

Multiple Coordinated Views on Massive Geo Data

Master's Thesis in partial fulfillment for the academic degree

“Master of Science” (M.Sc.)

in IT Systems Engineering

Hasso Plattner Institute | Faculty of Digital Engineering
University of Potsdam

submitted by

Robert Schäfer

robert.schaefer@student.hpi.de



Prof.-Dr.-Helmert-Str. 2-3
14482 Potsdam, Germany

Supervisors:

Prof. Dr. Jürgen Döllner
Dr. Benjamin Hagedorn
Jan Klimke

Hasso Plattner Institute
Potsdam, Germany
December 19, 2017

Abstract

Coordinated, multiple information visualizations can offer added value by combining the strengths of individual visualization techniques or data representations and by reflecting user interactions among multiple views simultaneously. Specifically, when visualizing multi-dimensional and hierarchical geodata, the combination of digital maps and treemap-based visualizations can facilitate the comprehensibility and interactivity. E.g. a linked geographic map can establish the geographical context if the used treemap layout algorithm does not take into account the spatial data. In addition, the user can select elements as desired based on the geographic proximity in the map or based on the technical context in the treemap.

This master thesis provides the design and implementation of a coordinated visualization of a treemap and a geographic visualization. The focus is on coordinated interactions, the information which is exchanged between individual views and the common intersection between data visualizations.

Based on existing scientific work and a series of plausible interactions in individual data visualizations, a conceptual framework for coordinated multiple views is specified. Furthermore, a formal concept of an interaction is defined, a common data model between visualizations is derived and a notification procedure is developed. This specification is implemented for the use case of the coordination of a treemap and a geographic visualization. Finally, common application scenarios and software-relevant requirements are evaluated and the system performance is analysed.

Zusammenfassung

Koordinierte, multiple Informationsvisualisierungen können einen Mehrwert bieten, indem sie die Stärken einzelner Visualisierungstechniken oder Datendarstellungen kombinieren und Benutzerinteraktionen gleichzeitig in mehreren Ansichten wiedergeben. Insbesondere bei der Visualisierung mehrdimensionaler und hierarchischer Geodaten kann die Kombination von digitalen Karten und Treemap-basierten Visualisierungen die Verständlichkeit und Interaktivität erleichtern. Z.B. kann eine verknüpfte Kartendarstellung den geografischen Kontext herstellen, wenn der verwendete Layoutalgorithmus der Treemap die räumlichen Daten nicht berücksichtigt. Darüber hinaus kann der Nutzer je nach Bedarf Elemente basierend auf der geographischen Nähe in der Karte oder basierend auf dem technischen Kontext in der Treemap auswählen.

Diese Masterarbeit beinhaltet den Entwurf und zeigt die Implementierung einer koordinierten Visualisierung einer Treemap und einer geographischen Visualisierung. Der Fokus liegt auf koordinierten Interaktionen, den Informationen, die dabei zwischen einzelnen Ansichten ausgetauscht werden, und den gemeinsamen Schnittmengen zwischen einzelnen Datenvisualisierungen.

Aufbauend auf bestehenden wissenschaftlichen Arbeiten und einer Reihe von plausiblen Interaktionen in eigenständigen Datenvisualisierungen wird ein konzeptueller Rahmen spezifiziert. Dabei wird ein formales Konzept einer Interaktion definiert, ein gemeinsames Datenmodell zwischen Visualisierungen abgeleitet und ein Benachrichtigungsverfahren entwickelt. Diese Spezifikation wird für den Anwendungsfall der Koordination einer Treemap und einer geographischen Visualisierung implementiert. Anschließend werden typische Anwendungsszenarien und software-relevante Anforderungen evaluiert und eine Performance-Analyse durchgeführt.

Glossary

AJAX Asynchronous JavaScript. 45

API Application programming interface. 3, 35, 45

BMI Body mass index. 10

CMV Coordinated Multiple View. 23

DOM Document Object Model. viii, 48, 50, 51

UI User Interface. 4

W3C World Wide Web Consortium. 4

WebGL Web Graphics Library. 47

XML Extensible Markup Language. 19

Contents

| | |
|--|----|
| Glossary | v |
| 1 Introduction | 1 |
| 1.1 Motivation | 2 |
| 1.2 Problem Statement | 3 |
| 1.3 Hypothesis | 3 |
| 1.4 Contributions | 4 |
| 1.5 Structure of the Work | 4 |
| 2 Related Work and Foundations | 7 |
| 2.1 Information visualization | 7 |
| 2.2 Data-driven Decision Support Systems | 8 |
| 2.3 Geovisualization | 8 |
| 2.3.1 Visual Variables | 8 |
| 2.3.2 Choropleth maps | 9 |
| 2.4 Information Visualization of Hierarchical Data | 10 |
| 2.4.1 Treemaps | 11 |
| 2.4.2 Spatially Consistent Treemaps | 12 |
| 2.4.3 3D treemaps and 2.5D treemaps | 13 |
| 2.5 Coordinated Multiple Views | 14 |
| 2.5.1 Brushing and Linking | 14 |
| 2.6 Modern Web Technologies | 15 |
| 2.7 Related Work on Coordinated Multiple Views | 17 |
| 2.7.1 Guidelines for Multiple Views | 17 |
| 2.7.2 Interaction Aspects | 18 |
| Interaction Categories | 18 |
| Formalization of Interactions | 19 |
| 3 Analysis | 23 |

| | | |
|----------|---|-----------|
| 3.1 | Existing Interactions in the Current Implementation | 23 |
| 3.2 | Software Requirements for Additional Views | 24 |
| 3.3 | Component Pattern and Web Frameworks | 25 |
| 3.4 | Arbitrary Data Visualization Techniques | 27 |
| 3.4.1 | Line Charts..... | 27 |
| 3.4.2 | Bar Charts and Multi-Set Bar Charts | 28 |
| 3.4.3 | Histograms | 29 |
| 3.4.4 | Bubble Charts and Scatter Plots | 30 |
| 3.4.5 | Stacked Bar Charts | 31 |
| 3.4.6 | Hierarchical visualizations | 32 |
| 3.4.7 | Geographic Data Visualizations | 33 |
| 3.4.8 | Temporal Data Visualizations | 34 |
| 4 | Conceptual Framework | 37 |
| 4.1 | The Conceptual Framework from the User Perspective | 37 |
| 4.1.1 | Interaction | 39 |
| 4.1.2 | Interaction in Coordinated Multiple Views..... | 39 |
| 4.1.3 | View..... | 39 |
| 4.1.4 | Trigger..... | 39 |
| 4.1.5 | Effect | 40 |
| 4.1.6 | Interaction Category | 40 |
| 4.1.7 | Interaction Purpose..... | 40 |
| 4.1.8 | Interaction Subject | 40 |
| 4.2 | Shared data model..... | 41 |
| 4.3 | Encoding of Interaction Subjects | 41 |
| 4.3.1 | Set definitions | 42 |
| 4.3.2 | Function declarations | 43 |
| 5 | Implementation | 45 |
| 5.1 | Implemented interactions | 45 |
| 5.2 | Implementation of the Conceptual Framework..... | 47 |
| 5.2.1 | PubSubJS as implementation for publish-subscribe | 47 |
| 5.2.2 | Encoding of category and purpose as topic and sub-topic | 47 |
| 5.2.3 | Encoding of the subject as a JavaScript callback..... | 48 |
| 5.2.4 | Data Model implemented as GeoJSON | 48 |
| 5.3 | Architecture | 49 |
| 5.3.1 | UAController | 49 |
| 5.3.2 | Geographic Visualization with React Components | 49 |
| 5.4 | Notifications | 50 |
| 5.5 | Querying the GeoJSON endpoint | 51 |

| | | |
|-------|--|----|
| 5.6 | Render views in the Document Object Model (DOM) | 52 |
| 5.7 | MapComponent | 52 |
| 5.7.1 | GeoJSON Component | 53 |
| 6 | Evaluation and Discussion | 57 |
| 6.1 | Use Case Scenarios | 57 |
| 6.1.1 | Explain outliers | 57 |
| 6.1.2 | Multi-Select in 2.5D treemap | 58 |
| 6.1.3 | Bounding Box Selection in Geographic visualization | 59 |
| 6.1.4 | Use case summary | 61 |
| 6.2 | Evaluation According to Guidelines | 61 |
| 6.3 | Performance Evaluation and Limitations | 62 |
| 6.3.1 | Immoscout | 63 |
| 6.3.2 | Landkreise | 65 |
| 6.4 | Discussion of Requirements | 65 |
| 7 | Summary and Conclusion | 69 |
| 7.1 | Future Work | 70 |

1

Introduction

The human brain processes visual information better than it processes text. As a result, data scientists, journalists and designers often communicate their results with data visualizations. These visualizations make the data easier accessible to readers and help them e.g. to spot correlations, disparities or anomalies.

Data visualizations on a computer allow the user to interact with the data and explore different levels of granularity. Users can use an input device like a mouse or a keyboard to trigger an interaction and the visual representation of the data will change accordingly. In many cases a great interactivity results in a great user experience, as the user can change the visual representation as desired.

There is a wide range of existing frameworks and implementations for data visualizations. These frameworks often support interactions with additional controls and provide an Application programming interface (API) and event handlers for input devices.

In some examples, these frameworks are extended to show multiple views next to each other, with controls and event handlers to manipulate multiple views. When visualizing the same data set, these visualizations are called coordinated multiple views. The user experience is often better compared to single data visualizations but the computational overhead for multiple views is high.

Research in the area of coordinated multiple views often presents a successful application of a coordinated multiple view for a specific use case. In most cases, the presented application is intended for the use case only and therefore these applications are not reused in any other environment. Currently, there is no dedicated or prevalent framework that helps to implement interactions in multiple visualizations of the same data.

1.1 Motivation

An existing VISUAL ANALYTICS PLATFORM visualizes multi-dimensional, hierarchical and spatial data in form of 3D treemaps. The input for this application is e.g. census data, real estate data or infrastructure data. Because this data has a geographic context in many cases, the treemap visualization should be supported by a second geographic visualization. Both visualizations should be incorporated in a coordinated multiple view layout and interactions in the treemap should be reflected in the geographic visualization and vice versa.

Besides the goal of implementing an application of a coordinated multiple view with a treemap and a geographic visualization, the resulting software should be easily extensible. Adding more interactions and more data visualizations should be possible with little effort. Until now, the type of additional data visualizations and interactions is unknown. Therefore the approach is to anticipate future data visualizations and interactions by analysing their characteristics.

Many research papers in the field of coordinated multiple views are more focused on visual representations than interaction aspects. Even though interaction aspects give a great user experience and enable the user to find complex connections involving multiple views. Ho (2013) assumes this may “originate from the fact that the implementation of interaction techniques and interactive features normally takes much more time than the implementation of visual representations”. There is a lot of existing research in the area but a lack of research regarding interactions in particular. This could also explain why hardly any general-purpose implementation exists to coordinate multiple views, that focuses on interactions. It is for this reason that this thesis develops an interaction model for coordinated multiple views, so future implementations can use this model as a specification.

Another motivation is a recent development in the industry: Many popular web frameworks have developed mechanisms to update User Interface (UI) elements during user interactions. These patterns and mechanisms have become so widespread and prevalent that they triggered even a web specification of the World Wide Web Consortium (W3C) called “web components”. These update mechanisms are a promising choice for coordinated multiple views and it seems, they have not been applied for this purpose.

1.2 Problem Statement

When visualizing e.g. administrative districts, an ordinary treemap adequately visualizes membership relationships. But the layout of nodes may be geographically inconsistent, especially if the treemap layout algorithm does not take into account the spatial data.

Problem Statement

A treemap of geographic data may lose the geographic context if the treemap layout algorithm is based on non-geographic attributes.

A node may be placed in the top left corner of the treemap although a node is geographically located in the south-east. Items that should belong together according to their geographic circumstances may be scattered across the treemap. The comprehensibility and interactivity can suffer in such a case. Users have a hard time to recognize geographic areas and locations. Selecting and grouping items based on their geographic proximity becomes increasingly difficult if the items are scattered across the treemap.

Some treemap layout algorithms try to layout nodes more geographically consistent as described in Section 2.4.2. But this consistency comes with a price. In some cases, these algorithms render severely fragmented treemaps that are even harder to understand.

1.3 Hypothesis

The hypothesis of this work is:

Hypothesis

A second, geographic visualization next to the treemap can preserve the geographic context if these two views are combined in a coordinated multiple view.

The user can relate an item in the treemap if the item is linked with the corresponding item in the geographic visualization and vice versa. Many items can be selected in the geographic map based on their proximity by dragging a bounding box around them. In the treemap the user can select many items based on their proximity in a non-geographic dimension and see the selection in the geographic visualization. Essentially, the limitations of a single treemap or a single

geographic visualization can be overcome by splitting up the interaction: The user can trigger the interaction in one view and see the effect, i.e. the change of visual representation, in another view.

1.4 Contributions

The contributions of this work are the following:

Contributions:

1. A formalization of interaction aspects in the field of data visualizations
2. A conceptual framework for coordinated multiple views of arbitrary data visualizations
3. An implementation of the conceptual framework for the present use case of geographic and hierarchical data
4. A proof of concept how treemap and geographic visualization can be combined to overcome limitations of each visualization respectively

1.5 Structure of the Work

In Section 2 the basic terminology of coordinated multiple views is introduced and the theoretical background in this area of research. This Section also covers the state of the art and research on multiple views and coordinated interactions. In Section 3 a set of data visualizations is analyzed and their interactions by example. Each interaction is further examined to identify the relevant information that need to be communicated in a coordinated multiple view. The gained knowledge from that section is used in the following Section 4 to develop a conceptual framework that can be used for future implementations of coordinated multiple views. This conceptual framework includes a data model shared between all views, a suggested communication protocol as well as a formalization of an interaction. In Section 5 the implementation of the conceptual framework is described for the present use case. This implementation serves as reference implementation, to prove feasibility of the conceptual framework. The implementation includes the necessary interactions to demonstrate the advantage of coordinated multiple views for hierarchical and geographic data. It also serves to validate or invalidate the hypothesis in Section 1.3. In Section 6 the implementation is tested for typical visual analytics tasks and it is demonstrated that addi-

tional value can be generated from the combination of treemap and geographic visualization. Based on these use case scenarios, the coordinated multiple view implementation is examined and evaluated for design criteria (Wang Baldonado, Woodruff, and Kuchinsky 2000). A performance analysis of the implementation is carried out to demonstrate the feasibility of the conceptual framework. The last part of the evaluation is a manual check of software requirements defined in Section 3.2. The types of evaluation are therefore:

- (1) Use case, (2) evaluation according to design criteria (3) performance profiling and (4) manual check of software requirements.

Finally, the main contributions in Section 7 are summarized and the future work is outlined.

Related Work and Foundations

This Chapter covers terminology, relevant techniques for data visualization and the related work in this area of research. Advantages and disadvantages of certain data visualization are emphasized. After introducing visualization techniques, related work on interaction aspects in coordinated multiple views is outlined.

2.1 Information visualization

Information visualization is a means of visual communication and has steadily developed since the 16th century (Friendly and Denis 2001). It is a generic term, expressing all kinds of effort to put data into visual context to help people understand the significance of data. Information visualization today goes beyond standard charts and graphs used in spreadsheet applications and covers also infographics, heat maps, geographic visualizations and treemaps (Rouse 2017). Otherwise abstract information is visually represented, making complex data more accessible, understandable and usable.

Kusinitz (2014) mentions that the human brain processes visual information 60,000 times faster than text and visual content makes up even 93% of all human communication. According to the Interaction Design Foundation the purpose of data visualizations is twofold: Sense-making and communication (Few 2013).

Statistical information is abstract and in data visualization “we must find a way to give form to that which has none.” (ibid.) Successful data visualizations helps the human user to derive knowledge and meta data from the visualization itself. Nocke and Schumann (2002) call this “visual data mining”.

2.2 Data-driven Decision Support Systems

Data-driven decision support systems are applications to support businesses and organizational decision-making activities in which data visualizations play a key part (Lavrač et al. 2007) (Poleto, Carvalho, and Costa 2015). A common expression by impatient managers who can not afford to wade through lengthy reports has even become the title of a book about decision support systems: Stephen Few's "Show me the numbers".

In the business context, sales managers demand a quick access on the latest data with all relevant visualizations at once. Examples of this kind of decision support systems are called e.g. "decision cockpit" or "business sphere" (Davenport 2013).

We can expect to see these technologies more and more in business applications. McAfee and Brynjolfsson (2012) from the MIT Center of Digital Business showed that organizations driven most by data-based decision making had 4% higher productivity rates and 6% higher profits.

However, little research has been done regarding the performance of coordinated multiple views in the field of decision making. There might be a great potential. In 1997 Mayer (1997) conducted eight studies to compare the effect of using multimedia on university students. The studies showed that when using combined visual and verbal explanations the generation of creative problem solutions increased by an average of more than 50%.

