spinifex: manual control of dynamic linear projections of high-dimensional data

true
true

Abstract

The set of dynamic linear projections of multivariate data collectively known as **tours** provides an important tool for extending the dimensionality of visuals. The R package **tourr** offers a variety of path generators and geometric displays for conducting tours. This paper discusses an extension package, **spinifex**, that adds support for the path generation of manual tours and extends the display of tours to use with the animation packages, **plotly** and **gganimate**. Manual tours are used to explore the sensitivity of structure as the contributions of a manipulation variable are changed. A recent paper {Wang et al. (2018)} visualizes the sensitivity of the hadronic experiments to nucleon structure. Sensitivity was characterized in non-linear 3D embeddings of the first 10 principal components. This research applies manual tours to this data showing that manual tours resolve information about cluster structure and identify outliers, information that is orthogonal to the original static viewing plane.

0.1 Introduction

A tour is a multivariate data analysis technique in which a sequence of linear (orthogonal) projections are viewed as an animation while the orientation of the projection basis is rotated across time. Each frame of the sequence corresponds to a small change in the projection for a smooth transition that perseveres continuity. This is a crucial tool extending the dimensionality of visualization which is a large boon as datasets and models exist in increasing cardinality. The package **tourr** (Wickham et al. 2011) builds a platform for generating tour paths and applying to a wide number of display geometrics in **base** graphics.

While there are numerous methods that generate tour paths, this research focuses on the manual tour. 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 how sensitive the projections structure is to these changes. This makes manual tours particularly useful once a feature of interest has been identified, for example, with the use of a guided tour (Hurley and Buja 1990). Rather than exploring the large phase space of the data, a guided tour selects a very specific path via projection pursuit(Diaconis and Freedman 1984), the optimization of an index function on the projection via a hill climbing algorithm. This 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, by controlling the contributions of individual variables.

Spinifex utilizes two new animation packages, **plotly** (Sievert 2018) and **gganimate** (Pedersen and Robinson 2019), to display tours, manual or other saved tours. From a given projection, the user can choose which variable to control, and the animation sequence is generated to remove the variable from the projection, and then extend its contribution to be the sole variable in one direction. This allows the viewer to assess the change in structure induced in the projection by the variable's contribution.

The paper is organized as follows. Section 0.2 explains the algorithm using a toy dataset. Section 0.3 discussed the display of the animation after the path has been generated. Section 0.4 illustrates how this can be used for sensitivity analysis applied to contemporary high energy physics. The last section, 0.6 summarizes the work and discusses future research.

0.2 Algorithm

The section below describes the algorithm for performing a 2D radial manual tour:

1. Provided with a 2D projection, choose a variable to explore. This is called the "manip" variable.
2. Create a 3D manipulation space, where the manip variable has the full contribution.
3. Generate a rotation sequence which increases the norm of the coefficient to 1 and zeros it.

The steps are described in more detail below. The R functions used below mentioned briefly, but more complete code example can be found in section **??**

0.2.1 Notation

This section describes the notation used in the algorithm for a 2D radial manual tour.

- $\mathbf{X}$, the data, an $n \times p$ numeric matrix to be embedded in two dimensions.
- $\mathbf{B} = (B_1,\ B_2)$, any orthonormal projection basis set, $p \times 2$ matrix, describing the projection from p to two dimensions
- $\mathbf{e}$, a zero column vector of length p with the $k-$th element set to one, where k is the number of the variable to manipulate.
- θ, the angle of in-projection-plane rotation, for example, on the reference axes.
- ϕ, the angle of out-of-projection-plane rotation, coming into the manipulation space.

The algorithm primarily operates on projection bases and utilizes the data only when making a display. The projection space can be viewed at any point in the process by pre-multiplying the data and plotting the first two variables.

0.2.2 Toy data set

The flea data, originally from Lubischew (1962), made available in **tourr** is used to illustrate the algorithm. The data contains 74 observations across 6 variables, physical measurements of the flea beetles. Each observation belonging to one of three species.

The data is defined. A basis set (ideally orienting to an interesting feature) must be provided as an initial orientation. One way to identify a projection containing interesting features is to apply a guided tour (Cook, Swayne, and Buja 2007). In a guided tour the projection sequence is selected by optimizing an index function via hill-climbing on the projection space. In this case, the holes index is selected and applied to standardized flea data. The holes index is maximized when the projected observations are furthest from the center. Figure @ref{fig:step0} shows a locally optimized projection of the data. The left panel displays the reference axes of the projection basis, a visual indication of the magnitude and direction each variable contributed to the projection. The right panel shows the data as projected through the basis set described by the reference axes (left). Data points are colored and given shape according to the species (while the guided tour was unsupervised with this information).

Call `view_basis()` on a basis to produce a **ggplot2** graphic similar to 1. Projection space is always available for display via the matrix multiplication $\mathbf{X}_{[n,\ p]}\ *\ \mathbf{B}_{[p,\ d]}\ =\ \mathbf{P}_{[n,\ d]}$.

```
# devtools::install_github("nspyrison/spinifex") # Development version
install.package("spinifex")
library("spinifex")

# Also see vignette:
vignette("spinifex") # vignette 'spinifex' not found
```

