MATHEMATICAL SCIENCES

FACE HISTOGRAMS

Bazarbayeva La.
Assistant-Professor, PhD
Zhumagulova A.
Bachelor's student in Mathematics
Nartbayeva M.
Bachelor's student in Mathematics
Nartbayeva B.
Bachelor's student in Mathematics
SDU University
Kaskelen, Kazakhstan
https://doi.org/10.5281/zenodo.20448080

Abstract

The paper proposes the concept of the Face Histogram as a technique for visualizing multivariate data using Chernoff Faces. Traditional histograms can only display one variable at a time, which is insufficient for multivariate data analysis. In contemporary data analysis, datasets often comprise multiple variables that need to be analyzed together to identify important patterns and relationships. The proposed method uses Chernoff Faces to map statistical variables to facial attributes such as eyes, mouth, nose, and face shape. Each variable is normalized and associated with a specific facial parameter, allowing each observation in the dataset to be depicted as a unique face. The Face Histogram aids in visual pattern recognition and allows for the simultaneous comparison of several variables within a single visualization. this method is a great alternative to traditional histograms for analysis of multivariate datasets, although it reduces numerical precision.

Keywords: Chernoff Faces, data visualization, multivariate analysis, Face Histogram, statistical graphics

1. Introduction

1. Background of the Study

One way to analyze data is by way of data visualization. This is one of the main tools used by many people to analyze data as part of their statistical analysis or learning using a machine learning system. The visualizations are used to help show the data, usually in the form of distributions or relationships, using such simple graphs as histograms, bar charts and scatter plots. In most cases these graphs are for use with low dimensional data and as such do not work well for higher dimensional data.

2. Problem of the Study

In many real-world situations, observations are multivariate, meaning each observation includes several variables that will be analyzed together. As said before, traditional histograms can only represent one variable, thus, it will not be possible to detect interactions or even patterns in high dimensional data spaces by means of histograms. Often, several graphs have to be generated and analyzed by the analyst in turn which again causes problems like lack of insight and high load on memory and attention.

3. Gap in Existing Methods

Even when using multivariate visualization there are many different ways to visualize data, the majority of graphs created for multivariate data are difficult to interpret by non-expert users. A number of strategies are in use to display multiple variables in one structure, to enable easier interpretation.

4. Purpose of the Study

A Chernoff Face is used to be used for a visualization of a statistical distribution. A Face Histogram is developed for such a purpose. The distribution of several statistical variables is encoded into the human facial features.

In the paper a method is proposed to compare multivariate observations in a straightforward way by a single display.

5. Method Overview

Each variable in the variables of a given dataset are normalized (i.e. transformed into a 0 to 100 scale) and then mapped to the features of a Chernoff Faces representation. Each face in the resulting Face Histogram represents an observation in the given data set. All the variables of the observation are displayed in parallel on the single structure of a face, thus enabling a visually intuitive comparison of observations.

6. Contribution of the Study

The main contribution of this work is the introduction of the Face Histogram framework for improving the pattern recognition of multivariate data by exploiting the ability of humans to recognize differences on faces. The method is more interpretable than other methods but it does not try to preserve the absolute numerical values. This method is better suited for exploratory data analysis than for exact statistical reporting.

7. Paper Outline

The remaining part of this paper is organized as follows: Section 2 introduces the concept of Face Histogram and methodology of Chernoff Faces . The algorithm to build the visualization is described in Section 3. Applications of the method are discussed in Section 4. Final remarks and limitations conclude the study in Section 5. (Considered studies ([1], [3], [5], [2], [4])

