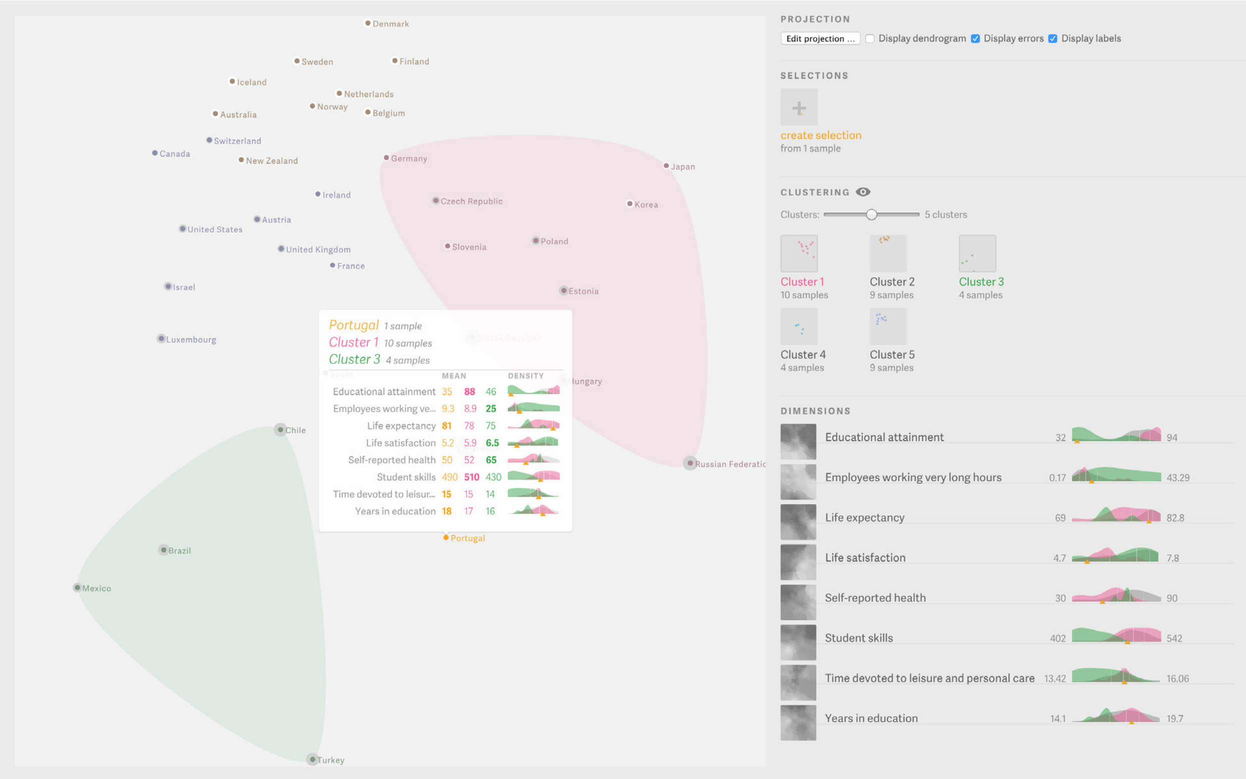
Probing Projections: Interaction Techniques for Interpreting Arrangements and Errors of Dimensionality Reductions

Julian Stahnke, Marian Dörk, Boris Müller, and Andreas Thom



Abstract—We introduce a set of integrated interaction techniques to interpret and interrogate dimensionality-reduced data. Projection techniques generally aim to make a high-dimensional information space visible in form of a planar layout. However, the meaning of the resulting data projections can be hard to grasp. It is seldom clear why elements are placed far apart or close together and the inevitable approximation errors of any projection technique are not exposed to the viewer. Previous research on dimensionality reduction focuses on the efficient generation of data projections, interactive customisation of the model, and comparison of different projection techniques. There has been only little research on how the visualization resulting from data projection is interacted with. We contribute the concept of probing as an integrated approach to interpreting the meaning and quality of visualizations and propose a set of interactive methods to examine dimensionality-reduced data as well as the projection itself. The methods let viewers see approximation errors, question the positioning of elements, compare them to each other, and visualize the influence of data dimensions on the projection space. We created a web-based system implementing these methods, and report on findings from an evaluation with data analysts using the prototype to examine multidimensional datasets.

Index Terms—Information visualization, interactivity, dimensionality reduction, multidimensional scaling

1 Introduction

A primary goal of information visualization is to find patterns and relationships in multivariate datasets. Many visualization techniques have been developed towards this goal such as multiple coordinated views [2], parallel coordinates [14], scatterplot matrices [28], and dimensionality reductions such as multidimensional scaling (MDS) and principal component analysis (PCA) [5]. Dimensionality reductions are a particular class of techniques that synthesise high-dimensional data spaces onto projection spaces with much fewer dimensions, typically the two dimensions of the plane. While most visualization techniques juxtapose the different data dimensions as matrices or columns, dimensionality reductions integrate them into a planar canvas. The projection results in a so-called spatialisation (i.e., embedding) of data elements that approximately represents similarity as proximity and in turn dissimilarity as distance. Considering that the human perceptional system comprises a well-developed capacity for spatial reasoning, the assumption is that spatialisation would be a more natural way [31] to analyse high-dimensional datasets since groupings, separations, and other patterns among data elements become immediately discernible.

However, there are two major caveats linked with dimensionality reduction: first, it can be challenging to interpret the positions of projected elements, and second, the errors that occur with any pro-

• *Julian Stahnke, Marian Dörk, Boris Müller and Andreas Thom are with Potsdam University of Applied Sciences, e-mail: hello@julianstahnke.com and {doerk, boris.mueller, andreas.thom}@fh-potsdam.de*

Manuscript received 31 Mar. 2015; accepted 1 Aug. 2015; date of publication 20 Aug. 2015; date of current version 25 Oct. 2015. For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.
Digital Object Identifier no. 10.1109/TVCG.2015.2467717

jection are seldom exposed to the viewer. In contrast to the mapping of a standard scatterplot, which can be traced back to the particular dimensions of a given dataset, the axes and positions of projections are the result of a complex algorithm whose iterative mechanism is highly opaque. Especially data analysts who have limited experience with dimensionality reduction have difficulty interpreting the meaning of the projection's axes and element positions [21]. While experts in data analysis and machine learning may understand the underlying principles and tend to agree on what constitutes a good projection [18], the specific process of using dimensionality reductions as part of analysis tasks remains very much a black box.

Intricately linked to the projection process is the inherent occurrence of errors. The relative distances among the projected data elements in the plane can only approximate their distances in high-dimensional space; projection techniques can accumulate significant errors. While there has been substantial research on optimising projection algorithms to increase computation speed and projection accuracy, approximation errors cannot be eradicated, they are built into the method. For example, when using multidimensional scaling, the magnitude of projection errors is typically represented in a single value called 'stress' [16], which is unlikely to be understood by novices as it is hard to relate to the positioning resulting from a projection. The lack of transparency can make it difficult to confidently analyse a given dataset and critically examine the quality of a projection. Additionally, stress values can be calculated for each point. There has already been some work on visualizing the distribution of distortions across projections [1, 23], however, the resulting visualizations tend to overpower the projection, do not provide point-specific corrections, and lack the support for crucial tasks: analysing point clusters and mapping dimensions [4]. There is a need for visualization environments that help the viewer in both analysing dimensionality-reduced data and assessing the quality of their visual representation.

With this research, we contribute the concept of *probing* as an integrated approach to interpreting the meaning and quality of data visualizations, especially those relying on dimensionality reduction. Probing aims to integrate the interrogation of the projection with the interpretation of the data and treat these as two necessarily linked activities. Towards this goal we designed a suit of interactive visualization techniques that are specifically designed to support the analysis of high-dimensional data using multidimensional scaling, as well as examination of projection errors resulting from this technique. We report on the implementation of a web-based visualization system and discuss the results from two brief studies with data analysts.

