Utrecht University

Computing Science
Algorithmic data analysis track

Master's Thesis

A VIEWPOINT-DRIVEN COMPARISON OF 3D VERSUS 2D PROJECTIONS OF HIGH-DIMENSIONAL DATA

Author
Wouter Castelein
5936608

Primary supervisor
Prof. dr. ir. Alexandru C. Telea

Secondary supervisor
dr. Tamara Mtsentlintze

Abstract

Dimensionality reduction is a popular data visualization technique that projects high-dimensional data to a low-dimensional space (2D or 3D) while preserving distance and/or neighborhood relations between points. The projected dataset can then be visualized in, for example, a scatterplot. This process greatly enhances interpretability of the dataset while minimizing information loss. While projections that target the 2D space have been studied in detail both quantitatively and qualitatively, 3D projections are far less well understood, with authors arguing both for and against the added value of a third visual dimension. More information can be stored in 3 dimensions, and point overlap in visualizations is reduced, but exploring and understanding a 3D projection adds complexity for users. A user can only ever see a 2D rendering of the 3D projection as seen from a certain viewpoint. In each view many points can be occluded, and therefore, in order to assess the entire projection, it is required to consider multiple views found by rotating it. Certain quality metrics can measure to what extent the structure of a dataset is preserved in a projection. But as of now, quantitative studies of 3D projections have disregarded this viewpoint limitation in 3D by using quality metrics that consider point neighborhoods and inter-point distances in 3D. We propose a different approach of measuring the quality of 3D projections, where we use quality metrics designed for 2D projections not on the entire 3D projection, but on multiple 2D views of a 3D projection. This tells us how the quality of a 3D projection changes as a function of the viewpoint, which we believe can give a better answer to the question of when and why 3D projections have added value over 2D projections from a user perspective. After a quantitative analysis of 30 3D projections we find that generally, most views of a 3D projection are of relatively high quality, with only a few considerably worse views. Therefore, users should not have trouble finding one of the better views. We furthermore find that, depending on the projection technique and chosen quality metric, many single views of a 3D projection can have higher quality than a 2D projection made with the same projection technique. We perform a user study to gain more insight in how users perceive the quality of single views of a 3D projection, and whether standard quality metrics can predict whether users will deem a view to be of good quality. Most importantly, we find that the strength of the correlation between measured quality of a viewpoint and user perceived quality depends on which dataset is projected. In some cases there appears to be no correlation at all. For projections where this correlation is strong, we observe an increased benefit of using a tool that suggests high quality viewpoints to users. In general, we find that in terms of user perceived quality, a 3D projection is just as good as or better than a 2D projection generated by the same projection technique. Furthermore, we find that users believe 3D projections to better display the dataset structure than their 2D counterpart.

1 Introduction

Generally, one of the first steps of data analysis is a cursory exploration of the dataset. Ideal for this is a visual representation of the dataset in a plot or graph. This gets complicated when datasets have tens, or even hundreds of variables and many samples. One common solution is dimensionality reduction (DR) [12]. DR techniques aid in the exploration of high-dimensional datasets by projecting them to a low-dimensional space (two-dimensional (2D) or three-dimensional (3D)), while preserving distance and/or neighborhood relations between the original data points. This low-dimensional projected space can then be visualized in a 2D or 3D scatterplot, where each point corresponds to a dataset entry. Projections[1] greatly enhance the interpretability of the dataset while retaining as much information as possible. Although the projected datasets never perfectly capture the original dataset, they can still provide significant insight in its quality and structure in terms of outliers, distinguishable clusters and variable correlations. If the dataset is annotated with class-labels, a color-coded projection thereof can display whether points within a class are indeed similar, and whether classes overlap. In a machine learning context such information is a useful indicator of which classes are easy to classify and which classes are more likely to be difficult to distinguish.

In figure 1 we offer an example of a dataset (569 samples, 30 dimensions) projected to the 2D space and visualised in a scatterplot. The dataset distinguishes between two classes, this is displayed in the projection by coloring each data point either blue or orange based on the class label. It is clear immediately that there is a strong correlation between attribute values and class labels, since we can clearly see distinct blue and orange clusters with little overlap. Since we assume that most of the data structure is preserved in the projection, we can conclude that this class distinction indeed exists in the original dataset.

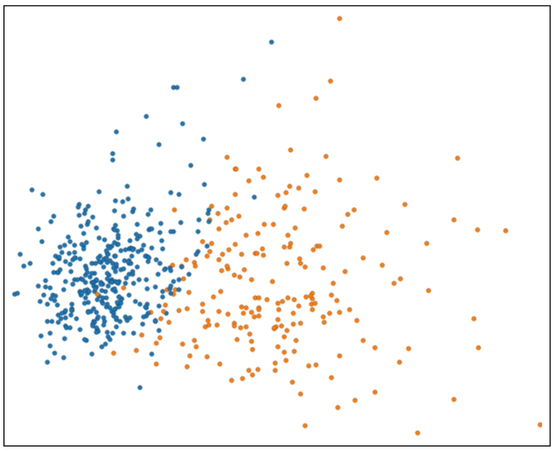

Figure 1: Example of a 2D projection of a dataset with clear structure

[1] In this document we use the term 'projection technique' when referring to a DR algorithm and the term 'projection' when referring to the resulting 2D or 3D scatterplot visualization. To disambiguate this term from a *graphical projection* of R^3 coordinates to R^2 for rendering 3D visualisations on a 2D screen, we use the words 'view' or 'rendering' for the latter.

Other techniques for the visualization of high-dimensional exist like scatterplot matrices [8] and parallel coordinates plots [16], but these are less scalable for large numbers of samples and dimensions, which makes them less interesting compared to projections. Therefore they are not the focus of this research.

1.1 3D projections

Generally projections target the two-dimensional space, which is more intuitive and easier to plot on a 2D screen. A projection targeting the three-dimensional space still needs to be rendered on a 2D screen, which means the depth dimension, aligned with the view vector, is flattened and not visible. This causes the appearance of the projection to change depending on the viewpoint. To illustrate this, take a look at figure 2, displaying two different views of a 3D projection of the same dataset shown in figure 1. The left image is very similar to the 2D projection in figure 1, and shows good separation of the two class clusters. In the right image, the projection is rotated, by changing the viewpoint we observe it from, such that the two clusters overlap completely, giving the impression that there is no correlation between attribute values and class labels. Furthermore, since all points are concentrated in a smaller area, there is more point overlap.

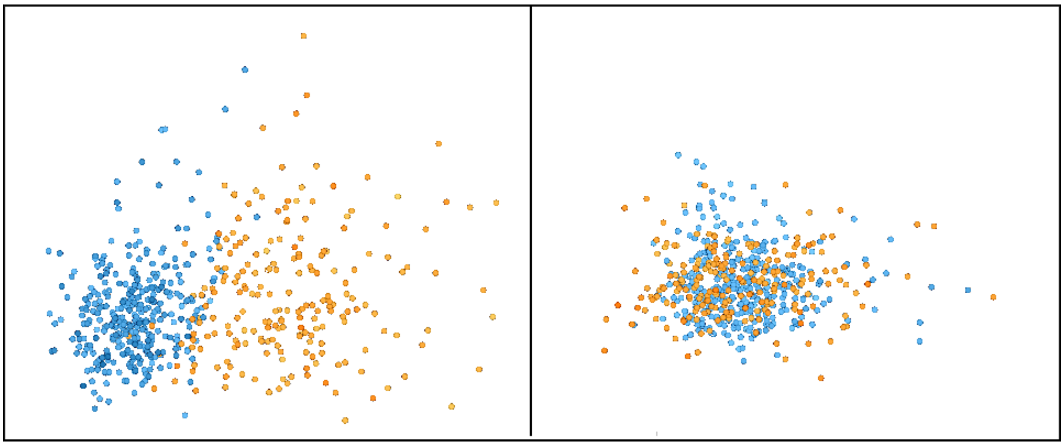

Figure 2: A good (left) and bad (right) view of a 3D projection of a dataset with clear structure

Clearly the quality of the 3D projection depends considerably on the viewpoint it is observed from, and the only way to observe a 3D projection entirely is through rotation. Something that is not required for 2D projections. Considering these simple examples, on could conclude that there is no reason to use 3D projections since, arguably, the 2D projection and a good view of the 3D projection are very similar, and both give the same insight in the structure of the dataset. Apparently, there is no gain in the 3D projection, but it adds complexity and even introduces the risk that a user does not find a good viewpoint, which could lead to false insights.

To give an intuition of a scenario where there is value in using a 3D projection over a 2D projection, we look at three other projections, shown in figure 3. Here, the leftmost image (labeled '2D') shows a 2D projection of a more complex and larger dataset with 6 different classes. In this 2D projection we can see multiple discernible point clusters. Points with similar colors are generally placed close to each other, however there is a lot of fuzziness and some overlap. In the other two images (labeled '3D-1' and '3D-2') of figure 3, two different views, of a 3D projection of the same dataset are shown. The images look different than the 2D projection. For example, we see a much clearer separation between the red and purple clusters in 3D than in 2D. (Classes D and E). Thus, the 3D projection indicates less similarity between these classes than its 2D counterpart. This difference likely occurs because in the 3D space, the projection algorithm has an extra dimension to make distinctions between points. There is more freedom to place points a structure preserving way, resulting in a clearer visualization if the right viewpoint is chosen. The fact that a 3D projection can offer different insights, raises the question of whether 2D or 3D is superior, and if so, in which situations or for what tasks this is the case.

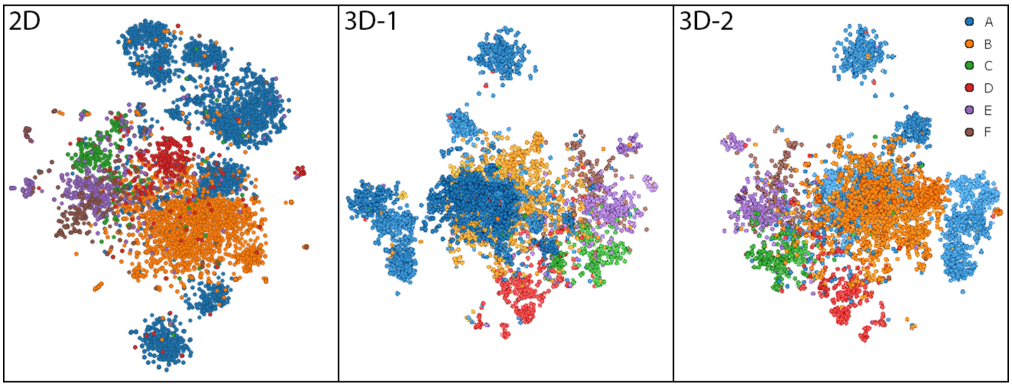

Figure 3: Example of two views of a 3D projection (two right-most images) that look different than a 2D projection (leftmost image),

To summarize, 2D and 3D projections both have advantages and disadvantages. 2D projections are more intuitive to observe on a 2D screen, and require no interaction. For larger datasets with more dimensions however, they lack space to place points. This increases projection errors, and causes many points to be plotted on the same pixel, which makes the projection harder to read.

3D projections target the 3D space, which means there is significantly more space to place points. This results in a decrease in point overlap and projection errors. Therefore, they can provide a clearer representation of the data structure. However, a user always observes a 3D projection from a viewpoint and can only ever see a 2D view (rendering) of the 3D projection. Therefore, in the eyes of the user, the 3D projection is essentially no more than a large set of 2D projections. In each view, the depth/view dimension is not visible, and point occlusion occurs depending on the viewpoint. Users are generally required to find multiple good views to analyze the entire projection, and combine these mentally into a map of the entire projection. This takes considerably more effort and might cause mistakes that lead to false insights.

Currently, many tools and techniques have been proposed that attempt to aid users in interpreting and benefiting from 3D projections (see chapter 2). Multiple recent works of literature have compared 2D and 3D projections, among which the primary inspiration for this work: A quantitative and qualitative comparison of 2D and 3D projection techniques [33]. It indicates that while 3D projections have a limited measurable added value over 2D projections, they can show more structure and motivate users to explore the data more than their 2D counterparts. It is however not clear how users pick good viewpoints in 3D projections, and how they value these viewpoints compared to 2D projections. This observation led us to the following research question.

Can we measure from a user perspective, for different projection techniques and datasets, whether, and by how much, a 3D projection is better than a 2D projection?

In this project we attempt to answer this question with the following contributions.

- We propose a tool to compare 2D and 3D projections, and display in real-time how the quality of the 3D projection changes as a function of the viewpoint.
- Using this tool, we perform a quantitative analysis of 3D projections and how they compare to 2D projections of multiple datasets, created with multiple projection techniques.
- We set up a user study showing how users pick good viewpoints in the presence or absence of our guiding tool and how they value these views compared to a 2D projection. Using this data, we test how predictive our measurements are of user perceived quality and if we can measure whether users prefer 3D or 2D.

In chapter 2 we provide a clear definition of datasets, projection techniques and quality metrics. Furthermore, we place this work in the context of other literature. We discuss projection techniques in general, visualization of 3D projections, tools that help understand 3D projections, methods to evaluate the quality of projections and studies that, like us, compare the quality and usefulness of 3D projections versus 2D.

In chapter 3 we state the objectives of this work and repeat our research question. We then split this research question into multiple subquestions.

In chapter 4, we go into a detailed description of our research. We specify how we measure the quality of individual views of 3D projections using four different quality metrics. We propose a tool that allows for a viewpoint-driven comparison of 3D projections versus 2D.

In chapter 5, we present the results of our research. We start with a quantitative analysis of projections of 6 different datasets, using 5 different projection techniques for a total of 30 projections. Secondly, we describe how we performed a user experiment and analyze its results.

In chapter 6, we discuss our most important findings and use them to give an answer to our research question. Lastly we give suggestions for future work.

For replication and verification purposes, all the code used in this project is open source and can be found in this GitHub repository. The repository also contains the datasets used in the experiment, the data gathered from the user experiment and some additional snapshots.

2 Related work

In this chapter we first formally explain dimensionality reduction (Section 2.1) and give a brief overview of different types of projection techniques (Section 2.2). We then discuss relevant literature where we highlight recent developments in dimensionality reduction. We look in more detail at how to best visualize 3D projections (Section 2.3), and how to convey more information through projections (Section 2.4). We discuss how to measure and compare projection quality (Section 2.5), and we go over previous works comparing 2D and 3D projections (Section 2.6).

2.1 Preliminaries

To help explain this project and related work we start by introducing some notations from Espadoto *et al.* [12]. Let $\mathbf{x} = (x^1, ..., x^n), x^i \in \mathbb{R}, 1 \leq i \leq n$ be an n-dimensional (nD) real-valued sample, and let $D = \{\mathbf{x}_i\}, 1 \leq i \leq N$ be a dataset of N samples. Let $\mathbf{x}^j = (x^j_1, ..., x^j_N), 1 \leq j \leq n$ be the j^{th} dimension of D. Thus D can be seen as a table with N rows (samples, elements) and n columns (dimensions, attributes, variables, features). A projection technique is a function

$$P : \mathbb{R}^n \rightarrow \mathbb{R}^q \tag{1}$$

where $q \ll n$. In this work, we consider $q \in \{2, 3\}$, the corresponding projections are denoted as P_2, respectively P_3. The projection $P(\mathbf{x})$ of a sample $\mathbf{x} \in D$ is a qD point. Projecting an entire dataset D yields a qD scatterplot, denoted as $P(D)$. The projection function P is also influenced by so-called *hyperparameters* which are typically fine-tuned by the user to optimize for specific quality metrics. The quality of a projection technique P can be gauged by several metrics defined as

$$M : \{(D, P(D))\} \rightarrow \mathbb{R}_+ \tag{2}$$

A metric M measures how well the projection $P(D)$ captures specific properties of the dataset D, the underlying idea being that a good projection will keep similar points in D close to each other in $P(D)$.

2.2 Projection techniques

Since Principal Component Analysis (PCA) was introduced [24, 15, 18], many other projection techniques have been proposed. Projection techniques can be categorized on the basis of multiple properties. First of all, projection techniques map data in either a linear or nonlinear fashion. Which one is more useful depends on whether the data is linearly or nonlinearly correlated inherently. Furthermore, a projection technique can optimize either for local neighborhoods, or globally. In local neighborhood optimization, only the K closest points to any point p are considered for computing the projection error. Therefore any structure between points that are far away from each other is disregarded. This generally results in better neighborhood preservation, but worse overall point-pair distance preservation. Lastly, some techniques offer out-of-sample quality, meaning that new observations can be added immediately to the projection, without having to recompute it entirely. For the most recent, and most extensive survey of (2D) projection techniques, refer to Espadoto *et al.* [12]. In our work, we do not target specific types of projection techniques. To the best of our knowledge, the dilemma of whether 2D or 3D is better applies to all projection techniques.

2.3 Visualizing 3D projections

A 3D projection can be visualized in a 3D scatterplot or point cloud, but finding a good rendering technique is less straightforward than for 2D since there is depth and occlusion to deal with. Picking the right visualization techniques has significant influence on the interpretability of the projection. For example, Piringer *et al.* [25] display how depth cues such as relative point size, halos and mapping color to depth allow for better discrimination of points in 3D point clouds. An image from their paper displaying this is shown in figure 4. The figure contains three snapshots of a 3D point cloud. One without depth cues (left), one with relative point size and halos (middle) and another with relative point size, halos and mapping of color to depth, such that close points are yellow and further points are increasingly blue (right).

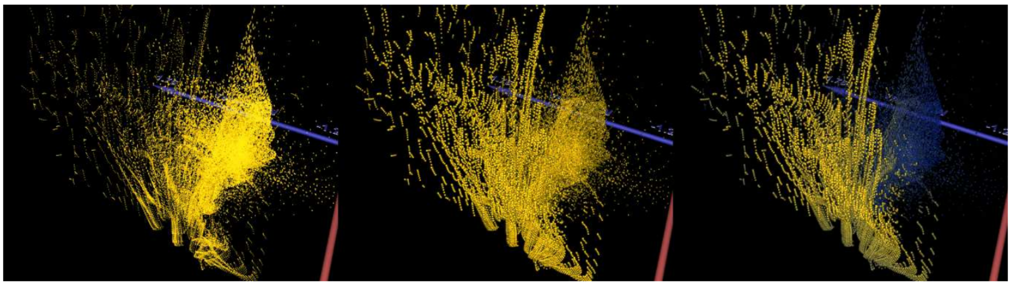

Figure 4: Figure from Piringer *et al.* [25], showcasing how depth perception can be improved. Left: No depth cues are used; Middle: Depth is indicated with point size and halos are used to ease the discrimination of single points; Right: Depth cueing using both color and point size, as well as halos

Furthermore, Sanftmann *et al.* [27] show how illumination techniques can highlight structures in 3D scatterplots with high point density, which supports the user in inferring shapes. Examples can be seen in figure 5 of a 3D scatterplot with and without illumination techniques.

