BEHAVIOUR & INFORMATION TECHNOLOGY, 1996, VOL. 15, NO. 2, 96–100

The use or misuse of three-dimensional graphs to represent lower-dimensional data

MICHAEL SIEGRIST

Abteilung Sozialpsychologie, Universität Zürich, Plattenstrasse 14, CH-8032 Zürich, Switzerland

Abstract. Some statisticians hold strong opinions regarding graphs with a 3-D look. However, in experiments little attention has been paid to the effects of adding decorative depth. The performance of subjects on pie charts and bar charts with and without 3-D was evaluated in the present experiment. Subjects were asked to make relative magnitude estimates for different graphs. For pie charts, better performance was observed for 2-D than for 3-D charts. For the bar charts, a more differentiated picture emerged: performance was dependent on the position, height and dimension of the bars. However, 3-D bar charts had the one disadvantage that subjects needed more time to evaluate this type of graph.

1. Introduction

With most commercial graphics software it is easy to produce 3-D graphics. The third dimension has a mere aesthetic function, as it does not present new information (see figure 1 for specific examples of graphs with and without 3-D effects). Three-dimensional pie charts and bar charts are used not only in the popular press and in business communications, but they can also be found in some scientific journals. The decision to use graphs with or without 3-D effects is difficult, as little research has been done to determine the effect of decorative perspectives on the viewer's perception of the data. Furthermore, the recommendations of statisticians vary and are not based on empirical findings.

Tufte (1983: 93) has proposed the data ink ratio, which is the 'proportion of a graphic's ink devoted to the non-redundant display of data-information'. According to Tufte (1983), the data ink ratio should be maximized in a good graphic. It is therefore not surprising that he refers to 3-D-displays as 'chartjunk'. Spence and Lewandowsky (1990) do not agree; based on empirical findings they believe that the addition of extra dimensions is not harmful. Irrelevant dimensions make a graph more attractive, and it is processed more quickly. An extra dimension which carries no information does not result in a loss in judgmental accuracy. Kosslyn (1994) argues that the visual system does not estimate volume very accurately. For this reason he recommends that 3-D perspectives not be used to communicate precise information.

Quantitative judgments on the basis of different kinds of graphs lead to recommendations for the construction of graphs (Cleveland *et al.* 1983, Cleveland and McGill 1984). Cleveland and McGill (1984) used a relative magnitude estimation task to evaluate the different elementary perceptual tasks performed by the viewer of a graph. They proposed a ranking order of these perceptual tasks (e.g., it is easier to estimate length than volume). Carswell (1992) conducted a meta-analysis of experiments on perception of graphics and performance. The results support the proposed ranking order only partly. The work of Cleveland and colleagues was criticized (Morris and Thisted 1986), because important sources of error were neglected without commentary. The performance of a task can vary systematically across different contexts, and the merits of a graph depend on the task performed (Hollands and Spence 1992).

Casali and Gaylin (1988) found that 3-D bar graphs were less effective than 2-D bars for reading specific values. However, the third dimension in this case carried information on a second factor and was therefore not a purely decorative dimension. Spence (1990) found that the accuracy for disks (which have a different angle of perspective than pies) was less than for pies, but the effect was not large. For other charts, dimensionality did not have a great effect on either accuracy or latency. He therefore recommends the use of bars, boxes, or cylinders for representing numerical quantities. Carswell *et al.* (1991) tested the effects of using 3-D graphs to represent lower dimensional data. They did not find significant differences for bar charts and pie charts. Only 3-D line graphs yielded reliable decrements in performance. There are several possible reasons why Carswell *et al.* (1991) could not find any differences between 2-D and 3-D pie charts and bar charts. They used only two segments or two bars, respectively. Furthermore, the perspective angle of the 3-D pie chart was almost the same as that of the 2-D pie chart; the front surface of the 3-D bar chart was black, while the other surfaces were white. It was therefore an easy task for subjects to compare the two black rectangles and to ignore the third dimension.

It is possible that their results are only valid for the specific graphs which Carswell *et al.* (1991) used in their experiments. The goal of the present study was to test the effects of using 3-D graphs to represent lower dimensional data on performance with varying stimuli. Furthermore, the influence of other factors, such as size and position of the bars, was evaluated.

The main hypothesis was that subjects would make better estimates for 2-D graphs than for 3-D graphs. For the bar charts an additional hypothesis was formulated. Subjects should be able to estimate the size of the second bar more accurately than the size of the third bar. We did not formulate an hypothesis regarding the size of the bars, but this factor was taken into account when analysing the data.

2. Method

2.1. *Subjects*

The subjects were 49 students (30 women and 19 men). Their ages ranged between 20 and 56 years (M = 26.1 and SD = 7.7).

