Hierarchical ascendant clustering with FactoMineR (Temperature example)

Magalie Houée-Bigot & François Husson

Import data

```
setwd("C:/Users/houee/Downloads") # select the working directory
temperature <- read.table("data_clustering_Temperature.csv",
        header=TRUE, sep=";", dec=".", row.names=1)
```

`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

It is important to check that the import is well done

```
summary(temperature)
```

```
##     January          February          March             April
##  Min.   :-9.300   Min.   :-7.900   Min.   :-3.700   Min.   : 2.900
##  1st Qu.:-1.550   1st Qu.:-0.150   1st Qu.: 1.600   1st Qu.: 7.250
##  Median : 0.200   Median : 1.900   Median : 5.400   Median : 8.900
##  Mean   : 1.346   Mean   : 2.217   Mean   : 5.229   Mean   : 9.283
##  3rd Qu.: 4.900   3rd Qu.: 5.800   3rd Qu.: 8.500   3rd Qu.:12.050
##  Max.   :10.700   Max.   :11.800   Max.   :14.100   Max.   :16.900
##       May              June            July            August
##  Min.   : 6.50   Min.   : 9.30   Min.   :11.10   Min.   :10.60
##  1st Qu.:12.15   1st Qu.:15.40   1st Qu.:17.30   1st Qu.:16.65
##  Median :13.80   Median :16.90   Median :18.90   Median :18.30
##  Mean   :13.91   Mean   :17.41   Mean   :19.62   Mean   :18.98
##  3rd Qu.:16.35   3rd Qu.:19.80   3rd Qu.:21.75   3rd Qu.:21.60
##  Max.   :20.90   Max.   :24.50   Max.   :27.40   Max.   :27.20
##    September        October         November         December
##  Min.   : 7.90   Min.   : 4.50   Min.   :-1.100   Min.   :-6.00
##  1st Qu.:13.00   1st Qu.: 8.65   1st Qu.: 3.200   1st Qu.: 0.25
##  Median :14.80   Median :10.20   Median : 5.100   Median : 1.70
##  Mean   :15.63   Mean   :11.00   Mean   : 6.066   Mean   : 2.88
##  3rd Qu.:18.25   3rd Qu.:13.30   3rd Qu.: 7.900   3rd Qu.: 5.40
##  Max.   :24.30   Max.   :19.40   Max.   :14.900   Max.   :12.00
##      Annual        Amplitude        Latitude       Longitude
##  Min.   : 4.50   Min.   :10.20   Min.   :37.20   Min.   : 0.00
##  1st Qu.: 7.75   1st Qu.:14.90   1st Qu.:43.90   1st Qu.: 5.05
##  Median : 9.70   Median :18.50   Median :50.00   Median :10.50
##  Mean   :10.27   Mean   :18.32   Mean   :49.04   Mean   :13.01
##  3rd Qu.:12.65   3rd Qu.:21.45   3rd Qu.:53.35   3rd Qu.:19.30
##  Max.   :18.20   Max.   :27.60   Max.   :64.10   Max.   :37.60
##     Area
##  East : 8
```

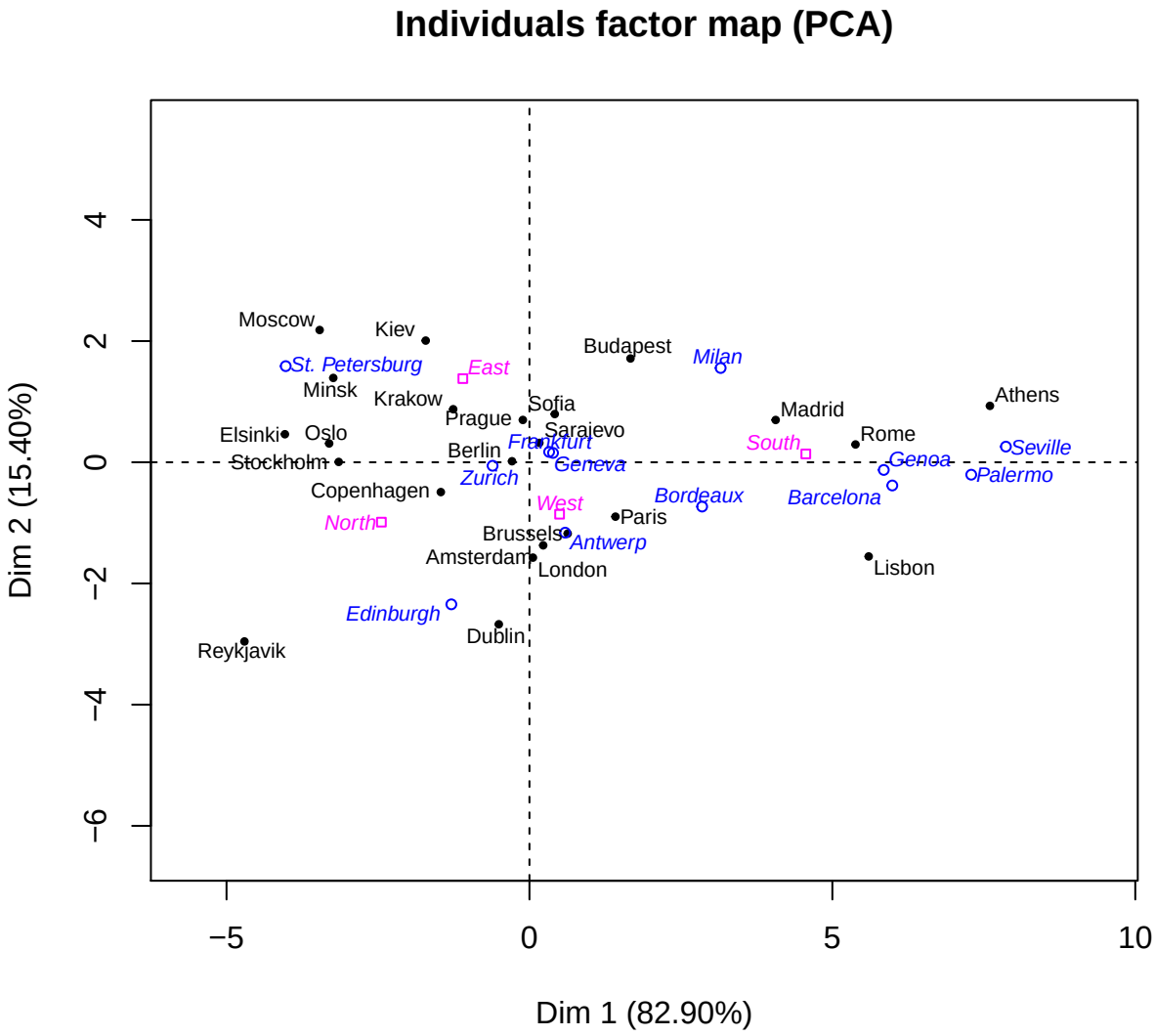
```
##  North: 8
##  South:10
##  West : 9
##
##
```

