Chapter 11
Visualization

Daniel Keim* and Matthew Ward**
*University of Konstanz, Germany and **Worcester Polytechnic Institute, USA

11.1. Introduction

The progress made in hardware technology allows today's computer systems to store very large amounts of data. Researchers from the University of Berkeley estimate that every year about 1.5 Exabytes (= 1.5 Million Terabytes) of data are generated, of which a large portion is available in digital form [359]. It is possible that in the next three years more data will be generated than in all of human history to date. The data are often automatically recorded via sensors and monitoring systems. Even simple transactions of everyday life, such as paying by credit card or using the telephone, are typically recorded by computers. Usually many variables are recorded, resulting in data with a high dimensionality. The data are collected because people believe that it is a potential source of valuable information, providing new insights or a competitive advantage (at some point). Finding valuable information hidden in the data, however, is a difficult task. With today's data management systems, it is only possible to examine quite small portions of the data. If the data are presented textually, the amount of data that can be displayed is in the range of some one hundred data items, but this is like a drop in the ocean when dealing with data sets containing millions of data items. Having no possibility to adequately explore the large amounts of data that have been collected because of their potential usefulness, the data becomes useless and the databases become data 'dumps'.

Information visualization and visual data analysis can help to deal with the flood of information. The advantage of visual data exploration is that the user is directly involved in the data analysis process. There are a large number of information visualization techniques that have been developed over the last two decades to support the exploration of large data sets. In this chapter, we present

an overview of information visualization and visual exploration using a classification based on the *data type to be visualized*, the *visualization technique*, and the *interaction technique*. We illustrate the classification using a few examples, and indicate some directions for future work.

Benefits of Visual Data Exploration

For data analysis to be effective, it is important to include the human in the data exploration process and combine the flexibility, creativity, and general knowledge of the human with the enormous storage capacity and the computational power of today's computers. Visual data mining aims at integrating the human in the data analysis process, applying human perceptual abilities to the analysis of large data sets available in today's computer systems. The basic idea of visual data mining is to present the data in some visual form, allowing the user to gain insight into the data, draw conclusions, and directly interact with the data. Visual data analysis techniques have proven to be of high value in exploratory data analysis. Visual data mining is especially useful when little is known about the data and the exploration goals are vague. Since the user is directly involved in the exploration process, shifting and adjusting the exploration goals can be done in a continuous fashion as needed.

Visual data exploration can be seen as a hypothesis generation process; the visualizations of the data allow the user to gain insight into the data and come up with new hypotheses. The verification of the hypotheses can also be done via data visualization, but may also be accomplished by automatic techniques from statistics, pattern recognition, or machine learning, as discussed earlier in this volume. In addition to the direct involvement of the user, the main advantages of visual data exploration over automatic data analysis techniques are:

- Visual data exploration can easily deal with highly non-homogeneous and noisy data.
- Visual data exploration is intuitive and requires no understanding of complex mathematical or statistical algorithms or parameters.
- Visualization can provide a qualitative overview of the data, allowing data phenomena to be isolated for further quantitative analysis.

As a result, visual data exploration usually allows a faster data exploration and often provides more interesting results, especially in cases where automatic algorithms fail. In addition, visual data exploration techniques provide a much higher degree of confidence in the findings of the exploration. These facts lead to a high demand for visual exploration techniques and make them indispensable in conjunction with automatic exploration techniques.

Visual Exploration Paradigm

Visual Data Exploration usually follows a three step process: *Overview first, zoom and filter, and then details-on-demand.* According to Ben Shneiderman