

AMADO-online, guide de l'Utilisateur

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

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

AMADO-online a été conçu par **Nguyen-Khang PHAM** (Univ. 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 21 décembre 2020*

Introduction. Développé dans le cadre du consortium "Paris Time Machine", *AMADO-online* est une application pour représenter et analyser des matrices de données (présence/absence, comptages, réponses sur des échelles, ou mesures de variables hétérogènes) selon les principes de Jacques BERTIN (Bertin 1967, 1973, 1999 et Bertin 1977).

AMADO-online réalise un affichage graphique de matrices de données où les nombres sont représentés par des rectangles dont la surface leur est proportionnelle. Pour mettre en évidence la structure des données, les lignes (et/ou les colonnes) peuvent être réordonnées à la main, ou automatiquement selon leurs coordonnées sur le premier axe factoriel de l'Analyse des Correspondances, ou en Composantes Principales, selon la nature des données ; une double Classification Ascendante Hiérarchique est disponible. Selon la nature des données et le type d'analyse choisie, on obtient une sériation (chronologique par exemple), ou des blocs exacts ou approximatifs, ou des classes relativement homogènes. De nombreuses options de mise en forme sont disponibles. Ces graphiques sont fidèles aux données et faciles à lire.

Le présent Guide de l'Utilisateur détaille les commandes sur de nombreux fichiers d'exemples, correspondant à divers types de tableaux ; on trouvera ici les données sources et les suites de commandes des menus d'*AMADO-online* qui permettant d'obtenir les graphiques présentés ; ceux-ci peuvent être sauvegardés en format image PNG ou vectoriel SVG.

Actuellement les menus sont proposés en 7 langues : anglais, français, espagnol, italien, russe, ukrainien et vietnamien ; le Guide de l'Utilisateur est disponible en français et en anglais.

AMADO-online complète et améliore le travail de Chauchat-Risson (1998).

Après la table des matières, la section 1 explique, pas à pas, l'ouverture d'un fichier de données, puis les commandes de base pour mettre en forme le graphique, changer ses dimensions et le sauvegarder.

Les sections 2 à 4 présentent, à partir de nombreux exemples, comment mettre en évidence la structure de tableaux de fréquences ou de présences/absences, de données numériques homogènes, puis de données numériques hétérogènes.

La section 4 donne replace brièvement *AMADO-online* dans la préhistoire et l'histoire des graphiques de Bertin.

Avant la bibliographie et les remerciements, les annexes présentent des compléments utiles aux utilisateurs d'*AMADO-online* : -1- « Sauvegarder les graphiques en format PNG ou SVG ? », -2- « Comment rogner un graphique SVG qui a été collé dans Word, Excel ou PowerPoint ? », -3- « L'Analyse des Correspondances comme codage optimal des lignes et colonnes » et -4- « Les distances ultramétriques associées à une classification hiérarchique donnent une vue déformée des proximités des éléments analysés ».

Table des matières

1	Commencer avec AMADO-online	3
1.1	Ouvrir un fichier de données.....	3
1.2	Les commandes de base.....	4
1.2.1	Fichier / Exporter.....	4
1.2.2	Édition / Annuler ; Refaire ; Copier le tableau ; Éditer le titre ; Supprimer les objets sélectionnés.....	4
1.2.3	Traitements / Transposer ; Trier selon les valeurs d'une ligne ou d'une colonne.....	5
1.2.4	Format / Légende lignes ; Légende colonnes ; Valeurs.....	6
1.2.5	Format / Ligne échelle commune ou Format / Ligne échelle propre	6
1.2.6	Format / Les 6 modes d'affichage des histogrammes	7
1.2.7	Format / Taille du graphique	8
1.2.8	Typographie / Intervalles entre lignes	8
1.2.9	Typographie / Couleurs et Tailles des libellés, Valeurs et Séparateur	8
1.2.10	Déplacer une ligne ou une colonne.....	8
2	Traitement de Données de fréquences ou de présence / absence	9
2.1	Exemple de « sériation » chronologique d'objet archéologiques	9
2.2	Exemple de classification : Couleurs des yeux et des cheveux de 592 personnes.....	10
2.3	Exemple d'Analyse factorielle des correspondances : Les instruments joués par les élèves et les professions de leurs parents	12
2.4	Exemple de données chronologiques : répartition des Juifs déportés depuis la France selon les convois et les villes de naissance	15
2.5	Exemple de classification : Les actifs de 25 à 54 ans à Paris en 2015, par PCS et par arrondissement.....	16
2.6	Recherche de blocs dans une matrice carrée de cooccurrence. Étude en marketing des territoires.....	19
3	Traitement de données numériques homogènes.....	22
4	Traitement de données numériques hétérogènes	25
5	Les graphiques de BERTIN.....	27
6	Éléments de bibliographie	28
7	Annexes	29
7.1	Annexe 1: Format de graphiques PNG ou SVG ?	29
7.2	Annexe 2: Comment rogner un graphique SVG qui a été collé dans Word, Excel ou PowerPoint ?.....	30
7.3	Annexe 3 : Analyse des Correspondances = Codage optimal des lignes et colonnes	31
7.4	Annexe 4 : L'arbre de classification hiérarchique montre une réalité déformée	33
8	Remerciements	34

1 Commencer avec AMADO-online

1.1 Ouvrir un fichier de données

La première cellule de la première ligne soit est vide, soit contient un titre ; les suivantes donnent les libellés de colonne. Chaque ligne suivante commence par le libellé de ligne. Le tableau ne doit pas inclure les totaux de ligne ni de colonne. Il y a deux façons d'importer les données dans *AMADO-online* :

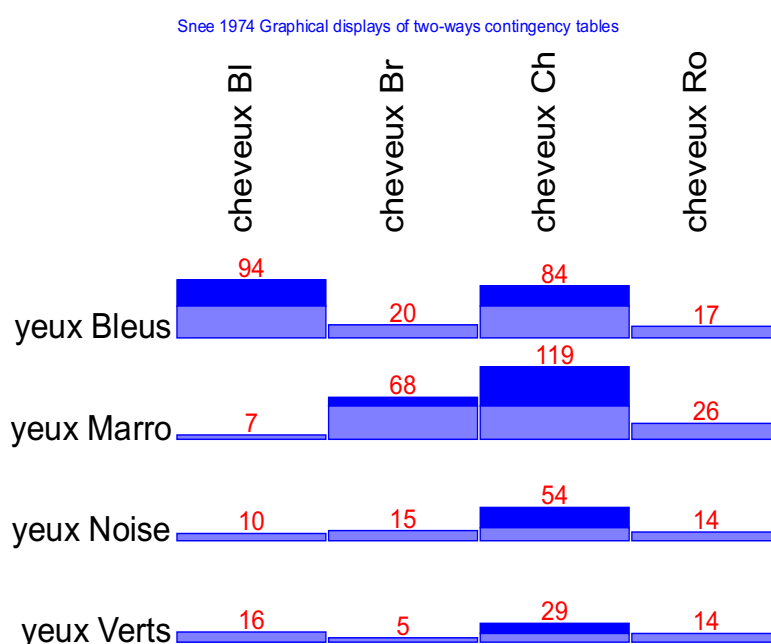
Fichier / Ouvrir un fichier en format TXT (codage UNICODE ou UTF-8, s'il y a des accents ou des caractères spéciaux) avec « séparateurs tabulations ».

Copier la zone du tableau dans un tableur, puis « **Edition / Coller** »¹ dans *AMADO-online*.

L'exemple suivant est tiré de Snee (1974). Le fichier peut être ouvert en choisissant « **Fichier / Exemples / Fr-Yeux-Cheveux.TXT** »

Snee 1974 Graphical displays of two-way contingency tables	cheveux Blonds	cheveux Bruns	cheveux Châtains	cheveux Roux
yeux Bleus	94	20	84	17
yeux Marron	7	68	119	26
yeux Noisette	10	15	54	14
yeux Verts	16	5	29	14

Dès le fichier chargé, *AMADO-online* affiche ce graphique



Les effectifs du tableau sont représentés par les hauteurs (et les surfaces) des rectangles. La valeur moyenne de chaque ligne sépare les bâtonnets en zones pâles et foncées (voir « **Format / Mode 6** »)

¹ Si vous utilisez le navigateur Firefox, allez dans le menu en haut à droite de Firefox

Personnalisez la barre d'outils / Déplacez les commandes « **Contrôle d'édition** » vers la barre d'outils. Vous pourrez alors copier un tableau.



1.2 Les commandes de base

1.2.1 Fichier / Exporter

Le menu « **Fichier / Exporter en SVG** » copie le graphique, au format SVG (*Scalable Vector Graphic*), dans le répertoire « Téléchargements » de l'ordinateur.

Le menu « **Fichier / Exporter en PNG** » copie le graphique, au format PNG (*Portable Network Graphics*), dans le dossier « Téléchargements » de l'ordinateur.

Les annexes 7.1 et 7.2 expliquent les différences entre ces deux formats de graphiques et comment rogner un graphique qui a été collé dans Word, Excel ou PowerPoint

1.2.2 Édition / Annuler ; Refaire ; Copier le tableau ; Éditer le titre ; Supprimer les objets sélectionnés

Les fonctions habituelles des logiciels de bureautique sont disponibles dans *AMADO-online* : « **Édition / Annuler** » (Ctrl Z) permet de revenir en arrière et la commande « **Édition / Refaire** » (Ctrl Y) permet de rétablir une commande annulée.

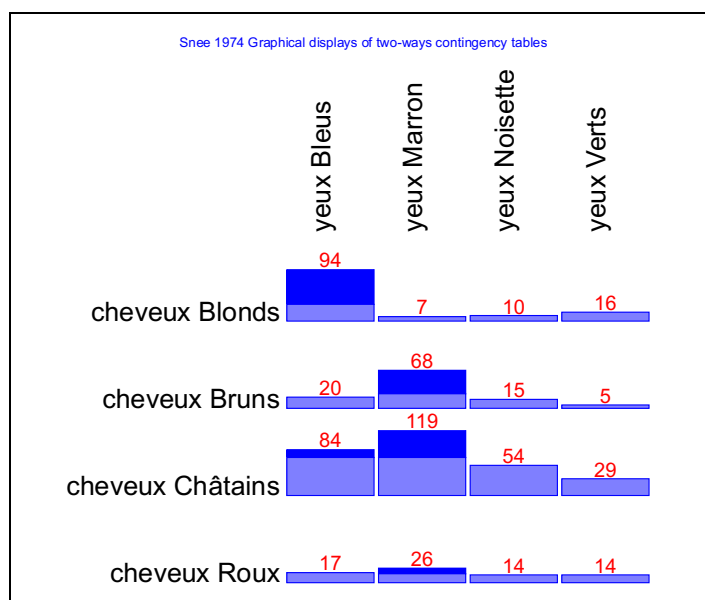
En utilisant le menu « **Édition / Copier le tableau** », on choisit le tableau courant qui peut ensuite être collé dans un fichier Texte, Excel ou Word.

Pour composer, modifier ou supprimer le titre du graphique, on utilise le menu « **Édition / Éditer le titre** ».

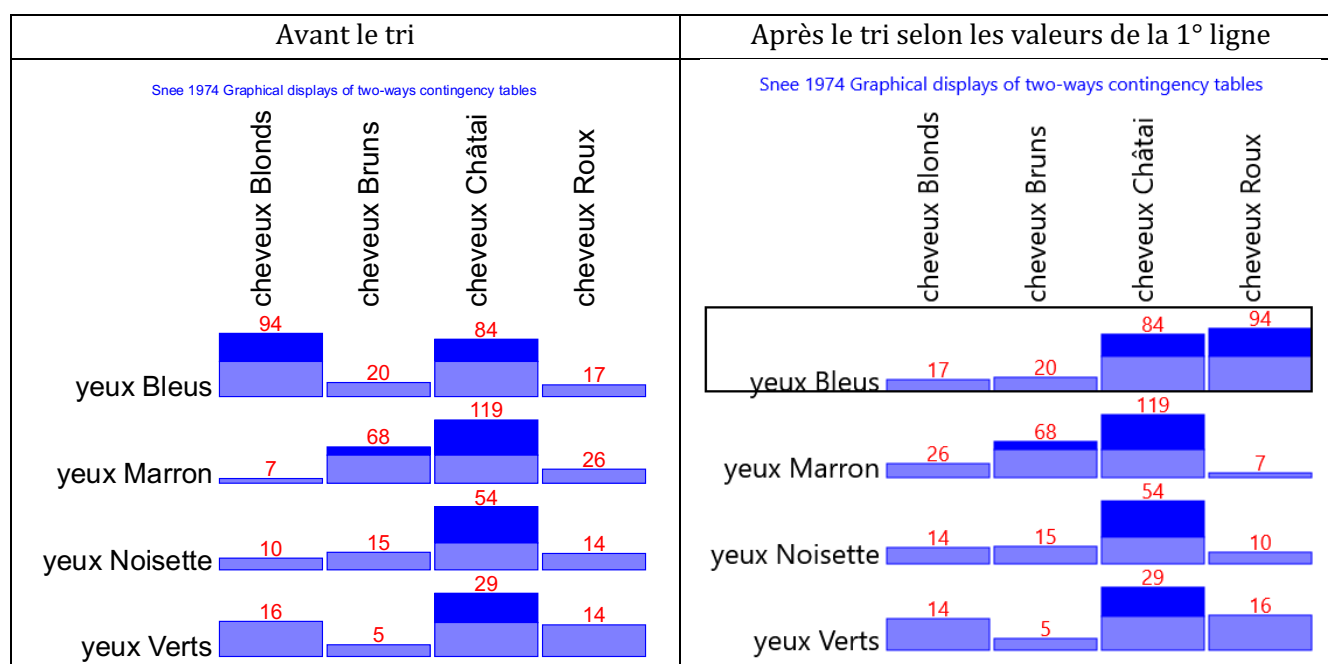
Pour supprimer une ligne, ou une colonne, il faut d'abord la sélectionner en cliquant sur le nom de la ligne ou de la colonne ; ensuite « **Edition / Supprimer les objets sélectionnés** ». Par « **Ctrl + Clic** » ; on peut aussi sélectionner plusieurs lignes, ou colonnes, pour les supprimer.

1.2.3 Traitements / Transposer ; Trier selon les valeurs d'une ligne ou d'une colonne

Pour transposer le graphique, on choisit le menu « **Traitements / Transposer** ».



AMADO-online permet de trier automatiquement les valeurs d'un tableau. Il faut d'abord sélectionner une ligne (respectivement colonne) en cliquant sur le nom de la ligne ou de la colonne. Ensuite, la commande « **Traitements / Trier ascendant** » permet de trier toutes les colonnes, respectivement lignes, dans l'ordre croissant des valeurs de la ligne, respectivement colonne, sélectionnée.



