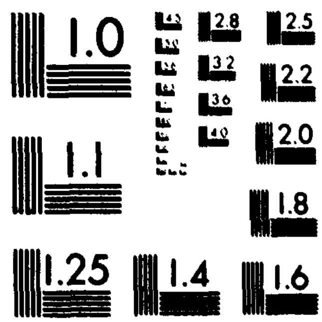
AD-A093 398   WHARTON SCHOOL PHILADELPHIA PA DEPT OF DECISION SCIENCES   F/G 9/2
PROVIDING AUTOMATIC GRAPHIC DISPLAYS THROUGH DEFAULTS.(U)
APR 80 S GANANMGARI, N I BADLER, H L MORGAN   N00014-75-C-0462

UNCLASSIFIED 80-03-09   NL

END
DATE FILMED
1 8
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A

AD A093398

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

14

1. REPORT NUMBER: 80-03-09	2. GOVT ACCESSION NO.: AD A093 398
3. RECIPIENT'S CATALOG NUMBER: 9	
4. TITLE (and Subtitle): PROVIDING AUTOMATIC GRAPHIC DISPLAYS THROUGH DEFAULTS.	5. TYPE OF REPORT & PERIOD COVERED: technical rept. 4/80-3/81 Apr 80-Mar 81
	6. PERFORMING ORG. REPORT NUMBER: 15
7. AUTHOR(s): Sakunthala Gananmgari, N.I. Badler, H.L. Morgan, Bonnie L. Webber	8. CONTRACT OR GRANT NUMBER(s): N00014-75-C-0462, MDA903-80-C-0093
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Decision Sciences, The Wharton School, U. of PA., Phila., PA 19104	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: Task NR049-272
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Darpa	12. REPORT DATE: Apr[redacted]80
	13. NUMBER OF PAGES: 7
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): 12 9	15. SECURITY CLASS. (of this report): unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

6 1 10 11

LEVEL II

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Distribution unlimited

DTIC ELECTE JAN 5 1981 D C

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

graphic displays; graphic representation; formats, graphic device characteristics; continuity, totality, cardinality; multiplicity, units;

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper addresses the problems in providing graphic displays automatically to serve a user naive with respect to computer graphic device. It identifies the properties of data that affect graphic representation and presents a formalism in which to view them. It also discusses and illustrates the selection of various graphic formats based on the data to be represented, its properties, and graphic device characteristics.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

408757 gm

80 12 30 004

PROVIDING AUTOMATIC GRAPHIC DISPLAYS THROUGH DEFAULTS *

Sakunthala Gnanamgari
N.I.Badler, H.L.Morgan, Bonnie L. Webber

Department of Computer and Information Sciences
Moore school of Electrical Engineering
University of Pennsylvania, Philadelphia, Pa. 19104, USA

Abstract

This paper addresses the problems in providing graphic displays automatically to serve a user naive with respect to computer graphic devices. It identifies the properties of data that affect graphic representation and presents a formalism in which to view them. It also discusses and illustrates the selection of various graphic formats based on the data to be represented, its properties, and graphic device characteristics.

I. Problem Statement

Broadly speaking, there are three phases of using computers: acquiring, processing and presenting information. As to the first two, many years of research and development have led to the availability of efficient ways of collecting and processing data. However, methods of presenting information are by and large limited to variations of tabular form. Reading a sequence of lines and understanding their import is a tedious job though, reminding people of the old proverb, "A picture is worth a thousand words." As a result, efforts are now being directed towards presenting such data graphically. Unfortunately, using graphic devices can be a complex process, requiring days or even weeks of training. Up to now, it has been almost impossible for a naive user to create a graphic display to view information.

* This research is partially supported by DARPA grant #MDA903-80-C-0093.

Our long range goal is to have an intelligent system helping users in the graphical display of data, performing the task of a graphic artist. Our objective, at present, is to facilitate automatic display of information by providing reasonable defaults for graphical representations and easy user modification of the resulting displays.

