

Middlesex University Research Repository

An open access repository of

Middlesex University research

<http://eprints.mdx.ac.uk>

Xu, Kai ORCID logoORCID: <https://orcid.org/0000-0003-2242-5440>, Zhang, Leishi ORCID logoORCID: <https://orcid.org/0000-0002-3158-2328>, Pérez, Daniel, Nguyen, Phong H. and Ogilvie-Smith, Adam (2017) Evaluating interactive visualization of multidimensional data projection with feature transformation. *Multimodal Technologies and Interaction*, 1 (3) , 13. ISSN 2414-4088 [Article] (doi:10.3390/mti1030013)

Final accepted version (with author's formatting)

This version is available at: <https://eprints.mdx.ac.uk/22196/>

Copyright:

Middlesex University Research Repository makes the University's research available electronically.

Copyright and moral rights to this work are retained by the author and/or other copyright owners unless otherwise stated. The work is supplied on the understanding that any use for commercial gain is strictly forbidden. A copy may be downloaded for personal, non-commercial, research or study without prior permission and without charge.

Works, including theses and research projects, may not be reproduced in any format or medium, or extensive quotations taken from them, or their content changed in any way, without first obtaining permission in writing from the copyright holder(s). They may not be sold or exploited commercially in any format or medium without the prior written permission of the copyright holder(s).

Full bibliographic details must be given when referring to, or quoting from full items including the author's name, the title of the work, publication details where relevant (place, publisher, date), pagination, and for theses or dissertations the awarding institution, the degree type awarded, and the date of the award.

If you believe that any material held in the repository infringes copyright law, please contact the Repository Team at Middlesex University via the following email address:

eprints@mdx.ac.uk

The item will be removed from the repository while any claim is being investigated.

See also repository copyright: re-use policy: <http://eprints.mdx.ac.uk/policies.html#copy>

Article

Evaluating Interactive Visualization of Multidimensional Data Projection with Feature Transformation

Kai Xu ^{1,†,*}, Leishi Zhang ^{1,†}, Daniel Pérez ^{2,†}, Phong H. Nguyen ^{3,†} and Adam Ogilvie-Smith ^{4,†}

¹ Middlesex University, UK; {k.xu, l.x.zhang}@mdx.ac.uk

² University of Oviedo, Spain; dperez@isa.uniovi.es

³ City, University of London, UK; Phong.Nguyen.3@city.ac.uk

⁴ CGI IT UK Limited & Robert Gordon University, UK; adam.ogilvie-smith@cgi.com

* Correspondence: k.xu@mdx.ac.uk; Tel.: +44-20-8411-5510

† These authors contributed equally to this work.

Academic Editor: name

Version July 3, 2017 submitted to *Multimodal Technologies and Interact.*; Typeset by L^AT_EX using class file mdpi.cls

Abstract: There has been extensive research on dimensionality reduction techniques. While these make it possible to present visually the high-dimensional data in 2D or 3D, it remains a challenge for users to make sense of such projected data. Recently, interactive techniques, such as *Feature Transformation*, have been introduced to address this. This paper describes an user study that was designed to understand how the feature transformation techniques affect user's understanding of multi-dimensional data visualisation. It was compared with the traditional dimension reduction techniques, both unsupervised (PCA) and supervised (MCML). Thirty-one participants were recruited to detect visually clusters and outliers using visualisations produced by these techniques. Six different datasets with a range of dimensionality and data size were used in the experiment. Five of these are benchmark datasets, which makes it possible to compare with other studies using the same datasets. Both task accuracy and completion time were recorded for comparison. The results show that there is a strong case for the feature transformation technique. Participants performed best with the visualisations produced with high-level feature transformation, in terms of both accuracy and completion time. The improvements over other techniques are substantial, particularly in the case of the accuracy of the clustering task. However, visualising data with very high dimensionality (i.e., greater than 100 dimensions) remains a challenge.

Keywords: Human-centered computing; Empirical studies; Visual analytics; Dimensionality reduction

1. Introduction

With the explosive growth in the size of available data (Big Data), there is an increasing demand to help users better understand the Big Data they have. A large portion of the Big Data is high dimensional and is notoriously difficult for humans to comprehend because of the lack of physical analogy of data with more than three dimensions. Various dimension reduction techniques have been developed to reduce the data dimensions, so they can be visually displayed [1,2]. Dimensionality Reduction (DR) techniques such as Principal Component Analysis (PCA) and Multidimensional Scaling (MDS) allow analysts to project multidimensional data to a lower dimensional (2D or 3D) visual display as scatterplot diagrams where patterns such as groups and outliers can be easily identified. The approach is widely used for explorative analysis of large information spaces.

29 However, most of these techniques are not designed for human perception, but rather optimising
30 for certain metrics such as minimising the distance distortion after the projection. While these
31 techniques have been shown to be very useful, they inadvertently introduced difficulties for data
32 visualisation and sense making in lower dimensions such as visual cluttering that affects the
33 interpretation of a projection. Moreover, with increasing dimensionality and noise in the data,
34 such methods become less effective due to the curse of the dimensionality problem [3]. When
35 the dimensionality is high, the distance measure becomes less meaningful as all objects tend to be
36 similar and dissimilar in many ways, leading to points being projected to similar locations in the
37 projection space (over-plotting problem). Given a particular pattern recognition task, often not all
38 the recorded information is relevant. The irrelevant information will obscure the patterns in the
39 visualization, leading to blurred group boundaries and patterns being hidden behind overlapping
40 group boundaries. A recent study by Etemadpour et al. [4] compared five different DR techniques
41 from the user perception perspective, and the results confirmed the two issues discussed earlier.

42 Recently, there have been a number of works that aim to improve the existing dimension
43 reduction techniques by producing more understandable visualisation or allowing user interaction
44 during the process [5–10]. These are later summarized by Sacha et al. in their survey [11]. Among
45 these, one approach is to use a supervised DR technique that employs class labels to compute the
46 projection. Supervised DR helps improve visual clarity of projections but an uncluttered projection
47 can hardly be guaranteed. On the other hand for explorative analysis, it is important to gain
48 an overview of the data before detailed analysis [12]. Schaefer et al. [8] proposed a feature
49 transformation approach that can be applied in conjunction with any existing DR technique to reduce
50 the over-plotting problem and improve group separation in the visual space. The essential idea is to
51 integrate prior knowledge in the projection process by extending certain features in the original data
52 space before projection to achieve projections that better reveal hidden patterns in the data. Schaefer's
53 work is further extended by Pérez et al. [9,13] where interactive visualizations are proposed to
54 provide analysts with more flexibility and user control over the feature transformation process.
55 Although the feature transformation approach "distorts" the original feature space to a certain
56 degree, testing results in both Schaefer's and Pérez's work demonstrate a good compromise can
57 often be made between maintaining the original characteristics of the data and achieving better visual
58 clarification in the final projection. This was demonstrated through the assessment of the projections
59 using quality measures that showed an improvement of visual overlapping with a small variation
60 of the structural preservation. However, both works do not include user studies that evaluate the
61 effectiveness of the feature transformation approach from the perspective of user perception and
62 comprehension.

63 This paper describes an experiment studying the effectiveness of feature transformation
64 techniques in supporting analysts making sense of high-dimensional data. The participants were
65 asked to perform common analysis tasks, i.e., cluster and outlier identification, using 2D projection
66 (i.e., visualisation) produced by feature transformation and other DR methods. The experiment used
67 a number of benchmark datasets that cover a wide range of size and dimensionality. Both task
68 accuracy and completion time were recorded, and the result analyses show significant difference
69 among these methods.

70 The remainder of the paper is organised as follows: Section 2 provides a more complete and
71 in-depth discussion on the existing work related to the study. The details of the feature transformation
72 are described in Section 3. This is followed by experiment design, hypotheses, data sets and protocol
73 (Section 4). The experiment results are reported in Section 5, followed by in-depth discussions in
74 Section 6. Section 7 concludes the paper.

75 2. Related Work

76 An extensive range of DR techniques exist [1] that estimate the structure of data in a low
77 dimensional space. Classical methods such as Principal Component Analysis (PCA) [14] or

78 Multidimensional Scaling (MDS) [15] are based on linear approaches. Later non-linear techniques
79 were developed, for example Sammon proposed a version of the MDS algorithm [16] to compute
80 a projection that is able to represent non-linear structures in the data. In the beginning of the
81 21st century, newer non-linear techniques, based on neighbour embedding, were proposed. These
82 algorithms compute a manifold in a low-dimensional space from high dimensional data with an
83 underlying structure. Some of the best known examples are isometric embedding mapping or
84 Isomap [17], Laplacian Eigenmaps (LE) [18], locally linear embedding (LLE) [19], local tangent
85 subspace alignment (LTSA) [20] and t-Distributed Stochastic Neighbour Embedding (t-SNE) [21].

86 Moreover there are methods that use class information to guide the computation of the
87 projection, that is, supervised dimensionality reduction. Available supervised methods include the
88 *Linear Discriminative Analysis* (LDA) [22] that extracts the discriminative features to the class labels
89 and uses them to generate embedding, the *Neighborhood Components Analysis* (NCA) [23] that learns
90 a distance metric by finding a linear transformation of input data such that the average classification
91 performance is maximized in the projection space, and the *Maximally Collapsing Metric Learning*
92 (MCML) [24] that aims at learning a distance metric that tries to collapse all objects in the same class
93 to a single point and push objects in other classes far away.

