

ISFA 2ème année 2003-2004 - Analyse des données – Théorie

*Le sujet comprend 5 questions indépendantes très simples et un problème à 15 questions.
Évitez les réponses sans justification et le remplissage. Merci d'utiliser l'espace imparti pour vos réponses.*

Pour les 5 premières questions on considère la matrice A :

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}$$

1. _____

Donner l'inertie autour de l'origine des 4 points de définis par A avec la pondération uniforme et la métrique canonique.

2. _____

Donner les axes principaux de l'ACP normée de ce tableau.

3. _____

Donner les éléments du triplet de l'analyse des correspondances de A .

4. _____

Donner les valeurs propres de où est la transposée de A .

5. _____

Donner les valeurs singulières de A .

Les données qui font l'objet du problème sont listées en annexe. Elles forment deux tableaux édités côte à côte. Le premier, appelé tab1 a 94 lignes et 9 colonnes. Le second, appelé tab2 , a 94 lignes et 16 colonnes. Les 94 lignes sont les 94 départements de la France continentale (Figure 1).

Le premier tableau donne les résultats du premier tour des élections présidentielles de 1988. Les candidats sont désignés par des lettres minuscules :

a	b	c	d	
F. Mitterand	J. Chirac	R. Barre	J.M. Le Pen	
e	f	g	h	i
A. Lajoinie	A. Waechter	P. Juquin	A. Laguillier	P. Bousssel

Le second donne les résultats du premier tour des élections présidentielles de 2002. Les candidats sont désignés par des lettres majuscules :

A	B	C	D
J. Chirac	J.M. Le Pen	L. Jospin	F. Bayrou
E	F	G	H
A. Laguillier	P. Chevènement	J. Saint Josse	N. Mamère
I	J	K	L
O. Besancenot	A. Madelin	R. Hue	B. Mégret
M	N	O	P
C.Taubira	C. Lepage	C. Boutin	D. Gluksten
CI.			

Les données sont éditées dans l'annexe en pour mille arrondi à l'entier le plus proche et conservées avec cette précisions pour tous les calculs ce qui est exprimé dans R par :

```
(apply(tab1,1,sum)==1000)
[1] TRUE
```

```
all(apply(tab2,1,sum)==1000)
[1] TRUE
```

Le premier tableau formait un exemple de référence dans le logiciel Cartographie-2D © de Argo Infographie S.A., aujourd'hui disparu. Le second tableau a été calculé à partir des informations disponibles sur le site du Ministère de l'Intérieur :

On peut considérer que F. Mitterrand (PS), J. Chirac (RPR), R. Barre (UDF), J.M. Le Pen (FN) et A. Laguillier (LO) sont des personnalités bien connues. Pour mémoire, on rappelle que A. Lajoinie était le candidat du PCF, A. Waechter le représentant du mouvement écologiste, P. Juquin était communiste dissident et P. Bussel un autre représentant de l'extrême gauche (MPPT).

Les moyennes par colonnes dans chacun des tableaux sont :

```
(tab1,2,mean)
```

a	b	c	d	e	f	g	h	i
343.064	202.670	164.681	137.606	67.106	38.638	21.670	20.574	3.989

```
apply(tab2,2,mean)
```

A	B	C	D	E	F	G	H	I	J	K
196.106	169.543	157.149	67.298	58.447	53.830	51.053	50.553	44.957	37.691	34.681
L	M	N	O	P						
23.660	19.713	18.479	11.915	4.926						

6.

Donner la somme des moyennes des variables de chaque tableau.

7.

Les variances des variables de chaque tableau sont :

```
apply(tab1,2,var)*93/94
```

```
      a      b      c      d      e      f      g      h      i
1578.507 1580.966 1058.515 1760.579  868.350  110.295  36.604  14.500  1.138
```

```
apply(tab2,2,var)*93/94
```

```
      A      B      C      D      E      F      G      H      I      J
1022.010 1886.674  657.488  207.252  106.694  311.865  690.178  91.481  47.232  79.320
      K      L      M      N      O      P
 183.068  64.990  41.098  13.526  13.057  0.686
```

On s'intéresse ici à la variabilité spatiale (entre départements) du pourcentage des voies obtenues. Les données centrées sont cartographiées et donnent les figures 2 et 3. Chaque carré représente l'écart entre le pourcentage obtenu dans le département et le pourcentage moyen. Ce carré est blanc si l'écart est négatif, noir si il est positif, et la surface du carré est proportionnelle à la valeur absolue de l'écart. L'échelle est la même pour toutes les cartes.

Combien vaut le plus grand écart négatif observé pour F. Mitterrand en 1988 et où se produit-il ? Même question pour J.M. Le Pen en 2002 ? Combien vaut le plus grand écart positif pour J. Chirac en 1988 et où se produit-il ? Même question pour J.P. Chevènement en 2002.

8.

Rédiger un commentaire bref et concentré pouvant servir de légende aux figures 2 et 3.

9.

On appelle `tab3` le tableau à 94 lignes et 25 colonnes formé par la juxtaposition de `tab1` et `tab2`. On peut faire :

A : l'ACP de <code>tab1</code> centrée	B : l'ACP de <code>tab1</code> normée
C : l'ACP de <code>tab2</code> centrée	D : l'ACP de <code>tab2</code> normée
E : l'ACP de <code>tab3</code> centrée	F : l'ACP de <code>tab3</code> normée.

A laquelle de ces analyses correspond chacun des graphes de valeurs propres (numérotés de 1 à 6 dans le cartouche) représentés dans la figure 4 ?

10.

A est la matrice 94-9 centrée par colonne issue du premier tableau, **B** est la matrice centrée 94-16 issue du second tableau. Donner un majorant aussi petit que possible du rang de ces deux matrices.

11.

C la matrice centrée 94-24 issue du troisième tableau. Donner un majorant aussi petit que possible des rangs de cette troisième matrice.

12.

Numériquement, les valeurs singulières de **C** valent :

```
svd(scalewt(tab3,scale=F))$d
```

```
[1] 6.886e+02 5.252e+02 4.134e+02 2.650e+02 2.424e+02 2.049e+02 1.772e+02
[8] 1.169e+02 9.887e+01 8.426e+01 6.511e+01 5.748e+01 3.938e+01 3.367e+01
[15] 3.116e+01 2.946e+01 2.632e+01 2.381e+01 2.071e+01 1.586e+01 1.144e+01
[22] 8.811e+00 4.229e+00 1.415e-13 2.551e-14
```

Donner le rang des matrices de covariances associées aux matrices **A**, **B** et **C**.

13.

On utilise maintenant uniquement des ACP centrées. La première coordonnée des lignes de l'ACP centrée de **A** est cartographiée à gauche sur la figure 6. On appelle cette variable `Ac001`. La première coordonnée des lignes de l'ACP centrée de **B** est cartographiée à droite sur la figure 6. On appelle cette

variable `Bcool`.

Le nuage bivarié de `Acool` et `Bcool` est tracé avec la droite de régression de Y/X . On donne la première valeur propre de l'ACP centrée de **A**, à savoir 2773 et la première valeur propre de l'ACP centrée de **B**, à savoir 2698. Quelle est la corrélation entre ces deux variables ? Choisir cette valeur dans la série de propositions ci-dessous qui contient la valeur exacte :

-1.302, -0.9832, -0.8032, -0.5392, -0.3082, -0.0832,
0.0023, 0.0832, 0.3082, 0.8032, 0.9832, 1.302, 130.2

14. _____

Donner l'équation de la droite de régression de `Bccool` sur `Accool`.

