doi.org/10.26398/IJAS.0029-016

DATA ANALYSIS: GOOD BUT …

John C. Gower[1]

Department of Mathematics and Statistics, The Open University, UK

Sugnet Gardner-Lubbe, Niël J. Le Roux

Department of Statistics and Actuarial Science, Stellenbosch University, RSA

Abstract. *In the 1950s Tukey and Benzécri led the Data Analysis movement, opposed to the perceived mathematisation of statistics. This has flourished but we perceive some troubling problems (Greenacre prefers challenges or opportunities) which are addressed in the paper. These problems are manifest in the simplest analyses of two-way arrays of data (Sections 2 - 3) and become worse with higher order arrays (Section 4). The most important thing about Data Analysis is the Data, its type (e.g. how data is collected, the physical kinds of variable, categories, counts, etc.) and the data structure (e.g. arrays, multiway tables, symmetry etc.). Analysis is concerned with models, distances, norms, measures of approximation and algorithms. Perhaps Data Analysis is in some danger of replicating the kind of mathematisation it was designed to supplant.*

Keywords*: Visualisation, Data, Algorithms, Data-structure*

1. INTRODUCTION

This paper arose from a presentation titled Bees in my Bonnet[2] that JCG gave at the 2015 Naples CARME conference but it also draws on the material discussed by NJL and SGL at the same conference. This version retains some of that material but is more concerned with issues involving the current position of the practice of Data Analysis. In no way is it an attempt to give an exhaustive review of developments in Data Analysis. Perhaps the following may be regarded as a State of the Union address, but rather than focusing on the undoubted successes of Data Analysis we shall focus on some problems (pace Greenacre) concerned with current practice. In Figure 1 we see the Millionaire calculator used by Fisher nearly 100 years ago. When Fisher said, perhaps apocryphally, I learned all my statistics at the computer he was not so much saying that it was the calculating process itself which gave him insights but it was the opportunity given for a close examination of the data. One

---

1 Corresponding author: John C. Gower, email: j.c.gower@open.ac.uk
2 To keep talking about something again and again because you think it is very important.

of us (JCG) recalls Yates, who succeeded Fisher at Rothamsted, talking about "data-sniffing" in which to great advantage he ran his fingers over a set of data and sniffed out errors or inconsistencies. So data analysis is very old and need not require sophisticated equipment.

We should state at the outset that our position is that the term Data Analysis is a synonym for Statistics and it was a reaction to the mathematisation of statistics in the nineteen fifties (or thereabouts), when the work of John Tukey (1962) in the USA and Jean-Paul Benzcri (1973) in France led to a resurgence in the analysis of data and the name Data Analysis. Along with an increased interest in data has been the phenomenal rise of computing, not forgetting the role that computers have played in supporting the possibilities for generating visualizations. Mathematics has a place in many disciplines (e.g. physics, engineering, astronomy, statistics,) and almost every scientist finds computers useful but we should recognise that this does not necessarily make them mathematicians or IT experts, although a few individuals have excelled in these fields.

Similar remarks pertain to visualization where geometric ideas are often supporting innovating ideas which are later developed in algebraic form (e.g. Fisher's derivation of several well-known probability distributions (e.g. Fisher, 1953), Karl Pearson's fits of lines and planes (Pearson, 1901), Stephen Hawking with cosmological models (see Mialet, 2012), Lew Pontryagin with algebraic and differential topology (Pontryagin, 1966). The place of visualization in scientific and mathematical innovation is one thing, and indeed one where some have strong opinions, but in the following we shall be more concerned with the visualization of data.

