TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

Master's Thesis

A Cartographic Approach to Spatio-temporal Data Visualization of Events in Visual Analytics Applications

submitted in fulfilment of the requirements
for the degree of *Master of Science*
to the

Research Group Cartography
Department of Geodesy and Geoinformation
Vienna University of Technology

Supervisor: Professor Georg Gartner

written in cooperation with the
GeoVISTA Center, Pennsylvania State University

by

Alice Rühl

October 31, 2014

Abstract

The increasing importance of Visual Analytics, particularly in response to the advent of 'big data', calls for greater attention to maps in analytic tools and applications. While maps provide a familiar, understandable interface to underlying data, not enough has been done in the field of Visual Analytics to advance the design and usage of maps. This study explores the development of map components within multi-view applications using multivariate and zoom-level dependent representations of news event data. Integrated in the system with a timeline, a word cloud and other view components, we consider the traditional 'rules' of cartography as well as aspects of human-computer interaction and interactive design research. This allows users to efficiently explore complex spatio-temporal data. Our approach was validated by re-developing the mapping component inside STempo, a project developed at the GeoVISTA Center, Pennsylvania State University. We found that the application of cartographic precepts to our spatio-temporal map visualizations aided in the development of more useful visual analytic mapping tools.

Acknowledgements

I would like to express my deepest appreciation to Georg Gartner for his helpful critique and feedback. He always supported and inspired me, during my studies and while working on my thesis.

I would like to thank Frank Hardisty, who made my stay in State College become real and opened me this great opportunity to work with amazing people at Penn State, the STempo team and especially: Scott Pezanowski, who patiently helped me learning more about programming in Java and who worked on my ideas when I reached the limit of my own skills. Donna Peuquet, kindly available for interviews and questions, always having an encouraging smile for me in our team meetings. And of course Anthony Robinson, for valuable comments on my design ideas and especially for advise and guidance on the evaluation (much more than I could put into practice within the limited time for this thesis).

Further, I would like to thank all participants of my user study. You've been great!

Declaration of Authorship

I hereby declare that the thesis submitted is my own unaided work. All direct or indirect sources used are acknowledged as references.

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig verfasst habe und ausschließlich die angegebenen Hilfsmittel und Quellen benutzt habe.

October 31, 2014

Contents

List of Figures

1 | Introduction

1.1 Problem and Motivation

The rapid technological development over the past decades, especially in the last years, made it increasingly easy to collect vast amount of data about basically everything (mobile apps, digital car control units, daily news online, digital photo libraries,...). Data itself consists only of numbers and letters, but to extract information and gain knowledge we need to process the data. Primarily the usage for knowledge construction makes data valuable. With an increasing amount of data (buzzword 'big data') the processing and interpretation gets more and more difficult, therefore new methods and tools have to be developed. The emergence of the field 'visual analytics' is one response to the big data problem. Visual analytics combines the computational power of a machine with the human cognition system to enable effective analytical reasoning. It is known that analyzing data in a visual display supports reasoning, since it is easier for the brain to discover patterns visually than in numeric values or text. With the increased knowledge, complex problems can be easier understand and therefore, better decisions can be made.

Most often, data has a location associated with: It might be important to know, *where* something happened in order to understand *why*. Maps are necessary to show these spatial patterns and relationships. Recent studies show that users prefer simple visualizations, and geographic representation was ranked the highest in preferation. Moreover, it is also shown that aesthetics is very important to engage the user with the interface (Lanfranchi et al., 2014).

Despite the extensive literature about cartographic design (Bertin, 1983, Slocum, 2009, Brewer, 2005, Krygier and Wood, 2011) developers of visual analytic applications are confronted with many problems when implementing a map component. Not only the special character of geospatial data makes it difficult to find appropriate representations, but also complexity of the data and the interactive environment leads to difficulties that have to be carefully considered.

1.2 Scientific Question

The following section gives a brief overview on the research goal of this thesis, and the derived research questions.

1.2.1 Research Goal

The goal is to create an interactive spatial visualization tool to analyse complex spatio-temporal event data, integrated in an application with other components such as time-line or charts. The data is extracted from news articles and social media streams and contains information about the event, like title, location, time, category, and URL to the source article. Event data is difficult to visualize, as it is inhomogeneous and the information extractions methods can involve relatively high inaccuracies. For example when geotagging an event, the coordinates could represent the country mentioned with its centre coordinates, or it is the location of a city mentioned, or maybe the geotag is actually as accurate as a street crossing. The characteristics of the data have to be considered for the visualization.

The interactive map is meant to support analytical reasoning and visualize the results of the computational methods such as the t-pattern analysis in STempo. The t-pattern analysis compares sequences of events looking for co-occurrences. The result is given as a set of events, a pattern, and also a dendogram.

This work demonstrates design suggestions for the map to visualize the spatio-temporal characteristics of the events itself as well as the result of statistical analysis methods. Being part of the multi-view application the map is tightly linked with the other components, in a way that the displayed data can, for example, be filtered by other components. This allows for analyzing complex data, enabling the user to find space-time patterns.

As recently noted, the field of cartography needs to find "broadly applicable golden rules, guidelines, or heuristics that inform cartographic interface design" (Roth, 2014). This work would like to contribute to that research by making a good example.

1.2.2 Research Questions

- How to use spatial visualization techniques tightly linked with inductive computational / statistical methods for the exploration of spatio-temporal event data?

  or: How to visualize event data extracted from news articles or social media streams on an interactive map showing spatio-temporal variations?

- Does the visualization created with arguments from traditional cartography improve the application?

The questions can be broken down to the following steps:

1. Research visualization methods and other visual analytics applications to understand important considerations
2. Design a map concept for the cartographic representation of to the complex spatio-temporal data
3. Think about appropriate interaction, e.g. zoom, dynamic filtering, or mouse-over
4. Evaluate the suitability of the suggested design in a qualitative user study

This thesis is written in coorperation with the GeoVISTA Center at the Pennsylvania State University. The work is applied to STempo, a project developed at GeoVISTA. The event data used for this work is created within the STempo project. Eventually, the results are implemented in the STempo application and tested in a qualitative user study using the application itself in a version of September 2014.

2 Theory of Cartography and Visual Analytics

To clarify what is assumed as common knowledge in this thesis, the following chapter describes some general terms and topics as definitions. Further, an outline of relevant research in the field of cartography and visual analytics is provided, highlighting projects related to the visualization of event data.

2.1 Definitions

Cartography

The science of cartography deals with the thoughtful representation of spatial data to efficiently communicate with humans. Cartography is the ‘magic’ that transforms raster values or vector points, lines, and polygons of sensors and surveyings into meaningful images that actually tell a data story a human can read and understand. Maps help the cognition and situational awareness by ordering and structuring spatial data, possibly allowing for interaction or exploration. Good cartographic methods enhance the understanding of a map and improve the spatial cognition of the viewer. For a general introduction to cartography, see Bertin (1967), MacEachren (2004) or Slocum (2009).

Visual Analytics

Thomas and Cook (2005) defines visual analytics as the science that combines the machines’ computational power with a human brain’s functionality to make sense of big and complex data. Complex data visually displayed is interpreted by a human who is using existing knowledge and intuition to draw conclusions. Visual analytics therefore supports analytical reasoning and sensemaking ('visual thinking'). The visualization does not always give the answers directly, but easily catches a user's attention on problems for further investigation. Visual analytics applications let the user prove existing hypothesis or discover new and unexpected ideas. For an extended definition and description on the research field of visual analytics, see also Keim et al. (2008).

The fact that visual analytics relies on communication with a human makes it very interesting for cartography, while other cartographic domains dealing with new technologies turn towards machine use only (e.g. navigation will be automated so a car can drive autonomous, no map reading necessary). The human will always be an important part of any visual analytics environment.

Cognition

Card et al. (1999) states, "cognition is the acquisition or use of knowledge". Graphics and visualizations are used to amplify cognition, and therefore lead to a better understanding of a matter. The research field of cognitive science deals with a human's process of thinking and understanding, trying to find best ways to efficiently support this process. The International Cartographic Association (ICA) has one commission on Cognitive Visualization dedicated to research on use and user issues, behavior, and representation.[1]

Analytical Reasoning

Analytical Reasoning is the process of thinking and gaining knowledge based on contents learned within this process. Based on given facts and information one can reason about implications and causes, logically creating new facts that can be further used for decision making. This process is essential within the use of visual analytics applications, where the user's knowledge and experience plays a role in decision making. In the context of analytical reasoning often the so-called *sensemaking loop* is mentioned (Ribarsky et al., 2009). It describes the reasoning process as a loop of foraging and sensemaking, while

[1] `https://www.geo.uzh.ch/microsite/icacogvis/mission.html`

creating hypothesis and searching for evidences to test the hypothesis and reevaluate to eventually make a decision. For more details see also Pirolli and Card (2005).

Spatio-temporal

The term spatio-temporal describes according to Keim et al. (2008) one of the most prominent and ubiquitous data types. The data contain both space and time aspects, leading to complex data models. While using and analyzing the data, the main tasks are the identification of linear or periodical patterns, trends and correlations over time. Observing change over time is one task when working with spatio-temporal data, but also discovering trends and its variations over an area (Andrienko and Andrienko, 2006). Facing spatio-temporal data, one challenge is the "complex nature of geographic and temporal spaces", its heterogeneity making it sometimes hard to quantify and express numerically (Andrienko et al., 2007a).

Event

An event is an occurrence that takes place, something that happens. It can be a sports event, or a political event, or else. It happens at a certain time, might have a duration, and can reoccur cyclically. Mostly, events have a location associated with. A famous textbook on geographic information systems and science states: "Almost everything that happens, happens somewhere. Knowing where something happens can be critically important." (Longley et al., 2010, page 4). Some events are trivial or personal ('I just had lunch'), but usually and according to the Oxford Dictionary of English, an event is seen as something of major importance for a larger group of people ('Ukraine Prime Minister Resigns'). It is the task of journalists to write about events in politics, economy, crime, weather, disasters, sports, arts, technology, and others, that are of interest for the general public. Their reports are published in daily news and other sources. There is too much happening on one day to print it all in one magazine, therefore events are filtered for the national newspapers according to the local interest, sometimes even based on local opinions. It is very difficult to get an objective impression of world wide events, as it would take much time and effort to read and compare various news sources.

The dataset used for this thesis and how it is gathered is described in chapter 3.2.

Pattern

A pattern is a recurring design or sequence in a regular structure. In the context of spatio-temporal visualization the term refers to the repeated occurrence of certain phenomena and characteristics (see also Erwig, 2004). Andrienko and Andrienko (2006) explain the term 'pattern' using it in differentiation to the term 'behavior', where behavior is a certain set of characteristics that can, for example, be objectively measured, while a pattern is what we observe and how we interpret the measurement: "our perception of the behaviour" (p. 8).

2.2 Literature Review

Being a research thesis in cartography, this work uses concepts of multiple disciplines, including visual analytics, interaction science and human-computer interaction, GIScience, graphic design, and information visualization.

MacEachren (2013) states that we need maps to support analytical reasoning, to interact with computational models, and to cope with massive data. A starting point for reading is the paper *Geovisual Analytics for Spatial Decision Support: Setting the Research Agenda* (Andrienko et al., 2007a) which identifies research areas in geovisual analytics, highlighting the specific nature of spatial data and problems. The authors describe the goals of geovisual analytics for spatial decision support. *Space, Time, and Visual Analytics* (Andrienko et al., 2010) gives an overview on visual analytics, what is done and what needs to be done in the future.

2.2.1 Cartography & Spatio-temporal Visualization

One of the oldest and most cited books in cartography is Bertin's *sémiologie graphique* from 1967, where he introduces the visual variables used to represent qualitative and quantitative data. These variables are *shape*, *orientation*, *color*, *texture*, *value (i.e. brightness or intensity)*, and *size*. An explanation of thematic mapping and spatial data visualization is given by Slocum (2009), where some common cartographic methods are described, for example choropleth maps, proportional symbol maps, or isarithmic maps.

Many basic cartographic rules and conventions are still valid today, so is the set of graphic variables described by Bertin (1967). Even though sometimes new variables are introduced as technological advances induce new possibilities, as sound or smell added to the map interaction, the basic principle stays. With the use in visual analytics applications, as with other interactive maps, the design has to adapt to the use on a screen. While printed on paper, a much higher resolution and object density can be used on a map, the same map on a screen needs to be simpler to read and strive to be less strenuous for the eyes (see also Kraak and Brown (2003)).

A very famous author in the field of graphic design and data visualization is Tufte. His book *The Visual Display of Quantitative Information* (2001) coined the terms data-ink ratio and chartjunk, that are representative for his work. The rules described in the book are supposed to improve the visual literacy for graphs and charts: "Above all else show the data. Maximize the data-ink ratio. Erase non-data ink. Erase redundant data ink. Revise and edit." Many findings and rules from graphic design and information visualization can be addressed to cartography as well (and vice versa).

There are several articles published about the various aspects of visualizing time. In this thesis, the focus is on mapping and visualizing time geographically. Other visualizations such as a timeline is a component of STempo, but will not discussed in detail. The recent book 'Mapping Time' by Kraak (2014) gives a comprehensive overview on how to depict temporal aspects in maps, though focusing on the famous graph of Napoleon's March (Minard's Map of Napoleon's Russian Campaign of 1812).

One possibility to visualize time, especially promoted by Kraak (2003), uses the third dimension (z-axis) for the temporal information. The resulting cube representation can be read like a 3-dimensional map, but giving time information of a path instead of the height of a surface area. Also, there are is plenty of literature about movement and its visualization (see for example Andrienko and Andrienko, 2013). But since an event typically does not move as such, these aspects will not be further discussed.

The article *Mapping Time* by Vasiliev (1997) describes the basic methods to visualize time on a map. Vasiliev gives also background understanding on the concept of 'time' with a historical perspective, and further discusses temporal precision and time scales. For representing time in cartography, the temporal aspects can be categorized as follows (Vasiliev, 1997, p. 17):

1. Moments: the dating of an event in a space,
2. Duration: the continuance of an event in a space,
3. Structured Time: the organization or standardization of space by time,
4. Time as Distance: the use of time as a measure of distance,
5. Space as Clock: the use of space as a measure of time.

Considering the nature of the temporal attribute, the following possibilities to show time on a map are described in the article:

- Time labels, to show moments in time when events occur ('pinpointing')
- Isochrones are isolines of equal travel time from a given point, using time as a measure of distance
- Map series or 'Snapshots' to represent the duration of events in space and time (small multiples)
- Animation, or using time to show time
- Flow chart with arrows indicating direction (movement implies time, by knowing what was first)
- Frequency indicated with color codes

For the complete structured framework, see page 45, table 3: 'Spatiotemporal phenomena and their symbolization'.

Buckley (2013) defines guidelines for spatio-temporal map design by discussing map examples. The article divides between dynamic displays that allow for animation and time series, and time indicators such as timelines. Further, basic guidelines are given on how to design legends, titles and text, and graphs: use a constant scale for the timeline, keep the legend clear and simple, place all text with care, repeat this rules for graphs.

Aigner et al. (2008) shows that a wide repertoire of interactive techniques for visualizing datasets with temporal components is available in the field of information visualization. The paper explains several types of time oriented data with examples, and demonstrates that it is important to choose visualization techniques and its parameters as well as interaction functionality according to the present data. Then it is shown how data abstraction and the combination of visual and analytics methods are useful, challenging the problem of 'accuracy vs. abstraction' by conducting many data analysis steps. Furthermore it is discussed, how the user can be supported or guided in his analytic task, e.g. by emphasizing task-relevant information. It gives interesting ideas, though no maps or geographic representations are considered.

Next to the pure spatio-temporal visualization using a map only, a coordinated multiple view application is a commonly used technique to corporate multiple aspects of a data set in one application. Separate views with different visualization methods simultaneously show the same data. This way, one component can focus on spatial visualization while another shows the data on a timeline. Vis brushing and linking techniques the interaction with one component is promoted to all other views. Some examples are presented in section 2.3.

2.2.2 Human-Computer Interaction & Interactive Design

When designing an interactive map it is important to look into related research areas, like human-computer-interaction (HCI) or interactive design, to learn from their findings and improve the cartographic interface. The book *Interacting with Geospatial Technologies* (Haklay, 2010) teaches about the common aspects of HCI and GIScience, considering spatial cognition, cartographic theory, user-centered design, and usability engineering. The book provides practical guidance and principles for the creation of various geospatial applications. One general path to follow when working on the development of an application is presented as the "user-centered design (UCD) cycle" (Haklay, 2010, van Elzakker and Wealands, 2007). This cycle describes the design as an iterative process, considering the user requirements and usage context to create solutions that are evaluated for refinement, before deploying the result (see figure 2.1). The path is followed by this thesis, culminating in an evaluation that creates new suggestions for future deployment.

To create a good interface, we need to understand how interaction works, and how to use it. But what even means interactive? Interactivity is "a cyclic process in which two actors alternately listen, think, and speak" (Crawford, 2002). The two actors in case of an visual analytics application are the computer, and the human using it. To

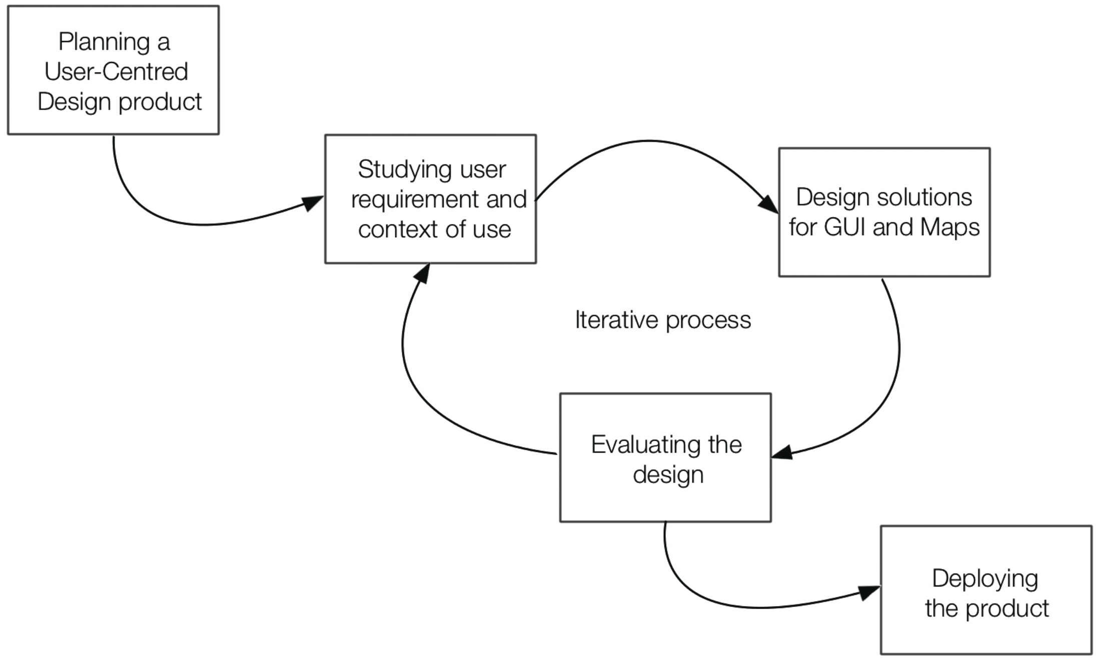


Figure 2.1: User-centered Design Cycle.
As presented by Haklay (2010, p. 100)

create a successful communication between computer and user, many researchers have thought about principles and guidelines that will help with this task. Graham (1999) published a book about the principles of interactive design, presenting 19 'rules' to follow: from knowing your audience, designing a clear navigation, and being forgiving of user error to a consistent and intuitive user interface that leaves as much control as possible with the user. Moreover, the design should kept simple. The author also gives suggestions for usability testing, which is recommended to confirm that the project meets the expectations: The participants of the test should try to understand the project themselves – It is a good sign, if they are able to.

