Dynamic visualization of high-dimensional data via low-dimension projections and sectioning across 2D and 3D display devices

A thesis submitted for the degree of

Doctor of Philosophy

by

Nicholas S Spyrison

B.Sc. Statistics, Iowa State University

Department of Econometrics and Business Statistics

Monash University

Australia

January 2019

Contents

Acknowledgements

I would like to thank . . .

Declaration

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma in 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.

Nicholas S Spyrison

Preface

The material in Chapter 1 has been submitted to *Something interesting jornal* for possible publication.

The contribution in Chapter 3 of this thesis was presented in the super awesome confrence held in Dublin, Ireland, in July 2015.

Abstract

This thesis is about …

Chapter 1

Introduction

This is where you introduce the main ideas of your thesis, and an overview of the context and background.

In a PhD, Chapter 2 would normally contain a literature review. Typically, Chapters 3–5 would contain your own contributions. Think of each of these as potential papers to be submitted to journals. Finally, Chapter 6 provides some concluding remarks, discussion, ideas for future research, and so on. Appendixes can contain additional material that don't fit into any chapters, but that you want to put on record. For example, additional tables, output, etc.

Chapter 2

Literature review

2.1 Touring

2.1.1 Overview

In univariate data sets histograms, or smoothed density curves are employed to visualize data. In bivariate data scatterplots and contour plots (2-d density) can be employed. In three dimensions the two most common techniques are: 2-d scatter plot with the 3rd variable as an aesthetic (such as, color, size, height, *etc*.) or rendering the data in a 3-d volume using some perceptive cues giving information describing the seeming depth of the image [1]. When there are 4 variables: 3 variables as spatial-dimensions and a 4th as aesthetic, or a scatterplot matrix consisting of 4 histograms, and 6 unique combinations of bivariate scatterplots.

Let p be the number of numeric variables; how do we visualize data for even modest values of p (say 6 or 12)? It's far too common that visualizing in data-space is dropped altogether in favor of modeling parameter-space, model-space, or worse: long tables of statistics without visuals (Wickham, Cook, and Hofmann, 2015). Yet, we all know of the risks inherent in relying too heavily on parameters alone (Anscombe, 1973; Matejka and Fitzmaurice, 2017). So why do we move away from visualizing in data-space? Scalability,

[1] Graphs of data depicting 3 dimension are typically printed on paper, or rendered on a 2-d monitor, they are intrinsically 2-d images. They are sometimes referred to as 2.5-d, or more frequently erroneously referred to as 3-d, more on this later.

in a word, we are not familiar with methods that allow us to concisely depict and digest $p \geq 5$ or so dimensions. This is where dimensionality reduction comes in. Specifically, we will be focusing on a specific group called touring. In the interest of time I will not belabor the diversity of dimensionality reduction, (see [Grinstein, Trutschl, and Cvek (2002); Carreira-Perpinán (1997); heer_tour_2010] for a quick summary). Suffice it to say that touring has a couple of salient features: linear transformations such that we can interpolate back to the original variable space and does not discard dimensions, something that is common to other linear techniques. By employing the breadth of tours we are able to preserve the visualization of data-space, and with it, the intrinsic understanding of structure and distribution of data that is more succinct or beyond the reach of statistic values alone.

Touring is a linear dimensionality reduction technique that orthogonally projects p-space down to $d(\leq p)$ dimensions. Many such projections are interpolated, each making local rotations in p-space. These frames are then viewed in order to the effect of watching an animation of the lower dimensional embedding changing as p-space is manipulated. Shadow puppets offer a useful analogy to aid in conceptualizing touring. Imagine a fixed light source facing a wall. When a hand or puppet is introduced the 3-dimensional object projects a 2-dimensional shadow onto the wall. This is a physical representation of a simple projection, that from $p = 3$ down to $d = 2$. If the object rotates then the shadow correspondingly changes. Observers watching only the shadow are functionally watching a 2-dimensional tour as the 3-dimensional object is manipulated.

Terminology

n, p (sometimes called d by Wegman, or n), d (sometimes called k by wegman, or d in tourr)

2.1.2 History

Touring was first introduced by Asimov in 1985 with his purposed Grand Tour(Asimov, 1985) at Stanford University. In which, Asimov suggested three types of Grand Tours:

torus, at-random, and random-walk. The specifics of which will be discussed below in the Typology section.