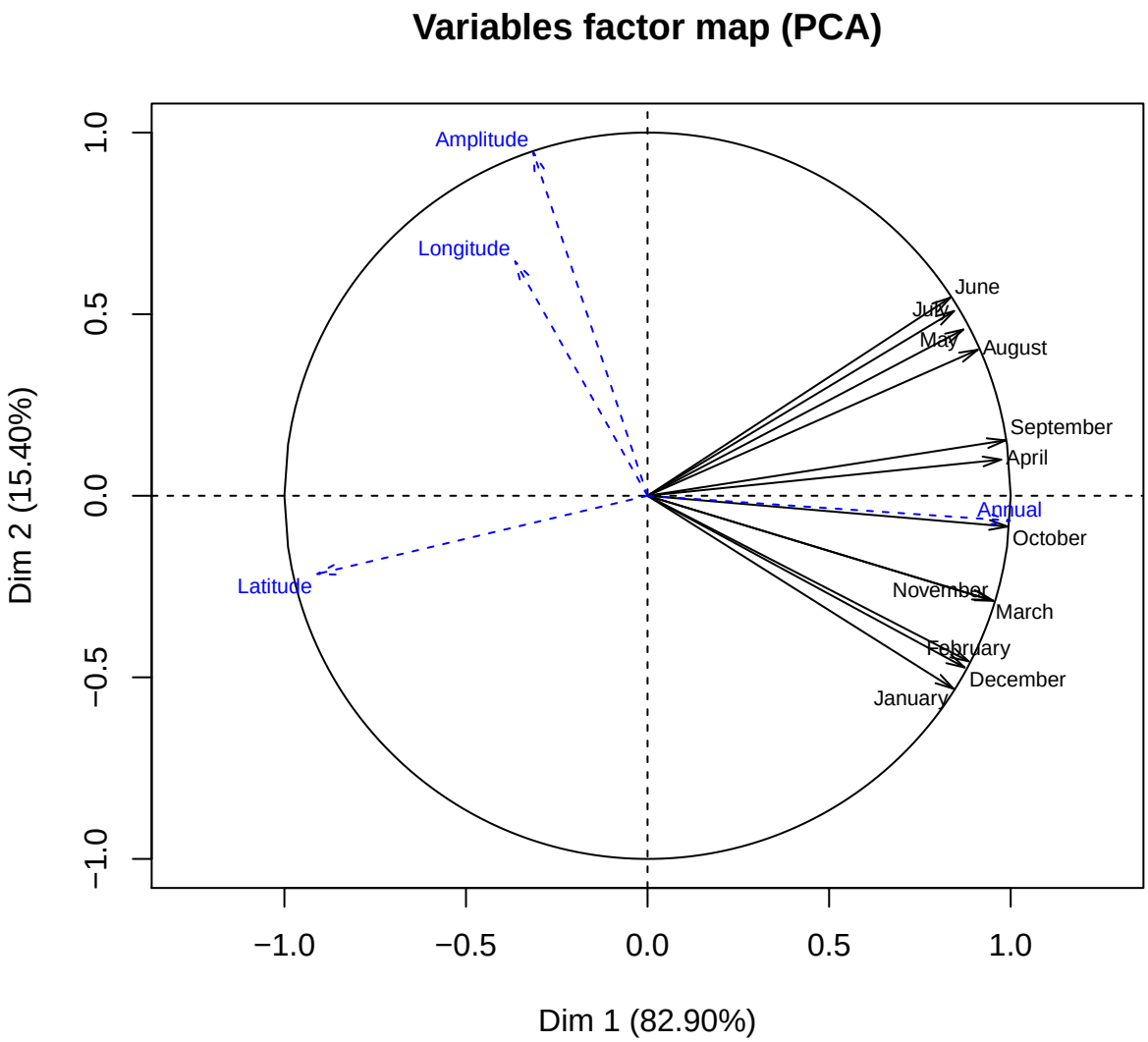
Loading FactoMineR

```
library(FactoMineR)
```

