Ph.D. Thesis

The Interactive Stardinates

An Information Visualization Technique Applied in a Multiple View System

Conducted for the purpose of receiving the academic title
'Doktorin der technischen Wissenschaften'

Supervisor
Silvia Miksch
Institute of Software Technology and Interactive Systems [E188]
Submitted at the Vienna University of Technology
Faculty of Science and Informatics

by

Monika Lanzenberger
9125087
Wiedner Gürtel 64/11
1040 Wien

Vienna, September 29, 2003 ______________________

Contents

Abstract, Zusammenfassung

Information Visualization uses visual metaphors to ease the interpretation and understanding of multi-dimensional data in order to provide the user with information. The relation of users' tasks, data, and interaction forms the background for developing a visualization technique.

In this thesis a new Information Visualization technique is introduced called the interactive Stardinates[1] capable of visualizing complex data. Characterized as geometric glyphs the Stardinates are equipped with sophisticated interaction mechanisms in order to support exploration and analysis tasks of the users effectively.

Finally, we integrated the Stardinates into a multiple view tool called LinkStar because the idea of providing the user with different views of the same data is appropriate in respect to the users, tasks, and data characteristics of our tool.

Findings of our evaluation studies indicate that the interactive Stardinates are a valuable visualization technique.

Informationsvisualisierung basiert auf der Anwendung von grafischen Methaphern und erleichtert so die Interpretation von multidimensionalen Daten mit dem Ziel, Informationen für die AnwenderInnen effizienter zugänglich zu machen. Die Wechselwirkungen von (1) Aufgaben und Probleme der AnwenderInnen, (2) Datenstruktur, und (3) Interaktionsmechanismen bilden den Hintergrund für die Entwicklung eines Informationsvisualisierungswerkzeuges.

In dieser Arbeit stellen wir eine neue Informationsvisualisierungstechnik vor, die 'Interactive Stardinates'. Diese Technik ist mit geeigneten Interaktionsmechismen ausgestattet und daher nicht nur für die Darstellung, sondern auch für eine explorative Analyse von komplexen Daten besonders gut geeignet.

Als eine von vier Visualisierungstechniken haben wir die Stardinates in LinkStar, einem 'Multiple View Tool' integriert. Dieses Tool ermöglicht den AnwenderInnen verschiedene Sichten auf die Daten und unterstützt dadurch den Interpretationsprozess.

Ergebnisse aus unserer Evaluierung bestärken uns darin, dass die 'Interactive Stardinates' eine nützliche Visualisierungstechnik darstellen.

[1] Stars and Coordinates

Chapter 1

Introduction

Information Visualization (InfoVis) uses visual metaphors to ease the interpretation and understanding of multi-dimensional data in order to provide the user with relevant information. In InfoVis three categories of visualizations are distinguished [65].

Three types of InfoVis

Visualizations are characterized by their purpose for ...

- Exploration: the user searches for structures and unknown relations which provide her or him with new insights about the data under investigation. A possible outcome of the exploration process is a derived hypothesis about the data which need to be analyzed next.
- Analysis: starting with certain hypotheses about the data the user tries to prove them by goal-oriented investigations. Finally, the hypotheses are proven by verifying or falsifying them.
- Presentation: presentation covers the static elements of the exploration or analysis process. After these processes are finished, possible findings need to be communicated to others. This task is handled by the presentation.

Interactive exploration

Whereas in the first two categories the user has the central and active role, in the third the author of the presentation is more important and defines the goals. Our focus is exploration and analysis, because we anticipate InfoVis mainly as a interactive exploration technique focusing on the users and their tasks. So in our opinion, InfoVis aims at making complex data accessible for interactive investigation by the user. We see exploration in the tradition of early approaches in statistics which emphasized in giving rapid statistical insight into data [74]. In this context exploratory data analysis is an approach how data analysis should be carried out and uses a variety of techniques in order to [52]:

1. maximize insight into a data set,

2. uncover underlying structure,

3. extract important variables,

4. detect outliers and anomalies,

5. test underlying assumptions,

6. develop parsimonious models, and

7. determine optimal factor settings.

Exploring data yields information which is processed into knowledge. Several definitions of data, information, and knowledge exist. In our context the following definition [63] seems appropriate:

- Data: input signals to sensory and cognitive processes.
- Information: data with an associated meaning.
- Knowledge: the whole body of data and information together with cognitive machinery that people are able to exploit to decide how to act, to carry out tasks, and to create new information.

About two years ago we started a project cooperation with the University of Vienna, Department of Pediatrics and our goal was to find an adequate visualization for psychotherapeutic data, particularly, data derived from a study on anorectic girls.

Beyond statistical analysis

One important characteristic of this kind of data is its complexity. Combining many dimensions of such data, in order to explore beyond statistical analysis, needs sophisticated visualization metaphors and interaction mechanisms because the interpretation of such data is complicated.

So, our goal was to enable the psychologists and physicians to explore the data in an experimental way in order to derive qualitative information on the states of the patients, especially, comparing patients of group therapy on the one hand and patients of individual therapy on the other hand. The results of testing relevant existing visualization tools did not satisfy their needs because the visualizations were either too fragmentary and biased or to complex for efficient and goal-oriented exploration. Each tool on its own did not provide sufficient visualizations. The diversity and complexity of the data and the demand to support efficient yet careful interpretation by the users, led us to the idea of combing three well-known InfoVis techniques. We developed LinkVis: a multiple view tool providing the user with different views on the the same data.

Developed a multiple view tool

In parallel we began to investigate a new visualization technique because we found shortcomings in the existing approaches which limited their usability in our context.

1.1 Problem Domain

Considering the characteristics of the underlying data the existing InfoVis techniques did not suffice in terms of achieving the goals of our users. In particular, existing InfoVis techniques did not allow to likewise accomplish the following tasks:

- Decomposing complexity without loosing relevant information
- Emphasizing similarities and differences
- Integrating functionality based on an artifact-driven approach to enable the user to explore and analyze the data directly
- Allowing for examining the data on different levels of details
- Avoiding interfering connotations such as caused by the Chernoff faces
- Grouping meaningful bundles of dimensions within one visualization item
- Characterized by traceable mapping
- High flexibility and consistency of the underlying visualization metaphors

Based on extensive discussions with the users and studies of their needs we found out that these properties are essential for our InfoVis technique in order to support the goals of the users efficiently. The basic parameters of our work are described by the relation between visualization, data, interaction, and users' tasks (compare Figure 1.1). In the following we list the **tasks of our users** because they form the starting point for our work.

- Observing the progress of the clinical study
- Visualizing the therapeutic course of a patient
- Comparing the states of patients
- Comparing the effects of individual and group therapy
- Finding relations of patients' states and family situation
- Deciding on the sufficient and best therapy forms

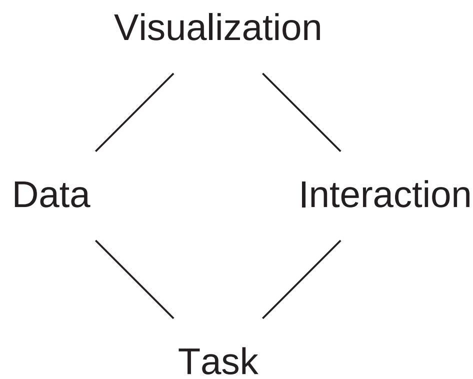


Figure 1.1: The task on the one hand is connected to the data categories and on the other hand requires certain interaction mechanisms. This is the basis for designing a visualization.

Generally speaking, these tasks are characterized by exploration and analysis of the data which typically take place when evaluating the results of a large number of diverse questionnaires (Chapter 10 describes some application domains).

The tasks of the users require capable interaction mechanisms in order to allow for a goal-oriented utilization of the visualization (compare Chapter 6). Moreover, the tasks deal with specific data which determine the visualization technique. (compare Section 5.1). So finally, the tasks of the users, the underlying data, and the interaction mechanisms constitute the requirements for our approach in finding a capable visualization technique.

1.2 Our Approach

Created the interactive Stardinates

Inspired by the glyphs and parallel coordinates, we created the *interactive Stardinates*[1] which is characterized as a geometric glyph with integrated functionality for exploration. The name results from a compound of 'Star' and 'Coordinates'. The axes are arranged in a circle, like rays. It aims at visualizing complex data in an efficient way by exploiting perceptual and cognitive features (compare Section 5.5). As a starting point for the design of the user interface we developed a framework, called the graphical object operation scheme, to get a comprehensive lineup of the needed interaction mechanisms. By integrating the interactive Stardinates in a multiple view system, called *LinkStar*, we expand our multiple view visualization tool.

Starting with an overview of the related InfoVis techniques in Chapter 2, the relevant approaches from cognitive psychology in Chapter 3, and

[1] http://www.ifs.tuwien.ac.at/~ mlanzenberger/Stardinates

two different frameworks on interaction in Chapter 4, we build a basis for the design of the interactive Stardinates which are covered in Chapter 5.

A comprehensive description of the user interaction mechanisms and associated functionality of the Stardinates is given in Chapter 6. Thereafter, we present the results of our evaluation of the interactive Stardinates in Chapter 7.

Chapters 8 and 9 deal with the integration of the Stardinates in a multiple view system, namely, LinkStar. Following this, three types of application domains of the Stardinates are sketched out in Chapter 10.

Finally, in Chapter 11 we summarize our work and outline some future plans, followed by our conclusions in Chapter 12.

Parts of this work were published at the following conferences: ACM Symposium on Applied Computing, SAC2003, March 9-13, 2003, Florida, USA; World Conference on Educational Multimedia, Hypermedia & Telecommunications, ED-Media'03, June 23-28,2003, Hawaii, USA; 7th IEEE International Conference on Information Visualization, IV03, July 16-18, 2003, London, UK; IFIP TC13 International Conference on Human-Computer Interaction, INTERACT 2003, Doctoral Consortium, September 1st, 2003, Zurich, Switzerland.

Chapter 2

Related Work in Information Visualization

This chapter introduces some InfoVis techniques which are relevant in our context. First, parallel coordinates and glyphs are described because both essentially inspired the Stardinates concept. Thereafter, we list visualization techniques characterized by using rays similar to the Stardinates but they do not offer mechanisms for interactive exploration. Therefore, they should be categorized as presentation techniques. Finally, we describe two techniques utilized by the Stardinates, namely: *multiple views* and *linking and brushing*.

2.1 Parallel Coordinates

Since more than twenty years this method is very popular in InfoVis and widely used in many tools (e.g., VisuLab [31], VisDB [36], Xmdvtool [77], XGobi [72], Parallax [4]). It is a geometric technique which performs non-projective mapping of N-dimensional data into 2-dimensional space [33]. The analysis of multidimensional data is thus transformed into a 2-D pattern recognition problem. This is realized by a number of N axes placed equidistant and in parallel to the y-axis in the Cartesian coordinate system. A data record, e.g., a point C with coordinates $(C_1, C_2, .., C_N)$ is represented by a polygonal line (compare Figure 2.1) whose N vertices are at position $(i-1, C_i)$ on the x_i-axis for $i = 1, ..., N$.

Mapping N dimensions into 2-D space

Obviously, the shape of the diagram and its expressiveness in terms of easily perceivable cues depends also on the arrangement of the axes. It is a difficult problem to find an axes permutation which is best suited for a specific data set. According to Wegman's Theorem [8] $\frac{N}{2}$ 'well chosen' displays suffice. If N is an odd value, $\frac{N}{2}+1$ such displays exist. But the best of these orderings is completely data set specific.

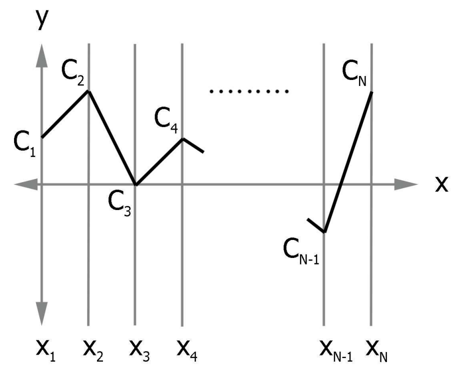


Figure 2.1: The concept of the parallel coordinates: parallel coordinates visualizing a N-dimensional point $C = (C_1, C_2, C_3, ..., C_N)$ by a polygonal line with N vertices.

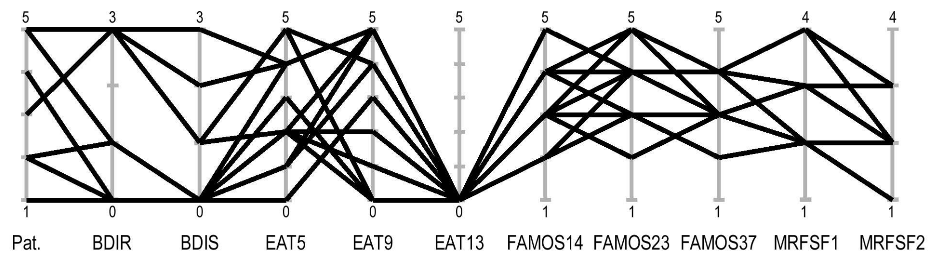


Figure 2.2: Parallel coordinates visualizing medical data derived from a clinical study on anorectic girls.

Without loss of information

What is special about this InfoVis method is its ability to represent multidimensional objects without loss of information. In contrast to many other InfoVis methods and techniques it is based on dimensionality selection instead of dimensionality reduction [32]. This feature is relevant in the context of the analysis processes of the data sets under investigation. Parallel coordinates offer great potential for interactive exploration and analysis by experienced users. Figure 2.2 visualizes data derived from a clinical study on anorectic girls. A selection of 10 attributes each represented by one axis provides the user with information on the states of the patients. Each polygonal line depicts the data of one patient. The dimensions have different ranges of values. In order make minimum and maximum values comparable the axes are of equal length. Since such a diagram is usually not understandable if only seen as a static image, skillful interaction is essential. However, in our opinion parallel coordinates are more convenient if supported by automatic pattern analysis. The sophisticated geometric properties of this InfoVis method allow to apply classifiers for dimensionality selection with low complexity, working nearly in real-time. Such algorithms [58; 23; 47] accomplish classification of data which is a basic task in data analysis and pattern recognition. It seems necessary to support the users by adequate classifiers since these complex geometric features and their meanings are not evident at the first glance. Properties like color, brightness, thickness, etc. of the lines are available for filtering, linking and brushing, and other methods of data set manipulation or could simply encode additional dimensions of the data.

Support data selection

2.2 Glyphs

Glyphs are graphical objects which represent data by using visual parameters such as color, size, position, shape, and so on. In contrast to symbols (e.g., traffic signs) and icons (e.g., stick figures [28; 56]) they are capable of visualizing high-dimensional data. Examples are the time-wheel glyph and the InfoBUG glyph [18; 19].

Basic concepts of the glyphs

Three basic concepts are associated to glyphs:

- Small multiples:

  A small number of glyphs is positioned on a grid. Each glyph is derived from the same design structure and gets an equal amount of space. They are effective because of three reasons: visual constancy, economy of perception, and uninterrupted visual reasoning.

- Established visualizations:

  Combinations of well-known visualizations, such as time series, histogram, and similar, are used to build a glyph.

- Information rich glyphs:

  A high number of dimensions is grouped within one glyph using clustered sets of simple graphical artifacts. Arranging these artifacts in a meaningful way by using certain metaphors eases perception: e.g., the InfoBug uses the metaphor of an insect.

A glyph often looks confusing in the first moment because in general mapping of the dimensions is rather arbitrary. Users need to learn the meaning of the graphical representation.

Chernoff faces

A totally different approach is the Chernoff face [17]. Developed in the early seventies it uses faces to visualize points in k-dimensional space. In particular, this glyph is capable of representing up to 18 dimensions. Such a glyph looks like a face whose characteristics and facial attributes are determined by the position of the point (compare Figure 2.3).

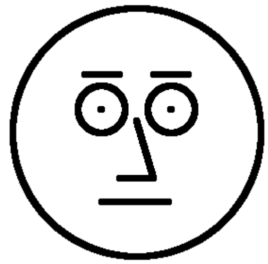


Figure 2.3: Chernoff face

In comparison to the numerical (textual) evaluation of the data, this kind of visualization bears several advantages:

Features

- Enhancing the user's ability to detect and comprehend important phenomena;
- Serving as a mnemonic device for remembering major conclusions;
- Communicating major conclusions to others;
- Making it possible to do relatively accurate calculations informally.
- Using the expected symmetry of the face to compare two different entities.

And problems of the Chernoff faces

However, this kind of visualization also causes some problems:

- Although a face consists of a rather high number of properties, we are not able to perceive each single property. For instance if seven data dimensions are mapped to properties of eyes and eyebrows: e.g., eye size, eye form, position of eyes, size of pupil, position of pupil, position of eyebrows, shape of eyebrows, we usually perceive the eyes as a whole and are unable to distinguish components adequately. So, unforseen interferences might occur which could result in misinterpreting values. Practically, the number of dimensions is limited to four or five.
- They are seen as oversimplified representation of data.
- There is no direct, intuitive way of mapping values to facial properties.
- Mapping values to facial properties could imply discrimination of people, especially, if trying to utilize common biases.

- The emotional connotations and meanings could mislead the user. Moreover, even small inconsistencies of the face result in an unfriendly, angry, sad or distorted mimic which imply specific associations.
- Perception of such faces is not pre-attentive [48].
- It is not possible to see absolute values, we perceive the values only in relation to the others. Therefore, a reference face is needed in order to compare.

In short, we think that Chernoff faces are a nice idea and users often find such a visualization appealing. However, they do not generate visualizations which are practical and suitable for in-depth examination of the data.

2.3 Star Coordinates, Starglyphs, Wheel Charts, Star Plots, and Spider Plots

Points or lines

Several geometry-based approaches exist which place axes in a circle similar to our interactive Stardinates. Two groups of such visualization techniques are distinguished.

Extension of scatterplots

Techniques of the first group visualizes data by points, for instance, **Star Coordinates** [35]. They are a simple extension of the 2D or 3D scatterplots to higher dimensions with normalization. The mapping of a data element is determined by the sum of all unit vectors on each coordinate multiplied by the value of the data element for that coordinate. So, mapping could not be verified directly because a single data point may correspond to a number of data values. This approach offers interesting attempts in visualizing a high number of dimensions on a relatively small space. However, compared to the Stardinates, the goals of this InfoVis technique are quite different, e.g., it is not intuitively understandable.

Focus on presentation data

The second group represents data values by lines or polygons. These visualization techniques have in common that their focus lies on presenting data. That is one important difference to the Stardinates which are a highly interactive visualization technique created for exploration but less for presentation of the data.

Moreover, the original forms of Starglyphs, star plots, and wheel charts only depict one data record per instance. Stardinates can visualize many data records within one instance because of the interactive approach.

Starglyphs and star plots are explicitly related to the concept of the glyph-technique, which arranges a number of items on a grid and motivates comparison. The **Starglyph** [70; 77] (compare Figure 2.4) uses position,

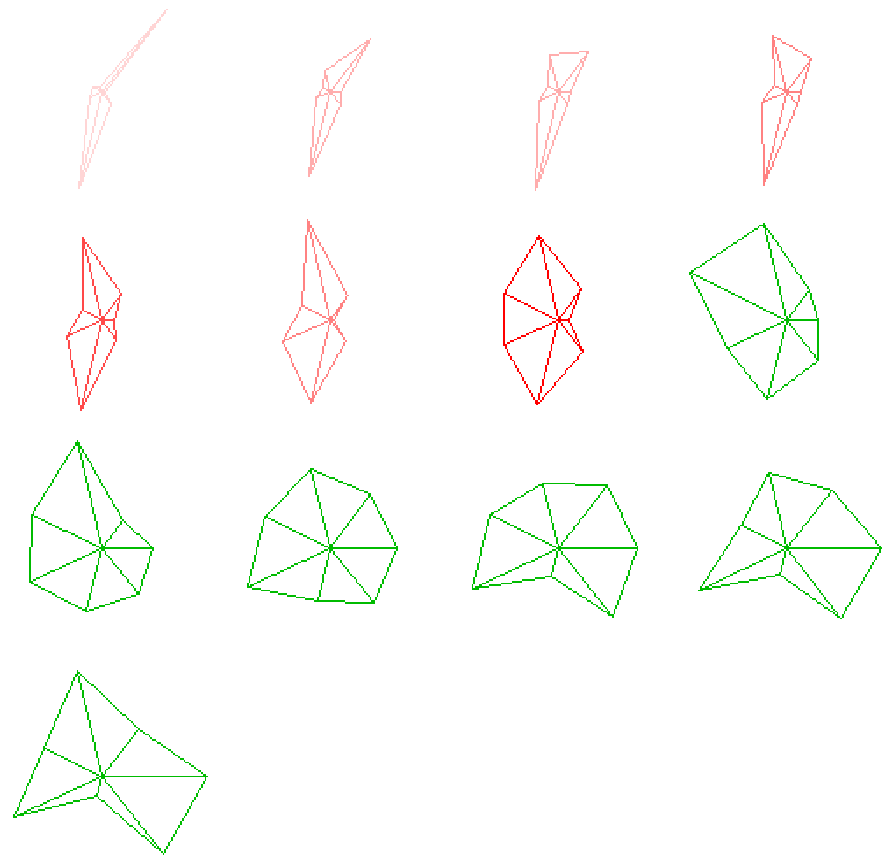


Figure 2.4: Starglyphs, Xmdvtool [77]

color, and length of the rays to encode information. The ends of the rays are connected which gives the Starglyph the characteristic shape. **Star plots** [16] use the same method for visualizing multivariate data. **Wheel charts** use colored polygonal areas for visualization which form distinct shapes.

Spider plots, also known as radar plots, are another presentation technique widely used. Either polygons or colored areas represent the data. In Figure 2.5 we see a spider plot which shows three colored areas each representing one data record with five parameters. However, their background is not related to the concept of the glyphs, they are used as single graphical presentations. In addition, the visualization does not offer integrated functionality such as sliders, linking and brushing, etc. which is essential for the interactive Stardinates.

At the first glance some of these techniques may look similar to the interactive Stardinates. However, apart from the similarity of arranging axes in a circle and utilizing the glyph-concept, Stardinates offer more sophis-

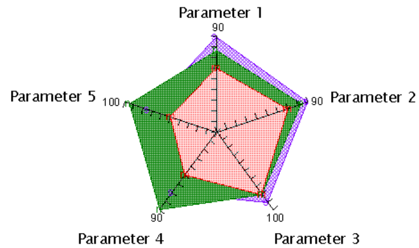


Figure 2.5: Spider plot

ticated explorative functionality. In contrast to the existing methods, the interactive Stardinates are an integrated tool for exploring and analyzing complex data. This is what distinguishes the interactive Stardinates from related visualization techniques discussed above.

2.4 Multiple View Visualizations

Advantages of multiple views

The idea of combining several views of usually the same data is well-known as so-called multiple view visualization [13; 34; 55; 76; 75; 60]. These multiple view visualizations offer a lot of advantages, such as improved user performance, discovery of unforeseen relationships, and unification of the desktop [54].

Exploring data by using InfoVis is characterized by abstraction. In addition, visualization itself often reduces information or emphasizes certain aspects of the data in order to ease goal-oriented interpretation. The combination of distinct visualizations yield different kinds of abstractions from the data which allow for diverse approaches of exploration. Figure 8.5, p. 92 shows such a multiple view tool. Stardinates and parallel coordinates provide the user with different views of the same data.

Definitions [5] of the components of such a multiple view visualization enable us to understand the main issues: in this context a single view of a conceptual entity could be seen as a set of data plus the specification how the data is displayed. The visualization technique acts like a lens which gives us a certain view of the data.

Different sights and different insights

Views are distinct if they enable the user to see different aspects of the conceptual entity. Different views provide the user with different sights and different insights. This could be done by emphasizing different aspects of the same information. Therefore, a multiple view visualization consists of two or more such distinct views in order to support the examination and exploration of a given conceptual entity.

Generally, the data can also differ, e.g., one data set could be a subset of the other or the data sets of different views could represent different selections. A typical application is combining overall information on the one hand, with detailed information on the other hand. However, different data increases complexity of a multiple view visualization.

Visualization pipeline

One way to structure the process of generating a multiple view visualization is to categorize at which stage of the visualization pipeline the distinct views are established. Generally, four stages are distinguished: (1) Raw Data, (2) Data Set, (3) Visual Structure, (4) Image. Thus, multiple views can be generated on three different levels [60] in between these stages:

The filter level: building different data sets by selecting data or filtering out data.

The mapping level: applying different visualization techniques to generate the main properties of an image.

The display level: changing the viewport, e.g., showing the front or the rear-view of the same object.

Multiple view and multiform

We distinguish multiple views and multiforms [60]. Multiple views means multiplicity in visualization, where various realizations of the data are depicted in separate windows. Multiform is a specific type of multiple view visualization which represent a change in the visual representation method, describing a representation that is depicted in a *different form*. It suggests that separate algorithms have been applied to the same data in order to generate a new view of that information, and that the representations are displayed in neighboring windows. In essence multiform images are generated by either altering the mapping of the data or by simplifying the data itself. Whereas multiple views could be generated on every level of the visualization pipeline, multiforms refer to the mapping level only. At this level the selected data is mapped into a geometric representation, this abstract visualization object [29] is then rendered into the final visualization.

Mapping level

In particular, generating different views (multiforms) on the mapping level can be done by these four methods [60] (compare Figure 2.6):

- The graphic-object being mapped, or the type of representation, can itself be altered. E.g., a surface of points can be changed to a surface of icons or glyphs.

- The appearance of the object can be altered by exchanging the retinal variables of size, texture, orientation, shape, and color.

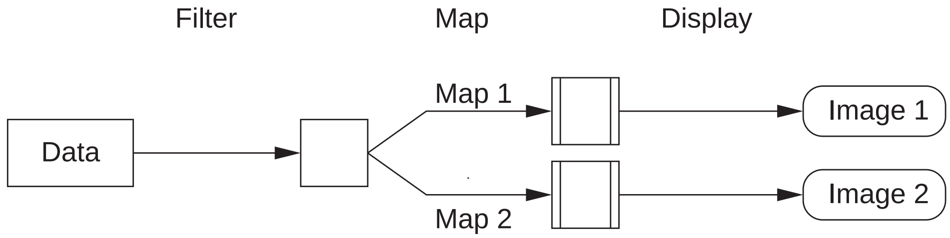


Figure 2.6: Mapping: generating different views on the mapping level.

- The mapping of the position can be changed, e.g., perspective or parallel projection versus non-linear mapping.
- The graphic-objects themselves can be simplified, by smoothing or regionalizing the components, e.g., a curved line can be transformed to a straight line.

Consistency of data and views

In general, an important challenge of an multiple view visualization is its complexity for both, users and designers. Designing interaction mechanisms needs to take consistency aspects of the views and the data into account. The user switches between different views and contexts. This is only useful if switching brings about more insights than confusion. Linking and brushing of associated data is one way to support the user.

2.5 Linking and Brushing

In General, three types of linking are distinguished:

- Linking data means to make relations within the data obvious by coloring connected data. This method is called linking and brushing.
- Navigational slaving means to connect viewpoints as described in Section 8.3
- Linking selection mechanisms, e.g., linking sliders in different Stardinates as described in Chapter 6

In the following we discuss the first category: linking and brushing, which is widely used in several tools [72; 27; 40]. In short, linking and brushing means that the user selects and highlights items in one view which are visual representations of the data. The associated items in the other views are automatically highlighted so that the user gets information about the relationship (compare Figure 8.5, p. 92).

