







Measuring and Interpreting the Quality of 3D Projections of High-Dimensional Data

Zonglin Tian , Wouter Castelein , Tamara Mchedlidze ,
and Alexandru C. Telea 

Department of Information and Computing Sciences, Faculty of Science, Utrecht University,
Utrecht, The Netherlands
a.c.telea@uu.nl

Abstract. Projections of high-dimensional data are among the most frequently used tools for exploring such data in information visualization. In contrast to 2D projections, which create static planar scatterplots, 3D projections create point clouds that can be visually explored from many viewpoints. The relative added value of using 3D projections is still a topic for debate in the community, having both proponents and critics. In this work, we propose several techniques to both increase the effectiveness of exploration of 3D projections and also measure their quality. We start by extending well-known definitions of 2D projection quality metrics to account for user-chosen viewpoints and inherent occlusion of 3D projections when viewed from such viewpoints. We also propose an interactive exploration tool for finding high-quality viewpoints from the perspective of such metrics. Using our tool, we show that 3D projections often allow viewpoints exhibiting higher quality than their 2D counterparts. Next, we enrich the interpretation of such viewpoints by explanatory techniques for 2D projections and show that good viewpoints, from the perspective of our metrics, also allow easy-to-interpret explanations of the depicted data. We use our tool in a user study to gauge how our computed quality metrics correlate with user-perceived quality for a cluster identification task. Our results show that our metrics can predict well viewpoints deemed good by users and that our tool increases the users' preference for 3D projections as compared to classical 2D projections.

Keywords: Multidimensional projections · Visual quality metrics · Perception · User studies · Projection explanations

1 Introduction

Dimensionality reduction (DR), also called projection, is a popular technique for visualizing high-dimensional datasets by low-dimensional scatterplots. Tens of different DR techniques [14] have been designed to address the several requirements one has for this class of methods, such as computational scalability, ease of use, robustness to noise

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-66743-5_16.

or small data changes, projecting additional points along those existing in an original dataset (out-of-sample ability), and visual quality.

Visual quality is a key requirement for DR methods. Globally put, a good projection scatterplot captures well the so-called *data structure* present in the original high-dimensional data in terms of point clusters, outliers, and correlations [14, 18, 26]. As such, high-quality projections are essential to allow users to reason about the data structure by exploring the visual structure of the scatterplot.

Projection techniques used for visualization purposes can typically create 2D or 3D scatterplots equally easily. For brevity, we call such scatterplots 2D and 3D projections respectively. In contrast to 2D projections, 3D projections have one extra dimension to project the data (thus, can in principle achieve higher quality). However, the user must choose a suitable viewpoint for analysis. As such, the perceived quality of 3D projections depends on what precisely the user sees from the chosen viewpoint.

In contrast to 2D projections, which do not have a viewpoint-selection challenge, the quality of 3D projections has been studied in far less detail. In this paper, we aim to extend our understanding of 3D projection quality by answering the following questions:

- Q1: How can we **measure** the quality of 3D projections by means of quantitative metrics?
- Q2: How do 3D projections **compare** with their 2D counterparts (generated on the same datasets by the same projection technique) from the perspective of these metrics?
- Q3: How do our proposed quality metrics correlate with quality as **perceived** by actual users?
- Q4: How do **explanatory** techniques help understanding 3D projections?

In earlier work [8], we aimed to answer these questions as follows. We measure 3D projection quality by a *function* (rather than a single value) that evaluates existing 2D quality metrics over a large set of 2D viewpoints of the 3D projection (Q1). Next, we quantitatively analyze 30 3D projections (five techniques run on six datasets) and find that most views of a 3D projection are of relatively high quality, with only a few poor views, and that these good views can have higher quality than a 2D projection made with the same technique for the same dataset (Q2). We propose an interactive tool for exploring the viewpoint-based quality. We perform a user study to test which viewpoints users *perceive* to be good for a cluster separation task and if these have high quality values as *measured* by our metric (Q3). We find a correlation of perceived vs computed quality, which suggests that the latter can be used to predict the former. This is further confirmed by seeing that users select even-higher quality views when they can explore the quality metrics.

In this paper, we refine the above insights along two main directions. First, we present a qualitative analysis of the user-selected viewpoints which visually confirms that these viewpoints are effective for the task at hand (Sect. 5.1A). We also present additional quantitative measurements that support the added value of our newly introduced metrics (Sect. 5.1C). Secondly, we address Q4 by using two so-called explanatory techniques to depict additional information atop of both 2D and 3D projections (Sect. 5.2). We show that such techniques succeed in explaining good viewpoints of 3D

projections very well and, in many cases, better than explaining the corresponding 2D projections.

2 Related Work

Let $D = \{\mathbf{x}_i\}$ be a dataset of n -dimensional data points $\mathbf{x}_i \in \mathbb{R}^n$. A projection P maps D to $P(D) = \{\mathbf{y}_i\}$, where $\mathbf{y}_i \in \mathbb{R}^q$ is the projection of \mathbf{x}_i . Typically $q \ll n$, yielding 2D projections ($q = 2$) and 3D projections ($q = 3$) that depict D by the respective scatterplots. Let next P_2 be a technique P that creates 2D projections ($q = 2$); P_3 for 3D projections ($q = 3$); we use P when the dimension q is not relevant for the discussion.

A *quality metric* is a function $M(D, P(D)) \rightarrow \mathbb{R}^+$ that tells how well the scatterplot $P(D)$ captures aspects of the dataset D . Such metrics exist both for 2D projections (Sect. 2.1) and 3D projections (Sect. 2.2).

2.1 Measuring the Quality of 2D Projections

Three types quality measurements exist for 2D projections, as follows.

Quantitative Metrics: Such approaches aim to compute a single (scalar) value M that gauges how well $P(D)$ captures the structure of D (see Table 1). Trustworthiness T measures the fraction of close points in D that are also close in $P(D)$ [43]. High T implies that visual patterns in $P(D)$ represent actual patterns in D , *i.e.*, the projection has few so-called false neighbors [21]. Conversely, continuity C measures the fraction of points in $P(D)$ that are also close in D [43]. High C implies that data patterns in D are captured by $P(D)$, *i.e.*, the projection has few so-called missing neighbors [21]. Normalized stress N measures how well inter-point distances in $P(D)$ reflect the same inter-point distances in D , where distances are typically measured by the L_2 metric. This tells how well a projection conveys, or preserves, distance information [16]. Distance preservation can also be measured by the Shepard diagram [16], a scatterplot of the L_2 distances between all points in $P(D)$ vs the corresponding distances in D . Points close to the plot's main diagonal show a good distance preservation. The diagram can be summarized by computing its Spearman rank correlation S , with $S = 1$ telling a perfect (positive) correlation of distances in D and $P(D)$. Many other metrics exist for 2D projections, *e.g.*, the neighborhood hit (NH) that captures how well $P(D)$ captures same-label clusters in D [4, 27, 29]; the Distance Consistency (DSC) [35] and Class Consistency Measures (CCM) [33, 36]), which tell how well $P(D)$ is separated into visually distinct, same-label, clusters. Additional visual separation metrics are given by [1, 25, 34].

Error Views: These methods offer a finer-grained insight of projection quality, typically at the level of each projection point \mathbf{y}_i . Views include the projection precision score [32], which aggregates the difference between distances of a point in $P(D)$ to its K nearest neighbors in D , respectively $P(D)$; stretching and compression [3, 18], which measure the increase (stretching), respectively decrease (compression) of distances of a point to all other points in $P(D)$ vs corresponding distances in D ; and the

Table 1. Metrics used to gauge the quality of 2D and 3D projections. All metrics range in $[0 = \text{worst}, 1 = \text{best}]$. $U_i^{(K)}$ are the K nearest neighbors of the projection points \mathbf{y}_i which are not among the K nearest neighbors of the data points \mathbf{x}_i ; and $r(i, j)$ is the rank of \mathbf{y}_j in the ordered-set of nearest neighbors of \mathbf{y}_i . $V_i^{(K)}$ are the K nearest neighbors of \mathbf{x}_i which are not among the K nearest neighbors of \mathbf{y}_i ; and $\hat{r}(i, j)$ is the rank of \mathbf{y}_j in the ordered-set of nearest neighbors of \mathbf{x}_i . Metric definitions taken from [14].

Metric	Definition (taken from [14])
Trustworthiness (T)	$1 - \frac{2}{NK(2N-3K-1)} \sum_{i=1}^N \sum_{j \in U_i^{(K)}} (r(i, j) - K)$
Continuity (C)	$1 - \frac{2}{NK(2N-3K-1)} \sum_{i=1}^N \sum_{j \in V_i^{(K)}} (\hat{r}(i, j) - K)$
Normalized stress (N)	$\frac{\sum_{ij} (\Delta^n(\mathbf{x}_i, \mathbf{x}_j) - \Delta^n(P(\mathbf{x}_i), P(\mathbf{x}_j)))^2}{\sum_{ij} \Delta^n(\mathbf{x}_i, \mathbf{x}_j)^2}$
Shepard goodness (S)	Spearman rank correlation of scatterplot $(\ \mathbf{x}_i - \mathbf{x}_j\ , \ P(\mathbf{x}_i) - P(\mathbf{x}_j)\), 1 \leq i \leq N, i \neq j$

average local error [21], which combines stretching and compression. We do not further use such views since they cannot be automatically compared against each other to assess the relative quality of several projections.

