A Visualization Approach for Cross-level Exploration of Spatiotemporal Data

Hans-Jörg Schulz
University of Rostock
Rostock, Germany
hjschulz@informatik.uni-rostock.de

Steffen Hadlak
University of Rostock
Rostock, Germany
hadlak@informatik.uni-rostock.de

Heidrun Schumann
University of Rostock
Rostock, Germany
schumann@informatik.uni-rostock.de

ABSTRACT

Spatiotemporal data often relates to different levels of granularity in space, time, and data. Yet, bringing these levels together for an integrated visual exploration across levels poses a challenge up to this day. With this paper, we aim to provide a first solution approach to this challenge, which decomposes the data in its various levels to be able to show them side-by-side. Based on this decomposition, we derive a visual exploration approach that consists of a novel multi-level visualization, adjoined traditional spatial and temporal views, as well as of tailored exploration techniques for their concerted use. We exemplify the utility of this approach by case studies on election and poll data from Germany's various administrative levels and different time spans.

Categories and Subject Descriptors

H.5.m [**Information Interfaces and Presentation**]: Miscellaneous; I.3.8 [**Computer Graphics**]: Applications

General Terms

Design

Keywords

multilevel visualization, exploratory data analysis, multiresolution analysis, spatiotemporal visualization

1. INTRODUCTION

Spatiotemporal phenomena can be observed at various levels of space (e.g., local, regional, global) and of time (e.g., hourly, daily, monthly). It is known that a chosen spatiotemporal frame of observation influences the level on which phenomena are observed: the price fluctuations of a single currency over the course of a day occur on a different level than the incline or decline of a country's standing in an annual ranking of world economies at large. So, unless the spatiotemporal frame of observation remains fixed, data is generated on multiple levels. Combined, these levels form hierarchies of different granularities in space, time, and data.

Since phenomena on different spatial and temporal levels are not isolated, the analysis of complex behavior requires to address several levels of granularity simultaneously. The Visual Analytics roadmap [9, p.82] states explicitly that therein lies a current research challenge, because such an analysis spanning multiple levels is not yet supported by visual analysis concepts for spatiotemporal data that traditionally focus on individual levels only. The state of the art in flat 2D map-based representations, as well as in 3D representations based on the space-time-cube [11] is to allow for an interactive selection of a desired granularity level at which to analyze space, time, and data. Current software tools, such as the Geospatial Visual Analytics Toolkit [2], reflect this state of the art. The need to investigate multiple levels side-by-side has furthermore been recognized in other domains, such as the field of biomedicine [21, 22]. Yet despite its need, it is so far not sufficiently supported.

Building on first ideas that we presented previously [18], this paper addresses this challenge by contributing a visualization approach that exposes different levels to the user for their simultaneous analysis. To this end, we propose a decoupling of the three aspects of space, time, and data, as well as of their levels of granularity. This way, the levels can be shown and explored side-by-side to facilitate their combined analysis. Their synchronization makes explicit for which combinations of levels data exists and how it compares to data from other level combinations. We provide traditional and novel forms of navigation to interactively seek out and investigate cross-level behavior of interest.

We exemplify our approach with case studies from the political domain, featuring various election data and voter sentiment data from polls for three administrative levels of Germany on three temporal levels. Apart from serving as a running example throughout this paper, we reproduce some of the findings we made in the data by using our visual analysis approach at the end of this paper.

2. SPATIOTEMPORAL MULTILEVEL DATA

The main idea of our visual analysis approach across multiple levels is already embodied in our multilevel data model. In accordance with the well-known Triad Representational Framework [16] of WHERE (space), WHEN (time), and WHAT (data), spatiotemporal data items can be regarded as a set of triples (S, T, v) with S being a spatial reference, T being a temporal reference, and v being a singular data value. The

references S and T can refer to multiple levels of granularity in space $s_1, s_2, ..., s_n$ and in time $t_1, t_2, ..., t_m$, e.g., $S = \{country\ code, zip\ code\}$ and $T = \{year, month, day\}$. For the same references S and T, multiple *data facets* $v_1, v_2, ..., v_f$ may be defined. These data facets are not confined to the spatiotemporal or even multivariate data models. They can also be multi-run (stem from different computations), multi-model (have been gathered w.r.t. different frames of reference, such as different coordinate systems), or even multi-modal (stem from different sources) [8]. As such, the data facets span a much broader range of information sources than comparable existing concepts, such as the "facets" known from the field of *faceted search* [24] or the "entity types" known from Jigsaw [19]. In consequence, multiple tuples with the same reference S and T can exist, each of them with a different data value v drawn from one of the available data facets. For the remainder of this paper, we only distinguish between the different domains (space, time, data) where necessary and otherwise subsume levels of granularity and facets under the general term "levels". To distinguish between the levels and the concrete elements they contain, we call these elements *instances*, regardless of them being spatial values (e.g., zip codes or GPS coordinates), temporal values (e.g., years or days of the week), or data values (e.g., parties or candidates).