PCA

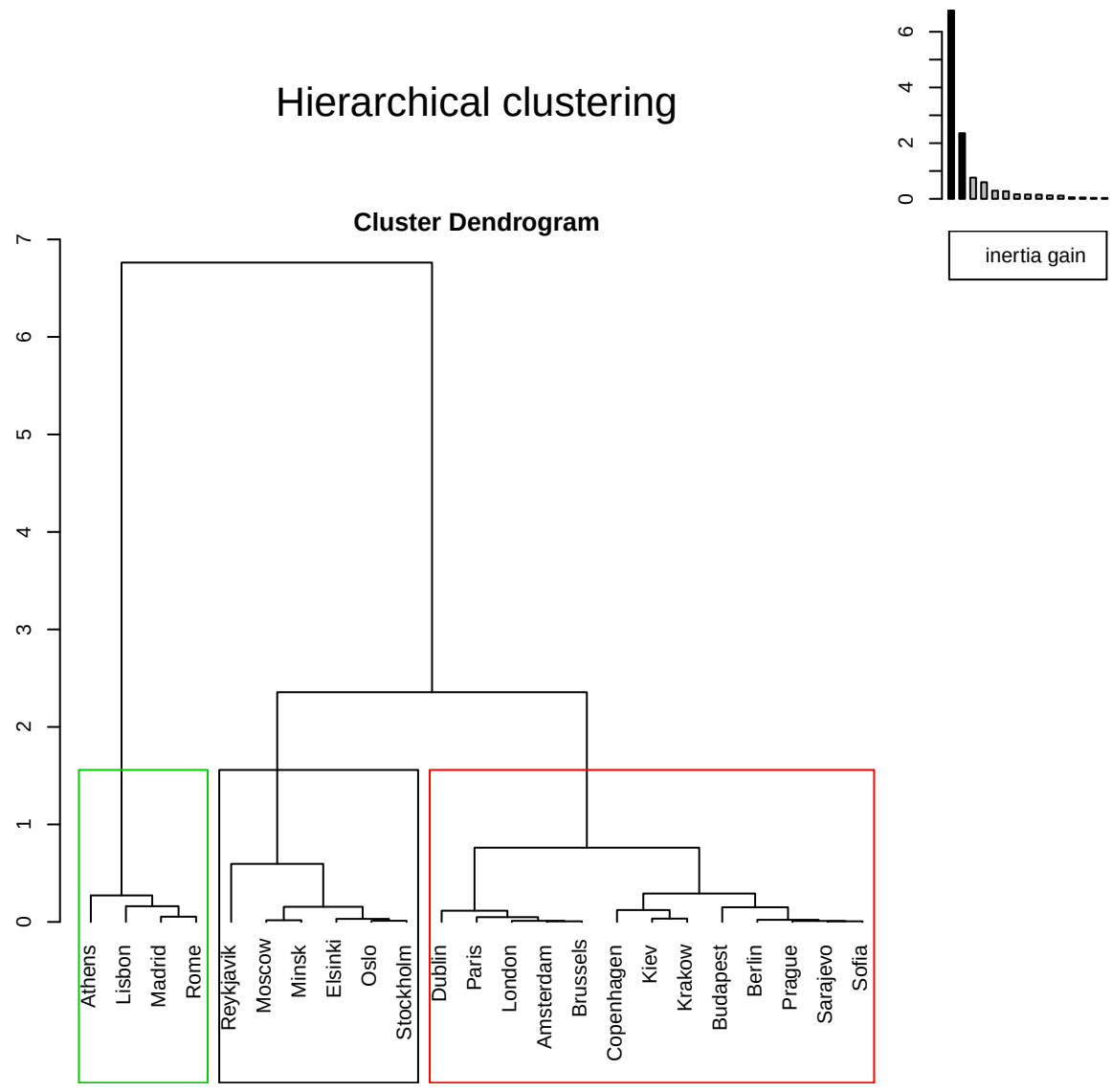
```
res <- PCA(temperature, ind.sup=24:35, quanti.sup=13:16,quali.sup=17, ncp=8)
```


Individuals factor map (PCA)
Moscow
Kiev
St. Petersburg
East
Minsk
Krakow
Budapest
Milan
Athens
Sofia
Prague
Madrid
Elsinki
Oslo
Sarajevo
Rome
Berlin
Frankfurt
South
Seville
Stockholm
Zurich
Geneva
Genoa
Palermo
Copenhagen
West
Bordeaux
Barcelona
Paris
North
Brussels
Antwerp
Amsterdam
London
Lisbon
Edinburgh
Dublin
Reykjavik
Dim 2 (15.40%)
Dim 1 (82.90%)
4
2
0
−2
−4
−6
−5
0
5
10

