Visual Methods for Analyzing Human Health Data

C. Tominski*, P. Schulze-Wollgast, and H. Schumann
University of Rostock, Institute for Computer Science
Albert-Einstein-Str.21, D-18059 Rostock, Germany
Tel: +49 381 4987490, Fax: +49 381 4987482
Email: {ct, psw, schumann}@informatik.uni-rostock.de

1 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, and by this, 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 article 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 Section 2. In the main part of this article we describe the visualization of human health data at different levels (see Section 3). 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 article concludes with remarks on future work and trends in Section 4 and a brief summary of the key issues described in this article (see Section 5).

2 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

---

Figure 1. The visualization pipeline (dos Santos & Brodlie, 2004).
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, 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 ground on a description of the addressed kind of data.

In this article, we address human health data that describe the number of cases of different 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 the three dimensions: time, space, and diagnosis. All these dimensions are of 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 sub-

![Figure 2. Different dimensions of health data modeled as a data-cube.](image-url)
ranges 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.

3 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, and
- Representation of diagnoses in time and space.

3.1 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, 1998) can be used to visualize frequencies of diagnoses (histograms for absolute frequency, pie charts for relative frequency). More sophisticated multivariate techniques like Scatter Plot Matrices (Cleveland, 1993) or Parallel Coordinates (Inselberg, 1998) can be used to compare several diagnoses to reveal correlations among them (see Figure 3).

Figure 3. Health data represented as (a) Scatter Plot Matrix and (b) Parallel Coordinates.
Although these first visual representations are quite abstract, they are nonetheless an initial step toward insight into the underlying data. The general number of cases as well as outliers can be recognized. The drawback is that the data’s frame of reference is not considered (e.g., where is a peak in the number of influenza cases).

### 3.2 Representing diagnoses in space

A representation of the spatial frame of reference is necessary to allow an easy interpretation of data features in their geographical context. Maps have been used since ages to illustrate information with respect to geographic realities. Maps have the advantage that they enable the representation of data with respect to natural or manmade spatial phenomena like continents or states, which in our case are given as irregular-shaped regions at different spatial granularities.

If only one diagnosis has to be represented on a map, Choropleth Maps are commonly applied. In Choropleth Maps, each geographical region is color-coded according to the number of cases that occurred for the respective region (see Figure 4 (a)). If multiple diagnoses must be visualized, icon-based techniques are a good choice. Icons are small graphical primitives capable of encoding multiple attributes simultaneously. Furthermore, icons can be placed on maps to communicate spatial dependencies of diagnoses. Figure 4 (b) shows icons representing three different diagnoses for user-selected regions. A challenge that arises when using icons on maps is to avoid occlusion of icons. Particularly for areas with many small regions this is a demanding task (Fuchs, 2004).

![Figure 4. Health data represented via Chorpleth Map (a) and via icons positioned on a map (b).](image)

### 3.3 Representing diagnoses in time

To gain better understanding of the data, it makes sense to consider temporal dependencies. Even though visualization of time-related data is a challenge in its own right (Aigner, 2006), a compact overview of visualization techniques suitable for health data will be given.

A differentiation that must be made when visualizing temporal data regards the elements used for modeling the temporal dimension. Basically, a temporal domain is made up of...
Figure 5. A single diagnosis represented a time diagram (a) and as a Spiral Display (b).

either time points or time intervals\(^1\). Frank (1998) suggests a further categorization of time that regards the “shape” of the time axis. Frank distinguishes between linear, cyclic, or branching time. Whereas linear time considers a straight time axis from the past to the future, a cyclic time axis comprises recurring temporal elements (e.g., seasons of the year). Branching time axes are used to describe alternative scenarios of which only one will actually happen. These distinctions imply that, for instance, diagnoses that show cyclically recurring patterns (e.g., influenza) should be represented differently from diagnoses that do not show such behavior (e.g., injuries). Whereas techniques for visualizing health data with respect to time intervals or branching time are rare (e.g., (Plaisant et al., 1998)), the visual representation with respect to time points given on linear or cyclic time axes is considered in various visualization approaches. From simple yet expressive diagram techniques (Harris, 1998) to more sophisticated techniques, like Spiral Displays (Weber et al., 2001), TimeSearcher (Hochheiser, 2002) or parallel 3D bar charts (Chitarro, 2003), a broad range of tools for visually analyzing human health data is available.

