Star Glyph Insets for Overview Preservation of Multivariate Data

Dominik Jäckle, Johannes Fuchs, Daniel A. Keim
University of Konstanz; Konstanz, Germany

Abstract

Exploring vast spatial datasets often requires to drill down in order to inspect details, thus leading to a loss of contextual overview. An additional challenge rises if the visualized data is of multivariate nature, which we encounter in various domains such as healthcare, nutrition, crime reports, or social networks. Existing overview-plus-detail approaches do provide context but only limited support for multivariate data and often suffer from distortion. In this paper, we dynamically integrate star glyphs as insets into the spatial representation of multivariate data thus providing overview while inspecting details. Star glyphs pose an efficient and space saving method to visualize multivariate data, which qualifies them as integrated data representative. Furthermore, we demonstrate the usefulness of our approach in two use cases: The spatial exploration of multivariate crime data collected in San Francisco and the exploration of multivariate whisky data.

Introduction

Multivariate data accompanies us in our day-to-day life. Prominent examples represent data from healthcare, nutrition, crime reports, or social networks, among others. We typically use spatial representations in order to determine patterns and correlations among dimensions. An example represents the exploration of a huge set of malt whiskies: Each whisky is assigned to the geo-location of its distillery and has several diverse taste categories. The task can be either to seek correlations between particular taste categories and geo-locations, or to find patterns of whiskies for certain taste categories. The latter case can be achieved by applying dimension reduction methods which project the data to a lower dimensional space. When exploring such vast amounts of spatial data, at some point we use zooming and panning interactions to focus on certain regions of interest to obtain a detailed view. However, due to the limited size of the display screen, zooming and panning interactions lead to an inevitable loss of the contextual overview. Overview can be regained by zooming out resulting in a continuous trade-off between overview and detail. Jerding and Stasko argue that the limited size of the display makes it difficult to create efficient global views [25].

Existing *Overview-and-Detail* and *Focus-plus-Context* approaches provide comprehensive methods that typically operate in image space. Overview-and-Detail techniques attach a second viewport to the visualization. Although overview is provided, the user is forced to split his attention, which can result in increased cognitive load [19]. In contrast, Focus-plus-Context techniques integrate overview and detail, but use image-based distortion which restricts the interface by means of zooming levels [36].

In this paper, we propose a novel data-driven *Off-Screen* visualization technique for spatial multivariate data. More specifi-

cally, we contribute a dynamic integration of star glyphs as efficient visual insets for the representation of multivariate off-screen data objects. To do so, we augment the viewport with a dedicated border region including star glyph insets. A result of our approach is depicted in Figure 1.

The remainder of this paper is organized as follows: First, we discuss related work. Then, we introduce the design of our approach and show the usefulness in two use cases, before we conclude and outline future work.

Related Work

In order to preserve the overview of multivariate data during exploration, we need to consider the potentials of both multivariate data visualization and overview preserving visualization. Following, we discuss related work of these areas.

Multivariate Data Visualization

Visual analysis of multivariate data has the objective of allowing the user to identify correlations and patterns among dimensions. Dimensions in multivariate data are not supposed to be considered independently but simultaneously, because they typically provide combined information that contributes to the overall understanding of the data [33]. Various techniques have been presented to visualize multivariate data. Prominent examples of geometric projections are parallel coordinates [22], Andrew curves[1], or star coordinates [27]. Pixel-oriented techniques include recursive patterns [28] and pixel barcharts [29]. However, aforementioned techniques are not optimal to be integrated as space efficient inset giving a coarse overview of dimensions; glyph-based techniques such as Chernoff faces [7] or star glyphs [5] meet these requirements.

Integration of Overview and Detail

In order to allow efficient navigation and provide support for data analysis, the integrated preservation of the contextual overview is crucial. In this paper, the term context refers to the overview of the multivariate nature of the data including information about location and in some cases topology of the data. Following, we give a brief overview of integrated techniques, namely Focus-plus-Context and Off-screen Visualization techniques.

