

AMADO-online, User Guide

<https://paris-timemachine.huma-num.fr/amado/>

English, Français, Español, Italiano, Русский, Український, Tiếng Việt

AMADO-online was written by **Nguyen-Khang Pham** (Can Tho University)
from **Alban Risson**'s original code

This Users Guide was prepared by **Jean-Hugues Chauchat** (Université Lyon 2)
The text was translated into English by **Jean Dumais** Version 22 December, 2020

Introduction. *AMADO-online* was developed as part of the « Paris Time Machine » consortium; *Amado-online* displays and analyses data matrices (binary, counts, responses to Lickert-type items, or measures of heterogenous variables) following the ideas laid out by Jacques BERTIN (Bertin 1967, 1973, 1999 and Bertin 1977).

AMADO-online displays bar charts of data matrices where the areas of the bars are proportional to the values they represent. Highlighting data structures can be achieved through sorting lines or columns either manually or automatically along their coordinates on the first factorial axis, be it from a Correspondence Analysis or a Principal Components Analysis; hierarchical clustering is also available. Depending on the nature of the data and on the type of analysis selected, data series (e.g. time series), exact or approximate data blocks, or homogeneous classes can be obtained. Several formatting options are available, making the displays faithful to the original data and easy to read.

This Users Guide details the various commands through worked-out examples. Source data and the sequence of commands used to derive the graphs are given herein; the final graphs produced by *AMADO-online* can be exported as PNG or SVG files.

AMADO-online menus are currently offered in seven languages: English, French, Spanish, Italian, Russian, Ukrainian and Vietnamese.

AMADO-online builds on and completes the earlier work of Chauchat-Risson (1998).

Following the Table of Contents, accessing a file, formatting and exporting a graph are covered step by step in Section 1.

Using several examples, highlighting the structure of binary, frequency, homogeneous and heterogeneous data files are presented in Section 2 to Section 4.

In Section 5, *AMADO-online* is set against the prehistory and history of Bertin graphs.

After Section 6, References, four annexes give additional information on: -7.1- "Exporting graphs in PNG or SVG?", -7.2- "How to resize a SVG image pasted in Word, Excel or PowerPoint?", -7.3- "Correspondence Analysis, an optimal coding of data rows and columns" and -7.4- "How ultrametric distances give a distorted view of item closeness"; Acknowledgements are on Section 8.

Table of contents

1	Starting with <i>AMADO-online</i>	3
1.1	Opening a data file	3
1.2	Basic Commands	4
1.2.1	File / Export	4
1.2.2	Edit / Cancel; Redo; Copy table; Set title; Delete	4
1.2.3	Process / Transpose; Sort.....	5
1.2.4	Format / Row labels ; Column labels; Value format	6
1.2.5	Format / Common scale; Scale by line	6
1.2.6	Six display modes	7
1.2.7	Format / Graph size	8
1.2.8	Typography / Spacing between lines or columns	8
1.2.9	Typography / Colour and Size of labels, values and separators	8
1.2.10	Moving a line or column	8
2	Processing Frequency or Binary data	9
2.1	Example: seriation of archaeological artefacts	9
2.2	Example of classification: Eye and Hair Colour of 592 people.....	10
2.3	Example of Correspondence Analysis: musical instruments played by children and the occupation of their parents	12
2.4	Example of time series: distribution of Jews deported from France by birthplace and convoy 16	
2.5	Example of hierarchical clustering: Employed labour force, age 25 to 54, Paris, 2015, by occupation and arrondissement	17
2.6	Identifying blocs in a square matrix of co-occurrences: marketing territories.	20
3	Processing homogeneous data	23
4	Processing heterogeneous data	26
5	BERTIN graphs	28
6	References	29
7	Appendices	30
7.1	Appendix 1: PNG or SVG format?	30
7.2	Appendix 2: Cropping a PNG or SVG picture file pasted in Word, Excel ou PowerPoint....	31
7.3	Appendix 3: Correspondence Analysis: optimal scoring of rows and columns	32
7.4	Appendix 4: The distorted view from a dendrogram	34
8	Acknowledgements	35

1 Starting with *AMADO-online*

1.1 Opening a data file

Working from a spreadsheet, the first cell of the first line is either empty or contains the table title; the following cells contain the column headers. Each following line starts with a line stub. The data table must not contain line totals or column totals. There are two ways to import data into *AMADO-online*:

“**File / Open**” a TXT file (UNICODE or UTF-8, if special characters) with TAB separators.

Copy a zone of a spreadsheet, then “**Edit/ Paste**”¹ in *AMADO-online*.

The following example is taken from Snee (1974). The file can be loaded by requesting “**File / Examples/ EN-Hair-colour_Eye-color.TXT**” or copy from the above table and past to *AMADO-online*.

Snee 1974 The Am. Statistician Volume 28	Black Hair	Brunette Hair	Red Hair	Blond Hair
Brown Eye	68	119	26	7
Blue Eye	20	84	17	94
Hazel Eye	15	54	14	10
Green Eye	5	29	14	16

As soon as the data are loaded, *AMADO-online* displays this graph:

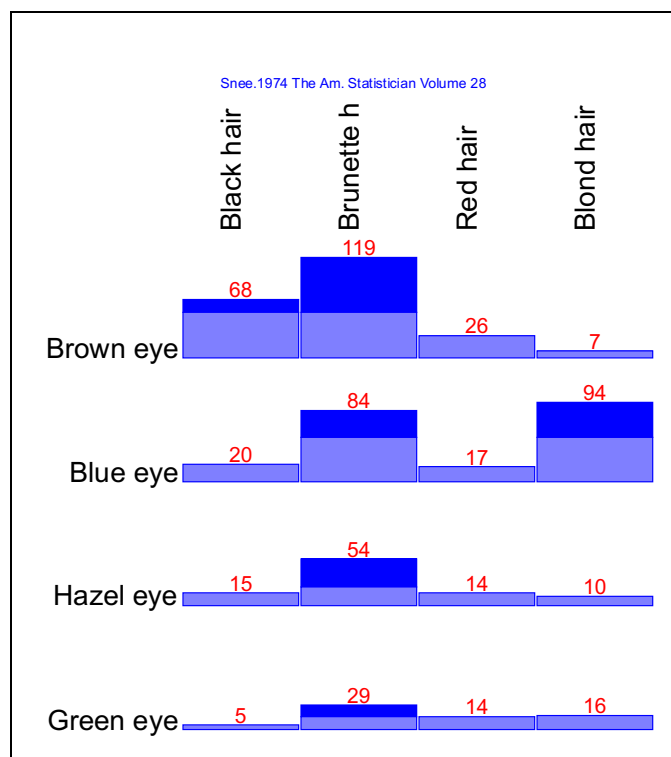




Table values are represented by the bar heights (and areas). The average value of each line splits each pair in light-coloured and dark-coloured zones (see “**Format / Mode 6**” on § 1.2.6).

¹ Under Windows OS, if you use the Firefox browser, you can add the  editing tools to your personal menu bar from the top right corner menu choices in .

1.2 Basic Commands

1.2.1 File / Export

The commands “**File / Export to SVG**” and “**File / Export to PNG**” copy the displayed graph to the « Downloads » folder. Differences between the two file formats are given in Appendix 7.1.

Appendix 7.2 to see how to resize a *Scalable Vector Graphic* (SVG) or *Portable Network Graphics* (PNG) pasted in *Word*, *Excel* or *PowerPoint*.

1.2.2 Edit / Cancel; Redo; Copy table; Set title; Delete

The familiar desktop publishing software editing commands are also available in *AMADO-online*: “**Edit / Cancel**” (Ctrl Z – Cmd Z) cancels the last command and “**Edit / Redo**” (Ctrl Y – Cmd Y) resets the last cancelled command.

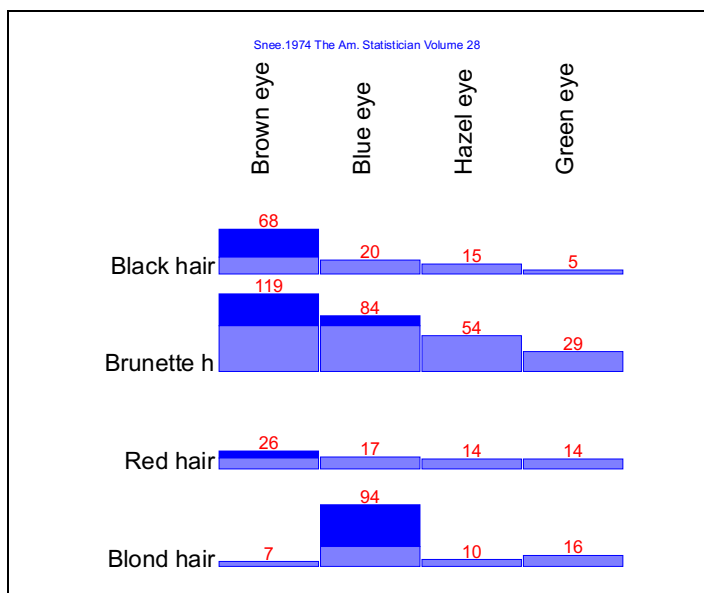
The command “**Edit / Copy table**” creates a copy the current data table which can then be pasted in a TXT, Excel or Word document.

To create, modify or delete the title of a graph, one uses “**Edit / Set title**”.

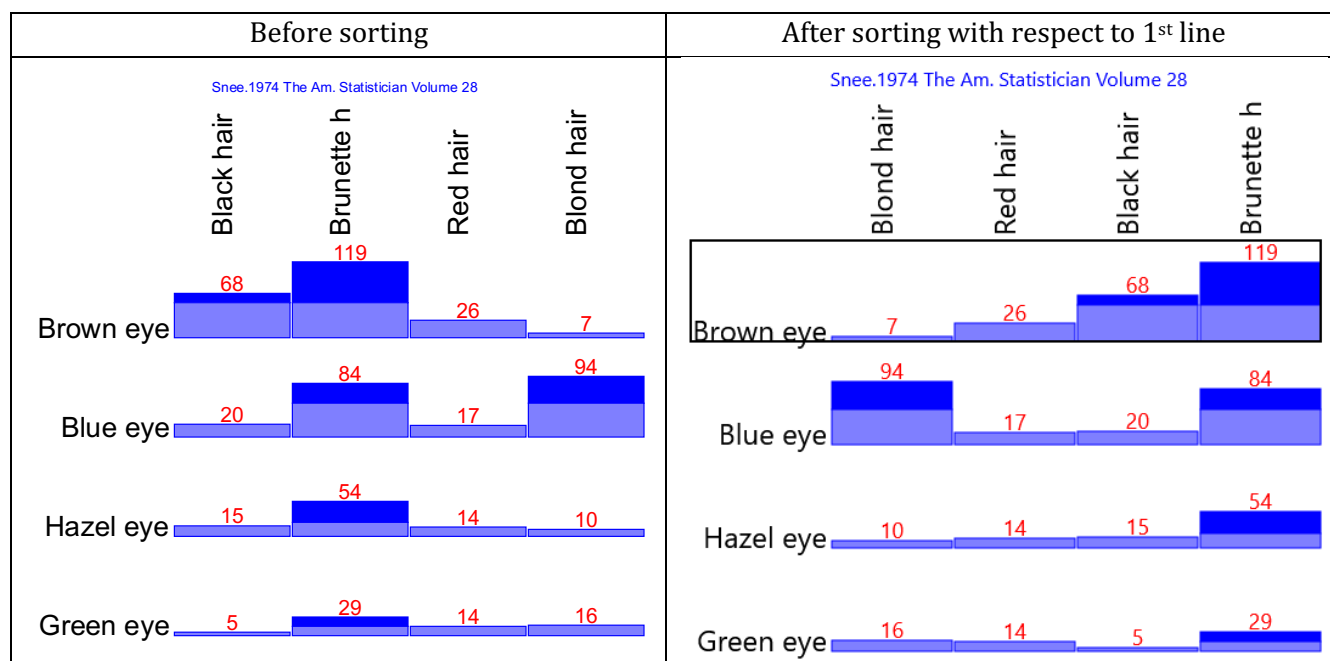
To delete a line or a column, first select it by clicking its name tag, then clicking the command “**Edit / Delete**”. As usual, several lines or columns can be deleted together if they are selected with “**Ctrl + Click**”.

1.2.3 Process / Transpose; Sort

The command “**Process / Transpose**” is used to transpose a graph.



AMADO-online can sort the values of a data table. First a line (respectively, a column) is selected by clicking its name. Then the command “**Process / Sort A to Z**” sorts the columns (respectively, the lines) with respect to the increasing values of the selected line (respectively, column).



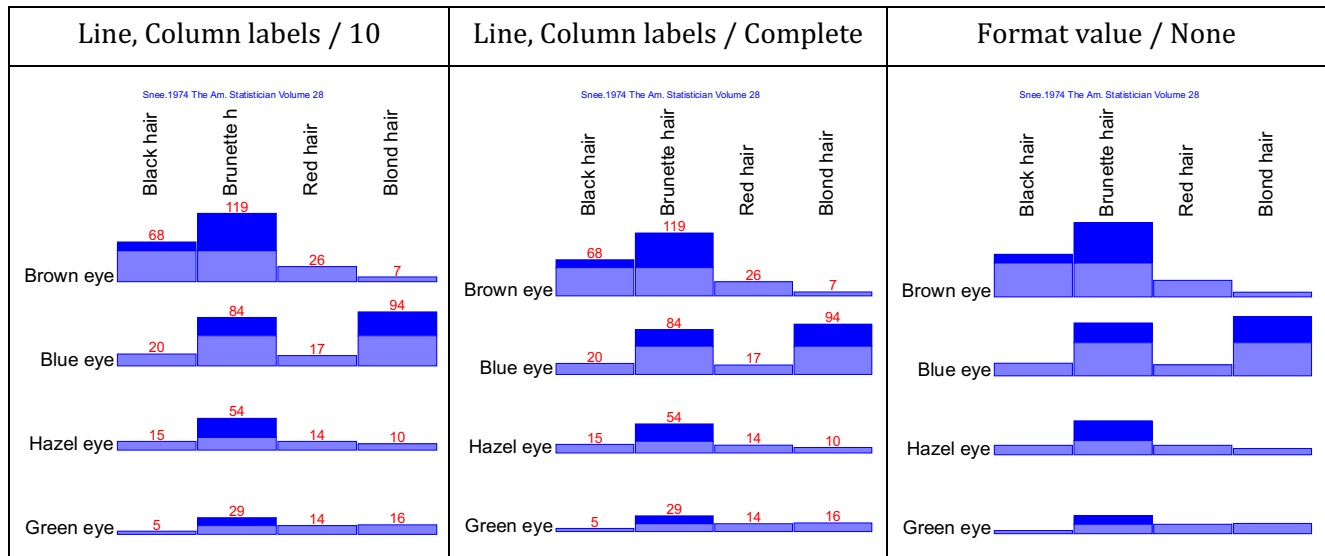
If desired, “**Process / Sort Z to A**” is also possible.

How to automatically sort lines and columns to highlight diagonal or block structures is detailed in Sections 2, 3 and 4.