8. Literature Review

Link	Author	Main idea	Connection to our work
Chernoff Faces	John Nelson	This work introduces Chernoff faces as a method for visualizing multivariate data using human facial features. Each variable is mapped to a facial attribute such as eye size, mouth curvature, or face shape. The approach leverages human sensitivity to facial expression to make complex datasets more interpretable.	Our work builds directly on this idea of encoding variables into facial features. However, we extend the classical approach by adding color encoding and additional shape transformations, improving expressiveness for multivariate datasets and allowing more flexible representation beyond standard facial mappings.
A Critique of Chernoff Faces	Robert Kosara	This critique discusses limitations of Chernoff faces, particularly issues with perceptual accuracy, inconsistent feature interpretation, and difficulty in comparing faces quantitatively. It argues that while visually appealing, Chernoff faces may not always be effective for precise data analysis.	Our approach addresses some of these limitations by introducing structured feature scaling, additional visual channels (colors and shapes), and controlled mapping rules, which aim to improve comparability between faces and reduce ambiguity in interpretation.
An application of the V-system to the clustering of Chernoff faces	Ruixia Song, Zhaoxia Zhao, Xiaochun Wang	This paper explores the use of the V-system to improve clustering of Chernoff faces, making it easier to group similar multivariate observations. It demonstrates how mathematical transformation systems can improve structure and interpretability in face-based visualization.	We extend this idea by not only clustering but also creating face histograms across countries, where each facial feature encodes statistical distributions of realworld multivariate datasets. Our work focuses more on comparative visualization across groups rather than only clustering.
Applied Multivariate Statistical Analysis	ichard A. Johnson, Dean W. Wichern	This textbook provides the theoretical foundation for multivariate statistical methods, including covariance structures, dimensionality reduction, and interpretation of high-dimensional datasets. It is a core reference for statistical modeling of complex data.	Our visualization method relies on principles of multivariate data representation described in this book. Each facial feature corresponds to a variable in a multivariate dataset, and our enhancements (color, shape, histograms) serve as an alternative visual encoding of high-dimensional statistical structures.

2. Chernoff Faces for Multivariate Data Visualization

Chernoff Faces are a graphical method to display multivariate data by making use of human facial features. Chernoff Faces were first introduced by Herman Chernoff in 1973. In Chernoff Faces every variable of a dataset is assigned to a particular feature of a face, such as the size of the eyes, the curvature of the mouth, the length of the nose, etc.

In this research a number of modifications to the Chernoff Faces have been made, leading to a so-called Face Histogram. The Face Histogram makes use of interval-based color mapping and changes in the facial features. With the Face Histogram, a number of variables can be displayed simultaneously.

3. Constructing a Face Histogram Using Chernoff Faces

3.1 Step 1: Preparing the Dataset

The first step is selecting a multivariate dataset. A multivariate dataset contains several variables for each observation. Let's take an example where four variables are used: **Example**

- GDP (Gross Domestic Product)
- Education Index
- Happiness Score
- Employment Rate

Each row in the dataset represents one country.

Table 2:

Initial Multivariate Dataset

Country	GDP	Education	Happiness	Employment
Kazakhstan	14005	0.82	6.2	65.0
Italy	40226	0.89	6.4	60.1
Japan	32476	0.93	6.1	77.0
Canada	55209	0.94	6.9	73.2

Traditional histograms would require separate graphs for each variable. The purpose of the Face Histogram method is to combine all variables into one visual structure.

3.2 Step 2: Creating Intervals

After selecting the dataset, intervals must be created. In this case GDP, happiness, education and employment index was studied globally. After that minimum and maximum values of each variables are fixed:

1. GDP → Face Shape

Minimum ≈ $154 (Burundi),

Maximum ≈ $138.935 (Bermuda)

2. Education Index → Eye Shape

Minimum ≈ 0.17 (Niger, Chad, South Sudan)

Maximum ≈ 0.97 (Norway, Germany, Switzerland)

3. Happiness Score → Mouth Shape

Minimum ≈ 1.7 (Afghanistan) Maximum ≈ 8.0 (Finland)

4. Employment Rate → Nose Shape

Minimum ≈ 35% (South Africa)

Maximum ≈ 85% (Switzerland, Iceland, Netherlands)

3.2.1 GDP → Face Shape

GDP connected to the overall face shape. GDP shows the general economic health and strength of a country and was picked as the main factor affecting the main structural feature of the Chernoff face. When GDP values changed, the facial outline shifted too. This helped visually set apart countries with different economic levels. Range for GDP is [$154; $138.935]