Perhaps more so in statistics than other disciplines, the benefits brought by computers have quickly become readily accessible but there is a perception that many users of the now readily available software are often not fully aware of its limitations nor in a position to give informed critical comment either on the numerical part of an analysis or on any associated visualizations. Ioannidis (2008) refers to severe defects in the medical/pharmaceutical literature and Gower, Groenen, Van de Velden and Vines (2014) comment similarly on the marketing literature but the shortcomings are widespread. The actual collection of data, such as in Sample Surveys and Designed Experiments gets lamentable attention with the current focus seeming to be on accumulating large amounts of data rather than on efforts to improve its quality. Thus, the general research population, which contains few persons well-versed in data analysis, now has ready access to software. Software may be misused in several ways – users may not understand it properly, they may use inappropriate default settings, the software itself may be defective, its

Figure 1: Fisher - I learned all my statistics at the computer. Reproduced with permission of Rothamsted Research.

accompanying user-manual may be inadequate or wrong, visualizations may be mis-scaled, fitted values may be inaccurate. The position is similar to the way the general population self-medicates, informed by astute advertising; the result may be good or harmless or disastrous. It is against this background that the following surveys where we think that there are problems (i.e. challenges and even opportunities) which we discuss in more detail below.

2. VISUALIZATIONS

Visualizations are often poor and sometimes very poor and on occasions they are deliberately misleading. The strong temptation to see patterns in random configurations of points is well-known (see Apophenia in Wikipedia). The risk of apophenia when using measured variables is very real but it is even worse with non-observable virtual variables (we use the term 'virtual variables' and not 'latent variables' to avoid the clash with 'latent vectors' – see the section about 'confusing of words'), where researchers have to be on special guard. Far from being opposed to the use of visualizations, even with virtual variables, we are enthusiastic users but we are also aware of its dangers.

Happily, visualizations have become common but they are often defective and often hard to interpret. We confess that, after a time lapse, we sometimes have difficulties in interpreting our own visualizations. As well as the difficulties about too much ink, inadequate labelling, unnecessary three-dimensional presentations of two-dimensional objects like pie charts and histograms etc. that Tufte (1983, second edition 2001) described so well in his classic book, there are many more worries that can be laid at the door of Data Analysis. In data analysis, a frequent problem concerns wrong aspect ratios either determined by the software or sometimes by editors who want figures to fit comfortably onto the pages of a journal. Figures that approximate distances or inner-products or use other measures which depend on angles, cannot be rescaled at will. Other aspects that deserve attention are:

(i) Axes are often placed concurrently at an origin which is at the centroid of exhibited sample points. Such axes can confusingly intermingle with the sample points and associated labelling shown in a visualization. Concurrency is unnecessary as axes may be shifted independently towards the margins of a visualization (Blasius, Eilers and Gower, 2009).

(ii) Numerical information should accompany every visualization to indicate where approximations are good and where they are poor (Gardner-Lubbe, le Roux and Gower, 2008).

Gower, Groenen, Van de Velden and Vines (2014) made some provisional proposals for a series of icons that could accompany every published graph, to give guidance in a coded form, on suitable measures (distances and angles among others) available for correct interpretation. Visualizations that use icons are referred to as being self-defining. Note that a system of icons not only helps define correct interpretation but also alerts an inexperienced reader to the fact that there may be something that needs attention.

Another of Tufte's displeasures was the unnecessary and inappropriate use of words, especially acronyms. Some of our favourite examples from the field of Data Analysis follow:

(i) The use of the word Classification (for forming classes) which should be carefully distinguished from the same word classification used to assign to classes (Discriminant Analysis in statistics).

(ii) A sample refers to a single example but in mathematical writings it is often a shorthand for denoting a set of samples? We had a major misunderstanding with a referee about this!

(iii) Acronyms should be used sparingly. In particular, ALS refers to an algorithm when it is the substantive method that matters more. The same applies to EM, IWLS and many more. In this paper we have found acronyms inescapable but have tried to give the full name on its first appearance.

(iv) We acknowledge the genesis of “inertia” and “mass” in applied mathematics, but they add little to the terms sum-of-squares and sample size, long used in statistics.

