

The Marker Cluster —

A critical analysis and a new approach to a common web-based cartographic interface pattern

Meier, Sebastian

meier@fh-potsdam.de

Interaction Design Lab,

University of Applied Sciences Potsdam,

Germany

ABSTRACT

The growing amount of gathered, stored and available data is creating a need for useful mass-data visualizations in many domains. The mapping of large spatial data sets is not only of interest for experts anymore, but, with regard to the latest advances in web cartography, also moves into the domain of public cartographic applications. One interactive web-based cartographic interface design pattern that helps with visualizing and interacting with large, high density data sets is the marker cluster; a functionality already in use in many web-based products and solutions. In this article, we will present our ongoing research on the problem of "too many markers". We will present an empirical evaluation and comparison of marker cluster techniques and similar approaches, including heatmaps and tiled heatmaps. We conclude with a first concept for overcoming some of the obstacles that we were able to identify in our study and thereby introduce a new direction for further research.

KEYWORDS

HCI, Usability, Web, Maps, Marker, Geovisualization, Evaluation.

INTRODUCTION

The overall growing amount of data is a driving force behind new fields of research in mass data visualization. One of many examples for recent research conducted in this area is the field of network visualizations, like e.g. the massive data plots by Hochman and Manovich (Hochman & Manovich, 2013). One might argue that mass data visualizations are not new. What might be new are the web-based or mobile technologies and communication channels that are available for developing visualizations on the production side as well as for the audiences on the reception side. The field of visualization is moving away from being an expert-only field towards being a field that is relevant for a broad audience. Thus, data and its representations are being made accessible and usable for scholars from many different fields as well as for a broad public audience.

The same applies for data visualization in cartography. Experts in the field have been working with mass data and geovisualization techniques since the introduction of the first Geo Information Systems (GIS) in the late 1960s (Coppock & Rhind, 2001). But especially with the advances in web cartography and the growing popularity of modern web mapping applications

like Google maps (Google Inc., 2014) or Bing (Microsoft Corporation, 2014), we observe that the public and academic fields are increasingly involved in the usage of modern web-based cartographic applications for private, professional and academic purposes. At the same time, we experience a growing number of services that collect or create large amounts of spatial data, for example services like Yelp (Yelp, 2014) or Foursquare (Foursquare, 2014). This combination of trends nourishes the need for new spatial mass data visualization methods for web-based cartographic applications.

In this paper, we analyze the marker cluster; a visualization method that is already being used by many web-based applications that integrate visualizations of large spatial datasets. From a scientific point of view, most of the research on marker clusters focuses on the more technical elements, like e.g. the algorithms generating the clusters (Bär & Hurni, 2011; Delort, 2010a; Kefaloukos, Vaz Salles, & Zachariassen, 2012; Stefanakis, 2005). In contrast to this related research, we investigate the functionality of the marker cluster method and take the human perception into account. Abstract representations of spatial data each require interpretation of the respective cartographic communication model. That is why also the marker cluster, just like every other layer of abstraction, will conceivably introduce errors in the user's interpretation. In order to better understand the user's perception of the marker cluster, we have conducted a series of experiments. They were aimed at testing the marker cluster's *performance and precision*, but also the accuracy in the users' interpretation, in other words their level of *understanding* the marker cluster as an abstract representation.

With our approach, we follow a trend towards more functional cartographic representations, away from purely design and artistic map creation (MacEachren, 2004), that has emerged over the last few decades. One of the earliest and most cited publications that have informed this trend towards more objective guidelines for cartographic representations is Arthur H. Robinson's 1952 book *The Look of Maps*, based on his dissertation, and his book *Elements of Cartography* (Robinson, 1952; Robinson et al, 1984). Since then, we have seen a whole body of research evolve that has been directed towards the analysis of perception of cartographic representations. Howard's (Howard, 1980) model for analyzing cartographic symbolization separates three distinct levels of cartographic representation: lexical, functional and cognitive. With our series of experiments, we mostly focussed on the latter: the individual perception of cartographic representation. Nonetheless, we also gained insights into the two more systemic aspects of Howard's model through the qualitative part of our analysis.