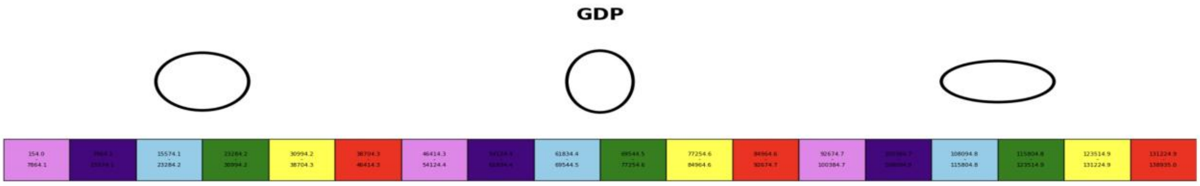

Figure 1: All intervals of GDP

This range was divided into 3 intervals by 3 different face shapes (see Figure 1). Each face has 6 subintervals by 6 colors in increasing order: violet, indigo, blue, green, yellow, red.

For example: there are 6 color intervals for first face shape (see Figure 2). If GDP falls in one certain interval, it means face edge will be the corresponding color.

If GDP is low(lies in 1st interval), then face shape is circular. If value is in 2nd interval, then face becomes oval. Finally, if GDP is high(3rd interval) face shape is wide.

3.2.2 Education Index → Eye Shape

Range for Education is [0.17; 0.97]

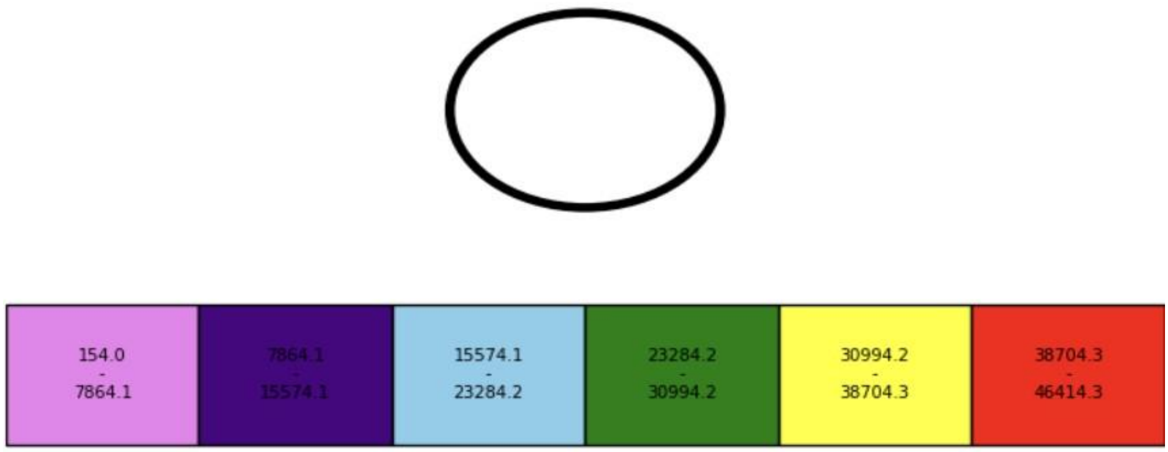

Figure 2: One interval of GDP

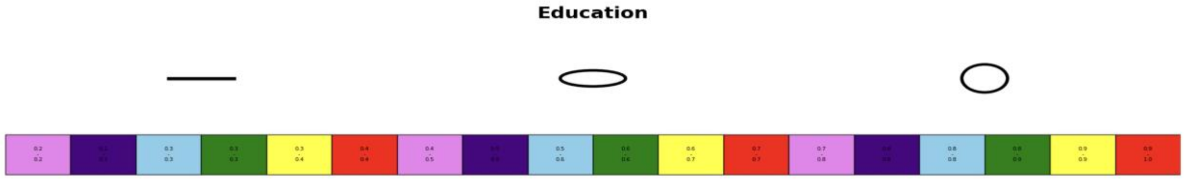

Figure 3: All intervals of Education

The same logic is applied to the eyes. If education index is low, the eyes are closed. The higher the value, the wider the eyes are opened.

