JDSSV

Journal of Data Science, Statistics, and Visualisation

MMMMMM YYYY, Volume VV, Issue II. *doi: XX.XXXXX/jdssv.v000.i00*

A study examining the benefit of the user-controlled radial tour for understanding variable contributions to structure visible in linear projections of high-dimensional data

Nicholas Spyrison
Monash University

Dianne Cook
Monash University

Kimbal Marriott
Monash University

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 radial, manual tour is one instance that rotates the contribution of a selected variable. This paper describes a winthin-participants user study evaluating the efficacy of using the radial tour compared with principal component analysis and the grand tour. We devise a supervised classification task where participants evaluate variable attribution of the separation between two classes. We define an accuracy measure as response variable. Data were collected from 108 crowdsourced participants, who performed two trials of 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.

Keywords: multivariate data, exploratory data analysis, grand tour, manual tour, dimension reduction, linear projections, linear embeddings, R.

1. Introduction

Multivariate data is ubiquitous though its visualize quickly becomes complex. At the same time, we know that data visualization is imperative for Exploratory Data Analysis, (EDA, Tukey 1977) the initial data summarization and testing model assumptions. Data visualization is more robust and informative than statistical summarization alone (Anscombe 1973; Matejka and Fitzmaurice 2017; Yanai and Lercher 2020). Data visualization is integral to EDA and our comprehension of the data. But how do we know which visual methods yield the best perception of information for multivariate data?

Dimension reduction is commonly used with visualization to provide informative low-dimensional summaries of high-dimensional data. In particular, Principal Component Analysis (PCA, Pearson (1901)) remains exceedingly popular. Visualization of PCA is typically static scatterplots of only a few of the leading components. Alternatively, a class dynamic linear projections known as a *tour* (Asimov 1985) was published 16 years prior. Tours also use a scatterplot and biplot display of the data. Instead of a static view of two orthogonal components, they are animated over small changes to the basis. Asimov originally animated between randomly selected bases in the *grand* tour. The *manual* tour (Cook and Buja 1997) novelly allows for user-control over the basis changes. A selected variable can be rotated to the desired contribution. The permanence of the data points from frame to frame and information held in intermediate interpolated frames or user-control of the basis could plausibly lead to more information perception than a static display. There have been several user studies comparing scatterplots across display dimensionality in dimensionally reduced spaces (Gracia et al. 2016; Wagner Filho et al. 2018). There are also empirical statistics, used to describe non-linear reduction, and how well and faithfully they embed the data in a lower space (Bertini, Tatu, and Keim 2011; Liu et al. 2017; Sedlmair, Munzner, and Tory 2013; Maaten and Hinton 2008). There is an absence of studies comparing visual techniques.

This paper describes a crowdsourced (prolific.co), user study conducted to assess the efficacy of different visualization methods: PCA as discrete pairs of two components, the grand tour, and the radial tour. *Tours* are animated linear projections over small changes to their bases and described in more detail in Section 2.4 Cluster data is simulated under several additional experimental factors: location, shape, and dimensionality. We task participants with identifying the variables attributing to the separation between two clusters. We define an accuracy measure for this task. Mixed models regressing on this measure finding strong evidence supporting the radial tour has for a considerable improvement in accuracy over alternatives.

Knowing which visuals and dimension reduction to use is also important for statistical modeling and their interpretations. Models are becoming increasingly complex involving many interacting, or otherwise non-linear terms cause an opaqueness to their interpretability. Exploratory Artificial Intelligence (XAI, Adadi and Berrada (2018); Arrieta et al. (2020)) is an emerging field that attempts to extend the interpretability of such black-box models by offering techniques to increase their interpretability. Multivariate data visualization is an essential part of 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 6.

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. Adding all combinations of pairs of variables can be viewed as a scatterplot matrix (Chambers et al. 1983). Figure 1 shows a scatterplot matrix of the first three components of simulated data. Looking at p univariate densities and many bivariate scatterplots quickly becomes burdensome as dimensionality increases. It also displays information containing in two orthogonal dimensions; features in three dimensions will never be resolved.

2.2. Parallel coordinates plot

Another common way to display multivariate data is with a parallel coordinates plot (Ocagne 1885), which displays observations by their quantile or normalized values for each variable connected by lines to the quantile value in subsequent variables. Parallel coordinates plot and other observations-based visuals, such as pixel plots or Chernoff faces scale well with dimensions and but poorly with observations. These are perhaps best used when there are more variables than observations.

Observations-based visuals have a couple of issues. One is that they are asymmetric across variable ordering, leading to different conclusions or features being focused on due to variable order. Another notable issue of observations-based visuals 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 most discerning visual channel.

At some point, then I think we will be forced to turn to dimension reduction to scale well. Non-linear transformations bend and distort spaces 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. Principal component analysis (Pearson 1901) defines

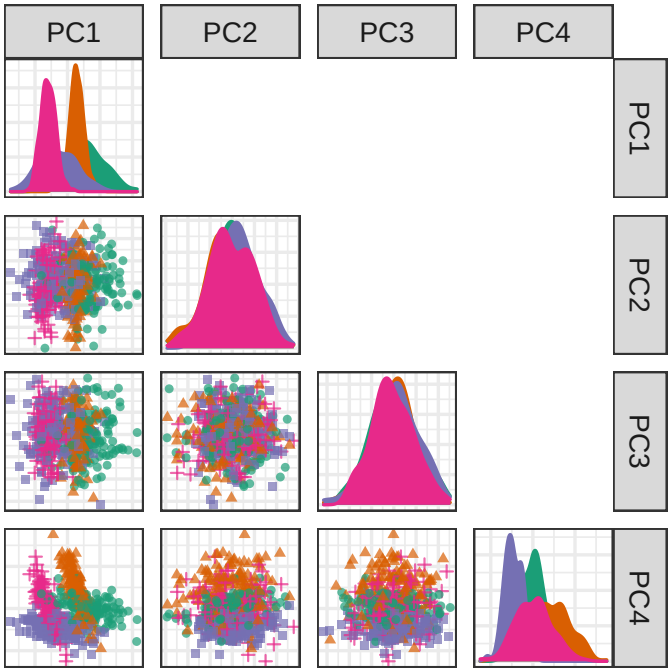

Figure 1: Scatterplot matrix of the first four principal components simulated data in six dimensions. An analyst would have to view both PC1 by PC2 and PC1 by PC4 to have a thorough take on which variables attribute to the separation between clusters.

new components, a linear combination 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 with the use of 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 holding less variation as well.

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 selection of their basis paths (Lee et al. 2021; Cook et al. 2008). To contrast with the discrete orientations of PCA, we compare with continuous changes of linear projection with grand and radial tours.

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 the radial tour but lacks the user control enjoyed by PCA and radial tours.

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. This 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. The target bases are then similarly orthogonally restrained, the data is projected through 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 method of steering the projection basis. It is advisable first to identify a basis of particular interest and then use the manual tour as 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 selected variable can change its magnitude of contribution but not its angle; 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 allow the user to steer via choosing the variable to rotate.

The recent implementation of manual tours us 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 **gganimte** (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.

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: cluster shape, location of the cluster separation, 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 common linear embedding. The grand tour will act as a secondary control that will help evaluate the benefit of animation but without influencing its path. Lastly, the radial tour should perform best as it benefits both from animation and being user-control of the contribution of individual variables.