User Evaluations: As discussed by several authors [14,26], ultimate suitability of a projection for a given analytic *task* requires executing specific user studies with that task in mind. Quantitative metrics are, however, an essential first step in this evaluation since (a) if these score poorly, the projection is very likely not suited for further use; and (b) they can be easily and automatically evaluated for many combinations of datasets and projection techniques. We follow this approach in our work.

2.2 Measuring the Quality of 3D Projections

Recently, Tian *et al.* [40] compared 29 projection techniques across 8 datasets using the T , C , and S metrics computed for the 2D, respectively 3D, scatterplots $P(D)$. They found that 3D projections show a small quality increase (on average, 3%) *vs* the 2D projections. A key problem of the comparison in [40] is that simply transposing 2D quality metrics (computed as in Tab. 1) to 3D has a major issue. Even if such metrics score highly on a 3D projection, this does not mean that *users* can see data patterns well. Indeed, the metrics ‘see’ $P(D)$ in three dimensions; users see only 2D *views* of $P(D)$ from chosen viewpoints. Information encoded along the viewing direction is used by the metrics but not seen by users, so the metrics could indicate artificial high-quality values that users do not perceive. We solve this issue by proposing quality metrics for 3D projections that take the viewpoint into account (Sect. 3).

Separately, Tian *et al.* also showed that adding *explanations* to the 3D projections are helpful in attracting users to explore these further. For this, they applied earlier-presented techniques that color a 2D projection by the data dimension having locally the least variance [12], respectively by the local intrinsic dimensionality of the data points [39], to 3D projections. However, they noted that finding good viewpoints from which such explanations are easy to understand is a laborious manual process. We solve this issue by showing that our new quality metrics also allow for the automatic generation of easy-to-understand explanations (Sect. 5.2).

3D projections have been also assessed by user studies. [28] compare 2D and 3D projections and show that 3D scores better than 2D for the NH and C metrics. They refine this insight by a user study where 12 participants were asked to count visual clusters seen in $P(D)$, order clusters by density, list all pairwise cluster overlaps, detect an object within a cluster, and find the cluster closest to a given point. Users provided better answers to these tasks in 3D (74.4%) than in 2D (64.3%). Yet, a statistically significant improvement was only found for the last task. Also, users needed around 50% more time for these tasks in 3D. Overall, this work suggests a slight, uncertain, advantage for 3D projections. [34] measured how well classes of 75 labeled datasets were separable in a 2D projection, an interactive 3D projection, and a scatterplot matrix. They found that the 2D projection was often good enough to visualize separate classes and was also the fastest method to use. The interactive 3D projection scored better than the 2D one and the scatterplot matrix only for highly synthetic (abstract) datasets. However, this study involved only two users. In contrast, our evaluation for similar tasks involves 22 users (Sect. 5).

Several works compared 2D with 3D scatterplots and argued that the latter better captures sample density variations [30,31] with less information loss [9]. However, 3D scatterplots whose axes directly encode data *dimensions* are very different from 3D projections where the three axes often have no meaning. As such, a further study specifically targeting 3D projections is warranted.

3 Viewpoint-Dependent 3D Projection Quality

As outlined in Sect. 2, quantitative metrics are a useful, scalable, generic, and accepted first step for evaluating 2D projections, but we lack such metrics for the 3D case. We address this by extending 2D projection metrics to the 3D case as follows.

Metric Definition: Take a 3D projection $P_3(D)$ to be explored from multiple viewpoints using a virtual trackball metaphor. Let $\mathbf{p} \in \mathbb{R}^3$ be a viewing direction pointing to the center of $P_3(D)$. Let $Q(\mathbf{p}, P_3(D))$ be the *view* of $P_3(D)$ from direction \mathbf{p} , *i.e.*, the 2D scatterplot of the orthographic projection of $P_3(D)$ on a plane orthogonal to \mathbf{p} . We can directly measure the quality of $Q(\mathbf{p}, P_3(D))$ by any quality metric M for 2D projections such the ones in Table 1. Hence, we can describe the quality of $P_3(D)$ by a function $M(D, Q(\mathbf{p}, P_3(D)))$ of the viewpoint \mathbf{p} . Note that we can ignore in-plane (around \mathbf{p}) rotations since these do not change the inter-point distances in $Q(\mathbf{p}, P_3(D))$ that all such metrics use.

To analyze M , we sample it over a set of viewpoints $V = \{\mathbf{p}_i | 1 \leq i \leq s\}$, yielding a dataset $\tilde{M} = \{M(D, Q(\mathbf{p}, P_3(D))) | \mathbf{p} \in V\}$. Samples \mathbf{p}_i are uniformly distributed over a sphere using the spherical Fibonacci lattice algorithm [15] with $s = 1000$. Other similar sampling methods can be used, *e.g.* [6, 19]. As outlined in Sect. 2.2, users do not see any information along the viewing direction \mathbf{p} . Hence, our metric $M(D, Q(\mathbf{p}, P_3(D)))$ does not account for points that are occluded along the viewing direction \mathbf{p} . Note that occlusion also happens for quality metrics for 2D projections – in that case, due to overflow.

To explore and compare 2D and 3D projections (and their quality metrics), we implemented an interactive tool having four views (Fig. 1) which we further deploy in

our user evaluation (see Sect. 5). Views (a) and (b) show the 3D projection $P_3(D)$ and the corresponding 2D projection $P_2(D)$ of a dataset D . Views (c,d) allow comparing P_2 and P_3 to decide which is better for the task at hand, as follows.

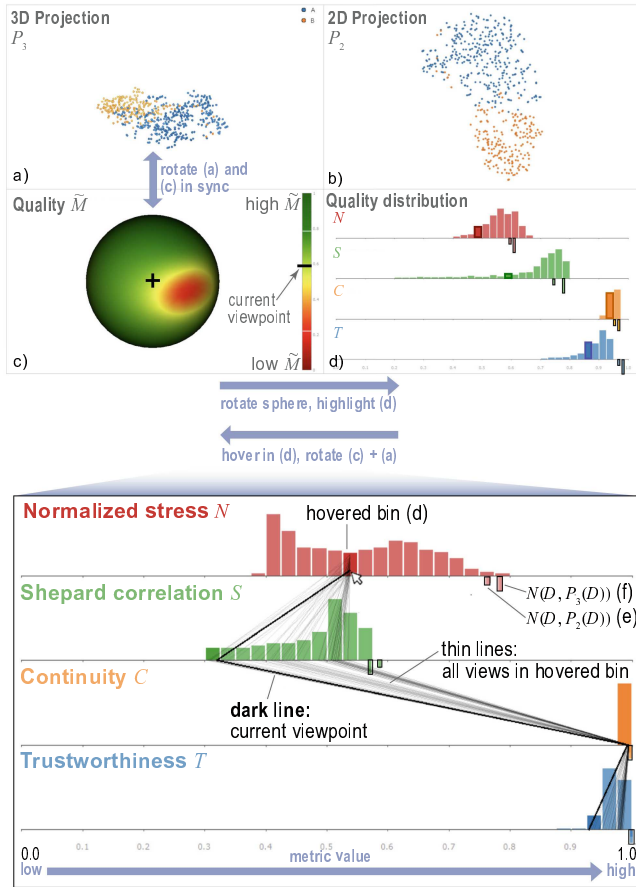


Fig. 1. Tool for exploring 2D and viewpoint-dependent 3D projection quality. Image taken from [8].

Quality Distribution: View (c) displays \tilde{M} (for a user-chosen $M \in \{N, S, C, T\}$) over all directions V by color-coding points \mathbf{p} on a sphere via an ordinal (red-yellow-green) colormap – e.g., red points show viewing directions \mathbf{p} from which \tilde{M} is low. The current viewpoint used in view (a) is at the sphere’s center, see black cross in (b). Rotations of the 3D projection (a) and sphere (b) are linked. Rotating the sphere allows finding viewpoints of high quality \tilde{M} and see how the 3D projection looks from them. Rotating the 3D projection allows users to see how much they can trust any viewpoints, i.e., if \tilde{M} is high. Our viewpoint exploration by sphere rotation is conceptually related

to the mechanism in [10]. However, the latter encodes *explanations* of the different viewpoints of a 3D projection, whereas we encode projection *quality*.

View (d) shows all quality metrics N , S , C , and T for both $P_3(D)$ and $P_2(D)$ using one annotated histogram per metric, as follows (see also inset in Fig. 1 bottom). For each metric, the histogram shows the number of views in V that have quality values \widetilde{M} falling in a given bin (we split the metric range $[0, 1]$ in 40 equal bins). Hence, long bars show \widetilde{M} values reached by many viewpoints; short bars tell \widetilde{M} values that only few viewpoints have. Left-skewed histograms, such as for C and T in Fig. 1 (inset), tell that the 3D projection has high quality from most viewpoints. Right skewed histograms, such as for S in Fig. 1 (inset), tell that the 3D projection has low quality from most viewpoints. Disagreement of the four histograms tells that it is hard to find views that are good in *all* four quality metrics.

Single-Value Metrics: A small, respective large, tick shown under a histogram tells the value of the quality metric for the 2D projection, *i.e.*, $M(D, P_2(D))$, and respectively the value of M computed *directly* on the 3D projection, *i.e.*, $M(D, P_3(D))$. We call these *single-value* metrics to distinguish them from viewpoint-dependent metrics which are, as explained, distributions. The place of the small tick in the $[0, 1]$ range tells how easy is to find viewpoints from which the 3D projection has a higher quality than the 2D projection. For example, Fig. 1 (inset) shows that the small-tick for N , marked (e), is close to the right end of the N histogram, with only two shallow bars right of it – so it is hard to find viewpoints where the 3D projection has a higher N than the 2D projection. The large tick shows why computing a single value for M in 3D can be deceiving: In Fig. 1 (inset), the large tick for N , marked (f), shows a very high value, larger than most per-viewpoint N values for the 3D projection and also larger than the N of the 2D projection (since the large tick is to the right of the small tick). However, as explained, users do not ‘see’ a 3D projection as such, but only 2D viewpoints thereof; as such, they see the distribution of quality values shown by the histogram, which has overall values much smaller than the one indicated by the big tick. Moreover, in some cases, the big tick indicates quality values which can *never* be reached by any of the viewpoints – see the C and T histograms in Fig. 1, inset. Concluding, using single-value 3D quality metrics is misleading.

