Chernoff Faces

Group 3 Hedwig Höller, Matthias Eichhaber, Thomas Nuschy, and Christof Steinkellner

21 May 2013

Abstract

Glyphs can be used as a form to represent one-, two- and multidimensional data visually. The design aspect is crucial to this kind of visualization form. Studies have shown that the most important glyph features are colour, shape and size. This survey discusses representation forms such as Parallel Coordinates, Star Glyphs, Stick Figures and Faces with a focus on the latter. Chernoff Faces are a tool to represent multivariate data as faces with dataset variables mapped to different facial features. The original face proposed by Hermann Chernoff in 1973 uses up to 18 different facial features that can be assigned to variables, whereas the new faces introduced by Flury und Riedwyl can handle up to 36 parameters and allow asymmetrical mappings of facial features. Another derivation are the faces proposed by Kabulov and Tashpulatova which feature up to 20 dimensions and introduce new facial features like beards, haircuts and coloring. Different applications include Matlab, UNISTAT, APLPACK, Wolfram Mathworld and Datavis. Research has shown that Chernoff faces, while providing a good tool for a first look at data are generally unreliable due to the brain's inability to interpret faces pre-attentive and the mismatch created by viewers which interpret faces based on their own personal perception.

Contents

Chapter 1

Introduction

1.1   Introduction

This survey elaborates multiple data visualisation in the form of Chernoff Faces. First of all it will give a short introduction on glyphs. The Questions what are glyphs and which different techniques are state of the art will be discussed. Note: Important terms in this theme are glyphs and icons. In the literature both terms are commonly used and describe the same thing. This survey will use the term glyphs.

In chapter two starts at explaining what Chernoff Faces are and will give a brief history of the Faces' inventor Herman Chernoff. The following questions will be addressed: How and when Chernoff Faces got introduced? In what form can Chernoff Faces be represented? How many dimensions can Chernoff Faces represent?

In the third chapter the further development of Chernoff Faces will be mentioned and a comparison and a representation of Chernoff Faces will he outlined.

The last Chapter concludes with a few of common Chernoff Faces applications. Those applications which can draw Chernoff Faces will be listed and summarised briefly.

Chapter 2

Information Visualization, Glyphs and Glyph Design

2.1 Naming Convention

Our research has show there is no consistent naming convention for glyphs. The most common names are Glyphs and Icons. We decided to stick to the word glyph in our paper.

2.2 Basics of Information Visualization

The huge amount of data that is produced every year makes it necessary to find methods to explorer this data. With the help of Information Visualization techniques it is possible to explorer large data sets and make conclusions. This is because IV techniques make use of the perceptual abilities of the human. Humans are capable to identify differences in the colour, shape or size very easily.
Some advantages of Information Visualization are are:

- non-homogeneous and noisy data can easily be deal with

- it is intuitive and does not require complex algorithms or parameters

- fast data exploration

- often better results

When dealing with Information Visualization we conquer different types of data: one-dimensional, two-dimensional, multidimensional and text and hypertext. The difference of the data dimension is because of the fact that not every group of items can be described by the same number of parameters. In the following chapters we will describe the different types of data.

2.2.1 One-Dimensional Data

One-dimensional data is described by only one parameter. An example for such data is temporal data.

2.2.2 Two-Dimensional Data

Two-dimensional data is described by only two parameters. An example for such data is geographical data.

2.2.3   Multidimensional Data

Multidimensional data consist of more than three parameters. Therefore no simple visualization is possible. We would need a transformation from the n-dimensional two a space that humans can interpret but since there is no such transformation that could do that without the loss of important information more advanced information visualization techniques are necessary. An example for such information visualization techniques would be Parallel Coordinate Techniques.
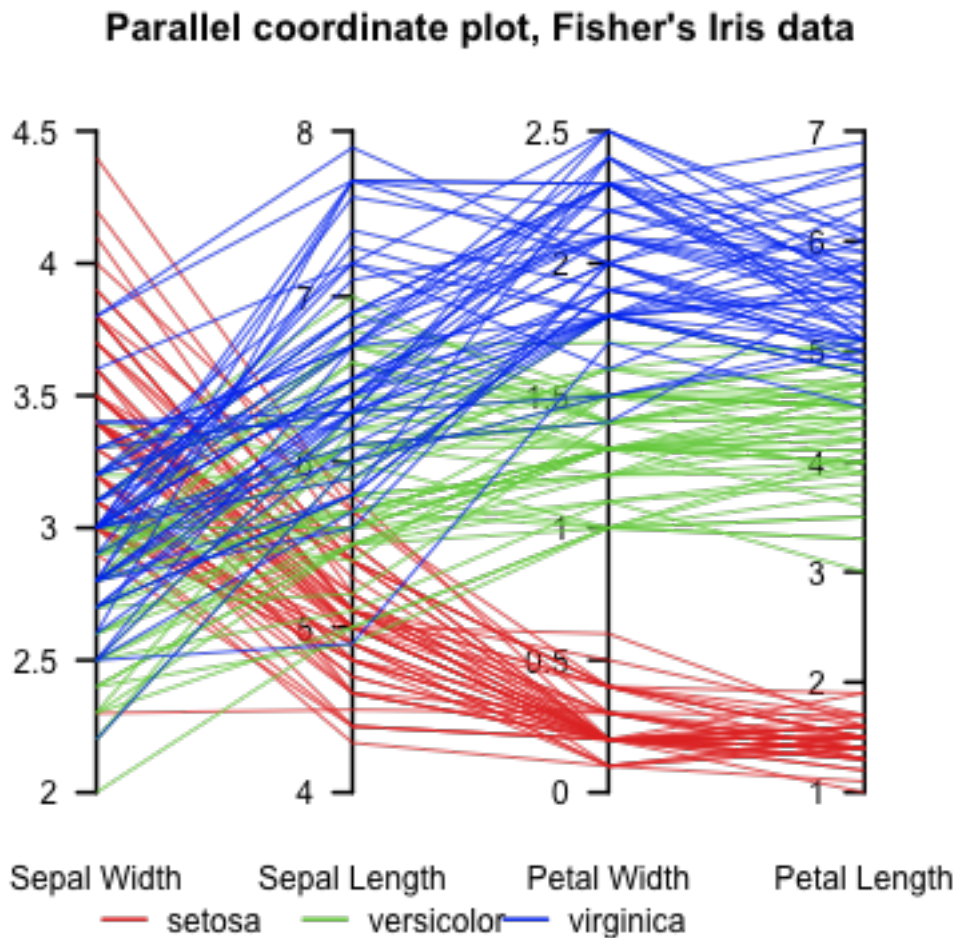

Figure 2.1: An example for Parallel Coordinates representing Multidimensional Data. `http://en.wikipedia.org/wiki/Parallel_coordinates`

2.2.4   Text and Hypertext

Some data cannot be represented in form of dimensionality. Some important of this type are text, hypertext and multimedia content. The problem with such data is that it is difficult to describe with numbers. Therefore we must find a way to transform it into a form that we can visualize. The simplest approach is to make a word-count-vector, where each word is a dimension and the total number of occurrence is the value of this dimension.

2.3 Why Glyphs?

Some data cannot be represented in form of dimensionality. Some important of this type are:

- text,

- hypertext and

- multimedia content.

The problem with such data is that it is difficult to describe with numbers. Therefore we must find a way to transform it into a form that we can visualize. The simplest approach is to make a word-count-vector, where each word is a dimension and the total number of occurrence is the value of this dimension.

Category	Technique	Tasks				
		overview	correlation	patterns	clustering	outlier
1D to 3D	Histogram			X		X
	Box Plot			X		X
	Scatter Plot		X	X		X
	Contour Plot		X			
Icon-based	Chernoff Faces			X	X	X
	Star Glyphs			X	X	X
	Stick Figure				X	
Geometrically transformed displays	Scatter Plot Matrix		X	X		X
	Parallel Coordinates	X	X	X		X
Pixel oriented	Query-independent techniques		X	X	X	
	Query-dependent techniques		X	X	X	
Graph-based orHierarchical	Graph	X	X			
	Cone Tree	X				
	Treemap	X	X	X	X	
	Dimensional Stacking			X	X	X
	Mosaic Plot	X		X		

Figure 2.2: This table shows the best uses of differen information visualization techniques. [**Maria_Madalena_Dias2012-09-12**]

2.4 Types of Glyphs

There are several types of glyphs. In the following chapters we will describe the most important glyphs. We will start with Faces, and then continue with stick figures, after them we will write about star glyphs and finally we will describe star glyphs.