2.2. *Design*

Different types of graph were evaluated in the present experiment. Subjects had to estimate the size of a slice of 2-D and 3-D pie charts. Bar charts with three bars were also investigated. Subjects had to compare two bars and to estimate the relative size of the second or the third bar as compared to the first bar. For some graphics the true value was smaller than 50%, and for some graphics the true value was greater than 50%.

2.3. *Graphs*

Stimuli were generated using the program EXCEL on a Macintosh computer. The size of the segments of the pie charts was determined by chance with the restriction, however, that the size of the test segment be between 10% and 49% of the whole pie. The start of the first segment was also determined by chance. The segments were labelled with the letters A, B, and C. In all instances, the subjects' task was to estimate the size of segment A. Examples of the pie charts used are presented in figure 1. Ten pie charts were generated. The same charts were presented in a 2-D and 3-D version.

For the construction of the bar charts two factors were taken into account: size (<50%/>50%) and position (second bar/third bar). The other restrictions were that the differences between the bars could not be smaller than 5% and that the height of the second and the third bar had to be between 10% and 90% of the first bar. Furthermore, the third bar had to be smaller than the second bar. Five different graphs of each type were generated. The same charts were presented in a 2-D and a 3-D version. Only the 3-D bar charts had a frame. This distinction between the 2-D and 3-D graphs was due to the fact that most 2-D graphs published do not have a frame, while 3-D graphs commonly do. Examples of the 2-D and 3-D bar charts used are presented in figure 1.

A total of 60 graphs was used. The charts were arranged in random order. The order of presentation remained the same for all subjects.

2.4. *Procedure*

An example of each type of graph was presented to the subjects. Pie and bar charts are used for different purposes, and therefore two different tasks were used to evaluate different versions of bar and pie charts. As we did not plan to compare pie charts with bar charts, it is not crucial to the present experiment that two different tasks were used. For the pie charts, subjects were told to estimate the percentage of the whole represented by the slice labelled with A. For the bar charts, subjects had to estimate the percentage of the second or the third bar relative to the first bar. Subjects were told that they should work rapidly yet accurately. To prevent subjects from rounding their estimates to the nearest tens, they were told that they could use all numbers between 1 and 100 for their estimates. A program written by the author presented the stimuli on the screen of a Macintosh II si. Subjects had to enter their estimates. The program recorded the estimates and reaction times in a separate file.

3. Results

Deviation errors were computed as percent-difference scores between subjects' judgments and the actual values. An exploratory data analysis yielded some outliers. This is a well known problem with the kind of data we used in the present study (Kruskal 1982). Most of the outliers were due to the fact that subjects evaluated the wrong bar or segment. For this reason we decided not to take into account differences which were greater than or equal to 20%. About 1% of the estimates were thus eliminated. The means

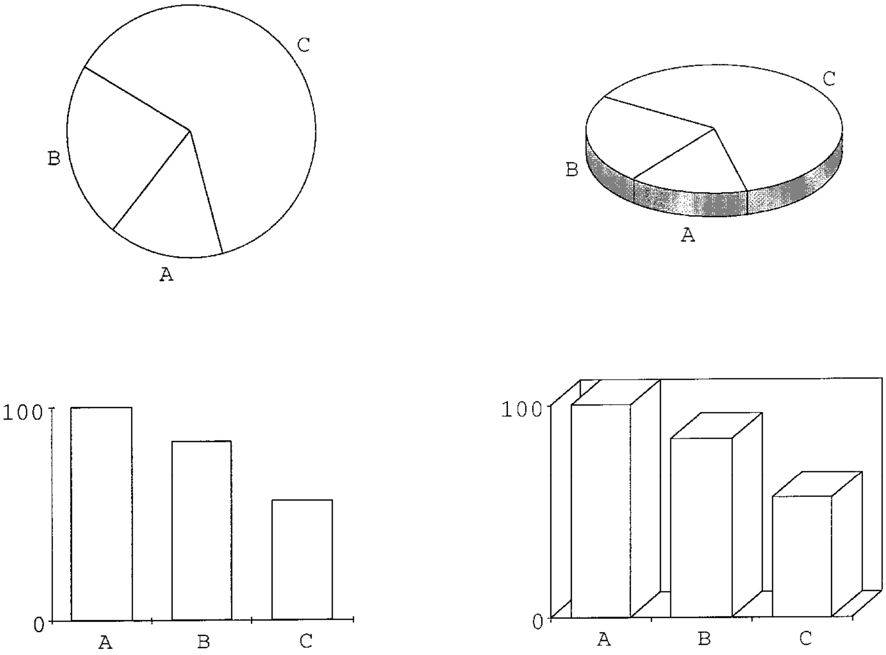

Figure 1. Examples of the 2-D and 3-D pie charts and bar charts which were used in the experiment.

of the deviation errors for each subject and graph type were then computed.

