Classification avec FactoMineR

François Husson

Importation des données

```
decathlon <- read.table("http://www.agrocampus-ouest.fr/math/livreR/decathlon.csv",
    header=TRUE, sep=";",dec=".", row.names=1, check.names=FALSE, fileEncoding="latin1")
summary(decathlon)
```

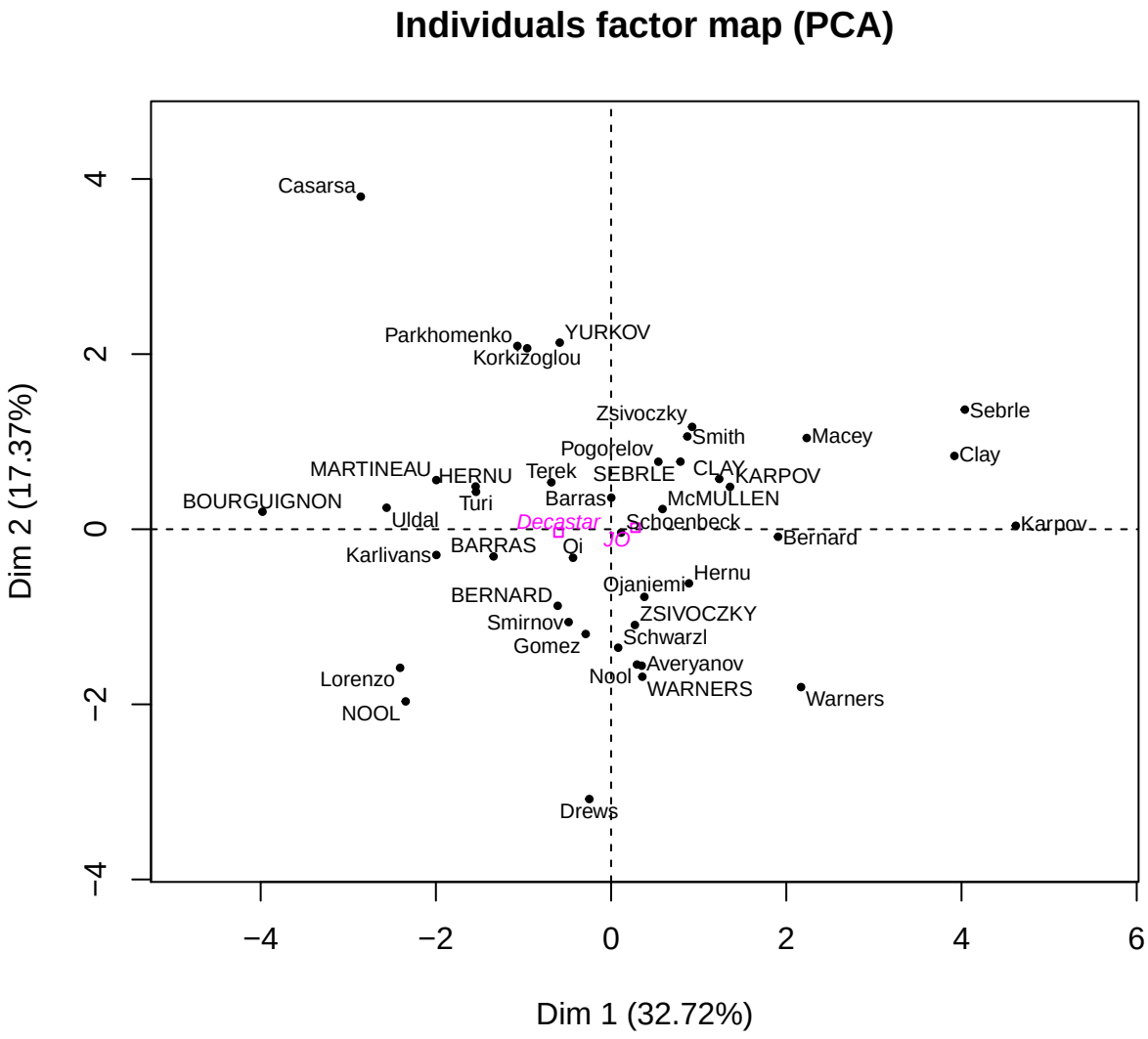
```
##       100m          Longueur        Poids          Hauteur
##  Min.   :10.44   Min.   :6.61   Min.   :12.68   Min.   :1.850
##  1st Qu.:10.85   1st Qu.:7.03   1st Qu.:13.88   1st Qu.:1.920
##  Median :10.98   Median :7.30   Median :14.57   Median :1.950
##  Mean   :11.00   Mean   :7.26   Mean   :14.48   Mean   :1.977
##  3rd Qu.:11.14   3rd Qu.:7.48   3rd Qu.:14.97   3rd Qu.:2.040
##  Max.   :11.64   Max.   :7.96   Max.   :16.36   Max.   :2.150
##       400m           110m H          Disque          Perche
##  Min.   :46.81   Min.   :13.97   Min.   :37.92   Min.   :4.200
##  1st Qu.:48.93   1st Qu.:14.21   1st Qu.:41.90   1st Qu.:4.500
##  Median :49.40   Median :14.48   Median :44.41   Median :4.800
##  Mean   :49.62   Mean   :14.61   Mean   :44.33   Mean   :4.762
##  3rd Qu.:50.30   3rd Qu.:14.98   3rd Qu.:46.07   3rd Qu.:4.920
##  Max.   :53.20   Max.   :15.67   Max.   :51.65   Max.   :5.400
##     Javelot          1500m         Classement        Points
##  Min.   :50.31   Min.   :262.1   Min.   : 1.00   Min.   :7313
##  1st Qu.:55.27   1st Qu.:271.0   1st Qu.: 6.00   1st Qu.:7802
##  Median :58.36   Median :278.1   Median :11.00   Median :8021
##  Mean   :58.32   Mean   :279.0   Mean   :12.12   Mean   :8005
##  3rd Qu.:60.89   3rd Qu.:285.1   3rd Qu.:18.00   3rd Qu.:8122
##  Max.   :70.52   Max.   :317.0   Max.   :28.00   Max.   :8893
##    Competition
##  Decastar:13
##  JO      :28
##
##
##
##
```

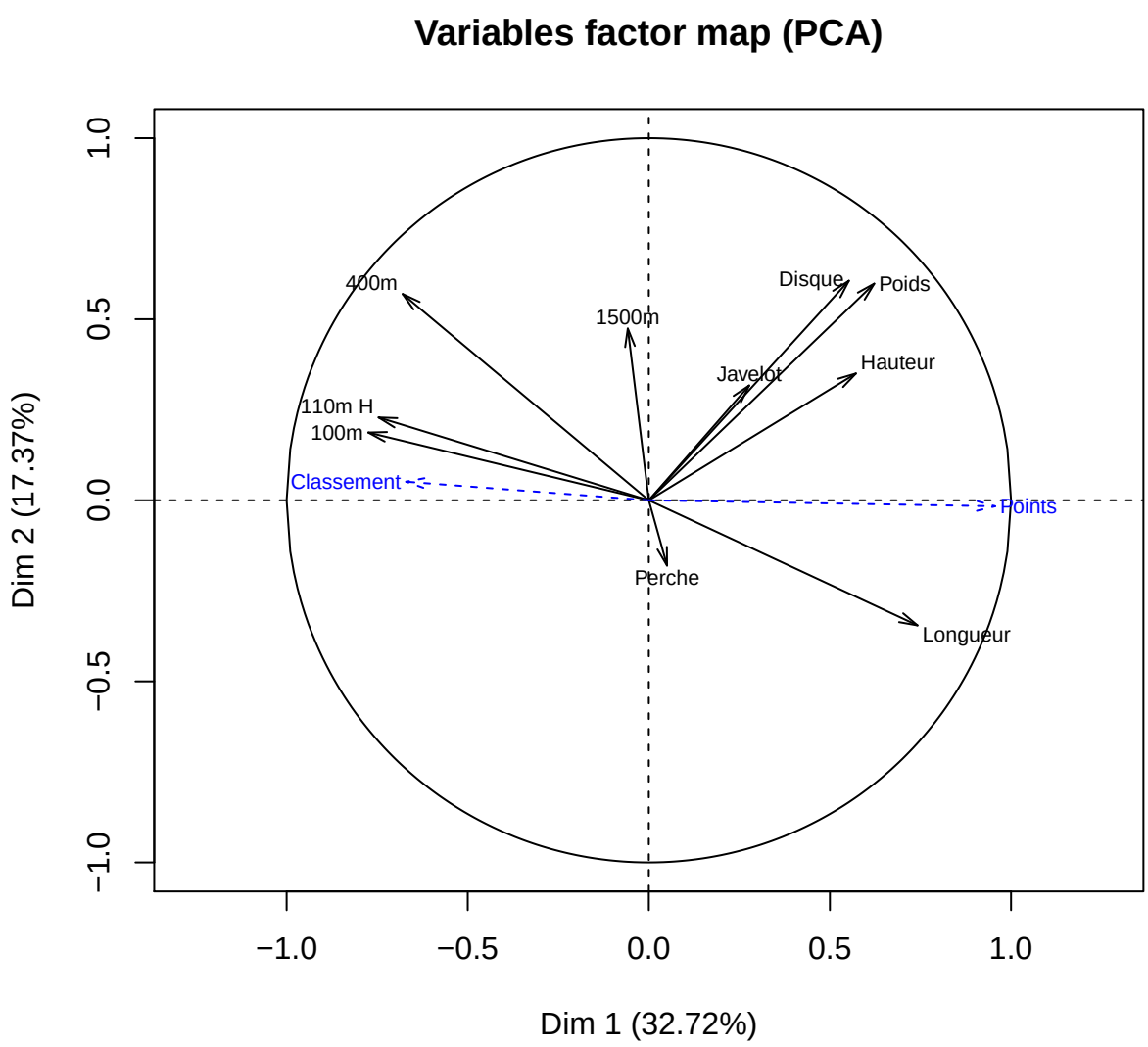
Chargement du package

```
library(FactoMineR)
```

Construction de l'analyse en composantes principales

```
res <- PCA(decathlon, quanti.sup=11:12, quali.sup=13, ncp=Inf)
```






```
res$eig
```

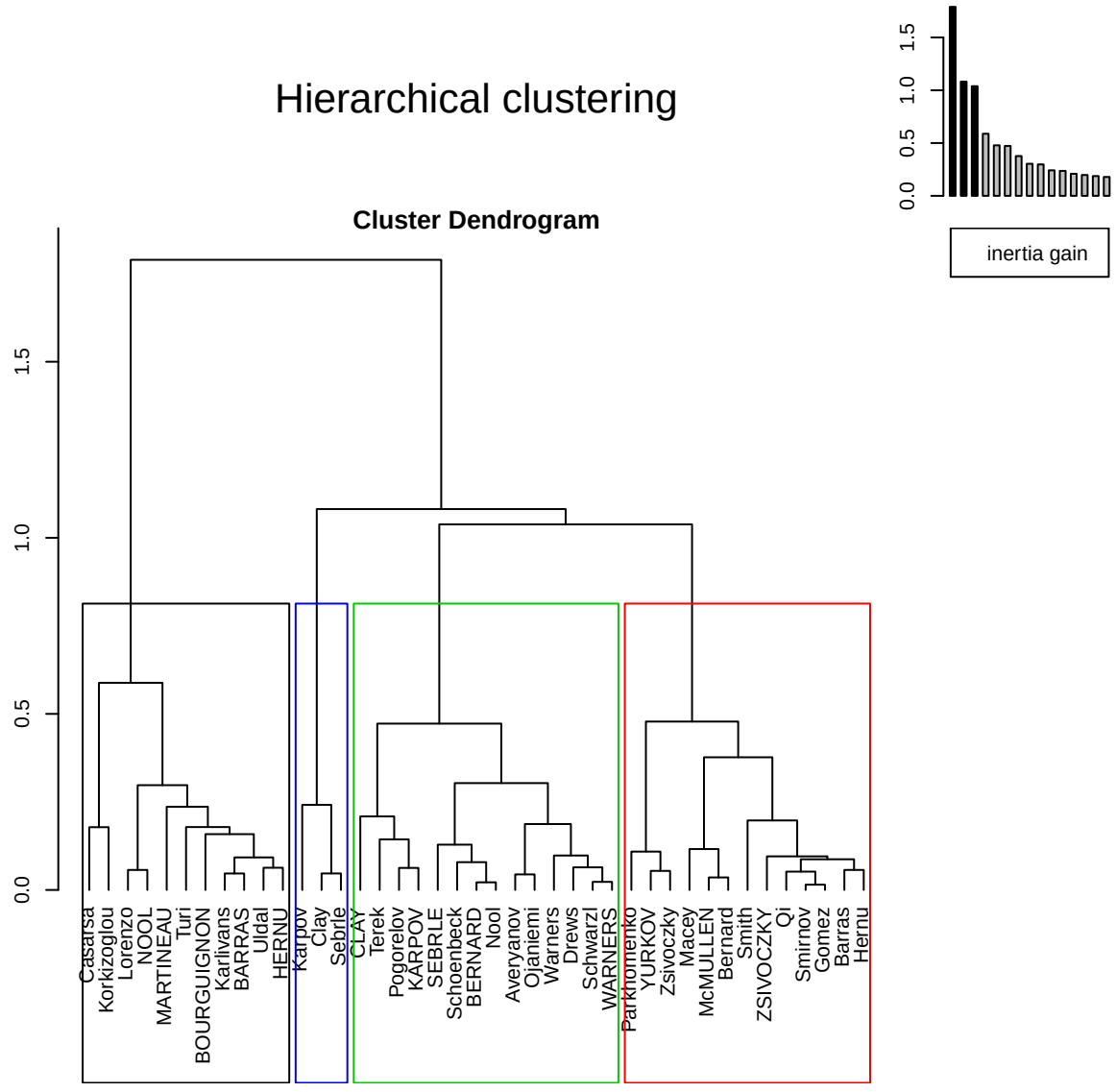
```