2.4.1 Faces

Faces where first introduced by H. Chernoff. Usally they have 18 parameters that can be manipulated. Therefore it is possible to represent data with 18 dimensions easily. We will discuss this glyph type in more detail later in this paper. In figure 2.3 you can see an exmaple of Chernoff faces.
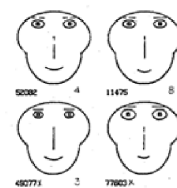

Figure 2.3: Faces like they where introduced by H. Cernoff. [Chernoff, 1973a]

2.4.2   Stick Figures

Stick figures consist of n lines. Each of these lines and the angles between the lines can be mapped to a parameter (see figure 2.4). Often there is one base line that cannot be mapped to a parameter and the other lines are docked to this line. When represented in an coordination system two parameters are mapped to the axis of the system (see figure 2.5 and figure 2.6).
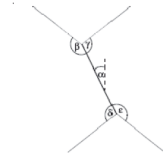

Figure 2.4: Basic stick figure scheme that shows which angles can be manipulated. [Ronald, 2007]
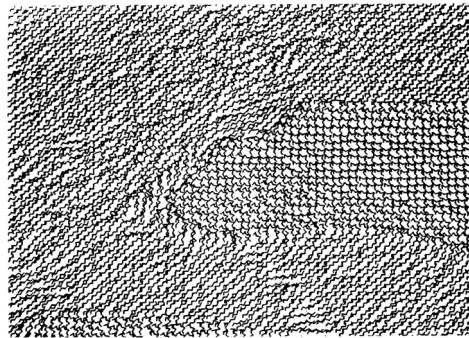

Figure 2.5: An example of a stick figure diagram that shows weather data. The position of the figure corresponds to the geographical coordinates. [Pickett and Grinstein, 1988]
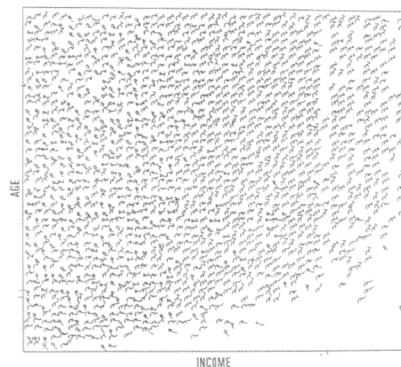

Figure 2.6: Another stick figurer example showing the income, gender, age, education etc. It can be seen that their are a view outlieres that are of young age and have a high income. [Ronald, 2007]

2.4.3   Star Glyphs

Star glyphs represent data in a shape that resemble stars. To get this shape each dimension is mapped to one direction. The value of the mapped parameter is representing for the length of the vector in this direction. Some examples of star glyphs can be seen in figure 2.7.
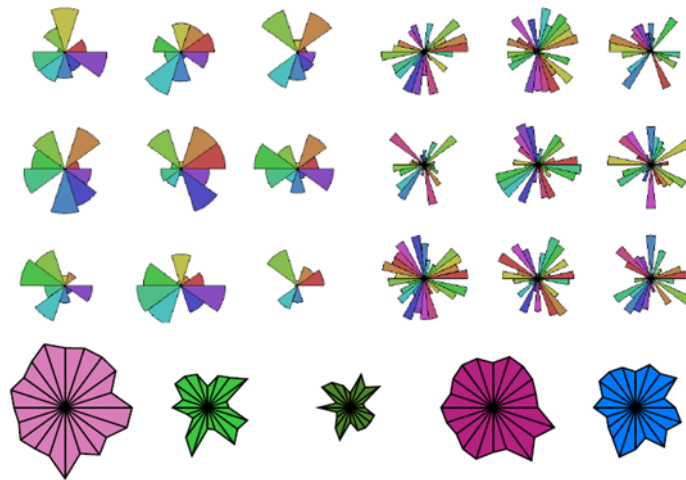
Figure 2.7: Examples of Star glyphs. [Fanea, Carpendale, and Isenberg, 2005] [**Maria Madalena Dias2012-09-12**]

2.4.4 Multi Glyphs

Multi glyphs are glyphs that consist of more than one glyph. Each of these glyphs can be manipulated separately. With multi glyphs it is possible to represent high dimensional data without getting unreadable glyphs.

2.5 Glyph Design

The glyph is an important part of glyph visualization because a bad glyph design results in a slow or ineffective distinction. It is very difficult to design meaningful glyphs. The common visual features that describe a glyph are:

- location,
- size,
- colour,
- shape,
- orientation,
- texture and
- brightness.

Out of these the most important are size, colour and shape. Different studies have shown that the ranking of these features is not absolute and varies for nominal and quantitative data. The rankings that where found can be seen in Table 2.1 and table 2.2. [Nowell, Schulman, and Hix, 2002]

	Mackinlay	Christ	Christ Ident.
Colour	1	1	1
Shape	2	3	3
Size	3	2	2

Table 2.1: Ranking of features of glyphs for nominal data.[Nowell, Schulman, and Hix, 2002]

	Christ Search Time	Mackinlay	Cleveland & McGill
Colour	1	2	2
Shape	3	3	3
Size	2	1	1

Table 2.2: Ranking of features of glyphs for quantitativ data.[Nowell, Schulman, and Hix, 2002]

Figure 2.8: Overview of different design options for multiglyphs. [**taxonomy-glyph-2012**]

Chapter 3

Chernoff Faces

3.1 Chernoff Faces Motivation

Herman Chernoff published the paper "The Use of Faces to Represent Points in k-Dimensional Space Graph-ically" in the Journal of the American Statistical of 1973. The content of this paper is the idea of displaying statistical data in a way of smiley-faces.

In an interview with Bather [Bather, 1996] he tried to explain how he got the idea to represent statistic data with human faces. At that time he was dealing with cluster analysis and pattern recognition. He said that the difficulty for him had been in dealing with multidimensional data. Due he could only represent them in a 2 dimensional way, instead of visualize the points of interest in an n-dimensional space, he started drawing faces instead.

The first reason of his motivation was based on the wish to describe n-dimensional statistic data in a 2-dimensionsal way. He wanted some visualization, which helps detecting patterns and correlations of data.

The second requirement was that he wanted to display the data in a human friendly readable way. Based on the fact that human beings are very good at recognizing faces he started drawing smiley-faces.

He began to work out an approach to convert any statistic data set into a bunch of face drawings. He defined 18 different variables, which are assigned to one facial characteristic (eyes, ear, nose, mouth and things like that) each and which are responsible for the variety of faces.

Chernoff discovered, by experiment, that people could comfortably interpret a face that expresses quite large amounts of data. Humans recognize even very small differences in faces.

He assumed that if humans are so good at interpreting facial expressions, they might be able to adopt this skill in interpreting graphs represented in such a way as well.

3.2 Herman Chernoff

Herman Chernoff was born on July 1, 1923 in New York. His competences are mathematics, statistics and physics. He graduated College in January 1943 and started as junior physicist with the U.S. Navy during the war. After his graduation in mathematics at the Brown University in 1945 with a Master Degree in mathematics he focused on statistic, where he completed his Ph.D theses at the Columbia University. He was a professor at Stanford University and founded the Statistics Center in 1974 at the MIT. Then he began working as professor in the Department of Statistics at Harvard where he is still teaching.

He invented the Chernoff faces, which are the most famous way in display multivariate data represented as icon/glyph. In 1973 he announced his idea with the Chernoff faces in a monograph called "The Use of Faces to Represent Points in K-Dimensional Space Graphically." [Chernoff, 1973b]
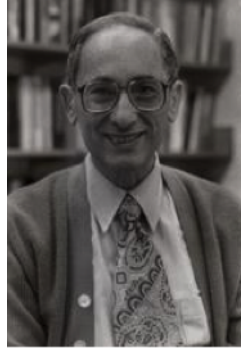
Figure 3.1: picture of hermann chernoff

3.3 Description of Chernoff Faces

Chernoff Faces were first introduced in 1973 as a method to visualize multivariate data in a way that is easily comprehendible by investigators. By representing variables of a dataset as different features of a face, small changes in the data set appear as changes in facial expression. Transitions and clusters are then easier detected and memorized. In figure 3.2 such transitions and clusters can be observed. The faces represent 36 mineral analysis datasets from a 4,500-foot core drilled from a Colorado mountainside[Chernoff, 1973b]. Each dataset is equally spaced from the core and listed sequentially. Transitions starting at specimen 220 and specimen 232 can be observed, as well as clustering of datasets from specimen 224 to specimen 231, which is indicated by a red margin.
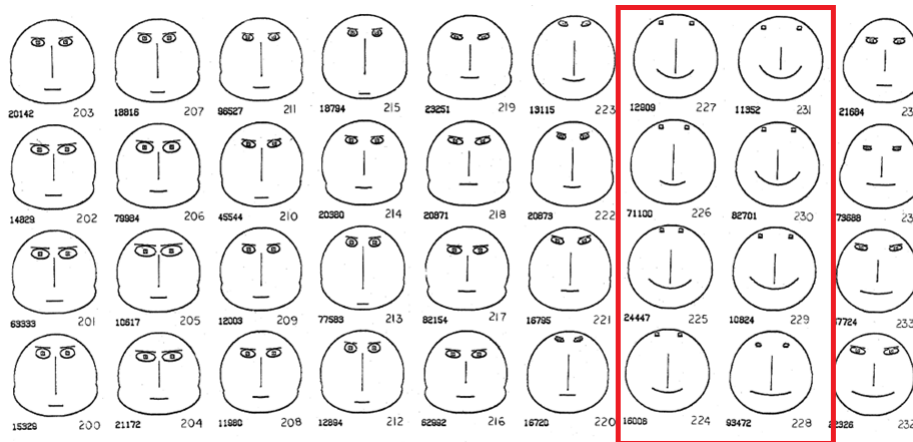

Figure 3.2: Chernoff Face Construction Example (taken from [Chernoff, 1973b])

