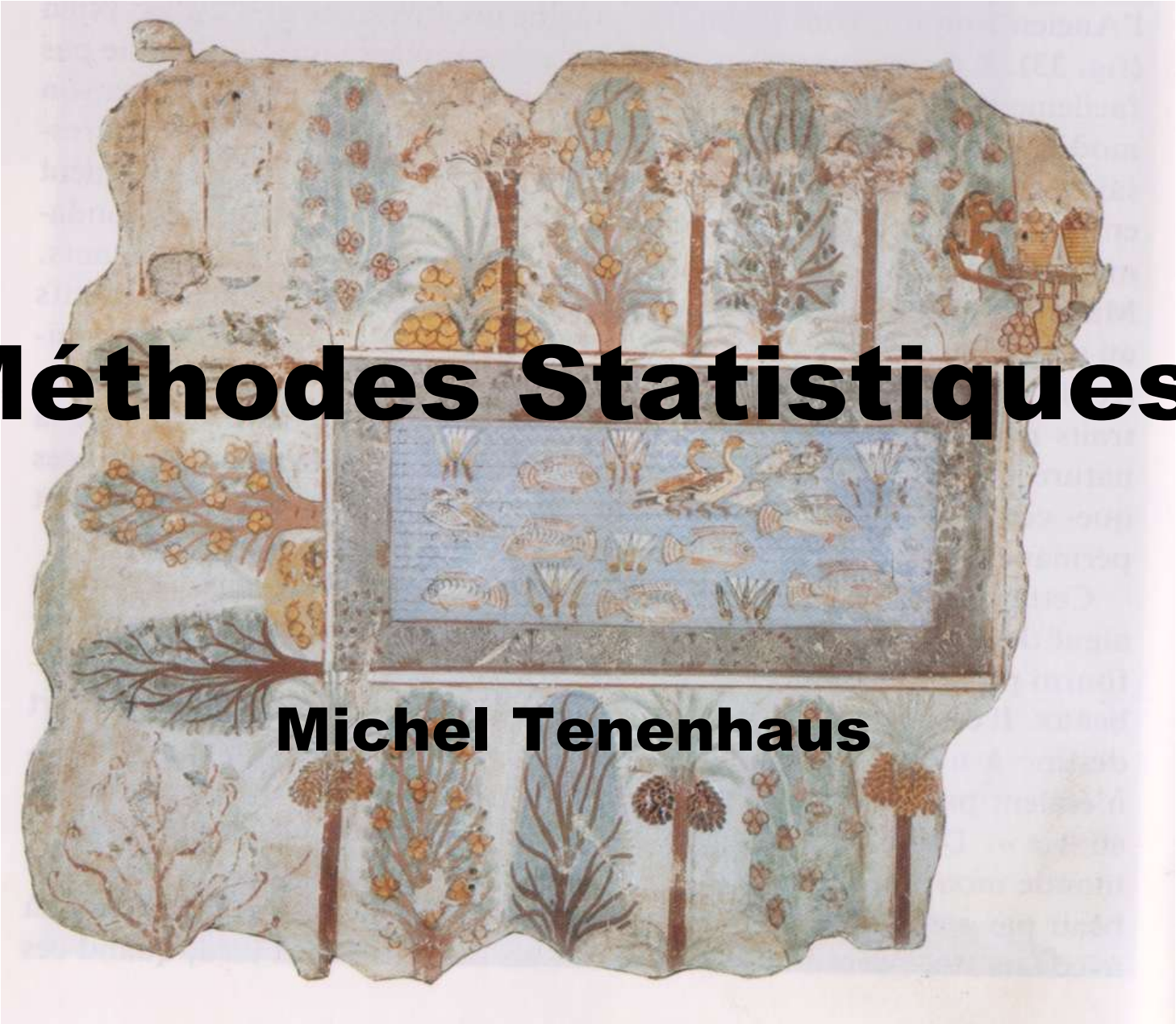


Méthodes Statistiques

Michel Tenenhaus

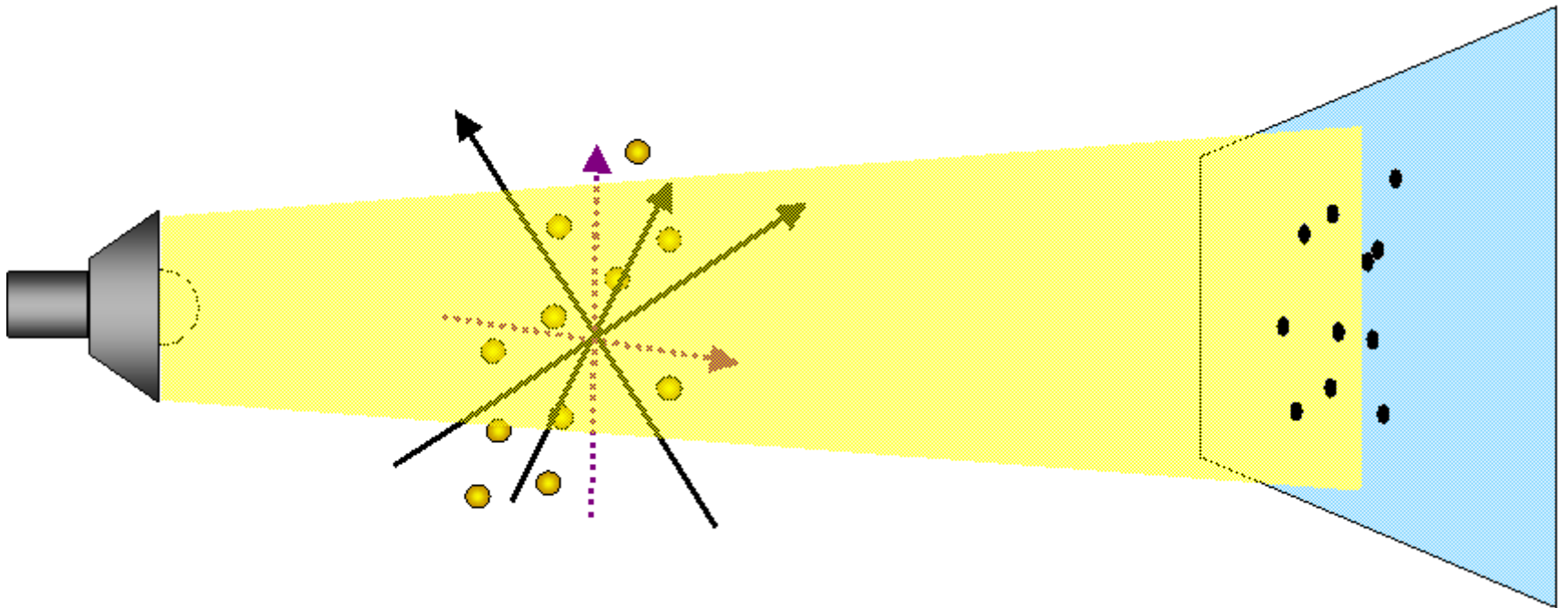


STATISTIQUE ?



- ⌘ Extraire des connaissances à partir de données pour décrire, expliquer, ou prévoir.
 - ☒ Simplifier une réalité complexe à l'aide de graphiques.
 - ☒ Simplifier une réalité complexe à l'aide de modèles mathématiques.
 - ☒ Outils de manipulation de grosses bases de données pour identifier et segmenter la clientèle d'une entreprise et orienter ses choix stratégiques (data mining).

Décrire ?



Exemple : les races canines

	Race	Taille	Poids	Vitesse	Intell.	Affect.	Agress.	Fonction
1	Beauceron	TA++	PO+	V++	INT+	AF+	AG+	Utilité
2	Basset	TA-	PO-	V-	INT-	AF-	AG+	Chasse
3	Berger-Allemand	TA++	PO+	V++	INT++	AF+	AG+	Utilité
4	Boxer	TA+	PO+	V+	INT+	AF+	AG+	Compagnie
5	Bull-Dog	TA-	PO-	V-	INT+	AF+	AG-	Compagnie
6	Bull-Mastiff	TA++	PO++	V-	INT++	AF-	AG+	Utilité
7	Caniche	TA-	PO-	V+	INT++	AF+	AG-	Compagnie
