

Cluster visualization for interactive maps

Seminar in Software Engineering

Josef Dabernig

Matrikelnummer 0927232

an der
Fakultät für Informatik der Technischen Universität Wien

Betreuung
Betreuer: Prof. Dr. Silvia Miksch

Wien, 3.5.2013

Abstract

This thesis investigates ways of visualization for presenting clustered information on maps. Performance and readability of digital mapping applications decreases when displaying large amounts of data. Clustering can be used to group overlapping items and reduce visual clutter of the map presentation.

Visualization techniques for putting clusters on a map are researched and evaluated in an exploratory analysis. Existing concepts from visual cluster analysis and geovisualization, as well as established classifications form the foundations of the presented study. *Map types* as well as *cluster visualization techniques* are presented. The *evaluation* classifies the stated techniques for cluster visualization on maps.

The presented map types for visualizing clusters range from standard, *geographic maps* with markers, *Heat maps* to *Dot Grid maps* and a *Voronoi* map. Cluster visualization techniques include *icon-based/glyphs*, *pixel-based*, as well as *geometric/diagrams*. The evaluation of visualization techniques for clusters on a map proposes a custom set of criteria. *Showing the number of items within clusters* is considered as a main feature of simple cluster visualizations on maps. Further, the study identifies *showing cluster areas* and *extra cluster information* as two additional aspects of primary interest when comparing approaches to cluster visualization on maps.

The results are discussed and put into practice by evaluating the visual aspects of *Geo-cluster*, a server-side clustering implementation for maps.

Contents

Contents	ii
1 Introduction	1
1.1 Structure	2
1.2 Related work	2
1.3 Method	3
2 Foundations	4
2.1 Visual variables	6
2.2 Visual data exploration techniques	8
2.3 Clutter reduction	9
3 Results	12
3.1 Map visualization types for clustering	12
3.2 Cluster visualization techniques for maps	16
3.3 Evaluation of visualization techniques for clusters on a map	21
4 Discussion	25
4.1 Geocluster visualization	26
4.2 Visual evaluation of Geocluster	28
4.3 Conclusions	30
List of Figures	31
Bibliography	33

Introduction

Digital mapping applications on the Internet are strongly emerging. Big players like Google Maps¹ and OpenStreetMap² provide online maps, that users can view and interact with. Maps allow telling stories and communicating data in a visual way. They can be used to get a quick overview of points-of-interest in a certain area. If a large amount of information is contained in such an area, visual clutter is causing problems. Obviously, when telling a story, information needs to be told in a compact way as the human brain can only process a limited amount of data at the same time [NÖ6, Del10].

Clustering³ is a technique for grouping objects with similarities that can be used to reduce visual clutter. Clustering points on a map enhances performance and readability of data-heavy map applications. Client-side clustering uses JavaScript to group overlapping items. Server-side clustering is needed when too many items slow down processing and create network bottle necks. This paper aims at discussing visual aspects of Geocluster [Dab13], a server-side clustering implementation for maps. It considers well-separated, spatial clusters based on Euclidean distance.

The primary goal is to *evaluate cluster visualization techniques for creating interactive maps based on large data sets*. Visual ways for presenting clustered data on maps should be researched and summarized. Combining existing concepts from visual cluster analysis and geovisualization should provide a better understanding of theoretic backgrounds and put into context. The idea is to draw practical conclusions from the overlap of these two complex areas of research. Given the lack of existing publication on visualizing clusters on maps for reference, the study primarily aims at exploration of concepts and ideas instead of being a complete reference.

¹<https://maps.google.at>

²<http://www.openstreetmap.org/>

³http://en.wikipedia.org/wiki/Cluster_analysis

1.1 Structure

The given introduction is expanded by a discussion of *related work* in the field of geovisualization and visual cluster analysis. In addition, the *methods* being used in this paper are presented.

Chapter 2 describes the theoretic *foundations* being researched. Basic concepts of geovisualization as visual variables, visual data exploration techniques and clutter reduction are explained.

Chapter 3 discusses the main *results* of the study. Map visualization types, as well as cluster visualization techniques are explained and summarized within an exploratory evaluation.

Chapter 4 is a *discussion* of the previously presented results and applies the evaluation of cluster visualization techniques for maps to the practical use case of Geocluster. Finally, conclusions of the presented results and discussions are made.

1.2 Related work

While many publications have been found on either cluster analysis or geovisualization, few of them touch upon both subjects at the same time. There hasn't been found any study that compares the different techniques for visualizing clusters on maps as its primary goal.

A primary resource instead are publications on individual approaches for cluster geovisualization: Heat maps [LCC11, Fis07], Voronoi Diagrams [Del10, PdOMA06], Convex Hulls [CPCM08]. A well researched blog article talks about different ways to visualize clusters on maps [Bet07]. A more complex visualization is provided by Bristle Maps [KMM⁺13]. Multiple cluster visualization techniques on maps can also be combined [AAH⁺12]. Another form are self-organizing maps, which represent clustered information in virtual space [GHKV09].

On a practical note, Leaflet.markercluster⁴, the OpenLayers Cluster Strategy⁵ and MarkerClusterer for Google Maps⁶ are common implementations for client-side clustering that visualize clusters on maps.

For classification of the researched techniques, 3 taxonomies have been considered. “Visual variables” by Jacques Bertin [Ber67, Ber83] with some adjustments by Alan

⁴<https://github.com/Leaflet/Leaflet.markercluster>

⁵<http://openlayers.org/dev/examples/strategy-cluster.html>

⁶<http://google-maps-utility-library-v3.googlecode.com/svn/trunk/markerclusterer/docs/reference.html>

M. MacEachren [Mac95] and Leland Wilkinson [Wil05]. “Classification of visual data exploration techniques” by Daniel A. Keim [Kei01] and the “Clutter Reduction Taxonomy” by Geoffrey Ellis and Alan Dix [ED07]. Further, related taxonomies can be found [LBWR94, Shn96].

In addition, publications on cluster visualization and multi-variate data visualization techniques [War02, Zha07, EDG⁺08, EF10], geovisualization [NÖ6, MK01, MK97] and diagrams [Lad12] are resources for the individual disciplines that are being combined in this paper.

1.3 Method

Scientific publications as well as real-world implementations have been researched in order to get an overview of existing technologies related to visualizing clusters on maps. Given the lack of comparative publications in that particular field, papers on the individual topics of visualization, geovisualization and cluster analysis have been compared for overlapping concepts. As no fixed criteria, despite the general context and goals of the paper have been set, the research was mainly conducted in an explorative way. The criteria of published classifications and taxonomies have been reviewed to match the use case for visualizing clusters on web maps.

Based on an overview of foundations in geovisualization and cluster analysis, the main part of the study tries to enumerate the diverse approaches for visualizing clusters on maps. This is done from two perspectives: *map visualization types* cover the whole visualization and *cluster visualization techniques* focus on individual clusters being represented on maps. Finally, an attempt for classification has been made as part of the evaluation based on exploratory analysis: custom criteria derived from existing taxonomies are defined and the researched technologies are classified amongst those criteria.

CHAPTER 2

Foundations

In the following chapter, foundations for visualizing clusters on a map will be discussed. Basics of geographic visualization and their driving forces lead to abstraction and clustering as a tool for simplifying information on maps. Visual variables, as well as classification approaches for geovisualization will be reviewed in order to discuss existing concepts for displaying aggregated data on a map.

Visualization is driven by the basic belief that ‘seeing’ is a good way of understanding and generating knowledge. Humans have a very well developed sense of sight, a fact which is underlined by more of 50 percent of the neurons in our brain being used for vision. [Fie08].

MacEachren & Kraak [MK01] define that “Geovisualization integrates approaches from visualization in scientific computing (ViSC), cartography, image analysis, information visualization, exploratory data analysis (EDA), and geographic information systems (GISystems) to provide theory, methods, and tools for visual exploration, analysis, synthesis, and presentation of geospatial data (any data having geospatial referencing).” In his lecture notes on “Geographic visualization”, Martin Nöllenburg adds that more human-centered definitions exist and observes that the user’s needs have to be taken into account for effective geovisualization techniques [NÖ6].

The **goals of geovisualization** can be summarized using the *map use cube* by MacEachren and Kraak [MK97], which is illustrated in figure 2.1. The goals *exploration, analysis, synthesis and presentation* are classified amongst three dimensions:

- The type of *task* varies from *knowledge construction* to *information sharing*. While the first is about revealing unknowns and constructing new knowledge, the latter will primarily share existing knowledge.

- The amount of *interaction* ranges from *high* to *low*. A low level of interaction means a rather passive consumption of knowledge, instead a high level will allow the user to actively influence the visualization.
- The *users* of visualization are classified between *public* and *private* audiences. A single, private user with specialized skills might require different visualization techniques than large, public audiences.

In the given model of the map use cube, the goals of exploring, analyzing, synthesizing and presenting shift between the extremes of the three defined aspects. The first goal of exploring is classified as a task of knowledge construction, based on high interaction and targeted at a rather private audience. On the other hand, presenting is a task of information sharing that requires a low amount of interaction but is suitable for public audiences [NÖ6].