TALK ABOUT maths Here::

Note that the the above methods have no input from the user aside from the starting basis. The bulk of touring development since has largely been around dynamic display, user interaction, geometric representation, and application.

This works well when the number of dimensions being toured is small (in the neighborhood of 5-10), yet the number of view, or 2-frames and we can produce from p-space suffers from the so called blessing/curse of dimensionality. In which the plethora of degrees of freedom either offer many (non-unique) solutions to a problem or something that becomes ever increasing unlikely,

2.1.3 Tour path

A fundamental aspect of touring is the path of rotation. Of which there are four primary distinctions(Buja et al., 2005): random choice, precomputed choice, data driven, and manual control.

- *grand tour*, a constrained random choice p-space. Paths are constrained for changes in direction small enough to maintain continuity and aid in user comprehension
  - torus-surface (Asimov, 1985)
  - Geodesic
  - at-random
  - random-walk
  - *local tour*, a sort of grand tour on leash, such that it goes to a nearby random projection before returning to the original position and iterating
- *guided tour*, data driven tour optimizing some objective function via (stochastic) gradient descent (Hurley and Buja, 1990).
  - holes (Cook, Buja, and Cabrera, 1993) - iterates projections that add more white space to the center of the projection.

- cmass (Cook, Buja, and Cabrera, 1993) - find the projection with the most density or mass in the center.
- lda (Lee et al., 2005) - linear discriminant analysis, seeks a projection where 2 or more classes are most separated.
- pda - principal component analysis finding where the data is most spread (1d only).
- other user-defined objective function (Wickham et al., 2011).

- *planned tour*, Precomputed choice, In which the path has already been generated or defined.
  - *little tour* (McDonald, 1982), where every permutation of variables is stepped through in order, analogous to a brute-force or exhaustive search.
  - a saved path of any other tour
- *manual tour* - Manual control, a constrained rotation on selected manipulation variable and magnitude(Cook and Buja, 1997). Typically used to explore the local area after identifying an interesting feature from another tour.
- *dependance tour*, combination of n independent 1d tours. A vector describes the axis each variable will be displayed on. **ie** $c(1, 1, 2, 2)$ is a 4 to 2d tour with the first 2 variables on on the first axis, and the remaining on the second.
  - *correlation tour* (Buja, Hurley, and McDonald, 1987), a special case of the dependence tour, analogous to canonical correlation analysis

2.1.4 Geometrics and display dimension

Up to this point we have been talking about 2d scatterplots, which offer the first and a simple case for viewing lower-dimensional embeddings of p-space. However, other geometrics (or geoms) offer perfectly valid orthonormal projections as well.

- 1d geoms

  - 1-d densities: such as histogram, average shifted histograms(**scott85**), and kernel density(**scott95**).
  - image: (**Wegman**)
  - time series: where multivariate values are independently lagged to view peak and trough alignment. Currently no package implementation, but use case is discussed in (Cook and Buja, 1997).

- 2d geoms
  - 2-d density (**NS**)
  - scatterplot
  -

- 2.5d, 3d geoms {ADD FOOTNOTE ABOUT 2.5d vs 3d}
  - Anaglyphs, sometimes called stereo, where (typically) red images are positioned for the left channel and cyan for the right, when viewed with corresponding filter glasses give the depth perception of the image.
  - Depth, which use some subset of depth cues, most commonly size and/or color of data points.

- *d*-dim geoms
  - Andrews curves (Andrews, 1972), smoothed variant of parallel coordinate plots, discussed below.
  - Chernoff faces (Chernoff, 1973), variables linked to size of facial features for rapid cursory like-ness comparison of observations.
  - Parallel coordinate plots (Ocagne, 1885), where any number of variables are plotted in parallel with observations linked to their corresponding variable value by polylines.
  - Scatterplot matrix (Becker and Cleveland, 1987), showing a triangle matrix of bivariate scatterplots with 1-d density on the diagonal.
  - Radial glyphs, radial variants of parallel coordinates including radar, spider, and star glyphs (Siegel et al., 1972).

