The effect of user interaction for understanding variable contributions to structure in linear projections

Nicholas Spyrison, Dianne Cook, Kimbal Marriott

Abstract

Viewing data in its original variable space is foundamental to the exploratory data analysis. For multivariate data this is an complex task. We perform a between-participant user study to evaluate 3 types of linear embeddings, namely, biplots of principal compoints, grand tours, and radial tours. Crowdsourced participants (N = 108) were asked to identify which variable(s) explain the difference between a pair of clusters within the data as the factor of visualization is changed within participant. Principal componants and grand tour are the discrete and animated benchmarks to beat. We find radial tours score higher and respond slightly faster than alternatives. Visual factor is significantly more important than the other block parameterizations (location, shape, dimension) or learning effects from the order of evaluation. Demographics, prior experience, and parameter values vary in significance, while the size of their coeffecients all relatively small as compared with visual and the random effect of the participants.

1 Introduction

Multivariate data is ubiquitous. Yet exploratory data analysis (EDA) (Tukey 1977) of such spaces becomes difficult, increasingly so as dimension increases. Numeric statistic summarization of data often doesn't explain the full complexity of the data or worse, can lead to missing obvious visual patterns (Anscombe 1973; Matejka and Fitzmaurice 2017; Goodman 2008; Coleman 1986). Data should be visually inspected in it's original variable-space before applying models or summarizations. This allows users to validate assumptions, identify outliers, and facilitates the identification of visual peculiarities.

For these reasons, it is important to use visualizations of data spaces and extend the diversity of its application. However, visualizing data containing more than a handful of variables is not trivial. Scatterplot matrices or small multiples (Chambers et al. 1983) looks at all permutation pairs of variables, but quickly becomes to vast a number of images to consider. On the other extreme, parallel coordinates plot (Ocagne 1885) and its radial variants, plot observations as lines varying across scaled variables as displayed in a line or circle. This scales well with dimensionality, while suffering from couple issues. The larger issue, being the loss of mapping multiple variables to graphic position, which is perhaps the most important visual cue for human perception (Munzner 2014). The lesser being that they suffer from asymmetry, as their interpretation is dependent on variable ordering.

Using a linear combinations of variables will allow us to keep position in 2 display axes while peering into information not contained in any one dimension. The idea of using a combination of variables may appear daunting at first, however we do it almost exclusively in the spatial dimensions. That is to say we are rarely completely aligned with rectangular objects at any one point in time. Consider a book or a filing cabinet any orientation that isn't fully a 2D rectangle, you are seeing as a linear combination of its variables, height, width, and depth. Generalizing this to arbitrary dimensionality we can project or embed a 2D profile of p-dimensional data. Its worth noting that the number of these embedded profiles, and consequently the time it takes to explore them, increase exponentially with the dimensionality of the data.

Non-linear embeddings, the compliment of the linear embedding, have also been well received recently especially with the emergence of t-Distributed stochastic neighbor embedding (Maaten and Hinton 2008). Such techniques distort the fully dimensionality on to a low, typically 2D plane. The issue with doing so is that unit of distance is not consistent with location in the embedded space, which severely hinders the interoperability of these embeddings. Additionally they often have hyperparameters that need tuning. Doing so results in completely different or contradicting embeddings. Suffice it to say we exclude their consideration for such broad application for multivariate EDA.

Additionally there are many methods for supervised data, that is, with observation classes known a priori. Linear discriminant analysis (Fisher 1936), for instance, is an example of linear projection, which creates it's components based on the mean separation and covariance shape between known classes. Because we which to extrapolate to generalized EDA of multivariate space we will not use any such supervised techniques to initialize components in the study.

In multivariate spaces, performance measures, and computational complexity are regularly compare to analogous algorithms and models. Human perception and inference from visuals is notably missing. We perform a within-participant, crowd sourced user study exploring the efficacy of 3 methods of linear embedding visualizations.

Section 2 discusses the visualization methods. Section 3 goes into the user study. The subsection 3.2 digs into the task and its evaluation. The results of the study are in section 4. Discussion is covered in section 5. An accompanying tool is discussed in section 6.

2 Background, visual methods

2.1 Linear projection notation

Consider a numeric data matrix with n observations of p variables,

$$\mathbf{X}_{[n,p]} = (\mathbf{x}_1,\ ...\ \mathbf{x}_p)$$
$$\mathbf{x}_i = (x_{1i},\ ...\ x_{ni}) \mid i \in [1, p]$$

Let $Y_{[n,d]}$ be the d-dimensional projection or embedding of $\mathbf{X}_{[n,p]}$ via matrix multiplication of a particular orthonormal basis matrix $\mathbf{B}_{[p,d]}$.

$$\mathbf{Y}_{[n,d]} = \mathbf{X}_{[n,p]}\mathbf{B}_{[p,d]} \mid \mathbf{B} \text{ is orthonormal}$$
$$\mathbf{y}_j = (y_{1j},\ ...\ y_{nj}) \mid j \in [i, d]$$

A matrix is said to be orthonormal if and only if they are 1) orthogonal, that is all column pairs are independent, having a cross product of 0, and 2) normal, each columns has a norm - Eucledian distance - of 1.

2.2 Principal component analysis

Considering that we want to explore multivariate data space, while maintaining position mapping of points. Linear combinations of variables becomes an ideal candidate. Principal component analysis (PCA) (Pearson 1901) creates new components that are linear combinations of the original variables. The creation of these variables is ordered by decreasing variation which is orthogonally constrained to all previous components. while the full dimensionality is in tact the benefit comes from the ordered nature of the components. For instance if nearly all of the variation in a data-space can be explained in the first half of its components than the complexity of viewing such a space is exponentially simplified.

2.3 Grand tours

Later, Asimov (Asimov 1985), coined data visualization *tour*, an animation of many linear projections across local changes in the basis. One of key features of the tour is the object permanence of the data points. That is to say by watching near by, orthogonally-interpolated frames one can track the relative changes of observations as variable contributions change.

Asimov originally purposed the *grand* tour. To start, several target bases are randomly selected. These target bases are then orthogonally-interpolated between with a fixed target distance between interpolation frames. The data matrix is premultiplied to the array of interpolated bases and rendered into an animation. There is no user interaction in a grand tour and the target.

2.4 Manual tours

The *manual* tour (Cook and Buja 1997; Spyrison and Cook 2020) defines its basis path by manipulating the basis contribution of a selected variable. A manipulation dimension is appended onto the projection plane, with a full contribution given to the selected variable. The target bases are then selected based on rotating this newly created manipulation space. The target bases are then similarly orthogonally-interpolated, data projected, and rendered into an animation. In order for variables to remain independent of each other the contributions of the other variables must also change, *ie.* dimension space should maintain its orthonormal structure. A key feature of the manual tour is that it affords users a way to control the variable contributions of the next target basis. This means that such manipulations can be selected and queued in advance or select on the spot for human-in-the-loop analysis (Karwowski 2006). However, this navigation is relatively time-consuming due to the huge volume of p-space (an aspect of the curse of dimensionality (Bellman 1957)) and the abstract method of steering the projection basis. It is advisable to first identify a basis of particular interest and then use a manual tour as a finer, local exploration tool to observe how the contributions of the selected variable does or does not contribute to the feature of interest.

