

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

true true true

Abstract

Principal component analysis is a long-standing go-to method for exploring multivariate data. Data visualization *tours* are a class of linear projections animated over small changes in the projection basis. The grand tour selects target bases at random to animate between. Alternatively, the manual tour rotates the contribution of a selected variable, offering analysts a unique means to control the basis. 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 designed where participants evaluate variable attribution of the separation between two classes. An accuracy measure is defined as a response variable. Data were collected from 108 crowdsourced participants, who performed two trials with each visual for 648 trials in total. Mixed model regression provides strong evidence that the radial tour increases accuracy over the alternatives. Participants subjectively prefer the use of the radial tour.

1 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), 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 developed, and it remains very popular. Visualization of PCA is typically in the form of static scatterplots of a few leading components. This is usually accompanied by a representation of the basis, where the magnitude and angles of the variable contributions are inscribed on a unit circle, called a biplot (Gabriel 1971).

The class of dynamic linear projections known as *tours* (Asimov 1985), can be used to animate over small changes to the basis. Instead of a static view of two orthogonal components, tours interpolate many intermediate frames. There are various types of tours that are distinguished by the generation of their basis paths. Asimov originally animated between randomly selected bases in the *grand tour*. The *manual tour* (Cook and Buja 1997; Spyrisson and Cook 2020) uniquely allows for user-control over the basis changes. A selected variable (or component) can be rotated to the desired contribution. A *radial tour* 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 and information held in intermediate interpolated frames and user-control of the basis could plausibly lead to more information perception than a static display. This is a hypothesis that a user study could assess.

Gracia et al. (2016) conducted an $n = 40$ user study comparing between 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 scatterplot display between 2D, 3D displays on monitors and 3D display with a head-mounted display on PCA reduced spaces. None of the tasks on any dataset lead to a significant difference in accuracy. However, the immersive display reduced effort and navigation while resulting in higher perceived accuracy and engagement. Sedlmair, Munzner, and Tory (2013) instead uses two expert coders to evaluate 75 datasets and four dimension reduction techniques for 2D scatterplots, interactive 3D scatterplots, and 2D scatterplot matrices. They suggest a tiered guidance approach finding that 2D scatterplots are often ‘good enough.’ 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.

Empirical studies scarcely compare tours. However, Nelson, Cook, and Cruz-Neira (1999) compare scatterplots of grand tours on 2D monitor with 3D (stereoscopic, not head-mounted) over $n = 15$ participants. Participants perform clusters detection, dimensionality and radial sparseness tasks on six dimensional data. They find that stereoscopic 3D leads to more accuracy in the cluster identification, though interaction time was much increased in the 3D case. In this work we extend the evaluation of tours which novelly compares the radial tour as benchmarked against the grand tour and discrete pairs of principal components.

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 to 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 non-linear interactions of the terms cause an opaqueness to model

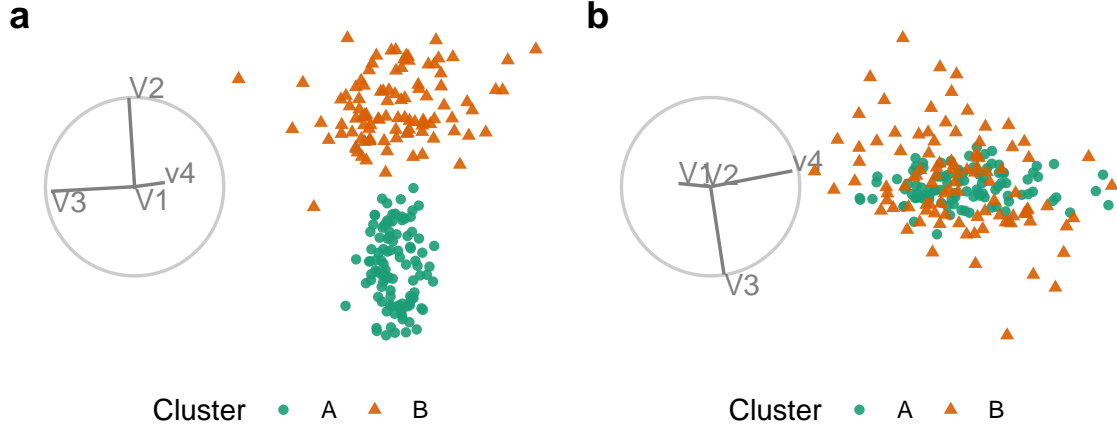


Figure 1: Illustration of cluster separation. Panel (a) shows clear separation in V2 and no separation in the direction of V3. While V1 and v4 have relatively small contributions to the frame. Panel (b) has a random basis with a minimal contribution from V2, and no separation between the cluster means is resolved.

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 2 discusses several visualization methods and orthogonal and observation-based visuals before arriving at the three linear dimension reduction techniques compared in the study. Section 3 describes the experimental factors, task, and accuracy measure used. The results of the study are discussed in Section 4. Conclusions and potential future directions are discussed in Section 5. The software used for the study is described in Section 8.1.

2 Background

Before discussing PCA, the grand tour, and the radial tour, this section covers orthogonal views and observation-based visuals of the full variable space. Consider data to be a complete matrix of n observations across p variables, $X_{n \times p}$.

2.1 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 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 first four components of simulated data. Such displays do not scale well with dimension, quickly becoming dense. Scatterplot matrices also display information in two orthogonal dimensions; features in three dimensions will never be fully resolved.

2.2 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 and other observations-based visuals, such as pixel plots or Chernoff faces scale well with dimensions but poorly with observations. These are perhaps best used when there are more variables than observations.

Observations-based visuals have a couple of issues. They are asymmetric across variable ordering in that a change in the order can lead to different conclusions. Another notable observation-based visual is the graphical channel used to convey information. Munzner suggests that position is the visual channel that is most perceptible to humans (Munzner 2014). In the case of parallel coordinates plots, the horizontal axes span variables rather than the values of one variable; the loss of a display dimension to be used by our most perceptible visual channel.

At some point, visualization will be forced to turn to dimension reduction to scale better with the dimensionality of the data. Non-linear transformations bend and distort spaces and are not entirely accurate or faithful to the original variable space. In light of this, we preclude non-linear techniques and instead decide on PCA, the grand tour, and the radial tour.

2.3 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 by each component, typically guided using 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 that hold less variation.

2.4 Animated linear projections, tours

A data visualization *tour* animates many linear projections over small changes in the projection basis. 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 moves, as opposed to discretely jumping to an orthogonal view 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.

2.4.1 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.

