

Analyse des correspondances (AFC) avec FactoMineR sur le jeu de données du didacticiel (naissances)

François Husson

Importation du jeu de données

```
## Père - 20 ans    Père 20 à 24 ans Père 25 à 29 ans Père 30 à 34 ans
## Min.   : 0.0    Min.   : 0      Min.   : 2      Min.   : 8
## 1st Qu.: 0.0    1st Qu.: 4      1st Qu.: 29     1st Qu.: 147
## Median : 17.0   Median : 394    Median : 1771   Median : 2685
## Mean   : 349.3  Mean   : 5523   Mean   : 20108  Mean   : 29378
## 3rd Qu.: 133.0  3rd Qu.: 7967   3rd Qu.: 21921  3rd Qu.: 20674
## Max.   :2085.0  Max.   :30070   Max.   :104614  Max.   :131262
## Père 35 à 39 ans Père 40 à 44 ans Père 45 à 49 ans Père 50 à 54 ans
## Min.   : 9      Min.   : 16     Min.   : 24     Min.   : 10
## 1st Qu.: 266    1st Qu.: 542    1st Qu.: 587    1st Qu.: 192
## Median : 4369   Median : 1631   Median : 884    Median : 415
## Mean   :20237   Mean   : 9704   Mean   : 3425   Mean   :1161
## 3rd Qu.:26259   3rd Qu.:16046   3rd Qu.: 7296   3rd Qu.:2340
## Max.   :82027   Max.   :35506   Max.   :10681   Max.   :3571
## Père 55 à 59 ans Père + 60 ans    Père + 55
## Min.   : 6.0    Min.   : 1.0    Min.   : 7.0
## 1st Qu.: 80.0    1st Qu.: 33.0    1st Qu.: 113.0
## Median : 172.0   Median : 53.0    Median : 225.0
## Mean   : 404.8   Mean   :159.3    Mean   : 564.1
## 3rd Qu.: 795.0   3rd Qu.:340.0    3rd Qu.:1135.0
## Max.   :1196.0   Max.   :479.0    Max.   :1675.0
```

```
setwd("C:/Users/husson/Documents/")
Naissance <- read.table("AnaDo_JeuDonnees_Naissances.csv",
  header=TRUE, sep=";", row.names=1, check.names=FALSE, fileEncoding="latin1")
summary(Naissance)
```

`header=TRUE` : précise que le nom des variables est présent

`sep=";"` : précise que le séparateur de colonnes est le point-virgule (fréquent dans les fichiers csv, pour une tabulation il faudrait écrire `sep="\t"`)

`row.names=1` : précise que le nom des individus est dans la première colonne du tableau

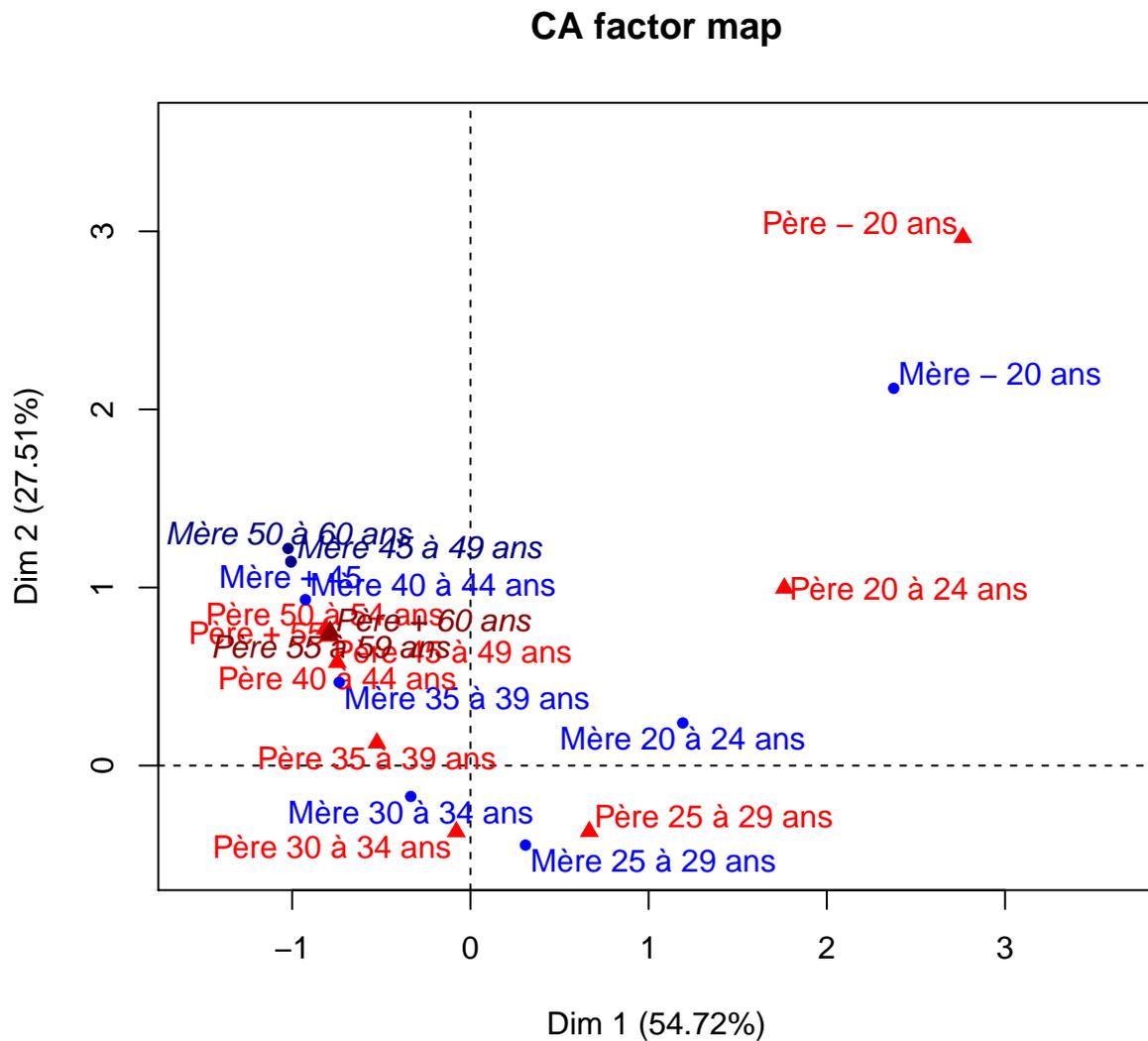
`check.names=FALSE` : impose que le nom des colonnes soit pris tel que dans le fichier (sinon les espaces sont remplacés par des points et des X sont mis avant les nombres)

