



# Guide de l'Utilisateur d'AMADO-online

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

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

AMADO-online a été conçu par **Nguyen-Khang PHAM** (Université de Can Tho) à partir du code d'**Alban Risson**  
Ce mode d'emploi a été écrit par **Jean-Hugues Chauchat** (Université Lyon 2) Version du 11 nov. 2020

**Débuter : Choisir la langue dans la page d'accueil, Démarrer.**

**Fichier : ouvrir un fichier de données.**

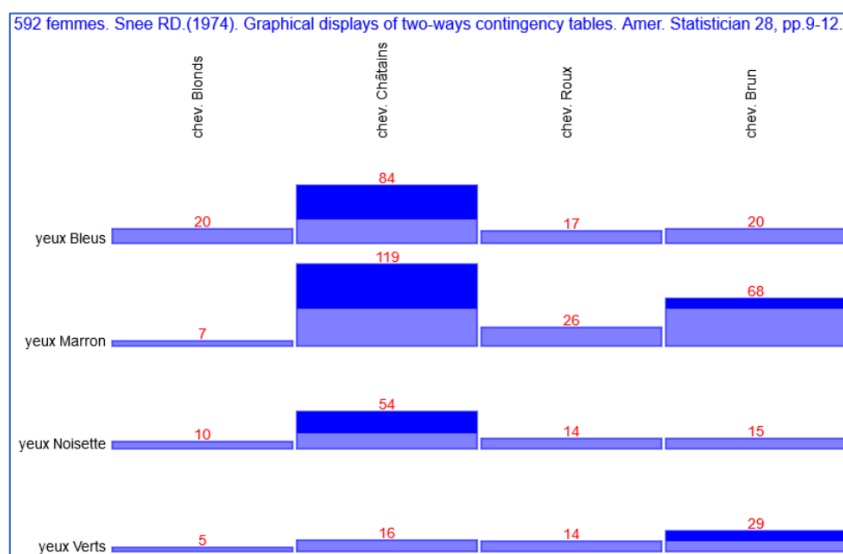
**Fichier / Ouvrir** un fichier TXT (en format UNICODE ou UTF-8, s'il y a des accents ou des caractères spéciaux, avec « séparateurs tabulations »).

Ou bien **Copier** la zone du tableau dans un tableur, puis « **Edition / Coller** » dans AMADO online <sup>1</sup>. Si la cellule en haut à gauche du tableau contient un texte, celui-ci apparaît comme titre du graphique.

**Exemple** : RD. (1974). Graphical displays of two-ways contingency tables. Amer. Statistician 28, pp.9-12

Fichier : Yeux-Cheveux.TXT

	A	B	C	D	E
1	592 femmes. Snee RD.(1974). Graphical displays of two-ways contingency tables. Amer. Statistician 28, pp.9-12.	chev. Brun	chev. Châtain	chev. Roux	chev. Blond
2	yeux Marron	68	119	26	7
3	yeux Noisette	15	54	14	10
4	yeux Verts	5	29	14	16
5	yeux Bleus	20	84	17	127

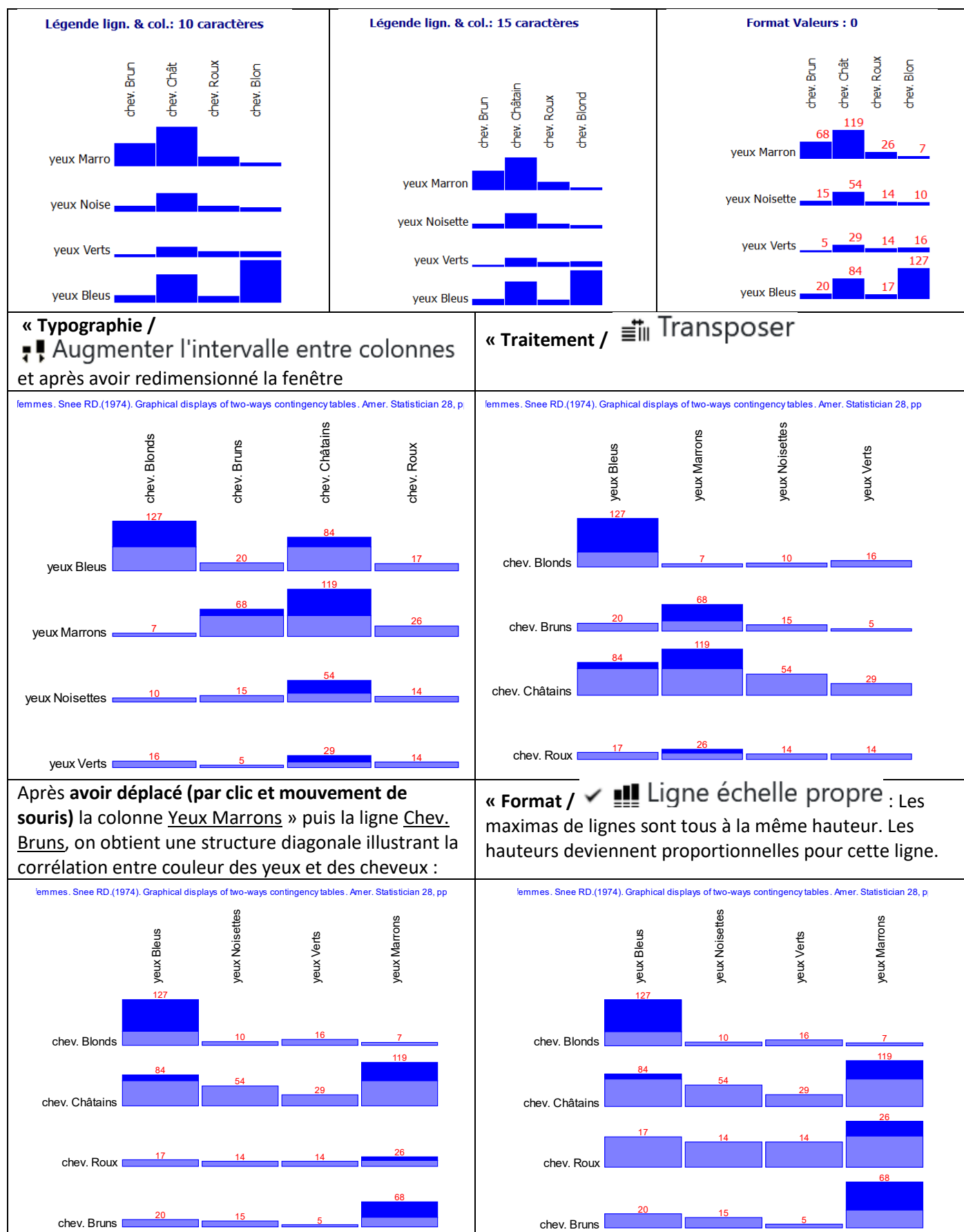


Les effectifs du tableau sont représentés par les hauteurs (et les surfaces) des rectangles.

<sup>1</sup> Si vous utilisez le navigateur Firefox, allez dans le menu en haut à droite de Firefox Personnalisez / Personnalisez la barre d'outils / Déplacez les commandes « Contrôle d'édition » vers la barre d'outils. Vous pourrez alors copier un tableau.


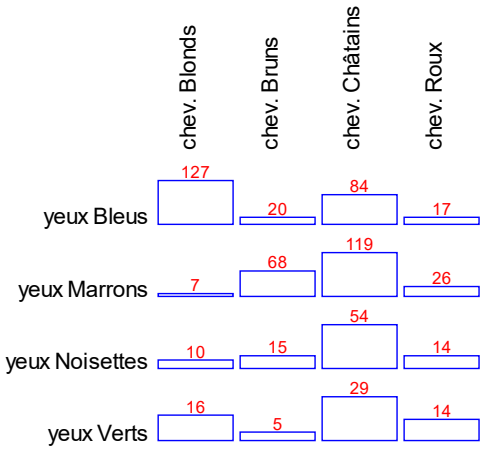
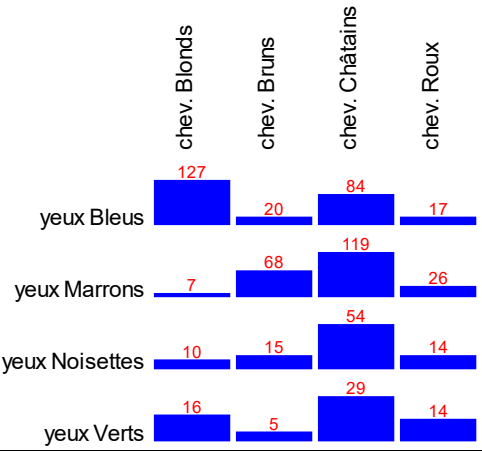

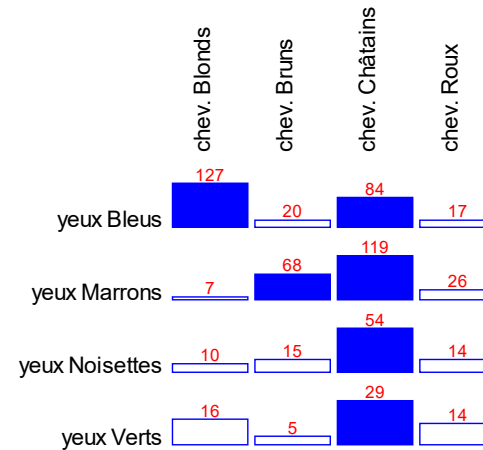
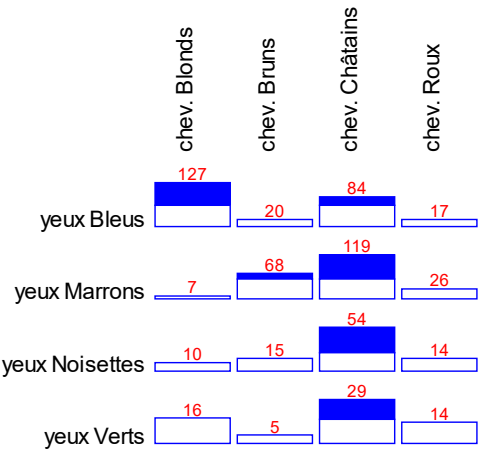

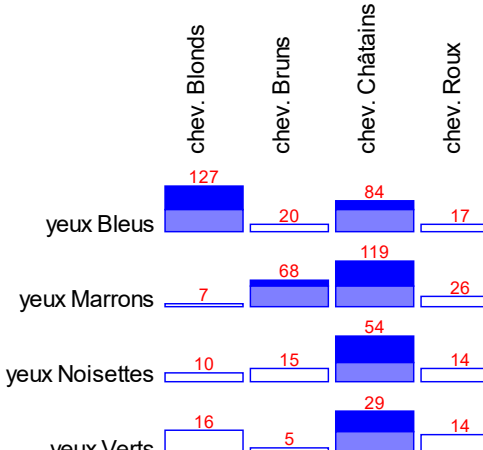
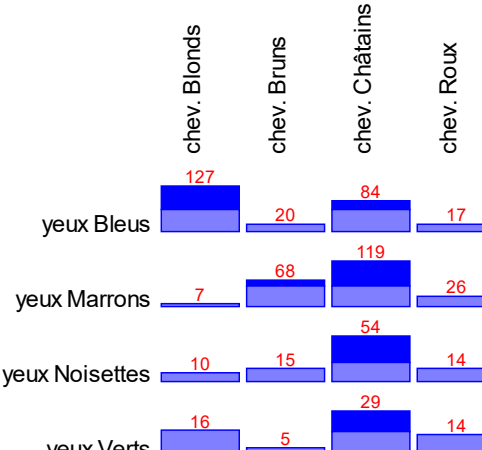
## Format : Légende lignes ; Légende colonnes ; Format valeurs

Par défaut, les libellés sont affichés avec 10 caractères. Pour afficher les libellés complets « **Format / Légende lignes / Complète** », et « **Format / Légende colonnes / Complète** ».



