EuroVis Workshop on Visual Analytics (2023)
M. Angelini and M. El-Assady (Editors)

Scaling Up the Explanation of Multidimensional Projections

J. Thijssen and Z. Tian and A. Telea

Department of Information and Computing Science, Utrecht University, the Netherlands

Abstract

We present a set of interactive visual analysis techniques aiming at explaining data patterns in multidimensional projections. Our novel techniques include a global value-based encoding that highlights point groups having outlier values in any dimension as well as several local tools that provide details on the statistics of all dimensions for a user-selected projection area. Our techniques generically apply to any projection algorithm and scale computationally well to hundreds of thousands of points and hundreds of dimensions. We describe a user study that shows that our visual tools can be quickly learned and applied by users to obtain non-trivial insights in real-world multidimensional datasets.

CCS Concepts

• ***Human-centered computing*** → ***Visualization techniques; Visual analytics;***

1. Introduction

High-dimensional data is present in many science and engineering fields and, as such, is a key target for information visualization techniques. A main challenge in this respect is *scalability*, that is, how to visually depict datasets having hundreds of thousands of observations and tens to hundreds of dimensions. *Dimensionality reduction*, also called projection, techniques are one of the solutions of choice in this area [vP09, NA18]. Compared to other multi-dimensional visualizations such as table lenses [RC94], parallel coordinate plots [ID90], and scatterplot matrices [EPF08], projections scale well on both sample and dimension counts, and have become the main technique for visualizing such data in *e.g.* biology, astronomy, chemistry, and machine learning.

A raw projection is just a scatterplot which is not directly useful. Several methods have been proposed to *explain* projections. Brushing and color-coding show all dimensions of a single point, respectively one dimension over all points. Other global explanations include biplot axes [Gre10, GLR11, CMN*16] and axis legends [BBT13]. A more recent family of global explanations computes how neighbor points in the projection relate to each other and color the projection accordingly. Such explanations include dimension variance [dRM*15], data local dimensionality [vDZTT20], and strongest correlated dimensions [vDZTT20, TZvD*21]. Neighborhood-based explanations are easy to interpret (as they use dimension names), work with any projection technique, and provide information over all projected points. Yet, they (1) do not *scale* to more than roughly 10 dimensions; and (2) do not explain *what* the patterns in the projection mean [TZvD*21].

In this paper, we extend neighborhood-based projection explanations to overcome the above two problems while keeping them computationally scalable and generic. For this, we propose to globally explain projection patterns by the *values* of their constituent points (Sec. 2) and several interactive techniques that allow scaling explanations to tens of dimensions locally (Sec. 3). We combine local and global explanations in an interactive data-analysis tool. A study with 23 participants asked to solve analysis tasks on datasets of increasing dimensionality (Sec. 4) shows that our combined explanatory techniques lead to coherent and correct findings, thereby supporting our claims of added value.

2. Extending global explanations

Variance explanation: We first recall the variance-based explanation of Da Silva [dRM*15] which forms the basis of our extension.

Let $D = \{\mathbf{p}_i\}$, $1 \leq i \leq N$, $\mathbf{p}_i = (p_i^1, \ldots p_i^n) \in \mathbb{R}^n$ be a high-dimensional dataset and $D^P = \{\mathbf{q}_i = P(\mathbf{p}_i)\}$ its 2D projection by a DR technique P. Let $\nu_i^P = \{\mathbf{q} \in D^P | \|\mathbf{q}_i - \mathbf{q}\| \leq \rho\}$ be a neighborhood of radius ρ around projected point $\mathbf{q}_i \in D^P$. Points in ν_i^P come from the projection of a neighborhood $\nu_i = \{\mathbf{p} \in P | P(\mathbf{p}) \in \nu_i^P\}$ in the dataset D. To explain a projected point $\mathbf{q}_i$, one first computes the local variance of every dimension $1 \leq d \leq n$ over ν_i as

$$LV_i^d = \frac{1}{|\nu_i|} \sum_{\mathbf{p}\in\nu_i} \left(p^d - \frac{1}{|\nu_i|} \sum_{\mathbf{p}\in\nu_i} p^d \right)^2. \tag{1}$$

