

## 2. Analyse en composantes principales

MAT8594

UQAM

- 1 2.1 Introduction
- 2 2.2 Petit rappel d'algèbre linéaire
- 3 2.3 Données
  - 2.3.1 Nuage des observations
  - 2.3.2 Nuage des variables
- 4 2.4 Analyse du nuage des observations
- 5 2.5 Analyse du nuage des variables
- 6 2.6 Interprétation des résultats
- 7 2.7 Applications
  - 2.7.1 Application 1 : Eurojobs
- 8 2.8 ACP : outil d'apprentissage machine
  - 2.8.1 Compression
  - 2.8.2 Application 2 : les petits chats

## 2.1 Introduction

# Base de données

	Agr	Min	Man	PS	Con	SI	Fin	SPS	TC
Belgium	3.3	0.9	27.6	0.9	8.2	19.1	6.2	26.6	7.2
Denmark	9.2	0.1	21.8	0.6	8.3	14.6	6.5	32.2	7.1
France	10.8	0.8	27.5	0.9	8.9	16.8	6.0	22.6	5.7
W. Germany	6.7	1.3	35.8	0.9	7.3	14.4	5.0	22.3	6.1
...	...	...	...	...	...	...	...	...	...

TABLE: Composantes de l'économie (%) pour certains pays.

# Objectifs

Une analyse doit être réalisée à partir d'une base de données contenant (ici !)  $k = 9$  variables numériques et  $n = 26$  sujets.

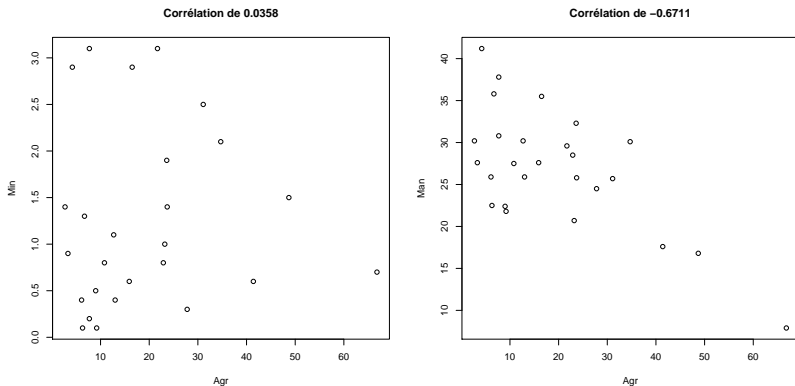


FIGURE: Nuages de points.

# Matrice des corrélations

	Agr	Min	Man	PS	Con	SI	Fin	SPS	TC
Agr	<b>1.00</b>	0.04	-0.67	-0.40	-0.54	-0.74	-0.22	-0.75	-0.56
Min	0.04	<b>1.00</b>	0.44	0.40	-0.03	-0.40	-0.44	-0.28	0.16
Man	-0.67	0.44	<b>1.00</b>	0.38	0.49	0.20	-0.16	0.15	0.35
PS	-0.40	0.40	0.38	<b>1.00</b>	0.06	0.20	0.11	0.13	0.38
Con	-0.54	-0.03	0.49	0.06	<b>1.00</b>	0.36	0.02	0.16	0.39
SI	-0.74	-0.40	0.20	0.20	0.36	<b>1.00</b>	0.37	0.57	0.19
Fin	-0.22	-0.44	-0.16	0.11	0.02	0.37	<b>1.00</b>	0.11	-0.25
SPS	-0.75	-0.28	0.15	0.13	0.16	0.57	0.11	<b>1.00</b>	0.57
TC	-0.56	0.16	0.35	0.38	0.39	0.19	-0.25	0.57	<b>1.00</b>

TABLE: Matrice de corrélation ( $R$ ).

# Objectifs

- L'*analyse en composantes principales*, ou ACP, est une technique ayant pour objectif principal la réduction du nombre de dimensions de l'échantillon initial en minimisant la perte d'information.
- On souhaite décrire la variabilité (l'inertie) présente dans un ensemble de variables initiales *corrélées*

$$[X_1 \quad \cdots \quad X_k]$$

à l'aide d'un nouvel ensemble de variables *non corrélées*

$$[Y_1 \quad \cdots \quad Y_k].$$

- Chacune des nouvelles variables  $Y_i$ ,  $i = 1, \dots, k$ , est une combinaison linéaire des variables initiales  $X_i$ ,  $i = 1, \dots, k$

$$Y_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ik}X_k.$$

# Projections

Projections  $\mathbb{R}^2 \rightarrow \mathbb{R}^1$

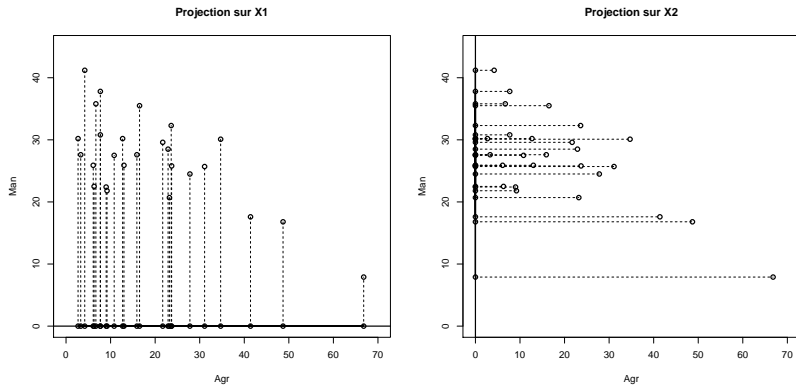


FIGURE: Projection sur une droite.



## 2.2 Petit rappel d'algèbre linéaire

# Matrices et vecteurs

On considère deux matrices  $\mathbf{A}$  et  $\mathbf{B}$  de taille  $n \times k$

$$\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nk} \end{bmatrix} \text{ et } \mathbf{B} = \begin{bmatrix} b_{11} & \cdots & b_{1k} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nk} \end{bmatrix}$$

et deux vecteurs  $\mathbf{X}$  et  $\mathbf{Y}$  de longueur  $k$

$$\mathbf{X} = \begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix} \text{ et } \mathbf{Y} = \begin{bmatrix} y_1 \\ \vdots \\ y_k \end{bmatrix}.$$

# Opérations

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} a_{11} + b_{11} & \cdots & a_{1k} + b_{1k} \\ \vdots & \ddots & \vdots \\ a_{n1} + b_{n1} & \cdots & a_{nk} + b_{nk} \end{bmatrix}$$

$$\begin{aligned} \mathbf{x}^T \mathbf{Y} &= \sum_{j=1}^k x_j y_j \\ &= \mathbf{Y}^T \mathbf{x}. \end{aligned}$$

# Définitions

- Le rang  $r$  d'une matrice  $\mathbf{A}$  est le nombre de lignes (ou de colonnes) indépendantes.
- Si  $n = k$  et que  $r(\mathbf{A}) = k$ , alors l'inverse de la matrice  $\mathbf{A}$  est unique et satisfait

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_k,$$

où  $\mathbf{I}_k$  est la matrice identité de taille  $k$ .

