THE USE OF FACES TO REPRESENT POINTS IN n-DIMENSIONAL SPACE GRAPHICALLY

BY

HERMAN CHERNOFF

TECHNICAL REPORT NO. 71
DECEMBER 27, 1971

PREPARED UNDER CONTRACT
N00014-67-A-0112-0051 (NR-042-993)
FOR THE OFFICE OF NAVAL RESEARCH

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) DEPARTMENT OF STATISTICS STANFORD UNIVERSITY STANFORD, CALIF.	2a. REPORT SECURITY CLASSIFICATION
	2b. GROUP

3. REPORT TITLE
THE USE OF FACES TO REPRESENT POINTS IN n-DIMENSIONAL SPACE GRAPHICALLY

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
TECHNICAL REPORT

5. AUTHOR(S) (Last name, first name, initial)
CHERNOFF, Herman

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
December 27, 1971	45	6

8a. CONTRACT OR GRANT NO. N00014-67-A-0112-0051	9a. ORIGINATOR'S REPORT NUMBER(S) #71
b. PROJECT NO. NR-042-993	
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	

10. AVAILABILITY/LIMITATION NOTICES
Unlimited. Reproduction in whole or in part is permitted for any purpose of the United States Government

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Statistics & Probability Program Office of Naval Research Arlington, Virginia 22217

13. ABSTRACT.

A new method of representing multivariate data graphically is described here. Briefly, it consists of representing a point in k-dimensional space by a picture of a face whose characteristics are determined by the position of the point. A sample of points in k-dimensional space is represented by a collection of faces. Two illustrations are sketched briefly. Detailed documentation, including the data for the illustrative examples and the method of generating the faces, is included.

DD FORM 1 JAN 64 1473

UNCLASSIFIED
Security Classification

UNCLASSIFIED

Security Classification

14. KEY WORDS	LINK A ROLE	LINK A WT	LINK B ROLE	LINK B WT	LINK C ROLE	LINK C WT
cluster analysis faces graphical representation multivariate data profile						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year, or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through ______________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through ______________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through ______________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, or *(U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, roles, and weights is optional.

DD FORM 1 JAN 64 1473 (BACK)

Unclassified

Security Classification

THE USE OF FACES TO REPRESENT POINTS IN n-DIMENSIONAL SPACE GRAPHICALLY

by

HERMAN CHERNOFF

TECHNICAL REPORT NO. 71

December 27, 1971

PREPARED UNDER CONTRACT

N00014-67-A-0112-0051 (NR-042-993)

FOR THE OFFICE OF NAVAL RESEARCH

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

THE USE OF FACES TO REPRESENT POINTS IN n-DIMENSIONAL SPACE GRAPHICALLY

by

Herman Chernoff

1. Introduction

Graphical representations serve to communicate essential information conveniently and effectively. They are also useful in exploratory work with data. In particular, a scatter diagram is a powerful device for indicating at a glance the essential relationships between two variables which have a bivariate distribution. In a field like cluster analysis, where the concept of cluster is not clear-cut, the actual data may have a profound effect on what the investigator would choose to call a cluster. In the bivariate case the scatter diagram can be used effectively to help decide which concept to use. For some problems the inadequacy of the classical linear techniques of normal multivariate analysis may be clearly revealed by graphs which can be used to suggest suitable transformations or other modifications of linear theory.

When dealing with multivariate data involving more than two variables, the scatter diagram can be used only in the limited form where two variables are studied at a time. This is unwieldy when the number of variables involved is large. Moreover, subtle relations or effects which require the simultaneous consideration of more than two variables may go undetected when this approach is used.

A new method of representing multivariate data graphically is described here. Briefly, it consists of representing a point in k-dimensional space by a picture of a face whose characteristics are determined by the position of the point. A sample of points in k-dimensional space is represented by a collection of faces.

In the next section, two illustrations are sketched briefly. In one of these where the investigator was interested in a cluster analysis, his task was merely to group together those faces which resemble each other. In the second, where the investigator was interested in detecting time points where a multivariate stochastic process changed character, he had to look at the sequence of faces corresponding to successive points in time to locate the places where the faces change character.

Following sections discuss the potential advantage of this graphical method over that of looking at numerical data and consider some alternative approaches to and predecessors of this method. Detailed documentation, including the data for the illustrative examples and the method of generating the faces, is contained in the appendix.

2. Illustrations

We present two examples illustrating this representation.

Example 1. Fossil Data

Eight measurements were made on each of 88 nummulited specimens from the Eocene Yellow Limestone Formation of northwestern Jamaica. Two measurements thought to be age-dependent were discarded. One specimen (Number 34) was rejected because of a permutation in an early copy of the measurements for that specimen which cast doubt upon its accuracy. The data and definition of the measurements appear in Table 2a of Appendix A4. The 87 faces corresponding to the 87 remaining specimens are presented in sequential order as indicated in Figure 1a. This order was selected after the data and had been grouped into three clusters.

The number at the bottom and left of each face is a randomly selected code number. Because the data were handled in two subgroups, these code numbers are repeated twice, but half were marked with a cross. The i.d. numbers were added for publication. It is immediately obvious how these faces divide into three distinct clusters. This division is obvious partly because of the special arrangement of the faces. When copies were made, separated and mixed up, and then given to people to cluster visually, these people selected the same clusters. On several occasions there were one or two discrepancies. The people, having the code number but not the sequence number, had no way of knowing except through the faces what grouping was expected.

A follow-up attempt to separate the large cluster of the first 40 faces (1-41 with 34 omitted) into subclusters seemed to be difficult, and the results of various individuals were inconsistent with one another. Since the ranges of the variables in the first 40 specimens were smaller than for the 87 specimens, it seemed reasonable to magnify the effects of variation by renormalizing the data according to the ranges in the first 40 specimens. A new set of faces was produced and is presented in Figure 1b. I clustered these visually. The groups were