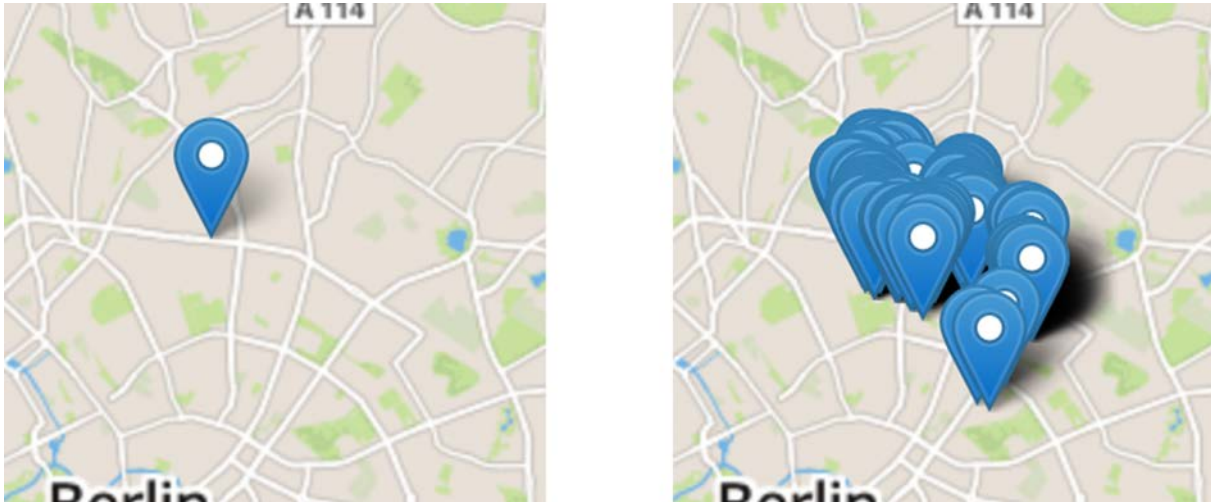


Figure 1. Left: 1 Marker, Right: 200 Markers.

PART 1: ANALYSIS

Challenge

In regards to small datasets, the conventional marker is still a sufficient interface solution. With an increasing number of data points, however, maps become cluttered and thereby unusable (Fig. 1). This is the starting point for our research. In the field of cartography, a large research corpus already exists when it comes to two-dimensional static maps and dense data set representation. Research has been conducted with focus on e.g. label placement (Christensen, Marks, & Shieber, 1995; Marks & Shieber, 1991), dot placement (Hey, 2011) or – when it comes to clustered data – choropleth maps. All of this shows us how to work with visual features in "dense" spatial data environments.

The problem that arises when we try to translate this knowledge into the field of web cartography is the factor of zoomability. While the old methods concentrated on displaying information at one specific zoom-factor, modern web mapping applications allow the user to zoom in and out and thereby reach different levels of detail. As framed in Ben Shneiderman's visualization mantra "overview first, zoom and filter, then details-on-demand" (Shneiderman, 1996), those web applications allow users to start at a low zoom-level and get an *overview* by seeing a large area of a big spatial data-set. In a second step, they can then *zoom in* and thereby *filter* the amount of data being displayed. In some cases, users can even zoom in until they are able to identify individual data-items and access their *detailed attributes on demand*.

The challenge with those techniques is to *overcome the cluttered marker concentration in low zoom levels but still provide the user a detailed overview*; or, as Woodruff et al. framed it, to provide "a constant information density" (Woodruff, Landay, & Stonebraker, 1998).

Existing Methods

Over the last few years, a couple of methods have been developed to overcome this problem of "too many markers" in web-based geovisualization applications. The most prominent methods are marker clusters, other solutions include heatmaps or tiled heatmap methods and choropleth based clustering approaches. In the following section we give a brief overview of such techniques:



Figure 2. Left to right: 1 Marker Cluster, 2. Heatmap, 3. Tiled Heatmap.

1. Marker Cluster

The marker cluster uses a grouping approach: if the density of markers is too high in a specified area, the markers are grouped and replaced by a cluster object (Fig. 2.1) (Delort, 2010b; Leaflet, 2014; Mahe & Broadfoot, 2010).

As a result of the clustering, it is required to introduce a new visual entity to the map that differentiates the cluster object from the single markers. In addition to simply differentiating the two entities, some of the existing solutions let the reader of the map know how many markers are merged into each cluster-object. The simplest way of doing this is by adding a text label with a number, indicating how many markers were merged. Sometimes, those solutions are combined with a "mouseover" interaction that allows the user to see the extent of each cluster by hovering over the cluster object with the pointer.

2. Heatmaps

The heatmap approach is a way of calculating the density of spatial data points per region and applying a colour from a predefined colour range to this region. Even though we will be discussing the influence of colour in this article, we will not go into detail in regards to the common palettes for web-based heatmaps, since this problem has already been thoroughly discussed in the visualization and cartographic community (Borland & Taylor, 2007; Brewer, 2013; Harrower & Brewer, 2003; Light & Bartlein, 2004) (Fig. 2.2).

3. Tiled Heatmaps

On a very rudimentary level, the tiled heatmap and the heatmap follow a similar approach. The tiled heatmap is also calculating the data points per region and applying a corresponding colour to the value, but, in contrast to the heatmap, the tiled heatmap has a lower resolution (Fig. 2.3).

4. Choropleth Maps

Similar to heatmaps and tiled heatmaps, choropleth maps also calculate data points per region (Fig. 2.4). In contrast to heatmaps and tiled heatmaps that use a raster that is independent from the map, the choropleth map is using shapes that are predefined. Those predefined shapes can for example be of political origin and depict national borders. Due to the zoomability of choropleth maps, layers of data with more detailed shapes for higher zoom levels are required. This data is not available in most cases, but crucial in the process of visualizing. We included the choropleth maps in this list of techniques for the sake of completeness and since it is a valid solution for clustering spatial data points in some cases. Nonetheless, we have excluded this type of

visualization from our further research due to the fact that additional data regarding the predefined shapes is required and thus results in a limited applicability of this technique.

Related Works

As briefly mentioned in the introduction, the existing research regarding marker clusters is mostly looking at applications (Komninos, Besharat, Ferreira, & Garofalakis, 2013) or at the algorithmic backbone of the clusters; that means the algorithms that decide how the clustering is computed (Bär & Hurni, 2011; Delort, 2010a; Kefaloukos et al., 2012; Stefanakis, 2005). In contrast to this existing computer-centered research, the human-centered hci research on marker clusters is still not very profound. This is why our research is focusing on gaining more insights into the usability aspects, and thereby the perceptual-cognitive tasks that those visualization methods bring with them, both of which were framed by Chen as 2 of the 10 unsolved information visualization problems (Chaomei Chen, 2005).

