Display of Quantitative Information: Are Grables better than Plain Graphs or Tables?

Jessica K. Hink
North Carolina State University
Raleigh, NC 27695-7801

Michael S.Wogalter
North Carolina State University
Raleigh, NC 27695-7801

Jason K. Eustace
American University
Washington, DC 20016-8019

ABSTRACT

Previous research is equivocal on the most efficient, effective methods for displaying quantitative information in tables and graphs. Guidelines suggest different display types are more appropriate for certain purposes but not others. However, there is little empirical evidence to support the recommendations. This study examines several methods of displaying quantitative information (e.g., line graphs, bar charts, tables) factorially crossed with different kinds of data extraction questions (i.e., inquiries about trends, comparisons, and exact numerical quantities). Results showed that tables, bar grables (combined bar graph and table) and line grables produced the most accurate responses, and line graphs and bar charts produced the fastest responses across question types. Results are discussed with respect to prior theoretical work and the potential benefits of hybrid forms of quantitative data displays for multiple kinds of data extraction inquiries.

INTRODUCTION

In recent years there has been increasing interest in how to best condense large amounts of data. The goal of most summarization techniques is to enable quick and accurate extraction of information about specific quantities, trends, and comparisons.

The most common way of conveying quantitative information is in the narrative form where trends and comparisons are described in words and the summaries of the numerical data (e.g., means, percentages) are contained within the text. However, information is not always effectively conveyed through the narrative form, especially when reporting more than a few individual pieces of data. *The Publication Manual of the American Psychological Association* (1994) suggests that when data are short and simple they should be narratively presented. More than three or four numbers should be reported in a non-narrative form.

There are two primary methods for visually displaying quantitative information in a non-narrative form: numerical and spatial. The numerical display, generally known as a table, presents specific (precise) quantities in an alphanumeric form. The spatial display, known as a graph or chart, presents the information in a picture form.

The issue of how to best present data was considered over 200 years ago when Playfair (1786) began to examine how graphs could be used to show trends in data. In the 1920s this problem still existed when a debate arose in the *Journal of American Statistical Association* regarding the merits of bar versus pie charts (Eells, 1926; Croxton, 1927; Croxton and Stryker, 1927; Von Huhn, 1927). The question of whether one type of graph is better than others still exists.

With tables, specific numerical quantities are easier to extract and they are less likely to produce misleading interpretations than graphs (Ehrenberg, 1975; Tufte, 1983). However, it is generally more difficult to discern visual patterns such as trends from multi-celled tables than from graphs (Spence and Lewandowsky, 1991; MacDonald-Ross 1977). Spence and Lewandowsky (1992) and Sanderson, Flach, Buttigieg, and Casey (1989) found that graphs can produce emergent features allowing for faster, more accurate data interpretation.

One explanation for the potential advantage of graphs compared to tables is Wicken's (1992) proximity-compatibility principle. This principle states that data integration processes are facilitated by an object-like presentation, or in this case, a graphical format. Object displays are advantageous for two reasons: (1) they foster parallel processing, and (2) they are more likely to allow pattern formation that serves to aid information integration.

Currently, there are many published guidelines on graph construction with recommendations and procedures for their creation (Kosslyn, 1994; Schmid and Schmid, 1979; Tufte, 1983). However, Cleveland and McGill (1984) concluded that the standard method for choosing a "... graph design for data analysis and presentation is largely unscientific." Indeed, there is sparse empirical research supporting most design principles, and the results are frequently equivocal. For example, some research supports the use of bar charts over pie charts; other research finds the opposite (Croxton, 1927; Von Huhn, 1926; Eells, 1926; Spence and Lewandowsky, 1991). These equivocal results might be due to the particular images used and the type of task that observers are asked to perform. Indeed, Spence and Lewandowsky (1991) found that pie charts were better than bar charts for making comparisons among proportions, but bar charts were better than pie charts when making direct magnitude estimates. They concluded that unless one is trying to transmit precise numerical values to the viewer, tables are inferior to charts and graphs. Carswell (1990) found that judgments requiring focused attention (e.g., seeking precise numerical values) are better performed with bar charts than line graphs. Thus it appears that the discrepant results between and within experiments can at least be partially attributed to the varied kinds of data acquisition tasks and displays employed in the different studies. These studies also illustrate that quantitative displays can be used in various ways for different tasks.