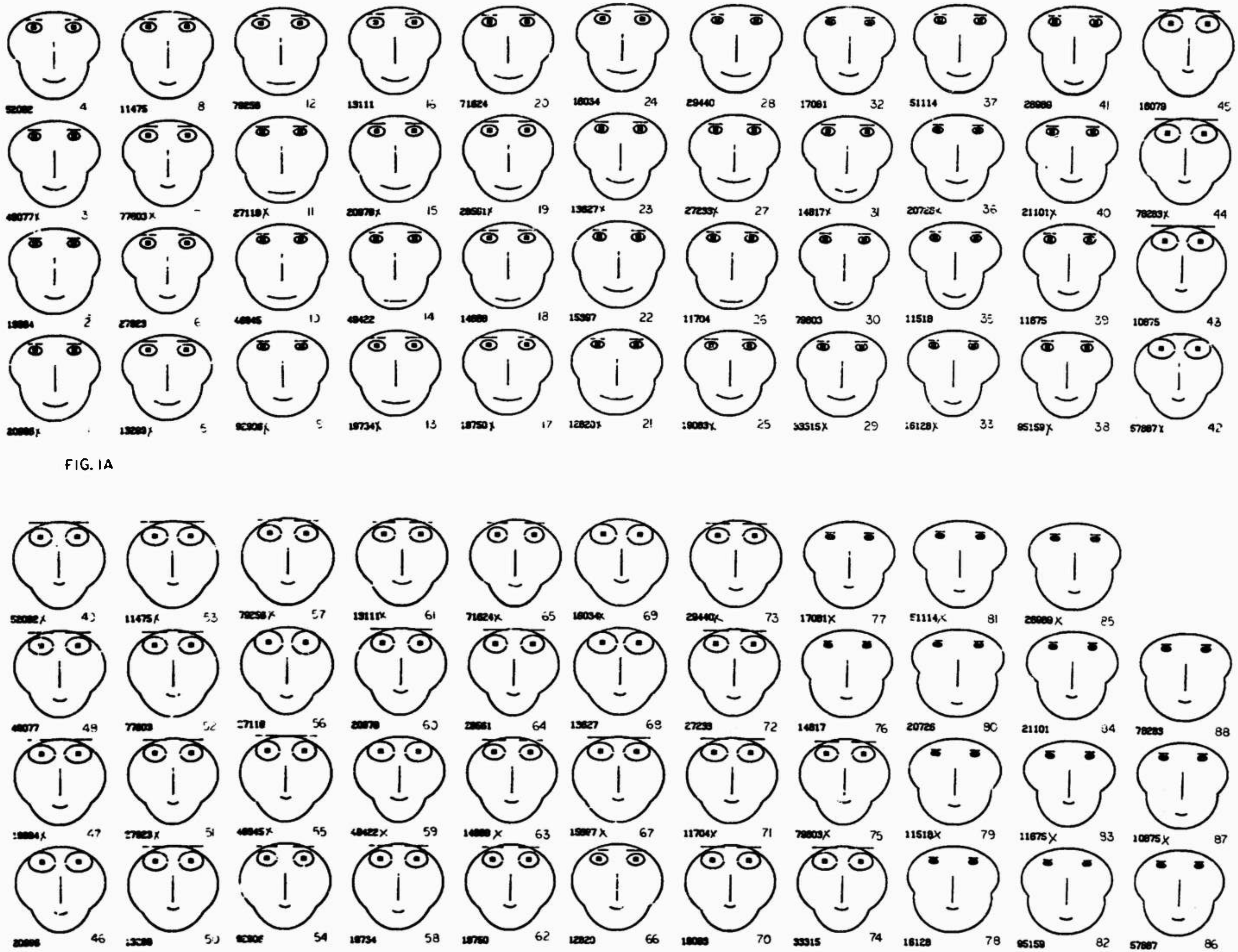


FIG. 1A

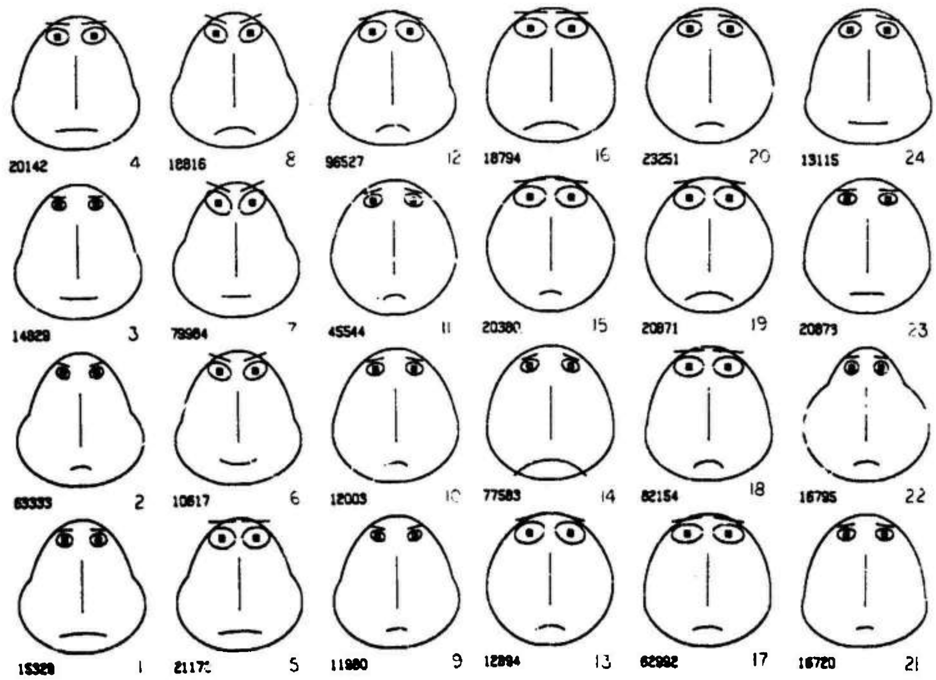

4
8
12
16
20
24
3
7
11
15
19
23
2
6
10
14
18
22
1
5
9
13
17
21

FIG. 1B

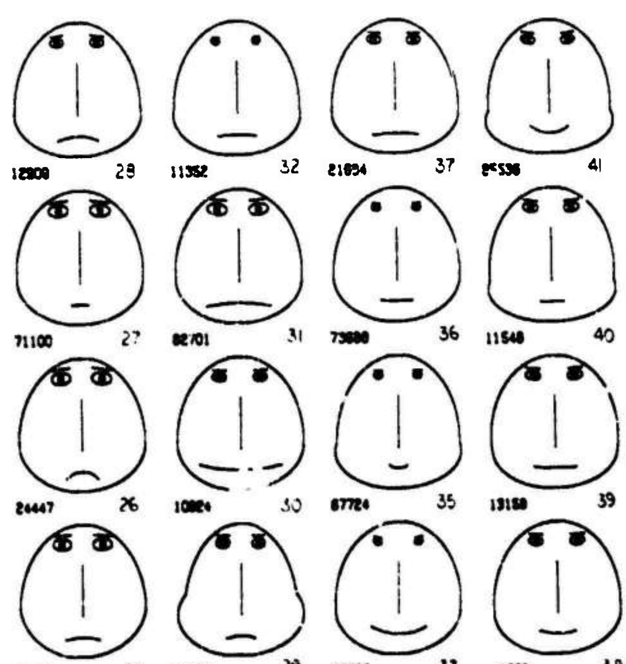
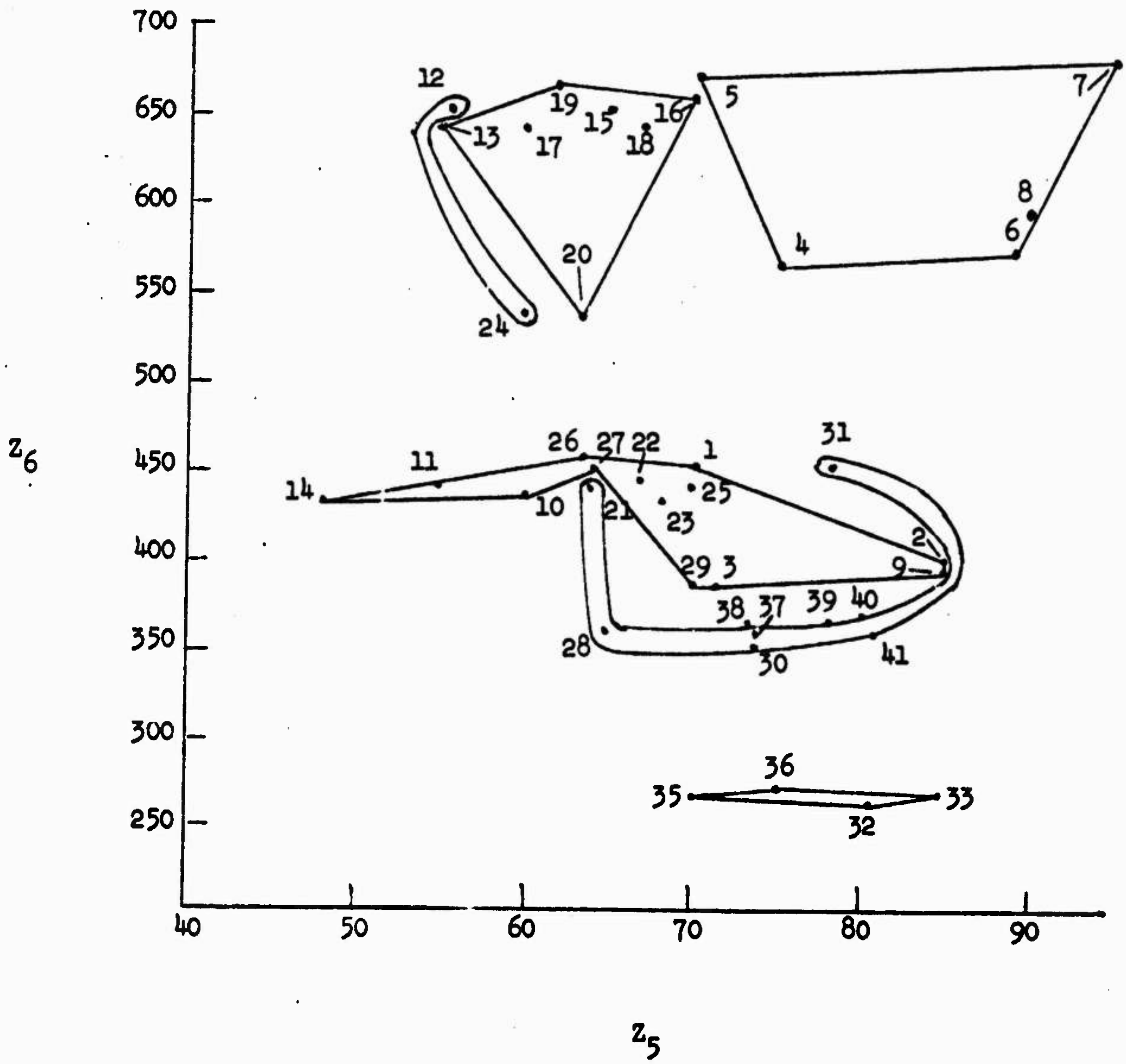

28
32
37
41
27
31
36
40
26
30
35
39
25
29
33
38

Figure 1C



700
650
600
550
500
450
400
350
300
250
40
50
60
70
80
90
z_6
z_5
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
35
36
37
38
39
40
41

I: (1,2,3,9,22,29)
II: (4,5,6,7,8)
III: (10,11,14,23,25,26,27)
IVa: (13,15,16,17,18,19,20)
IVb: (12,24)
V: (21,28,30,31,37,38,39,40,41)
VI: (32,33,35,36)

where IVb seemed to be similar to IVa but slightly different. Professor Switzer also clustered them visually. He obtained:

Ia: (1,2,3,9,22,29)
Ib: (4,5,12,24)
Ic: (13,15,16,17,18,19)
II: (6,7,8)
III: (10,11,14,20,21,23,25,26,27,28,31)
IV: (30,32,33,35,36,37,38,39,40,41)

which, though not in complete agreement with my groups, has substantial similarity. These grouping attempts were more ambitious than is ordinarily necessary for one can easily choose to leave peculiar cases out of the groupings. Furthermore these lists in numerical order do not indicate which specimens were obviously members of a group and which were regarded as borderline.

Finally, in connection with this example, a graph of (Z_5, Z_6) for these specimens is presented since these variables seemed important in the set of 87 specimens. See Figure 1C.

<u>Example 2. Geological Data</u>

Mineral analysis data from a 4,500-foot core drilled from a Colorado mountainside yielded 12 variables. These represent assays of 7 mineral

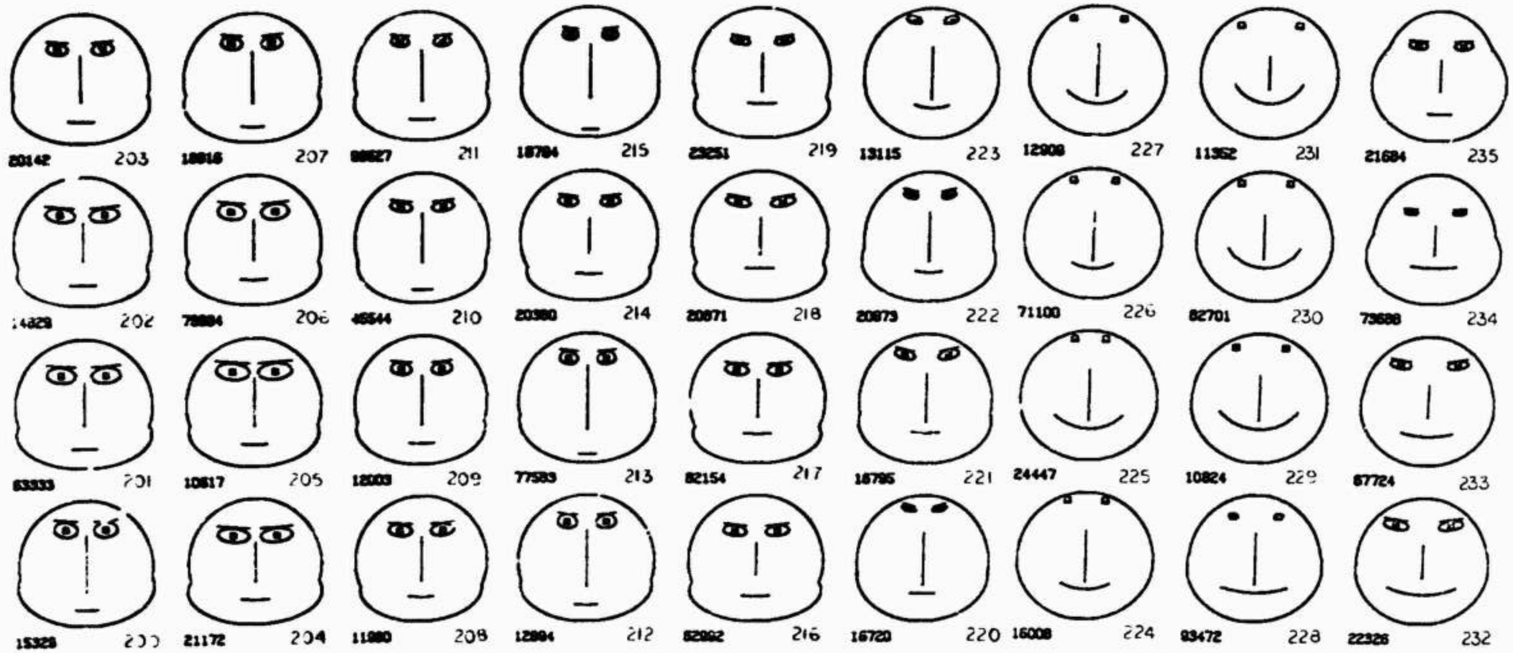
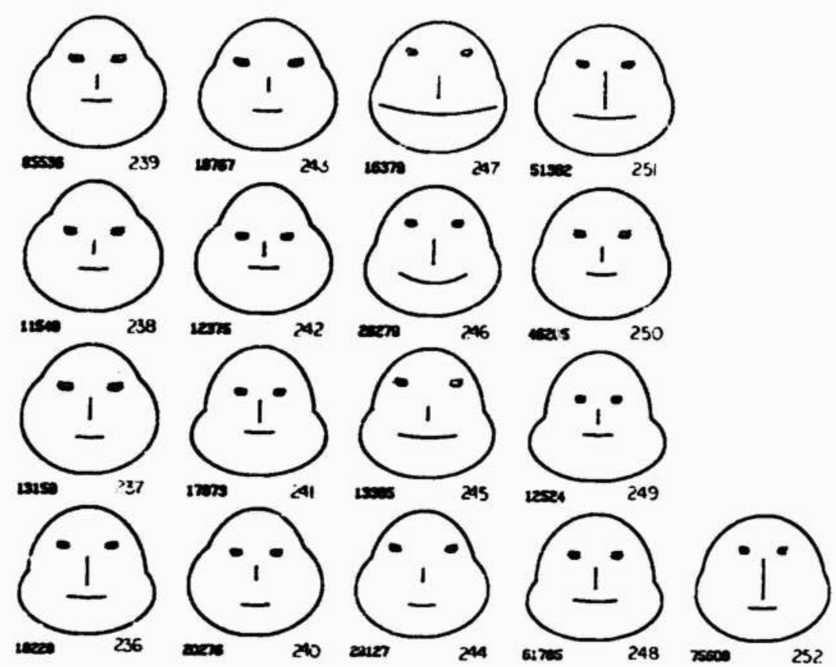


FIG. 2A



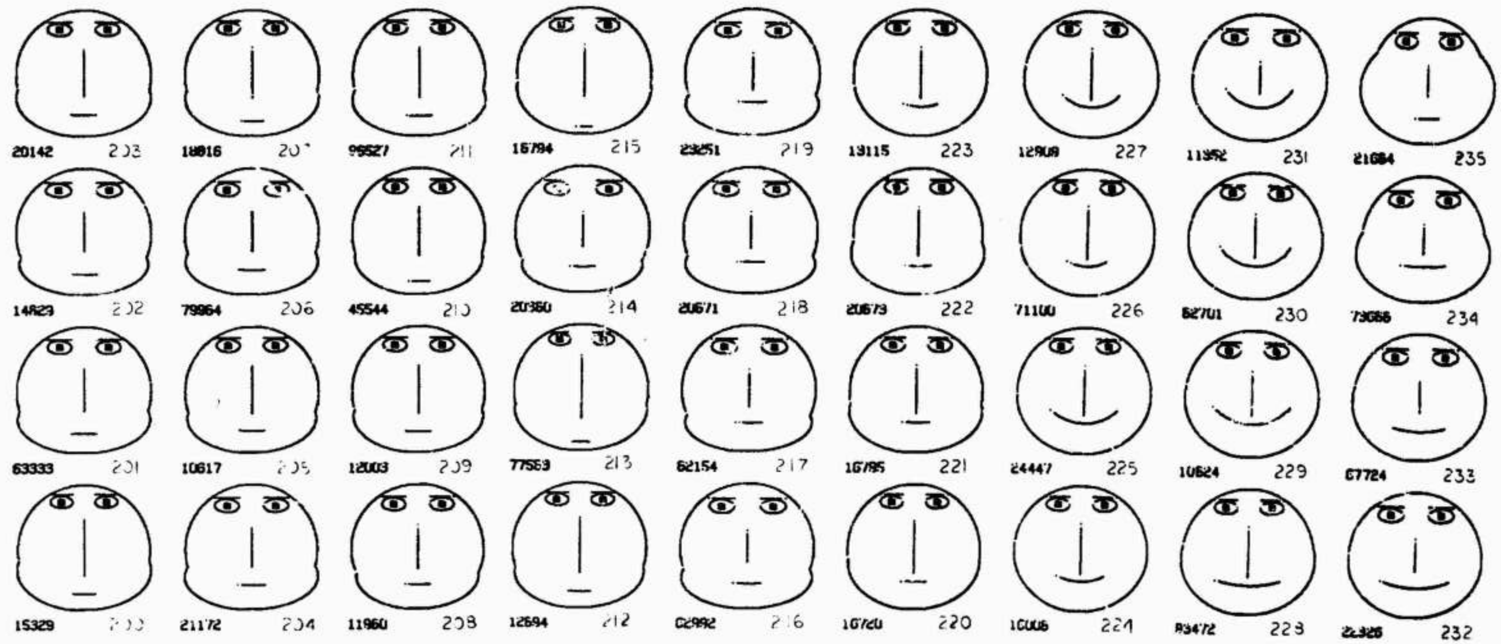
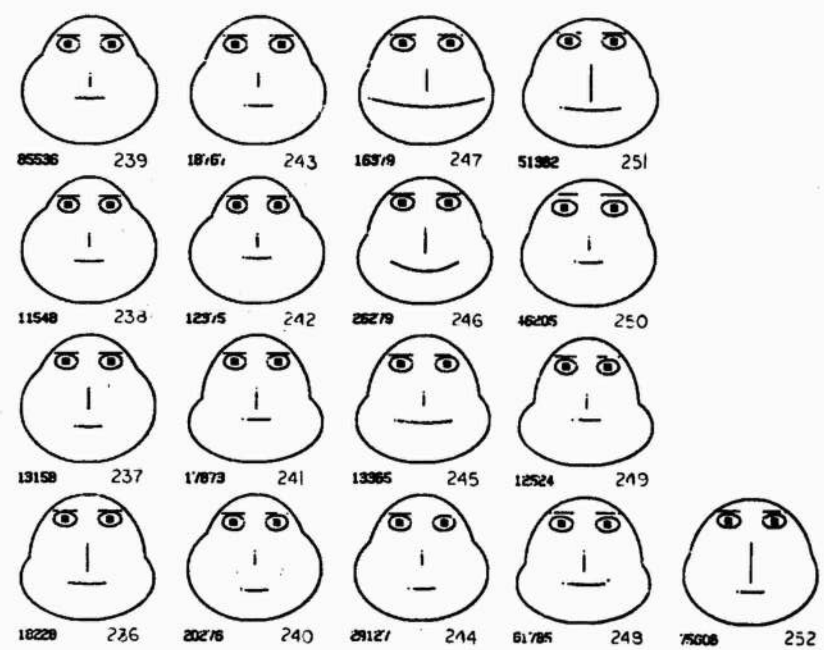