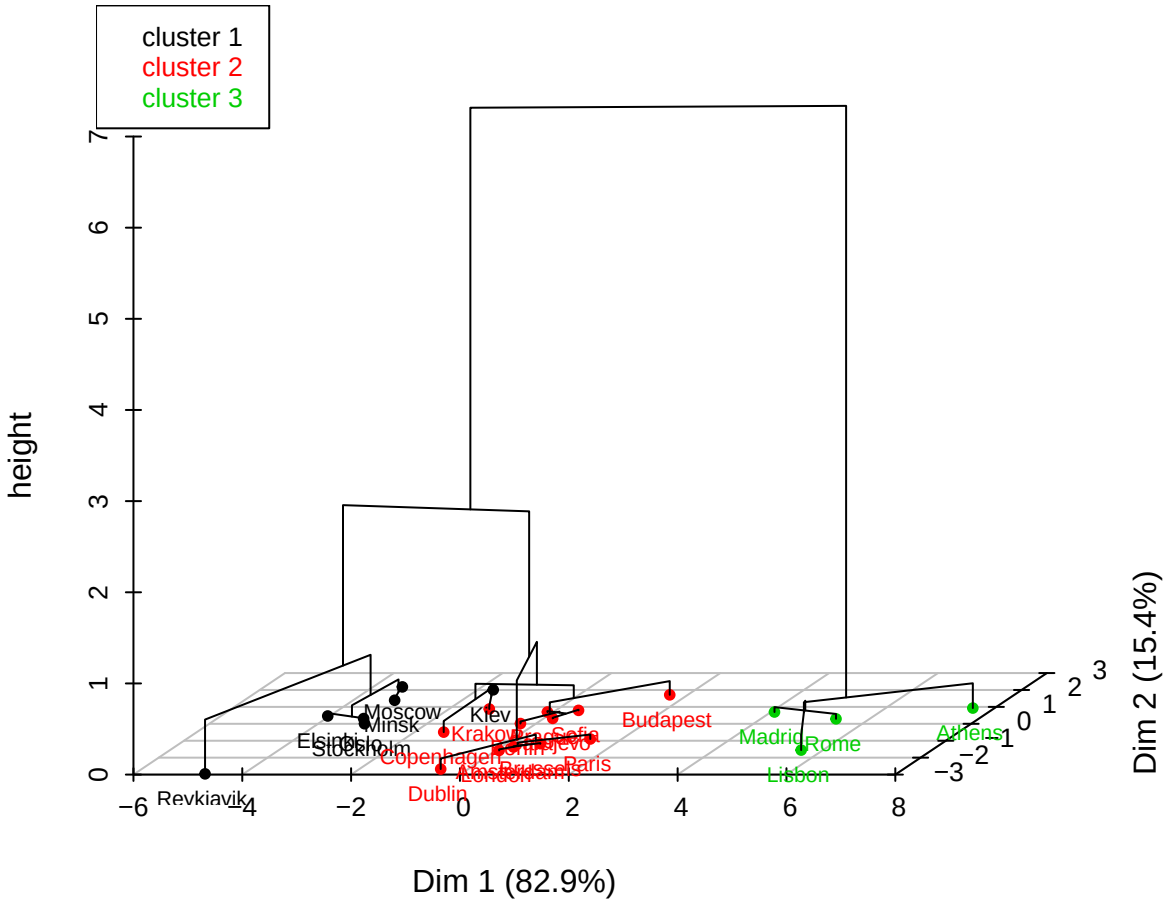


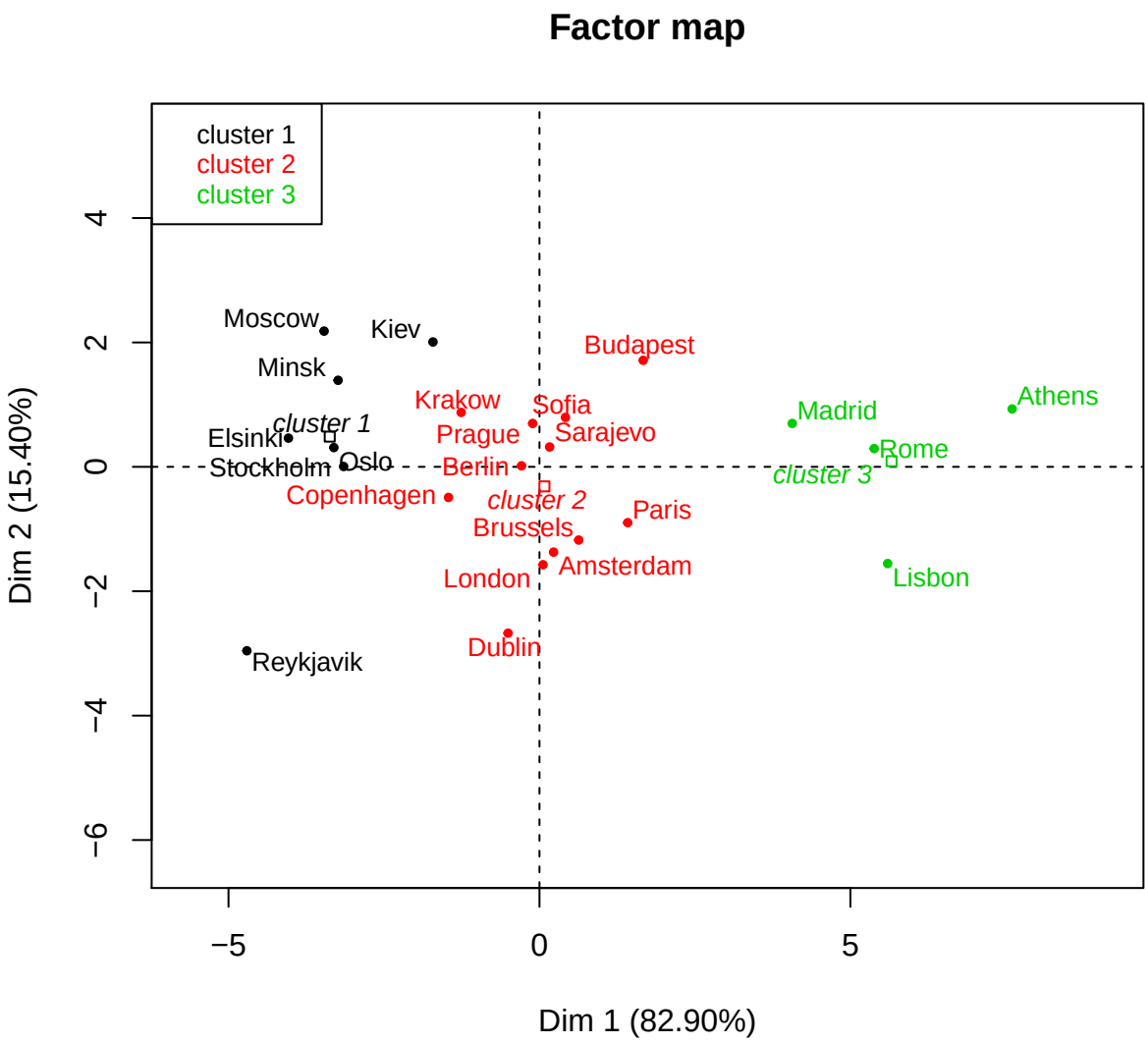
Hierarchical ascendant clustering

```
res.hcpc <- HCPC(res)
```


Hierarchical clustering
Cluster Dendrogram
0
1
2
3
4
5
6
7
Athens
Lisbon
Madrid
Rome
Reykjavik
Moscow
Minsk
Elsinki
Oslo
Stockholm
Dublin
Paris
London
Amsterdam
Brussels
Copenhagen
Kiev
Krakow
Budapest
Berlin
Prague
Sarajevo
Sofia
0
2
4
6
inertia gain