2.4.2 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-control-steering is a key aspect of *manual* tours that should facilitate testing 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 orthonormality of the basis and space. 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

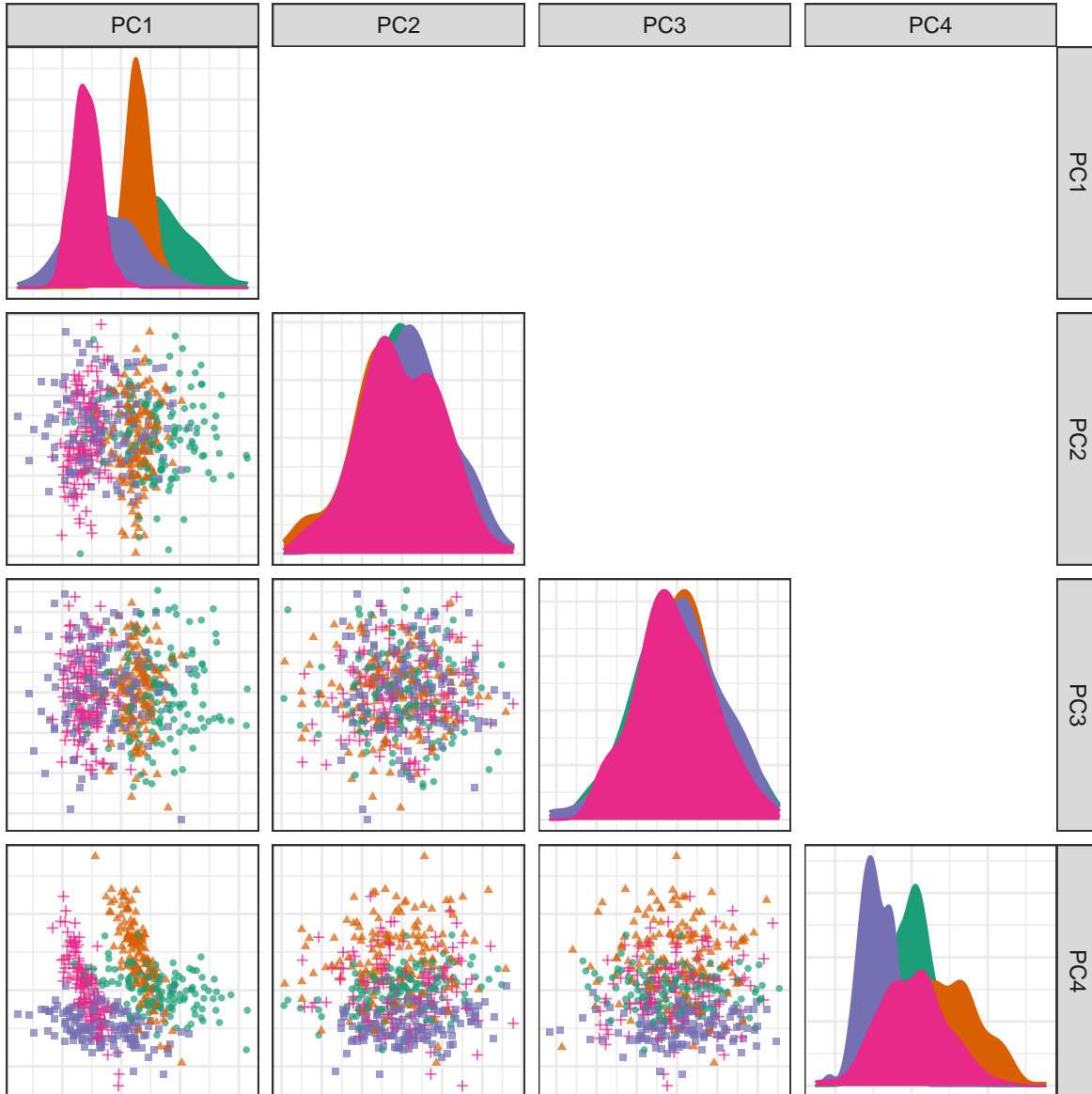


Figure 2: Scatterplot matrix of the first four principal components simulated data in six dimensions. An analyst would have to view both PC1 by PC2 and PC1 by PC4 to have a thorough take on which variables attribute to the separation between clusters.

method of steering the projection basis. It is advisable to first 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.

2.5 Empirical evaluation

Some studies compare visualizations across complete contributions of variables. C. Chang, Dwyer, and Marriott (2018) conduct an $n = 51$ participant study comparing parallel coordinate plots and SPLOM 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 SPLOM and three with parallel coordinates while coordinated view was usually marginally more accurate than the max of the separate visuals. Cao et al. (2018) compare unstandardized line-glyph and star-glyphs with standardized variants (with and without curve fill). Each of the $n = 18$ participant perform 72 trials across the six visuals, two levels of dimensions, and two levels of observations. Visuals with variable standardization outperform the unstandardized 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 an expert or cohort coding. 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 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 seven data sets. Expert opinion is reportedly more consistent than the novice groups, though this is confounded with the different sample sizes. Interestingly, cohort responses correlated with different quality metrics. Positive ratings from the expert group correlated strongest with the Trustworthiness metric.

Tours are absent from studies calculable quality measures. However, Nelson, Cook, and Cruz-Neira (1999) compare scatterplots of grand tours on a 2D monitor with 3D display (stereoscopic, not head-mounted) over $n = 15$ participants. Participants perform clusters detection, dimensionality, and radial sparseness tasks on six-dimensional data. They find that stereoscopic 3D leads to more accuracy for cluster identification, though interaction time greatly increased in the 3D case.

3 User study

An experiment was constructed to assess the performance of the radial tour relative to the grand tour and PCA for interpreting the variable attribution contributing to 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.

3.1 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, it seems plausible 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:

$$H_0 : \text{accuracy does not change across the visual methods}$$

$$H_\alpha : \text{accuracy does change across the visual methods}$$

3.2 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 *dimensionality* 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-containing variable are mixed. 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 *Equal* or *Vary*. 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.

Dimensionality 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.

