AMADO-online user's guide

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

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

AMADO-online was designed by **Nguyen-Khang PHAM** (University of Can Tho) from the code of **Alban Risson**

This manual was written by **Jean-Hugues Chauchat** (Université Lyon 2) Version of 11 Nov 2020

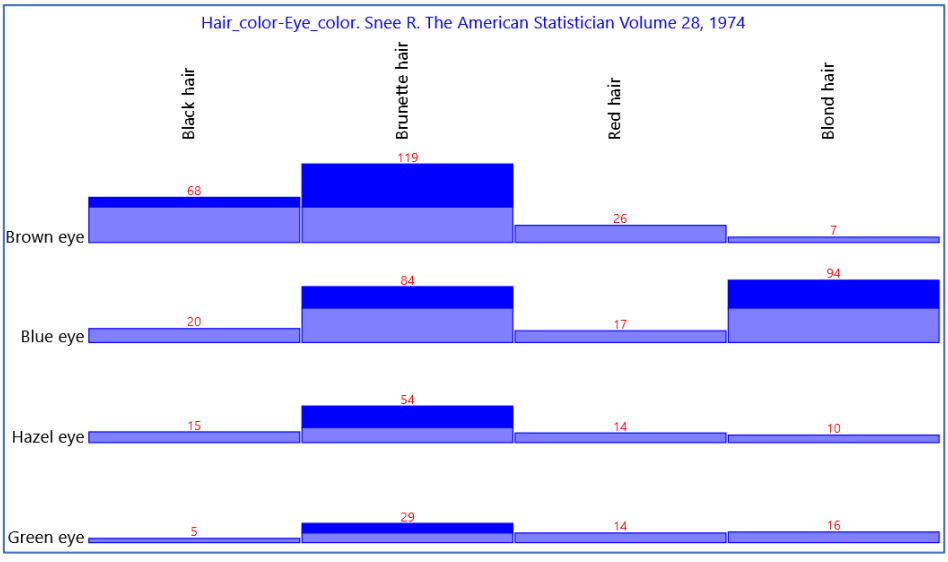
To start: Choose the language on the home page. Launch yourself.

File: open a data file

File / Open a file (in UNICODE format, or UTF-8 if there are accents or special characters), with "tab separators".
Open / Sample / EN_Hair_color-Eye_color.TXT

Or **Copy** the table area into a spreadsheet, then "Edit / Paste" in AMADO online [1]. **If the cell at the top left of the table contains text, it appears as the title of the chart.**

	A	B	C	D	E
1	Hair_color-Eye_color. Snee R. The American Statistician Volume 28, 1974	**Black hair**	**Brunette hair**	**Red hair**	**Blond hair**
2	**Brown eye**	68	119	26	7
3	**Blue eye**	20	84	17	94
4	**Hazel eye**	15	54	14	10
5	**Green eye**	5	29	14	16

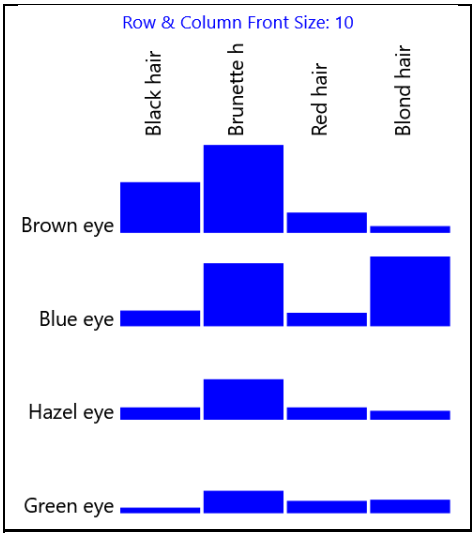


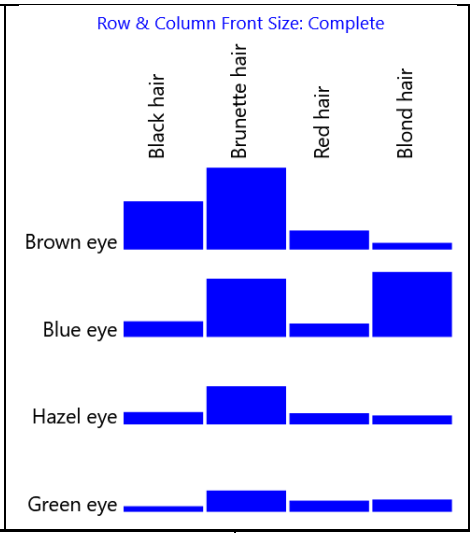
The numbers in the table are represented by the heights (and areas) of the rectangles.

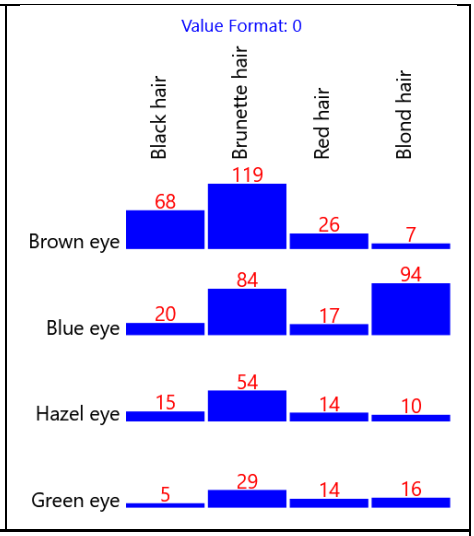
[1] If you are using the Firefox browser, go to the menu at the top right of Firefox Customize / Customize Toolbar / Move the "Edit Control" commands to the toolbar. You will then be able to copy a table.

Format: Row Legends; Column Legends; Value Format

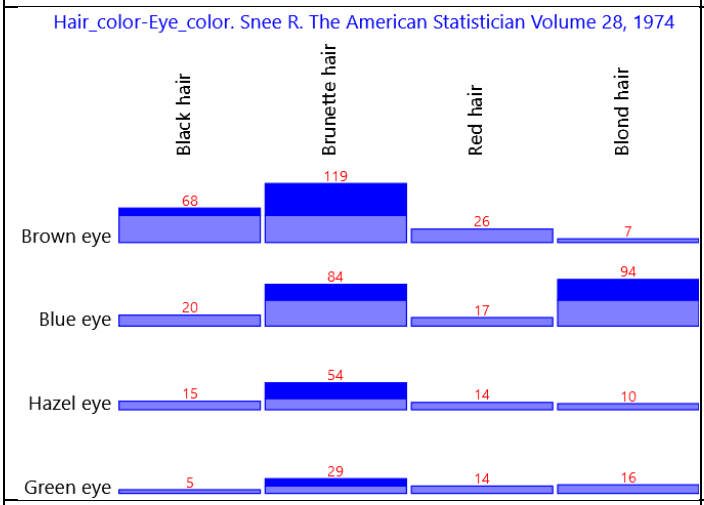
By default, labels are displayed with 10 characters. To display the full labels "**Format / Row Legend / Complete**", and "**Format / Column Legend / Complete**".



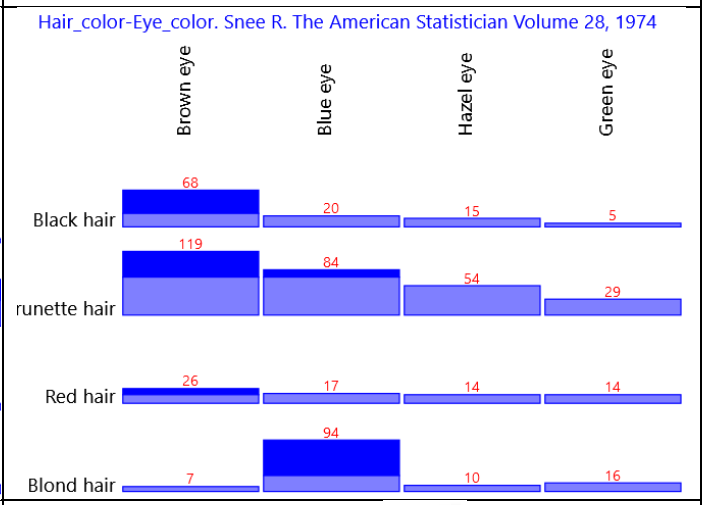




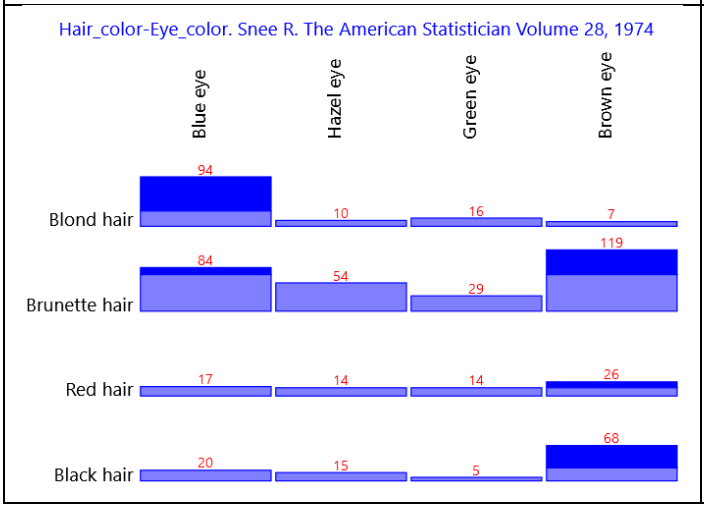
After « **Typography / Increase Column Spacing** and after resizing the window



After « **Process / Transpose** »

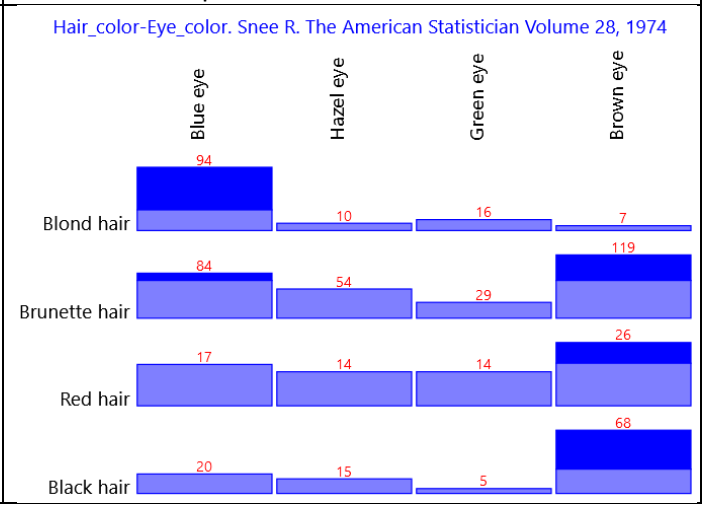


After having moved (by clicking and moving the mouse) the column "Brown Eyes" **then the line** "Black hair", a diagonal structure is obtained which illustrates the correlation between eye and hair colour:



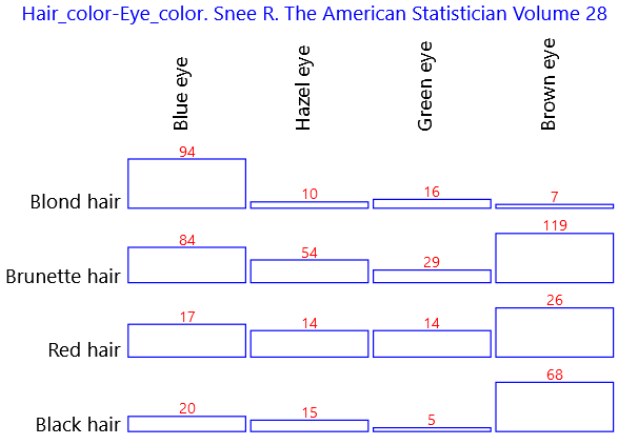
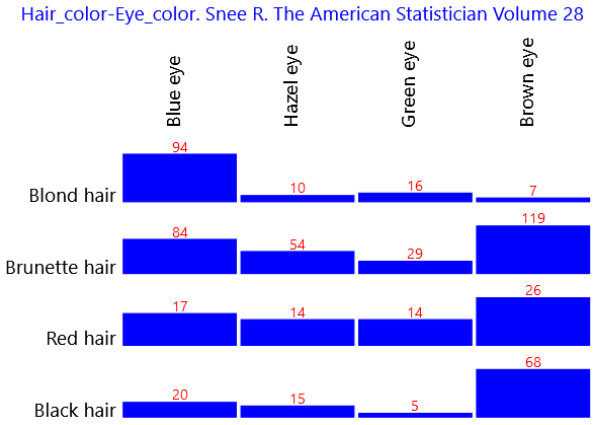
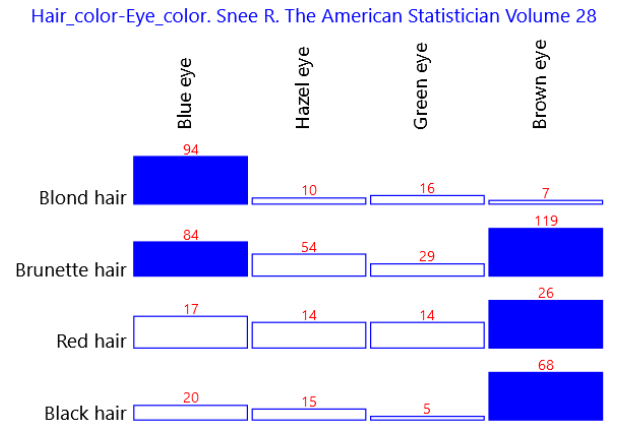
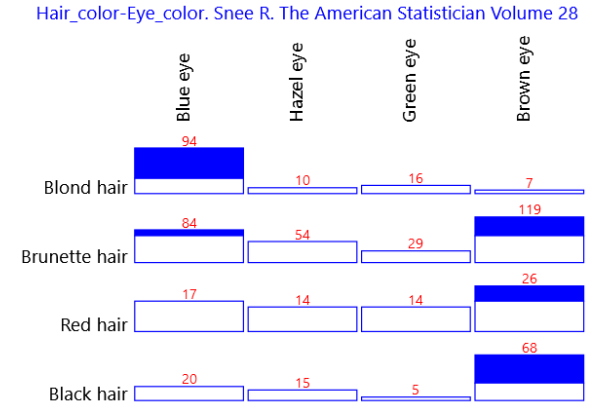
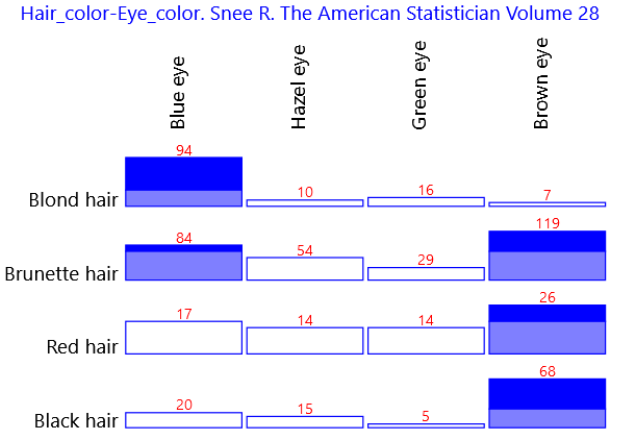
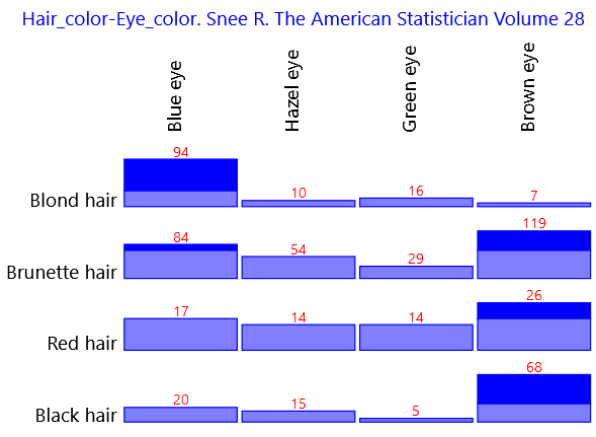
After « **Format / Scale by Line** ✓

The maxima of the lines all have the same height: in each of the lines, the heights become proportional to the numbers represented.



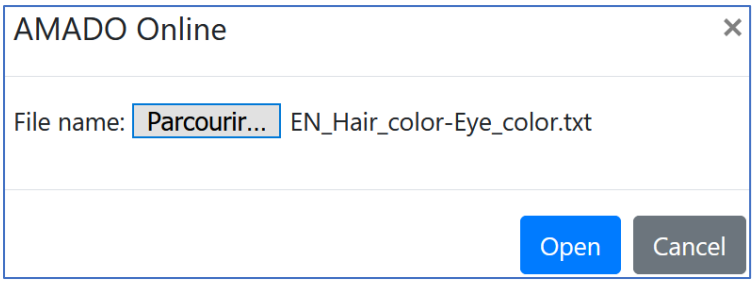
The 6 Display Modes

« Format / Mode 1, etc. »  allow, in different ways, to visually distinguish values below or above the average value of each line.

<table>
<tr>
<td>Modes
1 & 2</td>
<td> 
</td>
<td> 
</td>
</tr>
<tr>
<td>Modes
3 & 4</td>
<td> 
</td>
<td> 
</td>
</tr>
<tr>
<td>Modes
5 & 6</td>
<td> 
</td>
<td> 
</td>
</tr>
</table>

File: Open, Samples…, Export to SVG, Export to PNG

File / Open allows you to open a .TXT format file from your computer. The fields must be separated by "Tab" characters. If the title or the line or column labels contain accented characters, the file must have been saved in UTF or UNICODE format.



File / Sample: Some ".TXT" files are available, including those used in this Guide. The English file names begin with EN (Example: EN_Hair_color-Eye_color.TXT). If the cell at the top left of the .TXT table contains a text, this appears as the title of the graph.

Edit / Set Title allows you to create or change the title of the chart.

Export the Chart

File / Export to SVG copies the chart, in SVG (*Scalable Vector Graphic*) format, to the "Downloads" folder on the computer. See the appendix on how to crop an SVG chart that has been pasted into Word, Excel or PowerPoint.

File / Export to PNG copies the chart in PNG (*Portable Network Graphics*) format, to the "Downloads" folder on the computer, so that it can be pasted into Word, Excel, or PowerPoint.

SVG and PNG both are a type of image format to store images. SVG is a vector-based image format where an image is represented by set of mathematical figures and PNG is a binary image format and it uses lossless compression algorithm to represent image as pixels.

Following are the important differences between SVG and PNG.

Sr. No.	Key	SVG	PNG
1	Stands for	SVG stands for *Scalable Vector Graphics*.	PNG stands for *Portable Network Graphics.*
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

Moving a row or column

To show the data structure, rows and/or columns can be reordered.

Moves can be made "by hand": select a row by clicking on it and move it while keeping the mouse button pressed. You can also move a block of contiguous rows by clicking on one row and then on another using the Ctrl key. The same applies to one column or a block of contiguous columns.

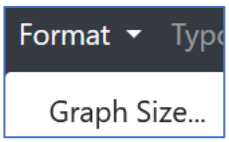
Automatic diagonalization (by Factorial Correspondent Analysis) and classification methods will be discussed later.

Sort the rows according to the values of a column (or sort the columns according to the values of a row)

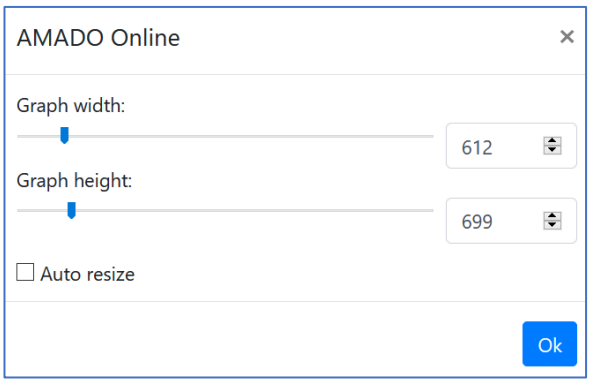
Select a column by clicking on its label, then **Process / Sort ascending (or descending)**.

Same when you select a line.

How to change the size, and the height/width ratio, of the graphic



Opens the window:



Uncheck: ☑ Auto resize ➔ ☐ Auto resize

Then choose the desired width and height for the graphic. At the end, validate by clicking on: Ok

Frequency Data Processing: diagonalisation & classification

Frequency Data… These commands are suitable for frequency cross-tabulation:

- in a population described by two qualitative variables, *N ij* is the number of individuals with the character "*i*" of the variable "rows" and the character "*j*" of the variable "columns" (examples: number of inhabitants of district "*i*" with the occupation "j". Number of people according to the colours "*i*" of their eyes and "*j*" of their hair),

- tables of *0/1*. *N ij*=1 if the character "j" of the variable "column" is present in the individual "*i*", *N ij*=0 if not (example: the pottery found in the archaeological site "*i*" has, or does not have, the shape or design "*j*"),

- square co-occurrence tables: *N ij* is the number of times characters "*i*" and "j" have been matched together.

For **Frequency Data / Processing by CA** (Factorial Correspondence Analysis) and **Hierarchical Clustering**.

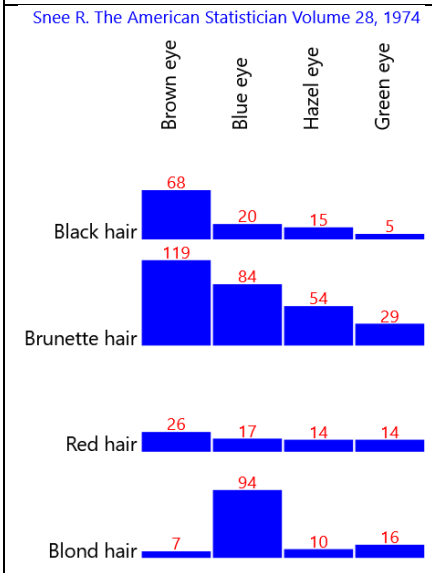
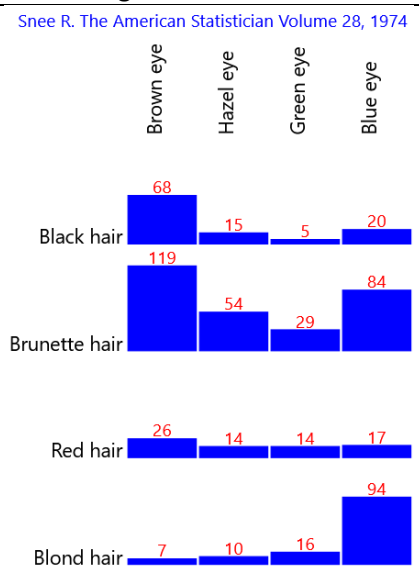
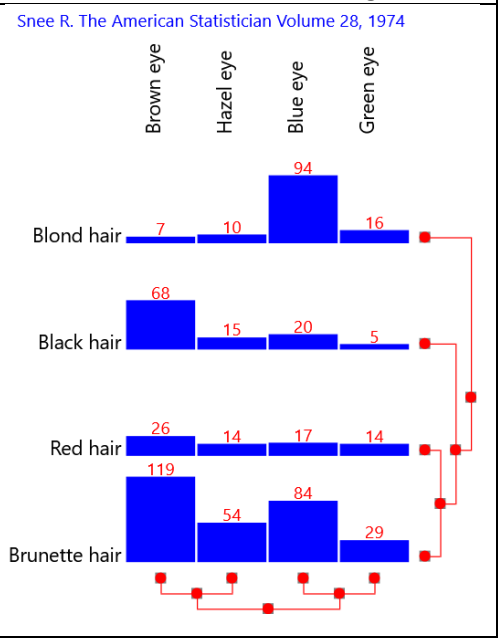
Example-1: eye and hair colours of 592 people

Open / Sample / EN_Hair_color-Eye_color.TXT

Hair_color-Eye_color. Snee R. The American Statistician Volume 28, 1974	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

By default, labels are displayed with 10 characters.

