

Cours 6

CLASSIFICATION

Master 1 – 2007/2008

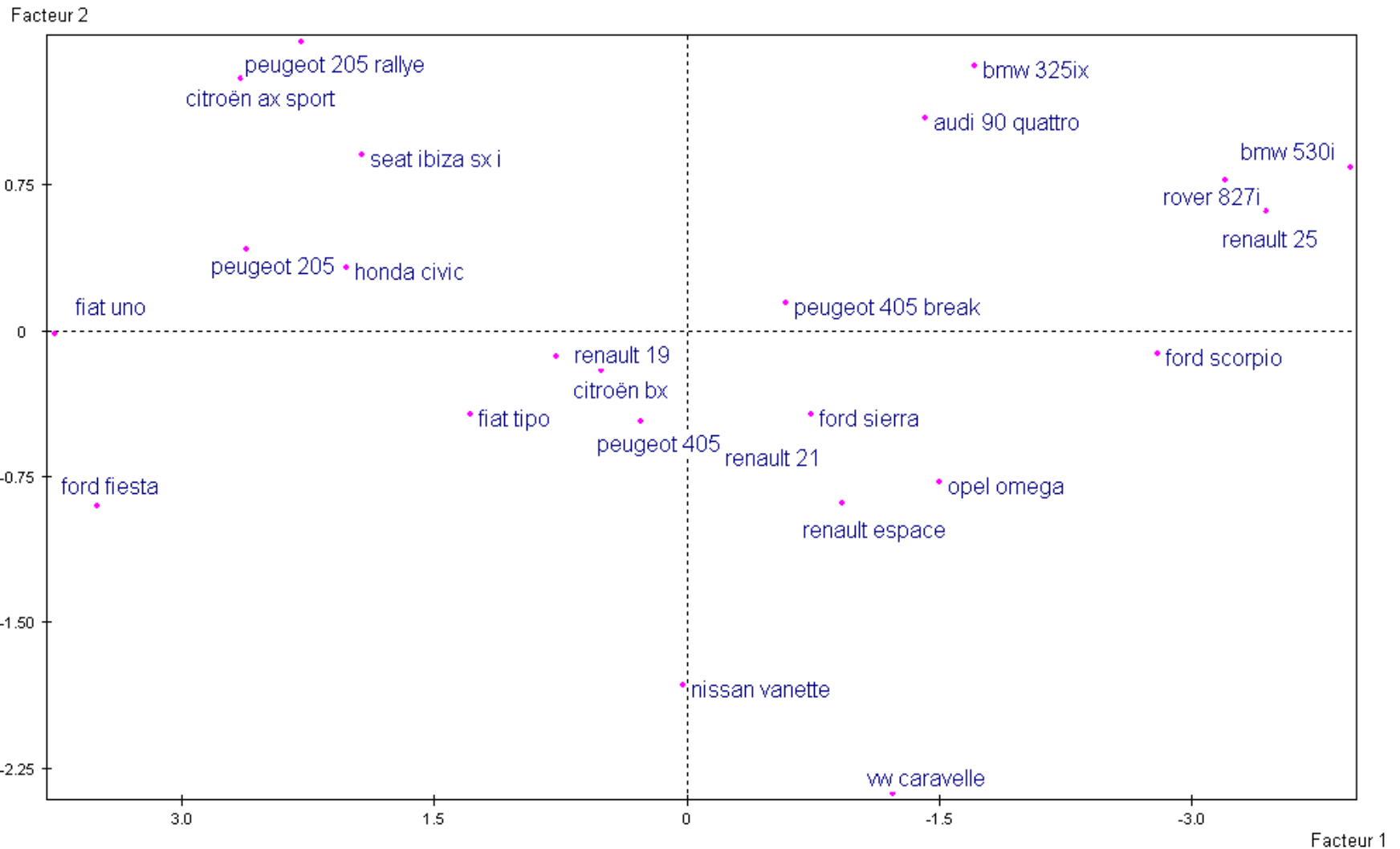


Les données

Modèle	Cylindrée	Puissance	Vitesse	Poids	Longueur	Largeur
Honda Civic	1396	90	174	850	369	166
Renault 19	1721	92	180	965	415	169
Fiat Tipo	1580	83	170	970	395	170
Peugeot 405	1769	90	180	1080	440	169
Renault 21	2068	88	180	1135	446	170
Citroën BX	1769	90	182	1060	424	168
BMW 530i	2986	188	226	1510	472	175
Rover 827i	2675	177	222	1365	469	175
Renault 25	2548	182	226	1350	471	180
Opel Omega	1998	122	190	1255	473	177
Peugeot 405 Break	1905	125	194	1120	439	171
Ford Sierra	1993	115	185	1190	451	172
BMW 325ix	2494	171	208	1300	432	164
Audi 90 Quattro	1994	160	214	1220	439	169
Ford Scorpio	2933	150	200	1345	466	176
Renault Espace	1995	120	177	1265	436	177
Nissan Vanette	1952	87	144	1430	436	169
VW Caravelle	2109	112	149	1320	457	184
Ford Fiesta	1117	50	135	810	371	162
Fiat Uno	1116	58	145	780	364	155
Peugeot 205	1580	80	159	880	370	156
Peugeot 205 Rallye	1294	103	189	805	370	157
Seat Ibiza SXI	1461	100	181	925	363	161
Citroën AX Sport	1294	95	184	730	350	160

