

---

# Traitement de données de catégorisation

---

**Marine Cadoret, Sébastien Lê, Jérôme Pagès**  
**AGROCAMPUS OUEST, France**



**Sens&Co**  
2 avril 2009, Versailles

# Introduction

- ✿ La catégorisation consiste à grouper des objets en fonction de leur ressemblance.
- ✿ Après cette tâche, une tâche de verbalisation peut être demandée pour décrire les groupes (catégorisation “qualifiée”).

# Données

- 98 consommateurs ont effectué une catégorisation “qualifiée” sur 12 parfums :



Angel



Lolita  
Lempicka



L'instant



Cinéma



J'adore  
(ET)



J'adore  
(EP)



Shalimar



Aromatics  
Elixir



Coco  
Mademoiselle



Chanel  
n°5



Pure  
Poison



Pleasures

« gourmand,  
vanille, boisé »



« épice, aldéhydé »



« fleur blanche,  
vanille, orange »



« oriental,  
ostentatoire,  
boisé,  
Patchouli »



« fleur,  
floral,  
vert »

# Tableau de données (1)

	Shalimar	Aromatics Elixir	Chanel n°5	Angel	Lolita Lempicka	Cinéma	L'instant	Pure Poison	Coco Mademoiselle	Pleasures	J'adore (EP)	J'adore (ET)
Shalimar	98	42	30	21	9	10	13	11	9	6	6	7
Aromatics Elixir	42	98	51	27	6	8	13	12	12	11	12	7
Chanel n°5	30	51	98	15	8	9	10	21	11	14	12	14
Angel	21	27	15	98	36	18	14	10	10	11	11	12
Lolita Lempicka	9	6	8	36	98	42	22	18	21	18	18	18
Cinéma	10	8	9	18	42	98	26	28	30	22	23	24
L'instant	13	13	10	14	22	26	98	25	20	23	28	22
Pure Poison	11	12	21	10	18	28	25	98	33	30	29	28
Coco Mademoiselle	9	12	11	10	21	30	20	33	98	28	28	38
Pleasures	6	11	14	11	18	22	23	30	28	98	38	48
J'adore (EP)	6	12	12	11	18	23	28	29	28	38	98	56
J'adore (ET)	7	7	14	12	18	24	22	28	38	48	56	98

Les données sont généralement regroupées dans un tableau de cooccurrences

(ou de dissimilarités) et traitées par MDS non-metrique

# Tableau de données (2)

produit	juge 12	juge 13	juge 14	juge 15	juge 16
Angel	1	4	1	5	2
Aromatic Elixir	3	3	5	2	1
Chanel n°5	4	3	4	1	3
Cinéma	2	5	6	4	2
Coco Mademoiselle	1	5	2	4	3
J'adore (EP)	1	6	2	3	3
J'adore (ET)	2	6	2	3	3
L'instant	1	4	6	2	4
Lolita Lempicka	1	5	1	5	2
Pleasures	3	4	6	3	4
Pure Poison	1	1	2	4	4
Shalimar	2	2	3	2	1

**Chaque consommateur peut aussi être considéré  
comme une variable qualitative**

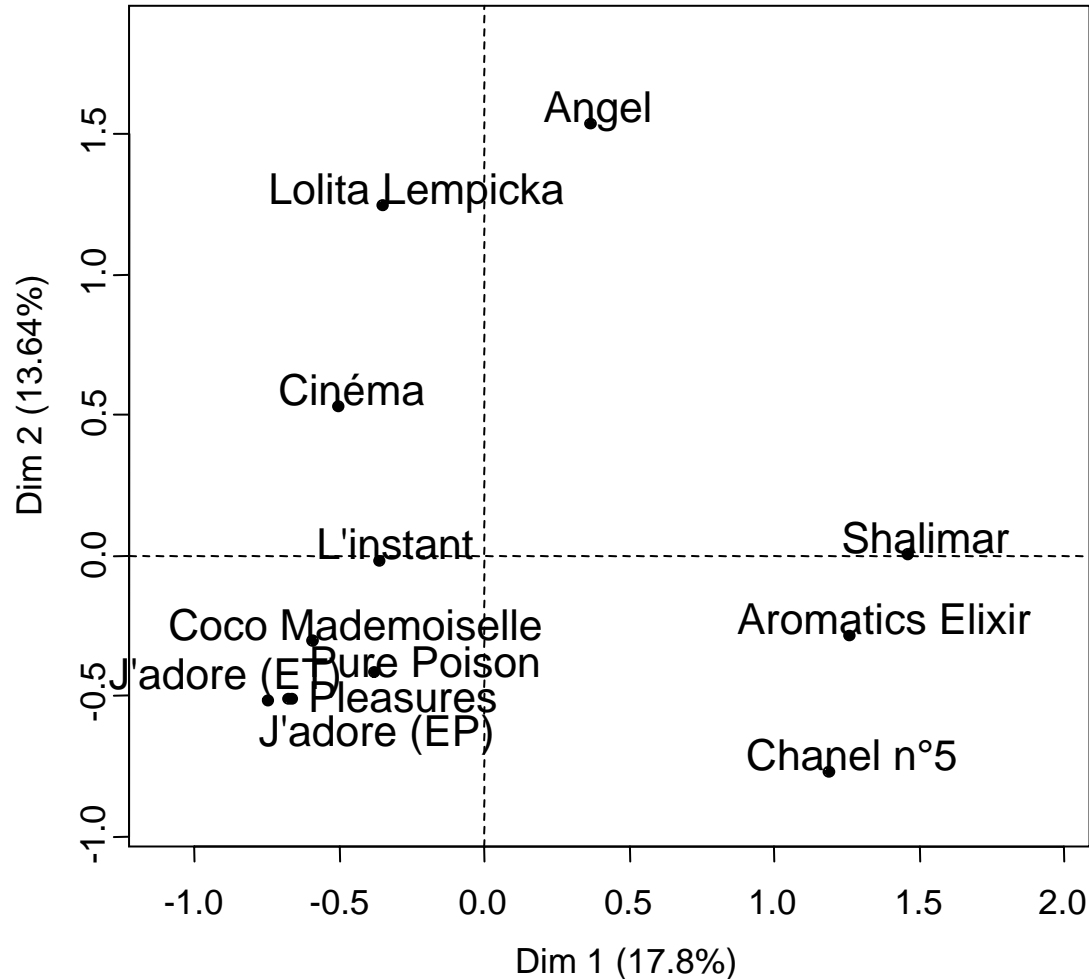
# Tableau de données (2)