Distortion-oriented Techniques

The pioneering approach of Apperley et al. [2] provides a maximum focus region while all surrounding areas are distorted. Variations of this approach apply the technique for example to one-dimensional visualizations [32, 39]. Furnas [14] further introduced the degree-of-interest (DOI) function as basis for the well-known Focus-plus-Context systems [40, 4]. Additional Focus-
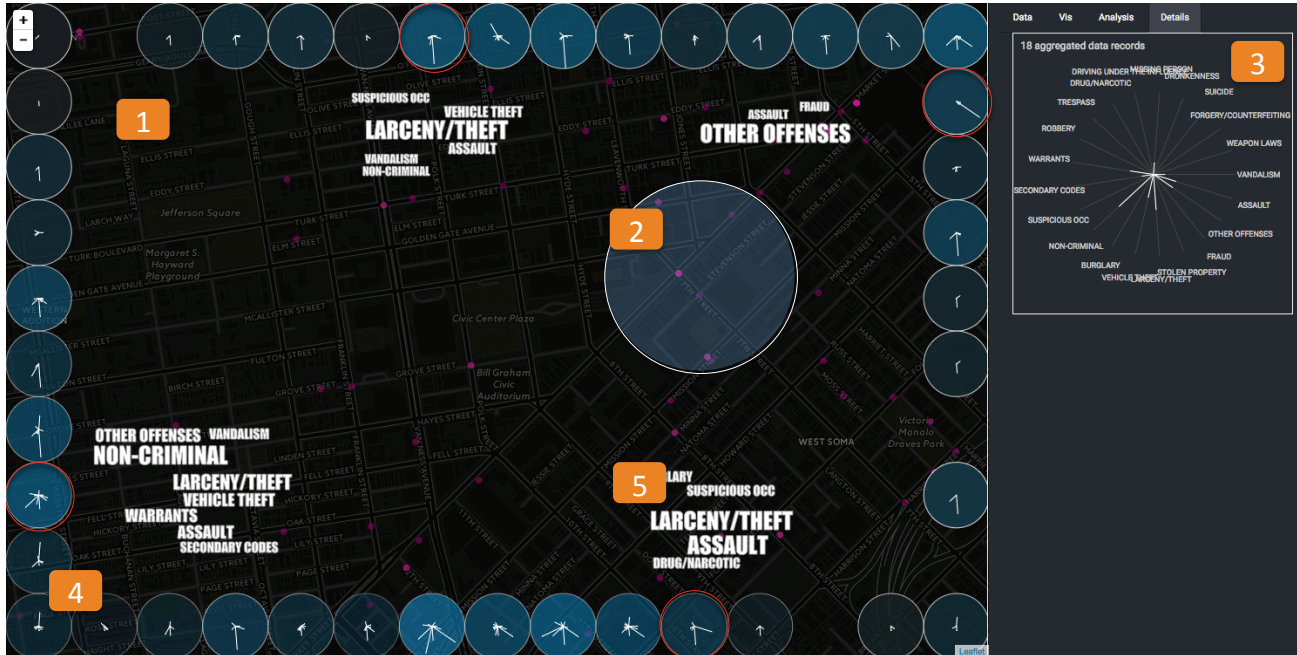
Figure 1. *Visualization of multivariate crime data collected from the official San Francisco Open Data repository [8]. The user first filtered the data and then shows all crimes committed on Monday June 6, 2015. (1) The user drills down to street-level around the area of the Civic Center Plaza for further analysis. (2) Via brushing and linking, the user can analyze (3) details of a certain area. In this case, the user selects some committed crimes along the main street of San Francisco. Each axis of the star glyph corresponds to one crime category. For the selection, the categories assault, other offenses, theft, non-criminal, and warrants are very prominent. (4) At the same time, aggregated overview of the surrounding is preserved using the very same glyph layout. When the user hovers (marked in red) an inset, (3) the detail view is updated and (5) a word cloud is presented next to the inset. The font size of visualized categories corresponds to the amount of crimes committed, respectively. Using the star glyph insets, the user can efficiently identify dimensions of interest.*

plus-context systems can be found in a comprehensive review of existing techniques, which was carried out by Cockburn et al. [9]. However, due to the image-based distortion, some weaknesses are inevitable: Even on high-resolution images, the amount of zoom-levels is restricted by the resolution of the visualization [36]. Furthermore, distortion impedes the ability of precise judgment about scale or distance [3]. Naturally, advantages and drawbacks are always task dependent. In this paper, we provide overview of multivariate data, this is why we aim at a data-driven approach which dynamically adapts to zoom level and position.

Off-screen Visualization Techniques

Unlike distortion-oriented techniques, off-screen visualizations also distort distances to objects that lie outside the viewport (focus region), but consider the data characteristics separately. Besides the application of arrows, Zellweger et al. [46] went one step further and introduced a family of visual proxies that point to off-screen located objects: City Light cues and Halos [3]. While City Light cues simply indicate the presence of any off-screen located object, Halos intersect the display with an arc whose origin is the location of the off-screen object. This way, distance can also be perceived. However, Halo heavily suffers from overplotting. Wedge [20] and HaloDot [18] were designed to improve scalability. Wedge uses isosceles triangles instead of arcs, which results in less overplotting. In contrast, HaloDot enhances Halo and aggregates off-screen located objects. The relevance of the aggregation is shown by thickness or transparency. Yet, different object classes are not considered [17]. Games and Joshi [15] carried on the concept of off-screen visualization and applied visual cues to statistical diagrams. However, aforementioned approaches lack of providing appropriate support for visualization of multivariate data and the preservation of visual topology of off-screen located objects. The approach by Jäckle et al. [23] is based on the concepts of EdgeRadar [21] and Ambient Grids [24] and integrates glyphs for the visualization of off-screen located objects considering topology, data value, and uncertainty.

Besides the application to unconnected objects, various approaches have been presented for the application to node-link diagrams [11, 35, 36]. Ghani et al. [16] propose Dynamic Insets for the exploration of large networks. The idea of insets is to have a coarse overview of off-screen located objects of interest. In our scenario, we aim at providing snapshots of off-screen located multivariate data.

Insets for Overview Preservation

In the following, we discuss and describe design decisions taken in order to provide integrated contextual overview. This work is inspired by the approach of combining several visualization techniques [45] to a system that addresses the problem of limited screen real estate. We adapt the idea of aggregation and integration of glyphs, which was introduced by Jäckle et al. [23], and the use of insets to provide a data-driven snapshot of off-screen objects proposed by Ghani et al. [16].

Glyph Design

Many different techniques exist to visually represent multivariate data. Besides scatterplot matrices, parallel coordinate plots, or pixel-visualizations are glyph-based visualization techniques well-established alternatives. Their main advantages are compact design and flexibility by means of layouting them on the screen. This makes glyph-based visualizations a perfect fit for off-screen visualizations. Glyphs can be integrated in different basic visualization techniques and positioned independently because their reading performance is not influenced by the amount of context information displayed in the background [34]. Following, we first describe the derived glyph and its layout. Then, we outline the integration of data aggregation into the glyph design.