FIG. 2B



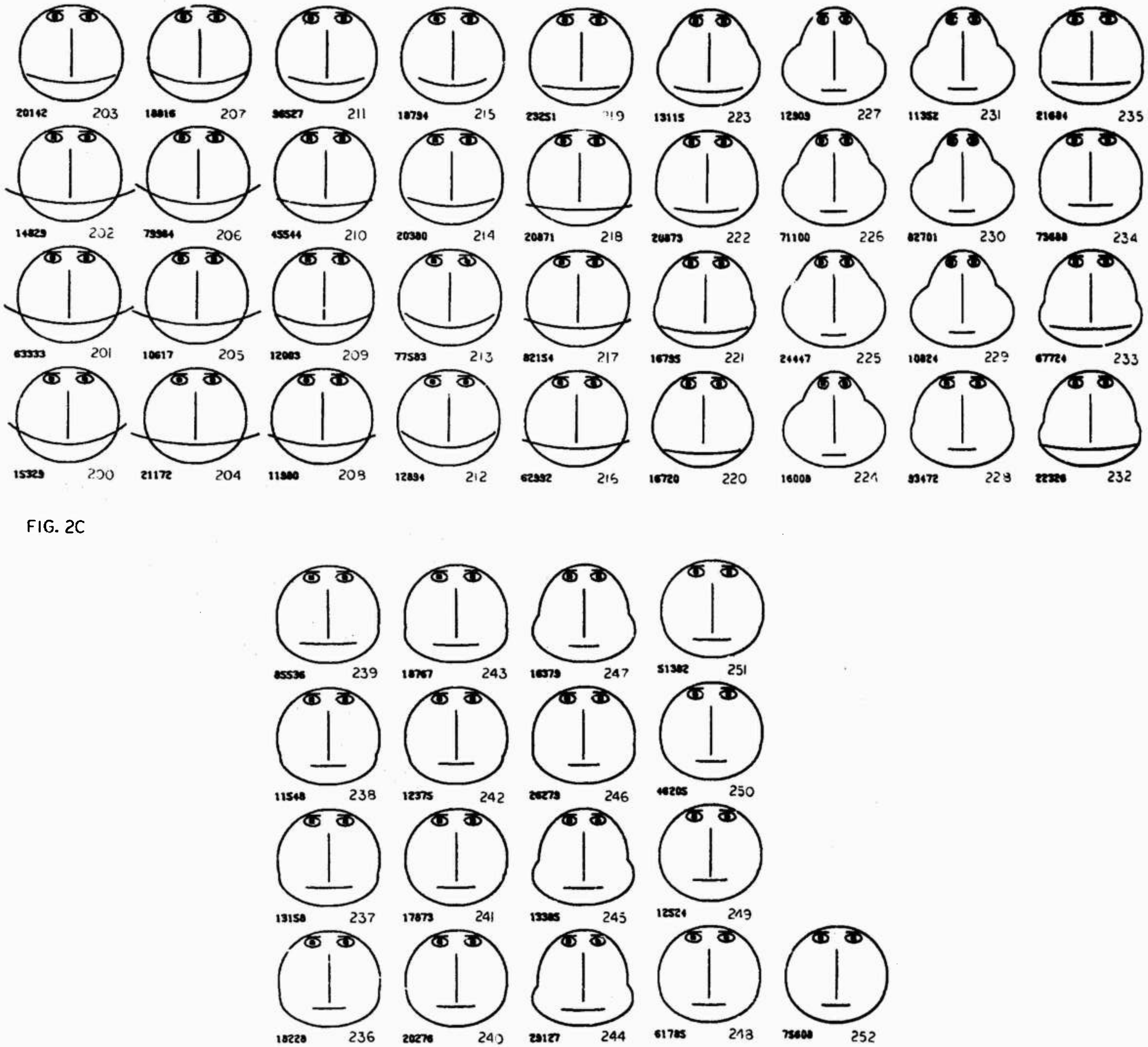

203
207
211
215
219
223
227
231
235
202
206
210
214
218
222
226
230
234
201
205
209
213
217
221
225
229
233
200
204
208
212
216
220
224
228
232
239
243
247
251
238
242
246
250
237
241
245
249
236
240
244
248
252

FIG. 2C

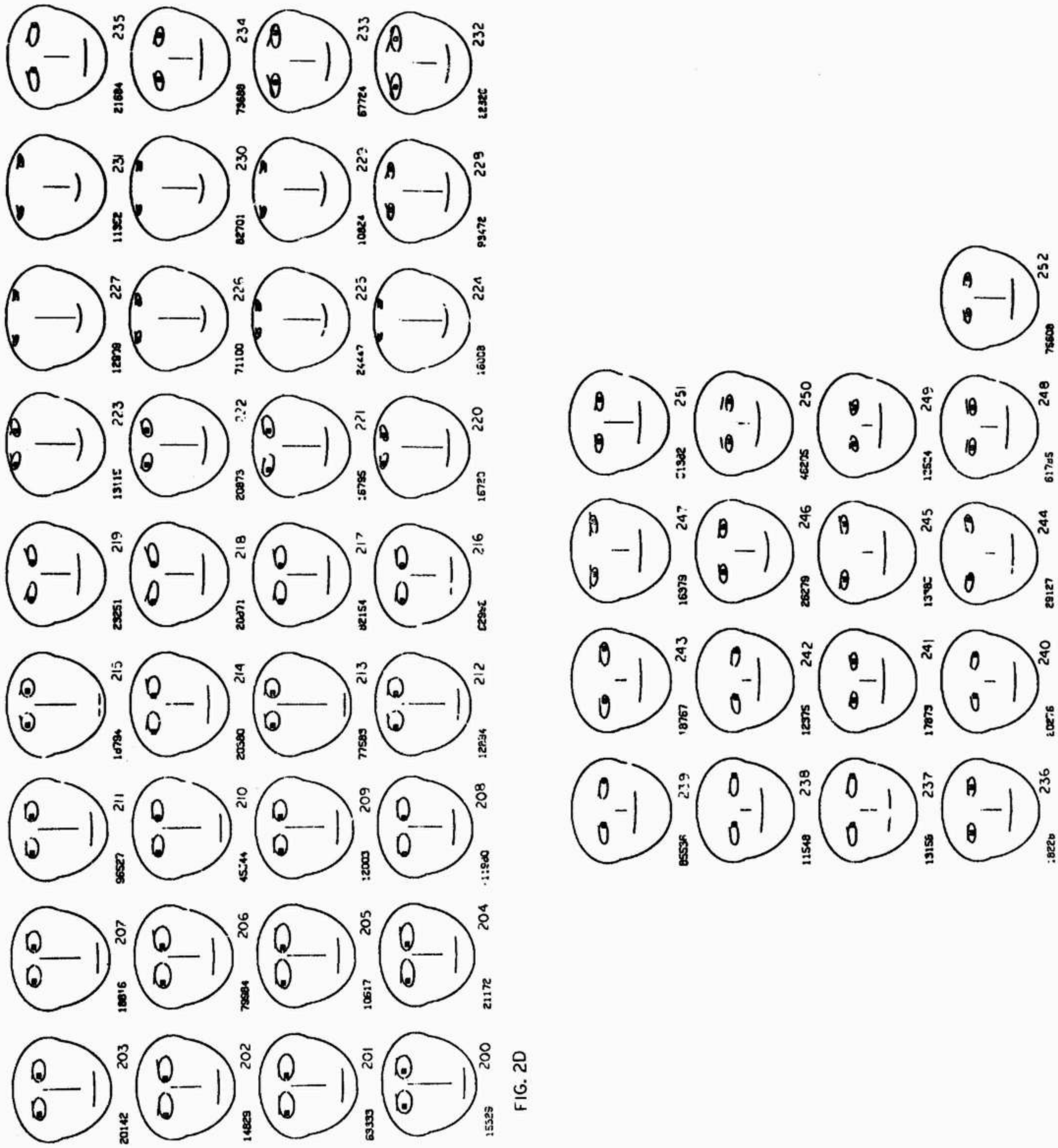

200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252

FIG. 2D

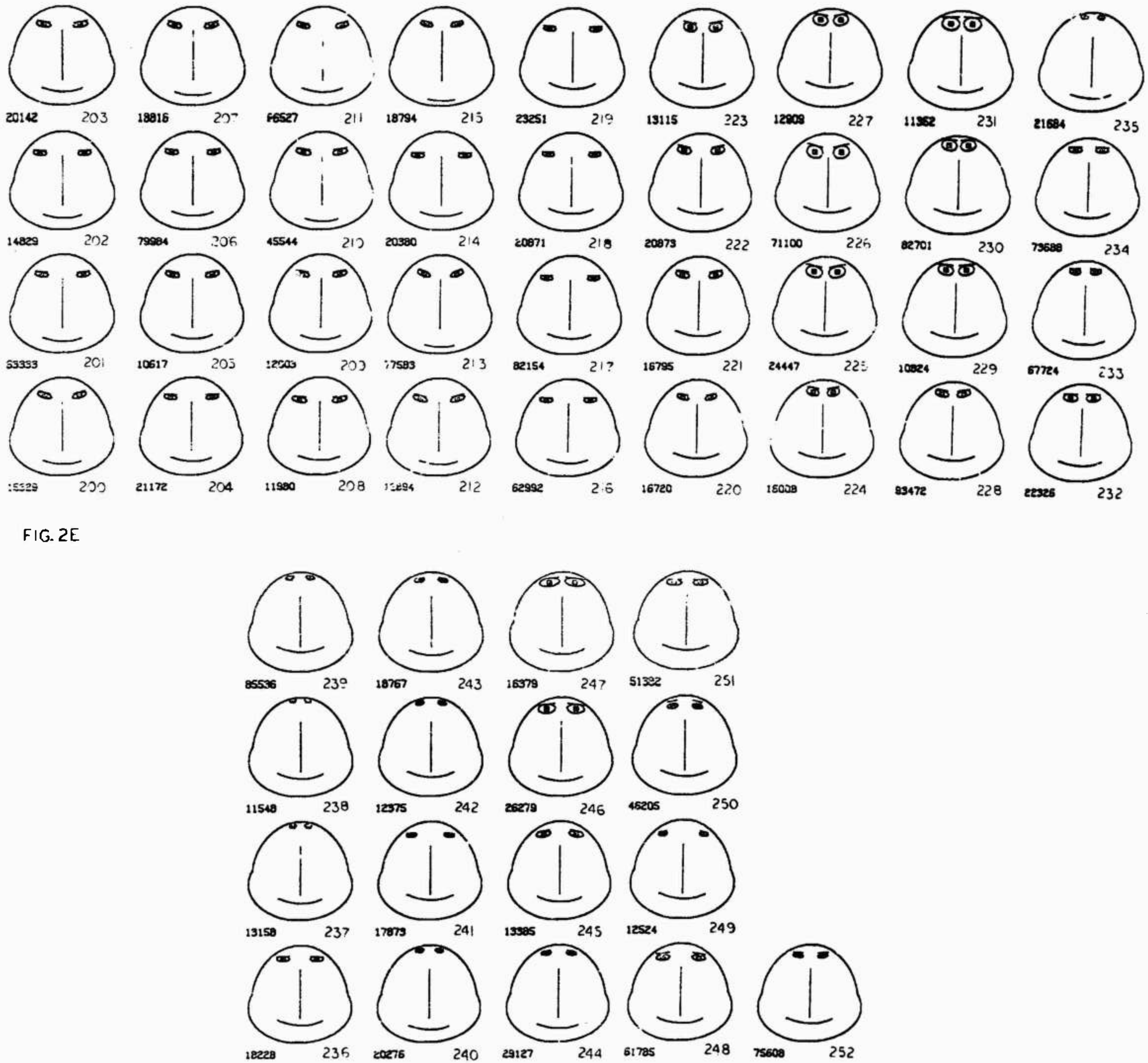

20142 203
18816 207
211
18794 215
23251 219
13115 223
12909 227
11362 231
21684 235
14829 202
79984 206
45544 210
20380 214
20871 218
20873 222
71100 226
82701 230
73688 234
53333 201
10617 205
209
17583 213
82154 217
16795 221
24447 225
10824 229
67724 233
200
21172 204
11980 208
212
62992 216
16720 220
16008 224
93472 228
22326 232
85536 239
18767 243
16379 247
51382 251
11548 238
12375 242
26279 246
46205 250
13158 237
17873 241
13385 245
12524 249
18228 236
20276 240
29127 244
61785 248
75608 252

