all the categories produced by at least one Subject during the course of the experiment). Appendix C.5 ("Sample Categorization of 'Insights'") exhibits an example of a categorization sheet for "insights".

6 - PRESENTATION OF THE RESULTS

6.1 - STATISTICAL ANALYSIS OF THE RESULTS

The null hypothesis was that there would be no differences between the scores of Subjects looking at a graphic representation of data and the scores of Subjects looking at a numeric representation of the same dataset, for each of four levels of "insight" complexity. The results for the statistical significance tests of the relationship between the dependent and independent variables, for each of the four null hypotheses, are presented in Figure 18 on page 133. The statistics for strength and magnitude of the relationships are presented in Figure 19 on page 134.

The four null hypotheses were rejected, i.e., there were statistically significant differences between the scores for "insight" generation—for each of the four operational definitions of the dependent variable—for Subjects looking at a graphic, as opposed to numeric, representation of the paradigm problem dataset. The strength of the four relationships represented by the operational hypotheses were relatively weak. For the first three hypotheses, which measure increasingly higher concentrations of complex "insights", there was a corresponding increase in relationship strength, indicating that the majority of the treatment effect is associated with the 'multiple field-groups' type of "insight" (tested primarily by the third hypothesis).

STATISTICAL SIGNIFICANCE OF RELATIONSHIPS

ASSUMPTIONS

1	**Model:**	$X_j = e_{i.}$ $Y_j = e_{m+j} + \Delta$ Independent random samples $N = n + m$ (observations)	$i = 1, \ldots, m$ $j = 1, \ldots, n$
2	**Level of Measurement:**	Ordinal scale (for dependent variable "insight generation".)	
3	**Hypotheses:**	$H1_0$ $\Delta1 = 0$ $H2_0$ $\Delta2 = 0$ $H3_0$ $\Delta3 = 0$ $H4_0$ $\Delta4 = 0$	($H1$ $\Delta1 > 0$) ($H2$ $\Delta2 > 0$) ($H3$ $\Delta3 > 0$) ($H4$ $\Delta4 > 0$)
4	**CRITICAL REGION:**	One-tailed test.	

5 SIGNIFICANCE LEVELS:

		$H1_0$	$H2_0$	$H3_0$	$H4_0$
Levels	α	.10	.05	.025	.025

6 COMPUTATION OF TEST STATISTIC:

		$H1_0$	$H2_0$	$H3_0$	$H4_0$
Wilcoxon test	W	1663	1718	1739	1714
Mean value of W	$E_0(W)$	1501	1501	1501	1501
Variance/W (w/ties)	$var_0(W)$	10006	10006	10005	10002
No. of cases of X	m	40	40	40	40
No. of cases of Y	n	38	38	38	38
Wilcoxon test (lg. N's)	W^*	1.624	2.169	2.379	2.185
Std. dev. units/(α's)	$z_{(\alpha)}$	1.283	1.645	1.960	1.960

7 DECISION:

	$H1_0$	$H2_0$	$H3_0$	$H4_0$
1 - Reject H_0 [if $W^* \geq z_{(\alpha)}$]	X	X	X	X
0 - Accept H_0 [if $W^* < z_{(\alpha)}$]				

Figure 18. Statistical Significance of Null Hypotheses: Presentation of results

STRENGTH OF RELATIONSHIPS

Computation of Test Statistic:

		$H1_a$	$H2_a$	$H3_a$	$H4_a$
No. of rows	r	6	5	7	4
No. of cols.	c	2	2	2	2
Chi-square	x^2	4.318	10.50	12.13	4.634
Test statistic	V^2	.0554	.1346	.1556	.0594

MAGNITUDE OF RELATIONSHIPS

Computation of Statistic:

		$H1_a$	$H2_a$	$H3_a$	$H4_a$
Treat. Group Mean	X_1	16.03	7.63	5.18	4.87
Contr. Group Mean	X_2	13.20	5.28	3.10	4.13
Diff. bet. means	Δx	2.83	2.35	2.08	.74
Proportion/diff.	Rx	.24	.45	.67	.18

Figure 19. Strength and Magnitude of Relationships: Presentation of results

6.2 - INTERPRETATION OF OTHER FINDINGS

A detailed breakdown of the two groups, by "insight" levels (fourteen in all) and time groups (six intervals of 10 minutes) shed some more light into the picture. The major source of variation between the treatment ("graphic") and control ("numeric") groups can be traced to "insights" belonging to at least two "field-groups" ('G' through 'N'), as seen in Figure 20 on page 136. Particularly noteworthy is the last category, 'N', corresponding to observations encompassing the whole dataset. The actual cumulative frequencies for the "single-field" ('A') and "single-group, multiple-fields" categories ('B' through 'F') were practically the same for the two groups.

The initial expectation was confirmed that there would be primarily a qualitative rather than quantitative difference between the treatment and control groups "insight" scores. The use of straight fluency measurement—combining all levels of "insights"—can safely be said to be a very coarse instrument, at best, for revealing anything about the nature of the underlying phenomena studied. The refinement of a classification schema for analysis of the differences between the two samples seems to be the more profitable approach to take for extensions to the present study, particularly since the "Class 4" scores seemed to have fared not much better than "Class 1" scores, which a better classification schema would lead to expect.

GRAPHIC MODE (N=38)

		10	20	30	40	50	60 min	Total	
M-Group M-Fld.	N	13	2	6	4	1	2	28	197
	M				1			1	
	L				1		1	2	
	K				1		1	2	
	J	1				2		3	
	I	1				2		3	
	H	3	6	5	7	4	3	28	
	G	10	18	29	18	30	25	132	
S-Group M-Fld.	F	5	2	2	2	1		12	93
	E		1					1	
	D	4	11	8	8	4	2	37	
	C	1	2	3	3			9	
	B	7	13	5	4	3	2	34	
S-Fld.	A	67	68	62	42	46	34	319	
		112	123	120	90	93	70	609 Total	

NUMERIC MODE (N=40)

		10	20	30	40	50	60 min	Total	
M-Group M-Fld.	N	2	2	2	3		2	11	124
	M			1				1	
	L								
	K								
	J				2			2	
	I			1	1	1	3	6	
	H	1	4	1	3	1	2	12	
	G	8	15	13	16	18	22	92	
S-Group M-Fld.	F	1	3		4	4	1	13	87
	E								
	D	4	3	7	4	5	3	26	
	C	2	2	1	5	2	1	13	
	B	9	6	8	3	6	3	35	
S-Fld.	A	56	74	60	52	46	29	317	
		83	109	94	93	83	66	528 Total	

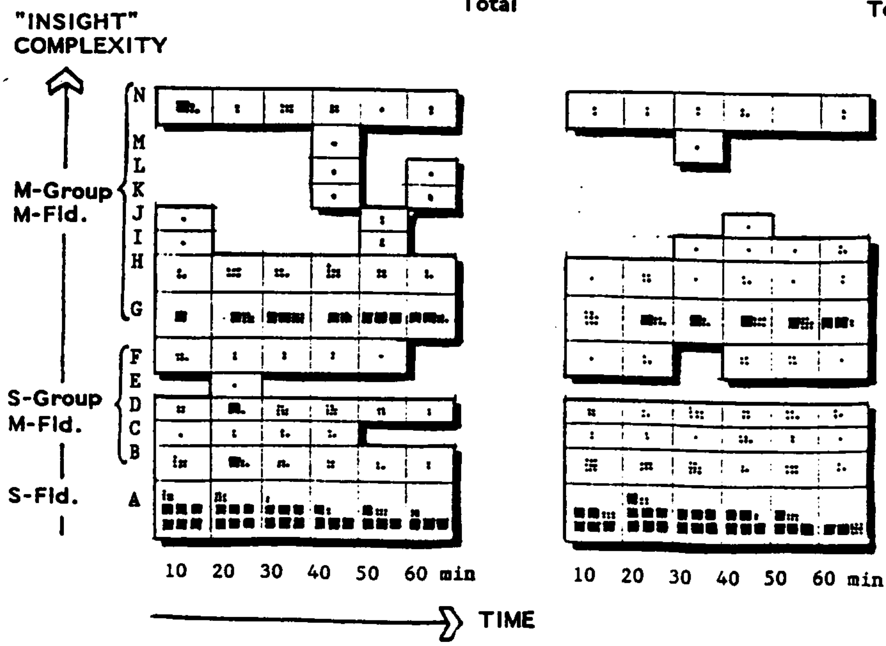


Figure 20. Numeric and Graphic Summary of the Data: Breakdown of overall "insight" scores by group, "insight" category, and time period.

As stated in the "Characterization of the Problem", in the first chapter, it was assumed that:

> there are products, similar to those derived from more complex and formal 'data' analysis, which are indispensable if that 'data' is to prove useful for problem-solving: compression (summary) of one field (or variable), relationships between two or more fields, observations focusing on "data" subsets or on the whole dataset.

The implicit ranking in those products was translated into a "complexity" scale, where the higher the number of data elements (variables or records, or both) correctly abstracted, the more appropriate, or useful, that particular "insight" was assumed to be for problem solving. As it turned out, a major difference between the two groups was the number of observations focusing on large, not directly related data subsets (i.e., belonging to different "field-groups".) Of special note is the category level of observations encompassing the whole dataset (labeled 'N'). A total of 21 Subjects (out of 38) from the treatment ("graphic") group made a total of 28 such observations, as opposed to only 10 Subjects (out of 40) from the control ("numeric") group making a total of 11 observations. The separate lists of "insights" for the two groups is presented in Appendix C.7.

A review was also made of the number of well formed but wrong "insights" (i.e., contradicted by the paradigm problem data. Approximately the same number of Subjects (24 out of 38 in the "graphic" group, 25 out of 40 in the "numeric" one) generated such observations, whose frequency was higher for the "graphic" (63) than for the "numeric" group (47). Most of these observations occurred, however, in the first 20-30 minutes into problem analysis and were also, in most instances, superseded by correct observations. The higher frequency of wrong observations by the "graphic" group might have been due to the lack of familiarity with the particular type of graphic representation used (multiple bar charts), a problem less perceptible after 30 minutes into problem analysis.

7 - GENERAL CONCLUSIONS

7.1 - RESTATEMENT OF OBJECTIVES

The study problem was to determine whether there was a discernible pattern of responses for "insight" complexity generation for Subjects looking at graphic, as opposed to numeric, interactive computer displays of quantitative data.

The present study was also set to investigate the quantitative and qualitative differences, if any, in the written responses of the two group samples. One group was assigned to an active treatment (use of a graphic display of the paradigm problem); the other group was assigned to a control treatment (use of a numeric display of the paradigm problem). The experimental units (the "Subjects") consisted of students (at the undergraduate and graduate levels) at Virginia Tech.

7.2 - CONCLUSIONS AND SUMMARY

In order to examine the implications of the present study, it must be reiterated that the study did not attempt to answer the general question of which mode (graphic or numeric) is the best one for data analysis. Arguably, there can be no simple, single answer to this problem. The question asked was rather how appropriate were "graphic and numeric representations of quantitative data to foster a more holistic, flexible examination of the problem space at the early stages of problem solving."

Both types of representation yielded approximately the same frequencies and rates of production for simpler types of "insight", i.e., observations or inferences associated with only one field (variable) or field-group. It can be argued that this type of "insight" simply forms the initial stepping stone through which to reach more complex types of observations. Therefore, for simple questions or inquiries related to the given dataset, there was no clear advantage of one type versus another, although specific data point values obviously could not be extracted from the graphic representation, nor was it ever intended to do so.

For more complex "insights", the outcome was different. There was an edge favoring the graphic data representation, in terms of frequency and even rates of "insight" generation. The implications of such results must be considered against the background of decision-making and creativity stages examined earlier. Exploratory data analysis techniques, it must be again emphasized, are not meant to replace more formal and refined analytical procedures. If time is of essence for professional DSS users, then any techniques that foster more expeditious invention, development and analysis of possible courses of action (the "design" phase of Simon's decision-making model) are preferable over other alternatives not so conducive.

The particular graphic representation used, within the confines of the paradigm problem utilized in the present experiment, seemed to be more effective for Subject understanding and verbalization of the problem space than the numeric representation. The problem space defined for this experiment was deliberately information-rich, unstructured, and novel, in terms of the participating Subjects. Although no specific questions were investigated concerning the effectiveness and efficiency for problem solution of the two types of data representation, it is expected nonetheless that the initial advantages of multivariate graphic representation can be carried towards the final stages of decision-making and problem solving.

The implications of this study with regard to its external validity must be, per force, more restrained. Although the Subjects who took part in the experiment can be considered a representative subset of the student body at Virginia Tech (chosen purely for research convenience purposes), no systematic sampling of that group was attempted. Only replicative studies of this kind, systematically investigating other groups, can fully address this issue. It is reasonable to expect, however, that the results of such studies would not markedly differ from the present one, inasmuch as Virginia Tech students are representative the reference population for this study.

It should be also noted that, very recently, a commercial software company (Javelin Software Corporation, Cambridge, Massachussets) released the first version of a spreadsheet program which includes the same type of graphic representation of data (multiple bar charts) used in the present experiment. Although this type of representation is not new (Bertin advocated it, for instance, in 1967), its very inclusion in such a commercial application product—subject to consumer acceptance—may be an indication that there are indeed advantages for having multivariate graphic representations in personal decision support systems, other than ubiquitous pie charts and simple bivariate graphics capabilities.

In light of the evidence examined during this study, it appears that at least one type of multivariate graphic representation can induce deeper (and moderatly faster) understanding of quantitative datasets than a numeric (tabular) representation. This advantage, it is expected, can be further extended towards the achievement of more creative—novel and useful—solutions.

7.3 - SUGGESTIONS FOR RESEARCH

The immediate necessary extension to the present study is in the area of the final stages of decision-making and problem solving. It is suggested that any such studies be divided in two stages, with the first one replicating the present experiment (for calibration purposes) and a second stage with specific goal statements that can be objectively measured and correlated with scores obtained in the first stage. Different paradigm problems, of varying (particularly larger) sizes, are another obvious extension to this study, as well as other types of multivariate graphic representations, such as multiple normalized bivariate scatterplots (Chambers et al., 1983).

In terms of control variables for further replicative studies, Guilford's "Structure-of-Intellect" model, as well as the cognitive style dichotomies examined earlier, present a useful framework for analysis of characteristics of people most likely to benefit from either a graphic or numeric mode of quantitative data representation.

LIST OF REFERENCES

ANDERSON, Edgar [1957]. "A Semigraphical Method for the Analysis of Complex Problems" Proceedings of the National Academy of Sciences 13: 923-7. Reprinted in Technometrics 2 (August 1960): 387-91.

BAKER, Sheridan (1976). The Complete Stylist and Handbook. New York: Thomas Y. Crowell.

BENIGER, J. R., and **ROBYN**, D. L. (1976). "The History and the Future of Graphics in Statistics." In AMERICAN STATISTICAL ASSOCIATION. 1976 Proceedings of the Social Statistics Section, American Statistical Association. Washington, D.C.: American Statistical Association.

BENNETT, John L. (1983). "Overview." In John L. BENNETT (ed.). Building Decision Support Systems. Reading, Ma.: Addison-Wesley.

BERTIN, Jacques. (1980). "The Basic Text of the Graph: A Matrix Theory of Graph Construction and Cartography." In Paul A. KOLERS, Merald E. WROLSTAD, and Herman BOUMA (eds.). Processing of Visible Language 2. Proceedings of the Second Conference on Processing of Visible Language, 3-7 September 1979, Niagara-on-the-Lake, Ontario, Canada. New York: Plenum Press.

BERTIN, Jacques (1981). Graphics and Graphic Information Processing. Translated by William J. Berg and Paul Scott. New York: Walter de Gruyter.

BERTIN, Jacques (1983). Semiology of Graphics. Translated by William J. Berg. Madison: University of Wisconsin Press.

BLALOCK, Hubert M., Jr. (1972). Social Statistics. Second Edition. New York: McGraw-Hill.

BRINTON, Willard C. (1914). Graphic Methods for Presenting Facts. New York: Engineering Magazine.

BRISSON, David W. (1978). "Visual Comprehension of N-Dimensions." In David W. BRISSON (ed.). Hypergraphics: Visualizing Complex Relationships in Art, Science, and Technology. AAAS Selected Symposia Series vol. 24. Boulder, Co.: Westview Press for the American Society for the Advancement of Science.

CAMPBELL, Donald T. (1957). "Factors Relevant to the Validity of Experiments in Social Settings." Psychological Bulletin 54: 297-311. Reprinted in Dennis P. FORCESE and Stephen RICHER (eds.). Stages of Social Research: Contemporary Perspectives. Englewood Cliffs, N.J.: Prentice-Hall, 1970.

CARLSON, Eric D. (1983). "An Approach for Designing Decision Support Systems." In John L. BENNETT (ed.) Building Decision Support Systems. Reading, Ma.: Addison-Wesley.

CARTER, Launor F. (1947a). "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 No. 19. Washington, D.C.: Government Printing Office.

CARTER, Launor F. (1947b). "An Experiment on the Design of Tables and Graphs Used for Presenting Numerical Data." Journal of Applied Psychology 31 (December 1947): 640-50.

CHACHRA, Vinod, and HETERICK, Robert C. (1982). "Anatomy of a Decision Support System." CAUSE/EFFECT 5 (September 1982): 6-11.

CHAMBERS, John M.; CLEVELAND, William S.; KLEINER, Beat; and TUKEY, Paul A. (1983). Graphical Methods for Data Analysis. Belmont, Ca.: Wadsworth International; Boston: Duxbury Press.

CHERNOFF, Herman (1972). "The Use of Faces to Represent Points in N-Dimensional Space Graphically." In Michael E. TARTER (ed.). Computer Science and Statistics. Sixth Annual Symposium on the Interface. 16-17 October 1972. University of California at Berkeley. Berkeley, Ca.: University of California, Division of Measurement Sciences and University Extension.

CLEVELAND, William S., and McGILL, Robert. (1985). "Graphical Perception and Graphical Methods for Analysing Scientific Data." Science 229 (August 1985): 828-33.

De BONO, Edward (1971). Lateral Thinking for Management: A Handbook of Creativity. N.p.: American Management Association.

DEWEY, John (1910). How We Think. Boston: D. C. Heath.

DEWEY, John (1933). How We Think: A Restatement of the Relation of Reflective Thinking to the Educative Process. Boston: D. C. Heath.

DICKSON, Gary W.; SENN, James A.; and CHERVANY, Norman L. (1977). "Research in Management Information Systems: The Minnesota Experiments." Management Science 23 (May 1977): 913-23.

DREW, Clifford J. (1980). Introduction to Designing and Conducting Research. Second Edition. St. Louis, Missouri: C. V. Mosby.

FIENBERG, Stephen E. (1979). "Graphical Methods in Statistics." American Statistician 33 (November 1979): 165-78.

FREEMAN, James; BUTCHER, H. J.; and CHRISTIE, T. (1971). Creativity: A Selective Review of Research. Second Edition. London: Society for Research into Higher Education.

GALTUNG, Johan (1969). Theory and Methods of Social Research. Revised Edition. New York: Columbia University Press.

GILMAN, Leonard, and ROSE, Allen J. (1984). APL: An Interactive Approach. Third Edition. New York: John Wiley & Sons.

GLOVER, John A. (1980). Becoming a More Creative Person. Englewood Cliffs, N.J.: Prentice-Hall.

GORRY, G. Anthony, and SCOTT MORTON, Michael S. (1971). "A Framework for Management Information Systems". Sloan Management Review 13 (Fall 1971): 55-70.

GUILFORD, J. P. (1959). "Three Faces of Intellect." American Psychologist 14 (August 1959): 469-79.

GUILFORD, J. P. (1976a). "Creativity: Yesterday, Today, and Tomorrow." In Dean C. DAUW and Alan J. FREDIAN (eds.). Creativity and Innovation in Organizations. Third Edition. Dubuque, Iowa: Kendall/Hunt.

GUILFORD, J. P. (1976b). "Factors that Aid and Hinder Productivity." In Dean C. DAUW and Alan J. FREDIAN (eds.). Creativity and Innovation in Organizations. Third Edition. Dubuque, Iowa: Kendall/Hunt.

GUILFORD, J. P. (1983). "Transformation Abilities or Functions." Journal of Creative Behavior 17 (October 1983): 75-83.

GUILFORD, J. P. (1984). "Varieties of Divergent Production." Journal of Creative Behavior 18 (January 1984): 1-10.

HOLLANDER, Myles, and WOLFE, Douglas A. (1973). Non-Parametric Statistical Methods. New York: John Wiley & Sons.

HOLT, Robert W. (1982). Theory, Evidence, Inference: A Handbook on the Scientific Method. Washington, D.C.: University Press of America.

International Business Machines Corporation (1982). An Introduction to APL2. SB21-3039. San Jose, Ca.: IBM.

IVERSON, Kenneth E. (1976). Elementary Analysis. Swarthmore, Pa.: APL Press.

IZENMAN, Alan J. [1978]. "Developments in Statistical Graphics 1960-1980." In U. S. DEPARTMENT OF COMMERCE. Bureau of the Census. Proceedings of the First General Conference on Social Graphics. Leesburg, Va., 22-24 October 1978. Technical Paper No. 49. Washington, D.C.: Government Printing Office, 1980.

KEEN, Peter G. W. (1980). "Decision Support Systems: A Research Perspective." In Goran FICK and Ralph H. SPRAGUE Jr. (eds.). Decision Support Systems: Issues and Challenges. Proceedings of an International Task Force Meeting. 23-25 June 1980. IIASA Proceeding Series Vol. 11. New York: Pergamon Press.

KEEN, Peter G. W., and SCOTT MORTON, Michael S. (1978). Decision Support Systems: An Organizational Perspective. Reading, Ma.: Addison-Wesley.

KICKERT, Walther. (1980). Organization of Decision-Making: A Systems Theoretical Approach. Amsterdam: North-Holland.

KRUECKEBERG, Donald A. (1978). "Practical Demand for Analytic Methods." In Robert W. BURCHELL and George STERNLIEB (eds.). Planning Theory in the 1980s: A Search for Future Directions. New Brunswick, N.J.: Center for Urban Policy Research, Rutgers University.

LABOVITZ, Sanford. (1970). "Criteria for Selecting a Significance Level: A Note on the Significance of .05." American Sociologist 3: 220-2. Reprinted in Dennis P. FORCESE and Stephen RICHER (eds.). Stages of Social Research: Contemporary Perspectives. Englewood Cliffs, N.J.: Prentice-Hall, 1970.

McCLEARY, George F., Jr. (1983). "An Effective Graphic 'Vocabulary'." IEEE Computer Graphics and Applications 3 (March/April 1983): 46-53.

McCOSH, Andrew M., and SCOTT MORTON, Michael S. (1978). Management Decision Support Systems. New York: MacMillan Press.

MANHEIM, Henry L. (1977). Sociological Research: Philosophy and Methods. Homewood, Il.: Dorsey Press.

MEZZICH, J. E., and WORTHINGTON, D. R. L. (1978). "A Comparison of Graphical Representations of Multidimensional Psychiatric Diagnostic Data." In Peter C. C. WANG (ed.). Graphical Representation of Multivariate Data. New York: Academic Press.

MILLER, George A. [1956]. "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information." Reprinted in Ralph N. HABER (ed.). Contemporary Theory and Research in Visual Perception. New York: Holt, Rinehart and Winston, 1968.

MINTZBERG, Henry; RAISINGHANI, Duru; and THEORET, Andre (1976). "The Structure of 'Unstructured' Decision Processes". Administrative Science Quarterly 21 (June 1976): 246-75.

MITCHELL, William J. (1977). Computer-Aided Architectural Design. New York: Van Nostrand Reinhold.

MOORE, Jeffrey H., and CHANG, Michael G. (1983). "Meta-Design Considerations in Building DSS." In John L. BENNETT (ed.) Building Decision Support Systems. Reading, Ma.: Addison-Wesley.

MORSE, Alan C. (1981). "Using Computer Graphics as an Aid to Understanding Simulation Data." Ph.D. Dissertation. University of Chicago.

NEGROPONTE, Nicholas, ed. (1975). Reflections on Computer Aids to Design and Architecture. New York: Petrocelli/Charter.

NEWELL, Allen, and SIMON, Herbert A. (1972). Human Problem Solving. Englewood Cliffs, N.J.: Prentice-Hall.

NEWTON, Carol M. (1976). "Graphic Data Analysis: Optimizing Tradeoffs Between Richness and Simplicity." In AMERICAN STATISTICAL ASSOCIATION. 1976 Proceedings of the Statistical Computing Section. Washington, D.C.: American Statistical Association.

PARNES, Sidney J. (1976). "Fact-Finding, Problem-Finding, Idea-Finding, Solution-Finding, and Acceptance-Finding--Applying the Total Process." In Dean C. DAUW and Alan J. FREDIAN (eds.). Creativity and Innovation in Organizations. Third Edition. Dubuque, Iowa: Kendall/Hunt.

PATRICK, Catharine (1938). "Scientific Thought". Journal of Psychology 5: 55-83.

PATRICK, Catharine (1941). "Whole and Part Relationship in Creative Thought". American Journal of Psychology 54: 128-131.

PICK, Herbert L., Jr., and **SALTZMAN**, Elliot (1978). "Modes of Perceiving and Processing Information." In Herbert L. PICK and Elliot SALTZMAN (eds.). Modes of Perceiving and Processing Information. Volume based on conferences sponsored by the Committee on Cognitive Research of the Social Science Research Council. Hillsdale, N.J.: Lawrence Erlbaum Associates.

POINCARE, Henri (1946). The Foundations of Science: Science and Hypothesis; The Value of Science; Science and Method. Authorized translation by George Bruce Walsted. Lancaster, Pa.: Science Press.

POWERS, Matthew; **LASHLEY**, Conda; **SANCHEZ**, Pamela; and **SCHNEIDERMAN**, Ben (1984). "An Experimental Comparison of Tabular and Graphic Data Presentation." International Journal of Man-Machine Studies 20: 545-66.

SCHMID, Calvin F., and **SCHMID**, Stanton E. (1979). Handbook of Graphic Presentation. Second Edition. New York: John Wiley & Sons.

SCHMID, Calvin F. (1983). Statistical Graphics: Design Principles and Practices. New York: John Wiley & Sons, Wiley Interscience.

SEYMOUR, Philip H. K. (1979). Human Visual Cognition. New York: St. Martin's.

SHOUKSMITH, George (1970). Intelligence, Creativity and Cognitive Style. New York: John Wiley & Sons, Wiley-Interscience.

SIMON, Herbert A. (1959). "Theories of Decision-Making in Economics and Behavioral Science." American Economic Review 49:253-83. Reprinted in Herbert A. SIMON. Models of Bounded Rationality. Vol. 2: Behavioral Economics and Business Organization. Cambridge: MIT Press, 1982.

SIMON, Herbert A. (1960). The New Science of Management Decision. New York: Harper & Bros.

SIMON, Herbert A. (1964). "Rationality." J. GOULD and W. L. KOLB (eds.) A Dictionary of the Social Sciences. Glencoe, Ill.: Free Press. Reprinted in Herbert A. SIMON. Models of Bounded Rationality. Vol. 2: Behavioral Economics and Business Organization. Cambridge: MIT Press, 1982.

SIMON, Herbert A. (1979). Models of Thought. New Haven: Yale University Press.

SPRAGUE, Ralph H. (1980). "A Framework for Research on Decision Support Systems." In Goran FICK and Ralph H. SPRAGUE Jr. (eds.). Decision Support Systems: Issues and Challenges. Proceedings of an International Task Force Meeting. 23-25 June 1980. IIASA Proceeding Series Vol. 11. New York: Pergamon Press.

TORRANCE, E. Paul (1962). Guiding Creative Talent. Englewood Cliffs, N.J.: Prentice-Hall.

TUFTE, Edward R. (1983). The Visual Display of Quantitative Information. Cheshire, Ct.: Graphics Press.

TWYMAN, Michael [1977]. "A Schema for the Study of Graphic Language (Tutorial Paper)." In Paul A. KOLERS, Merald E. WROLSTAD, and Herman BOUMA (eds.). Processing of Visible Language. Volume 1. Proceedings of the Conference Held at the Institute for Perception Research, IPO, Heindhoven, the Netherlands, 5-8 September 1977. New York: Plenum Press, 1979.

WASHBURNE, John N. (1927). "An Experimental Study of Various Graphic, Tabular and Textual Methods of Presenting Quantitative Material." Journal of Educational Psychology 18 (September 1927): 361-76; (October 1927) 465-76.

WITKIN, Herman A., and **GOODENOUGH**, Donald R. (1981). Cognitive Styles: Essence and Origins. Psychological Issues, Monograph 51. New York: International Universities Press.