- La trace d'une matrice  $\mathbf{A}$  est donnée par

$$\text{tr}(\mathbf{A}) = \sum_{j=1}^k a_{jj}.$$

# Définitions

On peut également définir

- un produit scalaire obtenu à partir d'une métrique  $\mathbf{M}$  :  
 $\langle \mathbf{X}, \mathbf{Y} \rangle_{\mathbf{M}} = \mathbf{X}^T \mathbf{M} \mathbf{Y}$  ;
- une norme :  $\|\mathbf{X}\|_{\mathbf{M}} = \sqrt{\langle \mathbf{X}, \mathbf{X} \rangle_{\mathbf{M}}}$  et
- une distance :  $d_{\mathbf{M}}(\mathbf{X}, \mathbf{Y}) = \|\mathbf{X} - \mathbf{Y}\|_{\mathbf{M}}$ .

# Décomposition spectrale d'une matrice

- La décomposition spectrale d'une matrice  $\mathbf{R}$  ( $k \times k$ ) consiste à rechercher son **squelette** en réorganisant l'information de manière **hiérarchique**.
- Les  $k$  couples de **valeurs propres** ( $\lambda_i$ ) et de **vecteurs propres** ( $\mathbf{a}_i$ ) forment la décomposition spectrale d'une matrice donnée.
- Les valeurs propres correspondent aux  $k$  solutions possibles de l'équation :

$$\det(\mathbf{R} - \lambda \mathbf{I}_k) = 0$$

avec  $k = \text{rang}(\mathbf{R})$  le nombre maximal de lignes/colonnes indépendantes dans la matrice considérée. Le vecteur propre associé à la valeur propre  $\lambda_i$  est donné par

$$\mathbf{R}\mathbf{a}_i = \lambda_i \mathbf{a}_i.$$

# Propriétés

On note que

$$\sum_{j=1}^k \lambda_j = \text{tr}(\mathbf{R})$$

et

$$\prod_{j=1}^k \lambda_j = \det(\mathbf{R}).$$

# Exemple 1

Déterminer la décomposition spectrale de la matrice

$$\mathbf{A} = \begin{bmatrix} 7 & 4 \\ 3 & 2 \end{bmatrix}.$$



# Quelques fonctions en R

- $t(A)$  : transposée de  $A$  ;
- $solve(A)$  : avec un seul argument (matrice carrée) : inverse de la matrice  $A$  ; avec deux arguments (une matrice carrée et un vecteur) : solution du système d'équation  $Ax = b$  ;
- $diag(A)$  : avec une matrice en argument : diagonale de la matrice  $A$  ; avec un vecteur en argument : matrice diagonale formée avec le vecteur ; avec un scalaire  $k$  en argument : matrice identité  $k \times k$  ;
- $eigen(A)$  : permet d'obtenir les valeurs propres et les vecteurs propres de la matrice  $A$  ;
- $A \% * \% B$  : le produit de la matrice  $A$  et de la matrice  $B$ .

## 2.3 Données

# Données

Pour simplifier la présentation, on ne conserve que  $n = 6$  pays et  $k = 3$  variables : matrice  $\mathbf{X}$  de taille  $(6 \times 3)$ .

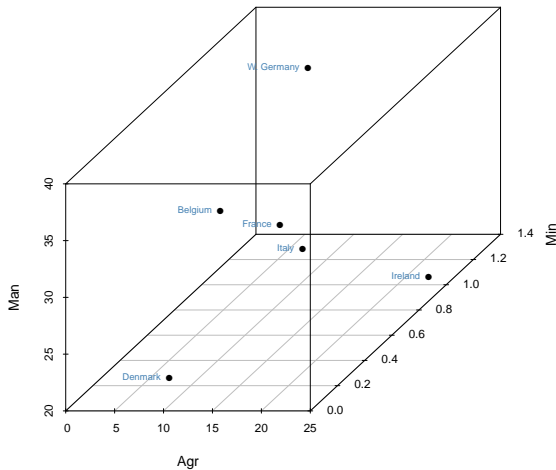
	Agr	Min	Man
Belgium	3.3	0.9	27.6
Denmark	9.2	0.1	21.8
France	10.8	0.8	27.5
W. Germany	6.7	1.3	35.8
Ireland	23.2	1.0	20.7
Italy	15.9	0.6	27.6

TABLE: Composantes de l'économie (%) pour certains pays.

On note  $\mathbf{x}_i$ , le vecteur correspondant à la  $i^{\text{e}}$  ligne et  $\mathbf{x}^j$ , le vecteur correspondant à la  $j^{\text{e}}$  colonne.

# Nuage des observations

Nuage des observations



# Nuage des observations

- Les  $n$  lignes de  $\mathbf{X}$ ,  $\mathbf{x}_i$  pour  $i = 1, \dots, n$ , définissent un nuage de points dans  $\mathbb{R}^k$ .
- Pour réaliser une analyse en composantes principales (ACP), on aura besoin du *nuage centré* et/ou du *nuage centré-réduit*.

## Nuage centré des observations

- La matrice  $\mathbf{Y}$  des observations centrées est composée des éléments  $y_{ij}$  tels que

$$\bar{x}^j = \frac{1}{n} \sum_{i=1}^n x_{ij}$$

$$y_{ij} = x_{ij} - \bar{x}^j.$$

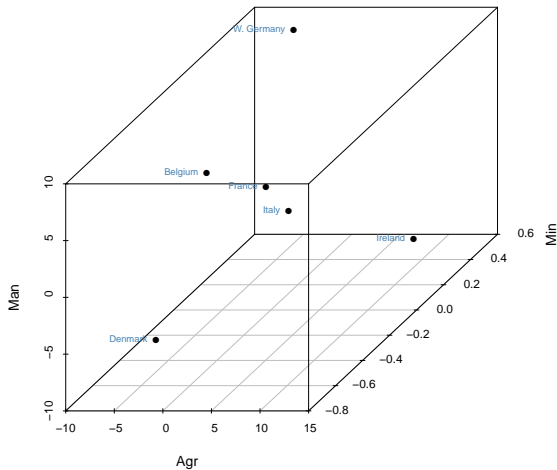
- Le nouveau centre de gravité du nuage est le point d'origine.
- Les distances entre les observations sont préservées, c'est-à-dire que

$$d_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) = d_{\mathbf{M}}(\mathbf{y}_i, \mathbf{y}_j),$$

où  $d_{\mathbf{M}}()$  est une mesure de distance.

