

Correspondance Analysis (CA) with FactoMiner (Birth dataset)

Magalie Houée-Bigot

Import data

Upload the Birth dataset on your computer

```
setwd("C:/users/houee")
Birth <- read.table("data_CA_birth.csv", header=TRUE, sep=";", row.names=1,
                    check.names=FALSE)
```

header=TRUE : indicates that the file contains the names of the variables

sep=";" : indicates the fields separator (usually ";" or "," for csv files)

row.names=1 : indicates the column of the table which contains the row names

check.names=FALSE : indicated that the names of the variables in the data frame are unchecked

It is important to check that the import is well done

```
summary(Birth)
```

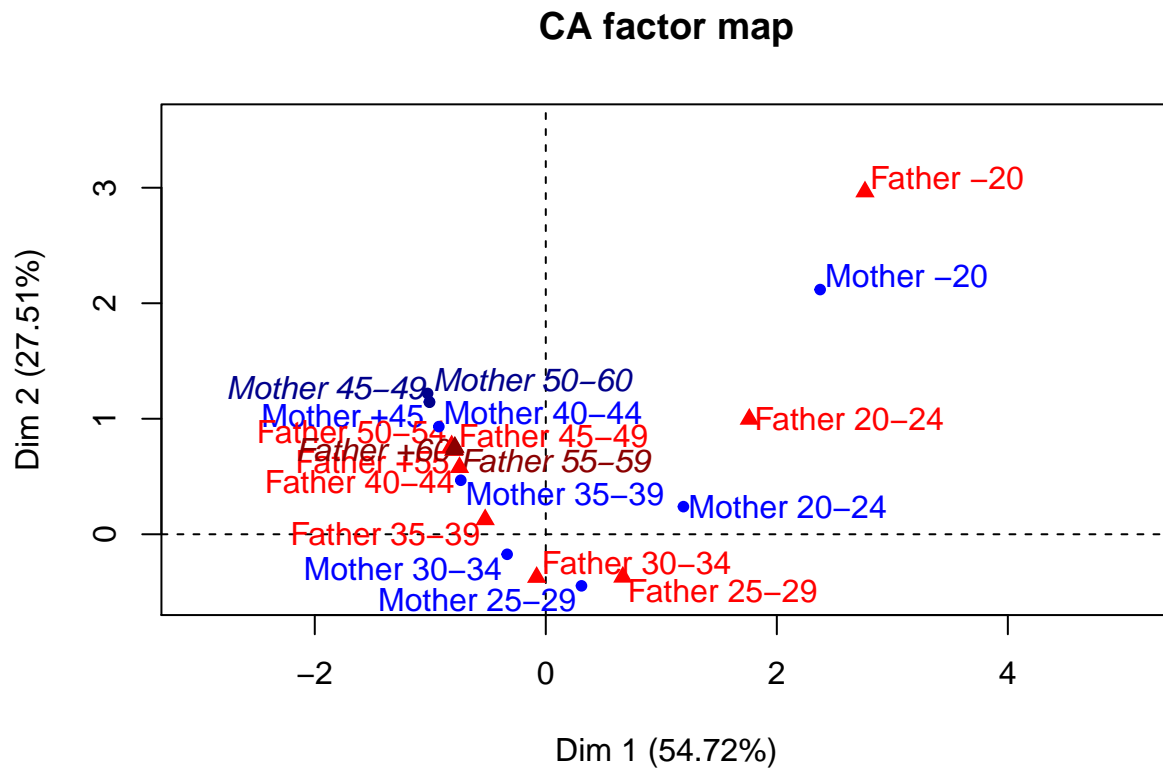
```
##      Father -20      Father 20-24      Father 25-29      Father 30-34
## Min.   : 0.0      Min.   : 0      Min.   : 2      Min.   : 8
## 1st Qu.: 0.0      1st Qu.: 4      1st Qu.: 29     1st Qu.: 147
## Median : 17.0     Median : 394     Median : 1771    Median : 2685
## Mean   : 349.3     Mean   : 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
##      Father 35-39      Father 40-44      Father 45-49      Father 50-54
## 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
##      Father 55-59      Father +60      Father +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
```

Loading FactoMineR

```
library(FactoMineR)
```

Correspondence Analysis

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



Outputs can be summarized with the function `summary`.

```
summary(res)
```

```
## function summary.CA
```

```
##  
## Call:  
## CA(X = Birth, 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  
##
```