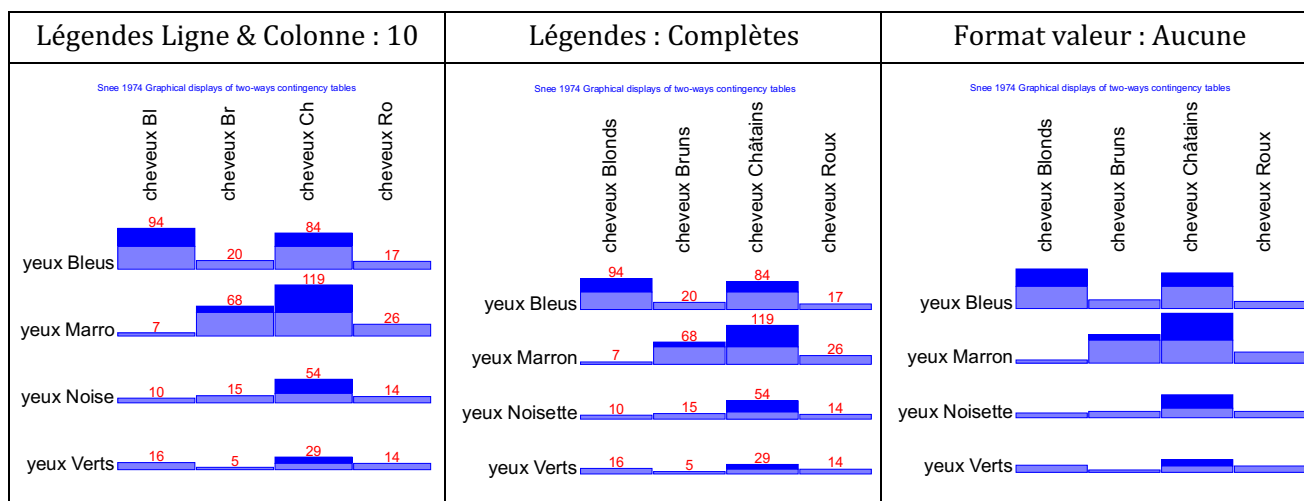
Il est aussi possible de faire « **Traitements / Trier descendant** » si cela est préférable.

On verra plus loin, dans les sections 2, 3 et 4, comment trier automatiquement lignes et colonnes pour mettre en évidence une structure diagonale, ou en blocs, selon le cas.

1.2.4 Format / Légende lignes ; Légende colonnes ; Valeurs

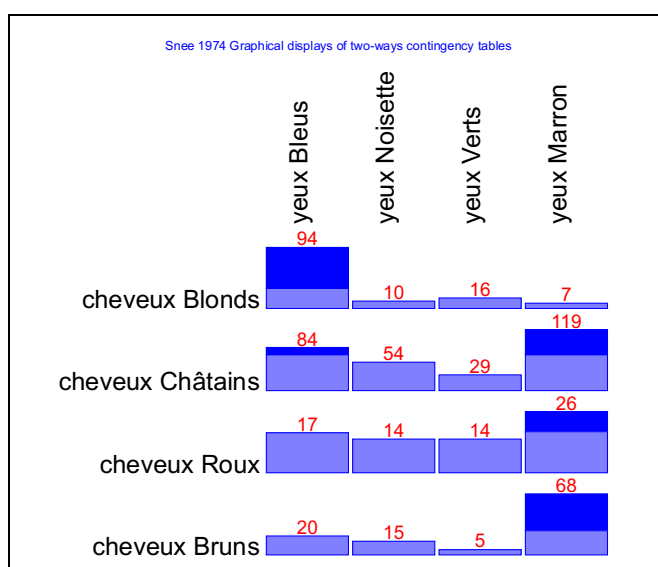
Par défaut, les libellés sont affichés avec 10 caractères. Pour afficher les libellés complets, on choisit « **Format / Légende lignes / Complète** », et « **Format / Légende colonnes / Complète** ».

Les valeurs peuvent être effacées (**Format / Légende colonnes / None**) ou bien apparaître dans différents formats : 0 ; 0.0 ; 0.00 ; 0.000 ou 0% ; 0.0% ; 0.00% ; 0.000%.



1.2.5 Format / Ligne échelle commune ou Format / Ligne échelle propre

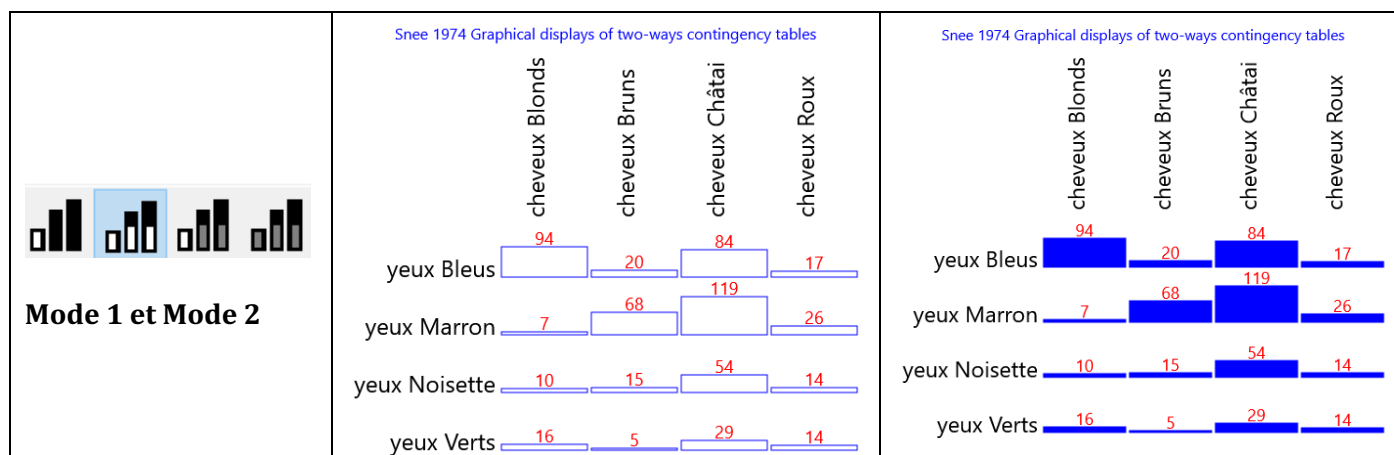
Dans les graphiques ci-dessus, toutes les lignes sont à la même échelle et les surfaces des bâtons sont proportionnelles aux nombres représentés dans l'ensemble du tableau. Si certaines lignes ont des valeurs beaucoup plus petites que d'autres, la commande « **Format / Ligne échelle propre** » rend égaux les maxima de hauteur des lignes. Les hauteurs des bâtonnets deviennent proportionnelles pour cette ligne ; c'est un peu comme des % en ligne.



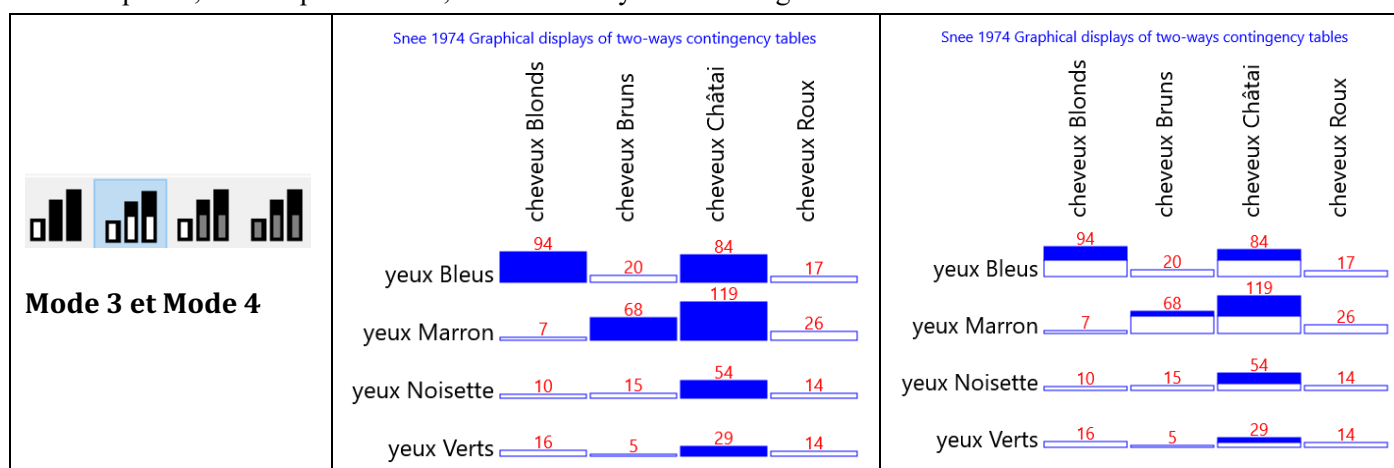
1.2.6 Format / Les 6 modes d'affichage des histogrammes

Le menu **Format** permet de modifier l'apparence des bâtonnets. Six modes d'affichage sont proposés.

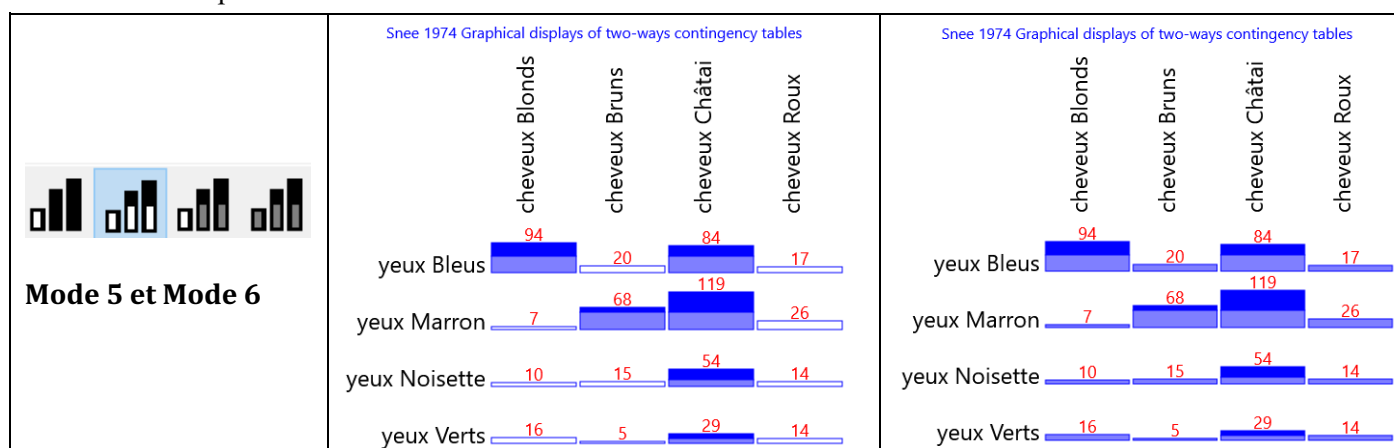
Les commandes « **Format / Mode 1** » et « **Format / Mode 2** » diffèrent dans la façon de remplir chaque bâtonnet : en clair (« **Mode 1** ») ou plein (« **Mode 2** »).



Les commandes « **Format / Mode 3** » et « **Format / Mode 4** » permettent de distinguer visuellement, les valeurs inférieures (en clair) ou supérieures (en plein) à la valeur moyenne de chaque ligne. L'option « **Mode 4** » indique où, sur chaque bâtonnet, se situe la moyenne de la ligne.



Finalement, les « **Mode 5** » et « **Mode 6** » ajoutent au « **Mode 4** » en colorant la portion inférieure à la moyenne de la ligne. Sous le « **Mode 6** », les bâtonnets entièrement inférieurs à la moyenne de leur ligne sont aussi colorés en pâle.

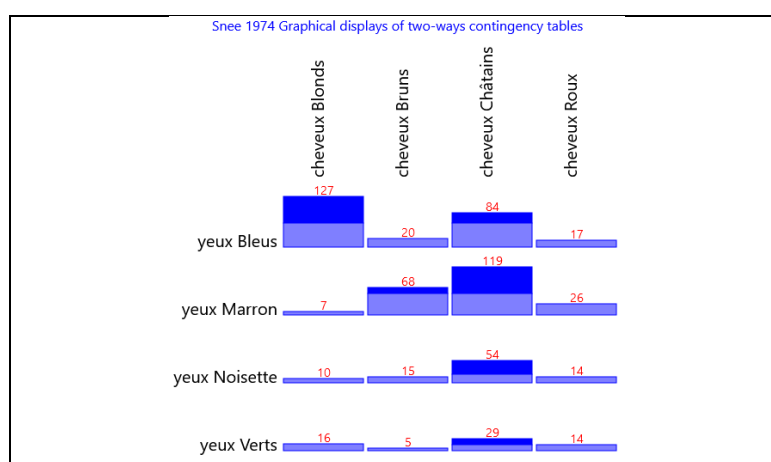


1.2.7 Format / Taille du graphique

On peut modifier la largeur et la hauteur d'un graphique au moyen du « **Format / Taille du graphique** ». Par défaut, les dimensions sont fixées automatiquement par *AMADO-online*. Pour les modifier, il faut d'abord décocher l'option « **Redimensionnement automatique** ». Ensuite, soit par bouton radio soit par curseur, on fixe les dimensions souhaitées. La commande est validée en cliquant « **D'accord** ».

1.2.8 Typographie / Intervalles entre lignes

On peut augmenter ou diminuer l'espacement entre les colonnes ou les lignes dans le menu « **Typographie / Augmenter l'intervalle entre les colonnes** ». Par exemple, après avoir redimensionné la fenêtre (voir 1.2.7) et modifié l'intervalles séparant les colonnes, on peut obtenir ce graphique :

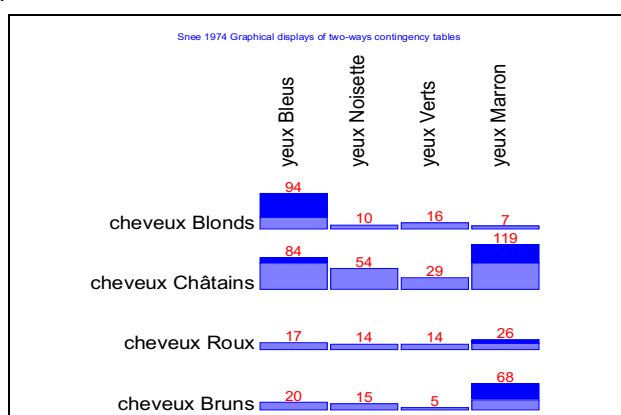


1.2.9 Typographie / Couleurs et Tailles des libellés, Valeurs et Séparateur

On peut changer les couleurs et les tailles des libellés des lignes, des colonnes et des valeurs numériques. On peut aussi changer la couleur et l'épaisseur des séparateurs.