2 Related Work

While not strictly a visualization technique, dimensionality reduction has been utilised in various subareas of visualization, for example to lay out network nodes based on their attribute similarity [7] and collections of documents based on their content [31].

Because of its high computational complexity, considerable work on dimensionality reduction has focused on the optimisation of projection techniques, for example, by utilising the graphics card [13] or guiding the algorithm towards regions of interest [30]. In light of both their computational and technical complexity, tools have been developed that support workflows around dimensionality reductions [12]. However, an open problem is that people using dimensionality reductions may not understand their inner working.

There have been some efforts in evaluating the different visualization techniques for representing dimensionality-reduced data. Two recent studies have shown that 2-dimensional scatterplots, i.e. dot maps, outperform 3-dimensional visualizations [27] as well as scatterplot matrices [22]. The authors suggest that depending on the structure and clustering of the data, several dimensionality-reduction techniques should be compared. However, a recent study on projection evaluation suggests that data analysts still unfamiliar with dimensionality reduction have difficulty providing reliable evaluations [18].

To better understand the needs for visualization design, there is a growing interest in the different ways dimensionality reduction is being used in the field. A two-year qualitative study of data analysts who use dimensionality reduction led to a characterisation of two main types of task sequences: first, analysts tended to verify the explicit and implicit clusters, and second, they sought the mapping between original and synthetic data dimensions [4]. The same study has also shown that the participating analysts who had a limited understanding of dimensionality reduction tended to lack the mental model for the underlying mechanism to trust the quality of a projection [21]. This lack of trust may actually turn away viewers from global projections towards more local arrangements that may be clearer to interpret [6].

This may be in part due the fact that errors are seldom displayed as part of the projection but rather as a global stress metric. It has already been demonstrated how to visualize distortions of an embedding, for example by using a voronoi tessellation around the data points [1, 17]. Such meta visualizations clearly indicate the areas of compression and stretching, which is ideal for a critical assessment of a given projection. Similarly, the global stress of a projection can be broken down to local stress measures which can be integrated with a regular projection [23]. However, the problem with error visualizations is that they tend to prioritise the examination of projection quality at the expense of analysing the characteristics of the underlying data. It is also possible to interactively inspect specific neighbourhoods in a projection to see tears and false neighbourhoods (i.e., areas of stretching and compression) [10]. While this technique is designed for exploration of local distortions, it may be beneficial to consider the potential of interaction more broadly as a way to examine the projection as well as the projected data.

A range of interaction techniques have already been proposed for the visual analysis of dimensionality-reduced data. In fact, especially for the purpose of coping with uncertainties of MDS, interaction techniques for animation, prioritising, and focusing the algorithm [5]. The viewer could prioritise the dimensions to reduce distortions in the important structures [15] weight the underlying model by moving data elements in the projection as a form of feedback [8]. A related idea is to allow the viewer to 'craft' projection functions according to their interest [9]. While these interactive approaches to customising projections are very promising, they assume a certain understanding of either the dataset or its dimensions. However, as aforementioned studies indicate, a recurring task for data analysts is actually understanding the dimensions in the first place [4, 21].

Existing visualization techniques for dimensionality-reduced data tend to either emphasise the exploration of clusters and dimensions or the examination of projection distortions. Considering the poor confidence in dimensionality reductions, there is a need to help analysts make better sense of data arrangements and projection errors.

3 Probing Projections

We introduce the notion of *probing* as a general interaction approach to information visualization that is aimed at both exploring the data as well as examining its representation. The goal is to couple these activities to build confidence in using and judging sophisticated visualization techniques such as those relying on dimensionality reduction.

Dimensionality-reduced data is typically visualized as a scatterplot, with the distances between points approximating the dissimilarities of the samples. Unlike regular scatterplots, the axes of projections have no apparent meaning, which can make them hard to grasp. As the underlying data is not directly shown in the presentation of the dataset, it can be challenging to explore why certain samples are placed where they are and why certain points are far or close from each other. The challenge of making sense of the positioning is further exacerbated due to approximation errors that make it unclear whether the distances are truthful in the first place.

Before these visualizations can be explored to analyse the data, they have to be examined to assess their quality and build trust in them. We propose to integrate these two steps, building on previous empirical research on dimensionality reductions in use [4, 18, 21] and visualization techniques for exposing their distortions [1, 10, 23].

Examining the projection Any visualization is a reduced representation of the underlying data. To confidently use the visualization for exploring the underlying data, the viewer needs to be able to develop trust in the representation [6]. For common representations such as bar charts this trust may have built up by frequent use, but for specialised techniques such as dimensionality reductions, novice viewers need particular support to judge how reliable the positioning is and how the visualization relates to the 'original' data. In the context of dimensionality-reduced data, the trust in the fidelity of the projection is usually established by looking at the total error (*stress*) of the projection, the distribution of errors across the projection [1], or at the relationship between dissimilarities and distances, usually in a Shepard diagram [3]. However, these methods show information *about* the projection, but not as a *part of* the interaction with the projection; the visualization of errors is typically disconnected from the data exploration. The idea behind probing is to embed the examination of errors into the interaction with the projection itself, for example, by revealing which elements exhibit the biggest error, and interactively considering locally corrected representations for them.

Exploring the data As the projection visually conveys the similarity of the samples, the main questions when analysing the samples revolve around understanding *why* they are similar, and why they are not. Analysts ask these questions again and again on different scales and during different tasks, making it important to be able to easily analyse and compare data elements and groups. Visualising the impact of dimensions on the projection space is important as well. Without traditional axes, it can be hard to come to an understanding of the spatial relationships between elements or clusters. The dynamic integration of additional visualisations representing dimensional distribution could aid the understanding of the projection space.

The combination of these principles together with an understanding of how projections are used [4] informed the following design goals for the interactive visualization of dimensionality-reduced data:

- **Show and correct approximation errors.** The user should be aware of distortions in the projection, as well as be able to view the true high-dimensional distances and compare them to the approximated distances to allow for a more careful reading and interpretation of it.
- **Allow for multi-level comparisons.** Comparing elements and groups of elements needs to be easy. At a glance, the viewer should be able to see the aspects different elements have in common, and what sets them apart. The comparison should also adapt to what is being compared; comparing a sample to the dataset is different form comparing two clusters.
- **Spatial orientation.** Without clearly labelled axes, there needs to be another way of finding correlations between dimensions, or visually matching clusters to dimensional values. Furthermore, it may help to expose the dimensional distribution across the space.
- **Consistent design.** The visual appearance and interaction techniques of the analysis environment should be carefully integrated so that error and data visualization do not impede the understanding of the respective other.

4 Design towards Probing

To explore the concept of probing, we have designed and implemented a web-based visualization environment.[1] The interface is subdivided into two main areas: a large projection area and a sidebar, which features a range of visualization controls allowing dynamic selections, clustering, and dimensions, (see Figure 1). These components are interactive, and they support brushing and linking to support the analysis of the project and data from different perspectives.

The projection shows the output of the dimensionality-reduction algorithm. (For the prototype, we decided to use MDS, as it exhibits many of the problems we were dealing with, such as projection errors and no meaningful axes, and we were familiar with it.) Information about projection errors is easily available for the general projection as well as for specific points. Points are displayed as dots whose visual appearance can be changed to convey additional information such as class membership or values for specific dimensions. A

[1] `http://julianstahnke.com/probing-projections/`

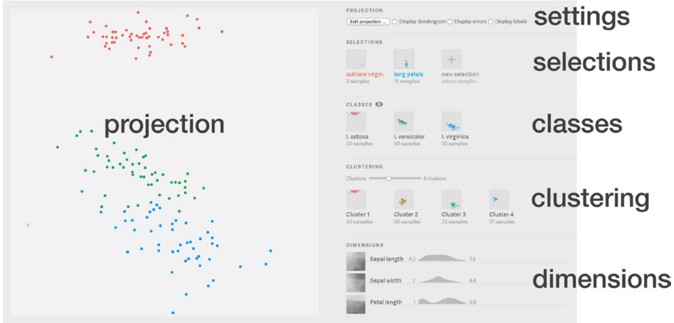


Fig. 1. Overview of the main interface components: projection on the left and sidebar with controls for the projection, groupings, and dimensions.

dendrogram can be overlaid onto the projection to aid understanding of the clustering hierarchy. The distributions of variable values for dimensions are shown as small embedded visualizations in the sidebar and are also available as heatmaps. Individual or multiple points can be selected, and their values inspected and compared.

