A Study on the Benefit of a User-Controlled Radial Tour for Variable Importance for Structure in High-Dimensional Data

Nicholas Spyrison
Monash University
Australia
nicholas.spyrison@monash.edu
ORCiD: 0000-0002-8417-0212

Dianne Cook
Monash University
Australia
dicook@monash.edu
ORCiD: 0000-0002-3813-7155

Kim Marriott
Monash University
Australia
kim.marriott@monash.edu
ORCiD: 0000-0002-9813-0377

***Abstract*—Principal component analysis is a long-standing go-to method for exploring multivariate data. The principal components are linear combinations of the original variables, which are ordered from largest to smallest variance. The first few typically provide a good visual summary of the data. *Tours* also make linear projections of the original variables but offer many different views, like examining the data from different directions. The grand tour shows a smooth sequence of projections as an animation following interpolations between random target bases. The manual radial tour rotates the selected variable's contribution, into and out of a projection. This allows the importance of the variable to structure in the projection to be assessed. This work describes a within-participants user study evaluating the radial tour's efficacy compared with principal component analysis and the grand tour. A supervised classification task is assigned to participants who evaluate variable attribution of the separation between two classes. Their accuracy in assigning the variable importance is measured across various factors. Data were collected from 108 crowdsourced participants, who performed two trials with each visual for 648 trials in total. Mixed model regression provides evidence that the radial tour increases accuracy over the alternatives. Participants also reported a preference for the radial tour in comparison to the other two methods.**

***Index Terms*—Linear Dimension Reduction; Visual Analytics; Grand Tour; Data Science; Machine Learning; Data Clustering; Explainable Artificial Intelligence**

I. Introduction

Despite decades of research, multivariate data continues to provide fascinating challenges for visualization. Data visualization is important because it is a key element of exploratory data analysis (EDA, Tukey 1977) for assessing model assumptions, and as a cross-check on numerical summarization (Anscombe 1973; Matejka and Fitzmaurice 2017; Yanai and Lercher 2020). One of the challenges is determining whether a new technique yields a better perception of information than current practices for multivariate data.

Dimension reduction is commonly used with visualization to provide informative low-dimensional summaries of quantitative multivariate data. Principal component analysis (PCA) (Pearson 1901) is one of the first methods ever developed, and it remains very popular. Visualization of PCA is typically in the form of static scatterplots of a few leading components. When accompanied by a representation of the linear combination of the original variables (magnitude and angles of the variable contributions are inscribed on a unit circle), they are called biplots (Gabriel 1971).

Dynamic visualizations called *tours* (Asimov 1985), animate through a sequence of linear projections (orthonormal bases or frames). Instead of a static view, tours provide a smoothly changing view by interpolating between frames. There are various types of tours distinguished by the way the paths are generated. Asimov originally animated between randomly selected bases in the *grand* tour. The *manual* tour (Cook and Buja 1997) allows for user control over the basis changes. A selected variable (or component) can be rotated into or out of view or to a particular value. The *radial tour* (Spyrison and Cook 2020) is a variant of the manual tour, that fixes the contribution angle and changes the magnitude along the radius. The permanence of the data points from frame to frame holds information between intermediate interpolated frames, and the user-control of the basis could plausibly lead to more information being perceived than a static display. This is a hypothesis that a user study could assess.

Empirical studies have rarely assessed tours. An exception is Nelson, Cook, and Cruz-Neira (1999), who compares scatterplots of grand tours on 2D monitors with 3D (stereoscopic, not head-mounted) over $n = 15$ participants. Participants perform cluster detection, dimensionality estimation, and radial sparseness tasks on six-dimensional data. They find that stereoscopic 3D leads to more accuracy in the cluster identification, though the time to interact with the display was much higher in the 3D environment. In this work, we extend the evaluation of tours which compares the radial tour as benchmarked against the grand tour and discrete pairs of principal components.

The contribution of this paper is an empirical user study comparing the radial tour against PCA and the grand tour for assessing variable attribution on clustered data. This is the first empirical evaluation of the radial or manual tour. We discuss how this fits in with other multivariate data visualization techniques and coordinated views of linear projections.

We are particularly interested in assessing the effectiveness of the new radial tour relative to common practice with PCA and grand tour. The user influence over a basis, uniquely available in the radial tour, is crucial to testing variable sensitivity to the structure visible in projection. If the contribution of a variable is reduced and the feature disappears, then we say that the variable was sensitive to that structure. For example, Figure 1 shows two frames of simulated data. Panel (a) has identified separation between the two clusters. The contributions in panel (b) show no such cluster separation. The former has a large contribution of V2 in the direction of separation, while it is negligible in the right frame. Because of this, we say that V2 is sensitive to the separation of the clusters.

Knowing which variables to use is also important for statistical modeling and their interpretations. Models are becoming increasingly complex, and the nonlinear interactions of the terms cause opaqueness to the model's interpretability. Exploratory Artificial Intelligence (XAI, Adadi and Berrada 2018; Arrieta et al. 2020) is an emerging field that extends the interpretability of such black-box models. Multivariate data visualization is essential for exploring features spaces and communicating interpretations of models (Biecek 2018; Biecek and Burzykowski 2021; Wickham, Cook, and Hofmann 2015).

The paper is structured as follows. Section II provides background on common visualization methods and linear dimension reduction techniques. Section III describes the experimental factors, task, and accuracy measure used. The results of the study are discussed in Section IV. Conclusions and potential future directions are discussed in Section VI. More results, participant demographics, and analysis of the response time are available in the appendix, Section IX.

II. Related work

Consider the data to be a matrix of n observations (rows) and p variables (columns), denoted as $X_{n\times p}$.

A. Orthogonal multivariate visualization

Grinstein, Trutschl, and Cvek (2002) illustrate many multivariate visualization methods. In particular, this work shows examples of actual visuals. Liu et al. (2017) give a good classification and taxonomy of such methods. The content below focuses on the most common visuals that use the full data space before discussing linear combinations of those variables in projections.

1) Scatterplot matrix: One could consider looking at p histograms or univariate densities. Doing so will miss features in two or more dimensions. A scatterplot matrix (Chambers et al. 1983) is a $p \times p$ matrix typically with univariate densities on the diagonal and all combinations of pairs of variables in off-diagonal elements. Figure 2 shows a scatterplot matrix of the four principal components of simulated data. Such displays do not scale well with dimensions because each plot would get less and less space. Scatterplot matrices can only display information in two orthogonal dimensions, so features in three dimensions may not be fully resolved.

2) Parallel coordinates plot: Another common way to display multivariate data is with a parallel coordinates plot (Ocagne 1885), which shows observations by quantile or normalized values for each variable connected by lines to the quantile value in subsequent variables. Parallel coordinates plots scale well with dimensions but poorly with observations as the lines overcrowd the display.

Parallel coordinate plots are asymmetric across variable ordering, in that shuffling the order of the variable can lead to different conclusions. Another shortcoming is the graphical channel used to convey information. Munzner (2014) suggests that position is the visual channel that is most perceptible to humans. In the case of parallel coordinates plots, the horizontal axes span variables rather than the values of one variable, causing the loss of a display dimension to be used by our most perceptible visual channel.

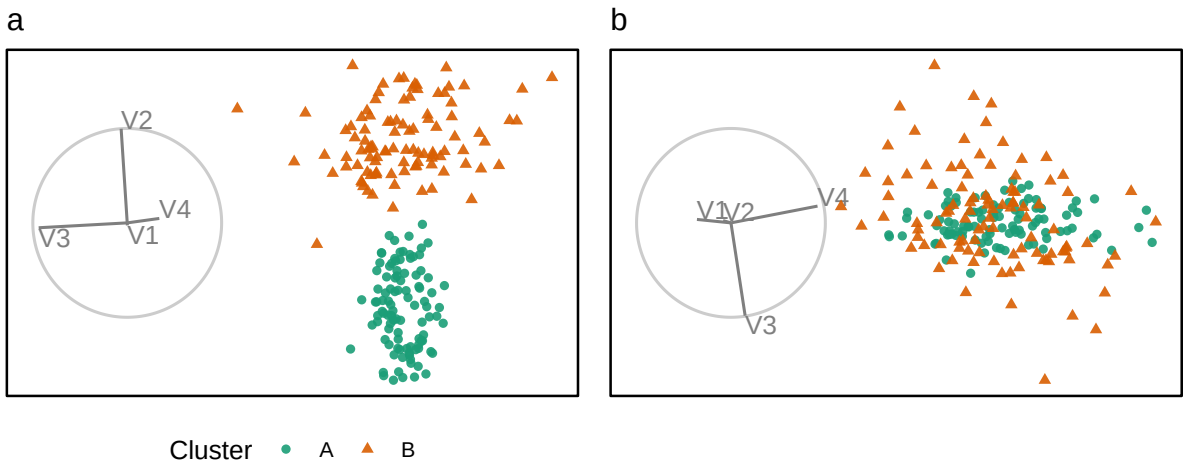

Fig. 1. Illustration of cluster separation, in relation to variable importance. Panel (a) is a projection mostly of V2 and V3, and the separation between clusters is in the direction of V2, not V3. This suggests V2 is important for clustering, but V3 isn't. Panel (b) shows a projection of mostly V3 and V4, with no contribution from V2, and little from V3. That there is no separation between the clusters indicates that V3 and V4 are not important.

B. Multivariate projections

At some point, visualization will be forced to turn to dimension reduction to scale better with the dimensionality of the data. Below we introduce linear projections and the common principal component analysis. Then we touch on nonlinear projections and exclude them from consideration.

1) Linear: Linear projections map a higher p-dimensional space onto a smaller d-space with an affine mapping (where parallel lines stay parallel). A projection is the resulting space of the data multiplied by a basis $Y_{n\times d} = X_{n\times p} \cdot A_{p\times d}$, this is essentially a reorientation of the original variable. The intuition of this is conveyed by thinking of a shadow as 2D projection of a 3D object. Rotating the the object reflects a change in the basis and which side/variable of the object are exposed and cast a shadow.