Then for some subset of tasks, we expect to find that the radial tour performs most accurately, as it enjoys the persistence of points and user-control to explore specific variables. In the appendix, section 9, we also regress on the last response time. We may expect the grand tour performs faster than the alternatives as its absence of inputs will allow focusing all of their attention on interpreting the fixed path. 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 was modest, it seems plausible that grand tour regularly crossed frames with the correct information for the task.

We measure the accuracy and response time over the support of the discussed experimental factors. The null hypotheses can be stated as:

H_0 : task accuracy does not change across visualization method
H_α : task accuracy does not change across visualization method

3.2. Experimental factors

In addition to visual factor, we simulate the data across three aspects: 1) The *location* of the difference between clusters by mixing a signal and a noise variable at different ratios, we vary the number of variables and their magnitude of cluster separation, 2) the *shape* of the clusters, to reflect varying distributions of the data, and 3) the *dimension*-ality of the data. Below we describe the levels within each factor, while Figure 2 gives a visual representation of the levels.

The *location* of the separation of the clusters is at the heart of the measure. It would be good to test a few varying levels. To test the sensitivity to this, we mix a noise variable with the signal-containing variable such that the difference in the clusters is mixed at the following percentages: 0/100% (not mixed), 33/66%, 50/50% (evenly mixed).

In selecting the *shape* of the clusters, we follow the convention given by Scrucca et al. (2016), where 14 variants of model families containing three clusters are defined. The name of the model family is the abbreviation of its respective volume, shape, and orientation of the cluster, which are either equal or vary. We use the models EEE, EEV, and EVV. The latter is further modified by moving four-fifths of the data out in a "V" or banana-like shape.

Dimension-ality is tested at two modest levels, namely, in four dimensions containing three clusters and six dimensions with four clusters. Such modest dimensionality is required to bound the difficulty and search space to keep the task realistic for crowd-

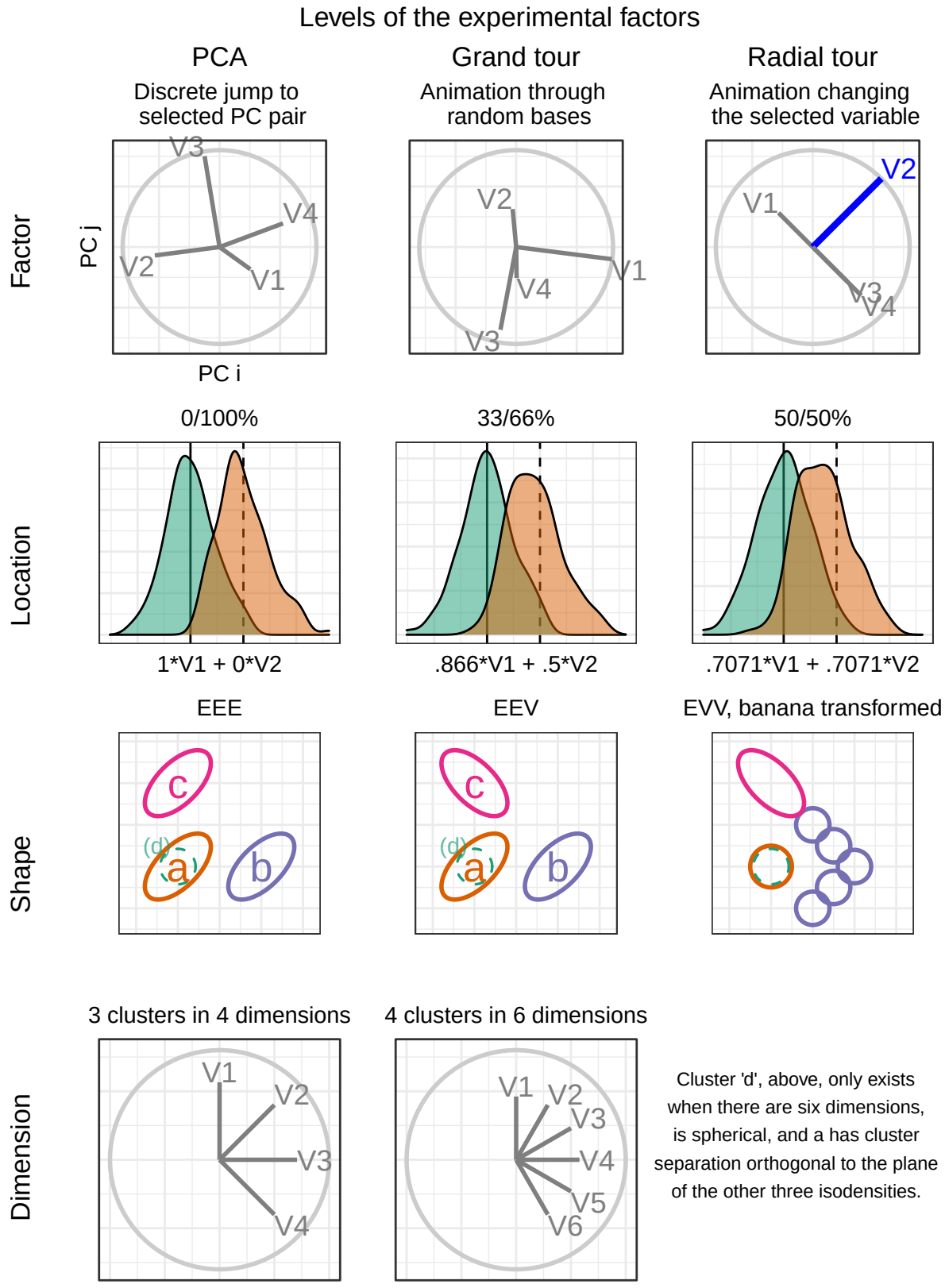

Figure 2: Illustration of the experimental factors, the parameter space of the independent variables, the support of our study.

sourcing.

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 elements of the display are held constant shown as a 2D scatterplot with an axis biplot to its left. Observations were supervised with the cluster level mapped to color 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 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 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 uniformed distributed cluster separation in the scatterplot. Independent with factor and block values, 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 evaluation measure of this task was designed with a few of features in mind: 1) the sum of squares of the individual variable weights should be one, 2) symmetric about zero, that is, without preference to under- or over-guessing 3) heavier than linear weight with increasing distance from a uniform height. With these in mind, we define the following measure for evaluating the task.

Let a data $\mathbf{X}_{n\times p}$ 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 will indicate the number of the cluster an observation belongs to. Cluster membership is exclusive; an observation cannot belong to more than one cluster.