We take this tiered notion of multi-faceted spatiotemporal data as a basis to separate each of the three aspects into the levels of granularity they exhibit. So, for example, we consider years, months, and days instead of a set of dates, as well as streets and cities instead of addresses, basically by decomposing S and T in their individual components s_i and t_j. Principally, this can be thought of as a flattening of all volume cells at all possible resolutions of the space-time-cube. The data levels are constructed from each v that is encountered for the same S and T, such as parties and candidates as in the above example. The different facets v_k can hold numerical data, categorical data, or nominal data. If v_k contains too many instances v, as it would be the case for a continuous numerical variable, its instances are binned into adequately, yet not necessarily uniform sized intervals as they suffice for a concrete analysis.

It is noteworthy, that for this separation to happen, it is not necessary that for all spatial and temporal levels, exactly one corresponding data level exists – it can also be multiple data levels or none at all. This information is encoded in the tuples, which we interpret as edges of a multipartite graph. The nodes of this graph are formed by the instances of each level, with each level forming one partition of the graph. As it is characteristic for multipartite graphs, there are no edges (tuples) that connect nodes of the same partition (level), but only run between different partitions. For example, no two years are directly connected with each other through an edge, but only with instances of places and data values, each forming an independent set. Our synchronization view and the visual analysis mechanisms built on top of it are grounded in this interpretation of the data.

3. VISUAL DESIGN

The above way of perceiving spatiotemporal multilevel data is reflected in a novel view, called *synchronization view*. It realizes the same decoupling of the levels of space, time, and data to be able to represent and navigate them simultaneously. While it provides a way to access the data and
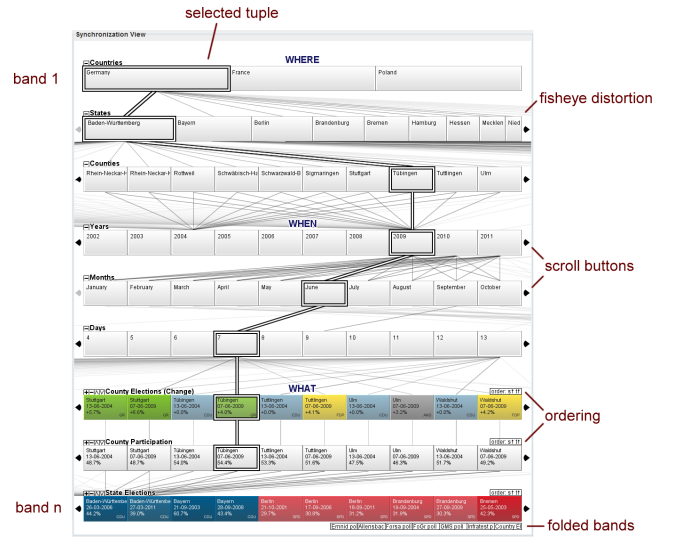

Figure 1: The synchronization view showing multiple levels of space, time, and data. Each level is represented by a horizontal band of all instances of that level. We use the hue of the colors to encode categorical values, such as parties, and we reflect quantitative values, such as percentages, through their saturation. Data tuples are shown as edges connecting instances on all levels a tuple contains. Scroll buttons and fisheye-like distortion allow the user to navigate and explore each level.

get an overview of their multiple levels, adjoined *data views* are needed to show the data in their spatial and temporal context. Hence, we provide a map view and a calendar view alongside the synchronization view. Each of these components of our visual design – the synchronization view and the data views – are introduced and discussed in the following.

3.1 Synchronization view

Typically, one does not know where in space and when in time, which data source exhibits behavior of interest. Searching through all possible level combinations in space, time, and data is tedious and time consuming. Therefore, our synchronization view shows all levels of space, time, and data simultaneously. To this end, it reflects the multipartite model by aligning all instances of each level in a horizontal *band*. To distinguish between different instances in a band, appropriate labeling and color-coding is applied. The bands are laid out in parallel and the tuples are shown as edges connecting the instances from all bands a tuple contains. This basic setup is shown in Figure 1. It is this connection via tuples that realizes the actual synchronization, as levels are shown simultaneously, even though they may never occur together in a tuple. This is made explicit through the linking: iff a link is connecting two instances, they co-occur. This way, yearly data is only linked to the level of years, while monthly data is connected to both, months and years, as a month must be part of some year as well.

Constructing a view from the proposed flattening of the space-time-cube poses a scalability challenge, if the number of levels or the number of instances per level grow large. We solve the former by allowing the user to interactively

adapt the number of shown bands by folding those of lesser interest and by duplicating others of higher interest. For example, in an investigation on a monthly granularity, bands of finer temporal scales (weeks, days, etc.) and bands in space and data that relate only to those finer temporal scales can be folded. Yet, a band that relates to many other bands can be duplicated to see its direct relation to more than two neighboring bands. For the case of a large number of instances, the user can scroll horizontally through the bands and apply a fisheye distortion, to maximize only instances that are currently explored and minimize all others. This is illustrated in Figure 1 for the spatial level of states.

With this view, we gain a first approach for an overview of all data on all levels of granularity. It relieves the user from having to choose an appropriate level of granularity to investigate the data on, before actually seeing the possibly unknown data. And even if the user already knew the data and which levels to look at, our approach avoids the need to switch back and forth between them in case of a comparative investigation involving data on two or more levels.

To achieve this, we paid a price, though. The separation of levels has dissolved the common spatial and temporal contexts in which a data item would usually be embedded. This could be a map or a globe for the spatial context, as well as a calendar or timetable for the temporal context. While the bands contain all the data, they do not show it in their spatiotemporal frame of reference. Therefore, we add these as linked data views to complement our synchronization view.