Shneiderman and Plaisant (2004) specify eight "golden rules" of interface design:

- Strive for consistency
- Cater to universal usability
- Offer informative feedback
- Design dialogs to yield closure
- Prevent errors

- Permit easy reversal of actions
- Support internal locus of control
- Reduce short-term memory load

The relationship between interaction and cognition is described by Pike et al. (2009). The author explains best practices in the representation and manipulation of visual displays. Herewith, ‘the science of interaction’ and its challenges in visual analytics are introduced, a field that studies how humans interact with machines to support analytical reasoning. Through interaction, questions are expressed and visually answered, considering not only control mechanisms of the software but moreover prior knowledge, input of colleagues, or the environment. Endert et al. (2013) discusses that visual analytics applications have the chance to evolve beyond control panels, such as sliders and check boxes, by allowing the user to directly interact with the displayed data using visual metaphors and hereby influence the computational model. It is what he calls “direct manipulation”. Direct manipulation could remove the additional mental process of connecting a control panel to the visual output and therefore reduce short-term memory load, being easier and more intuitive to use.

2.3 Related Work

The following systems and applications were found to be related to the STempo project and ideas. While some projects focus on the technologic advance, others deal with ways of visualization of news events.

Descartes

Descartes is an interactive environment for exploratory cartographic analysis, proposed by Andrienko and Andrienko (1999). The Java application automatically creates maps from data tables. Based on the selection of variables the system is able to produce several different map presentations. The user can create and compare different maps, and also interactively manipulate the visual properties with slider controls. See also Andrienko et al. (2007b).

GeoViz Toolkit

The GeoViz Toolkit developed by Hardisty and Robinson (2011) is a coordinated multiple view environment that enables analysis of complex data sets. The system offers many different visualization possibilities, among which one is a geo-map. Interaction with one component is propagated to the other views, so that a hover on the line of Pennsylvania in the Parallel Coordinate Plot also highlights the State of Pennsylvania on the map (see figure 2.2). The visualizations can be created on demand and arranged within the main window by the user.

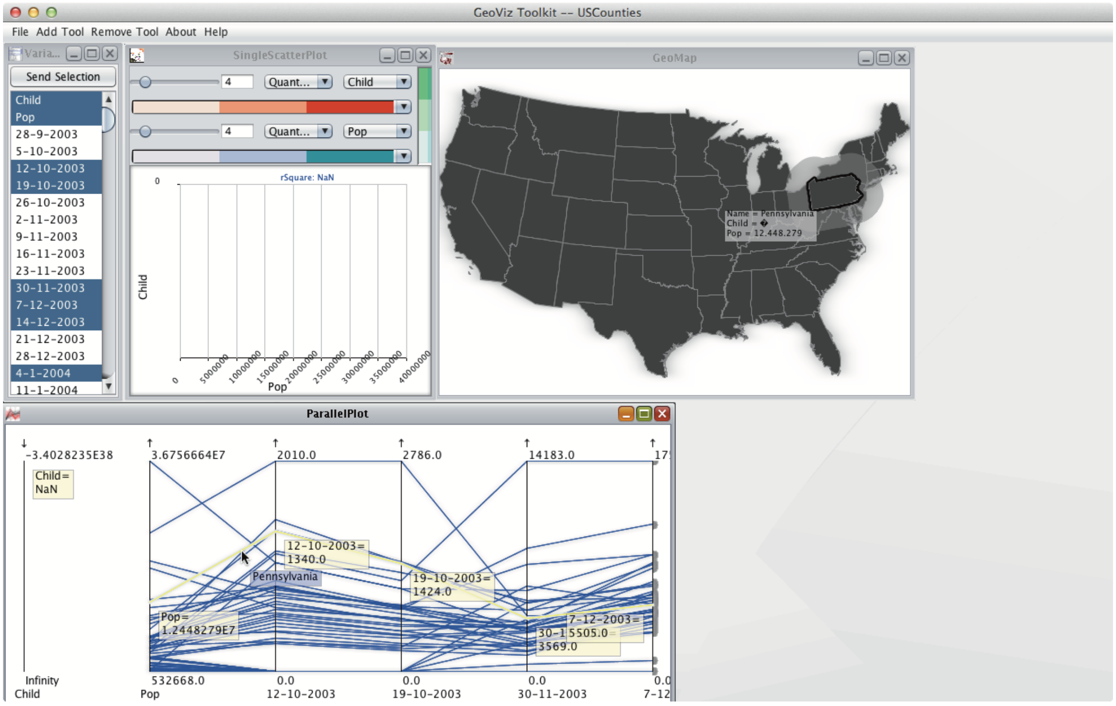


Figure 2.2: Screenshot of the GeoViz Toolkit.
Taken from a developer version in October, 2014.

Snap-Together Visualization

North and Shneiderman (2000) developed a multiple coordinated visualization system which they used to study the value of coordination usage. The system incorporates interaction techniques such as 'brushing', 'overview and detail', and 'multi-level browsing'. This work demonstrated that coordination is very important when the user is required to access details within data. Also, many tasks were conducted faster (30–80 %) with the overview and detail coordination, as tested in a usability study.

Assure Analytics (SATURN)

The communications services company BT is using visual analytics to discover problems in their infrastructure, like copper theft or network threatening. They developed a system to analyze real time data and visualize it in different ways, among one is also a 'geo-map' (BT website, 2014). The map shows the source and destination of network attacks. In the timeline, the displayed data can be limited to a certain time frame. A Play button enables the user to view the events occurring in an animated way, moving through the timeline while displaying the respective events. The basemap is a satellite image, not providing any further assistance for orientation in the map e.g. with labels (see figure 2.3).

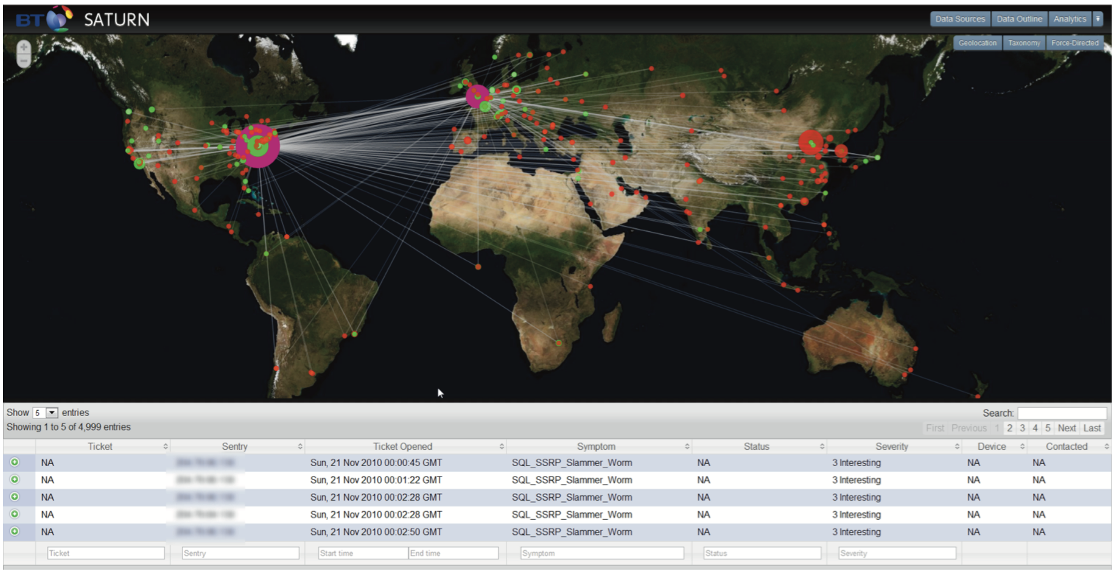


Figure 2.3: Screenshot of Assure Analytics (SATURN).
Taken from a presentation available online (Hercock, 2013)

Newsmap

The web application `http://newsmap.jp` shows a real time visualization of articles based on the Google News news aggregator. It displays the headlines of articles structured in a tree map. The fields of the tree map are colored according to a category (world, national, business, sports, ...), a darker shade means the news is from more than 10 minutes ago or older. Only the latest news are represented, there are no means of history. The size of the field is determined by the number of related articles, therefore more relevant stories are visualized bigger in size. It is possible to select certain countries from a list to only display news of this country, also a combination of several countries is possible. If more than one country is selected, a red frame with a small label indicates the country where the article is published. A search box filters the whole display and shows only articles related to the search term. Hovering a field displays a info box with the original story and a click links to the source where it was published. Both countries of interest and news category can be selected or deselected using the checkbox, and a click on one option deselects all others. A screenshot can be seen in figure 2.4.

Figure 2.4: Screenshot of newsmap.jp.
Taken on July 14, 2014. Sports seems to be the most important category in Germany this day, as it is the day Germany won the finals of the FIFA World Cup.

Article Threads

Krstajić et al. (2010) developed an approach to analyze news streams with article threads. News feeds are monitored in real time using a timeline representation. Events are visualized in threads based on category, displayed as on little block on the timeline (see figure 2.5). The color is set according to "the tonality value of the news article, where saturated green represents high positive tonality score and saturated red represents very negative tonality values." (Krstajić et al., 2010). Interesting about their approach is the implementation of a filtering algorithm, that removes old, short, and sparsely populated threads to reduce the data to the relevant only.

Figure 2.5: Article Threads.
The Visual Analysis of News Streams with Article Threads by Krstajić et al. (2010)
.

Spatiotemporal Social Media Analytics

Chae et al. (2012) uses semi-automated methods to analyze social media streams like Twitter, first creating topics ranked by importance, then exploring anomalies to gain insight in the large data body. The analyst can interactively refine the results in an iterative process to gain first a general overview and then a specific and detailed insight. Their approach is illustrated in figure 2.6, showing how the analyst steers the whole analysis process. The visualization consists of maps, graphs, word clouds and a list of the actual text content. Several use cases demonstrate the value of the application.

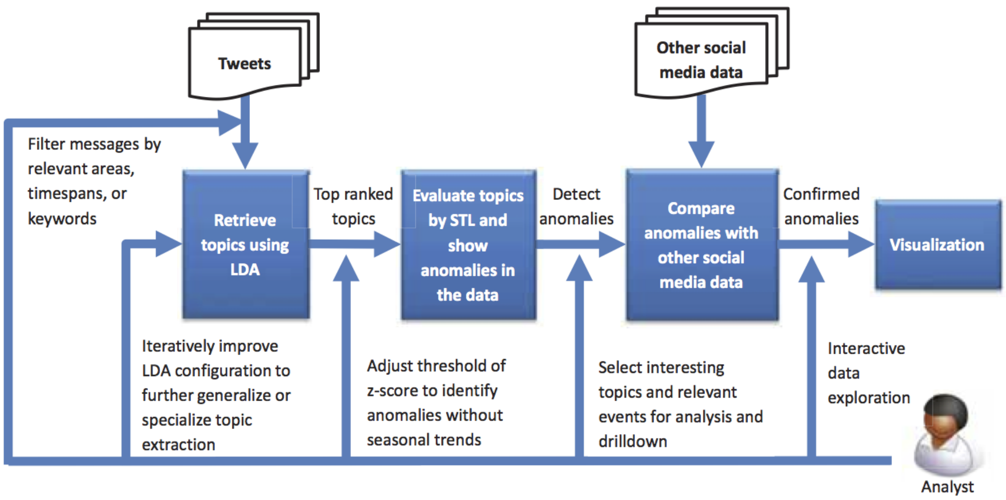


Figure 2.6: Spatiotemporal Social Media Analytics
for Abnormal Event Detection and Examination using Seasonal-Trend Decomposition
by Chae et al. (2012).

3 | STempo – Visual Data Exploration of News Events

STempo is a Project at GeoVISTA, first started in 2010. The team developed a visual analytics application that allows users to explore complex data for statistical co-occurrences and spatio-temporal patterns. It is a coordinated multiple view tool, offering various visualizations in the form of timeline, charts, dendogram, treemap, tag cloud, and map. This thesis focuses on the enhancement of the map component, discussing data visualization and interaction design. For a general overview on STempo, see also Swedberg et al. (2014).

Literature on design principles, of no matter what domain (graphic design, interactive design, cartographic design) gives the same advise: Before designing and implementing the map, its purpose and function have to be determined (e.g. Tufte, 2001, Graham, 1999, Haklay, 2010). The map needs to have a clear message and a justification for its existence. 'Concept before Compilation' is also the first principle of the five that should lead to good map design according to the Society of British Cartographers. In order to find the purpose and describe the function of the map component in STempo, in the following chapter the use and user of STempo is examined. It is analyzed *who* will use the application and *how*, and what is the role of the map. Next, the data is described in detail, from data gathering and mining to storage for use in the software, highlighting important aspects for the later visualization. Then the technology is described which builds the basis of STempo, with special interest to the visualization components. Following these three steps, the data visualization is created and sketched, also describing how the user would be able to interact with the system.

3.1 The Use and Users of STempo

Around the world, events are being continuously reported by news and social media. This ongoing stream of information can be very useful, when utilized appropriately. For example a user could learn *where* activities are taking place, *when* something happened before and *what* were the possible consequences.

According to Robinson et al. (2014), the main goal of STempo is the identification and visualization of potentially interesting spatial event sequences. STempo enables detection and analysis of co-occurrences of events based on a statistical significance. Those sequences might have been previously unknown and possibly no clear hypothesis was held by the user – the space-time patterns are supposed to support the analytical reasoning by enabling exploration and understanding of the patterns.

This can become relevant knowledge for various people, from journalists and news consumers to intelligence analysts and decision makers. Humans are the main factor in the loop of interpretation, context-awareness and decision making, because unlike machines, humans can rely on their experience as well as intuition (see also Andrienko et al., 2007a). This ability is used in STempo, by combining the computational power of the machine with a human who validates the result.

STempo is thought to be a 'general purpose tool', providing data analysis functionality for various data topics, e.g. the complexity of the world's economy or the spread of diseases. It is built for expert users, not for the general public. Analysts may use STempo to visually explore known patterns, but also to understand previously unknown patterns (Peuquet, 2014, personal communication on July 22). STempo targets for example political analysts (Robinson et al., 2014), assisting them in analyzing news articles of a certain topic, while being able to change the considered time frame. With its possibility to analyze current and ongoing events, but also to go back in time and compare events in the past, STempo can be an important tool in the sensemaking process.

The work flow, as described by Peuquet (2014, personal communication on July 22), includes a) opening a dataset and running the statistical analysis for t-patterns, b) discover significant patterns, and c) drilling down to get details about events involved in a pattern of interest.

In Peuquet's triad model of spatio-temporal data, the *where* is an very important part of the visual representation, next to *what* and *when* (Peuquet, 2002). It helps people to understand space and explore the spatial data. The role of the map, according to Peuquet (2014, personal communication on July 22), is to unravel big data from a spatial perspective, to show how many events create spatial patterns. The map is used to browse and explore the event data, first showing the whole data set and then allowing the user to filter in multiple modes. Further the map shows results of the statistical analysis, highlighting a t-pattern and showing its spatial distribution. This allows the user to identify important patterns that, based on probability, are associated with certain following events. For example, a certain combination could indicate political instability. If the beginning of this pattern occurs, it can be used to predict riots. The user is able to find patterns in differing temporal intervals, also it is possible to compare patterns to other locations.

We imagine the analyst to sit in front of a large screen, possibly with touch capabilities, browsing the news data for a current situation in, for example, the middle east. What she learns about the situation is collected and stored for a later report. Her first question could be 'Where does something happen?' to then ask 'How often does it happen?' or 'Which area is most active?'. The analyst needs to be able to load data into the application and conduct the statistical analysis. The data, as well as the analysis results are displayed visually in graphs, timelines, maps. But the display is not static, it only shows a current view of the data. This view can be changed by the user any time, as she could zoom into an area of interest or highlight an object to get more information ('What is this?'). Meanwhile, the computer runs the statistical analysis of events in real-time. If a specified pattern occurs, an alarm is started to draw the analyst's attention to that pattern. She can then explore the events and decide about the validity of the pattern.

What the user expects as a result of his work with STempo is an improved or new understanding of events around the world. An existing hypothesis was approved or declined, while new hypothesis were made. Eventually, a report can be created using the data visualizations and information gained. With this report, a decision maker can obtain a situational awareness and understanding, and make decisions based on the gained knowledge.

A set of use cases has been identified by the STempo project sponsor:

Humanitarian Assistance/Disaster Relief

There has been a flood, hurricane, earthquake, fire, drought,... recently in Blueland. What events have occurred, in what relevant time frame before the disaster and what pattern is associated with those events? What other events have a probability of occurrence given this pattern without consideration of the disaster event? Is there any data about the precursor pattern with disaster events?

The Red Cross is deploying food to the disaster zone. What event pattern has recently been associated with airports and ports? ...with locations where personnel are likely to be housed? ...with building sites likely to be used for warehouse storage? ...with roadways, major roadways around key locations? Is there temporal pattern/periodicity with roadway events?

The UN is about to deploy humanitarian relief. What precursor political event pattern exists and what likely downstream events from this pattern should be taken into account?

Doctors Without Borders wants to select a staging area somewhere in the country for relief and medical aid. What pattern events suggest good and bad sites?

Failed States

There has been a riot in Orangeland. What events are likely to follow, in what time frames, and where?

Military/Tactical

We are about to have a US Naval Port call in Yellowland. What are the events that occurred recently and what events are likely to occur while the vessel is there given those recent events?

Terrorism

There's been a kidnapping, bombing, village intimidation,... in ... What are the implications? What events are likely to follow given this event? What larger pattern is associated with this event? Can future kidnappings, bombings, etc. be predicted given the event pattern observed?

The following questions could be answered using the map component of STempo:

- Where do most events happen?
- What category of events is occurring the most in a certain area?
- What is the spatial extend of a pattern?
- At what times occurred events in a certain area?
- Where. . . ?

3.2 Event Data Generation and Explanation

As described in section 2.1, events are occurrences that take place somewhere, at some time, for some duration. Events are happening all the time, everywhere around the world. The daily news are a good source to learn about the occurrences around the world. The data used in STempo for current development and testing is gathered from RSS news feeds focusing on events associated with the Arab Spring in the Middle East in 2011/2012. The web-scraping tool developed in the GeoVISTA labs queries the feed and collects event data, which is possible to do in real-time. The data is stored in a table, one event per row, with ID, timestamp, event type code and name, location name and coordinates, and URL to the source report. Gathered from real world event reports, the data set is large and very complex. The information is most often inhomogeneous and needs abstraction to be standardized and eventually visualized. For the development, data sets containing between 3000 and 300000 entries are used, which can not be considered 'big data'. Though the planning keeps in mind to be able to process 'big data' eventually, and also deal with real time input that is highly dynamic.