For example if applied to scatterplots it allows to increase the dimensions by linking the data and their dimensions in terms of color [6].

Two features

Two features of linking and brushing should be mentioned:

- Insights are provoked by seeing an object or item in relation to different contexts. Linking and brushing allows to access items of the visualization according to group selections or individually. To really benefit from multiple views it is essential to make the data items graspable and distinguishable from the rest.
- The user engages in a process of repeated selection and highlighting and learns about the connection of her or his actions in one view with the data in another view. So actions of the user are propagated in distinct views of the data. This process helps the user in exploring the structure of the data.

Chapter 3

Related Work in Cognitive Psychology

This chapter discusses two different cognition theories which we believe to be relevant in the context of InfoVis mainly based on literature by Benjafield and Anderson [7] [3], a theory of visual perception by Gibson [25; 26], and some other important issues like attention, pre-attentive processes, automatic and controlled processes, perception and associated concepts. These topics should form a sound basis for statements on perception and cognition applied in the domain of InfoVis.

3.1 Approaches to Vision

Seeing is a process mainly done by our brain. The eyes enable us to generate electrical impulses related to the visual properties of our environment. However, assembling the picture from the pieces of the puzzle involves a lot of complex actions and cognitive processes. Seeing also means to picture it. Moreover, we visualize and anticipate it. In the following, we describe the most important approaches, namely the Gestalt Tradition and the Information Processing Tradition. Thereafter, we adumbrate the theory of visual perception by Gibson.

3.1.1 The Gestalt Tradition

The whole is different from the sum of its parts

The main point of the Gestalt psychology is that experience could not be understood by breaking it down into simpler units. The whole is different from the sum of its parts. In the opinion of the Gestalt psychologists experience follows certain laws:

The Gestalt Law of Organization

Figure 3.1: A line of points demonstrates the Gestalt law of organization.

’When we are presented with a number of stimuli, we do not as a rule experience *a number* of individual things ... Instead larger wholes ... are given in experience, their arrangement and division are concrete and definite’ [82]. Figure 3.1 shows an example. Usually you perceive a line formed by pairs of dots. It is difficult to see the first dot by itself, then the second, and then the third. The Gestalt psychologists thought that the way in which parts of a figure are influenced by such factors as proximity, good continuation, and closure, reflects a natural tendency toward good forms in our experience [21].

The Influence of Past Experience

This tendency toward good forms also means that the organization of the currently perceived experience is of more importance than experiences of the past. If you see a picture showing unusually designed letters for instance, it is difficult to recognize them as letters although you have learned that since childhood. Figure 3.2 shows the word ’sun’. Did you recognize it? If you cannot see it, cover up the bottom half of the figure. This example shows that previous experience is less relevant than the here and now.

Asch’s Concept of Unit Formation

It is easier to remember two or more objects, if you see them as one unitary pattern. Gestalt psychologists believed the reason is that unitary patterns are much more coherent perceptually, and can be remembered as units.

The Gestalt Theory of Thinking

A very famous experiment performed by Köhler, involving chimpanzees in the process of solving problems, showed that the animals were able to use tools in order to solve a certain task. In particular, one of the chimpanzees wanted to grasp a banana which was too far away to reach it. Köhler meant that the animal was able to understand what the situation requires. He thought it had insight in the form of: ’Aha! In order to get that banana, I will use a stick to bring it closer!’ Beside the question, if an animal could refer to itself with ’I’, this ’Aha experience’ of the animal is seen controversially by many psychologists.

Figure 3.2: Here-and-Now organization: What is it?

The findings of the Gestalt Theory have an important influence on perception. If we think of glyphs, for example, Chernoff faces [17] or icons like stick figures [28; 56], or even of scatterplots (compare Section 8.1), Cam Trees [61], or parallel coordinates [33] it is obvious that forms are perceived differently from the single symbol. That happens long before we think of meaning and association, nevertheless these cognitive processes also have high impact on perception of visualized data.

Likelihood versus simplicity

Research in Gestalt psychology ist still going on [30]. One example is the *Minimum Principle* [42]. It compares the *simplicity principle* with the *likelihood principle*. The simplicity principle states that there is a tendency for experience to presume an object or information according to its simplest possible state, whereas the likelihood principle declares that perception tends to pick up the representation of the most likely event. The experiments suggested that simplicity is a more fundamental principle than likelihood. Thus, the most likely event is seen if there is no contradiction by the simplicity principle.

3.1.2 The Information Processing Tradition

Any cognitive theory tries to describe the relation between knowledge and behavior. How is behavior influenced by knowledge? In contrast to the Stimulus-Response model of Behaviorism which goes strictly from left to right, in the information processing tradition a top-down model is propagated. This model takes two different units of behavior into account: molar units focus on certain goals, they are called strategies; Molecular units are tactics, concrete actions or ideas to reach a certain goal. Thus, behavior is organized hierarchically. Behavior can be described on several different levels. Strategies can be relatively constant, while tactics can vary widely. The processes that regulate behavior form the top down are called plans. A plan is a set of instructions for carrying out actions. Information processing psychology compares these plans to computer programs. The structure of such a plan is represented by the TOTE mechanism [46]. TOTE stands for Test Operate Test Exit. This is a feedback loop, action is instigated by an incongruity between the state that exists and the state the person is seeking. The person continues ’operating’ until the incongruity is removed. Thus, the person tests to see if an incongruity exists. If it does the person operates. After that the person tests again to see if the incongruity has been eliminated. If it has, the person exits from the routine. This reminds us of a simple do-while loop. Like such loops, TOTE units can be nested in other TOTE units. This kind of representation tries to simulate human behavior on computers.

There were several trials, most entertaining of these were early programs which imitated psychotherapists. Besides this tendency to simplify human behavior and cognition there are some interesting points. For example, that our behavior is regulated by the goals, or standards, that we set ourselves.

Scientists like Neisser argued that the analogy between cognition and a computer program is very powerful [50]. It allowed to find formal representations, which brought more clearness. Cognition was seen as the study of mental software. Neisser’s approach to the study of cognition followed the path of information processing from the time the person first began processing a stimulus. In the area of visual cognition the icon is the starting point. An icon is a visual object perceived by a person. Sperling investigated the effects of briefly exposing a stimulus to a subject [71]. Three rows of three letters were presented to the subject for a very short time - 50 milliseconds. After another 50 milliseconds the subjects hears a low, medium, or high tone. According to this tone the subject should report the letters of the third, the second, or the first row. It turned out that the subject still remembers the letters of the correct row, even if it is no longer being presented. There is a copy of the visual stimulus, in a form like a visual image, a visual icon that outlasts the stimulus itself. This experiment was very influential, partly because it implied that one could analyze cognition as a sequence of information processing stages. Thus, after the stimulus was stored briefly as an icon, it could be processed further. Different stages of information processing through which information flows were considered.

In order to read a letter as a letter, like the subjects did in the Sperling experiment, pattern recognition has to take place. This involves the rela-

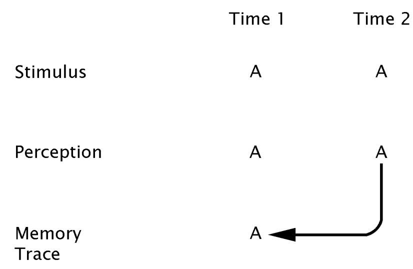


Figure 3.3: The Höffding function: a perception makes contact with a memory trace.

tion between perception and memory. If I want to recognize for instance the letter 'a' as the letter 'a', my perception of 'a' must somehow make contact with the memory trace of 'a'. The process whereby a perception makes contact with a memory trace is called the Höffding function (compare Figure 3.3), after a 19th-century Danish psychologist. Two different approaches of pattern recognition are described here, namely Template Matching and Feature Detection. They were important in the information processing tradition.

Pattern recognition

Template Matching

The process of template matching needs to compare the current configuration with the standard, or prototypical forms that we have in memory. Thus, a letter can take any one of many different forms. The prototypical pattern would differ somewhat from the particular pattern we perceive. It is hard to specify the way in which a template can match not only patterns that are identical to it, but also patterns that are similar enough to it. For this reason template models have often been criticized.

Feature Detection

This means that patterns are identified on the basis of their features or attributes, for example, size, color, shape, ... But a feature could also be that an object can be used as a tool to deal with a specific task. A bundle of features build up an object. A very prominent approach of the feature detection theories was called *Idealized pandemonium* [67]. It is based on so-called cognitive demons. Each one is able to detect a specific pattern and reports by shouting if it recognizes its pattern. So all demons may be shouting at the same time but the level of intensity depends on the degree of similarity to the object. The more similar the object is to the pattern the demon is looking for, the louder the demon shouts. On the top is the decision demon which selects the loudest cognitive demon. So the pattern is recognized in accordance to the loudest demon.

The information processing model (compare Figure 3.4) based on the model of Norman and Bobrow [53] describes the different stages of the information flow. At the beginning a physical signal like light energy is changed into a form that can be used by the cognition system, it is transduced. After that it is stored as an icon in the sensory information store. Then pattern recognition takes place, therefore long-term memory is involved. The next stage is short-term memory which contains what you are consciously aware of at any time. Obviously, the information of the sensory information storage would be lost after a very short time without pattern recognition.

This model was very influential in cognitive psychology, but it was also criticized for being too linear. Especially Broadbent argued that this one direction relation - from stimulus onward - is an unrealistically simple notion [9]. Information may be interacting in a variety of ways in a person, and not simply moving from one stage to the next. A second problem of the simple information processing model is that different people may process the same information in different ways, or the same person may process the same information in different ways on different occasions. The third complaint concerns the definition of the stages. Where does one end and the other begin? Cascading processes are not possible in this model. Cascading means one stage begins before another is finished. To solve these problems Broadbent proposed the Maltese cross model (compare Figure 3.5). Information can travel in a variety of ways through the system. The motor output store is added which holds programs of planned and / or executed actions. Sensory transduction and pattern recognition is not named explicitly anymore. In our opinion this model also has shortcomings because it is too simple.

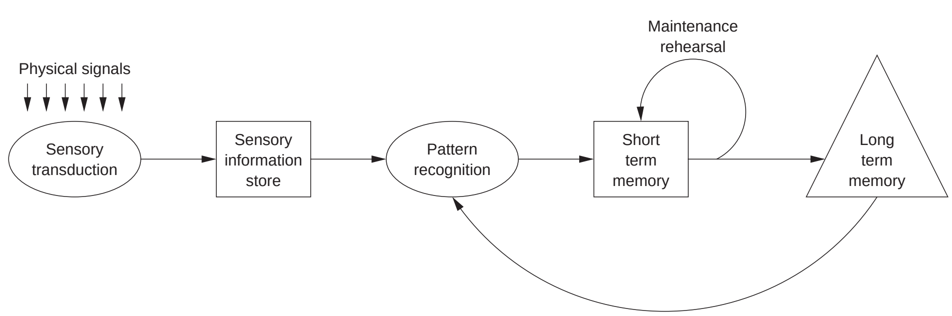


Figure 3.4: A typical information processing model describing the different stages of the information flow.

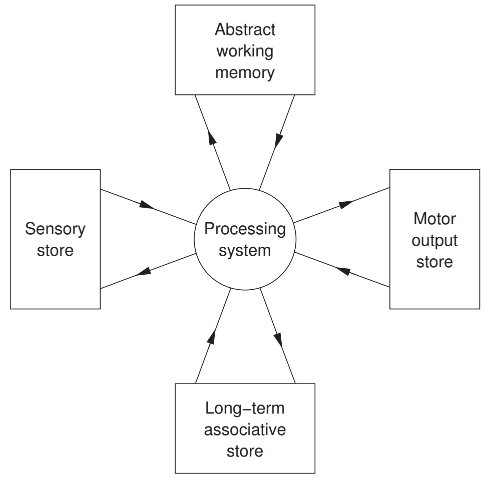


Figure 3.5: Another information processing model is the Maltese cross model. Cascading information processes can travel on various ways through the system. Therefore, it attempts to overcome linearity and simplicity of earlier information processing models

3.1.3 The Ecological Approach to Perception

Seeing by directly attending

This theory is called ecological approach to perception [25; 26] because it states that the visual environment provides the human with information. The human who is naturally an active observer perceives information about the things in the environment directly by attending to the information of the ambient optic array (compare Figure 3.6). This emphasizes the position of the observer because the structure of such an ambient optic array is described in terms of visual solid angles with a common apex at the point of observation. It structures the ambient light which is environing at every point.

What distinguishes this theory from others is its attempt to understand visual perception as a process which includes not only our eyes and our brain, but also the fact that the eyes are related to the head on our body and that we are supported by the ground. Visual perception is done in parallel to moving and changing the point of observation.

Medium, substances, and surfaces build up the environment, and their different textures for instance represent the distance to an object.

In contrast to other concepts which describe human vision similar to the mechanism of a camera, it argues that perception is direct and not mediated by a process of inference, such as sensations constructing percepts.

Radical different approach

This theory could be seen as an radically different approach, disagreeing with the ideas of the retinal image and that we see the environment in dimensions because we know Descartes's concept of three axes describing

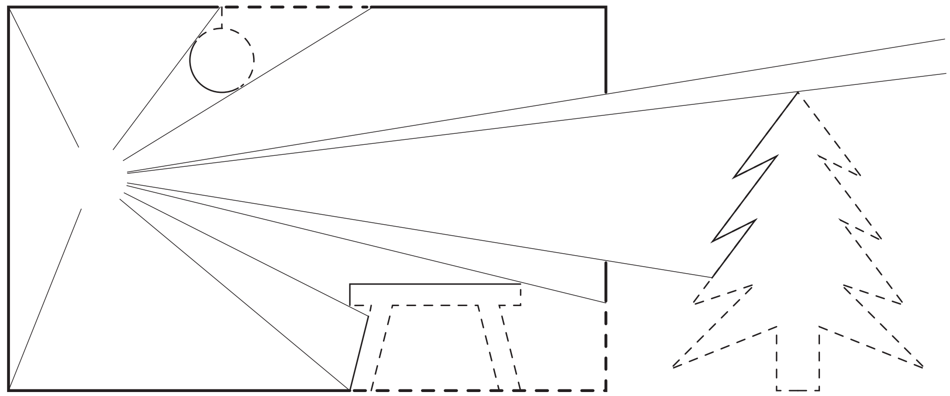


Figure 3.6: Ambient optic array: viewing the environment.

the space. Moreover, it states that the visual perception depends on moving in the environment.

In our context, the observation aspect seems interesting. Exploration of data is often compared to moving in the information space using the metaphor of locomotion in space. We think that the ecological approach to perception yields some interesting questions and ideas for the perception of artifacts utilized for interaction with the visualization (compare Chapter 6).

3.2 Consciousness and Some Rest of the Iceberg

For the cognitive complexity of visualization, properties of attention, awareness, and unconscious processing are relevant. Therefore, we discuss these issues in the following. Finally, we will see that awareness of people is deliberately structured so that they yield the most useful organization in a particular situation. This plays an important role in terms of exploration and analysis of data.

3.2.1 Selective Attention

Theories of the information processing tradition (cf. Section 3.1.2) state that one central processor handles all tasks. That means that a person can only pay attention to one thing at a time. Obviously, we are able to do more things at any given time. Automatic and overlearned tasks (cf. Section 3.4), like walking or driving or other tasks which do not need full attention, are often done synchronously. So doing more than one thing requires to rapidly switch from one task to another and select to which task to attend at a time.

Switching attention

Divided attention

On the contrary, divided attention which is a learnable technique enables people to really attend to more than one task at a time. One example is listening to a conversation and reading at the same time. Multi-modality – that means signals use different channels – seems to be one condition. So, it would not be possible to read two texts synchronously.

And looking at two pictures at the same time? This might be an interesting question if using multiple view visualizations. Actually, reading and looking is not really comparable. Although there are some exceptions involving specific techniques, reading usually requires attention and is done sequentially because we read to understand or even remember what was written. Looking at something appears to be a bundle of different processes – parallel and sequential – with a varying degree of attention (cf. pre-attentive processing Section 3.2.2) and influenced by previous experience (cf. associated concepts Section 3.3) and expectation. Obviously, there is a difference between glancing at two images on the one hand and examining them on the other hand. The latter would require selective attention.

3.2.2 Pre-attentive Processing

Visually perceived objects which can be processed pre-attentively are best characterized as popping-out patterns. The perception of such objects usually takes not more than 10 milliseconds per item. Non pre-attentive objects need at least 40 milliseconds per item [73]. This is based on certain properties, like orientation, size, motion, basic shape, lightness, and color which are perceived by a person without being fully aware of it (compare Figure 3.7). Especially, fast motion is a very striking feature, although there is no well-known order which property is stronger than the others. Since these properties are usually combined, this would be an interesting question.

Properties like orientation, size, motion, shape, color

Outside awareness

Pre-attentive perception happens outside of awareness. This is an interesting feature, as Colin Ware points it out: ’In displaying information, it is often useful to be able to show things *at a glance*. If you want people to be able to instantaneously identify some mark on a map as being of type A, it should be differentiated form all other marks in a pre-attentive way’ [78]. InfoVis benefits from the possibility to guide attention, but on the other hand this could also be a problem. If pre-attentive processing is the basis for navigation through an information space, people are actually not aware of their navigation decisions. Maybe this is in contrast to carrying out sophisticated problem solving processes like getting new insights. On the other hand it could be effective for immediate and spontaneous navigation which could be seen as the usual interaction process of searching for information in the WWW.

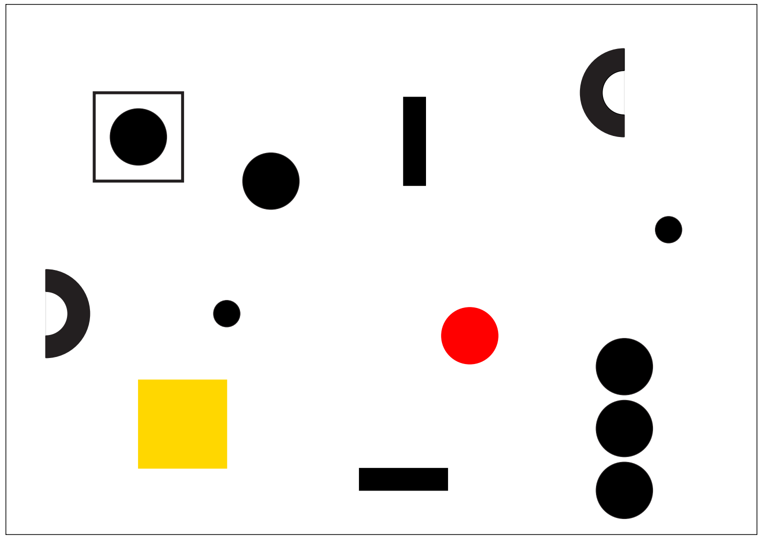


Figure 3.7: Pre-attentive cues: each of the 10 symbols is pre-attentively distinct from the others.

A pre-attentive process can be described by the time it needs to find an item. As mentioned above, this time is very short and optimally does not depend on the number of distractors. Such distractors are objects which are presented among the targets but do not convey any information. However, a high number of such objects, each one pre-attentively different form the others would be very confusing. Beside that, conjunction search tasks, such as looking for a blue circle within a picture of blue squares, red circles and other objects, are usually not pre-attentive. There are some exceptions, e.g., Nakayoma and Silverman [49] showed that the combination of stereoscopic depth and color, or stereoscopic depth and movement, is pre-attentively processed. Most research work on pre-attentive processes deals with the detection of isolated targets. There are clear advantages, if you want to guide attention. Since every kind of InfoVis has pre-attentive properties because they are all based on spatial location, shape, orientation, and size, it is both necessary and interesting to take the findings [73] on pre-attentiveness into account.

3.2.3 Attention and Awareness

Figure and ground

According to the Gestalt theory people distinguish between *figure* and *ground*, which means they organize their percepts into a part to which they pay attention and another part to which they do not attend. A lot of experiments exist in which an image is presented for a short time (e.g., for 100 milliseconds) and then the person is asked what she or he remembers. Often certain aspects of the image influence the perception, although they were part of the background to which the person did not pay full attention. This phenomenon is known as the effect of concealed figures [22]. It seems that we encode information from the background without being fully aware of it.

Encoding means to transform information into one or more forms of representation. E.g., a person hearing the word *horse* encodes this information into the broader categories of beasts of burden, four-legged creatures, mammals, warm blooded animals, and finally of animals in general [83]. It seems that this process is done rather automatically because we encode more dimensions of the world than we can be fully aware off.

Functional way of organizing

In this context consciousness could be seen as a functional way of organizing what we are given [43; 44]. The border between unconscious and conscious perception is fuzzy. Normally, we can only consciously attend to one thing a time but we utilize a lot of other information. Awareness could be described as deliberate structuring of information so that it yields the most useful organization of information in a particular situation.

3.3 Activation of Associated Concepts

Experiments showed that an object or word was recognized faster if another associated object or word was presented previously, e.g., 'dog' could be associated with 'bone'. One explanation for this result is the spread of activation within the memory which is seen as an interlinked network of nodes [2]. This network represents semantic concepts and is built up of neurons which hold specific levels of energy.

The activation of one concept in memory causes the activation of associated concepts. This process is called spread of activation. Objects belonging to activated concepts are recognized faster. Moreover, concepts which are activated more often or were activated recently are easier to access. The method of priming utilizes this effect. By prior presentation of a certain word or object the corresponding nodes of the network are activated, so the related concepts are also ready for faster access.

Two determining factors are interesting in this context: First, the associative strength between two concepts, e.g., the association between 'house' and 'window' is normally stronger than the association between 'house' and 'attic'. Second, the number of associations. Interferences among competing associations of concepts can occur which limit the spread of activation. For instance the fan effect states that the level of activation depends on the number of associations emanating from it. Particularly, the degree of activation of one node is inversely proportional to the number of associations. So recognizing words or objects of concepts with a high number of competing associations takes longer. This corresponds to the idea that expectation and experience influence perception automatically.

3.4 Automatic, Controlled, and Overlearned Processes

We distinguish between automatic processes on the one hand and controlled processes on the other hand. The first one is done without attention and highly autonomous. For instance, reading a word – in contrast to reading a longer text and understanding its meaning – is such a process. It has a tendency to be executed whether or not one wishes to do so. Automatic processes seem to avoid attention or control. Information tends to be picked up automatically [51]. It is an interesting question whether we tend to perceive the color of the word *green*, written in blue color, or its meaning. The perception of the color seems to be masked by the meaning we read. Supressing the automatic process appears to be awkward.

In contrast, controlled processes [68] need our attention to be executed accurately. Examples are learning to drive a car or speak a foreign language. Typically learners' attention is completely engaged in such tasks, but controlled processes could become automatic if they are overlearned. Therefore, practice can influence the degree of attention needed.

Chapter 4

A Glance at the Neighborhoods

When designing an InfoVis technique several other aspects not directly related to the visualization itself become relevant because each visualization of data should serve specific purposes. In particular, its main job is to support the user in accomplishing her or his tasks. Therefore, it needs to be task-specific and user-oriented. That does not imply an individual tool for every user but it requires to get in touch with the users, their tasks and goals. In order to design appropriate interaction techniques – which we assume to be an integrated part of the visualization technique – we want to focus on the users' tasks. Thus, we discuss three different approaches. Each provides us with categories of tasks in order to build a comprehensive structure for our visualization technique. The first emphasizes a more formal approach, the second emanates from a more practical point of view on users' needs, and finally, the third brings out the cyclic and interrelated nature of the relevant tasks. So the users, the tasks, and the data constitute the environment in which the visualization tool will be deployed (compare Figure 1.1).

4.1 The Basic Visualization Interaction Framework

The 'Basic Visualization Interactions' (BVI) [20] build up a framework for characterizing interactive visualization techniques. Such a basic visualization interaction is described by its input, output, and operation. We distinguish three kinds of operations:

Three types of operations

Graphical operations can be divided into encode-data, set-graphical-value, and manipulate-objects. These operations change graphical attributes, change the mapping of graphical objects and data or operate on the graphical object as a unit of manipulation (e.g., zooming, coloring, ...) and affect the graphical representations of data objects (graphical state).

Set operations expand the underlying data by classifying information and selecting data. These operations include creating, deleting, summarizing, joining, intersecting data sets, and so forth. A set operation manipulates the data in order to create or alter a selection of data to be visualized. It affects the control state.

Data operations affect the data state (database). Users can discover new facts during analysis or as mentioned above create useful classification information. It is useful to enrich the data with such new information contents.

Changing a certain state can cause secondary effects to other states. A set operation usually leads to a change of the graphical state. The manipulation of the graphical state could cause a modification of the control state and vice versa. An example is the handling of a graphical object in order to select data. Although there are some borderline cases which are difficult to relate to this structure, this framework is a good starting point for further examining the requirements for an interactive visualization.

We used this framework as a formal background for the implementation of LinkVis (compare Section 8.1) which is a multiple view tool. In addition, we started our considerations on interacting with the Stardinates (compare Section 6) by utilizing this scheme.

4.2 The Visual Information Seeking Mantra

This framework —based on a **task by data type taxonomy** [69] – is another useful starting point for designing user interfaces and interaction mechanisms in InfoVis tools. It is rather too general and for designing concrete tools it needs to be supplemented by other approaches (compare our approach for the Stardinates, the graphical object operation scheme 6.4).

In order to understand the overall concepts of interaction with an InfoVis tool this taxonomy builds up a useful categorization of tasks. Moreover, it suggests a procedure of repeated tasks for using InfoVis tools to find information, the visual information seeking mantra.

The task by data type taxonomy lists seven types of data and seven types of tasks which should be supported by an InfoVis tool:

Seven types of data

Data Types

- 1-dimensional:

For instance linear data types, such as textual documents, program source code, and alphabetical lists of names, organized in a sequential manner.

- 2-dimensional:

  This kind of data could be planar or map data including geographic maps, floorplans, plans for public transport or streets, layouts of websites or newspapers.

- 3-dimensional:

  Meaning items with volume such as buildings, objects, the human body, and space in general. This category refers to real-world objects.

- Temporal:

  Indicating that the data has a specific start and finish. Such events or processes could overlap or occur repeatedly.

- Multi-dimensional:

  This category refers to data characterized by a number of attributes, which could be seen as a point in a n-dimensional space. In InfoVis such data usually is mapped into 2D space utilizing additional retinal properties, such as color, shape, orientation, and size to represent the remaining dimensions.

- Tree:

  Hierarchical data characterized by a clear structure as parent and child items could be represented as a tree.

- Network:

  This data type consists of items with complex relationships, so that such data is semi-structured because there is no clear hierarchy within the structure. It requires adjusted techniques for representing the underlying structure.

Seven types of tasks

Task Types

- Overview:

  Means to zoom out in order to see the overall structure of the data. Therefore, the data is placed in abstract categories which can be understood at once. For instance, Figure 5.3, p. 40 shows an overview sight of the Stardinates.

- Zoom:

  Refers to magnifying or scaling down by utilizing a zoom factor and a zoom position usually by mouse cursor. This form of interaction is very intuitive. Smooth zooming helps to preserve the sense for position and context within the information space.

- Filter:

  Users can eliminate or filter out uninteresting data in order to reduce the amount of data and allow for a more efficient search of relevant information.

- Details-on-demand:

  After selecting individual items or groups sharing specific attributes the user can get details when needed. So that especially attributes which are not sufficient for visualization become accessible in the context of a visualization. Pop-up windows appear adequate for displaying details.

- Relate:

  Helps to examine relationships within the data by highlighting, linking and brushing or otherwise relating data items. This task is useful either when doing complex queries for selection purposes or using multiple view visualizations.