A natural goal when representing human health data with respect to time is to reveal relations, and by this, to find correlations in the underlying dataset. How this goal can be accomplished is illustrated by the example of two visualization techniques – time diagrams and spiral displays. A time diagram is useful to compare different diagnoses with respect to a linear time axis (see Figure 5 (a)). A simple line-plot is used to facilitate an intuitive interpretation of the evolution of diagnoses in time. However, for detecting periodic patterns, Spiral Displays are better suited. Spiral Displays represent the time domain as a spiral-shaped time axis (see Figure 5 (b)). To enable the detection of periodic patterns of different length, Spiral Displays allow users to interactively adjust the number of time points encoded per spiral cycle.

Since the two illustrated visualizations, and also most of the techniques known in the literature, allow the visualization of only a few diagnoses, there is the need to restrict the visualization to important portions of the data. It makes sense to apply statistical analyses in the filtering step of the visualization pipeline to pre-determine diagnoses that potentially correlate, and choose highly correlated diagnoses for visualization.

\(^1\) A formal description of the terms time point and time interval will not be presented here; interested readers are referred to (Hajnicz, 1996) or (Aigner, 2006).
Techniques that are suited for representing multiple diagnoses with respect to a linear time axis are, for instance, the TimeWheel and the KiviatTube (Tominski et al., 2005). The TimeWheel comprises a central, horizontally aligned time axis around which several axes (representing diagnoses) are arranged (see Figure 6 (a)). To represent data values in a TimeWheel, lines descent from the time axis to all other axes. Clearly speaking, for each time step, individual line segments establish direct visual connections between time and all corresponding time-dependent values. The TimeWheel enables users to easily discern the number of cases of the represented diagnoses with respect to a particular time step. The TimeWheel also allows determining where in time certain numbers were measured. Correlations among the diagnoses can be revealed using the KiviatTube. It is constituted by a central time axis and several axes (representing diagnoses) that emanate perpendicular to the time axis. This arrangement of axes makes it possible to render a dataset as a closed three-dimensional surface (see Figure 6 (b)). The KiviatTube can be understood as a data representation that has been extruded along the time axis in 3D (similar to expanding an accordion). As such, the Kiviat Tube encodes not only the evolution of diagnoses over time, but also correlations among diagnoses.

3.4 Representing diagnoses in time and space

Even though it is sufficient in certain analysis scenarios to consider time and space separately, it is more often the case that both spatial and temporal aspects have to be visualized. To consider all three dimensions of the data-cube (see Section 2), three basic tasks must be addressed:

1. Represent the spatial frame of reference as map,
2. Create visual representations of the temporal aspects of the data, and
3. Embed representations of time (2.) into the representation of space (1.).

To represent the spatial frame of reference, the two-dimensional map (as introduced in Section 3.2) is represented in a 3D presentation space, in which the map can be zoomed and rotated using common 3D interaction. Users that prefer a two-dimensional map representation can switch easily to a respective map projection.

To visualize temporal aspects, pencil icons are used for linear time axes and helix icons are applied for cyclic time (Tominski et al., 2005).
**Pencil icons:** Since everybody is familiar with the shape of a pencil, it is a very useful visual metaphor. The natural shape of a pencil provides multiple faces that evolve from a common tip. This shape excellently serves as a 3D icon for visualizing multiple diagnoses with dependencies on linear time. To map data onto a pencil icon, linear time is encoded along the pencil’s faces starting from the common tip. Each face of a pencil is used to color-code one diagnosis (cp. Figure 7 (a)). The number of faces of a pencil can be adjusted according to the number of diagnoses to be visualized.

**Helix icons:** A spiral helix provides a geometric shape that allows an emphasis of the cyclic character of diagnoses. In order to construct a helix icon, a ribbon is created: For each time step the ribbon extends in angle and height depending on the number of temporal primitives per cycle and the number of cyclic passes. Again color-coding is used to encode data values along the ribbon (cp. Figure 7 1(b)). In order to represent multiple diagnoses, the ribbon can be divided into narrower sub-ribbons.

The described pencil and helix icons can be easily embedded into a 3D map display. This can be realized by positioning 3D icons at the centroids of geographical regions and aligning them with the z-axis of the 3D presentation space. By doing so, the representation of the temporal dependencies is shifted from the map to a dedicated dimension in the 3D presentation space. However, the embedding in 3D involves new problems compared to a 2D representation:

- Undesired changes of the icon view upon user interaction, and
- Loss of information due to icon occlusion and hidden surfaces.