Glyph Layout

Several data glyph designs do exist and hence it is crucial to distinguish between three different mapping categories [44]. Firstly, many-to-one mappings, which support the intra record comparison by mapping all data values to the same visual variable. Well-established representatives are profile glyphs [10]. Profile glyphs linearly arrange dimensions and use position or length encoding to represent respective attribute values. The second category comprises one-to-one mappings. Such designs encode data values with different visual variables. A famous example are Chernoff faces where face characteristics (e. g. angle of the eyebrow, size of the nose, etc.) are adjusted based on the underlying data values [7]. This category offers a nearly endless design space with the most flexible way of assigning data values to visual features. However, since the attribute dimensions are represented with different visual variables, a comparison of two separate dimensions is more difficult compared to many-to-one mappings. The third group contains one-to-many mappings, which represent data values redundantly using at least two visual variables. Colored star glyphs [31], for example, make use of length and color of data rays to encode the respective attribute dimensions. Therefore, important dimensions can be visually boosted by assigning more than one visual variable to this attribute.

Based on the chosen mapping strategy, different analysis tasks are supported. Since we do not want to restrict ourselves in the analysis process, we aim for a design which supports both intra record and inter record comparisons. Besides detecting similar data objects, we are also interested in the comparison of single attribute dimensions. Hence, elementary tasks and synoptic tasks should be supported by the data glyph design. This makes many-to-one mappings an appropriate choice, although their design space is limited.

To derive the final glyph design, we first discuss the layout of dimensions. Results from quantitative experiments suggest to use circular designs rather than linear ones, because they facilitate the detection of single dimensions [12]. The ranking of visual variables by Cleveland and McGill suggest to use position or length for displaying the data value rather than color. In combination radial length encodings are more efficient compared to circular color encodings [37]. The most prominent design with length encoding and radial layout are star glyphs [42]. Figure 2 outlines the three different existing variations to choose from: (a) First, the common star glyph which uses data lines radiating from the center of the glyph and are connected via a surrounding contour line to create a closed shape. (b) Second, the whisker glyph using
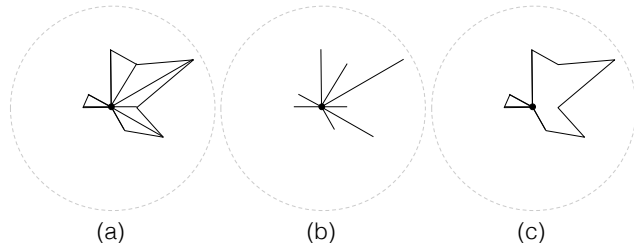

Figure 2. The three considered star glyph variations [42]: (a) The common star glyph uses data lines radiating from the center; maximum values of the data lines are connected and form a contour line. (b) In contrast, the whisker glyph only shows data lines. (c) The sensitivity star glyph only shows the contour line. Quantitative experiments suggest to use the whisker glyph to improve similarity judgments [13].

the exact same encoding, however, the surrounding contour line is removed. (c) Third, the polygon or sensitivity star glyph [6] that displays the contour line only and thus hides the single data rays.

Results from quantitative experiments suggest to use the whisker glyph without the contour line to improve similarity judgments on multivariate data [13]. Additionally, the data rays can be colored to avoid wrong data similarity judgments based on salient shapes [31]. However, enormous amounts of dimensions impede efficient color mappings.

Furthermore, we keep the size of data glyphs as small as possible. Labels are removed from the design in order to allow compact representation. However, little research has yet been carried out on the minimal size of glyphs. Because of this lack of guidance, we decide to use a minimum size of 30x30 pixels for our quadratic aspect ratio, which has been considered a convenient size based on previous comments from expert users.

Data Aggregation

When visualizing vast amounts of data, at some point overplotting is inevitable. Star glyph insets require additional space compared to pixel-based visualizations of single data objects. This is why, we propose to aggregate off-screen located data objects according to the idea of Gonçalves et al. [18]: Off-screen objects are binned with respect to the overlaying grid. The grid size depends on the level of detail the user wants to see; the bigger the grid cell, the less detail is shown, which in some cases is desired. For example, if the user wants to see a coarse overview without calling for too much details. In accordance with the visual analytics mantra, the user can change the parameter settings of the binning and adjust the visualization to her needs [30].

```
Function aggregate(D, isScale)
    result ← Map < Dimension, Value >
    foreach data object O in D do
        foreach dimension d in O do
            result[d] ← result[d] + O[d]
        end
    end
    return isScale ? buildAVG(result, D.size) : result
End
```

Algorithm 1: Dimension-wise aggregation.

We implemented two aggregation methods for the application to star glyphs: the sum of values as well as the average of values per dimension. Calculating either the average or the sum value per dimension depends on the data. If the data dimensions are assigned to a particular scale, as it is the case for whisky taste categories, the average value per dimension expresses which taste category is likely to occur in certain areas. In contrast, if the data dimensions state countable numerical amounts, the sum of values per dimension is the preferable choice. For example in crime analysis, dimensions such as theft or attack are absolute numbers that describe an amount of committed crimes. Thus, we require total crime numbers that occur in certain areas. Algorithm 1 shows the implementation. Parameters are: the set of data objects D which will be aggregated, and a boolean value *isScale* which determines if the values are assigned to a scale. We then initialize an empty map *result* that contains the dimension as key and the sum or average value as value. This map is used later to build the glyph. We iterate all data objects, and for each object O, we iterate the dimensions. For each dimension d, we add the value to the corresponding dimension contained in the result map $result[d]$. Finally, if the data is assigned to a scale, the average value per dimension is returned, the already built sum per dimension otherwise.

We provide the user the extent of the aggregation through the background color of the star glyph. We apply $min-max$ normalization to the amount of aggregated data objects among all bins, and then derive the color value for the star glyph background, respectively. We use a linear colormap from black to light blue, whereas black means low and light blue high aggregation.