8	Chihuahua	TA-	PO-	V-	INT-	AF+	AG-	Compagnie
9	Cocker	TA+	PO-	V-	INT+	AF+	AG+	Compagnie
10	Colley	TA++	PO+	V++	INT+	AF+	AG-	Compagnie
11	Dalmatien	TA+	PO+	V+	INT+	AF+	AG-	Compagnie
12	Doberman	TA++	PO+	V++	INT++	AF-	AG+	Utilité
13	Dogue Allemand	TA++	PO++	V++	INT-	AF-	AG+	Utilité
14	Epagneul Breton	TA+	PO+	V+	INT++	AF+	AG-	Chasse
15	Epagneul Français	TA++	PO+	V+	INT+	AF-	AG-	Chasse
16	Fox-Hound	TA++	PO+	V++	INT-	AF-	AG+	Chasse
17	Fox-Terrier	TA-	PO-	V+	INT+	AF+	AG+	Compagnie
18	Grd Bleu de Gascogne	TA++	PO+	V+	INT-	AF-	AG+	Chasse
19	Labrador	TA+	PO+	V+	INT+	AF+	AG-	Chasse
20	Lévrier	TA++	PO+	V++	INT-	AF-	AG-	Chasse
21	Mastiff	TA++	PO++	V-	INT-	AF-	AG+	Utilité
22	Pékinois	TA-	PO-	V-	INT-	AF+	AG-	Compagnie
23	Pointer	TA++	PO+	V++	INT++	AF-	AG-	Chasse
24	Saint-Bernard	TA++	PO++	V-	INT+	AF-	AG+	Utilité
25	Setter	TA++	PO+	V++	INT+	AF-	AG-	Chasse
26	Teckel	TA-	PO-	V-	INT+	AF+	AG-	Compagnie
27	Terre-Neuve	TA++	PO++	V-	INT+	AF-	AG-	Utilité