3.3 Task and evaluation

With our hypothesis formulated and data at hand, let us turn our attention to the task and how to evaluate it. Regardless of the visual method, the display elements are held constant, shown as a 2D scatterplot with a

Levels of the experimental factors

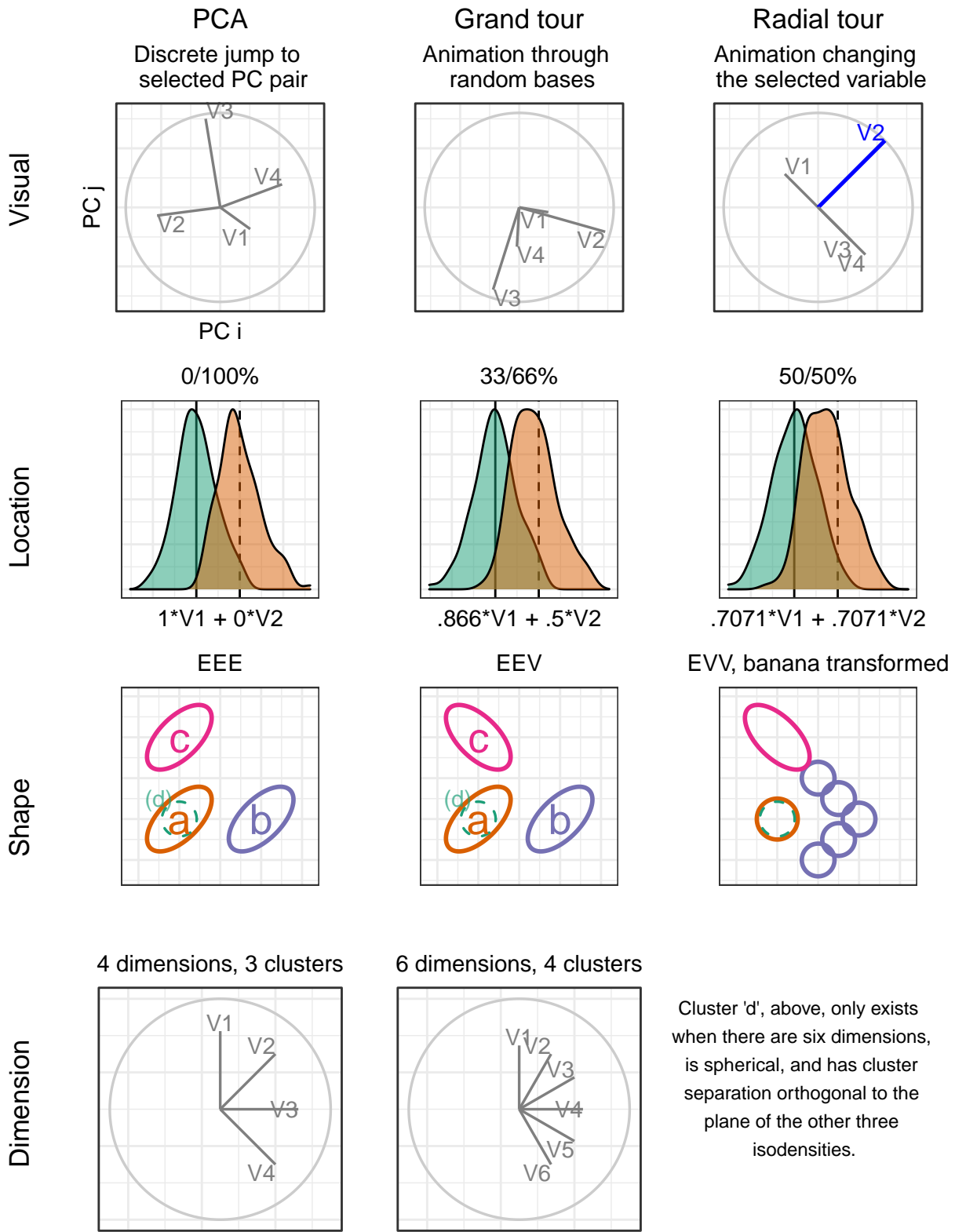


Figure 3: Illustration of the experimental factors, the parameter space of the independent variables.

biplot (Gabriel 1971) to its left. A biplot is a visual depiction of the variable contributions from the basis inscribed in a unit circle. Observations were supervised with cluster membership mapped to (colorblind safe) colors and shape.

Participants were asked to “check any/all variables that contribute more than average to the cluster separation green circles and orange triangles,” which 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 were iterated several times in the video was: 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 gray circle (biplot axes orientation), and 3) select all variables that contribute more than uniform 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.

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

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

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 lines).

3.4 Visual design standardization

The visual methods are tested within-participant, with each visual being evaluated twice by each participant. The order that 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 (colors, shapes, sizes, absence of legend, and axis titles) were constant. Variable contributions were 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

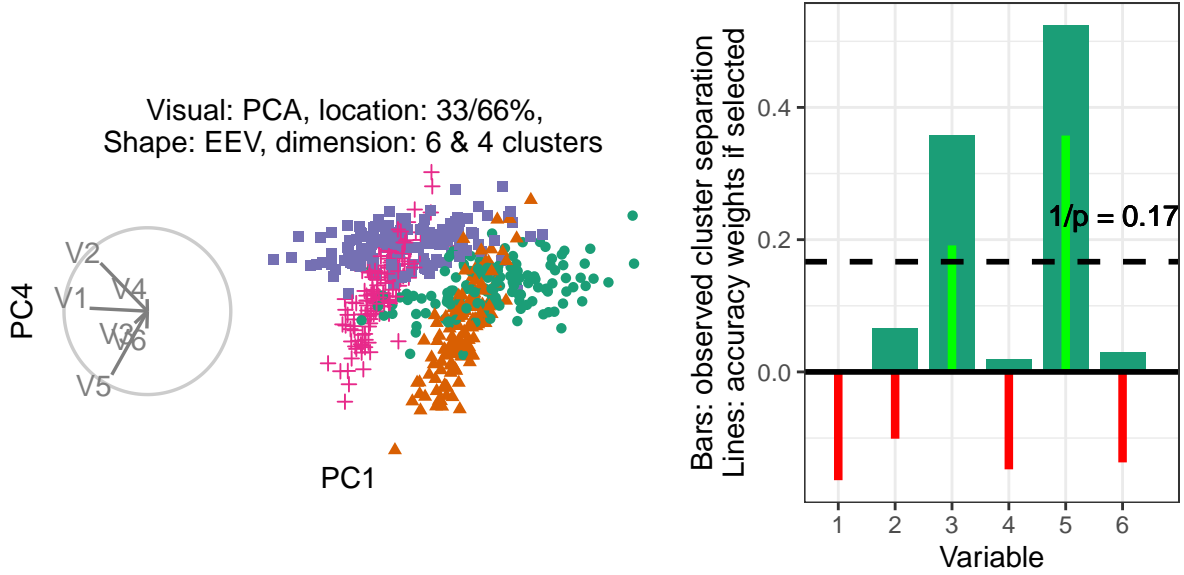


Figure 4: Illustration of how accuracy is measured. (L), Scatterplot and biplot of PC1 by PC4 of a simulated data set (R) illustrates cluster separation between the green circles and orange triangles. Bars indicate observed cluster separation, and (red/green) lines show the accuracy weight of the variable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

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).