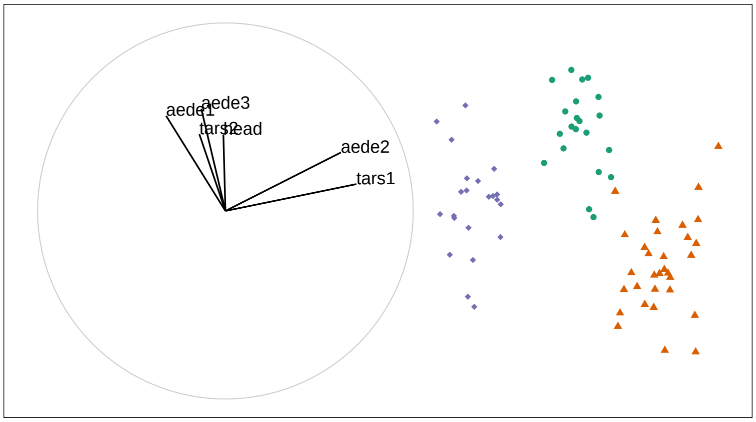

Figure 1: Basis reference axes (left) and projected data (right) of standardized flea data. Data points color and shape are mapped to beetle species. Basis identified by a holes-index guided tour. The variables `aede2` and `tars1` contribute mostly orthogonal to the other variables.

```r
f_data  <- tourr::rescale(flea[,2:6])
f_path  <- save_history(f_data, guided_tour(holes()))
f_basis <- matrix(f_path[,, max(dim(f_path)[3])], ncol=2)
f_cat   <- factor(flea$species)

view_basis(basis = f_basis,
           data = f_data,
           labels = colnames(f_data))
```

0.2.3 Step 1) Choose variable of interest

In figure 1 the contributions of the variables `tars1` and `aede2` are mostly orthogonal to the contributions of the other four variables. These two variables explain the variation of the data distinguishing the purple group from the rest of the sample. Select `aede2` as the manip var, the variable to be manipulated, as it has a larger contribution to the projection. The question that will be explored is how important the variable `aede2` is to the separation of the clusters.

0.2.4 Step 2) Create the manip space

Initialize a zero vector $\mathbf{e}$ of length p. Set the fifth element to one, as `aede2` is the fifth variable in the data, giving the manip var a full contribution in this dimension. Use the Gram-Schmidt process to orthonormalize the zero vector onto the basis yielding the 3D manipulation space, $\mathbf{M}$.

$$
\begin{aligned}
\mathbf{e} &\leftarrow \mathit{Orthonormalize}_{GS}(\mathbf{e}) \quad \mathit{w.r.t.\ the\ basis} \\
&\leftarrow \mathbf{e} - \langle \mathbf{e}, \mathbf{B}_1 \rangle \mathbf{B}_1 - \langle \mathbf{e}, \mathbf{B}_2 \rangle \mathbf{B}_2
\end{aligned}
$$

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

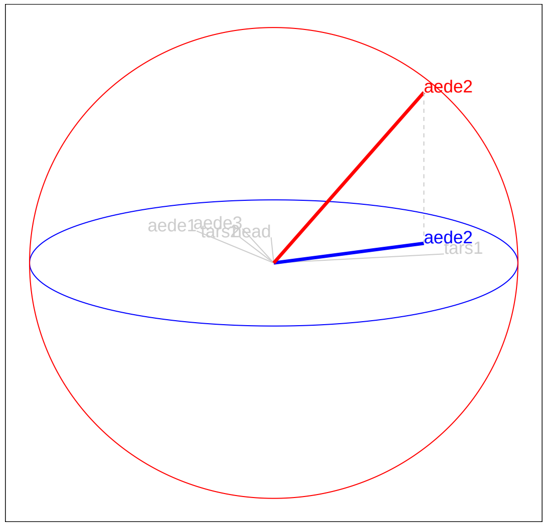

Figure 2: Manipulation space for controlling the contribution of `aede2` of standardized flea data. Basis selected by a holes-index optimized guided tour lies on the horizontal projection plane, shown in blue. The manip var axis, in red, extends into the vertical manipulation space, allowing the coefficients of the manip var to be changed by rotation around the origin.

Adding this new dimension to our projection space allows for the coefficients of the manip var to be changed via rotation about the origin. For example, the ability to lift a piece of paper, rather than being constrained to a 2D plane. Orthonormalizing rescales the new vector while the projection down to 2D remains the original basis. Place the plane horizontally, with the new dimension extending vertically, with axes projecting back onto the reference axes. Figure 2 illustrates this 3D manipulation space with the manip var highlighted (height of the other variables are not depicted.)

The representation in figure 2 can be produced by calling the function `view_manip_space()`.

```
view_manip_space(basis = f_basis,
                 manip_var = 4,
                 labels = colnames(f_data))
```

0.2.5 Step 3) Generate rotation

Imagine holding the manip var, the red axis, one end fixed to the origin. As it is controlled the manipulation space rotates about the origin, the projection onto the horizontal projection plane correspondingly moves. This is what happens in a manual tour. Generating a sequence of values for the horizontal and vertical, angles produces a path for the rotation of the manipulation space. This defines the (orthonormally-constrained) rotation on the coefficients of the variables.

For a radial tour fix the (horizontal) angle within the projection plane, θ, and define a sequence for the (vertical) angle coming out of the projection plane, ϕ, bringing the initial `XY` contributions of the manip var to a maximum and then to zero before returning to the initial position. Dynamic capture of user manipulation is typically performed directly on the projection plane (without depiction of the manipulation space.)

For i in 1 to n_slides:

Post-multiply the manipulation space by the pre-defined rotation matrix producing **RM**, the rotated manip space.

