

# Analytical, Visual, and Interactive Concepts for Geo-Visual Analytics

Heidrun Schumann<sup>a,\*</sup>, Christian Tominski<sup>a</sup>

<sup>a</sup>University of Rostock, Albert-Einstein-Straße 21, D-18059 Rostock, Germany

---

## Abstract

Supporting the visual analysis of structured multivariate geo-spatial data is a challenging task involving many different aspects. In this paper, we describe a systematic view of this task based on Chi's data state reference model. The analytical, visual and interaction components of the systematic view will be instantiated with specific examples that demonstrate how their tight interconnection facilitates exploration and analysis of geo-spatial data. In particular, we address the visualization of hierarchical structures on maps applying an extended focus+context concept. Moreover, we introduce an approach to extracting association rules from geo-spatial data and visualizing them on maps.

*Keywords:* Visual analytics, geo-spatial data, hierarchical data, extended focus+context, association analysis  
*2000 MSC:* 68U05, 68U35

---

## 1. Introduction

Visualizing data in a geo-spatial frame of reference is at the core of research in cartography and geo-visualization. For decades, many powerful methods have been developed in both fields. The focus of cartography is on the generation of expressive and aesthetic maps, where interactive manipulation is not a primary objective. Geo-visualization provides highly capable tools that allow users to navigate in space and time, to explore associated spatiotemporal data, as well as to interactively manipulate the visual representation and the underlying data. Today's web-based mapping applications (e.g., OpenStreetMap, Google Maps) and virtual globes (e.g., NASA World Wind, Google Earth) most impressively demonstrate that spatiotemporal data analysis and exploration is for everyone, and not just for experts.

The newly established research field of visual analytics, defined as “*the science of analytical reasoning facilitated by interactive visual interfaces*” [21], goes another big step forward. It aims at providing support for gaining insight into massive, heterogenous, dynamic, and ambiguous data by means of analytical computations, interactive exploration, and visualization of data, models, alternatives, and state information. Although first approaches to achieving this goal have already been developed, there are a variety of open questions and problems to be solved, specifically when considering visual analytics in space and time (see [15], Chapter 4).

In this work, we consider visual analytics support for the analysis of multivariate geo-spatial data and their struc-

tural relationships. We use a human health dataset as an example for such data. The data domain contains *data records* with quantitative and qualitative variables. Additionally, *data abstractions* (e.g., aggregates or clusters) and *structural information* (e.g., a cluster hierarchy or links between data records) associated with the data are of interest.

Such data can be analyzed with regard to many different goals and questions [4, 23]. Here we focus on two aspects in particular. First, we discuss the problem of representing hierarchies on maps and demonstrate our solution based on an extended focus+context concept. And second, we introduce a novel method for representing association rules on maps, which allows not only for communicating interesting patterns, but also for reducing the volume of data to be visualized. Although our work integrates analytical and visual methods as well as appropriate interaction techniques – the fundamental functionality of visual analytics tools – it clearly has a bias towards the visual part, as documented by a number of examples of map-based visualizations.

This article is structured as follows. To establish a systematic view, we will briefly describe a conceptual framework based on Chi's data state reference model in Section 2. In Section 3, the focus will be on an extended focus+context approach to geo-visual analytics. Section 4 serves to introduce the data we are dealing with. The novel techniques for visualizing hierarchies as well as for extracting and visualizing association rules in a spatial context are described in Sections 5 and 6. We conclude with a brief summary of the introduced concepts and directions for future work in Section 7.

---

\*Corresponding Author

Email addresses: [schumann@informatik.uni-rostock.de](mailto:schumann@informatik.uni-rostock.de)  
 (Heidrun Schumann), [ct@informatik.uni-rostock.de](mailto:ct@informatik.uni-rostock.de) (Christian Tominski)

## 2. Design of a Conceptual Framework

Next, we describe a conceptual framework that can serve as a basis for designing tailored visual analytics solutions for multivariate geo-spatial data. We do not propose a specific instantiation of the framework, but describe the general functionality involved. The data state reference model by Chi [6] is an excellent basis for this purpose. It describes the visualization process with all its facets in a generic manner. As Figure 1 illustrates, the model systematically describes: (1) the mapping process for transforming the raw data into image data via analytical and visual abstractions and (2) the application of a broad range of operators at different levels of the data stages. The visualization process is organized as a pipeline of four stages that the data have to pass through:

1. *Raw data*: In our case, the input data are multivariate data elements and relationships among them as well as the geo-spatial context.
2. *Analytical abstractions*: Abstractions are obtained from raw data values, for instance, by calculating aggregated values or by computing statistical moments. Analytical abstractions represent the major features of a dataset.
3. *Visual abstractions*: Visual variables such as size, color, or shape encode data values visually. They define the specifics of a visualization technique and decide about the design of a visual representation. Commonly, the geo-spatial frame of reference is given in terms of a geometrical description, which is a visual abstraction.
4. *Images*: The result of the rendering process are pixel data to be displayed on a visual output device.

Data is transformed and propagated through the pipeline by operators of two different classes: stage operators and transformation operators. Stage operators work within a single data stage, while transformation operators transfer data from one stage to another. Since we have four data stages, there are three different types of transformation operators (see top of Figure 1):

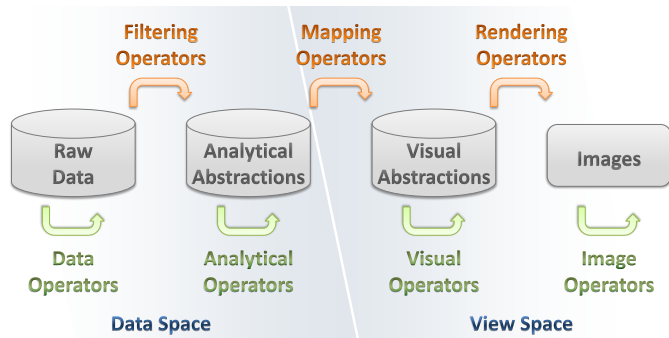


Figure 1: Schematic view of Chi's data state reference model.

- Filtering operators transform raw data values into analytical abstractions. In the light of very large datasets, this step is important, because one first has to condense the data down to a displayable size. In our case abstractions can be defined in terms of data, relationships, and/or the geo-spatial frame of reference. Thus, a particular challenge is to answer the question regarding which abstraction should be computed for which kind of raw data and how to relate both kinds of data.
- Mapping operators take analytical abstractions or subsets of the raw data as input and map these onto visual abstractions. This step is crucial and has to be handled carefully. In our case, visual encodings have to be found for data, structures, and geographic space, and these encodings should be visually distinguishable. For example, it should be clear to the user whether a line represents a part of a boundary in the geographic space or a relation between data elements.
- Rendering operators process the visual abstractions in order to obtain the visual representation. Here, an interesting question is how to manage different rendering facilities, e.g., representing the geographic space in Google Maps, but defining the multivariate data display by another rendering system.