2) Principal component analysis: PCA is a good baseline of comparison for linear projections because of its frequent and broad use across disciplines. PCA (Pearson 1901) defines new components, linear combinations of the original variables, ordered by decreasing variation through the help of eigenvalue matrix decomposition. While the resulting dimensionality is the same size, the benefit comes from the ordered nature of the components. The data can be said to be approximated by the first several components. The exact number is subjectively selected given the variance contained in each component, typically guided from a scree plot (Cattell 1966). Features with sizable signal regularly appear in the leading components that commonly approximate data. However, this is not always the case, and component spaces should be fully explored to look for signal in components with less variation. This is especially true for cluster structure (Donnell, Buja, and Stuetzle 1994).

3) Nonlinear: Nonlinear transformations bend and distort spaces are not entirely accurate or faithful to the original variable space. There are various quality metrics, such as Trustworthiness, Continuity, Normalized stress, and Average local error, have been introduced to describe the distortion of the space (Espadoto et al. 2021; Gracia et al. 2016). Unfortunately these metrics are hard to visualize and communicate make the distortions introduced opaque to the analyst. The intuition of this can be shown with map projections. Snyder (1987) lists over 200 different projections that distort the surface of the earth to display as a 2D map, each with unique properties and use cases.

Because of the difficultly to interpreting the distortions of nonlinear spaces and the added subjectivity of hyperparemeter selection, we exclude nonlinear techniques and instead decide to compare three linear techniques.

C. Tours, animated linear projections

A *tour* animates through many linear projections. One of the insightful features of the tour is the object permanence of the data points; one can track the relative changes of observations as the basis changes, as opposed to discretely jumping to an orthogonal view

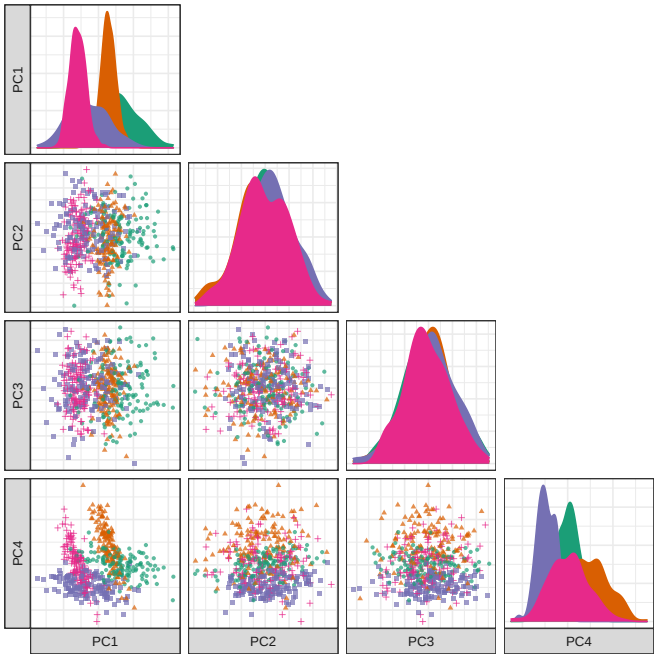

Fig. 2. Scatterplot matrix of the first four principal components of 6D simulated data containing four classes. The separation between classes is primarily in PC1 and PC4. This is not uncommon because PCA is summarizing variance, not cluster structure.

angle with no intermediate information. Types of tours are distinguished by the generation of their basis paths (Lee et al. 2021; Cook et al. 2008). To contrast with the discrete orientations of PCA, we compare continuous linear projection changes with grand and radial tours.

1) Grand tours: Target bases are selected randomly in a grand tour (Asimov 1985). These target bases are then geodesically interpolated for a smooth, continuous path. The grand tour is the first and most widely known tour. The random selection of target bases makes it a general unguided exploratory tool. The grand tour will make a good comparison that has continuity of data points similar to the radial tour but lacks the user control enjoyed by PCA and radial tours.

2) Manual and radial tours: Whether an analyst uses PCA or the grand tour, they cannot influence the basis. They cannot explore the structure identified or change the contribution of the variables. User-controlled steering is a key aspect of *manual* tours that helps to test variable attribution.

The manual tour (Cook and Buja 1997) defines its basis path by manipulating the basis contribution of a selected variable. A manipulation dimension is appended onto the projection plane, giving a full contribution to the selected variable. The target bases are then chosen to rotate this newly created manipulation space. This manipulation space is similarly orthogonally restrained. The data is projected through its interpolated frames and rendered into an animation. When the contribution of one variable changes, the contributions of the other variables must also change, maintaining the basis's orthonormality. A key feature of the manual tour is that it allows users to control the variable contributions to the basis. Such manipulations can be queued in advance or selected in real-time for human-in-the-loop analysis (Karwowski 2006). Manual navigation is relatively time-consuming due to the vast volume of resulting view space and the abstract method of steering the projection basis. First, it is advisable to identify a basis of particular interest and then use the manual tour as a more directed, local exploration tool to explore the sensitivity of a variable's contribution to the feature of interest.

To simplify the task and keep its duration realistic, we consider a variant of the manual tour called a *radial* tour. In a radial tour, the magnitude of along the radius with a fixed angle of contribution to the frame; it must move along the direction of its original contribution radius. The radial tour benefits from both continuity of the data alongside grand tours and user-steering via choosing the variable to rotate.

Manual tours have been recently made available in the **R** package **spinifex** (Spyrison and Cook 2020), which facilitates manual tours (and radial variant). It also provides an interface for a layered composition of tours and exporting to `gif` and `mp4` with **gganimate** (Pedersen and Robinson 2020) or `html` widget with **plotly** (Sievert 2020). It is also compatible with tours made by **tourr** (Wickham et al. 2011). Now that we have a readily available means to produce various tours, we want to see how they fare against traditional discrete displays commonly used with PCA.

D. Other animated linear projections

The work of Elmqvist, Dragicevic, and Fekete (2008) allows users to interactively change the face of a local display by navigating to adjacent faces on a global overview scatterplot matrix. This offers analysts a way to geometrically explore the transition between adjacent faces of a scatterplot matrix as though rotating the face of dice at right angles. The interpolated frames between the orthogonal faces display linear combinations of three variables at varying degrees. This is what McDonald (1982) called a *little tour* with the addition of user control. It is a particular type of manual tour where only horizontal or vertical rotation is allowed.

Star Coordinates (Kandogan 2000) also arrive at the biplot scatterplot displays starting from the perspective of radial parallel coordinates. Lehmann and Theisel (2013) extend this idea, mapping it back to orthogonal projections. They provide a means to interpolate through PCA components, the orthogonal contributions of scatterplot matrix, and the grand tour. It also defines user-controlled interaction similar to manual and radial tours.

TripAdvisor (Nam and Mueller 2012) is an interactive application that plans sequential interpolation between distant target frames. It also provides an additional global context of a subset of possible frames with glyph representation and an overview of variable attribution by summarizing the top ten principal components. It allows for use steering by using a "touchpad polygon." This touchpad allows the magnitude of the contributions to be changed, similar to an incremental change with the manual tour.

E. Empirical evaluation

FIT IN AFTER RELATED WORK RESTURCTURE Gracia et al. (2016) conducted an $n = 40$ user study comparing PCA reduced spaces as 2D and 3D scatterplots on traditional 2D monitors. Participants perform point classification, distance perception, and outlier identification tasks. The results are mixed and mostly small differences. There is some evidence to suggest a lower error in distance perception from 3D scatterplot. Wagner Filho et al. (2018) performed an $n = 30$ within participants using a scatterplot display between 2D, 3D displays on monitors, and 3D displays on a head-mounted device on PCA reduced spaces. None of the tasks on any dataset lead to a significant difference in accuracy. However, the immersive display reduced the effort of navigation, which led to a perception of improved accuracy and engagement. Sedlmair, Munzner, and Tory (2013) instead use two expert coders to evaluate 75 datasets and four dimension reduction techniques across the displays of 2D scatterplots, interactive 3D scatterplots, and 2D scatterplot matrices. They suggest a tiered guidance approach finding that 2D scatterplots are often sufficient to resolve a feature. If not, try 2D scatterplots on a different dimension reduction technique before going to scatterplot matrix display or concluding a true negative. They find that interactive 3D scatterplots help in very few cases.

Some studies compare visualizations across complete contributions of variables. Chang, Dwyer, and Marriott (2018) conducted an $n = 51$ participant study comparing parallel coordinate plots and scatterplot matrix either in isolation, sequentially, or as a coordinated view. Accuracy, completion time, and eye focus were measured for six tasks. Three tasks were more accurate with scatterplot matrix and three with parallel coordinates, while the coordinated view was usually marginally more accurate than the max of the separate visuals. Cao et al. (2018) compare nonstandardized line-glyph and star-glyphs with standardized variants (with and without fill under the curve). Each of the $n = 18$ participants performed 72 trials across the six visuals, two levels of dimensions, and two levels of observations. Visuals with variable standardization outperformed the nonstandardized variants and the radial star-glyph reportedly outperformed line-variant.

Other studies have investigated the relative benefits of projecting to 2- or 3D scatterplots in PCA-reduced spaces. Gracia et al. (2016) conducted an $n = 40$ user study comparing 2- and 3D scatterplots on traditional 2D monitors. Participants perform point classification, distance perception, and outlier identification tasks. The results are mixed and primarily have small differences. There is some evidence to suggest a lower error in distance perception from a 3D scatterplot. Wagner Filho et al. (2018) performed an $n = 30$ within-participants study on PCA reduced space using scatterplot displays between 2D on monitors, 3D on monitors, and 3D display with a head-mounted display. None of the tasks on any dataset lead to a significant difference in accuracy. However, the immersive display reduced effort and navigation, resulting in higher perceived accuracy and engagement.

