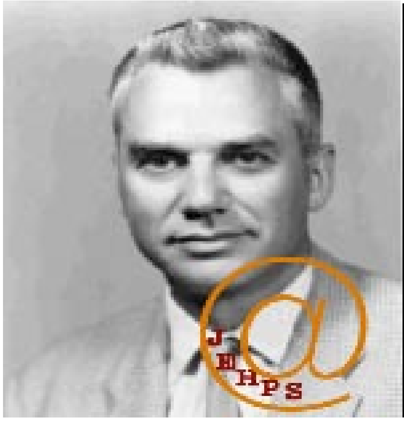


Journ@l Electronique d'Histoire des Probabilités et de la Statistique

Electronic Journ@l for History of Probability and Statistics

Vol 4, n°2; Décembre/December 2008

www.jehps.net

L'analyse des données ; before and around

Antoine de Falguerolles *

Résumé

Quelques étapes du développement exceptionnel de l'analyse des données en France dans les années 1965-1985 sont rappelées. S'agit-il d'une école spécifique et durable de la statistique descriptive multivariée ou seulement de la convergence de considérations méthodologiques sur des problèmes posés dans un environnement calculatoire complètement transformé par l'usage de l'ordinateur ? Des exemples montrent que, si les réponses sont nouvelles, les problèmes peuvent être anciens. Il semble aussi que les réponses de l'analyse des données tendent à se fondre dans l'arsenal commun des outils statistiques. Un effet de la Mondialisation en statistique ?

Abstract

Some milestones of the influential development of multivariate data analysis in France in the years 1965-1985 are recalled. Is *l'analyse des données* a distinctive and long-lived French school of multivariate descriptive statistics or simply the convergence of methodological concerns in response to questions arising in a drastically transformed computing environment? Examples show that if the answers are novel, the problems can be ancient. It seems also that the answers of *l'analyse des données* tend to merge into the common arsenal of statistical tools. An effect of Globalization on statistics?

Remerciements : L'auteur voudrait remercier le groupe informel mais actif des chevaliers des albums de statistique graphique pour son soutien amical, Marc Barbut pour lui avoir signalé le travail de René Suaudeau, les organisateurs de la cinquième conférence CARME tenue à Rotterdam (Pays-Bas) les 25-27 juin 2007 pour leur aimable invitation à présenter une partie de ce travail. L'auteur s'excuse de son anglais que les français liront certainement beaucoup plus facilement que tous autres.

*Institut de Mathématiques, Laboratoire de statistique et probabilités, Université de Toulouse III, 118 route de Narbonne, F-31062 Toulouse cedex 9, France.

1 Introduction

What is *l'analyse des données*? Several French speaking authors have drawn the contours of what can be considered a French connotated extension of descriptive statistics: Edwin Diday and Ludovic Lebart (1977, [28]), Jean-Marie Bouroche and Gilbert Saporta (1978, [14]), Michel Volle (1978, [85]), Gilles Celeux and Edwin Diday (1980, [20]), Gilbert Saporta (1984, [78]) ... Is the existence of a French school of data analysis known as *l'analyse des données* a scientific reality? To what extent the existence of this School is related to its publication language? to the emergence of new tools?

In this paper, the following working definition is used: *l'analyse des données* is a set of methods for the descriptive analysis of multivariate data observed on large sets of units. These methods are commanded by the following ingredients which in practice are blended in varying proportions:

1. A firm belief in multidimensional descriptions.
2. A search for latent variables giving sense to the observed data and allowing dimension reduction.
3. A conviction that proper graphical representations best convey the structure of either the original data or the result of their analyses.
4. A manifest (and possibly overplayed) claim from their authors to avoid any modeling driven by probabilistic considerations.

There is some form of agreement on dividing the methods of *l'analyse des données* in two groups: "clustering methods" (*classification automatique*) and "factorial methods" (*méthodes factorielles* [1]).

This paper is written by neither a pioneer nor a strict follower of *l'analyse des données*. It addresses the "before" and the "around" of *l'analyse des données*. The "before" is an excuse to present early material sharing some of the characteristics of *l'analyse des données*, as just defined. The "around" is an opportunity to describe one of the evolutionary branches of multivariate statistics. The article is organised as follows. Section 2 consists in an overview of the development of *l'analyse des données* in the period 1965-1985. Section 3 gives some examples of multivariate quantification published even before the word "statistics" existed. Section 4 describes some early attempts of data visualization where dimensionality is a real concern. Section 5 briefly consider a French "school of statistics" recognized in its time but foregone after a quarter of a century of existence. A bad omen for the future of *l'analyse des données*! Finally, Section 6 depicts the motivations behind an evolutionary branch of *l'analyse des données*.

2 An overview of the history of l'analyse des données

In its French flavour, *l'analyse des données* emerged in the late sixties with seminal lectures by Jean-Paul Benzécri. However Benzécri did not arise by spontaneous generation. Except for a few radical French researchers, *l'analyse des données* has always been carefully replaced in the main corpus of statistics and due respect has always been

1. A word to word translation of the French *méthodes factorielles* into the English *factorial methods* to refer to methods such as principal component analysis, correspondence analysis, metric multidimensional scaling ... is somewhat ambiguous. The terminology "eigen decomposition based methods" would be clearer.

paid to the founding fathers [2]. The same period also saw the rise of main frame computers and, some years later, of personal computers. The coupling of computer access with *l'analyse des données* is frequently emphasized by French authors [3]. The novelty was that statistical methods which required "heavy-for-their-time" computations could be effectively investigated. Moreover, the number of users having data (on decks of punched cards or magnetic tapes) and not knowing precisely what to do with them (except to reject the null hypotheses in simple tests) did increase. This dialog between users and statisticians under the protecting wing of computers is often advanced to explain the success of the methods of *l'analyse des données*. One celebrated example is the reiterated use of correspondence analysis by Pierre Bourdieu [4].

2.1 Development in the period 1965-1985

L'analyse des données developed substantially in France in the years 1965-1985. This explosion can be measured by the number of publications and even more by the diversity of fields in which, nowadays, its methods are fully used: economy, sociology, geography, data mining, text mining, bibliometry, environmetrics ... *L'analyse des données* even became fashionable in weekly magazines!

2.1.1 Books and handouts

Excellent comprehensive books, with frequent reeditions, have been written by Ludovic Lebart and Jean-Pierre Fénelon (1971, [53]), Jean-Paul Benzécri (1973, [5] and [6]), Patrice Bertier and Jean-Marie Bouroche (1975, [8]), Ludovic Lebart, Alain Morineau and Nicole Tabard (1977, [54]), Michel Volle (1978, [86]), Edwin Diday *et al.* (1980, [27]), Michel Masson (1980, [65]), Edwin Diday, Jacques Lemaire, Jean Pouget, Françoise Testu (1982, [29]), ...

Many handouts circulated: the neatly calligraphied notes by Israël-César Lerman, the series of booklets issued in 1971 by the *Centre d'Études Économiques d'Entreprises* (C3E), the *Cours d'analyse des données 1978-1979* by Yves Escoufier (1978, [35]) ...

Israël-César Lerman published his book on clustering in 1970 [57]. The C3E booklets, edited by Francis Cailliez and Jean-Pierre Pages, were combined into a book entitled *Introduction à l'analyse des données* [5] (1976, [16]), covering

2. In his preface dated 1971 to *l'introduction à l'analyse des données*, Georges Morlat quotes G. Darmois, B. de Finetti, R. A. Fisher, J. Neyman, J.W. Tukey, A. Wald ([16, pages I-V]). Not surprisingly, John W.Tukey's leading paper on "the future of data analysis" in the *annals of statistics* (1962, [81]) is explicitedly referred to.

3. This coupling is reflected in the title of an early book on *l'analyse des données*, namely *statistique et informatique appliquée* by Ludovic Lebart and Jean-Pierre Fénelon (1971, [53]). See also Georges Morlat in the Preface dated 1971 of *l'introduction à l'analyse des données* (Cailliez and Pages [16], 1976), Jean-Paul Benzécri in the first issue of *les Cahiers de l'analyse des données* (1976, [7, Part I, page 9, first line]), Jacques Dauxois and Alain Pousse (1976, [22, Introduction, page 1]) ...

4. Pierre Bourdieux (1930-2002) is perhaps the most noted French sociologist of recent times. An example of use of correspondence analysis can be found in his famous analysis of academic sociology *Homo academicus* (1984,[13]). Further details on scientific and personal links between Pierre Bourdieu and the French statisticians Jean-Paul Benzécri and Henry Rouanet (1932-2008) can be easily obtained from a search on the web.

5. The names of co-authors listed on the Front page of *Introduction à l'analyse des données* are F. Cailliez, J.P. Pages (editors), G. Morlat (direction), J.-C. Amiard, J. Andres, M.-F. Bara, J.-M. Braun, J. Brenot, P. Cazes, J. Dehedin, B. Diop, Y. Escoufier, C. Guegen, N. Lacourly, J.-P. Mailles, B. Marchadier, M. Pietri, E. Roy, G. Saporta, F. Testu, R.Thomas. In the Foreword to this book, Cailliez and Pagès acknowledge the leading role of T.W. Anderson and J.-P. Benzécri and mention also a few names of researchers and research teams in France: Y. Escoufier in Montpellier, J. Dauxois and A. Pousse in Toulouse, G. Saporta and

both clustering and factorial methods [6].

2.1.2 Statistical Software packages

The development of statistical software packages dedicated to *l'analyse des données* was stimulated by the accessibility of personal computers after 1980. Examples of softwares are SPAD, CHADOC ...

2.1.3 French statistical journals

It would be a difficult task to list all journals which published either theoretical articles or applications of *l'analyse des données*. Numerous examples of the flexibility of *l'analyse des données* can be found in the series of *Cahiers de l'analyse des données* published between 1976 and 1997 under the editorship of Jean-Paul Benzécri. The content of *Revue de statistique appliquée* (published since 1953 and merged in 2004 with another statistical journal) is also a fair indicator of the attention given to *l'analyse des données* in France. Under the supervision of Pierre Cazes, this journal addressed, along with more traditional topics, an astonishing variety of theoretical and practical aspects of factorial and clustering methods. In addition, *Statistique et analyse des données* [7], the journal of the *Association des statisticiens universitaires* [8] (ASU) published several theoretical papers on *l'analyse des données*.

More remarkably, *Économie et Statistique*, a monthly journal edited and published by the *Institut National de la Statistique et des Études Économique* (INSEE) also published numerous articles where the methods of *l'analyse des données* were used.

2.1.4 French statistical meetings

In the seventies and the eighties, the latest progresses in *l'analyse des données* were thoroughly discussed during the *Journées de Statistique*, the annual meeting organised by the ASU since 1969. During the *Journées de Statistique* special sessions were dedicated to cluster analysis in close contact with the *Société francophone de classification* [9] (SFC). In 1992, the SFC decided to hold its own annual meeting called *les rencontres de la SFC* where *l'analyse des données* is obviously at home. The official (and dominant) language was and still is the French language but both meetings are

M. Masson in Paris, E. Diday at the IRIA [Institut de Recherche en Informatique et Automatique created in 1967, nowadays the INRIA with a national N], J.M. Bouroche and collaborators ([16, pages VII-VIII]).

6. Similarly in the Netherlands, the mimeographed text of Albert Gifi, a pen name for an active group in the Netherlands, that was circulated in English since 1981 became a book in 1990 [41].

7. The first volume of *Statistique et analyse des données* appeared in 1976; the journal ceased publication in 1991. Its editors were Henri Caussinus and Guy Romier, Yves Escoufier, and Bernard Van Cutsem. While serving on the Board of the *Association des statisticiens universitaires*, I remember with mixed feelings the discutions at the meeting when the title of its journal was chosen: *Statistique et Analyse des Données*. Clarification or pleonasm? The alleged opposition between *Statistique* and *Analyse des données* is thoroughly discussed by Henri Caussinus in the section *Statistique, modèles et analyse des données* of a methodological paper (1993,[18]).

