Interactive and dynamic visualization of high-dimensional data via animated linear projections, their efficacy, and their application to local explanations of non-linear models

Nicholas Spyrison

B.Sc. Statistics, Iowa State University

Contents

CONTENTS

```
## Warning: package 'knitr' was built under R version 4.1.2
```

Copyright notice

Abstract

Visualizing data space is crucial to exploratory data analysis, checking model assumptions, and validating model performance. Yet visualization quickly becomes difficult as the dimensionality of the data or features increases. Traditionally, static, low-dimensional linear embeddings are used to mitigate this complexity. Such embedded space is a lossy approximation of the full space. However, viewing many embeddings with interactive supporting views, maximizes the information conveyed. Data visualization *tours* are a class of dynamic linear projections that animate many linear projections over small changes to the projection basis.

Tours are categorized by the path of their bases. *Manual tours* allow for User-Controlled Steering (UCS) of bases path, where the contributions of individual variables can be controlled. I create a free, open-source **R** package, **spinifex** that facilitates the creation of such tours. These manipulations of the variables can be precomputed or made in real-time. The animations are interoperable with the existing package **tourr** and can be rendered with recent graphics interfaces **plotly** and **gganimate**. The former offers interactive use with HTML widgets while the latter can render to static formats such as .gif or .mp4 files. An interactive application is included to illustrate the use of the manual tour.

Theoretically, UCS of the manual tour should enable an analyst to better explore the variable attribution to the structure identified in an embedding. I find strong evidence to support that radial tour leads to a significant and large increase in accuracy as compared with Principal Component Analysis (PCA) and an alternative tour variant, the grand tour. I conduct a within-participant, crowd-sourced user study comparing these visualization methods. Each of the $n = 108$ participants performs two trials with each visual for 648

total trials. The task is a supervised classification problem asking participants which variables attribute to the separation of 2 clusters.

Modern modeling techniques are sometimes referred to as black-box models due to the uninterpretable nature of model terms which are often many and non-linear. Recent research in Explainable Artificial Intelligence (XAI) tries to bring interpretability to these models through the use of *local explanations*. Local explanations are a class of techniques that approximate the linear-variable importance at one point in the data. I build an **cheem**, an **R** package to streamline model and local explanation preprocessing and two novel visuals. First the global view pits the data- and explanation-spaces side-by-side with a residual plot. Interactive features such as linked bushing and tooltips aid in the identification of observations to compare. The local explanation from the selected observation becomes a basis used in the creation of the radial tour. This allows us to evaluate the explanation and see explore how sensitive the explanation is to changes in the variable contribution.

Declaration

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Three publications are included in this thesis, one of which has been accepted for publication in *The R Journal* (Spyrison and Cook, 2020). The user study described in Chapter 3 will shortly be submitted to the *Journal of Data Science, Statistics, and Visualisation.* The research covered in Chapter 4 will shortly be submitted to *IEEE Transactions on Knowledge and Data Engineering.*

All of the papers in the thesis were conceptualized, developed, and written by myself, the student, while working in the Department of Human-Centered Computing under the supervision of professor Kimbal Marriott and professor Dianne Cook. When co-authors are listed, it indicates that the work was produced through active collaboration between researchers and that their contributions to team-based research have been recognized. The following table details the work, including my and my fellow co-authors' contributions.

I have renumbered and updated sections of submitted or published papers to generate a consistent presentation within the thesis.

Student name: Nicholas Spyrison

Student signature:

Thesis Chapter	Publication Title	Status (published, in press, accepted or returned for revision)	Nature and % of student contribution	Co-author name(s) Nature and % of Co-author's contribution	Co-author(s), Monash student Y/N
3	spinifex: An R Package for Creating a Manual Tour of Low-dimensional Projections of Multivariate Data	Published	75%. Concept, development, writing first draft and software development	(1) Dianne Cook, concept, methodology, writing 25%	N
4	The benefit of user-controlled radial tour for understanding variable contributions to structure in linear projections	To be submitted	80%. Concept, development, writing first draft, application, and conducting user study	(1) Dianne Cook, Concept, methodology, editing 12% (2) Kimbal Marriott, concept, methodology, editing 8%	N
5	Interrogating the linear variable importance of local explanations of black-box models with animated linear projections	To be submitted	85%. Concept, development, writing first draft and software development	(1) Dianne Cook, methodology, editing 10% (2) Kimbal Marriott, methodology, editing 5%	N

Date: 2021-12-12

The undersigned hereby certify that the above declaration correctly reflects the nature and extent of the student's co-authors' contribution to this work. In instances where I am not the responsible author I have consulted with the responsible author to agree on the respective contributions of the authors.

Main Supervisor name: Kimbal Marriott

Main Supervisor signature:

Date: 2021-12-12

Acknowledgments

I would like to express my sincere gratitude to my supervisors professor Kimbal Marriott and professor Dianne Cook for their support of my Ph.D studies and research, their subject expertise, and their careers of supervising and teaching in addition to performing research. Thank you for continuously pushing me and my research to new levels. I have enjoyed teaching data visualization which has been strongly shaped by Di's practical, data-first, visualization-often style. I will continue to imagine hearing Kim's persistent question "what did we learn from this?" as a reminder not to get lost in the details of implementation, but also to regularly step back and analyze if this is the correct aspect to be checking.

A special thanks to professor Przemyslaw Biecek for giving his input in the formulation stages of my project and his easy-to-understand explanations of complex modeling aspects. Thank you for responding to cold emails and being available for collaboration.

I thank my fellow Ph.D students, lab members, and Pomodoro partners especially on the occasional of stimulating discussions and for the positive peer pressure of knowing others are around and working. Thanks immensely to those that have empathized with me and others especially through the hard ships of studies and COVID-19. Gratitude to Ying Zhou for her enduring support though the thick of my studies and wavering mental health.

Last but not the least, I would like to thank my parents, Doug and Terry, for their support and concerns at odd hours of the day. Thanks to Alan and Claire for their companionship and support now and in our more formative years. I am looking forward to seeing you all in person shortly.

Preface

This thesis has been written using R Markdown with the bookdown package (Xie, 2016). All materials required to compile the thesis are available at <https://github.com/nspyrison/thesis_monash_phd>.

I recognize that terminology is often overburdened by ambiguous use, with several changing meanings coming from different fields. My background comes from Statistics and I will default to terms from statistics and geometry.

Chapter 2 has been published in *The R Journal*. Chapters 3 and 4 have not been submitted yet.

Chapter 1

Introduction

Exploratory data analysis (EDA) is a term that encompasses the initial summarization and visualization of a data set. This is a critical first step of checking for realistic values, finding improper data formats, and revealing insights (Tukey, 1977). Early and frequent data visualization is key to the data analyst's workflow cycle (Wickham and Grolemund, 2017). As modern datasets grow in complexity use of multivariate data is ubiquitous. The number of variables in the data increases the difficulty of creating and portraying an accurate representation of the data which is central to the iterated workflow.
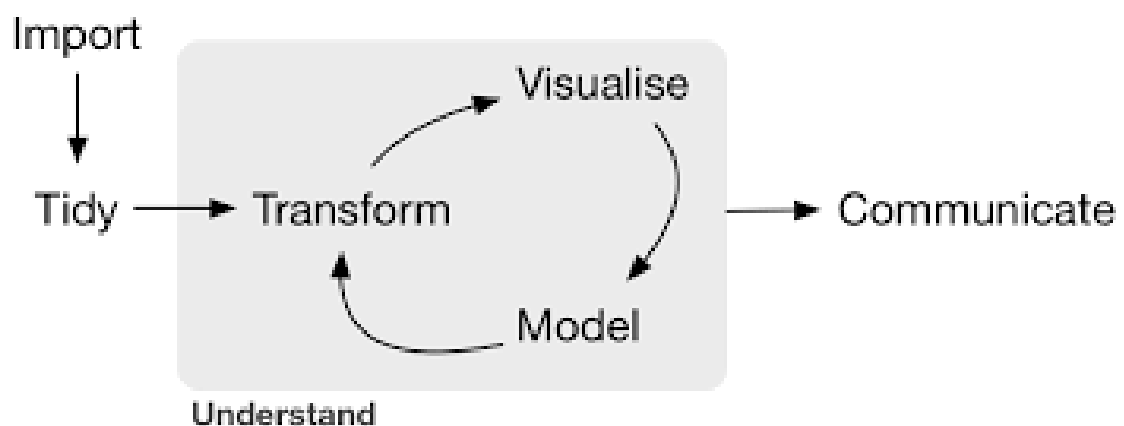

Figure 1.1: *Data analysis workflow (Wickham and Grolemund, 2017). This work focuses primary on multivariate data visualization, but also facilitates transforming data and interpretation of models.*

Early attempts at high-dimension visuals in scatterplot matrices (Chambers et al., 1983) and parallel coordinate plots (Ocagne, 1885). Principal component analysis (PCA, Pearson (1901)) reorients the basis to components, linear combinations of the original variables ordered by descending amount of variation explained. These components are typically viewed as discrete orthogonal pairs.

A data visualization *tour* (Cook et al., 2008; Lee et al., 2021) is a class of linear projections that animate small changes to the projection basis. Intuitively, they show a projection of the data object down to lower-dimensional space, in the same way that a 3D object casts a 2D shadow. A key feature of the tour is the persistence and tractability of the points through many frames. In the shadow analogy an object, such as a bar stool may be casting a circular shadow that could come from any number of shapes. However, if the stool were rotated the legs would show in the shadow quickly giving an intuitive interpretation of the object. Similarly looking at a data object over the rotation of its variables yields information about its structure. In this way and aided with linked brushing and other interactions, such continuous tours excel at extending the interpretability of multivariate spaces. Chapter 1.4 walks through some of the previous alternatives and why we ultimately continue with this class of animated linear dimension reduction.

Component spaces and tours optimize for certain objectives or selected randomly. None of the previous methods allow for an analyst to control the contributions of variables to the basis. This is crucial component for *human-in-the-loop* analysis (Karwowski, 2006). If the analyst wants to explore what happens to the structure of one variable were removed or contribution of another were increased they had no way to do so.

1.1 Research questions

Understanding variable sensitivity to the structure in a projection frame is crucial factor analysis after identifying a feature of interest. If an identified feature is the *what* of an analysis then its variable attributions are its *how*. The user interaction afforded by the radial tour should allow for a more precise exploration of this structure.

RQ 1. **How do we define user interaction for the radial tours to add and remove variables smoothly from a 2D linear projection of data?**

Neither component spaces nor grand tour provide a means for changing the contributions of a desired variable. To facilitate variable interaction, we need to have a means of manipulating the basis. This is crucial to exploring the sensitivity of the variable contributions to the structure in a frame.

RQ 2. **Does use of the radial tour improve analysts understanding of the relationship between variables and structure in a 2D linear projections?**

How can we be certain that having a means control leads to a meaningfully better analysis? A comparison of alternative methods should lend insight into which methods perform well under specific tasks.

RQ 3. **Can the radial tours be used in conjunction with the local explanation, SHAP, to improve the interpretability of black-box models?**

The tension from the trade-off between accuracy and interpretability of black-box models is rising. There is a clear need to be able to explain black-box models. After extrapolating local explanations of all observations we want to explore that space with the help of the radial tour.

1.2 Methodology

The research corresponding with RQ #1 entails *algorithm/software design* adapting the algorithm from Cook and Buja (1997). This allows for interactive control of 2D projections and serves as a foundation for the remaining work.

To address RQ #2, a controlled *experimental study* has explored the efficacy of interactive radial tours as compared with 2 benchmark methods: Principal Component Analysis (PCA, Pearson (1901)) and the grand tour (Asimov, 1985). This was a within-participant user study where each participant experienced each visual. Trials were balanced across 3 other experimental factors: location of the signal, the shape of the cluster distributions, and the dimensionality of the data. Ethics approval was obtained for this work. The approval and

disclosure forms are available at the project repository: `https://github.com/nspyrison/spinifex_study/tree/master/ethics`.

The research for RQ #3 involves *fundamental visualization design.* We know that the SHAP value is a local explanation for one observation. This SHAP value will also serve as the 1D basis for the radial tour. While using SHAP as a projection basis is novel it is not particularly insightful by itself. We provide tracking marks on the tour as well as showing the within-class distributions of the SHAP components as parallel coordinate marks on the basis. We also offer a global view and related statistics to map to color to aid in exploring the sensitivity of the SHAP-space relative to the sensitivity of the original data space.

1.3 Contributions

1. **spinifex**, a package offering a consistent framework for performing radial tours and the rendering of any tours to various formats:

   - Perform radial tours to explore the variable sensitivity to structure
   - Transform numeric variable in the data
   - Extract various bases exposing features of the data
   - Interoperable with tours generated with **tour** (Wickham et al., 2011)
   - Layered interface for producing tours that mirror the layered approach of **ggplot2** (Wickham, 2016)

2. A user study comparing the efficacy of the radial tour against the alternatives of PCA the grand tour includes:

   - Creation of supervised classification task to assess the variable attribution to the separation of clusters performed under various levels of experimental factors: location, shape, and dimensionality.
   - Definition of an accuracy measure to evaluate this task.
   - Results: strong evidence that the radial tour increases the accuracy of this task by a large amount while the task is performed fastest with the grand tour.

- Attribution of the error, by performing a mixed model regression helps to explain the source of the error term into the variability of participants skill aptitude for this task and variability of the difficultly of a simulation by random chance.

- Introduces an interactive application to preprocess data and explore with the radial tour. Users can choose from six supplied datasets or upload their own.

3. **cheem**, a package applying the radial tour to local explanations of black-box models, including:

- Preprocessing; creation of a random forest model, the extraction of all observation's tree SHAP local explanation, and extraction of other statistics for display.

- Visualization of approximations of the data- and attribution-space side-by-side with linked brushing, hover tooltips, and tabular display of selected points.

- Evaluating the local explanations with the use of the radial tour we can explore the sensitivity to the structure identified better see what variable support the explanation holds realistic.

1.4 Background

This section starts with setting the scope for the type of data we are focusing on, gives a brief motivation, then works through previous multivariate visualizations and their scalability. By the end, the focus narrows to linear projections, and in particular, the class of animated linear projections known as the tour.

For our purposes I will be focusing on the case where data X_{nxp} contains n observations of p variables, is complete with no missing values, variables are numeric (ideally not ordinal levels), and $n > p$ typically many more observations than variables. While I write as though always operating on the original variable space these methods could similarly be applied to feature decomposition of data not fitting this description. In the case of the linear projection let, $Y_{nxd} = X_{nxp} * A_{pxd}$ be the embedding of the data mapped by the basis A, where $d < p$. When p is large, say over 10 or 20 variables, the viewing space is quite large. In these cases, a PCA initialization step is commonly used where the variables are approximated as fewer principal components and this reduced space can be viewed, albeit with the disadvantage of having another linear mapping back to the original space.
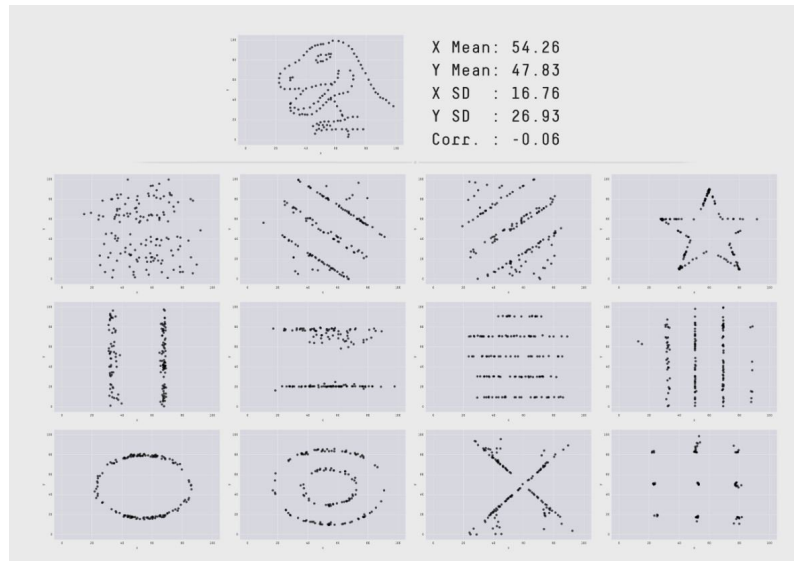
Figure 1.2: *Starting with the datasarus plot, observations are allowed to drift (by iterated simulated annealing) toward 12 patterns provided that they stay close to the originally listed statistics (Matejka and Fitzmaurice, 2017).*

Wickham, Cook, and Hofmann (2015) also view model spaces and features while urging to have a preference for visualizing data-space directly.

Visualization is much more robust than numerical summarization alone (Anscombe, 1973; Matejka and Fitzmaurice, 2017). In these studies the data have the same summary statistics, yet contains obvious visual trends and shapes that could go completely unheeded if plotting is foregone. Data visualization is fundamental to EDA and quickly evaluating data supports ensuring that models are suitable.