We define weights, W to be a vector explaining the variable-wise difference between two clusters. Namely, the difference of each variable between clusters, as a proportion of the total difference, less $1/p$, the amount of difference each variable would hold if it were uniformly distributed. 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. Then Y is a vector of variable accuracy.

$$W_i = \frac{(\overline{X_{i=1,k=1}} - \overline{X_{1,2}},\ \ldots\ (\overline{X_{p,1}} - \overline{X_{p,2}})}{\sum_{i=1}^{p}(|\overline{X_{i,k=1}} - \overline{X_{i,2}}|)} - \frac{1}{p}$$

$$\text{Accuracy}, Y = \sum_{i=1}^{p} I(r_i) * sign(w_i) * \begin{cases} \dfrac{w_+^2}{\sum w_+^2} & \Big| w_+ \text{ the positive elements of } w \\ \dfrac{w_-^2}{\sum w_-^2} & \Big| w_- \text{ the negative elements of } w \end{cases}$$

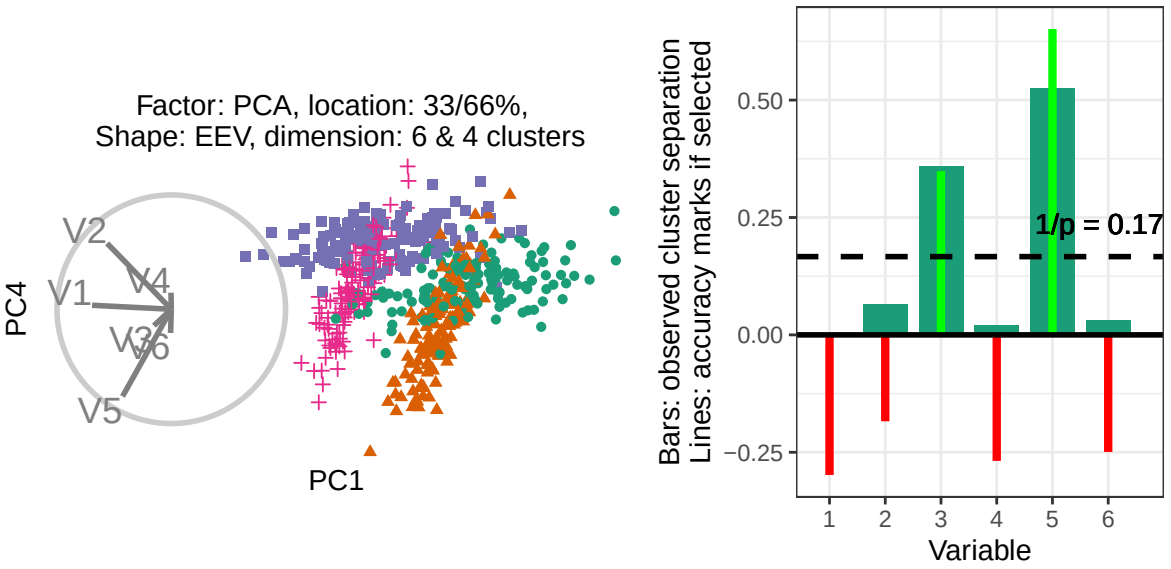

Figure 3: (L), PCA biplot of the components showing the most cluster separation with (R) illustration of the magnitude of cluster separation is for each variable (bars) and the weight of the variable accuracy if selected (red/green lines). The horizontal dashed line is $1/p$, the amount of separation each variable would have if evenly distributed. The weights equal the signed square of the difference between each variable value and the dashed line.

Where $I(r_i)$ is the indicator function on the response for variable i.

3.4. Visual design standardization

The factors are tested within-participant, with each visual being evaluated twice by each participant. The order that factors are experienced is controlled with the block assignment as illustrated below in Figure 4. Below we cover the visual design standardization, as well the input and display within each factor.

The visualization methods were standardized wherever possible. Data was displayed as 2D scatterplots with biplots (Gabriel 1971), a visual with variable contributions inscribed on a unit circle. 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 factors were their inputs.

CA inputs allowed users to select between the top four principal components for both the x- and y-axis regardless of the data dimensionality (four or six). Data was simulated to have cluster separation within the 2nd to 4th components, ideally not simple enough to be entire observed in PC1 by PC4. While also not buried in 5th and 6th components in the interest of simplicity and time. There was no user input for the grand tour; users were instead shown a 15-second animation of the same randomly selected path. Participants were able to view the same clip up to four times within the time limit. Radial tours were also displayed at five frames per second with a step size of 0.1 radians between interpolated frames. Users were able to swap between variables. The

display would change the start of radially increasing the contribution of the selected variable until it was full, zeroed, and then back to its initial contribution. The complete animation of any variable takes about 20 seconds and is almost entirely in the projection frame at around six seconds. The starting basis of each is initialized to a half-clock design, where the variables were evenly distributed in half of the circle which is then orthonormalized. This design was created to be variable agnostic while maximizing the independence of the variables.

3.5. Data simulation

Each dimension is originally distributed as $\mathcal{N}(2*I(signal), 1) \mid$ covariance Σ, a function of the shape factor. Signal variables had a correlation of 0.9 when they have equal orientation and -0.9 when their orientations vary. 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. Then location mixing is applied by post-multiplying a (2x2) rotation matrix to the signal variable and a noise variable for the clusters in question. All variables are then standardized by standard deviation. The rows and columns are then shuffled randomly. The observation's cluster and order of shuffling are attached to the data and saved.

Each of these replications is then iterated with each level of the factor. For PCA, every pair of the top four principal components and saved as 12 plots. We first save two basis paths of differing dimensions for the grand tour before each replication is projected through the common basis path. Each simulations variable order was previously shuffled effectually randomizing cluster separation shown, while mitigating bias from fringe selection of target bases. The resulting animations were saved as .gif files. The radial tour starts at either the four or six variable "half-clock" basis, where each variable has a uniform contribution in the right half and no variable contributing in the opposite direction (to minimize the dependence between variable contributions), 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, we explain how the experimental factors are assigned and illustrate how this is experienced from a participant's perspective.

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

Each period starts with an untimed training task at the simplest remaining block 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 is evaluated on two tasks with the same factor*location level across the increasing difficulty of dimension*shape. The plot was removed after 60 seconds, though this limit was rarely reached by participants.

The order of the factor and location levels is randomized with a nested Latin square

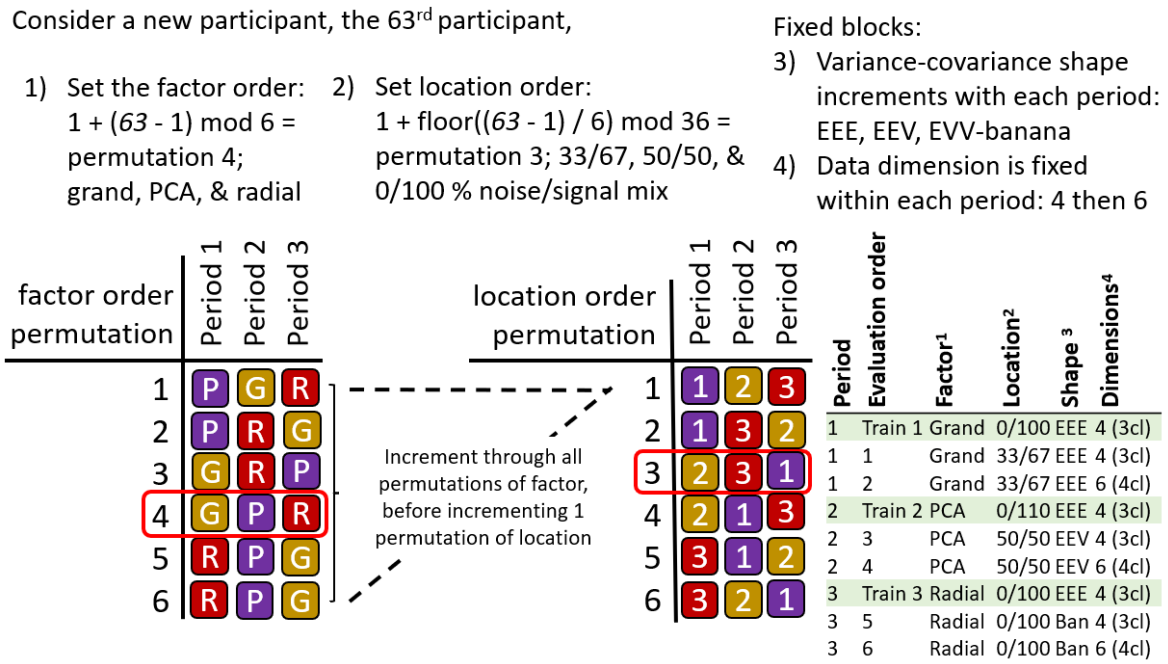

Figure 4: Illustration of how a hypothetical participant 63 is assigned factor and block parameterizations. Each of the 6 factor order permutations is exhausted before iterating to the next permutation of location order.