To display the full labels "**Format / Row Legend / Complete**", and "**Format / Column Legend / Complete**".

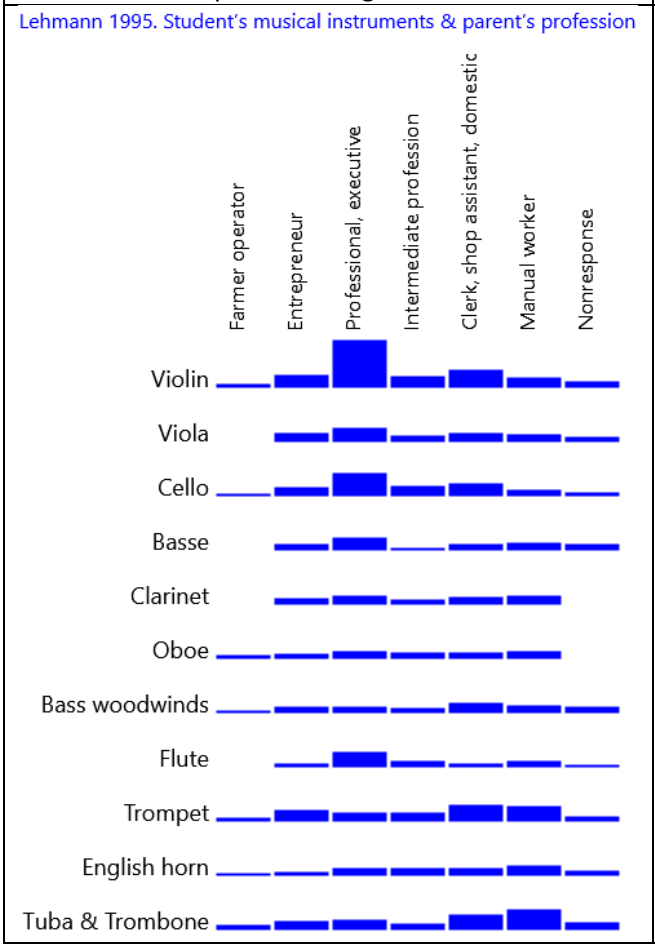
Graph. Original data	Frequency Data / Processing with CA	Frequency Data / Hierarchical Clustering
		

Example-2: Instruments played by students of the Paris National Conservatory of Music and Dance (CNSM) and their parents' professions and socio-professional categories (PCS)

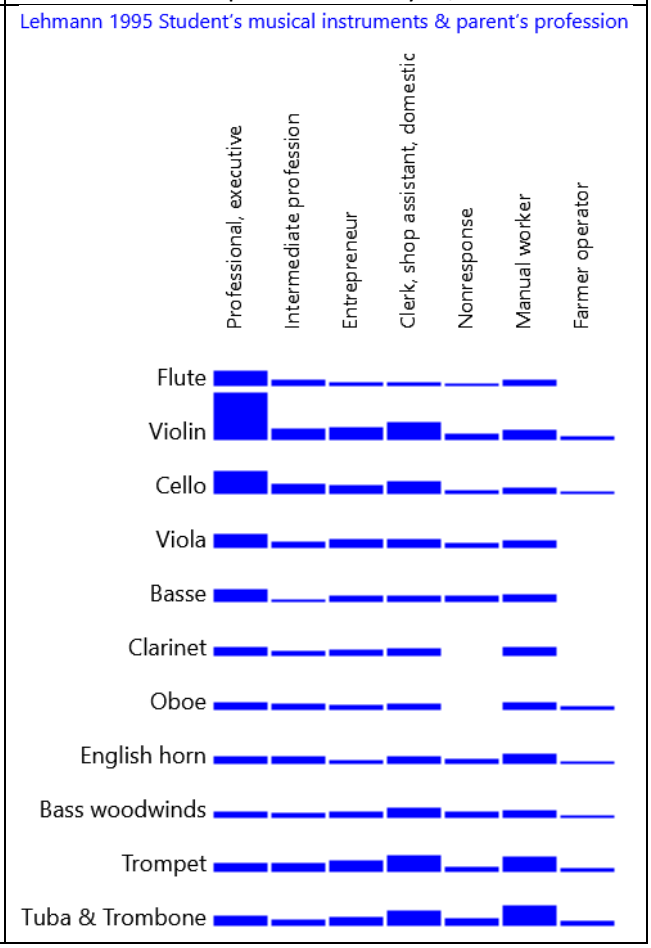
L'orchestre dans tous ses éclats : sociologie de la profession de musicien, by B. Lehmann. Doctoral thesis in Sociology defended in 1995 in Paris, EHESS. **File: EN_musical_instruments-parents_professions.TXT**

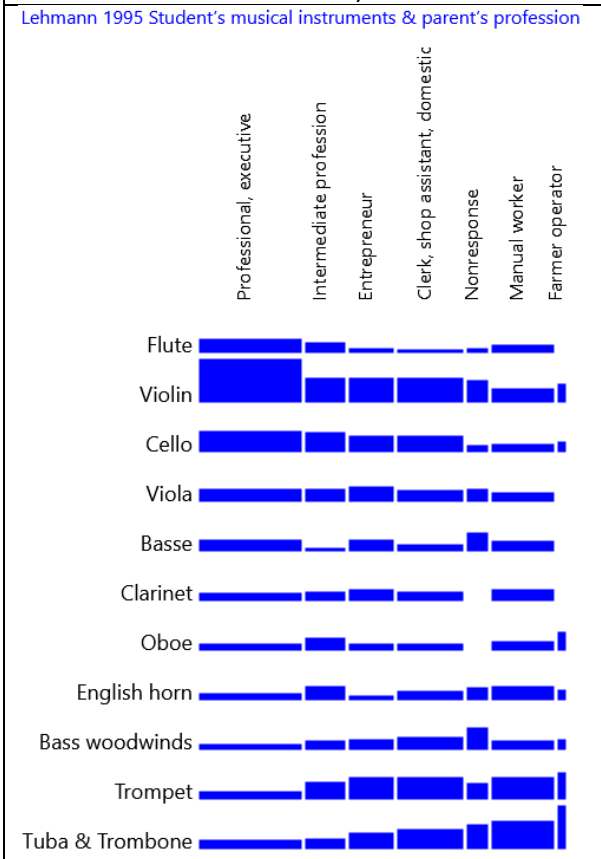
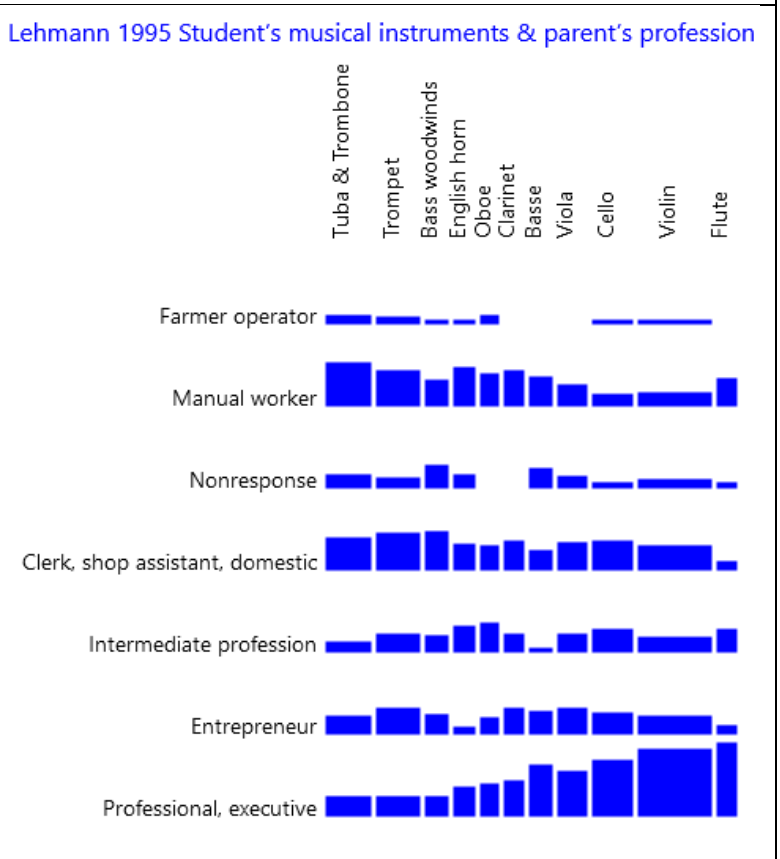
Sociology of the music profession. Lehmann 1995. Student's musical instruments & parent's profession	Farmer operator	Craftsman, shopkeeper, entrepreneur	Professional, executive, higher intellectual profession	Intermediate profession	Clerk, shop assistant, domestic worker	Manual worker	Nonresponse
Violin	2	17	69	15	24	12	7
Viola	0	10	18	7	11	8	4
Cello	1	11	32	12	16	6	2
Basse	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

Graph of the original table



After Correspondence Analysis, row data



After Correspondence Analysis & Columns (professions) % & Weight Width of Columns (to have, again, rectangle areas proportional to the original numbers)	After Correspondence Analysis (transposed) & Columns (music instruments) % & Weight Width of Columns (to have, again, rectangle areas proportional to the original numbers)
	

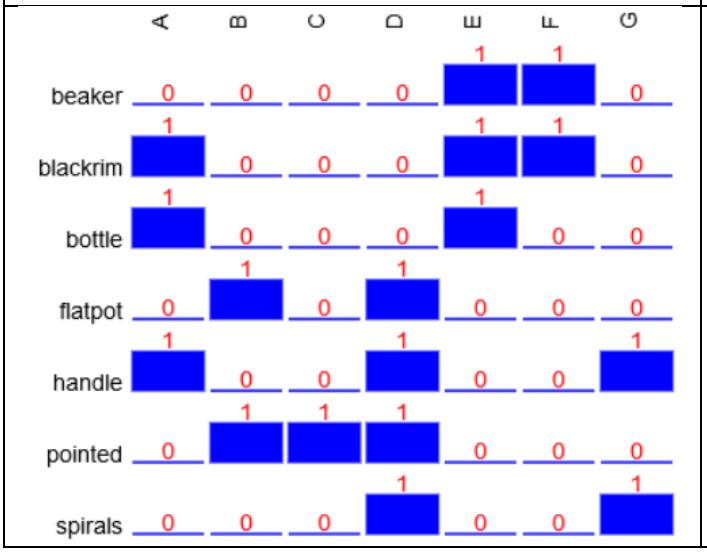
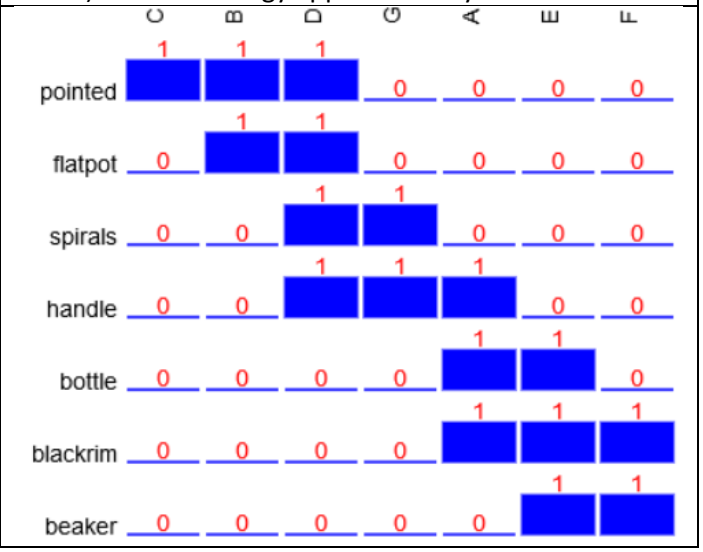
The graph on the right shows that the flute and stringed instruments are chosen more by children from the well-to-do and "cultured" categories. On the other hand, brass and woodwind instruments are played more by the children of employees and workers (these instruments are played in the harmonies and brass bands thanks to which they often began their musical education); the children of farmers are few in number at the CSNS and their share of flute or string players is nil or very low.