1.4.1 Scatterplot matrices

The work in Grinstein, Trutschl, and Cvek (2002) gives a good taxonomy of high-dimensional visualization. We will follow a few examples building up to why we conclude with tour methods. Broadly speaking, we are concerned with the question "How can an analyst visualize arbitrary $p-$ dimensions?". To illustrate some of the options for data I use the penguins data(Gorman, Williams, and Fraser, 2014, horst_palmerpenguins_2020). It contains 333 observations of 4 physical measurements for 3 species of penguins observed near Palm Station, Antarctica.

Viewing as many univariate histograms or density curves is one method. Similarly, one could look at all variable pairing as scatter plots. This forms the crux of the scatterplot matrices, also known as SPLOM (Chambers et al., 1983). In a scatterplot matrix variables are displayed across the columns and row, the diagonal elements show the univariate densities while one or both sides of the off-diagonal show scatterplot pairs. This is useful for getting a handle on the support and shapes of the variables, but is not going to scale well with dimension and is not a suitable audience-ready display as it is very busy and doesn't draw attention to any one spot. Munzner reminds us to abstract all of the cognitive work out of the visual allowing the audience to focus on seeing the evidence supporting the claim (Munzner, 2014).

1.4.2 Parallel coordinate plots

Alternatively, we could consider a class of observation-based visuals. In parallel coordinate plots (Ocagne, 1885) variables are arranged horizontally and observations are connected by lines with the height mapping to the quantile or z-value for each variable. This scales much better with dimensions, but poorly with observations. It also suffers from an asymmetry with the variable order, that is, changing the order of the variable will lead people to very different conclusions. The x-axis is also used to display variables rather than the values of the observations. This restricts the amount of information that can be gleaned between variables. Munzner asserts that position is the more human-perceptible channel for encoding information; we should like to reserve it for the values of the observations. The same issues persist across other observation-based displays such as radial variants, pixel-oriented visuals, and Chernoff faces [Keim (2000); chernoff_use_1973]. These visuals are better suited for the $n < p$ case where there are more variables than observations.

1.4.3 Dimension reduction

Ultimately, we will need to turn to dimension reduction to create a compelling visual allowing audiences to focus on features with contributions from multiple variables. Dimension reduction is separated into two categories, linear and non-linear. The linear case spans all affine mathematical transformations, essentially any mapping where parallel lines stay
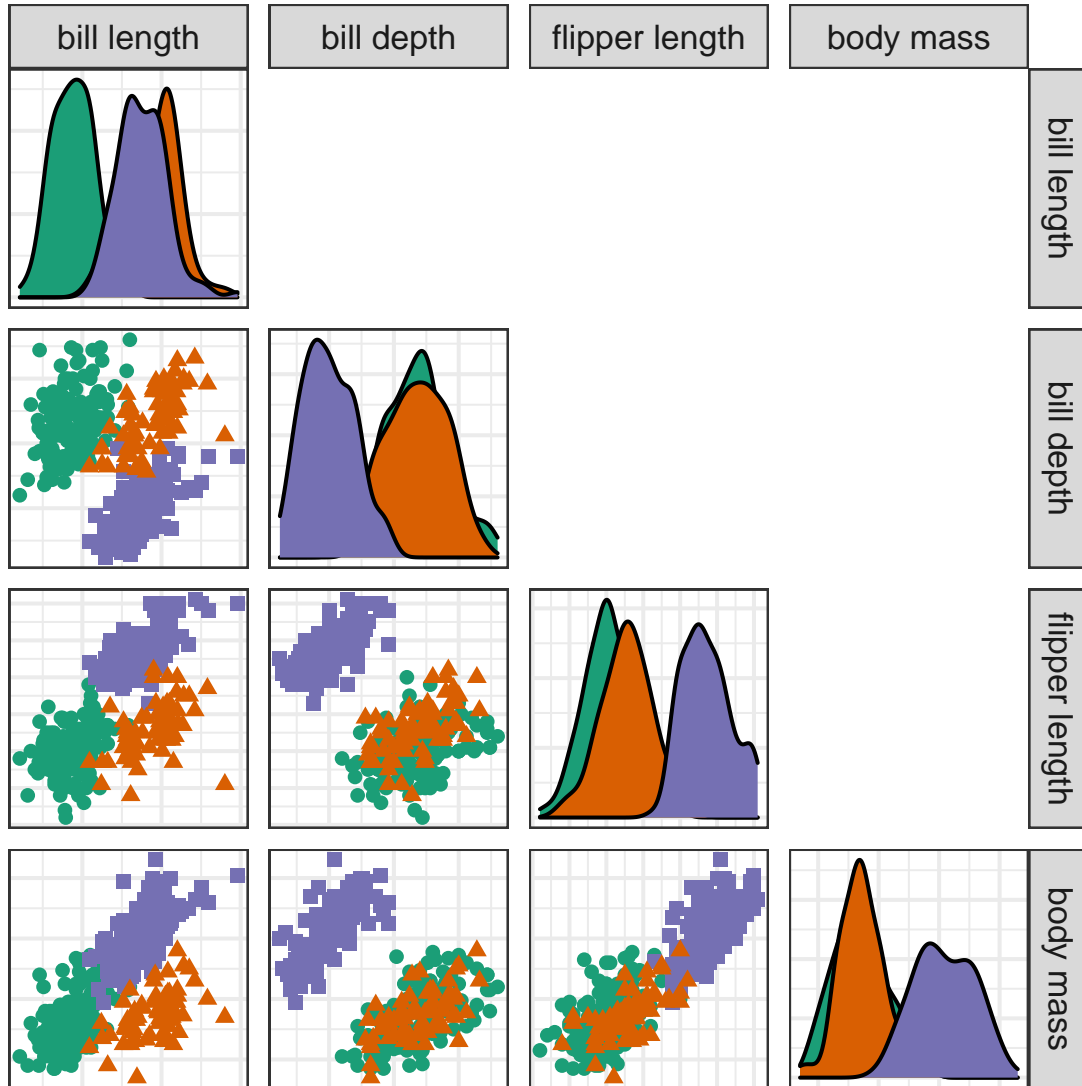
Figure 1.3: *(ref:penguinsmplom-cap)*

parallel. Non-linear transformations are the complement of that, think transformations with exponents or interacting terms. Examples in low dimensions are relatable. For instance, shadows are examples of linear projections where a 3-dimensional object casts a 2D projection, the shadow. Our vision at one instance in time is also a 2D projection. An example of a non-linear transformation is that of 2D representation of the globe. There are many different ways (and features to optimize) to distort the surface to display as a map. The most common may be rectangular displays where the area is distorted more and more with distance away from the equator. Other distortions are created when the
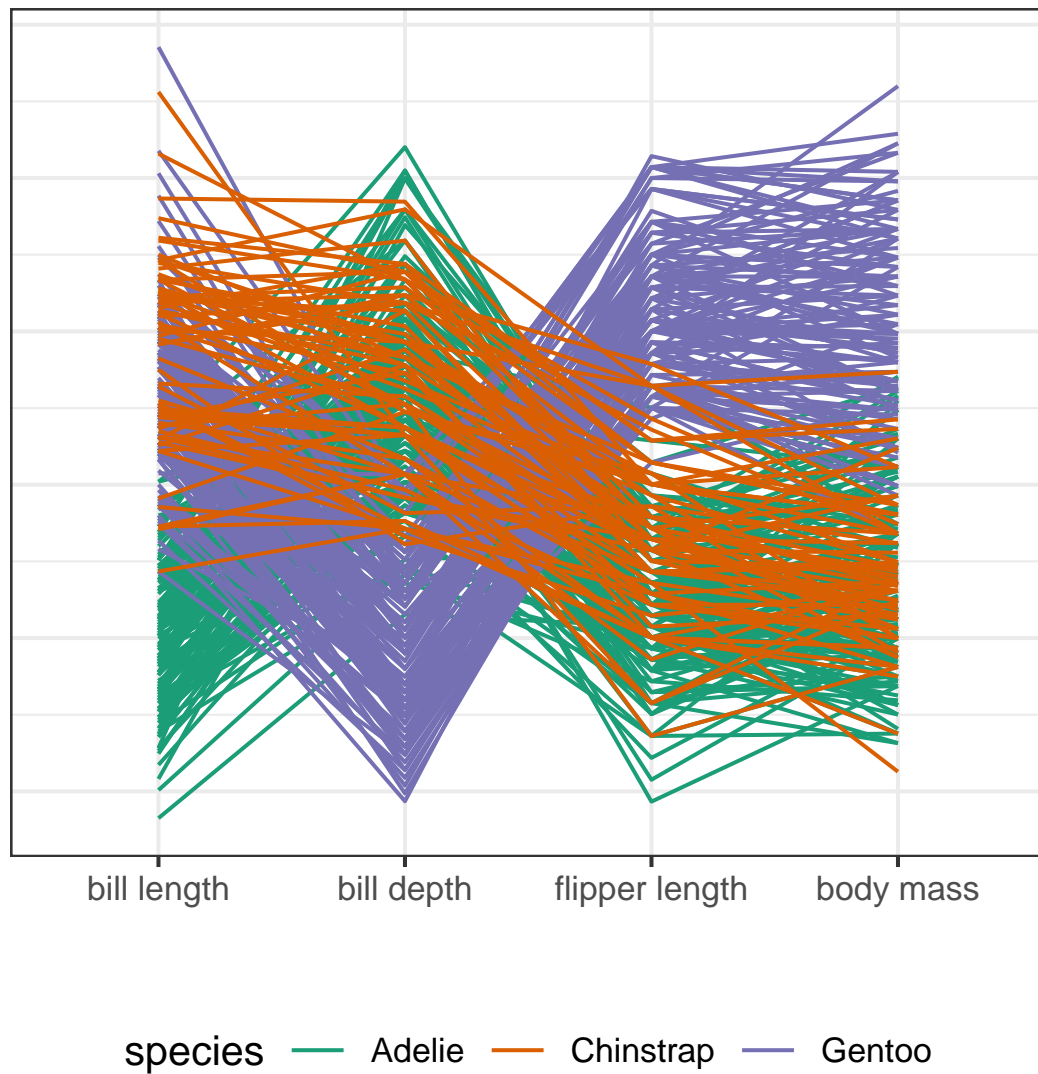
Figure 1.4: *Parallel coordinate plots of penguins data. Doesn't scale well with observations, suffers from asymmetry with the variable ordering, and horizontal position is used for variable rather than observation levels.*

surface is unwrapped as a long ellipse. Yet others create non-continuous gaps in oceans to minimize the distortion of countries.

Non-linear techniques often have hyperparameters that affect how the spaces are distorted to fit into a lower-space. To quote Anastasios Panagiotelis "All non-linear projections are wrong, but some are useful", a play on George Box quote about models. Non-linear techniques distort the space in unclear ways, and what is more, they can introduce features not in the data depending on the selection of hyperparameters. The presence of structure in a non-linear model is necessary, but not sufficient to conclude the existence of a structure in the data.

Unfortunately, there is no free lunch here. An increase in data space will lead to a $p - d$-dimensional viewing space in the linear case and an increasingly perturbed and distorted space in non-linear techniques. The intrinsic dimensionality of data is the number of variables needed to minimally represent the data (Grinstein, Trutschl, and Cvek, 2002). This is an important aspect of dimension reduction that does not have to end in visual, but is also a common part of factor analysis and preprocessing data. Consider a Psychology survey consisting of 100 questions about the Big Five personality traits. The data consists of 100 variables, while the theory would suggest the intrinsic dimensionality is five. The data likely picks up on other aspects and may better be summarized in with 8 or 10 dimensions. If this were the case, reducing the data to this space will be necessary to gate the exponentially increasing view time.

1.4.4 Tours, animated linear projections

A single linear projection is a resulting space from the data multiplied by the basis where the basis is an orthonormal matrix (the orientation of the unit origin) mapping the data space to a lower dimension. A data visualization tour animates many such projections through small changes in the basis. In the bar stool shadow analogy structural information about a hidden object was gained by watching its shadow change as a result of its rotation. An analyst similarly gains information about the data object by watching continuous changes to the basis (the rotation of the data). Originally in a *grand* tour (Asimov, 1985)
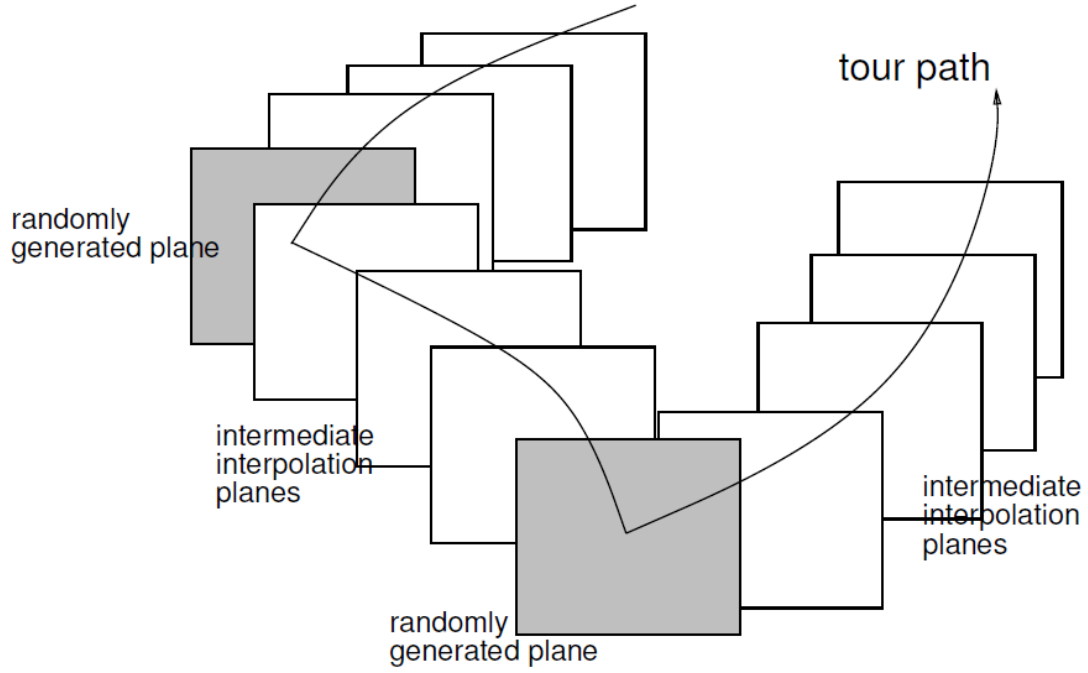
Figure 1.5: *Illustration of the grand tour selecting random target frames (grey) connected via geodesically interpolated frame (white). Figure from Buja et al. (2005).*

several target frames are randomly selected and then interpolated along their geodesic path.

There are various types of tours which are classified by the generation of their basis paths. A *guided* tour uses simulated annealing to move progressively closer to an objective function in the embedded space (Hurley and Buja, 1990). A more comprehensive discussion and review of tours can be found in the works of Cook et al. (2008) and Lee et al. (2021).

Tours are used for a couple of salient features: maintains transparency back to the original variable space and the persistence data points from frame to frame convey more information than looking at discrete jumps to other bases. Because of these features tours are a good method to extend the visualization of data space as dimensionality increases.

The work below covers manual tours (Cook and Buja, 1997; Spyrison and Cook, 2020) in Chapter 2. Radial tours are one type of manual tour that where the contribution of one variable is extended radially to a full contribution, removed completely, then restored to its original contribution. Chapter 3 compares the efficacy of the radial tour as compared

with PCA and the grand tour in a user study. Lastly, Chapter 4 extends the use of the radial tour to evaluate the local explanation of black-box models.

1.5 Thesis structure

The remainder of the thesis is organized as follows: Chapter 2 discusses the theory and implementation of the manual tour in the package **spinifex**. Chapter 3 discusses a user study evaluating the efficacy of the radial, manual tour as compared with PCA and the grand tour. There is large evidence that the use of the radial tour increases the accuracy of responses for the supervised variable-attribution task describing the separation between two variables. Chapter 4 extends the use of manual tours to increase the interpretability of non-linear models. Where manual tours explore variable sensitivity to the structure identified in local explanations of the model. Lastly, Chapter 5 concludes with some takeaways and a discussion of possible extensions.

Chapter 2

spinifex: an R package for creating user-controlled animated linear projections

TODO:XXX Segue. Bring over abstract?

2.1 Introduction

Exploring multivariate spaces is a challenging task, increasingly so as dimensionality increases. Traditionally, static low-dimensional projections are used to display multivariate data in two dimensions including principal component analysis, linear discriminant spaces or projection pursuit. These are useful for finding relationships between multiple variables, but they are limited because they show only a glimpse of the high-dimensional space. An alternative approach is to use a tour (Asimov, 1985) of dynamic linear projections to look at many different low-dimensional projections. Tours can be considered to extend the dimensionality of visualization, which is important as data and models exist in more than 3D. The package {tourr} (Wickham et al., 2011) provides a platform for generating tours. It can produce a variety of tours, each paired with a variety of possible displays. A user can make a grand, guided, little, local or frozen tour, and display the resulting projected

data as a scatterplot, density plot, histogram, or even as Chernoff faces if the projection dimension is more than 3.