Research Question

With the clustering techniques listed above, we are able to create visualizations of constant information density at every zoom level. But when integrating those techniques in complex cartographic interfaces, new human computer interaction (HCI) problems arise. In the following, we will compare the clustering and visualization techniques in an empirical evaluation in regards to *performance and precision*, but also in regards to *understanding*. With this analysis we are contributing to the ongoing discourse on the visualization of big spatial data sets and thereby support academics and practitioners in choosing best practice solutions. Furthermore, we are trying to push our research towards better understanding cognitive tasks in spatial data visualization.

Experiment

In the following section, we describe two experiments that were conducted on the basis of the first three techniques (marker cluster, heatmaps, and tiled heatmaps) described above. The first experiment is looking at the performance of the subject group in terms of time and precision, while the second experiment is looking at the way people read and understand the visualizations. We used a real-world data set for the experiment, containing data on restaurants in the inner city of Berlin, Germany. The dataset was holding 15.329 positions in the urban area of Berlin. The experiments were conducted through amazon's mechanical turk (AMT) (Amazon Inc., 2014). AMT is an online platform that allows so-called workers to perform small tasks and in return be paid by the so-called requesters. Completing academic surveys by means of the mechanical turk system is still a very new way of conducting surveys, but has already generated good results and positive feedback from a wide range of research fields (Buhrmester, Kwang, & Gosling, 2011; Paolacci, Chandler, & Ipeirotis, 2010). As long as several requirements are met, like e.g. keeping tasks short, easy to understand, and including verifiable questions, as for example Kittur et al. suggested, mechanical turk can generate valid results (Kittur, Chi, & Suh, 2008). Joel Ross et al. presented an overview of the demographic development of the mechanical turk participants (Ross, Irani, Silberman, Zaldivar, & Tomlinson, 2010), which shows that workers come mostly from the US and India and that there is a slight bias towards female workers. The demographic data we collected showed that the participants we reached were more men (60%) and more Indian participants than workers from other countries (50%). The age of our participants ranged from 20 to 40 years with a peak in the mid 20s.

With our experiment we did not only want to test the existing visualizations against each other, but also wanted to go a step further and test how adding visual variables to the cluster-objects would influence the participants' performance of comparing the cluster-objects with each other. Visual variables most prominently introduced by Bertin in 1983 (Bertin, 1983; Bertin, Dodge, Kitchin, & Perkins, 2011) help the reader to not only differentiate between two objects, but also to compare and rank objects for example by size or colour (Carpendale, 2008), which is one of the most fundamental features of visualizing spatial data points (locate, read, classify, group and compare) (Heidmann, 2013).

Experiment I: Performance

A. Experiment Design:

In our first experiment, we compared the performance of marker clusters - extended by visual variables (size and colour) - with heatmaps and tiled heatmaps regarding the time needed by the subject group to accomplish the task and accuracy in regards to detecting values (Fig. 3).

Through the variation of variables we reached a total number of 11 visualizations that were each tested by 30 participants, adding up to 330 participants in total.

After answering a demographic questionnaire, the participants received a short explanation on how the presented geovisualization works, which means they were explained which visual variables indicate high and low values on the map. Before they were allowed to see the map, they were briefed on their first task: Click on the area with the highest/lowest values. Then the participants were directed to a page showing the visualization on top of a map. After the participant selected an area they were redirected to the next map. Every participant had to work on 6 maps.

On the last page the participants had to perform a semantic differential (Hassenzahl, Platz, Burmester, & Lehner, 2000) and a Subjective Mental Effort Questionnaire (SMEQ) which has proven to be a good tool for letting participants rate the difficulty of a task in a post-task questionnaire (Sauro & Dumas, 2009).

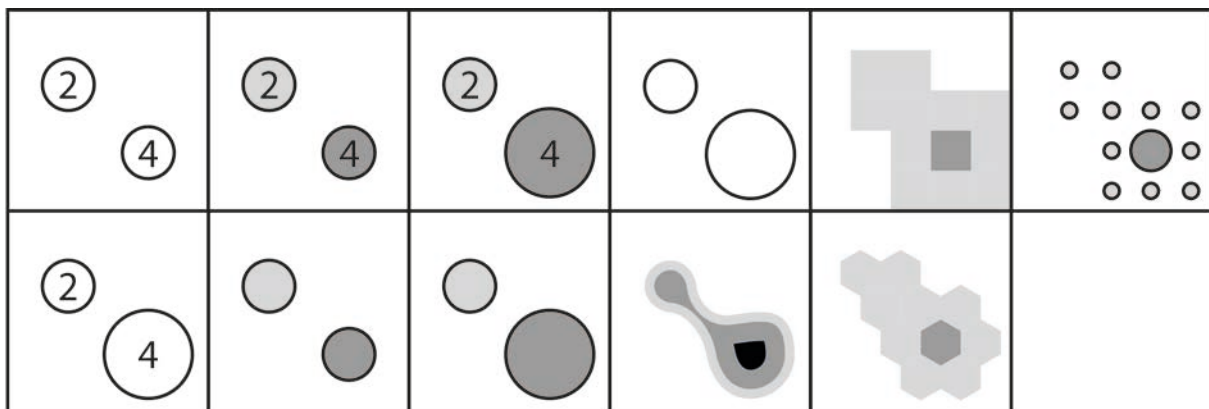


Figure 3. Visualization-types used in Experiment I.