In order 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 is allowed to change its magnitude of contribution, but not it's angle; it must move along the direction of its original contribution radius.

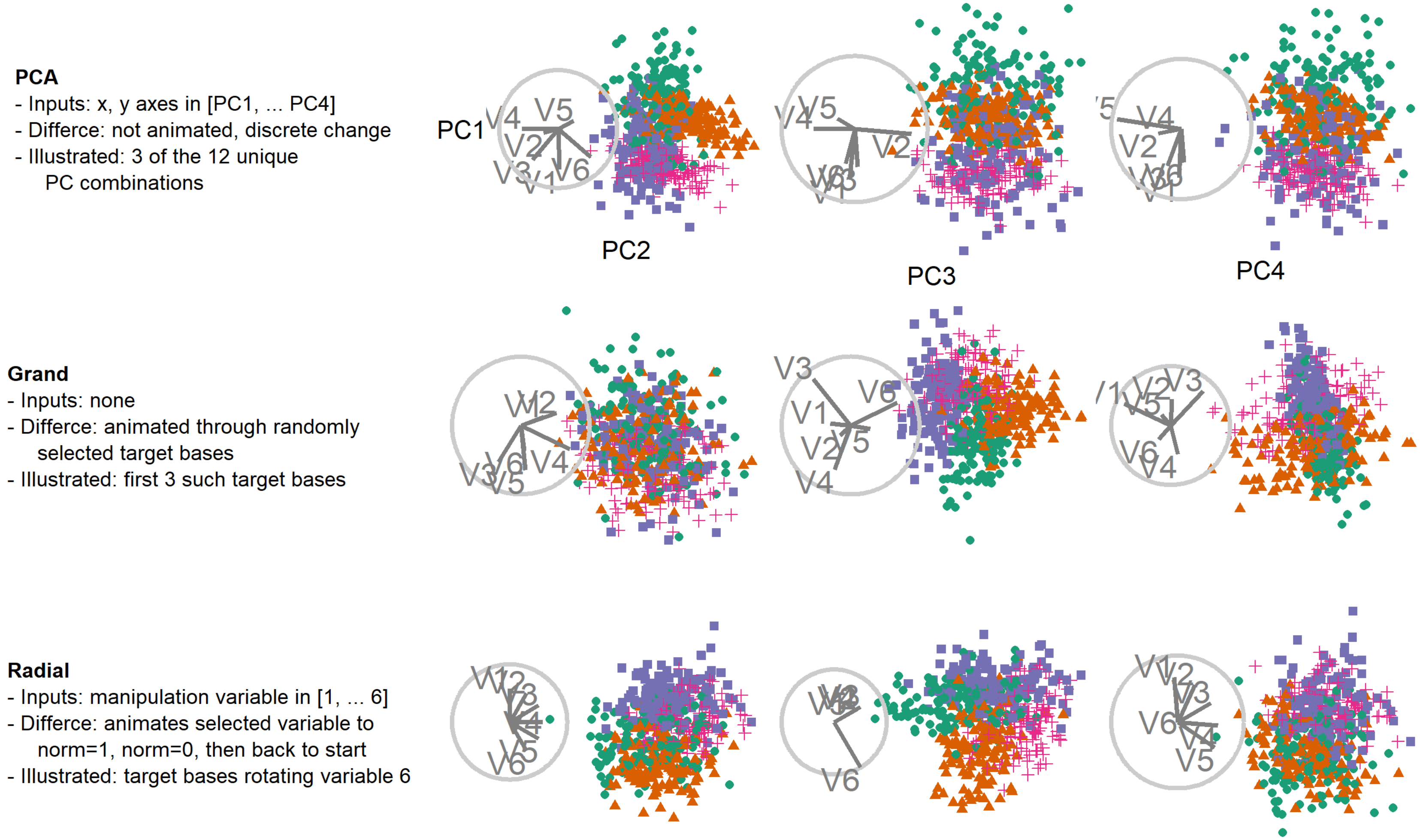


Figure 1: Example of the different visual factors. All use the same sort of biplot display to view linear projections of multivariate data. They differ in which bases are viewed which is influenced by the different factor inputs and whether or not are animated to convey the coninuity of data points from 1 frame to the next.

3 User study

3.1 Hypothesis

Does the animated removal of single variables via the radial tour improve the ability of the analyst to understand the importance of variables contribution to the separation of clusters?

H_o : visualization factor does not impact task performanceH_a : visualization factor does impact task performance

PCA will be used as a baseline for compared 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, where the object permanence of the data points, but without the ability influence it's path. Lastly the radial tour should perform best as it benefits both from animation and being able to select an individual variable to change the contribution of.

3.2 Task and evaluation

The display was a 2D scatterplot with observations supervised. Cluster membership was mapped to shape and color. There were either 3 or 4 clusters each with the number of observations within each cluster. 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 and all variable that carry more than their fair share of the weight, or 1 quarter in the case of 4 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, a visual depiction of basis, and 3) select all variables that contribute more than average in the direction of the separation in the scatterplot. Regardless of factor and block values participants were limited to 60 seconds for each evaluation of this task.

The evaluation measure of this task was designed to have a few of features: 1) the sum of squares of the individual variable marks should be 1, and 2) symmetric about 0. With these in mind we define the following measure for evaluating the task:

Let a dataset $\mathbf{X}$ be a simulation containing clusters of observations of different distributions. Let $\mathbf{X}_k$ be the subset of observations in cluster k containing the p variables.

$$\begin{aligned}
&\mathbf{X}_{[n,\ p]} = (x_1,\ ...\ x_p) \\
&\mathbf{X}_{[n_k,\ p]k} = (x_1,\ ...\ x_p) \mid n_k \in [1, n],\ \text{is an observation subset of } \mathbf{X}
\end{aligned}$$

where

$$x_{i,j,k} \text{ is scalar; the observation } i \in (1,\ ...\ n),\ \text{of variable } j \in (1,\ ...\ p),\ \text{of cluster } k \in (1,\ ...\ K)$$

We define weights, W to be a vector explaining the variable-wise difference between 2 clusters. Namely the difference of each variable between clusters, as a proportion of the total difference, less $1/p$ the amount of different each variable would hold if it were uniformly distributed.

$$\begin{aligned}
W &= \frac{(\overline{X_{j=1,k=1}} - \overline{X_{j=1,k=2}},\ ...\ (\overline{X_{j=p,k=1}} - \overline{X_{j=p,k=2}})}{\sum_{j=1}^{p}(|\overline{X_{j,k=1}} - \overline{X_{j,k=2}}|)} - \frac{1}{p} \\
&= (w_1,\ ...\ w_p)
\end{aligned}$$

Participant responses, R are a vector of logical values, whether or not participant thinks the variable separates the two clusters more than if the difference uniformly distributed. Then M is a vector of variable marks.

$$M = I(r_j) * sign(w_j) * \sqrt{|w_j|}$$
$$= (m_1, \ ... \ m_p)$$

where I is the indicator function. Then the total marks for this task is the sum of this marks vector.