produit	juge 12	juge 13	juge 14	juge 15	juge 16
Angel	fleuri doux	fruité fort	vanillé épicé esprit des îles	à manger sucré	nourriture épice
Aromatic Elixir	fort homme	capiteux grand-mère	rude fort	le vieux	ménager cire
Chanel n°5	Gr 4	capiteux grand-mère	toilettes	savon	connu classique
Cinéma	fleuri artificiel herbe	fruité moyen	sucré	doux	nourriture épice
Coco Mademoiselle	fleuri doux	fruité moyen	douceur fleuri	doux	connu classique
J'adore (EP)	fleuri doux	sucré faible	douceur fleuri	fleuri	connu classique
J'adore (ET)	fleuri artificiel herbe	sucré faible	douceur fleuri	fleuri	connu classique
L'instant	fleuri doux	fruité fort	sucré	le vieux	fleuri
Lolita Lempicka	fleuri doux	fruité moyen	vanillé épicé esprit des îles	à manger sucré	nourriture épice
Pleasures	fort homme	fruité fort	sucré	fleuri	fleuri
Pure Poison	fleuri doux	acidulé désodorisant	douceur fleuri	doux	fleuri
Shalimar	fleuri artificiel herbe	fort lavande eau de cologne	renfermé agressif	le vieux	ménager cire