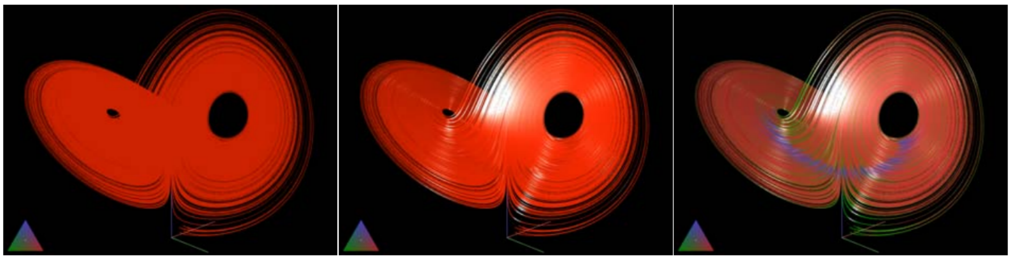

Figure 5: Figure from Sanftmann *et al.* [27] of a Lorenz attractor. Left: traditional 3D scatterplot; middle: illuminated scatterplot; right: linear, planar, and spherical structures highlighted through mapping to green, red, and blue colors respectively. The base colors are chosen to have equal intensities

Such techniques likely have some influence when qualitatively comparing 2D projections versus 3D projections, however, we believe that for the most part, the question of whether 3D or 2D is better is independent of such adjustments. Furthermore, the more complex visualization techniques are not widely used or publicly available, which means using them would make our results less relevant. We therefore only use basic visualization techniques in this work.

2.4 Explaining projections

Beyond the topic of how to render 3D projections, there is also an extensive body of work on how to aid users in exploring and understanding projections. Numerous tools have been proposed of which we will highlight some in this section.

2.4.1 Interactively linked 2D and 3D scatterplots

The increased complexity of gathering useful information from 3D scatterplots has been acknowledged in the past. Early on Piringer *et al.* [25] created a tool combining both 2D and 3D point clouds with interactive linked views to better convey information. An image of the tool can be seen in figure 6. The tool offers multiple extensions of the 3D point clouds such as rendering 3D histograms of the point density on the surface of a cube around the point cloud (bottom right widget), or displaying the principal component axes in the point cloud. Furthermore, the tool allows a user to view three 2D views alongside the 3D view, each showing an orthographic projection of the X-, Y- and Z-axis respectively. In these views a user can brush (highlight) points in square sections of the 2D point clouds, the same points are highlighted in the corresponding cube section in the 3D view. It is also possible to highlight points with attribute values in a certain range. Combining the more intuitive 2D views with the more information-rich 3D view and all extensions allows for a very thorough analysis of a 3D point system. The authors clearly argue in favor of using a third visual dimension when using the right visualisation tool, but there is no quantification of the added value, nor did they investigate how the effectiveness of the tool varies for different datasets or use-cases. Furthermore the tool was designed for 3D datasets, where the dimensions have a clear meaning, and not *3D projections* of *nD* datasets. For a 3D projection, where the meaning of the axes is lost, it would not make sense to display their pair-wise correlations in 2D views. as these correlations hold no meaning for us.

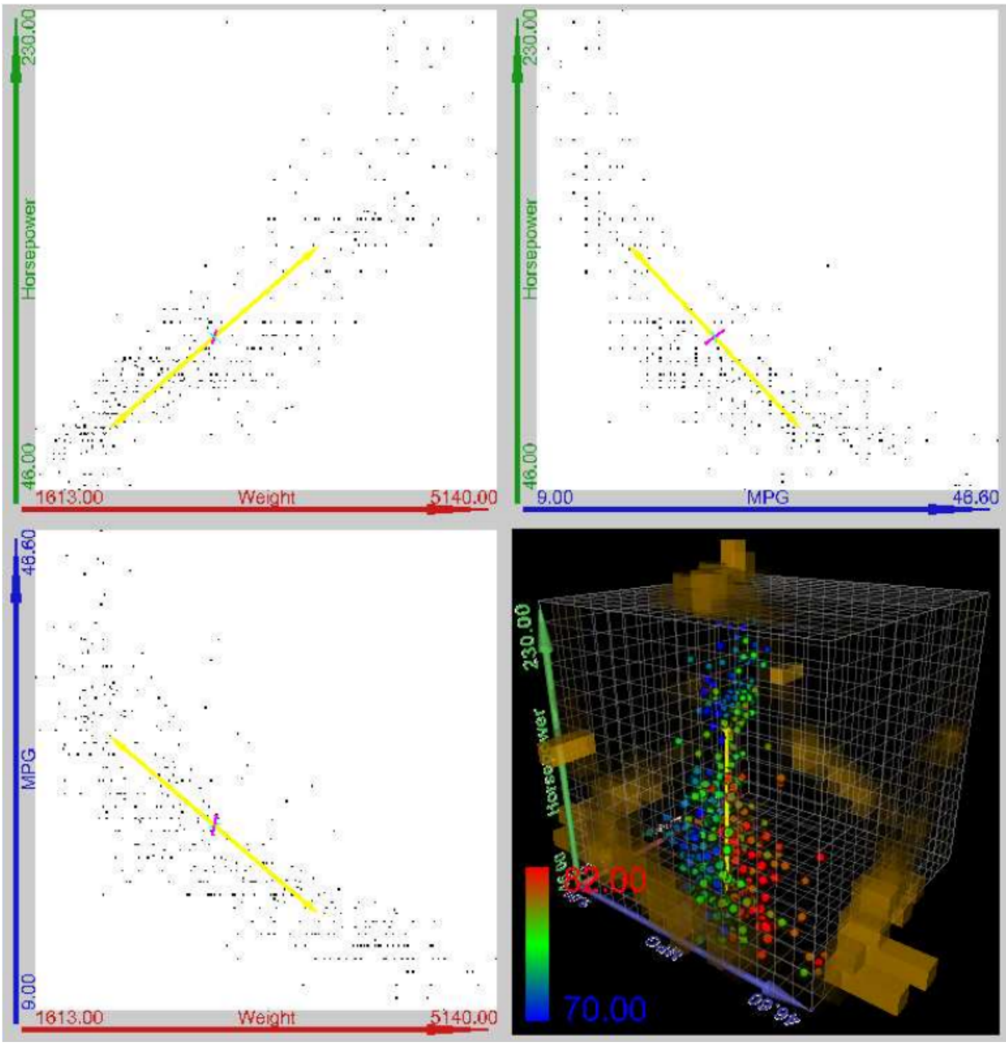

Figure 6: Figure from Piringer *et al.* [25], showcasing their tool displaying a projected InfoVis dataset. Three widgets plot all axis pairs against each other. The bottom right widget contains a 3D projection, with 3D histograms emphasizing point density. The attribute 'year' is mapped to color, the attributes 'weight', 'horsepower' and 'miles per gallon' are mapped to the axes. The yellow line is the first principle component axis, the other two axes are hardly visible as correlation occurs mostly in one direction

2.4.2 Enhanced biplots

Coimbra *et al.* [9] take a different, approach more specific to DR that involves a tool that draws enhanced biplots in the scatterplot. Standard biplots have linear axes, but often projection techniques are not linear. Furthermore, such biplots cannot show the direction and scaling of the n variables. By projecting, for each variable $j \in n$, custom data points that vary uniformly in variable j while fixing other variable values to the average observed in D, and then linking these points with a black line they create enhanced biplot axes that display the spread and nonlinearity of the projection for each distinct variable in D. An example is shown in figure 7. These biplot axes can then be used to find good viewpoints in 3D projections. If the biplot axis varies mostly in the X and Y dimensions of the current viewport, and not in the Z (view) dimension, we know that most of its variation is currently visible. This even allows for the calculation of the optimal viewing angle to display the variation of one specific variable, or the optimal view for correlation between two variables, by aligning them with respectively the X and Y axis of the viewport. A tool like this restores the connection with the original dataset dimensions in the projection and greatly enhances a users ability to draw information from it, however it requires some experience with the tool.

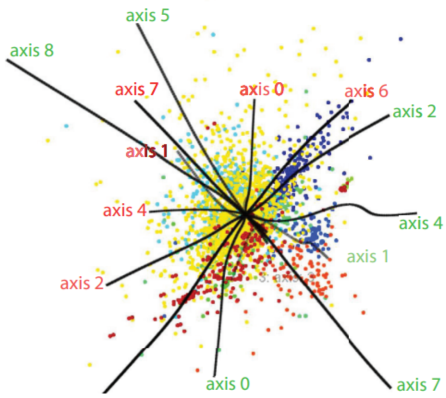

Figure 7: Figure from Coimbra *et al.* [9], showcasing how enhanced biplot axes are drawn in a 3D projection. Opposite ends of the the same axes are labeled with different colors (red and green)

2.4.3 Da Silva explanations

For projections that can not be color-coded based on class annotations (when the dataset is not annotated), it is often not clear what properties of the original data points contribute to the placement of that point in a certain position in the projection. Several explanatory techniques have been proposed to aid in the understanding of such projections. To better understand which dimensions contribute the most to similarity of points in point-neighborhoods, Da Silva *et al.* [10] proposed to color-code points based on which dimension best explained the placement of the point in that particular position. To find this dimension for each point, two techniques were proposed. The first being to rank dimensions based on the least average euclidean distance to other points in the neighborhood. The second technique ranked dimensions by the least variance in the local point-neighborhood. We show an example of the second technique in figure 8. Here an explained projection of a Wine dataset can be seen, it is clear immediately that the dataset dimensions that most influence point placement in the projection are 'Residual Sugar', 'Alcohol' and 'Sodium Chloride'. These visual explanations give insight in why points are close to each other, but they only highlight the single, best ranking dimension in the projection, whereas often one dimension is not enough to interpret the projection structure. Therefore van Driel *et al.* [38] extended this work with three additional explanation techniques to nuance the Da Silva [10] explanations. van Driel *et al.* [38] add views that color-code the local dimensionality (how many dimensions are needed to explain a significant amount of the variance in a point neighborhood) and local attribute correlations (which two dimensions are most strongly correlated in a particular region). Tian *et al.* [32] add another variation of a technique for explaining local dimensionality and provide additional examples of the usefulness of combining the mentioned visual explanations. The advantage of these explanatory views is that the connection with the dimensions of the original dataset is brought back visibly in the projection, at the cost of having a more complex tool.

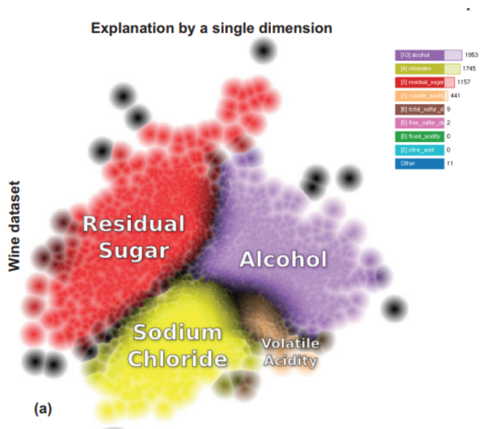

Figure 8: Figure from Da Silva *et al.* [10], showcasing how they explain projections by coloring points based on which original dimension has the least variance in the local point neighborhood. Brightness of points is reduced as the confidence decreases. (When there are few points available, or when the top ranking dimension varies much in an area.)

We have shown multiple techniques that help users in the interpretation of projections. Some work for 2D projections, others for 3D and some in both. It is good to be aware of the existence of such techniques, however they do not help us answer the question of whether 3D or 2D is better, and which is more preferred by users, which is our main topic of research.

2.5 Evaluation methods for projections

Multiple evaluation metrics have been proposed in the literature. Most of them can be placed in the following categories. (1) Metrics based on the difference between inter-point distances before and after projection, (2) metrics based on the proportion of correct local neighbors in the projection, (3) metrics for class consistency in point neighborhoods (for annotated datasets) and (4) metrics for how well clusters are separated visually. We introduce a number of common metrics that will be relevant for our research (definitions taken from [12]. These metrics are Trustworthiness (class 2), Continuity (class 2), Normalized stress (class 1) and Shepard diagram correlation (class 1). Next, we highlight some other evaluation methods.

2.5.1 Trustworthiness

Trustworthiness M_t, defined in equation 3 below, measures the fraction of close points in D that are also close in $P(D)$ [39]. Which tells us how much one can trust that clusters in a projection represent actual data patterns. Here, $U_i^{(K)}$ is the set of points that are among the K nearest neighbors of point i in $\mathbb{R}^q$ but not among the K nearest neighbors of point i in $\mathbb{R}^n$. and $r(i,j)$ is the rank of the point j in the ordered set of nearest neighbors of i in $\mathbb{R}^q$. $K = 7$ is commonly used [33]

$$M_t = 1 - \frac{2}{NK(2N - 3K - 1)} \sum_{i=1}^{N} \sum_{j \in U_i^{(K)}} (r(i,j) - K) \tag{3}$$

2.5.2 Continuity

Continuity M_c, defined in equation 4 below, measures the fraction of close points in $P(D)$ that are also close in D [39]. $W_i^{(K)}$ is the set of points that are among the K nearest neighbors of point i in $\mathbb{R}^n$ but not among the K nearest neighbors in $\mathbb{R}^q$ and $\hat{r}(i,j)$ is the rank of the $\mathbb{R}^n$ point j in the ordered set of nearest neighbors of i in $\mathbb{R}^\ltimes$. As for M_t, $K = 7$ is commonly used

$$M_c = 1 - \frac{2}{NK(2N - 3K - 1)} \sum_{i=1}^{N} \sum_{j \in W_i^{(K)}} (\hat{r}(i,j) - K) \tag{4}$$

2.5.3 Normalized stress

Normalized stress M_σ, defined in equation 5 below, measures the preservation of point-pair distances from D to $P(D)$ [23]. Any inter-point distance metric Δ^n and Δ^q can be used, but this is usually the euclidean distance. A lower stress value means better preservation of the point distances of D.

$$M_\sigma = \frac{\sum_{ij} (\Delta^n(\mathbf{x}_i, \mathbf{x}_j) - \Delta^q(P(\mathbf{x}_i), P(\mathbf{x}_j)))^2}{\sum_{ij} \Delta^n(\mathbf{x}_i, \mathbf{x}_j)^2} \tag{5}$$

2.5.4 Shepard diagram correlation

The Shepard diagram correlation M_s is the Spearman rank correlation of the Shepard diagram S [17]. The Shepard diagram is a scatterplot that plots the point-pair distances in $P(D)$ against the corresponding distances in D. The coordinates for each point can be calculated as defined in equation 6. In a perfect projection all point-pair distances scale linearly from $P(D)$ to D, which means that in the Shepard diagram all points lie on a single diagonal. The Spearman rank correlation measures to what degree this ideal correlation exists.

$$S = \{(||\mathbf{x}_i - \mathbf{x}_j||, ||P(\mathbf{x}_i) - P(\mathbf{x}_j)||)\}, 1 \leq i,j \leq N, i \neq j \tag{6}$$

2.5.5 Local error views

Arguing that general projection quality metrics only assess the overall quality of a projection, and not local quality variations, Martins *et al.* [19] improve upon other error visualization techniques to discover projection errors locally instead of globally over the entire projection. They first propose a view that color-codes points by how well distance relations to all other points are preserved. This highlights areas where the projected structure is not representative of the structure in nD. For extra insight in the more specific errors of false neighbors and missing neighbors some additional views are proposed. One view highlights points that have many false neighbors. Other views help find where the missing neighbors are of points that have many of them. Finally some views are proposed that help compare between projections created with different DR techniques. In figure 9 an example can be seen of a projection where points are color-coded based on their aggregate error with respect to all other points. Points that have many missing neighbors are connected to their missing neighbors using bundled lines that are drawn using a grayscale color map to differentiate between higher and lower errors. Thus it is easy to spot areas of points that should have been close to each other but are not. Error visualization techniques like this give more insight in whether information can be drawn reliably from local areas in a projection, which could also be a relevant factor when picking a good viewpoint $Q(P_3, p)$ for a 3D projection.

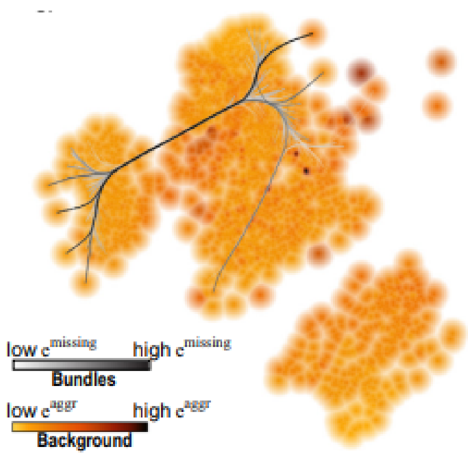

Figure 9: Figure from Martins *et al.* [19], Projection where points are color-coded based on their aggregate error with respect to all other points. Points that have many missing neighbors are connected to their missing neighbors using bundled lines that are drawn using a grayscale color map to differentiate between higher and lower errors.

2.5.6 Projection Inspector

Another tool for assessing the quality of projections is ProjInspector by Pagliosa *et al.* [22]. The tool allows users to interpolate between different projections created by different projection techniques (figure 10), and shows what the effect thereof is on the projection quality according to multiple metrics, one of which is an adaptation of Neighborhood Preservation the authors call Smooth

Neighborhood preservation. Which instead of measuring the percentage of missing neighbors measures how far the missing neighbors are away from the neighborhood of a point. Instead of capturing *whether* the projection is wrong, the enhanced metric captures *how* wrong the projection is when neighbors are missing. The tool offers an interesting approach to see the changes in quality metrics for different projection techniques and interpolated combinations of them, or for the same projection with different parameter settings. However, only projection methods that rely on user-specified control points are used, because otherwise it is unlikely that different projection techniques place points in similar locations, which would make interpolating between multiple of them useless.

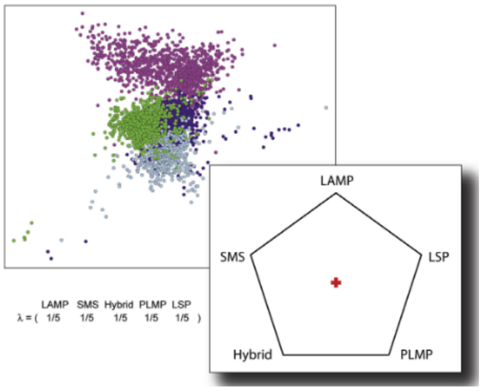

Figure 10: Figure from Pagliosa *et al.* [22], snapshot of how the ProjInspector tool interpolates between projections created by five different projection techniques

2.5.7 Visual cluster separation

In recent years, human perception of projection quality has gained increasing attention. One of the most important factors for user-perceived projection quality is visual cluster separation, which is the extent to which individual clusters in a projection are visible and discernible. Clusters are formed if groups of points are separated from other groups of points by empty space or space with lower point density in the case of unlabeled data. When projections are color-coded according to class labels the coloring of the points can also create clusters even if the point groups are not otherwise separated.

Sedlmair *et al.* [29] created a taxonomy of factors that characterize cluster separation of annotated datasets. Using this taxonomy they show how two studied cluster separation metrics fail miserably in their judgement when compared with human judgement. For over half of the projections of 75 datasets the metrics were considered to wrongly quantify class separation, and the authors show what factors cause these metrics to fail. Because of these findings a later study [6] evaluates 2002 systematically generated visual quality metrics (VQMs) using a machine learning approach. 58% of the evaluated metrics were found to better predict human judgement than the former state of the art Distance Consistency (DSC) measure [30], demonstrating the need for metrics that better capture user-perceived quality. This work inspired Wang *et al.* [40] to invent a new DR approach using simulating annealing to find projections close to the global optimum for two different human perception driven quality metrics. The used metrics are the best metric found in [6] and, for the sake of comparison, the DSC measure. Arguing that these measures lack the ability to model class density, they adapt these metrics to a density-aware version and demonstrate projections with significantly improved class separation values according to both quantifiable metrics and human judgement. While these results are impressive, the work lacks an evaluation of how these metrics affect objective quality such as neighborhood and point-pair distance preservation. In other words, while the projected points might be separated very well into clusters, it is not clear how representative the projections are of the original dataset compared to other DR techniques. The work does inspire a DR approach that searches for an optimal balance in both quantitative and perception-based metrics, however, the simulated annealing technique is linear, which means that it will have trouble projecting non-linear data, furthermore the perception-based metric is only applicable with labeled data. For this reason Abbas *et al.* [5] construct a monochrome VQM they call ClustMe, which, compared to other metrics, has significantly more agreement with human judgement in the task of ranking scatterplots on their cluster pattern complexity.