B. Evaluation

We collected the time it took participants to select the maximum or minimum value of each map they were shown as well as the precision of their task result, that is how close their selection was to the actual maximum or minimum.

The distribution of time needed for completing all tasks was very narrow, the maximum differences between the averages were 4 seconds for the task of finding the maximum, and 5.5 seconds for the minimum. The bigger variation for the minimum goes along with the finding that the time for identifying the minimum took an average of 35% longer across all visualizations. But we could not find correlations between the visual variables or the visualization type and their influence on the time.

The finding that the determination of the minimum was more time-consuming than identifying the maximum was also visible in the precision data. The task of finding the maximum had a maximum offset of 0.8%, the task of finding the minimum had a maximum offset of 5%. Furthermore, we could identify 2 patterns from the precision results (Fig. 4).

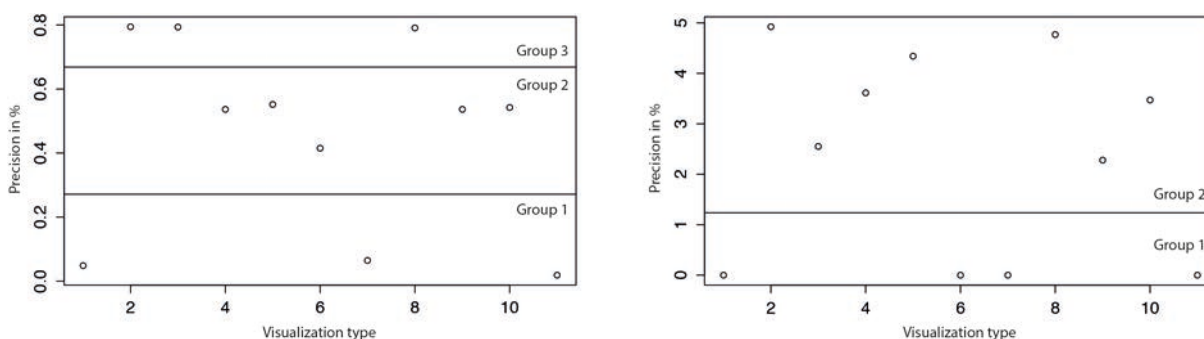


Figure 4. Precision offset in % for finding the maximum (left) and minimum (right).

- Heatmaps and tiled heatmaps performed better than marker cluster maps:* This finding applies to the tasks of finding the maximum and finding the minimum. The tiled heatmap visualizations performed best, followed by the heatmap, which did perform equally good for finding the minimum but slightly worse for finding the maximum.
- Size is a good visual variable for identifying the maximum:* size as visual variable has proven to be significant for identifying maximums (see Group 2 in Fig. 4). Regarding the task of finding the minimums, size has also been significant but not as outstanding as in the task of finding the maximum.
- Regarding the collected SMEQ and semantic differential values:* We received very similar values for each visualization and couldn't find patterns or correlations regarding the visual variables nor the variable time or precision.

C. Discussion

Our results show that size is a valid visual variable when using marker clusters. A previous study by Garlandini and Fabrikant who looked at visual variables in spatial data visualization in general (Garlandini & Fabrikant, 2009) also supports this finding. Furthermore, we saw that the heatmaps and tiled heatmaps performed very well, here we need further research through qualitative studies to find out why those visualizations performed better. Additionally, we need further research on the role of interaction and how interacting with the visualization might influence the performance and understanding of it.

Experiment II: Understanding

In *Experiment I* we gave each participant an introduction in order to explain how the visualization works. In the second experiment we were interested in finding out what participants would actually see in the visualizations and how they would make sense of them when they were given no introduction.

A. Experiment Design

For the second experiment we transferred the visualization types from *Experiment I* and extended them by the visual variable colour. While in *Experiment I*, we only used red as a visual variable due to its high contrast to the underlying map, we added blue and green as additional colours in our second experiment. Our hypothesis said that colour would influence the understanding of what is shown in the visualization.

The second experiment also started with a demographic questionnaire. After the first round of questions, the participants were shown one of 15 visualizations (Fig. 5). One group was shown just the visualization and one group was shown the visualization with a title that said “Restaurants in the city of Berlin” to test how contextualization of the visualization would effect the results. Thereafter, they were all asked to describe what they think the visualization on top of the map was indicating. The test was conducted on 239 participants, but due to errors we had to exclude 29 data sets and came down to 210 participants, 14 per visualization, 7 per group.

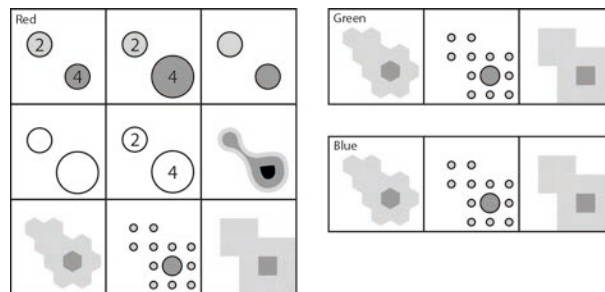


Figure 5. Visualization-types used in Experiment II.