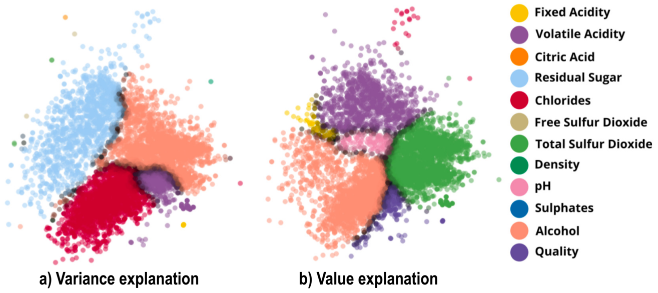
Next, a ranking of all n dimensions $\{\xi_i^d\}$, $1 \leq d \leq n$, is computed over ν_i as

$$\xi_i^d = \frac{LV_i^d / GV^d}{\sum_{j=1}^n LV_i^j / GV^j}, \tag{2}$$

where GV^d, the global variance of dimension d computed by replacing ν_i by D in Eqn. 1, is used to normalize across dimensions

DOI: 10.2312/eurova.20231098

with different variances. Intuitively put, low ξ_i^d values tell dimensions d which vary very little over ν_i (as compared to their variance over D), which is a way to explain why points in ν_i are similar. As such, the lowest-rank dimension $\lambda_i = \arg\min_{1 \leq j \leq n} \xi_i^j$ is picked to explain point $\mathbf{q}_i$. The C most-frequent such lowest-ranks λ_i over the whole projection D^P are next mapped to the C colors of a categorical colormap. Less-frequent ranks are mapped to a separate 'other dimensions' color. In our work, we use the $C = 20$ colormap proposed by Kelly [Kel65], excluding black and white. Finally, a *confidence* value is computed per point $\mathbf{q}_i$ as the fraction of points in ν_i^P which have the lowest-rank dimension equal to λ_i, and encoded in the projection via luminance.



a) Variance explanation

b) Value explanation

Figure 1: *Variance (a) and value (b) explanation of a projection.*

Figure 1a illustrates the variance explanation on the Wine dataset (further described in Sec. 4). Variance ranking helps explaining why points in the projection are close together in certain areas – for example, all red points have similar values of the *chlorides* dimension. Dark areas, close to the borders of same-color (same-explanation) regions, indicate points where the single-dimension explanation is less confident. However, this explanation does not tell *what* such points represent. To alleviate this, we next propose an explanation based on dimension values.

Value explanation: Similar to the variance explanation, we also compute ranks of all dimensions $\{\xi_i^d\}$, $1 \leq d \leq n$, over each neighborhood ν_i. The key idea behind value ranking is to find dimensions which have *outlier* values over such neighborhoods. For this, we first compute the local average

$$LA_i^d = \frac{1}{|\nu_i|} \sum_{\mathbf{p} \in \nu_i} p^d \tag{3}$$

of dimension d over ν_i. Using this, we compute the value ranking of dimension d as

$$\xi_i^d = \frac{1}{r(d)} \frac{LA_i^d - GA^d}{\sum_{j=1}^n \frac{1}{r(j)} |LA_i^j - GA^j|}, \tag{4}$$

where $r(d) = \max_{1 \leq i \leq N} p_i^d - \min_{1 \leq i \leq N} p_i^d$ and GA^d are the range, respectively, variance, of dimension d over D. Dimensions d with positive ranks ξ_i^d are unusually high in the neighborhood ν_i; dimensions with negative ranks are unusually low – the higher or lower the rank values are, the more unusual the dimension values are in that neighborhood. Depending on the application, one can choose whether they are interested in unusually high or unusually low dimensions or both of these. For space limitations, we discuss next in this paper only examining unusually high dimension values. Hence, we pick the highest-rank dimension $\lambda_i = \arg\max_{1 \leq j \leq n} \xi_i^j$ to explain point $\mathbf{q}_i$. These dimensions are next color-and-luminance mapped as for the variance ranking.

