

2. AFC – Analyse Factorielle des Correspondances

Plan

1. Exemples, objectifs
2. L'ACP du tableau de contingence ?
3. Les deux tableaux analysés par l'AFC
4. Une analyse graphique d'un lien de dépendance
5. L'ajustement des deux nuages de profils
6. Interprétation des résultats d'une AFC

2.1 – Exemples, objectifs

Exemple 1. Présidentielles 2012

Num	Département / Candidat										
		Eva Joly	Marine Le Pen	Nicolas Sarkozy	Jean-Luc Mélenchon	Philippe Poutou	Nathalie Arthaud	Jacques Cheminade	François Bayrou	Nicolas Dupont-Aignan	François Hollande
1	Ain	7268	66540	97722	30898	3323	1794	860	32650	7208	73096
2	Aisne	3455	78452	72090	30360	3860	2490	738	19895	5853	80751
3	Allier	3232	37736	49477	27969	2584	1482	457	17814	4068	61130
4	Alpes de Haute-Provence	2933	20875	25668	15269	1394	487	283	7483	1845	24551
5	Hautes-Alpes	3147	15359	22655	12175	1152	488	212	8559	1782	21248
6	Alpes-Maritimes	12556	136982	216738	49493	4048	1576	1238	38980	9241	111990
7	Ardèche	5621	40216	47687	28247	2750	1257	521	18373	3890	52156
8	Ardennes	1868	37628	37524	14260	1981	1185	366	11551	2787	44441
9	Ariège	2742	16125	17979	16197	1396	528	221	6411	1446	33003
10	Aube	2294	40740	49196	12860	1632	990	407	13575	3524	36967
11	Aude	4516	50234	46801	28456	2724	1116	458	13076	3184	65783
12	Aveyron	4243	25619	46351	22282	2417	969	435	22809	3095	53493
13	Bouches-du-Rhône	21977	243348	286175	139719	8293	3641	2249	66082	14087	255052
14	Calvados	8569	65126	109745	42396	5385	2641	1008	40562	8227	117773
15	Cantal	1534	14877	28151	8836	1218	587	240	10980	1632	30353
	:										
95	Val-d'Oise	10907	83102	139863	63679	5109	2340	1325	44683	9049	172658

Tableau de contingence : tableau croisé entre deux variables qualitatives

Variable 1

The diagram illustrates the relationship between a matrix A and its row and column sums. Matrix A is a 6x6 grid with columns labeled $1, \dots, j, \dots, J$ and rows labeled $1, \dots, i, \dots, I$. The element a_{ij} is highlighted in blue. To the right, a column vector Σ is shown with elements $\Sigma_1, \dots, \Sigma_i, \dots, \Sigma_I$, where Σ_i is highlighted in blue. Below matrix A , a row vector Σ is shown with elements $\Sigma_1, \dots, \Sigma_{\bullet}, \dots, \Sigma_I$, where Σ_{\bullet} is highlighted in blue.

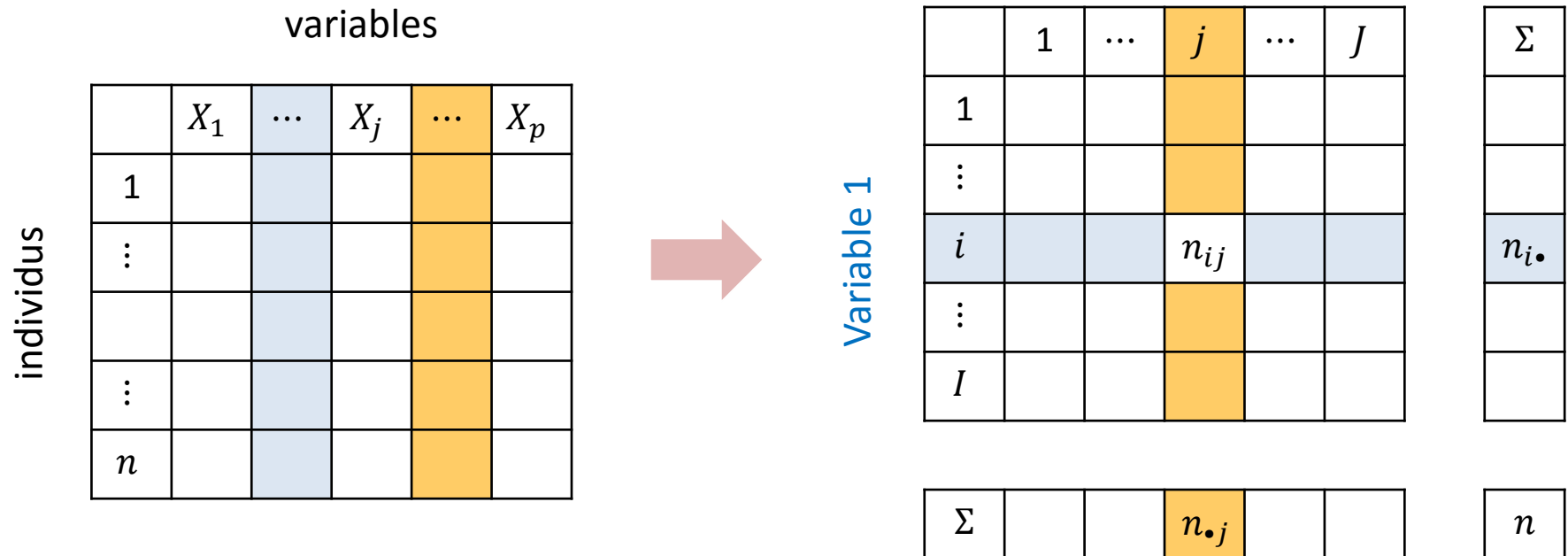
$$n_{ij} =$$

$$n_{i\bullet} =$$

$$n_{\bullet j} =$$

$$n =$$

Un tableau de contingence est souvent établi à partir d'un tableau *individus x variables*



Les objectifs d'une AFC

Comme dans toute analyse factorielle : les objectifs peuvent être explicités au niveau des lignes et des colonnes du tableau analysé

- Peut-on établir des « **ressemblances** » entre **lignes** ?

Quels sont les départements présentant des résultats similaires ?

- Peut-on établir des « **ressemblances** » entre **colonnes** ?

Certains candidats présentent-ils des résultats similaires ?

- **Liaison** entre les deux études

Si certains départements se ressemblent, peut-on expliquer leur ressemblance en termes de quelques candidats ?

- Quels sont les **principaux facteurs de variabilité** du tableau de données ?

Qu'entend-on par « ressemblance » ?

Exemple : quels sont les départements qui se ressemblent le plus ?

Num	Département	EvaJoly	Marine Le Pen	Nicolas Sarkozy	Jean-Luc Mélenchon	Philippe Poutou	Nathalie Arthaud	Jacques Cheminade	François Bayrou	Nicolas Dupont-Aignan	François Hollande
2A	Corse.du.Sud	1658	19081	23623	7191	873	220	163	4059	797	16540
67	Bas.Rhin	16188	124264	196968	42302	5993	3779	1655	69940	10141	114702
93	Seine.Saint.Denis	11781	72335	104010	90710	4936	2708	1259	32661	6978	206537

- Ce qui nous intéresse : **importance relative** des candidats dans les différents départements
- On doit donc considérer les **profils ou %** en lignes et non les effectifs bruts
- Si on s'intéresse aux valeurs brutes : faire une ACP !

Exemples de tableaux de contingence

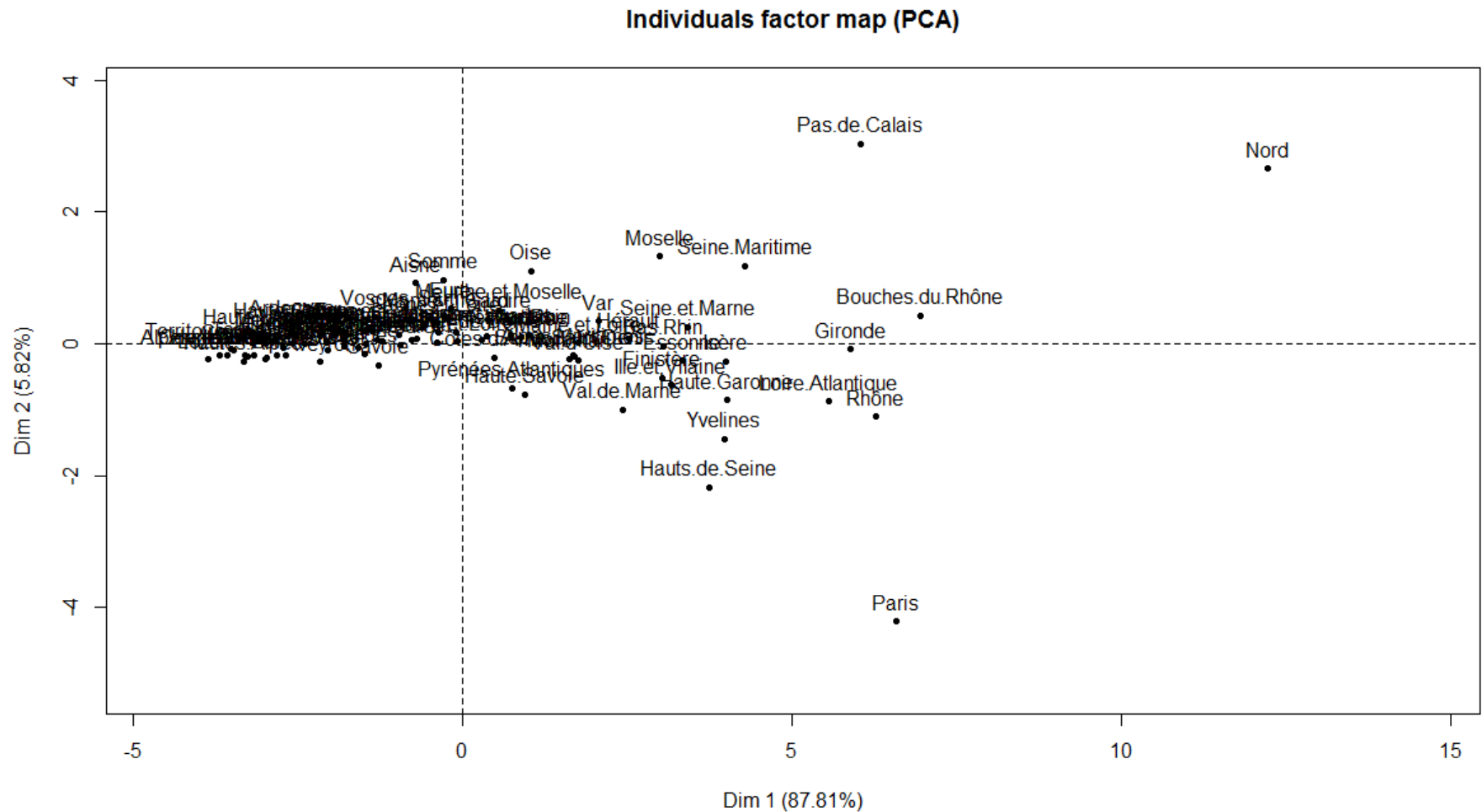
Variable 1	Variable 2	Effectif n_{ij}
PCS	Arrondissement	Nombre d'habitants de profession (i) dans l'arrondissement (j)
Couleur des yeux	Couleur des cheveux	Nombre d'individus avec cheveux (i) et yeux (j)
Candidat	Mot	Nombre de fois où le mot (i) est cité dans le discours (j)
Vin	Mot	Nombre de fois où le mot (i) est cité pour caractériser le vin (j)
Cause de décès	Classe d'âge	Nombre d'individus morts de la cause (i) dans la classe d'âge (j)

Exemple 2.

