Star Plots: A Literature Survey

Christian Partl
Patrick Plaschzug
Dieter Ladenhauf
Gernot Fernitz

Contents

Abstract

This literature survey is about star plots, which are a visualization method, to display multivariate data. We will give a description, including a historical abstract and a overview of the different representations. In the historical part, there is a discrepancy in the given literature between the early star plots and the definition of a star plot. There are major advantages and disadvantages of star plots, that restrict and influence the usage in different context. Different applications and website extenstions exist, that are capable of creating a star plot. These applications and their features will be discussed.

Keywords: Star plot, star glyph, wheel chart, spider plot, radar plot, radar chart, net chart, web chart, kiviat diagram, coweb chart,polar chart, irregular polygon, data visulaization, multivariate data.

Chapter 1

Introduction

In the last decades the progresses in technology have rapidly increased the speed and memory capacities of computers yielding a rise of the amount of data being processed. The field of information visualization deals with the representation of data that have no obvious spatial mapping and can be defined as “the use of computer-supported, interactive, visual representations of abstract data to amplify cognition” [2]. So, different visualization methods have been developed in order to ease the analysis and exploration of data. Since each of the methods has its inherent strengths and weaknesses they are often used in conjunction in a so called multiple view system, a system where two or more distinct views are used to explore one conceptual entity [1]. The exploration of the data is facilitated by different interaction techniques. For example, brushing is a technique where display elements are interactively selected in order to perform certain operations on them, such as highlighting or masking [18]. Brushing is typically implemented together with linking, a concept which connects different views in a multiple view system by exchanging different parameters such as brushing information [14]. This way, for example corresponding elements of different views are highlighted upon selection. Generally the visualization method is determined by the nature of the data. Hierarchical data for instance can be presented by methods such as the Hyperbolic Browser [15; 16] or tree maps [11]. Parallel Coordinates [10] and star plots [3] are suitable methods to display multivariate data. In this survey we will provide an overview of the star plot visualization method and regard it with respect to different perspectives.

Chapter 2

Description

2.1 What is a Star Plot?

A star plot is a visualization method that is used to represent multivariate data [19; 20; 3], *"data for which each observation involves values for more than one random variable"*[1]. In its original form [3], each data item is portrayed by one star symbol, which are positioned in a grid (see figure 2.1). The rays of a star (axes), which are equally spaced around a circle, represent the different variables of a data item. Their lengths are proportional to the data item's values of the variables. Therefore, the data values should all be non-negative and quite similar in size, which often requires rescaling the variables to a uniform range. Chambers et al. draw connections between neighboring rays in a star, however, in some representations these connection lines are omitted (e.g. [13], [25]). Since there is not enough space to label the rays on each star, Chambers et al. [3] added a key showing variable assignments of the rays (see figure 2.1).

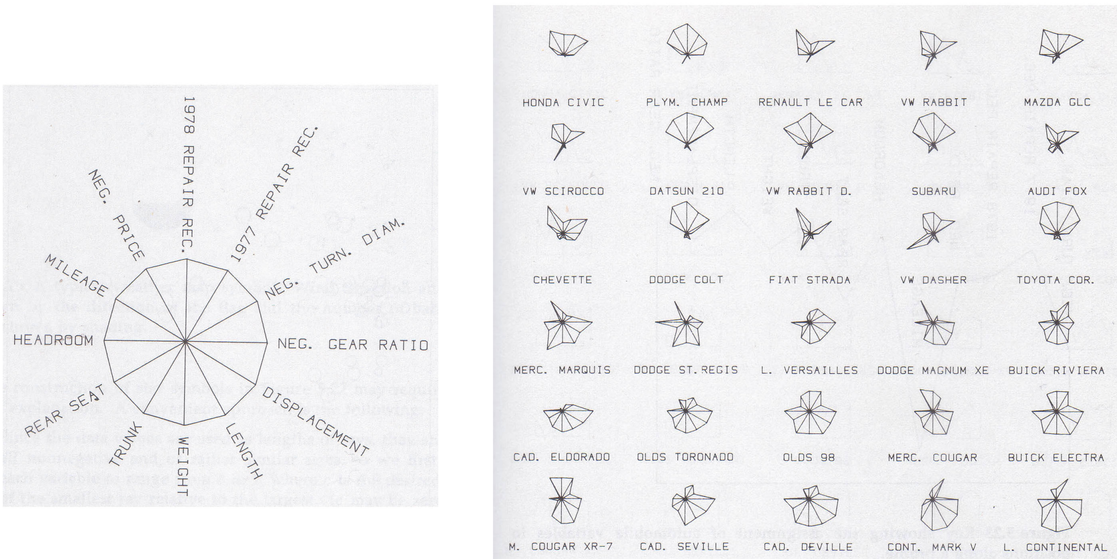


Figure 2.1: Scan of the star plot by Chambers et al. [3]. The key for variable assignment is displayed on the left, the star plot for a subset of the car dataset on the right.

Star plots are most useful when the scales of the rays face the same direction, for example such that "good" characteristics are represented by large rays and "bad" ones by small rays [5; 3]. In order to achieve this representation, sometimes axes need to be inverted. In the example of Chambers et al. [3] who illustrated star plots with the car dataset, the axis that corresponded to the price of a car was inverted because low prices are generally considered to be better than high prices.

[1] http://dictionary.bnet.com/definition/multivariate+data.html

One of the main purposes of star plots is to form distinct shapes for each data item that can be easily compared [3]. This is affected by the ordering of axes and arrangement of the stars [5]. Therefore it is recommended to try several arrangements, because they possibly lead to noticing different features of the data [5; 3].

In the literature a lot of different terms are used to designate star plots. Star glyphs [19] display one star per data item arranged in a grid. They are typically presented without labels and therefore mainly refer to the original form of star plots described by Chambers et al. [3]. According to Monika Lanzenberger [20], wheel charts use a colored polygonal area to depict one data item per instance. She also depicts, that spider plots or radar plots are used to represent more than one data item per instance by drawing one polygon or colored area for each of them (see figure 2.2). The terms radar plot or radar chart comes from their resemblance to radar screens and are also known as measure matrix, net chart [22], web chart [28], kiviat diagram [8], cobweb chart, polar chart or irregular polygon [32]. Though there are some efforts that try to distinguish all the different terms, most of them are used quite synonymously. Due to this fact, we made the decision, to use the term "star plot" from now on in this survey.

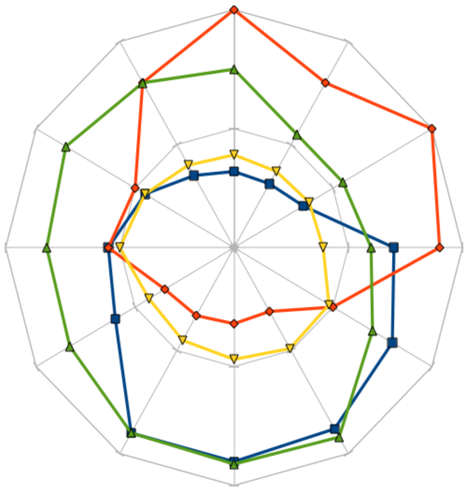


Figure 2.2: A Spider plot.

Another term that denotes a special form of a star plot is the petal chart[2], which shouldn't be mixed up with petal diagram used by Yee Swian Tan [29]. In this representation the circle of the star plot is split into N segments, where N is the amount data items you want to represent in one plot. N shouldn't be too large (like four to six), it depends on how many variables each data item has. In each segment only one data item is drawn (see figure 2.3). This method avoids possible clutter when drawing all items conventionally in one plot and saves screenspace compared to a matrix of plots. However, if you are not used to read a petal charts it's probably very hard to interpret and it's not as intuitive as normal star plots.

Some programs offer the possibility to use stacked (see figure 2.3)or percental stacked star plots, which avoid overlapping of multiple data items in one plot. In the stacked star plot the position of the value of a data item's variable on the axis is determined by adding this value to the position of the value of the previous data item for that variable. For example if there are two items whose variables have the same value, let's say two, the first variable is drawn at position two and the second is drawn at four. Percental stacked star plots work quite similar, just that an axis' end represents 100% and the values for that variable are added according to their percental contribution of the sum of all values for that variable. For example, if there are two items whose variables have the values 3 and 7, the first one is drawn at 30%, the second one at 100%. Generally, stacked and percental stacked representations make he comparison of data items harder.

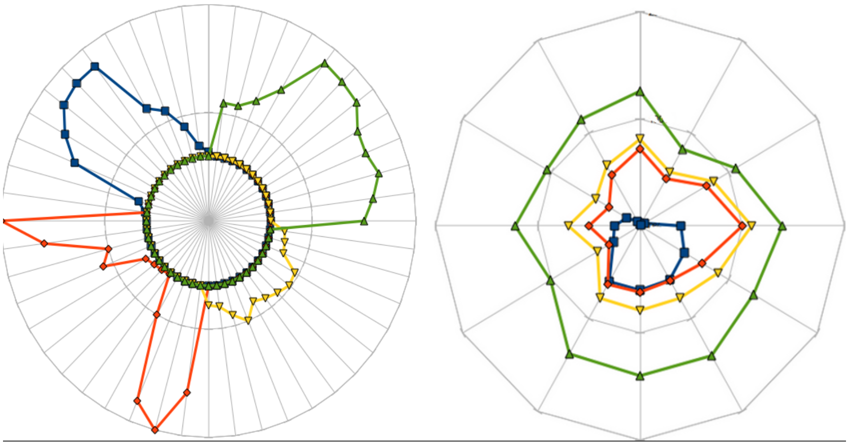


Figure 2.3: A petal chart (left) and a stacked star plot (right).