Overall however, it can be agreed that human judgement for cluster separation is an important aspect for projections that has recently gained considerable attention. In a benchmark by, Aupetit *et al.* [7] state of the art cluster separation metrics are compared in terms of how well they align with human findings. The authors argue as well for using human perception as a basis for the creation of new clustering/projection techniques. There are however, a few problems with visual cluster separation metrics that caused us to look for a different approach on measuring projection quality from a user perspective. We already mentioned that these metrics only consider how the projection looks visually, and they disregard to what extent the projection resembles the original dataset. Seeing two clearly separated clusters in the projection is useless if these clusters don't exist in the dataset. A solution could be to use a combination of the visual cluster separation metrics with some of the aforementioned objective methods, but the VQMs that require class labels can not be used for unlabelled data, and the monochrome VQM ClustMe [5] only works on unlabeled data, and will be less effective on labeled data. It is not possible to design a single VQM that similarly handles all projections, because the problem of measuring visual cluster separation changes depending on whether and how points are colored.

2.6 Comparing 2D and 3D projections

It is clear that for some contexts and some tasks 3D representations outperform 2D representations. Tavanti *et al.* [31] show that 3D can help with tasks that require spatial memory. Similarly Tory *et al.* [35] conclude that 3D views with additional depth cues like shadows are effective for tasks involving relative position estimation and orientation. Whether these findings also indicate an advantage in the context of projections is not sure, but they advocate the possibility. While in some scenarios 3D projections do not seem advantageous for 3D projections [36], other literature does suggest a benefit of 3D projections. We follow with an overview of some relevant works on this topic.

2.6.1 A Framework for Exploring Multidimensional Data with 3D Projections

Poco *et al.* [26] compare the performance of 2D vs 3D projections on an annotated dataset of scientific documents. Analysis of the projections shows that 3D scores better than 2D in the two objective metrics neighborhood hit and neighborhood preservation. They next performed a user study where 12 participants were asked to do a number of tasks: Count the clusters, order the clusters by density, list all pairwise overlaps of clusters, detect an object within a cluster and twice find the cluster closest to a specific point. The study showed that users were better able to provide the correct answer for these tasks in 3D (74.4%) than in 2D (64.3%), however the only statistically significant improvement was found in the last task of finding the cluster closest to a point. Users required around 50% more time for these tasks in 3D. Overall the work suggests a slight improvement when using 3D, but it lacks certainty.

2.6.2 Qualitative comparison of 2D, interactive 3D and scatterplot matrices for class separation

Sedlmair *et al.* [28] empirically studied the effects of different visualisation techniques on the specific task of visual cluster/class separation. Two experienced coders were asked to rate how well classes of 75 different classified datasets were separable in either a 2D scatterplot, an interactive 3D scatterplot or a scatterplot matrix. The results indicate that in many cases a 2D scatterplot is good enough to visualize separate classes. Since a 2D scatterplot allows for the easiest exploration of the data in terms of time and effort, it is therefore the preferable visualization method in most cases. In some cases however the scatterplot matrices showed better separation. The interactive 3D projection was not compared individually to the other 2 visualization methods, but an analysis was made on when it was better than both the 2D scatterplot and the scatterplot matrix. It turned out that this was only the case for highly synthetic datasets, specifically designed to give an edge to 3D projections. This work therefore provides evidence that 3D projections are in most cases not preferable to the other visualization methods for the specific task of class separation, but it is subjective since most results are the opinions of only two persons.

2.6.3 Quantitative and qualitative 2D versus 3D comparison

As the main inspiration for this project, Tian *et al.* [33] performed a quantitative study of 3D projections that discovered that certain projections show more structure in 3D than in their 2D counterparts, which could be of benefit in the visual exploration of a dataset. This study had two primary contributions. The first being to quantify the quality of a large number of projection techniques using four common quality metrics. These metrics excel at gauging objectively whether a projection technique was successful at its intended purpose, which is to preserve the structure of the dataset in the projection. However, this does not necessarily mean that a projection is useful from a user perspective. One core problem that calls for a different measuring approach is that these metrics 'see' in three dimensions, whereas a user can only look at the projection from a certain viewpoint. This viewpoint limitation means that a user can only ever see a 2D rendering of the 3D projection, which introduces problems such as occlusion of projected points and a flattening of the view dimension, which hides the true distance between points.

The second contribution was a qualitative study, showing that 3D projections either show the same structure as their 2D counterpart, or no structure at all. However, when augmented with the Da Silva [10] explanation, 3D projections can show more insights than 2D projections in terms of more separate zones explained by more data dimensions, which supports the hypothesis that with the right tools 3D projections can be advantageous.

2.7 Summary

Many projection techniques exist. Generally the 2D space is preferred for projections, because 2D visualisations are more intuitive and do not require searching for multiple good viewpoints. Some argue in favour of using 3D projections because the added dimension results in less projection errors. Furthermore three dimensions can capture more of the structure of D when its intrinsic dimensionality is more than two dimensional. However 3D projections are more complex and require interaction to explore. Because data points can be occluded in 3D, a combination of multiple viewpoints is needed to assess the entire projection, which requires the user to maintain a mental map of the projection. Albeit limited, there is evidence that 3D projections can have added value over 2D projections. [26, 28, 33]. The right tools and visualization techniques can help exploit this added value [10, 38, 32, 19], but they don't help us answer whether 3D or 2D is better or more preferred by users, and by how much. As of now, it is still unclear when 3D projections are better and how much of a difference they make. The few studies that have quantitatively compared 2D vs 3D [33, 26] show a subtle increase in quality of 3D projections, but we argue that the quality of 3D projections should not be measured in the same way as 2D projections because a user can not see in 3D. Therefore, in the next chapter we propose a different approach to measure the quality of 3D projections, that we believe to be more in line with how the user assesses quality.

3 Goal and Objectives

Now that we have introduced the topic and demonstrated its context in recent literature we will more formally define the research goals of this project. We have shown that there is debate on whether there can be an advantage for using 3D projections instead of 2D. This question has gained attention in some previous works, with both quantitative and qualitative comparisons of 2D and 3D projections. In these projects the same quantitative metrics where used for 2D as for 3D, but we believe that this approach does not give a representative view of how a user would perceive the quality of a 3D projection. After all, a user can not see in 3D whereas these metrics operate on three dimensions. This led us to the following research question.

Can we measure from a user perspective, for different projection techniques and datasets, whether, and by how much, a 3D projection is better than a 2D projection?

Innovative in this question is the user perspective part. As we have argued, measuring the quality of 3D and 2D projections in the same way is not fair, since a user observes them differently. We therefore propose a different method of quantifying the quality of 3D projections. Instead of calculating the quality of the entire projection P_3 at once, we create a uniformly distributed sample of *views* of P_3, and measure the quality of each 2D view individually, using the 2D projection quality metrics. This will tell us how the quality of a view of a 3D projection changes as a function of the viewpoint, which we believe to be more similar to how a user would evaluate a 3D projection. More details on this method are provided in section 4.1. We analyze our findings and see if patterns emerge, and whether these patterns are specific to a dataset or projection technique. Furthermore, we make a comparison to 2D projections, and see whether 2D views of P_3 can be of a higher quality than 2D projections. Thus we define the following subquestions concerning the quantitative analysis of P_3 vs P_2. After each subquestion, we reference the section that covers the answering of it.

- A1 How can we measure the quality of a view of a 3D projection? (Section 4.1)
- A2 How do quality metric values vary as a function of the viewpoints of a 3D projection? (Sections 5.2.1 - 5.2.4)
- A3 Can we find a recurring pattern in the viewpoint quality distribution for the same dataset using different projection techniques? (Section 5.2.5)
- A4 Can we find a recurring pattern in the viewpoint quality distribution for the same projection technique over different datasets? (Section 5.2.5)
- A5 Are there specific projection techniques, or specific types of datasets that score consistently better in 3D or in 2D, and if so can we find out why this is the case? (Section 5.2.6)

Naturally, the only way to confirm that our measurements indeed better capture the user perspective, is a user study. Therefore, we test if our quality metrics are predictive of the user-perceived quality. We furthermore investigate how users pick good viewpoints and whether it helps them to know the metric values. This part of the research can be broken down into the following subquestions. We again reference the sections that answer them.

- B1 Are our quantitative metrics predictive of user perceived quality? (Sections 5.3.4 - 5.3.7)
- B2 Does knowing the metric values help users find better viewpoints? (Sections 5.3.4 - 5.3.7)
- B3 Can users find single views of a 3D projection that they prefer over its 2D counterpart? (Section 5.3.8)

4 Methodology

In the previous chapter we defined the goals of this research project. In this chapter we will provide a more exact specification of how we aim to achieve these goals.

4.1 Quality measurement of 3D projection views

For this work we will use four evaluation metrics defined earlier in section 2.5, that are also used in [12]. These metrics are specifically Trustworthiness, Continuity, Normalized stress, and the Shepard diagram correlation.

Answering research subquestion A1 in chapter 3, we propose a viewpoint oriented method of measuring the quality of a 3D projection. We explain this approach now with some additional notation, on top of the notation introduced earlier in section 2.1. A 3D projection P_3 is always observed from a certain viewpoint $\mathbf{p} \in \mathbb{R}^3$. In this work viewport operations on the 3D projection are constrained, thus the distance $\Delta(\mathbf{p}, \mathbf{c})$ from $\mathbf{p}$ to the center of the projection $\mathbf{c}$ is always the same and the view vector always points from $\mathbf{p}$ towards $\mathbf{c}$. Our quality measurements are independent of rotation in the view plane, which side is up or down does not affect how the projection is perceived. Therefore defining an upvector is obsolete. Let $Q(P_3, \mathbf{p})$ denote the resulting 2D screen rendering (view, projection) of the 3D projection P_3 observed by looking at the projection P_3 from viewpoint $\mathbf{p}$. Q is defined as

$$Q : \mathbb{R}^3 \times \mathbf{p} \rightarrow \mathbb{R}^2 \tag{7}$$

A view $Q(P_3, \mathbf{p})$ is essentially a two-dimensional scatterplot, so we can use the same metrics M defined in equation 2 to gauge the quality of $Q(P_3, \mathbf{p})$ for a dataset D:

$$M : \{(D, Q(P_3, \mathbf{p}))\} \rightarrow \mathbb{R}_+. \tag{8}$$

Let $V = \{Q(P_3, \mathbf{p}_i) \,|\, 1 \leq i \leq s\}$ be a sample of s views of a 3D projection. Applying a quality metric M on each view in V yields a distribution of the quality of P_3 over all viewpoints. We denote this distribution as $S(V, M)$

We mentioned the problem of occluded points when observing 3D projections, which gives rise to the question, should we remove occluded points when computing Q? After all, a user can't see these points and will therefore not include them when assessing the quality of a view. We decide to ignore occlusion for a few reasons. For 2D projections, occlusion depends on the order in which points are drawn, Since, theoretically there is no drawing order for points in a 2D projection, it would be impossible to decide which points get ignored and which do not, and therefore the comparison to the 3D projection view where we do know the drawing order, would be unfair. Furthermore, occlusion depends on the point size, which is a parameter that users can change freely. Taking occlusion due to point size into account for the measurement of projection quality would add another degree of freedom to this study that is out of our scope.

4.2 Analysis of the viewpoint quality distribution

In order to analyze the viewpoint quality distribution $S(V, M)$ and perform the user evaluation we created a tool. This tool contains four different widgets. Two of these widgets are a 3D projection and a 2D projection of the same dataset, by the same projection technique. The projections are made to have more or less the same point size and the same scale. In the 3D projection we use depth cues inspired by the previously mentioned Piringer *et al.* [25] (Section 2). All points have dark halos, and we map color to depth by slightly brightening colors of points further away and darkening colors of closer points. We call the other two widgets the *quality sphere widget* and the *histogram widget*, which we explain in the next sections. An overview of the complete tool can be seen in Figure 11

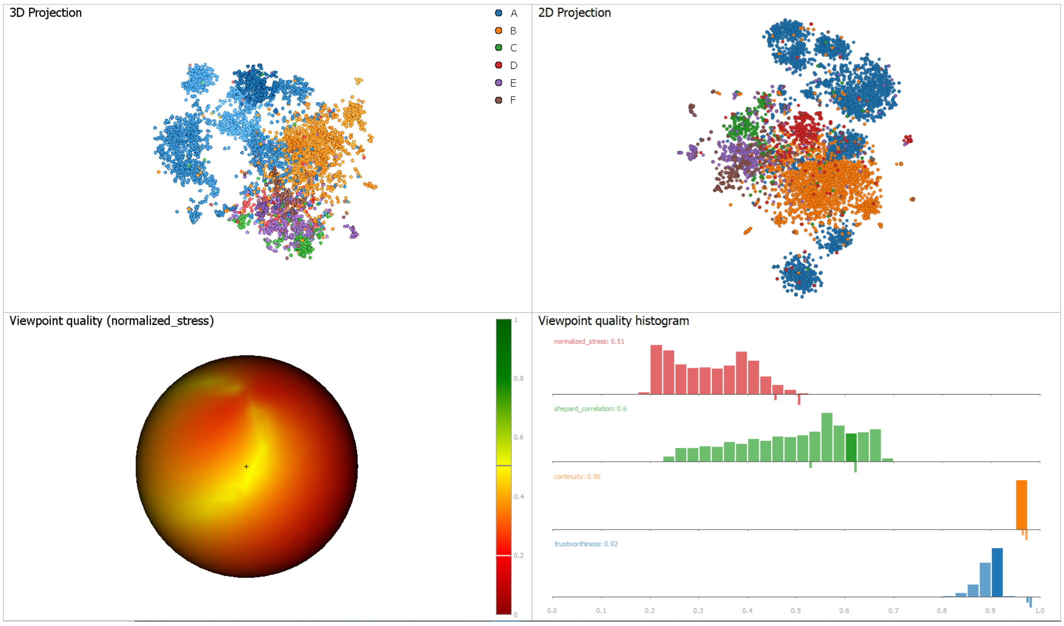

Figure 11: Image of the projection comparison tool containing four different widgets. **Top left:** A rotatable 3D projection scatterplot of D. **Top right:** A 2D projection scatterplot of D. **Bottom left:** A colored sphere showing the quality of different views of P_3 according to the selected metric. The sphere has its viewport linked with the 3D projection. **Bottom right:** Four histograms, showing the quality distribution $S(V, M)$ of 1000 different views of P_3 for four different quality metrics M

4.2.1 Quality sphere widget

We form all the views in V by generating $s = 1000$ approximately evenly spaced viewpoints on a sphere around the center of the projection using the Spherical Fibonacci Lattice algorithm [13]. Therefore, we can visualize the quality measurements (Equation 8) of said viewpoints in a 3D rendering of a sphere, its surface color-coded by the measured quality values in $S(V, M)$.

Very similar to what was done in [9]. We use a dark green color for optimal values, yellow for medium values, and dark red for the worst values. Using the crosshair in the center, targeted at the quality of the current 3D projection view, a user can quickly search for good views according to one selected quality metric, by rotating the sphere to a greener area. The rotation of the sphere is linked to the rotation of the 3D projection. Adjusting either of them affects both. The legend next to the sphere contains a black horizontal line that indicates the quality of the current viewpoint. Two white lines indicate the lower and upper bound of $S(V, M)$.

4.2.2 Histogram widget

The last widget contains four histograms, one for each quality metric. These histograms display the same quality distribution $S(V, M)$ of the views of P_3. This visualization shows directly what influence changing the viewpoint can have on the quality of the projection. We can see whether all viewpoints have similar quality, or whether there is a significant difference between viewpoints, which shows that picking a good viewpoint is crucial. Furthermore, by displaying two small ticks/lines under each histogram axis, we show the quality measured directly on P_3 (longer tick) and P_2 (shorter tick). This visualization allows us to see the proportion of views that have better or worse quality than P_2 or P_3. For example, look at figure 12, which shows a zoomed in version of the Shepard correlation histogram from figure 11. Here, the blue rectangle highlights the proportion of views with better quality than the 3D projection P_3. Similarly the orange rectangle highlights the views with better quality than the 2D projection P_2. This image shows that there are multiple views with better quality than both P_3 and P_2, according to the Shepard correlation metric. Lastly, with one slightly darker bar we highlight the quality interval that contains the current view of the 3D projection. Therefore, rotating the 3D projection affects which bar is highlighted.

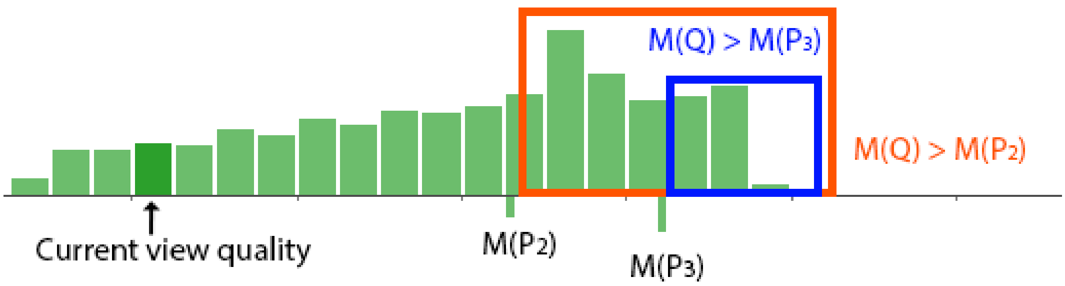

Figure 12: Visual explanation of how one glance at a histogram shows the proportion of views with better quality than P_3 (blue rectangle) and P_2 (orange triangle). The darker bar highlights the current view quality.

4.2.3 Histogram widget hovering

For the histograms we also implemented a hovering feature. Two things happen when a histogram bar is hovered over. The first is that the sphere and 3D projection widget rotate to a view that has a quality value within the hovered bar's interval. Hovering at the bottom of the bar will select a view with a quality closest to the lower end of the interval, hovering at the upper end of the bar will select the viewpoint with the quality value closest to the higher end of the interval. Thus, by moving the mouse pointer from the bottom to the top of the bar allows a user to quickly scan, in order, all views with qualities within a specific interval. Using two snapshots we show how the hovering affects the tool. In figure 13 we show a snapshot of what the tool looks like initially. The 3D projection view does not look good as there is much overlap, in the quality sphere we see the crosshair targeting a yellow area, indicating that the value for Shepard diagram correlation of the current view is quite low. In the histogram widget we see, by looking at the darker bars, that all of the metrics are at the average or lower end of their quality range. In the second snapshot, shown in figure 14, we see how the view shifts by hovering at the top end of the right-most bar in the Shepard diagram correlation histogram. The 3D projection is rotated towards the view that has a Shepard correlation value corresponding to the value that is being hovered. This is also apparent in the quality sphere, where a dark green area is now targeted. In the histograms, we see that the current view has higher quality values for all metrics.