- History:

  Means that a history of user actions is available which allows for undoing or redoing interactions. Since exploration usually is a process of repeated interactions this feature is absolutely necessary and widely used nowadays.

- Extract:

  After selecting data sets or single data items users should be able to extract this data set, for instance by printing it, saving it to a file or sending it by email to others. In other words gained information should be further processable.

We do not think that a visualization tool necessarily needs to support all of these data types. In our case the data could be characterized as network data partly showing hierarchical structure (compare Section 5.1). However, the seven tasks described above are widely accepted because all of them are really useful in practice. Based on these tasks a procedure for searching for information is defined—the mantra: overview first, zoom and filter, then details-on-demand. And repeated for several times. Clearly, interactive exploration is an iterative process.

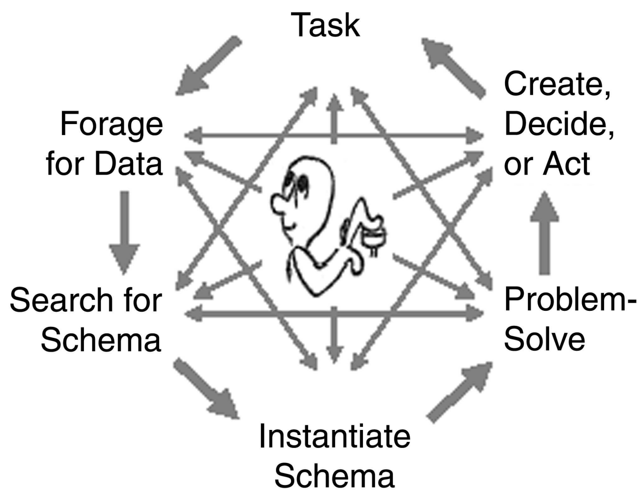


Figure 4.1: Knowledge crystallization [14]

.

4.3 Knowledge Crystallization

The process of searching and getting information in order to achieve a certain goal can be seen as knowledge crystallization [14] and means to get insight about data relative to some task. Figure 4.1 shows the different stages of this process. Each stage could be described by typical elements shown in Table 4.1.

Process of abstraction

Knowledge crystallization usually requires finding schemes or representations for the data that are effective in terms of accomplishing certain tasks. Data are coded by this representation. But some rest of the data is either unencoded or encoded inefficiently. In order to encode all required data sufficiently repeated improvements of the schema are needed. This process of abstraction or schematization and omission of information is a fundamental principle of how a mass of information can be made manageable.

Characteristics of knowledge crystallization

In short, a knowledge crystallization task is characterized by the following three steps:

- Gathering information for some purposes.
- Making sense of it [62] by constructing a representational framework (schema).
- Packaging it into some form for communication or action.

Stage	Elements
Forage for data	Overview Zoom Filter Details-on-demand Browse Search query
Search for schema	Reorder Cluster Class Average Promote Detect pattern Abstract
Instantiate Schema	Instantiate
Problem-solve	Read fact Read comparison Read pattern Manipulate Create Delete
Author, decide, or act	Extract Compose

Table 4.1: Knowledge crystallization: stages and their elements

Moreover, it is characterized by the use of large amounts of heterogeneous information and ill-structured problem solving, but a relatively well-defined goal.

InfoVis facilitates knowledge crystallization

So, knowledge crystallization requires data, a task, and a schema. If the data are not at hand, InfoVis can aid in the search for it. If there is a satisfactory schema, knowledge crystallization reduces to information retrieval. If there is no adequate schema, InfoVis is one of the methods by which an improved schema can be obtained. Thus, InfoVis offers useful functionality for handling such an interrelated process of foraging for data, searching and instantiating schemes, solving problems, and applying these solutions. Therefore, capable interaction mechanisms are required.

Chapter 5

The Stardinates

5.1 Data Characteristics

Generally, InfoVis deals with the visualization of different types of data from different sources with no inherent spatial meaning. Every visualization is optimized for some types of data. Therefore, we want to give a short description of the data the Stardinates can handle. The Stardinates are adequate for ordinal and nominal values. Ordinal means that the values are in a specific order or rank, e.g., the number system of natural numbers. Whereas nominal indicates that the data is classified or related to specific categories which do not show any quantity or ranking among them.

Ordinal and nominal values

On the one hand the complexity of the data structure could be very high, since every axis of one Stardinate can depict another dimension, on the other hand, the amount of data (frequency) and the parameter space is restricted because of the glyph-size of the Stardinates. If the number of lines or values is too high, the shapes of the Stardinates lack expressiveness so that they are hardly distinguished. We tested the Stardinates with up to 20 axes considered as the maximum of axes of one Stardinate. Therefore, this method is not suitable for high frequency data.

High complexity

5.2 The Method

The Stardinates (compare Figure 5.1) are an InfoVis method which could be described as geometric glyph capable of the exploration and analysis of complex data [41]. We use axes (x_1 to x_n) to represent the various dimensions of our data. The axes are arranged in a circle. Each Stardinate consists of a number of axes with scales and labels. The data values are displayed as points on the scales. In order to emphasize the data points, two data points on adjacent axes are connected by a line. So one data record is displayed by a line with its vertices on the axes, we call this the

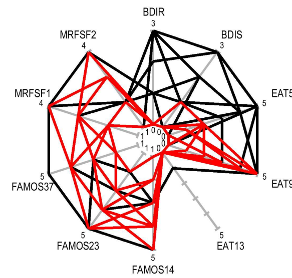


Figure 5.1: A Stardinate visualizing 10 Dimensions, due to selection some data lines are colored red. For description of dimensions compare Table 6.1, p. 47

.

Typical star-shape

data line. This results in the typical star-shape which is perceived efficiently. In contrast to the parallel coordinates we use closed star-shapes. The last axis x_n is placed next to the first axis x_1 and its values are connected. The same property holds for the other adjacent axes. According to the Gestalt laws [7] such shapes are perceived intuitively. Like glyphs a number of Stardinates are displayed next to each other in order to extract one dimension. Typically, meta dimensions, like time or other dimensions of special interest, are extracted this way to make the visualization more understandable. We will illustrate this in an concrete application (compare Figure 8.5, p. 92). Generally, the search for multivariate relations in n-dimensional data is transformed into a 2-D pattern recognition problem. Large shapes usually indicate high values, small shapes imply low values.

Pattern recognition

The data line is characterized by color, saturation, brightness, sharpness, thickness and mode (dotted, etc.). This is the basis for selecting and filtering data by the user. To provide the user with detailed information every data line offers context information on demand if the user focuses on it. In contrast to the parallel coordinates equal values cannot be recognized by straight lines. In order to ease the recognition we use a reference line which connects equal values on the axes on demand. The user can display, move, or hide the reference line according to her or his needs.

All axes are of equal length. Since the ranges of values could be different among the dimensions we emphasize the relative comparison of minimum and maximum values by using equal lengths. This could be seen as distortion of the axis, which prevents the user from associating absolute values of different ranges of values.

In order to enable the user to explore the data interactively, the visualization tool needs to offer direct interaction techniques based on the users' tasks and combined with immediate respond times.

In the following we specify the Stardinates in detail by mathematical definitions for the axes, the data lines (record lines) and its vertices (data points).

5.3 Formal Description of the Stardinates

The number of the axes of one Stardinate is derived from the number of dimensions of the visualized data. So we define the axes by:

$$x_1, \ldots, x_n \quad \text{with } n \in \mathbb{N} \text{ and } 3 \leq n \leq n_{max} \tag{5.1}$$

Number of axes

n_{max}: As mentioned above an upper limit for the number of dimensions to $n_{max} \leq 20$ seems useful.

Angle of the axes

The similarity the Stardinates and the polar coordinates is obvious. From a geometrical point of view the axes of the Stardinates are placed on the coordinate lines which emanate from the pole of the polar coordinates radially. We start with x_1 at twelve o'clock position and then place the rest of the axes clockwise. So we describe the angle for each axis by:

$$\varphi_{x_k} = \frac{\pi}{2} - (k-1) \cdot \frac{2\,\pi}{n} \quad \forall\, k \in \mathbb{N},\ 1 \leq k \leq n_{max} \tag{5.2}$$

Length of the axes

The axes have a specific length according to the size of the window dedicated to one Stardinate. Based on the size of the window $window_x$(width),$window_y$ (height) we define the length of the axis by:

$$l = \frac{3}{7} \cdot \min(window_x, window_y) \tag{5.3}$$

The coefficient $\frac{3}{7}$ results from the ideal case $n = 4, 6, 8, \ldots$ when the center (pole) of the Stardinate is placed at $\left(\frac{window_x}{2}\,;\frac{window_y}{2}\right)$ in the center of the window. If $n = 3, 5, 7, \ldots$ the lower space is used inefficiently because the lower axes are inclined and therefore, do not need the same vertical space like straight axes. But this is nearly irrelevant when $n \geq 7$ because then it converges rapidly to the ideal case. In the center of the window $\frac{1}{7}$ of size of the window is reserved for labels etc. That means we shift each axis out for $\frac{1}{6}$ of the axis length. Together with the $\frac{6}{6}$ of the axis we get the total size of $\frac{14}{6}$. The reciprocal value gives us the length of the axis in relation to the size of the window. Finally, we derive the definition of the axes by:

Definition of the axes

$$x_k : \ \frac{l}{6} \leq \varrho_{x_k} \leq \frac{7\,l}{6} \quad \text{and } \varphi_{x_k} \text{ from (5.2)} \tag{5.4}$$

Next we define the data points and the lines connecting the points on adjacent axes. Therefore, we specify minimum and maximum scale values for each axis: $x_{i\ min}$ = minimum scale value on axis x_i and $x_{i\ max}$ analog.

To display the value of a specific data record (data record a) we need to calculate the position in polar coordinates. So our input $d_{a,i}$ is a value of the dimension i of the data and results in the point $D_{a,i}$ which is displayed on the axis x_i. In our definition we distinguish the negative and the positive sections of the scale in order to enable the user to select different scales or different min/max values for every axis.

We derive the position of the point at the coordinates:

Definition of a data point

$$D_{a,i}(\varrho_{a,i};\varphi_{x_i}) =$$

$$D_{a,i}\left(r_{fix} + \frac{(d_{a,i} - x_{i\ min})\cdot l}{x_{i\ max} - x_{i\ min}}\ ;\varphi_{x_i}\right) \tag{5.5}$$

with $r_{fix} = \frac{l}{6}$ analog to (5.4)

The data points of one record on adjacent axes x_i and x_j are connected by a line. For the values of $D_{a,i}$ on axis x_i and $D_{a,j}$ on axis x_j we define the line $l_{a,i,j}$ connecting these two points:

Definition of a data line

$$D_{a,i}, D_{a,j} \quad \text{with} \quad 1 \le i \le n \tag{5.6}$$

$$\text{and} \quad j = \begin{cases} i+1 & \text{if } i \le n-1 \\ 1 & \text{if } i = n \end{cases} \tag{5.7}$$

$l_{a,i,j}$ given by $\varrho_{a,i,j}$ within the range of $\varphi_{a,i,j}$:

$$\varrho_{a,i,j} =$$

$$\frac{\varrho_{a,i}\cdot\varrho_{a,j}\cdot\sin(\varphi_{x_j} - \varphi_{x_i})}{\varrho_{a,i}\cdot\sin(\varphi_{a,i,j} - \varphi_{x_i}) - \varrho_{a,j}\cdot\sin(\varphi_{a,i,j} - \varphi_{x_j})} \tag{5.8}$$

$$\varphi_{a,i,j} = \begin{cases} \varphi_{x_i} \ge \varphi_{a,i,j} \ge \varphi_{x_j} & \text{if } i < j \\ \varphi_{x_i} \ge \varphi_{a,i,j} \ge -\frac{3\pi}{2} & \text{if } i > j \end{cases} \tag{5.9}$$

In only one case the value of i is higher than the value of j. That happens when comparing the axis x_n with x_1.

Equation (5.8) describing the data line is undefined if the denominator is '0'. By an indirect proof (compare detailed derivation at the end of this subsection) we can show that this never happens if we stick to the formulas above (hint: $n \geq 3$ and $r_{fix} > 0$). Thus, the definition of the line connecting two data points of the Stardinates is proved to be valid.

This mathematical definition describes how Stardinates are constructed and provides us with a good starting point for the implementation of a Stardinates tool.

5.4 Swiss Cheese and the Stardinates

Perception of texture and patterns

What does cheese have to do with the Stardinates you might ask. These two have a lot in common as we will show. Imagine a fine tasting, spicy piece of Swiss Cheese (e.g., Figure 5.2) with its characteristic holes. Hungry yet? However, for this purpose it does not matter if you like this kind of food or tend to see it as uneatable. We are just interested in the visual properties of Swiss Cheese. When looking at it we perceive a visual *texture* based on the manufacturing method and the ingredients on the one hand and *patterns* based on color, size, shape, position, closeness, occurrence, regularity, deepness, and bend of the holes on the other hand. In general, many small, similar but distinguishable objects placed next to each other on a relatively small area result in a textured image. If these objects are larger or more complex or simply scattered on a larger area, they are rather not perceived as texture anymore but we recognize patterns among them. Texture [79] conveys information about the surface, e.g., rough, smooth, soft, crispy material. So, if the single object recedes behind the accumulation of the same or rather the perception of the whole we talk about texture. On the contrary, segmentation and repeated depictions of similar visual primitives characterized by placing the primary emphasis on the single object are usually perceived as patterns [80].

Figure 5.2: Swiss cheese

Interesting research on pattern perception was done by a group of German psychologists. In particular, they invented the Gestalt laws (compare Section 3.1.1). Early iconographic approaches which are interesting in our context, not only because they can be considered as the roots of the glyph visualizations (compare Section 2.2), e.g., the stick-figure icon [28; 56], refer to visual texture perception. In the following development the glyphs became more complex and could encode more dimensions than icons. So in contrast to their ancestors – the icons – the glyphs rely on pattern perception. Although the holes of a cheese are usually round they are in some respects similar to our Stardinates. Both form patterns. Figure 5.3 shows 30 Stardinates, depicting fictive medical data of six patients, based on 50 indicators obtained at three different times. Particularly, four attributes of the Stardinates stand out which are seen at the first glance[1] just like the properties of the holes in the cheese:

Pre-attentive features of the Stardinates

Shape: the shape is more precisely known as 'Gestalt' which is built up from the lines connecting the data values on the axes. Intuitively we see bulges or caves and indentation which affect the orientation and accentuation of the item. In order to intensify the perception of the shape the axes and labels of the Stardinates can be hidden for experienced users.

Size: closely related to the shape we perceive the size of the Stardinate. Although we are not able to distinguish different sizes at a fine granularity, maximum and minimum values and a small number of intermediate stages are easy to recognize and form a strong attribute.

Relative Position: information regarding the position in relation to other objects is also perceived in the early stages. This contains attributes like closeness, occurrence, and regularity of the primitives.

Diversity and Accumulation of Lines: last but not least we identify crowded or single lines shown by the contour of the Stardinate. So a Stardinate looks fringed if the data values are of high diversity. It appears smooth if the data values are more steady.

Three stages of perception

Generally speaking, we distinguish three different levels of perception characterized by pre-attentive or attentive, parallel or sequential processing of a certain number of visual objects in relation to specific knowledge about these objects. Pattern recognition occurs between perception of simple properties (e.g., form, motion, color, and stereo depth) on the one hand and full object identification on the other hand [81]. First level perception is pre-attentive and done in parallel. On the contrary, third level perception is done consciously and sequentially. In other words, pattern recognition on the second level of this scheme is done without being fully aware of it. It connects processes of the first and the third level, which means it can initiate conscious recognition processes involving working memory.

[1] However, when perception becomes consciously the results seem more relevant for analysis and exploration in InfoVis since these complex tasks are normally conducted by aware users

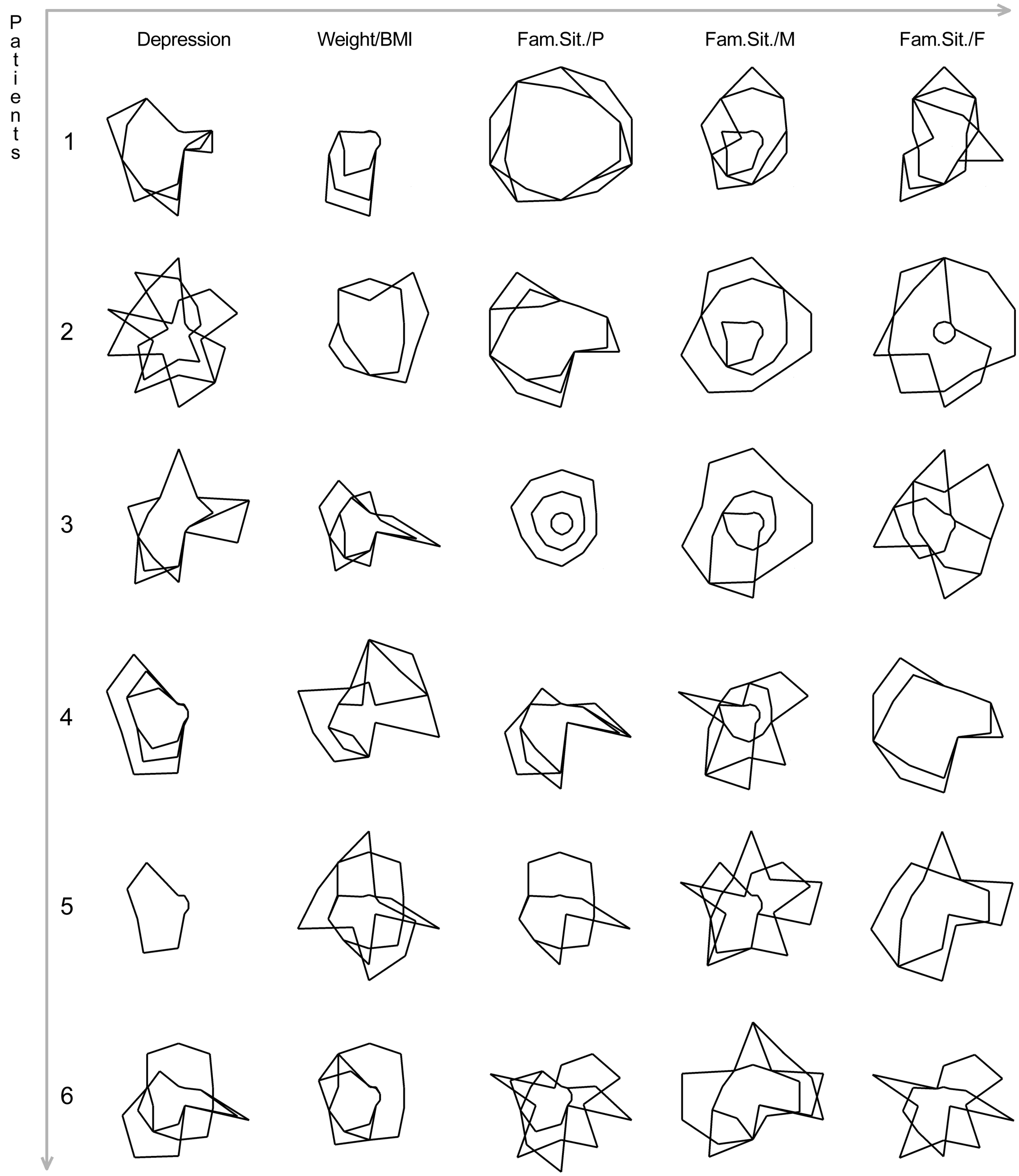


Figure 5.3: Stardinates visualizing five bundles of fictive data for six patients. Groups of five Stardinates visualize the patients' states. Each Stardinate displays 10 indicators collected at three different times and each indicator is assigned to one axis. But these 10 axes are hidden in order to focus on Stardinates' features, like size, shape, relative position, and accumulations or diversification of data values. To obtain more details the user can display axes, scales, and labels or alter the level of magnification.

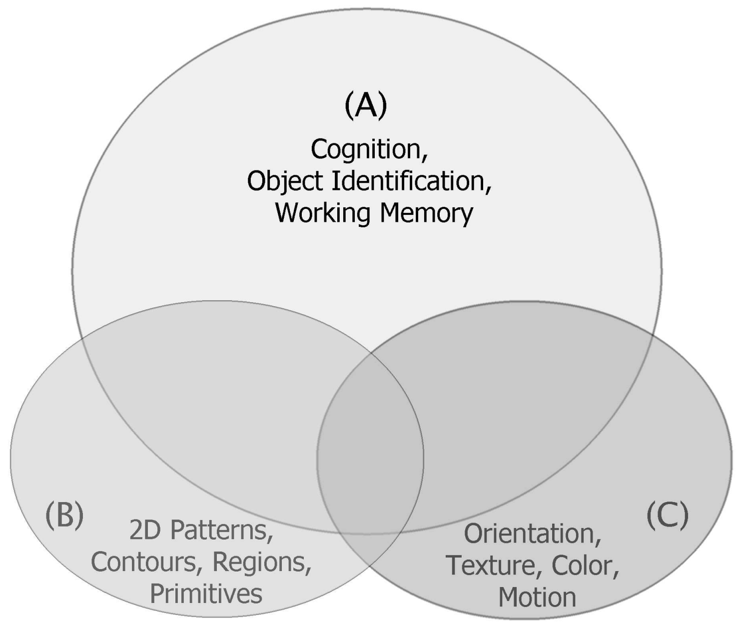


Figure 5.4: The three zones of perception

In addition, it is influenced by cognition and experience. This classification gives us a simplified idea of perception. On the one hand it is useful and generally accepted to describe *perception* and analyze its different processes according to their durations or attentive and pre-attentive parts. On the other hand it oversimplifies and tends to ignore *cognition* and its interweaved effects on how we perceive and see things. In our opinion the serial order of the stages is not suitable because processes of all three stages could take place simultaneously, partly influenced by each other or rather independent. To take these characteristics into account we suggest another illustration (compare Figure 5.4) of this classification. Instead of stages we talk about three different zones of perception.

Zones of perception instead of stages

In the center we see the zone A (light gray) which is the former third stage, left hand zone B (middle gray) representing the former second stage, and finally right hand zone C (dark gray) which is the former first stage. Beneath we sketch some examples[2] of such processes located in the intersection areas:

[2]As already stated such classifications are always arbitrary. Moreover, they are subjective simplifications according to their purposes. In our case we try to reflect the properties of the Stardinates in respect to common classifications used in InfoVis.

Color and Memory Intersections A/C or A/B/C: this could occur if a certain color reminds you of a situation. For instance a warm yellow color could let you think about your last vacation and all the bright sunshine there. But cognition influences perception of color also, e.g., a motorist is trained to watch out for combinations of red, yellow, and green lights on the street (like traffic lights), therefore, such colors would get attention earlier.

Experience and Pattern Recognition Intersections A/B: it seems that experience and expectations influence pattern recognition processes. One such example, according to the likelihood principle (described in Section 3.1.1) our experience can influence the way we see objects. Another theory (compare Activation of Associated Concepts Section 3.3) in cognitive psychology states that objects of associated concepts in long-term memory are activated concertedly. So, objects which usually occur in conjunction are recognized faster. Full object identification is usually also affected by expectations and experience.

This does not solve the problem of arbitrary abstraction but it makes the scheme more tolerable. Obviously, we give the third zone a more important role in this scenario.

The cognitive aspects of the Stardinates become interesting in this context. This brings us to a more comprehensive view of the features of the Stardinates which we sketch in the following.

5.5 Benefits and Features

Gestalt principles

Generally, visualizations have several advantages [14] from a cognitive point of view. The Stardinates seem to support some of these cognitive mechanisms in a fairly efficient manner. They represent some Gestalt principles very well, especially the principles of Closure and 'Prägnanz'. As a consequence, Stardinates form very distinct and memorable patterns which make abstraction and aggregation much easier.

They also support perceptual inferences (compare [15]) because the meaning of the elements of Stardinates are intuitively fairly clear. In contrast to that, glyphs, for example, require much more effort to learn the meaning of the elements of visualization.

The possibility to manipulate the data and interact with the form of representation is another considerable advantage of the Stardinates. Manipulating representations enables users to produce visualizations which are specifically tailored to their needs. In many cases, such representations are more meaningful for the users and better adapted at helping the users to solve their problems.

In summary, the interactive Stardinates offer a number of benefits and features:

Features of the Stardinates

- The interactive Stardinates visualize overviews very effectively.
- They make differences or similarities within the data obvious. So, comparing different Stardinates is done intuitively.
- They are capable of visualizing data with high complexity.
- Stardinates motivate user interaction in order to examine the data at hand more thoroughly.
- They build distinctive shapes which support pattern recognition.
- Such glyphs decompose complexity of data into more manageable pieces.
- The interactive Stardinates allow for visualizing data on different levels of details with consistent visualization metaphors utilizing cognitive features (compare 'Activation of Associated Concepts' 3.3).

The interactivity of Stardinates (compare Chapter 6) also helps users to formulate and test hypotheses, thus gaining deeper insights into the subject at hand. Stardinates share some of these advantages with other methods of visualization. What is perhaps unique about Stardinates is the combination of these advantages.

5.6 Shortcomings

Some problems and shortcomings of the Stardinates are not solved so far. Similarly to the parallel coordinates, data lines with equal values cover each other. So the user can not see, if a line represents one or more data records. Another problem occurs if the data are incomplete. This could happen easily with data derived from questionnaires. A possible approach is to add an extra point on the scale for missing values. However, this does not seem to be the best solution. Also, other approaches like drawing only parts of the polygon do not fit properly.

Incomplete data

To mention two other restrictions: The parameter space and the number of axes is limited. The number of Stardinates itself is also limited, because it is not possible to compare hundreds of Stardinates efficiently. Selecting and highlighting techniques help to deal with such problems.

Chapter 6

Interacting with the Stardinates

To communicate the importance of interaction and show its appealing possibilities for exploring and analyzing data, a theoretical approach is not really sufficient. However, it is our intention to point out that interaction is a significant part of an InfoVis technique, like the Stardinates. So, before we describe interaction processes with the Stardinates in detail, we start with a concrete example which gives a more vivid impression, although static images imply several restrictions for this demonstration. You can find animated demos and news about the Stardinates on our website[1].

6.1 Example: Selecting by Using Sliders

One typical interaction task is selecting data sets according to certain attributes of the data. Such selections help the user to find similarities or common patterns within the data.

Concrete example of interaction

In order to describe interaction more vividly we start with a concrete example. Particulary, how the users makes selections of data sets by the use of sliders. Sliders allow for filtering out the uninteresting data and highlighting the relevant data.

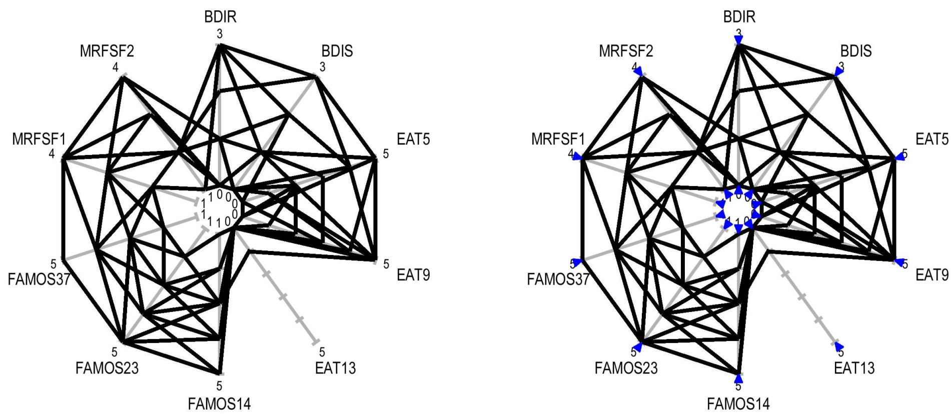
We use sliders in form of little arrows attached to the axes. They are always shown on top of the data lines and stick out because of their eye-catching color. Moving the slider is done by mouse intuitively. The top of the arrow points to the selected data lines. If the user turns around the arrow the selection is inverted. Both arrows attached to one axis are turned simultaneously.