**Ce tableau peut être traité par ACM**

# Représentation des parfums

MCA factor map



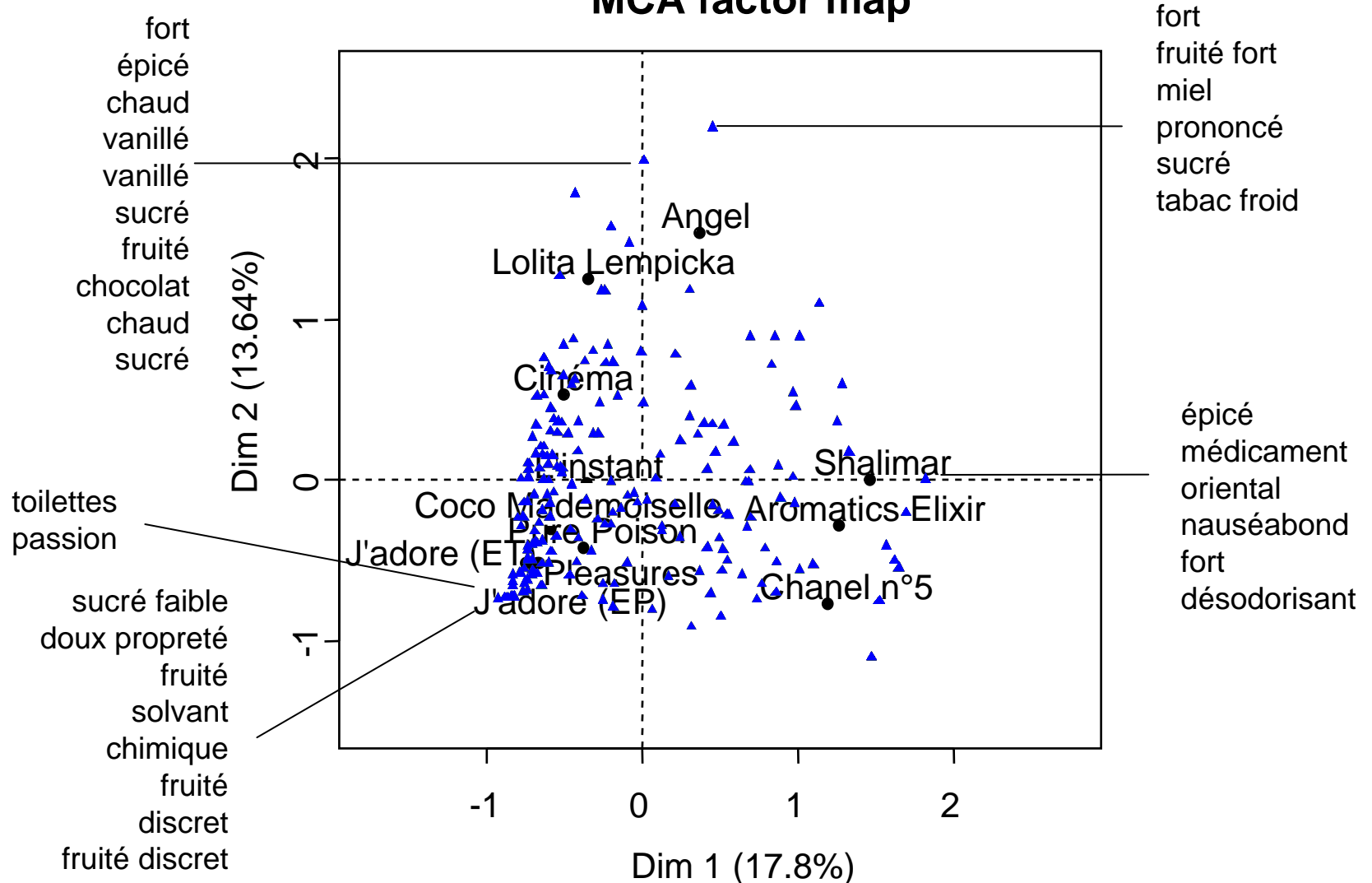


# Tableau de co-occurrences

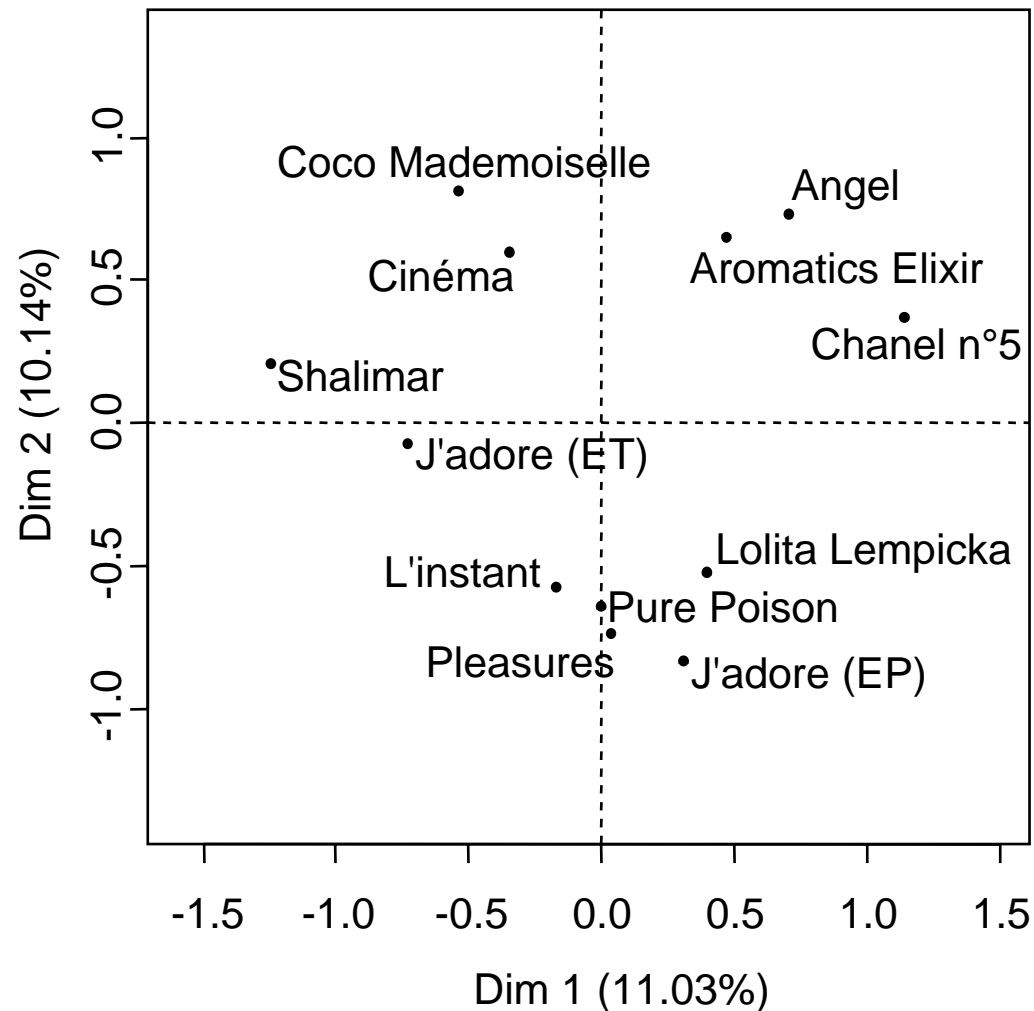
	Shalimar	Aromatics Elixir	Chanel n°5	Angel	Lolita Lempicka	Cinéma	L'instant	Pure Poison	Coco Mademoiselle	Pleasures	J'adore (EP)	J'adore (ET)
<b>Shalimar</b>	98	42	30	21	9	10	13	11	9	6	6	7
<b>Aromatics Elixir</b>	42	98	51	27	6	8	13	12	12	11	12	7
<b>Chanel n°5</b>	30	51	98	15	8	9	10	21	11	14	12	14
<b>Angel</b>	21	27	15	98	36	18	14	10	10	11	11	12
<b>Lolita Lempicka</b>	9	6	8	36	98	42	22	18	21	18	18	18
<b>Cinéma</b>	10	8	9	18	42	98	26	28	30	22	23	24
<b>L'instant</b>	13	13	10	14	22	26	98	25	20	23	28	22
<b>Pure Poison</b>	11	12	21	10	18	28	25	98	33	30	29	28
<b>Coco Mademoiselle</b>	9	12	11	10	21	30	20	33	98	28	28	38
<b>Pleasures</b>	6	11	14	11	18	22	23	30	28	98	38	48
<b>J'adore (EP)</b>	6	12	12	11	18	23	28	29	28	38	98	56