Some studies use cohort encoding. Sedlmair, Munzner, and Tory (2013) instead use two expert coders to evaluate 75 datasets and four dimension reduction techniques across 2D scatterplots, 2D scatterplot matrices, and interactive 3D scatterplots. They suggest a tiered guidance approach finding that 2D scatterplots are often sufficient to resolve a feature. If not, try an alternative dimension reduction technique before going to scatterplot matrix display, or lastly, concluding a true negative. They find that interactive 3D scatterplots help in relatively rare cases. Lewis, Van der Maaten, and Sa (2012) compare

across three cohorts: experts, uninformed novices, and informed novices ($n = 5 + 15 + 16 = 36$). Participants were asked their opinion of the quality of nine different embedding for each of several data sets. Expert opinion is reportedly more consistent than the novice groups, though this is confounded by a difference in sample sizes. Interestingly, cohort responses correlated with different quality metrics. Positive ratings from the expert group correlated strongest with the Trustworthiness metric.

F. Conclusion

Orthogonal visualizations either scale poorly with dimensionality or introduce an asymmetry of the variable ordering. Projections visualize the full p-data as fewer dimensions, typically 1-3 at a time. In linear projections the resulting dimension are composed of a linear combination of the orginal variables that maintains variable interpretability. While Nonlinear techniques distort and bend space in different ways which is hard to visualize and communicate.

Tours are linear projections that are animated over changes in the basis. There are several more recent other methods that animate or facilitate manual change the basis that were related back to tour techniques. Below we conduct a user study to compare the radial tour with PCA and the grand tour on a variable attribution task of clustered data.

III. User study

The experiment was designed to assess the performance of the radial tour relative to the grand tour and PCA for interpreting the variable attribution to the separation between two clusters. Data were simulated across three experimental factors: location of the cluster separation, cluster shape, and data dimensionality. Participant responses were collected using a web application and crowdsourced through prolific.co, (Palan and Schitter 2018) an alternative to MTurk.

A. Objective

PCA will be used as a baseline for comparison as it is the most commonly used linear embedding. It will use static, discrete jumps between orthogonal components. The grand tour will act as a secondary control that will help evaluate the benefit of observation trackability between nearby animation frames but without user-control of its path. Lastly, the radial tour will be compared, which benefits from the continuity of animation and user control of the basis.

Then for some subset of tasks, we expect to find that the radial tour performs most accurately. Conversely, we are less sure about the accuracy of such limited grand tours as there is no objective function in selecting the bases; it is possible that the random selection of the target bases altogether avoids bases showing cluster separation. However, given that the data dimensionality is modest, and it is probable that the grand tour coincidentally regularly crossed bases with the correct information for the task.

Experimental factors and the definition of an accuracy measure are given below. The null hypothesis can be stated as:

H_0 : accuracy does not change across the visual methods

H_α : accuracy does change across the visual methods

B. Visual factors

The visual methods are tested within-participants, with each visual being evaluated twice by each participant. The order in which experimental factors are experienced is randomized with the assignment, as illustrated in Figure 5. Below discusses the design standardization and unique input associated with each visual.

The visualization methods were standardized wherever possible. Data were displayed as 2D scatterplots with biplots. All aesthetic values (color-blind safe colors, shapes, sizes, absence of legend, and axis titles) were constant. The variable contribution biplot was always shown left of the scatterplot embeddings with their aesthetic values consistent. What did vary between visuals were their inputs.

PCA allowed users to select between the top four principal components for each axes regardless of the data dimensionality (four or six). Upon changing an axis, the visual would change to the new view of orthogonal components without displaying intermediate bases. There was no user input for the grand tour; users were instead shown a 15-second animation of the same randomly selected path (variables containing cluster separation were shuffled after simulation). Participants could view the same clip up to four times within the time limit. Radial tours allowed participants to select the manipulation variable. The starting basis was initialized to a half-clock design, where the variables were evenly distributed in half of the circle. This design was created to be variable agnostic while maximizing the independence of the variables. Selecting a new variable resets the animation where the new variable is manipulated to a complete contribution, zeroed contribution,

and then back to its initial contribution. Animation and interpolation parameters were held constant across grand and radial tour (five frames per second with a step size of 0.1 radians between interpolated frames).

C. Experimental factors

In addition to the visual method, data are simulated across three experimental factors. First, the *location* of the separation between clusters is controlled by mixing a signal and a noise variable at different ratios. Secondly, the *shape* of the clusters reflects varying distributions of the data. And third, the *dimension*-ality of the data is also tested. The levels within each factor are described below, and Figure 3 gives a visual representation.

The *location* of the separation between the clusters is at the heart of the measure. It would be good to test a few varying levels. To test the sensitivity, a noise and signal variable are mixed at different ratios. The separation between clusters are mixed at the following percentages: 0/100% (not mixed), 33/66%, 50/50% (evenly mixed).

In selecting the *shape* of the clusters, the convention given by Scrucca et al. (2016) is followed. They describe 14 variants of model families containing three clusters. The model family name is the abbreviation of the clusters' respective volume, shape, and orientation. The levels are either *E*qual or *V*ary. The models EEE, EEV, and EVV are used. For instance, in the EEV model, the volume and shape of clusters are constant, while the shape's orientation varies. The EVV model is modified by moving four-fifths of the data out in a "$>$" or banana-like shape.

Dimension-ality is tested at two modest levels: four dimensions containing three clusters and six with four clusters. Such modest dimensionality is required to limit the difficulty and search space to make the task realistic for crowdsourcing.

D. Task and evaluation

With our hypothesis formulated, let us turn our attention to the task and how it is evaluated. Participants were asked to "check any/all variables that contribute more than average to the cluster separation green circles and orange triangles." This was further explained in the explanatory video as "mark any and all variable that carries more than their fair share of the weight, or one quarter in the case of four variables."

The instructions iterated several times in the video were: 1) use the input controls to find a frame that contains separation between the clusters of green circles and orange triangles, 2) look at the orientation of the variable contributions in the grey circle (biplot axes orientation), and 3) select all variables that contribute more than uniformed distributed cluster separation in the scatterplot. Independent with experimental level, participants were limited to 60 seconds for each evaluation of this task. This restriction did not impact many participants as the 25th, 50th, 75th quantiles of the response time were about 7, 21, and 30 seconds, respectively.

The accuracy measure of this task was designed with a couple of features in mind. 1) symmetric about the expected value, that is, without preference to under- or over-guessing. 2) heavier than linear weight with an increasing difference from the expected value. The following measure is defined for evaluating the task.

Let the data $\mathbf{X}_{n,\ p,\ k}$ be a simulation containing clusters of observations of different distributions. Where n is the number of observations, p is the number of variables, and k indicates the observation's cluster. Cluster membership is exclusive; an observation cannot belong to more than one cluster.

The weights, w, is a vector, the variable-wise difference between the mean of two clusters of less $1/p$, the expected cluster separation if it were uniformly distributed. Accuracy, A is defined as the signed square of these weights if selected by the participant. Participant responses are a logical value for each variable — whether or not the participant thinks each variable separates the two clusters more than uniformly distributed separation.

$$w_j = \frac{(\overline{X}_{\cdot,j=1,k=1} - \overline{X}_{\cdot,1,2},\ \ldots\ (\overline{X}_{\cdot,p,1} - \overline{X}_{\cdot,p,2})}{\sum_{j=1}^{p}(|\overline{X}_{\cdot,j,k=1} - \overline{X}_{\cdot,j,2}|)} - \frac{1}{p}$$

$$A = \sum_{j=1}^{p} I(j) \cdot sign(w_j) \cdot w^2$$

Where $I(j)$ is the indicator function, the binary response for variable j. Figure 4 shows one projection of a simulation with its observed variable separation (wide bars), expected uniform separation (dashed line), and accuracy if selected (thin vertical lines).

E. Data simulation

Each dimension is distributed initially as $\mathcal{N}(0, 1)$, given the covariance set by the shape factor. Clusters were initially separated by a distance of two before location mixing. Signal variables had a correlation of 0.9 when they had equal orientation and -0.9 when their orientations varied. Noise variables were restricted to zero correlation. Each cluster is simulated with 140