The two page vita has been removed from the scanned document. Page 1 of 2

The two page vita has been removed from the scanned document. Page 2 of 2

APPENDIX A. STUDY-RELATED GENERAL MATERIALS

Contents:

A.1 - REVIEW EXEMPTION

The next page (section A.1.1) contains a reduced copy of the form requesting exemption from the need for formal research review by the University's Institutional Review Board. The following page (section A.1.2) presents a reduced copy of the statement on specific procedures to ensure subject data anonymity, submitted along with the Certificate of Exemption.

A.1.1 - Certificate of Exemption

CERTIFICATION OF EXEMPTION OF PROJECTS
INVOLVING HUMAN SUBJECTS

Principal Investigator(s) FRED W. LACERDA JR.

Department(s) Environmental Design & Planning, Ph.D. Program, Coll. Architecture & Urb. Stu.

Project Title "Comparative Advantages of Graphic vs. Numeric Representation of Quantitative Data"

Source of Support: Departmental Research X Sponsored Research __ Proposal No. ____

1. The criteria for "exemption" from review by the IRB for a project involving the use of human subjects and with no risk to the subject is listed below. Please initial all applicable conditions and provide the substantiating statement of protocol.

☐ a. The research will be conducted in established or commonly established educational settings, involving normal education practices. For example:

a) Research on regular and special education instructional strategies;
b) Research on effectiveness of instructional techniques, curricula or classroom management techniques.

☐ b. The research involves use of education tests (☐ cognitive, ☐ diagnostic, ☐ aptitude, ☐ achievement), and the subject cannot be identified directly or through identifiers with the information.

☒ c. The research involves survey or interview procedures, in which:

☒ a) Subjects cannot be identified directly or through identifiers with the information;
☒ b) Subject's responses, if known, will not place the subject at risk of criminal or civil liability or be damaging to the subject's financial standing or employability;
☒ c) The research does not deal with sensitive aspects of subject's own behavior. (illegal conduct, drug use, sexual behavior or alcohol use);
☐ d) The research involves survey or interview procedures with elected or appointed public officials, or candidates for public office.

☐ d. The research involves the observation of public behavior, in which:

☐ a) The subjects cannot be identified directly or through identifiers;
☐ b) The observations recorded about an individual could not put the subject at risk of criminal or civil liability or be damaging to the subject's financial standing or employability;
☐ c) The research does not deal with sensitive aspects of the subject's behavior (illegal conduct, drug use, sexual behavior or use of alcohol).

☐ e. The research involves collection or study of existing data, documents, records, pathological specimens or diagnostic specimens, or which:

☐ a) The sources are publicly available; or
☐ b) The information is recorded such that the subject cannot be identified directly or indirectly through identifiers.

2. I further certify that the project will not be changed to increase the risk or exceed the exempt condition(s) without filing an additional certification or application for approval by the Human Subjects Review Board.

Note: If children are in any way at risk while this project is underway, the chairman of the IRB should be notified immediately in order to take corrective action.

1985.11.01

Signature: Principal Investigator(s) Date / Signature: Principal Investigator(s) Date

11/1/85

(Optional Approval) Signature: Board Chairman/Authorized Reviewer Date

A.1.2 - Procedures to Ensure Data Anonymity

Experiment on Quantitative Data Display
Ph.D. Dissertation, FRED W. LACERDA JR
1985.11.01

PROCEDURES TO ENSURE DATA ANONYMITY

All data collected relative to each Subject will be used exclusively for research purposes. Only modified data, without possibility of identification of individual subjects, will be retained for final report (dissertation) and eventual publication.

There will be only one written document associating each Subject with his or her Experimental Subject Code Number; the document in question is a Subject Control Sheet, and its purpose is to keep track of satisfactory conclusion of the data collection process for the dissertation experiment. The data collection process includes not only the data collected during experimental sessions but the manipulation and interpretation of that data. Subjects may need to be contacted, after the end of an experimental session, to "translate" their written notes if their handwriting proves difficult to read. After conclusion of the data analysis, and presentation of final results, the Subject Control Sheet will be destroyed, eliminating all links (identifiers) between experimental Subject data and individual Subject information (name, or address, or any other personal identification item.)

A.2 - RANDOM SAMPLING PROCEDURES

Simple random sampling has been used for selection of Subjects to be included in the treatment group. The tables presented in the next two pages (from Blalock, 1972, p.554) were used for the choice (through arbitrary procedures) of random numbers leading to Subject assignment in the treatment group.

The numerical basis for the assignment was the order in which Subjects will take part in the experiment (one at a time,) a set of positive integers from 1 to 80, corresponding to the Subject participation number to be assigned on a "first-come, first-test" basis. The specific problem was the determination of which of those 80 Subjects will be part of the treatment group, i.e., a sample of size 40 from a population consisting of 80 individuals. Since the number 80 consists of two digits, the first two adjacent columns in every one of the four pages were arbitrarily chosen, followed by the eleventh and twelfth columns in each page, if need be. Every second two-digit number in those columns (in the order mentioned) were scanned for selection, provided they ranged from 01 to 80, and that they were not repeated. The process stopped when the 40th two-digit number was obtained. Listed below are the forty numbers selected:

02 05 07 08 09 10 11 12 19 20 21 23 24 30 31 32 33 34 37 38

39 46 48 51 52 53 59 60 61 64 65 66 67 69 73 74 75 78 79 80

A.2.1 - Random Table

Table B Random numbers

10 09 73 25 33	76 52 01 35 86	34 67 35 48 76	80 95 90 91 17	39 29 27 49 45
37 54 20 48 05	64 89 47 42 96	24 80 52 40 37	20 63 61 04 02	00 82 29 16 65
08 42 26 89 53	19 64 50 93 03	23 20 90 25 60	15 95 33 47 64	35 08 03 36 06
99 01 90 25 29	09 37 67 07 15	38 31 13 11 65	88 67 67 43 97	04 43 62 76 59
12 80 79 99 70	80 15 73 61 47	64 03 23 66 53	98 95 11 68 77	12 17 17 68 33
66 06 57 47 17	34 07 27 68 50	36 69 73 61 70	65 81 33 98 85	11 19 92 91 70
31 06 01 08 05	45 57 18 24 06	35 30 34 26 14	86 79 90 74 39	23 40 30 97 32
85 26 97 76 02	02 05 16 56 92	68 66 57 48 18	73 05 38 52 47	18 62 38 85 79
63 57 33 21 35	05 32 54 70 48	90 55 35 75 48	28 46 82 87 09	83 49 12 56 24
73 79 64 57 53	03 52 96 47 78	35 80 83 42 82	60 93 52 03 44	35 27 38 84 35
98 52 01 77 67	14 90 56 86 07	22 10 94 05 58	60 97 09 34 33	50 50 07 39 98
11 80 50 54 31	39 80 82 77 32	50 72 56 82 48	29 40 52 42 01	52 77 56 78 51
83 45 29 96 34	06 28 89 80 83	13 74 67 00 78	18 47 54 06 10	68 71 17 78 17
88 68 54 02 00	86 50 75 84 01	36 76 66 79 51	90 36 47 64 93	29 60 91 10 62
99 59 46 73 48	87 51 76 49 69	91 82 60 89 28	93 78 56 13 68	23 47 83 41 13
65 48 11 76 74	17 46 85 09 50	58 04 77 69 74	73 03 95 71 86	40 21 81 65 44
80 12 43 56 35	17 72 70 80 15	45 31 82 23 74	21 11 57 82 53	14 38 55 37 63
74 35 09 98 17	77 40 27 72 14	43 23 60 02 10	45 52 16 42 37	96 28 60 26 55
69 91 62 68 03	66 25 22 91 48	36 93 68 72 03	76 62 11 39 90	94 40 05 64 18
09 89 32 05 05	14 22 56 85 14	46 42 75 67 88	96 29 77 88 22	54 38 21 45 98
91 49 91 45 23	68 47 92 76 86	46 16 28 35 54	94 75 08 99 23	37 08 92 00 48
80 33 69 45 98	26 94 03 68 58	70 29 73 41 35	53 14 03 33 40	42 05 08 23 41
44 10 48 19 49	85 15 74 79 54	32 97 92 65 75	57 60 04 08 81	22 22 20 64 13
12 55 07 37 42	11 10 00 20 40	12 86 07 46 97	96 64 48 94 39	28 70 72 58 15
63 60 64 93 29	16 50 53 44 84	40 21 95 25 63	43 65 17 70 82	07 20 73 17 90
61 19 69 04 46	26 45 74 77 74	51 92 43 37 29	65 39 45 95 93	42 58 26 05 27
15 47 44 52 66	95 27 07 99 53	59 36 78 38 48	82 39 61 01 18	33 21 15 94 66
94 55 72 85 73	67 89 75 43 87	54 62 24 44 31	91 19 04 25 92	92 92 74 59 73
42 48 11 62 13	97 34 40 87 21	16 86 84 87 67	03 07 11 20 59	25 70 14 66 70
23 52 37 83 17	73 20 88 98 37	68 93 59 14 16	26 25 22 96 63	05 52 28 25 62
04 49 35 24 94	75 24 63 38 24	45 86 25 10 25	61 96 27 93 35	65 33 71 24 72
00 54 99 76 54	64 05 18 81 59	96 11 96 38 96	54 69 28 23 91	23 28 72 95 29
35 96 31 53 07	26 89 80 93 54	33 35 13 54 62	77 97 45 00 24	90 10 33 93 33
59 80 80 83 91	45 42 72 68 42	83 60 94 97 00	13 02 12 48 92	78 56 52 01 06
46 05 88 52 36	01 39 09 22 86	77 28 14 40 77	93 91 08 36 47	70 61 74 29 41
32 17 90 05 97	87 37 92 52 41	05 56 70 70 07	86 74 31 71 57	85 39 41 18 38
69 23 46 14 06	20 11 74 52 04	15 95 66 00 00	18 74 39 24 23	97 11 89 63 38
19 56 54 14 30	01 75 87 53 79	40 41 92 15 85	66 67 43 68 06	84 96 28 52 07
45 15 51 49 38	19 47 60 72 46	43 66 79 45 43	59 04 79 00 33	20 82 66 95 41
94 86 43 19 94	36 16 81 08 51	34 88 88 15 53	01 54 03 54 56	05 01 45 11 76

SOURCE: The RAND Corporation, *A Million Random Digits*, Free Press, Glencoe, Ill., 1955, pp. 1–3, with the kind permission of the publisher.

Table B Random numbers (*Continued*)

98 08 62 48 26	45 24 02 84 04	44 99 90 88 96	39 09 47 34 07	35 44 13 18 80
33 18 51 62 32	41 94 15 09 49	89 43 54 85 81	88 69 54 19 94	37 54 87 30 43
80 95 10 04 06	96 38 27 07 74	20 15 12 33 87	25 01 62 52 98	94 62 46 11 71
79 75 24 91 40	71 96 12 82 96	69 86 10 25 91	74 85 22 05 39	00 38 75 95 79
18 63 33 25 37	98 14 50 65 71	31 01 02 46 74	05 45 56 14 27	77 93 89 19 36
74 02 94 39 02	77 55 73 22 70	97 79 01 71 19	52 52 75 80 21	80 81 45 17 48
54 17 84 56 11	80 99 33 71 43	05 33 51 29 69	56 12 71 92 55	36 04 09 03 24
11 66 44 98 83	52 07 98 48 27	59 38 17 15 39	09 97 33 34 40	88 46 12 33 56
48 32 47 79 28	31 24 96 47 10	02 29 53 68 70	32 30 75 75 46	15 02 00 99 94
69 07 49 41 38	87 63 79 19 76	35 58 40 44 01	10 51 82 16 15	01 84 87 69 38
09 18 82 00 97	32 82 53 95 27	04 22 08 63 04	83 38 98 73 74	64 27 85 80 44
90 04 58 54 97	51 98 15 06 54	94 93 88 19 97	91 87 07 61 50	68 47 66 46 59
73 18 95 02 07	47 67 72 52 69	62 29 06 44 64	27 12 46 70 18	41 36 18 27 60
75 76 87 64 90	20 97 18 17 49	90 42 91 22 72	95 37 50 58 71	93 82 34 31 78
54 01 64 40 56	66 28 13 10 03	00 68 22 73 98	20 71 45 32 95	07 70 61 78 13
08 35 86 99 10	78 54 24 27 85	13 66 15 88 73	04 61 89 75 53	31 22 30 84 20
28 30 60 32 64	81 33 31 05 91	40 51 00 78 93	32 60 46 04 75	94 11 90 18 40
53 84 08 62 33	81 59 41 36 28	51 21 59 02 90	28 46 66 87 95	77 76 22 07 91
91 75 75 37 41	61 61 36 22 69	50 26 39 02 12	55 78 17 65 14	83 48 34 70 55
89 41 59 26 94	00 39 75 83 91	12 60 71 76 46	48 94 97 23 06	94 54 13 74 08
77 51 30 38 20	86 83 42 99 01	68 41 48 27 74	51 90 81 39 80	72 89 35 55 07
19 50 23 71 74	69 97 92 02 88	55 21 02 97 73	74 28 77 52 51	65 34 46 74 15
21 81 85 93 13	93 27 88 17 57	05 68 67 31 56	07 08 28 50 46	31 85 33 84 52
51 47 46 64 99	68 10 72 36 21	94 04 99 13 45	42 83 60 91 91	08 00 74 54 49
99 55 96 83 31	62 53 52 41 70	69 77 71 28 30	74 81 97 81 42	43 86 07 28 34
33 71 34 80 07	93 58 47 28 69	51 92 66 47 21	58 30 32 98 22	93 17 49 39 72
85 27 48 68 93	11 30 32 92 70	28 83 43 41 37	73 51 59 04 00	71 14 84 36 43
84 13 38 96 40	44 03 55 21 66	73 85 27 00 91	61 22 26 05 61	62 32 71 84 23
56 73 21 62 34	17 39 59 61 31	10 12 39 16 22	85 49 65 75 60	81 60 41 88 80
65 13 85 68 06	87 64 88 52 61	34 31 36 58 61	45 87 52 10 69	85 64 44 72 77
38 00 10 21 76	81 71 91 17 11	71 60 29 29 37	74 21 96 40 49	65 58 44 96 98
37 40 29 63 97	01 30 47 75 86	56 27 11 00 86	47 32 46 26 05	40 03 03 74 38
97 12 54 03 48	87 08 33 14 17	21 81 53 92 50	75 23 76 20 47	15 50 12 95 78
21 82 64 11 34	47 14 33 40 72	64 63 88 59 02	49 13 90 64 41	03 85 65 45 52
73 13 54 27 42	95 71 90 90 35	85 79 47 42 96	08 78 98 81 56	64 69 11 92 02
07 63 87 79 29	03 06 11 80 72	96 20 74 41 56	23 82 19 95 38	04 71 36 69 94
60 52 88 34 41	07 95 41 98 14	59 17 52 06 95	05 53 35 21 39	61 21 20 64 55
83 59 63 56 55	06 95 89 29 83	05 12 80 97 19	77 43 35 37 83	92 30 15 04 98
10 85 06 27 46	99 59 91 05 07	13 49 90 63 19	53 07 57 18 39	06 41 01 93 62
39 82 09 89 52	43 62 26 31 47	64 42 18 08 14	43 80 00 93 51	31 02 47 31 67

Table B Random numbers (*Continued*)

59 58 00 64 78	75 56 97 88 00	88 83 55 44 86	23 76 80 61 56	04 11 10 84 08
38 50 80 73 41	23 79 34 87 63	90 82 29 70 22	17 71 90 42 07	95 95 44 99 53
30 69 27 06 68	94 68 81 61 27	56 19 68 00 91	82 06 76 34 00	05 46 26 92 00
65 44 39 56 59	18 28 82 74 37	49 63 22 40 41	08 33 76 56 76	96 29 99 08 36
27 26 75 02 64	13 19 27 22 94	07 47 74 46 06	17 98 54 89 11	97 34 13 03 58
91 30 70 69 91	19 07 22 42 10	36 69 95 37 28	28 82 53 57 93	28 97 66 62 52
68 43 49 46 88	84 47 31 36 22	62 12 69 84 08	12 84 38 25 90	09 81 59 31 46
48 90 81 58 77	54 74 52 45 91	35 70 00 47 54	83 82 45 26 92	54 13 05 51 60
06 91 34 51 97	42 67 27 86 01	11 88 30 95 28	63 01 19 89 01	14 97 44 03 44
10 45 51 60 19	14 21 03 37 12	91 34 23 78 21	88 32 58 08 51	43 66 77 08 83
12 88 39 73 43	65 02 76 11 84	04 28 50 13 92	17 97 41 50 77	90 71 22 67 69
21 77 83 09 76	38 80 73 69 61	31 64 94 20 96	63 28 10 20 23	08 81 64 74 49
19 52 35 95 15	65 12 25 96 59	86 28 36 82 58	69 57 21 37 98	16 43 59 15 29
67 24 55 26 70	35 58 31 65 63	79 24 68 66 86	76 46 33 42 22	26 65 59 08 02
60 58 44 73 77	07 50 03 79 92	45 13 42 65 29	26 76 08 36 37	41 32 64 43 44
53 85 34 13 77	36 06 69 48 50	58 83 87 38 59	49 36 47 33 31	96 24 04 36 42
24 63 73 87 36	74 38 48 93 42	52 62 30 79 92	12 36 91 86 01	03 74 28 38 73
83 08 01 24 51	38 99 22 28 15	07 75 95 17 77	97 37 72 75 85	51 97 23 78 67
16 44 42 43 34	36 15 19 90 73	27 49 37 09 39	85 13 03 25 52	54 84 65 47 59
60 79 01 81 57	57 17 86 57 62	11 16 17 85 76	45 81 95 29 79	65 13 00 48 60
03 99 11 04 61	93 71 61 68 94	66 08 32 46 53	84 60 95 82 32	88 61 81 91 61
38 55 59 55 54	32 88 65 97 80	08 35 56 08 60	29 73 54 77 62	71 29 92 38 53
17 54 67 37 04	92 05 24 62 15	55 12 12 92 81	59 07 60 79 36	27 95 45 89 09
32 64 35 28 61	95 81 90 68 31	00 91 19 89 36	76 35 59 37 79	80 86 30 05 14
69 57 26 87 77	39 51 03 59 05	14 06 04 06 19	29 54 96 96 16	33 56 46 07 80
24 12 26 65 91	27 69 90 64 94	14 84 54 66 72	61 95 87 71 00	90 89 97 57 54
61 19 63 02 31	92 96 26 17 73	41 83 95 53 82	17 26 77 09 43	78 03 87 02 67
30 53 22 17 04	10 27 41 22 02	39 68 52 33 09	10 06 16 88 29	55 98 66 64 85
03 78 89 75 99	75 86 72 07 17	74 41 65 31 66	35 20 83 33 74	87 53 90 88 23
48 22 86 33 79	85 78 34 76 19	53 15 26 74 33	35 66 35 29 72	16 81 86 03 11
60 36 59 46 53	35 07 53 39 49	42 61 42 92 97	01 91 82 83 16	98 95 37 32 31
83 79 94 24 02	56 62 33 44 42	34 99 44 13 74	70 07 11 47 36	09 95 81 80 65
32 96 00 74 05	36 40 98 32 32	99 38 54 16 00	11 13 30 75 86	15 91 70 62 53
19 32 25 38 45	57 62 05 26 06	66 49 76 86 46	78 13 86 65 59	19 64 09 94 13
11 22 09 47 47	07 39 93 74 08	48 50 92 39 29	27 48 24 54 76	85 24 43 51 59
31 75 15 72 60	68 98 00 53 39	15 47 04 83 55	88 65 12 25 96	03 15 21 92 21
88 49 29 93 82	14 45 40 45 04	20 09 49 89 77	74 84 39 34 13	22 10 97 85 08
30 93 44 77 44	07 48 18 38 28	73 78 80 65 33	28 59 72 04 05	94 20 52 03 80
22 88 84 88 93	27 49 99 87 48	60 53 04 51 28	74 02 28 46 17	82 03 71 02 68
78 21 21 69 93	35 90 29 13 86	44 37 21 54 86	65 74 11 40 14	87 48 13 72 20

Table B Random numbers (*Continued*)

41 84 98 45 47	46 85 05 23 26	34 67 75 83 00	74 91 06 43 45	19 32 58 15 49
40 35 23 30 49	69 24 89 34 60	45 30 50 75 21	61 31 83 18 55	14 41 37 09 51
11 08 79 62 94	14 01 33 17 92	59 74 76 72 77	76 50 33 45 13	39 66 37 75 44
52 70 10 83 37	56 30 38 73 15	16 52 06 96 76	11 65 49 98 93	02 18 16 81 61
57 27 53 68 98	81 30 44 85 85	68 65 22 73 76	92 85 25 58 66	88 44 80 35 84
20 85 77 31 56	70 28 42 43 26	79 37 59 52 20	01 15 96 32 67	10 62 24 83 91
15 63 38 49 24	90 41 59 36 14	33 52 12 66 65	55 82 34 76 41	86 22 53 17 04
92 69 44 82 97	39 90 40 21 15	59 58 94 90 67	66 82 14 15 75	49 76 70 40 37
77 61 31 90 19	88 15 20 00 80	20 55 49 14 09	96 27 74 82 57	50 81 69 76 16
38 68 83 24 86	45 13 46 35 45	59 40 47 20 59	43 94 75 16 80	43 85 25 96 93
25 16 30 18 89	70 01 41 50 21	41 29 06 73 12	71 85 71 59 57	68 97 11 14 03
65 25 10 76 29	37 23 93 32 95	05 87 00 11 19	92 78 42 63 40	18 47 76 56 22
36 81 54 36 25	18 63 73 75 09	82 44 49 90 05	04 92 17 37 01	14 70 79 39 97
64 39 71 16 92	05 32 78 21 62	20 24 78 17 59	45 19 72 53 32	83 74 52 25 67
04 51 52 56 24	95 09 66 79 46	48 46 08 55 58	15 19 11 87 82	16 93 03 33 61
83 76 16 08 73	43 25 38 41 45	60 83 32 59 83	01 29 14 13 49	20 36 80 71 26
14 38 70 63 45	80 85 40 92 79	43 52 90 63 18	38 38 47 47 61	41 19 63 74 80
51 32 19 22 46	80 08 87 70 74	88 72 25 67 36	66 16 44 94 31	66 91 93 16 78
72 47 20 00 08	80 89 01 80 02	94 81 33 19 00	54 15 58 34 36	35 35 25 41 31
05 46 65 53 06	93 12 81 84 64	74 45 79 05 61	72 84 81 18 34	79 98 26 84 16
39 52 87 24 84	82 47 42 55 93	48 54 53 52 47	18 61 91 36 74	18 61 11 92 41
81 61 61 87 11	53 34 24 42 76	75 12 21 17 24	74 62 77 37 07	58 31 91 59 97
07 58 61 61 20	82 64 12 28 20	92 90 41 31 41	32 39 21 97 63	61 19 96 79 40
90 76 70 42 35	13 57 41 72 00	69 90 26 37 42	78 46 42 25 01	18 62 79 08 72
40 18 82 81 93	29 59 38 86 27	94 97 21 15 98	62 09 53 67 87	00 44 15 89 97
34 41 48 21 57	86 88 75 50 87	19 15 20 00 23	12 30 28 07 83	32 62 46 86 91
63 43 97 53 63	44 98 91 68 22	36 02 40 09 67	76 37 84 16 05	65 96 17 34 88
67 04 90 90 70	93 39 94 55 47	94 45 87 42 84	05 04 14 98 07	20 28 83 40 60
79 49 50 41 46	52 16 29 02 86	54 15 83 42 43	46 97 83 54 82	59 36 29 59 38
91 70 43 05 52	04 73 72 10 31	75 05 19 30 29	47 66 56 43 82	99 78 29 34 78

A.3 - SUBJECT RECRUITMENT MATERIALS

The next six sub-sections contain reduced copies of samples of the following recruitment materials:

- A.3.1 Subject Participation Handout
- A.3.2 Schedule Participation Sheet
- A.3.3 Recruitment Notice
- A.3.4 List of locations for placement of Recruitment Notices
- A.3.5 Map of locations for placement of Recruitment Notices

A.3.1 - Subject Participation Handout

SUBJECT PARTICIPATION AND SCHEDULING
Experiment on Quantitative Data Displays

NATURE OF THE EXPERIMENT

People handle quantitative data in different ways, reflecting their aptitudes and relative skills learned for manipulation of that type of data. The present experiment is an attempt to better understand the effect of different data presentation modes (graphic vs. numeric) on how people analyse quantitative data. The experiment consists of manipulating quantitative data in an interactive computer system, under different data displays. Eighty (80) Subjects are needed for this experiment, and will be accepted on a "first-come, first-test" basis.

TIME REQUIREMENTS

The duration of an experimental session ranges from one (1) to two (2) hours, with most people taking one and a half (1 1/2) hours. In principle, two hours should be set aside for participation in an experimental session. The available times are 17:15, 19:30, and 21:30 hrs, seven days per week, as well as 9:00, 11:00, 13:00, and 15:00 hrs on weekends.

Before scheduling your participation, avoid selecting times when you might feel too tired, sleepy, hungry, or too pressed for time afterwards. It is best to allocate an extra fifteen minutes for unexpected changes in schedule, like those due to computer malfunctioning.

CONFIDENTIALITY OF DATA

All data collected relative to each Subject will be used exclusively for research purposes. Only modified data, without possibility of identification of individual subjects, will be retained for final report (dissertation) and eventual publication.

PARTICIPATION STIPEND

Each Subject will receive a total of up to US $ 8.00 upon completion of an experimental session. For participation in the experiment, there will be a $ 7.00 payment. For timely participation, i.e., no missed dates or late arrivals (more than 10 minutes), there will be an extra bonus of $ 1.00.

No charges or bonuses will be in effect if rescheduling of parts of the experiment becomes necessary due to circumstances beyond the Subject's or the test administrator's control, as computer malfunctioning, illness, and so on.

APPLICATION FORM

Any students from Virginia Tech (undergraduate and graduate) are encouraged to participate. If interested, please fill out the enclosed form and return it as soon as possible to the **FORM BOX** at **201 Architectural Annex** (c/o Nora Bentley) or at **202-A Cowgill Hall** (c/o Kim O'Rourke or Janice Huffman). For further information please call FRED LACERDA (951-0080) or leave messages w/ Nora Bentley (961-5387), from 9:00 to 5:00, with reference to "Data Display Experiment".

Note: The Architectural Annex (the Old High School Bldg) is located behind the Donaldson-Brown Center and Squires Student Center. Cowgill Hall (the College of Architecture) is located behind Burruss Hall.

A.3.2 - Schedule Participation Sheet

SUBJECT APPLICATION FORM
Experiment on Quantitative Data Displays

Name:

Phone # 1: Best time(s) to contact:

Phone # 2: Best time(s) to contact:

POSSIBLE TIME CONFLICTS (Mark with a cross conflicting hours)

Time of Day	Sat	Sun	Mon	Tue	Wed	Thu	Fri
9:00	.	.	.				
10:	.	.	.				
11:	.	.	.				
12:	.	.	.				
13:	.	.	.				
14:	.	.	.				
15:	.	.	.				
16:	.	.	.				
17:00	.	.	.	.	.	.	.
18:	.	.	.	.	.	.	.
19:	.	.	.	.	.	.	.
20:	.	.	.	.	.	.	.
21:	.	.	.	.	.	.	.
22:00	.	.	.	.	.	.	.

(cross conflicting day(s))

MARCH 1986	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
	29	30	31				

APRIL 1986				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	29	30		

SCHEDULED DATE(S) (Do not write below this line)

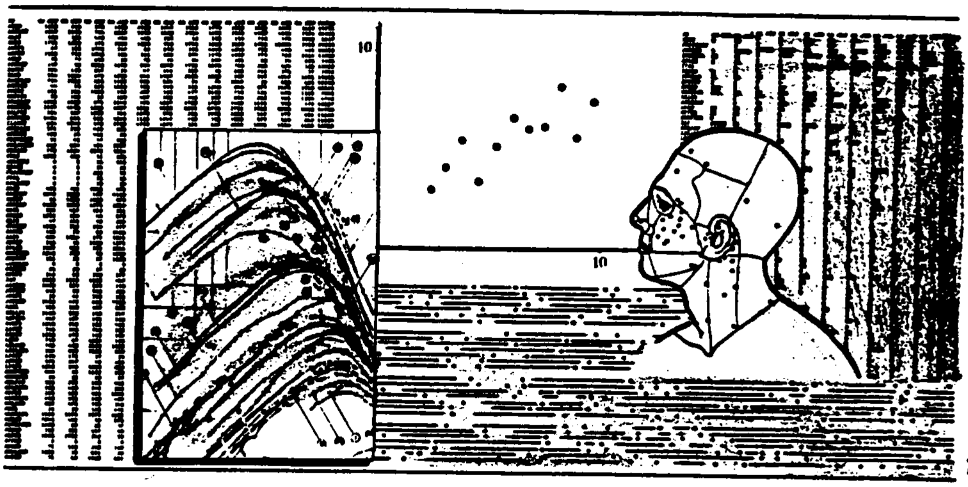
. at : . at : . at :

A.3.3 - Recruitment Notice

Please do not remove this notice before April 15, 1986

DATA DISPLAY

Please do not remove this notice before April 15, 1986



Subjects needed for experiment on the use of different modes of data display.