94 DR techniques estimate the underlying structure and reveal relationships in multidimensional
95 data. However, due to noise and irrelevant attributes, a satisfactory projection is not always
96 obtained. Feature selection and transformations have been developed to improve performance of
97 many applications in several research fields [25,26]. A recent approach [8] transforms the feature
98 space by extending specific features of selected dimensions. The result can be applied to improve
99 group separation and reduce visual cluttering in the final embedding.

100 Furthermore, with the increasing size and complexity of data, it becomes more difficult to
101 generate meaningful projections in a fully automatic way. This leads to the development of interactive
102 multidimensional data projection techniques that facilitate interactive analysis by integrating the
103 analyst's knowledge about the data with the knowledge gained during the learning process.
104 Examples include the iPCA approach [6] that provides coordinated views for interactive analysis
105 of projections computed by PCA method and the iVisClassifier system [7] which improves data
106 exploration based on a supervised DR technique (LDA). Moreover, the DimStiller framework [27]
107 analyzes dimension reduction techniques with interactive controls that guide the user during the
108 analysis process and Dis-Function [28] provides an interactive visualization to define a distance
109 function. Similarly, AxiSketcher [10] allows user to change the projection dimensions interactively.
110 Perez et al. [9] proposed an interactive framework for feature space extension that allows the user to
111 incorporate class labels into the projection gradually. A hierarchical interpretation can be done using
112 the clusters of the initial projection and the class labels that are revealed by the method. More details
113 of this technique can be seen in Section 3.

114 The previously mentioned techniques are only part of a rich body of research that exists on
115 multidimensional data visualization. Integrating human knowledge into the analysis loop requires
116 understanding of the usability of the techniques mentioned. There are metrics for comparing the
117 quality of visualisation layouts, but they do not consider human perception. Examples include
118 the rank-based criteria framework by Lee and Verleysen [29] that is scale independent and many
119 high-dimensional data visualization quality metrics discussed in the survey by Bertini et al. [30].

120 There are a number of experiments studying the effectiveness of the projections from a user's
121 perspective. Different quality measures were proposed to evaluate scatterplots based on visual
122 perception, for example in terms of correlation [31], cluster separation [32], or both [33]. Lewis
123 et al. [34] investigated whether human evaluations of the projections are reliable, showing that user
124 experts are reasonably consistent about layout quality, but novices disagree on the quality. Recently,
125 a controlled user experiment [4] was performed to evaluate the human performance on multiple
126 tasks with different projection techniques. The results demonstrated that performance of projection
127 techniques varies with cognition task and is also data dependent. As far as we know, there has been

128 no user evaluation on the effectiveness of interactive visualization techniques for DR, which this work
129 aims to address.

130 3. Feature Transformation

The main idea of the interactive feature transformations proposed in [9] is to extend the attributes based on prior knowledge such as class labels. Assuming a data matrix \mathbf{X} where rows correspond to objects, columns are features, and the labels \mathbf{y} describe the categorical class of each object:

$$\mathbf{X} = [x_{ij}] \in \mathbb{R}^{n \times d} \quad \mathbf{y} = [y_i] \in \mathbb{N}^n \quad (1)$$

Being $i = 1, \dots, n$ and $j = 1, \dots, d$, where n is the number of points and d the number of dimensions. Then a new data matrix \mathbf{X}' is defined using the original data matrix \mathbf{X} and a new extended part $\tilde{\mathbf{X}}$ as follows:

$$\mathbf{X}' = [\mathbf{X} \mid \tilde{\mathbf{X}}] \quad (2)$$

This extended part corresponds to the statistical value based on the class labels. Here we use the mean values of each class member. Using the extension of the full feature space, then this part $\tilde{\mathbf{X}}$ corresponds to the centroids of each class member.

$$\tilde{\mathbf{X}} = [\tilde{\mathbf{x}}_i] \in \mathbb{R}^{n \times d} \quad \text{being } \tilde{\mathbf{x}}_i = \frac{1}{|C_{y_i}|} \sum_{i \in C_{y_i}} x_{ij} \quad (3)$$

131 where C_{y_i} is the set of objects belonging to class y_i .

A real parameter $\lambda \in [0, 1]$ allows the transition between original data (\mathbf{X}) and the extended part ($\tilde{\mathbf{X}}$) by applying simple changes in the metrics of the feature space using the matrix $\mathbf{W}_\lambda \in \mathbb{R}^{2d \times 2d}$. This matrix allows a weighted feature extension of the both parts of the matrix:

$$\mathbf{X}_{weight} = \mathbf{X}' \mathbf{W}_\lambda \quad (4)$$

where the matrix \mathbf{W}_λ is defined as follows:

$$\mathbf{W}_\lambda = \left(\begin{array}{c|c} (1-\lambda) \mathbf{I} & \mathbf{0} \\ \hline \mathbf{0} & \lambda \mathbf{I} \end{array} \right), \lambda \in \mathbb{R} \quad (5)$$

132 The parameter λ controls the changes between the original data structure and the centroids of
133 the introduced classes. These changes are independent of the technique used for computing the
134 projection. They produce a better separation of the introduced groups in the projections. Therefore
135 a visual improvement is achieved by means of a controlled modification of the original structure,
136 essentially a trade-off between visual clarification and structural preservation.

Below is an example using the *iris* flower data [35] that contains three species of iris: *setosa*, *virginica* and *versicolor*. Each species has four features: the length and width of the sepals and petals, measured in centimetres. This data set has been used in data analysis, as an example by many classification techniques in machine learning. Below is part of this data set represented as a matrix as described in Eq. 1:

$$\mathbf{X} = \left(\begin{array}{cccc|c} \vdots & \vdots & \vdots & \vdots & \vdots \\ 5.3 & 3.7 & 1.5 & 0.2 & setosa \\ 5.0 & 3.3 & 1.4 & 0.2 & setosa \\ 7.0 & 3.2 & 4.7 & 1.4 & virginica \\ 6.4 & 3.2 & 4.5 & 1.5 & virginica \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{array} \right) \quad (6)$$

The new data matrix X' (as in Eq. 2) is composed by the original data and the extended part using the class information from the species of *iris*. This extension is built using the mean feature vector for each class. For instance, if the mean feature vector for setosa is $m_{setosa} = (5.01, 3.43, 1.46, 0.25)$ and for virginica $m_{virginica} = (5.93, 2.77, 4.26, 1.32)$, then the new data matrix is as follows:

$$X' = \left(\begin{array}{cccc|cccc} \vdots & \vdots \\ 5.3 & 3.7 & 1.5 & 0.2 & 5.01 & 3.43 & 1.46 & 0.25 \\ 5.0 & 3.3 & 1.4 & 0.2 & 5.01 & 3.43 & 1.46 & 0.25 \\ 7.0 & 3.2 & 4.7 & 1.4 & 5.93 & 2.77 & 4.26 & 1.32 \\ 6.4 & 3.2 & 4.5 & 1.5 & 5.93 & 2.77 & 4.26 & 1.32 \\ \vdots & \vdots \end{array} \right) \quad (7)$$

137 The two parts of this new matrix are then weighted using the λ parameter defined in Eq. 5), where
 138 $\lambda = 0$ corresponds to the original matrix and $\lambda = 1$ leaves the extended part only. Finally, embeddings
 139 can be computed with a DR technique. Figure 1 shows the resulting projections with a series of λ
 140 values using a supervised DR technique MCML (as discussed in Section 2).

141 4. Experiment

142 A controlled experiment was conducted to evaluate the effectiveness of the interactive feature
 143 transformation technique. The goal is to understand its impact on high dimensional data
 144 visualisation, and consequently the user's ability to gain insight from the data. The experiment
 145 followed a within-subject design, and task accuracy and completion time were collected for
 146 comparison.

147 4.1. Pilot

148 A pilot study was conducted with three participants using the three conditions:

- 149 1. **Visualisation generated by PCA.** This is the same as the first condition in the final experiment
 150 (as described in Section 4.2).
- 151 2. **Static Feature Transformation.** The visualisation in this condition included the distortion
 152 introduced by the feature transformation. However, the user was not allowed to change the
 153 level of distortion, so the visualisation was static.
- 154 3. **Interactive Feature Transformation.** This is similar to the previous condition, but users could
 155 interactively change the level of distortion introduced by feature transform. This is achieved
 156 through a slider that changes the λ value.

157 Two issues were identified after analysing the results from the pilot study:

- 158 • Both Feature Transformation conditions performed better than the PCA condition. However,
 159 this is partly due to the fact that they utilise the clustering information, whereas PCA does not.
 160 We believe that this gave the two Feature Transformation conditions unfair advantage. As a
 161 result, we decided to introduce a new DR technique that also uses the clustering information.
- 162 • There was large variation in the performance of the interaction feature transformation
 163 condition. One participant always set the λ to the maximum value. As a result, each cluster
 164 transformed into a single point and the tasks became trivial. To avoid this scenario, we
 165 removed the interactive feature transformation condition, and replaced it with two static feature
 166 transformation conditions that have low and high level of distortion respectively.