where all levels of the visual factor are exhausted before advancing to the next level of location. That means we need $3!^2 = 36$ participants to evaluate all permutations of the experimental factors once. This randomization controls for potential learning effects the participant may receive. Figure 4 illustrates how an arbitrary participant experiences the experimental factors.

Through pilot studies sampled by convenience (information technology and statistics Ph.D. students attending Monash University), we predict that we need three full evaluations to properly power our study; we set out to crowdsource $N = 3 * 3!^2 = 108$ participants.

3.7. Participants

We recruited $N = 108$ participants via prolific.co (Palan and Schitter 2018). We filtered participants 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); we apply this filter under the premise that linear projections and biplot displays used 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. We cannot preclude previous knowledge or experience with the factors but validate this assumption in the follow-up survey asking about familiarity with the factors. The appendix 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, we throttled the number of participants and over-collect survey trials until we received our target three evaluations of all permutation levels.

The processing steps were minimal. First, we format to an analysis-ready form, decoding values to a more human-readable state, and add a flag indicating whether the survey had complete data. We filter to only the latest three complete studies of each block parameterization, those which should have experienced the most minor 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 on GitHub https://github.com/nspyrison/spinifex_study.

4. Results

To recap, the primary response variable is task accuracy, as defined in section 3.3. The parallel analysis of the log response time is provided in the appendix. We have two primary data sets; the user study evaluations and the post-study survey. The former is the 108 participants with the explanatory variables: visual factor, location of the cluster separation signal, the shape of variance-covariance matrix, and the dimensionality of the data. Block parameterization and randomization were discussed in section 3.2. The survey was completed by 84 of these 108 people. It collected demographic information (preferred pronoun, age, and education), and subjective measures for each factor (preference, familiarity, ease of use, and confidence).

Below we look at the marginal performance of the block parameters and survey responses. After that, we build a battery of regression models to explore the variables and their interactions. Lastly, we look at the subjective measures between the factors.

4.1. Accuracy regression

To more thoroughly examine explanatory variables, we regress against accuracy. All models have a random effect term on the participant, which captures the effect of the individual participant. After we look at models of the block parameters, we extend to compare against survey variables. Last, we compare adding a random effect for data and regressing against time till the last response fares against benchmark models. The matrices for models with more than a few terms quickly become rank deficient; there

Table 1: Model performance of random effect models regressing accuracy. Each model includes a random effect term of the participant explaining the individual's influence on accuracy. Complex models perform better in terms of R^2 and RMSE, yet AIC and BIC penalizes their large number of fixed effects in favor of the much simpler model containing only the visual factor. Conditional R^2 includes the random effects, while marginal does not.

Fixed effects	No. levels	No. terms	AIC	BIC	R2 cond.	R2 marg.	RMSE
a	1	3	1096	1123	0.197	0.018	0.492
a+b+c+d	4	8	1121	1170	0.206	0.028	0.491
a*b+c+d	5	12	1130	1197	0.214	0.041	0.488
a*b*c+d	8	28	1161	1300	0.253	0.078	0.478
a*b*c*d	15	54	1212	1467	0.287	0.113	0.469

is not enough data to explain the effect terms.

In building a set of models to test, we include all single term models with all independent terms. We also include an interaction term of factor by location, allowing for the slope of each location to change across each level of the factor. For comparison, an overly complex model with all interaction terms is included.

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

where
α_i, fixed term for factor | $i \in$ (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 | $k \in$ (EEE, EEV, EVV banana) model shapes
δ_l, fixed term for dimension | $l \in$ (4 variables & 3 cluster, 6 variables & 4 clusters)
μ is the intercept of the model including the mean of random effect
$\mathbf{Z} \sim \mathcal{N}(0,\ \tau)$, the error of the random effect of participant
$\mathbf{W} \sim \mathcal{N}(0,\ \upsilon)$, the error of the random effect of simulation
$\epsilon \sim \mathcal{N}(0,\ \sigma)$, the remaining error in the model

We also want to visually explore the conditional variables in the model. Figure 5 explores violin plots of accuracy by factor while faceting on the location (vertical) and shape (horizontal). Use of the radial visual, on average, increase the accuracy received, and especially so when there is no signal/noise mixing.

4.2. Subjective measures

The 84 evaluations of the post-study survey also collect four subjective measures for each factor. Figure 6 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