1.2.10 Déplacer une ligne ou une colonne

Pour mettre en évidence la structure des données, on peut déplacer une ligne ou une colonne, ou un bloc de ligne (colonnes) contiguës, avec la souris (on verra plus loin des méthodes automatiques de diagonalisation et de classification). Par exemple, après **avoir déplacé (par clic et mouvement de souris) la colonne « Yeux Marron »** puis la ligne « Chev. Bruns », on obtient une structure diagonale illustrant la corrélation entre couleur des yeux et des cheveux :



2 Traitement de Données de fréquences ou de présence / absence

Dans cette section, on montre comment traiter tout tableau croisé de comptage, ou de présence-absence. Il s'agit du menu « **Traitements / Données de fréquences ou 0/1 / Prétraitement par AFC** » (Analyse Factorielle des Correspondances) et « **Traitements /.../ Classification supervisée** ».

2.1 Exemple de « *sériation* » chronologique d'objet archéologiques

L'exemple suivant de données binaires est inspiré par celles utilisées par l'archéologue Sir Flinders Petrie pour dater des tombes fouillées à Diospolis Parva en Égypte à la fin du XIX^e 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 de 1991, *Archaeology: Theories Methods and Practice*, C. Renfrew et P. G. Bahn donnent l'exemple pédagogique ci-dessous. **EN-Egyptian_pottery.TXT**

On ouvre le fichier en choisissant « **Fichier / Exemples / EN-Egyptian_pottery.TXT** ». *AMADO-online* montre alors la matrice brute de présence/absence des éléments de décors sur les poteries A, B, C...

	◀	Ⓜ	⓪	◻	Ⓜ	⓪	⓪
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

En permutant lignes et colonnes, on peut rendre la matrice diagonale ; mais on peut obtenir le même résultat en une commande « **Traitement / Données de fréquences ou 0/1 / AFC** ».

	⓪	Ⓜ	◻	⓪	◻	Ⓜ	⓪
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

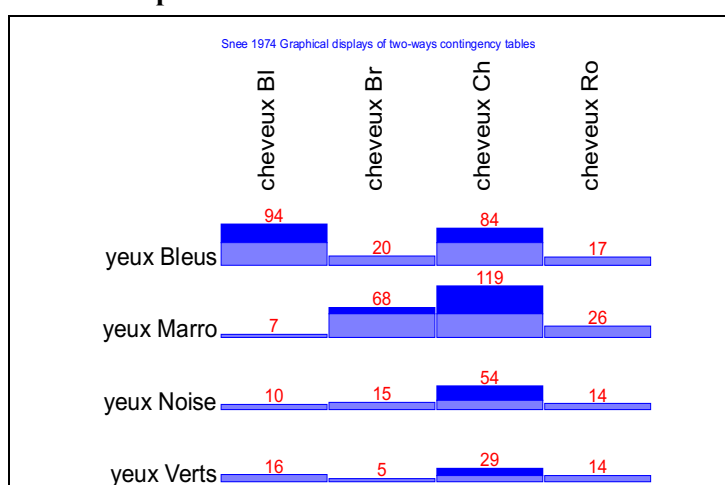
Le graphique montre les poteries (C, B, D, G, A, E et F) réordonnées, selon un axe de présence-absence des éléments de décor repérés par l'archéologue. Cet ordre correspond probablement à l'ordre chronologique (direct ou inverse) d'invention puis d'abandon de ces créations artistiques.

2.2 Exemple de classification : Couleurs des yeux et des cheveux de 592 personnes

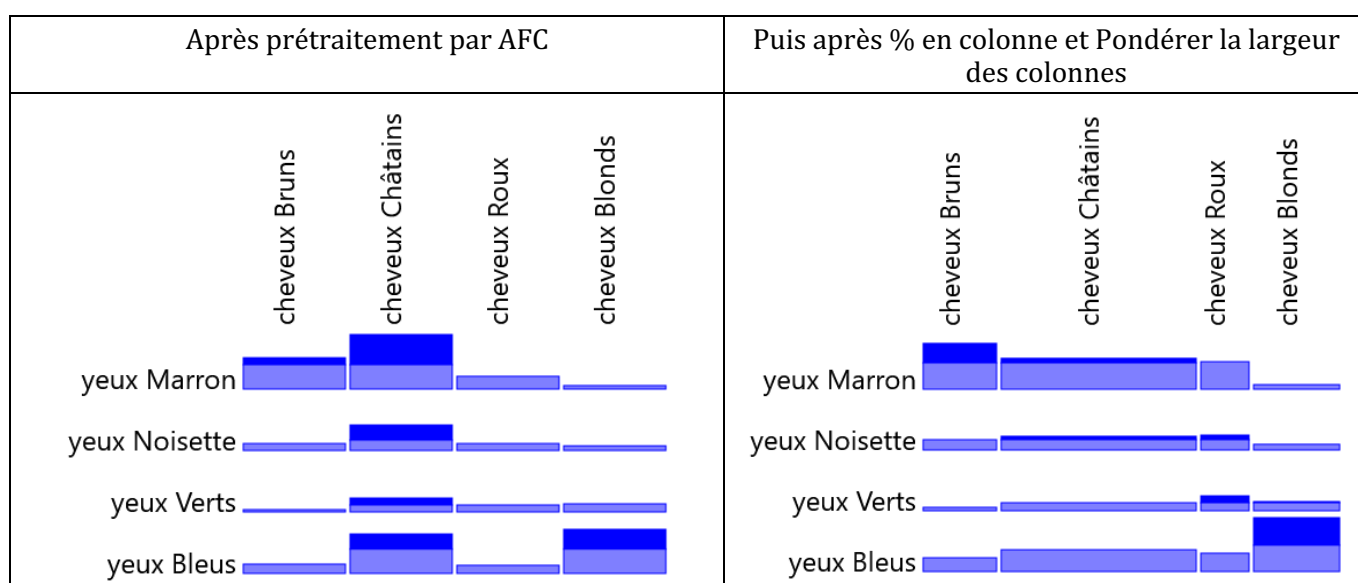
Les données de l'exemple suivant sont tirées de l'article *Graphical display of two-way contingency tables* de R.D. Snee (1974). En choisissant la matrice de données et en la copiant (Ctrl+C ou Cmd+C), on passe à *AMADO-online* où on la dépose (Ctrl+V ou Cmd+V). *AMADO-online* affiche alors le graphique ci-dessous :

Snee 1974 Graphical displays of two-way contingency tables	cheveux Blonds	cheveux Bruns	cheveux Châtains	cheveux Roux
yeux Bleus	94	20	84	17
yeux Marron	7	68	119	26
yeux Noisette	10	15	54	14
yeux Verts	16	5	29	14

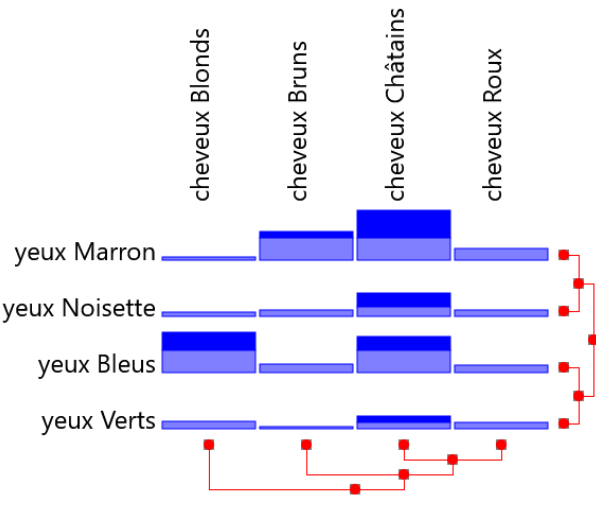
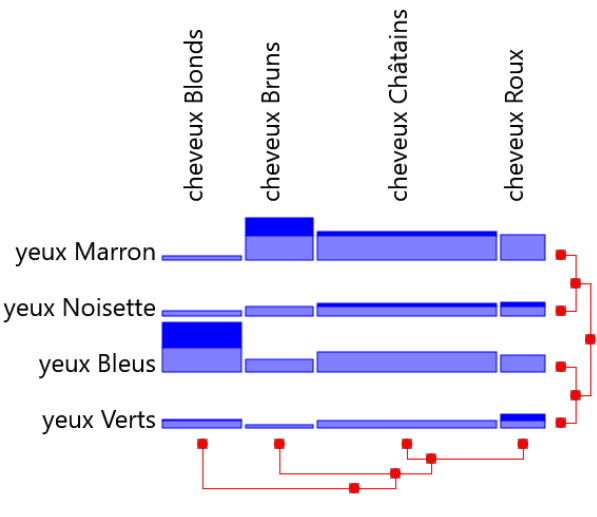
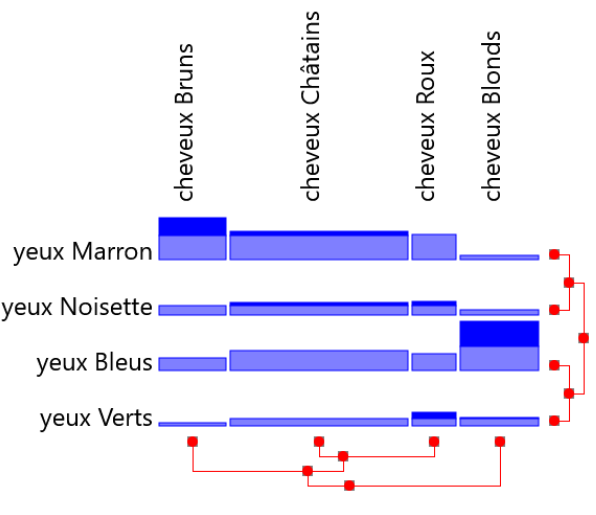
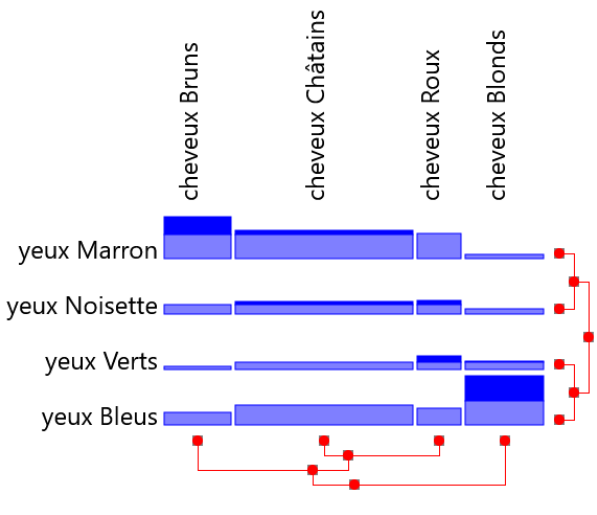
On peut aussi utiliser : « **Fichiers / Exemples / Fr-Yeux-Cheveux.TXT** ».



La commande « **Traitements / Données de Fréquences ou 0/1 / Prétraitement par Analyse des correspondances** » donne une diagonalisation de la matrice de données, diagonalisation qu'on voit mieux sur le graphique de gauche après « **Traitements / % en colonne** » et « **Format / Pondérer la largeur des colonnes par la somme de chaque colonne** »



La commande « **Traitements / Données de Fréquences ou 0/1 / Classification** » met plutôt en évidence la proximité entre lignes et colonnes, tel que montré ci-dessous.

-1- Après Classification	-2- Après % en colonne et Pondérer la largeur des colonnes
	
-3- Après clic sur le nœud le plus bas, pour mettre à droite les cheveux Blancs	-4- Après clic sur le 2° nœud le plus à gauche, pour faire redescendre les yeux Bleus
	

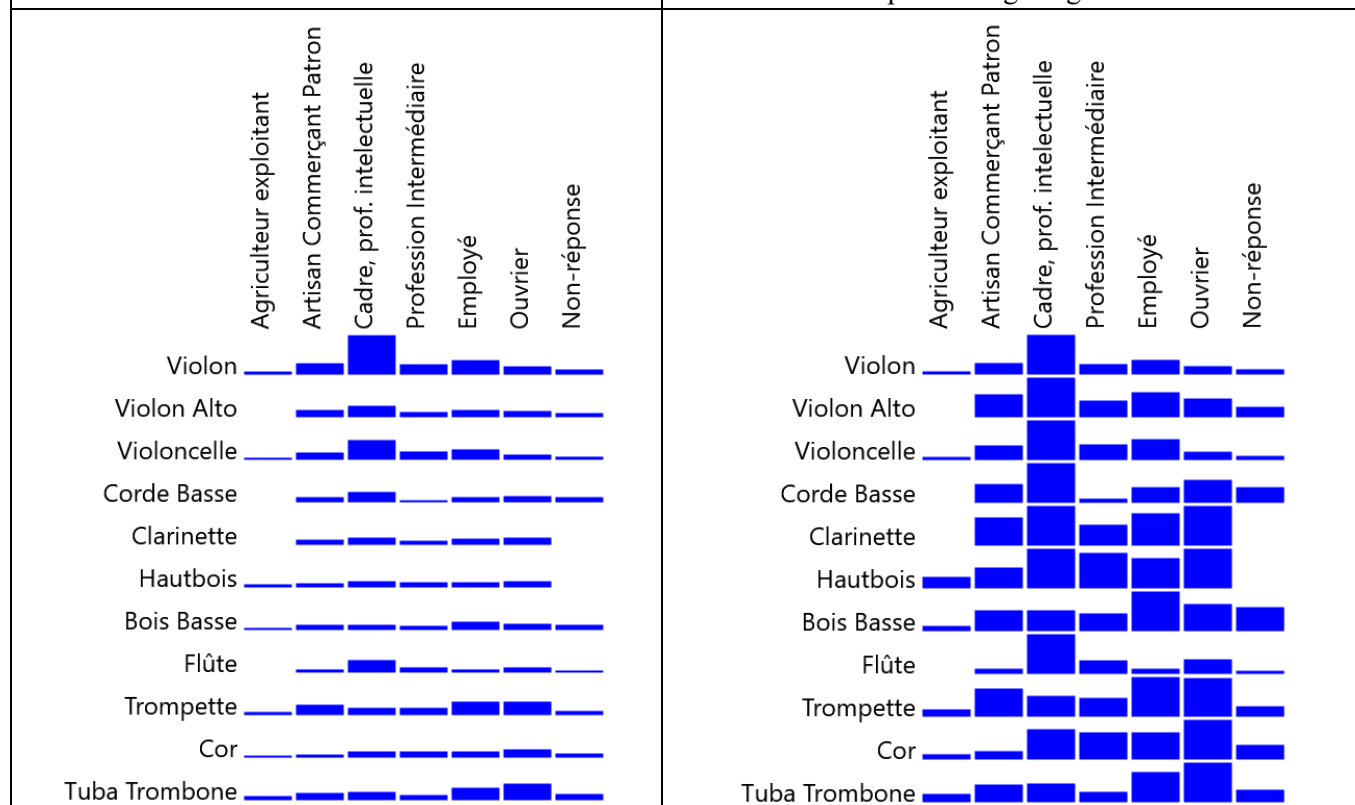
2.3 Exemple d'Analyse factorielle des correspondances : Les instruments joués par les élèves et les professions de leurs parents

Les données utilisées dans l'exemple donné dans cette section sont tirées de la thèse soutenue en 1995 par Lehman à l'EHESS *L'orchestre dans tous ses éclats : sociologie de la profession de musicien*. Il s'agit des élèves du Conservatoire National de Musique et de Danse de Paris (CNMD). On accède aussi aux données en passant la commande « Fichiers / Exemples / Fr-InstrumentsMusique-PCS.TXT ».