167 4.2. Conditions

168 Four revised conditions were included in the main experiment:

- 169 1. Visualisation generated by **PCA**. The PCA is used as an example of DR technique that does
 170 not utilize clustering information. While it is possible to include additional DR method such
 171 as MDS, it will make the experiment overly long (it is close to one hour already with the four
 172 conditions) and it is not the focus of this study to compare DR techniques that do and do not
 173 use clustering information.
- 174 2. Visualisation generated by **MCML**. This represents supervised techniques that take into
 175 account the class labels information during dimension reduction, since feature transformation
 176 also requires class information. This should produce visually more separated results than
 177 PCA because of the additional class labels information. Because feature transformation is
 178 independent of the DR technique used, any technique that uses class label can be used, so long
 179 as it is also used in the two feature transformation conditions.
- 180 3. Visualisation generated by *low-level* feature transformation distortion (**FT-low**), based on the
 181 results of MCML. The visualisation in this condition includes low level distortion introduced
 182 by the feature transform, and the user was not allowed to change the level of distortion. A
 183 small λ value was selected manually to ensure considerable visual difference from the MCML
 184 condition. This is to emulate the scenario when a low level of distortion is introduced through
 185 interactive feature transformation.
- 186 4. Visualisation generated by *high-level* feature transformation distortion (**FT-high**), based on the
 187 results of MCML. This is similar to the last condition except that the distortion level was higher.
 188 A larger λ value was selected manually to a) ensure considerable visual difference from the
 189 FT-low condition, and 2) avoid reducing the question to a trivial task, e.g., every cluster is
 190 reduced to a single point. This is to emulate the scenario when a high level of distortion is
 191 introduced through interactive feature transformation.

192 We selected $\lambda = 0.1$ and $\lambda = 0.3$ for the FT-low and FT-high condition respectively after
 193 considering different λ levels for all the datasets used. This ensures for all datasets enough visual
 194 difference between these two conditions and from the MCML only condition (Condition 2), without
 195 reducing the question to a trivial task. For example, Figure 1 shows the distorted projections of the
 196 *iris* dataset with different λ values. Please note that the colour here is to help demonstrate the effect of
 197 feature transformation. All the data points appear black in the experiment; no clustering information
 was provided through colour.

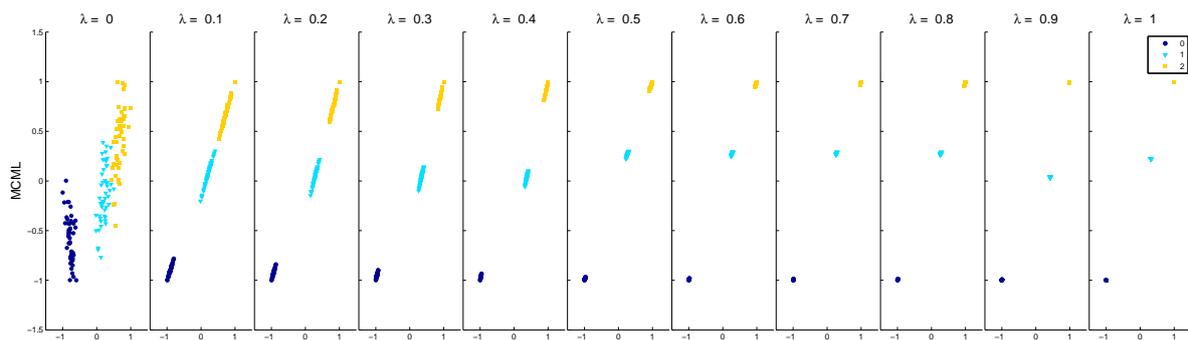


Figure 1. Projections of the *Iris* dataset with λ value from 0 to 1. The colour is used to help illustrate different clusters here, and was not used in the actual experiment.

198

199 4.3. Tasks

200 The participants were asked to complete two types of tasks during the experiment: identifying
 201 clustering and outlier. They are common in high-dimensional data analysis, and usually form the
 202 basis of more complex analysis tasks.

203 **Clustering:** The participants were asked to identify visually the number of clusters in the display.
 204 This is to test how well the resulting visualisation reveals the clustering structure within the
 205 original high-dimensional dataset.

206 **Outlier:** Similarly, this task requires participants to identify visually an outlier within the original
 207 dataset, which is another important property of high-dimensional data. To simplify the
 208 accuracy measurement, each dataset has exactly one outlier, so the answer can be either correct
 209 or incorrect. This avoids the case of ‘partially correct’ answers when there are two or more
 210 outliers.

211 We deliberately did not give formal definition of ‘clustering’ and ‘outlier’ during the training stage of
 212 the experiment. We wanted to see the participants’ intuition about these concepts, and its impact on
 213 task performance. As it turned out, all participants were able to grasp these concepts easily with the
 214 examples given during the training stage, and apply them successfully in the following tasks.

215 4.4. Datasets

216 We used a number of benchmark and synthetic datasets in the experiment. The goal was to cover
 217 a wide range of data size, dimensionality, and number of clusters in the dataset. The benchmark
 218 datasets are widely used by machine learning and visualization communities, and their details are
 219 in Table 1. The projections of all four conditions were checked before the experiment to ensure that
 220 the datasets do not favour any particular condition. We manually checked all the projections to make
 sure there were no trivial cases where clusters collapse into points.

Table 1. Experiment Datasets

Dataset	Points	Dimensions	Classes	Reference
HIV	78	159	6	[36]
Iris	147	4	3	[35]
Bbdm13	200	13	5	[37]
Tse300	244	46	8	[38]
Gaussian	500	10	5	[32]
Yeast	1452	8	10	[35]

221 For each dataset, a new point was added as the outlier. For half of the datasets, we added an
 outlier with extremely large value, using the formula below:

$$x > Q_3 + IQR \times 1.5$$

For the rest of the datasets, we added an outlier with extremely small value:

$$x < Q_1 - IQR \times 1.5$$

222 where Q_1 is the lower quartile (or the 25th percentile), Q_3 is the upper quartile (or the 75th percentile),
 223 and IQR the inter-quartile range ($Q_3 - Q_1$). This computation was applied to all dimensions in the
 224 corresponding dataset.

225 4.5. Participants and Procedure

226 We recruited 41 participants, with valid data collected from 31 of them. In several cases,
 227 the participant did not complete the experiment (participant can quit the experiment at any time
 228 without giving a reason) or there was a software error, so their data were not included for analysis.
 229 The participants were of mixed age range and technical background, including university students,
 230 administration staffs, and family and friends. It is voluntary to provide demographic information.
 231 In total, 11 participants chose to provide information about their age group (one under 19, six 19–25,
 232 and four 26–39) and gender (ten males, one female).

233 The study lasted approximately 45 minutes and consisted of three sections: training, experiment,
 234 and feedback. The training section started with the consent and demographic information form.
 235 After that, the two experiment tasks were explained using one example each. This part also showed
 236 the participants how to answer questions using the experiment software interface. The last part
 237 of training was practice, during which participants needed to complete one question for each task
 238 type. During practice, feedback was given if the participant did not answer correctly. Figure 2 is a
 239 screen-shot of the training interface.

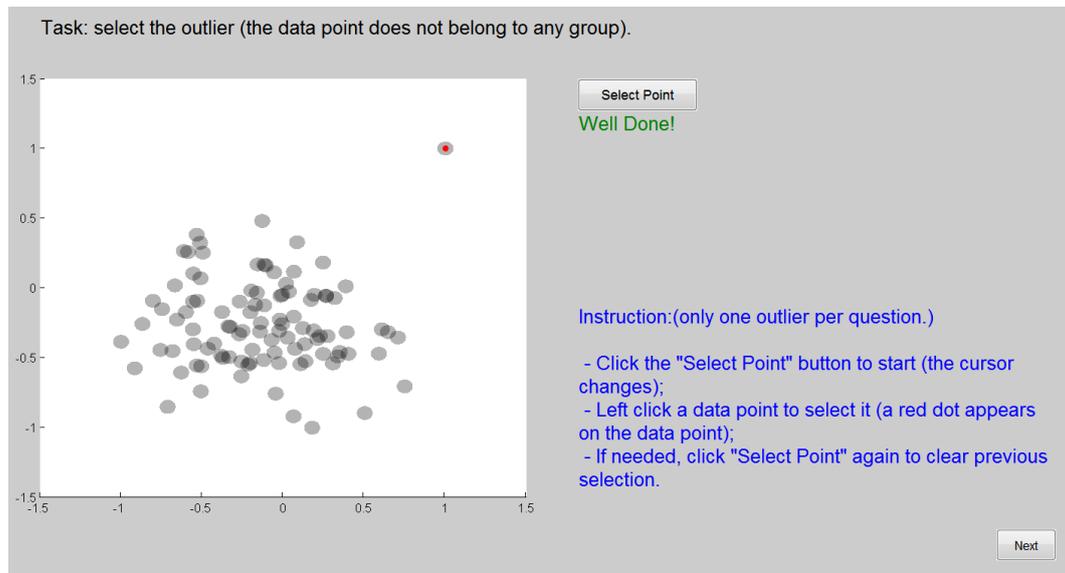


Figure 2. The training interface for the outlier task that includes the instructions (bottom right corner) and feedback ('Well Done!' for a correct answer).

240 The second section was the main experiment. The interface was the same as the training stage,
 241 except without feedback. As a within-subject design, each participant completed the two tasks on all
 242 six datasets under all four conditions. This led to in total $2 \times 6 \times 4 = 48$ questions. The order of the
 243 questions were counter balanced using Incomplete Block Design to avoid learning effect. Also, the
 244 same dataset appears quite differently under the four conditions, so it is unlikely that participants
 245 can recognise them under different conditions. Figure 3 shows the four conditions of the *HIV* dataset.
 246 Please note the data point colour and shape are for illustration only and they are not used in the actual
 247 experiment. It is not easy to recognize that these four projections are the same dataset, even when
 248 placing them next to each other with the colour and shape. The chance is very small that a participant
 249 can recognize so during the experiment when they appear randomly and without colour or shape.
 250 The task accuracy and completion time were recorded for further analysis.