Example-3: Chronological "seriation" of archaeological objects.

Example of data inspired by those used by the archaeologist Sir Flinders PETRIE to date tombs excavated at Diospolis Parva in Egypt at the end of the 19th century. He had hypothesised that the type of objects and decorative elements characterised their period and that, consequently, their variations reflected the chronology.

In their book *Archaeology: Theories Methods and Practice*, 1991, Renfrew, C., & Bahn, P. G. give the following pedagogical example. https://en.wikipedia.org/wiki/Seriation_(archaeology)

File: Egyptian_pottery_example.TXT

Raw matrix of presence/absence of decorative elements on pottery according to Tombs A, B, C ...	Matrix becomes diagonal by changing the rows and columns ranks (according Correspondence Analysis first factor): the chronology appears clearly.
	

The 2nd graphic is obtained, in one click, by the commands:

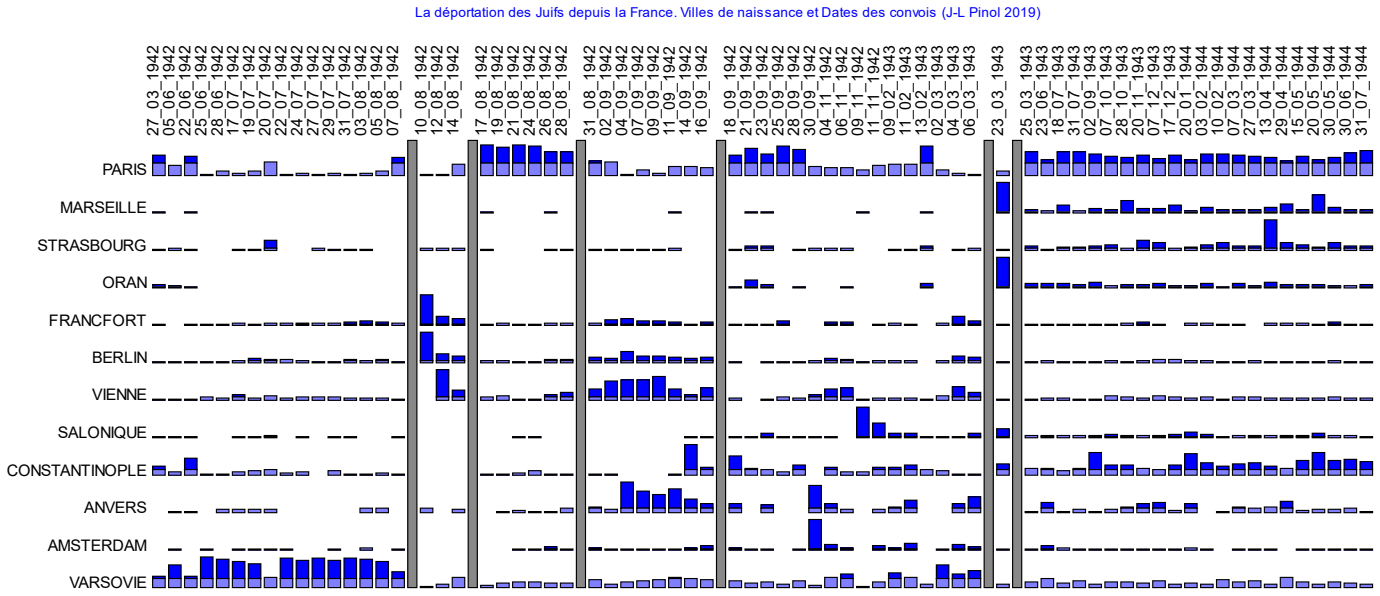
Process / Frequency Data or 0/1 / Processing with Correspondence Analysis

It shows the pottery (C, B, D, G, A, E and F) reordered along an axis of presence-absence of the decorative elements identified by the archaeologist; this rank probably corresponds to the chronological order (direct or inverse) of invention, then of abandonment, of these artistic creations.

Example-4: Chronological data: distribution of Jews deported from France according to convoys and cities of birth

In this example, we do not change the chronological order; the graph produced by AMADO online visualises the data table and supports the historian's commentary.

The graph represents the distribution of deportees to the Nazi death camps, born in a representative subset of cities, according to the convoys (% in columns) that left France. See Jean-Luc Pinol, Convois, *La déportation des Juifs de France*, Paris, Éditions du Détour, 2019.



On this graph, it is clear that the first 19 convoys (from March to August 7th 1942) included mostly Jews refugees in France, born in Warsaw (*Varsovie*); then from Germany (Berlin, Frankfurt) and Austria (Vienna) on August 10th, 12th and 14th 1942; after the *Rafle du Vel d'Hiv* (more than 13,000 French Jews were rounded up in Paris and herded into Paris velodrome), the deportees of the convoys from August 17th to 28th were mostly born in Paris.

From 31 August 1942 until 16 September, eight convoys transported people arrested during the great round-up of 26 August in the *unoccupied zone* where many Jews from Germany (Frankfurt, Berlin) and Austria (Vienna) and Belgium (Antwerp) had taken refuge.

The convoy that left on March 23rd, 1943 deported mostly Jews born in Marseille or Algeria, arrested after the destruction of the *Vieux Port* district of Marseille in January 1943.

Example-5. The working population of the Paris districts (25 to 54 years old) in 2015

File: EN_Paris2015_Districts-Professions.TXT

Paris labour force (2015) 25 to 54 y. old	entrepreneur	Professional, executive	Intermediate profession	Clerk, shop assistant, domestic	Manual worker
Paris 01	601	3651	1491	991	325
Paris 02	647	5969	2239	1404	599
Paris 03	1161	9005	3811	2206	713
Paris 04	871	5908	2621	1738	418
Paris 05	1145	12188	4699	2627	849
Paris 06	1062	7971	2842	1813	487
Paris 07	1532	10642	3417	2836	638
Paris 08	1203	7546	2461	2236	633
Paris 09	1880	15821	6028	3534	1247
Paris 10	2322	21477	9936	7228	2937
Paris 11	3211	35521	17426	10837	4143
Paris 12	2172	29285	14779	10556	3550
Paris 13	2604	28770	16884	15580	5503
Paris 14	2116	26809	11685	9988	2832
Paris 15	4066	50880	21436	15601	4676
Paris 16	4730	27917	10677	8916	2382
Paris 17	4095	37101	15304	11636	4051
Paris 18	4299	37529	21749	17266	8808
Paris 19	3213	25364	18532	17815	7803
Paris 20	3096	30256	21563	18382	7829

File / Open / Parcourir / EN_Paris2015_Arrondissements.TXT

Process / Transpose

Format / Row Legends / Complete

Typography / Increase Column Spacing (2 times)

Format / Graph Size… Graph width = 950 Graph height = 400 (Uncheck: ☑ Auto resize ➔ ☐ Auto resize **)**

Format / Mode 3

Process / Frequency Data / Hierarchical Clustering

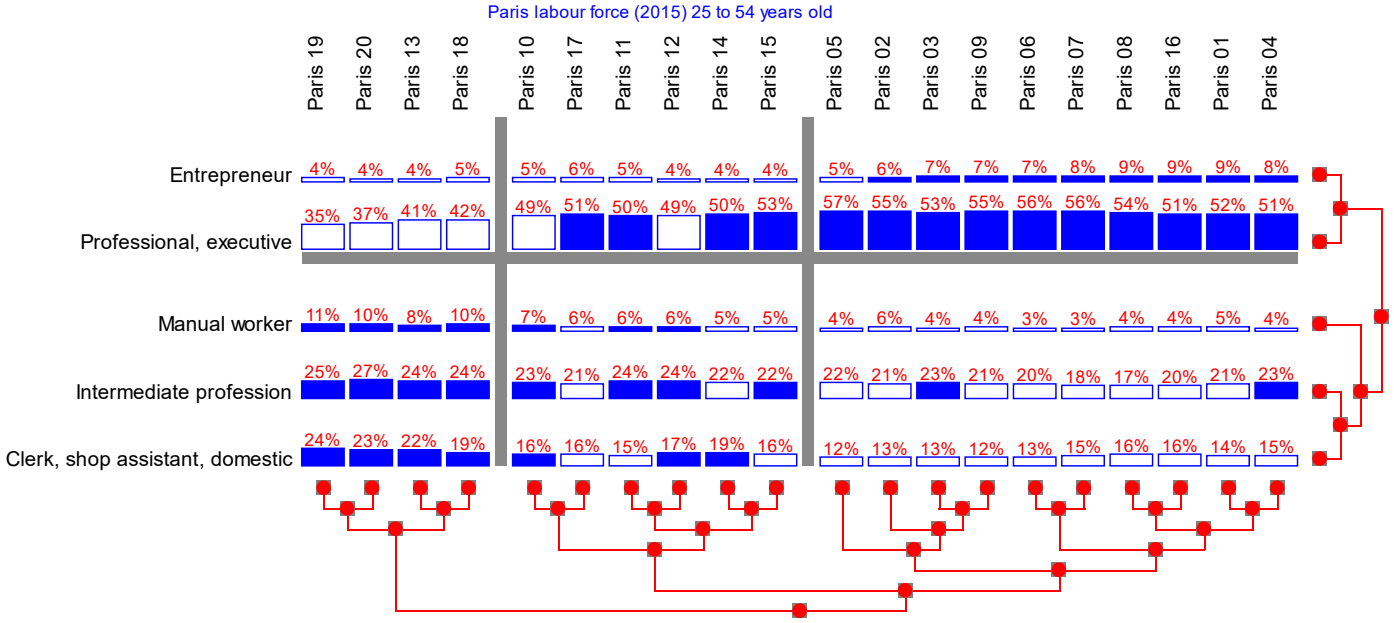
Process / Compute Column Percentage

Format / Value Format / 0%

Click on the line "Manual Worker", then **Process / Insert separator** (the insertion is done before the selected column or above the selected line).