Figure 1b shows the value explanation of the Wine dataset. We see, for instance, that most of the red points in the variance-explanation (a), *i.e.*, wines having similar *chloride* values, are pink, *i.e.*, are wines with unusually high *alcohol* values. Section 4 further shows how the variance-and-value explanations can be combined to get more insights on a projection.

3. Adding local explanations

Even if we can compute explanations for many dimensions (Eqns 2 and 4), we can only show C of these *simultaneously* by our categorical colormap. Also, explaining projection patterns by a *single* dimension (whether via variance or values) only tells a part of the full story, since close points are placed so because of *multiple* dimensions. We address these limitations by several mechanisms that explain fewer points at a time, but in more detail (see also Fig. 2).

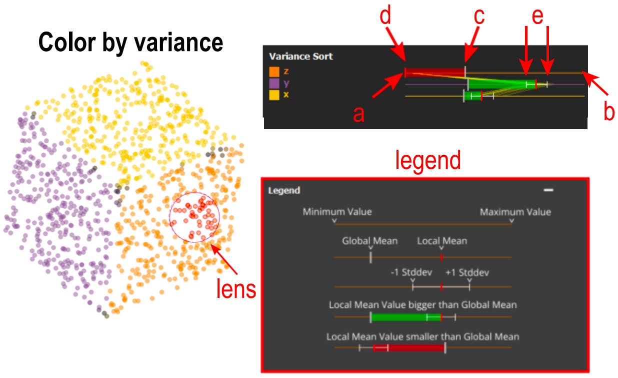


Figure 2: *Local explanation of lensed points (Sec. 3).*

Lens brushing: All projection points $\mathcal{S}$ within a given radius to the mouse pointer are selected to be the focus of the detailed (local) explanations described next. The radius is adjustable via the mouse wheel. For these selected points, we compute the variance and value rankings as for the global explanations (Eqns. 2 and 4) by substituting ν_i with the user selection $\mathcal{S}$. Users can interactively switch between the variance explanation (which tells *why* points in $\mathcal{S}$ are close in the projection) and the value explanation (which tells *what* these points are, data-wise).

Local analysis: We display detailed explanations of the lensed points $\mathcal{S}$ in a widget next to the projection. Figure 2 shows this widget for a simple 3D cube dataset projected using PCA. The widget is organized as a table with one row per dataset dimension. For each dimension, we show its name, assigned color (by variance or value ranking, see above), and a set of statistics for that dimension (described further below). In variance mode, dimensions are sorted top-to-bottom from lowest rank (lowest ratio of variance in the selected points $\mathcal{S}$ *vs* the whole projection) to highest rank (highest ratio of variance). In contrast to the global variance explanation (Sec. 2), we now not only show the least varying dimension (at the top), but all dimensions, sorted on variance. In value mode, we sort dimensions top-to-bottom from highest rank (highest mean value in $\mathcal{S}$ *vs* mean value over the whole projection) to lowest rank (lowest mean value). In contrast to the global value explanation, this shows not only the most outlier-like dimension (at the top), but all dimensions, sorted on their outlier-ness.

Dimension statistics: The above dimension sorting helps one find the most salient dimensions (in variance or value) but does not explain *how much* these contribute to the lensed points $\mathcal{S}$. To address this, we show both local and global statistics for each dimension d in the widget. A *range line* (same categorical color as the dimension) indicates the full extent $r(d)$ of the dimension over all projection points from the global minimum (Fig. 2, a) to the global maximum (Fig. 2, b). A large grey tick shows the dimension's global mean $\sum_{1 \leq i \leq N} p_i^d / N$ (Fig. 2, c). A red tick shows the dimension's local mean over the lensed points $\sum_{\mathbf{q}_i \in \mathcal{S}} p_i^d / |S|$ (Fig. 2, d) When the local mean is greater than the global mean, we draw a green bar between the two means to indicate a dimension which has higher than usual values over the lensed points. Similarly, when the local mean is smaller than the global mean, we draw a red bar between the two means, indicating a dimension having lower than usual values over the lensed points. Finally, we show the standard deviation of the dimension over $\mathcal{S}$ with white whiskers drawn left and right of the local mean (Fig. 2, e). Close whiskers indicate that the lensed points vary little over the analyzed dimension, thus the respective dimension is important for why the points are close in the projection. This is the same information as the top-to-bottom sorting in variance mode. However, in value mode, whiskers add the variance information which is not present in that mode.

