McDONALD, J. A. and PEDERSEN, J. (1985a). Computing environments for data analysis. I: Introduction. *SIAM J. Sci. Statist. Comput.* **6** 1004–1012.

McDONALD, J. A. and PEDERSEN, J. (1985b). Computing environments for data analysis. II: Hardware. *SIAM J. Sci. Statist. Comput.* **6** 1013–1021.

McDONALD, J. A. and PEDERSEN, J. (1988). Computing environments for data analysis. III: Programming environments. *SIAM J. Sci. Statist. Comput.* **9** 380–400.

TIERNEY, L. (1988). XLISP-STAT: A statistical environment based on the XLISP language. Technical Report 528, School of Statistics, Univ. Minnesota.

Comment: Industrial Strength VEDA

Forrest W. Young

Multivariate visual exploratory data analysis (**VEDA**) has withstood its "test of fire": Weihs and Schmidli are the first to try multivariate **VEDA** methods in an industrial applied statistics setting, and the methods proved useful. They are to be commended for their bravery in implementing and carrying out such a project, and are to be congratulated both on their successful application and on providing us with a model paper which shows how to turn the process of data visualization into a readable and informative report.

As one of the developers of multivariate **VEDA** methods, I am, naturally, very pleased with this paper. It is exciting to see that our methods can be used, in the words of Weihs and Schmidli, by "the investigator faced with an ongoing stream of many data sets, limited time and the need for a fairly general single routine strategy," and not just by developers who are "presenting just one more method · · · (with) examples particularly fitted to demonstrate their usefulness."

My excitement stems from three aspects of the Weihs and Schmidli paper: 1) the example, 2) the confirmatory use of exploratory methods, and 3) the emphasis on the independence of the visual methods from the multivariate methods. I will discuss these points in the next three sections of this comment. My excitement is tempered somewhat, however, by one major shortcoming: When "variables can be naturally attached to more than one group, and the predictability of one group by another is of interest" (to quote the authors), then redundancy analysis (Lambert, Wildt and Durand, 1988) should be used, not canonical analysis as suggested by the authors. I will discuss this shortcoming in the fourth section of this comment. Since no plotting tools have been proposed for redundancy analysis, in the fifth section I present the **triplot** a new **VEDA** tool for redundancy analysis with certain similarities to the biplot (Gabriel, 1971), comparing it to biplots and to the authors' approach to simplification.

Forrest W. Young is Professor of Psychometrics at University of North Carolina at Chapel Hill. His mailing address is Psychometric Laboratory, CB 3270 Davie Hall, University of North Carolina, Chapel Hill, North Carolina 27599-3270.

1. ILLUSTRATION

The application used by Weihs and Schmidli to illustrate **OMEGA** involves searching for structure in multivariate data arising in the context of a major pharmaceutical, dyestuffs and agrochemical company. The data, which concern the quality of dyestuffs, are used by Weihs and Schmidli to illustrate the kind of problem for which a routine online multivariate **VEDA** strategy is required in the industrial data analysis context.

The illustration of multivariate **VEDA** methods provided by Weihs and Schmidli is exciting because it reports the process of a real visual exploratory data analysis, not just the conclusions of the process nor a "cleaned-up" mythical version of the process. The illustration shows the dead-ends, the surprises, and the excitement of **VEDA** being applied to a typically messy set of data.

One of the major strengths of the analysis is that the authors begin with Principal Components Analysis (PCA), even though the fact that the variables fall into two groups suggests immediately that Canonical Correlation Analysis (CCA) be used. They ask the rhetorical question "But is it really justified to impose variables grouping at the beginning of the analysis?" to which they answer no, saying that they might "miss something." Thus, "following this feeling" they postponed CCA until later. My own experience is that this strategy is the best to follow. PCA is, I believe, the single most powerful multivariate exploratory tool that we have, and is nearly always my first choice with a new and unfamiliar set of data. I also find it very refreshing to see a phrase such as "following this feeling" being presented without embarrassment, since informed, scientifically based feelings—hunches, if you will—are a very important aspect of **VEDA**. They note that "we were lucky" that PCA helped them