The major problem in developing such a system is that there is a gap between the way a user conceives of a graphic display and the way the machine does. For the user, it is a meaningful picture made up of certain particular pieces; for the machine, it is the sequence of operations needed to create such a display. A second problem is that a user will not think to make explicit what s/he does not care about or what s/he believes the system already knows or is able to infer. What is needed is a graphic expert system that, on the one hand, is at an appropriate conceptual level for user to state things that s/he cares about, but, on the other, provides appropriate defaults to take care of everything else.

Research has proven that graphic presentation of information is better than tabular form. Tabular form merely presents raw data without interpretation [Gene Zelazny, 1972], whereas pictorial form conveys the relationship between the data items.

To illustrate this contrast between tabular and graphic presentation, consider the following example. Using the Harvest system [Harvest, 1979], a database query system, a naive user can type in

WHERE YEAR = 1980 DISPLAY BUDGET

and get a formatted output as shown below:

BUDGET FOR 1980

	ITEMS	AMOUNT*
1.	SALARIES	35
2.	TRAVEL	10
3.	EQUIPMENT	25
4.	MAINTENANCE	18
5.	MISCELLANIOUS	12
	TOTAL	100

* Thousands of dollars

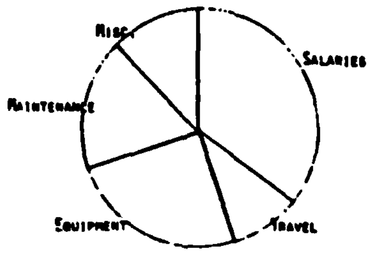
For tabular form output, systems such as HARVEST can provide default formats. This relieves a naive user of the need to provide detailed format specifications, a burdensome task especially when the user may not care more about the format than it be easy to read.

However, it is not currently possible to request a graphic display in the same easy terms - i.e., to type

WHERE YEAR = 1980 DISPLAY BUDGET GRAPHICALLY

and get a graphic display as shown here:

BUDGET FOR 1980



No existing system provides the default graphical formats needed to provide such a service.

There are some "high level" software packages commercially available, such as PLOT-10 and DISPLA [ISSCO], that allow an applications programmer to use a graphic device at a programming language level. Interactive systems like Tell-a-Graf [ISSCO] requires users to enter data and specify their preferences completely. But none of these systems can provide default displays for either completely or incompletely specified choices. What is needed is, highly automated graphics systems to meet the needs of naive users who either do not want to specify any preferences about the graphic display or give incomplete specifications.

This paper discusses appropriate defaults for those aspects of a display the user has failed to specify and how those defaults depend on three factors: the data to be displayed, the device on which it is to be displayed and the users it is displayed for. Two different types of defaults are considered: defaults affecting the choice of graph through which to display the data and defaults affecting the choice of "attributes" for that graph, such as color, size, orientation, order and other factors. These defaults are used to provide a naive user with the ability to see his or her numeric data (which would otherwise be presented as a table of numbers) in the form of a pie chart, bar graph or trends graph.

II. Definitions

Before introducing the system and basic assumptions for the system, we shall define the concepts we will be using:

1. CONTINUITY: a boolean value that represents whether or not the members of an ordered set represent an interval of a continuum with respect to the given ordering. Example: A set of days, {Sunday, Monday, Tuesday, Wednesday, Thursday, Friday, Saturday} could be defined to represent a WEEK, an interval of time, and have the property continuity, while {Sunday, Tuesday, Saturday} may not, and {Sunday, Tuesday, Friday, Wednesday, Monday, Saturday, Thursday} may not.

2. TOTALITY: is a boolean value that represents whether or not the members of a set represent ALL the component parts of an object or an abstract concept. Example: the set of items {Salaries, Travel, Equipment, Maintenance, Miscellaneous} could be defined to represent the parts of which BUDGET is composed, and have the property totality. The subset {Salaries, Travel, Maintenance} would not have totality.

3. CARDINALITY: is the number of elements in a set. Example: the cardinality of range the set of days is 7.

4. MULTIPLICITY: is the number of values assigned to each element in a domain set by a mapping. Example: the mapping "square root" from real numbers into complex numbers has the multiplicity of 2.

5. UNITS: is the set of labels specifying the unit of measurement associated with each numerical value. Example: Thousands of dollars, Hundreds of tons, etc.