3. DATA STRUCTURE

Computer software does not usually, if ever, give information on any limitations to be placed on the suitability of data for which it may be used (Nishisato, 1994). Many Data Analysis methods analyse two-way tables but computer languages are concerned mainly with array dimensions. Statisticians are well-aware that a two-way table is not the same as a data matrix; nested and crossed (and other) classifications are crucial; as well as Boolean variables, counts, measured variables (ratio and interval), mixed variables, (ordered) categorical variables have to be distinguished; diverse kinds of symmetric and asymmetric matrix are common. It is easy for a non-statistical user of software to plug-in all these kinds of data and get output from the computer, unaware that it may be complete nonsense or defective to some degree.

We shall use Principal Components Analysis (PCA) to draw attention to some problems that occur throughout data analysis. PCA is one of the foundations of Data Analysis but, despite its popularity and undoubted virtues, it is fraught with problems. PCA is concerned with a data matrix $\boldsymbol{X}$ with n rows and p columns. The rows refer to samples or cases and the columns to variables. $\boldsymbol{X}$ has come to be known as a data matrix. PCA was described by Karl Pearson (1901) as “*On Lines and Planes of Closest Fit to Systems of Points in Space*” showing a strong geometric attachment to data analysis. Of course, what Pearson was doing was to solve the least-squares minimisation problem $min \| \boldsymbol{X} - \hat{\boldsymbol{X}} \|^2$ where $\hat{\boldsymbol{X}}$ is a rank r approximation to $\boldsymbol{X}$**.** Indeed $\hat{\boldsymbol{X}}$ is given by the orthogonal projection of $\boldsymbol{X}$ onto an r-dimensional plane. Pearson believed that his geometric approach “can be easily applied to numerical problems” though calculations become ‘cumbersome’ for four or more variables. Indeed, given modern computing power, the approximation $\hat{\boldsymbol{X}}$ sought by Pearson can be found by the singular value decompostion (SVD) of $\boldsymbol{X}$ for a large number of variables although the latter is often arrived at via a simple algebraic eigenvalue algorithm operating on the inner-product $\boldsymbol{X}^T\boldsymbol{X}$. This approach has been at the bottom of much misunderstanding, ever since Hotelling (1933) wrote of "*Analysis of a Complex of Statistical Variables into Principal Components*" in which he wished to solve the problem $min \| \boldsymbol{X}^T\boldsymbol{X} - \hat{\boldsymbol{X}}^T\hat{\boldsymbol{X}} \|^2$ which is concerned with fitting a symmetric correlation matrix. Pearson was concerned with approximating $\boldsymbol{X}$ while Hotelling was approximating $\boldsymbol{X}^T\boldsymbol{X}$. To use PCA to describe both problems is a confusion compounded by the fact that the same eigenvalue algorithm may be used for both minimisations. There is another issue discussed by Bailey and Gower (1990) who points out how the double entry of each off-diagonal term of a symmetric matrix weighs the least-squares fit. This is admissible when the eigenvectors of $\boldsymbol{X}^T\boldsymbol{X}$ are used as a step in solving Pearson's PCA or indeed the SVD but strictly speaking it is suboptimal for Hotelling's PCA. At the base of this muddle is that the same algorithm is applied to two different data structures. It is also why some PCA software highlights $\hat{\boldsymbol{X}}$ (Pearson) and some highlights the eigenvectors $\boldsymbol{V}$ of $\boldsymbol{X}^T\boldsymbol{X}$ (Hotelling). This problem explodes when three-way and higher order structures are considered (see Section 4, below).

The initial scaling of raw data is very important indeed. For obvious reasons Hotelling's analysis of correlations implies that each variable is normalised to zero mean and unit sum-of-squares. Note that there is no question of normalising the row (or column) sums of $\boldsymbol{X}^T\boldsymbol{X}$. The variables of Pearson's PCA are also minimised to have zero mean but this is most easily seen as a response to the discovery by Huygens that the best fitting least squares plane to $\boldsymbol{X}$ passes through the centroid of the n points. In addition, when the p variables have different scales of measurement, some kind of initial scaling is necessary to induce commensurability. Normalising by dividing by the square root of a unit sum-of-squares is often used but is not obligatory; for example, among other possibilities, with positive measurements a logarithmic transformation is invariant to changes in ratio-scale measurements, and therefore has much to commend it.

