AUDITORY AND VISUAL ANALYSES

Chaired by Doris Aaronson, *New York University*

The empirical investigation of factors affecting graphical visualization

RUSSELL W. JONES

University of Melbourne, Parkville, Victoria, Australia

and

ISMAEL E. CARERAS

Boston College, Boston, Massachusetts

An estimated 2.2 trillion graphs were published during 1994. These graphs used a multitude of different visual techniques in an effort to communicate information to a reader. Despite the many important applications for which graphs are used and the frequency of their use, little is known about what techniques make a graph most effective at communicating information to a reader. This paucity of knowledge stems largely from a dearth of empirical investigations pertaining to graphs. A major problem encountered by researchers who wish to undertake sound empirical research into the effectiveness of graphical presentations is the shortage of suitable tools. This article presents a computer software tool, *Graphics Visualization*, which offers the researcher an inexpensive yet accurate and reliable method of performing empirical research into factors that affect the accuracy and speed by which a reader can interpret data encoded within a graph.

Tufte (1983), one of the foremost experts on graphs and graphical presentation, estimated that 900 billion statistical graphs were printed during 1983. This figure has increased substantially over the past 12 years, and, indeed, Feeney (1994) has presented evidence for a 240% increase in the number of graphs printed between the early 1980s and 1994. This means that in 1994 an estimated 2.2 trillion graphs were printed (note that this figure does not include the enormous number of graphs produced for display on television or computer terminals). The makers of these graphs have used a multitude of different visual techniques in an effort to communicate information to a reader. Despite the enormous number of graphs produced each year, their increasingly widespread use throughout all facets of our society (Feeney, 1994), and their involvement in the communication of information within many fields (including psychology, the media, private industry, the military, government, and educational organizations), little is known about what visual techniques make a graph most effective at communicating information to a reader. This is largely because of a dearth of empirical investigations into the comparative effectiveness of graphs and graphical techniques. Recognition of this

paucity of empirical research is not new. As early as 1874, Mayr drew attention to the need for more empirical research. More recently, Morley and Adams (1991); Dempsey, Fisher, and Marousky (1994); Lewandowsky and Spence (1989); Tufte (1983, 1990); Cleveland (1985); Cleveland and McGill (1986); Kosslyn (1985); Kruskal (1982); and Lohse, Walker, Biolsi, and Reuter (1991) are among the many to call for more empirical research.

It is worth noting that although there is a shortage of empirical research, there is no shortage of published opinion as to what makes a graph effective at communicating information to a reader (see, e.g., American Statistical Association, 1982; Cleveland, 1984, 1985; Cleveland & McGill, 1984; Cox, 1978; Ehrenberg, 1977; Morley & Adams, 1991; Schmid, 1983; Tufte, 1983). Yet because most of these opinions lack the foundation of solid empirical experimental research, the effects of the techniques these authors recommend are largely unknown. Again, this serves to highlight the need for sound experimental research.

One of the substantial problems encountered by researchers who wish to perform sound empirical research into the effectiveness of graphical displays is the shortage of suitable tools to enable them to investigate the effectiveness of graphs at communicating information to a reader. Indeed, much of the research hitherto performed on the effectiveness of graphical displays would have been improved if researchers had had access to an accurate and sensitive tool for displaying graphs and recording the effectiveness by which graphs communi-

We would like to thank Walt Haney for his suggestions during the construction of this stack. Correspondence should be addressed to R. W. Jones, Department of Science and Mathematics Education, University of Melbourne, Parkville, Victoria 3052, Australia (e-mail: r.jones@edfac.unimelb.edu.au).

cate information (see, e.g., Croxton & Stryker, 1927; Culbertson & Powers, 1959; Eells, 1926; Morley & Adams, 1991; Peterson & Schramm, 1955).

Statement of Problem

Two important questions for those concerned with the publication of graphs for the purposes of communicating information are as follows: (1) How accurately can a graph convey information to a reader? and (2) How fast can a graph convey information to a reader? The purpose of this article is to present a computer software tool that offers the researcher an inexpensive yet accurate and reliable method of performing empirical research into factors affecting the accuracy and speed by which a reader can interpret data encoded within a graph. The tool can be used to compare the effectiveness of different graphical techniques, different graphing formats, variations in symbols and shading within a graph, graph size, display intensity, the effects of color, and many other aspects pertinent to those interested in graphical visualization. A computerized tool such as this offers many advantages over other methods of collecting similar data. These advantages include the consistent objectivity and accuracy with which a computer can measure and record a subject's responses and the fact that a computer (unlike a person) is not prone to fatigue.

