

Analyse de Données

CTD 1 : Analyse en Composantes Principales

Méthode d'analyse multidimensionnelle

- Tableau individus/variables

	Archi	Prog. Impérative	Proba	Stats	EDP	Anglais	Sport
Bastien	10	12	18	17	17	15	14
Damien	19	6	17	11	4	7	4
Thibault	9	16	17	14	15	14	17
Sonia	18	18	8	10	8	12	13

Méthode d'analyse multidimensionnelle

- Représenter 7 notes pour 4 élèves

	Archi	Prog. Impérative	Proba	Stats	EDP	Anglais	Sport
Bastien	10	12	18	17	17	15	14
Damien	19	6	17	11	4	7	4
Thibault	9	16	17	14	15	14	17
Sonia	18	18	8	10	8	12	13

Visualisation des données

- Représenter 7 notes pour 4 élèves

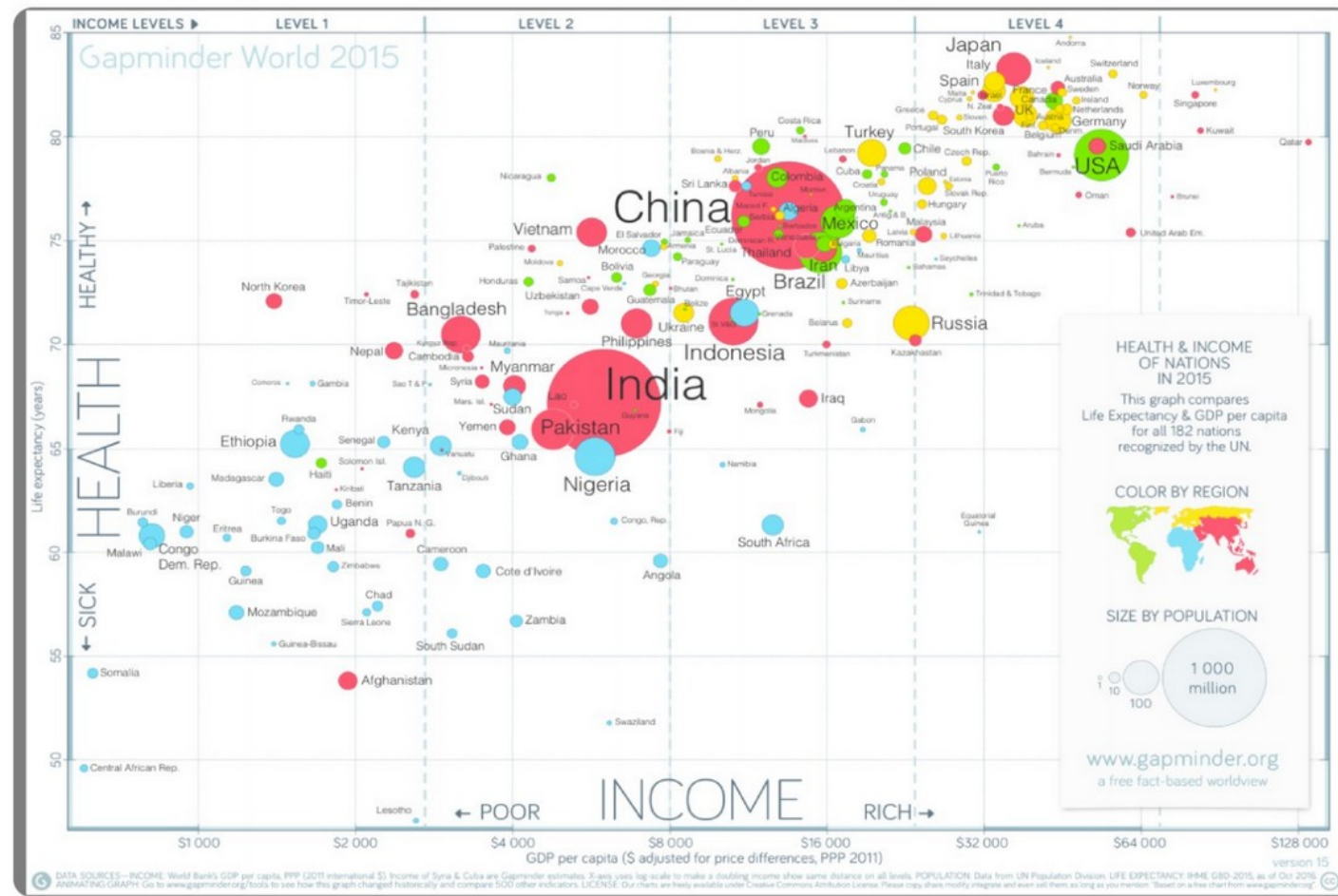


Visualisation des données

- 5D : revenu, espérance de vie, nombre d'habitants, continent, année

Visualisation des données

- 5D : revenu, espérance de vie, nombre d'habitants, continent, année : gapminder.org