FIG. 2E

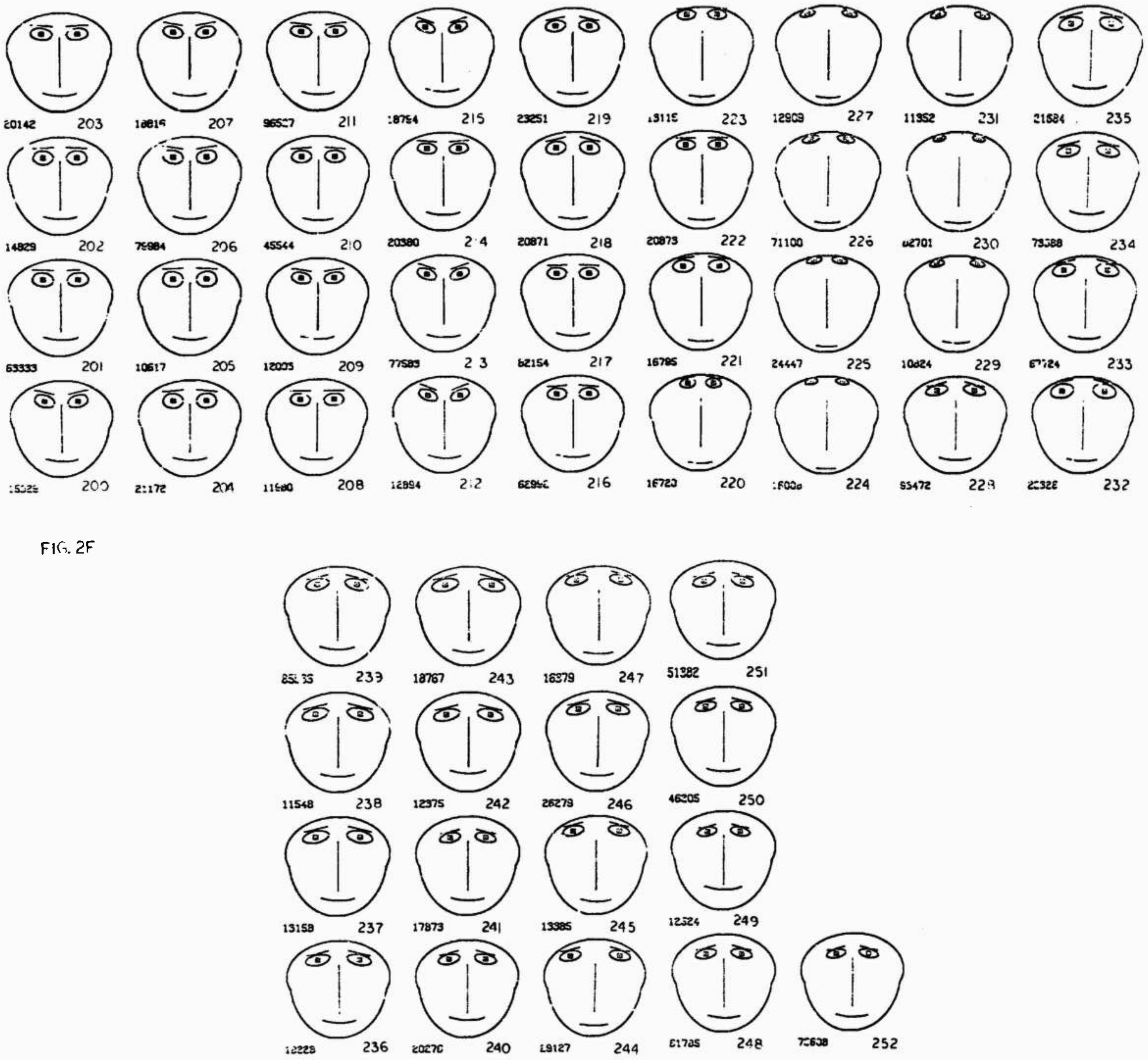

203 207 211 215 219 223 227 231 235
202 206 210 218 222 226 230 234
201 205 209 217 221 225 229 233
200 204 208 216 220 224 232
239 243 247 251
238 242 246 250
237 241 245 249
236 240 244 248 252

FIG. 2F

contents by one method and repeated assays of 5 of these by a second method. These 12 variables were observed on each of 53 equally spaced specimens along the core and are presented in Table 2a of Appendix A4.

The 53 faces obtained are shown in Figure 2a in the sequence as marked. They clearly indicate the sequence number where certain critical changes take place. One substantial change begins to take place after the 20th specimen, and those from 25 to 32 are quite distinct from the others. Another substantial change evolves from speciments 32 to 35. Particularly characteristic of the group from 25 to 32 are the tiny and high eyes, round face, broad smile, with mouth close to the relatively long nose. The group from 36 to 53 are characterized by a different constellation of special features, suggesting that a traditional linear analysis of this 12-dimensional time series may disguise some of the phenomena clearly observable.

In a casually designed experiment to determine (1) whether the large number of variables used interfered with comprehension, and (2) whether certain features had more impact than others, two additional sequences of faces were run. In Figure 2b, only the first seven variables were used. In Figure 2c, only the last 5 variables were used. These variables, which mainly controlled the eyes in Figure 2a, were made to control the face and mouth. A mistake gave the mouth too large a range, resulting in some peculiar idiosyncracies. The results seemed to indicate that the additional variables add richness to the picture and seem to help the viewer. This cannot be regarded as a serious test, especially since the last 5 variables are supposed to measure 5 of the quantities in the first 7.

Mr. Elliott, the geologist who provided me with the data, seemed to feel that the shape of face carried the essential information and that there was an element of luck in the particular choice of facial parameters selected to be controlled by the variables. He was challenged to make selections which he felt would be least informative. This resulted in Figures 2d, 2e and 2f for the 12, first 7 and last 5 variables respectively.

3. Potential Advantages

Graphical representations have many uses. These include (1) enhancing the user's ability to detect and comprehend important phenomena, (2) serving as a mnemonic device for remembering major conclusions, (3) communicating major conclusions to others, and (4) providing the facility of doing relatively accurate calculations informally. The representation by faces seems to have potential in the first two of these uses.

People grow up studying and reacting to faces all of the time. Small and barely measurable differences are easily detected and evoke emotional reactions from a long catalogue buried in the memory. Relatively large differences go unnoticed in circumstances where they are not important. This implies that the human mind subconsciously operates as a high-speed computer, filtering out insignificant visual phenomena and focusing on the potentially important. Particularly valuable is this flexibility in disregarding non-informative data and searching for useful information. It is this flexibility which is lacking in canned computer programs.

Moreover, this ability is great when applied to the study of faces. Experience with caricatures and cartoons would seem to indicate that the need for realistic faces on pictures is not great and that lack of realism is compensated, at least in part, by the ability to caricaturize.

The ability to relate faces to emotional reactions seems to carry a mnemonic advantage. For example, in looking at the numerical data from the geological problem, major changes in individual variables are readily apparent. The author found that when studying these numerical data visually with no background in the scientific problem, many changes were observed but attention would be distracted quickly by other effects. After a substantial time, a confusion of reactions remained with little useful memory. Certain major characteristics of the faces are instantly observed and easily remembered in terms of emotions and appearance. Finer details and correlations become apparent after studying the faces for a time. The awareness of these does not drive out of mind the original major impressions.

I would anticipate that the faces would have relatively little usefulness as a communication device. If results from a study of data were translated from the data to the faces, then the mnemonic advantages of the faces could conceivably make it desirable to use faces to communicate a relatively large assortment of results of varying degrees of importance.

Anyone who uses graph paper to analyze data is aware of how precise one can be with rough drawings which are strategically arranged. It would seem that the faces should not be expected to be useful except in the grossest types of calculation.

4. Alternative Representations

One is led to ask two questions. First, if this simple idea is so good, why wasn't it thought of before? Second, what alternative representations are there for points in high-dimensional space?

Introspection would suggest that this idea must have been considered before in a simpler form. However, the effective application in this form would require a computer technology which has only recently become available. Thus it is unlikely to have been used in this or similar form in spite of the growing need for a useful representation.

Several more primitive versions have come to my attention. Anderson [1] developed a method of using "glyphs", which are circles of fixed radius with rays of various lengths and directions extending from the boundary. The length of the ray represents the value of a variable. Pickett and White [4] used triangles which represent 4 variables (the three lengths of the sides and the orientation[1]). Both the glyphs and triangles can raise the dimensionality by 2 by locating the center on a point in two-dimensional space. I have some memory of being told of a scheme to convert cardiograms or brain waves to sound in the hope that the human processing of sound would be more revealing than looking at graphs. This idea seems interesting, but no follow-up has come to my attention.

Several alternative representations have been considered. The most standard is the use of profiles. Here one represents a point in k-dimen-

---

[1] I have felt greatly indebted to Pickett for many conversations we had in which he emphasized how the human ability to process subconsciously large amounts of information of textures is fundamental to the ability to locomote and, indeed, to exist. After having developed the faces, I noticed his paper containing the triangle representation and realized that I had seen it before but had not paid special attention to it in the form presented.

sional space by a series of k bars at heights corresponding to the values of the variables. It would seem desirable to standardize each variable so that the ranges either go from 0 to 1 or center about the mean. In some variations the bars are replaced by a polygonal line.

A relatively novel variation of the profile method is one due to Daetz [3] where a circle is drawn and along k equally spaced rays from the center, points are marked whose distance from the circumference are equal to standardized distances from the means of the k variables. These points are connected to form a polygon.

The polygons resulting from this variation seem to be more readily translatable to human experience than the simpler profiles. They assume "meaningful" shapes, and the tendency to lean in certain directions has mnemonic force.

A new technique of Andrews [2] consists of generating a Fourier Series of the form

$$f(t) = \frac{x_1}{\sqrt{2}} + x_2 \cos t + x_3 \sin t + x_3 \cos 2t + \cdots$$

where the x_i are the observed variables. This method has the interesting property that if x generates f, and y generates g, then

$$\int_0^1 [f(t)-g(t)]^2 dt = \sum_{i=1}^{k} (x_i - y_i)^2 ,$$

suggesting that the method could be useful for expressing moderately refined calculations relevant to linear analysis. Andrews has applied this method using the principal components, in place of the original observations, for the x_i. In the normal multivariate model, these x_i would be independent, and the distances in the above expression would be quite meaningful. A value of t which consistently and widely separates the $f(t)$ of two classes of observations provides an effective linear function for discriminating between the two classes.

It seems reasonable to conjecture that one may, in the spirit of Daetz, achieve more suggestive curves by plotting $(f(t) + C, t)$ in polar coordinates.

5. Summary

The use of the face representation provides a promising approach for a first look at multivariate data which is effective in revealing rather complex relations not always visible from simple correlations based on two-dimensional linear theories. It can be used to aid in cluster analysis, discrimination analysis, and to detect substantial changes in time series.

The study of faces does not seem to become more difficult as the number of variables increases. Example 2 indicated that the information content transmitted seems to become richer as the number of variables increases. At this point, one can treat up to 18 variables[1], but it would be relatively easy to increase that number by adding other features such as ears, hair, facial lines, and even possibly by taking pairs of faces.

This approach is an amusing reversal of a common one in artificial intelligence. Instead of using machines to discriminate between human

[1] We shall note in the appendix that the normalization of the width and length of the faces almost eliminates two of these variables.

faces by reducing them to numbers, we discriminate between numbers by using the machine to do the brute labor of drawing faces and leaving the intelligence to the humans, who are still more flexible and clever.

One question frequently asked is whether some features are more informative than others. The individuals who worked on Example 1 felt that one only looked at eyes. Elliott was convinced that only the shapes of the head are relevant. In my opinion, the human will tend to concentrate on what is important in the data. However, this question requires serious study. At present an experiment is under way to determine whether permuting the variables has an effect on the ability of subjects to separate data from a mixture of the two normal multivariate distributions into the appropriate families.