To ease operation, we show all above encodings in a legend below the widget. Note also that, while similar to boxplots, our overall design is different – our whiskers represent standard deviations, and our bars represent local-*vs*-global mean differences.

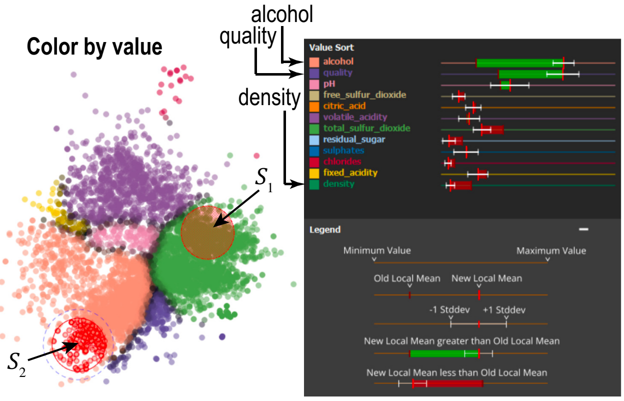


Figure 3: *Differential analysis of sets of points (Sec. 3).*

Parallel coordinates plot: We show more detailed information by a PCP of all lensed points $\mathcal{S}$ drawn half-transparently atop of the horizontal range lines of all dimensions.

Differential analysis: While local explanations show detailed information over a selected projection detail, one inherently needs to explore several such details in a sequence to understand a projection. This puts a certain burden on the user's memory. We alleviate this by offering a way to *compare* two different such user-selected details, as follows. The user selects a set of points $\mathcal{S}_1$, then presses a modifier key and selects a different set $\mathcal{S}_2$. The statistics that are normally shown in the analysis widget are now replaced by statistics showing the differences between $\mathcal{S}_1$ and $\mathcal{S}_2$. Figure 3 shows this for the Wine dataset using the value-ranking mode. The widget shows that the two top-most dimensions (*alcohol*, pink in the projection; and *quality*, dark purple in the projection) have long green bars, while the bottom-most dimension (*density*, dark green in the projection) has a red bar. This tells that wines in $\mathcal{S}_2$ have much higher alcohol and quality, but lower density, than wines in $\mathcal{S}_1$.

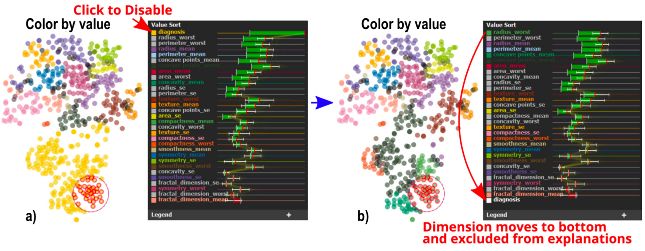


Figure 4: *Selective dimension disabling (Sec. 3).*

Dimension exclusion: Local analysis allows handling higher-dimensional data than global analysis as it displays details of all the dimensions over a selected data subset. Still, datasets can contain dimensions that do not convey much information for a given analysis. These can take up valuable colors from our limited $C = 20$ categorical colormap and also clutter the explanation widget. Excluding them upfront from the entire analysis is undesirable as users may wish to examine different sets of dimensions – and keep the same projection – depending how the analysis unfolds. To address this, we allow users to click on dimensions in the widget to temporarily exclude them from the generated explanations. Doing so reassigns colors to the remaining dimensions and instantly re-creates the global and local explanations. Clicking on an excluded dimension adds it back to the generated explanations. Figure 4 illustrates this. In image (a), about half of the projection points are explained by unusual high values of the *diagnosis* dimension (yellow, topmost in the rank-by-value widget). To get more insight on what else makes these points different, we click on this dimension and disable it. The dimension turns white in the widget and moves to the bottom to indicate disabling. The regenerated explanation (Fig. 4b) splits the big yellow blob into differently-colored groups that provide more insights of how these points differ.