15. _____

Donner l'équation de la droite de régression de `Accool` sur `Bccool` et tracer cette droite sur la figure.

16. _____

Donner l'équation de l'axe principal de ce nuage bivarié et placer cet axe sur la figure.

17. _____

Les deux ACP faites avec la fonction `dudi.pca` s'appellent respectivement `pca1` et `pca2`. Le premier axe principal de `pca1` est défini par :

a	b	c	d	e	f	g	h	i
0.532	0.335	-0.136	-0.760	0.052	-0.075	0.022	0.029	0.001

Le premier axe principal de `pca2` est défini par :

A	B	C	D	E	F	G	H	I	J	K
0.379	-0.816	0.305	0.000	0.023	-0.131	0.228	0.011	0.067	-0.026	0.065
L	M	N	O	P						
-0.135	0.020	0.001	0.009	-0.001						

Pour chaque vecteur donner la somme des composantes et la somme des carrés des composantes.

18. _____

Représenter graphiquement ces scores numériques des candidats et interpréter. On pourra commenter en même temps, le résultat numérique suivant :

```
((pca1$eig/sum(pca1$eig))[1:7],dig=3)
[1] 0.396 0.296 0.224 0.067 0.012 0.004 0.001
round((pca2$eig/sum(pca2$eig))[1:7],dig=3)
[1] 0.498 0.189 0.122 0.069 0.053 0.024 0.019
```

19. _____

On donne dans la figure 5 les cartes des variables des deux ACP (plans 1-2 des deux premières composantes principales). On retrouve l'information liée au facteur 1. Commenter la position relative des points sur la seconde dimension en même temps que la valeur :

```
cor(pca1$li[,2],pca2$li[,2])
[1] -0.7812
```

20. _____

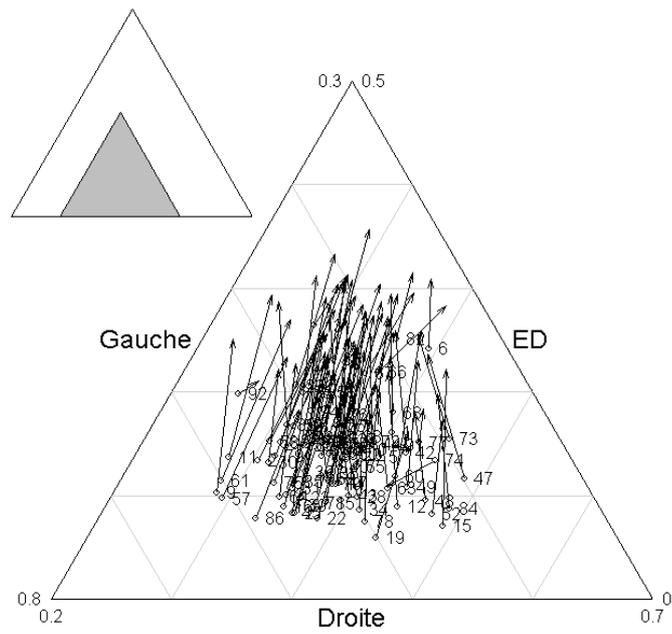
On utilise :

```
x1=tab1$a+tab1$e+tab1$f+tab1$g+tab1$h+tab1$i
y1=tab1$b+tab1$c
z1=tab1$d
tril=(cbind.data.frame(x1,y1,z1))
names(tril)=c("Gauche","Droite","ED")
x2=tab2$c+tab2$e+tab2$f+tab2$h+tab2$i+tab2$k+tab2$m+tab2$p
y2=tab2$a+tab2$d+tab2$j+tab2$n+tab2$o
```

```

z2=tab2$G+tab2$L+tab2$B
triangle.biplot(tri1,tri2)

```



Donner une légende.

Annexes

	a	b	c	d	e	f	g	h	i	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
D1	314	191	212	161	44	41	18	15	4	172	219	127	77	48	61	40	51	39	52	21	35	19	22	12	5
D2	395	172	133	134	90	33	14	25	4	188	212	154	53	82	47	41	36	48	32	39	26	12	14	9	7
D3	303	195	150	101	181	30	21	15	4	209	141	160	64	63	52	44	39	47	34	81	18	15	15	13	5
D4	313	189	162	172	64	47	32	18	3	164	166	136	59	55	59	90	56	51	37	44	28	21	18	11	5
D5	291	206	198	137	65	53	28	18	4	173	144	132	69	49	59	73	67	55	52	34	28	24	23	13	5
D6	244	243	150	242	62	30	15	12	2	220	260	122	68	38	48	24	48	26	48	30	25	15	18	7	3
D7	306	200	190	129	80	41	29	20	5	173	168	142	68	53	56	77	52	51	33	42	26	19	19	16	5
D8	372	173	142	151	82	36	17	23	4	179	229	162	56	73	42	39	39	44	29	34	36	11	13	9	5
D9	414	182	104	103	105	35	32	21	4	143	151	237	41	55	47	84	53	56	20	52	16	19	13	9	4
D10	337	205	178	144	65	35	13	19	4	209	217	137	62	56	47	50	37	36	41	32	27	15	18	11	5
D11	401	176	103	137	103	31	26	18	5	150	198	214	43	54	50	70	40	45	24	48	25	15	12	8	4
D12	336	258	185	89	45	39	23	22	3	222	121	163	71	49	50	82	48	49	42	29	16	19	19	16	4
D13	270	148	139	264	112	29	23	13	2	169	224	140	58	49	55	39	51	33	39	48	50	17	16	8	4
D14	374	199	182	111	45	41	20	23	5	195	142	158	70	74	48	69	53	48	39	23	20	21	22	12	6
D15	318	375	115	71	57	24	18	19	3	309	111	144	55	50	44	93	33	43	28	28	18	14	16	9	5
D16	414	196	153	89	70	32	20	22	4	196	137	177	58	69	44	74	50	54	33	38	19	18	18	10	5
D17	365	202	181	112	60	37	20	21	2	203	134	161	62	60	45	92	52	46	40	32	19	22	18	9	5
D18	333	187	166	116	118	32	21	23	4	207	158	147	59	69	49	55	40	43	35	65	22	18	16	11	6
D19	289	392	48	59	137	24	32	16	3	342	88	164	33	56	39	62	32	44	20	69	12	14	14	7	4

