

Chernoff face graphs as an efficient way of creating comprehensive Patient Profiles in SAS®

Adam Amborski, Quanticate, Warsaw, Poland

ABSTRACT

Patient Profile listings usually contain a large range of data domains and can run for many pages. With such a large volume of information, it may not be simple or quick to draw together a holistic view of a patient's health or group patients together. But it doesn't need to be that way! Instead, imagine a comprehensive graph containing a patient's key data in one place, making interpretation of Patient Profiles much easier. One graph can be used to reflect various health characteristics such as vital signs and laboratory data. Just as the face can reveal one's health, why not present the patient's 'face' on a graph? Chernoff faces can display multivariate data in the shape of a human face. The aim of the proposed paper is to present useful tips on how to create Patient Profile face graphs in SAS®.

INTRODUCTION

In this paper an attempt is made at describing the use of face graphs, proposed first by Chernoff, as a means of summarizing a Patient Profile. Some ideas are brought up for further consideration. Furthermore, some difficulties in interpreting face graphs are discussed.

THE IDEA

The human brain has a fantastic ability to discern faces at a quick glance. The identification takes place at a subconscious level and is very fast. This ability develops in early childhood and is the same across continents and cultures. The idea is to use this well-developed subconscious ability to allow for the quick assessment of a Patient Profile by showing a "face" built not on an actual picture, but on data collected in the course of a clinical study. For every person in a study, a multitude of data is gathered, such as adverse events, concomitant and previous drug and non-drug treatments, primary and secondary diagnosis, physical examination results, etc. It is impossible to clearly convey all of this information, but we could show some basic information, depending on the focus we choose.

Chernoff originally selected 18 features of a face to represent multivariate data. We do not need to use all 18 every time, in one of the original analysis only 8 features were used, the others remain unchanged between graphs.

There is code available from 'Friendly, M. (2007)', which allows for the drawing of customized, possibly asymmetric faces. The macro uses an Annotate Dataset with PROC GSLIDE to produce a matrix of graphs.

This allows pattern recognition and can group patients by their "looks". Hence, the subconscious process of categorizing people into groups by their external features may be put to good use. Due to special sensitivity, changes in relations between parameters that were previously unconnected across a group of subjects may show in the recognition of a human

face. This may lead to a better understanding of the data and help to propose more appropriate models.

FACES MACRO DEFINITION STATEMENT

```
%macro FACES(
    data=_last_,      /* Name of input data set          */
    out=asym,         /* Name of output anno set          */
    id=,              /* Character ID variable            */
    idnum=,           /* Numeric ID variable              */
    blks=1,           /* Blocks per page                  */
    rows=4,           /* Rows per block                   */
    cols=4,           /* Columns per block                */
    res=3,            /* resolution: 1=high/3=low         */
    frame=Y,          /* frame around each face?          */
    color='BLACK',    /* color of each face: variable     */
    hcolor='BLACK',   /* name or string in quotes         */
    row=,             /* use to assign particular         */
                    /* locations to faces               */
    col=,             /* column variable                  */
    blk=,             /* block variable                   */

    /* Variables can be assigned to features either by          */
    /* listing 18 variable names for LEFT and RIGHT or          */
    /* by assigning individually to L and R parameters.          */
    /* Variable names can appear more than once.                */
    /* Use . in LEFT= or RIGHT= to skip a parameter (leave     */
    /* unassigned).                                              */
    left=,
    right=,
    /* Names of variables assigned to features                    */
    r1=, r2=, r3=, r4=, r5=, r6=, r7=, r8=, r9=,
    r10=,r11=,r12=,r13=,r14=,r15=,r16=,r17=,r18=,
    l1=, l2=, l3=, l4=, l5=, l6=, l7=, l8=, l9=,
    l10=,l11=,l12=,l13=,l14=,l15=,l16=,l17=,l18=,
    gout=GSEG,        /* name of graphics catalog         */
    name=FACES         /* name for graphic catalog entry    */
);
```

While the macro definition may seem extensive, most of the macro variables have sensible default values or are optional. As mentioned earlier, there is no need to assign a variable to all 18 features, alternatively, multiple features can be assigned one variable, thus increasing the visual effect thereof.