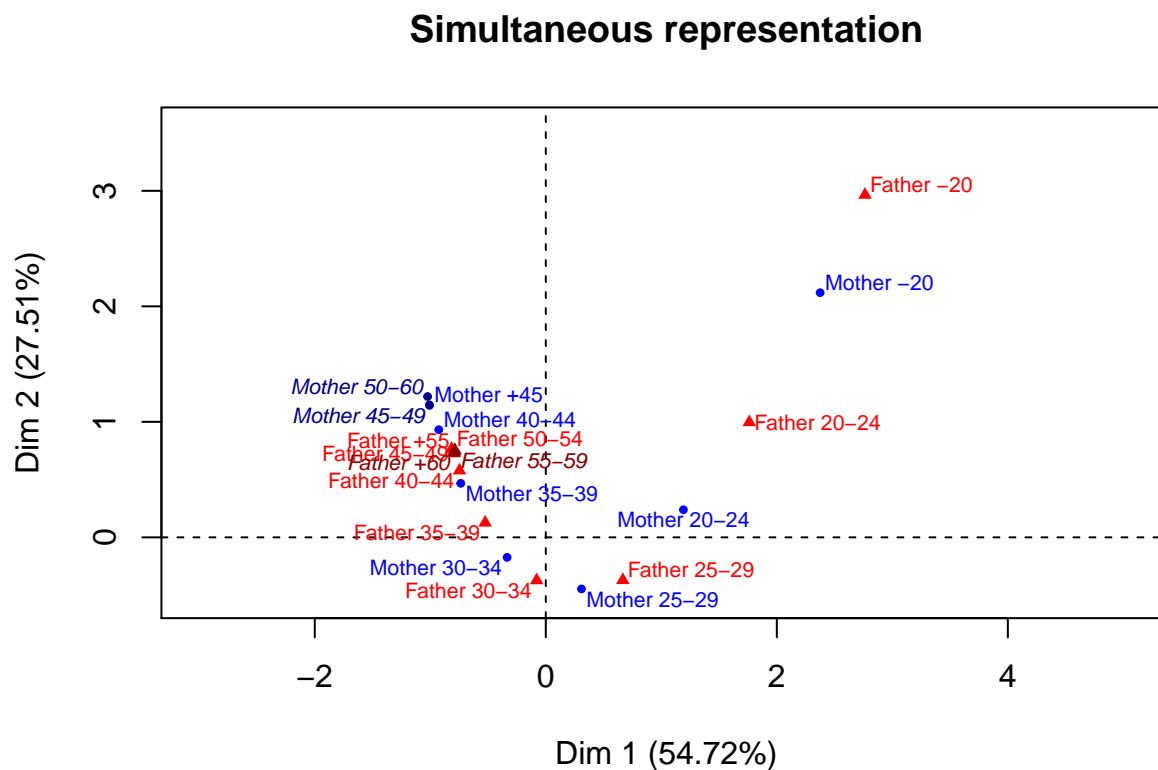
```

## Rows
##          Iner*1000      Dim.1      ctr      cos2      Dim.2      ctr
## Mother -20 | 209.626 | 2.375 21.138 0.482 | 2.118 33.449
## Mother 20-24 | 199.127 | 1.192 36.604 0.878 | 0.239 2.933
## Mother 25-29 | 104.178 | 0.309 6.095 0.279 | -0.446 25.286
## Mother 30-34 | 85.815 | -0.335 7.841 0.436 | -0.173 4.187
## Mother 35-39 | 140.420 | -0.736 19.228 0.654 | 0.468 15.443
## Mother 40-44 | 113.997 | -0.926 8.431 0.353 | 0.932 16.990
## Mother +45 | 19.649 | -1.008 0.664 0.162 | 1.147 1.712
##          cos2      Dim.3      ctr      cos2
## Mother -20 0.383 | -1.136 22.053 0.110 |
## Mother 20-24 0.035 | 0.254 7.574 0.040 |
## Mother 25-29 0.583 | 0.170 8.420 0.085 |
## Mother 30-34 0.117 | -0.325 33.769 0.412 |
## Mother 35-39 0.264 | 0.048 0.377 0.003 |
## Mother 40-44 0.358 | 0.729 23.867 0.219 |
## Mother +45 0.209 | 1.148 3.939 0.210 |
##
## Columns
##          Iner*1000      Dim.1      ctr      cos2      Dim.2      ctr
## Father -20 | 93.775 | 2.763 6.193 0.315 | 2.965 14.186
## Father 20-24 | 257.192 | 1.762 39.802 0.739 | 0.997 25.345
## Father 25-29 | 162.234 | 0.667 20.771 0.612 | -0.370 12.726
## Father 30-34 | 72.545 | -0.079 