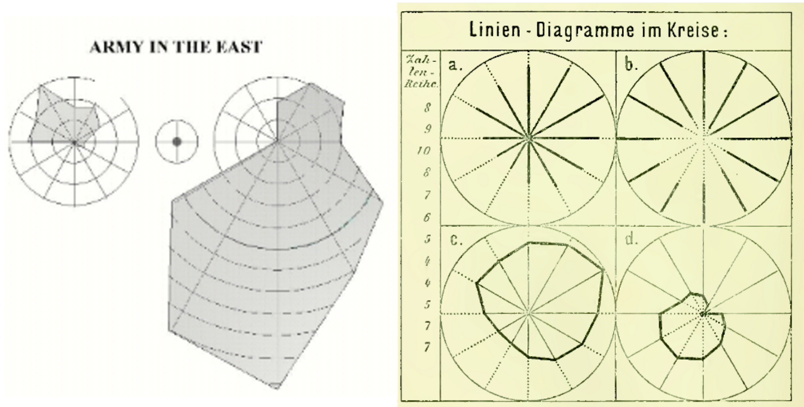


Figure 2.4: Florence Nightingale's Army in the East Bat's Wing Diagram [27] and Georg von Mayr's "Linien Diagramm im Kreise" [30].

2.2 Historical abstract

Florence Nightingale used a number of statistical graphics to illustrate her data. Her bat's wing diagram from 1858 illustrates how the Sanitary Commission that was sent out during the war dramatically reduced the death rate. In figure from Florance Nightingale the circle on the right has twelve sectors representing the first twelve months of the war. The circle on the left represents the second twelve months [24].

Georg von Mayr [30] used a "Linien-Diagramm im Kreise" in 1877 for data representation. He stated, that it is useful to display data which represent a circuit. In his example he showed the average "death rate" per month over more than one year. In this special case the January is as close to the December as the February is.

He used four different ways to represent his data in a star plot.

a.) every line represents the value of his data set starting from the middle point
c.) same as a.) but every point is connected to each other
b.) starts to draw is lines starting from the outside of the circle
d.) same as b.) but every point is connected to each other

According to Michael Friendly [6] Georg von Mayr was the first one who used polar diagrams and star

[2] http://chandoo.org/wp/2008/09/18/better-radar-charts-excel/

plots in 1877 for data representation, the same information can be found on the Infovis page[3] and wikipedia [32].

If we compare these two representations, the explanation and data they used, we find many similarities. So, if the plot from Georg von Mayr is considered as star plot, also the plot from Florence Nightingale must be a star plot. This means that she was actually the first one who used star plots.

However, star plots are used to represent multivariate data, but neither von Mayer's nor Nightingale's plot represent multivariate data. So we could come to the conclusion that actually none of them is a star plot.

[3] http://www.infovis.info/index.php?words=diagrams

Chapter 3

Advantages and Disadvantages

3.1 Advantages

This section focuses on the advantages of star plots. Therefore self made examples will be shown and discussed.

3.1.1 Creation process of a star plot

The following example shows the creating process of the star plot in figure 3.1. The figure shows the comparison between 4 different notebooks.

Information gathering

At first the notebooks and the corresponding data, which has to be plotted into the star plot, was gathered from the official webpages of the producer of the notebooks. Table 3.1 shows the gathered information. To show the advantages of the resulting star plot, the data set has to contain expensive high end notebooks, (like the `Sony Vaio VPCZ11X9E` and the `MacBook Air`) and cheaper notebooks, with less power (like the `Dell Inspiron 13z` and the `HP Pavilion dm3-1110eg`).

Information rating

The next step is the rating of the data, which was gathered before. That has to be done, because the multivariate data has no uniform unit. The rating scale in this example goes from one to four. One means, that the data value is very bad, four means, that it is very good.

In this particular example the rating can be easily done for the data in "Hard Disk", "RAM", "Display Resolution" and "CPU". For example if we look at the four data values, in the "Hard Disk" column, the best value gets the highest rating. The following sequence shows the ordered data values: 128 GB SSD (`Sony Vaio VPCZ11X9E`) $>$ 320 GB (`HP Pavilion dm3-1110eg`) $>$ 250 GB (`Dell Inspiron 13z`) $>$ 120 GB (`MacBook Air`). The Symbol "$>$" means in this case, that the value on the left is better than the value on the right. Thereby the corresponding rating for the notebooks is: `Sony Vaio VPCZ11X9E`: 4,

Model	Price	Hard Disk	RAM	Display Res.	CPU	Weight
MacBook Air [1]	1399 €	120 GB	2 GB DDR3	1280 x 800	1,86 GHz	1,36 kg
Sony Vaio VPCZ11X9E [2]	1899 €	128 GB SSD	4 GB DDR3	1600 x 900	2,4 GHz	1,41 kg
Dell Inspiron 13z [3]	549 €	250 GB	2 GB DDR3	1366 x 768	1,3 GHz	1,79 kg
HP Pavilion dm3-1110eg[4]	599 €	320 GB	3 GB DDR2	1366 x 768	1,6 GHz	1,91 kg

Table 3.1: Information of the four different notebooks.

Model	Price	Hard Disk	RAM	Display Res.	CPU	Weight
MacBook Air	2	1	2	2	3	4
Sony Vaio VPCZ11X9E	1	4	4	4	4	4
Dell Inspiron 13z	4	2	2	3	1	2
HP Pavilion dm3-1110eg	4	3	2	3	2	1

Table 3.2: Rating of the information of the four different notebooks.

`HP Pavilion dm3-1110eg`: 3, `Dell Inspiron 13z`: 2 and `MacBook Air`: 1.

The data in "Price" and "Weight" has to be managed in a different way. For example it is very bad, if the price of a notebook is high and it's very good, if the notebook is cheap. According to the previous scale (4 is very good, 1 is very bad), the rating of the "Price" has to be inverted. It's the same with the weight (a light notebook is good, a heavy one is bad). In that way all the good data values in the plot are on the outside and the bad values are on the inside of the plot [5] [3]. The complete rating can be found in table 3.2.

Plotting of the data

Finally the rating was plotted with Microsoft Excel. The resulting star plots can be seen in figure 3.1. They are, are arranged as a matrix.

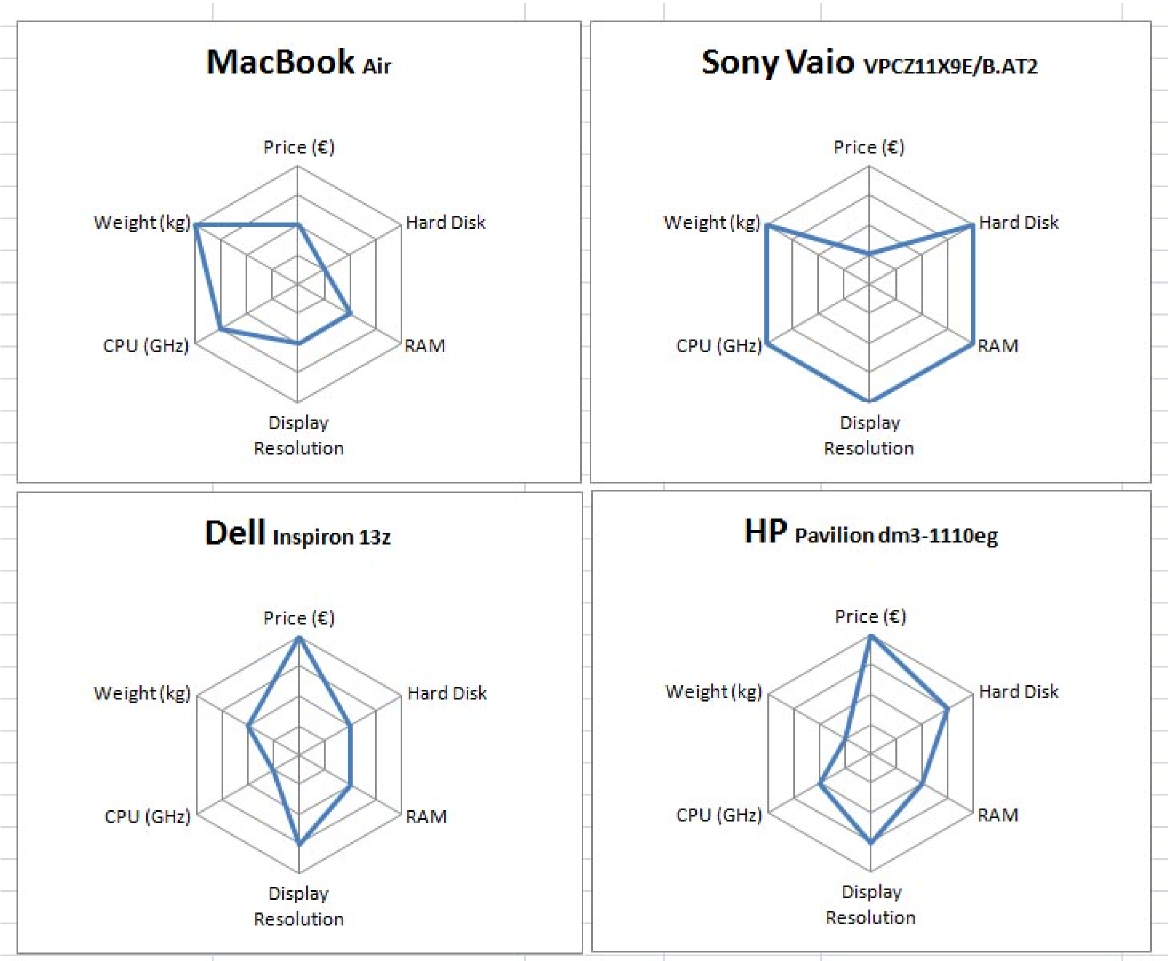


Figure 3.1: Comparison of four different 13" notebooks.