In the meantime, there are a few obvious limitations which require care. When the eyes are very small, the position of the pupil becomes hard to detect. The zero point in the variable which controls the curvature of the mouth may have unusual significance and hence has been avoided in some studies. The corner points where the ellipses of the face meet disappears when the face is circular, losing some information. These are minor points and can easily be avoided.

While the method looks promising, it still remains to be seen whether it can produce results not easily obtained by standard computations on the part of an investigator well versed in statistics and the field of application. One minor success was on the clustering of a randomly selected subset of Fisher's iris data which yielded poor results under the King stepwise clustering algorithm [5]. However, nothing that I would regard as a convincing major success for this method has yet been obtained.

Appendix

A1. Construction of Faces

Given 18 numbers $(x_1, x_2, \ldots, x_{18})$ in appropriate ranges (which will usually be 0 to 1), we define a face (see Fig. 3) as follows. Let H be a nominal distance and let $h^* = \frac{1}{2}(1+x_1)H$ be the distance from the origin to a "corner" point P. As x_1 varies from 0 to 1, h^* varies from $H/2$ to H. Let $\theta^* = (2x_2-1)\pi/4$ be the angle of OP with the horizontal. Let P' be a point symmetric to P about the vertical axis through O. Let $h = \frac{1}{2}(1+x_3)H$ represent the distance from O to U the top of the head and L the bottom of the head, both on the vertical line through O. The upper part of the head is an ellipse which is determined by $P', U,$ and P and an eccentricity x_4. Let x_4 represent the ratio of the width to height of the upper ellipse. Similarly, x_5 is the same ratio for the ellipse through $P', L,$ and P. The nose is a vertical line of length $2hx_6$ with O as center. The mouth intersects the vertical line extended through the nose at a point P_m whose distance below O is $h[x_7 + (1-x_7)x_6]$. This represents a point x_7 part of the way from the bottom of the nose to U. The mouth is part of a circle whose center is h/x_8 above P_m. Thus a positive value of x_8 yields a smile. The mouth is symmetric about the vertical axis through O. Its projection on the horizontal axis has the half-length $a_m = x_9(h/|x_8|)$ unless $(h/|x_8|)$ exceeds the half-width w_m of the face at the height of P_m. In that case $x_9 w_m$ is used. The eyes are located at a height $y_e = h[x_{10} + (1-x_{10})x_6]$ above O and at centers which are $x_e = w_e(1+2x_{11})/4$ from the vertical axis

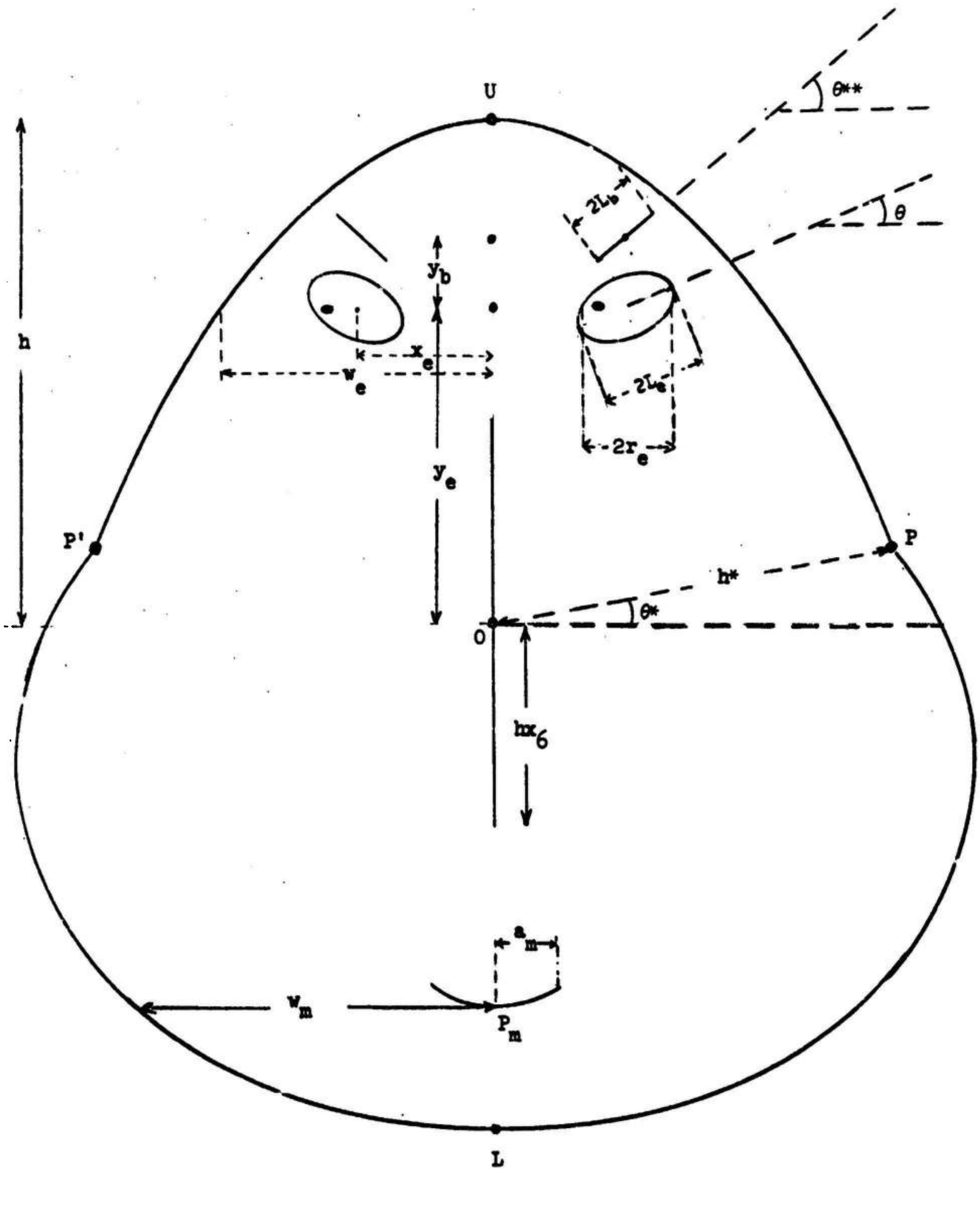


Figure 3

where w_e is the half-width of the face at the height y_e. They are symmetrically slanted at an angle $\theta = (2x_{12}-1)\pi/5$ with the horizontal. The eyes are ellipses with eccentricity x_{13} (height/length before slanting) and half-length $L_e = x_{14}\min(x_e, w_e - x_e)$.

The only asymmetry appears in the location of the pupils which move together an amount $r_e(2x_{15}-1)$ from the center of the eye where $r_e = (\cos^2\theta + \sin^2\theta/x_{13}^2)^{-1/2}L_e$ is the horizontal half-length of the slanted eye at height y_e.

Finally the eyebrows are symmetrically located with centers at a height $y_b = 2(x_{16} + .3)L_e x_{13}$ above the eye centers and slant $2(x_{17}-1)\pi/5$ with respect to the eye, i.e., $\theta^{**} = \theta + (2x_{17}-1)\pi/5$ with respect to the horizontal and half-length $L_b = r_e(2x_{18} + 1)/2$.

One final step taken by the programmer and which has been left intact, is to normalize both horizontal and vertical axes, each by a multiplicative factor, so that the width of the head at its widest part and its height are both equal to a specified constant. This step, which essentially removes two degrees of freedom, was left unaltered for intuitive and aesthetic reasons that are somewhat vague and may require reconsideration when dealing with 18-dimensional data. In the meantime, the effects of x_1 and x_3 are almost but not completely eliminated because of the secondary effects of the normalization, which will adjust all of the other features at the same time as the width and height are normalized.

Most of the parameters x_i are adjusted to range within a subinterval of $(0,1)$. The exceptions are two of the eccentricities, x_4 and x_5, and the parameter controlling curvature of the mouth, x_8.

Ordinarily x_4 and x_5 are kept within 1/2 to 2, and x_8 is kept within (-5,5). The eccentricity of the eye x_{13} has usually been kept within (.4,.8). Some of the ranges must be controlled carefully. We do not want negative length eyes. Others need not be so carefully controlled. It is no calamity to have eyes extend beyond the face.

When the two ellipses of the head meet smoothly, the corner point P is lost, and the variable x_2 loses effect. Restricting x_4 and x_5 to widely separated ranges seems to avoid this problem.

Data are converted to the x parameters as follows. If the variable Z is used to control the parameter x_i, which is to be allowed to range from a_i to b_i, we let

$$x_i = a_i + (b_i - a_i)\left(\frac{Z - m}{M - m}\right)$$

where m and M are the observed minimum and maximum of Z.

A2. Formulae Used on the Construction

We describe a few of the less trivial formulae used in the construction of the faces.

The point P has coordinates $x_o = h^* \cos \theta^*$ and $y_o = h^* \sin \theta^*$. The ellipse through PUP' has equation

$$\frac{x^2}{a_u^2} + \frac{(y - c_u)^2}{b_u^2} = 1$$

where $b_u = h - c_u$, $a_u = x_4 b_u$ and

$$c_u = \frac{1}{2}\left[(h+y_o) - \frac{x_o^2}{x_4^2(h - y_o)}\right] .$$

The ellipse through PLP' has equation

$$\frac{x^2}{a_L^2} + \frac{(y-c_L)^2}{b_L^2} = 1$$

where $b_L = h + c_L$, $a_L = x_5 b_L$ and

$$c_L = \frac{1}{2}\left[(-h+y_o) - \frac{x_o^2}{x_5^2(-h-y_o)}\right]$$

The head is then described by $(\pm x(y), y)$ where

$$x(y) = x_4[b_u^2 - (y-c_u)^2]^{1/2} \qquad y_o \leq y \leq h$$

$$= x_5[b_L^2 - (y-c_L)^2]^{1/2} \qquad -h \leq y \leq y_o$$

The mouth is a circular arc with curvature $|x_8/h|$ through $(0,y_m)$ where $y_m = -h(x_7 + (1-x_7)x_6)$. It is described by

$$y = y_m + (\text{sgn } x_8)\left[\frac{h}{|x_8|} - \sqrt{\left(\frac{h}{x_8}\right)^2 - x^2}\right] , \qquad 0 \leq x \leq a_m$$

where

$$a_m = x_9 \min[x(y_m), h/|x_8|] \quad .$$

The eyes are nominally centered at (x_e, y_e) where

$$y_e = h[x_{10} + (1-x_{10})x_6]$$

$$x_e = x(y_e)[1 + 2x_{11}]/4$$

and have half-length

$$L_e = x_{14} \min[x_e, x(y_e) - x_e]$$

Let (u,v) be the coordinates of an ellipse with center at the origin, half-length L_e and eccentricity x_{13} . Then $v = x_{13}(L^2 - u^2)^{1/2}$ describes part of the ellipse. A similar part of the slanted eye can be described for $0 \leq u \leq L$ by

$$x = x_e + u \cos \theta - v \sin \theta$$

$$y = y_e + u \sin \theta + v \cos \theta$$

and symmetry is used to complete both eyes.

To place the pupils within the eyes, both are moved a distance $r_e(2x_{15}-1)$ from the center of the eye, where r_e, the horizontal half-length of the slanted eye at height y_e, is $(u^2+v^2)^{1/2}$ when $v/u = \tan\theta$. This yields