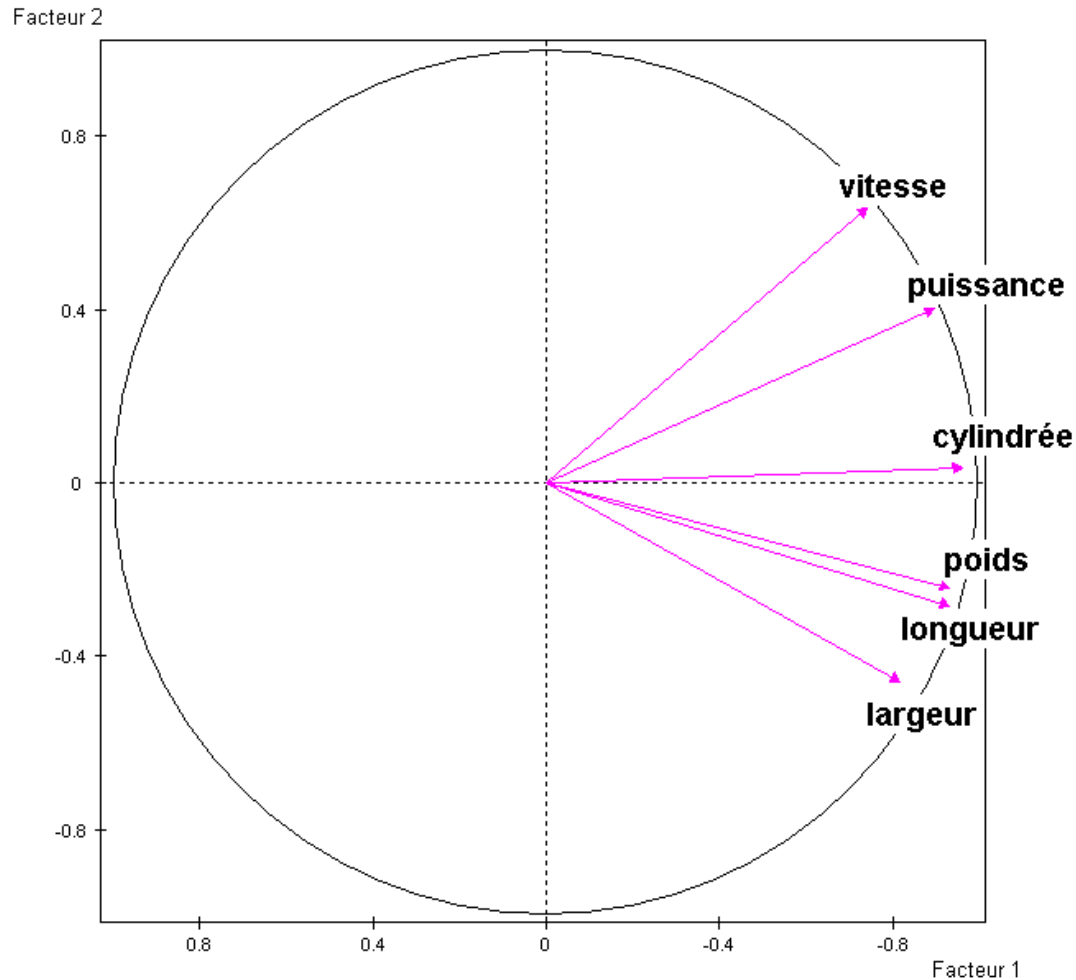
Exemple Auto88

Le premier plan principal



Exemple Auto88

Le cercle des corrélations



La classification :