251 The last section is feedback, during which the participants were asked to provide subjective
 252 comments about the tasks and visualisation. Because the participants are not aware of the four
 253 conditions (the information is not provided in the experiment), the feedback was not specific to
 254 experiment conditions.

255 4.6. Hypotheses

256 1. We hypothesise that participants will perform significantly better, in terms of both accuracy and
 257 completion time, with MCML than with PCA, because MCML takes advantage of additional
 258 clustering information. We hypothesise that this will be the case for both the clustering and
 259 outlier tasks, because the two require similar visual information, i.e., it is easier to identify
 260 outliers if the clustering is visually clear.

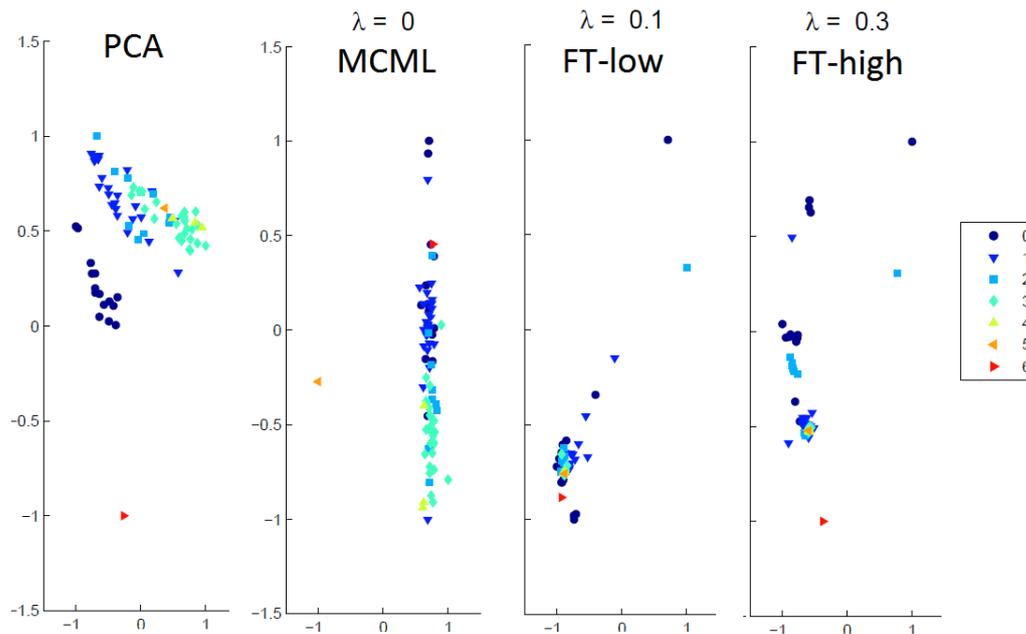


Figure 3. The four conditions of the *HIV* dataset. Please note the data point colour and shape are for illustration only and they are not used in the actual experiment. It is not easy to recognize that these four projections are the same dataset, even when placing them next to each other with the colour and shape. So when they appear randomly and without colour or shape, the chance that a participant could recognize them during the experiment was very small.

- 261 2. Similarly, we hypothesise that participants will perform significantly better with FT-low than
 262 MCML, in terms of both accuracy and completion time. The only difference between the two
 263 is the distortion introduced by the feature transformation, which makes the clustering/outlier
 264 structure visually more obvious.
- 265 3. Finally, We hypothesise that participants will perform significantly better with FT-high than
 266 FT-low, but only in accuracy. The higher level of distortion in FT-high will usually result in
 267 even clearer clustering/outlier structure, thus better accuracy. While it is likely the completion
 268 time will be shorter with FT-high, it can be already quite short with FT-low. As a result, the
 269 difference may not be significant.

270 5. Results

271 We used a repeated-measure analysis of variance (RM-ANOVA) to analyse the task accuracy
 272 and completion time of 31 participants with valid collected data. Accuracy was measured as the
 273 percentage of correct answers. Completion time was measured in seconds; however, it was not
 274 normally distributed as shown by the result of a Shapiro–Wilk test. We used the logarithm of
 275 completion time to normalize the skewed distribution.

276 5.1. Accuracy

277 Figure 4a shows the mean accuracy. A RM-ANOVA test showed a significant main effect of
 278 method ($F(3, 90) = 97.78, p < 10^{-27}$), task ($F(1, 30) = 32.01, p < 10^{-5}$), and the interaction method
 279 \times task ($F(3, 90) = 28.56, p < 10^{-12}$). Follow-up paired t-tests with Holm correction revealed that
 280 FT-high was significantly more accurate than FT-low ($p < 10^{-13}$), and both FT-low ($p < 10^{-8}$) and
 281 PCA ($p < .02$) were significantly more accurate than MCML. FT-low ($M = .54, SD = .50$) was more
 282 accurate than PCA ($M = .48, SD = .50$), but the difference was insignificant ($p = .09$). The results are

283 summarized in Figure 5a, where each line indicates a significant difference, pointing towards the less
 284 accurate condition.

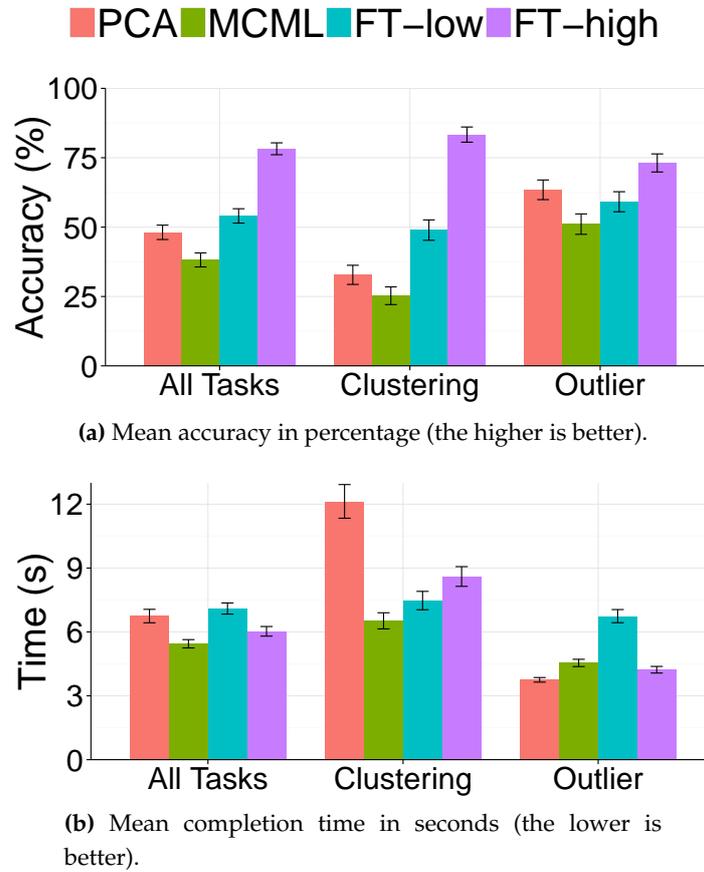


Figure 4. Mean accuracy and completion time overall and for each task.

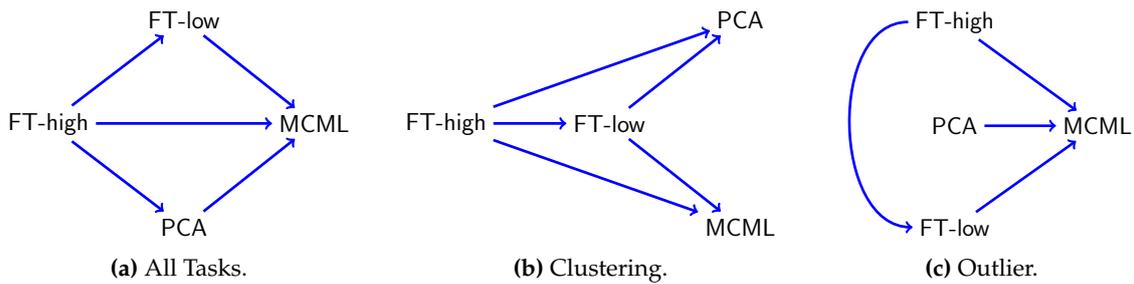


Figure 5. Significant results of paired t-tests for task accuracy. An arrow from condition A to condition B indicates that participants performed significantly more accurately under A than under B.

285 For Clustering task, a RM-ANOVA test showed a significant effect of method ($F(3,90) =$
 286 $74.52, p < 10^{-23}$). Follow-up paired t-tests with Holm correction revealed that FT-high was
 287 significantly more accurate than FT-low ($p < 10^{-14}$), and FT-low was significantly more accurate
 288 than PCA ($p < .001$). PCA ($M = .33, SD = .47$) was more accurate than MCML ($M = .25, SD = .44$),
 289 but the difference was insignificant ($p = .08$). The results are summarized in Figure 5b, following the
 290 same notation as in Figure 5a.

291 For Outlier task, a RM-ANOVA test showed a significant effect of method ($F(3,90) = 28.67, p <$
 292 10^{-12}). Follow-up paired t-tests with Holm correction revealed that FT-high was significantly more

293 accurate than FT-low ($p < 10^{-5}$), and FT-low was significantly more accurate than MCML ($p = .01$).
 294 PCA ($M = .63, SD = .48$) was more accurate than FT-low ($M = .59, SD = .49$), but the difference was
 295 insignificant ($p = .3$). Again, the results are summarized in Figure 5c, following the same notation.