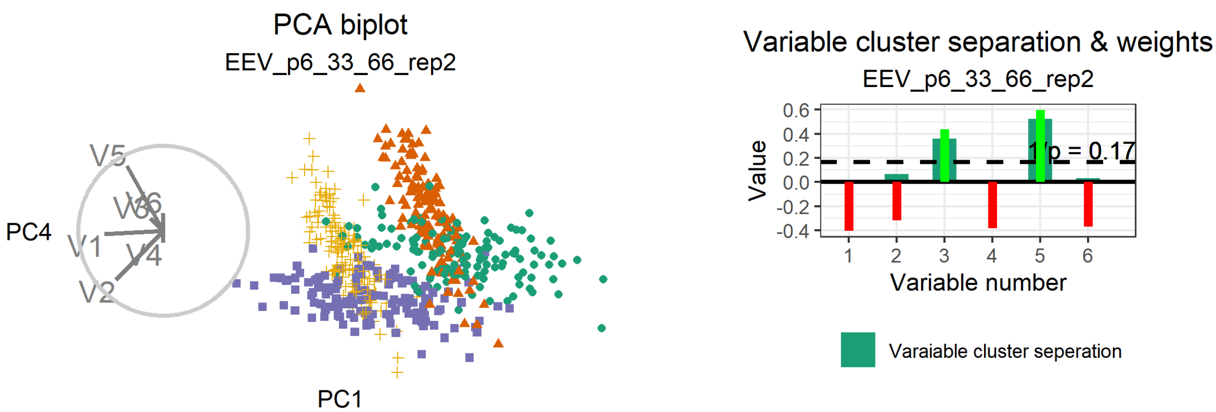


Figure 2: (L), PCA biplot of the components showing the most cluster separation with (R) illustration of the magnitude of cluster separation is for each variable (wide bar) and the weight of the marks given if a variable is selected (red/green line). The horizontal dashed line is 1 / dimensionality, the ammount of separation each variable would have if evenly detributed. The weights equal the signed square of the differce between each variable value and the dashed line.

Each of the 3 periods introduced a new factor, where participants were first able to explore an untimed task with data under the simplest parameterizations of block. The training allows the participant to become familiar with the inputs and visual specific to the factor. Upon clicking a button to proceed text containing the correct answer displays with visual still intact to explore further. After the training, participant performed 2 evaluation trials. After 60 second the display was removed, though in application, few participants elapsed this time. These evaluation trials where performed under different parameterizations as explained in section 3.4.

3.3 Factor application

Section 2 gives the sources and a description of the visual factors PCA, grand tours, and radial manual tours. The factors are tested within participant, with each factor being evaluated by each participant. The order that factors are experienced in is controlled with the block assignment as illustrated below in Figure 4. Below we cover the aesthetic standardization, as well the unique input and display within each factor.

The visualization methods were standardized wherever possible. each factor was shown as a biplot, with variable contributions displayed on a unit circle. All aesthetic values (colors, shapes, sizes, absence of legend, and absence axis titles) were held consistent. Variable contributions where always shown left of the scatterplot embeddings with their aesthetic values consistent as well. What did vary between factors were their inputs which caused a discrete jump to another pair or principal components, were absent for the grand tour with target bases to animate through selected at random, or for radial tour which variable should have its contribution animated. Key frames have of each factor have been illustrated above in Figure 1.

PCA inputs allowed for users to select between the top 4 principal components for both the x and y axis regardless of the data dimensionality (either 4 or 6). There were no user input for grand tour, users were instead shown a 15 second animation of the same randomly selected path. Users were able to view the same clip up to 4 times within the time limit. Radial tours were also displayed at 5 frames per second with in interpolation step size of 0.1 radians. Users where able to swap between variables, upon which the display would change the the start of radially increasing the contribution of the selected variable till it was full, zeroed, and then back to its initial contribution. The complete animation of any variable takes about 20 seconds, and is almost fully in the projection frame at around 6 second. 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.

3.4 Blocks and parameterization

In addition to factor we vary the data across 3 aspects: 1) The dimensionality of the data. 2) the shapes of the clusters, by changing the variance-covariance of the clusters. 3) The location of the difference between clusters, by mixing a signal and a noise variable at different ratio we control which variables contain cluster separation.

Dimensionality is tested at 2 modest levels, namely, in 4 dimensions containing 3 clusters and 6 dimensions with 4 clusters. Each cluster samples 140 observations. Each dimension is originally distributed as $\mathcal{N}(2 * I(signal), 1)$ | covariance Σ, before applying location mixing and standardizing by standard deviation). Signal variables have correlation 0.9 when they have equal orientation and -0.9 when their orientations vary. Noise variables were restricted to 0 correlation. Within each factor-period dimension is fixed with increasing difficulty, 4 then 6.

For choosing the shape of the clusters we follow the convention given by scrucca, (Scrucca et al. 2016) who named and categorize 14 variants of model families containing for 3 clusters. The name of the model family is the abbreviation of its respective volume, shape, and orientation, which are either equal or vary. We use the models EEE, EEV, and EVV, the latter is further modified by moving 4 fifths of the data out in an "V" or banana-like shape. Figure 3 shows the principal component biplot of the 3 model variants applied here. The training always uses 4 dimensions, while the 2 evaluations always contain 4 and 6 dimensions in order of increasing difficulty. The evaluation periods use EEE, EEV, and EVV-banana respectively in increasing order of difficulty.