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

Clusters of the EVV shape are transformed to 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 each of 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.

3.6 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.

Randomization for the 63rd participant

- 1) Set the visual order:
 $1 + (63 - 1) \bmod 6 =$
 permutation 4;
 grand, PCA, & radial

- 2) Set location order:
 $1 + \text{floor}((63 - 1) / 6) \bmod 36 =$
 permutation 3; 33/67, 50/50,
 & 0/100 % noise/signal mix

Fixed factors:

- 3) Variance-covariance shape increments with each period:
 EEE, EEV, EVV-banana
- 4) Data dimension is fixed within each period: 4 then 6

visual order permutations	Period 1	Period 2	Period 3	location order permutations	Period 1	Period 2	Period 3
1	P	G	R	1	1	2	3
2	P	R	G	2	1	3	2
3	G	R	P	3	2	3	1
4	G	P	R	4	2	1	3
5	R	P	G	5	3	1	2
6	R	P	G	6	3	2	1

Increment through all permutations of visual, before incrementing 1 permutation of location

Period	Evaluation order	Visual ¹	Location ²	Shape ³	Dimensions ⁴
1	Train 1	Grand	0/100	EEE	4 (3cl)
1	1	Grand	33/67	EEE	4 (3cl)
1	2	Grand	33/67	EEE	6 (4cl)
2	Train 2	PCA	0/110	EEE	4 (3cl)
2	3	PCA	50/50	EEV	4 (3cl)
2	4	PCA	50/50	EEV	6 (4cl)
3	Train 3	Radial	0/100	EEE	4 (3cl)
3	5	Radial	0/100	Ban	4 (3cl)
3	6	Radial	0/100	Ban	6 (4cl)

Figure 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 full evaluations are needed to power the study properly, a total of $N = 3 \times 3!^2 = 108$ participants.

3.7 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 ?? 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.

3.8 Data collection

Data were recorded in **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 were 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.

4 Results

To recap, the primary response variable is accuracy, as defined in Section 3.3. The parallel analysis of the log response time is provided in the appendix, Section ?. 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 3.2. A followup 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.

4.1 Accuracy

To quantify the contribution of the experimental factors to the accuracy, a 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 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.

Table 1: 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

Fixed effects Full model

$$\begin{aligned}
\alpha & \hat{Y} = \mu + \alpha_i + \mathbf{Z} + \mathbf{W} + \epsilon \\
\alpha + \beta + \gamma + \delta & \hat{Y} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + \mathbf{Z} + \mathbf{W} + \epsilon \\
\alpha \times \beta + \gamma + \delta & \hat{Y} = \mu + \alpha_i \times \beta_j + \gamma_k + \delta_l + \mathbf{Z} + \mathbf{W} + \epsilon \\
\alpha \times \beta \times \gamma + \delta & \hat{Y} = \mu + \alpha_i \times \beta_j \times \gamma_k + \delta_l + \mathbf{Z} + \mathbf{W} + \epsilon \\
\alpha \times \beta \times \gamma \times \delta & \hat{Y} = \mu + \alpha_i \times \beta_j \times \gamma_k \times \delta_l + \mathbf{Z} + \mathbf{W} + \epsilon
\end{aligned}$$

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

Table 1 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 2 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 examines violin plots of accuracy by visual with panels distinguishing location (vertical) and shape (horizontal).

4.2 Subjective measures

The 84 evaluations of the post-study survey also collect four subjective measures for each visual. Figure 7 shows the Likert plots, or stacked percentage bar plots, alongside violin plots with the same non-parametric, ranked sum tests previously used. Participants preferred to use radial for this task. Participants were also more confident of their answers and found radial tours easier than grand tours. All visuals have reportedly low familiarity, as expected from crowdsourced participants.

5 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.