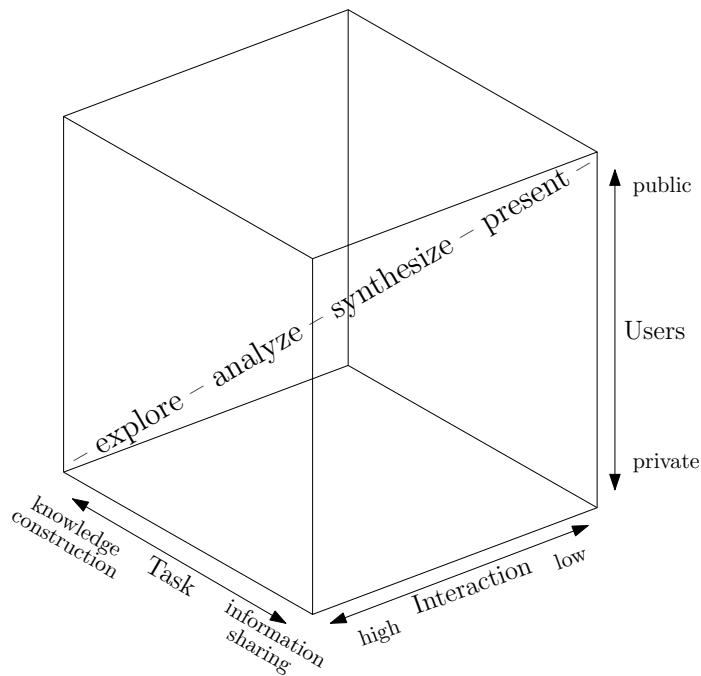


Figure 2.1: The map use cube after MacEachren and Kraak [MK97] characterizing geovisualization goals in a three-dimensional space by their level of interaction, their audience, and the addressed tasks. [NÖ6].

Cognitive aspects of visualization help us understand how visual thinking works. Complex input is abstracted on the retina of the human eye and matched against a vast collection of patterns from experience. Despite generating realistic images, visualization

can help generate new ideas by using abstraction to communicate patterns. The idea is to allow the user to join insight, draw conclusions and interact with the data by presenting it in a visual form that reduces the cognitive work needed to perform a given task [MG90, Kei01, NÖ6].

Various scientific publications [PN00, MG90, Kei01, Har08, ED07, Del10, NÖ6] that have been researched for this thesis mention the importance of using abstraction for efficiently visualizing information. Especially maps can only highlight interesting information by filtering out unnecessary details of the environment. For example, a road map is better visualized on a clear background instead of satellite images that would distract the user from the primary goal of finding directions. The challenge is to balance realism and abstraction in geovisualization depending on the problem [NÖ6].

Nöllenburg [NÖ6] names three “Driving forces in geovisualization”: The advent of high speed parallel processing and technology advances in computer graphics, today allows us to grasp enormous amounts of information. Besides the advances in graphics and display technologies, the second main driving force in geographic visualization is the increasing amount of geospatial data being collected and available. Finally, the third force is the *rise of the Internet*, which significantly pushes web mapping and contributes to geovisualization technologies.

As a logical consequence of broader audiences having access to geospatial visualizations using the Internet, it appears that there is a shift from technology-driven visualization towards more human-centered approaches. Interactive and highly dynamic interfaces have helped the map evolve from its traditional role as a presentational device towards exploring geospatial data [NÖ6, Fie08].

2.1 Visual variables

Information on a map is represented by symbols, point, lines or areas with properties such as color and shape. Bertin [Ber67, Ber83] has described the fundamental graphic variables for map and graphic design. While being written for hand-drawn maps on paper, the concepts described by Bertin are still applied in todays digital mapping applications and have been further developed by MacEachren [Mac95].

The main variables, introduced by Bertin are: *location, size, value, texture, color, orientation* and *shape*. MacEachren adds *crispness, resolution, transparency* and *arrangement* to the list and splits up color into its values of *brightness, saturation* and *hue*. Figure 2.2 describes a similar list of aesthetic attributes by different geometries and groups the variables regarding *form, color, texture* and *optics* [Wil05].

Each variable has a different potential for visualizing data of categorical information (nominal and ordinal) or numerical information (including intervals and ratios). For

example, the size of a point can be used to describe a numeric ratio and different shapes may be used to distinguish items based on categories. On the other hand, different texture patterns only offer a limited set of possibilities and size shouldn't be used for describing nominal properties [NÖ6, Mac95].

	Point	Line	Area	Surface	Solid
Form					
Size	• • ●	=====	□ □ □	~~~~~	█████
Shape	● ■ ▲	=====	△ ○ ◇	~~~~~	
Rotation	◀ ▶ ▷	=====	□ △ ○	~~~~~	█████
Color					
Brightness	● ● ○	=====	██ ██ □	~~~~~	█████
Hue	● ● ○	=====	■ ■ ■	~~~~~	█████
Saturation	● ● ●	=====	■ ■ ■	~~~~~	█████
Texture					
Granularity	◆ ◆ ◆	====		~~~~~	█████
Pattern	◆ ◆ ◆	=====	■■■	~~~~~	█████
Orientation	◆ ◆ ◆	=====		~~~~~	█████
Optics					
Blur	● ● ●	=====	██ ██ █	~~~~~	█████
Transparency	● ● ●	=====	██ ██ █	~~~~~	█████

Figure 2.2: Aesthetic Attributes by Geometry [Wil05].

2.2 Visual data exploration techniques

Depending on the task and the type of data to be shown, different forms of visualization and techniques for exploring the data exist. Daniel A. Keim [Kei01] classifies such techniques using three criteria as depicted in figure 2.3: the *data type* to be visualized, the *technique* itself and the *interaction and distortion* method [Del10]:

- The **data type** is classified into *one-dimensional* (as for example temporal data), *two-dimensional* (geographical maps), *multi-dimensional* (relational tables), *text and hypertext* (articles and web documents), *hierarchies and graphs* (networks) as well as *algorithms and software* (such as debugging operations).
- The **visualization techniques** are classified into *standard 2d/3d displays* (x-y plots and landscapes), *geometrically-transformed displays* (parallel coordinates), *iconic displays* (glyphs), *dense pixel displays* (recursive pattern) and *stacked displays* (tree maps).
- Finally, the third dimension describes the **interaction technique** being used, such as *dynamic projection*, *interactive filtering*, *zooming*, *distortion* and finally *link & brush* approaches.

With regards to visualizing clusters on a map, it appears that the visualization technique may be considered from two different view points:

- the visualization of the *entire map* with clustered points on it, as well as
- the visualization of an *individual cluster*, placed on the map.

A typical map for representing spatially clustered data is based on a least *two-dimensional data*, containing latitude and longitude information of cluster items. The map is presented in planar space, which classifies it amongst *standard 2d/3d displays*. Common interaction techniques for maps are *zooming* and *panning* which allow the user to explore the 2-dimensional space and reveal details.

The visualization of individual clusters on the map is likely to be classified amongst *iconic displays*. During the clustering process, individual points get aggregated, which potentially leads to multivariate aggregate information of item properties. Depending on the level of detail that clusters should expose, more complex visualization techniques may be possible.

Approaches for visualization techniques of the entire map and individual clusters will be discussed further in chapter 3.

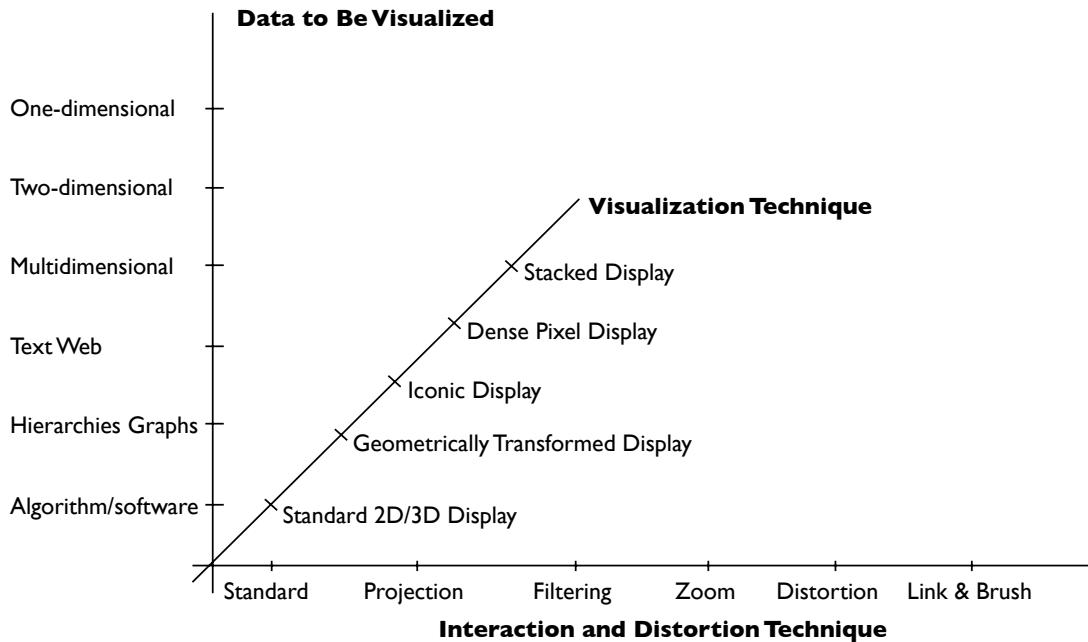


Figure 2.3: Classification of visual data exploration techniques based on [Kei01].

2.3 Clutter reduction

Clutter reduction is a way to enhance readability and general performance of maps. An early publication about *visual clutter* on maps by Richard J. Phillips and Liza Noyes [PN00] states that “reducing visual clutter improves map reading performance”. Clutter reduces the background visibility and prevents the user from understanding structure and content of the data being presented. This especially becomes true for visualizing large data sets on maps, so that properties of the data items are hardly visible [Har08, Del10].

Clustering of course is the approach for clutter reduction that is primarily being investigated for this thesis. In order to review the effectiveness and limitations of clustering, as well as the relationship that it has to other techniques in that field, it is helpful to review the “Clutter Reduction Taxonomy” by Geoffrey Ellis and Alan Dix [ED07]. It distinguishes between three main types of clutter reduction techniques:

- **appearance:** alter the look of data items by using techniques like *sampling*, *filtering*, *changing point sizes*, *changing opacity* or *clustering*.

- **spatial distortion:** displace the data items in ways as *point/line displacement*, *topological distortion*, *space-filling*, *pixel-plotting* or *dimensional reordering*.
- **temporal:** use *animation* to reveal additional information.

In a next step, the stated techniques have been evaluated against a list of 8 high-level criteria [ED07]. For this thesis, the relevant information for the clustering technique have been extracted and will be outlined by each criterium:

1. *avoids overlap*

The major benefit is to reduce clutter, provide the ability of seeing and identifying patterns, have less hidden data, as well as giving more display space to points. Clustering can be used in such a way to avoid *overplotting* by representing groups of points as single points.

