Understanding Bias in Perceiving Dimensionality Reduction Projections

Seoyoung Doh[1*] Hyeon Jeon[1†] Sungbok Shin[2‡] Ghulam Jilani Quadri[3§] Nam Wook Kim[4¶] Jinwook Seo[1‖]
[1]Seoul National University [2]Inria, Université Paris-Saclay [3]University of Oklahoma [4]Boston College

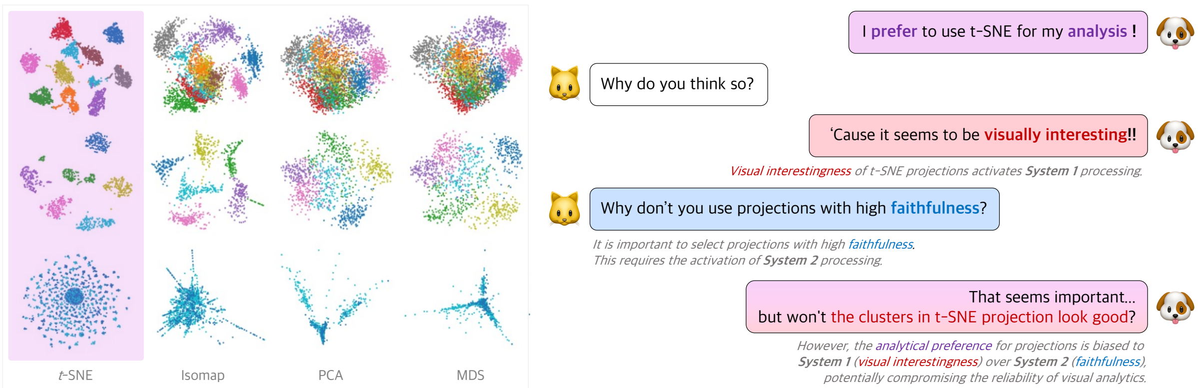

Figure 1: The illustration of our motivating scenario and its alignment with dual-system theory (Sect. 2.2). We assume a scenario where a practitioner wants to select DR techniques that produce projections suitable for analyzing their dataset. It is crucial to select projections with high faithfulness for reliable analysis, but practitioners tend to be more biased towards the visual interestingness. We verify the existence of such bias and investigate its underlying causes, providing a grounded basis for mitigating the biases.

Abstract

Selecting the dimensionality reduction technique that faithfully represents the structure is essential for reliable visual communication and analytics. In reality, however, practitioners favor projections for other attractions, such as aesthetics and visual saliency, over the projection's structural faithfulness, a bias we define as *visual interestingness*. In this research, we conduct a user study that (1) verifies the existence of such bias and (2) explains why the bias exists. Our study suggests that visual interestingness biases practitioners' preferences when selecting projections for analysis, and this bias intensifies with color-encoded labels and shorter exposure time. Based on our findings, we discuss strategies to mitigate bias in perceiving and interpreting DR projections.

Index Terms: Dimensionality reduction, Visual interestingness, Faithfulness, Bias, Dual-system theory

1 Introduction

Dimensionality reduction (DR) is a popular technique for visually interpreting and analyzing high-dimensional data. To visualize data with DR, practitioners should first select DR techniques that align with their analytical tasks. Here, *faithfulness* [26, 25, 16], the degree to which the structural characteristics of the original data are preserved without distortions, should be prioritized to ensure that the projection reliably supports the tasks. However, analysts may prioritize other factors, such as aesthetics or visual saliency, preferring projections with cleanly separated clusters and visually distinct boundaries (Fig. 1).

This bias towards *visual interestingness* [28, 35, 8], the degree to which projections exhibit perceptually appealing patterns, may prompt analysts to select unsuitable techniques. This issue is particularly concerning for researchers and data analysts who regularly perform visual analytics with high-dimensional datasets but have limited DR literacy, as their vulnerability to the bias threatens the reliability of their scientific discoveries [5].

Our work aims to empirically investigate the impact of visual interestingness on analytical preference of DR projections in practice. For this, we address the following research questions:

- (**RQ1**) Do practitioners tend to make biased selections by favoring visual interestingness of projections over faithfulness?
- (**RQ2**) If such bias is observed, why does such bias occur?

We conduct a two-phase user study for the purpose. In the first phase, we obtain the ranking of DR projections based on visual interestingness by asking the participants. In the second phase, we present participants with the same projections alongside artificially created faithfulness scores that contradict the visual interestingness rankings from phase 1. By doing so, we examine which factor participants prioritize when selecting projections for analysis.

Our results verify the bias towards visual interestingness over faithfulness in determining analytical preference of DR projections, i.e., the degree to which analysts prefer to use projections for their analysis. When selecting projections, participants exhibit a strong tendency to rely on visual interestingness, which is particularly pronounced with color-encoded class labels. We also find that the bias intensifies with clear class or cluster boundaries. These findings prompt discussions on design strategies that can help mitigate bias in perceiving DR projections, enabling more reliable use of DR in visual analytics and communications.