This work adds a manual tour to the collection. The manual tour was described in Cook and Buja (1997) and allows a user to control the projection coefficients of a selected variable in a 2D projection. The manipulation of these coefficients allows the analyst to explore their sensitivity to the structure within the projection. As manual tours operate on only one variable at a time, they are particularly useful once a feature of interest has been identified.

One way to identify "interesting" features is with the use of a guided tour (Cook et al., 1995). Guided tours select a very specific path, which approaches a projection that optimizes an objective function. The optimization used to guide the tour is simulated annealing (Kirkpatrick, Gelatt, and Vecchi, 1983). The direct optimization of a function allows guided tours to rapidly identify interesting projection features given the relatively large parameter-space. After a projection of interest is identified, an analyst can then use the "finer brush" of the manual tour to control the contributions of individual variables to explore the sensitivity they have on the structure visible in the projection.

The paper is organized as follows. Section 2.2 describes the algorithm used to perform a radial manual tour as implemented in the package {spinifex}. Section 2.2.2 explains how to generate an animation of the manual tour using the animation frameworks offered by {plotly} (Sievert, 2020) and {gganimate} (Pedersen and Robinson, 2020). Package functionality and code usage following the order applied in the algorithm follows in section 2.3.3. Section 2.4 illustrates how this can be used for sensitivity analysis applied to multivariate data collected on high energy physics experiments (Wang et al., 2018). Section 2.5 summarizes this paper and discusses potential future directions.

2.2 Algorithm

The algorithm to conduct a manual tour interactively, by recording mouse/cursor motion, is described in detail in Cook and Buja (1997). Movement can be in any direction and magnitude, but it can also be constrained in several ways:

- *radial*: fix the direction of contribution, and allow the magnitude to change.

- *angular*: fix the magnitude, and allow the angle or direction of the contribution to vary.

- *horizontal, vertical*: allow rotation only around the horizontal or vertical axis of the current 2D projection.

The algorithm described here produces a **radial** tour as an *animation sequence*. It takes the current contribution of the chosen variable, and using rotation brings this variable fully into the projection, completely removes it, and then returns to the original position.

2.2.1 Notation

The notation used to describe the algorithm for a 2D radial manual tour is as follows:

- $\mathbf{X}$, the data, an $n \times p$ numeric matrix to be projected to 2D.

- $\mathbf{B} = (B_1, \; B_2)$, any 2D orthonormal projection basis, $p \times 2$ matrix, describing the projection from $\mathbb{R}^p \Rightarrow \mathbb{R}^2$. This is called this the "original projection" because it is the starting point for the manual tour.

- k, is the index of the variable to manipulate, called the "manip var".

- $\mathbf{e}$, a 1D basis vector of length p, with 1 in the k-th position and 0 elsewhere.

- $\mathbf{M}$ is a $p \times 3$ matrix, defining the 3D subspace where data rotation occurs and is called the manip(ulation) space.

- $\mathbf{R}$, the 3D rotation matrix, for performing unconstrained 3D rotations within the manip space, $\mathbf{M}$.

- θ, the angle of in-projection rotation, for example, on the reference axes; c_θ, s_θ are its cosine and sine.

- ϕ, the angle of out-of-projection rotation, into the manip space; c_ϕ, s_ϕ are its cosine and sine. The initial value for animation purposes is ϕ_1.

- $\mathbf{U}$, the axis of rotation for out-of-projection rotation orthogonal to $\mathbf{e}$.

- **Y**, the resulting projection of the data through the manip space, **M**, and rotation matrix, **R**.

The algorithm operates entirely on projection bases and incorporates the data only when making the projected data plots, in light of efficiency.

2.2.2 Steps

Step 0) Setup

The flea data (Lubischew (1962)), available in the {tourr} package, is used to illustrate the algorithm. The data contains 74 observations on 6 variables, which are physical measurements made on flea beetles. Each observation belongs to one of three species.

An initial 2D projection basis must be provided. A suggested way to start is to identify an interesting projection using a projection pursuit guided tour. Here the holes index is used to find a 2D projection of the flea data, which shows three separated species groups. Figure 2.1 shows the initial projection of the data. The left panel displays the projection basis (**B**) and can be used as a visual guide of the magnitude and direction that each variable contributes to the projection. The right panel shows the projected data, $\mathbf{Y}_{[n, \ 2]} = \mathbf{X}_{[n, \ p]}\mathbf{B}_{[p, \ 2]}$. The color and shape of points are mapped to the flea species. This plot is made using the `view\_basis()` function in {spinifex}, which generates a {ggplot2} (Wickham, 2016) object.

Step 1) Choose manip variable

In figure 2.1 the contribution of the variables tars1 and aede2 mostly contrast the contribution of the other four variables. These two variables combined contribute in the direction of the projection where the purple cluster is separated from the other two clusters. The variable aede2 is selected as the manip var, the variable to be controlled in the tour. The question that will be explored is: how important is this variable to the separation of the clusters in this projection?
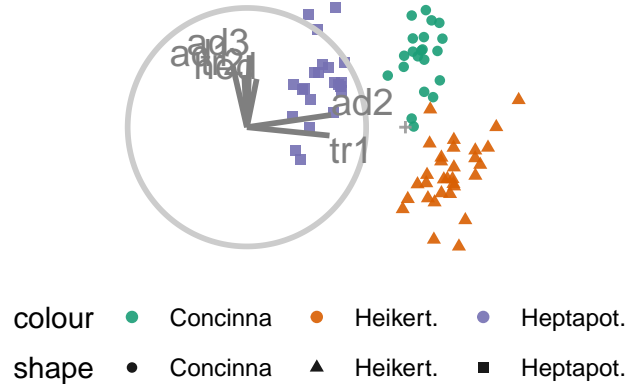
| colour | ● Concinna | ● Heikert. | ● Heptapot. |
| shape | ● Concinna | ▲ Heikert. | ■ Heptapot. |

Figure 2.1: *Biplot of the initial 2D projection: representation of the basis (left) and resulting data projection (right) of standardized flea data. The color and shape of data points are mapped to the species of flea beetle. The basis was produced by a projection pursuit guided tour, with the holes index. The contribution of the variables aede2 and tars1 approximately contrasts the other variables. The visible structure in the projection are the three clusters corresponding to the three species.*

Step 2) Create the 3D manip space

Initialize the coordinate basis vector as a zero vector, $\mathbf{e}$, of length p, and set the k-th element to 1. In the example data, aede2 is the fifth variable in the data, so $k = 5$, set $e_5 = 1$. Use a Gram-Schmidt process to orthonormalize the coordinate basis vector on the original 2D projection to describe a 3D manip space, $\mathbf{M}$.

$$e_k \leftarrow 1$$

$$\mathbf{e}^*_{[p,\ 1]} = \mathbf{e} - \langle \mathbf{e}, \mathbf{B}_1 \rangle \mathbf{B}_1 - \langle \mathbf{e}, \mathbf{B}_2 \rangle \mathbf{B}_2$$

$$\mathbf{M}_{[p,\ 3]} = (\mathbf{B}_1, \mathbf{B}_2, \mathbf{e}^*)$$

The manip space provides a 3D projection from p-dimensional space, where the coefficient of the manip var can range completely between $[0, 1]$. This 3D space serves as the medium to rotate the projection basis relative to the selected manipulation variable. Figure 2.2 illustrates this 3D manip space with the manip var highlighted. This representation is produced by calling the `view\_manip\_space()` function. This diagram is purely used to help explain the algorithm.
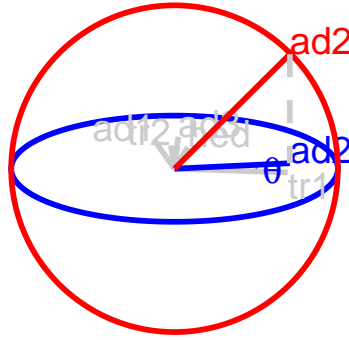
25

Figure 2.2: *Illustration of a 3D manip space, the projection plane is shown as blue circle extending into and out of the display. A manipulation direction is initalized, the red circle, orthogonal to the projection plane. This allows the selected variable, aede2, a means to change it's contribution back to the projection plane. The other variables contributions rotate into this space as well, preserving the orthogonal structure, but are omitted in the manipulation dimension for simplicity.*

Step 3) Defining a 3D rotation

The basis vector corresponding to the manip var (red line in Figure 2.2), can be operated like a lever anchored to the origin. This is the process of the manual control, that rotates the manip variable into and out of the 2D projection (Figure 2.3). As the variable contribution is controlled, the manip space rotates, and the projection onto the horizontal projection plane correspondingly changes. This is a manual tour. Generating a sequence of values for the rotation angles produces a path for the rotation of the manip space.

For a radial tour, fix θ, the angle describing rotation within the projection plane, and compute a sequence for ϕ, defining movement out of the plane. This will change ϕ from the initial value, ϕ_1, the angle between **e** and its shadow in **B**, to a maximum of 0 (manip var fully in projection), then to a minimum of $\pi/2$ (manip var out of projection), before returning to ϕ_1.

Rotations in 3D can be defined by the axes they pivot on. Rotation within the projection, θ, is rotation around the Z axis. Out-of-projection rotation, ϕ, is the rotation around an axis on the XY plane, **U**, orthogonal to **e**. Given these axes, the rotation matrix, **R** can be written as follows, using Rodrigues' rotation formula (originally published in Rodrigues (1840)):

$$\mathbf{R}_{[3,\,3]} = \mathbf{I}_3 + s_\phi \mathbf{U} + (1 - c_\phi)\mathbf{U}^2$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 & c_\theta s_\phi \\ 0 & 0 & s_\theta s_\phi \\ -c_\theta s_\phi & -s_\theta s_\phi & 0 \end{bmatrix} + \begin{bmatrix} -c_\theta(1 - c_\phi) & s_\theta^2(1 - c_\phi) & 0 \\ -c_\theta s_\theta(1 - c_\phi) & -s_\theta^2(1 - c_\phi) & 0 \\ 0 & 0 & c_\phi - 1 \end{bmatrix}$$

$$= \begin{bmatrix} c_\theta^2 c_\phi + s_\theta^2 & -c_\theta s_\theta(1 - c_\phi) & -c_\theta s_\phi \\ -c_\theta s_\theta(1 - c_\phi) & s_\theta^2 c_\phi + c_\theta^2 & -s_\theta s_\phi \\ c_\theta s_\phi & s_\theta s_\phi & c_\phi \end{bmatrix}$$

where

$$\mathbf{U} = (u_x, u_y, u_z) = (s_\theta, -c_\theta, 0)$$

$$= \begin{bmatrix} 0 & -u_z & u_y \\ u_z & 0 & -u_x \\ -u_y & u_x & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & -c_\theta \\ 0 & 0 & -s_\theta \\ c_\theta & s_\theta & 0 \end{bmatrix}$$

Step 4) Creating an animation of the radial rotation

The steps outlined above can be used to create any arbitrary rotation in the manip space. To use these for sensitivity analysis, the radial rotation is built into an animation where the manip var is rotated fully into the projection, completely out, and then back to the initial value. This involves allowing ϕ to vary between 0 and $\pi/2$, call the steps ϕ_i.

1. Set initial value of ϕ_1 and θ: $\phi_1 = \cos^{-1}\sqrt{B_{k1}^2 + B_{k2}^2}$, $\theta = \tan^{-1}\frac{B_{k2}}{B_{k1}}$. Where ϕ_1 is the angle between $\mathbf{e}$ and its shadow in $\mathbf{B}$.

2. Set an angle increment (Δ_ϕ) that sets the step size for the animation, to rotate the manip var into and out of the projection. (Note: Using angle increment, rather than
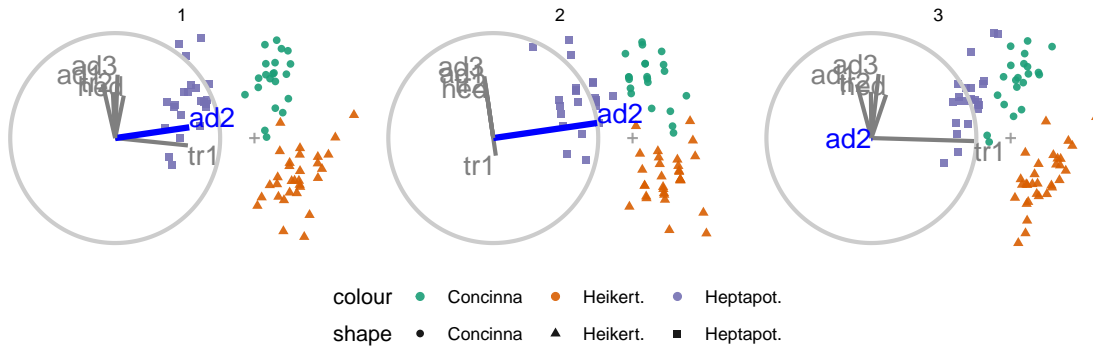
Figure 2.3: *Select frames highlight the animation of a radial manual tour manipulating aede2: (1) original projection, (2) full contribution, (3) zero contribution, before returning to the original original contribution.*

a number of steps, to control the movement, is consistent with the tour algorithm as implemented in the {tourr}).

3. Step towards 0, where the manip var is completely in the projection plane.

4. Step towards $\pi/2$, where the manip variable has no contribution to the projection.

5. Step back to ϕ_1.

In each of the steps 3-5, a small step may be added to ensure that the endpoints of ϕ (0, $\pi/2$, ϕ_1) are reached.

Step 5) Projecting the data

The operation of a manual tour is defined on the projection bases. Only when the data plot needs to be made is the data projected into the relevant basis.

$$\mathbf{Y}_{[n,\ 3]}^{(i)} = \mathbf{X}_{[n,\ p]}\mathbf{M}_{[p,\ 3]}\mathbf{R}_{[3,3]}^{(i)}$$

where $\mathbf{R}_{[3,3]}^{(i)}$ is the incremental rotation matrix, using ϕ_i. To make the data plot, use the first two columns of $\mathbf{Y}$. Show the projected data for each frame in sequence to form an animation.

Tours are typically viewed as an animation. The animation of this tour can be viewed online at https://github.com/nspyrison/spinifex_paper/blob/master/paper/gifs/flea_radialtour_mvar5.gif. The page may take a moment to load. Animations can be produced using the function play_manual_tour().

2.3 Package structure and functionality

This section describes the functions available in the package, and how to use them.

2.3.1 Installation

The {spinifex} is available from CRAN, and can be installed by:

```r
# Setup:
install.package("spinifex") ## Install from CRAN
library("spinifex") ## Read into session


# Getting started:
## Shiny app for visualizing basic application
run_app("intro")
## View the code vignette
vignette("spinifex_vignette")
```

2.3.2 Functions

Table 2.1 lists the primary functions and their purpose. These are grouped into four types: construction for building a tour path, render to make the plot objects, animation for running the animation, and specialty for providing illustrations used in the algorithm description.

29

Table 2.1: *Summary of available functions.*

Type	Function	Description
construction	create_manip_space	forms the 3D space of rotation
construction	rotate_manip_space	performs 3D rotation
construction	manual_tour	generates sequence of 2D frames
render	array2df	turn the tour path array into long form, for plotting
render	render_	render long form as a ggplot2 objection for animation
render	render_plotly	render the animation as a plotly object (default)
render	render_gganimate	render the animation as a gganimate object
animation	play_tour_path	composite function animating the specified tour path
animation	play_manual_tour	composite function animating the specified manual tour
specialty	print_manip_space	table of the rotated basis and manip space
specialty	oblique_frame	display the reference axes of a given basis
specialty	view_manip_space	illustrative display of any manip space

2.3.3 Usage

Using the flea data from the {tourr} package, we will illustrate generating a manual tour to explore the sensitivity of the cluster structure is to the variable aede2.

The `play\_manual\_tour()` function is a composite function handling interaction between `manual\_tour()`, `array2df()`, and `render\_plotly()`. This will generate an html animation using plotly. Switching the renderer to `render\_gganimate()` alternatively creates an animated gif. Each of these formats allows for the animation to be made available on a web site, or directly visible in an html formatted document.

2.3.4 Making illustrations

The function `oblique_frame` can be used to show a projection of the basis, or with the data overlaid. For example, the plots in Figures 2.1 and 2.3 were made with code similar to this:

An illustration of the manip space (as shown in Figure 2.2) is made with the `view_manip_space` function, as follows:

```
## Displays the projection plane and manipulation space for the
view_manip_space(basis = f_basis,
                 manip_var = f_mvar,
                 lab = colnames(f_data))
```

2.4 Use cases

Wang et al. (2018) introduces a new tool, PDFSense, to visualize the sensitivity of hadronic experiments to nucleon structure. The parameter-space of these experiments lies in 56 dimensions, $\delta \in \mathbb{R}^{56}$, and are visualized as 3D subspaces of the 10 first principal components in linear (PCA) and non-linear (t-SNE) embeddings.

Cook, Laa, and Valencia (2018) illustrates how to learn more about the structures using a grand tour. Tours can better resolve the shape of clusters, intra-cluster detail, and better outlier detection than PDFSense & TFEP (TensorFlow embedded projections) or traditional static embeddings. This example builds from here, illustrating how the manual tour can be used to examine the sensitivity of structure in a projection to different parameters. The specific 2D projections passed to the manual tour were provided in their work.