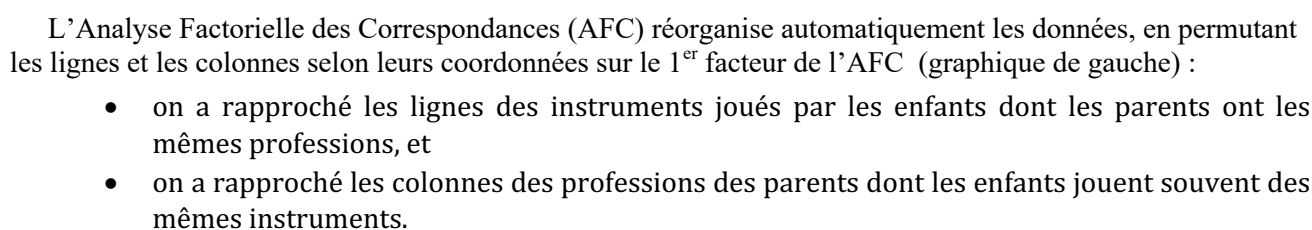
Elèves du Conservatoire Supérieur de Musique de Paris selon l'instrument joué et la PCS des parents	Agriculteur exploitant	Artisan Commerçant Patron	Cadre, prof. Intellectuelle	Profession Intermédiaire	Employé	Ouvrier	Non-réponse
Violon	2	17	69	15	24	12	7
Violon Alto	0	10	18	7	11	8	4
Violoncelle	1	11	32	12	16	6	2
Contrebasse	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

AMADO-online affiche alors ce graphique

Graphique des fréquences relatives en ligne. Les effectifs des cases étant très différents, on voit mieux les distributions en demandant « Traitements / Calculer des pourcentages lignes »



Avec « **Traitements / Transposer** », on obtient celui de droite.

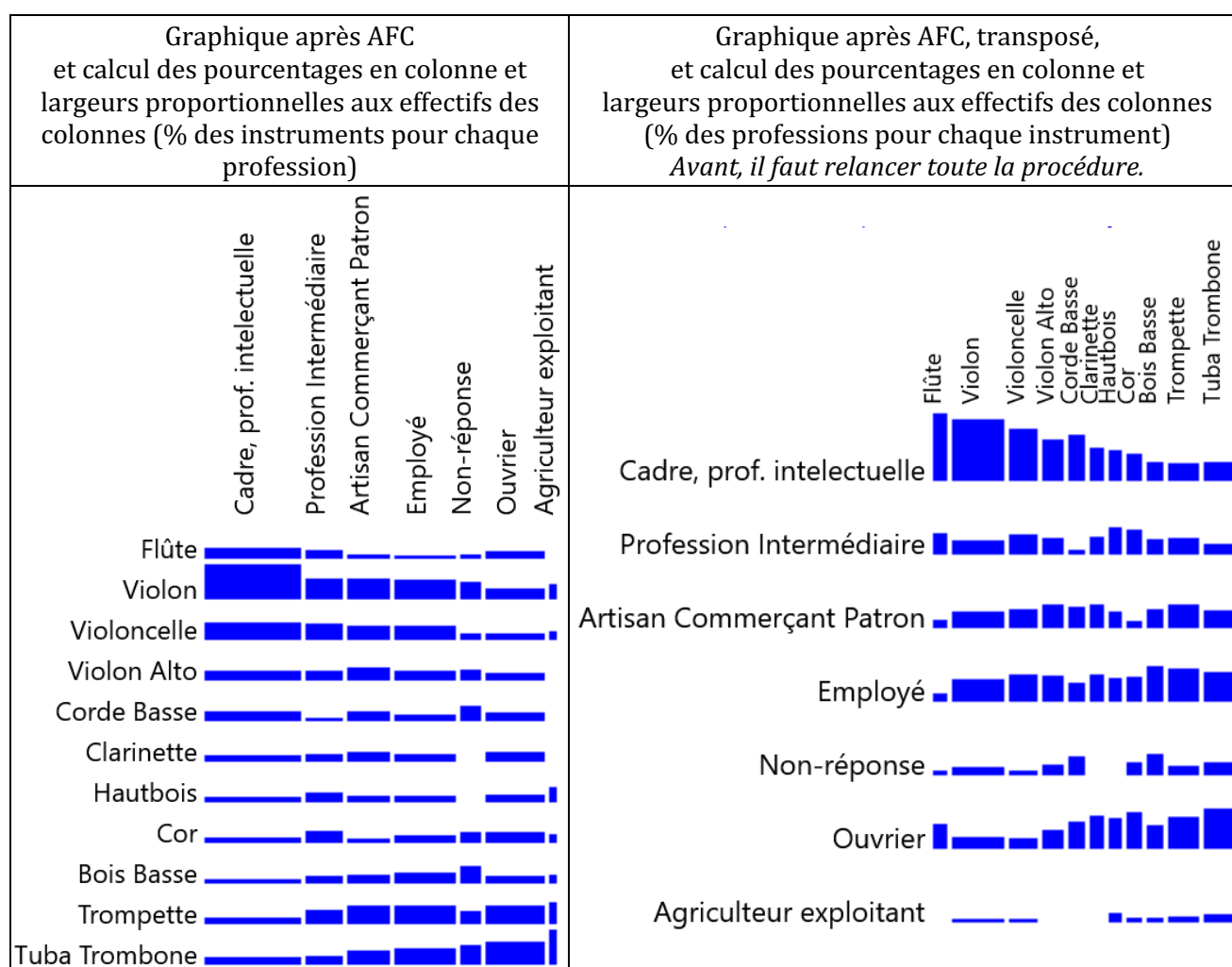


Quelque fois, une transposition des données donne un meilleur aperçu des structures (graphique de droite).

Finalement, on passe les commandes « **Traitements / Calculer des pourcentages lignes** » et « **Traitements / Calculer des pourcentages colonnes** » suivies de la commande « **Format / Pondérer la largeur des colonnes par la somme de chaque colonne** ». Cette dernière commande fera que les surfaces de rectangles représentent, de nouveau, les nombres originels (graphique de gauche).

Pour obtenir le graphique de droite, il faut repartir des données brutes puis activer « **Traiter / Transposer** », « **Traiter / Fréquence des données ou 0/1 / Traitement avec analyse des correspondances** », « **Traiter / Calculer les pourcentages des colonnes** » suivi de « **Format / Poids de la largeur de la colonne** ».

Ainsi, *AMADO-online* peut donner l'un ou l'autre des deux graphiques suivants :



Chaque nombre d'étudiants du Conservatoire est représenté par deux surfaces égales, à gauche et à droite.

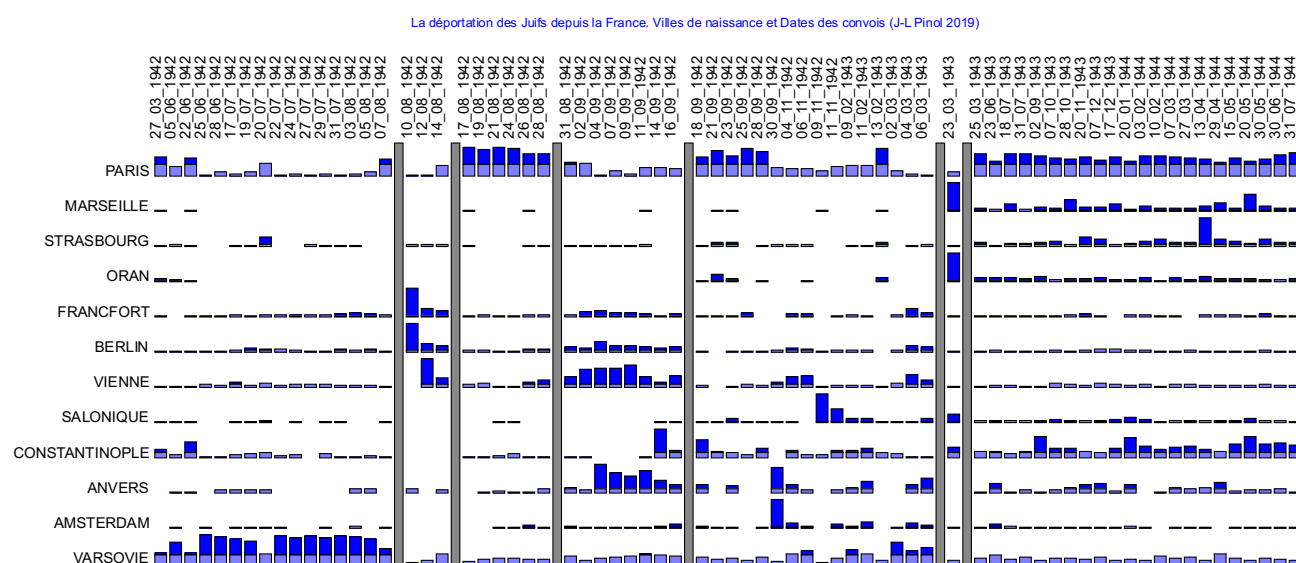
Avec ces graphiques, on visualise bien que les enfants des catégories aisées et « cultivées » sont les plus nombreux et qu'ils choisissent majoritairement la flûte ou les instruments à cordes ; à l'opposé, les cuivres et des bois sont plus joués par les enfants d'employés, d'ouvriers et d'agriculteurs, moins nombreux, (ces instruments sont joués dans les harmonies et les fanfares grâce auxquelles ils ont souvent commencé leur éducation musicale).

2.4 Exemple de données chronologiques : répartition des Juifs déportés depuis la France selon les *convois* et les *villes de naissance*

Dans cet exemple (Pinol, 2019), 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.

On accède au fichier en faisant « **Fichier / Exemples / Fr-Deportation.TXT** ».

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. On obtient les pourcentages colonnes en demandant « **Traitements / Calculer les pourcentages colonnes** ».



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 les huit convois partis entre le 31 août et le 16 septembre 1942.

Ensuite, le but de la police a été de remplir les trains pour atteindre les objectifs quantitatifs fixés.

Le convoi parti le 23 mars 1943 est particulier ; 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.

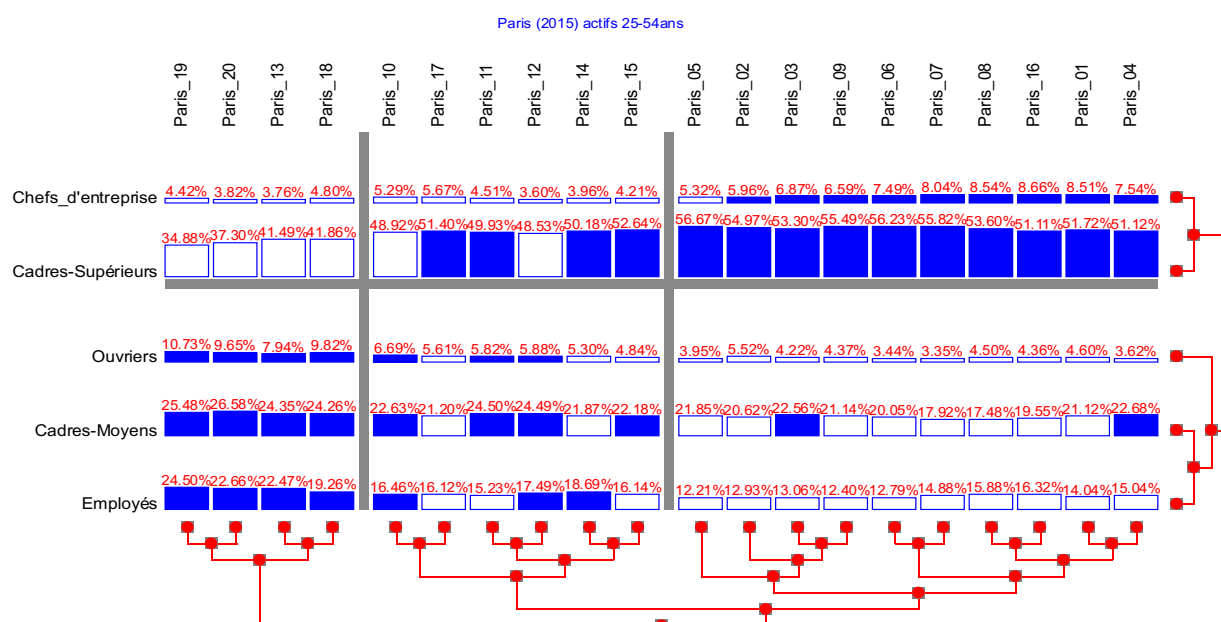
2.5 Exemple de classification : Les actifs de 25 à 54 ans à Paris en 2015, par PCS et par arrondissement

Le tableau ci-dessous montre la distribution de la population active de Paris, âgée de 25 à 54 ans, selon la catégorie socioprofessionnelle (PCS) et l'arrondissement.