Scalability: Our explanation system, implemented in C++ in the ManiVault framework [Div23], scales computationally well. It computes global explanations of datasets of hundreds of thousands of points and hundreds of dimensions in tens of seconds, and next interacts with these in real-time, on a commodity PC, and is openly available [TTT23]. Figure 5 illustrates the visual scalability in sample (a) and dimension (b) counts. Image (a) shows a dataset consisting of 22 registered images of the same brain-cortex tissue patch, each image mapping a gene. Pixel brightnesses encode where in the tissue the gene is expressed. We treat each image pixel as a sample having 22 dimensions (one from each image). This yields 115K 22-dimensional samples which we project with t-SNE [vH08] and next explain the projection. In Fig 5a, the global value explanation shows us how the projection is split into clearly separated point groups. We next lens over several points in the orange region, which corresponds to the Cux2 gene. The local explanation in the widget tells us that Cux2 is, indeed, unusually high in this region, and that only a few other dimensions have outlier values here. Figure 5b shows another dataset [ZMP*22] of gene expressions in the brain cortex. This dataset has 2400 samples (cells from the analyzed brain region) each with 314 dimensions (gene expressions). Finding high-expression genes in specific structures is one of the key analysis tasks such datasets are involved in. The projection shows the spatial layout of these cells. Even though the dataset has hundreds of dimensions, the global value-ranking explanation is able to assign colors to unravel a salient band-like structure in the projection, which is actually in line with the known layer structure of the cortex. Using the lens, we selected points in the purple band. The local explanation widget tells us that these have an unusually high expression of the Foxp2 gene (top-most bar in the widget), as well as showing other genes having high expressions in this area.

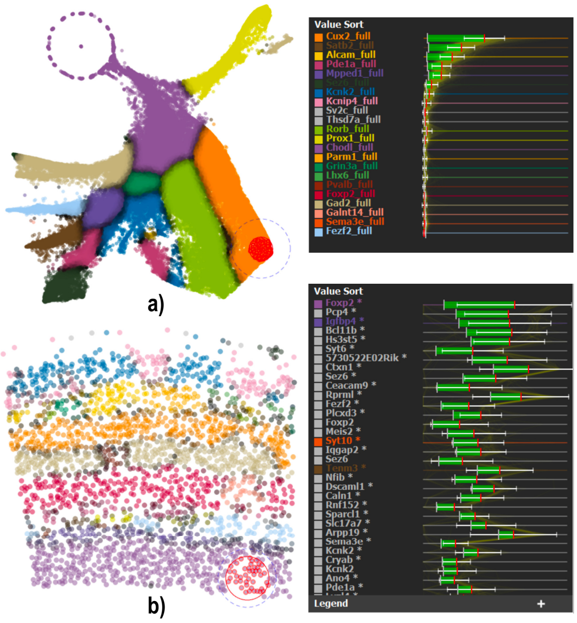


Figure 5: *Scalability of explanations in number of points (a) and dimensions (b) (Sec. 3).*

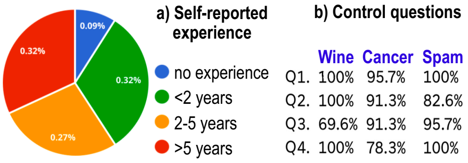
4. User Evaluation

To evaluate the effectiveness and ease of use of our interactive system for projection explanations, we conducted a user study (details in the supplementary material).

Participants: We invited about 60 people to take part in the study (and/or further spread the invitation). Of these, 23 completed the study. Participation was fully anonymous, *i.e.*, we did not collect nor trace the participants' identities. Participants self-reported (at the end of the study) experience with multidimensional data between none and several years (see also Fig. 6a).