The Software Tool

The software tool is written in HyperCard (Apple Computer, 1989) and will run on any Apple Macintosh computer with one or more megabytes of memory and, preferably, an 800K disk drive and a hard disk, although two 800K disk drives may be used. The tool requires an extended keyboard with an Enter key. Applications written using HyperCard are called "stacks." These stacks are composed of a number "cards." A user browses through cards within a stack in a sequential order. The user views each card as a discrete visual display that can contain a variety of stimuli, including text, artwork, and symbols.

The software tool described in this article, Graphics Visualization, was designed to collect both survey and experimental data. It offers the researcher a flexible and inexpensive method of presenting subjects with a series of cards, and each card is used to visually present a subject with information and/or to elicit a response from a subject. An important characteristic of the tool is that it is designed for experimental subjects to be able to work through it with minimal interaction with, and supervision from, an investigator. In its present form, Graphics Visualization contains 89 cards. Researchers may wish to add or delete cards depending on their specific experimental design. The remainder of this section describes the software tool and explains how subjects work through the stack.

The stack begins with a series of four cards that introduce the subject to the stack. The first card gives the reader three pieces of information: (1) it gives a brief explanation about the purpose of the "program,"¹ stating that the stack is intended to measure how people read information from graphs, (2) it informs subjects that they will be asked a series of questions about themselves, and (3) it provides instructions on how to use the mouse to progress through the stack.² Following these instructions the reader clicks on an

arrow in the lower right portion of the card and the second card in the stack is displayed. The second card gives additional instructions regarding keyboard use. The third card gives a visual display of the types of graphs within the stack and also orients subjects to the format of the graphs contained within the stack. For example, if the stack is to be used to investigate the properties of bar charts, line charts, and scatterplots, this card would display examples of each of these graph formats. Researchers who wish to investigate questions relating to other graphing formats can easily substitute a card displaying other graphs. The fourth and final introductory card prepares subjects for the next three sections of the stack and simply states:

After this introduction, this program has three parts:

1. You will be given several questions about yourself.
2. You will be given six "practice" questions, to make sure you are familiar with how this program operates.
3. You will be asked to "read" information from a series of 60 graphs.

HyperCard offers the opportunity for researchers to readily make editorial changes on existing cards; to add cards containing additional information, instructions, or questions; or to delete cards considered unnecessary or inappropriate to a specific investigation. Hence the stack may be tailored to meet the needs of any researcher investigating a wide range of research questions pertaining to graphs.

Demographic and background information. Data pertaining to subject demographics is an important component of many investigations focusing on human subjects. Hence, following the initial explanatory introduction, the stack presents a number of cards that ask subjects a series of demographic questions such as, "What is your gender?" and "What is your age?" It is likely that the educational experiences of subjects may influence their ability to read and interpret graphs; hence the stack includes a series of cards that also ask questions pertinent to the educational background of a subject. Such questions include, "How many courses have you taken in the graphic arts?" "How many courses have you taken in the visual arts?" and "How many courses have you taken in statistics?" The stack does not permit the subject to progress to the next card until each question is answered. If a subject attempts to ignore a question, the following message is displayed: "Please select your answer to the question before clicking on the arrow to continue." All data are recorded on the final card in the stack. Subjects do not have access to this card, but researchers can quickly access this card by simultaneously pressing the Command and L keys.

Practice tasks. Once subjects have answered the demographic and background questions, a series of four cards is displayed. These cards explain the experimental tasks to the subject and provide the subject with instructions for completing the tasks. The content and number of these cards may be easily changed using the editing features available in HyperCard. Figure 1 shows the first of these four cards as they presently appear in the stack. The next three cards give the reader the following instructions:

For each graph you will be asked to estimate the numerical value of one of the letters on the bottom axis. For example, in the graph below you may be asked, "What is the value of C?" or "What is the value of F?"³

Reading Graphs Part 2

You will be presented a series of graphs like the one below. Notice the 6 letters A to F on the bottom axis and the numbers 0 to 100 on the left axis.

Figure 1. The first of the cards explaining an experimental task within a stack.

Once you have estimated the value of C, press Enter. The graph will then disappear. Enter the value of your estimate into the computer using the numeric keys. Make sure you estimate the value to the nearest whole number and that you do not round to the nearest 5 or 10. That is, if you estimate a value at 36 then enter "36" and do *not* round to 35 or 40.

