3-Rapport-Secteurs-Activite

# Lecture des données

x\_matrix <- read.csv("csv/generated/year-activity-format.csv", header = T, sep = ";", row.names = 1)

## Nombre de colonne

ncol(x\_matrix)

## [1] 5

## Nombre de ligne

nrow(x\_matrix)

## [1] 30

## Affichage des 10 premières lignes (pour uniquement 2 colonnes)

x\_matrix[1:10,1:2]

## NA38.TAZ.Agriculture NA38.TBE.Industrie  
## X2018.p. 240762 3105002  
## X2017.p. 242032 3094187  
## X2016 238996 3095842  
## X2015 238764 3123779  
## X2014 235193 3159226  
## X2013 233705 3193554  
## X2012 229178 3231683  
## X2011 225602 3252758  
## X2010 221317 3270035  
## X2009 224872 3354146

# Informations basiques

## Résumé (pour uniquement 2 colonnes)

summary(x\_matrix[,1:2])

## NA38.TAZ.Agriculture NA38.TBE.Industrie  
## Min. :212438 Min. :3094187   
## 1st Qu.:228630 1st Qu.:3257077   
## Median :238880 Median :3832804   
## Mean :243188 Mean :3733275   
## 3rd Qu.:258187 3rd Qu.:4048762   
## Max. :284929 Max. :4549924

## Covariance (pour uniquement 2 colonnes)

cov(x\_matrix[,1:2])

## NA38.TAZ.Agriculture NA38.TBE.Industrie  
## NA38.TAZ.Agriculture 357604300 1108722275  
## NA38.TBE.Industrie 1108722275 209417229113

## Variance (pour uniquement 2 colonnes)

var(x\_matrix[,1:2]);

## NA38.TAZ.Agriculture NA38.TBE.Industrie  
## NA38.TAZ.Agriculture 357604300 1108722275  
## NA38.TBE.Industrie 1108722275 209417229113

## Corrélation (pour uniquement 2 colonnes)

cor(x\_matrix[,1:2])

## NA38.TAZ.Agriculture NA38.TBE.Industrie  
## NA38.TAZ.Agriculture 1.0000000 0.1281195  
## NA38.TBE.Industrie 0.1281195 1.0000000

# Données centrées réduites

centree\_reduite <- scale(x\_matrix, center = T, scale = T);

summary(centree\_reduite[,1:2])

## NA38.TAZ.Agriculture NA38.TBE.Industrie  
## Min. :-1.6261 Min. :-1.3965   
## 1st Qu.:-0.7698 1st Qu.:-1.0406   
## Median :-0.2278 Median : 0.2175   
## Mean : 0.0000 Mean : 0.0000   
## 3rd Qu.: 0.7932 3rd Qu.: 0.6894   
## Max. : 2.2073 Max. : 1.7846

## Covariance (pour uniquement 2 colonnes)

cov(centree\_reduite[,1:2])

## NA38.TAZ.Agriculture NA38.TBE.Industrie  
## NA38.TAZ.Agriculture 1.0000000 0.1281195  
## NA38.TBE.Industrie 0.1281195 1.0000000

## Variance (pour uniquement 2 colonnes)

var(centree\_reduite[,1:2]);

## NA38.TAZ.Agriculture NA38.TBE.Industrie  
## NA38.TAZ.Agriculture 1.0000000 0.1281195  
## NA38.TBE.Industrie 0.1281195 1.0000000

## Corrélation (pour uniquement 2 colonnes)

cor(centree\_reduite[,1:2])

## NA38.TAZ.Agriculture NA38.TBE.Industrie  
## NA38.TAZ.Agriculture 1.0000000 0.1281195  
## NA38.TBE.Industrie 0.1281195 1.0000000

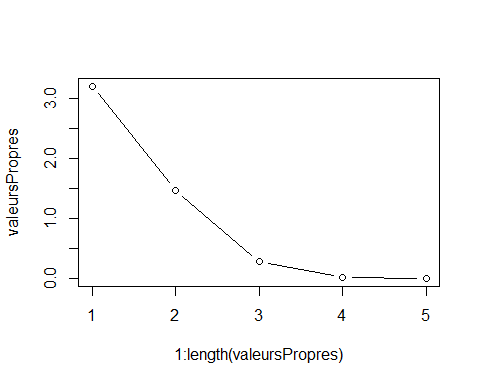
# Analyse en composante principale

## Valeurs propres

propres <- eigen(cor(centree\_reduite));  
valeursPropres <- propres$values;  
vecteursPropres <- propres$vectors;

## Graphique des valeurs propres (éboulis et coude)

plot(1:length(valeursPropres), valeursPropres, type = "b");



## Composantes principales

data\_acp <- centree\_reduite %\*% vecteursPropres;  
composante\_principale\_1 <- data\_acp[, 1];  
composante\_principale\_2 <- data\_acp[, 2];  
totalInfo <- sum(valeursPropres, na.rm = FALSE);  
qte <- (valeursPropres[1] + valeursPropres[2]) / totalInfo;  
message("Quantité d'information avec deux composantes : ", toString(qte \* 100),"%");