Before we start with details about the sliders, we describe the data which is used in this example: Figure 6.1 (a) shows a Stardinate visualizing (partly fictive) psychotherapeutic data derived from questionnaires.

Interpretation of the data

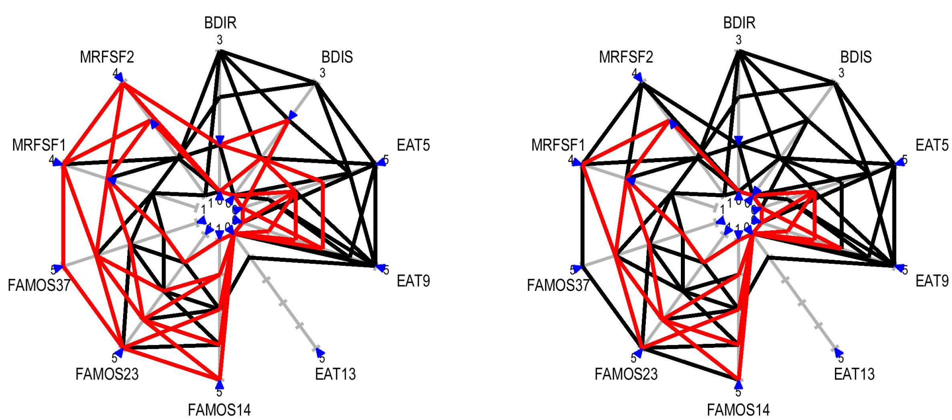
At the first glance we recognize a dichotomy within this Stardinate.

[1]http://www.ifs.tuwien.ac.at/~ mlanzenberger/Stardinates



(a) Stardinate with disabled sliders (b) Stardinate with blue sliders

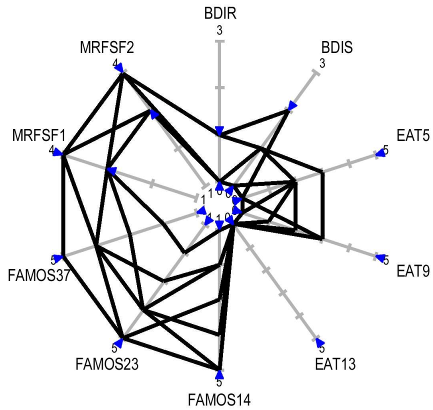
Figure 6.1: Integrated sliders: Stardinate visualizing data lines of 21 patients partly covering each other



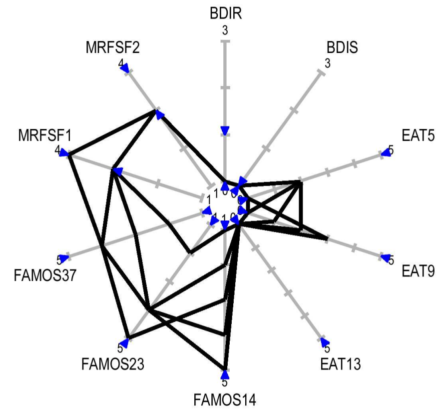
(a) Sliders attached to the axes restrict the values in BDIR from ’0’ to ’1’, in BDIS from ’0’ to ’2’, and in MRFSF1 and MRFSF2 from ’3’ to ’4’

(b) Sliders at BDIS were moved to value ’0’, so the selection is further reduced. Is there a correlation between the value ’4’ in MRFSF2 and higher BDIS values?

Figure 6.2: Selecting data lines by the use of sliders. Selected data lines are colored red, the rest is black.



(a) Values are limited in BDIR from ’0’ to ’1’, in BDIS from ’0’ to ’2’, and in MRFSF1 and MRFSF2 from ’3’ to ’4’. Same selection as in Figure 6.2 (a)



(b) Sliders on BDIS were moved to the value ’0’, so the selection is further reduced. Same selection as in Figure 6.2 (b)

Figure 6.3: Selecting data using sliders: unselected data lines are faded out. By moving the sliders the shape of the Stardinate changes.

The axes BDIR, BDIS, EAT5, and EAT9 on the right hand side form a shape similar to a wing of a butterfly. On the left hand side the axes MRFSF2, MRFSF1, FAMOS37, FAMOS23, and FAMOS14 show the other wing. EAT13 builds a borderline between these two wings. Although this Stardinate looks rather crowded we also recognize that there is a significant change between the axes MRFSF2 and BDIR. High values on the one axis are connected to lower values on the other axis and vice versa. The labels of these axes represent specific questions and indicators which are described in Table 6.1.

So, low values on the right hand side combined with high values on the left hand side indicate that the state of the patient is relatively good. The Stardinate in Figure 6.1 (a) shows data of 21 patients. Obviously, the user needs to interact with the graphical representation in order to get more insight because the number of data lines is rather high showing diverse values. Sliders enable the user to select data according to specific dimensions. Figure 6.1 (b) depicts our Stardinate described above with enabled sliders. Since there were no selections done the sliders are positioned on the ends of the axes.

Selecting by sliders

Figure 6.2 (a) shows highlighted data lines which results from using the sliders to select specific criteria accordingly. The upper arrow on BDIR was moved to value ’1’, the upper arrow on BDIS was moved to ’2’, the lower arrows on MRFSF1 and MRFSF2 were moved to ’3’. Data lines

Label	Question / Indicator	Answers
BDIR	’Loss of appetite’	(0) ’No’, (1) ’Rather No’, (2) ’Rather Yes’, (3) ’Yes’
BDIS	’Loss of weight’	analog
EAT5	’I avoid to eat when i am hungry’	(0) ’Never’, (1) ’Seldom’, (2) ’Sometimes’, (3) ’Often’, (4) ’Very Often’, (5) ’Always’
EAT9	’I look out for the calories of my food’	analog
EAT13	’I feel sick after eating’	analog
FAMOS14	’To relax is ...’	(1) ’Totally Unimportant’, (2) ’Unimportant’, (3) ’Does not matter’, (4) ’Rather Important’, (5) ’Extremely Important’ ... to me
FAMOS23	’To do something just for me is ...’	analog
FAMOS37	’To treat oneself to something is ...’	analog
MRFSF1	’I treat myself to tranquility and recreation’	(1) ’No’, (2) ’Rather No’, (3) ’Rather Yes’, (4) ’Yes’
MRFSF2	’To do something just for me is’	analog

Table 6.1: Psychotherapeutic data: 10 dimensions each one representing a question derived from questionnaires used in a clinical study.

meeting these criteria are colored red. Immediately, we see that the highlighted data lines are characterized by rather low values on the right hand side and rather high values on the left hand side. One possible conclusion is that good values in BDIR, BDIS, MRFSF1, and MRFSF2 imply rather good values for the rest of the dimensions.

Moving the sliders

By moving the sliders the user gets additional information. Figure 6.2 (b) visualizes the same data but the colored data lines are reduced because the slider on BDIS was moved to '0'. Surprisingly, there seems to be a correlation between the value '4' in MRFSF2 and higher values in BDIS. Patients who have '4' in MRFSF2 – which is very good – tends to have a value higher than '0' in BDIR. That means they lost weight, although they really care about themselves. Anorectic patients tend to orient themselves on socially expected behavior. So, one explanation could be, that MRFSF2 was answered with a very high value, although a more objective reflection would result in a low value. This might be a first hypothesis. So, there is an interesting discrepancy which needs further investigation.

Figure 6.3 depicts the same selection process as described above but instead of coloring the selected data lines, the rest of the data lines are faded out. Compared to Figure 6.2 the shapes of the Stardinates are more distinctive. Moves the sliders is immediately reflected in the shape of the Stardinates. This direct connection between the 'Gestalt' of the Stardinate and user's interaction bears great potential for the exploration processes. However, the user is not able to compare selected data lines with the rest of the data—the excluded data lines. According to her or his needs the user chooses the appropriate selection method.

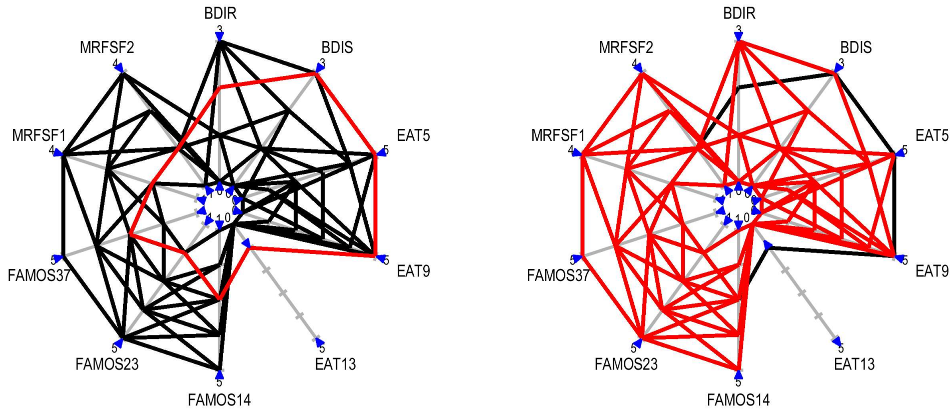
Inverted Selection

There is an easy way to reverse the selection by turning around the slider. Figure 6.4 (a) shows the usual form: the arrows are directed against each other. So, data lines between these two arrows are selected. In this case we see only one data line which has the value '1' in EAT13. Figure 6.4 (b) visualizes the reversed selection: the arrows are turned around. Thus, the data lines outside of the arrows are selected.

6.2 Cognitive Aspects of Artifacts

When conceptualizing a visualization technique the users and their tasks of operating and interacting with concrete data are the starting point. There is no common solution in InfoVis that fits all tasks, but some aspects of the tasks, of the data structure, and the user groups could be generalized. This allows us to find other domains with similar requirements.

Although we focus on a specific application domain, in particular, psychoanalytic data, our visualization technique is also applicable in other domains (cf. Chapter 10). If a visualization tool is designed by taking



(a) Positive Selection: sliders attached to axis EAT13 define the selection within the values ’1’ and ’5’

(b) Negative Selection: sliders attached to axis EAT13 define the selection outside the area between ’1’ and ’5’ which means ’0’ in this case.

Figure 6.4: Positive / negative selection by using sliders. The orientation of the arrow indicates whether the data next to it is selected or deselected.

Visualization facilitates cognition

the users, their tasks, and the data into account it can amplify cognition in several ways [15]:

- by increasing the memory and processing resources available to the users,
- by reducing the search for information,
- by using visual representations to enhance the detection of patterns,
- by enabling perceptual inference operations,
- by using perceptual attention mechanisms for monitoring, and
- by encoding information in a manipulable medium.

Constructivists point of view

One of the most important strengths of InfoVis is enabling the user to interact with the data under investigation. From a constructivists point of view cognition, particularly, learning and getting new insights is an active process in which new hypotheses and concepts are constructed which go beyond the information given [10; 11; 12].

Exploration, analysis, interpretation, or understanding of meaning of the data could be seen as a process of interaction on different levels of detail and should be supported by direct interaction. In this context direct interaction means that the user handles the graphical representation like a real object which offers inherent functionality. As a matter of course the view needs to change immediately after the manipulation, so that the user recognizes the change as a result of her or his manipulation.

Integrated functionality

Integrating all the needed functionality within an object confronts us with several challenges which are subject to sophisticated design considerations and go far beyond the scope of this thesis. Nevertheless, we start with some fundamental considerations of these issues.

Overloaded

Evidently, graphical representations need appropriate functionality but should not appear overloaded. Experienced users have other expectations than novices. An InfoVis tool should offer adequate functionality for both groups.

Functionality should not bother

On the one hand active areas (with integrated functionality) should not bother the user or distract from data, on the other hand operating should be intuitive and direct. It is necessary to integrate switches, sliders, labels, and other active areas. This causes changes of the graphical representation itself. So, we see interactivity as an integral part of the Stardinates.

Category	Manipulation
Instance	Grasp it Approach or dissociate Reject it
Location and view point	Look at it from different points of view Navigate or walk through the surroundings Place it in a new environment Rotate and flip it
Operating	Dis- or reassemble it Annotate something or add drawings Upgrade and enhance it Test it in different situations (How does it behave?) Compare or relate it to other objects Attend to inherent information (cf. Gibsons' theory in Section 3.1.3) Conclude from the present state on past of the object and make predictions

Table 6.2: Exploring and analyzing by manipulating objects.

Artifact-driven approach

As a first approach to interacting with the Stardinates we focus on manipulation of an artifact. Manipulation originates from the Latin word *manus* and refers to something done manually. We call our approach *artifact-driven* because things which primarily exist virtually are subject to objectification and become artifacts. It seems that such ascertainable data are easier to understand. Generally speaking, exploring and analyzing data correlates to exploring and analyzing an object which we are able to grasp. Keeping this background in mind we will take a closer look at the possible interactions with the Stardinates in the following.

6.3 Applying the BVI Framework

According to the BVI framework described in Section 4.1 interaction processes could be distinguished in respect to their input, output, and the states which are affected by the operations. This results in three types of interactions: (1) data operations, (2) set operations, and (3) graphical operations. Similarly to the idea of the visualization pipeline this scheme helps to abstract from the concrete interaction and to conceptualize the user interface. Based on this classification we sketch the following operations:

Data operations

- Selecting the database (source);
- Exporting the data of selected dimensions in a new database.

Set operations

- Adding or excluding dimensions and the corresponding axes;
- Retrieving detailed information about a specific dimension;
- Highlighting accumulations of data values;
- Using sliders attached to the axes in order to filter irrelevant data or to extract interesting data;
- Excluding or recalling certain Stardinates individually;
- Manually adjusting the arrangement of the Stardinates on the grid;
- Choosing up to two meta-dimensions which are used to sort the Stardinates on the grid along the x- and y-axis automatically;
- Saving or recalling interesting bundles of dimensions.

Graphical operations

- Arranging axes;
- Changing the orientation of an axis;
- Displaying or hiding scales, axes, sliders, and labels;

- Defining the size of the scales and their minimum and maximum values;
- Magnifying or zooming out;
- Displaying or removing the reference line;
- Configuring the thickness, color, scale, and labelling of the x- and y-axis;
- Altering the space available per Stardinate;
- Setting the bandwidth of color, thickness, and mode of the data lines;
- Setting the display of the scales, values, sliders, and labels;
- Setting the color, thickness, and mode of the axes and the reference line;
- Setting the color of the background.

Several interactions are missing here because they do not fit in this scheme. Moreover, from the users' point of view the boundary between graphical and set operations is rather arbitrary—for instance, a sophisticated user interface should support set operations by direct manipulation of graphical objects this in turn means actually doing graphical operations. Therefore, we add another perspective on interaction with the Stardinates by focussing on the operations with the visual representations called the *graphical object operation scheme*. Using different schemes provides us with more information about the interaction processes.

6.4 The Graphical Object Operation Scheme

Four types of operations

From a technical point of view we classify users' interactions with the visual representation by the object of interaction [38]. We think this is a good basis for both, the implementation and the concept of the user interface. Manipulation of one object could cause manipulations of the dependent objects. We distinguish four groups of interactions: (1) Manipulating the axes; (2) Manipulating the data line; (3) Manipulating one Stardinate; (4) Manipulating the grid (that means all Stardinates at once). In the following we describe these interactions in more detail:

Manipulating the axes each representing one selected data dimension. This includes for instance,

- Adding or excluding dimensions and the corresponding axes;
- Arranging axes;

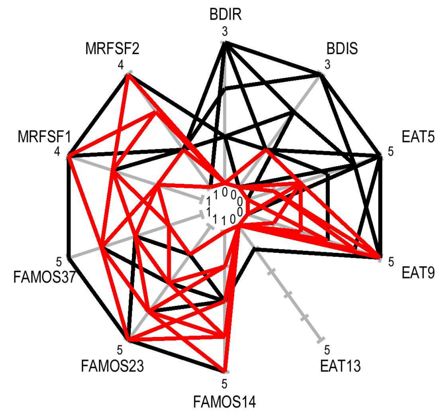


Figure 6.5: Selection according to the weight: BDIR is restricted to those values which exceed 40% of occurrences in the data. Here only one value, in particular, '0' is crossed by at least 40% of the data lines.

- Changing the orientation of an axis in order to handle negative polarity of dimensions;
- Displaying or hiding scales, axes, sliders, and labels;
- Defining the size of the scales and their minimum and maximum values;
- Retrieving detailed information about a specific dimension or concrete axis;
- Emphasizing accumulations of data values on a specific axis in respect to the weight of the data – common values are heavier – and highlighting those lines. Due to the fact that lines of equal values cover each other, a user cannot distinguish at the first glance, how often a specific value occurs. Accumulations of values are expressed by the corresponding weight of the line. Similarly to the parallel coordinates this needs to be supported by the visualization tool. In order to select the user states the minimum or maximum weight. Weight $w_x = \frac{o_x * 100}{n}$, o_x is the number of occurrences of the value x, n is the number of data records. All data lines which cross values of this weight are highlighted (compare Figure 6.5);
- Using sliders attached to the axes in order to filter irrelevant data or to extract interesting data by marking areas of interest. The orientation of the slider button indicates whether the area between two such buttons is excluded or selected (compare Figure 6.2, p. 45). One possible application would be

excluding common values in order to ease the search for abnormal patterns;

- Linking sliders on different axes: moving a slider on one axis causes corresponding changes on the linked axes. These axes could either belong to the same Stardinate or to different Stardinates. Linking sliders provides the user with information on the relation of one specific dimension to the other dimensions in different contexts.

Manipulating the data line which encodes the concrete values of all the selected dimensions of one data record showed in one Stardinate. Every manipulation of an axis results in alteration of data lines. Examples are:

- Highlighting or fading out one or more data lines manually. Highlighted data lines are depicted in front of the others in order to avoid occlusion by unselected data;
- Linking and brushing means selecting data lines in one Stardinate causes the highlighting of the corresponding data lines in the other Stardinates automatically;
- Comparing with the reference line. This could be done either manually or supported by the tool by highlighting correlating data lines;

Manipulating one Stardinate which depicts a bundle of data lines according to the dimensions which are combined within one Stardinate. One Stardinate represents a specific occurrence in respect to the meta-dimension(s) selected. So, if we choose the patient I.D. as a meta-dimension, each Stardinate represents one patient.

- Magnifying or zooming out;
- Scrolling through the data lines (e.g., the data lines of time-oriented data are highlighted one after the other in order to display changes over time);
- Excluding or recalling certain Stardinates individually;
- Displaying or removing the reference line;
- Setting the values of the reference line.

Manipulating the grid means to change general settings of the view or use tools which affect all Stardinates at once. Beside well-known drag and drop interactions which are associated naturally with their objects, several operations utilize the root of the grid in the top left corner.

- Manually adjusting the arrangement of the Stardinates on the grid;
- Choosing up to two meta-dimensions which are used to sort the Stardinates on the grid along the x- and y-axis automatically (cf. Figure 5.3, p. 40: On the y axis top down data of different patients are depicted by their IDs. On the x axis, left to right, we see specific categories of parameter bundles, e.g., one shows detailed information about 'depression', another shows the family situation, and so on.);
- Configuring the thickness, color, scale, and labelling of the x- and y-axis;
- Setting the size, position, and color of the data line label and the data line details;
- Altering the space available per Stardinate (formally defined in Equation (5.3));
- Setting the bandwidth of color, thickness, and mode of the data lines (defined in Equation (5.8)) used to distinguish different classes of records according to criteria of filtering or brushing processes or simply depicting a data line in its initial (unselected) state;
- Setting the display of the scales, values, sliders, and labels;
- Setting the color, thickness, and mode of the axes and the reference line;
- Setting the color of the background;
- Saving or recalling interesting bundles of dimensions;
- Comparing, saving or recalling specific visualizations (including concrete data) by using a so-called hypotheses tree (compare 9.1);
- Selecting the database (source);
- Exporting the data of selected dimensions in a new database to ease exploration;
- Accessing the history list of manipulations done in order to undo or redo specific steps.

This scheme based on four groups of interactions gives us a comprehensive description of possible interactions in the context of our visualization technique. Moreover, the interaction tasks identified above are not only relevant for the Stardinates but for all geometric glyphs. The graphical object operation scheme is a practical framework with the purpose of supporting the conceptualization of similar InfoVis techniques and their user interfaces.

Chapter 7

Evaluation of the Stardinates

Two different approaches

We utilize two different methods to evaluate the Stardinates. First we describe our user study conducted with 30 participants, which we characterize as concept testing. The tests were carried out by the use of a software tool we implemented for that purpose. Mainly we collected qualitative data which we evaluate by categorization. Based on our research questions we analyze our data by quantifying the qualitative data [57]. Second, we apply a metric, called ViCo [24], to measure the complexity of the Stardinates in comparison to parallel coordinates. These two complementing approaches provide us with detailed information on the limitations and strengths of the Stardinates.

7.1 Concept Testing

Concept testing differs form ordinary user testing insofar as it allows for focusing on the concept of the visualization itself. Thus, we reduced the need for extensive preparation of the participants and narrowed complexity of user interaction processes. Therefore, we restricted interaction to a bare minimum, namely highlighting data lines in order to view details, such as date or patient ID. Total elimination of interaction would have been counterproductive because the Stardinates are not conceptualized for presentation but for exploration and analysis. So, controlled restriction of interaction aspects seemed to be appropriate in order to cope with the characteristics of an explorative visualization technique on the one hand and receiving informative feedback from the subjects on the other hand.

Used two different visualizations

We started our study by designing two concrete visualizations – one using spatial data of aircrafts similar to Inselbergs' example [33], the other based on psychotherapeutic data. Thereafter, we formulated our research questions for the study and prepared the testing procedure. In order to consecutively improve the testing procedure and check its practical suitability

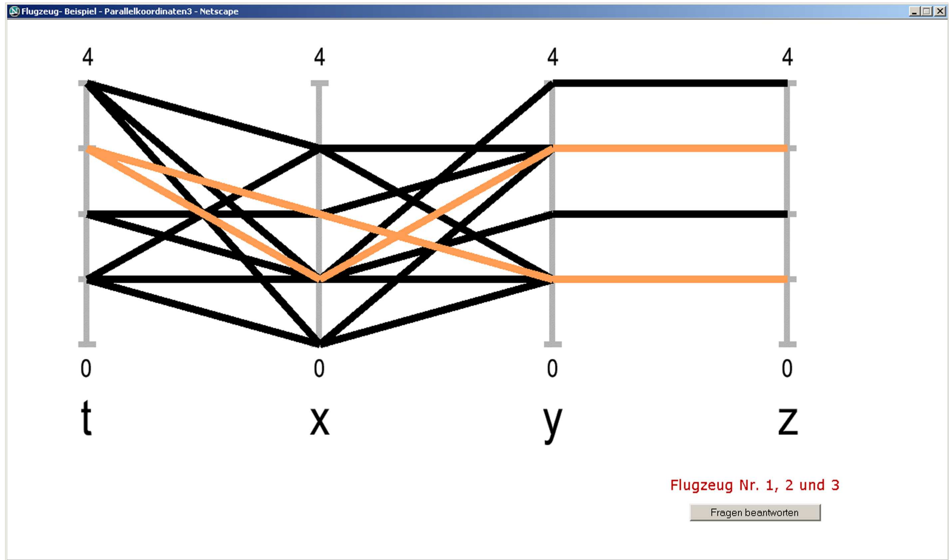


Figure 7.1: Pretesting: the parallel coordinates (first example) show spatial data of three aircrafts at four different times.

we adopted pretesting.

7.1.1 Pretesting

The goal of pretesting was to get first feedback on our visualization technique and the practicability of our testing procedure. Five people participated in these pretests.

7.1.1.1 Setting

Each of the five subjects individually tested two examples implemented as websites. Both depicted by two visualization techniques, namely, the Stardinates and parallel coordinates.

Example 1: spatial data

Figure 7.1 shows the first example using parallel coordinates. This visualization depicts spatial data of three aircrafts at four different times. Subjects were asked to decide whether a collision of aircrafts occurred. In Figure 7.3, p. 65 the same data is visualized by the Stardinates. The second example (compare Figure 7.2 and Figure 7.4, p. 66) uses (partly fictive) psychotherapeutic data, in particular, the states of five patients are depicted based on 10 parameters measured at three different times. More details about the questions of the study are described in Section 7.1.2.1.

Example 2: psychotherapeutic data

Participants viewed the first example (aircrafts) with both visualization techniques and thereafter the second example (patients' data) with both visualization techniques. These tests took between 40 and 60 minutes per

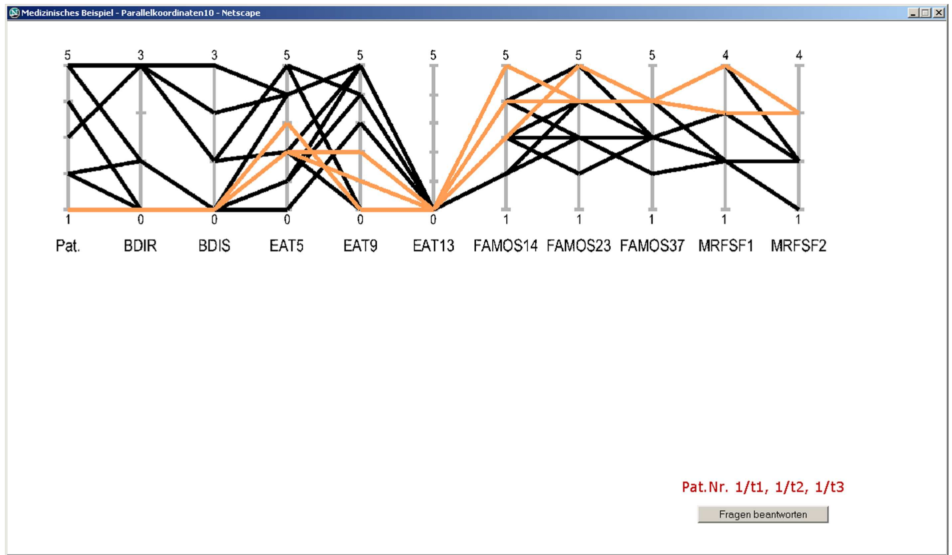


Figure 7.2: Pretesting: the parallel coordinates (second example) show psychotherapeutical data of five patients at three different times.

person, including reading information material, doing the investigations, and completing the questionnaires. Time measurement was done manually by an observer who also provided information and support to the subjects if needed.

The pretesting procedures slightly differed from one subject to the other, insofar as the information material given to the participants was revised and adapted in between in order to solve identified shortcomings. So, comparing the results of the five participants is not really possible. But each participant gave valuable feedback to us in order to improve the testing procedure.