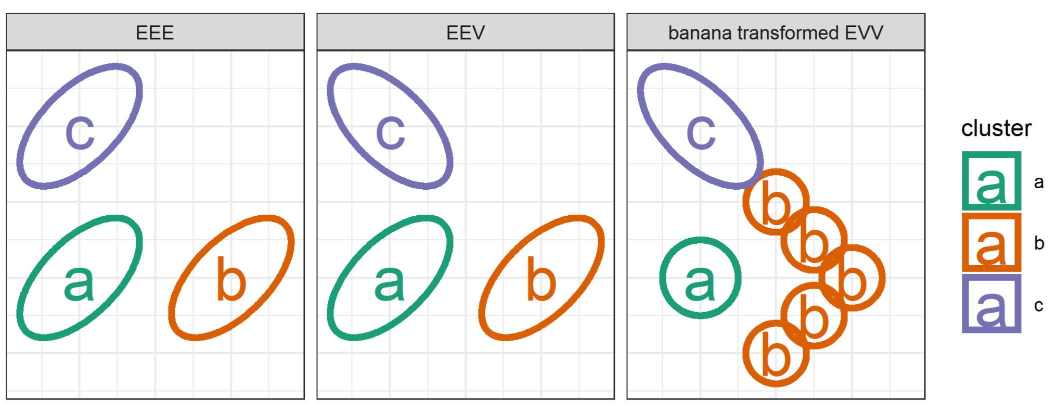
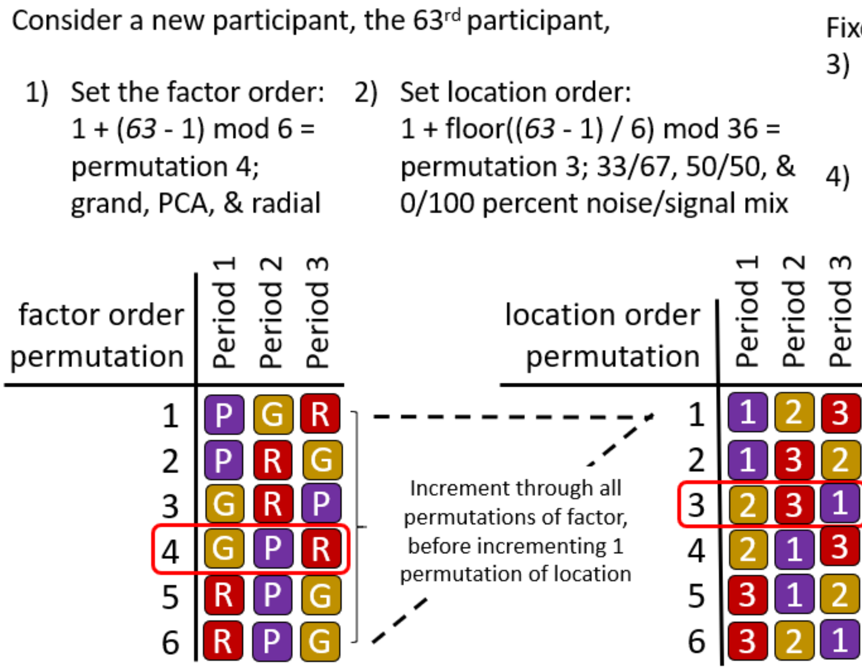


Figure 3: Ellipses of the isodensities of the model families used. Family labels are are the abbreviation for the clusters volume, shape, and orientation respectively, which are either equal or vary. We futher change the EVV model by shifting fifths of the data in banana or chevron arrow shape.

The separation of any pair of clusters is currently contained fully within a single variable at this point. In order 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). The training always uses 4 dimensions, while the 2 evaluations always contain 4 and 6 dimensions in order of increasing difficulty. The training data does not mix signal Location mixing within an evaluation period is held constant and rotated through the 6 permutations of their order. Randomizing the order of the location mixing is controlled by iterating once after each of the 6 factor order permutations are evaluated. This is illustrated in Figure 4.



Period	Evaluation order	Factor[1]	Location[2]	Shape [3]	Dimensions[4]
1	Train 1	Grand	0_1	EEE	4 (3cl)
1	1	Grand	33_67	EEE	4 (3cl)
1	2	Grand	33_67	EEE	6 (4cl)
2	Train 2	PCA	0_1	EEE	4 (3cl)
2	3	PCA	50_50	EEV	4 (3cl)
2	4	PCA	50_50	EEV	6 (4cl)
3	Train 3	Radial	0_1	EEE	4 (3cl)
3	5	Radial	0_1	Ban	4 (3cl)
3	6	Radial	0_1	Ban	6 (4cl)

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

With this setup we test the parameter space $dimension \in (4,\ 6)$, $shape \in (EEE,\ EEV,\ EVV-banana)$, $location \in (0/100,\ 33/66,\ 50/50)$ percent noise/signal mixing in order to evaluate the graphic display across the $factors \in (PCA,\ grand,\ radial)$. As we iterate though the possible permutations of these factors and location we perform an even evaluation of the full parameter space every 36 participants. Via pilot studies we estimate that 3 even block evaluations should be sufficient to identify difference between the factors; we targeted for $N = 108$ participants.

In addition to the explicitly controlled block parameters we will also be discussing each participants evaluation order regardless of factor or location experiences. This will expose an learning effect from the repetition of being expose to this data or problem. Keep in my that the fixed blocks shape and location are always experienced in order of increasing difficulty.

3.5 Post study survey

After the evaluation section of the study, participants were given a short survey containing questions gauging demographics, experience, and subjective evaluation of each factor on a 5-point Likert scale. The questions and possible responses are as follows:

Demographic:

- What are your preferred pronouns? [decline to answer, he/him, she/her, they/them or other]
- Which age group do you belong to? [decline to answer, 18 to 24, 25 to 35, 36 to 45, 45 to 60, 60 and up]
- What is your highest completed education? [decline to answer, Undergraduate degree (BA/BSc/other), Graduate degree (MA/MSc/MPhil/other), Doctorate degree (PhD/other); prolific.co participants were filtered to those stating they had an least an undergraduate degree]

Within participant bias:
Likert scale [1-5], least agreement to most agreement.

- I understand the how to perform the task.
- I am experienced with data visualization.
- I am educated in multivariate statistical analysis.

Subjective by factor:

- I was already familiar with visualization.
- I found this visualization easy to use.
- I felt confident in my answers with this visualization.
- I liked using this visualization.

The code, response files, their analyses, and study application are made publicly available at on GitHub at github.com/nspyrison/spinifex_study.

3.6 Sampling population

We recruited $N = 108$ via prolific.co (Palan and Schitter 2018). We make the assumption that interpretation of biplot displays used will not be commonly used for consumption by the general population and apply a single filter on education; that participants have completed at least an undergraduate degree (some 58,700 of the 150,400 users at the time). There is also the implicit filter that Prolific participants must be at least 18 years of age. Participants were compensated for their time at £7.50 per hour, whereas the mean duration of survey was about 16 minutes. We can't preclude previous knowledge or experience with the factors, but instead try to control for this in the user study. Figure 5 shows distributions of age and preferred pronouns of the participants that completed the post-study survey who are relatively young and well educated.