The above shows some of the things that should be borne in mind when contemplating what may look like a PCA problem. Even within the confines of two-way arrays, many of the problems seen with PCA remain, possibly in broadened form:

(i) A two-way structure may be a data matrix (asymmetric), or a table whose rows and columns are interchangeable (symmetric), or it may be of within-between type in which row-totals are irrelevant and columns may be of different lengths. Our use of the words 'asymmetric' and 'symmetric' to describe different or same treatments of rows and columns should be carefully distinguished from the use of 'asymmetric' and 'symmetric' to describe different forms of matrix – another potential confusion of the use of words which requires attention.

(ii) Data types are of the upmost importance. Numerical values may be on ratio scales or interval scales, categorical variables may be ordinal or nominal and both may be coded in different ways and transformed into numerical scores. Table 1 is a modified form of a figure given in Gower, Gardner-Lubbe and le Roux (2016) who give more detailed information on variant ways of coding categorical variables.

(iii) The choice of metric may be seen as part of the initial transformation of raw data (e.g. Correspondence Analysis (CA) chi-squared metric, Canonical Variate Analysis (CVA) Mahalanobis metric, Optimal scores for categorical variables, the L_1, or any other, norm are potential candidates for scaling). For example a PCA of $\boldsymbol{R}^{-1}\boldsymbol{X}\boldsymbol{C}^{-1/2}$ for row chi-squared metric (or $\boldsymbol{R}^{-1/2}\boldsymbol{X}\boldsymbol{C}^{-1}$ for column chi-squared metric) gives one of many variants of CA (see e.g. Gower, Lubbe and le Roux, 2011). Similarly, the two-sided eigenvalue equation $\boldsymbol{BZ} = \boldsymbol{AZ}\boldsymbol{\Lambda}$ with identification constraint t $\boldsymbol{Z}^{\boldsymbol{T}}\boldsymbol{AZ} = \boldsymbol{I}$, implies that t $\boldsymbol{Z}^{\boldsymbol{T}}\boldsymbol{BZ} = \boldsymbol{\Lambda}$, and that $\boldsymbol{ZZ}^{\boldsymbol{T}}= \boldsymbol{A}^{-1}$, so giving the Mahalanobis metric. When the between group sums-of-squares and products matrix $\boldsymbol{B}= \boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}$, then $\boldsymbol{XZ}$ is a PCA of the between groups means which have inter-distances in the Mahalanobis metric (for further details see Gower, Lubbe and Le Roux, 2011). The point which we wish to make here is that many multivariate methods are essentially the PCA of a non-arbitrary transformation of the initial raw data-set. Note that $\boldsymbol{R}^{-1/2}\boldsymbol{X}\boldsymbol{C}^{-1/2}$ gives another version of CA which treats rows and columns symmetrically as discussed in (i), above.

(iv) The identification constraint of CVA points to another area of unease. The orthogonality of the eigenvectors of a symmetric matrix is a mathematical fact but the length of the vectors is more arbitrary. With CVA and CA and PCA itself the justification for unit standardisation is clear but, especially in methods of analysis that depend on purely algorithmic extensions, identification constraints can become substantive constraints and so are central to the basic methodology. It seems to us that substantive constraints are sometimes introduced purely for the algorithmic convenience of improving convergence or speed. We note that while poor convergence properties may be something that numerical analysts may be concerned with, statisticians may see them as evidence of poor data or unexpected linearities that should be reported and not artificially eliminated.

(v) Similar comments can be made about weighting. Just as we have identification constraints and substantive constraints, also we have explicit weighting and implicit weighting. By explicit weighting we mean that the researcher concerned has deliberately chosen weights, for each sample or variable (or both) and these will be taken into account in any subsequent analysis. Implicit weighting often occurs with count data where the same category may have many occurrences. Rather than writing $1 + 1 + \ldots 1 = n$ or even worse $1^2 + 1^2 + \ldots 1^2 + 1^2 = n$, this type of summation is often condensed to give algebraic expressions that look like as if they are concerned weights but are more properly to be regarded as a convenient way of writing an unweighted mean or sum-of-squares.