You will now be presented with six practice questions.

1. Press the Enter key on the numeric pad when you are ready to answer the question.
2. Type in your answer using the numeric keypad, and press Enter again to record your response and be shown the next graph.
3. You will not need to use the mouse to answer the questions.

Prior to commencing the experimental portion of the stack it is useful for subjects to be given a series of practice tasks to familiarize them with the nature of the tasks they are to perform. Hence, each of the next six cards presents a practice task. These practice tasks serve a dual purpose. First, they ensure that subjects are familiar with the operation of the stack, and second, they expose subjects to all the tasks with which they will be presented in the experimental portion of the stack. Additional or alternative practice tasks may be included simply by adding or editing the cards within the stack. The exact nature of the practice tasks will depend on the tasks placed in the experimental component of the stack.

Experimental component. The experimental component of the stack contains a series of 60 tasks similar to those demonstrated during the practice tasks. Prior to the presentation of the first experimental task, the stack is programmed to present the subject with a card containing a warning that the experimental portion of the stack is about to begin. For this portion of the stack, the computer is instructed to randomly select a card from among a pool of available cards (up to 60 at this time, although more can be added), each containing graphical stimuli prepared by the researcher. Once the computer has selected the card, the stimuli contained on the card are presented to the subject on the computer screen. Once the subject has responded to the stimulus, the computer eliminates the card from the pool of available stimuli and randomly selects another card.

A large variety of experimental tasks may be pasted onto these cards. Typically, the sorts of tasks that have been used

with the stack thus far involve a display of three pieces of information on the screen (Figure 2, top panel). The central portion of the screen displays a graphical stimulus. The top portion of the screen displays a simple question concerning the graph, such as, "What is the value of [a specified point on the graph]?" The lower portion of the screen prompts the subject to press the Enter key once he/she has decided on an answer. The example in Figure 2 is an excerpt from an investigation designed to determine the influence of bar shading, within bar charts, on the speed and accuracy with which a subject can estimate the value of data encoded in a graph.

The time interval is calculated between the moment the graph is first displayed on the computer screen and the moment the subject strikes the Enter key, indicating that he/she has decided on an answer. This time interval is automatically recorded on the last card in the stack. To ensure that the subject decides on an answer to the task *prior* to pressing the Enter key (and therefore *prior* to the computer's calculation of elapsed time), the graph disappears when the Enter key is struck (Figure 2, bottom panel). This ensures that subjects complete their study of the graphical display before they press the Enter key and therefore before the elapsed time is calculated.

Researchers may design tasks that require subjects to estimate a numerical value (such as the task shown in Figure 2). If so, the computer is programmed to automatically compare the subject's numerical answer with the real value (which can be programmed into the stack prior to giving the stack to a subject), calculate the subject's estimation error (i.e., calculate the accuracy of the subject's answer), and to record both the estimate and the error on the last card in the stack.

After the graph has disappeared, the cursor automatically moves to the box in the central lower part of the screen and the subject types his/her answer into the box. If a subject fails to type an appropriate response into the box, the following message is displayed on the screen: "Please type your answer

What is the value of E?

Press ENTER to
type in your answer

What is the value of E?

Press ENTER to
type in your answer

Figure 2. To ensure that subjects complete their study of the graph prior to calculation of the time interval, the graph shown in the card (top) disappears once the Enter key is struck. Subjects type their answers into the box at the bottom of the card (bottom).

into the box." Once the subject records an answer, the subject again presses the Enter key, whereupon the computer records the subject's answer on the last card in the stack and randomly selects another task for display on the screen. This process continues until the subject has worked through all of the experimental tasks.

To counteract the effects of fatigue, which will affect subjects during any experiment requiring them to perform a number of similar tasks for an extended period of time, the stack is programmed to periodically pause and display the following message: "Please take a few moments to rest and exercise your eyes, neck, wrists, hands, and fingers. Get up, stretch and walk around the room. When you are rested, return to your seat and press the Enter key to continue." A card with these written instructions is programmed to be displayed each time the subject has completed 20 consecutive experimental tasks.

Figures 3 and 4 show examples of cards that were part of a series of investigations on the effects of tick placement (Figure 3) and legend placement (Figure 4). Note that both the graphical stimulus and the question may be changed to suit the needs of different researchers as they investigate different research questions.⁴ For example, the questions in Figures 3 and 4 ask, "How many points are greater than ____?" and the question in Figure 2 asks, "What is the value of ____?"