8. The *Association des Statisticiens Universitaires* was founded in 1969 mostly by academics, mainly from "provincial" universities, for a variety of reasons: the enthusiasm in the 1968 era for the creation of new structures; the Malthusianism of existing statistical societies, e.g. *Société de Statistique de Paris*; the rejection of Societies suspected of "imperialistic" views on statistics, e.g. *Société mathématique de France* ... In particular, the ASU was a forum where activities linked to the mass emergence of statistical curriculum in newly created universities were discussed. Its name was changed in 1987 into *Association pour la Statistique et ses Utilisations* to reflect the growing number of members who were non academic statisticians.

9. The *Société francophone de classification* was founded in 1977. See Lerman (1998, [58]) for an historical note.

regularly attended by a few non-francophones, either sympathizers or invited speakers, delivering their talks in English. An international forum for *l'analyse des données* was also the series *Journées d'Analyse des Données et Informatique* organized by the *Institut National de Recherche en Informatique et Automatique* (INRIA) in 1977, 1979, 1983 . . . .

2.1.5 International projects

It would be informative to establish an exhaustive list of international projects in which French teams of researchers in *l'analyse des données* were involved. This would require access to the archives of research institutions and has not be done. A special issue of the *Journal de la Statistique Appliquée* (1987, Vol. 35(3)) addresses some of the issues discussed below in subsection 2.3. It reflects the methodological results obtained in a joint program of the Economic and Social Research Council (ESRC) and the *Centre National de la Recherche Scientifique* (CNRS). Surely, there must have been a number of similar projects. Their joint analysis would be of real interest for the history of multivariate data analysis.

2.1.6 The come back of probability calculus

As anticipated by Georges Morlat in his Preface to the *introduction à l'analyse des données* [16] the probabilistic framework was to come back [10] . Broadly speaking, its reintroduction followed two distinct paths, a "within" and a "between". The former consists in extending the framework of data analysis to answer either mathematical questions such as the mathematical generality of methods or traditional statistical questions or . . . It includes detailed treatment of the convergence of eigen elements or of tests of equality for covariance operators. Seminal works for the "within" are by Michel Masson, Jean-Claude Naouri, Jacques Dauxois and Alain Pousse, Jean-Claude Deville . . . A typical example is Dauxois, Pousse and Romain (1982, [23]).

The "between" consists in building bridges between methods of *l'analyse des données* and statistical modeling methods which were developed in the same period. Anglo-Saxon publications which had a real impact in France on statistical data analysis are by Yvonne Bishop, Stephen Fienberg and Paul Holland (1975, [11]), John Nelder and Robert Wedderburn (1972, [68]), Peter McCullagh and John Nelder (1983, 1989, [66]), Murray Aitkin, Dorothy Anderson, Brian Francis and John Hinde (1988, [2]). At the junction of the "between" and "within" are useful considerations on the role of models in *l'analyse des données* (Henri Caussinus, 1986 [17], 1993 [18]) with applications to robustness and to metric or dimension choice.

2.2 Nowadays in France?

An observer of the present-day landscape of research in statistical methodology in France, may be disconcerted to notice that the portion of *l'analyse des données* has been substantially reduced from it was in its glorious period. Methodological progresses in *l'analyse des données* are no more considered as "cutting edges" statistics by the academic world. Is *l'analyse des données* nothing else than an useful extension of descriptive statistics widely researched in the period 1965-1985?

10. To assess the stability of results from *l'analyse des données*, validation procedures such as the jackknife, the bootstrap or permutation tests were naturally considered. But these probabilistic methods are peripheral to the methods of *l'analyse des données*.

A provocative observer can be tempted to plagiarize Paul F. Lazarsfeld [11]'s conclusion on Le Play and his *méthode d'observation*. Not only the text nicely transposes to *l'analyse des données* (and Jean-Paul Benzécri) but it raises a series of impertinent questions:

> "We are thus faced with the following situation. Here is a school created by a charismatic personality who makes an important innovation in social methodology and intertwines it with very strong and activitistic ideological beliefs. One group of his disciples share his beliefs and want at the same time to make methodological progress. Under normal circumstances, a scientific innovator is respected by his students, and it is taken for granted that his successors make continuous advances beyond the teacher. In a charismatic context, this leads to a tension between the scientific and the sectarian element in the tradition. This would be a matter of only secondary interest if it were not for the fact that rather suddenly, the LePlayistes disappeared from the French sociological scene. At least as far as one can see from a distance, the school which was so extensive and articulate up to the First World War has been completely replaced by the Durkheim tradition to which they paid only casual and rare attention in the "Science Sociale". In various reviews of French sociology which Frenchmen have recently written the LePlayistes are not even mentioned. Do we face here a political phenomenon? Did the few relevant university posts all go to the Durkheim group at a time when the French government had an anti-clerical tendency? Did the descriptive fervor of the LePlayistes exhaust its potentialities, and make it less attractive than the conceptualizing of the Durkheim School to a younger generation? Do we face here the difficulty a charismatic movement has: in spite of their ambivalence to the founder, did the LePlayistes form too much of a sect to be acceptable to the regular academic bureaucracy?
>
> I cannot tell. Certainly the methodological ideas of the school were interesting and susceptible of further development as I shall discuss presently. But first, I must trace briefly the effect which LePlay had abroad, especially in England and in the United States and here, strangely enough, the theme of the purge can be continued. While we find outspoken and clamorous admirers of his in the two Anglo-Saxon countries, they changed his ideas even more than did his French disciples, each in his own way and perhaps without knowing it. . . . " Lazarsfeld [52, pages 320-321].

As seen in the Introduction, the aims of this paper are more modest. In particular, French research policy will not be discussed and "outspoken and clamorous admirers" in France and outside France will not be ranked!

2.3 Globalization versus French exception

What is the specificity of *l'analyse des données* as compared to descriptive multivariate data analysis?

11. Paul F. Lazarsfeld (1901-1976) is one of the leading figures in 20th century American sociology. Lazarsfeld was also influential in France. See subsection 6.1 below.

2.3.1 Globalization

In the 2008 augmented Spanish translation [46] of his English book, *correspondence analysis in practice* [45], Michael Greenacre gives 14 references in the bibliography (Appendix C). These divide into 12 English publications and 2 French publications, namely Jean-Paul Benzécri's two 1973 volumes ([5] and [6]) and the book published in honor of Brigitte Escoffier-Cordier after her death in 1994 [34]. However, all have a Benzécri connection. Expectedly, with time, the link with Benzécri will be through Ludovic Lebart, Michael Greenacre, Fionn Murthagh ... But the starting point is undoubtedly the seminal work of Benzécri. In contrast, the review by Pierre Cazes (2008, [19])) of the Proceedings (2006, [47]) of the fourth CARME [12] meeting held in Barcelona in 2003 clearly shows that correspondence analysis is not any more a French exception. Besides, globalization is at work for the whole spectrum of *l'analyse des données* as can be seen by numerous publications by non-French authors in international journals.

2.3.2 The French exception

Then, was there an identifiable French touch in *l'analyse des données*? In the discussion of a paper entitled "data analysis in Official socio-economic Statistics" [26] which Jean-Claude Deville and Edmond Malinvaud presented in 1983 at the Royal Statistical Society, John Gower made the following comment:

> Tonight's speakers are right in suggesting that British and French statisticians should know more of each others' work. However, I do not think that things are quite as bad as they make out, for although many of us are not as fluent as we would like to be in speaking French, many can and do read it. My feeling is that the style of French mathematical presentation is often more of a barrier than is the language, but this is not a problem with tonight's paper.

John Gower's comment addresses two questions, a language barrier and a difference in mathematical presentation which can be differently evaluated according to the nationality of the discussant. Moreover the situation may have changed since 1983. Both questions are reconsidered below.

The language barrier

The language barrier is sometimes advanced as an explanation of the gaps between existing schools of data-analytic thought. The French prevention to the use of English and the decline of French as an international scientific exchange language during the second half of the 20th century are well known [13]. Clearly, the French had difficulties to write articles in English or to present talks in an English which could be reasonably understood (by others than their French colleagues). Moreover a few couldn't read English at all. This barrier has certainly contributed to a form of isolated development of multivariate data analysis. The barrier also showed some porosity: Kanti V. Mardia, John T. Bibby and John M. Kent's *multivariate analysis* (1979, [63]) crossed the Channel and found its place in French university libraries while Ludovic Lebart,

12. Correspondence Analysis and Related Methods.

13. As an example of late resistance, the French government issued in 1982 regulations stating that no public funds should go to scientific activities held in France where the French language was not the official language! And the regulations were well accepted in academic circles.

Alain Morineau and Kenneth M. Warwick's *multivariate descriptive statistical analysis* (1984, [55]) ventured in the wilderness of non French speaking countries.

Note in passing that nowadays the use of English is not any more felt either as an impossible task or as a betrayal of the French culture. However the increasing number of French researchers who have improved their English does not balance the rapidly decreasing number of non Francophone researchers who "can and do read" French!

The mathematical presentation

An article published by Jean-Pierre Pagès, Francis Cailliez and Yves Escoufier (1979, [70]) in the *Revue de Statistique Appliquée* can illustrate this aspect. The paper addresses factorial analyses and is structured as follows. Firstly, the methods are placed in line with Anglo-Saxon forerunning research by Charles Spearman, Karl Pearson and Harold Hotelling [14]. Secondly, a classification of the types of data and the relevant variants of factorial methods are recalled. Thirdly, the central mathematical concept underlying these methods, namely the "duality scheme", is described in an involved statistical environment. When read from today's perspective, the French flavour in the paper resides in the use of the "duality scheme" which makes the paper look rather mathematical. The "duality scheme" certainly brings a clear understanding of the interplay of the metrics in the various spaces involved. But nowadays, generalized singular value decomposition is considered as a tool much easier to understand and to use than the "duality scheme".

Articles addressing clustering analysis show a comparable difference which may be due in France to the formalism introduced in the teaching of set and graph theory. Again, the understanding of the mathematical structures put at work in statistics is emphasized.

To a large extent the mathematical background in statistical papers do reflect the structure of statistics teaching and the process of career choice in the previous years. Certainly, Post-1968 students felt more at ease with *l'analyse des données* than with "traditional" statistics to which they had a limited exposure. Note that the specification of a proper articulation between mathematics and statistics is certainly an issue for the future of statistical research.

3 Multivariate quantification

Multivariate quantification existed long before the word "statistics". As expected, the early examples which are presented below call for modern analyses provided by *l'analyse des données*.

3.1 *La Felicissima Armada*, 1588

During the preparation of the Armada [15], several accounts of its formidable strength were circulated to reassure allied powers of Spain or to intimidate its enemies. How the

14. The remembrance of this filiation is the result of a constant concern for at least two reasons: appropriate scientific behaviour and genuine interest. An example of the former is the note by Gilbert Saporta concerning the fathers of correspondence analysis (1975, [77]). An example of the latter is the French translation, with a foreword by Georges Darmois, of the second edition of Charles Spearman's book entitled *The abilities of man* by François Brachet (1936, [79]).

15. The Spanish Armada was the fleet intended to invade England in 1588. *La Felicissima Armada* was under the command of the Duke of Medina Sidonia; second in command was the famous sailor Juan Martinez de Recalde. It was all destroyed by a week's fighting.

Table 1: General summary of the Armada. Rows (10) correspond to fleets forming the Armada: *Armada de Galeones de Portugal* (A1), *Armada de Viscaya* (A2), *Galeones de la Armada de Castilla* (A3) ... *Galeras* (A10). Columns (10 items) correspond to descriptors: *Numero de navios* (V1); *Toneladas* (V2); *Gente de guerra* (V3); *Gente de mar* (V4), *Numero de todos* (V5=V3+V4), *Pieças di artilleria* (V6), *Peloteria* (V7), *Poluora* (V8), *Plomo qui tales* (V9), *Cuerda qui tales* (V10). Total columns sometimes differ from the published total. See the last two lines in the Table: the first is taken from the book, the second is computed from the table; the discrepancies reveal typographic errors more than gross errors.

