Keywords
Information visualization, electronic health records, data mining, decision support

ABSTRACT
With the increased complexity of electronic health data the demand for supporting health data visualization applications has grown recently. We summarize the existing health data visualization tools, analyze their strengths and weaknesses, and identify the most important missing components in current systems. We argue that existing visualization systems satisfy most of the various requirements of the different stakeholders, but are restricted to a specific scope of application. Interesting lines for further work are incorporating data mining algorithms and decision support algorithms, as well as integrating different visualizations into larger scale, multi-functionality systems. Such an integration will likely bring along new issues to tackle, like usability restrictions and data format issues.

1. INTRODUCTION
With the trend towards increased electronic health data accumulation (e.g. from mobile health applications [1] or by releasing public databases which were previously held under lock [5]), health data visualization applications have become popular over recent years. These applications target clinical research and appliance at the same time as government agencies and the general public (e.g. through personal health tracking applications on mobile phones). The possibilities of welfare gains for these various stakeholders are enormous, and there has been considerable research effort in developing health data visualization tools over recent years. Against this background, the goal of this term paper is to summarize for each stakeholder the components of a valuable and successful health data visualization application. In the following, we focus on clinical research and appliance and government agencies. The paper is structured as follows. In Section 2 we define the requirements and needs for the two types of applications described above. In the third section we review several existing systems with respect to the requirements and needs from section two. The paper concludes with a proposition for future systems.

2. APPLICATION REQUIREMENTS
Health data visualization applications (HVA) are important tools to gain insight into Electronic Health Records (EHR). In the field of clinical research and appliance they are often used to understand patient histories in an efficient, timely and user friendly way [6]. Further, HVA can be useful in finding hidden patterns which reveal important cause-and-effect phenomena. Two kind of patterns are of interest. First, the patient centered patterns which reveal important information for one patient, second, the inter-patient patterns revealing common information among several patients [1, 5].

A patients health history consists of the clinical information obtained by a physician in order to make the correct diagnostics for that patient. This information enclose symptoms, diagnostics, illness, recent treatments and other data about the patient from the past till this day. Based on this information, a physician draws his conclusions about the actual state of the patient. Such health histories can be of high complexity and visualizations have to consider all health conditions in order to enable physicians to work with the patients health data [6]. For this reason, HVA in this field should provide abstract (e.g. clustered) views of univariate and multivariate health variables over time to discover anomalies and relations in an overall view [6]. In addition there must be some flexibility to dive deeper into a specific pattern to gain detailed knowledge about it [1, 5]. An important component is also the ability to compare specific variables of one patient with the health data of other patients who have the same or similar demographic characteristics and patterns. This component is particularly important for research [5]. An overview of the various requirements for systems used in the field of clinical research and appliance can be found in Table 1.

With respect to government agencies, the most important...
applications of HVA are in the field of visualizing population phenomena such as public health care access and infectious disease-spreading tendency [3, 4]. To visualize population health phenomena, government agencies are more interested in aggregated EHR categories which are often multivariate and at the same time geographically distributed [3, 4]. Target HVA should enable users to visualize multiple public health care indicators (e.g., performance, access, and quality) with respect to time and geospatially, to apply dynamic queries to filter regions of interest by demographic attributes, and to produce visual results which makes it possible to discover important patterns in an easy and fast way [4]. The right hand side of Table 1 summarizes the three mentioned criteria.

<table>
<thead>
<tr>
<th>Clinical research and appliance</th>
<th>Government agencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i) Patient history overview</td>
<td>(iv) Multivariate pattern visualization for a specific region</td>
</tr>
<tr>
<td>(ii) Single-patient multivariate pattern visualization</td>
<td>(v) Multivariate pattern visualization for multiple regions</td>
</tr>
<tr>
<td>(iii) Multi-patient multivariate pattern visualization</td>
<td>(vi) Dynamic query filtering</td>
</tr>
</tbody>
</table>

### Table 1: Application Requirements

3. REVIEW OF EXISTING SYSTEMS

We turn now to reviewing several existing systems with respect to the requirements listed in Table 1. As the visualization applications for the two target user groups are quite different, the analysis is divided into two separate paragraphs.