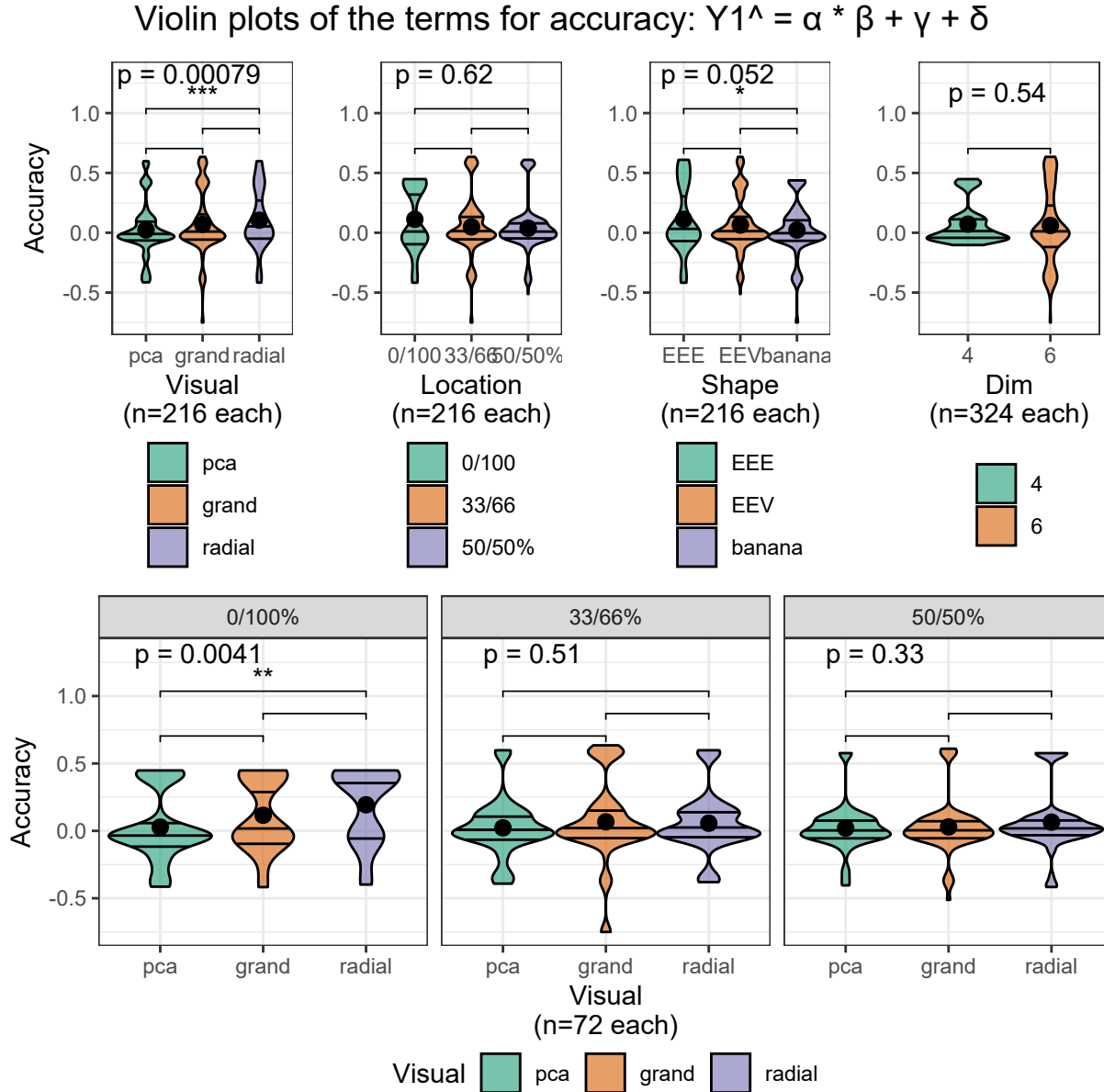


Figure 6: Violin plots of terms of the model $\hat{Y} = \alpha \times \beta + \gamma + \delta$. Overlaid with global significance from the Kruskal-Wallis test and pairwise significance from the Wilcoxon test, both are non-parametric, ranked-sum tests. Viewing marginal accuracy of the terms corroborate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

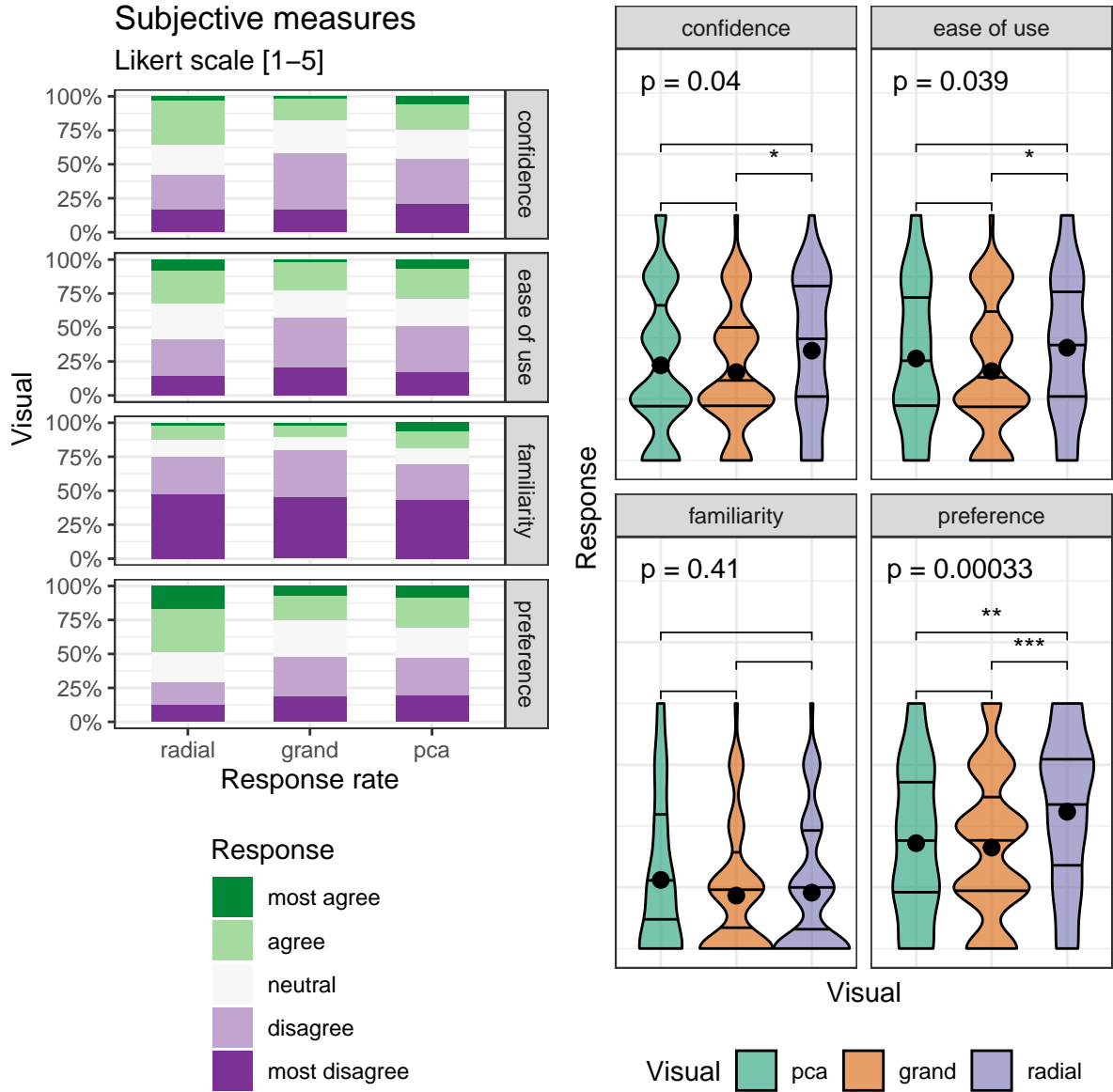


Figure 7: The subjective measures of the 84 responses of the post-study survey, five discrete Likert scale levels of agreement (L) Likert plots (stacked percent bar plots) with (R) violin plots of the same measures. Violin plots are overlaid with global significance from the Kruskal-Wallis test and pairwise significance from the Wilcoxon test. Both are non-parametric, ranked sum tests. 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 2: The task accuracy model coefficients for $\hat{Y} = \alpha \times \beta + \gamma + \delta$, 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	Pr(> t)	
(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 with-in 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 with the location of cluster separation 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 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.

6 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 Golemund 2018). Visuals were prepared with **spinifex**. All packages used are available from the Comprehensive R Archive Network [CRAN](#). The source files for this article, application, data, and analysis can be found on [GitHub](#). The source code for the **spinifex** package and accompanying application can be found [here](#). We thank Jieyang Chong for his help in proofreading this article.

7 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-Rodríguez, 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.
- Chang, Winston, Joe Cheng, JJ Allaire, Carson Sievert, Barret Schloerke, Yihui Xie, Jeff Allen, Jonathan McPherson, Alan Dipert, and Barbara Borges. 2021. *Shiny: Web Application Framework for r*. <https://CRAN.R-project.org/package=shiny>.
- 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.
- 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>.
- 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 Spyrisson, 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>.
- 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.
- 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>.