Visiting Viewpoints in Quality Order: We further link views (c) and (d) by interaction. When the user rotates the viewpoint sphere (c), the bins of the four histograms (d) in which the current viewpoint (crosshair in (c)) falls are rendered in a darker hue. This helps seeing all *four* quality metrics for that viewpoint. Conversely, when the user moves the mouse over a bar in (d), the sphere and 3D projection rotate to a viewpoint that has a quality value within the bar’s bin. Moving the mouse from the bottom to the top of the bar selects views with quality values increasing from the lower end to the higher end of the bin. This allows one to quickly scan, in increasing order, all 3D projection viewpoints with quality values in a given interval.

Combining All Quality Metrics: When hovering over a histogram bar, we draw a Parallel Coordinates Plot (PCP) from the hovered bar to the other three histograms. Given V_0 viewpoints in the hovered bar, the PCP shows V_0 polylines (rendered half-opaque to limit visual clutter), each showing the four quality values for all V_0 viewpoints. A thicker, more opaque, polyline shows the quality of the current viewpoint. This PCP

plot shows how, for a selected range of one quality metric (hovered bar), the other three metrics vary. For example, the PCP in Fig. 1 (inset) shows that all viewpoints with an N value around 0.53 (red hovered bar) have S values that cover almost the entire spectrum of S (since PCP lines fan out from the red bar to almost all green bars except the two rightmost ones), and very similar C and T values (since the lines fan in when reaching the orange and blue histograms respectively). Moving the mouse over the PCP plot selects the closest polyline and makes it the current viewpoint. This allows users to explore the viewpoint-space V using *all four* metric values jointly to find a viewpoint where one, or several, metrics have high values (if such a viewpoint exists).

4 Quantitative Analysis of Viewpoint-Dependent Metrics

We use our tool (Sect. 3) to study how the viewpoint-dependent quality of 3D projections varies among several datasets and projection techniques and also how it compares with the quality of corresponding 2D projections, thereby answering Q1.

4.1 Datasets and Techniques

We explored 6 different real-world datasets and 5 projection techniques, so a total of 30 2D-3D projection-pairs. Datasets come from the benchmark in [14] and have varying numbers of samples and dimensions; have categorical, ordinal, or no labels; and come from different application areas (Table 2). Projection techniques were selected from the same benchmark and include global-*vs*-local, linear-*vs*-nonlinear, approaches, using both samples and sample-pair distances as inputs.

Table 2. Datasets and techniques used to compare 2D and 3D projections. Tables taken from [8].

Dataset	Samples	Dims	Labels	Domain	Technique	Linearity	Input	Locality
AirQuality [44]	9357	13	-	physics	Autoencoders (AE) [5]	non linear	samples	global
Concrete [45]	1030	8	ordinal	chemistry	MDS [37]	non linear	distances	global
Reuters [20]	8432	1000	categories	text	PCA [17]	linear	samples	global
Software [24]	6773	12	ordinal	software	t-SNE [41]	non linear	distances	local
Wine [11]	6497	11	ordinal	chemistry	UMAP [23]	non linear	distances	local
WBC [13]	569	30	categories	medicine				

For each technique-dataset combination (P, D), we computed the 2D and 3D projections $P_2(D)$ and $P_3(D)$ and next measured the single-value metrics $M(D, P_2(D))$ and $M(D, P_3(D))$ and the viewpoint-dependent \bar{M} for the four metrics in Table 1. We compute T and C with $K = 7$ neighbors as in [14, 22, 42]. Our results and source code are publicly available [7].

4.2 Distribution of Metric Values

We first explore how \widetilde{M} varies over all evaluated combinations of datasets, projection techniques, and quality metrics. Figure 2 shows a table (one row per projection) ordered first by dataset and next by projection technique. Each row shows two snapshots of the quality sphere (as in Fig. 2c) for each quality metric, taken from two opposite viewpoints (chosen randomly), so that we can see nearly the whole sphere. We show also the four metric histograms (as in Fig. 1d). Figure 2 leads us to several findings.

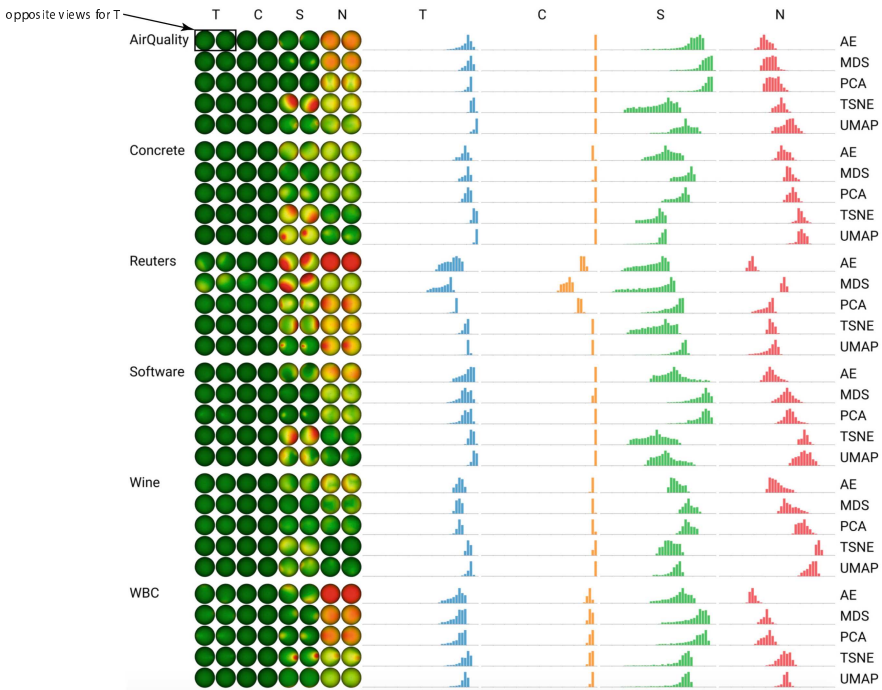


Fig. 2. Exploration of viewpoint-dependent metrics (Sect. 4.2). Image taken from [8].

Metric Ranges: We see that T and especially C have a (very) narrow range close to 1 (maximal quality), also shown by the C and T spheres which are almost fully green. Hence, C and T have very high values *regardless* of the viewpoint. In contrast, S and N vary much more over the viewpoints V – also seen by the larger color variation of the S and N spheres. This suggests that C and T cannot predict good viewpoints since, according to them, nearly all viewpoints are good. Yet, changes within the very small range of C and T could be just as significant as larger changes for the S and N metrics. To test this, we visually compare viewpoints with highest and second-highest, respectively lowest and second-lowest, T and C values, for two datasets and projection techniques (Fig. 3; similar results for all other projection-dataset combinations are in the supplementary material). We see that, even though C for AirQuality only differs

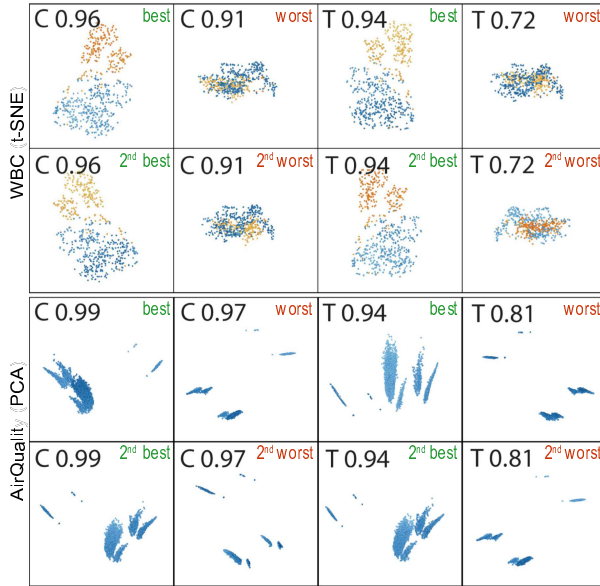


Fig. 3. Comparison of first and second best-*vs* first and second worst viewpoints of 3D projections using T and C (Sect. 4.2).

by 0.02 between the best and worst values; and T for WBC only differs by 0.22 for WBC, there is a clear difference in perceived quality in these projections in terms of separation of points into different clusters. The same pattern holds for the second-best and second-worst viewpoints which are visually almost identical, and have the same quality values as, the best, respectively worst, viewpoints.

Metric Distributions: In Fig. 2, we see no quality correlation with *datasets* but rather with projection *techniques*. Given this and our previous observation, we recreate Fig. 3 to use the *actual* ranges of the metric histograms and also to group projections by technique rather than dataset – see Fig. 4. While we now cannot compare the actual x positions of the metric histograms, we can (a) see much better the *shapes* of these histograms; and (b) how quality metrics correlate with projection techniques.