3.2 Data views

The data views aim to present the data in the more conventional ways the user is accustomed to. They address the spatial and the temporal domain independently by supplying a map view [13] and a calendar view [25] to capture them. These provide the contexts in which the values are then represented – usually by color-coding them for the regions of the map or the time slots in the calendar.

The implications of providing these views are apparent: they cannot show all the data, as the synchronization view can. Hence, the user has to choose a spatial level of resolution to show in the map view, a temporal level of resolution to show in the calendar view, and a data level to encode in both views. Since these choices can be made from within the synchronization view upon inspecting all levels, they are now informed and no longer based on trial-and-error. We go even further than selecting a flat level by permitting to choose a refinement of levels around a particular data item of interest. In this case, the map and/or the calendar are shown at a coarser level and only the neighborhood of a data item in question is then shown at a step-wise higher resolution – e.g., the calendar is shown on a yearly granularity level and only around a particular date, its year is shown on a monthly level, its month is shown on a weekly level, and its week is shown on a daily level. This is illustrated in Figure 2, which shows the map view and the calendar view with such a refinement of levels around a particular point in space and time. This way, we support exploring data items of interest simultaneously on different levels of granularity in space and time in the data views as well.

Both data views observe hierarchical dependencies between the levels. For example, spatial levels often exhibit a hierarchical dependency expressing the inclusion of administrative regions – i.e., countries, states, counties. Whereas
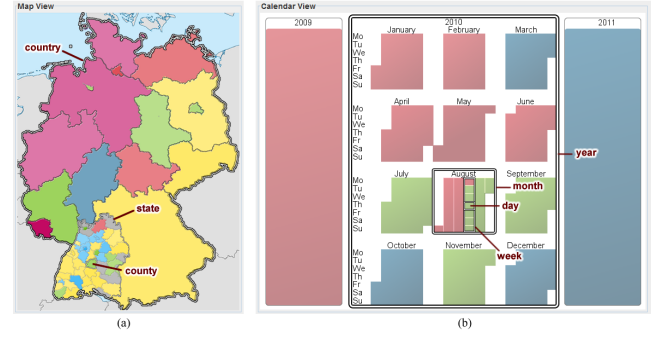

Figure 2: The data views show the data in more conventional ways. Here, synthetic data is color-coded in the views. A map view (a) addresses the spatial domain and a calendar view (b) addresses the temporal domain. Both views exemplify the refinement of levels around a particular data item.

levels that organize space in grid squares would be independent of these administrative levels, but form a hierarchical dependency among themselves. In case of such a hierarchical order among spatial and/or temporal levels, it depends very much on the semantics of the data if and how they can be projected onto other levels. This applies to all three of such possible projections:

- **Aggregation** of low level data to higher levels can be done iff an appropriate aggregation function is given. This is important as aggregating values is highly domain-dependent with many different methods in existence [4].

- **Duplication** of high level data at lower levels is possible iff the data given for the whole also applies to its parts. For example, this is the case for country-wide election results that hold true for all parts of the country and thus also for all states and counties.

- **Registration** of data between independent levels is only possible iff adequate mapping functions are given. For example, to map zip code level data to the county level and vice versa, one must know how much each zip code area contributes to each county, because at least in Germany, they do not strictly nest.

To prevent misinterpretation, the map or calendar in the view is by default colored gray to mark an invalid choice of levels if no data is given for their particular combination. If one wants to project data from one level onto another, our data views are conceptually able to do this by color-coding the projected data instead of graying out the level. In this case, one has to make the user aware of the projection, so that the projected data is not mistaken as actually collected data. This can happen easily, for example, when looking at higher level data (e.g., country level election results) on lower levels (e.g., county level) one may find that all counties appear to have voted the same. Yet, in fact merely the overall voting result of the whole country has been duplicated and color-coded in each individual county.

The synchronization view and the data views are mutually linked, so that interactive selections and adaptations in one view are reflected in the others as well. This enables their concerted use for an exploratory analysis of multilevel data, as it is further facilitated by the interaction concept presented in the following section.

4. INTERACTIVE EXPLORATION

The newly introduced synchronization view is built so that it effectively reflects our particular data model. It provides per-level visualization by aligning the bands and per-tuple visualization by connecting them accordingly. As such, it extends well beyond the data views that can only provide traditional per-level representations. Having the versatility of the data model embedded in our visualization technique, this section discusses the different modes of exploratory analysis that can be pursued by using this view. As a result of this discussion, we provide a tailored exploration mechanism by utilizing a combination of two orthogonal modes of interaction: the novel *tuple-based exploration* together with the common *level-based exploration*.

4.1 Modes of exploratory analysis

In the spirit of Bertin's Levels of Information [5], it is common in the context of spatiotemporal data to define the analysis interest with respect to its extent:

- **point-based extent**, e.g., for a particular point in time and space the associated value instance is sought;

- **local extent**, e.g., for a contiguous subset of points in time and space the development of value instances (temporal decline, spatial spread, etc.) is of interest;

- **global extent**, e.g., for all points in time and in space the overall behavior (distribution, global extremes, etc.) of value instances is to be determined.