Visualisation des données

- Enquêtes : très consommatrices d'ACP

Pensez-vous que le QR code est facile à utiliser ?	Tout à fait d'accord	Plutôt d'accord	Modérément en accord	Neutre	Modérément en désaccord	Plutôt en désaccord	Tout à fait en désaccord
Apprendre à utiliser le QR code serait facile pour moi							
De mon point de vue il est facile d'obtenir le QR code pour faire ce dont j'ai besoin							
Mon interaction avec le QR code est claire et compréhensible							
Il serait facile pour moi de devenir habile afin d'utiliser le QR code							
Je trouve le QR code facile à utiliser							
Je trouve le QR code flexible pour interagir avec							

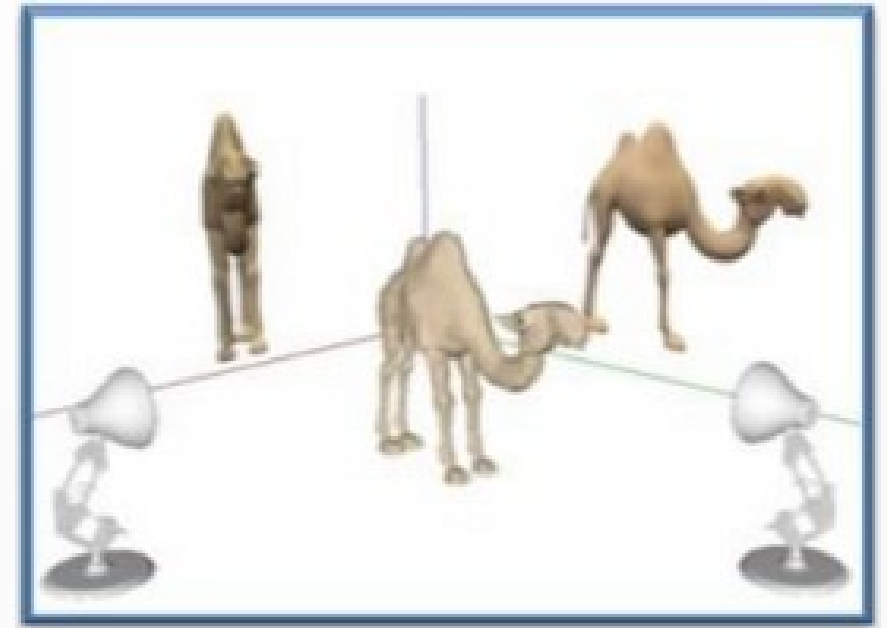
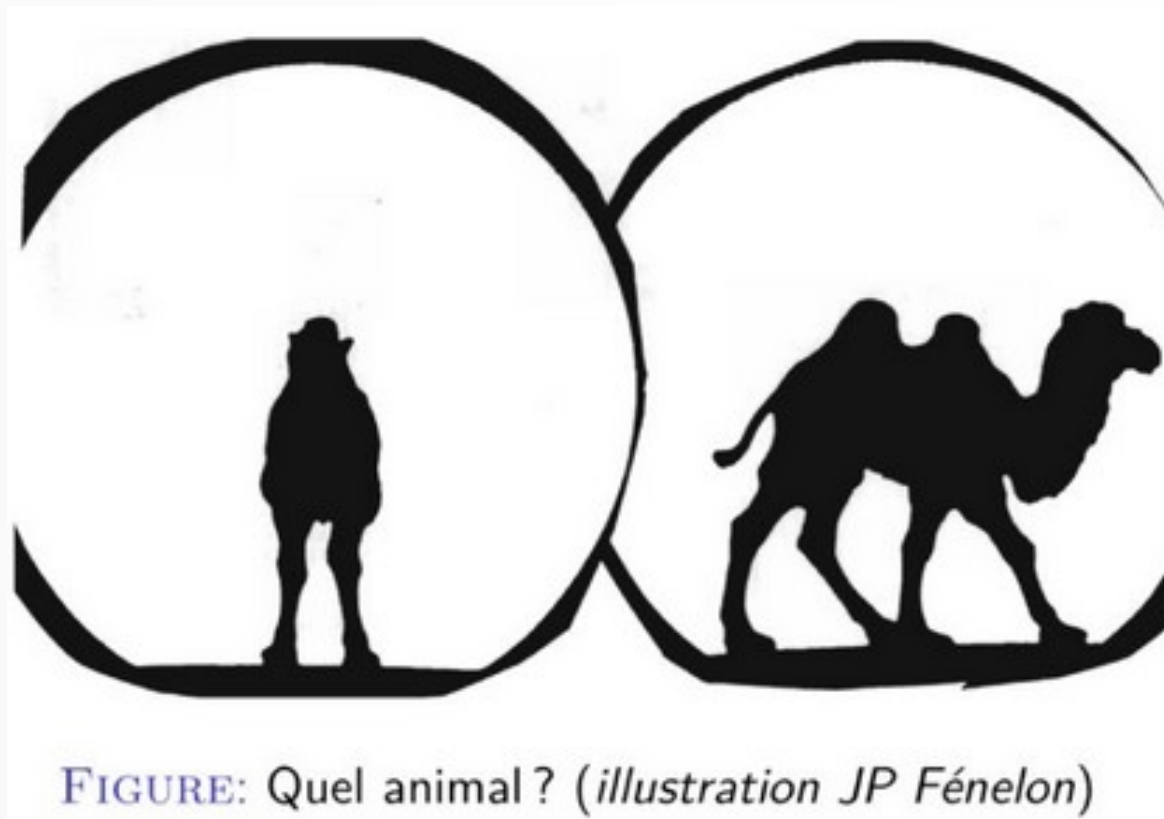
Visualisation des données