The events are automatically encoded using the TABARI and CAMEO (Schrodt et al., 2008) categorization tools, creating ten high-level event categories: 1. Armed Group, 2. Civil, 3. Demographic, 4. Diplomatic, 5. Economic, 6. Military, 7. Physical/Nature, 8. Political, 9. Religious, and 10. Tribal. There is no value or ranking associated with these categories, the data type is nominal. All ten categories are equally important and can not be further grouped or merged. Each category has up to a hundred sub-categories, that gives more detail information about the type of event. For example ‘Political’ has sub-categories like ‘make statement’, ‘reject’, ‘accuse’, or ‘threaten’, while ‘Military’ has sub-categories like ‘kill military casualty’, ‘use violent repression’, or ‘employ aerial weapons’.

The geographic location is retrieved using part-of-speech tagging: locations mentioned in the text are transformed into coordinates with help of the GeoTxt tool[1], resulting in latitude and longitude values. In MacEachren’s topology of data models, the events would be discrete objects with a varying spatial distribution from smooth to abrupt (see figure 3.1), depending on the type location mentioned and the georeferencing process. The implications of the data model have to be considered later on in the choice of cartographic representation methods.

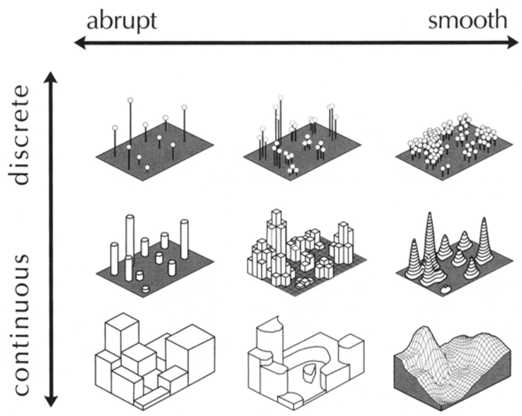


Figure 3.1: A Topology of Data Models (MacEachren, 1992).

When reporting events, the language can refer to a country, or a city, or be as precise as a street crossing. There are many pitfalls related to the georeferencing process, additionally to the difficulties with a vaguely defined spatial extend. For example several locations can be mentioned in one report. It has to be identified which place is most relevant and how to include all locations in the data model and representation. Also, ambiguities can occur when retrieving locations from written text, e.g. *New York* could refer to the state of New York, but is also commonly used for New York City; or even worse, *Paris* is not only the capital of France, but also the given name of a prominent person. Further, latitude/longitude coordinates imply a higher accuracy than is actually

[1]`http://www.geotxt.org`

given by the language in the report. This has to be kept in mind when analyzing and visualizing the data generated by automated processes.

An event is tagged with a timestamp. Even though, currently, only a simple point in time is assigned to each event in the data set, the temporal dimension of events can be quite complex. Other than just a point in time, an event can have a starting time, ending time, and duration. Events could recur periodically, e.g. every day or month.

The statistical analysis incorporated in STempo makes use of the T-pattern analysis technique, that detects significant co-occurrences of events. The method recognizes weather a temporal distance between two events is random or not, or if they recur in the same temporal sequence (Robinson et al., 2013). Eventually the significantly related events are organized in a hierarchical pattern, as can be seen in figure 3.2. In the main window of STempo, these complex hierarchies are reduced to patterns of color, indicating the involved event categories: .

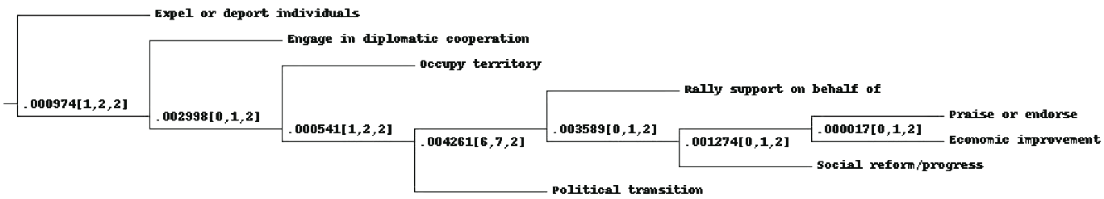


Figure 3.2: T-Pattern Dendrogram.
Example of a dendrogram as result of a t-pattern analysis.

The data set used within the development of STempo is, as mentioned earlier, related to the Arab Spring in 2011/2012. That time, many countries in the Middle East have experienced socio-political instability, some have undergone a political change. Some significant pattern was found for the data set of Yemen, that could be validated for different times in the last 25 years when the country was politically instable (Peuquet and Stehle, 2014). The pattern is described by Peuquet and Stehle as follows: "An accusation of human rights abuses is followed within one day by a demonstration for a change in leadership, followed within another day by engagement in negotiation, followed between three and eight days by an optimistic comment being made, concluding four to six days later with rejection of the plan that would have settled the dispute."(see p. 94).

3.3 The Technology behind STempo

STempo is a Java application, making use of standard Java objects (e.g. JFrame) to build a multiple view Graphical User Interface (GUI). It uses parts of the GeoViz Toolkit (Hardisty and Robinson, 2011), developed at the GeoVISTA Center for spatial, temporal and attribute data analysis. The multiple components of STempo are coordinated and information is exchanged, enabling to display the same data set in several views simultaneously.

A major task of STempo, the statistical analysis of events creating sequences vis pattern matching, is written in Python. The seamless integration of several programming languages within one application is considered as polyglot programming. This approach combines and leverages the best functionalities of both programming languages. The main interface of STempo consists of several components: a legend, a word cloud, a timeline, a sequence view, and a map. All these components are different representations of the data, linked with each other to consistently show even the same subset of the data. The legend explains the categories and their color representation. The word cloud displays words that are often mentioned in the reports. The timeline displays when events happen, also using the color scheme to indicate the category of an event. The sequence view lists the sequences found with the statistical analysis. The map component displays spatial distribution patterns. Figure 3.3 shows a screenshot taken at the begin of the work for this thesis in May 2014.

Further visualizations and settings can be found in tabs at the top of the window. All components are under development and being improved constantly. This thesis focuses on the development of the map component. But as the map is tightly integrated with the other components, their development is mentioned aside.

The STempo map component makes use of the Java component JMapViewer[2]. This component provides the framework and basic functionality to display a tile based map in a Java window. The open source code includes classes for loading tiles from a server or local file, controlling and caching the tiles for the fluid display of a slippy map. Likewise, zoom and pan interaction is already provided with the JMapViewer code. The component understands coordinates in latitude and longitude, transforming their values

[2] `http://wiki.openstreetmap.org/wiki/JMapViewer`

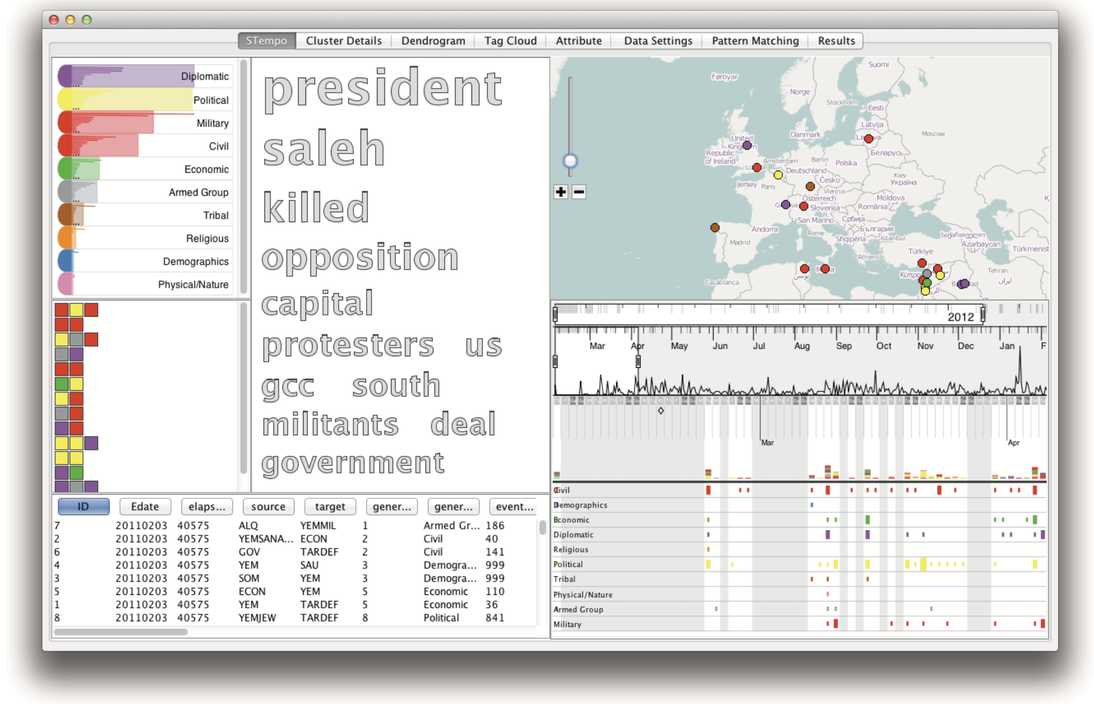


Figure 3.3: STempo Screenshot, May 2014.

to the correct position on the map. It also helps to create simple markers, rectangles, or polygons. The basic functionality can be extended and adapted to any needs, as all code is written in pure Java and openly available.

3.4 Implementation of the Data Visualization and Interactive Design

The data explained in section 3.2 is visualized for a possible user described in section 3.1 using the available technology, section 3.3. Based on the gained knowledge, the map component is designed. Starting with an efficient basemap, the visualizing process follows the ‘Visual Information-Seeking Mantra’ by Shneiderman (1996): Overview first, zoom and filter, then details on demand. The data visualization is tightly bound to the interactive design and user interaction, subsequently described. Eventually, possibilities on how to include contextual geographic information are discussed.

3.4.1 The Basemap

The basemap is important for orientation, it should provide known reference points to add context for situational understanding. To create an efficient basemap, it is necessary to know what questions the map is supposed to answer. STempo is visualizing news events, therefore administrative boundaries on country level are an important feature to be shown on the map. While political events are relevant on a global level, it is more important to see in which country events happen rather than to find the exact street or building within a city. Bertin (1983) divides these questions into elementary, intermediate and overall level; due to the zoom functionality of an interactive map the basemap should be created zoom-level dependent to provide optimal support on every stage of analysis. Bertin further suggests to choose reference points according to proximity, stability and prior knowledge of the average reader. A rapid and precise identification of the geographic area is desired. The basemap's visibility should be lower than the information content, establishing a good figure-ground contrast. The information carrying objects have to be clearly visible, while the basemap only supports the orientation without distracting from seeing what is important.

For STempo, the following zoom level, divided into three main groups, are considered: world, zoom 0 - 3 for an overall overview; country, zoom 4 - 6 for intermediate detail; and city, zoom 6 + for a detailed view. A zoom closer than city level is not required for this project, but could easily be included for other applications if necessary. For each level, different reference elements are important. For example, on the smallest scale (zoom 0), the most obvious reference is the coastline, dividing continents from ocean. Roads are not important on this zoom level, but they are a major feature for orientation on city level (zoom 7). Table 3.1 shows the selected features and their style for each zoom level group. The colors are pale and grayish, to reduce possible distraction from the main data visualization.

The basemap data used is from Natural Earth, a free-to-use dataset built by volunteers[3]. The shaded relief is provided as a raster file, all other data can be downloaded as vector data (conveniently all together in one sqlite database). The raster is manipulated using GIMP Image Manipulation, in a way that all pixel representing ocean are replaced with alpha transparent pixel. The 'Fuzzy select' tool with a threshold value of 0.0 can create

[3] `http://www.naturalearthdata.com`

Table 3.1: Geographic features of the basemap.

Features		Style	World	Country	City
Land		area #dcdcdc	✓	✓	✓
National borders		line #f8f8f8	0.5 pt	1 pt	1 pt
Ocean		area #f0f8ff	✓	✓	✓
Rivers		line #ddeeff	–	1.5 pt	3 pt
Glaciated areas		area #f8f8f8	✓	✓	✓
World cities		dot #646464	1 pt	3 pt	5 pt
Large cities		dot #646464	–	3 pt	3 pt
Label	Aa	Futura Med #646464	7 pt	10 pt	10 pt
Urban areas		area #ffd9bb	–	–	✓
Roads		line #d7a175	–	–	1 pt
Relief		shades of gray, opacity: 60%	–	✓	–

the whole selection with one click, as most seas around the world are connected to the ocean. The Black Sea and the Caspian Sea (with neighboring 'Garabogazköl Aylagy') and also the Great Lakes of North America have to be selected additionally. As the ocean and seas are transparent, an overall background color can be used later on to show them in blue.

The basemap is composed using TileMill, where the data layers can be styled with CartoCSS, a style sheet language similar to CSS but optimized for the use with spatial data in TileMill[4]. It is possible to load raster files and change the appearance of color range and transparency. The relief is displayed in gray tones with a transparency value of 0.6 to reduce its prominence. The vector layers are loaded and styled as indicated in table 3.1, making them appear zoom level dependent by specifying their zoom level range visibility.

The tiles are exported in the MBTiles format[5]. Considering the file size and since the basemap only uses a limited color range, the 8bit png format is suitable. It results in much smaller image files compared to 24bit png. Using the free and open source tool MBUtil[6], the image files and folder structure is unpacked, which is necessary for the implementation in STempo. The whole directory of all tile image files from zoom level 0 to 7 occupies approximately 40 MB disk space, increasing significantly for every

[4] `https://www.mapbox.com/tilemill`
[5] `https://www.mapbox.com/foundations/an-open-platform/#mbtiles`
[6] `https://github.com/mapbox/mbutil`

additional zoom level. There will be investigations on how to include MBTiles directly in a JMapViewer map, since MBTiles are a compressed file based on SQLite which has several advantages compared to the raw tile directory. One significant advantage is the handling of duplicate tiles, in a way that the very same tile (for example tiles of the ocean) is only stored once and then further referenced, which safes a huge amount of disk space. A relatively small file size allows to include the complete stack of tiles in the runnable version of STempo, enabling an offline use of the map. This makes it extremely fast to zoom and pan, as new tiles are loaded immediately from the local file, instead of from a server as it would be usually with slippy maps. Another advantage, next to the fast responsiveness for map interaction, is the possibility for a complete offline use of STempo. All that is needed is the data to be analyzed.

To include the local map tiles as a tile source for the JMapViewer, a new Java class is created. It extends the `AbstractOsmTileSource` written by the authors of the JMapViewer. In this class I named `OfflineTileSource`, the relative path to the resources directory where the tile images are stored is defined, and also the minimum (0) and maximum (7) zoom levels that are locally available as tiles. The class is then imported in the main file and can be used as a tile source for the JMapViewer map: `tileSource = new OfflineTileSource();`

The resulting basemap can be seen in figure 3.4 in three different zoom level.

Figure 3.4: Basemap in three Zoom Level.
The basemap in three different zoom level world, country, and city.

3.4.2 Overview First

The first goal of the map is to give an spatial overview of the data. The user should see where events happen, to be able to get an impression of the relevant geographic area.

She should be able to see the spatial distribution of the data. Questions addressing this early state of map inspection are: Where does something happen? How often?

As on some locations many events happen, other locations are empty: The data model is smooth, and events are represented as discrete objects on the map (see figure 3.1). Showing a point for each event results in clutter and much overlap. For a more meaningful representation, events can be aggregated based on their attributes or location.

We need a multivariate representation for an event that takes place at a location, having additional attributes category and time. Several events should be aggregated while counting the number of respective events. To show the amount of events happening at one location, the graphic variable *size* is used, as size is the only variable that allows for quantitative measurements (Bertin, 1983, p. 69).

Categories are of nominal data type, their qualitative difference is best shown using *color*. As there are currently 10 (possibly 12) different categories, color is the best variable to distinguish these. The human eye is able to distinguish up to 10 colors with a reliability rate of 98%, and also symbols differing in hue are better distinguished than variations in size or shape (MacEachren, 2004).

Color Scheme for Qualitative Data

The main function of color in the visualization of event data is the differentiation of categories, enabling the user to find features of the same category in other places on the map and in the timeline and other views. Color is a very important part of any good design, as is not only a strong actor for the judgment of beauty and attraction, it also carries meaning and may invoke associations. The color scheme must support the message of the data, guiding implications rather than obscure and confuse the reader. For example, an implication of any order is inappropriate, as all nominal categories are equally ‘important’ or are ‘worth’ the same. To create a good qualitative color scheme, only hue is varied while lightness and saturation are kept constant (Harrower and Brewer, 2003). It is important to ensure all colors can be distinguished (at least by people with full color vision) and they stand out from the background.

Finding a scheme that fulfills all requirements is challenging and needs careful consideration. Using the color brewer, an online tool helping with cartographic color selection

Figure 3.5: Color Brewer Color Scheme.
Set3, the only suggestions suitable for 12 qualitative classes for nominal data.

developed by Harrower and Brewer (2003), only one scheme suggestion is suitable for 10 (or 12) qualitative classes: ‘Set3’ (see table 3.5) . However, one of the colors in that scheme is a light gray, which does not stand out from the basemap’s relief. Also, this scheme is not color blind friendly, not LCD friendly, not photocopy friendly, and not print friendly, according to the color brewer advise.

Therefore a new color scheme was developed. Beginning was one suitable bright and saturated color, the next color is found using a split complementary scheme. This lead to 12 different hues with balanced brightness and saturation to achieve a harmonic color scheme. Using that many different colors soon creates a very playful and ‘decorative’ impression, not suitable for a serious data analysis task. Trying to avoid a too motley and colorful appearance, the color is mixed with light gray (#c0c0c0). Table 3.2 shows the result in both tones, in the left part of the swatch the hues mixed with gray, the right and larger part showing the full color. The categories are assigned to the colors sometimes based on personal association (Military = war = blood = red), but eventually it is random. The color scheme is tested only on a LCD screen (MacBook) and might be tested and approved for other digital displays. As people with color deficiency usually avoid color coded systems, these users are supported by providing interaction (e.g. selection) rather than trying to find a scheme that can be distinguished even by color vision deficient users.

Table 3.2: Qualitative color scheme for up to 12 different categories of news events.

Color	Hex value	Category	Color	Hex value	Category
	#f58403	Armed Group		#35b6e3	Physical/Nature
	#92cd2d	Civil		#355fe3	Political
	#e52a5f	Demographics		#29c05c	Religious
	#35e3b9	Diplomatic		#b935e3	Tribal
	#ffe009	Economic		#734be6	(not used here)
	#eb441a	Military		#e335b6	(not used here)

The previously described considerations lead in general to the following possibilities of graphical means:

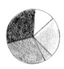
Choropleth, Tessellation, or Isoline Methods

Choropleth or tesselation, as well as isoline methods show the spatial distribution of data with frequency values, based on different areal units. A display using aggregated data on administrative units as in choropleth maps would not make much sense for the visualization of news events, as large parts of the map may not contain any data (where nothing is reported). The same is true for area-related composite symbols, for example pie charts placed for each country or state. A better possibility could be a method using point density, as with isolines (heat maps) or the geometric binning method (e.g. using hexagonal shapes). These methods are more flexible in dealing with inhomogeneous distribution patterns and show where many events occur.

