V

Visual Methods for Analyzing Human Health Data

C. Tominski
Institute for Computer Science, University of Rostock, Germany

P. Schulze-Wollgast
Institute for Computer Science, University of Rostock, Germany

H. Schumann
Institute for Computer Science, University of Rostock, Germany

INTRODUCTION

Day by day, large volumes of health-related data are collected by physicians, health insurance companies, and public authorities. These data are potentially useful to understand the history, monitor the present, and predict the future of the health situation in order to ensure a high level of human health protection. To take advantage of this potential, it is necessary to analyze the data. However, this is a demanding task when facing constantly growing volumes of data.

One approach to tackle the analysis of human health data is the application of **visual methods**. In recent years, visualization of data has become a commonly accepted and widely used tool for the extraction of relevant information from arbitrary data. In many cases, a better insight into complex processes and phenomena can be gained by means of visual representation.

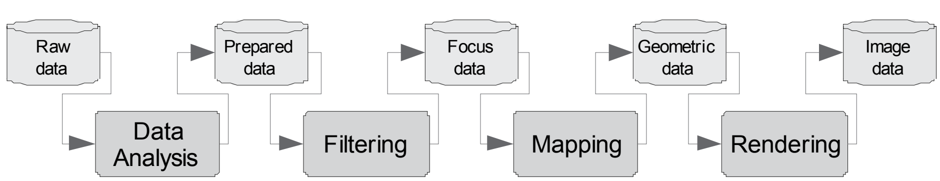
This chapter focuses on the **visual analysis** of human health data that describe the number of cases of different diagnoses in a spatial and temporal frame of reference. To build a common basis for the later description of different visualization methods, basic concepts of visualization as well as an abstract data model are illustrated in Background. In the main part of this chapter, we describe the visualization of human health data at various levels (see Visualizing Human Health Data). Whereas basic visual methods for representing human health data are presented only briefly, the visualization of data with respect to space and time is described in more detail. This chapter concludes with remarks on future work and trends in Future Trends and a brief summary of the key issues described in this article (see Conclusion).

BACKGROUND

Thanks to the capabilities of the human visual system, visualization is a promising tool to analyze larger volumes of data. If visualization is done properly, relevant information can be perceived intuitively, and the underlying data can be understood more easily. By proper visualization, it is meant that a visual representation has to be *expressive*, *effective*, and *appropriate*. Expressiveness relates to the requirement that all relevant information must be expressed in a visualization. Effectiveness depends on the degree to which a visualization supports easy and intuitive interpretation of the visualized facts. A visual representation is appropriate if gained benefit and required effort are balanced with respect to the task at hand.

Technically, the visualization process is implemented in four main steps: *data analysis*, *filtering*, *mapping*, and *rendering* (see Figure 1). They make up the **visualization pipeline** (dos Santos & Brodlie,

Figure 1. The visualization pipeline (dos Santos & Brodlie, 2004)

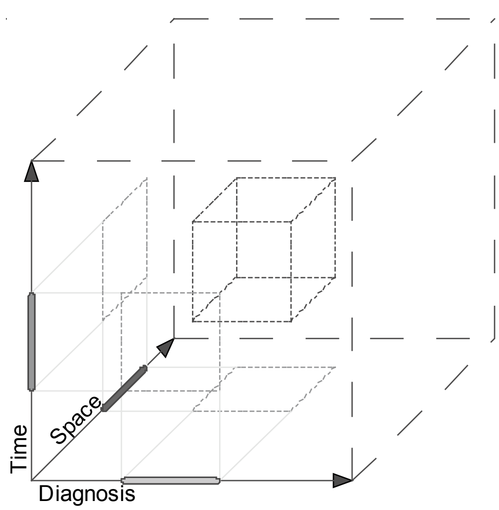


2004). To create a visual representation, a dataset is processed as follows. In the data analysis step, data are prepared for visualization (e.g., by applying a smoothing filter, interpolating missing values, or correcting erroneous measurements). The filtering step selects the data portions to be visualized (denoted as focus data). In the mapping step, focus data are mapped to geometric primitives (e.g., points, lines) and their attributes (e.g., color, position, size). The mapping step is the most critical one for achieving expressiveness and effectiveness, and hence, it is the most interesting one to visualization designers. Finally, geometric data are transformed to visual representations (e.g., images or animations).

Visualization aims at gaining insight by visually representing data. This implies that data characteristics are fundamental for any visualization. For this reason, research on visualization techniques, concepts, or methodologies must be grounded on a description of the addressed kind of data.

In this chapter, we address **human health data** that describe the number of cases of various diagnoses collected in a spatio-temporal frame of reference. Our data describe on a daily basis how many cases occur per diagnosis and per geographical region. From an abstract point of view, the data can be modeled as a data-cube (see Figure 2) that is constituted of three dimensions: time, space, and diagnosis. All these dimensions are of a hierarchical nature. Time uses days, months, quarters, and years; the spatial dimension comprises different administrative partitions of space (i.e., federal state, counties, and municipalities). The diagnoses are linked to the International Classification of Diseases (ICD10), which is hierarchical by definition. By relying on the data-cube model, it is relatively easy to reduce the volume of data to be visualized (e.g., by selecting only subranges of the dimensions [see Figure 2], or by using different levels of hierarchical abstraction). By this, only relevant data have to be extracted from the database. We will see in the next section that different visualization techniques vary in their usefulness for analyzing human health data depending on which dimensions of the data-cube are considered to what extent.

Figure 2. Different dimensions of health data modeled as a data-cube



VISUALIZING HUMAN HEALTH DATA

In the previous section, we indicated that human health data can be analyzed quite efficiently using visual methods. In this section, we suggest concrete **visualization techniques**. Depending on which dimensions of the described data-cube are addressed, different ways of representing human health data are possible:

- Multivariate representation of diagnoses
- Representation of diagnoses with respect to space
- Representation of diagnoses with respect to time
- Representation of diagnoses in time and space

Multivariate Representation of Diagnoses

In this first case, human health data are interpreted without spatio-temporal dependencies (i.e., only the diagnosis dimension of the data-cube is considered). In addition to diagnoses, it is also possible to enhance this abstract interpretation with derived statistics (e.g., maximum, minimum, average, and mean). The advantage of this interpretation is that classic visualization techniques can be applied to represent diagnoses and/or derived statistical characteristics visually. Simple diagram techniques (Harris, 1999) can be used to visualize frequencies of diagnoses (e.g., histograms for absolute frequency, pie charts for relative frequency). More sophisticated multivariate techniques such as Scatter Plot Matrices (Cleveland, 1993) or Parallel Coordinates (Inselberg, 1998) can be used to compare

Related Content