1.2.4 Format / Row labels ; Column labels; Value format

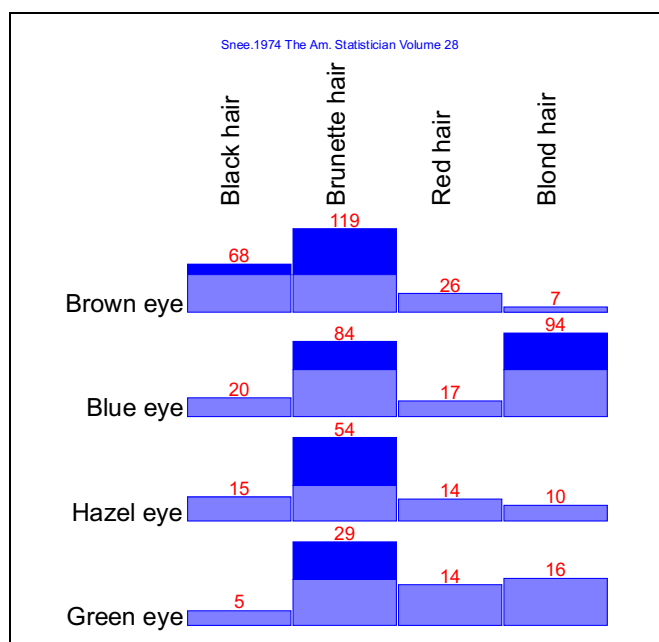
By default, labels are 10 characters long. Line stubs and column headers can be displayed in full using “**Format / Row labels / Complete**” or “**Format / Column labels / Complete**” respectively.

Values can be hidden by requesting “**Format / Value format / None**”; alternatively, numerical values can be displayed in one of the following formats: 0 ; 0.0 ; 0.00 ; 0.000 or 0% ; 0.0% ; 0.00% ; 0.000%.



1.2.5 Format / Common scale; Scale by line

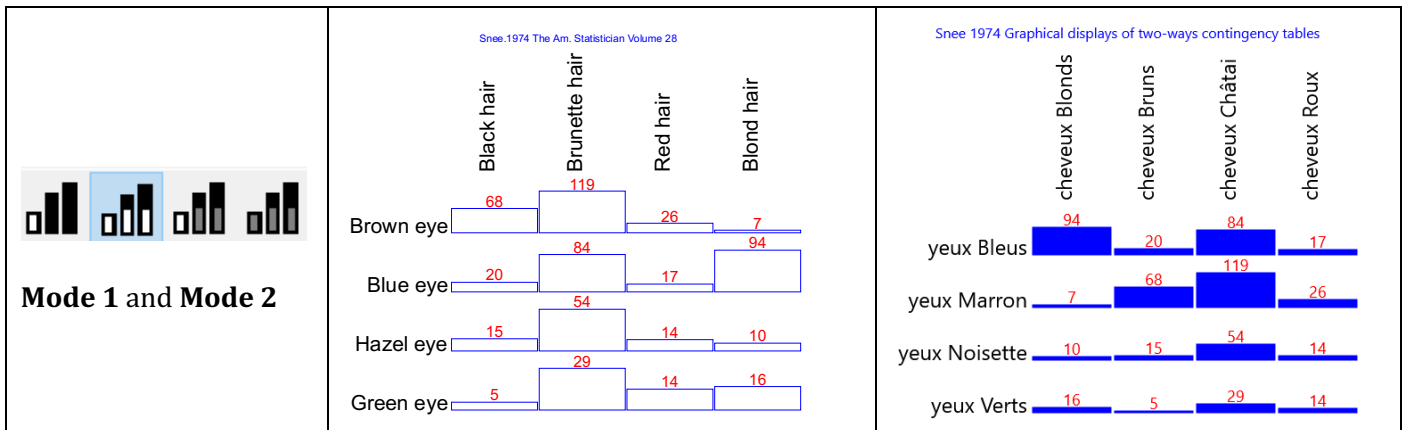
In the graphs shown above, all lines are displayed on a common scale making the area of the bars proportional to the values as a proportion of the table total. When very small values appear in a given line, the command “**Format / Scale by line**” makes the line maximums equal in height. Then the remaining bars of a line become proportional to the line total, tantamount to line percentages.



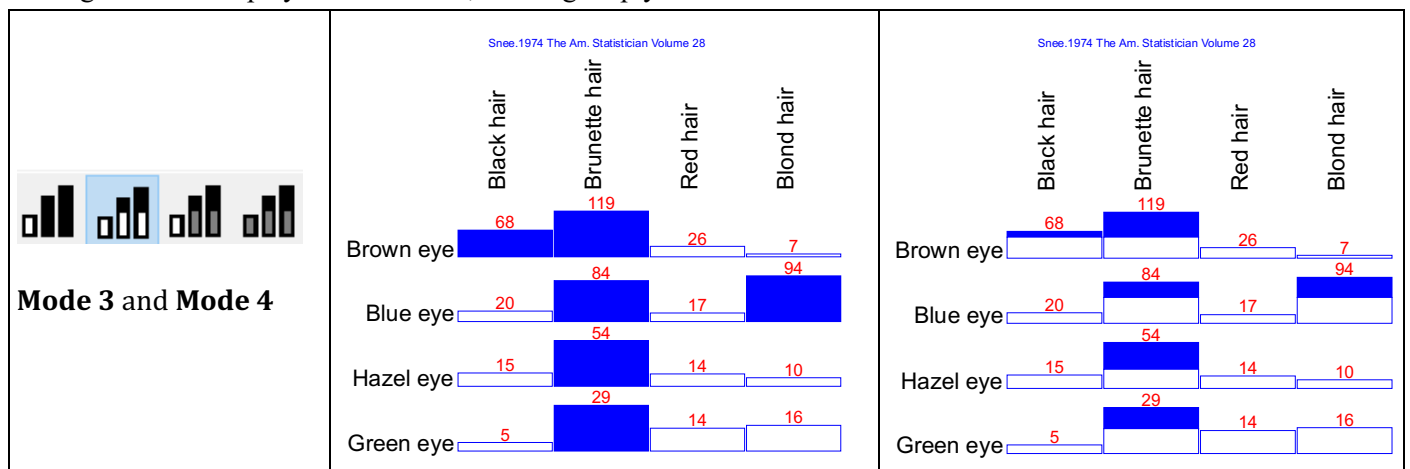
1.2.6 Six display modes

The command “**Format**” can be used to modify how histogram bars are displayed. There are six possible display modes.

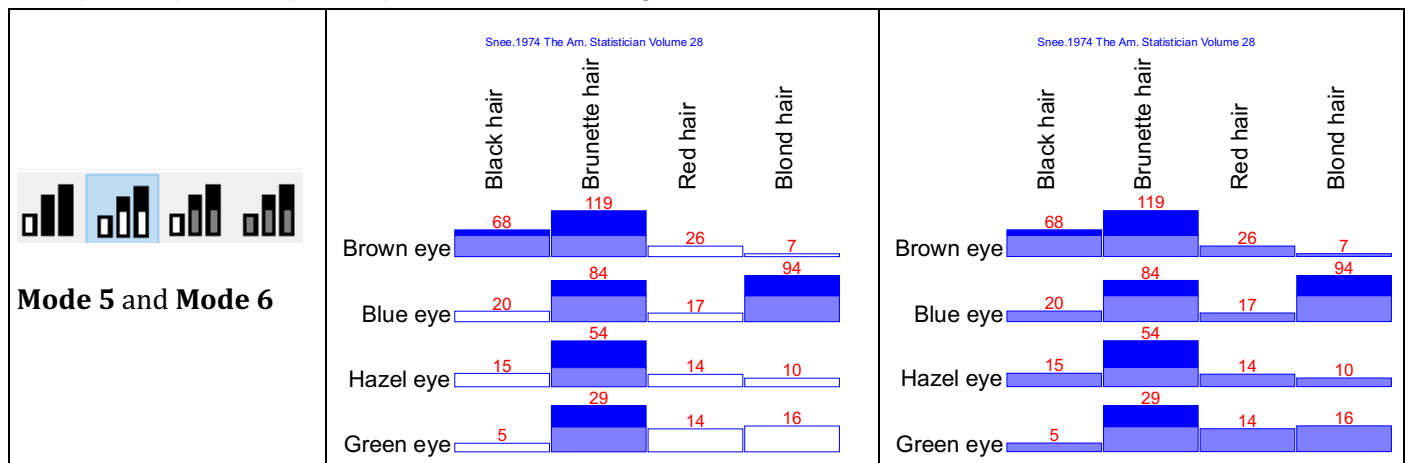
The options “**Format / Mode 1**” and “**Format / Mode 2**” display only the bar outline (“**Mode 1**”) or in solid colour (“**Mode 2**”).



Under the optional “**Format / Mode 3**” and “**Format / Mode 4**” it is possible to distinguish which values are less than the line average (bar outline) or above the line average (solid colour). Under “**Mode 4**”, the average value is displayed on each bar, creating empty and solid zones.



Finally, “**Mode 5**” and “**Mode 6**” add to “**Mode 4**” by bringing a lighter shade of colour to the portions of bars (**Mode 5**) or bars (**Mode 6**) below the line average.

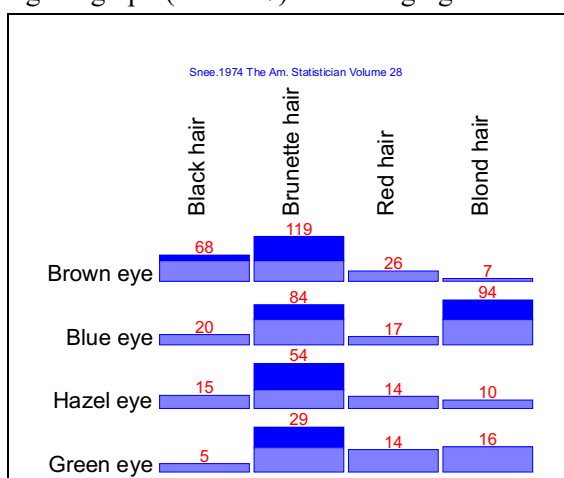


1.2.7 Format / Graph size

The width and height of a graph are set by default by *AMADO-online*. However, the dimensions can be changed using “**Format / Graph size**”. First, the box “**Auto resize**” must be switched off. Then, the preferred dimensions can be typed in or scrolled up/down. Once the preferred dimensions are reached, click “**OK**” to confirm the choice.

1.2.8 Typography / Spacing between lines or columns

The spacing between columns or lines can be modified by using the commands “**Typography / Increase column spacing**” or “**Typography / Decrease column spacing**”. Similar commands are available to adjust line spacing. For example, resizing the graph (see 1.2.7) and changing column spacing yields this graph:

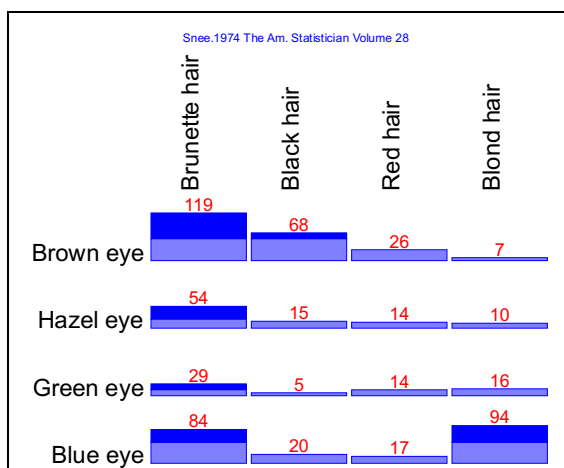


1.2.9 Typography / Colour and Size of labels, values and separators

The colour and font size of both column and row labels and of the values can be adjusted under “**Typography**”. The colour and width of separators (see 2.5) can also be adjusted under the same menu.

1.2.10 Moving a line or column

Individual lines and columns can be moved by hand to highlight data structures; contiguous lines or columns can also be moved in blocks by hand (automatic diagonalization and classification methods are shown later). For example, moving the line “Blue eyes” and the column “Dark hair” (**clicking on label, holding and moving with the mouse**), a diagonal structure is obtained that highlights the correlation between eye and hair colours:



2 Processing Frequency or Binary data

In this section, the processing of frequency tables and tables of binary entries are reviewed. There are two commands of interest: “**Process / Frequency data or 0/1 / Processing with Correspondence Analysis**” and “**Process / Frequency data or 0/1 / Hierarchical Clustering**”.

2.1 Example: seriation of archaeological artefacts

The following example was inspired by data used by archaeologist Sir Flinders Petrie when he dated the graves he excavated in Diospolis Parva, Egypt, at the end of the XIXth Century. Petrie hypothesised that the type of artefact and their ornamentation were characteristic of their era, and that a chronology could hence be deduced.

In their book, Renfrew and Bahn (1991), give the following pedagogical example. The data are stored in “**File / Examples / EN-Egyptian-pottery.TXT**”. Upon loading, *AMADO-online* displays the crude data matrix of presence (1) or absence (0) of décor elements on potteries A, B, C...

	◀	Ⓜ	○	◻	Ⓜ	Ⓜ	Ⓜ
beaker	0	0	0	0	1	1	0
blackrim	1	0	0	0	1	1	0
bottle	1	0	0	0	1	0	0
flatpot	0	1	0	1	0	0	0
handle	1	0	0	1	0	0	1
pointed	0	1	1	1	0	0	0
spirals	0	0	0	1	0	0	1

The matrix can be diagonalized by moving columns and rows by hand; the same result can be achieved by requesting “**Process / Frequency data or 0/1 / Processing with Correspondence Analysis**”.

	○	Ⓜ	◻	Ⓜ	◀	Ⓜ	Ⓜ
pointed	1	1	1	0	0	0	0
flatpot	0	1	1	0	0	0	0
spirals	0	0	1	1	0	0	0
handle	0	0	1	1	1	0	0
bottle	0	0	0	0	1	1	0
blackrim	0	0	0	0	1	1	1
beaker	0	0	0	0	0	1	1

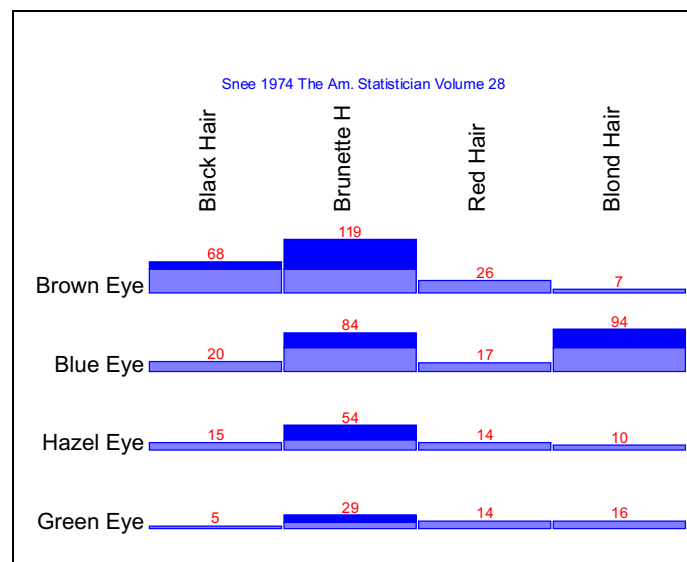
The graph now displays the potteries (C, B, D, G, A, E and F) sorted along a “presence / absence” axis of décor elements as identified by the archaeologist. This new ordering likely corresponds to the chronological order (direct or inverse) of the appearance, then obsolescence, of the artistic elements.

