An overview of scientific visualization techniques applied to experimental psychology

FRANK M. MARCHAK
TASC, Reading, Massachusetts

This paper presents an overview of current data visualization techniques. The various types of graphics, such as contour plots, surface plots, scatterplot matrices, and dynamic spinning, are described. Then, using a data set typical of the field, the potential merits and pitfalls of each visualization technique are presented in terms of what aspects of the data they make explicit. The results illustrate the benefits that a psychological researcher can expect from data visualization capabilities.

The graphics capabilities of personal computers has brought the power of data visualization techniques within reach of most scientists and researchers. The majority of applications of visualization have been in areas of the physical sciences, such as meteorology, astronomy, and molecular modeling, where large data sets abound. However, the application of these techniques is not limited to the realm of terabytes of data.

Psychologists collect experimental data by using a wide range of paradigms, but tend to restrict their graphical interactions to the standard line graphs, histograms, and bar charts. This rather myopic view has been noted by others in the field. Castellán (1991), in a review of the effects of computers in psychology over the last 20 years, discussed the relatively untapped benefits that these machines provide for exploratory data analysis. Butler (1993), in his presidential address to the Society for Computers in Psychology (SCiP), reviewed the use of conceptual graphs in the understanding of psychology, while also observing the relative dearth of advanced representation techniques.

But the benefits of graphical representation have also been extolled within SCiP. Loftus (1993) presented an approach that suggested eliminating the standard notion of hypothesis testing and replacing it with a more intuitive and informative graphical approach to data interpretation. F. M. Marchak and Whitney (1990), F. M. Marchak and L. C. Marchak (1991), and F. M. Marchak and Zulager (1992) examined the differences of static versus dynamic representations for the exploratory analysis of data, and described some benefits and drawbacks of each. Further, statisticians have discussed the merits of visualization not only for exploratory analysis (e.g., Becker & Cleveland, 1991), but for presentation of results (Cleveland, 1993). Many of these visualization techniques are available as a part of standard statistical analysis packages used by most researchers. However, overall, the field of ex-

perimental psychology has been slow to embrace this new collection of representation techniques.

The present paper represents an attempt at overturning this reticence by presenting an overview of current data visualization techniques and showing how they can be applied to the typical data obtained in a psychological experiment. The various types of graphics, such as raster images, scatterplot matrices, dynamic spinning, contour plots, surface plots, and overlays, are described. Using a data set that is typical of the field, the potential merits and pitfalls of each visualization technique, in terms of what aspects of the data they make explicit, is presented. The final outcome is a clear example of the benefits that a psychological researcher can expect from data visualization capabilities.

The Data Set

Similarity data are typical of the type of data collected in experimental psychology tasks. Figure 1 shows the Morse code sequence of dots and dashes for each letter in the alphabet. Figure 2 consists of a subset of a similarity matrix of confusions among Morse code signals for the letters of the alphabet, collected by Rothkopf (1957). One hundred fifty subjects, untrained in Morse code, made *same-different* judgments on the basis of auditory presentations of pairs of signals, separated by 1.4 sec. Each number in the matrix represents the percentage of *same* responses to the row signal followed by the column signal. For example, 24% of the subjects responded "same" when the signal for M (--) was followed by the signal for A (*-).

A	..	G	...	M	--	S	...	Y	----
B	----	H		N	-.	T	-	Z	----
C	----	I	..	O	---	U	..		
D	..	J	----	P	----	V	...		
E	.	K	-.	Q	----	W	---		
F		L		R	..	X	----		

Figure 1. Morse code representation of the alphabet.