B. Evaluation

In order to evaluate the qualitative data we received from our participants, we first screened all answers and then identified a set of variables or subjects:

- Location or Area*: 70% of the participants connected the visualizations to either location or area data.
- Density and Ranking*: 40% of the participants identified the visual variables as an unspecified ranking, 13% identified it as an indicator for density.
- Urban entity*: 70% of the participants identified some sort of urban entity, this ranked from hotels, restaurants, parks to more statistical data like crime rates.
- Restaurant*: 77% of the participants that were shown the title identified the visualized entities to be restaurants.
- Water or Nature*: 7% of the participants identified the represented entity to be some form of nature, like trees, parks or to be related to water.
- Cities*: 4% of the participants interpreted the markers as being labels for cities.
- Colour correlation*: 90% of the participants who identified water or nature did it in those cases when they were presented with green or blue maps.

From analysing the results of our clustering process by means of the variables listed above, we were able to identify some correlations and generate a series of insights:

- a) *Heatmaps and Tiled Heatmaps are more likely to be connected to area data than to point data*: 65% of the participants who received a heatmap or tiled heatmap interpreted it to be representing area data like e.g. population density or crime rates. Markers were instead put in relation to spatial location data, like the location of restaurants or hotels. The only exceptions were formed by the tiled heatmaps that used circles as visualizations; they were more likely to be interpreted as markers and not as heatmaps.
- b) *Density is an abstract concept; instead, people relate visual variable ranking to more common rankings*: only 13% of the participants identified the visualisations as indicating density, but 40% interpreted the visualization as being some sort of ranking. The most common association, especially from the participants who saw the title, was user rankings, e.g. from a social media platform.
- c) *Colour and label texts are used to connect the visualization to the mental models of the participants*: the 7% of participants who interpreted the visualisations as depicting nature and water were all participants who saw blue (water) or green (nature) visualizations. Beyond that, we noticed several participants who tried to use the label information to make sense of the visualization. The numbers, actually representing the number of clustered data points, were connected to distances, ratings or road numbers.

C. Discussion

Even though the results discussed in the evaluation above sound promising, we have to point out that only 13% of participants identified the visualization to be about density. However, we were able to conclude that heatmaps might not be the best visualization to show the density of spatial location data and should instead be used to depict areal data like population. Furthermore, the design of the visualization should take the data items' meaning or context into account, for example by visualizing data on water in blue. In the second part we will discuss an approach, which takes those visual relationships into account.

PART 2: BEYOND VISUAL VARIABLES — A NEW APPROACH

The study presented in Part 1 of this article showed that participants find it difficult to decode marker clusters or, more specific, their visual representation. At the same time, participants were trying to decode every single available visual information they received in order to make sense of the visualization. In regards to colour, for example: the colour blue was interpreted as information related to water, while green was interpreted as data related to nature. Building upon these findings, we will focus on proposing a new technique in the last chapter of this article that takes the characteristics of the data into account and extends the existing cluster-visualizations in order to help users to connect the visualization with the context of the underlying data, visual qualities or its meaning.



Figure 6. Abstraction from left to right: Pictorial-, Associative-, Geometric-representation.

Research Question

The problem that we were able to identify in our study originates in the fact that the real-world entity as the data's origin is separated from the visualization, e.g. through the design of the cluster symbol, by several layers of generalization and abstraction (Fig. 6); a process which is similar to what has been described by Robinson and Petchenik (Robinson & Petchenik, 1976) as the mimetic-arbitrary continuum. One of these steps of abstraction or steps from mimetic to arbitrary is the process of clustering and turning the singular data points into new visual cluster objects (symbols). Most techniques use abstract visual variables like colour and size to indicate differences in the number of data points being clustered in each cluster object (Fig. 7). In order to overcome this approach and the identified problems related to it we focused our research on the question *how we could develop and implement visual representations that are more closely related to the data points' meaning* and make it thereby easier for the reader to be decoded.

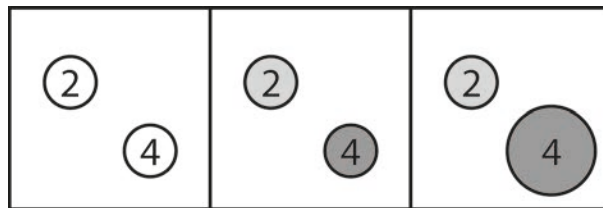


Figure 7. Visual variables from left to right: No visual variable, colour, colour + size.

The Generative Marker — A more iconic Representation

Before describing our actual approach and implementation, we want to briefly contextualize our research in the theoretical cartographic discourse on symbols and their semantic analysis. A large body of research covers the semantic analysis of cartographic elements from simple elements like lines and dots, down to specific symbols used in maps. MacEachren gives a good introduction to the topic of semiotics in the context of understanding map representation in his 2004 book “How Maps Work” (MacEachren, 2004). In regards to map symbols, which he describes in more semantic terms as *sign-vehicles*, he highlights the relationship of sign-vehicle, interpretant, and referent, and, building upon the work of Keates (Keates, 1982), emphasises that many map symbols have what he defines as an “conventional link with their referents”. In our case this would be, for example, the abstract map symbols in the marker cluster visualizations instead of an “iconic link”. A richer granularity on the same relationship is offered by Robinson et al. (Robinson, Sale, Morrison, Muehrcke, P. C., 1984), who defined three categories of map symbols: pictorial, associative and geometric.