The sidebar displays visual controls for and additional information about the projection, user-defined classes and clusters, and a list of the dataset's dimensions. Akin to scented widgets [29], the controls feature small dynamic visualizations that guide the viewer's interaction with various aspects of the dataset and projection. The views are coordinated allowing for cross-dimensional exploration.

4.1 Visual appearance

The visual design of the interface aims to emphasise the data, with controls staying in the background as much as possible, while remaining subdued in contrast to data points and clusters. This is accomplished by a restrained use of colour and shape.

The interface uses shades of grey to distinguish between elements. Colour is reserved to display the results of user interaction (pre-defined classes and user selections), separating them from elements generated by algorithms. In a wider sense, this can be thought of as a deliberate encoding of potential findings with colour.

The projection, the most important element, is placed on a bright rectangle, visually emphasising it, and separating it from the rest of the interface. The projection and its small counterparts in the sidebar are the only square boxes in the interface, aiming to make it recognisable at any size. All remaining corners are slightly rounded, and the rectangles filled instead of outlined, to make them more visually compact, emphasising the contents of the box, instead of the box itself.

Care was taken to visually separate additional overlays such as error representation and dimension distribution from the actual samples. Sample properties are always displayed via the dots in the projection, while approximation errors and heatmaps are additive and do not change the appearance of the dots.

Sections in the sidebar and in tooltips are separated by thin lines instead of putting them in boxes, to de-emphasise them compared to the projection. Buttons and other interface elements are lighter than their background and have slight drop shadows, visually separating them from content, which is generally darker than its background.

4.2 Grouping samples

When working with multidimensional data it is important to be able to create, compare, and name clusters or groups of samples. The probing interface supports three main ways of grouping samples:

- **Selections.** By dragging the mouse around a group of points on the projection or shift-clicking them selections of points can be selected that can then be saved and named.
- **Classes.** If the dataset contains classes, they are displayed in the sidebar under their own heading.
- **Clusters.** Using an agglomerative clustering algorithm with a Ward linkage criterion in the high-dimensional space, one can interactively generate varying numbers of clusters. While dragging the slider, the clusters are previewed by colouring the

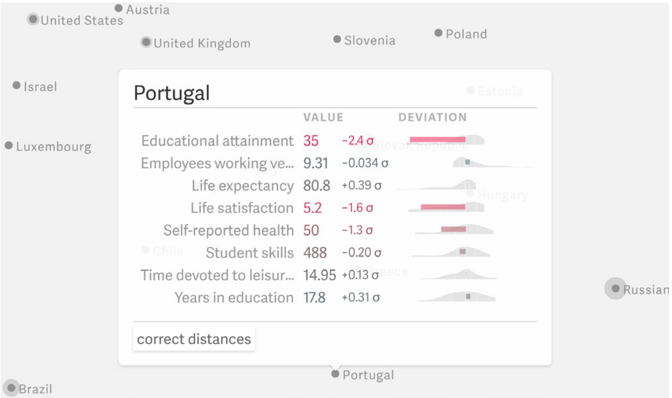


Fig. 2. A tooltip displays the sample's absolute values, standard deviations, and graphical representations for each dimension.

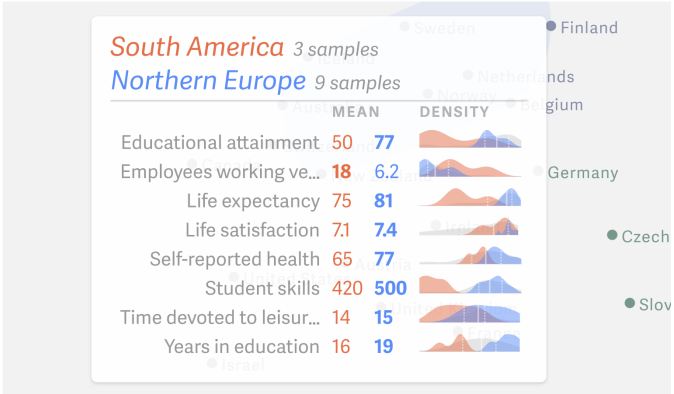


Fig. 3. After selecting one group of samples, hovering over another group shows a tooltip that compares these groups (here selections).

dots on the projection and drawing a convex hull around them. Clusters can also be saved and named as selections.

All of these groupings are displayed as panels in the sidebar. Each selection, cluster, or class is displayed with a thumbnail of its spatial distribution, providing a quick visual way of locating the relevant points in the projection. Some additional information, such as the name or the number of samples, is displayed below the thumbnail. Furthermore, hovering over a grouping's thumbnail displays small density plots in the list of dimensions, as well as a text-based preview of the most deviating dimensions per group.

On the projection, these groupings are coded by colour, with the user being able to switch between displaying classes, selections, or clusters using the respective eye icon.

4.3 Comparing elements

Elements can be analysed by viewing their values and comparing them to the dataset in general, or to other selections in particular. Even a single sample is never analysed in isolation; its values only make sense when compared to the rest of the dataset (see Figure 2).

Analysing a single sample is done by hovering the mouse pointer over a dot on the projection. The values for the corresponding sample are indicated in the list of dimensions. Additionally, a tooltip appears, showing the values for the various dimensions, and their standard deviations. They are displayed in text form for accuracy, as well as in a graphical representation for quick comprehension. The deviations from the mean are displayed as bar charts, with density plots of the whole dataset in the background to provide additional context. The colours of the bars reflect the deviation as well, either in red or blue, and with increasing saturation for higher deviations. If there are too many dimensions to display at once, only the dimensions are shown, in which the sample deviates most. An individual sample can also be compared to other samples by selecting it and hovering over other samples. A tooltip will appear and visualise the differences.

Analysing groups works similarly. When selecting a group of samples, density plots for them are shown in the list of dimensions, comparing the selection to the dataset. A tooltip comparison is displayed as well. Because there is no single value for the dimensions, the means are used instead. The graphical representation also takes this into account, showing a density plot instead of a bar. As shown in Figure 3, groupings can also be compared to each other, displaying density plots for each of them. The methods for comparing samples and groups work together, making it possible to compare a sample to multiple clusters to e.g. find out which of them it should belong to.

4.4 Analysing dimensions

It is important to be able to quickly reference original dimensions when analysing a dimensionality-reduced projection. Two things matter in this regard: the spatial distribution of values in the projection to account for clustering of the data, and the distribution of values in the dimension itself to see how elements compare to other elements within an individual dimension. For this purpose the interface features dynamic heatmaps in the projection and density plots in the sidebar.

4.4.1 Heatmaps

Projections created with most dimensionality-reduction techniques, such as MDS, have no meaningful axes, complicating spatial orientation because dimensional values are distributed nonlinearly. Yet, in order to assign meaning to clustering and find correlations between dimensions, it is important to know how those values correspond with the positioning of the dots. (For some techniques, such as PCA, the contribution of each original dimension can be mapped to the projected dimensions. It would then be possible to display this as a biplot, creating meaningful axes.)