The data state reference model distinguishes between data space and view space. Data operators, analytical operators, and filtering operators belong to the data space and provide automatic computations on data. They can be summarized as the *analytical component* of our conceptual framework. On the other hand, visual operators, image operators, and rendering operators belong to the view space; they define the *visual component* of the framework. The question is where to put the remaining mapping operators. Taking into account the importance of the mapping step, one option could be to define a separate component that provides the mapping functionality. However, we argue for including the mapping operators into the visual component. We do this because the mapping strategy determines the design of the visual representations, and therefore it is tightly interconnected with the operators of the view space.

All operators are characterized by their functionality as well as by the data they work on. Moreover, each operator also has its specific means to enable users to interactively manipulate the associated parameters and data. Conceptually, one can capture the interaction functionality by a separate *interaction component* to allow for a uniform access on data and operators through a consistent user interface.

In summary, the conceptual framework consists of three main building blocks: the analytical component, the visual component, and the interaction component. In order to be

broadly applicable, these components should provide operators for all kinds of data: multivariate data, geo-spatial data, structural relationships, and data abstractions.

In the literature one can find various instantiations of the conceptual framework. The OECD eXplorer [14] visualizes statistical data and supports collaborative web-enabled exploration and analysis. Other approaches address the challenge of extracting and visualizing meaningful patterns from massive movement data [3]. Not only clustering, but also self-organizing maps are utilized to make sense of geo-spatial data [16]. An application area particularly related to our work is the visual analysis of epidemics [20]. All these approaches have in common that they integrate analytical, visual, and interactive means. But each of them focuses on specific aspects and data, and hence, implements a different subset of operators of the conceptual framework.

In the following, we take a closer look at the visual component. It has to include diverse operators that are able to handle various kinds of data (e.g., multivariate geo-spatial data, structural relationships, and abstractions derived from the data), and to generate different kinds of visual output (e.g., 2D and 3D presentations). This way, data can be explored with respect to various tasks and analysis goals [4]. There are many powerful techniques to visualize the data to support users in finding answers to their questions. Choropleth maps [17], cartograms [8], and glyphs on maps [22] as well as small multiples [24] can be considered classic means to visualize multivariate spatial data. Because of the immense size of today’s datasets, more and more techniques address the visualization of data abstractions, most prominently the visualization of clusters [2].

In order to enhance the functionality of the visual component, we introduce three novel approaches. First, in the next section, we extend the well-known focus+context concept. This sets the stage for the visualization of complex data on maps. Second, in Section 5, we utilize the extended focus+context concept to visualize hierarchically organized information on maps. Third, in Section 6, we discuss the combination of analytical and visual methods using the example of extracting and visualizing association rules. Association rule analysis is a well-established data mining technique, but it is hardly applied in the context of geo-visual analytics. The goal is to extract relationships (i.e., associations) between specific data values, and in this way to crystallize patterns in large datasets.

### 3. Extended Focus+Context

The focus+context concept has been used in information visualization, geo-visualization, and cartography for a long time. Focus+context techniques (see [19, 7] for surveys) combine a *focus* representation that displays a part of a greater whole at high detail with a surrounding *context* representation that shows information at lower detail

to provide an overview. The different levels of granularity can be achieved in two ways:

- by semantic properties of the data, e.g., showing individual values in the focus, but only average values or cluster centroids in the context, or
- by applying a combination of different rendering styles to emphasize the region of interest and deemphasize the context.

The first case addresses the question “What has to be displayed?” and has to be handled by the analytical component. The second case focuses on the question “How does it have to be displayed?” and has to be realized by the visual component of the conceptual framework.

Usually, current approaches work on the basis of one unique domain of reference. In the context of geo-visual analytics, this concept has to be extended. Here, we have two domains of reference: the geographic spatial domain and the associated data domain. Consequently, there are now two foci: the *map focus* that corresponds to the current geographic region of interest and the *data focus* that encompasses the data associated with that region of interest. These foci are mutually dependent: Selecting a map focus defines the data focus, and vice versa.

Since we have two foci, we consequently consider two contexts: one for the map space and the other for the data domain. The *map context* consists of all geographic regions that are not focus regions. The *data context* includes all data outside the data focus, but at a coarser level of abstraction. However, it cannot always be taken for granted that the degree of abstraction is adequate for showing the data context on a map. Reasons for this might be that the data is not sufficiently abstractable or that the task at hand prohibits stronger abstraction. As a consequence, the data context might still be too complex (i.e., more than just a single data value per region) to keep visual clutter under control. Therefore, it makes sense to display the data context not for the entire map context, but only for a well-defined subset of the map context. We call these regions the *immediate context*.

In summary, full data detail is shown in the focus, coarser data abstractions are shown in the immediate context, and no data is shown for the map context (only the geographic regions are retained to maintain map coherence for better user orientation).

Because this limits the amount of information visible at a time, focus+context approaches have to support the flexible specification of the focus and the context by various interactive and automatic means. When using map representations, an intuitive way of specifying the map focus is the selection of regions of interest on the map. This means that the map focus is selected first, and then the corresponding data focus is derived by searching the data domain for data associated with the map focus. This might require additional operations if the granularity of the map does not match the level of abstraction of the data domain.