3.4 Construction of Chernoff Faces

The original face was proposed by Herman Chernoff in 1973 and uses 18 variables x_1, x_2,... x_{18} in appropriate ranges (0...1). The different facial features are as follows[Chernoff, 1973b]:

x_1 Chernoff1973 radius to corner of face, |OP|

x_2 angle of OP to horizontal

x_3 vertical size of face, |OU|

x_4 eccentricity of upper face

x₅ eccentricity of lower face

x₆ length of nose

x₇ vertical position of mouth

x₈ curvature of mouth

x₉ width of mouth

x₁₀ vertical position of eyes

x₁₁ separation of eyes

x₁₂ slant of eyes

x₁₃ eccentricity of eyes

x₁₄ size of eyes

x₁₅ position of pupils

x₁₆ vertical position of eyebrows

x₁₇ slant of eyebrows

x₁₈ size of eyebrows

The outline of the face consists of two ellipses. From the center point O a ray is being drawn to a corner point P. On the opposite side, a second corner point P' is taken so that OP and OP' are both of the same length and vertically symmetric. The top and bottom of the face, represented by U and L respectively, are being selected such that OU and OL are vertical and have equal length. The upper part of the face is an ellipse which is determined by U, P, P' and x_4. The lower part is similarly defined by L, P, P' and x_5. The nose is a vertical line through O whereas the mouth is a circular arc. Eyes are represented as ellipses with pupils that are both located at the same horizontal distance from the center of the eyes. Eyebrows are slanted line segments above eyes. Both eyes and eyebrows are vertically symmetric with respect to the vertical line through O. A normalization of the width and height of the face to two inches are applied subsequently to diminish the effects of x_1 and x_3. Except for the three eccentricities and the curvature of the mouth all other parameters are then normalized linearly to the range 0 to 1. An example face with the mentioned corner points can be seen in figure 3.3 below.

3.5 Critiques of Chernoff Faces

Chernoff Faces have often been subject to criticism over the years. Its effectiveness of displaying data in a meaningful and easily comprehensible way is still open to speculation and subject of studies. The results of Christopher J. Morris', David S. Ebert's and Penny Rheingans' studies indicated that facial feature perception is not, as widely assumed, pre-attentive, but rather a serial process[Morris, Ebert, and Rheingans, 2000]. This would suggest that Chernoff Faces do not take advantage of the human brain's capability to recognise faces subconsciously. Moreover, the personal perception of how to interpret a certain type of face can be extremely nonlinear [Ware, 2004]. Different viewers can interpret the same face in different ways due to their different perceptual mappings. In addition with the shortcomings discussed in the next section, Chernoff Faces make for an unreliable tool to view multivariate data and is only usable for providing a first look to narrow down the area of interesting datasets.
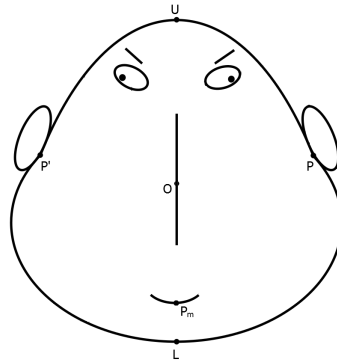
3.6   Comparison with other graphical visualisations

In this section we compare Chernoff Faces with other forms of graphical visualisation. The weaknesses and
strengths of Chernoff Faces will be discussed also. Most times Chernoff Faces are used to cluster datasets or to
detect outliers, so we will confine the comparison to those two aspects. At first the detection of outliers will be
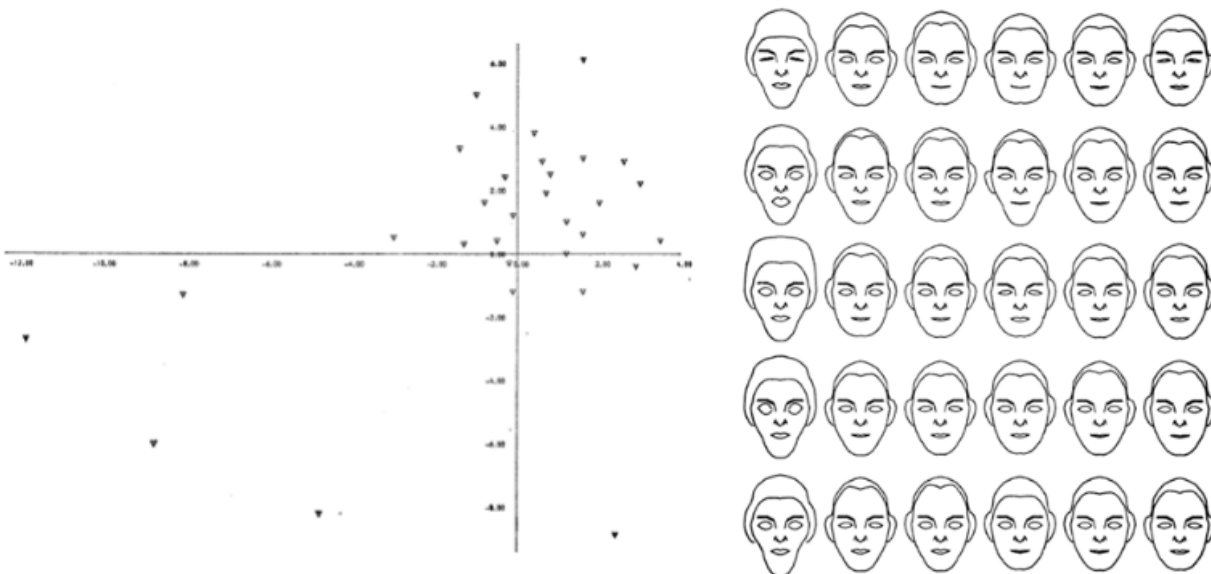discussed. An example with outliers in Non-linear Mapping and Chernoff Faces can be seen in figure 3.4:

Figure 3.4: Example of outlier detection (taken from [Everitt and Nicholls, 1975])

As can be seen in this figure outlier detection is quite easy with Chernoff Faces. Although the outlier de-
tection is quite easy in this example, because all outliers are on the left side of the table and the parameters are
well chosen. If the outliers were randomly distributed on the table, it wouldn't be so easy to see them at the
first glance.
The second case we will discuss is the detection of clusters in a dataset. An example of clustered data repre-
sented by Non-linear Mapping and Chernoff Faces is given in figure 3.5:
The plot on the left side clearly shows three clusters. On the right side in the Chernoff Faces plot it's a little
bit harder to see the three clusters, but it is still possible. However if the Faces were not in this specific order, it
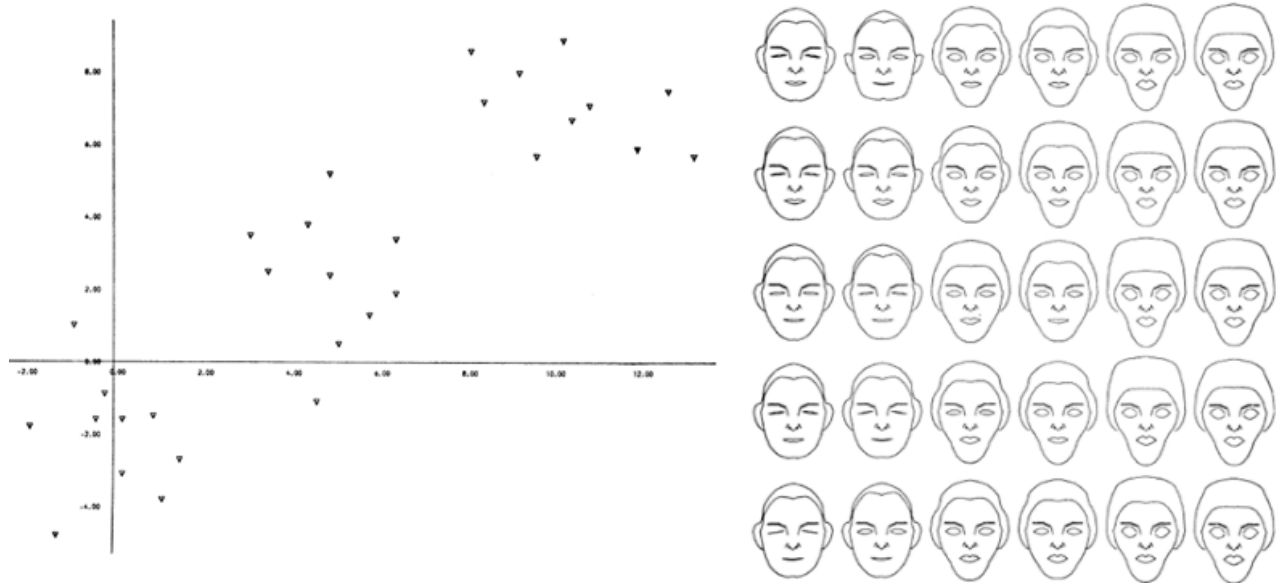
Figure 3.5: Example of cluster detection (taken from [Everitt and Nicholls, 1975])

would be a lot harder to see the clusters.

Recapitulatory we can observe, that the cluster detection and the outlier detection need a good ordering in the table of the Chernoff Faces to see the desired features properly. Due the fact that a representation of Chernoff Faces is rarely accompanied by a legend, Chernoff Faces are most times used for the two described scenarios above only, because without a legend it is hardly possible to read anything from a Chernoff Face.