However, all these methods are difficult when it comes to additional variables, since also we would like to differentiate *what* happened, i.e. the category of an event. There are maps using these methods with up to three different categories, where mixtures of the respective colors indicate the amount of that category in a certain area (for example the Pop-vs-Soda map by Alan McConchie[7]). But more than three colors get impossible to differentiate.

Figure 3.6: Choropleth Map and Hexagonal Tessellation.
Sketch of a choropleth map (left) and a hexagonal tessellation (right).

Pie Chart



A pie chart is a commonly used graph to visualize aggregated data, since it comes as a standard implementation with many tools (e.g. MS Excel). Unfortunately it is not very good. The size of the circle represents the total number of events, and the circle is divided into sections based on the number of events in a category and colored accordingly. The circle can be placed on the exact location of the event, only aggregating those at the very same location rather than combining data based on an administrative

[7] `http://popvssoda.com/`

unit. The latter would result in a quite empty map, as all information about smoothly distributed events will be lost as for example many countries do not contain data. The pie charts soon overlap each other, making it more and more difficult to read. Because of their low data-density and the difficulties readers have with ordering the values, Tufte (2001, p. 178) argues that pie charts "should never be used". Furthermore, "the only worse design than a pie chart is several of them [...]".

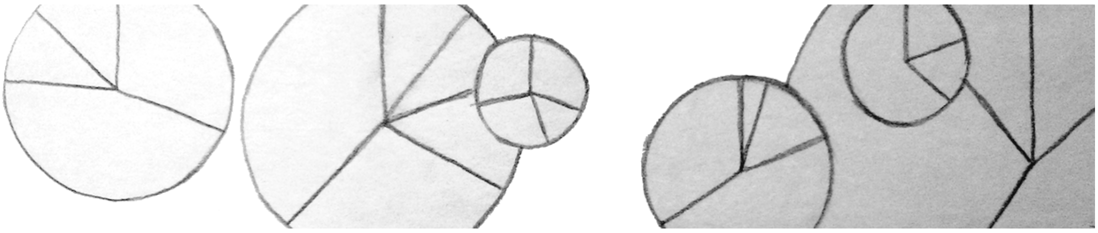


Figure 3.7: Pie Chart Map.
Sketch of spatially distributed Pie Charts, as it could look on a Map.

Star Plot and a variation, Flower Plot

A star plot uses radial axes around a center point, displaying quantities on each axis. As the individual quantities of every axis are connected, a graph forms with a shape that resembles a star. In our case, each axis represents one category displaying how many events occurs in the respective category. The star plot is aggregated and placed following the same principle described for a pie chart. However, the reader can look at the overall shape for visual comparison, so it might be easier to find similar shapes at different locations or to identify outliers.

The ten different categories are represented by an axis each, but they are difficult to identify with only the axis line colored in the respective hue. Therefore a graphic variation can be applied, where the area around the axis is filled with a colored shape, its size according to the quantity. As shown in figure 3.8, it would not be necessary to display all axes in each symbol, opposing to a star plot, which increases readability for those graphs anyways containing only a few categories.

Despite the usefulness of such glyphs for some purposes (Holzinger, 2014), several glyphs are probably better compared when overlaid instead of compared next to each other. Therefore it is suggested using these glyphs in a separate diagram instead of within the map (will be described in section 3.4.4).

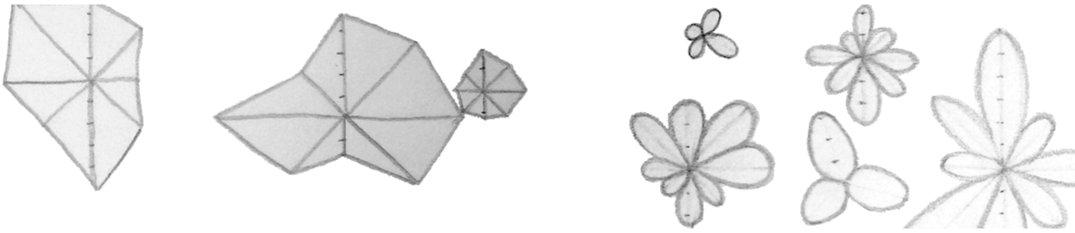


Figure 3.8: Star Plot and Flower Plot.
Sketch of star plot (left) and flower plot (right).

Proportional Ring Symbol

The proportional *ring* symbol is a simple variation of the Proportional Symbol explained by Slocum (2009). Events with the same category at the very same coordinate position are aggregated, the event count is mapped to the size of the symbol. The color of the symbol represents the category. Since circles tend to cover a lot of map space, and especially each other, only the stroke line of the circle is drawn leaving the filling blank, i.e. 100% transparent, creating a ring symbol. This way several colored rings for different categories and varying count can be placed at the very same location and still be differentiable (as long as the count differs). A concentric ring symbol emerges, with possibly up to 10 (or 12) differently colored rings at a location, one for each category with events at that location. If no event of a category occurs at this location, no ring symbol is drawn for that category. Elemental circles can be distinguished in size even when distributed over space (though, this needs to be validated by evaluation). Using a circle object let the user estimate, where happens something, and how often.

Figure 3.9: Concentric Proportional Ring Symbol.
Sketch of concentric proportional ring symbols.

The count can be mapped linear to the size $f(x) = x$, but this leads to a wide span of sizes: Many locations have symbols with size of 1 for one event, while on other locations many events occur, resulting in huge symbol sizes, rendering the ring too big to be even displayed on the map.

Therefore a better scaling function has to be found, that keeps the maximum size as small as possible while maintaining the minimum size readable. One suggestion is a square root function $f(x) = \sqrt{x}$, since the size increases much slower. A comparison to the linear function is shown in figure 3.10.

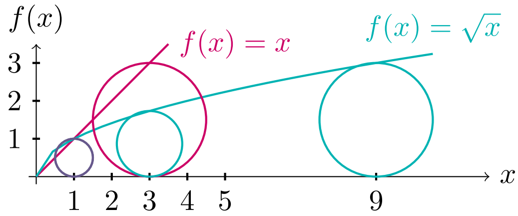


Figure 3.10: Scaling Functions.
Comparing linear with square root function.

Often, especially for paper maps, cartographers use a classification method to limit the number of different ring sizes to make it easier to actually read the value the size represents. Here, a classification is avoided to keep the variety of sizes as high as possible, reducing the graphic conflicts occurring when two rings have the same size. A comparison of size is possible, enabling the reader to make a statement about which ring is bigger than another. The size gives an overall impression and an estimation about the amount of events. To learn about a count value in detail, the user can interact with the map object and retrieve the information e.g. via pop up window or the connection to other components.

It remain graphical conflicts that need to be solved. Objects with the same location and event count would have the exact same ring size, making it impossible to distinguish between them. This is especially often occurring for single events, with a count of 1. How this would look like on the map is the same effect as can be seen in figure 3.10, where the ring for both graphs has the same size on position $x = 1$ and is therefore colored in dark violet (the sum of blue-green and red-violet added together).

Because already 10 different colors are difficult to differentiate on a map, a gradual mix of several colors to show the occurrence of several categories is not an option. To maintain all criteria of differentiation and still place the object on its location, a slight offset could help. The ring would be shifted to one direction by the width of the stroke, so both stoke lines of both otherwise overlapping rings can be seen. The overall impression of the spatial distribution would be the same, but all objects are visible without covering each other. Another idea, but maybe more complex to achieve technically, is showing the ring symbol striped in all involved category colors of this size and position. This option again makes the visual image very fuzzy and more difficult to see clear structures. Therefore we decided to use the offset variant.

Eventually, two additional variables can be used for representation of further attributes: brightness or saturation, and stroke width (see figure 3.11).

Figure 3.11: Proportional Ring Symbol.
Sketch of proportional ring symbols, varying in size, brightness, and stroke width.

It is apparent that on some locations seems to happen a multitude of events, even though there is no bigger city in that area somewhere in the middle of the country. This might be a hint that all these events refer to the country, instead of a city or more precise location. To indicate this situation for a clearer understanding of the spatial reference, the label with the location name is placed in the center of the rings. Capitals could be additionally indicated using a small star symbol, referring to the traditional symbology in atlases.

Using the presented ring symbol seems to create an impression of the spatial distribution of events, even though the accuracy of the geo-locations is very diverse. The proportional ring method is very flexible regarding inhomogeneous locations, as the symbol can be placed on any exact location or center coordinate of a polygon. Symbols placed on countries can be read and compared with rings placed on cities or even coordinates. While of course, the analyst has to consider that countries are probably mentioned quite often relatively to more specific locations: The value for the count of events is not normalized. Categories could be selected and deselected without affecting symbol sizes in other categories. Naturally, fewer categories and therefore objects to represent increase the readability, and this is what interaction helps with.

3.4.3 Zoom and Filter

To 'drill down', as Peuquet (2014) it calls, is possible by interacting with the components. The data visualization needs to consider the user interaction, both should consistently work together. Zoom and filter are basic and important interactions on an interactive map. As described in section 3.3, the map in STempo is built with tiles and natively provides zoom and pan interaction.

The user should be able to zoom in for seeing more detail, being able to discover even slight differences in an area of interest easily. Therefore the symbols should keep their size when zooming, i.e. always cover the same map space instead of being drawn in always the same pixel size. This gives the impression of seeing things closer when zooming in. To achieve this effect the value for meter per pixel is considered in the formula calculating the pixel size of a ring symbol. This value is reduced with a factor of 0.00001, a value found by experimenting. Though, after testing the visualization with a larger dataset, it was clear that only the square root function is not flexible enough. The biggest circle size should depend on maximum values, to ensure a readable symbology even with very large data sets. The actual scaling function looks as follows, where $c = const.$ is used for overall scaling, it is set depending on the maximum value for count, i.e. the total number of entries in the database:

$$size\ (in\ pixel) = \frac{\sqrt{count}}{meterPerPixel * c}$$

There are different ways of filtering, or *selecting* the objects displayed on the map. It is always executed based on user interaction, following the users intention and supporting his understanding. Possibilities for an interaction regarding feature selection within the map component are:

- simple click to select a feature,
- command + click to add features to the selection,
- alt + click and drag to create a rectangular selection,
- click on a geographic feature (e.g. country) to select all associated data features, or
- use of the other components to control the selection.

The selection is very dynamic and is linked simultaneously across all components of STempo. Consistently with the other components, the full data set is always displayed in transparent and subtle tone, while only the currently active selection is displayed in full color. To assist the user in interacting with the system, it should be obvious to the user how to interact with it (see also Shneiderman and Plaisant, 2004). Features that can be selected should be indicated, either by their appearance, or by being sensitive to mouse over. A hover effect, highlighting a feature when the mouse moves over it, indicates a responsive object for interaction. The highlighting effect is created be increasing the stroke width for a hovered ring symbol. While the symbol is temporarily enlarged it can also be better seen, with the mouse acting like a magnifying glass.

The rectangular selection by dragging the mouse is conflicting with panning the map, therefore the selection is created while holding a key down. There are two possibilities to create the selection: Either all ring symbols intersecting the rectangle will be selected, or only those having their center coordinate within the rectangle's boundaries. Behind the two ways of multiple selection lie different ideas and goals. While a simple click or drag over the rings considers them as objects to interact with, the rectangular selection of center coordinates, i.e. the actual event location, can be seen as a spatial query. Depending on which selection is desired for the current task, either the shift or the alt key is pressed while dragging to mouse to draw the rectangle.

The other components of STempo are tightly connected with each other and the map. The legend is the key to the colors for the different categories. Categories can be selected and deselected, resulting in the respective events being displayed accordingly also in any other components, including the map. The time frame of interest can be changed within the timeline, limiting the displayed events to events within this time frame. A click on a word in the tag cloud selects all events related to this word.

Visualizing Time

In a coordinated-view visualization, one primary visualization for temporal aspects of the data is a timeline. In STempo, the timeline component shows the occurrence of events over time, providing several layered timelines in combination, displaying an overview graph and additionally a detailed view of events, with days being the smallest unit and several years being displayed on a small scaled overview timeline. Via manual controls the timespan of data displayed in the other components can be interactively filtered by the user. The variable timespan creates a selection, it can be changed in size or moved along the timeline. This component gives also information about each category of events separately, but it does not indicate any spatial distribution. The selection made in the timeline is therefore transmitted to the map, where a spatial pattern of events in the specified time frame can be examined.

Directly incorporating temporal aspects in the map is a difficult, but interesting problem (as discussed in section 2.2). Some possibilities are already presented. The first problem that comes along with the STempo visualization is the fact, that the ring symbols are aggregates of several events, each having a different time attribute. Therefore their temporal aspect can not directly be visualized with, for example, lower saturation or lightness, unless an average value is calculated for all time occurrences. To gain more information, statistical values such as minimum, maximum, mean, standard deviation could be considered. The following figures 3.12 to 3.16 show various possibilities for visualizing time aspects within the map symbol. The result are spatially distributed tiny graphs based on a horizontal timeline (or radial, in case of figure 3.16), allowing for direct comparison of two or more locations.

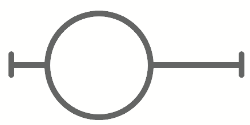


Figure 3.12: Visualizing Time: Ring Symbol with Horizontal Bar.
Ring symbol with horizontal bar indicating the time span in which the aggregated events take place.

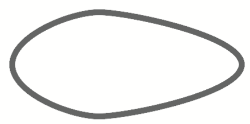


Figure 3.13: Visualizing Time: Distorted Ring Symbol.
Ring symbol distorted to left and/or right indicating amount of events older or newer than time x (defined by movable marker in timeline). The shape resembles an egg.

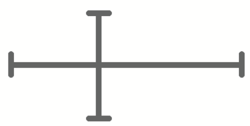


Figure 3.14: Visualizing Time: Two Orthogonal Bars.
Two orthogonal bars indicating the amount of overall events and their temporal extend, the intersection at time x (defined by movable marker in timeline) placed at the respective position on the map.

Figure 3.15: Visualizing Time: Small Histograms.
A small histogram indicating the temporal distribution of events, creating a shape that can be placed on the map likewise the ring symbol. This visualization would be redundant to the timeline and is therefore not desired.

Some major disadvantages of this method are an increased visual complexity and much graphic overlapping of symbols. It should only be used with a small selection of a

Figure 3.16: Visualizing Time: Clockwise Distortion. The ring symbol as a star plot where the axes represent time, clockwise indicating the age of events of that aggregation (top right: new, top left: old).

few symbols, or after a strong aggregation of map symbol locations. Alternatively, a histogram with the temporal distribution could be created on user command based on a current selection, and appear in a pop up window.

Another remark on the use of a timeline within the map symbology: On a north-oriented map, left and right represent west and east. One has to be careful to not confuse the map representation with the way of reading a timeline, as a map is not a linear representation but a two-dimensional surface.

Bak et al. (2009) found an interesting way of showing time variations in their sensor log data: Their method of 'Growth Ring Maps' shows the cage of mice equipped with sensors that keep track of the mice's favorite locations. By mapping onto the sensor location how often a mouse comes there and at which point in time, they can visually analyze the territorial behavior of the mouse. The visualization is created using a ring symbol with a gradient filling, one side of the color ramp representing the start time of the observation period, the other end of the color ramp the end time. Every time a mouse visits a sensor, a ring is added around the symbol in the color shade according to the time of visit. The symbol 'grows' with every visit, therefore frequently visited sensors are shown with larger symbols. Since the time is linear, the center of the symbol is always shown in the shade of the 'old' side of the color ramp (left or white), with its outer rings in the 'new' shade (right or black). A variation can be seen in the mean, i.e. gray value: Did the mouse visit the sensor location regularly in the early time, but rarely towards the end? Or continuously throughout the whole period? An example of four maps showing the same cage setting but different behavior of its inhabitant can be seen in figure 3.17.

This method could be easily transformed to the visualization of news events. The *size* to be the amount of events happening at a certain location is already introduced. Now, the graphic variable *brightness* is added to indicate the age of an event, always in relation to the selected time frame. A linear gradient from bright/desaturated to dark/intense is created, the extrema representing both ends of the time frame. Every event location

Figure 3.17: Growth Ring Maps.
By Bak et al. (2009), used to analyze the territorial behavior of mice over a certain time period.

exists only with a minimum of one event. The oldest event of a location is drawn in the shade of its position on the gradient in size 1, the next event is drawn in its shade with a little increase in size, and so on.

The algorithm described in Program Design Language (PDL):

```
Sort events according to their timestamp from old to new.
Group events based on their location.
For each location:
    Draw oldest event
        The color is determined by the category,
        the brightness determined by the timestamp,
        with a ring of size 1.
    Repeat with the next event, gradually increasing size.
```

T-Pattern as Flow Chart

STempo automatically analyzes the event data for statistically significant co-occurrence patterns (so called t-pattern). These patterns are the main focus of the whole application and should therefore be visually represented in the various views. One component contains all patterns in form of a list, selecting one pattern highlights the involved events in all other components.

Once one pattern is selected and the respective events are presented on the map, it is interesting to see the sequence of those events by highlighting their temporal order with arrows, i.e. if one event B follows event A, an arrow should lead from A to B. This cartographic method of flow charts is used and discussed by Slocum (2009).

Figure 3.18: T-pattern as Flow Chart.
Concept drawing of a flow chart, as it could be used to visualize sequences in t-pattern.

3.4.4 Then Details-on-Demand

Once an overview is created and some refinements are made using zoom and filter methods, the user probably wants more detail to the events of interest. The first possibility to get more information for a detail explanation uses the mouse over functionality: a pop up window is shown when a map object is hovered, containing further information to the object like the exact number of events aggregated. To not have the info box popping up continuously when browsing the map, a timer is implemented that triggers the pop up only after some delay of a few milliseconds. The mouse has to rest on an object to have the pop up displayed. The waiting period has to be carefully balanced, between the impression of 'it is not working' when the waiting time is too long, and having pop ups appearing unwanted when the timer runs to short.

Figure 3.19: Pop Up Window.
Concept drawing of a pop up window that shows up when the cursor rests on an object.

To retrieve further details for one object, it can be clicked and selected. This selection is reflected to all visualization components and therefore reveals all the information to this object. What was a data table in earlier versions of STempo is transformed to a news feed. It still lists all events, but in a more structured and better readable way. In this news feed, all information to a single event are given, like the title, category, timestamp, and a snippet of the first few words of the report. With a selection, the respective events in the news feed are highlighted, while all others move to the background.