The deviation errors associated with 3-D pie charts (M = 3.2) were higher than those associated with 2-D pie charts (M = 2.7). A *t*-test for paired samples yielded a significant result, $t(48) = 3.71$, $p = 0.001$.

The errors associated with the bar charts were submitted to a complete repeated measurement analysis of variance with three factors (2 dimensions $\times$ 2 positions $\times$ 2 heights). The analysis yielded a significant main effect for position ($F(1,48) = 8.16$, MSE = 1.69, $p < 0.01$), a significant interaction between position and height ($F(1,48) = 10.56$, MSE = 1.59, $p < 0.005$), and a significant interaction between dimension and height ($F(1,48) = 5.10$, MSE = 0.99, $p < 0.05$). No other significant main effects or interactions were observed. The means of the different graph types are presented in figure 2.

The absolute errors do not tell us whether the differences were the result of consistent underestimates or overestimates. To answer this question a second analysis was performed, whereby the mean differences between estimated magnitude and actual magnitude for each subject and graph type were computed. The analysis yielded that subjects overestimated the segments of the 2-D pie charts (M = 0.78) and of the 3-D pie charts (M = 1.28). The difference was significant, $t(48) = 2.56$, $p < 0.05$. For the bar charts, a complete repeated measurement analysis of variance with three factors (2 dimensions $\times$ 2 positions $\times$ 2 heights) yielded a significant main effect for position ($F(1,48) = 31.98$, MSE = 3.81, $p < 0.001$), a significant main effect for dimension ($F(1,48) = 14.16$, MSE = 3.95, $p < 0.001$), and an interaction between these two factors ($F(1,48) = 14.03$, MSE = 2.17, $p < 0.001$). This interaction was caused by underestimation of the second bar in the 3-D graphs and overestimation of the second bar in the 2-D graphs.

The time needed by subjects to make their decisions was measured. In the analysis, only the times for those graphs were used for which we had valid estimates. Furthermore, we eliminated from analysis times for estimates for which subjects first entered non valid characters (e.g., a decimal point). As the distributions of the times were skewed, we used logarithms of the times. The analysis yielded no significant difference between 2-D (M = 15.2 s, SD = 6.3) and 3-D (M = 15.4 s, SD = 6.4) pie charts, $t(48) = 0.32$, n.s. An analysis of variance with three factors (2 dimensions $\times$ 2 positions $\times$ 2 heights) was conducted to analyse the logarithms of the times for the bar charts. The analysis yielded a significant main effect for dimension ($F(1,48) = 5.41$, MSE = 0.03, $p < 0.05$) and a significant interaction between dimension and height ($F(1,48) = 7.20$, MSE = 0.02, $p = 0.01$). Subjects needed more time to evaluate 3-D graphs than 2-D graphs, and this difference was especially pronounced for small bars. Table 1 shows the means and standard deviations of the response times.

4. Discussion

Subjects' estimates were better for 2-D pie charts than for 3-D pie charts. The present experiment yielded results which were different from those of Carswell *et al.* (1991), who did not find any difference between these two types of graph. The reason for this is that we used a different

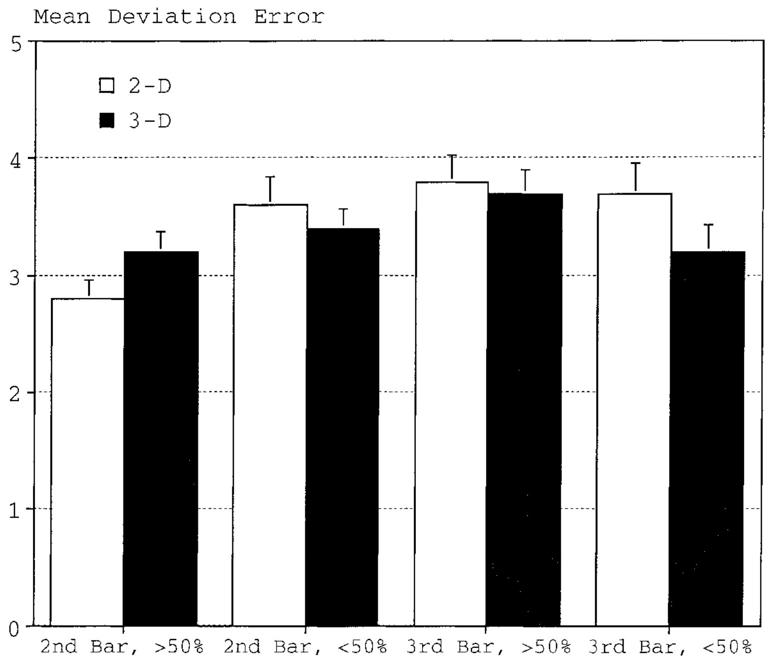

Figure 2. Mean deviation scores for bar charts as a function of dimension (2-D/3-D), position (2nd/3rd), and size (<50%/>50%). Deviation errors are percent-difference scores between subjects' judgements and actual value. Error bars indicate the standard errors of the means.

perspective angle for the 3-D pies than Carswell *et al.* (1991).