# Les 6 Modes d’affichage

« Format / Mode 1, etc. » permet de distinguer visuellement de différentes façons, les valeurs inférieures ou supérieures à la valeur moyenne de chaque ligne.

 <p>« Format / Mode 1, Mode 2 »</p>	<p>.(1974). Graphical displays of two-ways contingency tables. Amer.</p> 	<p>.(1974). Graphical displays of two-ways contingency tables. Ame</p> 
 <p>Mode 3 et Mode 4</p>	<p>.(1974). Graphical displays of two-ways contingency tables. Ame</p> 	<p>.(1974). Graphical displays of two-ways contingency tables. Ame</p> 
 <p>Mode 5 et Mode 6</p>	<p>.(1974). Graphical displays of two-ways contingency tables. Ame</p> 	<p>.(1974). Graphical displays of two-ways contingency tables. Ame</p> 

# Edition : Couper, Copier tableau, Coller tableau, Copier graphique.


**Annuler** (Ctrl Z) permet de revenir en arrière. **Refaire** (Ctrl Y) de rétablir

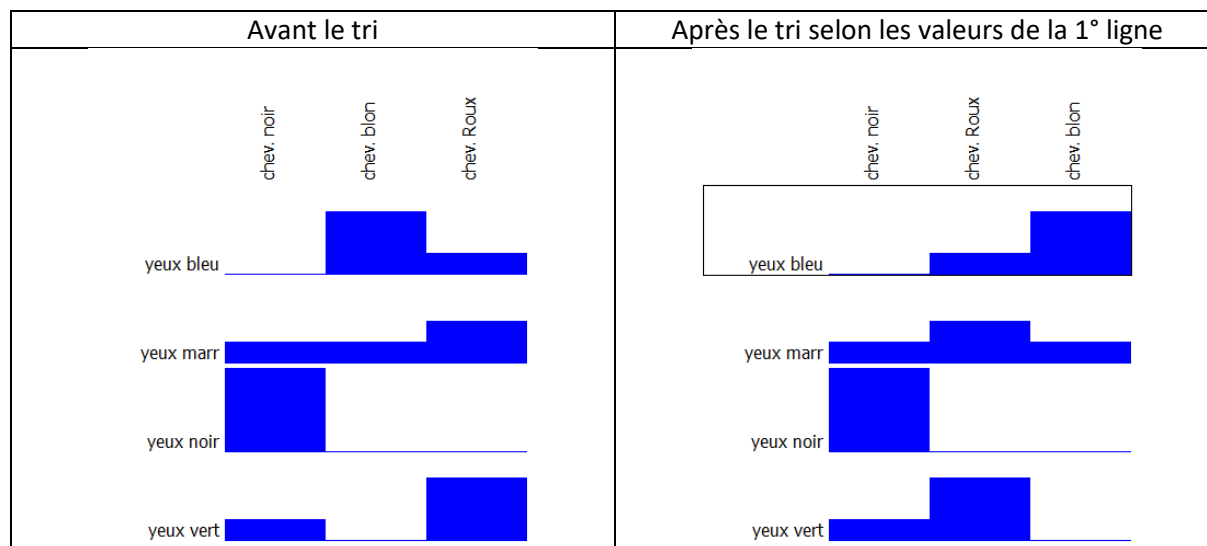
« **Edition / Supprimer les objets sélectionnés** » : Il faut d'abord sélectionner une ligne, ou une colonne, en cliquant sur le nom de la ligne ou de la colonne ; ensuite « **Edition / Supprimer les objets sélectionnés** ». Par « **Ctrl + Clic** », on peut aussi sélectionner plusieurs lignes, ou plusieurs colonnes, à supprimer.

**Copier tableau** : le tableau courant peut ensuite être collé dans un fichier Texte, Excel ou Word.

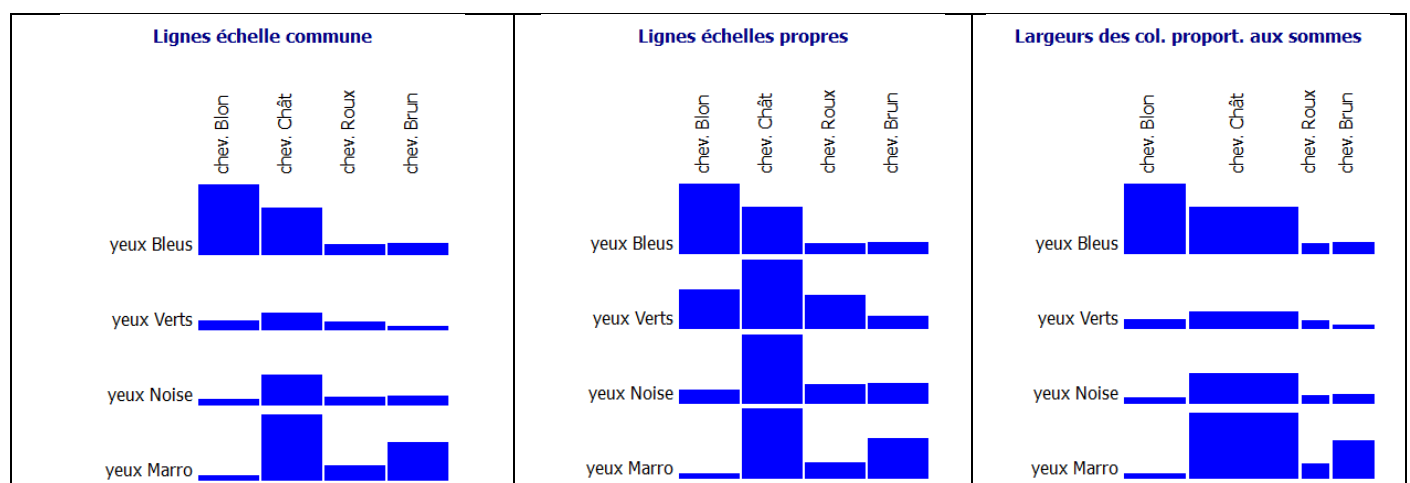
**Coller tableau** : un extrait de fichier Texte, Excel ou Word peut être collé dans AMADO online pour devenir un graphique ; c'est une alternative à Fichier / Ouvrir.

**Exporter le graphique en format SVG** : « **Fichier / Exporter en SVG** » copie le graphique dans le répertoire « Téléchargements » de l'ordinateur ; voir l'annexe expliquant comment rogner un graphique SVG qui a été collé dans Word, Excel ou PowerPoint.

 **Trier ascendant** : Il faut d'abord sélectionner une ligne (resp. colonne) en cliquant sur le nom de la ligne ou de la colonne. Ensuite, la commande « **Trier ascendant** » permet de trier toutes les colonnes, resp. lignes, dans l'ordre des valeurs de la ligne, resp. colonne, sélectionnée. « **Trier descendant** » : idem.



**Format** : Ligne échelle propre ; Ligne échelle commune ; Même largeur pour chaque colonne ; Pondérer la largeur de chaque colonne par la somme de la colonne

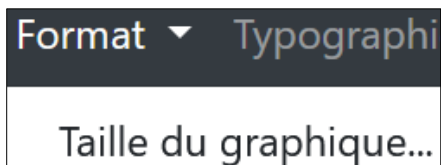


## Déplacer une ligne ou une colonne

Pour faire voir la structure des données, on peut réordonner des lignes et/ou des colonnes. On verra plus loin des méthodes automatiques de diagonalisation et de classification.

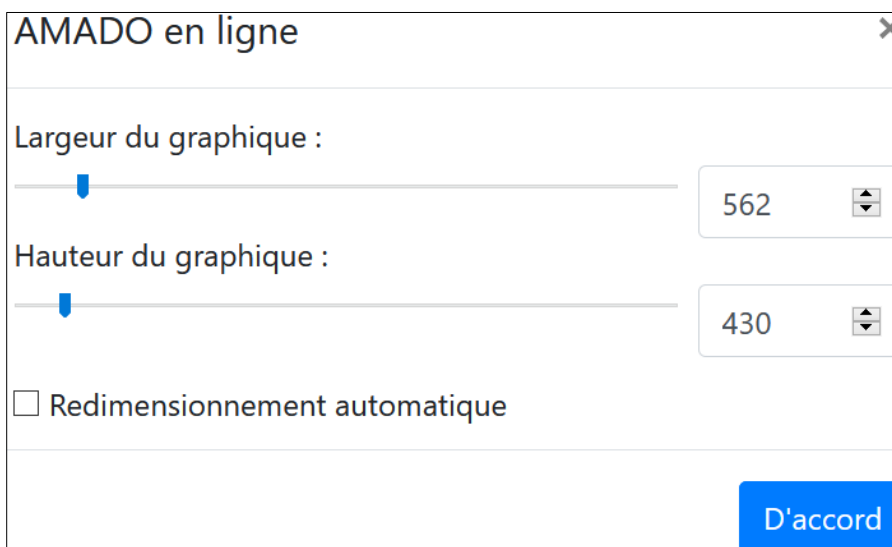
Les déplacements peuvent se faire « à la main » : sélectionner une ligne en cliquant dessus et la déplacer en conservant le bouton de souris enfoncé. On peut aussi déplacer un bloc de lignes contiguës en cliquant sur une ligne, puis sur une autre en utilisant la touche Ctrl. Idem pour une colonne ou un bloc de colonnes contiguës.

## Changer la taille, et le rapport hauteur / largeur, du graphique



Ouvrir la fenêtre, puis

- 1) **Décocher :** ☒ Redimensionnement automatique → ☐ Redimensionnement automatique
- 2) Puis choisir les largeur et hauteur souhaitées pour le graphique :