Décès / Âge

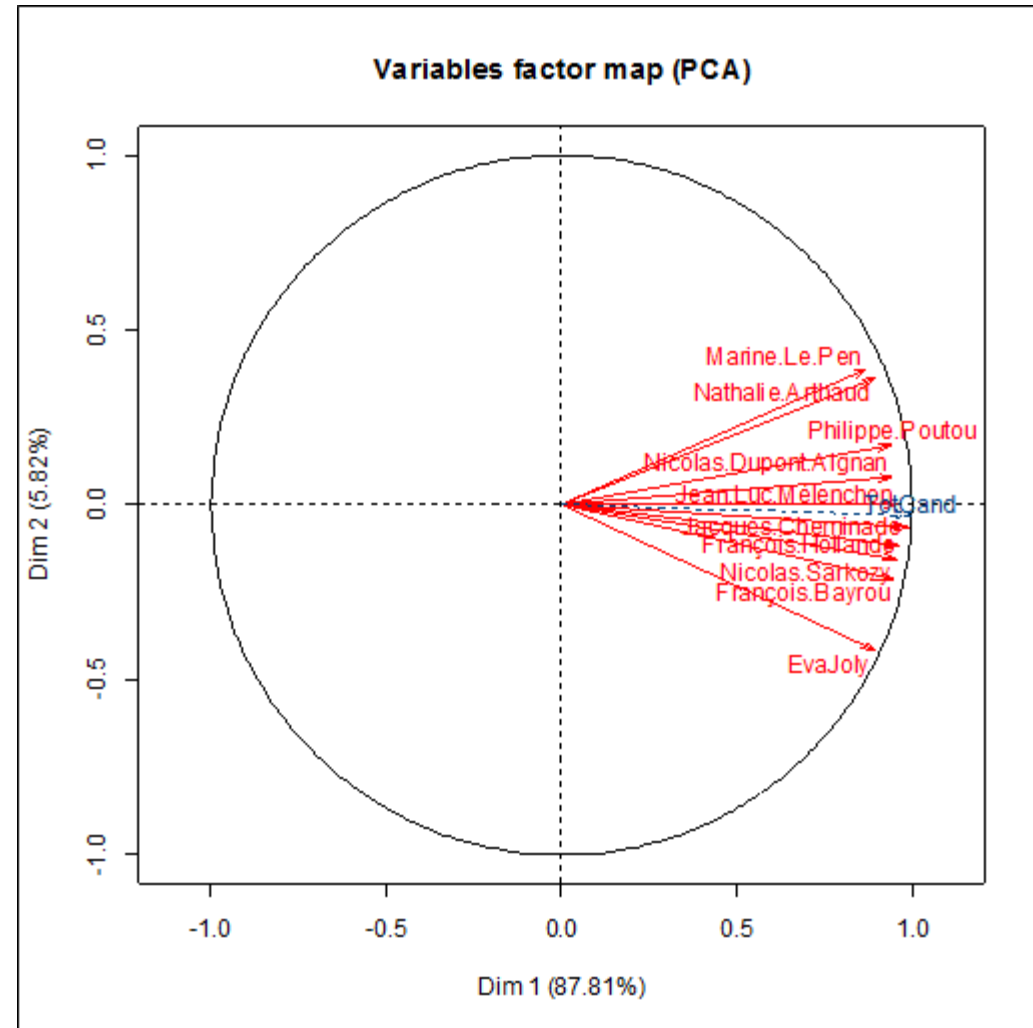
	a0-1	a1-4	a5-14	a15-24	a25-34	a35-44	a45-54	a55-64	a65-74	a75-84	a85-94	a95+
Abus alcool et psychose alcoolique	0	0	0	29	214	763	1839	1543	1203	636	95	5
Accidents de transport	52	194	655	4653	2451	1841	1927	1409	1479	1211	272	12
Arthrite rhumatoïde et ostéoarthrite	0	0	3	2	6	6	26	59	301	526	301	57
Asthme	9	6	9	29	39	59	183	266	535	846	464	67
Cardiopathies ischémiques	0	5	2	27	203	1212	4658	8436	20368	32305	18316	2813
Infections en période périnatale	3519	18	10	5	0	0	0	0	0	0	0	0
Chutes accidentelles	14	47	60	194	221	295	687	838	1933	6003	6511	1029
Complic. de grossesse accouch. et puerpéralité	0	0	0	30	80	39	1	0	0	0	0	0
Diabète sucré	1	1	5	25	73	162	570	1408	3945	7266	4060	574
Événements dont intention non déterminée	10	28	73	336	429	372	420	321	326	262	87	5
Grippe	5	3	5	11	7	13	19	35	143	409	453	76
Hépatites virales	4	2	3	19	22	91	154	140	227	311	89	6
Homicides	31	36	47	144	199	180	128	92	60	56	16	0
Infection à méningocoques	33	41	13	21	3	4	6	6	12	8	0	0
Infections peau et tissu cellulaire sous-cutané	0	3	2	7	8	15	49	78	345	1190	1369	378
Intoxications accidentelles	3	28	23	136	181	195	199	132	171	269	199	28
Maladies cérébrovasculaires	29	15	41	127	344	862	2826	5172	17402	40891	27554	3617
Maladies rein et uretère	11	8	15	24	42	104	288	650	1995	5433	4360	717
Maladies sang et organes hématopoïétiques	58	74	61	50	60	90	204	351	741	1480	1087	236
Mélanome malin peau	0	2	0	21	94	192	270	418	486	485	221	18
Méningites	102	24	18	24	22	37	52	79	125	105	37	0
Pharmacodépendance toxicomanie	0	0	0	25	81	74	21	10	4	7	0	0
Pneumonie	46	23	25	33	44	137	349	491	1434	4895	5707	1610
SIDA et maladies à VIH	0	0	2	9	40	317	252	121	48	14	4	2
Suicides	0	0	70	1431	2693	3280	4000	2968	2887	2336	685	57
Syndrome de mort subite nourrisson	854	0	0	0	0	0	0	0	0	0	0	0
Tuberculose	1	3	1	11	43	95	257	365	687	892	416	31
Tumeur maligne autres parties utérus	1	0	0	6	27	146	486	847	1394	1611	676	70
Tumeur maligne ovaire	0	0	5	20	42	159	694	1090	1578	1530	510	38
Tumeur maligne lèvre cavité buccale et pharynx	1	0	3	18	44	388	2391	2682	2316	1459	378	46
Tumeur maligne prostate	0	0	0	4	2	8	139	924	3710	7034	3375	318
Tumeur maligne vessie	0	0	2	1	7	56	371	1011	2230	3183	1343	120
Tumeur maligne estomac	0	0	0	10	63	241	868	1766	3890	4925	1896	124
Tumeur maligne col utérus	0	0	0	1	30	157	312	300	322	309	92	4
Tumeur maligne côlon	0	0	1	14	56	249	1094	2551	5470	8182	3743	394
Tumeur maligne foie et voies biliaires intrahépatiques	1	6	5	22	60	179	863	2031	3622	3570	1027	78
Tumeur maligne larynx trachée bronches et poumon	1	2	1	13	108	1286	7288	12059	15460	11747	2500	143
Tumeur maligne oesophage	0	0	0	0	14	205	1469	2206	2727	2064	547	35
Tumeur maligne rectum et anus	1	0	0	8	41	130	585	1336	2599	3122	1308	117
Tumeur maligne rein	3	16	21	11	22	96	435	863	1472	1653	604	49

2.2 – L'ACP du tableau de contingence





- Mise en évidence d'un **effet taille** prédominant (88% de l'information)
- La structure géopolitique du vote s'efface devant l'effet taille



2.3 – Les deux tableaux analysés en AFC

Poids des lignes et colonnes

Contrairement à l'ACP, les lignes et les colonnes sont affectées de **masses différentes**

Une modalité a d'autant plus de poids qu'elle a été observée un grand nombre de fois

	1	...	<i>j</i>	...	<i>J</i>	Σ
1						
⋮						
<i>i</i>			<i>n_{ij}</i>			<i>n_{i•}</i>
⋮						
<i>I</i>						

Masse ligne (*i*)

$$m_i = \frac{n_{i\bullet}}{n}$$

Σ			<i>n_{•j}</i>			<i>n</i>
---	--	--	-----------------------	--	--	----------

Masse colonne (*j*) : $m_j = \frac{n_{\bullet j}}{n}$

Les deux tableaux de profils



Tableau des profils lignes

	1	...	j	...	J	Σ
1						1
\vdots						
i	$\frac{n_{i1}}{n_{i\bullet}}$...	$\frac{n_{ij}}{n_{i\bullet}}$...	$\frac{n_{iJ}}{n_{i\bullet}}$	1
\vdots						
I						1

	1	...	j	...	J	Σ
1						
\vdots						
i			n_{ij}			$n_{i\bullet}$
\vdots						
I						

Σ			$n_{\bullet j}$			n
----------	--	--	-----------------	--	--	-----

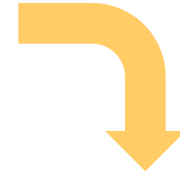
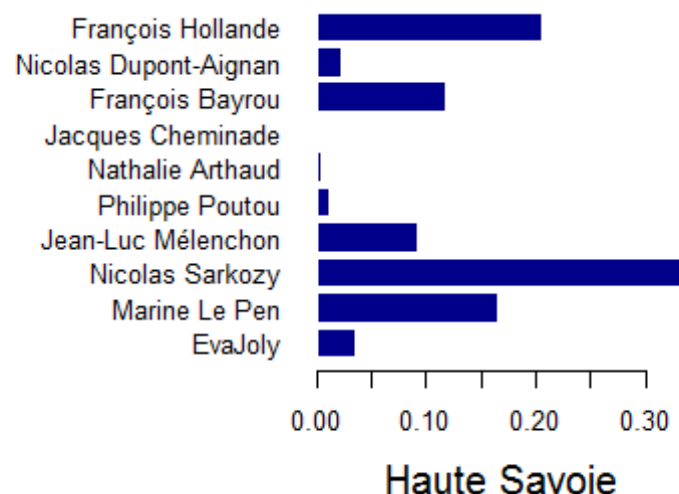
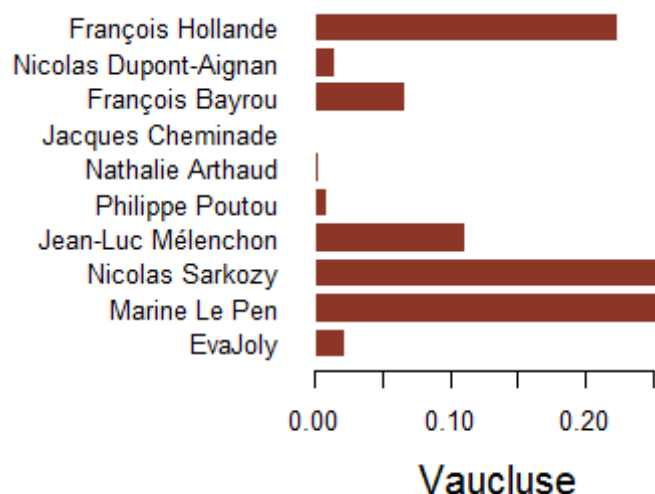
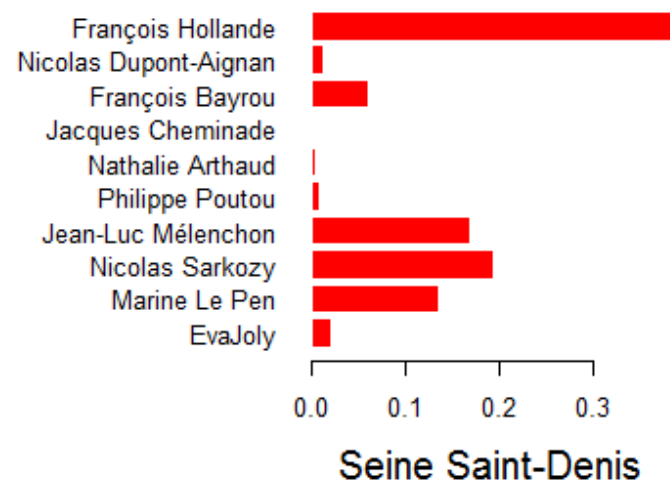
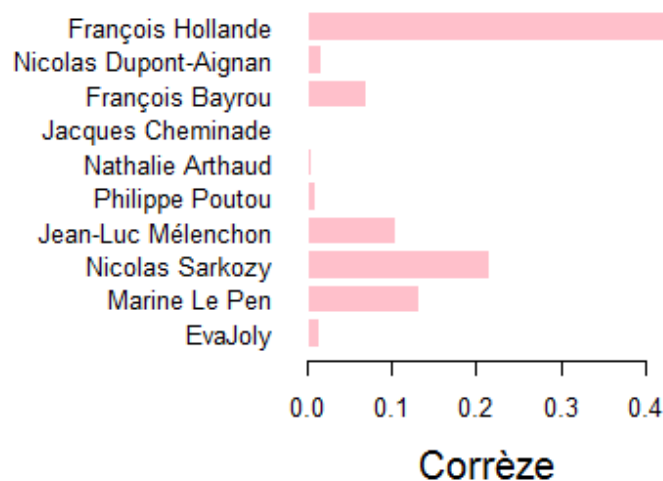