The questions of the pretesting were nearly the same as the questions used during the testing procedures later on. The only difference in pretesting was that we compared two visualization techniques and asked about both of them. Various problems in exactly comparing two techniques based on their suitability for solving a certain task resulted (compare Section 7.1.1.3) in the decision to eliminate the parallel coordinates from the study and ask each subject just about one visualization technique. One possible solution might have been to test one group of subjects with the Stardinates and the other group with the parallel coordinates. However, one result of pretesting was that subjects clearly preferred the Stardinates (compare page 62) over the parallel coordinates. Limited resources and already available studies [1] about parallel coordinates caused focusing on the Stardinates. So, pretesting was done with both visualization techniques, whereas during testing we concentrated on the Stardinates only.

The questionnaires were filled out by the subjects manually in parallel to investigating the examples. Each participant decided individually which visualization technique she or he chose to start with. Two subjects started with the Stardinates, three with the parallel coordinates (cf. Table 7.1). The order was relevant in terms of interpreting the data because subjects mainly used the first visualization for examining and the second for verifying. Subjects were asked to complete the examination with one technique before starting with the other. However, one subject switched between both techniques several times.

7.1.1.2 Participants

Five participants

Four men (students of Computer Science) and one woman (computer engineer) participated in the pretesting procedure. The age ranged from 24 to 35. Since the goal of pretesting was to check the testing procedure, whether it is practical and informative, a small number of a rather homogenous sample seemed adequate. None of the subjects had previous knowledge about the Stardinates or our particular application domain (psychotherapeutic data). In our analysis of the anonymous data, we refer to subjects as PS1, PS2, PS3, PS4, and PS5.

7.1.1.3 Results and Conclusions

The most important conclusions were that we (1) identified and solved the last bugs of the program, (2) formulated a comprehensive information flyer, so that the procedure was clear to all participants of the testing later on, and (3) developed an appropriate questionnaire.

Developed software tool for testing

Our software tool was enhanced, so that the testing procedure was fully supported. We decided to integrate time measurement and questionnaires into the software tool in order to give the process a more rigid structure. Total flexibility in answering questions and examining the visualizations seems to bias the results, but a really strict order is proved to be unpractical as well. In the following we list the most significant results in more detail.

All subjects answered the question whether a collision of aircrafts occurred correctly. In the end it was not clear which visualization technique actually was used to solve this question. Two visualizations seemed to be more helpful to accomplish such a task because after viewing both visualizations all subjects were able to find the relevant information.

Subjects preferred Stardinates

The second example was mainly investigated by the use of the Stardinates. Asked about obvious information on the states of the patients, four of five subjects only reported about the Stardinates. The fifth used both visualization techniques. One reason could be that the complexity of the

Subject	Example	Technique	Duration
PS1	Aircrafts	Parallel coordinates	4 Min.
	Aircrafts	Stardinates	2 Min.
	Patients' Data	Parallel coordinates	1 Min.
	Patients' Data	Stardinates	11 Min.
			18 Min.
PS2	Aircrafts	Stardinates	2 Min.
	Aircrafts	Parallel coordinates	3 Min.
	Patients' Data	Stardinates	11 Min.
	Patients' Data	Parallel coordinates	8 Min.
			24 Min.
PS3	Aircrafts	Parallel coordinates	8 Min.
	Aircrafts	Stardinates	1 Min.
	Patients' Data	Parallel coordinates	5 Min.
	Patients' Data	Stardinates	4 Min.
			18 Min.
PS4	Aircrafts	Parallel coordinates	8 Min.
	Aircrafts	Stardinates	5 Min.
	Patients' Data	Parallel coordinates	12 Min.
	Patients' Data	Stardinates	6 Min.
			31 Min.
PS5	Aircrafts	Stardinates	1 Min.
	Aircrafts	Parallel coordinates	5 Min.
	Patients' Data	Stardinates	2 Min.
	Patients' Data	Parallel coordinates	2 Min.
	Patients' Data	Stardinates	1 Min.
	Patients' Data	Parallel coordinates	1 Min.
			12 Min.

Table 7.1: Pretesting: visualizations and their durations of usage.

parallel coordinates was much higher. Glyphs like the Stardinates decompose complex information into more discernable pieces.

Table 7.1 shows how long each participant was engaged in testing an example by using one visualization technique. Moreover, we see the order of the tested visualization techniques.

Subjects' feedback

We asked the participants on their first impression of both visualization techniques. In the following we list their feedback. It is translated from German.

PS1 about the parallel coordinates:

- very complex because there are too many lines in example 2
- example 1 is not understandable

PS1 about the Stardinates:

- similar data easy to recognize
- data depicted as areas
- within one Stardinate it is difficult to distinguish between different data records

PS2 about the parallel coordinates:

- overcrowded

PS2 about the Stardinates:

- is adequate if the number of axes is it least 5
- looks nice

PS3 about the parallel coordinates:

- too complex
- needs getting used to it

PS3 about the Stardinates:

- good overview
- easy to perceive by glancing at it

PS4 about the parallel coordinates:

- confusing

- selection of one data line is nearly impossible

PS4 about the Stardinates:

- separated items support overview
- first I read the label of each Stardinate then I look at the clew

PS5 about the parallel coordinates:

- looks overcrowded
- separated diagrams are better
- generally rather awkward visualization
- in the 2nd example specifically problematic: first low values are better, but on the other axes high values are better

PS5 about the Stardinates:

- clear, appealing
- ordering the parameters helps to perceive patterns
- easy to distinguish between good and bad values

Feedback from the subjects shows us a clear preference of the Stardinates. But if we actually want to check which visualization technique helped to accomplish the task, it is a problem if both visualization techniques are tested by the same person and applied to the same data. Subjects investigated the second visualization technique totally different since they already knew what to look for. So, their solutions of the tasks were mostly influenced by both visualizations. Therefore, this tests were not practical for examining whether one visualization technique was more efficient than the other. This and other important reasons already mentioned above caused focusing on the Stardinates only during the tests later on.

7.1.2 Testing

Based on the findings of the pretesting we finalized the software, the testing procedure, and the information material for the subjects. The **goal of the concept testing** was to investigate the following questions.

Research questions of the concept testing

- If the users are able to find information at the first glance.
- If the users utilize the features of the interactive Stardinates for their search.

- If the users are able to find the crucial information.

In the following we describe the setting and the participants of the study in more detail. Thereafter, we discuss the results and summarize some conclusions.

7.1.2.1 Setting

In order to test different aspects of the Stardinates we developed two examples. Both are integrated in websites using simple PHP forms which build the questionnaires. Combined with automatic time measurements the results are stored in a MySQL database. We observe the time each subject (1) views the first example, (2) answers the corresponding questions, (3) views the second example, and (4) again answers the related questions.

The questionnaire is in German. Before the test started each subject was provided with a detailed information flyer (compare Appendix A) explaining the procedure, the examples, and the questions. The tests took place in our InfoVis lab, a maximum of five participants tested the concept of our visualization technique synchronously. The average duration of the tests was 38:30 minutes.

Stages of the testing procedure

The testing procedure consists of the following stages:

First the starting screen of the concept testings asks for general information on the subject, such as profession, age, and gender. After clicking the start button the subject gets the first example and time measurement starts.

This example shows position data of airplanes based on the x-, y-, and z-coordinates at different times. The subjects are asked whether a collision of aircrafts has occurred[1]. To understand this example the subjects need common knowledge of geometry. This first example gives an overall idea of how a simple Stardinate could look like. Just three axes and three data lines are depicted, each of the four Stardinates shows spatial data of three aircrafts at four different times. Although this kind of spatial data are not best suited, since they are very simple and other visualization methods are more appropriate, we chose it because tests of other geometric techniques used the same approach. Particularly, similar data were used for tests of the parallel coordinates [33].

As already mentioned interaction was limited to selection purposes: by moving the mouse cursor over a data line this line is highlighted and the number of the aircraft is shown. In Figure 7.3 we see a screen shot of the first example, the data line of aircraft number 3 is highlighted.

[1] Since this data describes simulation data the airplanes can continue their flights after a collision. Moreover, this is necessary to avoid unintended recognition of collision because of a reduces number of airplanes afterwards.

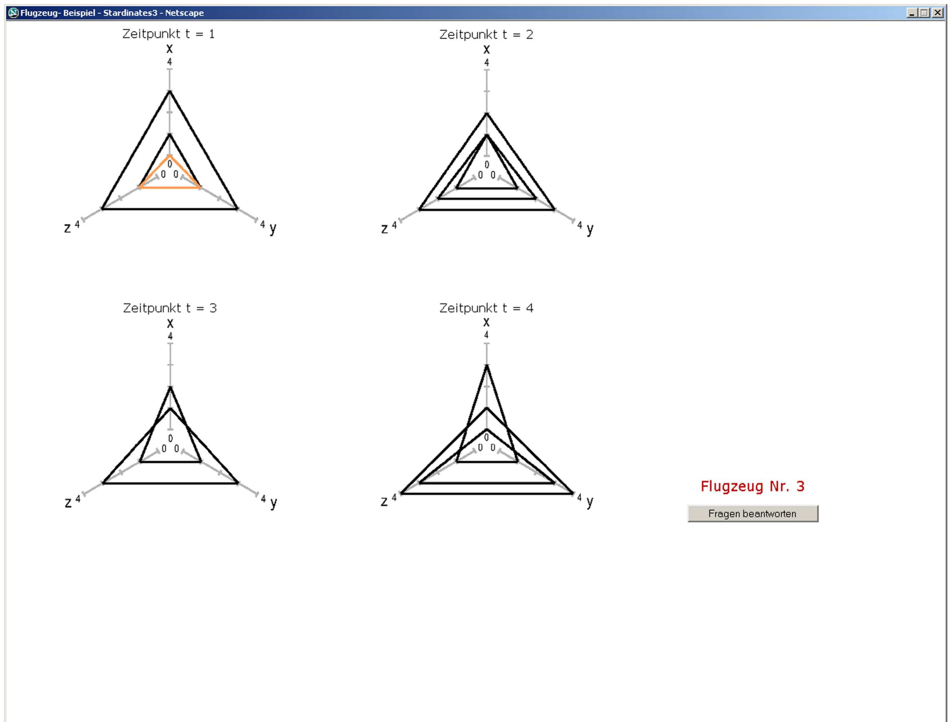


Figure 7.3: Concept testing: first example showing position data of three aircrafts at four different times visualized by the Stardinates.

By clicking the button 'Answer Questions' the subject opens an additional window containing the questionnaire. While answering the questions she or he can switch to the example again and examine it in more detail. It was not our goal to test whether the subject remembers the visualization. Therefore, we decided to allow for viewing the example and answering the questions synchronously. This affects time measurements and so their interpretation becomes harder. But it makes the tests more realistic.

First questions

The questions related to the first example are:

- Did a collision occur? If yes, which aircrafts were involved?
- Which visualization(s) helped you to examine the data? What information did you get there?
- Which problems / challenges occurred while interpreting the visualizations?

So, there is just one correct solution for the first question. The second and the third questions give additional information and help us to understand the answer to the first question.

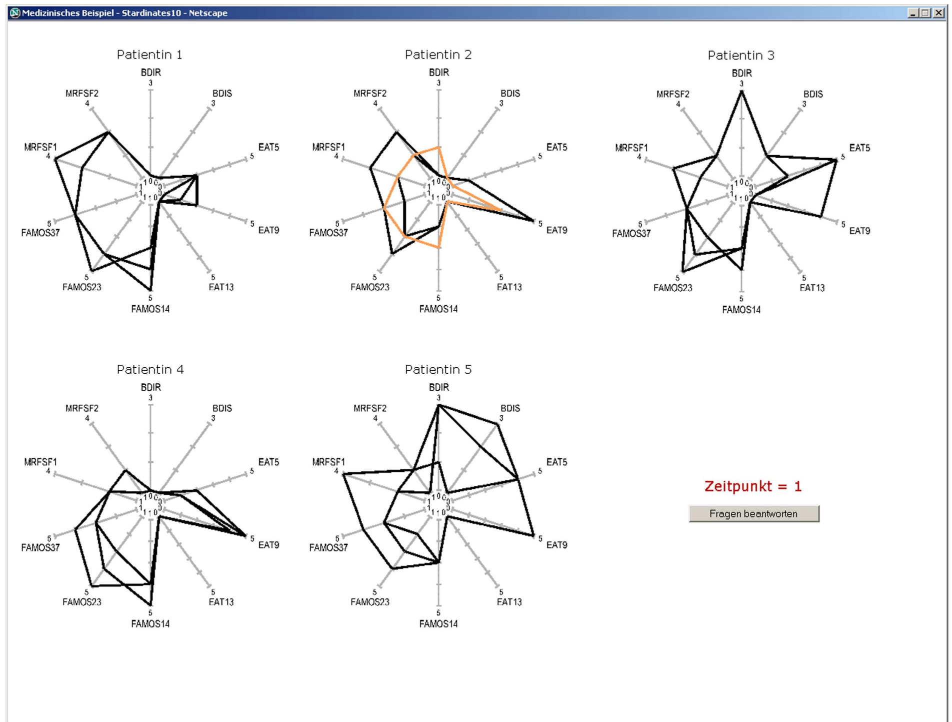


Figure 7.4: Concept testing: second example showing psychotherapeutical data of five patients at three different times visualized by the Stardinates.

The second example deals with totally different data. It visualizes (fictive) psychotherapeutic data of five patients at three different times (compare Figure 7.4). Each Stardinate depicts the data of one patient based on 10 parameters. The labels of the parameters are abbreviations the subjects were unfamiliar with. Moreover, all participants had no experience with this kind of data. Subjects were asked to interpret the data by searching for eye-catching similarities or varieties among the patients or significant changes over time.

Second questions

The questions related to the second example are:

- Are there any outstanding characteristics of the data?
- Which visualization(s) helped you to examine the data? What information did you get there?
- Which problems / challenges occurred while interpreting the visualizations?

These content-specific questions are followed by two more general items. In particular, the subject is asked to describe the first impression of the Stardinates and convey some feedback about the testing procedure.

Age	Number of Participants
Up to 20	3
21 - 25	7
26 - 30	8
31 - 35	7
36 - 40	1
41 - 45	3
Over 45	1
	30

Table 7.2: Participants' age distribution

7.1.2.2 Participants

30 participants

Following our intention we tried to build a representative sample by finding 30 participants with a high diversity in backgrounds, professions, ages, and different gender. Although the Stardinates are an InfoVis technique developed for a specific user group, we mean that concept testing should be done with a broader approach. This helps us to test the properties and features of the Stardinates more generally.

Table 7.2 shows the age distribution. The group consists of 17 female and 13 male participants. 15 among them study Computer Science or Business Informatics, the professions of the others range from nurses and kindergarten teacher to people holding a degree in drama, and from professors to pupils. In total 19 participants have a background in Computer Science or similar domains.

7.1.2.3 Results and Conclusions

Results of Example 1

73 % successful

The question if a collision occurred was answered correctly by 22 participants. So *73 percent* solved the first example successfully. 20 percent (six subjects) gave an incorrect answer and 7 percent (two subjects) could not find any solution. In the following we list some comments on question 2 of those subjects who answered the first question correctly describing which image provided them with the needed information to answer if a collision has occurred. Their statements show that most of the subjects responded to congruity of areas.

Subjects' statements

For the qualitative analysis we name the subjects from S1 to S30. Since we translated subjects' statements, they do not match literally with their answers, but we tried to illustrate the meaning of the statements.

S1: Glyph 3: Graphs of Aircrafts 1 and 2 are congruent.

S4: Times 1, 2, and 4: The triangles are not completely congruent. Time 3: Just 2 triangles are visible, so there must be a collision.

S9: Glyph number 3 shows identical x-, y-, and z-coordinates of two aircrafts. Since this glyph shows only two graphs, a collision happens for sure.

S10: At time 3 only two triangles are visible, that means two triangles are congruent.

S11: Aircraft 3 flies in one direction constantly, the altitude does not change. Aircraft 2 flies in contrary direction, at another altitude and different y-coordinates. Aircraft 1 moves in orthogonal direction to the other aircrafts doing a climb flight.

S26: In t=3 two triangles are identical, so they overlap in all three dimensions.

Subjects' comments on question 2 answering the first question incorrectly:

S6: Graphs of aircrafts 1 and 3 share one edge of the triangle, so they are on the same place at the same time -> collision.

S7 : Each glyph shows collisions.

S8: I cannot understand it.

S18: In time 1 there is a collision of the z- and y-coordinates of aircrafts 1 and 3. The second glyph shows a collision of the x-coordinate of aircrafts 1 and 3. The third glyph shows a collision because aircrafts 1 and 2 have identical coordinates. Time 4 shows no collision.

Challenges

Subjects faced several challenges in this first example. Many of them were unfamiliar with a visualization showing a point in 3D space as a triangle. They said it needs time to recognize the meaning of the triangle but once understood it is easy to work with. Some of them tried to translate x-, y-, and z-coordinates into longitude, latitude, and altitude. Since the educational backgrounds of the participants are very inhomogeneous, some had problems in basic issues of geometry. If a participant did not understand what the x-, y-, and z-coordinates represented she or he was not able to answer the question. All incorrect answers were motivated by misunderstandings concerning the meaning of the coordinates or how a collision is defined. So, these subjects simply were uncertain what to look for. However, most of the subjects solved the task successfully.

Two strategies

Those subjects who answered the first question correctly mainly applied the following two strategies: One group looked for congruent triangles without looking for the exact values of the coordinates. They focused on the shapes of the glyphs. The other group examined all axes sequentially and compared the values of the coordinates by imaging them as longitude, latitude, and altitude. They tried to relate the triangle with the point in 3D space. In a clear majority of cases participants adopted the first strategy utilizing features (compare Section 5.4) of the Stardinates to accomplish the task. This might be an answer to our second research question (compare Section 7.1.2).

Average usage time

At an average the subjects examined the first example for 3:11 minutes (minimum: 0:10, maximum: 13:13). Answering the first questionnaire (sometimes combined with further investigations of the Stardinates) took them another 8:53 minutes (minimum: 1:59, maximum: 18:32). Those subjects who answered the first question correctly needed on average 3:13 minutes for viewing the Stardinates and 8:44 for completing the form. The others examined the Stardinates for a average time of 3:05 and completed the form in 9:18 minutes. So there is no significant difference.

Results of Example 2

Subjects examined the visualization for an average time of 5:23 minutes. Answering the related questions took them another 21:04 minutes. So, participants nearly spent three fourths of the time with the second example.

However, the analysis of the results of the second example is quite different from the first one because there is not one correct solution of a problem such as whether a collision occurred. We are interested in the question if participants were able to find characteristic similarities or differences within the data and get information on the states of the patients. Although the domain, the questions, and terminology were uncommon and new to them, participants found a lot of interesting conclusions which we list beneath. But first we summarize the interpretations of the participants themselves:

63 % found information at the first glance

63 percent of the participants (19 subjects) said that they were able to get information at the first glance. So we can answer our first research question (compare Section 7.1.2) with yes. After becoming a little more familiar with the data and the visualization, 90 percent found new information about patients' states. The remaining 10 percent (3 subjects) either had problems to understand a geometric visualization (2 subjects) or found this kind of visualization with such

Findings of subjects

complicated labels too confusing (1 subject).

Typical findings are:

S1: None of the patients felt sick after eating.

S4: Patient 1: Quite good condition. Patient 2: Rather bad value in EAT9 but the patient seems to be well. Patient 3: BDIR is high all time, mean condition, partly good and partly bad values. Patient 4: EAT13 is stable, patient's state is also quite stable. Patient 5: Obvious variations, but values of EAT9, EAT5, FAMOS14, FAMOS13 are stable. Noticeable: EAT13 remains unchanged for all patients. Patients' states are quite different.

S6: Patient 4: Pronounced awareness concerning nutrition, controls calories. Tendency to restrain appetite. Patients would like to treat oneself to something without realizing these wishes. Patient 3: Loss of weight and appetite. Avoids to eat when hungry. Controls calories. Patient 5: Condition enhances clearly. Loss of weight and appetite decreases. Patients 1 and 2 are in good condition.

S7: Patients 3 and 5 serious states because BDIR, EAT5, and EAT9 show high values.

S8: Patients 2 and 4: A prong at EAT9.

S14: Patient 3 and patient 5 are in worse condition than the rest.

S23: Data of patients 2 and 4 indicates similarity. Patients 1, 2, and 4 quite stable, patients 3 and 5 significant variances.

These statements give a first impression how subjects examined the data and what their findings were. In order to evaluate the results more thoroughly we define **categories of statements**.

Categorization of Statements for Example 2

Quantifying by categorization

These categories of statements enable us to interpret which properties of the visualizations are primarily used by the subjects in order to find information about the patients. After seeing through the data collected during our study we developed these categories because they allow for a clear and comprehensive classification of our material. In particular, we check whether a participant made a statement associated with a certain category or not.

Comparing patients: such as, patient A is similar to patient B, both do not feel sick after eating.

Overview: for instance, Patient A seems to be in good condition.

Category	Number of Subjects	Percent of Subjects
1: Comparing Patients	19	63
2: Overview	19	63
3: Changes over Time	18	60
4: Examining Single Axes	16	53
5: General Conclusions	6	20
6: Finding Hypotheses	12	40
7: None	3	10

Table 7.3: Concept testing: categories of statements

Changes over time: as condition of patient A is rather stable.

Examining single axes: for instance, EAT9 is rather high indicating that ...

General conclusions: such as, patients with loss in weight and appetite tend to ignore their own well-being.

Finding hypotheses: for instance, patient A has a high BDIR value. In combination with a high value in MRFSF2 this patient would like to care more about herself but does not realize it because ...

None: this category refers to subjects who did not make any meaningful statements, because they were not able to understand the visualization.

These categories represent the way the visualization technique was used to search for information. In particular, we count whether a subject adopted a certain category or not. Categories 1 and 2 refer to information derived from the shape of the Stardinate. Whereas *category 2* deals with the shapes of single Stardinates, *category 1* covers comparisons of different shapes. *Category 3* indicates that interactive examination was important to the subject. *Category 4* contains statements about details which need intensive exploration of single axes. *Category 5* covers general conclusions the subject abstracted from the visualization. Finding hypotheses or reasons for specific data properties involving complex interpretation processes forms the *6th category*. Subjects who made none of the statements mentioned above and could not find any useful information are covered by *category 7*.

Table 7.3 shows the frequency scale of the various types of statements. We checked for each subject whether this kind of statement was made or not. For instance 18 subjects made statements associated to the third category ’changes over time’. The rest of the subjects, 12 participants did not adopt this category of statement. Table 7.4, p. 73 shows who made which type of statement.

It is not surprising that categories 1 (Comparing Patients) and 2 (Overview) have the highest rates followed by category 3 (Changes over Time) because the Stardinates support this kind of exploration very efficiently. Another indication that the participants utilized the features of the Stardinates which was one of our research questions (compare Section 7.1.2).

Key Statements for Example 2

Another approach is defining key statements from an expert's point of view. This is done in cooperation with a psychologist. In particular, we collected five patient-specific statements (one for each patient) and one statement about the whole group of patients, together representing the most significant information found within these data. These key statements outline the information an expert would conclude from these data using the same visualization technique. So we exploit these statements in order to evaluate whether the subjects were able to find the crucial insights.

Quantifying by key statements

Group: patients do not feel sick after eating.

Patient 1: good starting basis.

Patient 2: unstable.

Patient 3: contradicting answers.

Patient 4: positive progress in therapy. Cares more about herself.

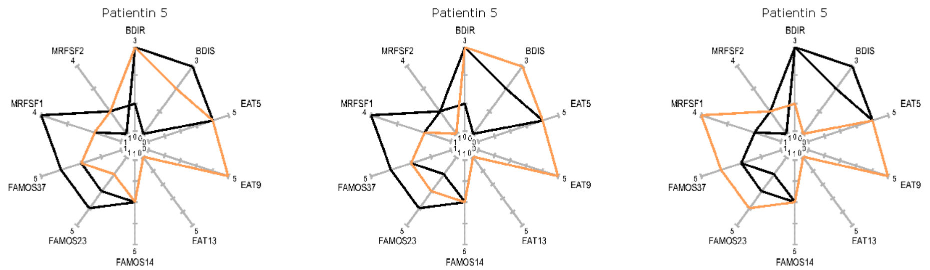
Patient 5: significantly positive progress in therapy between second and third time point.

Table 7.5 shows how many subjects found information analog to our key statements. Nearly two thirds of the subjects reported about the good starting basis of patient 1. Also noticeable is the high number of subjects (50%) who recognized the significant progress of patient 5, shortly followed by the group statement with 47% of the subjects. In Table 7.6, p. 75 we see who agreed to which key statement.

One reason might be that the Stardinates direct users' attention to differences and similarities. The visualizations of the data of patient 1 on the one hand and patient 5 on the other hand are obviously different. Participants were motivated to report this difference. Also the similarity of one parameter, in particular, EAT13, was informative to the participants.

Subject	Comparing	Overview	Time	Singe Axes	Conclusions	Hypotheses	None	Total
S1	•		•	•				3
S2		•				•		2
S3	•	•						2
S4	•	•	•	•				4
S5	•		•					2
S6		•	•			•		3
S7	•		•	•				3
S8	•	•		•		•		4
S9	•	•	•	•		•		5
S10	•	•	•	•				4
S11	•	•	•			•		4
S12	•	•	•	•		•		5
S13	•	•	•		•	•		5
S14	•	•	•	•		•		5
S15				•				1
S16	•	•	•		•	•		5
S17	•	•	•	•	•			5
S18				•		•		2
S19		•		•	•			3
S20		•			•			2
S21							•	1
S22		•	•					2
S23	•	•	•	•				4
S24	•		•	•				3
S25	•			•		•		3
S26	•	•			•	•		4
S27							•	1
S28			•					1
S29							•	1
S30	•	•	•	•				4
Total	19	19	18	16	6	12	3	

Table 7.4: Concept testing: categories of statements for each subject



(a) Time 1 (b) Time 2 (c) Time 3

Figure 7.5: Stardinates: time changes in patient data. The user highlights the data line of one time point by moving the mouse cursor over it.

Key Statement	Number of Subjects	Percent of Subjects
Patients do not feel sick after eating.	14	47
Pat. 1: good starting basis.	19	63
Pat. 2: unstable.	7	23
Pat. 3: contradicting answers.	4	13
Pat. 4: positive progress in therapy. Cares more about herself.	5	17
Pat. 5: significantly positive progress in therapy between second and third time point.	15	50

Table 7.5: Concept testing: key statements.

Both, patients 4 and 5 had a positive progress in therapy. However, only a few subjects reported the progress of patient 4. To recognize a therapy progress subjects needed to move the mouse cursor over the data lines in order to see which state was measured at which time (compare Figure 7.5). The Stardinate of patient 5 seemed to motivate further exploration of the data to a greater extent than the Stardinate of patient 4. Our interpretation is that the shape of the Stardinate of patient 5 was more eye-catching because the differences to the other Stardinates were obvious.

A relatively small number of subjects agreed with the key statements on patient 2 and 3. These two statements need much more knowledge about the data and require skilled interpretation. This result is not surprising since the participants had no experience in psychotherapy, moreover, most of them felt confused by the uncommon abbreviations (e.g., MRFSF, ...).