Le tableau disjonctif complet

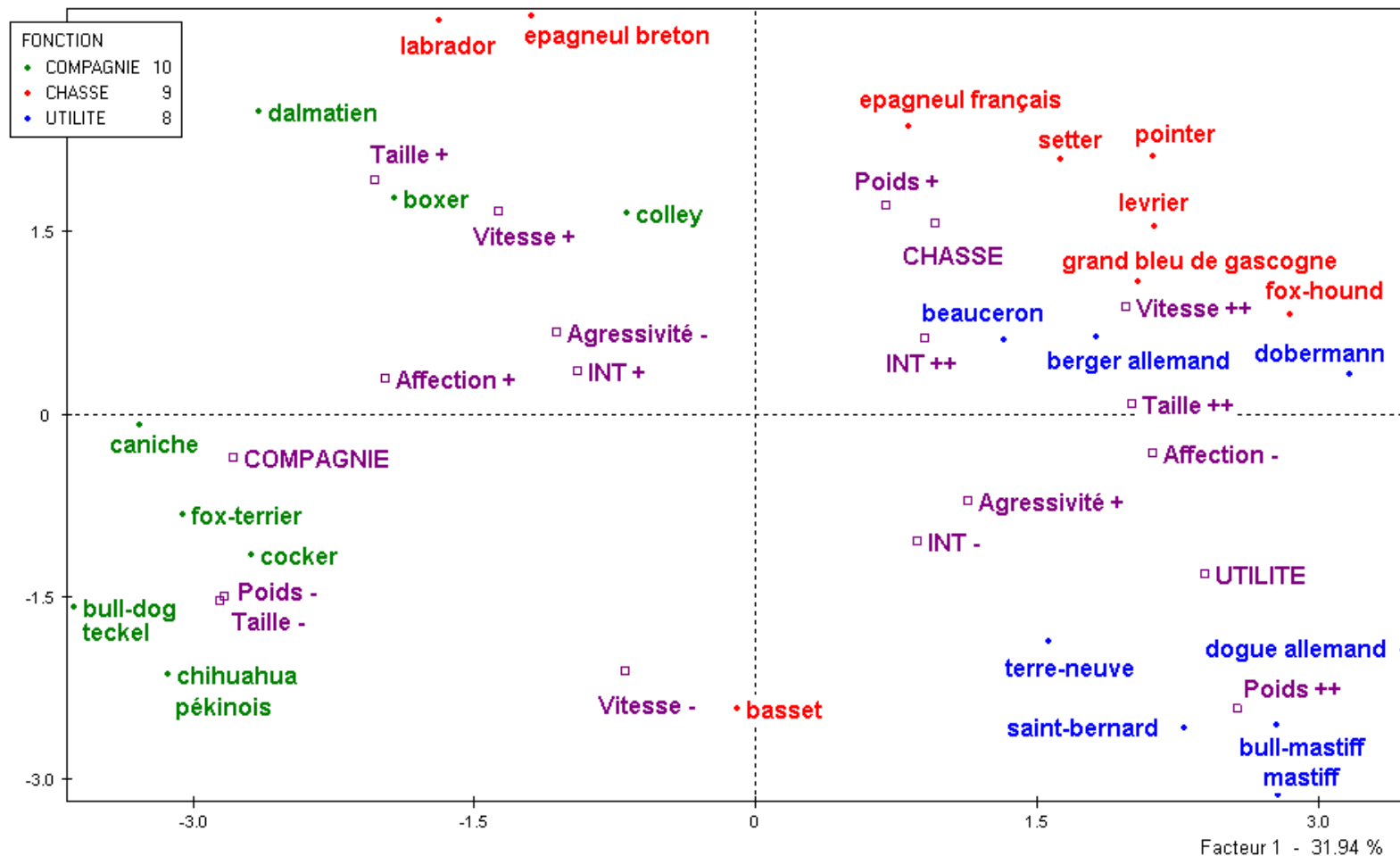
Race	T-	T+	T++	P-	P+	P++	V-	V+	V++	I-	I+	I++	Af-	Af+	Ag-	Ag+	Compagnie	Chasse	Utilité
Beauceron	0	0	1	0	1	0	0	0	1	0	1	0	0	1	0	1	0	0	1
Basset	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0
Berger all	0	0	1	0	1	0	0	0	1	0	0	1	0	1	0	1	0	0	1
Boxer	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	1	1	0	0
Bull-dog	1	0	0	1	0	0	1	0	0	0	1	0	0	1	1	0	1	0	0
Bull Mastiff	0	0	1	0	0	1	1	0	0	0	0	1	1	0	0	1	0	0	1
Caniche	1	0	0	1	0	0	0	1	0	0	0	1	0	1	1	0	1	0	0
Chihuahua	1	0	0	1	0	0	1	0	0	1	0	0	0	1	1	0	1	0	0
Cocker	0	1	0	1	0	0	1	0	0	0	1	0	0	1	0	1	1	0	0
Colley	0	0	1	0	1	0	0	0	1	0	1	0	0	1	1	0	1	0	0
Dalmatien	0	1	0	0	1	0	0	1	0	0	1	0	0	1	1	0	1	0	0
Doberman	0	0	1	0	1	0	0	0	1	0	0	1	1	0	0	1	0	0	1
Dogue all	0	0	1	0	0	1	0	0	1	1	0	0	1	0	0	1	0	0	1
Epagneul br	0	1	0	0	1	0	0	1	0	0	0	1	0	1	1	0	0	1	0
Epagneul fr	0	0	1	0	1	0	0	1	0	0	1	0	1	0	1	0	0	1	0
Fox-Hound	0	0	1	0	1	0	0	0	1	1	0	0	1	0	0	1	0	1	0
Fox-Terrier	1	0	0	1	0	0	0	1	0	0	1	0	0	1	0	1	1	0	0
Grd Bl de G	0	0	1	0	1	0	0	1	0	1	0	0	1	0	0	1	0	1	0
Labrador	0	1	0	0	1	0	0	1	0	0	1	0	0	1	1	0	0	1	0
Lévrier	0	0	1	0	1	0	0	0	1	1	0	0	1	0	1	0	0	1	0
Mastiff	0	0	1	0	0	1	1	0	0	1	0	0	1	0	0	1	0	0	1
Pékinois	1	0	0	1	0	0	1	0	0	1	0	0	0	1	1	0	1	0	0
Pointer	0	0	1	0	1	0	0	0	1	0	0	1	1	0	1	0	0	1	0
St-Bernard	0	0	1	0	0	1	1	0	0	0	1	0	1	0	0	1	0	0	1
Setter	0	0	1	0	1	0	0	0	1	0	1	0	1	0	1	0	0	1	0
Teckel	1	0	0	1	0	0	1	0	0	0	1	0	0	1	1	0	1	0	0
Terre neuve	0	0	1	0	0	1	1	0	0	0	1	0	1	0	1	0	0	0	1

$x_{ijl} = 1$ si l'individu i possède la modalité l de la variable j
 $= 0$ sinon