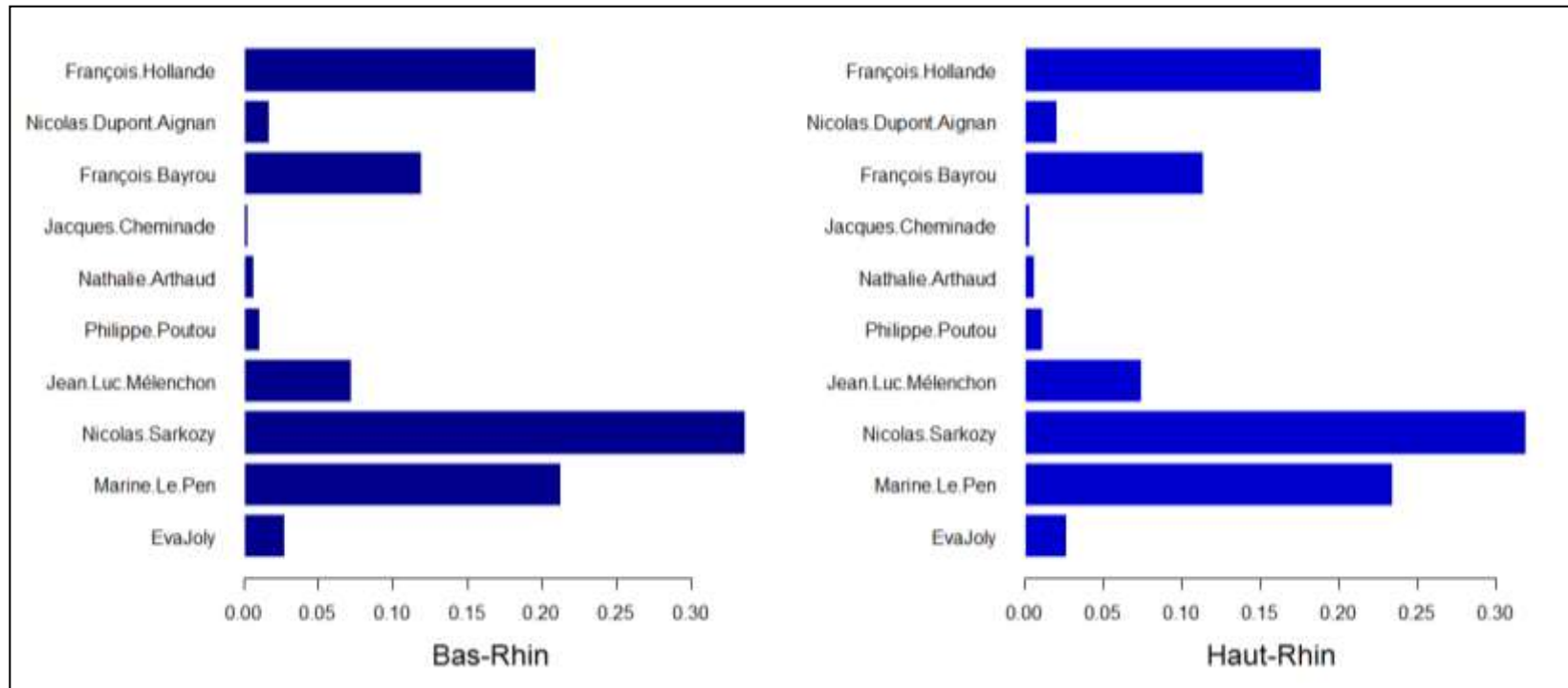
Tableau des profils colonnes

	1	...	j	...	J
1			$\frac{n_{1j}}{n_{\bullet j}}$		
\vdots					
i			$\frac{n_{ij}}{n_{\bullet j}}$		
\vdots					
I			$\frac{n_{Ij}}{n_{\bullet j}}$		

Σ	1		1		1
----------	---	--	---	--	---

Exemples de profils lignes (départements)

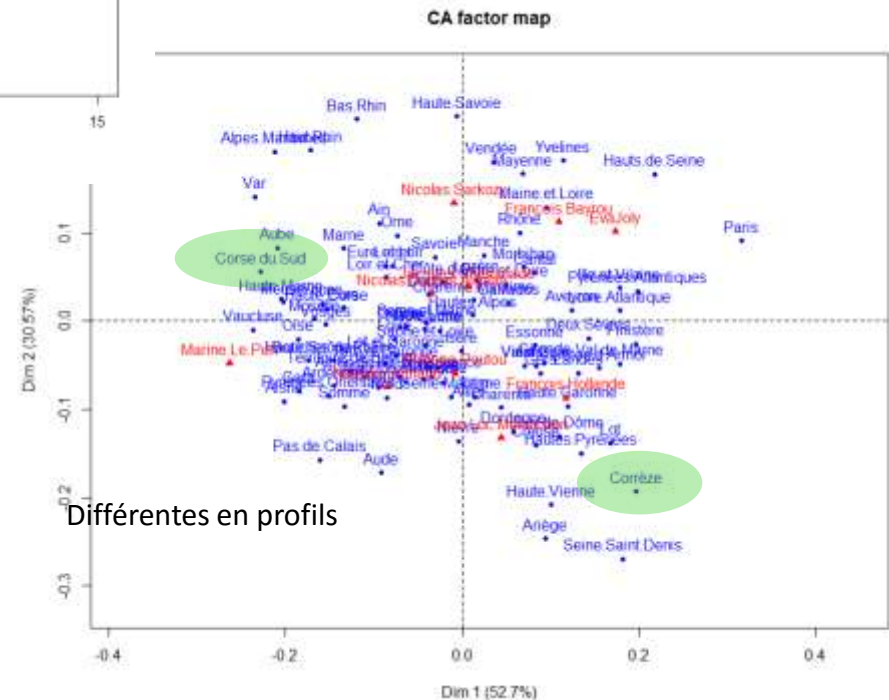
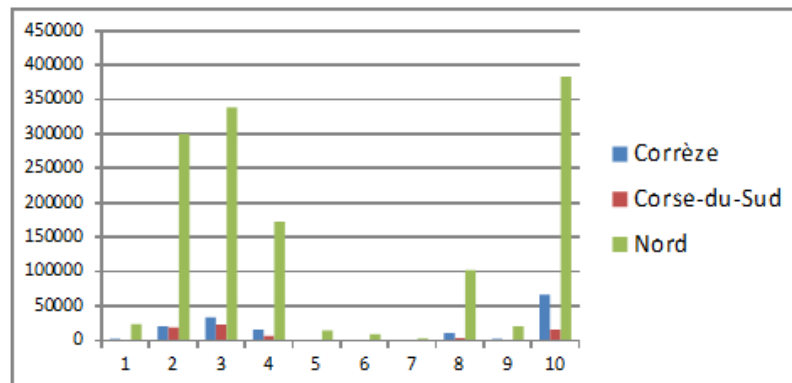
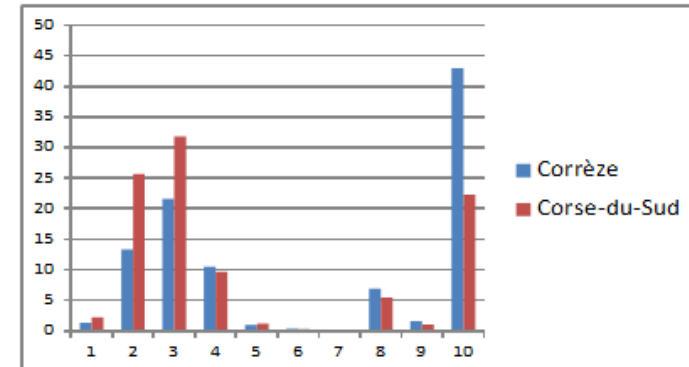
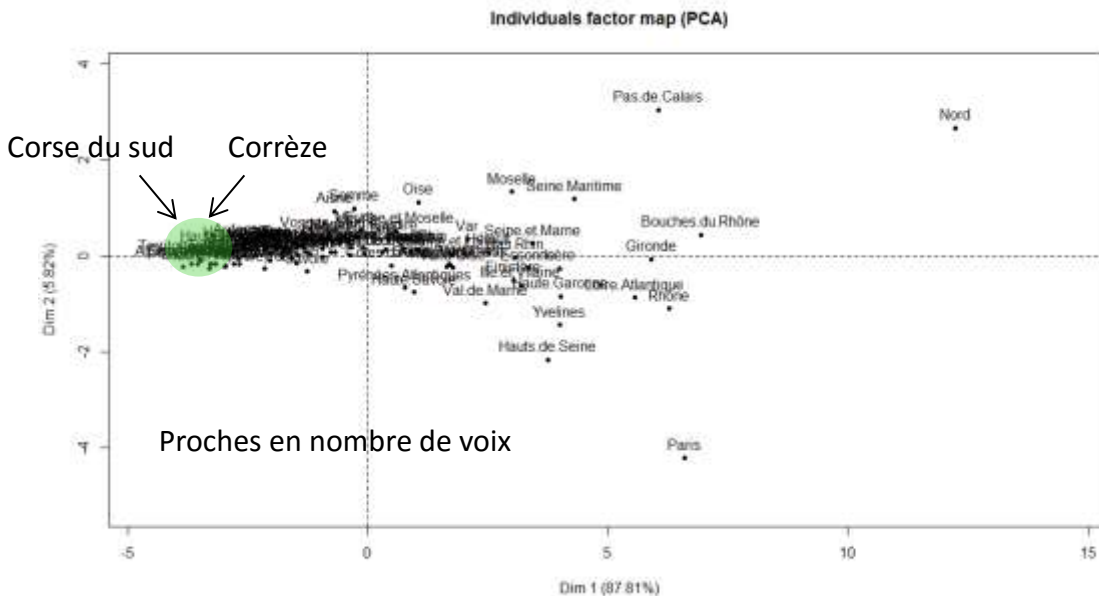




```
par(mar=c(5,10,4,2), mfrow=c(1,2))

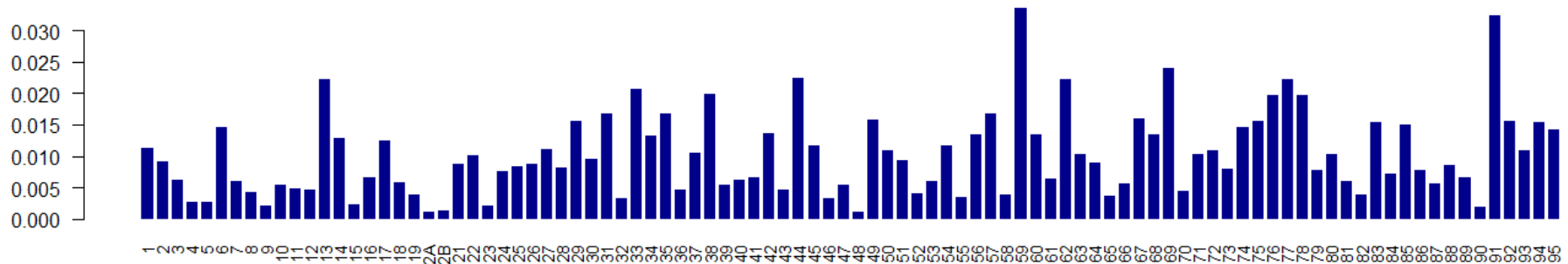
barplot(t(pres12)[1:10, 68]/sum(t(pres12)[1:10, 68]),
border="white", col="blue4", horiz=T, xlab="Bas-Rhin", cex.lab=1.5, las=1)

barplot(t(pres12)[1:10, 69]/sum(t(pres12)[1:10, 69]),
border="white", col="blue3", horiz=T, xlab="Haut-Rhin", cex.lab=1.5, las=1)
```