	a	b	c	d	e	f	g	h	i	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
D21	358	221	160	139	39	41	18	19	5	190	180	153	76	58	69	33	50	44	41	22	26	19	21	12	6
D22	382	189	172	83	76	41	28	25	4	203	117	193	64	67	44	39	61	57	39	48	11	20	20	12	5
D23	348	282	100	78	112	25	28	23	4	262	112	182	45	60	47	70	34	50	26	55	16	15	13	8	5
D24	349	238	121	98	113	32	28	18	3	216	122	174	52	57	45	78	46	52	28	62	21	17	16	9	5
D25	350	215	156	144	35	49	22	26	3	187	190	146	63	55	91	29	52	46	37	18	31	17	19	14	5
D26	316	183	171	167	64	46	30	19	4	159	208	139	69	49	60	51	58	45	38	31	29	21	20	19	4
D27	365	197	164	140	57	34	16	22	5	193	196	139	65	69	47	52	44	47	37	29	29	19	19	9	6
D28	357	193	176	154	45	34	16	20	5	202	191	144	68	59	49	53	43	39	43	23	27	20	19	14	6
D29	357	209	196	99	43	43	28	22	3	220	108	185	75	64	48	35	66	59	41	29	10	23	20	12	5
D30	291	154	146	206	121	33	30	18	1	151	249	138	61	51	52	53	43	39	33	53	31	17	15	9	5
D31	404	168	147	132	59	37	30	20	3	152	166	212	64	55	58	38	69	47	32	31	18	24	20	10	4
D32	401	203	144	107	60	38	24	20	3	179	133	199	64	49	44	108	46	46	27	35	17	22	16	11	4
D33	389	194	156	123	63	30	22	20	3	188	142	181	67	59	48	79	62	45	30	33	15	20	18	9	4
D34	311	173	137	199	90	33	35	17	5	150	230	161	56	50	53	58	55	40	34	41	24	20	15	9	4
D35	376	209	204	86	28	44	21	26	6	213	105	178	79	67	46	34	71	54	51	21	11	25	24	16	5
D36	376	199	147	113	84	29	21	26	5	212	153	165	53	66	43	75	37	47	32	44	21	16	17	13	6
D37	375	183	188	122	50	35	20	22	5	197	145	164	70	66	56	47	53	45	47	28	20	21	21	14	6
D38	339	166	169	161	68	46	28	18	5	153	183	160	68	58	68	27	70	45	44	36	27	21	23	12	5
D39	334	188	174	145	56	53	20	25	5	173	183	134	71	58	73	40	55	53	42	29	31	17	20	15	6
D40	421	210	150	90	69	23	20	15	2	192	105	219	70	48	39	121	42	43	25	45	12	14	13	9	3

	a	b	c	d	e	f	g	h	i	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
D41	355	194	182	128	64	31	19	22	5	191	184	146	75	61	48	66	43	42	37	33	24	17	18	10	5
D42	304	179	185	174	70	41	23	20	4	169	217	134	80	56	61	32	50	44	41	34	29	16	18	14	5
D43	300	210	214	141	45	40	23	23	4	204	184	128	85	58	54	49	46	52	35	24	26	17	16	16	6
D44	368	198	199	100	40	43	24	24	4	188	115	180	79	63	51	45	70	55	42	25	15	22	25	19	6
D45	318	210	183	149	60	37	18	20	5	197	194	141	71	52	53	49	47	36	42	30	30	20	21	12	5
D46	388	238	121	83	71	44	30	21	4	191	108	176	50	57	55	101	54	54	29	46	15	30	19	10	5
D47	336	185	160	154	85	35	23	18	4	175	189	157	67	48	41	93	45	43	30	41	26	17	14	10	4
D48	268	269	216	116	49	34	26	17	5	247	136	128	75	46	48	83	44	47	43	33	17	19	16	14	4
D49	326	221	243	95	27	41	16	26	5	221	117	150	100	60	44	45	62	47	46	19	18	21	23	22	5
D50	338	233	209	108	28	43	15	22	4	254	133	137	66	66	48	79	40	48	39	19	18	17	19	12	5
D51	346	205	172	140	58	40	14	20	5	206	193	138	77	69	47	35	47	39	43	27	29	15	19	11	5
D52	352	196	156	156	54	42	14	24	6	199	224	131	57	62	45	51	43	48	38	25	29	15	17	10	6
D53	336	241	234	82	22	41	16	23	5	258	119	144	84	55	39	38	57	53	49	16	16	18	22	26	6
D54	351	156	180	148	68	42	24	26	5	184	181	162	64	76	57	21	54	48	35	37	27	19	19	11	5
D55	351	181	184	150	45	45	14	26	4	197	203	143	73	66	45	45	45	45	36	23	28	15	19	12	5
D56	349	196	194	128	43	38	28	21	3	226	149	167	66	57	42	39	58	51	45	27	11	20	23	14	5
D57	333	163	167	199	41	48	14	30	5	190	237	145	65	71	46	13	52	43	33	21	34	15	17	12	6
D58	445	170	124	97	96	28	18	18	4	173	162	201	52	65	53	50	39	49	31	59	21	15	14	10	6
D59	367	150	156	152	103	29	15	23	5	177	194	168	66	72	46	36	48	42	34	49	26	13	14	10	5
D60	365	177	141	167	73	33	16	24	4	183	228	137	59	69	52	32	44	41	36	33	32	19	18	11	6

	a	b	c	d	e	f	g	h	i	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
D61	341	234	193	118	30	40	17	23	4	234	182	132	70	64	37	56	44	44	40	20	21	17	20	14	5
D62	413	145	139	114	116	27	14	27	5	166	184	175	49	83	53	64	38	47	29	54	22	11	11	9	5
D63	342	196	175	116	71	40	32	23	5	190	136	170	71	77	63	38	51	58	34	37	20	20	18	11	6
D64	338	245	181	107	50	34	23	19	3	187	104	178	131	52	38	93	57	50	24	30	10	18	15	9	4
D65	379	184	156	99	98	32	31	18	3	163	123	200	77	56	44	90	47	54	23	55	14	23	17	10	4
D66	314	178	134	205	94	32	24	16	3	170	209	166	61	53	52	58	44	43	32	48	24	16	13	7	4
D67	282	176	186	219	13	94	9	18	3	185	234	112	120	47	45	10	60	34	46	9	39	18	21	15	5
D68	295	171	181	218	15	93	8	17	2	181	235	108	91	52	63	10	60	35	41	11	50	17	23	18	5
D69	293	174	220	180	54	38	22	15	4	181	193	140	88	45	67	16	61	36	56	25	25	24	24	15	4
D70	374	201	155	138	46	41	16	23	6	180	223	134	53	55	87	47	39	47	34	22	34	15	15	10	5
D71	376	206	172	111	65	33	16	19	2	192	178	169	64	58	52	41	42	48	47	34	27	16	15	12	5
D72	381	193	181	93	62	36	22	27	5	213	150	166	62	74	47	37	53	51	36	33	21	18	20	13	6
D73	304	209	181	152	59	50	22	19	4	180	198	131	71	49	63	30	63	44	49	31	28	20	25	12	6
D74	272	225	228	155	32	54	16	15	3	186	208	114	83	45	57	23	72	38	62	17	24	21	32	14	4
D75	295	316	136	134	36	36	28	16	3	240	94	200	79	30	66	5	74	34	65	22	10	38	25	15	3
D76	390	168	161	112	87	34	21	24	3	191	162	168	60	79	47	37	49	51	32	46	27	19	17	9	6
D77	330	199	151	178	63	38	19	18	4	197	189	149	65	56	64	22	54	39	40	30	31	29	21	9	5
D78	294	245	185	151	47	38	20	16	4	226	147	150	88	42	65	16	58	31	54	23	20	26	27	23	4
D79	375	212	211	75	32	45	18	27	5	224	93	181	80	69	41	67	59	54	36	20	13	21	19	17	6
D80	367	169	148	138	101	29	18	26	4	176	163	139	64	80	42	121	37	44	26	45	22	12	16	8	5