Click on the column "Paris_10", then **Process / Insert separator**

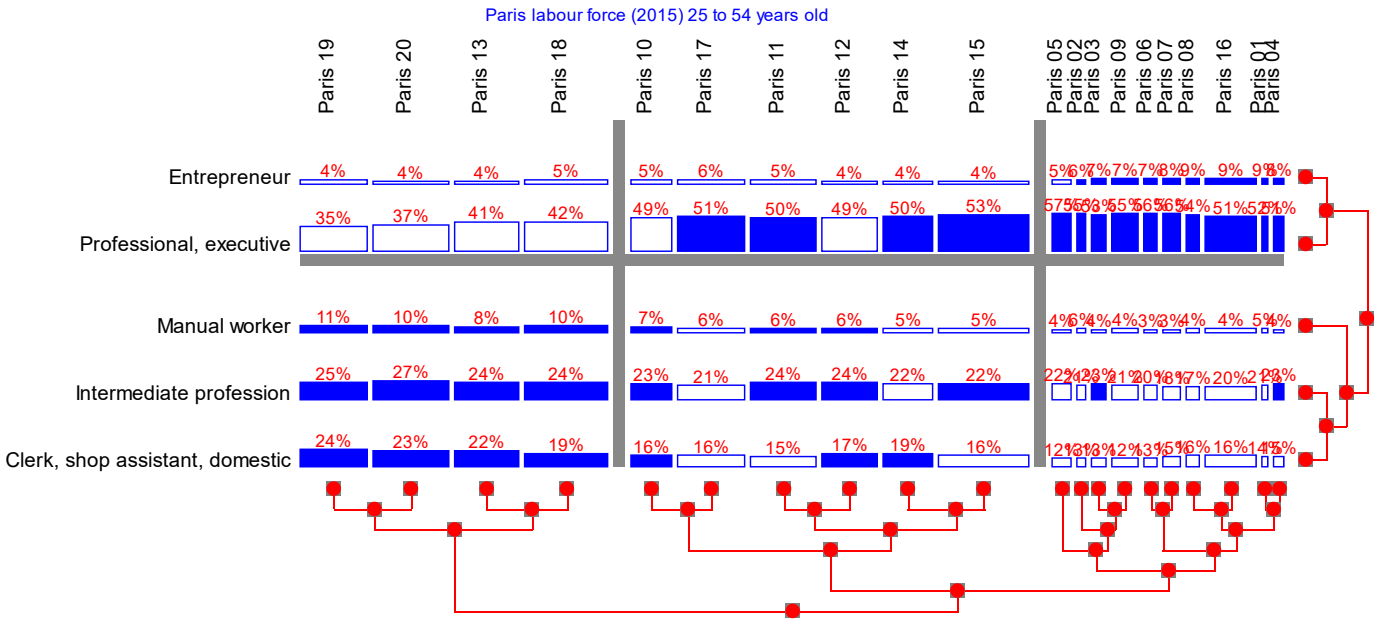
Click on the "Paris_05" column, then **Process / Insert separator**



On this graph, we can see the separation between - on the one hand, the eastern districts of Paris (19°, 20°, 13° and 18°) where relatively more manual workers, middle-class people and employees live; - on the other hand, the arrondissements of the Centre of Paris (5°, 2°, 3°, 9°, 6°, 16°, 1° and 4°) where more Managers and Senior Executives live; - and, in the middle of the graph, the intermediate arrondissements from the sociological point of view (10°, 17°, 11°, 12°, 14° and 15°). Here the numbers and heights represent the column %, i.e. the distribution of the working population aged 25 to 54 in each district.

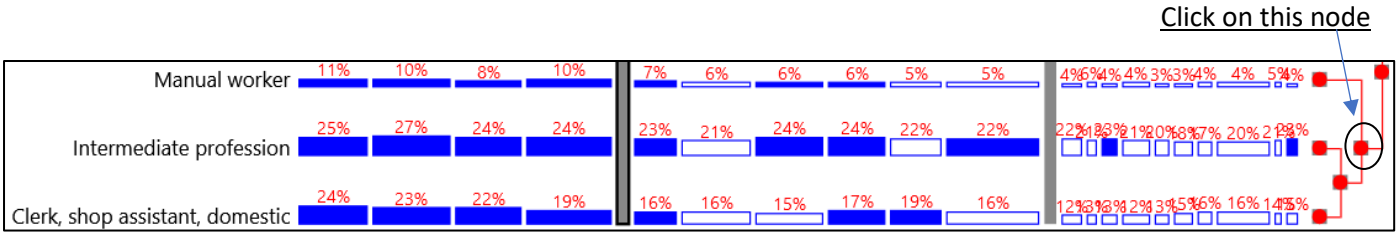
The graph can be enriched by introducing additional information: by making the column widths proportional to the sum of each column in the initial table (i.e. to the total population of 25-54 year olds in the district), the heights of the rectangles remain proportional to the % of each PCS in the district, but the surface area of the rectangles becomes proportional to the size of the sub-population concerned.

Format / Weight Width of columns

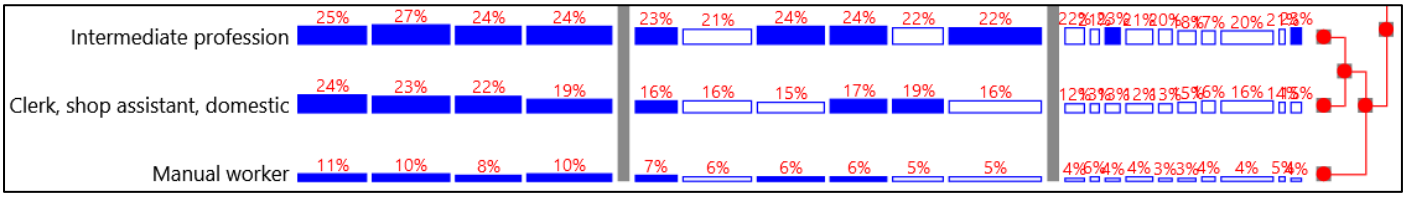


For example, in Paris_19, there are 3213 entrepreneurs, more than in Paris_09 where there are 1880. But, in proportion, there are aa% in Paris_19, less than the bb% in Paris_09. But, proportionally, there are 4% in Paris-19, less than the 7% in Paris-09.

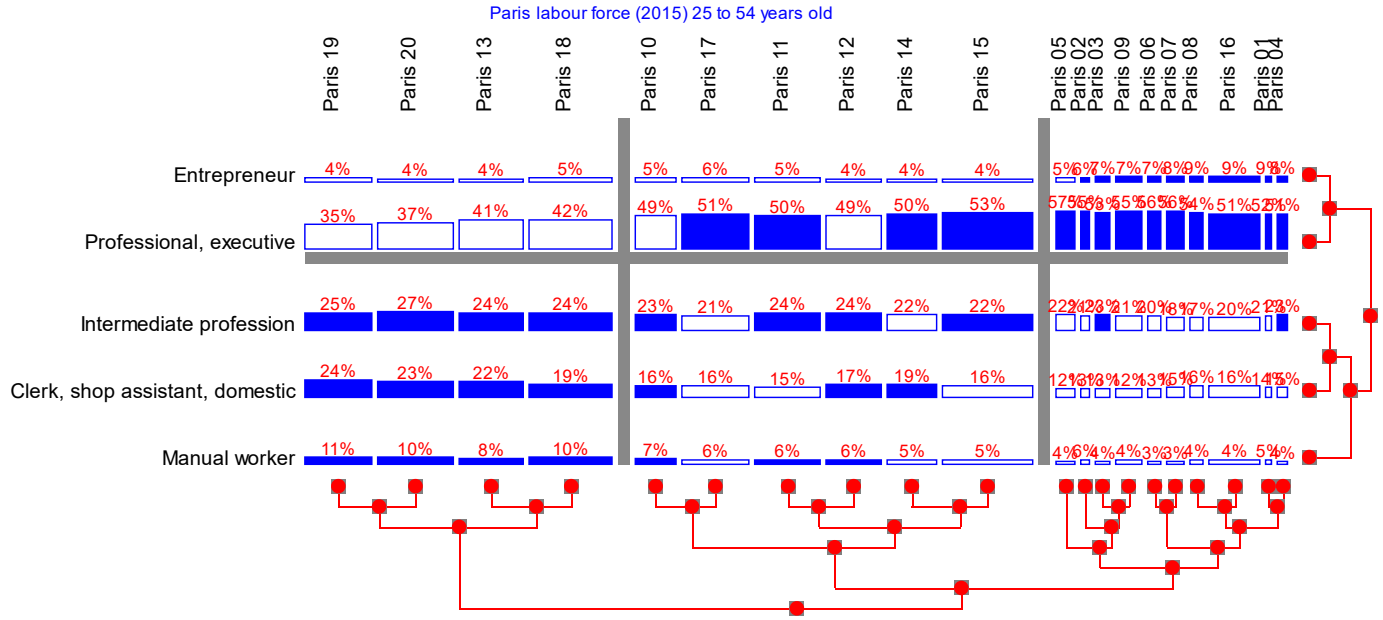
Since, at each node of the clustering tree, the order of the two classes is arbitrary, this order can be reversed by simply clicking on the small red square representing the node.



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



Finally, we get:



Example-6. Searching for blocks in a square co-occurrence matrix. A study in territory marketing

In order to attract new businesses to their territories, many towns and cities are marketing possible business location sites. These sites (brands) have various names (industrial zone, techno-park, etc.).
To help reduce the "range" of possible location names offered to businesses by local authorities, 72 business leaders were asked to group 49 cards on which were written the names of sites proposed by different French cities, with each group gathering the names that seemed synonymous to them. There were no constraints on the number and size of groups a respondent could form. Each interviewee also had the option of omitting any card bearing a site name that was unknown to him or her. The matrix gives the number of times two names have been put together[2]. It can be seen as a proximity type similarity matrix.

File / Open / Parcourir / MARKETING-Territorial.TXT

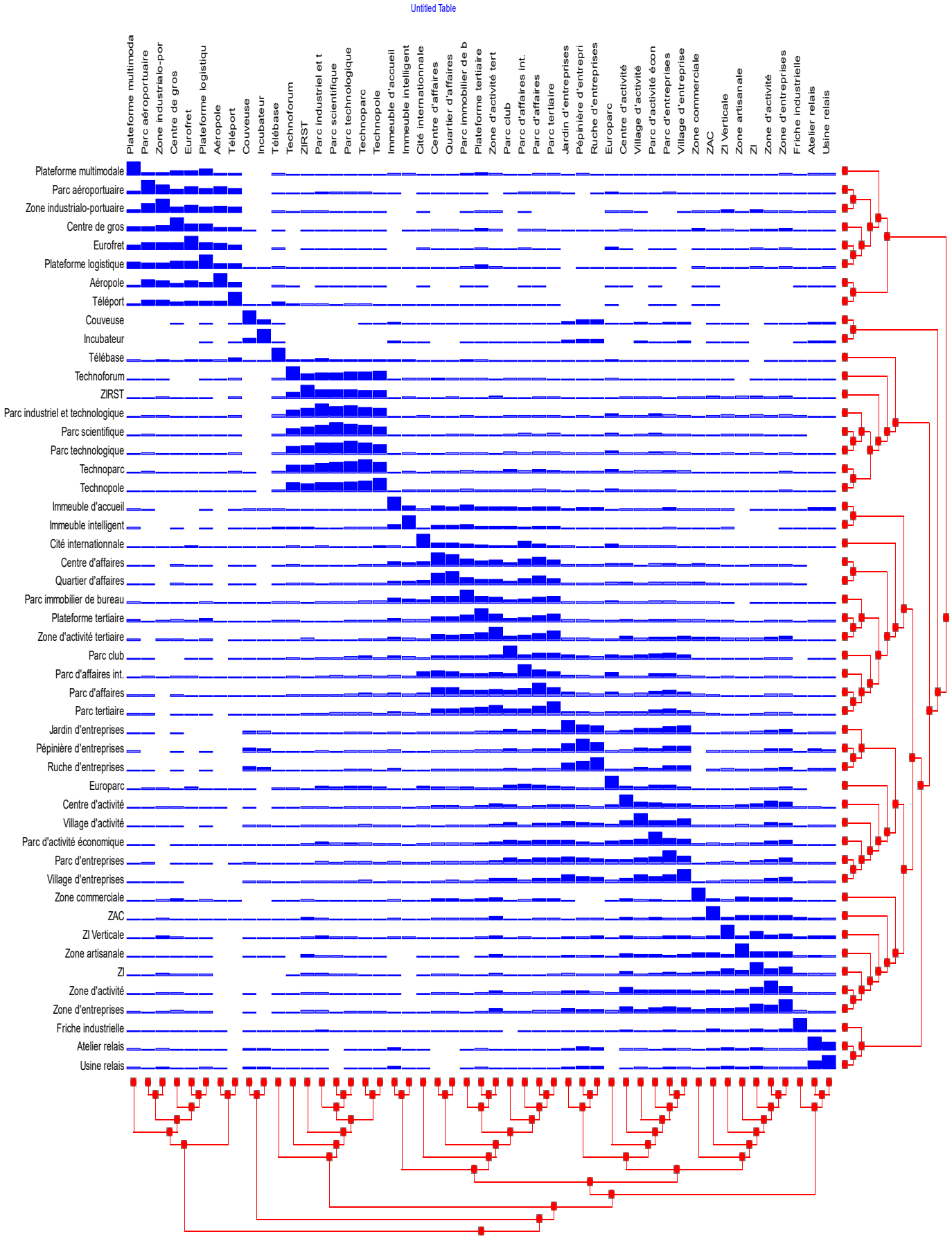
Ctrl + - => to decrease the menu font and access the last lines of the sub-menus.