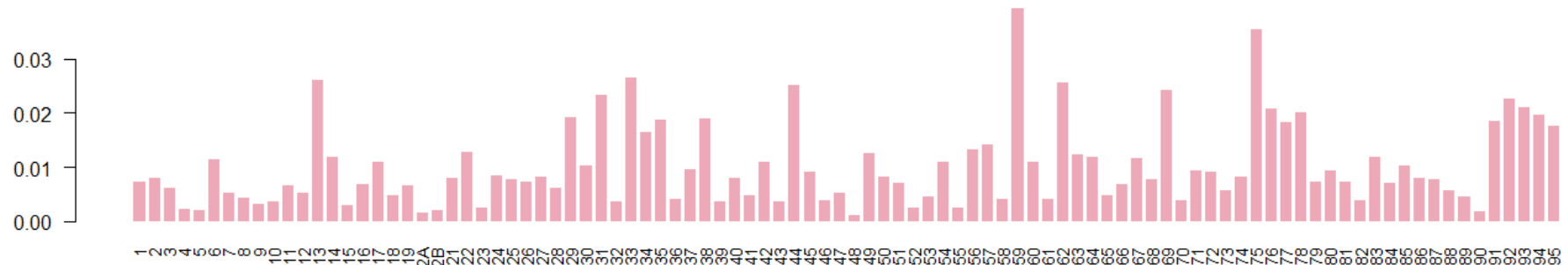



Exemples de profils colonnes (départements)

Nicolas Dupond Aignan



François Hollande



```
barplot(pres12[1:96,9]/sum(pres12[1:96,9]), names.arg=row.names(pres12)[1:96], border="white",
col="blue4", horiz=F, main="Nicolas Dupond Aignan", cex.main=1.5, cex.names=0.8, las=2)
```

```
barplot(pres12[1:96,10]/sum(pres12[1:96,10]), names.arg=row.names(pres12)[1:96], border="white",
col="pink2", horiz=F, main="François Hollande", cex.main=1.5, cex.names=0.8, las=2)
```

2.4 – L'AFC comme l'analyse d'un lien de dépendance

Tableau croisé ou de contingence

Lorsque des données s'inscrivent dans un tableau croisé

Question usuelle :

Existe-t-il un **lien de dépendance** entre les deux variables ?

Réponse :

Le test statistique du χ^2 (khi2)

	1	...	j	...	J	Σ
1						
\vdots						
i			n_{ij}			$n_{i\bullet}$
\vdots						
I						

Σ			$n_{\bullet j}$			n
----------	--	--	-----------------	--	--	-----

Qu'est-ce que l'indépendance ?

« Les valeurs prises par une variable ne dépendent pas des valeurs prises par l'autre variable »

$$P(A|B) = P(A)$$

$$P(A \cap B) = P(A) \times P(B)$$

Indépendance « intuitive »

- Égalité des profils lignes
- Égalité des profils colonnes
- Égalité avec profil marginal

Un exemple

	H	F	Σ
L			
\bar{L}			

Σ			
----------	--	--	--

En cas de lien de dépendance ?

Quelles sont les cases du tableau ayant fortement contribué au lien de dépendance ?

Calcul des **contributions au khi2** (ou D^2)

$$D^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(n_{ij} - t_{ij})^2}{t_{ij}}$$

Mesure globale des écarts entre

- Effectifs réellement observés n_{ij}
- Effectifs théoriques t_{ij} (attendus) si indépendance

Contribution de la case (i, j) $\frac{(n_{ij} - t_{ij})^2}{t_{ij}}$

Contribution en % $\frac{(n_{ij} - t_{ij})^2}{t_{ij}} / D^2$

Illustration : enquête OGM

V1 : « Couleur politique »

V2 : « Position / Culture OGM en France ? » : Favorable / Défavorable ?

Effectifs observés

	Defav.	Favor.	TOT
Centre	18	14	32
Ext.gauche	9	0	9
PS	37	10	47
UMP	16	24	40
Verts	7	0	7
TOT	87	48	135

Profils lignes

		Defav.	Favor.	TOT
	Centre	56,20	43,80	100
	Ext.gauche	100,00	0,00	100
	PS	78,70	21,30	100
	UMP	40,00	60,00	100
	Verts	100,00	0,00	100
	Profil ligne moyen	64,44	35,56	100

Effectifs théoriques

	Defav.	Favor.	TOT
Centre	20,62	11,38	32
Ext.gauche	5,80	3,20	9
PS	30,29	16,71	47
UMP	25,78	14,22	40
Verts	4,51	2,49	7
TOT	87	48	135

Contributions au chi2

	Defav.	Favor.	
Centre	0,33	0,6	
Ext.gauche	1,77	3,2	
PS	1,49	2,7	
UMP	3,71	6,72	
Verts	1,37	2,49	
			24,38

Pearson's Chi-squared test

X-squared = 24.3785,
df = 4,
p-value = 6.706e-05

Effectifs observés

	Defav.	Favor.	TOT
Centre	18	14	32
Ext.gauche	9	0	9
PS	37	10	47
UMP	16	24	40
Verts	7	0	7
TOT	87	48	135

Effectifs théoriques

	Defav.	Favor.	TOT
Centre	20,62	11,38	32
Ext.gauche	5,80	3,20	9
PS	30,29	16,71	47
UMP	25,78	14,22	40
Verts	4,51	2,49	7
TOT	87	48	135

Attirance entre deux modalités

L'effectif observé est **supérieur** à celui attendu en cas d'indépendance

Répulsion entre deux modalités

L'effectif observé est **inférieur** à celui attendu en cas d'indépendance

Nouveaux objectifs pour l'AFC

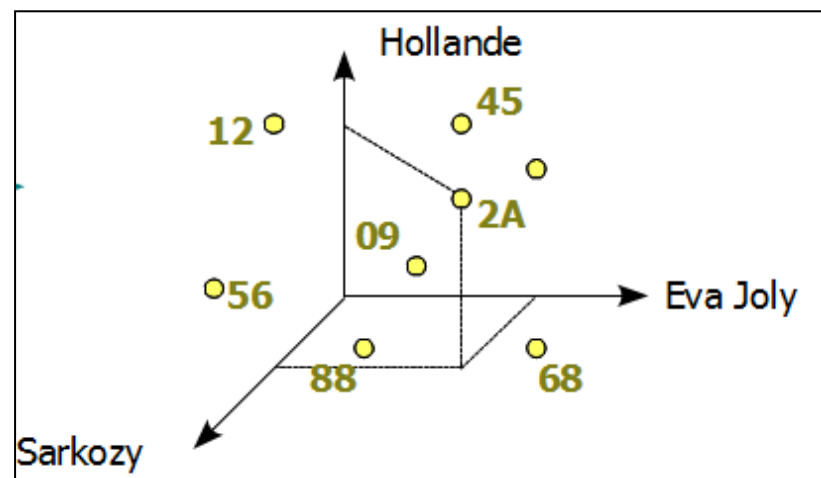
- Lignes et colonnes du tableau sont-elles indépendantes ?
Les % votes obtenus par les candidats dépendent-ils du département ?
- Comment **s'organise ce lien** de dépendance ?
- Y a-t-il des associations (ou **correspondances**) remarquables entre certaines modalités lignes ou colonnes ?
Notion **d'attraction** et de **répulsion** entre modalités
(cf. contributions au χ^2)
Certains départements sont-ils attirés par des candidats particuliers ?

2.5 – L'ajustement des nuages de profils

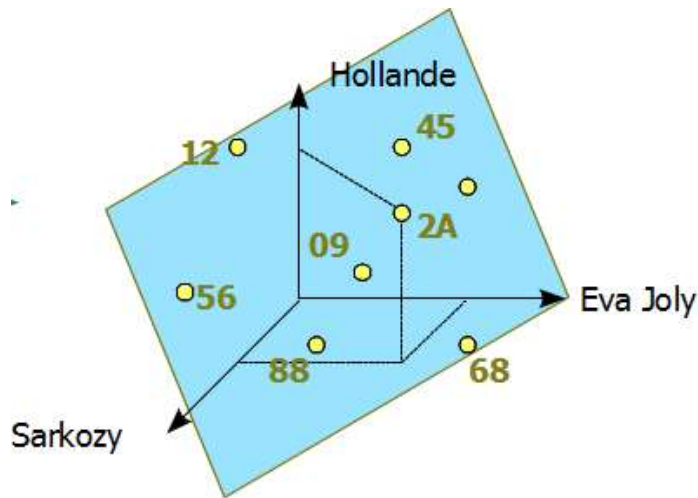
L'ajustement du nuage des profils lignes

- Les 96 départements forment un nuage de points dans un espace à 10 dimensions (9 dimensions en réalité !)
- Comment le visualiser dans un espace de plus faible dimension ?

	1	...	j	...	J	Σ
1						1
\vdots						
i	$\frac{n_{i1}}{n_{i\bullet}}$...	$\frac{n_{ij}}{n_{i\bullet}}$...	$\frac{n_{iJ}}{n_{i\bullet}}$	1
\vdots						
I						1



Détermination du meilleur plan de projection



En projection

- Les distances entre profils lignes sont aussi **peu déformées** que possible
- Le nuage projeté possède une **inertie maximale**

$$I = \sum_{i=1}^I m_i \times d^2(H_i, G)$$

Illustration dans des espaces de dimension 2 et 3

Quelle mesure de distance utiliser pour évaluer l'écart entre profils ?

La distance du khi2

$$d^2_{\chi^2}(i, l) = \sum_{j=1}^J \frac{1}{m_{\bullet j}} \left(\frac{n_{ij}}{n_{i\bullet}} - \frac{n_{lj}}{n_{l\bullet}} \right)^2$$

Deux propriétés importantes de la distance du khi2

- Les écarts portant sur des colonnes de faible effectifs sont revalorisés
- Principe **d'équivalence distributionnelle** : si deux lignes (resp. deux colonnes) ont un profil identique, leur agrégation ne modifie pas les mesures de distance entre profils colonnes (resp. entre profils lignes)

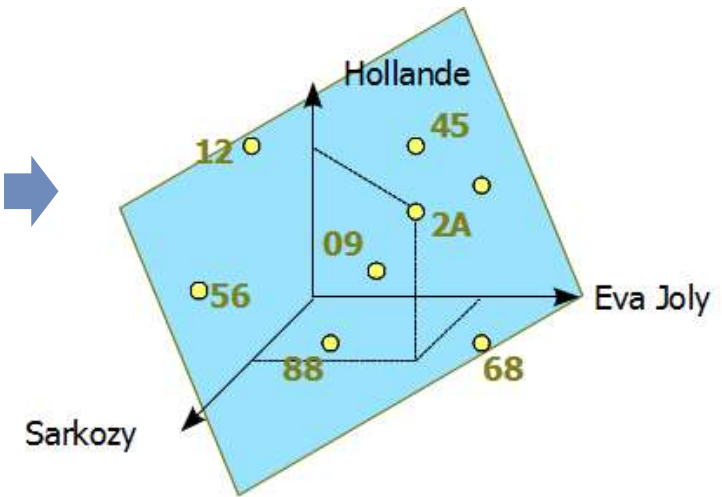
Application : robustesse vis-à-vis du découpage en modalités d'une variable nominale

Illustration

	D.A	Holl.
Dépt.1	0,04	0,26
Dépt.2	0,01	0,23
poids	0,05	0,25

L'ajustement du nuage des profils lignes

	1	...	j	...	J	Σ
1						1
\vdots						
i	$\frac{n_{i1}}{n_{i\bullet}}$...	$\frac{n_{ij}}{n_{i\bullet}}$...	$\frac{n_{iJ}}{n_{i\bullet}}$	1
\vdots						
I						1