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
A1	12	7737	3330	1293	4623	347	18450	789	186	150
A2	14	6567	1937	863	2800	238	11900	477	140	87
A3	16	8714	2458	1719	4171	384	23040	710	290	309
A4	11	8762	2325	780	3105	240	10200	415	63	119
A5	14	6991	1992	616	2608	247	12150	518	139	109
A6	10	7705	2780	767	3523	280	14000	584	177	141
A7	23	10271	3121	608	3729	384	19200	258	142	215
A8	22	1221	479	574	1093	91	4550	66	20	13
A9	4	0	873	468	1341	200	10000	498	61	88
A10	4	0	0	362	362	20	1200	60	20	20
	130	57868	19295	8050	27365	2431	123790	4575	1232	1151
	130	57968	19295	8050	27355	2431	124690	4375	1238	1251

data were structured? A book by Paz Salas et Álvarez published in 1588 gives a quantified description in the form of a data matrix where the rows correspond to the fleets forming the Armada and the columns to quantitative descriptors of its strength [25]. The data matrix, whose caption reads *summario general de toda el armada*, is reproduced in Table 1. Is the fact that there are exactly as many descriptors than statistical units just coincidental? Or was it believed that as many descriptors as objects were needed?

L'analyse des données offers several ways to analyse such a data matrix. The eigen decomposition of the correlation matrix associated to Table 1, descriptor 5 being omitted, shows a clear structure. The cumulative percentage of correlation are : 75, 88, 94, 97, 99, 100, ... A two dimensional representation is therefore suggested. As expected for such data, it turns out that the first dimension is a size factor for the fleets while the second is a fleet configuration factor.

3.2 *La balance des peintres*, 1708

In *la balance des peintres*, an appendix to a book entitled *cours de peinture par principes* (1708), Roger de Piles [16] scales fifty six painters according to four conceptual dimensions: composition, drawing, colour, expression [72] [17]. In the discussion, Piles

16. Roger de Piles (1635-1709) was a painter and an art critic. He was also the secretary of the French ambassador Michel Amelot de Gournay (1655-1724). This led Piles to visit Italy, Germany, Austria, Switzerland and Holland. On a diplomatic mission to Holland, in 1692, he was arrested and spent five years in prison. He then stayed in Paris. Piles is especially noted for his writings on art. Piles may also have used the pseudonym François Tortebat for some of his works.

17. Coincidence? I bought Piles's *cours de peinture par principes* some years ago from a book dealer established in Carcassonne. It turned out that the book dealer had been a student of Professor Jean-Paul

Table 2: *La balance des peintres*. This data set gives the four scores given by Piles on four dimensions. Note the two painters with a missing score. The painters'names are taken from the book of Piles [72]. In Davenport and Studdert-Kennedy's article [24], "Albane" reads "Albani", "Albert Dure" reads "Durer", "Le Guide" reads " Guido Reni", "Polid. de Carvagio" reads "Polidore da Caravagio ", "Tadée Zuccre" reads "T. Zuccaro", and "Frederic Zuccre" reads "F. Zuccaro".

NOMS des Peintres les plus connus	*Composition.*	*Dessein.*	*Coloris.*	*Expression.*
A				
Albane.	14	14	10	6
Albert Dure.	8	10	10	8
Andre del Sarte.	12	16	9	8
B				
Baroche.	14	15	6	10
⋮	⋮	⋮	⋮	⋮
Le Guide.	–	13	9	12
⋮	⋮	⋮	⋮	⋮
Polid. de Caravage	10	17	–	15
⋮	⋮	⋮	⋮	⋮
Z				
Tadée Zuccre.	13	14	10	9
Frederic Zuccre.	10	13	8	8

Table 3: Variable weights in the principal components based on the covariance matrix.

	Composition	Drawing	Colour	Expression	Percentage of inertia
component 1	-0.48	-0.42	0.38	-0.66	56
component 2	-0.38	0.19	-0.85	-0.33	29
component 3	0.78	-0.28	-0.21	-0.51	9
component 4	-0.10	-0.84	-0.31	0.43	6

Table 4: Variable weights in the principal components based on the correlation matrix.

	Composition	Drawing	Colour	Expression	Percentage of inertia
component 1	-0.50	-0.56	0.35	-0.56	57
component 2	-0.49	0.27	-0.77	-0.32	26
component 3	0.60	-0.59	-0.49	-0.25	10
component 4	-0.40	-0.52	-0.23	0.72	7

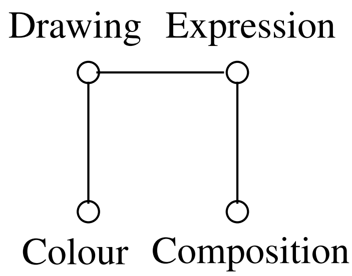


Figure 1: The independence graph of the Gaussian graphical model fitted to the painters' data (deviance 1.68 for 3 degrees of freedom).

specifies that composition and drawing are themselves bi-dimensional: invention and disposition for the former, taste and correctness for the latter. The four evaluations are each graded on 20, where 20 is the "sovereign perfection" and 19 the "not yet achieved". A partial fac-simile of the display in Piles' *balance des peintres* is given in table 2. With two missing data [18], only 54 painters are fully graded. The data set has been popularized by Davenport and Studdert-Kennedy (1972, [24]) [19].

Piles does not tell how to combine his four partial assessments into an overall ranking. But still, was such a multivariate approach for assessment well known in 1708? The first known rating of psychological variables is usually ascribed to Christian Thomasius (1655-1728) in 1692, see McReynolds and Ludwig (1984, [67]). Thomasius rated individuals on four "inclinations": sensuousness, acquisitiveness, social ambition and rational love. Thomasius proposed that a value of 60 be assigned to the dominant inclination, and a value of at least 5, and more if appropriate, to the weakest of the four. The values of the remaining inclinations were then to be rated in terms of their relative magnitude with respect to the first two. Intervals were allowed, an omen to the developers of Symbolic Data Analysis (see Billard and Diday, [10]). Did Piles come across

Benzécri!

18. A companion book by Piles is in my father-in-law's collection; it includes a hand written version of the data set on a sheet of paper which is glued on the last page of the book [73]. On this list, the two missing values are set to zero. However, these two values differ widely from estimated with conditional expectations derived from the graphical model described in Figure 1.

19. In their praiseworthy article, the authors have added the school to which belonged the painters considered by Piles. Their paper is a wonderful primer in *l'analyse des données* where, given the additional information on the school, unsupervised methods and supervised methods are simultaneously used: PCA, clustering, MANOVA … By discreticizing the scores given by Piles, multiple correspondence analysis can be compared to principal components analysis …

(the work of) Thomasius? The apparent connection is not documented.

The Piles data call for an analysis of either the variance matrix or the correlation matrix by factorial methods. As can bee seen in Tables 3 and 4, each eigen vector opposes in turn one of Piles criterion to the other three. The ordering is the same in both eigen decompositions: colour against composition, drawing and expression, drawing against composition, colour and expression …An alternative strategy is to look for conditional independence between colour, composition, drawing, and expression and to fit a Gaussian graphical model (Whittaker [88], Edwards [33]). The independence graph of the fitted model is given in Figure 1. Given the goodness of fit of the graphical model, the fitted covariance does not fail to replicate the pattern of oppositions found in the empirical covariance and correlation matrices.

3.3 Basville, the intendant of Languedoc

Nicolas de Lamoignon de Basville (1648-1724) was *Intendant* of the King of France, Louis the XIVth, for the Province of Languedoc. His administration took place during troubled times of the persecutions of the Protestant population whose religion had been banned in 1685. Basville is the author of a *Mémoires sur la province de Languedoc* written in the years 1696-1697 for the instruction of the Duke of Burgundy, a heir to the French Crown (1734, [4]). Basville had received a good mathematical education by the Jesuits which may explain his skilfulness in using numbers. His report contains 5 data sets displayed according to modern formats.

Two data sets are reproduced here, see Tables 5 and 6. They provide examples of quantification applied to the Catholic coverage of Languedoc by a diversity of female (Table 5) and male religious orders (Table 6). The quantification is itself multivariate: number of persons, number of establishments. Notwithstanding their identical format, the French captions slightly differ: in Table 5 the first word is *État* which is evocative of "Statistics", while in Table 6 the first word is *Carte* (Map) which emphasizes the geographical dimension of the row factor.

Basville's official report provides here examples of modern contingency tables [20]. No doubt that readers familiar with *l'analyse des données* will submit the two tables to correspondence analysis. However, the "null" situation of independence may not be appropriate? From the point of view of Basville, the marginal counts for the clergypersons should be proportional to the population in the diocese (possibly with reinforcement in the rebellious dioceses). On the other margin, no religious order should be dominant (dividing and reigning).

4 Visualization in reduced dimensions

The data sets considered in the previous section perfectly depict the statistical nature of the data: units $\times$ variables for the first two (Table 1 and Table 2) or cross-tabulations of counts for the last two (Table 5 or Table 6). In a sense, they are minimal attempts to visualize the data. The examples below present early attempts to go beyond the data, in the direction of visualization with special attention to their dimensionality. As expected, the displays are constrained by the technical limitations of their time. But they contain the germs of the very topics addressed in *l'analyse des données*.

20. A zero or one two-way contingency table where the presence is denoted by an asterisk can be found in *Statistique élémentaire, essai sur l'état géographique, physique et politique de la Suisse*, a 4 volumes description of Switzerland by François-Jacques Durand (1795, [31]) .

Table 5: Facsimile of the table inserted after page 64 of the 1734 edition of Basville's *Mémoires* .... The table is the cross tabulation of orders of nuns (columns) and dioceses (rows). The first count is number of convents (C.), the second is number of nuns (R.). The labels for the columns categories are *Bernardines, La Visitation, Malthoise, Ursulines, Réfuge, Tiers-Ordre, Saint Sulpice, Carmelites, Lepoiaffes, St. Dominique, Feüillantines, Sainte Claire, Notre-Dame, St. Sans Clou, Hospitalieres, Saint Servin, Descalses, La Magdelaine, Salingives, Filles de la Croix, Sainte Marthe, Benedictines, L'Anonciade* which are denoted here by F01, F02, ..., F23.

	F01		F02		F03		F04		F05		F06		F07		F08		F09		F10		F11		F12	
Diocèse	C.	R.	C.	R.	C.	R.	C.	R.	C.	R.	C.	R.	C.	R.	C.	R.	C.	R.	C.	R.	C.	R.	C.	R.
Narbonne	1	12	0	0	0	0	1	16	0	0	0	0	0	0	1	21	0	0	0	0	0	0	1	21
Toulouse	0	0	1	52	1	26	1	29	1	26	1	39	1	22	1	26	1	25	1	35	1	26	3	130
Alby	0	0	1	34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	28
Montpellier	1	5	1	35	0	0	1	30	1	60	0	0	0	0	0	0	0	0	1	20	0	0	2	45
Beziers	0	0	0	0	0	0	1	36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	28
Nîmes	0	0	1	16	0	0	3	70	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Viviers	0	0	3	71	0	0	1	36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	43
Carcassonne	1	22	0	0	0	0	1	31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Le Puy	3	45	1	25	0	0	2	25	1	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Montauban	0	0	0	0	0	0	1	18	0	0	0	0	0	0	0	0	0	0	1	25	0	0	1	21
St Papoul	0	0	0	0	0	0	1	11	0	0	0	0	0	0	0	0	0	0	1	87	0	0	0	0
Castres	1	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	33
Lavaur	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	26
Mirepoix	0	0	0	0	0	0	0	0	1	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Uzez	1	10	3	93	0	0	2	65	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
St Pons	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Lodève	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Mendes	1	5	1	18	0	0	3	46	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Agde	0	0	0	0	0	0	1	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Alais	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Alet	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Comenge	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Rieux	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	9	112	12	344	1	26	19	428	4	107	1	39	1	22	2	47	1	25	4	167	1	26	14	375

	F13		F14		F15		F16		F17		F18		F19		F20		F21		F22		F23	
Diocèse	C.	R.	C.	R.	C.	R.	C.	R.	C.	R.	C.	R.	C.	R.	C.	R.	C.	R.	C.	R.	C.	R.
Narbonne	1	21	0	0	1	14	0	0	0	0	1	39	1	20	1	8	1	26	0	0	0	0
Toulouse	1	80	1	38	1	50	1	39	1	20	0	0	0	0	0	0	0	0	0	0	0	0
Alby	0	0	0	0	2	17	0	0	0	0	0	0	0	0	0	0	0	0	1	14	2	55
Montpellier	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Beziers	2	56	0	0	1	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Nîmes	0	0	0	0	1	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Viviers	1	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	22	0	0
Carcassonne	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Le Puy	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Montauban	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
St Papoul	0	0	0	0	0	0	0	0	1	28	0	0	0	0	0	0	0	0	0	0	0	0
Castres	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	46	0	0
Lavaur	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	8	0	0	0	0	0	0
Mirepoix	1	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Uzez	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
St Pons	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Lodève	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	15	0	0
Mendes	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Agde	1	17	0	0	1	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Alais	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Alet	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Comenge	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Rieux	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	53
TOTAL	7	211	1	38	7	134	1	39	2	48	1	39	1	20	2	16	1	26	6	97	5	108

Table 6: Facsimile of the table inserted after page 64 of the 1734 edition of Basville's *Mémoires* .... The table is the cross tabulation of orders of clergymen (columns) and dioceses (rows). The first count is number of monasteries (C.), the second is number of clergymen (R.). The labels for the columns categories are *Ecclesïastiques, Jesuîtes, Peres de l'Oratoire, Cisteaux, Benedictins, Chartreux, Minimes, Feüillans, Jacobins, Cordeliers, Trinitaires, Augustins, Carmes, La Mercy, Capucins, Recolets, Saint Antoine, Prêtres Irlandois, Tierçaires ou Picpus, Dominicains réforméz, Relig. Saint Orans, Seminaires de Carmes, Prémontrez, Celestins, Seminaires.* They are abbreviated here as H00, H01, ..., H24. In the elementary columns, E., M., C., and R. stand for *ecclésiastiques, maisons, couvents and religieux*.