2.1.5 Aplication

Below is a non-exhaustive list of software implementing touring in some degree, ordered by descending year:

- Spinifex (**spinifex**) – for Linux, Unix, and Windows.
- Tourr (Wickham et al., 2011) – for Linux, Unix, and Windows. R package.
- CyrstalVision (Wegman, 2003) – for Windows.
- GGobi (Swayne et al., 2003) – for Linux and Windows.
- DAVIS (Huh and Song, 2002) – Java based, with GUI.
- VRGobi (Nelson, Cook, and Cruz-Neira, 1998) – for use with the C2 in stereoscopic 3d displays.
- ExplorN (Carr, Wegman, and Luo, 1996) – for SGI Unix.
- XGobi (Swayne, Cook, and Buja, 1991) – for Linux, Unix, and Windows (via emulation).
- XLispStat (Tierney, 1990) – for Unix, and Windows.
- Prim-9 (Asimov, 1985; Fisherkeller, Friedman, and Tukey, 1974) – on an internal operating system.

Support and maintenance of such implementations give them a particularly short life span, while conceptual abstraction and technically heavier implementations have hampered user growth. There have been notable efforts to diminish the barriers to entry and make touring more approachable as a data exploration tool [Huh and Song (2002); Swayne et al. (2003); Wegman (2003); Wickham et al. (2011); huang_tourrgui:_2012].

2.2 Virtual reality

Chapter 3

spinifex: extending *tourr* with manual tours and graphic display

3.1 Abstract

Touring techniques offer a great opportunity for data-space visualizations of ($\mathbf{X} \in \mathbb{R}^p,\ p > 3$) multivariate data sets. This paper discusses the *R* package *spinifex*, which adds support for the manual tour, which is particularly usefully for exploring the local structure after identifying a feature of interest, perhaps via guided tour. Additionally, *spinifex* extends graphic outputs to *plotly* and *gganimation*. This work extends the functionality of and is compatible with *tourr*.

Keywords: grand tour, projection pursuit, manual tour, tourr, touring, high dimensional visualization, high dim vis, dimensionality reduction, visualization, statistical graphics, data-space.

3.2 Introduction

Both classical and contemporary visualizations of data are presented in ($d = 2$) two dimensions, that of a computer monitor or in print. How is it that we come to view and share data that exists in $p > 3$ dimensions? In an appeal to brevity we shall ignore model and parameter summarization due to there shortcomings (Anscombe, 1973; Matejka and Fitzmaurice, 2017). Within the realm of data-space visualization we are left with projecting higher volumes and embedding them within lower dimensional spaces that we can visualize.

This is not a new phenomena, such linear projections have been in use for quite some time. (Pearson, 1901; Fisher, 1936) and the myriad of single value decomposition (SVD) techniques from numerous disciplines use such embeddings. Previous application look at data in one (or few) static orientations, after some objective optimization. For instance in PCA, we reorient p-dimensions such that we have a reference to the ordered components that describe a descending amount of variation held within the data. Yet we still have p components remaining to visualize. Where does the dimension reduction come in? From plotting only the first two or three and potentially another dimension tied to data point aesthetic. This is maximizes the amount of variation that can be display in an embedding, but regularly discards a large proportion of the variation held within the data.

More recently non-linear dimensionality techniques have become popular, such as t-distributed stochastic neighbor embedding (t-SNE) (Maaten and Hinton, 2008), building off of Sammon mappings (Sammon, 1969). Such non-linear methods make for astounding distinction when in lower embeddings, but contain inherent shortcomings. Namely: that the non-linear transformations break inter-operability back to the original data-space, and that they can suffer from overfitting. If there is no inherent clustering with the data, it's possible that noise within the variables may become the prominent feature and be displayed erroneously as group clustering.

TODO: Clean up touring, , below

(Asimov, 1985; Buja and Asimov, 1986) first suggested grand tours in which random walks in p-space can be interpolated and embedded in d dimensions which are then viewed in sequence. Imagine. Consider,

The broader scope of touring has some beneficial features, namely: touring keeps the original dimensionality in tact unlike tradition static linear-projections, and maintains inter-operability back to the original dimensions, a primary drawback of non-linear dimensionality reduction.

TODO: clean up above, talk about touring in general first

3.3 Manual tour

The manual tour(Cook and Buja, 1997; Cook et al., 2008) allows the user rotate a specified manipulation variable into and out of the current projection. This reveals the extent that this variable contributes to the current structure of the projection. In an appeal to the so called curse/blessing of dimensionality, volume contained within $\mathbb{R}^p$ increases exponentially as p increases. Exploring p-space by defining specific rotations on specific variables quickly becomes time-prohibited.