##         eigenvalue percentage of variance
## comp 1   3.2719055              32.719055
## comp 2   1.7371310              17.371310
## comp 3   1.4049167              14.049167
## comp 4   1.0568504              10.568504
## comp 5   0.6847735               6.847735
## comp 6   0.5992687               5.992687
## comp 7   0.4512353               4.512353
## comp 8   0.3968766               3.968766
## comp 9   0.2148149               2.148149
## comp 10  0.1822275               1.822275
##         cumulative percentage of variance
## comp 1                          32.71906
```

```
## comp 2                          50.09037
## comp 3                          64.13953
## comp 4                          74.70804
## comp 5                          81.55577
## comp 6                          87.54846
## comp 7                          92.06081
## comp 8                          96.02958
## comp 9                          98.17773
## comp 10                        100.00000
```

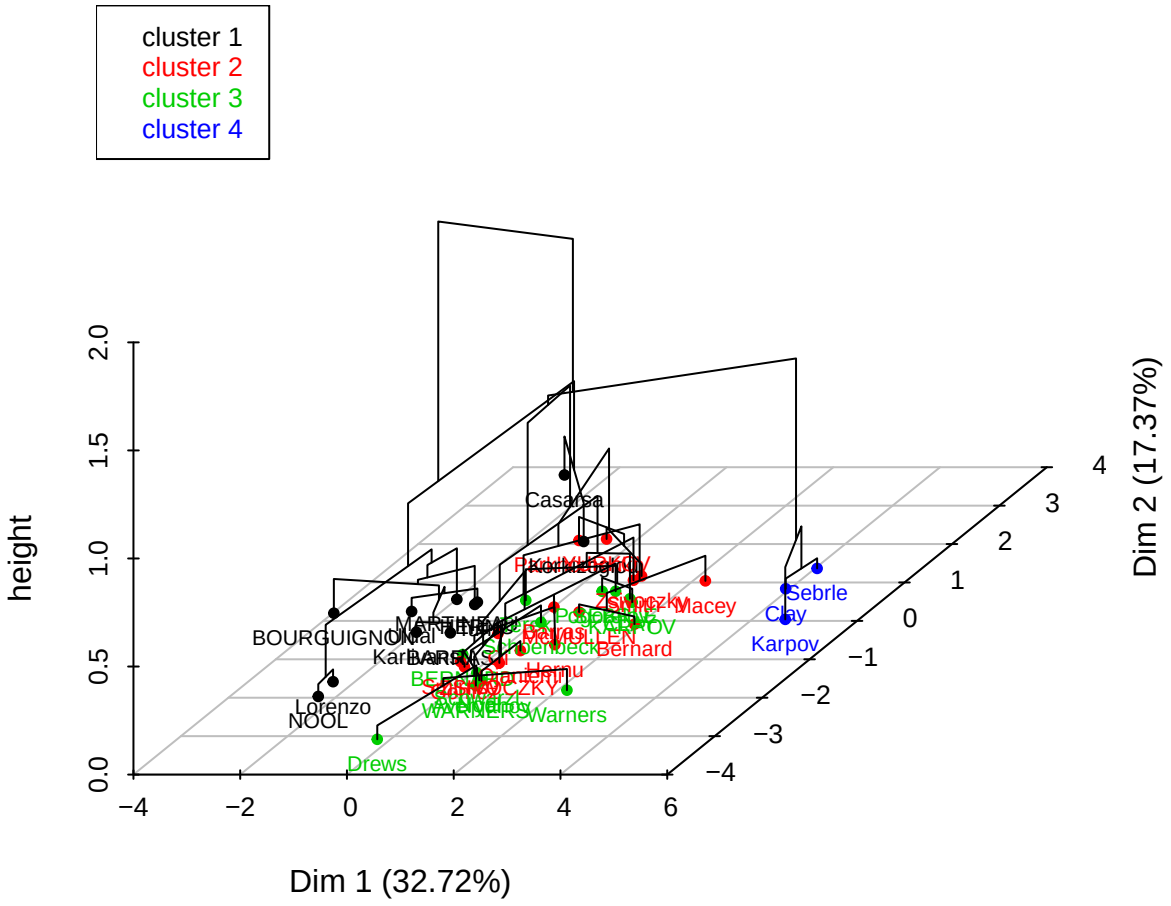
```
res <- PCA(decathlon, quanti.sup=11:12, quali.sup=13, ncp=8, graph=FALSE)
```

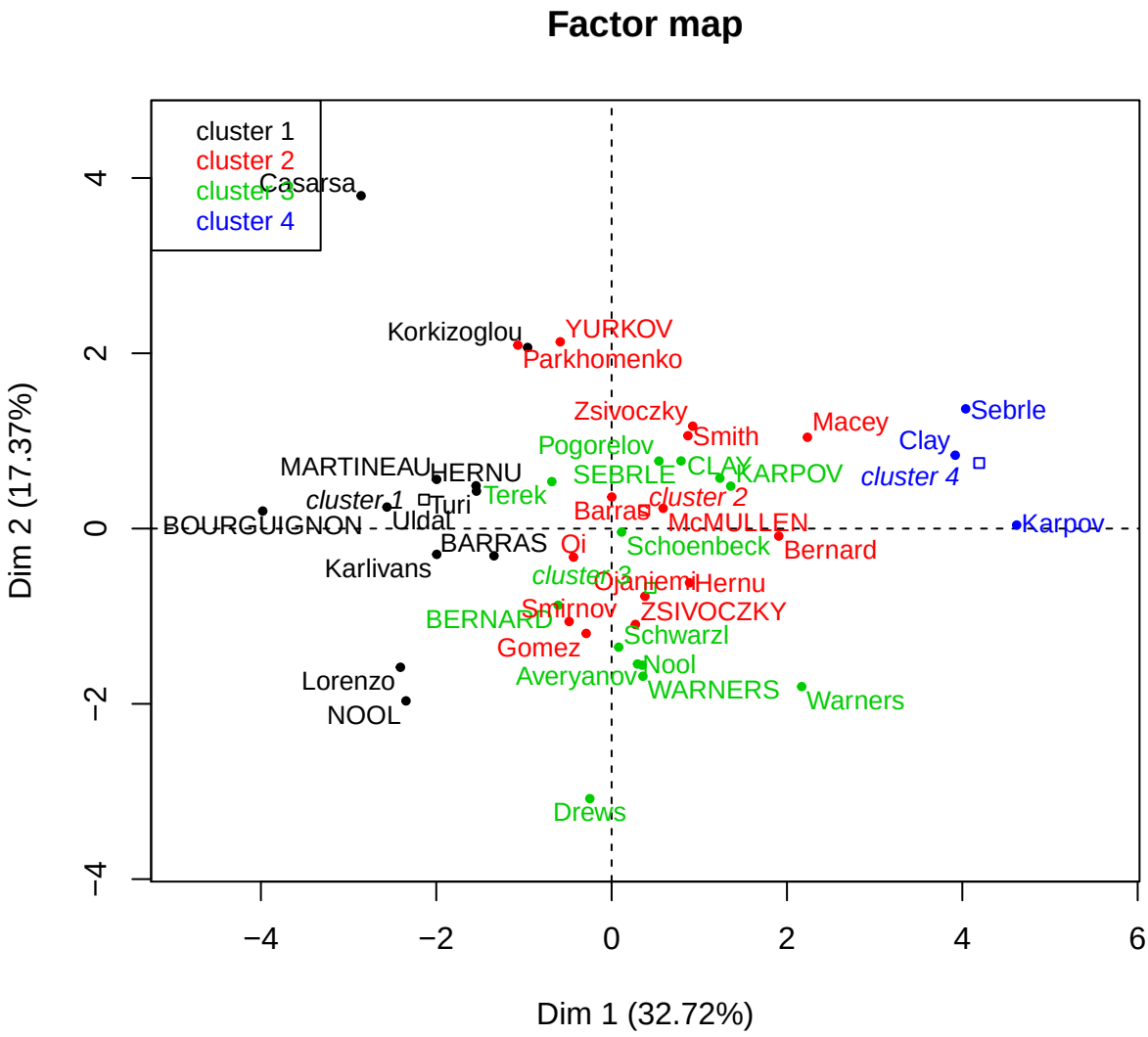
Construction de la classification ascendante hiérarchique

```
res.hcpc <- HCPC(res, kk=Inf, min=3, max=10, consol=TRUE)
```

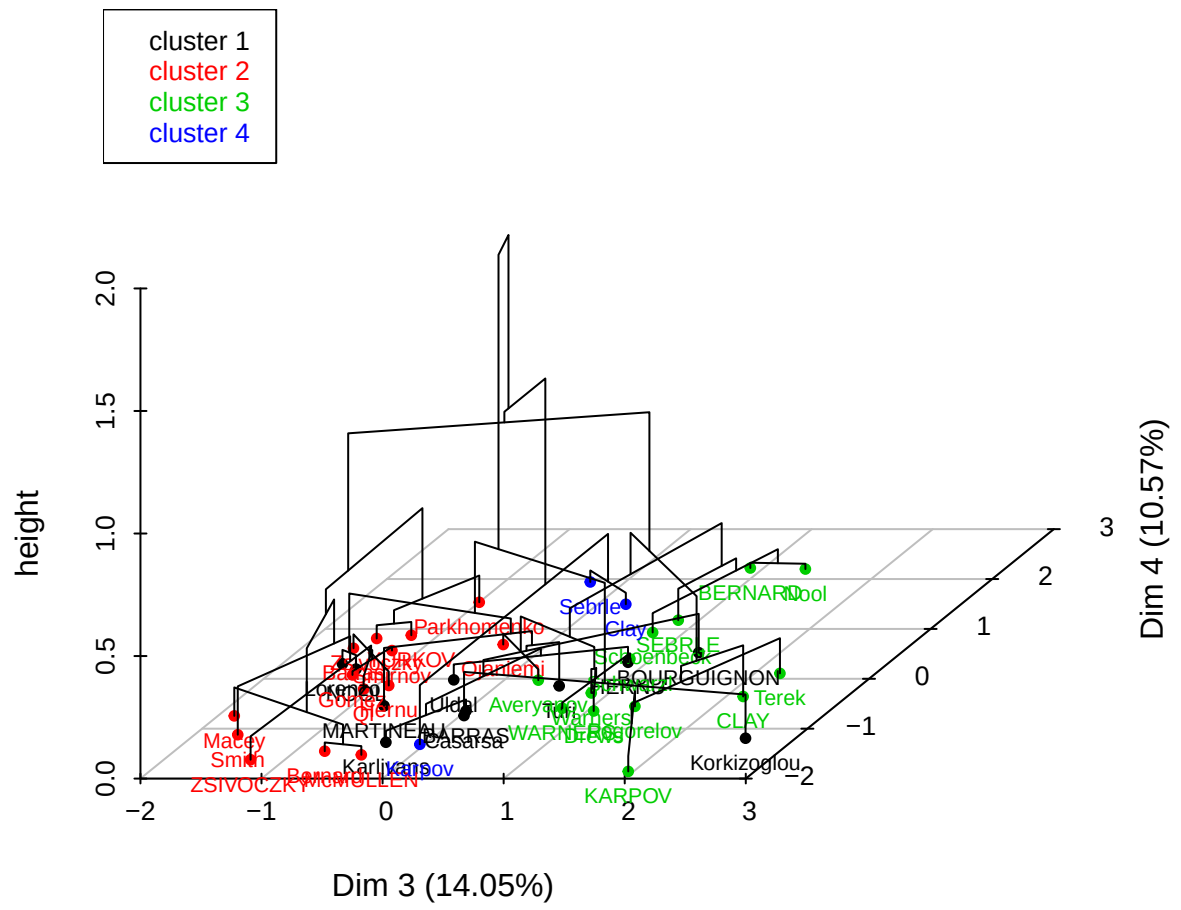


Hierarchical clustering on the factor map






```
plot(res.hcpc, axes=3:4)
```



Résultats de la classification

```
names(res.hcpc)
```

```
## [1] "data.clust" "desc.var"   "desc.axes"  "call"       "desc.ind"
```

```
res.hcpc$call$t
```

```
## $res
## **Results for the Principal Component Analysis (PCA)**
## The analysis was performed on 41 individuals, described by 13 variables
## *The results are available in the following objects:
##
##    name
## 1  "$eig"
## 2  "$var"
```

```
## 3  "$var$coord"
## 4  "$var$cor"
## 5  "$var$cos2"
## 6  "$var$contrib"
## 7  "$ind"
## 8  "$ind$coord"
## 9  "$ind$cos2"
## 10 "$ind$contrib"
## 11 "$quanti.sup"
## 12 "$quanti.sup$coord"
## 13 "$quanti.sup$cor"
## 14 "$quali.sup"
## 15 "$quali.sup$coord"
## 16 "$quali.sup$v.test"
## 17 "$call"
## 18 "$call$centre"
## 19 "$call$ecart.type"
## 20 "$call$row.w"
## 21 "$call$col.w"
##    description
## 1  "eigenvalues"
## 2  "results for the variables"
## 3  "coord. for the variables"
## 4  "correlations variables - dimensions"
## 5  "cos2 for the variables"
## 6  "contributions of the variables"
## 7  "results for the individuals"
## 8  "coord. for the individuals"
## 9  "cos2 for the individuals"
## 10 "contributions of the individuals"