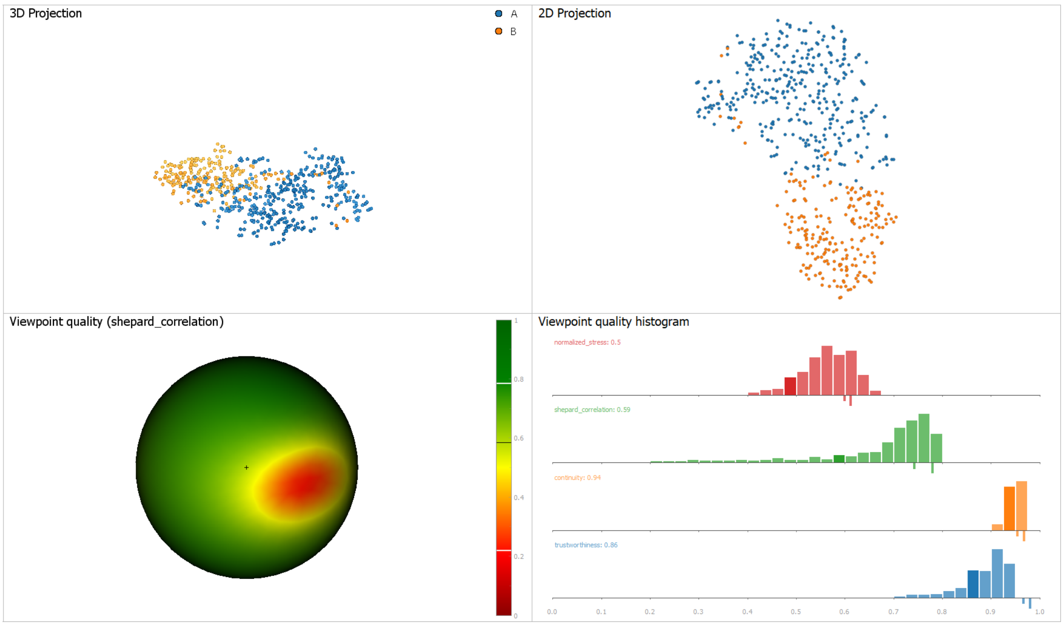

Figure 13: The first of two snapshots showing how hovering on a bar in the histograms affects the entire tool. This image shows the tool in its initial state.

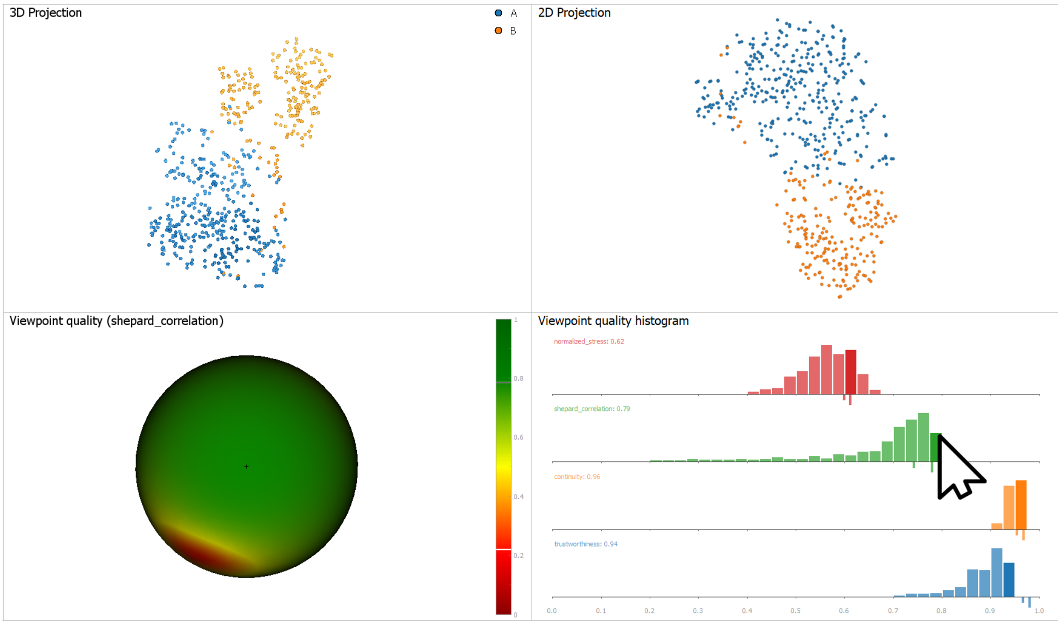

Figure 14: The second of two snapshots showing how hovering on a bar in the histograms affects the entire tool. This image shows the tool after a user has hovered on the highest end of the rightmost bar in the Shepard diagram correlation histogram

Secondly when hovering over a bar in the histogram, polylines are drawn from the hovered bar to other histogram axes, forming a PCP (Parallel Coordinates Plot). An example is shown in Figure 15. One polyline is drawn for each view contained in the hovered bar. It shows all quality metric values of this view by going through the axes of each metric. A thicker, and more opaque line highlights the currently hovered viewpoint. The PCP shows for the hovered quality interval (bar) of a metric, how quality is spread for the other three metrics. For example, the PCP in Figure 15 shows that all views with a normalized stress value around 0.53 (the red hovered bar), have very different values for Shepard correlation, very similar values for continuity, and quite similar values for trustworthiness. This visualization allows a user to dive deeper into high quality views. It is easy to find, for all views with a high value for a single metric, a view that also has high quality for one or multiple other metrics.

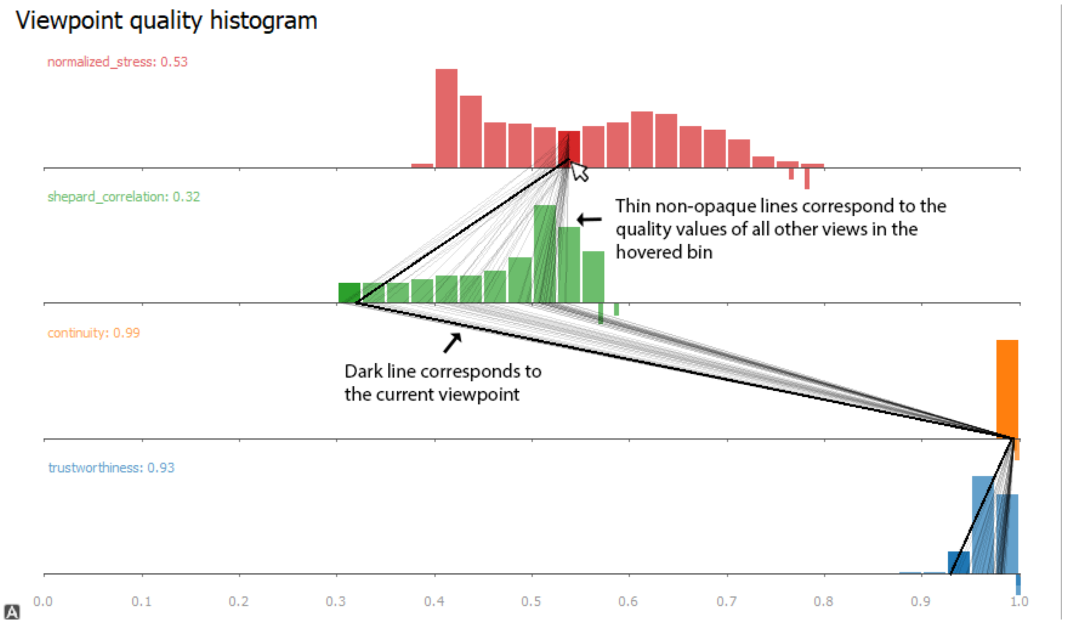

Figure 15: Image of the PCP formed by hovering over a red (normalized stress) bar in the histogram widget

5 Evaluation

Now that we have introduced the tool in the previous chapter we can use it to look for answers to our research question defined in section 3. We want to determine if our proposed viewpoint-driven approach of evaluating 3D projections can give us new insights in whether 3D or 2D is better for different datasets and different projection techniques. Therefore, we first need to make a varied selection of datasets and projection techniques to create our 3D and 2D projections. In section 5.1 we present this selection and give the reasoning behind our choices.

In order to find out if our new measurement approach can give us any new insights in how the quality of a 3D projection varies as a function of the viewpoint, we first quantitatively evaluate our projections in section 5.2. We analyze our measurements for recurring patterns and determine if these patterns are dataset or projection technique dependent. We furthermore compare how the single views of a 3D projection compare to its corresponding 2D projection.

Key in our research is the user perspective. We have proposed a new way to measure the quality of 3D projections that we believe to be more in line with how a user perceives them. We therefore perform a user experiment, described in detail in section 5.3, that has the primary purpose of correlating our quantitative measurements to user perceived quality, so that we can better reason about how quantifiable properties influence whether a user prefers a 3D or 2D projection, and determine what causes this. Furthermore, we test if the guiding widgets of our tool can help users exploit the value of 3D projections by suggesting higher quality views.

5.1 Choice of datasets and projection techniques

To create the 2D and 3D projections used in our experiments, we needed a number of datasets and projection techniques. The total number of projections had to be manageable for manual inspection, but large enough to find generalizable results. Ultimately, we used 6 different real-world datasets, and 5 different projection techniques yielding a total of 30 2D and 3D projection pairs.

The datasets we chose and their characteristics can be seen in table 1. They are a subset of the datasets used in [33], and chosen for their differing characteristics in terms of sample count N and dimensions n. For each dataset we specify its size in terms of samples and dimensions, and we specify whether they contain class labels. For the labels, we make a distinction between ordinal and categorical classes, since these will have different color coding in the projections. (e.g. a categorical color map vs an ordinal color map.)

Dataset	Samples N	Dimensions n	Class labels
AirQuality[11]	9357	13	-
Concrete[1]	1030	8	Ordinal
Reuters[2]	8432	1000	Categorical
Software[21]	6773	12	Ordinal
Wine[3]	6497	11	Ordinal
WisconsinBreastCancer[4]	569	30	Categorical

Table 1: Selected real-world datasets and their characteristics.

The projection techniques selected for our experiment are displayed in table 2. Again, we chose a diverse set of projection techniques in terms of linearity (nonlinear vs linear), input type (samples or point-pair distances) and whether they optimize for local or global neighborhoods. This ensures that our results are not specific to any kind of projection technique. Within these constraints, we favored more popular techniques.

Projection	Linearity	Input	Neighborhood
AE [14]	nonlinear	samples	global
MDS [34]	nonlinear	distances	global
PCA [18]	linear	samples	global
T-SNE [37]	nonlinear	distances	local
UMAP [20]	nonlinear	distances	local

Table 2: Selected projection techniques and their characteristics

5.2 Quantitative measurements

We start our analysis of all 30 projection pairs with an an image displayed in figure 16 that provides a visual overview of all our measurements using snapshots from the widgets of the tool. The figure displays a table, with a row for each projection, ordered primarily by dataset and secondarily by projection technique. Each row displays two snapshots of the quality sphere for each quality metric. These two snapshots are taken from two opposite viewpoints (chosen arbitrarily), so that nearly the entire sphere can be seen. Since there are four quality metrics, there is a total of 8 sphere snapshots. Next to these sphere images we display the corresponding histograms of each metric. These histograms display the quality distribution of the $s = 1000$ uniformly spread viewpoints in V. Therefore, they show how much the quality of the view of the 3D projection can change depending on the viewpoint, according to the four quality metrics. In this figure, the columns are named either T, C, S or N, for the respective quality metrics Trustworthiness, Continuity, Shepard diagram correlation and Normalized stress. The image showcases lots of interesting results, which we will next discuss.

5.2.1 Metric signal ranges

Arguably, one of the most notable observations is that the measurements for the metrics trustworthiness and especially continuity have a small range very close to the maximal value of 1, indicating that regardless of the viewpoint, the quality values are high. This is shown in almost all the spheres where only green or dark green colors are visible, and in the blue and orange histograms that are only one or a few bars wide. It would seem that these metrics can not be used as indicators of good viewpoint quality, since according to these metrics, all viewpoints are good. Whether that is truly the case depends on whether a small change in metric value in fact can denote a significant change in viewpoint quality. It could be that the metrics trustworthiness and continuity simply have a smaller effective signal range, but that relative within this range changes are just as significant as for the other metrics. To test this theory we visually compare views with the highest quality, and views with the worst quality. We do this for two projections. The T-SNE projection of the Wisconsin Breast Cancer dataset, which is a very clear projection, and the the PCA projection of the Airquality dataset, which is harder to read. The comparisons are shown in Figure 17. In this figure the first row shows the best view versus the worst view for both metrics for the Wisconsin Breast Cancer projection, the second row shows the the second-best versus the second-worst view. The third and fourth rows show the same information for the Airquality dataset. For each view we display the metric values rounded to two decimals.

In both figures we see the same trend. The views with maximal measured quality show significantly more of the point spread, whereas in the the worst views there is more overlap and and points are compressed in smaller spaces. This is the case even though the metric values only differ slightly. For example, the continuity value of the Airquality projection only differs by 0.02 between the best and worst view. The largest difference is 0.22 for the Trustworthiness metric of the Wisconsin Breast Cancer projection.

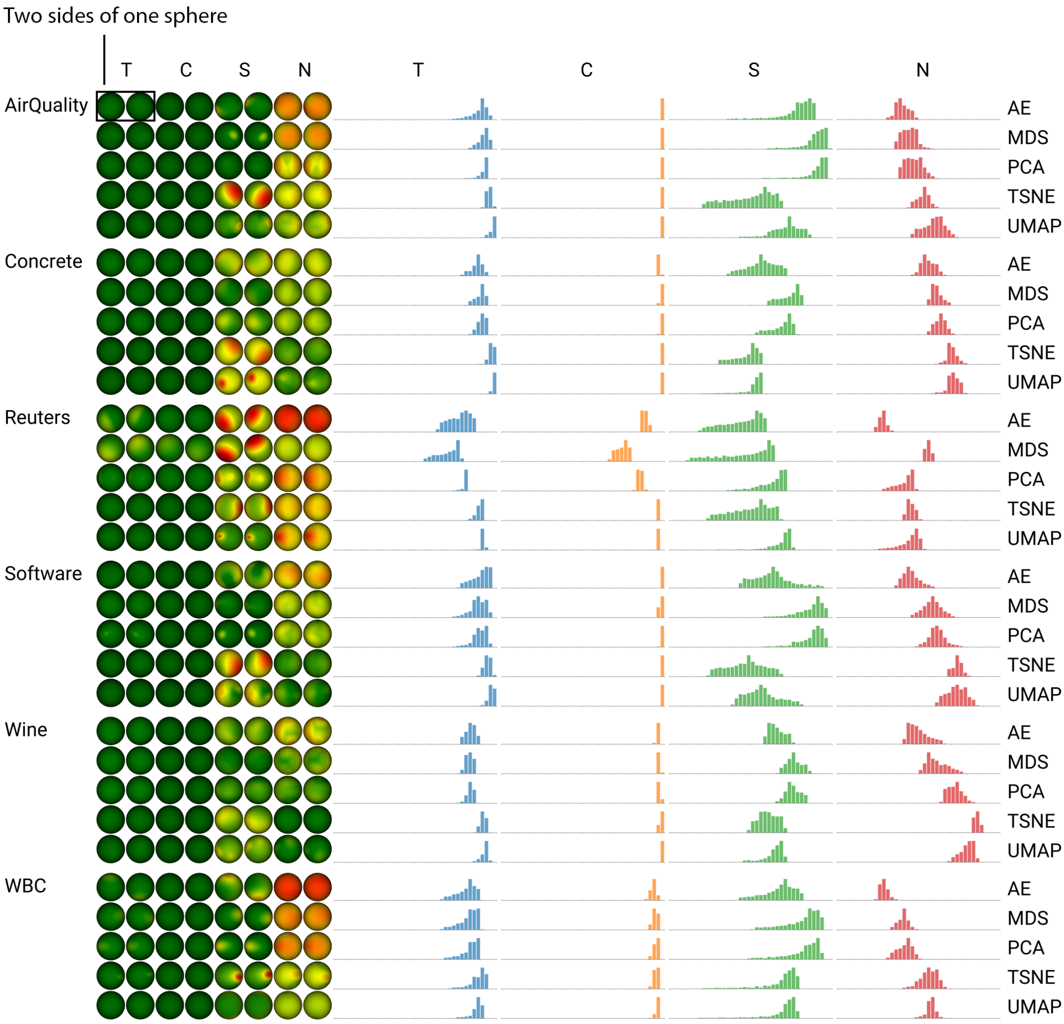

Figure 16: This figure contains images of the quality sphere for each metric, for each dataset, for each projection technique. For each sphere, two snapshots are taken from an arbitrary viewpoint, and its opposite viewpoint. Hence there are 8 sphere images for each dataset and projection technique configuration. Next to the spheres the corresponding viewpoint quality distribution histograms are shown. The letters T, C, S and N indicate the columns belonging to metrics Trustworthiness, Continuity, Shepard correlation and Normalized stress.

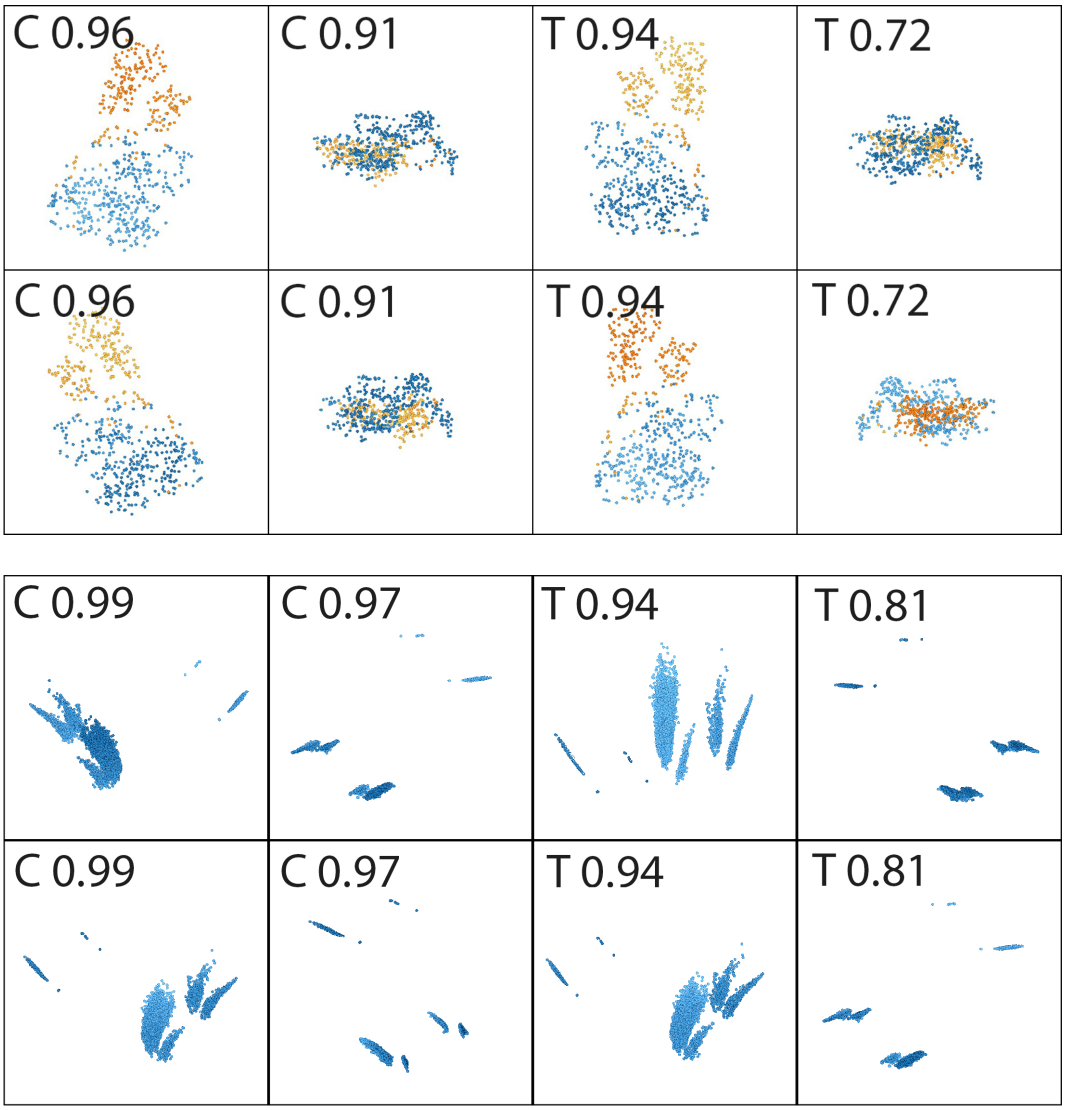

Figure 17: Comparison of the best and worst views of the Wisconsin Breast Cancer dataset T-SNE projection (top), and the Air Quality dataset PCA projection (bottom) according to the metrics Continuity (C) and Trustworthiness (T). For each dataset, the first row shows the best and worst views for both metrics, and the second row shows the second-best and second-worst views

We've only shown images of two projections, but we found this trend in the other projections as well. Clearly, these small differences in metric values can have significant impact on the view quality, so it stands to reason that we should consider these differences in our analysis. We therefore recreate figure 16, but change the axis bounds of each histogram to the minimum and maximum values observed within it, which essentially zooms in on each histogram. In this representation we lose the ability to fairly compare values of one metric with values of other metrics but also with values of the same metric in different projections, but we can better evaluate patterns for the metrics that have a condensed signal range. The result is shown in Figure 18.