Puis valider par : D'accord

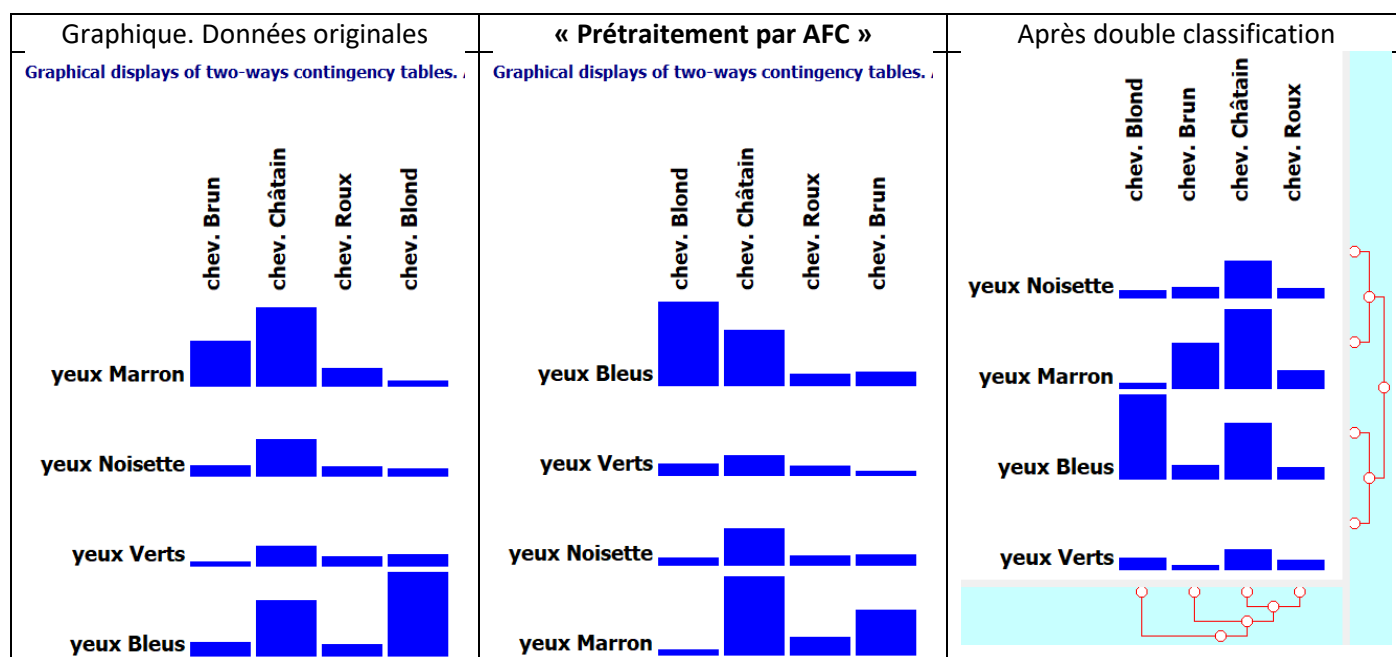
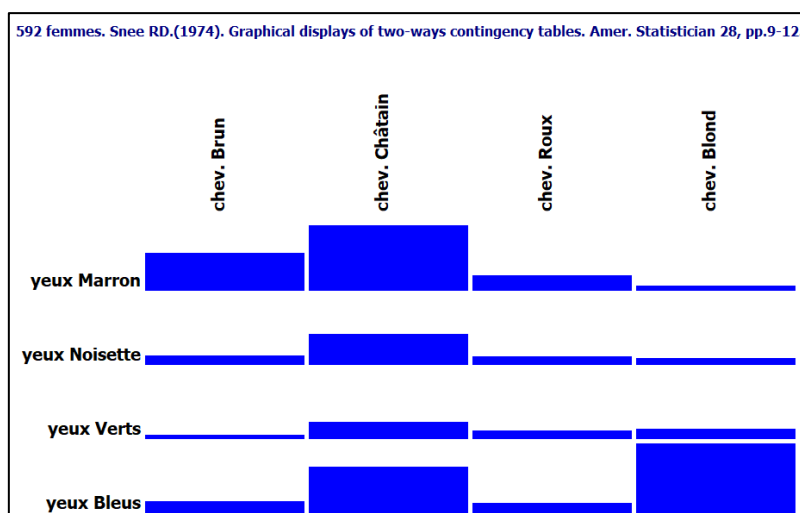
# Traitement de Données de fréquences : diagonalisation & classification

Pour tout tableau croisé de comptage, ou de présence-absence : Prétraitement par AFC (Analyse Factorielle des Correspondances) et Classification supervisée

## Exemple-1 : couleurs des yeux et des cheveux de 592 personnes

Snee (1974). Graphical displays of two-ways contingency tables. Amer. Statistician 28, pp.9-12.	chev. Brun	chev. Châtain	chev. Roux	chev. Blond
yeux Marron	68	119	26	7
yeux Noisette	15	54	14	10
yeux Verts	5	29	14	16
yeux Bleus	20	84	17	127

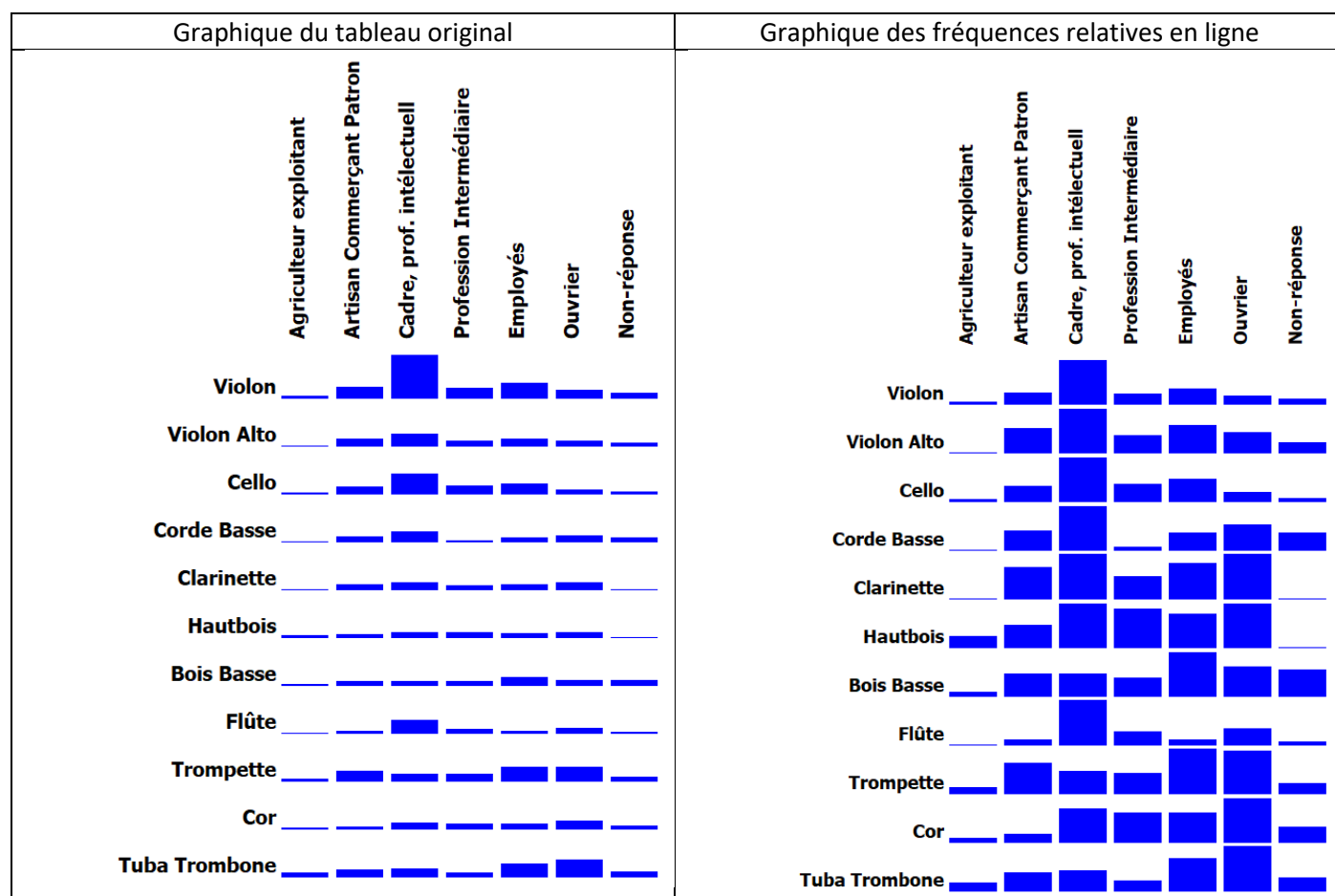
Copier le tableau depuis Excel et Coller Tableau dans AMADO online



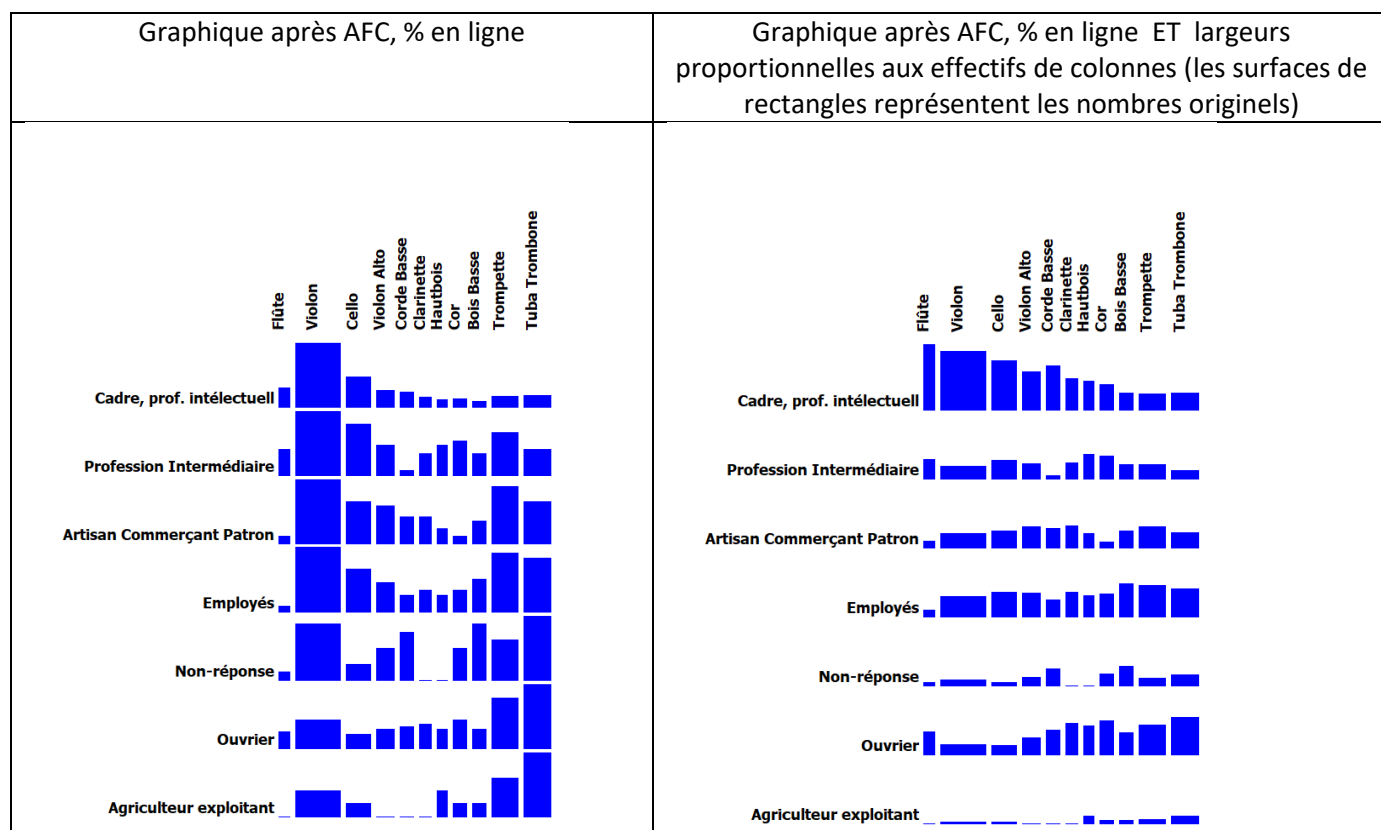
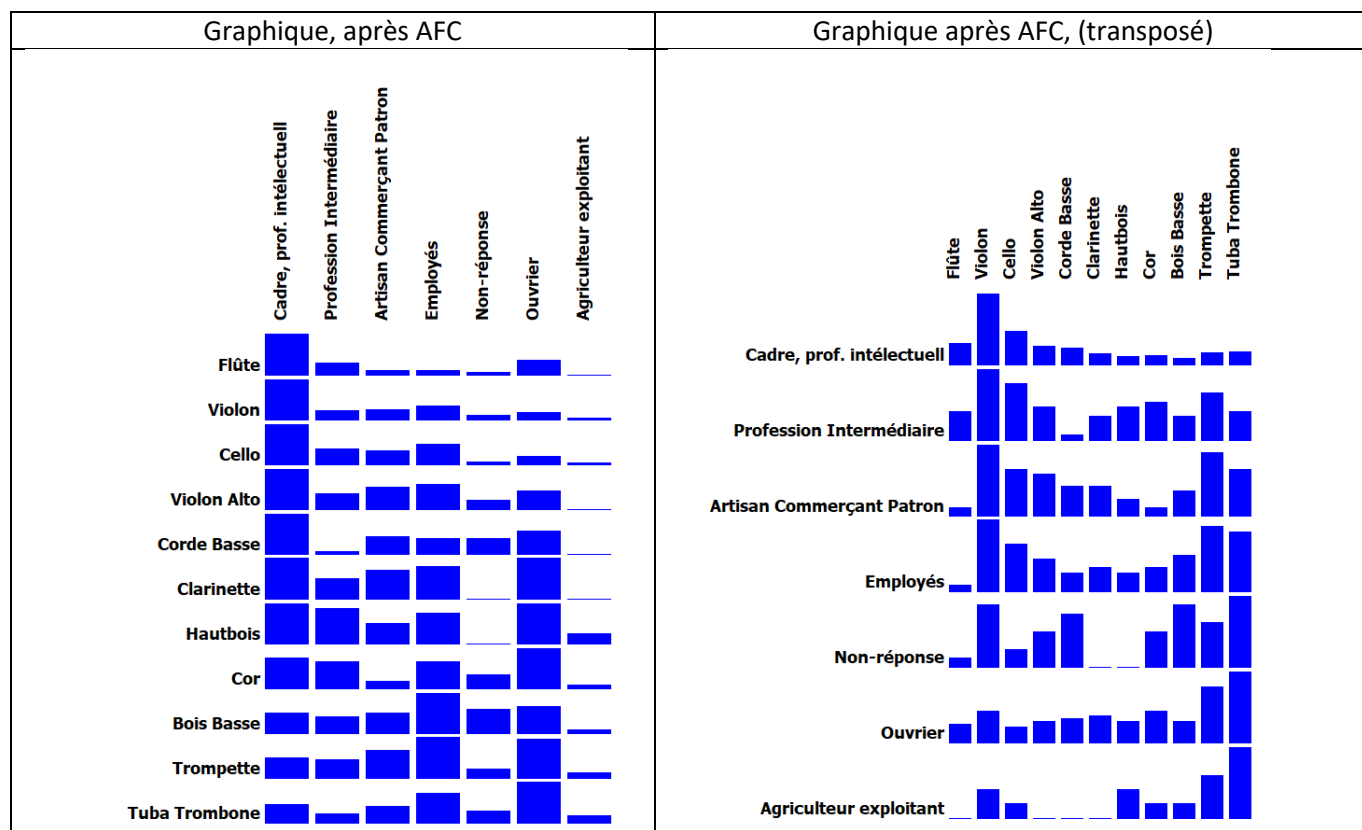
## Exemple-2 : les instruments joués par les élèves du Conservatoire National de Musique et de Danse de Paris et les professions et catégories socioprofessionnelles (PCS) des parents