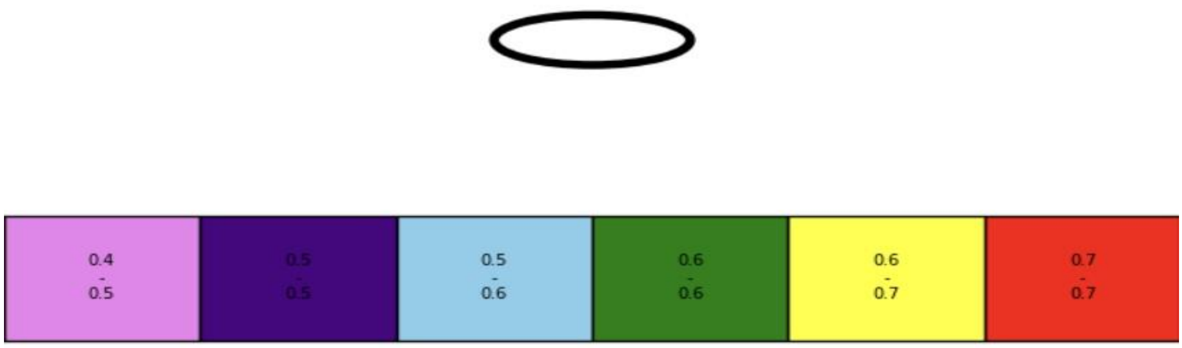

Figure 4: One interval of Education

3.2.3 Happiness Score → Mouth Shape

Range for Happiness is [1.7; 8.0]

Figure 5: All intervals of Happiness

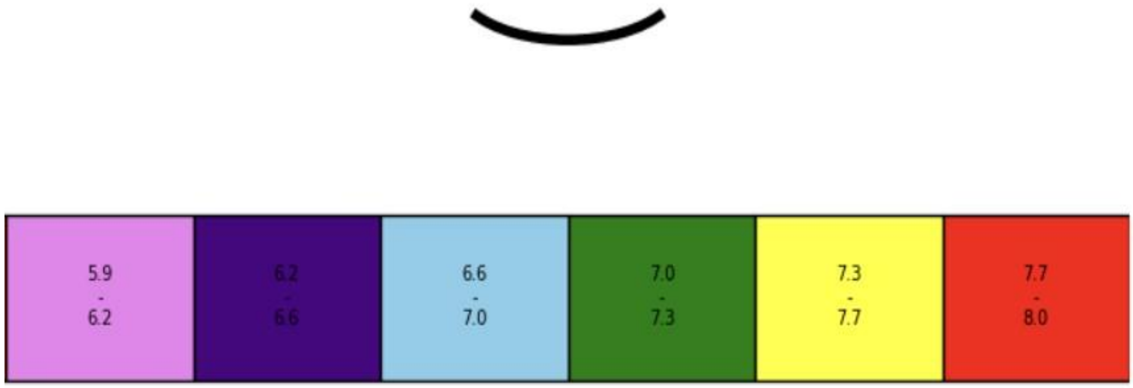

Figure 6: One interval of Happiness

3.2.4 Employment Rate → Nose Shape

Range for Employment is [%35;%85]