0.424 0.028 | -0.372 18.773
## Father 35-39 | 87.325 | -0.525 12.916 0.706 | 0.127 1.508
## Father 40-44 | 106.558 | -0.747 12.485 0.560 | 0.579 14.919
## Father 45-49 | 58.708 | -0.797 4.911 0.400 | 0.727 8.121
## Father 50-54 | 23.165 | -0.815 1.720 0.355 | 0.766 3.026
## Father +55 | 11.307 | -0.788 0.777 0.328 | 0.749 1.396
##          cos2      Dim.3      ctr      cos2
## Father -20 0.363 | -2.099 16.320 0.182 |
## Father 20-24 0.237 | -0.150 1.324 0.005 |
## Father 25-29 0.188 | 0.370 29.096 0.188 |
## Father 30-34 0.621 | -0.241 18.075 0.261 |
## Father 35-39 0.041 | -0.216 9.973 0.119 |
## Father 40-44 0.336 | 0.287 8.394 0.082 |
## Father 45-49 0.332 | 0.544 10.446 0.186 |
## Father 50-54 0.314 | 0.602 4.281 0.193 |
## Father +55 0.296 | 0.605 2.091 0.193 |
##
## Supplementary rows
##          Dim.1      cos2      Dim.2      cos2      Dim.3      cos2
## Mother 45-49 | -1.007 0.167 | 1.143 0.216 | 1.137 0.213 |
## Mother 50-60 | -1.023 0.061 | 1.219 0.087 | 1.356 0.108 |
##
## Supplementary columns
##          Dim.1      cos2      Dim.2      cos2      Dim.3      cos2
## Father 55-59 | -0.787 0.314 | 0.755 0.288 | 0.620 0.195 |
## Father +60 | -0.791 0.366 | 0.734 0.315 | 0.567 0.188 |