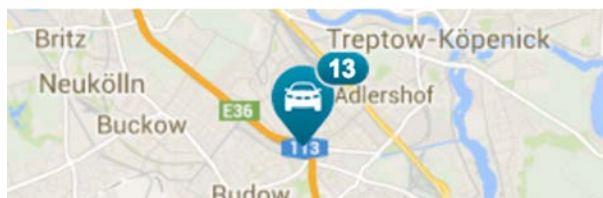


Figure 8. Cluster marker implementation by DriveNow.

Designing less arbitrary and abstract symbols as visual representations is a common task in cartographic map design. Tree-symbols (pictorial), as an example, are often representations for forests, or a cross for a church (associative). Research and guidelines for designing cartographic symbols, following the theory for cartographic communication, go back to e.g. Bertin's design principles (Bertin, 1983). Even though there is a little overlap between glyphs and static map symbols, most map symbols remain static, singular visual representations.

Our approach is trying to create more mimetic and pictorial or at least associative map symbols (as iconic links), that still allow us to communicate quantity; similar to the usage of glyphs which are abstract map symbols that are able to communicate quantitative data, in some cases even multivariate data.

The idea of combining the marker cluster with our new approach is inspired by existing marker cluster implementations like e.g. the visualization used by DriveNow (Fig. 8) ("DriveNow," 2014) which utilises conventional markers that do not only make use of abstract marker cluster symbols, but instead provide a more self explaining pictorial symbol in combination with a text-layer, displaying the number of clustered data points.

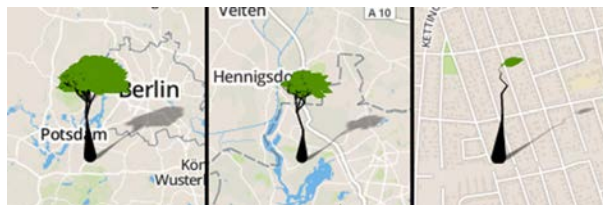


Figure 9. Generative Marker Cluster for number of trees.

Use Cases

For this article we have developed three use cases that aim at creating more pictorial or associative representations. The first example uses data from the city of Berlin that indicates where new trees are going to be planted in the forthcoming year. As a visual representation we chose a pictorial tree visualization. We individually generate custom tree visualizations for every cluster object. Thereby, we develop a new, generative solution in contrast to static map symbols. One leaf represents one new tree (Fig. 9). As the user zooms in, the cluster objects split up into several smaller trees with fewer leaves until each individual tree is represented by one leaf.



Figure 10. Generative Marker Cluster for number of restaurants.

The second use case employs the data set that was already implemented in the experiment described in part 1 of this article, depicting restaurants in the city of Berlin. We developed an associative, iconic symbol (knife and fork) with the visual variables size and colour, as well as a numeric indicator of the amount of clustered objects (Fig. 10). As the number of clustered objects grows, the marker's border-colour becomes more intense, while at the same time the size of the marker grows. With this extension we build upon a well-known iconic representation and add a visual, quantitative dimension to it.

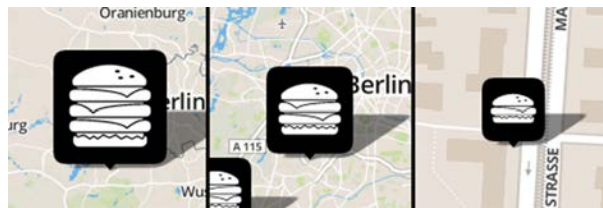


Figure 11. Generative Marker Cluster for number of burger bars.

The third use case employs data gathered from Foursquare (Foursquare, 2014) on restaurants serving burgers in the city of Berlin. Apart from the visual variable size, we used three different iconic representations to indicate an increase in number of clustered objects (Fig. 11). In this last example, similar to the second, we implemented a visual variable that represents quantity, but in addition we use a static symbols, slightly varying it depending on the quantity.

Technical Implementation

For this technique that we call "generative marker technique", we use the HTML5 canvas element to draw a marker image with javascript in real time, which is then turned into an image using the *dataToUrl* Method for better performance. This image can then be used as a marker image. The functionality can be extended to add a shadow on the image and thereby imitate the existing marker images that people are already used to, and, at the same time, achieve a higher contrast between map and marker.

Preliminary Study

A first preliminary study using the example of the tree visualization (Fig. 9) was conducted. It employed the same experimental design as the study presented in *Part 1* of this article. This preliminary study produced similar results in regards to efficiency as the markers that were extended by the visual variable size, but a better performance in regards to decoding. Proband identified the markers to be related to trees or nature.

CONCLUSION

Marker clusters as well as heatmaps and tiled heatmaps are viable solutions for overcoming the problem of displaying a large number of markers in a small area from a technical viewpoint. But none of the proposed visualization methods, as described in this article, has proven to be intuitively understood by our participants. From a design viewpoint, we saw that it is important to add visual variables to the marker cluster object in order to help users identify, classify and compare the visual entities on the map.

We introduced a new research direction by creating more “visually connected” data representations through “Generative Marker Clusters”, which showed good results in our preliminary study in terms of understanding. In regards to the proposed research direction, we do not only see a wide range of possible applications, but also the chance of contributing to the process of building more intuitive cluster or rather density visualizations.