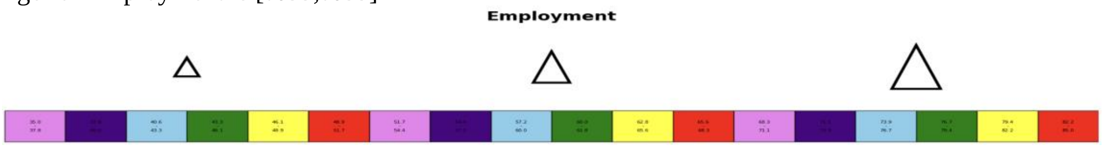

Figure 7: All intervals of Employment

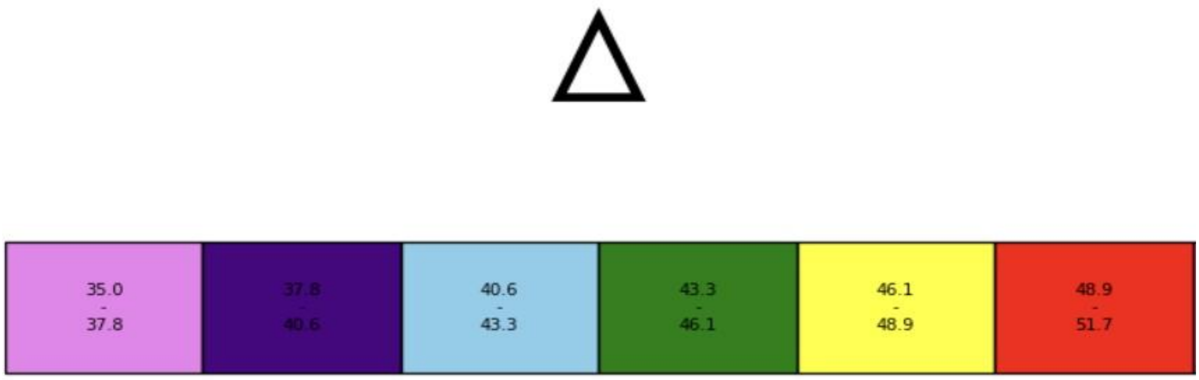

Figure 8: One interval of Employment

3.3 Step 3: Implementation and Interpretation Using Python

Creation of Chernoff faces was performed by Python via a procedure which converted basic socio-economic data into pictorial forms. In our example there was chosen four countries: Kazakhstan, Italy, Japan and Canada to visualize their GDP, Education index, Happiness score and Employment rate using Face Histogram.

```
import pandas as pd
import matplotlib.pyplot as plt
from matplotlib.patches import Circle, Ellipse, Arc, Polygon
```

After the division into groups, the system assigned every measurement to particular facial characteristics while following the Chernoff face model instructions. It has been observed that this linkage procedure operated inside Python by using logic rules and list positions, which connected every group type to visible parts like the outline of the head, optical organs, nasal feature, and oral opening. Each nation was changed into a distinct facial image that holds information for the four markers at the same moment because the code processed them together.

```
# SHAPE INTERVAL
# ------------------------------------------

shape_index = int((value - min_val) / main_interval_size)

if shape_index > 2:
    shape_index = 2

 # COLOR SUBINTERVAL
 # -----------------------------------------

 shape_start = min_val + shape_index * main_interval_size

 sub_size = main_interval_size / 6

 color_index = int((value - shape_start) / sub_size)

 if color_index > 5:
     color_index = 5

 color = COLORS[color_index]

 return shape_index, color
```

Software routines in Python produced the ultimate Chernoff faces, which created picture results where a single head belongs to one nation. Graphical results were made for the purpose of looking at many socio-economic dimensions together without the need to study the digits directly. Observers state that these pictures provide a way for the eye to see differences quickly when the user looks at the various socio-economic dimensions. Because the Python code uses specific visualization functions, the final Chernoff faces appear in a clear format that people can understand. It is often argued that the Chernoff faces allow for a faster comparison between the nations than standard tables of numbers.

