Paper 195-26

Wendy Dickinson, University of South Florida, Tampa, FL

Purpose

This study investigated the concept of multi-dimensional data display to enhance the integration and presentation of quantitative data as Chernoff type 'FACES.' Two research questions were investigated by this researcher:

1. How can FACES be modified to provide a more elegant and compelling image?
2. Can the problem of arbitrary assignment of features to variables be addressed through the development and use of a decision matrix?

The overwhelming premise of visual imagery is that of communication. Image content transcends the spoken word. For the viewer, this unspoken message can be more potent than spoken or written communication. Visual language is an effective vehicle of both context and content. Images can be visual renditions or representations of ideas, objects, dimensions, and events. Tapping the power of artistic imagery to transmit quantitative information is a natural extension of visual display. Visual data set images show patterns of response not readily apparent as a page of numbers.

Many methods of multivariate graphical display attempt to reduce the information in the data set to a relatively few variables. This reduction to two or three variables causes both a loss of original information and clouds the interpretation of results or research findings. In 1971, displaying multivariate information

was about to improve with the new ideas of Herman Chernoff.

Chernoff developed a unique graphing method to encode and display multivariable data by using iconic FACES to represent individual cases. A major advantage of the "FACES" program was the capability to store and show the values of up to 18 different variables per FACE display.

Friendly (1991) constructed an elegant SAS FACES algorithm incorporating the "Asymface" algorithm developed by Schuepbach (1989), which in turn had incorporated the original Chernoff algorithmic ideas. Friendly's FACES (1991) SAS algorithm was revised by this researcher to provide the different FACES output and results shown here.

A problem Chernoff noted in his algorithm was the arbitrary assignment of features to variables. Chernoff also suggested (but did not execute) providing a written explanation of the "FACES" for the viewer. This study addressed the problem of the arbitrary assignment of features to variables and completed the idea that was never executed by Chernoff: the providing of accompanying explanations with the visual display of information.

Procedure

Four aesthetic factors and one non-aesthetic enhancement factor were investigated in this study. As shown in Figure 2, each of the five factors utilized algorithmic strategies to investigate change. For Factor I, Variable Assignment, nonaesthetic (statistical) techniques were used to enhance facial construction. Pearson's product-moment correlation coefficients r, were calculated for each of the two research samples. The relationships among the variables were studied, and variables that were strongly correlated with each other were noted. These variable relationships, based on the statistical correlation initially calculated, formed the conceptual basis for assigning variable-to-feature in the FACES algorithm. (Stanford variable-to-feature assignments are documented in Table 1). Additionally, the variable values were transformed to z-scores to provide a constant scale for the graphing algorithm. SAS 6.12 code was written to provide a standard distribution of scores with a mean of zero and a standard deviation of one. The subsequent variable assignment was used to investigate the effects of deliberately violating this conceptual-statistical correlates approach to variable feature assignment.

This research began by examining the two multidimensional, large-scale databases. The first database was a national one, the NELS-88; while the second one examined was student achievement scores from a local Florida school district (in the form of Stanford Achievement Test Data). Each database was scrutinized by the researcher to determine the variables best suited for inclusion.

Shown in Figure 2 is the research design schematic.

FACES Family Tree
Shown in Figure 3 is a visual summary of the additive and subtractive algorithmic changes.

PARAMETER ENHANCEMENTS: FACTOR DESCRIPTIVES

Variable Assignment
conceptual - statistical correlates
deliberate violations (reordered)