*« Qui se ressemble
s'assemble »*

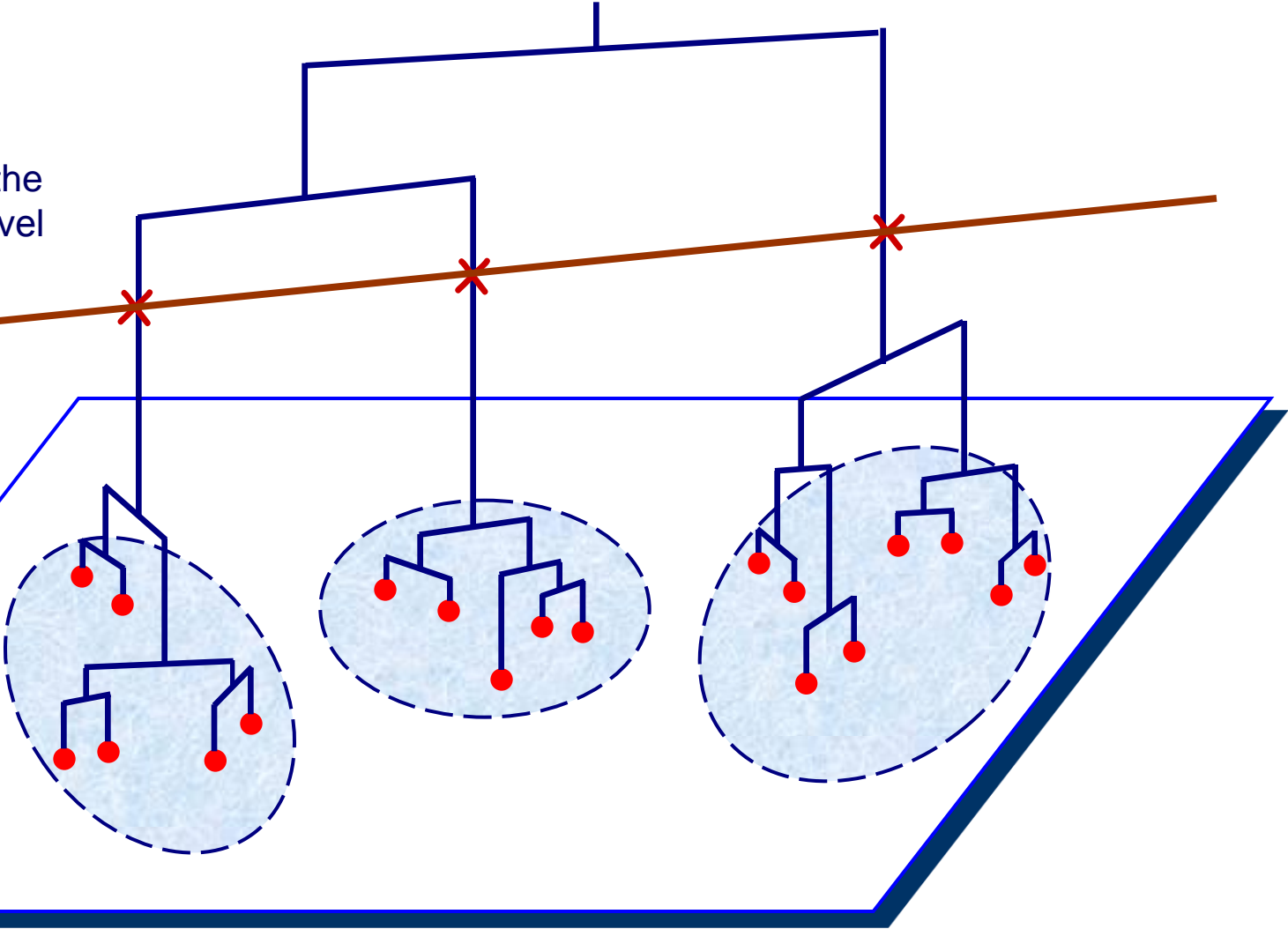
Dendrogramme

1  groupes

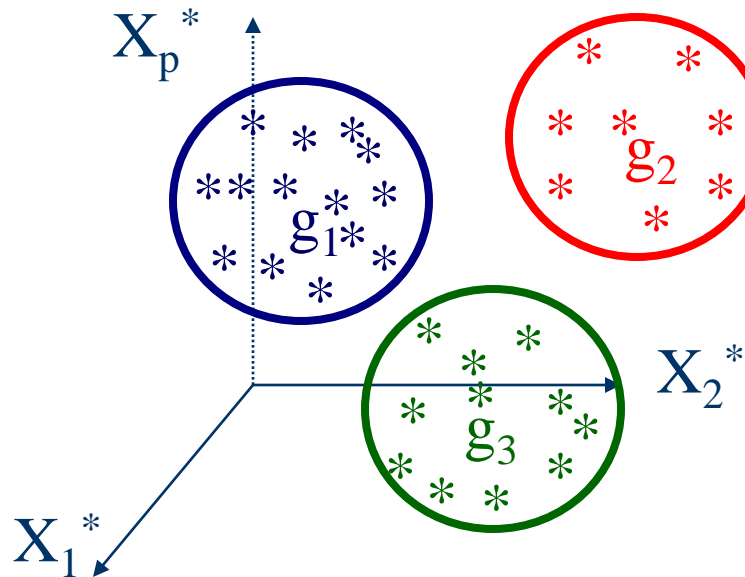


Choosing the "cutting" level

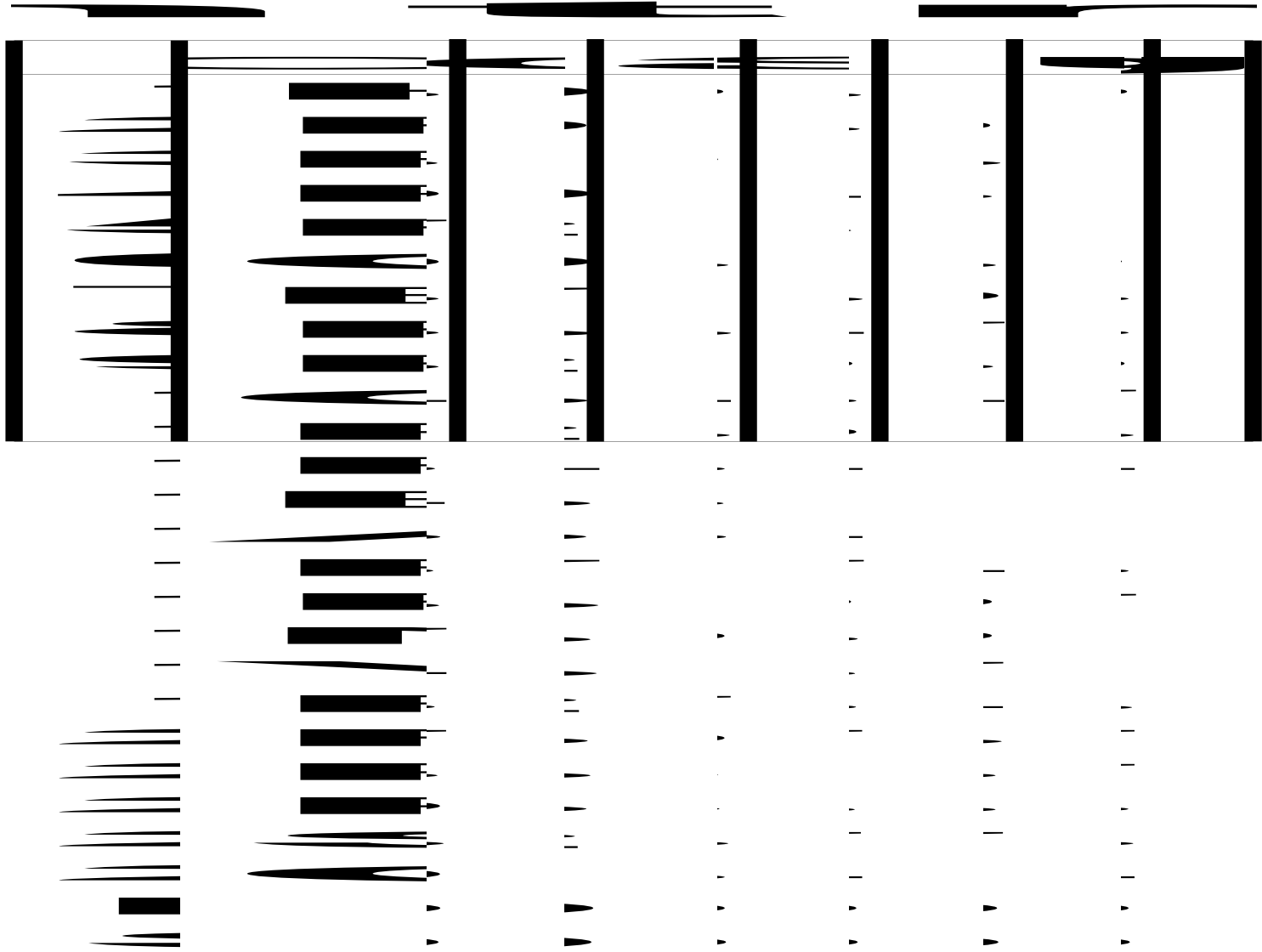
Definition of the clusters



Classification ascendante hiérarchique (Méthode de Ward)



$$\text{Distance de Ward : } D(G_i, G_j) = \frac{n_i n_j}{(n_i + n_j)} d^2(g_i, g_j)$$



Calculer la distance de Ward entre la BX et la 405 ?

Tableau des distances entre les voitures

The diagram illustrates a grid of cars with arrows indicating distances between them. The cars are arranged in a grid, and arrows point from one car to another, representing the distance between them. The grid is composed of several rows and columns of cars, with arrows connecting them in a systematic pattern.

$$D(\text{Citroën BX}, \text{Peugeot 405}) = \frac{1 \times 1}{(1+1)} \times .181 = .0905$$