2. *keeps spatial information*

Maintaining the correct coordinates of items is relevant, but the study also states that relative positions of points might have greater influence on orientation than just their exact coordinates. Clustering by definition looses the spatial information of individual points, but using aggregate values, like the centroid a cluster, as the position can be used for compensation.

3. *can be localized*

The term localization is used to specify, if the display can be reduced to a specific region. This is usually provided by *focus and context* techniques that reveal information underneath by zooming into an area. The study doesn't make a clear decision regarding the applicability of localization for clustering and states that different properties for spatial and non-spatial clustering apply. In the case of spatial clustering, localization is definitely possible and has been implemented, see chapter 4.1.

4. *is scalable*

Scalability of the clutter reduction technique with regards to large amounts of data is the goal of this criterium. The study admits that the meaning of large datasets is vaguely quantified. As one of the main goals for this thesis is to enhance performance for large data sets by using clustering, the technique is expected to satisfy this condition. Of course, the range of scalability depends on the implemented clustering algorithm.

5. *is adjustable*

If the user is able to control aspects of the visual display and adjust parameters of the system that influence the degree of display clutter. Scientific methods in

cluster analysis tend to offer a higher level of interactivity, while public facing clustering applications limit the amount of interactivity to controlling cluster sizes and the previously discussed localizing feature.

6. *can show point/line attribute*

The goal is to map attributes of the data to properties like color, shape or opacity of the displayed points or lines. For clustering, this feature can be used in order to display aggregates of the multivariate data results from the clustering process. Further information can be found in chapter [3.2](#).

7. *can discriminate points/lines*

Being to able to identify individual data items within a crowded display is the goal of this criterium. The study states the capability of clustering for detecting outliers as well as creating groups of points in order to satisfy this criterion. On the other hand, this classification appears unclear, as the grouping process of clustering aggregates individual information by definition and only makes it accessible by request or localization.

8. *can see overlap density*

This helps to gauge the amount of overplotting, see where higher density regions are and understand the distribution of data underlying the visualization. Clustering can show the amount of items within clusters by using visual indicators as point size, opacity and color.

CHAPTER 3

Results

In the previous chapter, foundations of visualization, visual variables and data exploration techniques, as well as the concept of clutter reduction have been introduced. As the main part of the presented study, existing visualization techniques for representing clustered data on maps will be discussed. In a first step, visualization concepts on the map level will be investigated. Later, approaches for visualizing individual clusters are added. An evaluation summarizes the discussed technologies.

3.1 Map visualization types for clustering

There exists a variety of map types, serving different purposes like standard *geographic maps*, *cartograms*, *geologic maps*, *linguistic maps* or *weather maps*. Some of these use distortion, for example the cartogram can be used to map the area of each country to the size of population. In this section, we try to identify those map types which are appropriate for visualizing clustered data [NÖ6, Wik13a]:

- **Geographic map with markers.** The default way of representing data is a standard 2-dimensional map with markers on top of it. Each marker represents a data point, or in the case of clustering, a cluster.

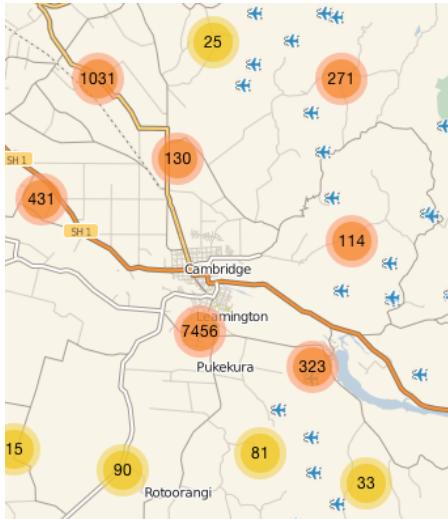


Figure 3.1: Leaflet map

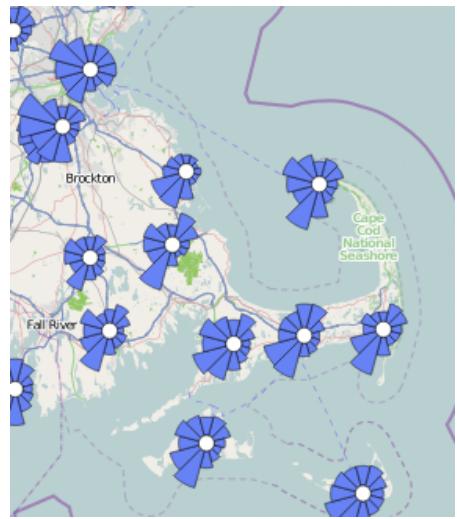


Figure 3.2: Wind history map

- Figure 3.1 depicts an example¹ of the Leaflet.markercluster library. Clustered results are displayed using markers of the same size. The amount of items within a cluster is indicated using text within the markers and by using a “hot-to-cold” color ramp [Bou96].
- Figure 3.2 shows a similar example, in this case a Wind history map² with markers for every wind measurement point. An advanced visualization technique is used for visualizing the amount of wind per cardinal direction as *polar area diagram*.

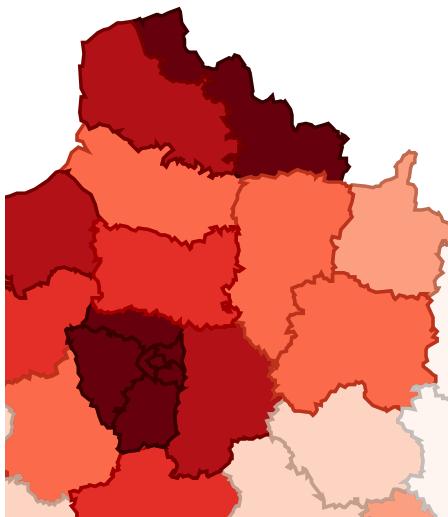
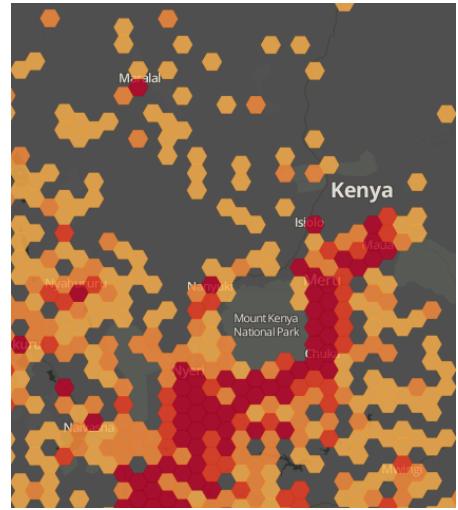
These two variations of standard maps shall indicate the potential of using advanced visualization techniques for displaying cluster items. Further ways for cluster visualization will be discussed in chapter 3.2.

• Geographic Heat map

Heat maps use colored, two-dimensional areas to express the value of each data entity on the map. *Choropleth maps* are the most common heat maps, which are often used for analysis of geographic and statistical data.

¹Leaflet.clustermarker example map from <http://leaflet.github.io/Leaflet.markercluster/example/marker-clustering-realworld.10000.html>

²Wind history map <http://windhistory.com/map.html>

Figure 3.3: Choropleth map³Figure 3.4: Heat map⁴

- Figure 3.3 visualizes an example choropleth map that shows population data for each of the departments of metropolean France. Color coding is used to indicate densely populated regions with heavier red tones.
- Figure 3.4 presents another example that uses *binning* for creating a hexagonal tessellation of the surface in order to visualize clustered results.

A problem with heat maps is that they require a non-overlapping tessellation of the surface to provide the areas for visualization. As the binning example indicates, such a tessellation can be done programmatically. Heat maps therefore can also be used for visualizing arbitrary clustered data without a need to calculate the exact boundaries of clusters. Another variation of heat maps is the *prism map*, which adds extrusion of areas as a third dimension [Lad12, Del10]. A publication on the evaluation of color schemes in choropleth maps can be found in [BMPH97].

• Dot grid maps

Dot grid maps are based on the suggestion by Jaques Bertin [Ber67, Ber83] to use graduated sizes in a regular pattern as an alternative to chloropeth maps. The advantage is that the map creator doesn't have to choose between quantity or density of a distribution value, because the dot grid map shows both at the same time. The user can understand the data distribution on a finer level of granularity, as opposed to where the chloropeth map usually creates larger areas of aggregated information [Ais12].

³Choropleth map example from Kartograph <http://kartograph.org/showcase/choropleth/>

⁴Heat map that uses binning <http://mapbox.com/blog/binning-alternative-point-maps/>

Figure 3.5 is an alternative version of the France map from figure 3.3, visualized as a dot grid map.

- **Voronoi map**

The Voronoi tessellation is a space partitioning technique. From a set of points it produces a Voronoi polygon for each point, such that the area covered is closest to that point in comparison to all other points. Jean-Yves Delort describes a technique that uses Voronoi polygons for “Vizualizing Large Spatial Datasets in Interactive Maps” [Del10]. It uses a hierarchical clustering technique to choose a subset of points per zoom level for proper visualization. Still, the effectiveness of this approach is questionable, as the scalability analysis of the studies shows that the technique can efficiently be used for datasets of up to 1000 items.

Figure 3.6 shows an exemplary voronoi map that displays all U.S. airports as of 2008. Besides the shown visualization, for Voronoi maps apply the same visualization possibilities as for choropeth maps.