One solution is to use a glyph plot, with the dots themselves being used to represent an additional dimension, for example by varying their size according to the values. This technique is available in the prototype and can be used to visualise a dimension spatially. Where dot size can only show the value distribution for the actual samples, the projection space can also be used to answer a more theoretical question: what values would a fictive sample have to have to be projected to a certain spot? Or, phrased differently: what are the interpolated values for the projection space? We used a heatmap to try to answer this question.

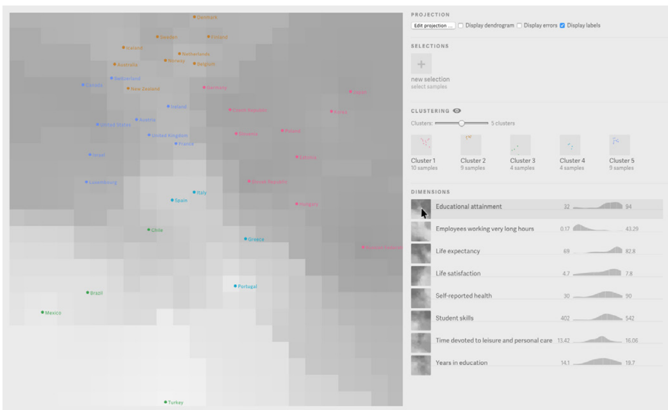


Fig. 4. Hovering over a dimension in the sidebar displays its distribution as a heatmap in the projection on the left.

The heatmap is a grid of cells each representing the value for a certain dimension at its position, with higher values being darker. Brightness is used to avoid confusion with the group colours. This allows to visually assess the value distribution for a given dimension, with smooth transitions between zones. All heatmaps are also shown as thumbnails in the list of dimensions, and on the projection itself

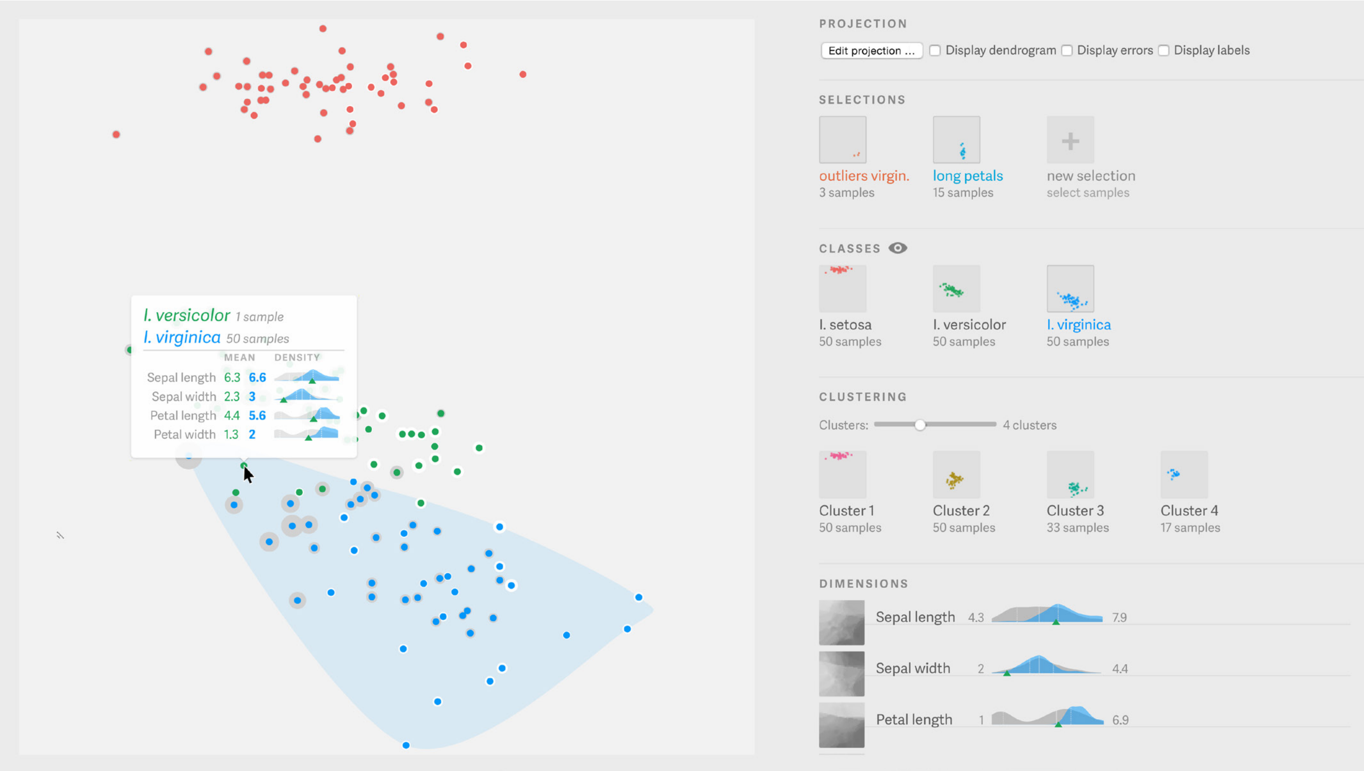


Fig. 5. Comparing an individual sample with a class from the well-known Iris flower data set. In addition to the distribution of dimensions of the class and the value of the sample, the visualization also indicates sample-centric distortions using grey and white halos.

when hovering a dimension in the list (see Figure 4). This allows both for an overview to spot dimensions with interesting patterns, and for a more detailed view together with the dots.

The grid is constructed by dividing up the projection space into a number of cells, with their size based on the mean of distances for the closest three points from each point. A minimum value is used to prevent the grid from getting too small to be useful, and calculations taking too long.

The values for each grid cell are calculated by averaging the values of the points in the cell, or, if there are none, the three closest points for the cell, weighting them according to their distances from the cell. This ensures smooth transitions over large gaps in the projection space, while being responsive to abrupt changes at the same time.

4.4.2 Density plots

While the heatmaps show how the values are spatially distributed in the projection space, kernel density plots in the list of dimensions show their value distributions. In the prototype, currently the plots are generated roughly equivalent to R's bw.nrd0 function which uses Silverman's 'rule of thumb' [24, p. 48]. Percentiles are indicated on the density plots to support the visual assessment. Used together with brushing and linking, it is possible to assess how a sample, or a group of samples, relates to the whole dataset.

Markers or sub-plots for selected elements are shown on the density plots in the list of dimensions, providing dynamic highlights of samples being examined (see Figure 5, lower right). Additional markers display the dimensional values mouse position in the projection space, based on the calculations done for the heatmap, making it possible to gradually track value progressions for multiple dimensions.

4.5 Examining projection errors

Besides exploring the distribution of samples and dimensions, the visualization environment allows for the integrated examination of projection errors by providing per-sample halos, distance corrections, and a dendrogram.

4.5.1 Error halos

The cumulative distance error for each point is displayed as a halo around the dot, with the radius corresponding to the relative amount of error and the value indicating the direction of the error (see Figure 6). This is intended to help the user visually understand the quality of the projection and find potentially unreliable spots. Hovering over a dot shows the errors in relation to the hovered point, to check if the distances between certain points are correct or just projection artefacts, and learn which points *should* be closer together or further apart.

Halos were chosen because their visual properties are a good match for the properties of the error they represent. The error does not belong to the sample, but to the projection, and as such should be displayed by the projection. A halo is clearly connected to the dot, but also part of the projection. Size was chosen as a very intuitive metric to display the amount of error, with points with a large error standing out easily.

The brightness of the halo displays the direction of the error. If the other points are farther away in the projection than in the high-dimensional space, the halo is white; if they are too close in the projection, the halo is dark. White was chosen for points that should be nearer because it stands out more, and while using the prototype ourselves, we often ended up looking for 'similar' samples.

Size and brightness were chosen over colour or shape, as using colour would have clashed with the colouring of the dots, and different shapes were not as glanceable as changes in brightness.