Classification Ascendante Hiérarchique

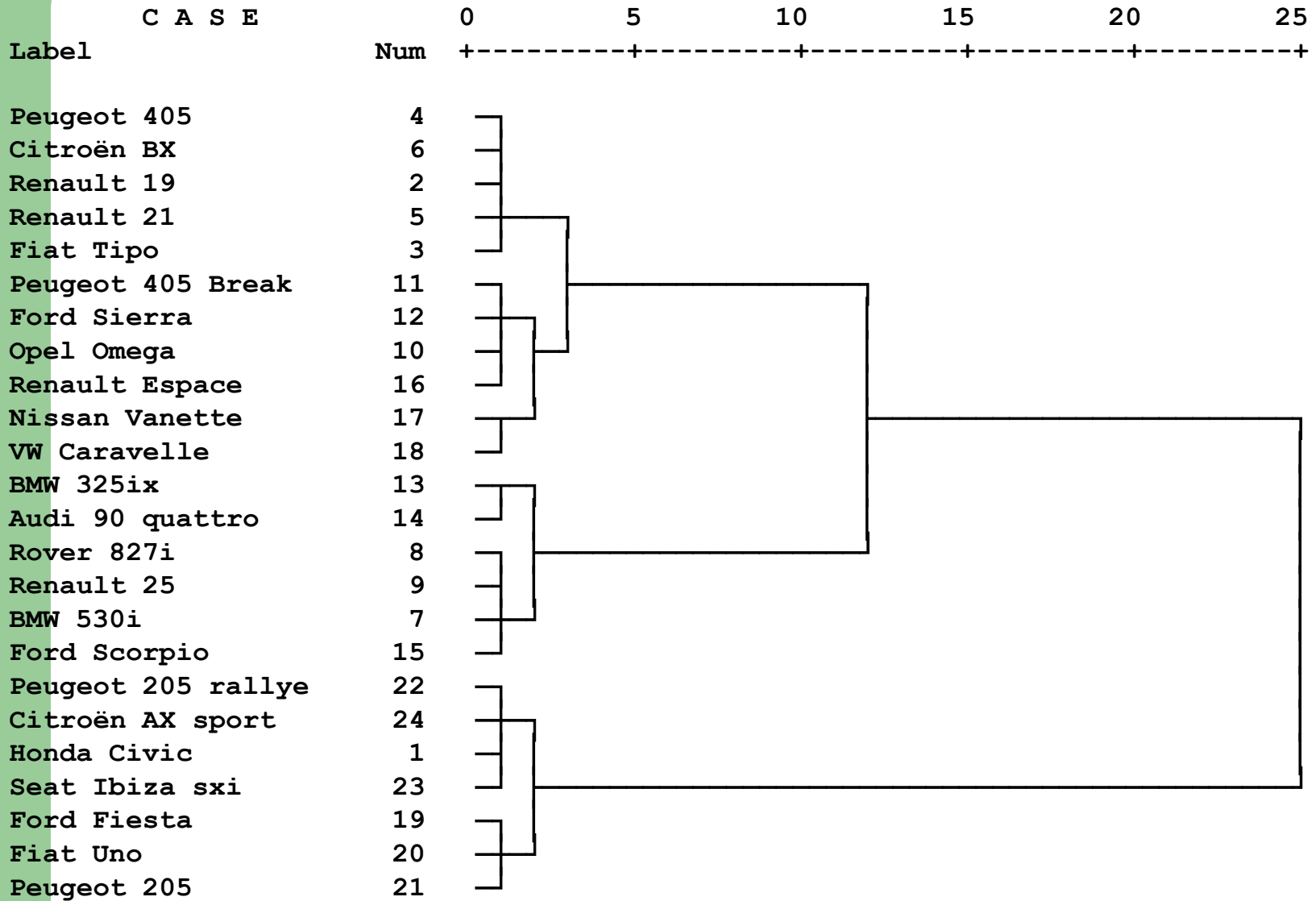
Etape initiale

__ Chaque individu forme une classe. On regroupe les deux individus les plus proches.

Etape courante

A chaque étape, on regroupe les deux classes G_i et G_j minimisant le critère de Ward $D(G_i, G_j)$.

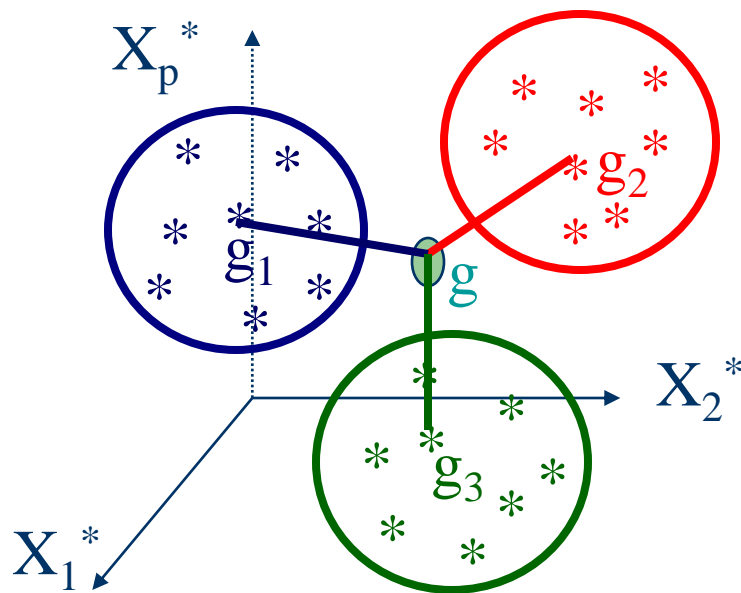
Rescaled Distance Cluster Combine



Qualité de la typologie en K classes

- La somme des carrés expliquée par la typologie en K classes est égale à la somme des carrés inter-classes de la typologie en K classes.
- La qualité de la typologie est mesurée par la part de la somme des carrés totale expliquée par la typologie.

Décomposition de l'inertie totale



$$I(N, g) = \sum_{i=1}^K \frac{n_i}{n} d^2(g_i, g) + \sum_{i=1}^K \frac{n_i}{n} I(G_i, g_i)$$