	a	b	c	d	e	f	g	h	i	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
D81	365	191	147	148	62	39	23	21	4	181	168	187	60	57	48	66	49	46	29	35	24	18	16	12	4
D82	364	204	137	152	59	39	21	20	4	177	201	156	58	52	45	84	46	42	29	29	27	24	15	11	4
D83	255	199	162	251	71	28	19	12	3	213	235	121	61	45	53	48	44	30	42	30	32	15	17	10	4
D84	294	167	152	231	78	37	20	16	5	167	258	126	57	45	49	60	47	37	36	31	39	18	17	9	4
D85	323	246	241	87	27	39	14	22	1	249	118	148	83	57	39	64	51	44	46	18	16	18	23	21	5
D86	379	218	164	93	58	38	23	23	4	215	111	173	66	71	45	73	57	52	32	31	15	21	19	13	6
D87	379	221	110	78	114	31	40	23	4	218	112	197	46	75	53	52	44	55	26	56	18	18	16	9	5
D88	354	192	174	149	39	48	14	25	5	192	209	142	64	68	55	32	45	51	39	21	28	16	19	14	5
D89	331	200	172	157	63	36	17	18	6	200	210	136	64	56	52	42	42	42	39	31	35	17	18	10	6
D90	368	173	139	168	45	53	21	29	4	142	225	99	51	55	194	22	43	40	24	18	40	14	16	12	5
D91	330	193	158	150	69	43	34	19	4	191	160	168	71	54	73	16	62	38	39	34	21	34	24	10	5
D92	291	247	165	148	69	37	24	16	3	239	119	167	89	38	65	8	58	30	59	34	17	34	25	15	3
D93	329	146	111	198	135	33	25	18	5	161	177	179	52	60	64	8	61	38	35	63	25	52	14	6	5
D94	315	193	140	156	110	38	27	16	5	194	143	172	72	47	69	9	63	36	49	54	19	41	20	8	4
D95	332	180	145	181	79	38	22	18	5	188	181	163	68	52	66	14	56	35	44	40	22	38	20	8	5

1988

b	c	d
J. Chirac	R. Barre	J.M. Le Pen
e	f	g
A. Lajoinie	A. Waechter	P. Juquin
h	i	
A. Laguillier	P. Bousset	

2002

B	C	D
J.M. Le Pen	L. Jospin	F. Bayrou
F	G	H
P. Chevènement	J. Saint Josse	N. Mamère
J	K	L
A. Madelin	R. Hue	B. Mégret
M	O	P
C. Taubira	C. Boutin	D. Gluksten

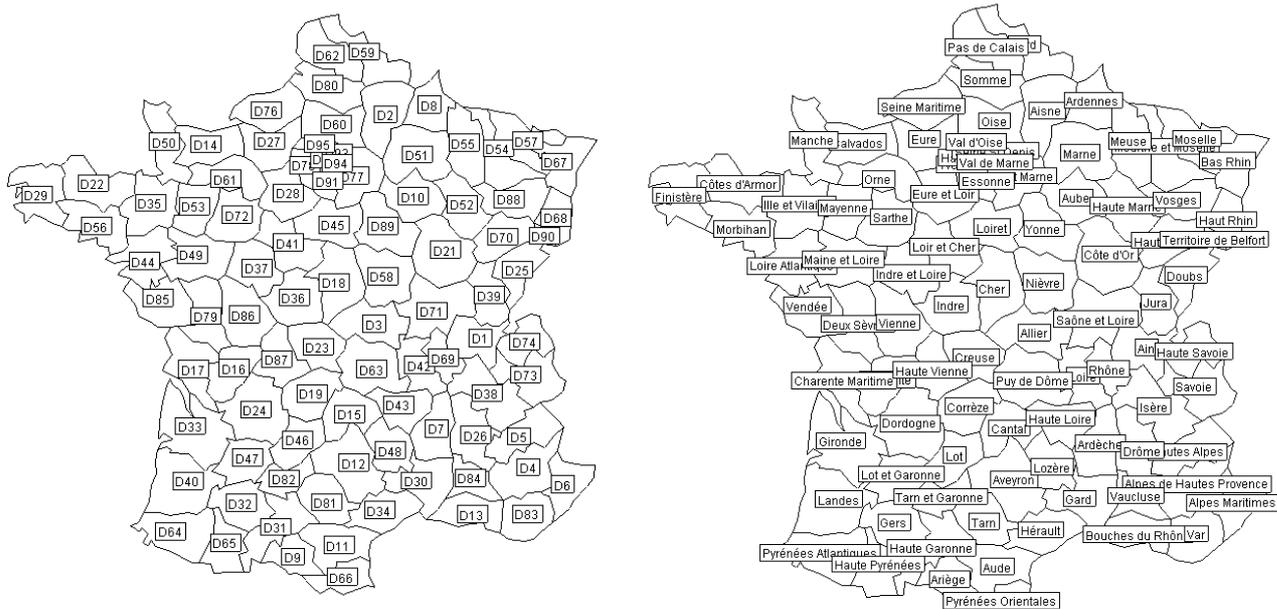


Figure 1 : 94 départements présents dans l'analyse.

