3-Rapport-Reference

# Lecture des données

x\_matrix <- read.csv("petrole.csv", header = T, sep = ";", row.names = 1)

## Nombre de colonne

ncol(x\_matrix)

## [1] 8

## Nombre de ligne

nrow(x\_matrix)

## [1] 16

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

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

## NET INT  
## 1969 17.93 3.96  
## 1970 16.21 3.93  
## 1971 19.01 3.56  
## 1972 18.05 3.33  
## 1973 16.56 3.10  
## 1974 13.09 2.64  
## 1975 13.43 2.42  
## 1976 9.83 2.46  
## 1977 9.46 2.33  
## 1978 10.93 2.95

# Informations basiques

## Résumé (pour uniquement 2 colonnes)

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

## NET INT   
## Min. : 9.46 Min. :2.330   
## 1st Qu.:12.38 1st Qu.:2.715   
## Median :13.23 Median :3.075   
## Mean :13.85 Mean :3.135   
## 3rd Qu.:16.30 3rd Qu.:3.570   
## Max. :19.01 Max. :3.960

## Covariance (pour uniquement 2 colonnes)

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

## NET INT  
## NET 8.423612 1.05828  
## INT 1.058280 0.28244

## Variance (pour uniquement 2 colonnes)

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

## NET INT  
## NET 8.423612 1.05828  
## INT 1.058280 0.28244

## Corrélation (pour uniquement 2 colonnes)

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

## NET INT  
## NET 1.0000000 0.6861014  
## INT 0.6861014 1.0000000

# Données centrées réduites

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

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

## NET INT   
## Min. :-1.5139 Min. :-1.5147   
## 1st Qu.:-0.5078 1st Qu.:-0.7903   
## Median :-0.2149 Median :-0.1129   
## Mean : 0.0000 Mean : 0.0000   
## 3rd Qu.: 0.8420 3rd Qu.: 0.8185   
## Max. : 1.7766 Max. : 1.5524

## Covariance (pour uniquement 2 colonnes)

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

## NET INT  
## NET 1.0000000 0.6861014  
## INT 0.6861014 1.0000000

## Variance (pour uniquement 2 colonnes)

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

## NET INT  
## NET 1.0000000 0.6861014  
## INT 0.6861014 1.0000000

## Corrélation (pour uniquement 2 colonnes)

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

## NET INT  
## NET 1.0000000 0.6861014  
## INT 0.6861014 1.0000000

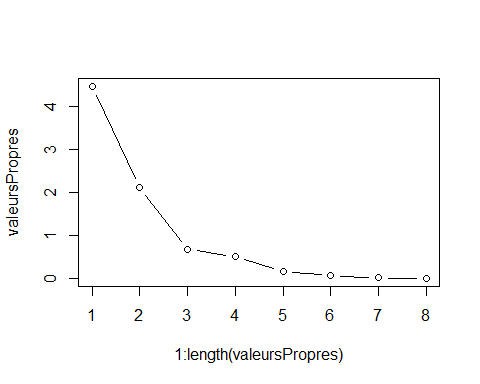
# Analyse en composante principale

## 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 : 82.3152157500261%

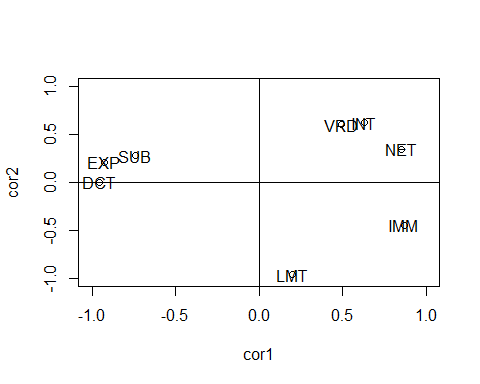
troisComposantes <- FALSE;  
if (qte < 0.8) {  
 composante\_principale\_3 <- data\_acp[, 3]  
 qte <- qte + valeursPropres[3];  
 message("Ajout d'une troisième composante pour améliorer la quantité d'information : ", toString(qte))  
 troisComposantes <- TRUE;  
} else {  
 message("On ne séléctionne 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 Corrélation entre chaque variable et les composantes principales

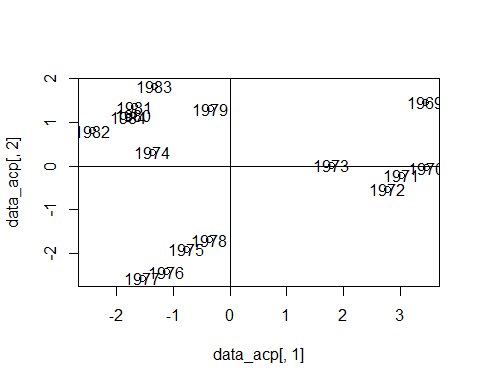
cor1 <- cor(composante\_principale\_1,centree\_reduite);  
cor2 <- cor(composante\_principale\_2,centree\_reduite);  
  
# Corrélation 1 - 2  
plot(cor1, cor2, xlim = c(-1, +1), ylim = c(-1, +1))  
abline(h = 0, v = 0)  
text(cor1, cor2, labels = colnames(x\_matrix))



if(troisComposantes){  
 cor3 <- cor(composante\_principale\_3, centree\_reduite);  
 # Corrélation 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))  
  
 # Corrélation 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))  
}

## 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)  
  
}