Visualization of Health Data

N. López de Chávez H. Schumann K.-U. Graw

University of Rostock
Institute of Computer Graphics
Albert-Einstein-Str.21
18059 Rostock, Germany
tel. +49 381 4983421
lopez@informatik.uni-rostock.de

Fraunhofer Institute
of Computer Graphics
Julius-Jungius-Str.9
18059 Rostock, Germany
tel. +49 381 4024 128
kugraw@egd.igd.fhg.de

Abstract

To ensure a high level of human health protection the huge data sets gathered by the physicians of the health service have to be analyzed in an efficient way. In this paper we discuss the visualization of human health data with geographical and temporal dependence and describe the prototype of the visualization system MedVis which has been developed for the visual analysis of health data of the German state Mecklenburg-Vorpommern as part of a tele-consultation-system for the monitoring of acute health risk.

1 Introduction

Day by day a huge amount of medical data describing the diseases of patients is collected by the physicians of the health service. An effective analysis of these data is necessary to ensure a high level of human health protection. Especially precautionary actions against communicable diseases can prevent the outbreak and spreading of epidemics. However, the current information flow in Germany is too slow and the analysis of these data consists mainly of reports with simple tables published by the regional public health departments. Especially for the prevention of epidemics a better monitoring and a faster data analysis is necessary. Therefore, we have to use better means of communication (WWW) and more effective analysis methods, e.g. Scientific Visualization. Especially graphical representations allow an intuitive analysis of the data for different user groups and offer an alternative way to the tedious and time consuming extraction of information from statistical tables.

2 Application Background

We have 2 different data sources. One source are the notifiable diseases reported weekly by the physicians to the Public Health Department. They are given as weekly tables

$$\mathcal{MD} = \left\{ \left\{ \begin{array}{ccc} k_{1\,1}^{t_j} & \dots & k_{1\,k}^{t_j} \\ \vdots & \dots & \vdots \\ k_{m\,1}^{t_j} & \dots & k_{m\,k}^{t_j} \end{array} \right\} t_j \in [t_a, t_e] \right\}$$

where $k_{i\,j}^{t_r}$ denotes the number of occurences of the disease K_i in the time interval $(t_{r-1}, t_r]$ in the geographical area o_j and $[t_a, t_e]$ denotes the whole time interval. Another data source are the patients data recorded by the health insurance company AOK according to the Health Structure Law §301. These data cover 50% of the population of Mecklenburg-Vorpommern and describe every patient in detail with specific information, e.g. age and sex, length of illness and treatment.

$$\mathcal{CD} = \left\{ f_{o_i, t_j}^{K_l, m_1, \dots, m_k},\ o_i \in \mathcal{O},\ t_j \in [t_a, t_e],\ K_l \in \mathcal{K} \right\}$$

where $\mathcal{O}$ denotes the set of geographical areas, $\mathcal{K}$ the set of all diseases and $m_1, \ldots, m_k$ the parameters describing the disease and the patient. The data sets $\mathcal{MD}$ and $\mathcal{CD}$ are different from each other and are analyzed and presented in a different way. Furthermore, we use additional semantic and demographic data, such as granularity structur, neighbouring areas, area size, population density or age structure to weight the data acccording to the spatial context.