Apparently, the application of combined data visualization techniques in decision making is a promising strategy.

2.3 Geovisualization

The umbrella term "Geovisualization" covers visualization techniques for the "visual exploration, analysis, synthesis and presentation of geospatial data (any data having geospatial referencing)" (MacEachren and Kraak 2001). The entities of geospatial data may include buildings, streets, landmasses, terrain, administrative districts and also moving entities like cars.

2.3.1 Visual Variables

French cartographer Jaques Bertin introduced seven visual variables in 1967 (Bertin 2010). Figure 2.1 shows these visual variables. These visual variables are used

in cartography but can also be applied to data visualization in general. Carpendale (2003) explains in detail their use in computational information instead of printed cartography. Garlandini and Fabrikant (2009) put these visual variables under systematic validation procedures. The authors conclude that the variable **size** provides the most accurate and efficient performance while the variable **orientation** provides the least performance. Bertin's visual variable play a role in interaction aspects of coordinated multiple views as they are used to communicate the effect of an interaction. A highlighted data point can be highlighted by changing the colour or increasing the size of a point.

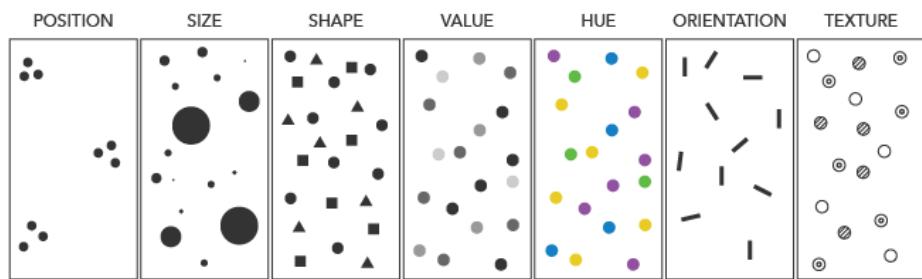


Fig. 2.1: Bertin's original visual variables (Foster 2017).

2.3.2 Choropleth maps

Choropleth maps are thematic maps in which areas are shaded or patterned in proportion to the statistical variables being displayed on the map. A popular use case is the display of population density or per-capita income. An example of a choropleth map is shown in Figure 2.2, visualizing the percentage of obese population in the US. Choropleth maps are very popular and therefore many people are familiar with them already. A downside of choropleth maps is that larger regions may appear more emphasized than smaller ones, since the entire area of regions is coloured. Another disadvantage of choropleth maps is the common error of incorrect encoding: Such an incorrect encoding would be the display of absolute numbers, e.g. total population or proceeds of crime, rather than relative numbers, e.g. population density or unemployment rate.

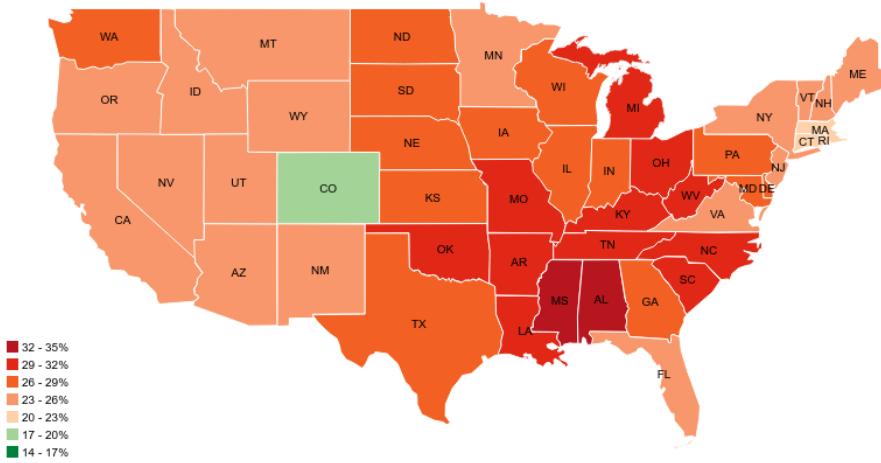


Fig. 2.2: Choropleth map of obese population, Body mass index (BMI) > 30, in the United States in 2008 (National Center for Chronic Disease Prevention and Health Promotion 2010).

2.4 Information Visualization of Hierarchical Data

The visualization of hierarchical data has a long tradition. The traditional visual representation of a tree is a directed graph with the root node at the top, as seen in Figure 2.3.

An common use case is a directory tree of a file system, e.g. a file browser or the command line utility `tree` on UNIX based operating systems. As Shneiderman (1992) mentions, this visualization becomes increasingly large when displaying more than one level and soon exceeds the entire screen size.

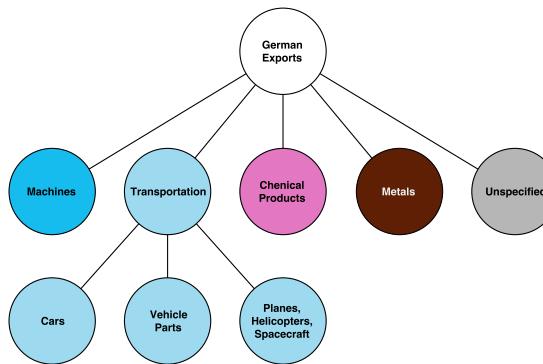


Fig. 2.3: Traditional visualization of a tree in form of a directed graph with edges and nodes and the root node at the top. This example visualizes equivalently a part of the treemap visualization of German exports in Figure 2.4.

2.4.1 Treemaps

Johnson and Shneiderman (1991) propose the treemap visualization technique, in which each node is a rectangle whose area is proportional to a specified dimension. In treemaps every node is visualized as a tile. The membership relationship is expressed with tiles containing other tiles, thus representing the hierarchy.

Figures 2.4 shows an example of a treemap and Figure 2.5 shows the same treemap in another level of detail. German exports are divided in generic groups like “Machines” and “Chemical Products” and include more specific groups like “Cars” and “Packaged Medicaments”. The user can click on a drop-down menu to change the current level of hierarchy, only leaf nodes are displayed at a time.

The advantage of treemaps is that they are space-filling visualizations, i.e. they make 100% use of the available screen size. A treemap will, unlike a graph representation of a tree, never exceed the size of the screen.

The area of the tiles can be mapped to a data attribute, e.g. the file size on disk or, in cases of Figures 2.4 and 2.5, the percentage of export quota. Thus, a treemap can display even more information than a traditional graph representation. A disadvantage of treemaps is the variable size of each node. If more and more nodes are displayed, the size of each tile will get smaller and smaller and e.g. there might not be enough space to display a label. You can see this problem occur in Figure 2.5.

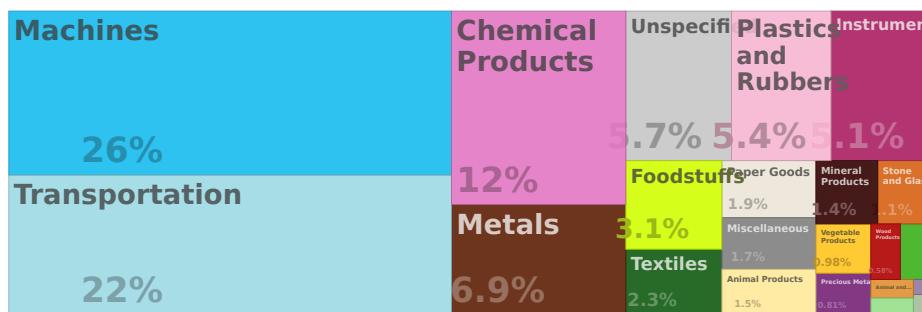


Fig. 2.4: A two dimensional treemap of Germany’s foreign trade quota of exports, showing only the first hierarchy level (Observatory of Economic Complexity n.d.).

If treemaps are used to visualize geographic data like municipalities, real estates or streets, the placement of nodes depends on the tiling algorithm and not the geographic location on a map. Let’s say a treemap visualizes a hierarchy of federal states and municipalities with the area of each tile mapped to the total population of administrative district. Even if the membership hierarchy is

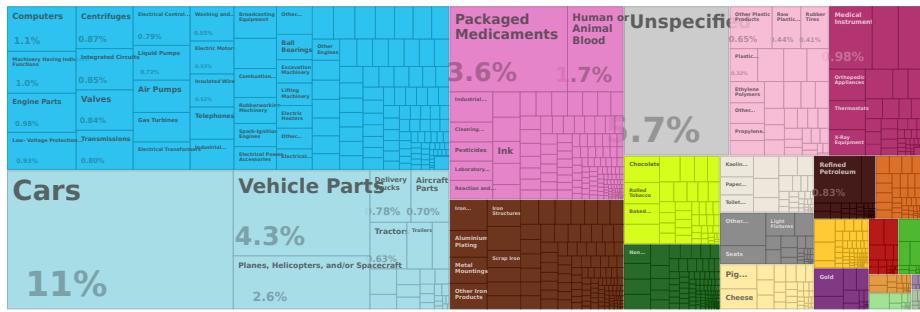


Fig. 2.5: Second level of detail of the treemap in Figure 2.4 (Observatory of Economic Complexity n.d.).

matched by the treemap, the placement of districts in the same hierarchy level is based on their total population and not their geographic location.

2.4.2 Spatially Consistent Treemaps

When visualizing geographical data with a treemap, e.g. census data, it is desired to have a spatially consistent visualizations. The layout of hierarchical administrative subdivisions should approximately match a geographic visualization of these subdivisions. An early algorithm for geographically consistent treemaps is called “Spatially Ordered Treemaps” by Wood and Dykes (2008). It is a modification of the “Squarified Treemap Algorithm” (Bruls, Huizing, and Van Wijk 2000). This modification places nodes based on their distance to the enclosing rectangle to be filled, rather than in the weight sequence order. As Ghoniem et al. (2015) demonstrate, this algorithm shows undesirable fragmentation for large, flat hierarchies. One especially serious manifestation of this problem is visible in Figure 2.6.

Less fragmented but also less consistent regarding the geographic correctness are “Histomaps” by Keim (2002). This treemap algorithm sorts nodes such that their direction in the layout approximates their direction with regard to latitude and longitude. This does not lead to fragmentation, but also does not guarantee geographic correctness.

Ghoniem et al. (2015) developed “Weighted Maps” which are considered to be a trade-off algorithm for “HistoMaps”. This trade-off is between aspect ratio and geographic correctness. Generally speaking, these treemap algorithms suffer because two unrelated circumstances are visualized in the same view. For that reason, the approach in this paper is to establish the geographic context in a separate view and let treemaps excel in what they are best: Visualizing hierarchical data.

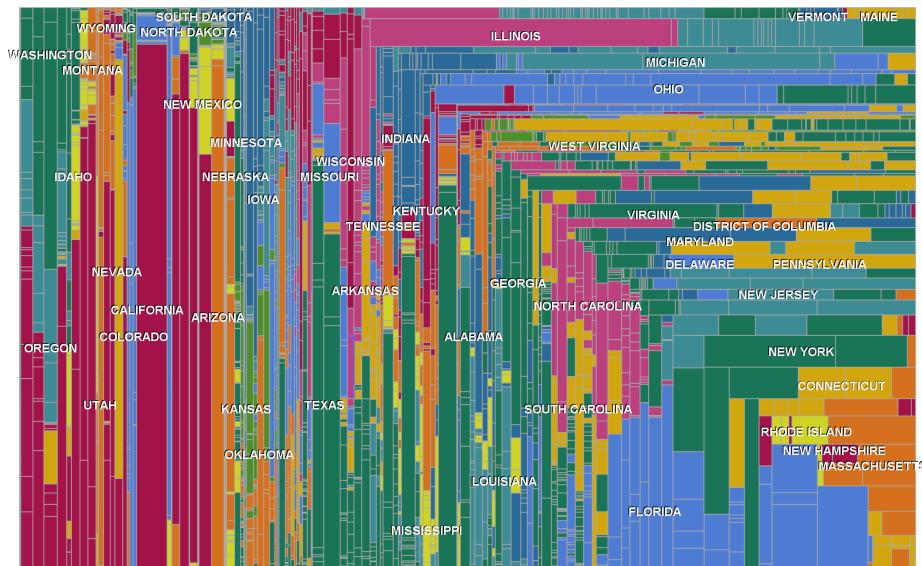


Fig. 2.6: A “Spatially Ordered Treemap” by Wood and Dykes (2008) visualizing the population in 3,109 counties in the USA (Ghoniem et al. 2015). This algorithm shows severe fragmentation for large, flat hierarchies.

2.4.3 3D treemaps and 2.5D treemaps

3D treemaps are a concept introduced by Bladh, Carr, and Scholl (2004) in 2004. The authors transfer the concept of treemaps from two dimensional into three-dimensional space, transforming tiles to blocks. They introduce “Step-Tree” (*ibid.*), which is a three-dimensional treemap to display a directory layout of a file system. It “differs from treemaps in that it employs three dimensions by stacking each subdirectory on top of its parent directory.”

3D treemaps are superior to 2D treemaps for tasks with a pronounced topological challenge. Users perform significantly better in interpreting the hierarchical structure. However, 3D visualizations also introduce some disadvantages. Blocks can superimpose each other, forcing the user to navigate the view point. The navigation of the view point itself is an increase of complexity not present in two dimensional treemaps.

The term 2.5D treemap was coined by Limberger et al. (2016) in 2016. A 2.5D treemap is just an ordinary 3D treemap, but it has all blocks attached to the ground, or more specifically, attached to the parent block. An example of a 2.5D treemap is shown in Figure 2.7.

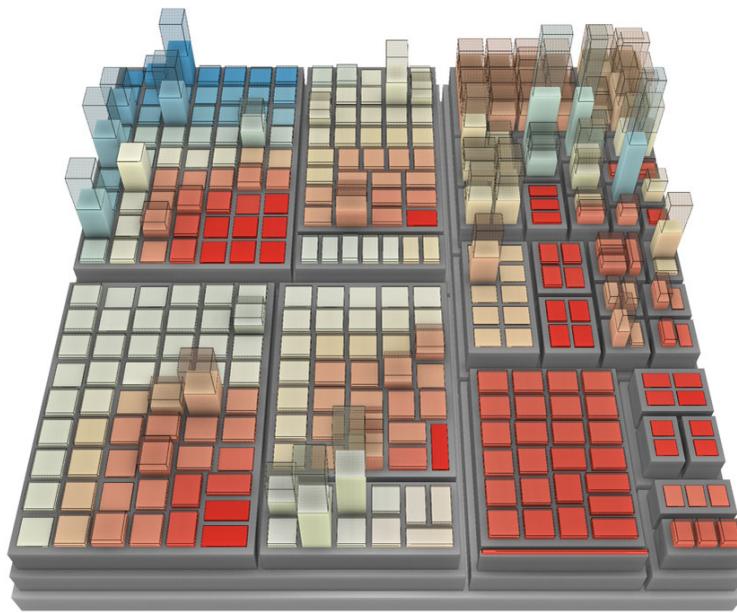


Fig. 2.7: Example of a 2.5D treemap (Döllner 2017).

2.5 Coordinated Multiple Views

Coordinated multiple views are a combination of data visualizations of the same data set in multiple views, often side-by-side. According to Roberts (2007) coordinated multiple views are just “a specific exploratory visualization technique that enables users to explore their data. The overall premise for the technique is that users understand their data better if they interact with the presented information and view it through different representations.” (ibid.)

Some coordinated multiple views are shown in Figure 2.8. They display spatial and temporal attributes of pictures from a picture database as well as continuous attributes like popularity and number of comments. The user can move the mouse cursor over each item in the scatter plot and the graduated symbol map and the corresponding item is highlighted with a larger stroke in all other views. On the time line below, the user can also filter for pictures in a certain time frame by dragging the mouse from lower to upper limit.

2.5.1 Brushing and Linking

Brushing and linking is a common interaction pattern found in coordinated multiple views and often a crucial part of these visualizations. “The technique of brushing is the principle approach, where elements are selected (and highlighted)

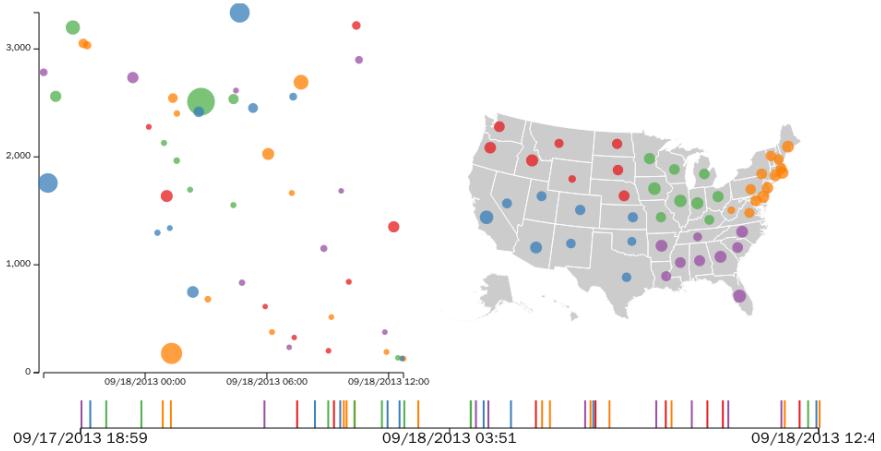


Fig. 2.8: coordinated multiple views displaying the popularity, number of comments, location and time of pictures in a picture data base (Bostock 2017b).

in one display, concurrently the same information in any other linked display is also highlighted.” (ibid.)