In case of multiple granularities, a "point in time and space" specifies value instances on a variety of levels, as a date specifies a year, a month, and a day and an address specifies a country, a state, a county, a city, a zip code, etc. Given multiple data levels, each combination of temporal and spatial scale can in addition yield multiple value instances.

Standard methods, such as maps or bar charts, can cope with this multitude of data only partially: either by reducing the number of scales (usually by selecting a few scales of interest) or by reducing the number of data items to show per scale (usually by summarizing the value instances of entire scales with a few statistical measures). Examples for the former are map views and calendar views, as we provide them as well. They permit a user to explore the data only at a single level (local data extent) – either for one point in time (point-based temporal extent) and at a single spatial level (local spatial extent) in the map view, or for one point in space (point-based spatial extent) and at a single temporal level (local temporal extent) in the calendar view. The latter, the reduction of data items, is for example done by computing means, quartiles, and maximal values and displaying them in one box-whisker-plot per scale [6]. This way, all scales can be shown in their global extent, but at the cost of losing the details of the individual value instances and showing only abstracted summary statistics. So, in each case the standard methods omit parts of the data (scales, value instances) to be able to show the rest. An access to all individual value instances on all scales (global extent without abstraction) is not possible from within these visualizations.

In contrast, our synchronization view can be used for such a global access without abstracting individual values. As such, it enables a user to follow two modes of exploration that cannot be pursued with standard techniques alone:

1. It provides a global view of all levels. Based on this full view, it permits an **exploration of levels** by making an informed selection of levels of interest to be shown in the adjoined traditional views (map and calendar).

2. It presents a global view of all tuples. By this, it is possible to **explore the tuples** across all scales directly in the synchronization view.

The following two sections describe these two modes of exploration and the particular interaction needed for them.

4.2 Level-based exploration

The level-based exploration steers the selection of the individually shown levels in adjoined map and calendar views directly from within the synchronization view. In terms of interaction, it comprises of the conventional browsing of multilevel data level by level [7]. This can be thought of as cutting "horizontally" along specified levels through the set of tuples to yield triples that contain only one spatial level, one temporal level, and one data level. In contrast to existing solutions, this is not a mere slider or mouse-wheel interaction performed on top of the data views, but a navigation of levels in the synchronization view that automatically updates the adjoined views through linking mechanisms. The difference is the informed choice of a level, as one does not have to go through all possible level combinations one by one to see whether they exhibit behavior of interest. Instead, the synchronization view shows all levels simultaneously and the user can pick level combinations of interest directly based on what he discovers in this overview. This is not only a convenient addition to the data views, but mutually benefitting to both data and synchronization views. The reason is that the power of the synchronization view of showing all levels comes for the price of a very compact representation that aligns all levels as 1-dimensional horizontal bands. Yet, these bands may not fully capture more complex behavior, such as the spatial or temporal spread of a certain pattern, which may be identified in the bands, but can more easily be verified in the data views.

4.3 Tuple-based exploration

The tuple-based exploration presents a new way of exploring multilevel data: tuple by tuple. It can be thought of as the orthogonal counterpart to the level-based exploration, as it cuts "vertically" along a specified tuple through the set of levels. This exploration mode is motivated by the fact that multilevel patterns are not to be found on individual levels. Or as it has been observed and formulated in the mid-nineties by Ahl and Allen [1, p.76]: "What makes levels interesting is the relationship between them." This relationship is encoded in the tuples that connect the levels in our multipartite data model. To find out all there is to know about this relationship, one must explore all tuples. Unfortunately, often the sheer number of tuples makes this a challenging undertaking, as they are too many to display them all at once without creating clutter and also to browse through all of them one by one. This is the downside of providing a detailed global view and we aim to solve this issue by two mechanisms: *sorting* and *pinning*.

The basic assumption that underlies both is that the user is not just looking at the tuples in general, but has instead a partial idea of what interests him. This can be either a *relative interest* (e.g., all tuples exhibiting a high voter turnout

in recent years) or an *absolute interest* (e.g., all tuples relating to a specific party in the years 2004 and 2005).

The **relative interest** can be translated into a partial **sorting** order for tuples. If the user looks for multiple criteria, these have to be prioritized to express in which the user is foremost interested – e.g., the user wants to sort first for high voter turnouts and only if values coincide, the more recent shall precede earlier tuples. Once sorted according to a user's criteria, the browsing would start with tuples of highest interest to the user – in the example, this would be the tuple with the highest voter turnout, from the most recent year if there were multiple tuples with the same turnout. This way, it is unlikely that the user will have to browse all tuples to explore the relations of the levels with respect to his particular interest. It is obvious that this method stands and falls with the availability of suitable sorting methods. While sorting is trivial for numerical levels and ordinal levels, in particular the sorting of spatial levels and nominal data levels is challenging. Usually, different application domains employ different orderings that are targeted towards their specific exploration tasks. This makes it hard to provide a general ordering strategy. So, we acknowledge that other application scenarios may require other ordering strategies, and we assume a simple north-east to south-west ordering for the geospatial domain and an alphabetical ordering for nominal data values in the context of this conceptual discussion. In the synchronization view, we show any sorting order (e.g., temporal first, then by value) and its direction (ascending, descending) in a small box at the top-right side of each band. Once, such an intra-level ordering has been determined, it is conceptually also possible to automatically determine an inter-level order that helps to show patterns among tuples more clearly [14, 15].