Analyse factorielle du tableau disjonctif complet

Modalité au barycentre des chiens la possédant

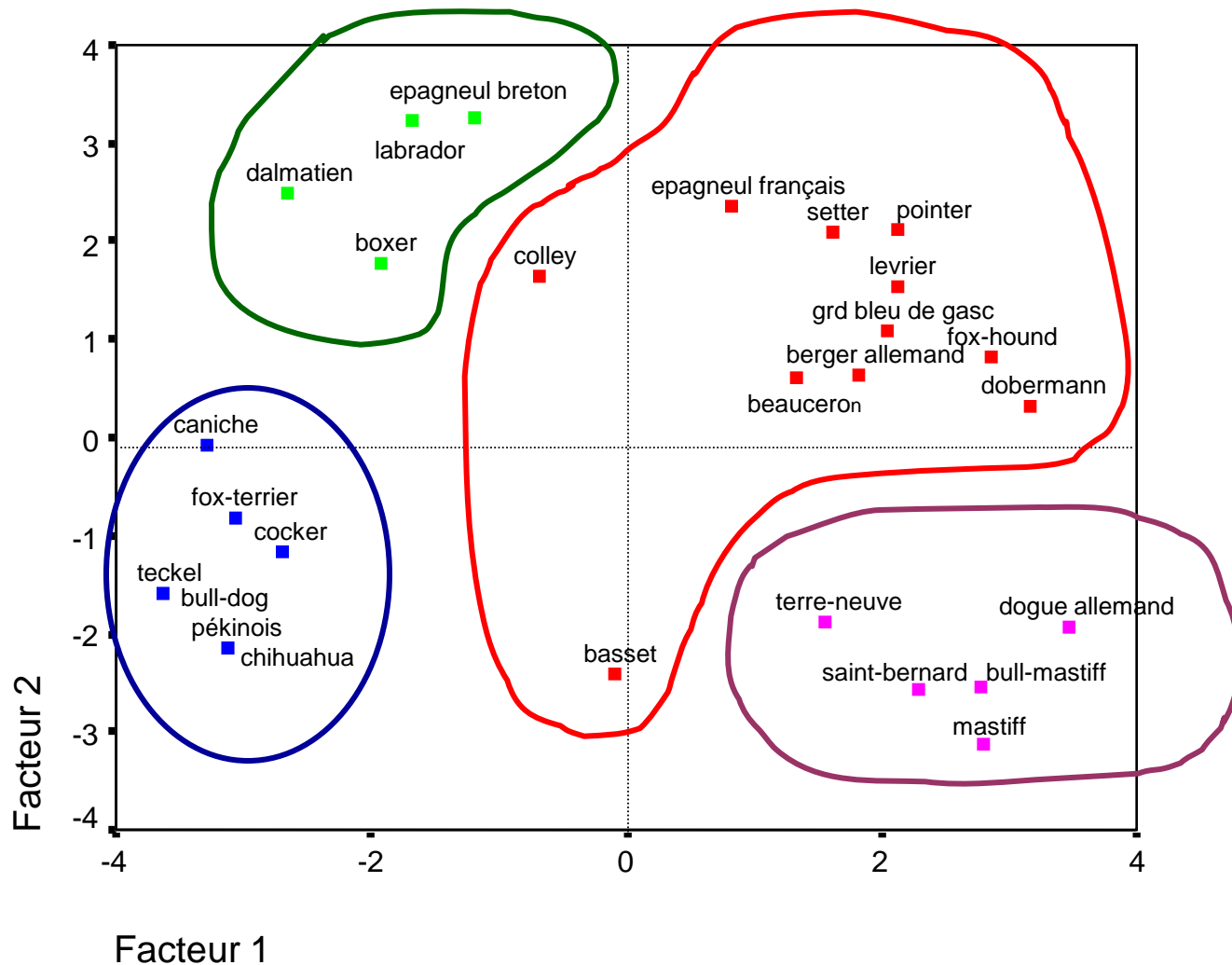
Facteur 2 - 20.49 %



Classification ascendante hiérarchique des chiens (sur le tableau disjonctif complet)

C A S E	0	5	10	15	20	25
Label	Num	+-----+	+-----+	+-----+	+-----+	+-----+
bull-dog	5	↓✕↓				
teckel	26	↓	↓			
chihuahua	8	↓✕↓	↓			
pékinois	22	↓	↓	↓	↓	↓
caniche	7	↓	↓	↓	↓	↓
cocker	9	↓	↓	↓	↓	↓
fox-terrier	17	↓	↓	↓		
epagneul breton	14	↓	↓	↓	↓	↓
labrador	19	↓	↓	↓	↓	↓
boxer	4	↓	↓	↓	↓	↓
dalmatien	11	↓				
dogue allemand	13	↓	↓	↓	↓	↓
mastiff	21	↓	↓	↓	↓	↓
saint-bernard	24	↓	↓	↓	↓	↓
terre-neuve	27	↓	↓	↓	↓	↓
bull-mastiff	6	↓	↓	↓	↓	↓
berger allemand	3	↓	↓	↓	↓	↓
dobermann	12	↓	↓	↓	↓	↓
beauceron	1	↓	↓	↓	↓	↓
pointer	23	↓				
setter	25	↓	↓	↓	↓	↓
levrier	20	↓	↓	↓	↓	↓
epagneul français	15	↓	↓	↓	↓	↓
colley	10	↓	↓	↓	↓	↓
fox-hound	16	↓	↓	↓	↓	↓
grd bleu de gasc	18	↓	↓	↓	↓	↓
basset	2	↓	↓	↓	↓	↓