Chargement de FactoMineR

```
library(FactoMineR)
```

L'AFC et le graphe simultanée par défaut

```
res <- CA(Naissance, row.sup=7:8, col.sup=9:10)
```



Résumé des principaux résultats

```
summary(res)
```

```
##
## Call:
## CA(X = Naissance, row.sup = 7:8, col.sup = 9:10)
##
## The chi square of independence between the two variables is equal to 708295.8 (p-value = 0 ).
##
## Eigenvalues
##              Dim.1   Dim.2   Dim.3   Dim.4   Dim.5   Dim.6
## Variance          0.478   0.240   0.105   0.032   0.014   0.004
## % of var.         54.724  27.510  11.988   3.700   1.626   0.452
## Cumulative % of var. 54.724  82.234  94.222  97.922  99.548 100.000
##
## Rows
##              Iner*1000   Dim.1   ctr   cos2   Dim.2   ctr
## Mère - 20 ans |    209.626 |   2.375  21.138  0.482 |   2.118  33.449
## Mère 20 à 24 ans |   199.127 |   1.192  36.604  0.878 |   0.239   2.933
## Mère 25 à 29 ans |   104.178 |   0.309   6.095  0.279 |  -0.446  25.286
## Mère 30 à 34 ans |    85.815 |  -0.335   7.841  0.436 |  -0.173   4.187
## Mère 35 à 39 ans |   140.420 |  -0.736  19.228  0.654 |   0.468  15.443
## Mère 40 à 44 ans |   113.997 |  -0.926   8.431  0.353 |   0.932  16.990
## Mère + 45      |    19.649 |  -1.008   0.664  0.162 |   1.147   1.712
##              cos2   Dim.3   ctr   cos2
## Mère - 20 ans  0.383 |  -1.136  22.053  0.110 |
## Mère 20 à 24 ans 0.035 |   0.254   7.574  0.040 |
## Mère 25 à 29 ans 0.583 |   0.170   8.420  0.085 |
## Mère 30 à 34 ans 0.117 |  -0.325  33.769  0.412 |
## Mère 35 à 39 ans 0.264 |   0.048   0.377  0.003 |
## Mère 40 à 44 ans 0.358 |   0.729  23.867  0.219 |
## Mère + 45      0.209 |   1.148   3.939  0.210 |
##
## Columns
##              Iner*1000   Dim.1   ctr   cos2   Dim.2   ctr
## Père - 20 ans |    93.775 |   2.763   6.193  0.315 |   2.965  14.186