2.2 Example of classification: Eye and Hair Colour of 592 people

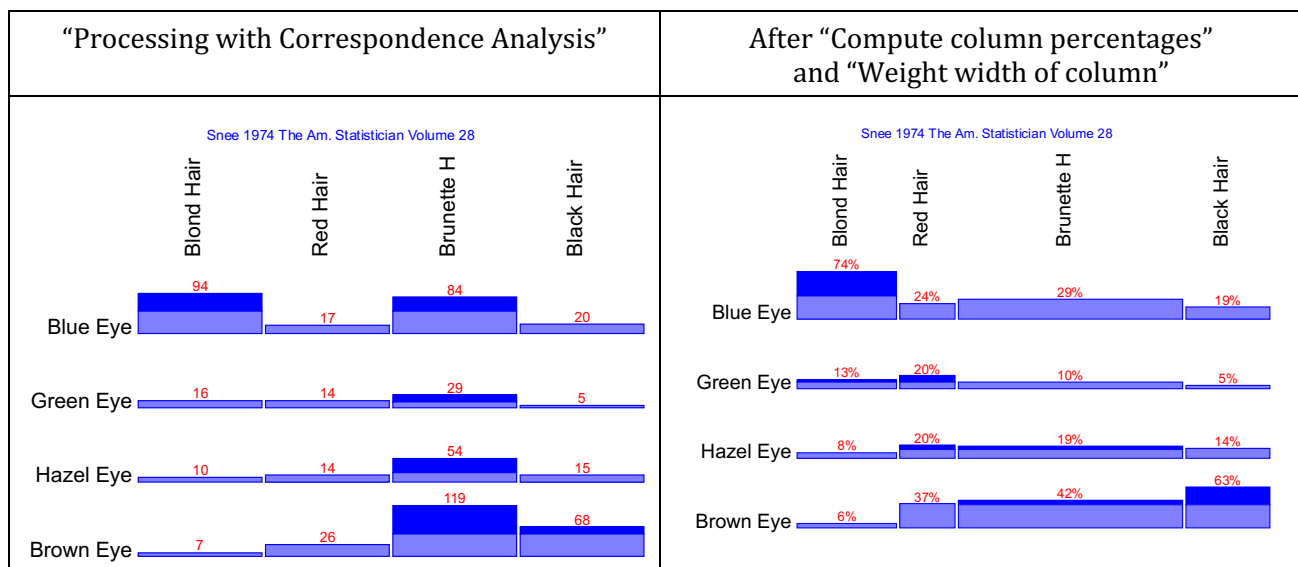
The following data table was taken from Snee's 1974 paper on graphical display of two-way tables.

Snee 1974 The Am. Statistician Volume 28	Black Hair	Brunette Hair	Red Hair	Blond Hair
Brown Eye	68	119	26	7
Blue Eye	20	84	17	94
Hazel Eye	15	54	14	10
Green Eye	5	29	14	16

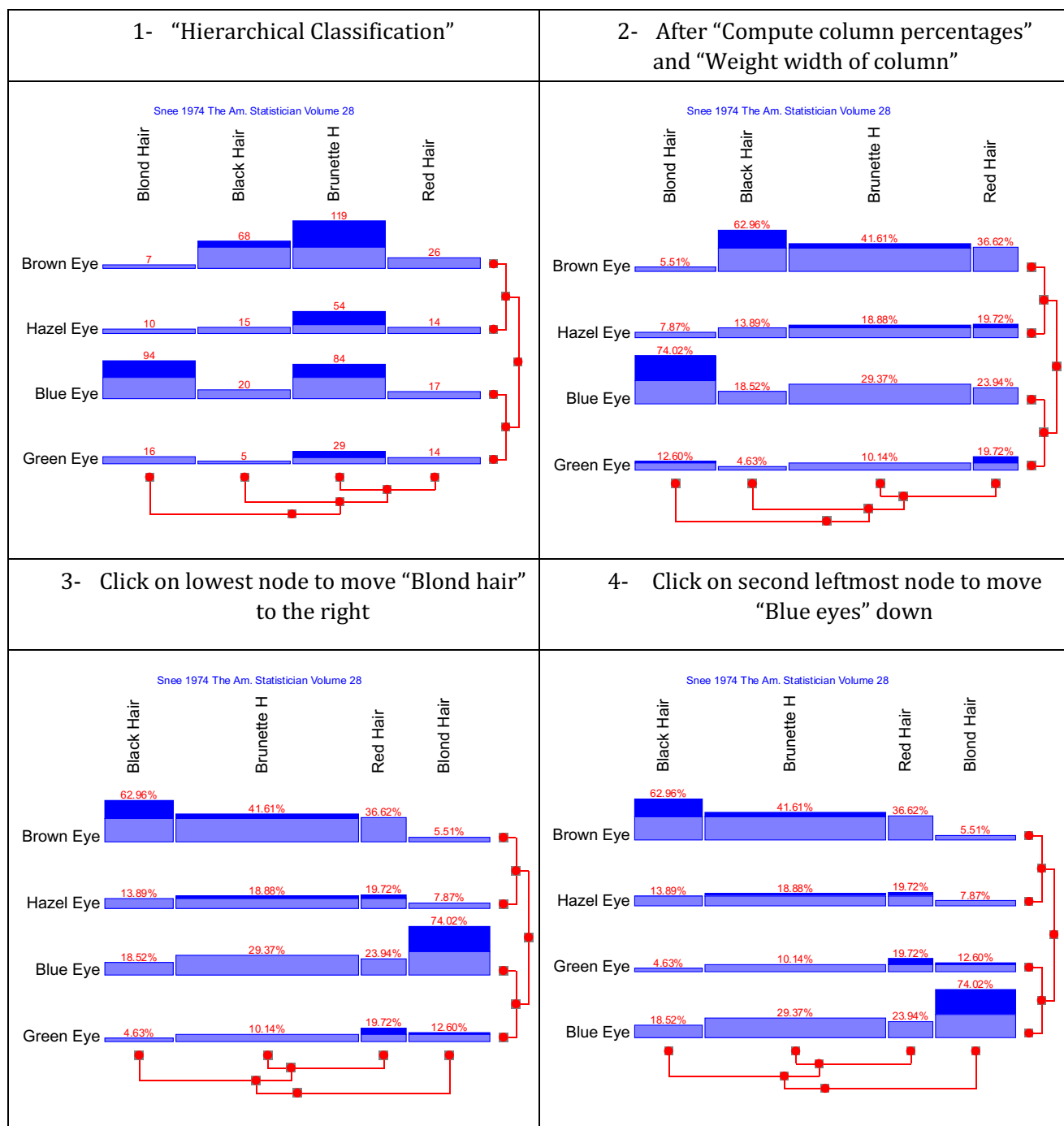
The data can be imported into *AMADO-online* by selecting and copying the table (Ctrl+C or Cmd+C), switching to *AMADO-online* where it is pasted (Ctrl+V or Cmd+V). Alternatively, the command “**File / Examples/ Eye Colour – Hair Color.TXT**” can be entered. *AMADO-online*, then shows the following graph:



The request “**Process / Frequency data or 0/1 / Processing with Correspondence Analysis**” yields a diagonalization of the data matrix as shown in the left-hand side graph, even more visible after “**Process / Compute column percentages**” and “**Format / Weight width of column**”.



The commands “**Process / Frequency data or 0/1 / Hierarchical Classification**” highlights the closeness of lines and columns as shown below:



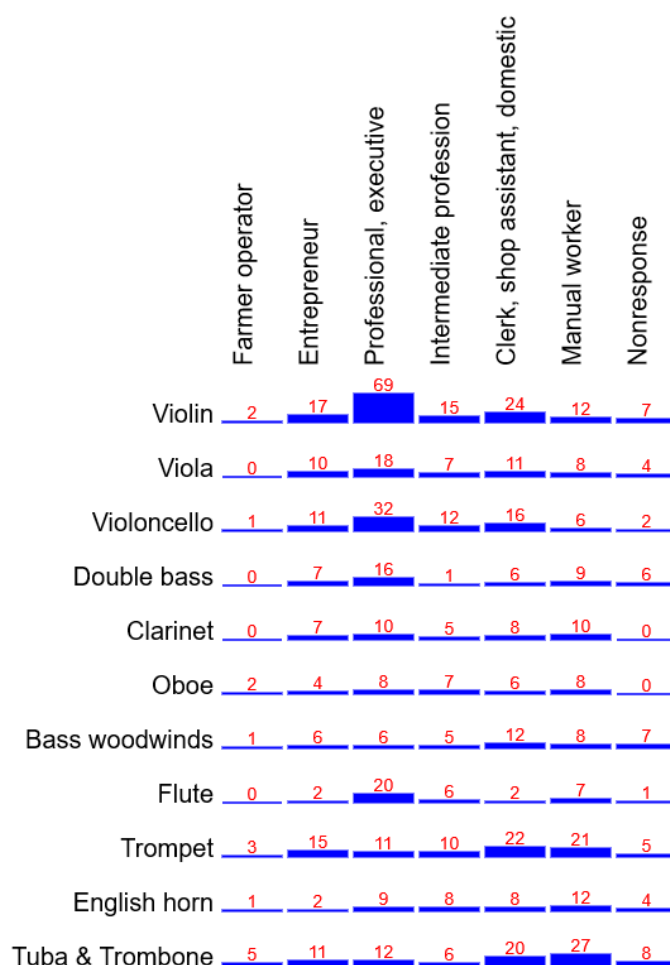
2.3 Example of Correspondence Analysis: musical instruments played by children and the occupation of their parents

The data used in this example are taken from Lehman's 1995 dissertation *L'orchestre dans tous ses éclats : sociologie de la profession de musicien*, (EHESS, Paris). The data were collected on students of the Conservatoire National de Musique et de Danse de Paris (CNMD). Data can be loaded with "File / Examples / EN-musical_instrument-parent_occupation.TXT".

Lehmann 1995. Student's musical instruments & parent's occupation	Farmer operator	Craftsman, shopkeeper, entrepreneur	Professional, executive, higher intellectual profession	Intermediate profession	Clerk, shop assistant, domestic worker	Manual worker	Non- response
Violin	2	17	69	15	24	12	7
Viola	0	10	18	7	11	8	4
Violoncello	1	11	32	12	16	6	2
Double bass	0	7	16	1	6	9	6
Clarinet	0	7	10	5	8	10	0
Oboe	2	4	8	7	6	8	0
Bass woodwinds	1	6	6	5	12	8	7
Flute	0	2	20	6	2	7	1
Trumpet	3	15	11	10	22	21	5
English horn	1	2	9	8	8	12	4
Tuba & Trombone	5	11	12	6	20	27	8

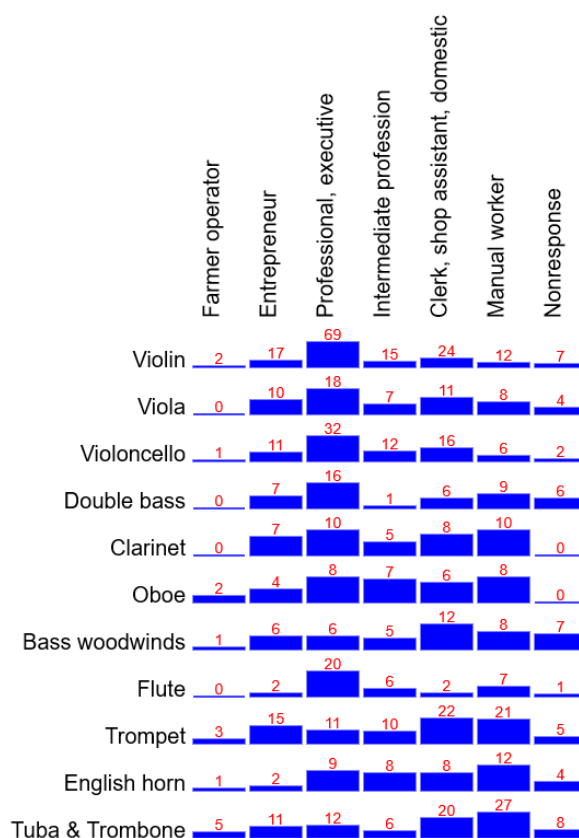
AMADO-online displays this graph upon data loading

Lehmann 1995. Student's musical instruments & parent's profession

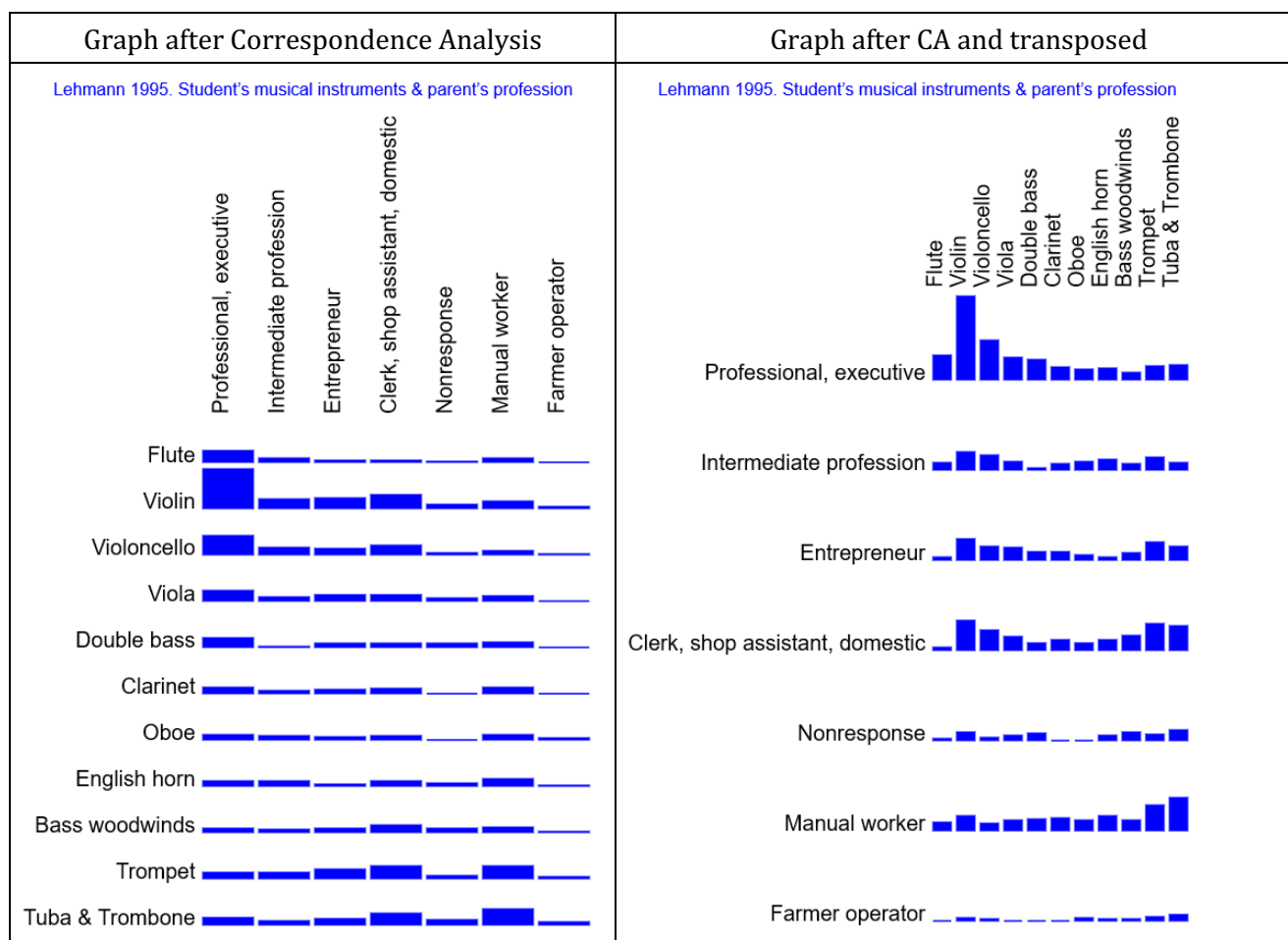


Graph of line-wise relative frequencies.
Since cell counts vary greatly,
distributions emerge after
“Format / Scale by row”

Lehmann 1995. Student's musical instruments & parent's profession



Now, starting afresh from the raw data, the commands “**Process / Frequency data or 0/1 / Processing with Correspondence Analysis**” and “**Process / Transpose**” produce the following two graphs:



Correspondence Analysis automatically reorganises the data, permuting rows and columns according to their coordinates on the first factor (left-hand side graph):