$$r_e = L_e(\cos^2\theta + x_{13}^{-2}\sin^2\theta)^{-1/2}$$

The program then normalizes all heights and widths by multiplicative factor k/h and $k/\max x(y)$ respectively. Currently k is set at 2 inches. A copy of the program follows.

```
//FACES JOB (J683,323,1.0,4),'BETTY'
//STEP1 EXEC ASMGC
//ASM.SYSIN DD *
RANDOM   CSECT
         USING *,15
         STM   2,5,28(13)
         LM    2,3,0(1)
         L     5,A
         M     4,0(2)
         D     4,P
         ST    4,0(2)
         SRL   4,7
         A     4,CHAR
         ST    4,0(3)
         LM    2,5,28(13)
         BR    14
CHAR     DC    F'1073741824'
A        DC    F'16807'
P        DC    F'2147483647'
         END
/*
//STEP2 EXEC FORTHCL
//FORT.SYSIN DD *
C     PROGRAM DRFACE ON SYS09
      REAL*4 XFACE(1000),YFACE(1000),CAPH,SMALLH,XO,YO,HSTAR,THSTAR,CU,
     *BU,AU,CL,BL,AL,LHSRHS(400),YSAME(201),XPLUS,BUSQ,BLSQ,AN,XNOSE(51)
     *,YNOSE(51),XMOUTH(51),YMOUTH(51),AM,YM,AX8,XLEYE(80),XREYE(80),XE,
     *YE,THETA,X13,L,LSQ,V,XSTAR,YSTAR,YEYES(80),SINTH,COSTH,PUPILX(2),
     *PUPILY(2),R,YB,THSTST,L9,XX,YY,XRBROW(41),XLBROW(41),YBROWS(41),
     *DATA(18),AMODES(200),ID1(100),
     *Y(19),MINY(19),RANGEY(18),MAXY(18),AI(18),BI(18),BIMAI(18)
      INTEGER*4 IFIX(18),IRAND(18)
      LOGICAL*4 IFMT(18)
      REAL*4 PI/3.141593/
      DATA CAPH/9.0/
      INTEGER*4 NU,NL,NFACE,NNOSE,NMOUTH,NEYES,NBROWS
     *,ISTR,IDENT(100)
      DATA NU,NL,NNOSE,NMOUTH,NEYES,NBROWS/400,400,51,51,80,41/
C
      WRITE(6,402)
  402 FORMAT('1  ID(1)     INTEGER')
      CALL MODESG(AMODES,'BETTY, BIN 323',14)
      READ(5,405)IFMT
      READ(5,1)NPLOTS
    1 FORMAT(18I4)
      NINCHS = NPLOTS/4 + 3
      AMODES(97)=250.0*NINCHS
C
C     READ IN STARTING PT. FOR RAND. NO. GENERATOR, IDENT. INFO
      READ(5,405)IFMT
      READ(5,406)INTRAN,IFID,IDNO
  406 FORMAT(I10,2I5)
C
C     READ IN ORDER NUMBERS OF VARIABLES THAT ARE FIXED
      READ(5,405)IFMT
      READ(5,1)NFIXED
      IF(NFIXED.EQ.0)GO TO 1122
      READ(5,1)(IFIX(I),I=1,NFIXED)
      READ(5,2)(DATA(IFIX(I)),I=1,NFIXED)
C
C     READ IN ORDER NUMBERS OF VARIABLES THAT ARE RANDOM
 1122 READ(5,405)IFMT
      READ(5,1)NRAND
      READ(5,1)(IRAND(I),I=1,NRAND)
C
C     CALCULATE NUMBER OF Y VALUES TO READ
      IF(IFID.EQ.0)GO TO 407
      NVREAD=NRAND+1
      GO TO 408
  407 NVREAD=NRAND
      IDCNT=IDNO-1
C
C     READ IN RANGE FOR RANDOM VALUES
  408 READ(5,405)IFMT
      DO 403 I=1,NRAND
      READ(5,2)AI(I),BI(I)
  403 BIMAI(I)=BI(I)-AI(I)
C
C     READ IN MINIMUM AND MAXIMUM OF RANDOM VALUES
      READ(5,405)IFMT
      DO 400 I=1,NRAND
      READ(5,2)MINY(I),MAXY(I)
      RANGEY(I)=MAXY(I)-MINY(I)
  400 CONTINUE
C
C     READ IN FORMAT FOR DATA
      READ(5,405)IFMT
      READ(5,405)IFMT
  405 FORMAT(18A4)
C
      IPLOT=0
      XRMAX=0.0
      DO 50 JPLOT4=1,NPLOTS,4
      XRMIN=XRMAX
      XRMAX=XRMAX+250.
      YRMIN=0.0
      YRMAX=250.
      JEND=MINO(NPLOTS,JPLOT4+3)
      DO 49 JPLOT1=JPLOT4,JEND
      READ(5,IFMT)(Y(I),I=1,NVREAD)
    2 FORMAT(9F8.2)
C
C     FORM 5-DIGIT NUMBER
      CALL RANDOM(INTRAN,UNIFOR)
      RSTR=INTRAN
      IF(IFID.EQ.0)GO TO 23
      XID1=Y(IDNO)
      IF(IDNO.EQ.NVREAD)GO TO 3
      DO 233 J=IDNO,NRAND
  233 Y(J)=Y(J+1)
      GO TO 3
   23 IDCNT=IDCNT+1
      XID1=IDCNT
C
    3 IF(RSTR.GT.99999.0)GO TO 4
      RSTR=RSTR*10.0
      GO TO 3
    4 RSTR=RSTR/10.0
      IF(RSTR.GT.99999.0)GO TO 4
      ISTR=RSTR
      IF(IPLOT.EQ.0)GO TO 46
   44 DO 45 J=1,IPLOT
      IF(ISTR.NE.IDENT(J))GO TO 445
      ISTR=ISTR+1
      GO TO 44
  445 IF(ISTR.LT.IDENT(J))GO TO 456
   45 CONTINUE
   46 IPLOT=IPLOT+1
```

```
      IDENT(IPLOT)=ISTR
      ID1(IPLOT)=XID1
      GO TO 459
  456 INDX=IPLOT+1
      DO 458 I=J,IPLOT
      INDXM1=INDX-1
      IDENT(INDX)=IDENT(INDXM1)
      ID1(INDX)=ID1(INDXM1)
      INDX=INDXM1
  458 CONTINUE
      IPLOT=IPLOT+1
      IDENT(J)=ISTR
      ID1(J)=XID1
C
C     FORM MISSING X VALUES FROM Y
  459 DO 460 J=1,NRAND
      XJ=(Y(J)-MINY(J))/RANGEY(J)
      DATA(IRAND(J))=XJ*BIMAI(J)+AI(J)
  460 CONTINUE
C
      WRITE(6,401)XID1,INTRAN
C
      HSTAR=.5*(1.0+DATA(1))*CAPH
      THSTAR=(2.0*DATA(2)-1.0)*PI*0.25
      SMALLH=.5*(1.0+DATA(3))*CAPH
      XO=HSTAR*COS(THSTAR)
      YO=HSTAR*SIN(THSTAR)
C
C     DRAW FACE
      CU=.5*(SMALLH+YO-XO**2/(DATA(4)**2*(SMALLH-YO)))
      BU=SMALLH-CU
      AU=DATA(4)*BU
      BUSQ=BU**2
      CL=.5*(-SMALLH+YO-XO**2/(DATA(5)**2*(-SMALLH-YO)))
      BL=SMALLH+CL
      AL=DATA(5)*BL
      BLSQ=BL**2
      XMAX=XO
      NFACE=NU+NL
C
      NUP1=NU+1
      YSAME(1)=YO
      LHSRHS(1)=-XO
      NSTEP=NU/2
      NSTPP1=NSTEP+1
      YSAME(NSTPP1)=SMALLH
      LHSRHS(NSTPP1)=0.0
      STPSIZ=(SMALLH-YO)/NSTEP
      ISTOP=NSTEP-1
C
      DO 5 I=1,ISTOP
      IP1=I+1
      YSAME(IP1)=YO+I*STPSIZ
      NUMI=NUP1-I
      XPLUS=DATA(4)*SQRT(BUSQ-(YSAME(IP1)-CU)**2)
      IF(XPLUS.GT.XMAX)XMAX=XPLUS
      LHSRHS(IP1)=-XPLUS
      LHSRHS(NUMI)=XPLUS
    5 CONTINUE
C
      XFACE(1)=LHSRHS(1)
      YFACE(1)=YSAME(1)
      NUP2=NU+2
      DO 6 I=2,NSTEP
      XFACE(I)=LHSRHS(I)
      YFACE(I)=YSAME(I)
      IX2=NUP2-I
      XFACE(IX2)=LHSRHS(IX2)
    6 YFACE(IX2)=YSAME(I)
      XFACE(NSTPP1)=LHSRHS(NSTPP1)
      YFACE(NSTPP1)=YSAME(NSTPP1)
C
      YSAME(1)=YO
      LHSRHS(1)=XO
      NLP1=NL+1
      YSAME(NSTPP1)=-SMALLH
      LHSRHS(NSTPP1)=0.0
      STPSIZ=(YO+SMALLH)/NSTEP
      DO 7 I=1,ISTOP
      IP1=I+1
      NLMI=NLP1-I
      YSAME(IP1)=YO-I*STPSIZ
      XPLUS=DATA(5)*SQRT(BLSQ-(YSAME(I)-CL)**2)
      IF(XPLUS.GT.XMAX)XMAX=XPLUS
      LHSRHS(IP1)=XPLUS
      LHSRHS(NLMI)=-XPLUS
    7 CONTINUE
C
      NLP2=NL+2
      XFACE(NUP1)=LHSRHS(1)
      YFACE(NUP1)=YSAME(1)
      DO 8 I=2,NSTEP
      XFACE(NU+I)=LHSRHS(I)
      YFACE(NU+I)=YSAME(I)
      IX2=NLP2-I
      XFACE(NU+IX2)=LHSRHS(IX2)
      YFACE(NU+IX2)=YSAME(I)
    8 CONTINUE
      XFACE(NU+NSTPP1)=LHSRHS(NSTPP1)
      YFACE(NU+NSTPP1)=YSAME(NSTPP1)
      XMIN=-XMAX
      YMAX=SMALLH
      YMIN=-SMALLH
C
C     DRAW NOSE
      AN=SMALLH*DATA(6)
      XNOSE(1)=0.0
      XNOSE(2)=0.0
      YNOSE(1)=AN
      YNOSE(2)=-AN
C
C     DRAW MOUTH
      YM=-SMALLH*(DATA(6)+(1.0-DATA(6))*DATA(7))
      XOFYM=DATA(5)*SQRT(BLSQ-(YM-CL)**2)
      AX8=SMALLH/ABS(DATA(8))
      AM=DATA(9)*AMIN1(XOFYM,AX8)
      NSTEP=NMOUTH/2
      NMP1=NMOUTH+1
      YMOUTH(NSTEP+1)=YM
      XMOUTH(NSTEP+1)=0.0
      STPSIZ=AM/NSTEP
      X8SQ=(SMALLH/DATA(8))**2
      HBY8=AX8
      IF(DATA(8).LT.0.0)SIGN=-1.0
      IF(DATA(8).GT.0.0)SIGN=1.0
      DO 11 I=1,NSTEP
      XPLUS=-AM+(I-1)*STPSIZ
      XMOUTH(I)=XPLUS
      NMMI=NMP1-I
      XMOUTH(NMMI)=-XPLUS
      YMOUTH(I)=YM+SIGN*(HBY8    -SQRT(X8SQ-XPLUS**2))
   11 YMOUTH(NMMI)=YMOUTH(I)
```

```
C
C     DRAW EYES
      YE=SMALLH*(DATA(6)+(1.0-DATA(6))*DATA(10))
      XOFYE=DATA(4)*SQRT(BUSQ-(YE-CU)**2)
      XE=XOFYE*(1.0+2.0*DATA(11))*0.25
      THETA=(2.0*DATA(12)-1.0)*PI*0.2
      X13=DATA(13)
      L=DATA(14)*AMIN1(XE,XOFYE-XE)
      LSQ=L**2
      SINTH=SIN(THETA)
      COSTH=COS(THETA)
      R=L/SQRT(COSTH**2+SINTH**2/X13**2)
      PUPILX(1)=-XE+R*(2.0*DATA(15)-1.0)
      PUPILX(2)=XE+R*(2.0*DATA(15)-1.0)
      PUPILY(1)=YE
      PUPILY(2)=YE
C
      NSTEP=NEYES/4
      STPSIZ=L/NSTEP
      I1=1
      I2=NSTEP+1
      I3=2*NSTEP+1
      I4=3*NSTEP+1
      U=0.0
      V=X13*L
      XSTAR=-V*SINTH
      YSTAR=V*COSTH
      XX=XE+XSTAR
      YY=YE+YSTAR
      XREYE(I2)=XX
      XLEYE(I2)=-XX
      YEYES(I2)=YY
      XX=XE-XSTAR
      YY=YE-YSTAR
      XREYE(I4)=XX
      XLEYE(I4)=-XX
      YEYES(I4)=YY
      U=L
      XSTAR=U*COSTH
      YSTAR=U*SINTH
      XX=XE+XSTAR
      YY=YE+YSTAR
      XREYE(I3)=XX
      XLEYE(I3)=-XX
      YEYES(I3)=YY
      XX=XE-XSTAR
      YY=YE-YSTAR
      XREYE(I1)=XX
      XLEYE(I1)=-XX
      YEYES(I1)=YY
      I1=I2
      I3=I4
      ISTOP=NSTEP-1
      DO 12 I=1,ISTOP
      U=I*STPSIZ
      V=X13*SQRT(LSQ-U**2)
      XSTAR=U*COSTH-V*SINTH
      YSTAR=U*SINTH+V*COSTH
      XX=XE+XSTAR
      YY=YE+YSTAR
      I2=I2+1
      I4=I4+1
      XREYE(I2)=XX
      XLEYE(I2)=-XX
      YEYES(I2)=YY
      XX=XE-XSTAR
      YY=YE-YSTAR
      XREYE(I4)=XX
      XLEYE(I4)=-XX
      YEYES(I4)=YY
      XSTAR=U*COSTH+V*SINTH
      YSTAR=U*SINTH-V*COSTH
      I1=I1-1
      I3=I3-1
      XX=XE-XSTAR
      YY=YE-YSTAR
      XREYE(I1)=XX
      XLEYE(I1)=-XX
      YEYES(I1)=YY
      XX=XE+XSTAR
      YY=YE+YSTAR
      XREYE(I3)=XX
      XLEYE(I3)=-XX
      YEYES(I3)=YY
   12 CONTINUE
C
C     DRAW EYEBROWS
      YB=YE+2.0*(0.3 + DATA(16))   *L*X13
      THSTST=THETA+PI*(2.0*DATA(17)-1.0)*0.2
      COSTH=COS(THSTST)
      SINTH=SIN(THSTST)
      LB=R*(2.0*DATA(18)+1.0)*0.5
      XX=LB*COSTH+XE
      YY=LB*SINTH+YB
      XRBROW(1)=XX
      XLBROW(1) =-XX
      YBROWS(1) =YY
      XX=-LB*COSTH+XE
      YY=-LB*SINTH+YB
      XRBROW(2) =XX
      XLBROW(2) =-XX
      YBROWS(2) =YY
C
C     ADJUST X AND Y MIN AND MAX TO ALLOW FOR MARGINS
      XLAB=XMIN
      DEL1=(XMAX-XMIN)/8.0
      XMIN=XMIN-DEL1
      XMAX=XMAX+DEL1
      DEL2=(YMAX-YMIN)/8.0
      YLAB=YMIN-DEL2
      YMIN=YMIN-2.0*DEL2
C
C     DRAW CURVES ON CALCOMP
      CALL SUBJEG(AMODES,XMIN,YMIN,XMAX,YMAX)
      CALL OBJECG(AMODES,XRMIN,YRMIN,XRMAX,YRMAX)
C
C     LABEL PLOT WITH 5 DIGIT NUMBER
      CALL INUMBG(AMODES,XLAB,YLAB,5,IST.)
C
C     DRAW FACE
      CALL LINESG(AMODES,NU,XFACE,YFACE)
      CALL LINESG(AMODES,NL,XFACE(NUP1),YFACE(NUP1))
C
C     NOSE
      CALL LINEG(AMODES,0,XNOSE(1),YNOSE(1))
      CALL LINEG(AMODES,1,XNOSE(2),YNOSE(2))
C
C     MOUTH
      CALL LINESG(AMODES,NMOUTH,XMOUTH,YMOUTH)
C
C     EYES
      CALL LINESG(AMODES,NEYES,XLEYE,YEYES)
```

```
      CALL LINESG(AMODES,NEYES,XREYE,YEYES)
C
C     EYEBROWS
      CALL LINEG(AMODES,0,XRBROW(1),YBROWS(1))
      CALL LINEG(AMODES,1,XRBROW(2),YBROWS(2))
      CALL LINEG(AMODES,0,XLBROW(1),YBROWS(1))
      CALL LINEG(AMODES,1,XLBPOW(2),YBROWS(2))
C
C     PUPILS
      CALL POINTG(AMODES,2,PUPILX,PUPILY)
      YRMIN=YRMAX
      YRMAX=YRMAX+250.
   49 CONTINUE
   50 CONTINUE
C
      WRITE(6,402)
  401 FORMAT(F10.3,2X,I10)
      WRITE(6,401)(ID1(J),IDENT(J),J=1,NPLOTS)
C
      CALL EXITG(AMODES)
C
      STOP
      END
/*
//FT16F001 DD UNIT=2314,VOLUME=SER=SYS03,DISP=(NEW,PASS),              XXX
//             DCB=(RECFM=F,BLKSIZE=600),SPACE=(TRK,(15,5),RLSE)
//PLOTTAPE DD DSNAME=PLOTTAPE,DISP=(NEW,KEEP),VOLUME=PRIVATE,         XXX
//             UNIT=TAPE7,LABEL=(,BLP)
//LKED.SYSLMOD DD DSNAME=J683.LIBRARY(FACES),UNIT=2314,               XXX
//             SPACE=(TRK,(10,5,1),RLSE),VOLUME=SER=SYS04,            XXX
//             DISP=(NEW,KEEP)
/*
```