```

Simultaneous representation with a title and a smaller size for the labels

```
plot(res, cex=0.7, title="Simultaneous representation")
```



Description of the dimensions

```
dimdesc(res)
```

```
## $`Dim 1`  
## $`Dim 1`$row  
##          coord  
## Mother 50-60 -1.0230623  
## Mother +45  -1.0075217  
## Mother 45-49 -1.0066959  
## Mother 40-44 -0.9258528  
## Mother 35-39 -0.7359407  
## Mother 30-34 -0.3345237  
## Mother 25-29  0.3091183  
## Mother 20-24  1.1916963  
## Mother -20   2.3752817  
##
```

```

## $`Dim 1`$col
##          coord
## Father 50-54 -0.81497107
## Father 45-49 -0.79730553
## Father +60   -0.79075017
## Father +55   -0.78809544
## Father 55-59 -0.78703920
## Father 40-44 -0.74677039
## Father 35-39 -0.52468621
## Father 30-34 -0.07889759
## Father 25-29  0.66703803
## Father 20-24  1.76185986
## Father -20    2.76326670
##
##
## $`Dim 2`
## $`Dim 2`$row
##          coord
## Mother 25-29 -0.4464193
## Mother 30-34 -0.1733188
## Mother 20-24  0.2391568
## Mother 35-39  0.4676267
## Mother 40-44  0.9318948
## Mother 45-49  1.1429222
## Mother +45    1.1467692
## Mother 50-60  1.2191707
## Mother -20    2.1184937
##
## $`Dim 2`$col
##          coord
## Father 30-34 -0.3720552
## Father 25-29 -0.3701847
## Father 35-39  0.1271161
## Father 40-44  0.5787896
## Father 45-49  0.7269325
## Father +60    0.7338045
## Father +55    0.7487134
## Father 55-59  0.7546451
## Father 50-54  0.7664040
## Father 20-24  0.9968244
## Father -20    2.9651350
##
##
## $`Dim 3`
## $`Dim 3`$row
##          coord
## Mother -20   -1.13553529
## Mother 30-34 -0.32492252
## Mother 35-39  0.04824263
## Mother 25-29  0.17005598
## Mother 20-24  0.25371576
## Mother 40-44  0.72911863
## Mother 45-49  1.13712215
## Mother +45    1.14815637

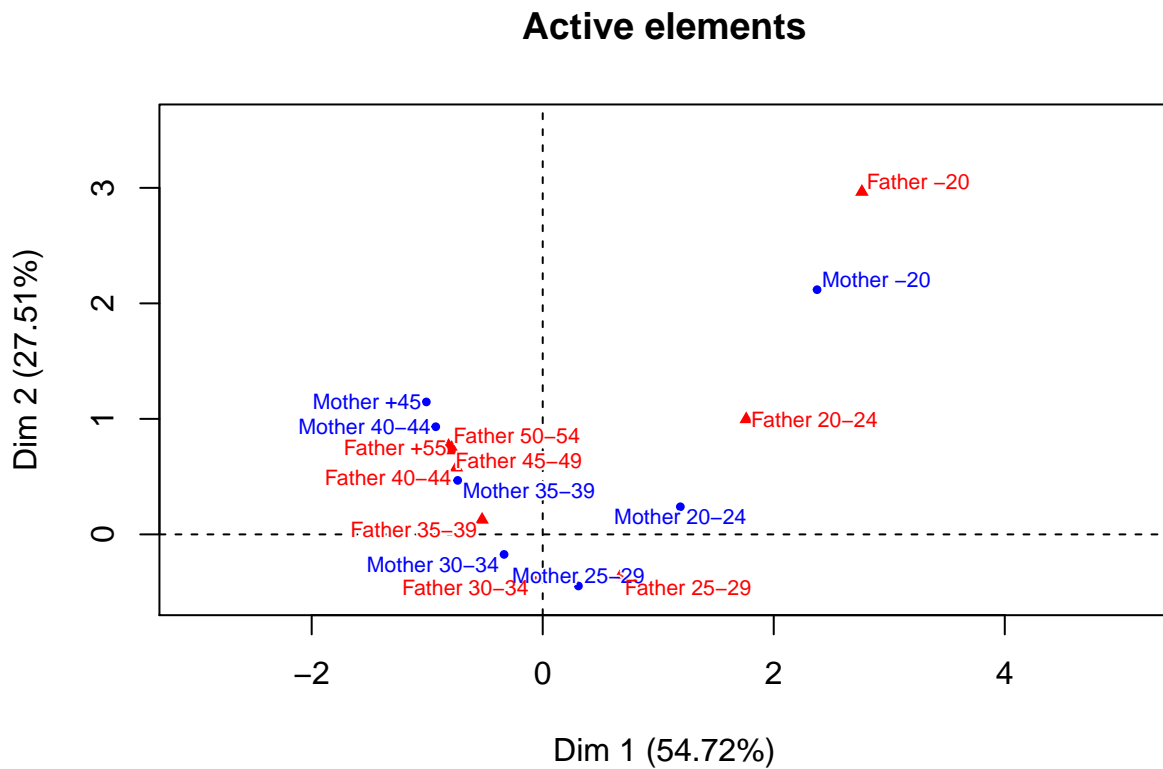
```

```
## Mother 50-60 1.35582383
##
## $`Dim 3`$col
##          coord
## Father -20 -2.0993859
## Father 30-34 -0.2409949
## Father 35-39 -0.2157827
## Father 20-24 -0.1504214
## Father 40-44 0.2865959
## Father 25-29 0.3695078
## Father 45-49 0.5442414
## Father +60 0.5669442
## Father 50-54 0.6017674
## Father +55 0.6049307
## Father 55-59 0.6200443
```