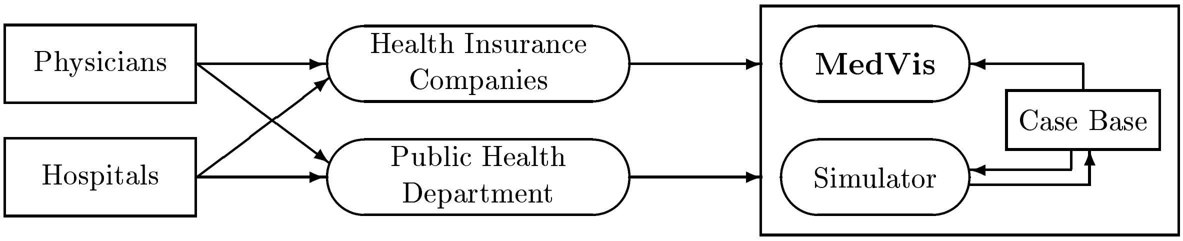

Figure 1: Data Flow in the Tele-Consultation-System

The prototype visualization system MedVis, currently developed in cooperation with the Institute of Medical Informatics and Biometrics, is part of the project 'TeCoMed - Tele-Consultation-System for acute health risks in Mecklenburg-Vorpommern' sponsored by the DFN (see Fig. 1). The goal is to provide physicians and pharmacists with information about current and future health risks in a geographical region (see [Gie96]). Our project partner use the Case-Base-Reasoning for the forecaste of future health risks. Our task is the visualization of the actual data $\mathcal{MD}$ and $\mathcal{CD}$ and the forcasted data. MedVis serves as a tool for the development and evaluation of the visualization techniques and the generation of the graphical representations. Finally, the users of TeCoMed will be able to access the data and their graphical representations through a WWW-Interface.

3 Visualization of Health Data

The prototype visualization system MedVis must be able to represent the geographical spatial dependence for different granularity levels and the temporal dependence in an adequate way to allow an easy assignment of relevant data features to their spatial and temporal context. Other demands are:

- Intuitive perception, i.e. the graphical representation have to be intuitive interpretable for different user groups with little experience in computer graphics
- No artifacts, because wrong conclusions due to misinterpreted artifacts in the graphical representation are very dangerous in medical applications. Thus, falsification and interpolation of the data are not allowed.
- Distinction of authentic and forecasted data, i.e. the graphical representation must show clearly which part of the data is authentic and which part is forecasted.

Both the generation of the graphical representation and the representation itself have to be manipulated interactively. We also need interactive focusing functionality, especially for accessing the data through the graphical representation (semantic interaction).

4 Underlying Concepts

4.1 Data Selection for visual Analysis of Health Data

The whole data contain too much information and can not be displayed in one picture. Furthermore, most users are only interested in a specific part of the data. Thus, the amount of data to display has to be reduced by focusing. This can be done by either a *projection* or a *selection*. A projection π reduces the amount of information by decreasing the number of parameters. The user chooses those parameters which are relevant for the visual analysis. Thus, he changes the structure of the data records to display:

$$\pi: \quad f_{o,t}^{K_l,m_1,\ldots,m_k} \longrightarrow f_{o,t}^{K_l,\tilde{m_1},\ldots,\tilde{m_{k_\pi}}}, \qquad 0 \le k_\pi < k,\ \tilde{m_i} \in \{m_1,\ldots,m_k\},\ i=1,\ldots,k_\pi$$

A selection σ reduces the number of data records to display.

$$\sigma: \qquad \mathcal{CD} \longrightarrow \mathcal{CD}_\sigma, \qquad |\mathcal{CD}_\sigma| < |\mathcal{CD}|$$

A selection σ is based on selection criteria defined with the operators $(=,<,>,\le,\ge)$ and composed with the logical operators $\neg,\wedge,\vee$. For instance, the selection

$$\sigma_{K=K_l \wedge o=o_r}: \qquad \mathcal{CD} \longrightarrow \mathcal{CD}_{\sigma_{K=K_l \wedge o=o_r}} = \left\{ f_{o_j,t_j}^{K_l,m_1,\ldots,m_k},\ t_j \in [t_a,t_e] \right\}$$

selects all data records with the disease K_l from the geographical area o_r.

Focusing in the medical data can be done with regard to

1. Parameter space (projection of parameters and selection of data records)
2. Time (selection of time steps, times and time intervals) or
3. Geographical space (selection of granularity level and geographical areas).

Whether to start focusing with regard to parameter space, geographical space or time depends mainly on the interpretation goals of the user, e.g. localization of a specific disease in a certain geographical area or time interval. Because dataflow oriented visualization systems do not support projection and selection in the data and the use of maps we developed and implemented the necessary modules for the visualization system IRIS-Explorer. Furthermore, we have integrated an information theoretical method for calculating correlations between parameters (see [Thei95]) to avoid an unfavourable projection of parameters.

For focusing by means of maps for different granularity levels we have to decide which information are to be shown for which granularity level and how to display data assigned to a lower granularity level on a rougher level? The aim is to present in every focusing step only those information which are necessary to make a decision. Only in the final graphical representation all details are shown. This approach reduces the cognitive demands on the user in the focusing process and data with relevant information can be located easier and quicker in the huge data set.