- **Requirements: Virginia Tech students, at least 18 yrs. old**
- **Time duration: approximately 2 hrs (± 15 min)**
- **PAYMENT: $ 7.00 + $ 1.00 bonus (for punctuality)**

If interested call FRED LACERDA (951-0080) or leave messages at 201-A Architectural Annex (in front of Donaldson-Brown Center) c/o NORA BENTLEY (961-5387).

Data Display Experiment
FRED LACERDÁ (951-0080)
or msg. c/o:
NORA BENTLEY (961-5582)

Data Display Experiment
FRED LACERDÁ (951-0080)
or msg. c/o:
NORA BENTLEY (961-5582)

Data Display Experiment
FRED LACERDÁ (951-0080)
or msg. c/o:
NORA BENTLEY (961-5582)

Data Display Experiment
FRED LACERDÁ (951-0080)
or msg. c/o:
NORA BENTLEY (961-5582)

Data Display Experiment
FRED LACERDÁ (951-0080)
or msg. c/o:
NORA BENTLEY (961-5582)

Data Display Experiment
FRED LACERDÁ (951-0080)
or msg. c/o:
NORA BENTLEY (961-5582)

Data Display Experiment
FRED LACERDÁ (951-0080)
or msg. c/o:
NORA BENTLEY (961-5582)

Data Display Experiment
FRED LACERDÁ (951-0080)
or msg. c/o:
NORA BENTLEY (961-5582)

Data Display Experiment
FRED LACERDÁ (951-0080)
or msg. c/o:
NORA BENTLEY (961-5582)

Data Display Experiment
FRED LACERDÁ (951-0080)
or msg. c/o:
NORA BENTLEY (961-5582)

Data Display Experiment
FRED LACERDÁ (951-0080)
or msg. c/o:
NORA BENTLEY (961-5582)

Data Display Experiment
FRED LACERDÁ (951-0080)
or msg. c/o:
NORA BENTLEY (961-5582)

A.3.4 - List of Locations for Recruitment Notices

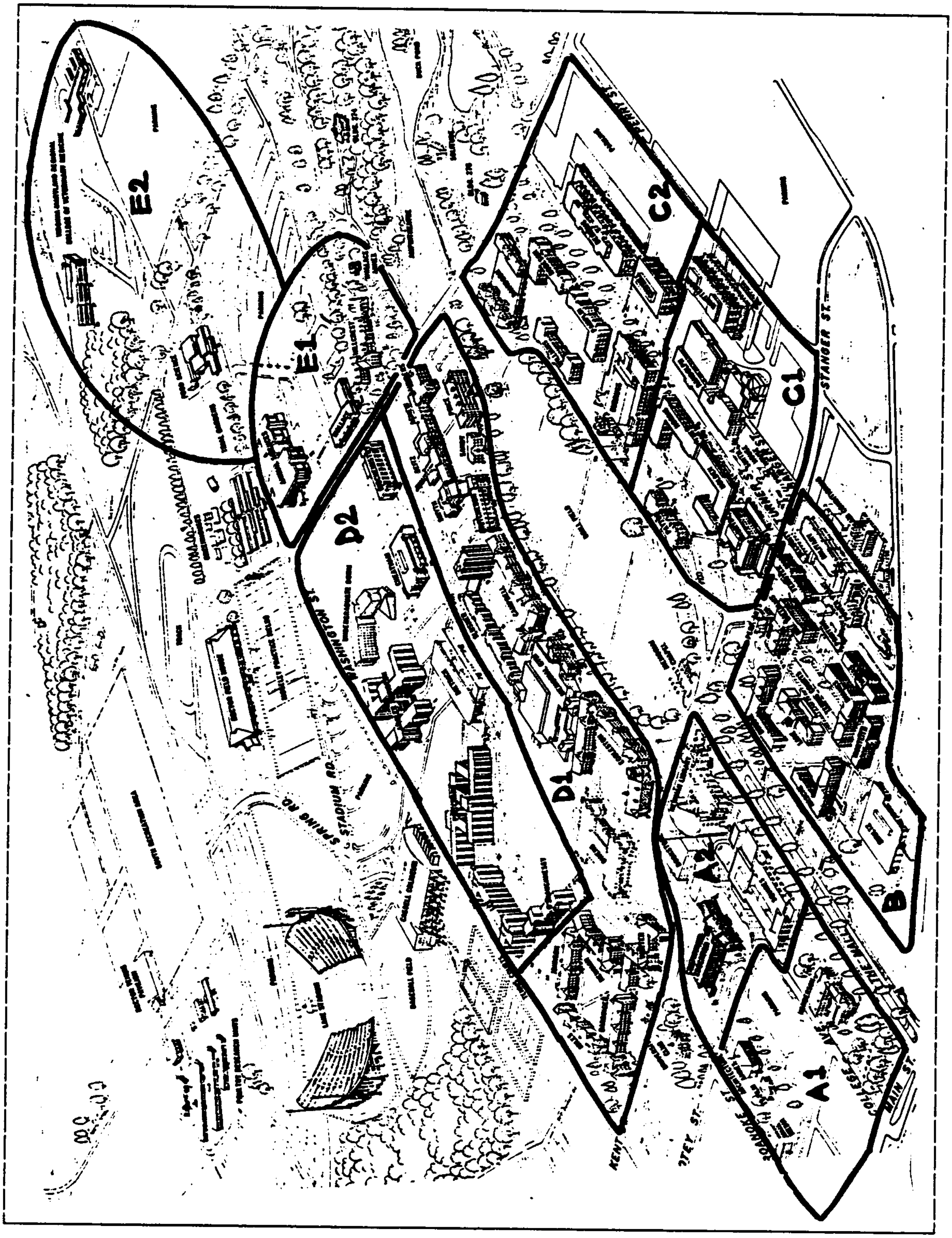
SECTOR — 86.03.10. Mon

SECTOR	LOCATION		DEPT.
A 11	DRAPER ROAD		MISH MISH
A 12	DRAPER ROAD		BOOK, STRING, AND THINGS
A 13	HENDERSON	103	STUDENT HEALTH SERVICES
A 14	COLLEGE AVENUE		KINKO'S
A 15	OLD BLACKSBURG ARMORY		ART AND ART HISTORY
A 16	MEDIA	101-E	EDUCATIONAL COMMUNICATIONS
A 17	ARCHITECTURAL ANNEX	201	ENVIRONMENTAL DESIGN AND PLANNING
A 17	ARCHITECTURAL ANNEX	201	LANDSCAPE ARCHITECTURE
A 17	ARCHITECTURAL ANNEX	201	URBAN AFFAIRS
A 17	ARCHITECTURAL ANNEX	201	URBAN AND REGIONAL PLANNING
A 21	SQUIRES		SQUIRES CAFETERIA
A 21	SQUIRES		SQUIRES DINING HALL
A 21	SQUIRES		SQUIRES STUDENT CENTER
A 21	SQUIRES	321-A	GRADUATE STUDENT ASSEMBLY
A 22	CAROL M. NEWMAN LIBRARY		LIBRARY
A 23	BOOKSTORE BUILDING		UNIVERSITY BOOKSTORE
B 11	SCHULTZ		SCHULTZ DINING HALL
B 12	BRODIE	143	MILITARY AFFAIRS
B 13	PERFORMING ARTS	201	ARTS MANAGEMENT INSTITUTE
B 13	PERFORMING ARTS	201	PERFORMING ARTS
B 13	PERFORMING ARTS	203	THEATRE ARTS
B 13	PERFORMING ARTS	203	UNIVERSITY THEATRE
B 14	FEMOYER	320	SYSTEMS RESEARCH CENTER
B 14	FEMOYER	419	ROTC, NAVY
B 15	MILITARY	226	ROTC, ARMY
B 15	MILITARY	228-A	ROTC, AIR FORCE
C 11	PATTON	120	MATERIALS ENGINEERING SCIENCE
C 11	PATTON	200	CIVIL ENGINEERING
C 11	PATTON	301	GEOGRAPHY
C 11	PATTON	309	PHILOSOPHY
C 12	McBRYDE	460	MATHEMATICS
C 12	McBRYDE	560	HISTORY
C 12	McBRYDE	562	COMPUTER SCIENCE
C 12	McBRYDE	610	INTERNATIONAL STUDIES
C 12	McBRYDE	660	SOCIOLOGY
C 12	McBRYDE	662	POLITICAL SCIENCE
C 13	HOLDEN	104	INSTITUTE FOR MAT. SCI. AND ENGINEERING
C 13	HOLDEN	108-A	FOREIGN LANGUAGES AND LITERATURES
C 13	HOLDEN	201	MATERIALS ENGINEERING
C 13	HOLDEN	213	MINING AND MINERALS ENGINEERING
C 14	NORRIS	225	ENGINEERING SCIENCE AND MECHANICS
C 14	NORRIS	337	ENGINEERING, COLLEGE OF
C 14	NORRIS	337	MINING AND MINERAL RESOURCES RESEARCH INSTITUTE
C 15	RANDOLPH	114	MECHANICAL ENGINEERING
C 15	RANDOLPH	133	CHEMICAL ENGINEERING
C 15	RANDOLPH	215	AEROSPACE AND ENGINEERING
C 15	RANDOLPH	332	ENGINEERING FUNDAMENTALS
C 16	WHITTEMORE	302	INDUSTRIAL ENGINEERING AND OPERATIONS RESEARCH
C 16	WHITTEMORE	340	ELECTRICAL ENGINEERING

List of Locations (cont.)

C 21	COWGILL	201	ARCHITECTURE
C 21	COWGILL	202	ARCHITECTURE AND URBAN STUDIES
C 22	DERRING	2125	BIOLOGY
C 22	DERRING	4044	GEOLOGICAL SCIENCES
C 22	DERRING	5088	PSYCHOLOGY
C 23	PAMPLIN	107	FINANCE, INSURANCE, AND BUSINESS LAW
C 23	PAMPLIN	107	MANAGEMENT SCIENCE
C 23	PAMPLIN	116	BUSINESS, COLLEGE OF
C 23	PAMPLIN	207	MANAGEMENT
C 23	PAMPLIN	207	MARKETING
C 23	PAMPLIN	307	ACCOUNTING
C 24	ROBESON	323-A	PHYSICS
C 25	WILLIAMS	126	ARTS AND SCIENCES, COLLEGE OF
C 25	WILLIAMS	200	ENGLISH
C 25	WILLIAMS	226	LIBERAL ARTS AND SCIENCES
C 26	DAVIDSON	107	CHEMISTRY
D 11	OWENS		OWENS DINING HALL
D 11	OWENS	20	BUILDING CONSTRUCTION
D 12	MEMORIAL GYMNASIUM	125-E	RECREATIONAL SERVICES
D 12	MEMORIAL GYMNASIUM	226	EDUCATION, COLLEGE OF
D 13	HUTCHESON	104-H	AGRICULTURE AND LIFE SCIENCES, COLLEGE OF
D 13	HUTCHESON	208-B	AGRICULTURAL ECONOMICS
D 13	HUTCHESON	406-A	STATISTICS
D 14	SANDY	100	GRADUATE SCHOOL
D 14	SANDY	206	ECONOMICS
D 15	PRICE	216	ENTOMOLOGY
D 15	PRICE	413	PLANT PATHOLOGY, PHYSIOLOGY, AND WEED SCIENCE
D 16	SAUNDERS	301	HORTICULTURE
D 17	AGNEW	11	COMMUNICATION STUDIES
D 18	SEITZ	200	AGRICULTURAL ENGINEERING
D 19	SMYTH	332	AGRONOMY
D 21	BIOCHEMISTRY	113	BIOCHEMISTRY AND NUTRITION
D 22	CHEATHAM	324	FORESTRY AND WILDLIFE RESOURCES
D 23	DIETRICK		DIETRICK DINING HALL
D 24	UNDERGRADUATE DORM		NEW UNDERGR. DORM DINING.HALL
E 11	HILLCREST		HILLCREST DINING HALL
E 12	WALLACE	103	CLOTHING AND TEXTILES
E 12	WALLACE	216	HUMAN RESOURCES, COLLEGE OF
E 12	WALLACE	218	HOUSING, INTERIOR DESIGN, AND RESOURCE MANAGEMENT
E 12	WALLACE	301	HUMAN NUTRITION AND FOODS
E 13	WALLACE ANNEX	102	FAMILY AND CHILD DEVELOPMENT
E 14	ANIMAL SCIENCES	2270	POULTRY SCIENCE
E 14	ANIMAL SCIENCES	2470	DAIRY SCIENCES
E 14	ANIMAL SCIENCES	3460	ANIMAL SCIENCE
E 21	FOOD SCIENCE	22	FOOD SCIENCE AND TECHNOLOGY
E 22	PHASE II		VETERINARY MEDICINE, COLLEGE OF
F 11	NORTH MAIN STREET		EATS FOOD COOPERATIVE
F 12	NORTH MAIN STREET		WADE'S SUPERMARKET
F 13			KROGER'S SUPERMARKET

A.3.5 - Map of Locations for Recruitment Notices



APPENDIX B. EXPERIMENTAL SESSION-RELATED MATERIALS

Contents:

B.1 - QUESTIONNAIRES FOR EXPERIMENTAL SESSIONS

The next two pages contain reduced copies of the two questionnaires presented to Subjects during experimental sessions. The first one (section B.1.1) was presented at the beginning of experimental sessions; its questions deal with previous experience in selected areas and with selected demographic variables. The second questionnaire (section B.1.2) was presented at the end of experimental sessions; its questions deal with Subject perceptions of the experimental sessions per se.

B.1.1 - Questionnaire No. 1 - Subject Data

SN ___ - ___ - ___

QUESTIONNAIRE NO. 1
Experiment on Quantitative Data Displays

Please take a couple of minutes to answer the questions below, which will remain confidential. The purpose of these questions is help form a composite picture, in terms of selected factors, of the group of Subjects volunteering to take part in this experiment.

1. What is your age? __________ And sex (M/F)? _____

2. How many years of higher education have you had? ________

3. In which program are you presently enrolled?
   ____

4. What is your academic level? [Circle one]

   Undergr. Master's Doctorate Extension Other:_______

5. How do you rank your familiarity with computers (as a user)?

   [circle one] 1 2 3 4 5 6 7
   low high

6. How do you rate your familiarity with statistics, both in terms of courses you took and experience in using statistics in your work or research?

   [circle one] 1 2 3 4 5 6 7
   low high

7. How did you find out about this experiment?

   ( ) Saw recruitment notice (Where? ________________________)
   ( ) Word-of-mouth
   ( ) Other (______________________________________)

B.1.2 - Questionnaire No. 2 - Session Evaluation

SN ___ - ___ - ___

QUESTIONNAIRE NO. 2
Experiment on Quantitative Data Displays

Please take a couple of minutes to answer the questions below. The purpose of these questions is to help in the evaluation of the design of the experimental session you are about to finish. This questionnaire ends your participation in this experiment.

1. How do you rate your enjoyment of the present experiment (both the tutorial and the hotel problem analysis) ?
   (Don't be shy, tell what you think...)

   [circle one] 1 2 3 4 5 6 7
   low ... high

2. Do you have any comments on the TUTORIAL (for instance, amount of information presented, sequence of screens, legibility of screens, and so on)?

3. Do you have any comments on the hotel PROBLEM ANALYSIS per se?

4. Do you have any last comments or observations to make?

Thank you very much for your participation!

B.2 - TUTORIAL SECTION OF EXPERIMENTAL SESSIONS

A reduced copy of a tutorial instruction sheet is presented in the next page (section B.2.1). The following page (section B.2.2) contains a reduced copy of a tutorial worksheet. A reduced copy of the supporting information for both the tutorial and problem analysis sections of experimental sessions is presented in the two subsequent pages. The first information sheet contains a summary of "Help" information for all computer system commands (section B.2.3). The second sheet contains a complete list of paradigm problem field names and descriptions (section B.2.4).

The last six pages (section B.2.5) contain the list of instructions for the "Hands-On" tutorial. Since the "Hands-On" tutorial sheets measure 11 by 11 inches, a sixty five percent reduction of each whole sheet is shown first, before each of the sheet's two component text columns is presented. The screens for the "Information Tutorial", which actually precede the "hands-on" tutorial, are presented in Appendix D.

B.2.1 - Tutorial Instructions

SN __ - __ - __

TUTORIAL INSTRUCTIONS
Experiment on Quantitative Data Displays

All information necessary to help in the analysis of the HOTEL problem, to be explained later on, will be made available in the screens of the TUTORIAL.

The tutorial screens will introduce, in a pre-arranged sequence, information about the system and the commands made available to support your problem analysis. Please remember that, ultimately, you remain the only judge of the usefulness of those commands and of how appropriate they are for your analysis. Approximately 20-25 min should be enough for both skimming through the few lines of text of the information tutorial (with approximately 20 screens) and for a hands-on introduction to the system.

Most of the information presented in the tutorial will be available, at any time, during problem analysis. You will be able to access HELP information about any command at, literally, the touch of a button. Therefore, you should spend just as much time in the tutorial as needed to familiarize yourself with the basic concepts and command description.

The two sheets accompanying these instructions consist of supporting information for both the tutorial and the "Hotel" problem analysis. The first information sheet contains a summary of all the commands available in the system; the second sheet contains the names and description of all the variables that will characterize the "Hotel" problem.

If you have any questions, please ask them now.

B.2.2 - Tutorial Worksheet

TUTORIAL WORKSHEET
Experiment on Quantitative Data Displays

SN ___ - ___ - ___

B.2.3 - "Help" Information Sheet

SN __ - __ - __

SYSTEM COMMANDS
Experiment on Quantitative Data Display

WINDOW.... If a table is larger than it can fit on screen, use the arrow keys to move the present window view of the data towards the direction desired. A maximum of 12 fields (columns) and 16 records (rows) can be seen on screen at any one time, out of a possible maximum of 20 fields and 24 records.

MENU...... The MENU list allows you to choose any of the ten commands below for execution by the system. With the help of the arrow keys, simply move the blinking cursor to the desired command and press the [ENTER] key.

DISPLAY... Displays any table stored in this system, including "HOTEL" and any others created by the user. Use reference number listed to the left of each name in list generated above the messages area.

RETURN.... Returns to any previous table on display (up to a maximum of 10), either saved or not.

SAVE...... Saves a new table (produced with the aid of other commands) under name chosen by user (up to 12 alphanumeric characters)

DROP...... Drops from the system any named table on display, with the exception of "HOTEL" (the basic dataset for the problem) which is displayed once more immediately after a successful DROP.

NAMES..... Generates a name list with all table names currently available (i.e., saved) under the system.

FIELDS.... Generates a list with description of all fields presently on screen.

PROJECT... From any table in display, it creates a new table containing only those fields (columns) specified by user.

SELECT.... From any table in display, it creates a new table containing the same original fields but only a subset of the records (rows) from any given field , specified by the user, according to a selection operation also user-specified (direct selection of records; or records greater-than, greater-than-or-equal-to, less-than, less-than-or-equal-to, or equal to specified record).

SORT...... From any table in display, it creates a new table containing the same original fields (columns) and records (rows), but with records rearranged according to sorting (ascending or descending order) of an user-specified field.

HELP...... Presents a summary of information for each command listed above, as well as for the data representation scheme adopted for this problem.

B.2.4 - Field Name and Description Sheet

SN ___ - ___ - ___

FIELDS (VARIABLES) OF THE "HOTEL" PROBLEM
Experiment on Quantitative Data Display

	Code	Description	Unit/Measurement
M	MONTHS	Months of past two years of Hotel operations	
1	FEM	Female guests	(% total no.guests)
2	LOCAL	Local (state) guests	(% total no.guests)
3	USA	Out-of-state guests	(% total no.guests)
4	EUR	European guests	(% total no.guests)
5	SOU	South-American guests	(% total no.guests)
6	AFR	African/Middle-Eastern guests	(% total no.guests)
7	ASI	Asian guests	(% total no.guests)
8	BUS	Business guests	(% total no.guests)
9	TOU	Tourist guests	(% total no.guests)
10	DIR	Direct reservations	(% total no.guests)
11	AGR	Agency reservations	(% total no.guests)
12	AIR	Air crews (permanent reservation)	(% total no.guests)
13	<20	Guests under 20 years of age	(% total no.guests)
14	20-34	Guests bet. 20 and 34 years of age	(% total no.guests)
15	35-54	Guests bet. 35 and 54 years of age	(% total no.guests)
16	>54	Guests above 54 years of age	(% total no.guests)
17	PRICE	Price of rooms	(U.S. dollar)
18	STAY	Length of stay	(Day)
19	OCC	Occupancy rate	(% total no. rooms)
20	CONV.	Conventions	(1=yes; 0=no)

B.2.5 - "Hands-On" Tutorial Sheets

SN ___ - ___ - ___

HANDS-ON TUTORIAL
Experiment on Quantitative Data Displays

COMMAND	HOW TO DO IT	WHAT WILL HAPPEN
		Note that the first field, after "Months", is no. 1, "FEM"; and that the first record, using the "Months" field as a reference, is "84.Jan", i.e., January of 1984.
WINDOW	☐ Press [→] once.	Now the first field, following "MONTHS", will be no. 2 ("LOC", i.e., "local guests").
	☐ Press [→] seven more times until you hear a bell sound.	The last field on the right side will now be no. 20 ("CONV", i.e., "conventions").
	☐ Press [→] a couple more times.	The window screen will not move and the bell sound will remind you that you have reached the right end of the table.
	☐ Press [←] eight times, until you hear a bell sound again.	You will have reached again the left-side limit of the table (the starting position for each table).
	☐ Press [↓] eight times, until you hear a bell.	Now you will have reached the bottom of the table. The last record displayed (the bottom row) corresponds to December 1985.
	☐ Press [↑] eight times, until you hear a bell.	Now you will have reached again the top of the table (the starting position for each table).
MENU	☐ Press [ENTER].	From the present WINDOW DISPLACEMENT screen you can request the MENU options screen The MENU options screen will be displayed.
HELP	☐ Move cursor (with the help of any of the arrow keys) to the HELP menu option and press [ENTER]. ☐ Press any key to return to WINDOW DISPLACEMENT. ☐ Press [ENTER] to return to MENU options screen.	The specific HELP information requested will be displayed, superimposed on the table display area.
PROJECT	☐ Move MENU cursor to PROJECT option and press [ENTER].	A message requesting you to choose fields for projection will be displayed.
	☐ Move cursor (with either the [→] or [←] keys) to field ___ and press the [TAB] key. ☐ Move cursor again to field ___ and press [TAB] key. ☐ Repeat for fields ___, ___, and ___. ☐ After all your field choices have been highlighted, press [ENTER].	The resulting table will now have only five fields, with the list of projected fields indicated in the messages area.
	☐ Press any key to return to WINDOW DISPLACEMENT, and press [ENTER] to return to the MENU options.	
FIELDS	☐ With cursor at the FIELDS option, press [ENTER]. ☐ Press [ENTER] twice to return to MENU options.	A descriptive list of the five fields on screen will be displayed.

HANDS-ON TUTORIAL
Experiment on Quantitative Data Displays

COMMAND	HOW TO DO IT	WHA
		Note that the first fiel and that the first re reference, is "84.jan",
WINDOW	☐ Press [→|] once.	Now the first field, fo ("LOC", i.e., "local g
	☐ Press [→|] seven more times until you hear a bell sound.	The last field on th ("CONV", i.e., "conve
	☐ Press [→|] a couple more times.	The window screen remind you that you table.
	☐ Press [|←] eight times, until you hear a bell sound again.	You will have reach table (the startin
	☐ Press [↓] eight times, until you hear a bell.	Now you will have record displayed 1985.
	☐ Press [↑] eight times, until you hear a bell.	Now you will starting pos
MENU	☐ Press [ENTER].	From the pres the MENU op The MENU option
HELP	☐ Move cursor (with the help of any of the arrow keys) to the HELP menu option and press [ENTER].	The specific H superimposed o
	☐ Press any key to return to WINDOW DISPLACEMENT.	
	☐ Press [ENTER] to return to MENU options screen.	
PROJECT	☐ Move MENU cursor to PROJECT option and press [ENTER].	A message will be di
	☐ Move cursor (with either the [→|] or [|←] keys) to field ____ and press the [TAB] key.	
	☐ Move cursor again to field ____ and press [TAB] key.	
	☐ Repeat for fields ____, ____, and ____.	
	☐ After all your field choices have been highlighted, press [ENTER].	The re the area.
	☐ Press any key to return to WINDOW DISPLACEMENT, and press [ENTER] to return to the MENU options.	
FIELDS	☐ With cursor at the FIELDS option, press [ENTER].	A d dis
	☐ Press [ENTER] twice to return to MENU options.	

SN __ - __ - __

HANDS-ON TUTORIAL
Experiment on Quantitative Data Displays

	WHAT WILL HAPPEN
	Note that the first field, after "Months", is no. 1, "FEM"; and that the first record, using the "Months" field as a reference, is "84.Jan", i.e., January of 1984.
til you hear a bell	Now the first field, following "MONTHS", will be no. 2 ("LOC", i.e., "local guests").
.	The last field on the right side will now be no. 20 ("CONV", i.e., "conventions").
til you hear a bell	The window screen will not move and the bell sound will remind you that you have reached the right end of the table.
l you hear a bell.	You will have reached again the left-side limit of the table (the starting position for each table).
il you hear a bell.	Now you will have reached the bottom of the table. The last record displayed (the bottom row) corresponds to December 1985.
	Now you will have reached again the top of the table (the starting position for each table).
	From the present WINDOW DISPLACEMENT screen you can request the MENU options screen
	The MENU options screen will be displayed.
he help of any of the arrow ...nd press [ENTER].	The specific HELP information requested will be displayed, superimposed on the table display area.
INDOW DISPLACEMENT.	
to MENU options screen.	
PROJECT option and press	A message requesting you to choose fields for projection will be displayed.
er the [→] or [←] keys) to he [TAB] key.	
ield ____ and press [TAB]	
__, ____, and ____.	
field choices have been ENTER].	The resulting table will now have only five fields, with the list of projected fields indicated in the messages area.
turn to WINDOW DISPLACEMENT, to return to the MENU options.	
FIELDS option, press [ENTER].	A descriptive list of the five fields on screen will be displayed.
ce to return to MENU options.	

HANDS-ON TUTORIAL (cont.)

SN ___ - ___ - ___

COMMAND	HOW TO DO IT	WHAT WILL HAPPEN
SORT	☐ Move cursor to SORT option and press [ENTER].	
	☐ Move cursor to field ____ and press [TAB].	After the field of your choice is highlighted, a small screen with the two sorting options (ascending and descending orders) will be presented.
	☐ Move cursor (with either the [→] or [←] keys) to "descending" order and press [ENTER].	The resulting table--with the same original number of fields--will now present a new order for the table records, based on the rearrangement of the specified field.
	☐ Press [ENTER] twice to return to MENU.	
SELECT	☐ Move cursor to SELECT option and press [ENTER].	
	☐ Move cursor to field ____ and press [TAB].	After the field of your choice is highlighted, a small screen with the six selection options will be presented.
	☐ Move cursor to less than/greater then option and press [ENTER].	
	☐ Move cursor (with either the [↓] or [↑] keys) to largest/smallest record among those presently displayed on screen, and press [TAB].	The resulting table will now display only those records less than/greater than specified record from chosen field.
	☐ Press [ENTER] twice to return to MENU.	
NAMES	☐ Move cursor to NAMES option and press [ENTER].	The list of available (i.e., saved) tables will be displayed (in this case, it will consist only of the original table, "HOTEL").
	☐ Press [ENTER] twice to return to MENU.	
SAVE	☐ Move cursor to SAVE option and press [ENTER].	A message will be displayed, requesting you to type the name under which you want the table to be stored.
	☐ Type one or more alphanumeric characters as the table name; press [ENTER] when finished.	The new name will be displayed in the table name area.
	☐ Press [ENTER] twice to return to MENU.	
RETURN	☐ Move cursor to RETURN option and press [ENTER].	
	☐ Press [TAB] once.	The previous table on display (indicate by the blinking number "-1") will be displayed again, temporarily.
	☐ Press [TAB] twice.	The blinking number "-3" will indicate that the "HOTEL" table is the third previous one.
	☐ Press [ENTER].	Now the first table, "HOTEL", will be effectively substituted for the starting table.
	☐ Press [ENTER] twice to return to MENU.	
DISPLAY	☐ Move cursor to DISPLAY option and press [ENTER].	A message will be displayed, requesting the number of the table to be brought back to the screen.
	☐ Type "1" (corresponding to the name of the table you previously saved) and press [ENTER].	The previously saved table will be once more displayed.
	☐ Press [ENTER] twice to return to MENU.	
DROP	☐ Move cursor to DROP option and press [ENTER].	
	☐ To confirm an intention of dropping the table in question from the table name list and screen, press [ENTER].	After any table is dropped from the system, the "HOTEL" table will automatically be recalled to the screen.
	☐ Press [ENTER] twice to return to MENU.	
	☐ Move cursor to NAMES option and press [ENTER].	The table name list will now have only one name on it, after the successful completion of the DROP operation.