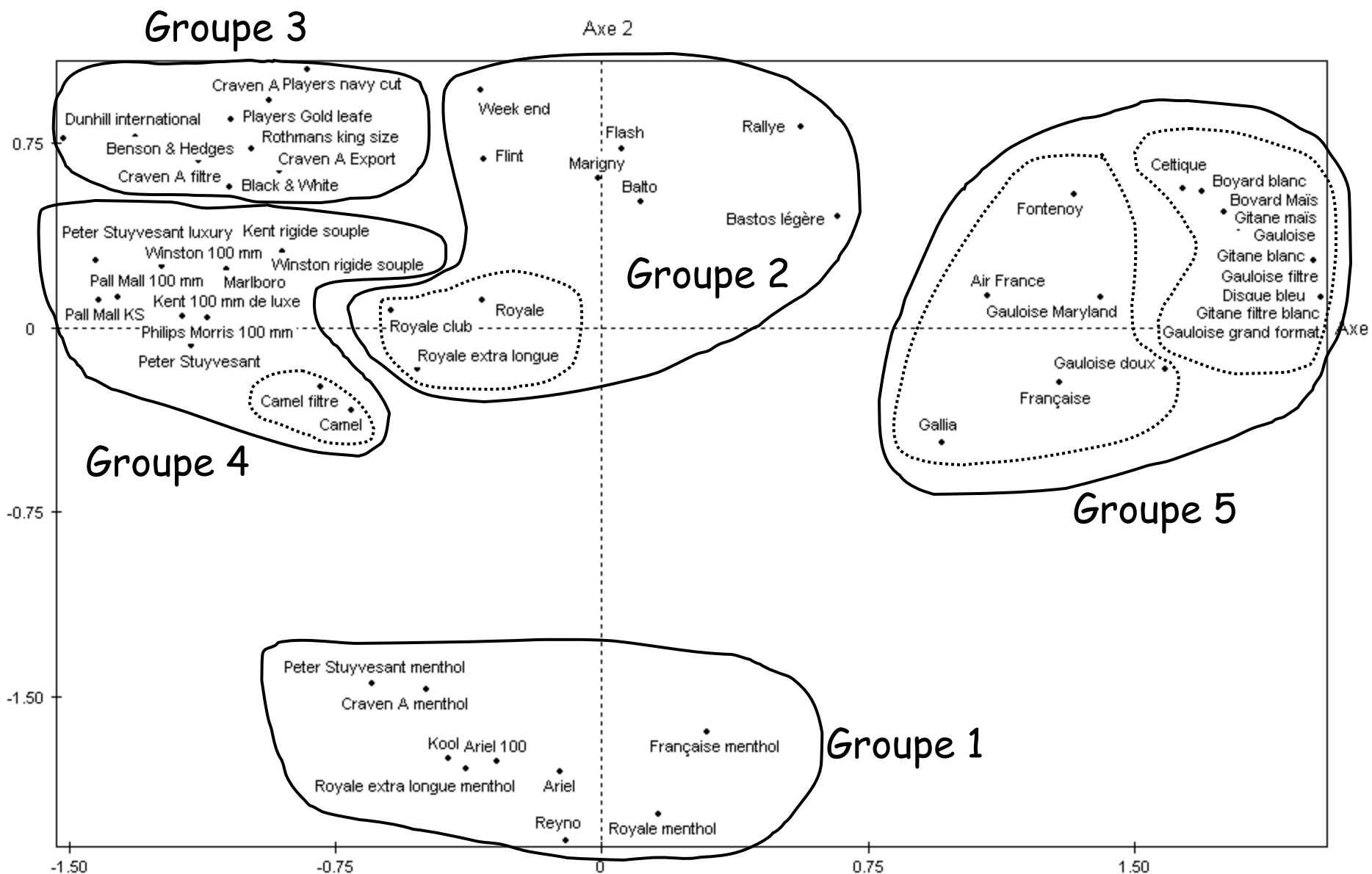
Visualisation de la classification des chiens en 4 groupes