Graphs with several selections for rows and/or columns

Plot with only the active rows and columns

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



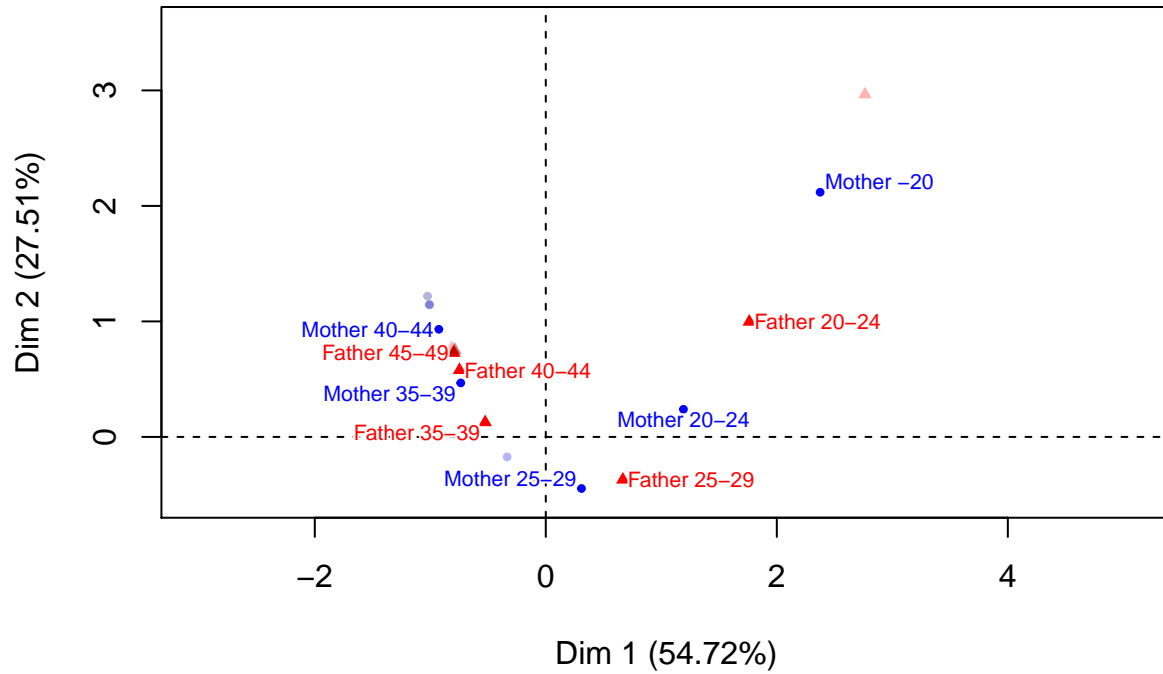
`selectRow="cos2 0.7"` : select the rows with a sufficiently good representation, with cos squared of at least 0.7 on the plane