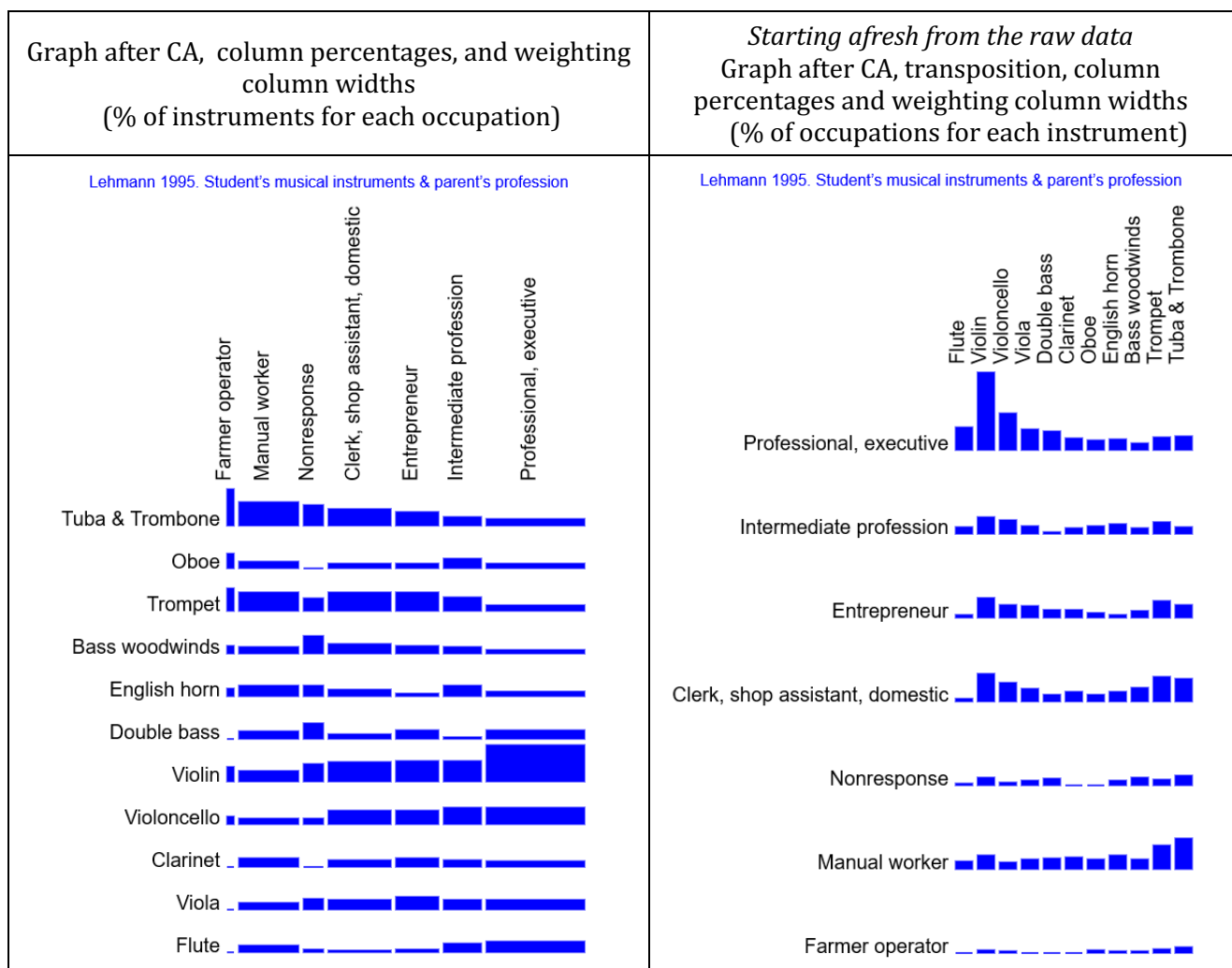
- Lines corresponding to instruments played by children whose parents have similar occupations are brought closer together, and
- Columns corresponding to the occupation of parents whose children often play the same instruments are brought closer together.

Transposition achieves a clearer display (right-hand side graph above).

Finally, one requests “**Process / Compute Column Percentages**” followed by “**Format / Weight Width of Column**” to reshape the bars so that they represent again the original numbers.

Then, starting afresh from the raw data, “**Process / Transpose**”, “**Process / Frequency data or 0/1 / Processing with Correspondence Analysis**”, “**Process / Compute Column Percentages**” followed by “**Format / Weight Width of Column**”.

Depending on the selected stream of commands, either representation displayed below can be achieved by *AMADO-online*:



Each group of Conservatoire students is represented by a bar of equal surface on the left-hand side and right-hand side graphs.

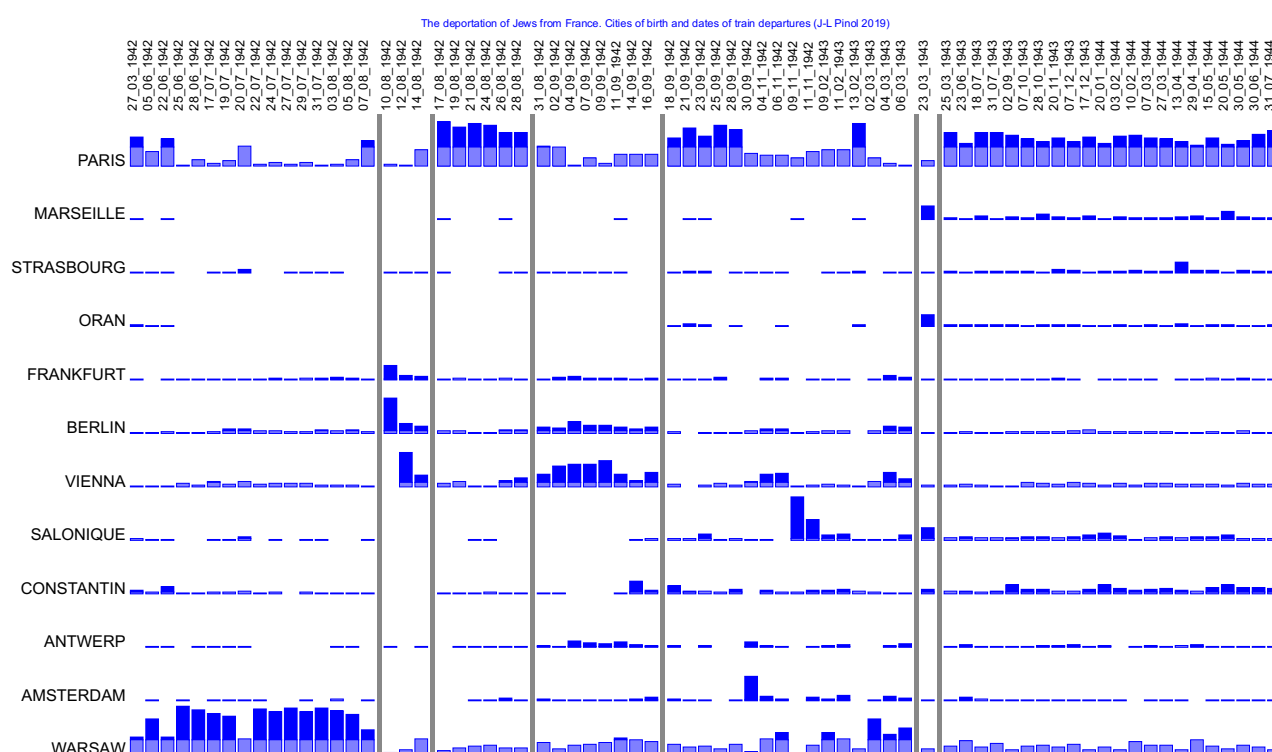
These graphs clearly show that children from more affluent and educated categories tend to take up string instruments and flute; at the opposite end, children from farm workers, skilled and unskilled labour tend to take up wind and brass instruments (those are preeminent in school bands where music education often began).

2.4 Example of time series: distribution of Jews deported from France by birthplace and convoy

In this example (Pinol, 2019), the chronological order is not modified; the graph displayed by *AMADO-online* merely shows the data and supports the historian's commentary.

Data can be loaded as “File / Examples / EN-Deportation_of_Jews_from_France.TXT”.

The graph shows the distribution of people born in a representative sub-sample of cities, deported to death camps, by trains leaving France (column %). Column percentages are obtained by requesting “Process / Compute Column Percentages”.



The graph above clearly shows that the 19 first deportation convoys, up to 7 August 1942, were mostly comprised of Jews born in Warsaw, having fled to France to escape the Nazi regime, then Jews from Germany and Austria on the 10, 12 and 14 August 1942.

The next 6 convoys, from 17 to 28 August 1942, deported Jews born in Paris and arrested during the “*Vel d’Hiv*” raid.

Jews born in Germany, Austria and Belgium, often Jews who had found refuge in the *unoccupied zone*, were arrested during the 26 August roundup and comprise the majority of those deported between 31 August and 16 September 1942.

Then, the police tried to fill up the train convoys to achieve the objectives set for them.

The convoy that left on 23 March 1943 is unique; it is comprised mostly of Jews born in Marseille, in Greece or in Algeria, arrested after the destruction of the old port area (*Le Vieux Port*) of Marseille in January 1943, who were deported that day to the death camps.

2.5 Example of hierarchical clustering: Employed labour force, age 25 to 54, Paris, 2015, by occupation and arrondissement

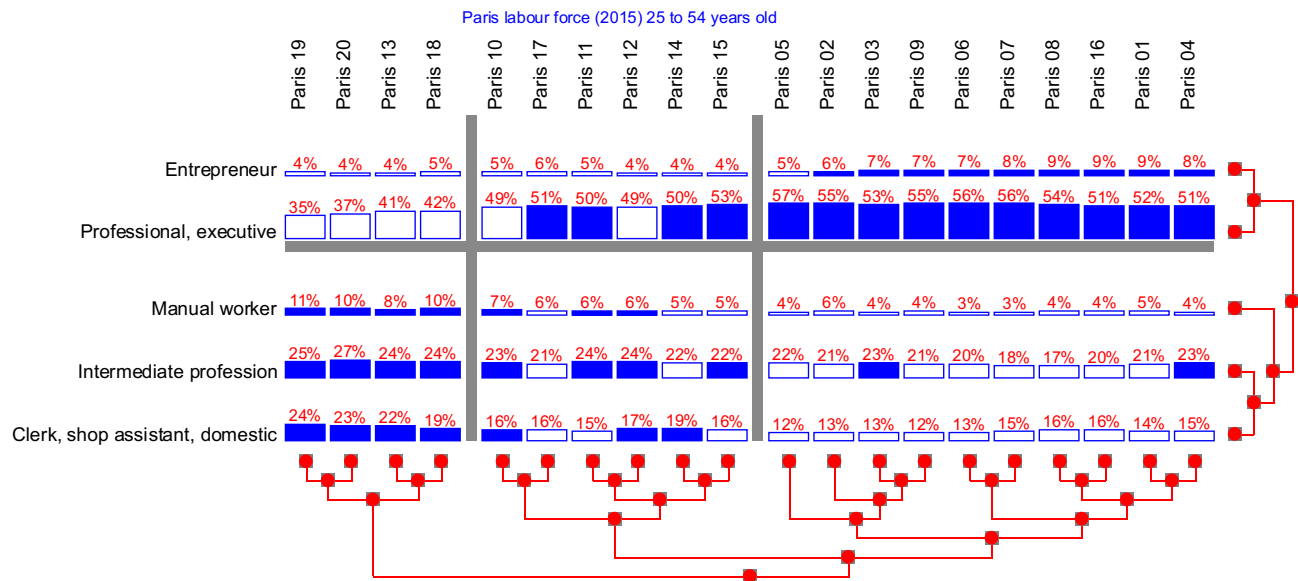
The following table shows the distribution of the employed labour force in Paris, 25 to 54 years old, by main occupation group and arrondissement.

Paris labour force (2015) 25 to 54 years old	Paris 01	Paris 02	Paris 03	Paris 04	Paris 05	Paris 06	Paris 07	Paris 08	Paris 09	Paris 10	Paris 11	Paris 12	Paris 13	Paris 14	Paris 15	Paris 16	Paris 17	Paris 18	Paris 19	Paris 20
Entrepreneur	601	647	1,161	871	1,145	1,062	1,532	1,203	1,880	2,322	3,211	2,172	2,604	2,116	4,066	4,730	4,095	4,299	3,213	3,096
Professional, executive	3,651	5,969	9,005	5,908	12,188	7,971	10,642	7,546	15,821	21,477	35,521	29,285	28,770	26,809	50,880	27,917	37,101	37,529	25,364	30,256
Intermediate profession	1,491	2,239	3,811	2,621	4,699	2,842	3,417	2,461	6,028	9,936	17,426	14,779	16,884	11,685	21,436	10,677	15,304	21,749	18,532	21,563
Clerk, shop assistant, domestic worker	991	1,404	2,206	1,738	2,627	1,813	2,836	2,236	3,534	7,228	10,837	10,556	15,580	9,988	15,601	8,916	11,636	17,266	17,815	18,382
	325	599	713	418	849	487	638	633	1,247	2,937	4,143	3,550	5,503	2,832	4,676	2,382	4,051	8,808	7,803	7,829

File / Open / Browse / EN-Paris2015_Districts-Occupations.TXT

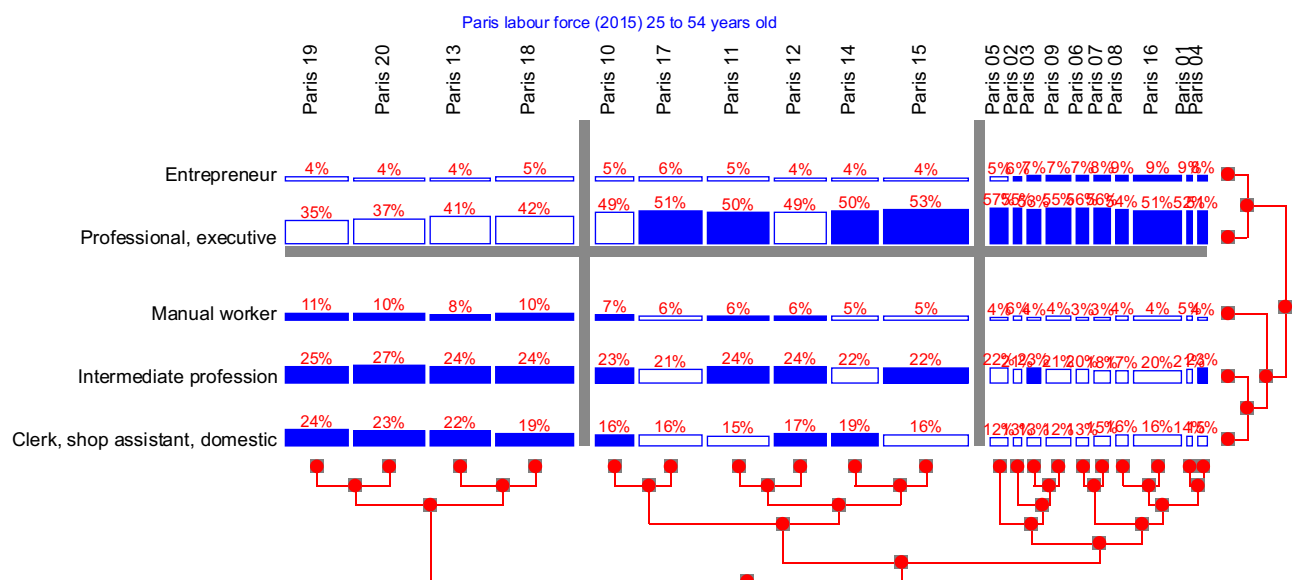
The following sequence of *AMADO-online* commands was used to derive the clustering and graphs shown next page:

- **File / Examples / Paris2015 Arrondissements-pcs_25-54 ans.TXT**
- **Format / Row labels / Complete**
- **Format / Graph Size / (switch off) Auto Resize; Graph width=950 ; Graph height=400 ; OK**
(Note : the graph width is a function of your screen width; 950 might be too narrow for your monitor)
- **Format / Mode 3**
- **Process / Frequency Data or 0/1 / Hierarchical Clustering**
- **Process / Compute Column Percentages**
- **Typography / Increase column spacing** (twice or more, to make all values distinct and legible)
- Click on line “*Manual workers*”, then **Process / Insert separator** (insertion appears above the selected line)
- Click on column “*Paris_10*”, then **Process / Insert separator** (insertion appears before the selected column)
- Click on column “*Paris_05*”, then **Process / Insert separator**



On the resulting graph above, the split between the eastern arrondissements (19°, 20°, 13° et 18°) where relatively more unskilled and skilled labour and middle managers are found, the central arrondissements (5°, 2°, 3°, 9°, 6°, 16°, 1° et 4°) where upper managers and executives live, and sociologically mid-ground arrondissements (10°, 17°, 11°, 12°, 14° et 15°). Here the values and bar heights represent column percentages, that is the distribution of the employed labour force in each arrondissement.