Subject	Group	Pat. 1	Pat. 2	Pat. 3	Pat. 4	Pat. 5	Total
S1	•						1
S2		•					1
S3		•					1
S4	•	•	•		•	•	5
S5		•				•	2
S6		•				•	2
S7		•			•	•	3
S8							0
S9	•	•				•	3
S10	•			•		•	3
S11	•	•	•		•	•	5
S12	•	•	•	•		•	5
S13	•					•	2
S14	•	•				•	3
S15		•	•				2
S16		•				•	2
S17	•	•				•	3
S18	•	•		•			3
S19	•	•	•				3
S20		•					1
S21							0
S22		•					1
S23	•	•			•	•	4
S24	•	•	•		•	•	5
S25	•						1
S26				•			1
S27							0
S28							0
S29							0
S30			•			•	2
Total	14	19	7	4	5	15	

Table 7.6: Concept testing: key statements of each subject

Conclusions

Although participants were not used to a visualization showing a 3-dimensional point as a triangle, a clear majority solved the fist task successfully by controlling whether the number of triangles corresponds to the number of aircrafts. If the number of triangles was lower, participants knew that a collision has happened. Mainly two features (cf. page 39) of the Stardinates were important in this context: the shape and the contour.

Overview and comparison

Subjects exploited the features of the Stardinates for getting an overview or comparing the states of the patients. Although the participants were unfamiliar with psychotherapeutic data, they were able to find crucial insights according to the key statements above.

Answers to our research questions

Finally, we can answer the three research questions defined in Section 7.1.2.

- The users were able to find information at the first glance.
- The users utilized the features of the interactive Stardinates for their search.
- The users were able to find the crucial information.

Shortcomings

Shortcomings reported indicate that the Stardinates are less appropriate for presenting data if the number of data lines is too high. If seen as static images they tend to look overcrowded and confusing. Most subjects said after a short time of getting used to this kind of visualization they felt comfortable but the first glance was confusing.

Some subjects stated that the use of color would improve usability. In addition, they argued that dimensions with negative polarization (for instance if a low value is better than a high value combined with other dimensions with contrary attribution) should be adjusted so that the dimensions are directly comparable.

Generally, participants asked for more interaction. We agree that the Stardinates need sophisticated interaction methods. Therefore, we took a close look at the special needs of such an interactive visualization method (compare Chapter 6) and created the interactive Stardinates.

User study indicate features

The following **strengths of the interactive Stardinates** have been identified during concept testing:

The interactive Stardinates ...

1. visualize overviews very effectively.
2. make differences or similarities within the data obvious. So, comparing different Stardinates is done intuitively.
3. are capable of visualizing data with high complexity.
4. motivate user interaction in order to examine data more thoroughly.
5. build distinctive shapes which support pattern recognition.
6. decompose complexity of data into more manageable pieces.
7. allow for visualizing data on different levels of details with consistent visualization metaphors utilizing cognitive features (compare 'Activation of Associated Concepts', 3.3).

7.2 Evaluation by ViCo

For the second type of evaluation we use a measurement tool. In particular, ViCo[2] [24] which is a metric to measure the complexity of visualizations by applying an algorithm with the following four steps:

Developing the complexity algorithm

1. Analyze the tasks to be accomplished by the use of a set of given visualizations and select those tasks to be taken as the basis of measurement.

2. Define minimal reading, writing, comparing, and calculating operations with respect to users' groups and variables of the data set to be visualized.

3. Develop an algorithm that accomplishes this task.

4. Develop the functions that describe the number of such operations needed to accomplish such a task.

The complexity of the algorithm gives us the complexity of the visualization. Moreover, we distinguish the complexity of different cases: best case, middle case(s), and worst case. For each one we can decide which visualization is more effective because less complex in the context of a

[2] Visualization and Complexity

specific application domain. We think that visualization should be discussed with respect to tasks and users. This metric takes such aspects into account. So the evaluation by ViCo is an approach to find out if the Stardinates are more appropriate for our specific application domain than other visualization techniques.

7.2.1 Defining Tasks and Operations

First we select the visualizations to compare. We decided to analyze the complexity of the Stardinates in relation to the parallel coordinates since they have a lot in common. Thereafter, we define the tasks, the required skills for the users, and the basic operations which then allows us to formulate an algorithm.

Generally, for both visualization techniques a lot of possible tasks exist. We chose two specific tasks representing two different groups of tasks which are situated in our application domain. One utilizes the shapes of visualizations, the other is based on examining single axes.

Reading, highlighting, comparing

The operations used are of the the following three types: Reading, Highlighting, Comparing. The tables beneath show the tasks, the user groups, the variables used in the algorithm, and the operations. In particular, Table 7.8, p. 80 characterizes the parallel coordinates. Operations needed to accomplish the tasks by utilizing parallel coordinates on the one hand or the Stardinates on the other hand are different. Table 7.7, p. 79 shows the definitions of tasks, users, and operations used for Stardinates' complexity measurement.

7.2.2 The Algorithms

In order to calculate the complexity of both visualization techniques we implemented algorithms for each task. The complexity of the algorithm complies with the complexity of the visualization. In Table 7.9, p. 81 we see both algorithms for the Stardinates, one for the first task which decides on the relation of two bundles of parameters, another one for the second task which determines similarities among the patients according to only one dimension. Table 7.10, p. 82 describes how the first task is accomplished by the parallel coordinates. The second task visualized by the parallel coordinates is described in Table 7.11, p. 83.

Kind	Name	Explanations
`Vis2`	The Stardinates	See Figure 7.4, p. 66
`T1`	Relate BDI with FAMOS	Decide the relation between loss in appetite and loss in weight with the ability of self gratification.
`T2`	Compare FAMOS14	Determine whether relaxation is important to the patient.
`Users`	Psychologists and Physicians	- knowledge about anorexia - familiar with the abbreviations and labels of the raw data - basic knowledge in geometry - able to highlight lines within the Stardinates by moving the mouse over it to see changes over time
`Var`	`#P`	- number of patients
	`#T`	- number of time series
	`#B`	- number axes of data bundles
`Scalar`	`a`	- a scalar identifies that we are only reading a subset of data series
	`b`	- a scalar identifies the complexity of a line shape in relation to an area shape
`Op1`	`Read Shape & Decide`	- read shape of the Stardinates and decide on their relationships
`Op2`	`Read Shape of Data Bundle & Decide`	- read shape of data bundles and decide on their relationships
`Op3`	`Read one Axis & Decide`	- read one dimension and decide on the relationships
`Op4`	`Highlight`	- select data line(s)
`Op5`	`Read Area Shape`	- read the shape of an area marked by the data lines
`Op6`	`Read Data Point`	- read the position of a data point
`Op7`	`Compare Area Shapes`	compare the shapes of areas
`Op8`	`Compare Data Point`	compare the positions of data points

Table 7.7: Complexity analysis of the Stardinates

Kind	Name	Explanations
Vis1	Parallel coordinates	See Figure 7.2, p. 59
T1	Relate BDI with FAMOS	Decide the relation between loss in appetite and loss in weight with the ability of self gratification.
T2	Compare FAMOS14	Determine whether relaxation is important to the patient.
Users	Psychologists and Physicians	- knowledge about anorexia - familiar with the abbreviations and labels of the raw data - basic knowledge in geometry - able to highlight lines within the parallel coordinates by moving the mouse over it
Var	#P	- number of patients
	#T	- number of time series
	#B	- number axes of data bundles
Scalar	a	- a scalar identifies that we are only reading a subset of data series
	b	- a scalar identifies the complexity of a line shape in relation to an area shape
Op1	Read Shape & Decide	- read and comprehend the overall shape and decide on the relationship
Op3	Read one Axis & Decide	- read one dimension and decide on the relationship
Op4	Highlight	- select specific data line(s) by mouse
Op5	Read Area Shape	- read the shape of an area
Op6	Read Data Point	- read the position of a data point
Op7	Compare Area Shapes	- compare shapes of areas
Op8	Compare Data Point	- compare positions of data points

Table 7.8: Complexity analysis of the parallel coordinates.

Task	Code

Task 1:
`The Algorithm`
	`Read Shape & Decide(Op1)`
	/*One may be able to recognize the relation if it is strong enough and the changes over time occur in a homogeneous way */
	`IF no clear Relation THEN`
	`  FOR MANY Data Bundles (a*#P)`
	`    Read Shape of Data Bundle & Decide(Op2)`
	`IF still no clear Relation THEN`
	`  FOR MANY Lines (a*#T)`
	`    Highlight(Op4)`
	`    Read Area Shape (Op5)`
	`    Compare Area (Op7)`
	`IF still no clear Relation THEN`
	`  FOR EACH Data Point (#P*#T*#B)`
	`     Read Data Point (Op6)`
	`     Compare Data Point (Op8)`

`Complexity for Task 1`	Best Case:
	`Op1`
	Middle Cases:
	`Op1 + a*#P*(Op2)`
	or: `Op1 + a*#P*(Op2)`
	`+ a*#T(Op4+Op5+Op7)`
	Worst Case:
	`Op1+ a*#P*(Op2) + a*#T(Op4+Op5+Op7)`
	`+#P*#T*#B (Op6+Op8)`

Task 2:
`The Algorithm`
	`Read one Axis & Decide(Op3)`
	`If not clear THEN`
	`  FOR MANY Data Points (a*#P)`
	`    Highlight (Op4)`
	`    Read Data Point (Op6)`
	`    Compare Data Point (Op8)`
	`  IF not clear THEN`
	`    FOR each Data Point (#P*#T)`
	`      Highlight (Op4)`
	`      Read Data Point (Op6)`
	`      Compare Data Point (Op8)`

`Complexity for Task 2`	Best Case: `Op3` Middle Case: `Op3 + a*#P*(Op4+Op6+Op8)` Worst Case: `Op3 + a*#P*(Op4+Op6+Op8)` `+ #P*#T(Op4+Op6+Op8)`

Table 7.9: Complexity analysis of the Stardinates: the algorithm

Task	Code

Task 1:
The Algorithm
	FOR each Patient (#P)
	  Highlight (Op4) /*Select every patient*/
	  /* Read Shape of one patient's data */
	  Read Shape & Decide(Op1)
	  IF no clear Relation THEN
	    FOR MANY Data Bundles (a*#P)
	      Highlight (Op4) /*Select every patient*/
	      Read Line (b*Op5)
	      Compare Line (b*Op7)
	  IF still no clear Relation THEN
	    FOR MANY Lines (a*#T*#P)
	      Highlight (Op4)
	      Read Line Shape (b*Op5)
	      Compare Line (b*Op7)
	  IF still no clear Relation THEN
	    FOR EACH Data Point (#P*#T*#B)
	       Highlight (Op4)
	       Read Data Point (Op6)
	       Compare Data Point (Op8)

Complexity for Task 1	Best Case:
	#P*(Op4+Op1)
	Middle Cases:
	#P*(Op4+Op1) + a*#P*(Op4+b*Op5+b*Op7)
	or: #P*(Op4+Op1) + a*#P*(Op4+b*Op5+b*Op7)
	    + a*#T*#P*(Op4+b*Op5+b*Op7)
	Worst Case:
	#P*(Op4+Op1) + a*#P*(Op4+b*Op5+b*Op7)
	             + a*#T*#P*(Op4+b*Op5+b*Op7)
	             + #P*#T*#B*(Op4+Op6+Op8)

Table 7.10: Complexity analysis of the parallel coordinates: the algorithm of task 1

Task	Code

Task 2:
The Algorithm
	/* If an accumulation of values exists */
	Read one Axis & Decide(Op3)
	If not clear THEN
	  FOR MANY Data Points (a*#P*#T)
	    Highlight (Op4)
	    Read Data Point (Op6)
	    Compare Data Point (Op8)
	  IF not clear THEN
	    FOR each Data Point (#P*#T)
	      Highlight (Op4)
	      Read Data Point (Op6)
	      Compare Data Point (Op8)

Complexity for Task 2	Best Case: Op3 Middle Case: Op3 + a*#P*#T*(Op4+Op6+Op8) Worst Case: Op3 + a*#P*#T*(Op4+Op6+Op8) + #P*#T*(Op4+Op6+Op8)

Table 7.11: Complexity analysis of the parallel coordinates: the algorithm of task 2

7.2.3 Results and Conclusions

The algorithms above give us the complexity for both visualization techniques. The best case for the first task – which is deciding on the relation of two bundles of parameters, in particular BDI and FAMOS, describing the relation between loss in appetite and loss in weight with the ability of self gratification – is `Op1` for the Stardinates and `#P*(Op4+Op1)` for parallel coordinates. These operations are described in Table 7.7, p. 79 for the Stardinates and in Table 7.8, p. 80 for the parallel coordinates. So, in the best case complexity of the parallel coordinates is much higher for the first task. The second task was to determine and compare one dimension (FAMOS14) for each patient. Here the best case is described by `Op3` for the Stardinates and by `Op3` for the parallel coordinates. The best case occurs for both visualizations if the values in FAMOS14 are accumulated at one point of the axis.

In the worst case the Stardinates have better results for both tasks. `Op1+ a*#P*(Op2) + a*#T*(Op4+Op5+Op7) + #P*#T*#B* (Op6+Op8)` gives us the complexity for the first task. `Op3 + a*#P*(Op4+Op6+Op8) + #P*#T*(Op4+Op6+Op8)` shows the complexity of the second task.

For parallel coordinates the results for the worst case are:

```
#P*(Op4+Op1) + a*#P*(Op4+b*Op5+b*Op7)
+ a*#T*#P*(Op4+b*Op5+b*Op7) +
```

`#P*#T*#B*(Op4+Op6+Op8)` for the first task, and `Op3 + a*#P*#T*(Op4+Op6+Op8) + #P*#T*(Op4+Op6+Op8)` for the second task.

The middle cases in complexity measurement also indicate the Stardinates as more efficient. One reason is that the Stardinates decompose the complex data into smaller pieces which could make perception more efficient because they are more 'graspable'.

Stardinates better than parallel coordinates

In summary, complexity measurement show us that the Stardinates are more effective than the parallel coordinates in the context of our application domain because they get predominantly better results. In the remaining cases they have at least equal results to the parallel coordinates.

Generally speaking, the Stardinates are proven to be more effective for the visualization of complex, multivariate data [3] (e.g., psychotherapeutic data) than the parallel coordinates.

[3] characterized by low frequency and rather small spans of values

Chapter 8

Multiple Views

Finding information within data is a complex cognitive process and cognitive psychology yielded several approaches to explain this challenging matter – the most relevant in our context are discussed in Chapter 3.

In our opinion, examining data by the use of InfoVis could be described as an insight problem [45]. An essential characteristic of such an insight problem is that the solution appears suddenly, without warning. In contrast, problems mainly solved without insight are worked out gradually rather than suddenly.

Insight by different views

So, insight by the use of InfoVis closely depends on the properties of the visualization. The Gestalt Theory (cf. Section 3.1.1) gives us some interesting hints regarding the impacts and effects of visualizations.

Biases cause misinterpretation

Each type of visualization emphasizes certain characteristics of the depicted data meaning it tends to produce a bias. Such biases could prevent the user from getting new insights or, moreover, cause misinterpretation.

Multiple view visualizations (cf. Section 2.4) offer an interesting potential to overcome such problems. They provide the user with two or more different views on the same data in order to support the investigation of a single conceptual entity. However, they need sophisticated coordination and interaction mechanisms and produce additional complexity for both, designers and users.

Therefore, the decision for a multiple view visualization is a cost/benefit trade-off between the advantages of such a multiple view system and the corresponding complexity that arises by applying it.

Four rules for multiple views

The following guidelines [5] should be considered:

Use multiple views if ...

Rule of Diversity:

... there is a diversity of attributes, models, user profiles, levels of abstraction, or genres.

Rule of Complementarity:

... different views bring out correlations and/or disparities.

Rule of Decomposition:

... the complexity of the data is too high for one visualization.

Rule of Parsimony:

... really necessary because a single view provides a user with a stable context for analysis, whereas multiple views bring about the effort of context switching.

When we started our project cooperation with the University of Vienna, Department of Pediatrics our task was to find an adequate visualization for psychotherapeutic data, particularly, data derived from a study on anorectic girls.

One important characteristic of this kind of data is its complexity because it consists of about 150 questionnaires summing up to about 6000 parameters per patient. But the range of values is rather low (for instance one question could be answered by the values 0–5 meaning grades from 'Yes' to 'No'). However, combining many different ranges needs sophisticated visualization metaphors because the interpretation of such data is complicated.

Our goal was to enable the psychologists and physicians to explore the data beyond doing simple statistical analysis. The results of testing single kinds of already existing visualizations did not satisfy their needs. The diversity and complexity of data and the demand to support interpretation by the users carefully but efficiently led us to the idea of combining three well-known InfoVis techniques. From this combination of different sights on the same data we expected significant benefits for the visualization of psychotherapeutic data. Based on the BVI framework [20] and taking the guidelines [5] for the design of multiple view systems into account we developed LinkVis which is described in the following.

8.1 LinkVis

Multiple views tool using three visualizations

LinkVis [40] is a multiple view tool based on three different techniques: Chernoff faces [17], scatterplots, and parallel coordinates [33]. These techniques are well known in InfoVis. Our tool offers the possibility to combine two of them. The user selects two methods for displaying the data set synchronously. Every technique offers advantages and disadvantages, moreover the data can be interpreted differently. If the user sees different aspects within the two visualizations, she or he will try to relate them and find analogies and contrasts. This offers an interesting potential for further

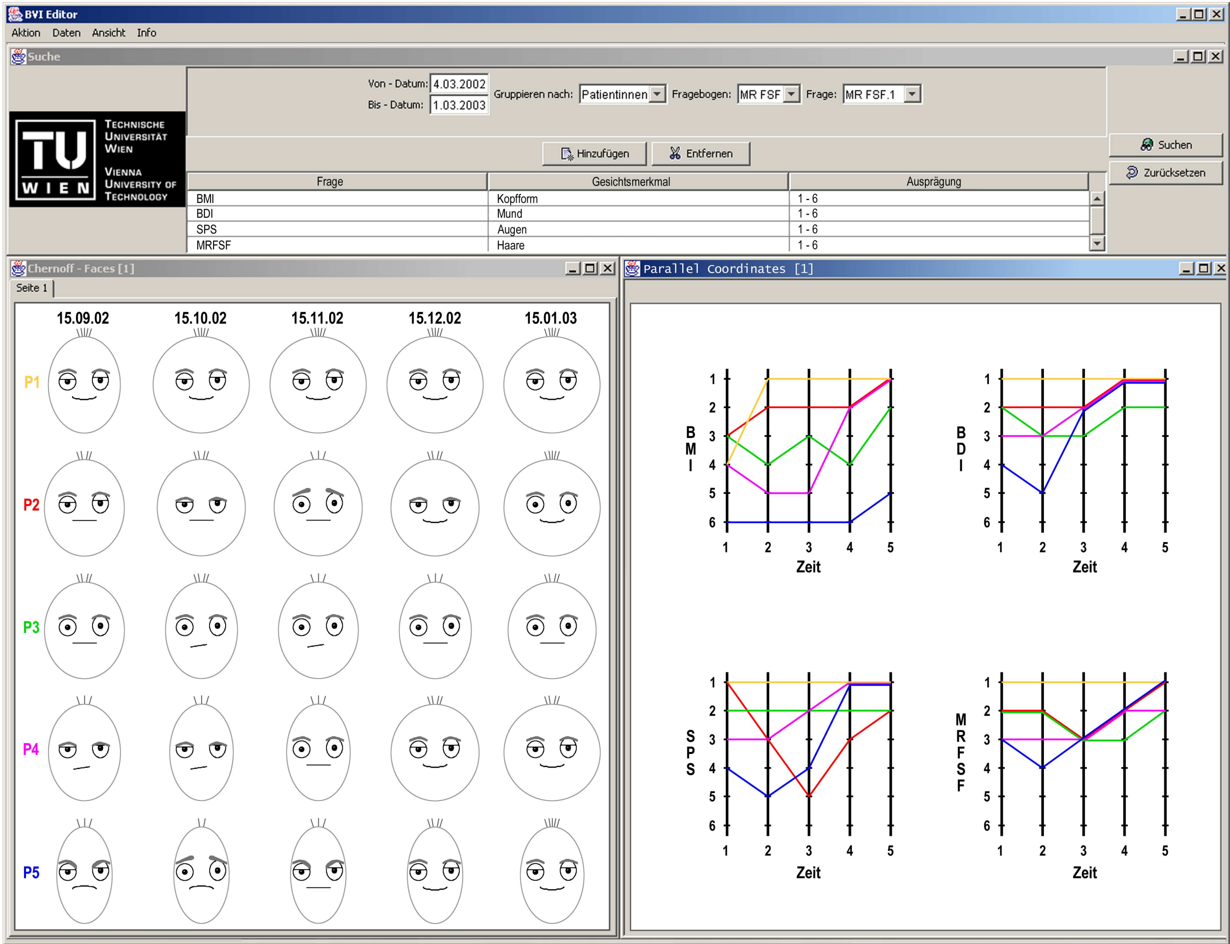


Figure 8.1: LinkVis, a multiple view tool is based on parallel coordinates, scatterplots and Chernoff faces.

exploration, especially to the advanced user. To give an impression of the benefits of our tool, we describe the techniques and their impact in detail followed by an example.

Chernoff faces

This face can represent up to 10 dimensions, e.g., by the size of eyes, slope of eyebrows, size and form of the mouth, and so forth. We tend to see the face as a whole, so this technique is very effective in communicating relative values and an overview. This kind of representation is based on the metaphor of emotions because the faces communicate different moods: visualizing specific data sets creates happy, sorrowful, or angry faces. Interpretation of these faces may cause certain problems. If the user wants to visualize the state of a patient, the facial expression of the Chernoff face could mislead the user by identifying every happy face with a good state and every unhappy face with a bad state of the patient. The facial expression basically depends on the combination of questions selected by the user and the allocation to facial attributes. The data, even the data set of one patient could result in very different facial expressions. However, the advanced user is able to get important information at a glance comparing these faces easily and perceiving relative values.

2D Scatterplots

This technique is well known and widely used. Two dimensions / questions are represented by the values of the x- and y-axis displayed as points or other symbols. In order to express temporal data we connected the points by directed edges according to their chronological order. The connecting lines emphasize the shape of the scatterplot.

Enhanced scatterplots

In contrast to the Chernoff faces, absolute values are easily recognized by looking at the values on the axes, although the user will detect clusters or data holes firstly. Scatterplots communicate relative and absolute values effectively, but the number of dimensions is limited. Since we reserve size, shape, and color of the symbol to provide the user with additional information, we need five scatterplots in order to display 10 dimensions by spatial arrangement. The user can configure grouping and size of the scatterplots, however, she or he tends to see five expressions separately. Users cannot get an impression of the whole. Therefore, the combination of Chernoff faces and Scatterplots will provide additional information because every technique presents the data differently.

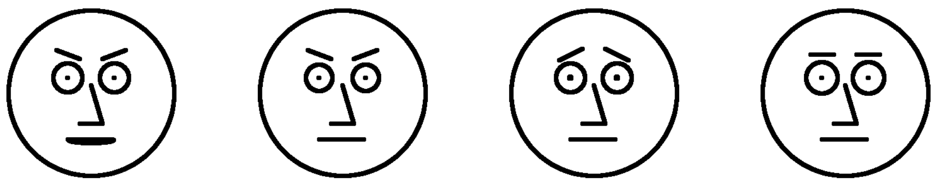
Parallel coordinates

In contrast to the 2D scatterplots, this technique places all axes vertically in parallel to each other. Thus we can display 10 dimensions on 10 axes. The data set is displayed by one line per record. We can rapidly record clusters of lines. Thus uncommon patterns will attract users attention. Arrangement of the axes is important. In order to analyze a record in detail, the line needs to be colored. Similar records will result in partly overlapping lines. It could be a problem that identical records are represented by a single line, and therefore, look like one record. Scatterplots cause the same problem. In contrast to Chernoff faces and scatterplots, parallel coordinates offer two important features: 10 dimensions can be combined in one graph. Similarly to the Chernoff faces, this method provides the user with overview information, but is not charged with the possibly confusing metaphor of emotions, especially if applied to psychotherapeutic data sets. In addition, the user can easily recognize absolute values.

Arrangement of axes is crucial

Dimension	Facial Attr.	Scatterplot	Par. coord.	P. A	P. B	P. C	P. D
1	Pupil	x/1	1	1	1	2	1
2	Eyebrows	y/1	2	2	2	4	3
3	Mouth width	x/2	3	2	3	3	3
4	Eye size	y/2	4	3	2	3	4

Table 8.1: Data set of four patients based on four questions / dimensions. Question 1: Do you often feel angry about your parents? Question 2: Do you feel strong and self-confident? Question 3: Do you like the way you look? Question 4: Do you talk to your colleagues about private problems?

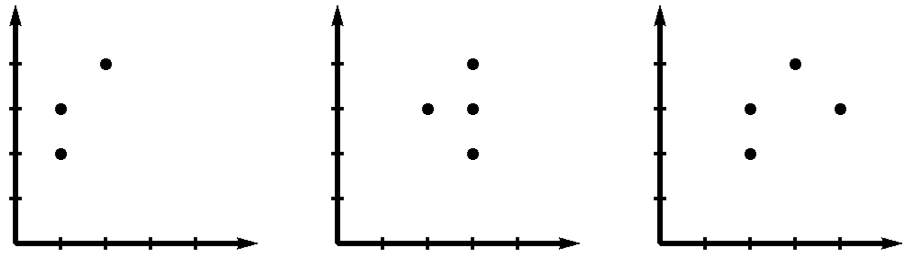


(a) Pat.A (b) Pat.B (c) Pat.C (d) Pat.D

Figure 8.2: Chernoff faces visualize the dimensions and values of Table 8.1, each describes a patient.

Finally, we want to show an example based on four dimensions / questions in order to ease interpretation. It works the same way with 10 dimensions, but the user needs more experience to understand the representations. The unused attributes of the Chernoff face are shown, but do not represent any data. Table 8.1 shows a data set which represents four patients. The range of the values is ’1’–’4’ and means ’Never’, ’Sometimes’, ’Mostly’, and ’Every time’.

Every face of Figure 8.2 depicts the data set of a patient. The same data set is represented by scatterplots in Figure 8.3(a) and 8.3(b) and by the parallel coordinates in Figure 8.4. The dimensions are modified in Figure 8.3(c) in order to show the combination of ’Question 2’ and ’Question 4’. If the user selects Chernoff faces and parallel coordinates, she or he can compare the lines to the faces. As mentioned before she needs to highlight a line in order to view the details. The faces communicate an overview at the first glance, but the user could have problems to recognize that the eye size varies from ’Sometimes’ to ’Every time’. In contrast this information is shown clearly on the second axis of the parallel coordinates. On the other hand, the user sees easily the differences of the eyebrows in the Chernoff faces. The changes of the display caused by interactions of the user have an important impact to the exploration of data. The data set



(a) x: Dimension 1 (Do you often feel angry about your parents?) / y: Dimension 2 (Do you feel strong and self-confident?)

(b) x: Dimension 3 (Do you like the way you look?) / y: Dimension 4 (Do you talk to your colleagues about private problems?)

(c) x: Dimension 2 (Do you feel strong and self-confident?) / y: Dimension 4 (Do you talk to your colleagues about private problems?)

Figure 8.3: Scatterplots visualize the dimensions and values of Table 8.1

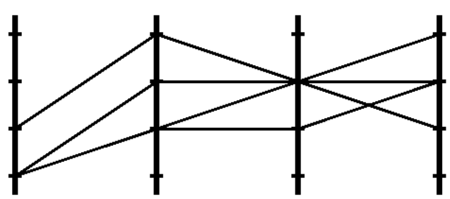


Figure 8.4: Parallel coordinates visualize the dimensions and values of Table 8.1, axes 1 (left) to 4 (right): Do you often feel angry about your parents? Do you feel strong and self-confident? Do you like the way you look? Do you talk to your colleagues about private problems?

becomes more meaningful if the user examines the changes interactively. Our example demonstrates that flexibility and an interactive design of the user interface is very important for visualization tools.