and similarly for the columns

```
selectCol="cos2 0.7"
```

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

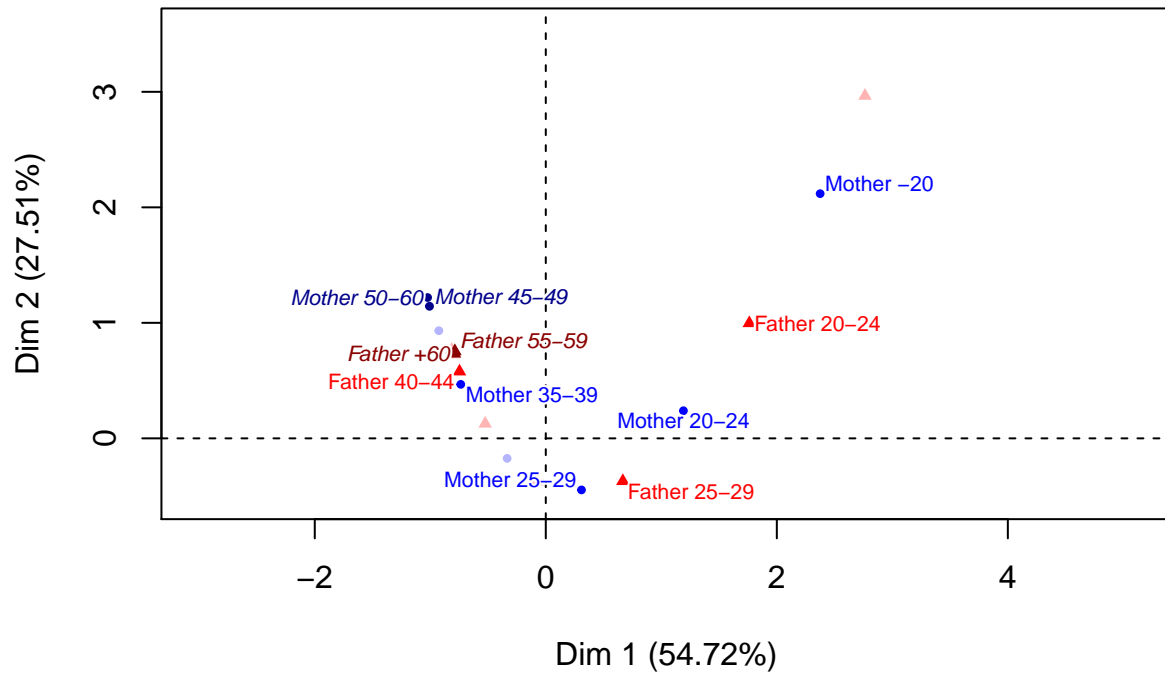
CA factor map



select the four best represented rows and the three best represented columns

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

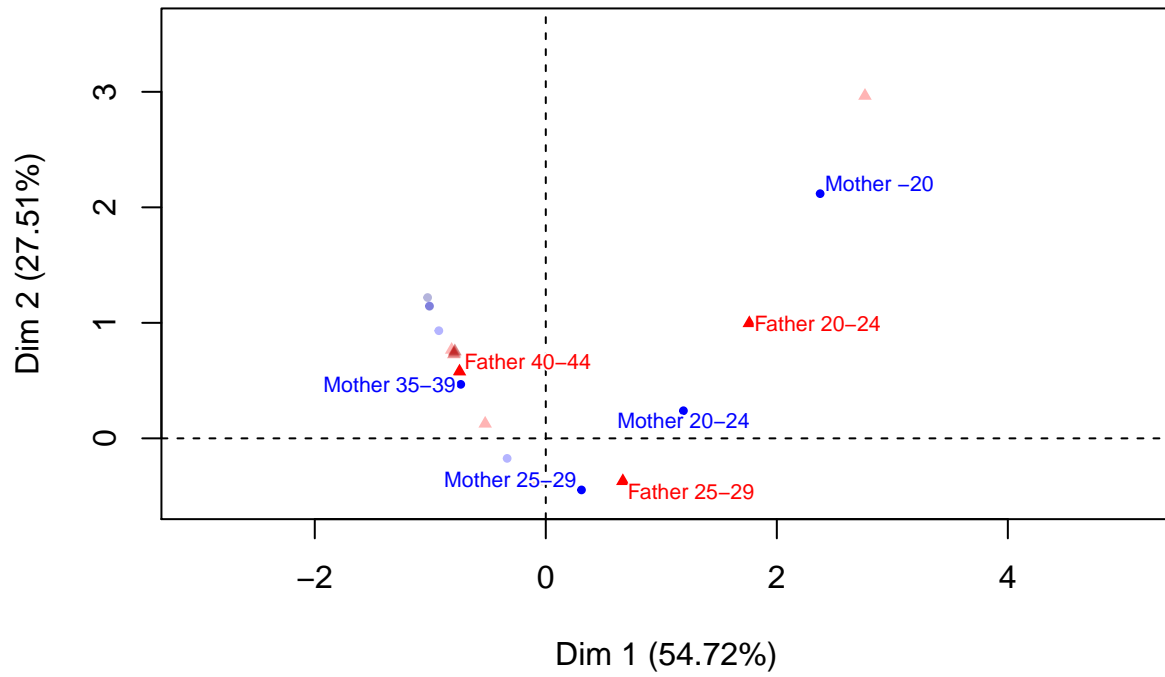
CA factor map



select rows and columns with the largest contribution to the plane's construction