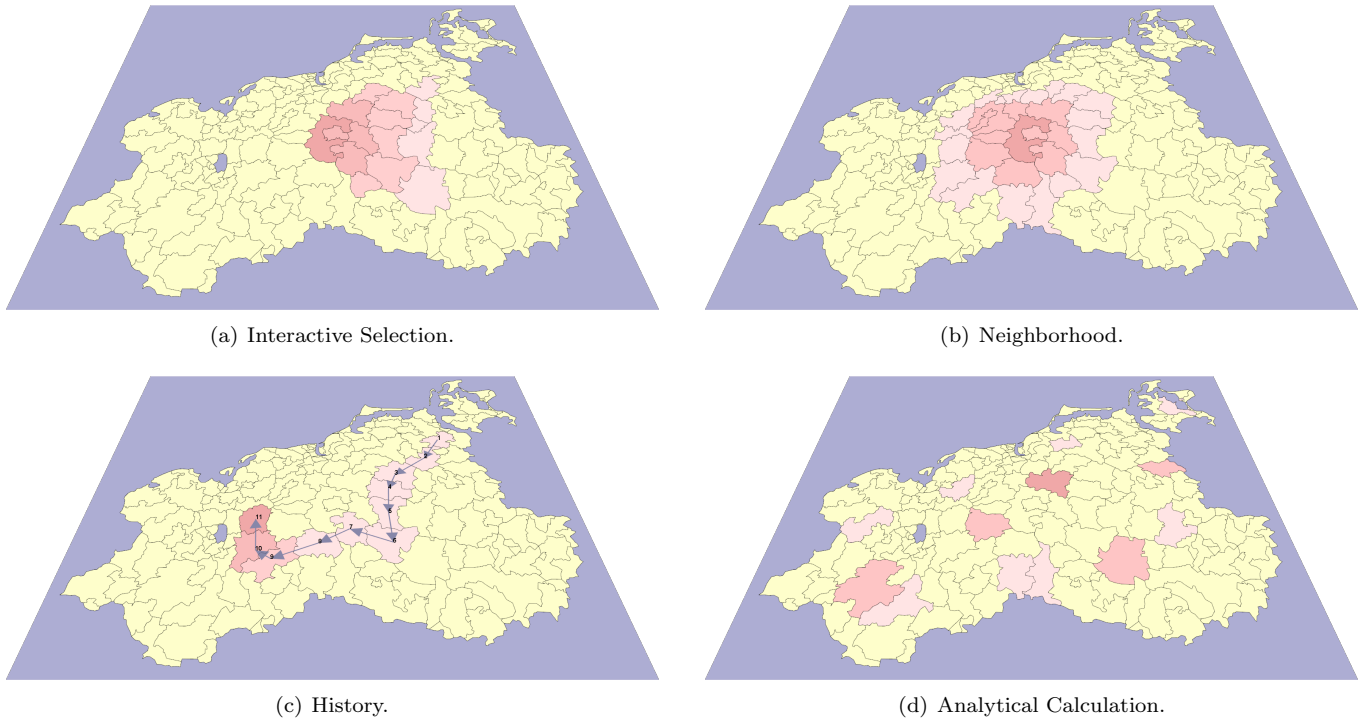


Figure 2: Different options to determine the immediate context.

In such cases, further aggregation of data records or automatic enlargement of the map focus can be applied to resolve inconsistencies between map granularity and data abstraction. The opposite way is that the user first selects data items of interest as the data focus. The map focus can then be derived directly from the selected data records, since each record is associated with a concrete geo-spatial reference.

These two ways of defining the focus (i.e., select regions and derive data vs. select data and derive regions) correspond to *direct lookup* and *inverse lookup*, which are two elementary tasks of exploratory analysis of spatiotemporal data as defined in [4].

Once the foci have been specified, the map regions of the immediate context have to be found. As illustrated in Figure 2, there are several options for determining the immediate context with respect to a user selected focus:

- *Interactive Selection* – The user manually selects regions for the immediate context.
- *Neighborhood* – The system determines a  $k$ -neighborhood of the focus and automatically sets it as the immediate context.
- *History* – The immediate context is constructed from regions that have previously been focus regions during the course of the user’s interactive exploration.
- *Analytical Calculation* – The immediate context is defined based on calculations in the data domain.

For instance, similarity measures can be computed and if the similarity value relative to the data focus is above a certain threshold, the corresponding map region is part of the immediate context. It would also be possible to cluster the data. Then, all regions whose data fall into the cluster that contains the data of the focus are part of the immediate context.

After the regions of the immediate context have been determined, the corresponding data context has to be extracted. To this end, appropriate analytical methods generate data abstractions at a coarser level of detail.

Finally, the representation style of the focus and the context need to be chosen. The goal is to communicate that the focus bears relevant detail information. For the map focus this can be achieved by means of accentuation of the focus region. A variety of visual modifications can be used to accentuate the focus, as long as there is no conflict with the actual visualization of the data. As examples of accentuation of the map focus one can imagine drawing wider contour lines or using a dedicated highlighting color.

For the data focus, a suitable visualization technique must be selected and the resulting visual representation is positioned within the map focus. We will illustrate this point in more detail in Section 5.

The depiction of the context should make clear that the information displayed there is not as relevant as that of the focus. The implication for the map regions of the immediate context is that the visual attributes have to be chosen so that they are less prominent than those used to

draw the map focus. Moreover, if the immediate context supports the notion of distance to the map focus, the visual representation should convey this fact. In Figure 2, we illustrate this by varying the color in the immediate context to attenuate the regions with decreasing proximity from the focus region. Similar visual effects can be achieved by modifying the width of contour lines or by alpha-blending entire context regions. Note that proximity can have different meanings, including recency of the interactive selection, distance in the neighborhood graph, or similarity of regions according to computations of an analytical method. The remaining map context is usually represented in a most abstract fashion, for example, by drawing only the boundaries of the associated regions.

For the presentation of the data context, the general goal is to avoid overloading the picture. This can be accomplished by semantic means (e.g., data abstraction or filtering) or graphical means (e.g., graphical simplification or de-accentuation by reducing opacity). Figure 5 in the next section demonstrates this with an example.

Taken together, the extended focus+context concept integrates visual, analytical, and interactive means to assist users in conducting geo-visual analytics. By focusing on relevant aspects and omitting less relevant ones, the exploration and analysis can be made scalable with regard to the size and complexity of geo-spatial data. The concept supports multiple focus regions and can be combined with other existing cartographic display techniques (e.g., cartographic lens). It is generally applicable and not restricted to a specific use case.

#### 4. Structured Spatiotemporal Data

In the next two sections, we will detail the discussion about representing data in their geo-spatial frame of reference. First, in Section 5, we describe a focus+context-based visualization of hierarchies on maps. Second, in Section 6, we include analytical methods into our considerations and introduce a visualization of data abstractions derived from association analysis. In this section, we describe the data that will be used in those sections.

Our examples are based on the scenario of visually analyzing public health data. Each data entry contains a time stamp, a geo-reference, and a diagnosis identifier (three independent variables), as well as the number of people suffering from that diagnosis at that location at that time (one dependent variable). More formally, this corresponds to a mapping  $f : \text{Time} \times \text{Space} \times \text{Diagnoses} \rightarrow \mathbb{N}$ .

In practice, we are interested in higher level aggregations of the data, say the number of people suffering from problems with the respiratory tract (aggregation of diagnoses) in February (aggregation in time) at the level of counties (aggregation in space). Hierarchies, irrespective of being natural or artificial, are an effective means to structure data along different levels of aggregation [9]. Our data contain three hierarchical structures (see Figure 3):

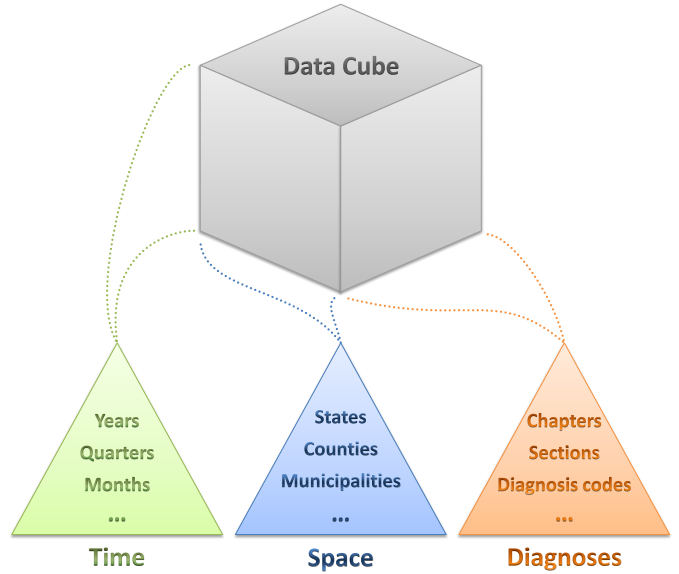


Figure 3: Data model structured along three different hierarchies.

- the hierarchically structured dimension of time (i.e., days, weeks, months, quarters, years),
- a hierarchically organized set of geographic regions (i.e., municipalities, counties, state), and
- the hierarchical classification of diseases based on the ICD-10 <sup>1</sup>.