Shortcomings of LinkVis

However, practical experience with LinkVis showed that interpretation of Chernoff faces becomes impossible if the number of dimensions is too high. Although the users found the faces appealing and exciting, they had problems in exploring complex relations. Using reference faces in order to utilize the perception of relative values is one idea to benefit from the features of this visualization technique. But we were skeptic about the bias which could emerge due to emotional connotations of faces.

But combining a number of five to 20 dimensions within one item representing a content-based bundle of parameters appeared to be a promising approach. Decomposing complexity and building meaningful combinations turned out to be an appropriate method to deal with this kind of data. So we began to think about another visualization technique which allows for both, detailed and exact exploration on the one hand and adequate handling of complex data on the other hand. However, the LinkVis idea was still promising, so our plan was to integrate this new visualization technique into such a multiple view system. This was the beginning of the *interactive Stardinates* an InfoVis technique applied in a multiple view system.

8.2 LinkStar: The Stardinates applied in LinkVis

A visualization technique such as the Stardinates can be seen as a module with definitions of input and output interfaces. In order to ease integration in a multiple view tool standardized interfaces are used. In particular, for data input we utilize Extensible Markup Language (XML). For graphical output on screen Open Graphic Library (Open GL) is our choice. The Stardinates visualization module is developed in Java.

The development of the multiple view tool with the integration of the Stardinates, called LinkStar, is an ongoing process. However, the mock-up showing LinkStar (compare Figure 8.5) gives a first impression. In the following we explain the visualization we see here in more detail.

Linking Stardinates with other visualizations

In this particular case, LinkStar visualizes (fictive) psychotherapeutic data based on a clinical study on anorectic girls. Nevertheless, LinkStar is appropriate for all application domains with similar data and user tasks. One example for another application area would be students' feedback on classes. Generally, data derived from questionnaires with the need for enhanced exploration and analysis processes could be visualized by LinkStar.

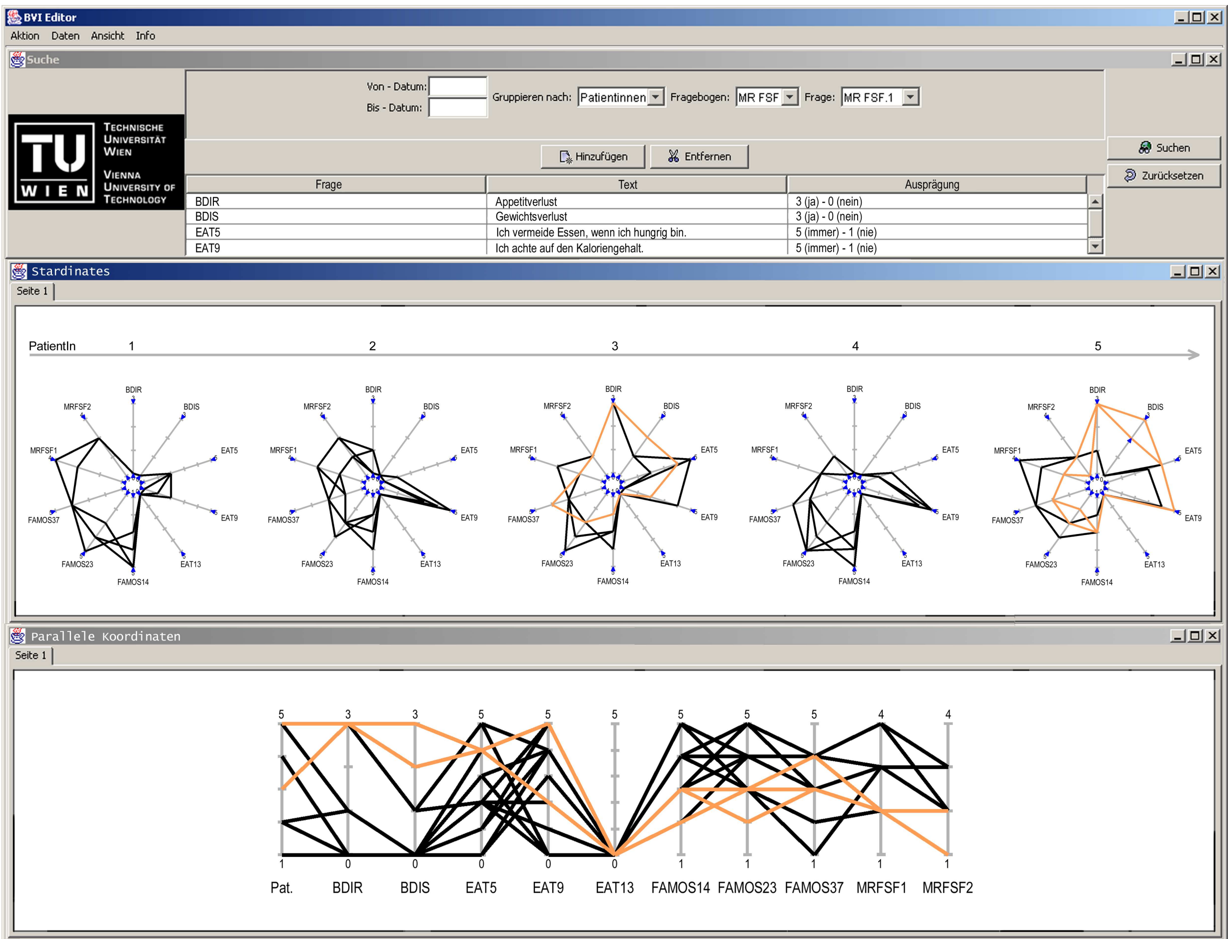


Figure 8.5: Mock-up of LinkStar, a multiple view tool showing parallel coordinates and the Stardinates.

Here two visualization techniques showing data from anorectic girls are depicted synchronously. The interactive Stardinates are shown in the upper output windows. Beneath we see the parallel coordinates. Above these two output windows we see some information about the data currently selected. The data is grouped according to patient ID, so that each Stardinate depicts the data of one patient. Parallel coordinates show the patient ID on the first axis.

In top of the Stardinates we see the axis showing the patient IDs, one for each Stardinate. Due to highlighting some data lines are colored orange, others are black. But before talking about highlighting and selecting we need to explain the data dimensions.

Each Stardinate consists of four bundles of dimensions. It also would be possible to separate the bundles so that one Stardinate only contains one type of dimensions or in other word shows only one bundle. This would result in a matrix of Stardinates: a second axis would be added as shown in Figure 5.3, p. 40. A more simple form is depicted in Figure 10.1, p. 102. However, the Stardinates are composed depending on users' needs. In this particular case, we see four bundles combined within one Stardinate. This emphasizes the relation between the bundles if the number of dimensions of each bundle is rather small.

Four data bundles

- The first bundle refers to the Body Mass Index and is described by BDIR ('loss in appetite'), and BDIS ('loss in weight'). Values range from 0–4: (0) 'no', (1) 'rather no', (3) 'rather yes', (4) 'yes'.

- The second bundle consists of parameters describing the eating behavior: EAT5 ('I avoid to eat when I am hungry'), EAT9 ('I look out for the calories of my food'), and EAT13 ('I feel sick after eating'). Values range from 0–5: (0) 'never', (1) 'seldom', (2) 'sometimes', (3) 'often', (4) 'very often', (5) 'always'.

- Third we see parameters describing self-gratification abilities of the patients. In particular, the dimensions are FAMOS14 ('To relax is...'), FAMOS23 ('To do something just for me is'...), and FAMOS37 ('To treat oneself to something is...'). The values range from 1–5: (1) 'totally unimportant', (2) 'unimportant', (3) 'does not matter', (4) 'rather important', (5) 'extremely important'.

- The last bundle shows if patients care about themselves represented by MRFSF1 ('I treat myself to tranquillity and recreation') and MRFSF2 ('I pay close attention to my body'). Here values range from 1–4: (1) 'no', (2) 'rather no', (3) 'rather yes', (4) 'yes'.

Each dimension is depicted as axis. Whereas the axes of the Stardinates are arranged in a circle, those shown in the parallel coordinates are placed in parallel to each other.

Shape indicates patient's state

For the interpretation of both visualizations it is useful to know that a good state of a patient results in low values in the first and the second bundle. In contrary, the third and the fourth bundle should have high values if indicating a good state of the patient. So a Stardinate is characterized by a bulge on the left hand side if the patient is in good state.

If the user switches the axes (compare Section 6.4) of the first and the second bundle in order to adjust polarity we would search for rather large polygons referring to a good state. However, here the axes are depicted in the original form as derived from the raw data.

Interpreting the data

Each Stardinate visualizes data of four times. So, in total we see 20 data lines partly covering each other. If we would use Chernoff faces for visualization this would result in 20 faces. The Stardinates do not need so much place but the exact changes over time are only accessible by mouse. However, the user gets a first impression if the state of the patient is stable or changing over time. The first one results in an image with a few lines like the Stardinate of patient 1, whereas the second is represented by crowded lines, like depicted in the Stardinates of patient 3 and patient 5. Moreover, the user gets overall information on patients' states at the first glance represented by the shape of the Stardinate. Patient 1 obviously is in good state, whereas patients 3 and 5 show characteristics of a critical state.

As already mentioned above some data lines are selected here. The sliders of the Stardinates are activated shown as small blue arrows. After activation sliders are positioned at the the ends of the axes meaning no specific selection was done. The arrows direct against each other which implies that the data line in between the arrows would be selected.

One possible task is analyzing dimensions with outstanding values in relation to other dimensions. In this case the user could select such data lines. The Stardinate of patient 5 obviously shows high values for BDIS indicating a critical state of the patient. If the user is interested in those patients characterized by high values in BDIS she or he can move the bottom slider attached to the BDIS axis of this Stardinate for instance to value '3'. This results in coloring all associated data lines. The color for highlighting data lines was set by the user previously.

Here two data lines are colored for patient 5 – meaning a high value was measured two times – and one line for patient 3. The other patients have better BDIS values, so they are not marked. Since we use linking and brushing, the same data lines are selected in all Stardinates and other visualizations synchronously. So, data lines are colored in the parallel coordinates immediately. In the following we list some concrete advantages of each visualization technique complementing each other.

Advantages of parallel coordinates

Parallel coordinates:

- Exploring a single dimension for all patients could be done easily. Especially, if the data lines accumulate at one point, here for instance in EAT13 all patients have the same value.
- Comparing one patient to the others: since all data lines are depicted in the same visualization the context information is directly available. In Figure 8.5, p. 92 we see that the highlighted lines show high values not only for BDIS, but also for BDIR and EAT5. We perceive the shape formed by the highlighted lines intuitively.

Advantages of the Stardinates

The Stardinates:

- First we recognize which patients belong to the group characterized by high values, namely, patients 3 and 5.
- We also perceive that the states of both patients are changing much more than the states of the other patients.
- It is obvious that patient 5 had high values in BDIS at two times.

In summary, the Stardinates focus on the single patients and show the overall state of each patient, whereas the parallel coordinates are sufficient for comparing all patients based on a single parameter. On the one hand the Stardinates decompose the data into smaller, meaningful pieces. Such data 'clews' seem more graspable. On the other hand parallel coordinates show the complete context helpful if examining dimensions individually.

A significant advantage of such multiple view systems is that they allow for using one view for examining the data and the other view for verifying the findings. This reduces misinterpretation due to biased visualizations.

Detailed exploration of the data shown in these visualizations requires user interaction. In addition to the considerations discussed in Chapter 6 interaction in multiple view tools brings specific challenges about.

8.3 Interacting with Multiple View Visualizations

Our graphical object operation scheme (cf. Section 6.4) is a good starting point and also helpful in this context because we can simply expand it by adding new graphical objects and get a comprehensive classification of interactions for multiple view tools. Moreover, some visualizations share components and fit into the exiting scheme, for instance all geometric techniques use axes.

But multiple view visualizations bear specific requirements concerning interaction mechanisms.

Coupling views

One challenge is coupling the visualizations [59], especially important if shown synchronously to the user. The following two common approaches [5] of coupling visualizations exist:

Slaving

Navigational slaving means that movements in one view are automatically propagated in the other views. This becomes relevant if switching between several levels of details.

Linking

This very common approach connects the data items of one view with the data items of the other views. A specific type of linking is *linking and brushing* in which the user selects and highlights items in one view and the corresponding items are highlighted automatically. This helps the user examining relations.

The design of interaction processes need to take increased complexity due to context switching into account. General design guidelines help us to come up with a clear user interface. In the following we list four design rules [5] which are useful in this context.

Four rules for user interfaces

- Rule of Space/ Time Resource Optimization:

  Balance the spatial and temporal costs of implementing multiple views with the spatial and temporal benefits of using the views. That means on the one hand computation time and display space needed for showing several views cause additional costs. On the other hand time is saved by using side-by-side views if the user wants to compare different views. Their relation should be balanced in order to get adequate benefits in comparison to the costs.

- Rule of Self-Evidence:

  Use perceptual cues to make relationships, similarities, and differences among multiple views obvious to the user.

- Rule of Consistency:

  Make the user interfaces for multiple views consistent. Moreover, make the state of the multiple views consistent. Inconsistency of views can cause misinterpretations.

- Rule of Attention Management:

  Use perceptual techniques to guide user's attention on the right view at the right time. Users can only focus on one visualization at one time. If both visualizations change at the same time, she or he can only attend to one change at a time.

These design guidelines build up a sound basis for the development of our multiple view tool, called LinkStar. Our intention was to outline our approach realizing LinkStar and to describe the background for the integration of the interactive Stardinates. The implementation of LinkStar is an ongoing process.

First results of LinkVis indicate that examining data by the use of multiple view tools is a promising approach but needs support for finding and verifying hypotheses in order to do it in a more structured and efficient way. The Hyperbolic Hypotheses Tree is one possible approach to deal with this matter.

Chapter 9

Exploration and Analysis in Multiple View Systems

In multiple view systems the need for context switching increases the risk of 'getting lost' in information space. Structuring the exploration and analysis tasks by creating a tree in which the nodes represent the gained insights helps the user to overlook the process of finding information. The nodes of the tree stand for the structure of the knowledge and enable the user to directly access the related visualizations. So the tree provides the user with an archive of meaningfully connected visualizations. Repeated adaptation of the tree supports the process of building knowledge.

In order to efficiently use the space needed for displaying the tree we apply a distortion technique explained beneath. We call our tool the Hyperbolic Hypotheses Tree [39]. In the following we describe its basic concepts and the structure of the tree.

9.1 The Hyperbolic Hypotheses Tree

Getting lost in information space

Generally, exploration without definition of goals is an unpredictable process in terms of results and time needed. Sometimes important insights happen by accident, this process is called serendipit [64]. However, the user could become frustrated if she is not able to get an overview of the content. In order to avoid the feeling of getting lost in data space our tree structure supports in getting an overview and defining exploration goals. Each node of such a tree represents an idea or, moreover, a hypothesis. Usually a tree is strictly hierarchical. It could be a problem to represent knowledge or data, which is semi-structured or not structured at all by a tree. Exploration processes are characterized by finding connections within semi-structured or unstructured data. Therefore, we need to adapt the structure of the tree. In order to get more flexibility we distinguish

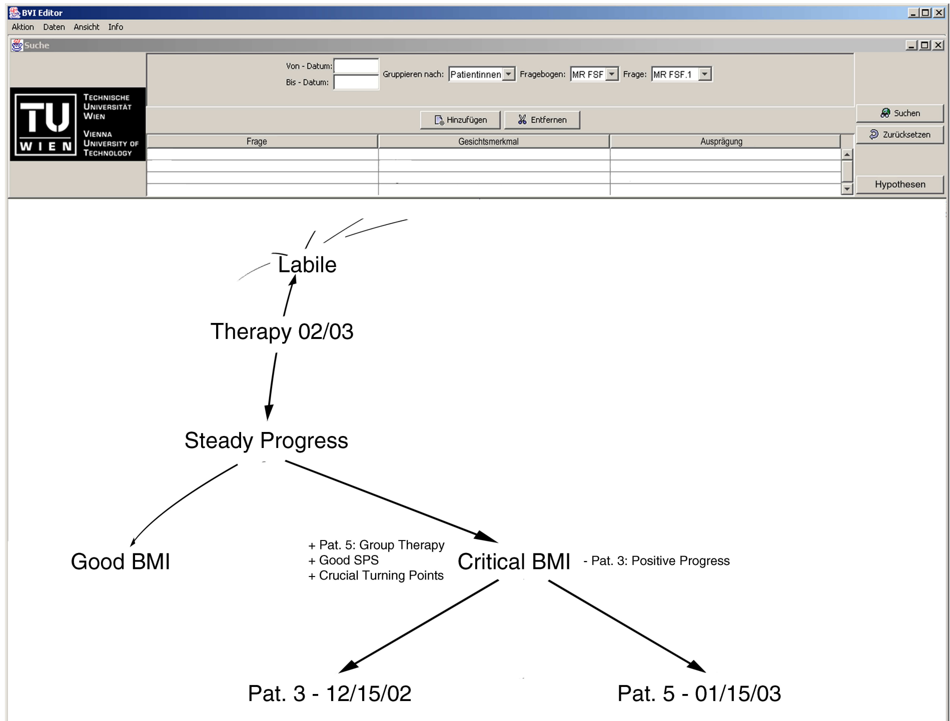


Figure 9.1: The hyperbolic hypotheses tree in LinkVis

different types of nodes we describe beneath. Every node is specified by a keyword and linked to certain visualizations the user generated before. In addition, the user can append notes describing her considerations about this visualizations or a hypothesis. The user constructs the tree in parallel to exploring the data. The idea of the Hyperbolic Hypotheses Tree is to represent the knowledge of the user and support extending and navigating it. These trees are usually complex. In order to handle such complex trees and use space efficiently, we decided to apply a focus and context technique, in particular, the Hyperbolic Tree [37]. The Hyperbolic Tree is a distortion-oriented technique, which looks like a tree with the nodes being appliqued on the surface of a globe. If the user focuses on a certain node, this node is placed in the center and gets more space than the peripheral nodes. Therefore, the user can see the details but does not lose the context at the same time.

9.2 Structure of the Tree

First of all every tree has a root node. We distinguish two types of nodes: hierarchical nodes and nodes of equal status. Hierarchical nodes, like parent or child nodes, and siblings describe clear hierarchies. Nodes of equal

Figure 9.2: The hyperbolic hypotheses tree is linked to visualizations: clicking on a specific node opens the associated visualization.

Defining the nodes

status are: subsidary nodes, which add information to another node (+); Restricting nodes, which limit the information of another node (-); Negation nodes, which are the antithesis of another node (!). A node of equal status is always linked to another node. If a node of equal status is not linked to any hierarchical node it is linked automatically to the root in order to give them higher priority and visibility. Furthermore, the brightness of a node indicates the relevance and correctness of its information.

Figures 9.1 and 9.2 gives an impression of the Hyperbolic Hypotheses Tree. Interaction with the tree, like navigating and extending it is crucial in order to understand the structure and build up a mental map of the information represented. Interaction with visualizations plays an important role in identifying meaningful patterns and active learning [14] and searching for information. Three types [66] of interaction can be distinguished: manipulation ('handling it'), navigation ('walking on it'), and issuing commands to it ('conversing with').

Chapter 10

Applications

In the following we mention three different application domains representative for others using the same types of data. First we sketch the visualization of questionnaires and ethnographic user studies, thereafter we describe our main application domain: Visualizing psychotherapeutic data.

10.1 Visualizing Questionnaires

Qualitative and quantitative data

The Stardinates are designed to visualize complex data – concerning numbers of dimensions and relationships within the data – but with a rather small range of ordinal or nominal values. Typically all kinds of questionnaires applied in extensive costumer studies or other kinds of surveys, e.g., asking for attitudes which also contain qualitative properties deliver such data. If common statistical methods do not suffice for evaluation the interactive Stardinates would offer adequate exploration techniques for finding relations and distinctive features within the data.

10.2 Visualizing Ethnographic Data

Another application domain would be visualizing data derived from ethnographic user studies which include many different types of data, combining quantitative and qualitative data. Especially, when quantifying qualitative data by categorization the Stardinates are appropriate for supporting the examination tasks. Typically, evaluation of this kind of data needs enhanced methods of aggregating and abstracting.

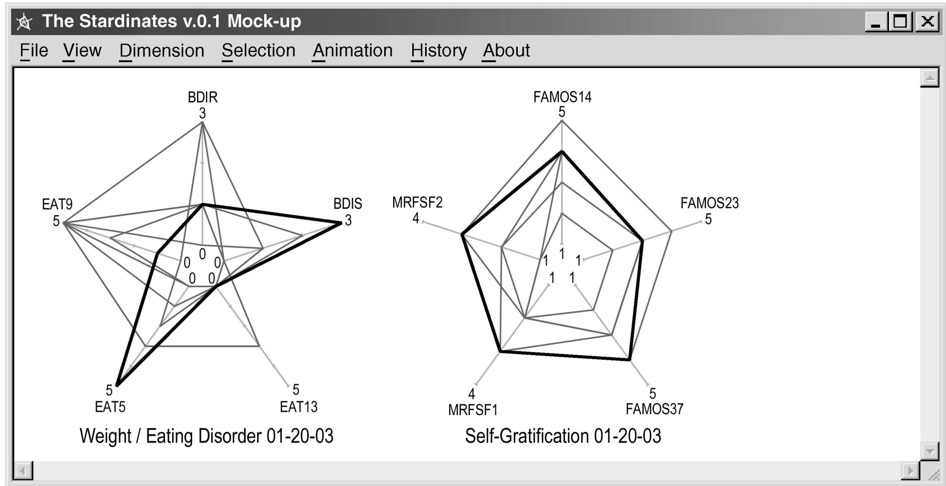


Figure 10.1: Mock-up of the Stardinates tool.

10.3 Visualizing Psychotherapeutic Data

Clinical study on anorectic girls

Meaningful bundles

We use the Stardinates to visualize psychotherapeutic data derived from a clinical study on anorectic girls. In close cooperation with two psychologists we are developing a visualization tool which aims to support the evaluation and analysis of the therapeutic progress with focus on the differences of group and individual therapy. Figure 10.1 shows a mock-up of this tool. The clinical study works with questionnaires—about 150 per patient. Each questionnaire consists of about 40 questions. Since some questionnaires are used several times, we reduce the number of dimensions respectively. This results in about 1000 dimensions per patient and is described as complex data. The questions have unique abbreviations which can be used as labels of the axes. The huge amount of questions makes up a number of dimensions to be visualized. We have defined meaningful bundles of questions which show the state of the patients according to significant characteristics of the ailment and the therapeutic progress, e.g., depression, loss of weight, self-confidence, social context, eating disorder, etc. So far 10 patients have started with individual therapy and three patients with group therapy.

The data derived from these questionnaires mainly ranges from the value ’0’ to ’6’ which is a scale from ’No’ to ’Yes’ and is classified as ordinal data. The granularity of the data varies. Some of the questions have negative polarity, therefore, their axes are flipped in order to fit in one Stardinate together with questions with positive polarity. Analysis and exploration of these data needs targeted interaction which implies knowledge on the underlying questionnaires and their meanings. The Stardinates support the user by intuitive distinguishable shapes, however, in order to really understand the data users need experience in psychology.

During the next two years the therapies of 60 patients will be evaluated. For quantitative analysis applicable techniques based on statistic methods already exist. Since the psychologists plan to pay detailed attention to qualitative exploration and analysis they need new techniques. Visualization offers excellent opportunities for efficient perception and recognition of relevant information because it makes use of our special visual abilities, e.g., pre-attentiveness. In this case visualization has crucial advantages over text or tables.

In the following we describe a mock-up (Figure 10.1) of the Stardinates tool which depicts two Stardinates. Each visualizes the state of seven patients according to five dimensions / questions. The right Stardinate deals with self-gratification abilities of the patients. The other one depicts data of weight / eating disorders. Both show data from Jan. 20th. One record line—patient id. 7 is marked in both Stardinates. This is well-known as linking and brushing. Surely, the Stardinates are not intended as static images. Interacting with them is absolutely necessary in order to build hypotheses and get more insight into the data under investigation. However, this mock-up gives an impression of our technique. Each axis has its own scale and minimum and maximum values drawn from the questionnaires. To normalize the parameter space the user can change this scales adequately.

The Stardinate on 'Self-Gratification' is based on the following questions:

- FAMOS14: 'To relax is': (1) totally unimportant (2) unimportant, (3) does not matter (4) rather important to (5) extremely important to me.
- FAMOS23: 'To do something just for me is': analog.
- FAMOS37: 'To treat oneself to something is': analog.
- MRFSF1: 'I treat myself to tranquillity and recreation': (1) no, (2) rather no, (3) rather yes, (4) yes.
- MRFSF2: 'I pay close attention to my body': analog.

The axes starting in twelve o' clock represents these questions.

The Stardinate on 'Weight / Eating Disorder' is based on the following questions:

- BDIR: 'loss in appetite': (0) no, (1) rather no, (3) rather yes, (4) yes.
- BDIS: 'loss in weight': analog.

- EAT13: 'I feel sick after eating': (0) never, (1) seldom, (2) sometimes, (3) often, (4) very often, (5) always.
- EAT5: 'I avoid to eat when i am hungry': analog.
- EAT9: 'I look out for the calories of my food': analog.

Interpretation of the data

The data result in easily perceivable shapes which need to be interpreted. Since we use five axes here, the data record of each patient is depicted as a pentagon. The Stardinate on self-gratification shows that patient id. 7 has high values, compared to other patients the highest values at FAMOS23, MRFSF1, and MRFSF2. The values of the other axes are also in the upper area. This means that this patients has – especially compared to other patients of this study – in principle the attempt and possibly the ability of self gratification because she rates them with high priority. However, the interpretation of the Stardinate about weight and eating disorders is more complicated. The patients rates BDIS and EAT5 very high, the other axes show the lowest value of '0'. In contrast to other patients, patient id. 7 has low values at EAT9 and BDIR. That means that calories are not important to her and she did not feel the loss of appetite but lost weight. This corresponds with her answer about avoiding food although she feels hungry. These conclusions are a first step into the analysis of

Conclusions

the state of this patient. So far we do not have more data, but it would be very interesting to show the changes over time by other Stardinates positioned beneath. Moreover, other meaningful bundles of dimensions offer additional opportunities for informed interpretations.

Chapter 11

Summary and Future Work

This chapter is divided into two parts. The first represents a summary of our work and the second part outlines some future plans.

First results of applying the interactive Stardinates for visualizing complex data, namely psychotherapeutic data, in order to explore the underlying structure and find new insights seem promising. They support the requirements preliminarily defined in Chapter 1 and offer the needed functionality for interactive investigation by the users.

Findings of the evaluation discussed in Chapter 7 give some support that the interactive Stardinates might be a valuable visualization technique.

We integrated the Stardinates into a multiple view tool because first results of our multiple view tool – LinkVis – and tests of the Stardinates indicate that the idea of providing the user with different views on the same data seemed appropriate. Especially, in combination with utilizing a geometric glyph our approach proved very efficient for the visualization of highly structured and multi-dimensional data. Furthermore, offering solutions to our application domain, psychotherapy, the interactive Stardinates could also benefit a number of other areas characterized by similar data types.

However, some open issues remain: in the near future we want to continue the implementation of LinkStar in order to thoroughly test the interactive Stardinates in such a multiple view system. With the goal of producing a comprehensive and practical tool for visualizing complex data.