4.5.2 Distance correction

After examining the approximation errors, the viewer might decide that the errors of a certain point warrant more attention. They can then visualize them by selecting to view the high-dimensional distances between the selected point and all others. This removes all projection errors when it comes to distance, but for the selected sample only.

The new, corrected positions for each point are calculated by taking the vector between the selected and the other point and multiplying it by the distance error ratio between the two. The angle between them is kept as is. As a result, the other point moves directly towards or away from the specified point, correcting the distance.

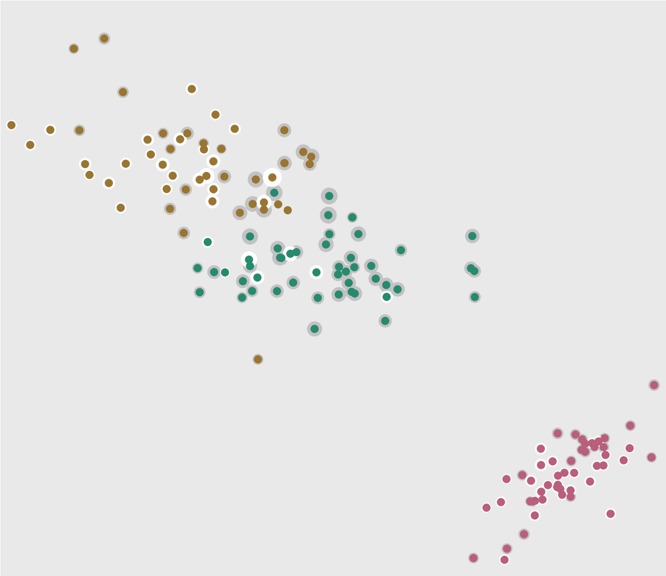


Fig. 6. Halos represent the cumulative error for the respective samples. White indicates that a majority of samples is more similar than indicated by their distance to the given sample; grey indicates the opposite.

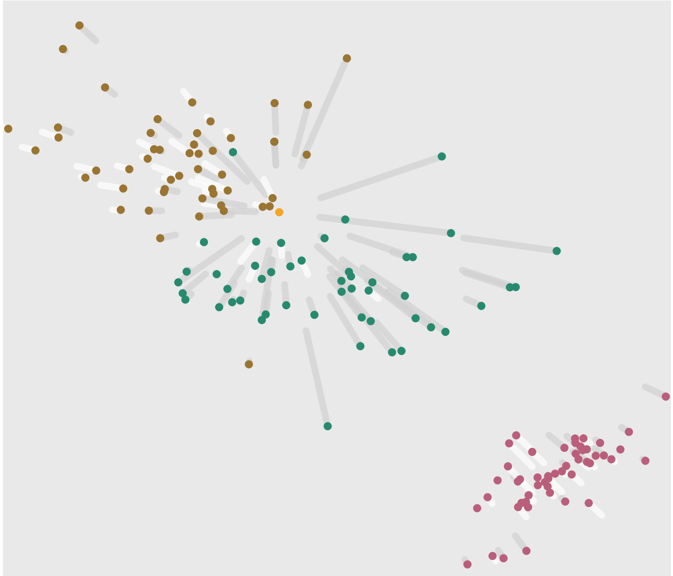


Fig. 8. Projection errors are corrected for the selected sample in orange; grey traces indicate that samples are more different in high-dimensional space, while white traces indicate a higher level of similarity.

The paths travelled by the points are shown as lines, leading from the points' original positions in the projection to the new, corrected positions (see Figure 8). This connects them to their original positions in the projection, and displays the size of the distance error at the same time. Resembling the brightness encoding of the halos, the brightness of the lines indicates whether they've moved closer or farther away.

A problem with this solution is that it introduces new distortions in the spatial relationship between all other points. Only the distances directly between the selected point and the other points are reliable, whereas all the other distances are distorted, and the new positioning might lead to wrong assumptions about potential clusterings. To mitigate this problem, the correction paths are shown.

Another solution would be to recompute the projection while preserving the distances from and to the selected point and being more generous with distance errors among the remaining points. This would somewhat reduce the introduced distortions. However, in a recomputed projection, the positions of the points might change significantly, most likely leading to completely different positions for all points, possibly confusing the observer even if an animation is used.

Fig. 7. Dendrograms mapped onto the projection. Left: projection with low projection error. Right: high projection error.

4.5.3 Dendrogram

In addition to the visualization of errors and corrections, a dendrogram can visualize the samples with regard to their position in the clustering hierarchy. Such a dendrogram (using the same agglomerative algorithm as the clusters) overlaid onto the projection may also help to visualise high-dimensional distances on the projection space [25]. It graphically emphasises clusters by connecting close dots through dense lines. Interestingly, the dendrogram is a surprisingly good indicator of goodness of fit: if many thick, long lines intersect, it is likely that the projection is of low quality.

5 Example: OECD countries

To illustrate the functionality of the interface we visualize the dataset of OECD countries in the prototype (see Figure 9). The dataset contains 8 dimensions for 36 countries[2]. First, the viewer is drawn to the projection and notices Turkey that seems to be a clear outlier, far away from all other countries. To explore why this is, the viewer can examine this sample by hovering over it. A tooltip relating Turkey to the rest of the dataset appears, showing that it deviates strongly from the mean in nearly every dimension. This indicates the positioning as outlier is probably correct.

To test this assumption and build up trust in the visualization, the viewer selects 'correct distances', showing the high-dimensional distances between Turkey and the other countries. This reveals that Turkey should be even farther apart from several of the other countries. Having confirmed that Turkey is an outlier in this dataset, the viewer uses the built-in clustering to get a sense of how the countries are grouped. Playing around with the number of clusters, they notice that there seem to be seven clusters roughly corresponding to the geographical and geopolitical placement of the countries.

Taking a closer look at the positioning of the clustered countries, they realise that the arrangement seems to roughly correspond to geographic directions: Northern and Southern countries are roughly distributed along the vertical axes, East and West along the horizontal. To find out if or how this correlates with the dimensions, the viewer first compares the different clusters. Here the differences along the dimensions are very much pronounced. Interestingly though, life expectancy is lower in Latin America than Asia, while the self-reported health is higher for the former than the latter.

After a few more comparisons between the clusters, the viewer becomes interested in the dimension life satisfaction and turns towards the heatmaps. They notice that the values for life satisfaction and self-reported health seem to be higher in the Western countries, whereas the value for employees working very long hours seems to be especially high in the countries of the far East and the South.