- Munzner, Tamara. 2014. “Visualization Analysis and Design.”
- 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>.
- 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 Grolemond. 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>.

8 Appendix

This section covers an accompanying graphical user interface for conducting radial tours and extended analysis. First, this analysis explores the demographics of the participants. Then a parallel modeling analysis is performed regressing log response time. Lastly, we look at the effect ranges and marginal effects of the random effect of the participants and data simulation.

8.1 Accompanying radial tour application

An accompanying application illustrates the radial tour. The **R** package, **spinifex**, (Syrison and Cook 2020) is an open-source and now contains a **shiny** (W. Chang et al. 2021) application allowing users to apply various preprocessing tasks and interactively explore their data via interactive radial tour. Example datasets are provided with the ability to upload data. The **html** widget produced is a more interactive variant relative to the one used in the user study. Screen captures and more details are provided in the appendix. Run the following **R** code will run the application locally.

```
install.packages("spinifex", dependencies = TRUE)
spinifex::run_app()
```

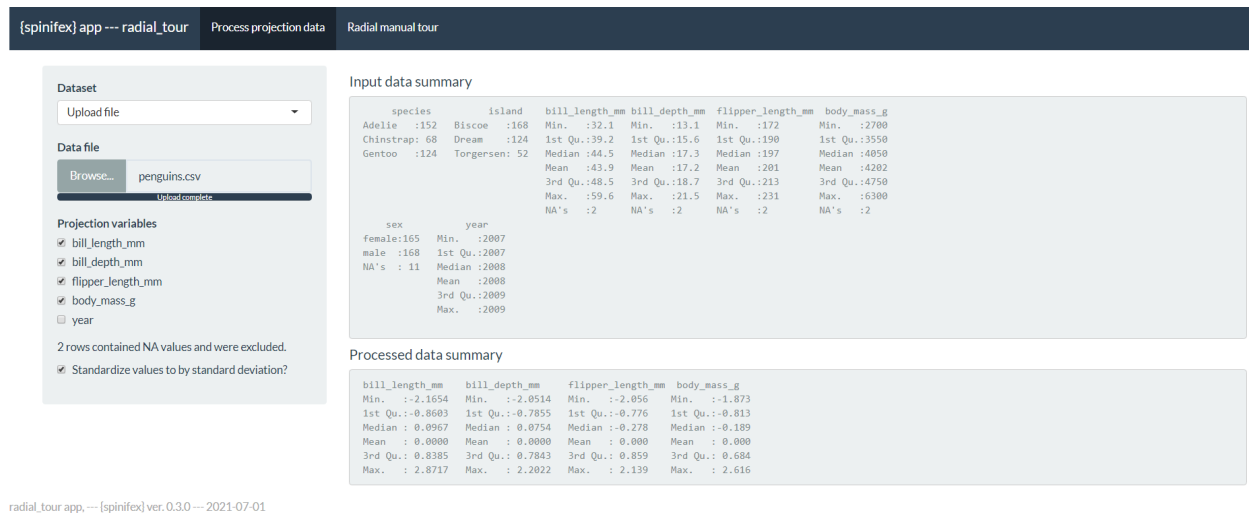


Figure 8: Process data tab, interactively loads or select data, check which variables to project, and optionally scale columns by standard deviation.

In the initial tab, Figure 8, users upload their own (.csv, .rds, or .rda) data or select from predefined data sets. The numeric columns appear as a list of variables to include in the projection. Below that, a line displays whether or not missing rows were removed. Scaling by standard deviation is included by default, as this is a common transformation used to explore linear projections of spaces. Summaries of the raw data and processed numeric data are displayed to illustrate how the data was read and its transformation.

The second tab, Figure 9 contains interaction for selecting the manipulation variable, and non-numeric columns can be used to change the color and shape of the data points in the projection. The radial tour is created in real-time, animated as an interactive **plotly html** widget. The application offers users a fast, intuitive introduction elucidating what the radial tour does and some of the features offered.

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,

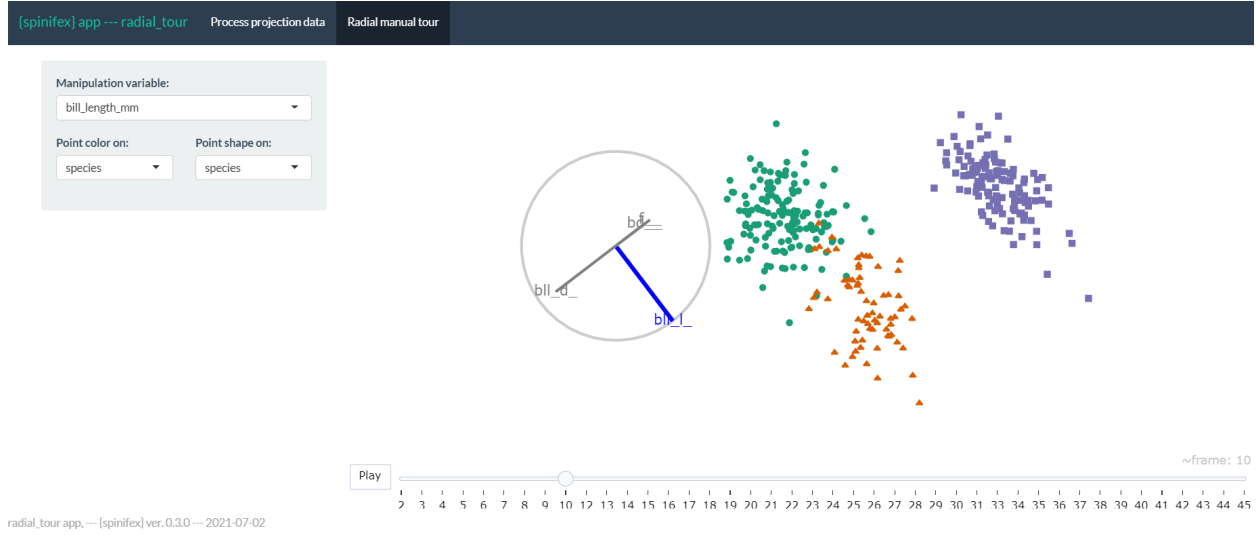


Figure 9: Radial tour tab, interactively create radial tours, changing the manipulation variable, color, or shape of the resulting manual tour. Here, the palmer penguins data is being explored, bill length was selected to manipulate as it is the only variable separating the green cluster from the orange. By mapping shape to island of observation, the green species can be noted to live on all three islands, while the other species live on only one island.

Table 3: 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

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 10 shows a heat map of the demographics for these 84 participants.