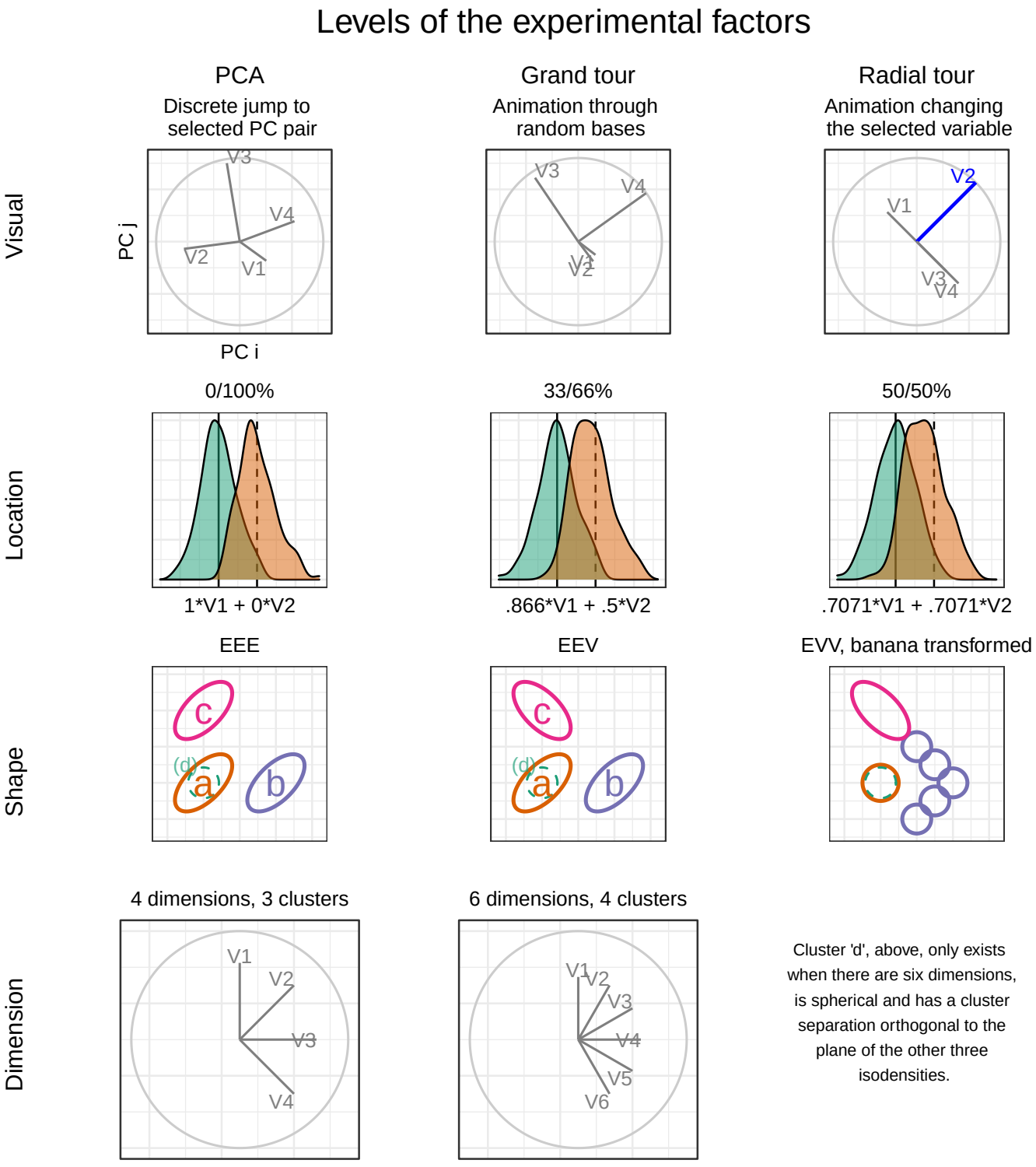

Fig. 3. Levels of the visuals and three experimental factors: location of cluster separation, the shape of clusters, and dimensionality of the sampled data.

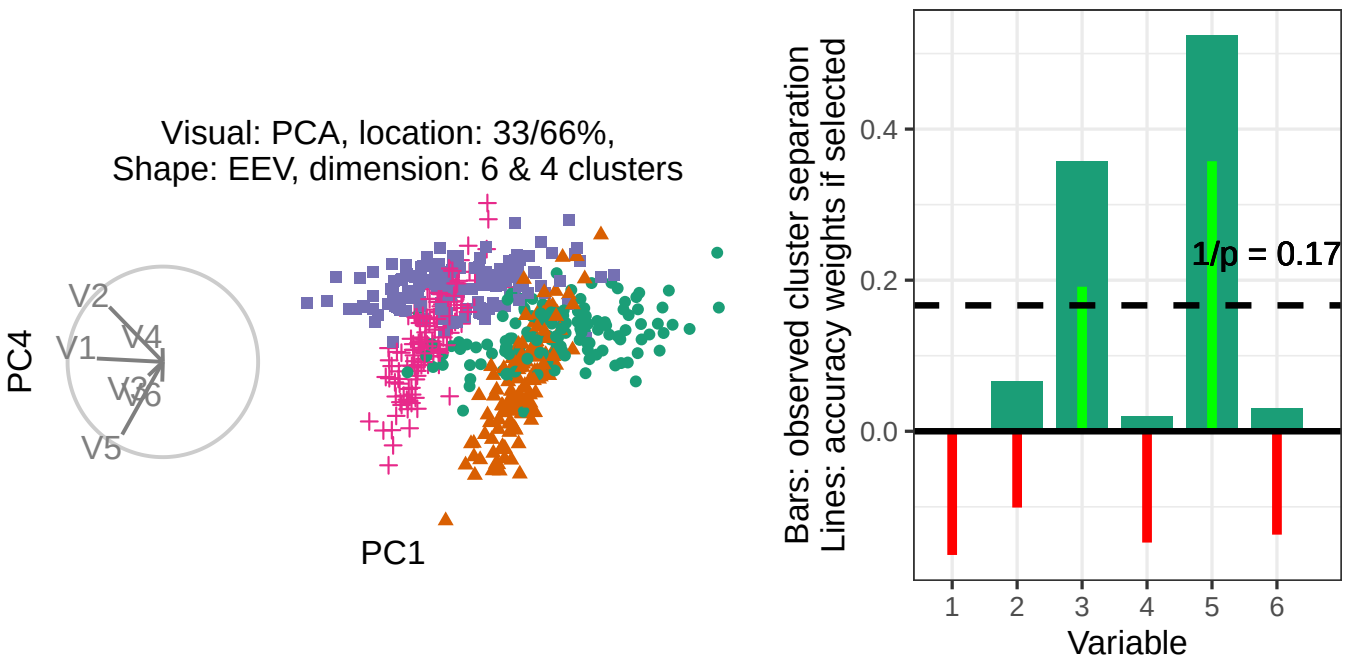

Fig. 4. Illustration of how accuracy is measured. (L), Scatterplot and biplot of PC1 by PC4 of a simulated data set (R) illustrate cluster separation between the green circles and orange triangles. Bars indicate observed cluster separation, and (red/green) lines show the accuracy weight of the variable if selected. The horizontal dashed line has a height $1/p$, the expected value of cluster separation. The accuracy weights equal the signed square of the difference between each variable value and the dashed line.

observations and is offset in a variable that did not distinguish previous variables.

Clusters of the EVV shape are transformed into the banana-chevron shape (illustrated in Figure 3, shape row). Then location mixing is applied by post-multiplying a rotation matrix to the signal variable and a noise variable for the clusters in question. All variables are then standardized by standard deviations away from the mean. The columns are then shuffled randomly.

Each of these replications is then iterated with each level of the visual. For PCA, projections were saved (to `png`) for the 12 pairs of the top four principal components. A grand tour basis path is saved for each dimensionality level. The data from each simulation is then projected through its corresponding bases path and saved to `gif` file. The radial tour starts at either the four or six-variable "half-clock" basis. A radial tour is then produced for each variable and saved as a `gif`.

F. Randomized factor assignment

Now, with simulation and their artifacts in hand, this section covers how the experimental factors are assigned and demonstrate how this is experienced from the participant's perspective.

The study is sectioned into three periods. Each period is linked to a randomized level of visual and location. The order of dimension and shape are of secondary interest and are held constant in increasing order of difficulty; four then six dimensions and EEE, EEV, then EVV-banana, respectively.

Each period starts with an untimed training task at the simplest remaining experimental levels; location = 0/100%, shape = EEE, and four dimensions with three clusters. This serves to introduce and familiarize participants with input and visual differences. After the training, the participant performs two trials with the same visual and location level across the increasing difficulty of dimension and shape. The plot was removed after 60 seconds, though participants rarely reached this limit.

The order of the visual and location levels is randomized with a nested Latin square where all levels of the visuals are exhausted before advancing to the next level of location. This requires $3!^2 = 36$ participants to evaluate all permutations of the experimental factors once. This randomization controls for potential learning effects the participant may receive. Figure 5 illustrates how an arbitrary participant experiences the experimental factors.

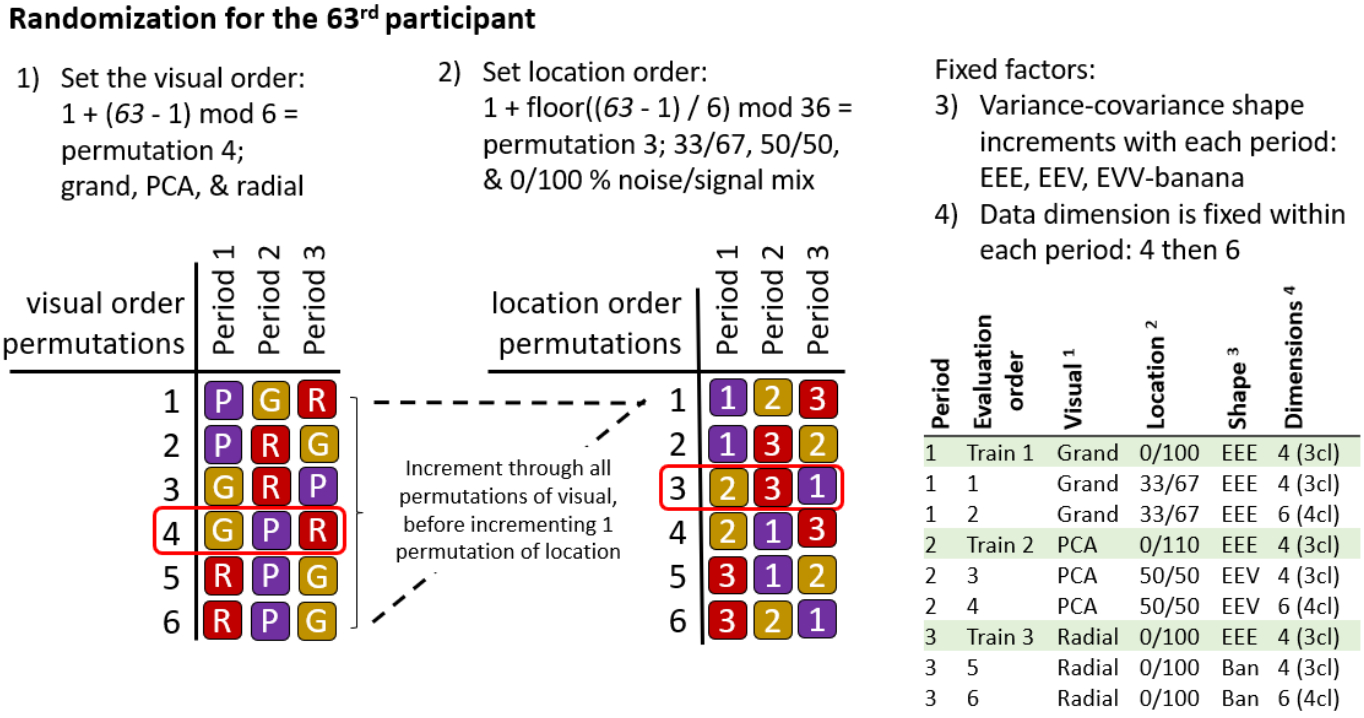

Fig. 5. Illustration of how a hypothetical participant 63 is assigned experimental factors. Each of the six visual order permutations is exhausted before iterating to the next permutation of location order.