Color
presence, absence

Features
presence, absence of ears
presence, absence of hair

Output
line quality (width),
presence, absence of frames

Scale
micro, macro

Figure 1. Parameter Enhancements: Factor Descriptives

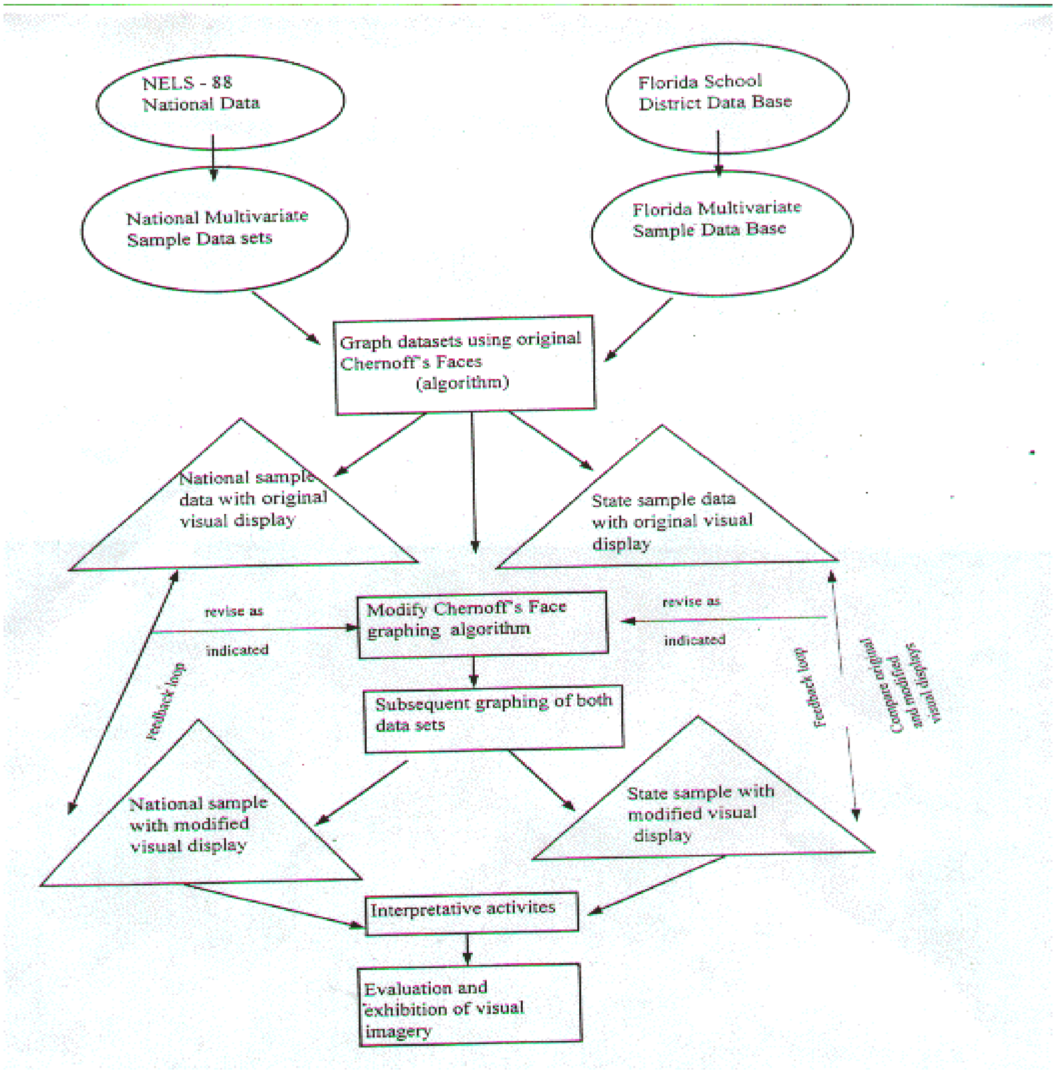

Figure 2. Research Design Schematic

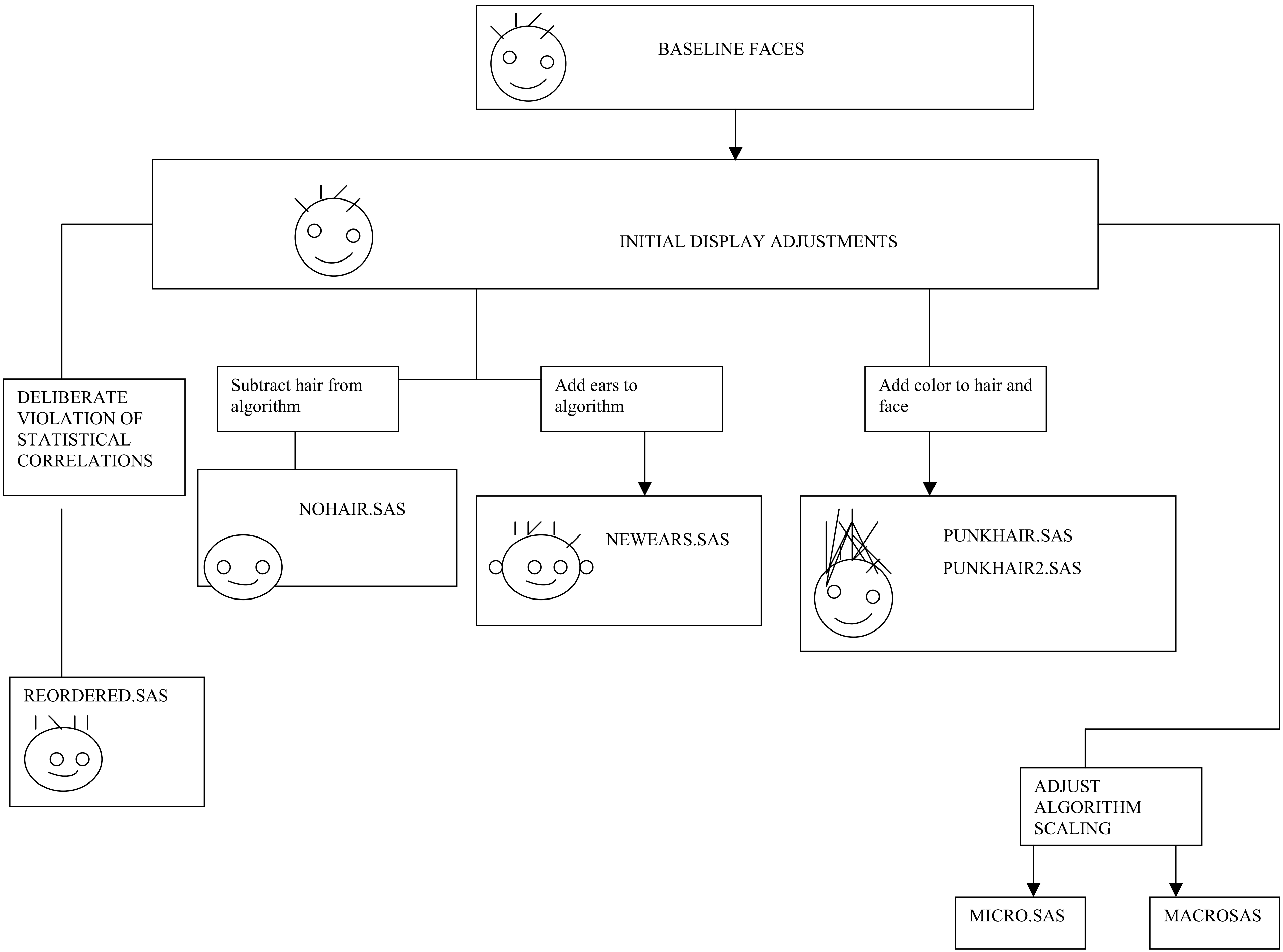

Figure 3. FACES Family Tree. Additive and Subtractive Changes

Baseline FACES: Stanford Datasets

Shown in Table 1 are the graphing algorithm specifications for the Stanford datasets. There were 250 male cases and 250 female cases for a total of 500 observations. Variable-to-feature assignments were allocated as described in Table 1. Using the calculated Mahalanobis Distance Values for each of the cases, the cases with the highest Mahalanobis Distance Value, the lowest Mahalanobis Distance Value, and the median Mahalanobis Distance value were identified and graphed for both Stanford samples.

Table 1
Stanford Baseline FACES Specifications

	Variable Assignments	
Parameter	Facial Feature	Variable
1 (EYSI)	Eye Size	Totread
2 (PUSI)	Pupil Size	Totread
3 (POPU)	Position of Pupil	Probsolv
(EYSL)	Eye Slant	Probsolv
(HPEY)	Horizontal position of eye	Mathproc
(VPEY)	Vertical position of eye	Mathproc
(CUEB)	Curvature of eyebrow	Lang
(DEEB)	Density of eyebrow	Lang
9 (HPEB)	Horizontal position of eyebrow	Prewrite
10 (VPEB)	Vertical position of eyebrow	Prewrite
11 (UPPHA)	Upper hair line	Compose
12 (LOHA)	Lower hair line	Compose
13 (FALI)	Face line	Editing
14 (DAHA)	Darkness of hair	Editing
15 (HSSL)	Hair shading slant	UseinfoA
16 (NOSE)	Nose line	UseinfoA
17 (SIMO)	Size of mouth	Basicbat
18 (CUMO)	Curvature of mouth	Basicbat
	Face Color = Black	
	Hair color = Black	