Set-up: Participants installed our tool (Windows or Linux) and followed a tutorial (15 minutes) covering loading data, switching between variance and value explanations, and using the lens and local-explanation widget. Next, they were asked to analyze three multidimensional datasets and report answers via Google Forms. These datasets, all from the UCI repository [DG22], had increasing dimensionalities to test our system's scalability in this respect. The *Wine* dataset ($N = 6500, n = 12$) contains wine samples with 11 measured physicochemical attributes and one dependent attribute (perceived quality). The *Cancer* dataset ($N = 569, n = 31$) contains 10 attributes describing the size, shape, and texture (mean, max, and standard deviation) values of cell nuclei in a lung tissue. An extra attribute tells whether cells are benign or malignant. The *Spam* dataset ($N = 4601, n = 57$) contains frequencies of selected words for classifying whether mails are spam or not, and also the classification result. The datasets were projected using LAMP [JCC*11] (*Wine*) and t-SNE (*Cancer*, *Spam*) to check suitability with different projection techniques.

Questions: For each dataset, the participants had to answer four *control* (C) and three *live exploration* (LE) questions. The C questions involved examining screenshots of the application (produced by us) and selecting one of four multiple-choice answers. The answers were designed so that there was a single unambiguously correct one. In each question, different points of the projection were selected by the lens and snapshots of both global and local explanations were provided. The goal of these questions was to see if the participants understood how to read a pre-computed visualization (without interaction) to come to a correct conclusion. The LE questions asked participants to analyze the datasets interactively using the tool on their machines and select one or several multiple-choice answers. In contrast to the C questions, and given also the freedom to explore any parts of the dataset using any tool mechanisms, there were no 100% right or wrong answers. Rather, the goal here was to test if users of our tool, when asked to do an analysis in an uncontrolled environment, would come up with similar insights.



	Wine	Cancer	Spam
Q1.	100%	95.7%	100%
Q2.	100%	91.3%	82.6%
Q3.	69.6%	91.3%	95.7%
Q4.	100%	78.3%	100%

Figure 6: *Users' experience (a) and correctness of answering control questions (b).*

Results: The 12 control questions were overwhelmingly correctly answered (see Fig. 6b), suggesting that users were able to learn to correctly use our tool. The 9 live-exploration questions had no 100% right or wrong answers, as explained. Hence, we ranked their answers on an 4-point ordinal scale (very likely, likely, unlikely, very unlikely) telling how likely we ourselves would provide an answer after having studied those datasets in depth. We also measured the coherence of the users' answers – high values tell that different people arrive at similar insights. Figure 7 shows the agreement scores for 5 of these 9 questions (for more, see supplementary material). Long-and-bright bars in this figure indicate consensus between users and also with our own assessment, as follows.

Single cluster: This simple analysis asks users to find very-low-density wines in the projection and find which other attribute is also