Address correspondence to F. M. Marchak, TASC, 55 Walkers Brook Drive, Reading, MA 01867-3297.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	92	4	8	13	3	14	10	13	46	5	22	3	25	34	6	6	9	35	23	6	37	13	17	12	7	3
B	5	84	37	31	5	28	17	21	5	19	34	40	6	10	12	22	25	16	18	2	18	34	8	84	30	42
C	4	38	87	17	4	29	13	7	11	19	24	35	14	3	9	51	34	24	14	6	6	11	14	32	82	38
D	8	62	17	88	7	23	40	36	9	13	81	56	8	7	9	27	9	45	29	6	17	20	27	40	15	33
E	6	13	14	6	97	2	4	4	17	1	5	6	4	4	5	1	5	10	7	67	3	3	2	5	6	5
F	4	51	33	19	2	90	10	29	5	33	16	50	7	6	10	42	12	35	14	2	21	27	25	19	27	13
G	9	18	27	38	1	14	90	6	5	22	33	16	14	13	82	52	23	21	5	3	15	14	32	21	23	39
H	3	45	23	25	9	32	8	87	10	10	9	29	5	8	8	14	8	17	37	4	36	59	9	33	14	11
I	64	7	7	13	10	8	6	12	93	3	5	16	13	30	7	3	5	19	35	16	10	5	8	2	5	7
J	7	9	38	9	2	24	18	5	4	85	22	31	8	3	21	63	47	11	2	7	9	9	9	22	32	28
K	5	24	38	73	1	17	25	11	5	27	91	33	10	12	31	14	31	22	2	2	23	17	33	63	16	18
L	2	69	43	45	10	24	12	26	9	30	27	86	6	2	9	37	36	28	12	5	16	19	20	31	25	59
M	24	12	5	14	7	17	29	8	8	11	23	8	96	62	11	10	15	20	7	9	13	4	21	9	18	8
N	31	4	13	30	8	12	10	16	13	3	16	8	59	93	5	9	5	28	12	10	16	4	12	4	6	11
O	7	7	20	6	5	9	76	7	2	39	26	10	4	8	86	37	35	10	3	4	11	14	25	35	27	27
P	5	22	33	12	5	36	22	12	3	78	14	46	5	6	21	83	43	23	9	4	12	19	19	19	41	30
Q	8	20	38	11	4	15	10	5	2	27	23	26	7	6	22	51	91	11	2	3	6	14	12	37	50	63
R	13	14	16	23	5	34	26	15	7	12	21	37	14	12	12	29	8	87	16	2	23	23	62	14	12	13
S	17	24	5	30	11	26	5	59	16	3	13	10	5	17	6	6	3	18	96	9	56	24	12	10	6	7
T	13	10	1	5	46	3	6	6	14	6	14	7	6	5	6	11	4	4	7	96	8	5	4	2	2	6
U	14	29	12	32	4	32	11	34	21	7	44	32	11	13	6	20	12	40	51	6	93	57	34	17	9	11
V	5	17	24	16	9	29	6	39	5	11	26	43	4	1	9	17	10	17	11	6	32	92	17	57	35	10
W	9	21	30	22	9	36	25	15	4	25	29	18	15	6	26	20	25	61	12	4	19	20	86	22	25	22
X	7	64	45	19	3	28	11	6	1	35	50	42	10	8	24	32	61	10	12	3	12	17	21	91	48	26
Y	9	23	62	15	4	26	22	9	1	30	12	14	5	6	14	30	52	5	7	4	6	13	21	44	86	23
Z	3	46	45	18	2	22	17	10	7	23	21	51	11	2	15	59	72	14	4	3	9	11	12	36	42	87

Figure 2. Confusion matrix of Morse code similarity judgments.

Raster Images

Raster images assign a color to each data point on the basis of its value. In Figure 3, the frequency of *same* responses are mapped into gray-scale values, with black representing low values, white representing high values, and grays representing the values in between. As can readily be seen, the diagonal, representing the similarity judgments corresponding to pairs that are truly the same, appears brightest. Reviewing other white points reveals other judged similarities. The bright white point in the lower left corner corresponds to the judgments between X (•••) and B (•••). Similarly, the black point in the far lower left corner corresponds to the judgment of Z (•••) and A (••).

A feature of this type of display is the ability to manipulate the color table mapping of the values. By stretching, compressing, or changing the mapping, different aspects of the data may be highlighted. Raster images give a good overview of the data, highlighting the maximums and minimums. However, such color mappings provide minimal quantitative information.

Scatterplot Matrices

Figure 4 shows a scatterplot matrix for the first seven letter pairs of the confusion matrix. In a scatterplot matrix, individual scatterplots of each pair of variables are arranged into rows and columns. It shows the correlations between all the variables in a global way. For example, the cluster of points for B (•••) and C (•••) show a stronger grouping than those for the other variable pairs.

Since the matrix is symmetric on the diagonal, it is possible to display just the lower half. Further, the diagonal cells can be used to display histograms of the data values. Scatterplot matrices provide a good overview of the data and allow detailed comparisons on all possible pairings. However, with a very large number of variables, they can become unwieldy.

Dynamic Spinning

Scientific visualization techniques are especially useful for fostering exploratory data analysis by allowing easy

Figure 3. Raster image.

Figure 4. Scatterplot matrix.

manipulation and display of multivariate data. Dynamic spinning adds motion to the analysis task by providing the ability to rotate a data set in each of the three dimensions. Figure 5 shows a snapshot of the 3-D coordinates obtained from a multidimensional scaling solution to the Morse code similarity data. By rotating the points in 3-D space, structure that may not be evident in the static 2-D plots can be revealed. The two outlying points on the left correspond to the letters E (•) and T (–).