An example is given in Figure 2.9. It displays an on-time performance of airlines, visualized with the “Crossfilter” JavaScript library.

Each of the flight in the data set has an hour and a date for departure, an arrival delay, which can also be negative, and a traveled distance in miles. The user can “brush” the data by selecting an interval by dragging the mouse. The respective view will become a primary view and display the deselected items with a grey colour.

All other views become secondary views and display only selected items. The visualization takes the most recent 80 flights from the database that match all given filters. The user can further filter for items by dragging another interval in one of the secondary views.

This technique of propagating interactions to other views is called “linking”.

Figure 2.9 shows a filter for travels with a long delay, i.e. from 120 minutes to the maximum value, see the selection in the upper center. In the view in the upper left corner in Figure 2.9 long delays correlate with the time of the day.

2.6 Modern Web Technologies

This section introduces some state-of-the-art libraries, frameworks and standards for web development, as the existing VISUAL ANALYTICS PLATFORM is

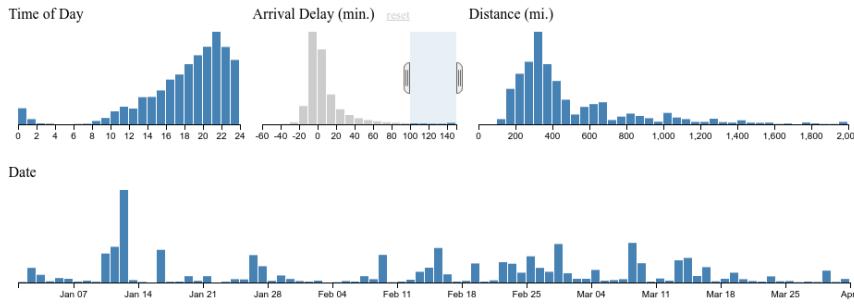


Fig. 2.9: Airline on-time performance: Correlation of time of day with arrival delay. Most recent flight with a delay of more than 100 minutes selected (Bostock 2017a).

a web application. While not dedicated for coordinated multiple views only, these technologies are either considered or used for the implementation of the additional geographic visualization, as described in Chapter 5.

CrossfilterJS is one of the very few libraries dedicated for coordinated multiple views. It simplifies the implementation of brushing and linking interactions among several timeline visualizations of tabular data. However, this library is unmaintained as of December 2017, the most recent commit dating back to March 2016.

Leaflet is the leading open-source JavaScript library for mobile-friendly interactive maps (Agafonkin and contributors 2017). It has support for GeoJSON which makes it very easy to display tiled web maps with interactive overlays.

Web components is a recent standard of the World Wide Web Consortium (W3C) (2017) to bring component-based software engineering to the world wide web. Web components are a set of web platform APIs to create new custom, reusable, encapsulated HTML tags that can be used in web pages and web applications (Li, Stockwell, and Mutton 2017). Because of their encapsulation and reusability, web components are a promising choice for coordinated multiple views that are implemented in JavaScript based web applications.

ReactJS is an open-source JavaScript library to build user interfaces and allows to create reusable UI components. React renders HTML on the client, it changes the page without reloading the page. The framework corresponds with the View in the Model-View-Controller pattern. React components are structured hierarchically, with each component having dedicated responsibility. React is explicitly not implementing web components and is not going to implement web components in the future. It has, in return, a well-known way of

integrating the component framework into a legacy application built with e.g. jQuery.

GlimmerJS is the rendering engine of EmberJS (Dale 2017). In 2017 it was released as a standalone component framework. Applications written in GlimmerJS can be exported as web components. These web components can be included in any website, which makes GlimmerJS a reasonable choice to build high-quality widgets for user interfaces.

Google Polymer is another popular library to build web components (Google Inc. 2017). With 18,469 stars on Github it is the most popular framework for web components at the time of writing. Polymer has a large community and comprehensive documentation and therefore more suitable than GlimmerJS to build coordinated multiple views.

PubSubJS is a topic-based publish/subscribe library written in JavaScript (Roderrick 2017). Topics can be registered hierarchically, with subtopics delimited by dots. A subscription to topic `mcv.select.focus` will be notified only for `focus` interactions whereas a subscription of `mcv.select` will be notified for both `focus` and `highlight` interactions. Furthermore, topics are published asynchronously, so if the user interacts with a visualization, that does not block code execution.

2.7 Related Work on Coordinated Multiple Views

The annual conference “International Conference on Coordinated and Multiple Views in Exploratory Visualization” provides a good selection of scientific papers in the area of coordinated multiple views. E.g. four papers of the conference of 2007 provide an introduction into the area of research: “Coordinated Multiple Views: a Critical View” by G. Andrienko and N. Andrienko (2007), “State of the Art: Coordinated & Multiple Views in Exploratory Visualization” by Roberts (2007), “The Future of CMV” by Erbacher (2007) and “Is coordination a means to collaboration?” by Weaver (2007). Other papers present specific applications of coordinated multiple views or user studies. Relevant for this thesis are interaction aspects, in order to develop a conceptual framework of coordinated multiple views, and guidelines for multiple views, to evaluate the system.

2.7.1 Guidelines for Multiple Views

Wang Baldonado, Woodruff, and Kuchinsky (2000) established 8 rules as guidelines for multiple views. These guidelines can be used to either choose a coor-

dinated multiple view for an application or to improve the user experience for existing coordinated multiple view applications. Each rule comes with a set of defined positive impacts on utility as well as negative impacts on utility. Following these rules, designers and developers can tell whether the benefits of multiple views outweigh their drawbacks.

The eight rules are as follows:

1. The rule of diversity: Use multiple views when there is a diversity of attributes.
2. Complementarity: Use multiple views when different views bring out correlations or disparities.
3. Decomposition: Partition complex data into multiple views to create manageable chunks.
4. Parsimony: Use multiple views minimally.
5. Space/Time Resource Optimization: Balance the costs of presenting multiple views with the benefits of using them.
6. Self-Evidence: Use perceptual cues to make relationships more apparent.
7. Consistency: Make the interfaces and states for multiple views consistent.
8. Attention Management: Focus the attention of the user on the right view at the right time.

2.7.2 Interaction Aspects

According to Ho (2013) interactions are a crucial part of data visualizations, yet most research in the area still focuses on visual representations. Roughly speaking, research on interaction falls into these groups: How to categorize interaction techniques? How to find new interaction techniques and apply those to visualizations?

The following sections will give an overview of relevant research in interaction aspects for each group.

Interaction Categories

This section covers high-level classification and categorization of interactions in coordinated multiple views.

In 1996 Shneiderman (1996) classified interactions into six groups: (1) Gain an *overview* of the entire collection, (2) *zoom* in on items of interest, (3) select an item or group and get *details* when needed, (4) view *relationships* among items, (5) keep a *history* of actions to support undo, (6) allow *extraction* of sub-collections and of the query parameters.

Two years later, Dix and Ellis (1998) identified these categories: (1) *Highlight and focus* particular subsets of the data, (2) instead of displaying everything simultaneously *access extra information* by drilling down the data, (3) zoom in and out to give an *overview and context*, (4) *change parameters* of the *same representation*, e.g. another baseline of a stacked bar chart, (5) *change representation* of the *same data* by switching the chart type, (6) *link representations* to determine the relationship between items.

In 2002, Keim (2002) defined the following classification: (1) *Dynamic projection* to show all combination of data attributes mapped to the axis of a diagram, (2) focus on a smaller subsets by *filtering* out parts of the data, (3) *zoom* into a subset of the data and get a higher level of detail, (4) drill-down operations to preserve an overview of the data are called *distortion* (5) and finally *link and brush* visualizations, to highlight the same data points in multiple visualizations.

The most recent classification was done in 2007 by Yi, Kang, and Stasko (2007) listing seven categories: (1) *Select* to mark something as interesting, (2) *explore* to show something else, (3) *reconfigure* to show a different arrangement, (4) *encode* to show a different representation, (5) *abstract/elaborate* to show more or less detail, (6) *filter* to show something conditionally, (7) *connect* to show related items.

It is noticeable that all of these classifications of interactions are redundant. E.g. there is a *zoom* interaction for the first three classifications and all of them have a category similar to *select*. In this work the classification of Yi, Kang, and Stasko (ibid.) is used for the remaining parts because it is the most recent classification and it is based on the precursors.

Formalization of Interactions

This section covers the smaller part of research in coordinated multiple views, research that is *not* related to a high-level classification of interactions. Not only interactions in coordinated multiple views are considered but any kind of a formalization of interaction that may be used as the starting point for a framework for coordinated multiple views.

ITlib (Figueroa, Green, and Watson 2001) is an architecture and a framework of interaction techniques for virtual reality applications, designed to be extensible and flexible. New interaction techniques can easily be added and application specific code is seamlessly integrated.

On a low level an interaction technique “is modeled as a set of filters connected in a small data flow” (ibid., p. 2). These filters are the smallest process unit in the data flow. Composed of input and output ports, they communicate with other filters, to receive data input from predecessors and send data output to successors.

The framework specifies and stores the interaction techniques along with its filters, the execution model and the scene in Extensible Markup Language (XML) documents.

Even though the system describes interactions in an abstract way, the domain of the framework is clearly the interaction of a human body within a 3D virtual reality. Certain assumptions are made, including the data model, which is the 3D scene, and human computer interaction devices, like the user’s hand or the user’s head.

The goal is not to better understand the data, which is the common goal in data visualizations. In this case, the data model is the 3D scene and the goal is to manipulate the 3D scene.

Most importantly, the framework describes interaction techniques for a single viewpoint but not for coordinated multiple views.

A framework for Focus+Context Visualization by Bjork, Holmquist, and Redstrom (1999) is one of the few formalizations of interactions in data visualizations.

The idea behind Focus+Context visualizations is to present the object of primary interest in full detail while at the same time giving a overview of the surroundings.

The authors of the paper distinguish visualizations and second-level visualizations: Visualizations, referred to as IV , are triples of a set $[D]$ of underlying data, a visual representation V and I which is the possible interaction or manipulation.

$$IV([D], V, I) \tag{2.1}$$

If I affects $[D]$ the underlying data set can be manipulated. Examples would be changes in a spreadsheet editor, or a change of the start and end date of an appointment in a calendar.

If I affects V the user can manipulate IV in order to change the way $[D]$ is represented. This statement holds e.g. for an interaction in which the user increases the visible level of hierarchy in a treemap. According to Yi, Kang, and Stasko (2007) this would be an *abstract/elaborate* interaction. The effect of such an interaction is depicted by the change from Figure 2.4 to Figure 2.5 in Section 2.5.

Second-level Visualizations are information visualizations consecutively applied. The underlying data set $[D]$ of the formula in Equation 2.1 is replaced with some information visualization IV' , which is compatible with IV .

$$IV'(IV, V', I') \quad (2.2)$$

Focus+context visualizations are second-level visualizations. An example given by the authors is the “rubbersheet” visualization, that visually distorts a first-level visualization similar to a magnifier. Regions of primary interest are distorted to appear magnified, while the remaining regions are minified.

The formalization of Bjork, Holmquist, and Redstrom (1999) is suitable to describe multiple information visualizations applied one after another. Yet, it is specialized on Focus+Context interactions. It does not describe any kind of interaction in arbitrary information visualizations like line or bar charts. This is also the reason why the formalization does not describe the exchanged data between views like identifiers of data points.

3

Analysis

The current treemap implementation and its set of possible interactions is described in Section 3.1. For the implementation of the geographic visualization several web frameworks are compared and advantages and disadvantages are contrasted in Section 3.3.

Section 3.4 describes essentials of interactions in information visualizations. To accomplish this, the approach is to deduce the essentials by a list of examples. What are the expected data structures for each visualization? Which visual variables can be used to show the effect of an interaction? A list of possible interactions is given for each visualization technique. Those interactions will be classified according to Yi, Kang, and Stasko (2007) and the relevant subject of the interaction is specified.

3.1 Existing Interactions in the Current Implementation

The current implementation comes with a visualization of a 2.5D treemap and should be complemented with a geographic visualization. In this section, the set of already possible interactions are classified according to Yi, Kang, and Stasko (*ibid.*) as in Section ???. They fall into these categories: *select*, *explore*, *reconfigure*, *encode* and *filter*.

The user can *select* one item in the view by clicking on it. The user can reveal a tooltip showing the item properties by moving the mouse cursor on the item, which is another *select* interaction.

The user can *explore* the map in the usual manner: If the user drags with the mouse on the map, a panning operation is performed with the viewpoint focused on the 2.5D treemap, like a turntable. The zoom factor can be changed by scrolling with the mouse on the canvas of the map.

Encode and *reconfigure* techniques are performed through a menu: The user can *reconfigure* different data sets and the displayed diagram, e.g. a treemap visualization based on the geometry shape, cubes or voronoi regions. In a submenu the user can *encode* properties of a data set in predefined visual variables, e.g. the height, color and texture of an item. A slider can be used to *filter* the data set on a data attribute. When the user drags the slider and changes an upper or a lower limit, items with an attribute beyond this interval are filtered out.

3.2 Software Requirements for Additional Views

The existing implementation should be complemented with a coordinated multiple view layout. Next to the 2.5D treemap a geographic visualization should appear, so that the user can understand the geographic context of items in the treemap.

To facilitate a future use of the software, this section provides a list of software requirements. In Chapter 6 these requirements are reused to evaluate the system.

1. Serialization

Serialization is the process of translating objects that can be stored or transmitted and reconstructed later. In order to coordinate interactions among views, information needs to be passed from one view to another. A framework for coordinated multiple views should therefore find a serialization format for interactions which has (1) small payloads and (2) fast serialization and deserialization.

2. Reversibility

Reversibility means, in the context of coordinated multiple views, the ability to undo the effect of an interaction. Ideally, every interaction should be undoable. If not every interaction is undoable, the cost to replay the interactions from the original state up to the point of the interaction should be minimized.

3. Software Extensibility

Software extensibility means, in this thesis, the costs of changing behaviour. How many views need to be touched, if a new interaction option is added to the system? How much time and effort is necessary to implement the new feature?

4. Maintainability

Maintainability means, in this thesis, the costs of changing the source code. How error-prone is the system, will other views be affected if a new interaction is added to the system?

3.3 Component Pattern and Web Frameworks

The current VISUAL ANALYTICS PLATFORM is a web application and the complementing Coordinated Multiple View (CMV) extension should be based on web technologies, too. Many popular web frameworks like Angular, Ember, React and Vue have developed mechanisms to update UI elements during user interactions. Because a view in a coordinated multiple view is a UI element as well, it is worth to consider one of these web frameworks for the implementation of the coordinated multiple view.

One software pattern, which is a commonly used update mechanism e.g. in Ember and React, is the so-called “component” pattern. It has become so widespread and prevalent that it triggered even a web specification called “web components”. This section evaluates the most suitable component based JavaScript framework for coordinated multiple views and whether or not to follow the “web components” web specification.

Three JavaScript frameworks have been evaluated: (1) GlimmerJS (2) Google Polymer and (3) ReactJS. GlimmerJS is the templating engine of Ember, the framework which the author of this thesis is most familiar with. Applications written with GlimmerJS can be built and exported as web components. Google Polymer is the most popular framework for web components. React is another very popular framework developed by Facebook, but does not support web components at all.

Most importantly, coordinated multiple views require a way to exchange data between views, which is specific to interactions. The web component specification, unfortunately, does not specify how arbitrary JavaScript objects can be passed to web components. String based attributes are supported, as seen in Listing 3.1.

To pass rich data to components however, web component frameworks have to roll their own data flow and syntax. Google Polymer’s own syntax to pass rich data to a components is shown in Listing 3.2. But this is a custom solution that abandons standard HTML.

Listing 3.1: An example of string based attributes of web components (Bidelman 2017).

```

1 <google-map fit-to-marker api-key="AIzaSyD3E1D9b-
   Z7ekrT3tbhl_dy8DCXuIuDRC">
2   <google-map-marker latitude="37.78" longitude="-122.4"
     draggable="true"></google-map-marker>
3 </google-map>
```

Listing 3.2: A small syntax example how Google Polymer passes rich data to a component.

```

1 <some-component some-prop="{{richData}}></some-component>
```

This raises some problems in existing applications: A particular component-based frontend framework can not be assumed, a lot of existing applications are also written without any JavaScript framework. Custom solutions like the one of Polymer reduce the main motivation of implementing against web components: Platform agnostic flexibility.