In summary, our contributions are as follows:

- We identify **perceptual bias** in selecting DR projections to analyze towards visual interestingness over faithfulness.
- We empirically verify the existence of such bias and investigate why such bias occurs through a **user study** with 32 participants.
- We recommend **strategies** to **mitigate** bias in perceiving DR projections.

*e-mail: 0303dsy@snu.ac.kr
†e-mail: hj@hcil.snu.ac.kr
‡e-mail: sungbok.shin@inria.fr
§e-mail: quadri@ou.edu
¶e-mail: nam.wook.kim@bc.edu
‖e-mail: jseo@snu.ac.kr, corresponding author

2 PRELIMINARIES

Our work draws upon two main areas of literature: (1) DR and (2) dual-system theory from cognitive psychology.

2.1 Dimensionality Reduction

We first review DR, then discuss the importance of assessing both the faithfulness and visual interestingness of DR projections.

Dimensionality reduction. DR techniques receive high-dimensional data as input and transform it into low-dimensional (typically two or three) representations that preserve structural characteristics [26, 5]. By doing so, analysts can visually investigate data characteristics such as relationships between different classes [21, 17]. However, even with the same dataset, DR projections can show different visual patterns depending on the chosen technique and hyperparameters [26, 36]. Therefore, analysts must carefully select both techniques and hyperparameters to align with their analytical objectives [36, 16].

Faithfulness of DR projections. When selecting a DR technique, evaluating *faithfulness*, the degree to which the original high-dimensional structure is preserved in the low-dimensional projection, is crucial for achieving reliable visual analytics [16, 26]. Various metrics have thus been proposed to assess DR faithfulness [14]. For example, *trustworthiness and continuity* [34] metrics evaluate how well neighborhood relationships are maintained in projections.

Visual interestingness of DR projections. Building on previous literature that defines interesting projections [35, 28, 8], we define visual interestingness as follows:

> **Definition.** *Visual interestingness* of a projection denotes the degree to which the projection exhibits visually salient, distinctive, or appealing patterns that elicit perceptual attention.

It is worth noting that our scope of visual interestingness does not cover cognitive interpretations of DR projections, such as engagement or alignment of patterns with analysts' expectations.

Our Contribution. We empirically verify that practitioners tend to prioritize visual interestingness over faithfulness when selecting DR projections for their analysis. Based on our findings, we discuss strategies to mitigate such bias.

2.2 Dual-System Theory: System 1 and System 2

Our hypothesis and study design is grounded on *dual-system theory*, a cognitive psychological framework that explains human thinking processes. We first introduce the theory, then discuss how this theory aligns with our problem statement and the bias in perceiving DR projections.

Dual-system theory. *Dual-system theory* [19, 32] posits that human thinking operates through two processing systems: System 1, which is fast and intuitive, and System 2, which is slower but more logical and analytical. These two systems work together in a complementary manner. For instance, System 1 enables quick judgments but may lead to premature or biased decisions, which can be corrected by more deliberate reasoning of System 2. Conversely, in time-limited or routine tasks, System 1 can help optimize cognitive processing by compensating for System 2's slower and more effortful thinking.

Alignment of the dual-system theory with our study. We align our study with dual-system theory by conceptualizing visual interestingness as a System 1 process and faithfulness as a System 2 process. We map visual interestingness to System 1 as it captures practitioners' attention instantaneously through visual elements [11], thereby activating fast and intuitive processing. Faithfulness requires practitioners to engage in rational evaluation of objective numerical scores [10]; we thus align the faithfulness with System 2's deliberate reasoning that requires a longer time.

Our Contribution. Building on this theoretical foundation, our experimental design examines System 1 and System 2 impacts the analytical preference of DR projections. We further align with established findings that shorter exposure times and higher visual saliency amplify System 1 processing [11, 24, 4] by investigating how exposure duration and color encoding of class labels influence such bias. By controlling these factors, we identify empirical evidence of the cognitive mechanisms of perceiving and interpreting DR projections.

3 USER STUDY

We detail our study, which reveals the existence of bias towards System 1 in perceiving DR projections (Fig. 2).

3.1 Objectives

We aim to understand the interplay between three key components: (1) Visual interestingness [28, 35, 8], which stands for the degree to which the projection exhibits visually salient, distinctive, or appealing patterns that elicit perceptual attention, (2) Faithfulness [26, 25, 16], i.e., the degree to which projections accurately represent the original high-dimensional data without distortions, and (3) Analytical preference [38], which is defined as the degree to which analysts want to use projections for their analysis. To do so, we verify the following hypotheses:

- **(H1)** *Visual interestingness* more strongly influences *analytical preference* of DR projections than *faithfulness* does.
- **(H2)** The tendency observed in H1 becomes more pronounced when the color encoding of classes is provided.
- **(H3)** The tendency observed in H1 becomes more pronounced with shorter exposure time to stimuli.