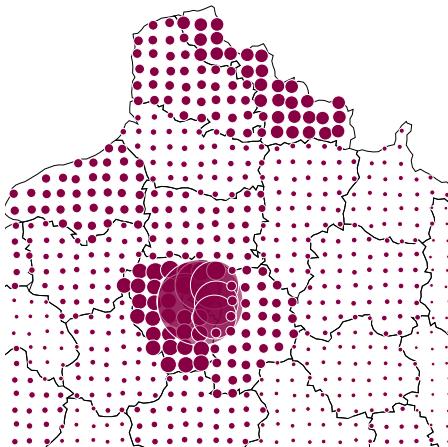


Figure 3.5: Dot Grid map⁵

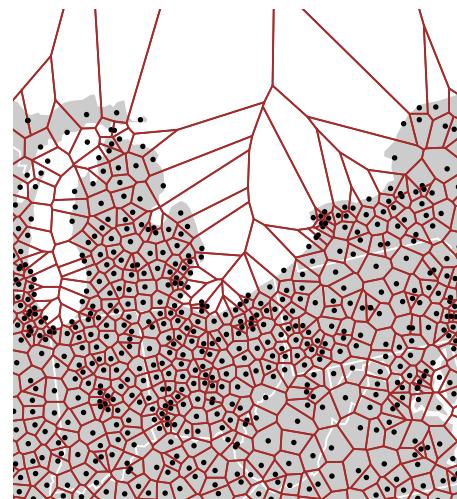


Figure 3.6: Voronoi map⁶

Self-organizing maps (SOM) can also be used to visualize clusters of data. But instead of displaying data on a geographic map, self-organizing maps create their own virtual space in order to represent information [NÖ6].

⁵Dot Grid map example from Kartograph <http://kartograph.org/showcase/dotgrid/>

⁶Voronoi map example <http://mbostock.github.io/d3/talk/20111116/airports-all.html>

3.2 Cluster visualization techniques for maps

The previous chapter has shown different kinds of map visualization techniques appropriate for displaying clustered data. In the end, each visualization will show (clustered) items on a map as visual objects with attributes like a particular shape or coloring. As clusters contain aggregated information, an important task is to find the right way for visualizing the cluster items themselves. From the examples provided so far, we have seen variations in size, shape and color which expose information on the cluster items being visualized on the map. In the following, multivariate data visualization techniques will be evaluated for visualizing cluster items on a map.

In chapter 2.2, a classification of visualization techniques by data type and interaction technique has been presented. Ke-Bing Zhang [Zha07] has written about “Visual Cluster Analysis in Data Mining”, where he lists an extensive list of multivariate data visualization techniques. Potentially, any such visualization technique can be used, but the frame of the map puts constraints in terms of space on the representation of individual items. *Iconic displays* are a simple way to visualize data, which also prevents clutter. *Dense Pixel* displays and *geometric visualizations* like charts can be used to encode more complex information.

- **Icon-based, Glyphs**

Matthew O Ward [War02] defines *glyphs* as “graphical entities that convey one or more data values via attributes such as shape, size, color, and position”. While the work of Otto Neurath on *ISOTYPE* [Jan09] (1930s) can be seen as fundamental for *pictorial statistics*, the best-known literature reference to glyphs is “Chernoff faces” [Che73]. As (e) figure 3.7 indicates, data is encoded into properties of the face icon, such as shape of nose, mouth, eyes. Other fundamental glyph-based techniques include *stick figures* [WSP89], *color icons* [Lev91], *Hyperbox* [AC91] and *shape coding* [Bed90]. Figure 3.7 extends this list by showing examples of glyphs that Ward collected for his taxonomy of glyphs placement strategies [War02].

Some glyph types have been created to identify clusters or similarities by plotting them side-by-side on a 2-dimensional plane, a technique which is referred to as *mosaic-based rendering*. *Stick figures* and *mosaic metaphors* are examples in that field [WSP89, NSS05]. On the other hand, reducing visual clutter as explained in chapter 2.3 also matters for glyphs, especially when putting them on a map. A trade-off between information-richness vs. simplicity and clarity has to be made. As Zhang writes, “with the amount of data increasing, the user hardly makes any sense of most properties of data intuitively, this is because the user cannot focus on the details of each icon when the data scale is very large” [Zha07]. We

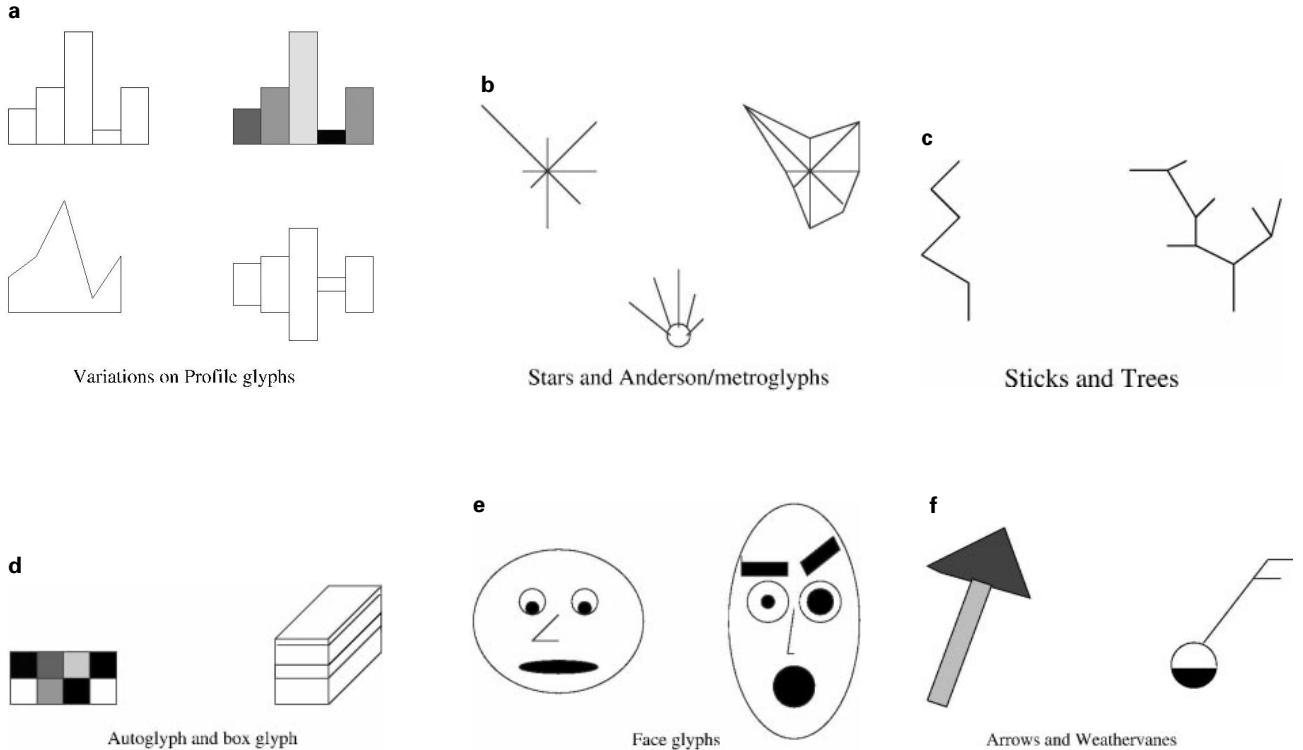


Figure 3.7: Examples of glyphs. Top row: (a) variations on profiles; (b) stars/metroglyphs; and (c) stick figures and trees. Bottom row: (d) autoglyphs and boxes; (e) faces; and (f) arrows and weathervanes. [War02].

can compare this to the map use cube presented in chapter 2: more complex glyph types seem to be better suited for scientific purposes which can be related to private uses, while simpler glyph types seem more appropriate for presenting data to a public audience.

Examples for uses of simple icons and glyph types for clustered data on maps can be found in JavaScript mapping libraries like Leaflet. These are usually based on a simple icon or geometric shape like a circle or marker and use color coding and size variations as indicators for underlying information. Further examples are scaled data values and scaled dots [Vil13], as well as the proportional symbol map [Fie08]. In contrast to those presented so far, figure 3.8 visualizes eight simple glyphs [EDG⁺08].

- **Pixel-oriented**

Pixel-oriented techniques display the most possible information at a time by map-

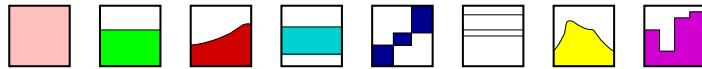


Figure 3.8: Eight different glyphs for aggregated edges (color shade, average, min-/max histogram, min/max range, min/max tribox, Tukey box, smooth histogram, step histogram) [EDG⁺08].

ping each attribute value of data to a single, colored pixel. Color mapping approaches such as *linear variation of brightness*, *maximum variation of hue* and *constant maximum saturation* are used to color pixels which are arranged within limited space. By providing an overview of large amounts of data, pixel-oriented display techniques are suitable for a variety of data mining tasks in combination with large databases [Zha07].

The first pixel-oriented technique was presented by Keim [KK94] as part of the *VisDB system*. Large amounts of multidimensional data are represented as *Spirals* and *Axes*. Figure 3.9 illustrates how a spiral would be constructed and figure 3.10 shows a rendered result of the axes technique. Further developments include the *Recursive Pattern Technique* [KK95] and the *Circle Segments Technique* [AKpK96]. Figure 3.11 visualizes such a circle which represents about 265,000 50-dimensional data items.

While no real-world examples have been identified during the research, using pixel-oriented techniques for visualizing complex clusters on a map seems possible. As the visualization relies on a large amounts of multidimensional data being present within clusters, the clustering algorithm would need to provide such required data. Performance implications also have to be considered, as potentially multiple clusters will be visualized on a map, sometimes even in real-time.