Let:

c_θ be the cosine of θ

c_ϕ be the cosine of ϕ

s_θ be the sine of θ

s_ϕ be the sine of ϕ

then

$$\begin{aligned}
\mathbf{RM}_{[p,\ 3,\ i]} &= \mathbf{M}_{[p,\ 3]} \ * \ \mathbf{R}_{[3,\ 3,\ i]} \\
&= \begin{bmatrix} M_{1,\ 1} & M_{1,\ 2} & M_{1,\ 3} \\ M_{2,\ 1} & M_{2,\ 2} & M_{2,\ 3} \\ \vdots & \vdots & \\ M_{p,\ 1} & M_{p,\ 2} & M_{p,\ 3} \end{bmatrix}_{[p,\ 3]} * \begin{bmatrix} c_\theta^2 c_\phi s_\theta^2 & -c_\theta s_\theta(1-c_\phi) & -c_\theta s_\phi \\ -c_\theta s_\theta(1-c_\phi) & s_\theta^2 c_\phi + c_\theta^2 & -s_\theta s_\phi \\ c_\theta s_\phi & s_\theta s_\phi & c_\phi \end{bmatrix}_{[3,\ 3,\ i]}
\end{aligned}$$

A note on application: ϕ is the angle relative to the initial value of ϕ, we find the transformation ϕ_i - ϕ_1 useful to think about ϕ relative to the basis plane. Additionally, the value of ϕ may be out of phase by a factor of pi. If the manip variable doesn't move as expected these are the first places to check.

Figure 3 illustrates a sequence with 15 projected bases, showing the reference axes on top the corresponding projected data points on the below. As a result, we can see that changes in the manip var controlled the distance between the purple cluster and the remaining sample, `aede2` is crucial in distinguishing this species. Tours are typically viewed as an animation such a dynamic version of this tour can be viewed online at https://nspyrison.netlify.com/thesis/flea_manualtour_mvar5/. The page may take a moment to load.

Animations can be produced using the function `play_manual_tour()`. This function defaults to an HTML5 widget produced from **plotly**. The `render_type` argument can be changed to `render_gganimate` for exporting to .gif or .mp4 files.

```
angle_speed <- .26

play_manual_tour(data = f_data,
                 basis = f_basis,
                 manip_var = 5,
                 angle = angle_speed,
                 col = f_cat,
                 pch = f_cat)
```

0.3 Data in projection-space

In an appeal to performance, the above operations are performed on the bases without the use of the larger datasets. After the bases are brought into the projection-space, however, it is helpful to observe them with data in the same space. Pre-multiply the data by basis frame bringing the data into the projection space.

$$\mathbf{P}_{[n,\ 3]} = \mathbf{X}_{[n,\ p]} \ * \ \mathbf{RM}_{[p,\ 3]} \tag{1}$$

$$= \begin{bmatrix} X_{1,\ 1} & \dots & X_{1,\ p} \\ X_{2,\ 1} & \dots & X_{2,\ p} \\ \vdots & \vdots & \vdots \\ X_{n,\ 1} & \dots & X_{n,\ p} \end{bmatrix}_{[n,\ p]} * \begin{bmatrix} RM_{1,\ 1} & RM_{1,\ 2} & RM_{1,\ 3} \\ RM_{2,\ 1} & RM_{2,\ 2} & RM_{2,\ 3} \\ \vdots & \vdots & \vdots \\ RM_{p,\ 1} & RM_{p,\ 2} & RM_{p,\ 3} \end{bmatrix}_{[p,\ 3]} \tag{2}$$

For a 2D scatterplot, plot the first two variables from each frame statically as in the previous figure, or in sequence, producing an animated scatterplot. The remaining variable is sometimes linked to a data point aesthetic (such as size or color) to produce depth cues used in conjunction with the XY scatterplot.