- Exemple de données en 12D

	Janv	Févr	Mars	Avri	Mai	Juin	juil	Août	Sept	Octo	Nove	Déce	Lati	Long
Bordeaux	5.6	6.6	10.3	12.8	15.8	19.3	20.9	21	18.6	13.8	9.1	6.2	44.5	-0.34
Brest	6.1	5.8	7.8	9.2	11.6	14.4	15.6	16	14.7	12	9	7	48.24	-4.29
Clermont	2.6	3.7	7.5	10.3	13.8	17.3	19.4	19.1	16.2	11.2	6.6	3.6	45.47	3.05
Grenoble	1.5	3.2	7.7	10.6	14.5	17.8	20.1	19.5	16.7	11.4	6.5	2.3	45.1	5.43
Lille	2.4	2.9	6	8.9	12.4	15.3	17.1	17.1	14.7	10.4	6.1	3.5	50.38	3.04
Lyon	2.1	3.3	7.7	10.9	14.9	18.5	20.7	20.1	16.9	11.4	6.7	3.1	45.45	4.51
Marseille	5.5	6.6	10	13	16.8	20.8	23.3	22.8	19.9	15	10.2	6.9	43.18	5.24
Montpellier	5.6	6.7	9.9	12.8	16.2	20.1	22.7	22.3	19.3	14.6	10	6.5	43.36	3.53
Nantes	5	5.3	8.4	10.8	13.9	17.2	18.8	18.6	16.4	12.2	8.2	5.5	47.13	-1.33
Nice	7.5	8.5	10.8	13.3	16.7	20.1	22.7	22.5	20.3	16	11.5	8.2	43.42	7.15
Paris	3.4	4.1	7.6	10.7	14.3	17.5	19.1	18.7	16	11.4	7.1	4.3	48.52	2.2
Rennes	4.8	5.3	7.9	10.1	13.1	16.2	17.9	17.8	15.7	11.6	7.8	5.4	48.05	-1.41
Strasbourg	0.4	1.5	5.6	9.8	14	17.2	19	18.3	15.1	9.5	4.9	1.3	48.35	7.45
Toulouse	4.7	5.6	9.2	11.6	14.9	18.7	20.9	20.9	18.3	13.3	8.6	5.5	43.36	1.26
Vichy	2.4	3.4	7.1	9.9	13.6	17.1	19.3	18.8	16	11	6.6	3.4	46.08	3.26