## 11 "results for the supplementary quantitative variables"
## 12 "coord. for the supplementary quantitative variables"
## 13 "correlations suppl. quantitative variables - dimensions"
## 14 "results for the supplementary categorical variables"
## 15 "coord. for the supplementary categories"
## 16 "v-test of the supplementary categories"
## 17 "summary statistics"
## 18 "mean of the variables"
## 19 "standard error of the variables"
## 20 "weights for the individuals"
## 21 "weights for the variables"
##
## $tree
##
## Call:
## flashClust::hclust(d = dissi, method = method, members = weight)
##
## Cluster method   : ward
## Distance         : euclidean
## Number of objects: 41
##
##
## $nb.clust
## [1] 4
```

```
##
## $within
##  [1] 9.60295766 7.81359025 6.73201549 5.69384289 5.10568403 4.62753793
##  [7] 4.15518969 3.77848179 3.47508130 3.17757353 2.93601063 2.69982282
## [13] 2.49103892 2.29370340 2.10659252 1.92797387 1.74973452 1.59132327
## [19] 1.44775789 1.31915516 1.20305119 1.09413928 0.99635559 0.90095691
## [25] 0.80853956 0.72143332 0.64254463 0.57837735 0.51529830 0.45271459
## [31] 0.39587815 0.33905840 0.28492784 0.23292321 0.18607536 0.13926531
## [37] 0.09510137 0.05959903 0.03662152 0.01495934
##
## $inert.gain
##  [1] 1.78936741 1.08157476 1.03817260 0.58815886 0.47814610 0.47234824
##  [7] 0.37670790 0.30340049 0.29750778 0.24156290 0.23618781 0.20878390
## [13] 0.19733552 0.18711087 0.17861865 0.17823935 0.15841125 0.14356538
## [19] 0.12860273 0.11610397 0.10891190 0.09778369 0.09539868 0.09241735
## [25] 0.08710624 0.07888869 0.06416729 0.06307905 0.06258371 0.05683644
## [31] 0.05681976 0.05413056 0.05200463 0.04684785 0.04681006 0.04416394
## [37] 0.03550234 0.02297751 0.02166218 0.01495934
##
## $quot
## [1] 0.8615777 0.8457858 0.8967027 0.9063502 0.8979267 0.9093404 0.9197031
## [8] 0.9143883
```

```
res.hcpc$data.clust
```

```
##              100m Longueur Poids Hauteur  400m 110m H Disque Perche
## Sebrle      10.85     7.84 16.36    2.12 48.36  14.05  48.72   5.00
## Clay        10.44     7.96 15.23    2.06 49.19  14.13  50.11   4.90
## Karpov      10.50     7.81 15.93    2.09 46.81  13.97  51.65   4.60
## Macey       10.89     7.47 15.73    2.15 48.97  14.56  48.34   4.40
## Warners     10.62     7.74 14.48    1.97 47.97  14.01  