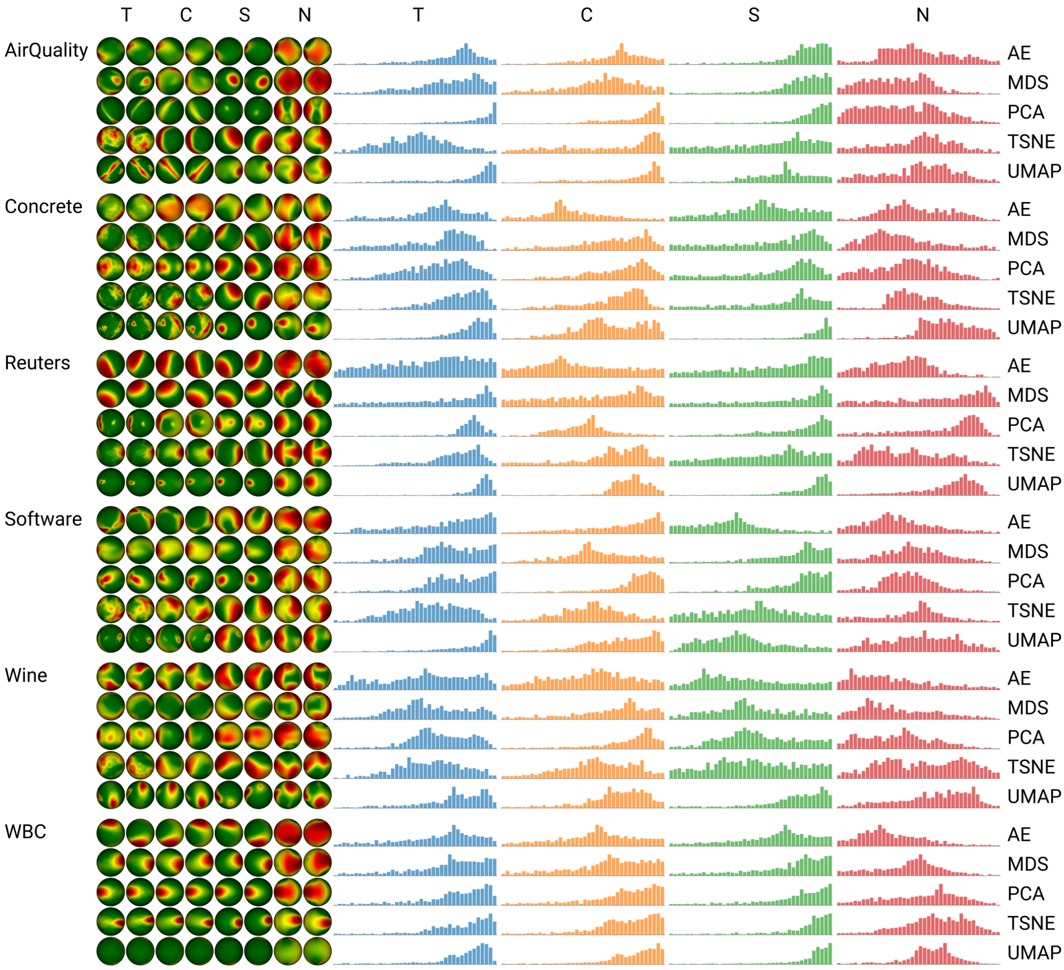

Figure 18: Adaptation of Figure 16 with axis bounds set to the minimum and maximum metric values for each individual sphere and histogram, to allow for better reasoning about patterns

Using both figures 16 and 18, we can search for an answer to the research subquestion A2 (Section 3). *How do quality metric values vary as a function of the viewpoints of a 3D projection?*. First of all we see how important it is to not measure the quality of a 3D projection as a whole, since its quality depends significantly on the viewpoint. Furthermore, we conclude that all metrics have a different effective signal range. The metric values for continuity are generally close to the maximum of one, regardless of the viewpoint, whereas the metrics

Normalized Stress and especially Shepard Diagram Correlation cover a large portion of the signal range. However, we've also seen that small changes in the metric value for continuity or trustworthiness have significant effect on the view quality, so we can not conclude that all views are good according to these metrics. Because the significance of a specific difference in metric value is relative to the metric, and probably also to the projection, we can disregard the signal range of 0 to 1, and only consider the effective signal range determined by the minimum and maximum values, as shown in figure 18.

5.2.2 Patterns in the view quality distribution

In figure 18, the spheres give a good intuition of how the quality of the views is distributed. If a sphere is mostly green, it means that most of the viewpoints have a quality value at the higher end of the signal range. This means that it is very easy to find a higher quality viewpoint. A good example here is the UMAP projection of the WBC dataset. The spheres for each metric are almost entirely green. (Note that there are still yellow and red areas at the edges that are barely visible due to the chosen viewpoint.) We see why this is the case by looking at the histograms, that have tall bars on the higher end of the effective signal range and a tail of low bars on the left. There are few cases where the opposite is true, and the spheres are mostly red. Generally, the worst we observe is that the quality values are spread more or less evenly. Only in a few cases do we observe that there are more bad viewpoints than good viewpoints. For example for the normalized stress metric in the Airquality MDS projection, or the Shepard Correlation metric in the Software AE projection. These findings indicate that for most projections, a user should not have a problem finding a high quality viewpoint, countering the argument against 3D projections that their value can be lost because the user might not find a good viewpoint. However, we do not yet know if users also perceive these views to be of good quality. Therefore we test whether there is a positive correlation between high metric values and good viewpoint quality according to users later in section 5.3

5.2.3 Similar patterns across metrics

If all metric values are similarly correlated with viewpoint quality, then one would expect the histograms of different metrics to be of similar shapes. After all, if most viewpoints are good, or if most viewpoints are bad, the metrics should all similarly reflect this. Looking at the histograms in figure 18, we find that in most cases the shapes of the histograms are similar, albeit sometimes they are a bit shifted. Because of the individual scaling of the histograms, this can happen very easily if there are a few outlying values. Only in a few cases do we see large differences. For example in the Airquality PCA projection (figure 18, third row), we see that the T, C and S histograms have very similar shapes, with a thin tail on the left, and a peak at the right end. Whereas the N metric is very evenly spread, with a small tail on the right end. Observing this difference raises the question of which metric is 'right', because there must be cases where normalized stress considers a viewpoint bad, and the other metrics consider it good. This is a difficult question to answer because each metric measures in its own way the preservation of the original dataset structure, and it depends on which properties of the original dataset a user deems more important to be preserved. One answer could be to measure which metric has a stronger correlation with user perceived viewpoint quality which, again, is covered in section 5.3.

5.2.4 Similarity of opposite sphere images

Another observation of note in figure 18 is that for all metrics, both images of the sphere are very similar. This is a logical result considering that the snapshots are taken from opposite viewpoints. Because the view vector is flattened in the 2D rendering of a 3D projection, two views from opposite viewpoints should be mirrored and equal in terms of distances between points. The only difference between these views lies in the tilting of the camera, and the order in which points are drawn on the screen. These things are not considered by the metrics. The opposite shapes on the spheres differ slightly because the Spherical Fibonacci Lattice algorithm [13] we used to generate the viewpoints does not ensure that each point has an opposite.

5.2.5 Patterns inherent to datasets or projection techniques

To answer the research subquestions A3 and A4 (Section 3), we search for patterns specific to a dataset or projection technique. Here we have to conclude that, in terms of histogram shapes and sphere patterns, there is no discernible pattern specific to any dataset. Therefore, any other findings we have are likely to be applicable to projections of other datasets as well.

In order to search for projection technique specific patterns we create another adaptation of figure 16, where the rows are sorted first by projection techniques, and secondly by dataset, such that all projections by the same projection technique are grouped together. This image can be seen in figure 19. This figure helps us to make the observation that the projection techniques PCA and especially UMAP tend to be more peaked. Many of their histograms have many similar values in a small part of the signal range, and long tails consisting of a few outlying values. The spheres reflect this by having small red dots for the few worst views, and being mostly green otherwise. The conclusion we can draw from this observation is that, according to the quality metrics, most views of these projection techniques are quite similar, with a few views that are considerably worse. So, for these projection techniques it should be easy to find one of the better viewpoints in terms of measured quality. However this does not imply that these projection techniques are better than the others, since we do now know how the actual quality values compare. Aside from this, there do not appear to be any patterns specific to projection techniques. Figure 16 shows that there are projection techniques that tend to score higher than others for certain metrics, but this is not a pattern within the projection technique itself.

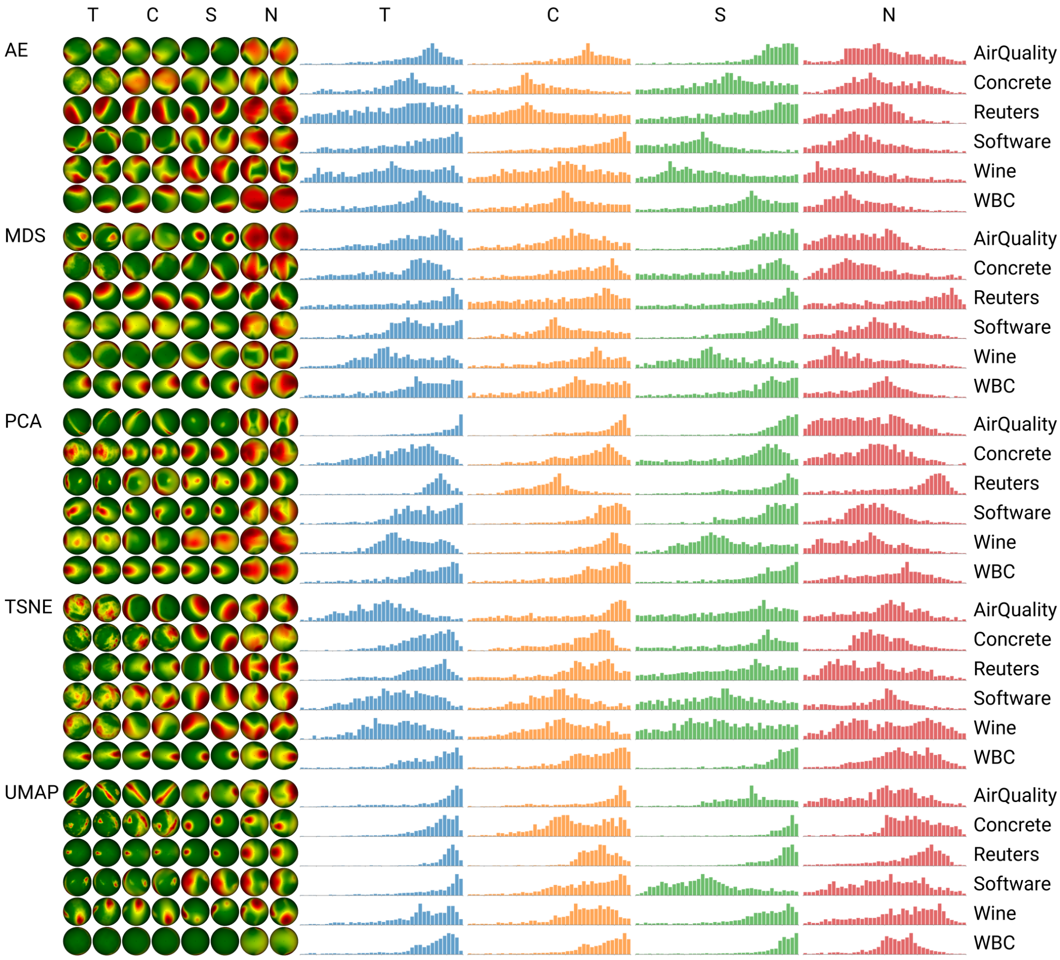

Figure 19: Adaptation of Figure 18 with rows ordered first by projection technique, and secondly by dataset.

5.2.6 Comparison of 3D versus 2D

Answering research subquestion A5 (Section 3) is not easy. Like Tian et. al. [33], we find that directly measuring the quality of the 3D projections, instead of its views, generally results in higher metric values than the 2D projections. In table 3 we show the average quality values for both the 2D projections and the 3D projections. We see for each metric that on average the 3D projection scores slightly higher.

	T	**C**	**S**	**N**
2D	0.945	0.961	0.732	0.556
3D	0.968	0.974	0.780	0.603

Table 3: Average metric values over all 2D projections and over all 3D projections for the metrics Trustworthiness (T), Continuity (C), Shepard diagram correlation (S) and Normalized stress (N)

However, we have argued that it is unfair to directly compare quality measurements of 3D projections with 2D projections, because users can not observe them in the same way. What we can find out is how the quality values $S(V, M)$ of the views of the 3D projection compare to the quality of the 2D projection. After all, both are essentially 2D point clouds so they can be compared fairly. In figure 20 we show a stacked bar plot that shows for each dataset, for each of the metrics, and through color-coding for each of the projections, the percentage of views (from the $s = 1000$ generated viewpoints) that score higher than the 2D projection. Because we stack the bars of projections by 5 projection techniques on top of each other, 20% corresponds to the number of views (1000) of a single projection technique and dataset pair.

The image contains a lot of interesting information. For the metric Trustworthiness (T), there are only a few cases where a single view of the 3D projection outperforms the 2D projection, but for all other metrics we see that many views have higher quality scores than the 2D projection. In some cases (For example, the Normalized stress metric for the Airquality dataset), more than 50% of the views of the 3D projections have higher quality than the corresponding 2D projection. Furthermore, for all datasets and all metrics except Trustworthiness, we see that the bars generally consist of multiple, differently colored, stacked bars. This means that for many projection techniques, their respective 3D projections have a considerable amount of views that score better than their 2D counterpart. Whether this is evidence in favor of preferring 3D projections over 2D projections is arguable. On the one hand one might say that, if there are single views of the 3D projection that are, on their own, better than the 2D projection, then looking at that view alone should give users better information than looking at a single 2D projection, and that is not even considering all the other vantage points that could give even more insights. On the other hand, there are still many views that are worse, and a user might prefer to look at these, which could lead to less reliable insights. Here, a tool like ours could make a significant difference by helping users to only consider the views that score better than 2D projection.

To give a definitive answer to research subquestion A5 (Section 3), we do not see a specific dataset, of which all projections score consistently better in

3D. We do observe however that some projection techniques score consistently better in 3D, for specific metrics. For example, for all but one of the AE projections (blue bars), close to all of the views (20% in the figure) have better normalized stress values than the 2D projection. One could argue that if every single view of the 3D projection has better quality than the 2D projection, then 3D is without a doubt better. Because no matter what the vantage point is, the view shows a better representation of the dataset than the 2D projection. Following this reasoning, we find that for AE projections, and considering only the Normalized stress metric, 3D is better for all but one of the datasets. For PCA and T-SNE (green and red) we see considerably less 2D views that outperform the 2D projection in terms of Normalized stress, this indicates that in terms of Normalized stress these latter projection techniques improve less in 3D, than the others. Of course, since this finding is relative, these 3D projections could still score higher than the AE projection, but there appears to be less reason to pick 3D instead of 2D for them. Following the same reasoning, it appears to only be beneficial to use 3D UMAP projections (purple bars) when it comes to Shepard Correlation or Normalized stress, but rarely for Continuity and never for trustworthiness. Similarly, for all MDS projections (orange bars) the views outperform the 2D projections mostly in terms of Continuity. Because we see these differences, we can conclude that it likely depends on the projection technique, and the metric, how much improvement, if any, one can expect from using a 3D projection instead of 2D.

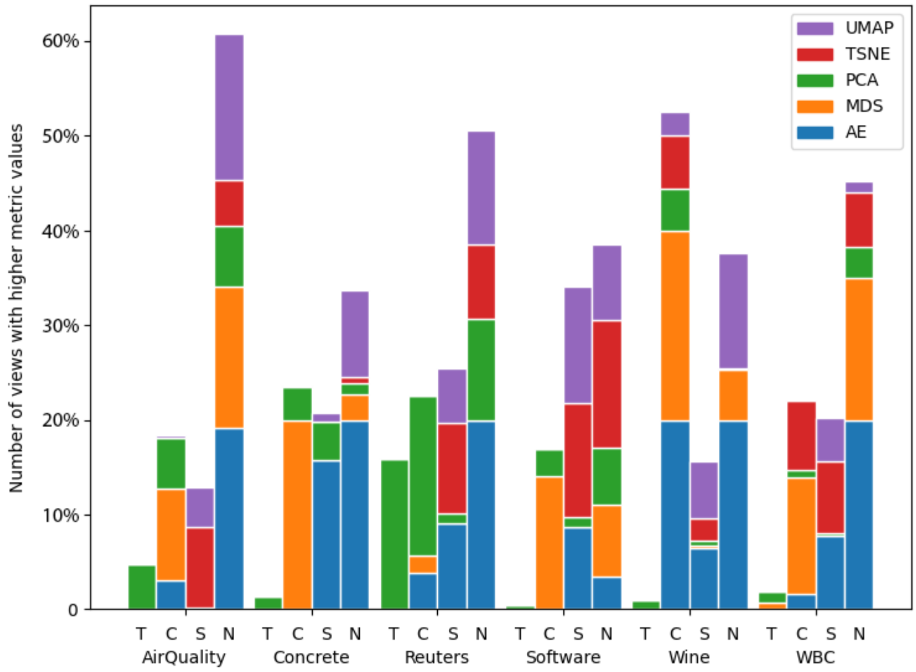

Figure 20: Stacked bar plot displaying for each dataset and for of the metrics Trustworthiness (T), Continuity (C), Shepard diagram correlation (S) and Normalized stress (N), the percentage of views of the 3D projection, that score higher than the measurement of the same metric of the corresponding 2D projection. For each projection technique, a different color is used and the corresponding bars are stacked on top of each other. Since there are five different projection techniques, and $s = 1000$ views per configuration, 100% corresponds to 5000 views.