We thus demonstrate the substantial impact of System 1 on selecting DR projections for analysis (H1). We further show that System 1's influence intensifies with supporting factors such as color encoding and limited stimulus exposure time, aligned with prior observations in the literature (H2 and H3) [24, 35]. Confirming these hypotheses establishes both the existence of bias in DR projection selection and its underlying mechanisms, thereby informing strategies for bias mitigation.

To achieve these objectives, we examine how the following two independent variables affect the visual interestingness and analytical preference of DR projections:

- Color encoding of class labels (COLORENCODING): *Monochrome* and *Polychrome*
- Time duration for exposing stimuli (EXPOSURETIME) : *seven seconds* and *15 seconds*

We select COLORENCODING because color-encoded class labels are commonly used for visualizing DR projections and significantly influence their perception [24, 4]. We also select EXPOSURETIME as time constraints are known to amplify System 1 processing [19]. We set exposure times of 7 and 15 seconds based on the time needed to read and interpret the metric scores in our pilot study.

Note that visual patterns of DR projections (e.g., class separability or the number of clusters) also affect their perception. For the completeness of our analysis, we also investigate their effect on visual interestingness in Sect. 4 (Analysis 2).

3.2 Study Design

Our experiment consists of two phases (Fig. 2). In the first phase, we present participants with pairs of projections and ask them to select the one that are more visually interesting (H1). In the second phase, we present different participants with pairs of projections and ask them to select the one they prefer more for their analysis, while presenting faithfulness scores that favor projections with

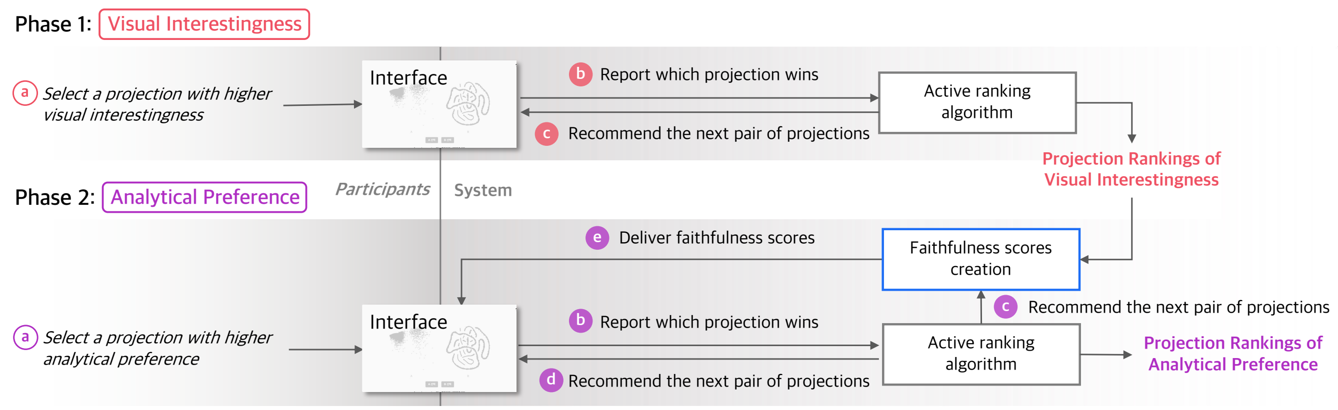

Figure 2: The illustration of our study design (Sect. 3). In Phase 1, participants view pairs of projections and select the one with higher visual interestingness. Phase 2 repeats this process for analytical preference, but with artificially generated faithfulness metrics that favor less visually interesting projections. This design enables us to quantify participants' bias toward visual interestingness when it conflicts with faithfulness. Participant actions are denoted in italics, while system actions appear in regular text.

lower visual interest. We thus examine how practitioners' analytical preference for DR projections is affected by visual interestingness and faithfulness (H2). We control our independent variables (ColorEncoding and ExposureTime) for both phases.

3.2.1 Phase 1: Visual Interestingness

We show participants pairs of projections and ask them to select a more visually interesting projection. Upon selection, the result feeds into an active ranking algorithm [12] that identifies optimal comparison pairs to efficiently converge on stable rankings with minimal iterations. For each session, we ask participants to perform 50 trials of comparisons. Then, the active ranking algorithm outputs the final ranking based on the results of all trials. Participants complete four sessions, during which we distribute the combinations of independent variables to the sessions using a Latin square design.

Procedure. One instructor manages all experiments. After participants make their consent and report demographics, the instructor details the objectives and tasks. We allow participants to freely ask questions during the introduction. Then, participants go through four sessions. After the sessions, we conduct interviews to gather qualitative insights into their selections.