[2] `http://www.oecdbetterlifeindex.org/`

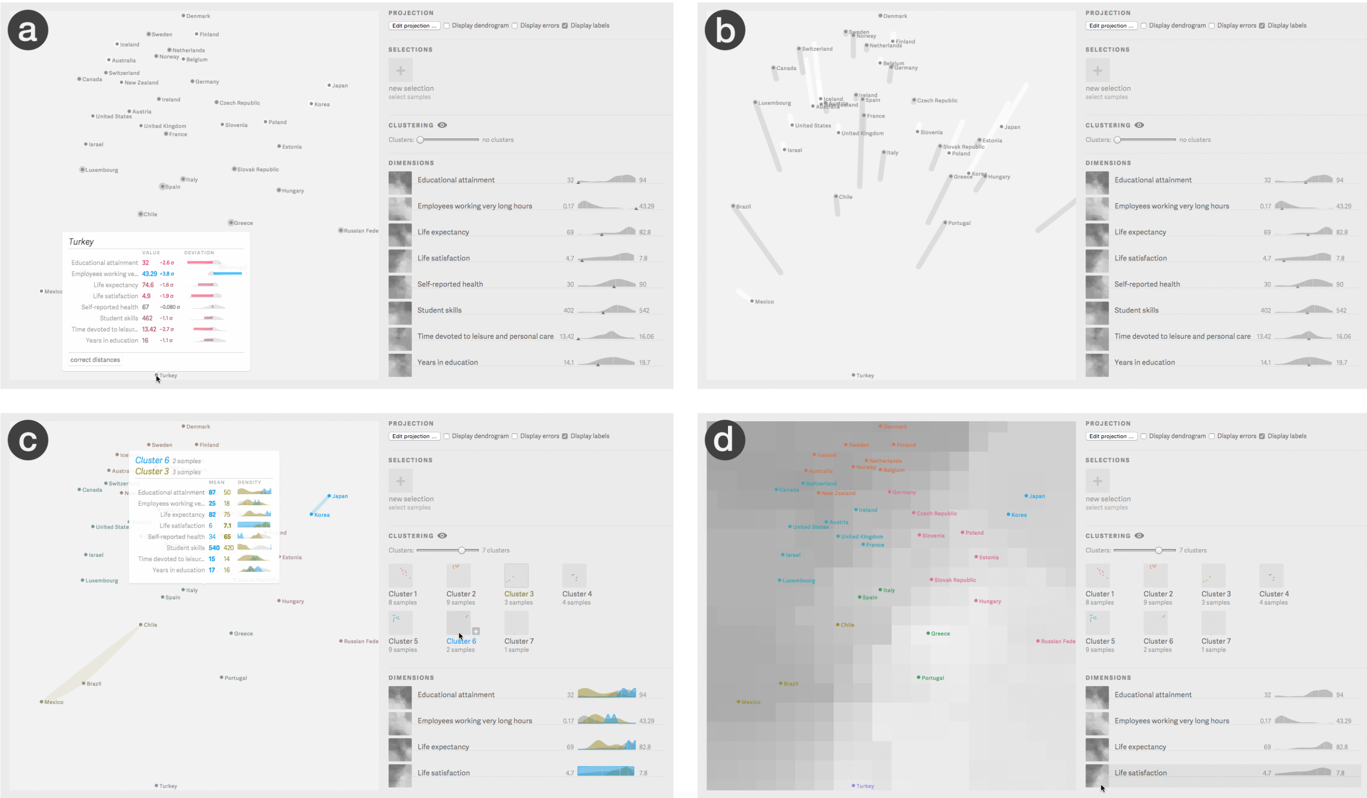


Fig. 9. Probing the OECD dataset: (a) analysing Turkey as an outlier at the bottom of the projection, and (b) its corrected distances, (c) comparing two diametrically positioned clusters, and (d) using heatmaps to relate the dimension 'Life satisfaction' to the projection plane.

6 Evaluation

To examine the potential of probing and get a first sense of how the prototype would be used, we carried out an evaluation in two phases. First, we conducted a brief user study of the web-based visualization tool to assess the general viability of the concept of probing and identify specific opportunities for refinement of the prototype; second, we ran a more focused study to better understand how people would use the different probing functions during the analysis.

6.1 Setup

Both studies consisted of three parts: pre-study interview, a set of predefined tasks to be carried out with the prototype preloaded with a given dataset, and a post-study interview. In the first study, the visualized dataset was about 36 countries and eight socio-economic dimensions, and the second study featured a whisky dataset consisting of 12 dimensions related to taste and smell for 86 brands. While the country dataset builds on participants' prior knowledge, the whisky dataset was entirely novel to all participants. The studies were conducted in a user-experience research lab at a university. The language used was German, with all participants as well as the moderator being native speakers. Video and audio recordings were made of the participants as well as the screen; eye tracking was employed during the first study. The study moderator was not involved in the design process of the prototype.

6.1.1 Participants

We carried out the evaluation with ten participants (aged 26-47, 9 male, one female), six taking part in the first study and four in the second study. Participants were recruited from the university campus and nearby research institutions with the criterion of having experience with data analysis. Short interviews were conducted before the studies to estimate the level of expertise: all participants indeed had experience with data analysis and/or information visualization. In the first study, four were at least somewhat familiar with the concept of MDS and three were using it at least sometimes. In the second study, participants were considerably less familiar with MDS. To introduce the concepts behind MDS and the functionality of the prototype to participants, a brief explanation was given verbally by the moderator (study 1) or as a brief video (study 2).

6.1.2 Tasks and procedure

After being set up at the test computer, a brief explanation of the dataset was provided by the moderator, and the participants were able to freely explore the prototype for two minutes. The exploration phase was intended for the participants to familiarise themselves with the interface, as well as to provide insight into which elements elicit the most attention. During the first study participants were asked to think aloud while using the prototype, providing insight into expectations and suggesting possible issues with the interface.

For the main part of each study, we respectively developed a set of tasks the participants were asked to complete (see Tables 1 and 2. The tasks of the first study were ordered so that each task built on the previous task and become increasingly more complex as the participants became more familiar with the prototype and the data. During the second study the tasks were carried out independently from each other; here we paid particular attention to the interaction sequences for each question. Both the moderator and an observer were noting markers for each main function of the prototype.

After the tasks of the first study, we conducted a follow-up interview, in which we used the retrospective think-aloud method and asked a semi-structured set of questions, requesting their opinion of the prototype and suggestions for improvement. They were also asked to grade the prototype along four general terms. During the closing interview of the second study, participants were asked to evaluate specifically the different functions of the prototype and the more general potential of probing. The average session length of both studies, including pre- and post-interviews, lasted for about 60 minutes. The tasks themselves took about 35-45 minutes.

Table 1. Tasks of the study 1 and their completion rates (C: Completed, H: With help, and N: Not completed).

#	Task	C	H	N
1.	You want to find out which data points about Turkey are most interesting. Why is it so far out from other countries?	6	0	0
2.	Now you want to compare Turkey to other countries. What are the differences?	5	1	0
3.	There may be approximation errors in projections. How can you display them?	4	2	0
4.	After completing the previous tasks, you want to know if the positioning of Turkey is correct.	4	2	0
5.	How would you go about clustering the data?	6	0	0
6.	After clustering the data, you want to find out about the differences between two clusters.	5	1	0
7.	How do you interpret the positioning of countries in this projection with regards to the dimensions?	2	4	0
8.	Can you discover correlating dimensions, or synthesise new dimensions?	1	3	2

6.2 Findings: Study 1

The first two tasks of the first study were completed quickly and by all participants. The error-reading and correction tasks (3-4) were problematic for participants unfamiliar with dimensionality reduction, and some participants had to receive help from the moderator. The clustering tasks (5-6) were completed almost without problems. The dimension-related tasks (7-8) were more challenging, with two participants not completing the last task even when helped by the moderator.

6.2.1 Observations

The eye tracking data reflected the general layout of the prototype, with clear hot spots around the middle of the projection and the list of dimensions. The tooltip received considerable attention during comparisons of elements, but participants often turned their gaze to the list of dimensions in the sidebar to compare the values of elements across dimensions. Some negatively remarked that the tooltips are not as useful as they could be, because they overlapped the projection area.

Reading the approximation errors was problematic to some participants, especially those without previous knowledge of dimensionality reduction techniques. Participants recognised the different white/grey colouring of the halos, but were first unsure about the meaning, looking for a legend or an explanation. Some participants thought white indicated higher errors than grey. After entering the distance correction mode, when they saw the differently coloured paths leading to and from the original positions, they were able to make sense of the colouring. Most participants were also confused by the fact that the error display was context-sensitive; changing to show errors in regards to hovered points: *"There's a difference in the meaning, there seems to be a general white and a country-related white?"* (P4). Generally, the participants appreciated being able to view the errors, with one participant even suggesting *"a list of the samples with the worst distance errors between them"* (P3). Another participant asked why the halos were round, expecting them to be shaped to express the directions of the errors in terms of the projection's coordinate system.

The eye tracking data showed that participants used both heatmaps and density plots to verify clusters. The dimensional heatmaps and their thumbnails were used to find correlations between dimensions as well, even though participants had trouble interpreting their exact meaning: *"The visualization of the heatmaps is difficult to read. What is dark and what is light? There is no legend."* (P3). The density plots were understood by all participants, but found to be confusing when too many overlapping sub plots were shown.