In Fig. 4, we first see why viewpoint-dependent metrics are important: Viewpoints are non-uniformly spread over the (wide or narrow) metric ranges; and Fig. 3 showed that small metric values can correspond to big visual differences. Hence, small metric-value changes *are* important predictors of visual quality. Secondly, we see that the histograms for T , C , and S are left skewed, *i.e.*, have many long bars for higher metric values, with only a few exceptions (N , Airquality MDS; S , Software AE). Hence, users should not have a problem in finding high-quality-metric viewpoints in 3D projections, which partially counters the argument in previous papers that viewpoint selection is hard for 3D projections [28]. We further show that such high-quality-metric viewpoints are indeed seen as high-quality by users in Sect. 5. Thirdly, if our four metrics inherently capture ‘quality’, their shapes should be similar (at least for the same dataset-technique

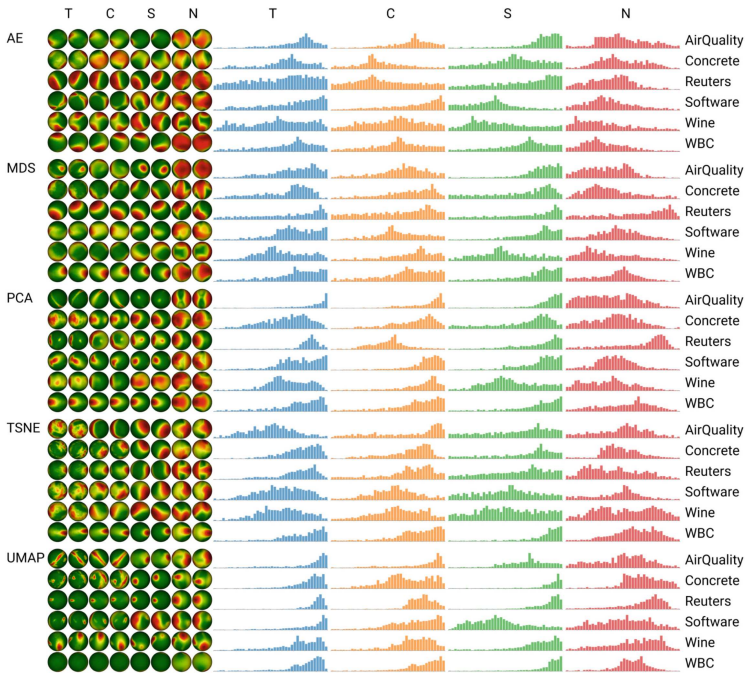


Fig. 4. Refinement of Fig. 2 with metrics clipped to their actual ranges and projections grouped per technique (Sect. 4.2). Image taken from [8].

combination). Figure 4 shows that this so in most cases for T , C , and S . In contrast, the N metric has quite different histogram shapes in most cases, tending to show lower values. This correlates with earlier papers [14, 16] indicating that N is not a good way to assess projection quality. We further explore how this correlates with the actual quality perception of users in Sect. 5.

Finally, we study Fig. 4 regarding projection *techniques*. We see that UMAP has more ‘peaked’ and left-skewed histograms than the other four techniques – see also the amount of green in its sphere snapshots. So, UMAP has many high-quality viewpoints, so picking a good viewpoint with UMAP is easier than for the other techniques. Also, UMAP yields higher quality than the other techniques for nearly all datasets and metrics. Hence, UMAP is the best technique to use for 3D projections from the perspective of our four quality metrics. Interestingly, Figs. 3 and 4 show that t-SNE does not yield higher quality values (spread over all viewpoints), nor many high-quality views. This is in line with earlier findings [39, 40] that noted that t-SNE generates ‘organic’, round, clusters which tend to fill the projection space. Such clusters will *overlap* in most 2D views of a 3D projection, *i.e.*, yield low values for our viewpoint-dependent metrics. Simply put: t-SNE may be best-quality for 2D projections [14] but not for 3D ones.

Comparing 2D and 3D Metric Values: Figure 5 next compares the quality of our 3D projections P_3 with their 2D counterparts P_2 . The top table shows the single-value metrics for the 2D and 3D projections, *i.e.*, $M(D, P_2(D))$ and $M(D, P_3(D))$, averaged

over all tested datasets and projection methods. We see that the 3D metrics are slightly higher than the 2D ones. As argued earlier, this is not relevant, since users do not ‘see’ 3D projections but only 2D orthographic views thereof. The stacked barchart shows, for each dataset, projection technique, and metric, the fraction of the total s computed viewpoints where 3D metrics exceeded the quality of the 2D projection. Since we stack the bars of 5 projection techniques atop each other, 20% in the figure means all views of a single technique-dataset pair. For T , viewpoints of 3D projections outperform 2D projections only in a few cases. For all other metrics, many viewpoints do this: For N on *AirQuality*, over 50% of the 3D viewpoints have higher quality than the 2D projection. For all datasets and all metrics except T , we see multiple, differently colored, non-zero-height, bars stacked atop each other. So, many techniques create 3D projections having viewpoints that score better than their 2D counterparts. As we can find such viewpoints using our tool (Sect. 3), 3D projections can effectively provide higher-quality results than 2D projections. Our user study in Sect. 5 further analyzes this.

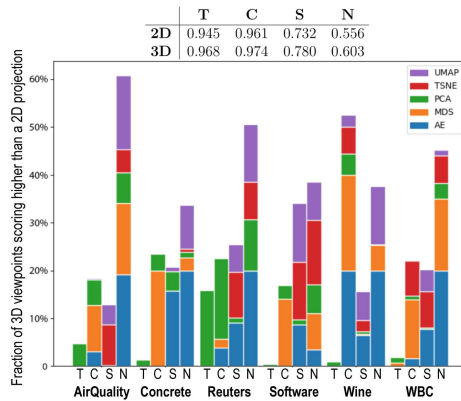


Fig. 5. Quality metrics of 2D vs 3D projections for 6 datasets, 5 projection techniques (Sect. 4.2). Image taken from [8].

There is no *dataset* where all techniques score better in 3D than 2D for any metric – that would be a bar in Fig. 5 with five stacked fragments each larger than 12.5% (since a 20% bar indicates that *all* views of a technique-dataset pair score better in 3D than 2D). Yet, some *techniques* score consistently better in 3D for some metrics: For all but one of the AE projections (blue bars), almost all viewpoints (20%) have better N than the 2D projection – so, if we trust N , 3D AE projections are better. For PCA and t-SNE (green and red), we see far fewer viewpoints with better N than the 2D projection. Also, UMAP (purple bars) is better in 3D only in terms of S or N , but rarely for C and never for T . For MDS (orange bars), 3D viewpoints outperform 2D projections mostly in C .

5 User Evaluation of Proposed Quality Metrics

We further analyze how our proposed metrics can predict good views of 3D projections by conducting a user study.

Projections and Datasets: To keep the study duration short (10–15 min), we picked a subset of the 30 (D , P) pairs used in our quantitative evaluation. This contains projections which (1) have discernible structure in terms of separated point-groups having the same labels; (2) finding a good viewpoint, showing strong visual cluster separation in the 3D projection, is not trivial; (3) the datasets have over 1000 samples, so their projections are arguably complex enough. Our subset contains six pairs: (Wine, t-SNE); (Wine, PCA); (Concrete, t-SNE); (Reuters, AE); (Reuters, t-SNE); and (Software, t-SNE). Each pair contains a 3D projection and the corresponding 2D projection.

Study Design: We aim to discover how users reason about the quality of views of a 3D projection in comparison to a 2D projection, and how this correlates to our metric values and findings in Sect. 4. Our study proceeded as follows (for full details, see supplementary material). First, we explained our tool (Fig. 1) to users. We next explained to users that T and C measure the quality of neighborhood preservation; that N and S measure how distances in the projection reflect data distances; and that all metrics range between 0 (worst) and 1 (best). We did not elaborate on the exact metric definitions (Table 1) since such knowledge was not needed for the study’s tasks. Further, we asked users to search for viewpoints in 3D projections which show well-separated point groups that have similar colors (labels). That is, users were *implicitly* tasked with finding views that have minimal overlap for different clusters and show most of the data structure in terms of class separation.

Usage of Metrics: To study how metrics correlate with the users’ choices of good viewpoints, we split the study into two parts. For the first three projections, further called the *blind* (B) set, users had to go through the first three (of the six) projection-pairs and select, for each pair, 3 different viewpoints of the 3D projection that they deemed good by using only views (a) and (b) in Fig. 1 – that is, without seeing the metrics. For the remaining three projections, further called the *guided* (G) set, users had to accomplish the same task but they also had access to the metric views (Fig. 1c, d). We explicitly stressed that the metric values are just *suggestions* for interesting viewpoints, as these metrics do not measure class and cluster separation (which the task aims to maximize) but only local structure preservation. We randomized the order in which users saw the projections so that projections in the B and G sets differed for each user. For each viewpoint users picked, we also asked whether they preferred it to the 2D projection. Finally, we asked users to tell their agreement on a 7-point Likert scale on whether a 3D projection, examined from various viewpoints, better displays data structure than a 2D projection. ‘We explained the users that ‘data structure’ in this context means seeing reasonably-well separated clusters of points (which we know to exist in the studied datasets).

5.1 Study Results

From the invited 50 people, 22 downloaded our tool and performed the study. At the end of the study, our tool saved the projections selected by the users as ‘good’ (a total of 66 per projection) and also the corresponding viewpoint-dependent quality metrics. These data were anonymously sent by the participants back to us. We next analyze these data to study how user preferences correlate with the computed quality metrics.