The macro requires parameters to be normalized to the range of [0,1], and this can be achieved in a number of ways with one of the simplest being:

```
proc stdize data=all out=all_sc method=range;
    var sbp dbp temp bmi sex age fgluc;
run;
```

The procedure STDIZE offers a multitude of options and statements that may allow a more sophisticated approach to the task.

EXAMPLE OF USE

Dummy data has been generated for the purpose of this article, with some clinical data assigned values close to average among healthy people. It has been slightly manipulated by superimposing conditions related to diabetes mellitus and pregnancy, while allowing for some differences between subjects that are not related to the two states mentioned. While all the face parameters have been assigned a variable, some variables are assigned to multiple parameters, and the faces remain symmetrical.

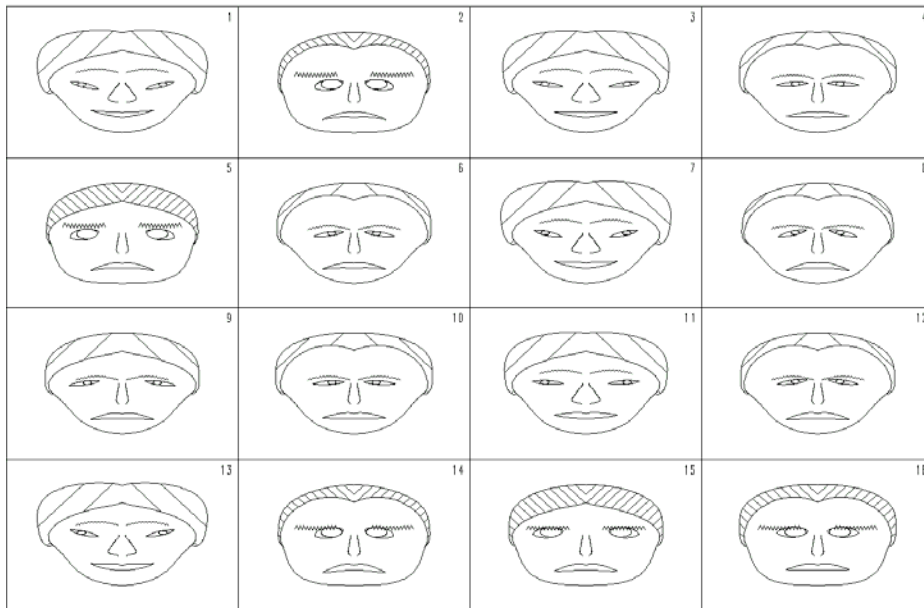


Figure 1. Faces representing condition of 16 hypothetical patients. Please try to discern groups of subjects with similar features.

Figure 1 presents faces representing some hypothetical condition of 16 patients (one face per patient). Numbers have been assigned to each face for identification purposes. Should the method work, a reader should easily find three groups of patients, each with their own common set of features. The grouping of patients with similar ‘faces’ is the initial step in pursuing reasons, names, and, possibly, appropriate treatment, for each specific group. The advantage of this method is that prior knowledge of the existence of a specific condition, that would make the graphs look similar, is not required.

EXAMPLE OF USE – DOES IT WORK?

The data used is rather “over-pure” – the patients are very close to what is presumed a norm, and only selected patients have been attributed traces of an extraordinary state. Figure 2 is shown to allow readers to check if their interpretation of Figure 1 was as expected. If it is not, it means that the scaling and selection of variables to face parameters, the method, or the example, were inappropriate. If it is as expected, further study of real-life examples are required, so that the method can be used confidently and the advantages over other methods utilized.