Paris (2015) Actifs 25-54 ans	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

Voici la séquence des commandes passées à *AMADO-online* pour obtenir la classification et le graphique présentés à la page suivante :

- **Fichier / Exemples / Fr_Paris2015.TXT**
- **Format / Légende ligne / Complète**
- **Format / Taille du graphique /** (décocher le ☒) **Redimensionnement automatique ; Largeur=950 ; Hauteur=400 ; D'accord** (Note : la largeur du graphique est fonction de la largeur de l'écran ; 950 peut être trop étroit pour votre écran)
- **Format / Mode 3** • **Traitements / Données de fréquences ou 0/1 / Classification**
- **Traitements / Calculer des pourcentages colonnes**
- **Traitements / Pondérer la largeur des colonnes par la somme de chaque colonne**
- **Typographie / Augmenter l'intervalle entre colonnes** (deux fois ou davantage, assez pour rendre les valeurs distinctes et lisibles)
- Cliquer sur la ligne « **Ouvriers** », puis **Traitements / Insérer séparateurs** (l'insertion se fait au-dessus de la ligne sélectionnée)
- Cliquer sur la colonne « **Paris_10** », puis **Traitements / Insérer séparateurs** (l'insertion se fait avant la colonne sélectionnée)
- Cliquer sur la colonne « **Paris_05** », puis **Traitements / Insérer séparateurs**,

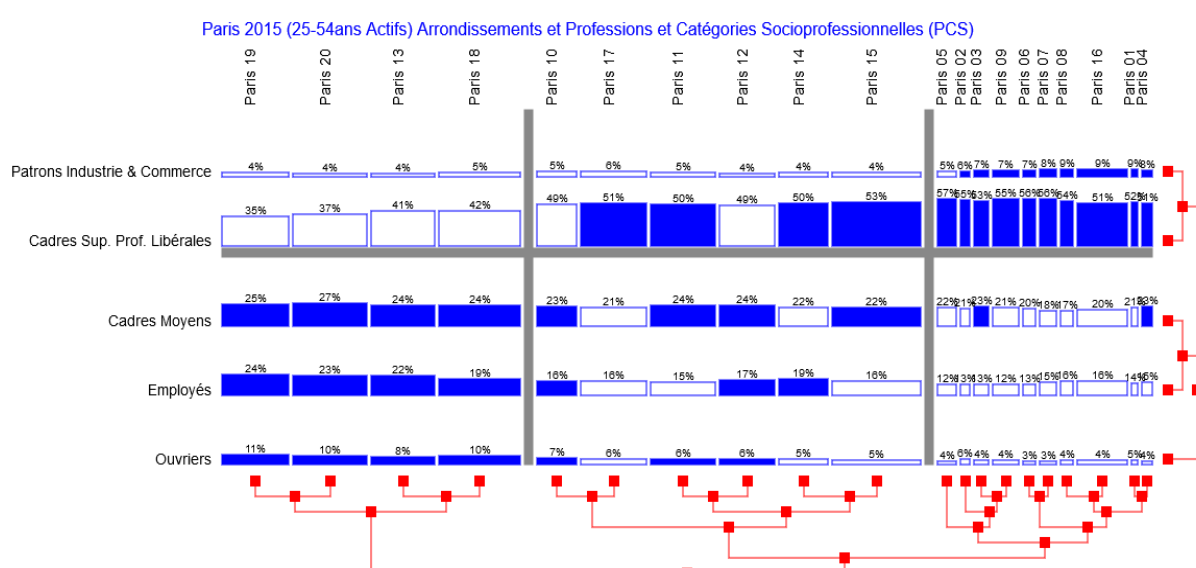


Sur ce graphique, on voit la séparation entre - d'un côté, les arrondissements de l'Est de Paris (19°, 20°, 13° et 18°) où habitent relativement plus des Ouvriers, des Cadres Moyens et des Employés ; - de l'autre côté, les arrondissements du Centre de Paris (5°, 2°, 3°, 9°, 6°, 16°, 1° et 4°) où habitent plus des Chefs d'entreprises et des Cadres supérieurs ; - et, au milieu du graphique, les arrondissements intermédiaires du point de vue sociologique (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.

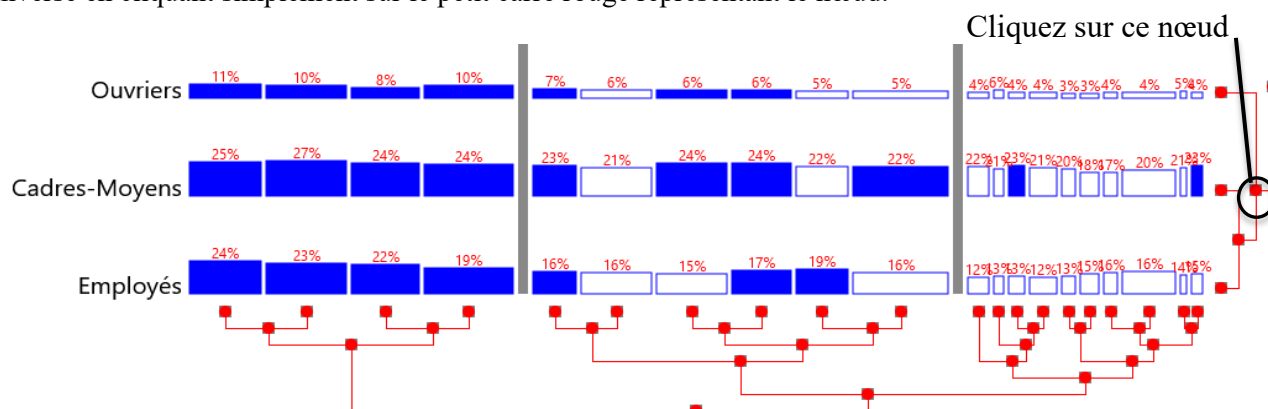
Format / Pondérer la largeur des colonnes par la somme de chaque colonne

Format / Format valeurs / 0%

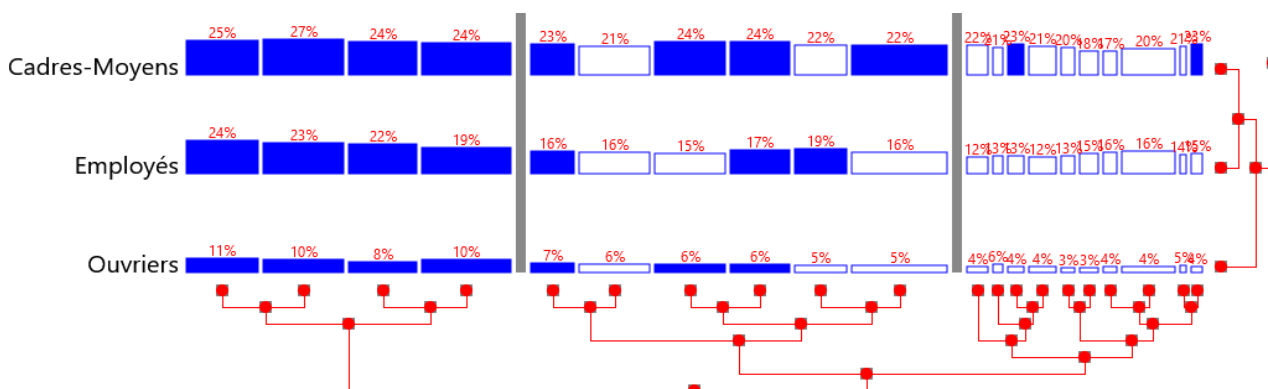


Avec **Format / Mode 3**, le graphique fait apparaître clairement dans quels arrondissements habitent le plus chaque catégorie socioprofessionnelle. REMARQUE : la page suivante explique comment la ligne « Ouvriers » a été déplacée pour se retrouver tout en bas de ce dernier graphique.

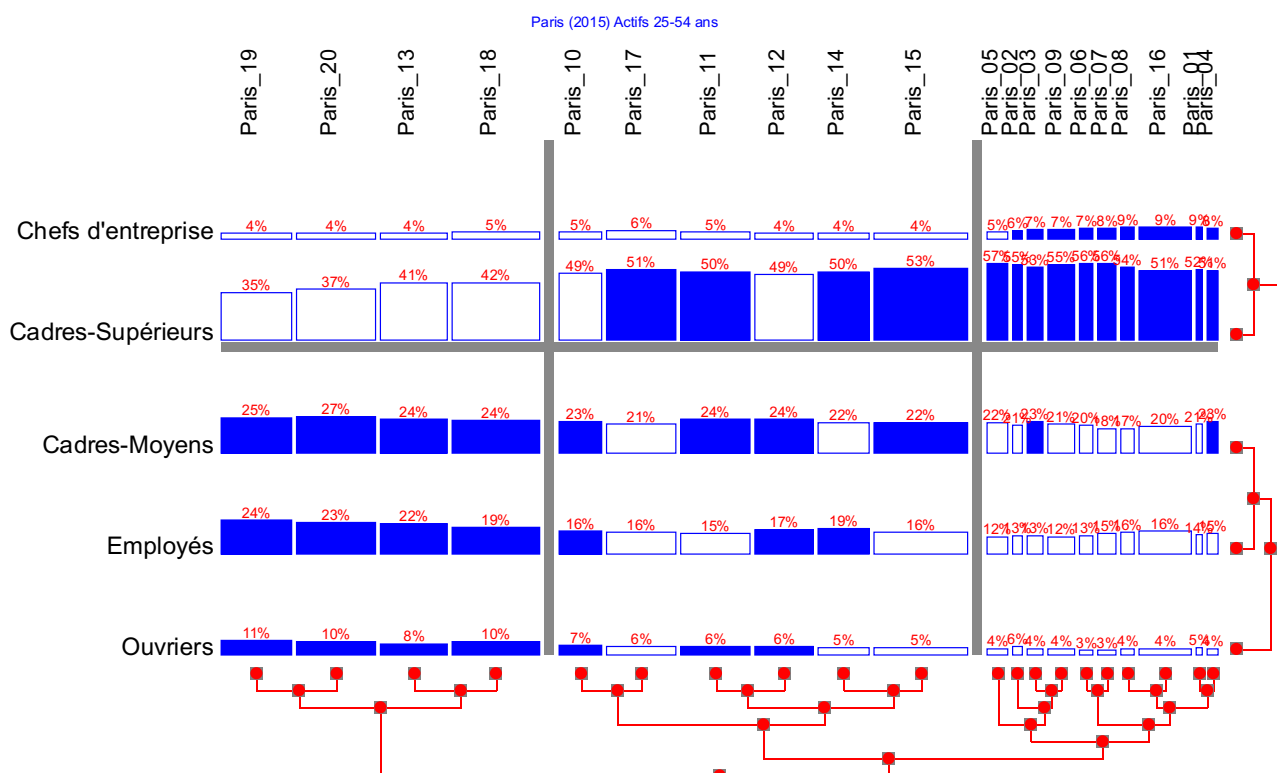
Comme, à chaque nœud de l'arbre de regroupement, l'ordre des deux classes est arbitraire, cet ordre peut être inversé en cliquant simplement sur le petit carré rouge représentant le nœud.



Ainsi, la ligne "Travailleur manuel" est déplacée en bas du graphique, à l'opposé de "Entrepreneurs".



Et, finalement, on obtient :



2.6 Recherche de blocs dans une matrice carrée de cooccurrence. Étude en marketing des territoires

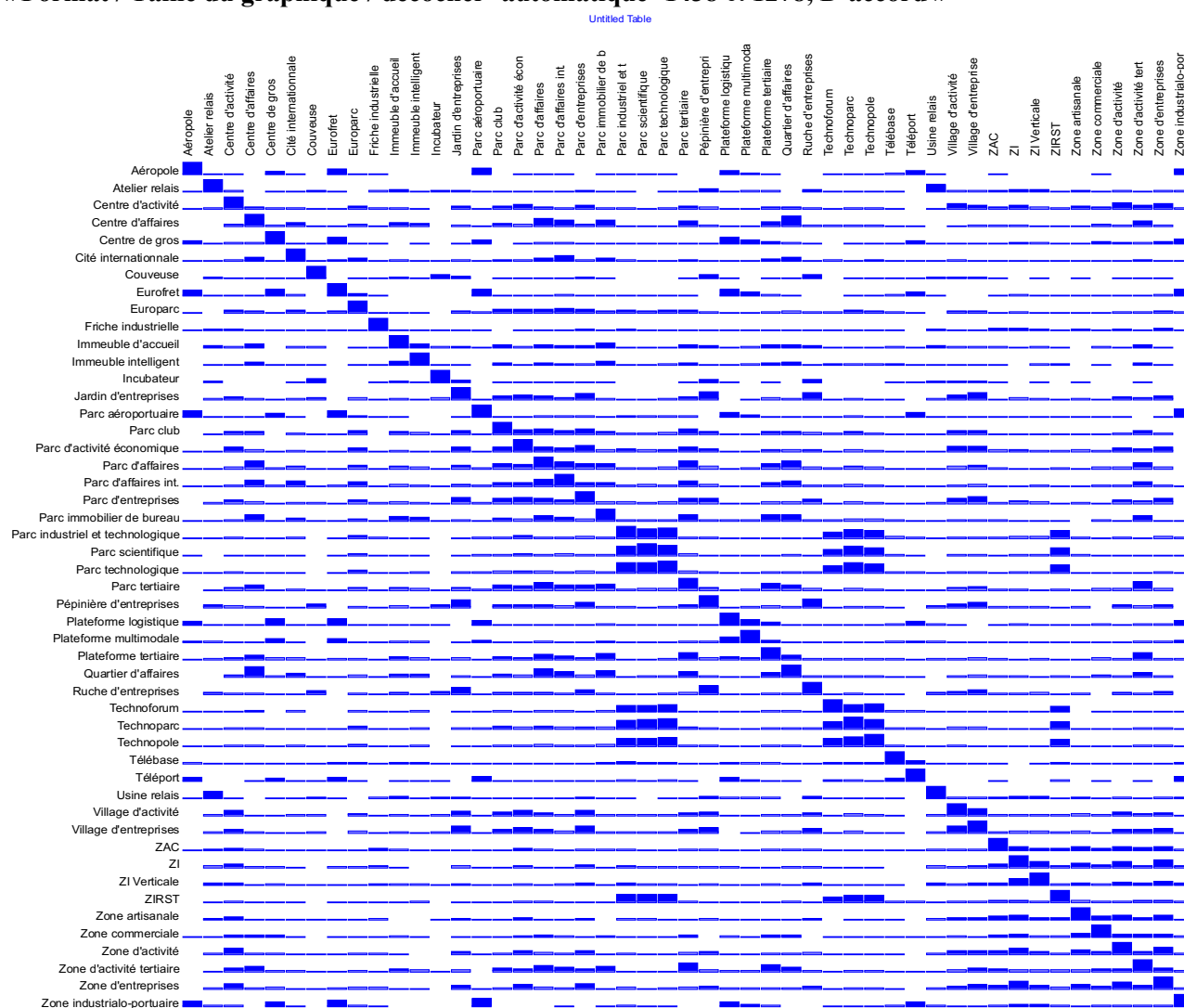
Pour attirer les entreprises sur leurs territoires, de nombreuses villes commercialisent des sites possibles d'implantation d'activités. Ces sites portent des noms divers (zone industrielle, parc technologique, ...).