The data has a hierarchical structure with top-level clusters; DIS, VBP, and jet. Each cluster is a particular class of experiments, each with many experimental datasets which, each have many observations of their own. In consideration of data density, we conduct manual tours on subsets of the DIS and jet clusters. This explores the sensitivity of the structure to each of the variables in turn and we present the subjectively best and worst variable to manipulate for identifying dimensionality of the clusters and describing the span of the clusters.

2.4.1 Jet cluster

The jet cluster resides in a smaller dimensionality than the full set of experiments with four principal components explaining 95% of the variation in the cluster (Cook, Laa, and Valencia, 2018). The data within this 4D embedding is further subsetted, to ATLAS7old
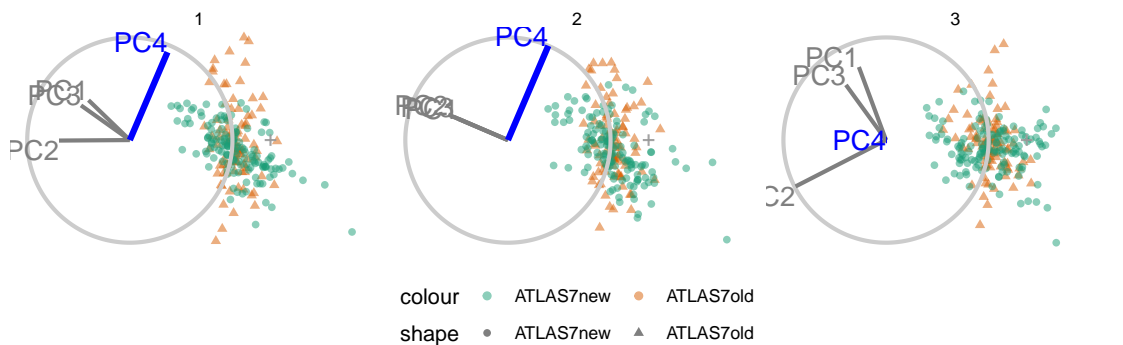
Figure 2.4: *Select frames from a radial tour of PC4 within the jet cluster, with color indicating experiment type: ATLAS7new (green) and ATLAS7old (orange). When PC4 is removed from the projection (frame 10) there is little difference between the clusters, suggesting that PC4 is important for distinguishing the experiments.*

and ATLAS7new, to focus on two groups that occupy different parts of the subspace. Radial manual tours controlling contributions from PC4 and PC3 are shown in Figures 2.4 and 2.5, respectively. The difference in shape can be interpreted as the experiments probing different phase-spaces. Back-transforming the principal components to the original variables can be done for a more detailed interpretation.

When PC4 is removed from the projection (Figure 2.4) the difference between the two groups is removed, indicating that it is important for distinguishing experiments. However, removing PC3 from the projection (Figure 2.5) does not affect the structure, indicating it is not important for distinguishing experiments. Animations for the remaining PCs can be viewed at the following links: PC1, PC2, PC3, and PC4. It can be seen that only PC4 is important for viewing the difference in these two experiments.

2.4.2 DIS cluster

Following Cook, Laa, and Valencia (2018), to explore the DIS cluster, PCA is recomputed and the first six principal components, explaining 48% of the full sample variation, are
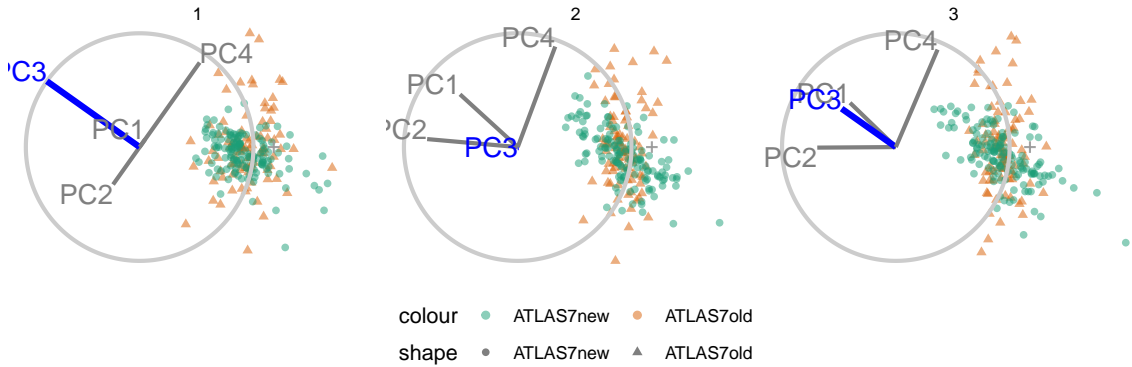
Figure 2.5: *Frames from the radial tour manipulating PC3 within the jet cluster, with color indicating experiment type: ATLAS7new (green) and ATLAS7old (orange). When the contribution from PC3 is changed there is little change in the separation of the clusters, suggesting that PC3 is not important for distinguishing the experiments.*

used. The contributions of PC6 and PC2 are explored in Figures 2.6 and 2.7, respectively. Three experiments are examined: DIS HERA1+2 (green), dimuon SIDIS (purple), and charm SIDIS (orange).

Both PC2 and PC6 contribute to the projection similarly. When PC6 is rotated into the projection, variation in the DIS HERA1+2 is greatly reduced. When PC2 is removed from the projection, dimuon SIDIS becomes more clearly distinct. Even though both variables contribute similarly to the original projection, their contributions have quite different effects on the structure of each cluster, and the distinction between clusters. Animations of all of the principal components can be viewed from the links: PC1, PC2, PC3, PC4, PC5, and PC6.

Figure 2.6: *Select frames from a radial tour exploring the sensitivity that PC6 has on the structure of the DIS cluster, with color indicating experiment type: DIS HERA1+2 (green), dimuon SIDIS (purple), and charm SIDIS (orange). DIS HERA1+2 is distributed in a cross-shaped plane, charm SIDIS occupies the center of this cross, and dimuon SIDIS is a linear cluster crossing DIS HERA1+2. As the contribution of PC6 is increased, DIS HERA1+2 becomes almost singular in one direction (frame 5), indicating that this cluster has very little variability in the direction of PC6.*
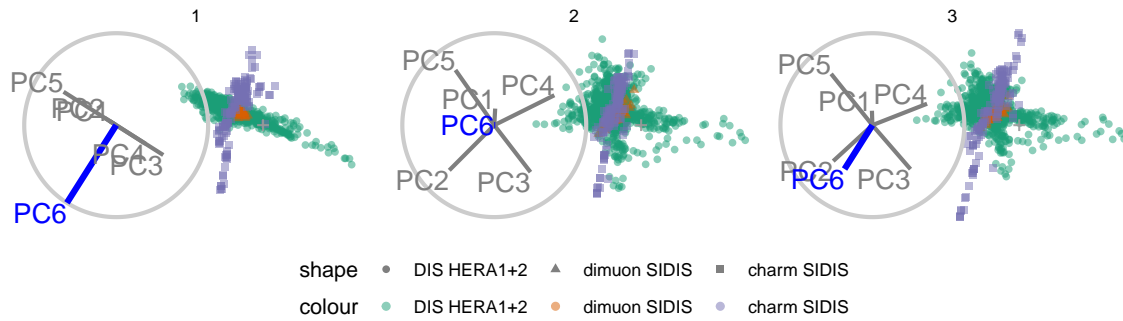
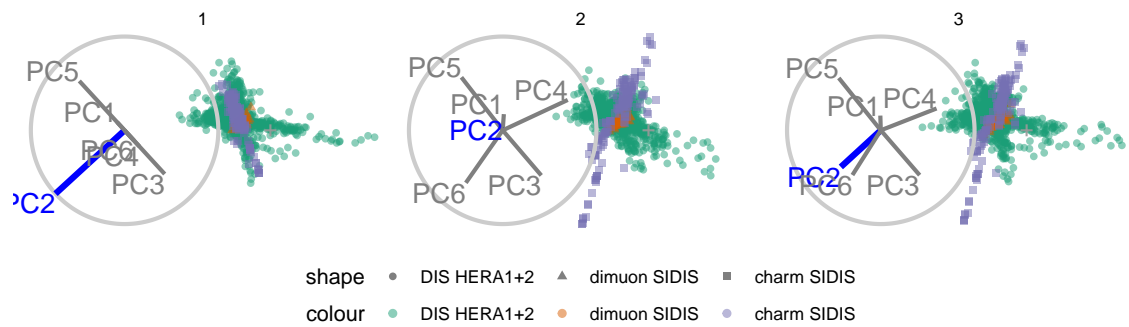**Figure 2.7:** *Frames from the radial tour exploring the sensitivity PC2 to the structure of the DIS cluster, with color indicating experiment type: DIS HERA1+2 (green), dimuon SIDIS (purple), and charm SIDIS (orange). As contribution from PC2 is decreased, dimuon SIDIS becomes more distinguishable from the other two clusters, indicating that in its absence PC2 is important for separating this cluster from the others.*

2.5 Discussion

Dynamic linear projections of numeric multivariate data, tours, play an important role in data visualization; they extend the dimensionality of visuals to peek into high-dimensional data and parameter spaces. This research has taken the manual tour algorithm, specifically the radial rotation, used in GGobi (Swayne et al., 2003) to interactively rotate a variable into or out of a 2D projection, and modified it to create an animation that performs the same task. It is most useful for examining the importance of variables, and how the structure in the projection is sensitive or not to specific variables. This functionality available in package {spinifex}. The work complements the methods available in the {tourr} package.

This work was motivated by problems in physics, and thus the usage was illustrated on data comparing experiments of hadronic collisions, to explore the sensitivity of cluster structure to different principal components. These tools can be applied quite broadly to many multivariate data analysis problems.

The manual tour is constrained in the sense that the effect of one variable is dependent on the contributions of other variables in the manip space. However, this can be useful to simplify a projection by removing variables without affecting the visible structure. Defining a manual rotation in high dimensions is possible using Givens rotations and Householder reflections as outlined in Buja et al. (2005). This would provide more flexible manual rotation, but more difficult for a user because they have the choice (too much choice) of which directions to move.

Another future research topic could be to extend the algorithm for use on 3D projections. With the current popularity and availability of 3D virtual displays, this may benefit the detection and understanding of the higher dimensional structure, or enable the examination of functions.

Having a graphical user interface would be useful for making it easier and more accessible to a general audience. This is possible to implement using {shiny} (Chang et al., 2020). The primary purposes of the interface would be to allow the user to interactively change the manip variable easily, and the interpolation step for more or less detailed views.

2.6 Acknowledgments

This article was created in R, using {knitr} (Xie, 2020) and {rmarkdown} (Allaire et al., 2020), with code generating the examples inline. The source files for this article be found at github.com/nspyrison/spinifex_paper/. The animated gifs can also be viewed at this site, and also in the supplementary material for this paper. The source code for the {spinifex} package can be found at github.com/nspyrison/spinifex/.

Chapter 3

The benefit of user-controlled radial tour for understanding variable contributions to structure in linear projections

THIS CHAPTER IS CURRENTLY BEING EDITED IN ITs OWN REPOSITORY, THIS VERSION IS OUT OF DATE.

TODO:XXX Segue. Bring over abstract?

3.1 Introduction

Multivariate data underlies most classification problems. Yet exploratory data analysis (EDA, Tukey, 1977) of such spaces is difficult, increasingly so as dimension increases, and often leads to the consideration of models for these problems being considered to be black-boxes. There is increasing emphasis on the need to provide explainers to improve the interpretability for black-box models (Biecek, 2018; Biecek and Burzykowski, 2021; Lundberg and Lee, 2017; Ribeiro, Singh, and Guestrin, 2016; Wickham, Cook, and Hofmann, 2015). Visualization is an important part of providing interpretations (Anscombe, 1973;

Coleman, 1986; Goodman, 2008; Matejka and Fitzmaurice, 2017). Tour methods (Lee et al., 2021; Cook et al., 2008) provide ways to visualize linear projections of high-dimensional spaces, to obtain an overview of shape (distributions and associations) and anomalies (outliers, clusters). The recently introduced radial tour (Spyrison and Cook, 2020) provides a user-controlled manual rotation of variables into and out of a projection, which might especially be useful for studying variable importance.

Dimension reduction is commonly used in conjunction with visualization to provide informative low-dimensional summaries of high-dimensional data. There have been several user studies for dimension reduction comparing across embeddings and display dimensionality (Gracia et al., 2016; Wagner Filho et al., 2018). There are also empirical metrics and comparisons used to describe non-linear reduction and how well and faithfully they embed the data (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 techniques for assessing variable importance, particularly, how best to convey information to the viewer.

This paper describes a user study conducted to assess the benefit of the radial tour, in comparison with principal component analysis and a grand tour for understanding variable importance. The experiment is a within-participant user study. The type of visualization is the primary factor of the study, corresponding to a null hypothesis that all techniques provide a similar ability for the user to determine variable importance. The techniques are compared by having subjects complete several tasks, where accuracy and speed are recorded.

The paper is structured as follows. Section 3.2 provides the background on the visualization methods being compared. Section 3.3 describes the user study, the tasks, evaluation, and measures used. The results of the study are in Section 3.4. Conclusions and potential future directions are discussed in Section 3.5. The software used for the study is described in Section 3.6.

3.2 Background

3.2.1 Principal component analysis

Principal component analysis is a good baseline of comparison for linear projections because of its frequent and broad use across disciplines. 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 intact, the benefit comes from the ordered nature of the components. The first 2 or 3 components are typically used to approximate the variation multivariate data set, while the rest are discarded.

3.2.2 Scatterplot matrix

An extension to showing only a few components is to show pairs of components as a scatterplot matrix (Chambers et al., 1983), where all pairs of the components are displayed on the upper or lower triangles of the matrix. This is a convenient way to fit many plots onto a page, but when there are many variables there is insufficient space on a page to display them all. Figure 3.1 shows the first three components of simulated data as a scatterplot matrix.

3.2.3 Data visualization tours

A data visualization *tour* uses time to animate local changes in the projection basis. One of the key features of the tour is the object permanence of the data points; that is to say by watching nearby frames one can track the relative changes of observations as the basis moves toward the next target basis. There are various types of tours that are distinguished by the selection or generation 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.
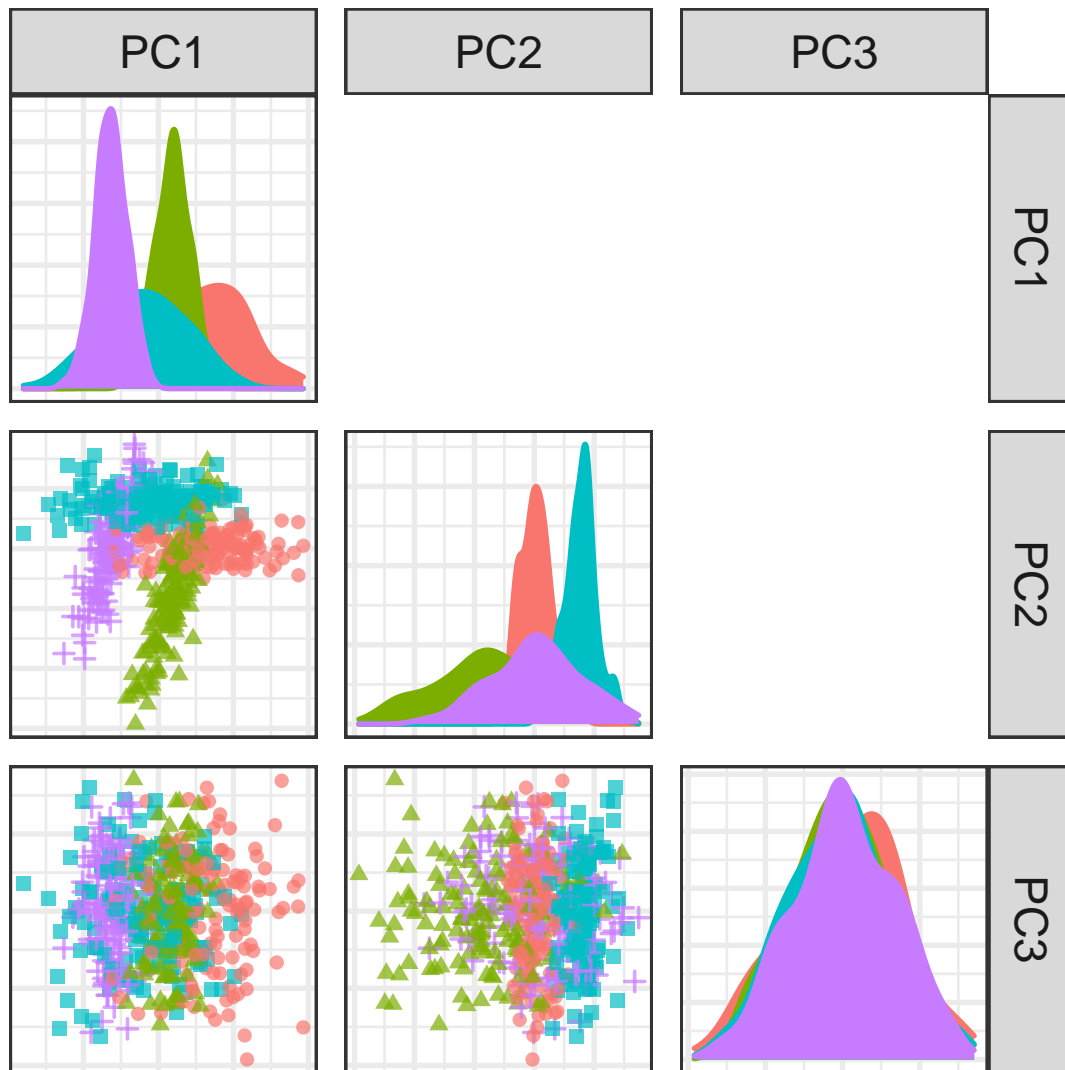
Figure 3.1: *Scatterplot matrix of the first 3 principal components for simulated data. Traditional multivariate visualization of *data*-space involes static display of the first several principal components.*

Grand tours

In a grand tour (Asimov, 1985) the target bases are selected randomly. The grand tour is the first and most widely known tour. It will serve as an intermediate unit of comparison which has continuity of data points in nearby frames along with the radial tour but lacks the user control enjoyed by PCA and radial tours. This lack of control makes grand tours more of a generalist exploratory tool.