Table 3:

Initial Multivariate Dataset

Country	GDP	Education	Happiness	Employment
Kazakhstan	14005	0.82	6.2	65.0
Italy	40226	0.89	6.4	60.1
Japan	32476	0.93	6.1	77.0
Canada	55209	0.94	6.9	73.2

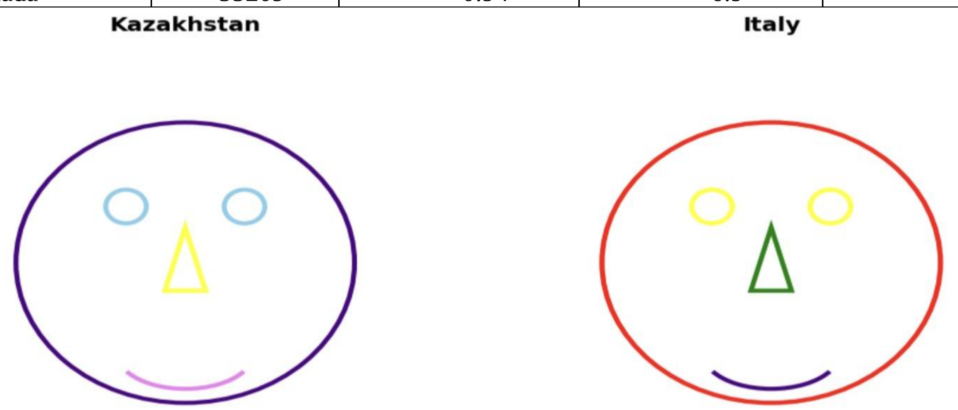

Figure 9: Face Histogram of Kazakhstan and Italy

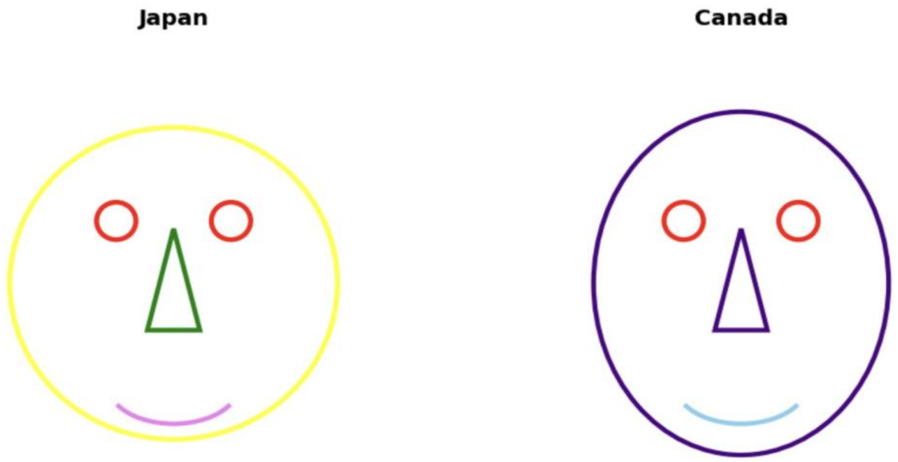

Figure 10: Face Histogram of Japan and Canada

The same steps can be done to any data, any size.

4 Conclusion

The Face Histogram method makes the long-standing Chernoff Faces larger by bringing together color mapping that stays within intervals and changes to parts of the face for showing many types of data. Experts claim that the Face Histogram method lets a person see many different pieces of information at the same time because one easy-to-understand picture is produced. Since the Face Histogram method uses the shape of a face to show data and allows several pieces of information to be shown in one image.

When compared to the old-style bar charts, the Face Histogram method makes finding patterns and looking at differences between things seen much better. Even though exact counting numbers are made smaller, the Face Histogram method works well for the activities of looking at data and showing things to students in an educational institution. Because scholars hope to do more in the coming days, work will be done on making the way features are mapped better and creating tools where people can touch and change the pictures. It has been observed that the Face Histogram method needs more ways to be understood easily by every person who looks at the data.

References:

1 Richard Johnson and Dean Wichern. *Applied Multivariate Statistical Analysis*. Sixth edition. Edinburgh Gate, 2014.

2 Robert Kosara. A critique of chernoff faces, 2007.

3 John Nelson. Chernoff faces, 2018.

4 Xiaochun Wang Ruixia Song, Zhaoxia Zhao. An application of the v-system to the clustering of chernoff faces, 2016.

5 Nazhan Yau. Crime chernoff faces by state, 2015.