<b>J'adore (ET)</b>	7	7	14	12	18	24	22	28	38	48	56	98

# Représentation des mots

MCA factor map



# Représentation de données aléatoires



# Significativité des résultats

- $H_0$  : Absence de consensus
- Indicateur : première valeur propre
- Evolution de l'indicateur sous  $H_0$

		Nombre de sujets			
		10	20	50	100
Nombre de produits	10	0,69566	0,624637	0,557816	0,524184
	20	0,471051	0,389234	0,319656	0,286282
	50	0,306195	0,228844	0,167827	0,14006
	100	0,236413	0,16401	0,109948	0,086396

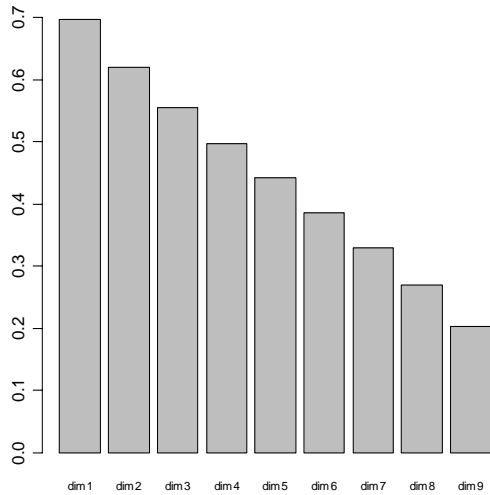
# Explication

- ✱ Nombre de colonnes >> nombre de lignes
- ✱ Production automatique de dimensions communes

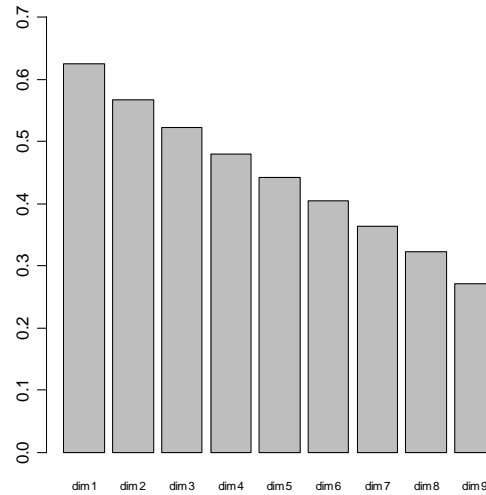
		Nombre de sujets			
		10	20	50	100
Nombre de produits	10	0,69566	0,624637	0,557816	0,524184
	20	0,471051	0,389234	0,319656	0,286282
	50	0,306195	0,228844	0,167827	0,14006
	100	0,236413	0,16401	0,109948	0,086396