These three hierarchies allow us to flexibly access the data and combine different levels of abstraction, and thus they are the primary means to support interactive information drill-down via focus+context mechanisms.

#### 5. Representation of Hierarchies on Maps

As with our data, hierarchical relationships are inherent in datasets of a variety of applications. The relationships can be explicit (e.g., hierarchical structure of time) or can be implicit in the data, which requires extraction by appropriate means (e.g., hierarchical clustering). Because of their widespread use, hierarchies and associated data values are often the subject of analytical investigations. By looking at the hierarchies in their geo-spatial context, we hope to find answers to a number of analytical questions in terms of the structural properties of a hierarchy as well as in terms of the data associated with the hierarchy:

- What pattern does the hierarchical structure (e.g., data abstraction or categorization) exhibit for a specific region?
- Are the hierarchies' structural properties similar across the regions of the map?

<sup>1</sup><http://www.who.int/classifications/icd/>

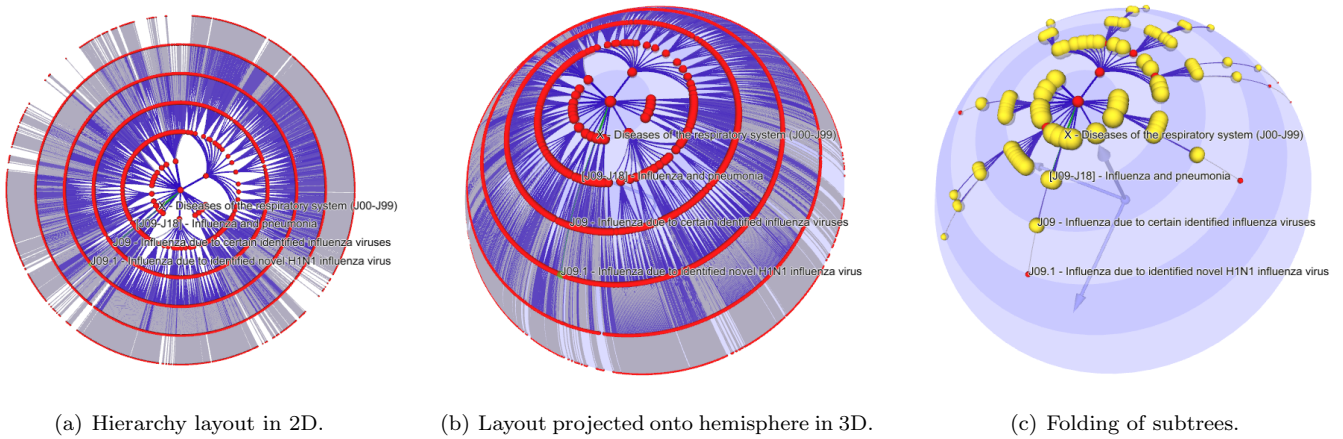


Figure 4: Magic Eye View showing the ICD-10 classification hierarchy with 43k nodes.

- How are the data values distributed in the hierarchy of a particular region?
- Is there a significant difference of the value distribution when comparing a specific region with its neighbors?
- Do certain patterns emerge when focusing on the data along a path through the map?

As a concrete example, one can consider an analyst who wants to find out how the total number of people being sick in February is distributed across different diagnoses, and whether there are any significant differences depending on where the people live. To satisfy the analyst’s needs, we have to display the hierarchical categorization of diseases (ICD-10) and the corresponding number of people afflicted in February as well as the geo-spatial context, all at an appropriate level of aggregation. From a conceptual point of view, this corresponds to the visualization of multiple attribute hierarchies (i.e., hierarchies where each node is attributed with one data value) on a geographic map.

Recent work has addressed techniques to display hierarchical relations on maps [11]. We extend this work and present an alternative display. Our approach uses an explicit 3D representation of hierarchy layouts, rather than an implicit 2D embedding of layouts into the regions of the map, and it makes use of the extended focus+context concept.

In order to integrate visual representations of hierarchies into a map, a number of requirements have to be met:

- The visual mapping of the hierarchy layout has to be compact.
- The hierarchy should be clearly distinguishable from the map.

- The relation between a hierarchy and the map region it is associated with must be visualized.
- Different levels of map granularity and data abstractions need to be considered.
- Appropriate interaction techniques are required including data selection, view manipulation, and information drill-down.

The combination of the extended focus+context concept with the Magic Eye View (MEV) approach [18] fulfills these requirements. In order to create a compact visual representation of the hierarchy, we first compute a layout of the hierarchy nodes in two dimensions. To this end, we adapt the Walker layout described in [5] so as to use polar coordinates, rather than Cartesian coordinates. As illustrated in Figure 4(a), this results in a layout where the hierarchy’s root node is located in the center, and all other nodes are arranged on concentric circles around the center, with the deepest hierarchy nodes being located on the outermost circle. This layout is then projected onto a three-dimensional hemisphere as depicted in Figure 4(b).

This compact representation is easily distinguishable from the map background and allows us to place several such hemispheres into the map display much like floating balloons anchored at individual map regions. However, Figure 4(b) also clearly makes a point that large structures such as the ICD-10 tree with its 43k nodes need additional effort to make them comprehensible. Therefore, we limit the number of nodes shown per MEV as illustrated in Figure 4(c), as well as the overall number of MEVs to be embedded into the map display. This reduction of the information load is steered depending on foci and contexts interactively selected by the user or automatically determined by analytical means (see previous section).

Figure 5 illustrates our approach. For the focus region, a larger MEV is shown with a high elevation above the map. Around the focus, the immediate context shows