As a summary, if there is (1) no obligation to implement web components and (2) an easy integration into an existing application is necessary, then React is the perfect choice for coordinated multiple views.

Table 3.1 shows the pros and cons of each framework for the use case.

Table 3.1: Comparison of component based web frameworks, advantages highlighted in green, disadvantages highlighted in red.

| Web Components | | Specific Framework |
|----------------|-------------------------|--------------------|
| standard | string-based attributes | no standard |
| GlimmerJS | Google Polymer | React |
| familiarity | maturity | maturity |
| | documentation | documentation |
| | community | large community |
| | | declarative style |
| | | small size |
| | | integrability |

3.4 Arbitrary Data Visualization Techniques

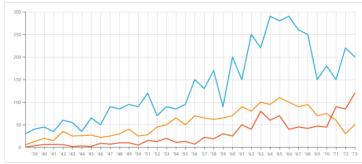
Even though the use case of the coordinated multiple view is focused on a 2.5D treemap and a geographic visualization, the implementation should allow any kind of data visualizations to be added and coordinated as well. Instead of implementing as many data visualizations as possible, the approach in this thesis is to analyse many data visualization techniques theoretically. The goal is to come up with a set of formal requirements and a specification that anticipates additional data visualizations. With regard to the software requirements in Section 3.2, the coordinated multiple view framework needs to be extensible and maintainable.

The data visualization catalogue by Severino Ribecca lists many of the most used data visualizations (Ribecca 2017). This section covers a selection of data visualization techniques from that catalogue. For each technique the expected data structure is examined and a list of plausible interactions is analysed. This analysis of an interaction specifies the required information, i.e. the information that would be given by an external view. The gathered knowledge is a preparation of the conceptual framework in Section 4.

3.4.1 Line Charts

Line charts display how quantitative values have changed over time. You can see an example in Figure 3.1. They are perfectly suited to show trends or compare multiple series of data with each other. Line charts visualize one or many series of data in parallel and therefore the expected data format is *tabular*.

Line charts are drawn in a Cartesian coordinate system, connecting subsequent points to each other. Thus, (1) position (2) orientation and (3) texture are constrained by the visualization technique. However, an interaction with the line chart can alter the (1) shape (2) color or (3) size of lines to communicate an interaction. It is further possible to highlight either the entire series of data or a single data point within that series, e.g. changing the shape and size of the point. Table 3.2 shows a list of plausible interactions in a line chart.



(a) Line charts

Fig. 3.1: Line charts connect data points in a cartesian coordinate system and are used to display trends.

Table 3.2: Plausible interactions for line charts.

| Category | Description | Required information |
|----------|-------------------------------|----------------------------|
| Select | Highlight a data point | id of data point |
| Select | Highlight a data series | id of data series |
| Encode | Change colours of data series | id of data series + colour |
| Filter | Restrict interval on x-axis | lower limit + upper limit |
| Filter | Hide a data series | id of data series |

3.4.2 Bar Charts and Multi-Set Bar Charts

Bar charts use either horizontal or vertical bars to show discrete, numerical comparisons across categories. The length of a bar displays a quantitative value of a category. You can see two examples in Figure 3.2.

Multi-Set bar charts display many data series next to each other. Every series is grouped by category and a colour can be used to identify a data series. Like line charts, bar charts expect a *tabular* data format. In contrast to line charts, bar charts are used to show a comparison rather than a trend.

The type of the visualization constrains the (1) shape, (2) size and, in case of a multi-set bar charts, (3) the colour of the visualization. An interaction can be shown by altering (1) position, (2) colour, (3) shape and (4) the texture of bars and columns. Table 3.3 lists some possible interactions.

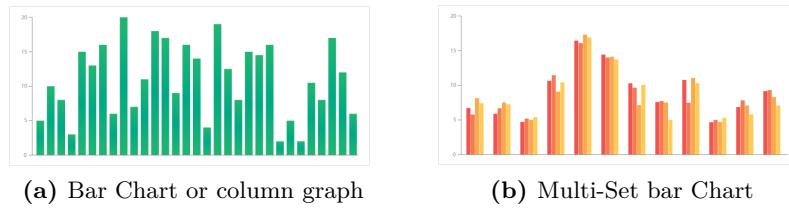


Fig. 3.2: A multi-set bar chart is a variation of a bar chart, showing data series grouped together under categories (Ribecca 2017).

Table 3.3: Plausible interactions for bar charts.

| Category | Description | Required information |
|-------------|----------------------------------|------------------------------------|
| Select | Highlight a bar | id of data point |
| Encode | Change colours of data series | id of data serie(s) + colour(s) |
| Reconfigure | Sort by attribute | name of data attribute |
| Reconfigure | Drag bars to reorder data series | ordered list of ids of data series |
| Filter | Hide a data series | id of data series |

3.4.3 Histograms

Histograms, as shown in Figure 3.3, visualize the distribution of data over a continuous interval or a certain time period. A special type is the population pyramid, which is a pair of back-to-back histograms, one for each sex.

Histograms and bar charts expect the same kind of data, i.e. a *tabular* data structure. Almost the same interactions as in Table 3.3 can be applied to histograms, except a re-ordering of bars along the x-axis. This is not possible, because the histogram constrains the position of bars along the interval.

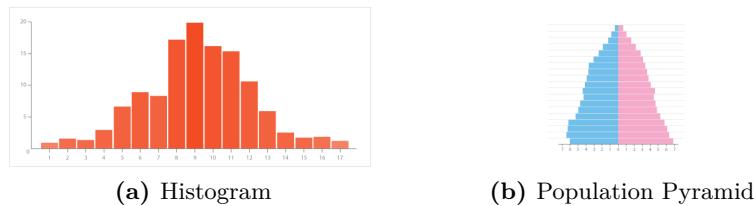


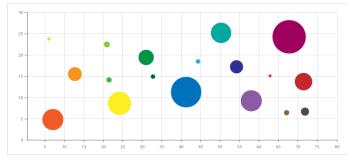
Fig. 3.3: A histogram is a bar chart over a continuous interval (Ribecca 2017).

3.4.4 Bubble Charts and Scatter Plots

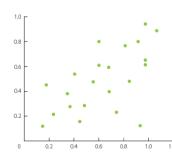
Both bubble charts and scatter plots are techniques to visualize continuous values from two data attributes. Points are placed with the two attributes in Cartesian coordinates in order to detect relationships and correlations. An example for each technique is displayed in Figure 3.4.

In case of bubble charts, each point is displayed as a bubble with a third value encoded in the size of bubbles. It is even possible to encode a fourth value in the colour of the bubble.

Like line charts, bar charts and histograms, a scatter plot expects *tabular* data. Each data point can take up to four values (in case of a coloured bubble chart). As seen in Table 3.4, interactions also include a zooming and movement of the viewpoint.



(a) Bubble Chart



(b) Scatter plot

Fig. 3.4: Bubble charts and scatter plots are similar regarding interactions (Ribecca 2017).

Table 3.4: Plausible interactions for bubble charts.

| Category | Description | Required information |
|-------------|----------------------------------|-------------------------------------|
| Select | Highlight a bubble | id of data point |
| Explore | Zoom in, zoom out | width and height of window |
| Explore | Move viewpoint position | x- and y-coordinates of viewport |
| Encode | Change colour mapping | id of data series + colour |
| Encode | Change colour function | mapping function of value to colour |
| Encode | Change mapping of size | mapping function of value to size |
| Reconfigure | Sort by attribute | data attribute |
| Reconfigure | Drag bars to reorder data series | ordered list of ids of data series |
| Filter | Hide a data series | id of data series |

3.4.5 Stacked Bar Charts

Unlike a multi-set bar graph which displays bars side-by-side, stacked bar graphs segment their bars of multiple datasets on top of each other. A baseline, as shown in figure 3.5 might be modeled as two back-to-back multi-set bar graphs. A reordering would e.g. move one data set from the left side to the right side. A stacked bar chart also expects *tabular* data.

If the stacked bar chart has a baseline, often the algebraic sign of the numeric value defines the placement of the segment on the left or on the right side. Table 3.5 shows possible interactions, including the highlighting of a data point, a change of color mapping or a reordering of the baseline.

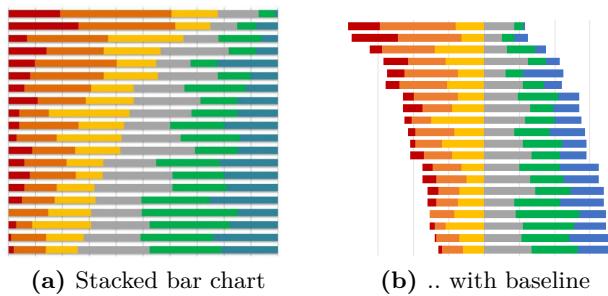


Fig. 3.5: Stacked bar charts can be ordered along a baseline or stretch to 100% width to show the percentage-of-the-whole of each group (Mann 2016) (Peltier 2016).

Table 3.5: Plausible interactions for stacked bar charts.

| Category | Description | Required information |
|-------------|------------------------|------------------------------------|
| Select | Highlight a bar | id of data point |
| Encode | Change colour mapping | id of data series + colour |
| Reconfigure | Sort by attribute | name of data attribute |
| Reconfigure | Specify stacking order | ordered list of ids of data series |
| Reconfigure | Flip data series | list of ids of data series |
| Filter | Hide a data series | id of data series |

3.4.6 Hierarchical visualizations

Tree maps show hierarchical data without ever exceeding the available screen. Each node is represented as a tile. Unlike a treemap a hierarchical ring diagram or sunburst diagram shows each level of the underlying tree as a series of rings.

Therefore, both treemap and ring diagram expect *hierarchical* data. Typically, each node will have at least one continuous value that can be used as input for the tiling algorithm or layout algorithm respectively. Additionally, each node can encode more attributes by colour.

As these visualization techniques are about hierarchies, the visible, maximal depth of the tree may be increased or decreased. Again, interactions could include a highlighting of data points and a change of color encoding. Both visualizations may show only a subtree. E.g. a click on a box in the treemap opens another treemap focused on the subtree. Similarly, a click on a slice of the ring

would surround the most external ring with the child nodes of the parent node. Table 3.6 gives a more comprehensive list of interactions.

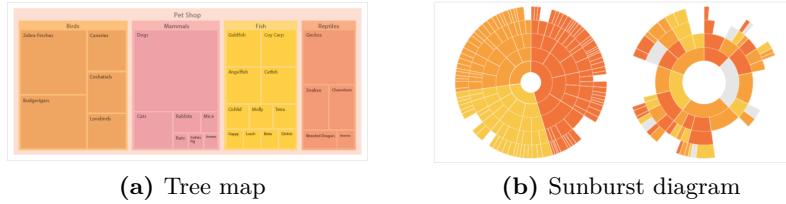


Fig. 3.6: Tree maps and sunburst diagrams are used to show hierarchies (Ribecca 2017).

Table 3.6: Plausible interactions for hierarchical visualizations.

| Category | Description | Required information |
|--------------------|--|------------------------------------|
| Select | Highlight a node | id of data point |
| Explore | Use another node as root of the visible tree | id of data point |
| Encode | Change mapping of category to colour | id of data series + colour |
| Reconfigure | Change data attribute used for layout | name of data attribute |
| Reconfigure | Sort by attribute | name of data attribute |
| Reconfigure | Specify order | ordered list of ids of data points |
| Abstract/Elaborate | Specify maximum depth of visible tree | number of hierarchy levels |

3.4.7 Geographic Data Visualizations

Choropleth maps and flow maps are thematic maps to visualize geographic data. Size, position and shape of a data point is determined by its geometry. Choropleth maps encode a continuous data attribute with relative values in the color of each region. Flow maps may display relationships between features, a data value defining the size, colour, direction or shape of each arrow. Figure 3.7 shows an example for each technique.

Non-geographic data may be given in a *tabular* form, assigned to each geographic feature. In contrast to tabular data, a flow map expects relationships between geographic features. Thus, it also expects *relational* data in form of a graph

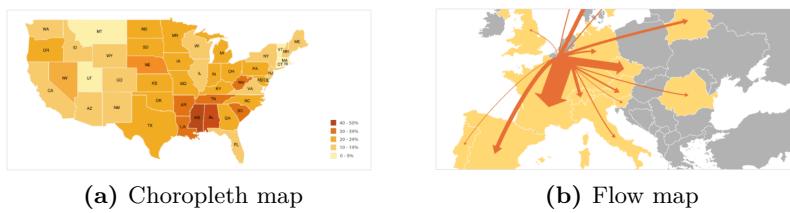


Fig. 3.7: Choropleth maps focus on a density while flow maps show a migration of data (Ribecca 2017).

Table 3.7: Plausible interactions for geographic visualizations.

| Category | Description | Required information |
|--------------------|---|-------------------------------------|
| Select | Highlight a feature | id of data point |
| Explore | Move viewport | latitude and longitude of viewpoint |
| Explore | Zoom in, zoom out | zoom factor |
| Encode | Change shape of marker | data id shape |
| Encode | Change mapping of category to colour | id of data series + colour |
| Encode | Change colour function | value + colour |
| Encode | Change data attribute used for colour | name of data attribute |
| Connect | Show relations of a feature | id of data point |
| Abstract/Elaborate | Change granularity of displayed regions | number of hierarchy levels |

3.4.8 Temporal Data Visualizations

Activity diagrams, like calendars and gantt charts, and timelines are temporal visualizations. Gantt charts and timelines have in common that each feature is represented as a rectangle, with the duration of the feature mapped to size and

position. A calendar, as shown in Figure 3.8, may show the event as a single marker for brevity. Calendars and gantt charts could not only read the data from the data source, but also add new features to the data set or update metadata of a feature, e.g. the progress of the activity. Calendars and gantt charts expect *tabular* data, although data points might reoccur on a regular schedule. So some data points, i.e. events, might repeat infinitely. A very common interaction is the filtering of the data set by selecting an interval over a timeline visualization, e.g. by dragging the mouse from left to right. A more comprehensive list is shown in Table 3.8.

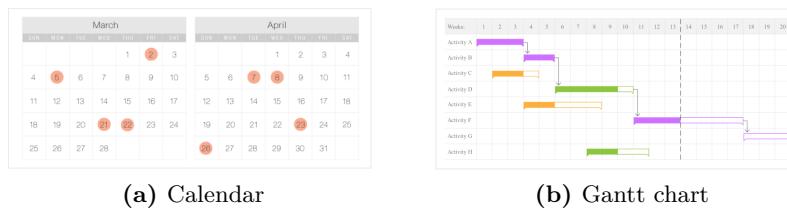


Fig. 3.8: Similar to a calendar, a gantt chart shows activities and the progress along a time line (Ribecca 2017).

Table 3.8: Interactions for temporal visualizations.

| Category | Description | Required information |
|----------|--|----------------------------|
| Select | Highlight a feature | id of data point |
| Explore | Show a different period of dates | start and end datetime |
| Explore | Show a different time interval | start and end hour |
| Encode | Change color of categories or activities | id of data series + colour |
| Encode | Change data attribute used for colour | data attribute |
| Filter | Remove a calendar or a category | id of data series |
| Filter | Filter data set by time interval | upper and lower limit |

Conceptual Framework

Based on the preparations in Chapter 3.4 this chapter specifies a conceptual framework for coordinated multiple views. The approach involves the definition of the terminology, a formalization of an interaction and the basic components of the conceptual framework. These components are derived from the interactions and the data structures discovered in Chapter 3.

4.1 The Conceptual Framework from the User Perspective

Figure 4.1 shows a component diagram of the conceptual framework. The user interacts via input devices, e.g. mouse or keyboard, with a computer. In order to see the effect of the interaction the user observes an output device like a screen.

The computer or the browser provides an API for these devices and the implementation of each view has access to this API. On the other side, views can communicate with each other through a coordinator. This is necessary in order to coordinate interactions.

Each view can have multiple triggers and multiple effects. A trigger is the handling of an event, caused by a user interaction, e.g. a mouse click. A effect is the change of the visual representation of the view, in order to communicate the interaction.

Each view is self-responsible for the implementation of triggers and effects. This is inevitable, as sometimes a view can not react to an interaction at all. E.g. a re-ordering in a parallel plot will not affect a scatter plot, where the position of items is constrained by coordinates.

Interactions of the same category also need to be distinguishable. E.g. the user could select a group of items with a bounding box. Additionally, the user moves the mouse cursor on an item within that group in order to highlight the item.

Therefore the message exchanged between views not only includes the interaction category and the relevant item but also an interaction purpose.

Every view can subscribe to named interactions at the coordinator. The coordinator notifies all subscribed views when a named interaction happens. In order to trigger an interaction, the visualization simply publishes the named interaction at the coordinator. This pattern is known as the publish-subscribe pattern and widely used in message queues.