# Diagrammes des valeurs propres

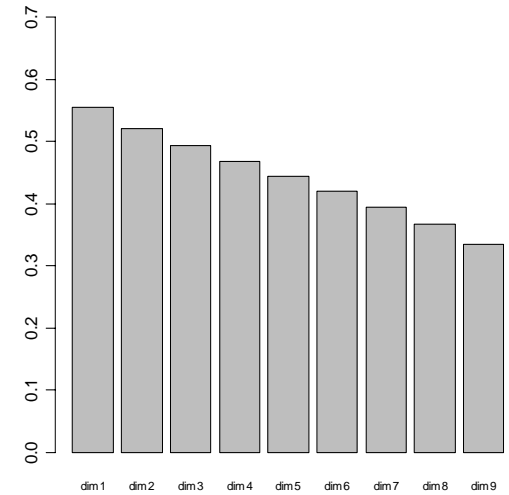
10 subjects



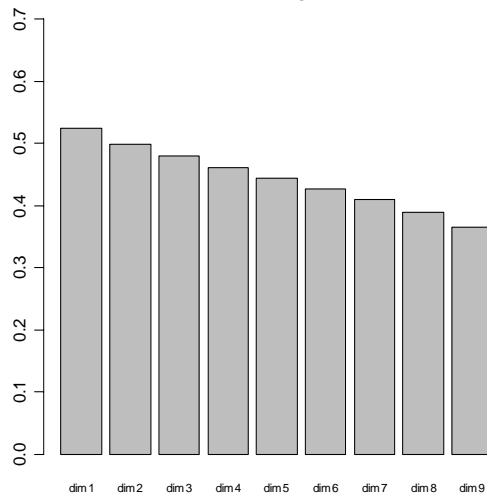
20 subjects



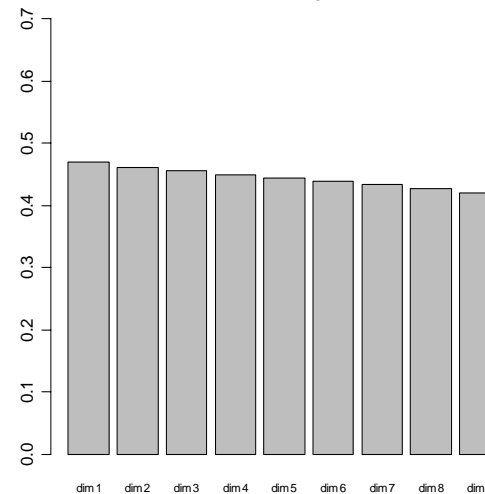
50 subjects



100 subjects



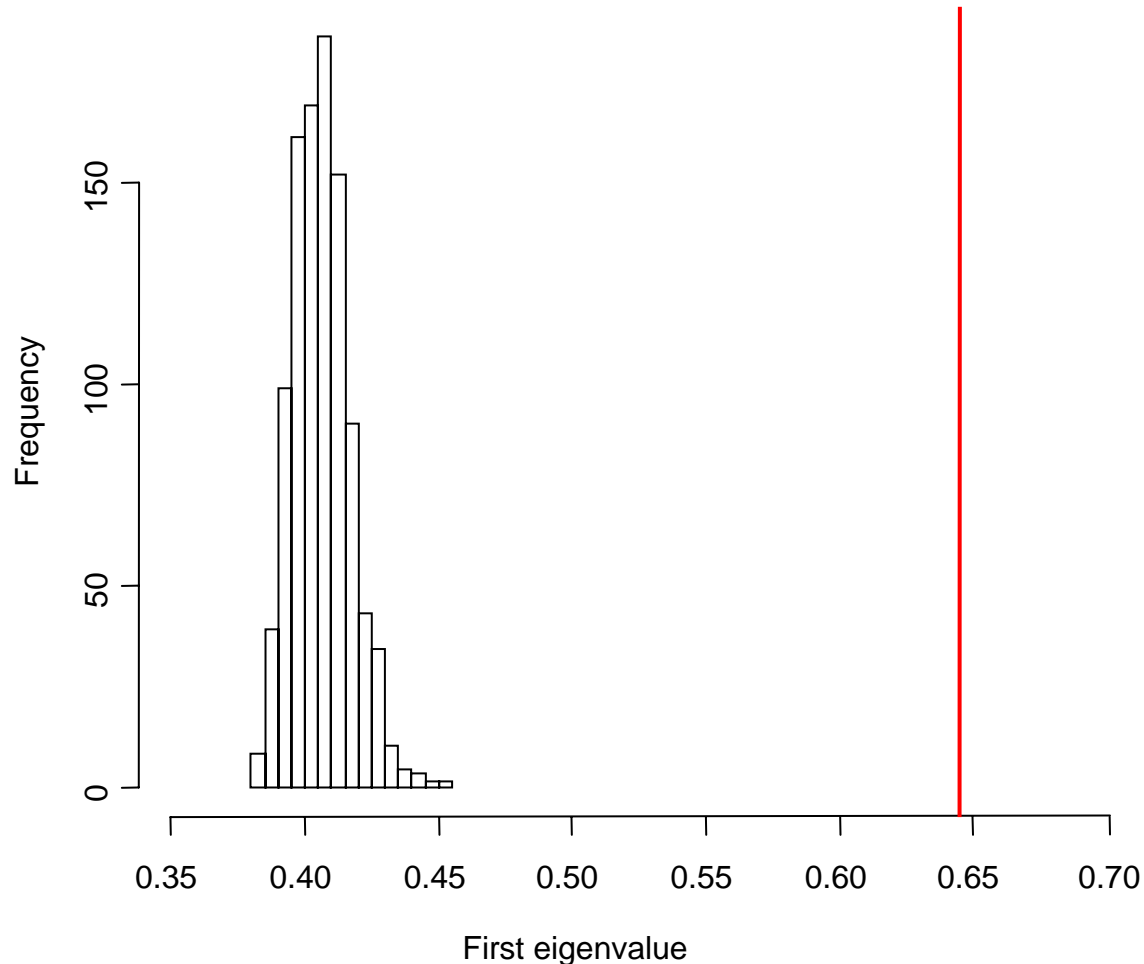
1000 subjects



# Significativité de l'indicateur pour un jeu de données

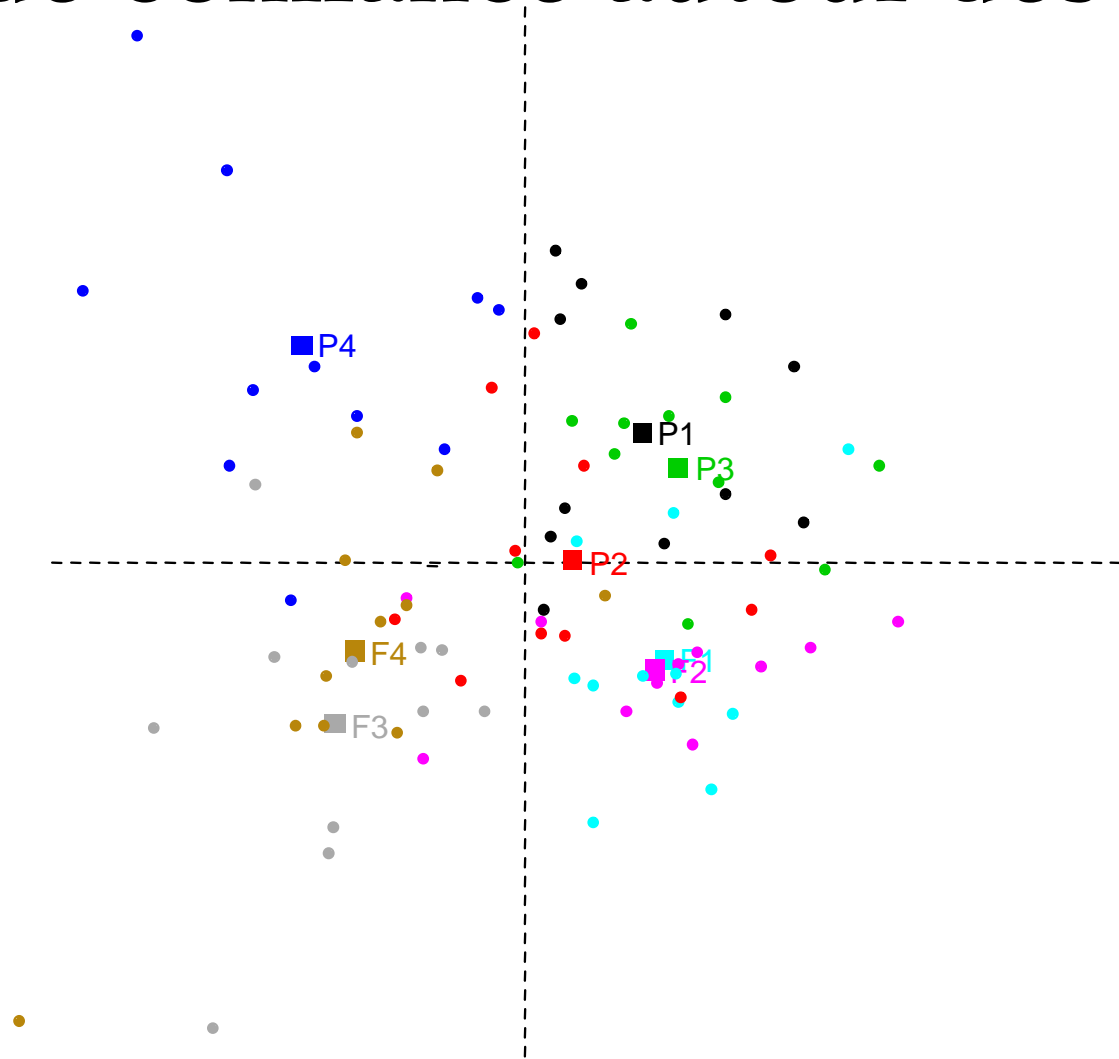
- ★ Calcul de la probabilité critique associée à la première valeur propre de l'ACM :
  1. Répéter un grand nombre de fois :
    1. Permutations indépendantes des lignes pour chaque colonne
    2. Calcul de la première valeur propre associée au jeu de données permuté
  2. Distribution des valeurs propres sous  $H_0$
  3. Localiser la valeur propre observée dans la distribution pour obtenir la probabilité critique