## Quantité d'information avec deux composantes : 93.5145146785711%

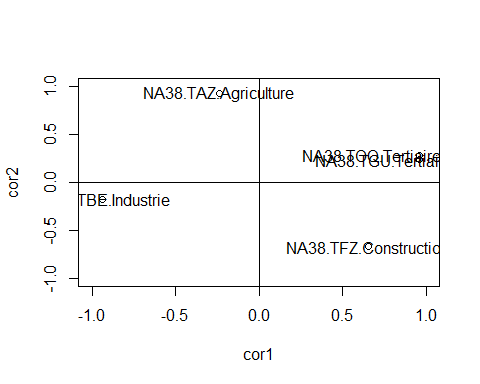
troisComposantes <- FALSE;  
if (qte < 0.8) {  
 composante\_principale\_3 <- data\_acp[, 3]  
 qte <- qte + valeursPropres[3];  
 message("Ajout d'une troisième composante pour améliorer la quantité d'information : ", toString(qte))  
 troisComposantes <- TRUE;  
} else {  
 message("On ne séléctionne que les deux première composantes principales car elles contiennent à elles seules plus de 80% des informations");  
 troisComposantes <- FALSE;  
}

## On ne séléctionne que les deux première composantes principales car elles contiennent à elles seules plus de 80% des informations

## Cercle de corrélation

Calcule de la correlation entre chaque variable et les composantes principales

cor1 <- cor(composante\_principale\_1,centree\_reduite);  
cor2 <- cor(composante\_principale\_2,centree\_reduite);  
  
# Correlation 1 - 2  
plot(cor1, cor2, xlim = c(-1, +1), ylim = c(-1, +1))  
abline(h = 0, v = 0)  
text(cor1, cor2, labels = colnames(x\_matrix))



if(troisComposantes){  
 cor3 <- cor(composante\_principale\_3, centree\_reduite);  
 # Correlation 3 - 1  
 plot(cor1, cor2, xlim = c(-1, +1), ylim = c(-1, +1))  
 abline(h = 0, v = 0)  
 text(cor1, cor2, labels = colnames(x\_matrix))  
  
 # Correlation 3 - 2  
 plot(cor1, cor2, xlim = c(-1, +1), ylim = c(-1, +1))  
 abline(h = 0, v = 0)  
 text(cor1, cor2, labels = colnames(x\_matrix))  
}

Dans un premier temps on peut dire que la plupart des variable sont proches du cercle et ainsi bien représentées par l’ACP.

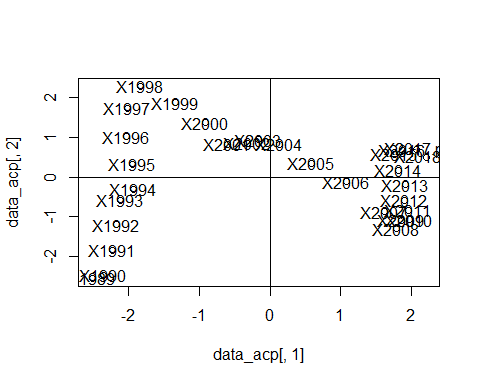
On observe que les secteur de l’industrie sont fortement négativement corrélés avec la composante principale 1 tandis que que les deux secteurs du tertiaire lui sont fortement corrélés. De plus, le secteur de la construction lui est également assez correlé, alors que le secteur de l’agriculture lui est légèrement négativement correlé.

On peut donc interpréter la composante principale 1 comme le fait d’être un secteur qui fournit un service plutôt qu’un secteur qui manufacture des produits.

Malheureseument, nous n’avons pas pu déterminer d’interprétation satisfaisante pour la composante principales 2.

## Graphe 2D

# Graphe 1 - 2  
plot(data\_acp[, 1], data\_acp[, 2])  
text(data\_acp[, 1], data\_acp[, 2], labels = rownames(data\_acp))  
abline(h = 0, v = 0)



if(troisComposantes){  
 # Graphe 3 - 1  
 plot(data\_acp[, 1], data\_acp[, 3])  
 text(data\_acp[, 1], data\_acp[, 3], labels = rownames(data\_acp))  
 abline(h = 0, v = 0)  
  
 # Graphe 3 - 2  
 plot(data\_acp[, 3], data\_acp[, 2])  
 text(data\_acp[, 3], data\_acp[, 2], labels = rownames(data\_acp))  
 abline(h = 0, v = 0)  
  
}

On observe que plus le temps passe, plus les valeurs de la composante principale 1 sont élevées. On peut donc en conclure que l’emploi dans les secteurs d’activités lié à la production industrielle à baissé entre les années 1989 et 2018 au profit des secteurs de service.