Task. We provide participants with pairs of DR projections and ask them to determine which one is more visually interesting. We ask: *"Given two projections, which projection catches your eye first?"*, following the definition of visual interestingness in Healey and Enns [11]. We employ pairwise comparison to obtain rankings that are unbiased by anchoring effects [33] and subjectivity inherent in Likert-scale-based evaluations [29].

Distributing the combinations of independent variables. We have $2 \times 2 = 4$ combinations of ColorEncoding and ExposureTime. We distribute these conditions across participants using a Latin square design for counterbalancing, requiring participant counts in multiples of four. We recruit 16 participants in total.

Stimuli. We generate 20 projections for each combination of ColorEncoding and ExposureTime, resulting in 80 projections. Here, we aim to maximize the diversity of visual patterns. To do so, we first select datasets with maximum variance in patterns using stratified sampling [1, 27] from the existing set of 96 real-world datasets [13]. Then, we generate various projections of the sampled datasets and again use stratified sampling to identify projections with diverse patterns (detailed procedure in Appendix A).

Determining the ranking of projections. After each session, we determine the visual interestingness rankings of the 20 projections based on the participants' 50 pairwise selections using an active ranking algorithm [12]. Note that this algorithm also determines which pairs of projections to present to participants in each trial.

Participants. We recruit 16 participants from four local universities (eight males and eight females, aged 21–33 [25.6 ± 3.5]). We recruit participants who have experience in data analysis using scatterplots to align our experiments with real-world data analysis. No participants report being experts, while 11 and five report being intermediates and novices, respectively. We limit participants to those with low literacy of DR, including those who have prior experience with DR but have not engaged with it deeply. Participants are compensated with the equivalent of $7.

Apparatus. We conduct experiments over recorded Zoom calls. We develop a website where participants can view stimuli and make their selections with a mouse click or keyboard arrows. We ask them to access the website and share their screens.

3.2.2 Phase 2: Analytical Preference

Phase 2 shares most design choices and stimuli with Phase 1, except that we request analytical preferences: participants are guided to select projections that they wish to use for cluster analysis. We select cluster analysis for two reasons: First, it is one of the most commonly applied analysis types for DR projections [38, 21, 17]. Second, cluster analysis is highly affected by our independent variable ColorEncoding. Specifically, we ask participants: *"Given two projections, which one do you prefer to use for analyzing the cluster structure of the original data?"*. Additionally, we present faithfulness scores for the projections. These scores are artificially generated to favor projections with low visual interestingness, allowing us to examine which factor participants prioritize when visual interestingness and faithfulness conflict (H2).

Participants. We recruit 16 participants from four local universities (11 males and five females, aged 23–30 [25.6 ± 2.1]) following the same criteria as Phase 1. One participant reports being an expert in scatterplot-based visual analysis, while five and 10 report being intermediate and novice practitioners, respectively.

Generating and presenting faithfulness scores. For each pairwise comparison, we present five pairs of faithfulness scores. We use artificially generated scores rather than actual faithfulness scores to examine how practitioners select projections when visual interestingness and faithfulness conflict. Out of the five metrics, we assign higher scores to the less visually interesting projection on three or four randomly selected metrics. All scores randomly range within the [0, 1], consistent with the typical range of widely used DR met-

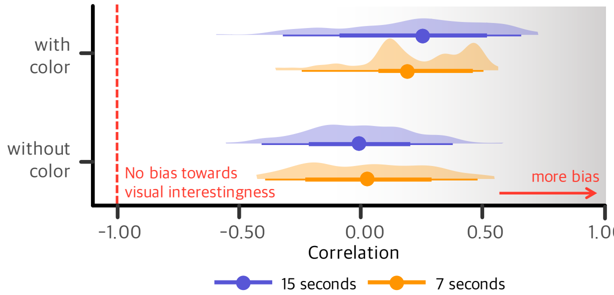

Figure 3: The correlations between visual interestingness and analytical preference in our experiment (Sect. 4.1). We find that the correlations becomes significantly higher with color encoded class labels.

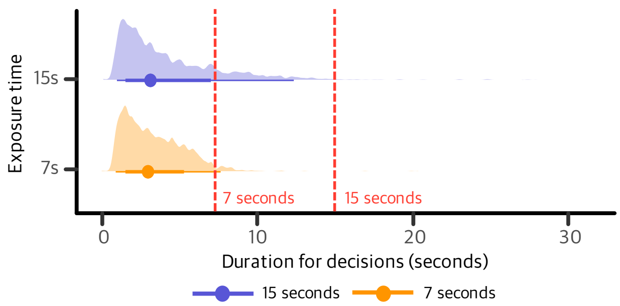

Figure 4: The time duration needed to select projections in our experiments (Sect. 4.1). Regardless of ExposureTime constraints, participants make a selection within approximately five seconds.

rics [14]. Also, to minimize potential bias from participants' background knowledge, we use generic labels (metrics A through E) for each pair of faithfulness scores, rather than actual metric names.