4.2 Visualization of Health Data in Space and Time

4.2.1 Visualization of Health Data

There are many visualization techniques for multiparameter data, e.g. geometric techniques (Prosection Views [STDS95]), pixel oriented techniques (recursive pattern), hierarchical techniques (Treemap [Joh93]), graph based techniques (SeeNet [BEW95]) and the iconographic

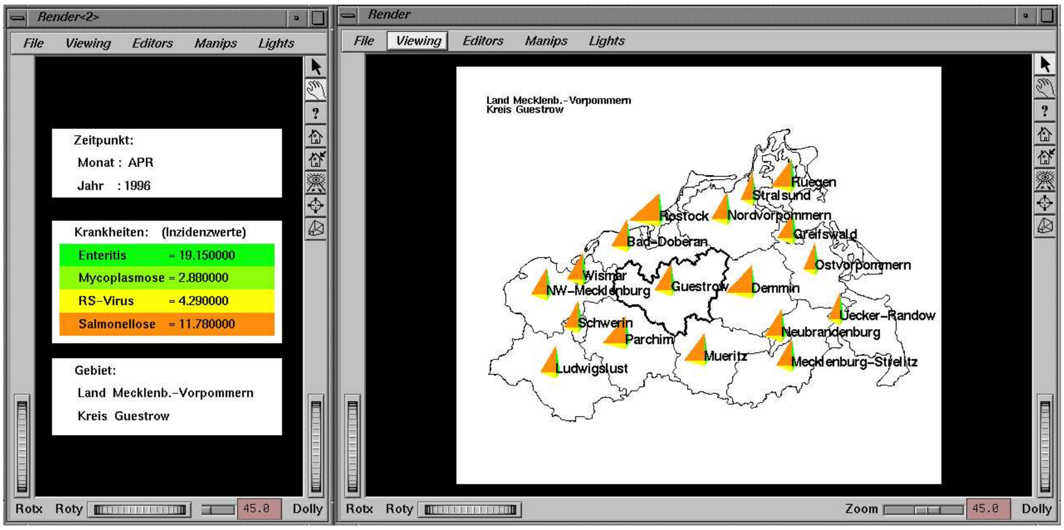

Figure 2: Display of the health data with icons

approach. All these techniques are suitable for different situations but only icons are suitable for the display of multivariate spatial relations because icons can be positioned exactly. Especially the identification of multivariate spatial correlations is an important interpretation goal for the analysis of the health data and icons can be used for the analysis of different kinds of correlations (see section 5), e.g. frequency of a disease for different age groups for a fixed time and geographical area or several diseases for a fixed time and geographical area.

4.2.2 Display of spatial Dependence

An adequate representation of the spatial frame of reference is necessary to assign important data features to their spatial context. Therefore, the main concern is the display of the different spatial granularities. Maps are very suitable for the display of geographic spatial dependence because they have a low level of abstraction which allows an intuitive perception of spatial relations. However, they are more suitable for the qualitative analysis of data with small numbers of data records and parameters. In our context we have to deal with different levels of granularity. On a rough granularity level there are usually small numbers of different geographical areas and icons can be used efficiently. On a more detailed granularity level we can still use icons if we zoom into the map and represent a smaller geographical region. However, the relation to the whole geographical region is lost. If we want to display a large geographical region on a more detailed granularity level and several parameters we have to use other methods where the display of spatial dependence is less intuitive.

Therefore we developed another approach. Instead of displaying the data corresponding to one geographical area with separate graphical representations we group areas according to their data and represent groups of areas with equal or similar data by one graphical representation, e.g. icon. Thus we reduce the number of graphical representations without loss of information and a display with icons could be possible. One icon represents the data of a group of several geographical areas and the areas which belong to a certain group are indicated by a specific encoding, e.g. colour coding.