The star glyph is used in two variations: Either directly plotted on the map, or the glyph is used as inset for off-screen objects which is described in the hereinafter section. In this paper, we motivate the use of star glyph insets, which is why we intentionally and solely apply the glyph to objects located off-screen. Using our prototype, it is also possible to use the glyph for data located inside the viewport (within the focus region).
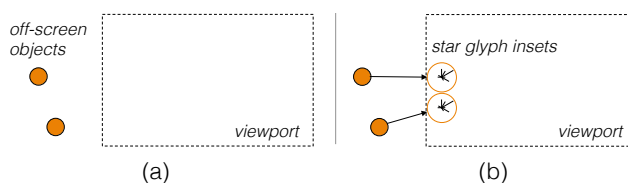
Star Glyph Insets

Figure 3. *Schematic presentation of our approach. Panning and zooming interactions move objects off-screen. In order to preserve the contextual overview, data objects are mapped back to the border region of the display. The use of a star glyph inset helps to maintain overview of multivariate data.*

Zooming, panning, and parameter steering interactions have one commonality: they are instruments used to explore the data at hand. However, once the user zooms in to move from overview to details according to Shneiderman [41], she loses the contextual overview. That is why the aim is to maintain overview while drilling down to details. Following the approach of Apperley et al. [2], we use a rectangular viewport that also provides maximum focus. To provide data-driven overview, off-screen visualization techniques have been proven to be efficient.These techniques are able to not only visualize up to full data topology, but also to con-

sider data characteristics such as its multivariate nature [23]. In addition, the integration of off-screen data objects, clearly shows dense and sparse data areas. As a result, we address the Desert Fog problem [26], because the user is aware of empty areas and consequently does not navigate there.

As described above, zooming and panning interactions lead to a movement of data objects beyond the viewport (off-screen) due to the limited size of the display screen. A schematic representation of how data objects are mapped to the border region is depicted in Figure 3. For each performed interaction, we dynamically update the off-screen visualization, because off-screen located data objects permanently move and thus the indicated position requires update. Each data object is mapped to the border region of the viewport and is visualized as star glyph inset. For aggregated data objects, we first compute the aggregation values per dimension and then visualize the star glyph as representative. We use the same mapping strategy as proposed by Jäckle et al. [23]: Assuming the user navigates from point to point, we draw a virtual line between viewport center and off-screen located object. Then, along this line the object is mapped to the border region of the viewport. In addition, we adapt the idea of HaloDot [17], and use transparency as indicator for the distance between off-screen object and viewport, namely the relevance; objects located near to the viewport are considered to be of higher interest than objects located far apart. This is because it is likely that the user is interested in the surrounding of the area she is currently exploring – once she drills down from overview to detail.
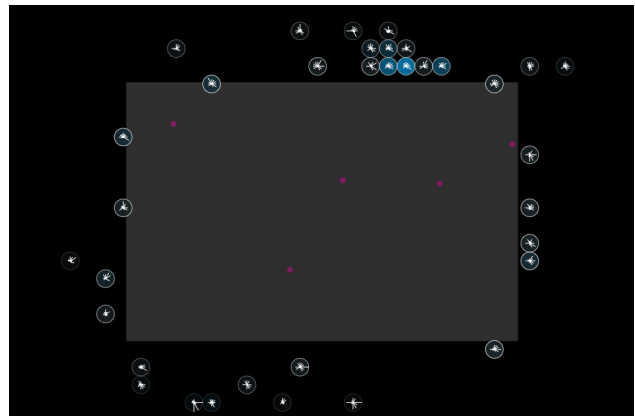

Figure 4. *Towards preservation of data topology. For the attempt to preserve data topology, we reserved a border region of 200 pixels and reduced the size of glyphs to 20x20 pixels. Furthermore, only 86 data records are considered. As we can see, data objects still need to be aggregated (indicated by blueish color), and the sheer amount of small glyphs distributed among the border region complicates efficient perception of dimensions.*

Star glyph insets are space consuming and cannot be perceived within the glimpse of an eye, compared to the simple visualization of data points only. However, they provide an efficient overview. Therefore, and contrary to the approach of EdgeRadar [21], we argue that for the use of multidimensional insets the preservation of the full data topology is neither possible nor adequately supports navigation. Preservation of full topology results in potentially small glyph sizes, so that dimensions are not efficiently perceivable, as depicted in Figure 4. The Figure highlights the drawback of combining topology preservation and multivari-

ate data – additional cognitive load is given by the fact that small glyphs are distributed according to the position of off-screen objects among the border region of the viewport. The Figure shows a small dataset consisting of 86 data objects. This effect intensifies with an increasing amount of data objects. This is why we rather increase the size of the insets and only show location of and distance to data objects.

Interaction

Interaction is crucial when it comes to the analysis of multivariate data, especially if data is represented as glyphs. We implemented two different interactions besides the possibilities to zoom/pan and to adapt all kinds of parameters including color coding, glyph size, border size, etc.

The first interaction is brushing and linking applied to objects located within the viewport. Figure 1 (2) shows the selection of objects through a lens. Objects that are brushed, are aggregated and the result visualized in (3). In this paper, we focus on data located off-screen, which is why we do not further consider interactions related to data visualized inside the viewport.

The second interaction refers to the visualization of star glyph insets. To help users reading exact data values, we decided to implement an enlarged version of the glyph in a detailed view (see Figure 1 (3)). This visualization offers a closer look at data values and adds labels to the dimensions. Since data glyph designs are used more often for synoptic tasks in overview visualizations, such a close up helps to read data values more accurately and, therefore, supports elementary tasks like direct lookups as well. Once the user clicks on a star glyph inset, the detail view is updated.
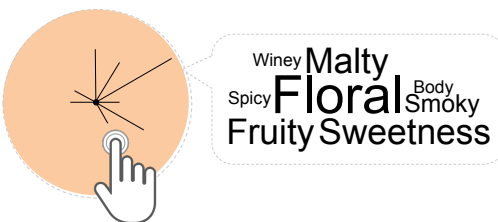
Figure 5. *A word cloud visualization is presented to the user next to the glyph representation when hovering it. The word cloud visualizes the dimensions of the star glyph according to their normalized data value. The font size of the words determines the importance of the corresponding dimensions, this is, the bigger the size of a word, the higher the value of the dimension is.*

When hovering an inset, a word cloud is visualized next to the glyph. The visualization of a word cloud facilitates overview, because important terms have bigger font sizes than unimportant terms. Each term in the word cloud corresponds to one dimensions of the data. Like in Wordle [43], each term is assigned a weight. We derive the weight from the value in the glyph, which is assigned to the term. Figure 5 shows an example for whisky data. In this example, the whiskies are mostly floral, but not that winey and spicy.