Through pilot studies sampled by convenience (information technology and statistics Ph.D. students attending Monash University), it was estimated that three complete evaluations are needed to power the study properly, a total of $N = 3 \times 3!^2 = 108$ participants.

G. Participants

$N = 108$ participants were recruited via prolific.co (Palan and Schitter 2018). Participants are restricted based on their claimed education requiring that they have completed at least an undergraduate degree (some 58,700 of the 150,400 users at the time). This restriction is used on the premise that linear projections and biplot displays will not be regularly used for consumption by general audiences. There is also the implicit filter that Prolific participants must be at least 18 years of age and implicit biases of timezone, location, and language. Participants were compensated for their time at £7.50 per hour, whereas the mean duration of the survey was about 16 minutes. Previous knowledge or familiarity was minimal, as validated in the follow-up survey. The appendix, Section IX-B, contains a heatmap distribution of age and education paneled across preferred pronouns of the participants that completed the survey, who are relatively young, well educated, and slightly more likely to identify as males.

H. Data collection

Data were recorded in a **shiny** application and written to a Google Sheet after each third of the study. Especially at the start of the study, participants experienced adverse network conditions due to the volume of participants hitting the application with modest allocated resources. In addition to this, API read/write limitations further hindered data collection. To mitigate this, the number of participants was throttled, and over-collect survey trials until three evaluations were collected for all permutation levels.

The processing steps were minimal. The data were formatted and then filtered to the latest three complete studies of each experimental factor, which should have experienced the least adverse network conditions. The bulk of the studies removed were partial data and a few over-sampled permutations. This brings us to the 108 studies described in the paper, from which models and aggregation tables were built. The post-study surveys were similarly decoded to human-readable format and then filtered to include only those 84 associated with the final 108 studies.

The code, response files, their analyses, and the study application are publicly available at https://github.com/nspyrison/spinifex_study.

IV. Results

To recap, the primary response variable is accuracy, as defined in Section III-D. Two primary data sets were collected; the user study evaluations and the post-study survey. The former is the 108 participants with the experimental factors: visual, location of the cluster separation signal, the shape of variance-covariance matrix, and the dimensionality of the data. Experimental factors and randomization were discussed in section III-C. A follow-up survey was completed by 84 of these 108 people. It collected demographic information (preferred pronoun, age, and education), and subjective measures for each visual (preference, familiarity, ease of use, and confidence).

Below a battery of mixed regression models is built to examine the degree of the evidence and the size of the effects from the experimental factors. Then, Likert plots and rank-sum tests to compare the subjective measures between the visuals.

A. Accuracy

To quantify the contribution of the experimental factors to the accuracy, mixed-effects models were fit. All models have a random effect term on the participant and the simulation. These terms explain the amount of error that can be attributed to the individual participant's effect and variation due to the random sampling data.

In building a set of models to test, a base model with only the visual term is compared with the full linear model term and progressively interacting with an additional experimental factor. The models with three and four interacting variables are rank deficient; there is not enough varying information in the data to explain all interacting terms.

Fixed effects	**Full model**
α	$\widehat{Y} = \mu + \alpha_i + \mathbf{Z} + \mathbf{W} + \epsilon$
$\alpha + \beta + \gamma + \delta$	$\widehat{Y} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + \mathbf{Z} + \mathbf{W} + \epsilon$
$\alpha \times \beta + \gamma + \delta$	$\widehat{Y} = \mu + \alpha_i \times \beta_j + \gamma_k + \delta_l + \mathbf{Z} + \mathbf{W} + \epsilon$
$\alpha \times \beta \times \gamma + \delta$	$\widehat{Y} = \mu + \alpha_i \times \beta_j \times \gamma_k + \delta_l + \mathbf{Z} + \mathbf{W} + \epsilon$
$\alpha \times \beta \times \gamma \times \delta$	$\widehat{Y} = \mu + \alpha_i \times \beta_j \times \gamma_k \times \delta_l + \mathbf{Z} + \mathbf{W} + \epsilon$

where
μ, the intercept of the model
α_i, visual term | $i \in$ (pca, grand, radial)
β_j, location term | $j \in$ (0/100%, 33/66%, 50/50%) mix
γ_k, shape term | $k \in$ (EEE, EEV, EVV banana)
δ_l, dimension term | $l \in$ (4 & 3, 6 & 4) var & clust,
$\mathbf{Z} \sim \mathcal{N}(0,\ \tau)$, error of the random effect of participant
$\mathbf{W} \sim \mathcal{N}(0,\ \upsilon)$, error of the random effect of simulation
$\epsilon \sim \mathcal{N}(0,\ \sigma)$, the remaining error of the model

Table I compares the model summaries across increasing complexity. The $\alpha \times \beta + \gamma + \delta$ model is selected to examine in more detail as it has relatively high condition R^2 and not overly complex interacting terms. Table II looks at the coefficients for this model. There is strong evidence suggesting a relatively large increase in accuracy from the radial tour, though there is evidence that almost of increase is lost under 33/66% mixing.

We also want to visually examine the conditional variables in the model. Figure 6 illustrates the violin plots of accuracy for each of the model terms.

B. Subjective measures

The 84 evaluations of the post-study survey also collect four subjective measures for each visual. Responses were 5-point Likert items on confidence, ease of use, familiarity, and preference with each of the visuals. Figure 7 shows the Likert plots, or stacked percentage bar plots, alongside violin plots with the same non-parametric, ranked sum tests previously used.

There was strong evidence to support that participants preferred the radial tour to either alternative. There is less evidence that the radial tour led to more confidence and found easier to use than the grand tour. In confirmation of expectations, crowd sourced participants had low familiarity with all visuals, with no difference in mean supported.

V. Discussion

TODO

VI. Conclusion

Data visualization is an integral part of understanding relationships in data and how models are fitted. However, thorough exploration of data in high dimensions becomes difficult. Previous methods offer no means for an analyst to impact the projection basis. The manual tour provides a mechanism for changing the contribution of a selected variable to the basis. Giving analysts such control should facilitate the exploration of variable-level sensitivity to the identified structure.