3.1.2 Discussion

According to Hinum star plots are primarily used as comarative tools [7]. As one can see in our example, the comparability of the displayed notebooks is very good if the ranked data is displayed as a star plot. The characteristics of each notebook are easy to see in a short period of time. This increases of course the ability of making a decision.
It's also very easy to find outliners in a star plot [32]. The price of the notebook `Dell Inspiron 13z` in figure 3.1 is a good example for that.
This self-made example shows clearly, that a star plot is very easy to create and it can be used in many different scenarios.

3.2 Disadvantages

This section is about the negative aspects of star plots. With the help of self-made examples, the disadvantages will be shown and discussed.

3.2.1 Clutter

Figure 3.2 shows a star plot, which contains the same notebook datasets from the previous section. The difference is, that this time all four datasets are shown within one plot. As one can see, the star plot is very hard to interpret, because the lines of the different data sets form clutter in the plot. In general, if a plot contains too much examples the shapes lack expressiveness so that they can be hardly distinguished [20].

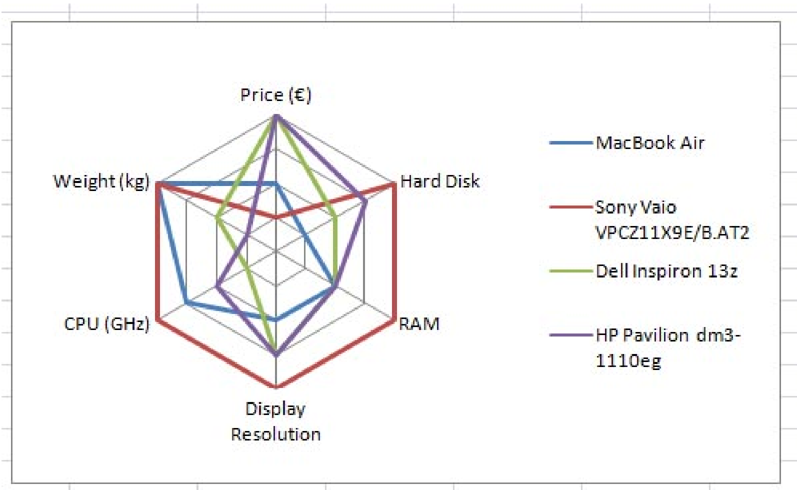


Figure 3.2: Comparison of four different 13" notebooks in one star plot.

3.2.2 Area

In figure 3.3 the same data is presented in three different star plots. The data itself in this example is fictional. This decision was made, because for showing the following disadvantage the data has to be structured in a special way. The mentioned figure indicates one of the major issues of a star plot: The area of the star plot representing an example changes with the order of the axes [29]. And because of this effect the perception of the viewer is affected [5].

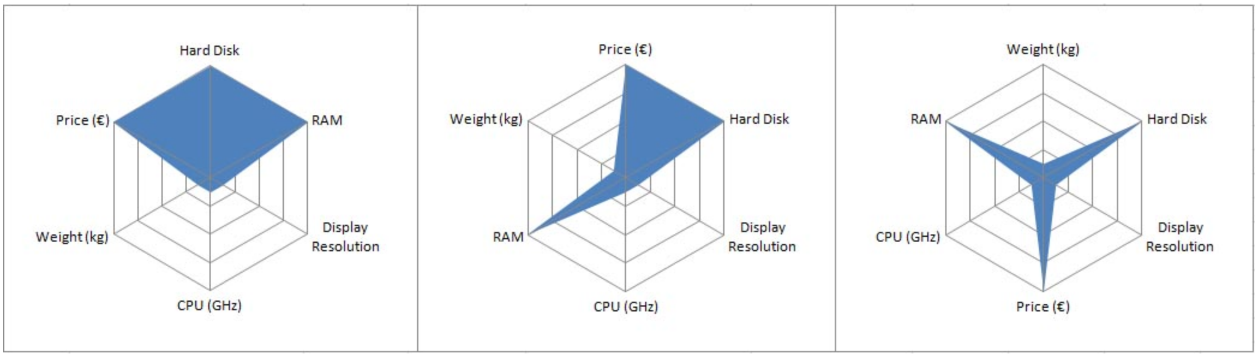


Figure 3.3: A plot of the same dataset with different order of the axis.

To proof that the areas in figure 3.3 vary, they were calculated using the formula of Yee Swian Tan [29]. The results of this calculation are listed in figure 3.4. It can be easily seen, that the area varies with the arrangement of the axis. The area A_j representing alternative j is given by

$$A_j = k[(w_1 x_{j1})(w_2 x_{j2}) + (w_2 x_{j2})(w_3 x_{j3}) + \ldots + (w_{(n-1)} x_{j(n-1)})(w_n x_{jn}) + (w_n x_{jn})(w_1 x_{j1})]$$

where $k = \frac{1}{n} \sin \Theta r^2 n^2$ is a constant, $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$. The meaning of each variable is listed below.

n ...number of axes (criteria)

m ...number of alternatives

Θ ...angle between two lines

r ...radius of the referece lines

w_i ...weight of the ith criterion (in this case the weight is always 1)

x_{ji} ...normalized value of alternative j of the ith criterion

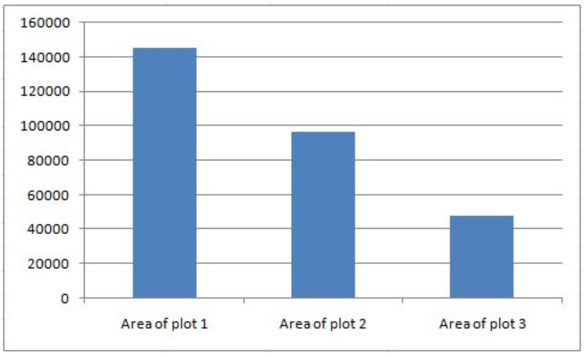


Figure 3.4: A comparison of the areas of the star plots in figure 3.3.

3.2.3 Equal values

The star plot in figure 3.5 shows the costs, spent on two different tables, which were built by Sheridan [26]. The costs, which are illustrated at the left side of the plot are equal. That means, that the blue line is covered from the red one in this example. This point shows another disadvantage of star plots: If two or more data values, whose axes are arranged side by side are equal, the resulting lines cover each other [20]. In this case it would be better to arrange the star plots as a matrix.

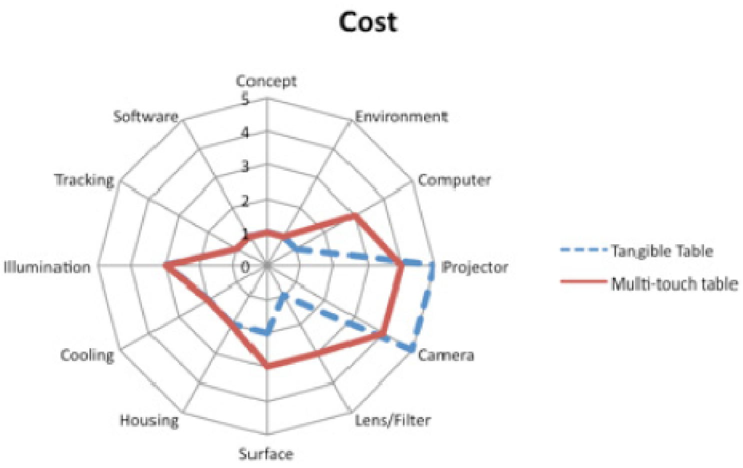


Figure 3.5: A comparison of the costs spent on two tables. [26]

3.2.4 Incomplete data

Lanzenberger mentiones, that another problem ocurrs, if the data, which should be plotted is incomplete [20]. Incomplete data is very common, if the data comes from questionnaires. There are two possible solutions for this problem: Adding an extra point, which represents the missing data, or not drawing the part of the polygon [20] [7].

3.2.5 Limits

Another big disadvantage of star plots according to Hinum is the limited parameter space, because of the limited number of axis [7]. Lanzenberger considered 20 as the maximum of axes within one stardinate [20].
Hinum mentioned furthermore, that in case of small multiples like star icons, star glyphs, or stardinates the number of polygons is limited by the screenspace [7].