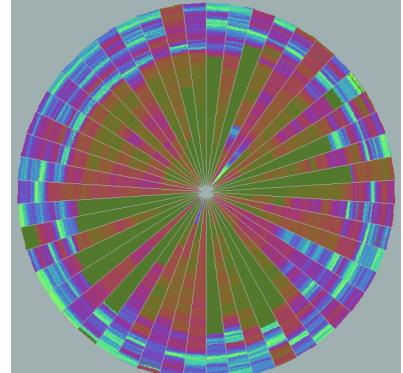
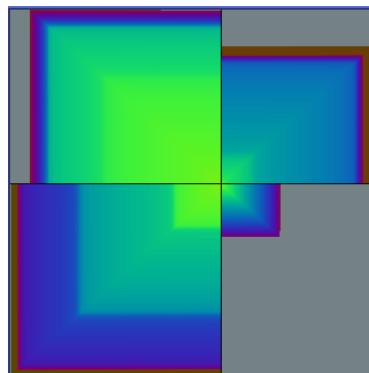
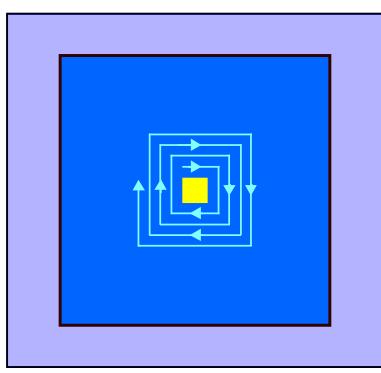


Figure 3.9: Spiral [KK94]

Figure 3.10: Axes [KK94]

Figure 3.11: Circle [AKpK96]

- **Geometric techniques & Diagrams**

Geometric techniques produce useful and insightful visualization by using geometric transformations and projections of the data. *Diagrams* are algorithmically drawn graphics that visualize data. This section lists a selection of geometric techniques for multivariate data presented by Ke-Bing Zhang [Zha07], diagram types described by Dieter Ladenhauf [Lad12] and related examples found in additional literature and on the web as stated in the individual references.

- *Line charts* visualize data as lines by connecting data points of the corresponding values. They are used to display trends over time. Figure 3.12 illustrates surface temperature anomalies from NASA's GISS⁷ as a *Sparkline* map [Sof08]. The Sparkline is a reduced line chart without axes and coordinates. It presents the general shape of variation in a simple and highly condensed way [Wik13b].
- *Bar charts* express data values by vertical or horizontal bars, in which the length of a bar indicates the data value. Figure 3.13 shows an example from UgandaWatch⁸, which displays economic indicators per region for the country Uganda.

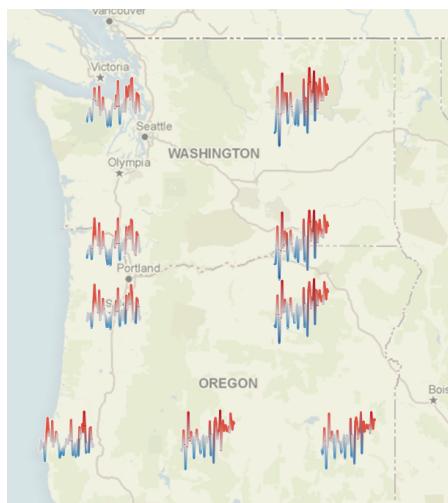


Figure 3.12: Sparkline map⁹

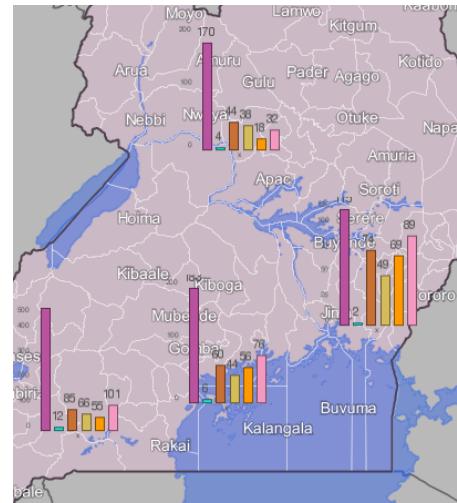


Figure 3.13: Bar chart map¹⁰

- *Pie charts* use a circle divided into sectors for expressing the proportional significance of data values. Variants of pie charts include *doughnut charts*,

⁷NASA Goddard Institute for Space Studies <http://www.giss.nasa.gov/>

⁸UgandaWatch: <http://www.ugandawatch.org/>

⁹Sparkline map example: <http://www.tableausoftware.com/about/blog/2008/08/sparklines-maps>

¹⁰UgandaWatch bar chart example: <http://www.ugandawatch.org/>

three-dimensional pie charts and *multi-level pie charts*. Also, the *polar area diagram* introduced in figure 3.2 is a special kind of pie chart and a further development of the *Bat's wing diagram* by Florence Nightingale [Sma98].

Figure 3.14 depicts a pie chart map example from the Kartograph¹¹ framework. It shows unemployment rates in Spain, providing an effective way to display ratios as opposed to a chloropeth map, where the user usually needs to consult a legend to understand the actual data values.

- *Container shapes* such as *bounding boxes* and *hulls* are an alternative to iconic displays as they can show the area covered by clusters [Del10]. The Leaflet.markercluster library visualizes the convex hull of a cluster to indicate the covered area on mouse-hover. Marco Cristani et al [CPCM08] use a hull-based technique for visualizing clusters from a geo-located image database on a map. As illustrated in figure 3.15, each cluster is represented by a hull that marks the boundaries of the area and contains a representative image of the clustered set of images.

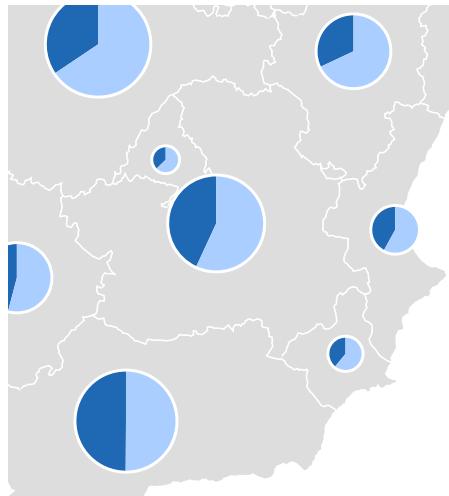


Figure 3.14: Pie chart map¹²

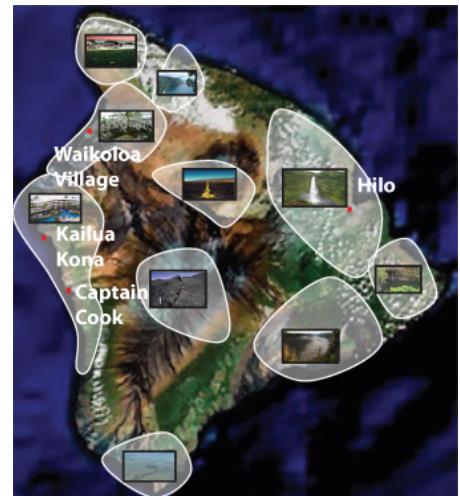


Figure 3.15: Convex hull map¹³

- Further chart types that can possibly be used for visualizing clusters on a map include *Area charts* and *Star plots* [Lad12] or more complex ones like *Parallel coordinates*, *Scatterplots*, *Treemaps* [Zha07] or the *Contingency Wheel++* [AAMG12]. *Bristle maps* are an interesting approach for visualizing spatio-temporal data on a map. Basically, histograms of the data are rendered onto linear map elements. If the data permits a mapping from aggregates to spatial line data, such a visualization technique could be of interest [KMM⁺13].

¹¹Kartograph <http://kartograph.org/>

¹²Pie chart map example from Kartograph: <http://kartograph.org/showcase/charts/>

¹³Convex hull map example from [CPCM08]

This concludes the investigative enumeration of cluster visualization techniques for maps. It is by no means a complete, but rather an exemplary listing that may provide a starting point when researching visual means of presenting clustered data on a map.

An interesting publication by Andrienko et al [AAH⁺12] presents a complex system for place-oriented analysis of movement data. This goes beyond the use case of simply visualizing clustered data on a slippy map, but it is a good show case for how effective the presentation of spatio-temporal data using a combination of techniques can be. *Time graphs*, *mosaic diagrams*, *space-time cubes* and *table lens displays* are used to create a powerful tool for inspecting movement data on maps.

3.3 Evaluation of visualization techniques for clusters on a map

This chapter summarizes the main visualization examples presented in the previous chapters. An evaluation of techniques for visualizing clusters on maps is created by proposing a set of key characteristics that are considered to be significant for this kind of visualization.

Evaluating information visualization techniques is a well-known problem. Undertaking an evaluation that is capable of “proving” the effectiveness is impossible in many situations as it would require too many tasks, data sets, implementations and users. Ellis and Dix state exploratory analysis as the most effective approach for evaluating visualization techniques [ED06, Del10]. In this sense, the following evaluation should primarily be treated as exploratory and as a help for understanding how cluster visualization on maps works, rather than a final, summative conclusion of which technique is superior than another.

Criteria. the following, custom criteria have been defined: *category*, *shows number of items within cluster (by shape size, by color or other)*, *shows cluster area*, *shows extra cluster info (extra cluster info complexity)*. Some criteria contain sub-criteria which are stated within parentheses.

In chapter 2, three taxonomies have been introduced: *visual variables*, a *classification of visual data exploration techniques* and a *clutter reduction taxonomy*. An attempt to classify the different visualization techniques for representing clusters on maps according to classes introduced by these taxonomies didn’t feel valid. Some criteria are too general while others are too specific to actually matter for visualizing clusters on maps, so the resulting data would not have much value. Similarly, a classification based on all *visual variables* would have become very complex. As the presented visualization examples describe general concepts, there are many possible variations that would in-

crease the data set even more. It is still helpful to rely on the aesthetic attributes for a better understanding of how each visualization is constructed. Especially the *shape*, *size* and *color* attributes are considered to have a strong effect on visualizing clusters on maps and are therefore included within the evaluation.