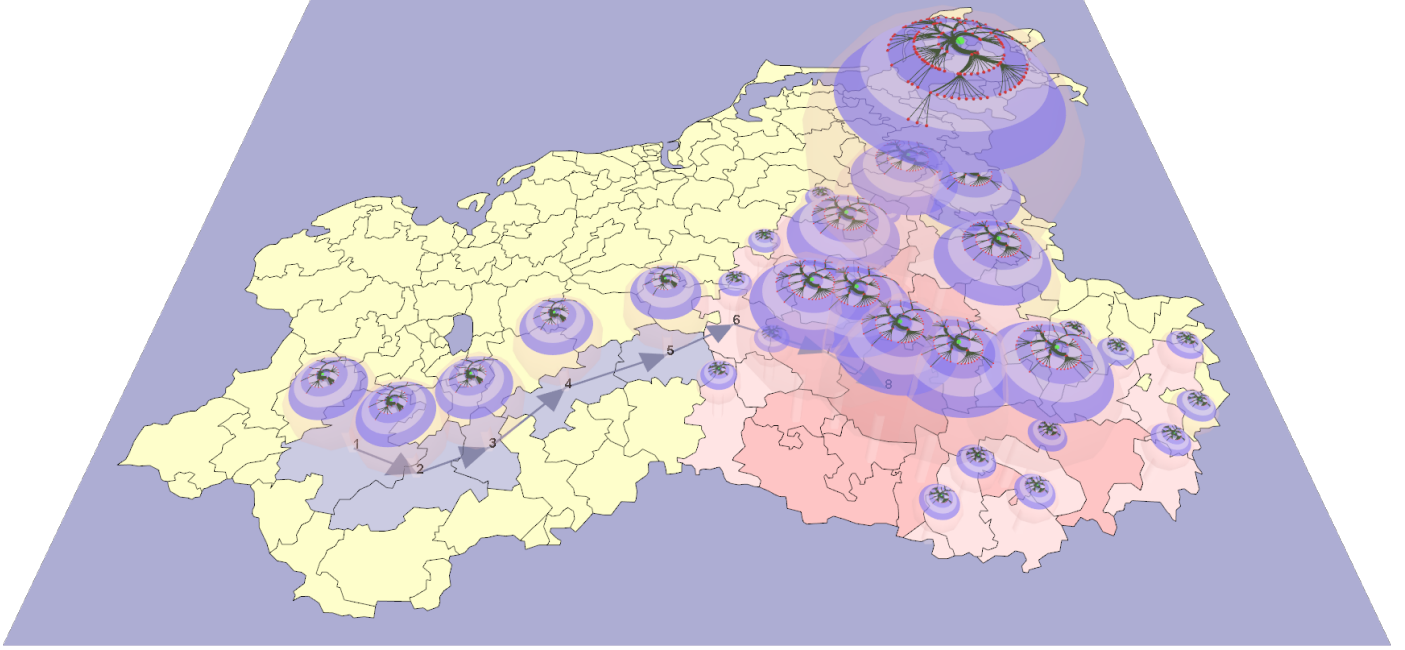


Figure 5: Illustration of the combination of the extended focus+context concept with Magic Eye Views.

increasingly smaller MEVs at increasingly lower elevation. Additional MEVs are displayed for regions that have been in the focus earlier during the visual exploration. Using size and elevation are purely graphical means to realize the focus+context concept. In terms of analytical means, our implementation utilizes automatic folding of subtrees based on the classic filter fisheye approach [10]. For the focus, the most relevant nodes are shown, while for the MEVs in the context, more and larger subtrees are automatically folded retaining just essential structural information. Moreover, it is quite easy to visualize attributes associated with nodes and edges of the hierarchy. Node attributes could be encoded with node color or size, and edge attributes could be visualized by varying edge width. The latter is shown in Figure 5, where complexity of subtrees (estimated by Strahler numbers [12]) is indicated by edges of different width.

A number of interaction techniques allow users to adapt the visual representation, including switching to different levels of spatial granularity, selecting a different focus, accessing the data of a different time period, and manually adjusting the level of folding of hierarchy nodes in the MEVs. It goes without saying that the map and the MEVs can be rotated and zoomed interactively. Optionally, the rotation of MEVs can be linked, meaning that rotating one MEV automatically rotates all other MEVs as well. Another useful option is to unlink map and MEV rotation. This way, the map can be viewed from different perspectives, but the view on the data (i.e., the nodes facing the user) remains constant.

The approach we presented here is generally applicable to visualize hierarchically structured *data* that are associated with the regions of a map. It can also be used to

complement the visual analysis of *data abstractions*, for example, to represent hierarchies generated through hierarchical clustering (see [25]), where individual regions may yield different cluster hierarchies.

While clustering is widely applied to support geo-visual analytics, there are alternative data abstractions whose usefulness in the context of visual analytics has not yet been sufficiently explored. Association rules represent one such data abstraction; they will be described in the next section.

## 6. Extraction and Visualization of Association Rules

The extraction of *association rules* from data is a well-established data mining technique. The goal is to detect relationships or associations between specific data values of categorical or classified variables, and in doing so, to enable analysts to uncover hidden patterns in large datasets. The goal is to arrive at statements such as “*When a large number of people suffer from diagnosis A in region B at time C it is likely that diagnosis X occurs in region Y at time Z*”.

### 6.1. Basics of Association Rule Analysis

Let us briefly review the basic notations behind association rule analysis [1, 13]. Let  $I = \{i_1, i_2, \dots, i_n\}$  be a set of  $n$  distinct literals, called items. Let  $\mathcal{D} = \{T_1, T_2, \dots, T_m\}$  be a set of  $m$  transactions, called the database. Each transaction  $T_i \in \mathcal{D}$  has a unique transaction ID  $i$  and contains a subset of the items, i.e.,  $T_i \subseteq I$ . Let  $X$  and  $Y$  be itemsets with  $X, Y \subseteq I$  and  $X \cap Y = \emptyset$ . Then, an association rule is an implication of the form

$X \Rightarrow Y$ , where  $X$  and  $Y$  are called *antecedent* (left-hand-side or LHS) and *consequent* (right-hand-side or RHS) of the rule, respectively.

In order to find interesting rules among all possible rules it is necessary to define measures of significance. Support and confidence are most commonly used for this purpose. They are defined as follows. The support  $supp(X)$  of an itemset  $X$  is defined as the proportion of transactions in  $\mathcal{D}$  that contain  $X$ , that is,  $supp(X) = \frac{|\{T \in \mathcal{D} | X \subseteq T\}|}{|\mathcal{D}|}$ . The support of a rule  $X \Rightarrow Y$  is defined as  $supp(X \Rightarrow Y) = supp(X \cup Y)$ . Using the support, the measure of confidence of a rule can be formalized as  $conf(X \Rightarrow Y) = \frac{supp(X \cup Y)}{supp(X)}$ . This formalization gives an estimate of the probability of finding  $Y$  in  $\mathcal{D}$  provided that  $\mathcal{D}$  contains  $X$ .

Using the two significance measures support and confidence as well as user-specified thresholds for both, association rule mining is usually conducted in two phases:

1. Find frequent itemsets, i.e., find all itemsets in  $\mathcal{D}$  whose support is above the minimum support threshold.
2. Use the frequent itemsets and the minimum confidence threshold to form rules.

The problem with the first step is that one has to search the power set  $\mathfrak{P}(I)$ , which is of size  $2^n - 1$  (empty set not included). Thanks to the downward-closure property of support [1], efficient algorithms can be applied to mine for frequent itemsets. The second step is a rather straightforward filtering process.

## 6.2. Determining Association Rules

These basic steps are commonly agreed on in association rule mining. However, finding expressive association rules is not an easy endeavor. In particular, we should raise awareness about the fact that conclusions can only be drawn with regard to the given literals (or items). For the human health data we consider here, no ad-hoc specification of the literals is possible. Our data are of multivariate quantitative nature, but literals have to be qualitative.

We suggest an interactive approach to derive literals from the data. Users specify different parameters and select statistical measures and transformation functions to be computed, and in this way interactively generate sensible literals, which later on can be used to find frequent itemsets, and eventually to mine to association rules complying with the task at hand.