43.73   4.90
## Zsivoczky   10.91     7.14 15.31    2.12 49.40  14.95  45.62   4.70
## Hernu       10.97     7.19 14.65    2.03 48.73  14.25  44.72   4.80
## Nool        10.80     7.53 14.26    1.88 48.81  14.80  42.05   5.40
## Bernard     10.69     7.48 14.80    2.12 49.13  14.17  44.75   4.40
## Schwarzl    10.98     7.49 14.01    1.94 49.76  14.25  42.43   5.10
## Pogorelov   10.95     7.31 15.10    2.06 50.79  14.21  44.60   5.00
## Schoenbeck  10.90     7.30 14.77    1.88 50.30  14.34  44.41   5.00
## Barras      11.14     6.99 14.91    1.94 49.41  14.37  44.83   4.60
## Smith       10.85     6.81 15.24    1.91 49.27  14.01  49.02   4.20
## Averyanov   10.55     7.34 14.44    1.94 49.72  14.39  39.88   4.80
## Ojaniemi    10.68     7.50 14.97    1.94 49.12  15.01  40.35   4.60
## Smirnov     10.89     7.07 13.88    1.94 49.11  14.77  42.47   4.70
## Qi          11.06     7.34 13.55    1.97 49.65  14.78  45.13   4.50
## Drews       10.87     7.38 13.07    1.88 48.51  14.01  40.11   5.00
## Parkhomenko 11.14     6.61 15.69    2.03 51.04  14.88  41.90   4.80
## Terek       10.92     6.94 15.15    1.94 49.56  15.12  45.62   5.30
## Gomez       11.08     7.26 14.57    1.85 48.61  14.41  40.95   4.40
## Turi        11.08     6.91 13.62    2.03 51.67  14.26  39.83   4.80
## Lorenzo     11.10     7.03 13.22    1.85 49.34  15.38  40.22   4.50
## Karlivans   11.33     7.26 13.30    1.97 50.54  14.98  43.34   4.50
## Korkizoglou 10.86     7.07 14.81    1.94 51.16  14.96  46.07   4.70
## Uldal       11.23     6.99 13.53    1.85 50.95  15.09  43.01   4.50
## Casarsa     11.36     6.68 14.92    1.94 53.20  15.39  48.66   4.40
```

```
## SEBRLE      11.04     7.58 14.83    2.07 49.81  14.69  43.75   5.02
## CLAY        10.76     7.40 14.26    1.86 49.37  14.05  50.72   4.92
## KARPOV      11.02     7.30 14.77    2.04 48.37  14.09  48.95   4.92
## BERNARD     11.02     7.23 14.25    1.92 48.93  14.99  40.87   5.32
## YURKOV      11.34     7.09 15.19    2.10 50.42  15.31  46.26   4.72
## WARNERS     11.11     7.60 14.31    1.98 48.68  14.23  41.10   4.92
## ZSIVOCZKY   11.13     7.30 13.48    2.01 48.62  14.17  45.67   4.42
## McMULLEN    10.83     7.31 13.76    2.13 49.91  14.38  44.41   4.42
## MARTINEAU   11.64     6.81 14.57    1.95 50.14  14.93  47.60   4.92
## HERNU       