# Nuage centré des observations

Nuage centré des observations



## Nuage centré-réduit des observations

- La matrice  $\mathbf{Z}$  des observations centrées-réduites est composée des éléments  $z_{ij}$  tels que

$$s_j^2 = \text{Var} [\mathbf{x}^j] = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}^j)^2$$

$$z_{ij} = \frac{x_{ij} - \bar{x}^j}{s_j}.$$

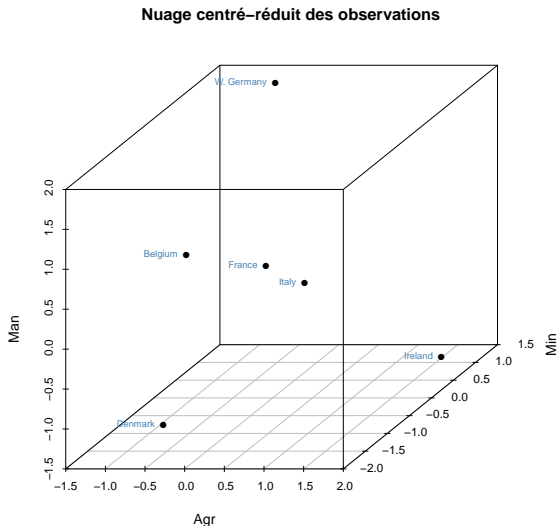
- La variance des variables centrées-réduites est 1.
- Les distances entre les observations sont modifiées, c'est-à-dire que

$$d_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) \neq d_{\mathbf{M}}(\mathbf{y}_i, \mathbf{y}_j),$$

où  $d_{\mathbf{M}}()$  est une mesure de distance.



# Nuage centré-réduit des observations



## Distance entre deux observations

- Il faut définir une distance  $d_{\mathbf{M}}()$  sur l'espace  $\mathbb{R}^k$ .
- Dans le cadre de l'analyse en composantes principales, on utilise essentiellement des métriques diagonales, c'est-à-dire  $\mathbf{M} = \text{diag}(m_1, \dots, m_k)$  pour permettre la pondération des variables :

$$d_{\mathbf{M}}(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{j=1}^k m_j (x_j - y_j)^2}.$$

## Distance entre deux observations

- Si on accorde la même importance à toutes les variables, on a alors  $\mathbf{M} = \mathbf{I}_k$  et on obtient la distance euclidienne

$$d_2(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{j=1}^k (x_j - y_j)^2}.$$

- Si on accorde moins d'importance aux variables ayant les plus fortes variances, on a alors  $\mathbf{M} = \text{diag}(1/s_1^2, \dots, 1/s_k^2)$  et on obtient

$$d_{\mathbf{M}}(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{j=1}^k \frac{1}{s_j^2} (x_j - y_j)^2}.$$

- Centrer et réduire les données permet d'accorder la même importance à toutes les variables dans les calculs.

# Inertie

- L'inertie d'un nuage d'observations est une mesure de la dispersion de ce même nuage. Elle est définie par

$$\mathcal{I}(\mathbf{X}) = \sum_{i=1}^n w_i d_{\mathbf{M}}^2(\mathbf{x}_i^T, \bar{\mathbf{x}}), \quad \text{avec } \bar{\mathbf{x}} = \begin{bmatrix} \bar{x}^1 \\ \vdots \\ \bar{x}^k \end{bmatrix}.$$

## Proposition 1.

Si on choisit une métrique diagonale  $\mathbf{M} = \text{diag}(m_1, \dots, m_k)$  et  $w_i = 1/n$ ,  $\forall n$ , on obtient

$$\mathcal{I}(\mathbf{X}) = \sum_{j=1}^k m_j \text{Var} [\mathbf{x}^j].$$

Si on utilise  $\mathbf{M} = \mathbf{I}_k$ , alors  $\mathcal{I}(\mathbf{Y}) = \sum_{j=1}^k \text{Var} [\mathbf{x}^j]$  et  $\mathcal{I}(\mathbf{Z}) = k$ .

# Nuages des variables

- Les  $k$  colonnes de  $\mathbf{X}$  forment un nuage de  $k$  points dans  $\mathbb{R}^n$ .
- On peut accorder un poids de  $m_j$  à chacune des variables mais en analyse en composantes principales, ce poids est  $m_j = 1, \forall j$ .
- Comme pour le nuage des observations, on va s'intéresser au nuage des variables centrées (si on travaille avec  $\mathbf{Y}$ ) et au nuage des variables centrées-réduites (si on analyse  $\mathbf{Z}$ ).

## Nuage des variables

- On définit une métrique diagonale  $\mathbf{N} = \text{diag}(1/n, \dots, 1/n)$  pour  $\mathbb{R}^n$ .  
On a alors

$$\text{Var} [\mathbf{x}^j] = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}^j)^2 = \|\mathbf{y}^j\|_{\mathbf{N}}^2.$$

- On remarque que la variance empirique d'une variable  $j$  est égale au carré de la norme de la variable centrée et que la norme d'une variable centrée-réduite  $\mathbf{z}^j$  est égale à 1, c'est-à-dire

$$\text{Var} [\mathbf{z}^j] = \|\mathbf{z}^j\|_{\mathbf{N}}^2 = 1.$$

# Covariance

- La covariance empirique  $c_{j_1 j_2}$  mesure le lien entre deux variables  $j_1$  et  $j_2$  :

$$\begin{aligned} c_{j_1 j_2} &= \frac{1}{n} \sum_{i=1}^n (x_{ij_1} - \bar{x}^{j_1}) (x_{ij_2} - \bar{x}^{j_2}) \\ &= \langle \mathbf{y}^{j_1}, \mathbf{y}^{j_2} \rangle_{\mathbf{N}}, \end{aligned}$$

où  $\langle \mathbf{a}, \mathbf{b} \rangle_{\mathbf{N}}$  indique le produit scalaire des vecteurs  $\mathbf{a}$  et  $\mathbf{b}$ .