	H00	H01		H02		H03		H04		H05		H06		H07		H08		H09		H10		H11	
Diocèse	E.	M.	R.	C.	R.	C.	R.	C.	R.	C.	R.	C.	R.	C.	R.	C.	R.	C.	R.	C.	R.	C.	R.
Narbonne	320	0	0	0	0	0	0	2	15	0	0	1	15	0	0	2	19	2	23	2	15	1	15
Toulouse	560	4	126	1	18	1	1	1	36	1	34	1	26	1	4	3	78	3	111	1	22	2	84
Alby	363	2	19	0	0	1	12	0	0	0	0	0	0	0	0	1	15	3	41	1	5	1	18
Montpellier	255	1	15	1	8	0	0	1	12	0	0	0	0	0	0	2	15	3	40	1	11	1	7
Beziers	240	1	20	0	0	0	0	0	0	0	0	2	15	0	0	2	13	2	19	0	0	1	8
Nîmes	202	1	15	0	0	0	0	1	4	0	0	0	0	0	0	1	10	2	12	0	0	1	4
Viviers	534	2	74	1	8	2	22	1	12	1	15	0	0	0	0	2	7	4	30	0	0	1	8
Carcassonne	257	1	16	0	0	1	5	3	36	0	0	1	9	0	0	2	15	1	15	0	0	1	11
Le Puy	396	2	34	0	0	0	0	1	6	2	50	0	0	0	0	1	10	1	20	0	0	0	0
Montauban	95	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
St Papoul	121	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	50	1	10	0	0	0	0
Castres	184	1	13	0	0	0	0	0	0	1	9	0	0	0	0	1	80	1	15	1	7	0	0
Lavaur	208	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	20	1	14	0	0	0	0
Mirepoix	293	0	0	0	0	1	14	0	0	0	0	0	0	0	0	1	50	1	14	1	5	0	0
Uzez	323	1	2	0	0	0	0	1	23	2	30	1	4	0	0	1	2	2	8	0	0	0	0
St Pons	88	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	10	0	0	0	0	0	0
Lodève	124	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	7
Mendes	357	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	5	3	27	0	0	1	4
Agde	136	0	0	0	0	1	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Alais	168	0	0	0	0	1	8	0	0	0	0	0	0	0	0	1	4	0	0	0	0	1	4
Alet	248	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Comenge	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Rieux	123	0	0	0	0	1	5	0	0	0	0	0	0	0	0	1	10	1	8	0	0	1	6
TOTAL	5625	16	334	3	34	9	76	11	144	7	138	6	69	1	4	25	233	31	407	7	65	13	176

	H12		H13		H14		H15		H16		H17		H18		H19		H20		H21		H22		H23		H24	
Diocèse	C.	R.	C.	R.	C.	R.	C.	R.	C.	R.	C.	R.	C.	R.	C.	R.	C.	R.	C.	R.	C.	R.	C.	R.	C.	R.
Narbonne	1	12	0	0	2	26	0	0	0	0	1	12	0	0	3	36	0	0	1	8	0	0	0	0	1	8
Toulouse	2	108	2	23	2	68	1	36	1	4	0	0	1	36	3	68	1	1	1	4	0	0	0	0	0	0
Alby	2	17	0	0	4	52	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Montpellier	3	19	1	8	4	60	1	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Beziers	1	12	0	0	2	32	2	40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Nîmes	1	6	0	0	2	28	3	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Viviers	2	7	0	0	2	26	4	52	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	18	0	0
Carcassonne	2	25	1	9	3	49	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Le Puy	1	12	0	0	2	35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	4	0	0	0	0
Montauban	1	8	0	0	2	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
St Papoul	1	12	1	5	1	13	0	0	0	0	0	0	0	0	1	10	0	0	0	0	0	0	0	0	0	0
Castres	0	0	0	0	1	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Lavaur	0	0	0	0	1	5	0	0	0	0	0	0	0	0	2	13	0	0	0	0	0	0	0	0	0	0
Mirepoix	0	0	0	0	1	10	0	0	0	0	0	0	1	6	0	0	0	0	0	0	0	0	0	0	2	4
Uzez	1	8	0	0	3	27	3	29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
St Pons	0	0	0	0	0	0	2	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Lodève	1	10	0	0	0	0	1	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Mendes	1	9	0	0	4	34	0	0	0	0	0	0	0	0	1	18	0	0	0	0	0	0	0	0	0	0
Agde	0	0	0	0	2	40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Alais	0	0	0	0	3	32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Alet	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Comenge	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Rieux	0	0	0	0	1	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	20	265	5	45	42	572	17	239	1	4	1	12	2	42	10	145	1	10	2	12	1	4	1	18	3	12

4.1 Jean Le Rond d'Alembert and the triangular diagram

Triangular diagrams appear under various names in the literature, including triangular plot, triplot, trilinear, triaxial, three-element maps ... According to Suaudeau (1958, [80]), the triangular plot should also be known under the name of *d'Alembert's triangle* [21]. However, the first use of "trilinear coordinates" is usually ascribed to Josiah Willard Gibbs (1839-1903) with a publication dated 1873.

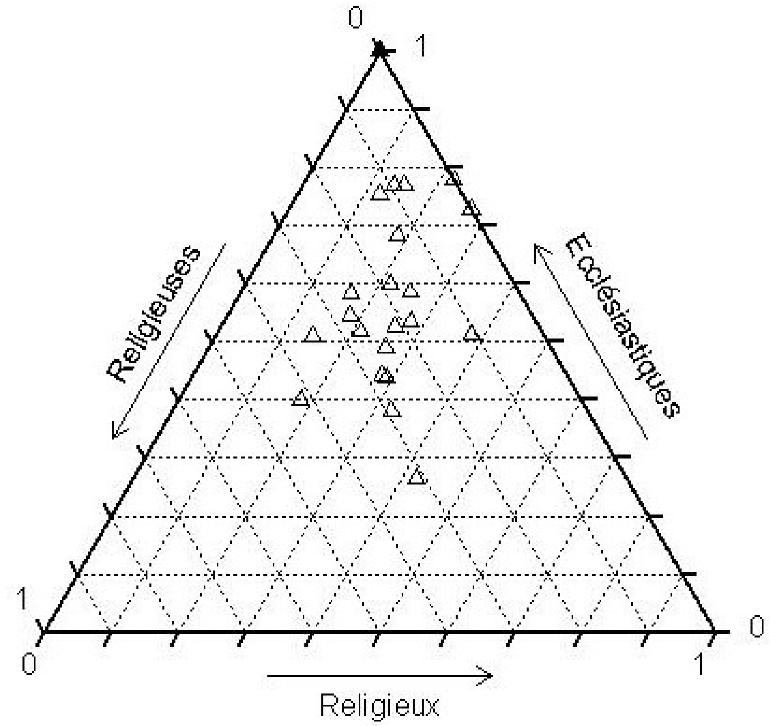


Figure 2: A "What if" in history. How d'Alembert could have visualized Basville's aggregated data derived from Table 5 and Table 6. Note that, some years later, William Playfair (1759-1823) would have drawn the diocese markers with symbols proportional to the total number persons concerned.

The principle underlying the diagram is the recognition that points $M_i, i = 1, \dots, n$, with coordinates (x_i, y_i, z_i), such that $x_i + y_i + z_i = K > 0$ and $x_i, y_i, z_i \geq 0$, in the Euclidian space of dimension 3 belong to remarkable region, an isoceles triangle, of the affine subspace of dimension 2, $x + y + z = K$. In the early use of the graphic described by Suaudeau, d'Alembert recognized that if one coordinate at a time was allowed to be negative, $x \leq 0$ say, the corresponding triangle could be also unfolded and connected to the all non negative triangle.

As reported by Suaudeau, d'Alembert considered his triangle in the context of budgets, a topic of lasting interest. Budget items are basic social measures used in quantitative Sociology. Besides, compositional data have been investigated with a variety of

21. Jean Le Rond d'Alembert (1717-1783) draw a triangular plot in the margins of Voltaire's famous book entitled *l'homme aux quarante écus* (1768), an essay in political economics, which he was reading.

methods of *l'analyse des données* under a variety of names. See for example, van der Heijden (1994, [84]).

4.2 Johann Heinrich Lambert and typographic graphics

326 Die tägliche Sonnenwärme.

Werthe von $\frac{2y}{\cos p. \cos \delta}$

Tageslängen in Stunden.

Tagesstunden.	6	8	10	12	14	16	18
3 Morgen							0,0000
4						0,0000	0,0482
5					0,0000	0,0572	0,1895
6				0,0000	0,0624	0,2172	0,4142
7			0,0000	0,0633	0,2302	0,4602	0,7071
8		0,0000	0,0620	0,2275	0,4749	0,7629	1,0482
9	0,0000	0,0524	0,2093	0,4572	0,7667	1,0993	1,4142
10	0,0413	0,1760	0,4084	0,7183	1,0756	1,4426	1,7802
11	0,1322	0,3317	0,6209	0,9785	1,3726	1,7661	2,1213
Mittag.	0,2324	0,4812	0,8147	1,2091	1,6315	2,0454	2,4142
1	0,3203	0,5954	0,9632	1,3858	1,8301	2,2697	2,6390
2	0,3356	0,6516	1,0456	1,4901	1,9512	2,3930	2,7802
3	0,2956	0,6342	1,0485	1,5102	1,9839	2,4350	2,8284
4		0,5355	0,9655	1,4397	1,9238	2,3800	2,7802
5			0,7973	1,2815	1,7733	2,2371	2,6390
6				1,0437	1,5414	2,0095	2,4142
7					1,2431	1,7144	2,1213
8						1,3717	1,7802
9							1,4142

Figure 3: A boundary or frontier graphic from the Pyrometrie (Lambert 1779, [49, page 326]). The complete version of the *Pyrometrie* can be downloaded from the *Service inter-établissement de coopération documentaire* of the University of Strasbourg.

In a well documented history of graphs and nomograms, Hankins (1999, [48]) quotes the mathematician Johann Heinrich Lambert (1728-1777), or rather Jean Henri Lambert for the French. Lambert's graphs published in a posthumous book entitled *Pyrometrie* [49] illustrate how ordinary typography suffices to produce flexible statistical graphics: see Figure 3 and Figure 4.

Lambert's typographic graphics [22] are proper precursors the graphics produced when most plotting devices connected to computers were line-printers. In what looks now prehistoric days, scree graphs for eigenvalues, scatter plots of statistical units on selected principal components, and hierarchical clustering, for *l'analyse des données*, or

22. Typographic graphics are often called "semi graphics" or "low resolution graphics". The difficulties and costs of reproduction have hindered the circulation of statistical graphics for a long time. Lithography which was in use around 1850 has somewhat alleviated that problem.