However, manual touring can be particularly useful in exploring the local structure once a feature of interest has been identified. Features of interest can be quickly identified by a guided tour(Hurley and Buja, 1990). In guided tours an index of interest is defined and gradient descent is performed on the projection, analogous to projection pursuit (Friedman and Tukey, 1974).

Let's explore the process behind the manual tour:

Given:

$\mathbf{X}_{[n,\,p]}$ A data set containing n observations of p numeric variables.

$\mathbf{B}_{[p,\,d]}$ An orthonormal [1] basis describing the current orientation projecting p down to d dimension.

$$\mathbf{X}_{[n,\,p]} = \begin{bmatrix} X_{1,\,1} & \dots & X_{1,\,p} \\ X_{2,\,1} & \dots & X_{2,\,p} \\ \vdots & \ddots & \vdots \\ X_{n,\,1} & \dots & X_{n,\,p} \end{bmatrix}$$

[1] Where each variable is both: orthogonal, at right angles (dot product is 0) to the other variables, and unit vectors, a norm = 1

$$\mathbf{B}_{[p,\ d]} = \begin{bmatrix} B_{1,\ 1} & \dots & B_{1,\ d} \\ B_{2,\ 1} & \dots & B_{2,\ d} \\ \vdots & \ddots & \vdots \\ B_{p,\ 1} & \dots & B_{p,\ d} \end{bmatrix}$$

For ease of computation we will be working mostly with the basis and not the data, once basis manipulation is done post multiply the data by the basis to get back to data-space.

Select a manipulation variable, k. Initialize a zero vector e, and set the k-th element set to 1. Use the Gram-Schmidt process to orthonormalize the concatenation of the basis and e yielding the manipulation space.

$$\begin{aligned} \mathbf{M}_{[p,\ d+1]} &= Orthonormalize_{GS}(\mathbf{B}_{[p,\ d]}|\mathbf{e}_{k\ [p,\ 1]}) \\ &= Orthonormalize_{GS}\left(\begin{bmatrix} B_{1,\ 1} & \dots & B_{1,\ d} \\ B_{2,\ 1} & \dots & B_{2,\ d} \\ \vdots & \ddots & \vdots \\ B_{k,\ 1} & \dots & B_{k,\ d} \\ \vdots & \ddots & \vdots \\ B_{p,\ 1} & \dots & B_{p,\ d} \end{bmatrix} \quad | \quad \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \right) \end{aligned}$$

TODO: (below) clean up the language of phi, tie to manipulation var. Make sure to paint the picture for the Z-direction of manip_sp.

Select a vector ϕ_i, the angle of out-of plane rotation, orthogonal to the projection plane (relative to ϕ_1, the transformation ϕ_i - ϕ_1 proved to be helpful to discuss ϕ relative to the Z axis).

For i in 1 to n_slides:

For each ϕ_i, post multiply the manipulation space by a rotation matrix, producing as many basis-projections.

$$\mathbf{P}_{b[p,\ d+1,\ i]} = \mathbf{M}_{[p,\ d+1]} \ * \ \mathbf{R}_{[d+1,\ d+1]}$$

For the $d = 2$ case:

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

Where:

θ is the angle that lies on the projection plane (*ie*. on the XY plane)

ϕ is the angle orthogonal to the projection plane (*ie*. in the Z direction)

c_θ is the cosine of θ

c_ϕ is the cosine of ϕ

s_θ is the sine of θ

s_ϕ is the sine of ϕ

To get back to data-space post multiply each projection basis by the data, for the data projection.

$$\mathbf{P}_{d[n,\ d+1]} = \mathbf{X}_{[n,\ p]} \ * \ \mathbf{P}_{b[p,\ d+1]} \tag{3.1}$$

$$= \begin{bmatrix} X_{1,\ 1} & \dots & X_{1,\ p} \\ X_{2,\ 1} & \dots & X_{2,\ p} \\ \vdots & \vdots & \vdots \\ X_{n,\ 1} & \dots & X_{n,\ p} \end{bmatrix}_{[n,\ p]} * \begin{bmatrix} P_{b:1,\ 1} & P_{b:1,\ 2} & P_{b:1,\ 3} \\ P_{b:2,\ 1} & P_{b:2,\ 2} & P_{b:2,\ 3} \\ \vdots & \vdots & \vdots \\ P_{b:p,\ 1} & P_{b:p,\ 2} & P_{b:p,\ 3} \end{bmatrix}_{b[p,\ d+1]} \tag{3.2}$$

Plot the first two variables from each projection in sequence for an XY scatterplot. The remaining variable is sometimes linked to a data point aesthetic to produce depth cues used in conjunction with the XY scatterplot.