Visualisation des données

- Visualisation = projection en 2D



Visualisation des données

- Vol d'oiseaux en V



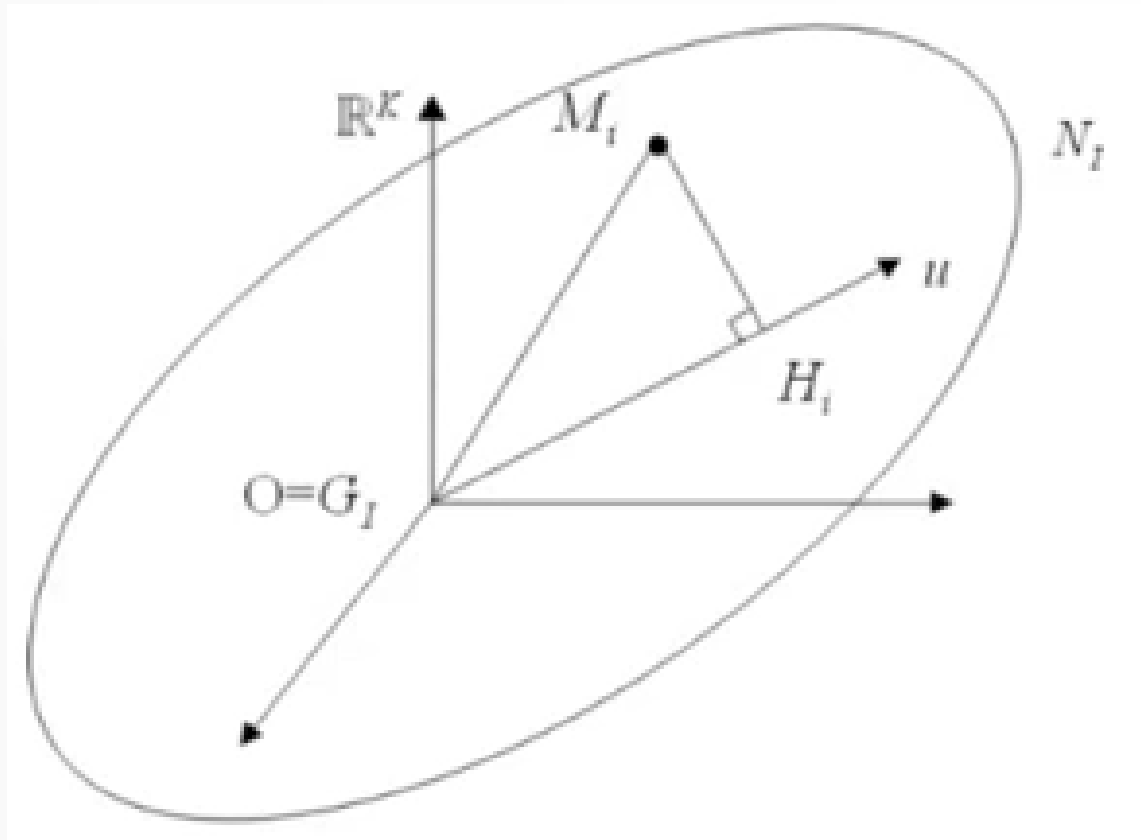
Matrice de variances-covariances

- Données stockées dans une matrice X
- n individus, p variables

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_i^T \\ \vdots \\ \mathbf{x}_n^T \end{pmatrix} = \begin{pmatrix} \vdots \\ \vdots \\ \dots & \dots & x_{ij} & \dots & \dots \\ \vdots \\ \vdots \end{pmatrix}$$

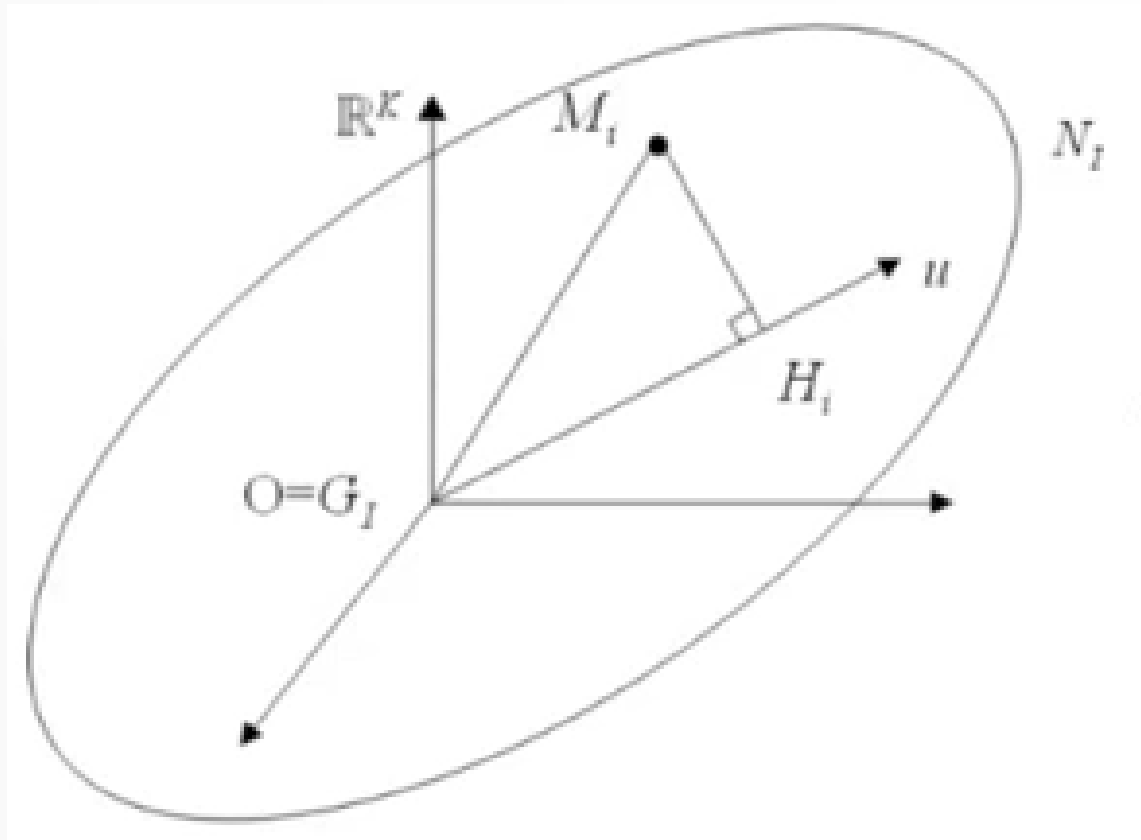
Matrice de variances-covariances

- Droite d'inertie maximale



Matrice de variances-covariances

- Droite d'inertie maximale



Trouver u tel que : $\sum_i \frac{1}{I} OH_i^2$ maximum

Matrice de variances-covariances

- Exercice

On compte le nombre de cerveaux endormis en séances de TP à deux moments différents de la journée : 7h45-9h45 et 10h-12h.

	7h45-9h45	10h-12h
TP 1	6	2
TP 2	5	3
TP 3	6	1
TP 4	4	3
TP 5	4	1

1. Calculer la matrice de variance-covariance Σ . Y a-t-il une dépendance sur le taux de cerveaux endormis entre 7h45-9h45 et 10h-12h ? Expliquer la réponse.
2. Calculer les composantes principales de ce jeu de données.
3. Représenter sur un même graphe les données et les composantes principales.