# Significativité de la première dimension pour les données parfums



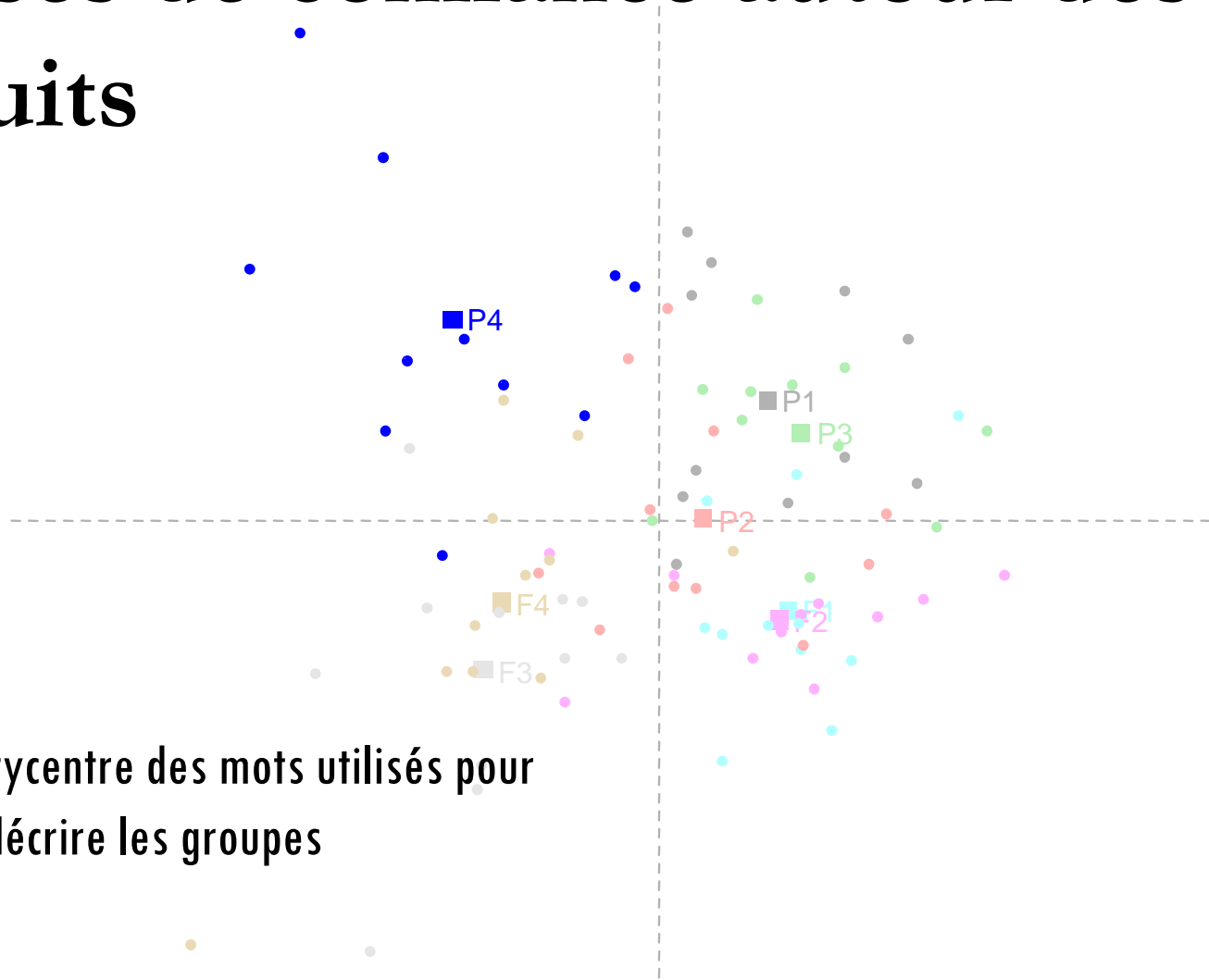


# Ellipses de confiance autour des produits



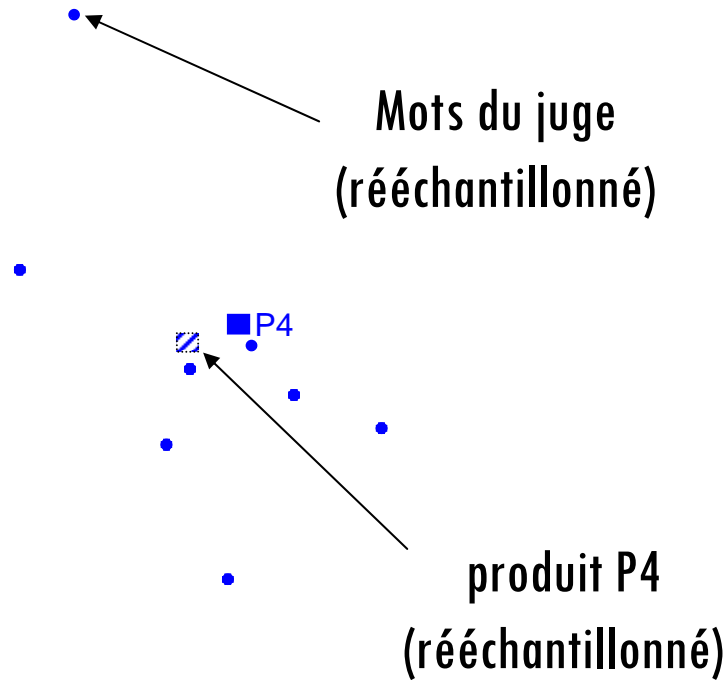
Représentation superposée des produits et de leurs descriptions

# Ellipses de confiance autour des produits

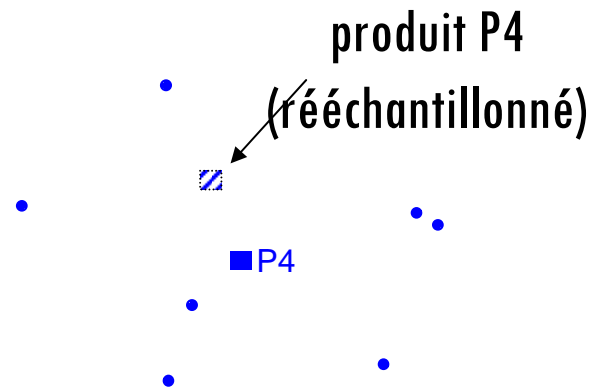


P4 est au barycentre des mots utilisés pour décrire les groupes

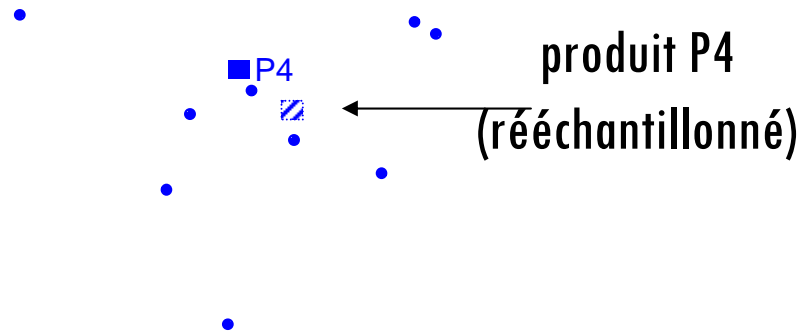
# Ellipses de confiance autour des produits



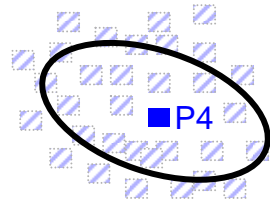
# Ellipses de confiance autour des produits



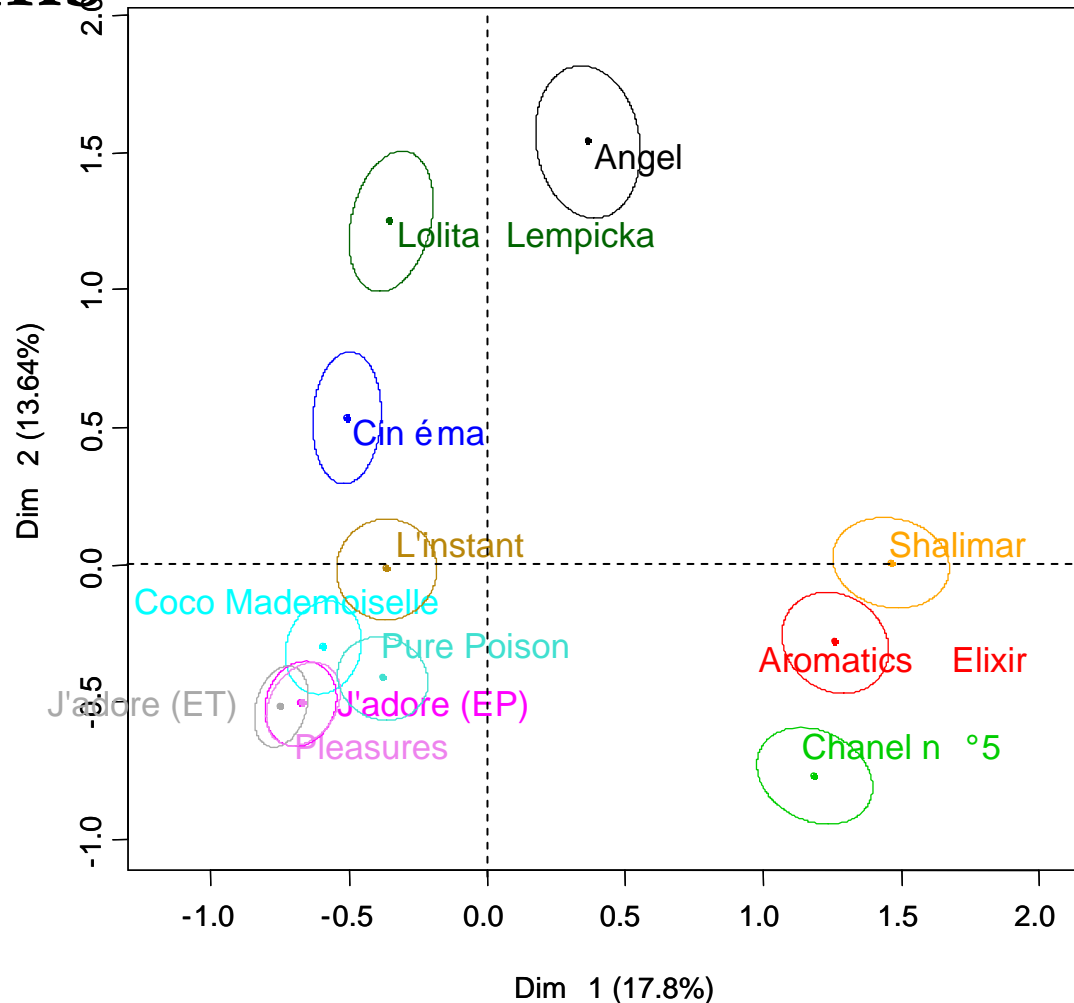
# Ellipses de confiance autour des produits