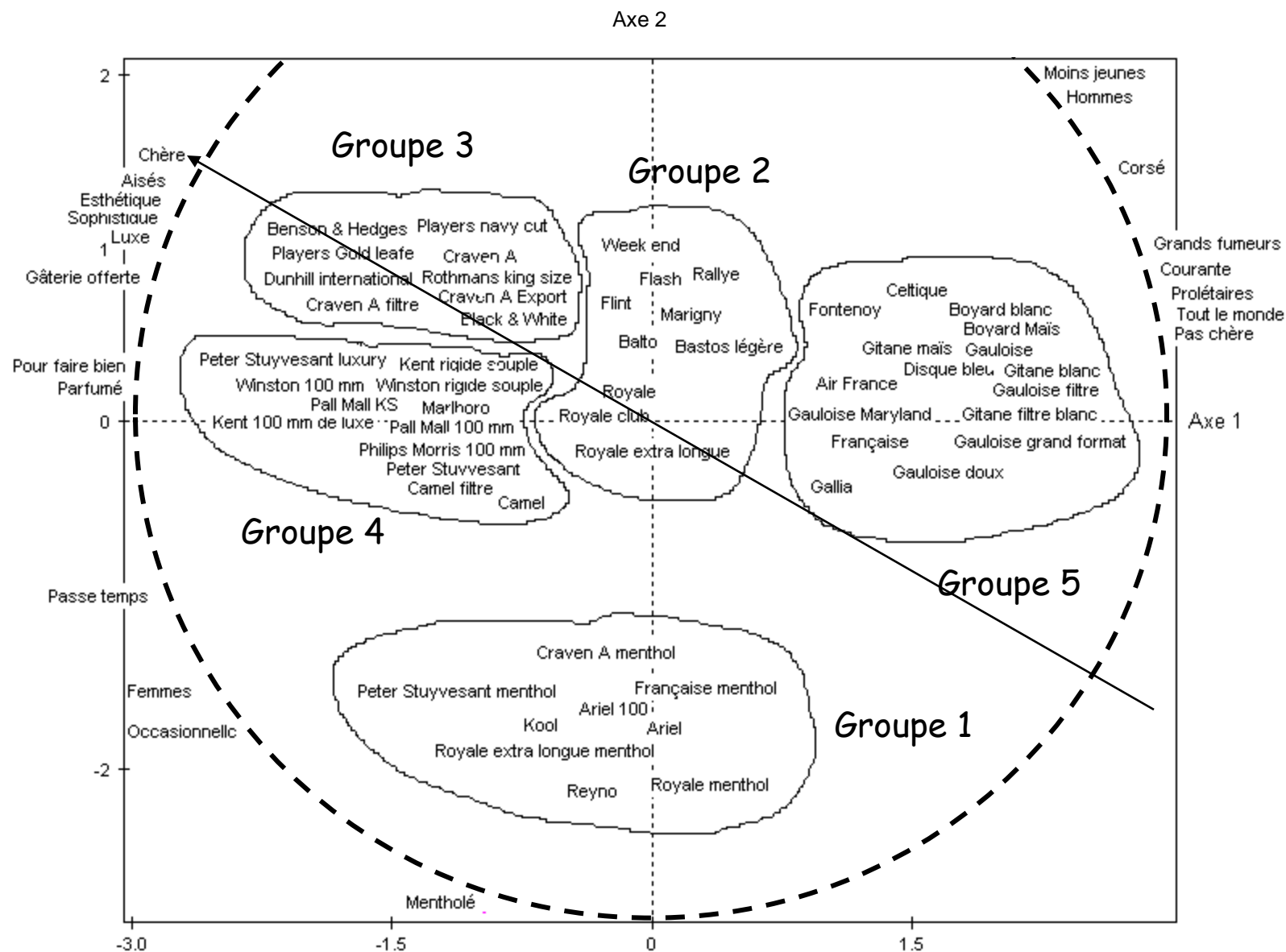


Exemple : Positionnement des cigarettes

Marque	1	2	3	4	5	6	7	8	9	10	11	12	13	56
1 Air France	100	12	5	33	29	5	7	25	6	6	37	4	8	25
2 Ariel	12	100	74	8	6	2	6	2	10	10	1	3	8	0
3 Ariel 100	5	74	100	6	10	8	9	2	7	7	3	3	11	2
4 Balto	33	8	6	100	53	9	19	10	41	31	12	19	17	11
5 Bastos légère	29	6	10	53	100	7	21	20	18	18	28	5	10	21
6 Benson & Hedges	5	2	8	9	7	100	67	2	12	19	3	40	37	2
7 Black & White	7	6	9	19	21	67	100	3	11	22	3	40	52	3
8 Boyard Maïs	25	2	2	10	20	2	3	100	6	4	58	3	4	81
9 Camel	6	10	7	41	18	12	11	6	100	82	10	24	15	6
10 Camel filtre	6	10	7	31	18	19	22	4	82	100	8	25	21	6
11 Celtique	37	1	3	12	28	3	3	58	10	8	100	4	6	71
12 Craven A	4	3	3	19	5	40	40	3	24	25	4	100	45	5
13 Craven A Export	8	8	11	17	10	37	52	4	15	21	6	45	100	5
⋮	⋮						⋮						⋮	⋮
56 Boyard blanc	25	0	2	11	21	2	3	81	6	6	71	5	5	100

- Chaque personne interrogée construit sa typologie des 56 marques.
- % de personnes mettant les marques i et j dans le même groupe.