A data object of our human health dataset is given by a tuple  $(t, r, d, n)$  where  $t$ ,  $r$ , and  $d$  are independent variables with  $t$  representing a point in time,  $r$  a geographic region,  $d$  a diagnosis, and  $n$  being a dependent variable holding the number of people suffering from  $d$  in region  $r$  at time  $t$ . While time, regions, and diagnoses can be easily interpreted as qualitative values from their respective aggregation hierarchies (e.g., time: 3rd quarter, region: city of Rostock, diagnosis: influenza), the overall tuple is not,

because the measured number of people is a quantitative value.

In order to determine qualitative literals  $(t, r, d, q)$ , where  $q$  is a qualitative value, first appropriate statistical measures are computed such as minimum, maximum of diseases, mean, variance, and others. Then, a set of qualitative literals is constructed. In this initial study, we used five options for  $q$ :

- $q = \text{THRES}$  – A literal  $(t, r, d, \text{THRES})$  is constructed, if for the tuple  $(t, r, d, n)$  the number of sick people  $n$  exceeds a certain threshold, that is,  $n$  is significantly large.
- $q = \text{MIN/MAX}$  – A literal  $(t, r, d, \text{MIN/MAX})$  is constructed, if the number of sick people reaches a minimum/maximum within a short time window around  $t$ .
- $q = \text{INC/DEC}$  – A literal  $(t, r, d, \text{INC/DEC})$  is constructed, if the number of sick people steadily increases/decreases during a short time period starting at  $t$ .

While THRES resorts to basic thresholding, the other options capture a notion of short-term trends in the data. These literals are then used to construct itemsets. The next steps follow the general procedure for association rule mining: frequent itemsets are determined based on the minimum support threshold and those itemsets are extracted that comply with the minimum confidence threshold. We experimented with itemsets with one or two literals only. This is sufficient to finally arrive at association rules such as *"If during a period of five days influenza reaches a peak in the city of Stralsund as well as in the county of Nordvorpommern, then there is a maximum in the city of Rostock one week later with a probability of 96.6 percent."*

Let us once more emphasize that the construction of literals is an interactive process, because the different diagnoses require individual thresholds and trend windows of individual length, which usually have to be determined by domain experts. Using different parameter values (e.g., for thresholds or for the minimal support) as well as different options for constructing literals also leads to different rules, and thus to different results and insights. This is why an interactive specification is indispensable in order to support application-specific and task-dependent association rules. Also note that our list of options to construct literals is not exhaustive. Alternative transformations could take spatial proximity into account to construct further literals that are particularly useful with regard to spatial analysis.

## 6.3. Visualizing Association Rules

While the association rules determined through the analytical component of our conceptual framework, the visualization of rules on a map has to be realized by the



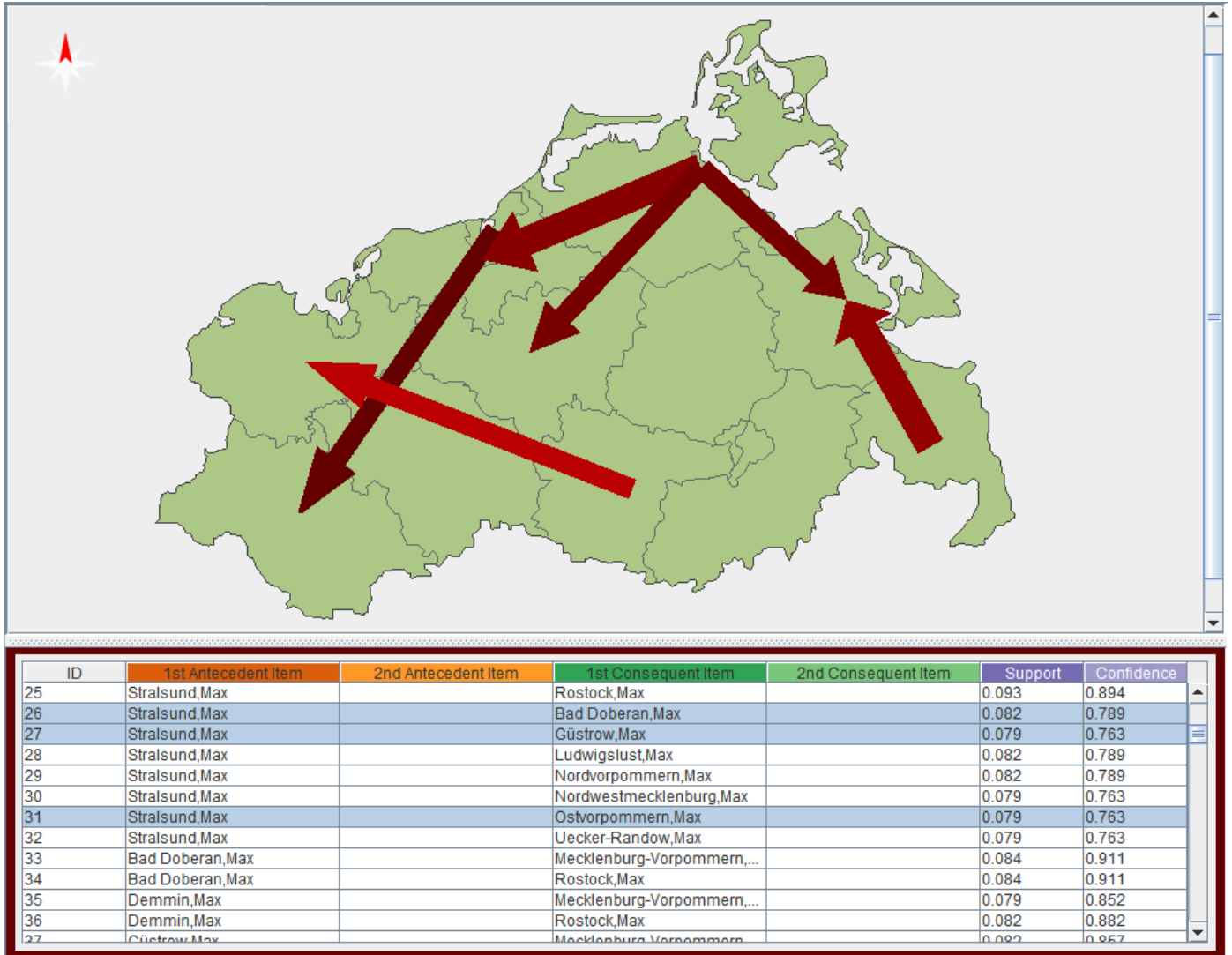


Figure 6: Visualization of association rules using arrows on a map.