Chapter 4

Variants

4.1 Star Glyphs with Spheres

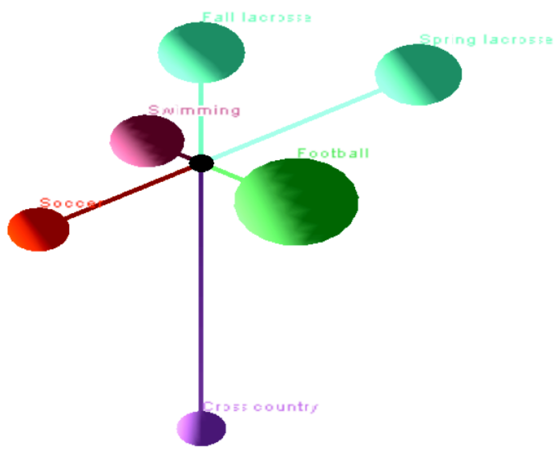


Figure 4.1: Star Glyph with Spheres [25].

Adrian Rusu [25] and his team tried to extend a normal star plot by adding a second attribute at the end of each ray. They decided to add a sphere which can represent an attribute value proportional to the radius of the sphere. To reduce cluttering they represented the plot in three dimensions. However, the addition of spheres still introduces clutter when a data item has many attributes. In order to amplify the perception of outliers of such a plot, they introduced the option to cluster all attributes whose values are close to their average at the center of the plot. The user can interact with this star plot by rotating the plot, changing the scale from every axis, also the scale of each axis and sphere could be changed, labels can be displayed or hidden, clicking on a sphere pops up detailed information.

4.2 Data Rose

Elmqvist et al. [4] concentrated on interactively representing a large set of multidimensional data to the user using a star plot in three different modes: color histogram (high brightness = high density - data divided among the values on each dimension), opacity bands (full opacity = minima and full transparency = maxima - see average and extreme values, distribution) and parallel coordinates. The data isn't reduced in the prefield, so the whole bunch of data is drawn into the plot by transforming the parallel coordinates "into polar space, mapping each axis on the radius of a circle"[4]. The user can now interact with the plot by defining visual queries over integrated sliders or by filtering the data. He can use different operations like union, intersection and

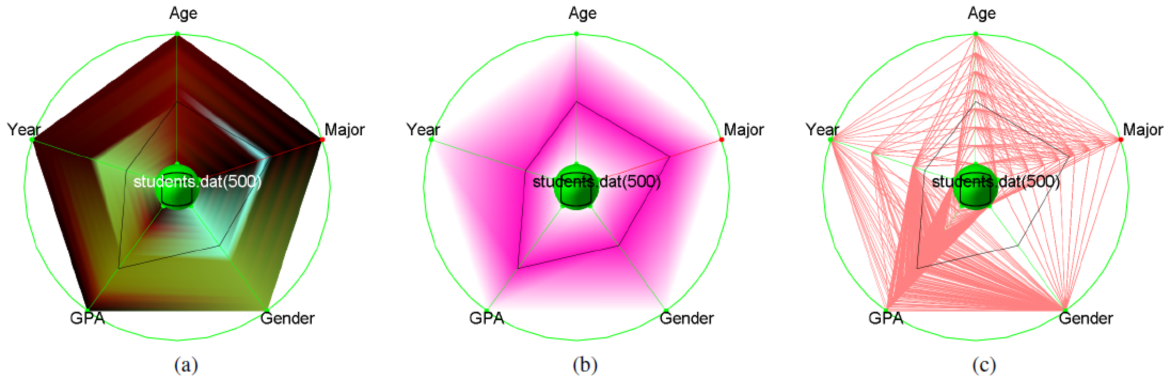


Figure 4.2: Data Rose (a) color histogram mode (b) Opacity bands mode (c) Parallel coordinate mode [4].

uniqueness, and create a new plot connected to the original one. To represent such a big amount of data high-dimensional representing techniques like `dense pixel displays`, `stacked displays`, `brushing`, `axis filtering`, `Dust and Magnet` are used. ”The design intention of the DataRose is to provide a self-contained visual entity that lends itself to side-by-side comparison to other data sets.“[4] The user can easily import data represented in excel or any .csv conform xml table, even if the data is incoherent or separated in different files only a unique identifier is required. In a user study they found out that participants liked the way of playing with the data by changing the filter settings and that the opacity band representation was the most efficient one. They could proof that their method is a good way representing multivariate data.

4.3 Stardinates

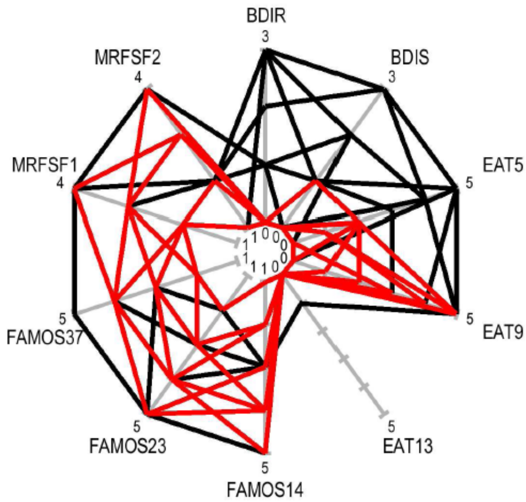


Figure 4.3: Stardinate visualization, selected lines are coloured red [17].

The stardinates by Monika Lanzenberger [20] do not differ much in their visual representation from ordinary star plots. However, interactive features and the possibility of filtering data are added. Stardinates can be arranged a matrix form, everything which disturbs that representation can be hidden, such as axis or labels. To adjust the filters they use sliders displayed as little arrows which are attached to the beginning and at the end of each axis. By turning an arrow, an axis can be inverted. Some sliders can also be used to variate the number displayed axes. The stardinates also support linking and brushing, you can zoom in and out or change the look of the lines. That makes the plot very interactive and it’s possible to handle a large set of data.[20]

The evaluation of a comparison of stardinates with parallel coordinates has shown: “In summary, complexity measurement show us that the Stardinates are better than parallel coordinates more effective than the parallel coordinates in the context of our application domain because they get predominantly better results.”[20]

4.4 Modified Star Graph, Petal Diagram

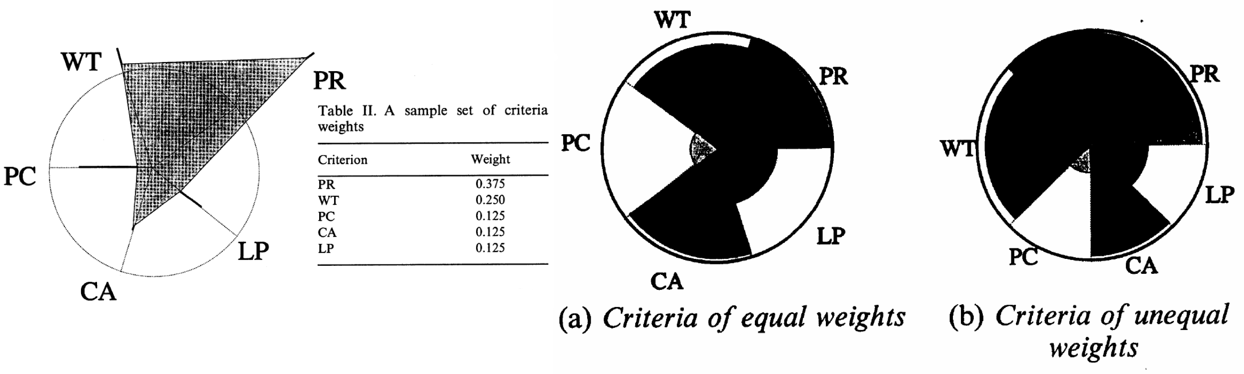


Table II. A sample set of criteria weights

Criterion	Weight
PR	0.375
WT	0.250
PC	0.125
CA	0.125
LP	0.125

Figure 4.4: Modified star graph on the left side, petal diagrams on the right side [29].

In their modified star graph Tan and Fraser [29] modified the star plot by varying the length of each axis to indicate a weight for each variable. To get a better understanding of the weight function, they always drew the original data with thick lines. One issue of this approach was that now the length of each data item represented both the weight and also the value which makes it hard and confusing when studying the effects of changing weights. Additionally, when increasing the weight of one variable and therefore increasing the length of its axis, at the same time the lengths of all other axes are decreased moving closer to zero. This results in very similar shapes for different data items making them hard to compare.

In their second method, the petal diagram, Tan and Fraser [29] tried to overcome the problems of the modified star graph and also the inherent problem of star plots, that the polygonal area of a star plot changes when changing the order of axes. The name comes from the similarity to petals of flowers. In a petal diagram the circle of the star plot is divided into n segments, each representing one variable. The angle of each segment is proportional to the weight of the corresponding variable and the variable's value is represented by the radius of the segment. That way the area of a petal diagram is proportional to the data item's worth[29].

Chapter 5

Usage in different context

5.1 NASA JPL Case Study

A case study of the Jet Propulsion Laboratory (JPL), California Institute of Technology, demonstrates a systematic methodology for an analytical framework that would provide design teams with a better basis upon which to make design decisions of a space mission [23]. They investigated how to automate the design of the MER IDD (Mars Exploration Rover Instrument Deployment Device), which is a robotic manipulator arm. A special algorithm evaluated 60,000 configurations and submitted the best design in each of seven metrics. The results of four selected runs are shown in figure 5.1. In this star plot, the center represents the most desirable results. The red line represents the handcrafted MER IDD, and is used for comparison and validation of the approach. The graph demonstrates the tradeoff between the metrics. The produced results would provide a good starting point for a human designer, this approach would use only about one-fifth as many workhours as a fully handcrafted design.

