# OpenSpat Clustering example

Pr Yves Brostaux, ULiège

Openspat ULisboa

June 2019

# Prepare data

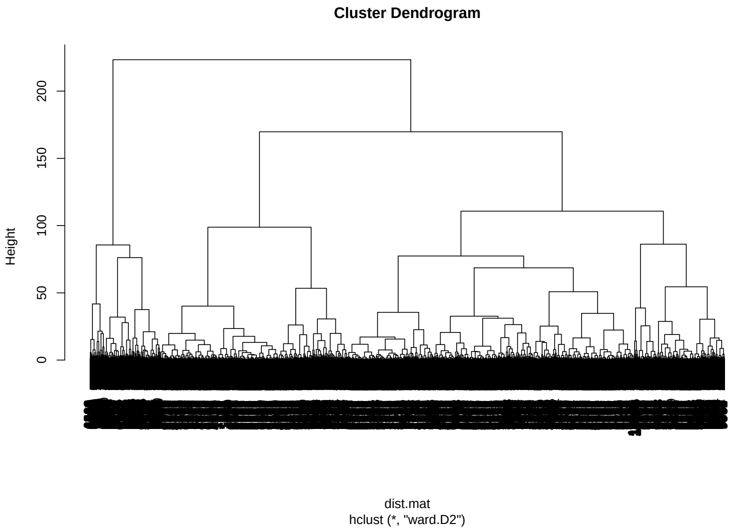
We use the same bioclimatic data as for PCA. But most clustering methods don't scale automatically the data to zero mean and unitary standard deviation. So we need to do it manually before starting the clustering.

```r
# Standardize the data (zero mean and unit
# standard deviation)
data.std <- scale(climdat)

# Euclidian distance matrix (on the scaled data)
dist.mat <- dist(data.std)
```

# Hierarchical clustering by Ward's method

```r
# Hierarchical clustering (Ward's method)
clim.hc <- hclust(dist.mat, method="ward.D2")

# Resulting dendrogram
plot(clim.hc)
```

# Hierarchical clustering by Ward's method



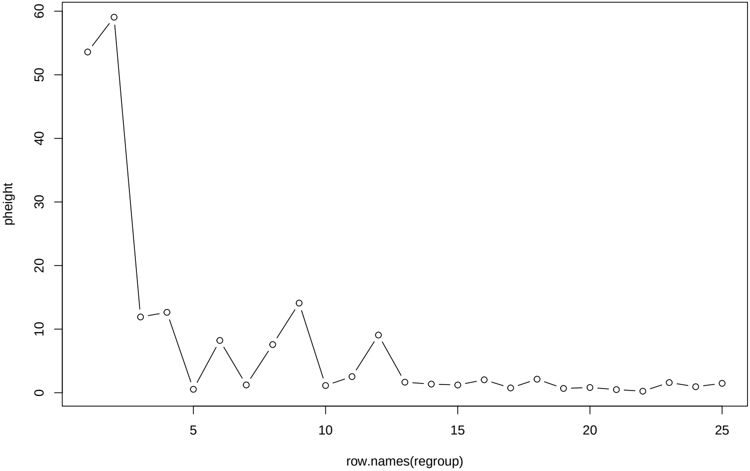
# Informations about successive mergings

```
# informations on successive merging steps
source("../hclust.info.R")
regroup <- hclust.info(clim.hc)

head(regroup)

plot(height~row.names(regroup), data=regroup,
     subset=1:25, type="b")

plot(pheight~row.names(regroup), data=regroup,
     subset=1:25, type="b")
```

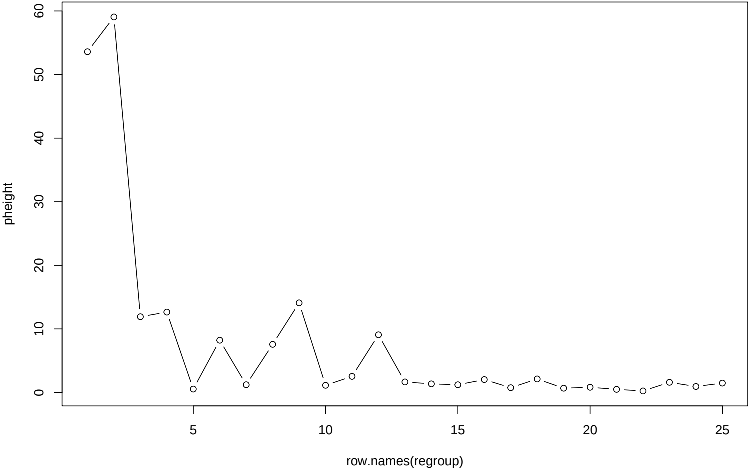
# Informations about successive mergings

```
##      height    pheight
## 1 223.36369 53.5875863
## 2 169.77611 59.0570374
## 3 110.71907 11.9151244
## 4  98.80394 12.6445357
## 5  86.15941  0.5402447
## 6  85.61916  8.2226364
```