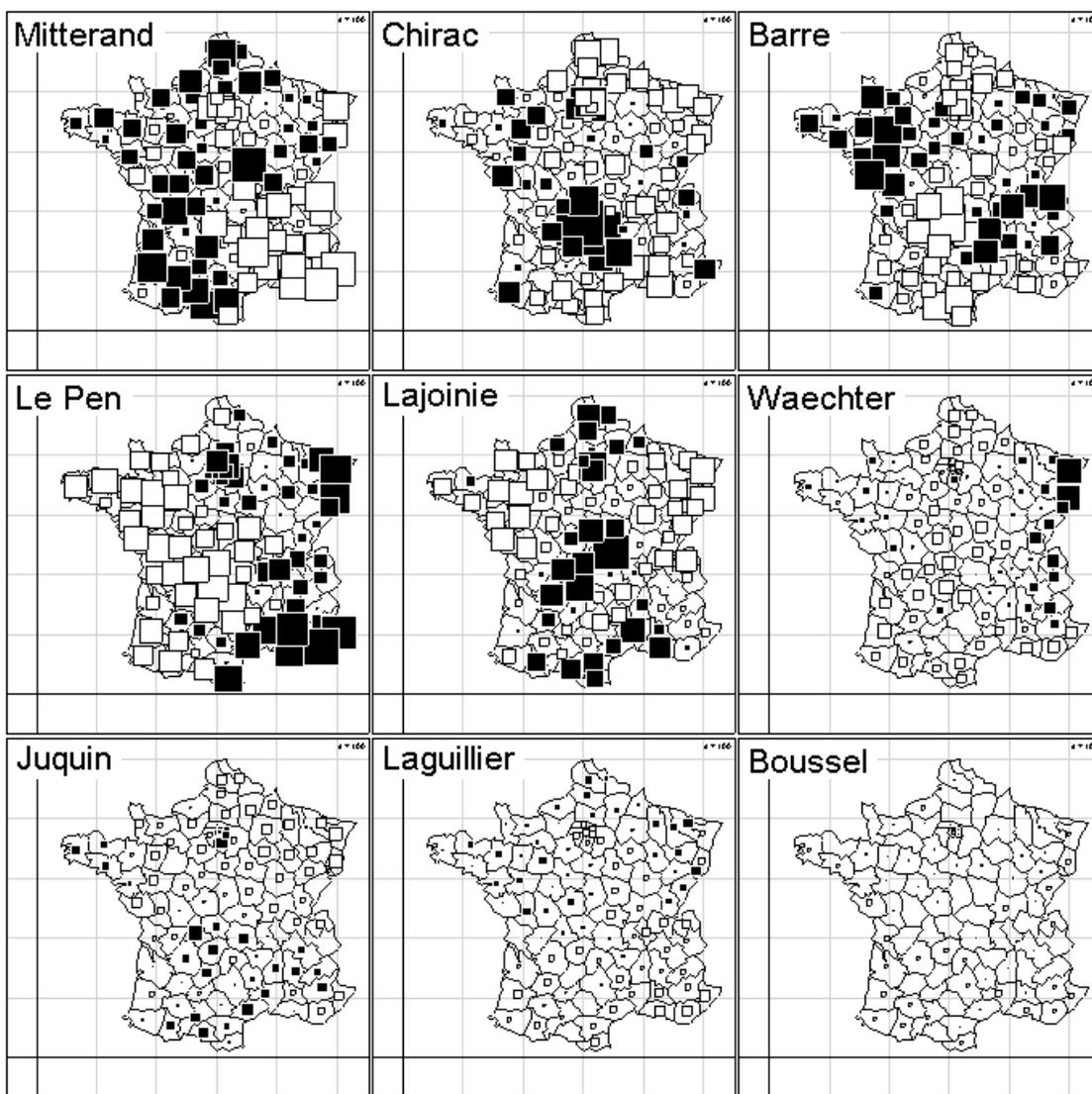


Figure 2 : Cartographie des résultats de l'élection de 1988 après centrage.

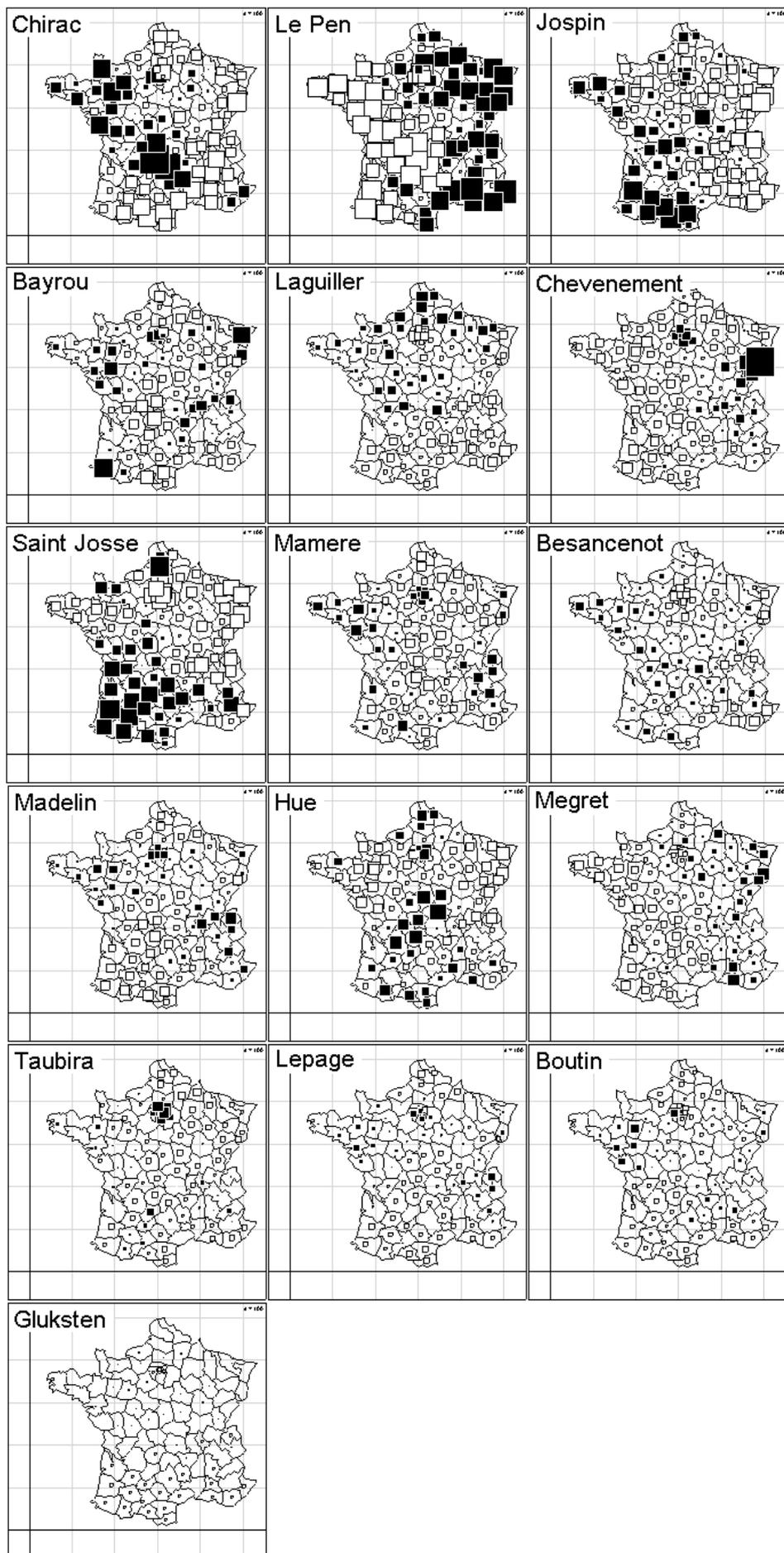
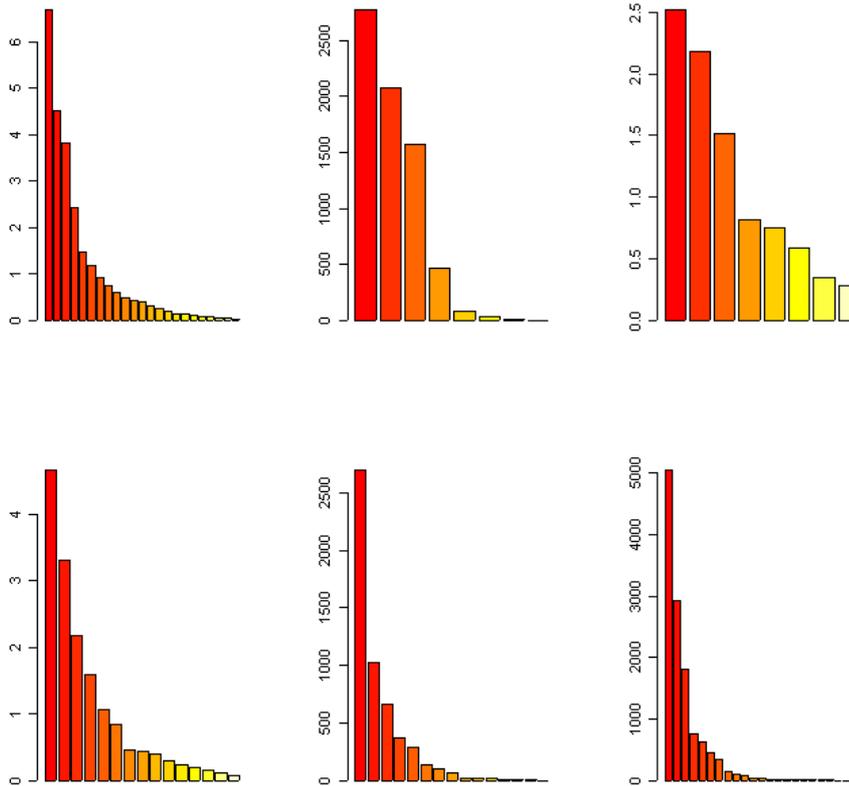


Figure 3 : Cartographie des résultats de l'élection de 2002 après centrage.