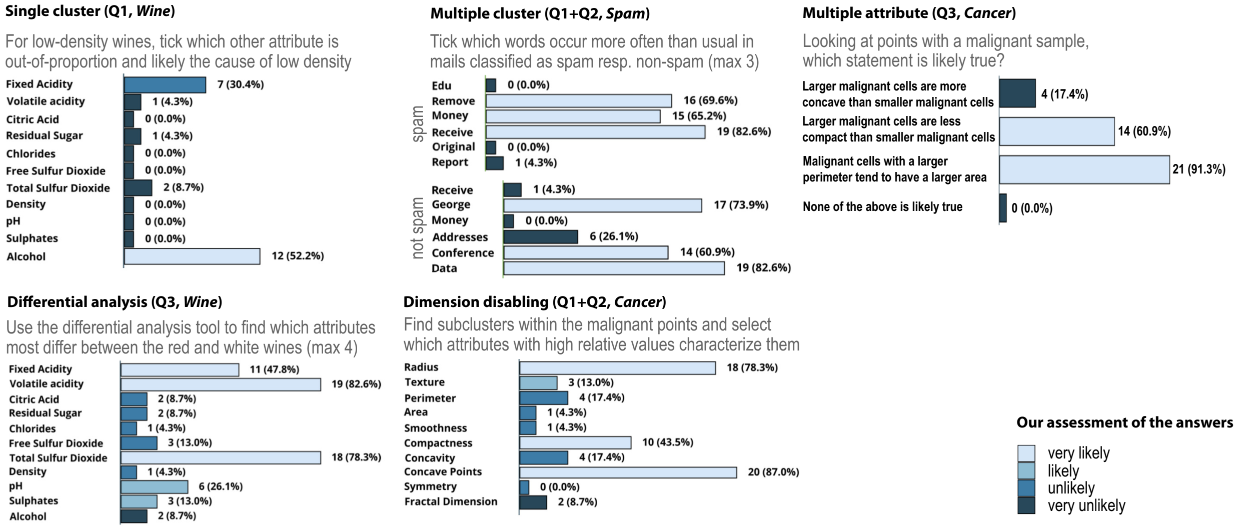


Figure 7: *Inter-user agreement (and our assessment of correctness likelihood) for answers from the live exploration analysis.*

out-of-proportion and thus likely causes the low density. This question can be easily answered using the lens and the value-ranking. Most users answered *alcohol* which is also our pick, with *fixed acidity* being a valid second option.

Multiple cluster: Users were asked to find which words occurred more often in non-spam and then in spam mails – thus, study at least 2 different clusters. This involved finding point clusters with spam, respectively non-spam, mails, via *e.g.* the variance global explanation, and then lensing in value-ranking mode to see which of the 6 words occurred there more often than elsewhere. Participants yielded very similar answers – and also similar to our own findings.

Multiple attributes: This question – arguably the most complex on our study – involved analyzing several attributes per point cluster. This requires interactively finding projection areas having low/high values of one attribute and then analyzing the other attributes in these areas. Again, we see strong inter-user agreement and also agreement with our own findings.

Differential analysis: Users were asked to tick up to four attributes that are most different between red and white wines. To do this, they had to find both red and white wines using the global explanation, and next use the differential analysis to find which attributes differ between these wines. We see again a strong agreement between users and also with ourselves.

Dimension disabling: Users were asked to find structure among the malignant points in the projection. For this, one could first disable the *diagnosis* dimension so as to see how malignant points split into sub-clusters described by other dimensions. When doing this, the malignant cluster indeed splits into three regions described well by outlier values of *radius*, *compactness*, and *concave points*.

The participants arrived at very similar answers, which we deem to be correct by our own independent analysis of the same datasets, which shows that our tool can help obtaining correct insights in high-dimensional data in a predictable way. Finally, we asked participants to provide feedback on the usefulness of our tool's features. The variance mode got a mean score of 4.83 (SD=1.63) and the value mode a mean of 6.52 (SD=0.77) on a Likert scale (1=not useful at all, 7=very useful). This supports our claim that the value mode is a useful addition. The PCP plot was found to provide additional explanatory value (60.9% users) and help understand the distribution of values (47.8% users). Yet, 17.4% of the participants found that it makes the local analysis widget more confusing and 8.7% found it brings no added value. Finally, the differential analysis tool's usefulness was ranked with a mean of 5.74 (SD=1.03) on the same scale as the variance and value. The dimension exclusion got an identical mean of 5.74 (SD=1.42).

5. Discussion and Conclusion

We have presented a set of interactive visual analysis techniques for the exploration and explanation of multidimensional projections. Our techniques extend existing global variance explanations with a global value-based explanation; local explanations based on variance, value, and detailed statistics on all dataset dimensions; a differential analysis tool allowing the comparison of two projection regions; and a dimension filtering mechanism. Our techniques handle any projection algorithm and scale computationally and visually to datasets of over 100K samples and over 300 dimensions. A user study showed that our techniques can be quickly learned, are found useful, and can be applied to answer non-trivial questions on real-world multidimensional datasets.