A3. Dictionary of Parameters and Features They Control

The following table provides a dictionary and ranges within which the x_i are typically restrained.

Table 1

Range				
(0,1)	x_1	constrols	h^*	distance from 0 to P
(0,1)	x_2	constrols	θ^*	angle between OP and horizontal
(0,1)	x_3	controls	h	half-height of face
(0.5,2)	x_4	is		eccentricity of upper ellipse of face (width/height)
(0.5,2)	x_5	is		eccentricity of lower ellipse of face (width/height)
(0,1)	x_6	controls		length of nose
(0,1)	x_7	controls	P_m	position of center of mouth
(-5,5)	x_8	controls		curvature of mouth (radius $= h/x_8$)
(0,1)	x_9	controls	a_m	length of mouth
(0,1)	x_{10}	controls	y_e	height of centers of eyes
(0,1)	x_{11}	controls	x_e	separation of centers of eyes
(0,1)	x_{12}	controls	θ	slant of eyes
(0.4,0.8)	x_{13}	is		eccentricity of eyes (height/width)
(0,1)	x_{14}	controls	L_e	half-length of eye (L_e also depends in part on x_{10} and x_{11})
(0,1)	x_{15}	controls		position of pupils
(0,1)	x_{16}	controls	y_b	height of eyebrow center relative to eye
(0,1)	x_{17}	controls	$\theta^{**}-\theta$	angle of brow relative to eye
(0,1)	x_{18}	controls		length of brow

A4. Data for Examples

Example 1: In Table 2a we present a list of variables $Z_1, Z_2, \ldots, Z_6$ preceded by a specimen number. At the end of the list are appended the minima m_i and maxima M_i for the six variables Z_i, $i=1,2,\ldots,6$ used in the faces. The 34th specimen was omitted from the faces because an error in copying had made it seem unreliable.

(In a second study, the 40 specimens from 1 to 41, omitting 34, were used. The minima and maxima used for that study are listed as m_i^*, M_i^* .)

The variables are measurements of 87 nummulited specimens from the Eocene Yellow Limestone Formation, Jamaica [6]. They represent

Z_1 = inner diameter of embryonic chamber (in microns)
Z_2 = total number of whorls
Z_3 = number of chambers in first whorl
Z_4 = number of chambers in last whorl
Z_5 = maximum height of chambers in first whorl (in microns)
Z_6 = maximum height of chambers in last whorl (in microns)

Table 2b identifies the feature variables controlled by the data and the ranges (a_i, b_i) which correspond to the minima (m_i, M_i) in the first set of faces and (m_i^*, M_i^*) in the second set. Tables 2c and 2d are the dictionary relating specimen identity (i.d.) numbers to random code numbers for the two studies.

Table 2a

6 Measurements on 87 Nummulited Specimens

from the Eocene Yellow Limestone Formation, Jamaica

ID	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6
1	160	51	10	28	70	450
2	155	52	8	27	85	400
3	141	49	11	25	72	380
4	130	50	10	26	75	560
5	161	50	10	27	70	665
6	135	50	12	27	88	570
7	165	50	11	23	95	675
8	150	50	9	29	90	580
9	148	48	8	26	85	390
10	150	45	7	31	60	435
11	120	40	6	33	55	440
12	120	51	8	32	56	650
13	100	42	8	30	55	640
14	100	44	9	35	48	430
15	150	40	7	29	65	650
16	90	46	9	30	70	655
17	75	42	8	28	60	640
18	120	47	7	35	67	645
19	200	43	9	30	62	660
20	120	41	8	28	63	530
21	105	50	7	27	64	435
22	210	52	9	26	67	440
23	90	40	10	25	68	430
24	110	52	11	25	60	530
25	100	43	9	25	70	440
26	90	44	7	36	63	454
27	70	45	8	23	64	450
28	100	48	9	27	65	355
29	130	52	9	25	70	380
30	90	45	11	37	74	350
31	80	46	10	32	78	450
32	95	49	10	25	82	260
33	70	44	12	30	85	262
35	95	51	15	31	70	270
36	100	46	11	24	76	270
37	95	48	10	27	74	355
38	85	47	12	25	73	360
39	70	48	11	26	78	365
40	80	54	10	21	80	370
41	85	55	13	33	81	355

Table 2a (Cont'd)

ID	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6
42	200	34	10	24	98	1210
43	260	31	8	21	110	1220
44	195	30	9	20	105	1130
45	195	32	9	19	110	1010
46	220	33	10	24	95	1205
47	220	30	8	25	90	1210
48	190	34	9	26	96	1070
49	285	30	11	19	100	990
50	300	30	9	20	102	1120
51	225	30	10	22	105	985
52	260	34	8	22	97	1090
53	280	30	8	20	112	1200
54	300	34	10	20	108	835
55	310	30	11	19	106	1055
56	290	31	12	26	94	1240
57	260	30	8	22	98	1015
58	290	33	9	25	100	1010
59	160	31	11	20	79	1170
60	240	35	11	20	88	990
61	195	31	8	21	81	975
62	290	34	10	19	94	860
63	210	35	9	22	96	950
64	180	30	11	22	97	990
65	205	29	11	23	90	805
66	215	34	8	21	100	700
67	270	31	8	20	111	1170
68	290	30	9	23	102	1350
69	320	32	10	19	87	1160
70	210	30	9	18	112	1010
71	210	30	8	21	95	1190
72	185	34	9	25	96	1055
73	200	32	8	26	98	980
74	170	29	9	20	95	1095
75	140	30	9	20	98	990
76	90	52	8	24	120	210
77	110	49	9	22	130	220
78	100	56	8	19	128	216
79	95	49	8	24	124	218
80	65	62	9	30	134	200
81	55	50	10	27	128	205
82	70	53	7	28	118	204
83	85	49	11	19	117	206
84	115	50	10	21	122	198
85	110	57	9	26	125	230
86	95	48	8	27	114	228
87	95	49	8	29	118	240
88	120	61	9	24	120	244
m	70	29	6	18	48	198
M	320	62	15	37	134	1350
m*	70	40	6	21	48	260
M*	210	55	15	37	95	675

Table 2b

Details on Faces for Fossil Data

Feature	First Set of Faces		Second Set of Faces	
	Range	Data	Range	Data
x_1	(0.3,0.8)	Z_1	0.9	
x_2	(0.1,0.5)	Z_2	(0.2,0.8)	Z_1
x_3	0.7		0.9	
x_4	(1.2,2.0)	Z_3	0.75	
x_5	1.0		(1.0,2.0)	Z_4
x_6	0.3		0.4	
x_7	(0.2,0.8)	Z_4	0.5	
x_8	(0.5,5.0)	Z_5	(-4.0,4.0)	Z_3
x_9	0.5		(0.2,0.8)	Z_4
x_{10}	0.5		0.5	
x_{11}	0.5		0.5	
x_{12}	0.5		(0.2,0.8)	Z_5
x_{13}	0.6		0.6	
x_{14}	(0.2,0.8)	Z_6	(0.2,0.8)	Z_6
x_{15}	0.5		0.5	
x_{16}	0.5		0.5	
x_{17}	0.5		0.5	
x_{18}	0.5		0.5	

Table 2c

Dictionary for 87 Faces in Figure 1A

ID(1)	Integer	ID(1)	Integer	ID(1)	Integer	ID(1)	Integer
1.000	2069660917	46.000	2069660917	43.000	10975	87.000	10875×
2.000	1998401560	47.000	1998401560	8.000	11475	53.000	11475×
3.000	490779840	48.000	490779840	35.000	11518	79.000	11518×
4.000	52082753	49.000	52082753	39.000	11675	83.000	11675×
5.000	1328985342	50.000	1328985342	26.000	11704	71.000	11704×
6.000	279230547	51.000	279230547	21.000	12820×	66.000	12820
7.000	776034734	52.000	776034734	16.000	13111	61.000	13111×
8.000	1147586107	53.000	1147586107	5.000	13289×	50.000	13289
9.000	929066642	54.000	929066642	23.000	13627×	68.000	13627
10.000	469454757	55.000	469454757	31.000	14317×	76.000	14817
11.000	271181821	56.000	271181821	18.000	14888	63.000	14888×
12.000	792566613	57.000	792566613	22.000	15697	67.000	15697×
13.000	1973485997	58.000	1973485997	24.000	16034	69.000	16034×
14.000	494223664	59.000	494223664	45.000	16079	78.000	16128
15.000	2097857899	60.000	2097857899	33.000	16128×	77.000	17081×
16.000	1311192047	61.000	1311192047	32.000	17081	62.000	18750
17.000	1875032062	62.000	1875032062	17.000	18750×	70.000	19083
18.000	1488829956	63.000	1488829956	25.000	19083×	58.000	19734
19.000	285615648	64.000	285615648	13.000	19734×	47.000	19984×
20.000	716244891	65.000	716244891	2.000	19984	46.000	20696
21.000	1282041602	66.000	1282041602	1.000	20696×	80.000	20726
22.000	1569774463	67.000	1569774463	36.000	20726×	60.000	20978
23.000	1362796246	68.000	1362796246	15.000	20978×	84.000	21101
24.000	1603411267	69.000	1603411267	40.000	21101×	85.000	26989×
25.000	1908361913	70.000	1908361913	41.000	26989	56.000	27118
26.000	1170403846	71.000	1170403846	11.000	27118×	72.000	27233
27.000	27233202	72.000	27233202	27.000	27233×	51.000	27923×
28.000	294409203	73.000	294409203	6.000	27923	64.000	28561
29.000	333152133	74.000	333152133	19.000	28561×	73.000	29440×
30.000	798031602	75.000	798031602	28.000	29440	74.000	33315
31.000	1481759299	76.000	1481759299	29.000	33315×	55.000	46945×
32.000	1708167681	77.000	1708167681	10.000	46945	48.000	49077
33.000	1612821471	78.000	1612821471	3.000	49077×	59.000	49422×
35.000	1151870663	79.000	1151870663	14.000	49422	81.000	51114×
36.000	2072638983	80.000	2072638983	37.000	51114	49.000	52082×
37.000	511149294	81.000	511149294	4.000	52082	86.000	57887
38.000	951596258	82.000	951596258	42.000	57887×	65.000	71624×
39.000	1167588997	83.000	1167588997	20.000	71624	52.000	77603
40.000	2110189940	84.000	2110189940	7.000	77603×	88.000	78283
41.000	269891375	85.000	269891375	44.000	78283×	57.000	79256×
42.000	578877161	86.000	578877161	12.000	79256	75.000	79803×
43.000	1087524017	87.000	1087524017	30.000	79803	54.000	92906
44.000	782834102	88.000	782834102	9.000	92906×	82.000	95159
45.000	1607930792			38.000	95159×		