As pointed out in the introduction, the research presented in this article is ongoing research. We see the two studies as a starting point for developing new methods, like the generative method presented above, and conducting further research on the cognitive processes involved in reading marker clusters, heatmaps and tiled heatmaps. Based on the findings presented in this article, we propose that more techniques should be developed that take the characteristics of the data into account and extend the existing visualizations in order to help users connect the visualization with the data beneath.

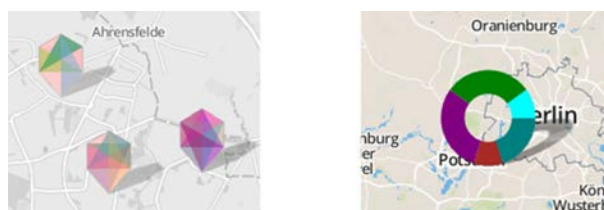


Figure 12. From left to right: Artistic- and Data-Vis-Marker Cluster objects.

FUTURE RESEARCH

In the implementation section we have shown three possible implementation scenarios. We are furthermore looking into using the method described in this article in two other ways. On the one hand we are interested to see if it is also possible to use the method to generate more complex visualizations, taking more data dimensions into account than just the number of clustered data points. Therefore, we have built a first prototype that generates custom doughnut charts for each set of clustered data items (Fig. 12 right).

On the other hand, we are also interested to see if the method can be used to foster more artistic and generative design approaches to the design of marker objects for marker clusters. A first prototype is generating custom markers based on the location data provided by the clustering technique (Fig. 12 left).

Besides experimenting with the actual visualization techniques and optimizing the code for better performance, we are also aiming at extending the study presented in the first part of this article in order to look deeper into the cognitive processes of decoding marker clusters and how this process can be optimized from a user centered design perspective. As mentioned in the closings of our empirical analysis, further experiments e.g. on how interactions with the visualization

changes the processes of understanding and decoding are needed for a more complete image. This will hopefully allow academics and practitioners to build more useable and engaging web mapping experiences.

ACKNOWLEDGEMENT

We would like to thank Ilmari Heikkinen for providing the free online tutorial on how to use image filters on canvas drawn graphics, which we used to create the shadows:
<http://www.html5rocks.com/en/tutorials/canvas/imagefilters/>

Our code for the generative markers is available under the MIT license as an open source project on GitHub: https://github.com/sebastian-meier/generative_marker

REFERENCES

- Amazon Inc. (2014). Amazon Mechanical Turk - Welcome. Retrieved March 10, 2014, from <https://www.mturk.com/mturk/welcome>
- Bär, H. R., & Hurni, L. (2011). Improved Density Estimation for the Visualisation of Literary Spaces. *Cartographic Journal, the*, 48(4), 309–316.
<http://doi.org/10.1179/1743277411Y.0000000022>
- Bertin, J. (1983). *Semiology of graphics*.
- Bertin, J., Dodge, M., Kitchin, R., & Perkins, C. (2011). General Theory, from *Semiology of Graphics. The Map Reader* (pp. 8–16). Chichester, UK: John Wiley & Sons, Ltd.
<http://doi.org/10.1002/9780470979587.ch2>
- Borland, D., & Taylor, M. R. (2007). Rainbow color map (still) considered harmful. *IEEE Computer Graphics and Applications*, 27(2), 14–17.
- Brewer, C. A. (2013). Spectral Schemes: Controversial Color Use on Maps. *Cartography and Geographic Information Science*, 24(4), 203–220.
<http://doi.org/10.1559/152304097782439231>
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science*, 6(1), 3–5. <http://doi.org/10.1177/1745691610393980>
- Carpendale, M. (2003). Considering visual variables as a basis for information visualisation. *Computer Science TR 2001-693*, 16.
- Chaomei Chen. (2005). Top 10 Unsolved Information Visualization Problems. *IEEE Computer Graphics and Applications*, 25(4), 12–16. <http://doi.org/10.1109/MCG.2005.91>
- Christensen, J., Marks, J., & Shieber, S. (1995). An empirical study of algorithms for point-feature label placement. *ACM Transactions on Graphics*, 14(3), 203–232.
<http://doi.org/10.1145/212332.212334>
- Coppock, J. T., & Rhind, D. W. (2001). The History of GIS. In D. J. Maguire, M. F. Goodchild, & D. W. Rhind (Eds.), (Vol. 1, pp. 21–43).
- Delort, J. Y. (2010a). Visualizing Large Spatial Datasets in Interactive Maps (pp. 33–38). Presented at the Advanced Geographic Information Systems, Applications, and Services (GEOPROCESSING), 2010 Second International Conference on, IEEE.
<http://doi.org/10.1109/GEOProcessing.2010.13>
- Delort, J.-Y. (2010b). Hierarchical cluster visualization in web mapping systems (p. 1241). Presented at the the 19th international conference, New York, New York, USA: ACM Press.
<http://doi.org/10.1145/1772690.1772892>
- DriveNow. (2014). DriveNow. Retrieved May 18, 2014, from <https://de.drive-now.com/>
- Foursquare. (2014, March 10). Berlin | Food, Nightlife, Entertainment. Retrieved March 10, 2014, from <https://foursquare.com/>
- Garlandini, S., & Fabrikant, S. I. (2009). Evaluating the Effectiveness and Efficiency of Visual Variables for Geographic Information Visualization. In *Lecture Notes in Computer Science* (Vol. 5756, pp. 195–211). Berlin, Heidelberg: Springer Berlin Heidelberg.
http://doi.org/10.1007/978-3-642-03832-7_12
- Google Inc. (2014, March 10). Google Maps. Retrieved March 10, 2014, from <https://maps.google.com/>
- Harrower, M., & Brewer, C. A. (2003). ColorBrewer.org: An Online Tool for Selecting Colour Schemes for Maps. *Cartographic Journal, the*, 40(1), 27–37.