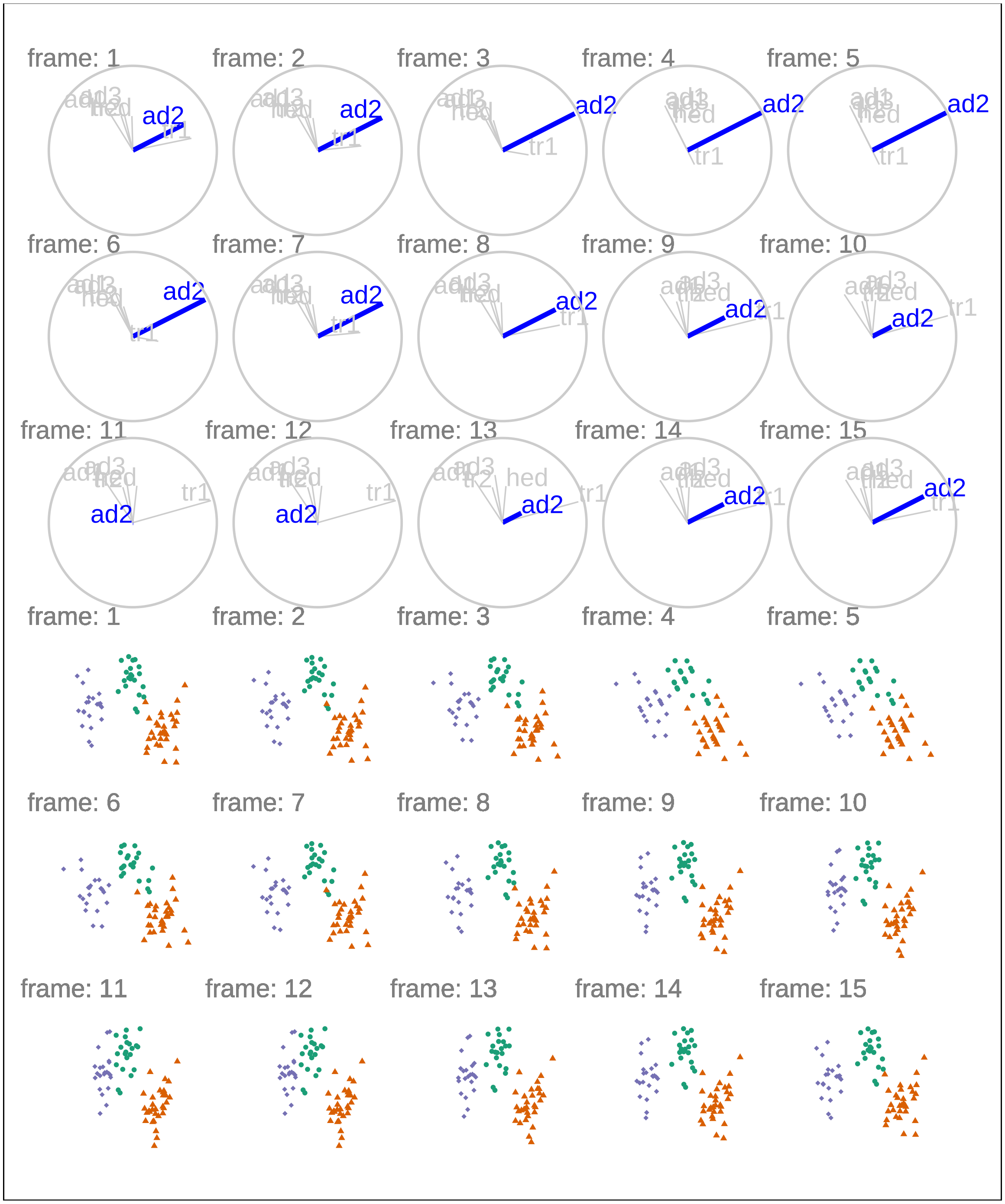

Figure 3: Radial manual tour manipulating `aede2` of standardized flea data. The contributions increase from its initial contribution to a full contribution to the projection before decreasing to zero and then returning to its initial value. The change in the projected data shows that `aede2` is important for distinguishing the purple species. An animated version can be viewed at https://nspyrison.netlify.com/thesis/flea_manualtour_mvar5/.

0.3.1 Rendering and sharing

The **tourr** package utilizes **base** graphics for the display of tours. **spinifex** allows tours to be used in rendered in **plotly** as an HTML5 object or **gganimate** as .gif or .mp4 objects. Both of which build off **ggplot2** objects in internal functions. Sharing of animations is not trivial especially in print and static file formats such as .pdf. Even with the use of computers and dynamic file formats capturing the correct resolution, aspect, and display is challenging and many formats quickly bloat file sizes. Keep in mind hosting options and exporting functions from **plotly**, **gganimate** and **tourr**.

0.3.2 Storage

Storing each data point for every frame of the animation is redundant. Just as operations are performed on the bases, so too should tour paths be stored as bases and a single instance of the data. Consider a radial manual tour, we can store the salient features in 3 bases, where ϕ is at its starting, minimum, and maximum values. The frames in between can be interpolated by supplying angular speed. With the use of the `tourr::save_history()` function, the target bases can be saved. From there geodesic interpolation can be used to populate the intermittent frames. This type of interpolation should not be used on manual tours, which have already been initialized into a 3D manipulation space where direct linear interpolation is appropriate.

0.4 Application

In a recent paper, Wang et al. (2018), the authors aggregate and visualize the sensitivity of hadronic experiments to nucleon structure. The authors introduce a new tool, PDFSense, to aid in the visualization of Parton distribution functions (PDF). 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.

Using the same data, another study, Cook, Laa, and Valencia (2018), applies grand tours (Asimov 1985) to the same subspaces. Grand tours are dynamic subspace projections of high dimensional where frames and selected at random and linked with geodesically interpolation of the intermediate frames. Grand tours are able to better resolve the distribution shape of clusters, intra-cluster detail, and better outlier detection than the use of PDFSense & TFEP (TensorFlow embedded projections). Before applying manual tours let's discuss the structure of the data.

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, in turn, have many observations. In the 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.

0.4.1 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 subset down to `ATLAS7old` and `ATLAS7new` to focus in on two groups with a reasonable number of observations that occupy different parts of the subspace. Radial manual tours controlling contributions from PC4 and PC3 are shown in figure 4 and figure 5 respectively. These components are selected to contrast the difference of information conveyed by touring different variables. Links to dynamic HTML5 animations controlling each of the four variables are also provided.

When manipulating PC4, there is a clear difference in the parameter space spanned by the experiment types `ATLAS7new` and `ATLAS7old`. Yet, when PC3 is manipulated there is no indication that the different experiments probe different parameter space. Performing a radial manual tour on PC4 is more insightful than for PC3.

Radial manual tours manipulating each of the principal components in the `jet` cluster can be viewed by following the links: PC1, PC2, PC3, and PC4.

0.4.2 DIS cluster

A different space is used to explore the `DIS` cluster; specifically the first six principal components, which explain 48% of the variation contained within the aggregated data(Cook, Laa, and Valencia 2018). Radial manual tours are performed on PC6 and PC2 in figure 6 and figure 7 respectively.

The selection of the manip variable is important as the manipulation spaces convey substantially different information. The manual tour of PC6 offers information about the dimensionality, shape, and orientations of the different experiment classes. whereas manipulating the contributions of PC2 only shows a subset of the dimensionality and shape information. Manipulating the contributions of PC6 turned out to be much more insightful than that of PC2. This result might seem counter-intuitive at first as PC2 should explain much more of the variation in the data. However, features and structure in the data regularly reside in finer detail that can be lost when looking only at static projections.

`DIS` cluster manual tours manipulating each of the principal components can be viewed from the links: PC1, PC2, PC3, PC4, PC5, and PC6.