11.37     7.56 14.41    1.86 51.10  15.06  44.99   4.82
## BARRAS      11.33     6.97 14.09    1.95 49.48  14.48  42.10   4.72
## NOOL        11.33     7.27 12.68    1.98 49.20  15.29  37.92   4.62
## BOURGUIGNON 11.36     6.80 13.46    1.86 51.16  15.67  40.49   5.02
##             Javelot  1500m Classement Points Competition clust
## Sebrle        70.52 280.01          1   8893          JO     4
## Clay          69.71 282.00          2   8820          JO     4
## Karpov        55.54 278.11          3   8725          JO     4
## Macey         58.46 265.42          4   8414          JO     2
## Warners       55.39 278.05          5   8343          JO     3
## Zsivoczky     63.45 269.54          6   8287          JO     2
## Hernu         57.76 264.35          7   8237          JO     2
## Nool          61.33 276.33          8   8235          JO     3
## Bernard       55.27 276.31          9   8225          JO     2
## Schwarzl      56.32 273.56         10   8102          JO     3
## Pogorelov     53.45 287.63         11   8084          JO     3
## Schoenbeck    60.89 278.82         12   8077          JO     3
## Barras        64.55 267.09         13   8067          JO     2
## Smith         61.52 272.74         14   8023          JO     2
## Averyanov     54.51 271.02         15   8021          JO     3
## Ojaniemi      59.26 275.71         16   8006          JO     2
## Smirnov       60.88 263.31         17   7993          JO     2
## Qi            60.79 272.63         18   7934          JO     2
## Drews         51.53 274.21         19   7926          JO     3
## Parkhomenko   65.82 277.94         20   7918          JO     2
## Terek         50.62 290.36         21   7893          JO     3
## Gomez         60.71 269.70         22   7865          JO     2
## Turi          59.34 290.01         23   7708          JO     1
## Lorenzo       58.36 263.08         24   7592          JO     1
## Karlivans     52.92 278.67         25   7583          JO     1
## Korkizoglou   53.05 317.00         26   7573          JO     1
## Uldal         60.00 281.70         27   7495          JO     1
## Casarsa       58.62 296.12         28   7404          JO     1
## SEBRLE        63.19 291.70          1   8217    Decastar     3
## CLAY          60.15 301.50          2   8122    