296 5.2. Time

297 Figure 4b shows the mean completion time. A RM-ANOVA test showed a significant main effect
 298 of method ($F(3, 90) = 13.97, p < 10^{-6}$), task ($F(1, 30) = 87.46, p < 10^{-9}$), and the interaction method
 299 \times task ($F(3, 90) = 51.55, p < 10^{-18}$). Follow-up paired t-tests with Holm correction revealed that
 300 FT-high was significantly faster than FT-low ($p < .02$), and MCML was significantly faster than PCA
 301 ($p < .001$). MCML ($M = 5.44, SD = .19$) was faster than FT-high ($M = 6.03, SD = 0.23$), but the
 302 difference was insignificant ($p = .06$). The results are summarized in Figure 6a.

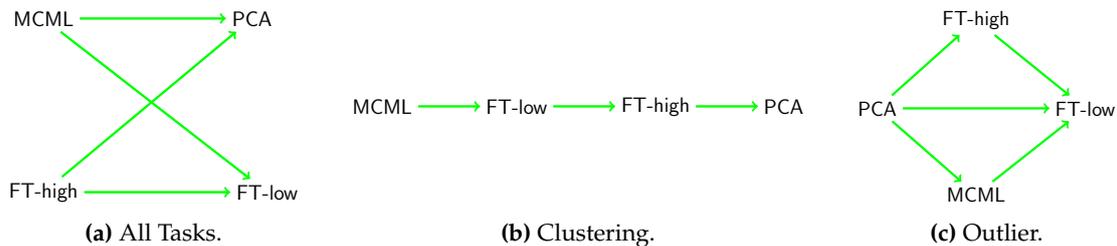


Figure 6. Significant results of paired t-tests for completion time. An arrow from condition A to condition B indicates that participants completed the tasks much faster under A than under B.

303 For Clustering task, a RM-ANOVA test showed a significant effect of method ($F(3, 90) =$
 304 $24.2, p < 10^{-10}$). Follow-up paired t-tests with Holm correction revealed that MCML was
 305 significantly faster than FT-low ($p < .023$), FT-low was significantly faster than FT-high ($p < .021$),
 306 and FT-high was significantly faster than PCA ($p < 10^{-5}$). The results are summarized in Figure 6b.

307 For Outlier task, a RM-ANOVA test showed a significant effect of method ($F(3, 90) = 55.46, p <$
 308 10^{-19}). Follow-up paired t-tests with Holm correction revealed that PCA was significantly faster than
 309 MCML ($p < 10^{-5}$), and MCML was significantly faster than FT-low ($p < 10^{-14}$). FT-high ($M =$
 310 $4.23, SD = .16$) was faster than MCML ($M = 4.54, SD = .17$), but the difference was insignificant
 311 ($p = .075$). The results are summarized in Figure 6c.

312 6. Discussions

313 6.1. Methods

314 Overall, FT-high performed the best: it is significantly more accurate than the three other
 315 conditions (Figure 5a) and took significantly less time than PCA and FT-low (Figure 6a). This supports
 316 our Hypothesis 3 and demonstrated that feature transformation can help users better understand
 317 multi-dimensional data. The improvement is more obvious in term of accuracy (Figure 4a) and less
 318 so for completion time (Figure 4b).

319 FT-low did not perform as well as we expected. It is significantly more accurate than MCML
 320 (Figure 5a), as in Hypothesis 2, but it required longer completion time than MCML (Figure 6a), which
 321 is different from what we hypothesised. Figure 7a and 7b shows the detailed completion time of
 322 clustering and outlier task respectively, ordered by dataset size. Figure 7a shows that the completion
 323 time under the FT-low is comparable to other conditions for the clustering task. However, its time is
 324 much longer than the rest for the outlier task (Figure 7b), especially the HIV dataset. As in Table 1,
 325 the HIV data has the highest dimensionality (159) among all the data sets, which can be the cause of
 326 the poor completion time of the outlier task under FT-low.

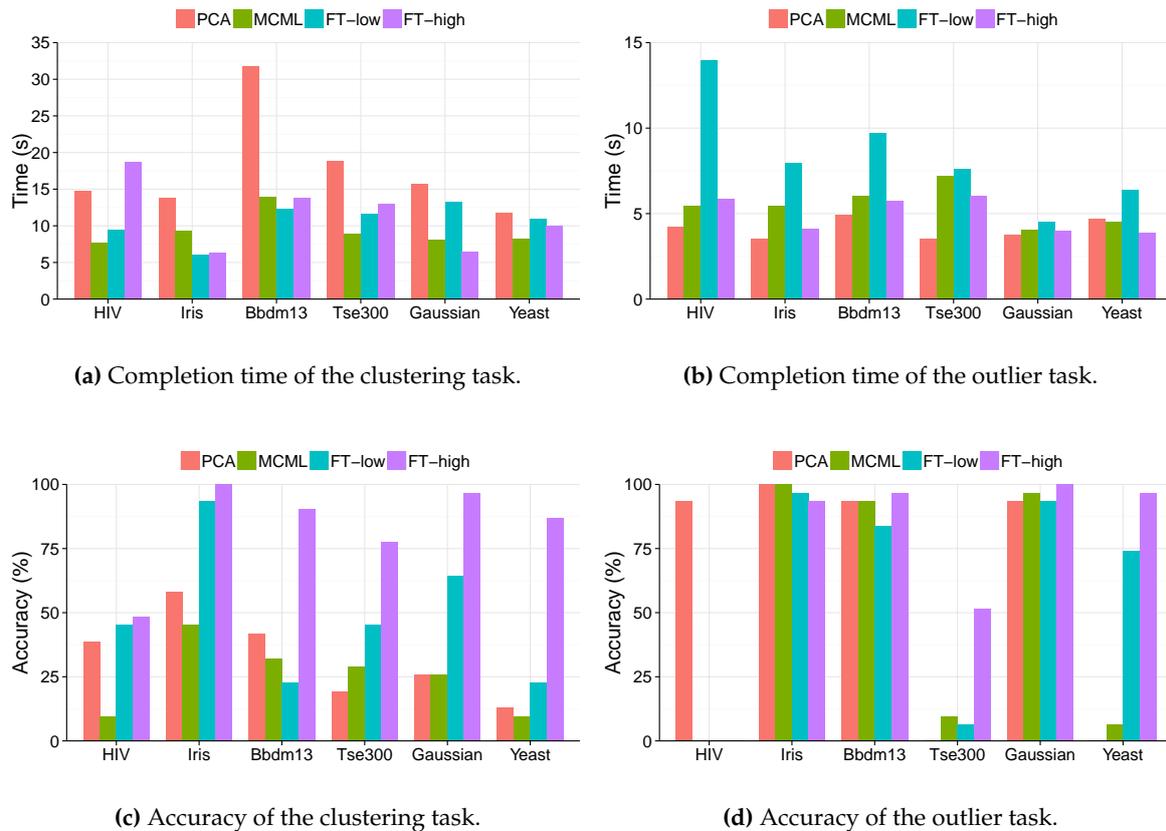


Figure 7. The results of the clustering and outlier task, ordered by data size.

327 The performance of the MCML condition is one of the surprises in the experiment results. It has
 328 the lowest task accuracy (Figure 5a), and this is the case for both the clustering (Figure 5b) and outlier
 329 task (Figure 5c). It was expected to out-perform PCA (Hypothesis 1), given that it takes advantage of
 330 the clustering information, i.e., which data point belongs to which cluster. Figure 7c and 7d show the
 331 accuracy of the clustering and outlier task respectively. For the clustering task, the accuracy of MCML
 332 is particularly poor for the HIV dataset. The results of the same dataset are even more extreme for the
 333 the outlier task (Figure 7d): except for PCA, the accuracy of the other three methods are all 0%. The
 334 high dimensionality of the HIV dataset may be the cause here, particularly for the outlier task; it also
 335 led to long completion times for the outlier task for FT-low (Figure 7b) as discussed earlier. Figure 3
 336 shows the four conditions of the HIV dataset with the outlier inserted. The outlier is marked as class
 337 6 (the red triangle). For clustering, it is obvious that the clusters are not well separated in all cases,
 338 particularly for MCML, which may explain the results in Figure 7c. Similarly, it is easy to see that the
 339 outlier is not well separated from other data points in MCML and FT-low, which makes it difficult to
 340 spot when the colouring is removed (no colouring was used in the experiment.) While the outlier is
 341 better separated in FT-high, the two data points in the top-right corner may make it difficult to select
 342 the true outlier. This can be the reason for the poor performance of these three conditions, as shown
 343 in Figure 7d.

344 The completion time of MCML is surprisingly fast. Overall there is no significant difference
 345 between MCML and FT-high, which was expected to have the fastest completion time (Figure 6a).
 346 However, the detailed results in Figure 7a and 7b show that the absolute difference is not that
 347 substantial, even if it is statistically significant.

348 Finally, PCA performed better than expected in the experiment. It was expected to be the least
 349 accurate method overall (Hypothesis 1), but this is not the case (Figure 5a). The poor performance

350 of other methods on the HIV dataset, particularly the outlier task (Figure 7d), can be a contributing
 351 factor. Also, it is interesting that its accuracy varied dramatically for the outlier task among the
 352 datasets (Figure 7d): while it performed extremely well for the HIV dataset, the accuracy dropped
 353 to 0% for the *Tse300* and *Yeast* dataset. Time-wise, PCA is comparable to other methods, except for
 354 the clustering task (Figure 4b). The detailed results in Figure 7a show that this may be the result
 355 of the large difference with the *Bbdm13* dataset. However, further investigation into the individual
 356 completion time did not reveal any anomaly. Overall, being one of the classic DR methods, PCA
 357 does a reasonably good job to support user understanding even though it was not designed for this
 358 purpose.