4 Quantitative Analysis

We present the results of the quantitative analysis of our user study.

4.1 Analysis 1: Correlation between Rankings

Objectives. We examine the correlation between rankings of DR projections based on visual interestingness and analytical preference (H1). As we intentionally assign higher faithfulness scores to projections with lower visual interestingness, the correlation will ideally have the worst value if there is no bias from visual interestingness. Conversely, positive correlations may occur when bias exists. We test this hypothesis and investigate how such correlations are affected by ColorEncoding and ExposureTime (H2, H3).

Analysis design. We compute rankings of DR projections based on visual interestingness and analytical preference, then examine how such correlations vary across ColorEncoding and ExposureTime. For each session, we obtain the rankings based on pairwise comparison results. This yields 16 (participants) $\times$ 4 (sessions) = 64 rankings each for visual interestingness and analytical preference. With 2 (ColorEncoding) $\times$ 2 (ExposureTime) = 4 conditions and four dataset variations, we obtain 16 combinations in total. Each combination contains four rankings for visual interestingness $\mathbf{V} = \{V_1, V_2, V_3, V_4\}$ and four rankings for analytical preference $\mathbf{A} = \{A_1, A_2, A_3, A_4\}$. For each combination, we compute the set of correlations between $\mathbf{V}$ and $\mathbf{A}$ as $\{\rho(V_i, A_i) \mid V_i \in \mathbf{V}, A_i \in \mathbf{A}\}$, where ρ denotes the Spearman correlation coefficient. We then analyze how these correlations vary across independent variables using two-way ANOVA with Tukey's HSD for post-hoc analysis.

Results and discussions. We find a significant effect on ColorEncoding ($F_{1,336} = 65.10$, $p < .001$) but not on ExposureTime ($F_{1,336} = 0.22$, $p = 0.64$), with no significant interaction effect ($F_{1,336} = 0.46$, $p = 0.50$), confirming H1. For ColorEncoding, we find that correlations range around 0.25 (Fig. 3). This confirms the ineffectiveness of adversarial depiction of faithfulness, supporting H2. We observe that there is no significant effect for ExposureTime because participants complete their tasks within shorter durations—approximately five seconds on average—regardless of the time constraint (Fig. 4).

4.2 Analysis 2: Effect of Visual Features

Objectives. We want to understand how visual patterns beyond color encoding (e.g., class separability) of DR projections influence the visual interestingness of DR projections. By doing so, we aim to identify not only which DR projections are susceptible to visual bias, but also which specific visual patterns drive this bias, thereby informing strategies to mitigate it.

Analysis design. We extract features representing visual patterns in DR projections and examine the extent to which these features affect visual interestingness. The *number of clusters* and *cluster quality* serve as our primary features, as prior work has demonstrated their significant influence on scatterplot perception [17, 27]. For the number of clusters, we employ Gaussian Mixture Models across varying cluster numbers and select the configuration with the lowest Bayesian Information Criterion score, following [17]. We measure cluster quality using Silhouette scores due to their widespread use in cluster quality assessment in scatterplot [2, 18, 21]. For polychrome projections, we additionally leverage *class separability*, as this feature significantly influences scatterplot perception [35, 24, 4]. We also incorporate Scagnostics [37], which substantially impacts how people perceive DR projections [24].

We examine their influence on visual interestingness using an ablation study. We first train a regression model to predict the visual interestingness of DR projections using all extracted features. Subsequently, we train models with individual features removed and measure the resulting accuracy degradation relative to the full model, using this as a proxy for feature importance. We use linear regression models as we have limited number of data points (which is 20), thus using other advanced model may suffer from data sparsity. We use R^2 as the target metric due to its interpretability [6].

We apply this procedure independently to each set of 20 projections. For each set, we aggregate all participant trials and input them to the active ranking algorithm [12] to derive a consensus ranking of visual interestingness. We assign visual interestingness scores (our target variable) using a linear transformation: projection with rank r receives score $21 - r$, to align with the use of linear regression.

Results and discussions. Table 1 and 2 depicts the results. For polychrome projections, *class separability* dominates the influence of visual features in predicting visual interestingness, with others showing substantially lower impact. In contrast, for monochrome projections, influences of features are more evenly distributed, where the impact of *class separability* is substantially reduced compared to the case of polychrome projections. We also find that *clumpy* exhibits the highest influence, with the *number of clusters* and *cluster quality* serve as runner-ups. These results indicate that when selecting DR projections, *class separability* in polychrome projections and *clumpy* in monochrome projections intensify visual interestingness of practitioners, potentially stimulating System 1 processing. We discuss the implications of such results in Sect. 6.

5 Qualitative Results

The following is our findings from the post-hoc interview study.