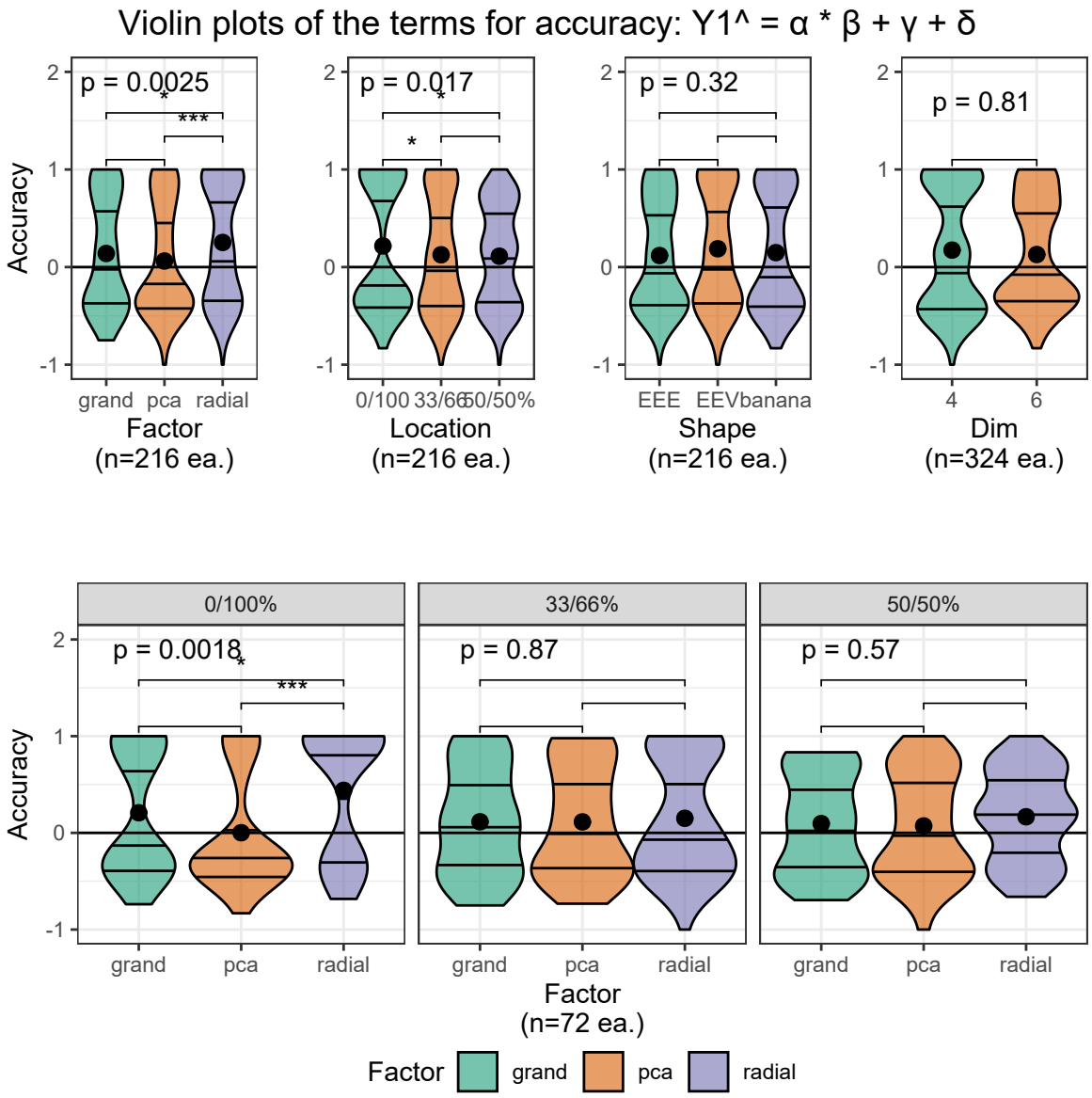

Figure 5: Violin plots of terms of the model $\hat{Y} = \alpha * \beta + \gamma + \delta$. Overlaid with global significance from the Kruskal-Wallis test and pairwise significance from the Wilcoxon test, both are non-parametric, ranked sum tests suitable for handling discrete data. Participants are more confident and find the radial tour easier to use than the grand tour. Participants claim low familiarity, as we expect from crowdsourced participants. Radial is more preferred compared with either alternative for this task.

Table 2: The task accuracy model coefficients for $\widehat{Y} = \alpha * \beta + \gamma + \delta$, with factor=pca, location=0/100%, and shape=EEE held as baselines. Factor being radial is the fixed term with the strongest evidence supporting the hypothesis. When crossing factor with location radial performs worse with 33/66% mixing relative to the PCA with no mixing. The model fit is based on the 648 evaluations by the 108 participants.

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	0.18	0.09	34.5	1.94	0.060	
Factor						
Factorpca	-0.15	0.09	622.9	-1.63	0.103	
Factorradial	0.25	0.09	619.7	2.62	0.009	**
Fixed effects						
Location33/66%	-0.05	0.11	61.9	-0.49	0.623	
Location50/50%	-0.08	0.11	60.5	-0.80	0.426	
ShapeEEV	0.07	0.07	11.3	1.01	0.333	
Shapebanana	0.03	0.07	11.3	0.43	0.674	
Dim6	-0.05	0.06	11.3	-0.81	0.436	
Interactions						
Factorpca:Location33/66%	0.13	0.14	592.0	0.95	0.341	
Factorradial:Location33/66%	-0.25	0.14	583.3	-1.76	0.079	
Factorpca:Location50/50%	0.10	0.14	594.3	0.70	0.482	
Factorradial:Location50/50%	-0.16	0.14	592.6	-1.16	0.247	

answers and found radial tours easier than grand tours. All factors have reportedly low familiarity, as expected from crowdsourced participants.

5. Conclusion

Data visualization is an integral part of EDA. Yet through exploration of data in high dimensions become 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. We find strong evidence that using the radial tour improves the accuracy relative to PCA or the grand tour on the supervised cluster task assigning variable attribution to the separation of the two clusters.

This paper discussed an $n = 108$, with-in participant user study comparing the efficacy of three linear projection techniques. The participants performed a supervised cluster task, explicitly identifying which variables contribute toseparating two target clusters. This was evaluated evenly over four experimental factors. In summary, we find strong evidence that using the radial tour leads to a sizable accuracy increas. There is also evidence for a small change response time, with increasing order of PCA, grand, and radial. The effect sizes on accuracy are large relative to the change from the other experimental factors, though smaller than the random effect of the participant. The radial tour was subjectively preferred, leading to more confidence in answers, and increased ease of use than the alternatives.