3.4 Usage

3.4.1 Flea

Let's start off with the flea data set from the R package *tourr* (Wickham et al., 2011), which performs different tours on the same data. The flea data contains 74 observations of flea beetles across 6 variables, physical measurements of the flea. Each individual belonged to one of three species being observed.

Let:

$d = 2$ For the sake of illustration we'll embed into two dimensions.

Given:

$\mathbf{X}_{[74,\ 6]}$ Flea data, 74 observations by 6 variables, $\mathbf{X} \in \mathbb{R}^{\mathbf{6}},\ \mathbf{p} = \mathbf{6}$.

$p = 6$

Let's initialize a random orthonormal basis of dimensions $[p,\ d]$, which describes a random orientation projected from six down to two dimensions. Check how each of the dimensions is contributing the XY components with `view_basis()`

```
##              X        Y norm_XY   theta
## [1,]  0.08073  0.45268 0.45982  1.3943
## [2,]  0.49166 -0.68137 0.84023 -0.9457
## [3,]  0.35320 -0.24222 0.42828 -0.6011
## [4,]  0.76457  0.51785 0.92344  0.5953
## [5,]  0.20421 -0.05406 0.21125 -0.2588
## [6,] -0.02704  0.03259 0.04235 -0.8783
```

```
##              X        Y norm_XY   theta
## [1,]  0.14689 -0.07195  0.1636 -0.4555
```

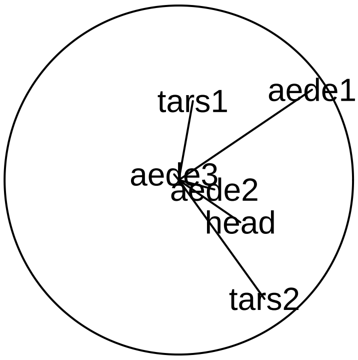

Figure 3.1: *Random basis, flea data*

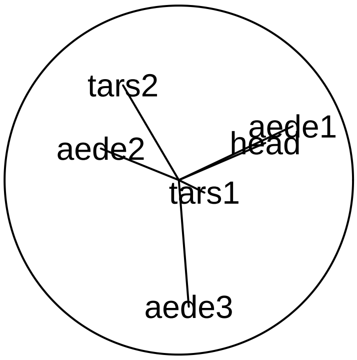

Figure 3.2: *Random basis, flea data*

```
## [2,] -0.32013  0.54295  0.6303 -1.0380
## [3,]  0.49601  0.21246  0.5396  0.4047
## [4,]  0.65298  0.30858  0.7222  0.4415
## [5,] -0.44751  0.18089  0.4827 -0.3841
## [6,]  0.05723 -0.72591  0.7282 -1.4921
```

3.4.2 Breast Cancer

Let's apply a manual tour to the Wisconsin Breast Cancer Database, formated from the machine learning benchmarking data sets in the *R* package (Leisch and Dimitriadou, 2010).