Finding 1. **Practitioners are biased towards visual interestingness over faithfulness in selecting DR projections.** We find qualitative evidence of the bias towards visual interestingness in selecting DR projections to analyze. This tendency appeared regardless of the participants' level of DR literacy. All participants report that

Table 1: The performance of linear regression models predicting visual interestingness from the features representing visual patterns of **polychrome** DR projections (number of clusters, cluster quality, class separation, and Scagnostics). We investigate how the performance degrades as we remove each feature representing visual patterns. The first row depicts the correlations of regression models, and the second row shows the performance decrement compared to the full model. The opacity of cells represents the amount of decrement, with opacity mapped from 0.8 to 0 corresponding to 0% to 15%. *Class separability* dominates the contribution of visual features in predicting visual interestingness, while other features show relatively small impact compared to their influence on monochrome projections.

	full	cluster #	cluster qual.	class sep.	outlying	skewed	clumpy	sparse	striated	convex	skinny	stringy	monotonic
R^2	0.7784	0.7666	0.7689	0.5855	0.7663	0.7509	0.7722	0.7372	0.7432	0.7624	0.7163	0.7653	0.7605
Change	·	1.18%	0.95%	19.29%	1.21%	2.75%	0.61%	4.11%	3.52%	1.60%	6.20%	1.30%	1.79%

Table 2: The performance of regression models predicting visual interestingness based on visual patterns of **monochrome** DR projections. The table shares visual encoding and arrangement with Table 1. *Clumpy* shows the largest impact on visual interestingness, while other features exhibit smaller, comparable effects.

	full	cluster #	cluster qual.	class sep.	outlying	skewed	clumpy	sparse	striated	convex	skinny	stringy	monotonic
R^2	0.7293	0.7062	0.7139	0.7294	0.7229	0.7499	0.6666	0.7265	0.7371	0.7381	0.7370	0.7532	0.7253
Change	·	8.09%	7.09%	5.07%	5.91%	2.40%	13.25%	5.45%	4.07%	3.94%	4.08%	1.98%	5.61%

they relied predominantly on the scatterplot's visual characteristics when making their selections. For example, P2 notes: "*I relied much more on the visual part rather than the faithfulness metrics. When there was a conflict, I made decisions based more on the visual elements*." Such findings provide additional evidence to confirm H1.

***Finding 2*. Practitioners exhibit stronger bias with color encoding and shorter exposure time.** We also find evidence that the bias towards visual interestingness intensifies with color-encoded labels and shorter exposure time. 16 participants report preferring polychrome projections over monochrome, noting that polychrome projections draw more visual attention. Moreover, we find that shorter exposure time compels practitioners to make selections based on the visual interestingness of projections. Ten participants report that they preferred more visually interesting projections when seven seconds are given, as they had limited time to read metric scores. These results provide weak support for H3.

***Finding 3*. Practitioners are strongly influenced by clear separation of classes and clusters.** Participants report that clearly separated classes and clusters draw their visual attention for polychrome and monochrome scatterplots, respectively, while selecting visually interesting projections. The results agree with our quantitative findings (Sect. 4.2) that *class separability* is an important feature determining the visual interestingness of polychrome projections, and that *clumpy* serves the same role for monochrome projections.

***Finding 4*. Practitioners are unaware of their bias.** We observe that all participants are unaware of their bias, regardless of their DR literacy level. Three participants also deny being biased. P2 states that they cannot agree that favoring visual interestingness over faithfulness scores should be considered bias.

6 Discussions: Strategies to Mitigate Bias

Our findings inform the following strategies to mitigate bias towards visual interestingness in selecting DR projections to analyze.

***Strategy 1*: Deactivating System 1 in perceiving DR projections** We recommend deactivating System 1 processing in perceiving DR projections by controlling visual factors in projections. As our findings reveal that (1) color encoding activates System 1 (Sect. 4.1), and (2) class separability plays an important role in such activation (Sect. 4.2 and 5), we suggest avoiding color encoding of class labels when depicting DR projections. One approach is to simply use monochrome representation. If displaying class labels is necessary (e.g., when showing DR performance for interactive labeling [3]), we recommend using shape encoding [31].

***Strategy 2*: Activating System 1 and 2 in reading faithfulness scores.** We also suggest activating System 1 processing when reading faithfulness scores. Our findings reveal that faithfulness scores are overlooked due to the visual interestingness of projections (Sect. 4.1). These results indicate that faithfulness scores are less visually salient compared to projections. To reduce this gap, we recommend visually highlighting texts representing faithfulness scores [30], e.g., by assigning background color with high opacity to higher scores. Another plausible approach is to visually encode faithfulness scores, e.g., by depicting their distributions with uncertainty [7] or leveraging word-scale visualizations [9]. These visual representations will not only increase System 1 engagement but also benefit System 2 processing, as interpreting visualizations requires additional cognitive load [10].