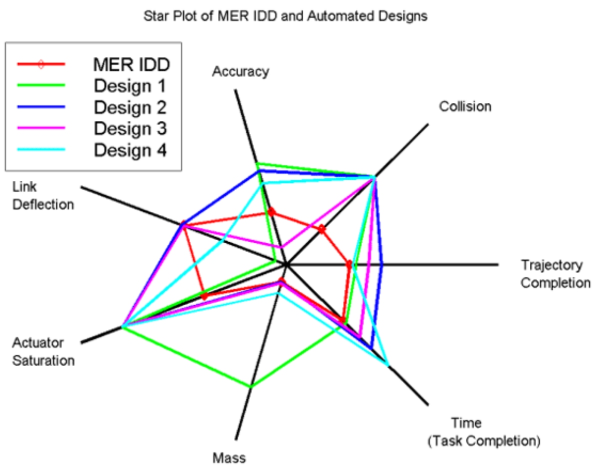


Figure 5.1: Star Plot of MER IDD and Automated Designs [23]

5.2 Basketball Statistics

Star plots are also used to analyze basketball statistics of players or teams of the NBA [9]. The graph represents FG%, points, assists as offensive skills in the upper half versus rebounds, blocks and steals as defensive skills in the lower half. The three upper right statistics are also associated with perimeter players, while the bottom left are associatet with big men. The graph uses normalized values to a scale from 0 to 100.

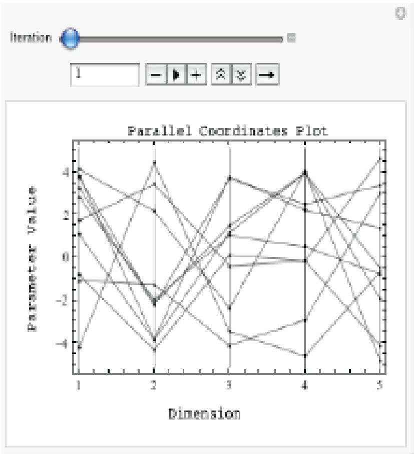
The example in figure 5.2 shows Dwight Howard as an strong defensive and offensive player and Kobe Bryant as a scorer and playmaker.

Figure 5.2: Comparison of two NBA basketball players [9]

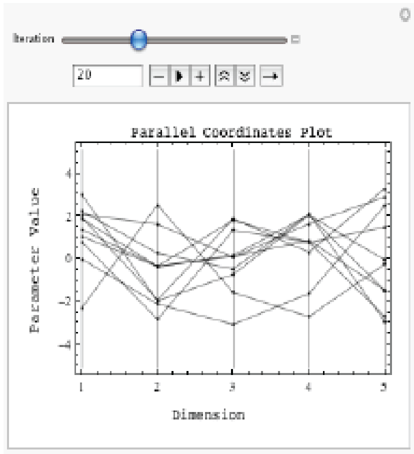
5.3 VISPLORE:A Toolkit to Explore Particle Swarms

Population-based optimization algorithms like particle swarm optimization generate large amount of data. A toolkit like VISPLORE helps the user to view the multidimensional data produced over time [13]. Interactive manipulations are supported by manual and automated sliders. VISPLORE is implemented as a palette of the scientific computational software program *Mathematica*.

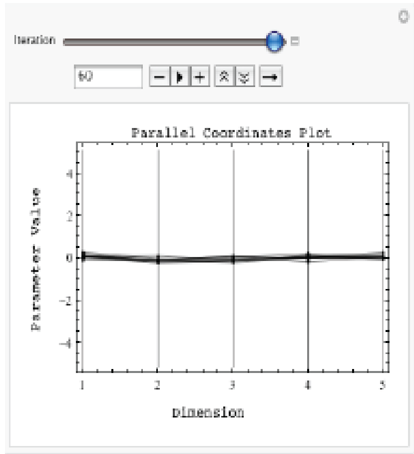
Particle Swarm Optimization is an optimization technique for search spaces, where particles could fly through the solution space and land on good solutions. The overall population moves through the problem space in time-discrete steps. A population at a given iteration consists of many particles that represent potential solutions to the optimization problem In VISPLORE different views of a particle at a given iteration are provided. Figure 5.3 shows parallel coordinates plots, density plots and star plots at three different time steps. Thus, we can see a convergence trend over the time steps, we can also observe areas of the search space that are covered at a particular iteration.



(a) Iteration 1

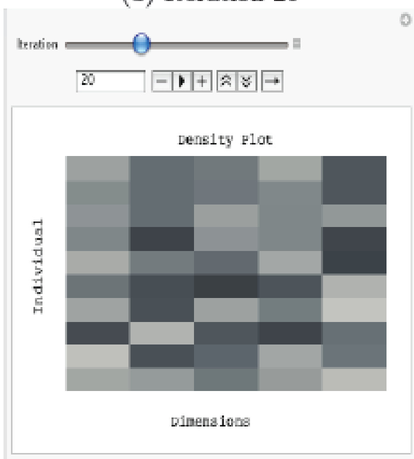


(b) Iteration 20



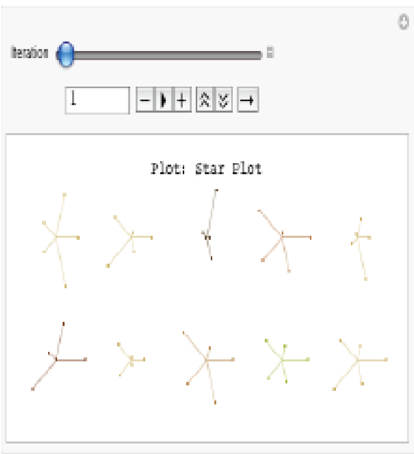
(c) Iteration 60

(d) Iteration 1

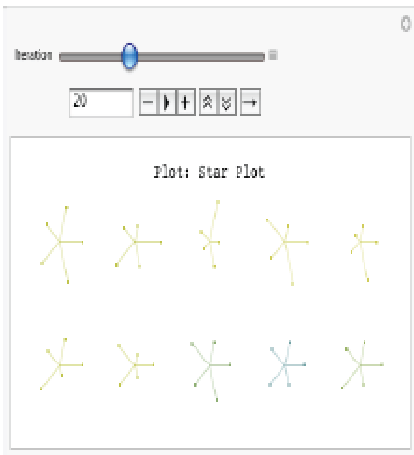


(e) Iteration 20

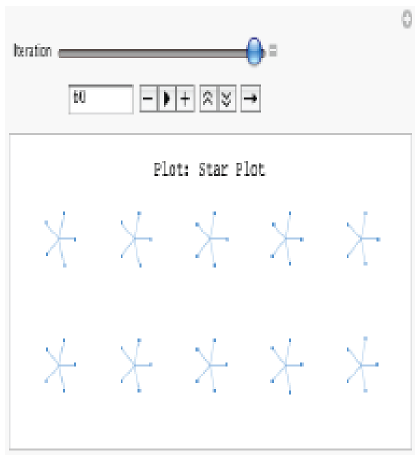
(f) Iteration 60



(g) Iteration 1



(h) Iteration 20



(i Iteration 60

Figure 5.3: Visualizing an particle swarm experiment [13]

5.4 Continuing Vocational Training Survey

The illustration and analysis of the results of the european questioning *Continuing Vocational Training Survey 2 (CVTS2)* by means of radar charts can be done by displaying different indicators in one single graphic [12]. Approximately 76000 companies where interviewed about the further educhation of their employees.

Each radar chart has 4 axes with benchmarked values betweeen 0 (worst) and 1 (best). An overall performance indicator SMOP (surface measure of overall performance) can be calculated by a special formula. Countries with similiar values can be outlined to groups.

The example plots show the participation of the companies (top), the participation of the employees (right) the costs (left) and the duration (bottom) of the courses. The star plots and SMOP of Germany, Norway and Italy are shown.