The **absolute interest** can be translated into fixing or, as we call it, **pinning** certain instances on some levels. In the above example, the specified party on the data level "party" and the years 2004 and 2005 on the temporal level "year" get pinned. This results in a filtering of tuples to only those that contain these instances and thus cutting down on the large number of tuples in a data set. The browsing of tuples would then only encompass those that run through these particular instances. Furthermore, pinning works also on the data levels, so that data instances of interest can be pinned to narrow down the number of tuples to those of all times and places that instance occurred. If the number of tuples is still too large, pinning can of course be coupled with sorting to impose an additional order on the tuples.

To reflect the tuple-based exploration in the adjoined data views that show only selected temporal and spatial levels, we utilize the aforementioned refinement of levels around a data item of interest to mimic showing the tuple across multiple levels in the otherwise flat display. The views are linked, so that the browsing of tuples in the synchronization view will update the adjoined views by moving the refined region around to reflect the currently viewed tuple. On the other hand, a simple click on a region in the map view or on a day, month, or year in the calendar view will trigger a pinning of the clicked instance on the appropriate spatial or temporal level in the synchronization view.

The interplay of the views and their use together with the described interaction concept are brought to life in the following section, which presents our realization of them.

5. IMPLEMENTATION & CASE STUDIES

The realization of such a novel visualization concept, which deviates in many aspects from established representations in the field of geovisualization, cannot be done without user feedback along the way. Hence, we included users throughout the process of developing our approach, as this is an established way in visualization design to prevent from building ad-hoc solutions [12, 23]. When involving users and gathering their feedback, it is important to do so in the context of data they can relate to. Since we wanted to include the opinions of a diverse group of people and prevent tailoring our prototype to a specific application domain, we chose a number of different, yet related data sets that most people have some basic knowledge about – various election results and voter polls from Germany. In the following, we describe the design of our implementation, give a description of the data used, and reproduce some cross-level findings that were made in the data while test-driving our visualization.

5.1 Implementation based on user feedback

Our implementation relies on Java 6 with the standard Java2D functionality for the synchronization view and the calendar, as well as on the Apache™ Batik SVG Toolkit for the rendering of the map view. It implements the multipartite data model and transforms input data to this model by partitioning it into levels and instances. The exploration strategies make heavy use of this data model for their fast interactive realization. For example, it would be very cumbersome to implement a pinning on the original data items (e.g., on full dates or addresses).

From the early stages on, eight users from different departments including a domain expert from the political sciences gave iterative feedback on our realization. Four out of these eight users had some prior exposure to information visualization. In short sessions, we let them use our visualization in various configurations and setups. Afterwards, we led structured interviews with them to evaluate their experiences with the prototype. To rule out learning effects, we slightly distorted the real data by a random process, so that the data was still reasonably realistic, yet no participant could thus claim prior knowledge of it. Besides the free exploration of the data, we had two recurring tasks to be solved by each participant with each presented setup: a generic overview task and a specific localization task. The two main aspects that we investigated with their help were the view setup (i.e., whether additional spatial and temporal views are needed and whether they should be simultaneously visible or be shown one by one) and the tuple-based exploration (i.e., how the ordering affects the utility of the tuple-based exploration and which orderings to provide).

With respect to the first design question, one user summarized the feelings of most participants as he does not "want to miss any of these views as they all have their benefits." This underlines that besides our synchronization view, both data views are necessary as a simultaneous display of the missing spatial and temporal contexts. Having then been given a simultaneous display of both types of views, many participants highlighted the importance of the synchronization view as a central element of the view setup, despite having to master the learning curve attached to such a novel view. One participant stated that "the synchronization view provides a good way to coordinate the analysis." In this way, the back and forth between the views seems to be a powerful

feature, as it allows for switching seamlessly between a more general overview and more detailed views of the data.

The benefit of the ordering of bands in the synchronization view was also highlighted, even if it only vaguely reflected a user's partial interest. The reason is that because an ordering forms blocks of similar tuples, they can thus easily be skipped in bulk when browsing them. So even if the tuples, which are the most interesting to a user, are not sorted to the very top, the user can fast-forward through the ones of lesser interest and quickly reach them. The possibility to freely switch the ordering at any time was received very well by the participants: "Being able to change the order in which the data is represented makes it very easy to find extraordinary features in the data." Our domain expert valued the idea of scrolling through the tuples in a particular order and thus seeing the relations between the different aspects of the data within a single view: "Normally, we can only see a couple of data aspects, but never all of the data we have gathered. Their interplay is then often only captured within statistical evaluations of the data. Yet, having the possibility to really see this interplay together with the data within a single visualization is a very promising feature."

In addition to the user feedback, we also collected design insights from projects having similar aims, such as the SO-LAP interface by Beaulieu and Bédard [3] or the early use of Polaris for exploring hierarchical data [20]. Altogether, this has helped us to put our conceptual design and our concrete implementation design on a broader basis.