The display of spatial dependence and the granularity level we have to choose depends also on the spatial interpretation goals like locating geographical areas with important data features or the spatial course of an epidemic. The display of the spatial course of communicable diseases, i.e. the spreading of epidemics is difficult. The data are assigned to areas of different size and do not carry explicit direction information. The data have to be weighted according to the number and the semantic and demographic data of the neighbouring areas to prevent falsification of the data. A possible way offers the arrow-icon (see Fig. 3) where the length of the arrows encodes the speed of the spreading in this direction and the colour encodes the actual incidence.

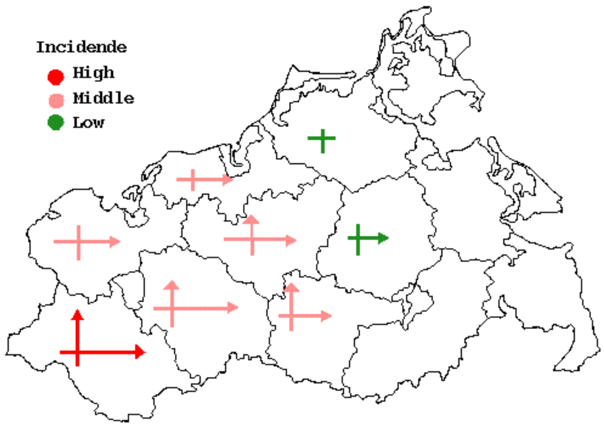

Figure 3: Arrow icon for the display of the spatial course of a disease

4.2.3 Display of temporal Dependence

We need visualization methods for identifying significant temporal changes and events by comparing static or dynamic representations for different times or time intervals, e.g. comparing data for selected geographical areas and diseases for different times or comparison of the temporal course of a specific disease for different geographical areas at a certain time interval.

The detection of a significant increase or decrease of a disease is an important goal and must be possible through visual analysis. A significant change is different for every disease. Sometimes it consists of only a minor change in the frequency which is difficult to display in an intuitive way. Therefore, we define events on the time axis to display temporal courses in a more striking way by using the concept of an event dependent visualization. Following [Lie96] an event dependent visualization is the display of a dynamic process where the graphical representation changes automatically when certain events occur. Thus, events are emphasized explicitely in the graphical representation which allows an easier detection of relevant changes (see Fig. 4). The alteration of the representation caused by an event can consist of different graphical attributes, e.g. different colour tables, icons or geometrical primitives.

An event dependent visualization reduces the cognitive demands significantly because the user has to deal only with the relevant part of the data which is less demanding than the analysis of a consistant graphical representation. However, we have to decide first which events and which shape of an event is going to change the graphical representation. Important events in our context are, for instance

- fast and significant changes of the frequency of a disease
- exceeding of disease related threshold values
- high incidence of a disease
- occurrence of exceptional and dangerous diseases

An event dependent graphical representation must display a specific event in such a way

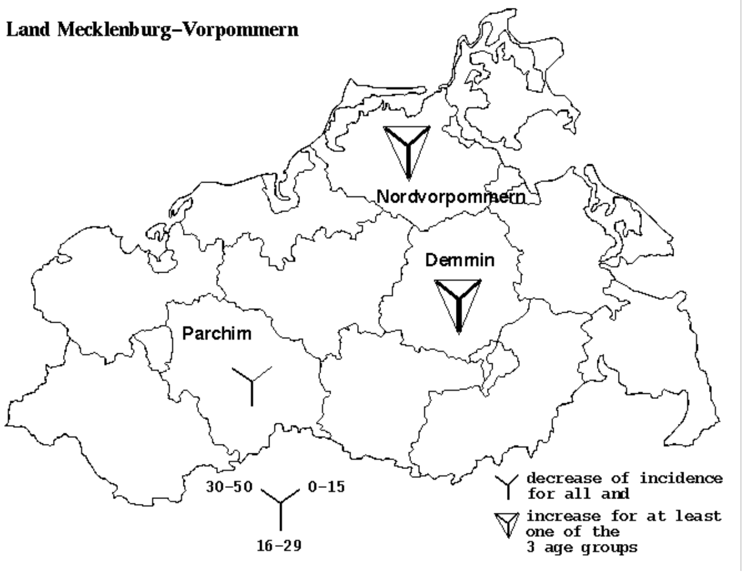

Figure 4: event dependent visualization with 2 different icons

that it can be distinguished easily from other events and the unimportant data. Different users are interested in different events and different event shapes. Events are data and user dependent which makes the definition of threshold values very difficult. Therefore, we focus only on few selected diseases, e.g. influenca and we want to include statistical methods to improve the numerical detection of events, especially for distinguishing between significant and insignificant changes.