A. What did the Users Select and Prefer? Figure 6 show several snapshots, arbitrarily selected from all those generated during the study, of views selected by users in the evaluation. We group these on whether users preferred the 2D projection or a view of the 3D projection. Per dataset-projection pair, we show five images:

- the *2D projection* shows the unique, static, 2D projection for the dataset. This is what users had to compare with to decide if they prefer a 3D viewpoint and, if so, which one that would be;
- the *2D preference* shows the best 3D viewpoint users found in cases where they still preferred the 2D projection, for both the B and G conditions;
- the *3D preference* shows the best 3D viewpoint users found in cases where they preferred it against the 2D projection, for both the B and G conditions.

Figure 6 leads us to the following qualitative observations. In all cases, users seem to have understood well the experimental task – that is, find viewpoints of 3D projections where the classes in the projection are well separated. Indeed, in all images in columns 1, 2, 4, and 5 of Fig. 6, we see that the respective viewpoints show more than reasonable visual separation of the classes in the projection.

If we compare the guided set (G) with the blind set (B) images, we cannot immediately say that one of the two sets consistently shows better visual separation than the other set. Hence, analyzing quality metrics of the two sets – discussed further below – is important to gain more fine-grained insights on which of the two sets scored higher.

Similarly, the ‘2D preference’ and the ‘3D preference’ snapshots (that is, columns 1 and 2, and 4 and 5 respectively in Fig. 6) look quite similar in terms of visual separation of the classes. Still, in some cases, users preferred the 2D projection and in others they preferred their own choice of a 3D viewpoint. This further suggests that analyzing quality metrics is of added value to understand the users’ selections.

The best-chosen viewpoints for the guided set (G) and blind set (B) are quite similar for the same dataset. This tells us that showing (or not) the quality metrics did not influence the users massively in deciding what they find to be a good viewpoint of a 3D projection. Of course, finding such a good viewpoint takes longer in the B condition since one does not have hints about where, in the viewpoint space, such a viewpoint is. Hence, if showing the quality metrics per viewpoint leads to very similar viewpoints as in the B condition, and finding such viewpoints is faster in the G condition (since one directly sees the metrics telling which viewpoint is good or not), this brings evidence towards the added-value of the quality sphere and quality distribution widgets.

B. Do Users Prefer Viewpoints with high Metric Values? Figure 7 shows the histograms of each metric and projection-pair in the evaluation set. The three box plots show the distributions of quality values in the actual histogram (H); for viewpoints in the blind set (B); and for viewpoints in the guided set (G). Comparing H and B , we see that, in almost all cases, users chose viewpoints with high values (for all metrics) in the B condition, *i.e.* without seeing these metric values. This is a first sign that quality metrics do correlate with what users see as good viewpoints.

Comparing the histograms H and G , we see that, in the G condition, users selected viewpoints with a higher quality than in the B condition. This must be interpreted with care. On the one hand, users could have been biased by the quality metrics displayed

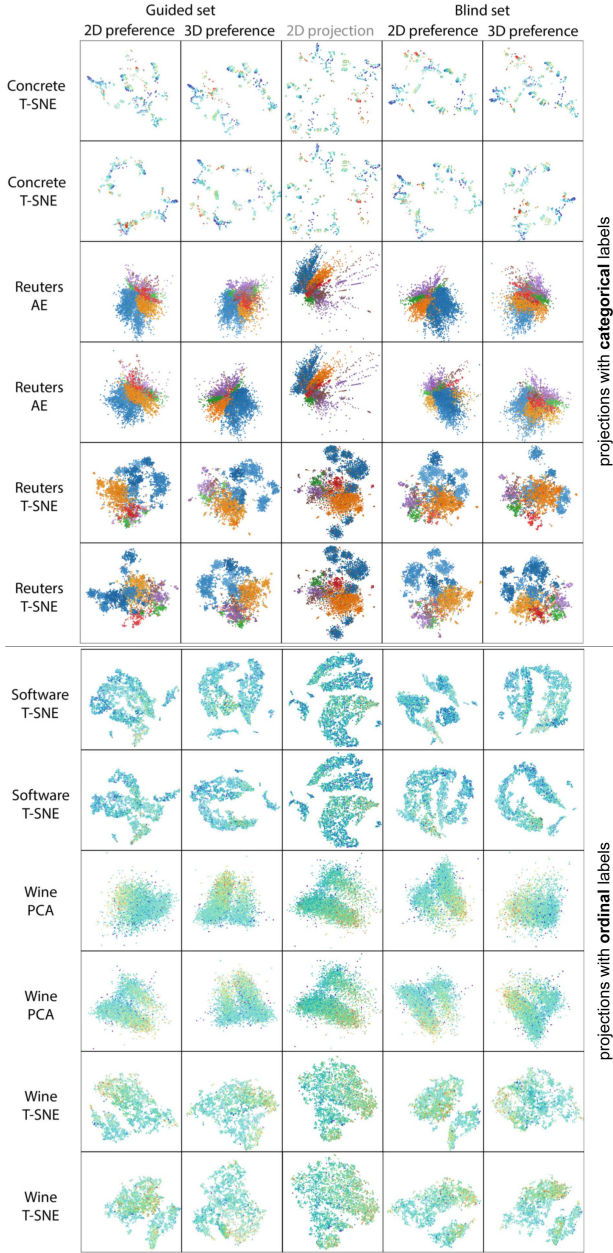


Fig. 6. Views selected by users as best to view a 3D projection. Projections are colored by their categorical labels (top), respectively ordinal labels (bottom; *Wine*: 11 values; *Software*: 12 values). The ‘guided set’ respectively ‘blind set’ labels denote the two conditions of our experiment. The ‘3D preference’ and ‘2D preference’ labels indicate whether the users preferred the selected 3D view or rather the 2D projection. See Sect. 5.1A.

during the *G* condition. On the other hand, as explained, we stressed that metrics are only hints for finding interesting viewpoints and explicitly told users that, if they find other viewpoints as being better, they should ultimately go by their own preference. As such, the *H-G* comparison suggests us that quality metrics are useful predictors of users’ preferences of good viewpoints. Hence, showing the metric widgets during actual exploration of 3D projections can be useful since it helps users find high-quality viewpoints (*G* boxplots show clearly that users selected the high-end of the quality ranges) and users find high-quality viewpoints to be good (correlation of *H* and *B* boxplots). Yet, the strength of this correlation is not equal for all (dataset, projection) pairs.

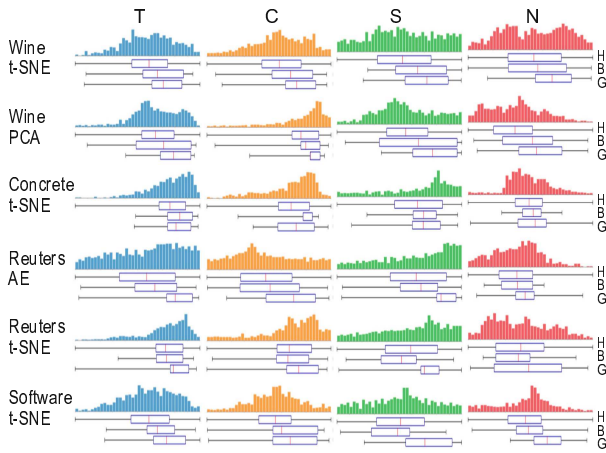


Fig. 7. Distribution of metric values for all viewpoints (histograms and boxplots *H*), viewpoints in the blind set (boxplots *B*), and viewpoints in the guided set (boxplots *G*). Image taken from [8].

Table 3 refines these insights by showing the p-values of a t-test (equal variance, one tail) for each projection, all four metrics. The test checks whether the average metrics for the selected views in the *B*, *G*, and combined (*B+G*) conditions are significantly higher than the average metrics over all viewpoints *V*. We see that, for nearly all cases, this is so for the guided set *G*. For the combined set *B+G*, this is slightly less often so.

C. Do Users Mostly Prefer 3D or 2D Projections? Figure 8 shows the percentage of 3D viewpoints that users preferred over the 2D projection of the same dataset for both the *B* and *G* conditions. In the *G* condition, users preferred the 3D projection over the 2D projection, and did so *more* than in the *B* condition. This, and the findings from Fig. 7 (users tend to pick high-quality views in the *G* condition), tell us that the metric widgets add to the user-perceived value of 3D projections. The 3D-*vs*-2D preference in the *G* condition was the strongest for the *Wine* dataset. For this dataset, we also found the strongest increase in metric values in the *G vs* the *B* condition (Table 3). For *Reuters* and *Software*, Fig. 8 shows a much smaller 3D-*vs*-2D preference in the *G* condition and also little correlation between metric values and user-perceived quality (Table 3). This

Table 3. p values of t-testing whether the average metrics for user-selected viewpoints in the blind (B), guided (G), and both (B+G) sets are significantly higher than average values for all viewpoints V . Significant values ($p < 0.05$) are in bold. Table taken from [8].

	T			C			S			N		
	B	G	B+G	B	G	B+G	B	G	B+G	B	G	B+G
Wine t-SNE	.029	.001	<.001	.01	<.001	<.001	.005	<.001	<.001	.24	.001	.003
Wine PCA	.015	<.001	<.001	.121	.041	.025	.003	<.001	<.001	<.001	<.001	<.001
Concrete t-SNE	.003	.035	.001	.004	.111	.004	.159	.156	.08	.148	.049	.029
Reuters AE	.087	<.001	<.001	.232	<.001	<.001	.214	<.001	<.001	.448	.025	.071
Reuters t-SNE	.418	.025	.059	.595	.029	.111	.935	.009	.23	.679	.056	.197
Software t-SNE	.04	<.001	<.001	.088	<.001	<.001	.689	<.001	.002	.279	<.001	<.001

further reinforces our claim that metric values are a good indicator of user-perceived 3D projection quality. For *Software*, we see that 3D was preferred much less than 2D in the G condition. We see in Fig. 7 that this is the only case of the six (P, D) studied combinations where *all* four quality metrics are right-skewed, *i.e.*, have only few 3D viewpoints with high metric values. This is likely so since *Reuters* is a much higher-dimensional dataset than all other studied ones (1000 vs a few tens of dimensions), so it is a harder dataset to project well. In such cases, it is indeed hard to argue about the advantages of a 3D projection.