Radial (manual) tours

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, 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 restrained, data projected, and rendered into an animation. For the 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 do or do not contribute 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 is allowed to 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, but also allows the user to steer via choosing the variable to rotate.

43

The recent implementation of manual tours us the R package {spinifex} (Spyrison and Cook, 2020), which facilitates manual tours (and radial variant). It is also compatible with tours made with {tourr} (Wickham et al., 2011) and facilitates exporting to .gif or .html widget, with recent graphic packages. 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.

3.3 User study

An experiment was constructed to assess the performance of the radial tour relative to the tour and scatterplots of principal components for interpreting the importance of variables to class separations.

The three methods were examined for three different cluster shapes, using different combinations of contributing variables, and data dimensionality. Data was collected using a specially constructed web app, through crowd-sourced with prolific.co (Palan and Schitter, 2018).

3.3.1 Experimental factors

In addition to visual factor, we vary the data across 3 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 different distributions of the data, and 3) the *dimension*-ality of the data.

The *location* of the separation of the clusters is a crucial aspect of analysis, it is the variables or combination their of that is important to the explanation of the structure. 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 3 clusters are defined. The name of the
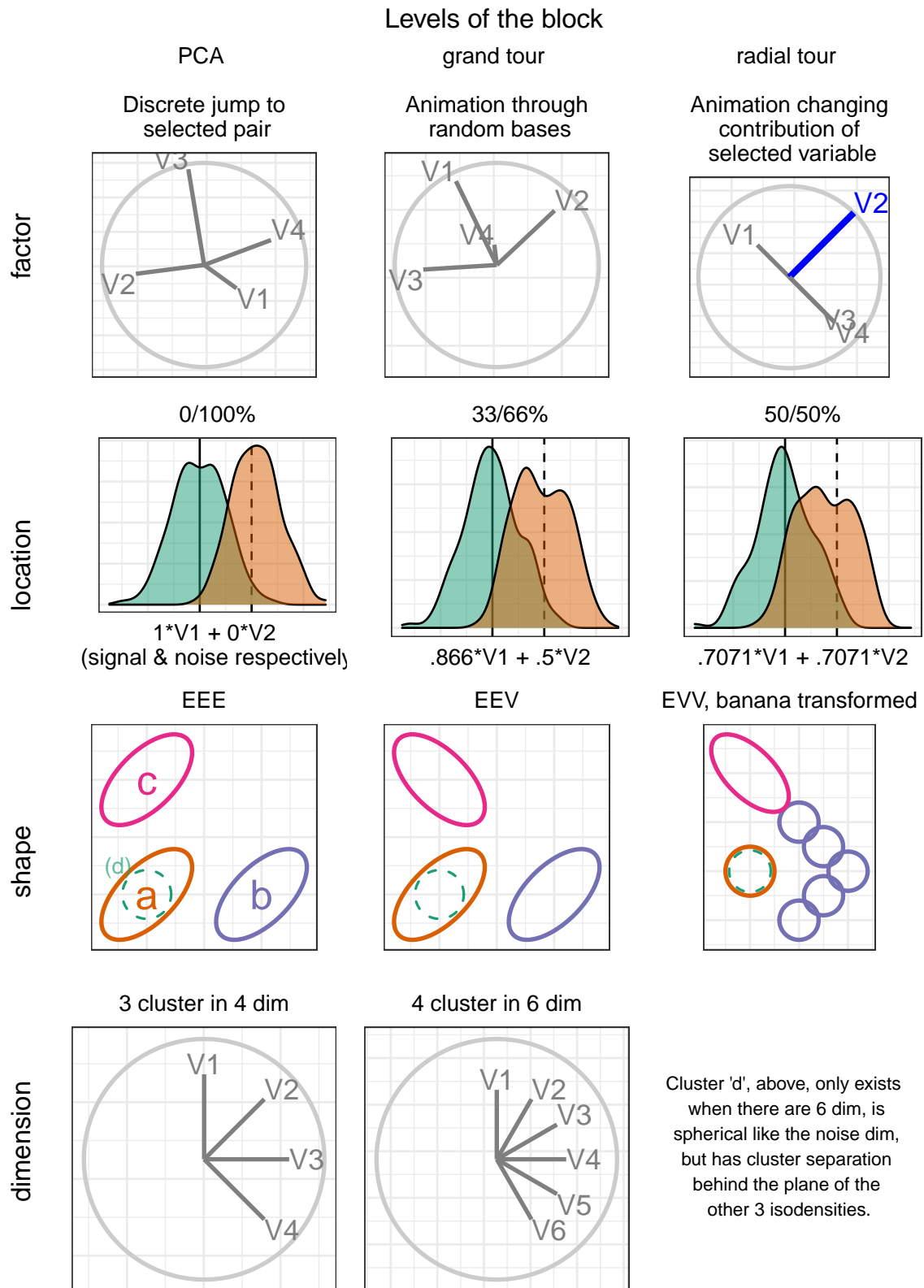
Figure 3.2: *Illustration of the experimental factors: visualization factor, the variable location of the cluster separationm, the shape of the clusters, and the dimensionality (and clusters) of the data.*

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 4 fifths of the data out in a "V" or banana-like shape. Figure 3.2 shows the principal component isodensity of the 3 model variants applied here.

Dimension-ality is tested at 2 modest levels, namely, in 4 dimensions containing 3 clusters and 6 dimensions with 4 clusters. We must do so to bound the difficulty and search space to keep the task realistic for crowdsourcing.

3.3.2 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, with the persistence of the data points across changes in basis, but without the ability to influence its 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. Next we cover how we expect them to perform and state the hypothesis to test.

Then for some subset of tasks, we expect to find that the radial tour performs most accurately, as it enjoys both the persistence of points and input control to explore specific variables. Secondly, it may be the case that grand performs faster than the alternatives with its absence of inputs, users can focus all of their attention on interpreting the fixed path. Conversely, we are less certain about the accuracy of such a limited grand tours as there is no objective function in the target bases; it is possible that, by chance, the planes completely avoid the information needed. However, given that the data dimensionality will be modest, it seems likely that grand tour regularly crosses frames with the correct information to perform the task quickly.

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

H_0 : visualization factor does not impact task *accuracy*, Y_1.

H_0 : visualization factor does not impact task *speed*, Y_2.

3.3.3 Task and evaluation

With our hypothesis formulated let's turn our attention to the task and how to evaluate it. Recall that the display was a 2D scatterplot with axis biplot to its left. Observations were supervised with the cluster level coded by 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 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 with a couple of features in mind: 1) the sum of squares of the individual variable marks should be 1, and 2) symmetric about 0, without preference to under- or over-guessing. 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.

$$W_j = \frac{(\overline{X_{j=1,k=1}} - \overline{X_{1,k=2}}, \; ... \; (\overline{X_{j=p,k=1}} - \overline{X_{j=p,k=2}})}{\sum_{j=1}^{p}(|\overline{X_{j,k=1}} - \overline{X_{j,k=2}}|)} - \frac{1}{p}$$

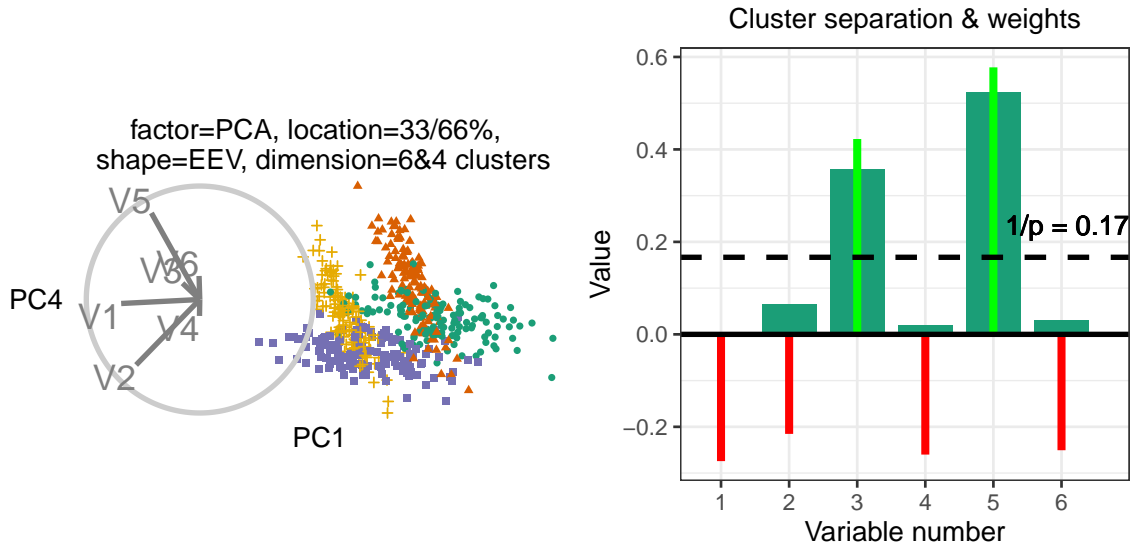$$\text{Accuracy}, Y = \sum_{j=1}^{p} I(r_j) * sign(w_j) * \sqrt{|w_j|}$$

Figure 3.3: *(L), PCA biplot of the components showing the most cluster separation with (R) illustration of the magnitude of cluster separation is for each variable (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 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 is the indicator function. Then the total marks for this task is the sum of this marks vector. We use the time till the last response as a secondary dependent variable Y_2.

3.3.4 Visual design standardization

Section 3.2 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 is controlled with the block assignment as illustrated below in Figure 3.4. Below we cover the visual design standardization, as well the 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 of axis titles) were held constant. Variable contributions were 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 the radial tour which variable should have its contribution animated.

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 was no user input for the 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 within the interpolation step size of 0.1 radians. Users were able to swap between variables, upon which the display would change 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 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.

3.3.5 Data simulation, task

Each dimension is originally distributed as $\mathcal{N}(2 * I(signal), 1) \mid$ covariance Σ, a function of the shape. Signal variables have a correlation of 0.9 when they have equal orientation and -0.9 when their orientations vary. Noise variables were restricted to 0 correlation. Each cluster is simulated with 140 observations and is offset in a variable that does not separate 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 are then iterated with each level of the factor. For PCA, every pair of the top 4 principal components and saved as 12 plots. For the grand tour, we

first save 2 basis paths (for 4 and 6 dimension), each replication is then projected through the common basis path as the variable(s) containing the were previously shuffled. The resulting animations were saved as .gif files. The radial tour starts at either the 4 or 6-variable "half-clock" basis, where each variable has a uniform contribution, and no variable contributing in the opposite direction (to minimize variable dependence), a radial tour is then produced for each variable and saved as a .gif.

3.3.6 Data collection, and factor assignment

Now, with simulation and their artifacts in hand. We explain how the experimental factors are assignment, and illustrate how this is experienced from a participant's perspective.

We section the study into 3 periods, each period is linked to a randomized level of both the factor visualization and the location. The order of dimension and shape are of secondary interest and are held constant in increasing order of difficultly; 4 then 6 dimensions and EEE, EEV, then EVV-banana respectively.

The period starts with an untimed training task at the simplest remaining block parameterization; location $= 0/100\%$, shape $=$ EEE, and 4 dimensions with 3 clusters. This serves to introduce and familiarize participants with input and visual differences. After the training, the participant is evaluated on 2 tasks with the same factor * location level, across the increasing difficulty of dimension * shape. These evaluations removed the plot after 60 seconds, though this limit was rarely reached by participants.

The order of the levels of the factor and location is randomized with a nested Latin square where all levels of factor are exhausted before advancing to the next level of location. That means we need $3!^2 = 36$ participants to perform a full block evaluation. This randomization is important to control for any potential learning effects the participant may receive. Figure 3.4 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 3 full evaluations to properly power our study; we set out to crowdsource $N = 3 * 3!^2 = 108$ participants.
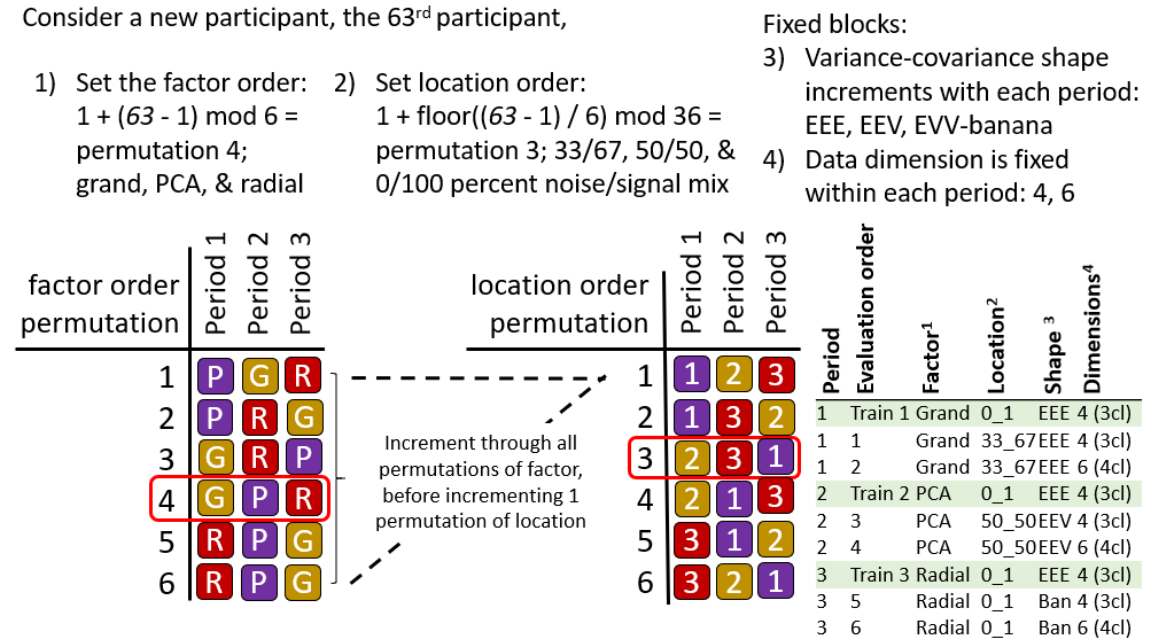
Consider a new participant, the 63rd participant,

1) Set the factor order:
1 + (*63* - 1) mod 6 =
permutation 4;
grand, PCA, & radial

2) Set location order:
1 + floor((*63* - 1) / 6) mod 36 =
permutation 3; 33/67, 50/50, &
0/100 percent noise/signal mix

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

Figure 3.4: *Illustration of the nested latin square and how a hypothetical participant 63 is assigned factor and block parameterizations. Each of the 6-factor permutations is exhausted before iterating to the next permutation of location.*

3.3.7 Recruiting subjects

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 location/language bias associated with. Participants were compensated for their time at £7.50 per hour, whereas the mean duration of the survey was about 16 minutes. We can't preclude previous knowledge or experience with the factors, but validate this assumption in the follow-up survey where we ask about familiarity with the factors (see Figure 3.6). 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 and well educated.

3.3.8 Collecting participant data

Data were recorded by a {shiny} application and were 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 volume of participant and over-collect survey trials until we had received our target 3 evaluations of our 36 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 to indicate if the survey had complete data. We filter to only the latest 3 complete studies of each block parameterization, those which should have experienced the least adverse network conditions. Of the studies removed the bulk were partial data and a few of 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 surveys that were 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.

3.4 Results

To recap, the primary response variable is task marks as defined in section 3.3.3, and the log of response time 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: visual factor, location of the cluster separation signal, the shape of variance-covariance matrix, and the dimension-ality of the data. Block parameterization and randomization were discussed in section 3.3.1. The survey was completed for 84 of these 108 trials and contains demographic information (preferred pronoun, age, and education), and subjective measures for each of the factors (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.

Random effect regression 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.