- La matrice de covariance empirique peut alors être calculée à partir de la matrice  $\mathbf{Y}$  selon la formule suivante

$$\mathbf{C} = \mathbf{Y}^T \mathbf{N} \mathbf{Y}.$$

# Corrélation

- La corrélation empirique  $r_{j_1 j_2}$  mesure également le lien entre les deux variables  $j_1$  et  $j_2$  et prend des valeurs dans l'intervalle  $[-1, 1]$

$$\begin{aligned}
 r_{j_1 j_2} &= \frac{1}{n} \sum_{i=1}^n \left( \frac{x_{ij_1} - \bar{x}^{j_1}}{s_{j_1}} \right) \left( \frac{x_{ij_2} - \bar{x}^{j_2}}{s_{j_2}} \right) \\
 &= \frac{\langle \mathbf{y}^{j_1}, \mathbf{y}^{j_2} \rangle_{\mathbf{N}}}{\|\mathbf{y}^{j_1}\|_{\mathbf{N}} \|\mathbf{y}^{j_2}\|_{\mathbf{N}}} \\
 &= \langle \mathbf{z}^{j_1}, \mathbf{z}^{j_2} \rangle_{\mathbf{N}} \\
 &= \cos \theta \left( \mathbf{z}^{j_1}, \mathbf{z}^{j_2} \right),
 \end{aligned}$$

où  $\theta(\mathbf{a}, \mathbf{b})$  indique l'angle entre les vecteurs  $\mathbf{a}$  et  $\mathbf{b}$ .

- La matrice de covariance empirique peut alors être calculée à partir de la matrice  $\mathbf{Z}$  selon la formule suivante

$$\mathbf{Z} = \mathbf{Z}^T \mathbf{N} \mathbf{Z}.$$



# En bref...

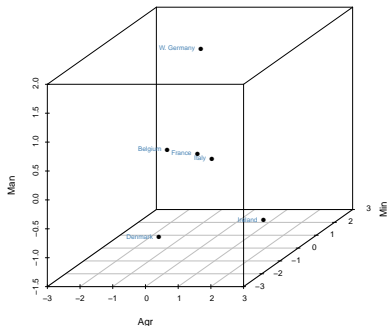
- Une analyse en composantes principales peut se faire à partir des données centrées ( $\mathbf{Y}$ ) ou des données centrée-réduites ( $\mathbf{Z}$ ).
- Une analyse en composantes principales se fait en étudiant deux nuages : le nuage des  $n$  observations dans  $\mathbb{R}^k$  avec la métrique  $\mathbf{M} = \mathbf{I}_k$  et le nuage des  $k$  variables dans  $\mathbb{R}^n$  avec la métrique  $\mathbf{N} = (1/n)\mathbf{I}_n$ .
- Dans la très grande majorité des cas, on réalise une *analyse en composantes principales normée*, c'est-à-dire à partir de la matrice des corrélations  $\mathbf{R}$  construites à partir des données centrées-réduites  $\mathbf{Z}$ .

## 2.4 Analyse du nuage des observations

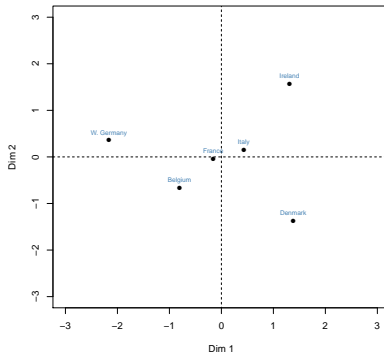
# Analyse du nuage des observations

**Objectif** : trouver un espace (ici un plan) tel que les distances entre les observations soient les mieux préservées.

Nuage centré-réduit des observations



Projection sur un plan



## Projection d'une observation

- La projection (**M**-orthogonale) d'un point  $\mathbf{z}_i \in \mathbb{R}^k$  sur un axe  $D_\alpha$  dont la direction est donnée par un vecteur  $\mathbf{v}_\alpha$  de norme 1, c'est-à-dire tel que

$$\mathbf{v}_\alpha^T \mathbf{M} \mathbf{v}_\alpha = 1,$$

a pour coordonnée

$$p_{i\alpha} = \langle \mathbf{z}_i^T, \mathbf{v}_\alpha \rangle_{\mathbf{M}} = \mathbf{z}_i \mathbf{M} \mathbf{v}_\alpha.$$

- Pour l'ensemble de l'échantillon, on notera

$$\mathbf{p}_\alpha = \begin{bmatrix} p_{1\alpha} \\ \vdots \\ p_{n\alpha} \end{bmatrix} = \mathbf{Z} \mathbf{M} \mathbf{v}_\alpha.$$

# Projection d'une observation

- $\mathbf{p}_\alpha$  est une combinaison linéaire des colonnes de  $\mathbf{Z}$ . Par exemple, avec  $\mathbf{M} = \mathbf{I}_k$ , on a

$$\mathbf{p}_\alpha = \mathbf{Z}\mathbf{v}_\alpha = \sum_{j=1}^k v_{j\alpha} \mathbf{z}^j.$$

- $\mathbf{p}_1, \mathbf{p}_2, \dots$  sont nommées *composantes principales*.

## Exemple 2

On considère la matrice centrée-réduite

$$\mathbf{Z} = \begin{bmatrix} -1.1566844 & 0.28663500 & 0.1423795 \\ -0.3261240 & -1.67886215 & -0.9347525 \\ -0.1008873 & 0.04094786 & 0.1238083 \\ -0.6780564 & 1.26938358 & 1.6652213 \\ 1.6446973 & 0.53232215 & -1.1390362 \\ 0.6170548 & -0.45042643 & 0.1423795 \end{bmatrix}$$

que l'on veut projeter, en utilisant la métrique  $\mathbf{M} = \mathbf{I}_3$  sur les axes orthogonaux définis par

$$\mathbf{v}_1 = \begin{bmatrix} 0.4783158 \\ -0.5237373 \\ -0.7049207 \end{bmatrix} \quad \mathbf{v}_2 = \begin{bmatrix} 0.739170360 \\ 0.673517525 \\ 0.001149914 \end{bmatrix}.$$

Calculer  $\mathbf{p}_1$  et  $\mathbf{p}_2$ .