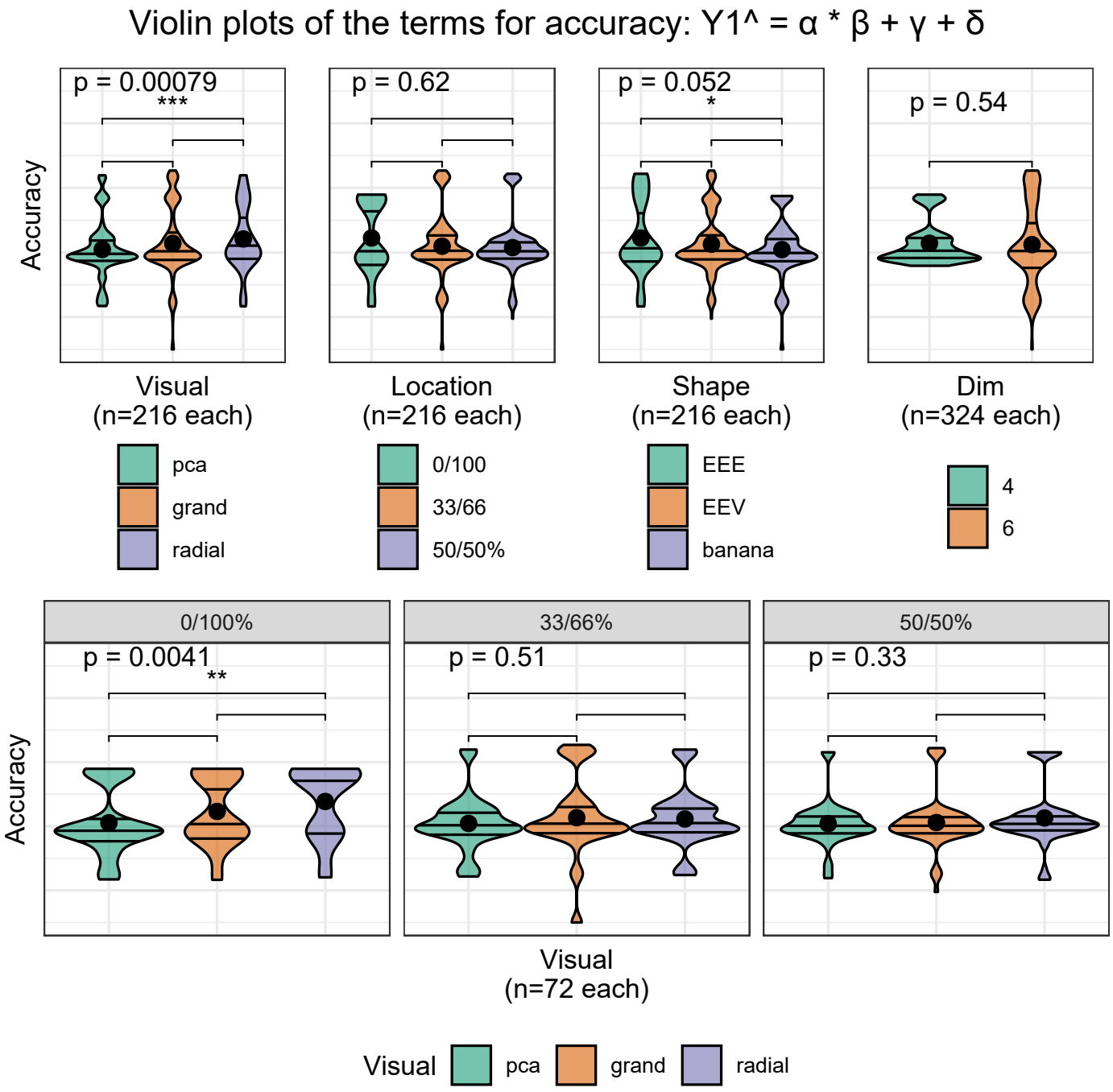

Fig. 6. Violin plots of terms of the model $\widehat{Y} = \alpha \times \beta + \gamma + \delta$. Overlaid with global significance from the Kruskal-Wallis test and pairwise significance from the Wilcoxon test, both are non-parametric, ranked-sum tests. Viewing the marginal accuracy of the terms corroborates the primary findings that the use of the radial tour leads to a significant increase in accuracy, at least over PCA, and this effect is particularly well supported when no location mixing is applied.

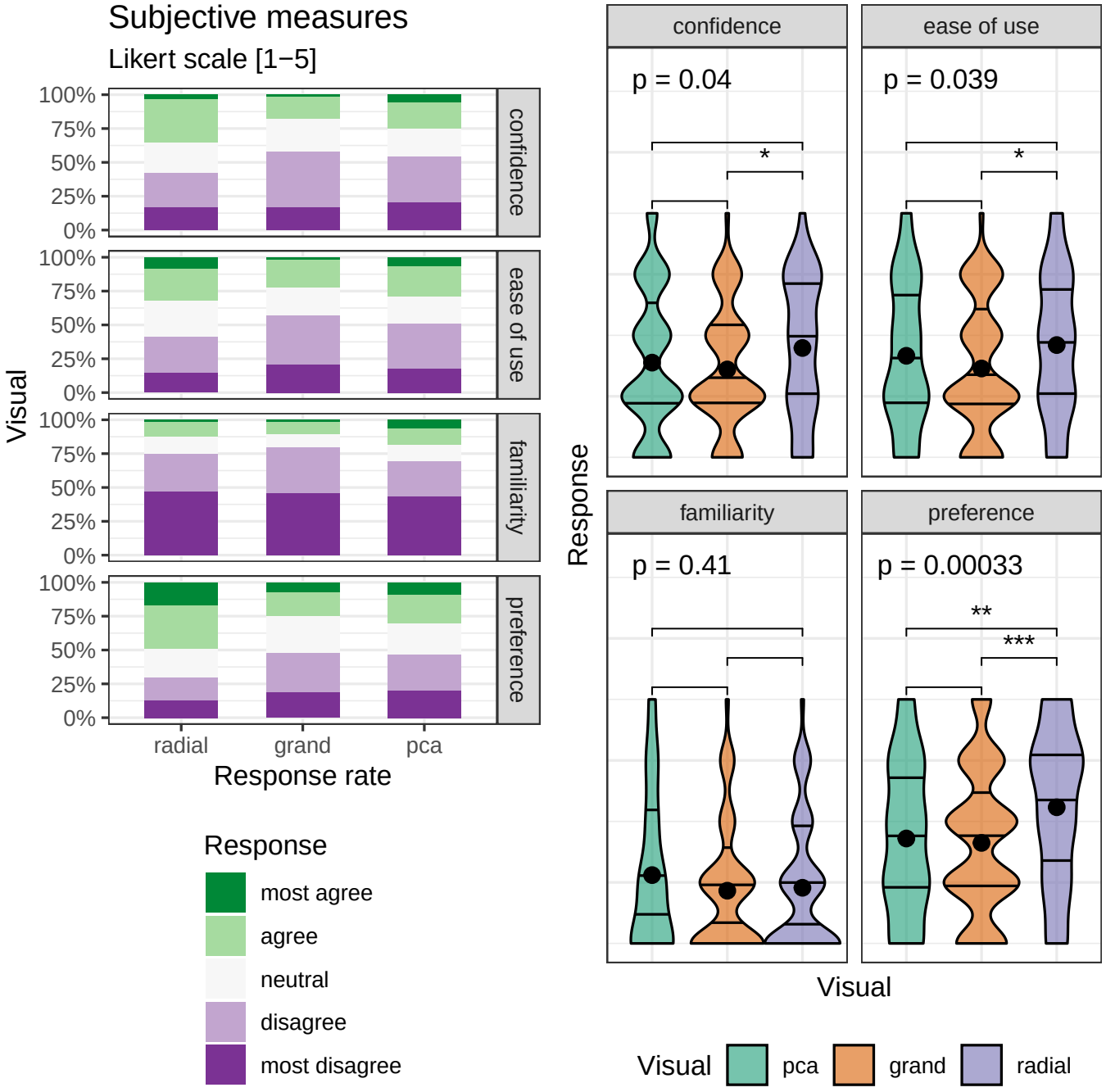

Fig. 7. The subjective measures of the 84 responses of the post-study survey with five-point Likert items levels of agreement. (L) Likert plots (stacked percent bar plots) with (R) violin plots of the same measures. Violin plots are overlaid with global significance from the Kruskal-Wallis test and pairwise significance from the Wilcoxon test. Participants are more confident using the radial tour and find it easier to use than the grand tour. The radial tour is the most preferred visual.

TABLE I

MODEL PERFORMANCE OF RANDOM EFFECT MODELS REGRESSING ACCURACY. COMPLEX MODELS PERFORM BETTER IN TERMS OF R^2 AND RMSE, YET AIC AND BIC PENALIZE THEIR LARGE NUMBER OF FIXED EFFECTS IN FAVOR OF THE MUCH SIMPLER MODEL CONTAINING ONLY THE VISUALS. CONDITIONAL R^2 INCLUDES ERROR EXPLAINED BY THE RANDOM EFFECTS, WHILE MARGINAL DOES NOT.

Fixed effects	No. levels	No. terms	AIC	BIC	R2 cond.	R2 marg.	RMSE
a	1	3	**-71**	**-44**	0.303	0.018	0.194
a+b+c+d	4	8	-45	4	0.334	0.056	0.194
a*b+c+d	5	12	-26	41	0.338	0.064	0.193
a*b*c+d	8	28	28	167	**0.383**	0.108	0.19
a*b*c*d	15	54	105	360	0.37	**0.222**	**0.185**

TABLE II

THE TASK ACCURACY MODEL COEFFICIENTS FOR $\widehat{Y} = \alpha \times \beta + \gamma + \delta$, WITH VISUAL = PCA, LOCATION = 0/100%, SHAPE = EEE, AND DIM = 4 HELD AS BASELINES. VISUAL BEING RADIAL IS THE FIXED TERM WITH THE STRONGEST EVIDENCE SUPPORTING THE HYPOTHESIS. INTERACTING WITH THE LOCATION TERM, THERE IS EVIDENCE SUGGESTING RADIAL PERFORMS WITH MINIMAL IMPROVEMENT FOR 33/66% LOCATION MIXING.

	Estimate	Std. Error	df	t value	Prob	
(Intercept)	0.10	0.06	16.1	1.54	0.143	
Factor						
Visualgrand	0.06	0.04	622.1	1.63	0.104	
Visualradial	0.14	0.04	617.0	3.77	0.000	***
Fixed effects						
Location33/66%	-0.02	0.07	19.9	-0.29	0.777	
Location50/50%	-0.04	0.07	20.0	-0.66	0.514	
ShapeEEV	-0.05	0.06	11.8	-0.82	0.427	
Shapebanana	-0.09	0.06	11.8	-1.54	0.150	
Dim6	-0.01	0.05	11.8	-0.23	0.824	
Interactions						
Visualgrand:Location33/66%	-0.02	0.06	588.9	-0.29	0.774	
Visualradial:Location33/66%	-0.12	0.06	586.5	-2.13	0.033	*
Visualgrand:Location50/50%	-0.03	0.06	591.6	-0.47	0.641	
Visualradial:Location50/50%	-0.06	0.06	576.3	-1.16	0.248	

This paper discussed a within-participant user study ($n = 108$) comparing the efficacy of three linear projection techniques: PCA, grand tour, and radial tour. The participants performed a supervised cluster task, explicitly identifying which variables contribute to the separation of two target clusters. This was evaluated evenly over four experimental factors. In summary, mixed model regression finds strong evidence that using the radial tour sizably increases accuracy, especially when cluster separation location is not mixed at 33/66%. The effect sizes on accuracy are large relative to the change from the other experimental factors and the random effect of data simulation, though smaller than the random effect of the participant. The radial tour was the most preferred of the three visuals.

There are several ways that this study could be extended. In addition to expanding the support of the experimental factors, more exciting directions include: introducing a new task, visualizations used, and experience level of the target population. It is difficult to achieve good coverage given the number of possible experimental factors.

VII. ACKNOWLEDGMENTS

This research was supported by an Australian Government Research Training Program (RTP) scholarship.

This article was created in **R** (R Core Team 2020) and **rmarkdown** (Xie, Allaire, and Grolemund 2018). Visuals were prepared with **spinifex**. All packages used are available from the Comprehensive R Archive Network CRAN at https://CRAN.R-project.org/. The source files for this article, application, data, and analysis can be found on GitHub, https://github.com/nspyrison/spinifex_study/. The source code for the **spinifex** package and accompanying application can be found at https://github.com/nspyrison/spinifex/. We thank Jieyang Chong for his help in proofreading this article.