6.2.2 Post-study interview

During the post-study interviews, many participants mentioned the density plots making it possible to compare elements quickly. The brushing-and-linking function was also repeatedly mentioned to be "well thought out" and "there when needed". The design of the interface was positively described as restrained, though some participants would have liked it to be "more colourful". Not so well received was the mouse interaction with the projection; elements were described as too small to easily select, and many participants preferred lasso instead of rectangular selections: *"I thought I could draw free-form shapes as well."* (P6). Among the possible improvements suggested were additional tools for comparing dimensions, such as correlation analysis, and the ability to zoom and pan the projection.

6.3 Findings: Study 2

The aim of the second study was to better understand how people would use the prototype when exploring unfamiliar data and how the different functionalities for data and error representations were used.

6.3.1 Techniques

The sequences per task in the second study shows that more complex questions also resulted in longer interaction sequences (see Table 2). Close inspection of how the different functions were used suggested diverse uses as well as differing preferences among the participants.

P - Projection was used by all participants during all tasks to locate outliers, analyse distances, and gain detailed information via tooltips.

E - Error display was almost exclusively used in task 4, and a little bit during free exploration. The participants often (3 out of 4 times) required help from the moderator to interpret the projection display.

D - Data distribution through the heatmaps was actively used to compare and interpret clusters and individual elements. The data distribution as part of the tooltip was used throughout all tasks.

C - Clustering was used moderately by all participants (except P7 who employed it extensively) used it moderately with a tendency for either 3 or 9 clusters; the latter was the maximum number of clusters possible. The clusters were often compared with dimensions in tooltips and analysed with the help of heatmaps.

T - Tooltips with data details were brought up by all participants throughout all tasks, mainly to see more detailed information for data points and also in combination with other techniques.

G - Manual groupings were created by several participants especially in tasks 1, 3 and 4, however, the grouping feature was sometimes triggered accidentally.

X - Misinterpretation occurred when participants tried to understand scales of the MDS axes; furthermore, sometimes the error display as halos and displacements was interpreted falsely.

H - Help by moderator was provided mostly when items could not be found, questions were not understood or forgotten during task completion, and techniques were not discovered fully (e.g., correct distances). Help was mostly sought during tasks 4 and 5.

6.3.2 Observations

While the current functionalities were sufficient to solve the tasks in the study setting, the amount of sequences and verbal statements by participants suggest that the current functionality is not sufficient for hypothesis-led analysis, but rather open exploration. During the multidimensional comparison tasks with clustering such as task 5, it seemed that the mental models of the experts differed from the visual representation. The participants expressed the opinions, that ad-hoc usage of the visual representation may need a lot of time and would be inefficient. *"It would be better, when the tool supports hypotheses by the researcher, since it is more goal-orientated."* (P10) However, during the follow-up interview participants frequently expressed that with these new representations of data it would be possible to compare datasets in a more natural 'narrative' way beyond mere statistical impressions. But simultaneously some participants questions whether the tool in its current form would support a scientific approach: *"It is important to know what is the projection algorithm and what are the activated dimensions for the clustering."* (P10)

Table 2. Tasks and markers of the interaction sequences during the second study. See Section 6.3.1 for description of functions used.

	1. Free exploration.	2. How would you try to characterise the whisky type Talisker?	3. In what way are the whiskies Scapa and Aberlour different in their properties?	4. Can you discover whether the data projections of the whiskies Oban and Scapa are correct or with error deviation? How do you interpret these visualizations?	5. Can you discover which parts of the cluster combinations are malty, nutty and spicy?
P7	P T T D D C D D C D T C H H C C H D	C C D T T T D T D D P P	P T C C C C C C C D D C T X C C C T T G C C C C C C H C C T C T T T T H P P T	E E E E T T E T E E T P P X T T E E P E T X E P	D C T C D D D C D C D C T C T C T T C D T D C D C D D C T C T T C C T C D T T C D C T T T T D
P8	G X T P G G G D D X X	P T X T X T X T D T D D X T	G X P X T G D D X D X X X X H H D X T T D	P P H H P X P X H X P E E H H H H X E G E X H E T E H T E E	P E P G D H D D C C D D T C P P X X X T H H H H X H
P9	D D P T T C P T P E T T D T P D T G X	D P T X T X	P D P G H P P T T G X G X G X P X D G D D D H H X T P P X T	H P T E X E X E E D E H T P T E E E H E H T H E H E T E E H E E	E T D D C D C T D T T C T T T C T C C H H C T T T D D X T D X D D
P10	T T T D P E P G X C D D T T G G	P P T H H D	P C P P P H P T H G T G H D H T X X	P P T E E X H H H T H E E E H H E T E H E T E E E T G X H H E T E T E E T H	E C D C D D C D C D C E D C D C H T H H D X C D C C D D C D T D T D T T D T D E X H H H C H C C C

7 Discussion

Following the idea of probing, our design goals for the visualization of dimensionality-reduced data were to reveal errors, allow for comparisons, and support spatial orientation, all while remaining consistent in the visual design and interactions. The largely successful completion of the tasks during the studies suggests that the implemented prototype generally succeeded in satisfying these aspirations. However, the post-study interviews revealed a range of specific issues of the prototype and more general challenges for visualizing dimensionality-reduced data. In the following, we briefly discuss three points: learnability and manipulation of projection methods, and the larger potential of probing for other types of visualization techniques.

Learnability The evaluation suggested specific problems with the prototype that can be in part be addressed with simple changes: using a legend to make the dimensional heatmaps clearer, better explaining the different types of error, and making the option to show error-corrected distances easier to find. There is a need for more effective mechanisms for grasping the concepts behind dimensionality reduction, recognising the involved errors, and pursuing dimension-related tasks. A built-in help explaining features of the prototype was something some participants missed and was then included for the second study. Something like this seems necessary for a tool that is designed to be simple to use, as data analysis and visualization are large fields, and not everyone is an expert in working with dimensionality-reduced data. But beyond that, there may be an opportunity to use the tool as an explanatory, didactic device; a tool to explain the basics of multidimensional scaling to interested practitioners of data analysis, by helping them analyse their own datasets. For this area the question is how to also support goal-directed, hypothesis driven research.

Manipulation A related area with great potential for further investigation is the support of dimension-related tasks as was inquired by several participants. Some improvements are simple; like the ability to reorder, group, or annotate dimensions, or to weigh them differently according to their importance. These types of manipulations relate to the notion of semantic interaction [8] as a method to adjust the underlying method for spatialisation. Visualising the differences between two or more dimensions on the projections space is another potentially interesting feature, leading up to the possibility of how to combine correlating dimensions into new ones. Furthermore, our prototype currently focusses on the process of probing, neglecting what to do with the results of this process. Viewers should be able to save their findings and share them with others. When doing so, they may want to focus on the presentation of their findings, which may require an entirely different interface.

Probing visualizations Applying the notion of probing to different types of variables and visualization techniques is another important area with great potential. Currently our design and prototype only supports scalar variables, but we are already thinking about what changes need to be made to support categorical variables or even entirely different types of datasets with different needs. For example, the visualizations of text corpora that need to deal with a number of dimensions far bigger than currently supported by the presented interface. In this context, it may be particularly important to 'bubble up' interesting dimensions when comparing elements, or think about how to present a big number of differences between elements. Currently, the interface can only accommodate about a dozen dimensions and at most a few hundred samples. While the integration of visualization of projection errors and projected data prompted favourable responses, it is not clear whether the idea of probing as an approach to integrate the visualization of data and its errors has wider applicability in visualization. For many people the visualization pipeline remains a black-box just as experienced data analysts may actually not sufficiently comprehend the working of MDS to confidently put it to use in their work.