Fixed effects:	**Full model:**
α	$\widehat{Y_1} = \mu + \alpha_i + \mathbf{Z} + \mathbf{W} + \epsilon$
$\alpha + \beta + \gamma + \delta$	$\widehat{Y_1} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + \mathbf{Z} + \mathbf{W} + \epsilon$
$\alpha * \beta + \gamma + \delta$	$\widehat{Y_1} = \mu + \alpha_i * \beta_j + \mathbf{Z} + \mathbf{W} + \epsilon$
$\alpha * \beta * \gamma + \delta$	$\widehat{Y_1} = \mu + \alpha_i * \beta_j * \gamma_k + \mathbf{Z} + \mathbf{W} + \epsilon$
$\alpha * \beta * \gamma * \delta$	$\widehat{Y_1} = \mu + \alpha_i * \beta_j * \gamma_k * \delta_l + \mathbf{Z} + \mathbf{W} + \epsilon\%$

%

where μ is the intercept of the model including the mean of random effect

α_i, fixed term for factor $\mid i \in (\text{pca, grand, radial})$

β_j, fixed term for location $\mid j \in (0\_1, 33\_66, 50\_50)$ % noise/signal mixing

%

γ_k, fixed term for shape $\mid k \in (\text{EEE, EEV, EVV banana})$ model shapes

δ_l, fixed term for dimension $\mid l \in (4 \text{ variables \& 3 cluster, 6 variables \& 4 clusters})$

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

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

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

```
#> # A tibble: 5 x 8

#>   `Fixed effects` `No. levels` `No. terms` AIC    BIC    `R2 cond. (on RE)`

#>   <chr>               <int>        <int> <chr>  <chr>  <chr>

#> 1 a                     1            3 *1000* *1027* 0.18

#> 2 a+b+c+d               4            8 1026   1075   0.187

#> 3 a*b+c+d               5           12 1036   1103   0.198

#> 4 a*b*c+d               8           28 1069   1207   0.238

#> 5 a*b*c*d              15           54 1125   1380   *0.255*

#> # ... with 2 more variables: R2 marg. (w/o RE) <chr>, RMSE <chr>


#>                     Estimate Std. Error    df t value Pr(>|t|)

#> (Intercept)          -0.12       0.08   43.9   -1.50    0.14

#> fct=grand             0.15       0.09  622.4    1.74    0.08

#> fct=radial            0.37       0.09  617.1    4.18    0.00 ***

#> loc=33_66             0.17       0.09   83.2    1.78    0.08

#> loc=50_50             0.14       0.09   84.8    1.52    0.13

#> shp=EEV               0.04       0.06   11.5    0.79    0.44

#> shp=ban              -0.03       0.06   11.5   -0.48    0.64

#> dim=6                -0.06       0.05   11.5   -1.39    0.19

#> fct=grand:loc=33_66  -0.06       0.13  587.3   -0.49    0.63

#> fct=radial:loc=33_66 -0.34       0.13  585.2   -2.65    0.01  **
```

```
#> fct=grand:loc=50_50    -0.09        0.13 589.6   -0.68      0.50

#> fct=radial:loc=50_50   -0.19        0.13 574.3   -1.43      0.15
```

We also want to visually explore the conditional variables in the model. Figure 3.5 explores violin plots of marks by factor while faceting on the location (vertical) and shape (horizontal). Radial tends to increase the marks received, and especially so when there is no signal/noise mixing.

3.4.1 Subjective measures

The 84 evaluations of the post-study survey also collect 4 subjective measures for each factor. Figure 3.6 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 answers and found radial tours easier to use compared with the grand tour. All factors have reportedly low familiarity something we expect from crowdsourced participants.
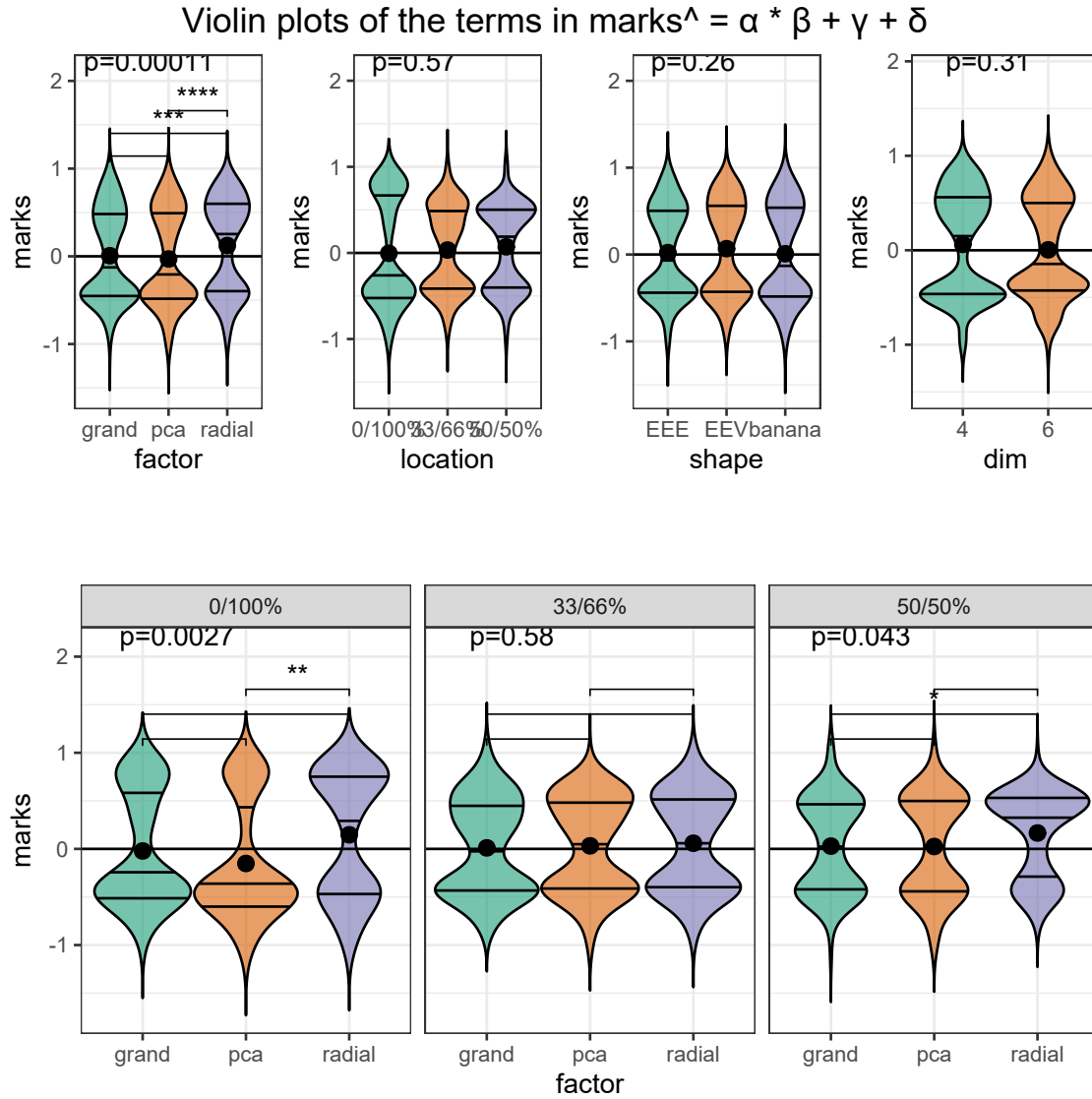
Figure 3.5: *Violin plots of terms of the model $\widehat{Y_1} = \alpha * \beta + \gamma + \delta$. Overlaid with global significance from the Kruskal-Wallis test, and pairwise significance from the Wilcoxon test, both are non-parametric, ranked sum tests suitable for handling discrete data. Participants are more confident and find the radial easier to use relative to the grand tour. Participants claim low familiarity as we expect from crowdsourced participants. Radial is more preferred compared with either alternative for this task.*
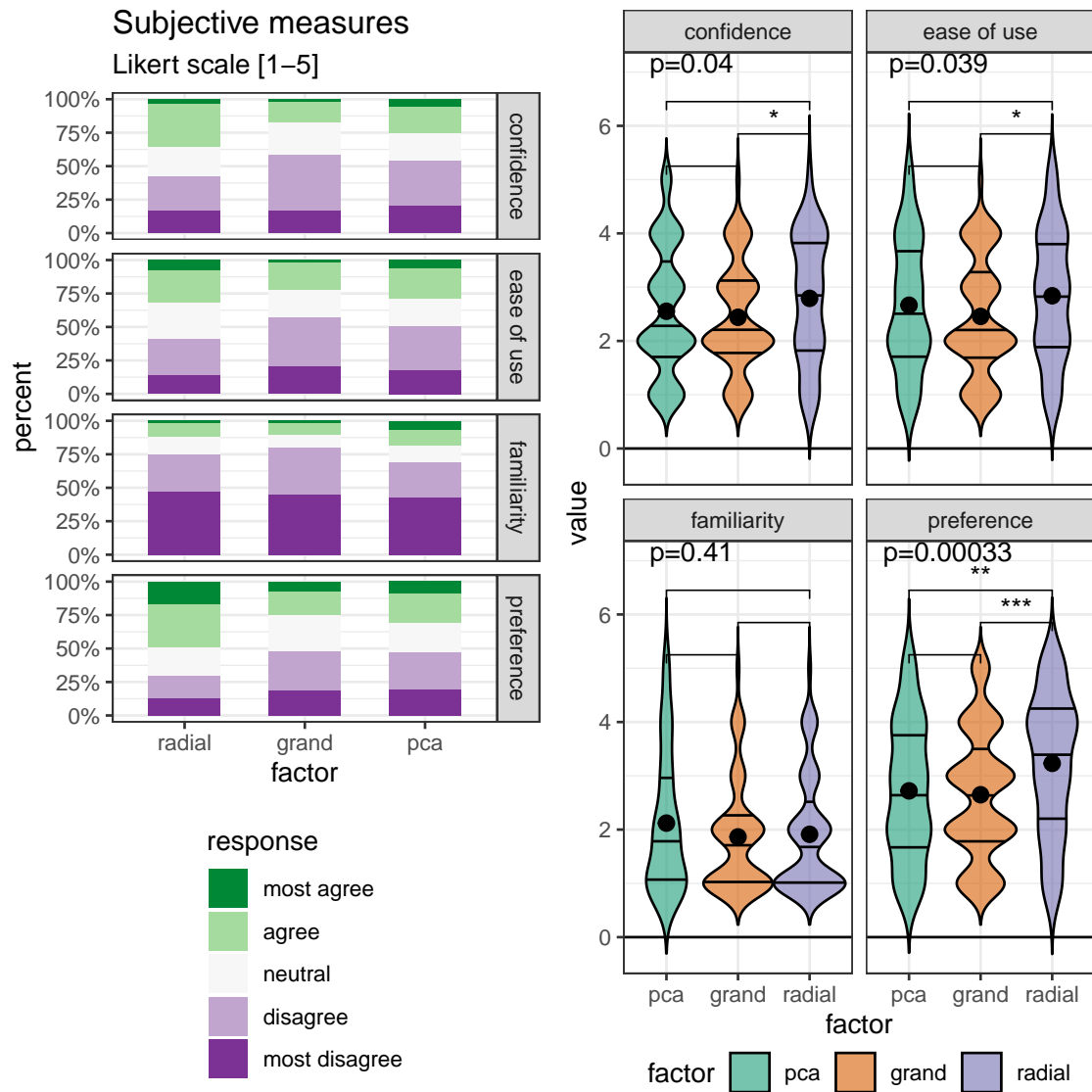
Figure 3.6: *The subjective measures of the 84 responses of the post-study survey, 5 discrete Likert scale levels of aggrement. (L) Likert plots (stacked percent bar plots) with (R) violin plots of the same measures. Violin plots are overlaid with global significance from the Kruskal-Wilcoxon test, and pairwise significance from the Wilcoxon test, both are non-parametric, ranked sum tests.*

3.5 Conclusion

Above we discussed an $n = 108$, with-in participant user study comparing the efficacy of 3 linear projection techniques. The participants performed a supervised cluster task, specifically the identification of which variables contribute to the separation between 2 target clusters. This was evaluated evenly over 4 block parameterizations. In summary, we find that radial tour increases accuracy while the grand tour decreases the time it takes to perform this task. These effects are large relative to the other block parameterizations, but smaller than the random effect of the participant. Radial tour was subjectively most preferred, lead to more confidence in answers, and is easier to use than alternatives.

There are several ways that this study could be extended. In addition to expanding the support of the block parameterizations, more interesting directions include: 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. Be sure to step back and plan the target support of your block parameters. Keep in mind the volume and quality of responses from participants especially when crowdsourcing. These planning steps are useful for navigating when the complexity of the application details.

3.6 Accompanying tool: radial tour 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) contains a free, open-source {shiny} (Chang et al., 2020) application. The application allows users to upload, perform limited preprocessing, and interactively explore their data via interactive radial tour. The .html widget produced is a more interactive variant relative to the application in the user study. Screencaptures and more details are provided in the appendix. Data can be imported from .csv, .rds, or .rda format, and projections. Run the following R code to run the application locally.

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

3.7 Acknowledgments

Chapter 4

Application of the radial tour to model-agnostic local explanations

THIS CHAPTER IS CURRENTLY BEING EDITED IN ITs OWN REPOSITORY, THIS VERSION IS OUT OF DATE.

TODO:XXX Segue.

4.1 Introduction

This work is concerned with linear projections of multivariate data. More specifically, we focus on the class of visualizations known as *tours* (Cook et al., 2008; Lee et al., 2021). Tours are viewed near-continuously through small changes to the projection basis. There are many variants of tours. We focus on one of the variants, *radial tours* (Cook and Buja, 1997; Spyrison and Cook, 2020). Radial tours can be used to control or steer the projection basis, while other tour variants select bases randomly or optimize an objective function. Because of this unique attribute, *user interaction* is another key aspect of interest in this work.

Radial tours change the basis by selecting one variable and specifying how to change its contribution to the current projection. By controlling the contribution of a single variable, a user can explore its sensitivity to the structure of the projection and identify which

variables are ultimately most important to the structure in question. In the work addressing RQ#1, we improve upon the geodesic interpolator for use in the radial tour, and apply it in an open-source R package, `spinifex`, this package facilitates making radial tours and extends the graphics packages interoperability for itself and the tours made from the existing package `tourr`.

Next, we substantiated the efficacy of radial tours as compared with user-selected, discrete combinations of principal components (Pearson, 1901) and continuous projections without interaction with the *grand tour* (Asimov, 1985). We conducted an $N = 108$ within-participant user study, where all participants use each of these visual factors. This is performed over balanced trials across the other experimental factors: location, shape, and dimension of the data. This addresses the second research question.

In our latest work, we want to see if we can apply the radial tour to aid the interpretability of black-box models. Such models have a non-linear space making them hard to interpret or understand. One recent branch in explainable artificial intelligence (XAI, Adadi and Berrada (2018); Arrieta et al. (2020)) is the use of local explanations or the attribution of the variables for one observation of a black-box model. We use such an explanation to form a 1D basis and perform radial tours to explore how the SHAP values behave differently for misclassified observations against neighboring correctly classified observations.

4.2 Motivation

The term exploratory data analysis (EDA) was coined by Tukey (1977), who leaves it as an intentionally broad term that encompasses the initial summarization and visualization of a data set before a testing hypothesis has been formulated. This is a critical first step for understanding and becoming familiar with data and validating model assumptions. It may be tempting to review a series of summary statistics to check model assumptions. However, there are known datasets where the same summary statistics miss glaringly obvious visual patterns (Anscombe, 1973; Matejka and Fitzmaurice, 2017). It is easy to look at the wrong, or incomplete set of statistics needed to validate assumptions. Data visualization is crucial in EDA, it *forces* you to see details and peculiarities of the data which are opaque to

numeric summarization, or more nefariously, obscure their true values. Data visualization does and must remain a primary component of data analysis and model validation.

studies/support –> While static documents are the norm, there are sizable benefits of user interaction. Interactive data visualization shift the locus of control back to the user, inviting them to explore and interact with the data, and offers a compact way to explore a wider range of dimensions, questions as they arise, which helped to keep the curiosity and the interest of the user.

With the emerging field of XAI, the constant tension between the interpretability of a model and its predictive power is receiving more attention. Linear models are the champions of interpretability with modest accuracy while increasing complex models improve accuracy but they can scarcely be interpreted even by experienced practitioners. One way to gain insight into a model is to focus on the local vicinity of one observation, and explain the variable weighting around that location, in an agnostic non-linear model. We call this observation level variable weights a *local explanation.* There are various such local explanations, many are tied to specific classes of models, while others are model-agnostic.

We know that data visualization is important in EDA and assumption validation. User interaction allows us to explore widely and quickly while allowing us to explore ideas as they arise. These 2 elements were especially important to consider from the work addressing RQ#1 as it forms a foundation to build on in further work. The efficacy of radial tours was supported by a user study in response to the second objective. In the latest work, we apply tours in tandem with local explanations to extend the interpretability of black-box models, to address RQ#3.