0.5 TODO:

0.5.1 Summary

0.6 Discussion

Tours, the dynamic linear projection of multivariate data, is an important aspect of data visualization extending the display of data-space as data dimensionality increases. This research has modified the algorithm producing manual tours, applied this functionality in R and offers extends the graphics offerings that can be used to display tours. The paragraphs below explore how this work might be extended.

Future research on the algorithm would include extending it for use in 3D projections. The addition of another dimension theoretically allows for improved perception. This could explore interactions in immersive virtual reality or mixed reality, which may further allow for a better perception of structure and aid in higher-dimensional function visualization. Functions with many parameters suffer from the same dimensionality problem as data while their possible values lie on a plane of values rather than discrete points. Occulation, or the closer surface blocking further surfaces, will likely be an issue that may be alleviated by the use of wire mesh, changing opacity, or looking at sections of the projections (Furnas and Buja 1994).

The **tourr** package provides many other geometric displays with the `tourr::display_*()` family. These geometric options could be integrated into the **ggplot2** framework for display on **plotly** and **gganimate**. Additionally, the **animation** package Xie et al. (2018) could be implemented for another graphics framework. However, **animation** builds from **base** graphs while **spinifex** utilizes **ggplot2** graphics.

The Givens rotations and Householder reflections as outlined in Buja et al. (2005) could also be added. Currently, Gram-Schmidt is the only form of frame interpolation used (not used in manual tours). In a Givens rotation, the x and y components (for example $\theta\ = 0,\ pi/2$) of the in-plane rotation are calculated separately and would be applied sequentially to produce the radial rotation. Householder reflections define reflection axes to project points on to the axes and generate rotations.

Having a script only interaction with tours causes a significant barrier to entry. To a lesser extent, **plotly** offers some static interactions with the contained object, such as tooltips, brushing, and linking without communicating back to the R console. The development of a dynamic graphical user interface, perhaps with

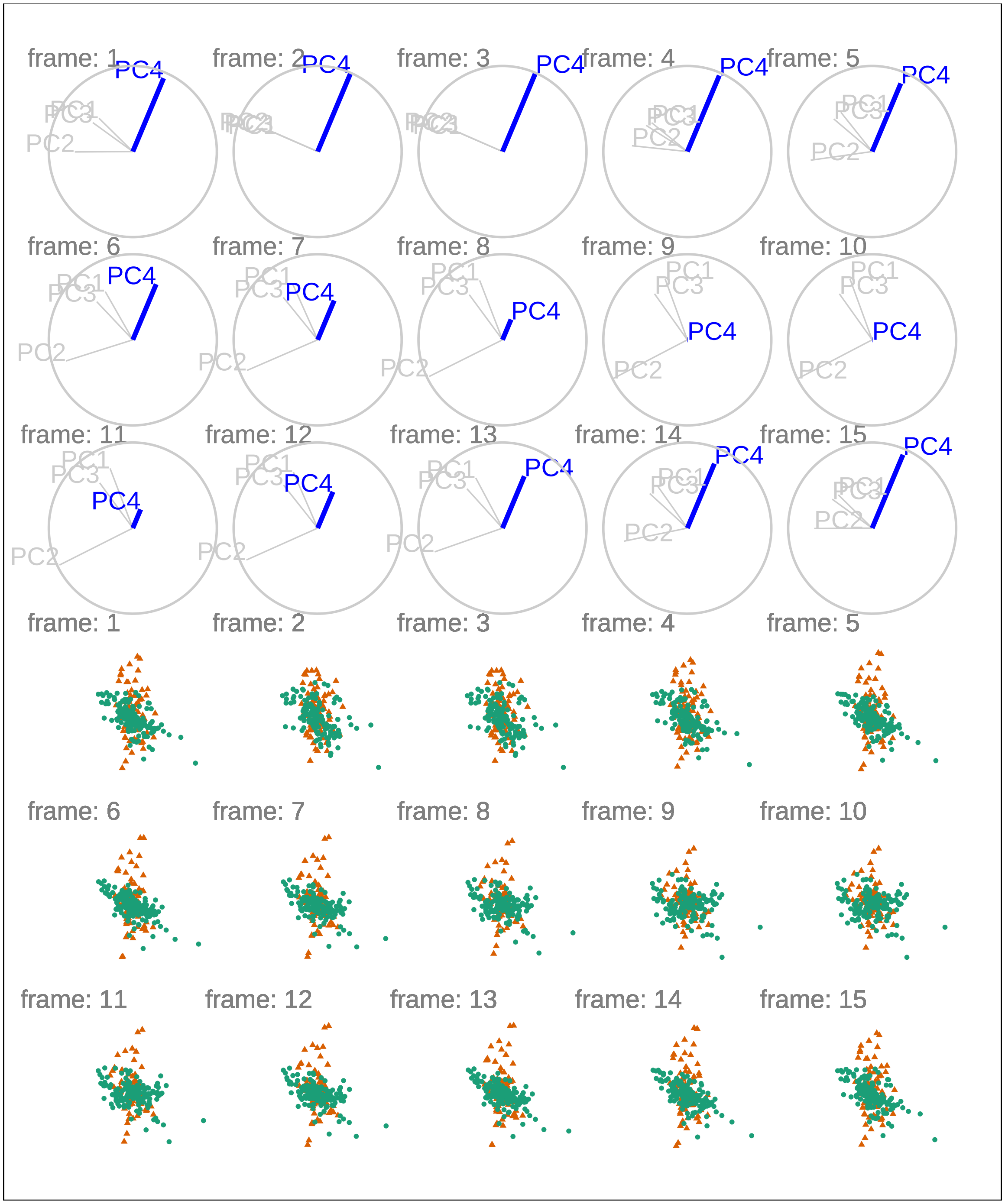

Figure 4: A radial manual tour of PC4 within the `jet` cluster. Colored by experiment type: `ATLAS7new` in green and `ATLAS7old` in orange. When PC4 fully/negligibly contributes to the projection `ATLAS7new` (green) spans the same space as the orange points. During the intermediate frames, the `ATLAS7new` is compressed in the direction radial to PC4. A dynamic version can be viewed at https://nspyrison.netlify.com/thesis/jetcluster_manualtour_pc4/.