## Père 20 à 24 ans |   257.192 |   1.762  39.802  0.739 |   0.997  25.345
## Père 25 à 29 ans |   162.234 |   0.667  20.771  0.612 |  -0.370  12.726
## Père 30 à 34 ans |    72.545 |  -0.079   0.424  0.028 |  -0.372  18.773
## Père 35 à 39 ans |    87.325 |  -0.525  12.916  0.706 |   0.127   1.508
## Père 40 à 44 ans |   106.558 |  -0.747  12.485  0.560 |   0.579  14.919
## Père 45 à 49 ans |    58.708 |  -0.797   4.911  0.400 |   0.727   8.121
## Père 50 à 54 ans |    23.165 |  -0.815   1.720  0.355 |   0.766   3.026
## Père + 55      |    11.307 |  -0.788   0.777  0.328 |   0.749   1.396
##              cos2   Dim.3   ctr   cos2
## Père - 20 ans  0.363 |  -2.099  16.320  0.182 |
## Père 20 à 24 ans 0.237 |  -0.150   1.324  0.005 |
## Père 25 à 29 ans 0.188 |   0.370  29.096  0.188 |
## Père 30 à 34 ans 0.621 |  -0.241  18.075  0.261 |
## Père 35 à 39 ans 0.041 |  -0.216   9.973  0.119 |
## Père 40 à 44 ans 0.336 |   0.287   8.394  0.082 |
```

```

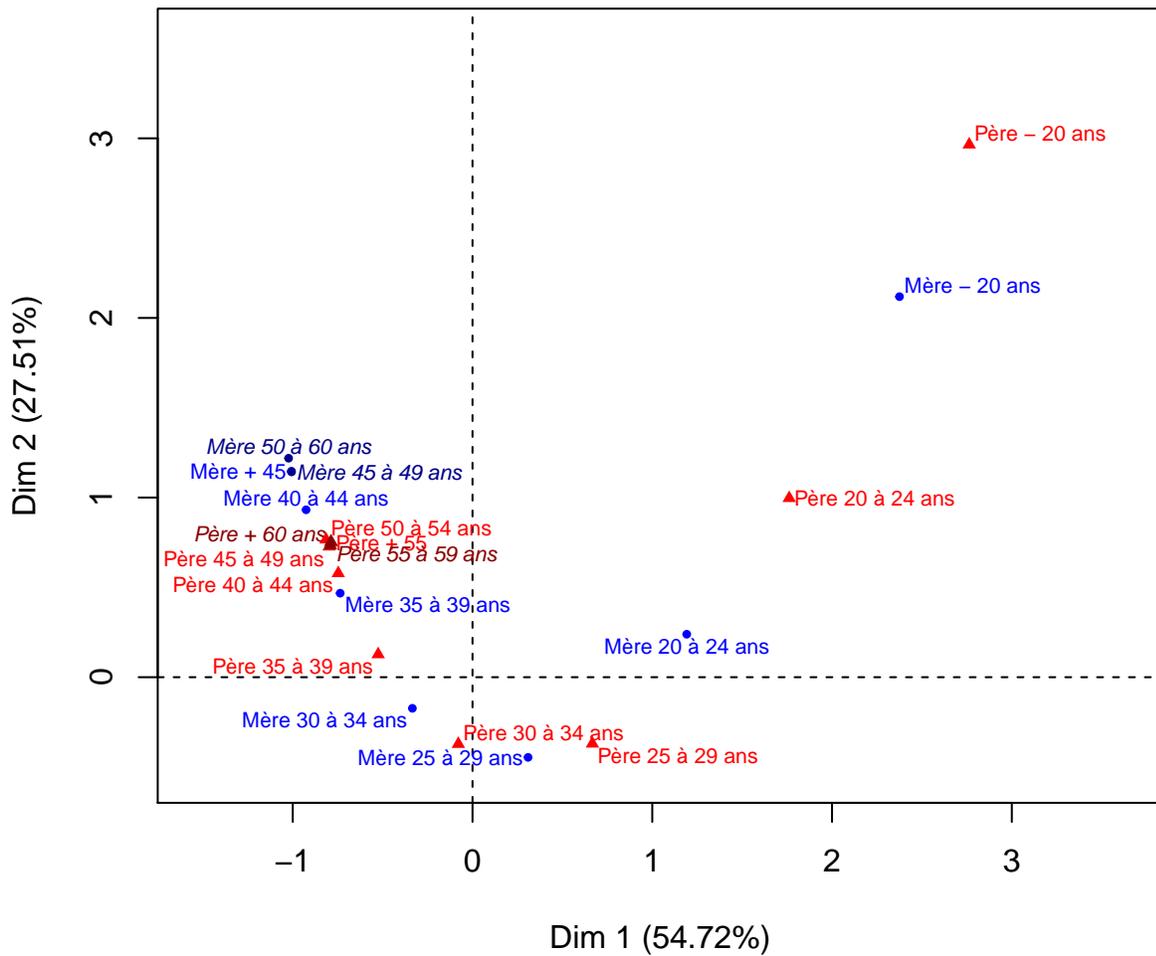
## Père 45 à 49 ans  0.332 |  0.544 10.446  0.186 |
## Père 50 à 54 ans  0.314 |  0.602  4.281  0.193 |
## Père + 55         0.296 |  0.605  2.091  0.193 |
##
## Supplementary rows
##           Dim.1  cos2  Dim.2  cos2  Dim.3  cos2
## Mère 45 à 49 ans | -1.007 0.167 |  1.143 0.216 |  1.137 0.213 |
## Mère 50 à 60 ans | -1.023 0.061 |  1.219 0.087 |  1.356 0.108 |
##
## Supplementary columns
##           Dim.1  cos2  Dim.2  cos2  Dim.3  cos2
## Père 55 à 59 ans | -0.787 0.314 |  0.755 0.288 |  0.620 0.195 |
## Père + 60 ans   | -0.791 0.366 |  0.734 0.315 |  0.567 0.188 |

```

Graphe de la représentation simultanée avec un titre et une taille de police plus petite : les libellés se chevauchent moins, le graphique est plus lisible

```
plot(res, cex=0.7, title="Représentation simultanée")
```