Implementation Details and Scalability

We implemented a fully web-based prototype that uses SVG for data visualization. The use of SVG affects the amount of data we can explore interactively, but already provides promising re-

sults. To increase the visual scalability, the visual representation can be replaced with efficient representations such as WebGL, among others. To improve scalability on the data processing side, we use a quadtree as data structure and further use a hash table for the aggregation of data objects. One limitation with respect to scalability is the amount of considered dimensions. Increasing the number of dimensions results in a reduced angle between data lines in the star glyph.

Evaluation

In this section, we showcase the usefulness of our technique in two different use cases. The first use case is crime analysis in San Francisco and the second use case is exploration of whiskies according to their taste categories. Both use cases are carried out using multivariate data. For the visualization of the data we used a dark matter tile server in order to provide maximum contrast.
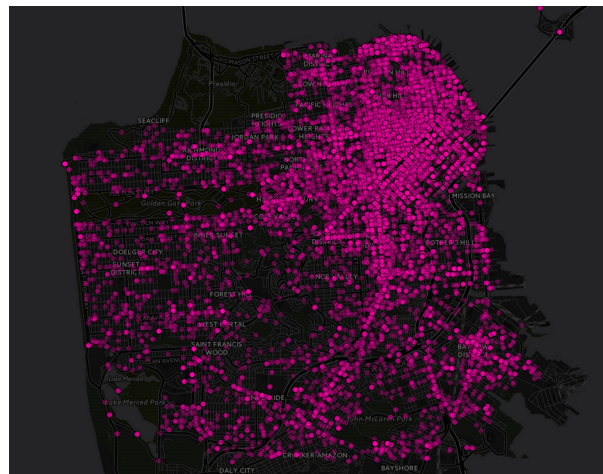
Use Case: San Francisco Crime Analysis

Figure 6. *Mapping of all committed crimes to the map of San Francisco for the period of June 1st, 2015 to June 30, 2015. In total, the dataset consists of 12480 committed and registered crimes.*

The city of San Francisco offers collected crime data through its open data portal [8]. We downloaded the data for the period of June 1st, 2015 to June 30, 2015, which consist of 12480 committed and registered crimes. Figure 6 shows all locations of committed crimes visualized on a map. We can clearly identify crime accumulations in the north eastern area of San Francisco. This area is also known as tourist attraction. The dataset contains various crime categories ranging from larceny/theft, to vehicle theft, up to assault and kidnapping. Figure 8 provides overview over all 36 considered categories that we use as dimensions within our star glyph inset.

This use case involves a user who wants to explore the crime situation in San Francisco before planning her holiday. To do so, she drills-down to the downtown area of San Francisco. In order to highlight the results of our star glyph insets, we intentionally do not show any visualizations related to data objects contained in the focus region. Firstly, the user analyzes the entire dataset. Secondly, she applies a time filter and analyzes the 6th of June.

Figure 7. *Overview preservation of committed crimes in San Francisco over a period of one month (June 1st, 2015 to June 30, 2015). The visualization shows a cutout of the city center of San Francisco, near Union Square. At the same time, by inspecting and hovering the insets, the user can see that in the northern as well as in the western directions, the category larceny/theft is very prominent. This is indicated on the one hand by the word cloud and on the other hand by the star glyph inset, which clearly points out this single dimension. However, the eastern and southern areas are different. While the eastern area seems balanced, the southern area shows more other offenses, assaults, and vehicle thefts.*
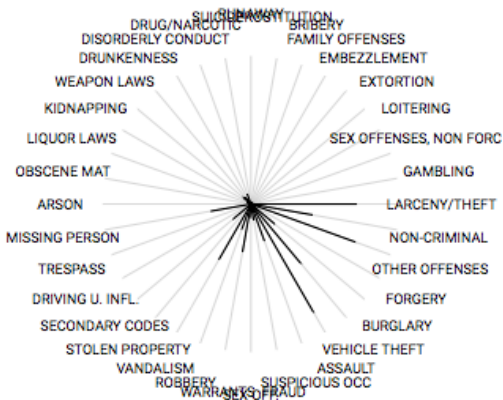

Figure 8. *Dimensions of the San Francisco crime dataset.*

Findings

While exploring the downtown area of San Francisco, the user is interested in surrounding areas to navigate to. Figure 7 shows the data for the entire month of June 2015. The combination of color and transparency provides great overview of the surrounding areas. The user can clearly see, that crimes in the north are in total nearer to her current position than crimes committed in the south of the city. Also, the blueish colored background of the star glyph insets in the south indicate a higher aggregation

level, thus revealing more committed crimes. In fact, the southern area (e.g. South Market) of San Francisco is well known to have high crime rates. This is also revealed by the inset word cloud generated from the glyph beneath. Larceny/theft, other offenses, assaults, and vehicle thefts are very prominent and salient dimensions of the data. In contrast, the northern and western areas are very salient with respect to larceny/theft. It cannot only be clearly seen with help of the word clouds, but also the insets highlight this very prominent dimension.

Then, the user applies a time filter and visualizes the crime data for June 6th, 2015. Also, she slightly adjusts the zoom level and pans to the area of the Civic Center Plaza. Figure 1 depicts the snapshot of this situation. For that day, crime dimensions seem to be a more distributed. However, larceny/theft still is salient in all areas. In the eastern area other offenses stick out as outlier.