4.3 Extending the interpretation of black-box models with the use of interactive continuous linear projections

Local explanations describe the linear variable weights in the vicinity of an observation in $p - dim$ space. Local explanations are point-measurements of the weights for each variable describing the importance to the prediction at that point. In a highly non-linear space, the importance of the variables may change rapidly with even a small change in the

explanatory variables. There are several *model-agnostic* local explanations such as LIME (Ribeiro, Singh, and Guestrin, 2016), and SHAP (Lundberg and Lee, 2017). In practice, this application can be used with any model and compatible local explanation. However, below we apply and discuss SHAP values extracted from a random forest model.

4.3.1 How a SHAP value is calculated

We use FIFA soccer data (Leone, 2020) to explain SHAP values. We use 5000 player-observations of 9 aggregate skill measures to predict that player's wages. We use SHAP to observe how the skill attribution changes for different players, which should reflect different fielding positions. Fielding position is not explicitly used in the model, but we would expect that SHAP finds a way to change the variable weights across differing skill distributions to improve the accuracy of its wage predictions.

We have trained a random forest model and wish to further explore the weightings of this non-linear model. Following the work in (Biecek, 2018; Biecek and Burzykowski, 2021) we can similarly extract SHAP values, highlighting that different skills are valued differently across player positions within the model. We also show "break down" profiles, that is additive prediction explanations, how much of each player's predicted wages is added by each of the skill evaluations. Figure 4.1 takes a look at the SHAP and break down profiles of a star offensive and defensive player.

At the top of figure 4.1 we see the relative wages of the 2 different players. SHAP uses permutations of the X variables to approximate the variable weights. In the middle, we see the distributions of the variable weightings for each player across 25 such permutations. We use the median of those distributions (large dot) as the SHAP value. We see a difference in the weights across the 2 players that makes sense considering their positions. Reaction skills are important for both players, while offensive and movement are weighted higher of the offensive player. Conversely, defensive, power, and accuracy skills are weighted higher for the defensive player. The bottom used the SHAP values to explain their relative wage predictions. We see a similar trend of the SHAP values additively explaining the difference of the global intercept to their predictions. We would expect that other offensive players have explanations closer to Messi and defensive players to lie closer to van Dijk's. The
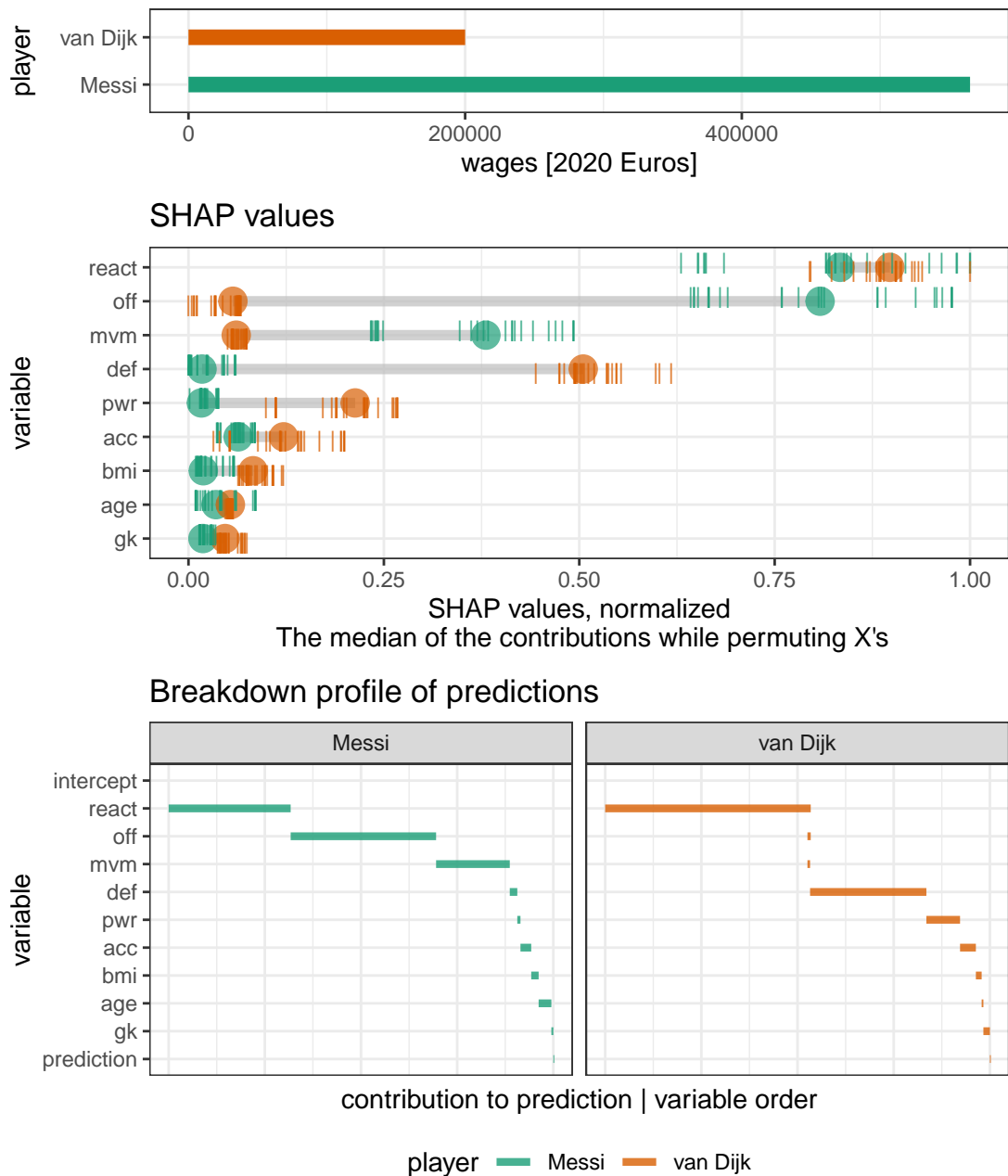
Figure 4.1: *SHAP values and prediction explanations of an offensive player (Messi, top) and a defensive player (van Dijk). SHAP values show a change in weights at the location of each player. Break down profiles show one order-sensitive explanation for the prediction of that observation.*

local explanations of middle fielders and goalkeepers would be expected to be different from both of these.

4.3.2 Our application; Trees of Cheem

Above, the FIFA data was used to regress continuous wages. In introducing this application, `Trees of Cheem`, we will be discussing the classification of simulated data. We simulation of 3 spherical clusters on the vertices of a triangle. The difference between the clusters is contained in the first 2 dimensions with no mean separation in another 2 noise dimensions. After extracting all observations' SHAP values, forming a SHAP *matrix*, of the original dimensionality, $(n \times p)$.

We want to show a global view of the SHAP matrix and show how its sensitivity differs from that of the original data space. In figure 4.2 we create an approximate of the data- and SHAP-spaces with their first 2 principal components.

This gives an orientation showing where selected observation lies. We facilitate exploration and observation identification by adding a hovering tooltip displaying row number and class (actual & predicted) with linked brushing highlighting selected points and displaying their data tabularly below the plot.

Looking at the SHAP space (right) the bulk of the correctly classified points are clustered in relatively small areas. This means that the distribution of their SHAP values is quite similar; the model is selecting a very tight variable contribution to explain the predicted class of each observation. Conversely, misclassified points tend to lie in between 2 clusters of correctly classified points. Using the interactive brushing and hover tooltips we confirm that these points lie between the actual and predicted classes.

Given the global view shown in figure 4.2, we want to look at the local explanation of primary and comparison points (highlighted as '*/x'). In this case, the primary observation is misclassified while the comparison point is a correctly classified nearby point. These 2 points that are classified differently, but otherwise have very similar values in the explanatory variables, lie quite far apart in SHAP space.
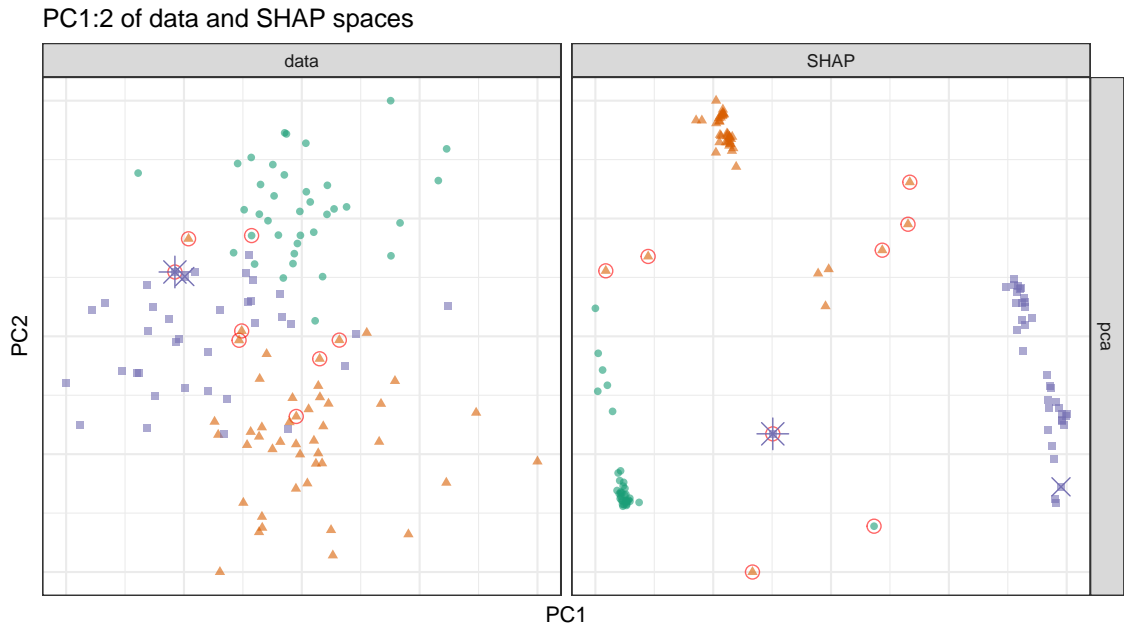
PC1:2 of data and SHAP spaces

Figure 4.2: *The first two principal conompents of the data- and SHAP-spaces (left and right respectively) of simulated data. The points are colored and shaped according to their predicted class, misclassified points are identified with a red circle. A target observation '*' is shown in contrast to a comparison point 'x', containing a nearby value in data space, but quite different SHAP space. These same 2 points are tracked in the proceeding tour.*

Initial contribution

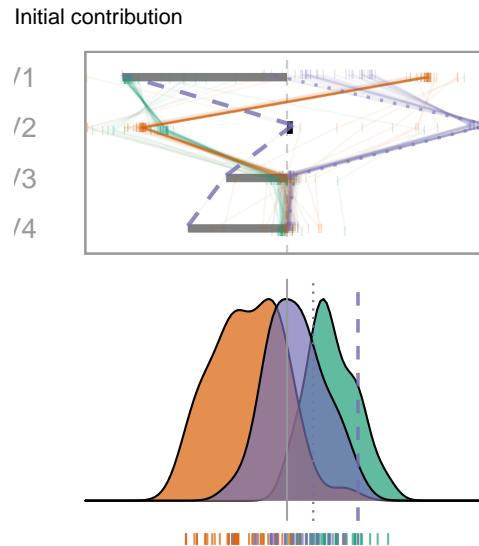**Figure 4.3:** *Top: 1D projection basis, the normalized SHAP values of the select observation are shown as grey/black bars. The distribution of the other observations is shown with parallel coordinate lines. The primary and comparison points ('*/x' previous) are shown as bolder dashed lines and fainter dotted lines respectively. Bottom: the 1D density and rug marks of the current projection basis.*

Now we have the global context and a comparison of the sensitivity of the SHAP spaces let's turn more toward the target observations. Figure 4.3 highlights the same selected points, bolder dashed and fainter dotted lines (previously '*/x' points respectively). The top of the figure shows the selected observation's SHAP values as grey/black bars. The other values of the SHAP matrix are displayed as short vertical bars and joining parallel coordinate lines and are similarly colored with their predicted class. The majority of the easy-to-classify points form a relatively tight spread in the first 2 variables (signal) and more chaotic behavior in the two noise variables as we would expect. Our dotted comparison lies with the bulk of the purple class. The dashed primary point is actually a green point misclassified as a purple, its weight for V1 looks like a green point, but its V2 weight deviates and is much closer to that of a purple. V2 is the variable that is crucial to the explanation of the model for why this particular point is misclassified. In the absence of V2, our target observation is weighted much closer to the that of its true class. This is the variable we want to explore the structure of in the radial tour as indicated with the black contribution bar.

Now that we have identified that V2 is the variable that deviated most from its true peer observation, we select it to rotate with the radial tour. In figure **??** illustrates the difference in the extremes of the contribution. The tour starts at the original contributions as described by the SHAP values (left frame). As the tour gives V2 a full contribution (middle frame) the comparison lies in the middle of the purples while the target observation is in the tails of both the purple and green (true class). In the right-most frame, the contribution from variable 2 is zero and the target variable lies in the middle of its true class. That is to say, without the contribution of variable 2 the local explanation does not seem reasonable; the selected observation looks much more like that of its true class.

4.3.3 Discussion

We have used radial tours to improve the interpretability of black-box models. We display 2D approximations of the data to keep the global context in mind and facilitate the selection of points to focus on. After a target point have been identified, its SHAP values are used as the initial projection basis. Where the within-class distributions of the value values are
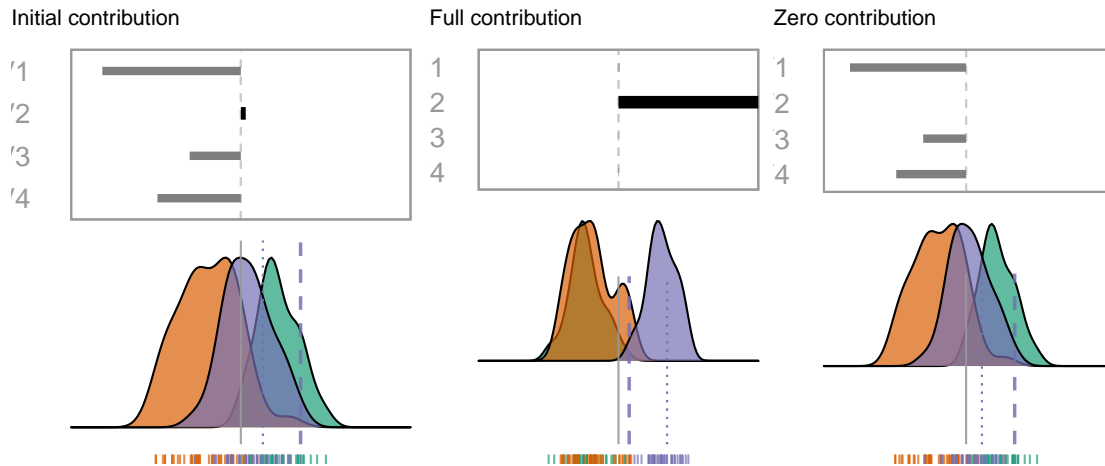
Figure 4.4: *The first frame of the radial tour. The SHAP values of the selected observation set the initial basis, shown as the grey and black bars on top. Within class distributions of the SHAP values are shown as parallel coordinate plots next to the variable contributions. The class densities and observation positions of the 1D projection are shown on the bottom. The tour animates over small changes in the basis (top bars) as the variable with the largest contribution (weight) is rotated to have a full contribution, zero contribution, and then back to the initial contribution. Variable 2 is critical to this local explanation; In the initial contribution, both selected observations are in the middle of the purple. With a full contribution of V2 typical of most purple, our primary observation is toward the tail. When there is no contribution of variable 2, the selected observation is in the middle of the green distribution, its true class.*

displayed. The default variable selected contains the largest difference from the median of SHAP values within that point's actual classification. That is that variable that deviates the most from what would be expected from a typical correctly classified observation.

Taking a step back, it is important to remember that this visualization and analysis is sensitive to the model and local explanation; it describes them as they are, independent of their validity and quality. The insight gleaned by this is predicated on meaningful selection of model and explanation.

Turning for a minute to real-world application, developing methods to better interpret black-box models is an important and impactful challenge. Private corporations and nation-states increasingly using complex models to classify and predict their customers and citizens from loans and insurance claims to employment and credit scores, the real-world impact of highly non-linear is here to stay. Being able to see the specifics of one observation

may seem small in the context of a model, but it is crucial to that instance; if misclassified, that observation may receive an outcome a world apart from its actual peers.

4.4 Acknowledgments

I would like to thank Professor Przemyslaw Biecek (Warsaw University of Technology) for his time and input in suggesting to look at SHAP local explanations and try applying to the FIFA dataset.

For transparency and reproducibility, the source files are made available at github.com/nspyrison/phd_milestones.

Chapter 5

Conclusion

We know that visualizing data is more robust numerical summarization alone. It provides a means to rapidly get a feel for the data, check for erroneous values, and confirm model assumptions. But visualizing data spaces becomes increasingly complex as dimensionality increase. Linear dimension reduction and especially dynamic animations of linear projections, tours, seems to be a fruitful way to extend data visualization as dimensionality increases. Previous we haven't had a means of testing the sensitivity of individual variables contribution to the structure in one embedding. With the **spinifex** package we create package facilitating manual tours. Then we explore the efficacy of the manual tour in a user study. Finally, we extended the use of the manual tour to the increase the interpretability of non-linear models by using it to explore their local explanations in the package **cheem**.