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 villes françaises, 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 ensemble (Texier, 1999). Elle peut être considérée comme une matrice de similarité de type proximité.

« Fichier / Exemples / MARKETING Territorial.TXT »²

« Format / Valeurs / None » « Format / Légende ligne / Complète » « Format / Légende Colonne / 20 »

« Format / Taille du graphique / décocher 'automatique' 1438 et 1278, D'accord »³

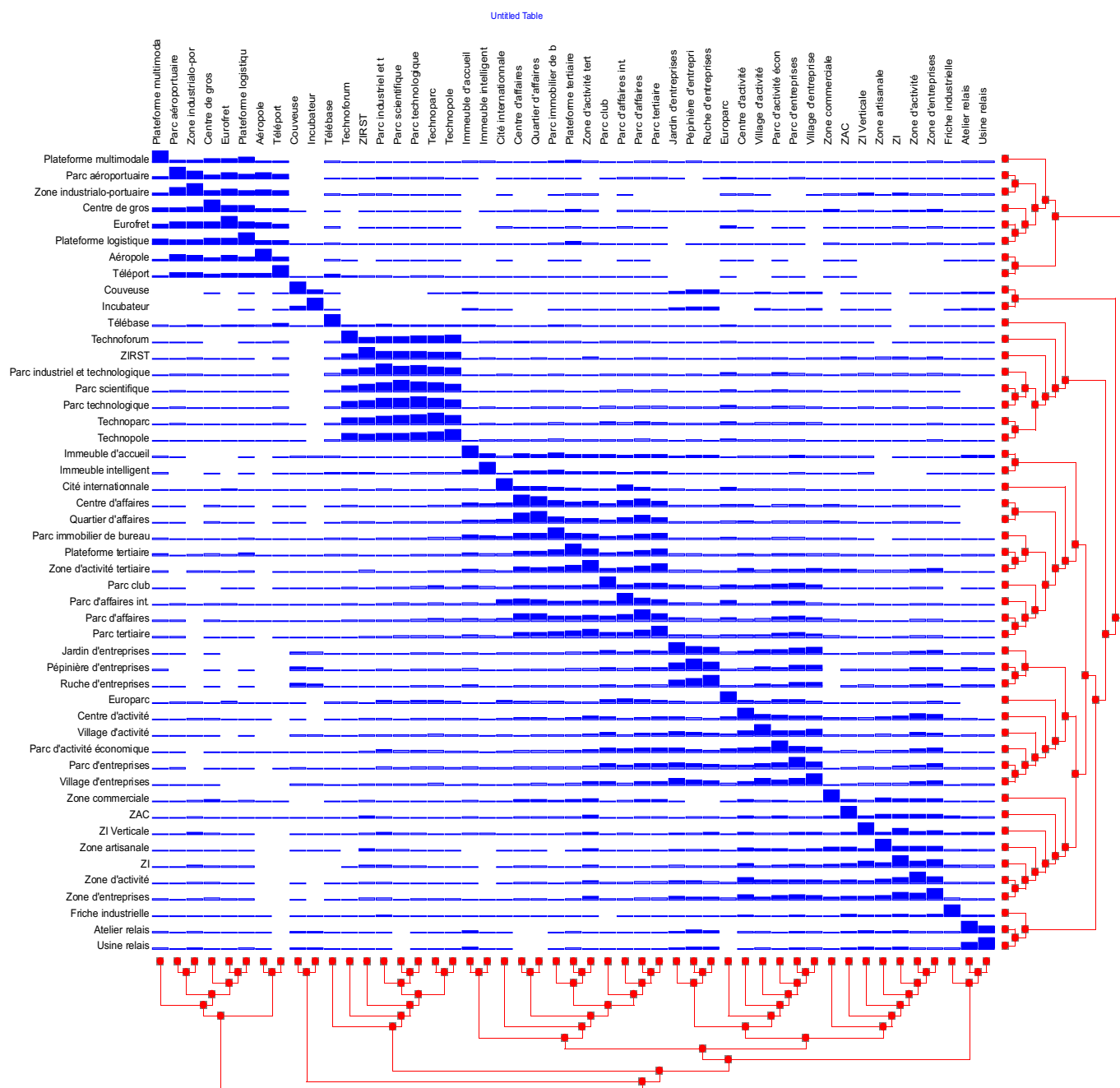


² Selon la taille du moniteur utilisé, il peut être utile ou nécessaire de faire CTRL – ou CMD – pour diminuer la police d'affichage et afficher tout le tableau

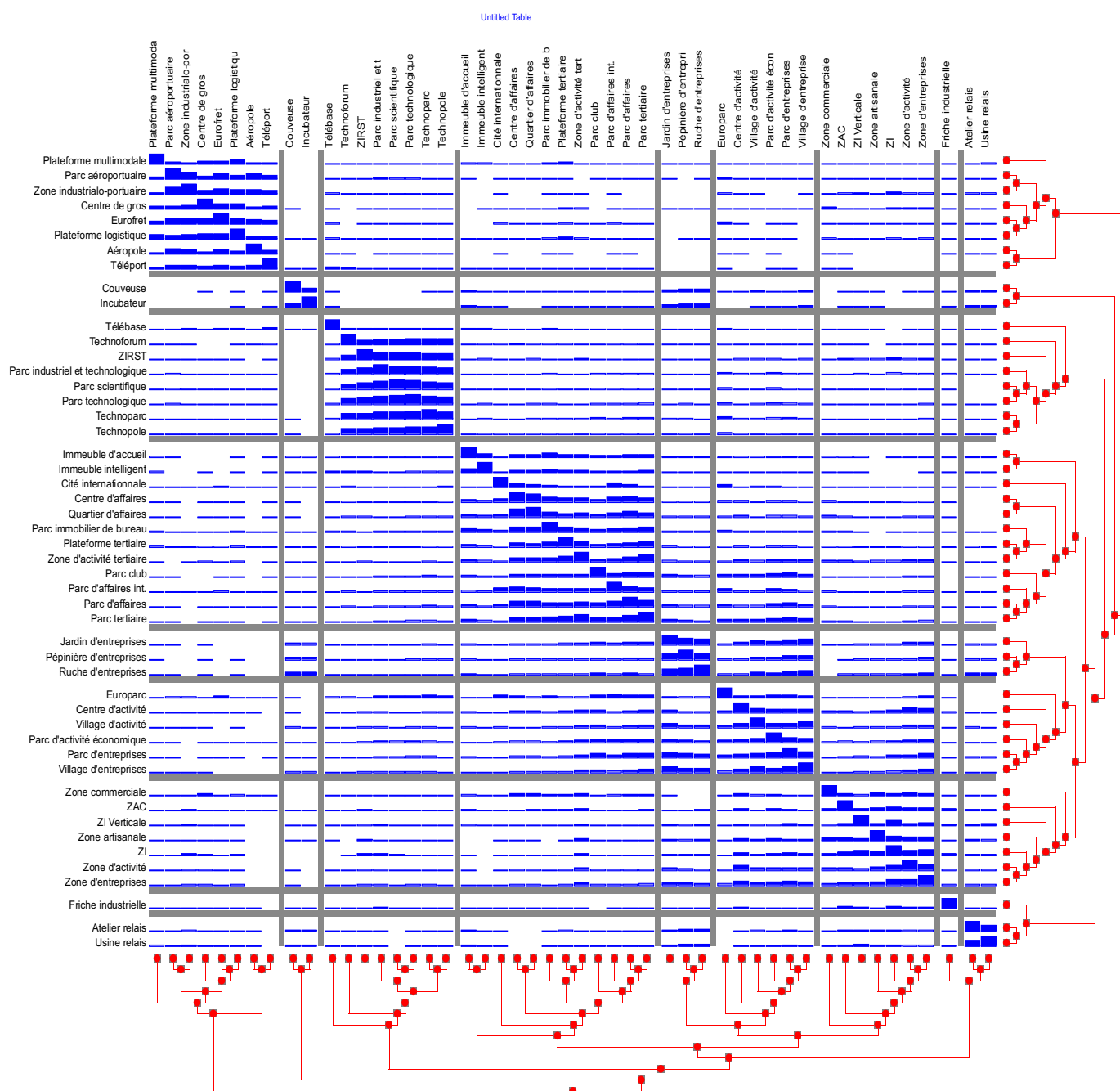
³ Les dimensions dépendent du moniteur utilisé

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* :

« **Traitements / Données de fréquences ou 0/1/ 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 : « **Traitements / Insérer séparateurs** »



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.

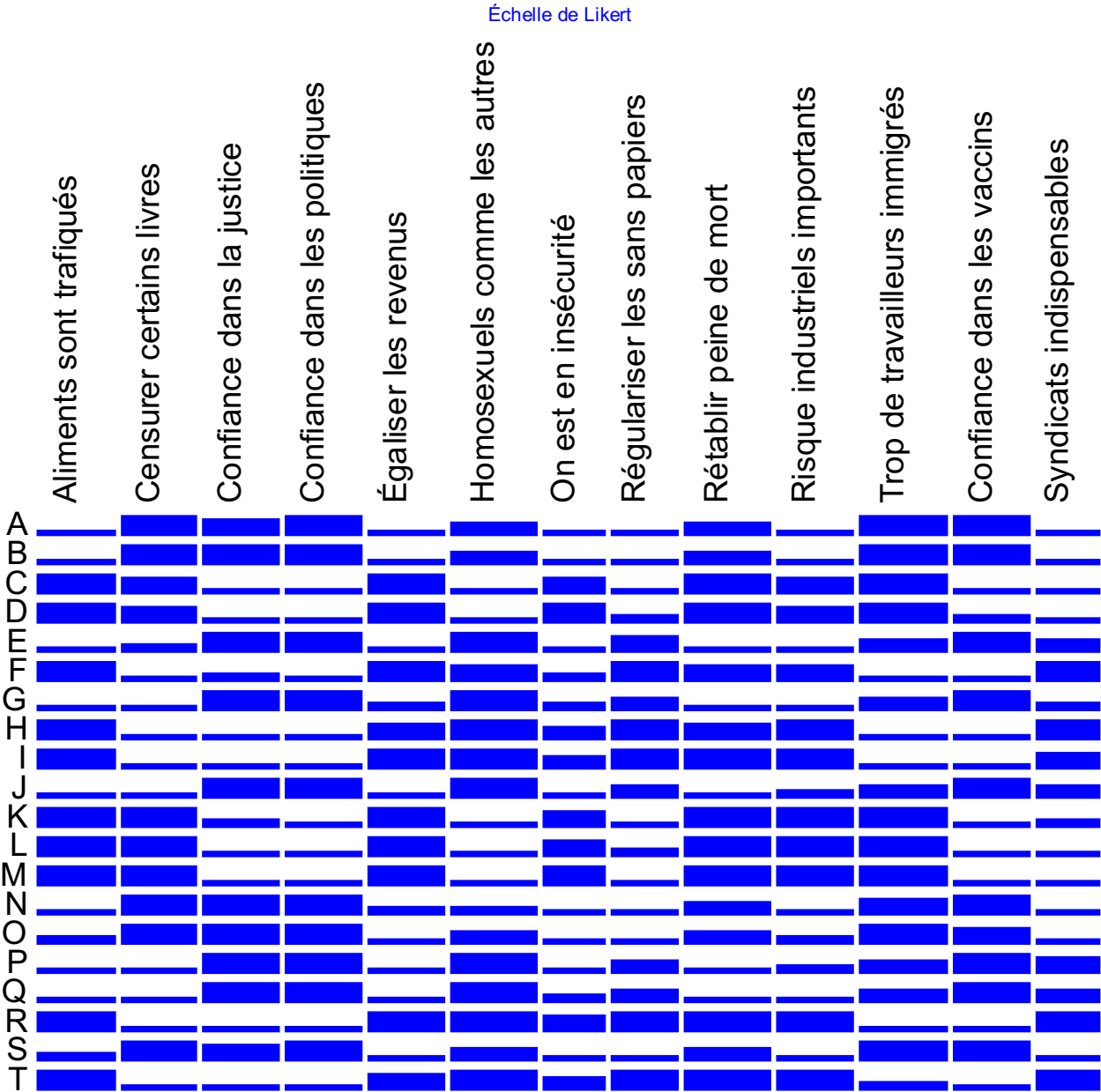
Les responsables d'une campagne de marketing d'un territoire devront réduire la gamme de sites d'implantation d'activités qu'ils offrent et n'utiliser qu'un seul nom par groupe, tout en sachant que d'autres territoires utilisent des synonymes.

3 Traitement de données numériques homogènes

Ces traitements sont adaptés aux tableaux dont les colonnes représentent des variables dont les unités sont homogènes. Il peut s'agir, par exemple, de prix unitaires, ou bien de températures, ou, comme dans notre exemple, de réponses à des items utilisant toutes la même échelle (pas du tout d'accord=1 ; pas d'accord=2 ; ... tout à fait d'accord=5) comme les échelles de Likert.