Illustration de l'ACP

- Données brutes

	Janv	Févr	Mars	Avri	Mai	Juin	juil	Août	Sept	Octo	Nove	Déce	Lati	Long
Bordeaux	5.6	6.6	10.3	12.8	15.8	19.3	20.9	21	18.6	13.8	9.1	6.2	44.5	-0.34
Brest	6.1	5.8	7.8	9.2	11.6	14.4	15.6	16	14.7	12	9	7	48.24	-4.29
Clermont	2.6	3.7	7.5	10.3	13.8	17.3	19.4	19.1	16.2	11.2	6.6	3.6	45.47	3.05
Grenoble	1.5	3.2	7.7	10.6	14.5	17.8	20.1	19.5	16.7	11.4	6.5	2.3	45.1	5.43
Lille	2.4	2.9	6	8.9	12.4	15.3	17.1	17.1	14.7	10.4	6.1	3.5	50.38	3.04
Lyon	2.1	3.3	7.7	10.9	14.9	18.5	20.7	20.1	16.9	11.4	6.7	3.1	45.45	4.51
Marseille	5.5	6.6	10	13	16.8	20.8	23.3	22.8	19.9	15	10.2	6.9	43.18	5.24
Montpellier	5.6	6.7	9.9	12.8	16.2	20.1	22.7	22.3	19.3	14.6	10	6.5	43.36	3.53
Nantes	5	5.3	8.4	10.8	13.9	17.2	18.8	18.6	16.4	12.2	8.2	5.5	47.13	-1.33
Nice	7.5	8.5	10.8	13.3	16.7	20.1	22.7	22.5	20.3	16	11.5	8.2	43.42	7.15
Paris	3.4	4.1	7.6	10.7	14.3	17.5	19.1	18.7	16	11.4	7.1	4.3	48.52	2.2
Rennes	4.8	5.3	7.9	10.1	13.1	16.2	17.9	17.8	15.7	11.6	7.8	5.4	48.05	-1.41
Strasbourg	0.4	1.5	5.6	9.8	14	17.2	19	18.3	15.1	9.5	4.9	1.3	48.35	7.45
Toulouse	4.7	5.6	9.2	11.6	14.9	18.7	20.9	20.9	18.3	13.3	8.6	5.5	43.36	1.26
Vichy	2.4	3.4	7.1	9.9	13.6	17.1	19.3	18.8	16	11	6.6	3.4	46.08	3.26

Illustration de l'ACP

- Données centrées réduites

	Janv	Févr	Mars	Avri	Mai	Juin	juil	Août	Sept	Octo	Nove	Déce	Lati	Long
Bordeaux	0.84	0.98	1.40	1.33	0.94	0.85	0.52	0.74	0.90	0.84	0.67	0.72	44.5	-0.34
Brest	1.10	0.54	-0.29	-1.30	-1.95	-1.98	-2.06	-1.83	-1.28	-0.18	0.62	1.14	48.24	-4.29
Clermont	-0.71	-0.63	-0.50	-0.50	-0.44	-0.31	-0.21	-0.24	-0.44	-0.63	-0.76	-0.66	45.47	3.05
Grenoble	-1.28	-0.90	-0.36	-0.28	0.05	-0.02	0.13	-0.03	-0.16	-0.52	-0.82	-1.35	45.1	5.43
Lille	-0.81	-1.07	-1.51	-1.52	-1.40	-1.46	-1.33	-1.27	-1.28	-1.09	-1.05	-0.71	50.38	3.04
Lyon	-0.97	-0.85	-0.36	-0.06	0.32	0.38	0.42	0.27	-0.05	-0.52	-0.70	-0.92	45.45	4.51
Marseille	0.79	0.98	1.20	1.48	1.63	1.71	1.69	1.66	1.63	1.52	1.30	1.09	43.18	5.24
Montpellier	0.84	1.03	1.13	1.33	1.22	1.31	1.39	1.41	1.30	1.29	1.19	0.87	43.36	3.53
Nantes	0.53	0.26	0.11	-0.13	-0.37	-0.37	-0.50	-0.50	-0.33	-0.07	0.16	0.35	47.13	-1.33
Nice	1.82	2.03	1.74	1.70	1.56	1.31	1.39	1.51	1.86	2.08	2.05	1.77	43.42	7.15
Paris	-0.30	-0.41	-0.43	-0.20	-0.09	-0.19	-0.36	-0.45	-0.55	-0.52	-0.47	-0.29	48.52	2.2
Rennes	0.43	0.26	-0.23	-0.64	-0.92	-0.94	-0.94	-0.91	-0.72	-0.41	-0.07	0.29	48.05	-1.41
Strasbourg	-1.84	-1.85	-1.78	-0.86	-0.30	-0.37	-0.41	-0.65	-1.06	-1.60	-1.74	-1.87	48.35	7.45
Toulouse	0.37	0.42	0.65	0.45	0.32	0.50	0.52	0.69	0.74	0.55	0.39	0.35	43.36	1.26
Vichy	-0.81	-0.79	-0.77	-0.79	-0.57	-0.42	-0.26	-0.39	-0.55	-0.75	-0.76	-0.76	46.08	3.26