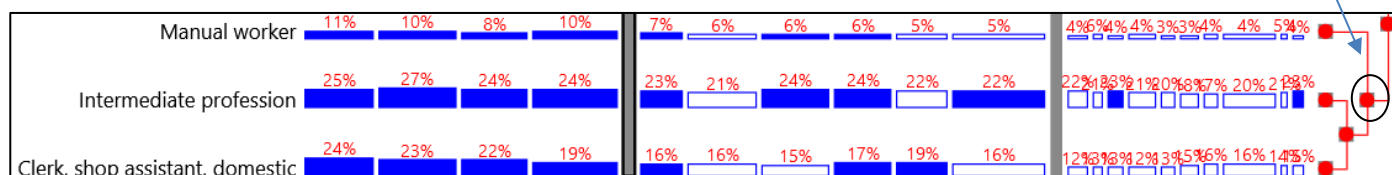
By requesting “**Format / Weight Width of column**” and “**Format / Value format / 0%**”, the graph conveys even more information: the heights of the bars remain proportional to the percent distribution of occupations in each arrondissement, and the area of each bar becomes proportional to the size of the corresponding sub-population.



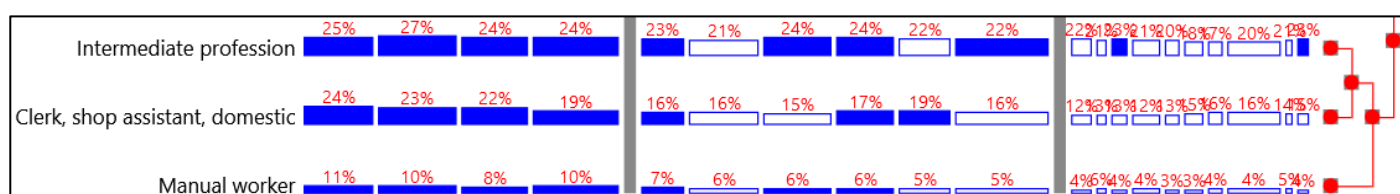
Using “**Format / Mode 3**” displays clearly where each occupation group mostly live. For example, in Paris_19, there are 3213 entrepreneurs, more than in Paris_09 where there are 1880. But, proportionally, there are 4% in Paris-19, less than the 7% in Paris-09.

Since, at each node of the clustering tree, the order of the two classes is arbitrary, this order can be reversed by simply clicking on the small red square representing the node.

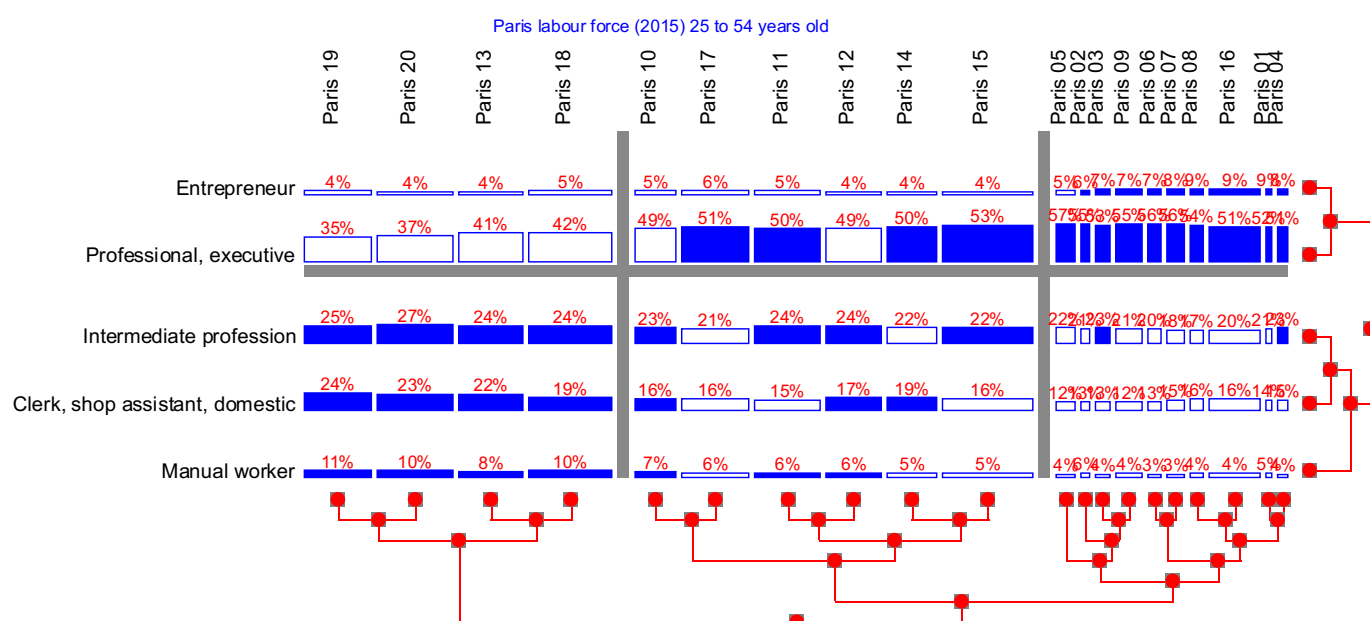
Click on this node



Below, the line "Manual worker" is moved to the very bottom of the graph, as opposed to "Entrepreneur".



Finally, we get:



2.6 Identifying blocs in a square matrix of co-occurrences: marketing territories.

In order to attract businesses, many cities brand and market their respective areas of industrial activity. These areas vary greatly in names (industrial zone, techno-park, ...).

With the aim of reducing the confusing variety of names, 72 business executives were asked to create groups of synonymous names from a set of 49. The number and size of the groups were left unconstrained. Each respondent was allowed to omit any name with which they felt was unknown to them. The results were gathered in a data matrix where an entry is the number of times line “name” and column “name” were in the same group (Texier, 1999). It can be seen as a similarity or proximity matrix.

“File / Exemples / EN-Territorial_Marketing.TXT”²

“Format / Value / None”; “Format / Row labels / Complete”; “Format / Column labels / 20”;

“Format / Graph size / switch off ‘auto resize’, W=1438, H=1278, OK”³

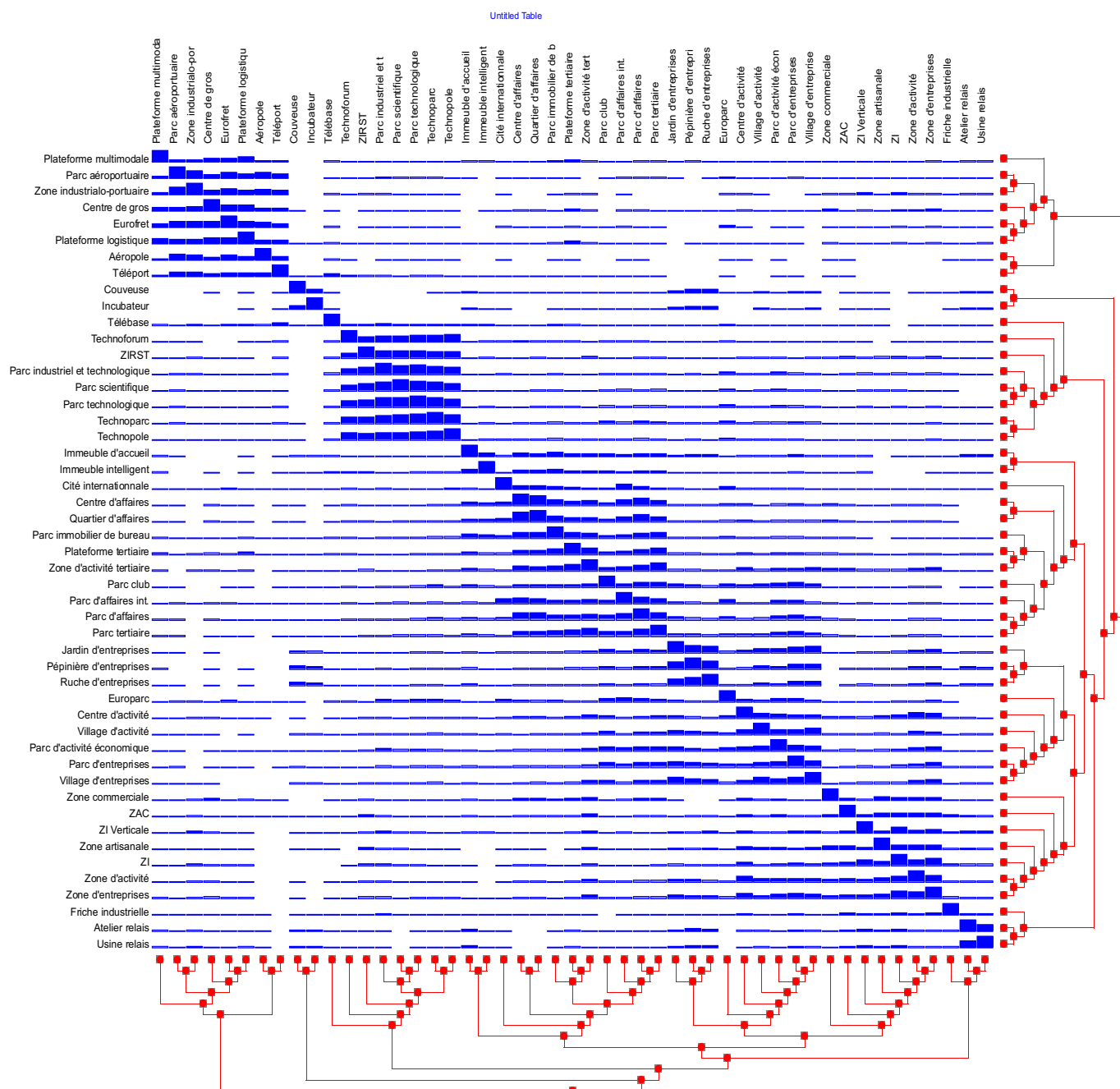


Hierarchical clustering available in *AMADO-online* is used to identify the blocs of names considered synonymous:

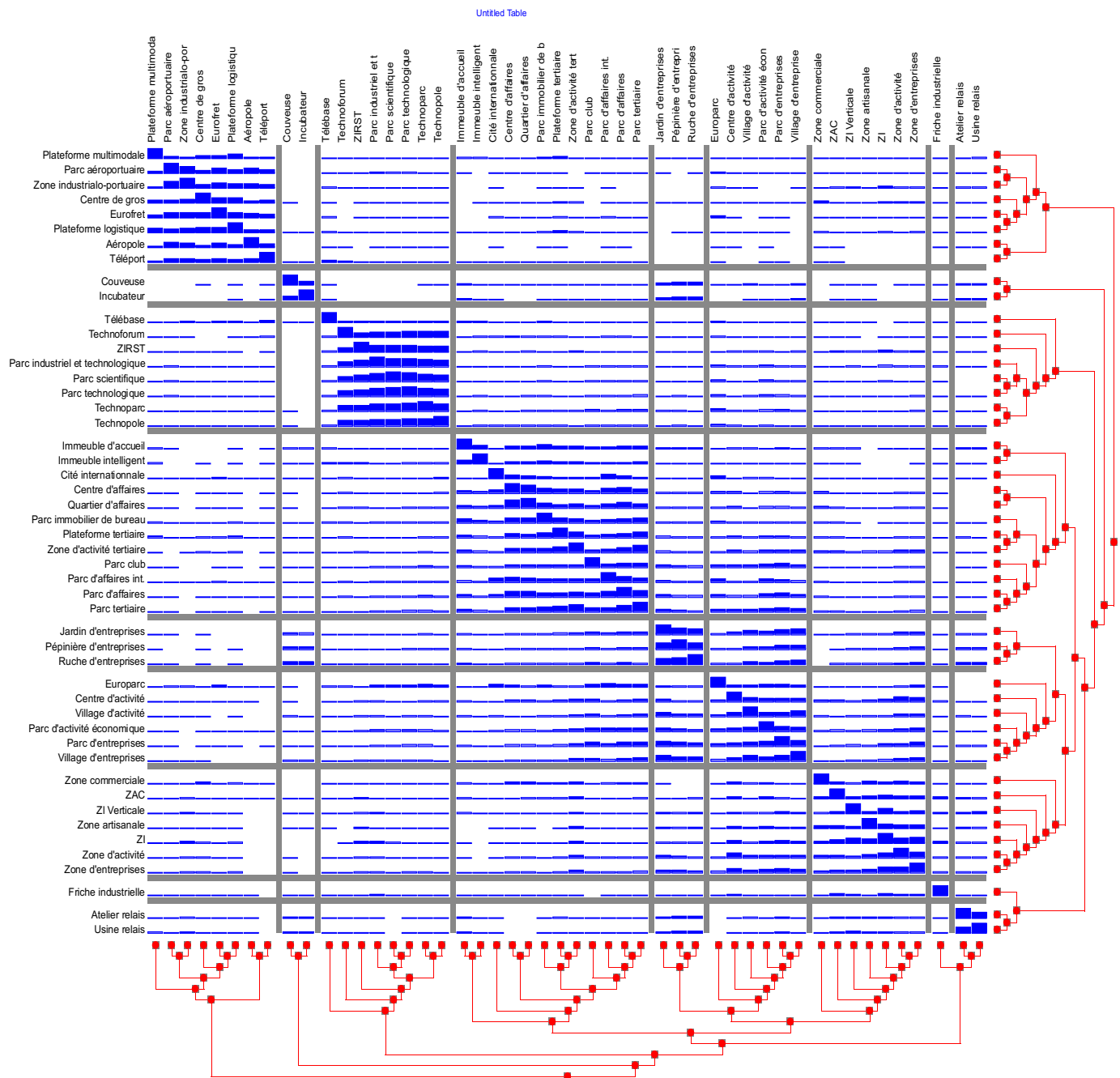
² Given the size of the monitor, it might be useful or necessary to do CTRL- (or CMD-) to decrease font size and display the complete graph

³ Dimensions depend on monitor used

“Process / Frequency Data or 0/1 / Hierarchical clustering”



Inserting separators improves how visible classes are. After clicking a line or column, “**Process / Insert separator**” inserts a separator just before the item selected.