L'orchestre dans tous ses éclats : sociologie de la profession de musicien, par [Bernard Lehmann](#). Thèse de doctorat en Sociologie soutenue en 1995 à [Paris, EHESS](#). **Fichier** : InstrumentsMusique-PCS\_parents.TXT

	Agriculteur exploitant	Artisan Commerçant Patron	Cadre, professions intellectuelles	Profession Intermédiaire	Employés	Ouvrier	Non-réponse
Violon	2	17	69	15	24	12	7
Violon Alto	0	10	18	7	11	8	4
Cello	1	11	32	12	16	6	2
Corde Basse	0	7	16	1	6	9	6
Clarinette	0	7	10	5	8	10	0
Hautbois	2	4	8	7	6	8	0
Bois Basse	1	6	6	5	12	8	7
Flûte	0	2	20	6	2	7	1
Trompette	3	15	11	10	22	21	5
Cor	1	2	9	8	8	12	4
Tuba Trombone	5	11	12	6	20	27	8



Après Analyse Factorielle des Correspondances (AFC) :



Avec le graphique de droite, on visualise bien que la flûte et les instruments à cordes sont plus choisis par les enfants des catégories aisées et « cultivées ». À l’opposé, les cuivres et des bois sont plus joués par les enfants d’employés et d’ouvriers (ces instruments sont joués dans les harmonies et les fanfares grâce auxquelles ils ont souvent commencé leur éducation musicale) ; les enfants d’agriculteurs sont peu nombreux au CSNS et leur part dans les joueurs de flûte ou de cordes sont nulles ou très faibles.



## Exemple-3 : « Sériation » chronologique d'objet archéologiques

Exemple de données inspirées par celles utilisées par l'archéologue Sir Flinders PETRIE pour dater des tombes fouillées à Diospolis Parva en Égypte à la fin du XIX<sup>e</sup> siècle. Il avait fait l'hypothèse que le type d'objets et les éléments de décor caractérisaient leur époque et que, en conséquence, leurs variations traduisaient la chronologie.

Dans leur livre *Archaeology: Theories Methods and Practice*, 1991, Renfrew, C., & Bahn, P. G. donnent l'exemple pédagogique ci-dessous [https://en.wikipedia.org/wiki/Seriation\\_\(archaeology\)](https://en.wikipedia.org/wiki/Seriation_(archaeology))

**Fichier :** Egyptian\_pottery\_example.TXT

Matrice brute de présence/absence des éléments de décors sur les poteries : A, B, C ...								Matrice rendue diagonale par permutation de lignes et des colonnes.							
	A	B	C	D	E	F	G		C	B	D	G	A	E	F
beaker	0	0	0	0	1	1	0	pointed	1	1	1	0	0	0	0
blackrim	1	0	0	0	1	1	0	flatpot	0	1	1	0	0	0	0
bottle	1	0	0	0	1	0	0	spirals	0	0	1	1	0	0	0
flatpot	0	1	0	1	0	0	0	handle	0	0	1	1	1	0	0
handle	1	0	0	1	0	0	1	bottle	0	0	0	0	1	1	0
pointed	0	1	1	1	0	0	0	blackrim	0	0	0	0	1	1	1
spirals	0	0	0	1	0	0	1	beaker	0	0	0	0	0	1	1

Le 2<sup>e</sup> graphique est obtenu, en un clic, par les commandes :

### Traitement / Données de fréquences / Classification

Il montre les poteries (C, B, D, G, A, E et F) réordonnées, selon un axe de présence-absence des éléments de décor repérés par l'archéologue ; cet ordre correspond probablement à l'ordre chronologique (direct ou inverse) d'invention puis d'abandon de ces créations artistiques.

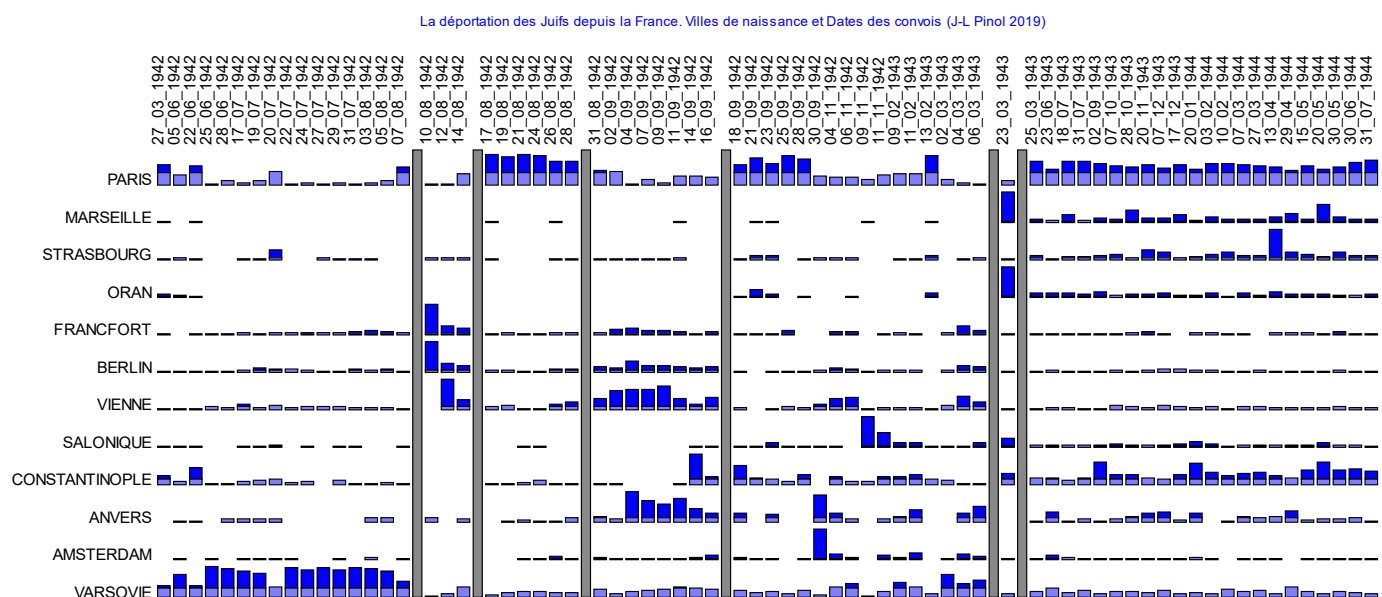
## Exemple-4 Données chronologiques : répartition des Juifs déportés depuis la France selon les convois et les villes de naissance

Dans cet exemple, on ne modifie pas l'ordre chronologique ; le graphique produit par AMADO online visualise le tableau de donnée et soutient le commentaire de l'historien.

Le graphique représente la répartition des déportés vers les camps de la mort nés dans un sous-ensemble représentatif de villes, selon les convois (% en colonnes) partis de France. Voir Jean-Luc Pinol, *Convois, La déportation des Juifs de France*, Paris, Éditions du Détour, 2019.

Fichier : Deportation\_Villes-naissance\_Convois.TXT

Traitements / Calculer les pourcentages en colonne



Sur ce graphique, on voit clairement que les 19 premiers convois ont déporté majoritairement des juifs réfugiés en France pour fuir le régime Nazi, natifs de Varsovie jusqu'au départ du 7 août 1942, puis d'Allemagne et d'Autriche les 10, 12 et 14 août 1942

Les 6 convois suivants, du 17 au 28 août 1942, ont emporté les juifs nés à Paris arrêtés lors de la rafle du Vel d'Hiv.

Les Juifs originaires d'Allemagne, d'Autriche et de Belgique, souvent réfugiés en Zone Non Occupée, ont été arrêtés en masse lors de la grande rafle du 26 août et représentent la grande majorité des huit convois partis entre le 31 août et le 16 septembre 1942.

Ensuite, l'objectif est de remplir les trains pour atteindre les objectifs fixés.

Le convoi parti le 23 mars 1943 est spécifique ; ce sont en majorité des juifs nés à Marseille ou en Algérie, arrêtés après la destruction du quartier du Vieux Port de Marseille en janvier 1943, qui ont été déportés ce jour-là vers les camps de la mort.