It is also possible for a researcher to restrict the numerical values the computer will accept as answers to the questions. For example, because there are only six points plotted in Figure 8, the researcher may wish to restrict answers to plausible values. In this example, the stack was programmed to refuse to accept answers greater than 6.

Additional opportunity to gather data. Once a subject has completed an experiment that has exposed him/her to a variety of graphical techniques, formats, or displays, the stack provides the researcher with an opportunity to gather additional related data from each subject. This opportunity occurs while the graphs are fresh in the subject's mind. This useful feature is placed immediately after the experimental component of the stack. A series of cards is displayed that seek to find out more about the subject's impressions, beliefs, and preferences for graphical displays. The content of these cards may be tailored to the particular investigation being performed by the researcher. So far the stack has been used to ask subjects to record their preference for such graph-

How many points are greater than 35?

Figure 3. Example of a card used to study the effects of tick placement.

How many points are greater than 50?

Figure 4. Example of a card used to study the effects of legend placement.

ical techniques as the placement of graph labels, legends, or grid lines; or for preference for particular graphical symbols (e.g., dots of various sizes, geometric shapes, and so forth); or for particular intervals of scale; or for preference for particular graph formats (e.g., bar chart, line charts, or scatterplots). The scope of these questions is limited only by the imagination of a researcher. The stack can be readily adapted to ask additional or fewer questions with similar or entirely different content. Hence, the content and quantity of questions may be varied to reflect the needs and interests of a researcher. The sorts of questions that are particularly appropriate to ask at this time are those that ask for information that the researcher could not request before the experiment commenced in case they should influence the subject's responses to the experimental component of the stack.

The stack does not permit a subject to continue working through cards in the stack until each of these questions is answered. If a subject attempts to ignore a question, the following message is displayed: "Please answer the question before clicking on the arrow to continue." The answers to all questions are recorded on the same card, which records the answers to the demographic questions and the responses to the experimental tasks.

Once all data have been gathered and recorded, the stack presents a final card with the following message: "Thank you for your participation in this project. This program is now complete." This card also displays a Quit button. When a subject clicks on this button, the stack closes and the subject is returned to the Finder.

Conclusions

The applications of this tool are limited only by the researcher's imagination. For example, the tool can be used to compare different graphical techniques, different graphing formats, variations in symbols and shading within a graph, graph size, display intensity, the effects of color, and many other pertinent graphical features. Typical examples of the sorts of tasks for which the tool can be used involve asking subjects to estimate (1) the value of a discrete piece of data within a graph, (2) any data trends within a graph, and (3) the number of data points within a specific area of a graphical display. The choice of task depends on the research question(s) under investigation. For example, if a researcher wishes to determine whether bar charts or line charts enable

subjects to more accurately and quickly read graphs, then the tasks can involve comparisons of bar charts and line charts. Similarly, if the investigator wishes to determine whether different colored lines within a line chart affect the accuracy or speed by which a subject can read the chart, the tasks can involve a comparison of line charts encoding data within different colored lines. Any number of graphical stimuli can be created in many graphics packages and pasted onto cards for inclusion in the tool. This offers the researcher a tremendous amount of freedom in the scope of research questions. So far the stack has been administered to more than 300 subjects and used to investigate the effects of (1) framing a graph, (2) tick placement, (3) extraneous data placement, and (4) encoding information within bar charts, line charts, or scatterplots on the estimation accuracy and estimation speed with which a graph communicates information to a reader.

Potentially, properly designed studies that make use of Graphics Visualization are applicable to all facets of our society where graphs are used to communicate information. The future will see an even greater number of graphs being used to communicate information to billions of people. As the number of graphs in our society increases, researchers will be able to use this tool to help identify those techniques that best facilitate the accurate and rapid communication of information encoded within graphical displays.

Availability

To obtain a copy of Graphics Visualization, write to R. W. Jones, DSME, The University of Melbourne, Parkville, Victoria 3052, Australia. Please include a check for \$5 to cover costs of the disk, postage, and handling.