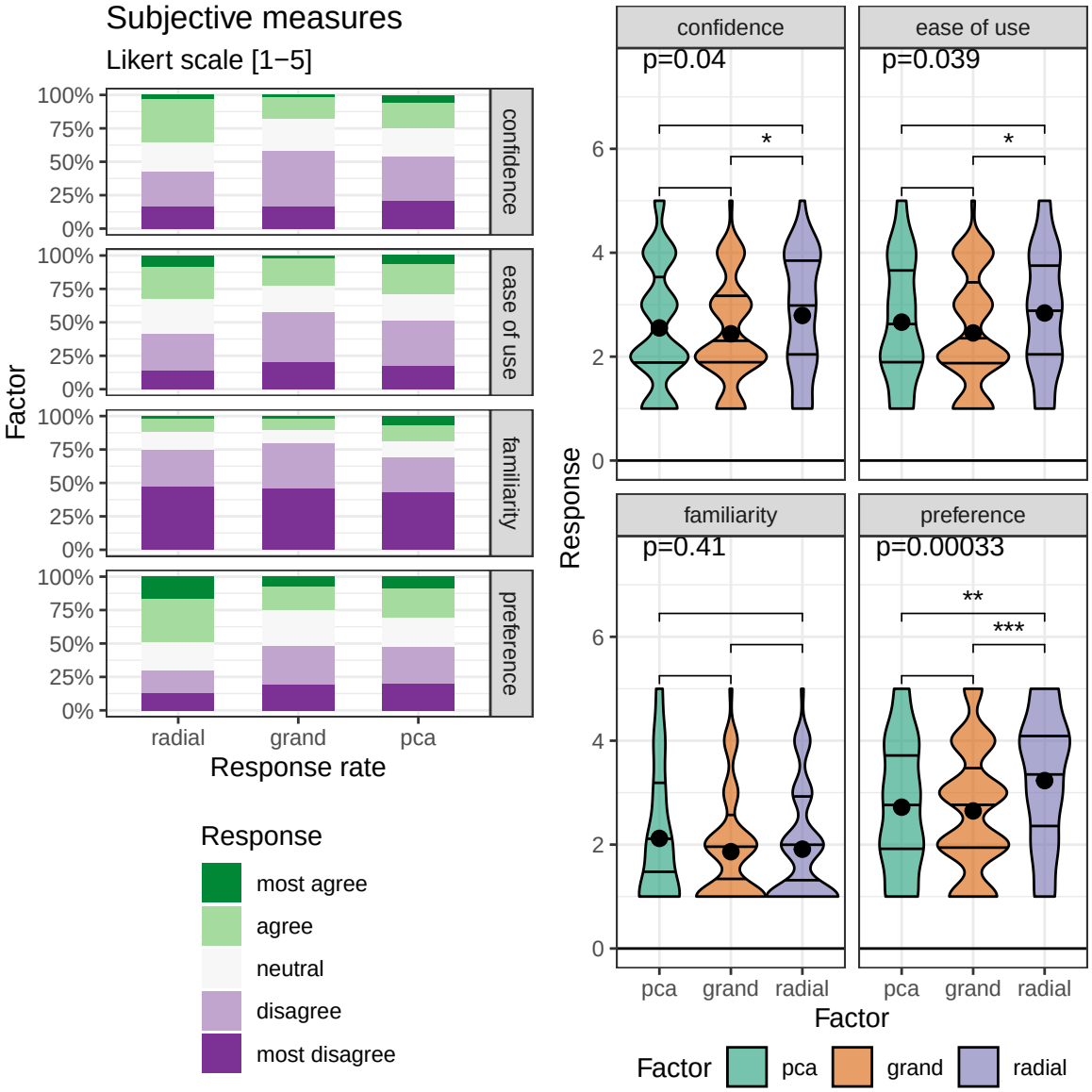

Figure 6: The subjective measures of the 84 responses of the post-study survey, five discrete Likert scale levels of agreement (L) Likert plots (stacked percent bar plots) with (R) violin plots of the same measures. Violin plots are overlaid with global significance from the Kruskal-Wallis test, and pairwise significance from the Wilcoxon test, both are non-parametric, ranked sum tests.

There are several ways that this study could be extended. In addition to expanding the support of the experimental factors, more exciting directions include: changing the type of the task, visualizations used, and experience level of the target population. It is difficult to achieve good coverage given the number of possible permutations. Keep in mind the traffic volume and low effort of responses from participants when crowdsourcing.

6. Accompanying tool: radial tour application

We have produced an application to illustrate the radial tour to accompany this study. The **R** package, **spinifex**, (Spyrison and Cook 2020) is free, open-source and now contains a **shiny** (Chang et al. 2020) application that allows 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.

```
## Download:
install.packages("spinifex", dependencies = TRUE)
## Run interactive app demonstrating the radial tour:
spinifex::run_app()
```

7. 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. The source files for this article, application, data, and analysis can be found on GitHub. The source code for the **spinifex** package and accompanying shiny application can be found here.

8. References

10 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.

Bertini, Enrico, Andrada Tatu, and Daniel Keim. 2011. "Quality Metrics in High-Dimensional Data Visualization: An Overview and Systematization." *IEEE Transactions on Visualization and Computer Graphics* 17 (12): 2203–12.

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.

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, Winston, Joe Cheng, J. J. Allaire, Yihui Xie, and Jonathan McPherson. 2020. *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 GonzÃ¡lez, VÃctor 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 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.

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.

Maaten, Laurens van der, and Geoffrey Hinton. 2008. "Visualizing Data Using t-SNE." *Journal of Machine Learning Research* 9 (Nov): 2579–2605.

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."

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 Grolemund. 2018. *R Markdown: The Definitive Guide*. Boca Raton, Florida: Chapman; Hall/CRC. https://bookdown.org/

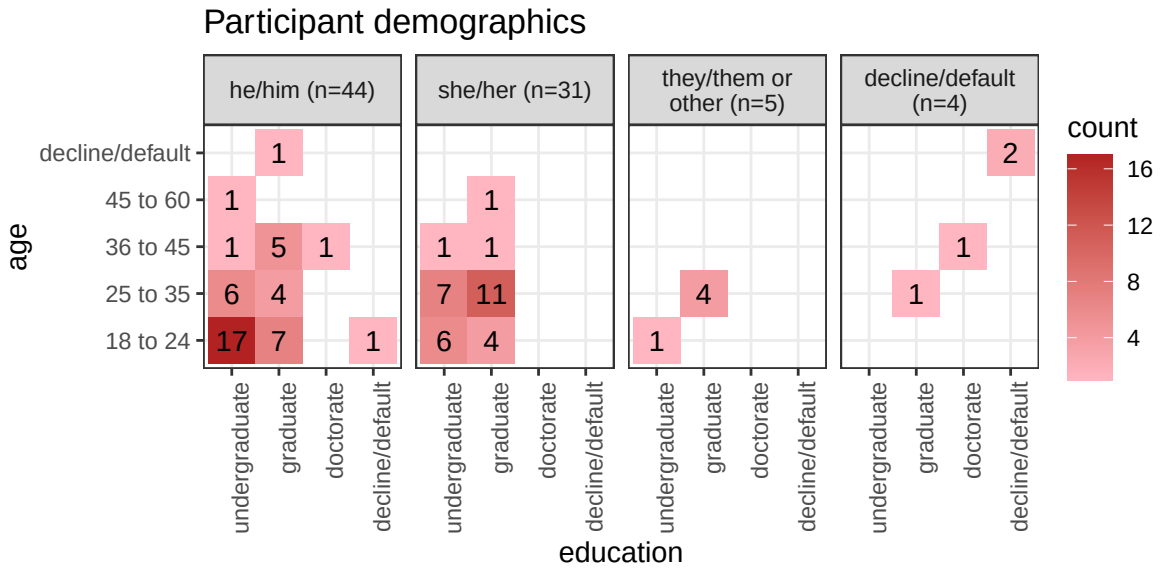

Figure 7: Heatmaps of survey participant demographics; counts of age group by completed education as faceted across preferred pronoun. Our sample tended to be between 18 and 35 years of age with an undergraduate or graduate degree.

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.

9. Appendix

Survey participant demographics

The target population is relatively well-educated people, as linear projections may prove difficult for generalized consumption. Hence we restrict Prolific.co participants 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, who are 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.

9.1. Response time regression

As a secondary explanatory variable, we also want to look at time. First, we take the log transformation of time as it is right-skewed. We repeat the same modeling procedure: 1) Compare the performance of a battery of all additive and multiplicative models. 2) Select the model with the same effect terms, $\alpha * \beta + \gamma + \delta$, and examine its coefficients.

Random effect ranges

Table 3: Model performance regressing on log response time [seconds], $\widehat{Y_2}$ random effect models, where each includes random effect terms for participants and simulations. We see the same trade-off where the simplest factor model is preferred by AIC/BIC, while R^2 and RMSE is largest in the full multiplicative model. We again select the model $\alpha * \beta + \gamma + \delta$ to explore further as it has relatively high marginal R^2 while having much less complexity than the complete interaction model. Conditional R^2 includes the random effects, while marginal does not.