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

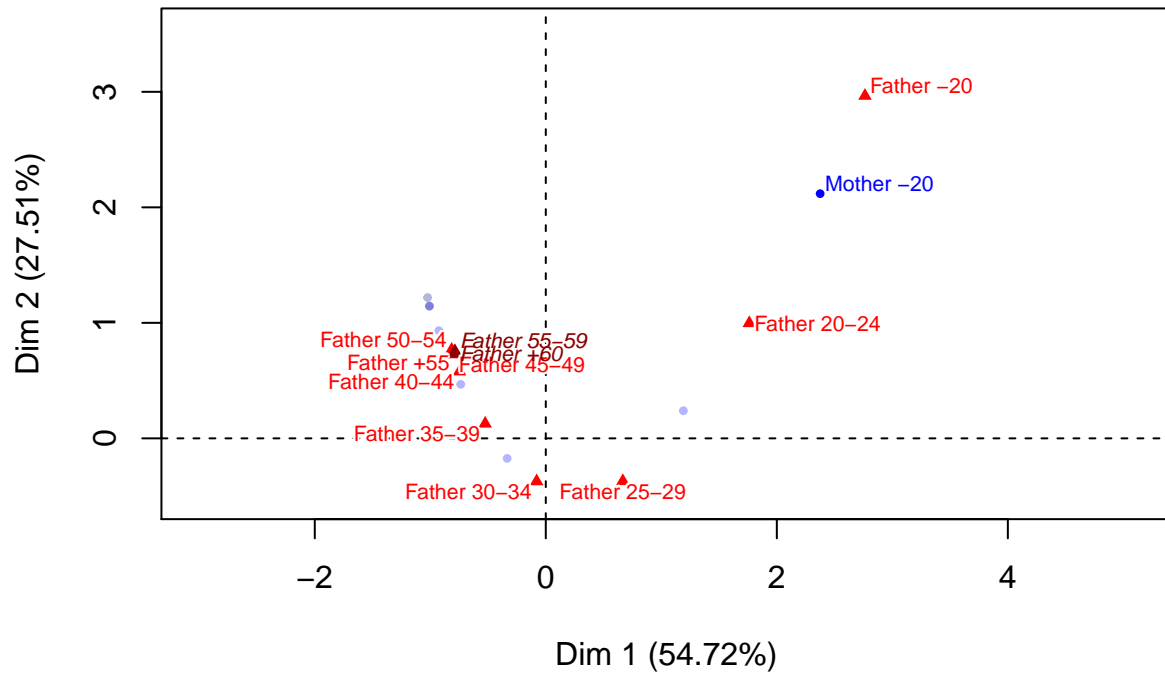

CA factor map



select by the categories name

```
plot(res, shadow=TRUE, cex=0.7, selectRow=c("Mother -20"))
```

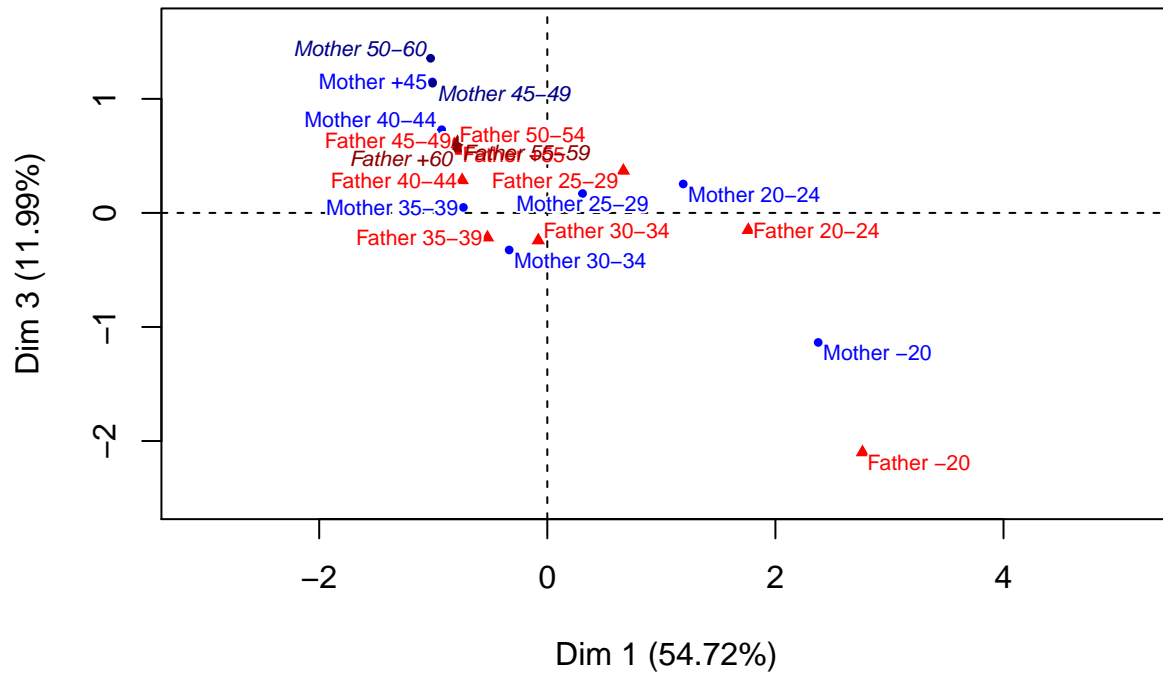
CA factor map



Graph with dimensions 1 and 3

```
plot(res, shadow=TRUE, cex=0.7, axes=c(1,3), title="Representation on dimensions 1-3")
```