Illustration de l'ACP

- ACP sur les données centrées réduites

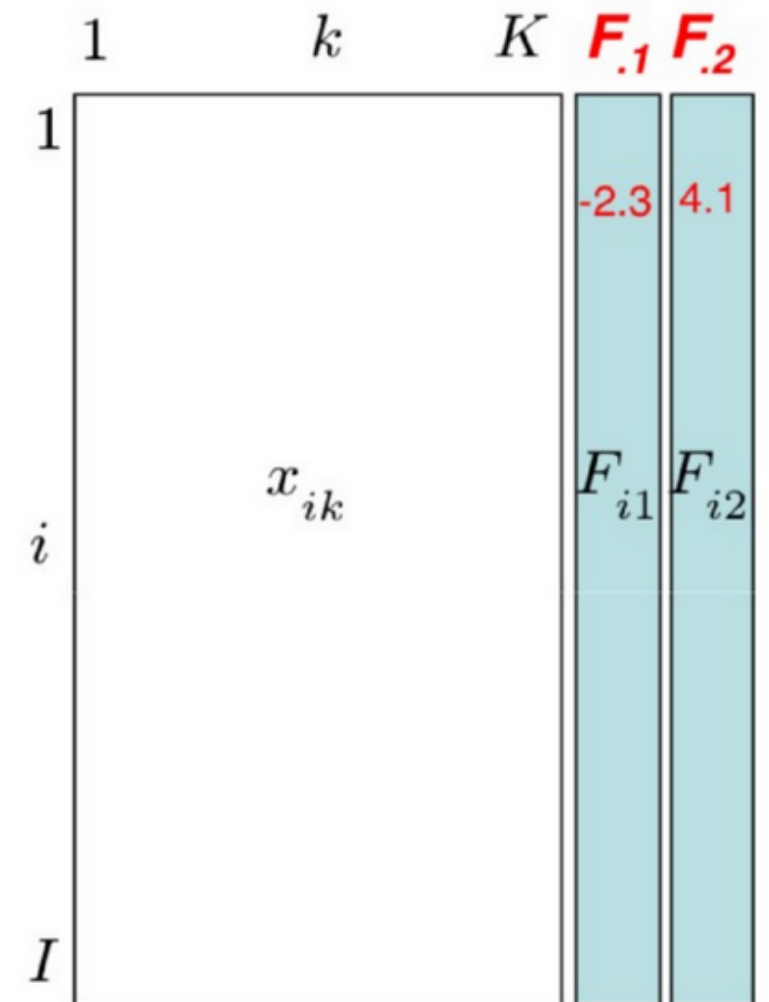
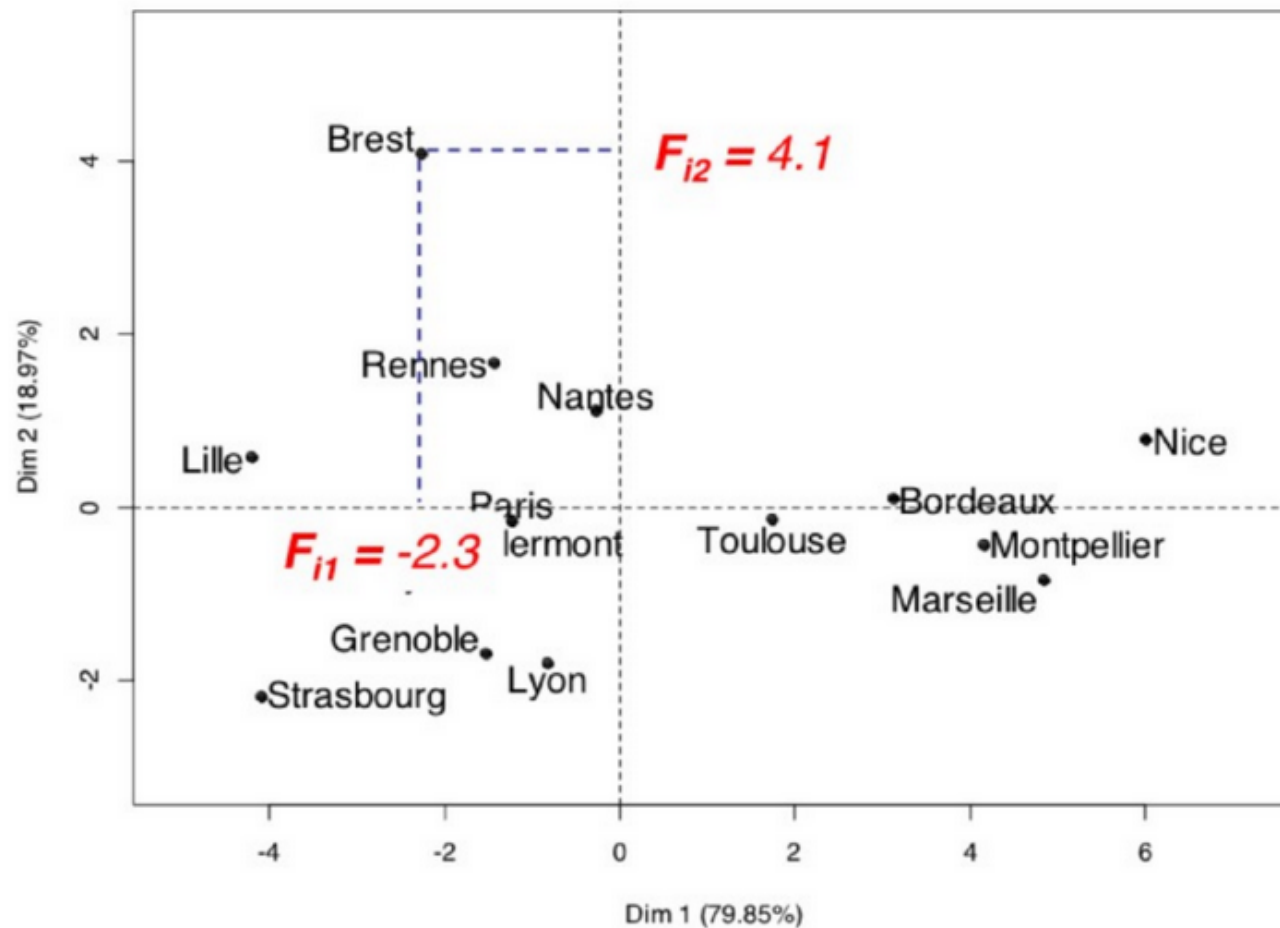