359 6.2. Data Size and Dimensionality

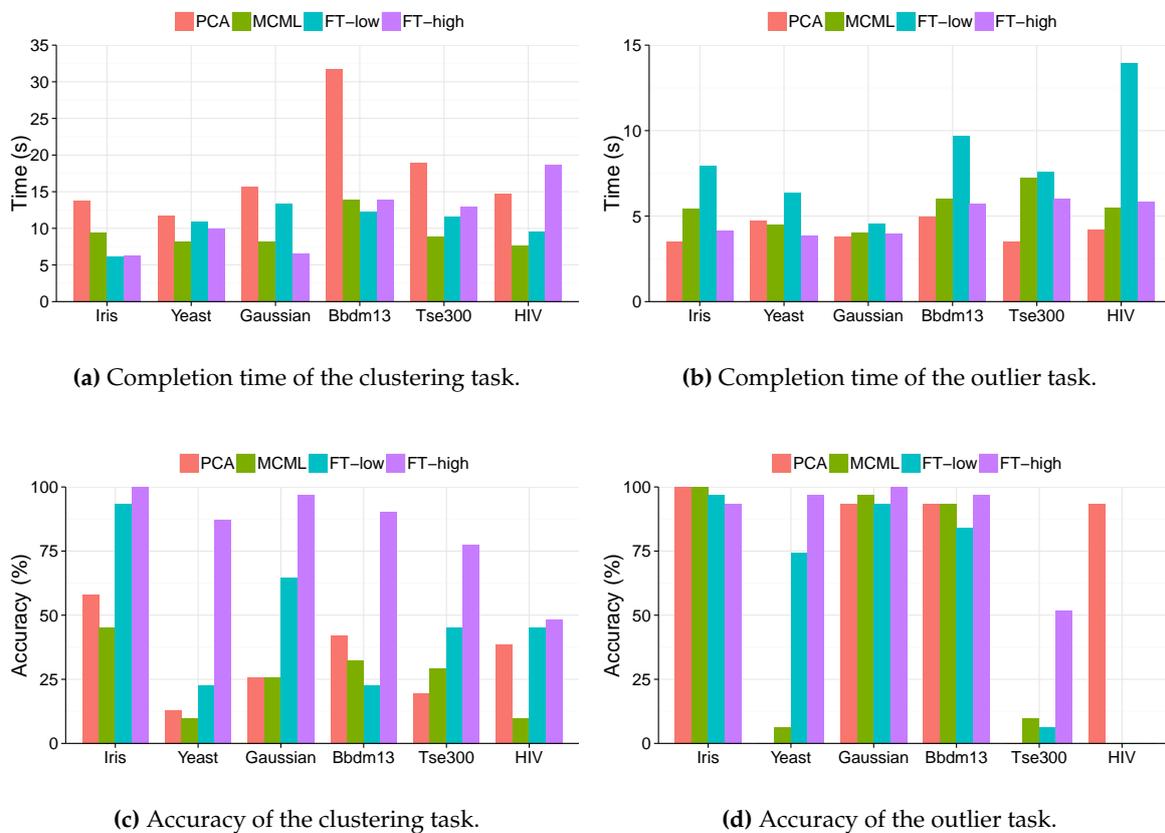


Figure 8. The results of the clustering and outlier task, ordered by data set dimensionality.

360 It is important to understand how the performance of different methods scale with data. This
 361 is particularly relevant if these approaches are to be applied to Big Data. There are two possible
 362 scaling: data size, i.e., number of data points, and data dimensionality. The data sets in Figure 7a
 363 to 7d are ordered by their sizes, i.e., increasing from left to right. Figure 7a and 7b show that the
 364 completion time does not increase with data size. In fact, it took longer with the HIV dataset, which
 365 has the smallest number of data points (78), than the Yeast dataset, which has the largest number
 366 of data points (1452). This is the result of *pre-attentative* visual processing [39]: users use the data
 367 point *location*, which is one of pre-attentative visual features, to decide clustering structure, and the
 368 processing of such visual feature takes constant time, regardless the number of points. This is one
 369 of the main advantages of data visualisation: information represented with pre-attentative visual
 370 features can be processed very quickly irrespective of the data size. There is no obvious trend in

371 the task accuracy (Figure 7c and 7d), either. Other factors, such as the complexity of the clustering
372 structure and appropriateness of the visualisation method, may have more of an impact on the task
373 performance than the data size does.

374 Figure 8 shows the same results as in Figure 7a to 7d, but ordered by the data set dimensionality,
375 increasing from left to right. There is a weak trend of increasing completion time with the data
376 dimensionality (Figure 8a and 8b), which is an indicator of the data set complexity. The trend is less
377 clear for the accuracy results (Figure 8c and 8d), possibly because the suitability of the visualisation
378 method is the main factor. For example, PCA led to low accuracy with the Yeast and Tse300 dataset,
379 and performed very well with the result of data sets (Figure 8d).

380 6.3. Tasks and Participants

381 While not the main goal of this study, we also examined the performance difference between the
382 two tasks used in the study. The results show that in general the clustering task is more difficult
383 than the outlier task, which is supported by both the performance metrics and user preference.
384 The clustering task has significantly lower accuracy than the outlier task (t -test, $p < 10^{-5}$), and
385 the difference is obvious as shown in Figure 10a. Similarly, the clustering task took significantly
386 longer to complete than the outlier task (t -test, $p < 10^{-6}$), and the difference is sizeable as shown in
387 Figure 10b. User preference data (Figure 10c) showed a similar pattern, with the clustering task being
388 perceived as significantly more difficult than the outlier task (Fisher's exact test, $p < 10^{-6}$). This
389 strengthens the argument for applying a Feature Transformation type of approach when visualising
390 high dimensional data: FT-high (high-level of feature transformation) was the only condition with
391 more than 50% percent accuracy for the clustering task and beat the second best option FT-low by a
392 healthy 30% margin (Figure 4a).

393 There is a weak correlation between user preference and performance. For the clustering task,
394 the Spearman's correlation coefficient is 0.0692892 (almost no relation) between rating and accuracy,
395 and 0.3012618 (a weak positive – more difficult, more time spent) between rating and completion time.
396 Similarly, for the outlier task, the Spearman's correlation coefficient is -0.2622217 (a weak negative –
397 more difficult, less accurate) between rating and accuracy and -0.1281551 (a weak positive) between
398 rating and completion time.

399 We analysed the relationship between participants' performance and their demographic
400 information such as age group. Both completion time and accuracy of the three age groups are shown
401 in Figure 9, and they appear to be similar across the groups. The small number of participants (11)
402 who provided their information does not allow any meaningful significance tests.

403 Finally, we checked the performance variations among the individuals participated the study.
404 Figure 11 shows the average completion time (Figure 11a) and accuracy (Figure 11b) of each
405 participant across all tasks. There appears to be larger variation among the performance of the
406 completion time than that of the accuracy, and this is confirmed by their coefficient of variation:
407 0.4198652 for time and 0.129759 for accuracy.

408 We further investigated participant 14 who had the longest completion time. For the clustering
409 task, his completion time (Figure 12a) appears to be similar to the average time (Figure 7a) except
410 for a few questions, such as Bbdm13-PCA and HIV-FT-high. We speculate that he struggled with
411 these questions and spent long time to find the right answers: he correctly answered four out of five
412 questions that he spent most time on (>40s). This is much higher than the average accuracy. Similarly
413 for outlier task, his completion time is also close to the average except for one question (Iris-MCML),
414 which he answered correctly.

415 6.4. Limitations

416 As with any user study, this experiment is not without its limitations. For example, the tasks
417 were simplified to make the experiment manageable, and thus less representative of the real-world
418 scenario: users were not able to interactively choose the λ value for the feature transformation and

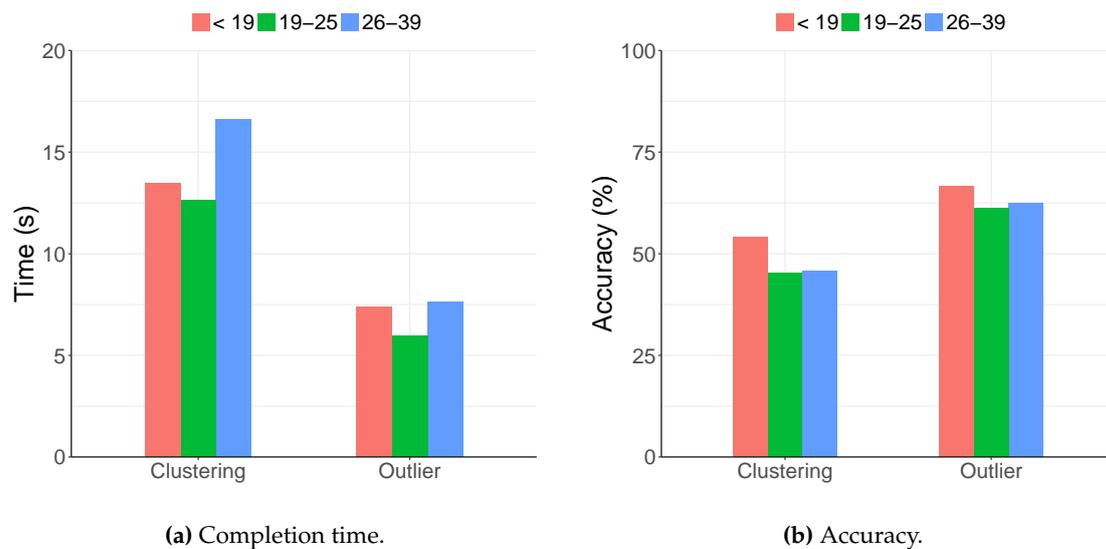


Figure 9. Performance in different age groups and tasks.