Typographic graphics use mostly ordinary characters and, sometimes, special characters. The latter are mobile printing squares or rectangles which can be typeset like ordinary printing characters to construct the columns or the rows of a diagram. Émile Levasseur, in a paper read at the International Statistical Congress, held at the Jubilee of the Statistical Society of London June 23, 1885, ([59, page 13] mentions their use in some Russian statistical publications. Lechartier, presumably a printer, presents such a printing system to the *Société de Statistique de Paris* in 1893. (See [56].) However, these cheap alternatives to lithography have not been widely used. In particular, the scalogram by Foville which is presented in Figure 7) is produced by lithography.

352 Anwendung der Theorie auf Beobachtungen.

		21	22	23	24	25	26	27	28	29	30	31	32
1 7 3 6.	Sept.					4	6	7	3				
	Oct.		1	1	3	8	6	10					
	Nov.	1	9	11	3	5							
	Dec.	5	15	6									
1 7 3 7.	Febr.			4	13	11							
	Merz						16	14					
	Apr.						1	8	15				
	May							1	14	15	1		
	Jun.									10	9	8	2
	Jul.							2	7	8	3		
	Aug.						4	12	8	5	1		
	Sept					1	5	10	5	8			
	Oct.	1	4	3	3	3	8	5					
	Nov.	3	6	11	6								
	Dec.		4	10	8								
1 7 3 8.	Febr.				7	11	7						
	Merz					11	19						
	Apr.						5	15	5				
	May								9	6	4	7	3
	Jun.								3	19	5		
	Jul.								5	13	12		
	Aug.					4	2	6	8	8			
	Sept						7	5	14	1			

Figure 4: An observed periodic function from the Pyrometrie (Lambert 1779, [49, page 352]).

stem and leaf displays and boxplots, for exploratory data analysis, were line printed on listings!

4.3 Toussaint Loua, 1873, and Alfred de Foville, 1888

Toussaint Loua (1824-1907) and, some years later, Alfred de Foville (1842-1913), both active members of the *Société de statistique de Paris*, have addressed the simultaneous visualization of a large set of discreticized variables. Their graphical solutions are presented below.

Loua's *Atlas statistique* [23] essentially presents current data on the population of Paris as characterized by 40 quantitative variables observed on the 20 *Arrondissements* (Districts) of Paris (1873, [61]). Each variable is used to rank the *Arrondissements* on a five-level scale (two levels lower than "average", "average", and two levels upper than "average") defined by *ad hoc* cut points; an interview of Loua on the computing process of their values was published by Georges Renaud in the *Revue géographique* and later reprinted in the *Journal de la Société de statistique de Paris* (1877, [74]). Paris cartograms corresponding to the 40 variable are drawn, the colour code being white, yellow, blue, orange, violet for the 5 ordered levels [24] . How to summarize these 40 cartograms? The answer is a "graphical general summary", see Figure 6, which simultaneously displays the values of 40 discreticized variables (40 columns) observed

23. A black and white scan of Louas'atlas can be downloaded from gallica.bnf.fr.

24. Loua's five levels scheme follows one of the scales recommended by Levasseur in his Royal Statistiscal Association 1885 Jubilee paper [59, page 24]. However, Loua's colour scheme is inappropriate for black-and-white photocopying or scanning (of which he was obviously not aware!). It is quite surprising that the color scheme (four colors and two metals) used in Heraldry for the black-and-white representation of a coat of arms have never been knowingly considered in statistical graphics.

```
          BOX PLOT OF VARIABLE: COMPOSIT    , N =    54

   0.00                                               18.00
MINIMUM                                             MAXIMUM
                         ---------------+--------
-------------------------+              +       +----------
                         ---------------+--------


          BOX PLOT OF VARIABLE:  DRAWING    , N =    54

6.00                                                  18.00
MINIMUM                                             MAXIMUM
                  ------------------+-------
------------------+                 +      +---------------
                  ------------------+-------


          BOX PLOT OF VARIABLE:   COLOUR    , N =    54

0.00                                                  18.00
MINIMUM                                             MAXIMUM
                      ----------+--------------------
----------------------+         +                   +------
                      ----------+--------------------


          BOX PLOT OF VARIABLE: EXPRESSI    , N =    54

0.00                                                  18.00
MINIMUM                                             MAXIMUM
            ------+--------------------
------------+     +                   +--------------------
            ------+--------------------
```

Figure 5: Semi-graphic boxplots of the painters data using MYSTAT (Version 2.1, a DOS operated popular statistical software program) distributed for free in 1992.

on the 20 *Arrondissements* (20 lines) by their colour code.

Foville's article addresses the evolution over 11 years (1877-1887) of French economic indicators (1888, [36]). Humorously, its title is "essay in economic and social meteorology". The idea is to record and visualize the simultaneous variations over time of the 32 descriptors. Again the answer is graphical, see Figure 7 which is nserted between pages 248 and 249 of the *Journal de la Société de Statistique de Paris* [36]. (For a recent presentation see [50].) Fovilles uses a four-level scale for each economic indicator $\times$ year. The four categories are ordered: black means a bad year, red a good year; light colouring and hatching shade mean mediocre.

Loua does not comment his summary. His objective is limited to an easy reading of the level achieved by the discreticized version of the 40 variables on any statistical unit (*Arrondissement*). Other statistical graphics are possible, e.g. a parallel coordinate

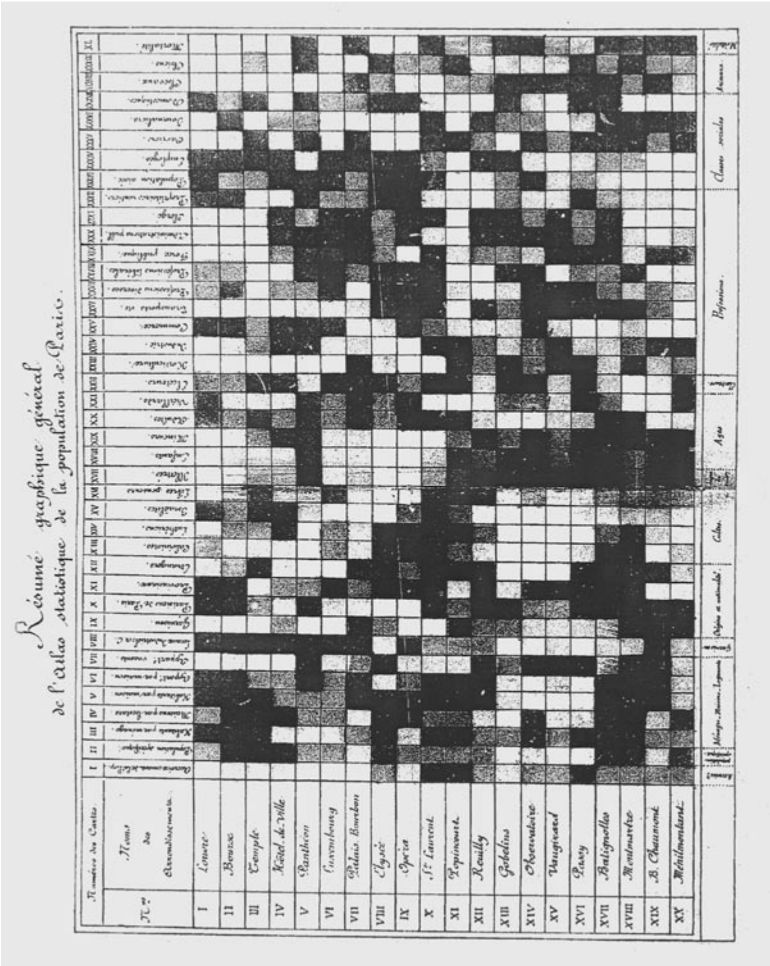


Figure 6: Loua's scalogram. The full *Atlas statistique de la population de Paris* (18773, [61]) can be downloaded from gallica.bnf

plot of the centered and standardized variables. Would they bring more? Foville goes further than that and derives a general view of the evolution of the French economy from the pattern of colour variations [25].

Both graphics exemplifies what is called now a coloured scalogram. The construction and the optimization of scalograms [26] has been considered by Jacques Bertin (1977, [9]), Jean-Paul Benzécri and other authors (See Valois 2000, [83, pages 100-101]).

25. Foville recognizes a difficulty in the interpretation of some variables: an increase for some items, for example the number of suicides, is negatively viewed while an increase for others, for example iron production, is positively viewed.

26. Neither Loua nor Foville discuss the optimization of their scalograms by reordering the columns (Loua) or the lines (Foville) since the ordering of the other dimension (*Arrondissements*, or years) can be considered as fixed.

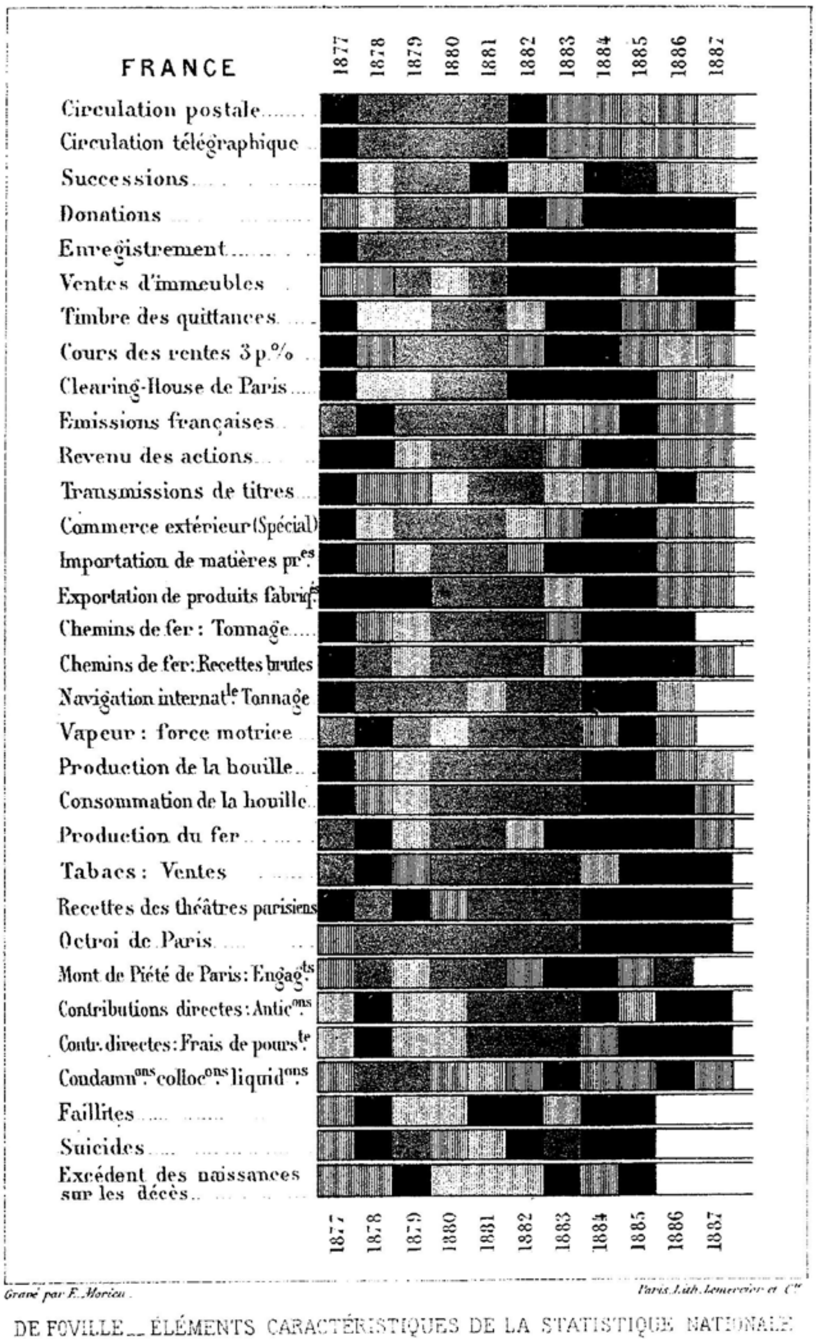


Figure 7: Foville's scalogram. The original scalogram uses two colours (black and red) in conjonction with two shades (dark and light). When light, the colour is superimposed onto hatched lines. The four categories are ordered: black means a bad year, red a good year; light color and hatching shade mean mediocre. White means missing.