After you are finished with this part of the tutorial, please call the test administrator.

HANDS-ON TUTORIAL (cont.)

COMMAND	HOW TO DO IT	WHAT
SORT	☐ Move cursor to SORT option and press [ENTER].	
	☐ Move cursor to field ____ and press [TAB].	After the field of your screen with the two descending orders) will b
	☐ Move cursor (with either the [→] or [←] keys) to "descending" order and press [ENTER].	The resulting table--w fields--will now present based on the rearrangem
	☐ Press [ENTER] twice to return to MENU.	
SELECT	☐ Move cursor to SELECT option and press [ENTER].	
	☐ Move cursor to field ____ and press [TAB].	After the field of y screen with the six
	☐ Move cursor to less than/greater than option and press [ENTER].	
	☐ Move cursor (with either the [↓] or [↑] keys) to largest/smallest record among those presently displayed on screen, and press [TAB].	The resulting tab less than/greate
	☐ Press [ENTER] twice to return to MENU.	
NAMES	☐ Move cursor to NAMES option and press [ENTER].	The list displayed origin
	☐ Press [ENTER] twice to return to MENU.	
SAVE	☐ Move cursor to SAVE option and press [ENTER].	A message name und
	☐ Type one or more alphanumeric characters as the table name; press [ENTER] when finished.	
	☐ Press [ENTER] twice to return to MENU.	The new n
RETURN	☐ Move cursor to RETURN option and press [ENTER].	
	☐ Press [TAB] once.	The previou number "-1"
	☐ Press [TAB] twice.	The blinkin table is th
	☐ Press [ENTER].	Now the substitute
	☐ Press [ENTER] twice to return to MENU.	
DISPLAY	☐ Move cursor to DISPLAY option and press [ENTER].	A message table to
	☐ Type "1" (corresponding to the name of the table you previously saved) and press [ENTER].	The prev
	☐ Press [ENTER] twice to return to MENU.	
DROP	☐ Move cursor to DROP option and press [ENTER].	
	☐ To confirm an intention of dropping the table in question from the table name list and screen, press [ENTER].	After tabl
	☐ Press [ENTER] twice to return to MENU.	
	☐ Move cursor to NAMES option and press [ENTER].	Th a

HANDS-ON TUTORIAL (cont.)

SN ___ - ___ - ___

	WHAT WILL HAPPEN
s [ENTER].	
s [TAB].	After the field of your choice is highlighted, a small screen with the two sorting options (ascending and descending orders) will be presented.
or [←] keys) to ER].	The resulting table--with the same original number of fields--will now present a new order for the table records, based on the rearrangement of the specified field.
o MENU.	
d press [ENTER].	
press [TAB].	After the field of your choice is highlighted, a small screen with the six selection options will be presented.
ater than option and	
[↓] or [↑] keys) to ong those presently ss [TAB].	The resulting table will now display only those records less than/greater than specified record from chosen field.
rn to MENU.	
ress [ENTER].	The list of available (i.e., saved) tables will be displayed (in this case, it will consist only of the original table, "HOTEL").
turn to MENU.	
press [ENTER].	A message will be displayed, requesting you to type the name under which you want the table to be stored.
ters as the d.	The new name will be displayed in the table name area.
rn to MENU.	
nd press [ENTER].	The previous table on display (indicate by the blinkling number "-1") will be displayed again, temporarily.
	The blinking number "-3" will indicate that the "HOTEL" table is the third previous one.
	Now the first table, "HOTEL", will be effectively substituted for the starting table.
rn to MENU.	
ion and press [ENTER].	A message will be displayed, requesting the number of the table to be brought back to the screen.
o the name of the table press [ENTER].	The previously saved table will be once more displayed.
turn to MENU.	
on and press [ENTER].	
of dropping the table in ble name list and screen,	After any table is dropped from the system, the "HOTEL" table will automatically be recalled to the screen.
return to MENU.	
option and press [ENTER].	The table name list will now have only one name on it, after the successful completion of the DROP operation.

B.3 - PARADIGM PROBLEM ANALYSIS

A reduced copy of the problem analysis instructions is presented in the next page (section B.3.1), followed by a reduced copy of the problem worksheets (section B.3.2). Following that, the time-series data (performance of a small hotel) for the paradigm problem is presented in two pages, with the first devoted to representation of that data in numeric mode (section B.3.3) and the second page devoted to their representation in graphic mode (section B.3.4).

B.3.1 - Problem Instructions

SN __ - __ - __

PROBLEM ANALYSIS INSTRUCTIONS
Experiment on Quantitative Data Display

PROBLEM BACKGROUND

Congratulations! You finally got a great summer job at a very pleasant, well located, middle-sized hotel, helping its assistant-manager with office work.

After a few days, you're asked to go through some data--collected and organized by the assistant-manager--relative to the hotel's performance for the last two years. The assistant-manager believes that, between the two of you, some new facts or ideas might come out after analysing that data. The assistant-manager's problem is to help the hotel manager--due back in a couple of days--to decide on what to do in order to improve the hotel's performance for next year.

PROBLEM ANALYSIS

Make yourself comfortable, and start analysing the data on the hotel performance, available through the table named HOTEL in the computer system. Please remember that the most important thing for YOU to do is to figure out what important RELATIONSHIPS you can identify from the data in question. No suggestions (solutions) on how to improve the Hotel performance are requested at this point.

NOTATION OF IDEAS

List the **Observations** you make on the **left** side of the worksheets made available to you. **Inferences** (further conclusions based on one or more of the observations you made) can be listed on the **right** side of the worksheets. Avoid passing judgement on the worth of your ideas; assume that ALL of your IDEAS are IMPORTANT, until you're asked to rank them.

Please write as legibly as possible; it will be quite helpful for the subsequent transcription of your notes. Number your observations and inferences or alternatively indicate, in any way you prefer, the separation between any two observations (or inferences).

Be as precise as you can when making comments about the data; always mention the variables you have in mind when writing your observations and inferences. When indicating a time period, specify the month(s) included in that period.

If you have any questions, please ask them now.

B.3.2 - Problem Worksheet

PROBLEM ANALYSIS WORKSHEET
Experiment on Quantitative Data Display

Page ___/___

SN ___-___-___

OBSERVATIONS

INFERENCES

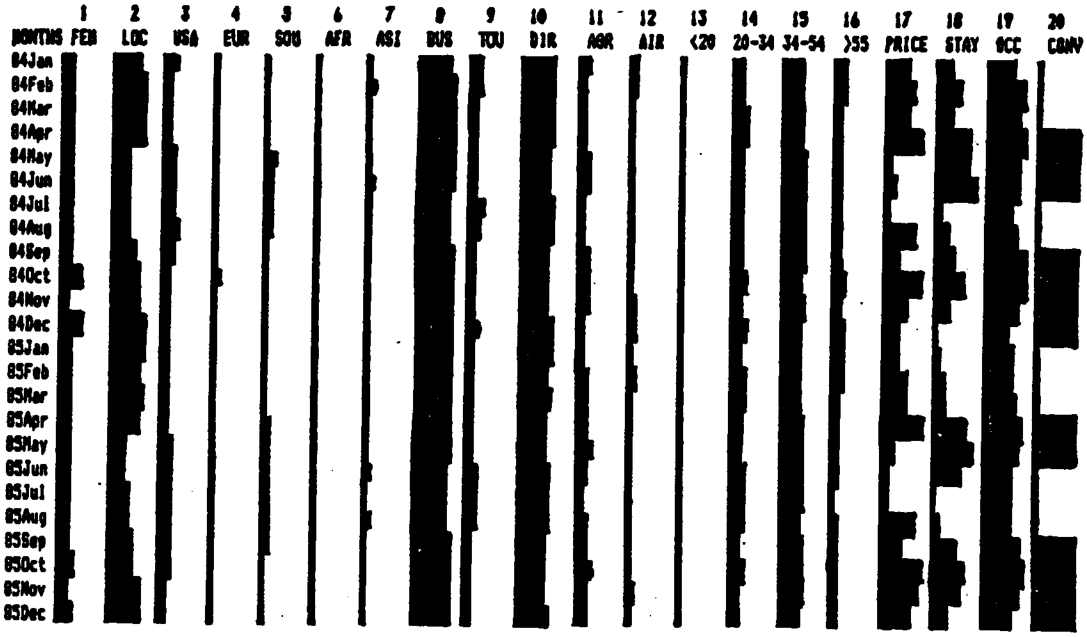
B.3.3 - "Hotel" Time-Series Data: Representation in Numeric Mode

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
MONTHS	FEM	LOC	USA	EUR	SOU	AFR	ASI	BUS	TOU	DIR	AGR	AIR	<20	20-34	34-54	>55	PRICE	STAY	OCC	CONV
84Jan	26	69	20	0	7	1	3	78	22	70	20	10	2	25	48	25	74	1.65	67	0
84Feb	21	70	15	0	6	0	10	80	20	70	18	12	2	27	49	22	76	1.71	82	0
84Mar	26	77	14	0	3	0	6	85	15	75	19	6	4	37	42	17	75	1.65	70	0
84Apr	28	71	15	0	6	8	0	86	14	74	17	9	2	35	48	15	79	1.91	83	1
84May	20	37	23	8	23	6	3	85	15	69	27	4	2	25	54	19	69	1.9	74	1
84Jun	20	36	-27	6	14	4	13	87	13	68	27	5	1	25	55	19	70	2	77	1
84Jul	20	39	22	6	19	6	8	70	30	74	19	7	1	27	53	19	66	1.54	56	0
84Aug	20	39	30	4	14	4	9	76	24	75	19	6	2	28	51	19	77	1.6	62	0
84Sep	20	55	27	-2	9	2	5	87	13	68	26	6	2	24	55	19	71	1.73	90	1
84Oct	40	60	19	12	6	1	2	85	15	68	27	5	4	30	46	20	79	1.82	92	1
84Nov	15	68	19	0	8	0	5	87	13	64	21	15	2	24	55	19	75	1.66	78	1
84Dec	40	72	17	0	8	1	2	80	20	75	15	10	5	30	43	22	72	1.44	55	1
85Jan	28	73	17	2	6	1	1	82	18	74	17	9	4	28	46	22	72	1.55	62	0
85Feb	26	68	18	0	8	1	5	81	19	69	21	10	1	31	48	20	74	1.6	68	0
85Mar	21	75	16	2	3	0	4	82	18	73	22	5	5	34	43	18	74	1.58	71	0
85Apr	26	67	16	3	4	7	3	86	14	68	25	7	3	27	51	19	82	1.88	87	1
85May	20	35	25	9	17	7	7	87	13	66	30	4	1	28	56	15	70	2.03	78	1
85Jun	20	35	28	5	19	3	10	78	22	71	23	6	1	30	52	17	68	1.8	67	0
85Jul	20	42	24	6	15	4	9	72	28	77	14	9	1	34	57	8	68	1.48	60	0
85Aug	20	43	22	3	17	1	14	74	26	71	22	7	2	31	49	18	78	1.54	64	0
85Sep	26	54	29	0	11	2	4	85	15	66	29	5	1	28	52	19	73	1.78	88	1
85Oct	38	58	23	3	5	4	7	88	12	64	31	5	5	32	51	12	81	1.9	92	1
85Nov	22	71	15	1	7	0	6	85	15	62	24	14	3	-29	52	16	79	1.85	73	1
85Dec	35	70	17	2	5	0	6	82	18	71	22	7	6	27	48	19	77	1.65	62	1

Where:

Code	Field Name	Unit
Months	Months of past two years of Hotel operations	
1	Female guests	(% total no.guests)
2	Local (state) guests	(% total no.guests)
3	Out-of-state guests	(% total no.guests)
4	European guests	(% total no.guests)
5	South-American guests	(% total no.guests)
6	African/Middle-Eastern guests	(% total no.guests)
7	Asian guests	(% total no.guests)
8	Business guests	(% total no.guests)
9	Tourist guests	(% total no.guests)
10	Direct reservations	(% total no.guests)
11	Agency reservations	(% total no.guests)
12	Air crews (permanent reservation)	(% total no.guests)
13	Guests under 20 years of age	(% total no.guests)
14	Guests bet. 20 and 34 years of age	(% total no.guests)
15	Guests bet. 35 and 54 years of age	(% total no.guests)
16	Guests above 54 years of age	(% total no.guests)
17	Price of rooms	(U.S. dollar)
18	Length of stay	(Day)
19	Occupancy rate	(% total no. rooms)
20	Conventions	(1=yes; 0=no)

B.3.4 - "Hotel" Time-Series Data: Representation in Graphic Mode



Where:

Code	Field Name	Unit
Months	Months of past two years of Hotel operations	
1	Female guests	(% total no.guests)
2	Local (state) guests	(% total no.guests)
3	Out-of-state guests	(% total no.guests)
4	European guests	(% total no.guests)
5	South-American guests	(% total no.guests)
6	African/Middle-Eastern guests	(% total no.guests)
7	Asian guests	(% total no.guests)
8	Business guests	(% total no.guests)
9	Tourist guests	(% total no.guests)
10	Direct reservations	(% total no.guests)
11	Agency reservations	(% total no.guests)
12	Air crews (permanent reservation)	(% total no.guests)
13	Guests under 20 years of age	(% total no.guests)
14	Guests bet. 20 and 34 years of age	(% total no.guests)
15	Guests bet. 35 and 54 years of age	(% total no.guests)
16	Guests above 54 years of age	(% total no.guests)
17	Price of rooms	(U.S. dollar)
18	Length of stay	(Day)
19	Occupancy rate	(% total no. rooms)
20	Conventions	(1=yes; 0=no)

B.4 - TIME CONTROL SHEET AND VERBAL INSTRUCTIONS

A reduced copy of an experimental session control sheet is presented in the next page (section B.4.1). A reduced copy of a list of the verbal instructions to Subjects is presented in the following pages (section B.4.2).

B.4.1 - Time Control Sheet

SN __ - __ - __

TIME CONTROL SHEET
Experiment on Quantitative Data Display

SESSION STARTED	:		
		Give sheet describing experiment	.
		Give RELEASE FORM	.
		Set WS for testing	.
QUESTIONNAIRE NO. 1	:	Type password to start	.
	:	Type password to end	.
TUTORIAL INFORMATION		Read instructions and info. sheet	.
	:	Type password to start	.
	:	[Subject call - end of tutorial]	.
HANDS-ON TUTORIAL		Read information	.
	:	Type password to start	.
	:	Type password to end	.
PROBLEM ANALYSIS		Read instructions	.
	:	Type password to start	.
10'	:	red line 1	.
20'	:	red line 2	.
30'	:	red line 3	.
40'	:	red line 4	.
50'	:	red line 5	.
	:	Type password to end	.
QUESTIONNAIRE NO. 2			
	:	Give last questionnaire	.
	:	Type password to end session	.
		Hand out payment & RECEIPT form	.
SESSION ENDED	:	Duration: :	

B.4.2 - Verbal Instructions

SESSION ACTIVITIES & VERBAL INSTRUCTIONS
Experiment on Quantitative Data Display

PRELIMINARY ACTIVITIES

[Check session handouts package for correct order and total number of sheets]

1	Subject Participation and Scheduling handout
1	Subject Application Form
1	Time Control Sheet
1	Session Notes
1	Release Form
1	Questionnaire No. 1
1	Tutorial Instructions
1	Tutorial Worksheet
1	System Commands handout
1	Fields (Variables) of the "Hotel" Problem (handout)
1	Hands-On Tutorial (first sheet: "Window" through "Fields")
1	Hands-On Tutorial (second sheet: "Sort" through "Drop")
1	Problem Analysis Instructions
4	Problem Analysis Worksheet
1	Questionnaire No. 2
1	Receipt Form

[Number each appropriate page with Subject Number]

[Attach yellow adhesive notesheet to first Hands-On Tutorial sheet]

[Check other materials needed]

3	No. 2 pencils (sharpened)
1	Red-ink pen
1	Blue-ink pen
1	Black-ink pen

SESSION ACTIVITIES and VERBAL INSTRUCTIONS 1

START SESSION

[Note down TIME in CONTROL sheet]

[Give SUBJ. PART. & SCHED. handout and SUBJ. APPL. form]

[Ask Subject to read handout and fill out form, if necessary]

RELEASE FORM

[Give RELEASE FORM]

Now, based on this handout, please read, date and sign this Form.

[Check DATE and rest of the information in form]

SET WORKSPACE

[Make sure WS is MTXCLR and that there are no pending fns]

[Execute 'xxEXPS TEST']

QUESTIONNAIRE NO. 1

Please fill out this questionnaire.

[Type short pw to start; note down TIME in CONTROL sheet]

[Get questionnaire back]

[Type LONG PASSWORD to end this stage; note down TIME in CONTROL sheet]

[Note down test mode]

[Request five numbers between 1 and 20, inclusive]

[Request one number between 1 and 5, inclusive]

[Note numbers down]

SESSION ACTIVITIES and VERBAL INSTRUCTIONS 2

TUTORIAL

Here are the tutorial instructions, which I'll read aloud with you.

[Hand out the TUTORIAL INSTRUCTIONS]

[Read aloud the TUTORIAL INSTRUCTIONS]

All information necessary to help in the analysis of the HOTEL problem, to be explained later on, will be made available in the screens of the TUTORIAL.

The tutorial screens will introduce, in a pre-arranged sequence, information about the system and the commands made available to support your problem analysis. Please remember that, ultimately, you remain the only judge of the usefulness of those commands and of how appropriate they are for your analysis.

[Ad lib here; reinforce idea that User must decide whether all these goodies are desirable or not]

Approximately 20-25 min should be enough for both skimming through the few lines of text of the information tutorial (with approximately 20 screens) and for a hands-on introduction to the system.

Most of the information presented in the tutorial will be available, at any time, during problem analysis. You will be able to access HELP information about any command at, literally, the touch of a button. Therefore, you should spend just as much time in the tutorial as needed to familiarize yourself with the basic concepts and command description.

The two sheets accompanying these instructions consist of supporting information for both the tutorial and the "Hotel" problem analysis. The first information sheet contains a summary of all the command available in the system; the second sheet contains the names and description of all the variables that will characterize the "Hotel" problem.

If you have any questions, please ask them now.

[Wait for any questions]

Here are some note sheets and pencils, if you need them.

Call me only if the cursor is at the bottom of the screen.

Please remember to press keys slowly, waiting for a new screen to show up before pressing any new key.

[Type **short pw** to start; note down **TIME** in CONTROL sheet]

SESSION ACTIVITIES and VERBAL INSTRUCTIONS 3

Intermediate Operations

[Prepare hands-on tutorial sheets for next stage]

[Write numbers in 1st "hands-on" sheet]

[Circle alternatives for SELECT option in 2nd "hands-on" sheet]

Numeric:	less than/largest/less than
Graphic:	greater than/smallest/greater than

[Write out check for full name given in Release Form]

[Attach check with paper clip to RECEIPT Form]

[Wait for Subject to call at the end of Information Tutorial]

[Note down TIME in CONTROL sheet]

HANDS-ON TUTORIAL

Here are the instructions for the hands-on tutorial.

[Hand out the two sheets for the HANDS-ON TUTORIAL]

Simply follow the directions on these two sheets. The exact things you have to do are listed to the left, under the "HOW TO DO IT" column; comments on what the results will be like are listed under the "WHAT WILL HAPPEN" column.

I will be around to answer brief questions, if necessary, at this stage. In principle, all you really need to do is to follow the "recipe", and there should be no problems. Please use this red pen to check each box after you complete the specified operation.

[Type short pw to start; note down TIME in CONTROL sheet]

[Type short pw to end; note down TIME in CONTROL sheet]

SESSION ACTIVITIES and VERBAL INSTRUCTIONS 4

PROBLEM ANALYSIS

Here are the problem instructions, in this sheet of paper.

[Hand out the PROBLEM INSTRUCTIONS]

[Read aloud the PROBLEM INSTRUCTIONS]

PROBLEM BACKGROUND

Congratulations! You finally got a great summer job at a very pleasant, well located, middle-sized hotel, helping its assistant-manager with office work.

After a few days, you're asked to go through some data--collected and organized by the assistant-manager--relative to the hotel's performance for the last two years. The assistant-manager believes that, between the two of you, some new facts or ideas might come out after analysing that data. The assistant-manager's problem is to help the hotel manager--due back in a couple of days--to decide on what to do in order to improve the hotel's performance for next year.

PROBLEM ANALYSIS

Make yourself comfortable, and start analysing the data on the hotel performance, available through the table named **HOTEL** in the computer system. Please remember that the most important thing for **YOU** to do is to figure out what important **RELATIONSHIPS** you can identify from the data in question. No suggestions (solutions) on how to improve the Hotel performance are requested at this point.

NOTATION OF IDEAS

List the **Observations** you make on the **left** side of the worksheets made available to you. **Inferences** (further conclusions based on one or more of the observations you made) can be listed on the **right** side of the worksheets. Avoid passing judgement on the worth of your ideas; assume that **ALL** of your **IDEAS** are **IMPORTANT**, until you're asked to rank them.

Please write as legibly as possible; it will be quite helpful for the subsequent transcription of your notes. Number your observations and inferences or alternatively indicate, in any way you prefer, the separation between any two observations (or inferences). A blank space between items would be fine, for instance.

Be as precise as you can when making comments about the data; always mention the variables you have in mind when writing your observations and inferences. When indicating a time period, specify the month(s) included in that period.

If you have any questions, please ask them now.

---->>> **CONTINUE ON NEXT PAGE**

SESSION ACTIVITIES and VERBAL INSTRUCTIONS 5

[Wait for, and briefly answer, any questions]

Here are the worksheets you will be using, and some pencils. Later on, if you need more sheets, let me know. Please write only on these sheets.

I will be stopping by, every now and then, to draw some lines in your worksheets, to help tie what you are writing with the computer log of your session.

[Type short pw to start; note down TIME in CONTROL sheet]

[Every 10 min, draw a red line after Subject's last written lines]

[Label five red lines as: A X Y Z B; note down TIME for each in CONTROL sheet]

[Type short pw to end; note down TIME in CONTROL sheet]

QUESTIONNAIRE NO. 2

[Give questionnaire; note down TIME in CONTROL sheet]

Please fill out this questionnaire, the last part of this experiment.

[Wait for Subject to fill out questionnaire]

[Type LONG PASSWORD to end; note down TIME in CONTROL sheet]

END SESSION

[Hand out check and receipt]

Please sign and date this receipt.

[Check DATE and other information in Receipt]

Please do not discuss the experiment with other people; one of them might be a potential volunteer, and it is very important that Subjects are not aware of the procedures and content of the experimental session.

Once more, thank you for participating.

[Note down TIME in CONTROL sheet]

SESSION ACTIVITIES and VERBAL INSTRUCTIONS 6

FINAL ACTIVITIES

[Note down session end TIME in CONTROL sheet]

[Check once more RELEASE and RECEIPT Forms]

[Execute 'xxCTCV' in the computer, to see if CONTROL TABLE looks OK]

[Arrange all sheets in order given below]

1	Time Control Sheet
1	Session Notes
1	Receipt Form
1	Questionnaire No. 2
1	Problem Analysis Instructions
n	Problem Analysis Worksheets
1	System Commands handout
1	Fields (Variables) of the "Hotel" Problem (handout)
1	Hands-On Tutorial (first sheet: "Window" through "Fields")
1	Hands-On Tutorial (second sheet: "Sort" through "Drop")
1	Tutorial Instructions
1	Tutorial Worksheet
1	Questionnaire No. 1
1	Release Form
1	Subject Participation and Scheduling handout
1	Subject Application Form

SESSION ACTIVITIES and VERBAL INSTRUCTIONS 7

B.5 - RELEASE AND RECEIPT FORMS

A reduced copy of a sample Subject Release Form is presented in the next page (section B.5.1), followed by a reduced copy of the Receipt Form used (section B.5.2).

B.5.1 - Subject Release Form

RELEASE FORM
Experiment on Quantitative Data Displays

I have read and understood the description of, and conditions for participation in, the experimental research on graphic display of data, undertaken by Fred W. Lacerda Jr. as partial fullfilment of the requirements for his Ph.D. degree in Environmental Design and Planning, College of Architecture and Urban Studies, Virginia Tech.

I agree to participate in that experiment, according to the description of activities presented in the "Subject Participation and Scheduling Handout" (the "handout"), subject to the basic participation payment and bonus stipulated in that document.

I understand that all the data collected relative to my participation in that experiment will remain confidential and, after the conclusion of the research, will not have any identifier linking it to me.

Blacksburg, __________ 1986
date

____
signature

____
LAST name FIRST name middle init.
(CAPITAL LETTERS)

____
Social Security number

B.5.2 - Receipt Form

RECEIPT FORM
Experiment on Quantitative Data Displays

For participation in the experimental research on graphic display of data, undertaken by Fred W. Lacerda Jr. as partial fullfilment of the requirements for his Ph.D. degree in Environmental Design and Planning, College of Architecture and Urban Studies, Virginia Tech, I received a total payment of $ ___.00 (_________ dollars), corresponding to:

$ 7.00 basic participation payment

$ __.00 punctuality

$ __.00 other:__________________

Blacksburg, _________ 1986
date

____
signature

____
LAST name FIRST name middle init.
(CAPITAL LETTERS)

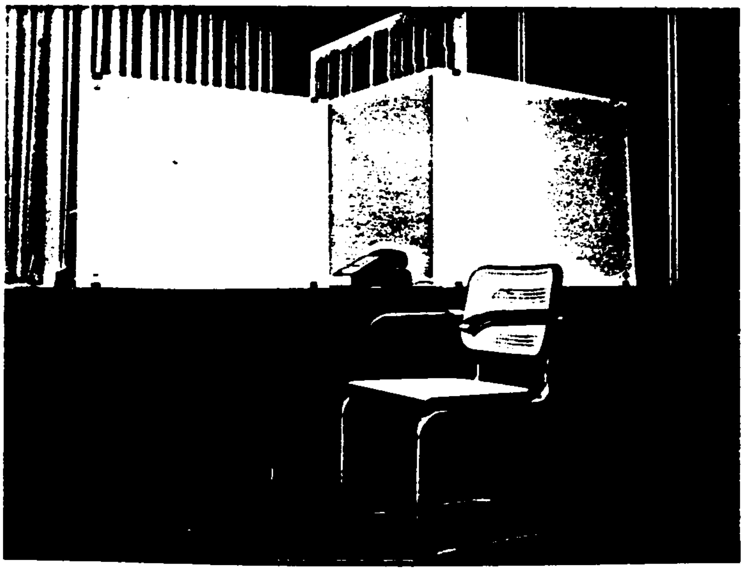
____
Social Security number

B.6 - WORKSTATION AREA FOR EXPERIMENTAL SESSIONS

Photographs of the workstation area are presented in the next two pages (section B.6.1). A floor plan of the office in question (202 Cowgill Hall) is presented in the following page (section B.6.2). The photographs and the floor plan try to convey the relative sense of enclosure of the workstation area used for the experimental sessions, and the relative amount of control available over the visual environmental conditions.