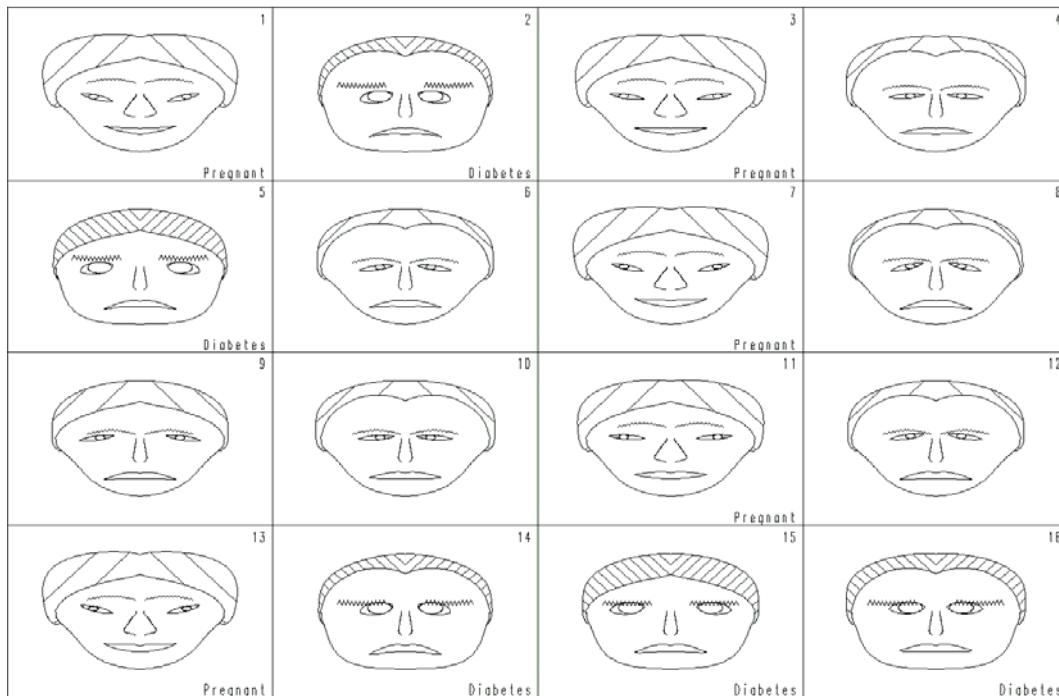


Figure 2. Faces as previously but with labels.

THE CRITIQUE

Robert Kosara provides an extensive critique of Chernoff faces. The author analyses phenomena such as face perception mechanism and pareidolia. In his critique, Kosara has proven that this technique does not currently allow an audience to read values attributed to certain features based on multivariate data. The conclusion is that there is more research on face perception needed to better understand the relationships the human brain makes between the elements of a face.

CONCLUSION

The critique, while acute, does not exclude the potential use for grouping patients with a common set of features. This could be an initial step in pursuing reasons, names, and, possibly, appropriate treatment, for a specific group of patients. The advantage of the method is that prior knowledge of the existence of a specific condition, that would make their graphs look similar, is not required.

REFERENCES

- Chernoff, H., "The use of faces to represent points in k-dimensional space graphically," J. Am. Stat. Assoc., v68, 361-368 (1973).
- Friendly, M. (2007). "Faces macro: Faces display of multivariate data", Version 1.5 (19 Apr 2005). <http://www.datavis.ca/sasmac/faces.html>
- Kosara, R., "A Critique of Chernoff Faces", <http://eagereyes.org/VisCrit/ChernoffFaces.html>

ACKNOWLEDGMENTS

I would like to thank Edyta Winciorek for her help and support in this work. Furthermore, I would like to thank Quanticate for encouragement and support.

CONTACT INFORMATION

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

Adam Amborski
Quanticate
Hankiewicza 2
02-103 Warsaw
POLAND
Phone: +48 22 576 21 47
Fax: +48 22 576 21 59
Email: adam.amborski@quanticate.com
Web: www.quanticate.com

Brand and product names are trademarks of their respective companies.