VIII. References

Adadi, Amina, and Mohammed Berrada. 2018. "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)." *IEEE Access* 6: 52138–60.

Anscombe, F. J. 1973. "Graphs in Statistical Analysis." *The American Statistician* 27 (1): 17–21. https://doi.org/10.2307/2682899.

Arrieta, Alejandro Barredo, Natalia Díaz-Rodriguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-López, Daniel Molina, and Richard Benjamins. 2020. "Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges Toward Responsible AI." *Information Fusion* 58: 82–115.

Asimov, Daniel. 1985. "The Grand Tour: A Tool for Viewing Multidimensional Data." *SIAM Journal on Scientific and Statistical Computing* 6 (1): 128–43. https://doi.org/https://doi.org/10.1137/0906011.

Biecek, Przemyslaw. 2018. "DALEX: Explainers for Complex Predictive Models in R." *The Journal of Machine Learning Research* 19 (1): 3245–49.

Biecek, Przemyslaw, and Tomasz Burzykowski. 2021. *Explanatory Model Analysis: Explore, Explain, and Examine Predictive Models*. CRC Press.

Cao, Nan, Yu-Ru Lin, David Gotz, and Fan Du. 2018. "Z-Glyph: Visualizing Outliers in Multivariate Data." *Information Visualization* 17 (1): 22–40. https://doi.org/10.1177/1473871616686635.

Cattell, Raymond B. 1966. "The Scree Test for the Number of Factors." *Multivariate Behavioral Research* 1 (2): 245–76.

Chambers, J. M., W. S. Cleveland, B. Kleiner, and P. A. Tukey. 1983. "Graphical Methods for Data Analysis."

Chang, Chunlei, Tim Dwyer, and Kim Marriott. 2018. "An Evaluation of Perceptually Complementary Views for Multivariate Data." In *2018 IEEE Pacific Visualization Symposium (PacificVis)*, 195–204. IEEE.

Cook, Dianne, and Andreas Buja. 1997. "Manual Controls for High-Dimensional Data Projections." *Journal of Computational and Graphical Statistics* 6 (4): 464–80. https://doi.org/10.2307/1390747.

Cook, Dianne, Andreas Buja, Eun-Kyung Lee, and Hadley Wickham. 2008. "Grand Tours, Projection Pursuit Guided Tours, and Manual Controls." In *Handbook of Data Visualization*, 295–314. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-33037-0_13.

Donnell, Deborah J., Andreas Buja, and Werner Stuetzle.

1994. “Analysis of Additive Dependencies and Concurvities Using Smallest Additive Principal Components.” *The Annals of Statistics* 22 (4): 1635–68. https://doi.org/10.1214/aos/1176325746.

Elmqvist, Niklas, Pierre Dragicevic, and Jean-Daniel Fekete. 2008. “Rolling the Dice: Multidimensional Visual Exploration Using Scatterplot Matrix Navigation.” *IEEE Transactions on Visualization and Computer Graphics* 14 (6): 1539–1148.

Espadoto, Mateus, Rafael M. Martins, Andreas Kerren, Nina S. T. Hirata, and Alexandru C. Telea. 2021. “Toward a Quantitative Survey of Dimension Reduction Techniques.” *IEEE Transactions on Visualization and Computer Graphics* 27 (3): 2153–73. https://doi.org/10.1109/TVCG.2019.2944182.

Gabriel, Karl Ruben. 1971. “The Biplot Graphic Display of Matrices with Application to Principal Component Analysis.” *Biometrika* 58 (3): 453–67.

Gracia, Antonio, Santiago González, Víctor Robles, Ernestina Menasalvas, and Tatiana von Landesberger. 2016. “New Insights into the Suitability of the Third Dimension for Visualizing Multivariate/Multidimensional Data: A Study Based on Loss of Quality Quantification.” *Information Visualization* 15 (1): 3–30. https://doi.org/10.1177/1473871614556393.

Grinstein, Georges, Marjan Trutschl, and Urska Cvek. 2002. “High-Dimensional Visualizations,” 14.

Kandogan, Eser. 2000. “Star Coordinates: A Multi-Dimensional Visualization Technique with Uniform Treatment of Dimensions.” In *In Proceedings of the IEEE Information Visualization Symposium, Late Breaking Hot Topics*, 9–12.

Karwowski, Waldemar. 2006. *International Encyclopedia of Ergonomics and Human Factors, -3 Volume Set*. CRC Press.

Lee, Stuart, Dianne Cook, Natalia da Silva, Ursula Laa, Nicholas Spyrison, Earo Wang, and H. Sherry Zhang. 2021. “The State-of-the-Art on Tours for Dynamic Visualization of High-Dimensional Data.” *WIREs Computational Statistics* n/a (n/a): e1573. https://doi.org/10.1002/wics.1573.

Lehmann, Dirk J., and Holger Theisel. 2013. “Orthographic Star Coordinates.” *IEEE Transactions on Visualization and Computer Graphics* 19 (12): 2615–24.

Lewis, Joshua, Laurens Van der Maaten, and Virginia de Sa. 2012. “A Behavioral Investigation of Dimensionality Reduction.” In *Proceedings of the Annual Meeting of the Cognitive Science Society*. Vol. 34.

Liu, Shusen, Dan Maljovec, Bei Wang, Peer-Timo Bremer, and Valerio Pascucci. 2017. “Visualizing High-Dimensional Data: Advances in the Past Decade.” *IEEE Transactions on Visualization and Computer Graphics* 23 (3): 1249–68. https://doi.org/10.1109/TVCG.2016.2640960.

Matejka, Justin, and George Fitzmaurice. 2017. “Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics Through Simulated Annealing.” In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI ’17*, 1290–94. Denver, Colorado, USA: ACM Press. https://doi.org/10.1145/3025453.3025912.

McDonald, John A. 1982. “Interactive Graphics for Data Analysis.” https://lib-extopc.kek.jp/preprints/PDF/1984/8410/8410042.pdf.

Munzner, Tamara. 2014. “Visualization Analysis and Design.”

Nam, Julia EunJu, and Klaus Mueller. 2012. “Tripadvisorˆ${$ND$}$: A Tourism-Inspired High-Dimensional Space Exploration Framework with Overview and Detail.” *IEEE Transactions on Visualization and Computer Graphics* 19 (2): 291–305.

Nelson, Laura, Dianne Cook, and Carolina Cruz-Neira. 1999. “XGobi Vs the C2: Results of an Experiment Comparing Data Visualization in a 3-d Immersive Virtual Reality Environment with a 2-d Workstation Display.” *Computational Statistics* 14 (1): 39–51.

Ocagne, Maurice d’. 1885. *Coordonne’es Paralle’les Et Axiales. Me’thode de Transformation Ge’ome’trique Et Proce’de’ Nouveau de Calcul Graphique de’duits de La Conside’ration Des Coordonne’es Paralle’les*. Paris: Gauthier-Villars.

Palan, Stefan, and Christian Schitter. 2018. “Prolific: A Subject Pool for Online Experiments.” *Journal of Behavioral and Experimental Finance* 17: 22–27.

Pearson, Karl. 1901. “LIII. On Lines and Planes of Closest Fit to Systems of Points in Space.” *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2 (11): 559–72.

Pedersen, Thomas Lin, and David Robinson. 2020. *Gganimate: A Grammar of Animated Graphics*. https://CRAN.R-project.org/package=gganimate.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Scrucca, Luca, Michael Fop, T. Brendan Murphy, and Adrian E. Raftery. 2016. “Mclust 5: Clustering,

Classification and Density Estimation Using Gaussian Finite Mixture Models." *The R Journal* 8 (1): 289–317. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5096736/.

Sedlmair, Michael, Tamara Munzner, and Melanie Tory. 2013. "Empirical Guidance on Scatterplot and Dimension Reduction Technique Choices." *IEEE Transactions on Visualization & Computer Graphics*, no. 12: 2634–43.

Sievert, Carson. 2020. *Interactive Web-Based Data Visualization with R, Plotly, and Shiny*. Chapman; Hall/CRC. https://plotly-r.com.

Snyder, John Parr. 1987. *Map Projections–A Working Manual*. Vol. 1395. US Government Printing Office.

Spyrison, Nicholas, and Dianne Cook. 2020. "Spinifex: An R Package for Creating a Manual Tour of Low-Dimensional Projections of Multivariate Data." *The R Journal* 12 (1): 243. https://doi.org/10.32614/RJ-2020-027.

Tukey, John W. 1977. *Exploratory Data Analysis*. Vol. 32. Pearson.

Wagner Filho, Jorge, Marina Rey, Carla Freitas, and Luciana Nedel. 2018. "Immersive Visualization of Abstract Information: An Evaluation on Dimensionally-Reduced Data Scatterplots." In.

Wickham, Hadley, Dianne Cook, and Heike Hofmann. 2015. "Visualizing Statistical Models: Removing the Blindfold." *Statistical Analysis and Data Mining: The ASA Data Science Journal* 8 (4): 203–25. https://doi.org/10.1002/sam.11271.

Wickham, Hadley, Dianne Cook, Heike Hofmann, and Andreas Buja. 2011. "Tourr: An R Package for Exploring Multivariate Data with Projections." *Journal of Statistical Software* 40 (2). https://doi.org/10.18637/jss.v040.i02.

Xie, Yihui, J. J. Allaire, and Garrett Grolemund. 2018. *R Markdown: The Definitive Guide*. Boca Raton, Florida: Chapman; Hall/CRC. https://bookdown.org/yihui/rmarkdown.