Hierarchical clustering on the factor map





Outputs

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

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

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

```
##             January February March April  May June July August September
## Amsterdam       2.9      2.5   5.7   8.2 12.5 14.8 17.1   17.1      14.5
## Athens          9.1      9.7  11.7  15.4 20.1 24.5 27.4   27.2      23.8
## Berlin         -0.2      0.1   4.4   8.2 13.8 16.0 18.3   18.0      14.4
## Brussels        3.3      3.3   6.7   8.9 12.8 15.6 17.8   17.8      15.0
```

```
## Budapest       -1.1      0.8   5.5  11.6 17.0 20.2 22.0   21.3      16.9
## Copenhagen     -0.4     -0.4   1.3   5.8 11.1 15.4 17.1   16.6      13.3
## Dublin          4.8      5.0   5.9   7.8 10.4 13.3 15.0   14.6      12.7
## Elsinki        -5.8     -6.2  -2.7   3.1 10.2 14.0 17.2   14.9       9.7
## Kiev           -5.9     -5.0  -0.3   7.4 14.3 17.8 19.4   18.5      13.7
## Krakow         -3.7     -2.0   1.9   7.9 13.2 16.9 18.4   17.6      13.7
## Lisbon         10.5     11.3  12.8  14.5 16.7 19.4 21.5   21.9      20.4
## London          3.4      4.2   5.5   8.3 11.9 15.1 16.9   16.5      14.0
## Madrid          5.0      6.6   9.4  12.2 16.0 20.8 24.7   24.3      19.8
## Minsk          -6.9     -6.2  -1.9   5.4 12.4 15.9 17.4   16.3      11.6
## Moscow         -9.3     -7.6  -2.0   6.0 13.0 16.6 18.3   16.7      11.2
## Oslo           -4.3     -3.8  -0.6   4.4 10.3 14.9 16.9   15.4      11.1
## Paris           3.7      3.7   7.3   9.7 13.7 16.5 19.0   18.7      16.1
## Prague         -1.3      0.2   3.6   8.8 14.3 17.6 19.3   18.7      14.9
## Reykjavik      -0.3      0.1   0.8   2.9  6.5  9.3 11.1   10.6       7.9
## Rome            7.1      8.2  10.5  13.7 17.8 21.7 24.4   24.1      20.9
## Sarajevo       -1.4      0.8   4.9   9.3 13.8 17.0 18.9   18.7      15.2
## Sofia          -1.7      0.2   4.3   9.7 14.3 17.7 20.0   19.5      15.8
## Stockholm      -3.5     -3.5  -1.3   3.5  9.2 14.6 17.2   16.0      11.7
##              October November December Annual Amplitude Latitude Longitude
## Amsterdam       11.4      7.0      4.4    9.9      14.6     52.2       4.5
## Athens          19.2     14.6     11.0   17.8      18.3     37.6      23.5
## Berlin          10.0      4.2      1.2    9.1      18.5     52.3      13.2
## Brussels        11.1      6.7      4.4   10.3      14.4     50.5       4.2
## Budapest        11.3      5.1      0.7   10.9      23.1     47.3      19.0
## Copenhagen       8.8      4.1      1.3    7.8      17.5     55.4      12.3
## Dublin           9.7      6.7      5.4    9.3      10.2     53.2       6.1
## Elsinki          5.2      0.1     -2.3    4.8      23.4     60.1      25.0
## Kiev             7.5      1.2     -3.6    7.1      25.3     50.3      30.3
## Krakow           8.6      2.6     -1.7    7.7      22.1     50.0      19.6
## Lisbon          17.4     13.7     11.1   15.9      11.4     38.4       9.1
## London          10.2      6.3      4.4    9.7      13.5     51.4       0.0
## Madrid          13.9      8.7      5.4   13.9      19.7     40.2       3.4
## Minsk            5.8      0.1     -4.2    5.5      24.3     53.5      27.3
## Moscow           5.1     -1.1     -6.0    5.1      27.6     55.7      37.6
## Oslo             5.7      0.5     -2.9    5.6      21.2     59.5      10.5
## Paris           12.5      7.3      5.2   11.2      15.3     48.5       2.2
## Prague           9.4      3.8      0.3    9.2      20.6     50.0      14.2
## Reykjavik        4.5      1.7      0.2    4.6      11.4     64.1      21.6
## Rome            16.5     11.7      8.3   15.4      17.3     41.5      12.3
## Sarajevo        10.5      5.1      0.8    9.4      20.3     43.5      18.3
## Sofia           10.7      5.0      0.6    9.6      21.7     42.4      23.2
## Stockholm        6.5      1.7     -1.6    5.8      20.7     59.2      18.0
##               Area clust
## Amsterdam     West     2
## Athens       South     3
## Berlin        West     2
## Brussels      West     2
## Budapest      East     2
## Copenhagen   North     2
## Dublin       North     2
## Elsinki      North     1
## Kiev          East     1
## Krakow        East     2
```

```
## Lisbon      South     3
## London      North     2
## Madrid      South     3
## Minsk        East     1
## Moscow       East     1
## Oslo        North     1
## Paris        West     2
## Prague       East     2
## Reykjavik   North     1
## Rome        South     3
## Sarajevo    South     2
## Sofia        East     2
## Stockholm   North     1
```