# Informations about successive mergings



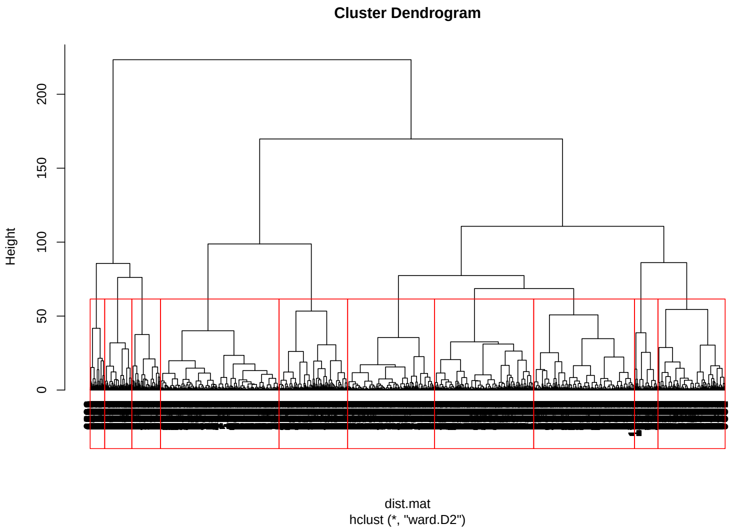
# Informations about successive mergings



# Create a partition

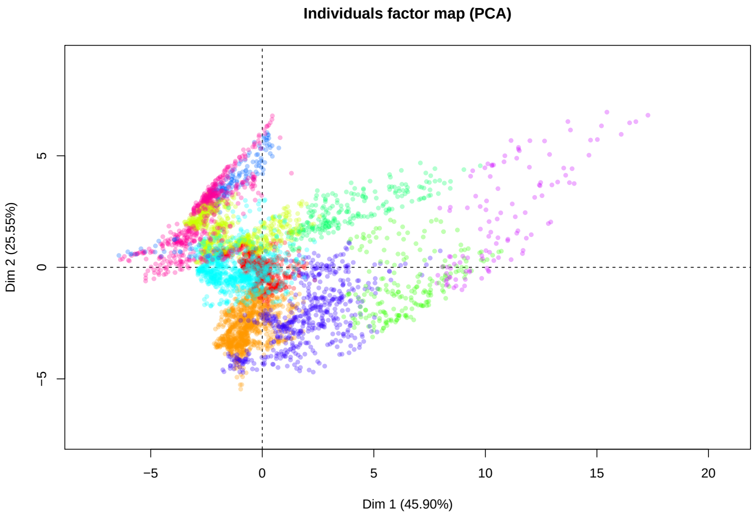
```
# partition in 10 groups
groups <- cutree(clim.hc, 10)
# visualize partition on dendrogram
plot(clim.hc, hang=-1)
rect.hclust(clim.hc,k=10)
```

# Create a partition



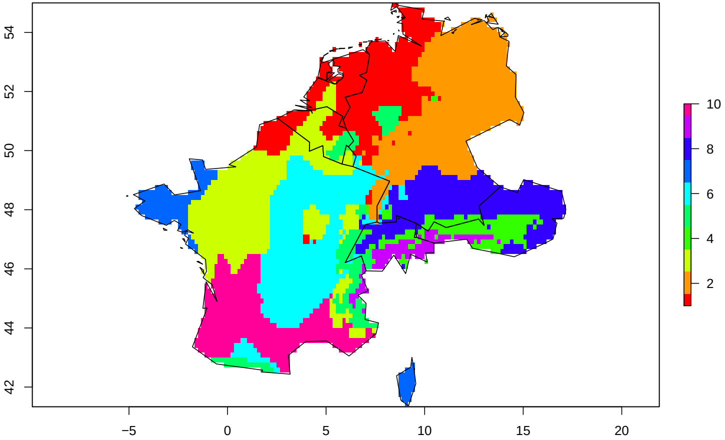
# visualize groups in the data space through PCA

```
# generate colors vector according to groups
col.gr <- rainbow(length(unique(groups)))[groups]

# plot groups in PCA space
plot(climdat.pca, choix="ind",
     col.ind=adjustcolor(col.gr, alpha.f = 0.3)
     label="none")
```


Individuals factor map (PCA)
Dim 2 (25.55%)
5
0
−5
−5
0
5
10
15
20
Dim 1 (45.90%)

# Visualize groups in the coordinate space

```r
climClust <- raster(climEur)
values(climClust)[-nai] <- groups
plot(climClust, col=rainbow(length(unique(groups))))
plot(eur_simpl, add=TRUE)
```


54
52
50
48
46
44
42
−5
0
5
10
15
20
10
8
6
4
2

```r
# Save the resulting typology map
writeRaster(climClust, "climTypo10.tif", "GTiff")
```