3.7 Further development of Chernoff Faces

Since Herman Chernoff postulated his idea of information visualization with the so called Chernoff Faces in 1973 researchers adopted and refined those faces. These researchers pointed out some shortcomings of the original Chernoff Faces and tried to eliminate those problems with their refined faces.

A widely known problem of the original Chernoff Faces is that if certain parameters are close to their extreme values, they change all other parameters as well. The other parameters might be still the same, but their graphical representation in the form of Chernoff Faces will change nonetheless. This problem is easiest illustrated with extreme values for the nose-length. In figure 3.6 there is an illustration of this example. While only the value for the length of the nose was changed, six other parameters of the Chernoff Face were influenced by this value. This example with the nose-length might be the example with the most apparent effect of this shortcoming, but the problem exists with other parameters as well.

Another critic on Chernoff Faces is their cartoon-like appearance. A main reason for developing the Chernoff Faces was that the human brain is very good at distinguishing human faces, which have only very slight differences. However the faces invented by Herman Chernoff do not resemble human faces very much. An observer will see that those images of Chernoff Faces are representing human faces, but they definitely do not look like any real person's face. So some researchers tried to give their faces a more humanly look [Flury and Riedwyl, 1981] [Kabulov and Tashpulatova, 2010].

Several scientists wanted more parameters in their enhanced Chernoff Faces. The right and left parts of Chernoff Faces are symmetrical. Hence an observer only has to look into one half of the face to see all parameters. Tue to this fact half of the total space is wasted, since one half is just a mirrored copy of the other half. In the following two examples of these more advanced Chernoff Faces are described.

Flury and Riedwyl invented their so called "new Faces" in their paper "Graphical Representation of Multivari-
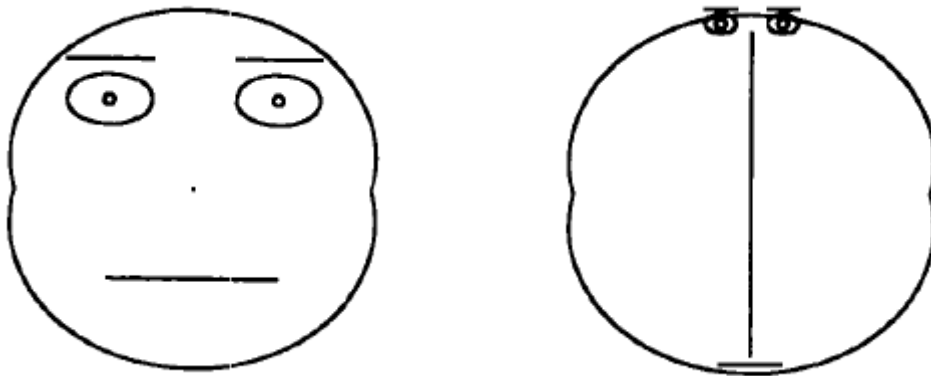
Figure 3.6: Effect of Changing the Parameter Length of Nose in a Chernoff Face taken from [Flury and Riedwyl, 1981]

ate Data by Means of Asymmetrical Faces" [Flury and Riedwyl, 1981]. They used asymmetrical faces to avoid the problem with double values in the original symmetric Chernoff Faces. The asymmetric version of their faces allows up to 36 different data points included into the resulting face. The faces of Flury and Riedwyl allow both symmetrical and asymmetrical representations as can be seen in figure 3.7.
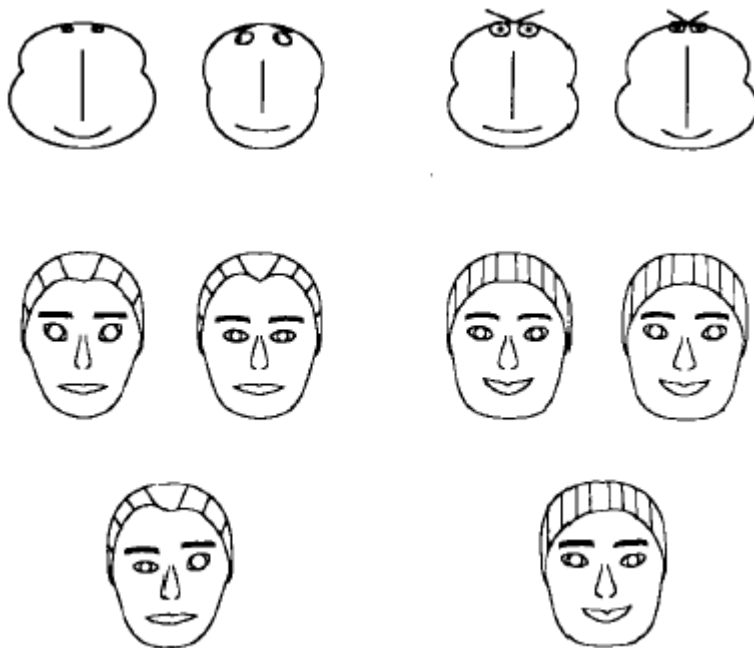

Figure 3.7: Comparison of Chernoff Faces with the symmetric and asymmetric Faces of Flury and Riedwyl (taken from [Flury and Riedwyl, 1981])

In figure 3.7, Chernoff Faces are compared to the symmetric and asymmetric representation of Faces designed by Flury and Riedwyl [Flury and Riedwyl, 1981]. The asymmetric representation is just the two symmetric Faces joined into one Face. The main goals of these new faces are more possible parameters and a better representation which is more tolerant to extreme parameter values than the original Chernoff Faces. The Effect seen in figure 3.6 wouldn't be so extreme with this faces as can be seen in figure 3.8 were both the face with all minimum values and the face with all maximum values are displayed.
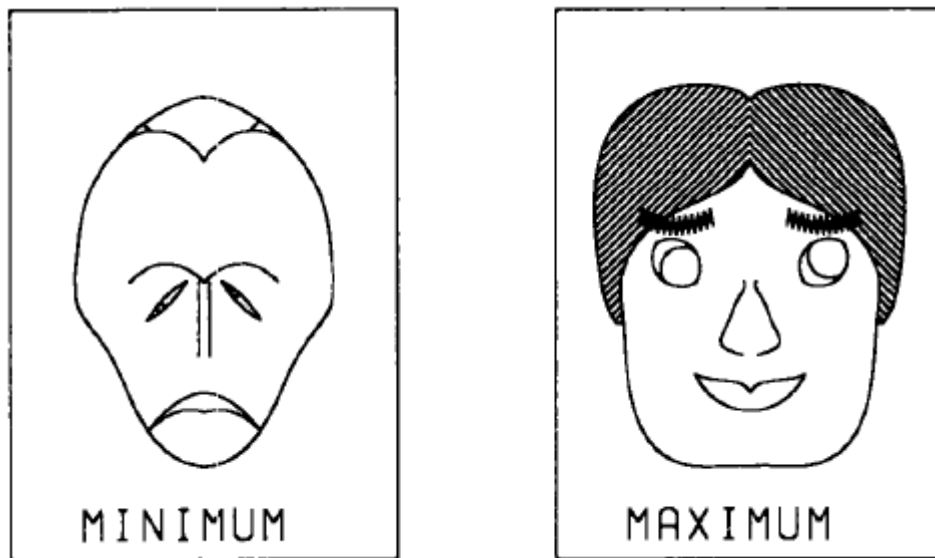
Figure 3.8: Minimum and Maximum Faces (from [Flury and Riedwyl, 1981])

Another extension to Chernoff Faces was done by Kabulov and Tashpulatova in [Kabulov and Tashpulatova, 2010]. They also tried to get more parameters into their so called "enhanced Chernoff Faces" by adding more facial features like beards and hair, color coding to highlight the variable's values and asymmetric features as well. Tue to their enhancements up to 20 dimensions can be displayed with a single face [Kabulov and Tashpulatova, 2010]. An example of their enhanced faces can be seen in figure 3.9.

All further developments of Chernoff Faces try to get rid of the shortcomings of the original Chernoff Faces. The most requested features seem to be more possible parameters and a better visualization of the faces.
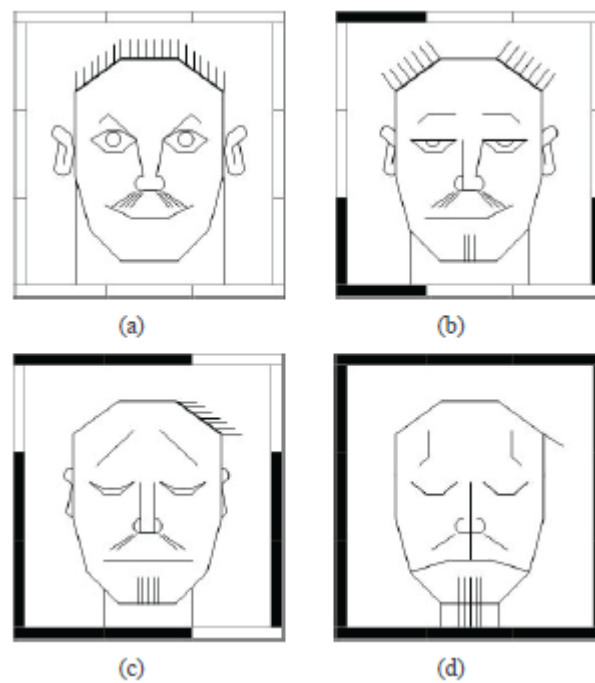
Figure 3.9: Enhanced Chernoff Faces (taken from [Kabulov and Tashpulatova, 2010])

Chapter 4

Chernoff Faces Applications

4.1 Chernoff Faces Applications

In this section the paper will list some applications which can draw chernoff faces.