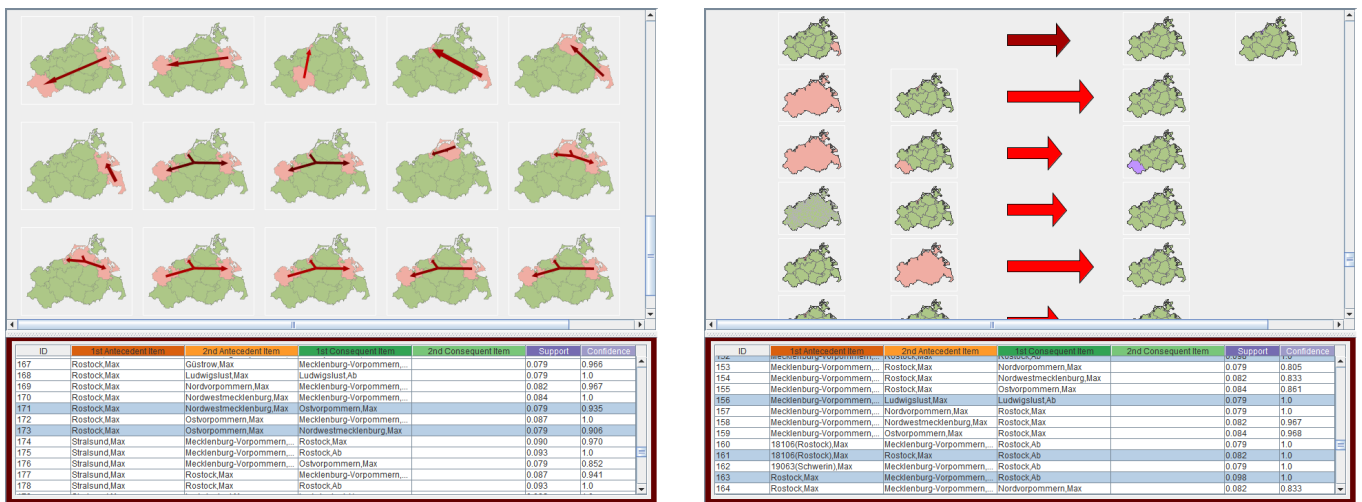


Figure 7: Small multiples for visualizing association rules.

visual component. This is a novel challenge, because we have to represent association rules in their spatial frame of reference, rather than plain data values. It makes sense to utilize multiple views to present the various aspects involved. The association rules are represented by a table view (see bottom of Figure 6). Each row of the table represents one rule. The first two columns contain the literals of the left-hand side of the rule, the columns three and four contain those of the right-hand side. The last two columns of the table show the values for support and confidence. Antecedent, consequent, and significance measures are indicated to the user by differently colored column headers. This table representation provides an initial overview of the given rules. The rule table provides all the functionality that is typically included within such table views: scrolling, sorting, as well as selecting and highlighting of rules.

The visualization of the rules selected within the table view integrates different visual means with a map display:

- arrow plots (see Figure 6),
- small multiples (see Figure 7), and
- color coding (see Figure 8).

The arrow plot shows the implications  $LHS \Rightarrow RHS$  as straight arrows superimposed on the map (see top of Figure 6). Arrow plots communicate well the spatial relationships of rules, and each arrow encodes rule support and confidence with arrow width and saturation, respectively.

However, the number of rules that can be presented on a single map is limited. Moreover, rules with self-reflections (circular arrows) and rules with multiple literals in LHS and/or RHS ( $n$ -to- $m$  arrows) increase clutter and thus cognitive efforts during the analysis. Rules that contain geographic regions from different levels of granularity (e.g., state vs. counties) cannot be represented on a single map, because a map usually shows only the regions of one selected level of granularity.

Therefore, our prototype provides different arrangements of small multiples [24]. Figure 7(a) shows the presentation of small multiples where each map visualizes a single rule. Here, arrow-arrow and arrow-map occlusions are minimized to clear the view on a per-rule basis at the cost of reducing the map display to a small thumbnail. Although details cannot be recognized by these small-sized images, the general dependences of regions can be communicated. Consider, for instance, the fourth and fifth thumbnails in the top row. We can infer that both rules originate from the same region, but the rules' destinations as well as their significance are different. This might indicate a stronger relation between the regions highlighted in the fourth thumbnail.

Figure 7(b) shows an arrangement that is suitable for representing rules with  $n$ -to- $m$  dependencies and dependencies across different levels of spatial granularity. Each row of the arrangement illustrates one rule as follows. The

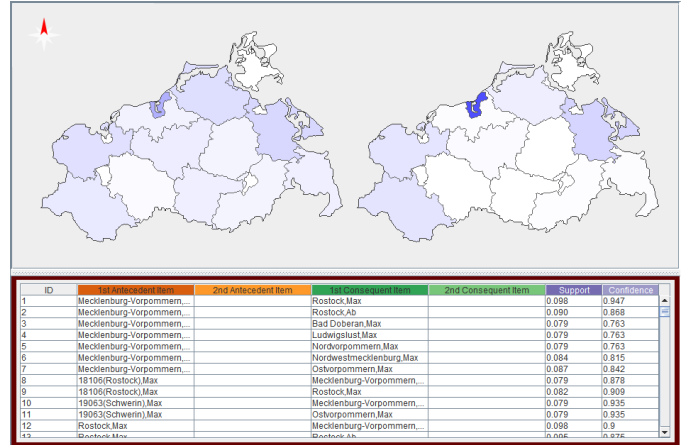


Figure 8: Color coding the frequency of regions in antecedents (left) and consequents (right) of association rules.

left side shows  $n$  map thumbnails, each of which highlights a region of the antecedent, and the right side shows  $m$  map thumbnails highlighting regions of the consequent. Different hues are used to indicate THRES, MIN, and so forth. The length of the central arrow encodes the rule's support, and the saturation reflects the confidence associated with the rule. For both arrangements, sorting the small multiples according to significance measures offers additional assistance to the user.

In addition to just sorting according to rule significance, we can also use color-coding to emphasize regions that are particularly relevant with regard to their frequency in itemsets. Figure 8 shows the color coded frequency of the regions of the antecedent (left) and of the consequent (right). One fact that can be derived from the figure is that the city of Rostock (saturated blue region in the center) occurs more frequently in the consequent than any other region. Another derivable fact is that the island of Rügen (top-right-most region) rarely occurs in rules at all. Such insights can be quite useful for adjusting the construction of qualitative literals from the data. The color-coding could also be extended to find correlations. For example, when marking one region, all rules that include this region as antecedent could be selected automatically and the regions belonging to consequents of this antecedent could visualize the rules' confidence by varying color saturation.

Association analysis is a well-established method for knowledge extraction. Doubtless, it makes sense to consider this method in a geo-visual analytics framework as well. Our concept shows that, in addition to the analytical extraction of association rules, the interactive definition of appropriate literals as well as the visualization of rules and corresponding significance measures have to be considered. We have developed a very first approach in this direction and believe that it is worth conducting further research on this topic.

## 7. Conclusion and Future Work

In this article, we discussed several aspects of geo-visual analytics. From talks and discussions with prospective users of visual analytics we recognized the need for clarification of the process of generating interactive visual representations of data. Therefore, we started with a conceptual view. We used Chi's data state reference model for this purpose because it describes data states and operators in a generic way. The set of operators is conceptually not limited, and thus any functionality of visual analytics frameworks can be smoothly related to this model. Moreover, the operators can be connected arbitrarily, and in this way a tight connection of analytical and visual methods can be achieved encompassed by a common interaction layer.