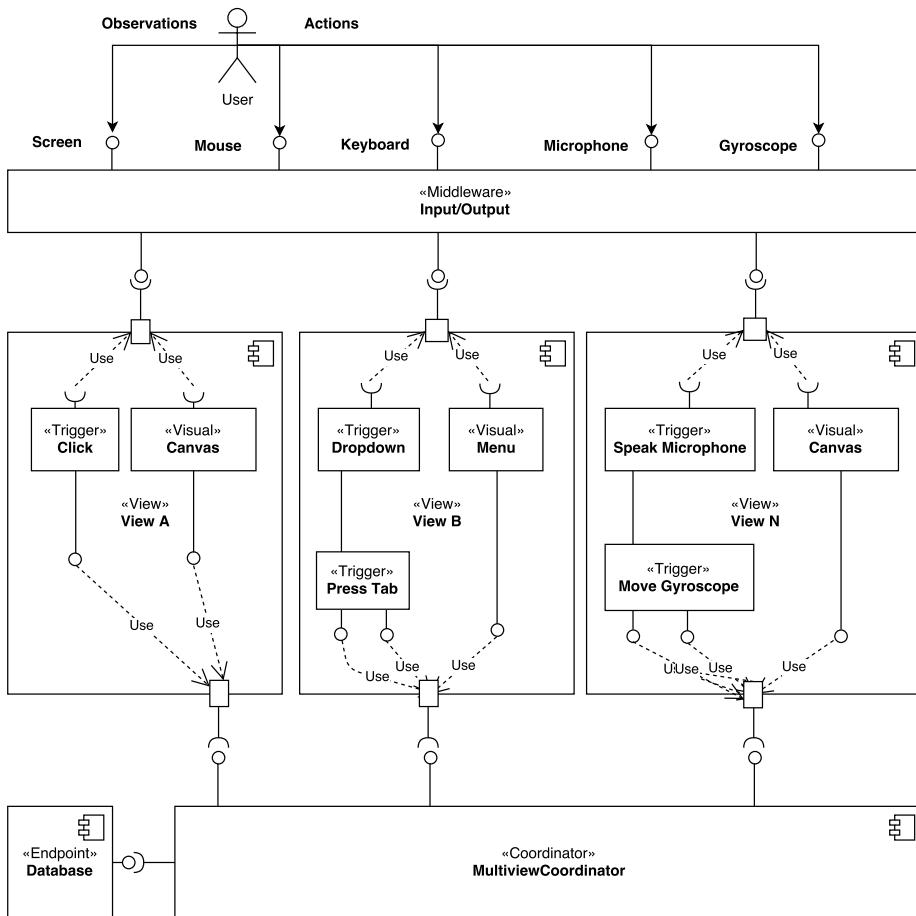


Fig. 4.1: Component diagram of the conceptual framework.

4.1.1 Interaction

In single views, an interaction I consists of at least a trigger and an effect:

$$I(T, E) \quad (4.1)$$

The purpose of the interaction in a single view does not need to be explicitly stated as such. E.g. hovering over a geographic area changes the background colour of the polygon and the user can identify the interaction as a highlighting. This implicit meaning gets explicit in the described conceptual framework, so that views are able to react reasonably to interactions in other, outside views.

4.1.2 Interaction in Coordinated Multiple Views

An interaction I in coordinated multiple views is formalized as

$$I(T, M(C, P, S), E) \quad (4.2)$$

for a trigger T , a message M and the effect E . The message M is defined by a category C , an application specific purpose P as well as an interaction subject S . The shared part of the interaction among multiple views is the message M .

4.1.3 View

A view is the section of the screen in which the user can see a data visualization. A coordinated multiple view system shows multiple views, either side by side on a single screen or on different screens which are physically placed next to each other.

4.1.4 Trigger

If the event of a user action is handled by a view and causes an interaction, this handling is called a trigger. The user e.g. clicks on a shape in the view, hovers over an area, selects an item from a dropdown menu, turns around a mobile device, speaks into the microphone or makes a particular gesture. As described in the introduction, views are responsible of their triggers.

4.1.5 Effect

The effect is the change of the visual representation subsequent to an interaction. In order to be perceivable by the user, the interaction must have some visual effect, e.g. a change of a visual variable according to Bertin (2010).

Some examples are: A change of colour of a selected bubble, a movement of the viewpoint, a rearrangement of attributes in a parallel plot or a higher level of detail in a 2.5D treemap.

Similar to a trigger, a view is self-responsible of its visual effects. Obviously visual effects are not shared, as a visual variable might be constrained due to the nature of the visualization technique itself. A re-ordering interaction in a parallel plot will not have an effect in bubble charts, as position is constrained by the type of visualization.

4.1.6 Interaction Category

An interaction category is the declaration and definition how the subject of the interaction should be changed. The aforementioned explicit meaning of the interaction is the smallest unit of information of the interaction. Some examples of categories include: Selection, Deletion, Point-of-Interest, Filtering, Reordering, Re-encoding. Categories can be classified with the interaction categories of Yi, Kang, and Stasko (2007).

4.1.7 Interaction Purpose

The interaction purpose describes the interaction in the context of a task or an application specific intention. A developer may want to have many select interactions. E.g. the user desires to select a detail view of an item under the mouse cursor from a previously selected set of already highlighted items. Therefore two interactions of the same category can be distinguished by a user-defined purpose.

4.1.8 Interaction Subject

The interaction subject refers to the target of the interaction. It defines what data or meta-data is affected by the interaction. E.g. when a user moves the mouse cursor on a line in the line diagram, that could highlight the data point under the cursor as well as the entire data series. Therefore the object affected

of an interaction is called the interaction subject. A subject can be a data point, a list of data points, a position of the viewpoint, a certain order of attributes or a mapping of attributes to visual variables.

4.2 Shared data model

The shared data model is a model of the data on which all visualizations must agree on. To account for data structures discovered in Chapter 3, the shared data model is inclusive enough for tabular, hierarchical and relational data. You can see a class diagram of this data model in Figure 4.2.

The entity class is used to model the smallest distinguishable unit. All entities can be identified and retrieved via the id. An entity is defined to be any object that can have data attributes attached as dimensions.

While entities describe what an object *is*, a dimension describes what it *has*.

An entity can have arbitrary many attributes and each value can be accessed by the name of the attribute. So if you want to get the latitude value of an entity, you can retrieve the value with a call to the dimension `latitude`.

Entities can also be series of other entities. A series contains an ordered list of contained entities. Series can also contain other series, this way a hierarchy relation can be modeled.

Every entity has a parent which is the series it is contained in. The root entity of the hierarchy has a parent which is `nil`. Every series has a special attribute `height` that describes the number of nested series or the height of the subtree.

To display tabular data only one or two levels of hierarchy are necessary. E.g. one level of hierarchy for a histogram and two levels of hierarchy for a stacked bar chart.

Other relations than hierarchical relations can be modeled as a relation entity. It represents a directed edge in a graph and must have incoming and outgoing entity. Since every relation is an entity as well, it can have attributes. These attributes may describe e.g. the weight of an edge in a flow map.

4.3 Encoding of Interaction Subjects

As described in Section 4.1.2, a message consists of the interaction category, purpose and subject. Category and purpose are just identifiers and can be encoded

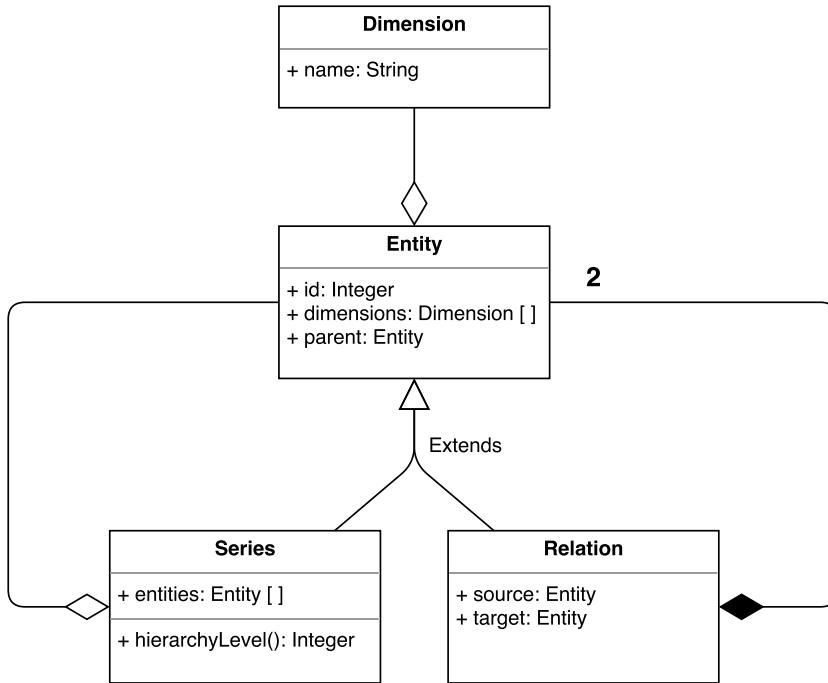


Fig. 4.2: A data structure for tabular, hierarchical and relational data.

as a number or a string. The subject is a highly variable object. It could define some state, e.g. explicitly stating the id of an entity or the name of an attribute. Alternatively, it could define some behaviour, e.g. implicitly defining if an entity is filtered based on its dimensions. Therefore, the approach in this thesis is to model the subject as a mathematical function. If the function has no input parameter, it just returns some state, i.e. the ids of entities, series, and relations or their respective attributes. If the function depends on input parameters, it may return e.g. `true` or `false` or define a sort order.

In this section, a couple of function declarations are derived from the examples in Section 3.4. Domain and range of these functions refer to the objects defined in the data model in Section 4.2.

4.3.1 Set definitions

The set of all entities \mathbb{E} in our subject space is defined as:

$$\mathbb{E} : \mathbb{E} \subseteq \mathbb{N} \quad (4.3)$$

Each entity can be represented by its `id`, so for simplicity \mathbb{E} is a subset all natural numbers in \mathbb{N} .

The set of all data dimensions D in our shared data model is defined as:

$$\mathbb{D} : \mathbb{D} \subseteq \Sigma^* \quad (4.4)$$

For simplicity, each dimension is represented as a sequence of characters, with the set of all sequences of characters written as Σ^* .

The set of all values of a data dimension d is defined as:

$$Space(d), d \in \mathbb{D} \quad (4.5)$$

So for an attribute `name`, $Space(name)$ would be the set of all strings.

Visual variables according to Bertin (2010) as in in Section 2.3.1 are defined as:

$$\mathbb{V} = \{position, size, shape, value, hue, orientation, texture\} \quad (4.6)$$

4.3.2 Function declarations

The functions *Select* and *Filter* operate on entities and can be used interchangeably:

$$Select : \emptyset \rightarrow \mathcal{P}(\mathbb{E}) \quad Filter : \mathbb{E} \rightarrow \{\perp, \top\} \quad (4.7)$$

The empty set is written as \emptyset and the set of all subsets of E is written as $\mathcal{P}(\mathbb{E})$.

Select just returns a subset of entities explicitly and expects no input. It can be used to highlight or focus entities, mark them for deletion or show details of an entity.

Filter returns for every entity either `true` or `false`, written as \top and \perp , depending on the entity and the value of its dimensions. Instead of an explicitly defined set of filtered entities, entities are defined implicitly. A good example would be a filter function checking upper and lower limit for each entity and a dimension called `price`.

$$Order : \emptyset \rightarrow \mathbb{E}^* \quad Sort : \mathbb{E} \times \mathbb{E} \rightarrow \mathbb{R} \quad (4.8)$$

The functions *Order* and *Sort* are used to order two entities $e1, e2 \in \mathbb{E}$. The following equivalencies are valid for *Sort*:

$$\begin{aligned} e1 < e2 &\Leftrightarrow \text{Sort}(e1, e2) < 0 \\ e1 = e2 &\Leftrightarrow \text{Sort}(e1, e2) = 0 \\ e1 > e2 &\Leftrightarrow \text{Sort}(e1, e2) > 0 \end{aligned}$$

Similar to *Select* and *Filter*, these order functions can be implemented either explicitly or implicitly. *Order* explicitly returns a sequence of entities, e.g. $(e4, e2, e5, e1, e3)$. *Sort* implicitly orders two entities $e1$ and $e2$ based on their dimensions. E.g. ordering entities based on the alphabetical order of their name would be an example of the latter.

$$\text{Encode} : \mathbb{D} \rightarrow \mathbb{V} \quad (4.9)$$

The *Encode* function can be used to change the mapping of dimension to a visual variable. E.g. bar charts, line diagrams, histograms and bubble charts can change the dimension mapped to their coordinate axes. Bubble charts can encode a different dimension in the **size** of the bubbles. Choropleth maps, treemaps and bubble charts can map a different attribute to **colour**. A specialized version of this function may return the attribute that is used for the layout algorithm in treemaps.

$$\text{Window} : \mathbb{D} \rightarrow W, \quad W = \{w | w \subseteq \text{Space}(d), d \in \mathbb{D}\} \quad (4.10)$$

For each of the dimensions in \mathbb{D} , the function returns the currently visible subset. E.g. a zooming or panning operation in a geographic visualization would result in a change of **latitude**, **longitude** and **zoom** of the viewpoint and therefore the currently visible section of the geographic vector space. The subset w can be encoded implicitly or explicitly. For continuous values two representatives **from** and **to** could define an interval. E.g. in a calendar the attributes **fromDay**, **toDay**, **fromHour** and **toHour** define the currently visible time section. If a dimension has discrete values, subset w could also be encoded with all values explicitly stated.

Implementation

This chapter describes the actual implementation, starting with a list of implemented interactions. After that, the components of the conceptual framework of Chapter 4 are translated into software components. Finally, the resulting architecture is explained for every component and the integration into the existing 2.5D treemap implementation is shown.

5.1 Implemented interactions

Figure 5.1 shows a screenshot of the final coordinated multiple view application. Different views are labeled and separated by red lines. The 2.5D treemap is on the left, the geographic visualization is on the right and some additional views to inspect and manually publish messages are below the geographic visualization.

In the course of this thesis the following interactions have been implemented:

- Select
 - Highlighting an entity is possible by moving the mouse cursor on a geographic feature in the geographic visualization or a block in the 2.5D treemap. The corresponding block or geographic feature will be highlighted respectively.
 - A click on a block in the 2.5D treemap selects this entity in the geographic visualization and vice versa. In Figure 5.1, district “Magdeburg” is selected in the geographic visualization. This district has an unusually high percentage of graduates compared to other districts with a similar unemployment rate.
 - Holding the control key, multiple clicks on a block in the 2.5D treemap creates a group of selected entities. Corresponding geographic feature or

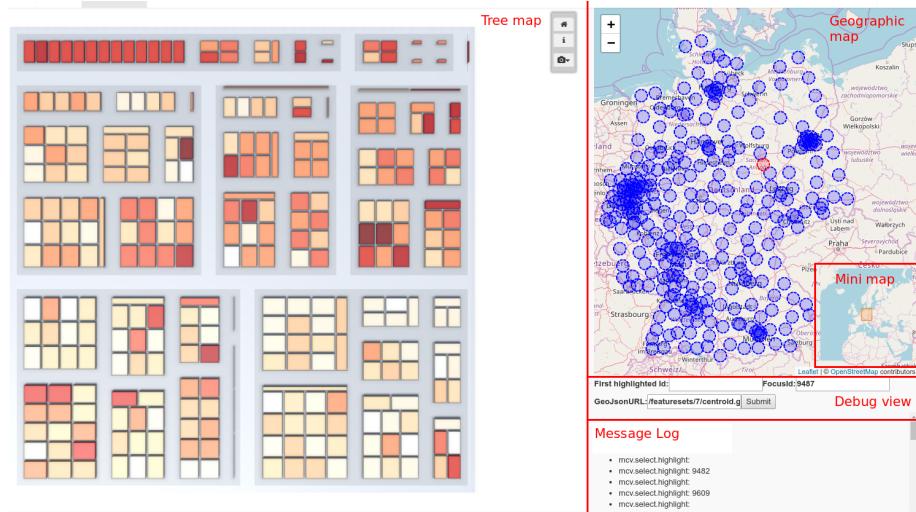


Fig. 5.1: This Figure shows the layout of the coordinated multiple view, red lines separating different views. The currently visualized data set is called “Wahlkreise” and consists of German administrative districts. The layout of the 2.5D treemap is based on the unemployment rate, the colour is based on the percentage of high-school graduates.

blocks in the 2.5D treemap or geographic visualization are selected as well.

- Explore
 - Selecting a block in the 2.5D treemap centers the viewpoint in the geographic visualization on the corresponding geographic feature.
 - Selecting a group of blocks in the 2.5D treemap centers the viewpoint in the geographic visualization on the boundaries of all corresponding geographic features.
- Reconfigure
 - Choosing a different data set updates the geometries in the geographic visualization.
 - Selecting another shape for the 2.5D treemap (e.g. point instead of polygon geometries) changes the visual representation in the geographic visualization.