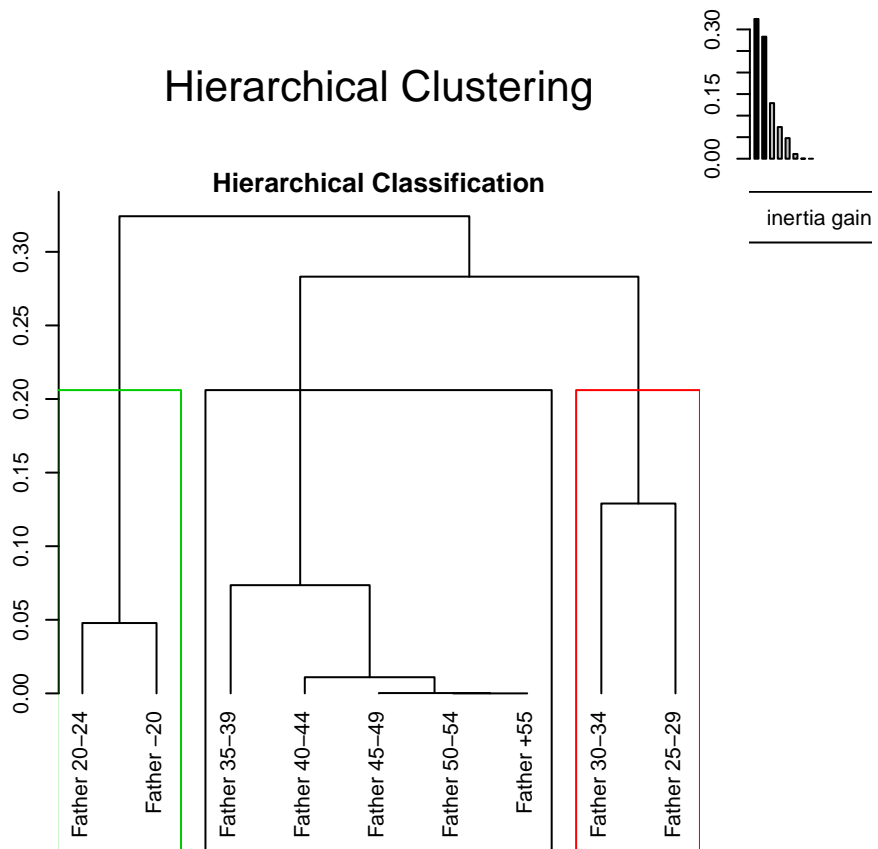
Representation on dimensions 1-3



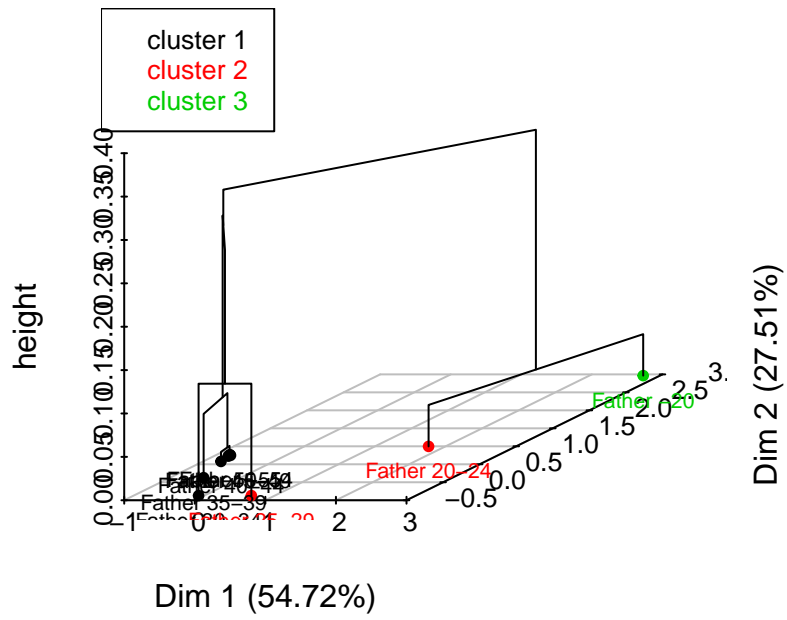
Clustering on the columns

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

Hierarchical Clustering



Hierarchical clustering on the factor map



Factor map