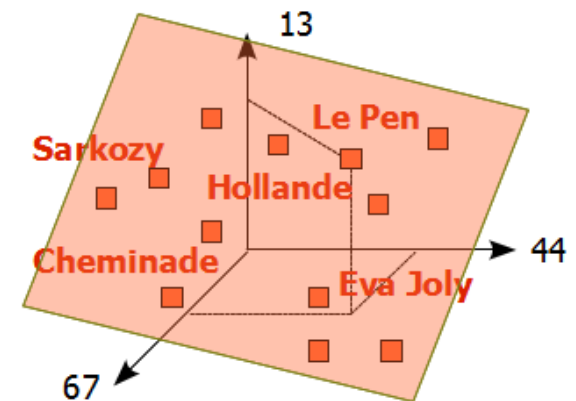


L'ajustement du nuage des profils colonnes

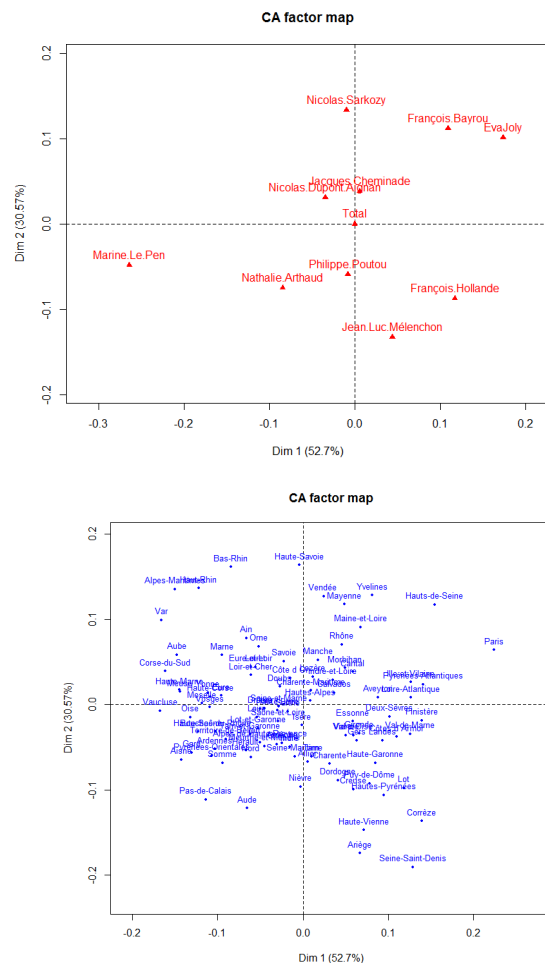
	1	...	j	...	J
1			$\frac{n_{1j}}{n_{\bullet j}}$		
\vdots					
i			$\frac{n_{ij}}{n_{\bullet j}}$		
\vdots					
I			$\frac{n_{Ij}}{n_{\bullet j}}$		



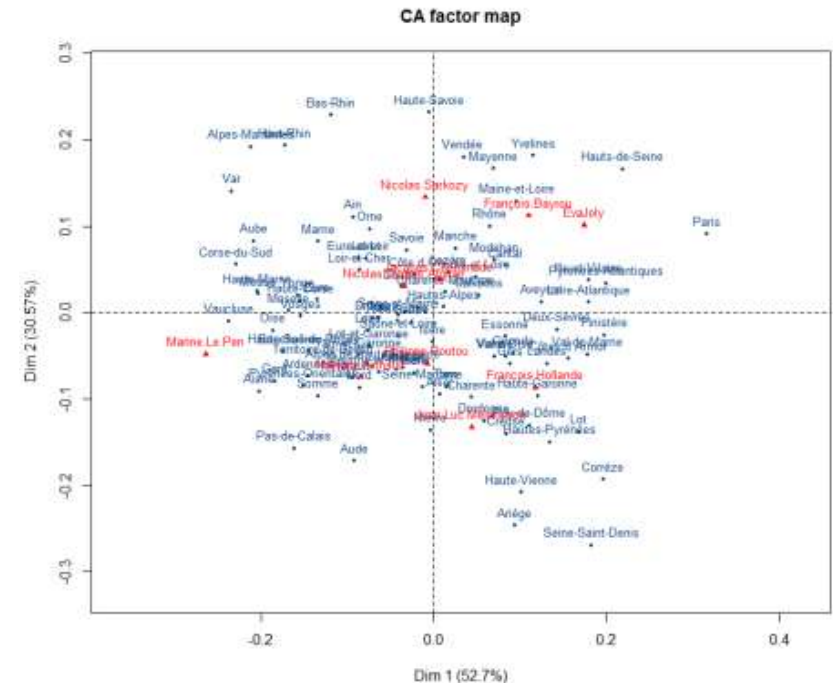
Σ	1		1		1
----------	---	--	---	--	---



Liaison entre les deux ajustements

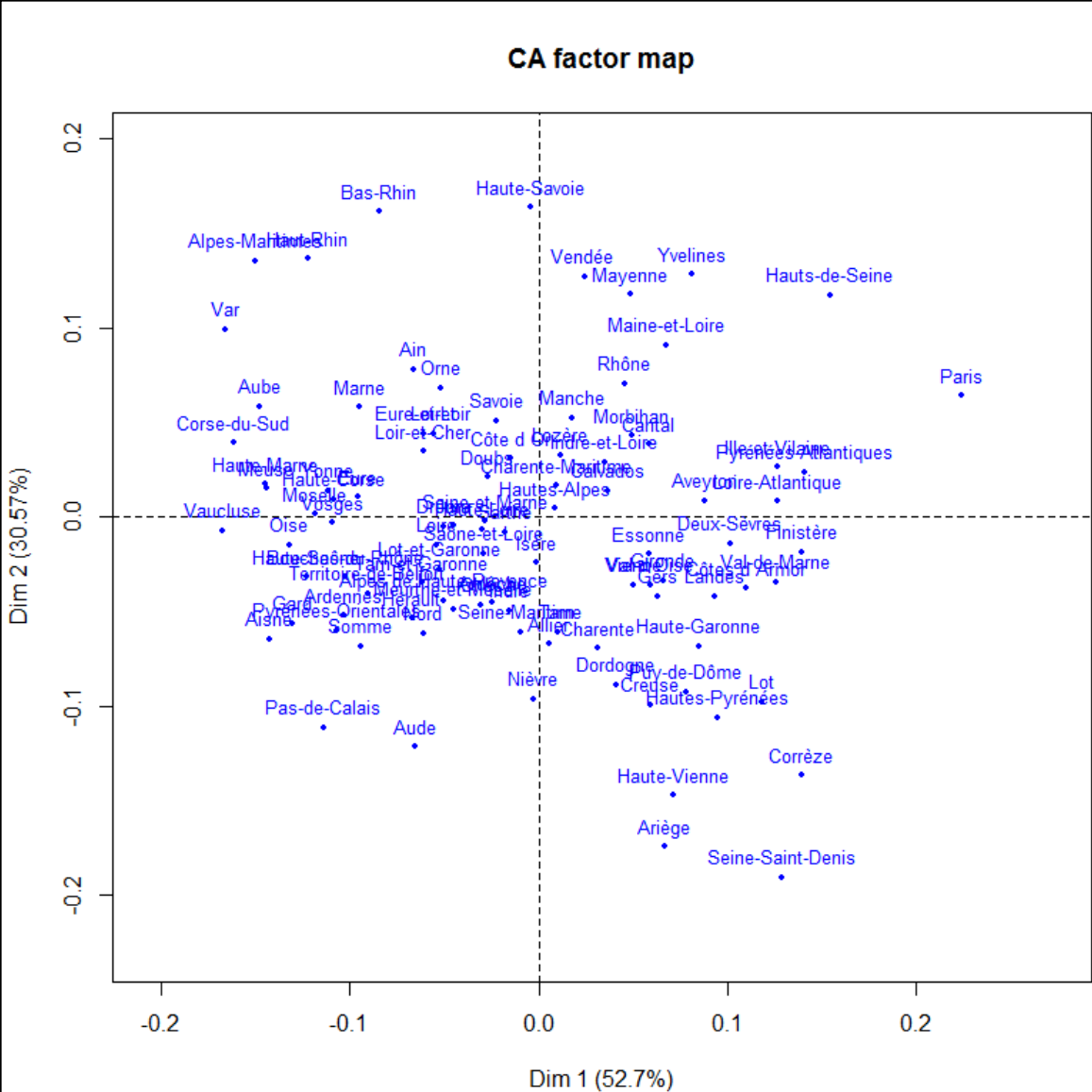


La représentation superposée

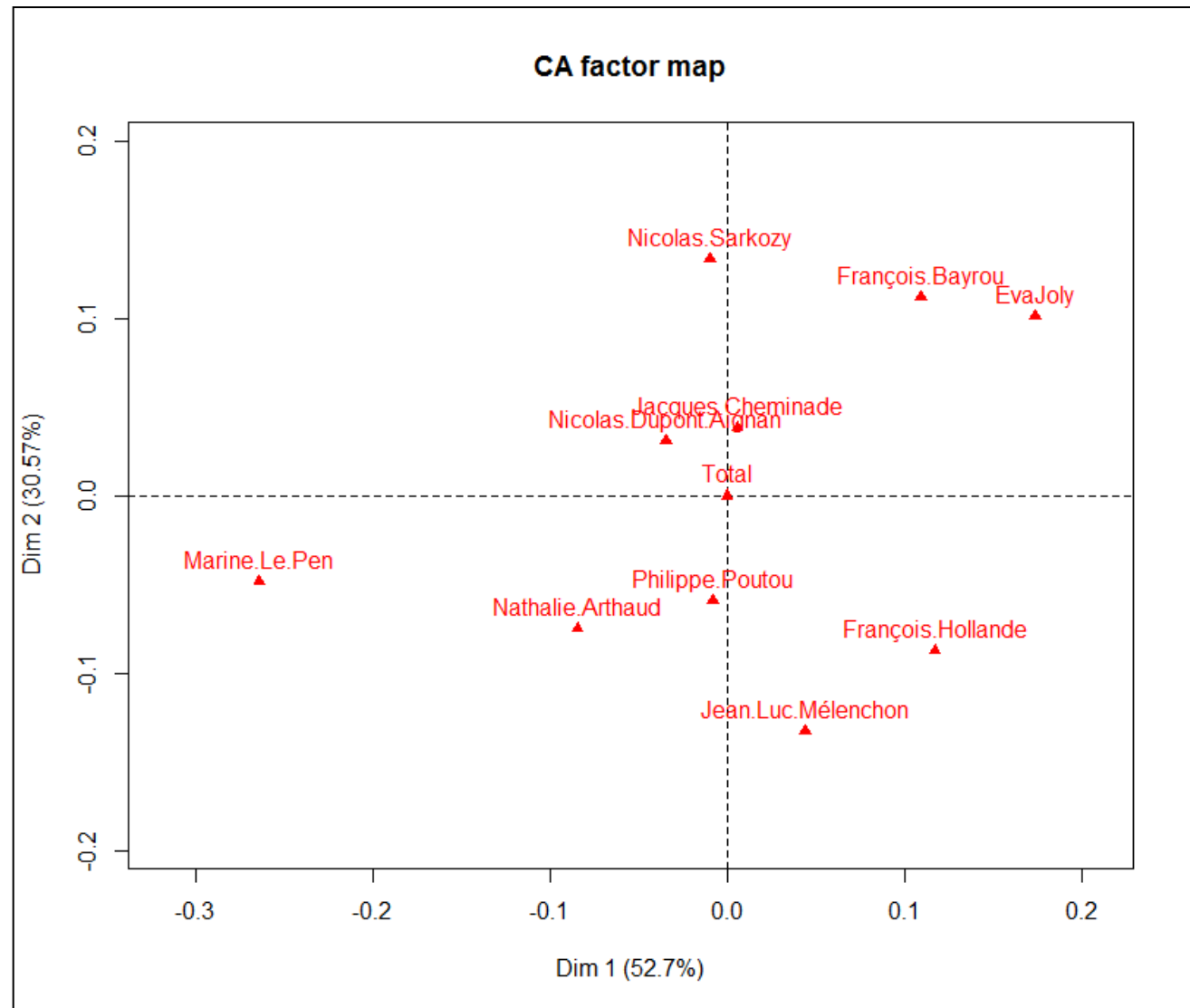


- Elle est « naturelle » car les objets lignes et colonnes sont de même nature
- Symétrie dans les deux ajustements réalisés par l'AFC

Départements



Le plan (1,2) des profils colonnes *Candidats*



2.6 – Interprétation des résultats d'une AFC

L'inertie totale

L'inertie totale du nuage des profils lignes est définie par

$$I(N_I) = \sum_{i=1}^I m_i \times d^2_{\chi^2}(i, G_I) \quad G_I : \text{centre de gravité des profils lignes}$$