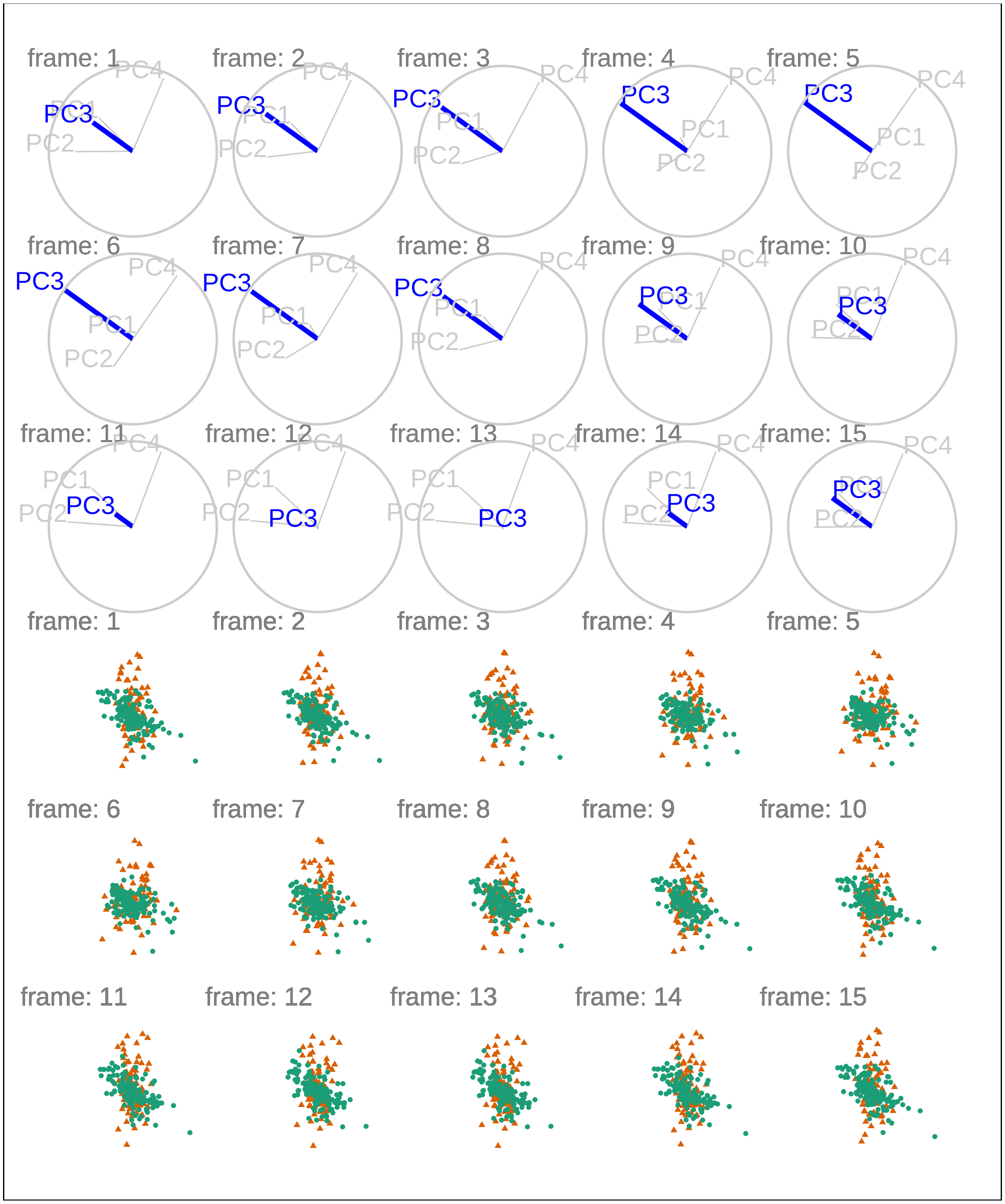

Figure 5: A radial manual tour of PC4 within the `jet` cluster. Colored by experiment type: `ATLAS7new` in green and `ATLAS7old` in orange. Data from `ATLAS7new` (green) spans mostly the same space as `ALTLAS7old` (orange) with no evident difference in cluster structure across varying contributions of PC3. A dynamic version can be viewed at https://nspyrison.netlify.com/thesis/jetcluster_manualtour_pc3/.