Format / Value Format / None **Format / Row Legends / Complete**

Format / Column Legends / 20 **Format / Graph Size / 1438 et 1278**

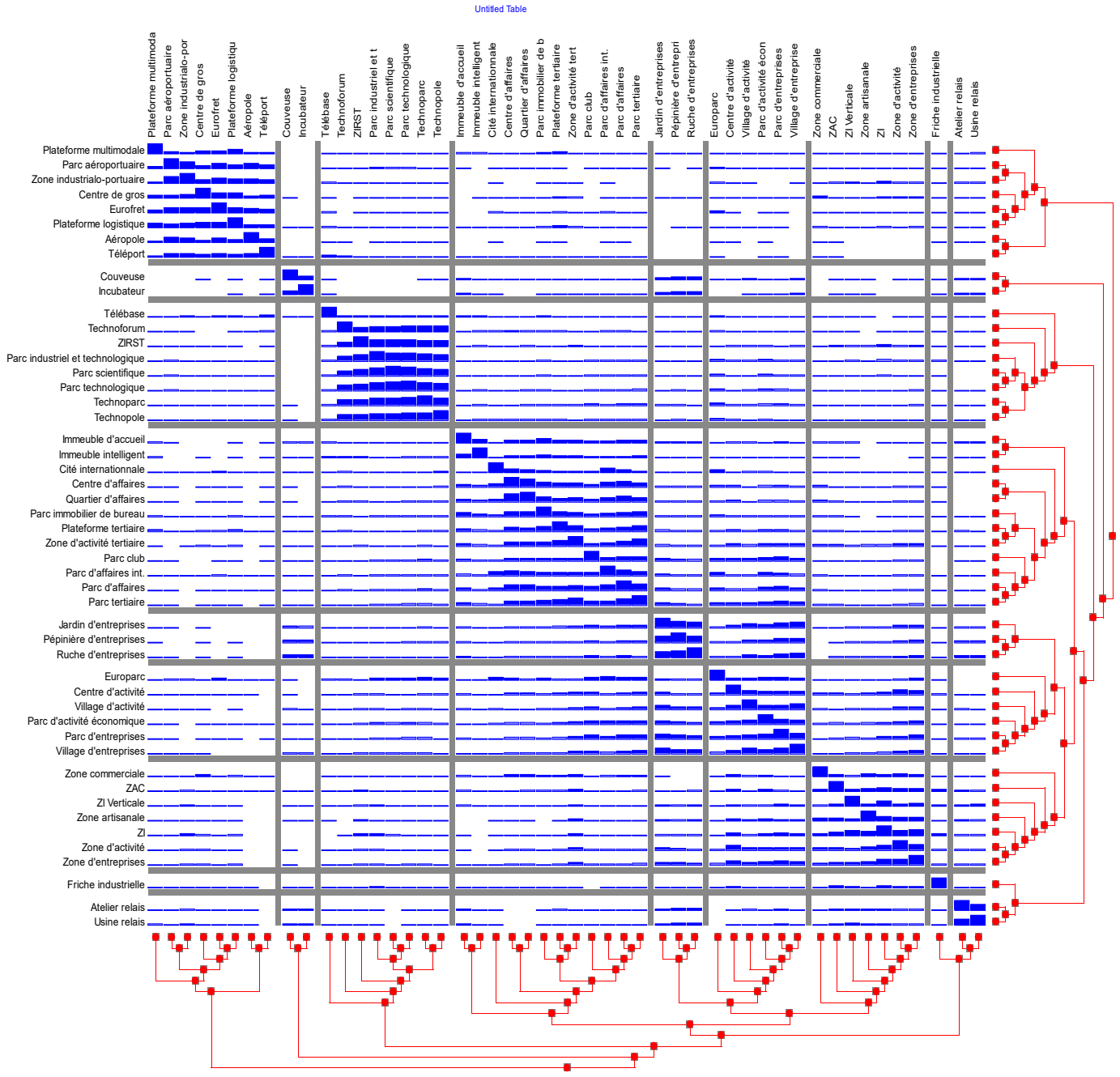
[2] TEXIER Laurence. 1999, « Une clarification de l'offre d'implantation en marketing territorial : produit de ville et offre de territoire », RERU Revue d'économie régionale et urbaine, no 5, p. 1021-1036

To find blocks of names that are often classified together as synonyms, the classification (hierarchical ascending clustering) integrated in AMADO online is used: **Process / Frequency Data / Hierarchical Clustering**



Classes are best seen by isolating them with separators:

Click on a column (or a row) to select it and **Process** / **insert a separator** between it and the previous one:



On this graph, we can see that certain groups of names are almost synonymous for the heads of companies:

- *Plateforme multimodale, Parc aéroportuaire, Zone industrialo-portuaire, Centre de gros, Eurofret, Plateforme logistique*; then slightly separated: *Aéropole, Téléport*.
- *Couveuse* (hatchery), *Incubateur*.
- *Technoforum, ZIRST (Zones d'Innovation et de Recherche Scientifique et Technique), Parc industriel et technologique, Parc scientifique, Parc technologique, Technoparc, Technopole*. It can be observed that the names *Parc Scientifique et Parc* T*echnologique* are almost confused in the minds of business leaders.
- *Jardin d'entreprises, Pépinière d'entreprises* (Business Nursery)*, Ruche d'entreprises* (Business Beehive)*,*
- Atelier relais (*Bridging Plant*) et Usine Relais (*Bridging Factory*).

Those responsible for a territory's marketing campaign will have to reduce the range of activity sites they offer and use only one name per group, while being aware that other territories use synonyms.

Homogeneous Numerical data processing: diagonalisation & clustering

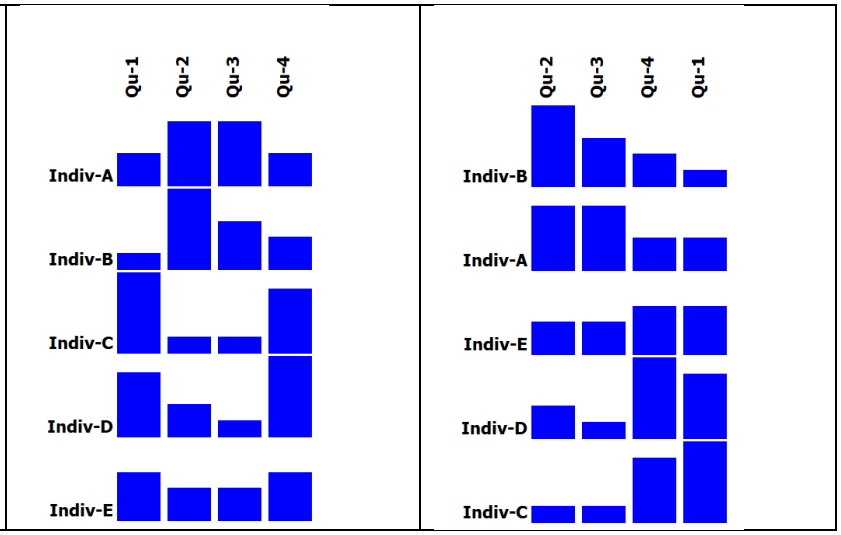
These statistical methods are suitable for tables where the columns represent variables with homogeneous units, e.g. the responses of a set of individuals to questions all using the same scale (for instance: strongly disagree=1; disagree=2; ... strongly agree=5), or unit prices, or temperatures, etc.

With "**Process / Homogeneous numerical data / Processing by PCA** (Principal Component Analysis)", the rows-individuals are reordered according to their coordinates on the first factor (i.e. the first principal component), and the columns-variables according to their correlations with this first factor.

Principal Component Analysis.

Coding:
1 = strongly disagree;
2 = disagree;
...
5 = strongly agree.

	Qu-1	Qu-2	Qu-3	Qu-4
Indiv-A	2	4	4	2
Indiv-B	1	5	3	2
Indiv-C	5	1	1	4
Indiv-D	4	2	1	5
Indiv-E	3	2	2	3

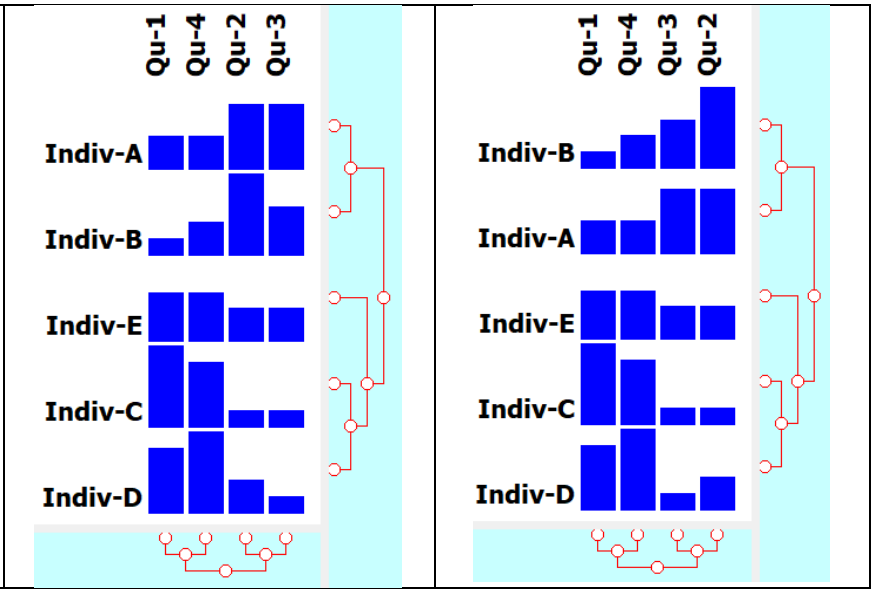


With **« Process / Homogeneous numerical data / Hierarchical Clustering », t**he lines are permuted as the nodes of the Hierarchical Ascending Classification tree and this tree is drawn to the right of the graph. As the columns are swapped as the nodes of the Hierarchical Ascending Classification tree and this tree is drawn at the bottom of the graph.

Note: since at each node of the 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**. Below, the columns Qu-2 and Qu-3, and the rows Indiv-A and Indiv-B have been swapped.

Coding:
1 = strongly disagree;
2 = disagree;
...
5 = strongly agree.

	Qu-1	Qu-2	Qu-3	Qu-4
Indiv-A	2	4	4	2
Indiv-B	1	5	3	2
Indiv-C	5	1	1	4
Indiv-D	4	2	1	5
Indiv-E	3	2	2	3



Heterogeneous Numerical data processing: diagonalisation & clustering: 24 car models according to 6 characteristics

These treatments are suitable for tables with columns representing variables of different units. Example for cars: Engine capacity in cm^3, Power (horsepower), Speed in Km/h, Weight in Kg, Width and Length in cm.