Results

Due to space considerations, only one sample of FACES are shown. For complete results, see Poster Presentation. Looking at Stanford Optimal Faces, Figure 4, we see displayed from left to right the FACES corresponding to minimum Mahalanobis distance (closest to centroid), median Mahalanobis distance (typical case), and maximum Mahalanobis distance (extreme case). Stanford female (case #87, MD value = 0.87) exhibits the most human-like proportions and structural uniformity of these displayed FACES. Eye size (EYSI) and pupil size PUSI) reflect high scores on the Total Reading subtest. Dense, well-placed eyebrows indicate high scores on both the Language subtest and the Pre-Writing subtest; resulting in increased eyebrow density (DUEB). The uniform eyebrow placement is controlled by HPEB, horizontal position of eyebrow, and VPEB, vertical position of eyebrow. This uniform placement indicates moderately high scores on the Mathematics procedures subtest.

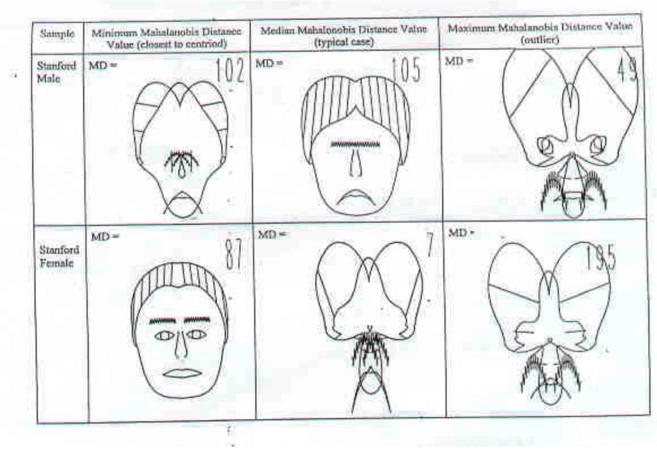

Figure 4. Stanford Optimal FACES

Components of the Optimal FACE: An Aesthetic Analysis

The application of the FACES graphing algorithm to these selected student data samples created a unique visual event: the translation of data into a meaningful visual correlate. Unlike other 2 and 3-dimensional graphical display devices, the FACES algorithm provides one summative visual image per student, rather than separate data pieces. Each student's 'FACE' is his or her own personal graphic representation combining up to 20 pieces of information in a highly recognizable, usable format. Each student's 'FACE' is a visual embodiment of his or her academic performance - a holistic summary of the individual's academic strengths and weaknesses. In the following sections, Tables 2, 3, 4, 5, 6, and 7 summarize the changes and enhancements for optimal FACES display for the aesthetic factors of FACE shape, hair, graphical output, color, and ears.

To enhance the viewer's accessibility of this graphical visual language, the following composite factors are presented to describe the optimal face. Optimal FACES are shown in Figure 4, Stanford Sample Optimal FACES.

FACE Shape

FACE shape is the most important visual attribute of the FACES composition. Conceptually, the first thing one attends to aesthetically is the overall image composition and its attendant figure - ground relationship. The shape of the FACE demarcates the boundary within the figure - ground relationship. The printed image (FACE) occupies area within space, and the FACE shape (created by variables FALI, FACE Line, and variable LOHA, lower hair line) contributes greatly to the initial impression of the audience/viewer. Data values were transformed to improve algorithmic efficiency (see Table 2).

Table 2

Transform data into Standardized Scores: Graphical FACES Display: Changes and Enhancements

Original Algorithmic Code	Modified Algorithmic Code Component	Resultant Change	Aesthetic/Operational Basis for Change
None	proc standard; mean = 0 sdt = 1 out = one; var ______; proc print; run	Transform Data into z-scores	Scaling of variable values

Hair
The hair attribute of the FACES is included for the optimal FACE because it reads as pattern. As pattern, the hair is important to the overall composition because it adds visual interest. As part of the multivariate display device, the hair functions as an indicator of inherent values from the data set. As complex versus simplistic patterns emerge in the hair, the ability to detect unusual cases is made easier. For example, a dense, overall hair pattern (variable DAHA; darkness of hair, and variable HSSL, hair shading slant) is more energetic, vibrant, and active; whereas little or no hair is simpler and less stimulating to the viewer.