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

```
## $test.chi2
##          p.value df
## Area 0.001195843  6
##
## $category
## $category$`1`
## NULL
##
## $category$`2`
## NULL
##
## $category$`3`
##            Cla/Mod Mod/Cla   Global      p.value   v.test
## Area=South      80     100 21.73913 0.0005646527 3.448048
##
##
## $quanti.var
##                Eta2      P-value
## Annual    0.9177156 1.422898e-11
## October   0.8989724 1.107640e-10
## March     0.8864732 3.556182e-10
## November  0.8707485 1.301232e-09
## September 0.8559706 3.841596e-09
## April     0.8353281 1.466202e-08
## February  0.8246130 2.754070e-08
## December  0.7730136 3.630759e-07
## January   0.7476664 1.046528e-06
## Latitude  0.7378117 1.535089e-06
## August    0.7159888 3.414722e-06
## July      0.6309204 4.690301e-05
## May       0.5860137 1.478632e-04
## June      0.5752594 1.910900e-04
## Longitude 0.3720657 9.531078e-03
##
## $quanti
## $quanti$`1`
##              v.test Mean in category Overall mean sd in category
## Latitude   3.173406        57.485714   50.2956522      4.2809974
```

```
## Longitude  2.858559        24.328571   15.4521739      8.1046220
## Amplitude  2.142021        21.985714   18.8000000      4.8430889
## July      -1.994799        16.785714   18.9260870      2.4549824
## June      -2.058250        14.728571   16.7652174      2.5172061
## August    -2.479950        15.485714   18.3043478      2.2585890
## May       -2.549115        10.842857   13.2739130      2.4289075
## September -3.137128        10.985714   14.7086957      1.6685568
## January   -3.255703        -5.142857    0.1739130      2.6250753
## December  -3.269173        -2.914286    1.8434783      1.8302448
## November  -3.359278         0.600000    5.0782609      0.9396048
## Annual    -3.373602         5.500000    9.3739130      0.7671841
## April     -3.387118         4.671429    8.3782609      1.5489298
## February  -3.442119        -4.600000    0.9565217      2.3366643
## October   -3.453788         5.757143   10.0652174      0.9194053
## March     -3.675091        -1.142857    4.0608696      1.1043513
##           Overall sd      p.value
## Latitude    7.029223 0.0015066157
## Longitude   9.633631 0.0042557039
## Amplitude   4.614062 0.0321917637
## July        3.328822 0.0460648072
## June        3.069855 0.0395661536
## August      3.526111 0.0131400659
## May         2.958733 0.0107996531
## September   3.681790 0.0017061164
## January     5.066447 0.0011311207
## December    4.515079 0.0010786223
## November    4.135841 0.0007814643
## Annual      3.562512 0.0007419150
## April       3.395260 0.0007063103
## February    5.008152 0.0005771764
## October     3.869795 0.0005527723
## March       4.392854 0.0002377646
##
## $quanti$`2`
##              v.test Mean in category Overall mean sd in category
## Longitude -2.060647             11.4     15.45217       7.449161
##           Overall sd    p.value
## Longitude   9.633631 0.03933669
##
## $quanti$`3`
##              v.test Mean in category Overall mean sd in category
## Annual     3.851788           15.750    9.3739130       1.393736
## September  3.808964           21.225   14.7086957       1.536839
## October    3.717610           16.750   10.0652174       1.911151
## August     3.705134           24.375   18.3043478       1.883315
## November   3.692832           12.175    5.0782609       2.264260
## July       3.603579           24.500   18.9260870       2.089258
## April      3.531688           13.950    8.3782609       1.175798
## March      3.448552           11.100    4.0608696       1.274755
## February   3.434969            8.950    0.9565217       1.744276
## June       3.389407           21.600   16.7652174       1.864135
## December   3.387321            8.950    1.8434783       2.337199
## January    3.292484            7.925    0.1739130       2.076505
## May        3.183059           17.650   13.2739130       1.553222
```

```
## Latitude  -3.328226           39.425   50.2956522       1.523770
##           Overall sd      p.value
## Annual      3.562512 0.0001172584
## September   3.681790 0.0001395501
## October     3.869795 0.0002011160
## August      3.526111 0.0002112792
## November    4.135841 0.0002217705
## July        3.328822 0.0003138648
## April       3.395260 0.0004129164
## March       4.392854 0.0005636006
## February    5.008152 0.0005926217
## June        3.069855 0.0007004406
## December    4.515079 0.0007057887
## January     5.066447 0.0009930662
## May         2.958733 0.0014572796
## Latitude    7.029223 0.0008740097
##
##
## attr(,"class")
## [1] "catdes" "list "
```

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

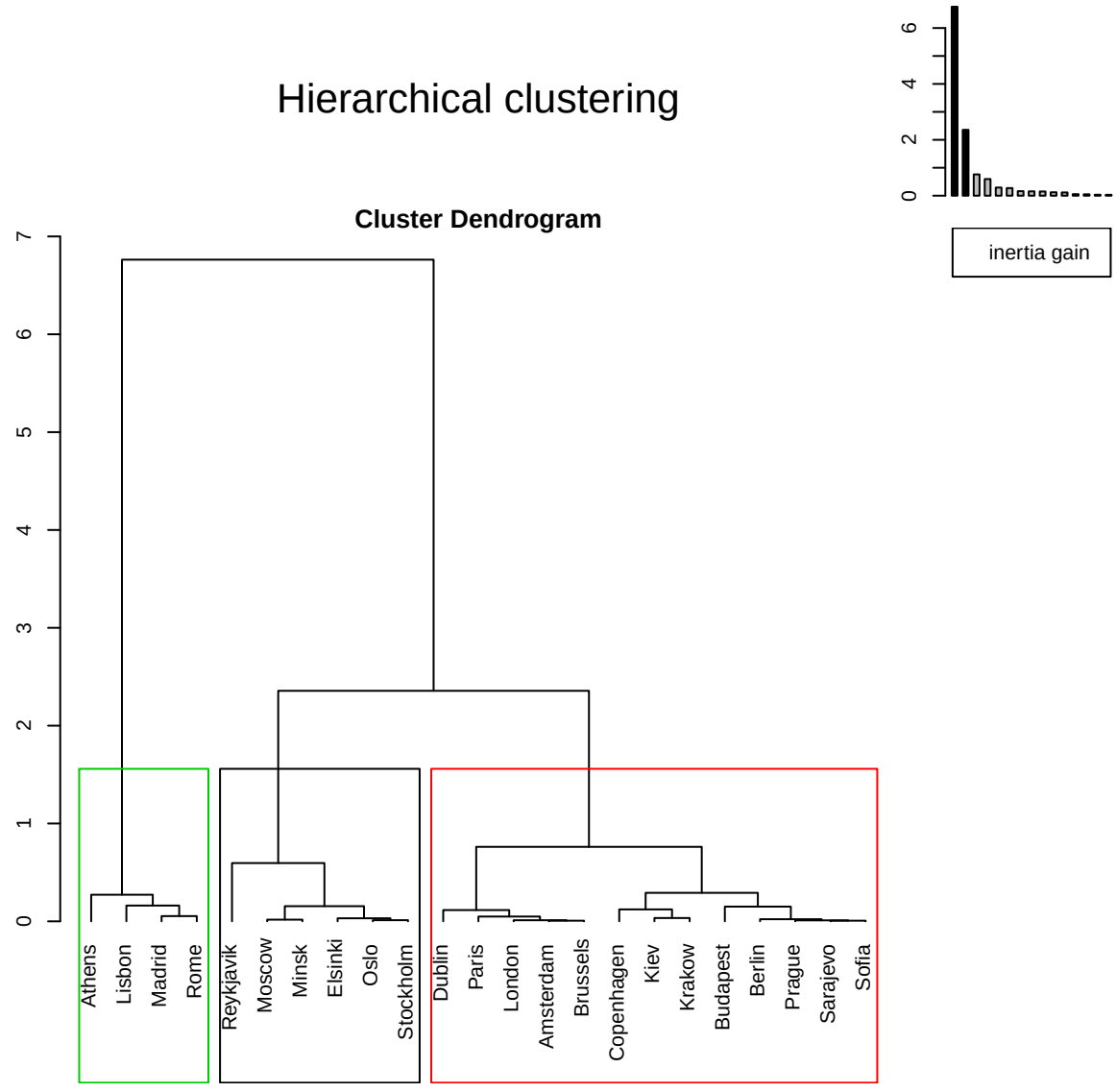
```
## $quanti.var
##            Eta2      P-value
## Dim.1 0.9086821 4.032342e-11
## Dim.3 0.2691961 4.345188e-02
##
## $quanti
## $quanti$`1`
##          v.test Mean in category Overall mean sd in category Overall sd
## Dim.1 -3.317358        -3.372519  9.98899e-16      0.8494564   3.154005
##            p.value
## Dim.1 0.0009087321
##
## $quanti$`2`
##          v.test Mean in category  Overall mean sd in category Overall sd
## Dim.3 -2.414108       -0.1750962 -2.289835e-16      0.2178523  0.3553249
##          p.value
## Dim.3 0.01577378
##
## $quanti$`3`
##         v.test Mean in category Overall mean sd in category Overall sd
## Dim.1 3.863084         5.661507  9.98899e-16       1.264335   3.154005
##            p.value
## Dim.1 0.0001119644
##
##
## attr(,"class")
## [1] "catdes" "list "
```