The benefits of motion cues can greatly enhance the perception of structure in a data set by employing visual pattern-recognition skills. However, caution must be exercised, as it is also possible to misidentify spurious collections of points as actual groupings.

Contour Plots

Contour plots are constructed by drawing lines at discrete data levels, similar to a topographical map. The lines form groupings within the overall data set. In Figure 6, the contours on the diagonal are very apparent. The center point of the large, X-shaped contour in the bottom left of the figure corresponds to the high-similarity judgments between letters E (•) and T (–).

The area between the contour lines can be shaded to highlight the differences between the layers. In addition, labels can be put on the lines to indicate the value at each level, providing quantitative information about the plot. The contour plot also provides detailed information about slope, so if the contour lines are close together, this indicates that values are changing rapidly. However, such plots can provide little feel for the overall form of the data.

Surface Plots

In a surface plot, the data values determine the height of the surface points above the plane. Such plots can often provide a good overview of the data as well as quantita-

tive information about data values. For the current data set, the surface plot (Figure 7) displays some of the drawbacks of this technique. With the large number of greatly disparate values in the confusion matrix, the sharp spikes in the plot provide little detail information; some features are obscured by the intervening peaks. In addition, the appearance of the surface plot is dependent on the viewpoint chosen; different viewpoints reveal different aspects of the structure. Despite the shortcomings of this technique with this data set, surface plots can be powerful tools for revealing underlying structure in data.

Overlays

Figure 8 shows a contour plot overlaid on an interpolated raster image. By combining multiple graphics, different features of the data can be made explicit. Through registration of images and highlighting, it is possible to display several variables on one graphic and present additional features of a data set. In this example, the contours emphasize the high-similarity judgment values. If labels were placed on the contours showing the levels, this method would bring quantitative information to the general, overall view of the raster image.

Conclusions

One element of scientific visualization techniques that cannot be conveyed in a paper is the interconnections between the graphic displays and the data. Although the

Figure 5. Dynamic spinning.

Figure 6. Contour plot.

Figure 7. Surface plot.

Figure 8. Overlay plot.

static graphics can convey a great amount of information, one of the true benefits of these techniques is the ability to look at related data points across images. Many visualization software packages support linking, in which selecting a point in one display highlights the same point in the

other graphics and the raw data. This ability is an extremely powerful tool for analyzing and understanding the relationships in the data.

In closing, one point needs to be kept in mind about the efficacy and usefulness of scientific visualization capabilities. Like any tool, these techniques have appropriate areas of application and can easily be abused. The impetus here is not to advocate the wholesale application of these techniques, but to highlight relevant uses. The principal goal is to promote understanding and present accurate information, not merely to create pretty pictures. It is extremely important to carefully consider the goal of a visualization before selecting a particular technique, and to be certain not to distort the information.

When properly employed, scientific visualization techniques can supplement and extend existing graphical techniques. These methods are especially useful for finding clusters of related data points and for detecting outliers or atypical points. The exploratory analysis of data from experimental psychology can be greatly enhanced by the judicious use of these powerful graphical techniques.

REFERENCES

- BECKER, R. A., & CLEVELAND, W. S. (1991). Viewing multivariate scattered data. *Pixel*, **2**, 36-41.
- BUTLER, D. L. (1993). Graphics in psychology: Pictures, data, and especially concepts. *Behavior Research Methods, Instruments, & Computers*, **25**, 81-92.
- CASTELLAN, N. J., JR. (1991). Computers and computing in psychology: Twenty years of progress and still a bright future. *Behavior Research Methods, Instruments, & Computers*, **23**, 106-108.
- CLEVELAND, W. S. (1993). A model for studying display methods of statistical graphics. *Journal of Computational & Statistical Graphics*, **2**, 323-344.
- LOFTUS, G. R. (1993). A picture is worth a thousand *p* values: On the irrelevance of hypothesis testing in the microcomputer age. *Behavior Research Methods, Instruments, & Computers*, **25**, 250-256.
- MARCHAK, F. M., & MARCHAK, L. C. (1991). Interactive versus passive dynamics and the exploratory analysis of multivariate data. *Behavior Research Methods, Instruments, & Computers*, **23**, 296-300.
- MARCHAK, F. M., & WHITNEY, D. A. (1990). Dynamic graphics in the exploratory analysis of multivariate data. *Behavior Research Methods, Instruments, & Computers*, **22**, 176-178.
- MARCHAK, F. M., & ZULAGER, D. D. (1992). The effectiveness of dynamic graphics in revealing structure in multivariate data. *Behavior Research Methods, Instruments, & Computers*, **24**, 253-257.
- ROTHKOPF, E. Z. (1957). A measure of stimulus similarity and errors in some paired-associate learning tasks. *Journal of Experimental Psychology*, **53**, 94-101.