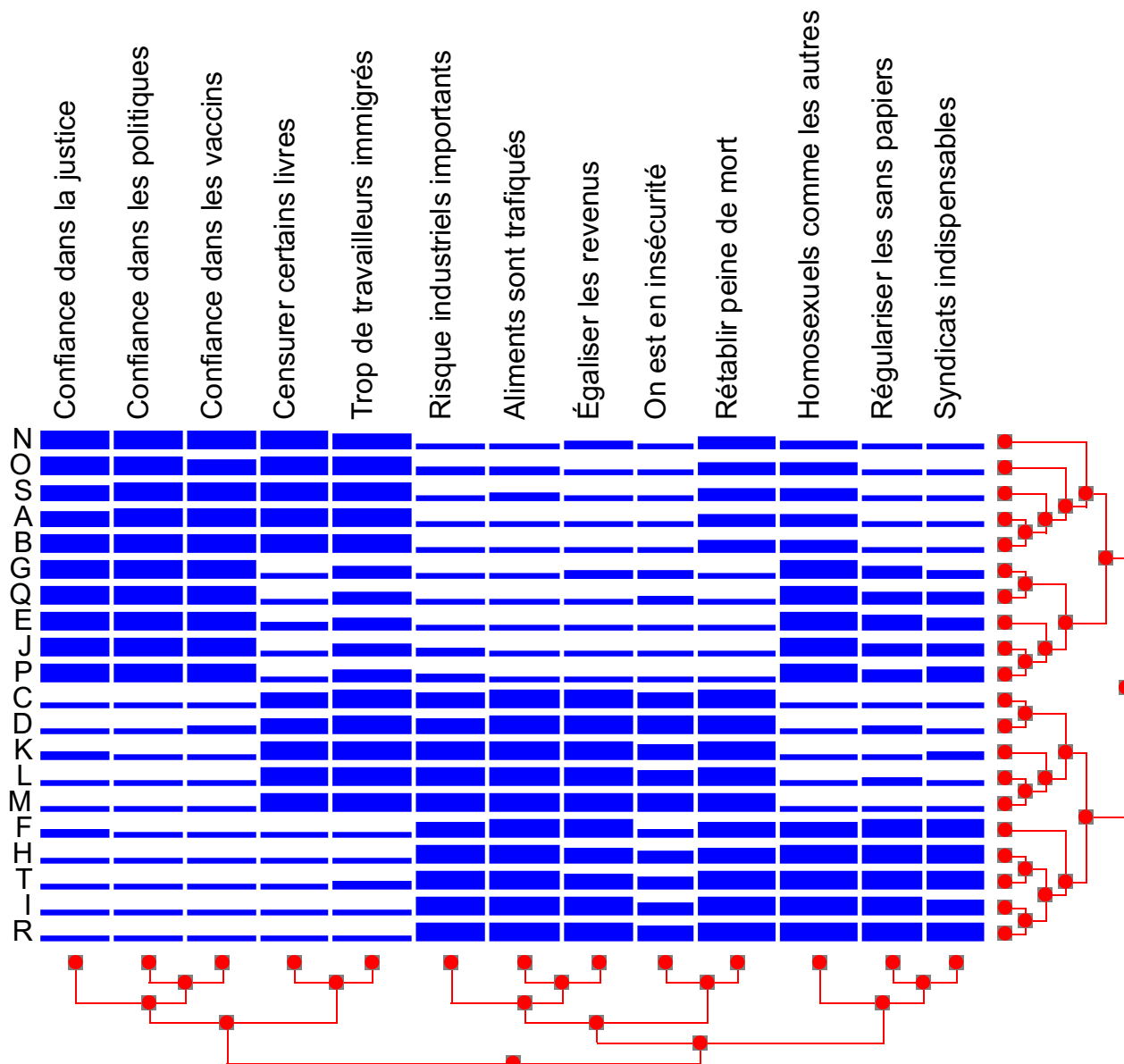
Ouvrons « **Fichier / Exemples / Fr-Enquete_Echelle_Linkert.TXT** ».

Après avoir collé le tableau dans AMADO-online, demandé l'affichage des libellés complets des lignes par « **Format / Légende Colonnes / Complète** » et « **Typographie / Taille des Colonnes / 20** », « **Format / Format values / None** » puis « **Format / Mode 3** », on obtient la représentation :

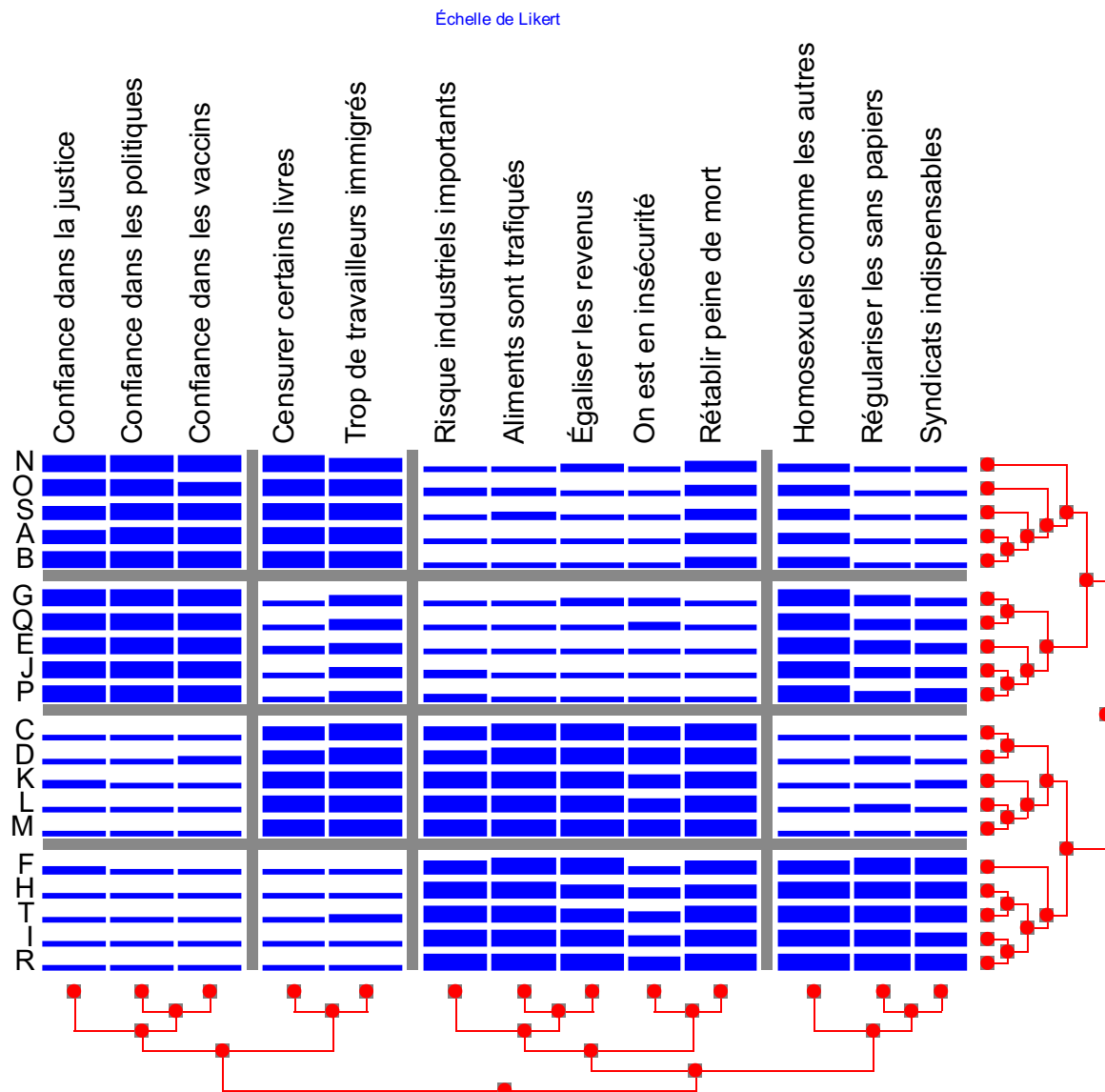


Après « **Traitements / Données numériques homogènes / Classification** », les lignes-individus et les colonnes-variables sont automatiquement réordonnées :

Échelle de Likert



Et, après l'insertion de séparateurs, on peut obtenir le graphique de la page suivante :



Ce graphique met en évidence 4 groupes d'opinion (un peu caricaturaux) qu'on pourrait schématiser ainsi, du haut en bas de l'image, chaque groupe étant caractérisé autant par les idées qu'il approuve (rectangles hauts bleu foncé), que par celles avec lesquelles il est en désaccord (rectangles bas et blancs sur le graphique). :

- la droite conservatrice ;
- le centre libéral ;
- la droite extrême ;
- la gauche extrême.

4 Traitement de données numériques hétérogènes

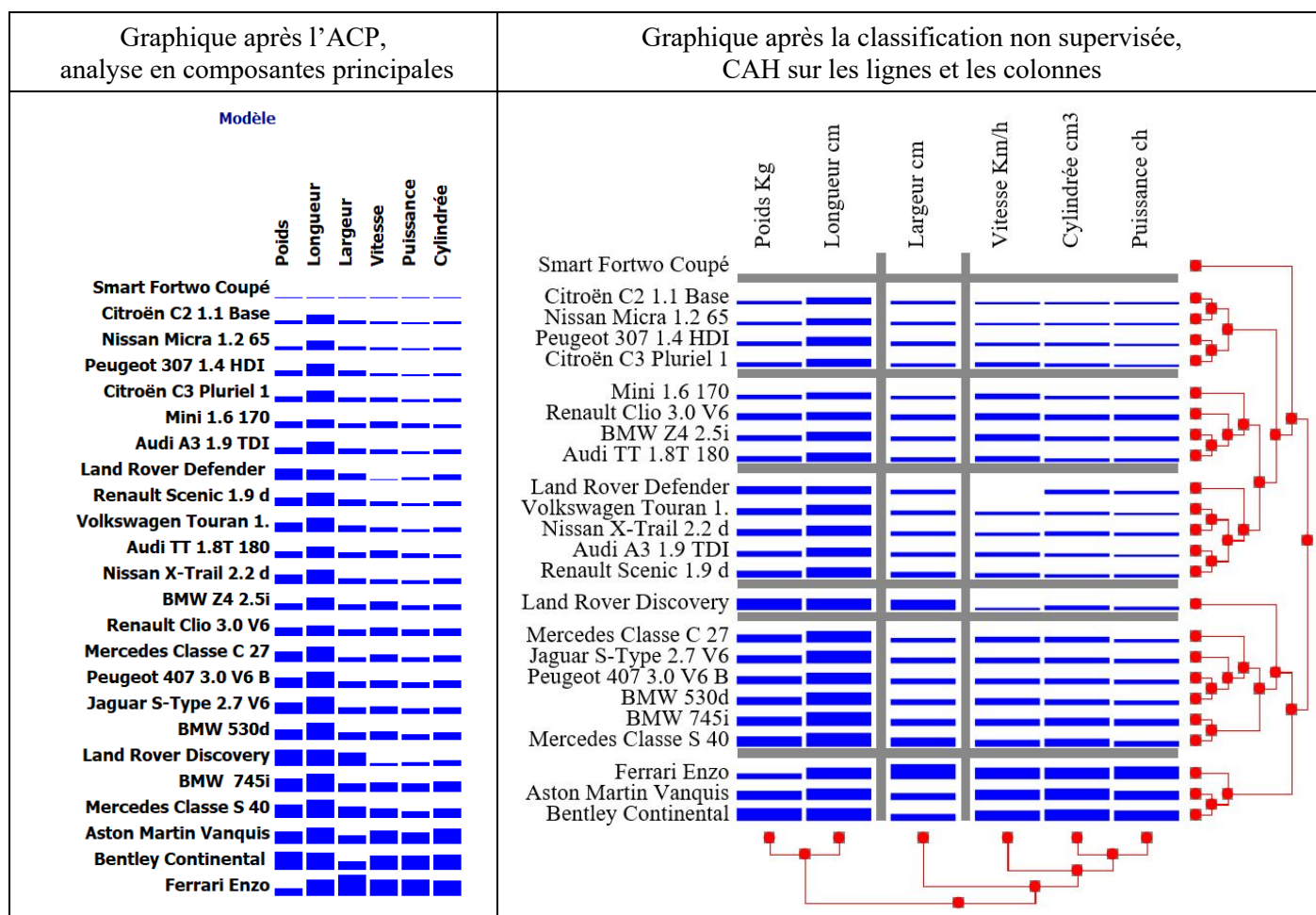
Ces traitements sont adaptés aux tableaux avec des colonnes représentant des variables d'unités différentes. En demandant « **Fichier / Exemples / Fr-Voitures 2004 Tenenhaus.TXT** », AMADO-online affiche le graphique de 24 voitures repérées selon 6 critères : Cylindrée en cm³, Puissance en cheval-vapeur, Vitesse en Km/h, Poids en Kg, Largeur et Longueur en cm.

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

Ici, les colonnes ont des unités différentes (cm³, CV, 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 $(X_{ij} - \min_j) / \sigma_j$; 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 deviendront 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 *longueur* 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).

Et le graphique permet de voir que ce sont les particularités de la *Land Rover Discovery* (lourde, longue et large, mais lente) et de la *Ferrari* (légère, longue et très large, mais très rapide et très puissante) qui expliquent pourquoi la *largeur* est peu corrélée aux quatre autres variables.

5 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)⁴ 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 plus d'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 Philipps Arabie, Scott Boorman et 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 et 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 ésotérisme même ... De leur côté, les arbres de classification donnent une représentation utile mais déformée (*ultramétrique*⁵) 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 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.

⁴ 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).

⁵ Une distance est *ultramétrique* si tous les triangles sont isocèles, le 3^e 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*.

6 Éléments de bibliographie

- 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>
- Caraux 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=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>
- 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 Annexes

7.1 Annexe 1: Format de graphiques PNG ou SVG ?

La commande « **Fichier / Exporter en SVG** » copie le graphique, au format SVG (*Scalable Vector Graphic*), dans le dossier "Téléchargements" de l'ordinateur. On verra à l'annexe 6.2 la façon de recadrer un graphique SVG qui a été collé dans Word, Excel ou PowerPoint.

La commande « **Fichier / Exporter en PNG** » copie le graphique, au format PNG (*Portable Network Graphics*), dans le dossier "Téléchargements" de l'ordinateur.

Le SVG et le PNG sont tous deux un type de format d'image permettant de stocker des images. SVG est un format d'image vectoriel dans lequel une image est représentée par un ensemble de figures mathématiques et PNG est un format d'image binaire qui utilise un algorithme de compression sans perte pour représenter l'image sous forme de pixels.

Voici les différences importantes entre le SVG et le 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 are made up of paths and shapes.	PNG images are 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>



L'annexe 7.2 explique comment remodeler de tels graphiques collés en Word, Excel ou PowerPoint

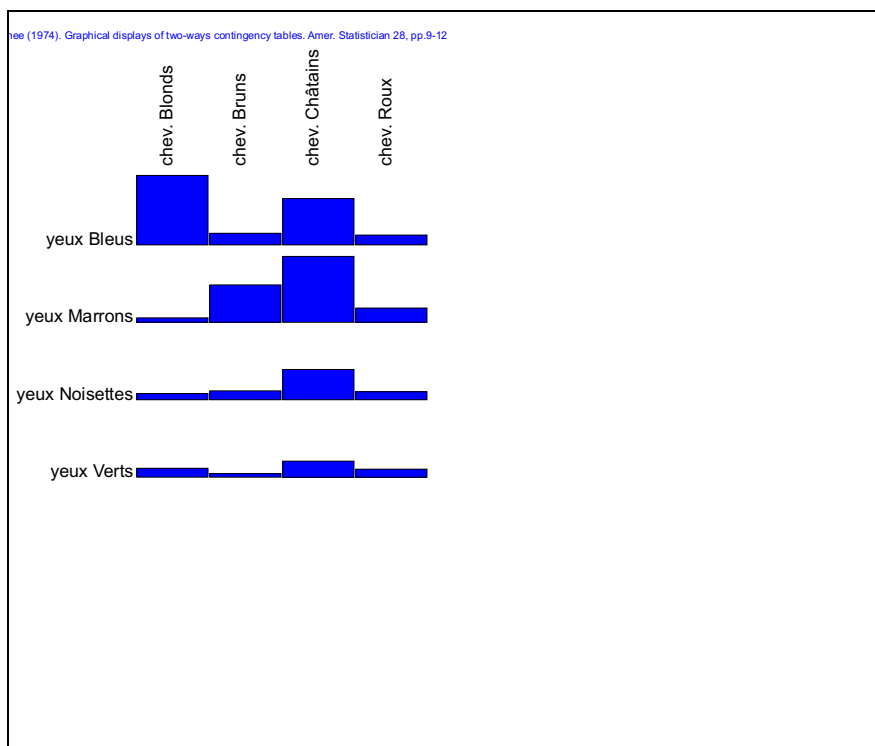
7.2 Annexe 2: Comment rogner un graphique SVG qui a été collé dans Word, Excel ou PowerPoint ?