3.7 Evenness of block evaluation

From pilot studies through a sample of convenience (primarily PhD students) we predict that we wanted 3 even block evaluations to support differences in our factor and block parameterizations. Given that factor and location each have 6 permutations we targeted $N = 108 = 3*(6*6)$ evaluations before data was collected. In data collection we experience a number of adverse conditions, primarily: limited control of application server network configuration, throughput thresholds on data read/write API, and repeat attempts from users when experiencing disconnects. To mitigate this we over collect survey trials, exclude all partial trials, and remove the oldest attempts (mostly likely to experience adverse network conditions) from over evaluated permutations until we have our desired 3 even evaluations under each permutation.

4 Results

To recap, the primary response variable are marks as defined in section 3.2, while time till last response will be used as a secondary response variable. We have 2 primary data sets; the user study evaluations and post study survey. The former is contains the 108 trials with explanatory variables: *factor*, *location* of the cluster separation signal, *shape* of variance-covariance matrix, *dim*-ensionality of the data, and *order* of evaluation within the participant. Block parameterization and randomization was discussed in section 3.4. The survey was completed for 84 of these 108 trials and contains demographic information (*pronoun*, *age*, and *education*), explanatory variables (*task understanding*, *visualization experience*, and *analysis experience*),

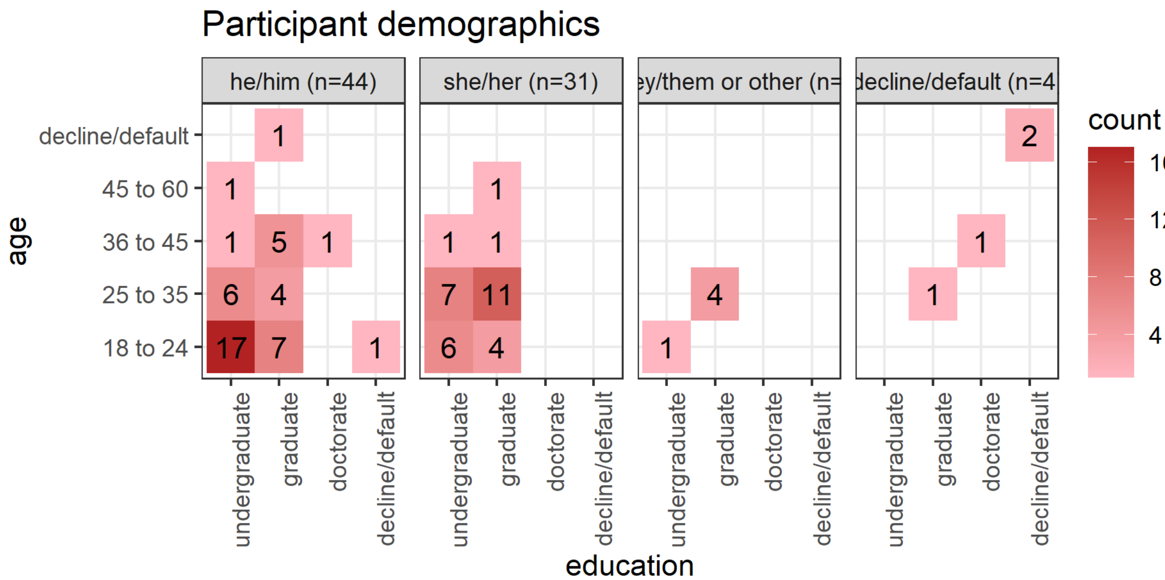


Figure 5: Heatmaps of participant demopgraphics, counts of age group by completed eduation as faceted across prefered pronoun. The average parcipant was a young adult with an undergraduate or graduate degree.

and subjective measures for each of the factors (*preference*, *familiarity*, *ease of use*, and *confidence*). The survey was covered in more detail in 3.5.

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 Random effects regression model against marks

To more thoroughly examine explanatory variables we regress against marks. 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 how adding a random effect for data and regressing against time till last response fares against benchmark models. The matrices for models with more than a few terms quickly become rank deficient; there is not enough information in the data to explain all of the effect terms. In which case the least impactful terms are dropped.

In building a set of models to test we include all single term models, a model 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, which is feasible. For comparison an overly complex model with many interaction terms is included.

Terms:	**Expanded Model:**
α	$\widehat{marks} = \mu + \alpha_i + \mathbf{Z} + \mathbf{W} + \epsilon$
$\alpha * \beta * \gamma$	$\widehat{marks} = \mu + \alpha_i * \beta_j * \gamma_k + \mathbf{Z} + \mathbf{W} + \epsilon$
$\alpha * \beta * \gamma * \delta$	$\widehat{marks} = \mu + \alpha_i + \beta_j * \gamma_k * \delta_l + \mathbf{Z} + \mathbf{W} + \epsilon$

where

Table 1: Model comparison of our random effects models regressing marks. Each model includs a random effect term of the participant, which explains the individuals influence on their marks. Complex models perform better in terms of R2 and RMSE, yet AIC and BIC penialize their large number of fixed effect in favor of the much simpler model containing only factor.

Fixed effects	No. levels	No. terms	AIC	BIC	R2 cond. (on RE)	R2 marg. (w/o RE)	RMSE
alpha	1	3	*1000*	*1027*	0.18	0.022	0.462
alpha + beta + gamma + delta	4	8	1026	1075	0.187	0.03	0.46
alpha * beta	3	9	1021	1075	0.194	0.036	0.458
alpha * beta * gamma	7	27	1064	1198	0.237	0.077	0.447
alpha * beta * gamma * delta	15	54	1125	1380	0.255	*0.115*	*0.438*

μ is the intercept of the model including the mean of random effects

$\epsilon \sim \mathcal{N}(0,\ \sigma),$ the error of the model

$\mathbf{Z} \sim \mathcal{N}(0,\ \tau),$ the random effect of participant

$\mathbf{W} \sim \mathcal{N}(0,\ \upsilon),$ the random effect of simulation

α_i, fixed term for factor | $i \in$ (pca, grand, radial)

β_j, fixed term for location | $j \in$ (0_1, 33_66, 50_50) percent noise mixing of a noise and signal variable respectively

γ_k, fixed term for shape | $k \in$ (EEE, EEV, EVV banana) mclust model family shapes described above

δ_l, fixed term for dim | $dim \in$ (4 variables with 3 cluster, 6 variables with 4 clusters)

Residual plots have no noticeable non-linear trends and contain striped patterns as an artifact from regressing on discrete variables. Figure 6 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 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 of the mean marks for participation and simulation respectively.

We also want to visually explore the conditional variables in the model. Figure XXX explores violin plots of marks by factor while faceting on the location (vertical) and shape (horizonal). Notice that the conditional variable regularly have much smaller p-value, with seemingly so much more evidence in their favor, this is because they are not "diversifying" their explanatory power across numerous interaction terms, so to speak.