Fixed effects	No. levels	No. terms	AIC	BIC	R2 cond.	R2 marg.	RMSE
a	1	3	1096	1123	0.197	0.018	0.492
a+b+c+d	4	8	1121	1170	0.206	0.028	0.491
a*b+c+d	5	12	1130	1197	0.214	0.041	0.488
a*b*c+d	8	28	1161	1300	0.253	0.078	0.478
a*b*c*d	15	54	1212	1467	0.287	0.113	0.469

Table 4: Model coefficients for log response time [seconds] $\widehat{Y_2} = \alpha * \beta + \gamma + \delta$, with factor=PCA, location=0/100%, and shape=EEE 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)	2.48	0.14	42.8	17.41	0.000	***
Factor						
Factorpca	0.23	0.12	567.6	1.97	0.049	*
Factorradial	0.39	0.12	571.9	3.29	0.001	**
Fixed effects						
Location33/66%	0.29	0.14	42.0	2.06	0.046	*
Location50/50%	0.07	0.14	40.5	0.54	0.594	
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						
Factorpca:Location33/66%	-0.24	0.18	580.9	-1.34	0.181	
Factorradial:Location33/66%	-0.48	0.18	583.1	-2.63	0.009	**
Factorpca:Location50/50%	-0.12	0.18	578.6	-0.69	0.491	
Factorradial:Location50/50%	-0.08	0.18	580.9	-0.43	0.668	

Residual plots have no noticeable non-linear trends and contain striped patterns as an artifact from regressing on discrete variables. Figure 8 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 9 shows the Gelman simulations and marginal effects of the simulation and participants for Y_2, the same model regressing on log response time.

Radial tour application details

Below we describe the **shiny** app available in the **spinifex** package which runs locally. It streamlines creating and interacting with radial tours more interactive than the application used in the user study.

In the initial tab, 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 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 animating 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.

Affiliation:

Nicholas Spyrison
Monash University
Faculty of Information Technology
ORCiD: 0000-0002-8417-0212
E-mail: nicholas.spyrison@monash.edu

Dianne Cook
Monash University
Department of Econometrics & Business Statistics,
ORCiD: 0000-0002-3813-7155

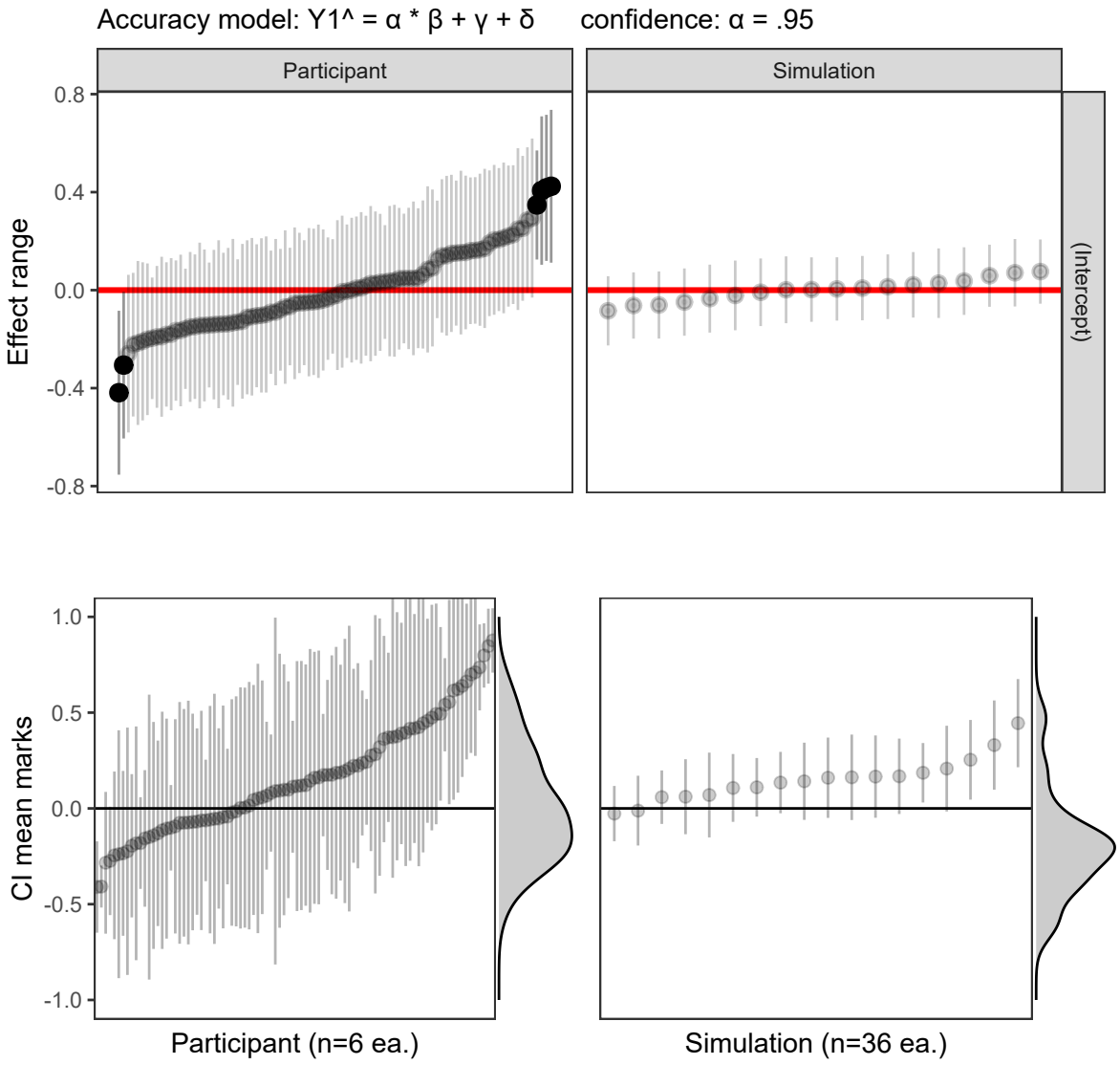

Figure 8: Accuracy model: (T) Estimated effect ranges of the random effect terms participant and data simulation of the accuracy model, $\widehat{Y_1} = \alpha * \beta + \gamma + \delta$. Confidence intervals are created with Gelman simulation on the effect posterior distributions. The effect size of 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.