5.2 Data description

The used data for the user-driven design, for all examples and screenshots in this paper, as well as for the following case studies is based on various data sources. It contains German election data compiled from different online sources, such as the statistical offices of the states, as well as poll data from different institutes taken from the website http://www.wahlrecht.de/umfragen and reaching back as far as 1998. Overall, this data consists of 16 levels structured in three hierarchical time levels (years, months, days), three hierarchical space levels (countries, states, counties), and ten data levels covering different results and aspects of elections. This data is comprised of the percentage achieved by each party in elections and polls, and its delta of percentage points as compared to the last election/poll. Additionally, we have included the voter participation numbers for elections where they were available. Election results are available on country, state, and county level, while polls are only available on country level and for a few selected states at selected time points – e.g., a week before a state election.

Elections and polls are conducted on a specific date, which is encoded as a temporal reference in their tuples. Yet their results remain in effect until the next election or poll yields a new outcome – i.e., a country is governed by the leading party/coalition until a new election is held, and journalists refer to the most recent poll until a new poll is published. Hence, we chose to extend the data from the actual day they were gathered to the entire time interval until the next election/poll comes into effect. When a tuple is selected in the synchronization view, all other data instances are connected to it, if its date lies within the interval in which they are valid. For data that is truly given for specific dates, their extension into intervals and their connection would not be performed and the tuples are simply shown as they are.
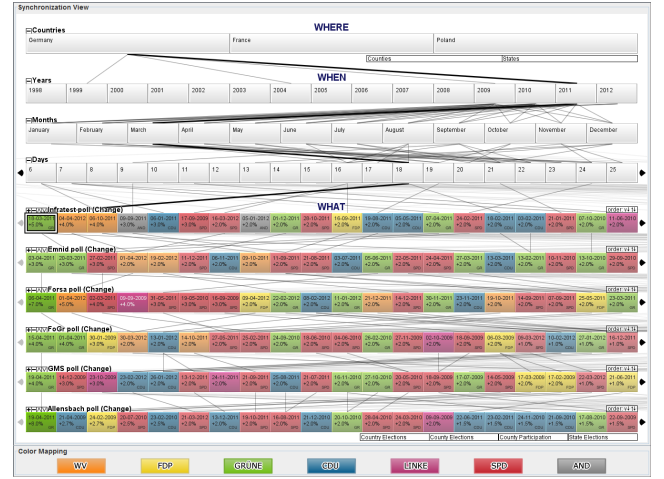

Figure 3: Investigating the largest increase in voter popularity for each poll in the synchronization view. All data levels (all polls) are sorted to show the instances of largest increase gained by a party on the left side. A clear pattern emerges there: across all polls, the Green party (green) has the largest increases among all instances. When looking at the dates of these polls, one can notice that they are from shortly after the Fukushima disaster.

5.3 Case studies

Through the exploration of this election data, we revealed two interesting patterns of voter behavior, which would be cumbersome to find without our approach. These patterns relate to two events in recent years: the Fukushima Daiichi nuclear disaster (temporal effects) and the Stuttgart21 railway project controversy (temporal and spatial effects).

The first example concerns the change of voter sentiment and is depicted in Figure 3. The shown synchronization view contains monthly or weekly poll data collected by different survey institutes. As each data level is derived from a different data source, it constitutes an instance of multi-modal data. All data levels relate to the country level only, which is why all other spatial levels have been folded away. The data levels encode the party with the highest increase in voter popularity since the last poll. They are sorted in descending order of percentage points of each increase. If instances coincide, these are sorted in descending order of the time points they are associated with. From this sorting, a clear pattern emerges on the left side of the synchronization view. It shows the biggest increase for the Green party throughout all the different polls, with a huge jump of up to 8 percentage points in the Allensbach poll (bottom band). Pinning individual value instances reveals that these polls were all taken shortly after the Fukushima disaster on March 11, 2011.

To confirm this finding, we re-order the data bands with respect to time only and we pin the year, month, and day to the date of the Fukushima disaster. According to the described extension to intervals, the pinned date connects to all prior polls, as March 11, 2011 falls into the respective intervals for which these poll results are considered valid. The result of this adaptation can be seen in Figure 4. The calendar view shows the increase in voter popularity from the weekly poll results of the first band, the Infratest poll. One can observe in the synchronization view that the dis-
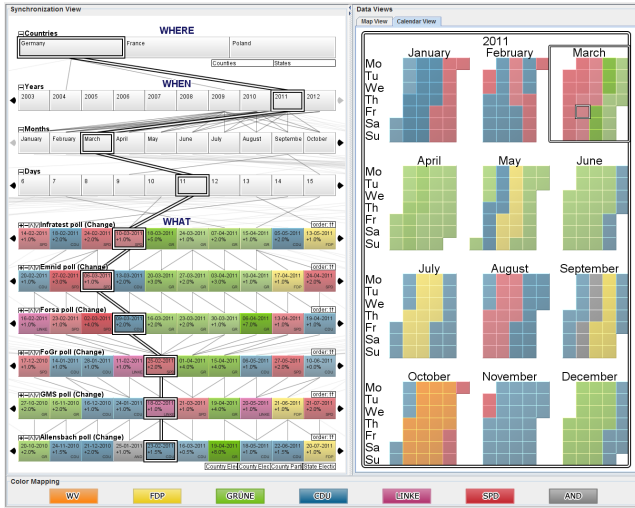
Figure 4: Effect of the March 11, 2011 Fukushima Daiichi nuclear disaster on German voter behavior. The synchronization view shows a strong increase in popularity for the oppositional Green party (green) in polls following the disaster. This trend was only broken after the ruling parties themselves announced to end nuclear energy production in Germany. The calendar view shows this period of time (4 weeks) for the weekly Infratest poll from the first band in the temporal context of 2011 at large.