Représentation simultanée



Description des dimensions

```
dimdesc(res)
```

```
## $`Dim 1`  
## $`Dim 1`$row  
##          coord  
## Mère 50 à 60 ans -1.0230623  
## Mère + 45        -1.0075217  
## Mère 45 à 49 ans -1.0066959  
## Mère 40 à 44 ans -0.9258528  
## Mère 35 à 39 ans -0.7359407  
## Mère 30 à 34 ans -0.3345237  
## Mère 25 à 29 ans  0.3091183  
## Mère 20 à 24 ans  1.1916963  
## Mère - 20 ans    2.3752817  
##  
## $`Dim 1`$col  
##          coord  
## Père 50 à 54 ans -0.81497107  
## Père 45 à 49 ans -0.79730553  
## Père + 60 ans   -0.79075017  
## Père + 55       -0.78809544  
## Père 55 à 59 ans -0.78703920  
## Père 40 à 44 ans -0.74677039  
## Père 35 à 39 ans -0.52468621  
## Père 30 à 34 ans -0.07889759  
## Père 25 à 29 ans  0.66703803  
## Père 20 à 24 ans  1.76185986  
## Père - 20 ans    2.76326670  
##  
##  
## $`Dim 2`  
## $`Dim 2`$row  
##          coord  
## Mère 25 à 29 ans -0.4464193  
## Mère 30 à 34 ans -0.1733188  
## Mère 20 à 24 ans  0.2391568  
## Mère 35 à 39 ans  0.4676267  
## Mère 40 à 44 ans  0.9318948  
## Mère 45 à 49 ans  1.1429222  
## Mère + 45         1.1467692  
## Mère 50 à 60 ans  1.2191707  
## Mère - 20 ans    2.1184937  
##  
## $`Dim 2`$col  
##          coord  
## Père 30 à 34 ans -0.3720552  
## Père 25 à 29 ans -0.3701847  
## Père 35 à 39 ans  0.1271161  
## Père 40 à 44 ans  0.5787896  
## Père 45 à 49 ans  0.7269325
```

```

## Père + 60 ans      0.7338045
## Père + 55         0.7487134
## Père 55 à 59 ans  0.7546451
## Père 50 à 54 ans  0.7664040
## Père 20 à 24 ans  0.9968244
## Père - 20 ans     2.9651350
##
##
## $`Dim 3`
## $`Dim 3`$row
##                                coord
## Mère - 20 ans      -1.13553529
## Mère 30 à 34 ans  -0.32492252
## Mère 35 à 39 ans   0.04824263
## Mère 25 à 29 ans   0.17005598
## Mère 20 à 24 ans   0.25371576
## Mère 40 à 44 ans   0.72911863
## Mère 45 à 49 ans   1.13712215
## Mère + 45          1.14815637
## Mère 50 à 60 ans   1.35582383
##
## $`Dim 3`$col
##                                coord
## Père - 20 ans      -2.0993859
## Père 30 à 34 ans  -0.2409949
## Père 35 à 39 ans  -0.2157827
## Père 20 à 24 ans  -0.1504214
## Père 40 à 44 ans   0.2865959
## Père 25 à 29 ans   0.3695078
## Père 45 à 49 ans   0.5442414
## Père + 60 ans      0.5669442
## Père 50 à 54 ans   0.6017674
## Père + 55          0.6049307
## Père 55 à 59 ans   0.6200443

```

Graphes avec différentes sélections de lignes et/ou colonnes

On peut sélectionner certaines lignes ou certaines colonnes pour améliorer la lisibilité des graphes. Pour cela, deux arguments sont utilisés, un argument pour les lignes `selectRow`, l'autre pour les colonnes `selectCol`. Ces deux arguments fonctionnent selon le même principe. On décrit ici la sélection sur les lignes uniquement.

`selectRow = "cos2 0.7"` : sélectionne les lignes actives et supplémentaires qui ont, sur le plan tracé, une qualité de projection supérieure à 0.7

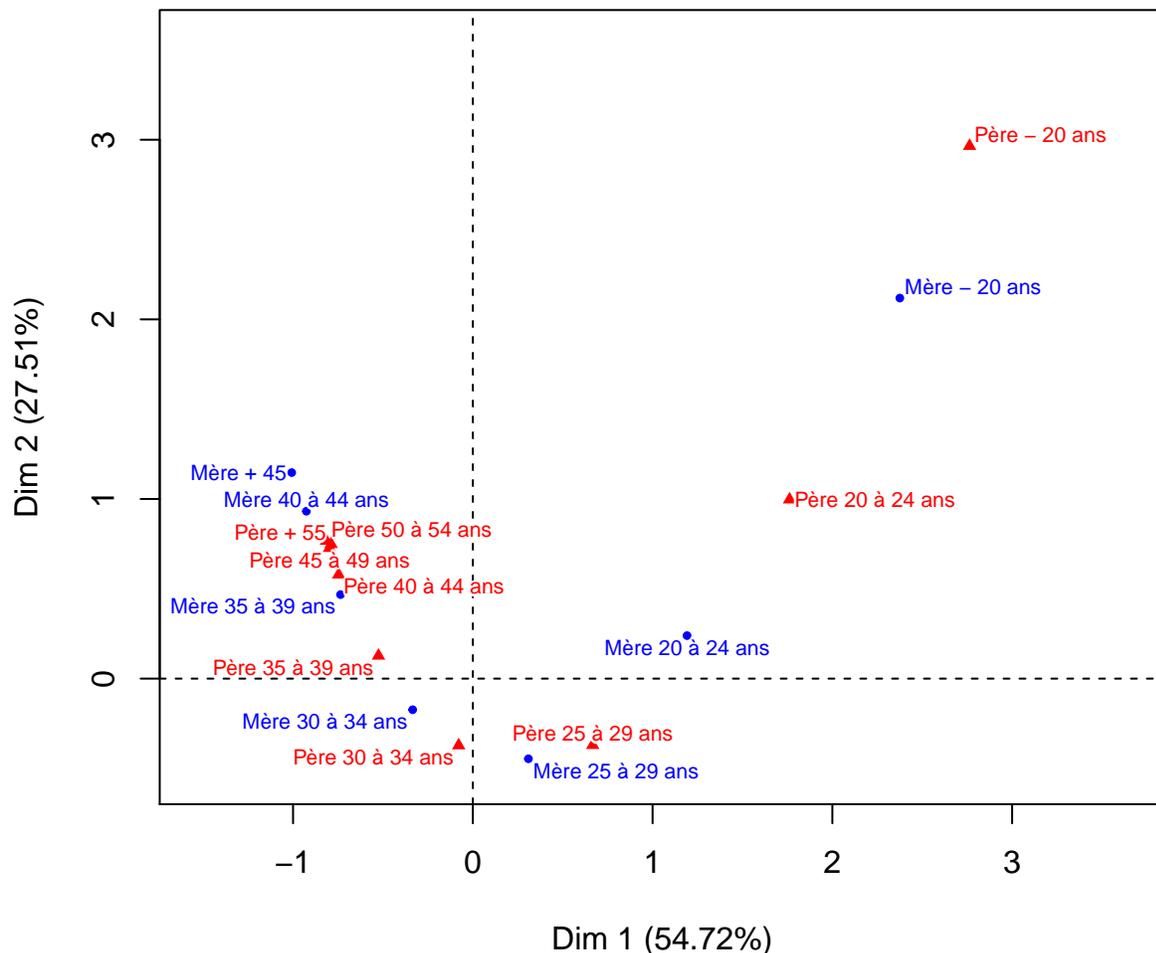
`selectRow = "cos2 5"` : sélectionne les 5 lignes actives et les 5 lignes supplémentaires qui ont la meilleure qualité de projection sur le plan tracé