Powers, Lashley, Sanchez, and Shneiderman (1984) also attempted to determine which form of data display is the easiest to comprehend. They hypothesized that more usable information can be conveyed using both a table and graph together than by using either alone. They found that tables alone as opposed to graphs alone increased comprehension. However, when both tables and graphs were provided together, slower but more accurate performance was produced. Powers et al. (1984) recommended using the display form most familiar and comfortable to target users.

Given that displays might be used for multiple purposes (i.e., determining exact quantities, forecasting trends, making comparisons) by the same or different persons, display effectiveness might be enhanced if aspects of both graphs and tables were combined into a single form. This display which we call the *grable*, combines features of graphs and tables. By having both kinds of features, grables may efficiently accommodate a wider variety task goals than either graphs or tables by themselves. Alternatively, the added material might clutter the display hindering data extraction and reducing accuracy (Tufte, 1983).

The present experiment evaluates seven display types: three conventional forms of graphs (line, bar, and pie), three forms of graph-table combinations (line grable, bar grable, and pie grable), and the table form in three types of information extraction tasks (determining numerical values, analyzing trends, and making comparisons). It is predicted that grables, due to the combined nature of both tables and graphs, will allow for more accurate and faster interpretation than graphs or tables across question type.

METHOD

Participants

Participants were 63 undergraduates between the ages of 18 and 36 years of age (67% female) from North Carolina State University who fulfilled a course requirement.

Materials

Tables and graphs were produced by Microsoft Excel 5.0 and were laser printed on 21.6 cm by 28 cm (8.5 by 11 inch) white paper in landscape orientation.

A total of 441 sheets of tables and/or graphs were produced based on 21 different scenarios, 3 question types, and 7 display types. Scenarios covered a variety of quantifiable situations including: stock prices, number of mountain bikes sold during a two year period, weight loss methods, types of student housing, movie ratings, and miles two salespersons traveled. An example of one of the scenarios in each of the seven presentation methods is shown in reduced form in Figure 1.

The three question types were: (a) numerical, e.g., what was the price of stock 2 during week 5?, (b) trends, e.g., if you bought stock 1 during the first week and sold it during the fourth week, would you have made any money?, and (c) comparisons, e.g., which stock was less expensive—stock 2 during the third week or stock 1 during the fourth week?

The seven presentation methods were (a) line graph, (b) line grable (line graph with adjacent numbers), (c) bar graph (vertically oriented), (d) bar grable (vertical bar graph with adjacent numbers), (e) pie chart, (f) pie grable (pie chart with adjacent numbers), and (g) table.

All alphanumeric characters were printed in Times font. Data labels and axis labels were 12 point. The exact numbers in the grables were 10 point. All questions were 18 point.

Procedure

Each participant was provided with a packet of 21 stimulus sheets, a response sheet, and blank paper. Every participant viewed all 21 scenarios which were balanced through the seven display types and three question types across participants. Participants viewed a number, trend, and comparison question for each of the seven display types.

Participants were instructed to answer the question located on the bottom of each display as quickly and as accurately as possible, and to perform any work necessary to formulate their answer on the blank paper provided before writing their final response. Time was recorded from the turn of the each stimulus page to the turn of the next page. Changing of answers was not permitted.

Responses were scored as being either correct (1) or incorrect (0) as the accuracy measure. When numbers were not provided (and because some numerical answers were quite large), participants could not answer any of the questions precisely; therefore, answers to the number questions were considered correct if the responses were within 10 percent of the correct answer. For example, if the correct answer was 9786, an answer plus or minus 97 was considered correct. Time (in s) was recorded by the experimenter using a stopwatch.