On montre que l'inertie totale des deux nuages de profils est aussi égale à

$$I(N_I) = I(N_J) = \frac{D^2}{n}$$

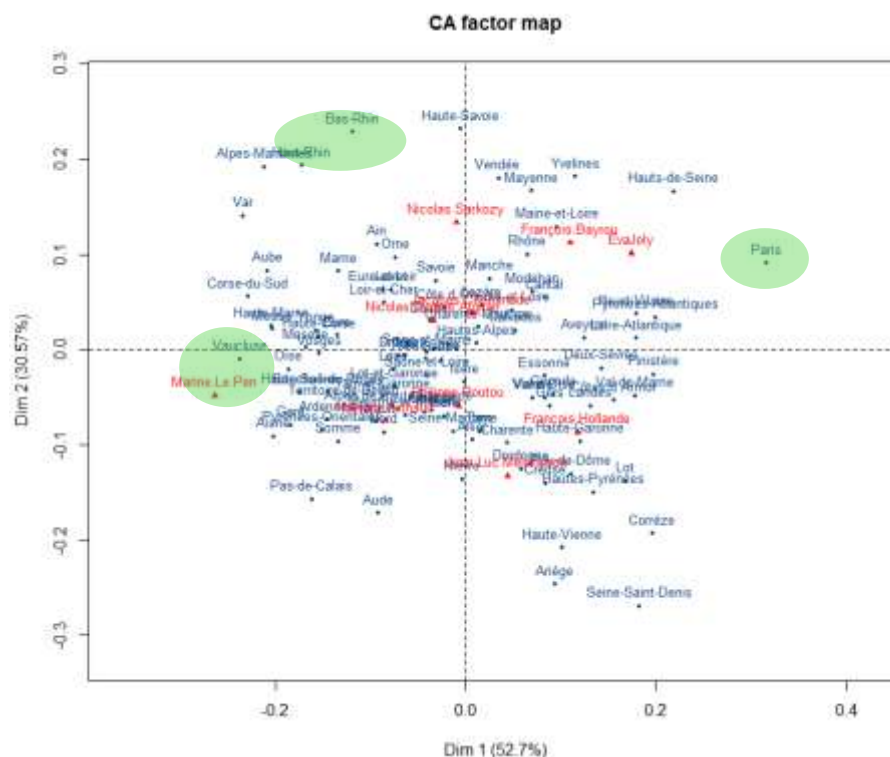
L'inertie totale est donc d'autant plus élevée que

- Les différences entre profils lignes (ou colonnes) sont importantes
- Le lien de dépendance entre les deux variables est élevé

En recherchant des **axes d'inertie élevée**, on met en évidence les éléments ayant fortement contribué au Khi^2 , *i.e.* au lien de dépendance :

- Modalités de **profils similaires ou très différents**
- **Associations remarquables** entre modalités

L'AFC permet de **visualiser** comment s'organise le **lien de dépendance** structuré au sein d'un tableau de contingence

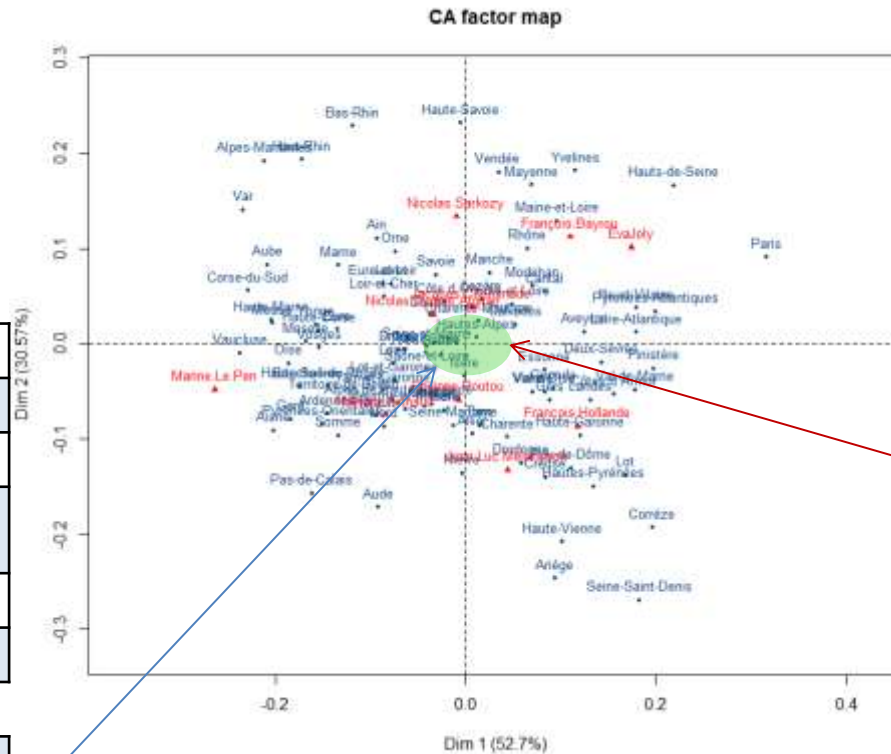


Centre de gravité, profil ligne moyen, profil colonne moyen

Plus un élément s'écarte du centre de gravité,
plus son profil est atypique par rapport au profil moyen

Barycentre des
profils lignes
départements

	1	...	j	...	J	Σ
1						1
\vdots						
i	$\frac{n_{i1}}{n_{i\bullet}}$...	$\frac{n_{ij}}{n_{i\bullet}}$...	$\frac{n_{iJ}}{n_{i\bullet}}$	1
\vdots						
I						1
G_I	$\frac{n_{\bullet 1}}{n}$...	$\frac{n_{\bullet j}}{n}$...	$\frac{n_{\bullet J}}{n}$	1

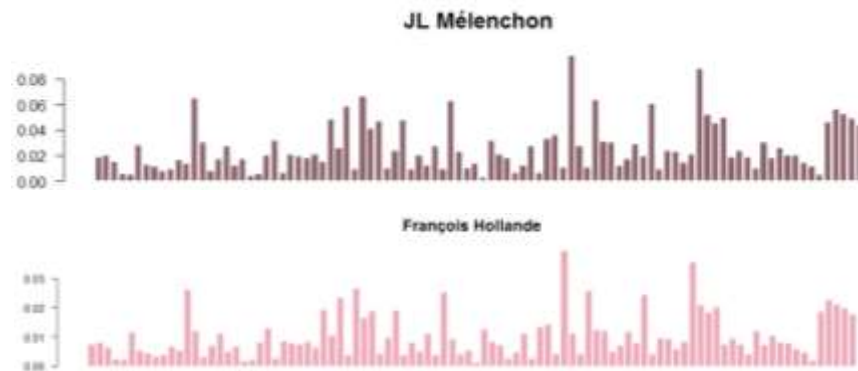
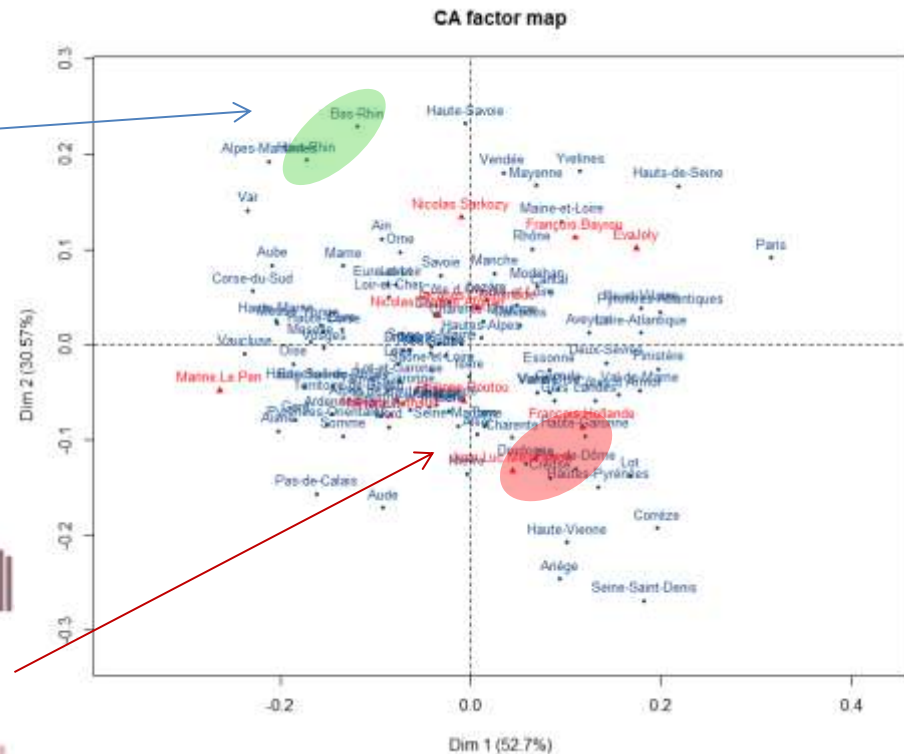
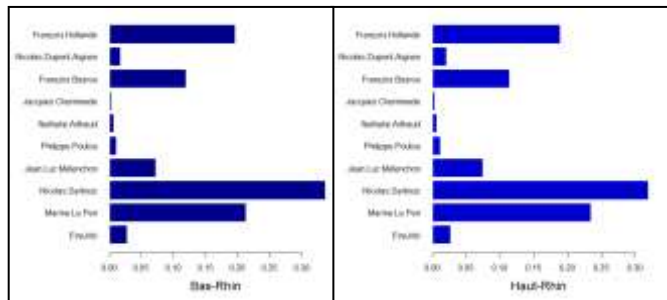


Barycentre des
profils colonnes
candidat

	1	...	j	...	J	G_j
1			$\frac{n_{1j}}{n_{\bullet j}}$			$\frac{n_{1\bullet}}{n}$
\vdots						
i			$\frac{n_{ij}}{n_{\bullet j}}$			$\frac{n_{i\bullet}}{n}$
\vdots						
I			$\frac{n_{Ij}}{n_{\bullet j}}$			$\frac{n_{I\bullet}}{n}$
Σ	1		1		1	1

Proximité entre profils lignes ou entre profils colonnes

Deux points proches (et bien représentés !) ont un profil similaire



Relations quasi barycentriques ou de transition

Proximité entre lignes et colonnes

- **Une ligne (i)** est au « quasi-barycentre » de l'ensemble des colonnes (j) où chaque colonne est affectée du poids $\frac{n_{ij}}{n_{i.}}$

$$F_k(i) = \frac{1}{\sqrt{\lambda_k}} \sum_{j=1}^J \left(\frac{n_{ij}}{n_{i.}} \right) G_k(j)$$

- **Une colonne (j)** est au « quasi-barycentre » de l'ensemble des lignes (i) où chaque ligne est affectée du poids $\frac{n_{ij}}{n_{.j}}$

$$G_k(j) = \frac{1}{\sqrt{\lambda_k}} \sum_{i=1}^I \left(\frac{n_{ij}}{n_{.j}} \right) F_k(i)$$

Illustration d'une relation barycentrique

Un département
(respectivement un candidat)
est « au quasi barycentre »
des dix candidats
(respectivement des 96
départements)

Phénomènes *d'attraction* et de
répulsion !