`selectRow = "contrib 5"` : sélectionne les 5 lignes (nécessairement actives) qui ont le plus contribué à la construction du plan tracé

`selectRow = c("nom1", "nom2")` : sélectionne les lignes par leur nom

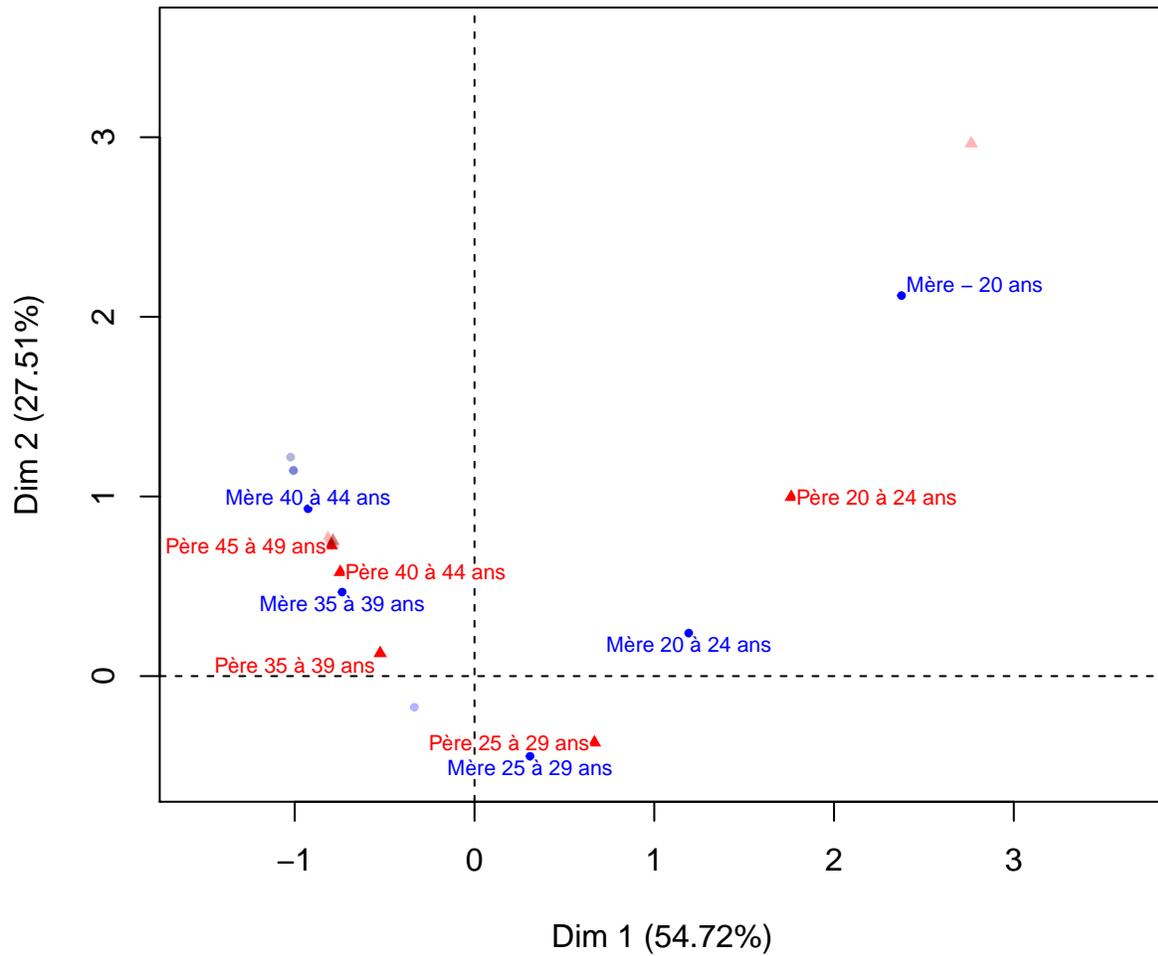
```
plot(res, invisible=c("row.sup", "col.sup"), cex=0.7, shadow=TRUE, title="Eléments actifs")
```