8 Conclusion

As dimensionality reductions have been gaining popularity for visual analysis of high-dimensional data, there is a growing need to support data-related tasks concerned with the interpretation of the clusters and dimensions, while also enabling the analyst to confidently judge the quality of the projection. In this paper we introduced the concept of probing as an approach that integrates these two aspects: exploring the data and examining the errors. We created a visualization environment for dimensionality-reduced data that provides a suite of interaction and visualization techniques especially designed for probing activities. The results of a brief user study are promising, yet also raise specific issues related to the learnability of projection interfaces as well as directions for future research on the visualization of dimensionality-reduced data. Considering growing concerns in the visualization community about trust [6], rhetoric [11], and uncertainty [19] probing could be one way to let viewers develop visualization literacy, empowering them to look beneath the surface of visualizations. The aim to turn the inner workings outside and disclose inaccuracies, distortions, and omissions may not just be useful and necessary for error-prone dimensionality reductions, but also for any visualization in general. Considering recent studies of the persuasive power of visualizations [20, 26], we need to develop methods for interrogating not just the data, but the visualization itself. The concept of probing can be seen as one step into this direction.

ACKNOWLEDGMENTS

We are very grateful to the reviewers and our study participants for their thoughtful comments on this work. We would like to particularly thank Michaël Aupetit and Michael Sedlmair for insightful discussions early during this project and their feedback on the prototype. Many thanks also to Till Nagel for his feedback on drafts of this paper and Jens Rauenbusch for his helpful assistance during the evaluation. This research was in part funded by HERE, a Nokia company.

REFERENCES

[1] M. Aupetit. Visualizing distortions and recovering topology in continuous projection techniques. *Neurocomputing*, 70(7):1304–1330, 2007.

[2] M. Q. W. Baldonado, A. Woodruff, and A. Kuchinsky. Guidelines for using multiple views in information visualization. In *AVI 2000: Advanced Visual Interfaces*, pages 110–119. ACM, 2000.

[3] I. Borg and P. J. F. Groenen. *Modern multidimensional scaling: Theory and applications*. Springer Verlag, 2005.

[4] M. Brehmer, M. Sedlmair, S. Ingram, and T. Munzner. Visualizing dimensionally-reduced data: Interviews with analysts and a characterization of task sequences. In *Proc. ACM BELIV Workshop*, 2014.

[5] A. Buja and D. F. Swayne. Visualization methodology for multidimensional scaling. *Journal of Classification*, 19(1):7–43, 2002.

[6] J. Chuang, D. Ramage, C. Manning, and J. Heer. Interpretation and trust: designing model-driven visualizations for text analysis. In *CHI '12: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 443–452. ACM, 2012.

[7] M. Dörk, S. Carpendale, and C. Williamson. Visualizing explicit and implicit relations of complex information spaces. *Information Visualization*, 11(1):5–21, 2012.

[8] A. Endert, P. Fiaux, and C. North. Semantic interaction for visual text analytics. In *CHI '12: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 473–482. ACM, 2012.

[9] M. Gleicher. Explainers: Expert explorations with crafted projections. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2042–2051, 2013.

[10] N. Heulot, M. Aupetit, and J. Fekete. Proxilens: Interactive exploration of high-dimensional data using projections. In *Proceedings of the EuroVis Workshop on Visual Analytics using Multidimensional Projections*, pages 11–15, 2013.

[11] J. Hullman and N. Diakopoulos. Visualization rhetoric: Framing effects in narrative visualization. *TVCG: Transactions on Visualization and Computer Graphics*, 17(12):2231–2240, Dec 2011.

[12] S. Ingram, T. Munzner, V. Irvine, M. Tory, S. Bergner, and T. Moller. Dimstiller: Workflows for dimensional analysis and reduction. In *Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on*, pages 3–10. IEEE, 2010.

[13] S. Ingram, T. Munzner, and M. Olano. Glimmer: Multilevel mds on the gpu. *Visualization and Computer Graphics, IEEE Transactions on*, 15(2):249–261, 2009.

[14] A. Inselberg and B. Dimsdale. Parallel coordinates: a tool for visualizing multi-dimensional geometry. In *Proceedings of the 1st conference on Visualization '90*, pages 361–378. IEEE, 1990.

[15] S. Johansson and J. Johansson. Interactive dimensionality reduction through user-defined combinations of quality metrics. *TVCG: Transactions on Visualization and Computer Graphics*, 15(6):993–1000, Nov/Dec 2009.

[16] J. Kruskal. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1):1–27, 1964.

[17] S. Lespinats and M. Aupetit. Checkviz: Sanity check and topological clues for linear and non-linear mappings. In *Computer Graphics Forum*, volume 30, pages 113–125. Wiley Online Library, 2011.

[18] J. M. Lewis, L. Van Der Maaten, and V. R. de Sa. A behavioral investigation of dimensionality reduction. In *Proc. of the 34th Annual Conference of the Cognitive Science Society*, pages 671–676, 2012.

[19] A. M. MacEachren, R. E. Roth, J. O'Brien, B. Li, D. Swingley, and M. Gahegan. Visual semiotics & uncertainty visualization: An empirical study. *TVCG: Transactions on Visualization and Computer Graphics*, 18(12):2496–2505, 2012.

[20] A. Pandey, A. Manivannan, O. Nov, M. Satterthwaite, and E. Bertini. The persuasive power of data visualization. *TVCG: Transactions on Visualization and Computer Graphics*, 20(12):2211–2220, Dec 2014.

[21] M. Sedlmair, M. Brehmer, S. Ingram, and T. Munzner. Dimensionality reduction in the wild: Gaps and guidance. *Dept. Comput. Sci., Univ. British Columbia, Vancouver, BC, Canada, Tech. Rep. TR-2012-03*, 2012.

[22] M. Sedlmair, T. Munzner, and M. Tory. Empirical guidance on scatterplot and dimension reduction technique choices. *TVCG: Transactions on Visualization and Computer Graphics*, 19(12):2634–2643, 2013.

[23] C. Seifert, V. Sabol, and W. Kienreich. Stress maps: analysing local phenomena in dimensionality reduction based visualisations. In *EuroVAST 2010: International Symposium on Visual Analytics Science and Technology*, 2010.

[24] B. W. Silverman. *Density estimation for statistics and data analysis*, volume 26. CRC press, 1986.

[25] A. Skupin and S. Fabrikant. Spatialization methods: A cartographic research agenda for non-geographic information visualization. *Cartography and Geographic Information Science*, 30(2):99–120, 2003.

[26] A. Tal and B. Wansink. Blinded with science: Trivial graphs and formulas increase ad persuasiveness and belief in product efficacy. *Public Understanding of Science*, 2014.

[27] M. Tory, C. Swindells, and R. Dreezer. Comparing dot and landscape spatializations for visual memory differences. *TVCG: Transactions on Visualization and Computer Graphics*, 15(6):1033–1039, Nov/Dec 2009.

[28] M. O. Ward. Xmdvtool: Integrating multiple methods for visualizing multivariate data. In *Proceedings of the Conference on Visualization'94*, pages 326–333. IEEE Computer Society Press, 1994.

[29] W. Willett, J. Heer, and M. Agrawala. Scented widgets: Improving navigation cues with embedded visualizations. *TVCG: Transactions on Visualization and Computer Graphics*, 13(6):1129–1136, Nov/Dec 2007.

[30] M. Williams and T. Munzner. Steerable, progressive multidimensional scaling. In *InfoVis 2004: Symposium on Information Visualization*, pages 57–64. IEEE, 2004.

[31] J. A. Wise, J. J. Thomas, K. Pennock, D. Lantrip, M. Pottier, A. Schur, and V. Crow. Visualizing the non-visual: Spatial analysis and interaction with information from text documents. In *InfoVis 1995: Symposium on Information Visualization*, pages 51–58. IEEE, 1995.