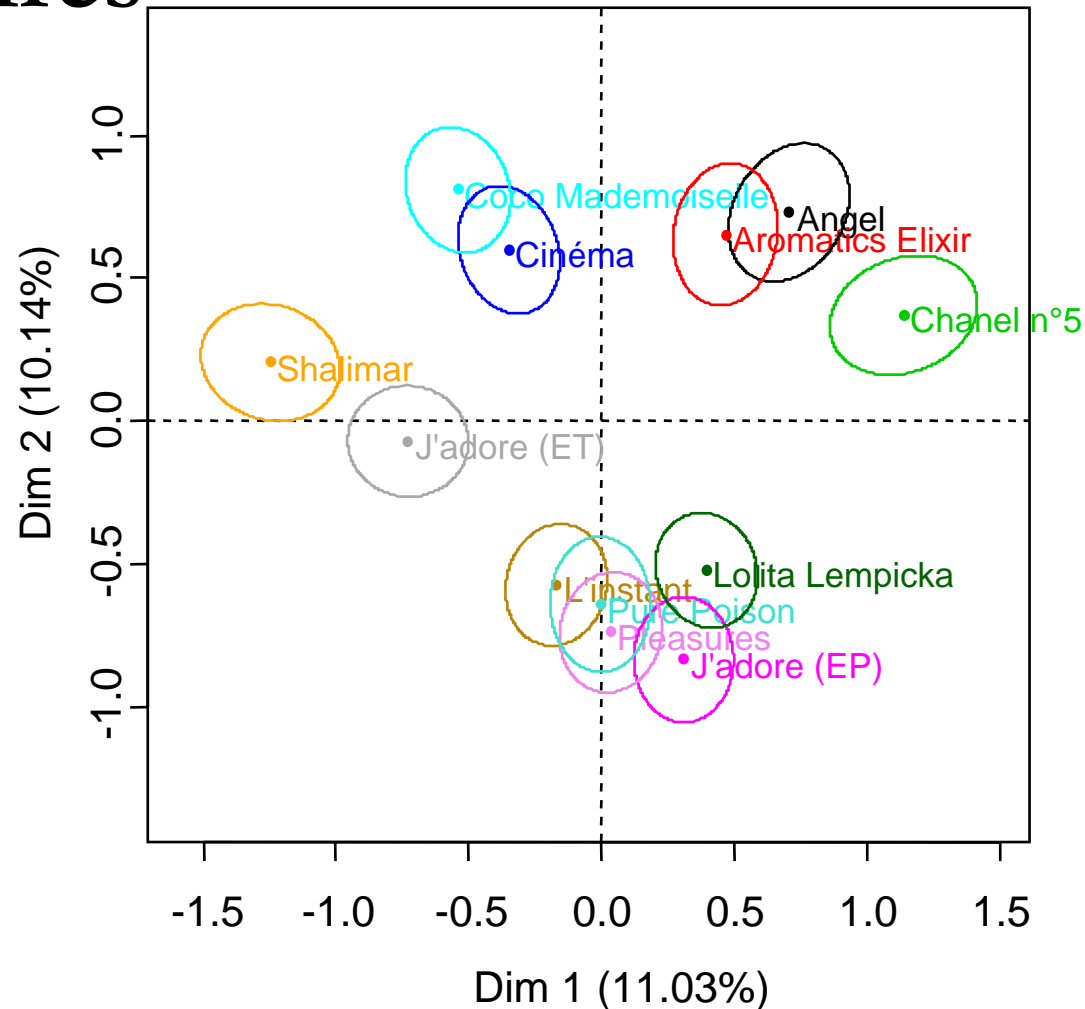
# Ellipses de confiance autour des produits



# Ellipses de confiance pour les données parfums



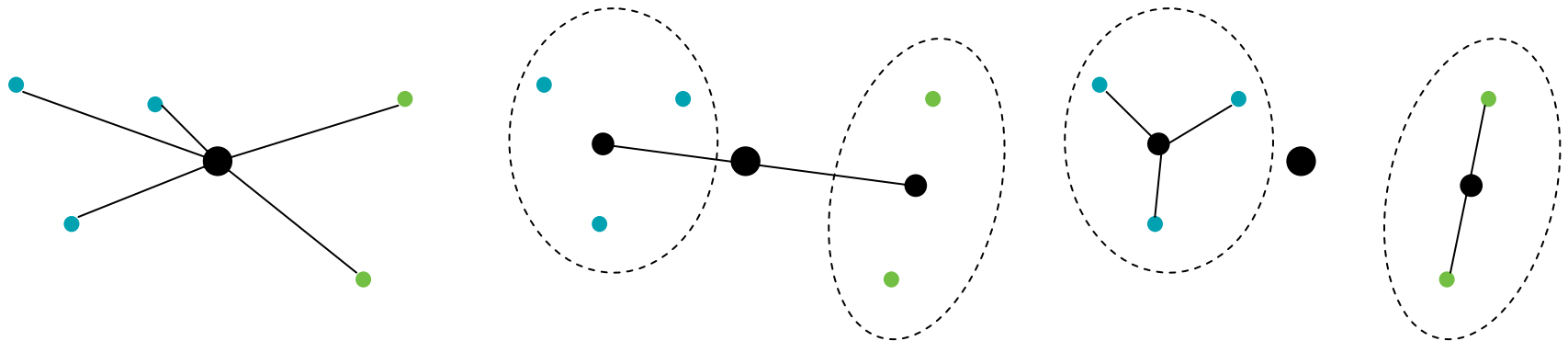
# Ellipses de confiance pour des données aléatoires