4.2 Time regressing models

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. Now we repeat the same modeling procedure, namely, 1) build a battery of all additive and multiplicative models. 2) Compare their performance, reporting a some top performers. 3) Select a model to examine it's coefficients.

4.3 Subjective measures

TODO: XXX

Table 2: TODO: XXX FILL IN.

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	-0.06	0.12	47.6	-0.49	0.62	
factor						
factorgrand	-0.02	0.15	611.5	-0.14	0.89	
factorradial	0.32	0.15	601.3	2.13	0.03	*
fixed effects						
location33__66	0.11	0.17	45.4	0.65	0.52	
location50__50	-0.02	0.17	46.0	-0.13	0.90	
shapeEEV	-0.12	0.17	47.0	-0.72	0.47	
shapebanana	-0.12	0.17	46.7	-0.73	0.47	
interactions						
factorgrand:location33__66	0.12	0.21	611.8	0.57	0.57	
factorradial:location33__66	-0.47	0.21	609.2	-2.19	0.03	*
factorgrand:location50__50	0.16	0.21	612.0	0.76	0.45	
factorradial:location50__50	0.01	0.21	610.3	0.05	0.96	
factorgrand:shapeEEV	0.21	0.21	611.9	0.98	0.33	
factorradial:shapeEEV	0.35	0.21	608.6	1.62	0.11	
factorgrand:shapebanana	0.30	0.21	609.4	1.42	0.16	
factorradial:shapebanana	-0.21	0.21	606.4	-0.97	0.33	
location33__66:shapeEEV	-0.02	0.24	46.2	-0.07	0.95	
location50__50:shapeEEV	0.13	0.24	46.9	0.55	0.59	
location33__66:shapebanana	0.18	0.24	46.8	0.73	0.47	
location50__50:shapebanana	0.35	0.24	45.6	1.47	0.15	
factorgrand:location33__66:shapeEEV	0.01	0.29	603.3	0.02	0.98	
factorradial:location33__66:shapeEEV	0.01	0.29	607.7	0.04	0.96	
factorgrand:location50__50:shapeEEV	-0.10	0.29	606.9	-0.35	0.72	
factorradial:location50__50:shapeEEV	-0.41	0.29	603.9	-1.40	0.16	
factorgrand:location33__66:shapebanana	-0.53	0.29	608.7	-1.81	0.07	
factorradial:location33__66:shapebanana	0.34	0.29	606.7	1.17	0.24	
factorgrand:location50__50:shapebanana	-0.62	0.29	606.8	-2.12	0.03	*
factorradial:location50__50:shapebanana	-0.18	0.29	607.4	-0.62	0.54	

Table 3: Model comparison of our random effects models regressing marks. Each model includs a random effect term of the participant, which explains the individuals influence on their marks. Complex models perform better in terms of R2 and RMSE, yet AIC and BIC penialize their large number of fixed effect in favor of the much simpler model containing only factor.

Fixed effects	No. levels	No. terms	AIC	BIC	R2 cond. (on RE)	R2 marg. (w/o RE)	RMSE
alpha	1	3	*1000*	*1027*	0.18	0.022	0.462
alpha + beta + gamma + delta	4	8	1026	1075	0.187	0.03	0.46
alpha * beta	3	9	1021	1075	0.194	0.036	0.458
alpha * beta * gamma	7	27	1064	1198	0.237	0.077	0.447
alpha * beta * gamma * delta	15	54	1125	1380	0.255	*0.115*	*0.438*

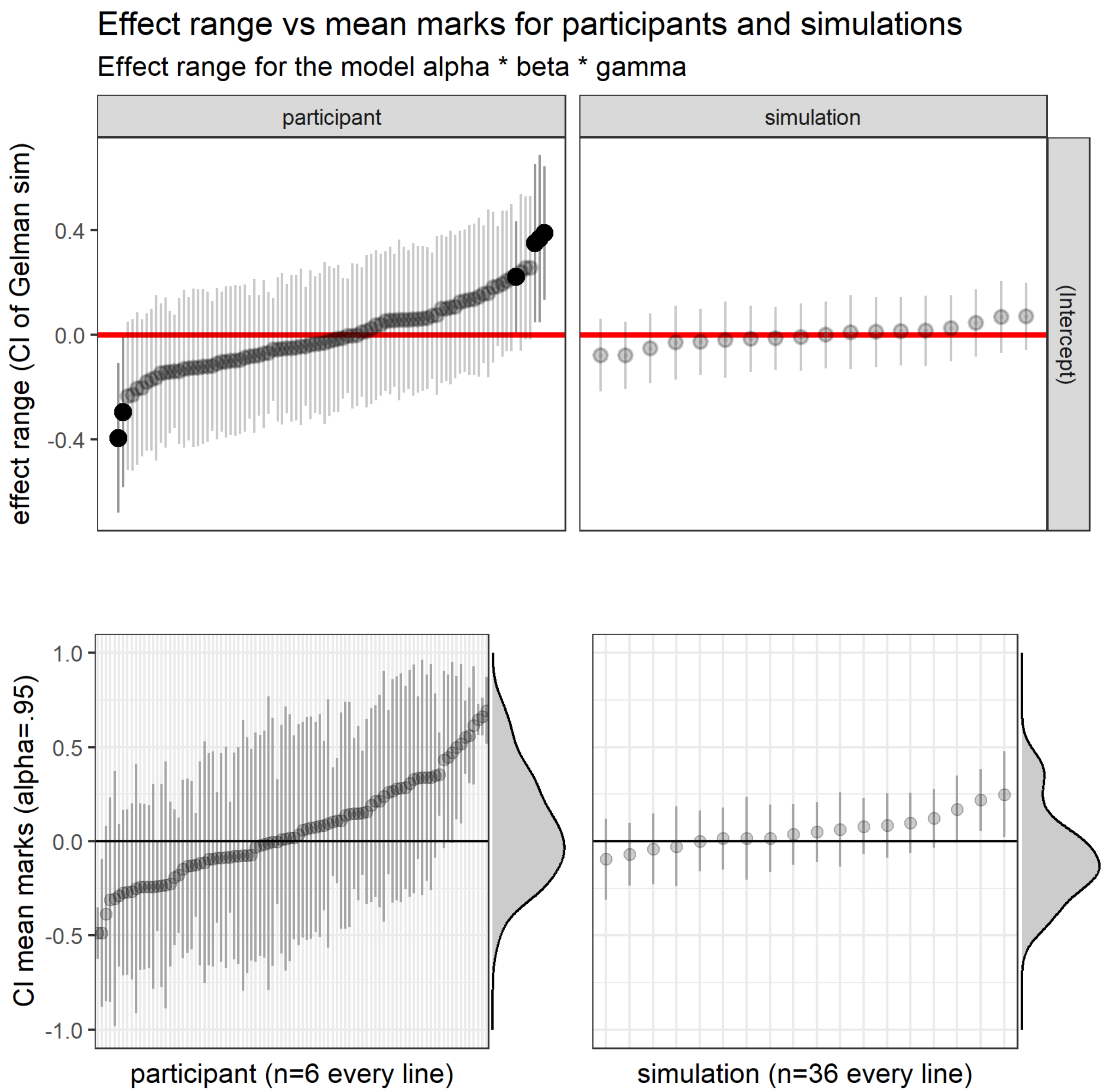


Figure 6: (T) Estimated effect ranges of the random effect terms partitcipant and data simulation of the alpha * beta * gamma model. Confidence intervals are created via Gelman simulation on the effect posterior distributions. The effect size of participant is relatively large, with several significant extrema (alpha = .95). None of the simulations deviate significantly. (B) The ordered distributions of the marks for the