Developers often wish to include as much information as possible in one view, but from a user perspective it is more important to have various interaction possibilities to get a diverse view from different perspectives. Often, it is not effective to have many complex symbols on one map to compare them with each other. We should rather provide the possibility to dynamically create a complex diagram symbol based on a current selection. Here, a star plot or other glyphs could be used for multivariate visualization of either attributes or temporal aspects. Allowing the user to store this diagram also gives the possibility to create and compare several complex graphics for different selections. This way, the selection is created based on the direct interest of the user, and comparisons can be made using very flexible views. Furthermore, the complex visualization could be exported for a later report, including explanation on the underlying selection and details for further comparison.

Figure 3.20: Create Diagram using Context Menu.
Concept drawing of a context menu dialog to create a diagram.

3.4.5 Contextual Geographic Information

For many occurrences of events, some sort of contextual information is crucial for a deeper understanding of spatial patterns. Some patterns gain meaning when compared to, for example, population density or climate maps. A user might be interested in certain relationships and therefore might want to use contextual geographic information.

To be flexible for the users needs and allow the usage of a wide range of already existing maps and geographic data, Web Map Service (WMS) layers will be implemented. Web map services are provided for example by the US National Atlas[8], with a number of map layers to a variety of topics (vegetation, hydrography, transportation, natural hazards,...). Allowing the user to add and remove WMS layers as desired leaves the

[8] `http://nationalatlas.gov/infodocs/wms_intro.html`

control over the actual research direction and analysis with the user. Depending on the specific interest, a certain layer can be selected and individually combined with others.

By providing a button within the map to access the layers, a context menu could be opened that allows to select and deselect layers, change the basemap, or create new layers based on WMS URLs.

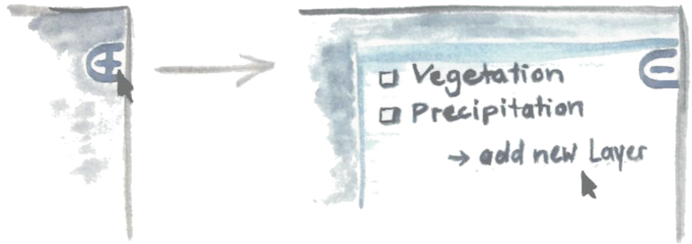


Figure 3.21: Adding Web Map Service Layers.
Concept drawing of how to access additional WMS layers in a secondary menu on the map.

3.5 Further Ideas and Considerations

The following section is a documentation of additional ideas for other components, the user interface, and the design, that came up during the work on the STempo project.

3.5.1 Improvement of the GUI Layout

The arrangement of the single components in the GUI could be changed, to make use of the space efficiently. The map should be much larger as it currently is, bringing it to the user's attention as the main component for browsing and selecting. The legend and the timeline could be aligned and combined, so the categories have to be labeled only once. The timeline could be designed more efficiently by removing the redundancy of the colored blocks and overlaying information of the same time scale. This makes the timeline smaller and gives more space to the map. The data table listing all events (as seen in figure 3.3) is only used for development and will be transformed to become a news feed, similar to a Twitter feed. The news feed will then show category, timestamp, text and source URL for the events in an aesthetic layout that allows the user to scroll the list and get an overview on selected events.

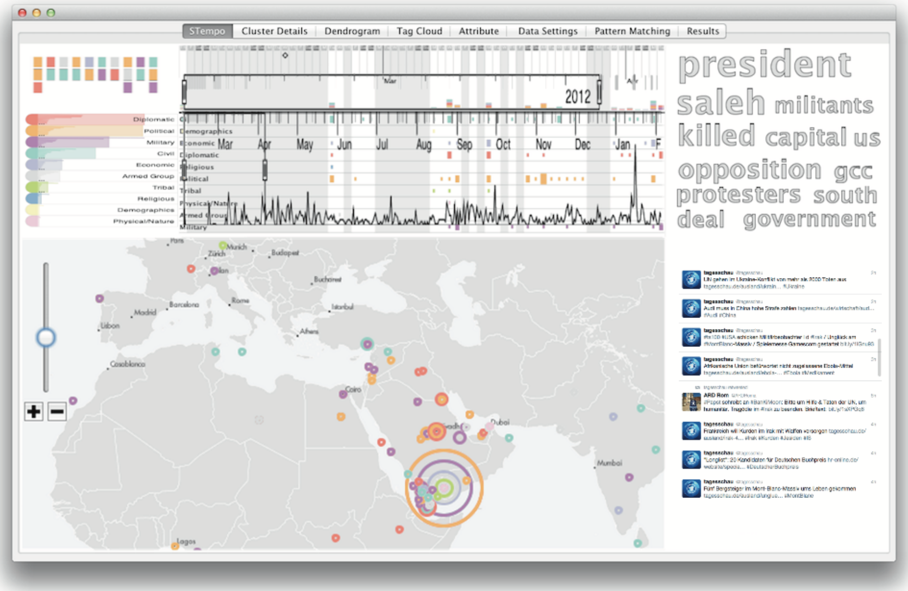


Figure 3.22: STempo GUI Mock-up.
Mock-up for an improved arrangement of components in the Graphical User Interface.

3.5.2 Graphical Suggestions for other Components

The pattern overview visualization should indicate the event types that are identified to occur in dependency. Currently, the colored squares representing an event type have the same dimension and distance horizontally as well as vertically. It is not clear, how the boxes should be read or understand. With just a small difference in the distances a grouping effect can be achieved that helps reading the boxes as groups of event types. Also, it is important to understand that the row of event types is not a sequence over time, which was mistakenly understood even by STempo team members. Rotating the whole list of patterns for 90° showing the patterns in a vertical order breaks the association with a timeline and could improve the cognition.

Figure 3.23: STempo Pattern Overview Mock-up.
Suggestion for a new style of the pattern overview
(left: current style, middle: grouping effect, right: rotated).

The world cloud currently wastes a lot of space that could be used more efficiently by squeezing smaller words in the empty space. It would interfere with the current ordering according to the size, but also makes words within one line easier to compare in their size. Furthermore, the words are currently filled according to their relevance from the bottom up. It might be a better idea to translate the amount of filling in a gray value, and color the whole inside of the word accordingly. It seems like the words are better readable when filled evenly, as can be observed in figure 3.24. The current filling method

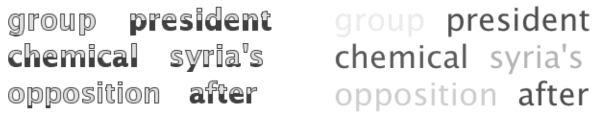


Figure 3.24: STempo Word Cloud Mock-up.
Comparison of the current style of the word cloud (left) and a suggestion for a new style with gray tone filling and thinner font (right).

is also problematic, as different words have or do not have ascenders and descenders, maybe only one or maybe many: Depending on the shape of the word itself, a different impression is created by the amount of filling. Additionally, using the gray value instead of partial filling allows the usage of different fonts. A smaller, thinner font can be used that is still easily readable. It should be a font without serifs that is simple and clear. Also, the difference in size is currently hard to see, setting the minimum and maximum size further apart would help.

3.5.3 Hardware Possibilities

Currently, STempo is developed by team members using their laptop and desktop computers. The layout and various sizes are chosen to fit on a small 13-inch display. Having more space available allows user to simultaneously access information and is therefore facilitating their mental resources. Andrews et al. (2010) argues that by organizing spatial layouts, users can externalize their insights about a dataset on the basis of the information's positions on the screen. The screen becomes 'external memory' and furthermore adds semantic meaning to the information by using spatial relationships like ordering, proximity, or alignment. Or more complex structures like clusters, or lists can be created. Using this principles, the screen is "replacing memorization and computation with perception" (Andrews et al., 2010). The user benefits by the flexibility of the layout and the evolving organization with understanding.

Thinking of latest development in technology, one can also think of a tangible display to be used for STempo, or visual analytics applications in general. A tangible or touch interface requires different design of user interaction, as for example a 'right-click' is not possible, a context menu should be called differently by e.g. using a 'long-click'.

Since tablet computers are widely used and their interface is known to many people, it might be convenient to look into details of the interaction possibilities and stick to similar conventions when developing new applications. Zoom in and out with two fingers is a very common and intuitive interaction and should be implemented with any map application. The use of a large touch display has many advantages. The user can directly and intuitively interact with data as with physical objects, which requires less training and results in faster response time. Wigdor and Wixon (2011) describe the goal of a natural user interface to lead to "unreflective expertise in its use". By offering such interface, the user does not need to think about how to use the technology, but can purely focus on his actual task, the data analysis. The need for natural interfaces to geospatial applications is also declared by Cartwright et al. (2001). According to Cartwright et al. it would make these systems accessible to a wider audience.

4 Results

The following chapter gives an overview on what is implemented in STempo already by showing the application in use with screenshots (4.1), and also on what needs to be done in short (4.2.1) and long terms (4.2.2).

4.1 STempo as of September 2014

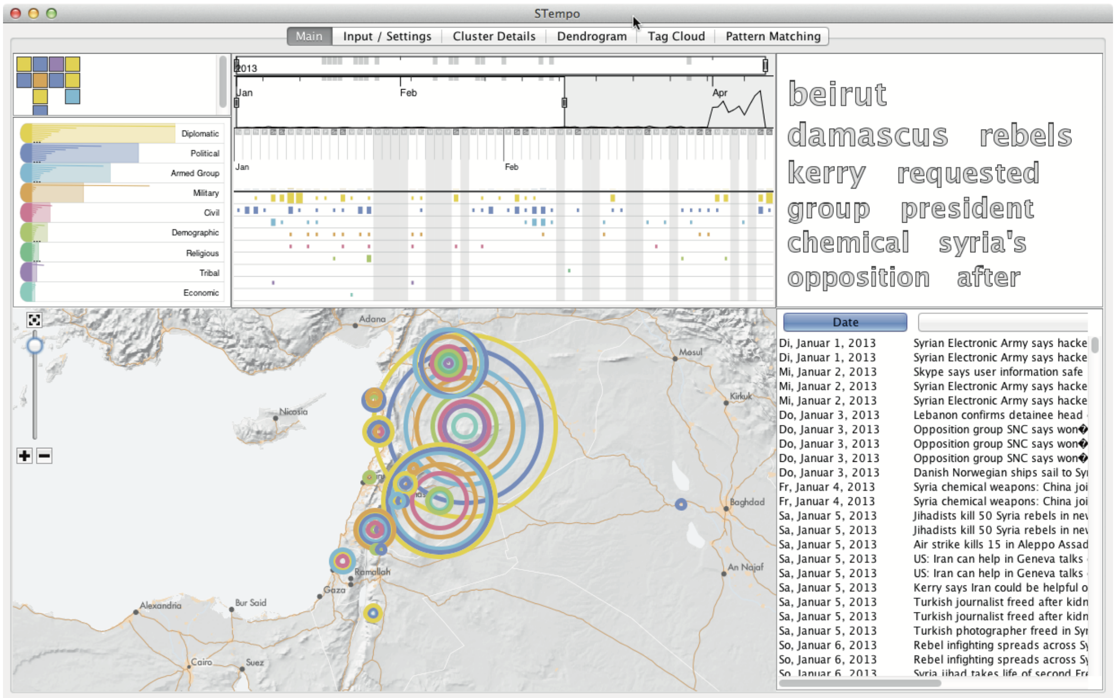


Figure 4.1: STempo Start Screen.
The STempo start screen, the view on application startup.

The layout of the main STempo view was completely rearranged to use the screen space more efficiently (see figure 4.1 in comparison to figure 3.3, page 27). Now, the map component takes a larger part of the screen as it contains the most complex visualization.

Being resized to a vertical shape, the data table is displaying more entries simultaneously. The columns are rearranged to display the relevant information first. Expected to appear in an improved and more efficient way in the near future, the space for the word cloud has been reduced already. The legend is aligned with the timeline, so the categorization can be read for both components using the labels of the legend only. The timeline is stretched to support the linear character of the time visualization, showing a broader time frame. The window for the pattern overview can be found in the top left corner, being the main component of interest for the analysis of statistical co-occurrences of events. The drawing direction of the pattern is changed from horizontal to vertical, not only to better fit in the layout but mostly to break the association with a time-sequence even some of the project team members possessed.

The first goal achieved within the STempo project map component was the implementation of a new basemap. The basemap offers spatial orientation in higher zoom level. Due to the reduced zoom level from 0 to 7, the map tiles could be included within the Java built path and directly incorporated in the runnable version. These tiles have together a size of 40,7 MB, while zoom level 8 alone has a size of 26,6 MB and zoom level 9 already 66,6 MB, likewise increasing for all further zoom level. We would like to keep the application package as small as possible, including only as many zoom level as necessary. The offline use of the tiles make the basemap extremely fast, reacting to zoom and pan interaction without any noticeable delay. A complete redraw of the tiles with a delay as it would occur using remote tiles on a server might take a second, but this short moment could already cause the user to lose orientation between the two states of display (van Elzakker et al., 2008). This problem is compensated with an extremely fast transition that will not be visible for a human's eye. Also, the application can run completely without an internet connection, which can be important for a data analyst, or even be useful for demonstration purposes.

The spatial data visualization answers the question 'what happens where, and how often?'. The symbology is connected and coordinated with the other views, dynamically changing the displayed data. It is possible to 'drill down' and get more detailed information of a map object, causing a pop up window by mouse over (see figure 4.2) or creating a selection via click. The pop up appears also over other components, like the pattern overview (figure 4.3), or the legend (figure 4.4). The pop up contains the category of the object and the number of events aggregated in that symbol. When the timeline sliders are changed, the selection is reflected to all other components including the map (figure 4.5). After the time frame of interest is chosen, unnecessary categories can be deselected by clicking their label (figure 4.6). Moving the time frame always selects *all* events within the time set, no matter which categories where selected before. Using the news feed and pop up window information, the user can drill down to understand more of the content displayed (figure 4.7), it is possible to analyze details for map symbols and objects in other components.

When the user clicks on a pattern in the pattern overview component, all respective events belonging to this pattern are selected (figure 4.8). The events are highlighted in the legend, the map, and the news feed. The words of the word cloud are filled according to their relevance in the current pattern. The timeline highlights the events on the days of occurrence, though the time sliders are not adapted to reflect the time frame of the events of that pattern. There is a button on the map interface that allows to zoom to the extent of the current selection (figure 4.9), a possible start to a detail analysis. Using this button, the user can understand the spatial extend after just one click, and in case one event is far away from the focus area it won't be missed when using the button. In another tab of the STempo application, the user can get information on cluster details (figure 4.10). The selected pattern is also highlighted in that view, a dendrogram and a tree map visualizing all patterns found by the statistical analysis of STempo.

Clicking a key word in the word cloud selects all events that relate to this term. Figure 4.11 shows the selection of 'damascus'. It reveals that most events are geo-referenced to the administrative division of Damascus ("Rif Dimashq Governorate"). But also events in other places are somehow related to the term 'damascus', mostly events of the category 'armed group', but also political, military, civil, and diplomatic. Most of these events happen in the beginning of April, but this is true for the whole data set and therefore not significant.

It is possible to analyze events based on their geo-location: Shift plus click&drag creates a rectangular selection of all map symbols within the shape, as demonstrated in figure 4.12 and 4.13. The corresponding events of selected objects are highlighted in the data table, and the bar chart visualization of the legend is adapted to reflect the current count of events in that selection. It is differentiated between a regional selection based on locations of interest, and selections based on the map symbol itself. The first process selects all events that have their center coordinate *within* the rectangular shape. This allows to limit the scope to a small area of interest (e.g. Tartus) and still retrieve all events related, even when the symbol is very large and not obviously related to the area of interest (figure 4.12). The second process allows to select several map symbols at once, to gain information about them simultaneously, which enables better comparison (figure 4.13).

To summarize, this is a list of things I achieved and implemented (some with help of Scott Pezanowski, programmer):

- creation of a basemap and implementation for offline use
- creation of a new color scheme
- map visualization to answer 'what happens where and how often?'
  - consideration of Bertin (1967) to find appropriate map symbols for the attributes category and count
  - discussion of cartographic methods
- map visualization is connected and coordinated with other views
- 'drill down' with selections and pop ups is possible
- a button zooms to the selections spatial extend
- a timed pop up shows up over map symbols

4.2 Identified Tasks To Do in the Future

STempo is an extensive project that runs already for 4 years and will continue in the future. Many of my previously described ideas are not implemented yet. But to give a summary of what, from a design perspective, needs to be done to develop STempo further, the following 'to-do lists' are created. The tasks are divided into short-term actions, smaller tasks that can be done in a short time frame, and long-term plans that are broader ideas for a possible change in the future.

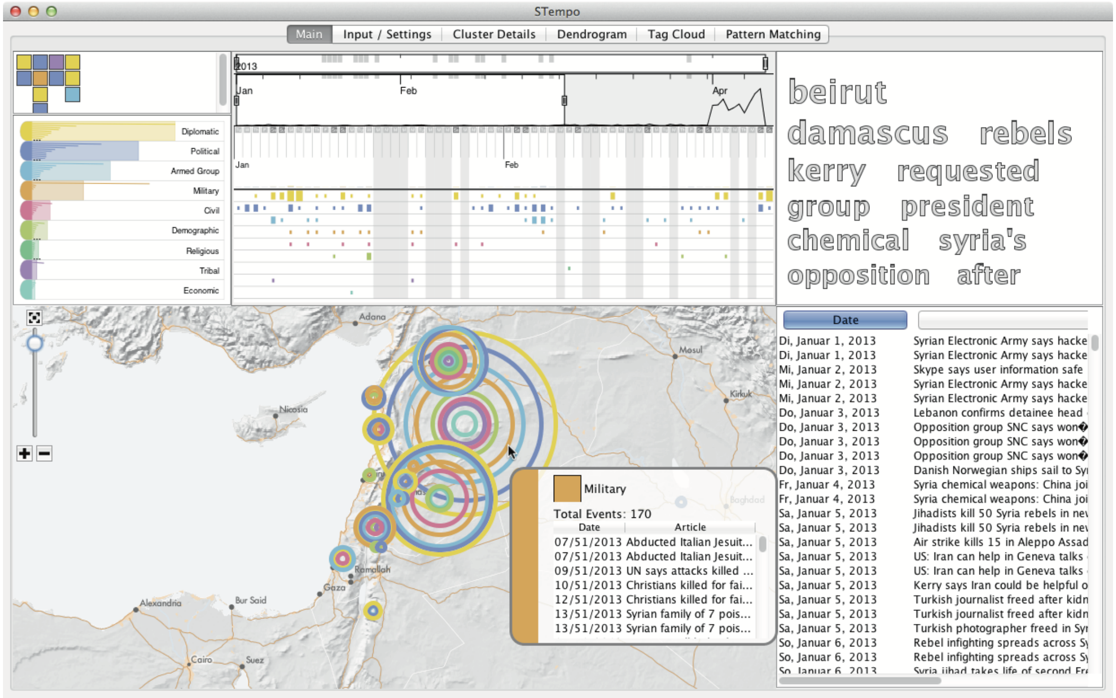


Figure 4.2: STempo Pop Up Window on Map Object.
A pop up window shows up when the mouse rests on a map object.