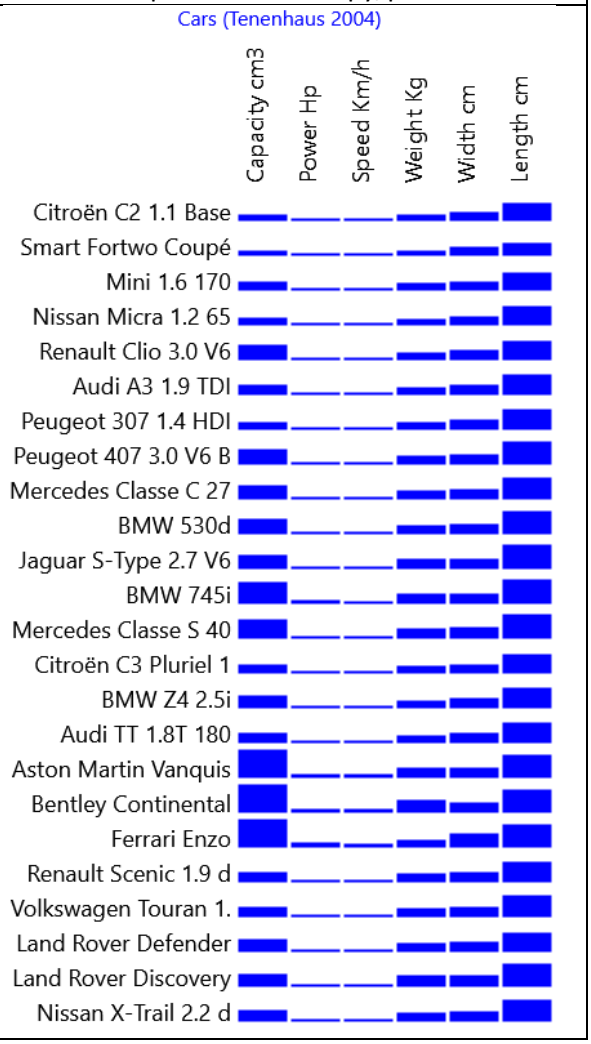
In order to avoid the choice of units, and to make them homogeneous for the calculations, each variable is "normalized". **File: EN_Cars_2004_Tenenhaus.TXT**

Format / Row Legends / Complete & Format / Column Legends / Complete
Format / Value Format / None
Typography / Increase Column Spacing (2 times)
Format / Graph Size… Graph width = 480 Graph height = 690 (Uncheck: ☑ Auto resize ➔ ☐ Auto resize **)**
Format / Mode 3
Process / Frequency Data / Hierarchical Clustering
Process / Compute Column Percentage
Format / Value Format / 0%

24 car models according to 6 characteristics

Cars (Tenenhaus 2004)	Capacity cm3	Power Hp	Speed Km/h	Weight Kg	Width cm	Length cm
Citroën C2 1.1 Base	1124	61	158	932	1659	3666
Smart Fortwo Coupé	698	52	135	730	1515	2500
Mini 1.6 170	1598	170	218	1215	1690	3625
Nissan Micra 1.2 65	1240	65	154	965	1660	3715
Renault Clio 3.0 V6	2946	255	245	1400	1810	3812
Audi A3 1.9 TDI	1896	105	187	1295	1765	4203
Peugeot 307 1.4 HDI	1398	70	160	1179	1746	4202
Peugeot 407 3.0 V6 B	2946	211	229	1640	1811	4676
Mercedes Classe C 27	2685	170	230	1600	1728	4528
BMW 530d	2993	218	245	1595	1846	4841
Jaguar S-Type 2.7 V6	2720	207	230	1722	1818	4905
BMW 745i	4398	333	250	1870	1902	5029
Mercedes Classe S 40	3966	260	250	1915	2092	5038
Citroën C3 Pluriel 1	1587	110	185	1177	1700	3934
BMW Z4 2.5i	2494	192	235	1260	1781	4091
Audi TT 1.8T 180	1781	180	228	1280	1764	4041
Aston Martin Vanquis	5935	460	306	1835	1923	4665
Bentley Continental	5998	560	318	2385	1918	4804
Ferrari Enzo	5998	660	350	1365	2650	4700
Renault Scenic 1.9 d	1870	120	188	1430	1805	4259
Volkswagen Touran 1.	1896	105	180	1498	1794	4391
Land Rover Defender	2495	122	135	1695	1790	3883
Land Rover Discovery	2495	138	157	2175	2190	4705
Nissan X-Trail 2.2 d	2184	136	180	1520	1765	4455

Graph after table copy/paste



In the graph, the heights of the rectangles are proportional to the numbers in the table, which makes no sense here; as these numbers are in different units (cm3, Hp, Km/h, cm...) they are not comparable.

For all subsequent calculations, it will therefore be necessary to normalise them (i.e. each value in the table is centred on its column mean and then divided by the column standard deviation.; we then obtain pure numbers, i.e. “without dimension”) X_{ij} -> $[(X_{ij} - m_{j}) / \sigma_j]$ In this case of "heterogeneous numerical data", the calculation procedures, Principal Component Analysis and Clustering, are performed on these "pure numbers".

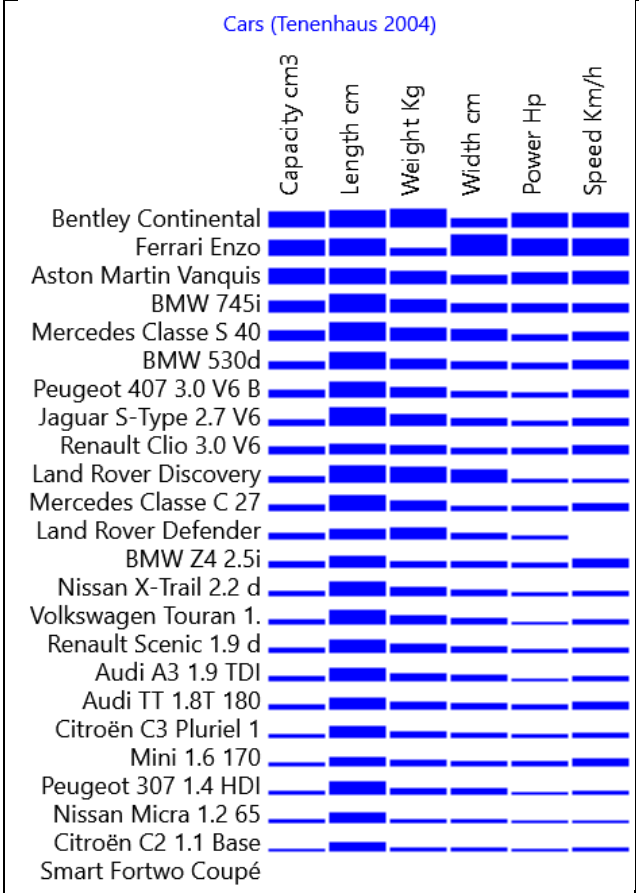
Because *AMADO online* can only represent positive numbers, we add to these "normalised" numbers their minimum in the column; the smallest value becomes zero; in our example, the "*Smart Fortwo Coupé*" is the smallest car out of the 6 variables, the 6 values become zero for it in the following graphs.

Process / Heterogeneous numerical data / Processing with normalized PCA

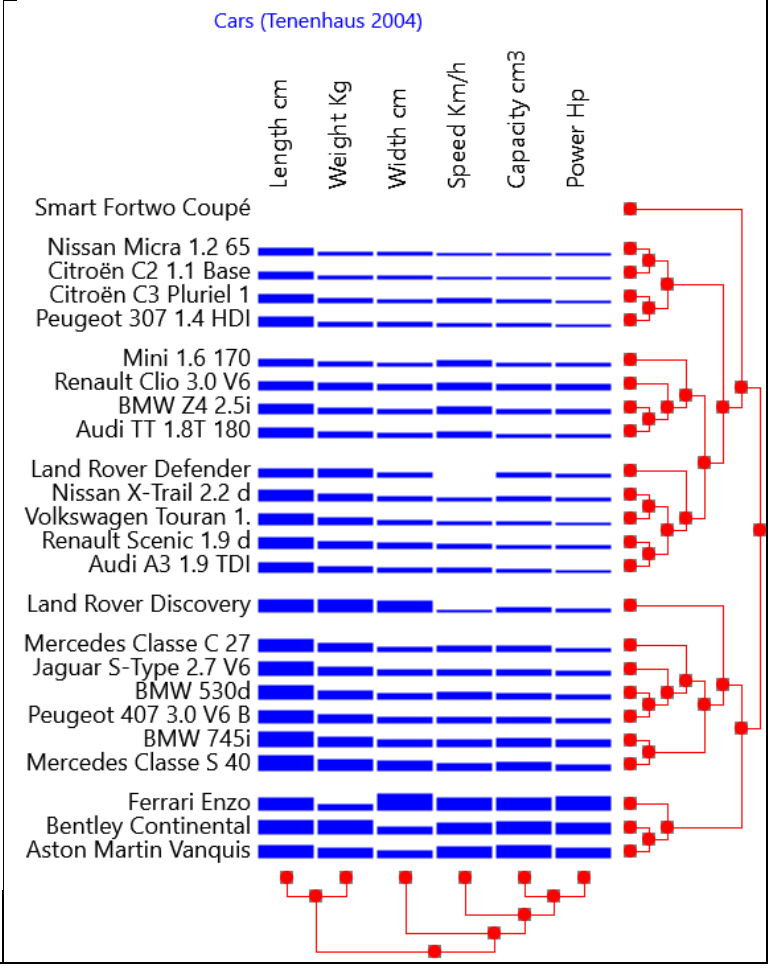
Process / Normalize Columns **Process / Heterogeneous numerical data / Hierarchical Clustering**

Format / Separators / 10 **Format / Separators / Invisible**

Graph after PCA, Principal Component Analysis



Graph after Hierarchical Clustering



<u>On the classification tree, you can see car classes:</u> - The *Smart Fortwo Coupé* is alone, clearly the smallest for all variables; - Then *Citroën C2 1.1 Base, Nissan Micra 1.2 65, Citroën C3 Pluriel 1,* and *Peugeot 307 1.4 HDI* form a homogeneous group of 4 small cars; - As are (but more sporty) *Mini 1.6 170, Renault Clio 3.0 V6, BMW Z4 2.5i* and *Audi TT 1.8T 180*; - Next comes a group of 4 large family cars *Land Rover Defender, Nissan X-Trail 2.2 d; Volkswagen Touran 1; Renault Scenic 1.9 d* and *Audi A3 1.9 TDI*; - The Land Rover Discovery is specific, being long, wide and heavy, relatively unpowerful for its size and rather slow; - Next comes a group of 6 luxury sedans, large, nervous and fast: *Mercedes Classe C 27; Jaguar S-Type 2.7 V6; BMW 530d; Peugeot 407 3.0 V6 B; BMW 745i; Mercedes Classe S 40* ; - Ultimately, cars, both large, very powerful and extremely expensive, form the last group: *Ferrari Enzo, Bentley Continental and Aston Martin Vanquis*.

As for the variables, we can see that *weight* and *size* are strongly correlated, as are *capacity* and *power* and, a little less so, *speed*.

BERTIN's Graphic

AMADO-online allows you to graphically represent a numerical cross-table and then to permute the rows (and columns) to reveal the data structure: - either a diagonal structure (seriation) if it exists, - or a class structure of the rows and columns, or even blocks.

This user's guide presents several types of tables with, for each one, the source data and the sequence of commands in the AMADO-online menus that allow the reproduced graphics to be obtained.