419 there is always one outlier in the outlier-detecting task. We were aware of these limitations, and
 420 consulted the end users during the experiment design stage. While not fully realistic, they thought
 421 the simplified tasks were good enough approximation of the real-world analysis as the first step to
 422 explore the performance difference among these techniques. More realistic set-up will be explored in
 423 the further studies.

424 7. Conclusions

425 This paper described a user study that was designed to understand how feature transformation
 426 technique affects the user's understanding of multi-dimensional data visualisation. Four different
 427 conditions were included: PCA, MCML, low-level feature transformation (FT-low), and high-level
 428 feature transformation (FT-high). Thirty-one participants were recruited to detect clusters and
 429 outliers using visualisation of six different datasets. Both task accuracy and completion time were
 430 recorded for comparison.

431 7.1. Techniques

- 432 • There is a strong case for the feature transformation technique. Participants performed best with
 433 the visualisation produced with high-level feature transform (FT-high), in term of both accuracy
 434 and completion time. The improvements over other techniques were substantial, particularly
 435 in the case of the accuracy of the clustering task.
- 436 • Low-level feature transformation has a lesser impact on visualisation readability, and as a result
 437 does not have a clear advantage over existing techniques, represented by MCML (supervised
 438 DR) and PCA (un-supervised DR).
- 439 • Very high dimensional data seems to be a challenge for all the techniques, but particularly
 440 MCML and to certain extend FT-low. MCML performed poorly with the HIV dataset, which
 441 has a much higher dimensionality (139) than the rest of the data sets.
- 442 • The results of PCA were better than expected; its performance was close to that of the FT-low
 443 and MCML. Also, it performed surprisingly well on the very high-dimensional HIV dataset,
 444 matching the results of FT-high.

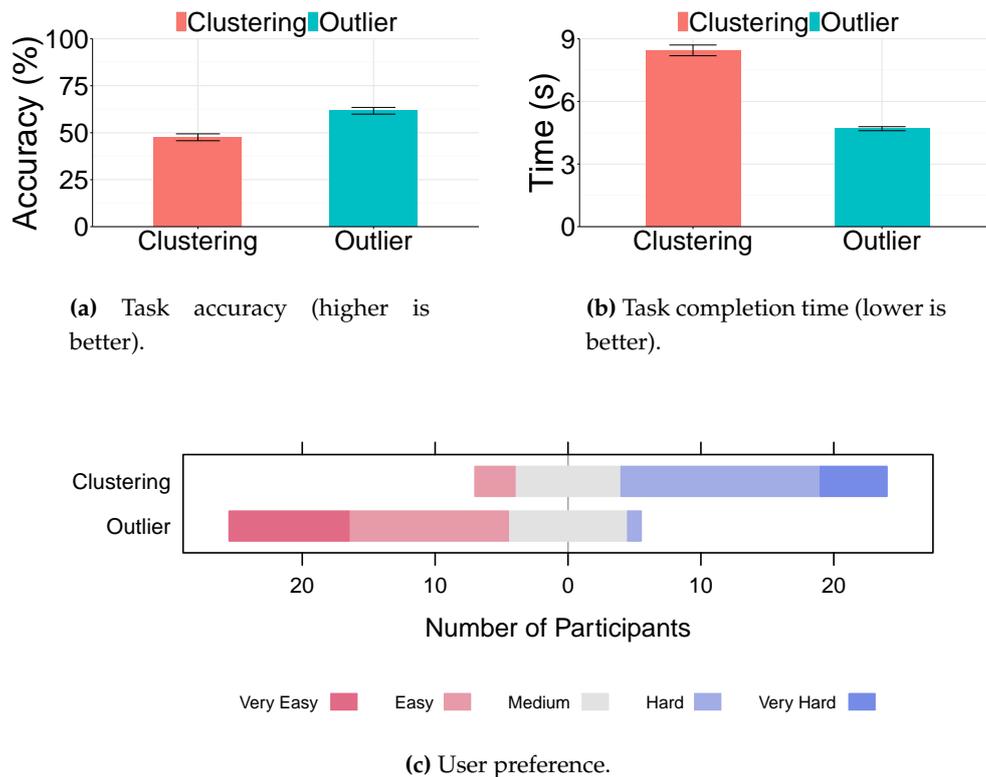


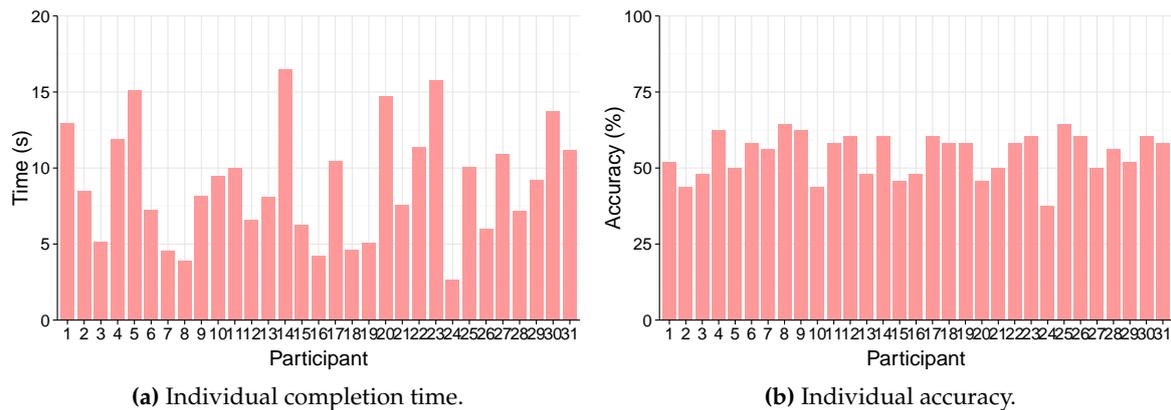
Figure 10. Clustering vs. outlier task

445 7.2. Scalability

- 446 • All the visualisation methods scaled well with data size, particularly with completion time.
 447 There is no apparent increase in completion time as the number of data points grow (20 fold
 448 difference between the size of the smallest and largest dataset). This is the result of human
 449 pre-attentative visual processing, which requires constant time regardless of data size. This
 450 makes visualisation an effective tool for understanding large data.
 451 • The data dimensionality appears to have a larger impact on the user performance than the data
 452 size. It leads to an increase in completion time as the data dimensionality grows. The effect
 453 on the accuracy is less clear, with the performance of a certain method changes dramatically
 454 between data sets. This indicates that the suitability of a visualisation method to a particular
 455 data set can be the dominant factor for task accuracy.

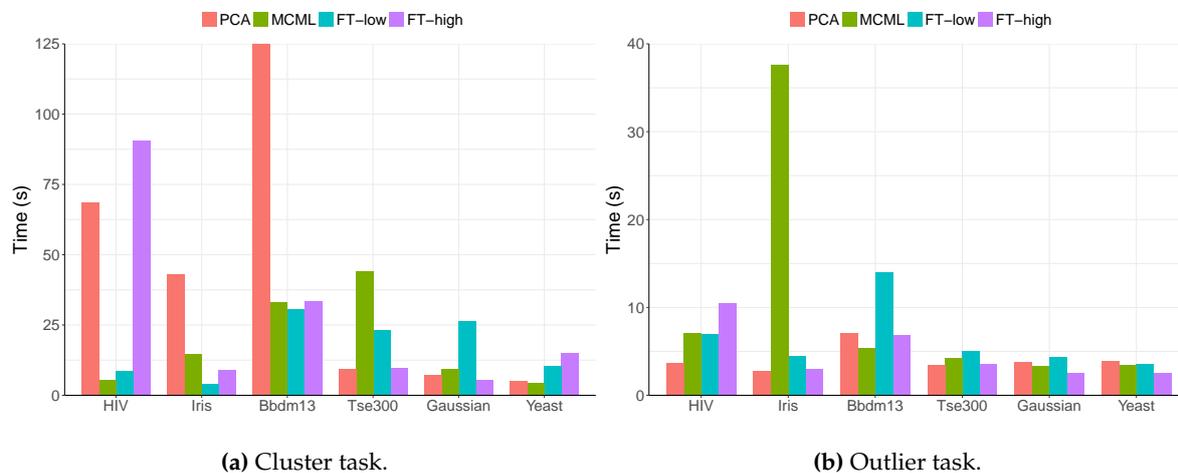
456 7.3. Tasks and Participants

- 457 • Clustering is a more difficult task than outlier identification. Its accuracy is significantly lower
 458 and took significantly longer to complete. Except for FT-high, all techniques led to accuracy of
 459 only around 25%. This demonstrates that it is almost impossible to perform visual clustering
 460 analysis without feature transformation.
 461 • Outlier detection is the relatively easier task, with faster completion time and higher accuracy.
 462 However, its accuracy varies dramatically between data sets and techniques. One technique
 463 can have close to 100% accuracy on one dataset, but 0% on another data set with similar size
 464 and dimensionality. Therefore, selecting an effective visualisation method is important for a
 465 successful analysis.



(a) Individual completion time.

(b) Individual accuracy.

Figure 11. Individual performance.

(a) Cluster task.

(b) Outlier task.

Figure 12. Time completion of participant 14 broken down by condition and dataset.