One of the factors upon which effective automatic data display depends comprises particular characteristics of the data itself. By abstracting out these characteristics, one can form a well defined bijection mapping that can help one to understand the complex phenomenon of data and its manipulations.

Let this abstract form of data be represented by the word title, a mapping from the domain set of labels into the range set of quantities. That is,

$$\text{TABLE: } \{l_1, l_2, \ldots, l_m\} \longrightarrow \{(q_{11}, q_{12}, \ldots, q_{1n}),\ (q_{21}, q_{22}, \ldots, q_{2n}), \ldots, (q_{m1}, q_{m2}, \ldots, q_{mn})\}$$

where, for every $i=1$ to m, l_i is the ith element in the domain set and for every $i=1$ to m and $j=1$ to n, q_{ij} is the jth component of the ith tuple in the range set. Each column also has an entity called units and another ENTITY called column-label.

In other words, the data in the range set is a matrix of size m rows and n columns.

The cardinality of row-labels and multiplicity of the mapping can be derived from input data. However, two additional properties of this mapping, that are necessary to select a display format are not directly derivable from the input data itself. These are:
(i) whether elements of either row-labels or column-labels form component parts of some whole with respect to the quantities represented by each member of the column-labels and row-labels respectively: that is, whether either set has totality.

(ii) whether elements of either row-labels or column-labels denote to a continuum with respect to the quantities represented by each member of the column-labels and row-labels respectively: that is, whether either set has continuity. For example,

NET INCOME PER SHARE

	COMPANY-1	COMPANY-2
1972	0.86	0.60
1973	1.01	0.90
1974	1.22	1.15
1975	1.35	1.45
1976	1.60	1.80
1977	1.93	2.24
1978	2.44	1.90
1979	2.70	2.01

In this example, the row-labels are 1972, 1973, 1974, 1975, 1976, 1977, 1978 and 1979 and the column-labels are COMPANY-1 and COMPANY-2. The continuity of row-labels could be true or false with respect to each column-label. If the comaparison of incomes for two companies over the period of time is prefered, then the continuity of row-labels would be {true, true} with respect to each of the column-labels. If an absolute comparison of incomes is prefered then the continuity of row-labels would be {false,false}. The totality of row-labels could be true or false. If a relative comparison of each year's income with respect to the total income of each company is prefered, then the totality of row-labels would be {true,true}; otherwise, it would be {false,false}. Similarly, the continuity and totality could be defined for column-labels. The cardinality of row-labels is 8. The multiplicity of the mapping is 2. The units are dollars for each column-label.

III. Examples

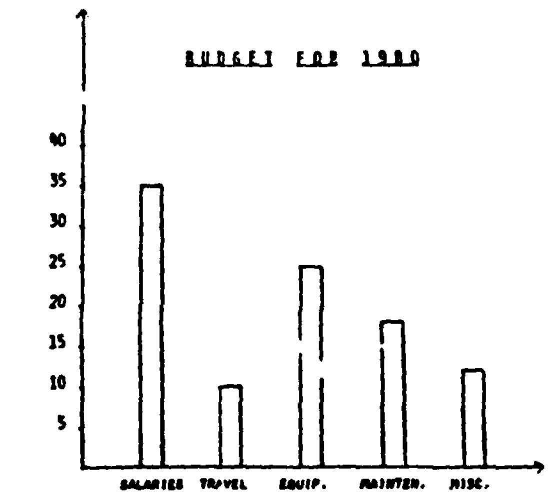
Having defined the concepts that we will be using, to demonstrate how the above mentioned ideas can be used to provide a graphic display, consider the BUDGET FOR 1980 example given earlier. Here the mapping is BUDGET FOR 1980, the set of row-labels is {Salaries, Travel, Equipment, Maintenance, Miscellaneous}, the range set of quantities is {(35), (10), (25), (18), (12)}, the set of column-labels is {Amount} and the set of units is {Thousands of dollars}. Let the totality and continuity of row-labels be {True} and {False} respectively. Given this information and no preferences on the user's part, the system's task is to observe the data and its characteristics, decide what type of graphic format is both suitable and feasible with respect to the graphic device that is available, decide its attributes and then display the picture. (Although it should also allow the user to modify the resulting display, this aspect of the user interface will not be discussed.) For this example, the system selects a pie chart representation to express the totality of the row-labels. This pie chart representation is an appropriate choice as confirmed in the literature:

"Because a circle gives such a clear impression of being a total, a pie chart is ideally suited for the one purpose it serves - showing the relative sizes of the components of some whole." - [Zelazny, 1972]

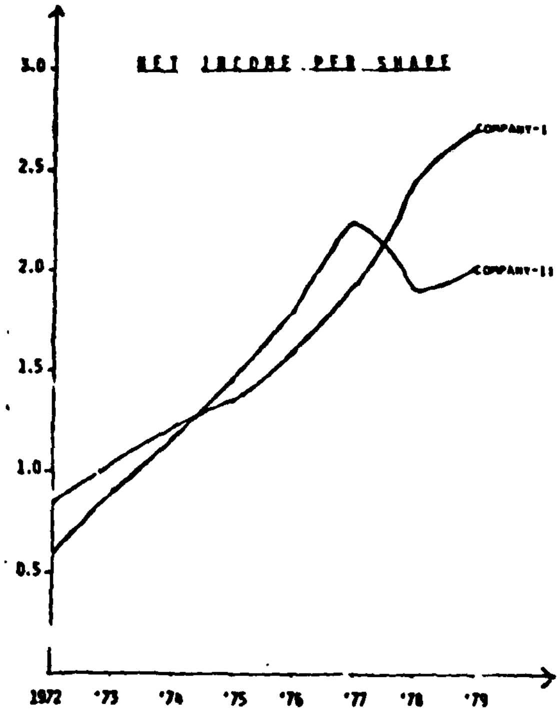
"...the separation of a whole amount in terms of its component quantities. In the graphic figure, a circular form can be used to represent a whole amount, and can be divided into segments which represent proportional quantities, or percentages, of the whole." - [Bowman, 1968].

As we noted above, the user has not stated any preferences regarding the display. This being the case, the choice of whether or not to color the different segments of the pie (and if so, what colors) is left as another set of defaults. These choices/defaults depend partially on device capabilities but also on whether colors would be an effective way of communicating information to the user. For a device such as the printed page, the choice of colors is black and white.

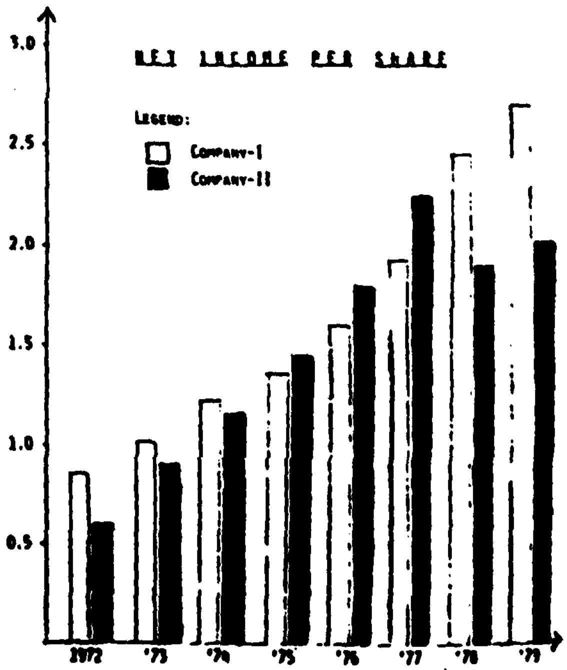
In this example, suppose the totality is {False}, the system whould have opted for a bar chart. The reasons for this option are: (i) continuity being false, a line graph is not selected, (ii) totality being false, a pie representation is not selected, and (iii) "In a graphic figure, quantity can be shown in comparative relation to other quantities, through the extension of abstract parallel bar forms." - Bowman [1968]. The resulting figure is shown below:



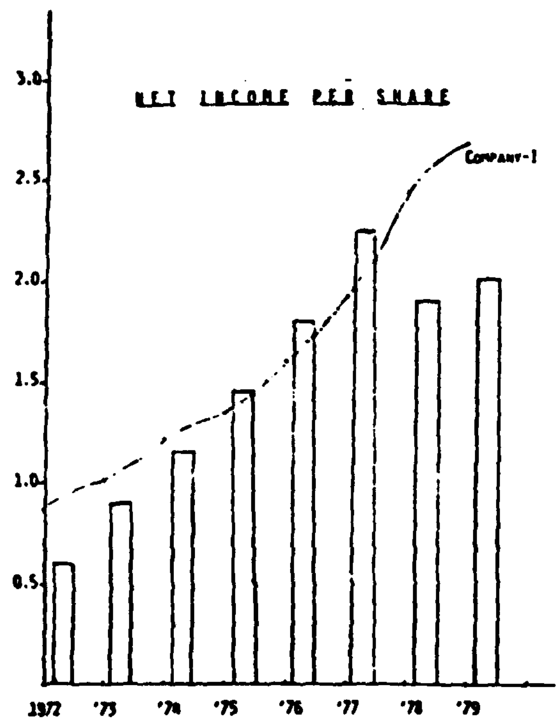
As another example consider the mapping INCOME PER SHARE. We will look at five cases. case 1. If the continuity of row-labels is {true, true}, the totality of row-labels is {false, false} and the units is {dollars, dollars} the graphic format selected would be a LINE graph. That is,

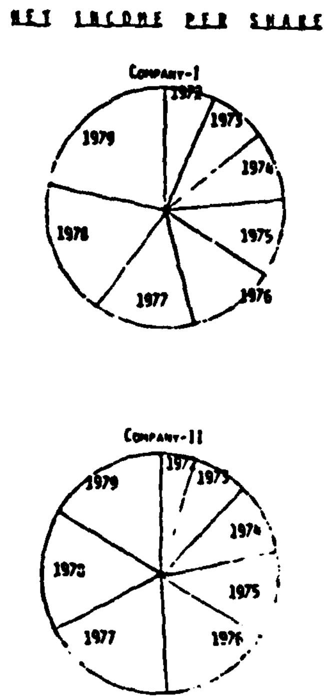


Case 2. If the continuity of row-labels is {false, false}, the selected graphic format would look



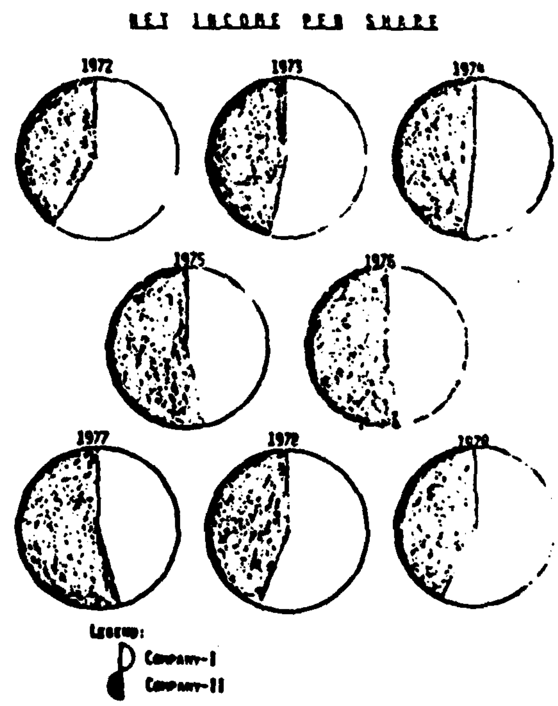
Case 3. If the continuity of row-labels is {true, false}, the graphic format selected would be





Case 4. If the continuity and totality of row-labels are {false,false} and {true, true} respectively, then the data would be presented in the form on the top right.