AMADO-online is a tool adapted to small or medium sized tables (up to fifty rows and columns) [3] such as those constructed in the Social and Hum an Sciences where each element has been precisely defined and must be clearly positioned in the overall picture.

The graphs produced by AMADO online are faithful to the data and simple to understand; they give the reader direct access to the result: each piece of information - each number in the data table - is restored in its original form, the numbers are represented by rectangles whose heights are proportional to the values in the original table, either in absolute numbers or in percentages.

The idea of permuting the rows and columns of a matrix in order to reveal a hidden structure in a data matrix is an old one: Sir W. M. Flinders Petrie (1899) presented a century ago a "sequence in prehistoric remains", i.e. a chronological " seriation " of the shapes and decorative elements of objects found during archaeological excavations in Egypt. As Phipps Arabia, Scott Boorman & Paul Levitt (1978), Giles Caraux (1984) and Jean-François Marcotorchino (1987) have pointed out, this idea is increasingly influential in applied mathematics and the cognitive sciences.

Jacques Bertin (1967, 1977) came up with the idea of putting histograms side by side, using an appropriate scale, and permuted the elements to reveal the underlying structures in the data. Since then, this approach has gained considerable momentum (Bord 1997, Palsky 2017, Harvey 2019). Originally, Bertin and his team worked with sets of wooden cubes that they moved several times by hand, first the rows, then the columns, then the rows, etc. Then, the spread of multidimensional data analysis methods (Cordier 1965, Benzécri 1973, Arabie & al. 1978, Greenacre 1984, Caraux 1984, Tenenhaus & Young 1985, Hoffman DeLeeuw 1992) somewhat overshadowed this empirical approach.

Of course, the numerical techniques of data analysis make it possible to quickly discover the main features of the structure of the table. This saves a considerable amount of time in the search for the best pair of permutations of the *n* rows and *p* columns of the table among the *n! p!* possible solutions. However, in factor analysis, lists of coordinates and other "*numerical aids to interpretation*" are useful to the statistician but often incomprehensible to the social scientist; the same applies to *factorial graphs, cloud of individuals, circle of correlations, simultaneous representation,* etc. Their interpretation requires a trained eye, and they may owe part of their success with the general public to their very esotericism ... For their part, classification trees give a useful but distorted ("*ultrametric*")[4] representation of the original table, and almost always for only one side of the table, either the rows or the columns. But many lists of *averages, marginal and conditional means, standard deviations, contributions*, etc., are necessary to clarify the meaning of such a tree.

On the contrary, as will be seen in the examples presented here, AMADO online uses factor analysis or classification and then gives the reader direct access to the result: each piece of information - each number in the data table - is

[3] The method of graphic representation proposed by Bertin has been adapted to the very large data tables by Jean Daniel Fekete and his collaborators (2015, 2016).

[4] A distance is ultra-metric if all the triangles are isosceles, the 3° side being smaller than the two equal sides. This is the case when the distance of two elements is measured by the height of the smallest node that joins them together in a hierarchical classification tree. This type of distance is very special; for example, it is impossible to place more than 3 points in a plane such that their distances (in the usual geometric sense) meet this ultrametric condition.

restored in its original form, either as an absolute number or as a percentage. It is only the order of the rows and columns that has changed, but it is all there.

Elements of bibliography

Petrie (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

Cordier Brigitte,1965, *L'Analyse des Correspondances*, Thèse de Doctorat (3° cycle), Université de Rennes.

Bertin J. 1967. *Sémiologie Graphique. Les diagrammes, les réseaux, les cartes*, Paris, La Haye, Mouton, Gauthier-Villars. 2e édition : 1973, 3e édition : 1999, EHESS, Paris.

Benzécri Jean-Paul, 1973, *L'Analyse des Données, t. I : Taxinomie ; t. II : L'Analyse des Correspondances*, Bordas, Paris (1re édition 1973, 2e édition 1976, 3e édition 1980, 4e édition 1982)

Bertin J., 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/

Arabie Ph, 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

Greenacre M. 1984. *Theory and applications of correspondence analysis*. Academic Press. https://www.ogi-nic.net/CARME-N/download/theory%20and%20applications%20of%20correspondence%20analysis.pdf

Caraux, G., 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/

Tenenhaus, M., Young, F.W. 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

Hoffman, D.L., De Leeuw, J. 1992. *Interpreting multiple correspondence analysis as a multidimensional scaling method*. Marketing Letters 3, 259–272. https://doi.org/10.1007/BF00994134

Bord J-P. 1997. *30 years of graphic semiology in honour of Jacques Bertin* *https://journals.openedition.org/cybergeo/501?lang=en*

Chauchat J-H & A. 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=YEjKNYBvUfsC&printsec=frontcover&dq=Visualization+of+Categorical+Data,+1998&hl=fr&sa=X&ved=0ahUKEwj38KiN5LrlAhUHtRoKHVF9DhMQ6AEIKzAA#v=onepage&q=Visualization%20of%20Categorical%20Data%2C%201998&f=false

Fekete J-D. & J. 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#*

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.

Palsky G. 2017. *La Sémiologie graphique de Jacques Bertin a cinquante ans* *https://visionscarto.net/la-semiologie-graphique-a-50-ans*

Harvey F. 2019. *Jacques Bertin's legacy and continuing impact for cartography*. https://doi.org/10.1080/15230406.2019.1533784

Appendix: Correspondence Analysis = Optimal Coding of Rows and Columns

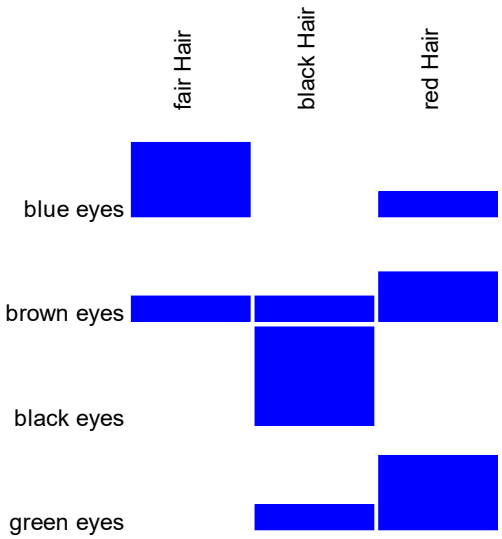
Why is Correspondence Analysis an effective tool for reordering rows and columns in a table?

Michel Tenenhaus and M., Young, F.W. showed in their 1985 article (*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) that there are multiple ways of understanding Factorial Correspondence Analysis; identical, but differently named, methods have been proposed during the development of science in different 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.

Example. Let us start with a matrix that crosses the eye and hair colours of 16 people.

This data can also be in the form of a "database" with the identifier of individuals and two qualitative fields: eye colour and hair colour.

	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**



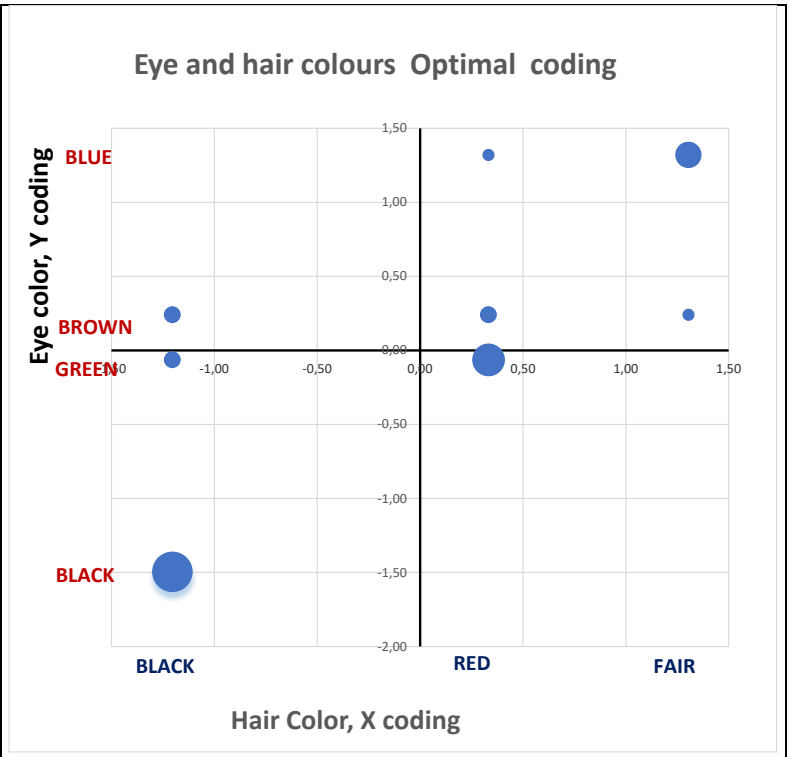
identifier	X	Y
1	brown eyes	red Hair
2	green eyes	red Hair
3	green eyes	black Hair
4	black eyes	black Hair
5	black eyes	black Hair
6	blue eyes	fair Hair
7	brown eyes	red Hair
8	brown eyes	fair Hair
9	brown eyes	black Hair
10	black eyes	black Hair
11	green eyes	red Hair
12	green eyes	red Hair
13	blue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

The numerical coding of the two qualitative variables is a set of numerical values associated with the modalities of each of the two qualitative variables. (« X coding » and « Y coding » in the table below).

Optimal coding is that which maximises the "R" correlation between the two numerical variables thus created.

Individu	X	Y	codage X	codage Y
1	brown eyes	red Hair	0,24	0,33
2	green eyes	red Hair	-0,06	0,33
3	green eyes	black Hair	-0,06	-1,20
4	black eyes	black Hair	-1,50	-1,20
5	black eyes	black Hair	-1,50	-1,20
6	blue eyes	fair Hair	1,32	1,31
7	brown eyes	red Hair	0,24	0,33
8	brown eyes	fair Hair	0,24	1,31
9	brown eyes	black Hair	0,24	-1,20
10	black eyes	black Hair	-1,50	-1,20
11	green eyes	red Hair	-0,06	0,33
12	green eyes	red Hair	-0,06	0,33
13	blue eyes	fair Hair	1,32	1,31
14	black eyes	black Hair	-1,50	-1,20
15	blue eyes	red Hair	1,32	0,33
16	blue eyes	fair Hair	1,32	1,31
		Moyenne =	0,00	0,00
		Sigma =	1	1

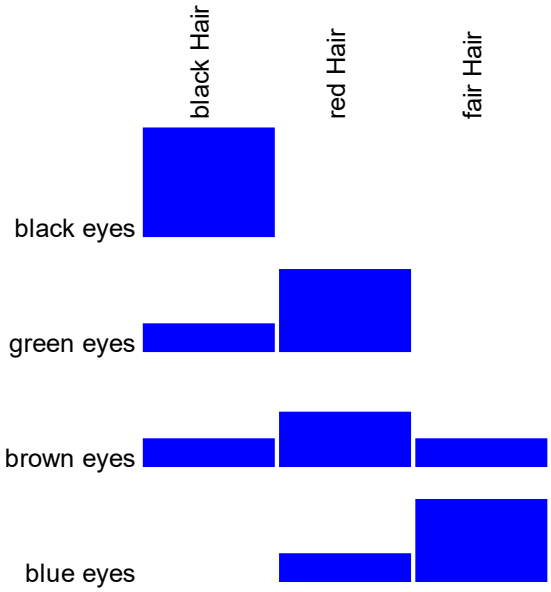
R =	0,80
R^2 =	0,65



The numerical values of the optimal coding are identical to the coordinates of the rows and columns of the table on the Correspondence Analysis 1st factor.

By reordering the rows and columns of the table according to the values of their numerical coding, we obtain the graph that best shows the structure of the data, i.e. the links between the two qualitative variables, rows and columns (the rows of the rows are reversed here) :

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



Contents

Table des matières