Use Case: Whisky Exploration

In this second use case, we make use of a whisky dataset [38], which consists of 86 distilleries and 12 taste categories, respectively. This use case involves a user, who explores the different whiskies without particular analysis task at hand. Figure 10 shows the overview of all distilleries, distributed across Scotland. Especially in the area of Dufftown, many distilleries seem to be located, which is indicated by the high saturation of the magenta color. Figure 11 gives an overview of all included taste categories.

Figure 9. *Visualization of multivariate whisky data. The user drills-down to town-level to see details in the area of Dufftown. Through interactions, she can analyze details and explore the surrounding of the focus area. In the northern part, the user selects sherry-intensive whiskies – this can be identified by looking at the star glyph which reveals whiskies are mostly malty, fruity, floral, and sweet. This is also shown in the detail view. With the help of star glyph insets, the user can seek for taste distributions that she enjoys most, or she can even efficiently compare them.*
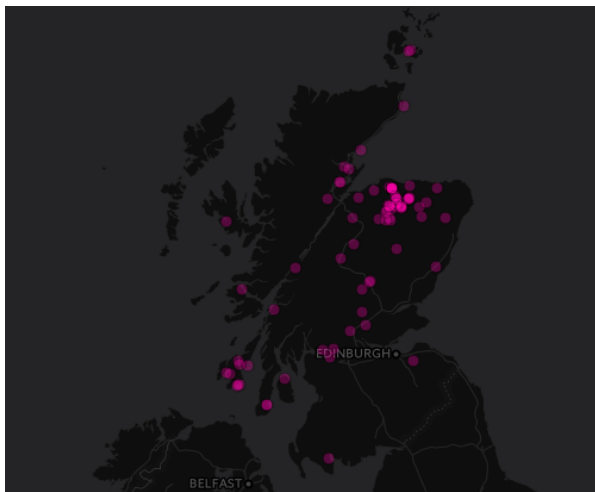

Figure 10. *Overview of 86 whisky distilleries located across Scotland. The dataset [38] contains the exact locations of the distilleries as well as 12 taste categories.*
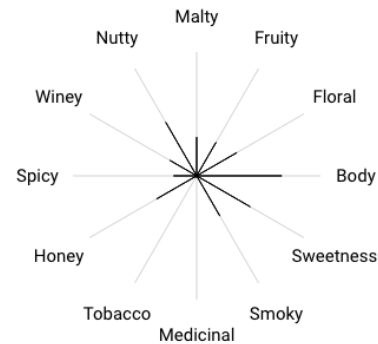
Firstly, the user explores whiskies on the map. Secondly, the user applies Multidimensional Scaling (MDS) [33] to the data and explores the corresponding scatterplot for similar whisky varieties.

Findings

In the first exploration phase, the user explores the map of Scotland for different whisky taste categories. Therefore, she zooms the most promising area: Dufftown. Figure 9 shows the

Figure 11. *Taste categories a whisky can satisfy.*

result of the performed interaction. While having Dufftown in the focus region, she can still explore the surrounding areas. The blueish colored background of the star glyph insets indicate the level of aggregation. In the northern part she selects the very blue glyph. Both detail view and word cloud immediately confirm her assumption based on the star glyph inset and reveal that 12 different distilleries have been aggregated leading to an overall result of mostly sherry-intensive taste categories: sweetness, floral, body, fruity, malty.

At the same time, the user detects a whisky in the south that seems very malty. It is interesting, that the neighbor star glyph on the right has nearly the same shape. Also, the user detects in the western area a star glyph that reveals more body-intensive whiskies and in the eastern part a star glyph that reveals winey-intensive whiskies.
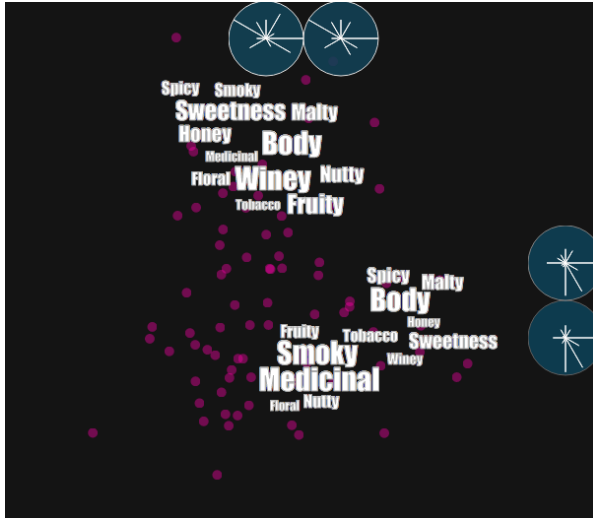
Figure 12. *Application of Multidimensional Scaling (MDS) [33] to the whisky data. While examining a cluster of diverse whiskies in the focus region, the star glyph insets clearly reveal that in the northern part of the MDS scatterplot the taste categories winey and body rule, whereas in the eastern part the taste categories body, smoky, and medicinal protrude.*

However, the user decides to seek similar whiskies and therefore applies MDS to the data. Figure 12 illustrates the result. The result is a scatterplot that contains all distilleries mapped according to their similarity value to each other. The user zooms to a very diverse cluster as can be seen in the Figure. Star glyph insets are generated and visualized for the northern and eastern areas of the scatterplot. It is protruding that the star glyphs in the north as well as the star glyphs in the east look very similar. By hovering the respective glyphs, the word clouds reveal that the northern ones are ruled by the taste categories winey and body. The eastern glyphs are ruled by the taste categories body, smoky, and medicinal.