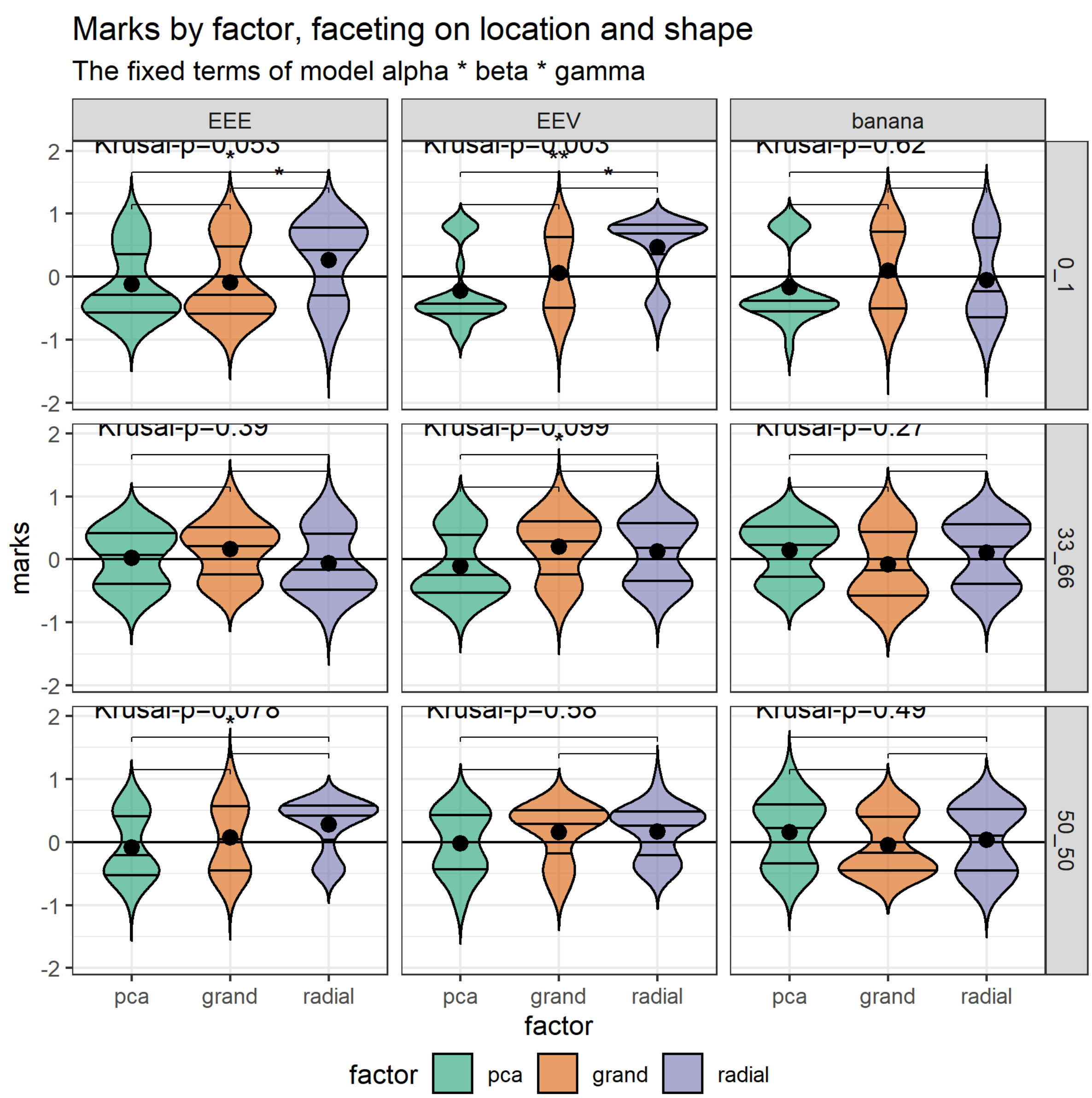


Figure 7: TODO: XXX

Table 4: TODO: XXX FILL IN.

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	28.95	3.52	51.7	8.23	0.00	***
factor						
factorgrand	-8.99	3.85	570.6	-2.34	0.02	*
factorradial	-0.51	3.99	580.9	-0.13	0.90	
fixed effects						
location33_66	-5.11	4.64	39.8	-1.10	0.28	
location50_50	-10.62	4.66	40.4	-2.28	0.03	*
shapeEEV	-10.96	4.69	41.4	-2.34	0.02	*
shapebanana	-4.63	4.71	42.0	-0.98	0.33	
interactions						
factorgrand:location33_66	11.90	5.42	567.6	2.19	0.03	*
factorradial:location33_66	-0.57	5.52	572.4	-0.10	0.92	
factorgrand:location50_50	8.69	5.39	565.7	1.61	0.11	
factorradial:location50_50	7.98	5.49	571.6	1.45	0.15	
factorgrand:shapeEEV	7.00	5.38	561.8	1.30	0.19	
factorradial:shapeEEV	0.42	5.54	573.4	0.08	0.94	
factorgrand:shapebanana	-2.23	5.51	572.7	-0.40	0.69	
factorradial:shapebanana	3.35	5.58	575.7	0.60	0.55	
location33_66:shapeEEV	3.12	6.60	40.5	0.47	0.64	
location50_50:shapeEEV	23.32	6.61	40.9	3.53	0.00	**
location33_66:shapebanana	6.90	6.62	41.2	1.04	0.30	
location50_50:shapebanana	6.35	6.56	39.8	0.97	0.34	
factorgrand:location33_66:shapeEEV	-8.58	7.32	549.4	-1.17	0.24	
factorradial:location33_66:shapeEEV	7.25	7.41	553.2	0.98	0.33	
factorgrand:location50_50:shapeEEV	-13.79	7.40	551.9	-1.86	0.06	
factorradial:location50_50:shapeEEV	-14.42	7.32	550.0	-1.97	0.05	*
factorgrand:location33_66:shapebanana	-6.70	7.44	554.2	-0.90	0.37	
factorradial:location33_66:shapebanana	-4.84	7.39	552.2	-0.65	0.51	
factorgrand:location50_50:shapebanana	-3.32	7.39	552.7	-0.45	0.65	
factorradial:location50_50:shapebanana	-9.06	7.41	552.4	-1.22	0.22	