## Exemple-5 Les actifs de 25 à 54 ans à Paris en 2015, par PCS et par arrondissement

Fichier : Fr Paris2015\_Arrond-PCS\_25-54ans.TXT

Paris (2015) actifs 25-54ans	Chefs d'entreprise					Cadres-Supérieurs					Cadres-Moyens					Employés				
Ouvriers																				
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															

Fichier / Ouvrir/ Parcourir / Paris2015\_Arrondissements-PCS\_25-54ans.TXT

Traitement / Transposer / Légende ligne / Complète

Format / Taille du graphique : Largeur=950 ; Hauteur=400 (décocher le Redimensionnement automatique)

Typographie / Augmenter l'intervalle entre colonnes (deux fois)

Format / Mode 3

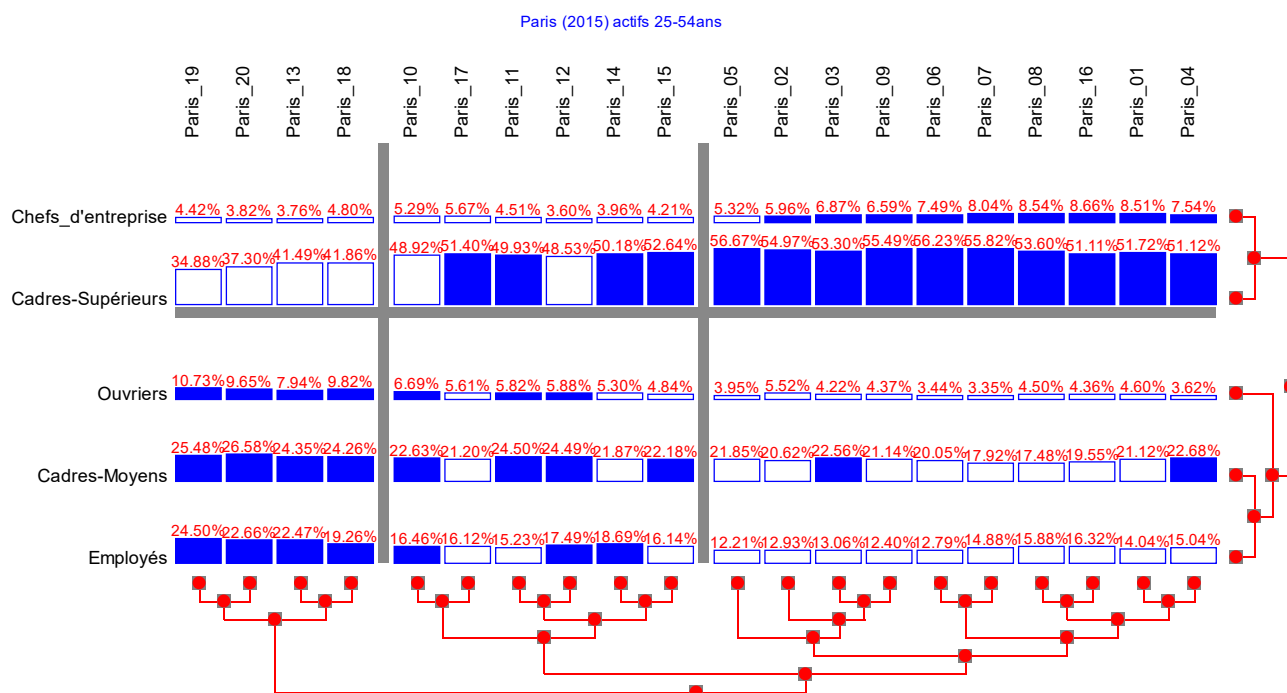
Traitement / Données de fréquences / Classification

Traitement / Calculer les % colonnes

Cliquer sur la ligne « Ouvriers » , puis Traitement / Insérer séparateur

Cliquer sur la colonne « Paris\_10 » , puis Traitement / Insérer séparateur

Cliquer sur la colonne « Paris\_05 », puis **Traitement / Insérer séparateur** (l'insertion se fait avant la colonne sélectionnée, ou au-dessus la ligne sélectionnée)

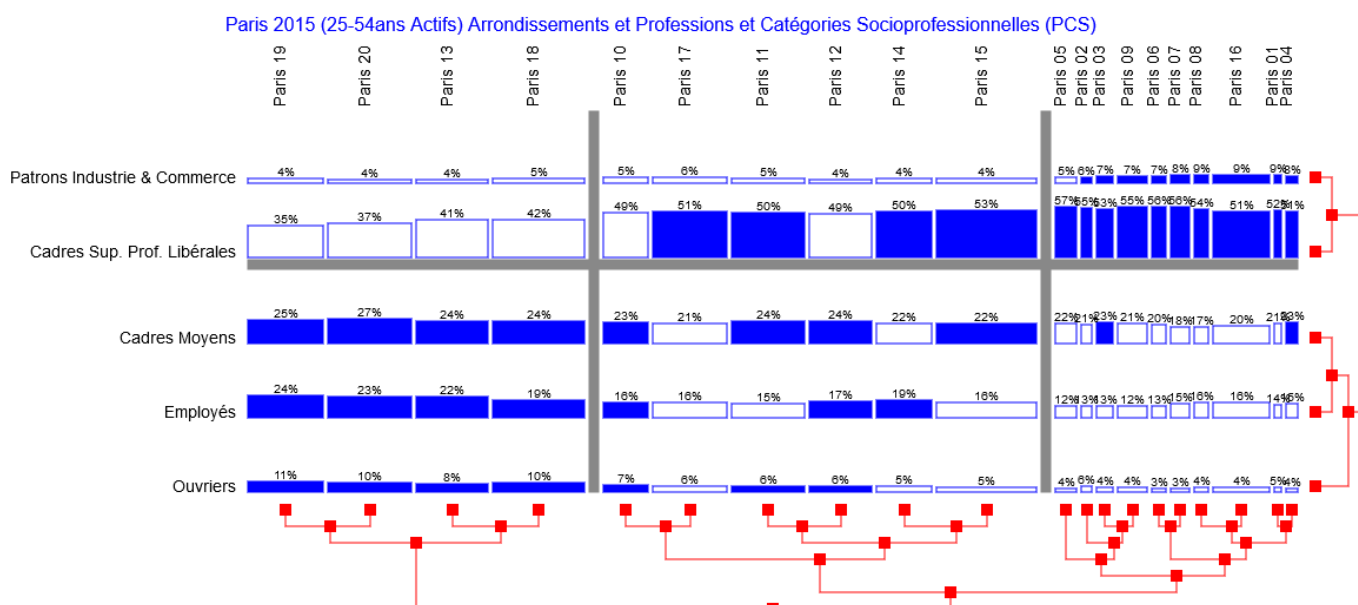


Sur ce graphique, on voit la séparation entre - d'un côté, les arrondissements de l'Est de Paris (19°, 20°, 13° et 18°) où habitent relativement plus des Ouvriers, des Cadres Moyens et des Employés ; - de l'autre côté, les arrondissements du Centre de Paris (5°, 2°, 3°, 9°, 6°, 16°, 1er et 4°) où habitent plus des Chefs d'entreprises et des Cadres supérieurs ; - et, au milieu du graphique, les arrondissements intermédiaires du point de vue sociologiques (10°, 17°, 11°, 12°, 14° et 15°). Ici les nombres et les hauteurs représentent les % en colonne, c'est-à-dire la répartition des actifs de 25 à 54 ans dans chaque arrondissement.

On peut enrichir le graphique et introduisant une information supplémentaire : les hauteurs des rectangles restent proportionnelles au % de chaque PCS dans l'arrondissement ; et, en plus, la surface des rectangles devient proportionnelle à l'effectif de la sous-population concernée.

**Format / Pondérer la largeur des colonnes par la somme de chaque colonne**

**Format / Format valeurs / 0%**



## Exemple-6 Recherche de blocs dans une matrice carrée de cooccurrence. Étude en marketing des territoires

Pour attirer les entreprises sur leurs territoires, de nombreuses villes commercialisent des sites possibles d'implantation d'activités. Ces sites (marques) portent des noms divers (zone industrielle, parc technologique, ...).

Pour aider à réduire la « gamme » des lieux d'implantations possibles offertes aux entreprises par les collectivités locales, on a demandé à 72 chefs d'entreprise de regrouper en tas 49 cartons sur lesquels étaient inscrits les noms de sites proposés par différentes ville française, chaque tas réunissant les noms qui lui paraissaient synonymes. Aucune contrainte n'a été imposée quant au nombre et à la taille des groupes qu'un répondant pouvait former. Chaque interviewé avait également la possibilité d'omettre toute carte portant un site industriel qui lui était inconnu. La matrice donne les nombres de fois où deux lieux ont été classés ensembles <sup>2</sup>. Elle peut être considérée comme une matrice de similarité de type proximité.

Fichier / MARKETING-Territorial.TXT

Ctrl & « - » pour diminuer la police des menus et accéder aux dernières lignes des sous-menus