Concluding Remarks and Perspectives

In this paper, we presented star glyph insets for overview preservation of multivariate data in two-dimensional space. We derived a star glyph design backed up by several quantitative experiments. Star glyphs are space saving and provide an efficient overview of multivariate data. The star glyph is used as inset to visualize off-screen located data, and hence to provide data-driven contextual overview if and when the data space exceeds the available size of the display screen.

We proposed to either use single row or multi row insets. The choice significantly depends on the available size of the display screen as well as the level-of-detail aimed at. In this paper we argue, that due to the sheer amount of dimensions and amount of data, full topology preservation results in additional cognitive load. Also, evaluations of glyph choice and qualification to be used as inset for off-screen visualizations have yet not been carried out and poses promising work for future research. However, we could show in our use cases that star glyphs support certain analysis tasks and provide an useful contextual overview of multivariate data.

References

[1] D. F. Andrews. Plots of high-dimensional data. *Biometrics*, 28(1):pp. 125–136, 1972.

[2] M. D. Apperley, I. Tzavaras, and R. Spence. A bifocal display technique for data presentation. In *Proceedings of Eurographics*, volume 82, pages 27–43, 1982.

[3] P. Baudisch and R. Rosenholtz. Halo: a technique for visualizing off-screen objects. In *Proceedings of the 2003 Conf. on Human Factors in Computing Systems, CHI 2003, Ft. Lauderdale, Florida, USA, April 5-10, 2003*, pages 481–488, 2003.

[4] M. S. T. Carpendale and C. Montagnese. A framework for unifying presentation space. In *Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology*, UIST '01, pages 61–70, New York, NY, USA, 2001. ACM.

[5] J. Chambers. *Graphical methods for data analysis*. Chapman & Hall statistics series. Wadsworth International Group, 1983.

[6] Y.-H. Chan, C. Correa, and K.-L. Ma. The Generalized Sensitivity Scatterplot. *IEEE TVCG*, 19(10):1768–1781, Oct. 2013.

[7] H. Chernoff. The Use of Faces to Represent Points in K-Dimensional Space Graphically. *Journal of the American Statistical Association*, pages 361–368, 1973.

[8] City and Country of San Francisco. *Sf Open Data*. Socrata, 2015. Available at https://data.sfgov.org.

[9] A. Cockburn, A. K. Karlson, and B. B. Bederson. A review of overview+detail, zooming, and focus+context interfaces. *ACM Comput. Surv.*, 41(1), 2008.

[10] S. H. Du Toit, A. G. W. Steyn, and R. H. Stumpf. *Graphical Exploratory Data Analysis*. Springer, 1986.

[11] M. Frisch and R. Dachselt. Off-screen visualization techniques for class diagrams. In *Proceedings of the ACM 2010 Symposium on Software Visualization, Salt Lake City, UT, USA, October 25-26, 2010*, pages 163–172, 2010.

[12] J. Fuchs, F. Fischer, F. Mansmann, E. Bertini, and P. Isenberg. Evaluation of alternative glyph designs for time series data in a small multiple setting. In *Proc. CHI*, pages 3237–3246. ACM, 2013.

[13] J. Fuchs, P. Isenberg, A. Bezerianos, F. Fischer, and E. Bertini. The Influence of Contour on Similarity Perception of Star Glyphs. *IEEE TVCG*, 20(12):2251–2260, Dec 2014.

[14] G. W. Furnas. Generalized fisheye views. In *Proceedings of the SIGCHI Conf. on Human Factors in Comp. Systems*, CHI '86, pages 16–23, New York, NY, USA, 1986. ACM.

[15] P. S. Games and A. Joshi. Visualization of off-screen data on tablets using context-providing bar graphs and scatter plots. *Proc. SPIE*, 9017:90170D–90170D–15, 2013.

[16] S. Ghani, N. H. Riche, and N. Elmqvist. Dynamic insets for context-aware graph navigation. *Comput. Graph. Forum*, 30(3):861–870, 2011.

[17] T. Gonçalves, A. P. Afonso, M. B. Carmo, and P. P. de Matos. Evaluation of halodot: Visualization of relevance of off-screen objects with over cluttering prevention on mobile devices. In *Human-Computer Interaction - INTERACT 2011 - 13th IFIP TC 13 Internat. Conf., Lisbon, Portugal, September 5-9, 2011, Proc., Part IV*, pages 300–308. 2011.

[18] T. Gonçalves, A. P. Afonso, M. B. Carmo, and P. Paulo. Halodot:

Visualization of the relevance of off-screen objects. *SIACG 2011*, pages 117–120, 2011.

[19] J. Grudin. Partitioning digital worlds: focal and peripheral awareness in multiple monitor use. In *Proceedings of the CHI 2001 Conference on Human Factors in Computing Systems, Seattle, WA, USA, March 31 - April 5, 2001.*, pages 458–465, 2001.

[20] S. Gustafson, P. Baudisch, C. Gutwin, and P. Irani. Wedge: clutter-free visualization of off-screen locations. In *Proceedings of the 2008 Conference on Human Factors in Computing Systems, CHI 2008, 2008, Florence, Italy, April 5-10, 2008*, pages 787–796, 2008.

[21] S. G. Gustafson and P. P. Irani. Comparing visualizations for tracking off-screen moving targets. In *CHI '07 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '07, pages 2399–2404, New York, NY, USA, 2007. ACM.

[22] A. Inselberg and B. Dimsdale. Parallel coordinates: A tool for visualizing multi-dimensional geometry. In *IEEE Visualization*, pages 361–378, 1990.

[23] D. Jäckle, H. Senaratne, J. Buchmüller, and D. A. Keim. Integrated Spatial Uncertainty Visualization using Off-screen Aggregation. In E. Bertini and J. C. Roberts, editors, *EuroVis Workshop on Visual Analytics (EuroVA)*. The Eurographics Association, 2015.