Decastar     3
## KARPOV        50.31 300.20          3   8099    Decastar     3
## BERNARD       62.77 280.10          4   8067    Decastar     3
## YURKOV        63.44 276.40          5   8036    Decastar     2
## WARNERS       51.77 278.10          6   8030    Decastar     3
## ZSIVOCZKY     55.37 268.00          7   8004    Decastar     2
## McMULLEN      56.37 285.10          8   7995    Decastar     2
## MARTINEAU     52.33 262.10          9   7802    Decastar     1
## HERNU         57.19 285.10         10   7733    Decastar     1
## BARRAS        55.40 282.00         11   7708    Decastar     1
## NOOL          57.44 266.60         12   7651    Decastar     1
```

```
## BOURGUIGNON   54.68 291.70         13   7313    Decastar     1
```

```
res.hcpc$desc.var
```

```
## $quanti.var
##                 Eta2      P-value
## Points     0.8437383 5.573335e-15
## Perche     0.5764242 4.800164e-07
## 100m       0.5333868 2.775598e-06
## Longueur   0.5021046 8.970185e-06
## 400m       0.4212961 1.339694e-04
## 110m.H     0.4063919 2.110412e-04
## Poids      0.3774738 4.924878e-04
## Classement 0.3754467 5.217813e-04
## Javelot    0.3150077 2.670122e-03
## Hauteur    0.3129253 2.816321e-03
## Disque     0.2724198 7.661623e-03
## 1500m      0.2122280 3.007475e-02
##
## $quanti
## $quanti$`1`
##               v.test Mean in category Overall mean sd in category
## 100m        4.035700        11.271818    10.998049     0.19272298
## 400m        3.716021        50.721818    49.616341     1.11838210
## 110m.H      3.605283        15.044545    14.605854     0.38662558
## Classement  3.323127        18.909091    12.121951     7.39108241
## Hauteur    -2.239384         1.925455     1.976829     0.05836619
## Longueur   -2.796211         7.031818     7.260000     0.24052285
## Poids      -2.837963        13.873636    14.477073     0.69580918
## Points     -4.629619      7596.545455  8005.365854   140.51293285
##              Overall sd      p.value
## 100m         0.25979560 5.443985e-05
## 400m         1.13929751 2.023847e-04
## 110m.H       0.46599998 3.118133e-04
## Classement   7.82178048 8.901453e-04
## Hauteur      0.08785906 2.513092e-02
## Longueur     0.31251927 5.170562e-03
## Poids        0.81431175 4.540237e-03
## Points     338.18394159 3.663385e-06
##
## $quanti$`2`