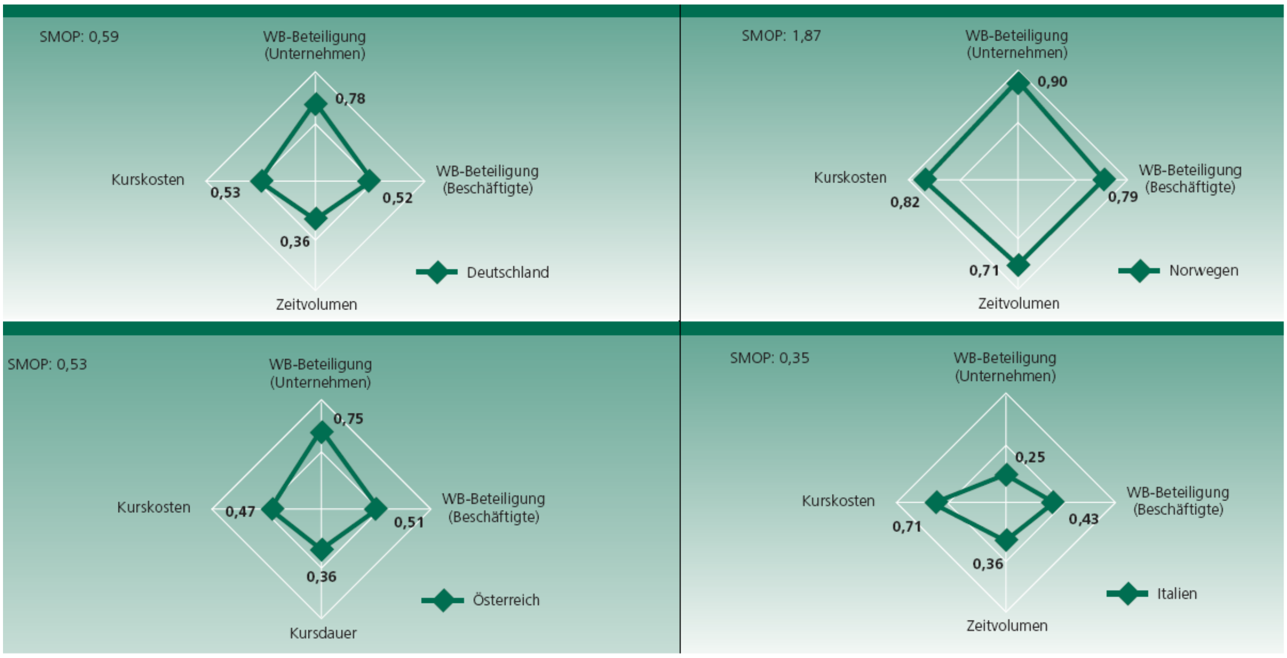


Figure 5.4: Radar charts of different european countries [12]

5.5 Gravi++

Gravi++ is a Flash application to visualize high dimensional, time-oriented data [7]. It is part of the Interactive Information Visualization (in2vis) project at the Vienna University of Technology.

Gravi++ was developed for the analysis of psychotherapeutic questionnaires, but it can be useful to visualize highly structured, temporal, categorical data of any kind. It integrates a spring-based core visualization to display the multidimensional data set, furthermore, a star plot view, attraction rings view. Additionally there are two overview visualizations called ListVis and TableVis to handle larger amounts of data.

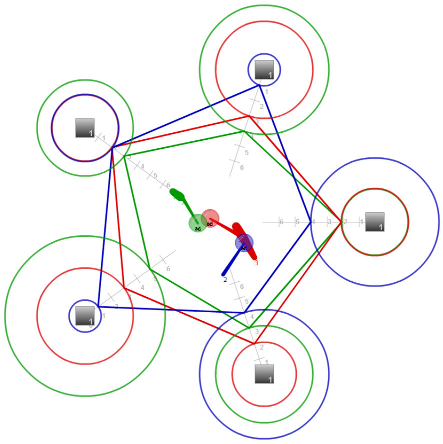


Figure 5.5: Gravi++ Plot [7]

5.6 Architectural Styles

Morisawa used star plots to display the characteristics of nine different architectural styles of applications as a help for selecting an appropriate style for less experienced developers [21]. According to Morisawa an architectural style assignes the processing style and the processing structure for an application system and has a large influence in cost, quality, administration and development period of the system.
In figure 5.6 three different architectural styles (`Centralized Query Style`, `Distributed Transaction St` and `Asynchronous Notification Style`) are shown. Morisawa rated the different styles according to "Data Security", "Reply to User", "Scope of Client/Server", "Independency among Servers", "Data Distribution", "Budget of the System" and "Delivery of the System" using a scale from one to four.

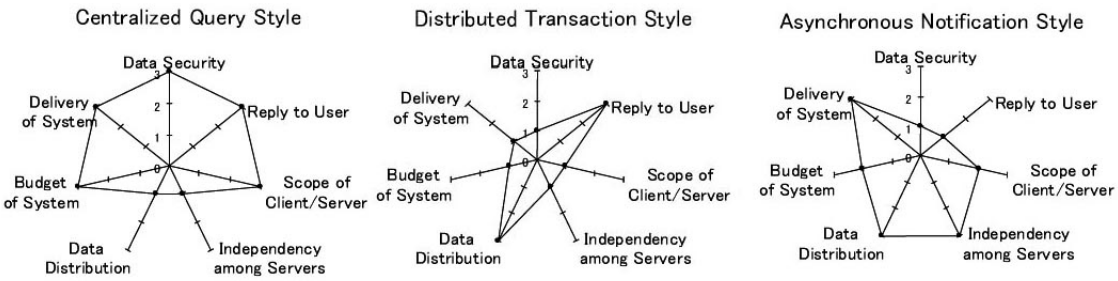


Figure 5.6: Comparison of three different architectural styles [21].

Chapter 6

Applications and Website Extensions

6.1 Applications

This section gives an introduction to a number of different applications that can be used to create star plots with user specified data. Since most of the applications have a large feature set, we will only focus on the parts of the programs that deal with star plots.

6.1.1 Microsoft Office Excel 2007

Microsoft Office Excel 2007 is a spreadsheet application that supports creating several types of charts, including star plots. Every star plot is based on a data table, which can be created in the program or loaded from different sources such as files in comma-separated values (CSV) format. There are several possibilities to adjust the appearance of the plot like explicitly specifying the range of axes, showing them with a logarithmic scale, setting the colours of data item lines, axes and grid, showing different parts of the plot in three dimensions (see figure 6.1) and many more. The program also provides a large number of design templates. In Microsoft Office Excel 2007 there is no direct way to save plots as images, but they can be simply copied and pasted into a graphics program in order to save them there.

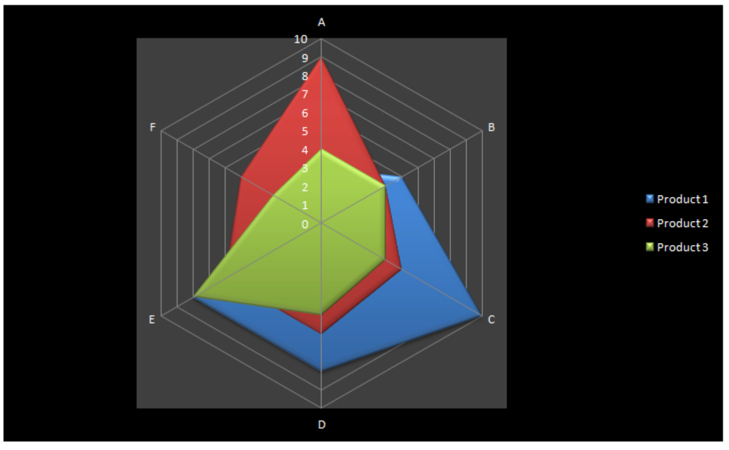


Figure 6.1: Microsoft Office Excel 2007: Data polygons of star plot are shown in three dimensions.

OpenOffice Calc represents an alternative to Microsoft Office Excel 2007. It basically works the same way but has a little less design options. However, OpenOffice calc additionally supports the creation of stacked and percental stacked star plots.

6.1.2 AmCharts Visual Editor

AmCharts is a set of Adobe Flash based chart tools that are supposed to be integrated into web applications. However, amCharts also provides an online visual editor[1] that can be used to create the different chart types, including star plots, directly.

For creating a star plot using this editor the data to be visualized has to be specified either in a CSV or extensible markup language (XML) format. Basically, the star plot can be displayed in a normal line mode, a stacked mode and a percental stacked mode. The editor provides various settings to modify the plot's appearance. For example, different attributes of lines (data item lines, axis, grid) such as colour, opacity or width can be specified, the polygons formed by data lines can be filled with any colour/opacity and texts like title or axis labels can be formatted. There are many more options for modifying the plot's design, but listing them all would exceed the scope of this document. Though the visualization is based on Adobe Flash, direct user interactions are limited to displaying the value of an item's attribute when moving the mouse over an intersection point of a data line with an axis. Since the editor is still in its beta stage, some features like exporting the plot as image and printing it do not work properly. So when the star plot created should be used elsewhere, a screenshot of it has to be captured.

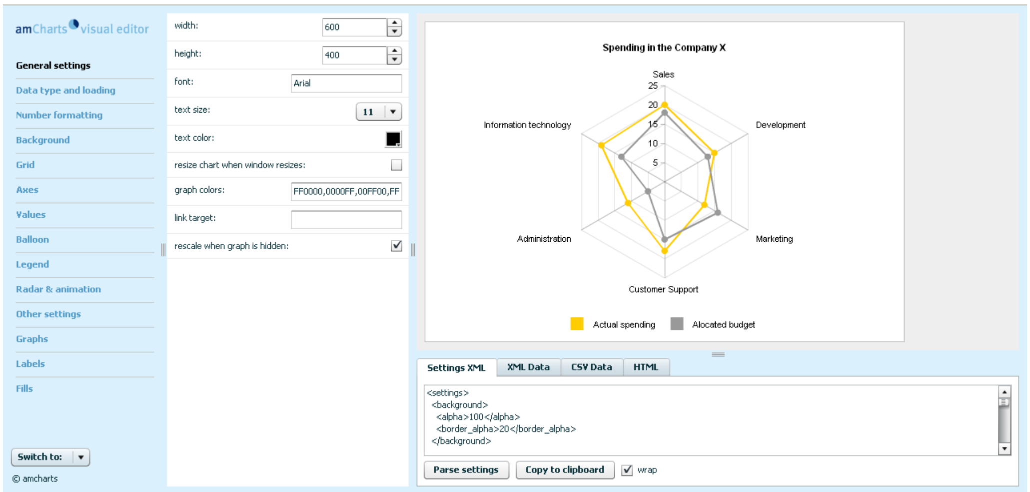


Figure 6.2: AmCharts Visual Editor.

