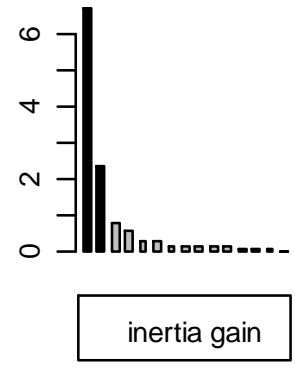
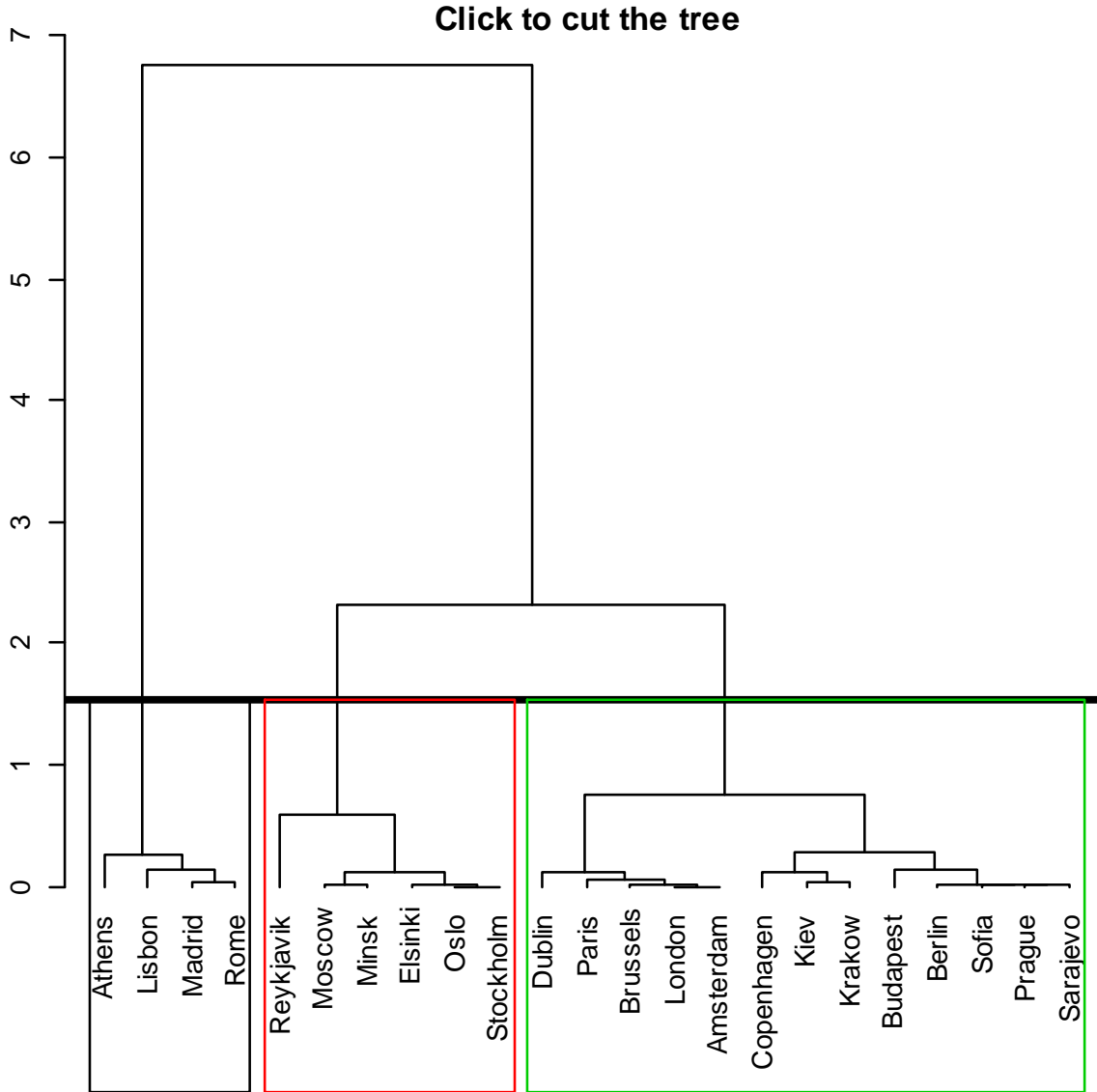