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    675 obs. of  10 variables:
##  $ Id             : Factor w/ 630 levels "1000025","1002025",..: 1 4 5 6 8 9 10 11 15
##  $ Cl.thickness   : num  5 5 3 6 4 8 1 2 2 4 ...
##  $ Cell.size      : num  1 4 1 8 1 10 1 1 1 2 ...
##  $ Cell.shape     : num  1 4 1 8 1 10 1 2 1 1 ...
##  $ Marg.adhesion  : num  1 5 1 1 3 8 1 1 1 1 ...
##  $ Epith.c.size   : num  2 7 2 3 2 7 2 2 2 2 ...
##  $ Bare.nuclei    : num  1 10 2 4 1 10 10 1 1 1 ...
##  $ Bl.cromatin    : num  3 3 3 3 3 9 3 3 1 2 ...
##  $ Normal.nucleoli: num  1 2 1 7 1 7 1 1 1 1 ...
##  $ Class          : Factor w/ 2 levels "benign","malignant": 1 1 1 1 1 2 1 1 1 1 ...
```

TODO: Address figure output. PDF output is a static image w/ play slider.

Chapter 4

Display dimensionality

- XGobbi vs the C2

Chapter 5

Human-computer interaction of 3d projections

- Tour in 3D
- ImAxes / IATK

Appendix A

Additional stuff

You might put some computer output here, or maybe additional tables.

Note that line 5 must appear before your first appendix. But other appendices can just start like any other chapter.

Bibliography

Andrews, DF (1972). Plots of High-Dimensional Data. *Biometrics* **28**(1), 125–136. (Visited on 12/19/2018).

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

Asimov, D (1985). The grand tour: a tool for viewing multidimensional data. *SIAM journal on scientific and statistical computing* **6**(1), 128–143.

Becker, RA and WS Cleveland (1987). Brushing Scatterplots. *Technometrics* **29**(2), 127–142. (Visited on 01/10/2019).

Buja, A and D Asimov (1986). Grand Tour Methods: An Outline. In: *Proceedings of the Seventeenth Symposium on the Interface of Computer Sciences and Statistics on Computer Science and Statistics*. New York, NY, USA: Elsevier North-Holland, Inc., pp.63–67. http://dl.acm.org/citation.cfm?id=26036.26046 (visited on 01/09/2019).

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

Buja, A, C Hurley, and JA McDonald (1987). A data viewer for multivariate data. In: *Colorado State Univ, Computer Science and Statistics. Proceedings of the 18 th Symposium on the Interface p 171-174(SEE N 89-13901 05-60)*.

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

Carreira-Perpinán, MA (1997). A review of dimension reduction techniques. *Department of Computer Science. University of Sheffield. Tech. Rep. CS-96-09* **9**, 1–69.

Chernoff, H (1973). The Use of Faces to Represent Points in K-Dimensional Space Graphically. *Journal of the American Statistical Association* **68**(342), 361–368. (Visited on 01/05/2019).

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).

Cook, D, A Buja, and J Cabrera (1993). Projection Pursuit Indexes Based on Orthonormal Function Expansions. *Journal of Computational and Graphical Statistics* **2**(3), 225–250. (Visited on 01/07/2019).

Cook, D, A Buja, EK Lee, and H Wickham (2008). "Grand Tours, Projection Pursuit Guided Tours, and Manual Controls". en. 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).

Fisher, RA (1936). The use of multiple measurements in taxonomic problems. *Annals of eugenics* **7**(2), 179–188.

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

Friedman, J and J Tukey (1974). A Projection Pursuit Algorithm for Exploratory Data Analysis. en. *IEEE Transactions on Computers* **C-23**(9), 881–890. (Visited on 06/22/2018).

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

Huh, MY and K Song (2002). DAVIS: A Java-based Data Visualization System. en. *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).

Lee, EK, D Cook, S Klinke, and T Lumley (2005). Projection Pursuit for Exploratory Supervised Classification. *Journal of Computational and Graphical Statistics* **14**(4), 831–846. (Visited on 01/07/2019).

Leisch, F and E Dimitriadou (2010). *mlbench: Machine Learning Benchmark Problems*.

Maaten, Lvd 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. en. 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).

McDonald, JA (1982). INTERACTIVE GRAPHICS FOR DATA ANALYSIS.

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. en. *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, ...* French. OCLC: 458953092. Paris: Gauthier-Villars.

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.

Sammon, JW (1969). A nonlinear mapping for data structure analysis. *IEEE Transactions on computers* **100**(5), 401–409.

Siegel, JH, EJ Farrell, RM Goldwyn, and HP Friedman (1972). The surgical implications of physiologic patterns in myocardial infarction shock. English. *Surgery* **72**(1), 126–141. (Visited on 01/05/2019).

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

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).

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

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

Wickham, H, D Cook, and H Hofmann (2015). Visualizing statistical models: Removing the blindfold: Visualizing Statistical Models. en. *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. en. *Journal of Statistical Software* **40**(2). (Visited on 11/23/2018).