6.1.3 R

R is a software environment and programming language for statistical computing and graphs. In R, star plots can be drawn by using the command "stars". The command takes a matrix containing the data as parameter and draws one star plot for each row. The values of the matrix must be within the range of 0 and 1. If this is not the case, an automatic scaling can be applied, mapping the minimum value of each column to 0 and the maximum value to 1. Though this approach is quite suitable for comparison of different items, it does not show the real quantity of an attribute well in most cases and therefore can lead to a wrong interpretation of the plot. For example, if the minimum value of a column is 50 and the maximum is 100, a mapping to 0.5 for the minimum and 1 for the maximum value would be more intuitive in most cases. Therefore a manual scaling is sometimes necessary. By default, the plots are arranged in a rectangular grid, but the positions can also be specified manually by an argument. When specifying just one position, all plots are superimposed resulting in a single star plot. However, since there are very few parameters to adjust the design of the plot, this approach often leads to star plots that are very hard to interpret. After the plots have been drawn they can be saved in different image- and other file formats.

[1] http://extra.amcharts.com/editor/radar/

There is also an online version[2] [31] for drawing star plots, where the user can use an interface of textboxes instead of having to write R code.

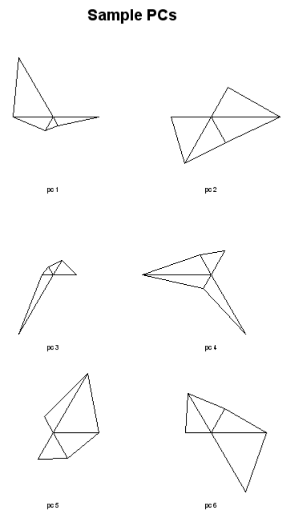


Figure 6.3: Star plots created with R using automatic scaling of values.

6.1.4 Online Chart Tool

As its name suggests, using the Online Chart Tool[3] a number of different chart types such as the star plot can be crated online in a wizard-style step by step creation process. At first, the chart type, in this case the star plot has to be chosen and some general settings like normal or stacked mode can be configured. In the second step the data has to be specified by first setting the number and labels of items and groups and then entering the actual values. In this step the diagram title and data item line colours can be set too. Some settings regarding grid labels such as font, size and colour can be configured in the next step. Finally, the resulting image is displayed and can be saved in different image file formats. Note that the steps do not have to be taken strictly in this order, which enables the user to modify all settings at any step of the creation process.

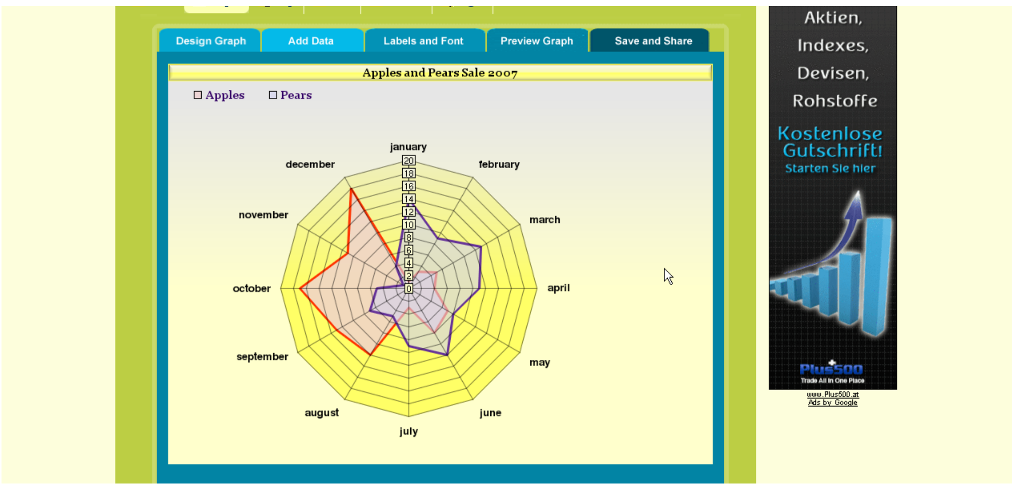


Figure 6.4: Sample star plot of the Online Chart Tool.

[2] http://www.wessa.net/rwasp_starplot.wasp/
[3] http://www.onlinecharttool.com

6.1.5 MATLAB

MATLAB is a high level technical computing language and environment that also supports drawing star plots. Similarly to R (see section 6.1.3), star plots can be created with the command “glyphplot” that takes a matrix of the data to be visualized as parameter. The matrix is automatically standardized, mapping all values onto the range of 0 and 1. The standardization method can be set via parameter. Since one star plot is drawn for each row of the matrix, these plots are positioned in a rectangular grid, whose rows and columns can be specified via parameter. Generally, there are hardly any other possibilities to adjust the appearance of the star plots aside from positioning them. When the plots are drawn, they can be saved in various image- and other file formats.

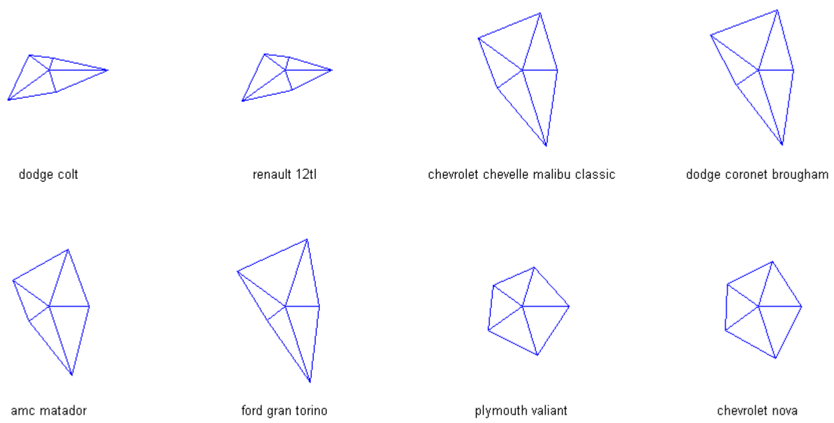


Figure 6.5: Star plots created with MATLAB.

6.1.6 XmdvTool

The XmdvTool [19] integrates several visualization methods for the representation of multivariate data, including star plots. The data that shall be visualized has to be loaded from a file in the XmdvTool datasets format. When the glyph visualization is chosen, the data is represented as matrix of star plots (each plot represents one item), which can be saved in the portable pixmap (PPM) format. The main goal of the XmdvTool is to provide a system that allows an interactive exploration of the data in order to find patterns, dependencies or anomalies [19], rather than creating plots of high visual quality. This is achieved for example through the concept of linking and brushing: When selecting one star plot representing one item, other star plots and elements of other visualization methods that represent similar items are highlighted.

6.1.7 Summary

The visual quality of star plots created with Microsoft Office Excel 2007 is probably the highest among the discussed tools and it offers most design options. In contrast to Excel, AmCharts Visual Editor also supports stacked modes. It is available online, has many design options and also produces plots of high quality. The second online tool, the Online Chart Tool, does not require an Adobe Flash plug-in like AmCharts Visual Editor, but it does not have as many design options and the visual quality of the plots is a bit lower. R and MATLAB are very similar. They offer the fewest design options, the plots have a low visual quality and are only recommended for experienced users, because their usage requires a little programming. However, with R or MATLAB a matrix of star plots where each plot represents one item can be created easily. This is also true for the XmdvTool. Though the visual quality of plots created with this tool is low and its design options are very limited, it is the only one among the discussed tools that allow an interactive exploration of the data.

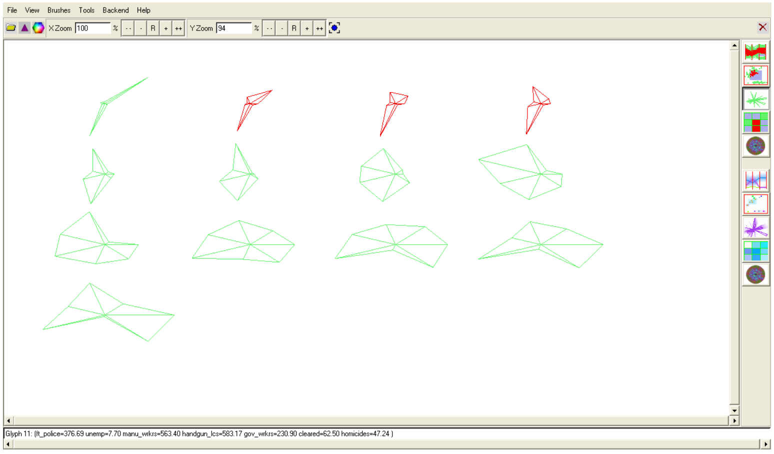


Figure 6.6: XmdvTool showing one of its sample datasets: Similar star plots are highlighted through user interaction.