The graph above displays which names are quasi synonymous for business executives:

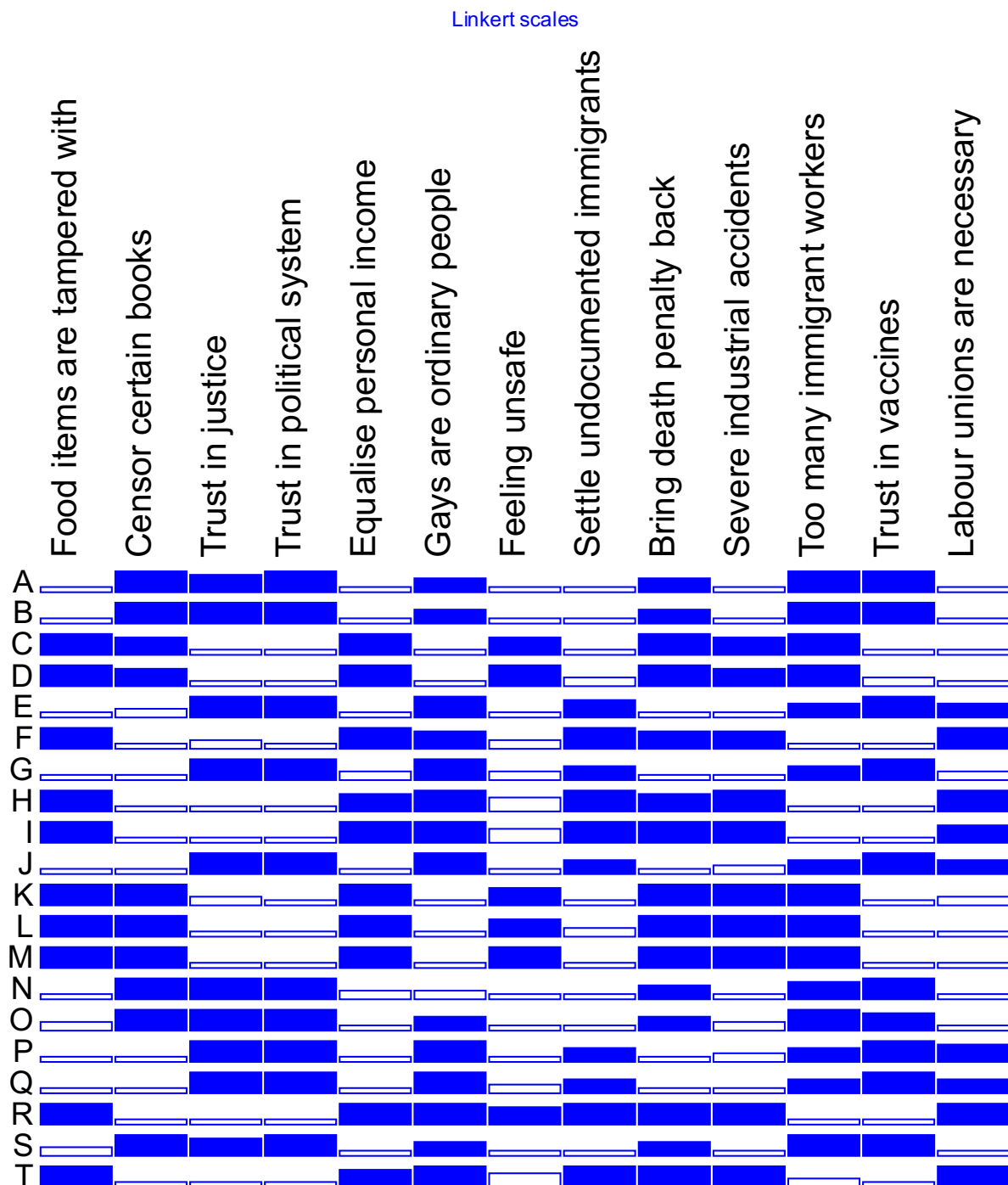
- *Plateforme multimodale*, *Parc aéroportuaire*, *Zone industrialo-portuaire*, *Centre de gros*, *Eurofret*, *Plateforme logistique* are grouped together, then *Aéroport* and *Téléport* seem to only correspond to one another.
- *Couveuse* and *Incubateur* naturally regroup.
- *Technoforum*, *ZIRST* (scientific and technical research and innovation zone), *Parc industriel et technologique*, *Parc scientifique*, *Parc technologique*, *Technoparc*, *Technopole* for a separate group. It can be noted that *Parc Scientifique* and *Parc technologique* are nearly identical in the mind of the executives.
- Another group is comprised of *Jardin d'entreprises*, *Pépinière d'entreprises* and *Ruche d'entreprises*
- And finally, *Atelier relais* and *Usine Relais* form the last group.

A territory marketing strategy could use a reduced set of site name or labels, possible one per group, recognising that other localities might very well use one of the synonyms.

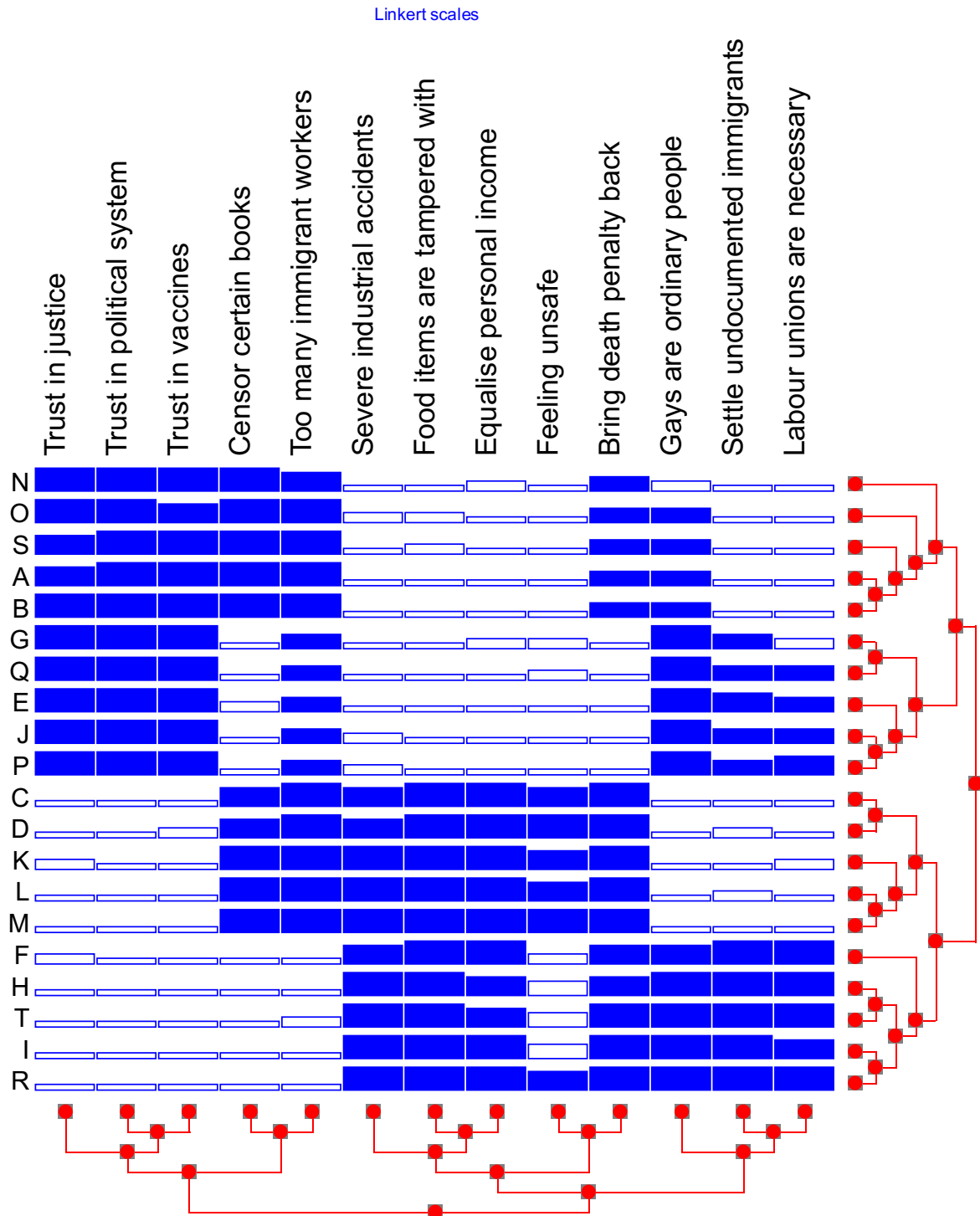
3 Processing homogeneous data

Homogeneous data matrices are those whose lines (say) are repeated measurements (e.g. survey respondents) and columns are variables or items measured on a unique scale, for example, a Lickert-type scale (1= totally disagree; 2= disagree; ... 5=totally agree), unit prices or temperatures.

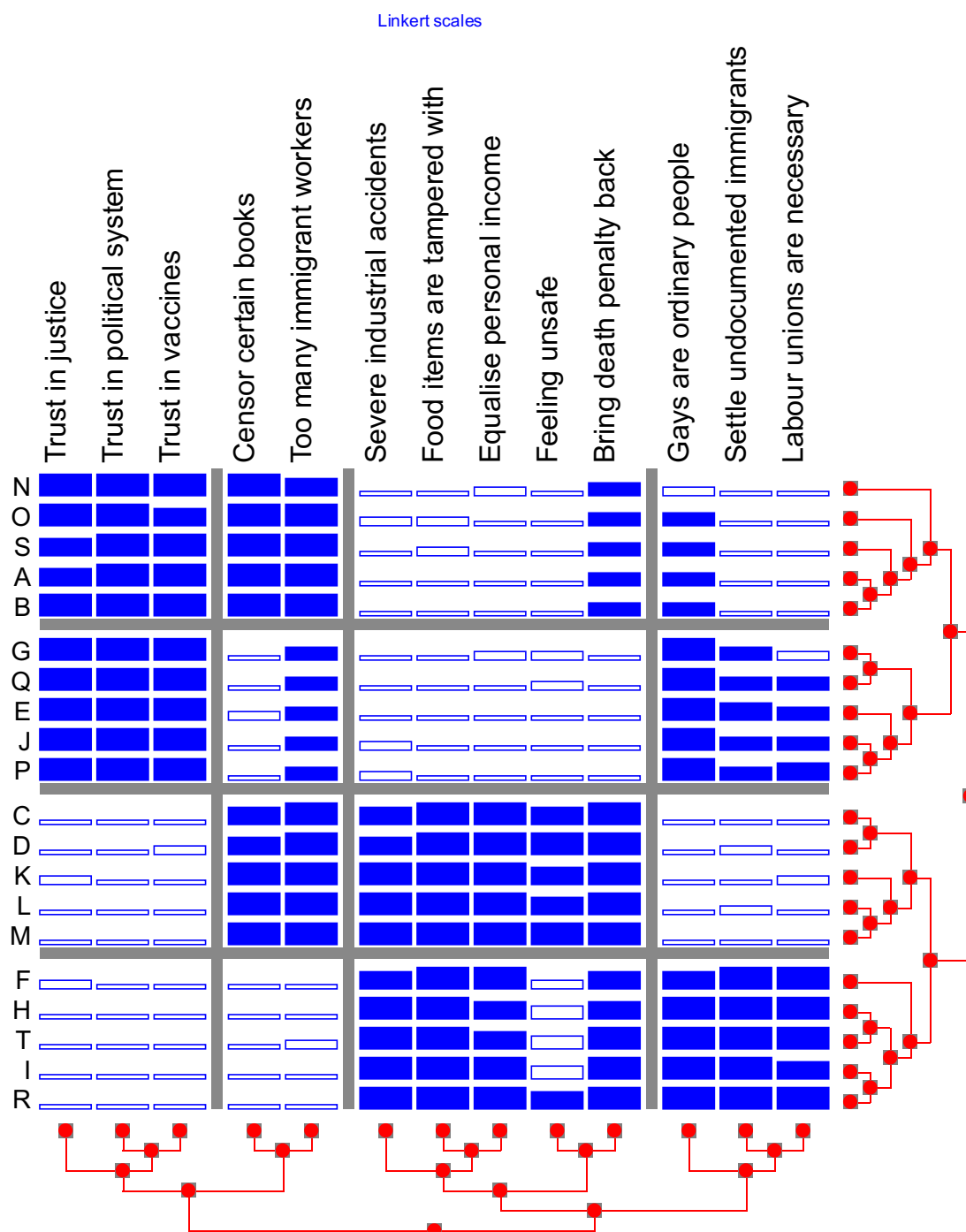
Open “**File / Examples / EN-Linkert_Scale_Survey.TXT**” then request “**Format / Format values / None**”, “**Format / Column label / Complete**”, “**Typography / Column Legend / 20**”, and “**Format / Mode 3**”, this yields this representation:



Now, after “**Process / Homogeneous data / Hierarchical clustering**” row-items and column-individuals are automatically reordered:



Then, after inserting separators, this graph can be obtained:



The graph above highlights 4 (archetypical) opinion groups, each group being characterised as much by the ideas it approves (high dark blue rectangles), as by those it disagrees with (low white rectangles on the graph):

- conservatives,
- liberals,
- far right,
- far left.

4 Processing heterogeneous data

The options offered under this menu item are best suited to process tables whose columns are variables measured in different units or measured on different scales.

File / Samples / EN-Cars_2004_Tenenhaus.TXT

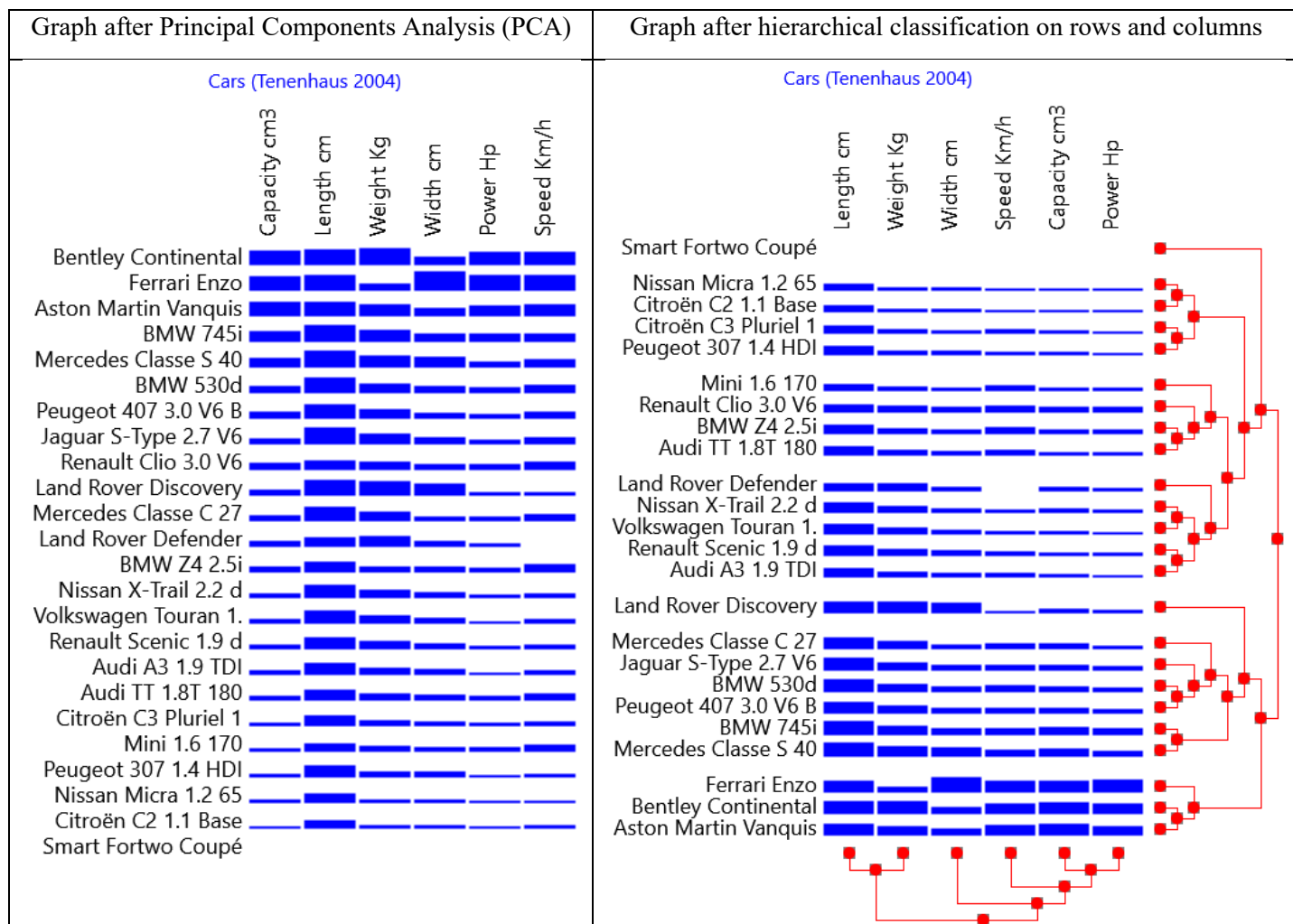
Original data							Graph after export to PNG and Cut/Paste						
Cars (Tenenhaus 2004)							Cars (Tenenhaus 2004)						
	Capacity cm ³	Power Hp	Speed Km/h	Weight Kg	Width cm	Length cm	Capacity cm ³	Power Hp	Speed Km/h	Weight Kg	Width cm	Length cm	
Citroën C2 1.1 Base	1124	61	158	932	1659	3666	Citroën C2 1.1 Base						
Smart Fortwo Coupé	698	52	135	730	1515	2500	Smart Fortwo Coupé						
Mini 1.6 170	1598	170	218	1215	1690	3625	Mini 1.6 170						
Nissan Micra 1.2 65	1240	65	154	965	1660	3715	Nissan Micra 1.2 65						
Renault Clio 3.0 V6	2946	255	245	1400	1810	3812	Renault Clio 3.0 V6						
Audi A3 1.9 TDI	1896	105	187	1295	1765	4203	Audi A3 1.9 TDI						
Peugeot 307 1.4 HDI	1398	70	160	1179	1746	4202	Peugeot 307 1.4 HDI						
Peugeot 407 3.0 V6 B	2946	211	229	1640	1811	4676	Peugeot 407 3.0 V6 B						
Mercedes Classe C 27	2685	170	230	1600	1728	4528	Mercedes Classe C 27						
BMW 530d	2993	218	245	1595	1846	4841	BMW 530d						
Jaguar S-Type 2.7 V6	2720	207	230	1722	1818	4905	Jaguar S-Type 2.7 V6						
BMW 745i	4398	333	250	1870	1902	5029	BMW 745i						
Mercedes Classe S 40	3966	260	250	1915	2092	5038	Mercedes Classe S 40						
Citroën C3 Pluriel 1	1587	110	185	1177	1700	3934	Citroën C3 Pluriel 1						
BMW Z4 2.5i	2494	192	235	1260	1781	4091	BMW Z4 2.5i						
Audi TT 1.8T 180	1781	180	228	1280	1764	4041	Audi TT 1.8T 180						
Aston Martin Vanquis	5935	460	306	1835	1923	4665	Aston Martin Vanquis						
Bentley Continental	5998	560	318	2385	1918	4804	Bentley Continental						
Ferrari Enzo	5998	660	350	1365	2650	4700	Ferrari Enzo						
Renault Scenic 1.9 d	1870	120	188	1430	1805	4259	Renault Scenic 1.9 d						
Volkswagen Touran 1.	1896	105	180	1498	1794	4391	Volkswagen Touran 1.						
Land Rover Defender	2495	122	135	1695	1790	3883	Land Rover Defender						
Land Rover Discovery	2495	138	157	2175	2190	4705	Land Rover Discovery						
Nissan X-Trail 2.2 d	2184	136	180	1520	1765	4455	Nissan X-Trail 2.2 d						