#### Clinical Research and Appliance.

Several visualization approaches have been made to overcome the complexity of patients health histories. Interactive timelines representing temporal events of a patients life history became very popular recently. EHR are divided into several categories (such as symptoms, diagnoses, results, etc.) and attached to the timeline as labeled pins which display the patients health data at a specific time. Users have the possibility to query patterns either by filters next to the timeline, by clicking on legends to hide or show certain categories, or by directly manipulating the timeline (e.g. zooming in or out). A review of several systems implementing this approach can be found in [5]. Visualization applications based on the timeline approach are very useful to gain insight into time-related health data and can help physicians and researchers make inferences about which symptoms leads to which diseases and what medical treatments have to be taken into account to examine it. The main benefit of this approach is, that users can easily filter target indicators and thereby switch from a general view of a patients health variables to a detailed view revealing important patterns of the queried variables in a fast and clear way [5] (requirement (ii)). The main drawback however is that there exists no possibility to compare target variables among several patients (requirement (iii)). Furthermore, only a limited amount of health events can be displayed for a given time-range. For this reason, users have to either scroll forth and backward or to remember a lot of information in order to gain an overall picture of the patients health history [5] (requirement (i)).

An extension to this approach is the system *Lifeline2* [5] which, instead of visualizing health variables of a single patient, allows users to select a subset of EHRs from multiple patients. Records are attached to the timeline as colored triangles and can be clearly assigned to patients by IDs. This allows users to query patterns among multiple patients. Furthermore, users can rearrange events to reveal previously unknown patterns resulting in additional knowledge [5]. Such a multivariate and multi-patient comparison can, for example, help physicians and researchers link symptoms to specific diseases and find treatments which were successful in helping other patients. This approach is clearly the best solution regarding requirement (iii), but it is very restricted regarding requirements (i) and (ii).

A completely different approach was implemented in the system named *AnamneVis* [6] in which a patients health history is visualized as a radial sunburst (see Figure 1). Nodes representing health events such as diagnoses, symptoms, treatments, etc. (encoded as ICD9 codes) are drawn as wedges or bars around the human body in a circular fashion. The relative placement to neighbor nodes represents the relationship in the hierarchy. The angle of a node represents its number of incidents whereas the severity of a specific health event is color encoded. Three layers of nodes are used to cluster categories. Layer one stands for the highest hierarchy level, layer two represents more detailed categories and layer three contains the discrete incident node (symptom or diagnostic etc.). Red dots on the human body encodes incident locations. The whole sunburst contains the patients entire history. Having three different abstraction layers available, users can easily switch from an overall view of the patients health conditions to a very detailed view about a specific symptom, diagnostic or treatment. Moreover, the color encoding makes it easy to examine the severity of a specific health event. Because neighboring boxes represent relations between the displayed incidents, inferences from previously made decisions and the resulting effects are easily accessible. This features clearly address the requirements (i) and (ii).

*Figure 1: Patient health history visualization as radial sunburst (taken from [6])*
The main downside of AnamneVis is the lack of multipatient comparison which violates requirement (iii) and that the overall chronological order of health events is not available [6].

Taking all this information into account we can say that a timeline based visualization containing features for multivariate and single-/multi-patient comparison possibilities belong to the most valuable components of a HVA in this area of application. But particularly for physicians it is also of great interest to have a complete overview of a patients health history at hand, like it is implemented in AnamneVis.