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

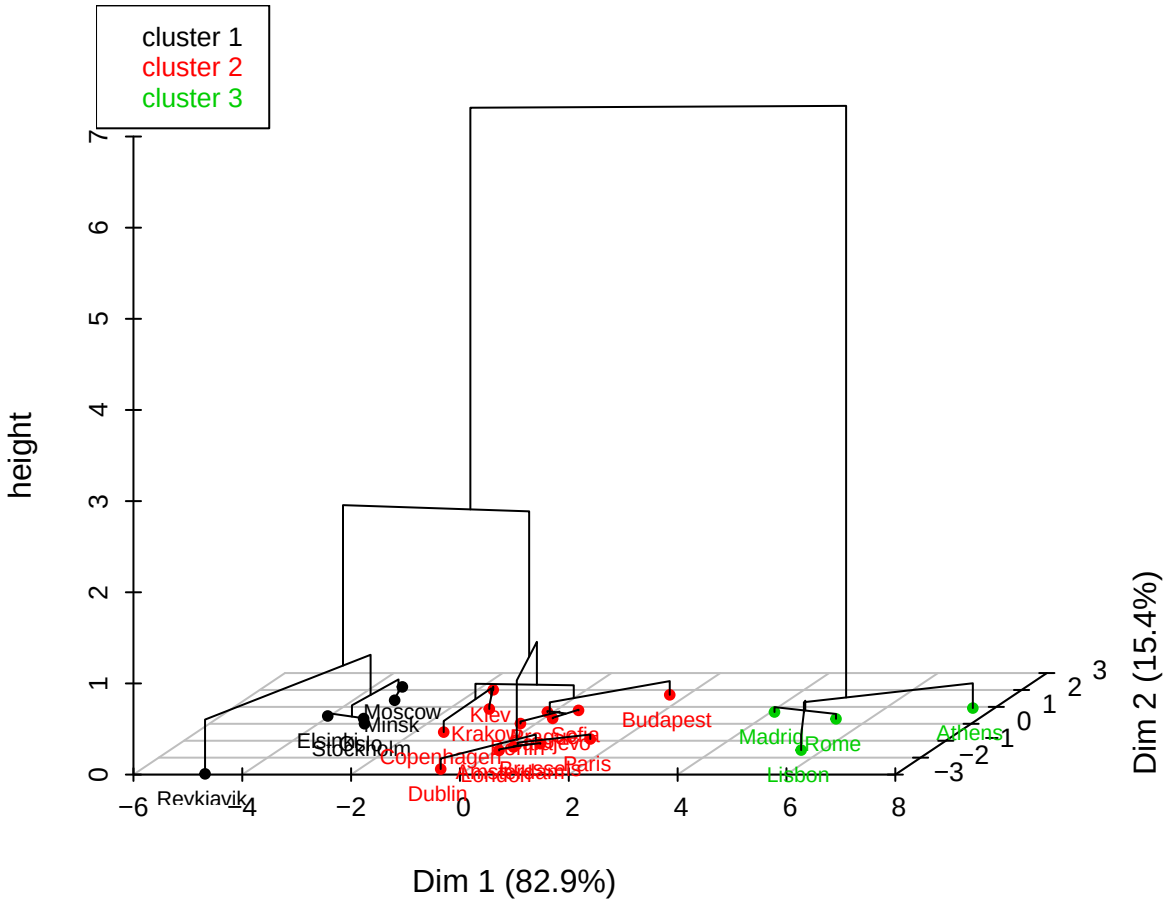
```
## $para
## Cluster: 1
##      Oslo     Elsinki Stockholm      Minsk      Moscow
##  0.3335877  0.8818784  0.9221233  0.9647962  1.7659010
## ------------------------------------------------------
## Cluster: 2
##    Berlin   Sarajevo    Brussels    Prague  Amsterdam
##  0.5741752  0.7123843  1.0336686  1.0542124  1.1229268
## ------------------------------------------------------
## Cluster: 3
##     Rome    Lisbon    Madrid    Athens
## 0.3573358 1.7366354 1.8348415 2.1666454
##
## $dist
## Cluster: 1
## Reykjavik      Moscow    Elsinki     Minsk       Oslo
##   5.473906   4.339616   4.280132   3.738331   3.483769
## ------------------------------------------------------
## Cluster: 2
##     Paris   Budapest    Brussels    Dublin  Amsterdam
##   4.380902   4.374246   4.351300   4.284171   4.077312
## ------------------------------------------------------
## Cluster: 3
##  Athens   Lisbon     Rome   Madrid
## 7.665137 5.660708 5.350332 4.216008
```