5.3 User evaluation

In our quantitative analysis we have seen how the quality metrics change for different views of a 3D projection, and we have seen that small changes for one metric can be just as significant in terms of view quality as large changes for another metric. As of now however, we are not sure to what extent our quality metrics align with user perceived quality of a projection. We have seen for example, how projections of different datasets can have similar quality values, but these projections will likely be very different in the eyes of the user. Overall, on multiple occasions we have seen that we need to take into account how users reason about quality of views before we can draw any conclusions on how effective quality metric values are to characterize the user perceived quality of a viewpoint. Therefore, we performed a user study with the specific purpose of testing how quality metric values relate to user perceived quality in terms of single views of a 3D projection. With this user study we specifically target the research subquestions B1-B3 (Section 3). For our user study we selected a subset of projections. This process is described in section 5.3.1. We next describe and motivate how we set up the experiment in the section 5.3.2, and in the rest of this section we discuss the results.

5.3.1 Evaluated projections

To limit the duration of the user study, we manually selected a subset of the 30 2D and 3D projection pairs introduced in 5.1. These projections were selected based on three criteria. (1) They should have some discernible structure, e.g. distinguishable point clusters with similar coloring based on the class labels. (2) Finding a good view, with strong visual cluster separation for the 3D projection should not be trivial. (3) The datasets should be of sufficient size so that the added space of a third dimension can have value. The chosen configurations are specified in Table 4.

Dataset	Projection	Samples N
Wine	TSNE	6497
Wine	PCA	6497
Concrete	TSNE	1030
Reuters	AE	8432
Reuters	TSNE	8432
Software	TSNE	6773

Table 4: Selected dataset and projection technique configurations to create the 2D and 3D projection pairs used in the user evaluation

5.3.2 Task definition

We will now describe what we asked users to do in this experiment. The exact evaluation procedure that users received is given in the appendix (Section 7). The primary purpose of this evaluation was to discover how users reason about the quality of views of a 3D projection in comparison to a 2D projection, and how their reasoning correlates to our metric values and findings in section 5.2.

We therefore had to let users use our tool, to see to what extent their judgement of the viewpoints was in line with the measurements of the tool. Therefore, we first explained our tool to users and described what the different widgets were for, and how they could use them, as described in section 4.2. To understand the widgets, users needed some intuition of what our quality metrics mean. We kept this explanation very basic, because deep understanding of the metrics is not required to use the tool and this also decreased the required experience so that we could reach a broader group of participants. We told users only that the metric values range from 0 (worst) to 1 (best), that Trustworthiness and Continuity measure neighborhood preservation, and that Normalized stress and Shepard diagram correlation measure distance preservation between point pairs. After that, we described in detail what each of the four widgets of our tool display and how they can be used.

Since we wanted to know what viewpoints are deemed good by users, the main aspect of the experiment was for users to search for good viewpoints of the 3D projection. We therefore first needed a definition of what a 'good' viewpoint was, that a user could read and understand. We ended up describing a good view as a view that has *visually well-separated point groups, that have similar colors internally.* This means that indirectly, we asked users to find views that have minimal overlap for different clusters and show most of the interesting structure in terms of class separation.

The widgets of our tool tell users what good viewpoints are according to measured quality metrics, however we are also interested in what users themselves think are good viewpoints if they are unaware of the measured quality, since this removes bias in the results. We therefore decided to show them half of the projections in table 4, without access to the guiding widgets, and the other half with access to these widgets. This allows us to see how knowledge of the metric value of the observed view and the metric value distribution affects the decision making of the user. We let users go through these projections one by one, and asked them to select 3 different viewpoints that they deemed good according to the definition we just gave. For the first three projections, the sphere and histogram widgets were invisible, and for the second three projections users could use the sphere and histogram widgets to guide them in selecting a good view. This means that for the last three projections, they could easily find views that had high quality values for any or multiple chosen metrics. We stressed here that the users should not rely solely on the high quality values, but ultimately use their own judgement for picking good viewpoints. We did this to remind users that the quality metrics don't measure class and cluster separation, but only structure preservation. Using these widgets, users were motivated to look mostly at high quality views until they found one that appealed to them. From there they could make slight adjustments if they so preferred. We randomized the order in which users saw the projections, so that the projections seen with or without the widgets differ for each user. Because we wanted some indication of how the 3D projection compares to the 2D projection in the eyes of the users, we asked them, each time they found a good viewpoint, whether, *considering only the current view of the 3D projection*, they preferred the 3D or the 2D projection.

Our evaluation ultimately left us with, for each projection pair in table 4, a list of three viewpoints per user that they deemed good. For each viewpoint we stored the quality metric values, whether users preferred the view of the 3D projection over the 2D projection, and whether users had access to the guiding widgets or not when choosing the viewpoint. Furthermore, at the end of the evaluation we asked users to rank their agreement with the following statement on a 7 point scale. *Considering the entire 3D projection, instead of just one viewpoint, the 3D projection better displays the data structure than the 2D projection.*

5.3.3 Result analysis

22 participants responded to our user evaluation. This means that per projection we have 66 viewpoints that users deem good. Due to the random order in which users saw the projections, approximately half of these viewpoints were selected by users that did not have access to the guiding widgets, and for the other half users did have access to guiding widgets. For future reference we call these sets of viewpoints the *blind set* and the *guided set*, respectively. We call the union of these sets the *combined set*. To get a feeling for what the users saw, we show snapshots of an arbitrary selection of these views in section 7.2 of the appendix. Our GitHub repository contains additional snapshots of all views grouped by set and specified preference over the 2D projection.

For an initial analysis of the results, we created figure 21. It contains the histograms displayed in the tool for each metric and each projection pair in the evaluation set. Since we found in our quantitative analysis that the significance of small changes in metric values depends on the metric and projection, we again zoom in on the histograms by using the histogram signal bounds as the axis bounds. We would like to note here that during the user evaluation, we did not use the zoomed in versions of the histograms in our tool. Therefore, it might be that users paid more attention to metrics with a larger signal range. This is something that could be addressed in future work. Below each histogram three box plots are drawn labeled with, in order from top to bottom, 'histogram', 'users-blind' and 'users-guided'. The boxplots summarize the distribution of, respectively, the quality values $S(V, M)$ viewpoints that make up the histogram, the quality values of the blind set and the quality values of the guided set.

5.3.4 Do users prefer higher quality viewpoints?

Figure 21 allows for an easy analysis of how the means and spread of the different viewpoint sets compare to each other. By looking at the red lines in the boxplots, corresponding to the mean metric values, we observe that except for a few occasions, the viewpoints in the blind set are of higher average quality than the histogram average. So it appears that users tend to prefer viewpoints with above average quality. The viewpoints from the guided set are usually of even higher quality, so, as might have been expected, our guiding widgets cause users to pick even higher quality viewpoints. Before we can draw any conclusions from these observations, we have to test whether these differences are indeed statistically significant. Therefore we perform a T-test (equal variance, one tail) for each projection, for each metric and for each of three sets of user selected views. We display the p-values in table 5. Here, each significant p-value ($p < 0.05$) is displayed in bold. We see that for nearly all projections and metrics, the views from the guided set have a significantly higher average quality than the histogram average. For the combined set this is slightly less often.

Figure 21: Zoomed in histograms corresponding to all projections used in the user evaluation. Under each histogram boxplots are drawn for the quality of, in order from top to bottom, **1)** the 1000 viewpoints in the histogram, **2)** the viewpoints from the blind set, **3)** the viewpoints from the guided set.

	T			C		
	Blind	**Guided**	**Comb**	**Blind**	**Guided**	**Comb**
Wine TSNE	**.029**	**.001**	**<.001**	**.01**	**<.001**	**<.001**
Wine PCA	**.015**	**<.001**	**<.001**	.121	**.041**	**.025**
Concrete TSNE	**.003**	**.035**	**.001**	**.004**	.111	**.004**
Reuters AE	.087	**<.001**	**<.001**	.232	**<.001**	**<.001**
Reuters TSNE	.418	**.025**	.059	.595	**.029**	.111
Software TSNE	**.04**	**<.001**	**<.001**	.088	**<.001**	**<.001**

	S			N		
	Blind	**Guided**	**Comb**	**Blind**	**Guided**	**Comb**
Wine TSNE	**.005**	**<.001**	**<.001**	.24	**.001**	**.003**
Wine PCA	**.003**	**<.001**	**<.001**	**<.001**	**<.001**	**<.001**
Concrete TSNE	.159	.156	.08	.148	**.049**	**.029**
Reuters AE	.214	**<.001**	**<.001**	.448	**.025**	.071
Reuters TSNE	.935	**.009**	.23	.679	.056	.197
Software TSNE	.689	**<.001**	**.002**	.279	**<.001**	**<.001**

Table 5: P-values for a test of whether the mean metric values of the viewpoints picked by the users are significantly higher than the mean metric values of the $S = 1000$ sampled viewpoints, calculated using the equal variance, one tail t-test. We show the values for the blind set, the guided set and the combined set. Significant values ($p < 0.05$) are displayed in bold.

We can however, only use the viewpoints from the blind set to fairly answer the question of whether users prefer viewpoints that are of higher quality according to our metrics. This is because users were biased to pick higher quality viewpoints for the guided set because our tool said that they were good. Perhaps they themselves did not think them better. Whether the views from the blind set have a significantly higher average quality varies, and seems to depend on the dataset. For both Reuters projections, users did not pick viewpoints with significantly higher average quality for any metric. We see this also in the box plots in figure 21 where the mean metric values are often very similar, or even lower than average in the blind set. The Concrete and especially Wine dataset projections do show a significant increase in metric values for viewpoints of the blind set.

Following these observations, we have to conclude for now that users do not prefer higher quality viewpoints in all cases. This could be because the number of participants is too low for statistically significant results, but likely the correlation between measured quality and user-perceived quality depends on the datasets. After all, the strength of the correlation between high metric values and good visual cluster separation is inherently dataset dependent. If class clusters are not separated well in the dataset D then a good projection of it will also have bad cluster separation. In this case there might be little difference between a high and low quality view in the eyes of the user. All our datasets are real world datasets, which means clear separation of classes is not to be expected. We specifically selected projections for which finding a view with clear cluster separation was not trivial. The Reuters dataset, which appears to score the worst when it concerns correlation with metric quality and user perceived quality, also happens to be one of the most complex datasets we have. With $N = 8432$ samples and $n = 1000$ dimensions. A topic of future research could be how the quality or complexity of the dataset affects the strength of the correlation between user-perceived quality and objectively measured quality. In this user evaluation, we lack the number of different datasets and number of participants to make relevant statements on this topic.

5.3.5 Quality improvement for views of the guided set

A conclusion we can make from our data is that the guiding widgets cause users to select viewpoints with significantly higher than average quality metric values. Generally the selected views also have higher quality metric values than those they would have picked without the guiding widgets. Metric values, however do not measure visual cluster separation, which was the task given to the users. They only measure dataset preservation of dataset properties. Therefore, We can not assume that users actually believed these views more visually appealing than the views users picked in the blind set. But we argue that, even if we stick with the weaker assumption that users only believed the views equally visually appealing, then the views with higher quality are still preferable. Finding a good view is a multi-objective problem, since objectively, the best viewpoints should be of high visual appeal to the user, and of high quality according to metrics. Because the latter means the dataset is more strongly represented, resulting in more reliable insights. It appears that our tool helps users find such viewpoints.

5.3.6 User perceived quality difference between the blind and guided set

So we have found that our tool causes users to pick higher quality viewpoints, but we can not be sure whether users indeed prefer them over the lower quality viewpoints they selected in the blind dataset. We have one other measurement that indirectly captures how users perceived the quality of the 3D projection view, which is whether they preferred it over the 2D projection or not. If users prefer high quality viewpoints, then the views they prefer over the 2D projection should reasonably be of higher quality than the views they do not prefer over 2D, since clearly they deem the former set better than the latter. Figure 22 contains a bar graph that shows for both the blind and guided set the percentage of views that users preferred over the 2D projection of the same dataset. Here we see that generally, users were more inclined to prefer the 3D projection view for the viewpoints they picked when guided by the sphere and histogram widgets. Especially for the wine dataset projections do we see a substantial increase for the guided set. This is interesting, since for the wine dataset we also observed the strongest correlation between metric values and user perceived quality according to table 5. For the Reuters and Software dataset projections, where we found little or no correlation between metric values and user perceived quality there is a substantially smaller increase in 3D over 2D preference in the guided set. In fact for the Software TSNE projection, we see a substantial decrease in preference of the 3D projection view for the guided set, which seems odd. Users were allowed to ignore the suggestions of the tool to pick viewpoints they liked, they were only given more information. So one would expect that, in the worst scenario, this extra information was of no benefit to the user in which case the user was essentially blind again. This should result in viewpoints with more or less equal user-perceived quality compared to the blind set, and therefore equal 3D over 2D preference. It could be that the suggested viewpoints were mostly bad in the eyes of the user, and that users were therefore mislead into picking worse viewpoints. But if they deemed the suggested viewpoints bad, why did they not search for a better one? Either way, we observe that for datasets where we've measured a stronger correlation between metric values and user perceived quality, users appear to benefit more from the suggestions of high quality viewpoints.

We observe that totalled over all projections, there is a slight increase in 3D over 2D preference for the guided set. If we correlate this observation with the finding in figure 21 and table 5 that the viewpoints in the guided set are of higher quality, there appears to be more evidence towards a correlation between higher metric quality and user-perceived quality. To further investigate this we performed another T-test to calculate whether there is a significant increase in metric values for 3D views that users preferred over the 2D projection, compared to the quality of views they did not prefer over the same 2D projection. The P-values of the tests are displayed in table 6. Here, only for the wine dataset projections and the continuity metric, do we see a significant increase in values compared to the viewpoint qualities where users preferred the 2D projection. P-values above 0.5 indicate a negative correlation between high metric values and 3D view preference, which happens quite often, especially for the Reuters and Software dataset projections. Though only significantly ($p > 0.95$) for the the continuity metric in the Reuters AE projection. These results, appear to support our earlier findings that it depends on the dataset whether there is a positive correlation between metric values and perceived quality, but since most results are not significant, we can not make any relevant claims using them.

5.3.7 Answering research subquestion B1 and B2

Following the previous discussion, we answer research subquestion B1 (Section 3) like this. We have found evidence that depending on which dataset was projected, users prefer viewpoints that are of higher quality according to our metrics, when it concerns visual cluster separation. Therefore, depending on the dataset the objective metrics are predictive of user perceived quality. Supporting this statement, we have also observed that, for the projections where we have found a stronger correlation between metric values and user perceived quality, we see an increased benefit for users from the suggestions of higher quality views by our guiding widgets.

A number of observations concern research subquestion B2 (Section 3). Initially, we have seen in figure 21 and table 5 that users pick higher quality viewpoints when the quality metrics are known to them. This is undoubtedly an improvement, but it does not prove that users also perceived these viewpoints to be of higher quality. However, it seems in figure 22 that users are generally more inclined to prefer the 3D projection view over the 2D projection in the guided set where users had access to the metric values. For some datasets this correlation is stronger than others. This indicates that knowing the metric values can indeed help users find viewpoints of not only higher metric quality but also better perceived quality. We therefore answer research subquestion B2 as follows. Yes, knowing the metric values, users pick higher quality viewpoints, that are more representative of the dataset and can therefore be considered as better, and depending on the dataset the user-perceived quality of these views also increases.

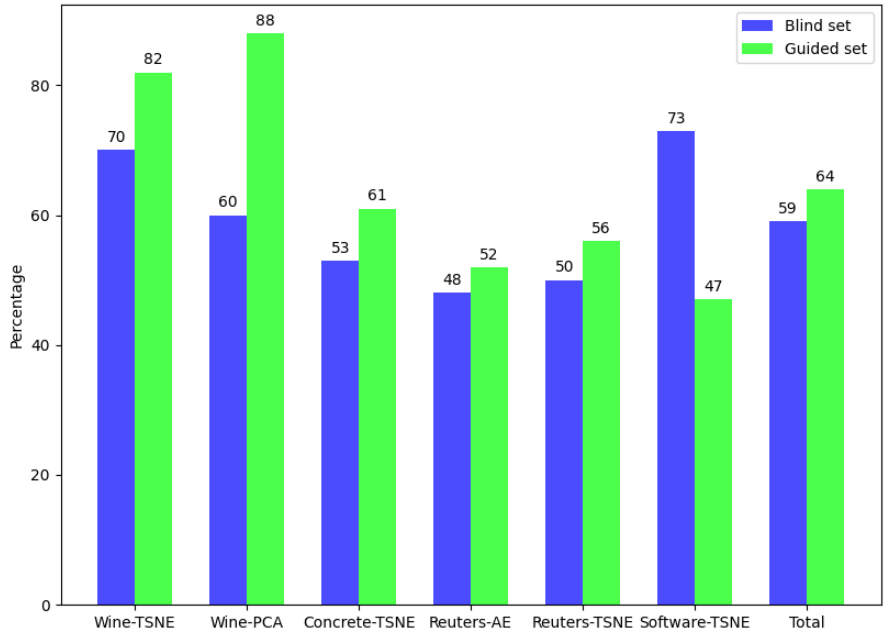

Figure 22: Bar graph displaying the percentage of views of the 3D projection where users specified a preference over the corresponding 2D projection, for each projection pair in the evaluation, for both the blind and guided set of viewpoints

	T	**C**	**S**	N
Wine TSNE	0.104	**0.037**	0.258	0.302
Wine PCA	0.058	**0.037**	0.104	0.308
Concrete TSNE	0.556	0.341	0.321	0.186
Reuters AE	0.916	0.954	0.877	0.455
Reuters TSNE	.411	.317	.557	.361
Software TSNE	.945	.695	.93	.842

Table 6: P-values for an equal variance, one-tail t-test of whether the set of views users preferred over the 2D projection have significantly higher metric quality than the set of views users did not prefer over the 2D projection, for each metric and each projection pair. Significant values ($p < 0.05$) are displayed in bold.

5.3.8 Is 3D better than 2D?

An interesting observation in figure 22 is that, regardless of whether they were guided or not, the users preferred most of their selected views of the 3D projection over the 2D projection. About 60% averaged over all projections. Thus, it is quite likely that a 3D projection contains single views that a user prefers over the 2D projection of the same dataset. One reason of this occurring might be that users don't get a choice for the the 2D projection. There is just a single, unalterable projection they simply have to accept, whereas for the 3D projection they can essentially search through thousands of 2D projections until they find one that appeals to them. While, this is time consuming, the result of a more preferable view could be an argument in favor of 3D projections. After all, if a single view outperforms a 2D projection, then one could argue that the combination of multiple such views of the 3D projection can definitely outperform a 2D projection. Either way, we can now answer research subquestion B3 (Section 3). It is likely that a user can find a view they deem better than or at least equally good as the 2D projection. Naturally it depends on the projection. For some projections, like the Wine dataset projections, users reported significantly more often to prefer their chosen view of the 3D projection, but for none of the projections is there a significant preference for 2D. The worst we observe is a more or less 50/50 ratio, which we can interpret as that users simply do not have a preference, and they picked one at random. That means that overall, the selected views are at least as good as 2D, and can be better. So in terms of user perceived quality, there is no reason not to use a 3D projection.

On a side note, we should not disregard the possibility that users were biased into reporting they preferred their selected view. They might have been more inclined to say they prefer it over the 2D projection, since they spent time and effort to find it. To test if this bias exists, we would need a separate user experiment where unbiased users are asked to rate the views picked by other users compared to the 2D projection. Which is a topic we suggest for future work.

