3-Rapport-FemmeHomme

# Introduction

Cette analyse en composante principale se penche sur l’employabilité des femmes et des hommes en France métropolitaine.

# Lecture des données

x\_matrix <- read.csv("CSV/generated/f&h-t-format.csv", header = T, sep = ";", row.names = 1)

## Nombre de colonne

ncol(x\_matrix)

## [1] 2

## Nombre de ligne

nrow(x\_matrix)

## [1] 30

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

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

## hommes femmes  
## X2018.p. 12478958 12130012  
## X2017.p. 12376710 12082791  
## X2016 12184011 11956942  
## X2015 12059880 11887246  
## X2014 12018119 11835683  
## X2013 12053882 11796718  
## X2012 12044639 11732477  
## X2011 12095423 11750655  
## X2010 12071802 11713497  
## X2009 12042035 11699139

# Informations basiques

## Résumé (pour uniquement 2 colonnes)

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

## hommes femmes   
## Min. :10873025 Min. : 8825899   
## 1st Qu.:11266319 1st Qu.: 9703512   
## Median :12056881 Median :11054322   
## Mean :11855366 Mean :10779178   
## 3rd Qu.:12245559 3rd Qu.:11734159   
## Max. :12478958 Max. :12130012

## Covariance (pour uniquement 2 colonnes)

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

## hommes femmes  
## hommes 273807190415 5.075695e+11  
## femmes 507569518761 1.211118e+12

## Variance (pour uniquement 2 colonnes)

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

## hommes femmes  
## hommes 273807190415 5.075695e+11  
## femmes 507569518761 1.211118e+12

## Corrélation (pour uniquement 2 colonnes)

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

## hommes femmes  
## hommes 1.000000 0.881414  
## femmes 0.881414 1.000000

# Données centrées réduites

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

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

## hommes femmes   
## Min. :-1.8773 Min. :-1.7749   
## 1st Qu.:-1.1257 1st Qu.:-0.9774   
## Median : 0.3851 Median : 0.2500   
## Mean : 0.0000 Mean : 0.0000   
## 3rd Qu.: 0.7457 3rd Qu.: 0.8678   
## Max. : 1.1917 Max. : 1.2275

## Covariance (pour uniquement 2 colonnes)

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

## hommes femmes  
## hommes 1.000000 0.881414  
## femmes 0.881414 1.000000

## Variance (pour uniquement 2 colonnes)

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

## hommes femmes  
## hommes 1.000000 0.881414  
## femmes 0.881414 1.000000

## Corrélation (pour uniquement 2 colonnes)

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

## hommes femmes  
## hommes 1.000000 0.881414  
## femmes 0.881414 1.000000

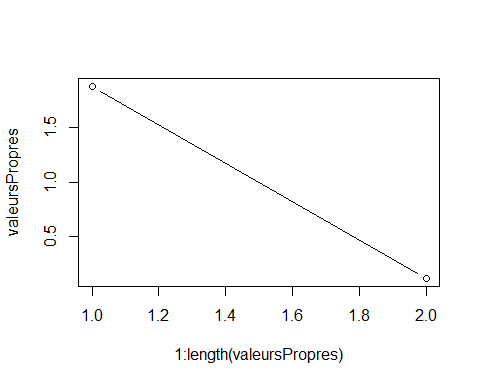
# Analyse en composante principale

## 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 : 100%

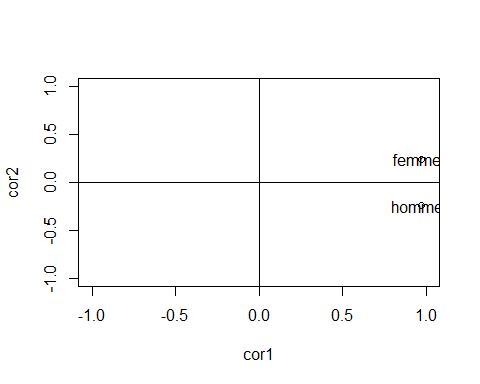
troisComposantes <- FALSE;  
if (qte < 0.8) {  
 composante\_principale\_3 <- data\_acp[, 3]  
 qte <- qte + valeursPropres[3];  
 message("Ajout d'une troisième composante pour améliorer la quantité d'information : ", toString(qte))  
 troisComposantes <- TRUE;  
} else {  
 message("On ne séléctionne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

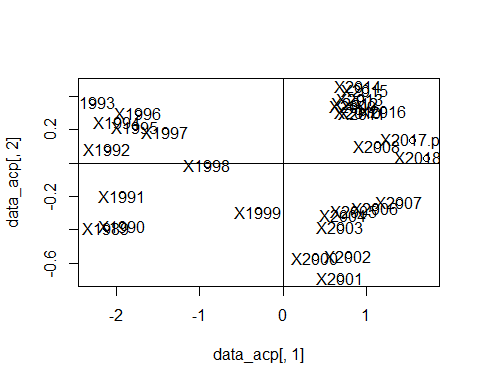
Dans un premier temps on peut dire que la plupart des variable sont proches du cercle et ainsi bien représentées par l’ACP. On observe que nos deux variables sont fortement corrélées à notre première composante principale. Cependant, on peut également constater que la variable *homme* est négativement corrélée à la seconde composante principale, et que la variable *femme* est quand à elle légerement corrélée à cette dernière.

Étant donné que la quantité d’information portée par la première composante principale est bien supérieure à celle portée par la seconde, on peut dire que la première composante principale peut être analysée comme **La quantité de personnes employés cette année** De plus, la seconde composante principale peut être analysée comme **La quantité de femmes employées cette année**.

Comme on pouvait s’y attendre, étant donnée la faible quantité de variables dans cette analyse, cela ne nous a pas permis de synthétiser des variables.

## 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)  
  
}

L’observation de ce graphique du nuage des individus, nous permet de déterminer quatres groupes d’années : - Le groupe 1 : Les années 1992, 1993, 1994, 1995, 1996, 1997

Ce groupe présente des valeurs relativement élevées dans la composante principale 2 et des valeurs plus faibles dans la composante principale 1. On peut donc dire que ce sont les années où ont été employés moins de personnes, et un peu plus de femmes.

* Le groupe 2 : Les années 1989, 1990, 1991

Ce groupe présente des valeurs relativement faibles dans les deux composantes principales. On peut donc dire que durant ces années peu de personnes ont été employées et également peu de femmes.

* Le groupe 3 : Les années 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007

Ce groupe présente des valeurs très faibles dans la composante principale 2 et des valeurs relativement élevées dans la composante principales 1. On peut donc dire que ces années sont celles ou ont été employés le moins de femmes mais beaucoup de personnes.

* le groupe 4 : Les années 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018

Ce groupe présente des valeurs très élevées dans les deux composantes principales. On peut donc dire que ce sont les années où ont été employés le plus grand nombre de personnes, et dans une moindre mesure les années où ont été employées le plus de femmes.