B.5.1 - Workstation Area Photographs

VIEW OF THE OFFICE AREA WITH EXPERIMENTAL WORKSTATION AT THE END OF THE CORRIDOR

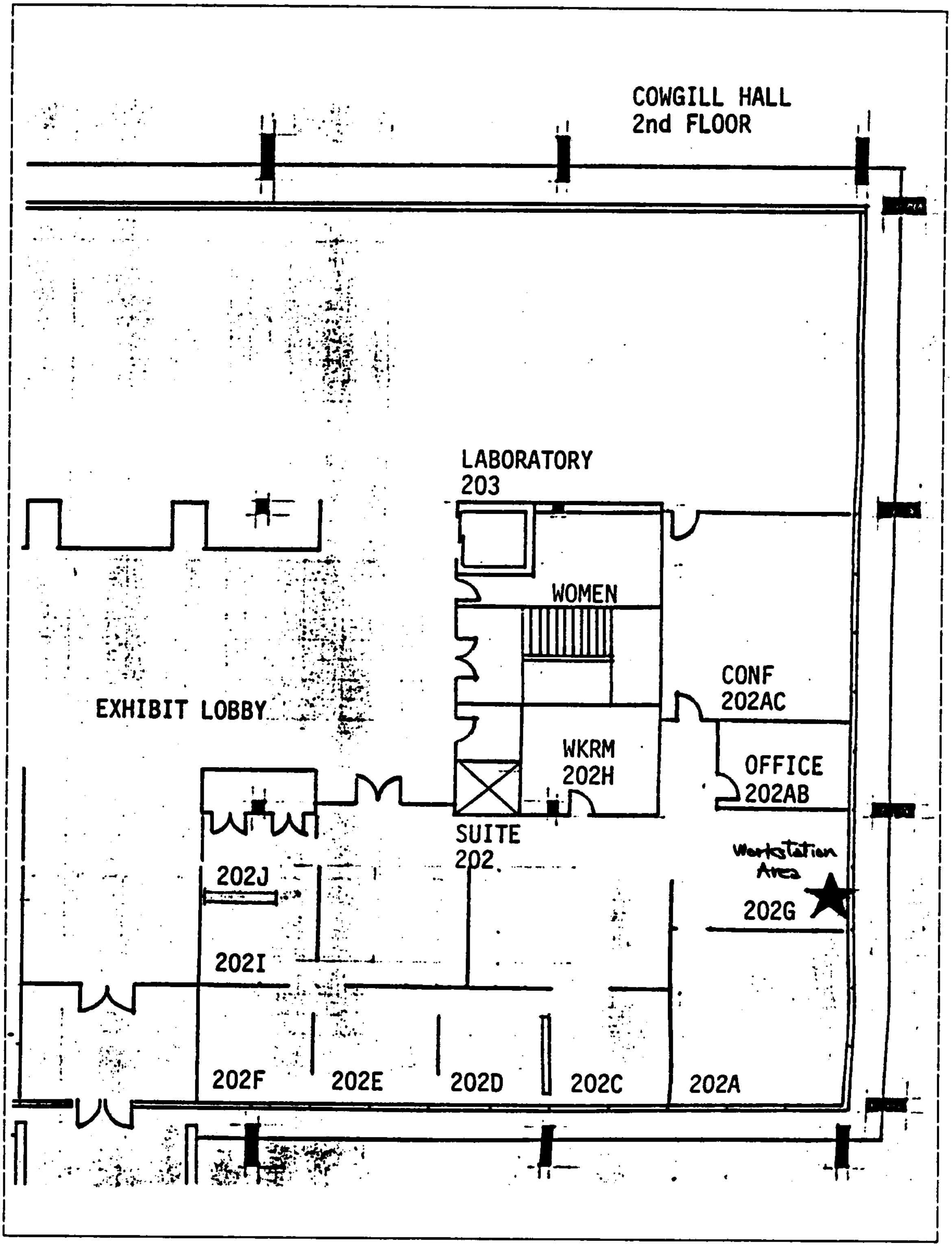


PARTITION FOR VISUAL SCREENING OF WORKSTATION

TEST ADMINISTRATOR READING VERBAL INSTRUCTIONS TO SUBJECT DURING SIMULATED EXPERIMENTAL SESSION

SUBJECT READING INSTRUCTIONS DURING SIMULATED EXPERIMENTAL SESSION

B.5.2 - Workstation Area Floor Plan



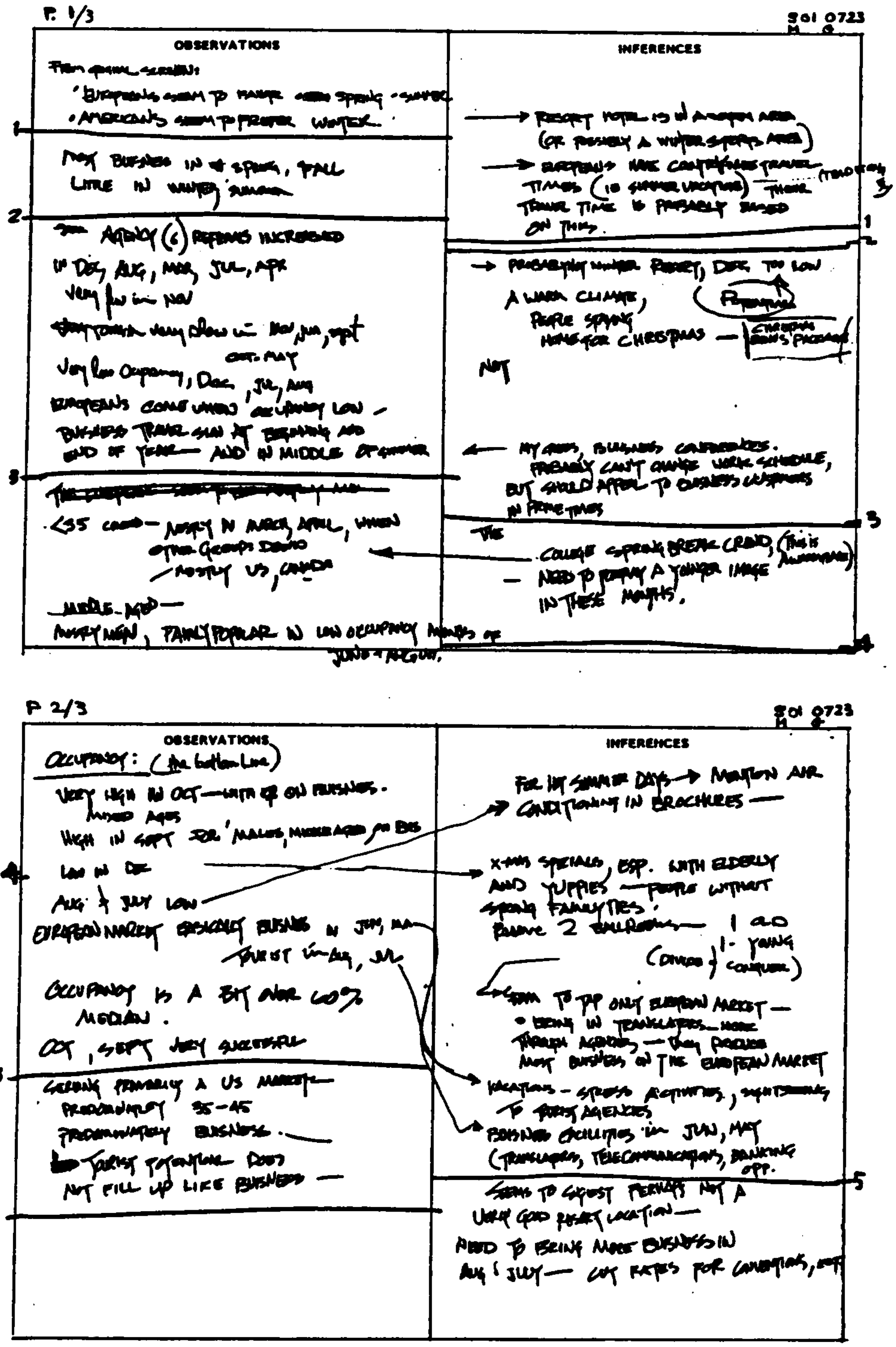
APPENDIX C. SAMPLE TREATMENT OF THE RAW DATA

Contents:

C.1 - SAMPLE VERBATIM TRANSCRIPTION OF "INSIGHTS"

A sample set of worksheets filled with handwritten notes, generated during a pilot experimental session (Subject Code 231), is presented in the next two pages (section C.1.1), followed by the verbatim transcription of those notes (section C.1.2).

C.1.1 - Sample Handwritten Notes (Pilot Session/July 1985)



P. 1/3

S01 0723

OBSERVATIONS	INFERENCES
From [illegible]: · Europeans seem to [illegible] spring - summer · Americans seem to prefer winter	→ Resort hotel is in [illegible] area (or possibly a winter sports area)
Most business in spring, fall Little in winter, summer	→ Europeans have [illegible] travel times (ie summer vacation) — their travel time is probably based on this.
Agency (6) referrals increased in Dec, Aug, Mar, Jul, Apr Very few in Nov [illegible] very slow in Nov, Jun, Sept Very low occupancy, Dec, Oct-May, Jul, Aug Europeans come when occupancy low - business travel [illegible] at beginning and end of year — and in middle of summer	→ Probably not winter resort, Dec too low A warm climate, people staying home for Christmas — [illegible] Not ← My [illegible], business conferences. Probably can't change work schedule, but should appeal to business customers in prime times
<35 [illegible] — mostly in March, April, when other groups down / mostly US, Canada Middle-aged — mostly men, fairly popular in low occupancy months of June + August.	The college spring break crowd, (this is [illegible]) — Need to portray a younger image in these months.

P 2/3

S01 0723

OBSERVATIONS	INFERENCES
Occupancy: (the bottom line) Very high in Oct — with [illegible] on business. Mixed ages High in Sept for males, middle-aged on bus. Low in Dec Aug + July low European market basically business in Jun, Ma[illegible] Tourist in Aug, Jul Occupancy is a bit over 60% median. Oct, Sept very successful	For hot summer days → mention air conditioning in brochures — X-mas specials, esp. with elderly and yuppies — people without strong family ties. Have 2 ballrooms — 1 old, 1 - young (divide + conquer) Seem to tap only European market — · bring in translators — more through agencies — they produce most business on the European market Vacations - stress activities, sightseeing to tourist agencies Business facilities in Jun, May (translators, telecommunications, banking opp.
Serving primarily a US market — predominantly 35-45 predominantly business. Tourist potential does not fill up like business —	Seems to suggest perhaps not a very good resort location — Need to bring more business in Aug + July — cut rates for conventions, [illegible]

Sample Handwritten Notes (cont.)
Pilot Session/July 1985

P 3/3

S 010723
M 6

OBSERVATIONS	INFERENCES
	WHY SUCH A DIVISION FROM MEN TO WOMEN IN OCT/SEPT.? FOR BUSINESS. BUSINESS AMENITIES: · GOLF · SAUNA · GYM · BARS

6

C.1.2 - Sample Verbatim Transcription of "Insights" (Pilot Session/July 1985)

PROBLEM DATA

S010723

OBSERVATIONS

from original screen:

- europeans seem to favor spring - summer
- americans seem to prefer winter.

TIME: 10 min.

- most buisness [sic] in spring, fall
  little in winter, summer

TIME: 20 min.

- agency (6) referals [sic] increased
  in dec, aug, mar, jul, apr
  very few in nov
- tourism very slow in nov, jun, sept
  oct, may
- very low occupancy, dec., jul, aug
- europeans come when occupancy low -
- buisness [sic] travel slow at beginning and
  end of year -- and in middle of summer.

TIME: 30 min.

- < 35 crowd -- mostly in march, april, when
  other groups down
  - mostly US, canada
- middle aged -
  mostly men, fairly popular in low occupancy months of
  june & august
- occupancy: (the bottom line)
  very high in oct -- with females on buisness [sic]
  mixed ages
  high in sept for males, middle aged, on buis [business]

TIME: 40 min.

low in dec.
aug & july low

- european market basically buisness [sic] in jun, ma
  tourist [market] in aug, jul
- occupancy is a bit over 60%
  median
  oct, sept very successful

TIME: 50 min.

- serving primarily a US market -
  predominantly 35 - 45
  predominantly buisness [sic]
- tourist potential does
  not fill up like buisness [sic] -

TIME: 60 min.

INFERENCES

- -->> resort hotel is in a warm area
  (or possibly a winter sports are)
- -->> europeans have countrywide travel
  times (ie summer vacations). their
  travel time is probably based
  on this

TIME: 10 min.

TIME: 20 min.

- -->> probably winter resort, dec. too low.
  a warm climate, c
  people staying potential--!
  home for christmas -- christmas bonus packages.
- <<-- my guess, buisness [sic] conferences.
  probably can't change work schedule,
  but should appeal to buisness [sic] customers
  in prime times

TIME: 30 min.

- <<-- the college springbreak crowd,
  (this is a warm place)
  - need to portray a younger image
  in these months

TIME: 40 min.

- -->> for hot summer days --> mention air
  conditioning in brochures -
- -->> x-mas specials, esp. with elderly
  and yuppies -- people without
  strong family ties.
  reserve 2 ballrooms -- 1 old 1 young
  (divide and conquer)
- -->> seem to tap only european market -
  o bring in translators - work
  through agencies -- they produce
  most buisness [sic] on the european market
- -->> vacations -- stress activities, sightseeing
  to tourist agencies
- -->> buisness [sic] facilities in jun, may
  (translators, telecommunications, banking opp.

TIME: 50 min.

- -->> seems to sugest [sic] perhaps not a
  very good resort location -
- Need to bring more buisness [sic] in
  aug & july -- cut rates for conventions, etc.
- why such a division from men to women
  in oct/sept.? for buisness [sic]
- buisness [sic] amenities:
  . golf
  . sauna
  . gym
  . bars

TIME: 60 min.

C.2 - SAMPLE PARSING AND COMPRESSION OF "INSIGHTS"

The verbatim transcription and the respective parsing and compression of each "insight", generated during a pilot experimental session (Subject Code 231), are presented in the next two pages (section C.2.1).

The compressed form of each "insight" presents the following structure:

Subject - Description of State - Specifier

The "core" information (the sentence subject, i.e., field name or names) is presented capitalized with bold letters. The "state specifier" is capitalized, with abbreviations for the most common time periods (presented in section C.2.2). Marginally important information is placed between parenthesis. A list of common expressions (employed by subjects of the pilot experimental sessions) and their "translation" is presented in section C.2.3.

At this stage of raw data manipulation, a first block of identifying labels is attached to each sentence. The labels are composed of one of a maximum of five parts, from left to right, in the order given below:

[01-99] Provenance (number of original verbatim sentence).

[S/F/G] Type of "insight" ("S" - observation; "F" - inference; "G" - suggestion).

[1-6] Time period during which a "insight" will have been written.

1. Between 0 - 10 min
2. Between 11 - 20 min
3. Between 21 - 30 min
4. Between 31 - 40 min
5. Between 41 - 50 min
6. Between 51 - 60 min

[a-z] Order of decomposition of sentences, from an originally complex sentence or from a block of simple sentences originally lumped together by Subject.

["/F" or "/S"] An optional set of symbols, to be included only when the original placement of sentence in the worksheet (either under the "Observations" or the "Inferences" headings) is found to be incorrect; this new symbol indicates the correct "insight" type, to be included in the next block of identifying symbols.

C.2.1 - Verbatim Transcription and Sample Parsing and Compression of "Insights" (Pilot Session/July 1985)

PROBLEM DATA	D.E.S. PILOT-TEST DATA - JUL 85 SP 231 - OBSERVATIONS (2nd Revision)
S010723	
OBSERVATIONS	
from original screen:	
•. europeans seem to favor spring - summer	01S1a EUROPE/ASIA Gs high SPRING-SUMMER.
• americans seem to prefer winter.	02S1a USA/CANADA Gs high WINTER.
TIME: 10 min.	TIME 10 min.
• most buisness [sic] in spring, fall little in winter, summer	03S2a BUSINESS high SPRING FALL low WINTER SUMMER.
TIME: 20 min.	TIME 20 min.
• agency (6) referels [sic] increased in dec, aug, mar, jul, apr very few in nov	04S3a AG. RESERV. high DEC AUG MAR JUL APR very few NOV.
• tourism very slow in nov, jun, sept oct, may	05S3a TOURIST Gs very slow NOV JUN SEP OCT MAY.
• very low occupancy, dec., jul, aug	06S3a OCCUPANCY very low DEC JUL AUG.
• europeans come when occupancy low -	07S3a EUROPE/ASIA Gs inv. rel. OCCUPANCY.
• buisness [sic] travel slow at beginning and end of year -- and in middle of summer.	08S3a BUSINESS Gs low BEG/YR END/YR MIDSUMMER.
TIME: 30 min.	TIME 30 min.
• < 35 crowd -- mostly in march, april, when other groups down - mostly US, canada	09S4a < 35 YRS Gs highest MAR APR when USA/CANADA lowest.
• middle aged - mostly men, fairly popular in low occupancy months of june & august	10S4a > 55 YRS Gs mostly MALE Gs highest JUN AUG when OCCUPANCY low.
• occupancy: (the bottom line) very high in oct -- with females on buisness [sic] mixed ages	11S4a OCCUPANCY highest OCT w/ FEMALE Gs mixed AGES on BUSINESS.
high in sept for males, middle aged, on buis [business]	11S4b OCCUPANCY highest SEP w/ MALE 35-55 YRS Gs on BUSINESS.
TIME: 40 min.	TIME 40 min.
low in dec. aug & july low	11S5c OCCUPANCY low DEC AUG JUL.
• european market basically buisness [sic] in jun, ma	12S5a EUROPE/ASIA Gs mainly BUSINESS Gs in JUN MAR.
tourist [market] in aug, jul	12S5b EUROPE/ASIA Gs mainly TOURIST Gs in AUG JUL.
• occupancy is a bit over 60% median	13S5a OCCUPANCY median 60% +.
oct, sept very successful	13S5b OCCUPANCY high OCT SEP.
TIME: 50 min.	TIME 50 min.
• serving primarily a US market - predominantly 35 - 45 predominantly buisness [sic]	14S6a/F HOTEL main market USA/CANADA, 35 - 55 YRS, BUSINESS Gs.
• tourist potential does not fill up like buisness [sic] -	15S6a/F TOURISM secondary to BUSINESS.
TIME: 60 min.	TIME 60 min.

Sample Parsing and Compression (cont.)
Pilot Session/July 1985

INFERENCES

- -->> resort hotel is in a warm area (or possibly a winter sports are)
- -->> europeans have countrywide travel times (ie summer vacations). their travel time is probably based on this

TIME: 10 min.

TIME: 20 min.

- -->> probably winter resort, dec. too low.
  a warm climate, C
  people staying potential--I
  home for christmas -- christmas bonus packages.
- <<-- my guess, buisness [sic] conferences. probably can't change work schedule, but should appeal to buisness [sic] customers in prime times

TIME: 30 min.

- <<-- the college springbreak crowd, (this is a warm place)
  - need to portray a younger image in these months

TIME: 40 min.

- -->> for hot summer days --> mention air conditioning in brochures -
- -->> x-mas specials, esp. with elderly and yuppies -- people without strong family ties.
  reserve 2 ballrooms -- 1 old 1 young (divide and conquer)
- -->> seem to tap only european market -
  o bring in translators - work through agencies -- they produce most buisness [sic] on the european market
- -->> vacations -- stress activities, sightseeing to tourist agencies
- -->> buisness [sic] facilities in jun, may (translators, telecommunications, banking opp.

TIME: 50 min.

- -->> seems to sugest [sic] perhaps not a very good resort location -
- Need to bring more buisness [sic] in aug & july -- cut rates for conventions, etc.
- why such a division from men to women in oct/sept.? for buisness [sic]
- buisness [sic] amenities:
  . golf
  . sauna
  . gym
  . bars

TIME: 60 min.

D.E.S. PILOT-TEST DATA - JUL 85

SP 231 - INFERENCES (2nd Revision)

01F1a HOTEL warm-area or winter-sports resort.

02F1a EUROPEANS have countrywide travel times (summer vacations).

TIME 10 min.

TIME 20 min.

03F3a HOTEL probably winter resort since OCCUPANCY too low DEC.

03F3b HOTEL also probably warm climate since people home DEC (XMAS).

03F3c/G Potential XMAS bonus packages for low OCCUPANCY in DEC.

04F3a/S BUSINESS Gs overall majority.

04F3b BUSINESS Gs overall majority due probably to conferences.

04F3b/G No changes possible in work schedule but encourage BUSINESS Gs in prime times.

TIME 30 min.

05F4a < 35 YRS Gs probably college springbreak crowd (warm climate).

05F4b/G Need to portray a younger image for college springbreak crowd.

TIME 40 min.

06F5a/G Encourage OCCUPANCY in hot SUMMER; mention air conditioning in brochures.

07F5a/G Encourage XMAS OCCUPANCY w/ specials for elderly and yuppies--people without strong family ties. Reserve 2 ballrooms -- 1 old 1 young (divide and conquer).

08F5a HOTEL tapping only EUROPE/ASIA market.

08F5b/G Since HOTEL tapping only EUROPE/ASIA market, bring in translators, work through agencies--they produce BUSINESS in EUROPE.

09F5a/G SUMMER vacations; stress activities and sightseeing to encourage AG. RESERV. Gs.

10F5a/G Provide BUSINESS facilities JUN MAY (translators, telecommunications, banking opp.).

TIME 50 min.

11F6a BUSINESS Gs majority suggest poor resort location.

12F6a/G Encourage BUSINESS Gs in AUG JUL (cut rates for conventions, etc.).

14F6a/G Provide BUSINESS Gs w/ amenities (golf, sauna, gym, bars).

TIME 60 min.

C.2.2 - List of Abbreviations for Parsing and Compression of "Insights".

- **Time periods:**

Months	JA FE MA AP MY JN JL AU SE OC NO DE	
	Sequence (e.g.):	AU-NO, JA-AP
Others	SUMMER WINTER SPRING FALL 1ST SEM 2D SEM 2 YRS BEG/YR END/YR ALL YR REST/YR XMAS HOLIDAYS N., S. HEMISPHERE	

- **Miscellaneous abbreviations:**

w/	with
dir rel	direct relationship
inv rel	inverse relationship
no rel	no relationship
corr	correlation with
max	maximum
min	minimum
med	medium
%(s)	percentage(s)
b/c	because
b/w	between

- **Field Names**

FEMALE Gs	Female guests.
MALE Gs	Male guests.
LOCAL Gs	Guests from the local state
STATE Gs	Guests from the local state
USA Gs	Guests from the rest of the U.S.S.
OUT/STATE Gs	Guests from the rest of the U.S.S.
EUROPEAN Gs	Guests from Europe.
S. AMER. Gs	Guests from South America.
AFR/M.EAST Gs	Guests from Africa or the Middle-East.
ASIAN Gs	Guests from Asia.
BUSINESS Gs	Guests on business.
TOURIST Gs	Guests on tourism.
DIR. RESERV. Gs	Guests making direct hotel reservations.
AG. RESERV. Gs	Guests making hotel reservations through travel agencies.
AIR CREW Gs	Air Crew guests with continuous reservations.
<20 YRS Gs	Guests less than 30 years of age.
20-34 YRS Gs	Guests between 20 and 34 years of age.
35-54 YRS Gs	Guests between 35 and 54 years of age.
>55 YRS Gs	Guests with more than 55 years of age.
Gs	Guests

C.2.3 - Sample List of Expressions and Their "Translations" (Pilot Sessions/July 1985)

... Guests	... Gs ...
... [guests] travel in the ...	... Gs highest ...
... drops off considerably ...	... sharp drop ...
... most active months ...	... highest ...
... least active month ...	... lowest ...
... particularly ...	... mainly ...
... did not fluctuate ...	... stable ...
... switched times of the year w/ ...	... inv rel ...
Most of business is ...	... overall majority.
Most people (throughout year) ...	... majority.
... usually means ...	... all parts/year.
... implies ...	... ALL YR.
Slump in ... during ...	... lowest ...
... seem to favor ...	... high.
Most ... in ...	... high ...
... increased in ...	... highest ...
... mostly in ...	
... basically ...	... mainly
... primarily ...	... mainly ...
... potential does not fill up like ...	... secondary to ...
... overwhelming %.	... overall majority.
... large %.	... great majority.
... 1/4 of guests ...	... 1/4 overall.
... fluctuate widely over course of a year.	... wide variation ALL YR.
... is particularly active in ...	... high ...
... % is stronger in ...	... high ...
... appear in ...	... high ...
... the rest of the year ...	... REST/YR.
... consistently over 60% ...	... overall 60%. +.
... popular months ...	... high ...
... between ...	... bet ...
Most of ... is from ...	... mainly ...

Expressions and Their Translations (cont.)
Pilot Session/July 1985

... peak ...	... highest ...
... high stable ...	... high.
... mostly conducted on months ...	... highest ...
... first months of the year.	... 1ST MOS. YR.
... basically not of a nature ...	... imply [negation] ...
... heaviest mths. ...	... highest ...
... lighest mths. ...	... lowest ...
... 75-80% ...	... 75/80% [implied proportion].
All people are coming from ... in ...	... only ones ...
No. of people from ... are minimum.	... minority.
No. of people from ... are maximum.	... majority.
All ... are ... in ...	... only ones ...
No ... in ...	... only ones ...
All have come for ... in ...	... only ones ...
Not many people are ...	... minority.

C.3 - SAMPLE REORDERING OF "INSIGHTS"

An example of reordered, parsed "insights", generated during a pilot experimental session (Subject Code 231), is presented in the next page.

A new block of symbols is attached, at this stage, to the left of the sentence identification label for each "insight". The new symbols, separated from the first block by a dash and placed to its left, are divided in three parts, arranged from left to right in the sequence given below:

[S/F/G]	Final "insight" type determined for the sentence (S - observations; F - inferences; G - suggestions).
[A-Z]	One-character code for identification of leading field name (the first one mentioned) in sentence.
[01-99]	Number indicating order of appearance of sentence within the "insight" type and leading field-name cluster.

The alphanumeric labels used for identification of "insights" fulfill essentially a procedural purpose, namely, the possibility of quick retrieval of original information location.

Sample Reordering of "Insights" (cont.)
Pilot Session/July 1985

D.E.S. PILOT-TEST DATA - JUL 85

SP 231 - OBSERVATIONS (3d Revision)

U01-02Sa USA/CANADA Gs high WINTER.

SE01-01Sa EUROPE/ASIA Gs high SPRING-SUMMER.

SE02-07Sa EUROPE/ASIA Gs inv. rel. OCCUPANCY.

SE03-12Sa EUROPE/ASIA Gs majority when BUSINESS Gs high JUN MAR.

SE04-12Sb EUROPE/ASIA Gs majority when TOURIST Gs high AUG JUL.

SB01-03Sa BUSINESS Gs high SPRING FALL low WINTER SUMMER.

SB02-08Sa BUSINESS Gs low BEG/YR END/YR MIDSUMMER.

SB03-04Fa BUSINESS Gs overall majority.

ST01-05Sa TOURIST Gs lowest NOV JUN SEP OCT MAY.

SA01-04Sa AG. RESERV. Gs high DEC AUG MAR JUL APR very few NOV.

SL01-09Sa < 35 YRS Gs highest MAR APR when USA/CANADA Gs lowest.

SG01-10Sa > 55 YRS Gs mostly MALE Gs highest JUN AUG when OCCUPANCY low.

SO01-06Sa OCCUPANCY very low DEC JUL AUG.

SO02-11Sa OCCUPANCY highest OCT when FEMALE Gs mixed AGES on BUSINESS.

SO03-11Sb OCCUPANCY highest SEP when MALE 35-55 YRS Gs on BUSINESS.

SO04-11Sc OCCUPANCY low DEC AUG JUL.

SO05-13Sa OCCUPANCY median 60% +.

SO06-13Sb OCCUPANCY high OCT SEP.

SP 231 - INFERENCES (3d Revision)

FU01-14Sa USA/CANADA, 35-55 YRS, BUSINESS Gs main market.

FE01-02Fa EUROPEANS countrywide travel times (summer vacations).

FE02-08Fa EUROPE/ASIA market only one tapped.

FB01-04Fb BUSINESS Gs overall majority due probably to conferences.

FB02-11Fa BUSINESS Gs majority suggest poor resort location.

FT01-15Sa TOURISM secondary to BUSINESS.

FL01-05Fa < 35 YRS Gs probably college springbreak crowd (warm climate).

FC01-01Fa HOTEL warm-area or winter-sports resort.

FC02-03Fa HOTEL probably winter resort since OCCUPANCY too low DEC.

FC03-03Fb HOTEL also probably warm climate since people home DEC (XMAS).

D.E.S. PILOT-TEST DATA - JUL 85

SP 231 - SUGGESTIONS (3d Revision)

GZ01-03Fa Potential XMAS bonus packages for low OCCUPANCY in DEC.

GZ02-04Fb No changes possible in work schedule but encourage BUSINESS Gs in prime times.