5.2 Implementation of the Conceptual Framework

This section is intended to show precisely, how the specification of Chapter 4 is implemented and translated to actual code. As you recall, an interaction is defined as $I(T, (M(C, P, S), E))$, the message M is the part that is communicated between views. This message is published to a **publish-subscribe** broker. In Chapter 4 in Figure 4.1 this component is called **MultiviewCoordinator** and conveniently, the class has the same name in the actual code base. The coordinator has access to a database backend, just like in the component diagram, and the data will be fetched at runtime with an Asynchronous JavaScript (AJAX) call. The **Input/Output** middleware in our case is the browser because a web based application is developed. The browser provides an API for the mouse, keyboard, microphone and even the webcam.

5.2.1 PubSubJS as implementation for publish-subscribe

In accordance to Chapter 4, the implementation makes use of the **publish-subscribe** pattern for coordination. To achieve this, it uses the JavaScript library **PubSubJS** which is an topic-based, asynchronous implementation of the pattern. **PubSubJS** allows nested topics, i.e. topics and sub-topics, and a subscriber will be notified if a topic or any of its sub-topics is published. This means, a view can subscribe to a group of interactions. E.g. the group `mcv.select` includes both `mcv.select.highlight` and `mcv.select.focus`.

5.2.2 Encoding of category and purpose as topic and sub-topic

PubSubJS's nested topics can be used to encode categories and purposes. This approach gives views more flexibility, because they can either subscribe to every interaction, all interactions of a category, or just a few specific interactions. Figure 5.2 gives an example of a select interaction with the purpose to focus on three entities.

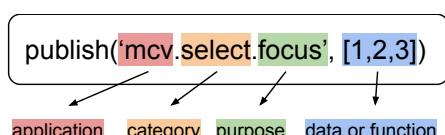


Fig. 5.2: Example of the encoding of interaction category and purpose.

5.2.3 Encoding of the subject as a JavaScript callback

Not only basic data types like numbers and strings can be published, as seen in Listing 5.1, but also functions. In javascript, these functions are called “callbacks”. If an interaction of type `filter` is published, the `data` parameter of the subscription callback is itself a callback. This callback would be a threshold function expecting an entity as parameter and returns whether or not the entity is filtered.

Listing 5.1: A simplified example how to subscribe to an interaction.

```

1 let coordinator = new MultiviewCoordinator();
2
3 // subscribe to a `select` interaction with purpose `highlight`
4 coordinator.subscribe('mcv.select.highlight', (msg, data) => {
5   console.log(msg, data);
6 });
7
8 // publish a topic asynchronously
9 coordinator.publish( 'mcv.select.highlight', 4711);

```

5.2.4 Data Model implemented as GeoJSON

GeoJSON is used as implementation of the shared data model. The format is used to encode a variety of geographic data structures (Internet Engineering Task Force 2016). Based on JSON, it can represent simple geographic features like points, lines and areas and reserves a properties object for non-spatial attributes.

GeoJSON allows to encode an `id` as a member in the feature object, which can be used to encode the `id` of the entity. Data attributes of entities are encoded as properties and hierarchical relationships are expressed via feature collections.

A GeoJSON example can be seen in listing 5.2, coordinates are omitted for brevity. The GeoJSON property `user_count_normalized` encodes a percentage of a user base per region. This is one of the attributes that can be mapped to the colour of a 2.5D treemap.

Listing 5.2: A GeoJSON example of a user distribution across German federal states. Coordinates are omitted.

```

1 {
2   "type": "FeatureCollection",
3   "features": [
4     {
5       "type": "Feature",

```

```

6     "id": 0,
7     "geometry": {
8       "type": "MultiPolygon",
9       "coordinates": [...]
10    },
11    "properties": {
12      "NAME_1": "Baden-Württemberg",
13      "state_code": "BW",
14      "user_count_total": "34",
15      "user_count_normalized": "0.10149253731343283"
16    }
17  }
18 ]
19 }
```

5.3 Architecture

The architecture of the implementation is depicted in the class diagram in Figure 5.3. For each of the views in Figure 5.1 you can see a corresponding class in Figure 5.3, i.e. `Treemap`, `MapComponent`, `MessageLog` and `DebugView`.

All views have a reference to `MultiviewCoordinator` in order to subscribe to interactions. The rendering 2.5D treemap is controlled by `UAController`, thus it is connected to `MultiviewCoordinator` through this class. The `MultiviewCoordinator` itself does not have a visual representation.

5.3.1 UAController

Written in plain TypeScript, the `UAController` handles mouse events and updates the 3D Canvas with Web Graphics Library (WebGL). The treemap is shown on the left side in Figure 5.1. Click and mouse move events already highlight the respective item or show a detail view of the item. It is part of the existing code base and so the only applied change to the `UAController` is the support of publications and subscriptions. The geographic visualization on the other hand needs to be implemented from scratch.

5.3.2 Geographic Visualization with React Components

The geographic visualization is implemented with React components, Figure 5.3 shows the superclass `ReactComponent` of the classes `MapComponent`, `MessageView` and `DebugView`. Every React components have a `render` method

which is called for ever re-rendering of the component in the virtual Document Object Model (DOM). This re-rendering gets triggered if the properties of a component change. Thus, if the properties of the parent component get passed to a child component, this child component will also re-render.

E.g. if the position of the view point of the `MapComponent` changes, that will re-render the included `Map` which depends on the position. If the Geometries of the visualized data set are updated, that will change the `GeoJSON` component but not the `Map`.

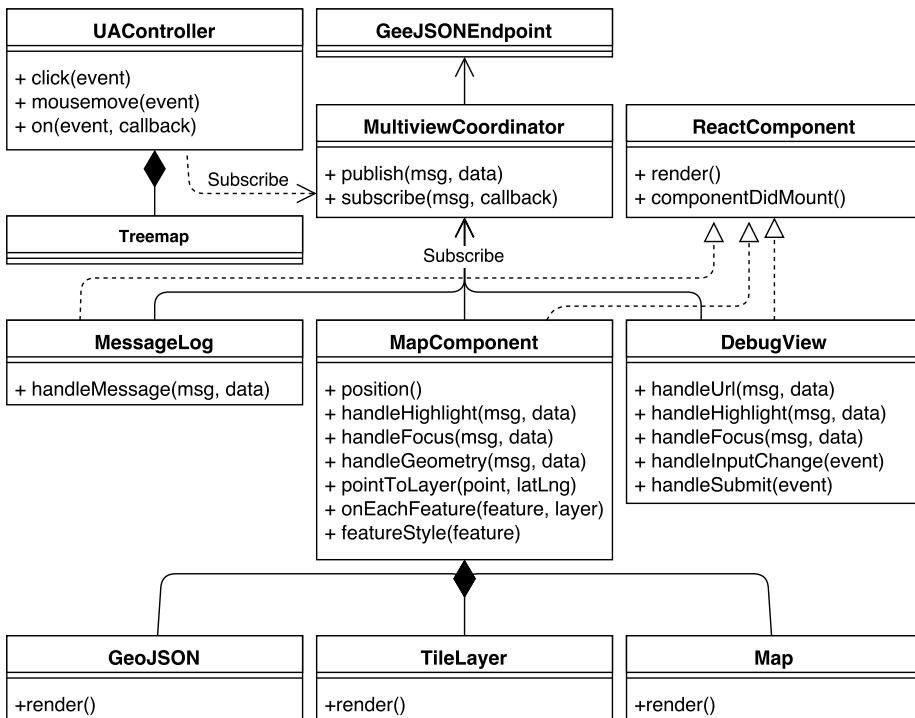


Fig. 5.3: Architecture of components. Every class with a `render` method is a React component. `MultiviewCoordinator` is used to coordinate 2.5D treemap and geographic visualization as well as to load geometry data.

5.4 Notifications

The automatic update mechanism works only within the virtual DOM of React components. Interactions among coordinated multiple views need to be implemented manually. For this purpose, the class `MultiviewCoordinator` wraps the library PubSubJS and works as a broker for interactions across coordinated multiple views.

Let's take an example from a common usage scenario: During initialization, all views subscribe to `MultiviewCoordinator` in order to re-render after interactions. Then the user chooses another data set and the geometries are updated. After that, the user clicks with the mouse on a geographic feature and the corresponding item is focused. You can see this scenario in the sequence diagram in Figure 5.4.

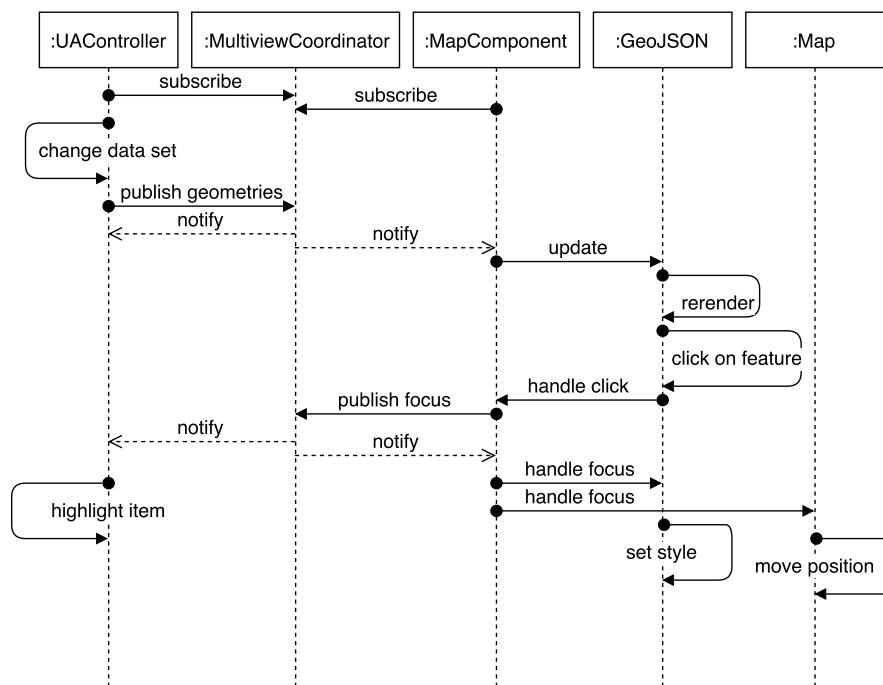


Fig. 5.4: This sequence diagram shows the notification of different components during a common usage scenario: First a change of the data set followed by a click on a feature.

5.5 Querying the GeoJSON endpoint

Each view gets a reference to the `MultiviewCoordinator`, e.g. during initialization, and can subscribe to interactions as seen in Listing 5.1. As a central part of the implementation, the `MultiviewCoordinator` is also used for certain performance optimizations and data integrity considerations. Therefore, it is responsible to query geometry data.

If the user selects a different data set, the `url` of the GeoJSON endpoint changes and an update of the geometry data is required. This behaviour is implemented by the `MultiviewCoordinator` which observes all changes to the `url` of the

GeoJSON endpoint. If the `url` changes, the coordinator fetches geometry data, publishes a change of the geometries and all subscribed views can re-render. This is also a performance optimization, as it reduces the number of requests to the GeoJSON endpoint.

5.6 Render views in the DOM

Adding the geographic visualization to the DOM is straightforward. Listing 5.3 shows an example how to use React's low-level API to render components inside the DOM. It uses a HTML id `multiview-map-component` to reference a particular node in the HTML document. The HTML document needs to import three JavaScript files, i.e. two imported libraries required by React and the compiled JavaScript application.

Listing 5.3: Example application written in TypeScript. Views can be added to the DOM individually. The implementation exposes convenient TypeScript declarations, here `MultiviewCoordinator` and `MapComponent` are imported.

```

1 import * as React from "react";
2 import * as ReactDOM from "react-dom";
3 import { MapComponent, MultiviewCoordinator } from 'urban-
  analytics-multiview-map-component';
4
5 let coordinator = new MultiviewCoordinator();
6
7 ReactDOM.render(
8   <MapComponent coordinator={coordinator}>,
9   document.getElementById('multiview-map-component')
10 );

```

The implementation of the conceptual framework is written in `TypeScript`. `TypeScript` is a typed superset of `JavaScript` that compiles to plain `JavaScript`. It provides optional static typing, classes and interfaces and helps to reduce errors by raising type errors at compile time. The implementation exports type-script declaration and you can see an import of the classes `MapComponent` and `MultiviewCoordinator` on lines 1 to 3 in Listing 5.3.

5.7 MapComponent

The geographical visualization makes use of three components of the React Leaflet library: `Map`, `TileLayer` and `GeoJSON`. Listing 5.4 shows the `render` method of the component.

Listing 5.4: render method of the Map component of the geographical visualization.

```

1  render() {
2    return (
3      <div className="multiview-map-component">
4        <Map ref={(m) => this._map = m} bounds={this.bounds()}>
5          <TileLayer
6            attribution='&copy; <a href="http://osm.org/copyright">
7              OpenStreetMap</a> contributors'
8            url='http://{s}.tile.osm.org/{z}/{x}/{y}.png'
9            />
10           { this.state.geojsonUrl && this.state.geojson &&
11             <GeoJSON
12               key={this.state.geojsonUrl}
13               data={this.state.geojson}
14               style={this.featureStyle}
15               pointToLayer={this.pointToLayer}
16               onEachFeature={this.onEachFeature}
17             >
18             </GeoJSON>
19           }
20         </Map>
21       </div>
22     );
23   }

```

The `render` method is the only required method of a React component. It will be invoked on the initial rendering of the component of the DOM and on every update of the component's properties.

React's templating language "JSX" allows to nest other child components into the React parent component. In this case the `Map` component includes a `TileLayer GeoJSON` component from the `react-leaflet` library. This library conveniently provides "React components for Leaflet maps." (Cam 2017).

The subscriptions to the coordinator happen during the method `didComponentMount`, a lifecycle callback method provided by React. You can see the respective interaction handlers `handleHighlight`, `handleFocus` and `handleGeometry` in the class diagram in Figure 5.3.

5.7.1 GeoJSON Component

The `GeoJSON` component is provided by the `Map` component with a couple of properties: It gets a (1) `geojsonURL` as well as a (2) `geojson` as data attribute. Furthermore a couple of callbacks is passed into the child component, including (3) `featureStyle`, (4) `pointToLayer` and (5) `onEachFeature`.

This way, the parent `Map` component controls the data flow and without a `geojson` object, no polygons are placed on the map. A changed `geojsonURL` will always update the child component as it is used a `key` on the `GeoJSON` component. The callbacks passed into the `GeoJSON` component control the visual representation of each polygon and they add event handlers for a mouse click or a mouse move on each polygon. Listing 5.5 shows the event handlers added to the map.

Listing 5.5: `onEachFeature` callback, adding handlers for mouse events.

```

1  onEachFeature(feature: geojson.Feature<geojson.GeometryObject>,
2    layer: Leaflet.Layer){
3    this.state.layerList.push(layer);
4    layer.on({
5      mouseover: () => {
6        this.state.controller.publish('mcv.select.highlight', [
7          Number(feature.id)]);
8      },
9      click: (event: Leaflet.LeafletMouseEvent) => {
10        if(event.originalEvent.ctrlKey){
11          this.state.controller.publish('mcv.select.focus', this.
12            xor(this.state.focusedIds, Number(feature.id)));
13        } else {
14          this.state.controller.publish('mcv.select.focus', [
15            Number(feature.id)]);
16        }
17      }
18    });
19  }

```

First all layers in the internal state of the `Map` component are cached. On each `mouseover` event, the `id` of the feature is published as `mcv.select.highlight` interaction. A `click` event is distinguished if the control key is pressed or not. In the latter case, the `id` of the feature is either added or removed from the list of focused ids and then the list of focused ids is published as `mcv.select.focus` interaction.

Listing 5.6: `featureStyle` callback, configuring the visual appearance depending on the currently highlighted or focused feature ids.

```

1  featureStyle(feature: geojson.Feature<geojson.GeometryObject>):
2    Leaflet.PathOptions{
3      const focused = (this.state.focusedIds.includes(Number(
4        feature.id)));
5      const fillColor = focused ? Color('red') : Color('blue');
6      let color = fillColor;
7      let weight = 2;
8      let dashArray = '3';
9      if (this.state.highlightedIds.includes(Number(feature.id))) {

```

```
8     weight = 4;
9     color = 'white';
10    dashArray = '';
11  }
12  return { color, weight, fillColor, dashArray };
13}
```

The `featureStyle` in Listing 5.6 method is very straightforward. If the feature is currently focused, the `fillColor` of the polygon is red, otherwise blue. Likewise, if the feature is currently highlighted, the polygon has a white, solid stroke.