Format / Valeurs / None

Format / Légende ligne / Complète

Format / Légende Colonne / 20

Format / Taille du graphique / 1438 et 1278

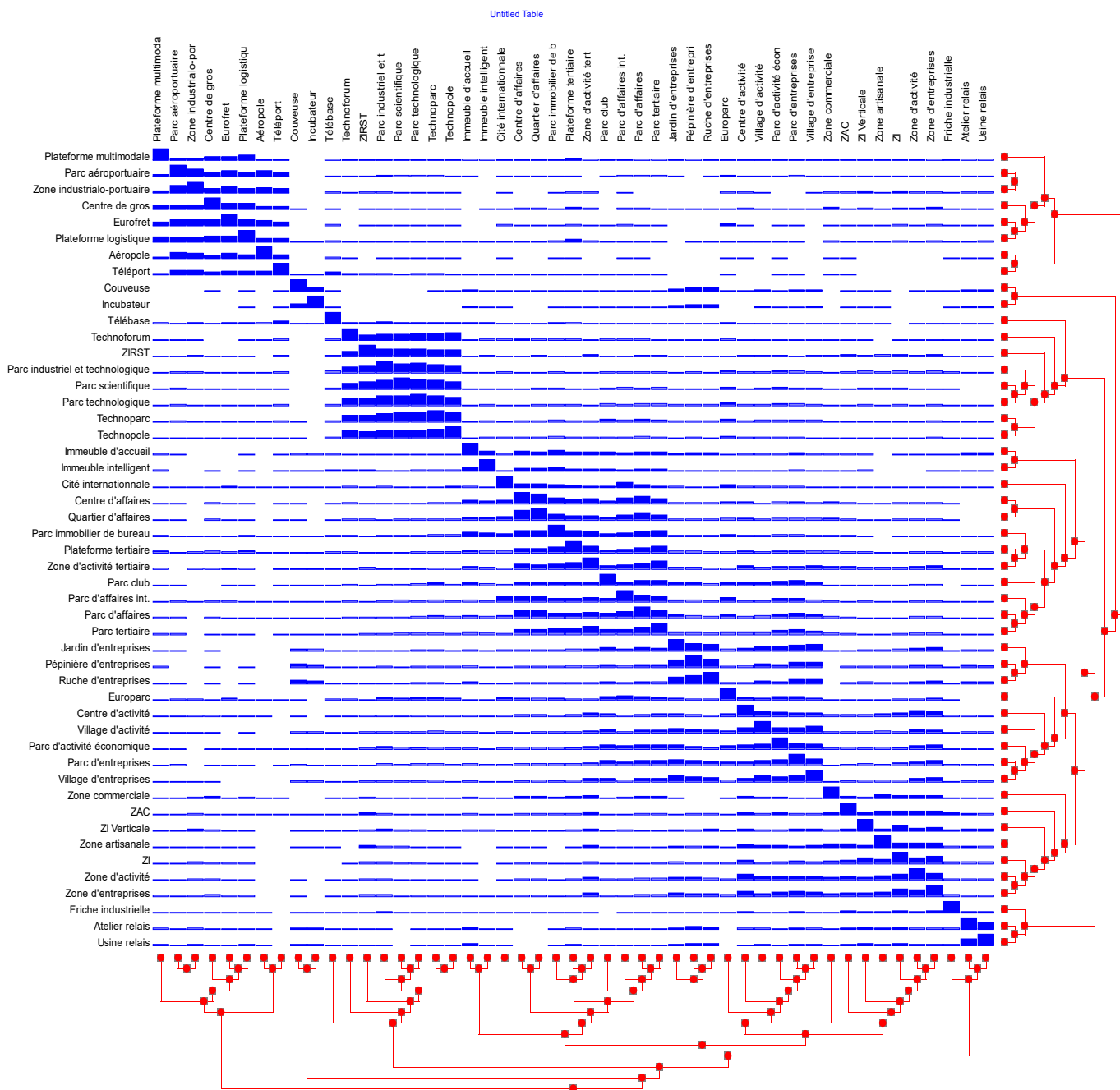
Untitled Table



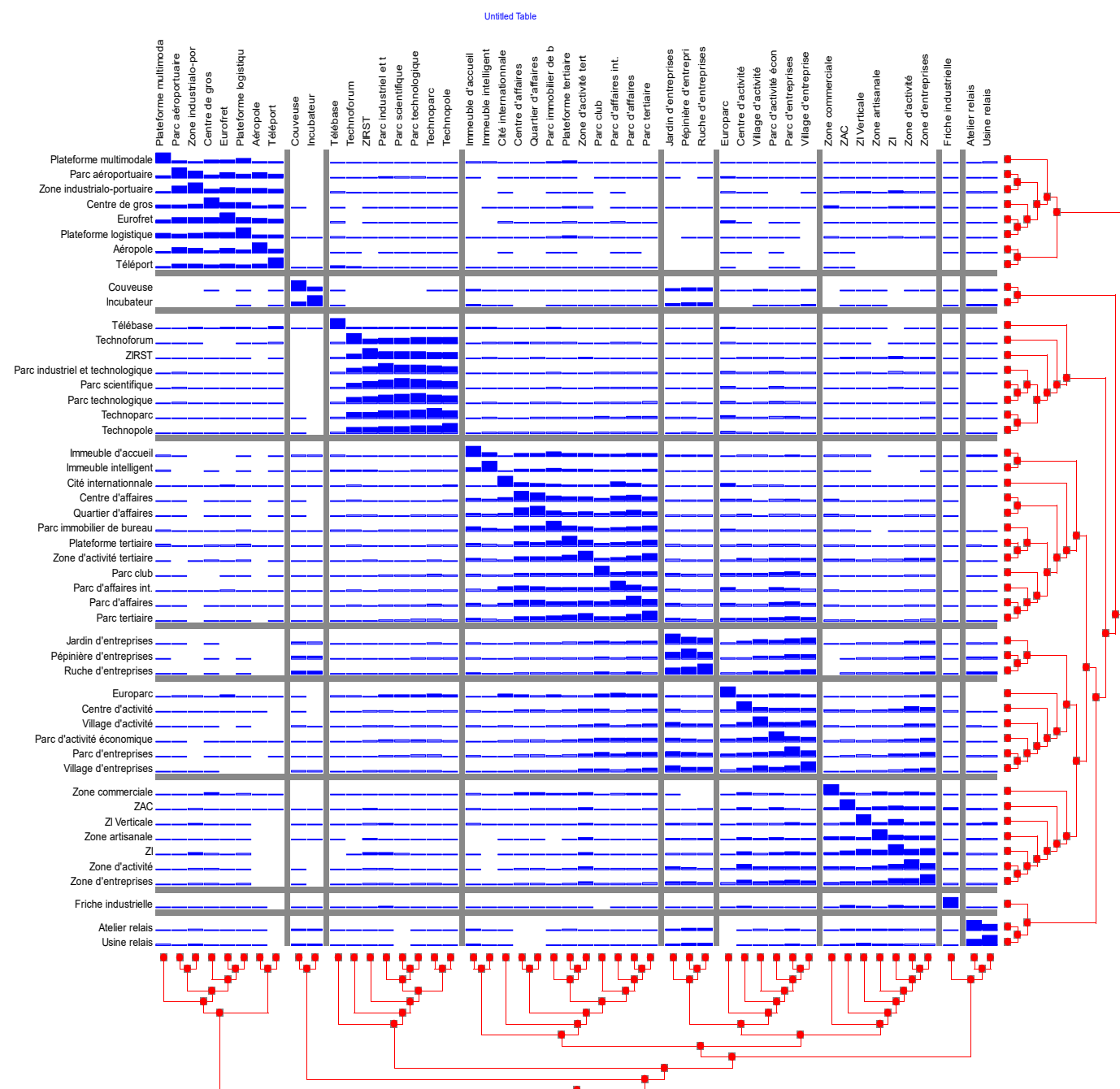
<sup>2</sup> TEXIER Laurence. 1999, « Une clarification de l'offre d'implantation en marketing territorial : produit de ville et offre de territoire », RERU Revue d'économie régionale et urbaine, no 5, p. 1021-1036

Pour trouver les blocs de noms qui sont souvent classés ensemble comme étant synonymes, on utilise la classification (CAH, classification ascendante hiérarchique) intégrée à AMADO online :

## Traitement / Données de fréquences / Classification



On voit mieux les classes en les isolant avec des séparateurs : Il faut cliquer sur une colonne (ou une ligne) pour la sélectionner et insérer un séparateur entre celle-ci et la précédente : **Traitement / Insérer séparateur**



Sur ce graphique, on voit que certains groupes de noms sont quasi-synonymes pour les chefs d'entreprises :

- Plateforme multimodale, Parc aéroportuaire, Zone industrialo-portuaire, Centre de gros, Eurofret, Plateforme logistique ; puis un peu séparés : Aéropole, Téléport.
- Couveuse, Incubateur.
- Technoforum, ZIRST (*Zones d'Innovation et de Recherche Scientifique et Technique*), Parc industriel et technologique, Parc scientifique, Parc technologique, Technoparc, Technopole. On observe que les noms Parc Scientifique et Parc technologique sont presque confondus dans l'esprit des chefs d'entreprise.
- Jardin d'entreprises, Pépinière d'entreprises et Ruche d'entreprises
- Atelier relais et Usine Relais.

Les responsables d'une campagne de marketing d'un territoire devront réduire la gamme de sites d'implantation d'activités qu'ils offrent et n'utiliser qu'un seul nom par groupe, tout en sachant que d'autres territoires utilisent des synonymes.

# Traitement de Données numériques homogènes : diagonalisation & classification

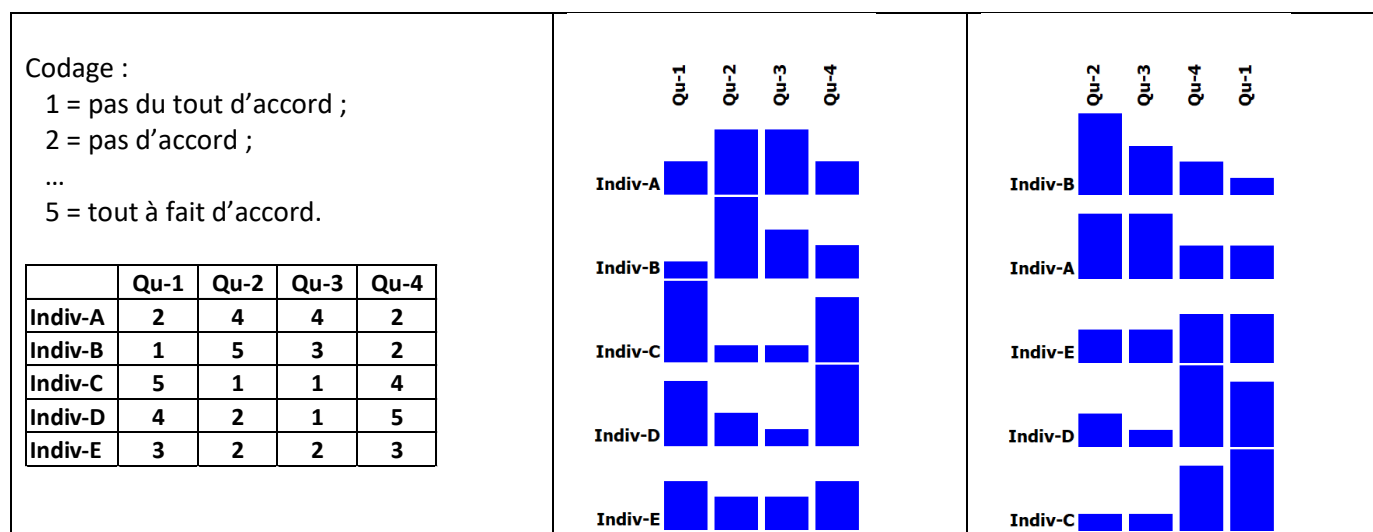
Ces traitements sont adaptés aux tableaux dont les colonnes représentent des variables dont les unités sont homogènes.

## Exemples : réponses à des réponses utilisant toutes la même échelle

(pas du tout d'accord=1 ; pas d'accord=2 ; ... tout à fait d'accord=5), ou bien des prix unitaires, ou bien des températures, etc.

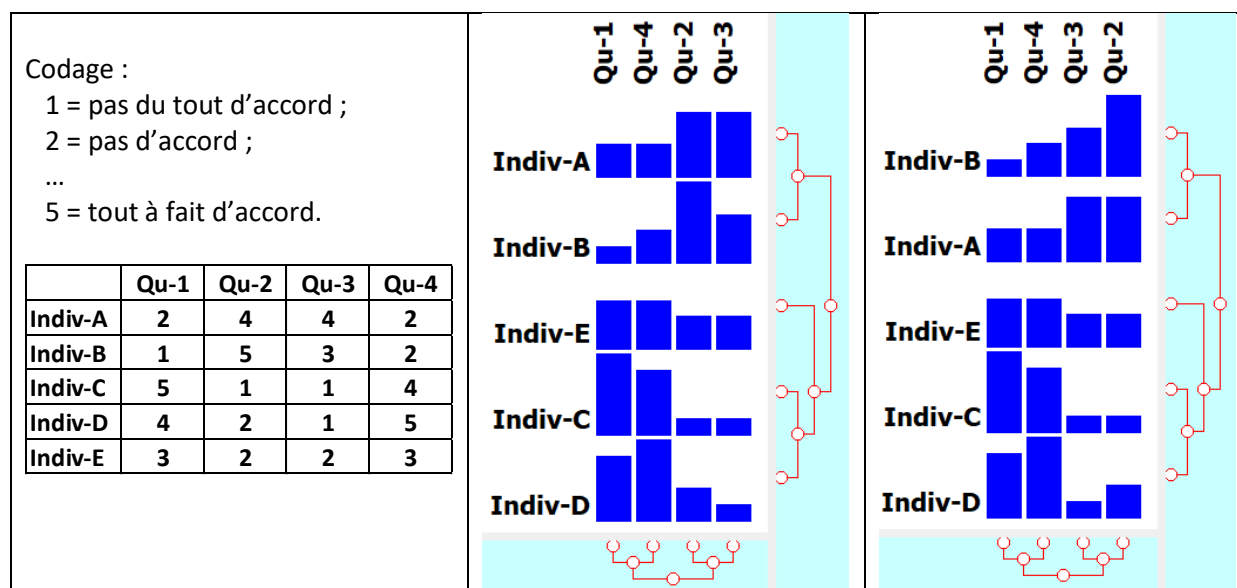
Après « **Traitement / Données numériques homogènes / Prétraitement par Analyse en Composantes Principales** », les lignes-individus sont réordonnées selon leurs coordonnées sur le premier facteur, et les colonnes-variables selon leurs corrélations avec ce premier facteur.

### Analyse en Composantes Principales.



Après « **Traitement / Données numériques homogènes / Classification automatique** », les lignes sont permutées comme les nœuds de l'arbre de Classification Ascendante Hiérarchique et cet arbre est dessiné à droite du graphique. De même les colonnes sont permutées comme les nœuds de l'arbre de Classification Ascendante Hiérarchique et cet arbre est dessiné en bas du graphique.