Therefore, further refinement of the interaction mechanisms in LinkStar is necessary. Our evaluation of LinkVis indicated that a multiple view visualization supports *examine/verify strategies* may need further investigation.

The development process is evaluated by ethnographic user studies, starting in parallel to the implementation, by involving the users at an early stage. In addition, some open questions of the Stardinates are not solved so far, for instance how to deal with incomplete data. A possible approach is to add an extra point on the scale for missing values. However, this does not seem to be the best solution.

Another unresolved issue is that data lines with equal values cover each other. Thus, the user can not recognize, if a line represents one or more data records. Depicting the lines slightly displaced results in a rather overcrowded visualization. Additional properties of our visualization technique could help to encode accumulations by using the weight of the line, indicating the number of covered lines.

Chapter 12

Conclusions

Applied Information Visualization which yields useful tools must be based on the practical needs of the users. This requires the developers and designers to stay in touch with the users and their goals. That was one aspect of our work I really appreciated. We spent several days and nights discussing our issues with the psychologists and physicians. In the process of this we developed fantastic ideas and visions but at the same time we made sure to prove their practicability in terms of solutions for the users' tasks. In the end we implemented some of the more promising visions: the interactive Stardinates applied in a multiple view system.

Our user study indicated that the interactive Stardinates might offer distinct advantages.

In particular, the interactive Stardinates ...

1. visualize overviews very effectively.
2. make differences or similarities within the data obvious. So, comparing such Stardinates is done intuitively.
3. are capable of visualizing data with high complexity.
4. motivate user interaction in order to examine data more thoroughly.
5. build distinctive shapes which support pattern recognition.
6. decompose complexity of data into more manageable pieces.
7. allow for visualizing data on different levels of details with consistent visualization metaphors utilizing cognitive features.

In accordance to our premises on the relation of users' tasks, data, interaction, and visualization (compare Section 1.1) we introduce the interactive Stardinates capable of explorative visualization of complex data.

Chapter 13

Acknowledgements

There are some people I would like to thank. First my supervisors Silvia Miksch and Margit Pohl who asked me a lot of crucial questions and inspired me to take my work to new and higher levels. They were always very motivated to discuss my research and eager to help me overcome problems. In particular, **Silvia Miksch** kept pointing out new challenges which motivated me to do more and thorough research and always had an open ear for my worries. **Margit Pohl** confronted me with a lot of interesting psychological background. She further spend much time discussing my work with me and giving valuable feedback. The continuous support of my supervisors made writing my thesis an exciting and satisfying experience to me.

Furthermore, I would like to thank **Susanne Ohmann** and **Christian Popow** for the interesting project cooperation yielding a lot of useful raw material for my research. They contributed freely with their endless knowledge in our application domain for which I am grateful.

Thanks to my colleagues at the **Institute of Software Technology and Interactive Systems** for their support and letting me use their resources.

The students **Herbert Herzog** and **Rami El-Berry** gave valuable inputs by coding the 'BVI-Editor', a multiple view tool, according to our specifications. Furthermore, I would like to thank **Bernhard Rengs** who helped me with preparations of the user study. **Heike Adler**, **Gerti Lanzenberger**, **Gabi** and **Barbara Hayda** amongst others deserve thanks for participating in our user study and providing us with crucial feedback.

Special thanks to **Cliff Kapatais** for proofreading and endless patience with my complicated sentence structure.

I would like to thank my friends and family for their patience and support. Especially, I am grateful to **Markus Rester** who helped me not only by valuable LaTeX 2$_\varepsilon$ knowledge, support with the graphics, and enduring interest in the discussions I elicited but also for motivating and supporting me throughout all the mood swings when writing a thesis.

Parts of this work have been carried out within the Asgaard Project which is supported by *Fonds zur Förderung der wissenschaftlichen Forschung* (Austrian Science Fund), grant P12797-INF.

Appendix A

Stardinates Concept Testing 2003 Info-Blatt

Im Rahmen dieses Tests wird eine Visualisierungsmethode getestet. Nährere Informationen zu dieser Visualisierungsmethode findest du weiter hinten. Bei diesem Test geht es darum, möglichst viel Information aus den Graphiken abzulesen. Grundsätzlich sind solche Graphiken eher dafür geeignet, Überblicksinformation statt Details abzulesen. Während des Tests kannst du jederzeit Fragen stellen, falls etwas unklar ist. Der gesamte Test dauert etwa 15 - 20 Minuten. Im Laufe des Tests sind Fragen am Computer zu beantworten. Zur Information vorab sind die Fragen auch am Ende dieses Info-Blattes abgedruckt.

Ablauf des Concept Testing:

1. Info-Blatt durchlesen.
2. Die Fragen auf der Webseite zu Alter, Gechlecht und Beruf/Studienrichtung beantworten und danach mit dem Button "Los gehts" zum 1. Beispiel weitergehen.
3. Schau dir das erste Beispiel (Flugzeug-Beispiel) im Detail an und versuche möglichst viel Information abzulesen. Wenn Du mit der Analyse fertig bist, gehst du mit "Fragen beantworten" weiter.
4. Es wird ein neues Browserfenster geöffnet, in dem die Fragen 1, 2 und 3 beantwortet werden sollen. Während der Beantwortung bleibt die Webseite mit dem Flugzeugbeispiel im Hintergrund geöffnet. Durch einfaches Klicken auf das jeweilige Fenster kannst du zwischen den beiden Webseiten hin und her wechseln, falls du für deine Antworten noch etwas nachschauen möchtest. Wenn du alle Fragen beanwortet hast, gehst du mit dem Button "Antworten abspeichern und weiter zum nächsten Teil" weiter.
5. Dann wird im Hauptfenster das zweite Beispiel (Medizinisches Beispiel) geöffnet. Schau dir auch dieses Beispiel im Detail an und versuche möglichst viel Information abzulesen. Wenn Du mit der Analyse fertig bist, gehst du mit "Fragen beantworten" weiter.
6. Es wird wieder ein neues Browserfenster mit den Fragen 4 - 8 geöffnet. Bitte beantworte die Fragen und schliesse den Test dann mit dem Button "Antworten abspeichern und beenden!" ab.
7. Zum Schluß bitte das Browserfenster schliessen.

Die Stardinates

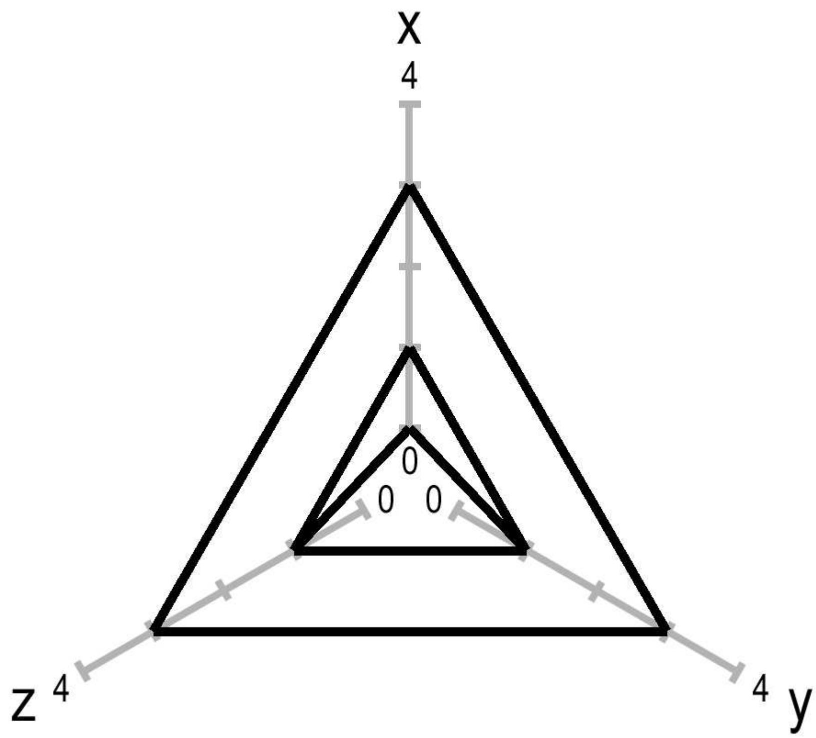


Figure A.1: Info-Blatt: Beispiel mit 3 Achsen und 3 Datensätzen

Bei dieser Visualisierungsmethode werden die Achsen im Kreis angeordnet. Die wichtigste Dimension wird herausgehoben. Wenn zum Beispiel die Zeit herausgehoben wird, wird für jeden Zeitpunkt eine eigene Graphik erzeugt. Die Werte eines Datensatzes werden miteinander verbunden. Wenn du mit der Maus auf eine Linie gehst, wird der gesamte Verlauf des Datensatzes farbig markiert. Gleichzeitig wird auf der Webseite unten Zusatzinformation angezeigt.

Die Beispiele

Zum Flugzeug-Beispiel: Die Dimension t ist der Zeitpunkt, x, y, z sind die drei Raumkoordinaten. Eine Kollision tritt dann auf, wenn mindestens zwei verschiedene Flugzeuge zum selben Zeitpunkt am selben Ort sind. Es wurden die Daten von 3 Flugzeugen zu 4 Zeitpunkten abgebildet (dh 12 Datensätze bzw. Linien).

Zum Medizinischen Beispiel: Die hier verwendeten Abkürzungen stammen von einer klinischen Studie, die im AKH durchgeführt wird. In dieser Studie wird der Therapiefortschritt von Anorexie- Patientinnen erhoben. Für dieses Beispiel wurden Fragen von mehreren verschiedenen Fragebögen miteinander kombiniert, wie sie z.B. von den betreuenden PsychologInnen ausgewertet werden. Mögliche Fragestellungen sind z.B.: Gibt es besondere Auffälligkeiten in den Daten, die für alle gleich sind? Ändert sich der Zustand einer Patientin oder bleibt er eher gleich? Weisen die Meßwerte einer Patientin eher auf einen guten oder eher auf einen schlechten Zustand hin? Gibt es Ähnlichkeiten zwischen einzelnen Patientinnen? Für 5 Patientinnen wurden drei Zeitpunkte abgebildet (dh 15 Datensätze bzw. Linien). Die Fragebögen wurden unverändert für diesen Test übernommen. Die Abkürzungen und zugehörigen Werte bedeuten folgendes:

Fragebogen BDI:
BDIR: Appetitverlust
BDIS: Gewichtsverlust
3 = ja, 2 = eher ja, 1= eher nein, 0= nein
Das bedeutet: Kleinere Werte sind besser als größere Werte.

Fragebogen EAT:
EAT5 Ich vermeide Essen, wenn ich hungrig bin.
EAT9 Ich achte auf den Kaloriengehalt der Nahrung, die ich esse.
EAT13 Ich erbreche nach der Mahlzeit.
5 = immer, 4 = sehr häufig, 3 = häufig, 2 = manchmal, 1 = selten, 0 = nie
Das bedeutet: Kleinere Werte sind besser als größere Werte.

Fragebogen FAMOS:
FAMOS14: Es ist mir wichtig, mich zu entspannen
FAMOS23: Es ist mir wichtig, etwas für mich zu machen
FAMOS37: Es ist mir wichtig, mir selber etwas zu gönnen
5 = außerordentlich, 4 = ziemlich, 3 = mittelmäßig, 2 = weniger, 1 = überhaupt nicht
Das bedeutet: Größere Werte sind besser als kleinere Werte.

Fragebogen MRFSF:
MRFSF1: Ich gönne mir Zeiten der Ruhe und Entspannung
MRFSF2: Ich achte genau auf meine körperlichen Bedürfnisse.
4 = ja, 3 = eher ja, 2 = eher nein, 1 = nein
Das bedeutet: Größere Werte sind besser als kleinere Werte.

Die Fragen

Zuerst ein paar Infos zu deiner Person: Alter, Geschlecht, Studienrichtung / Beruf.

Fragen zum Beispiel 1 (Flugzeug-Beispiel), 4 Dimensionen:
1. Ist deiner Meinung nach eine Kollision aufgetreten? Wenn ja, welche Flugzeuge waren beteiligt?
2. Bei welcher Graphik (welchen Graphiken) konntest du etwas ablesen? Wenn ja, was hast du dort abgelesen?
3. Welche Probleme / Schwierigkeiten hattest du bei der Interpretation?

Fragen zum Beispiel 2 (Medizinisches Beispiel), 11 Dimensionen:
4. Gibt es Aussagen, die auf den ersten Blick auffallen?
5. Bei welcher Graphik (welchen Graphiken) konntest du etwas ablesen? Wenn ja, was hast du dort abgelesen?
6. Welche Probleme / Schwierigkeiten hattest du bei der Interpretation?

Allgemeines:
7. Beschreibe kurz deinen ersten Eindruck von den Stardinates.
8. Deine Meinung zu diesem Test ist ...

List of Figures

List of Tables

Bibliography

[1] P. Ahonen-Rainio. Concept testing of some visualzation methods for geographic metadata. In *Proceedings of the ScanGIS'2003*, pages 103–117, 2003.

[2] J.R. Anderson. *Spreading Activation*, pages 61–90. Freeman, San Francisco, 1984.

[3] J.R. Anderson. *Kognitive Psychologie, 2.Auflage*. Spektrum der Wissenschaft, 1989.

[4] T. Avidan and S. Avidan. Parallax - a data mining tool based on parallel coordinates. *Comput.Stat.*, 13:1–65, 1998.

[5] M.Q. Baldonado, A. Woodruff, and A. Kuchinsky. Guidelines for using multiple views in information visualization. In *Advanced Visual Interfaces (AVI2000)*, pages 110–119. ACM Press, 2000.

[6] R.A. Becker and W.S. Cleveland. Brushing scatterplots. *Technometrics*, 29:127–142, 1987.

[7] J.G. Benjafield. *Cognition*. Prentice-Hall International Editions, 1992.

[8] M. Bolorfoush and E.J. Wegman. On some graphical representations of multivariate data. In *Proceedings of the Interface'88, Amer.Stat.Assoc.*, Alexandria, Virginia, 1988.

[9] D.E. Broadbent. The maltese cross: A new simplistic model for memory. *Behavioral and Brain Sciences*, 7:55–94, 1984.

[10] J. Bruner. *Toward a Theory of Instruction*. Harvard University Press, Cambridge, MA, 1966.

[11] J. Bruner. *Going Beyond the Information Given*. Norton, NY, 1973.

[12] J. Bruner. *Acts of Meaning*. Harvard University Press, Cambridge, MA, 1990.

[13] A. Buja, D. Cook, and D.F. Swayne. Interactive high-dimensional data visualization. *Journal of Computational and Graphical Stastics*, 5(1):78–99, 1996.

[14] S.K. Card, J.D. Mackinlay, and B. Shneiderman, editors. *Readings in Information Visualization*, chapter 1, pages 1–34. Morgan Kaufman, 1999.

[15] S.K. Card, J.D. Mackinlay, and B. Shneiderman, editors. *Readings in Information Visualization*, chapter 1, page 16. Morgan Kaufman, 1999.

[16] J.M. Chambers, W.S. Cleveland, B. Kleier, and P.A. Tukey. *Graphical Methods for Data Analysis*. Chapman & Hall, 1983.

[17] H. Chernoff. The use of faces to represent points in k-dimensional space graphically. *Journal of the American Statistical Association*, 68:361–368, 1973.

[18] M.C. Chuah and S.G. Eick. Glyphs for software visualization. In *Proceedings of the Fifth International Workshop on Program Comprehension (IWPC'97)*, pages 183–191, 1997.

[19] M.C. Chuah and S.G. Eick. Information rich glyphs for software management data. *IEEE Computer Graphics and Applications*, 18(4):24–28, 1998.

[20] M.C. Chuah and S.F. Roth. On the semantics of interactive visualization. In *Proceedings IEEE Symposium on Information Visualization 1996*, pages 29–36, San Francisco, CA, October 1996.

[21] S. Coren and J.S. Girgus. Principles of perceptual organization and spatial distortion: The gestalt illusion. *Journal of Experimental Psychology*, 6:404–412, 1980.

[22] M. Eagle, D.L. Wolitzky, and G.S. Klein. From 'imagery': The effect of a concealed figure in a stimulus. *Science*, 151:837–839, 1966.

[23] U.M. Fayad, P. Smyth, G. Piatesky-Shapiro, and R. Uthurusamy. *Advances in Knowledge Discovery and Data Mining*. AAAI/MIT Press, 1996.

[24] J. Gärtner, S. Miksch, and S. Carl-MacGrath. Vico: A metric for the complexity of information visualizations. In *Second International Conference on Theory and Application of Diagrams (Diagrams 2002)*, pages 249–263, Berlin, 2002. Springer.

[25] J.J. Gibson. *The Perception of the Visual World*. Houghton Mifflin, 1950.

[26] J.J. Gibson. *The Ecological Approach to Visual Perception*. Lawrence Erlbaum Associates, 1986.

[27] D.L. Gresh, B.E. Rogowitz, R.L. Winslow, D.F. Scollan, and C.K. Yung. Weave: A system for visually linking 3-d and statistical visualizations, applied to cardiac simulation and measurement data. In *Proceedings of the 2000 IEEE Conference on Visualization (Vis'2000)*, pages 489–492. IEEE Computer Society Press, 2000.

[28] G. Grinstein, R. Pickett, and M.G. Williams. Exvis: An exploratory visualization environment. In *Proceedings of the Graphics Interface '89*, 1989.

[29] R.B. Haber and D.A. McNabb. Visualization idioms: A conceptual model for scientific visualization systems. In B. Shriver, G.M. Nielson, and L.J. Rosenblum, editors, *Visualization in Scientific Computing*, pages 74–93. IEEE Computer Society Press, 1990.

[30] M. Henle. Koffka's principles after fifty years. *Journal of the History of the Behavioral Sciences*, 25:14–21, 1987.

[31] H. Hinterberger. *Data Density: A Powerful Abstraction to Manage and Analyze Multivariate Data*. PhD thesis, ETH ZUrich, Zurich, Switzerland, 1987.

[32] A. Inselberg. Visualization and knowledge discovery for high dimensional data. In *User Interfaces to Data Intensive Systems (UIDIS2001)*, pages 5–24. IEEE Computer Society Press, 2001.

[33] A. Inselberg and B. Dimsdale. Parallel coordinates: A tool for visualizing multi-dimensional geometry. In *Proc. of the First IEEE Conference on Visualization*, pages 361–378, 1990.

[34] A.S. Jacobson, A.L. Berkin, and M.N. Orton. Linkwinds: Interactive scientific data analysis and visualization. *Communications of ACM*, 37(4):42–52, 1994.

[35] E. Kandogan. Visualizing multi-dimensional clusters, trends, and outliers using star coordinates. In *Proceedings of the ACM SIGKDD '01*, pages 107–116. ACM, 2001.

[36] D.A. Keim and H.P. Kriegel. Visualization techniques for mining large daabases: A comparison. *Trans. Knowledge and Data Engineering*, 8(6):923–938, 1996.

[37] J. Lamping, R. Rao, and P.A. Pirolli. Focus + context technique basen on hyperbolic geometry for visualizing large hierarchies. In *Proceedings of ACM CHI95 Conference on Human Factors in Computing Systems*, pages 401–408, New York, 1995.

[38] M. Lanzenberger. The interactive stardinates. In *Proceedings of the Ninth IFIP TC13 International Conference on Human-Computer Interaction, Doctoral Consortium, September 1 2003, Zurich, CH*, pages 688–690. IOS PRess, 2003.

[39] M. Lanzenberger and S. Miksch. The hyperbolic hypotheses tree. In *Proceedings of the ED-Media 03, World Conference on Educational Multimedia, Hypermedia & Telecommunications, June 23-28 2003, HI, Honolulu*, pages 1062–1065. AACE, 2003.

[40] M. Lanzenberger, S. Miksch, S. Ohmann, and C. Popow. Applying information visualization techniques to capture and explore the course of cognitive behavioral therapy. In *Proceedings of the ACM Symposium on Applied Computing SAC2003, Florida, March 9-12 2003*, pages 268–274. ACM Press, 2003.

[41] M. Lanzenberger, S. Miksch, and M. Pohl. The stardinates—visualizing highly structured data. In *Proceedings of the Int. Conference on Information Visualization (IV03), July 16-18 2003, London, UK*, pages 47–52. IEEE Computer Science Society, 2003.

[42] E. Leeuwenberg and F. Boselif. Against the likelihood principle in visual form perception. *Psychological Reviews*, 95:485–491, 1988.

[43] A.J. Marcel. Conscious and unconscious perception: An approach to the relations between phenomenal experience and perceptual processes. *Cognitve Psychology*, 15:238–300, 1983.

[44] A.J. Marcel. Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitve Psychology*, 15:197–237, 1983.

[45] J. Metcalfe and D. Wiebe. *Intuition in insight and non-insight problem solving*, chapter 15, pages 238–246. 1987.

[46] G.A. Miller, E. Galanter, and K. Pribram. *Plans and the structure of behavior*. Holt, Rinehart & Winston, New York, 1960.

[47] T.M. Mitchell. *Machine Learning*. McGraw-Hill, 1997.

[48] C.J. Morris, D.S. Ebert, and P. Rheingans. An experimental analysis of the effectiveness of features in chernoff faces. In *Proc. of Applied Imagery Pattern Recognition '99: 3D Visualization for Data Exploration and Decision Making*, October 1999.

[49] K. Nakayoma and G.H. Silverman. Serial and parallel processing of visual feature conjunctions. *Nature*, 320:264–265, 1986.

[50] U. Neisser. *Cognitive Psychology*. Appleton-Century-Crofts, New York, 1967.

[51] U. Neisser. *Cognition and Reality*. Freeman, San Francisco, 1976.

[52] NIST/SEMATECH. Engineering statistics handbook. Internet address: http://www.itl.nist.gov/div898/handbook/.

[53] D.A. Norman and D.G. Bobrow. *Active memory processes in perception and cognition*, pages 114–132. Freeman, San Francisco, 1976.

[54] C. North and B. Shneiderman. A taxonomy of multiple window coordinations. Technical Report CS-TR-3854, 1997. citeseer.nj.nec.com/north97taxonomy.html.

[55] C. North and B. Shneiderman. Snap-together visualization: A user interface for coordinating visualizations via relational schemata. In *Proc. of Advanced Visual Interfaces AVI 2000, Palermo, Italy*, pages 128–135, 2000.

[56] R.M. Pickett and G.G. Grinstein. Iconographic displays for visualizing multidimensional data. In *Proc. IEEE Conf. on Systems, Man, and Cybernetics*, pages 514–519, Piscataway, NY, 1988.

[57] J. Preece, Y. Rogers, and H. Sharp, editors. *Interaction Desing*, chapter 12, pages 379–386. J. Wiley & Sons, 2002.

[58] J.R. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmannl, 1993.

[59] J.C. Roberts. On encouraging coupled views for visualization exploration. In *Visual Data Exploration and Analysis VI, Proceedings of SPIE*, pages 14–24, 1999.

[60] J.C. Roberts. Multiple-view and multiform visualization. In *Visual Data Exploration and Analysis VII, Proceedings of SPIE*, pages 176–185, 2000.

[61] G.G. Robertson, J.D. Mackinlay, and S.K. Card. Animated 3d visualizations of hierarchical information. In *Proceedings of SIGCHI'91*, pages 189–194. ACM, 1991.

[62] D.M. Russell, M.J. Stefik, P. Pirolli, and S.K. Card. The cost structure of sensemaking. In *Proceedings of INTERCHI'93, ACM Conference on Human Factors in Computing Systems*, pages 269–276, 1993.

[63] G. Schreiber, H. Akkermans, A. Anjewierden, R. de Hoog, N. Shadbolt, W. van de Felde, and B. Wielinga. *Knowledge Engineering and Management: The CommonKADS Methodolgy*. MIT Press, 1999.

[64] R. Schulmeister. *Grundlagen Hypermedialer Lernsysteme*. München, Germany, 1997.

[65] H. Schumann and W. Müller. *Visualisierung – Grundlagen und allgemeine Methoden*. Springer, 2000.

[66] K. Sedig and J. Morey. Facilitating learning through different forms of interaction with visual abstractions. In *Proc. of ED-MEDIA 2002: World Conference on Educational Multimedia and Hypermedia, Denver, USA, July 2002*, pages 1776–1777. AACE, 2002.

[67] O. Selfridge. Pandemonium: A paradigm for learning. *The mechanization of thought processes*, 1959.

[68] R.M. Shiffrin and W. Schneider. Controlled and automatic human information processing. *Psychological Review*, 84:155–171, 1977.

[69] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the IEEE Symposium on Visual Languages*, pages 336–343. IEEE Computer Society Press, 1996.

[70] J.H. Siegel, E.J. Farrell, R.M. Goldwyn, and H.P. Friedman. The surgical implication of physiologic patterns in myocardial infarction shock. *Surgery*, 72:126–141, 1972.

[71] G. Sperling. The information available in brief visual presentations. *Psychological Monographs*, 74(11), 1960.

[72] D.F. Swayne, D. Cook, and A. Buja. Xgobi: Interactive dynamic data visualization in the x window system. *Journal of Computational and Graphical Statistics*, 7(1):113–130, 1998.

[73] A. Treisman and S. Gormican. Feature analysis in early vision: Evidence from search asymmetries. *Psychological Reviews*, 95:15–48, 1988.

[74] J.W. Tukey. *Exploratory Data Analysis*. Addison-Wesley, 1977.

[75] L. Tweedie. *Characterizing Interactive Externalizations*, chapter 7.4, pages 616–623. Morgan Kaufman, 1999.

[76] L.A. Tweedie, R. Spence, H. Dawkes, and H. Su. Externalising abstract mathematical models. In *Proceedings of the ACM SIGCHI 1994, Vancouver, BC*, pages 406–412. ACM, 1994.

[77] M. Ward. Xmdvtool: Integrating multiple methods for visualizing multivariate data. In *Proceedings of the Visualization '94*, pages 326–333, 1994.

[78] C. Ware. *Information Visualization: Perception for Design*, chapter 5, page 165. Morgan Kaufmann Publishers, 2000.

[79] C. Ware. *Information Visualization: Perception for Design*, chapter 5, pages 180–188. Morgan Kaufmann Publishers, 2000.

[80] C. Ware. *Information Visualization: Perception for Design*, chapter 6, pages 201–240. Morgan Kaufmann Publishers, 2000.

[81] Colin Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann Publishers, 2000.

[82] M. Wertheimer. *Laws of organization in perceptual forms*, pages 71–88. Humanities Press, New York, 1967.

[83] D.D. Wickens. Encoding categories of words: An empirical approach to meaning. *Psychological Review*, 77:1–15, 1970.

Curriculum Vitae

Address	Monika Lanzenberger Wiedner Gürtel 64/11 1040 Vienna, Austria lanzenberger@ifs.tuwien.ac.at
Date of Birth	April 21st, 1971
Education	October 2001 to November 2003 Ph.D. studies in Computer Science at the Vienna University of Technology October 1996 to November 1997 Studies in Political Science (unfinished) at the University of Vienna October 1991 to July 2001 M.S. studies in Computer Science at the Vienna University of Technology

Job Experience	Since September 2001 Faculty member of the Institute of Software Technology and Interactive Systems August 1999 to August 2001 Conducted a project for women in informatics funded by the European Union (esf) Oktober 1996 to December 2002 Coach for Vocational Training in IT
Professional Memberships	ACM (Association for Computing Machinery) ÖGAI (Austrian Society for Artificial Intelligence) OCG (Austrian Computer Society) GI (German Informatics Society)
Publications	Please see: http://www.ifs.tuwien.ac.at/˜ mlanzenberger/publications.html