1	2	3
4	5	6

Figure 4. Graphes des valeurs propres de 6 ACP (Analyses en Composantes Principales).

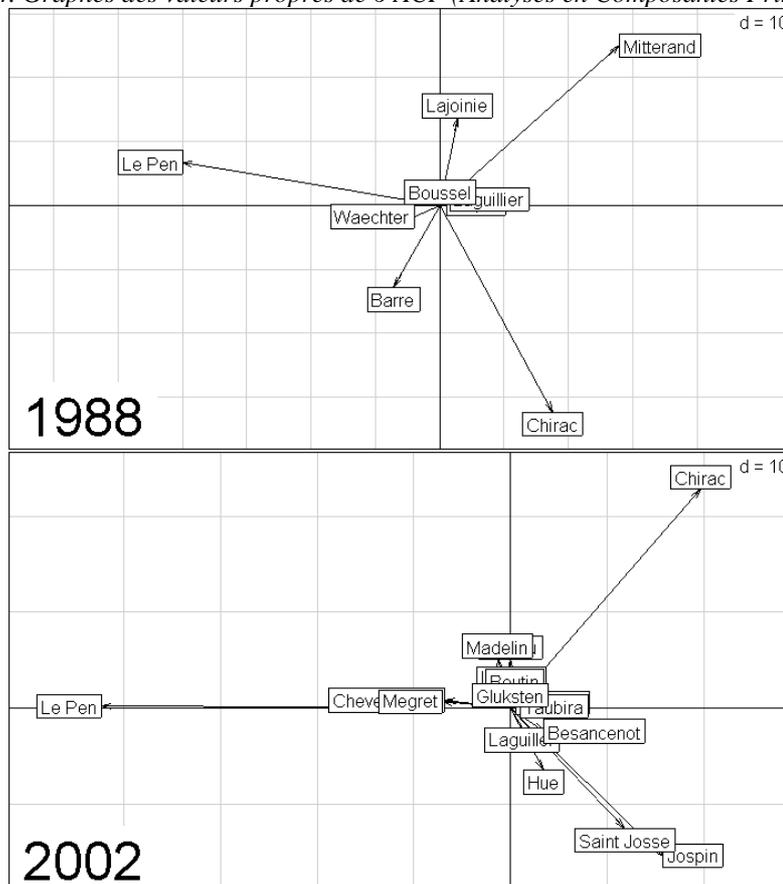


Figure 5 : Cartes factorielles des variables de deux ACP centrées.

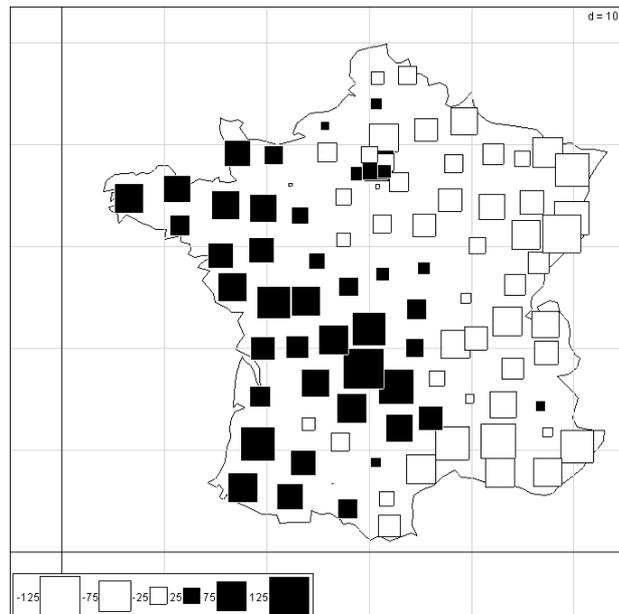
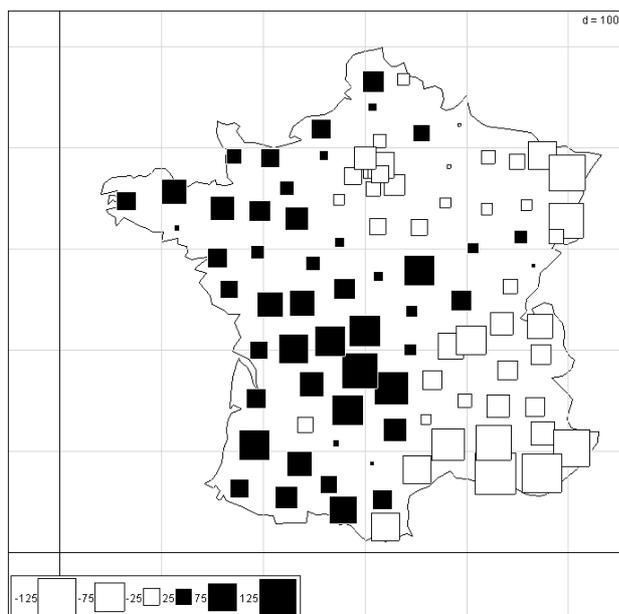


Figure 6 : Cartographie des coordonnées de deux ACP centrées.

1. Par définition : $\frac{1}{4} \times (1^2 + 0^2) + \frac{1}{4} \times (1^2 + 1^2) + \frac{1}{4} \times (1^2 + 1^2) + \frac{1}{4} \times (0^2 + 1^2) = \frac{3}{2}$

2. Les poids des lignes valent 1/4. Les moyennes valent 3/4. Les variances valent $\frac{\sqrt{3}}{4}$. La matrice de corrélation est :

$$\mathbf{R} = \begin{bmatrix} 1 & -\frac{1}{3} \\ -\frac{1}{3} & 1 \end{bmatrix} = \begin{bmatrix} -\frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} 4/3 & 0 \\ 0 & 2/3 \end{bmatrix} \begin{bmatrix} -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix}$$

Les axes principaux sont les deux vecteurs propres.

3. En suivant les définitions :

$$\mathbf{D}_I = \begin{bmatrix} 1/6 & 0 & 0 & 0 \\ 0 & 1/3 & 0 & 0 \\ 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 1/6 \end{bmatrix} \quad \mathbf{D}_J = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix} \quad \mathbf{P}_0 = \begin{bmatrix} 1 & -1 \\ 0 & 0 \\ 0 & 0 \\ -1 & 1 \end{bmatrix}$$

4. Il faut diagonaliser $\mathbf{A}^T \mathbf{A} = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}$ qui a les valeurs propres 5 et 1. Donc $\mathbf{A} \mathbf{A}^T$ a les valeurs propres 5, 1 et 0 (valeur propre associée à un sous-espace propre de dimension 2).

5. Par définition, les racines des valeurs propres de la précédente, soit $\sqrt{5}$ et 1.

6. La somme vaut 1000. En effet, comme $\sum_{j=1}^p x_{ij} = 1000$:

$$\sum_{j=1}^p \bar{x}_j = \frac{1}{n} \sum_{j=1}^p \sum_{i=1}^n x_{ij} = \frac{1}{n} \sum_{i=1}^n 1000 = 1000 \quad \square$$

7. Examiner les cartes pour retrouver la position ou examiner le tableau.

Mitterrand dans les Alpes Maritimes, en 1988 : -99.