We introduced a number of novel concepts (operators of the conceptual framework if you will) that are related to analysis, visualization, and interaction. The extended focus+context concept is quite useful when larger geo-spatial datasets have to be explored. It combines a visualization strategy with interactive and automatic means to steer the foci and contexts. We also addressed a new visualization challenge, namely the integration of structural information (hierarchical structures in our case) with a map display. For this purpose we combined the extended focus+context with the Magic Eye View technique. Finally, we shed some light on association rule analysis as an alternative to the most often used clustering. Association rule analysis has not received much attention in previous work on geo-visual analytics. With our work we hope to have made a first step that illustrates the potential of this novel combination of analytical and visual means.

For all techniques described here we see that a whole ensemble of tightly connected operators is involved. While each operator has its specific functionality, the ensemble always follows the same goal: "support analytical reasoning facilitated by interactive visual interfaces" [21].

What will be useful for the future development is to investigate new combinations of operators and to increase reusability and exchangeability of operators – conceptually, but also implementation-wise. We have seen that the interactive focus+context concept can be combined with a hierarchy visualization to enable the exploration and analysis of hierarchically structured data in a geo-analytics scenario. But it is not yet possible to simply plug our concept into other frameworks and tools. The same holds for analytical operators. While existing frameworks are often based on clustering, there are other analytical means such as association analysis, principal component analysis, trend analysis, and others. But it is not yet clear to us how to combine these methods flexibly in order to assist analysts in extracting meaningful information from large and heterogenous geo-referenced data. Finally, considering alternative analysis methods or novel combinations of them will lead to a need for new visual representations and new ways for interacting with them. Therefore, it makes

sense to investigate concepts that allow us to flexibly plug in and combine diverse operators much like in the spirit of Chi's model.

## Acknowledgements

We gratefully acknowledge conceptual and implementation contributions by Lars Kornelsen, Matthias Kreuseler, Arne Klaassen, Thomas Nocke, Uta Rennau, and Petra Schulze-Wollgast. This work has been partly conducted in the context of the EU coordination project "VisMaster".

## References

- [1] Rakesh Agrawal, Tomasz Imieliński, and Arun Swami. Mining Association Rules Between Sets of Items in Large Databases. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*. ACM, 1993.
- [2] G. Andrienko, N. Andrienko, S. Rinzivillo, M. Nanni, D. Pedreschi, and F. Giannotti. Interactive Visual Clustering of Large Collections of Trajectories. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST)*, pages 3–10, 2009.
- [3] Gennady Andrienko and Natalia Andrienko. A General Framework for Using Aggregation in Visual Exploration of Movement Data. *The Cartographic Journal*, 47(1):22–40, 2010.
- [4] N. Andrienko and G. Andrienko. *Exploratory Analysis of Spatial and Temporal Data*. Springer, Berlin, Germany, 2006.
- [5] Christoph Buchheim, Michael Jünger, and Sebastian Leipert. Improving Walker's Algorithm to Run in Linear Time. In *Proceedings of the International Symposium on Graph Drawing (GD)*, pages 344–353. Springer, 2002.
- [6] Ed H. Chi. A Taxonomy of Visualization Techniques Using the Data State Reference Model. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis)*, pages 69–76, Washington, DC, USA, 2000. IEEE Computer Society.
- [7] Andy Cockburn, Amy Karlson, and Benjamin B. Bederson. A Review of Overview+Detail, Zooming, and Focus+Context Interfaces. *ACM Computing Surveys*, 41:2:1–2:31, January 2009.
- [8] Danny Dorling, Anna Barford, and Mark Newman. Worldmapper: The World as You've Never Seen it Before. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):757–764, 2006.
- [9] Niklas Elmqvist and Jean-Daniel Fekete. Hierarchical Aggregation for Information Visualization: Overview, Techniques, and Design Guidelines. *IEEE Transactions on Visualization and Computer Graphics*, 16(3):439–454, 2010.
- [10] G. W. Furnas. Generalized Fisheye Views. In *CHI '86: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 16–23, New York, NY, USA, 1986. ACM.
- [11] Steffen Hadlak, Christian Tominski, and Heidrun Schumann. Visualization of Attributed Hierarchical Structures in a Spatio-Temporal Context. *International Journal of Geographical Information Science*, 24(10), 2010.
- [12] Ivan Herman, Guy Melançon, and M. Scott Marshall. Graph Visualization and Navigation in Information Visualization: A Survey. *IEEE Transactions on Visualization and Computer Graphics*, 6(1), 2000.
- [13] Jochen Hipp, Ulrich Güntzer, and Gholamreza Nakhaeizadeh. Algorithms for Association Rule Mining – A General Survey and Comparison. *SIGKDD Exploration Newsletter*, 2(1):58–64, 2000.
- [14] Mikael Jern. Collaborative Web-Enabled Geoanalytics Applied to OECD Regional Data. In *Proceedings of the 6th International Conference on Cooperative Design, Visualization, and Engineering (CDVE)*, pages 32–43, Berlin, Heidelberg, 2009. Springer.

- [15] Daniel Keim, Jörn Kohlhammer, Geoffrey Ellis, and Florian Mansmann, editors. *Mastering The Information Age – Solving Problems with Visual Analytics*. Eurographics Association, 2010.
- [16] E. L. Koua and M.-J. Kraak. An Integrated Exploratory Geovisualization Environment Based on Self-Organizing Map. In P. Agarwal and A. Skupin, editors, *Self-Organising Maps: Applications in Geographic Information Science*. John Wiley & Sons, 2008.
- [17] Menno-Jan Kraak and Ferjan Ormeling. *Cartography: Visualization of Spatial Data*. Longman Singapore Publishers, Singapore, 1996.
- [18] Matthias Kreuseler and Heidrun Schumann. Information Visualization Using a New Focus+Context Technique in Combination With Dynamic Clustering of Information Space. In *Proceedings of the Workshop on New Paradigms in Information Visualization and Manipulation (NPIVM)*, pages 1–5, New York, NY, USA, 1999. ACM.
- [19] Ying K. Leung and Mark D. Apperley. A Review and Taxonomy of Distortion-Oriented Presentation Techniques. *ACM Transactions on Computer-Human Interaction*, 1(2):126–160, 1994.
- [20] Anthony C. Robinson. A Design Framework for Exploratory Geovisualization in Epidemiology. *Information Visualization*, 6(3):197–214, 2007.
- [21] J. J. Thomas and K. A. Cook. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE Press, 2005.
- [22] Christian Tominski, Petra Schulze-Wollgast, and Heidrun Schumann. 3D Information Visualization for Time Dependent Data on Maps. In *Proceedings of the International Conference Information Visualisation (IV)*, pages 175–181. IEEE Computer Society, 2005.
- [23] Christian Tominski, Petra Schulze-Wollgast, and Heidrun Schumann. Visual Methods for Analyzing Human Health Data. In Nilmini Wickramasinghe and Eliezer Geisler, editors, *Encyclopedia of Healthcare Information Systems*, pages 1357–1364. Information Science Reference, 2008.
- [24] Edward R. Tufte. *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, CT, 1983.
- [25] Jarke J. Van Wijk and Edward R. Van Selow. Cluster and Calendar Based Visualization of Time Series Data. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis)*, pages 4–9, Los Alamitos, CA, USA, 1999.