## Sélection des vecteurs $\mathbf{v}$

- Dans un premier temps, on cherche un axe défini par un vecteur  $\mathbf{v}_1 \in \mathbb{R}^k$  tel que

$$\mathbf{v}_1 = \arg \max_{\mathbf{v} \in \mathbb{R}^k; \|\mathbf{v}\|=1} (\text{Var} [\mathbf{Z}\mathbf{v}]) .$$

- Ce problème d'optimisation sous contraintes peut se réécrire sous la forme

$$\mathbf{v}_1 = \arg \max_{\mathbf{v} \in \mathbb{R}^k; \|\mathbf{v}\|=1} \left( \mathbf{v}^T \mathbf{R} \mathbf{v} \right) ,$$

où  $\mathbf{R} = (1/n)\mathbf{Z}^T \mathbf{Z}$  est la matrice de corrélation.

- On peut démontrer (pas nécessaire dans le cadre du cours) que la solution à ce problème d'optimisation est le vecteur propre  $\mathbf{a}_1$  obtenu par une décomposition spectrale de la matrice  $\mathbf{R}$ .

## Sélection des vecteurs $\mathbf{v}$

- Dans un second temps, on cherche un axe défini par un vecteur  $\mathbf{v}_2 \in \mathbb{R}^k$  tel que

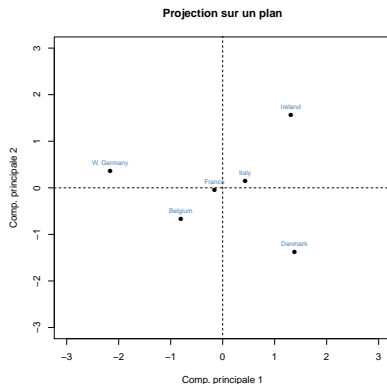
$$\mathbf{v}_2 = \underset{\mathbf{v} \in \mathbb{R}^k; \|\mathbf{v}\|=1; \mathbf{v} \text{ et } \mathbf{v}_1 \text{ non corr.}}{\arg \max} (\text{Var}[\mathbf{Z}\mathbf{v}]) .$$

- On peut démontrer (pas nécessaire dans le cadre du cours) que la solution à ce problème d'optimisation est le vecteur propre  $\mathbf{a}_2$  obtenu par une décomposition spectrale de la matrice  $\mathbf{R}$ .



## Sélection des vecteurs $\mathbf{v}$

Pour  $1 \leq q \leq r = \text{rang}(\mathbf{Z})$ , on construit les composantes principales  $\mathbf{p}_1, \dots, \mathbf{p}_q$  qui formeront une droite ( $q = 1$ ), un plan ( $q = 2$ ), etc. sur lequel on projettera les observations.



# Inertie

- Les composantes principales ( $\mathbf{p}_\alpha, \alpha = 1, \dots, q$ ) sont  $q$  nouvelles variables telles que
  - les corrélations entre les composantes principales sont nulles
  - la variance est  $\text{Var}[\mathbf{p}_\alpha] = \lambda_\alpha$ .
- Les  $q$  premières composantes principales représentent une inertie totale de

$$\mathcal{I}(\mathbf{p}_1, \dots, \mathbf{p}_q) = \lambda_1 + \dots + \lambda_q.$$

## 2.5 Analyse du nuage des variables

# Nuage des variables

Les 3 variables centrées-réduites forment un nuage de  $\mathbb{R}^6$  dans l'hypersphère de rayon 1.

	Belgium	Denmark	France	W. Germany	...
Agr	-1.1566844	-0.3261240	-0.10088728	-0.6780564	...
Min	0.2866350	-1.6788622	0.04094786	1.2693836	...
Man	0.1423795	-0.9347525	0.12380828	1.6652213	...

**Objectif** : trouver un espace (ici un plan) tel que les angles entre les variables (c'est-à-dire les corrélations) soient les mieux préservés.

## Projection d'une variable

- La projection (**N**-orthogonale) d'une variable  $\mathbf{z}^j \in \mathbb{R}^n$  sur un axe  $G_\beta$  décrit par un vecteur  $\mathbf{u}_\beta$  de norme 1, c'est-à-dire tel que

$$\mathbf{u}_\beta^T \mathbf{N} \mathbf{u}_\beta = 1,$$

a pour coordonnée

$$t_{j\beta} = \langle \mathbf{z}^j, \mathbf{u}_\beta \rangle = (\mathbf{z}^j)^T \mathbf{N} \mathbf{u}_\beta.$$

- Pour l'ensemble de l'échantillon, on notera

$$\mathbf{t}_\beta = \begin{bmatrix} t_{1\beta} \\ \vdots \\ t_{k\beta} \end{bmatrix} = \mathbf{Z}^T \mathbf{N} \mathbf{u}_\beta.$$

## Exemple 2 (suite)

En utilisant les données de l'Exemple 2, calculer  $\mathbf{t}_1$  et  $\mathbf{t}_2$  en utilisant la métrique  $\mathbf{N} = (1/6)\mathbf{I}_6$  et les axes orthogonaux définis par

$$\mathbf{u}_1 = \begin{bmatrix} 0.2655155 \\ -0.4566114 \\ 0.0518568 \\ 0.7145390 \\ -0.4330243 \\ -0.1422757 \end{bmatrix} \quad \mathbf{u}_2 = \begin{bmatrix} -0.29747573 \\ -0.61713072 \\ -0.02106045 \\ 0.15987825 \\ 0.70705671 \\ 0.06873194 \end{bmatrix}.$$

# Sélection des vecteurs $\mathbf{u}$

- Dans un premier temps, on cherche un axe défini par un vecteur  $\mathbf{u}_1 \in \mathbb{R}^n$  tel que

$$\mathbf{u}_1 = \arg \max_{\mathbf{u} \in \mathbb{R}^n; \|\mathbf{u}\|=1} \left( \|\mathbf{Z}^T \mathbf{N} \mathbf{u}\|^2 \right).$$

- En posant  $\mathbf{N} = (1/n)\mathbf{I}_n$ , on peut démontrer (pas nécessaire dans le cadre du cours) que la solution à ce problème d'optimisation est le vecteur propre  $\mathbf{b}_1$  obtenu par une décomposition spectrale de la matrice  $(1/n)\mathbf{Z}^T \mathbf{Z}$ .