4.1.1 MATLAB and Simulink

Matlab is commercial software by the MathWorks, Inc. Company. This software is used to solve mathematical problems and graphical representation of mathematical things. It also supports drawing of Chernoff Faces. The glyphplot function encodes the data values into facial features.
It is possible to represent 17 dimensions:

- Size of face

- Forehead/jaw relative arc length

- Shape of forehead

- Shape of jaw

- Width between eyes

- Vertical position of eyes

- Height of eyes

- Width of eyes (this also affects eyebrow width)

- Angle of eyes (this also affects eyebrow angle)

- Vertical position of eyebrows

- Width of eyebrows (relative to eyes)

- Angle of eyebrows (relative to eyes)

- Direction of pupils

- Length of nose

- Vertical position of mouth

- Shape of mouth

- Mouth arc length

Using the software:

- 30-day free test version

- free student test version

- very easy handling

- Tutorial: `http://www.mathworks.de/de/help/stats/glyphplot.html` (retrieved 2013-05-13)

In 4.1 you can see an exmaple of MATLAB. Image-Source: `http://www.mathworks.de/products/statistics/examples.html?file=/products/demos/shipping/stats/mvplotdemo.html` (retrieved 2013-05-13)

```
glyphplot(X(models77,:), 'glyph','face', 'centers',Y, ...
          'varLabels',varNames, 'obslabels',Model(models77,:));
title('1977 Model Year');
```
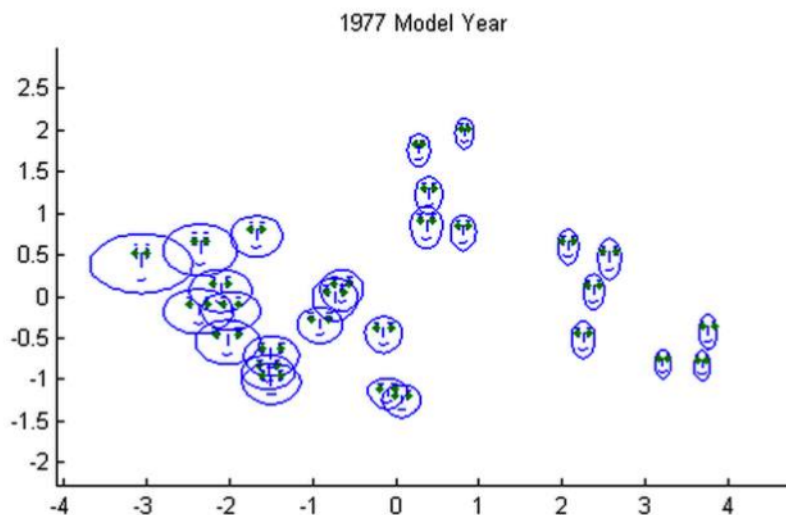

Figure 4.1: Matlab and Simulink, Image-Source: `http://www.mathworks.de/products/statistics/examples.html?file=/products/demos/shipping/stats/mvplotdemo.html` (retrieved 2013-05-13)

4.1.2   UNISTAT

Unistat is a statistical software package featuring a Microsoft Excel add-in. Besides a lot of other features for data analysis, charting and data presentation it enables the user to draw Chernoff Faces as well. This application draws a Chernoff Face for each row in an provided matrix. Features on the face represent the relative value of variables.
Unistat provides 18 dimensions:

- Curvature of mouth

- Angle of eyebrow

- Eccentricity of upper face

- Eccentricity of lower face

- Length of nose

- Height of centre of mouth

- Position of pupils

- Length of mouth

- Height of eyes

- Separation of eyes

- Face height

- Face width

- Nose width

- Eccentricity of eyes

- Length of eyes

- Radius of ears

- Length of eyebrows

- Height of eyebrows

Using the software:

- Microsoft Excel Add-in : produces output to Excel or Word

- Free license

- Only for Windows

- Download: `http://www.unistat.com/download`

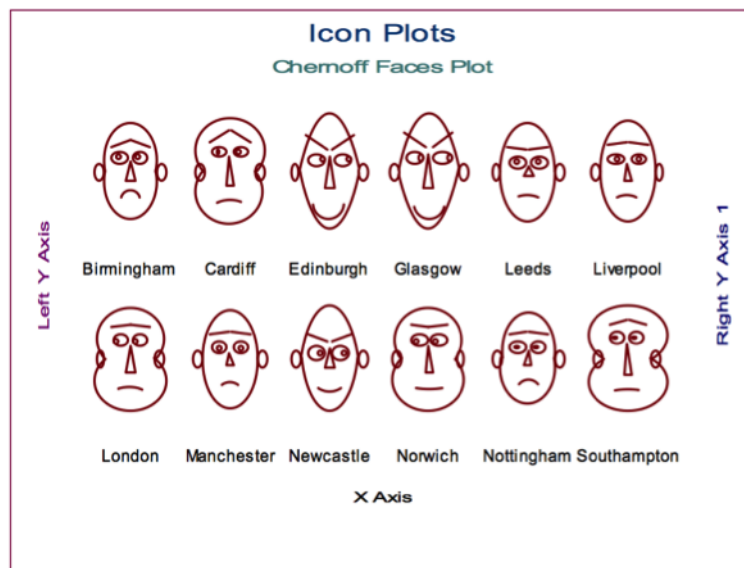In 4.2 you can see an exmaple of UNISTAT. Image-Source: Unistat 6.0 User Guide (PDF)

Figure 4.2: Unistat, Unistat 6.0 User Guide (PDF)

4.1.3 APLPACK

APLPACK is the acronym for "Another PLotting PACKage". It is a package for the Software R. R is a statistic software for data manipulation, calculation and graphical displaying. APLAPACK provides some functions for drawing plots, containing Chernoff Faces ('faces').
It can deal with 15 Dimensions:

- Height of face

- Width of face

- Structure of face

- Height of mouth

- Width of mouth

- Smiling

- Height of eyes

- Width of eyes

- Height of hair

- Width of hair

- Style of hair

- Height of nose

- Width of nose

- Width of ear

- Height of ear

Using the software:

- Free Software

- Tutorial: `http://cran.r-project.org/web/packages/aplpack/aplpack.pdf`

- Easy: Load csv file, draw face

Sample Code:

```
test <- read.csv("test123.csv")
test[1:6,]
library(aplpack)
faces(test[,2:8])
```

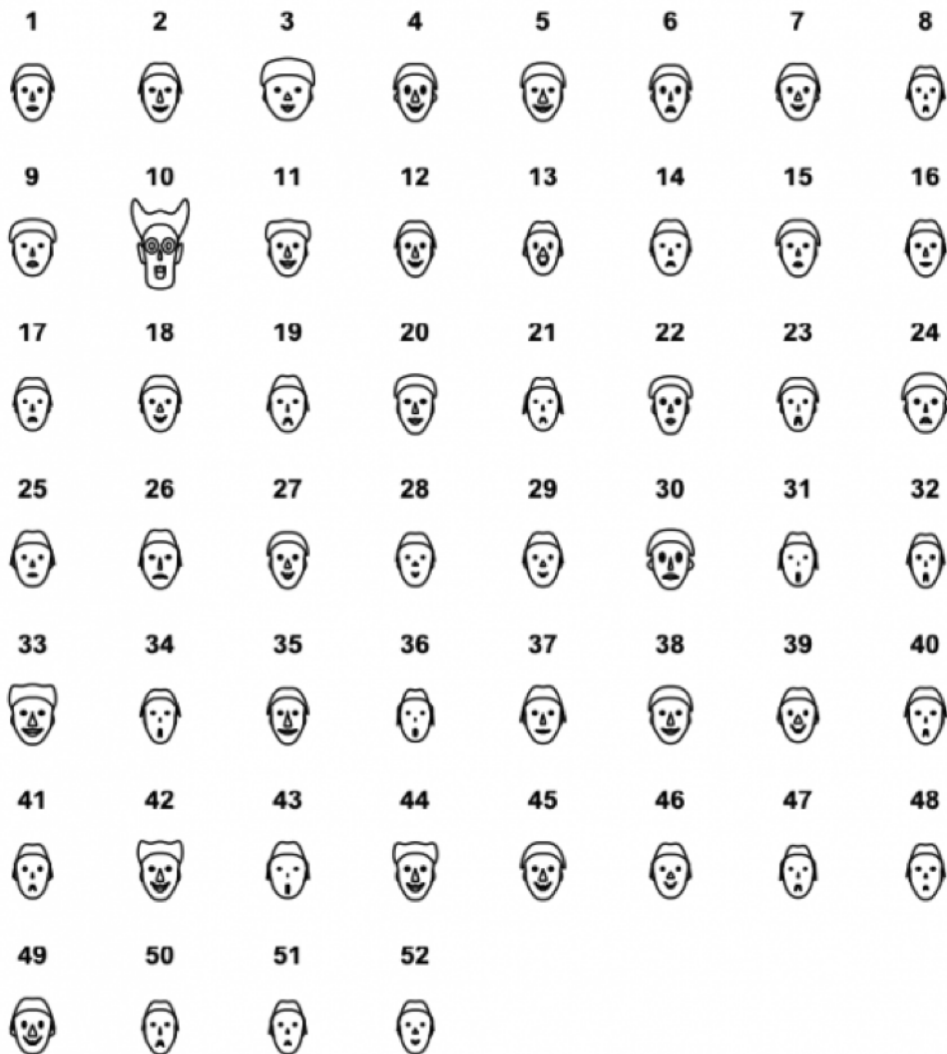In 4.3 you can see an exmaple of APLPACK.Image-Source: `http://flowingdata.com/2010/08/31/how-to-visualize-data-with-cartoonish-faces/` (retrieved 2013-05-13)