Government agencies. The most common approach to visualize multivariate and geographically distributed health data is the use of interactive choropleth maps2 together with a filter to select the target health categories. Recently developed choropleth maps can be split up into two kinds of systems. The first one allows users to select a target health category and to display it as a color encoded attribute on an interactive map for multiple regions (e.g. Dartmouth Atlas [4]). Hovering a certain region compares the selected category relative to the average of the state or the country (using different kind of charts). The second kind of systems allows users to select and compare multiple health categories, but only for one specific region (e.g. Health Profiles Interactive [4]). An overview of such visualization applications can be found in [4]. As one can clearly see, both approaches violate requirement (v). In addition, the first type of systems also violates requirement (iv). The newly developed visualization application Community Health Map [4] tries to combine the two approaches by including several views on a single page (see Figure 2). This page contains a map view (1) to visualize one specific health category as color encoded attribute, a selection panel to select the health categories of interest (2), a dynamic filter panel to filter health categories by demographic variables (3) as well as a chart or table view to compare multiple health categories among different regions (4). Changing one of the demographic variables or one of the health categories automatically updates the map and table view. With the combination of several different views, Community Health Map addresses all three requirements and therefore represents a great visualization system for government agencies. Users can investigate target variables for a single region through dynamic query filtering among with the choropleth map display and for multiple regions through the chart or table view. One drawback is that the integration of multiple views into one visualization restricts the usability somewhat, as users have to switch the visualization context if they want to shift from a single-region to a multi-region representation. Moreover a chart or table view is not really suitable for pattern recognition as the process of searching for interesting information has to be done manually and is not supported by the visualization.

Against the background of these critics, another interesting system is MediMap [3]. Although it does not introduce new visualization techniques, it enhances the visualization by including data mining and decision support algorithms. Using different clustering methods to make similarity groups out

2A choropleth map is a map displaying color encoded statistical variables (e.g. public health access) for predefined regions

3For a detailed description of the applied algorithms please see [3].
be used to collect and store huge amount of standardized EHR. By adapting the API of the previously discussed visualization systems to the standardized EHR the data format issue is solved.

One thing that is missing in all reviewed HVA systems is the possibility to visualize multivariate and multi-patient health variables in a large scale manner by using data mining and decision support algorithms. The previously discussed systems allow users to find (hidden) patterns by manually searching for such events or by rearranging elements of the visualization (also randomly). However, one of the main purposes of good data visualization tools is to automate this process of revealing hidden patterns by using data mining and decision support algorithms together with data visualizations which can handle large amount of data and explain anomalies in the data by visualizing them in a clear and meaningful way without any involvement of users. Such a visualization system would be especially important for research, in which large amount of patients are compared by multiple variables. An interesting approach to address this problem is the data visualization tool Circos [2]. Circos targets the identification and analysis of similarities and differences in large scale data. With its efficient way of displaying variations or anomalies in either clustered or detailed views of related data (see Figure 3), Circos is a promising visualization application which combined with preprocessed health data is likely a benefit for researchers. As data mining and decision support algorithms are not part of Circos, they have to be applied on the dataset separately. The preprocessed data can then be loaded into the visualization tool. A possible example of use is to visualize all patients who have a disease of interest along with several other common characteristics like age, gender, etc. This could be done by using well-known dimensionality reduction and feature extraction algorithms (e.g. PCA or LDA) together with classification algorithms (e.g. KNN or ANN) to extract target patients out of the large scale dataset and to group them based on how good their characteristics match. Such preprocessed data would perfectly suite the visualization pattern of Circos which could then be used to visualize the similarities between those patients as described above. This would not only help physicians and researchers identify important patterns among multiple patients, but also to automatically select the relevant patients.

5. CONCLUSION

With the increased amount of collected electronic health data and an accompanied rising complexity in understanding and evaluating them, the demand for supporting health data visualization applications has grown recently. Our summary of existing systems shows that several good visualization systems have been implemented which satisfy the main requirements and needs of the various different stakeholders. Overall, we identify two key avenues for further work: First, one limitation that these current systems share is that they are in general very specific to one field of application, e.g. single-patient multivariate pattern recognition or multi-patient multivariate pattern recognition. However, for users, it would likely be beneficial to integrate these different functionalities and features into one flexible system. This combination of multiple visualizations will likely bring along new complications, such as usability restrictions and data format issues. There is room for interesting work to be done. Second, a feature lacking most systems is data mining and decision support. Given the increasing and complex amount of data available, the potential gains of such algorithms are enormous. Interesting work in this direction is done by Circos [2] which targets to identify and analyze similarities and differences in large scale data by providing an API to include such preprocessed data. This opens up a new and interesting line of work.

6. REFERENCES