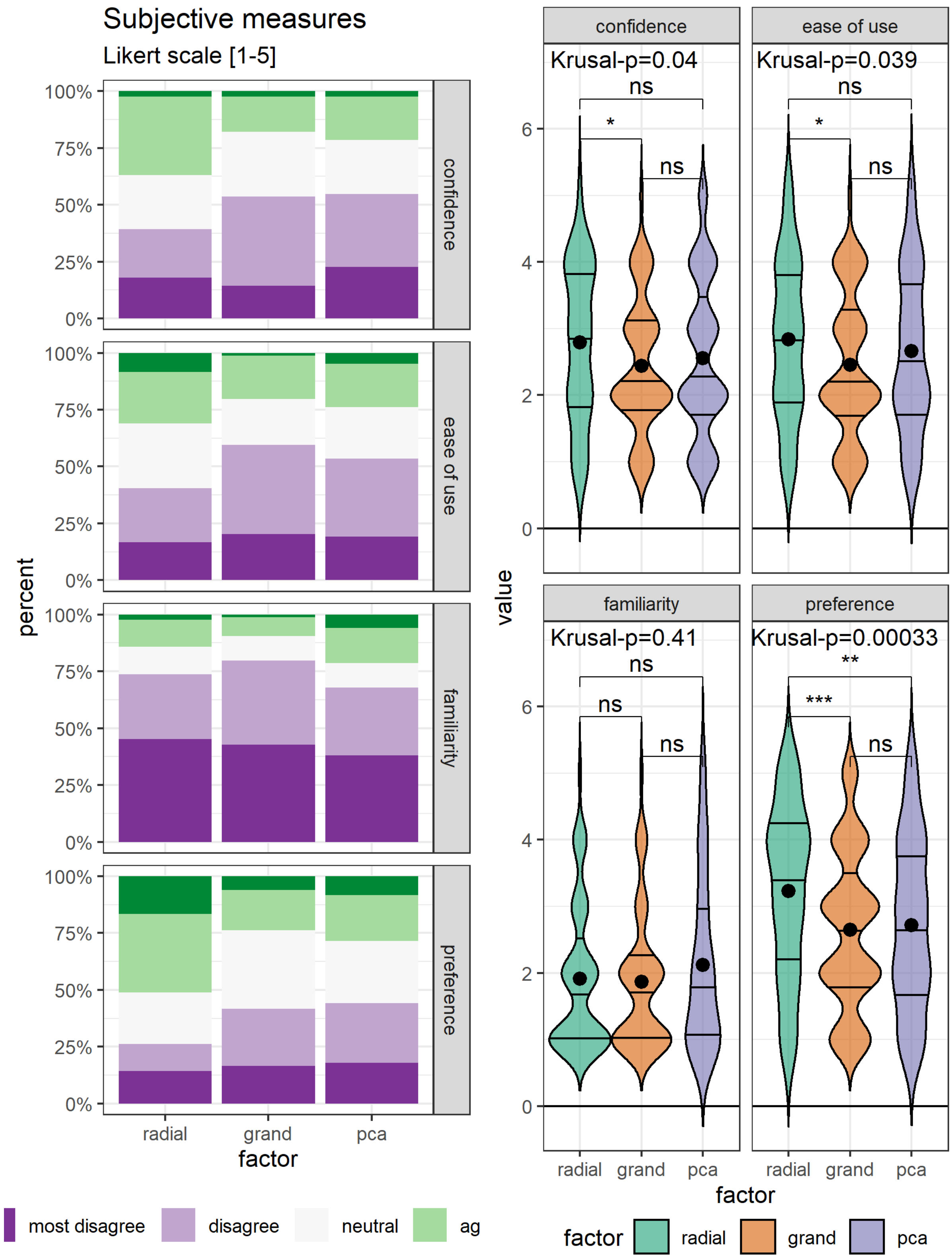


Figure 8: TODO: XXX caption needed. need significance test; Nothing is significiant using coin... go to permutation tests next?

5 Discussion

6 Accompanying tool: spinifex application

To accompany this study we have produced a more general use tool to perform such exploratory analysis of high dimensional data. The R package, `spinifex`, (Spyrison and Cook 2020) R package contains a free, open-source `shiny` (Chang et al. 2020) application. The application allows users to explore their data with either interactive or predefined manual tours without the need for any coding. Limited implementations of grand, little, and local tours are also made available. Data can be imported in .csv and .rda format, and projections or animations can be saved as .png, .gif, and .csv formats where applicable. Run the following R code for help getting started.

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), using `knitr` (Xie 2014) and `rmarkdown` (Xie, Allaire, and Grolemund 2018). The source files for this article, application, data, and analysis can be found at github.com/nspyrison/spinifex_study/. The source code for the `spinifex` package and accompanying shiny application can be found at github.com/nspyrison/spinifex/.

8 Bibliography

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

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.

Bellman, Richard Ernest. 1957. *Dynamic Programming.* Princeton University Press. https://books.google.com.au/books?id=fyVtp3EMxasC&redir_esc=y.

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.

Coleman, David. 1986. "Geometric Features of Pollen Grains." In. Statistical Computing Statistical Graphics. http://stat-computing.org/dataexpo/1986.html.

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.

Fisher, Ronald A. 1936. "The Use of Multiple Measurements in Taxonomic Problems." *Annals of Eugenics* 7 (2): 179–88. https://doi.org/10.1111/j.1469-1809.1936.tb02137.x.

Goodman, Steven. 2008. "A Dirty Dozen: Twelve p-Value Misconceptions." *Seminars in Hematology*, Interpretation of quantitative research, 45 (3): 135–40. https://doi.org/10.1053/j.seminhematol.2008.04.003.

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

Maaten, Laurens van der, and Geoffrey Hinton. 2008. "Visualizing Data Using t-SNE." *Journal of Machine Learning Research* 9: 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.* CRC press.

Ocagne, Maurice d'. 1885. *Coordonnées Parallèles Et Axiales. Méthode de Transformation géométrique Et Procédé Nouveau de Calcul Graphique déduits de La Considération Des Coordonnées Parallèles, Par Maurice d'ocagne, ...* Paris: Gauthier-Villars.

Palan, Stefan, and Christian Schitter. 2018. "Prolific. Ac-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.

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/.

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.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in r." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. http://www.crcpress.com/product/isbn/9781466561595.

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