Figure 4.3: example in APLPACK, Image-Source: `http://flowingdata.com/2010/08/31/how-to-visualize-data-with-cartoonish-faces/` (retrieved 2013-05-13)

4.1.4   Wolfram MathWord

MathWord Statistical Plots supporting drawing Chernoff Faces. A free Mathematica plugin (ChernoffFace.nb) is necessary to draw the faces with that application. Mathematica is commercial software for Data Science, Finance, Modelling and lots of other mathematical based solutions.
The plugin can deal with 10 dimensions:

- head eccentricity

- eye size

- eye spacing

- eye eccentricity

- pupil size

- eyebrow slant

- nose size

- mouth shape

- mouth size

- and mouth opening

Using the software:

- Commercial Software

- Students of TUGRAZ can get a free Mathematica Licence in case of no commercial use: `http://portal.tugraz.at/portal/page/portal/zid/software/student/mathematica` (retrieved 2013-05-13)

- Download: `http://mathworld.wolfram.com/ChernoffFace.html`

In 4.4 you can see an exmaple of the Wolfram MathWord application.Image-Source: `http://mathworld.wolfram.com/ChernoffFace.html` (retrieved 2013-05-13)
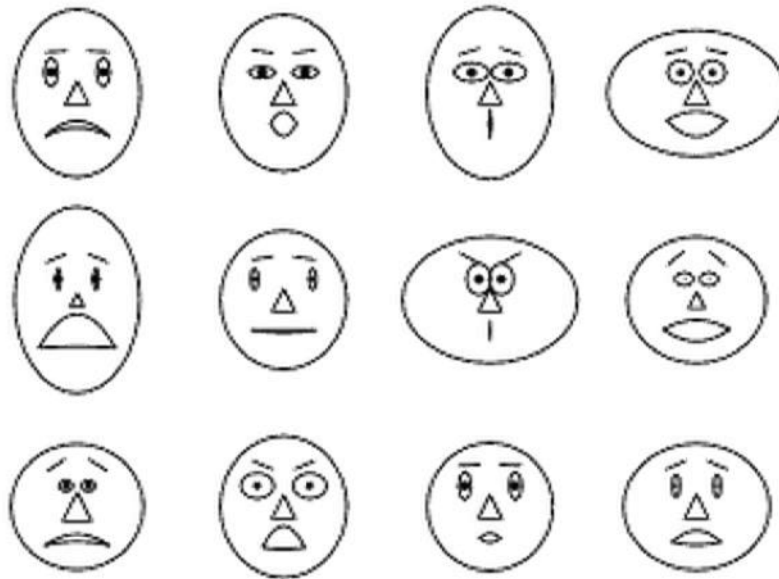
4.1.5 Wolfram Alpha Emoticon

Emoticon is a model which shows how you can experiment with a lot of emotions using only two parameters and some simple graphics primitives. It is based on Chernoff Faces and gives a good introduction /feeling in how data can be drawn as faces.

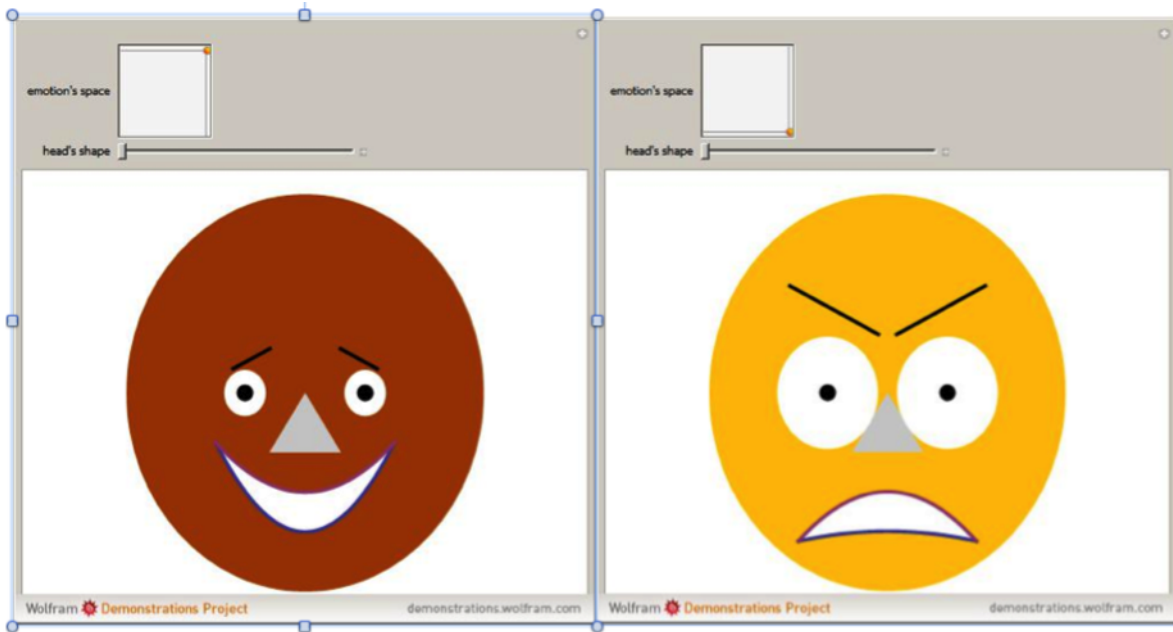In 4.5 you can see an exmaple of the Wolfram Alpha Emoticon application. Image-Source:`http://demonstrations.wolfram.com/Emoticon/` (retrieved 2013-05-15)

Figure 4.5: example in Wolfram Emoticon, `Image-Source:` `http://demonstrations.` `wolfram.com/Emoticon/` (retrieved 2013-05-15)

4.1.6 Datavis

Datavis provides a SAS macro program 'faces' which can draw asymmetric Chernoff Faces and can deal with 18 dimensions:

- Eye size

- Pupil size

- Position of pupil

- Eye slant

- Horizontal position of eye

- Vertical position of eyebrow

- Upper hair line

- Lower hair line

- Face line

- Darkness of hair

- Hair shading slant angle

- Nose line

- Size of mouth

- Curve ature of mouth

Using the software:

- Free plugin

- Download and Tutorial: `http://www.datavis.ca/sasmac/faces.html`

In **??** you can see an exmaple of the DATAVIS application. Image-Source:`http://demonstrations.wolfram.com/Emoticon/` (retrieved 2013-05-15)
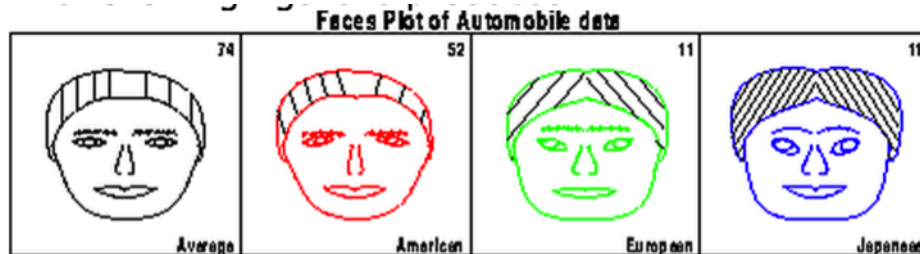

Figure 4.6: example in Datavis

4.1.7 Faces 2.2 Screensaver

Faces 2.2 is a screensaver module for Mac OS X that generates and displays random Chernoff faces. It is possible to change colour and frequency of it to get an individual screensaver. It is just for fun, free and useful to get a feeling for Chernoff faces.
Download: `http://bradandkathy.com/software/faces.html`
In 4.7 you can see an exmaple of the Faces2.2 screensaver.