Table 2d

Dictionary for 40 Faces in Figure 1B

ID(1)	Integer	ID(1)	Integer
6.000	10617	1.000	1532932388
30.000	10824	2.000	633332057
32.000	11352	3.000	1482927467
40.000	11548	4.000	2014214434
9.000	11980	5.000	2117264577
10.000	12003	6.000	1061714849
13.000	12894	7.000	799844220
28.000	12909	8.000	1881658967
24.000	13115	9.000	1198072647
39.000	13158	10.000	1200303857
3.000	14829	11.000	45544681
1.000	15329	12.000	965275235
25.000	16008	13.000	1289405207
21.000	16720	14.000	775832172
22.000	16795	15.000	2038093867
38.000	18228	16.000	1879453019
16.000	18794	17.000	629926610
8.000	18816	18.000	82154560
4.000	20142	19.000	2087188546
15.000	20380	20.000	232518877
19.000	20871	21.000	1672011846
23.000	20873	22.000	1679574727
5.000	21172	23.000	2087380521
37.000	21684	24.000	1311559055
33.000	22326	25.000	1600884577
20.000	23251	26.000	244472376
26.000	24447	27.000	711006721
11.000	45544	28.000	1290947939
17.000	62992	29.000	934725132
2.000	63333	30.000	1082415719
35.000	67724	31.000	827015496
27.000	71100	32.000	1135277888
36.000	73688	33.000	223260021
14.000	77583	35.000	677241638
7.000	79984	36.000	736880766
18.000	82154	37.000	216841913
31.000	82701	38.000	182282832
41.000	85536	39.000	1315876802
29.000	93472	40.000	1154814408
12.000	96527	41.000	8553670

Example 2: Table 3a contains specimen number (i.d. and 12 measurements) for 53 specimens taken at intervals from a 4500-foot core drilled from a Colorado mountainside to locate a deposit of molybdenum. The 12 variables Z_i, $i=1,2,\ldots,12$ represent mineral contents. The last 5 variables are measurements by different methods of 5 of the minerals covered in the first seven measurements. At the bottom of the table are the minima and maxima m_i and M_i. Further identification of the data has not been furnished me. When no trace of the element appeared, the nominal value 0.001 was used.

Six different sets of faces were obtained. The feature variables and the ranges (a_i, b_i) corresponding to each set are identified in Table 3b.

Table 3c indicates the two-way dictionary between the specimen number (ID) and the randomly generated code number. In the second part, the first five digits of the code number is presented, but in the first part the code has up to ten digits.

A copy of the program is presented in Table 4. This program was put into a memory file in compiled form and requires a simpler program to enter the relevant data and parameters to drive the main program. This program was constructed by Mrs. Elizabeth Hinkley. At this time the cost of drawing these faces is about 20 to 25 cents per face on the IBM 360-67 at Stanford University using the Calcomp Plotter. Most of that cost is in the computing. I believe that with some attention to cost-cutting it may be possible to reduce this cost considerably. For example, many square roots needed for neighboring points could be replaced by linear approximations derived from Taylor Expansions based on the preceding point.

I wish to thank Mrs. Hinkley for her excellent work in assembling the program so that it could be used conveniently.

Table 3a

Data on 12 Variables Representing Mineral Contents From a 4500-Foot Core Drilled from a Colorado Mountainside

ID	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9	Z_{10}	Z_{11}	Z_{12}
200	320	105	057	050	001	001	001	060	020	250	210	370
201	280	150	040	050	001	001	001	060	040	210	130	420
202	260	165	033	050	001	001	001	060	010	250	090	440
203	305	110	044	040	001	001	001	050	050	260	140	250
204	290	160	035	035	001	001	001	050	020	210	060	510
205	275	130	047	035	001	001	001	050	020	230	090	570
206	230	155	035	035	001	001	001	080	020	270	170	400
207	300	115	050	060	001	001	001	120	010	280	190	300
208	250	130	041	030	005	001	001	070	030	250	110	330
209	285	120	047	040	001	001	001	070	010	240	170	280
210	280	105	047	070	001	001	001	060	020	370	070	300
211	300	135	050	040	001	001	001	120	060	250	160	200
212	280	110	056	050	001	001	001	150	010	280	270	280
213	305	080	065	080	005	001	001	130	010	300	260	260
214	230	175	029	035	001	001	001	270	030	250	140	240
215	325	060	052	090	001	001	001	160	010	280	260	170
216	270	170	025	040	001	001	001	160	010	290	070	330
217	250	185	031	025	001	001	001	120	001	260	080	330
218	260	185	030	015	001	001	001	270	080	480	010	330
219	270	185	032	010	005	001	001	180	040	450	020	220
220	325	045	053	005	020	001	001	600	080	660	020	250
221	315	090	047	005	020	001	001	410	200	600	060	260
222	335	100	047	010	040	001	001	360	080	590	110	170
223	310	010	049	005	080	018	001	640	240	630	060	190
224	410	001	049	001	075	032	001	760	440	800	001	001
225	360	001	048	001	080	055	001	770	260	770	010	010

Table 3a (Cont'd.)

ID	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9	Z_{10}	Z_{11}	Z_{12}
226	310	015	051	001	105	036	001	660	380	640	001	010
227	420	005	049	001	095	056	001	620	520	680	001	001
228	415	020	049	005	025	036	001	370	220	340	001	001
229	420	005	041	001	070	060	001	630	510	580	001	001
230	450	005	040	001	090	070	001	690	570	630	001	001
231	395	001	025	015	100	071	001	580	530	560	001	010
232	380	010	027	025	035	039	001	350	320	400	001	270
233	430	010	025	030	030	025	001	340	340	360	001	200
234	410	075	022	010	005	015	001	170	170	170	001	060
235	520	055	024	040	005	001	001	210	190	190	001	180
236	385	135	018	010	005	008	001	140	200	260	001	020
237	535	065	010	020	001	001	001	110	230	270	001	070
238	550	095	001	010	001	001	001	050	230	270	001	030
239	510	100	001	001	001	001	001	190	150	230	001	110
240	510	095	001	040	001	001	001	140	100	150	001	040
241	385	180	010	001	001	001	001	050	050	300	001	050
242	505	125	001	001	001	001	001	001	200	130	001	030
243	470	090	001	020	001	001	001	160	300	380	001	060
244	465	110	001	035	001	001	001	260	440	500	001	060
245	400	140	001	015	001	023	001	330	400	390	001	040
246	415	105	015	025	040	032	001	220	190	270	001	010
247	435	075	010	015	001	069	001	370	360	500	001	010
248	370	145	010	010	005	012	040	130	080	330	001	030
249	380	210	001	001	001	001	020	070	001	050	001	030
250	430	065	001	005	020	001	075	130	070	300	001	020
251	420	080	030	001	005	026	001	050	100	350	001	050
252	425	060	035	005	001	001	030	100	010	340	001	010
m	250	001	001	001	001	001	001	001	001	050	001	001
M	520	210	065	090	105	071	075	770	570	800	270	570

Table 3b

Details on Data for Example 2

Feature Variable	Set A Range	Set A Data	Set B Range	Set B Data	Set C Range	Set C Data	Set D Range	Set D Data	Set E Range	Set E Data	Set F Range	Set F Data
x_1	.8		.8		.8		.8		.8		.8	
x_2	(.2,.8)	z_1	(.2,.8)	z_1	(.2,.8)	z_8	.5		.5		.5	
x_3	.9		.9		.9		.9		.9		.9	
x_4	.9		.9		.9		1.4		.8		1.4	
x_5	(1,2)	z_2	(1,2)	z_2	(1,2)	z_9	.8		1.4		.8	
x_6	(.1,.5)	z_3	(.1,.5)	z_3	.5		(.1,.5)	z_3	.5		.5	
x_7	(.2,.8)	z_4	(.2,.8)	z_4	(.1,.5)	z_{10}	(.2,.8)	z_4	(.2,.8)	z_4	(.2,.8)	z_{10}
x_8	(.3,2.0)	z_5	(.3,2.0)	z_5	(.2,.8)	z_{11}	(.3,2.0)	z_5	1.0		1.0	
x_9	(.2,.9)	z_6	(.2,.9)	z_6	(.3,2.0)	z_{12}	.5		.5		.5	
x_{10}	(.1,.8)	z_8	.5		.5		(.1,.8)	z_8	(.1,.8)	z_1	(.1,.8)	z_8
x_{11}	(.2,.8)	z_9	.5		.5		(.2,.8)	z_9	(.2,.8)	z_2	(.2,.8)	z_9
x_{12}	(.5,.8)	z_{10}	.5		.5		(.3,.8)	z_{10}	(.3,.8)	z_3	(.3,.8)	z_{11}
x_{13}	(.4,.8)	z_{11}	.6		.6		(.4,.8)	z_{11}	(.4,.8)	z_5	(.4,.8)	z_{12}
x_{14}	(.2,.8)	z_{12}	.5		.5		(.4,.8)	z_{12}	(.4,.8)	z_6	.6	
x_{15}	.5		.5		.5		(.1,.9)	z_1	.5		.5	
x_{16}	(.3,.6)	z_7	(.3,.6)	z_7	.4		(.2,.8)	z_7	(.2,.8)	z_7	.4	
x_{17}	.5		.5		.5		(.2,.8)	z_2	.5		.5	
x_{18}	.5		.5		.5		(.2,.8)	z_6	.5		.5	

Table 3c

Dictionary for Specimens in Figures 3A-3F

ID(1)	Integer	ID(1)	Integer	ID(1)	Integer	ID(1)	Integer
200.000	1532932388	227.000	1290947939	205.000	10617	218.000	20871
201.000	633332057	228.000	934725132	229.000	10824	222.000	20873
202.000	1482927467	229.000	1082415719	231.000	11352	204.000	21172
203.000	2014214434	230.000	827015496	238.000	11548	235.000	21684
204.000	2117264577	231.000	1135277888	208.000	11980	232.000	22326
205.000	1061714849	232.000	223260021	209.000	12003	219.000	23251
206.000	799844220	233.000	677241638	242.000	12375	225.000	24447
207.000	1881658967	234.000	736880766	249.000	12524	246.000	26279
208.000	1198072647	235.000	216841913	212.000	12894	244.000	29127
209.000	1200303857	236.000	182282832	227.000	12909	210.000	45544
210.000	45544681	237.000	1315876802	223.000	13115	250.000	46205
211.000	965275235	238.000	1154814408	237.000	13158	251.000	51382
212.000	1289405207	239.000	8553670	245.000	13385	248.000	61785
213.000	775832172	240.000	2027610988	202.000	14829	216.000	62992
214.000	2038093867	241.000	1787364720	200.000	15329	201.000	63333
215.000	1879453019	242.000	1237594804	224.000	16008	233.000	67724
216.000	629926610	243.000	1876749633	247.000	16379	226.000	71100
217.000	82154560	244.000	291274695	220.000	16720	234.000	73688
218.000	2087188546	245.000	1338567352	221.000	16795	252.000	75608
219.000	232518877	246.000	262799092	241.000	17873	213.000	77583
220.000	1672011846	247.000	1637961012	236.000	18228	206.000	79984
221.000	1679574727	248.000	617857791	243.000	18767	217.000	82154
222.000	2087380521	249.000	1252460092	215.000	18794	230.000	82701
223.000	1311559055	250.000	462058350	207.000	18816	239.000	85536
224.000	1600884577	251.000	513820898	203.000	20142	228.000	93472
225.000	244472376	252.000	756088099	240.000	20276	211.000	96527
226.000	711006721			214.000	20380		

References

[1] Anderson, E. (1960), "A semigraphical method for the analysis of complex problems", Technometrics 2, 387-391.

[2] Andrews, D.F., "Plots of high dimensional data", to be published in Biometrika.

[3] Daetz, D., "A graphical technique to assist in sensitivity analysis", unpublished report.

[4] Pickett, R., and White, B.W. (1966), "Constructing data pictures", Proc. of 7th Nat'l. Symp. of Soc. for Info. Display, 75-81.

[5] Solomon, H. (1970), "Numerical Taxonomy", Stanford Technical Report No. 167, 1-44.

[6] Wright, R.M., and Switzer, P., "Numerical Classification Applied to Certain Jamaican Eocene Nummuliteds", submitted for publication to the Journal of the Int. Assoc. for Math. Geology.