## Sélection des vecteurs $\mathbf{u}$

- Dans un second temps, on cherche un axe défini par un vecteur  $\mathbf{u}_2 \in \mathbb{R}^n$  tel que

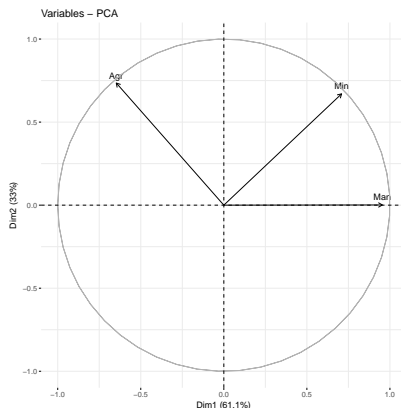
$$\mathbf{u}_2 = \underset{\mathbf{u} \in \mathbb{R}^n; \|\mathbf{u}\|=1; \mathbf{u} \text{ et } \mathbf{u}_1 \text{ non corr.}}{\arg \max} \left( \|\mathbf{Z}^T \mathbf{N} \mathbf{u}\|^2 \right).$$

- La solution à ce problème d'optimisation est le vecteur propre  $\mathbf{b}_2$  obtenu par une décomposition spectrale de la matrice  $(1/n)\mathbf{Z}^T \mathbf{Z}$ .



## Sélection des vecteurs $\mathbf{u}$

Pour  $1 \leq q^* \leq r^* = \text{rang}(\mathbf{Z}^T \mathbf{Z})$ , on construit les composantes principales  $\mathbf{t}_1, \dots, \mathbf{t}_{q^*}$  qui formeront une droite ( $q^* = 1$ ), un plan ( $q^* = 2$ ), etc. sur lequel on projettera les variables.



## 2.6 Interprétation des résultats

# Qualité de la représentation

- La qualité de la représentation est mesurée par le pourcentage de l'inertie initiale des données que les  $q$  premières composantes principales permettent d'expliquer :

$$\mathcal{I}(\mathbf{Z}) = \lambda_1 + \dots + \lambda_r = p \quad (\text{si on travaille avec } \mathbf{R})$$

$$\mathcal{I}(\mathbf{p}_1, \dots, \mathbf{p}_q) = \lambda_1 + \dots + \lambda_q \leq \mathcal{I}(\mathbf{Z}).$$

- La  $i^{\text{e}}$  composante principale permet d'expliquer une proportion de

$$\frac{\lambda_i}{\lambda_1 + \dots + \lambda_r}$$

de l'information initiale.

- Le passage de  $k$  à  $r$  dimension(s) se fait en préservant une proportion de

$$\frac{\lambda_1 + \dots + \lambda_k}{\lambda_1 + \dots + \lambda_r}$$

de l'information initiale.

## Exemple 3

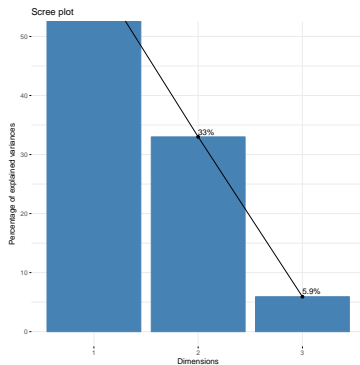
Avec les données de l'Exemple 2, calculer la qualité de la représentation des trois premières composantes principales. Vérifier les résultats en R en utilisant le code

```
res.pca <- PCA(data, scale.unit = TRUE, ncp = 3,  
               graph = FALSE)  
summary(res.pca)
```

# Combien de composantes principales ?

- 1 On peut décider de conserver le nombre de composantes principales nécessaires pour préserver une proportion déterminée d'inertie (généralement 80% ou 90%).
- 2 On peut décider de conserver les composantes principales dont l'inertie  $\lambda_\alpha$  est supérieure à l'inertie moyenne par variable (1 lorsque l'on travaille avec la matrice des corrélations  $\mathbf{R}$ ).
- 3 Visualiser le graphique des valeurs propres et chercher un « coude ».

# Graphique des valeurs propres (Exemple 3)



## Qualité de la représentation des observations

- Pour deux observations « bien projetées », la distance en projection est similaire à la distance dans le nuage initial  $\mathbb{R}^k$ .
- La qualité de la projection de l'observation  $i$  sur l'axe  $D_\alpha$  est mesurée par le carré du cosinus de l'angle  $\theta_{i\alpha}$  formé entre le vecteur  $\mathbf{z}_i$  et l'axe  $D_\alpha$  :

$$\cos^2(\theta_{i\alpha}) = \frac{\mathbf{p}_{i\alpha}^2}{\|\mathbf{z}_i\|^2}.$$

- Si l'espace final est un plan formé des axes  $D_1$  et  $D_2$ , la qualité de la projection de l'observation  $i$  sur ce plan est donnée par

$$\cos^2(\theta_{i(1,2)}) = \frac{\mathbf{p}_{i1}^2 + \mathbf{p}_{i2}^2}{\|\mathbf{z}_i\|^2}.$$

- Plus cette valeur est près de 1, meilleure est la projection.

## Exemple 3 (suite)

Avec les données de l'Exemple 2, on obtient

Individuals

		Dim.1		Dim.2		Dim.3
Belgium		0.449		0.304		0.247
Denmark		0.503		0.496		0.001
France		0.906		0.081		0.013
W. Germany		0.966		0.026		0.008
Ireland		0.401		0.577		0.022
Italy		0.307		0.039		0.654



# Contribution des observations

- Les observations qui contribuent de manière élevée à la construction des axes sont sources d'instabilité. La contribution d'une observation à un axe est mesurée par la proportion de l'inertie de l'axe expliquée par cette observation.
- Pour l'axe  $D_\alpha$ , l'inertie totale est  $\lambda_\alpha = \sum_{i=1}^n \mathbf{p}_{i\alpha}^2$ .
- La contribution de l'observation  $i$  sur l'axe  $D_\alpha$  est mesurée par

$$C(i, \alpha) = \frac{\mathbf{p}_{i\alpha}^2}{\lambda_\alpha}.$$

- Si l'espace final est un plan formé des axes  $D_1$  et  $D_2$ , la contribution de l'observation  $i$  sur ce plan est donnée par

$$C(i, (1, 2)) = \frac{\mathbf{p}_{i1}^2 + \mathbf{p}_{i2}^2}{\lambda_1 + \lambda_2}.$$

## Exemple 3 (suite)

Avec les données de l'Exemple 2, on obtient (en %)

Individuals

	Dim.1	Dim.2	Dim.3
Belgium	7.050	8.849	40.153
Denmark	20.849	38.085	0.380
France	0.269	0.044	0.039
W. Germany	51.057	2.556	4.394
Ireland	18.751	49.993	10.528
Italy	2.024	0.472	44.507

# Cercle des corrélations

- Si deux variables sont bien projetées, alors leur angle en projection est similaire à celui dans l'espace initial ( $\mathbb{R}^n$ ).
- Corrélation entre deux variables = cosinus de l'angle entre les variables centrées-réduites.
  - un angle droit (90 degrés) correspond à une corrélation nulle ;
  - un angle nul correspond à une corrélation de 1 ; et
  - un angle plat (180 degrés) correspond à une corrélation de  $-1$ .
- La corrélation entre la variable  $i$  et la composante principale  $j$  est  $\sqrt{\lambda_j} a_{ij}$ , où  $a_{ij}$  est le  $i^{\text{e}}$  élément du vecteur propre correspondant à la  $j^{\text{e}}$  plus grande valeur propre.