Eléments actifs



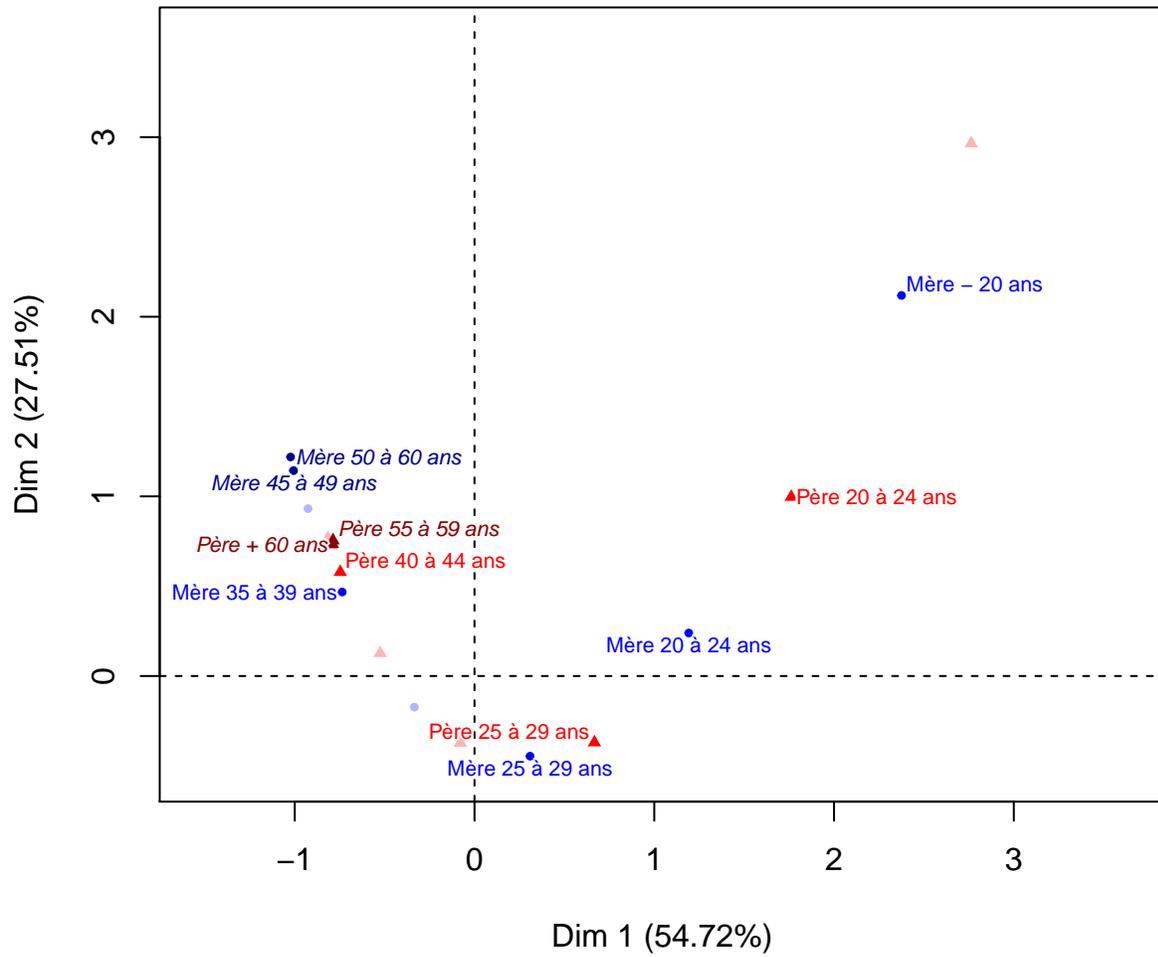
```
plot(res, shadow=TRUE, cex=0.7, selectRow="cos2 0.7", selectCol="cos2 0.7")
```