GZ03-05Fb Need to portray a younger image for college springbreak crowd.

GZ04-06Fa Encourage OCCUPANCY in hot SUMMER; mention air conditioning in brochures.

GZ05-07Fa Encourage XMAS OCCUPANCY w/ specials for elderly and yuppies--people without strong family ties. Reserve 2 ballrooms -- 1 old 1 young (divide and conquer).

GZ06-08Fb Since HOTEL tapping only EUROPE/ASIA market, bring in translators, work through agencies--they produce BUSINESS in EUROPE.

GZ07-09Fa SUMMER vacations; stress activities and sightseeing to encourage AG. RESERV. Gs.

GZ08-10Fa Provide BUSINESS facilities JUN MAY (translators, telecommunications, banking opp.).

GZ09-12Fa Encourage BUSINESS Gs in AUG JUL (cut rates for conventions, etc.).

GZ10-14Fa Provide BUSINESS Gs w/ amenities (golf, sauna, gym, bars).

C.4 - SAMPLE SCORING OF "INSIGHTS"

A sample of a scoring sheet for "insights", generated during a pilot experimental session (subject J231), is presented in the next page.

The "insights" to be removed are identified, in the appropriate form, by the the first letter of the label that identifies each rule:

P	Partial "insights".
R	Repeated "insights".
I	Incomplete "insights"
S	"Spot", or data-element specific, "insights".
W	Wrong "insights".
T	Tautological "insights".

In the example on the next page, only one "insight" was rejected (a "wrong" observation, unsupported by the data).

Another symbol is attached, at this stage, to the left of the sentence identification label for each "insight". The new symbol corresponds to the Subject identification and is separated from the original block by a dash and placed to its left.

[SN]	Subject number code.
[01-80]	Two-digit unique identification code.

Sample Scoring of "Insights" (cont.)
Pilot Session/July 1985

D.E.S. PRE-TEST DATA - JUL 85

J231-SU01-02Sa	●	USA/CANADA Gs high WINTER.
J231-SE01-01Sa	●	EUROPE/ASIA Gs high SPRING-SUMMER.
J231-SE02-07Sa	●	EUROPE/ASIA Gs inv. rel. OCCUPANCY.
J231-SE03-12Sa	●	EUROPE/ASIA Gs majority when BUSINESS Gs high JUN MAR [MAY]
J231-SE04-12Sb	●	EUROPE/ASIA Gs majority when TOURIST Gs high AUG JUL.
J231-SB01-03Sa	●	BUSINESS Gs high SPRING FALL low WINTER SUMMER.
J231-SB03-04Fa	●	BUSINESS Gs overall majority.
J231-ST01-05Sa	●	TOURIST Gs lowest NOV JUN SEP OCT MAY.
J231-SA01-04Sa	●	SP. [DIR.] RESERV. Gs high DEC AUG MAR JUL APR very few NOV.
J231-SL01-09Sa	●	< 35 YRS Gs highest MAR APR when USA/CANADA [EUROPE/ASIA] Gs lowest.
J231-S?01-10Sa	●	[35-55] YRS Gs mostly MALE Gs highest JUN AUG when OCCUPANCY low.
J231-S001-06Sa	●	OCCUPANCY very low DEC JUL AUG.
J231-S002-11Sa	●	OCCUPANCY highest OCT when FEMALE Gs mixed AGES on BUSINESS.
J231-S003-11Sb	●	OCCUPANCY highest SEP when MALE 35-55 YRS Gs on BUSINESS.
J231-S005-13Sa	●	OCCUPANCY median 60% +.
J231-FU01-14Sa	●	USA/CANADA, 35-55 YRS, BUSINESS Gs main market.
J231-FE01-02Fa	●	EUROPEANS countrywide travel times (summer vacations).
J231-FE02-08Fa	W	EUROPE/ASIA market only one tapped. ???
J231-FB01-04Fb	●	BUSINESS Gs overall majority due probably to conferences.
J231-FB02-11Fa	●	BUSINESS Gs majority suggest poor resort location.
J231-FL01-05Fa	●	< 35 YRS Gs probably college springbreak crowd (warm climate).
J231-FC01-01Fa	●	HOTEL warm-area or winter-sports resort.

check original ✓
check original ✓
check original ✓

1 (W)

C.5 - SAMPLE CATEGORIZATION OF "INSIGHTS"

A sample of a categorization sheet for "insights", generated during a final experimental session in April 1986 (Subject Code 14), is presented in the next page.

Sample Categorization of "Insights" (cont.)
April 1986

FWL/DES - Experimental Data - MARCH/APRIL 1986

			A (1)	B (2)	C (3)	D (4)	E (5)	F (6)	Summary	G (2)	H (3)	I (4)	J (5)	K (6)	L (7)	M (8)	N (9)	O (10)	#	Table Summary
	(7) SC14 - OBSERVATIONS (Rewritten: C - 2nd Revision)																			
SN14-SF01-02S2	FEMALE Gs highest WINTER	S 2 A	●																	
SN14-SU01-08S6	USA Gs highest SUMMER	S 6 A	●																	
SN14-SB01-03S3	BUSINESS Gs inv rel TOURIST Gs in SUMMER	S 3 B		●																
SN14-SD01-04S4	DIR. RESERV. Gs majority	S 4 CS			●				●											
SN14-SZ01-07S5	35-54 YRS Gs majority	S 5 DS				●			●											
SN14-S001-01S2/C	OCCUPANCY high ALL YR; highest AP SE-OC/84/85	S 2 A	●																	
SN14-S003-06S5	OCCUPANCY dir rel CONVENTIONS	S 5 G								●										
			3	1	1	1			(2)	1										
	(1) SC14 - INFERENCES (Rewritten: C - 2nd Revision)																			
SN14-FT01-02F6	TOURIST Gs and USA Gs highest SUMMER: possibly TOURIST Gs from OUT OF STATE	F 6 G								●										
										1										

C.6 - SUMMARY OF PILOT-STUDY RESULTS

The next two pages (section C.6.1) present a complete list of all "multiple-field" "insights" generated by Subjects of the pilot-study. The list of "insights" is divided according to Subject participation in either the treatment (graphic mode of data representation) or control group (numeric mode). Section C.6.2 presents a manuscript/graphic summary of the results of the pilot study.

C.6.1 - Complex "Insights" Generated by Pilot-Study Subjects

GRAPHIC MODE

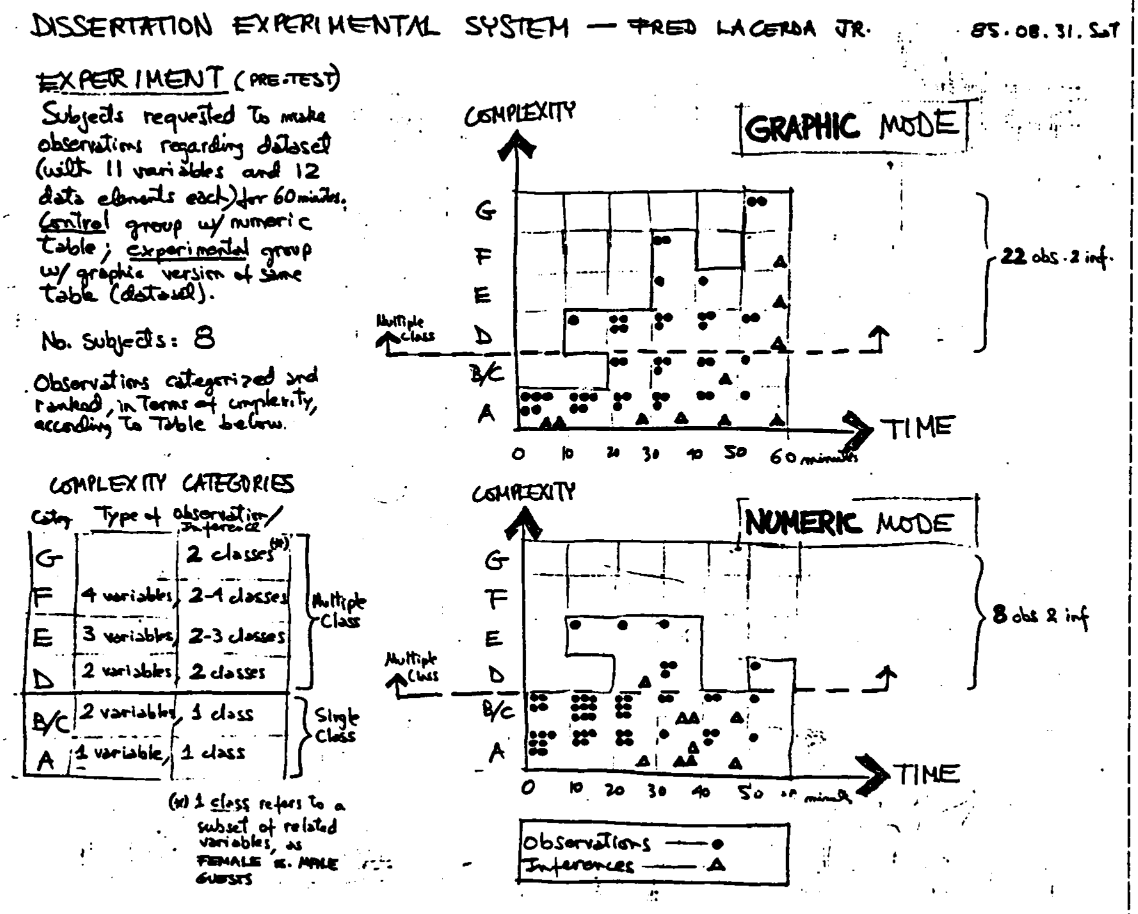
Group	Code	Insight
	J091-CFB01-06S9f	Attracting BUSINESS in SUMMER (when TOURISM high) unlikely.
C	J152-CSF02-06F6a	FEMALE and MALE Gs balanced 1ST SEM OCT mostly MALE Gs 2ND SEM.
C	J152-CSU02-03F4a	USA/CANADA Gs majority 1ST SEM mixed with EUROPE/ASIA Gs 2ND SEM.
C	J152-CSB01-04S3a	BUSINESS Gs overall majority except SUMMER when TOURIST Gs high.
C	J152-CSM02-05F5a	[35-55 YRS Gs] nul MAR when < 35 YRS Gs highest.
C	J152-CFZ03-05F5b	If > 55 YRS Gs nul MAR when < 35 YRS Gs highest, possible reason young business training.
D	J091-DSE03-06S9a	EUROPE/ASIA Gs, TOURIST Gs majority SUMMER (JUL-AUG).
D	J091-DS002-02S9a	OCCUPANCY inv. rel. < 35 YRS Gs. JAN JUL AUG DEC.
D	J091-DS004-07S9a	OCCUPANCY GE 4 (FEB APR SEP-NOV) BUSINESS Gs. majority.
D	J091-DS005-07S9b	OCCUPANCY GE 4 (FEB APR SEP-NOV) USA/CANADA Gs. majority.
D	J091-DS006-07S9c	OCCUPANCY GE 4 (FEB APR SEP-NOV) EUROPE/ASIA Gs. 20-60%.
D	J091-DS007-07S9d	OCCUPANCY GE 4 (APR SEP-NOV) TOURIST Gs. 20% -.
D	J091-DS008-07S9f	OCCUPANCY GE 4 (APR) DIR. RESERV. 100%.
D	J152-DST01-03F5b	TOURIST Gs highest SUMMER when OCCUPANCY low.
D	J152-DSD01-02S3a	DIR. RESERV. Gs inv. rel. OCCUPANCY.
D	J152-DS002-04S3b	OCCUPANCY lowest JUL mostly TOURIST Gs.
D	J152-DFZ01-02F3b	OCCUPANCY nul DEC implies BUSINESS HOTEL.
D	J172-DSB02-03F6b	BUSINESS Gs dir. rel. AG. RESERV. Gs.
D	J172-DST02-08S6a	TOURIST Gs dir. rel. DIR. RESERV. Gs.
D	J172-DFT01-02F6a	TOURIST Gs high SUMMER from EUROPE/ASIA.
D	J231-DSE02-07S3a	EUROPE/ASIA Gs inv. rel. OCCUPANCY.
D	J231-DSE03-12S5a	EUROPE/ASIA Gs majority when BUSINESS Gs high JUN [MAY].
D	J231-DSE04-12S5b	EUROPE/ASIA Gs majority when TOURIST Gs high AUG JUL.
D	J231-DSL01-09S4e	< 35 YRS Gs highest MAR APR when EUROPE/ASIA Gs lowest.
D	J261-DSB02-07S4a	BUSINESS Gs highest SPRING mostly 35-55 YRS.
D	J261-DST02-01F2a	~~[illegible]~~ TOURIST Gs increasing dir. rel. DIR. RESERV. Gs.
D	J261-DS003-02F4e	Highest OCCUPANCY some corr. BUSINESS Gs high SPRING (MAR-JUN) FALL (early, SEP-NOV).
E	J152-EFZ02-04F5a	AGE of Gs imply BUSINESS rather than FAMILY HOTEL.
E	J231-ESM01-10S4a	35-55 YRS Gs mostly MALE Gs highest JUN AUG when OCCUPANCY low.
E	J231-EFU01-14S6a	USA/CANADA, 35-55 YRS, BUSINESS Gs main market.
F	J091-FF001-07S9e	OCCUPANCY GE 4 (APR SEP-NOV) TOURIST Gs 20% -; EUROPE/ASIA Gs possibly also BUSINESS Gs..
F	J172-FF001-01F6a	OCCUPANCY fluctuations due to demands by different AGE, TOURISM, and BUSINESS groups at different YR periods.
F	J231-FS002-11S4a	OCCUPANCY highest OCT when FEMALE Gs mixed AGES on BUSINESS.
F	J231-FS003-11S4b	OCCUPANCY highest SEP when MALE 35-55 YRS Gs on BUSINESS.
G	J152-GSZ01-09S6a	Rel. found: AGE vs. BUSINESS.
G	J152-GSZ05-09S6e	Rel. found: SEX vs. ORIGIN.

Complex "Insights" (cont.)
Pilot Study/July 1985

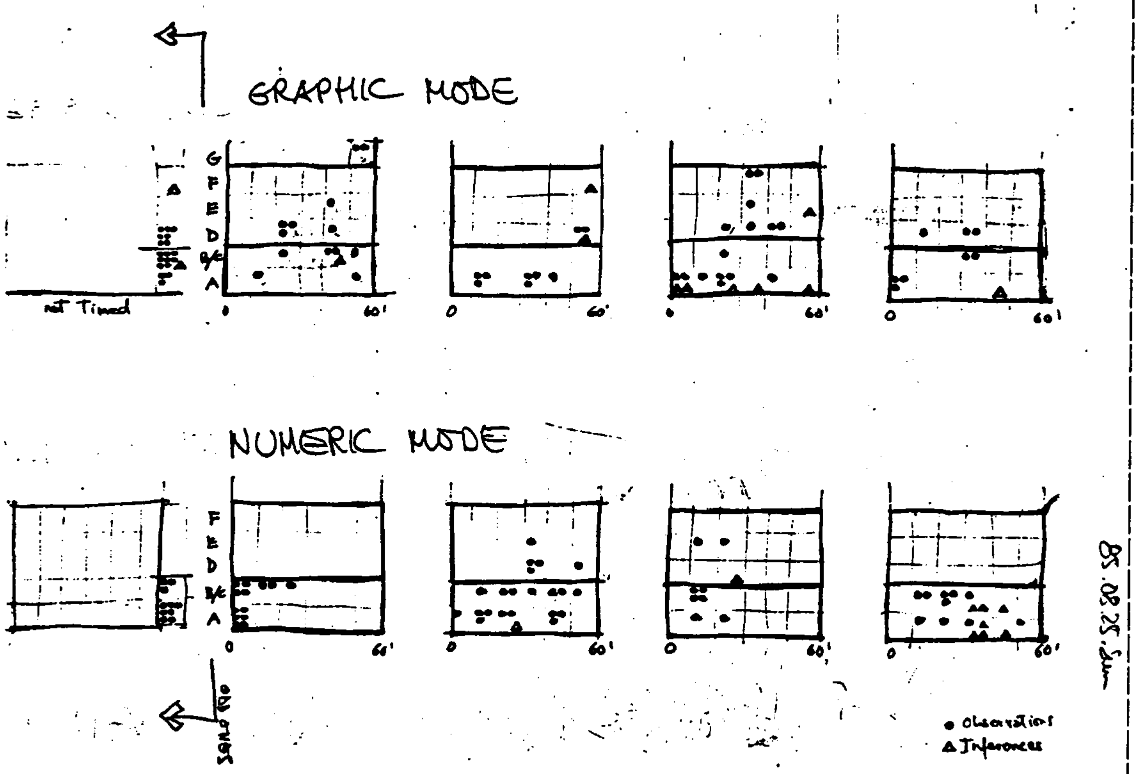
NUMERIC MODE

	J151-CSU03-11S2a	USA/CANADA and EUROPE/ASIA Gs balanced (50/50%) SEP OCT.
C	J171-CSU02-02F2a	USA/CANADA Gs inv. rel. EUROPE/ASIA Gs.
	J171-CSB01-05S3a	BUSINESS Gs more then TOURIST Gs.
	J171-CST02-04F3a	TOURIST Gs inv. rel. BUSINESS Gs.
	J171-CSL01-11S5a	< 35 YRS Gs more than > 55 YRS.
	J251-CSU01-02S2a	USA/CANADA Gs (JAN-APR SEP-DEC) inv. rel. EUROPE/ASIA Gs (MAY-AUG).
D	J171-DSB03-05F4a	BUSINESS Gs dir. rel. USA/CANADA Gs.
	J171-DSB04-08F6a	BUSINESS MALE Gs majority.
	J171-DST03-06F4a	TOURIST Gs dir. rel. EUROPE/ASIA Gs.
	J171-DSD02-09S4a	DIR. RESERV. Gs high DEC MAR AUG (low OCCUPANCY months) low FALL SUMMER (early).
	J251-DSB02-06S2a	BUSINESS Gs dir. rel. DIR. RESERV. Gs.
	J251-DSB03-07S3a	BUSINESS MALE Gs majority ALL YR.
	J251-DFS01-07S3b	BUSINESS MALE Gs ALL YR probable reason international CONVENTIONS.
E	J171-ESD03-07F4a	DIR. RESERV. more by TOURIST Gs than by BUSINESS Gs JUL AUG.

C.6.2 - Graphic Summary of Pilot-Study Results



Graphic Summary (cont.)
Pilot Study/July 1985



C.7 - STUDY RESULTS: CATEGORY LEVEL "N" "INSIGHTS"

The next two pages present a list of the most complex type of "insights" (category level "N") produced by Subjects in the treatment ("graphic") group and the control ("numeric") group (sections C.7.1 and C.7.2, respectively).

C.7.1 - List of Level "N" "Insights" for "Graphic" Group

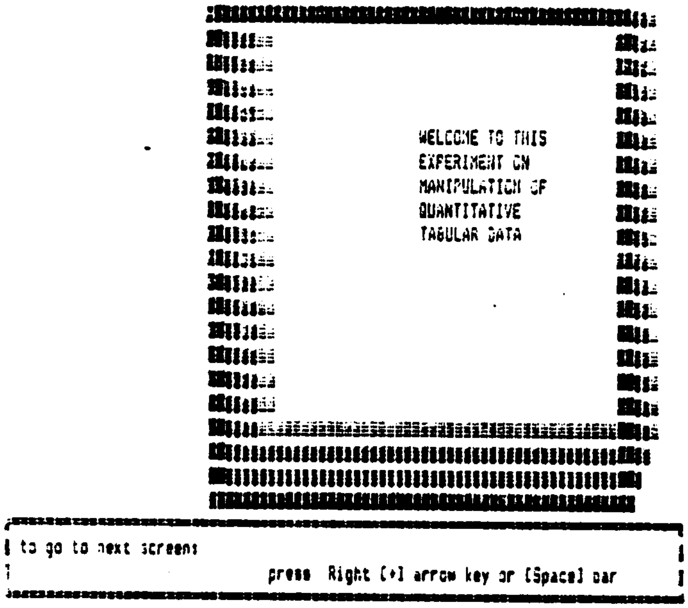
SN02-SQ01-05F4	Gs: majority BUSINESS, 20-34 YRS Gs
SN05-SQ01-01S1	Gs: majority STATE, BUSINESS, DIR. RESERV. Gs
SN05-SQ02-02S1	Gs: minority ASIA, < 20 YRS, AIR CREW RESERV., AFR/M.EAST, S. AMER., EUROPEAN Gs
SN07-SB01-08S1	BUSINESS Gs overall majority; lowest SUMMER
SN07-SB03-07F4	BUSINESS Gs and CONVENTIONS mainstay HOTEL business
SN09-SQ01-02S3	Gs: majority STATE, BUSINESS, DIR. RESERV., 35-54 YRS Gs
SN10-SQ02-06F6	HOTEL main business: 35-54 YRS, STATE BUSINESS MALE Gs, CONVENTIONS
SN11-SQ01-07S4	Gs: majority BUSINESS Gs dir rel CONVENTIONS
SN19-SQ01-11S6b	Gs: majority STATE, 35-54 YRS, BUSINESS, MALE, DIR. RESERV. Gs
SN20-SQ01-02F1	Gs: majority BUSINESS DIR. RESERV. Gs
SN21-SZ02-04F3a	35-54 YRS, BUSINESS Gs majority
SN23-SB01-03S2	BUSINESS Gs overall majority
SN30-SQ01-02F4	Gs: majority 35-54 YRS STATE BUSINESS MALE Gs
SN31-SB04-11F3	BUSINESS Gs majority overall
SN31-SQ01-15F4	Gs: majority 35-54 YRS, BUSINESS, MALE Gs
SN37-SB01-01F2	BUSINESS Gs and DIR. RESERV. Gs constant majority
SN37-SQ01-01S1	Gs: majority LOCAL Gs, BUSINESS Gs, DIR. RESERV. Gs
SN37-SQ02-02S1	Gs: med. import. FEMALE Gs, USA Gs, TOURIST Gs, AG. RESERV. Gs
SN37-SQ03-03S1	Gs: minority AIR CREW Gs, ASIAN Gs, EUROPEAN Gs, S. AMER. Gs, AFR/M.EAST. Gs
SN39-SQ01-01F3	Gs: majority BUSINESS Gs, LOCAL Gs
SN46-SQ01-01S1	Gs: minority AFR/M.EAST. Gs, ASIAN Gs, < 20 YRS Gs, AIR CREW Gs, EUROPEAN Gs
SN46-SQ02-02S1	Gs: majority BUSINESS Gs, LOCAL Gs, next USA Gs, TOURIST Gs
SN64-SB02-02F3	BUSINESS Gs overall majority
SN65-SQ02-01F1	Gs: majority STATE, BUSINESS, MALE, 35-54 YRS, DIR. RESERV. Gs
SN66-SC01-05S3a	STATE, BUSINESS Gs overall majority
SN67-SQ01-03S1a	Gs: majority BUSINESS DIR. RESERV. Gs
SN78-SQ01-01F1	Majority/largest fields: LOCAL Gs, BUSINESS Gs, 20-34 YRS Gs, 35-54 YRS Gs, > 54 YRS Gs, PRICE, STAY, OCCUPANCY
SN80-SQ01-01S1	Gs: majority DIR. RESERV., STATE, BUSINESS, 35-54 YRS Gs in CONVENTIONS, highest JN SE-DE/84 MA AP SE-DE/85

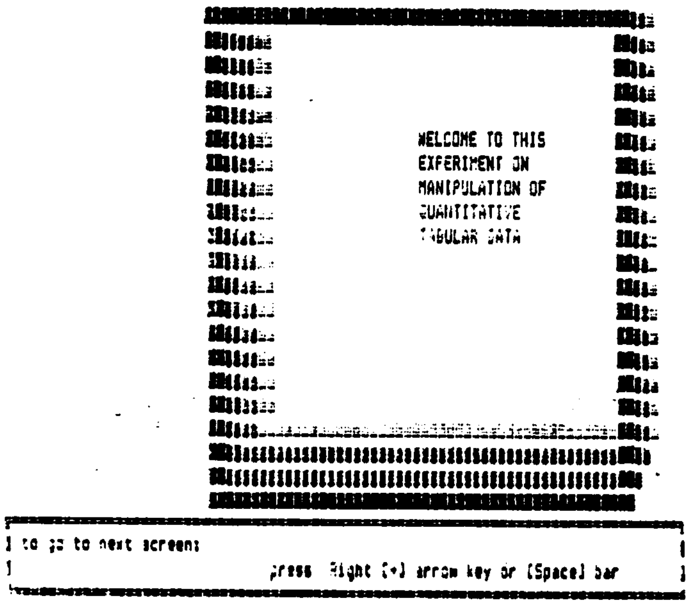
C.7.2 - List of Level "N" "Insights" for "Numeric" Group

SN06-SB04-13S3	BUSINESS Gs overall majority
SN17-SB01-06S4	BUSINESS Gs overall majority JA-MA
SN25-SQ02-04F2b	Gs: majority STATE, 35-54 YRS, BUSINESS MALE Gs
SN27-FV01-03F4	CONVENTIONS major attraction factor
SN28-SQ01-01S1	Gs: majority 35-54 YRS, BUSINESS Gs during CONVENTIONS
SN40-SB01-09S3	BUSINESS Gs overall majority
SN50-SB02-03S2	BUSINESS Gs overall majority
SN50-SZ01-11F6a	35-54 YRS Gs second overall majority
SN58-SQ01-01S1	Gs: majority 35-54 YRS BUSINESS MALE Gs
SN68-SB02-06F4	BUSINESS Gs overall majority; lowest SUMMER
SN70-SQ01-12F6	Gs: majority BUSINESS Gs > 80% 7-8 MO./YR

APPENDIX D. INFORMATION-TUTORIAL SCREENS

The screens for the "Information Tutorial" are presented in the remanining pages. (For more information on other "Tutorial" components, refer to Appendix B.2.) The top screen in each page corresponds to the "graphic" mode of data representation; the bottom one, to the "numeric" mode.


WELCOME TO THIS
EXPERIMENT ON
MANIPULATION OF
QUANTITATIVE
TABULAR DATA
to go to next screen:
press Right [→] arrow key or [Space] bar


WELCOME TO THIS
EXPERIMENT ON
MANIPULATION OF
QUANTITATIVE
TABULAR DATA
to go to next screen:
press Right [→] arrow key or [Space] bar

This experiment is concerned with
the generation of ideas to improve
the performance of a small hotel.

During the next 10 or 15 minutes,
you will be introduced to basic
concepts and procedures available
to analyze the "hotel problem".

After this introductory tutorial,
you will be given more specific
instructions for problem analysis.
[take as long as you need; there
is no time limit for the analysis]

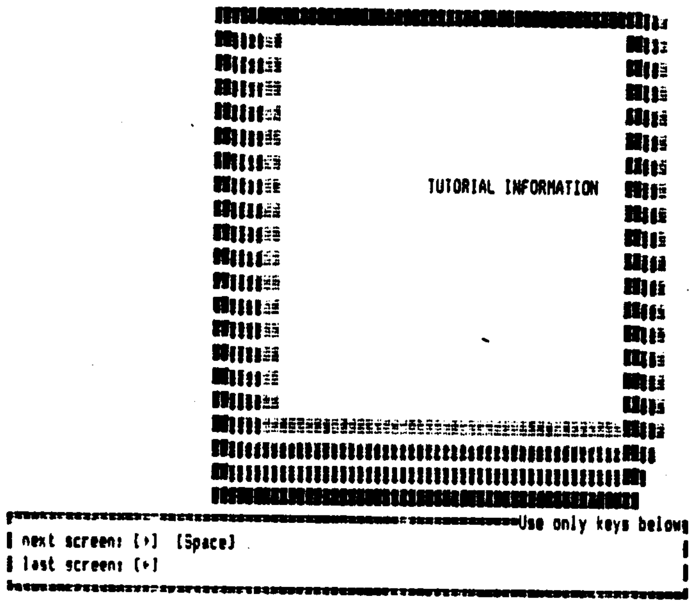
next screen: press Right [→] arrow key or [Space] bar
last screen: press Left [←] arrow key

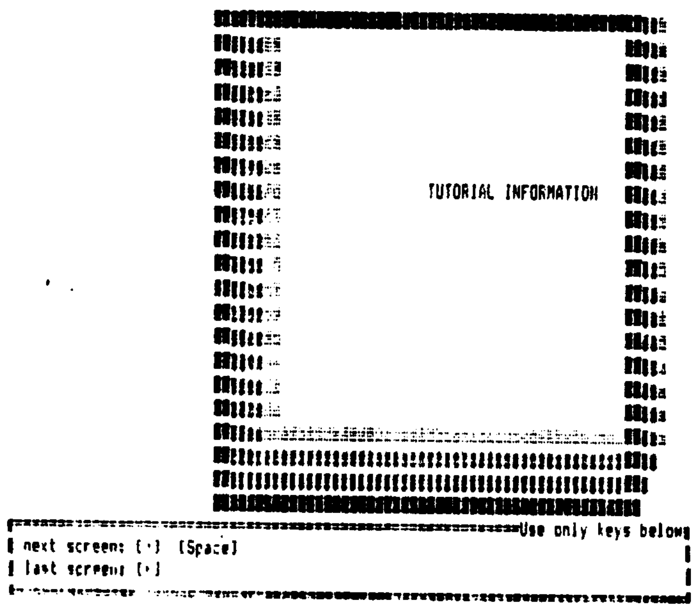
This experiment is concerned with
the generation of ideas to improve
the performance of a small hotel.