Exemple du « Rhône »

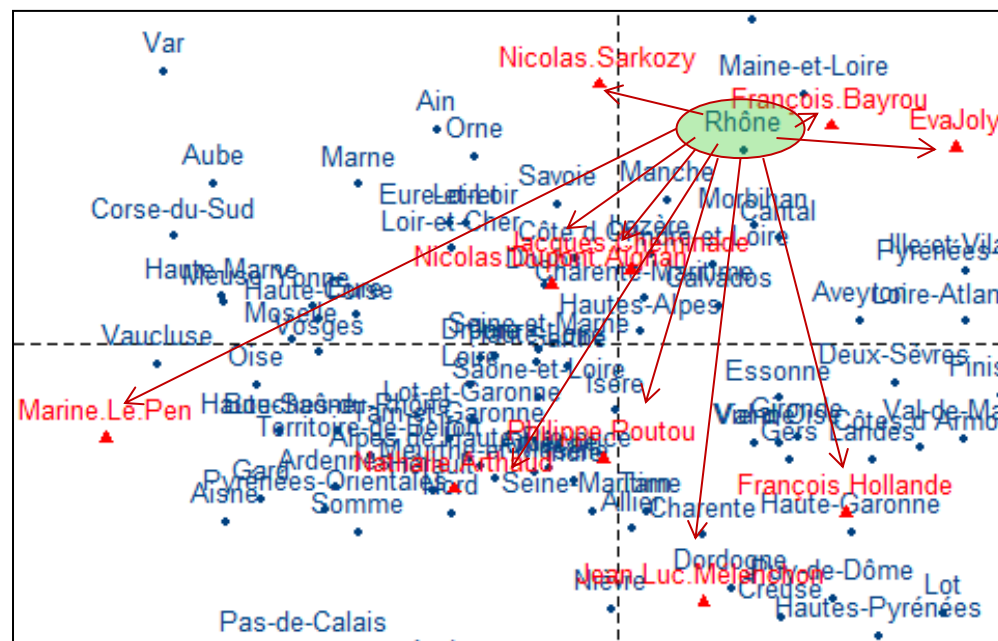
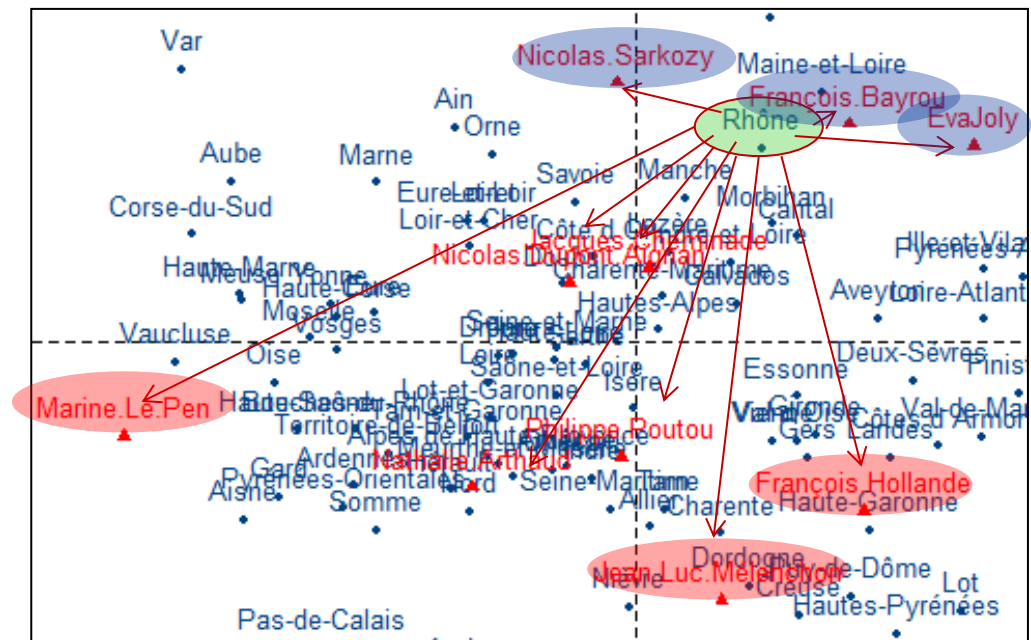


Illustration d'une relation barycentrique

Exemple du « Rhône »

Le Rhône est

- « attiré » par les modalités NS, FB, EJ
- « repoussé » par les modalités MLP, JLM, FH



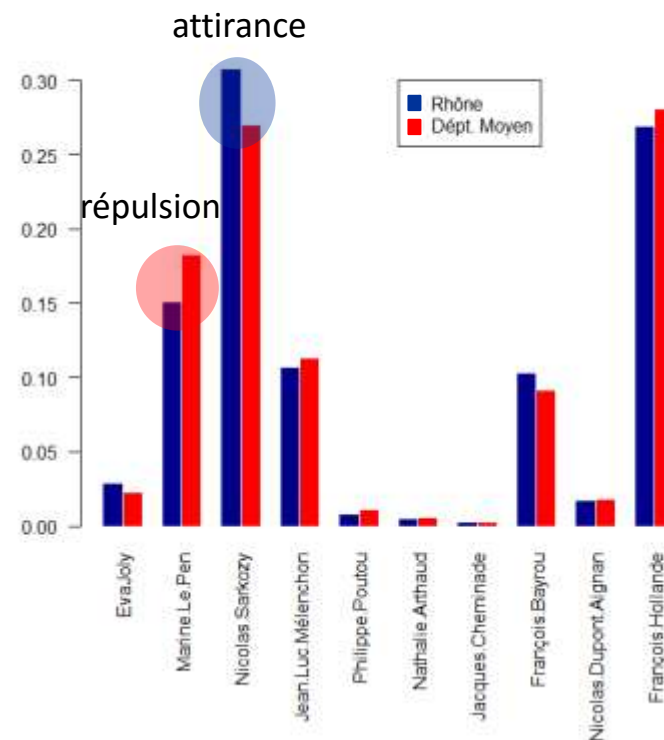
Attirance / Répulsion entre profils

- **Attirance.** Une modalité est attirée par une autre si l'effectif observé pour ce couple de modalités est **supérieur** à celui que l'on observerait en cas d'indépendance

$$n_{ij} > t_{ij} \Leftrightarrow \frac{n_{ij}}{n_{i\bullet}} > \frac{n_{\bullet j}}{n}$$

- **Répulsion.** Une modalité est repoussée par une autre si l'effectif observé pour ce couple de modalités est **inférieur** à celui que l'on observerait en cas d'indépendance

$$n_{ij} < t_{ij} \Leftrightarrow \frac{n_{ij}}{n_{i\bullet}} < \frac{n_{\bullet j}}{n}$$



Nombre d'axes, inertie des axes

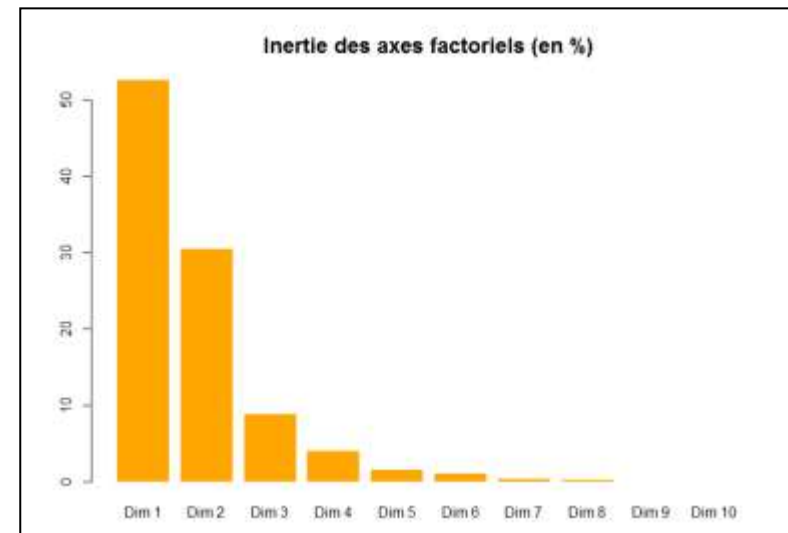
Nombre d'axes extraits par l'AFC

$$\min(I - 1, J - 1)$$

```
> res$eig
      eigenvalue percentage of variance cumulative percentage of variance
dim 1  1.870174e-02          5.270047e+01          52.70047
dim 2  1.084876e-02          3.057119e+01          83.27167
dim 3  3.165261e-03          8.919530e+00          92.19120
dim 4  1.450863e-03          4.088450e+00          96.27965
dim 5  5.938874e-04          1.673542e+00          97.95319
dim 6  4.096087e-04          1.154255e+00          99.10744
dim 7  1.687078e-04          4.754093e-01          99.58285
dim 8  1.340162e-04          3.776503e-01          99.96050
dim 9  1.401680e-05          3.949856e-02          100.00000
```

Décroissance de l'inertie « Éboulis » des valeurs propres

```
barplot (res$eig[,2],
names=paste("Dim",1:length(res$eig[,2])),
main="Inertie (en %) des axes factoriels",
col="orange", border="white")
```



Combien d'axes retenir ?

- Critère de Kaiser : l'inertie moyenne**

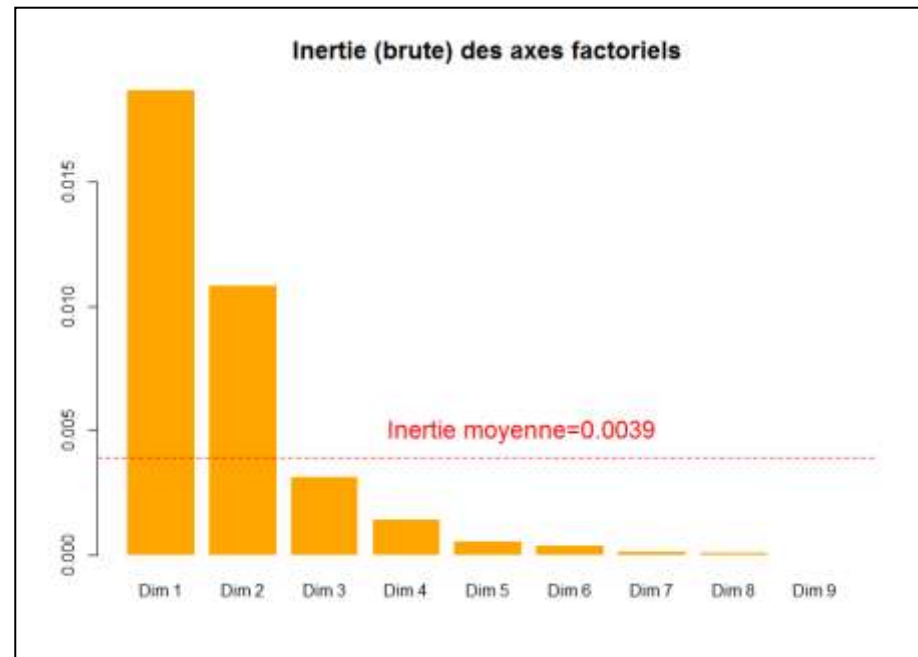
Les axes intéressants sont ceux dont l'inertie est supérieure à l'inertie moyenne

```
> sum(res$eig[,1])/9
[1] 0.003942984
```

Valeur propre moyenne = 0,0039

→ on retient 2 axes

```
barplot(res$eig[1:9,1], names=paste("Dim",1:9),
main="Inertie (brute) des axes factoriels", col="orange",
border="white", cex.main=1.5)
abline(h=sum(res$eig[,1])/9, lty=2, col="red")
text(6,0.005,"Inertie moyenne=0.0039", col="red", cex=1.5)
```



- **Scree – test de Cattell** : la recherche d'un **coude**

Illustration

Axe	inertie	diff 1e	diff 2e
1	0,01870		
2	0,01085	7,85E-03	
3	0,00317	7,68E-03	1,69E-04
4	0,00145	1,71E-03	5,97E-03
5	0,00059	8,57E-04	8,57E-04
6	0,00041	1,84E-04	6,73E-04
7	0,00017	2,41E-04	-5,66E-05
8	0,00013	3,47E-05	2,06E-04
9	0,00001	1,20E-04	-8,53E-05
Total	0,0355		

Un critère incontournable : la possibilité **d'interpréter** les dimensions...