Inertie totale = p

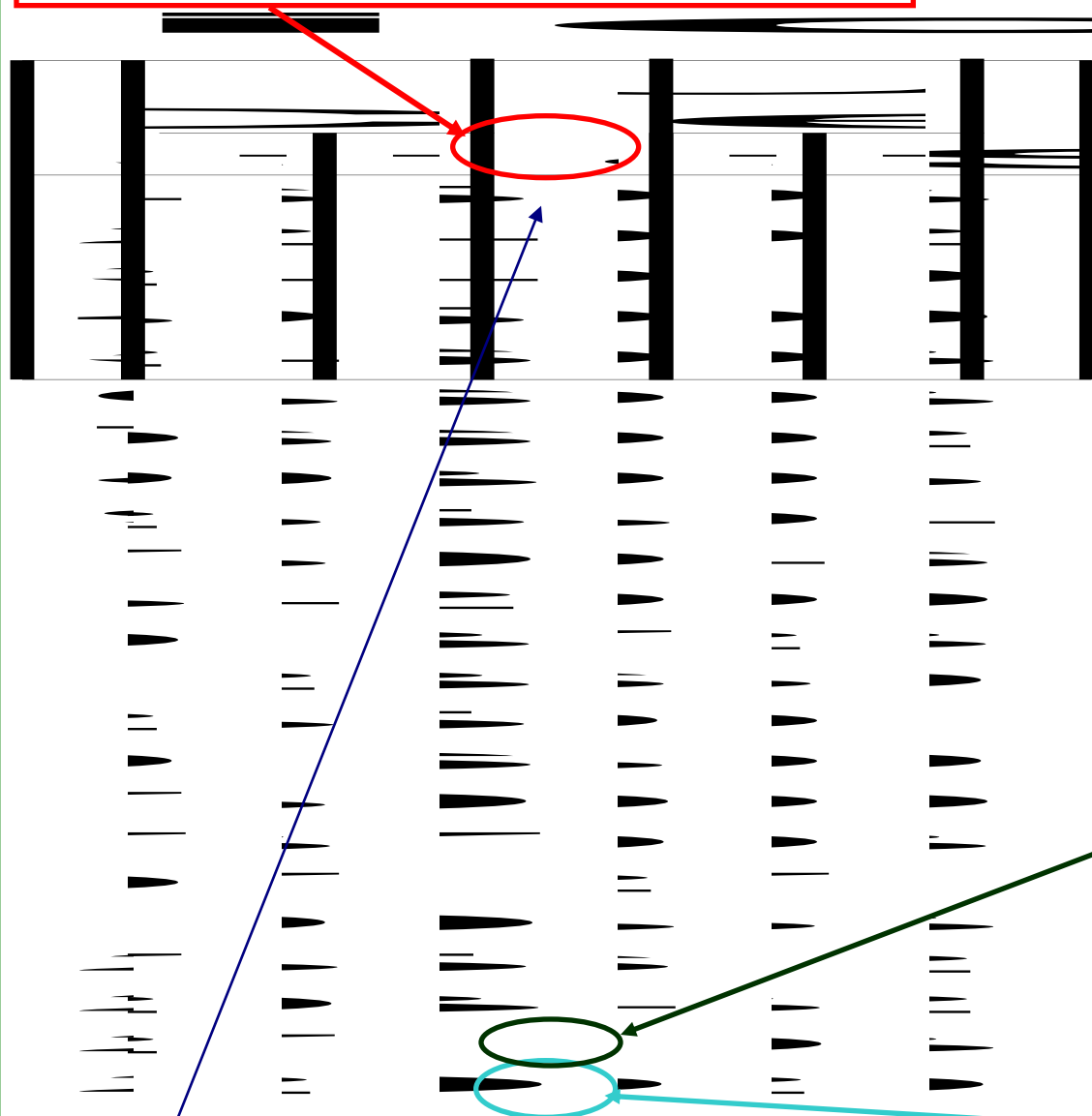
Inertie inter-classes

+

Inertie intra-classes

Coefficient : Somme des carrés
intra-classes de la typologie en K classes

Résultats SPSS :
Somme des carrés intra-classes



Part de somme des carrés
totale expliquée par la
typologie en K classes :
 $(138 - \text{Coeff}[n-K])/138$

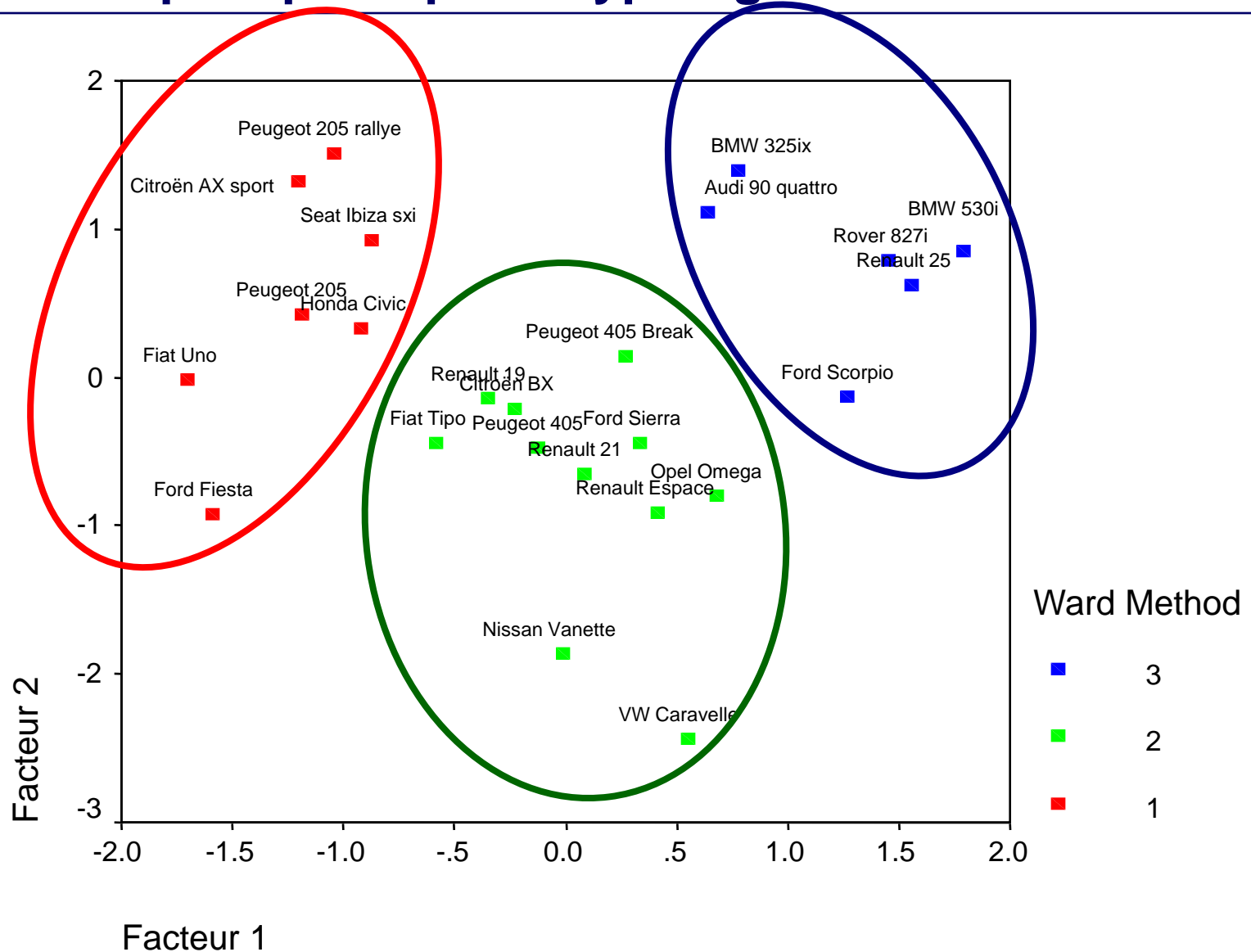
Part de somme des carrés
totale expliquée par la
typologie en 2 classes :
 $(138 - 67.344)/138 = 0.512$