RESULTS

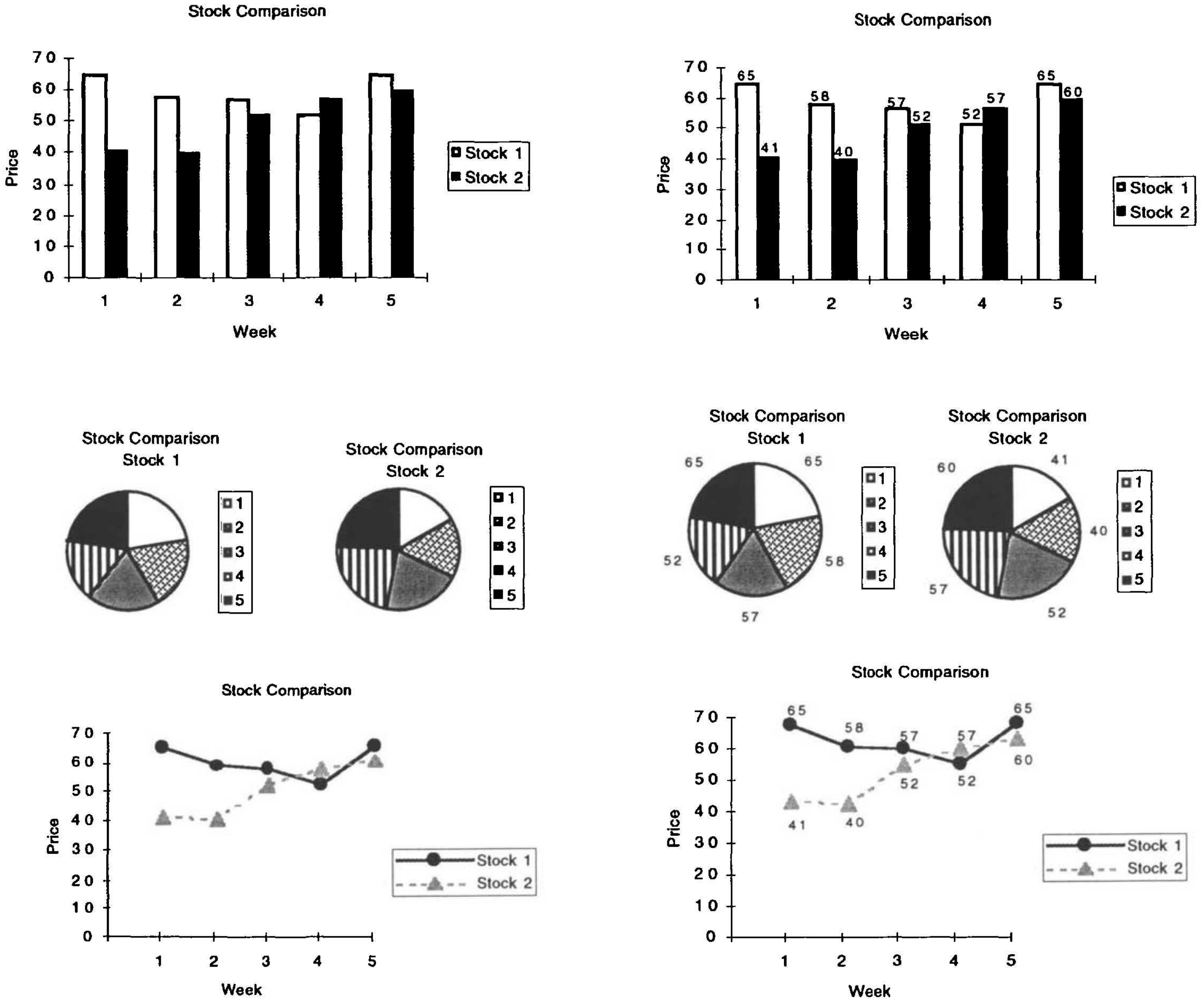
Separate 3 (display type) x 7 (question type) repeated-measure analyses of variance (ANOVAs) were performed on the accuracy and time data.

Accuracy

Table 1 shows the proportion correct means. The ANOVA indicated a significant main effect of question type, $F(2, 124) = 3.85$, $p < .03$. Comparisons among the means using Tukey's Honestly Significant Difference (HSD) test ($p < .05$) showed that accuracy was significantly higher for the comparison questions compared to the number questions. Accuracy on the trend questions was intermediate and not significantly different from the other two question types.

The ANOVA also showed a significant main effect of display type, $F(6, 372) = 55.26$, $p < .0001$. Comparisons among the means using the Tukey's HSD test indicated that the four displays with numbers (line grables, bar grables, pie grables, and tables) were not significantly different from one another, but all produced significantly more accurate

Figure 1
Example Scenario in the Seven Display Forms: Bar Graph, Bar Grable, Pie Chart, Pie Grable, Line Graph, Line Grable, and Table (in reduced size).



Stock Comparison

		Stock 1	Stock 2
	1	$65	$41
	2	$58	$40
Week	3	$57	$52
	4	$52	$57
	5	$65	$60

Table 1
Proportion accuracy means as a function of question and display type.