Figure 5: Effect of the controversial governmental development project Stuttgart21 on local voter behavior in the German state Baden Württemberg. The synchronization view shows the gains of the Green party (green) in response to the project's start of construction. The map view indicates a similar trend in Stuttgart's neighboring counties where oppositional citizens' action committees (orange) won the elections in the same period of time.

aster took a few days to arrive in the political debate, as the Emnid poll (second data band from the top) still shows the largest increase for the governing party (blue) for March 13, 2011, two days after the disaster. But as the temporal sorting of the data levels already hints at and as browsing through the tuples with the tuple-based exploration confirms: the popularity of the Green party increases steadily over the next weeks. This was most certainly due to an increased support for their stance against nuclear power. After four weeks, the impact of the event became weaker, which can be explained with the fact that the ruling parties in Germany followed suit and decided to abandon nuclear power, so that the Green party lost its edge over them.

The second example is shown in Figure 5 and presents our findings regarding the controversial railway and urban development program Stuttgart21. Even though it concerned mainly the German state Baden Württemberg, it sparked debate and grassroots protests all across Germany. The county election in Stuttgart and the state election for all of Baden Württemberg held in 2009 and 2011, respectively, reflect the strong discontent of the local population with this project. This can be seen in the synchronization view, where we pinned the state and county, as well as the date of the 2009 county election in Stuttgart. The synchronization view shows a similar pattern to the first example: the winner of the county election held on that date was the Green party that opposed the project (first data band). This was not directly reflected in the state election two years later (second data band, the instance on the right side of the selected one) where the Christian Democratic Union (CDU, shown in blue) that traditionally governs the state received the most votes. Yet, its impact can be seen in the change of voter popularity (third data band, the instance on the right side
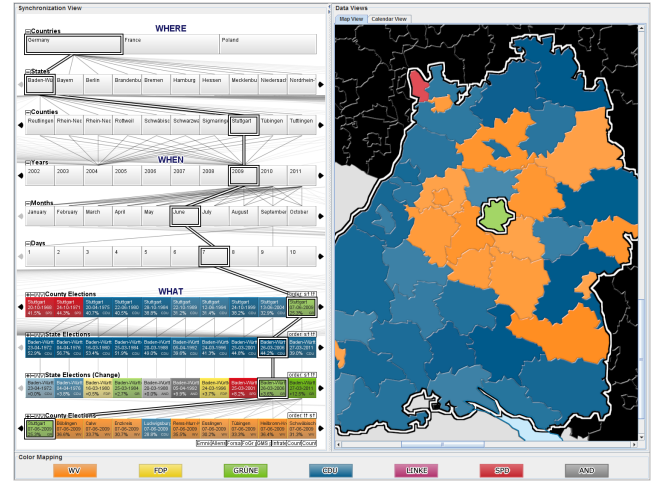
of the selected one) where the Green party had the largest increase of 12.5 percentage points. Together with a loss on the side of the CDU, this actually led to the situation that the CDU could not form a majority coalition. Instead, the Green party formed a coalition and is now heading the state.

Another interesting observation can be made in the remaining band (fourth data band), which is a duplicate of the first data band, but differently sorted. The first band was sorted first by space (alphabetical by county names) to bring together all of Stuttgart's elections and second by time to have these elections line-up in ascending order from earliest (left) to latest (right). Yet, the fourth band is sorted first by time to bring together all county elections of the same date and second by space (ascending by each county's distance to Stuttgart) to have these elections line-up in ascending order from those closest to Stuttgart (left) to those farthest from Stuttgart (right). It can be seen from the ordering of the fourth band that on the day of county elections in 2009, numerous counties close to Stuttgart have voted oppositional citizens' action committees (shown in orange) into office. While it can only be hinted at by the linear arrangement of the band, this observation can be verified in the map view. This effect is hardly coincidental, as it is very strong around Stuttgart and decreases with distance, so that it is very likely also a consequence of the Stuttgart21 project.

6. CONCLUSION AND FUTURE WORK

Discoveries and findings, which are based on a multitude of levels, are very difficult to make using traditional approaches alone. With our synchronization across space, time, and data levels, we provide a first handle on the problem of visually exploring spatiotemporal multilevel data. This is made possible by a versatile graph-based data model for capturing and aligning information on multiple levels. The visu-

alization built on top of this data model passes the model's versatility on to the user to use it for interactively exploring data on multiple levels simultaneously with tailored exploration approaches. It does so by flattening the dimensions of the space-time-cube into bands that only show value instances that actually occur and connecting them if they co-occur. This effectively removes the "empty" cells of the space-time-cube and only shows those that actually contain data. As a result, we yield a very compact overview visualization that permits a global exploration of multilevel data.