Somme des carrés
intra-classes pour
la typologie en K=2 classes

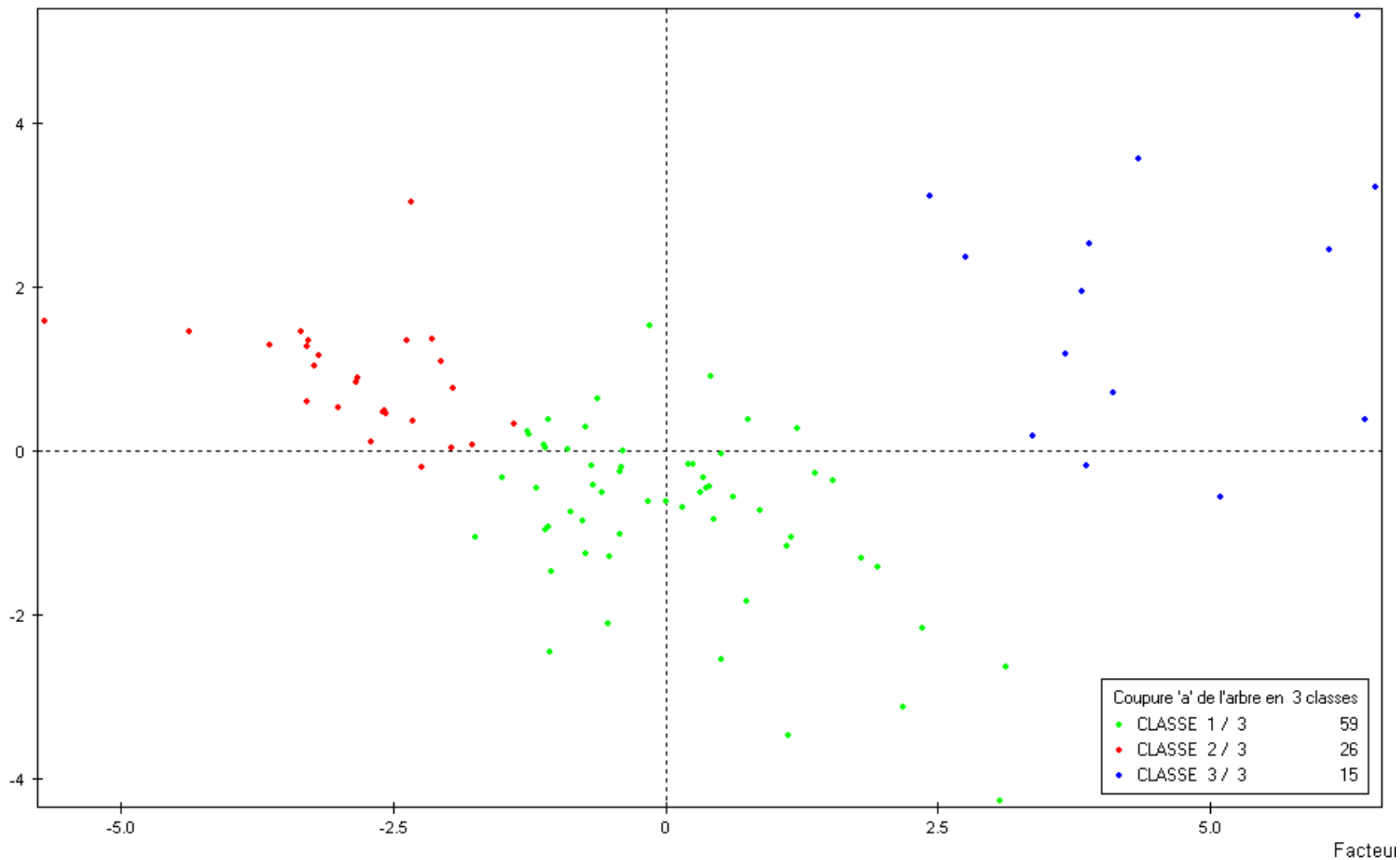
Somme des carrés
totale = $p*(n-1)$

14 Distance de Ward(4,6)