Yanai, Itai, and Martin Lercher. 2020. "A Hypothesis Is a Liability." *Genome Biology* 21 (1): 231. https://doi.org/10.1186/s13059-020-02133-w.

IX. Appendix

This section covers extended analysis. First, it illustrations of the different visuals are provided. Then, the participant demographics are covered. Lastly, a parallel modeling analysis on log response time is conducted.

A. Illustrations of visual methods

Below illustrates the three visual methods evaluated in the user study. Data was collected from a **shiny** application, and pre-rendered `gif` files were displayed based on the selected inputs.

B. Survey participant demographics

The target population is relatively well-educated people, as linear projections may prove difficult for generalized consumption. Hence, Prolific.co participants are restricted to those with an undergraduate degree (58,700 of the 150,400 users at the study time). From this cohort, 108 performed a complete study. Of these participants, 84 submitted the post-study survey, represented in the following heatmap. All participants were compensated for their time at £7.50 per hour, with a mean time of about 16 minutes. Figure 9 shows a heat map of the demographics for these 84 participants.

C. Response time

As a secondary explanatory variable, response time is considered. Response time is first log-transformed to remove its right skew. The same modeling procedure is repeated for this response. 1) Compare the performance of progressively more complex models. Table III shows the higher level performance of these models over increasing model complexity. 2) Select the model with the same effect terms, $\alpha \times \beta + \gamma + \delta$, with relatively high conditional R^2 without becoming overly complex from variable interactions. The coefficients of this model are displayed in Table IV.

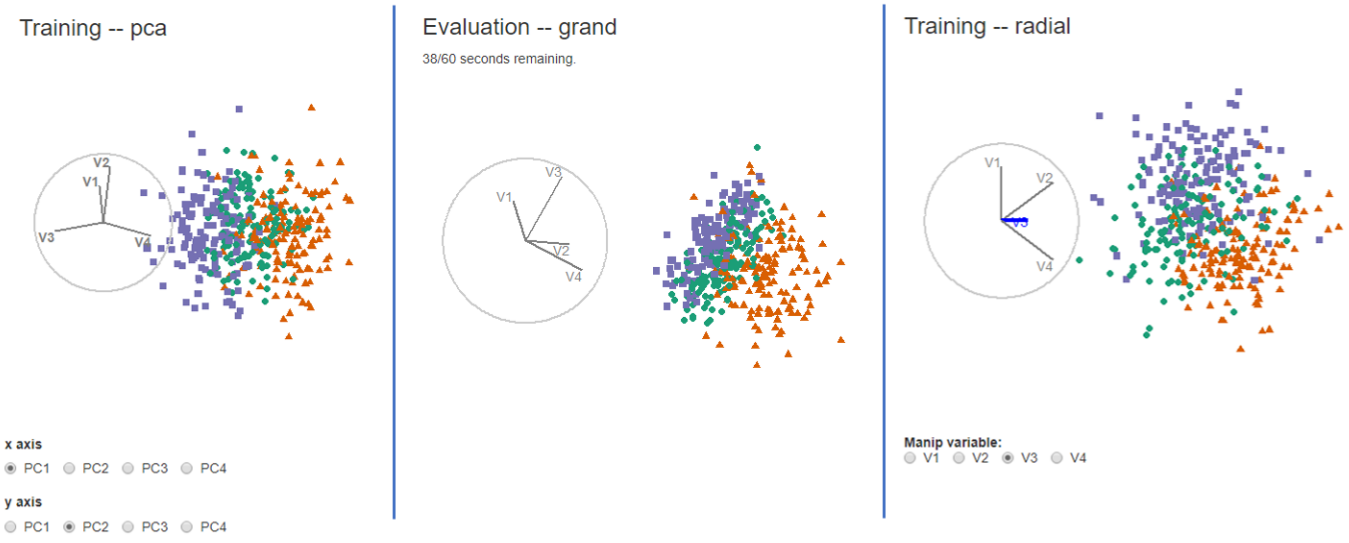

Fig. 8. Examples of the application displays for PCA, grand tour, and radial tour.

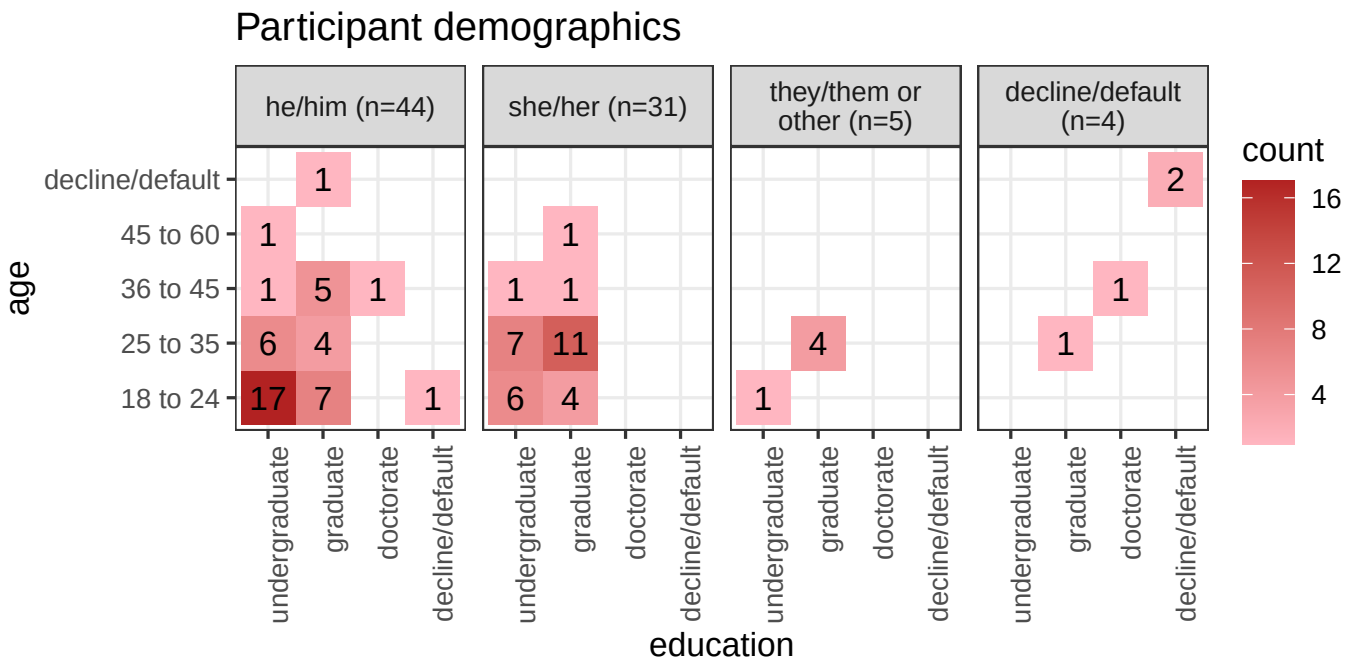

Fig. 9. Heatmaps of survey participant demographics; counts of age group by completed education as faceted across preferred pronoun. Our sample tended to be between 18 and 35 years of age with an undergraduate or graduate degree.

TABLE III

MODEL PERFORMANCE REGRESSING ON LOG RESPONSE TIME [SECONDS], $\widehat{Y_2}$ RANDOM EFFECT MODELS. CONDITIONAL R^2 INCLUDES THE RANDOM EFFECTS, WHILE MARGINAL DOES NOT. THE MODEL $\alpha \times \beta + \gamma + \delta$ MODEL IS SELECTED TO EXAMINE FURTHER AS IT HAS RELATIVELY HIGH MARGINAL R^2 WHILE HAVING MUCH LESS COMPLEXITY THAN THE COMPLETE INTERACTION MODEL.

Fixed effects	No. levels	No. terms	AIC	BIC	R2 cond.	R2 marg.	RMSE
a	1	3	**1448**	**1475**	0.645	0.007	0.553
a+b+c+d	4	8	1467	1516	0.647	0.017	0.552
a*b+c+d	5	12	1474	1541	0.656	0.024	0.548
a*b*c+d	8	28	1488	1627	0.673	0.054	0.536
a*b*c*d	15	54	1537	1792	**0.7**	**0.062**	**0.523**

TABLE IV

MODEL COEFFICIENTS FOR LOG RESPONSE TIME [SECONDS] $\widehat{Y_2} = \alpha \times \beta + \gamma + \delta$, WITH FACTOR = PCA, LOCATION = 0/100%, SHAPE = EEE, AND DIM = 4 HELD AS BASELINES. LOCATION = 50/50% IS THE FIXED TERM WITH THE STRONGEST EVIDENCE AND TAKES LESS TIME. IN CONTRAST, THE INTERACTION TERM LOCATION = 50/50%:SHAPE = EEV HAS THE MOST EVIDENCE AND TAKES MUCH LONGER ON AVERAGE.

	Estimate	Std. Error	df	t value	Prob	
(Intercept)	2.71	0.14	42.6	19.06	0.000	***
Factor						
Visualgrand	-0.23	0.12	567.6	-1.97	0.049	*
Visualradial	0.16	0.12	573.5	1.34	0.181	
Fixed effects						
Location33/66%	0.05	0.14	40.9	0.34	0.737	
Location50/50%	-0.05	0.14	42.1	-0.35	0.729	
ShapeEEV	-0.15	0.09	8.3	-1.61	0.145	
Shapebanana	-0.13	0.09	8.3	-1.42	0.192	
Dim6	0.14	0.08	8.3	1.90	0.093	
Interactions						
Visualgrand:Location33/66%	0.24	0.18	580.9	1.34	0.181	
Visualradial:Location33/66%	-0.24	0.18	582.4	-1.32	0.188	
Visualgrand:Location50/50%	0.12	0.18	578.6	0.69	0.491	
Visualradial:Location50/50%	0.05	0.18	584.4	0.25	0.800	