## Qualité de la projection des variables

- La qualité de la projection de la variable  $j$  sur l'axe  $G_\beta$  est mesurée par le carré du cosinus de l'angle  $\theta_{j\beta}$  formé entre le vecteur  $\mathbf{z}^j$  et l'axe  $G_\beta$  :

$$\cos^2(\theta_{j\beta}) = \frac{\mathbf{t}_{j\beta}^2}{\|\mathbf{z}^j\|^2} = \mathbf{t}_{j\beta}^2.$$

- Si l'espace final est un plan formé des axes  $G_1$  et  $G_2$ , la qualité de la projection de la variable  $j$  sur ce plan est donnée par

$$\cos^2(\theta_{j(1,2)}) = \mathbf{t}_{j1}^2 + \mathbf{t}_{j2}^2.$$

- On obtient ainsi une flèche de longueur  $\sqrt{\cos^2(\theta_{j(1,2)})}$  sur le cercle des corrélations pour chacune des variables. Plus la flèche est près de la circonférence du cercle, meilleure est la représentation de la variable.

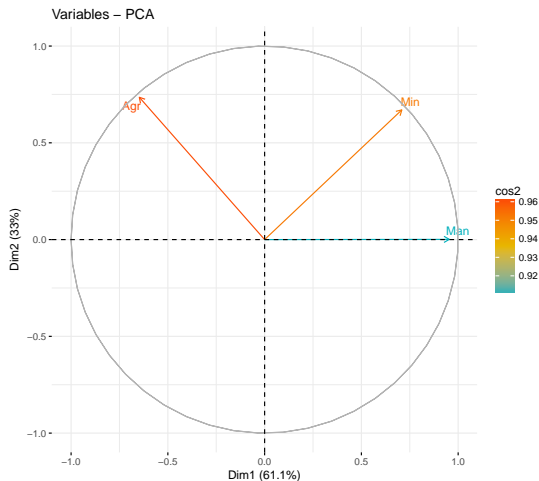
## Exemple 3 (suite)

Avec les données de l'Exemple 2, on obtient

Variables

	Dim.1	Dim.2	Dim.3
Agr	0.419	0.541	0.040
Min	0.503	0.449	0.048
Man	0.911	0.000	0.089

# Cercle des corrélations (Exemple 3)



# Contribution des variables

- Les contributions des variables aux axes permettent de donner une interprétation à ceux-ci. Pour déterminer la contribution d'une variable à un axe, on évalue la proportion de l'inertie totale de l'axe expliquée par la variable.
- Pour l'axe  $G_\beta$ , l'inertie totale est  $\lambda_\beta = \sum_{j=1}^p \mathbf{t}_{j\beta}^2$ .
- La contribution de la variable  $j$  sur l'axe  $G_\beta$  est mesurée par

$$C(j, \beta) = \frac{\mathbf{t}_{j\beta}^2}{\lambda_\beta}.$$

- Si l'espace final est un plan formé des axes  $G_1$  et  $G_2$ , la contribution de la variable  $j$  sur ce plan est donnée par

$$C(j, (1, 2)) = \frac{\mathbf{a}_{j1}^2 + \mathbf{a}_{j2}^2}{\lambda_1 + \lambda_2}.$$

## Exemple 3 (suite)

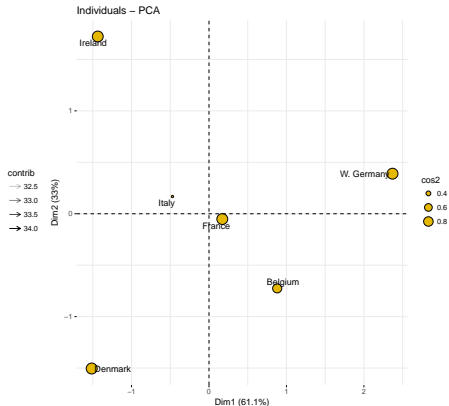
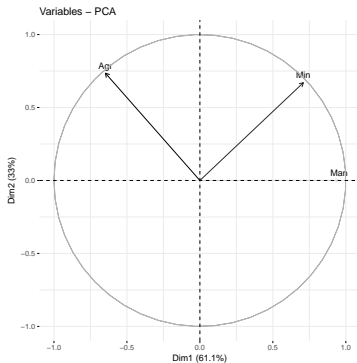
Avec les données de l'Exemple 2, on obtient (en %)

Variables

	Dim.1	Dim.2	Dim.3
Agr	22.879	54.637	22.484
Min	27.430	45.363	27.207
Man	49.691	0.000	50.309



# Interprétation globale (Exemple 3)



## 2.7 Applications

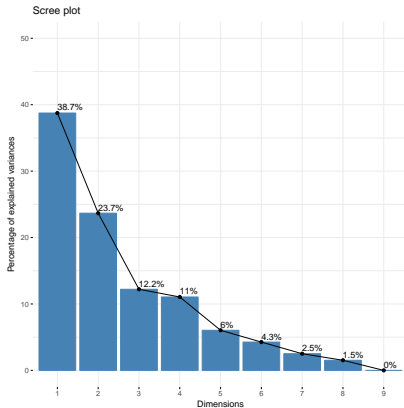
## Application 1 : *Eurojobs*

```
### Préparation des données
data <- Eurojobs
rownames(data) <- data[,1]
data <- data[,-1]

### Analyse
res.pca <- PCA(data, scale.unit = TRUE, ncp = 5,
               graph = FALSE)
summary(res.pca)

### Graphique des valeurs propres
fviz_eig(res.pca, addlabels = TRUE, ylim = c(0, 50))
```