Illustration de l'ACP

- Cercle de corrélation

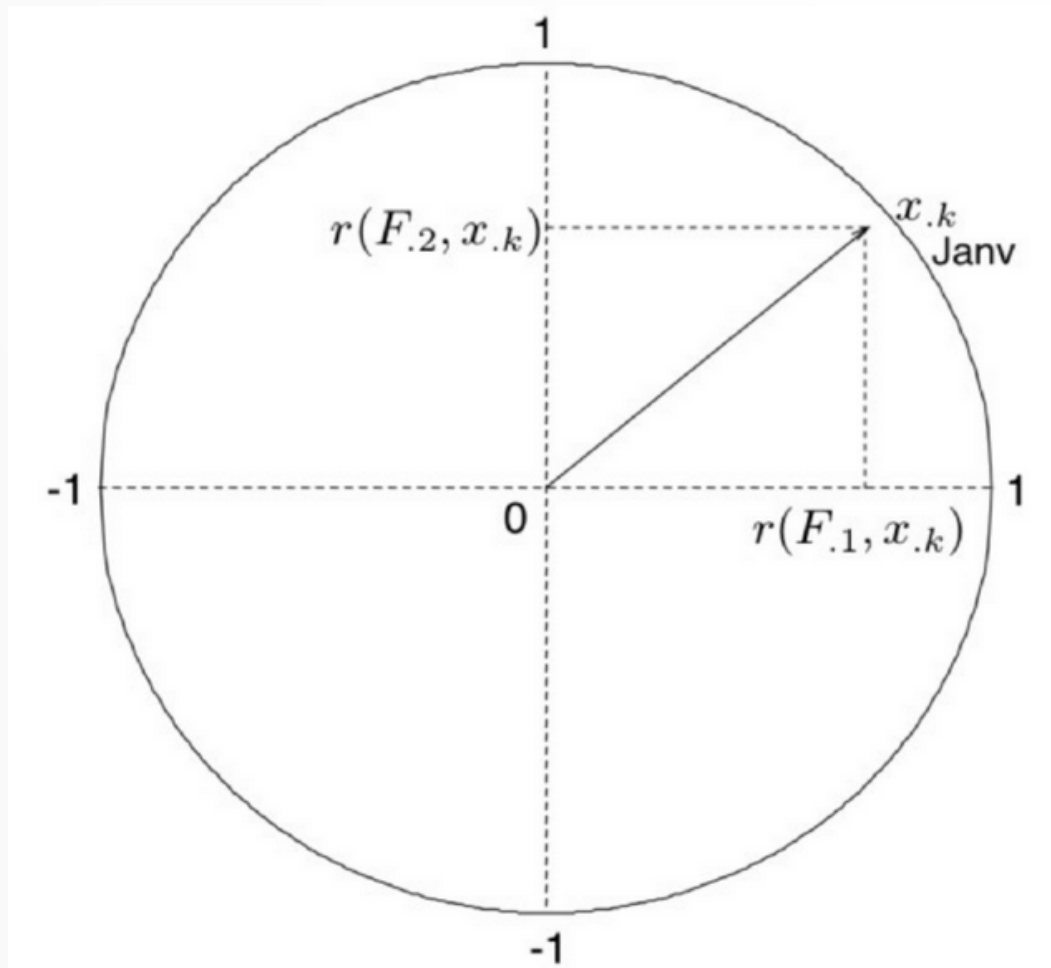


Illustration de l'ACP

- Cercle de corrélation

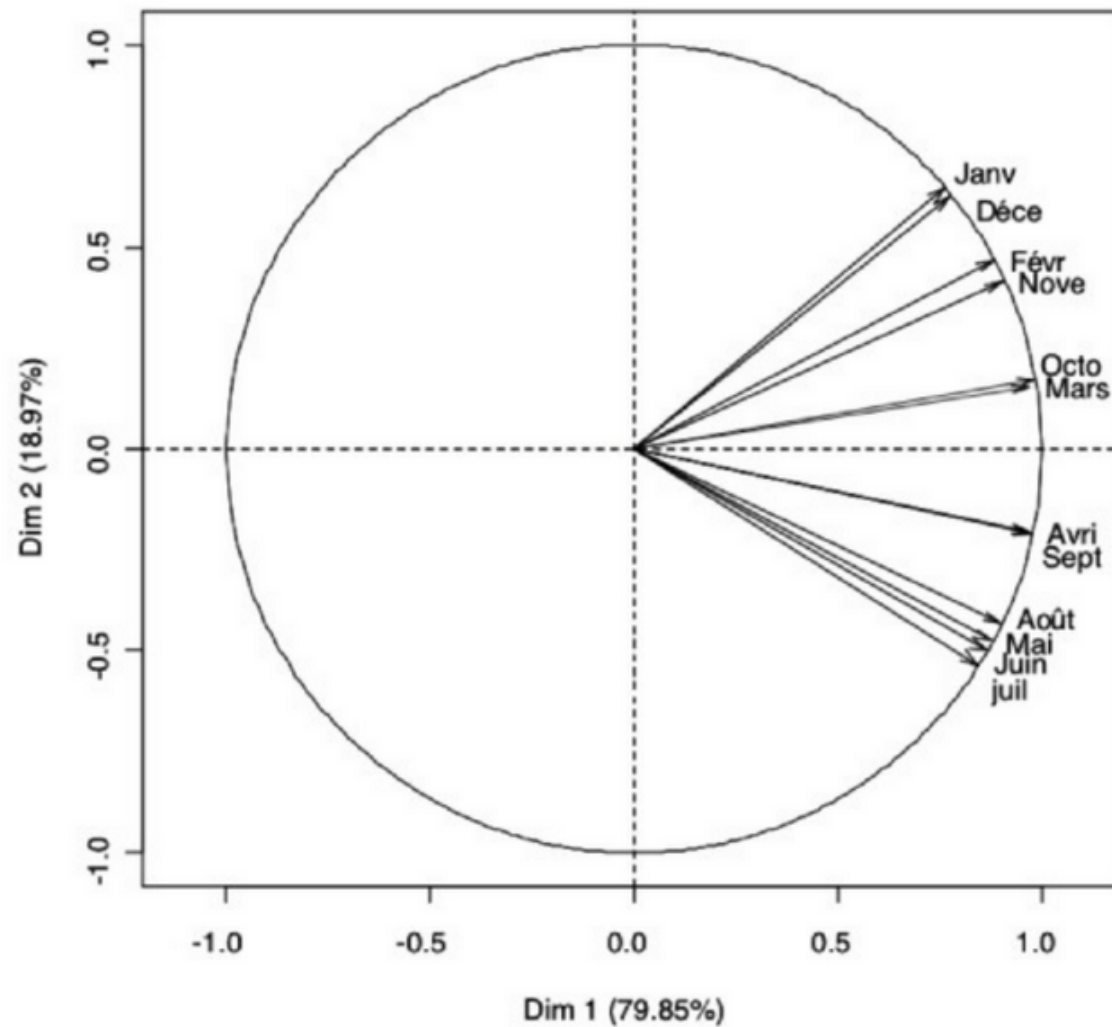


Illustration de l'ACP

- Cercle de corrélation

	janv	févr	nove	déce	moy	ampli	F1	F2
Bordeaux	5.6	6.6	9.1	6.2	13.33	15.4	3.12	-0.11
Brest	6.1	5.8	9.0	7.0	10.77	10.2	-2.27	-4.09
Clermont	2.6	3.7	6.6	3.6	10.94	16.8	-1.73	0.59
Grenoble	1.5	3.2	6.5	2.3	10.98	18.6	-1.53	1.69
Lille	2.4	2.9	6.1	3.5	9.73	14.7	-4.22	-0.6
Lyon	2.1	3.3	6.7	3.1	11.36	18.6	-0.83	1.79
Marseille	5.5	6.6	10.2	6.9	14.23	17.8	4.83	0.83
Montpellier	5.6	6.7	10.0	6.5	13.89	17.1	4.15	0.44
Nantes	5.0	5.3	8.2	5.5	11.69	13.8	-0.28	-1.11
Nice	7.5	8.5	11.5	8.2	14.84	15.2	6.01	-0.79
Paris	3.4	4.1	7.1	4.3	11.18	15.7	-1.24	0.16
Rennes	4.8	5.3	7.8	5.4	11.13	13.1	-1.44	-1.67
Strasbourg	.4	1.5	4.9	1.3	9.72	18.6	-4.11	2.17
Toulouse	4.7	5.6	8.6	5.5	12.68	16.2	1.74	0.14
Vichy	2.4	3.4	6.6	3.4	10.72	16.9	-2.2	0.58



Illustration de l'ACP

- Matrice de corrélation

	janv	févr	mars	avr	mai	juin	juil	août	sept	octo	nove	déce
janvier	1.00											
février	.97	1.00										
mars	.84	.93	1.00									
avril	.61	.76	.92	1.00								
mai	.36	.55	.77	.95	1.00							
juin	.34	.52	.76	.94	.99	1.00						
juillet	.30	.49	.72	.91	.98	.99	1.00					
août	.41	.59	.80	.95	.98	.99	.99	1.00				
septembre	.60	.76	.91	.98	.94	.94	.93	.97	1.00			
octobre	.85	.94	.97	.91	.77	.76	.74	.81	.93	1.00		
novembre	.95	.99	.93	.78	.59	.57	.55	.64	.80	.96	1.00	
décembre	.99	.97	.83	.62	.38	.36	.32	.43	.62	.87	.96	1.00