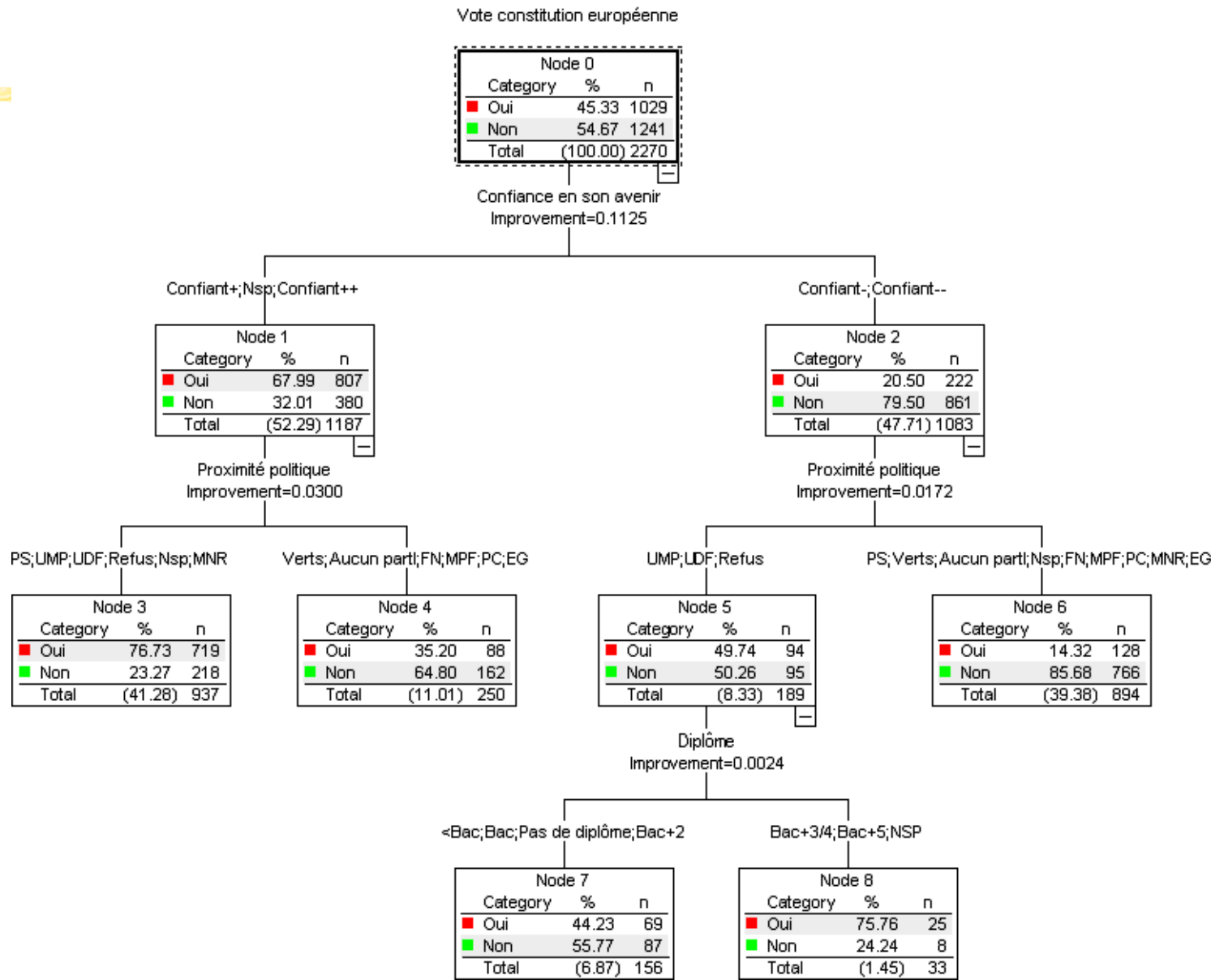


Expliquer ?

Référendum sur la constitution européenne

Vote constitution européenne	Sexe	Classe d'age	Proximité politique	Dernier diplôme	Confiance en son avenir
Oui	Femme	25-34	PS	Bac+3/4	Confiant+
Oui	Homme	60 et +	PS	< Bac	Confiant-
Oui	Femme	35 à 44 ans	UMP	Bac+3/4	Nsp
Oui	Homme	45-59	PS	Bac	Confiant++
Oui	Femme	35 à 44 ans	UMP	Bac+5/Grande école	Confiant++
Oui	Homme	25-34	UMP	Bac	Confiant+
Oui	Femme	25-34	UMP	Bac	Confiant+
Oui	Homme	35 à 44 ans	PS	Bac+5/Grande école	Confiant+
Oui	Femme	35 à 44 ans	UDF	Pas de diplôme	Confiant+
Oui	Homme	45-59	UDF	< Bac	Confiant--
Oui	Homme	25-34	UMP	Bac+5/Grande école	Confiant+
Oui	Homme	60 et +	UMP	< Bac	Confiant+
Oui	Femme	35 à 44 ans	PS	< Bac	Confiant+
Oui	Homme	18-24	UMP	Bac+3/4	Confiant-
Oui	Femme	35 à 44 ans	PS	Bac+2	Confiant-
Oui	Femme	18-24	Verts	Bac	Confiant++
Oui	Femme	60 et +	UMP	< Bac	Confiant+
Oui	Homme	35 à 44 ans	PS	Bac+2	Confiant+
Oui	Homme	60 et +	UMP	< Bac	Confiant+

Arbre de segmentation avec Answer Tree



Prévoir ?