Here, columns carry different units (cm³, CV, Km/h, Kg, cm) and are not directly comparable. Bar heights, which are proportional to the data values, have no meaning here.

First, columns (variables) must be standardised: each data value is centred on the variable minimum then divided by the column standard deviation. This operation yields “pure numbers”, or “scale-less” numbers: if i is a row-car and j a column-variable, then X_{ij} becomes $(X_{ij} - \text{Min}_j)/\sigma_j$.

Since *AMADO-online* can only display and process positive numbers, each column is centred on its minimum and the smallest value of the column becomes zero. All remaining computations are performed on these “pure numbers”.

In this example, the "Smart Fortwo Coupé" is the smallest car and all values associated with that car will become zero in the graphs.



The classification tree outlines clear automobile classes:

- the *Smart Fortwo Coupé*, the smallest of the lot on all parameters, is alone;
- the *Citroën C2*, *Nissan Micra*, *Citroën C3* and *Peugeot 307* appear to form a homogeneous group of 4 small cars;
- the small sport cars, *Mini*, *Renault Clio*, *BMW Z4* and *Audi TT*, are grouped together;
- the next class is comprised of large family cars *Land Rover Defender*, *Nissan X-Trail*, *Volkswagen Touran*, *Renault Scenic* and *Audi A3*;
- the *Land Rover Discovery*, a long, wide, heavy and relatively low performer for its size and rather slow, is in a class of its own;
- then the six large, fast and responsive cars, *Mercedes Classe C*, *Jaguar S*, *BMW 530d*, *Peugeot 407*, *BMW 745i*, *Mercedes Classe S* form the next class;
- lastly, *Ferrari*, *Bentley* and *Aston Martin*, the larger, more expensive and faster cars, form their own class.

The graph also shows that the unique characteristics of the *Land Rover Discovery* and of the *Ferrari* explain why *width* is not well correlated with the four other variables.

On the variable side, *weight* and *length* are highly correlated, as are *cubic capacity* and *power*. *Speed*, on the one hand, and *width*, on the other, are less correlated with the other characteristics, and we can see why on the graph: the *Land Rover Discovery* is wide, heavy and not very fast, while the *Ferrari Enzo* is light but very wide, while being very powerful and very fast.

5 BERTIN graphs

With *AMADO-online*, one can graphically display two-way tables of data, and highlight data structures by permuting rows and columns: a diagonal (seriation) if it exists, a block diagonal, or a structure of crossed groups of lines and columns.

This User Guide shows a variety of table types, with source data and the sequence of commands required to reproduce the graphs shown here. *AMADO-online* is best suited to small and mid-sized tables (up to about 50 rows and columns)⁴, similar to those often found in social sciences where each element is precisely defined and must be interpretable in context.

The graphs produced by *AMADO-online* are easy to read, giving the user a direct access to the results: each piece of information, i.e. each data entry, is displayed as captured, yet presented as rectangles whose heights are proportional to the original data values, in absolute or relative (%) terms.

Permuting rows and columns of a data matrix to make underlying structures emerge is old: Sir W. M. Flinders Petrie showed in 1899 a "sequence in prehistoric remains", that is, a seriation of shapes and elements of ornamentation on objects found during a dig in Egypt. Many have highlighted how this idea had a growing influence on applied mathematics, especially on behavioural sciences (Arabie, Boorman and Levitt, 1978, Caraux, 1984 and Marcotorchino, 1987).

Jacques Bertin (1967, 1977) revealed underlying structures in data matrices by aligning histograms on a common suitable scale and permuting their elements. Since then, that approach has grown considerably in France and around the world (Bord 1997, Palsky 2017, Harvey 2019). At first, Bertin and his team at the École des Hautes Études used to work with rows of cubes that were moved by hand. Then the development of numerical methods for the analysis of multivariate data (Cordier 1965, Benzécri 1973, Arabie et al. 1978, Greenacre 1984, Caraux 1984, Tenenhaus and Young 1985, Hoffman and DeLeeuw 1992) made that purely visual approach more and more obsolete.

Surely numerical methods of analysis are quick at finding the main traits of a data structure which will be legible on a graph. This is a considerable saving in time when trying to determine the best set of permutations of n rows and p columns among the $n!p!$ possible solutions. But, in correspondence analysis, if the list of coordinates and other numerical aids to interpretation, as *contribution to inertia*, *relative contribution of an axis to a point*, are useful to the statistician, they may not be readily understood, even by an educated reader; the same can be said of *simultaneous display*, *display of supplementary rows and column profiles*, etc. Their correct interpretation requires a trained eye, and their very esotericism might be the source of their popularity... On the other hand, classification trees give a useful but distorted (*ultrametric*⁵) view of the data matrix, and almost always for one of the margins of the original two-way table. Then, a large number of means, marginal and conditional, of standard deviations, contributions, etc., are necessary to derive the meaning of such a tree.

Conversely, graphs produced by *AMADO-online* use factor analysis or hierarchical clustering while giving the user readily access to the results: each piece of information, that is, each table entry, is displayed as captured, in absolute or relative (%) terms. Only the order of rows or columns has changed, but all is there.

⁴ Jean Daniel Fekete et al. (2015, 2016) adapted Bertin's methods to large tables.

⁵ A distance is *ultrametric* if all triangles are isocèles, the 3rd side shorter than the two equal sides. This is the case when the distance in hierarchical clustering when the distance between two items is measured as the height of the lowest node that connects them. This type of distance is really peculiar; for example, it is impossible to have more than three points on a plane such that their distances (in a classic geometric sense) complies with the *ultrametric* condition; see §0.

6 References

- Arabie Phipps, Scott A Boorman & Paul R Levitt (1978) *Constructing blockmodels: How and why?* Journal of Mathematical Psychology, Vol.17-1, PP 21-63 [https://doi.org/10.1016/0022-2496\(78\)90034-2](https://doi.org/10.1016/0022-2496(78)90034-2)
- Behrisch Michael, Benjamin Bach, Nathalie Henry Riche, Tobias Schreck & Jean-Daniel Fekete (2016) *Matrix Reordering Methods for Table and Network Visualization*. Computer Graphics Forum, Wiley, 35, pp.24.
- Benzécri Jean-Paul (1973) *L'Analyse des Données, t. I : Taxinomie ; t. II : L'Analyse des Correspondances*, Bordas, Paris (1^{re} édition 1973, 2^e édition 1976, 3^e édition 1980, 4^e édition 1982)
- Bertin Jacques (1967) *Sémiologie Graphique. Les diagrammes, les réseaux, les cartes*, Paris, La Haye, Mouton, Gauthier-Villars. 2e édition : 1973, 3e édition : 1999, EHESS, Paris.
- Bertin Jacques (1977) *La graphique et le traitement graphique de l'information*. Réédition, Zones sensibles, 2017 <http://www.zones-sensibles.org/jacques-bertin-la-graphique-et-le-traitement-graphique-de-linformation/>
- Bord Jean-Paul (1997) *30 years of graphic semiology in honour of Jacques Bertin* <https://journals.openedition.org/cybergeol/501?lang=en>
- Carau, Gilles (1984) *Réorganisation et représentation visuelle d'une matrice de données numériques : un algorithme itératif*. R. de Stat. Appliquée 32-4, pp. 5-23. http://www.numdam.org/item/RSA_1984__32_4_5_0/
- Chauchat Jean-Hugues & Alban Risson (1998) *Bertin's Graphics and Multidimensional Data Analysis*, in Visualization of Categorical Data, J. Blasius, M. Greenacre Editors. <https://books.google.fr/books?id=YEjKNYBvUfC&printsec=frontcover&dq=Visualization+of+Categorical+Data,+1998&hl=fr&sa=X&ved=0ahUKEwj38KiN5LrIAhUHTRoKHVF9DhMQ6AEIKzAA#v=onepage&q=Visualization%20of%20Categorical%20Data%2C%201998&f=false>
- Cordier Brigitte (1965) *L'Analyse des Correspondances*, Thèse de Doctorat (3^e cycle), Université de Rennes.
- Fekete Jean-Daniel, Jeremy Boy (2015) *Recherche en visualisation d'information ou Dataviz : pourquoi et comment ?* I2D – Information, données & documents 2015/2 <https://www.cairn.info/revue-i2d-information-donnees-et-documents-2015-2-page-32.htm#>
- Greenacre Michael (1984) *Theory and applications of correspondence analysis*. Academic Press. <https://www.ogi-nic.net/CARME-N/download/theory%20and%20applications%20of%20correspondence%20analysis.pdf>
- Harvey Francis (2019) *Jacques Bertin's legacy and continuing impact for cartography*. <https://doi.org/10.1080/15230406.2019.1533784>
- Hoffman Donna L. & Jan De Leeuw (1992) *Interpreting multiple correspondence analysis as a multidimensional scaling method*. Marketing Letters 3, 259–272. <https://doi.org/10.1007/BF00994134>
- Lehman Bernard (1995) *L'orchestre dans tous ses éclats : sociologie de la profession de musicien*, Thèse de doctorat, EHESS, Paris
- Palsky Gilles (2017) *La Sémiologie graphique de Jacques Bertin a cinquante ans* <https://visionscarto.net/la-semiologie-graphique-a-50-ans>
- Petrie Sir W. M. Flinders (1899) *Sequences in Prehistoric Remains*, The Journal of the Anthropological Institute of Great Britain & Ireland 29 pp.295–301. <https://babel.hathitrust.org/cgi/pt?id=uiug.30112089727678&view=1up&seq=1>
- Pinol Jean-Luc (2019) *Convois, La déportation des Juifs de France, Paris, Éditions du Détour*
- Renfrew Colin & Paul Bahn (1991), *Archaeology: Theories, Methods and Practice*, London UK: Thames and Hudson.
- Snee Ronald D. (1974) Graphical displays of two-way contingency tables, *Am. Statistician*, 28, pp. 9-12.
- Tenenhaus Michel & Forrest W. Young (1985) *An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis and other methods for quantifying categorical multivariate data*. *Psychometrika* 50, 91–119. <https://doi.org/10.1007/BF02294151>
- Texier Laurence (1999), « Une clarification de l'offre d'implantation en marketing territorial : produit de ville et offre de territoire », *RERU Revue d'économie régionale et urbaine*, no 5, p. 1021-1036

7 Appendices

7.1 Appendix 1: PNG or SVG format?

The command “**File / Export to SVG**” copies the displayed graph to the Downloads folder as a *Scalable Vector Graphic* (SVG) file.

The command “**File / Export to PNG**” copies the displayed graph to the Downloads folder as a *Portable Network Graphics* (PNG) file.

SVG and PNG are both image formats. SVG is a vectorial format where an image is represented by a set of mathematical figures while PNG is a binary image format that uses a compression algorithm to represent the image as a set of pixels.

These are important differences between SVG and PNG.

Sr. No.	Key	SVG	PNG
1	Stands for	SVG stands for <i>Scalable Vector Graphics</i> .	PNG stands for <i>Portable Network Graphics</i> .
2	Image type	SVG image is vector based.	PNG image is pixel based.
3	On Zoom	SVG image quality remains same while zooming.	PNG image quality degrades while zooming.
4	Basis	SVG images is made up of paths and shapes.	PNG images is made up of pixels.
5	Editable	SVG images are editable.	PNG images are not editable.
6	Extensions	SVG images use .svg extension.	PNG images use .png extension.
7	Usage	SVG images are used in devices using high resolution images.	PNG images are generally used in image creation.

<https://www.tutorialspoint.com/difference-between-svg-and-png>

How to reshape such graphs pasted in Word, Excel or Powerpoint is covered in Annex 7.2.