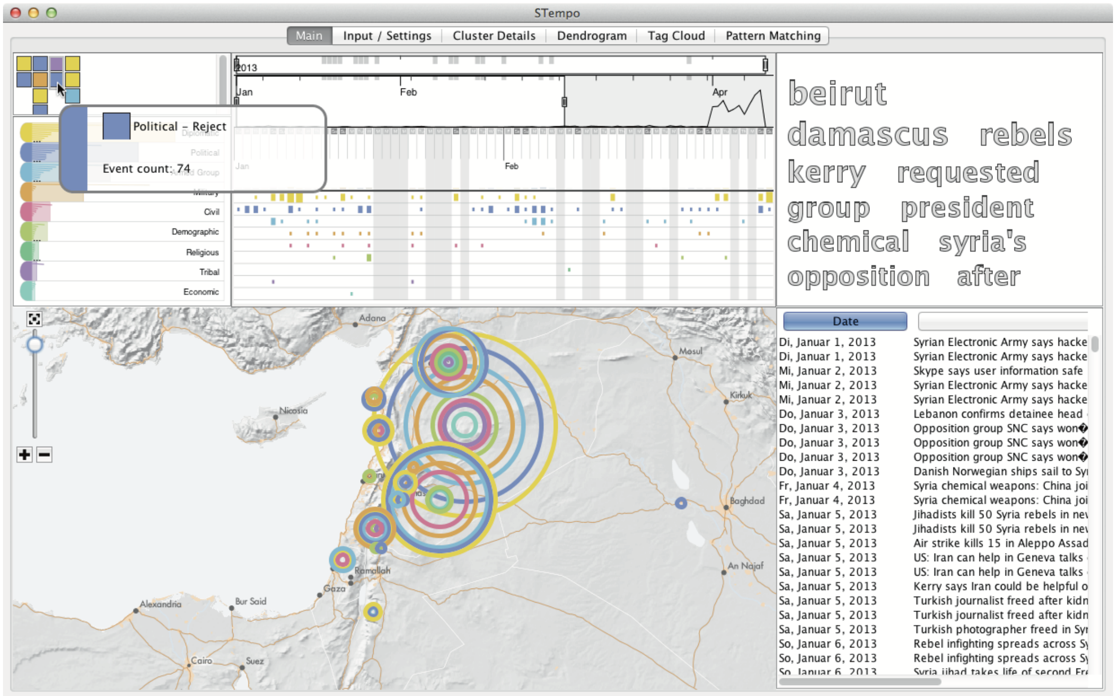


Figure 4.3: STempo Pop Up Window on Pattern.
A pop up window shows up when the mouse rests on a pattern.

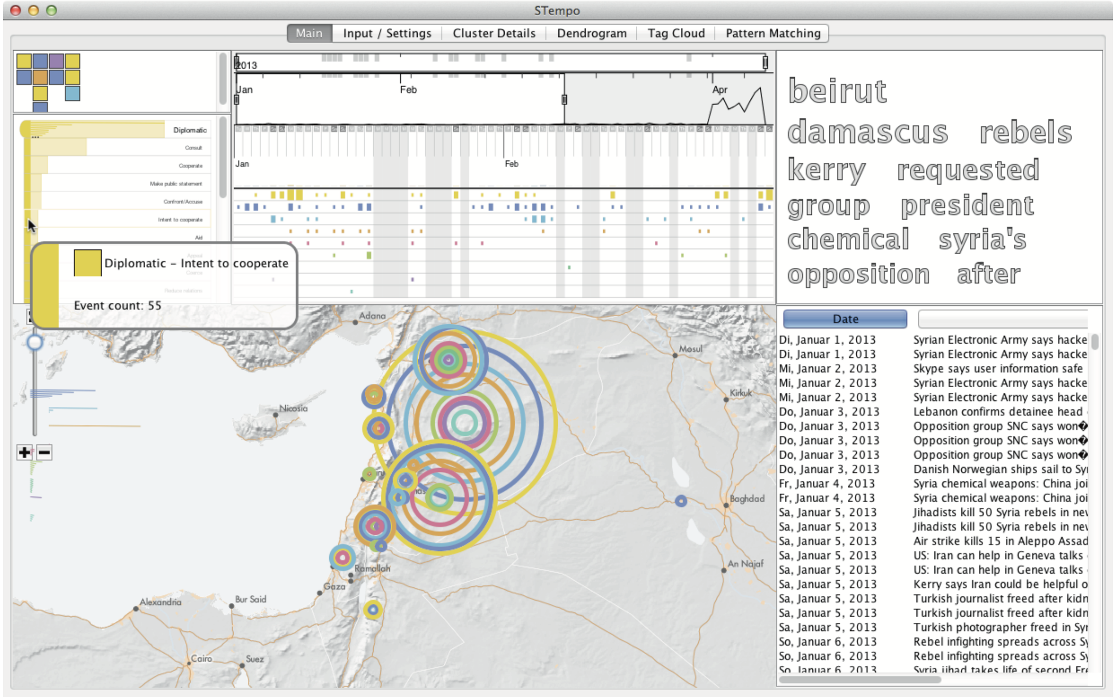


Figure 4.4: STempo Legend Detail with Subcategories.
The legend contains detail information on subcategories, opened as a drop down view.

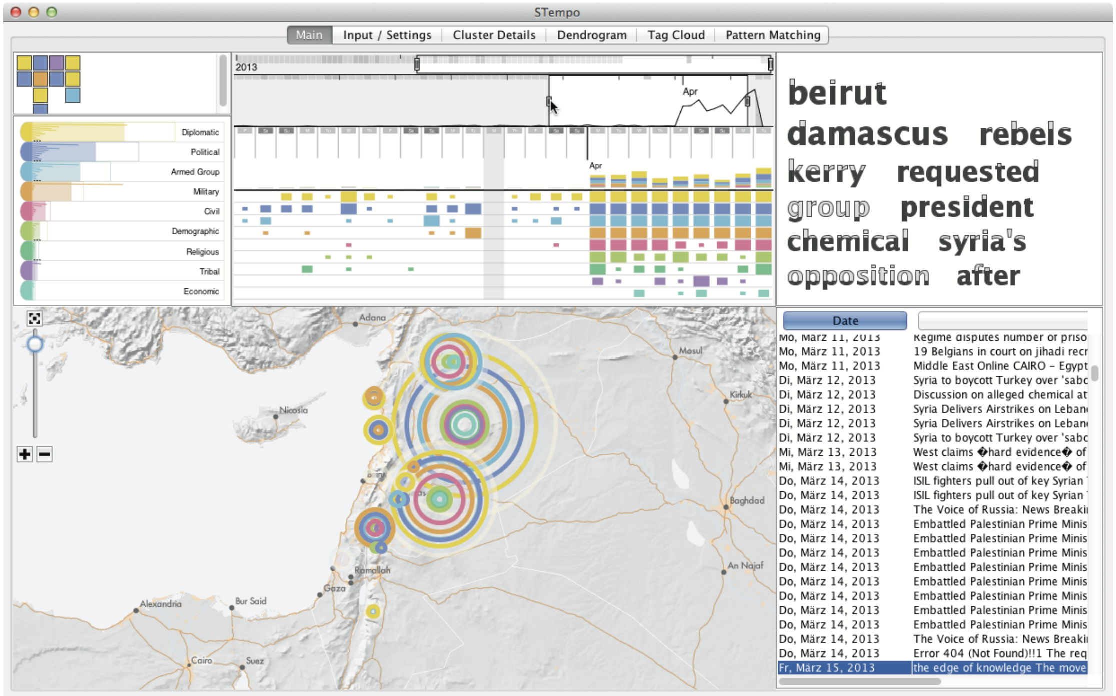


Figure 4.5: STempo Changing the Time Frame.
Changing the time frame using the handles on the slider narrows down the events selected to the specified time range.

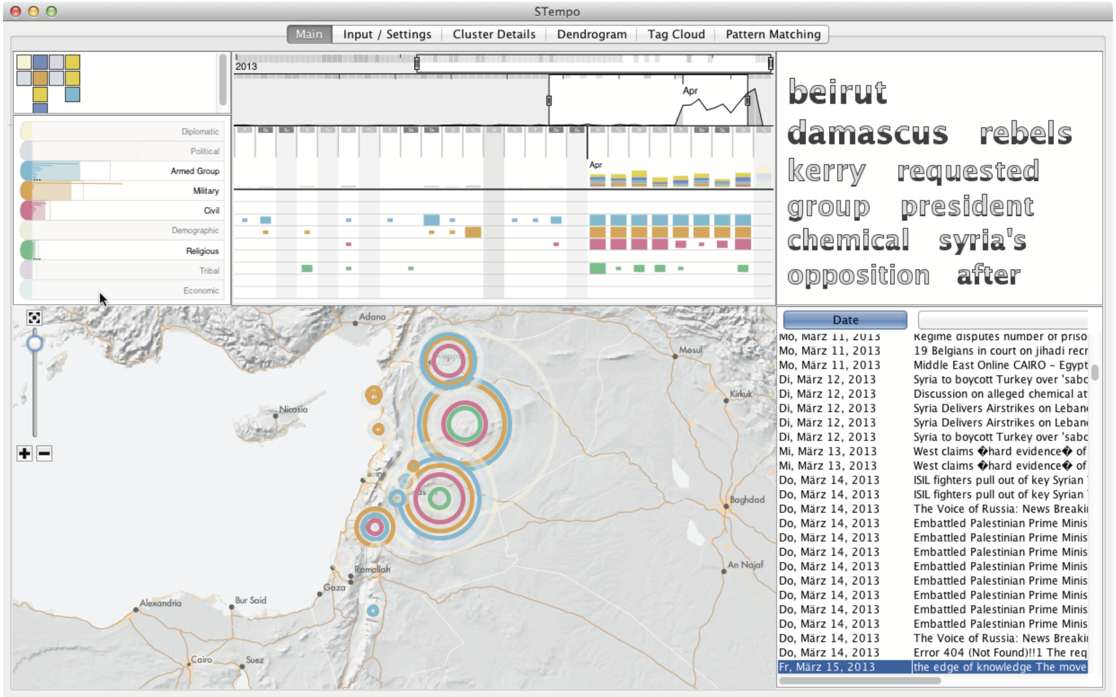


Figure 4.6: STempo Category Selection via Legend.
Categories can be disabled using the legend component, also it is possible to select one category only.

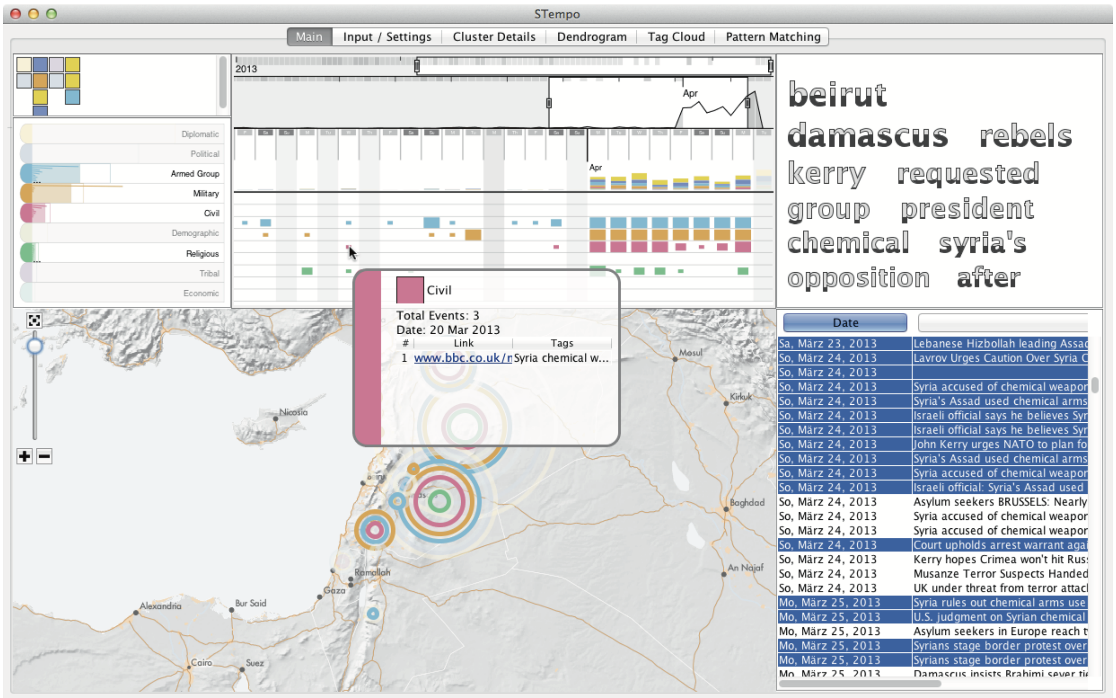


Figure 4.7: STempo Drill Down to Analyze Detail.
To retrieve detail information on the current selection, the selected events are highlighted in the news feed. A pop up opens when hovering an object with the cursor.

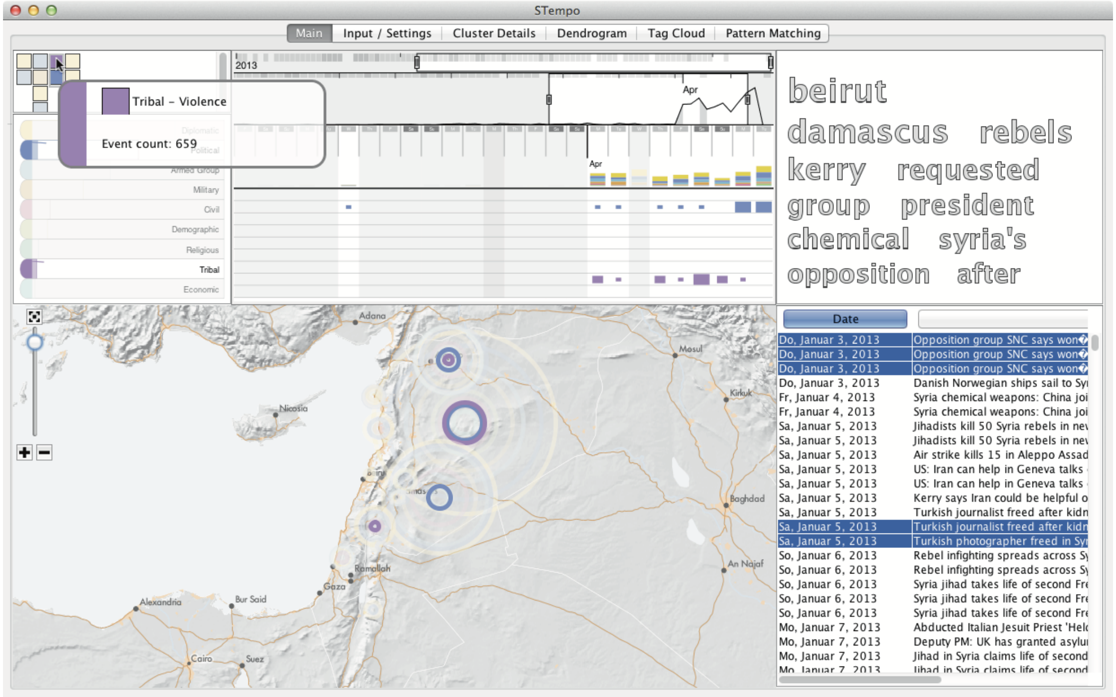


Figure 4.8: STempo Selection of a Pattern.
The user can select a pattern from the pattern overview. The related events of this pattern is highlighted as a selection in all components.

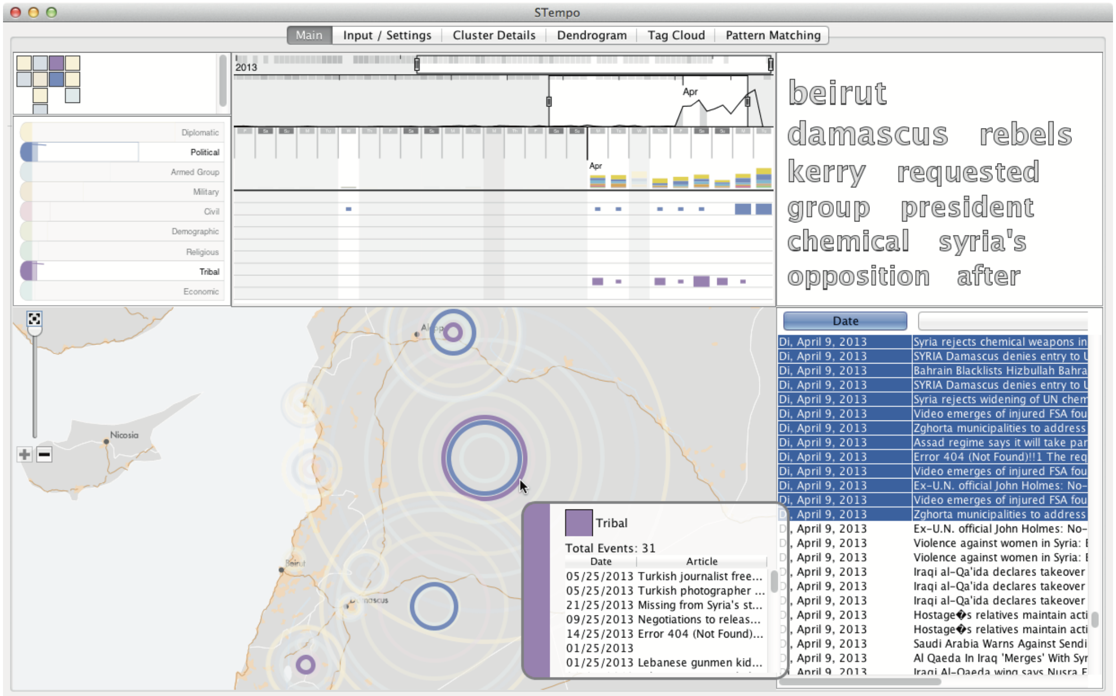


Figure 4.9: STempo A Map Button Zooms to Selection.
A button within the map (top left corner) allows the user to quickly understand the spatial extend of a pattern or selection.

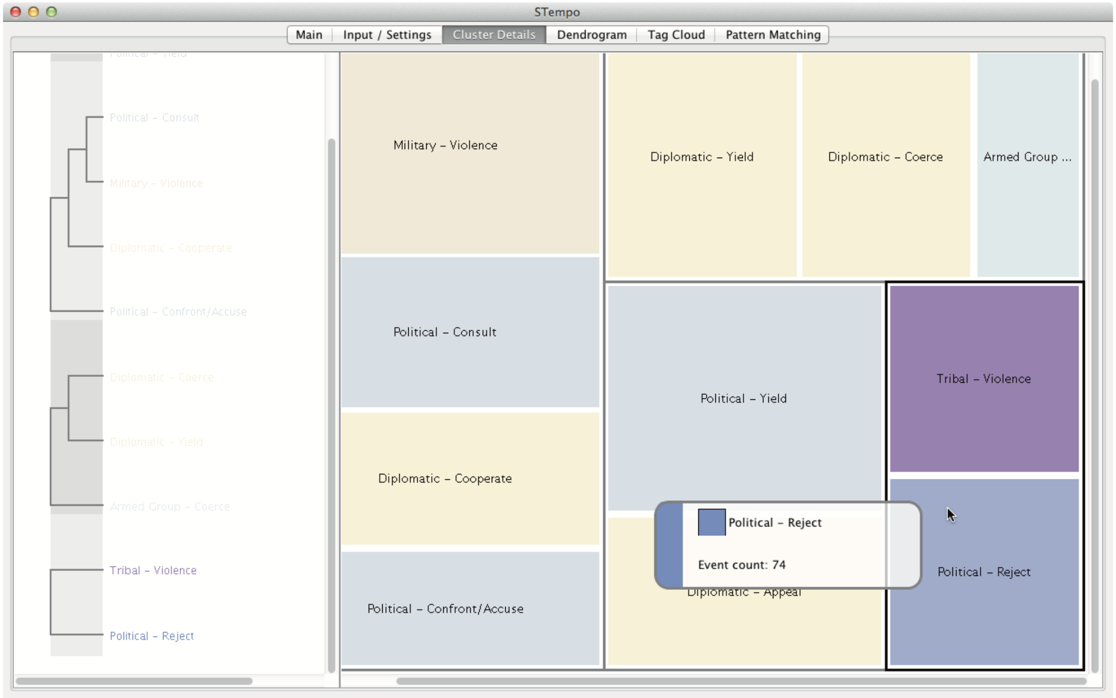


Figure 4.10: STempo Cluster Details in Separate Tab.
In a separate tab, cluster details are visualized with a dendrogram and a tree map. The selection is also highlighted in this view.