During the next 10 or 15 minutes,
you will be introduced to basic
concepts and procedures available
to analyze the "hotel problem".

After this introductory tutorial,
you will be given more specific
instructions for problem analysis.
[take as long as you need; there
is no time limit for the analysis]

next screen: press Right [→] arrow key or [Space] bar
last screen: press Left [←] arrow key


TUTORIAL INFORMATION
Use only keys below
next screen: [ ] [Space]
last screen: [ ]


TUTORIAL INFORMATION
Use only keys below
next screen: [ ] [Space]
last screen: [ ]

TUTORIAL INFORMATION

The tutorial information in the next screens is intended to give you a quick introduction to basic concepts, terms, and a few system commands available to help you with the analysis of the "Hotel ABC" problem.

The tutorial is divided in two parts. The first one deals with the nature of the data available, as well as with representation of that data.
The second part deals with the operation of the system and how information is presented in the screens.

Use only keys below

next screen: [+] [Space]
last screen: [-]

TUTORIAL INFORMATION

The tutorial information in the next screens is intended to give you a quick introduction to basic concepts, terms, and a few system commands available to help you with the analysis of the "Hotel ABC" problem.

The tutorial is divided in two parts. The first one deals with the nature of the data available, as well as with representation of that data.
The second part deals with the operation of the system and how information is presented in the screens.

Use only keys below

next screen: [+] [Space]
last screen: [-]

Information Tutorial (cont.)

NOTE ON KEYS TO BE USED

All the available commands--like table display, window movement, and so on--can be executed simply by pressing the appropriate key on the keyboard.

For each command, the keys available for use will be listed inside a box placed at the bottom of the screen; only those keys are to be used for that particular command.

Please beware of pressing too many keys quickly in sucession. Wait for the result of any particular operation to show up on screen before requesting the next operation.

Use only keys below:

next screen: [→] [Space]

last screen: [←]

NOTE ON KEYS TO BE USED

All the available commands--like table display, window movement, and so on--can be executed simply by pressing the appropriate key on the keyboard.

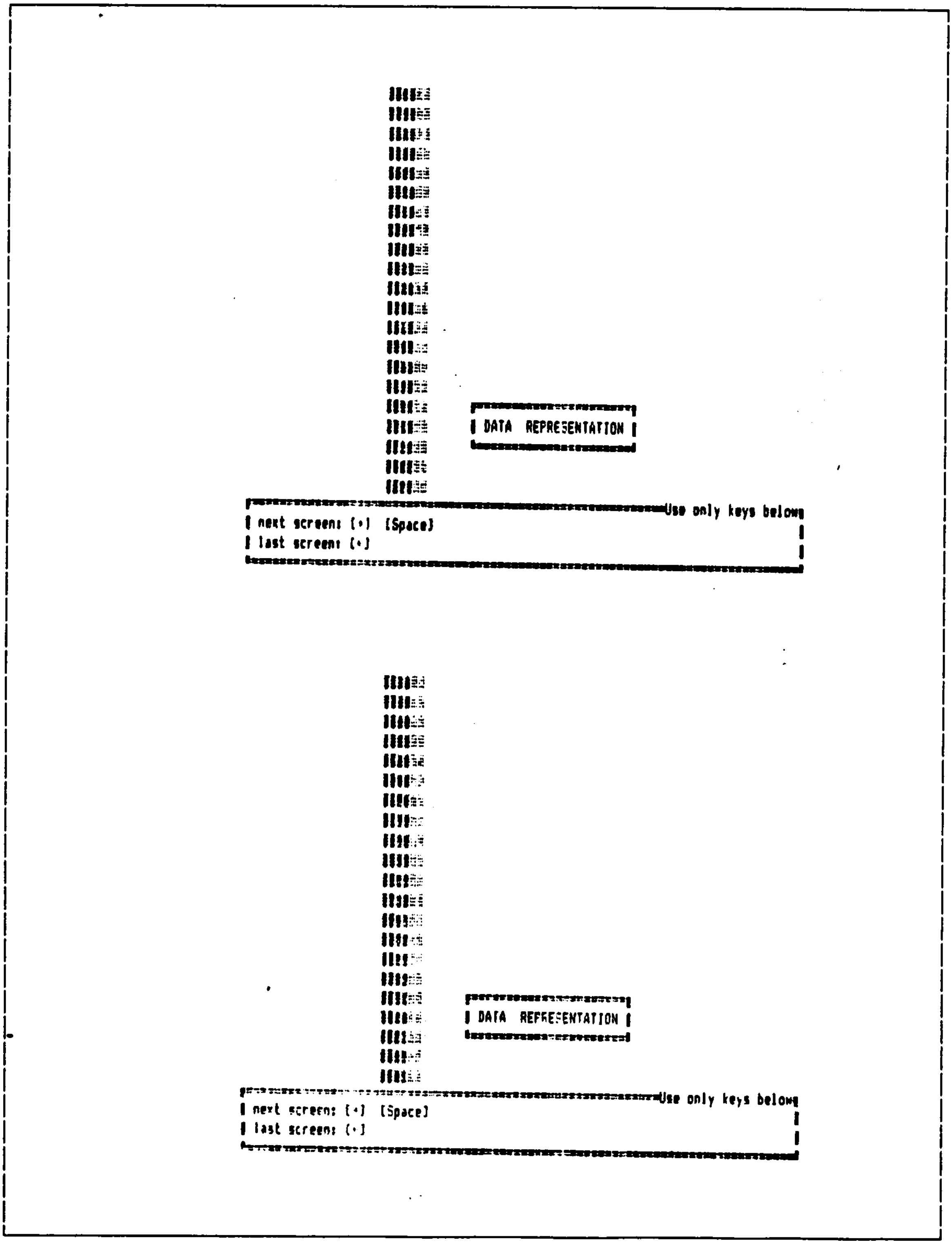
For each command, the keys available for use will be listed inside a box placed at the bottom of the screen; only those keys are to be used for that particular command.

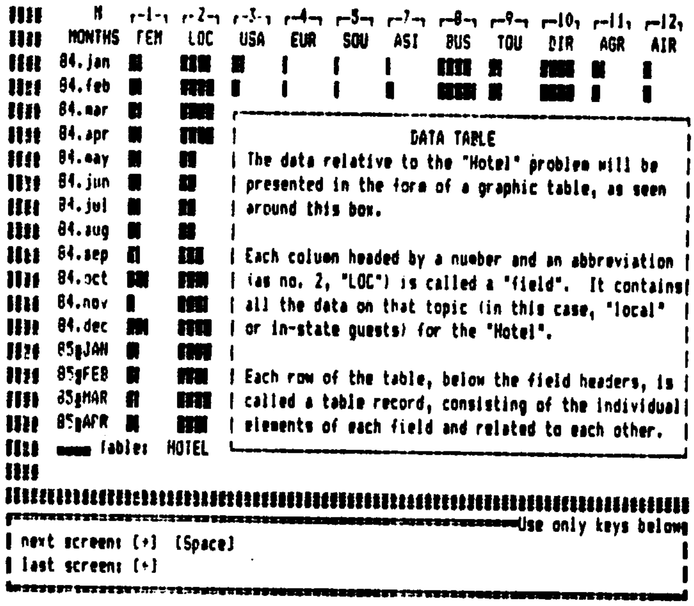
Please beware of pressing too many keys quickly in sucession. Wait for the result of any particular operation to show up on screen before requesting the next operation.

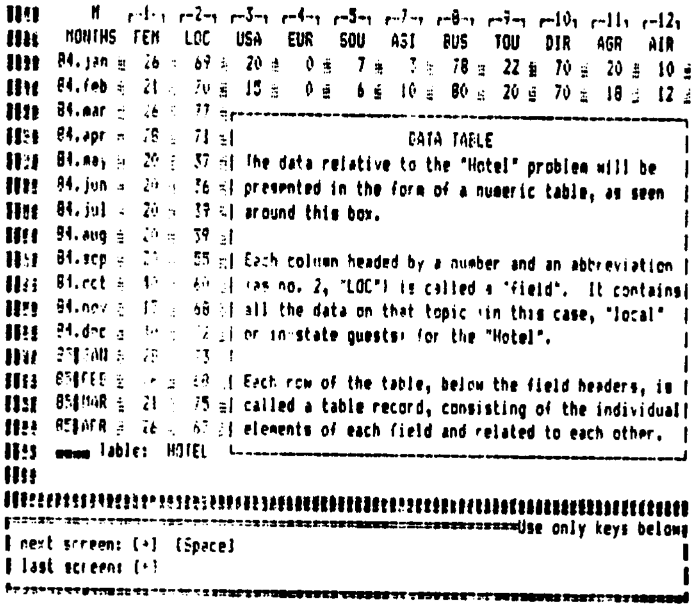
Use only keys below:

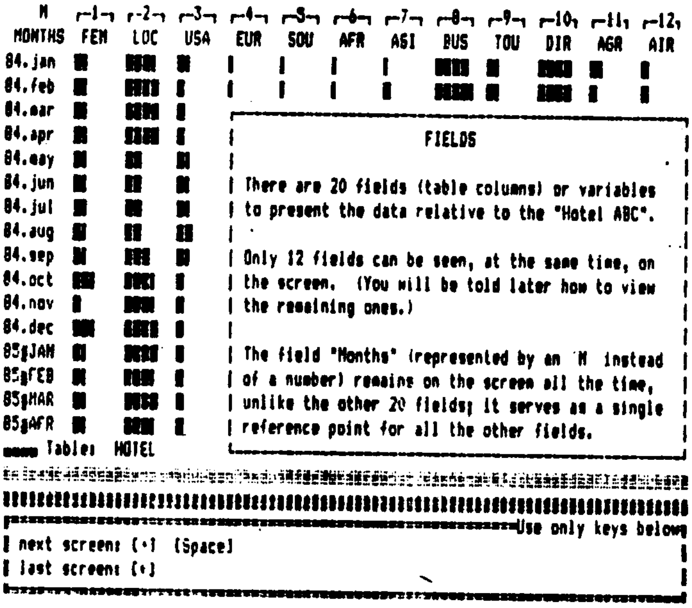
next screen: [→] [Space]

last screen: [←]


DATA REPRESENTATION
Use only keys below
next screen: [→] [Space]
last screen: [←]
DATA REPRESENTATION
Use only keys below
next screen: [→] [Space]
last screen: [←]


M
MONTHS FEM LOC USA EUR SOU ASI BUS TOU DIR AGR AIR
84.jan
84.feb
84.mar
84.apr
84.may
84.jun
84.jul
84.aug
84.sep
84.oct
84.nov
84.dec
85JAN
85FEB
85MAR
85APR
Table: HOTEL
DATA TABLE
The data relative to the "Hotel" problem will be presented in the form of a graphic table, as seen around this box.
Each column headed by a number and an abbreviation (as no. 2, "LOC") is called a "field". It contains all the data on that topic (in this case, "local" or in-state guests) for the "Hotel".
Each row of the table, below the field headers, is called a table record, consisting of the individual elements of each field and related to each other.
Use only keys below
next screen: (+) (Space)
last screen: (+)


M
MONTHS FEM LOC USA EUR SOU ASI BUS TOU DIR AGR AIR
84.jan 26 69 20 0 7 78 22 70 20 10
84.feb 21 70 15 0 6 10 80 20 70 18 12
84.mar
84.apr
84.may
84.jun
84.jul
84.aug
84.sep
84.oct
84.nov
84.dec
Table: HOTEL
DATA TABLE
The data relative to the "Hotel" problem will be presented in the form of a numeric table, as seen around this box.
Each column headed by a number and an abbreviation (as no. 2, "LOC") is called a "field". It contains all the data on that topic (in this case, "local" or in-state guests) for the "Hotel".
Each row of the table, below the field headers, is called a table record, consisting of the individual elements of each field and related to each other.
Use only keys below
next screen: (+) (Space)
last screen: (+)



M MONTHS	1 FEM	2 LOC	3 USA	4 EUR	5 SOU	6 AFR	7 ASI	8 BUS	9 TOU	10 DIR	11 AGR	12 AIR
84.jan	26	67	20	0	7	1	3	78	22	70	20	10
84.feb	21	70	15	0	6	0	10	80	20	70	18	12
84.mar	26	77	14									
84.apr	28	71	15									
84.may	20	57	23									
84.jun	20	36	27									
84.jul	20	39	22									
84.aug	20	37	30									
84.sep	20	55	27									
84.oct	40	60	17									
84.nov	15	68	15									
84.dec	40	72	17									
85 JAN	28	73	17									
85 FEB	26	68	18									
85 MAR	21	75	16									
85 AFR	26	67	16									

FIELDS

There are 20 fields (table columns) or variables to present the data relative to the "Hotel ABC".

Only 12 fields can be seen, at the same time, on the screen. (You will be told later how to view the remaining ones.)

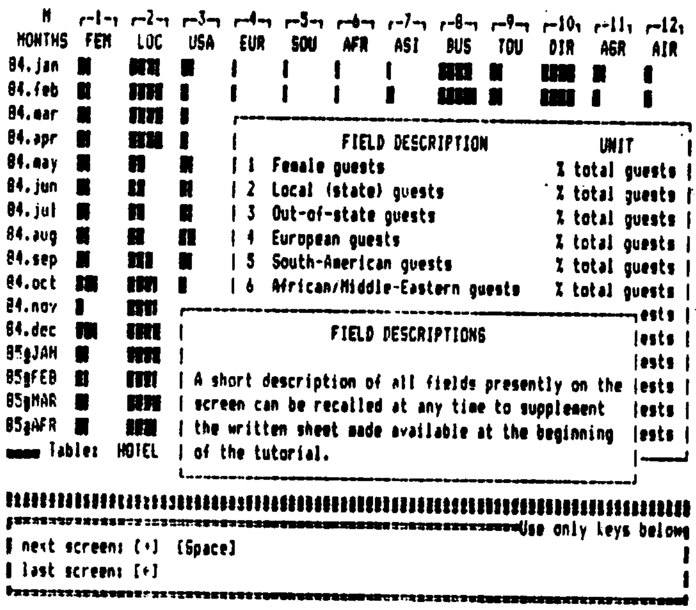
The field "Months" (represented by an 'M' instead of a number) remains on the screen all the time, unlike the other 20 fields; it serves as a single reference point for all the other fields.

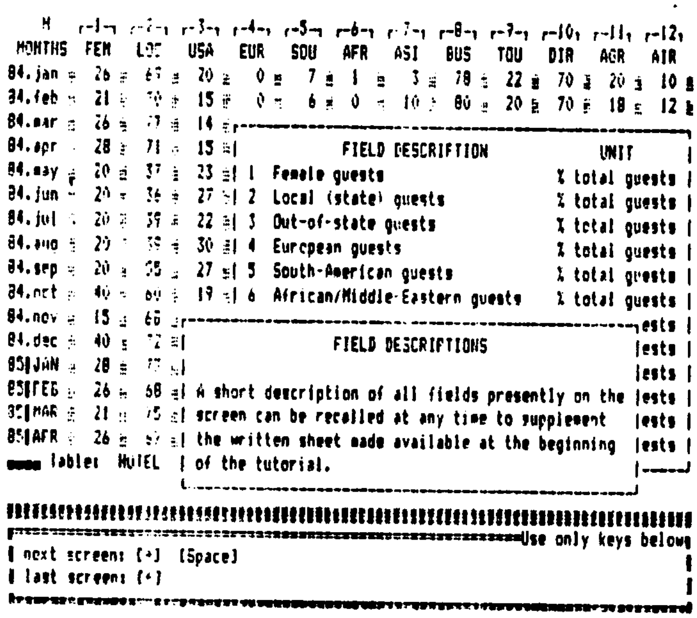
Table: HOTEL

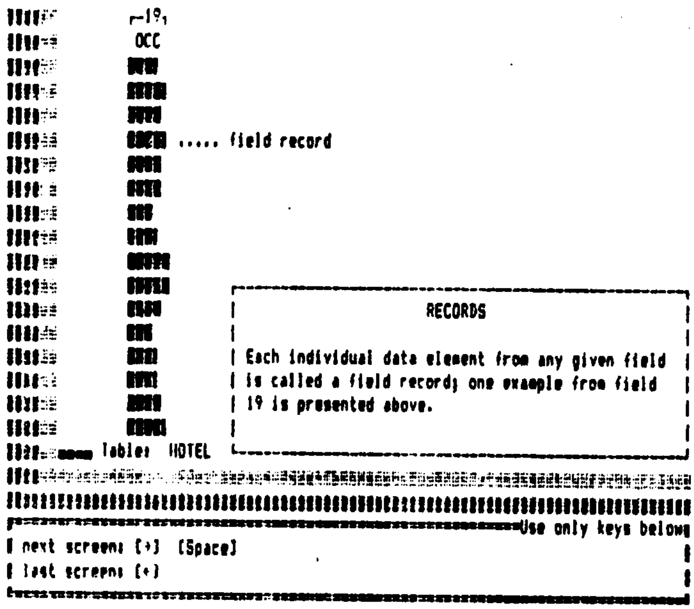
Use only keys below

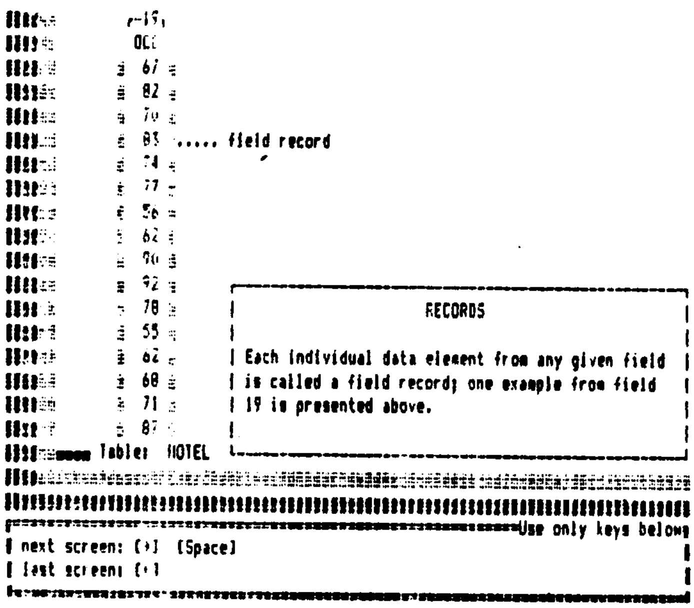
next screen: (+) (Space)

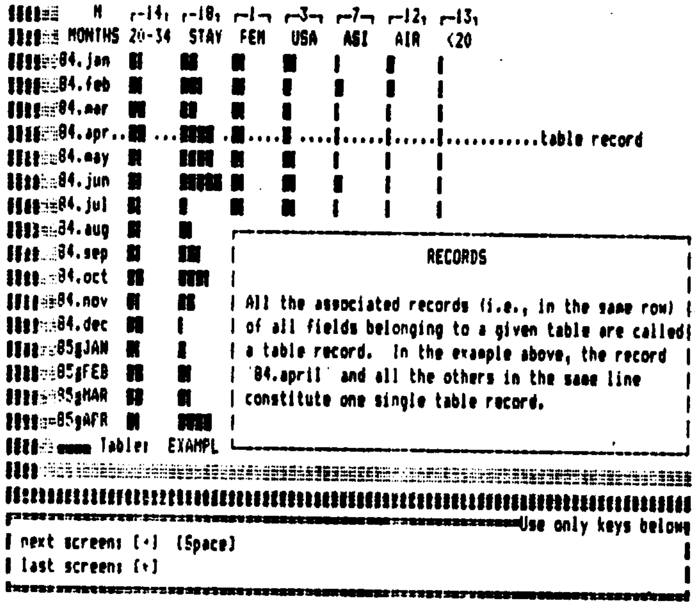
last screen: (+)

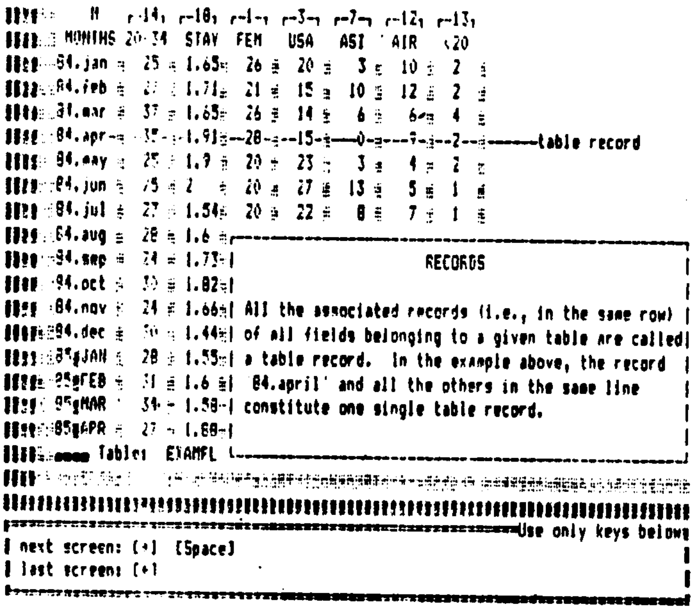

M
MONTHS FEM LOC USA EUR SOU AFR ASI BUS TOU DIR AGR AIR
84.jan
84.feb
84.mar
84.apr
84.may
84.jun
84.jul
84.aug
84.sep
84.oct
84.nov
84.dec
85 JAN
85 FEB
85 MAR
85 APR
Table: HOTEL
FIELD DESCRIPTION UNIT
1 Female guests % total guests
2 Local (state) guests % total guests
3 Out-of-state guests % total guests
4 European guests % total guests
5 South-American guests % total guests
6 African/Middle-Eastern guests % total guests
FIELD DESCRIPTIONS
A short description of all fields presently on the screen can be recalled at any time to supplement the written sheet made available at the beginning of the tutorial.
Use only keys below
next screen: [↓] [Space]
last screen: [↑]


M
MONTHS FEM LOC USA EUR SOU AFR ASI BUS TOU DIR AGR AIR
84.jan 26 20 0 7 1 3 78 22 70 20 10
84.feb 21 15 0 6 0 10 80 20 70 18 12
84.mar 26 14
84.apr 28 71 15
84.may 20 37 23
84.jun 36 27
84.jul 20 39 22
84.aug 30
84.sep 20 27
84.oct 40 19
84.nov 15 66
84.dec 40
85 JAN 28
85 FEB 26 68
85 MAR 21 75
85 APR 26
Table: HOTEL
FIELD DESCRIPTION UNIT
1 Female guests % total guests
2 Local (state) guests % total guests
3 Out-of-state guests % total guests
4 European guests % total guests
5 South-American guests % total guests
6 African/Middle-Eastern guests % total guests
FIELD DESCRIPTIONS
A short description of all fields presently on the screen can be recalled at any time to supplement the written sheet made available at the beginning of the tutorial.
Use only keys below
next screen: [↓] [Space]
last screen: [↑]


19
OCC
..... field record
RECORDS
Each individual data element from any given field is called a field record; one example from field 19 is presented above.
Table: HOTEL
Use only keys below
next screen: (→) (Space)
last screen: (←)

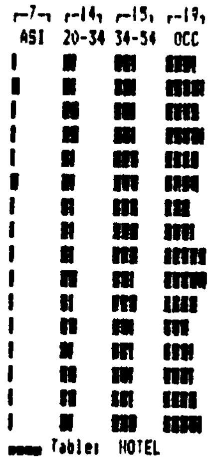

19
OCC
67
82
83 ..... field record
77
62
92
78
55
68
71
RECORDS
Each individual data element from any given field is called a field record; one example from field 19 is presented above.
Table: HOTEL
Use only keys below
next screen: (→) (Space)
last screen: (←)


M MONTHS 20-34 STAY FEM USA ASI AIR <20
84.jan
84.feb
84.mar
84.apr
table record
84.may
84.jun
84.jul
84.aug
84.sep
84.oct
84.nov
84.dec
85JAN
85FEB
85MAR
85APR
Table: EXAMPL
RECORDS
All the associated records (i.e., in the same row) of all fields belonging to a given table are called a table record. In the example above, the record '84.april' and all the others in the same line constitute one single table record.
Use only keys below
next screen: [→] [Space]
last screen: [←]


M MONTHS 20-34 STAY FEM USA ASI AIR <20
84.jan 25 1.65 26 20 3 10 2
84.feb 1.71 21 15 10 12 2
84.mar 37 1.65 26 14 6 6 4
84.apr 1.91 28 15 0 7 2
table record
84.may 25 1.7 20 23 3 4 2
84.jun 25 2 20 27 13 5 1
84.jul 27 1.54 20 22 8 7 1
84.aug 28 1.6
84.sep 24 1.73
84.oct 1.82
84.nov 24 1.66
84.dec 1.44
85JAN 28 1.55
85FEB 31 1.6
85MAR 34 1.58
85APR 27 1.68
Table: EXAMPL
RECORDS
All the associated records (i.e., in the same row) of all fields belonging to a given table are called a table record. In the example above, the record '84.april' and all the others in the same line constitute one single table record.
Use only keys below
next screen: [→] [Space]
last screen: [←]

```
          M     -1-   -2-   -3-   -4-              -16-  -17-  -18-  -19-  -20-
     MONTHS   FEM   LOC   USA   EUR              >55   PRICE STAY  OCC   CONV
1.....84.jan  █     ███   █     ▌     .... + .... █     ███   ██    ███   ▌
2.....84.feb  █     ███   ▌     ▌     .... + .... █     ███   ██    ███   ▌
3.... 84.mar  █     ███   ┌───────────────────────────────────────────────┐
4...  84.apr  █     ███   |                   RECORDS                     |
5...  84.may  █     ██    |                                               |
6..   84.jun  █     ██    | All the table rows presently available in a given |
7.    84.jul  █     ██    | table are called (surprise!) table records.  There |
↓     ↓       ↓     ↓     | are 24 records in the original table ("HOTEL")  |
↓     ↓       ↓     ↓     | made available in the system; they range from   |
18    85¦JUN  █     ██    | 1984 January to 1985 December.                  |
19.   85¦JUL  █     ██    |                                               |
20..  85¦AUG  █     ██    | With 20 fields (not including the "Months" field) |
21..  85¦SEP  █     ███   | there are, as a result, a total of 480 elements |
22... 85¦OCT  █     ███   | (field records, or individual data points) in the |
23....85¦NOV  █     ███   | original dataset. [As such datasets go, it could |
24....85¦DEC  █     ███   | be called medium size]                          |
                          └───────────────────────────────────────────────┘
==========================================================Use only keys below
| next screen: [→] [Space]
| last screen: [←]
```

```
          M     -1-   -2-   -4-   -5-                  -16-  -17-  -18-  -19-  -20-
     MONTHS   FEM   LOC   EUR   SOU                  >55   PRICE STAY  OCC   CONV
1.....84.jan   26    69     0     7   .... + ....     25    74   1.65   67    0
2.....84.feb   21    70     0     6   .... + ....     22    76   1.71   82    0
3.... 84.mar   25    77   ┌───────────────────────────────────────────────┐
4...  84.apr   25    71   |                   RECORDS                     |
5...  84.may   20    37   |                                               |
6..   84.jun   20    36   | All the table rows presently available in a given |
7.    84.jul   20    39   | table are called (surprise!) table records.  There |
↓     ↓       ↓     ↓     | are 24 records in the original table ("HOTEL")  |
↓     ↓       ↓     ↓     | made available in the system; they range from   |
18    85¦JUN   20    35   | 1984 January to 1985 December.                  |
19.   85¦JUL   20    42   |                                               |
20..  85¦AUG   20    43   | With 20 fields (not including the "Months" field) |
21..  85¦SEP   26    54   | there are, as a result, a total of 480 elements |
22... 85¦OCT   38    58   | (field records, or individual data points) in the |
23....85¦NOV   22    71   | original dataset. [As such datasets go, it could |
24....85¦DEC   35    70   | be called medium size]                          |
                          └───────────────────────────────────────────────┘
==========================================================Use only keys below
| next screen: [→] [Space]
| last screen: [←]
```



RECORD RANGE

Almost all the fields (with the exception of "17", "18", and "20") are represented in terms of the percentage of the total number of "Hotel ABC" guests. The records for those fields, therefore, range from 0 to 100, at least in principle, and are directly comparable with each other.