Table 3

Features: Graphical FACES Display: Changes and Enhancements

Original Algorithmic Code	Modified Algorithmic Code Component	Resultant Change	Aesthetic/Operational Basis for Change
Do-hair	delete do-hair block output	Remove hair from FACES	Composition (NOHAIR.SAS)

Color

Table 4 displays the color modifications for the FACES algorithm. Color was not chosen for the optimal FACES presentation for several reasons: detection issues for color-blind individuals, color field vibrations, reproducibility concerns (both hardware and software), and archival concerns for the life of the printed FACES in a visual display (light-fastness of ink and paper).

Table 4

Color: Graphical FACES Display: Changes and Enhancements

Original Algorithmic Code	Modified Algorithmic Code Component	Resultant Change	Aesthetic/Operational Basis for Change
hcolor = black	hcolor = red	Hair color changed from black to red	investigate primary color options - punkhair.sas
fcolor = black	fcolor = blue	Face color changed from black to blue	investigate primary color options
hcolor = black	hcolor = green	Hair color changed from black to green	investigate complementary color options - punkhair2.sas
fcolor = black	fcolor = red	Face color changed from black to red	investigage complementary color options - punkhair2.sas

Ears

The Ear modification is shown in Table 5. Ears were a peripheral appendage to the FACES (Table 5). The ears did not add anything to the overall FACE presentation; rather they functioned as chart-junk. "Chart-junk" was described by Tufte as unnecessary additions to the graphic display. Chart-junk is a piece of the graphical display which can be removed without causing a loss of information.

Table 5

Features: Graphical FACES Display: Changes and Enhancements

Original Algorithmic Code	Modified Algorithmic Code Component	Resultant Change	Aesthetic/Operational Basis for Change
None	Do ear ; (see Poster for complete code)	Add ears to FACES	Composition (NEWEARS.SAS)

Scale
After investigating several scale options (BASELINE.SAS, MicroFACE.SAS, and MacroFACE.SAS), the scale of the FACES was chosen to be eight FACES represented on a page. The optimal FACES scale is the more physically manipulable size and thus invites viewer interaction (Table 7).

Table 6

Scale: Graphical FACES Display: Changes and Enhancements

Original Algorithmic Code	Modified Algorithmic Code Component	Resultant Change	Aesthetic/Operational Basis for Change
blks = 1 rows = 4 cols = 4	Blks = 2 Rows = 2 Cols = 2	Changed output from sixteen to eight FACES per page	investigate options for scale optimal.sas
blks = 1 rows = 4 cols = 4	Blks = 1 Rows = 1 Cols = 1	Changed output from sixteen FACES per page to one FACE per page (MacroFACE.SAS)	investigate options for scale macroFACE.sas
blks = 1 rows = 4 cols = 4	Blks = 4 Rows = 4 Cols = 4	Changed output from sixteen FACES to 64 FACES per page	Investigate options for scale MicroFACE.sas

Summary

The research question: How can FACES be modified to provide a more elegant and compelling image?, was addressed using the following modifications FACES were enhanced by incorporating the following modifications into the graphing algorithm:

(a) Adding a frame around each FACE - by adding a frame (box) around each graphed FACE, the image was enhanced per the Gestalt principle of whole object. The frame enhanced the spatial separation between figure (FACE) and ground (page).

(b) Switching output specifications from low to high resolution - by adjusting the output specifications from low to high resolution, the graphic output catalog is enhanced by a more definitive presence.

(c) Adjusting the scale of the graphical output to produce eight FACES per page - using the information displays obtained from scale modifications (micro.sas, macro.sas, initial display adjustments), an optimal size of FACES was selected. This FACE size was the best choice based on the considerations outlined previously: reproducibility, featural definition, and viewer interaction.

(d) Increasing the line factor width of the output - increasing the line factor width resulted in a more defined, darker image which aids both in viewing and reproduction of imagery.

(e) Displaying the resultant FACES in black-and-white only - by displaying FACES only in black-and-white, the researcher avoids visual vibration (complementary colors), enhances clarity of presentation, enhances ease of photo-reproduction duplication), and avoids the difficulties a color-blind person would encounter with colored imagery.