Le Pen dans la Corrèze en 2002 : -81.5

Chirac dans la Corrèze en 1988 : +189.3

Chevènement dans le Territoire de Belfort en 2002 : +140.2

8. La figure donne la carte des écarts au résultat moyen de chaque candidat dans chaque département. En général, ces cartes sont structurées et donne des indications sur les tendances régionales en faveur ou en défaveur des candidats. On peut comparer les cartes d'une même élection, les cartes des candidats homologues des deux élections, toutes les cartes entre elles. Des faits marquants : les bastions (la Nièvre de F. Mitterrand, la Corrèze de J. Chirac, le territoire de Belfort de J.P. Chevènement, l'apparition de l'écologisme en Alsace), les composantes régionales fortes (le Sud-Ouest des chasseurs pour Saint-Josse, le déficit des socialistes dans l'Est et le Sud-Est, les régions traditionnelles du communisme des mines Nord-Centre-Sud, l'installation de l'extrême droite sur le pourtour méditerranéen, en Alsace, autour de Paris et Lyon) et la grande fracture de l'est pour l'extrême droite contre l'ouest pour la gauche.

9. Les valeurs propres comprises entre 0 et 25 sont celles d'ACP normées (inertie totale = trace = nombre de variables). Le nombre de valeurs propres ne peut dépasser le nombre de variables, respectivement 9 pour tab1, 16 pour tab2 et 25 pour tab3. Il n'y a alors qu'une possibilité, dans l'ordre **F A B D C E**.

10. Les rangs ne peuvent dépasser le nombre de colonnes (respectivement 9 et 16). Mais pour **A**, la somme par lignes est nulle (1000-1000, voire question 1). On a $\mathbf{A}\mathbf{1}_9 = \mathbf{0}_{94}$ et le rang est majoré par

8. Pour **B**, la somme par lignes est nulle (même argument). On a $\mathbf{B}\mathbf{1}_{16} = \mathbf{0}_{94}$ et le rang est majoré par **15**.

11. Par juxtaposition $\mathbf{C} = [\mathbf{A}|\mathbf{B}]$. Si **u** est le vecteur à 25 composantes définies, dans l'ordre, par 1 (9 fois) et 0 (16 fois) et si **v** est le vecteur à 25 composantes définies, dans l'ordre, par 0 (9 fois) et 1 (16 fois), **u** et **v** sont indépendants (orthogonaux) et vérifient $\mathbf{C}\mathbf{u} = \mathbf{C}\mathbf{v} = \mathbf{0}_{94}$. Le rang est au plus égal à **23**.

12. On sait que le rang d'un tableau et le rang de la matrice de covariances associée sont les mêmes. On sait que les carrés des valeurs singulières sont proportionnelles aux valeurs propres, car :

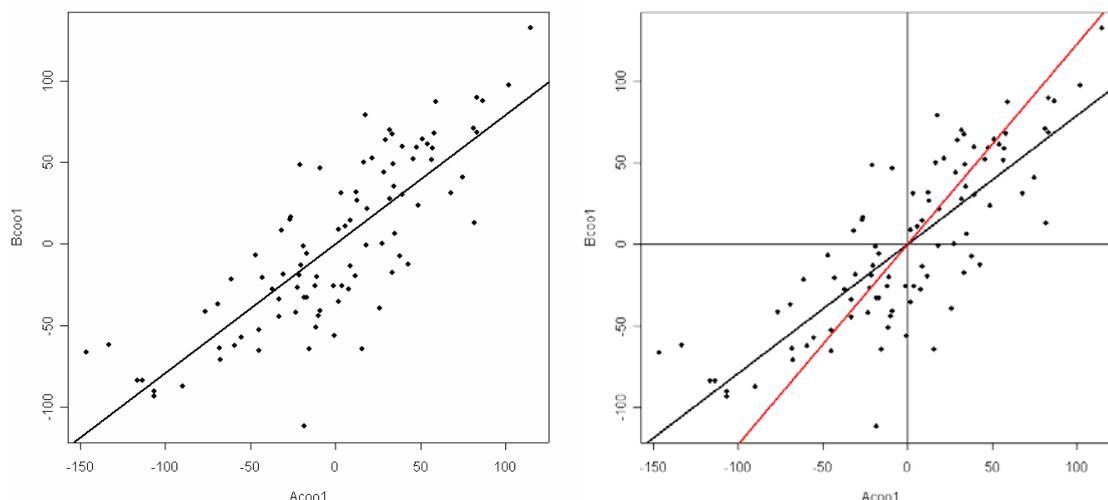
$$\mathbf{X} = \mathbf{UDV}^T \Rightarrow \mathbf{C} = \frac{1}{n} \mathbf{X}^T \mathbf{X} = \frac{1}{n} \mathbf{VDU}^T \mathbf{UDV}^T = \frac{1}{n} \mathbf{VD}^2 \mathbf{V}^T = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^T$$

Donc **C** a 23 valeurs propres non nulles et est de rang **23**. Alors **B** est de rang **15** et **A** est de rang **8** (sinon le rang de **C** serait strictement inférieur à 23).

13. La corrélation est manifestement positive et on peut éliminer les valeurs négatives. Elle est bornée par 1 (c'est un cosinus), elle est élevée (au moins 0.5) mais pas caricaturale (moins de 0.95). Il reste la valeur **0.8032**.

Pour refaire la figure de base :

```
plot(Acoo1,Bcoo1,pch=20,cex=1.5)
abline(h=0,v=0)
abline(lm(Bcoo1~Acoo1),lwd=2)
```



14. L'équation de la droite de Y/X est $y = ax$ (les coordonnées sont des variables centrées) avec :

$$a = \frac{\text{cov}(x, y)}{\text{var}(x)} = \frac{\text{corr}(x, y)\sqrt{\text{var}(y)}}{\sqrt{\text{var}(x)}} = \frac{0.8032\sqrt{2698}}{\sqrt{2773}} = 0.7923$$

car les variances des coordonnées sont les valeurs propres.

15. L'équation de la droite de Y/X est $x = ay$ donc $y = \frac{1}{a}x$ avec :

$$a = \frac{\text{var}(y)}{\text{cov}(x, y)} = \frac{\sqrt{\text{var}(y)}}{\text{corr}(x, y)\sqrt{\text{var}(x)}} = \frac{\sqrt{2698}}{0.8032\sqrt{2773}} = 1.228$$

16. Les deux variables sont centrées et l'axe principal est porté par le premier vecteur propre de la matrice de covariances. La covariance est :

$$\text{cov}(x, y) = \text{corr}(x, y) * \sqrt{\text{var}(x)\text{var}(y)} = 0.8032 * \sqrt{2773} * \sqrt{2698} = 2197$$