An explanation of each criterion follows:

- **category:** Determines the type of visualization technique being evaluated. Possible values are *type of map* (see chapter 3.1), a visualization *example*, as well as abbreviations for cluster visualization techniques (see chapter 3.2): *glyph*: Icon-based, Glyphs, *pixel*: Pixel-oriented techniques and *geom*: Geometric techniques & Diagrams.
- **shape:** Defines the type of shape being used for the visualization of clusters on the map. For example *circle* or *area*. Refer to the visual variable shape in chapter 2.1.
- **shows # of items within cluster:** If the visualization provides an indicator of the amount of items per clusters. This relates to the ‘can see overlap density’ criterion of the clutter reduction taxonomy, see 2.3. Two sub-criteria are used to differentiate between visual means of encoding the number of items within clusters:
 - **by shape size:** classifies visualization techniques that use the shape size for indicating the number of items within clusters.
 - **by color or other:** classifies visualization techniques that use color or other visual indicators to describe the number of items within a cluster.
- **shows cluster area:** Determines, if the visualization indicates the spatial area that is covered by the cluster or the items within a cluster.
- **shows extra cluster info:** Besides the two characteristics of number of items within a cluster and the cluster area, the technique might provide means of visualization additional information of clusters such as aggregates.
 - **extra cluster info complexity:** This sub-criterion expands of an intuitive notion of complexity that can be visualized as extra cluster info by the technique. *low* indicates a maximum of three dimensions. *medium* is used to describe up to 12 dimensions of additional data and *high* classifies cluster visualization techniques that go beyond this number of dimensions.

	category	shape	shows # of items within cluster	by shape size	by color or other	shows cluster area	shows extra cluster info	extra cluster info complexity
Geographic map	map type	~	~	~	~	~	~	~
Leaflet.markercluster	example	circle	x	-	x	~ (1)	-	-
Wind history	example	polar area diag.	-	-	-	-	x	low
Choropleth	map type	area	x	-	x	x	~	low (7)
Heatmap	map type	area (hexagon)	x	-	x	~ (2)	~	low (8)
Dot-grid map	map type	circle	x	x	~	~ (3)	~	low (8)
Voronoi	map type	area (convex)	~	~	~	~ (2)	~	low (7)
Face glyph	glyph	ellipsis	~ (4)	~ (4)	-	-	x	medium
Step histogram glyph	glyph	rectangle	~ (4)	~ (4)	-	-	x	low
Spiral	pixel	spiral	~ (4)	~ (4)	- (6)	-	x	high
Axes	pixel	rectangle	~ (4)	~ (4)	- (6)	-	x	high
Circle	pixel	circle	~ (4)	~ (4)	- (6)	-	x	high
Spark line	geom	line	~ (4)	~ (4)	- (6)	-	x	low
Bar chart	geom	rectangles	~ (4)	~ (4)	- (6)	-	x	low
Pie chart	geom	circle	x	x	- (6)	-	x	low
Hull	geom	area (hull)	~ (5)	-	~ (5)	x	x	low

Figure 3.16: Evaluation of visualization techniques for clusters on a map. Legend: ‘x’: yes, ‘~’: possibly, ‘-’: no. Numbers in parentheses reference additional notes within the accompanying text.

The results of the evaluation of visualization techniques for clusters on a map based on the stated criteria are illustrated in figure 3.16. Note that the classifications for each technique being evaluated primarily represent the according examples shown in the previous chapters. Where it seemed obvious, the optional possibility of fulfilling a criterion has been marked as such. As an extreme example, the geographic map as its general concept has been marked with the optional possibility of fulfilling each criterion. It is up to the actual implementation to satisfy them individually. While the Leaflet.markercluster example provides a visual indicator for the number of items within clusters, the wind history example doesn’t.

Some classifications contain a number referring to additional notes as presented in the following: (1) The Leaflet.markercluster example shows cluster on demand, based on user interaction. By mouse-hovering over an item, it will display the convex hull of the items being clustered. (2) In the case of the binned heat map example and the Voronoi map, cluster areas are approximated by the tessellation which is part of the clustering algorithm, see figure 3.4 and 3.6. (3) The dot-grid map doesn’t provide a mean of showing cluster areas by themselves, but the density of items still supports the notion of recognizing cluster areas, see figurefig:map-type-dotgrid. (4) For various visualization techniques, the amount of items within a cluster could be visualized by simply scaling the visual entity. (5) The hull example uses an area shape defined by the data, similarly to the choropleth map. Without a distortion technique, the shape therefore can’t be used

to indicate the number of items within a cluster. Still, a non-shape visual aspect like color could be used as an indicator. (6) In the case of pixel-oriented techniques and the provided chart examples, the color attribute will likely be used for showing extra cluster info instead of indicating the number of items within a cluster. Modifying the shape size can be used as an alternative in this cases. (7) The choroleth and voronoi map examples would rely on representing extra information within the defined area and therefore rely on the variation of visual variables related to color and texture. (8) The same restrictions as in the previous note apply, but for even smaller areas.

Map visualization types and cluster visualization techniques for maps have been introduced and were summarized within the given evaluation.

CHAPTER 4

Discussion

The previously presented evaluation classifies 7 different map visualizations that contain clusters (5 map types and 2 examples), as well as 9 cluster visualization techniques (2 glyphs, 3 pixel-based and 4 geometric / diagrams). Each visualization has been evaluated against a limited set criteria which are considered as important for showing clustered, multi-variate data on maps.

All visualizations, with the exception of the Wind history example, fulfill the primary criterion of **being able to indicate the number of items within clusters**. This makes sense, as in many cases *seeing overlap density* is an important feature of the visualization to indicate the amount of overplotting and tell the user an estimate of what to expect within the cluster. In the case of the Wind history example this is not necessary because all clusters are of the same size and visualize their inner data ratio as part of the polar area chart. Given the many classifications of “possibly” and considering the presented examples, it seems like fulfilling this criterion heavily depends on the needs of the use case. The fact, that all evaluated visualization techniques support showing cluster sizes, apparently correlates with their selection process. Naturally, those visualization approaches have been evaluated that seemed logical for visualizing clustered data.

On the contrary, few visualization techniques fulfill the criterion of **showing the cluster area**. The presented glyphs, pixel-based and diagram techniques do not have this feature as their shape is predefined. On the other hand, the Hull example indicates that a shape that indicates area can also be combined with an arbitrary shape inside.

While most cluster visualization techniques fail at indicating areas, they are perfectly suited for **showing extra cluster info**. Glyphs, geometric-techniques and diagrams appear suitable for visualizing cluster info of lower complexity while pixel-based techniques are geared towards multi-variate data of higher complexity.

Given the general discussion of the study results, a practical evaluation of Geocluster is added to the discussion.

4.1 Geocluster visualization

Geocluster¹ is a server-side clustering implementation for mapping with Drupal based on Geohash. Refer to the complete thesis on Geocluster for an explanation of related concepts as Drupal, the Drupal mapping stack, Bounding Box strategies etc. [Dab13]. Two visualizations for Geocluster have been implemented:

- **Geocluster default visualization**

A simple Geocluster visualization component has been built to support the display of clustered markers on interactive maps based on the output of the server-side clustering implementation. It extends the Bounding Box strategy of the Leaflet GeoJSON module in order to create numbered markers that visualize the cluster sizes. Clicking on a clustered marker will zoom into the map in order to explore the data on a more granular level. Figure 4.1 demonstrates an example screenshot of the Geocluster visualization. Compare this with an unclustered map, containing the same amount of items in figure 4.2.

Geocluster test mysql json map



Figure 4.1: Geocluster visualization

Geocluster test default json map

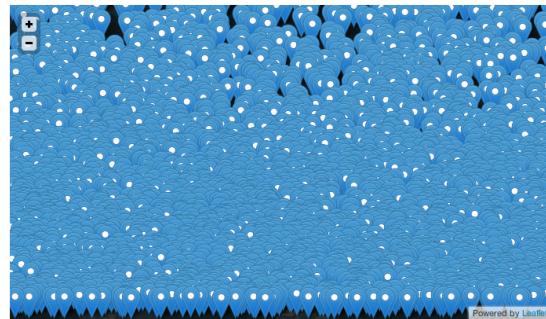


Figure 4.2: Unclustered Leaflet map

The visualization component of Geocluster is based on a GeoJSON-feed provided by the server-side clustering implementation. It uses the JavaScript Bounding Box strategy to dynamically update results for the current viewport after zooming or panning the map. The cluster visualization is very simple and based on a code snippet for numbered markers on github².

¹<http://drupal.org/project/geocluster>

²<https://gist.github.com/comp615/2288108>

- **GeoRecruiter**

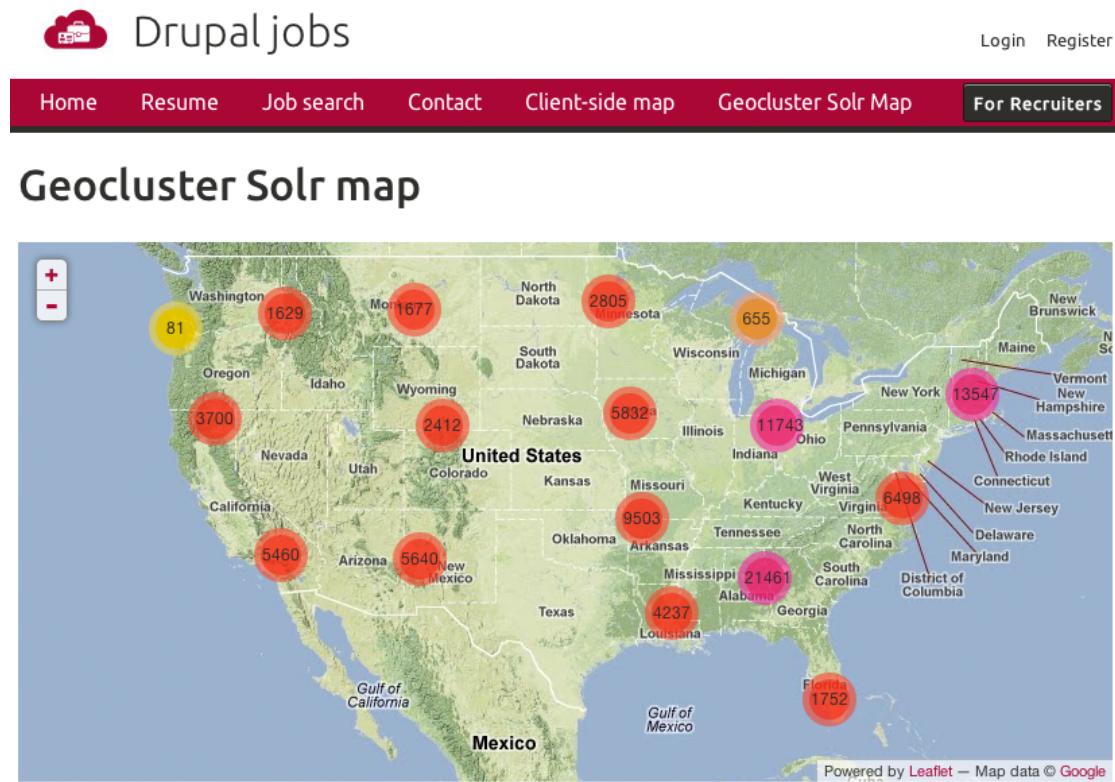


Figure 4.3: Screenshot of map that visualized job search results on a map using Solr-based clustering on a Drupaljobs test installation.