At the end of the evaluation, we asked users to rate, on a 7 point scale, their agreement with the following statement: *Considering the entire 3D projection, instead of just one viewpoint, the 3D projection better displays the data* *structure than the 2D projection.* Every single participant responded with a value on the positive side of the scale (4 or higher), with an average of 5.94. Thus all users agreed that the 3D projection in its entirety has a better representation of the structure than the 2D projection. All our measurements up till now are focused on single views of 3D projections, since this allows for a fairer comparison to 2D, but it does complicate the question of whether 3D is better than 2D. The answers to this question indicate that, although we have not found strong evidence that users prefer 3D views over 2D, they do believe the 3D projection in its entirety shows more of the dataset structure.

6 Discussion and conclusions

In this project we proposed a new, viewpoint-driven way of measuring and analyzing the quality of 3D projections, hoping that it could give us better insights in finding out if and why a 3D projection is better than a 2D projection. Traditionally, the quality of 3D projections is measured as how well the structure of D is preserved in the 3D space, just like for 2D projections in the 2D space. We have argued that this is an unfair measurement, because a user can not see in three dimensions, and the 3D projection is displayed on a 2D screen anyway. We therefore proposed a new tool that instead measures the quality of 3D projections as a function of the viewpoint, where each 3D projection view is evaluated as a 2D projection. Two different widgets allow users to see in real-time what the quality of the current view of the 3D projection is. Using this tool we performed a quantitative analysis of 30 projections created using 6 datasets and 5 different projection techniques. It becomes clear that the quality of a 3D projection depends significantly on the viewpoint it is observed from. Our findings include that it is difficult to compare quality values of different metrics or different projections, since the effective range of quality metrics depends strongly on the metric and projection. For example, all continuity metric values were close to the maximum of 1, whereas for Shepard diagram correlation the values are spread out much more over the 0-1 range. We have shown that relative within its own signal range, changes in metric values can be just as important as other metrics. A small change in continuity likely has a stronger effect on the projection quality than the same change in the Shepard correlation metric.

Using images of the histogram and sphere widgets of our tool, we searched for patterns in how the view quality of 3D projections is distributed for different projections. We observe that generally, most views of a 3D projections are of relatively high quality, with a smaller number of views that have significantly worse quality. This means that in general, it should not be hard for a user to find one of the better viewpoints in terms of metric quality.

We have not found any patterns in the way that the quality of viewpoints is distributed that are specific to any one dataset. Which indicates that the quantitative findings of this work are likely generalizeable over all datasets. We do observe patterns that appear specific to some projection techniques. For example, the view quality distributions of UMAP projections are more peaked, which means that, relative to projections by other techniques, their outliers are significantly further away. The conclusion we can draw from this is that the projection technique influences how easy it is for users to find a viewpoint in the higher quality ranges, and how much worse it is if a user picks a lower quality viewpoint.

With our measurement method, we can still not directly compare 3D versus 2D projections, so we can not objectively state when 3D or 2D is better for specific projection techniques or datasets. But by looking at how the quality of 3D projection views compares to the quality of the 2D projection, we are able to give an intuition of which is better. We have seen that depending on the projection technique and the metric, a significant portion of the views of a 3D projection score better than the 2D projection. Therefore, one way to measure whether 3D or 2D is better could be to compute the percentage of views of a 3D projection, that score better than the 2D projection. To further refine this metric one could also measure the actual quality increase for these better views.

In our user evaluation we found that generally users appear to have a preference for views with above average metric values. We found that the strength of this correlation appears to depend on the dataset, sometimes there is no correlation at all. The guiding widgets of our proposed tool cause users to select significantly higher quality views, which offer a better representation of the dataset and also appear to be better, or at least as good in the eyes of the user. For datasets that have a stronger correlation with metric quality and user perceived quality, we notice that suggesting high quality views using our tool is of more benefit to users than for datasets where we did not find a such a correlation. Therefore,whether we can measure if a user will prefer one projection over the other, and subsequently the effectiveness of a tool like ours, seems to depend on the dataset. Since we only have a few datasets to base these findings on, more research is required to validate them. An interesting topic would be to investigate how the quality or complexity of datasets affects the correlation of metric values with user perceived quality.

We do find that in most cases, users can find a view of the 3D projection that they deem better than, or at least equally good as the 2D projection. Which indicates that purely in terms of the user perceived quality, there is no reason not to use a 3D projection. On top of that, all participants of our user evaluation unanimously reported that they believe the 3D projection in its entirety to better display the dataset structure than the 2D projection, a strong argument in favor of using 3D projections.

Our research question stated in Section 3 was the following. *Can we measure from a user perspective, for different projection techniques and datasets, whether, and by how much, a 3D projection is better than a 2D projection?*. We answer it with the following summary. Our new way to measure and compare the quality of 3D and 2D projections can in some sense give an objective measure which of the two is better, and by how much, by looking at how the quality of the views of the 3D projection compare to the quality of the 2D projection. If a significant proportion of the views have higher quality values than the 2D projection, then one could argue that it is better than the 2D projection. A threshold for the proportion, could be defined for this. However, the extent to which users prefer views with higher quality metric values, and therefore the extent to which we can measure whether a user will prefer 3D or 2D, depends on the dataset. But overall, users can generally find a view of a 3D projection they like at least equally well as the 2D projection, arguably, a 3D projection in its entirety can then only be better. A statement supported by the result that all users in our experiment believe the 3D projection in its entirety to better represent the dataset structure than the 2D projection. Of course, here we look purely in terms of quality and we do not take into account that the 3D projection takes more time and interaction to explore.

There are many directions in which to continue or expand this line of research. Naturally, one could be to set up a user evaluation with more participants for more statistically relevant results, and instead of letting users rate their own found views, let them rate those of others to get rid of any bias. Experiments could be done with different or more projection techniques, datasets and metrics. We already mentioned the relevance of an investigation of how dataset quality and complexity affects the correlation between user perceived quality and metric values. Part of this investigation could be to define a metric that is normalized by dataset complexity, so that metric values of projections of different datasets can be compared more fairly. Further research could be done on whether users prefer higher quality views in general by, for example, letting them repeatedly choose which of two random views they prefer, and observe how often they pick the higher quality views. We also have not investigated how much 3D or 2D preference depends on the users. Likely there are some users that are more inclined to prefer either of the two. It would be interesting to see how many users always prefer 3D, how many always prefer 2D and how many choose differently each time, over different projections. Furthermore, users could be given a different task than rating visual cluster separation. Lastly, it would be interesting to gauge the effect of techniques like the Da Silva explanations [10] or Enhanced Biplots [9] on how users pick good viewpoints.

7 Appendix

7.1 User evaluation

Here we display the instructions given to all participants of our user experiment. We skip the parts on downloading, running the tool and uploading evaluation data.

7.1.1 Introduction

Dimensionality reduction is a popular data visualization technique that projects high-dimensional data to a low-dimensional space (2D or 3D) while preserving distance and/or neighborhood relations between points. The projected dataset can then be visualized in a 2D or 3D scatterplot. In this evaluation we compare 2D projections versus 3D projections of the same dataset from a user perspective.

For this purpose we created a tool and set up this user evaluation. We will next explain how to download the tool and get it running

If you encounter any problems, or have any questions during the survey you can contact me here: w.m.castelein@students.uu.nl

7.1.2 How to get the program running

Once running, the tool should look like the image in figure 23. In the next section we will explain what you see, and how the tool can be used.

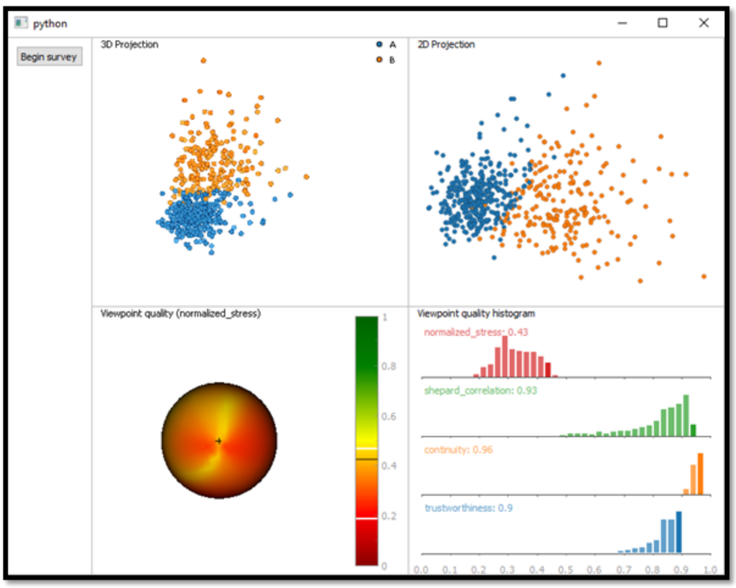

Figure 23:

7.1.3 Projection widgets

The top left widget, marked with a red A in the image in figure 24, contains a scatterplot of the 3D projection of a high dimensional dataset. Each point is color-coded based on a class label. The colors and labels are displayed in a legend in the top right. In our datasets, the classes can be either categorical or ordinal. Categorical classes are denoted with letters from the alphabet, and have very distinct colors. Ordinal classes are denoted with numbers and smooth color transition between nearby classes. For this survey, it only matters that different classes exist, not what these classes are specifically. The projection can be rotated by holding and dragging the mouse. To give a feeling depth, closer points are colored slightly darker and further points slightly brighter.

The top right widget, marked with a red B in the image below, contains a scatterplot of a 2D projection of the same high dimensional dataset. These points are color-coded exactly the same as the 3D projection.

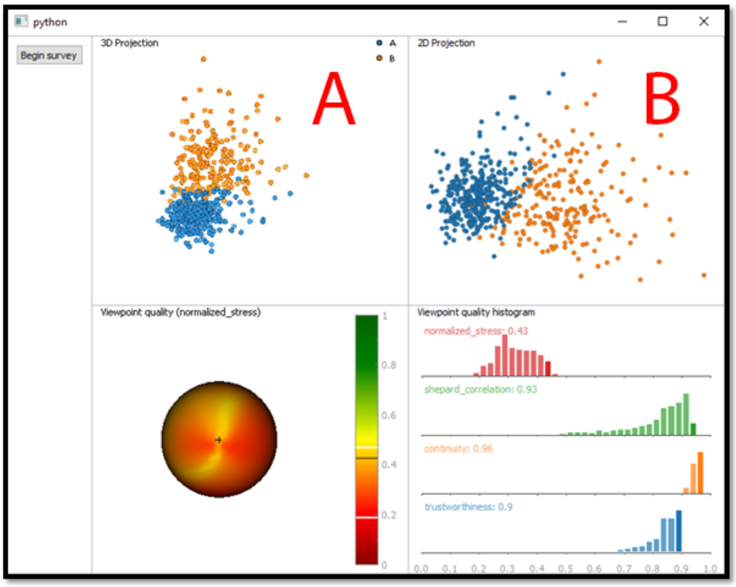

Figure 24:

7.1.4 Quality metrics

Before we look at the two bottom widgets, we need to give an intuitive explanation of four different quality metrics for projections. These metrics capture how well structure from the original dataset is preserved in the projection. Each quality metric has a value ranging between 0 and 1. Where 0 is the worst quality, and 1 the best.

The metrics Normalized stress and Shepard diagram correlation relate to how well a projection preserves distances between data points.

The other two metrics are called Continuity and Trustworthiness. These metrics relate to how well a projection keeps similar data points close to each other in the projection

7.1.5 The quality sphere widget

The quality metrics we just mentioned are used in the bottom two widgets. The bottom left widget contains a colored sphere. As with the 3D projection, you can rotate this sphere. You will notice that the view of the sphere and the 3D projection are linked, moving one also moves the other. This is because the sphere shows the quality of the current view of the 3D projection, as measured by one of the previously described quality metrics.

Look at the image in figure 25 while reading what is explained next. The color under the crosshair in the center of the sphere shows the quality of the current viewpoint. Therefore, if we rotate the sphere such that the crosshair targets a green area, we know that the current view of the 3D projection is of high quality according to the selected quality metric. Likewise, a red area indicates a poor score.

The colored bar on the right of the sphere serves as a legend that displays the full range of colors from best (quality 1) to worst (quality 0). In this bar a black line indicates the quality of the current viewpoint. Two white lines indicate the minimum and maximum quality values observed over the entire sphere for the current metric.

It is only possible to look at one metric at a time. To change the metric shown by the sphere, click on the name (label) of the metric you want to show in the Quality histogram widget to the right of the Sphere widget.

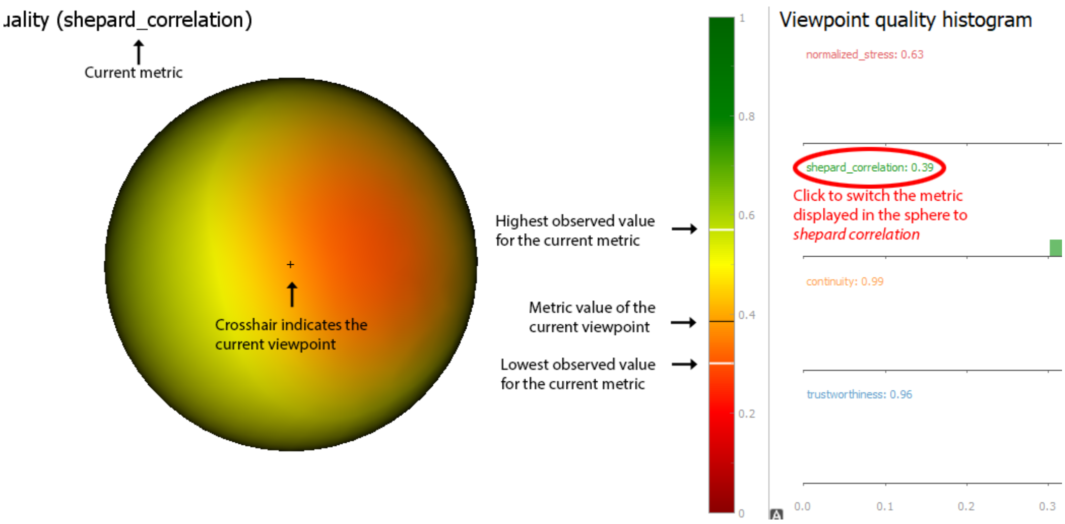

Figure 25:

7.1.6 Histogram widget

The bottom right widget contains one histogram for each of the four different quality metrics. The histograms show the quality distribution of 1000 different views of the 3D projection. Each bar is placed on a quality value interval, the height of the bar indicates how many views exist that have a quality within this interval. The name of a metric is shown in the title above its histogram. Next to that, we show the metric value for the currently selected view of the 3D projection.

Each histogram contains a slightly darker bar. This is the interval in which the quality of the current view of the 3D projection falls. Rotate the 3D view to see how this bar changes.

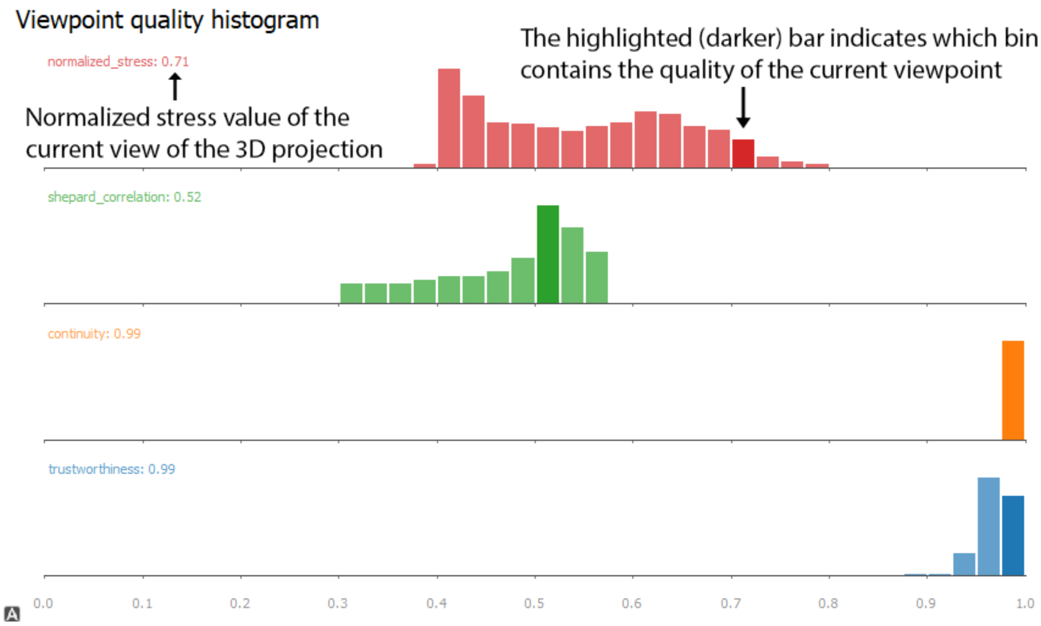

Figure 26:

7.1.7 Histogram hovering

You can hover the mouse pointer over a histogram bar to do several things. First, the 3D projection and colored sphere rotate to a view that has a quality value within the hovered bar. Hovering at the bottom of the bar will select a viewpoint with quality on the lower end of the bar's interval. Hovering at the top of the bar will select a viewpoint with quality on the higher end of the bar's interval. For example: to find a view with highest overall quality for a metric, move the pointer to the top of the rightmost bar of that metric's histogram.

A number of polylines are drawn from the hovered bar to the other histograms, forming a parallel coordinates plot (PCP). Each polyline corresponds to one viewpoint contained in the hovered bar. The polyline shows all quality metric values of that viewpoint by going through the axes of each metric. A thicker, and more opaque line highlights the currently hovered viewpoint. The PCP plot shows, for the hovered quality interval (bar) of a metric, how quality is spread for the other three metrics.

For example, the PCP plot in the image below shows that all views with a normalized stress value around 0.35 (red selected bar) have very different values for Shepard correlation (green bars), quite similar values for Continuity (orange bars), and quite different values for Trustworthiness (blue bars). To test if you understand the hovering tools, try to use it to find a viewpoint that has high quality for multiple metrics.

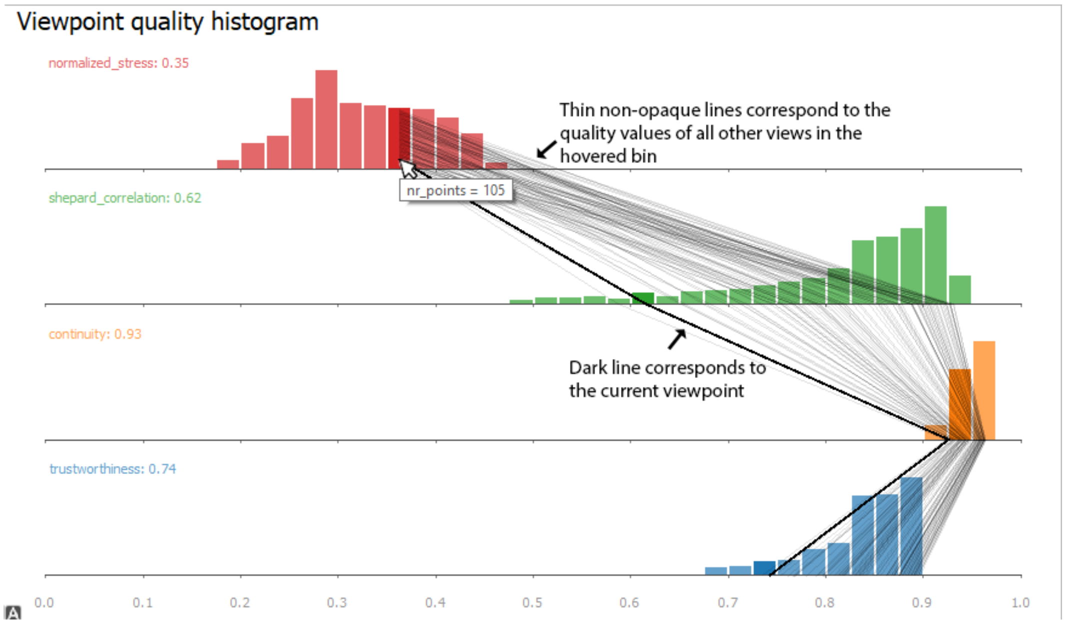

Figure 27:

7.1.8 Survey

The tool starts off with projections of an example dataset. This allows you to familiarize yourself with the tool before beginning the survey. Make sure the tool is in fullscreen view, if it isn't already so you can see more detail. Try to understand each of the widgets and how they can be used. When you are ready press the 'begin survey' button in the top left menu. (figure 28)

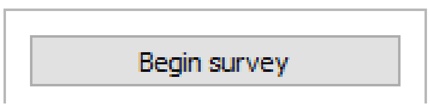

Figure 28:

This will load a new set of projections, and the bottom two widgets disappear for now.