The present data suggest that the perspective angle of the pie has an influence on the accuracy of estimates. With a low perspective angle it is especially difficult to estimate the size of the segments. It becomes more difficult to estimate size, because a change in the perspective angle also changes the angle of the segments. The addition of a third dimension leads to a qualitative change in the graph; it changes the angles of the segments. There are aesthetic reasons for using a 3-D chart, of course. The use of 3-D graphs can be appropriate, especially when it is important to draw attention to the graph. However, the perspective angle of the 3-D graph should not be too different from the 2-D pie chart. Otherwise, a decrease in the accuracy of the estimates is very probable.

The 2-D bar charts were not superior to the 3-D bar charts. However, the position of the bar was important: analysis yielded a significant main effect for this factor.

Table 1. The mean and standard deviation of the response latencies in seconds for bar charts with the factors dimension (2-D/3-D), position (2nd/3rd) and size (<50%/>50%).

		2nd bar		3rd bar	
		<50%	>50%	<50%	>50%
2-D	M	10.8	10.9	10.8	11.9
	SD	4.1	4.0	3.8	5.7
3-D	M	11.5	11.3	11.9	11.6
	SD	4.4	4.8	4.6	4.7

Furthermore, data analysis yielded two significant interactions: position $\times$ size and dimension $\times$ height. A different picture emerged when differences, and not deviation scores, were analysed. An underestimation of the second bar in the 3-D graphs and an overestimation of the second bar in the 2-D graphs were observed. Depending on the aim to be achieved by a graph, a 2-D or a 3-D representation can be more appropriate. However, 3-D graphs had a clear disadavantage, in that the subjects required more time in order to evaluate them. This was especially the case for small bars.

The results of the present experiment suggest that one should not investigate just one single factor, such as dimensionality alone. The number of bars and their relative height have an impact on the accuracy of the estimates as well. Possible interactions between different parameters may also be found. The results thus lead to an important conclusion: stylized graphs with two bars, as were used in many experiments (e.g., Cleveland and McGill 1984, Spence 1990, Carswell *et al.* 1991), do not allow evaluation of different types of graph. More realistic graphs must be used in order to gain results that can serve as guidelines for the production of good, readable graphs.

These results can be used to evaluate Tufte's (1983) hypothesis that charts with a high data ink ratio are superior to charts with a low data ink ration. This hypothesis is not supported by our data. We therefore agree with Spence (1990) and Kosslyn (1985), who had some doubt as to the wisdom of Tufte's recommendations for empirical and theoretical reasons.

Acknowledgement

The author gratefully acknowledges the assistance of Pierre Martin in data collection.

References

Carswell, C. M. 1992, Choosing specifiers: an evaluation of the basic tasks model of graphical perception, *Human Factors*, **34,** 535–554.

Carswell, C. M., Frankenberger, S. and Bernhard, D. 1991, Graphing in depth: perspectives on the use of three-dimensional graphs to represent lower-dimensional data, *Behaviour & Information Technology*, **10,** 459–474.

Casali, J. G. and Gaylin, K. B. 1988, Selected graph design variables in four interpretation tasks: a microcomputer-based pilot study, *Behaviour & Information Technology*, **7,** 31–49.

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

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.

Hollands, J. G. and Spence, I. 1992, Judgments of change and proportion in graphical perception, *Human Factors*, **34,** 313–334.

Kosslyn, S. M. 1985, Graphics and human information processing, *Journal of the American Statistical Association*, **80,** 499–512.

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

Kruskal, W. H. 1982, Criteria for judging statistical graphics, *Utilitas Mathematica, Ser. B.*, **21B,** 283–310.

Morris, M. and Thisted, R. A. 1986, Sources of error in graphical perception: a critique and an experiment, *Proceedings of the Section on Statistical Graphics/American Statistical Association*, 43–48.

Spence, I. 1990, Visual psychophysics of simple graphical elements, *Journal of Experimental Psychology: Human Perception and Performance,* **16**, 683–692.

Spence, I. and Lewandowsky, S. 1990, Graphical perception, in J. Fox and J. S. Long (eds) *Modern Methods of Data Analysis* (Sage, Newbury Park), 13–57.

Tufte, E. R. 1983, *The Visual Display of Quantitative Information* (Graphics Press, Cheshire, Connecticut).