A practical use case for server-side geo clustering has been implemented for the Recruiter job board solution. It supports spatial search capabilities of the Recruiter distribution by visualizing a large amount of job offers on e-recruitment websites. The prototype being discussed has been developed, based on a copy of the Drupaljobs website³.

Recruiter is a Drupal distribution for building Drupal based e-recruitment platforms. Users can register either as recruiter and post job classifieds or they can register as applicants and fill out their resume. A faceted search helps users to find jobs and possible job candidates.⁴. *Drupaljobs* is provided as a show case for the Recruiter distribution. Its

³<http://drupaljobs.epiqo.com>

⁴<http://drupal.org/project/recruiter>

base features allow to create and search for job offers by companies as well as resumes of registered applicants on the e-recruitment platform.

A map visualizes the clustered job search results. In order to enhance the representation of clusters and to experiment with interaction, the CSS styles of the client-side clustering library Leaflet.markercluster have been adapted and extended with additional colors for large clusters. A visualization of a map within the Drupaljobs test installation is provided in figure 4.3.

4.2 Visual evaluation of Geocluster

Two visualizations have been provided: first, the Client-side Geocluster Visualization component as visualized in figure 4.1 and second, an alternative visualization similar to the Leaflet.markercluster library for the GeoRecruiter use case, see figure 4.3.

Both implementations fulfill most of the clutter reduction criteria from chapter 2.3:

1. *Overlap* is avoided by visualizing a non-overlapping clustering of points, as created by geohash-based clustering algorithm of Geocluster.
2. *Spatial information* is expressed as an aggregate of latitude and longitude values for each cluster, approximating its centroid. One problem with the current implementations is, that clusters aren't "stable". In some experiments, changes to the bounding box will cause a change of cluster assignment. Intuitively, this leads to confusion of the user and should be investigated upon further.
3. The Leaflet Bounding Box strategy implementation allows to *localize* the view by panning and zooming and therefore reduce the display to a specific region.
4. *Scalability* is provided by the underlying clustering algorithm, as evaluated in the publication on Geocluster [Dab13].
5. The configuration options of Geocluster allow to *adjust* the minimum distance between clusters. Additional configuration options are provided by the Views integration of the Geocluster module. Still, the configuration options could be expanded for example to control visual parameters of clusters [Dab13].
6. The criterion of *showing point/line attributes* is fulfilled only in a very limited way. Both implementations are currently restricted to displaying the number of items within a cluster as the only aggregate value. In order to support complex visualization techniques for multivariate data as discussed in chapter 3.2, the clustering

implementation needs to provide the required, aggregate values of the underlying data.

7. As explained in the discussion of the *discriminating points/lines* criterion, its definition seems unclear. Clustered items and individual points are visualized in a different way, which can be seen as a fulfillment. On the other hand, the visualization currently doesn't provide any means of inspecting clusters. Only a list of identifiers of the items within a cluster is provided. Based upon the identifiers, a popup could be used to visualize the items in a more detailed way.
8. With regards to *overlap density*, the number of items within clusters is indicated by both visualizations, but using different means. The Geocluster visualization doesn't make use of visual attributes like size or color, instead it creates a marker glyph that contains the number of items using textual representation. The GeoRecruiter prototype uses a color ramp that indicates low cluster densities from green and yellow to high densities indicated by tones of red and violet.

	category	shape	shows # of items within cluster	by shape size	by color or other	shows cluster area	shows extra cluster info	extra cluster info complexity
Geocluster default	example	marker	x	-	-	-	-	-
GeoRecruiter	example	circle	x	-	x	-	-	-

Figure 4.4: Evaluation of Geocluster visualization techniques for clusters on a map.
Legend: ‘x’: yes, ‘~’: possibly, ‘-’: no.

Next, the evaluation of visualization techniques for clusters on map presented in 3.3 is applied to the two Geocluster implementations as illustrated in figure 4.4. It reiterates some key aspects identified by the previous discussion of clutter reduction criteria.

Cluster sizes: Both visualizations show the number of items within a cluster, but only the GeoRecruiter example uses color and none of them encodes the size of a cluster into the shape size. As naturally, a cluster with more items can be visualized larger than smaller clusters, the algorithm could be improved for growing clusters by their size. The bigger size of a cluster would therefore reduce the distance to its neighbor clusters, potentially merging additional neighbors into it. Andrew Betts describes a similar approach under the term “*Grid based viral growth algorithm*” [Bet07].

Cluster areas: Neither implementation provides a visual indicator for showing cluster areas. Again, the server-side clustering implementation would need to provide this information. Instead, for performance reasons, only the number of items within clusters is provided as an aggregate value.

Extra cluster info: Similarly, no additional data about aggregates within clusters is exposed by the clustering implementation. This might suffice the use case of clustering points on maps for reducing clutter, but potentially hides a lot of information that could be displayed as part of the cluster visualization.

On a note on performance of the interaction, for low zoom levels, the roundtrip to the server for fetching a separate clustered result on every bounding box change can be an overhead. Christopher Calid proposes a way of “Progressively enhance server-side with client-side clustering”⁵. The intention is to switch from server-side clustering at higher zoom levels to client-side clustering for lower zoom levels.

4.3 Conclusions

A simple taxonomy for classification of cluster visualization techniques for maps has been presented and applied to a selection of techniques and examples. For the particular use case of web maps, visualizing *cluster sizes* was proposed as a primary goal for displaying clusters, followed by *showing the cluster area* and *showing extra cluster info*. The techniques have been enumerated and evaluated from two different view points: *map visualization types for clusters* and *cluster visualization techniques for maps*. In this way, a structured overview of existing approaches could be given.

The discussion of the practical implementation of Geocluster visualization for server-side clustering with maps illustrates how the same dataset can be visualized in different ways. Indicating cluster sizes by visual means like shape size and color seems to generate a more intuitive result than simply putting numbers of cluster sizes within markers. In comparison to advanced cluster visualization techniques, the Geocluster visualization is a very simple one that doesn't show cluster areas and extra cluster infos.

A brief overview has been given by enumerating examples from different research disciplines like geovisualization, cluster visualization and chart techniques. Complex cluster visualizations seem to be specific for scientific purposes of exploration and analysis of rather private audiences. Simpler cluster visualizations can be found in JavaScript mapping libraries that are used to present information to public audiences.

The study only touches briefly on further aspects like interaction, usability and animation. Also the impact of visual variables and the exact data types being presented offers room for additional investigation.

⁵<http://drupal.org/node/1914704>

List of Figures

2.1	The map use cube after MacEachren and Kraak [MK97] characterizing geo-visualization goals in a three-dimensional space by their level of interaction, their audience, and the addressed tasks. [NÖ6].	5
2.2	Aesthetic Attributes by Geometry [Wil05].	7
2.3	Classification of visual data exploration techniques based on [Kei01].	9
3.1	Leaflet map	13
3.2	Wind history map	13
3.3	Choropleth map ⁶	14
3.4	Heat map ⁷	14
3.5	Dot Grid map ⁸	15
3.6	Voronoi map ⁹	15
3.7	Examples of glyphs. Top row: (a) variations on profiles; (b) stars/metro-glyphs; and (c) stick figures and trees. Bottom row: (d) autoglyphs and boxes; (e) faces; and (f) arrows and weathervanes. [War02].	17
3.8	Eight different glyphs for aggregated edges (color shade, average, min/max histogram, min/max range, min/max tribox, Tukey box, smooth histogram, step histogram) [EDG ⁺ 08].	18
3.9	Spiral [KK94]	18
3.10	Axes [KK94]	18
3.11	Circle [AKpK96]	18
3.12	Sparkline map ¹⁰	19
3.13	Bar chart map ¹¹	19
3.14	Pie chart map ¹²	20
3.15	Convex hull map ¹³	20

3.16 Evaluation of visualization techniques for clusters on a map. Legend: ‘x’: yes, ‘~’: possibly, ‘-’: no. Numbers in parentheses reference additional notes within the accompanying text.	23
4.1 Geocluster visualization	26
4.2 Unclustered Leaflet map	26
4.3 Screenshot of map that visualized job search results on a map using Solr-based clustering on a Drupaljobs test installation.	27
4.4 Evaluation of Geocluster visualization techniques for clusters on a map. Legend: ‘x’: yes, ‘~’: possibly, ‘-’: no.	29

Bibliography

- [AAH⁺12] G. Andrienko, N. Andrienko, C. Hurter, S. Rinzivillo, and S. Wrobel. Scalable analysis of movement data for extracting and exploring significant places. *IEEE Trans Vis Comput Graph*, 2012.
- [AAMG12] Bilal Alsallakh, Wolfgang Aigner, Silvia Miksch, and Eduard Gröller. Reinventing the contingency wheel: Scalable visual analytics of large categorical data. *IEEE Transactions on Visualization and Computer Graphics (Proceedings of IEEE VAST 2012)*, 18(12):2849–2858, 12/2012 2012.
- [AC91] Bowen Alpern and Larry Carter. The hyperbox. In *Proceedings of the 2nd conference on Visualization '91, VIS '91*, pages 133–139, Los Alamitos, CA, USA, 1991. IEEE Computer Society Press.
- [Ais12] G. Aisch. Dot Grid Maps. <http://kartograph.org/showcase/dotgrid/>, 2012. [Online; accessed 26-April-2013].
- [AKpK96] Mihael Ankerst, Daniel A. Keim, and Hans peter Kriegel. Circle segments: A technique for visually exploring large multidimensional data sets. 1996.
- [Bed90] Jeff Beddow. Shape coding of multidimensional data on a microcomputer display. In *Proceedings of the 1st conference on Visualization '90, VIS '90*, pages 238–246, Los Alamitos, CA, USA, 1990. IEEE Computer Society Press.
- [Ber67] Jacques Bertin. *Sémiologie graphique*. Mouton [u.a.], Paris [u.a.], 1967.
- [Ber83] Jacques Bertin. *Semiology of graphics*. University of Wisconsin Press, 1983.