Listing 5.7: `pointToLayer` callback, if a feature of `GeoJSON` has a point geometry, it will be shown as a circle.

```
1 pointToLayer(geoJsonPoint:any, latlng: Leaflet.LatLng){
2   return new Leaflet.CircleMarker(latlng);
3 }
```

Finally, it is configured how to display point geometries in callback `pointToLayer`. Since normal markers do not have a configurable color and style, the `GeoJSON` component is instructed to render a `CircleMarker` for each point geometry instead. This way, the same options of `featureStyle` can be applied to both point and area geometries.

Evaluation and Discussion

In this chapter the reference implementation of the conceptual framework is evaluated. First, a couple of use cases are described and how a geographical visualization can create more value. A performance evaluation is carried out, to discover technical limitations and to find performance bottlenecks. As a last step, the requirements of Chapter 3 are used to validate the conceptual framework and the reference implementation.

6.1 Use Case Scenarios

The following three use case scenarios demonstrate how more insights can be gained with geographic visualization next to the 2.5D treemap.

6.1.1 Explain outliers

2.5D treemaps can have outliers, i.e. unusual local maxima of a certain attribute. If the attribute is mapped to the height of the corresponding block, a local maximum is identifiable by a block that protrudes from a group of evenly leveled blocks. Such an example is shown in Figure 6.1.

The data set of the visualizations in the figure includes gas stations in Berlin. The 2.5D treemap is configured as follows: The layout is based on the brand name of a gas station, i.e. gas stations of the same brand like “Total” or “Aral” are grouped together. Height and colour of the blocks are mapped to the price of “Diesel” and “E10” respectively.

A 2.5D treemap with this configuration is suited to show a correlation of brand and price. Gas stations of brand “Total” are located on the center-left side of

the 2.5D treemap and they are rather expensive in general but with smaller price variations.

This is in strong contrast to gas stations of brand “ARAL” in the lower left: 10 gas stations can be identified as outliers, which are more expensive than other gas stations in the group. “Esso” and “Shell”, located in the center and on the lower right, show outliers, too.

Gas stations of the brands “HEM”, “Star” and “Sprint” are generally inexpensive and located in the groups with lighter colours at the top and on the right.

However, the 2.5D treemap alone is not able to explain the reason for certain outliers. What is special about the protruding blocks, i.e. the outliers within a group?

The geographic visualization can give a possible explanation: Many of those outliers are gas stations located next to a highway. These gas stations are slightly more expensive in general and significantly more expensive if they belong to the brand “ARAL”.

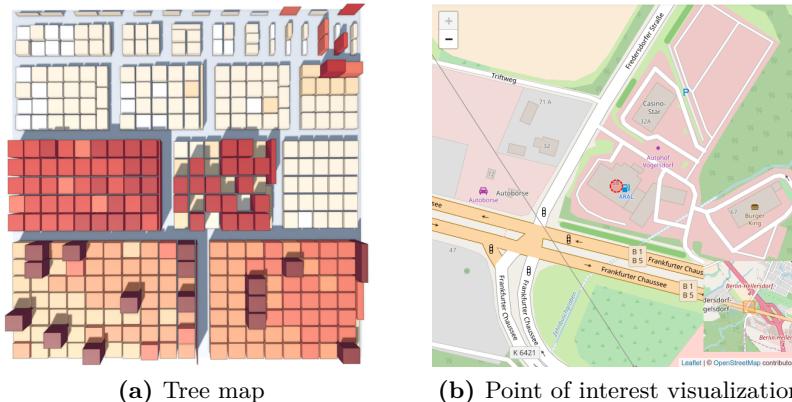


Fig. 6.1: Expensive gas stations, compared to other gas stations of the same brand, are often located next to highways.

6.1.2 Multi-Select in 2.5D treemap

This use case demonstrates the benefit of a multiple select and the ability to relate interesting features with their geographic context. Figure 6.2 (a) shows a 2.5D treemap visualizing German constituencies with the following configuration: The layout is based on the population density, large clusters on the right

are sparsely populated districts. Colour is based on the increase or decrease of inhabitants, a red colour indicating a growth of inhabitants.

There are three groups of items in the 2.5D treemap that catch our interest: (1) A group of orange items on the left side of treemap, (2) many scattered, white coloured items in the large cluster on the right and (3) three purple coloured items in the large cluster on the right.

Districts in the orange group are rather densely populated and show a decline of population. But as seen in Figure 6.2 (b) the districts of this group have a geographic context: They are all located in the Ruhr area.

A similar observation can be made about the white group, i.e. sparsely populated districts with a serious decline of population: These districts are all located in the east of Germany, see Figure 6.3 (a).

And finally, even the three purple districts in the cluster of sparsely populated districts are geographically related as well: These districts are in the vicinity of Munich, as seen in Figure 6.3 (b).

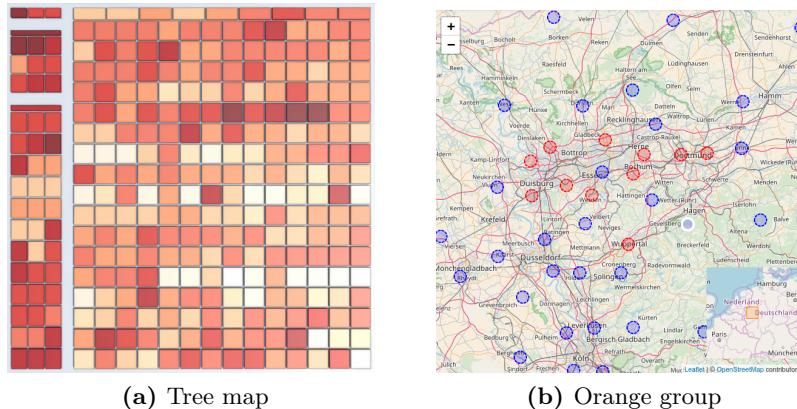


Fig. 6.2: The orange coloured group on the left, districts with a moderate decline of inhabitants and large population density, are all districts of the Ruhr area.

6.1.3 Bounding Box Selection in geographic visualization

Similar to the multiple select in Section 6.1.2, multiple selects can be carried out by a bounding box in the geographic visualization.

The layout of the 2.5D treemap in Figure 6.4 (a) is based on construction activity, i.e. the number of completed accommodations per capita. The colour is

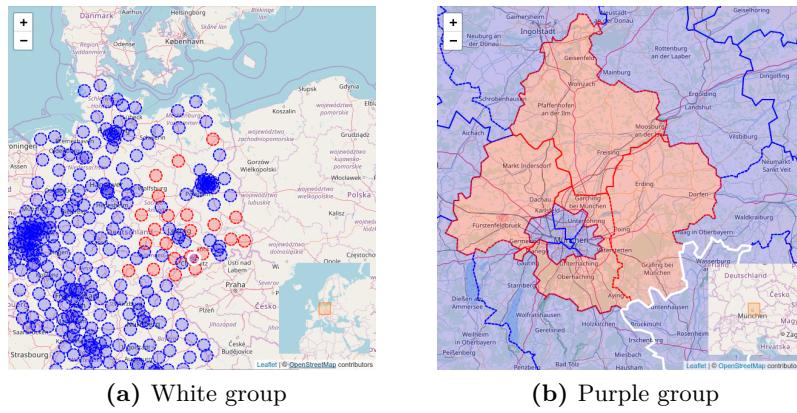


Fig. 6.3: The sparsely populated districts with a serious decline of inhabitants are all located in the east of Germany. The few districts with a low population density and high increase of population are all located in the vicinity of Munich.

mapped to the increase of population. As Berlin is known for quickly rising rents and real estate speculation, all districts in Berlin are selected with a bounding box. All selected districts are placed in the center right group of the 2.5D treemap. They are the dark red items next to the currently highlighted item in Figure 6.4 (a). But as seen in Figure 6.4 (a), the districts are not in the group with the highest construction activity at the very top.

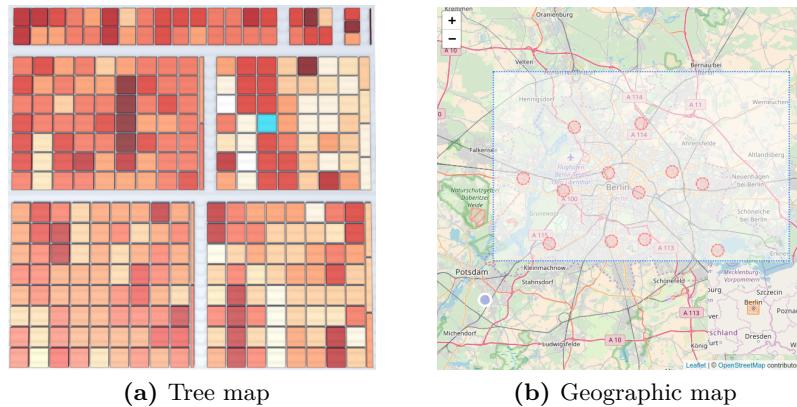


Fig. 6.4: A bounding box multiple select in the geographic visualization can reveal a non-geographic context in the 2.5D treemap

6.1.4 Use case summary

The scenario in Section 6.1.1 is a great example how a geographic visualization can create more insights. The 2.5D treemap alone allows to identify the outliers and also shows a possible correlation of price and brand. However, without the geographic visualization it is not possible to draw a hypothesis of the proximity to a highway being the reason for the high prices. The geographic context is just not present in the raw data.

Deliberately selecting a particular area of interest in the geographic visualization by clicking on an interesting item in the 2.5D treemap can be considered to be a new technique.

Similarly, in the scenario in Section 6.1.2, the 2.5D treemap alone can show outliers or group of outliers. However, it is not possible to see the geographic relation between those items in that group immediately. The geographic visualization makes the geographic context visible. In this scenario, the multiple selection really helps to select items of the white group, which is scattered across the 2.5D treemap.

Finally, in the scenario of Section 6.1.3, the reverse is also possible: Identify a non-geographic relation in the 2.5D treemap by selecting geographically related items with a bounding box. In this case we can see a very similar construction activity.

6.2 Evaluation According to Guidelines

Wang Baldonado, Woodruff, and Kuchinsky (2000) introduced eight rules for multiple views. This section checks the use case scenarios in Section 6.1 according to these guidelines and if a positive impact on utility could be achieved.

A positive example are the rules of diversity, complementary and decomposition to improve memory and comparison (*ibid.*). Their benefits are apparent in the use case of the visualization of gas stations in Section 6.1.1. The geographic visualization reveals a hidden correlation, decomposes complex data into chunks and provides additional insights. Less apparent are the benefits of the rules of self-evidence and parsimony. In the mentioned use cases in Section 6.1 insights are not self-evident but must be revealed with the right configuration. Another positive example is the rule of attention management. If the user selects one or many administrative districts in the treemap in Section 6.1.2, the geographic

visualization next to it zooms on it. In summary, the benefit of a support of learning and comparison outweighs the drawback of computational overhead.

6.3 Performance Evaluation and Limitations

The following performance evaluation was carried out with the built-in runtime performance analysis feature of the Chrome browser. In particular, a Chromium Browser was used in Version 62.0 (64 Bit). The hardware specifications of the machine are listed in Table 6.1.

Table 6.1: Hardware specifications.

| | |
|----------------|---|
| Device name: | LENOVO ThinkPad L540 |
| CPU type: | Intel i3-4100M CPU @ 2.50GHz |
| #CPUs: | 4 |
| Main memory: | 8GiB |
| Graphics card: | Intel 4th Gen Core Processor Integrated Graphics Controller |

A couple of data sets were used in three different scenarios: (1) The 2.5D treemap without a geographical visualization, just publishing interactions, (2) an example application of the geographical visualization without a 2.5D treemap and (3) both visualizations together.

In the first and second scenario, the data set is loaded, some data points are highlighted and then some data points are focused with a single-select and a multi-select holding the control-key. In the second scenario and third scenarios, which have a geographical visualization, many items are selected with a select box while holding the shift-key. For every scenario there are six data sets, that is 18 profilings, and each scenario profiling took about 60 seconds to finish.

Table 6.2 shows the list of data sets used for profiling the performance of the reference implementation. The largest data set consists of German administrative districts called “Landkreise Deutschland” with a total size of 2.13 MiB. The data set with the highest number of features is called “Immoscout Wohnungsangebote” with 8601 coordinates German real estates, totalling 2.11 MiB.

Table 6.2: Data sets used for performance profiling, ordered by file size.

| Data Set Name | #Features | Type | Size (MiB) |
|-------------------------------|-----------|--------|------------|
| Bundesländer Deutschland | 16 | Areas | 0.64 |
| Tankstellen Berlin | 366 | Points | 0.75 |
| Wahlkreise BT 2009 | 299 | Areas | 0.91 |
| Regierungsbezirke Deutschland | 31 | Areas | 0.94 |
| Immoscout Wohnungsangebote | 8601 | Points | 2.11 |
| Landkreise Deutschland | 402 | Areas | 2.13 |

6.3.1 Immoscout

The slowest profiling is the visualization of data set “Immoscout”. Chrome’s runtime analysis shows a red bar at the top of the screen if the frames per second drop in such a way that it impairs the perceived interactivity. You can see a screenshot of the analysis in Figure 6.5.



Fig. 6.5: During profiling of both 2.5D treemap and geographic visualization visualizing the “Immoscout” data set, the frame per second rate drops to 1 FPS.

As you can see in Figure 6.6 the 2.5D treemap spends almost the entire CPU time in scripting. The geographic visualization has a more balanced CPU time, spending time for painting and rendering during focusing interactions.

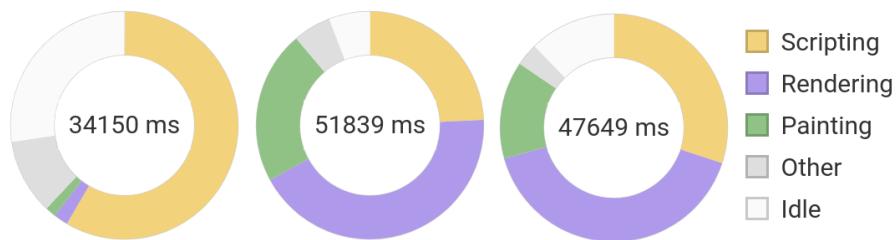


Fig. 6.6: Summary of profile data for data set “Immoscout” of 2.5D treemap only (left), both 2.5D treemap + geographic visualization (center) and only geographic visualization (right).

Going through the timeline, the handling of the “mousemove” event can be identified to be the likely cause of this slow scripting. The 2.5D treemap constantly checks the point or polygon which is under the mouse cursor. This is likely a performance problem.

The profile summary looks totally different for the scenario of a 2.5D treemap along with a geographical visualization. Compared to just a 2.5D treemap only, much more time is spent during painting and rendering. This is caused by the fact, that LeafletJS moves the viewpoint and zooms if a feature is focused. This can cause a network request and will re-render background tiles.

Note that not the communication between views hits the performance but rather the change of visual representation of views.



Fig. 6.7: The sawtooth pattern of the callstack during a `highlight` interaction indicates an expensive iteration of all features of the `GeoJSON`.

Figure 6.5 shows spikes whenever an interaction is made. Most of these interactions are `highlight` interactions, when the user moves the mouse cursor. Figure 6.7 shows the lower part of the callstack during such a `highlight` interaction.

The dominating subroutine is identified as `setStyle` which spans almost the entire callstack. Below this subroutine, you can see a lot of quick calls for each layer. LeafletJS iterates through all geographic features in order to update the style, e.g. change the stroke width. Therefore, `setStyle` seems to be costly operation for a large number of features. This would explain why many small features have a stronger performance impact than fewer but larger features.

6.3.2 Landkreise

The profiling of data set “Landkreise” seems to support that assumption. This data set is larger than “Immoscout” but has fewer features. Nevertheless, the frame per second rate rarely drops in a way which has an impact on interactivity, as you can see in Figure 6.8.

Fig. 6.8: Larger but fewer features seem to have positive effects on the frame rate.



The geographic visualization alone idles almost 50% of the CPU time, as you can see on the right side of Figure 6.9.

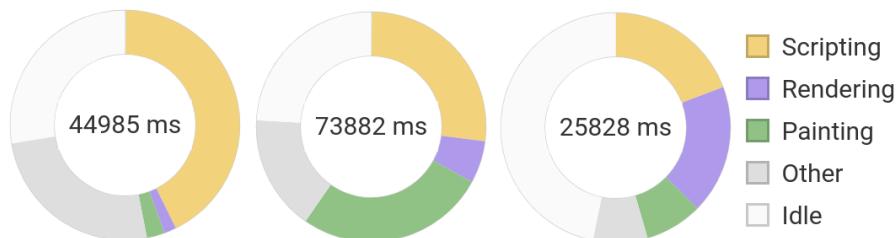


Fig. 6.9: Summary of profile data for data set “Landkreise” of 2.5D treemap only (left), both 2.5D treemap + geographic visualization (center) and only geographic visualization (right).

6.4 Discussion of Requirements

In this section, the reference implementation is evaluated based on the requirements from Section 3.2.