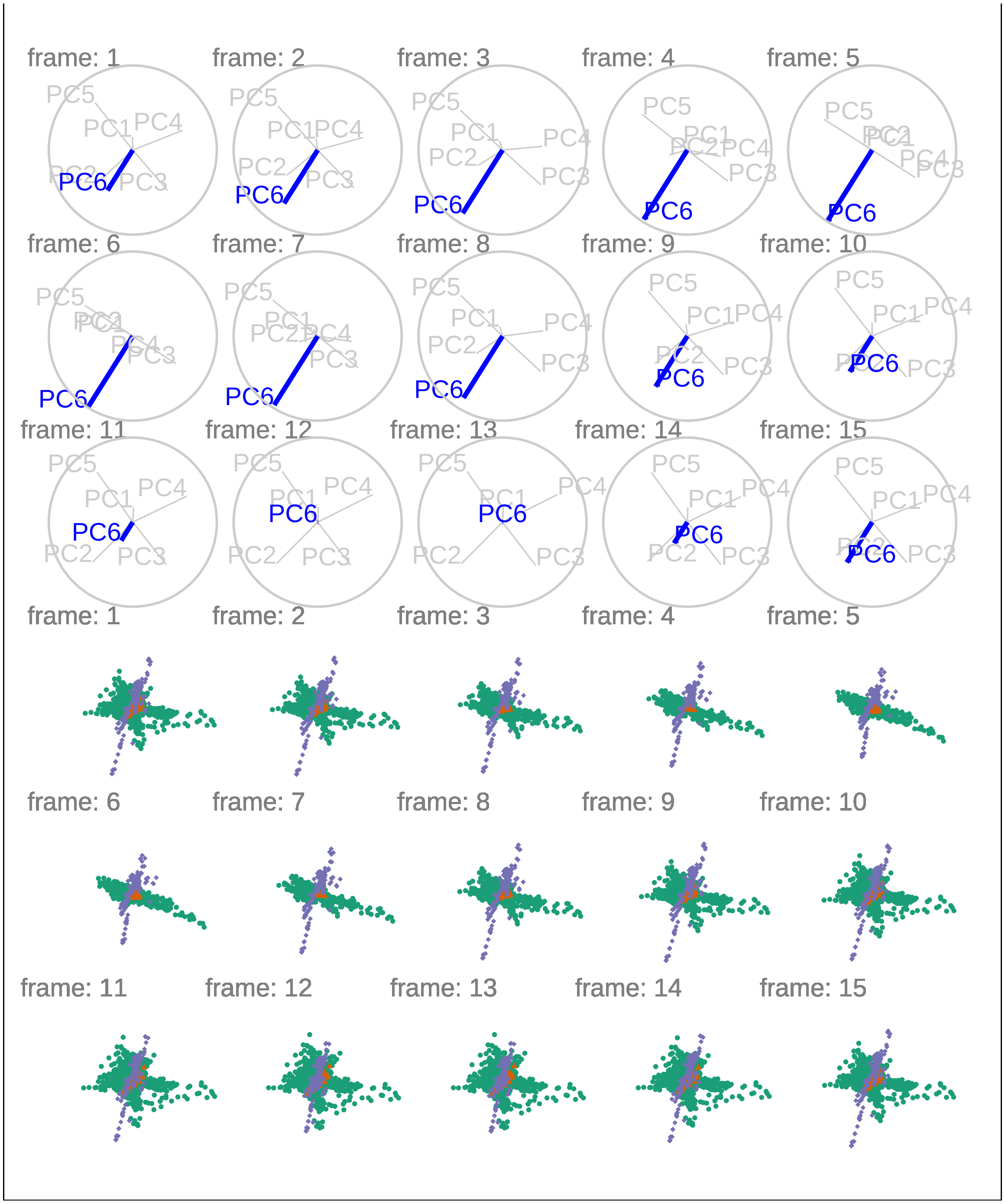

Figure 6: A radial manual tour manipulating the contribution of PC6 within the `DIS` cluster. Points are colored by experiment type: `DIS HERA1+2` in green, `dimuon SIDIS` in purple, and `charm SIDIS` in orange. The cluster `DIS HERA1+2` (green) is distributed in a cross-shaped plane, `charm SIDIS` (orange) occupies the center space of this cross, with the plane projecting into the field of view when the contribution of PC6 is max. Less evident is the linear `dimuon SIDIS` (purple) observations approaching the line of view for intermediate values of PC6. A dynamic version can be viewed at https://nspyrison.netlify.com/thesis/discluster_manualtour_pc6/.