Table 1: Diagram showing possible two-way structures, especially adapted for categorical variables indicated by the coding matrices $\mathbf{G}_i$. The diagonal matrices $\mathbf{L}_i = \mathbf{G}_i^T \mathbf{G}_i$ give the frequencies of the categories of the *j*th variable. The light grey region indicates that row totals do not apply for variation within groups.

Objects	1 …	*j*th variable		… *p*	Combined	S.S
1		$\mathbf{G}_j\mathbf{Z}_j$	$\mathbf{G}_j\mathbf{z}_j$		$\mathbf{Gz}$	$\mathbf{z}^T\mathbf{G}^T\mathbf{Gz}$
…						
n						
Total		$\mathbf{1}^T\mathbf{L}_j\mathbf{Z}_j\ (= \mathbf{0}^T)$	$\mathbf{1}^T\mathbf{L}_j\mathbf{z}_j\ (= 0)$		$\mathbf{1}^T\mathbf{Lz}\ (= 0)$	
S.S			$\mathbf{z}^T\mathbf{L}_j\mathbf{z}_j = 1$		$\mathbf{z}^T\mathbf{Lz} = p$	

4. THREE-WAY EXTENSIONS

With three, and higher order arrays, the confusions encountered between data-matrices and two-way tables are even worse. Multiway-table extensions, for three or more way tables, using linear models have been routine tools for over a century and nowadays are often supplemented by generalised linear, or additive models. The distinction between blocks (local controls) and treatments (including treatments structure), not to mention the distinction between dependent and independent variables, are fundamental to many statistical methods. Multiplicative terms to represent interactions start with Fisher and Mackenzie (1923) but initially they were not much used because of the challenging cost of the time required for extensive eigenvalue calculations. Computers have made slight work of computing eigenvalues, so making bilinear modelling routine. Nowadays, triple product models are becoming more common (see Kroonenberg, 2008, for an excellent discussion). Multi-way structure for sets of data-matrices, or sets of distance matrices or sets of correlation matrices gives an especially useful class of three-way data. Then the visualizations of the two-way group average may be displayed in the usual way and the grouping factor displayed, separately or superimposed on the average. These methods include CVA, sets of data-matrices, generalised canonical correlation and notably INDSCAL, which handles sets of inner-products or distance matrices. The basic structure behind INDSCAL is the decomposition $y_{ijk} = \sum_{r=1}^{R} u_{ir} v_{jr} w_{kr}$ proposed as a generalisation of the SVD decomposition $y_{ij} = \sum_{r=1}^{R} u_{ir} v_{jr}$ which is basic for developing many methods of analysis for two-way arrays of data. Something similar (perhaps in tensor theory) might pertain to higher order arrays of data and hence the name of the algorithm CANDECOMP (see also DEDICOM) used for fitting data of type y_{ijk}. It proved simple to provide algorithms like CANDECOMP to fit three-way data but nothing similar to the least-squares properties of SVD seem to be available (see Chen and Saad, 2009, for a recent account of current progress). With linear models it was recognised that even when it is mathematically possible, it was inappropriate to fit three-way interactions without also including main effects and two-way interactions (see the marginality principle, Nelder, 1977). A concern is that algorithms developed for three-way arrays have little place for the marginality principle and do not make special provision for data arrays with symmetric matrix components. We have already seen (Bailey and Gower, 1990) that the approximation of a correlation matrix as if it were a square symmetric matrix has consequences and we would expect similar consequences when concerned with multiway arrays.