5 The System MedVis

The prototype visualization system MedVis has been developed for the visual analysis of human health data of the German state Mecklenburg-Vorpommern. It serves as a tool for the development and evaluation of visualization techniques, focusing techniques and interactive functionality and is based on the visualization system IRIS-Explorer. We are using the NetCDF-format (see [RDE93]) for data input which allows the storage of additional meta information about the data.

Focusing in the Health Data

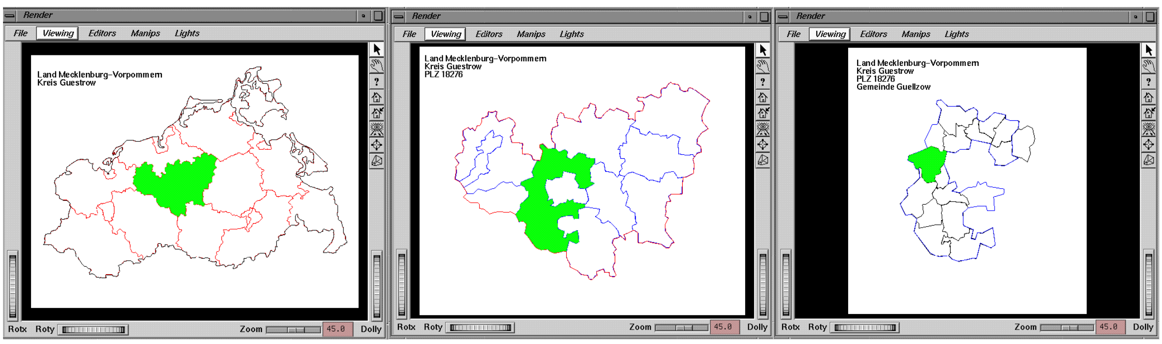

Figure 5: The 3 granularity levels of Mecklenburg-Vorpommern levels

Diseases, time steps and intervals and areas can be selected from lists or by explicit input. The data are scaled automatically according to the chosen time step. Geographical areas can also be selected by picking on the map. The map can be displayed in the 3 granularity levels district and district towns, post code areas and communities. The user can select the granularity level interactively and decide whether the whole region or only the selected areas are displayed (see Fig. 5).

Visual Representation of the Health Data

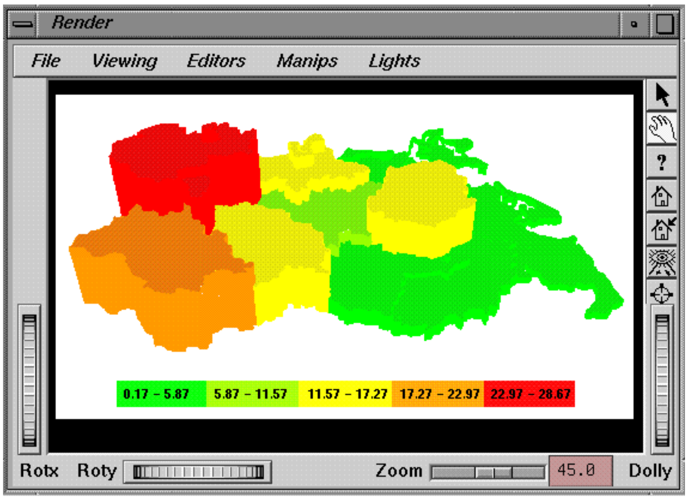

Figure 6: 3D colour coding

The visualization techniques have been implemented in close cooperation with the users. Colour coding is used for the display of one disease. Since we use 5 colour classes for easier detection of significant areas we offer 3D colour coding where the exact value is encoded in the height (see Fig. 6). We use icons (see Fig. 2) for the display of either one disease at different time steps or different diseases at a fixed time. MedVis offers different icon types and the parameter can be encoded in colour or size of the icon. Clicking at an icon displays the exact value in a separate window. Parallel Coordinates are used for the analysis of larger numbers of parameters and data records.