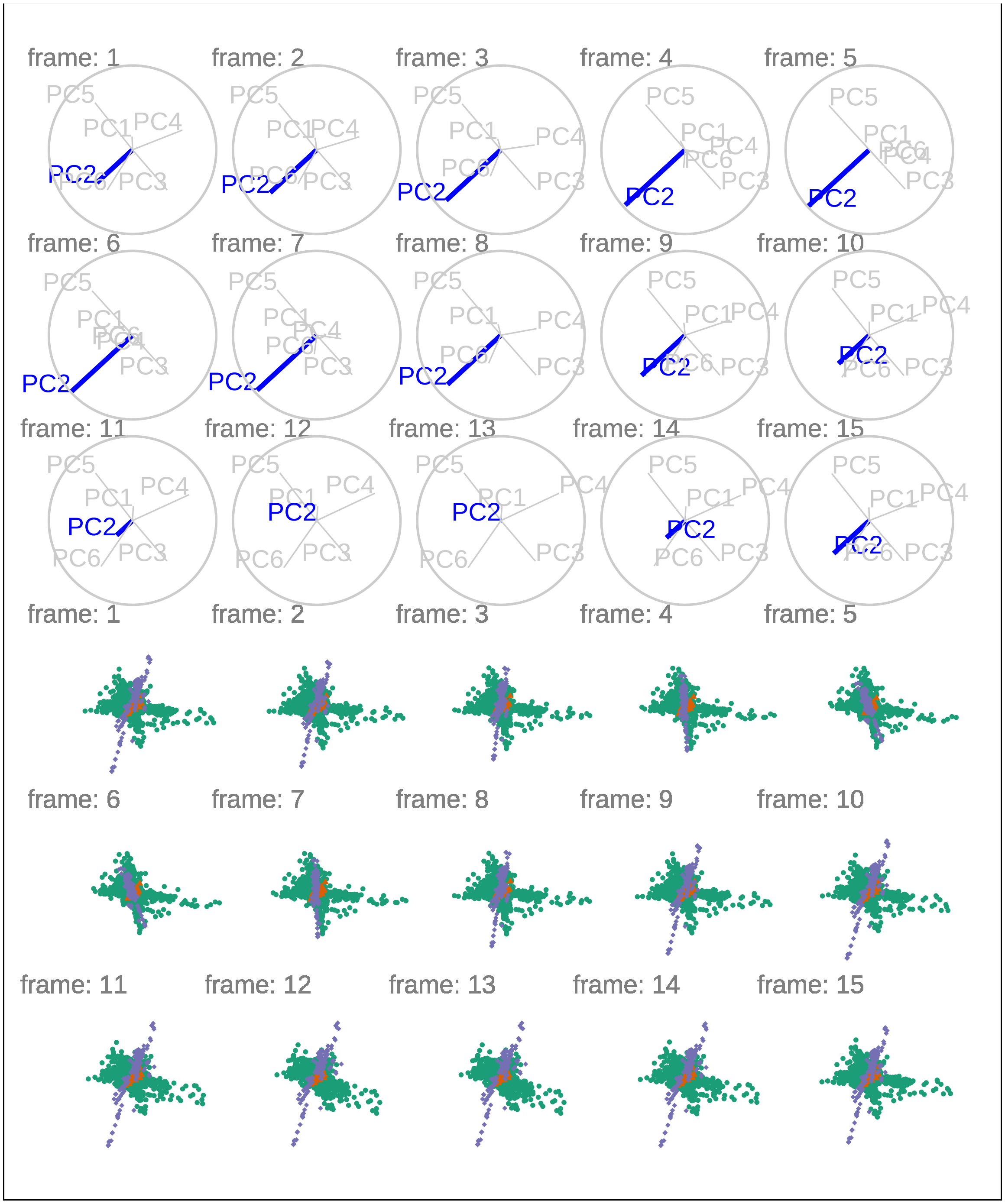

Figure 7: A radial manual tour manipulating the contribution of PC2 within the `DIS` cluster. Points are colored by experiment type: `DIS HERA1+2` in green, `dimuon SIDIS` in purple, and `charm SIDIS` in orange. The plane of cross distributed `DIS HERA` data (green) and a nearly orthogonal jet of `dimuon SIDIS` (purple) is present. This jet does extend more in the plane of view when the contribution of PC2 is full, giving insight to its orientation. However, less information about the shape of `DIS HERA` (green) and `charm SIDIS` (orange) is available. A dynamic version can be viewed at https://nspyrison.netlify.com/thesis/discluster_manualtour_pc2/.

the use of a **shiny** (Chang et al. 2018) application, would mitigate the barrier to entry, allow for more rapid analysis, and offer an approachable demo tool. The user could easily switch between variables to control, adjust interpolation step angle, or flag/save specific frame basis sets.

0.7 Acknowledgments

This article was created in R (R Core Team 2018), using **bookdown** (Xie 2016) and **rmarkdown** (Xie, Allaire, and Grolemund 2018), with code generating the examples inline. The source files for this article be found at github.com/nspyrison/confirmation/. The source code for the **spinifex** package can be found at github.com/nspyrison/spinifex/.

Bibliography

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

Buja, Andreas, Dianne Cook, Daniel Asimov, and Catherine Hurley. 2005. "Computational Methods for High-Dimensional Rotations in Data Visualization." In *Handbook of Statistics*, 24:391–413. Elsevier. https://doi.org/10.1016/S0169-7161(04)24014-7.

Chang, Winston, Joe Cheng, J. J. Allaire, Yihui Xie, and Jonathan McPherson. 2018. *Shiny: Web Application Framework for R.* https://CRAN.R-project.org/package=shiny.

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

Cook, Dianne, Ursula Laa, and German Valencia. 2018. "Dynamical Projections for the Visualization of PDFSense Data." *Eur. Phys. J. C* 78 (9): 742.

Cook, Dianne, Deborah F. Swayne, and A. Buja. 2007. *Interactive and Dynamic Graphics for Data Analysis: With R and GGobi.* Springer Science & Business Media.

Diaconis, Persi, and David Freedman. 1984. "Asymptotics of Graphical Projection Pursuit." *The Annals of Statistics* 12 (3): 793–815.

Furnas, George W., and Andreas Buja. 1994. "Prosection Views: Dimensional Inference Through Sections and Projections." *Journal of Computational and Graphical Statistics* 3 (4): 323–53. https://doi.org/10.2307/1390897.

Hurley, C., and A. Buja. 1990. "Analyzing High-Dimensional Data with Motion Graphics." *SIAM Journal on Scientific and Statistical Computing* 11 (6): 1193–1211. https://doi.org/10.1137/0911068.

Lubischew, Alexander A. 1962. "On the Use of Discriminant Functions in Taxonomy." *Biometrics*, 455–77.

Pedersen, Thomas Lin, and David Robinson. 2019. *Gganimate: A Grammar of Animated Graphics.* http://github.com/thomasp85/gganimate.

R Core Team. 2018. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Sievert, Carson. 2018. *Plotly for R.* https://plotly-book.cpsievert.me.

Wang, Bo-Ting, T. J. Hobbs, Sean Doyle, Jun Gao, Tie-Jiun Hou, Pavel M. Nadolsky, and Fredrick I. Olness. 2018. "Mapping the Sensitivity of Hadronic Experiments to Nucleon Structure." *Physical Review D* 98 (9): 094030.

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

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

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

Xie, Yihui, Christian Mueller, Lijia Yu, and Weicheng Zhu. 2018. *Animation: A Gallery of Animations in Statistics and Utilities to Create Animations*. https://yihui.name/animation.