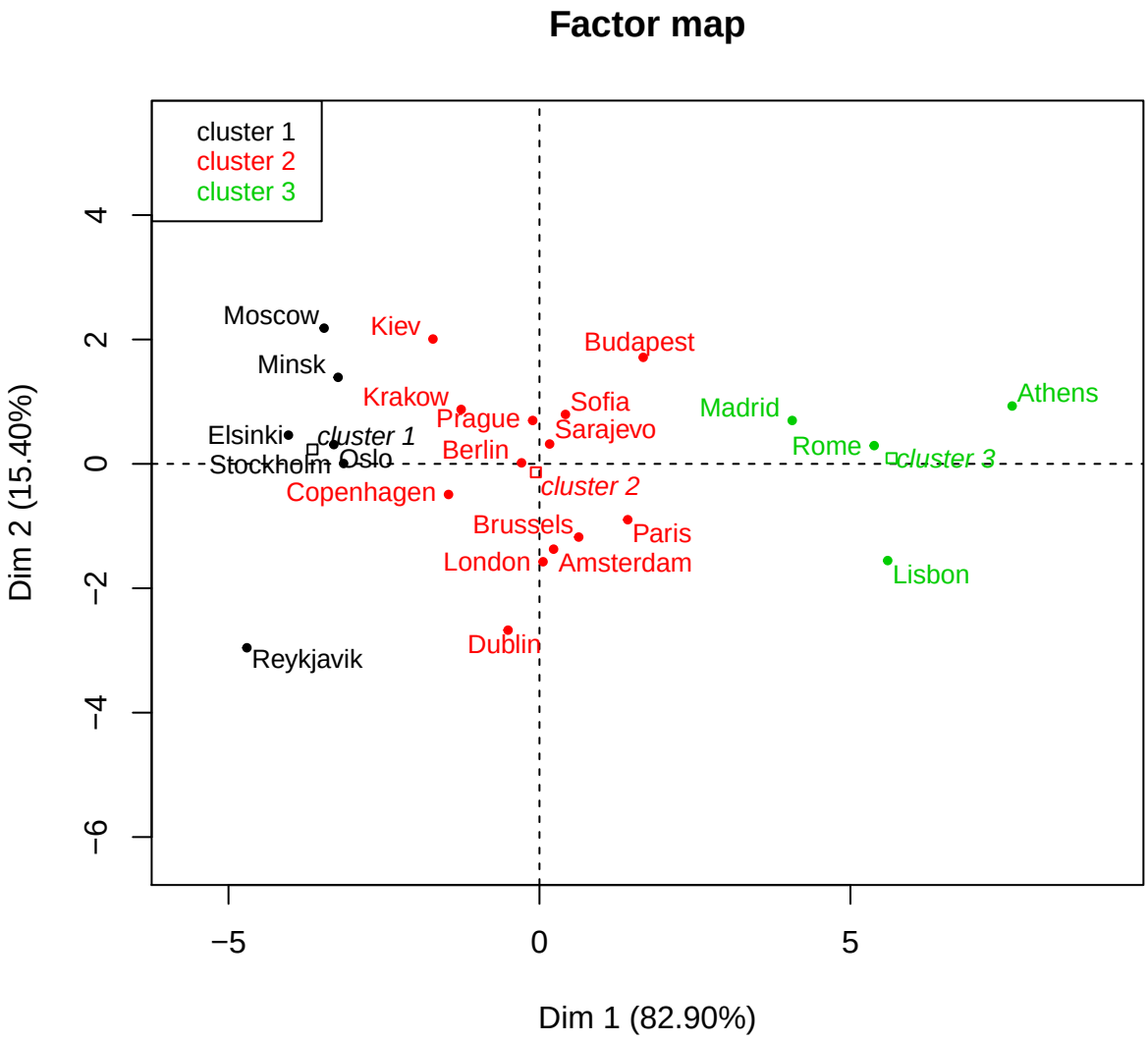
Hierarchical ascendant clustering without consolidation

```
res.hcpc <- HCPC(res, consol=FALSE)
```


Hierarchical clustering
Cluster Dendrogram
0
1
2
3
4
5
6
7
Athens
Lisbon
Madrid
Rome
Reykjavik
Moscow
Minsk
Elsinki
Oslo
Stockholm
Dublin
Paris
London
Amsterdam
Brussels
Copenhagen
Kiev
Krakow
Budapest
Berlin
Prague
Sarajevo
Sofia
0
2
4
6
inertia gain

Hierarchical clustering on the factor map




Factor map
cluster 1
cluster 2
cluster 3
Moscow
Kiev
Budapest
Minsk
Krakow
Sofia
Madrid
Athens
Prague
Sarajevo
Elsinki
cluster 1
Berlin
Rome
Oslo
cluster 3
Stockholm
Copenhagen
cluster 2
Brussels
Paris
London
Amsterdam
Lisbon
Dublin
Reykjavik
Dim 2 (15.40%)
4
2
0
−2
−4
−6
−5
0
5
Dim 1 (82.90%)