##            v.test Mean in category Overall mean sd in category  Overall sd
## Hauteur  2.089667         2.017143     1.976829     0.09191833  0.08785906
## 1500m   -2.880815       271.731429   279.024878     5.89983760 11.53001177
## Perche  -3.570803         4.547143     4.762439     0.17527238  0.27458865
##              p.value
## Hauteur 0.0366477460
## 1500m   0.0039664846
## Perche  0.0003558891
##
## $quanti$`3`
##          v.test Mean in category Overall mean sd in category Overall sd
## Perche 4.452686         5.046154     4.762439      0.1763536  0.2745887
##             p.value
```

```
## Perche 8.480264e-06
##
## $quanti$`4`
##               v.test Mean in category Overall mean sd in category
## Points      4.242103       8812.66667  8005.365854    68.78145745
## Longueur    3.468581          7.87000     7.260000     0.06480741
## Disque      3.107539         50.16000    44.325610     1.19668988
## Poids       2.974272         15.84000    14.477073     0.46568945
## Javelot     2.586808         65.25667    58.316585     6.87867397
## Hauteur     2.289003          2.09000     1.976829     0.02449490
## 110m.H     -2.119695         14.05000    14.605854     0.06531973
## Classement -2.299627          2.00000    12.121951     0.81649658
## 400m       -2.333955         48.12000    49.616341     0.98634004
## 100m       -2.745523         10.59667    10.998049     0.18080069
##              Overall sd      p.value
## Points     338.18394159 2.214348e-05
## Longueur     0.31251927 5.232144e-04
## Disque       3.33639725 1.886523e-03
## Poids        0.81431175 2.936847e-03
## Javelot      4.76759315 9.686955e-03
## Hauteur      0.08785906 2.207917e-02
## 110m.H       0.46599998 3.403177e-02
## Classement   7.82178048 2.146935e-02
## 400m         1.13929751 1.959810e-02
## 100m         0.25979560 6.041458e-03
##
##
## attr(,"class")
## [1] "catdes" "list "
```

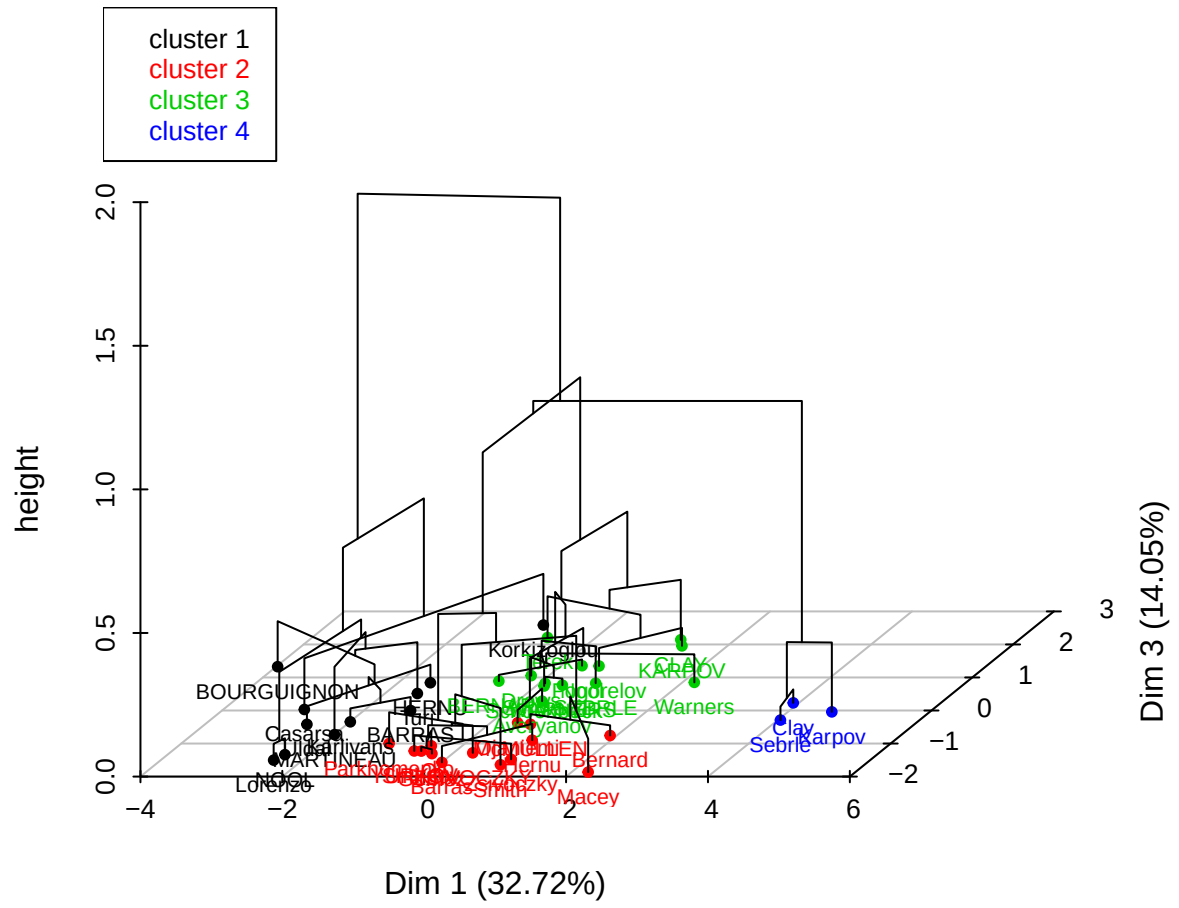
```
res.hcpc$desc.axes
```

```