# Application 1 : *Eurojobs*



## Application 1 : *Eurojobs*

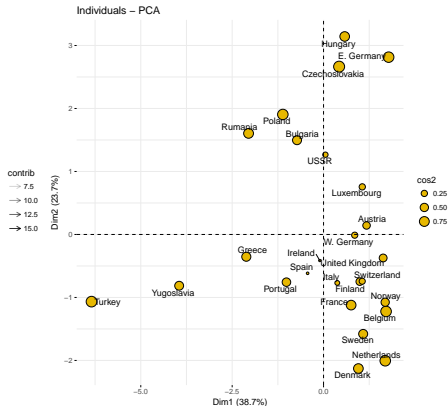
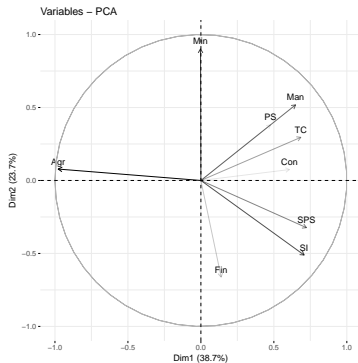
```
### Cercle des corrélations
fviz_pca_var(res.pca, col.var = "black") #Noir et blanc
fviz_pca_var(res.pca, col.var = "cos2", #Couleurs (cos2)
              gradient.cols = c("#00AFBB",
                                "#E7B800", "#FC4E07"),
              repel = TRUE))
```

## Application 1 : *Eurojobs*

```
### Carte des observations
fviz_pca_ind (res.pca, col.ind = "cos2",      #Couleurs
              gradient.cols = c("#00AFBB",
                                "#E7B800", "#FC4E07"),
              repel = TRUE)

fviz_pca_ind (res.pca, pointsize = "cos2",    #Tailles
              pointshape = 21,
              fill = "#E7B800",
              repel = TRUE)
```

# Application 1 : Eurojobs



## En bref...

- L'analyse en composantes principales (ACP) consiste en une décomposition spectrale d'une matrice particulière :
  - soit la matrice de variance-covariance ( $\Sigma$ );
  - soit la matrice de corrélation ( $\mathbf{R}$ ).
- L'objectif est de réorganiser l'information de manière hiérarchique avec l'idée que l'on veut réduire le nombre de dimensions.
- L'utilisation de la matrice de corrélation permet généralement d'améliorer le résultat de l'analyse en composantes principales en mettant toutes les variables sur le même pied.



## 2.8 ACP : outil d'apprentissage machine

# Analyse de données complexes

- L'analyse en composantes principales peut aussi être utilisée comme outil pour la compression de données complexes.
- On pourra l'utiliser pour traiter des images en niveaux de gris (tableau de taille  $n_1 \times n_2$ ), des images en couleur (tableau de taille  $n_1 \times n_2 \times 3$ ), des données de télématique (tableau de taille  $n_1 \times \dots \times n_k$ ), etc.

# Le problème

- On a des données initiales  $\mathbf{X}$  de taille<sup>1</sup> ( $k \times n$ ) que l'on souhaite simplifier, c'est-à-dire projeter dans  $\mathbb{R}^d$  avec  $d \ll k$ .
- On définit une matrice de compression  $\mathbf{W}$  de taille ( $d \times k$ ) et une matrice de restauration  $\mathbf{U}$  de taille ( $k \times d$ ).
- On pourra ainsi
  - réduire la taille des données :  $\mathbf{X}^* = \mathbf{W}\mathbf{X}$  et
  - restaurer la taille initiale des données (après une perte d'information plus ou moins importante) :  $\tilde{\mathbf{X}} = \mathbf{U}\mathbf{X}^* = \mathbf{U}\mathbf{W}\mathbf{X}$ .
- On cherche donc à résoudre le problème d'optimisation suivant :

$$\arg \min_{\mathbf{W}, \mathbf{U}} \sum_{i=1}^n \|\mathbf{x}_i - \mathbf{U}\mathbf{W}\mathbf{x}_i\|_2^2, \quad (1)$$

où les  $\mathbf{x}_i$  sont les colonnes de la matrice  $\mathbf{X}$ .

1. J'ai transposé la matrice  $\mathbf{X}$  utilisée dans les sections précédentes pour simplifier la présentation.

# Un nouveau problème ?

- On peut démontrer que si  $\mathbf{U}$  et  $\mathbf{W}$  sont des matrices solutions de l'Équation (1), alors les colonnes de  $\mathbf{U}$  sont orthonormales<sup>2</sup> et on a  $\mathbf{W} = \mathbf{U}^T$ .
- On définit la matrice  $\mathbf{A} = \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T = \mathbf{X}^T \mathbf{X}$ . On peut démontrer que

$$\mathbf{U} = [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \cdots \quad \mathbf{v}_d],$$

où  $\mathbf{v}_i$ ,  $i = 1, \dots, d$  est le vecteur propre de taille  $(k \times 1)$  correspondant à la  $i^{\text{e}}$  plus grande valeur propre obtenue par la décomposition spectrale de la matrice  $\mathbf{A}$ .

## Application 2 : images

```
install.packages("BiocManager")  
BiocManager::install("EBImage")  
library(EBImage)
```

```
### Préparation des données (100 images)  
fnames <- paste0("cat.", 1001:1100, ".jpg")  
original_dataset_dir <- "~/..."
```

```
### taille des images (200 x 200 pixels)  
n <- 200  
XX <- matrix(NA, ncol = 100, nrow = n*n)
```

## Application 2 : images

### Lecture et reformatage des images

```
img_read <- function(x){  
  f <- file.path(original_dataset_dir, fnames[x])  
  y <- resize(readImage(f), w = n, h = n)  
  XX[,x] <- matrix(imageData(getFrame(y, i = 1))[1:n, 1:n],  
                    ncol = 1)  
}
```

```
sapply(1:100, function(x) img_read(x))
```

## Application 2 : images

```
### Affichage de l'image 2
```

```
m <- Image(matrix(XX[,2], ncol = n, byrow = FALSE),  
            colormode = Grayscale)  
display(m, method="raster", all=FALSE, frame = 1)
```

## Application 2 : images





## Application 2 : images

```
### Décomposition spectrale
```

```
X <- t(XX)
```

```
spec <- eigen(t(X) %*% X)
```

```
### Décomposition spectrale (méthode accélérée)
```

```
B <- X %*% t(X)
```

```
VN <- sapply(1:100, function(x)
```

```
  t(X)%*%eigen(B)$vectors[,x]/
```

```
  norm(t(X) %*% eigen(B)$vectors[,x],
```

```
  type = "2"))
```

## Application 2 : images

```
### Compression en dimensions 100
```

```
d <- 100
```

```
U <- VN[,1:d]
```

```
W <- t(U)
```

```
XXdim <- W %*% XX
```

```
### Restauration en dimensions 10 000
```

```
XXnew <- U %*% XXdim
```

## Application 2 : images



## Application 2 : images