In Chapter 2 we discussed **spinifex**. How it facilitates the preprocess transformations with `scale_*` functions, identifies features with `basis_*` functions, and creates manual tours. It creates a framework for layered creation of tour visuals with `proto_*` functions that will feel at home to **ggplot2** users. Manual tours and those created from **tourr** can be rendered and exported as interactive HTML widgets, .gif, or .mp4 files. It has been downloaded over 14,400 times from CRAN between 09 April 2019 and 28 November 2021. my contributions **spinifex** and **tourr** won the ACEMS Impact and Engagement Award, 2018.

Chapter 3 discussed a user study comparing the radial tour against PCA and the grand tour. Participants performed a variable attribution task of the supervised cluster separation. We

define an accuracy measure for this task and the additional experimental factors location, shape, and dimensionality. I find strong evidence that use of the radial tour leads to high accuracy for this task. Using mixed model regression, I partition the error term of the model into the variation from participant's skill, and variation in difficult of the simulations due to random draws.

Now that we have evidence to support the radial tours use for variable attribution we extend their use to the further explore the local explanation of black-box models. Chapter 4 introduces an interactive global view that compares approximations of data- and attribution-spaces next to model performance. The explanations of selected observation from this global view are used as the initial basis for a radial tour to explore what support of variable contributions agree with the explanation view. Preprocessing functions as well as interactive application are provided in a free, open-source package **cheem**.

5.1 Further extensions

There are direct natural extension to **spinifex** such as adding adding protos, such as adding a text table of the basis, 2D density contours, convex hulls, alpha hulls, and an 'high density region' display where the bulk of the data is shown as density contour, while the outer most observations are displayed as points (Hyndman, 1996).

Similarly, **cheem** could be generalized to a broader range models and local explanations. The models and local explanations facilitated by `DALEX::explain()` seems to be a good starting point [Biecek (2018); biecek_explanatory_2021]. Alternatively there may be other statistics that better show the structure identified by explanations that could be added.

Experiencing tours as 3D scatterplots on 2D display, or in extended reality with steroscopically true head tracking may be fruitful. This has been partial explored [Nelson, Cook, and Cruz-Neira (1998); yang_3d_1999; yang_interactive_2000], though it would be interesting see an modern implementations using WebGL, Mozilla A-frame, or Unity. One concern would be trying to keep hardware and software as generalized as possible.

In addition to application that involve one tour variant such as the applications we have discussed or **liminal** (Lee, Laa, and Cook, 2020), a general more consolidated application

with multiple linked brushing would be nice to have. There have been many takes on applications to facilitate the exploration of multivariate data and the audience that know of and uses tours remains small. Tour-based applications (Fisherkeller, Friedman, and Tukey, 1974; Swayne, Cook, and Buja, 1991; Härdle, Klinke, and Turlach, 1995; Carr, Wegman, and Luo, 1996; Nelson, Cook, and Cruz-Neira, 1998; Sutherland et al., 2000; Swayne et al., 2003) and other multi-variate data visualization applications [Carr and Nicholson (1988); Tierney (1990); Huh and Song (2002); Wegman (2003); Jeong et al. (2009), hadjar_webvr_2018; cordeil_iatk_2019].

5.2 Other contributions

In addition to the research discussed around the thesis of the manual and radial, manual tour other notable contributions during my candidature include:

- *The state-of-the-art on tours for dynamic visualization of high-dimensional data* (Lee et al., 2021), WIREs Computational Statistics. Review of current tour methods. I contributed writing and manual tour frame discussing manual tours.

- *"Is IEEE VIS that good?" On key factors in the initial assessment of manuscript and venue quality* (Spyrison, Lee, and Besançon, 2021). A survey IEEE VIS authors, how they source articles, decide which to read, and evaluate venue quality. We find low evidence that sentiment changes across academic positions for these topics and provide commentary and discussion on the effects of "publish or perish" environment, standard author and journal metrics, and the need to publish "null" findings and replication studies.

- *Intraday effect of COVID-19 restrictions on Melbourne electricity consumption* (Barrow, Chong, and Spyrison, 2020). We corroborate that the Victorian interday effect on energy consumption did not change, and novelly find that the intraday distribution of energy consumption does change. Namely we find a statistically significant change in the height of the morning and evening peak energy consumption that we posit is due to less strict schedules associated with working from home, absence of commute time, or other employment changes. We were awarded 1st place of hundreds of entries in the insights category of the Melbourne 2020 Datathon.

- Student volunteering at conferences: UseR!2021 - Online, CHI Down Under 2020 - Online, and UseR!2018 - Brisbane, Australia.

Chapter 6

Appendix

ADD SPINIFEX VIGNETTE GGPROTO API HERE? after first paper?

Bibliography

Adadi, A and M Berrada (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE access* **6**, 52138–52160.

Allaire, JJ, Y Xie, J McPherson, J Luraschi, K Ushey, A Atkins, H Wickham, J Cheng, W Chang, and R Iannone (2020). *rmarkdown: Dynamic Documents for R.* `https://github.com/rstudio/rmarkdown`.

Anscombe, FJ (1973). Graphs in Statistical Analysis. *The American Statistician* **27**(1), 17–21. (Visited on 12/19/2018).

Arrieta, AB, N Díaz-Rodríguez, J Del Ser, A Bennetot, S Tabik, A Barbado, S García, S Gil-López, D Molina, and R Benjamins (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion* **58**, 82–115.

Asimov, D (1985). The Grand Tour: a Tool for Viewing Multidimensional Data. *SIAM journal on scientific and statistical computing* **6**(1), 128–143.

Barrow, M, J Chong, and N Spyrison (2020). Changes to Victorian Intraday Electricity Demand Following COVID-19 Restrictions. In: *Melbourne Datathon 2020, Insights category*, pp.4. `https://www.overleaf.com/project/5f614799515ac0000119daf7` (visited on 08/21/2021).

Bellman, RE (1957). *Dynamic Programming.* Princeton University Press. `https://books.google.com.au/books?id=fyVtp3EMxasC&redir_esc=y` (visited on 02/08/2021).

Bertini, E, A Tatu, and D Keim (2011). Quality metrics in high-dimensional data visualization: An overview and systematization. *IEEE Transactions on Visualization and Computer Graphics* **17**(12), 2203–2212.

BIBLIOGRAPHY

Biecek, P (2018). DALEX: explainers for complex predictive models in R. *The Journal of Machine Learning Research* **19**(1), 3245–3249.

Biecek, P and T Burzykowski (2021). *Explanatory Model Analysis: Explore, Explain, and Examine Predictive Models.* CRC Press.

Buja, A, D Cook, D Asimov, and C Hurley (2005). "Computational Methods for High-Dimensional Rotations in Data Visualization". In: *Handbook of Statistics.* Vol. 24. Elsevier, pp.391–413. `http : / / linkinghub . elsevier . com / retrieve / pii / S0169716104240147` (visited on 04/15/2018-04-15).

Carr, DB and WL Nicholson (1988). 'Explor4: A Program for Exploring Four-Dimensional Data Using Stereo-Ray Glyphs, dimensional constraints, rotation, and masking. *Cleveland and McGill (1988)*, 309–329.

Carr, D, E Wegman, and Q Luo (1996). ExplorN: Design considerations past and present. **129**.

Chambers, J, W Cleveland, B Kleiner, and P Tukey (1983). Graphical Methods for Data Analysis.

Chang, W, J Cheng, JJ Allaire, Y Xie, and J McPherson (2020). *shiny: Web Application Framework for R.* `https://CRAN.R-project.org/package=shiny`.

Coleman, D (1986). Geometric Features of Pollen Grains. In: Statistical Computing Statistical Graphics. `http://stat-computing.org/dataexpo/1986.html` (visited on 02/05/2021).

Cook, D and A Buja (1997). Manual Controls for High-Dimensional Data Projections. *Journal of Computational and Graphical Statistics* **6**(4), 464–480. (Visited on 04/15/2018-04-15).

Cook, D, A Buja, J Cabrera, and C Hurley (1995). Grand Tour and Projection Pursuit. *Journal of Computational and Graphical Statistics* **4**(3), 155. (Visited on 05/27/2018-05-27).

Cook, D, A Buja, EK Lee, and H Wickham (2008). "Grand Tours, Projection Pursuit Guided Tours, and Manual Controls". In: *Handbook of Data Visualization.* Berlin, Heidelberg: Springer Berlin Heidelberg, pp.295–314. `http://link.springer.com/10.1007/978-3-540-33037-0_13` (visited on 08/21/2018).

Cook, D, U Laa, and G Valencia (2018). Dynamical projections for the visualization of PDFSense data. *Eur. Phys. J. C* **78**(9), 742.

Fisherkeller, MA, JH Friedman, and JW Tukey (1974). PRIM-9: An Interactive Multidimensional Data Display and Analysis System.

Goodman, S (2008). A Dirty Dozen: Twelve P-Value Misconceptions. *Seminars in Hematology*. Interpretation of Quantitative Research **45**(3), 135–140. (Visited on 09/08/2020).

Gorman, KB, TD Williams, and WR Fraser (2014). Ecological sexual dimorphism and environmental variability within a community of Antarctic penguins (genus Pygoscelis). *PloS one* **9**(3), e90081.

Gracia, A, S González, V Robles, E Menasalvas, and T 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. (Visited on 08/20/2018).

Grinstein, G, M Trutschl, and U Cvek (2002). High-Dimensional Visualizations, 14.

Härdle, W, S Klinke, and B Turlach (1995). *XploRe: An Interactive Statistical Computing Environment.* https://dial.uclouvain.be/pr/boreal/object/boreal:172984 (visited on 01/05/2019).

Huh, MY and K Song (2002). DAVIS: A Java-based Data Visualization System. *Computational Statistics* **17**(3), 411–423. (Visited on 01/06/2019).

Hurley, C and A Buja (1990). Analyzing High-Dimensional Data with Motion Graphics. *SIAM Journal on Scientific and Statistical Computing* **11**(6), 1193–1211. (Visited on 11/27/2018).

Hyndman, RJ (1996). Computing and graphing highest density regions. *The American Statistician* **50**(2), 120–126.

Jeong, DH, C Ziemkiewicz, B Fisher, W Ribarsky, and R Chang (2009). iPCA: An Interactive System for PCA-based Visual Analytics. In: *Computer Graphics Forum.* Vol. 28. Wiley Online Library, pp.767–774.

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

Keim, DA (2000). Designing pixel-oriented visualization techniques: Theory and applications. *IEEE Transactions on visualization and computer graphics* **6**(1), 59–78.

Kirkpatrick, S, CD Gelatt, and MP Vecchi (1983). Optimization by simulated annealing. *science* **220**(4598), 671–680.

Lee, S, D Cook, N da Silva, U Laa, N Spyrison, E Wang, and HS Zhang (2021). The state-of-the-art on tours for dynamic visualization of high-dimensional data. *WIREs Computational Statistics* **n/a**(n/a), e1573. (Visited on 12/10/2021).

Lee, S, U Laa, and D Cook (2020). Casting Multiple Shadows: High-Dimensional Interactive Data Visualisation with Tours and Embeddings. *arXiv preprint arXiv:2012.06077.*

Leone, S (2020). *FIFA 20 complete player dataset.* https : / / kaggle . com / stefanoleone992/fifa-20-complete-player-dataset (visited on 08/23/2021).

Liu, S, D Maljovec, B Wang, PT Bremer, and V Pascucci (2017). Visualizing High-Dimensional Data: Advances in the Past Decade. *IEEE Transactions on Visualization and Computer Graphics* **23**(3), 1249–1268.

Lubischew, AA (1962). On the use of discriminant functions in taxonomy. *Biometrics*, 455–477.

Lundberg, S and SI Lee (2017). A unified approach to interpreting model predictions. *arXiv preprint arXiv:1705.07874.*

Maaten, L van der and G Hinton (2008). Visualizing Data Using t-SNE. *Journal of machine learning research* **9**(Nov), 2579–2605.

Matejka, J and G 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.* Denver, Colorado, USA: ACM Press, pp.1290–1294. http://dl.acm.org/citation.cfm?doid=3025453.3025912 (visited on 12/19/2018).

Munzner, T (2014). *Visualization analysis and design.* AK Peters/CRC Press.

Nelson, L, D Cook, and C Cruz-Neira (1998). XGobi vs the C2: Results of an Experiment Comparing Data Visualization in a 3-D Immer- sive Virtual Reality Environment with a 2-D Workstation Display. *Computational Statistics* **14**(1), 39–52.

Ocagne, Md (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, S and C Schitter (2018). Prolific. A Subject Pool for Online Experiments. *Journal of Behavioral and Experimental Finance* **17**, 22–27.

Pearson, K (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), 559–572.

Pedersen, TL and D 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/.

Ribeiro, MT, S Singh, and C Guestrin (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *arXiv:1602.04938 [cs, stat]*. (Visited on 08/20/2018).

Rodrigues, O (1840). Des lois géométriques qui régissent les déplacements d'un système solide dans l'espace: et de la variation des cordonnées provenant de ces déplacements considérés indépendamment des causes qui peuvent les produire. *Journal de Mathématiques Pures et Appliquées* **5**, 380–440.

Scrucca, L, M Fop, TB Murphy, and AE Raftery (2016). mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. *The R journal* **8**(1), 289–317. (Visited on 08/30/2020).

Sedlmair, M, T Munzner, and M Tory (2013). Empirical Guidance on Scatterplot and Dimension Reduction Technique Choices. *IEEE Transactions on Visualization & Computer Graphics* (12), 2634–2643.

Sievert, C (2020). *Interactive Web-Based Data Visualization with R, plotly, and shiny*. Chapman and Hall/CRC. https://plotly-r.com.

Spyrison, N and D Cook (2020). spinifex: an R Package for Creating a Manual Tour of Low-dimensional Projections of Multivariate Data. *The R Journal* **12**(1), 243. (Visited on 10/16/2020).

Spyrison, N, B Lee, and L Besançon (2021). "Is IEEE VIS *that* good?" On key factors in the initial assessment of manuscript and venue quality. *OSF preprints*. (Visited on 08/21/2021).

Sutherland, P, A Rossini, T Lumley, N Lewin-Koh, J Dickerson, Z Cox, and D Cook (2000). Orca: A Visualization Toolkit for High-Dimensional Data. *Journal of Computational and Graphical Statistics* **9**(3), 509–529. (Visited on 01/10/2019).

Swayne, DF, DT Lang, A Buja, and D Cook (2003). GGobi: evolving from XGobi into an extensible framework for interactive data visualization. *Computational Statistics & Data Analysis*. Data Visualization **43**(4), 423–444. (Visited on 12/19/2018-12-19).

Swayne, DF, D Cook, and A Buja (1991). *Xgobi: Interactive Dynamic Graphics In The X Window System With A Link To S.*

Tierney, L (1990). *LISP-STAT: An Object Oriented Environment for Statistical Computing and Dynamic Graphics*. Wiley Series in Probability and Statistics. New York, NY, USA: Wiley-Interscience.

Tukey, JW (1977). *Exploratory data analysis*. Vol. 32. Pearson.

Wagner Filho, J, M Rey, C Freitas, and L Nedel (2018). Immersive Visualization of Abstract Information: An Evaluation on Dimensionally-Reduced Data Scatterplots. In:

Wang, BT, TJ Hobbs, S Doyle, J Gao, TJ Hou, PM Nadolsky, and FI Olness (2018). Mapping the sensitivity of hadronic experiments to nucleon structure. *Physical Review D* **98**(9), 094030.

Wegman, EJ (2003). Visual data mining. *Statistics in Medicine* **22**(9), 1383–1397. (Visited on 12/19/2018).

Wickham, H and G Grolemund (2017). *R for Data Science*. O'Reily Media. `https://r4ds.had.co.nz/index.html`.

Wickham, H (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. `https://ggplot2.tidyverse.org`.

Wickham, H, D Cook, and H Hofmann (2015). Visualizing statistical models: Removing the blindfold: Visualizing Statistical Models. *Statistical Analysis and Data Mining: The ASA Data Science Journal* **8**(4), 203–225. (Visited on 03/16/2018).

Wickham, H, D Cook, H Hofmann, and A Buja (2011). tourr: An R Package for Exploring Multivariate Data with Projections. *Journal of Statistical Software* **40**(2). (Visited on 11/23/2018-11-23).

Xie, Y (2016). *bookdown: Authoring Books and Technical Documents with R Markdown.* Boca Raton, Florida: Chapman and Hall/CRC. https://github.com/rstudio/bookdown.

Xie, Y (2020). *knitr: A General-Purpose Package for Dynamic Report Generation in R.* https://yihui.org/knitr/.

Xie, Y, JJ Allaire, and G Grolemund (2018). *R Markdown: The Definitive Guide.* Boca Raton, Florida: Chapman and Hall/CRC. https://bookdown.org/yihui/rmarkdown.