[24] D. Jäckle, F. Stoffel, B. C. Kwon, D. Sacha, A. Stoffel, and D. A. Keim. Ambient Grids: Maintain Context-Awareness via Aggregated Off-Screen Visualization. In E. Bertini, J. Kennedy, and E. Puppo, editors, *Eurographics Conference on Visualization (EuroVis) - Short Papers*. The Eurographics Association, 2015.

[25] D. F. Jerding and J. T. Stasko. The information mural: a technique for displaying and navigating large information spaces. In *IEEE Symposium On Information Visualization 1995, InfoVis 1995, 30-31 October 1995, Atlanta, Georgia, USA*, pages 43–50, 1995.

[26] S. Jul and G. W. Furnas. Critical zones in desert fog: Aids to multiscale navigation. In *ACM Symposium on User Interface Software and Technology*, pages 97–106, 1998.

[27] E. Kandogan. Star coordinates: A multi-dimensional visualization technique with uniform treatment of dimensions. In *In Proceedings of the IEEE Information Visualization Symposium, Late Breaking Hot Topics*, pages 9–12, 2000.

[28] D. A. Keim, M. Ankerst, and H. Kriegel. Recursive pattern: A technique for visualizing very large amounts of data. In *IEEE Visualization*, pages 279–286, 1995.

[29] D. A. Keim, M. C. Hao, U. Dayal, and M. Hsu. Pixel bar charts: a visualization technique for very large multi-attribute data sets? *Information Visualization*, 1(1):20–34, 2002.

[30] D. A. Keim, F. Mansmann, J. Schneidewind, and H. Ziegler. Challenges in Visual Data Analysis. In *Information Visualization (IV 2006)*. IEEE, IEEE Press, 2006.

[31] A. Klippel, F. Hardisty, R. Li, and C. Weaver. Colour-Enhanced Star Plot Glyphs: Can Salient Shape Characteristics Be Overcome? *Cartographica*, 44(3):217–231, 2009.

[32] J. D. Mackinlay, G. G. Robertson, and S. K. Card. The perspective wall: Detail and context smoothly integrated. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 173–176, 1991.

[33] B. Manly. *Multivariate Statistical Methods: A Primer, Third Edition*. Taylor & Francis, 2004.

[34] J. P. Martin, J. E. Swan, R. J. Moorhead, Z. Liu, and S. Cai. Results of a User Study on 2D Hurricane Visualization. *Computer Graphics Forum*, 27(3):991–998, 2008.

[35] T. May, M. Steiger, J. Davey, and J. Kohlhammer. Using signposts for navigation in large graphs. *Comput. Graph. Forum*, 31(3):985–994, 2012.

[36] T. Moscovich, F. Chevalier, N. Henry, E. Pietriga, and J. Fekete. Topology-aware navigation in large networks. In *Proceedings of the 27th International Conference on Human Factors in Computing Systems, CHI 2009, Boston, MA, USA, April 4-9, 2009*, pages 2319–2328, 2009.

[37] E. Nelson and P. Gilmartin. An evaluation of multivariate, quantitative point symbols for maps. *Cartographic design: Theoretical and practical perspectives*, pages 191–203, 1996.

[38] NESSIE. *Classification of whiskies*. IAS programme Complex Networks, 2015. Available at `https://www.mathstat.strath.ac.uk/outreach/nessie/`.

[39] U. Rauschenbach, S. Jeschke, and H. Schumann. General rectangular fisheye views for 2d graphics. *Computers & Graphics*, 25(4):609–617, 2001.

[40] G. G. Robertson and J. D. Mackinlay. The document lens. In *Proceedings of the 6th Annual ACM Symposium on User Interface Software and Technology*, pages 101–108, 1993.

[41] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *VL*, pages 336–343, 1996.

[42] J. Siegel, E. Farrell, R. Goldwyn, and H. Friedman. The Surgical Implications of Physiologic Patterns in Myocardial Infarction Shock. *Surgery*, 72(1):126–141, 1972.

[43] F. B. Viégas, M. Wattenberg, and J. Feinberg. Participatory visualization with wordle. *IEEE Trans. Vis. Comput. Graph.*, 15(6):1137–1144, 2009.

[44] M. O. Ward. Multivariate Data Glyphs: Principles and Practice. In *Handbook of Data Visualization*, pages 179–198. Springer, 2008.

[45] J. Wood, J. Dykes, A. Slingsby, and K. Clarke. Interactive visual exploration of a large spatio-temporal dataset: Reflections on a geovisualization mashup. *IEEE Trans. Vis. Comput. Graph.*, 13(6):1176–1183, 2007.

[46] P. Zellweger, J. D. Mackinlay, L. Good, M. Stefik, and P. Baudisch. City lights: contextual views in minimal space. In *Extended abstracts of the 2003 Conference on Human Factors in Computing Systems, CHI 2003, Ft. Lauderdale, Florida, USA, April 5-10, 2003*, pages 838–839, 2003.

Author Biography

Dominik Jäckle is PhD student at the Data Analysis and Visualization Group at the University of Konstanz. His main research interests include developments and applications of new visualization techniques and algorithms for the exploration of vast amounts of multivariate data.

Johannes Fuchs is PhD student at the Data Analysis and Visualization Group at the University of Konstanz. His main research interests are quantitative evaluation methods and glyph-based visualization techniques.

Daniel A. Keim is a full professor and the head of the Information Visualization and Data Analysis Research Group in the University of Konstanzs Computer Science Department. Keim received a habilitation in computer science from the University of Munich. He has been program cochair of the IEEE Information Visualization Conference, the IEEE Conference on Visual Analytics Science and Technology (VAST), and the ACM SIGKDD Conference on Knowledge Discovery and Data Mining. He is on the steering committees of IEEE VAST and the Eurographics Conference on Visualization.