Serialization of interactions depends on the serialization of the published message (M, C, P) as defined in Chapter 4. The interaction category and purpose are just strings and therefore trivial to serialize, as shown in Section 5.2.2. The interaction subject, however, can be an arbitrary JavaScript object. The estimated size and data type of that object determines how graciously a message can be serialized.

In our case, the largest JavaScript object that is transmitted between views are geometries during initialization. The largest data set in Table 6.2 is called “Immoscout Wohnungsangebote” with a file size of 2.1 MiB. Geometries and meta data serialize well in form of GeoJSON. The performance analysis in Section 6.3 does not indicate a large impact during initialization. Apparently, geometries seem not to be the bottleneck.

Instead, the analysis shows a performance impact for very frequent interactions, e.g. highlighting of entities. Yet, the interaction subject is a list of a few ids, in most cases just one, and is therefore easy to serialize with a small resulting size. The impact on performance is caused rather by data sets with a high number of features.

As an alternative to explicitly specifying the interaction subject, a interaction subject can be specified as a function. In JavaScript, functions can be serialized with `toString()` and deserialized with `eval`. So the serialization for callbacks is straightforward, although the use of `eval` is considered unsafe.

Reversibility is connected to the serialization of interaction messages, too. The undoing of an interaction needs to be implemented in each view separately. If the interaction framework is used to publish an `undo` interaction, there are two possible implementations: (1) Each subscribed view keeps a record of every received interaction in the past and can replay these interactions up to the desired step (2) if the effect of the interaction is reversible, e.g. the inverse of a function exists, the inverse of a function is published. Both options have an undesired impact on the required memory, because a record of published messages needs to be kept in memory. If that record is long, it will take a long time to replay every interaction up to the desired step. As a result, reversibility is non-trivial to implement and has a potential performance impact.

Extensibility The coordinated multiple view system is designed for loose coupling, independent views and almost no shared state. It is lightweight and has a low complexity. The downside of this approach is that it does not come with a huge simplification of the implementation effort in individual views. Therefore, the system is very extensible and scales well, but the development effort for each interaction remains high.

Maintainability as described in Section 3.2 means how much other parts of the code are impacted by an interaction and how error-prone the framework is. The framework benefits of the main advantages of the publish-subscribe pattern, i.e. loose coupling and scalability. On the other hand, it suffers from the main disadvantages of this software pattern, the decoupling of publisher and subscriber. Therefore, a proper debugging of all components in the whole

system is necessary. Nevertheless, each view in the coordinated multiple view system has small dependencies and is therefore easy to maintain and test.

Summary and Conclusion

In thesis a coordinated multiple view was developed to visualize multi-dimensional, hierarchical and geographical data. An existing treemap implementation was complemented with a geographic visualization. Interactions in the categories *Select*, *Explore* and *Reconfigure* were coordinated between these views.

Apart from the specific use case, interactions in coordinated multiple views were formalized. Precisely, a shared data model, an encoding of the different parts of the interaction and a communication pattern was specified. This specification provides the basis for the actual implementation, making the system extensible for additional data visualizations and interactions. The specification itself may help developers of future coordinated multiple view frameworks.

The actual implementation is lightweight, scales well and has loose coupling. With respect to the requirements of a framework of coordinated multiple views it shows good serialization and extensibility. Yet, the framework is responsible only to coordinate interactions among views, it does not reduce effort of implementing interactions within views.

Some issues became apparent in the performance analysis. Re-rendering of views gets slow for large data sets with many features, as every feature is iterated. The 2.5D treemap spends too much time during picking which is caused by an unnecessary expensive lookup of feature ids. Especially for frequent interactions like highlighting, this can impair the interactivity and leads to poor user experience. This performance issue could be fixed by selectively updating only those features which have been changed.

It was demonstrated how a coordinated multiple view can be used to improve comprehensibility and interactivity of 2.5D treemaps. A treemap provides the technical context while the geographic visualization next to it establishes the geographic context. Focusing on a section of a map by clicking on an item in the

treemap was discovered as an investigation technique. It was possible to guess the reason for outliers in the data set by exploring the geographic surrounding. Furthermore, selecting a group of items in the treemap could reveal correlations in the geographic visualization.

Data analysts can use the new knowledge by applying treemaps in the context of decision support systems. This applies especially in application scenarios with a strong geographical context, e.g. local administration and urban planning.

7.1 Future Work

The developed coordinated multiple view framework provides a basis for additional data visualizations and more coordinated interactions. In the future, the conceptual framework should be validated and tested with more data visualizations. This is necessary as new use case scenarios often reveal unidentified problems.

Out of scope of this thesis is the configuration and layout of multiple views next to each other. Information visualization often provides the ability to explore and discover unanticipated correlations and anomalies. More opportunities for data exploration would be generated if the user could add additional data visualizations on the fly. In this case, a change of the use case or group of coordinated visualization techniques would not require a change to the code base. This functionality could be completed with a feature to save these layouts to and load them from disk respectively.

References

- Agafonkin, Vladimir and OpenStreetMap contributors (Aug. 1, 2017). *Leaflet*. URL: <http://leafletjs.com/> (visited on 11/08/2017).
- Andrienko, Gennady and Natalia Andrienko (2007). “Coordinated multiple views: a critical view”. In: *Coordinated and Multiple Views in Exploratory Visualization, 2007. CMV'07. Fifth International Conference on*. IEEE, pp. 72–74.
- Bertin, Jacques (Jan. 2010). *Semiology of Graphics: Diagrams, Networks, Maps*. ISBN: 978-1-58948-261-6.
- Bidelman, Eric (Nov. 18, 2017). *google-map*. URL: <https://www.webcomponents.org/element/GoogleWebComponents/google-map> (visited on 11/20/2017).
- Bjork, Staffan, Lars Erik Holmquist, and Johan Redstrom (1999). “A framework for focus+ context visualization”. In: *Information Visualization, 1999.(InfoVis' 99) Proceedings. 1999 IEEE Symposium on*. IEEE, pp. 53–56.
- Bladh, Thomas, David A. Carr, and Jeremiah Scholl (2004). “Extending Tree-Maps to Three Dimensions: A Comparative Study”. In: *Computer Human Interaction: 6th Asia Pacific Conference, APCHI 2004, Rotorua, New Zealand, June 29-July 2, 2004. Proceedings*. Ed. by Masood Masoodian, Steve Jones, and Bill Rogers. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 50–59. ISBN: 978-3-540-27795-8. DOI: 10.1007/978-3-540-27795-8_6. URL: http://dx.doi.org/10.1007/978-3-540-27795-8_6.
- Bostock, Mike (July 2017a). *Crossfilter*. URL: <http://square.github.io/crossfilter/> (visited on 07/10/2017).
- (Mar. 27, 2017b). *d3 Tutorial: Coordinated Views*. URL: <http://bl.ocks.org/dukevis/6768900> (visited on 12/04/2017).
- Bruls, Mark, Kees Huizing, and Jarke J Van Wijk (2000). “Squarified treemaps.” In: *VisSym*. Springer, pp. 33–42.
- Cam, Paul Le (Nov. 10, 2017). *react-leaflet*. URL: <https://github.com/PaulLeCam/react-leaflet/commits/master> (visited on 11/17/2017).

- Carpendale, MST (2003). “Considering visual variables as a basis for information visualisation”. In:
- Dale, Tom (Apr. 5, 2017). *Introducing Glimmer.js*. URL: https://emberjs.com/blog/2017/04/05/emberconf-2017-state-of-the-union.html#toc_introducing-glimmer-js (visited on 12/16/2017).
- Davenport, Thomas H. (Apr. 3, 2013). *How P&G Presents Data to Decision-Makers*. URL: <https://hbr.org/2013/04/how-p-and-g-presents-data> (visited on 12/04/2017).
- Dix, Alan and Geoffrey Ellis (1998). “Starting Simple: Adding Value to Static Visualisation Through Simple Interaction”. In: *Proceedings of the Working Conference on Advanced Visual Interfaces*. AVI '98. L'Aquila, Italy: ACM, pp. 124–134. DOI: 10.1145/948496.948514. URL: <http://doi.acm.org/10.1145/948496.948514>.
- Döllner, Jürgen (July 2017). *Visualization Techniques for Big Data*. URL: <https://hpi.de/doellner/masterarbeiten/visual-software-analytics.html> (visited on 07/11/2017).
- Erbacher, Robert F (2007). “Panel Position Statement: The Future of CMV”. In: *Coordinated and Multiple Views in Exploratory Visualization, 2007. CMV'07. Fifth International Conference on*. IEEE, pp. 75–75.
- Few, Stephen (2013). “Data visualization for human perception”. In: *The Encyclopedia of Human-Computer Interaction, 2nd Ed.*
- Figueroa, Pablo, Mark Green, and Benjamin Watson (2001). “A framework for 3D interaction techniques”. In: *CAD/Graphics*. Vol. 8, pp. 22–24.
- Foster, Mike (2017). *Life as a Modern Day Cartographer - Redefining Cartography for the 21st Century*. URL: <http://mjfoster83.github.io/modern-cartography/#/> (visited on 12/13/2017).
- Friendly, Michael and Daniel J Denis (2001). “Milestones in the history of thematic cartography, statistical graphics, and data visualization”. In: URL <http://www.datavis.ca/milestones> 32.
- Garlandini, Simone and Sara Fabrikant (2009). “Evaluating the effectiveness and efficiency of visual variables for geographic information visualization”. In: *Spatial information theory*, pp. 195–211.
- Ghoniem, Mohammad et al. (2015). “Weighted maps: treemap visualization of geolocated quantitative data.” In: *Visualization and Data Analysis*, 93970G.
- Google Inc. (Sept. 22, 2017). *Polymer Project*. URL: <https://www.polymer-project.org/> (visited on 09/23/2017).
- Ho, Quan (May 2013). *Architecture and Applications of a Geovisual Analytics Framework*. Chap. 2. ISBN: 978-91-7519-643-5.
- Internet Engineering Task Force (Aug. 1, 2016). *GeoJSON*. URL: <http://geojson.org/> (visited on 11/20/2017).

- Johnson, Brian and Ben Shneiderman (1991). "Tree-Maps: A Space-filling Approach to the Visualization of Hierarchical Information Structures". In: *Proceedings of the 2Nd Conference on Visualization '91. VIS '91*. San Diego, California: IEEE Computer Society Press, pp. 284–291. ISBN: 0-8186-2245-8. URL: <http://dl.acm.org/citation.cfm?id=949607.949654>.
- Keim, Daniel A. (Jan. 2002). "Information Visualization and Visual Data Mining". In: *IEEE Transactions on Visualization and Computer Graphics* 8.1, pp. 1–8. ISSN: 1077-2626. DOI: 10.1109/2945.981847. URL: <http://dx.doi.org/10.1109/2945.981847>.
- Kusinitz, Sam (July 18, 2014). *12 Reasons to Integrate Visual Content Into Your Marketing Campaigns*. URL: <https://blog.hubspot.com/marketing/visual-content-marketing-infographic> (visited on 07/08/2017).
- Lavrač, Nada et al. (2007). "Data mining and visualization for decision support and modeling of public health-care resources". In: *Journal of Biomedical Informatics* 40.4. Public Health Informatics, pp. 438–447. ISSN: 1532-0464. DOI: <http://dx.doi.org/10.1016/j.jbi.2006.10.003>. URL: <http://www.sciencedirect.com/science/article/pii/S153204640600116X>.
- Li, Sam, Douglas Stockwell, and Andy Mutton (Nov. 17, 2017). *What are web components?* URL: <https://www.webcomponents.org/introduction> (visited on 11/20/2017).
- Limberger, Daniel et al. (2016). "Dynamic 2.5D Treemaps Using Declarative 3D on the Web". In: *Proceedings of the 21st International Conference on Web3D Technology*. Web3D '16. Anaheim, California: ACM, pp. 33–36. ISBN: 978-1-4503-4428-9. DOI: 10.1145/2945292.2945313. URL: <http://doi.acm.org/10.1145/2945292.2945313>.
- MacEachren, Alan M and Menno-Jan Kraak (2001). "Research challenges in geovisualization". In: *Cartography and geographic information science* 28.1, pp. 3–12.
- Mann, Robert (Aug. 5, 2016). *Who Lies More: a Comparison*. URL: <http://mannmetrics.com/who-lies-more/#more-1> (visited on 12/13/2017).
- Mayer, Richard E. (1997). "Multimedia learning: Are we asking the right questions?" In: *Educational Psychologist* 32.1, pp. 1–19. DOI: 10.1207/s15326985ep3201_1. URL: http://dx.doi.org/10.1207/s15326985ep3201_1.
- McAfee, Andrew and Erik Brynjolfsson (Oct. 2012). *Big Data: The Management Revolution*. URL: <https://hbr.org/2012/10/big-data-the-management-revolution> (visited on 07/08/2017).
- National Center for Chronic Disease Prevention and Health Promotion (2010). *Choropleth Map of Obesity in the United States (1995 to 2008)*. URL: <https://www.cdc.gov/obesity/strategies/choropleth-map.html>

- //homes.cs.washington.edu/~jheer//files/zoo/ex/maps/choropleth.html (visited on 10/09/2017).
- Nocke, Thomas and Heidrun Schumann (2002). "Meta Data for Visual Data Mining". In: *Proceedings Computer Graphics and Imaging, CGIM'02*. Citeseer.
- Observatory of Economic Complexity (n.d.).
- Peltier, Jon (Aug. 4, 2016). *Diverging Stacked Bar Charts*. URL: <https://peltiertech.com/diverging-stacked-bar-charts/> (visited on 12/13/2017).
- Poleto, Thiago, Victor Diogho Heuer de Carvalho, and Ana Paula Cabral Seixas Costa (2015). "The Roles of Big Data in the Decision-Support Process: An Empirical Investigation". In: *International Conference on Decision Support System Technology*. Ed. by Boris Delibašić et al. Cham: Springer International Publishing, pp. 10–21. ISBN: 978-3-319-18533-0. DOI: 10.1007/978-3-319-18533-0_2. URL: http://dx.doi.org/10.1007/978-3-319-18533-0_2.
- Ribecca, Severino (Sept. 11, 2017). *The Data Visualisation Catalogue*. URL: <http://www.datavizcatalogue.com/index.html> (visited on 09/23/2017).
- Roberts, J. C. (July 2007). "State of the Art: Coordinated Multiple Views in Exploratory Visualization". In: *Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization (CMV 2007)*, pp. 61–71. DOI: 10.1109/CMV.2007.20.
- Roderick, Morgan (May 21, 2017). *PubSubJS*. URL: <https://github.com/mroderick/PubSubJS> (visited on 11/09/2017).
- Rouse, Margaret (July 1, 2017). *Guide to telling stories with data: How to share analytics insights*. URL: <http://searchbusinessanalytics.techtarget.com/definition/data-visualization> (visited on 12/04/2017).
- Shneiderman, Ben (Jan. 1992). "Tree Visualization with Tree-maps: 2-d Space-filling Approach". In: *ACM Trans. Graph.* 11.1, pp. 92–99. ISSN: 0730-0301. DOI: 10.1145/102377.115768. URL: <http://doi.acm.org/10.1145/102377.115768>.
- (1996). "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations". In: *Proceedings of the 1996 IEEE Symposium on Visual Languages*. VL '96. Washington, DC, USA: IEEE Computer Society, pp. 336–. ISBN: 0-8186-7508-X. URL: <http://dl.acm.org/citation.cfm?id=832277.834354>.
- Wang Baldonado, Michelle Q., Allison Woodruff, and Allan Kuchinsky (2000). "Guidelines for Using Multiple Views in Information Visualization". In: *Proceedings of the Working Conference on Advanced Visual Interfaces*. AVI '00. Palermo, Italy: ACM, pp. 110–119. ISBN: 1-58113-252-2. DOI: 10.1145/345513.345271. URL: <http://doi.acm.org/10.1145/345513.345271>.

- Weaver, Chris (2007). “Is coordination a means to collaboration?” In: *Coordinated and Multiple Views in Exploratory Visualization, 2007. CMV’07. Fifth International Conference on*. IEEE, pp. 80–84.
- Wood, Jo and Jason Dykes (2008). “Spatially ordered treemaps”. In: *IEEE transactions on visualization and computer graphics* 14.6.
- World Wide Web Consortium (W3C) (Feb. 13, 2017). *Web components current status*. URL: <https://www.w3.org/standards/techs/components> (visited on 07/24/2017).
- Yi, Ji Soo, Youn ah Kang, and John Stasko (2007). “Toward a deeper understanding of the role of interaction in information visualization”. In: *IEEE transactions on visualization and computer graphics* 13.6, pp. 1224–1231.

Declaration of Authorship

I declare that I have developed and written the enclosed thesis completely by myself, and have not used sources or means without declaration in the text. Any thoughts from others or literal quotations are clearly marked. The Master Thesis was not used in the same or in a similar version to achieve an academic grading or is being published elsewhere.

Potsdam, December 19, 2017

Robert Schäfer