CA factor map



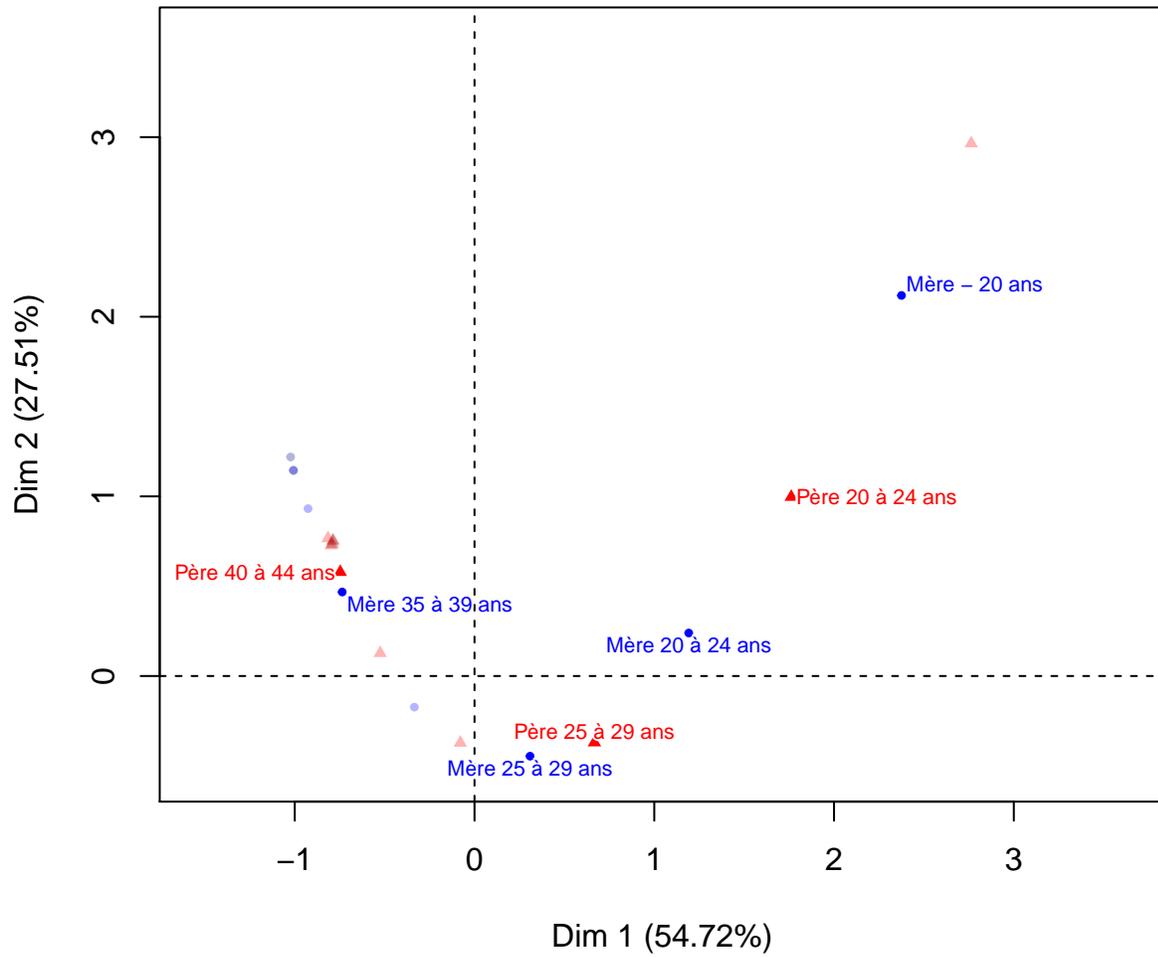
```
plot(res, shadow=TRUE, cex=0.7, selectRow="cos2 4", selectCol="cos2 3")
```

CA factor map



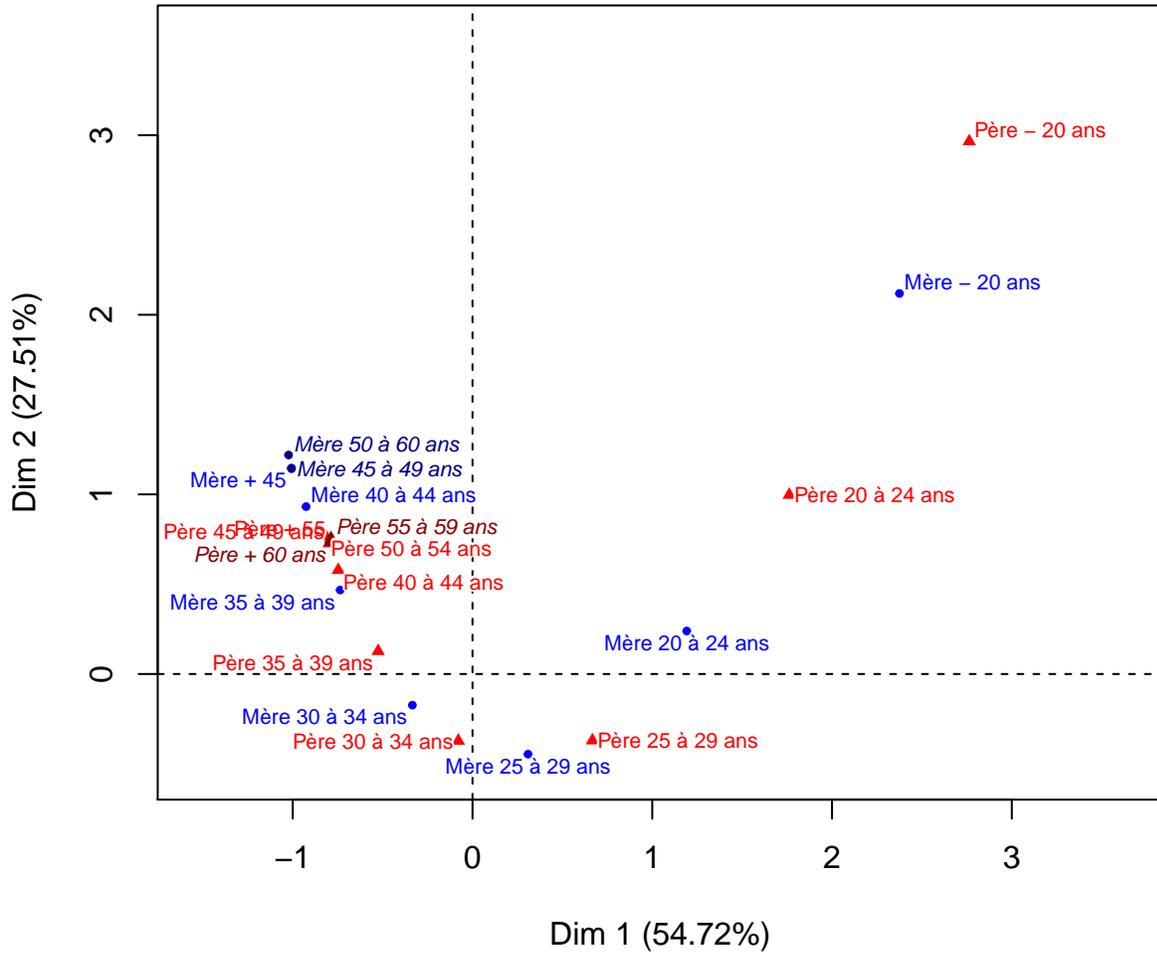
```
plot(res, shadow=TRUE, cex=0.7, selectRow="contrib 4", selectCol="contrib 3")
```