***Strategy 3*: Activating System 2 by enhancing DR literacy.** Our findings indicate that projections with high clumpiness and well-separated clusters are perceived as visually interesting (Sect. 4.2). Regarding the fact that participants are unaware of their bias (Sect. 5), such tendency may indicate that practitioners may erroneously favor DR techniques that exaggerate cluster structure, such as t-SNE [22] and UMAP [23], as documented in the literature [36, 16, 20, 15]. To address this issue, we argue to invest community efforts in enhancing DR literacy. This will lead practitioners to be more cautious in selecting DR projections to analyze (i.e., more influenced by System 2 processing). This direction will require practical efforts beyond academic papers, such as creating tutorials and approachable web articles [36, 16].

7 Conclusion

In this study, we empirically demonstrate that practitioners are biased towards visual interestingness over faithfulness when selecting DR projections. Based on our findings, we recommend three strategies to mitigate this bias that deactivate System 1 processing and activate System 2 processing. In future work, we plan to investigate this bias more deeply and evaluate the effectiveness of our proposed strategies. Investigating how DR literacy impacts practitioners' perception of visual interestingness and faithfulness in DR projections would also be an interesting avenue to explore.

References

[1] M. M. Abbas, M. Aupetit, M. Sedlmair, and H. Bensmail. Clustme: A visual quality measure for ranking monochrome scatterplots based on cluster patterns. *Computer Graphics Forum*, 38(3):225–236, 2019. doi: 10.1111/cgf.13684 3

[2] S. S. Bae, T. Fujiwara, C. Tseng, and D. Szafir. Uncovering how scatterplot features skew visual class separation. In *ACM CHI*, 2025. 4

[3] J. Bernard, M. Hutter, M. Zeppelzauer, D. Fellner, and M. Sedlmair. Comparing visual-interactive labeling with active learning: An exper-

imental study. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):298–308, 2018. doi: 10.1109/TVCG.2017.2744818 5

[4] A. Bibal and B. Frénay. Learning interpretability for visualizations using adapted cox models through a user experiment, 2016. 2, 4

[5] D. Cashman, M. Keller, H. Jeon, B. C. Kwon, and Q. Wang. A critical analysis of the usage of dimensionality reduction in four domains. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–20, 2025. doi: 10.1109/TVCG.2025.3567989 1, 2

[6] A. Colin Cameron and F. A. Windmeijer. An r-squared measure of goodness of fit for some common nonlinear regression models. *Journal of Econometrics*, 77(2):329–342, 1997. doi: 10.1016/S0304-4076(96)01818-0 4

[7] M. Correll and M. Gleicher. Error bars considered harmful: Exploring alternate encodings for mean and error. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2142–2151, 2014. doi: 10.1109/TVCG.2014.2346298 5

[8] J. H. Friedman. Exploratory projection pursuit. *Journal of the American statistical association*, 82(397):249–266, 1987. 1, 2

[9] P. Goffin, J. Boy, W. Willett, and P. Isenberg. An exploratory study of word-scale graphics in data-rich text documents. *IEEE Transactions on Visualization and Computer Graphics*, 23(10):2275–2287, 2017. doi: 10.1109/TVCG.2016.2618797 5

[10] T. M. Green, W. Ribarsky, and B. Fisher. Visual analytics for complex concepts using a human cognition model. In *2008 IEEE Symposium on Visual Analytics Science and Technology*, pp. 91–98, 2008. doi: 10.1109/VAST.2008.4677361 2, 5

[11] C. Healey and J. Enns. Attention and visual memory in visualization and computer graphics. *IEEE Transactions on Visualization and Computer Graphics*, 18(7):1170–1188, 2012. doi: 10.1109/TVCG.2011.127 2, 3

[12] K. G. Jamieson and R. Nowak. Active ranking using pairwise comparisons. *Advances in neural information processing systems*, 24, 2011. 3, 4

[13] H. Jeon, M. Aupetit, D. Shin, A. Cho, S. Park, and J. Seo. Measuring the validity of clustering validation datasets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(6):5045–5058, 2025. doi: 10.1109/TPAMI.2025.3548011 3

[14] H. Jeon, A. Cho, J. Jang, S. Lee, J. Hyun, H.-K. Ko, J. Jo, and J. Seo. Zadu: A python library for evaluating the reliability of dimensionality reduction embeddings. In *2023 IEEE Visualization and Visual Analytics (VIS)*, pp. 196–200. doi: 10.1109/VIS54172.2023.00048 2, 4

[15] H. Jeon, H. Lee, Y.-H. Kuo, T. Yang, D. Archambault, S. Ko, T. Fujiwara, K.-L. Ma, and J. Seo. Unveiling high-dimensional backstage: A survey for reliable visual analytics with dimensionality reduction. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25, 2025. doi: 10.1145/3706598.3713551 5

[16] H. Jeon, J. Park, S. Shin, and J. Seo. Stop misusing t-sne and umap for visual analytics, 2025. 1, 2, 5