# Second indicateur empirique

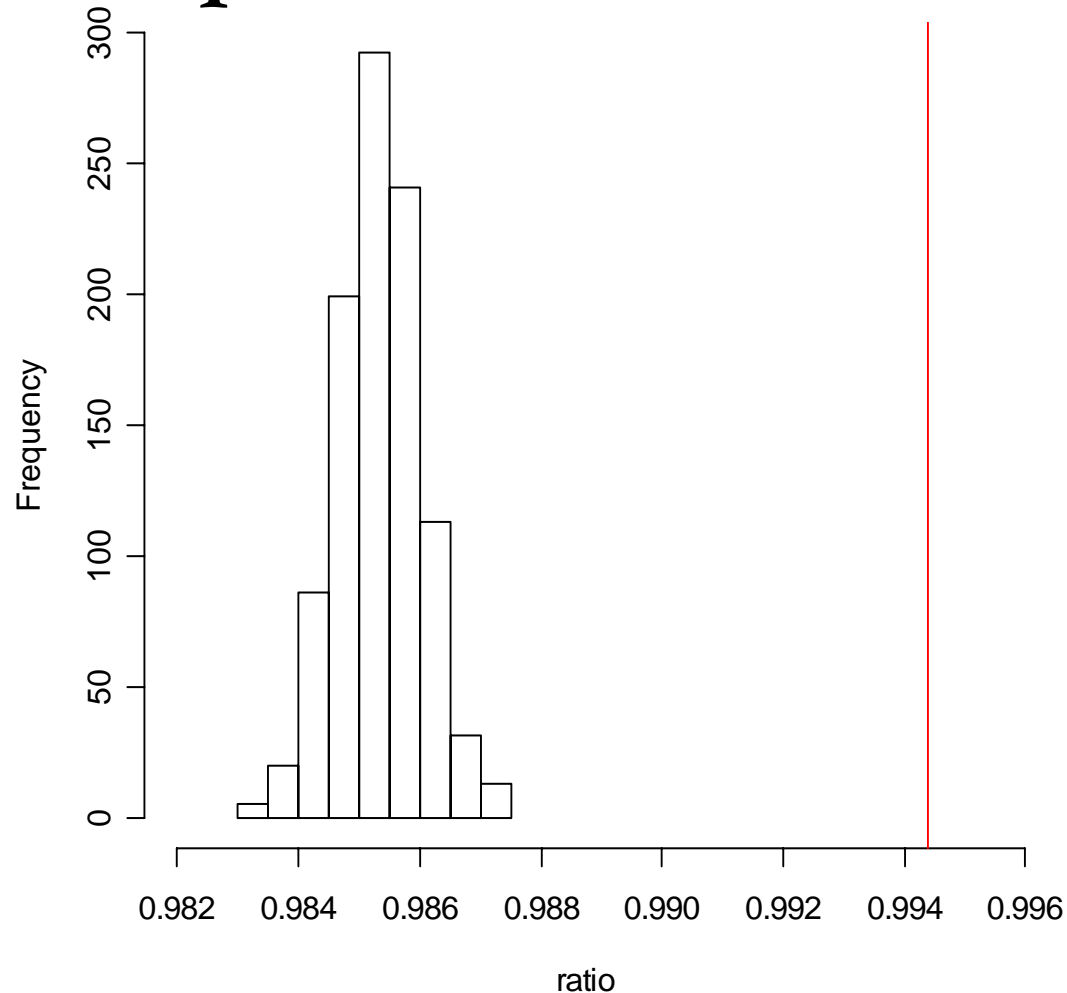
## ★ Recouvrement des ellipses



**Inertie totale = Inertie inter-groupe + Inertie intra-groupe**

★ Calcul du rapport d'inertie :  $0 \leq \frac{\text{inertie inter}}{\text{inertie totale}} \leq 1$

# Significativité du rapport d'inertie pour les données parfums



# Conclusion

- ✿ ACM : méthode appropriée pour traiter des données de catégorisation
- ✿ Données sensorielles : peu de lignes et beaucoup de colonnes
- ✿ Par construction, beaucoup de liaisons
- ✿ Validation externe nécessaire

# SensoMineR

**SensoMineR a package for sensory data analysis**  
*Journal of sensory studies (2008)*

# FACTOMINER

**FactoMineR: an R package for multivariate analysis**  
*Journal of statistical software (2008)*

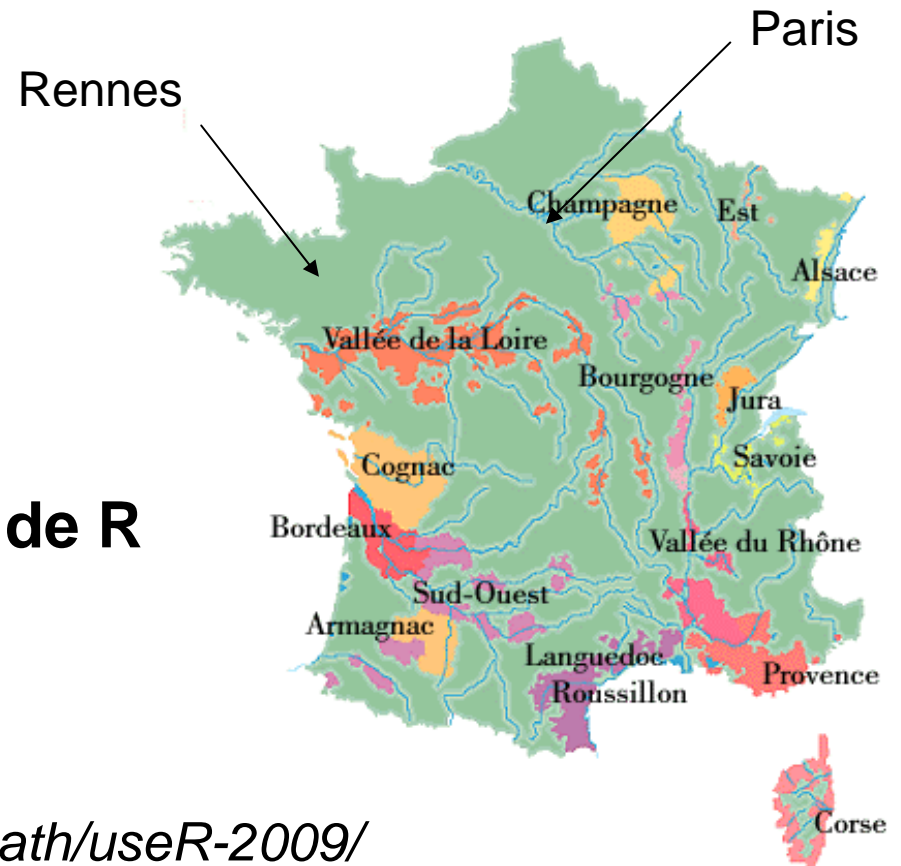
<http://www.agrocampus-rennes.fr/math/>

# Le laboratoire de mathématiques appliquées d'Agrocampus organise

*use* **R!**

**Le congrès des utilisateurs de R**

**8-10 juillet à Rennes**



<http://www.agrocampus-rennes.fr/math/useR-2009/>