Several directions can be explored next. Global explanations are still limited as they inherently show a *single* dimension. Further studying the original idea proposed – but not elaborated – by Da Silva [dRM*15] to use dimension-sets, computed *e.g.* by subspace clustering, and possibly complemented by dimension-value-ranges, can improve such explanations. Separately, we could incorporate knowledge on the specific projection method used to make the explanatory metrics more insightful than generic variance and outlier-value computations. Finally, deploying our tool in a long-term analysis involving a real use-case and domain experts would bring additional evidence of its practical value.

References

[BBT13] BROEKSEMA B., BAUDEL T., TELEA A.: Visual analysis of multidimensional categorical datasets. *Computer Graphics Forum 32*, 8 (2013), 158–169. 1

[CMN*16] COIMBRA D., MARTINS R., NEVES T., TELEA A., PAULOVICH F.: Explaining three-dimensional dimensionality reduction plots. *Information Visualization 15*, 2 (2016), 154–172. 1

[DG22] DUA D., GRAFF C.: UCI machine learning repository, 2022. http://archive.ics.uci.edu/ml. 4

[Div23] DIVISION OF IMAGE PROCESSING, LUMC: ManiVault visualization framework, 2023. https://github.com/hdps. 3

[dRM*15] DA SILVA R., RAUBER P., MARTINS R., MINGHIM R., TELEA A. C.: Attribute-based visual explanation of multidimensional projections. In *Proc. EuroVA* (2015), pp. 97–101. 1, 5

[EPF08] ELMQVIST N., P P. D., FEKETE J. D.: Rolling the dice: multidimensional visual exploration using scatterplot matrix navigation. *IEEE TVCG*, 14 (2008), 1141–1148. 1

[GLR11] GOWER J., LUBBE S., ROUX N.: *Understanding biplots*. Wiley, 2011. 1

[Gre10] GREENACRE M.: *Biplots in practice*. Fundacion BBVA, Bilbao, 2010. 1

[ID90] INSELBERG A., DIMSDALE B.: Parallel coordinates: A tool for visualizing multi-dimensional geometry. In *Proc. IEEE VIS* (1990), pp. 361–378. 1

[JCC*11] JOIA P., COIMBRA D., CUMINATO J. A., PAULOVICH F. V., NONATO L. G.: Local affine multidimensional projection. *IEEE TVCG 17*, 12 (2011), 2563–2571. 4

[Kel65] KELLY K. L.: Twenty-two colors of maximum contrast. *Color Eng 3*, 26 (1965), 26–27. 2

[NA18] NONATO L., AUPETIT M.: Multidimensional projection for visual analytics: Linking techniques with distortions, tasks, and layout enrichment. *IEEE TVCG 25*, 8 (2018), 2650–2673. 1

[RC94] RAO R., CARD S. K.: The table lens: Merging graphical and symbolic representations in an interactive focus+context visualization for tabular information. In *Proc. ACM SIGCHI* (1994), pp. 318–322. 1

[TTT23] THIJSSEN J., TIAN Z., TELEA A.: Visual explanation system for multidimensional projections, 2023. https://github.com/JulianThijssen/ProjectionExplorer. 3

[TZvD*21] 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* (2021), 93–104. 1

[vDZTT20] VAN DRIEL D., ZHAI X., TIAN Z., TELEA A.: Enhanced attribute-based explanations of multidimensional projections. In *Proc. EuroVA* (2020), Eurographics, pp. 37–41. 1

[vH08] VAN DER MAATEN L., HINTON G. E.: Visualizing data using t-sne. *JMLR 9* (2008), 2579–2605. 3

[vP09] VAN DER MAATEN L., POSTMA E.: *Dimensionality Reduction: A Comparative Review*. Tech. rep., Tilburg University, Netherlands, 2009. Tech. report TiCC TR 2009-005. 1

[ZMP*22] ZHANG Y., MILLER J. A., PARK J., LELIEVELDT B. P., ET AL.: Reference-based cell type matching of spatial transcriptomics data. In *bioRxiv* (2022). https://doi.org/10.1101/2022.03.28.486139. 4