

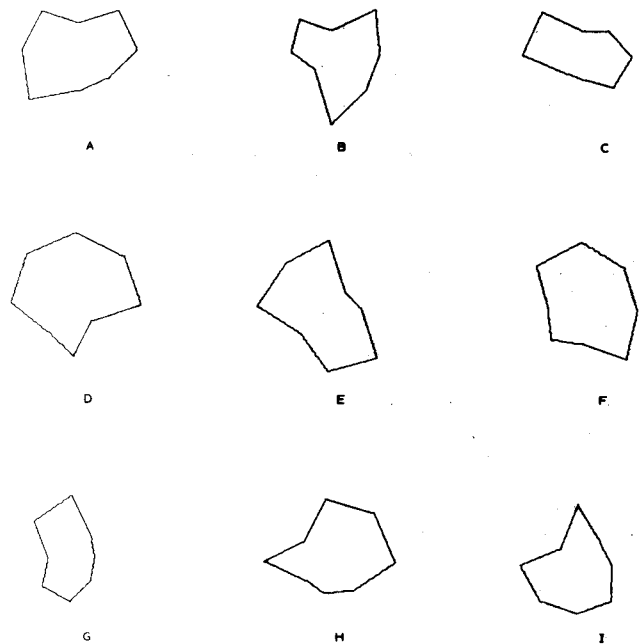
Stephen E. Fienberg's interesting article inadvertently provides an illustration of the importance of using care in applying graphical representations for multivariate data. He applies two such representations (STARS and FACES) to the same set of data in his article. If STARS are used, the points fall readily into three meaningful clusters, but this does not obtain with FACES. Further, Fienberg recounts that he found it difficult to obtain a satisfactory assignment of data components to figure parameters for FACES but not for STARS. It appears to me, however, that the differences he found between FACES and STARS arose from his inappropriate use of two different procedures for normalizing the components of the data before plotting.

Careful examination of Figures C and E of Fienberg's article reveals that each of the data components for the STARS was normalized by using the overall minimum and maximum of all the data components, whereas each component for the FACES was normalized by using the minimum and maximum values of that component only. The former method emphasizes the larger components (in this case, the first two, which most clearly define the three clusters); and the latter reveals variations in the smaller components. I submit that this difference alone explains why the STARS in the example seem to be easier to cluster visually than the FACES. The STARS in Figure A are plotted by using the same normalization procedure as that used to plot Fienberg's FACES. These STARS appear just as difficult to cluster as the FACES in his Figure E. Conversely, Figure B shows a set of FACES normalized by using the same procedure as that used for Fienberg's STARS. The clustering of interest is as clear for these FACES as it is for Fienberg's STARS (his Figure C). To create these FACES, I chose parameter assignments as similar to those of the original plot as possible, given that I was using a different, improved face-plotting routine (Jacob 1976).

The choice of an appropriate representation depends, of course, on what one is studying. Here it might be the main division into clusters or the structure within them. Nevertheless, these illustrations show that the differences Fienberg observed between two types of graphical representations were actually caused by two different normalization procedures. Applying the normalization procedure he used for the STARS to either STARS or FACES (and probably many other displays) gives a representation in which the main clustering is obvious; applying the procedure he used for the FACES to either display obscures the clusters (but emphasizes other information).

My own use of these displays has revealed another important difference between STARS and FACES. The visual effects of STARS are strongly influenced by which data components are adjacent to which other ones (because large variations in adjacent components result in sharp points). The visual effects of the parameters of FACES, by contrast, are interrelated in much more complex and less well known ways; but the main relationship is certainly not this simple, sequential ordering. The STARS, therefore, are a good choice when the principal organization of the data components is a sequential ordering (as in Fienberg's example). If it is some other organization, FACES are likely to be preferable. FACES are also particularly appropriate when the Euclidean distances between points, rather than their coordinate values, are of interest, as with factor analysis solutions (Wainer 1979).

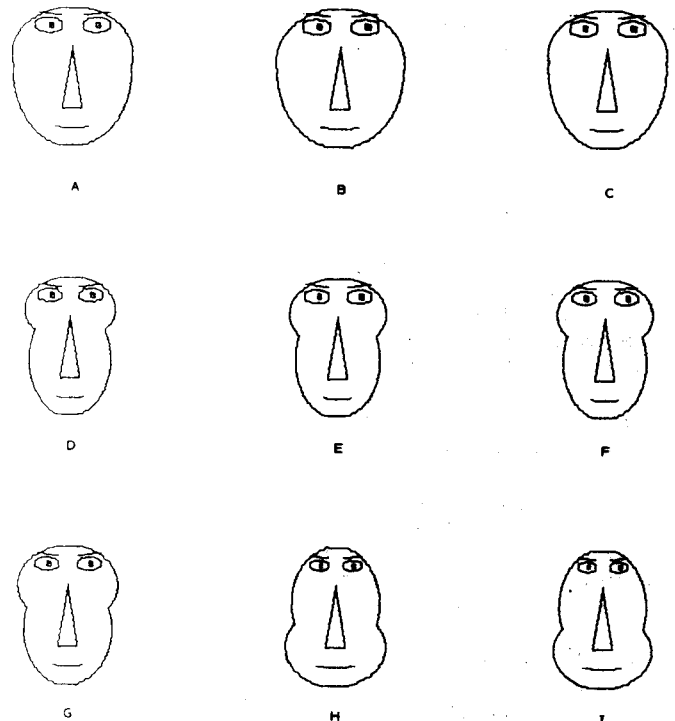
Robert J.K. Jacob
 Naval Research Laboratory
 Washington, DC 20375



A. STARS, Normalized the Same Way As Fienberg's FACES

References

Jacob, R. (1976), "Computer-produced Faces As an Iconic Display for Complex Data," unpublished PhD thesis, Johns Hopkins University, Dept. of Electrical Engineering.
 Wainer, H. (1979), "About Faces in Factor Analysis," unpublished paper.



B. FACES, Normalized the Same Way As Fienberg's STARS