Figure 4.7: example in Faces2.2

4.2 Representation of Chernoff Faces

The different Chernoff Faces gained by a set of data must be represented in a certain way. Most times Chernoff Faces are just represented in a table. This kind of representation can be seen in Figure 4.8. But there are other types of representations of Chernoff Faces as well, which will be described later in this section.
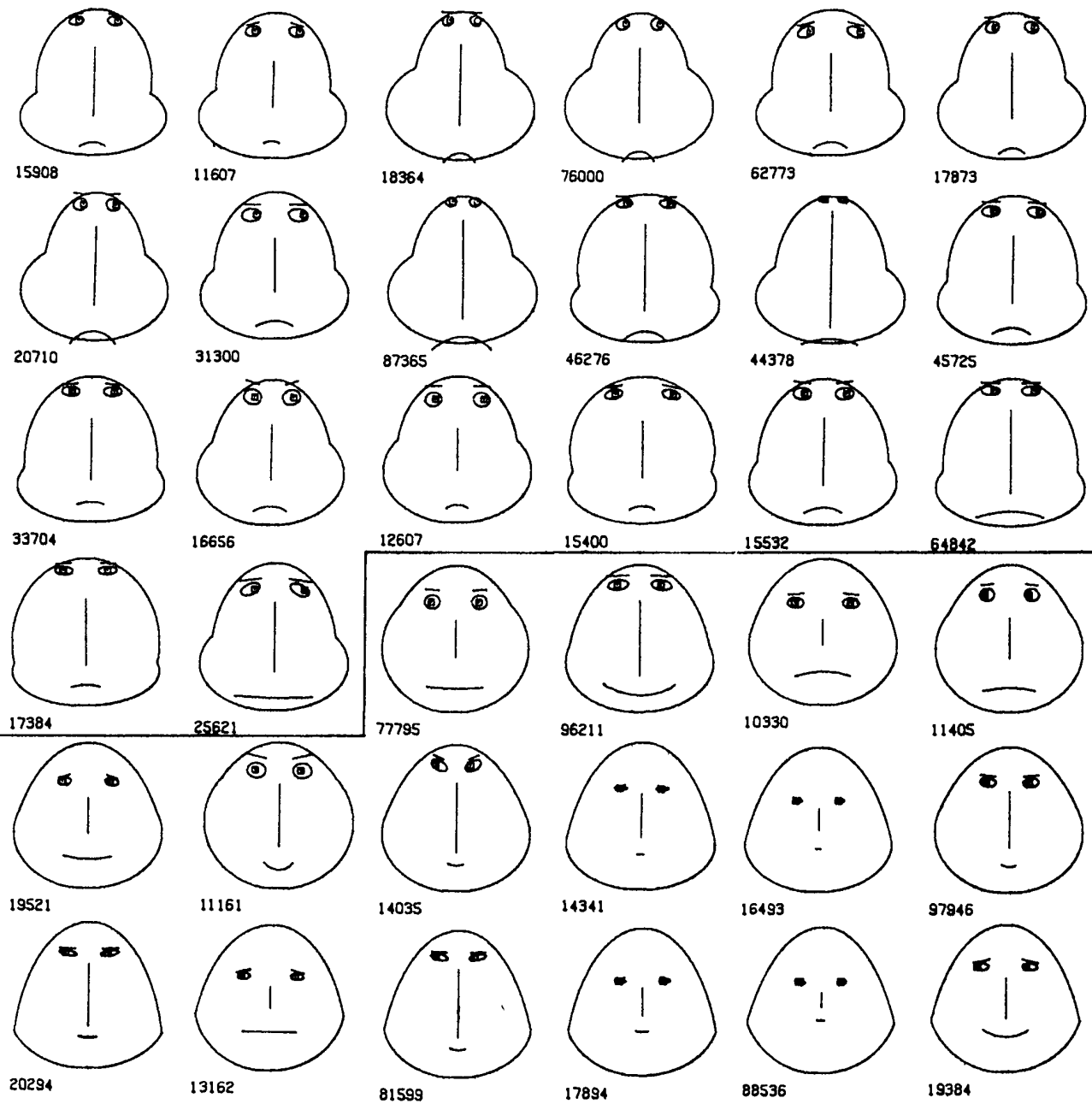
Figure 4.8: Chernoff Faces in a Table (taken from [Chernoff and Rizvi, 1975])

In Figure 4.8 like in many other representations of Chernoff Faces in a table there is no legend to understand how the faces could be interpreted. However the Faces seem to be ordered in a certain way, because the Chernoff Faces on the top of the table have broader lower half of the face and the Faces on the bottom of the table have a more circular face. This observation is another characteristics of the table representation. All those representations are seemingly ordered, so that the observer can easily see clusters, outliers and things like that. Another representation of Chernoff Faces can be done with maps. In that case the Chernoff Faces are positioned on a map to represent certain characteristics in the designated area. An Example of this kind of representation can be seen in figure 4.9.
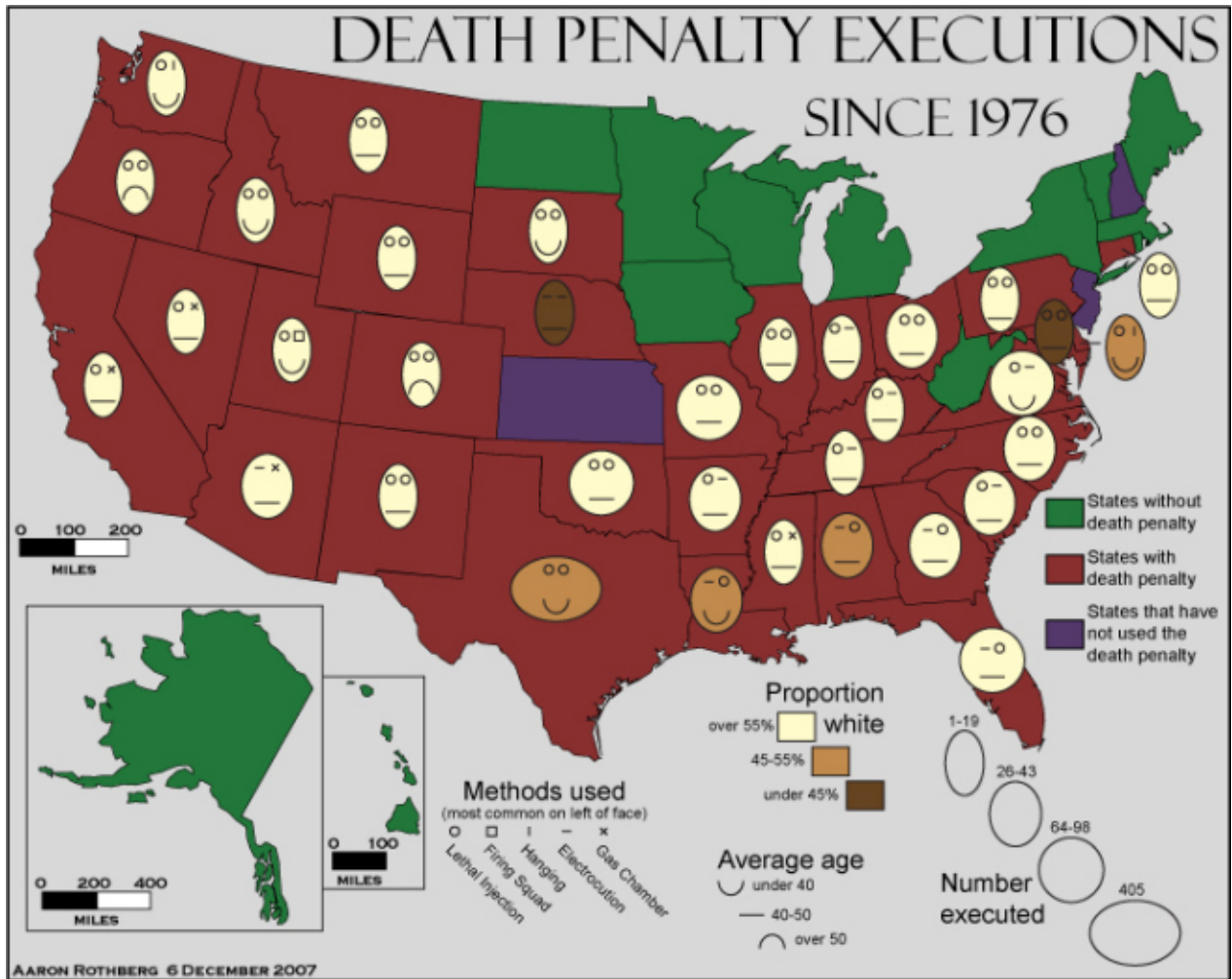
Figure 4.9: Example of Chernoff Faces on maps (taken from http://cartastrophe. wordpress.com/2010/06/16/on-the-abuse-of-chernoff-faces/ (retrieved 2013-05-13))