- <http://doi.org/10.1179/000870403235002042>
- Hassenzahl, M., Platz, A., Burmester, M., & Lehner, K. (2000). Hedonic and ergonomic quality aspects determine a software's appeal (pp. 201–208). Presented at the the SIGCHI Conference, New York, New York, USA: ACM Press.
<http://doi.org/10.1145/332040.332432>
- Heidmann, F. (2013). Interaktive Karten und Geovisualisierungen. In W. Weber, M. Burmester, & R. Tille (Eds.), *Interaktive Infografiken* (pp. 39–69). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Hey, A. (2011). Automated Dot Mapping - How to generate Dot Clusters (pp. 1–6). Presented at the Proceedings of the 22nd International Cartographic
- Hochman, N., & Manovich, L. (2013). Zooming into an Instagram City: Reading the local through social media. *First Monday*, 18(7).
- Howard, V. A. (1980). Theory of representation: Three questions. In P. A. Kolers, M. E. Wrolstad, & H. Bouma (Eds.), (Vol. 2, pp. 501–515). Presented at the Visible Language.
- Keates, J. S. (1982). Understanding maps. Halsted Press, New York.
- Kefaloukos, P. K., Vaz Salles, M., & Zachariassen, M. (2012). TileHeat (1389023194 ed., pp. 349–358). Presented at the the 20th International Conference, New York, New York, USA: ACM Press. <http://doi.org/10.1145/2424321.2424366>
- Kittur, A., Chi, E. H., & Suh, B. (2008). Crowdsourcing user studies with Mechanical Turk (p. 453). Presented at the Proceeding of the twenty-sixth annual CHI conference, New York, New York, USA: ACM Press. <http://doi.org/10.1145/1357054.1357127>
- Komninos, A., Besharat, J., Ferreira, D., & Garofalakis, J. (2013). HotCity (pp. 1–10). Presented at the the 12th International Conference, New York, New York, USA: ACM Press.
<http://doi.org/10.1145/2541831.2543694>
- Leaflet. (2014). github.com/Leaflet/Leaflet.markercluster.
- Light, A., & Bartlein, P. J. (2004). The End of the Rainbow? Color Schemes for Improved Data Graphics. *Eos*, 85(40), 385–391.
- MacEachren, A. M. (2004). How maps work: representation, visualization, and design.
- Mahe, L., & Broadfoot, C. (2010, December). Too Many Markers! - Google Maps API — Google Developers. Retrieved March 2, 2014, from <https://developers.google.com/maps/articles/toomanymarkers>
- Marks, J., & Shieber, S. (1991). The Computational Complexity of Cartographic Label Placement, 1–28.
- Microsoft Corporation. (2014). Bing Karten – Anfahrtsbeschreibungen, Verkehrsinfos und Straßenbedingungen. Retrieved from <http://maps.bing.com>
- Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running Experiments on Amazon Mechanical Turk. *Judgement and Decision Making*, 5(5), 411–419.
- Robinson, A. H. (1952). The Look of Maps. Madison: University of Wisconsin Press.
- Robinson, A. H., & Petchenik, B. B. (1976). The Nature of Maps. University of Chicago Press, Chicago.
- Robinson, A. H., Sale, R. D., Morrison, J. L., Muehrcke, P. C. (1984). Elements of Geography (5 ed.). Wiley, New York.
- Ross, J., Irani, L., Silberman, M. S., Zaldivar, A., & Tomlinson, B. (2010). Who are the crowdworkers? (pp. 2863–10). Presented at the the 28th of the international conference extended abstracts, New York, New York, USA: ACM Press.
<http://doi.org/10.1145/1753846.1753873>

- Sauro, J., & Dumas, J. S. (2009). Comparison of three one-question, post-task usability questionnaires (pp. 1599–1608). Presented at the the SIGCHI Conference, New York, New York, USA: ACM Press. <http://doi.org/10.1145/1518701.1518946>
- Shneiderman, B. (1996). The eyes have it: a task by data type taxonomy for information visualizations (pp. 336–343). Presented at the 1996 IEEE Symposium on Visual Languages, IEEE Comput. Soc. Press. <http://doi.org/10.1109/VL.1996.545307>
- Stefanakis, E. (2005). Clustering Dynamic Map Objects Based on Density Measures. Presented at the Proceedings of the 22nd International Cartographic
- Woodruff, A., Landay, J., & Stonebraker, M. (1998). Constant information density in zoomable interfaces. *The Working Conference*, 57–65. <http://doi.org/10.1145/948496.948505>
- Yelp. (2014, March 10). Berlin Restaurants, Dentists, Bars, Beauty Salons, Doctors. Retrieved March 10, 2014, from <http://www.yelp.com/berlin>