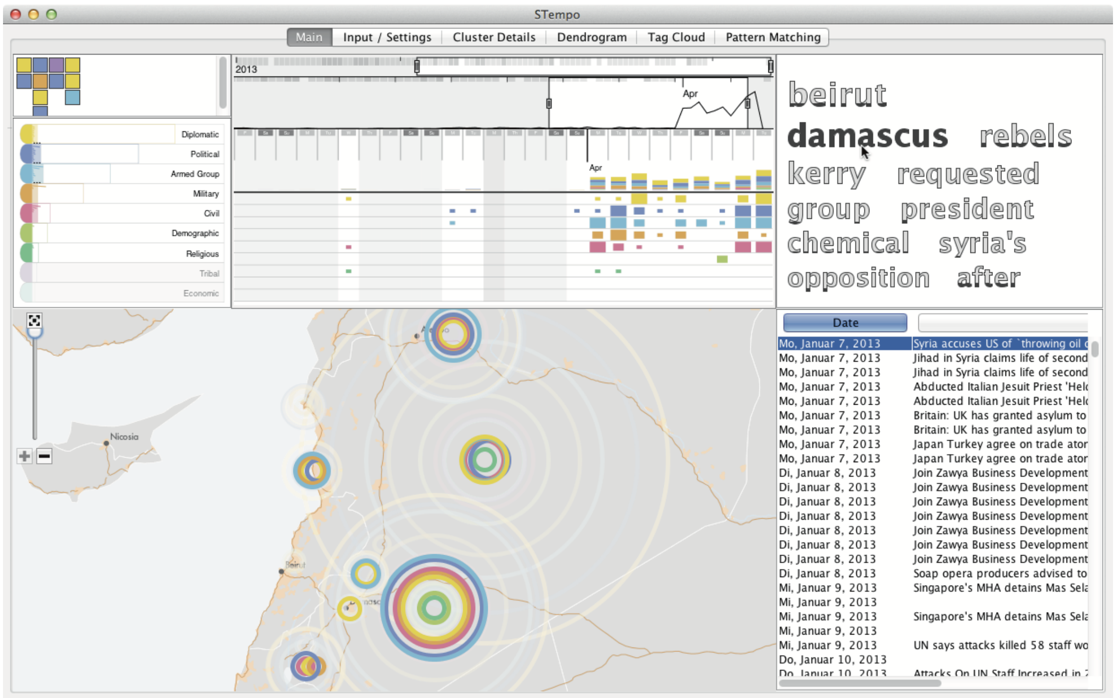


Figure 4.11: STempo Selection of a Key Word.
A word in the word cloud can be selected, which selects all related events.

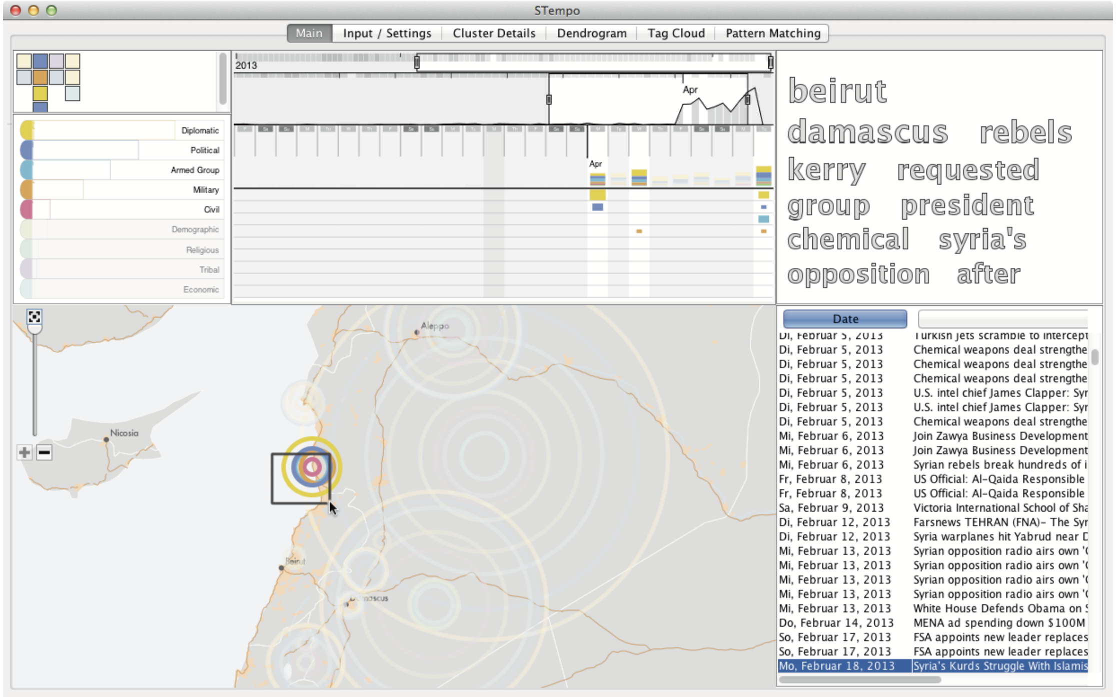


Figure 4.12: STempo Selection of a Geographic Region.
With alternative key plus click&drag, a rectangular selection of all events georeferenced to that region is made.

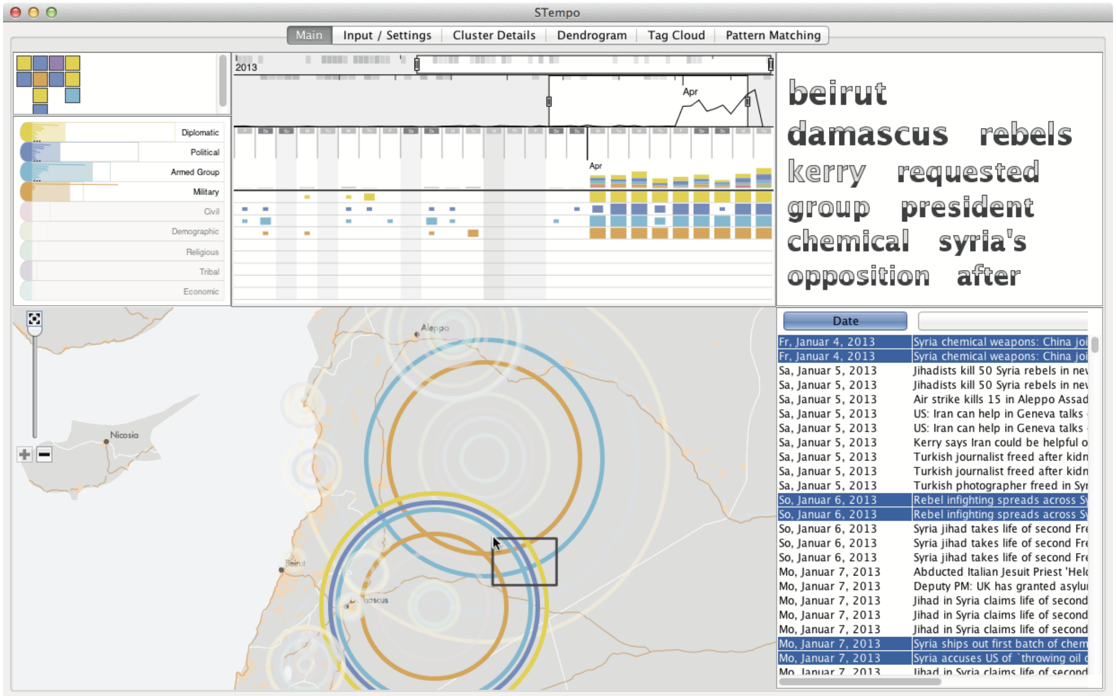


Figure 4.13: STempo Selection of Several Map Objects.
With shift plus click&drag , a rectangular selection of all map symbols intersecting the shape is made.

4.2.1 Short-Term Actions

Within the map component:

- ☐ Display a label in the center of the map ring symbol to name the location.
- ☐ Re-arrange the drawing order of map symbols: selected rings should always be drawn on top of the unselected.

Regarding the UI Design:

- ☐ Clear the pop up contents, only display necessary information.
- ☐ Move selected items to the top of any list (news feed, but also legend).
- ☐ Re-design of the word cloud, to be better readable. Also allow for scrolling or panning if not all words fit in the view.
- ☐ Adapt legend size to the timeline, in a way that the categories match in one row.
- ☐ Be sure to keep font sizes readable (minimum 8 pt).
- ☐ The patterns (pattern overview) need to be visually grouped and separated from each other.
- ☐ The news feed could contain more information (category, source URL) and should be structured to fit all information to one event in the same view, without horizontal scrolling. One event can take two or three lines of text.

The coordination between components, and user interaction:

- ☐ The time slider should adapt to a selection made in any other component, and therefore reflect the temporal extend of the selection.
- ☐ Allow the user to edit a selection: add/remove single items when holding the ctrl key.

4.2.2 Long-Term Plans

- ☐ Include more detailed zoom level to the basemap.
- ☐ Implement design ideas to visualize time & pattern within the map.
- ☐ Add functionality to include WMS layers with contextual geographic information.
- ☐ Add an additional layer with labels and borderlines on top of the data visualization, to ensure they are always readable and provide a good spatial orientation.
- ☐ Implement buttons to select all and clear the selection.
- ☐ 'Pin a selection': allow the user to set a time frame of interest and keep that frame throughout further analysis.
- ☐ Be forgiving of accidental clicks, and allow to undo actions (crtl/cmd + z).
- ☐ Brushing: Highlight items on hover throughout the whole application and all of its components.
- ☐ Functionality to create a diagram or chart from selection, allow to export the graphic.

5 Evaluation

To evaluate if the suggested visual changes of the STempo application are actually an improvement for the goal of STempo, interviews are held with 5.1) the project sponsors and main users of the system, and 5.2) with various people from different backgrounds including novices and visualization experts. We would like to find out, how well a design support higher-level thinking and analysis. All tests and interviews are based on the methods of thinking aloud and semi-structured interviewing, which are common methods in qualitative research (for example used by van Elzakker et al. (2008) to evaluate mobile geo-applications). The STempo project is still ongoing, the evaluation can therefore be considered a formative assessment. User feedback will be used for further development and implementation of new ideas and summarized to extend the list of future tasks (section 4.2).

5.1 Expert Feedback

The first feedback resulted from a meeting with the project sponsor on August 19, 2014. It has to be mentioned that the sponsor was involved in the development process from the beginning of the project in 2010 with regular meetings and discussions on the subject. The tool and its purpose are known to the sponsor, as them as experts requested features and expressed needs during all phases of development. This time, they saw a complete make-over of the user interface with a new arrangement of the components and some graphical changes. The first reaction was very positive, as the interface was described as much more aesthetic and is therefore inviting to use it rather than make the user want to “run away” (Project Sponsor, 2014).

It was striking that the main focus during the meeting was not the data visualization itself, rather than the basemap as well as contextual geographic information. The basemap being subtle and gray, containing only relevant information for orientation was mentioned positive, though some more information was requested especially regarding political situations, such as administrative borders. Also, more detail in closer zoom ranges was requested. One interesting remark was about the labeling: Currently, all map labels are part of the basemap tiles and the elements of the data visualization are plotted on top of it, hiding the labels. We thought of using a new layer containing labels only, to be placed on top of all graphics. The sponsors support the idea of implementing WMS layers, as they currently use these services in external sources to gain information about climate, landuse, vegetation, or transportation systems, etc. To incorporate these sources directly in the map component is a very important aspect for the expert user to improve their workflow. It was mentioned it would be very interesting to see how geography affects events. For example, simply said: 'do more riots occur when the weather is nice – because people tend to stay at home when it is raining?'.

The temporal aspects of the data visualization was only briefly spoken about and did not provide much feedback. An idea that came up was to include a small histogram about time variations in the information pop up, so every observed object can be analyzed by itself, but not directly compared.

The graphical suggestions for other components, i.e. a better grouping for the pattern overview, a transformation of the table in a news feed, and especially the space-saving design of the word cloud was welcomed and supported.

5.2 Semi-structured Interviews

Semi-structured interviews were conducted with test persons from various fields of expertise. It gives an idea of how possible users use the application, what are their experiences with it, and eventually, what is their opinion. The main focus is on first reaction of the participant, does he understand what he sees even without explanation, what are the first conclusions. People with a background in visualization as well as completely without any prior knowledge will be asked, and differences in their understanding are examined.

For the test, a MacBook Pro running the STempo application as a runnable version from September 4, 2014 will be used. In total 12 participants (6 male) from different backgrounds were interviewed: 5 cartographers, 4 civil engineers, 1 mountain guide, 1 economics professional, 1 economics engineer. Each participant had experience with at least basic visualization tools like Excel charts. The interviews were conducted with single participants or in groups of two. The computer was handed to the participant and the application started. Without any explanation or guidance, their behavior and interaction with the application is examined. The participants, in some cases two participants simultaneously, are asked to speak about what they see, and what they think about it. Active “playing” with the application is encouraged. At first, the role of the observer is limited to note taking only, the communication should only take place in a monologue of the participant or dialog between both participants to reduce any possible influence. That way possible bias by the observer shall be reduced. It is interesting to see at what point a participant can not continue with her analysis, and for what reason. Eventually, latest when the participant does not know any further, specific questions like where do most events happen, what category of events is occurring most in a certain area, what is the spatial extend of a pattern, at what times occur events in a certain area, will be asked. It was noted that groups of two participants are easier to test, as they can interact with each other and discuss problems and ideas. Also it is easier for them to ‘think aloud’ as they can tell their thoughts to another person instead of just themselves, which leads to more creativity in discussing the application and its possible usage. As there was no fixed set of questions or fixed structure of the test, every interview was different from the others. Sometimes, issues were discussed in detail that were not mentioned at all during other sessions with different participants. With the first interviews, many new ideas were formed. The later interviews were then influenced by previous insights from the first sessions, so more specific questions could be asked.

This experiments reveal interesting usability issues.

5.2.1 Without Instructions

All interviews started without instructions on the application itself, no background information was given on the topic or purpose. Participants were invited to run the application and see if they get any idea about what they see, and if they can learn anything.

Many participants started reading some labels and words they could find, which gave them a brief overview on what is the main topic. Relating the key words to the map, for most participants it was clear that the task is related to crisis and conflicts in Syria. The list of reports in the table view sounds like news reports to most participants. While some participants started clicking on objects anywhere, or went strait to the next tabs, half of the participants used the map at first. The ring visualization attracts a lot of attention. After just a short time, all participants understood that the ring size represents the count of events taking place at this location, while the color is relating to the topic given in the legend. That all components are connected and show the same data in different views was naturally and intuitively clear to everyone.

While people found the explanation of the ring size in the pop up window (event count), 8 of 12 participants questioned if the ring might refer to the area covered. Some were reminded to seismologic maps visualizing earth quakes. For others, it was obvious that the ring is placed on the location of interest and the size just represents a quantity. But many participants had trouble to find the corresponding location name (at least 5), as for example the largest ring symbol is placed in the middle of Syria where no city can be found. Only one guessed that the symbol relates to the whole country. Eventually, when asked to point the area with most e.g. military events, all participants were able to answer correctly. It was noted by half of the participants that they need more detail on the basemap to understand the spatial context. They wished to see more administrative borders, and more cities.

Many problems came up with the timeline: Only 4 participants used the handles on the time slider to change the time frame. All other participants did not change the time at all, and did not even realize it was possible. The time slider is set to a certain time frame on application start up not representing the actual time frame of the full, and at this point selected, data set. Even when changing the selection in any component, the time sliders do not adapt – they do not change at all, unless actively moved by the user. This was considered problematic during the interviews, as people did not understand this behavior. In general, any text was very important for the participants to create some idea about the subject. While only 5 participants understood the word cloud as important key words of the news reports, and its filling has some relation to the current selection, it gave an idea of the subject to most people with only its presence in terms like ‘damascus’, ‘rebels’, or ‘syria’. Also the first few words of the news reports gave an impression on the theme, as well as the categories in the legend and their relative importance given by the number of occurrences (all participants understood the bar chart in the legend as such). The pop up window as a source for more detailed information was found and used by all participants. However, many complained about the pop up showing up only sometimes, or not at all, or for the wrong object. It was noted that in the timeline, when hovering a colored box representing events at that day a pop up shows up. But to get a new pop up for a different object, the cursor first has to leave any object and than return to the new. As long as one pop up is still open, a new pop up would not appear. At least 9 of 12 participants were annoyed by the pop up and its malfunction. While all participants found that they could click an object to select it, or click on a word in the word cloud to select all related objects, 9 participants had trouble controlling the selection according to their wish. They either accidentally clicked somewhere and the previously made selection was gone, in which case they wished to have an undo button (crtl/cmd + z). Or they wanted to select all or none of the objects, or at least add some objects to a selection. No one guessed that they could create a multi-selection by using an additional key. Also, none of the participants used the button to zoom to the selection. Other issues people had were with reading the labels in the legend (too small) and finding the possibility to unfold the legend (not indicated). Also, 3 participants stumbled upon multiple entries of the same news report, that was confusing: Do multiple entries of the very same report still increase the size of a symbol?

Table 5.1 shows the main insights gained during the interviews summed up in statements, as well as the number of participants agreeing to a statement. The number of agreements is not necessarily a true value, as sometimes issues did not came up during a session, so the respective question was not even asked. The semi-structured methodology has to be kept in mind when reading the table: Some statements were only created in the later interviews, and could therefore not be evaluated for earlier sessions.

Many participants felt frustration as they did not understand what they should do with the application, or had the impression the system is not responsive and behaves unexpectedly. When the participant was stuck and did not know any further things to do, specific questions were asked, for example "can you tell me where the most military events happen?", or "what type of event is most prominent in the north?", or "what time was most active in terms of event occurrences?". This kind of questions could be answered correctly by every participant.