The Tucker-2 model $y_{ijk} = \sum_{p,q,r=1}^{P,Q,R} u_{ip} v_{jq} w_{kr} z_{pqr}$ is sometimes preferred to the INDSCAL model. In this model, the three-way array with terms z_{pqr} is termed the core matrix and its size PQR is chosen to be much smaller than the size of the data y_{ijk}. Note that the core matrix itself is a three-way array and so is a potential candidate for analysis by, say, INDSCAL. The Tucker-2 model was developed as a three-way generalisation of PCA and its properties, including its ability to distinguish among main effects and interactions, are perplexing. We deduce that while considerable progress has been made with developing algorithms for fitting three-way models, much work is needed to assimilate their potential.

Even with a good understanding of the properties, there remains the challenge of interpreting three (or more)-way interactions and how these combine with main and bilinear effects. Visualizations of three way arrays are poor, except when we are concerned with sets of two-way arrays, (Generalised canonical correlation, Sets of data-matrices, CVA, INDSCAL, …). Then some kind of group average gives a visualizable two-way summary and a third classification can give information on departures from the average. Albers and Gower (2014) and Williams and Gardner-Lubbe (2016) have given general methods for visualizations of rank-three three-way arrays, but even with this special case they are not visually compelling. Our feeling is that that is about as far one can go with giving visually useful depictions of interactions and, indeed, probably with understanding complicated interactions in more general terms.

The rise in computing power has carried with it a rise in the development of algorithms and this has been especially notable in data analysis. This trend has been touched upon in this section about three-way arrays but it pervades all of data analysis and probably extends far beyond to other fields of application. Algorithms for computing matrix inverses, eigenvalues and singular values are used freely in many methods of data analysis and have well-attested mathematical and numerical validity. However, they may be incorporated into software to give extended forms of analysis whose numerical properties are not so well attested. Sometimes data analysis is said to allow the data to speak for themselves, but it is not always clear what language or dialect they are speaking. Some new methodology is defined entirely in terms of computer code, which in some sense is said to give results that their authors find pleasing or informative but may be less compelling to others. Quite a lot of time has been spent not so much in understanding how an algorithm works but more on what information it is trying to convey. At one time, closed form solutions were sought but algorithms incorporated in much modern software cannot be regarded as a valid substitute for giving well attested solutions in closed form. How an algorithm works is fairly simple to discover but what it is intended to achieve is less clear. We are concerned that although the algorithmic approach is welcome, there is a need for a more rigorous validation of its objectives and how well they are met, than is commonly given.

5. CONCLUSIONS

Most of the above is about Analysis and a very restricted range of analytical methods at that, although we hope that our comments have a wider bearing. Rather than Analysis, perhaps the most important thing about Data Analysis is the Data. A well-designed experiment is a prerequisite to a good analysis. Parolini (2015) has drawn attention to how much effort Fisher put into the daily conduct and recording of experiments. Fisher's ideas soon spread from field experiments, to animal experiments, to horticulture and engineering and thence to clinical trials and pharmaceutical trials. The different fields of application threw up new challenges and opportunities for innovations and they all had their own problems with gathering data. Certainly, collecting data for observational studies in the social sciences, where direct experimentation is challenging, needs more attention. Yates

(1949) did something similar for surveys to Fisher's work on experimental design, using structured techniques such as stratified random sampling, multi-stage and multi-phase sampling. Yates pointed out that a well conducted survey was much less costly than an exhaustive census and gave comparable results. In contrast to the kind of data often used in Data Analysis, for Yates sample data was mostly quantitative; count data was only incidental and used mainly for estimating error bounds on means. We are not sure where big-data fits in here but we feel that big does not necessarily imply better; as much attention is needed to the collection of data as to keeping an eye on its analysis.