The values for each record from a graphic field are represented by horizontal bars of different length (there are 10 intervals in all). The smallest value is shown as the 1st record for fld. "7"; the largest one is shown as the 10th record (from the top) for field "19". The other highlighted records correspond to the remaining intervals.

Use only keys below

next screen: (→) (Space)
last screen: (←)

7 ASI	14 20-34	15 34-54	19 OCC
3	25	48	67
19	27	49	82
6	37	42	70
11	35	48	83
3	25	54	74
13	25	55	77
8	27	53	56
9	28	51	62
5	24	55	90
2	30	46	72
5	24	55	78
2	30	[illegible]	55
1	28	46	62
3	31	43	68
4	34	15	71
3	27	51	87

Table: HOTEL

RECORD RANGE

Almost all the fields (with the exception of "17", "18", and "20") are represented in terms of the percentage of the total number of "Hotel ABC" guests. The records for these fields, therefore, range from 0 to 100, at least in principle, and are directly comparable with each other.

Use only keys below

next screen: (→) (Space)
last screen: (←)

M MONTHS	1 FEM	2 LOC	3 USA	4 EUR	5 SOU	7 ASI	8 BUS	9 TOU	10 DIR	11 AGR	12 AIR
84.jan											
84.feb											
84.mar											
84.apr											
84.may											
84.jun											
84.jul											
84.aug											
84.sep											
84.oct											
84.nov											
84.dec											
85 JAN											
85 FEB											
85 MAR											
85 APR											

Table: HOTEL

TABLE MANIPULATION

The initial table (named "Hotel") made available to you can be manipulated at will, i.e., you can create new, smaller tables (subsets) carved out of the original dataset. You may also change the order of the records by sorting (grading) any given field, in ascending or descending order.

Even if you decide no other manipulations are necessary for your analysis, you will probably want to move around the "window" showing the parts of a data table which fit on the screen.

Use only keys below

next screen: [+] (Space)
last screen: [-]

M MONTHS	1 FEM	2 LOC	3 USA	4 EUR	5 SOU	7 ASI	8 BUS	9 TOU	10 DIR	11 AGR	12 AIR
84.jan	26	69	20	0	7	3	78	22	70	20	10
84.feb	21	70	15	0	6	10	80	20	70	18	12
84.mar	26	77									
84.apr	29	71									
84.may	24	37									
84.jun	20	36									
84.jul	20	39									
84.aug	20	37									
84.sep	20	55									
84.oct	10	60									
84.nov	15	68									
84.dec	40	72									
85 JAN	28	73									
85 FEB	26	68									
85 MAR	21	75									
85 APR	26	67									

Table: HOTEL

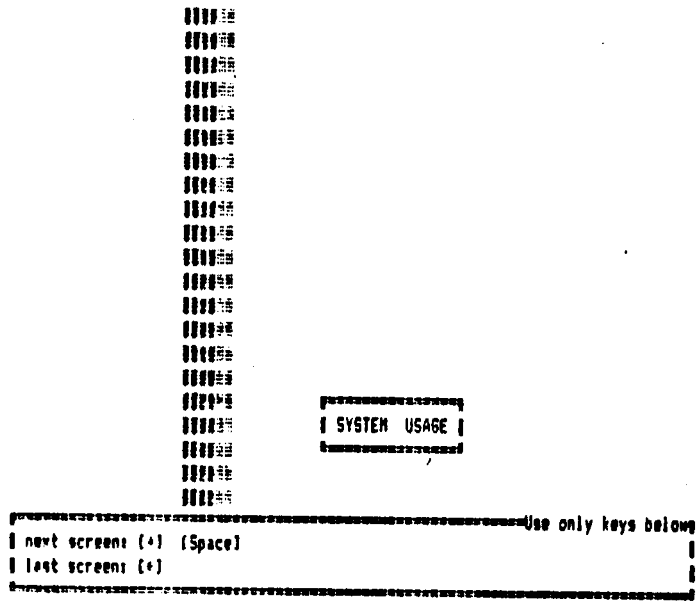
TABLE MANIPULATION

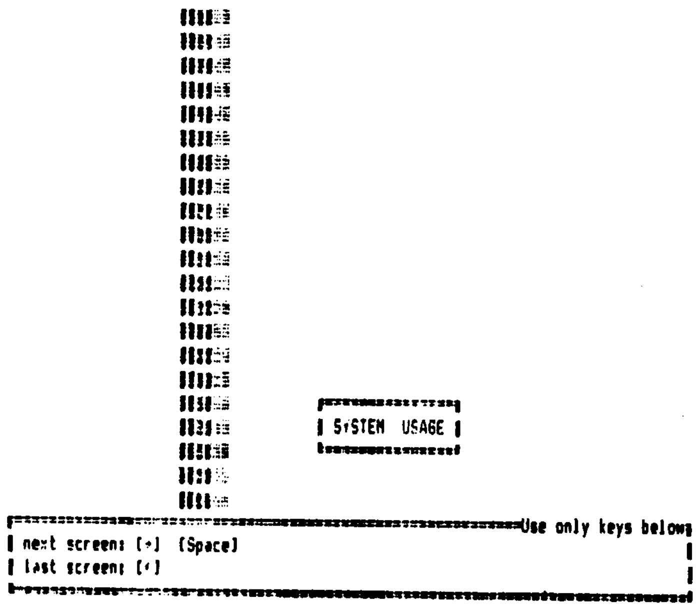
The initial table (named "Hotel") made available to you can be manipulated at will, i.e., you can create new, smaller tables (subsets) carved out of the original dataset. You may also change the order of the records by sorting (grading) any given field, in ascending or descending order.

Even if you decide no other manipulations are necessary for your analysis, you will probably want to move around the "window" showing the parts of a data table which fit on the screen.

Use only keys below

next screen: [+] (Space)
last screen: [-]


SYSTEM USAGE
Use only keys below
next screen: [→] [Space]
last screen: [←]


SYSTEM USAGE
Use only keys below
next screen: [→] [Space]
last screen: [←]

```
                    SCREEN AREAS

 There are four major screen areas in the displays
 for the present system.  The first one is the
 table area, where table fields and records are
 displayed.  Occasionally, an overlapping box like
 the present one can be placed inside this area, if
 there is not enough space in the messages area to
 present, at that moment, relevant information.

 The messages area will present all the commands,
 their requirements, and any resulting messages,
 necessary to operate the system.      (continue →)
TABLE NAME
          MESSAGES

Use only keys below
 next screen: [→]  [Space]
 last screen: [←]


          TABLE
```



```
                    SCREEN AREAS

 There are four major screen areas in the displays
 for the present system.  The first one is the
 table area, where table fields and records are
 displayed.  Occasionally, an overlapping box like
 the present one can be placed inside this area, if
 there is not enough space in the messages area to
 present, at that moment, relevant information.

 The messages area will present all the commands,
 their requirements, and any resulting messages,
 necessary to operate the system.      (continue →)
TABLE NAME
          MESSAGES

Use only keys below
 next screen: [→]  [Space]
 last screen: [←]
```

TABLE

SCREEN AREAS

The third one is the table name area, presenting, as the name suggests, the name (if any) of the current table.

The fourth, and last, area corresponds to the list of available keys at any given moment. The list is presented inside the box at the bottom of each screen. Only the keys indicated are in use at the time; use of any others will simply result in a warning message.

TABLE NAME

MESSAGES

Use only keys below

next screen: [→] [Space]
last screen: [←]

TABLE

SCREEN AREAS

The third one is the table name area, presenting, as the name suggests, the name (if any) of the current table.

The fourth, and last, area corresponds to the list of available keys at any given moment. The list is presented inside the box at the bottom of each screen. Only the keys indicated are in use at the time; use of any others will simply result in a warning message.

TABLE NAME

MESSAGES

Use only keys below

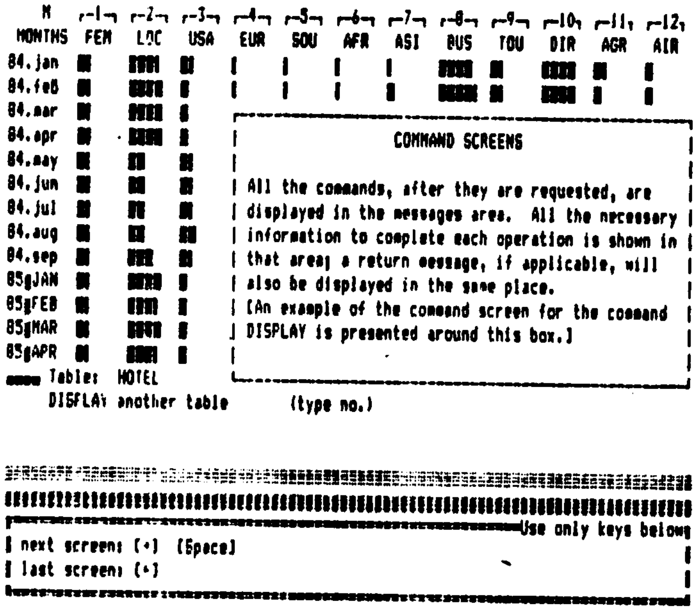
next screen: [→] [Space]
last screen: [←]

```
              M   ┌-1-┐ ┌-2-┐ ┌-3-┐ ┌-4-┐             ┌-16┐ ┌-17┐ ┌-18┐ ┌-19┐ ┌-20┐
         • MONTHS  FEM   LOC   USA   EUR                >55  FRICE STAY  OCC  CONV
1.....84.jan   █    ███   █     ▌    .... ♦ ....   █    ██    █    ███   ▌
2.....84.feb   █    ███   ▌     ▌    .... ♦ ....   █    ███   ██   ████  ▌
3.... 84.mar   █    ███ ┌──────────────────────────────────────────────────┐
4...  84.apr   █    ███ │               WINDOW DISPLACEMENT                │
5...  84.may   █    ██  │                                                  │
6..   84.jun   █    ██  │ The original data table (named "Hotel") is larger│
7.    84.jul   █    ██  │ than can be accomodated in the screen.  What you │
♦      ♦       ♦    ♦   │ are able to see is a fraction or a subset of, or │
♦      ♦       ♦    ♦   │ a "window" on the data, so to speak.             │
18    85▌JUN   █    ██  │                                                  │
19.   85▌JUL   █    ██  │ To move this window around, the four arrow keys  │
20..  85▌AUG   █    ███ │ ([→][←][↑][↓]) will be made available.  After the│
21..  85▌SEP   █    ███ │ completion (or cancellation) of each command, the│
22... 85▌OCT   ██   ███ │ "WINDOW DISPLACEMENT" command will be displayed  │
23....85▌NOV   █    ████│ once more, to allow you to move the window around.│
24....85▌DEC   ██   ████│                                                  │
                        └──────────────────────────────────────────────────┘

┌──────────────────────────────────────────────────────────Use only keys below
│ next screen: [→]  [Space]                                                  │
│ last screen: [←]                                                           │
└────────────────────────────────────────────────────────────────────────────┘
```

```
              M   ┌-1-┐ ┌-2-┐ ┌-4-┐ ┌-5-┐             ┌-16┐ ┌-17┐ ┌-18┐ ┌-19┐ ┌-20┐
           MONTHS  FEM   LOC   EUR   SOU                >55  FRICE STAY  OCC  CONV
1.....84.jan     26    67     0     7 .... ♦ ....    25    74   1.65   67    0
2.....84.feb     21    70     0     6 .... ♦ ....    22    76   1.71   82    0
3.... 84.mar     26    77 ┌──────────────────────────────────────────────────┐
4...  84.apr     28    71 │               WINDOW DISPLACEMENT                │
5...  84.may     20    37 │                                                  │
6..   84.jun     20    36 │ The original data table (named "Hotel") is larger│
7.    84.jul     20    39 │ than can be accomodated in the screen.  What you │
♦      ♦       ♦      ♦   │ are able to see is a fraction or a subset of, or │
♦      ♦       ♦      ♦   │ a "window" on the data, so to speak.             │
18    85▌JUN     24    35 │                                                  │
19.   85▌JUL     20    42 │ To move this window around, the four arrow keys  │
20..  85▌AUG     20    43 │ ([→][←][↑][↓]) will be made available.  After the│
21..  85▌SEP     26    54 │ completion (or cancellation) of each command, the│
22... 85▌OCT     38    58 │ "WINDOW DISPLACEMENT" command will be displayed  │
23....85▌NOV     27    71 │ once more, to allow you to move the window around.│
24....85▌DEC     35    70 │                                                  │
                          └──────────────────────────────────────────────────┘

┌──────────────────────────────────────────────────────────Use only keys below
│ next screen: [→]  [Space]                                                  │
│ last screen: [←]                                                           │
└────────────────────────────────────────────────────────────────────────────┘
```

```
  M     ┌-1-┐ ┌-2-┐ ┌-3-┐ ┌-4-┐ ┌-5-┐ ┌-6-┐ ┌-7-┐ ┌-8-┐ ┌-9-┐ ┌-10┐ ┌-11┐ ┌-12┐
MONTHS   FEM   LOC   USA   EUR   SOU   AFR   ASI   BUS   TOU   DIR   AGR   AIR
84.jan
84.feb
84.mar        ┌──────────────────────────────────────────────────┐
84.apr        │                  COMMAND MENU                    │
84.may        │                                                  │
84.jun        │ You may request, from the "WINDOW DISPLACEMENT"  │
84.jul        │ screen, the command menu options.  There are ten │
84.aug        │ commands in all, listed in the messages area.    │
84.sep        │ A short description of each one is presented in  │
85.JAN        │ one of the written sheets made available to you, │
85.FEB        │ at the beginning of the tutorial.  [More detailed│
85.MAR        │ information on each command: use HELP command to │
85.APR        │ access HELP menu]                                │
Table: HOTEL  └──────────────────────────────────────────────────┘
DISPLAY table   SAVE new table  NAMES-tables    PROJECT fields  SORT field
RETURN prev. tb.DROP table      FIELDS-descrip. SELECT records  HELP

══════════════════════════════════════════════════Use only keys below
│ next screen: (→)  [Space]                                          │
│ last screen: (←)                                                   │
└────────────────────────────────────────────────────────────────────┘
```

```
  M     ┌-1-┐ ┌-2-┐ ┌-3-┐ ┌-4-┐ ┌-5-┐ ┌-6-┐ ┌-7-┐ ┌-8-┐ ┌-9-┐ ┌-10┐ ┌-11┐ ┌-12┐
MONTHS   FEM   LOC   USA   EUR   SOU   AFR   ASI   BUS   TOU   DIR   AGR   AIR
84.jan    26    69    20     0     7     1     3    78    22    70    20    10
84.feb    21    76    15     0     6     0    10    80    20    70    18    12
84.mar    26    77    14 ┌──────────────────────────────────────────────────┐
84.apr    28    71    15 │                  COMMAND MENU                    │
84.may    20    37    23 │                                                  │
84.jun    20    36    27 │ You may request, from the "WINDOW DISPLACEMENT"  │
84.jul    20    35    22 │ screen, the command menu options.  There are ten │
84.aug    20    37    30 │ commands in all, listed in the messages area.    │
84.sep    20    35    27 │ A short description of each one is presented in  │
85.JAN    28    73    17 │ one of the written sheets made available to you, │
85.FEB    26    65    18 │ at the beginning of the tutorial.  [More detailed│
85.MAR    21    75    16 │ information on each command: use HELP command to │
85.APR    26    67    16 │ access HELP menu]                                │
Table: HOTEL             └──────────────────────────────────────────────────┘
DISPLAY table   SAVE new table  NAMES-tables    PROJECT fields  SORT field
RETURN prev. tb.DROP table      FIELDS-descrip. SELECT records  HELP

══════════════════════════════════════════════════Use only keys below
│ next screen: (→)  [Space]                                          │
│ last screen: (←)                                                   │
└────────────────────────────────────────────────────────────────────┘
```



MONTHS	1 FEM	2 LOC	3 USA	4 EUR	5 SOU	6 AFR	7 ASI	8 BUS	9 TOU	10 DIR	11 AGR	12 AIR
84.jan	26	49	20	0	7	1	3	78	22	70	20	10
84.feb	21	70	15	0	6	0	10	80	20	70	18	12
84.mar	26	77	14									
84.apr	28	71	15									
84.may	20	27	23									
84.jun	20	36	27									
84.jul	20	37	22									
84.aug	20	35	30									
84.sep	20	55	27									
85JAN	28	73	17									
85FEB	26	68	18									
85MAR	21	75	16									
85APR	26	57	16									

COMMAND SCREENS

All the commands, after they are requested, are displayed in the messages area. All the necessary information to complete each operation is shown in that area; a return message, if applicable, will also be displayed in the same place.
[An example of the command screen for the command DISPLAY is presented around this box.]

Table: HOTEL
DISPLAY another table (type no.)

Use only keys below
next screen: [→] [Space]
last screen: [←]

```
   M    r-1-¬ r-2-¬ r-3-¬ r-4-¬ r-5-¬ r-6-¬ r-7-¬ r-8-¬ r-9-¬ r-10¬ r-11¬ r-12¬
MONTHS   FEM   LOC   USA   EUR   SOU   AFR   ASI   BUS   TOU   DIR   AGR   AIR
84.jan
84.feb                 |               PROJECT                    |
84.mar                 |                                          |
84.apr                 | An introductory explanation may be in order for   |
84.may                 | three commands: Project, Select, and Sort.  With  |
84.jun                 | the first one--Project field(s)--you can specify  |
84.jul                 | which of the 20 original fields to retain in order|
84.aug                 | to create a new table.                    |
84.sep                 |                                          |
84.oct                 | In the example around this box, the four fields   |
85 JAN                 | highlighted indicate that they have been chosen   |
85 FEB                 | for projection, and the resulting table (in a new |
85 MAR                 | screen) will present only those fields.   |
85 APR                 |                                          |
     Table:  HOTEL
     PROJECT chosen FIELDS (columns)

                                                     Use only keys below
| next screen: (→)  [Space]                                            |
| last screen: (←)                                                     |
```

```
   M    r-1-¬ r-2-¬ r-3-¬ r-4-¬ r-5-¬ r-6-¬ r-7-¬ r-8-¬ r-9-¬ r-10¬ r-11¬ r-12¬
MONTHS   FEM   LOC   USA   EUR   SOU   AFR   ASI   BUS   TOU   DIR   AGR   AIR
84.jan    26    69    20
84.feb    21    70    15 |               PROJECT                    |
84.mar    26    77    14 |                                          |
84.apr    28    71    15 | An introductory explanation may be in order for   |
84.may    20    37    23 | three commands: Project, Select, and Sort.  With  |
84.jun    29    36    27 | the first one--Project field(s)--you can specify  |
84.jul    20    32    22 | which of the 20 original fields to retain in order|
84.aug    20    35    30 | to create a new table.                    |
84.sep    20    55    27 |                                          |
84.oct    40    60    19 | In the example around this box, the four fields   |
85 JAN    28    73    17 | highlighted indicate that they have been chosen   |
85 FEB    26    58    18 | for projection, and the resulting table (in a new |
85 MAR    21    75    16 | screen, will present only those fields.   |
85 AFR    26    67    16 |                                          |
     Table:  HOTEL
     PROJECT chosen FIELDS (columns)

                                                     Use only keys below
| next screen: (→)  [Space]                                            |
| last screen: (←)                                                     |
```

M MONTHS	9 TOU	10 DIR	11 AGR	12 AIR	13 <20	14 20-34	15 34-54	16 >55	17 PRICE	18 STAY	19 OCC	20 CONV
84.jan												
84.feb												
84.mar												
84.apr												
84.may												
84.jun												
84.jul												
84.aug												
84.sep												
84.oct												
85JAN												
85FEB												
85MAR												
85APR												

SELECT

The second command--Select record(s)--allows you to specify which ones (from up to 24 original records of any given table) to retain in order to create a new table.

In the example to the left of this box, the six highlighted records from field 10 have been chosen for selection, and the resulting table (in a new screen) will present only those records. The original number of fields remains unchanged.

Table: HOTEL

SELECT one or several RECORDS (rows)

Use only keys below

next screen: (→) (Space)

last screen: (←)

M MONTHS	9 TOU	10 DIR	11 AGR	12 AIR	13 <20	14 20-34	15 34-54	16 >55	17 PRICE	18 STAY	19 OCC	20 CONV
84.jan	22	70	20									
84.feb	20	70	18									
84.mar	15	75	19									
84.apr	14	74	17									
84.may	15	69	27									
84.jun	13	68	27									
84.jul	30	74	19									
84.aug	24	75	17									
84.sep	13	66	26									
84.oct	15	68	27									
85JAN	18	74	17									
85FEB	17	69	21									
85MAR	18	73	22									
85APR	14	68	25									

SELECT

The second command--Select record(s)--allows you to specify which ones (from up to 24 original records of any given table) to retain in order to create a new table.

In the example to the left of this box, the six highlighted records from field 10 have been chosen for selection, and the resulting table (in a new screen) will present only those records. The original number of fields remains unchanged.

Table: HOTEL

SELECT one or several RECORDS (rows)

Use only keys below

next screen: (→) (Space)

last screen: (←)

```
     M     -10-  -17-  -18-  -19-  -20-
 MONTHS   DIR  PRICE  STAY  OCC   CONV
 85 APR
 85 OCT
 84.apr
 84.oct
 85 NOV
 85 AUG
 84.feb
 84.aug
 85 DEC
 84.jan
 84.mar
 84.nov
 85 FEB
 85 MAR
 ____ Tables
```

```
+--------------------------------------------------+
|                       SORT                       |
|                                                  |
| The third and last command to be explained here is |
| Sort field.  With this command, you may specify  |
| an order for rearrangement of the records of that |
| particular field (ascending or descending).      |
|                                                  |
| In the example to the left of this box, field 17 |
| has been sorted already, in ascending order, and |
| all the other field records have also be moved   |
| in conjunction with the records of field 17.     |
| The original number of fields and records remains |
| unchanged.                                       |
+--------------------------------------------------+
```

```
+=====================================Use only keys below+
| next screen: [→]  [Space]                              |
| last screen: [←]                                       |
+========================================================+
```

```
     M     -10-  -17-  -18-  -19-  -20-
 MONTHS   DIR  PRICE  STAY  OCC   CONV
 85 APR    68    82   1.88
 85 OCT    64    81   1.9
 84.apr    74    79   1.51
 84.oct    68    79   1.82
 85 NOV    62    77   1.85
 85 AUG    71    78   1.54
 84.aug    75    77   1.6
 85 DEC    71    77   1.65
 84.feb    70    76   1.71
 84.mar    75    75   1.65
 84.nov    64    75   1.66
 84.jan    70    74   1.65
 85 FEB    69    74   1.6
 85 MAR    73    74   1.58
 ____ Tables
```

```
+--------------------------------------------------+
|                       SORT                       |
|                                                  |
| The third and last command to be explained here is |
| Sort field.  With this command, you may specify  |
| an order for rearrangement of the records of that |
| particular field (ascending or descending).      |
|                                                  |
| In the example to the left of this box, field 17 |
| has been sorted already, in ascending order, and |
| all the other field records have also be moved   |
| in conjunction with the records of field 17.     |
| The original number of fields and records remains |
| unchanged.                                       |
+--------------------------------------------------+
```

```
+=====================================Use only keys below+
| next screen: [→]  [Space]                              |
| last screen: [←]                                       |
+========================================================+
```

SECOND NOTE ON KEYS TO BE USED

Please note that the key labeled [TAB] (in upper left corner of the keyboard, just below [EXIT] key) is to be used every time you indicate which particular field(s) or record(s) you chose in in response to a command request.

After you have indicated the desired field(s) or record(s), press the [ENTER] key, if necessary, as indicated inside the allowed keys box area. Failure to use the [TAB] key is equivalent to an [EXIT] request, i.e., to cancelling the requested operation.

Use only keys below

next screen: [→] [Space]
last screen: [←]

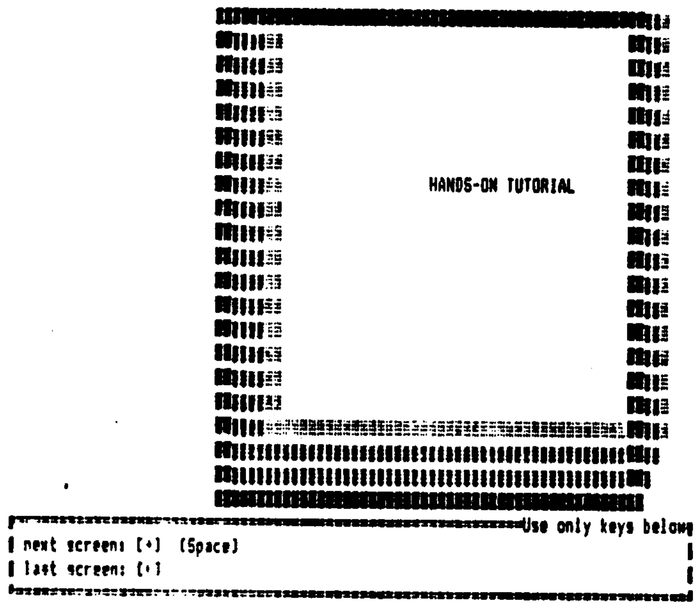
SECOND NOTE ON KEYS TO BE USED

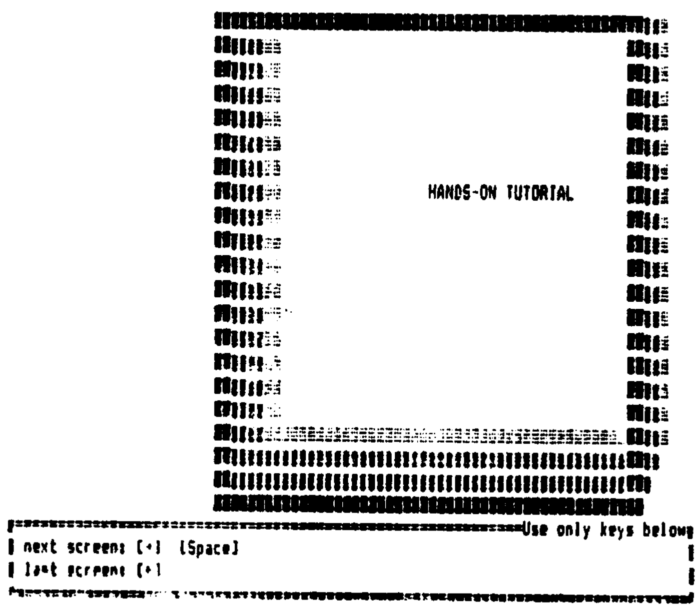
Please note that the key labeled [TAB] (in upper left corner of the keyboard, just below [EXIT] key) is to be used every time you indicate which particular field(s) or record(s) you chose in in response to a command request.

After you have indicated the desired field(s) or record(s), press the [ENTER] key, if necessary, as indicated inside the allowed keys box area. Failure to use the [TAB] key is equivalent to an [EXIT] request, i.e., to cancelling the requested operation.

Use only keys below

next screen: [→] [Space]
last screen: [←]


HANDS-ON TUTORIAL
Use only keys below
next screen: [→] (Space)
last screen: [←]


HANDS-ON TUTORIAL
Use only keys below
next screen: [→] (Space)
last screen: [←]

HANDS-ON TUTORIAL

The next section of this experiment consists of a quick, hands-on introduction to actual use of the system. You will be given a sheet with a list of commands to execute; simply follow the instructions on it. This last tutorial section should take from 5 to 10 minutes, at the most.

Please feel free to ask for help if you have trouble identifying a particular key to press, or operation to perform. The objective here, remember, is simply to familiarize you with the the operation of the system.

Use only keys below

next screen: (→) [Space]
last screen: (←)

HANDS-ON TUTORIAL

The next section of this experiment consists of a quick, hands-on introduction to actual use of the system. You will be given a sheet with a list of commands to execute; simply follow the instructions on it. This last tutorial section should take from 5 to 10 minutes, at the most.

Please feel free to ask for help if you have trouble identifying a particular key to press, or operation to perform. The objective here, remember, is simply to familiarize you with the the operation of the system.

Use only keys below

next screen: (→) [Space]
last screen: (←)

Information Tutorial (cont.)