(f) Displaying the FACES with hair- the hair with the FACE aids in ease of identification of outliers and completes the image visually. Using statistical correlations to assign variables to features - using the conceptual statistical correlates provided an empirical solution to the arbitrary assignment.

(g) Transforming the data into standard scores and scaling parameter values to ensure full graphic capability - transforming the data into standard scores and scaling the parameter values resulted in complete algorithmic FACES presentations.

For the Variable Assignment Decision Matrix, please see the poster presentation.

Concluding Remarks

Tufte (1990) asked, "How are we to represent the rich visual world of experience and measurement on mere Flatland?" This study used criteria of elegance, simplicity,and clarity to determine the most compelling images produced with SAS 6.12 algorithmic analysis. Escaping from Flatland requires innovative methods and unique problem resolution. This study investigated and provided possible solutions for "escaping Flatland."

REFERENCES

Abbot, E. A. (1953). Flatland: A romance of many dimensions, with illustrations by the author. NY, New York: Dover.

Bower, G. H. & Hilgard, E. R. (1981). Theories of learning (5th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Braham, R. (1995). Math and visualization: New tools, new frontiers. IEEE Spectrum, 19-36.

Bruckner, L. A. (1979). The interactive use of computer drawn faces to study multidimensional data. Los Alamos, NM: National Technical Information Service.

Bruckner, L. (1976). The looks of some companies involved in offshore oil and gas leaks. Paper presented at the American Statistical Association, Boston, MA.

Chernoff, H. (1978). Graphical representations as a discipline, in Graphical Representation of Multivariate Data, 1-11. NY: Academic Press.

Chernoff, H. (June 1973). The use of faces to represent points in k-dimensional space graphically. Journal of the American Statistical Association, 68(342).

Chernoff, H., Rizvi, M. H. (1975). Effect on classification error of random permutations of features in representing multivariate data by faces. Journal of the American Statistical Association, 70, 548-554.

Flury, B., & Riedwyl, H. (1981). Graphical representation of multivariate data by means of asymmetrical faces. Journal of the American Statistical Association, 76(376), 757-765.

Katzer, J., Cook, K. H., & Crouch, W. W. (1978). Evaluating information. Reading, MA: Addison-Wesley Publishing Co., Inc.

Koffke, K. (1963). Principles of gestalt psychology. New York, NY: Harcourt, Brace & World.

Kohler, W. (1969). The task of gestalt psychology. Princeton, NJ: Princeton University Press.

Rencher, A. C. (1995). Methods of Multivariate Analysis. NY: John Wiley & Sons.

Tufte, E. R. (1992). The visual display of quantitative information. Cheshire, CT: Graphics Press.

Tufte, E. R. (1990). Envisioning information. Cheshire, CT: Graphics Press.

Tufte, E. R. (1974). Data analysis for politics and policy. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Tukey, J. W. (1974). Mathematics and the picturing of data. Proceedings of the International Congress of Mathematics, Vancouver, Canada.

Tukey, J. W. (1972). Some graphic and semigraphic displays. Arnes, IA.

Wainer, H. (January, 1979). Some supporting evidence for accurate multivariate perceptions with Chernoff Faces, Project 547. Washington, D.C.: Bureau of Social Science Research, Inc.

Wainer, Howard, & Thissen, D. (1987). Graphical data analysis, pp. 391-457. City, State: Publisher.

Wainer, H. (1997). Some multivariate displays for NAEP results. Psychological Methods, 2(1), 34-63.

Wainer, H. (March, 1995). A study of display methods for NAEP results. Technical Report, 95(1), 51.

Wainer, H. (February, 1992). Understanding graphs and tables. Educational Researcher, 21(1), 14-23.

Wainer, H. (November, 1982). Multivariate display for quipus to faces. Bureau of Social Science Research, Washington, D.C. Princeton, NJ: Educational Testing Service.

Wang, P. C. (1978). Application of graphical multivariate techniques in policy sciences. New York, NY: Academic Press.

Wang, P. C. (Ed.) (1978). Graphical representation of multivariate data. New York, NY: Academic Press.

Whiting, R. (1999). Data made visible. Information week, 722, 108.

CONTACT INFORMATION

Your comments and questions are valued and encouraged.
Contact the author at:

Wendy B. Dickinson
University of South Florida
4202 E. Fowler Ave. EDU 162
Tampa, FL 33620
(813) 974-3220 statgirl@aol.com