Inertie maximale pour un axe

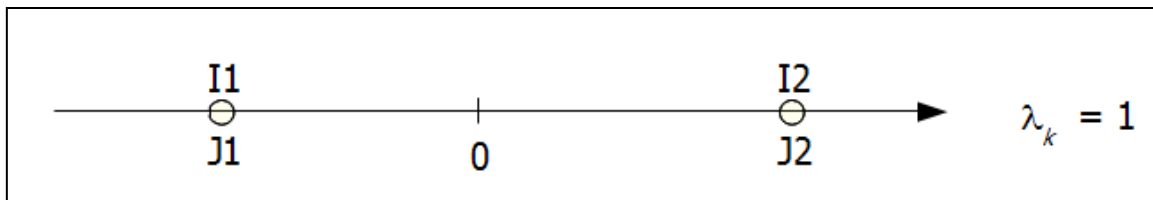
$$\lambda_k \leq 1$$

- Le cas d'une inertie maximale**

On peut réarranger lignes et colonnes du tableau de contingence de façon à obtenir la structure ci-contre :

Les modalités lignes et colonnes sont partitionnées en deux sous-ensembles qui s'opposent le long de la dimension

	J1	J2
I1		0
I2	0	



Illustration

Données

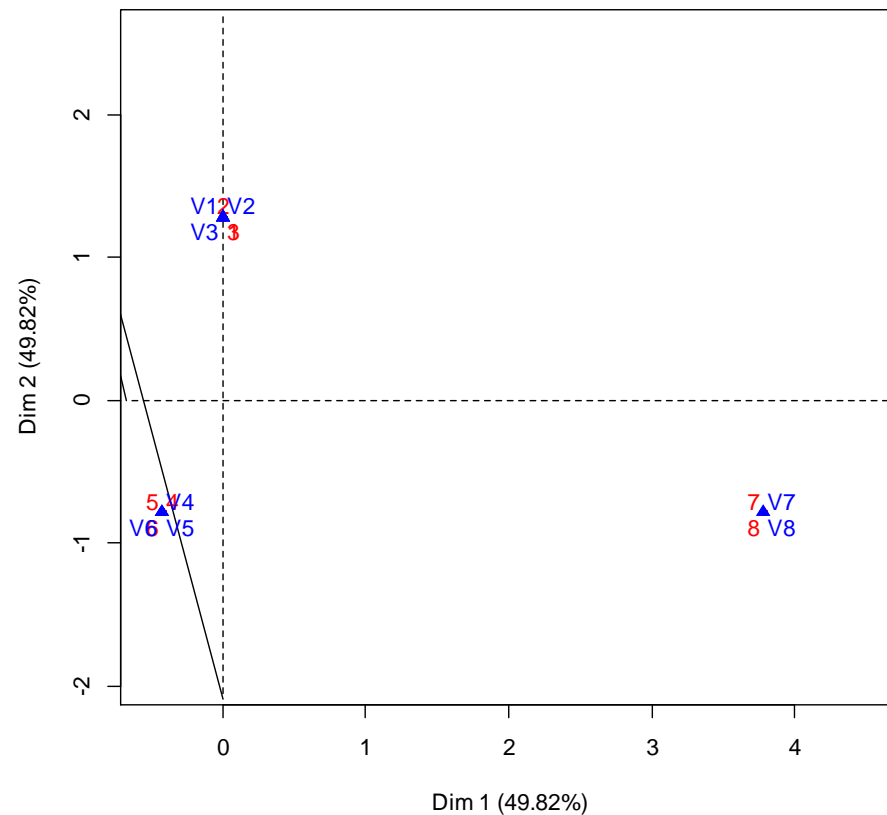
	V1	V2	V3	V4	V5	V6	V7	V8
1	5	5	5	0	0	0	0	0
2	5	6	5	0	0	0	0	0
3	6	5	6	0	0	0	0	0
4	0	0	0	8	8	7	0	0
5	0	0	0	7	8	8	0	0
6	0	0	0	8	9	8	0	0
7	0	0	0	0	0	0	2	2
8	0	0	0	0	0	0	2	2

Inerties

```

> res$eig
      eigenvalue
dim 1 1.000000e+00
dim 2 1.000000e+00
dim 3 5.055147e-03
dim 4 1.890359e-03
dim 5 1.482042e-04
dim 6 5.467097e-33
dim 7 5.145157e-36
  
```

CA factor map



Contribution d'une ligne ou d'une colonne

- **Contribution d'une modalité (i) à l'axe (k)**

$$CTR_k(i) = \frac{m_i \times c_{ik}^2}{\lambda_k}$$

avec

$$\left\{ \begin{array}{l} m_i = \frac{n_{i\bullet}}{n} \\ c_{ik} \\ \lambda_k \end{array} \right.$$

masse de (i)

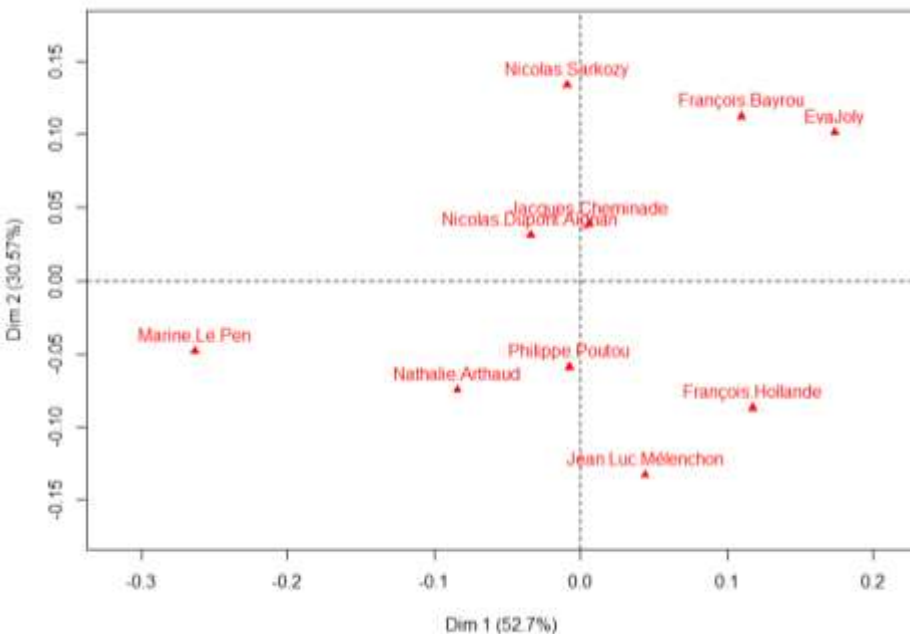
coordonnée de (i)
sur l'axe (k)

inertie de l'axe (k)

*L'individu le plus éloigné le long d'un axe
n'a pas nécessairement la plus forte contribution !*

Comparaison des coordonnées et des contributions des colonnes le long de l'axe 1

	Poids (en %)	Coordonnée	Contribution (%)
Eva.Joly	0,0228	0,1737	3,6729
Marine.Le.Pen	0,1830	-0,2638	68,0729
Nicolas.Sarkozy	0,2704	-0,0095	0,1298
Jean.Luc.Mélenchon	0,1129	0,0440	1,1688
Philippe.Poutou	0,0116	-0,0079	0,0039
Nathalie.Arthaud	0,0056	-0,0843	0,2143
Jacques.Cheminade	0,0025	0,0058	0,0004
François.Bayrou	0,0919	0,1097	5,9086
Nicolas.Dupont.Aignan	0,0182	-0,0343	0,1141
François.Hollande	0,2813	0,1174	20,7143
		Total CTR	100,0000



Contribution à l'inertie totale

> **res\$col\$inertia**

EvaJoly	Marine.Le.Pen	Nicolas.Sarkozy
1.756837e-03	1.322583e-02	5.481239e-03
Jean.Luc.Mélenchon	Philippe.Poutou	Nathalie.Arthaud
3.032689e-03	5.208093e-04	4.322629e-04
Jacques.Cheminade	François.Bayrou	Nicolas.Dupont.Aignan
2.291488e-05	3.887175e-03	7.576775e-04
François.Hollande		
6.369421e-03		

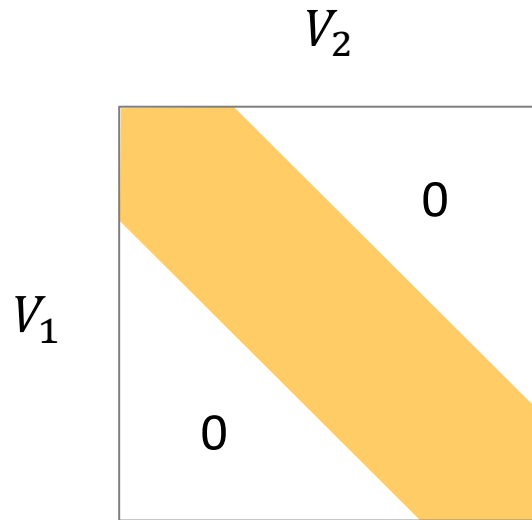
> **sort(res\$row\$inertia, decreasing=T)**

Paris	Hauts.de.Seine	Seine.Saint.Denis
3.705535e-03	1.819151e-03	1.763658e-03
Alpes.Maritimes	Var	Pas.de.Calais
1.663282e-03	1.478459e-03	1.334815e-03
Bas.Rhin	Bouches.du.Rhône	Yvelines
1.202429e-03	1.052391e-03	1.025199e-03
Pyrénées.Atlantiques	Haut.Rhin	Loire.Atlantique
8.835909e-04	8.641772e-04	7.605381e-04
Haute.Savoie	Finistère	Val.de.Marne
7.128989e-04	6.975364e-04	6.734772e-04

2.7 – L'effet Guttman en AFC

Un tableau de données présentant une **structure diagonale**

- Il existe une structure d'ordre sur les lignes et sur les colonnes
- Ces deux structures sont étroitement associées

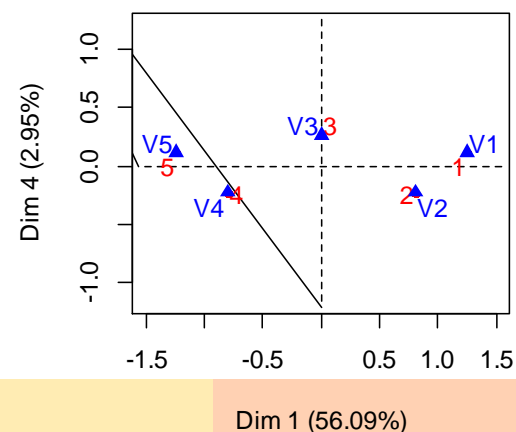
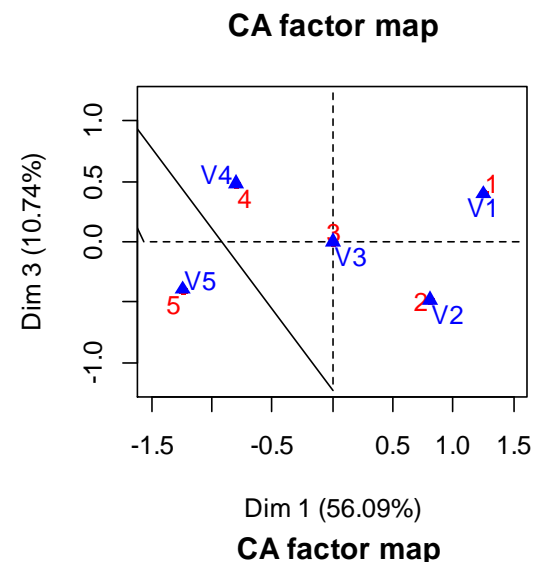
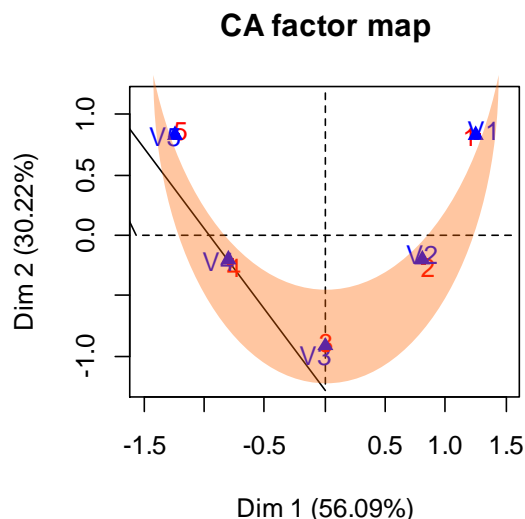


- Des cases présentant des valeurs élevées sur une bande diagonale
- Des valeurs nulles ou très faibles en dehors

Conséquence, illustration

Le facteur de rang (s) est un **polynôme de degré (s)** du premier facteur

	V1	V2	V3	V4	V5
1	8	3	0	0	0
2	3	8	3	0	0
3	0	3	8	3	0
4	0	0	3	8	3
5	0	0	0	3	8



- Tableau de données de type « scalogramme »
- Les modalités sont ordonnées selon l'axe 1
- Sur le plan (1,2) : un **effet « parabole »**

Dans quelle situation apparaît l'effet Guttman ?

- En présence de variables dont les modalités présentent un « ordre naturel »
- Lorsqu'on analyse des variables provenant de variables quantitatives découpées en classes