REFERENCES

- AMERICAN STATISTICAL ASSOCIATION (1982). Preparation of manuscripts for ASA journals. *Journal of the American Statistical Association*, **77**, 226-229.
- APPLE COMPUTER (1989). *HyperCard* [Computer program]. Cupertino, CA: Author.
- CLEVELAND, W. S. (1984). Graphs in scientific publications. *American Statistician*, **38**, 261-269.
- CLEVELAND, W. S. (1985). *The elements of graphing data*. Monterey, CA: Wadsworth.
- CLEVELAND, W. S., & MCGILL, R. (1984). The many faces of a scatterplot. *Journal of the American Statistical Association*, **79**, 807-822.
- CLEVELAND, W. S., & MCGILL, R. (1986). An experiment in graphical perception. *International Journal of Man-Machine Studies*, **25**, 491-500.
- COX, D. R. (1978). Some remarks on the role in statistics of graphical methods. *Applied Statistics*, **27**, 4-9.
- CROXTON, F. E., & STRYKER, R. E. (1927). Bar charts versus circle diagrams. *Journal of the American Statistical Association*, **22**, 473-482.
- CULBERTSON, H. M., & POWERS, R. D. (1959). A study of graph comprehension difficulties. *Audio-Visual Communication Review*, **7**, 97-100.
- DEMPSEY, J. V., FISHER, S., & MAROUSKY, R. (1994). *Preference and use of information in instructional graphs*. Paper presented at the annual conference of the National Consortium for Instruction and Cognition, New Orleans.
- ELLIS, W. C. (1926). The relative merits of circles and bars for representing component parts. *Journal of the American Statistical Association*, **21**, 119-132.
- EHRENBERG, A. S. C. (1977). Rudiments of numeracy. *Journal of the Royal Statistical Society: Series A*, **140**, 277-297.
- FEENEY, M. (1994, March 11). Beyond the voodoo stick: If a picture is worth a thousand words, how many words is an information graphic worth? *Boston Globe Magazine*, pp. 10-18.
- KOSSLYN, S. M. (1985). Graphics and human information processing. *Journal of the American Statistical Association*, **80**, 499-512.
- KRUSKAL, W. H. (1982). Criteria for judging statistical graphics. *Utilitas Mathematica*, **21B**, 283-310.
- LEWANDOWSKY, S., & SPENCE, I. (1989). The perception of statistical graphs. *Sociological Methods & Research*, **18**, 200-241.
- LOHSE, G., WALKER, N., BIOLSI, K., & REUTER, H. (1991). Classifying graphical information. *Behaviour & Information Technology*, **10**, 419-436.
- MAYR, G. (1874). *Gutachten über die Anwendung der graphischen und geographischen Methoden in der Statistik*. Munich: J. Gotteswinter & Mossel.
- MORLEY, S., & ADAMS, M. (1991). Graphical analysis of single case time series data. *British Journal of Clinical Psychology*, **30**, 97-115.
- PETERSON, L. V., & SCHRAMM, W. (1955). How accurately are different kinds of graph read? *Audio-Visual Communication Review*, **2**, 178-189.
- SCHMID, C. F. (1983). *Statistical graphics*. New York: Wiley.
- TUFTE, E. R. (1983). *The visual display of quantitative information*. Cheshire, CT: Graphics Press.
- TUFTE, E. R. (1990). *Envisioning information*. Cheshire, CT: Graphics Press.

NOTES

1. Because subjects tend to be familiar with the concept of a computer program yet are often confused by the word *stack*, the word "program" was used when describing the stack to subjects. No subjects were confused by this interchange of terminology.
2. An early version of the stack included a tutorial designed to show subjects how to use a mouse and to familiarize subjects with the keyboard. However, because the first 300 subjects who used the stack were familiar with both the mouse and keyboard and because all of these subjects chose to skip the tutorial, this tutorial was eliminated from subsequent versions.
3. Although the specific question associated with each graph can be changed (e.g., from "What is the value of . . . ?" to "How many points are greater than . . . ?"), it is best if the researcher resists the urge to mix questions requiring different tasks in the same stack. This is because the stack is designed as a tool to investigate the comparative differences in speed and accuracy by which subjects can read different graphs. By including a variety of different tasks within a specific stack, the researcher would introduce a confounding variable into the experiment that would affect a subject's cognitive processing as he/she works through the stack. Hence it is best to keep these questions simple and consistent through each stack so as to minimize the need for the reading of text and subsequent cognitive processing. Any such textual reading and cognitive processing would have the effect of adding "noise" to the experimental results.
4. Note that the instructional text and practice tasks may need to be changed to ensure that they are consistent with the particular experimental task under investigation.

(Manuscript received November 13, 1995;
revision accepted for publication February 2, 1996.)