Quand le graphique a été collé dans Word, Excel, PowerPoint, etc., **cliquer sur sa frontière** pour le sélectionner, puis **Clic droit** pour ouvrir la boîte de dialogue et **cliquer tout en bas** sur **Format de l'image** ;

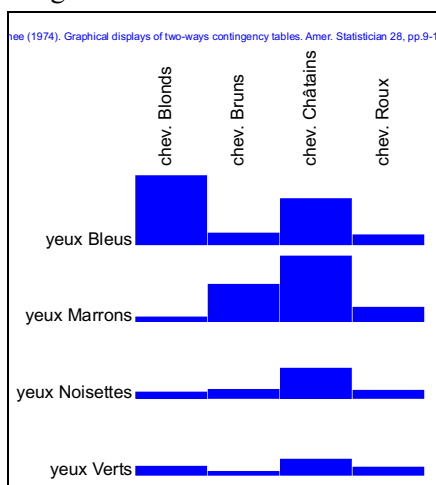
Format de l'image



on obtient  **Cliquer sur l'icône de droite** , puis sur la ligne « **Rogner** » ; on peut alors modifier les largeurs des marges de droite et du bas en changeant les valeurs de **Position du rognage**, **Largeur**, **Hauteur** qui sont en bas de la boîte de dialogue.



On obtient le graphique suivant. On peut alors modifier la taille de cette image en cliquant sur le bord et en bougeant la souris.



7.3 Annexe 3 : 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 ?

Tenenhaus et Young (1985) ont montré qu'il y a de multiples façons de comprendre l'Analyse Factorielle des Correspondances (AFC) 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.

Pour illustrer l'AFC, partons d'une matrice croisant, pour 16 personnes, les couleurs des yeux et des cheveux.

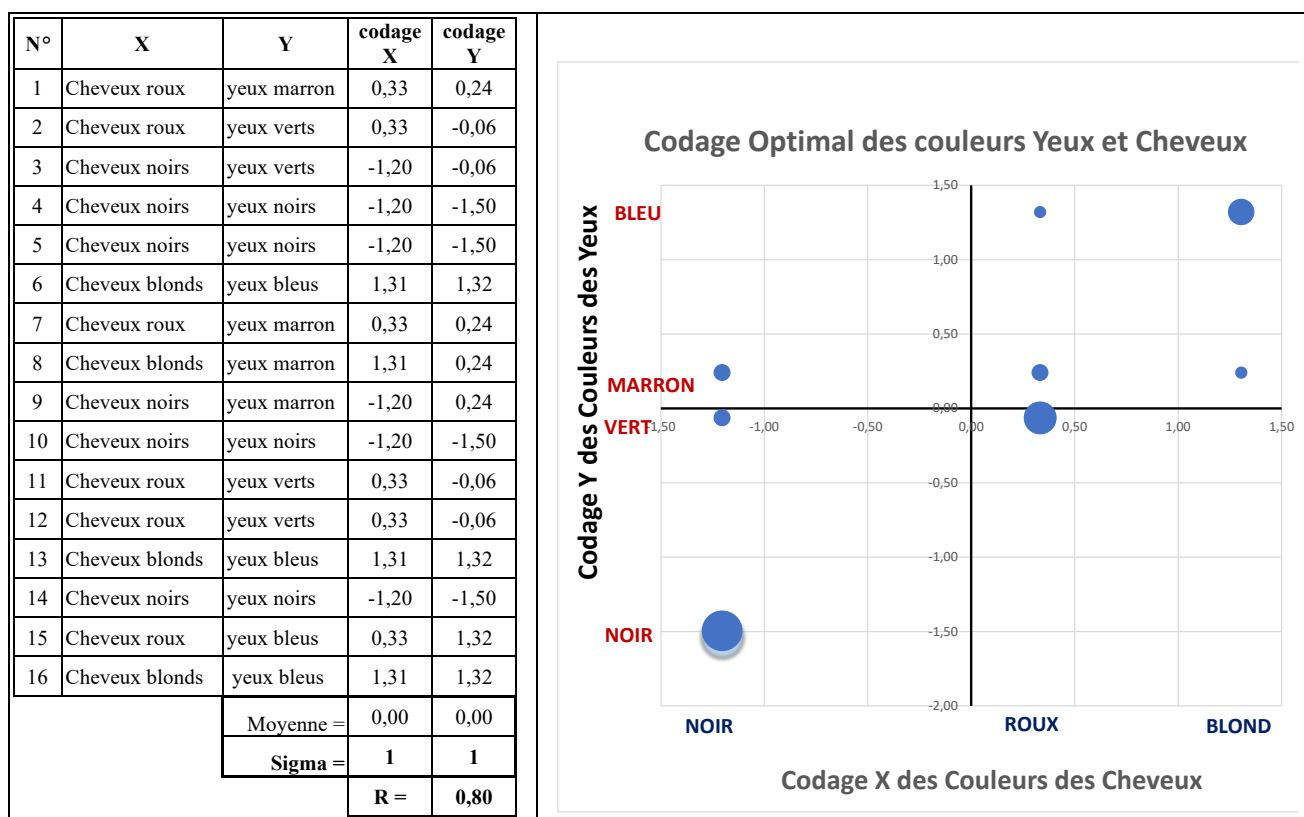
Ces données peuvent aussi se mettre sous la forme d'un tableau du type « base de données » avec l'indicatif (N°) de l'individu et deux champs qualitatifs : la couleur de ses cheveux et la couleur de ses yeux.

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

	cheveux Blonds	cheveux Noirs	cheveux Roux
yeux Bleus	3	0	1
yeux Marron	1	1	2
yeux Noirs	0	4	0
yeux Verts	0	1	3

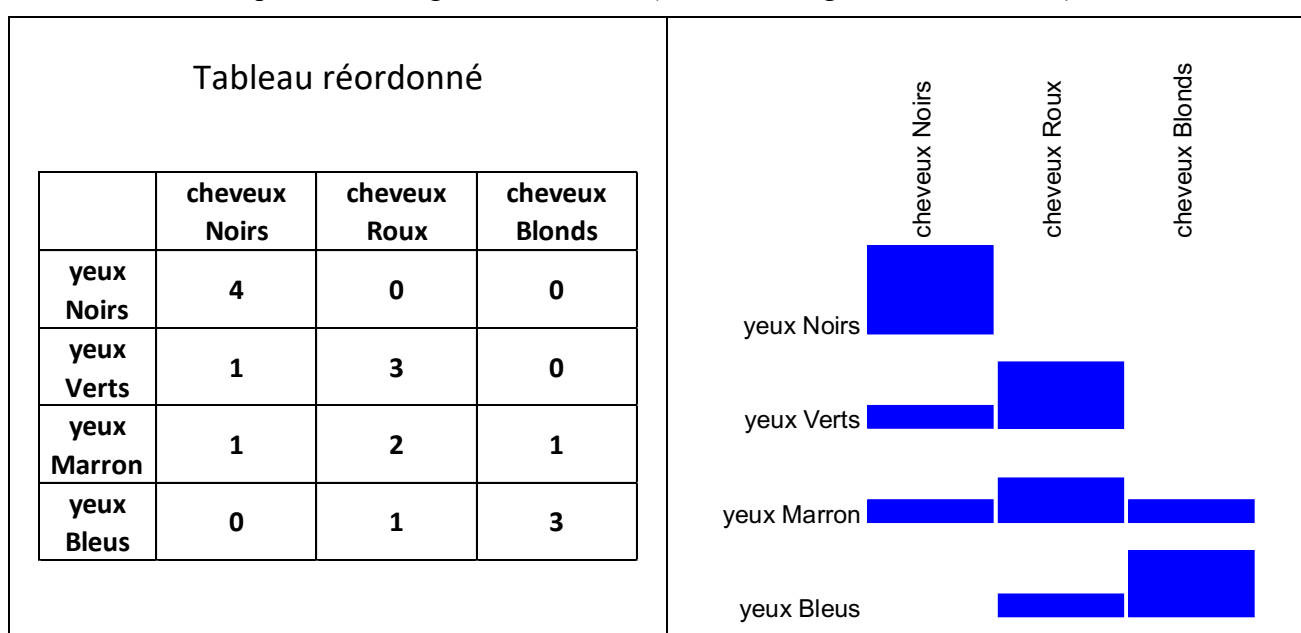
Un codage numérique est un ensemble de valeurs numériques associées aux modalités de chacune des deux variables qualitatives (« codage X » et « codage Y » dans le tableau ci-dessous).

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^{er} facteur de l'Analyse des Correspondances.

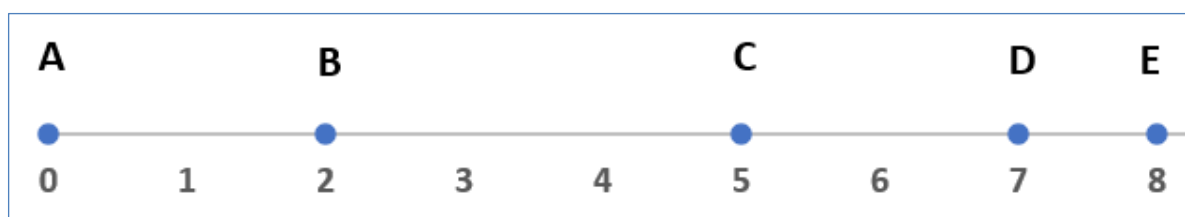
En réordonnant les lignes et les colonnes du tableau selon les valeurs de leur codage numérique, *on obtient le graphique qui montre le mieux la structure des données*, c'est-à-dire les liaisons entre les deux variables qualitatives, lignes et colonnes (l'ordre des lignes est ici inversé) :



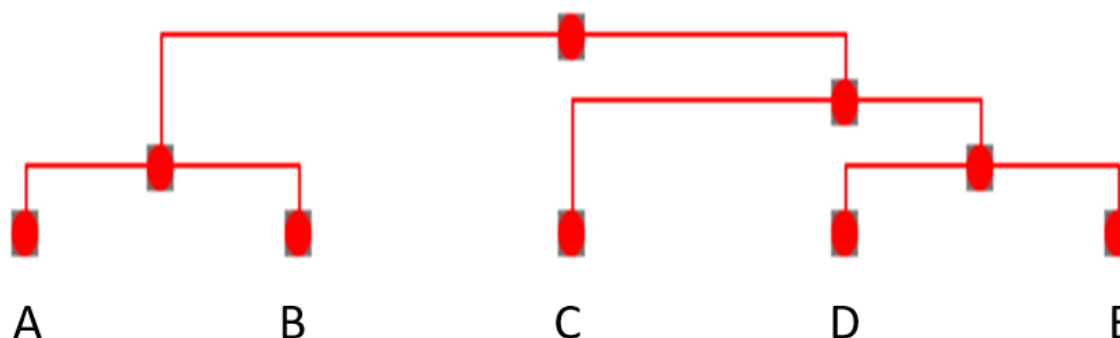
7.4 Annexe 4 : L'arbre de classification hiérarchique montre une réalité déformée

Toute classification simplifie la réalité au prix d'une déformation de celle-ci. Les graphiques produits par *AMADO-online* permettent de bénéficier de la simplification tout en restant fidèle aux données.

En regardant un arbre de classification, la distance perçue entre deux éléments est la hauteur jusqu'à laquelle il faut monter dans l'arbre pour trouver le nœud qui réunit ces deux éléments. Illustrons cela par un exemple simple avec 5 points placés sur un axe :



La classification hiérarchique va produire un arbre du type ci-dessous ; pour réunir B et C il faut aller tout en haut de l'arbre, comme pour réunir A et E ; on a donc l'impression que B et C sont aussi loin l'un de l'autre que le sont A et E. Ceci ne rend pas du tout compte des distances entre nos 5 points sur l'axe !



La notion de distance ultramétrique rend compte de cette déformation : *tous les triangles ont deux côtés égaux et plus longs que le 3^e côté (ou bien les trois côtés sont égaux)*. On retrouve bien cette propriété dans la matrice des distances ci-dessous, correspondant à la hauteur jusqu'à laquelle il faut monter dans l'arbre pour trouver le nœud qui réunit deux éléments : on a réuni d'abord D et E (au niveau 1), puis A et B (au niveau 2) puis C à (D, E) (au niveau 3), puis (A, B) à (C, D E) (au niveau 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

Cette situation « ultramétrique » ne correspond à aucune situation géométrique usuelle à partir de 3 points distincts sur un axe, ou de 4 points distincts sur un plan, etc.

L'arbre de classification donne une vue déformée de la réalité ; comme dans les exemples présentés dans ce guide, les graphiques de *AMADO-online* permettent de bénéficier à la fois des réordonnances des lignes et colonnes selon la classification, et à la fois de la restitution des données originelles.

8 Remerciements

Nous remercions chaleureusement :

- l'équipe du consortium Paris Time Machine qui a soutenu ce travail : Jean-Luc Pinol, Hélène Noizet, Paul Rouet, Laurent Costa, Julien Avinain, Éric Mermet et le conseil scientifique,
- nos collègues universitaires qui nous ont aidé dans la traduction des menus : Annie Morin, Alexis Rouet, Annie Le Gloahec et Jairo Cugliari pour l'espagnol, Linda Gattuso pour l'italien, Iryna Zolotariova pour l'ukrainien et Olena Orobinska-Goncharova pour le russe,
- Jean Dumais pour ses relectures attentives du texte, des menus et des exemples proposés, la traduction en anglais et ses nombreux conseils d'amélioration,
- Basu Tallur pour son assistance à la création de la vidéo,
- Sylvain Clément qui a prêté sa voix pour la version anglaise de la vidéo,
- Alban Risson qui, alors étudiant en Licence d'informatique, a réalisé la première version d'Amado.