7.2 Appendix 2: Cropping a PNG or SVG picture file pasted in Word, Excel ou PowerPoint

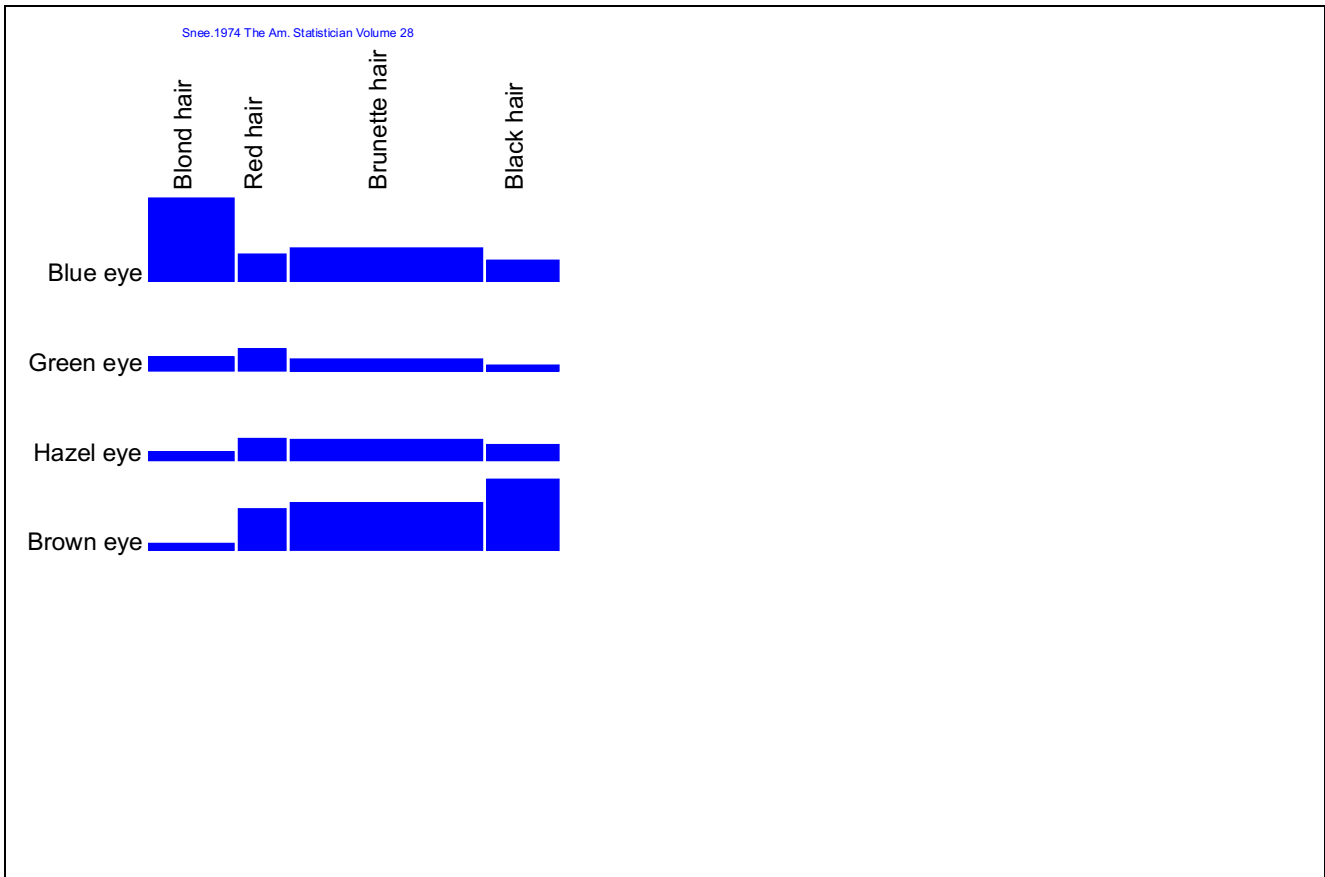
Once the selected picture file is pasted in Word, Excel, PowerPoint, etc., **click on its edge** to select it, then

[Format Graphic](#)

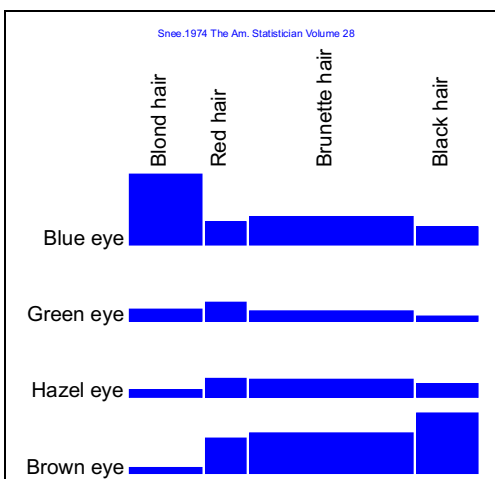
click right to open a dialog box and click on **picture format**; one gets this sub-menu selection



. **Click on the rightmost icon** , then on **Crop**; margins can then be changed in the dialog box.



These actions yield the next graph. Its size can now be adjusted by clicking the picture edge and moving the mouse.



7.3 Appendix 3: Correspondence Analysis: optimal scoring of rows and columns

Why is CA efficient at reordering rows and columns of a two-way table?

Tenenhaus and Young (1985) showed many ways to understand Correspondence Analysis of a two-way table. Identical methods, bearing different names, have emerged over time in various countries: *Optimal Scaling*, *Optimal Scoring & Appropriate Scoring methods* in the USA; *Dual Scaling* in Canada; *Homogeneity Analysis* in the Netherlands; *Scalogram Analysis* in Israel; *Quantification Method* in Japan, etc.

As an illustration of CA, consider the table comprised of eye and hair colour of 16 people.

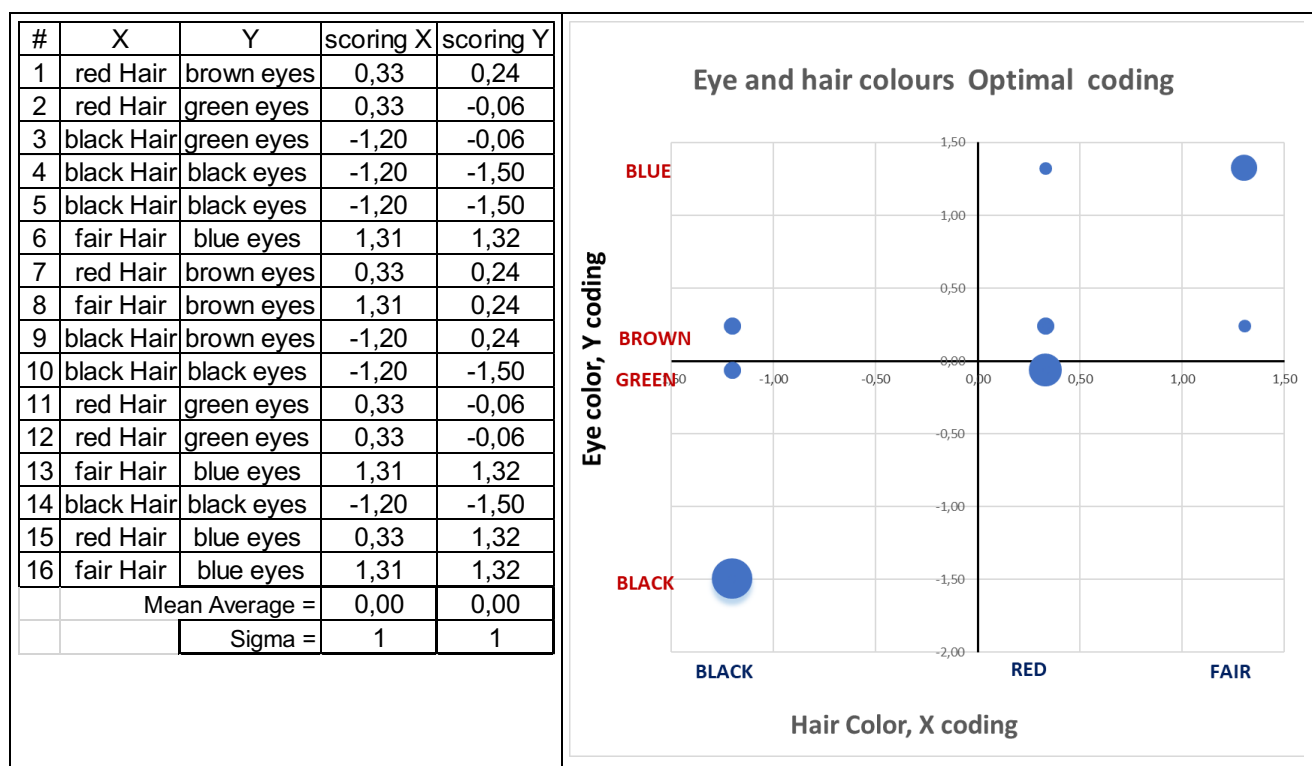
These data can be summarised in a two-way table crossing eye and hair colours.

#	X	Y
1	red Hair	brown Eyes
2	red Hair	green Eyes
3	black Hair	green Eyes
4	black Hair	black Eyes
5	black Hair	black Eyes
6	fair Hair	blue Eyes
7	red Hair	brown Eyes
8	fair Hair	brown Eyes
9	black Hair	brown Eyes
10	black Hair	black Eyes
11	red Hair	green Eyes
12	red Hair	green Eyes
13	fair Hair	blue Eyes
14	black Hair	black Eyes
15	red Hair	blue Eyes
16	fair Hair	blue Eyes

	fair Hair	black Hair	red Hair
blue Eyes	3	0	1
brown Eyes	1	1	2
black Eyes	0	4	0
green Eyes	0	1	3

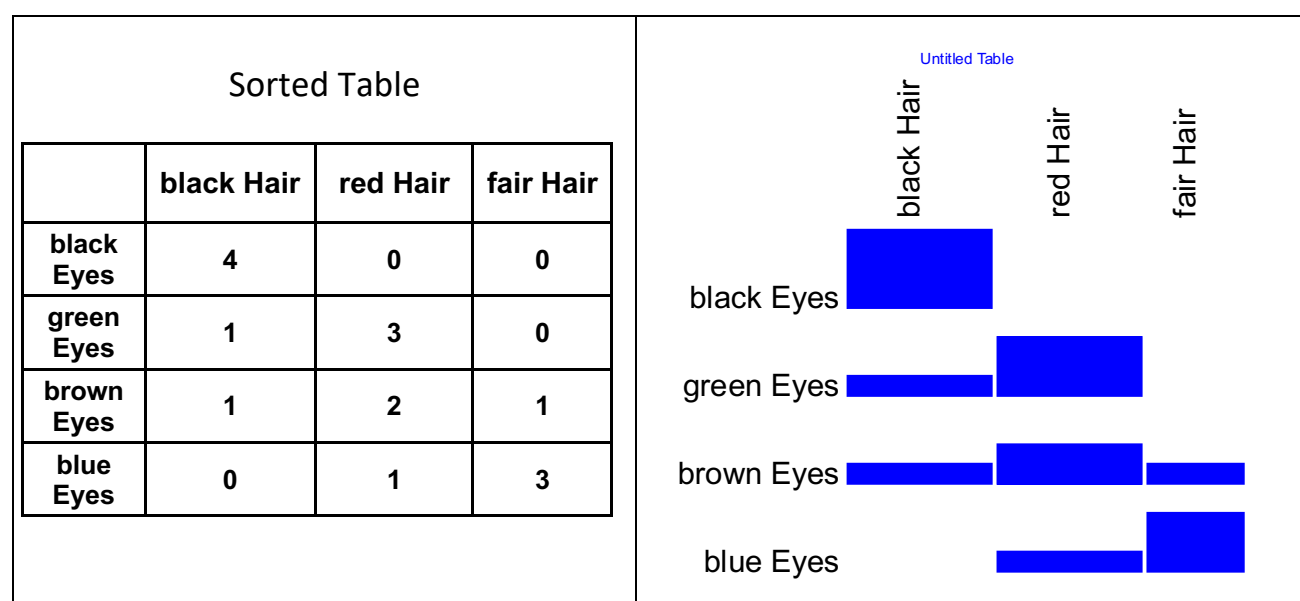
A numerical scoring of the two qualitative variables is a set of numbers associated to the values of each of the qualitative variables (“X score” and “Y score” in the table below).

The optimal scoring is that which maximises the correlation “R” between the two series of scores. Among the infinite number of possible solutions, one chooses those of zero means and variances equal to one.



The values of the optimal scoring are identical to the coordinates of the table rows and columns on the first factor of the Correspondence Analysis.

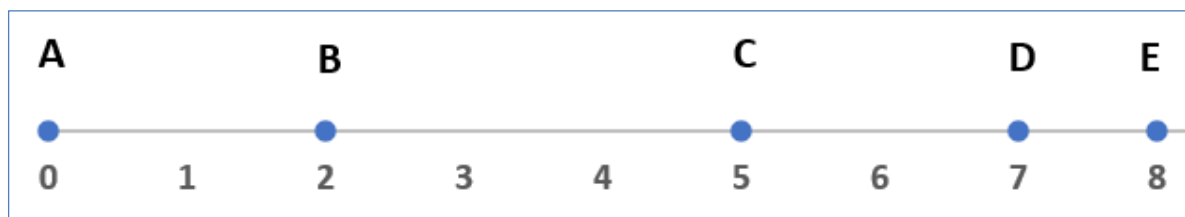
Sorting the table according to the scores gives *the graph that best displays the structure of the data*, that is, the relationship between the two qualitative variables, the rows and the columns (here, the data are transposed):



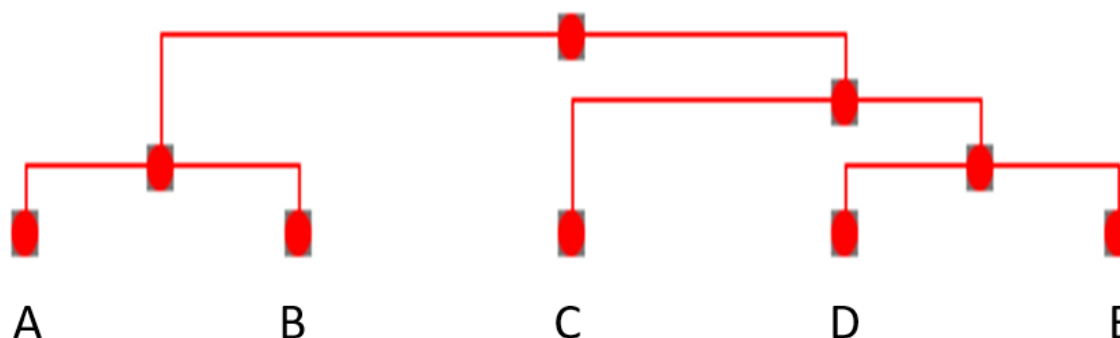
7.4 Appendix 4: The distorted view from a dendrogram (or classification tree)

All classifications simplify the reality at the cost of a distorted representation. *AMADO-online* produce graphs that benefit from the simplification while remaining true to the data.

On a dendrogram, the distance between two items is the height at which a node connects the branches that end in the selected items. Consider this simple example of 5 points on an axis:



A hierarchical classification will yield a tree similar to that shown below; items B and C connect at the top of the tree, just like items A and E. The impression is that B and C are as far from one another as A and E are. This is not quite the reality on the original axis!



The notion of ultrametric distance captures this distortion: *all triangles have two sides equal in length and longer than the third side (or else, all sides are equal)*. This characteristic is visible in the data matrix shown below where an entry corresponds to the height of the node connecting two items: items D and E connect one level up, then A and B one level higher, then C and the pair (D, E) connect at level 3, and lastly (A, B) connects to (C, (D,E)) at level 4.

	A	B	C	D	E
A	0	2	4	4	4
B	2	0	4	4	4
C	4	4	0	3	3
D	4	4	3	0	1
E	4	4	4	1	0

This “ultrametric” situation does not correspond to any usual geometric situation from 3 distinct points on a line or 4 distinct points on a plane, etc.

The dendrogram gives a distorted view of the reality; in the examples shown in this Guide, *AMADO-online* graphs have the benefit of the reordering of the items according to the dendrogram and be true to the original data.

8 Acknowledgements

Heartfelt thanks go to:

- The Paris Time Machine consortium team who supported this work: Jean-Luc Pinol, Hélène Noizet, Paul Rouet, Laurent Costa, Julien Avinain, Éric Mermet and the scientific board,
- Our academic colleagues who helped with translating the menus: Annie Morin, Alexis Rouet, Annie Le Gloahec and Jairo Cugliari (Spanish), Linda Gattuso (Italian), Iryna Zolotariova (Ukrainian) and Olena Orobinska-Goncharova (Russian),
- Jean Dumais for his careful proofreading of the text, the menus and examples proposed, the translation into English and his many recommendations for improvement,
- Basu Tallur for his assistance in the creation of the video,
- Sylvain Clément who dubbed the English version of the video,
- Alban Risson who, while an undergraduate student in computer science, created the first version of AMADO.