## $quanti.var
##            Eta2      P-value
## Dim.1 0.8016334 4.493387e-13
## Dim.3 0.6337463 3.403421e-08
##
## $quanti
## $quanti$`1`
##          v.test Mean in category Overall mean sd in category Overall sd
## Dim.1 -4.528722        -2.138998 5.507535e-16      0.7963714   1.808841
##            p.value
## Dim.1 5.934161e-06
##
## $quanti$`2`
##          v.test Mean in category  Overall mean sd in category Overall sd
## Dim.3 -4.421665        -1.150798 -3.467755e-15      0.4209538   1.185292
##            p.value
## Dim.3 9.794314e-06
##
## $quanti$`3`
##          v.test Mean in category  Overall mean sd in category Overall sd
## Dim.3  4.263406         1.1726255 -3.467755e-15      0.5733819   1.185292
## Dim.2 -2.214701        -0.6773432 -8.316349e-15      1.2114992   1.318003
```

```
##            p.value
## Dim.3 0.0000201334
## Dim.2 0.0267805884
##
## $quanti$`4`
##         v.test Mean in category Overall mean sd in category Overall sd
## Dim.1 4.118906           4.1926 5.507535e-16      0.3060936   1.808841
##            p.value
## Dim.1 3.806749e-05
##
##
## attr(,"class")
## [1] "catdes" "list "
```

Construction de l'arbre sur le plan 1-3

```
plot(res.hcpc,axes=c(1,3))
```



Description des classes par les individus

```
res.hcpc$desc.ind
```

```
## $para
## Cluster: 1
##       Uldal      BARRAS   Karlivans       HERNU BOURGUIGNON
##    1.435595    1.538945    1.548147    1.757527    2.451654
## ------------------------------------------------------------
## Cluster: 2
##     Hernu        Qi    Barras   Smirnov Zsivoczky
##  1.503952  1.656106  1.663715  1.745507  1.819434
## ------------------------------------------------------------
## Cluster: 3
##   Schwarzl Schoenbeck    WARNERS  Pogorelov  Averyanov
##   1.088834   1.562252   1.651209   2.072228   2.143060
## ------------------------------------------------------------
```

```
## Cluster: 4
##     Clay   Sebrle   Karpov
## 1.535603 1.692349 2.569576
##
## $dist
## Cluster: 1
##     Casarsa BOURGUIGNON Korkizoglou   MARTINEAU        NOOL
##    5.083775    4.582906    4.041168    3.897578    3.884405
## ------------------------------------------------------------
## Cluster: 2
##       Smith Parkhomenko   Zsivoczky       Macey     Bernard
##    4.400988    3.607607    3.489667    3.365448    3.291086
## ------------------------------------------------------------
## Cluster: 3
##     CLAY    Drews   KARPOV     Nool    Terek
## 4.165973 4.111160 3.918168 3.894664 3.545632
## ------------------------------------------------------------
## Cluster: 4
##   Karpov   Sebrle     Clay
## 4.780916 4.610899 4.508333
```