The menu now looks like this: (figure 29)

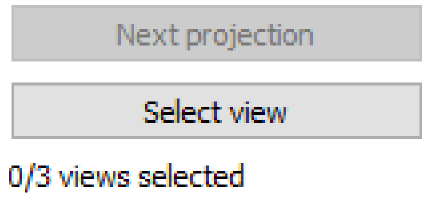

Figure 29:

We now ask you to repeatedly take the following steps:

Find a viewpoint in the 3D projection that best displays, according to you, visually well-separated point groups, that have similar colors internally.

Once you have found such a viewpoint, press the button 'Select view' in the top left menu. This will freeze the tool. (It is no longer possible to change the view). Two buttons have appeared with the texts '3D preference' and '2D preference'. (32)

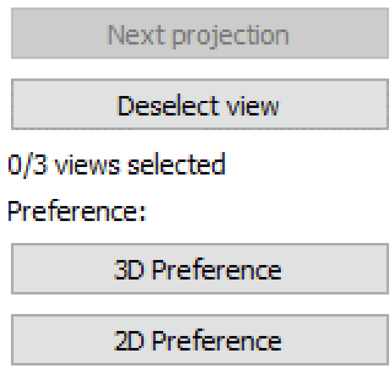

Figure 30:

Now look at the 3D projection (from the current view) and 2D projection and select the one which you think has clearer point-group and color separation. To do this, click one of the buttons shown above (3D preference or 2D preference)

Once you have done this it is possible to select a second viewpoint as long as it is not too close to a previously selected viewpoint. Repeat the previous steps until you have selected three different viewpoints, then press the 'next projection' button. Repeat these steps for three different projections, until the two bottom widgets reappear.

We now ask you to repeat the previous steps, but this time you can use the sphere and histogram widget to help finding good view points. Remember that the quality metrics do not directly measure visual group separation and similar internal coloring, but only preservation of the original dataset structure. Therefore you should see the tool as a recommender system for good viewpoints, but ultimately use your own judgement to pick a viewpoint.

Once you are finished, the program will close automatically.

7.2 Snapshots of user selected views

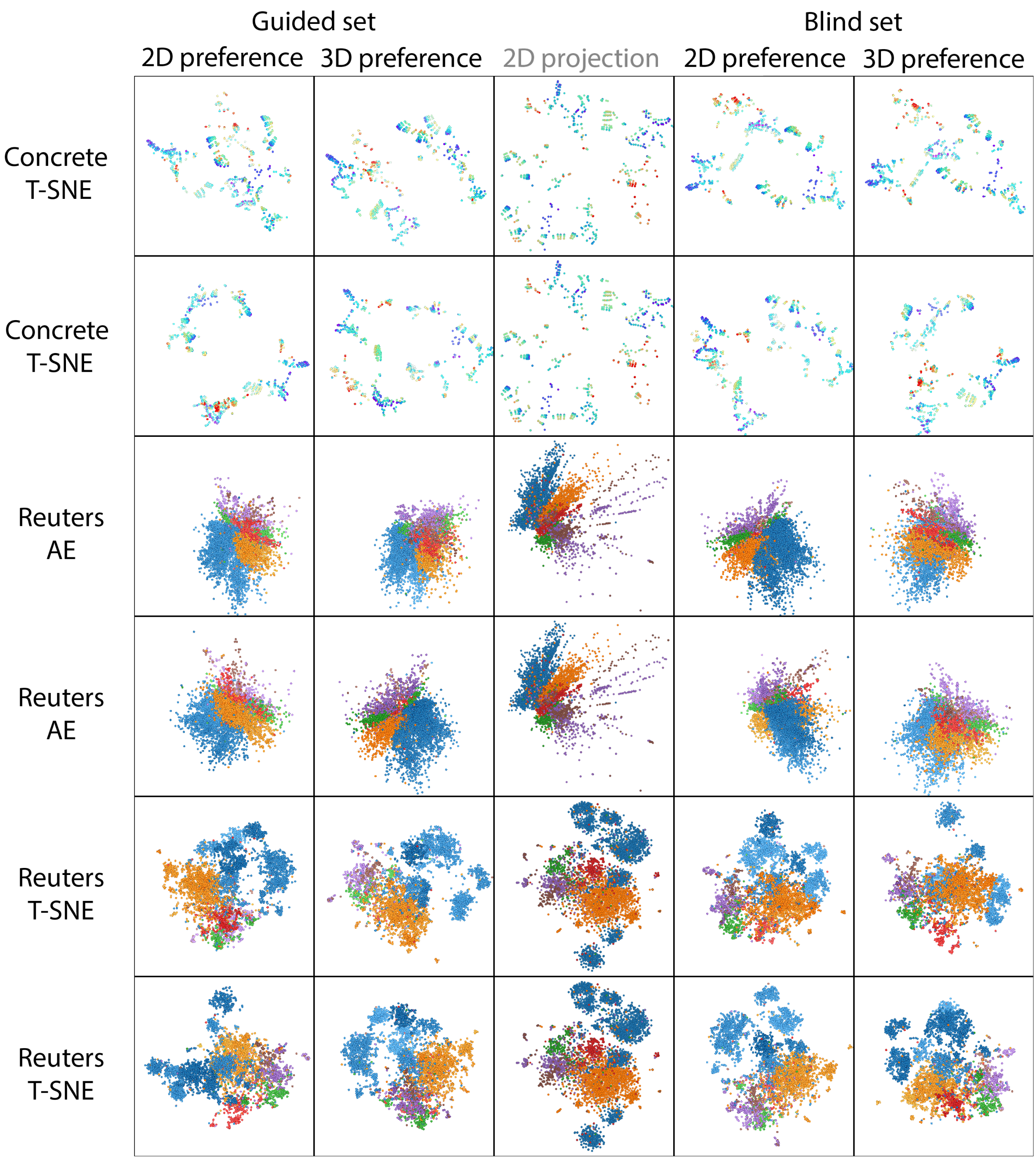

Figure 31: Arbitrarily chosen snapshots of the views selected by users in the user evaluation, for both the guided set and the blind set. A distinction is made between views where users preferred the 2D projection (2D preference) and the view of the 3D projection (3D preference). Per dataset/projection pair, we show 2 snapshots of each category

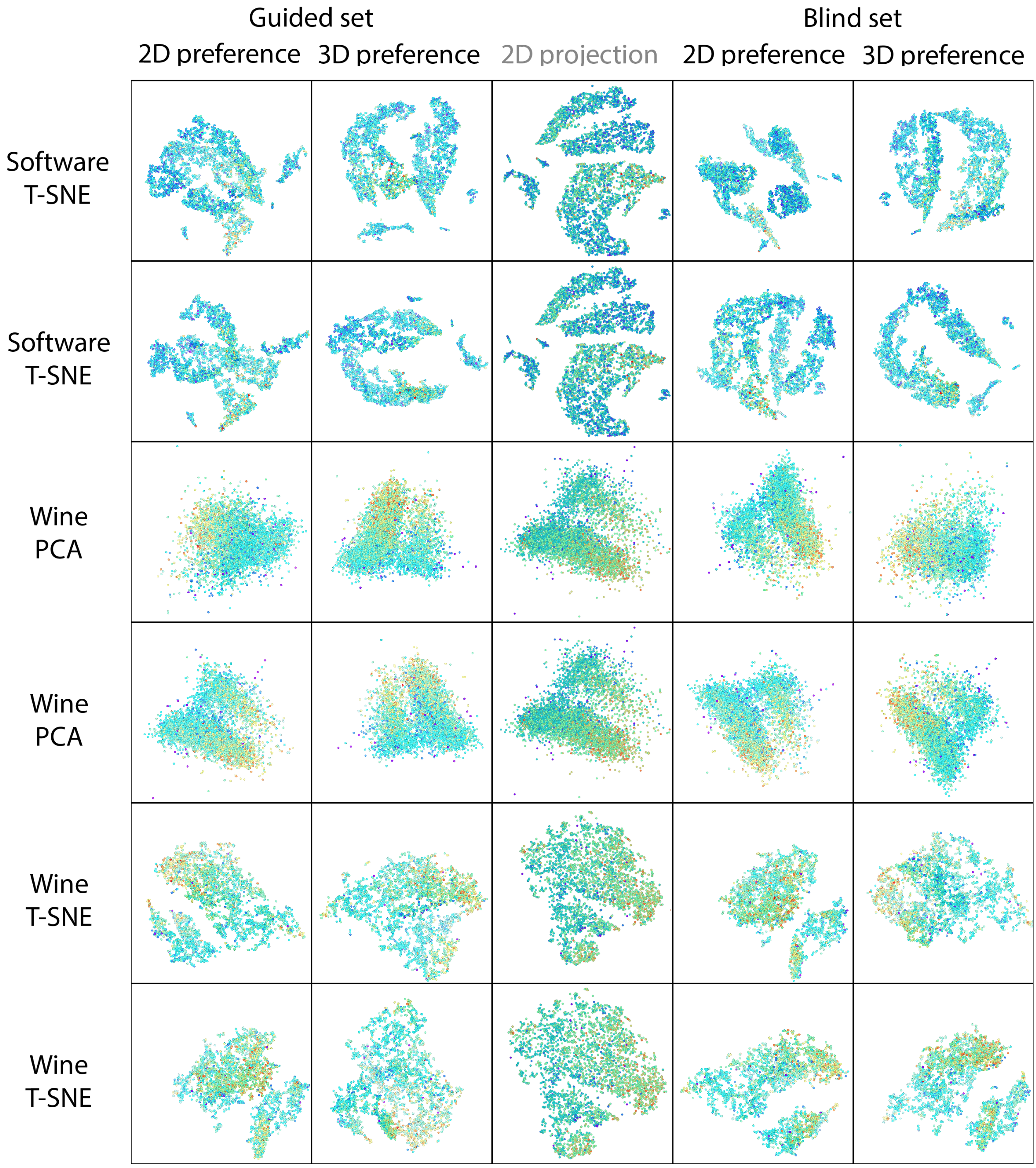

Figure 32: Arbitrarily chosen snapshots of the views selected by users in the user evaluation, for both the guided set and the blind set. A distinction is made between views where users preferred the 2D projection (2D preference) and the view of the 3D projection (3D preference). Per dataset/projection pair, we show 2 snapshots of each group

References

[1] Concrete compressive strength data set. 2022. Available online: https://archive.ics.uci.edu/ml/datasets/concrete+compressive+strength.

[2] Reuters newswire classification dataset. 2022. Available online: https://keras.io/api/datasets/reuters/.

[3] Wine quality data set. 2022. Available online: https://archive.ics.uci.edu/ml/datasets/wine+quality.

[4] Wisconsin breast cancer dataset. 2022. Available online: https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)].

[5] Mostafa M Abbas, Michaël Aupetit, Michael Sedlmair, and Halima Bensmail. Clustme: A visual quality measure for ranking monochrome scatterplots based on cluster patterns. In *Computer Graphics Forum*, volume 38, pages 225–236. Wiley Online Library, 2019.

[6] Michael Aupetit and Michael Sedlmair. Sepme: 2002 new visual separation measures. In *2016 IEEE pacific visualization symposium (PacificVis)*, pages 1–8. IEEE, 2016.

[7] Michaël Aupetit, Michael Sedlmair, Mostafa M Abbas, Abdelkader Baggag, and Halima Bensmail. Toward perception-based evaluation of clustering techniques for visual analytics. In *2019 IEEE Visualization Conference (VIS)*, pages 141–145. IEEE, 2019.

[8] Richard A Becker, William S Cleveland, and Ming-Jen Shyu. The visual design and control of trellis display. *Journal of computational and Graphical Statistics*, 5(2):123–155, 1996.

[9] Danilo B Coimbra, Rafael M Martins, Tácito TAT Neves, Alexandru C Telea, and Fernando V Paulovich. Explaining three-dimensional dimensionality reduction plots. *Information Visualization*, 15(2):154–172, 2016.

[10] Renato RO da Silva, Paulo E Rauber, Rafael Messias Martins, Rosane Minghim, and Alexandru C Telea. Attribute-based visual explanation of multidimensional projections. In *EuroVA@ EuroVis*, pages 31–35, 2015.

[11] Saverio De Vito, Ettore Massera, Marco Piga, Luca Martinotto, and Girolamo Di Francia. On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario. *Sensors and Actuators B: Chemical*, 129(2):750–757, 2008. Available online: https://archive.ics.uci.edu/ml/datasets/Air+Quality [Online; accessed 19-July-2008].

[12] Mateus Espadoto, Rafael M Martins, Andreas Kerren, Nina ST Hirata, and Alexandru C Telea. Toward a quantitative survey of dimension reduction techniques. *IEEE transactions on visualization and computer graphics*, 27(3):2153–2173, 2019.

[13] Álvaro González. Measurement of areas on a sphere using fibonacci and latitude–longitude lattices. *Mathematical Geosciences*, 42(1):49–64, 2010.

[14] Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks. *science*, 313(5786):504–507, 2006.

[15] Harold Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6):417, 1933.

[16] Alfred Inselberg and Bernard Dimsdale. Parallel coordinates: a tool for visualizing multi-dimensional geometry. In *Proceedings of the First IEEE Conference on Visualization: Visualization90*, pages 361–378. IEEE, 1990.

[17] Paulo Joia, Danilo Coimbra, Jose A Cuminato, Fernando V Paulovich, and Luis G Nonato. Local affine multidimensional projection. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2563–2571, 2011.

[18] Ian T Jolliffe. *Principal component analysis for special types of data*. Springer, 2002.

[19] Rafael Messias Martins, Danilo Barbosa Coimbra, Rosane Minghim, and Alexandru C Telea. Visual analysis of dimensionality reduction quality for parameterized projections. *Computers & Graphics*, 41:26–42, 2014.

[20] Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.

[21] Paulo Meirelles, Carlos Santos Jr, João Miranda, Fabio Kon, Antonio Terceiro, and Christina Chavez. A study of the relationships between source code metrics and attractiveness in free software projects. In *2010 Brazilian symposium on software engineering*, pages 11–20. IEEE, 2010.

[22] Paulo Pagliosa, Fernando V Paulovich, Rosane Minghim, Haim Levkowitz, and Luis Gustavo Nonato. Projection inspector: Assessment and synthesis of multidimensional projections. *Neurocomputing*, 150:599–610, 2015.

[23] Fernando V Paulovich, Luis G Nonato, Rosane Minghim, and Haim Levkowitz. Least square projection: A fast high-precision multidimensional projection technique and its application to document mapping. *IEEE Transactions on Visualization and Computer Graphics*, 14(3):564–575, 2008.

[24] Karl Pearson. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, 2(11):559–572, 1901.

[25] Harald Piringer, Robert Kosara, and Helwig Hauser. Interactive focus+context visualization with linked 2d/3d scatterplots. In *Proceedings. Second International Conference on Coordinated and Multiple Views in Exploratory Visualization, 2004.*, pages 49–60. IEEE, 2004.

[26] Jorge Poco, Ronak Etemadpour, Fernando Vieira Paulovich, TV Long, Paul Rosenthal, MCF d Oliveira, Lars Linsen, and Rosane Minghim. A framework for exploring multidimensional data with 3d projections. In *Computer Graphics Forum*, volume 30, pages 1111–1120. Wiley Online Library, 2011.

[27] Harald Sanftmann and Daniel Weiskopf. Illuminated 3d scatterplots. In *Computer Graphics Forum*, volume 28, pages 751–758. Wiley Online Library, 2009.

[28] Michael Sedlmair, Tamara Munzner, and Melanie Tory. Empirical guidance on scatterplot and dimension reduction technique choices. *IEEE transactions on visualization and computer graphics*, 19(12):2634–2643, 2013.

[29] Michael Sedlmair, Andrada Tatu, Tamara Munzner, and Melanie Tory. A taxonomy of visual cluster separation factors. In *Computer Graphics Forum*, volume 31, pages 1335–1344. Wiley Online Library, 2012.

[30] Mike Sips, Boris Neubert, John P Lewis, and Pat Hanrahan. Selecting good views of high-dimensional data using class consistency. In *Computer Graphics Forum*, volume 28, pages 831–838. Wiley Online Library, 2009.

[31] Monica Tavanti and Mats Lind. 2d vs 3d, implications on spatial memory. In *IEEE Symposium on Information Visualization, 2001. INFOVIS 2001.*, pages 139–145. IEEE, 2001.

[32] Zonglin Tian, Xiaorui Zhai, Daan van Driel, Gijs van Steenpaal, Mateus Espadoto, and Alexandru Telea. Using multiple attribute-based explanations of multidimensional projections to explore high-dimensional data. *Computers & Graphics*, 98:93–104, 2021.

[33] Zonglin Tian, Xiaorui Zhai, Gijs van Steenpaal, Lingyun Yu, Evanthia Dimara, Mateus Espadoto, and Alexandru Telea. Quantitative and qualitative comparison of 2d and 3d projection techniques for high-dimensional data. *Information*, 12(6):239, 2021.

[34] Warren S Torgerson. Theory and methods of scaling. 1958.

[35] Melanie Tory, Arthur E Kirkpatrick, M Stella Atkins, and Torsten Moller. Visualization task performance with 2d, 3d, and combination displays. *IEEE transactions on visualization and computer graphics*, 12(1):2–13, 2005.

[36] Melanie Tory, Colin Swindells, and Rebecca Dreezer. Comparing dot and landscape spatializations for visual memory differences. *IEEE transactions on visualization and computer graphics*, 15(6):1033–1040, 2009.

[37] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.

[38] Daan van Driel, Xiaorui Zhai, Zonglin Tian, and Alexandru C Telea. Enhanced attribute-based explanations of multidimensional projections. In *EuroVA@ Eurographics/EuroVis*, pages 37–41, 2020.

[39] Jarkko Venna and Samuel Kaski. Visualizing gene interaction graphs with local multidimensional scaling. In *ESANN*, volume 6, pages 557–562. Citeseer, 2006.

[40] Yunhai Wang, Kang Feng, Xiaowei Chu, Jian Zhang, Chi-Wing Fu, Michael Sedlmair, Xiaohui Yu, and Baoquan Chen. A perception-driven approach to supervised dimensionality reduction for visualization. *IEEE transactions on visualization and computer graphics*, 24(5):1828–1840, 2017.