In future work, we aim to compact our synchronization view even further in order to lower the interaction cost for investigating the data in its entirety. This can principally be done in two ways: by reducing the number of levels and/or by reducing the number of tuples. Yet at the same time, we want to preserve all of the information necessary for the current exploration. Since no preconceived overall reduction can achieve this, we plan to tie the reduction to the interactive exploration modes and show all information related to a current level/tuple under investigation while hiding others. This would couple the reduction of levels with the level-based exploration, so that levels are automatically folded away if they do not co-occur in any tuple with the currently selected level. Likewise, the tuple-based exploration would be used to reduce the number of shown tuples – e.g., by bundling all tuples into ribbons that have no value instance in common with the currently selected tuple. The result would look similar to a Sankey Diagram [17] or a Parallel Sets visualization [10], but with the tuple currently under scrutiny and all related tuples being shown as individual lines. These future additions promise to even further improve the utility of our multilevel visualization approach.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] V. Ahl and T. Allen. *Hierarchy Theory: A Vision, Vocabulary, and Epistomology.* Columbia University Press, 1996.

[2] N. Andrienko and G. Andrienko. Visual analytics of movement: An overview of methods, tools, and procedures. *Information Visualization*, 12(1):3–24, 2013.

[3] V. Beaulieu and Y. Bedard. Interactive exploration of multi-granularity spatial and temporal datacubes: Providing computer-assisted geovisualization support. In *Proc. of GeoVA'08*, 2008.

[4] G. Beliakov, A. Pradera, and T. Calvo. *Aggregation Functions: A Guide for Practitioners.* Vol.221 of Studies in Fuzziness and Soft Computing. Springer, 2007.

[5] J. Bertin. *Graphics and Graphic Information Processing.* Walter de Gruyter, 1981. page 12.

[6] J. Dykes and C. Brunsdon. Geographically weighted visualization: Interactive graphics for scale-varying exploratory analysis. *IEEE TVCG*, 13(6):1161–1168, 2007.

[7] N. Elmqvist and J.-D. Fekete. Hierarchical aggregation for information visualization: Overview, techniques, and design guidelines. *IEEE TVCG*, 16(3):439–454, 2010.

[8] J. Kehrer and H. Hauser. Visualization and visual analysis of multi-faceted scientific data: A survey. *IEEE TVCG*, 19(3):495–513, 2013.

[9] D. Keim, J. Kohlhammer, G. Ellis, and F. Mansmann, editors. *Mastering the Information Age: Solving Problems with Visual Analytics.* Eurographics, 2010.

[10] R. Kosara, F. Bendix, and H. Hauser. Parallel Sets: Interactive exploration and visual analysis of categorical data. *IEEE TVCG*, 12(4):558–568, 2006.

[11] M.-J. Kraak. The space-time cube revisited from a geovisualization perspective. In *Proc. of ICC'03*. International Cartographic Association, 2003.

[12] D. Lloyd and J. Dykes. Human-centered approaches in geovisualization design: Investigating multiple methods through a long-term case study. *IEEE TVCG*, 17(12):2498–2507, 2011.

[13] A. M. Maceachren and M.-J. Kraak. Exploratory cartographic visualization: Advancing the agenda. *Computers & Geosciences*, 23(4):335–343, 1997.

[14] H. Makwana, S. Tanwani, and S. Jain. Axes re-ordering in parallel coordinate for pattern optimization. *International Journal of Computer Applications*, 40(13):43–48, 2012.

[15] W. Peng, M. Ward, and E. Rundensteiner. Clutter reduction in multi-dimensional data visualization using dimension reordering. In *Proc. of IEEE InfoVis'04*, pages 89–96. IEEE, 2004.

[16] D. Peuquet. *Representations of Space and Time.* The Guilford Press, 2002.

[17] P. Riehmann, M. Hanfler, and B. Fröhlich. Interactive Sankey diagrams. In *Proc. of IEEE InfoVis'05*, pages 233–240. IEEE, 2005.

[18] R. Rosenbaum, H.-J. Schulz, S. Hadlak, and H. Schumann. Visual analysis of spatiotemporal multilevel data. In *Proc. of GeoVA(t)'12*, 2012.

[19] J. Stasko, C. Görg, and Z. Liu. Jigsaw: Supporting investigative analysis through interactive visualization. *Information Visualization*, 7(2):118–132, 2008.

[20] C. Stolte, D. Tang, and P. Hanrahan. Query, analysis, and visualization of hierarchically structured data using Polaris. In *Proc. of ACM KDD'02*, pages 112–122. ACM, 2002.

[21] M. Streit, H.-J. Schulz, D. Schmalstieg, and H. Schumann. Towards multi-user multi-level interaction. In *Proc. of CoVIS'09*, pages 5–8, 2009.

[22] D. Testi, D. Giunchi, X. Planer, R. Cardenes, G. Clapworthy, N. McFarlane, S. R. Aylward, and R. Christie. Multiscale spatiotemporal visualisation. Technical report, 2012. White paper.

[23] M. Tory and T. Möller. Human factors in visualization research. *IEEE TVCG*, 10(1):72–84, 2004.

[24] D. Tunkelang. *Faceted Search.* No. 5 of Synthesis Lectures on Information Concepts, Retrieval, and Services. Morgan&Claypool Publishers, 2009.

[25] J. J. van Wijk and E. R. van Selow. Cluster and calendar based visualization of time series data. In *Proc. of IEEE InfoVis'99*, pages 4–9. IEEE, 1999.