CA factor map



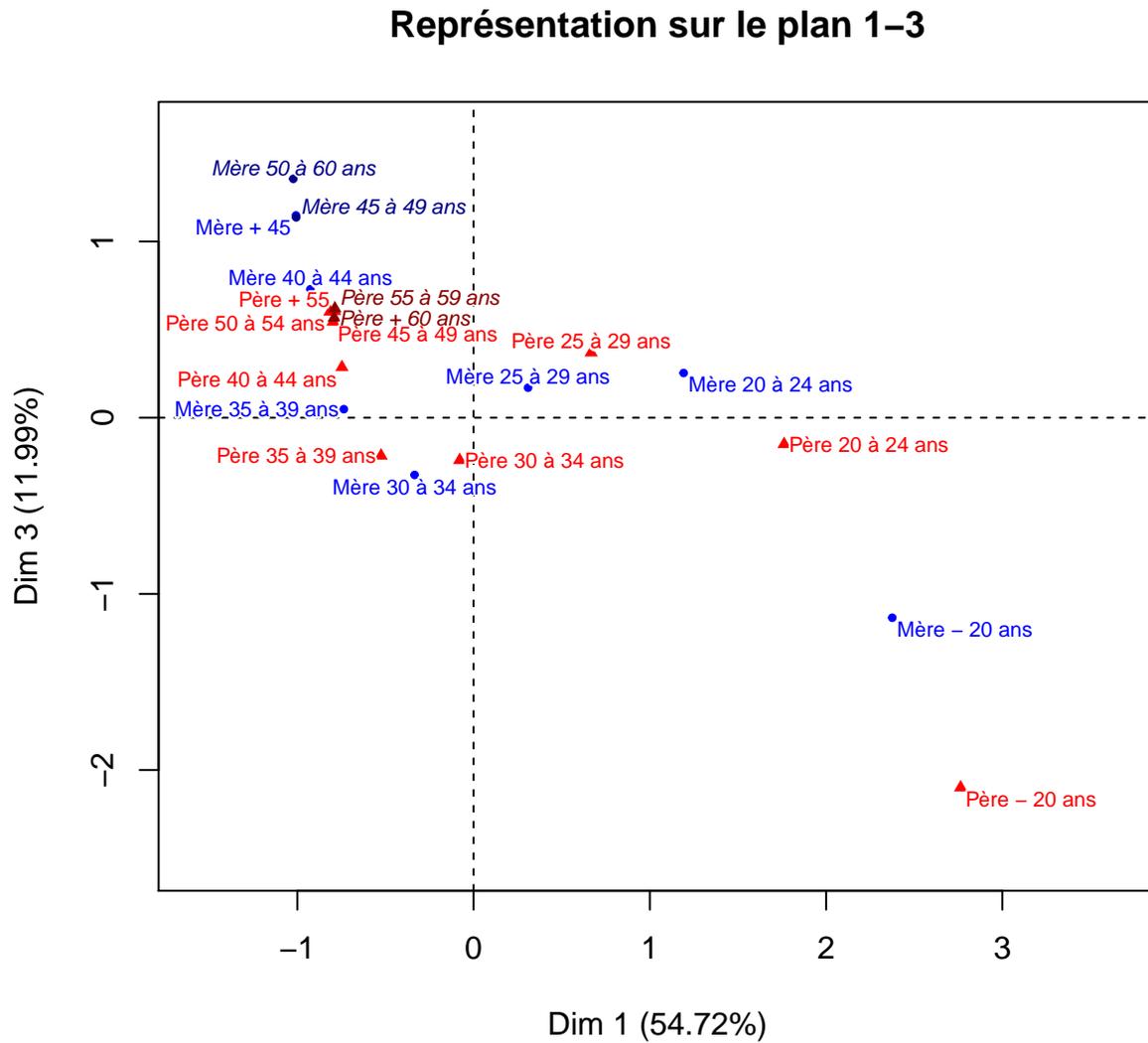
```
plot(res, shadow=TRUE, cex=0.7, selectRow=c("M?re - 20 ans"))
```

CA factor map



Graphes sur les dimensions 1 et 3

```
plot(res, shadow=TRUE, cex=0.7, axes=c(1,3), title="Représentation sur le plan 1-3")
```



Classification sur les colonnes

```
res.hcpc <- HCPC(res, cluster.CA="columns")
```