Table 4. p values for an equal variance, one-tail t-test of whether the 3D views users preferred over the 2D projection have significantly higher metric quality than the views users did not prefer over the 2D projection, for each metric-projection pair. Significant values ($p < 0.05$) indicating correlation are displayed in bold.

	T	C	S	N
Wine t-SNE	0.104	0.037	0.258	0.302
Wine PCA	0.058	0.037	0.104	0.308
Concrete t-SNE	0.556	0.341	0.321	0.186
Reuters AE	0.916	0.954	0.877	0.455
Reuters t-SNE	0.411	0.317	0.557	0.361
Software t-SNE	0.945	0.695	0.93	0.842

To further investigate the observed correlation between measured quality and user-perceived quality, 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. Table 4 shows the p -values of this test. Here, only for the *Wine* dataset projections and the *C* metric, do we see a significant increase in values compared to 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, especially for the *Reuters* and *Software* dataset projections. However, this negative correlation is only statistically significant ($p > 0.95$) for the C metric in the *Reuters AE* projection. Combining these results with our earlier observations, we can say that, overall, our metrics (and their display in the G condition) do help users in finding good viewpoints of 3D projections with higher *measured* quality than 2D projections, but such views are not necessarily *always* preferred by users compared to the respective 2D projections.

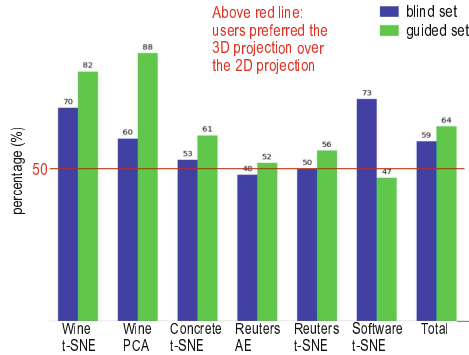


Fig. 8. Percentage of cases where users preferred viewpoints of 3D projections vs 2D projections. Image taken from [8].

We next consider the last question we asked our participants – whether, in the end, they preferred a 3D interactive projection to a static 2D projection for the task of assessing data structure. *All* 22 participants responded with a value on the positive side of the scale (4 or higher), with an average of 5.94. This is additional evidence that, when aided by interaction and by tools that help finding interesting viewpoints (like our quality metrics), 3D projections are a viable alternative to classical 2D projections.

5.2 Explaining the Viewpoints of 3D Projections

In our study, projections were color-coded by the data class labels. This matches the task asked to the users, *i.e.*, finding viewpoints from which same-label points would group visually well in the 3D projection. Figure 6 showed that, in most cases, users could find such good viewpoints. An important follow-up task for someone using such viewpoints could be to *explain* the visible point groups in terms of the underlying data dimensions (question Q4, Sect. 1). Indeed, if one found that a dataset is well-separated into samples of different classes, finding out which dimensions contribute most to the separation of each perceived sample group is a frequently-asked question in information visualization. Conversely, having a (2D or 3D) projection well-separated into visual groups is by itself not useful if we don't know what the groups mean.

We address the above by using two projection explanation techniques. The first one – *variance* explanation – colors groups of points in a projection by the data dimension which varies least over them, thus, the dimension which explains best why such points are similar [12]. The second one – *dimensionality* explanation – colors point groups by the local intrinsic dimensionality of the respective data points, *i.e.*, how many of the data dimensions are needed to explain 90% of the data variance [39]. This explanation is useful in understanding how many dimensions we need to understand to explain a given local pattern in the projection. For each of the dataset-projection combinations and selected viewpoints in Fig. 6, we render both 2D and 3D projections using the above explanations. We omit the Reuters dataset (Fig. 6, bottom four rows) based on our earlier finding that the 3D projection of dataset has hardly any good viewpoints (Sect. 5.1).

Variance Explanation: Fig. 9 shows the variance explanations. We see that virtually all images in this figure show a good separation of the projection into different-color groups. That is, the viewpoints deemed good by users to visually separate points having different *classes* are also good to explain these separated groups by the *data dimensions*. This is a non-trivial insight since it implies that, if one can find the desired class-based visual separation, then one can next explain what makes the respective classes separate from each other. For example, consider the *Wine* dataset projected with PCA (Fig. 9, rows 5 and 6). The first and last two columns of these rows show together 8 different viewpoints of this dataset (selected as ‘good’ by different users). When colored by explanations (Fig. 9), we see much easier than in Fig. 6 that these viewpoints separate well the 3D projection in three large groups explained by the red, yellow, and pink color-coded dimensions (wine residual sugar, volatile acidity, and alcohol) plus a smaller fourth group of lower explanation confidence using the green dimension (wine volatile acidity). Importantly, this good separation that the variance explanation shows was *not* seen by the users when they chose the respective viewpoints the tool in the study did not include explanations. This supports our hypothesis that a good *separation* of a 3D projection into visual point clusters favors next a good *explanation* of these clusters in terms of dimension variance.

For the *Wine* (projected by PCA) rows in Fig. 9, we also see that the eight selected viewpoints show very *similar* insights when using the variance explanation. This is also interesting when considering that users did not see this explanation when selecting good viewpoints. This makes us hypothesize that viewpoints which are good for visual separation in a 3D projection will also be good for examining the explained projection. That is, users can use our viewpoint-selection widgets (Sect. 3) also to select good viewpoints for viewing the explained projection.

Thirdly, if we compare the aforementioned eight viewpoints with the 2D projection of *Wine* (using PCA) in Fig. 9, we see that the explanation images are very similar. This can clarify why in some cases users would prefer the 2D projection whereas in others their self-selected viewpoint of the 3D projection. Indeed, if differences between these are small, then it is more likely that either of them can be preferred.

Finally, we consider the task of finding a good viewpoint to *understand* a 3D projection. By understanding, at a high level, we mean being able to separate the projection into point groups which one can reason about in terms of their dependent or independent

variables. In our actual user study, this took the form of separating the projection into groups which are distinct from each other and also have different dependent variable (label) values. As visible from Fig. 6 (bottom), this is harder to do when one has *ordinal* label values with many values than when having a small set of categorical values (Fig. 6 (top)). In contrast, when using variance explanations, understanding the projection (as *e.g.* three groups determined by three data dimensions for the *Wine* dataset) becomes much easier.

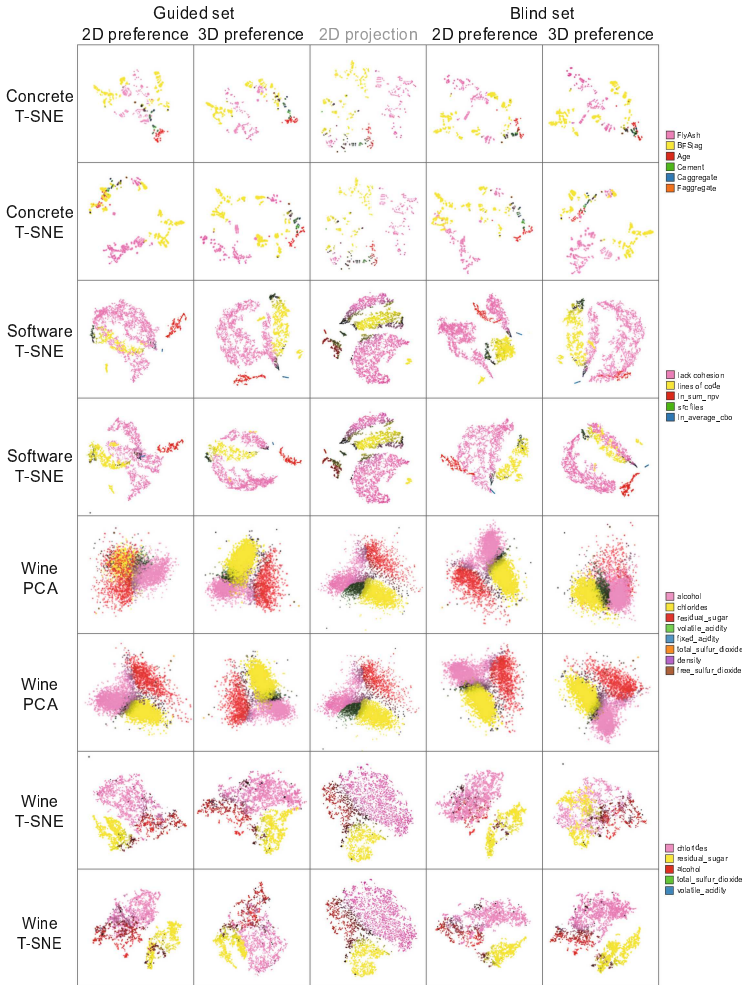


Fig. 9. Views of the 3D projections shown in Fig. 6 color coded by variance explanation (Sect. 5.2).