8.2 Response time regression

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 a battery of all additive and multiplicative models. Table 3 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 interaction. The coefficients of this model are displayed in Table 4.

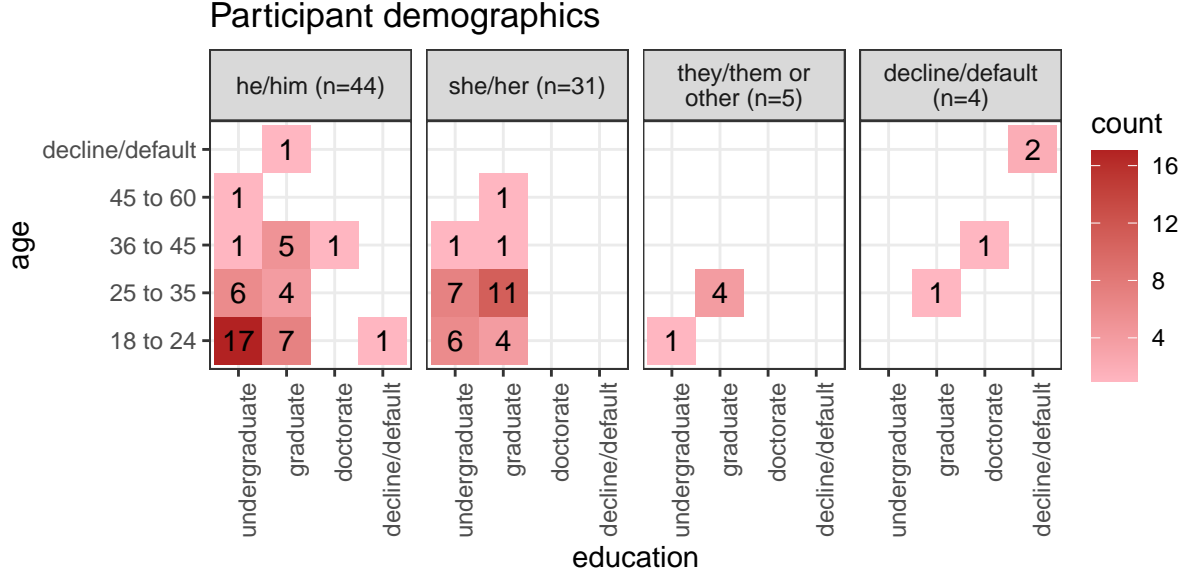


Figure 10: Heatmaps of survey participant demographics; counts of age group by completed education as faceted across preferred pronoun. Our sample tended to be between 18 and 35 years of age with an undergraduate or graduate degree.

Table 4: Model coefficients for log response time [seconds] $\hat{Y}_2 = \alpha \times \beta + \gamma + \delta$, 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	Pr(> t)	
(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	

Random effect ranges

The random effect terms further clarify the source of the error. Below is a comparison the effect ranges attributed to the participant and to the simulations next to their marginal effect on the response, a sort of upper bound of the error they could explain. This was performed for the models regressing accuracy and then log response time.

The residual plots have no noticeable nonlinear trends and contain striped patterns as an artifact from regressing on discrete variables. Figure 11 illustrates (T) the effect size of the random terms participant and simulation, or more accurately, the 95% CI from Gelman simulation of their posterior distribution. The effect size of the participant is much larger than simulation. The most extreme participants are statistically significant at $\alpha = .95$, while none of the simulation effects significantly deviate from the null of having no effect size on the marks. In comparison, (B) 95% confidence intervals participation and simulation mean accuracy, respectively.

Residual plots have no noticeable nonlinear trends and contain striped patterns as an artifact from regressing on discrete variables. Figure 11 illustrates (T) the effect size of the random terms participant and simulation, or more accurately, the 95% CI from Gelman simulation of their posterior distribution. The effect size of the participant is much larger than simulation. The most extreme participants are statistically significant at $\alpha = .95$, while none of the simulation effects significantly deviate from the null of having no effect size on the marks. In comparison, (B) 95% confidence intervals participation and simulation mean accuracy, respectively.

Similarly, figure 12 shows the Gelman simulations and marginal effects of the simulation and participants for the model with the same terms regressing on log response time.