**Remarque** : comme, à chaque nœud de l'arbre, l'ordre de deux classes est arbitraire, **on peut inverser cet ordre par un double clic sur petit rond qui représente le nœud**. Ci-dessous, on a ainsi permuté les colonnes Qu-2 et Qu-3, et les lignes Indiv-A et Indiv-B.





# Traitement des Données numériques hétérogènes : diagonalisation & classification.

## 24 modèles de voitures selon 6 caractéristiques.

Fichier / Voitures 2004 Tenenhaus.TXT

Ces traitements sont adaptés aux tableaux avec des colonnes représentant des variables d'unités différentes.

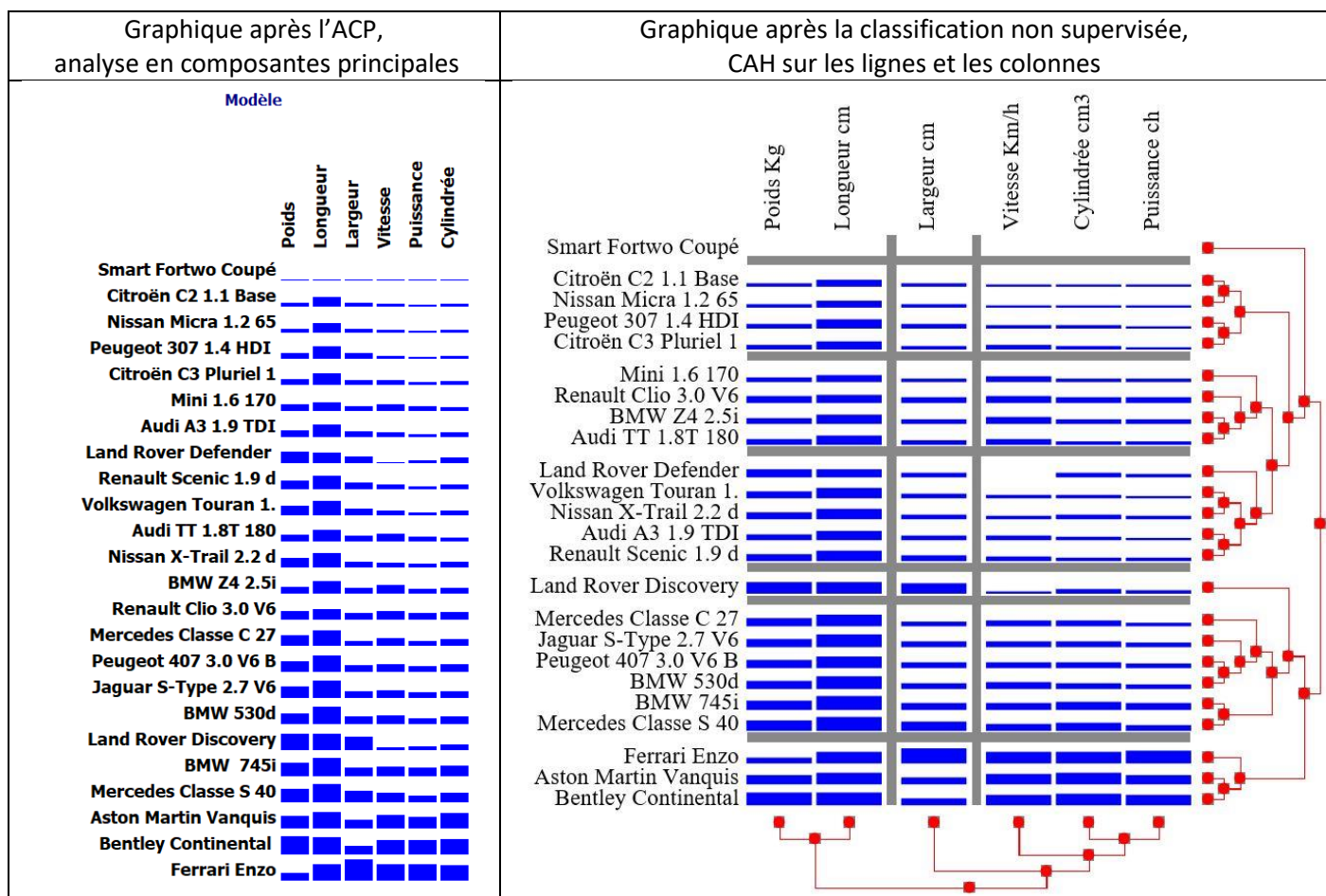
Exemple pour les voitures : Cylindrée en cm<sup>3</sup>, Puissance (cheval-vapeur), Vitesse en Km/h, Poids en Kg, Largeur et Longueur en cm. Pour s'affranchir du choix des unités, on « normalise » chaque variable.

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

Ici, les colonnes ont des unités différentes (cm<sup>3</sup>, ChV, Km/h, Kg, cm) et ne sont pas directement comparables. Les hauteurs de rectangles proportionnelles aux nombres du tableau (graphique ci-dessus) n'ont pas de sens ici.

Pour tout calcul ultérieur, il faut normaliser les colonnes : chaque valeur du tableau est centrée sur le minimum de colonne puis divisée par l'écart-type de colonne ; on obtient alors des nombres purs, c'est-à-dire "sans dimension" : si  $i$  représente une ligne-voiture et  $j$  une colonne-mesure,  $X_{ij}$  devient  $(X_{ij} - \text{Min}_j) / \sigma_j$  ; comme AMADO online ne peut représenter que des nombres positifs, on centre chaque colonne sur son minimum et la plus petite valeur devient zéro. Ensuite, les calculs sont effectués sur ces "nombres purs".

Dans notre exemple, la "Smart Fortwo Coupé" est la plus petite voiture parmi les 6 variables, les 6 valeurs deviennent zéro pour elle dans la suite sur les graphiques.



Sur l'arbre de classification, on distingue bien les classes de voitures :

- la *Smart Fortwo Coupé* est seule, la plus petite pour toutes les variables ;
- les *Citroën C2*, *Nissan Micra*, *Citroën C3* et la *Peugeot 307* forment un groupe homogène de 4 petites voitures ;
- petites (mais plus sportives) les *Mini*, *Renault Clio*, *BMW Z4* et *Audi TT* ;
- grandes voitures familiales *Land Rover Defender*, *Nissan X-Trail*, *Volkswagen Touran*, *Renault Scenic* et *Audi A3* ;
- la *Land Rover Discovery* est spécifique, étant longue, large et lourde, relativement peu puissante pour sa taille et plutôt lente ;
- les 6 berlines grandes, nerveuses et rapides : *Mercedes Classe C*, *Jaguar S*, *BMW 530d*, *Peugeot 407*, *BMW 745i*, *Mercedes Classe S* ;
- enfin les grandes, très puissantes et extrêmement chères : *Ferrari*, *Bentley* et *Aston Martin*.

Du côté des variables, le *poids* et la *taille* sont fortement corrélés, tout comme la *cylindrée* et la *puissance* et, un peu moins, la *vitesse* (car il y a des petites voitures rapides et des grosses voitures lentes).

## Les graphiques de BERTIN

AMADO online permet de représenter graphiquement un tableau croisé de nombres, puis de permuter les lignes et les colonnes pour faire apparaître la structure des données : - soit une structure diagonale (sériation) si elle existe, - soit une structure en classes croisées des lignes et des colonnes, voire en blocs.

Le présent guide d'utilisation présente plusieurs types de tableaux avec, pour chacun, les données sources et les suites de commandes des menus d'*AMADO on line* permettant d'obtenir les graphiques reproduits. *AMADO online* est un outil adapté aux tableaux petits ou moyens (jusqu'à une cinquantaine de lignes et colonnes)<sup>3</sup> tels que ceux qui sont construits en Sciences Humaines et Sociales où chaque élément a été défini précisément et doit être resitué facilement dans l'ensemble.

Les graphiques produits par AMADO online sont simples à lire ; ils donnent au lecteur un accès direct au résultat : chaque élément d'information - chaque nombre du tableau de données - est restitué dans sa forme originelle, les nombres sont représentés par des rectangles dont les hauteurs sont proportionnelles aux valeurs du tableau original, soit en nombre absolu, soit en pourcentage.

L'idée de permuter les lignes et les colonnes d'une matrice dans le but de révéler une structure cachée dans une matrice de données est ancienne : Sir W. M. Flinders Petrie (1899) a présenté il y a un siècle une "séquence dans les vestiges préhistoriques", c'est-à-dire une "sériation" chronologique des formes et éléments de décor d'objets trouvés lors de fouilles en Égypte. Comme l'ont souligné Phipps Arabie, Scott Boorman & Paul Levitt (1978), Giles Caraux (1984) et Jean-François Marcotorchino (1987), cette idée a une influence croissante dans les mathématiques appliquées, en particulier dans les sciences du comportement.

Jacques Bertin (1967, 1977) a mis côte à côte des histogrammes, en utilisant une échelle appropriée, et a permuté les éléments pour révéler les structures sous-jacentes dans les données. Depuis lors, cette approche a connu un essor considérable en France et dans le monde (Bord 1997, Palsky 2017, Harvey 2019). À l'origine, Bertin et son équipe de l'École des Hautes Études travaillaient avec des rangées de cubes que l'on déplaçait à la main. Ensuite, la diffusion des méthodes d'analyse des données multidimensionnelles (Cordier 1965, Benzécri 1973, Arabie & al. 1978, Greenacre 1984, Caraux 1984, Tenenhaus & Young 1985, Hoffman DeLeeuw 1992) a quelque peu éclipsé cette approche purement visuelle.

Certes, les techniques numériques de l'analyse des données permettent de découvrir rapidement les grands traits de la structure du tableau, structure qui sera rendu lisible sur le graphique. On économise ainsi un temps considérable dans la recherche du meilleur couple de permutations des  $n$  lignes et des  $p$  colonnes du tableau parmi les  $n! p!$  solutions possibles. Mais, en analyse factorielle, les listes de coordonnées et autres « aides numériques à l'interprétation » sont utiles au statisticien mais souvent incompréhensibles pour le chercheur en sciences sociales ; il en est de même pour les graphiques factoriels, nuage des individus, cercle des corrélations, représentation simultanée, etc. Leur interprétation demande un œil averti, et ils doivent peut-être une partie de leur succès auprès du grand public à leur érotisme même ... De leur côté, les arbres de classification donnent une représentation utile mais déformée (*ultramétrique* <sup>4</sup>) du tableau originel, et presque toujours pour un seul côté du tableau, soit les lignes, soit les colonnes. Mais de nombreuses listes de moyennes, marginales et conditionnelles, d'écarts-types, de contributions, etc., sont nécessaires pour préciser le sens d'un tel arbre.

Au contraire, les graphiques construits par AMADO online utilisent l'analyse factorielle ou la classification tout en donnant au lecteur un accès direct au résultat : chaque élément d'information - chaque nombre du tableau de

---

<sup>3</sup> La méthode de représentation graphique proposée par Bertin a été adaptée aux très grands tableaux de données par Jean Daniel Fekete et ses collaborateurs (2015, 2016).

<sup>4</sup> Une distance est *ultramétrique* si tous les triangles sont isocèles, le 3<sup>e</sup> côté étant plus petit que les deux côtés égaux. C'est le cas quand on mesure la distance de deux éléments par la hauteur du plus petit nœud qui les réunit dans un arbre de classification hiérarchique. Ce type de distance est très particulière ; par exemple il est impossible de placer plus de 3 points dans un plan tels que leurs distances (au sens géométrique habituel) respectent cette condition *ultramétrique*.

données - est restitué dans sa forme originelle, soit en nombre absolu, soit en pourcentage. C'est uniquement l'ordre des lignes et des colonnes qui a changé, mais tout est là.