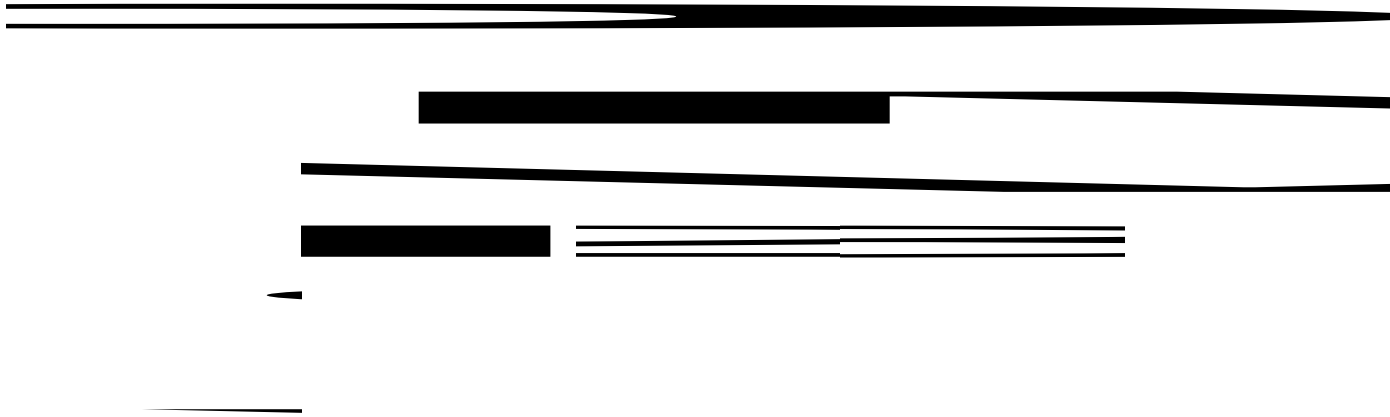
Premier plan principal et typologie



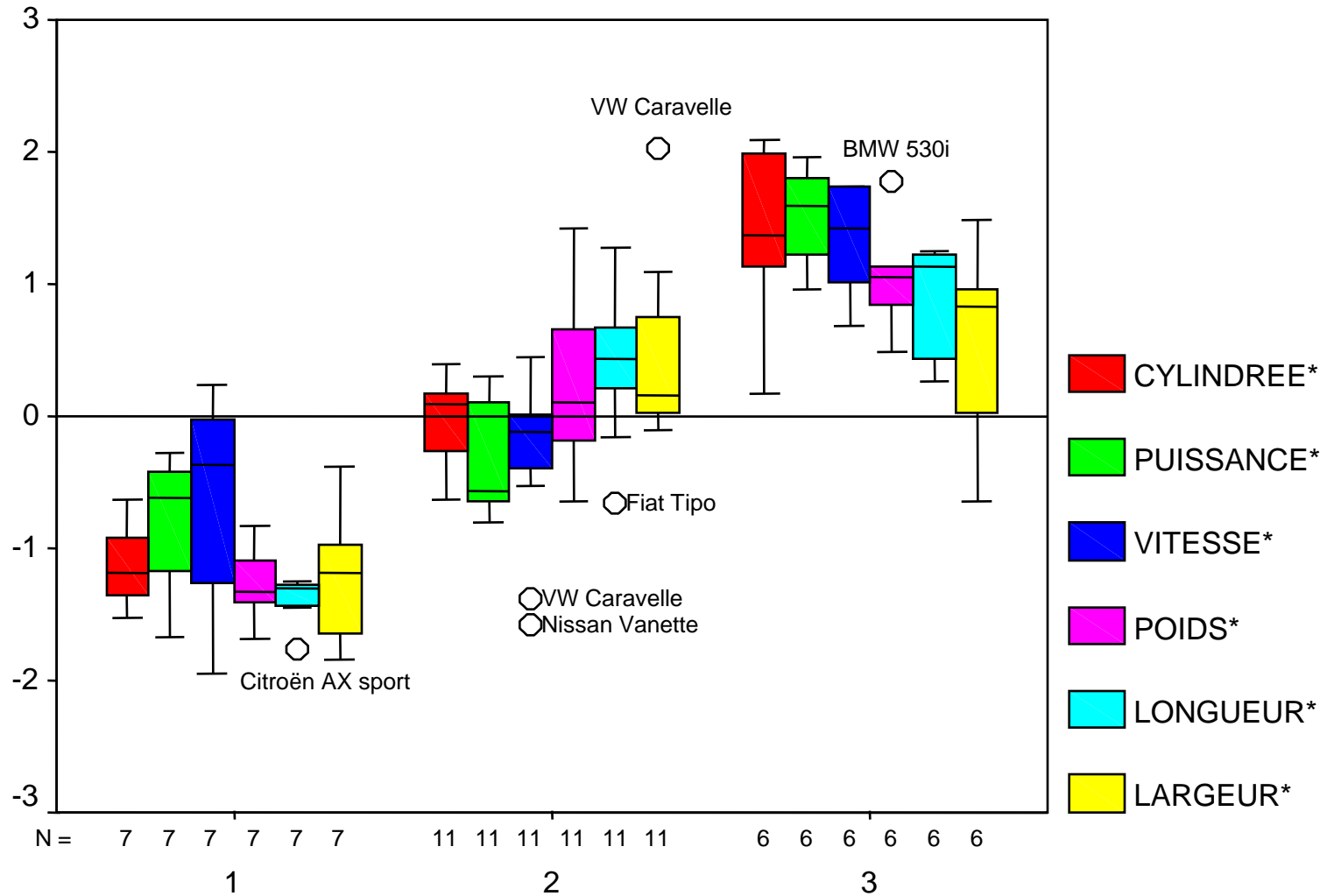
Facteur 2



Interprétation des classes



Interprétation des classes



Ward Method