4.4 Early 3D

During the XIXth century, pedagogical devices were developed for educating the view in three dimensions. Several strategies were explored. Firstly, the XIXth century saw the development of “descriptive geometry”, theoretical and applied (engineering drawing), which was introduced by Gaspard Monge (1728-1777). But descriptive geometry was found difficult by many and especially in education.

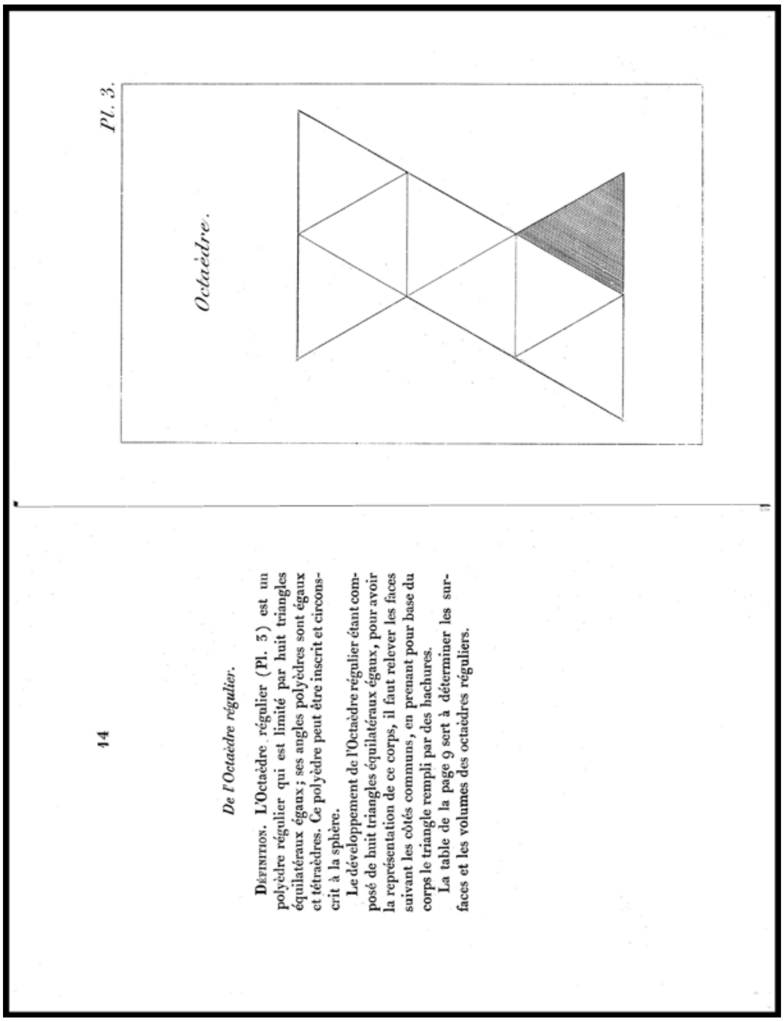
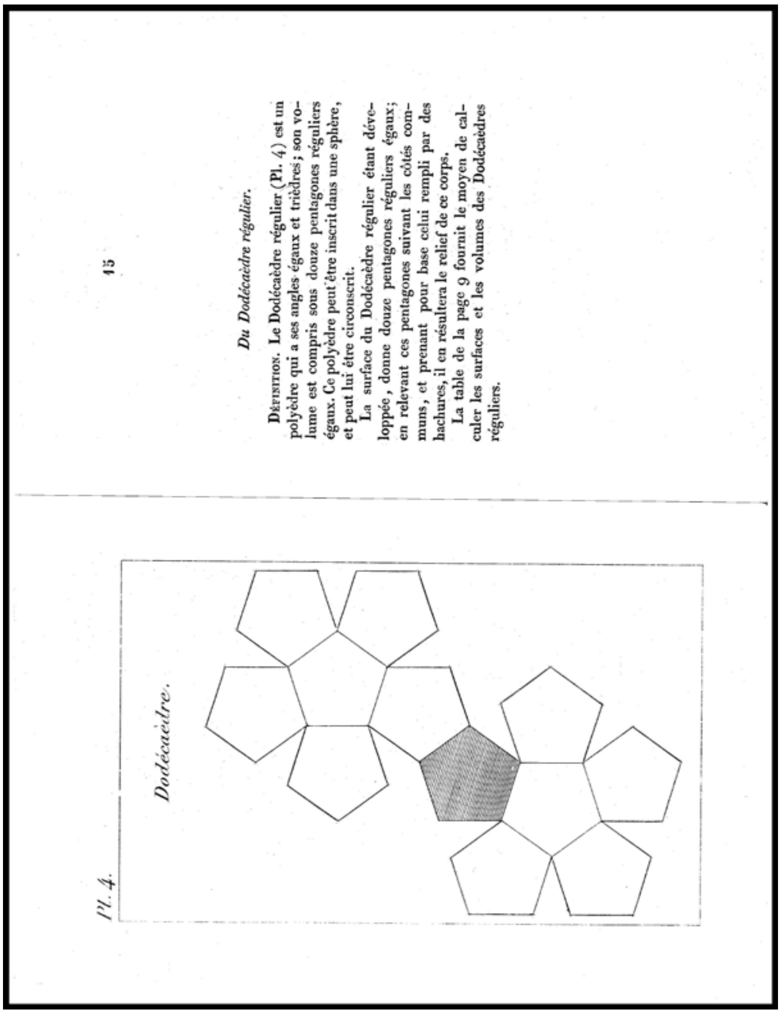


Figure 8: Marie’s regular octahedron to be cut out and folded (see [64, p. 14-Pl. 3].)

Less mathematics loaded solutions were also explored. One was to construct collections of solids. This approach is exemplified in France by the translation by François-Charles-Michel Marie (1835, [64]) of the English book by John-Lodge Cowley [27]. Marie’s book contains 24 engraved plates. These are all partly cut out except the part

27. John-Lodge Cowley (1719-1787?) was a professor of mathematics at the Royal Military Academy at Woolwich. Cowley was interested in applied mathematics, especially solid geometry. A presentation of the Cowley’s textbook (1787, [21]) is available on line from the library and information center of Georgia Tech (USA). How was his work introduced in France?

that corresponds to the base of the figure. The form of the figure is exhibited by raising up the different parts upon their base (see Figure 8 and Figure 9). Figure 8 is obviously connected with d'Alembert's triangle while Figure 9 can be regarded as a step in the direction of "dodecahedral views" of multivariate data discussed Paul Tukey and John Tukey (1981,[82]).


15

Du Dodécaèdre régulier.

Définition. Le Dodécaèdre régulier (Pl. 4) est un polyèdre qui a ses angles égaux et trièdres; son volume est compris sous douze pentagones réguliers égaux. Ce polyèdre peut être inscrit dans une sphère, et peut lui être circonscrit.

La surface du Dodécaèdre régulier étant développée, donne douze pentagones réguliers égaux; en relevant ces pentagones suivant les côtés communs, et prenant pour base celui rempli par des hachures, il en résultera le relief de ce corps.

La table de la page 9 fournit le moyen de calculer les surfaces et les volumes des Dodécaèdres réguliers.

Figure 9: Marie's regular dodecahedron to be cut out and folded (see [64, Pl. 4-p. 5]).

The construction of solid stereograms in plaster was also tried. Éric Brian tells the story of a solid stereogram similar to that of Luigi Perozzo [28] (1881) in which André Breton saw the very example of a surrealist object (2001, [15]).

Another vein was to exploit stereoscopic vision. Henry Vuibert, a mathematician by training and the founder of the eponym French publishing house, *Vuibert Éditions*, did his best to promote the anaglyphs considered by Louis Ducos du Hauron [29]. Vuibert

28. There are few details on Luigi Perozzo's life such as date of birth, date of death, career ...
29. Louis Ducos du Hauron (1837-1920) is also credited of the invention of colour photography.

(1912, [87]) was foreseeing that, by improving the printing process [30], low cost perfect 3D graphics would be seen through the special red and green pair of glasses. So far, this looks like a dead lead in statistics or at best a vague protohistoric form of "Grand tours" [31] or, more ambitiously, of "virtual reality" [32].

5 The "Golden Age" of statistical graphics

Both Loua and Foville belonged to the "Golden Age" of statistical graphics which is rapidly evoked below. This terminology was introduced by Funkhouser to describe a period of time during which graphical representation of statistical data were the obliged companion of published official tables of data. A wide variety of graphics were considered and advocated by talented statisticians: Émile Levasseur (1828-1911) and Émile Cheysson (1836-1910) in France, and Georg van Mayr (1841-1925) in Germany to name but a few. Specialized articles on statistical graphics written in French were widely read. Articles in German were translated. For example, Mayr was translated into French and Italian. (See a bilingual German and French text (1897, [1]), and an Italian translation (1879, [76]) with statistical diagrams.)

Notwithstanding their degree of sophistication, the graphics used during the Golden Age have not much in common with those of *l'analyse des données*. Still they can teach us lessons in the semiology [33] of data visualization.

5.1 The Golden age

The "Golden age of Graphics", between 1850-1900 ($\pm$ 10), is well documented. See Funkhouser (1937,[38]), Robinson (1982,[75]), Palsky (1996, [71]), Friendly (2009, [37]).

A monument of the Golden Age, its "pinnacle" according to Friendly (2009, [37]), is the series of annual *Album de statistique graphique* published by the *Ministère des travaux publics* (Ministry of Public Works) and edited by Émile Cheysson (1836-1910) [34]. The first *Album* [35] was published in July 1879 and the series was discontinued in 1900. The *Albums* portray recurrent data for the largest part but also occasional ones. Most statistical graphics in them are of an outstanding quality.

Other French administrations published similar albums [36], e. g. an *Album de statistique graphique du Service Vicinal* (1883), or *Nouvelle évaluation du revenu foncier des propriétés non bâties de la France* (1884), or Jacques Bertillon's *Atlas de Statistique Graphique de la ville de Paris* (1888). Individual researchers also published graphical

30. As reported by Vuibert ([87, page 15]): "During the International meeting of mathematicians, held in August 1912 at Cambridge, M. Richard has exhibited about fourty anaglyphs which have been greatly admired. We shall be able, I hope, to produce perfect ones by printing but we do not yet arrived to. The difficulty lies in the complexity of the composition of printing inks . . . "

31. The image grand tour is a method for visualizing multispectral images or multiple registered images.

32. Virtual reality describes a variety of applications, commonly associated with its immersive, highly visual, 3D environments. it is associated with a technology which allows a user to interact with a computer-simulated environment.

33. For a comprehensive book on image semiology see Bertin, (1977, [9]).

34. Émile Cheysson has been an active member of the *Société de statistique de Paris*.

35. Toussaint Loua did not forget to announce the publication of the first volume in the *Journal de la Société de statistique de Paris* [62].

36. One can see in these albums a distant sign of the use of *l'analyse des données* by French Official Statistics?

statistics. Toussaint Loua's *Atlas statistique de la population de Paris* (1873) has already been considered in subsection 4.3 [37]. But there are many others around the world and an exhaustive worldwide list of albums is yet to be be made.

5.2 The objectives of graphics

Another leading statistical "graphist" in France is Émile Levasseur (1828-1911) who was invited to review *la statistique graphique* at the meeting held on the occasion of the Jubilee of the Statistical Society of London in 1885 (1885, [59]). Levasseur also made a related presentation of the topic at the *Société de Statistique de Paris* (1886, [60]). In these papers, Levasseur distinguishes two main objectives and an aside for graphics:

- demonstration,
- invention,
- control.

These can be summarized as follows. Demonstration graphics are intended for wide circulation. Invention graphics are temporary graphics used in the process of research. Control graphics are used to detect outliers in the data. While a demonstration graphic must obey the rule of clarity (they are to reveal on first sight a general tendancy and on further investigation possible irregularities of interest), an invention graphic is ephemeral and dictated by the personal feelings of its user. Obviously, the object of a graphic impacts its construction.