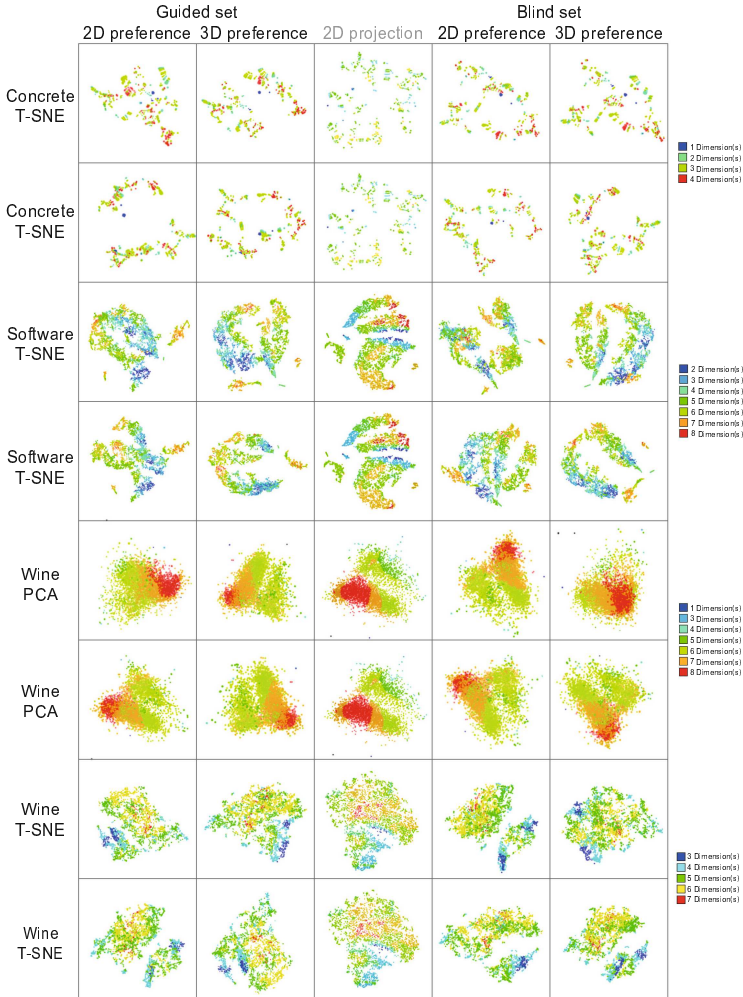


Fig. 10. Views of the 3D projections shown in Fig. 6 color coded by local dimensionality explanation (Sect. 5.2).

Dimensionality Explanation: Fig. 10 repeats the design of Fig. 9 but using local dimensionality instead of dimension variance explanations. All our earlier observations made on Fig. 9 also hold for local dimensionality: The user-selected good viewpoints (for a given dataset-projection technique pair) show quite similar explanations among themselves and also with the 2D projection; and these explanations help us understand how to ‘parse’ the projection beyond the color-coding by label values shown in Fig. 6 (bottom). The two explanations can be also *combined* to better understand the 3D projection. Take, for instance, the *Wine* (projected by PCA) – row 5, second column from the left. The dimensionality explanation (Fig. 10) shows a red zone in the left of the pro-

jection – that is, about 8 dimensions are needed to explain the projection in this area. The variance explanation of the same viewpoint in Fig. 6 shows that the projection is explained here by the pink variable (alcohol). Hence, we know that we need seven other dimensions to understand well what makes samples similar to each other in this part of the 3D projection. While seven additional dimensions seem a lot, let us consider using the 2D projection to understand this dataset. The dimensionality explanation (Fig. 9, rows *Wine*, middle column) shows a much larger red spot in the same left area than in the abovementioned 3D projection view. So, if we used the 2D projection, we would have *more points* that are hard to explain due to their high local dimensionality. This is due to the fact that a 3D projection spreads the points better (as it has 3D space to its avail) than a 2D projection. The amount of intrinsically high-dimensional points is the same, since it is determined by the dataset. However, depicting them in 3D allows finding viewpoints from which fewer such points are visible (due to the ‘depth’ dimension) – thus, viewpoints which are easier to explain. Interestingly, the *occlusion* inherently present in a 3D projection works in our favor rather than against us in this case: From certain viewpoints, certain projection points (which are hard to explain due to their high intrinsic dimensionality) will not be visible in the 3D projection due to occlusion, therefore yielding an image which is easier to explain using e.g. variance explanations.

6 Discussion

Our results answer our questions Q1-Q4 as follows:

Q1: Simply *reusing* 2D projection quality metrics for 3D projections is misleading. These metrics will score higher values than their 2D counterparts but the respective 3D projections appear massively different from different viewpoints. To address this, we need viewpoint-dependent quality metrics. Using such viewpoint-dependent quality metrics, as proposed in this paper, helps assessing the quality of 3D projections as these show significant variations between viewpoints of different projection techniques for different datasets. Such metrics are as simple and fast to compute as their 2D counterparts.

Q2: Viewpoint-dependent quality metrics can reach *higher values* than their 2D counterparts, albeit for a small number of viewpoints. Such viewpoints can be easily found using our proposed interactive visual metric-and-projection exploration. Hence, 3D projections can generate 2D images which are of higher quality than typical static 2D projections with minimal effort.

Q3: Users’ definition of “good viewpoints” (for the task of separating a 3D projection into distinct same-label clusters) correlates well with high values of our viewpoint-based quality metrics. This correlation is little influenced by the projection technique but more so by the dataset being explored. Separately, enabling visual exploration of the quality metrics increases the users’ preference for a 3D projection *vs* a 2D one for performing the same task. Summarizing the above, using our quality metrics during the visual exploration helps using 3D projections in multiple ways.

Q4: Adding variance and dimensionality explanations helps interpreting the visible point-groups present in viewpoints of 3D projections. Once good viewpoints are found,

we observed that their explanations are also easy to read. In contrast to earlier work [40] where users searched for good viewpoints based on the perceived quality of the explanations, we search for these viewpoints based on our computed quality metrics which, as already explained, is a fast process.

Limitations: Computing quality metrics (Table 1) is linear in the number of viewpoints s and dataset points N . For $s = 1000$ and our studied datasets (N in the thousands), this takes a few minutes. This is not an issue for our study goal as we can precompute all the metric values for all tested datasets prior to the actual study. Using our metric-based exploration tool (Fig. 1) at interactive rates on *unseen* datasets would require faster metric computation. As recently shown in separate work [38], this can be trivially implemented by GPU parallelization.

Our limited study sample – 6 datasets, 5 projection techniques, and 22 users who evaluated only 6 of the 30 dataset-projection combinations – could not lead to statistically significant results in terms of whether user preference (of selected views of 3D projections) correlates with the measured quality metrics for these views. That is, we cannot say, with the current information we have, that the respective metrics *predict* quality as perceived by users. Apart from the relative small study sample, this is however not surprising. Our quality metrics are essentially viewpoint-dependent extensions of *local* quality metrics used for 2D projections, namely T , C , S , and N . As extensively covered by earlier work [14, 26], such local metrics only gauge how well a projection captures data structure over local neighborhoods. Our evaluation task asked users a higher-level question, namely, to rank projections in terms of how well they succeed in separating the visualized dataset into distinct visual clusters having different labels. While, of course, a poor-quality projection will not be able to do so, it is not immediate that a good-quality projection (in the sense of our metrics) will *always* achieve this separation. To capture this, more refined, task-specific, quality metrics are needed. However, we argue that, whichever these metrics should be, they should always be measured in a *viewpoint-dependent* way, following our proposal in Sect. 3).

Given the above, the practical added-value of our visual metrics-exploration tool (Sect. 3) can be summarized as follows. The tool can be used to easily find high-metric-value viewpoints for a 3D projection. All other aspects considered, these are the best viewpoints that one should consider for exploring a given 3D projection. However, whether any of these viewpoints is indeed ideal for a given exploration task involving such a 3D projection – as opposed to using other exploration techniques for the same dataset and task – is not answered by our metrics. Simply put, our metrics and visual exploration tool simplify the task of finding a small set of viewpoints to be further examined, but do not tell if these are effective in answering the exploration task.

We have shown that adding explanations to high-quality viewpoints of 3D projections is an effective way in telling *what* the respective viewpoints mean in terms of the data dimensions (Sect. 5.2). However, earlier-identified limitations of such explanations still remain valid for their usage for our 3D viewpoint-dependent explanation [39, 40]. In detail, such explanations can, by construction, only highlight a *few* data-dimensions (under roughly 10) to explain a given 3D projection viewpoint. This means that they are less effective for high-dimensional datasets having hundreds of dimensions and/or datasets in which dimensions do not have a clear significance, such as latent dimensions

extracted by machine learning techniques. However, recent developments in explanatory techniques [38] show the ability to handle datasets with higher dimension counts. Studying whether such explanation techniques would integrate well with our viewpoint-dependent 3D projection exploration is a low hanging fruit for future work.

Using more datasets to study this correlation can bring valuable insights and be used to improve our visual tool to recommend good viewpoints as a function on these traits. Also, using more quantitative tasks to gauge how users select suitable 3D viewpoints (and measuring the time needed for this) is an important direction for future work.

As mentioned, our findings are restricted to the five projection techniques we studied. However, as [14], average quality metrics evaluated on a total of 45 projection techniques show quite similar values. As such, we believe that our findings – and the added-value of our proposed visual tool for choosing good viewpoints for 3D projections – will hold for most, if not all, such techniques.