[17] H. Jeon, G. J. Quadri, H. Lee, P. Rosen, D. A. Szafir, and J. Seo. Clams: A cluster ambiguity measure for estimating perceptual variability in visual clustering. *IEEE Transactions on Visualization and Computer Graphics*, 30(1):770–780, 2024. doi: 10.1109/TVCG.2023.3327201 2, 3, 4

[18] P. Joia, D. Coimbra, J. A. Cuminato, F. V. Paulovich, and L. G. Nonato. Local affine multidimensional projection. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2563–2571, 2011. doi: 10.1109/TVCG.2011.220 4

[19] D. Kahneman. Thinking, fast and slow. 2011. 2

[20] D. Kobak and G. C. Linderman. Initialization is critical for preserving global data structure in both t-sne and umap. *Nature biotechnology*, 39(2):156–157, 2021. doi: 10.1038/s41587-020-00809-z 5

[21] B. C. Kwon, B. Eysenbach, J. Verma, K. Ng, C. De Filippi, W. F. Stewart, and A. Perer. Clustervision: Visual supervision of unsupervised clustering. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):142–151, 2018. doi: 10.1109/TVCG.2017.2745085 2, 3, 4

[22] L. v. d. Maaten and G. Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008. 5

[23] L. McInnes, J. Healy, and J. Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018. 5

[24] C. Morariu, A. Bibal, R. Cutura, B. Frénay, and M. Sedlmair. Predicting user preferences of dimensionality reduction embedding quality. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):745–755, 2023. doi: 10.1109/TVCG.2022.3209449 2, 4

[25] Q. Nguyen, P. Eades, and S.-H. Hong. On the faithfulness of graph visualizations. In *2013 IEEE Pacific Visualization Symposium (PacificVis)*, pp. 209–216, 2013. doi: 10.1109/PacificVis.2013.6596147 1, 2

[26] L. G. Nonato and M. Aupetit. Multidimensional projection for visual analytics: Linking techniques with distortions, tasks, and layout enrichment. *IEEE Transactions on Visualization and Computer Graphics*, 25(8):2650–2673, 2019. doi: 10.1109/TVCG.2018.2846735 1, 2

[27] A. V. Pandey, J. Krause, C. Felix, J. Boy, and E. Bertini. Towards understanding human similarity perception in the analysis of large sets of scatter plots. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, p. 3659–3669, 2016. doi: 10.1145/2858036.2858155 3, 4

[28] J. Seo and B. Shneiderman. A rank-by-feature framework for interactive exploration of multidimensional data. *Information visualization*, 4(2):96–113, 2005. 1, 2

[29] L. South, D. Saffo, O. Vitek, C. Dunne, and M. A. Borkin. Effective use of likert scales in visualization evaluations: A systematic review. *Computer Graphics Forum*, 41(3):43–55, 2022. doi: 10.1111/cgf.14521 3

[30] H. Strobelt, D. Oelke, B. C. Kwon, T. Schreck, and H. Pfister. Guidelines for effective usage of text highlighting techniques. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):489–498, 2016. doi: 10.1109/TVCG.2015.2467759 5

[31] C. Tseng, A. Z. Wang, G. J. Quadri, and D. A. Szafir. Shape it up: An empirically grounded approach for designing shape palettes. *IEEE Transactions on Visualization and Computer Graphics*, 31(1):349–359, 2025. doi: 10.1109/TVCG.2024.3456385 5

[32] A. Tversky and D. Kahneman. Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *science*, 185(4157):1124–1131, 1974. 2

[33] A. C. Valdez, M. Ziefle, and M. Sedlmair. Priming and anchoring effects in visualization. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):584–594, 2018. doi: 10.1109/TVCG.2017.2744138 3

[34] J. Venna and S. Kaski. Local multidimensional scaling. *Neural Networks*, 19(6):889–899, 2006. Advances in Self Organising Maps - WSOM'05. doi: 10.1016/j.neunet.2006.05.014 2

[35] Y. Wang, K. Feng, X. Chu, J. Zhang, C.-W. Fu, M. Sedlmair, X. Yu, and B. Chen. A perception-driven approach to supervised dimensionality reduction for visualization. *IEEE Transactions on Visualization and Computer Graphics*, 24(5):1828–1840, 2018. doi: 10.1109/TVCG.2017.2701829 1, 2, 4

[36] M. Wattenberg, F. Viégas, and I. Johnson. How to use t-sne effectively. *Distill*, 2016. doi: 10.23915/distill.00002 2, 5

[37] L. Wilkinson, A. Anand, and R. Grossman. Graph-theoretic scagnostics. In *Information visualization, IEEE symposium on*, pp. 21–21. IEEE Computer Society, 2005. 4

[38] J. Xia, Y. Zhang, J. Song, Y. Chen, Y. Wang, and S. Liu. Revisiting dimensionality reduction techniques for visual cluster analysis: An empirical study. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):529–539, 2022. doi: 10.1109/TVCG.2021.3114694 2, 3