In this figure the statistical data of the death penalty executions since 1976 in the USA can be seen represented as Chernoff Faces. There is also a legend to understand the meanings of the Chernoff Faces. In this kind of representation a legend is used quite often. This figure is also a good example that the linking of data to face-features must be well chosen. Here the Face in Texas looks very happy, but actually Texas is the state with the most executions and the executed are under 40 years old, as can be seen with aid of the legend. Another example of the representation of Chernoff Faces on maps can be seen in figure 4.10.
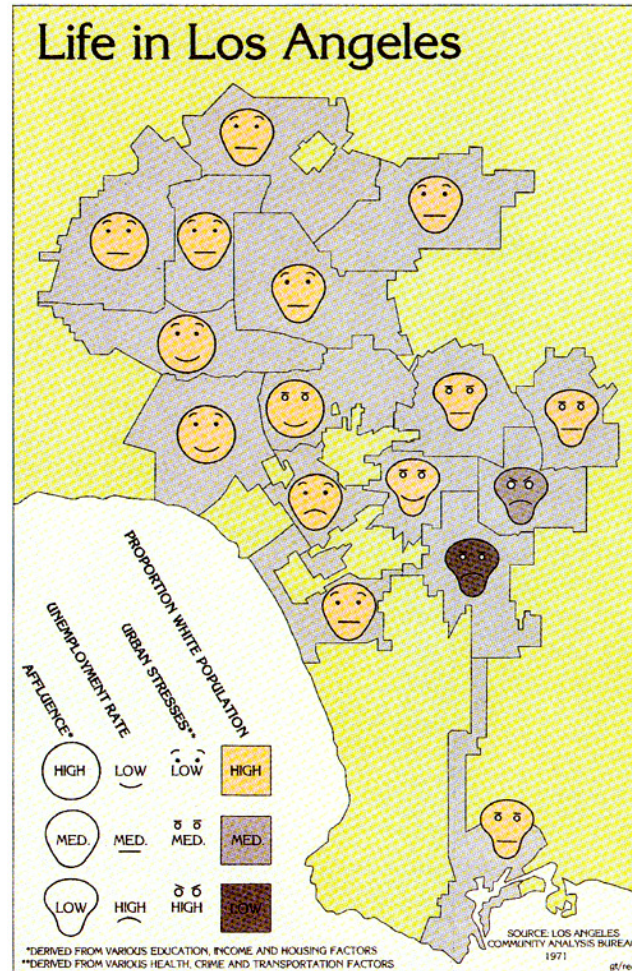
Figure 4.10: Example of Chernoff Faces on maps (taken from `http://kspark.kaist.ac.kr/Human%20Engineering.files/Chernoff/Chernoff%20Faces.htm` (retrieved 2013-05-13))

In this figure the quality of life in Los Angeles is pictured. Here the facial features seem to be well chosen, because the happy faces represent good areas to live and the sadder faces represent areas where the populations is less educated, the employment rate is bad and the crime rate is rather high. So if the representation of Chernoff Faces on maps is well done, an observer with no further knowledge can see patterns fairly easily and can connect the different aspects of the data.

The third representation of Chernoff Faces is the representation in scatterplots. In a Scatterplot two further dimensions are possible. These two dimensions are the x and y axis of the scatterplot. For a good data-visualization those two dimensions must be well chosen, because they determine the position of the Chernoff Faces in the resulting plot. This representation is the least commonly used form. In order to see a scatterplot with Chernoff Faces you have to do a long search. An example of those scatterplots can be seen in figure 4.11.
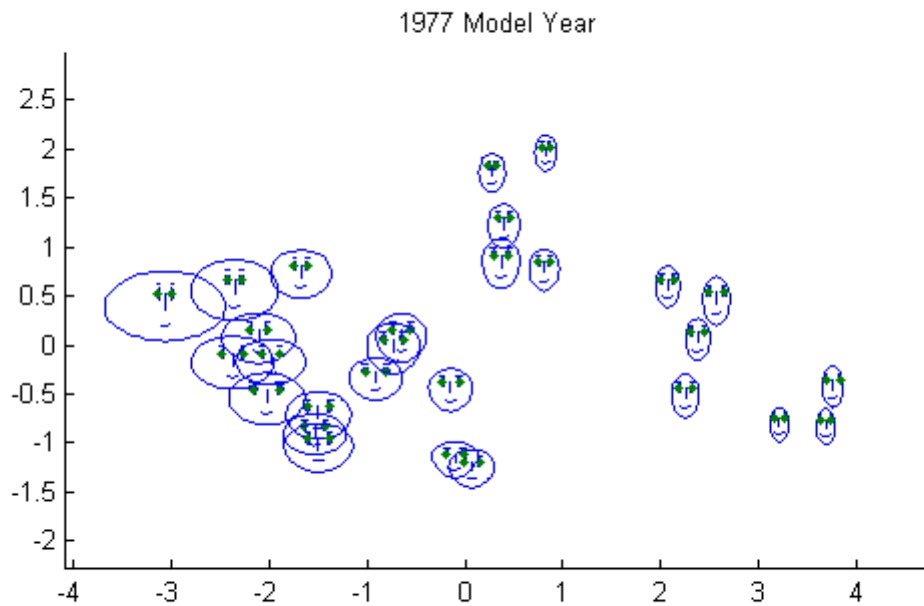
Figure 4.11: Example of Chernoff Faces in Scatterplots (taken from `http://www.mathworks.` `de/products/statistics/examples.html?file=/products/demos/` `shipping/stats/mvplotdemo.html` (retrieved 2013-05-13))

All representations of Chernoff Faces have in common that all parameters must be well chosen, because otherwise the result might convey an unintended message. An example was described above at figure 4.9. A legend is not very common, but if present it helps a lot to understand the data.

Chapter 5

Conclusion

Chernoff Faces can provide a good tool for presenting a first look at multivariate data. They rely on the ability of the human brain to process faces very quickly and notice even small differences. However, research has shown that the recognition of faces is not, as widely assumed, pre-attentive but rather a serial process and does in fact not use the brain's precognition to detect interesting faces. Moreover, the personal perception of how to interpret faces can differ from viewer to viewer. Choosing the features in a meaningful way is crucial to the effectiveness of the displaced faces. If the features are well chosen it is a lot easier to see clusters, outliers and other interesting areas, but on the other side if the choice is done poorly, this tasks gets a lot harder. Another problem of the Chernoff Faces is that they are very rarely accompanied by a legend. Without a legend they can only be used for detection of clusters and transitions, because it is impossible to get any meaning out of the data. Yet, if there is a legend present and the features are well chosen, interesting areas can be observed and interpreted more efficiently. Very few representation methods for the Chernoff Faces exist, notably tables, maps and scatterplots. If there were more common representations, the area of application could be broadened significantly.

Bibliography

Bather, John [1996]. "A conversation with Herman Chernoff." English. *Stat. Sci.* 11.4 (1996), pages 335–350. doi:10.1214/ss/1032280306 (cited on page 9).

Chernoff, Herman [1973a]. "The Use of Faces to Represent Points in K-Dimensional Space Graphically". *Journal of the American Statistical Association* 68.342 (1973), pages 361–368. ISSN 01621459. doi: 10.2307/2284077. http://dx.doi.org/10.2307/2284077 (cited on page 5).

Chernoff, Herman [1973b]. "The Use of Faces to Represent Points in K-Dimensional Space Graphically". English. *Journal of the American Statistical Association* 68.342 (1973), pages. ISSN 01621459. http://www.jstor.org/stable/2284077 (cited on pages 9, 10).

Chernoff, Herman and M. Haseeb Rizvi [1975]. "Effect on Classification Error of Random Permutations of Features in Representing Multivariate Data by Faces". English. *Journal of the American Statistical Association* 70.351 (1975), pages. ISSN 01621459. http://www.jstor.org/stable/2285931 (cited on page 26).

Everitt, B S and P Nicholls [1975]. "Visual Techniques for Representing Multivariate Data". *The Statistician* 24 (1975), pages 37–49 (cited on pages 12, 13).

Fanea, Elena, Sheelagh Carpendale, and Tobias Isenberg [2005]. "An Interactive 3D Integration of Parallel Coordinates and Star Glyphs". In: *IN PROC. IEEE INFOVIS*. 2005, pages 149–156 (cited on page 7).

Flury, Bernhard and Hans Riedwyl [1981]. "Graphical Representation of Multivariate Data by Means of Asymmetrical Faces". *Journal of the American Statistical Association* 76.376 (1981), pages 757–765. doi: 10.1080/01621459.1981.10477718. eprint: http://www.tandfonline.com/doi/pdf/10.1080/01621459.1981.10477718. http://www.tandfonline.com/doi/abs/10.1080/01621459.1981.10477718 (cited on pages 13–15).

Kabulov, B. T. and N. B. Tashpulatova [2010]. "Enhanced Chernoff faces". In: *Application of Information and Communication Technologies (AICT), 2010 4th International Conference on*. 2010, pages 1–4. doi: 10.1109/ICAICT.2010.5612059 (cited on pages 13, 15, 16).

Morris, Christopher J., David S. Ebert, and Penny L. Rheingans [2000]. "Experimental analysis of the effectiveness of features in Chernoff faces" (2000), pages 12–17. doi:10.1117/12.384865. +%20http://dx.doi.org/10.1117/12.384865 (cited on page 11).

Nowell, L., R. Schulman, and D. Hix [2002]. "Graphical encoding for information visualization: an empirical study". In: *Information Visualization, 2002. INFOVIS 2002. IEEE Symposium on*. 2002, pages 43–50. doi: 10.1109/INFVIS.2002.1173146 (cited on pages 7, 8).

Pickett, R.M. and G.G. Grinstein [1988]. "Iconographic Displays For Visualizing Multidimensional Data". In: *Systems, Man, and Cybernetics, 1988. Proceedings of the 1988 IEEE International Conference on*. Volume 1. 1988, pages 514–519. doi:10.1109/ICSMC.1988.754351 (cited on page 6).

Ronald, Peikert [2007]. *Information Visualization*. 2007. http://graphics.ethz.ch/teaching/former/scivis_07/Notes/Slides/10-infoVis.pdf%20R.%20Peikert (cited on page 6).

Ware, C. [2004]. *Information Visualization: Perception for Design*. Interactive Technologies. Elsevier Science, 2004. ISBN 9780080478494. `http://books.google.at/books?id=ZmG%5C_FiqqyqgC` (cited on page 11).