7 Conclusions

We have presented a set of techniques for the measurement of the quality, exploration, and explanation of 3D projections of high-dimensional data. We defined (and measured) quality as a viewpoint-dependent function that uses well-known quality metrics for 2D projections applied on the viewpoints of a 3D projection. We showed that our viewpoint-dependent metrics capture the visual variability in 3D projections much better than the naive extension of single-value 2D quality metrics to 3D projections. Moreover, we proposed a visual interactive tool for finding high-quality viewpoints of 3D projections as defined by the above metrics. We showed, by means of a user study involving 22 participants, 6 dataset-projection technique combinations, that our proposed metrics do indeed generate good viewpoints for a generic task related to projections – finding well-separated same-label point groups. Our study also showed that users agree well with the predictions of the metrics both when the latter are shown (and when not) during the exploration. Also, our study showed that, when supported by our exploration tool, users prefer 3D projections as compared to classical, static, 2D projections. Finally, we showed that good viewpoints – as assessed by our quality metrics – lead to effective explanations of the explored datasets in terms of their data variables.

Several future work directions exist. First, we plan to extend our evaluation with more datasets, tasks, and projections to find more accurately when, and how much, 3D projections can bring added value atop their 2D counterparts. We also aim to extend our evaluation to involve more sophisticated visualizations of high-dimensional data using pre-conditioned 3D projections that display data using density estimation rather than raw scatterplots such as the Viz3D system [2]. Last but not least, we plan to include the explained views of 3D projections in a controlled user study to gauge more accurately their effectiveness in explaining 3D projections seen from different viewpoints.

References

1. Albuquerque, G., Eisemann, M., Magnor, M.: Perception-based visual quality measures. In: Proc. IEEE VAST, pp. 11–18 (2011)
2. Artero, A.O., de Oliveira, M.C.F.: Viz3D: Effective exploratory visualization of large multi-dimensional data sets. In: Proc. SIBGRAPI (2004)
3. Aupetit, M.: Visualizing distortions and recovering topology in continuous projection techniques. *Neurocomputing* **10**(7–9), 1304–1330 (2007)
4. Aupetit, M.: Sanity check for class-coloring-based evaluation of dimension reduction techniques. In: Proc. BELIV, pp. 134–141. ACM (2014)
5. Bank, D., Koenigstein, N., Giryas, R.: Autoencoders (2020). [arXiv:2003.05991](https://arxiv.org/abs/2003.05991) [cs.LG]
6. Camahort, E., Lerios, A., Fussell, D.: Uniformly sampled light fields. In: Proc. EGSR, pp. 117–130 (1998)
7. Castelein, W., Tian, Z., Mchedlidze, T., Telea, A.: Viewpoint-based comparison of 2D and 3D projections – datasets, software, and results (2022). https://github.com/WouterCastelein/Proj3D_views
8. Castelein, W., Tian, Z., Mchedlidze, T., Telea, A.: Viewpoint-based quality for analyzing and exploring 3D multidimensional projections. In: Proc. IVAPP. SCITEPRESS (2023)
9. Chan, Y., Correa, C., Ma, K.L.: Regression cube: a technique for multidimensional visual exploration and interactive pattern finding. *ACM TIS* **4**(1) (2014)
10. Coimbra, D., Martins, R., Neves, T., Telea, A., Paulovich, F.: Explaining three-dimensional dimensionality reduction plots. *Inf. Vis.* **15**(2), 154–172 (2016)
11. Cortez, P., Cerdeira, A., Almeida, F., Matos, T., Reis, J.: Modeling wine preferences by data mining from physicochemical properties. *Decision Support Syst.* **47**(4), 547–553 (2009). <https://archive.ics.uci.edu/ml/datasets/wine+quality>
12. da Silva, R., Rauber, P., Martins, R., Minghim, R., Telea, A.C.: Attribute-based visual explanation of multidimensional projections. In: Proc. EuroVA (2015)
13. Dua, D., Graff, C.: Wisconsin breast cancer dataset (2017). [https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+\(Diagnostic\)](https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic))
14. Espadoto, M., Martins, R., Kerren, A., Hirata, N., Telea, A.: Toward a quantitative survey of dimension reduction techniques. *IEEE TVCG* **27**(3), 2153–2173 (2019)
15. Gonzalez, A.: Measurement of areas on a sphere using Fibonacci and latitude-longitude lattices. *Math. Geosci.* **42**(1), 49–64 (2010)
16. Joia, P., Coimbra, D., Cuminato, J.A., Paulovich, F.V., Nonato, L.G.: Local affine multidimensional projection. *IEEE TVCG* **17**(12), 2563–2571 (2011)
17. Jolliffe, I.: *Principal Component Analysis*. Springer (2002)
18. Lespinats, S., Aupetit, M.: CheckViz: sanity check and topological clues for linear and non-linear mappings. *CGF* **30**(1), 113–125 (2011)
19. Levoy, M.: Light fields and computational imaging. *Computer* **39**(8), 46–55 (2006)
20. Lewis, D., Shoemaker, P.: Reuters dataset (2021). <https://keras.io/api/datasets/reuters>
21. Martins, R., Coimbra, D., Minghim, R., Telea, A.C.: Visual analysis of dimensionality reduction quality for parameterized projections. *Comput. Graph.* **41**, 26–42 (2014)
22. Martins, R., Minghim, R., Telea, A.C.: Explaining neighborhood preservation for multidimensional projections. In: Proc. CGVC. pp. 121–128 (2015)
23. McInnes, L., Healy, J., Melville, J.: UMAP: Uniform manifold approximation and projection for dimension reduction (2018). [arXiv:1802.03426v2](https://arxiv.org/abs/1802.03426v2) [stat.ML]
24. Meirelles, P., Santos, C., Miranda, J., Kon, F., Terceiro, A., Chavez, C.: A study of the relationships between source code metrics and attractiveness in free software projects. In: Proc. SBES, pp. 11–20 (2010)

25. Motta, R., Minghim, R., Lopes, A., Oliveira, M.: Graph-based measures to assist user assessment of multidimensional projections. *Neurocomputing* **150**, 583–598 (2015)
26. Nonato, L., Aupetit, M.: Multidimensional projection for visual analytics: linking techniques with distortions, tasks, and layout enrichment. *IEEE TVCG* (2018). <https://doi.org/10.1109/TVCG.2018.2846735>
27. Paulovich, F.V., Nonato, L.G., Minghim, R., Levkowitz, H.: Least square projection: a fast high-precision multidimensional projection technique and its application to document mapping. *IEEE TVCG* **14**(3), 564–575 (2008)
28. Poco, J., Etemadpour, R., Paulovich, F.V., Long, T., Rosenthal, P., Oliveira, M.C.F., Linsen, L., Minghim, R.: A framework for exploring multidimensional data with 3D projections. *CGF* **30**(3), 1111–1120 (2011)
29. Rauber, P.E., Falcão, A.X., Telea, A.C.: Projections as visual aids for classification system design. *Inf. Vis.* **17**(4), 282–305 (2017)
30. Sanftmann, H., Weiskopf, D.: Illuminated 3D scatterplots. *CGF* **28**(3), 642–651 (2009)
31. Sanftmann, H., Weiskopf, D.: 3D scatterplot navigation. *IEEE TVCG* **18**(11), 1969–1978 (2012)
32. Schreck, T., von Landesberger, T., Bremm, S.: Techniques for precision-based visual analysis of projected data. *Inf. Vis.* **9**(3), 181–193 (2010)
33. Sedlmair, M., Aupetit, M.: Data-driven evaluation of visual quality measures. *CGF* **34**(3), 545–559 (2015)
34. Sedlmair, M., Munzner, T., Tory, M.: Empirical guidance on scatterplot and dimension reduction technique choices. *IEEE TVCG*, pp. 2634–2643 (2013)
35. Sips, M., Neubert, B., Lewis, J., Hanrahan, P.: Selecting good views of high-dimensional data using class consistency. *CGF* **28**(3), 831–838 (2009)
36. Tatu, A., Bak, P., Bertini, E., Keim, D., Schneidewind, J.: Visual quality metrics and human perception: an initial study on 2D projections of large multidimensional data. In: *Proc. AVI*, pp. 49–56. ACM (2010)
37. Tenenbaum, J.B., De Silva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. *Science* **290**(5500), 2319–2323 (2000)
38. Thijssen, J., Tian, Z., Telea, A.: Scaling up the explanation of multidimensional projections. In: *Proc. EuroVA. Eurographics* (2023)
39. Tian, Z., Zhai, X., van Driel, D., van Steenpaal, G., Espadoto, M., Telea, A.: Using multiple attribute-based explanations of multidimensional projections to explore high-dimensional data. *Computers & Graphics* **98**(C), 93–104 (2021)
40. Tian, Z., Zhai, X., van Steenpaal, G., Yu, L., Dimara, E., Espadoto, M., Telea, A.: Quantitative and qualitative comparison of 2D and 3D projection techniques for high-dimensional data. *Information* **12**(6) (2021)
41. van der Maaten, L., Hinton, G.E.: Visualizing data using t-sne. *JMLR* **9**, 2579–2605 (2008)
42. van der Maaten, L., Postma, E.: Dimensionality reduction: A comparative review. Tech. rep., Tilburg Univ., Netherlands (2009), tech. rep. TiCC 2009-005
43. Venna, J., Kaski, S.: Visualizing gene interaction graphs with local multidimensional scaling. In: *Proc. ESANN*, pp. 557–562 (2006)
44. Vito, S., Massera, E., Piga, M., Martinotto, L., Francia, G.: On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario. *Sensors Actuators B* **129**(2), 750–757 (2008). <https://archive.ics.uci.edu/ml/datasets/Air+Quality>
45. Yeh, I.C.: Concrete compressive strength dataset (2021). <https://archive.ics.uci.edu/ml/datasets/concrete+compressive+strength>