For the display of the temporal course of one parameter we use a multiple window display or an animation of a colour coding representations where the user controls the animation interactively. Pie-chart icons are used for the display of one disease for different times at the map. So far only a threshold based display has been implemented for event dependent visualization where areas are only colour coded when the parameter exceeds a user defined threshold value.

Grouping of geographical Areas

To allow the display with icons on a more detailed granularity level we have implemented the concept of grouping (see section 4.2.2). Areas, whose graphical representations would be equal, are grouped together and are represented by one icon, e.g. bar icon (see Fig. 7). The groups are formed automatically according to disease dependent threshold values. The areas belonging to one group are indicated by the same grey colour.

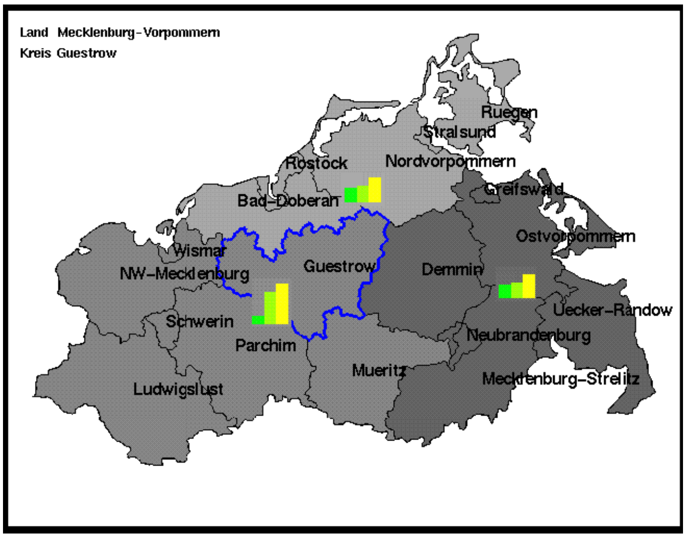

Figure 7: Grouping of geographical areas with similar data

6 Conclusions and future work

To improve the analysis of human health data we discussed the visualization of these data especially the display of the temporal and spatial dependence on different granularity levels. We introduced the concept of the event dependent visualization and the grouping of geographical areas for an intelligent visualization by emphasizing relevant information in the graphical representation. Furthermore, we described the prototype visualization system MedVis for the visual analysis of human health data as a part of a tele-consultation-system for the monitoring of acute health risk.

Future work will focus on the further development of the concept of an event depent visualization for the display of temporal changes and the practical implementation of this approach in MedVis. Especially the data dependent threshold values are still in the evaluation process. Furthermore, statistical methods which are currently developed by our project partners will be included in MedVis in the future.

References

[BEW95] Becker, R. A.; Eick, S.; Wilks, A. R.: Visualizing Network Data, Trans. on Visualization and Computer Graphics, Vol. 1, No. 1, 1995

[Gie96] Gierl, L.; Bull, M.: Architecture of an Early Warning System for Regional Health Risks, Rostocker Informatik Berichte (1996) 19.

[Joh93] Johnson, B.: Visualizing Hierarchical and Categorical Data, Ph.D. Thesis, Department of Computer Science, University of Maryland, 1993

[Lie96] personal talk with Robert van Liere at the 7th Eurographics Workshop, April 1996, Prag

[RDE93] Rew, R.; Davis, G.; Emmerson, S.: NetCDF User's Guide. An Interface for Data Access, University Corporation for Atmospheric Research, 1993

[STDS95] Spence, R.; Tweedie, L.; Dawkes, H., Su, H.: Visualization for functional design, Proc. Int. Symp. on Information Visualization, Atlanta, GA, 1995

[Thei95] Theisel, H.: Analyse und Visualisierungshilfe fuer mehrdimensionale wissenschaftliche Daten, Informatik - Forschung und Entwicklung, Heidelberg, 1995,10

Acknowledgements

The authors wish to thank Andrea Schumann for design and implementation work and providing some of the pictures. The work is part of the TeCoMed project supported by the German Research Network (DFN).