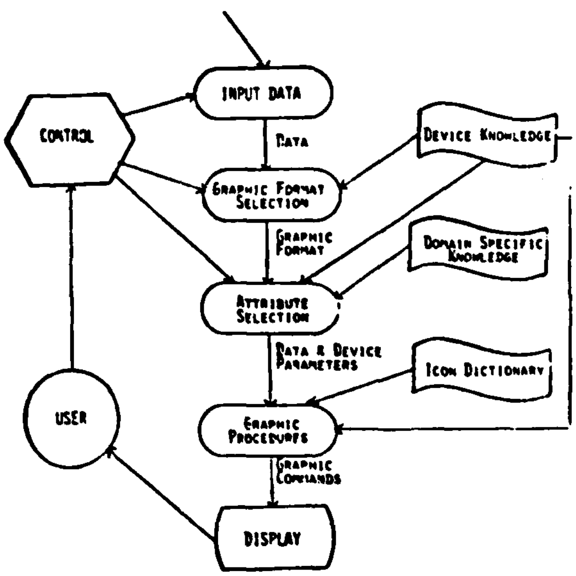
Case 5. If the continuity and totality of row-labels are {false, false} and {false, false}; and the totality of column-labels is {true, true, true, true, true, true, true, true}, then the graphic format on the bottom right represents the input data.



IV. System Overview

The overview of the proposed system currently under development is given in the following figure.

OVERVIEW OF THE PROPOSED SYSTEM



We are making the following three assumptions with respect to this system design:

(i) DATA is expected from an existing database. The system expects a table of information which has both row-labels and column-labels. Either of these sets may be tagged with the properties of continuity and/or totality. These two properties of the mapping are expected as input to the system along with the data mapping and information on units of measurement for the quantities in the range set.

(ii) DEVICE is expected to have a set of routines for drawing and erasing points, lines and characters, and for setting colors or grey values.

(iii) USER is expected to be able to type in the request for a graphic display.

The information from a database enters the system at the node INPUT DATA. The data is passed to the next node FORMAT SELECTION.

Depending on the characteristics of input data such as multiplicity, cardinality, units, continuity and totality; and of graphic device such as device type, spatial and intensity or color resolution; a default graphic format (such as a pie chart) will be selected to display information. These rules of selecting a particular display format are defined after consulting Bertin [1973], Bowman [1968] and Gene Zelazny [1972 and 1980] and studying various graphic representations.

Once the appropriate graphic format has been selected, the format and the information to be displayed are passed to the next node, the ATTRIBUTE SELECTION. This state consults the device knowledge and domain specific knowledge to determine the attributes of the display such as color and icons. The output of this state consists of data and device parameters.

Depending upon these parameters, the next node, GRAPHIC PROCEDURES, generate the graphic commands to a particular device that realizes the display.

DISPLAY is the actual display of information, the final output of the system, in the graphic format.

The graphic display is obtained by simply requesting the system to present tabular information graphically. If the display is not satisfactory to the user, it may be modified. The modifications are provided at three levels: (i) input data could be modified by selecting or grouping the row-labels to be displayed, (ii) the properties such as totality or continuity could be changed thereby changing the format of the display and (iii) attributes of display could be changed.

V. Summary

In summary, this paper has discussed the system [Gnanamgari, 1980] which we have designed to provide appropriate defaults for those aspects of the presentation of a user's data that s/he either does not care about or assumes the system would "obviously" infer. The underlying structures of input data have been studied and abstracted and relavant properties of data have been recognized. A reasonably large set of graphic formats have been defined for presenting data. Currently we are working on, knowledge representation issues of the system.

BIBLIOGRAPHY

1. Bertin, J.: "La Graphique et Le Traitement Graphique de L'Information",Flammarion, Paris, 1977.

2. Bowman, W.J.: "Graphic Communication", John Wiley, 1968.

3. DISPLA: a software product of Integrated Software Systems Corporation, San Diego, Ca. 92121.

4. Gnanamgari, S.: "Automatic Generation and Presentation of Graphic Information Displays", Ph.D. thesis, Department of Computer and Information Sciences, University of Pennsylvania, 1980 (forthcoming).

5. HARVEST Reference Manual, International Data Base Systems, Philadelphia, Pennsylvania, 1979.

6. TELLAGRAF: a software product of Integrated Software Systems Corporation, San Diego, Ca. 92121.

7. Zelazny, G.: "Choosing and Using Charts", copyright 1972, Gene Zelazny, McKinsey and Company Inc., New York.

8. Zelazny, G.: Director, Visual Communications, McKinsey and Company Inc., New York, Personal Communication, 1980.