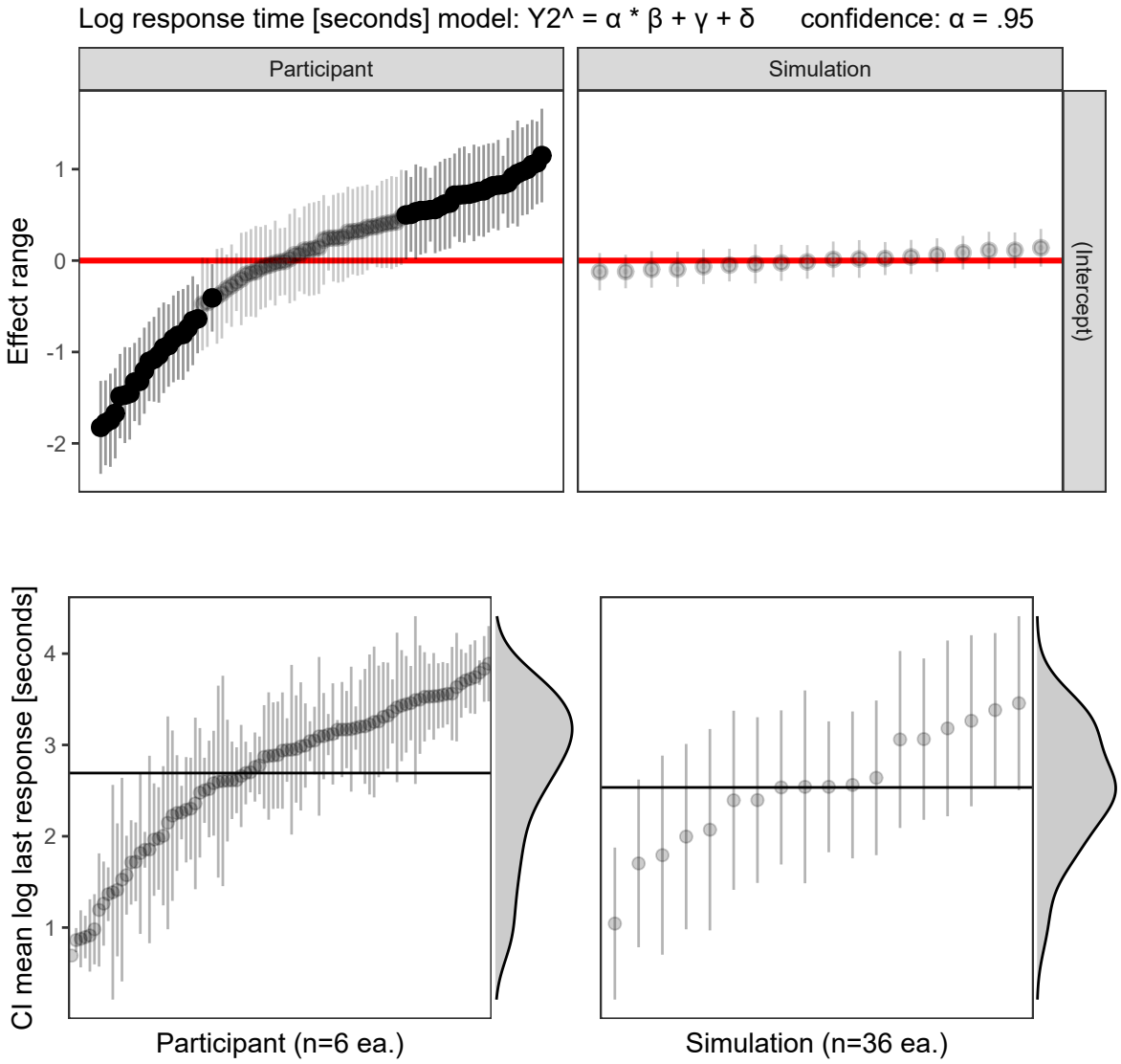

Figure 9: Log response time model: (T) The effect ranges of Gelman resimulation on posterior distributions for the time model, $\widehat{Y_2} = \alpha * \beta + \gamma + \delta$. These show the magnitude and distributions of particular participants and simulations. Simulation has a relatively small effect 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.

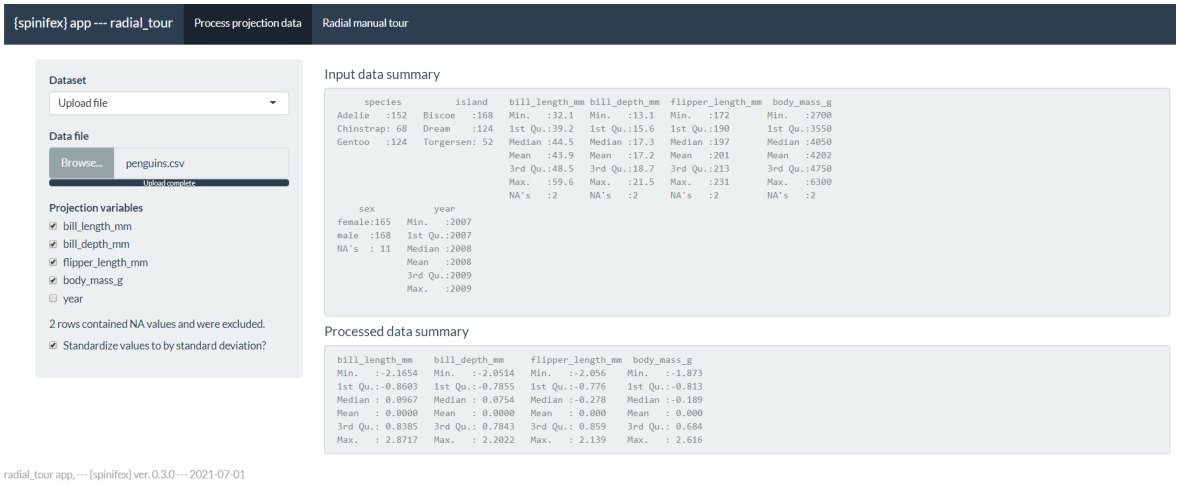

Figure 10: Process data tab, interactively loads or select data, check which variables to project, and optionally scale columns by standard deviation.

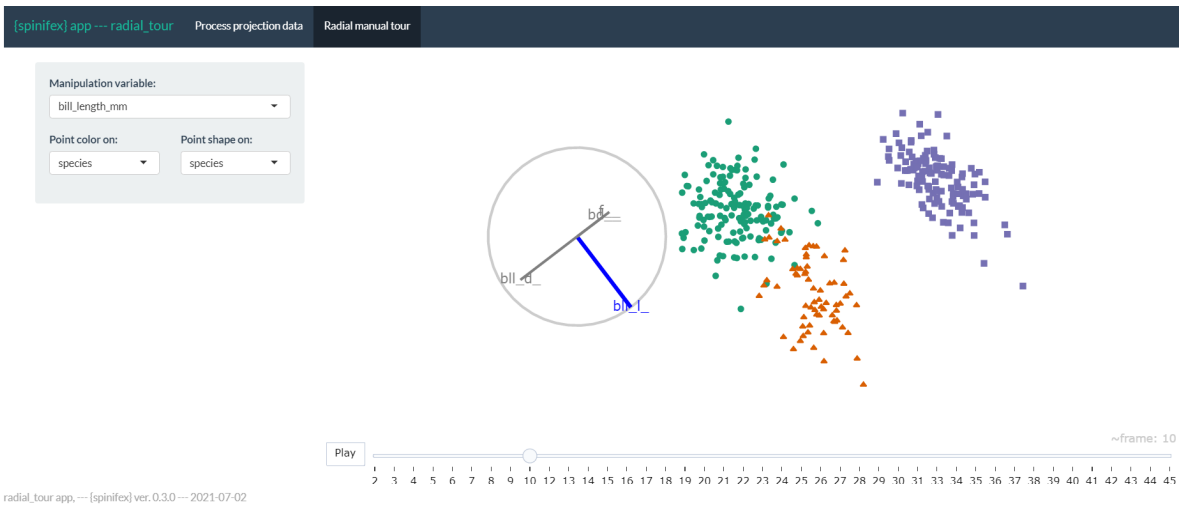

Figure 11: Radial tour tab, interactively create radial tours, changing the manipulation variable, and color or shape of the resulting manual tour. Here, the palmer penguins data is being explored, bill length was selected to manipulate as it is the only variable separating the green cluster from the orange. By mapping shape to island of observation, we also notice that the species in green live on all three islands, while the other species live only on one of the islands.

Kimbal Marriott
Monash University
Faculty of Information Technology,
ORCiD: 0000-0002-9813-0377

Journal of Data Science, Statistics, and Visualisation https://jdssv.org/
published by the International Association for Statistical Computing
http://iasc-isi.org/

MMMMMM YYYY, Volume VV, Issue II *Submitted:* yyyy-mm-dd
doi:XX.XXXXX/jdssv.v000.i00 *Accepted:* yyyy-mm-dd