5.2.2 After Introduction and Guidance

When the participants had the impression that they discussed all parts of the application, and all specific questions regarding the understanding of visualizations were asked, a short introduction to STempo and its ideas was given. Since it is impossible to learn from the user interface only, the idea of the pattern and sequence analysis was new to all participants. After explanation, most participants agreed with the ideas and thought of STempo as an useful tool. But training is definitely necessary to successfully use the application for knowledge construction and analysis. While the discussion came up only with one group of participants, they strongly agreed on the importance of a spatial query (select all symbols with their center coordinate in the rectangular selection box), rather than select several ring symbols. When the word cloud was discussed, the layout and also the filling was considered unattractive by participants, they welcomed the idea of using a gray tone instead of partial filling and also preferred to have more words in the view rather than a few huge letters. While half of the participants thought of the timeline in different time scales is a good idea, the other half did not like it and found it confusing. The sub-timelines should be clearer separated from each other, so the user realizes there are different scales.

Though the overall design and choice of colors was positively mentioned several times, it was also said that there is much work required in improving the user experience and interaction. Especially for functionalities like the pop up window or multiple selection, consistency with standard operating systems is requested. It seems very important that the application performs reliable, steady and responsive.

Table 5.1: Result of the Interview Sessions

Remark: Not all statements were tested in every interview. These statements were formed afterwards based on notes taken during all interview sessions. A checkmark is set when the behavior of the participant revealed the confirmation of a statement.

Number of Participants:	**2**	**2**	**1**	**1**	**1**	**1**	**1**	**2**	**1**	**groups 9**	**in total 12**
Understood topic (news visualization)	✓	✓	✓		✓		✓	✓	✓	7	10
Understood coordination between views	✓	✓	✓	✓	✓	✓	✓	✓	✓	9	12
Used map at first	✓		✓	✓	✓		✓			5	6
Understood ring size & color	✓	✓	✓	✓	✓		(✓)	✓	✓	8	11
— also saw ring related to area	✓		✓	✓	✓		✓	✓		6	8
Able to answer map questions (where...?)	✓	✓	✓	✓	✓	✓	✓	✓	✓	9	12
Needs more detail on the basemap	✓		✓		✓		✓		✓	5	6
Not sure what location symbol refers to (missing label or hint in pop up)	✓		✓					✓		3	5
Somehow understood the pattern overview as a subset							✓			1	1
Discovers inconsistencies in selection and interaction	✓	✓			✓		✓		✓	5	7
— was annoyed by pop up (especially in timeline)	✓	✓		✓	✓			✓	✓	6	9
Used and understood word cloud	✓		✓	(✓)					✓	4	5
Used the slider to change time frame	✓	✓								2	4
— was confused by slider not adapting to changes							✓	✓		2	3
Wishes to select all/none /add to selection or undo selection	✓	✓	✓	✓	✓	✓	✓			7	9
Wants spatial query instead of box selection								✓		1	2
Trouble reading font size in legend	✓			✓						2	3
Was confused by double entries (they bias count!)							✓	✓		2	3
Wants larger screen					✓					1	1
Positively mentioned design, colors			✓	✓	✓			✓	✓	5	6
Gave up due to frustration		✓	✓		✓	✓			✓	5	6
Eventually, all was fine and the basics understood, but what is the usage?	✓			✓			✓	✓		4	6

6 Conclusion

The work with this thesis focused on the improvement of the visual analytics application STempo, considering cartographic rules as well as lessons from HCI research. This work shows exemplary that any map in a visual analytics application does benefit from cartographic knowledge, the thoughtful use of guidelines does make it more attractive and therefore easier to use and understand. Many design choices do not relate to the map component only, but can also be translated for other components and improve the whole user interface experience. As presented in section 2.2.2 (see also figure 2.1), the user-centered design cycle is an iterative process where this work would be one iteration. Ways are presented to visualize spatio-temporal news data, considering the frame settings of STempo and its data generation process. The evaluation shows that the user needs some textual information to understand basics, as it is given in every map with a title and legend text. The design of colors and the choice of cartographic methods should support an intuitive understanding of the underlying message. Furthermore, the user needs a consistent interaction functionality and a responsive system to be able to work with it.

The design implemented so far was already a big improvement compared to the look of STempo before (see figure 3.3 on page 27). A new basemap was created to support spatial awareness and provide orientation. Because the basemap was firstly built for another data set that was more globally distributed, less detail in higher zoom level was required at first. For the future, additional zoom level with more detail are necessary. Cartographic methods for the data visualization within the map component were discussed, and the proportional ring symbol created. The symbol still needs a label, to allow the user to directly read the location name it is representing. The user interaction is coordinated with the other components, trying to be consistent within the application and known conventions in interaction design.

That it actually is an improvement was confirmed by team members, as well as participants of the evaluation when asked for their opinion on the differences. The identified tasks for future development (section 4.2) will hopefully provide some useful input for the STempo team as they will continue their work. However, the application still requires a lot of work to be done: implementing the 'to-do'-s at first to get a stable beta-version of STempo including all desired functionalities, than again run another evaluation to test the usability.

Regarding the evaluation conducted within this thesis, there is certainly room for improvement in future user testing. Rather than having a verbal discussion and interview sessions with participants, a structured test using screen logging methods and a usability survey in which the participants rate their experience on a numerical scale would give more detailed insight on the user experience. For this sort of test, a stable version of STempo is required, eliminating at least major usability issues identified in this thesis (see section 4.2.1). It is important to be able to access the actual news story, to understand an event itself.

This work has shown how a useful tool to analyze large amounts of news data could look like. With its multiple coordinated views, STempo gives a complex and versatile perspective on events around the world and in areas of interest. It has high potential in being very interesting tool for intelligence analysis, as the large body of news events is visualized for a broad overview, but drilling down to details is easily possible. While similar applications dealing with the visualization of news events often focus on either spatial, or temporal, or thematic aspects (see section 2.3), STempo combines several approaches. By incorporating thematic, spatial and temporal aspects, as well as the statistical analysis of t-pattern and frequent key words, STempo combines many useful features in one tool.

Finally, it has to be said that even though a coordinated multiple view application is quite common practice in complex interactive visualizations (see also section 2.3), this technique also has limitations regarding the amount and complexity of the data (Andrienko et al., 2010). According to Andrienko et al., more research has to be done in the related disciplines to integrate approaches from visualization, cognition, statistics, or data mining.

Bibliography

Wolfgang Aigner, Silvia Miksch, Wolfgang Muller, Heidrun Schumann, and Christian Tominski. Visual methods for analyzing time-oriented data. *IEEE Transactions on Visualization and Computer Graphics*, 14(1):47–60, 2008.

Christopher Andrews, Alex Endert, and Chris North. Space to think: large high-resolution displays for sensemaking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 55–64. ACM, 2010.

Gennady Andrienko and Natalia Andrienko. Interactive maps for visual data exploration. *International Journal of Geographical Information Science*, 13(4):355–374, 1999.

Gennady Andrienko, Natalia Andrienko, Piotr Jankowski, Daniel Keim, M-J Kraak, Alan MacEachren, and Stefan Wrobel. Geovisual analytics for spatial decision support: Setting the research agenda. *International Journal of Geographical Information Science*, 21(8):839–857, 2007a.

Gennady Andrienko, Natalia Andrienko, and Stefan Wrobel. Visual analytics tools for analysis of movement data. *ACM SIGKDD Explorations Newsletter*, 9(2):38–46, 2007b.

Gennady Andrienko, Natalia Andrienko, Urska Demsar, Doris Dransch, Jason Dykes, Sara Irina Fabrikant, Mikael Jern, Menno-Jan Kraak, Heidrun Schumann, and Christian Tominski. Space, time and visual analytics. *International Journal of Geographical Information Science*, 24(10):1577–1600, 2010.

Natalia Andrienko and Gennady Andrienko. *Exploratory analysis of spatial and temporal data*. Springer, 2006.

Natalia Andrienko and Gennady Andrienko. Visual analytics of movement: An overview of methods, tools and procedures. *Information Visualization*, 12(1):3–24, 2013.

Peter Bak, Florian Mansmann, Halldor Janetzko, and Daniel A Keim. Spatiotemporal analysis of sensor logs using growth ring maps. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):913–920, 2009.

Jacques Bertin. *Sémiologie Graphique*. Paris, Gauthier-Villars, 1967. Trans. as Semiology of Graphics (1983). Wisconsin: University of Wisconsin Press.

Jacques Bertin. *Semiology of graphics*. University of Wisconsin Press, 1983. ISBN 9780299090609.

Cynthia Brewer. *Designing better maps: a guide for GIS users*. Esri Press, 2005. ISBN 9781589480896.

BT website. Visibility of your security threats with assure analytics. `http://www.globalservices.bt.com/uk/en/virtualshowcase/visibility_of_your_security_threats_with_assure_analytics`, 2014. [Online; accessed 30-July-2014].

Aileen Buckley. Guidelines for the effective design of spatio-temporal maps. In *Proceedings of the 26th International Cartographic Conference*, 2013.

Stuart Card, Jock Mackinlay, and Ben Shneiderman. *Readings in information visualization: using vision to think*. Morgan Kaufmann, 1999.

William Cartwright, Jeremy Crampton, Georg Gartner, Suzette Miller, Kirk Mitchell, Eva Siekierska, and Jo Wood. Geospatial information visualization user interface issues. *Cartography and Geographic Information Science*, 28(1):45–60, 2001.

Junghoon Chae, Dennis Thom, Harald Bosch, Yun Jang, Ross Maciejewski, David S Ebert, and Thomas Ertl. Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition. In *Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on*, pages 143–152. IEEE, 2012.

Chris Crawford. *The Art of Interactive Design: a euphonious and illuminating guide to building successful software*. No Starch Press, 2002. ISBN 9781886411845.

Alex Endert, Lauren Bradel, and Chris North. Beyond control panels: Direct manipulation for visual analytics. *IEEE Computer Graphics and Applications*, 33(4):6–13, 2013.

Martin Erwig. Toward spatio-temporal patterns. In *Spatio-Temporal Databases*, pages 29–53. Springer, 2004.

Lisa Graham. *The Principles of Interactive Design*. Delmar Publishers, 2nd edition, 1999. ISBN 9780827385573.

Mordechai Haklay. *Interacting with geospatial technologies*. John Wiley & Sons, 2010. ISBN 9780470998243.

Frank Hardisty and Anthony Robinson. The geoviz toolkit: using component-oriented coordination methods for geographic visualization and analysis. *International Journal of Geographical Information Science*, 25(2):191 – 210, 2011. ISSN 13658816.

Mark Harrower and Cynthia Brewer. Colorbrewer.org: an online tool for selecting colour schemes for maps. *Cartographic Journal, The*, 40(1):27–37, 2003.

Robert Ghanea Hercock. Technology strategy board - innovateuk; workshop on assure analytics. `https://connect.innovateuk.org/c/document_library/get_file?groupId=6131285&folderId=8696610&title=Robert+Ghanea+Hercock.pptx`, 2013. [Online; accessed 30-July-2014].

Andreas Holzinger. Lecture 9 interactive information visualization and visual analytics. In *Biomedical Informatics*, pages 379–420. Springer, 2014.

Daniel Keim, Gennady Andrienko, Jean-Daniel Fekete, Carsten Görg, Jörn Kohlhammer, and Guy Melançon. Visual analytics: Definition, process, and challenges. In Andreas Kerren, JohnT. Stasko, Jean-Daniel Fekete, and Chris North, editors, *Information Visualization*, volume 4950 of *Lecture Notes in Computer Science*, pages 154–175. Springer Berlin Heidelberg, 2008. ISBN 978-3-540-70955-8. doi: 10.1007/978-3-540-70956-5_7.

J.M. Kraak and A. Brown. *Web Cartography*. Taylor & Francis, 2003. ISBN 9780203305768.

Menno-Jan Kraak. The space-time cube revisited from a geovisualization perspective. In *Proceedings of the 21st International Cartographic Conference*, pages 1988–1996, 2003.

Menno-Jan Kraak. *Mapping Time: Illustrated by Minard's Map of Napoleon's Russian Campaign of 1812*. Esri Press, 2014. ISBN 9781589483125.

Miloš Krstajić, Enrico Bertini, Florian Mansmann, and Daniel Keim. Visual analysis of news streams with article threads. In *Proceedings of the First International Workshop on Novel Data Stream Pattern Mining Techniques*, pages 39–46. ACM, 2010.

John Krygier and Denis Wood. *Making Maps: A Visual Guide to Map Design for GIS*. Guilford Press, 2011. ISBN 9781609181666.

Vitaveska Lanfranchi, Suvodeep Mazumdar, and Fabio Ciravegna. Visual design recommendations for situation awareness in social media. In *Proceedings of the 11th International ISCRAM Conference*, May 2014.

Paul Longley, Mike Goodchild, David Maguire, and David Rhind. *Geographic Information Systems and Science*. John Wiley & Sons, 2010. ISBN 9780470721445.

Alan MacEachren. Visualizing uncertain information. *Cartographic Perspectives*, 0(13), 1992.

Alan MacEachren. *How maps work: representation, visualization, and design*. Guilford Press, 2004. ISBN 9781572300408.

Alan MacEachren. Cartography as an academic field: A lost opportunity or a new beginning? *The Cartographic Journal*, 50(2):166–170, 2013.

Chris North and Ben Shneiderman. Snap-together visualization: Evaluating coordination usage and construction. *Int'l Journal of Human-Computer Studies special issue on Empirical Studies of Information Visualization*, 53(5):715–739, 2000.

Donna Peuquet. *Representations of space and time*. Guilford Press, 2002.

Donna Peuquet. Personal Conversation, July 22 2014.

Donna Peuquet and Sam Stehle. Validation of results for temporal patterns derived in STempo. In *Extended Abstract Proceedings of the GIScience, Vienna, AT*, September 2014.

William Pike, John Stasko, Remco Chang, and Theresa O'Connell. The science of interaction. *Information Visualization*, 8(4):263–274, 2009.

Peter Pirolli and Stuart Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of the International Conference on Intelligence Analysis*, volume 5, pages 2–4. Mitre McLean, VA, 2005.

Project Sponsor. Personal Conversation, August 19 2014.

William Ribarsky, Brian Fisher, and William M Pottenger. Science of analytical reasoning. *Information Visualization*, 8(4):254–262, 2009.

Anthony Robinson, Frank Hardisty, Sam Stehle, and Donna Peuquet. Geovisualization of event sequences with t-pattern analysis in STempo. In *GeoViz_Hamburg*, March 6-8 2013.

Anthony Robinson, Frank Hardisty, Sam Stehle, Wei Luo, and Donna Peuquet. Geovisualization of t-pattern event sequences in STempo. In revision, 2014.

Robert Roth. Interactive maps: What we know and what we need to know. *Journal of Spatial Information Science*, (6):59–115, 2014.

Philip Schrodt, Ömür Yilmaz, Deborah Gerner, and Dennis Hermreck. The cameo (conflict and mediation event observations) actor coding framework. In *2008 Annual Meeting of the International Studies Association*, 2008.

Ben Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the IEEE Symposium on Visual Languages*, pages 336–343. IEEE, 1996.

Ben Shneiderman and Catherine Plaisant. *Designing the User Interface: Strategies for Effective Human-Computer Interaction (4th Edition)*. Pearson Addison Wesley, 2004. ISBN 0321197860.

Terry Slocum. *Thematic Cartography and Geographic Visualization*. Prentice Hall series in geographic information science. Pearson Prentice Hall, 2009. ISBN 9780132298346.

Brian Swedberg, Anthony Robinson, Frank Hardisty, and Donna Peuquet. Geovisualization of spatio-temporal events in STempo. In *Workshop at GIScience, Vienna, AT*, September 2014.

James Thomas and Kristin Cook. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE Computer Society Press, 2005. ISBN 9780769523231.

Edward Tufte. *The Visual Display of Quantitative Information*. Graphics Press, 2nd edition, 2001. ISBN 9780961392147.

Corné van Elzakker and Karen Wealands. Use and users of multimedia cartography. In *Multimedia cartography*, pages 487–504. Springer, 2007.

Corné van Elzakker, Ioannis Delikostidis, and Peter van Oosterom. Field-based usability evaluation methodology for mobile geo-applications. *The Cartographic Journal*, 45(2): 139–149, 2008.

Irina Vasiliev. Mapping time. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 34(2):1–51, 1997.

Daniel Wigdor and Dennis Wixon. *Brave NUI World: Designing Natural User Interfaces for Touch and Gesture*. Elsevier Science, 2011. ISBN 9780123822321.

Appendix

Personal Conversation (notes) with Donna Peuquet, Principal Investigator of the STempo Project

on July 22, 2014

What is the Purpose of STempo?

STempo is meant as a general purpose tool, for example to examine the spread of a disease, world economics, and others... It addresses more expert users than the public. STempo could be used un unravel big data. It visualizes many events to show patterns. First step is to discover patterns, the second step drills down to get details. Once an important pattern is identified, for example one that indicated political instability, it can be used for prediction. There are indicator events in associate patterns, this event is always participating. Predicator event is when the occurrence of the event is an indicator.

What is the role of the map component?

In the triad of visual representation (What? When? Where?), the map answers one important part. It helps people to understand space. The map is used to explore interrelatedness, to browse the data at once or with selection/filtering in multiple modes. At the same time the map shows the result of a selection, for example the members of a probability based pattern.

The user is expected to drill down, understand and request more detail.

Important lesson: The t-pattern is not a sequence! In its overview window it just gives and idea on the overall categories involved in that pattern, not in any particular order.

Meeting (notes) with Project Sponsor of the STempo Project

on August 19, 2014

General Feedback

- great visual improvement
- much more aesthetic
- invites for exploration rather than make the user want to run away
- love the style of the new base map
- sequence grouping should be optimized (reading direction left to right, therefore preferred grouping)
- ring symbol could mistakenly refer to area covered instead of representing a quantity
- base map needs more detail in higher zoom level, political boundaries and administrative areas are of interest, relief is very good (beautiful, elegance, grace), labels should not be hidden by the data viz
- contextual geo info is important: climate/landuse/vegetation/...
- remove unimportant elements, everything displayed should have a purpose (referring to contents in pop up)
- star plots are disliked in general
- a histogram with time information could be placed in a pop up
- news feed instead of table is much welcomed

Interesting: main focus is on base map & contextual geo info, not much discussion on data viz itself

Study Participants

Name	Profession	Date	Signature
Florian Brabant	civil engineer	07.10.14.	[signature]
Sebastian Brabant	civil engineer	07.10.14	[signature]
Luyi Hua	Cartography Student	10.10.14	[signature]
Ieva Dobrāja	Cartography Student	10.10.2014.	[signature]
Mariana Damiani	cartographer MSc.	10.10.2014	[signature]
Lisa Clemens	Cartography	15.10.2014	[signature]
Franz Gerlich	Mountain Guide	15.10.14	[signature]
Malena Vargas Amado	Cartography student	15.10.2014	[signature]
Michael, Jenkner	Economics	17.10.2014	[signature]
Angela Schmidt	civil engineer	26.10.2014	[signature]
Thomas Trämer	civil engineer	26.10.2014	[signature]
Sophie Emmrich	economics engineer	26.10.2014	[signature]