We have been asked us to forecast how Data Analysis will develop in the future. We are wary of forecasting the future of data analysis. M. G. Kendall gave two forecasts 25 years apart and both wide of the mark; Tukey wasn't much better and JCG's attempt to look ten years into the future was woeful. Since the seventeenth century statistics has been led by the problems of the day and our forecast is that this will continue. BIG Data and DNA are currently at the forefront of statistical developments. Data visualization is another matter. The technology of visualization has made giant strides and no doubt will continue to do so. The resources of the entertainment industries are largely untouched in statistics (but see Hans Rosling and his TED project). Apophenia, not to mention deliberate distortion (see Tufte), will remain a problem. Statisticians will continue to use mathematics as a tool and may even generate some of their own new mathematics, but keeping a balance between applicable mathematics and mathematical abstraction will need watching.

REFERENCES

Albers, C.J. and Gower, J.C. (2014). A contribution to the visualization of three-way arrays. In *Journal of Multivariate Analysis*. 132: 1-8.

Bailey, R.A. and Gower, J.C. (1990). Approximating a symmetric matrix. In *Psychometrika*. 55: 665-675.

Benzécri, J-P (1973). *L'Analyse des Données. Vol. 1: Taxinomie; Vol. 2: Analyse des Correspondances*. Dunod, Paris.

Blasius, J., Eilers, P.H.C. and Gower, J.C. (2009). Better biplots. In *Computational Statistics and Data Analysis*. 53: 3145-3158.

Chen, J. and Saad, Y. (2009). On the tensor SVD and the optimal low rank orthogonal approximation of tensors. In *SIAM Journal on Matrix Analysis and Applications*. 30: 1709-1734.

Fisher, R. (1953). Dispersion on a sphere. In *Proceedings of the Royal Statistical Society A*. 217: 295-305.

Fisher, R.A. and Mackenzie, W.A. (1923). Studies in crop variation. II. The manurial response of different potato varieties. In *The Journal of Agricultural Science*. 13: 311-320.

Gardner-Lubbe, S., Le Roux, N.J. and Gower, J.C. (2008). Measures of fit in principal component analysis and canonical variate analysis. In *Journal of Applied Statistics*. 35: 947-965.

Gower, J.C., Gardner-Lubbe, S. and Le Roux, N.J. (2016). Biplots: qualitative data. In *WIREs Computational Statistics*. 8: 82-111.

Gower, J.C., Lubbe, S. and Le Roux, N.J. (2011). *Understanding Biplots*. John Wiley, Chichester.

Gower, J.C., Groenen, P.J.F., Van de Velden, M. and Vines, S.K. (2014). Better perceptual maps: Introducing explanatory icons to facilitate interpretation. In *Food Quality and Preference*. 36: 61-69.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. In *Journal of Educational Psychology*. 24: 417–441.

Ioannidis, J.P.A. (2008). Why most discovered true associations are inflated. In *Epidemiology*. 19: 640-648.

Kroonenberg, P. (2008). *Applied Multiway Data Analysis*. John Wiley & Sons, Hoboken, NJ.

Mialet H. (2012). *Hawking Incorporated: Stephen Hawking and the Anthropology of the Knowing*. University of Chicago Press, Chicago.

Nelder, J.A. (1977). A reformulation of linear models. In *Journal of the Royal Statistical Society A*. 140: 48-77.

Nishisato, S. (1994). *Elements of Dual Scaling: An Introduction to Practical Data Analysis*. Lawrence Erlbaum Associates, Publishers, Hillsdale, NJ.

Parolini, G. (2015). The emergence of modern statistics in agricultural science: analysis of variance, experimental design and the reshaping of research at Rothamsted experimental station, 1919–1933. In *Journal of the History of Biology*. 48: 301–335.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. In *Philosophical Magazine*. 2(6): 559–572.

Pontryagin, L.S. (1966). *Topological Groups*. Gordon and Breach Sicience Publishers, New York.

Tukey, J.W. (1962). The future of data analysis. In *The Annals of Mathematical Statistics*. 33: 1-67.

Tufte, E.R. (1983). *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, Connecticut.

Williams, D. and Gardner-Lubbe, S. (2016). Visualising three-way arrays. In *Chemometrics and Intelligent Laboratory Systems*. 158: 180-186.

Yates, F. (1949). *Sampling Methods for Censuses and Surveys*. Griffin, London.