6.2 Website Extensions

In order to represent user specified data with star plots on a website, it is either necessary to create the visualization on one's own, or to utilize already existing software. For example the Adobe Flash based star plots from AmCharts[4] or Open Flash Chart[5] can be used, whereas the star plots from IBM ILOG Elixir[6], AmCharts for Flex[7], Google Birdeye[8] or the Kap Lab Radar Chart component[9] provide an Adobe Flex based solution. Another possibility is to use the image charts or interactive charts from Google Chart Tools[10]. Interactive charts can be created using a Google developed JavaScript library, images of charts can be generated by sending a special encoded URL request to a Google chart server. For example the URL

```
http://chart.apis.google.com/chart?cht=r&chs=400x400&chco=FF0000,FF9900
&chls=2.0,4.0,0.0|2.0,4.0,0.0&chxt=x&chxl=0:|0|1|2|3|4|5|6|7
&chxr=0,0.0,360.0&chd=t:10,20,60,40,50,90,70,80,10|80,70,70,50,40,30,40,
10,80&chtt=Google+Radar+Chart
```

results in the star plot of figure 6.7.

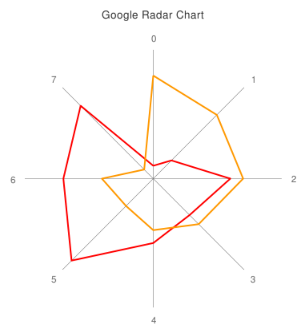


Figure 6.7: Star plot created via URL by Google chart server.

[4] http://www.amcharts.com/
[5] http://teethgrinder.co.uk/open-flash-chart-2/
[6] http://www-01.ibm.com/software/integration/visualization/elixir/
[7] http://flex.amcharts.com/
[8] http://code.google.com/p/birdeye/
[9] http://lab.kapit.fr/display/radarchart/Radar+Chart
[10] http://code.google.com/intl/en/apis/charttools/index.html

Chapter 7

Concluding Remarks

This literature survey has presented an overview of the visualization method star plot. It's used to represent multivariate data and in order to achieve a good representation of the characteristics, sometimes the axis need to be inverted. One of the main purposes of star plots is to form distict shapes for each data item for a better comparability and there are a lot of synonyms for the term star plot.

Within a short historical abstract a comparison between Florence Nightingale's "Bat's Wing diagram" from 1858 [24] and Georg von Mayr's "Linien Diagramm im Kreise" from 1877 [30] was made and it was mentioned, that von Mayr's diagram is referenced in the current literature as the first star plot, even though Nightingale's diagram was published before Mayr's. If the definition of a star plot is taken seriously, none of the diagrams can be denoted as a star plot, because both of them do not display multivariate data.

There are different representations of star plots like star glyphs with spheres, data roses, stardinates, modified star graphs and petal diagrams. Each of them has different characteristics, or adds extenstions to the plots.

An example of the creation process of a star plot was shown and the advantages and disadvantages were discussed. A star plot is mainly used as a comparative tool, because it can be created very easily and outliners can be seen very quickly. It was shown, that the area depends on the order of the axes and too much data forms clutter within the plot. Other disadvantages are the difficulty to handle incomplete data, the limited amount of axes and screenspace and the problem of equal values within one star plot.

Star plots are used in different context, like car comparison, basketball statistics, particale swarm exploration and comparison of architectural styles. There are a lot of different applications (like Microsoft Excel, AmCharts Visual Editor, R, Online Chart Tool, MATLAB and XmdvTool) and website extenstions, that are capable of creating a star plot.

References

[1] Michelle Q. Wang Baldonado, Allison Woodruff, and Allan Kuchinsky. Guidelines for using multiple views in information visualization. In *AVI 00: Proceedings on Advanced visual interfaces*, pages 110–119. ACM, 2000.

[2] Stuart K. Card, Jock D. Mackinlay, and Ben Shneiderman. *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann, 1999.

[3] John M. Chambers, William S. Clevland, Beat Kleiner, and Paul A. Turkey. *Graphical Methods For Data Analysis*. Wadsworth Statistical/Probability Series, 1983.

[4] Niklas Elmqvist, John Stasko, and Philippas Tsigas. Datameadow: A visual canvas for analysis of large-scale multivariate data. In *Proceedings of IEEE VAST '07*, pages 187–194, 2007.

[5] Michael Friendly. Statistical graphics for multivariate data. In *SAS SUGI 16 Conference*, 1991.

[6] Michael Friendly. Milestones in the history of thematic cartography, statistical graphics, and data visualization. Technical report, National Sciences and Engineering Research Council of Canada, 2009.

[7] Klaus Alexander Hinum. *Gravi++ - An Interactive Information Visualization for High Dimensional, Temporal Data*. PhD thesis, Vienna University of Technology, Faculty of Informatics, 2006.

[8] Dirk W. Hoffmann. *Software-Qualität*. Springer Berlin Heidelberg, 2008.

[9] How to watch sports. Spider Graphs: Charting Basketball Statistics, 2009.

[10] Alfred Inselberg. The plane with parallel coordinates. *The Visual Computer*, 1(4):69–91, December 1985.

[11] Brian Johnson and Ben Shneiderman. Tree-maps: a space-filling approach to the visualization of hierarchical information structures. In *Proceedings of the IEEE Conference on Visualization (Vis '91)*, pages 284 – 291, 1991.

[12] Bernd Kaepplinger. Radar-charts als instrument zum vergleich der betrieblichen weiterbildung in europa. *BIBB BWP 5/2006*, 5:26–29, 2006.

[13] Namrata Khemka and Christian Jacob. Visplore: a toolkit to explore particle swarms by visual inspection. In *GECCO '09: Proceedings of the 11th Annual conference on Genetic and evolutionary computation*, pages 41–48, New York, NY, USA, 2009. ACM.

[14] Robert Kosara, Helwig Hauser, and Donna L. Gresh. An interaction view on information visualization. In *State-of-the-Art Proceedings of EUROGRAPHICS 2003 (EG 2003)*, pages 123–137, 2003.

[15] John Lamping and Ramana Rao. Laying out and visualizing large trees using a hyperbolic space. In *UIST '94: Proceedings of the 7th annual ACM symposium on User interface software and technology*, pages 13–14, New York, NY, USA, 1994. ACM.

[16] John Lamping, Ramana Rao, and Peter Pirolli. A focus+context technique based on hyperbolic geometry for visualizing large hierarchies. In *CHI '95: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 401–408, New York, NY, USA, 1995. ACM Press/Addison-Wesley Publishing Co.

[17] Lanzenberger M., Miksch S., and Pohl M. The stardinates - visualizing highly structured data. In *In Proceedings of the Seventh International Conference on Information Visualization (InfoVis '03)*, pages 47–52. IEEE Computer Society, 2003.

[18] Allen R. Martin and Matthew O. Ward. High dimensional brushing for interactive exploration of multivariate data. In *Proceedings of the 6th conference on Visualization (Vis '95)*, page 271, Washington, DC, USA, 1995. IEEE Computer Society.

[19] Ward M.O. Xmdvtool: integrating multiple methods for visualizing multivariate data. In *Visualization, 1994., Visualization '94, Proceedings., IEEE Conference on*, pages 326 –333, oct 1994.

[20] Lanzenberger Monika. *The Interactive Stardinates - An Information Visualization Technique Applied in a Multiple View System*. PhD thesis, Vienna University of Technology, Faculty of Science and Informatics, 2003.

[21] Yoshitomi Morisawa and Koji Torii. An architectural style of product lines for distributed processing systems, and practical selection method. In *ESEC/FSE-9: Proceedings of the 8th European software engineering conference held jointly with 9th ACM SIGSOFT international symposium on Foundations of software engineering*, pages 11–20, New York, NY, USA, 2001. ACM.

[22] Hugh Mosley and Antje Mayer. Benchmarking national labour market performance: A radar chart approach. Technical report, Report prepared for European Commission, Directorate-General V, Employment, Industrial Relations and Social Affairs (DG V/A2), 1999.

[23] NASA JPL. Automation tool for rapid design of space systems.

[24] Florence Nightingale. *Notes on Matters Affecting Health, Efficiency and Hospital Administration of the British Army*. Harrison and Sons, 1858.

[25] A. Rusu, C. Santiago, A. Crowell, and E. Thomas. Enhanced star glyphs for multiple-source data analysis. In *Information Visualisation, 2009 13th International Conference*, pages 183 –190, july 2009.

[26] Jennifer G. Sheridan, James Tompkin, Abel Maciel, and George Roussos. Diy design process for interactive surfaces. In *BCS HCI '09: Proceedings of the 2009 British Computer Society Conference on Human-Computer Interaction*, pages 485–493, Swinton, UK, 2009. British Computer Society.

[27] Hugh Small. Florence nightingale's statistical diagrams. In *In Proc. Stats. & Lamps Research Conference*, 1998.

[28] Nancy R. Tague. *Quality Toolbox*. Amer Society for Quality, 2 edition, March 2005.

[29] Yee Swian Tan and Niall M. Faser. The modified star graph and the petal diagramm: Two new visual aids for discrete alternative multicriteria decision making. In *Journal of Multi-Criteria Decision Analysis*, volume 7, pages 20–33, 1998.

[30] Georg von Mayr. *Die Gesetzmaessigkeit im Gesellschaftsleben*. R. Oldenbourg, 1877.

[31] P. Wessa. Star plot (v1.0.3) in free statistics software (v1.1.23-r6), 2008.

[32] Wikipedia. Radar chart, may 2010.