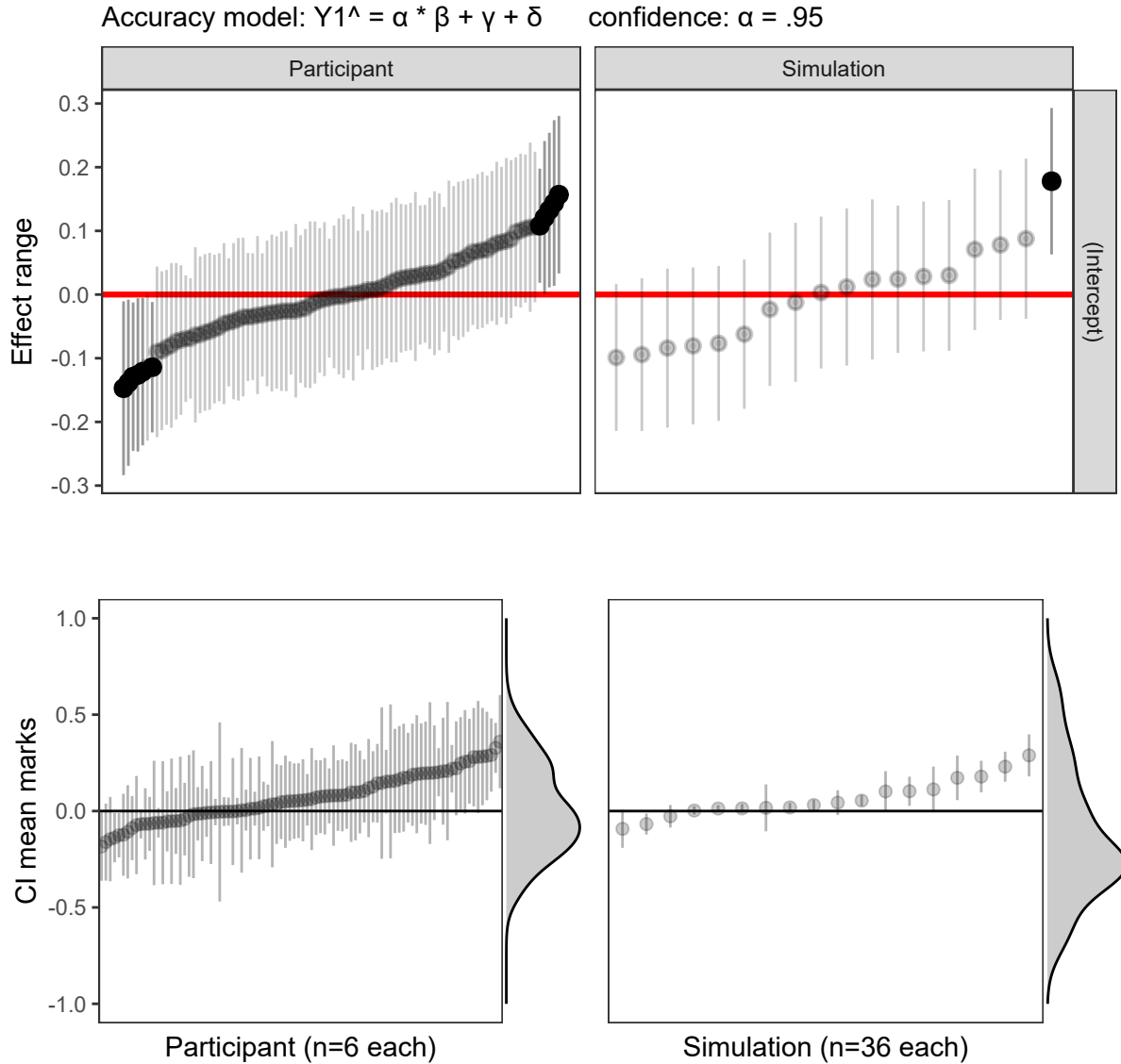


Figure 11: Accuracy model: (T) Estimated effect ranges of the random effect terms participant and data simulation of the accuracy model, $\hat{Y}_1 = \alpha \times \beta + \gamma + \delta$. Confidence intervals are created with Gelman simulation on the effect posterior distributions. The effect size of the participant is relatively large, with several significant extrema. None of the simulations deviate significantly. (B) The ordered distributions of the CI of mean marks follow the same general pattern and give the additional context of how much variation is in the data, an upper limit to the effect range. The effect ranges capture about two-thirds of the range of the data without the model. All intervals for $\alpha = .95$ confidence.

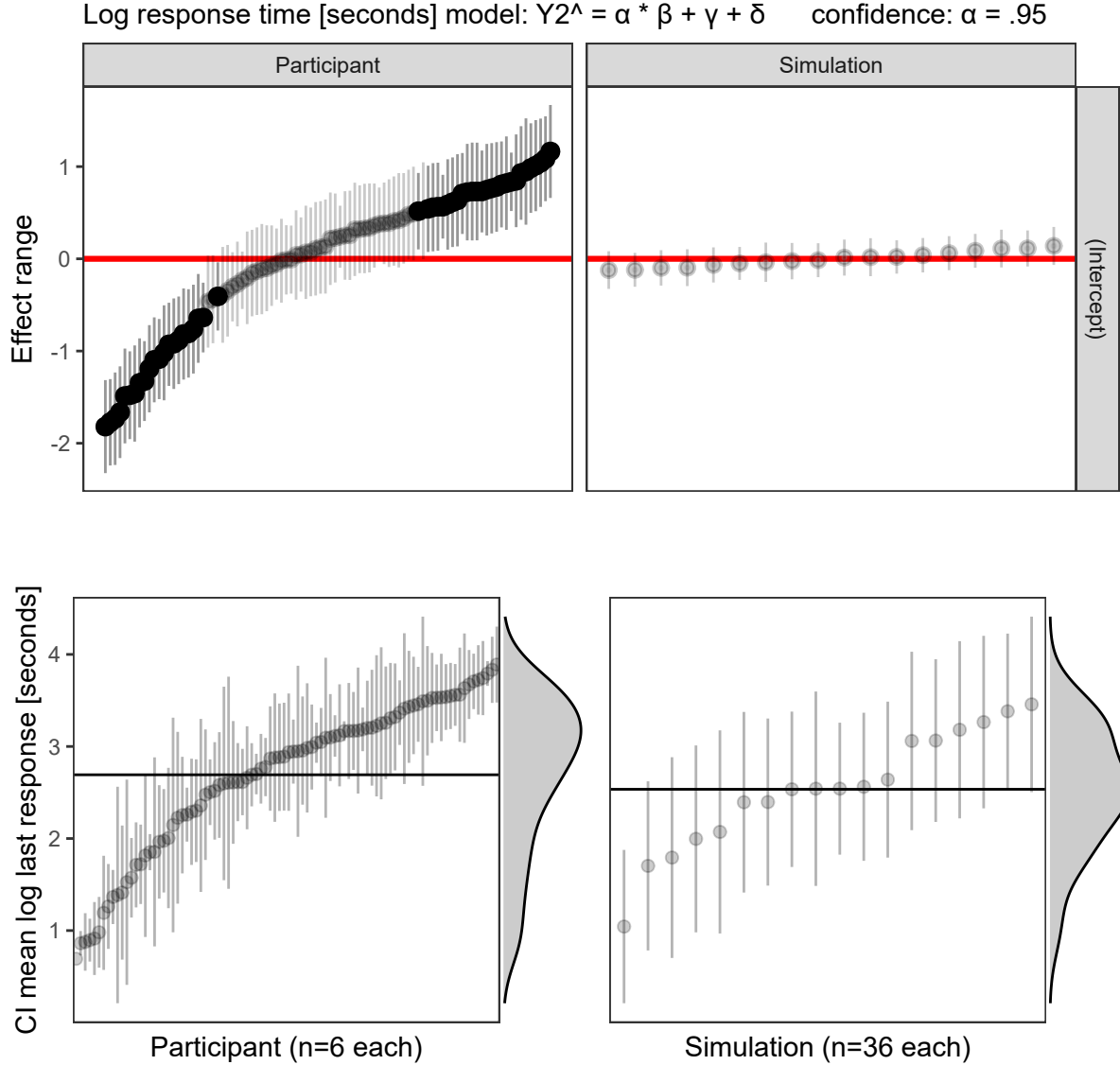


Figure 12: Log response time model: (T) The effect ranges of Gelman resimulation on posterior distributions for the time model, $\hat{Y}_2 = \alpha \times \beta + \gamma + \delta$. These show the magnitude and distributions of particular participants and simulations. Simulation has a relatively small effect on response time. (B) Confidence intervals for mean log time by participant and simulation. The marginal density shows that the response times are left-skewed after log transformation. Interpreting back to linear time there is quite the spread of response times: $e^1 = 2.7$, $e^{2.75} = 15.6$, $e^{3.75} = 42.5$ seconds. Of the simulations on the right, the bottom has a large variation in response time, relative to the effect ranges which means that the variation is explained in the terms of the model and not by the simulation itself.