Surely, invention graphics play an important in the course of *l'analyse des données*. Statistical softwares have widely facilitated their production [38]. It would be of pedagogical interest to have an estimate of the ratio of demonstration to invention graphics in large multivariate data analyses.

5.3 And nomograms?

A nomogram, nomograph, or abac is a graphical calculating device. This planar diagram is designed to allow the approximate computation of the value of a function with multiple variables. As reported in Hankins (1999, [48]) and Friendly (2009, [37]), innovative nomograms were designed in France [39] under the impulsion of creative mathematical engineers, Léon Lalanne (1811-1892) and Maurice d'Ocagne (1862-1938), to name but two. In his Statistical Society of London Jubilee article (1885, [59]), Levasseur briefly mentions the work of Lalanne but does not seem to really recognize the potential of nomograms. To some extent, this is still the case. To my knowledge, there has been no attempt to combine explicitly the methods of *L'analyse des données* with the geometrical principles underlying the construction of nomograms. Yet, a calibrated biplot graphic, used to approximate the elements of a matrix, performs like a nomogram. The same apply to more elaborate situations: skew-symmetry in square tables (Gower and Digby, [44, section 3]), complex decompositions (Dossou-Gbété and Grorud, [30]), or distances in hierarchical classification trees.

37. Renaud (1877, [74]), in his "interview" of Loua's discretization process, mentions two atlas: Loua's atlas and an *Atlas industriel* published by the Ministry of Trade.

38. Some researchers are old enough to remember the long computer listings on which typographic graphics were printed: series of scatter diagrams on axes 1 and 2, axes 1 and 3, axes 2 and 3 … in factorial analyses, or huge classification trees in cluster analysis.

39. Hankins's article [48] is motivated by a nomogram which resulted from a collaboration between the French Maurice d'Ocagne and the American Lawrence Joseph Henderson. Henderson's famous nomogram is an attempt to quantify the relationships in a physico-chemical system which models mammalian blood.

5.4 The potential sin of the Golden Age

In the instructions of the *Album de Statistique Graphique* established by an administrative decision dated 12 mars 1878, it is stated that statistical maps and diagrams are to be published on a yearly basis in order to ease comparisons over time of economic and technical facts. It is also recalled in the instructions of several following albums that the duty of the Administration is to publish data of good quality and that the methodology used in each graph is to be recalled. It is further emphasized that in no ways the Administration is to provide interpretations. On the contrary, it is stated that the readers (Public and Press) are to draw conclusions at their own risks and dangers (see for example, the *Instructions* for the year 1883).

This seems at best a hazardous standpoint or worse a potential sin. The potential sin might be the emphasis on storing in a graphical format the numbers making the data rather than patterns in the data. First, it seems at least conjectural to profess that data or graphical representations of data can be collected without any objective on mind. The early examples discussed so far showed some form of purpose in their production which was additionally reflected by the layout of their publications. In the same line, the selection of data presented in the Album are telling something about economic progress in France as viewed by high ranked civil servants. Finally, are these maps strictly and accurately reproducing the data they visualized? No one has really checked. Some data may have been "smoothed" to provide attractive diagrams if not to disentangle the information they carry. A statistical graphic should be understood as the transposition of an abstract reality (the data) to a concrete fiction (its graphical representation). In this process, it may be helpful to paraphrase the statistical concept of sufficiency: the data given the graphical representation should be independent of the objective of the work. Obviously, most of the remarks above apply to *l'analyse des données* and its graphics.

6 The "between" introduction of probabilistic models

In this section, two problems illustrate the second form of the unavoidable reintroduction of probability in *l'analyse des données* introduced in subsection 2.1. One relates to cluster analysis and the other to factorial methods.

6.1 Example 1: central partition and latent class

Let $\mathcal{F}$ be a finite set of qualitative variables observed on a given (sample of a) population $\mathcal{S}$. Let $|\mathcal{F}|$ be the cardinality of $\mathcal{F}$. Each observed variable, A say, is equivalent to an observed partition of $\mathcal{S}$ whose classes correspond to the levels of A.

Following a descriptive approach, one can look for a "central" partition which would be to the observed partitions what is an ordinary mean to a statistical series of numbers. This problem has been investigated by Simon Régnier (1932-1980): the central partition P should minimize the average of the cardinal of the symmetric difference between P and partitions A in $\mathcal{F}$ like the mean minimizes the average of the sum of squared differences. Let $\mathbb{1}_A(i', i'')$ be the indicator of i' and i'' being in the same class of a given partition A. Then the central partition $\overline{P}$ of the family $\mathcal{F}$ of partitions is a solution of the following optimization problem:

$$\overline{P} \quad \in \quad \arg\min_P \sum_{i' \in \mathcal{S}} \sum_{i'' \in \mathcal{S}} \mathbb{1}_P(i', i'') \left(d(i', i'') - \frac{1}{2} \right)$$

where

$$d(i', i'') \;=\; 1 - \frac{1}{|\mathcal{F}|} \sum_{A \in \mathcal{F}} \mathbb{1}_A(i', i'')$$

Régnier elicited the connection of this problem with linear programming and proposed an algorithm leading to a "local" optimum (over the set of partitions obtained by transfer of one and only one element to any other class).

In the tradition of statistical modeling one can alternatively consider a latent class model. The latent class model, introduced by Paul Lazarsfeld (1950,[51]), was well received in France through the work of Raymond Boudon[40] (1967, [12]). See also Erling Andersen[41] (1994, [3, section 12]). Let y_f^F be the observed counts in cell f, f being an element of the Cartesian product of variables A in $\mathcal{F}$. A latent class model is an unobserved qualitative variable P (thus a partition) defined on $\mathcal{S}$ with levels p such that the variables in $\mathcal{F}$ are independent given any level p of P. Since P is unobserved, it is the solution of an optimization problem which can be solved by the EM algorithm (Goodman, [42]).

How do these two approaches relate? Is the latent partition approximately central? A precise answer is still unclear but the question is typical of the "within" approach to the introduction of probabilities in *l'analyse des données*. A dual investigation would consider the relationship between the eigenvectors in multiple correspondence analysis and continuous latent variable models.

6.2 Example 2: incomplete and unbalanced design of the data

A quantitative response (continuous, counts or 0/1) and two cross-classifying factors are observed. In an incomplete or in an unbalanced design, some of the cells resulting from the cross-classification may be structurally missing or the number of observed responses may not be constant over the cells. Therefore the data cannot be coerced into matrix form and statistical analyses using (generalized) singular value decomposition cannot be applied[42]. However, a flexible optimization problem based on the formulation of generalized linear model developed in statistical modeling can still be defined.

$$\underline{b} \;\in\; \arg\min_{\underline{\beta}} \sum_{\ell} w_\ell \left(y_\ell - m(\eta_\ell)\right)^2$$

where

$$\eta_\ell(\underline{\beta}) \;=\; \beta_0 + \beta^I_{I(\ell)} + \beta^J_{J(\ell)} + \sum_{k=1}^{K} \delta_k \beta^I_{k,I(\ell)} \beta^J_{k,J(\ell)}$$

where: ℓ is a statistical unit, y_ℓ denotes the corresponding response value, $I(\ell)$ and $J(\ell)$ the corresponding levels the cross-classifying factors I and J; m, the mean function, is a strictly monotone function; $w_\ell = p_\ell \phi V(m(\eta_\ell)))$ is a weight which possibly combines an *a priori* weight p_ℓ, the variance function (a positively valued function of the mean $V(m(\eta_\ell)))$, and a dispersion parameter ϕ. Identification constraints need

40. Raymond Boudon visited Paul Lazarsfeld in the US. In the published version of his dissertation the first acknowledgements go to Paul Lazarsfeld (1901-1976) and Robert Fortet (1912-1998) and Jean Stoetzel (1910-1987).

41. Erlig Andersen (1939-2004) should be recognized in France for presenting both model based and correspondence analysis methods in his book on categorical data analysis.

42. The solution to the optimal approximation of a data matrix by a matrix of lower rank is usually ascribed to Carl Eckart and Gale Young (1936, [32]).

to be introduced for the parameter $\underline{b}$, e.g. the orthonormality constraints used in principal component analysis. Note that some parameters in the linear part of the model $(\beta_0 + \beta^I_{I(\ell)} + \beta^J_{J(\ell)})$ may be omitted to replicate particular methods.

This approach generalizes the problem of lower rank approximation to data sets which are not matrix coercible. It has been introduced by K. Ruben Gabriel (1979,[40]) for known variances. It comprises methods such as principal component analysis, a variety of correspondence analyses (symmetric, non-symmetric ...) and Goodman's RC association model. It can be further generalized to accommodate correlations between the observations [43] :

$$\underline{b} \quad \in \quad \arg\min_{\underline{\beta}} ||\underline{y} - m(\underline{\eta})||^2_{W(\underline{\beta})}$$

where $\underline{y}$ and $\underline{\eta}$ are the response and the predictor vectors.

In the class of extension above, the core of factorial methods is preserved, namely the row and column inner product form for the reduced rank interaction

$$\sum_{k=1}^{K} \delta_k \beta^I_{k,I(\ell)} \beta^J_{k,J(\ell)},$$

Benzécri's correspondence analysis, Nishisato's dual scaling (1970, [69]), Gabriel's biplot (1972, [39]), Goodman's RC association (1979, [43]) ... Curiously a simple algorithm for the estimation of the $\underline{\beta}$, is an adaptation of the old NIPALS algorithm considered by Herman Wold (1966, [89]) and central in PLS [44] analyses.

7 Concluding remarks

The main factors which have contributed, in the late sixties, to a remarkable revival of multivariate descriptive analysis in France have been considered. Clearly the impetus of *l'analyse des données* has stimulated statistical research in France and has produced novel methodologies. Did a distinctive and long-lived school of statistics emerge? To some extent this was true in the eighties. Nowadays, the globalization of statistical practices as well as the mobility of researchers hinders the very existence of isolated national developments. But affinity groups remain as can be seen from the programs of contemporary international meetings.

Without much risk, one can forecast a long live for *l'analyse des données* under the banner of multivariate data analysis.

References

[1] *Atlas graphique et statistique de la Suisse; Graphisch-statisticher Atlas der Schweiz.* Buchdruckerei Stämpfli & Cie, Bern, 1897.

[2] Murray A. Aitkin, Dorothy A. Anderson, Brian J. Francis, and John P. Hinde. *Satistical modelling in GLIM.* Oxford University Press, Oxford, 1989.

43. Other generalization are possible: explanatory variables can be introduced in the predictor $\eta_\ell(\underline{\beta}) = \beta_0 + \beta^I_{I(\ell)} + \beta^J_{J(\ell)} + \sum_{k=1}^K \delta_k \beta^I_{k,I(\ell)} \beta^J_{k,J(\ell)}$; the "row" $\beta^I_{k,I(\ell)}$ scores and the "column" scores $\beta^J_{k,J(\ell)}$ can be constrained to belong to separate explanatory spaces ...

44. PLS stands for Partial least squares.

[3] Erlig B. Andersen. *The statistical analysis of categorical data*. Springer-Verlag, Berlin, 3rd edition, 1994.

[4] Nicolas de Lamoignon de Basville. *Mémoires pour servir à l'histoire de Languedoc*. Pierre Boyer, Amsterdam, 1734.

[5] Jean-Paul Benzécri. *L'analyse des données. I la Taxinomie*, volume 1. Dunod, Paris, 1973.

[6] Jean-Paul Benzécri. *L'analyse des données. II l'analyse des correspondances*, volume 2. Dunod, Paris, 1973.

[7] Jean-Paul Benzécri. Histoire et préhistoire de l'analyse des données, I, II, III, IV. *Cahiers de l'analyse des données*, 1(1, 2, 3, 4), 1976.

[8] Patrice Bertier and Jean-Marie Bouroche. *Analyse des données multidimensionnelles*. PUF, Paris, 1975.

[9] Jacques Bertin. *La graphique et le traitement graphique de l'information*. Flammarion, Paris, 1977.

[10] Lynne Billard and Edwin Diday. *Symbolic data analysis: conceptual statistics and data mining*. Wiley, 2006.

[11] Yvonne M. M. Bishop, Stephen E. Fienberg, and Paul W. Holland. *Discrete multivariate analysis; Theory and practice*. The MIT Press, Cambridge (Massachusetts), 1975.