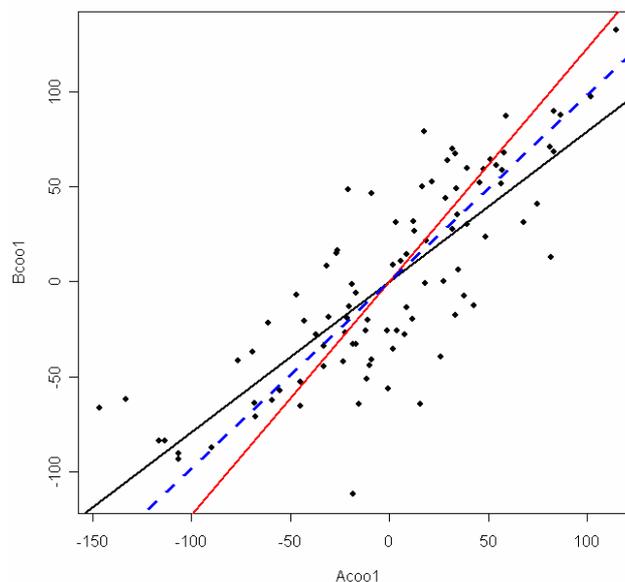
La matrice de covariances est :

$$C = \begin{bmatrix} \text{var}(x) & \text{cov}(x, y) \\ \text{cov}(x, y) & \text{var}(y) \end{bmatrix} = \begin{bmatrix} 2773 & 2197 \\ 2197 & 2698 \end{bmatrix}$$

La première valeur propre de cette matrice est :

$$\begin{aligned} \lambda &= \frac{1}{2} \left(\text{var}(x) + \text{var}(y) + \sqrt{4\text{cov}^2(x, y) + (\text{var}(x) - \text{var}(y))^2} \right) \\ &= \frac{1}{2} \left(2773 + 2698 + \sqrt{4 * 2197^2 + 75^2} \right) = 4933 \end{aligned}$$

Le premier vecteur propre vérifie $2773x + 2197y = 4932 * x$ et la pente vaut $\frac{4932 - 2773}{2197} = 0.983$.



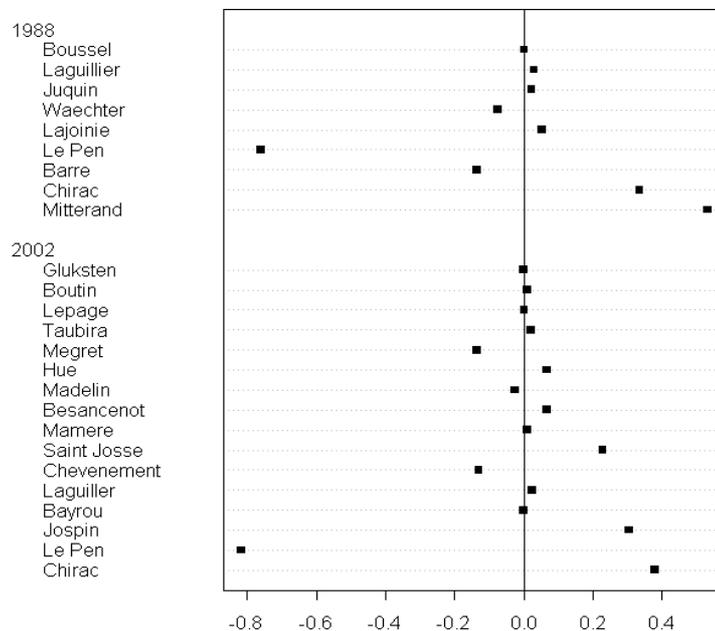
Pour refaire la figure

```
abline(c(0,1.228),lwd=2,col="red")
abline(c(0,0.983),lwd=3,col="blue",lty=2)
```

17. La somme des composantes est nulle. En effet, d'après la question 4, $\mathbf{1}_n$ est vecteur propre pour 0, donc les axes principaux lui sont orthogonaux, donc la somme des composantes d'un axe principal est nulle. La somme des carrés des composantes est égale à 1, car il s'agit de vecteurs unitaires.

```
sum(pca1$c1[,1])
[1] -3.563e-16
sum(pca1$c1[,1]^2)
[1] 1
sum(pca2$c1[,1])
[1] -2.516e-16
sum(pca2$c1[,1]^2)
[1] 1
```

18. Les deux scores des départements sont très liés (question 8). Les coefficients des variables sont donc associés. On peut les représenter par un « dotchart » :



La coordonnée exprime l'opposition Le Pen contre Chirac-Mitterand dans le premier cas, Le Pen contre Jospin-Chirac dans le second. C'est la même carte et la même signification politique. On a augmenté le nombre de candidats mais cette opposition qui prend en compte 40% du résultat en 1988 représente 50% en 2002. Saint Josse est le seul des candidats de seconde importance à prendre une place marquée dans ce schéma. Ceci est lié à la régionalisation commune dans le Sud-Ouest du candidat des chasseurs et du candidat socialiste. La position spatiale commune induit une covariance qui n'est pas forcément le résultat d'une alliance. La carte de l'opinion d'extrême droite se renforce entre les deux élections.

19. On observe une inversion des positions relatives Mitterand-Chirac et Chirac-Jospin. Mathématiquement, ceci n'a **aucune signification** car la position dans un sens ou dans l'autre d'un axe principal ou d'une composante principale est aléatoire. La corrélation négative entre les deux coordonnées de rang 2 a donc exactement la même signification que la corrélation positive entre les deux premières coordonnées. Noter l'association renforcée Saint-Josse-Jospin (Sud-Ouest) qui recouvre une dichotomie gauche-droite classique (particulièrement simple en 1988 sous la forme Barre-Chirac contre Mitterand-Lajoinie).

20. Représentation triangulaire de deux tableaux de résultats électoraux après simplification à trois composantes. Le premier contient par département les scores de « Mitterrand+Lajoinie+Waechter+Juquin+ Laguillier+ Boussel » appelé Gauche, les scores de « Chirac+Barre » appelé Droite et les scores de Le Pen appelé ED. Le second contient les scores de « Jospin+Laguillier+Chevènement +Mamère+Besancenot+Hue+Taubira+ Gluksten » appelé Gauche, les scores de « Chirac+Bayrou+Madelin+Lepage+ Boutin » appelé Droite et les scores de « Le Pen+Saint Josse+Mégret » appelé ED. On peut discuter de ce regroupement qui reprend les associations les plus courantes. Chaque département est alors représenté par un point (résultat de 1988) et un vecteur dont l'extrémité est le résultat de 2002. Globalement l'évolution est générale et sans ambiguïté.

Ou bien, si on se veut un peu plus technique :

Représentation triangulaire double de deux tableaux 94 lignes et 3 colonnes. Un point de coordonnées (x, y, z) est représenté par son profil $(x/s, y/s, z/s)$ où $s=x+y+z$. Seule une partie de la représentation définie par les trois vecteurs de la base canonique $(1,0,0)$, $(0,1,0)$ et $(0,0,1)$ est utilisée. C'est le plus petit triangle équilatéral de la grille du triangle (unité 0.1) qui contient tous les points (en haut et à gauche, le triangle gris). La variable x varie entre 30 et 80%, la variable y entre 20 et 70% et la variable z entre 0 et 50%. Le point 57 (Nièvre) a un profil voisin de $(0.6, 0.3, 0.1)$. Si on retourne aux données, on a exactement (la ligne 57 correspond à D58) :

D58 445 170 124 97 96 28 18 18 4

donc 1000-294-97 294 97 donc 609 294 97 donc $(0.609, 0.294, 0.097)$. La Nièvre est le fief de F. Mitterrand élu en 1988. L'extrémité du vecteur représente de la même manière le résultat de 2002. Par exemple pour D58

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
173	162	201	52	65	53	50	39	49	31	59	21	15	14	10	6

donc 487 280 233 ou $(0.487, 0.280, 0.233)$. L'extrême droite a gagné plus de 13% des voix exprimées essentiellement au détriment de la gauche. Il semble que ce soit le cas général.