Applying all interactions allowed in the 3D presentation space (i.e. rotation, translation, scaling) to both the map and the icons is disadvantageous, since it implies visual inconsistencies during the visual analysis. If users rotate the map while analyzing diagnoses represented by a pencil icon, the faces of the pencil are rotated as well, i.e. the view of diagnoses changes. An alternative is to unlink map interactions and icon interactions on demand. This can be achieved by aligning the icons with respect to the user’s current view into the 3D scene. This ensures that the icon view is preserved during

Figure 7. Visualizing monthly health data by means of 3D icons on a map: (a) Pencil icons representing the number of cases of six diagnoses; some diagnoses show a certain pattern over time; (b) Helix icons clearly reveal the cyclic characteristic of two selected diagnoses. Additional “tunnel views” mitigate, for a selected icon, the problem of hidden information.
map navigation. Users are allowed to change the icon view via separate icon rotation.

The second problem that must be addressed is loss of information due to occlusion of icons and on back faces of 3D icons. To alleviate this problem, the crude procedure of positioning icons at the areas’ centroids must be improved. An iterative approach that calculates occlusion conflicts, and alters the positions of icons locally can help to reduce occlusion. By applying an iterative algorithm that runs until no conflicts occur or until a maximum number of iterations has been exceeded, enhanced icon positions can be achieved and interactive frame rates are ensured. To mitigate the problems imposed by information lost on back faces, it makes sense to provide possibility to switch to special designated viewpoints. These viewpoints are located directly underneath each icon and the view directions are aligned with the icons’ axes. Switching to one of these viewpoints results in a “tunnel view” that reveals all data values represented by a selected 3D icon. Figure 7 shows how “tunnel views” support visual analysis in a 3D presentation space.

These enhanced viewing and interaction techniques in combination with common 3D interactions provide a rich basis for a visual exploration of multiple, time and space dependent diagnoses. Furthermore, it is possible to emphasize linear or cyclic characteristics of data by using dedicated 3D pencil or helix icons, respectively.

4 Future Trends

We presented a variety of visual methods for analyzing human health data. However, in the light of large datasets, visualizing all data in a comprehensible manner without burying possibly important information becomes more and more challenging. This challenge can be dealt with by conducting additional data analysis steps. Statistical and data mining methods like Principle Component Analysis or clustering are helpful analytical tools to support the identification of the important in human health data. Although the analysis step is indispensable, it is yet mainly data-driven. In the future, also the interests of users should be considered. This will allow for providing automatically adjusted visualizations that emphasize on relevant parts or hiding of non-relevant data. A further, particularly challenging problem is to find new ways of describing tasks of the visual exploration process and to automatically adapt the whole analysis procedure according to the users’ tasks at hand. This also includes providing specific interaction functions for investigating health data.

5 Conclusion

The analysis of human health data can be supported by interactive visual interfaces and appropriate visualization techniques. To achieve the main goal of visualization – gaining deeper insights into larger datasets – it is crucial to consider the characteristics of the data and the needs of the analyst. In this article we investigated the role of visualization in the context of human health data. We elaborated on choosing visualization techniques properly with respect to characteristics of the data. In particular, we described techniques for visually analyzing multivariate human health data in time and space. Although the explained techniques are promising tools to support a better health care, more work (as indicated in Section 4) has to be done to fully exploit the usefulness of visual analysis methods.
6 References


7 Terms and Their Definition

**Multivariate Data:** Data that contain multiple attributes are denoted multivariate data.

**Data Cube:** The dimensions of a dataset can be modeled as a hyper-cube, which allows an easy selection of relevant parts of the data.

**Visualization:** Visualization is a visual approach to gaining insight into data.

**Visualization Pipeline:** The visualization pipeline describes data analysis, filtering, mapping, and rendering as steps of visualization.

**Expressiveness:** A visual representation is expressive if it communicates all information that is relevant with respect to the current analysis task.

**Effectiveness:** A visual representation is effective if it has been generated in accordance with the capabilities of the human visual system.

**Appropriateness:** A visual representation is appropriate if its benefit and creation effort are balanced.

**Interaction:** Interaction denotes the use of specific techniques to adjust visual representations according to the task at hand.