- 466 • Participants perceived clustering as the significantly more difficult task, but there was only a
 467 weak correlation between user preference and actual performance. There is a larger variation
 468 among the individual completion time than that of the task accuracy.

469 In summary, the experiment results showed that visualisation is an effective approach for high
 470 dimensional data analysis, because it does not require additional time as the data size grows. The
 471 feature transformation technique can significantly improve user's understanding, increasing task
 472 accuracy and reducing completion time simultaneously. It is almost impossible to obtain meaningful
 473 results from visual clustering analysis without feature transformation. Visualising data with very
 474 high dimensionality (i.e., greater than 100 dimensions) remains a challenge. It will be an interesting
 475 future work to evaluate further the effectiveness of the feature transformation with more realistic task
 476 settings and when in combination with more advanced approaches such as t-SNE.

477 Acknowledgement

478 The authors would like to thank CGI Group for their financial support, without which the study
 479 would not be possible. They also would like to thank Dr. Peter Passmore for his careful proof reading
 480 of the manuscript.

481 References

- 482 1. Lee, J.; Verleysen, M. *Nonlinear dimensionality reduction*; Springer, 2007.
- 483 2. Van der Maaten, L. An introduction to dimensionality reduction using matlab. *Report* **2007**, 1201, 62.
- 484 3. Donoho, D.L. High-dimensional data analysis: the curses and blessings of dimensionality. *Proceedings*
- 485 *of American Mathematical Society Conf. Math Challenges of the 21st Century* (2000), 2000.
- 486 4. Etemadpour, R.; Motta, R.; de Souza Paiva, J.; Minghim, R.; Ferreira de Oliveira, M.; Linsen, L.
- 487 Perception-Based Evaluation of Projection Methods for Multidimensional Data Visualization. *IEEE*
- 488 *Transactions on Visualization and Computer Graphics* **2015**, 21, 81–94.
- 489 5. Paulovich, F.; Silva, C.; Nonato, L. User-Centered Multidimensional Projection Techniques. *Computing in*
- 490 *Science Engineering* **2012**, 14, 74–81.
- 491 6. Jeong, D.H.; Ziemkiewicz, C.; Fisher, B.; Ribarsky, W.; Chang, R. iPCA: An Interactive System for
- 492 PCA-based Visual Analytics. *Computer Graphics Forum* **2009**, 28, 767–774.
- 493 7. Choo, J.; Lee, H.; Kihm, J.; Park, H. iVisClassifier: An interactive visual analytics system for classification
- 494 based on supervised dimension reduction. *Visual Analytics Science and Technology (VAST), 2010 IEEE*
- 495 *Symposium on*, 2010, pp. 27–34.
- 496 8. Schäfer, M.; Zhang, L.; Schreck, T.; Tatu, A.; Lee, J.A.; Verleysen, M.; Keim, D.A. Improving
- 497 projection-based data analysis by feature space transformations. *IS&T/SPIE Electronic Imaging.*
- 498 *International Society for Optics and Photonics*, 2013, pp. 86540H–86540H.
- 499 9. Pérez, D.; Zhang, L.; Schaefer, M.; Schreck, T.; Keim, D.; Díaz, I. Interactive feature space extension for
- 500 multidimensional data projection. *Neurocomputing* **2015**, 150, Part B, 611–626.
- 501 10. Kwon, B.C.; Kim, H.; Wall, E.; Choo, J.; Park, H.; Endert, A. AxiSketcher: Interactive Nonlinear Axis
- 502 Mapping of Visualizations through User Drawings. *IEEE Transactions on Visualization and Computer*
- 503 *Graphics* **2017**, 23, 221–230.
- 504 11. Sacha, D.; Zhang, L.; Sedlmair, M.; Lee, J.A.; Peltonen, J.; Weiskopf, D.; North, S.C.; Keim, D.A. Visual
- 505 Interaction with Dimensionality Reduction: A Structured Literature Analysis. *IEEE Transactions on*
- 506 *Visualization and Computer Graphics* **2017**, 23, 241–250.
- 507 12. Keim, D.A.; Kohlhammer, J.; Ellis, G.; Mansmann, F. *Mastering The Information Age - Solving Problems with*
- 508 *Visual Analytics*; Eurographics, 2010.
- 509 13. Pérez, D.; Zhang, L.; Schaefer, M.; Schreck, T.; Keim, D.; Díaz, I. Interactive Visualization and Feature
- 510 Transformation for Multidimensional Data Projection. *Proc. EuroVis Workshop on Visual Analytics*
- 511 *Using Multidimensional Projections*, 2013.
- 512 14. Jolliffe, I. Principal component analysis. *Spring-verlag, New York* **1986**.
- 513 15. Torgerson, W. Multidimensional scaling: I. Theory and method. *Psychometrika* **1952**, 17, 401–419.
- 514 16. Sammon Jr, J.W. A nonlinear mapping for data structure analysis. *Computers, IEEE Transactions on* **1969**,
- 515 *100*, 401–409.
- 516 17. Tenenbaum, J.B.; De Silva, V.; Langford, J.C. A global geometric framework for nonlinear dimensionality
- 517 reduction. *Science* **2000**, 290, 2319–2323.
- 518 18. Belkin, M.; Niyogi, P. Laplacian eigenmaps and spectral techniques for embedding and clustering. *NIPS*,
- 519 *2001*, Vol. 14, pp. 585–591.
- 520 19. Roweis, S.T.; Saul, L.K. Nonlinear dimensionality reduction by locally linear embedding. *Science* **2000**,
- 521 *290*, 2323–2326.
- 522 20. Zhang, Z.y.; Zha, H.y. Principal manifolds and nonlinear dimensionality reduction via tangent space
- 523 alignment. *Journal of Shanghai University (English Edition)* **2004**, 8, 406–424.
- 524 21. Van der Maaten, L.; Hinton, G. Visualizing Data using t-SNE. *Journal of Machine Learning Research* **2008**,
- 525 *9*, 2579–2605.
- 526 22. Fisher, R.A. The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics* **1936**,
- 527 *7*, 179–188.
- 528 23. Goldberger, J.; Roweis, S.; Hinton, G.; Salakhutdinov, R. Neighbourhood components analysis. *NIPS'04*
- 529 **2004**.
- 530 24. Globerson, A.; Roweis, S. Metric learning by collapsing classes. *Nips*, 2005, Vol. 18, pp. 451–458.
- 531 25. Blum, A.L.; Langley, P. Selection of relevant features and examples in machine learning. *Artificial*
- 532 *intelligence* **1997**, 97, 245–271.

- 533 26. Guyon, I.; Elisseeff, A. An introduction to variable and feature selection. *The Journal of Machine Learning*
534 *Research* **2003**, *3*, 1157–1182.
- 535 27. Ingram, S.; Munzner, T.; Irvine, V.; Tory, M.; Bergner, S.; Möller, T. DimStiller: Workflows for dimensional
536 analysis and reduction. *Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on*, 2010,
537 pp. 3–10.
- 538 28. Brown, E.T.; Liu, J.; Brodley, C.E.; Chang, R. Dis-function: Learning distance functions interactively.
539 *Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on*, 2012, pp. 83–92.
- 540 29. Lee, J.A.; Verleysen, M. Scale-independent quality criteria for dimensionality reduction. *Pattern*
541 *Recognition Letters* **2010**, *31*, 2248–2257.
- 542 30. Bertini, E.; Tatu, A.; Keim, D. Quality Metrics in High-Dimensional Data Visualization: An Overview
543 and Systematization. *Proceedings of the IEEE Symposium on IEEE Information Visualization (InfoVis) 2011*,
544 *17*, 2203–2212.
- 545 31. Rensink, R.A.; Baldrige, G. The perception of correlation in scatterplots. *Computer Graphics Forum*.
546 *Wiley Online Library*, 2010, Vol. 29, pp. 1203–1210.
- 547 32. Sedlmair, M.; Tatu, A.; Munzner, T.; Tory, M. A taxonomy of visual cluster separation factors. *Computer*
548 *Graphics Forum*. *Wiley Online Library*, 2012, Vol. 31, pp. 1335–1344.
- 549 33. Albuquerque, G.; Eisemann, M.; Magnor, M. Perception-based visual quality measures. *Visual Analytics*
550 *Science and Technology (VAST), 2011 IEEE Conference on*. IEEE, 2011, pp. 13–20.
- 551 34. Lewis, J.M.; Van Der Maaten, L.; de Sa, V. A behavioral investigation of dimensionality reduction. *Proc.*
552 *34th Conf. of the Cognitive Science Society (CogSci)*, 2012, pp. 671–676.
- 553 35. Frank, A.; Asuncion, A. UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. Irvine,
554 CA: University of California. *School of Information and Computer Science* **2010**, 213.
- 555 36. Sips, M.; Neubert, B.; Lewis, J.P.; Hanrahan, P. Selecting good views of high-dimensional data using class
556 consistency. *Comput. Graph. Forum* **2009**, *28*, 831–838.
- 557 37. of Massachusetts, U. Statistical Data and Software Help, 2011.
558 <http://www.umass.edu/statdata/statdata/>.
- 559 38. Inc., V.T. VisuMap Data Repository., 2011. <http://www.visumap.net/>.
- 560 39. Ware, C. *Information Visualization: Perception for Design*, 2nd revised edition edition ed.; Morgan Kaufmann
561 Publishers In: San Francisco, CA, 2004.

562 © 2017 by the authors. Submitted to *Multimodal Technologies and Interact.* for possible open
563 access publication under the terms and conditions of the Creative Commons Attribution license
564 (<http://creativecommons.org/licenses/by/4.0/>)