[12] Raymond Boudon. *L'analyse mathématique des faits sociaux*. Plon, Paris, 1967.

[13] Pierre Bourdieu. *Homo academicus*. Editions de Minuit, Paris, 1984.

[14] Jean-Marie Bouroche and Gilbert Saporta. L'analyse des données. *Pour la Science*, 5:23–35, 1978.

[15] Éric Brian. Les objets de la chose. théorie du hasard et surréalisme au XXe siècle. *Revue de synthèse*, (2-3-4):473–502, 2001.

[16] Francis Cailliez and Jean-Pierre Pagès, editors. *Introduction à l'analyse des données*, Paris, 1976. SMASH.

[17] Henri Caussinus. Models and uses of principal component analysis; a comparison emphasizing graphical displays and metric choices. In *Multidimensional data analysis*, pages 149–170, Leiden, 1986. DSWO Press.

[18] Henri Caussinus. Modèles probabilistes et anlyse des données multidimensionnelles. quelques réflexions méthodologiques et applications. *Journal de la Société de Statistique de Paris*, tome 134, N˚2, 1993.

[19] Pierre Cazes. Compte rendu de lecture de "multiple correspondence analysis and related methods". *Journal de la Société de Statistique de Paris*, tome 149, N˚3, 2008.

[20] Gilles Celeux and Edwin Diday. L'analyse des données. *Bulletin de liaison de la recherche en informatique et automatique*, 62:2–12, 1980.

[21] John Lodge Cowley. *An illustration and mensuration of solid geometry in seven books: containing forty-two moveable copper-plate schemes for forming the various kinds of solids and their sections ..., by John Lodge Cowley ; revised, corrected, and augmented by William Jones*. printed by S. Cosnell ... for the editor, London, 1787.

[22] Jacques Dauxois and Alain Pousse. *Les analyses factorielles en calcul des probabilités et en statistique : essai d'étude synthétique*. PhD thesis, Laboratoire de statistique et probabilités, Université de Toulouse III, 1976.

[23] Jacques Dauxois, Alain Pousse, and Yves Romain. Asymptotic theory for the principal component analysis of a random vector fonction: Some applications to statistical inference. *Journal of multivariate analysis*, 12:136–154, 1982.

[24] Michael Davenport and W. Gerald Studdert-Kennedy. The statistical analysis of aesthetic judgment: an exploration. *Applied Statistics*, 21(3):324–331, 1972.

[25] Pedro de Paz Salas and António Álvares. *La felicissima armada que elrey Don Felipe nuestro Señor mandó juntar enel puerto de la ciudad de Lisboa enel Reyno de Portugal. El año de mil y quinientos y ohcenta y ocho*. Lisbao, 1588.

[26] Jean-Claude Deville and Edmond Malinvaud. Data analysis in official socio-economic statistics (with discussion). *Journal of the Royal statistical Society, series A*, 146(4):335–361, 1983.

[27] Edwin Diday and collaborateurs. *Optimisation en classification automatique, vol.I and vol.II*. INRIA-Sedis/Diffusion, 1980.

[28] Edwin Diday and Ludovic Lebart. L'analyse des données. *La Recherche*, 74:15–25, 1977.

[29] Edwin Diday, Jacques Lemaire, Jean Pouget, and Françoise Testu. *Éléments d'analyse des données*. Dunod-Bordas, Paris, 1982.

[30] Simplice Dossou-Gbété and Axel Grorud. Biplots for matched two-way tables. *Annales de la faculté des sciences de Toulouse*, 11(4):469–483, 2002.

[31] François-Jacques Durand. *Statistique élémentaire, essai sur l'état géographique, physique et politique de la Suisse*. Lausanne, 1795.

[32] Carl Eckart and Gale Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1, 1936.

[33] David Edwards. *Introduction to Graphical Modelling*. Springer-Verlag, New York, 2nd edition, 1995.

[34] Brigitte Escofier. *Analyse des Correspondances - Recherches au coeur de l'analyse des données*. Presses universitaires de Rennes, 2003.

[35] Yves Escoufier. Cours d'analyse des données. année scolaire 1978-1979, 1978.

[36] Alfred de Foville. Éssai de météorologie économique et sociale. *Journal de la Société de Statistique de Paris*, 1ère Série, 29e Volume, N°6:243–249, 1888.

[37] Michael Friendly. The golden age of statistical graphics. *Statistical Science*, to appear.

[38] H. Gray Funkhouser. Historical development of the graphical representation of statistical data. *Osiris*, 3(1):269–405, 1937.

[39] K. Ruben Gabriel. Biplot display of multivariate matrices for inspection of data and diagnosis. In Vic Barnett, editor, *Interpreting multivariate data*, pages 147–173. John Wiley & sons, 1981.

[40] K. Ruben Gabriel and Shmuel Zamir. Lower rank approximation of matrices by least squares with any choice of weights. *Technometrics*, 21(4):489–498, 1979.

[41] Albert Gifi. *Nonlinear multivariate analysis*. John Wiley and sons, Chichester, 1990.

[42] Leo A. Goodman. On the exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, 61:215–231, 1974.

[43] Leo A. Goodman. Simple models for the analysis of association in cross-classifications having ordered categories. *Journal of the American Statistical Association*, 74:537–552, 1979.

[44] John C. Gower and Peter G. Digby. Expressing complex relationships in two dimensions. In Vic Barnett, editor, *Interpreting multivariate data*, pages 83–118. John Wiley & sons, 1981.

[45] Michael Greenacre. *Correspondence analysis in practice*. Chapman and Hall, 1st edition, 1993.

[46] Michael Greenacre. *La práctica del análisis de correspondencias*. Fundación BBVA, 2008.

[47] Michael Greenacre and Jeorg Blaesius, editors. *Multiple correspondence analysis and related methods*. Chapmann & Hall, 2006.

[48] Thomas L. Hankins. Blood, dirt and nomograms. a particular history of graphs. *Isis*, 90:50–80, 1999.

[49] Johann Heinrich Lambert. *Pyrometrie oder vom masse des feuers und der wärme mit acht kupfertafeln*. Berlin, 1779.

[50] Dominique Ladiray. Conjoncture, statistique et économétrie. *Économie et Statistique*, 359-360:3–12, 2002.

[51] Paul F. Lazarsfeld. The logical and mathematical foundation of latent structure analysis. In Samuel A. Stouffer, Louis Guttman, Edward A. Suchman, Shirley A. Star, and John A. Clausen, editors, *Measurement and prediction*, Princeton, 1950. Princeton University Press.

[52] Paul F. Lazarsfeld. Notes on the history of quantification in sociology - trends, sources and problems. *Isis*, 52(2):277–333, 1961.

[53] Ludovic Lebart and Jean-Pierre Fénelon. *Statistique et informatique appliquée*. Dunod, Paris, 1971.

[54] Ludovic Lebart, Alain Morineau, and Nicole Tabard. *Techniques de la description statistique : Méthodes et logiciels pour l'analyse des grands tableaux*. Dunod, Paris, 1977.

[55] Ludovic Lebart, Alain Morineau, and Kenneth M. Warwick. *Multivariate descriptive statistical analysis: correspondence analysis and related techniques for larges matrices*. Wiley & sons, New York, 1984.

[56] Lechartier. Les graphiques typographiques. *Journal de la Société de Statistique de Paris*, pages 311–321, 1893.

[57] Israël César Lerman. *Les bases de la classification automatique*. Gauthier-Villars, 1970.

[58] Israël César Lerman. Historique de la société francophone de classification (sfc). *Bulletin de la SFC*, 10, 1998.

[59] Émile Levasseur. *La statistique graphique, read at the International Statistical Congress held at the jubilee of the Statistical Society of London, June 23, 1885*. Edward Stanford, London, 1885.

[60] Émile Levasseur. Quelques reflexions sur la statistique graphique. *Journal de la Société de Statistique de Paris*, 27:224–225, 1886.

[61] Toussaint Loua. *Atlas statistique de la population de Paris*. J. Dejey & C^{ie}, Paris, 1873.

[62] Toussaint Loua. L'album de statistique graphique pour 1879. *Journal de la Société de statistique de Paris*, 1re Série, Volume 20(11):308, 1879.

[63] Kanti V. Mardia, John T. Kent, and John M. Bibby. *Multivariate analysis*. Academic Press, London, 1979.

[64] François Charles Michel Marie. *Géométrie stéréographique, reliefs des polyèdres, pour faciliter l'étude des corps*. Bachelier, Paris, 1835.

[65] Michel Masson. *Méthodologie générale du traitement de l'information de masse*. Nathan, Paris, 1980.

[66] Peter McCullagh and John A. Nelder. *Generalized Linear Models*. Chapman & Hall, 2nd edition, 1989.

[67] Paul McReynolds and Klaus Ludwig. Christian Thomasius and the origin of psychological rating scales. *Isis*, 73(3):546–553, 1984.

[68] John A. Nelder and Robert W. M. Wedderburn. Generalized linear models. *Journal of the Royal Statistical Society, Series A*, 135:370–384, 1972.

[69] Shizuhiko Nishisato. *Analysis of categorical data. Dual scaling and its applications*. University of Toronto Press, Toronto, 1970.

[70] Jean-Pierre Pagès, Francis Cailliez, and Yves Escoufier. Analyse factorielle : un peu d'histoire et de géométrie. *Revue de Statistique Appliquée*, 27(1):5–28, 1979.

[71] Gilles Palsky. *Des chiffres et des cartes : naissance et développement de la cartographie quantitative française au XIX^{e} siècle*. Comité des Travaux Historiques et Scientifiques (CTHS), Paris, 1996.

[72] Roger de Piles. *Cours de peinture par principes composé par M^{r} de Piles*. Jacques Estienne, Paris, 1708.

[73] Roger de Piles. *Recueil de divers ouvrages sur la peinture et le coloris par M^{r} de Piles*. Charles Antoine Jombert, Paris, 1755.

[74] Georges Renaud. Procédé de cartographie statistique. *Journal de la Société de statistique de Paris*, 1re Série, Volume 18(1):27–28, 1877.

[75] Arthur H. Robinson. *Early Thematic Mapping in the History of Cartography*. University of Chicago Press, Chicago, 1982.

[76] Giovani Battista Salvioni. *La statistica et la vita sociale*. Ermanno Loescher, Torino e Roma, 1879.

[77] Gilbert Saporta. Données supplémentaires sur l'analyse des données. *L'écho des messaches*, 4:4–5, 1975.

[78] Gilbert Saporta. L'analyse des données. *MATAPLI, Bulletin de liaison de la Société de mathématiques appliquées et industrielles (SMAI)*, 2:4–6, 1984.

[79] Charles Spearman. *Les aptitudes de l'homme, leur nature et leur mesure (translated by François Brachet)*. Conservatoire national des arts et métiers, Paris, 1936.

[80] René Suaudeau. *Les représentations figurées des physiocrates*. Librairie du Recueil Sirey, Paris, 2^{e} edition, 1958.

[81] John W. Tukey. The future of data analysis. *Annals of mathematical statistics*, 33:1–67, 1962.

[82] Paul A. Tukey and John W. Tukey. Preparation; prechosen sequences of views. In Vic Barnett, editor, *Interpreting multivariate data*, pages 189–213. John Wiley & sons, 1981.

[83] Jean-Paul Valois. Approche graphique en analyse des données (avec discussions). *Journal de la Société Française de statistique*, 141(4):5–107, 2000.

[84] Peter van der Heijden. End-member analysis and latent budget analysis (letter to the editor). *Applied Statistics, Journal of the Royal Statistical Society (Series C)*, 43(3):527–528, 1994.

[85] Michel Volle. L'analyse des données. *Économie et Statistique*, 96:3–23, 1978.

[86] Michel Volle. *L'analyse des données*. Economica, Paris, 1978.

[87] Henry Vuibert. *Les anaglyphes géométriques*. Librairie Vuibert, Paris, 2e édition edition, 1912.

[88] Joe Whittaker. *Graphical models in applied multivariate Statistics*. John Wiley & Sons, Chichester, 1990.

[89] Herman Wold. Nonlinear estimation by iterative least squares procedures. In F. N. David, editor, *Research papers in statistics*, pages 411–444, New York, 1966. Wiley.