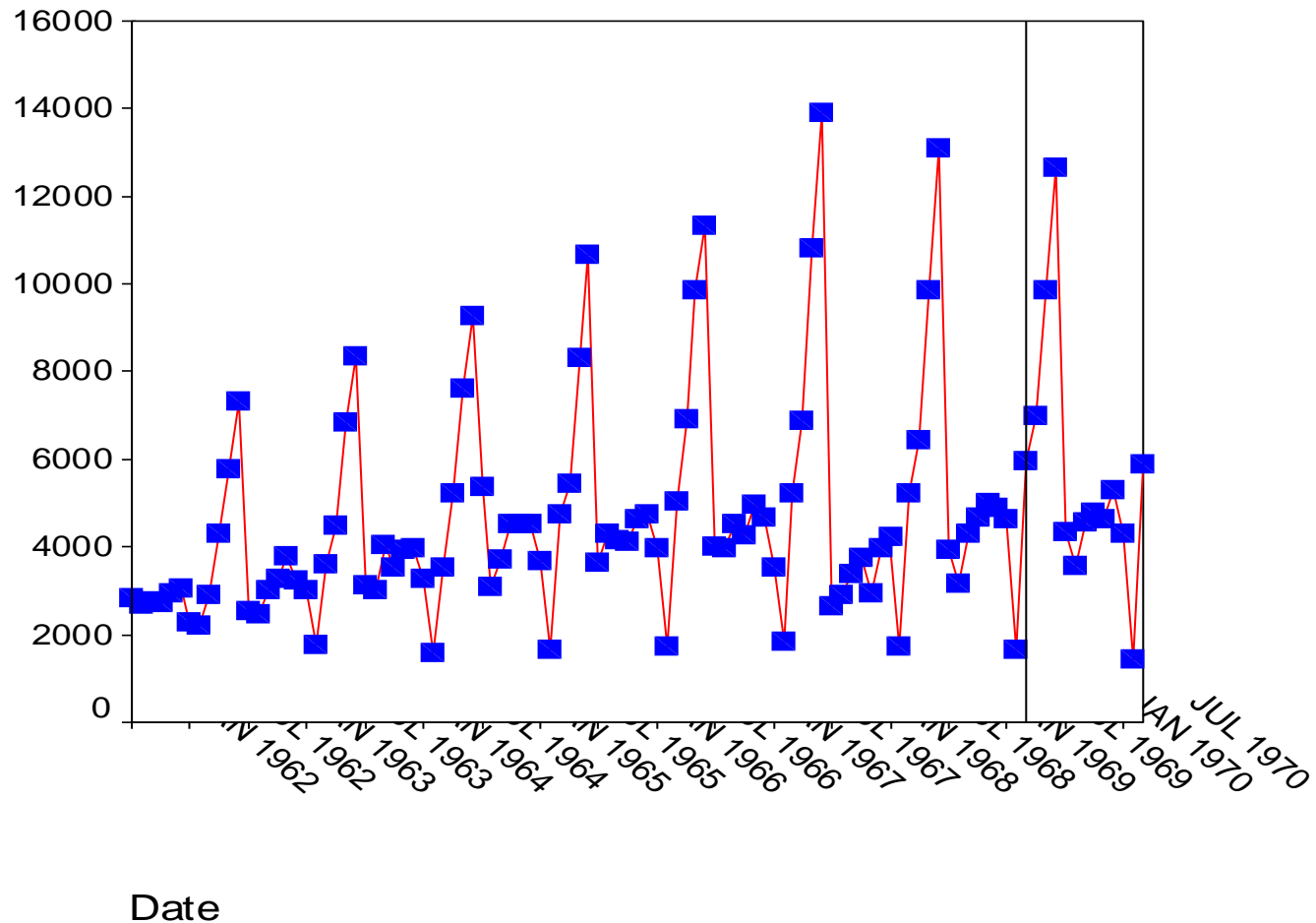
La méthode de Winters

Exemple : Ventes de Champagne

	1962	...	1968	1969	1970
Janvier	2815		2639	3934	4348
Février	2672		2899	3162	3564
Mars	2755		3370	4286	4577
Avril	2721		3740	4676	4788
Mai	2946		2927	5010	4618
Juin	3036		3986	4874	5312
Juillet	2282		4217	4633	4298
Août	2212		1738	1649	1431
Septembre	2922		5221	5951	5877
Octobre	4301		6424	6981	
Novembre	5764		9842	9851	
Décembre	7312		13076	12670	

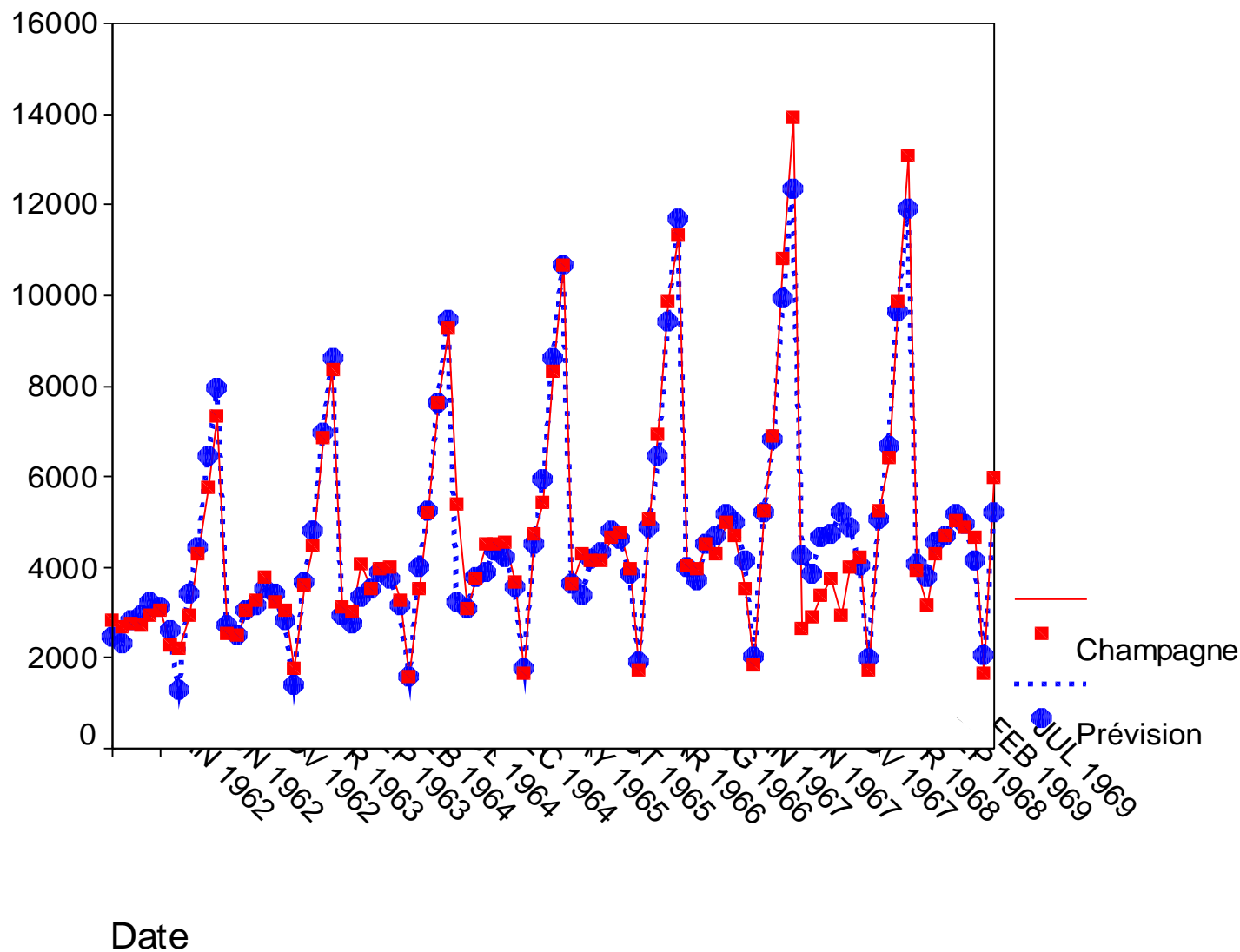
On exclut les douze derniers mois pour valider la méthode.

Ventes de Champagne



On exclut les douze derniers mois pour valider la méthode.

Résultats sur l'historique utilisé (prévision à l'horizon 1)



Résultats sur la période test (prévision sur l'horizon 1 à 12)