- [Bet07] A. Betts. Clustering points on a Google Map. <http://web.archive.org/web/20071121140547/http://trib.tv/tech/clustering-points-on-a-google-map/>, 2007. [Online; accessed 26-April-2013].
- [BMPH97] C.A. Brewer, Alan M. MacEachren, Linda W. Pickle, and Douglas Herrmann. Mapping mortality: Evaluating color schemes for choropleth maps. *Annals of the Association of American Geographers Annals of the Association of American Geographers*, 87(3):411–438, 1997.
- [Bou96] P. Bourke. Colour ramping for data visualisation. http://web.archive.org/web/20120415055120/http://local.wasp.uwa.edu.au/~pbourke//texture_colour/colourramp/, 1996. [Online; accessed 26-April-2013].
- [Che73] H. Chernoff. The use of faces to represent points in k- dimensional space graphically. *Journal of the American Statistical Association*, 1973.
- [CPCM08] Marco Cristani, Alessandro Perina, Umberto Castellani, and Vittorio Murino. Content visualization and management of geo-located image databases. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '08, pages 2823–2828, New York, NY, USA, 2008. ACM.
- [Dab13] Josef Dabernig. Geocluster: Server-side clustering for mapping in drupal based on geohash. Master’s thesis, Vienna University of Technology, 2013. <http://dasjo.at/thesis>.
- [Del10] Jean-Yves Delort. Vizualizing large spatial datasets in interactive maps. In *Proceedings of the 2010 Second International Conference on Advanced Geographic Information Systems, Applications, and Services*, GEOPROCESSING ’10, pages 33–38, Washington, DC, USA, 2010. IEEE Computer Society.
- [ED06] Geoffrey Ellis and Alan Dix. An explorative analysis of user evaluation studies in information visualisation. In *Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization*, BELIV ’06, pages 1–7, New York, NY, USA, 2006. ACM.
- [ED07] G. Ellis and A. Dix. A taxonomy of clutter reduction for information visualisation. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6):1216–1223, 2007.

- [EDG⁺08] Niklas Elmqvist, Thanh-Nghi Do, Howard Goodell, Nathalie Henry, and Jean-Daniel Fekete. Zame: Interactive large-scale graph visualization. In *IEEE VGTC Pacific Visualization Symposium 2008, PacificVis 2008, Kyoto, Japan, March 5-7, 2008*, pages 215–222. IEEE, 2008.
- [EF10] N. Elmqvist and J. Fekete. Hierarchical aggregation for information visualization: Overview, techniques, and design guidelines. *Visualization and Computer Graphics, IEEE Transactions on*, 16(3):439–454, 2010.
- [Fie08] Robert Fiedler. Visualization & cognition. *Simon Fraser University*, 2008. http://www.sfu.ca/gis/geog_x55/web355/icons/11_lec_vweb.pdf.
- [Fis07] D. Fisher. Hotmap: Looking at geographic attention. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6):1184–1191, 2007.
- [GHKV09] Emden Gansner, Yifan Hu, Stephen Kobourov, and Chris Volinsky. Putting recommendations on the map: visualizing clusters and relations. In *Proceedings of the third ACM conference on Recommender systems, RecSys ’09*, pages 345–348, New York, NY, USA, 2009. ACM.
- [Har08] F.J. Harvey. *A primer of GIS: fundamental geographic and cartographic concepts*. Guilford Publications, Incorporated, 2008.
- [Jan09] Wim Jansen. Neurath, arntz and ISOTYPE: The legacy in art, design and statistics. *J Design Hist*, 22(3):227–242, September 2009.
- [Kei01] Daniel A. Keim. Visual exploration of large data sets. *Commun. ACM*, 44(8):38–44, August 2001.
- [KK94] D.A. Keim and H.-P. Kriegel. Visdb: database exploration using multi-dimensional visualization. *Computer Graphics and Applications, IEEE*, 14(5):40–49, 1994.
- [KKA95] D.A. Keim, H.-P. Kriegel, and M. Ankerst. Recursive pattern: a technique for visualizing very large amounts of data. In *Visualization, 1995. Visualization ’95. Proceedings., IEEE Conference on*, pages 279–286, 463, 1995.
- [KMM⁺13] Sung Ye Kim, Ross Maciejewski, Abish Malik, Yun Jang, David S. Ebert, and Tobias Isenberg. Bristle maps: A multivariate abstraction technique for geovisualization. *IEEE Transactions on Visualization and Computer Graphics*, 99(PrePrints):1, 2013.

- [Lad12] Dieter Ladenhauf. Liquid diagrams – a suite of visual information gadgets. Master’s thesis, Graz University of Technology, Austria, January 2012. <http://www.iicm.tugraz.at/thesis/dladenhauf.pdf>.
- [LBWR94] Gerald L. Lohse, Kevin Biolsi, Neff Walker, and Henry H. Rueter. A classification of visual representations. *Commun. ACM*, 37(12):36–49, December 1994.
- [LCC11] Che-An Lu, Chin-Hui Chen, and Pu-Jen Cheng. Clustering and visualizing geographic data using geo-tree. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2011 IEEE/WIC/ACM International Conference on*, volume 1, pages 479–482, 2011.
- [Lev91] H Levkowitz. Color icons: Merging color and texture perception for integrated visualization of multiple parameters. In *Proc. Visualization ’91*, 1991.
- [Mac95] Alan M. MacEachren. *How maps work : representation, visualization, and design*. Guilford Press New York, 1995.
- [MG90] A MacEachren and J Ganter. A pattern identification approach to cartographic visualization. *Cartographica: The International Journal for Geographic Information and Geovisualization, Volume 27, Number, (2)*, 1990.
- [MK97] Alan M. MacEachren and Menno-Jan Kraak. Exploratory cartographic visualization: advancing the agenda. *Comput. Geosci.*, 23(4):335–343, May 1997.
- [MK01] A. M. MacEachren and M. J. Kraak. Research challenges in geovisualization. *Cartography and Geographic Information Science*, 28(1):3–12, 2001.
- [NÖ6] Martin Nöllenburg. Geographic visualization. In Andreas Kerren, Achim Ebert, and Jörg Meyer, editors, *Human-Centred Visualization Environments, GI-Dagstuhl Research Seminar, Dagstuhl Castle, Germany, March 5-8, 2006, Revised Lectures*, volume 4417 of *Lecture Notes in Computer Science*, pages 257–294. Springer, 2006.
- [NSS05] Thomas Nocke, Stefan Schlechtweg, and Heidrun Schumann. Icon-based visualization using mosaic metaphors. In *Proceedings of the Ninth International Conference on Information Visualisation, IV ’05*, pages 103–109, Washington, DC, USA, 2005. IEEE Computer Society.

- [PdOMA06] R. Pinho, M.C.F. de Oliveira, R. Minghim, and M.G. Andrade. Voromap: A voronoi-based tool for visual exploration of multi-dimensional data. In *Information Visualization, 2006. IV 2006. Tenth International Conference on*, pages 39–44, 2006.
- [PN00] Richard J. Phillips and Liza Noyes. An investigation of visual clutter in the topographic base of a geological map. *Cartographic Journal, The*, 19(2):122–132, 1982-12-01T00:00:00.
- [Shn96] B. Shneiderman. The eyes have it: a task by data type taxonomy for information visualizations. In *Visual Languages, 1996. Proceedings., IEEE Symposium on*, pages 336–343, 1996.
- [Sma98] Hugh Small. Florence Nightingale’s Statistical Diagrams. In *Stats and Lamps Research Conference*. Florence Nightingale Museum, March 1998.
- [Sof08] Tableau Software. Sparklines on Maps. <http://www.tableausoftware.com/about/blog/2008/08/sparklines-maps>, 2008. [Online; accessed 26-April-2013].
- [Vil13] Ian Villeda. Scaled Data Value Design in TileMill. <http://mapbox.com/blog/scaled-data-value-design-in-tilemill/>, 2013. [Online; accessed 26-April-2013].
- [War02] Matthew O. Ward. A taxonomy of glyph placement strategies for multidimensional data visualization. *Information Visualization*, 1(3/4):194–210, December 2002.
- [Wik13a] Wikipedia. Category:Map types — Wikipedia, the free encyclopedia. http://en.wikipedia.org/wiki/Category:Map_types, 2013. [Online; accessed 26-April-2013].
- [Wik13b] Wikipedia. Sparkline — Wikipedia, the free encyclopedia. <http://en.wikipedia.org/wiki/Sparkline>, 2013. [Online; accessed 26-April-2013].
- [Wil05] Leland Wilkinson. *The Grammar of Graphics (Statistics and Computing)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.
- [WSP89] M.G. Williams, S. Smith, and G. Pecelli. Experimentally driven visual language design: texture perception experiments for iconographic displays. In *Visual Languages, 1989., IEEE Workshop on*, pages 62–67, 1989.

- [Zha07] Ke-Bing Zhang. Visual cluster analysis in data mining. Master's thesis, Macquarie University, Australia, October 2007. http://comp.mq.edu.au/hdr/current/kebing_thesis.pdf.