	Question Type			
Display Type	*Number*	*Trend*	*Comparison*	*mean*
line	.49	.67	.68	.61
line grable	.97	.71	.76	.81
bar	.62	.75	.81	.72
bar grable	.84	.84	.90	.86
pie	.05	.32	.46	.28
pie grable	.86	.75	.87	.83
table	.90	.87	.76	.85
mean	.68	.70	.75	

Table 2
Response time means (in s) as a function of question and display type.

	Question Type			
Display Type	*Number*	*Trend*	*Comparison*	*means*
line	40.49	49.24	50.63	46.79
line grable	39.37	60.43	57.14	52.31
bar	36.84	53.06	47.21	45.70
bar grable	39.87	60.10	50.67	50.21
pie	52.98	47.37	57.62	52.66
pie grable	43.98	71.76	61.52	59.09
table	30.24	59.24	61.90	50.46
mean	40.54	57.31	55.24	

responses than the remaining displays. Bar charts and line graphs did not differ but both produced significantly greater accuracy than pie charts.

The ANOVA also showed a significant interaction of question type and display type, F (12, 744) = 5.51, $p < .0001$. Comparisons were made using simple effects analyses and contrasts among pairs of means ($p < .05$). Examination of Table 1 as well as the statistical comparisons among cell means show a pattern that is consistent with that already described for the display-type main effect (i.e., the displays with numbers produced greater accuracy than displays without numbers with the pie chart being the least accurate). However, there were a few exceptions: (a) for the trend questions, the table produced significantly more accurate responses than the line grable; and (b) for the comparison questions, the bar grable produced significantly more accurate responses than the line grable and the table.

Time

Table 2 shows the mean times (in s) for conditions. The ANOVA showed a significant main effect of question type, F (2, 124) = 41.42, $p < .0001$. Comparisons among the means using Tukey's HSD test ($p < .05$) showed that responses to number questions were significantly faster than trend and comparison questions, with the latter two question types not differing.

The ANOVA also showed a significant main effect of display type, F (6, 372) = 5.73, $p < .001$. Comparisons among the means showed that the pie grable produced significantly slower response times than all other display types except the pie chart. No other difference was significant.

The ANOVA also showed a significant interaction of question type and display type, F (12, 744) = 4.30, $p < .0001$. Comparisons were made using simple effects analyses followed by contrasts between pairs of means ($p < .05$). The cell means in Table 2 show a fairly complex pattern in which response time changes as a function of question and display type. The following description outlines the significant comparisons. (a) For number questions, the response times reflect a pattern similar to the main effect of display type already described. The pie chart produced significantly slower response times than all other displays except for the pie grable. The pie grable was significantly slower than the table (for which the fastest times were found). (b) For the trend questions, the pie grable produced significantly slower response times than all other displays. The line grable, the bar grable, and the table produced significantly slower response times than the line graph and pie chart. (c) For the comparison questions, the table and pie grable produced significantly slower response times than the line graph and the two bar displays.

DISCUSSION

Although the data show a fairly complicated set of results, several patterns are discernible. Pie charts consistently produced the worst accuracy rate regardless of question type. The displays with numbers (tables, and the three types of grables) produced the most accurate responses across question types. Thus grables, which combine the precise quantities of tables with the emergent features of graphs, produced more correct interpretations than conventional graphs across inquiry types.

The response time results were more complicated than the accuracy results. In general, these data show that the two pie displays produced the slowest response times. However, response time also depended on question and display type. For numerical inquiries, most of the displays with numbers (except for the pie grable) fared better than those without. In addition, bar graphs produced relatively fast times (similar to the numerical displays). For trend inquiries, response times for pie grables were particularly poor, while the conventional graph types (pie charts, line graphs, and bar graphs) did well. For comparison inquiries, line graphs and the two types of bar displays produced faster response times than tables and pie grables.

Thus looking across the data, there are hints of a speed-accuracy tradeoff. Thus, some of the displays produced greater accuracy but took longer. This pattern is more apparent for the trend and comparison questions then for the number questions. In the latter, displays with numbers (the tables and the grables) tend to produce both greater accuracy and faster times. This finding is not unexpected since specific numbers are requested and these displays have them. It is more difficult to ascertain exact quantities with graphs having categories without assigned numbers since using the ordinate scale axes generally provide approximations. Pie charts which just show relative category size are indeterminate in this regard and showed extremely poor accuracy not only for numerical but also for trend and comparison inquiries. Of the three grables, the bar grable (and to some extent, the line and pie grables) showed good accuracy for both the trend and comparison questions. This finding is not unexpected since these displays provide both a graphic representation and specific numbers. Because tables lack the visual qualities of graphs, it is somewhat surprising that tables produced relatively high accuracy for both the trend and the comparison questions. However, for these two types of questions, tables produced relatively slow response times. With tables, the formation of a visual mental representation requires additional processing. The longer latency might therefore reflect the higher cognitive load involved in transforming the numbers to a more usable representation.