## Éléments de bibliographie

- 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<sup>e</sup> 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 (1<sup>re</sup> édition 1973, 2<sup>e</sup> édition 1976, 3<sup>e</sup> édition 1980, 4<sup>e</sup> 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](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/](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/cybergeogeo/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=0ahUKEwj38KiN5LrIAhUhtRoKHVF9DhMQ6AEIKzAA#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>

## Annexe : Analyse des Correspondances = Codage optimal des lignes et colonnes

**Pourquoi l'Analyse des Correspondances est-elle un outil efficace pour réordonner lignes et colonnes d'un tableau ?**

Michel Tenenhaus et M., Young, F.W. ont montré dans leur article de 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) qu'il y a de multiples façons de comprendre l'Analyse Factorielle des Correspondances d'un tableau croisé ; des méthodes identiques, mais nommées différemment, ont été proposées au cours du développement des sciences dans différents pays : *Optimal Scaling, Optimal Scoring & Appropriate Scoring methods* aux USA ; *Dual Scaling* au Canada ; *Homogeneity Analysis* aux Pays-Bas ; *Scalogram Analysis* en Israël ; *Quantification Method* au Japon ; etc.

**Exemple.** Partons d'une matrice croisant, pour 16 personnes, les couleurs des yeux et des cheveux.

Ces données peuvent aussi se mettre sous la forme d'un tableau du type « base de données » avec l'indicatif de l'individu et deux champs qualitatifs : la couleur de ses yeux et la couleur de ses cheveux.

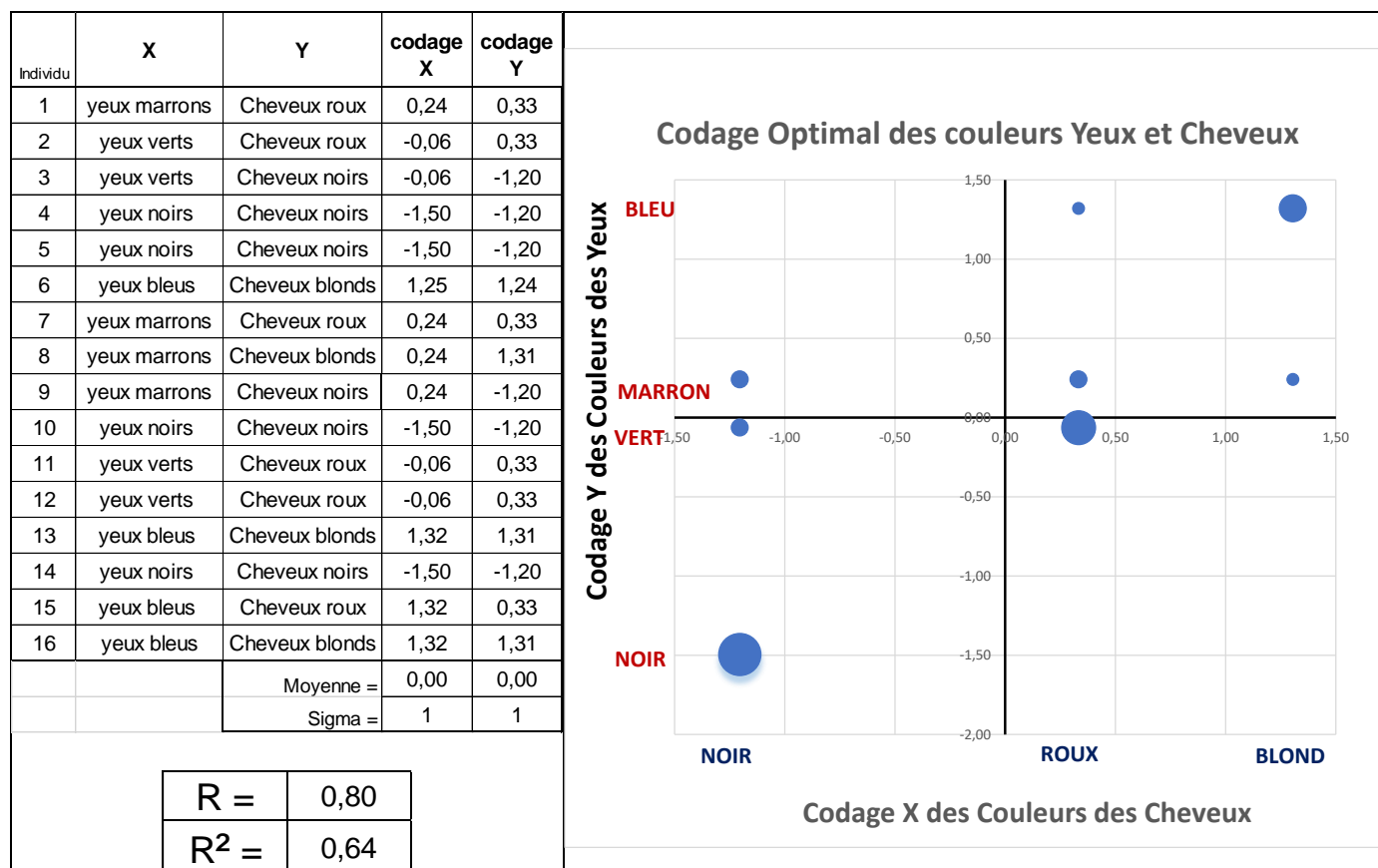
				Individu	X	Y
				1	yeux marrons	Cheveux roux
				2	yeux verts	Cheveux roux
				3	yeux verts	Cheveux noirs
				4	yeux noirs	Cheveux noirs
				5	yeux noirs	Cheveux noirs
				6	yeux bleus	Cheveux blonds
				7	yeux marrons	Cheveux roux
				8	yeux marrons	Cheveux blonds
				9	yeux marrons	Cheveux noirs
				10	yeux noirs	Cheveux noirs
				11	yeux verts	Cheveux roux
				12	yeux verts	Cheveux roux
				13	yeux bleus	Cheveux blonds
				14	yeux noirs	Cheveux noirs
				15	yeux bleus	Cheveux roux
				16	yeux bleus	Cheveux blonds

	Cheveux blonds	Cheveux noirs	Cheveux roux
yeux bleus	3	0	1
yeux marrons	1	1	2
yeux noirs	0	4	0
yeux verts	0	1	3

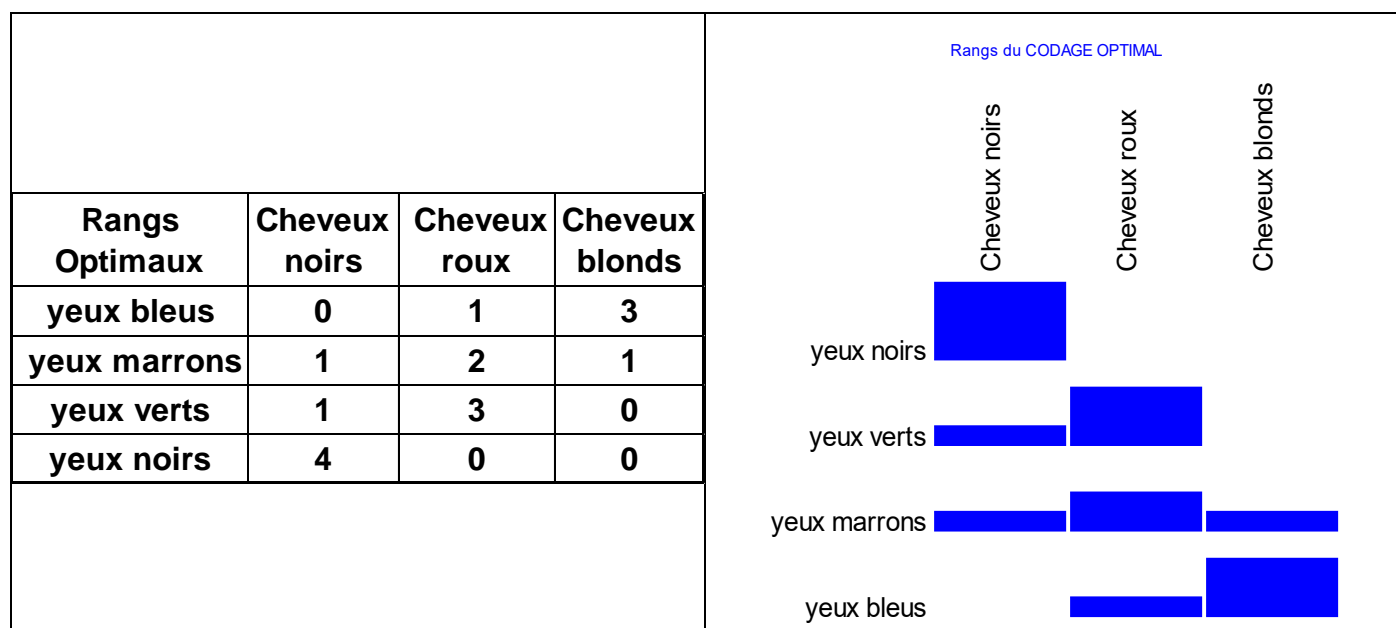
Le codage numérique des deux variables qualitatives est un ensemble de valeurs numériques associées aux modalités de chacune des deux variables qualitatives (« codage X » et « codage Y » dans le tableau ci-dessous).

Le codage optimal est celui qui maximise la corrélation « R » entre les deux variables numériques ainsi créées.



Les valeurs numériques du codage optimal sont identiques aux coordonnées des lignes et des colonnes du tableau sur le 1<sup>er</sup> facteur de l'Analyse des Correspondances.

En réordonnant les lignes et les colonnes du tableau selon les valeurs de leur codage numérique, on obtient le graphique qui montre le mieux la structure des données, c'est-à-dire les liaisons entre les deux variables qualitatives, lignes et colonnes (l'ordre des lignes est ici inversé) :



# Table des matières

<b>Débuter : Choisir la langue dans la page d'accueil, Démarrer.....</b>	<b>1</b>
Fichier : ouvrir un fichier de données. ....	1
Les 6 Modes d'affichage .....	3
Edition : Couper, Copier tableau, Coller tableau, Copier graphique. ....	4
Déplacer une ligne ou une colonne .....	5
Changer la taille, et le rapport hauteur / largeur, du graphique .....	5
<b>Traitement de Données de fréquences : diagonalisation &amp; classification .....</b>	<b>6</b>
Exemple-1 : couleurs des yeux et des cheveux de 592 personnes .....	6
Exemple-2 : les instruments joués par les élèves du Conservatoire National de Musique et de Danse de Paris et les professions et catégories socioprofessionnelles (PCS) des parents .....	7
Exemple-3 : « Sériation » chronologique d'objet archéologiques.....	9
Exemple-4 Données chronologiques : répartition des Juifs déportés depuis la France selon les convois et les villes de naissance.....	10
Exemple-5 Les actifs de 25 à 54 ans à Paris en 2015, par PCS et par arrondissement .....	11
Exemple-6 Recherche de blocs dans une matrice carrée de cooccurrence. Étude en marketing des territoires ....	13
<b>Traitement de Données numériques homogènes : diagonalisation &amp; classification .....</b>	<b>16</b>
Exemples : réponses à des réponses utilisant toutes la même échelle.....	16
<b>Traitement des Données numériques hétérogènes : diagonalisation &amp; classification. ....</b>	<b>17</b>
24 modèles de voitures selon 6 caractéristiques.....	17
<b>Éléments d'histoire des représentation graphiques de tableaux.....</b>	<b>19</b>
<b>Éléments de bibliographie .....</b>	<b>20</b>
<b>Annexe : Analyse des Correspondances = Codage optimal des lignes et colonnes .....</b>	<b>21</b>