Two of the combined displays, line grables and bar grables, took longer to answer in some cases than the line graph or bar chart; this is not surprising because these displays contain more information. The extra few seconds produced more accurate interpretations than the conventional graphs. So, although there is some indication of a speed-accuracy tradeoff, if interpretation accuracy is the primary goal, then bar and line grables appear to be the best choice. The bar grable appears best for comparison and trend questions, and the line grable appears best for number questions.

The results have implications with respect to existing recommendations and empirical research. Tufte's (1983) data-ink ratio guideline predicts that redundant information such as the inclusion of numerical values (as seen in grables) would degrade performance. Kosslyn (1994) recommends leaving off specific values because they force the reader to perform more work increasing the likelihood of interpretation errors. Moreover, Wickens and Andre (1990) showed no effect of adding numbers to displays. Whereas these guidelines and research are supported to some extent by the current study's response time results (depending on inquiry type), they are not supported by the accuracy results.

With respect to specific display types, the results failed to support Wicken's (1992) suggestion that a bar display would degrade performance relative to a line display. However, the results support Sanderson et al.'s (1989) finding that bar graphs support task performance better than line displays.

Further research on the best ways of display quantitative information is needed, particularly for the hybrid graph-table that we have termed the grable. One area that needs further investigation includes the size and positioning of the alphanumerics in grables. For example, in bar grables should the numbers be placed inside, above or on the side of the bar? If the numbers are placed inside the bar, then the emergent features that they might form would not be disrupted. At the same time, bars frequently have shading and so contrast could be reduced without white space surrounding the number.

The present research suggests that the grable form of quantitative display has potential for communicating information across varied types of inquiry. Systematic research in this area will likely benefit display decisions and users' performance in interpreting them.

REFERENCES

American Psychological Association. (1994). *Publication Manual of the American Psychological Association (4th ed.).* Washington, DC: Author.

Bennett, B., and Flach, J. (1992). Graphical displays: Implications for divided attention, focused attention, and problem solving. *Human Factors, 34,* 513-533.

Carswell C. M., (1992). Choosing specifiers: An evaluation of the basic tasks model of graphical perception. *Human Factors, 34,* 535-554.

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

Croxton, F. E. (1927). Further Studies in the Graphics Use of Circles and Bars II: Some Additional Data. *Journal of the American Statistical Association, 22,* 36-39.

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

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

Kosslyn, S. M. (1989). Understanding Charts and Graphs. *Applied Cognitive Psychology, 3,* 185-193.

Kosslyn, S. M. (1994). *Elements of Graph Design,* New York: W. H. Freeman.

MacDonald-Ross, M. (1977). How numbers are shown: a review of research on the presentation of quantitative data in texts. *Audio Visual Communication Review, 25,* 359-407.

Playfair, W. (1786). *The Commercial and Political Atlas*. London.

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

Robinson, A. H., Sale, R. D., and Morrison, J. (1978). *Elements of Cartography.* New York: John Wiley.

Sanderson, P. M., Flach, J. M., Buttigieg, M. A., and Casey, E. J. (1989). Object displays do not always support better integrated task performance. *Human Factors, 31,* 183-198.

Spence, I., and Lewandowsky, S. (1991). Displaying proportions and percentages. *Applied Cognitive Psychology, 5,* 61-77.

Tufte, E. R. (1983). *The Visual Display of Quantitative Information*. Chesire, CT : Graphics Press.

Von Huhn, R. (1927). Further studies in the graphic use of circles and bars: A discussion of Eels experiment. *Journal of the American Statistical Association, 22,* 31-36.

Wickens, C. D. (1992). *Engineering Psychology and Human Performance.* New York: Harper Collins.

Wickens, C., and Andre, A., (1990). Proximity compatibility and display: Effects of color, space, and objectness on information integration. *Human Factors, 32,* 61-77.