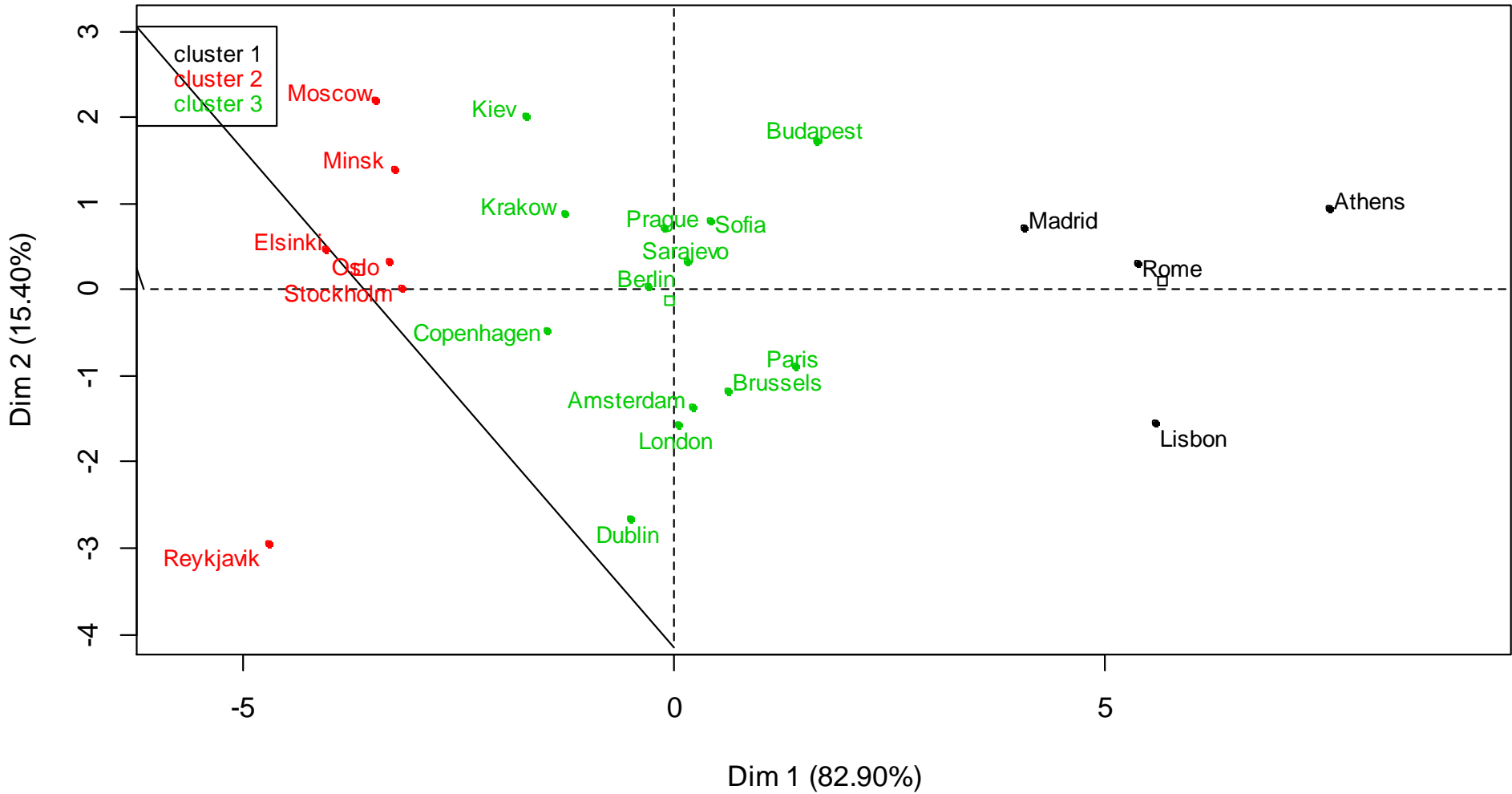
# Classification ascendante hiérarchique

# Hierarchical Clustering

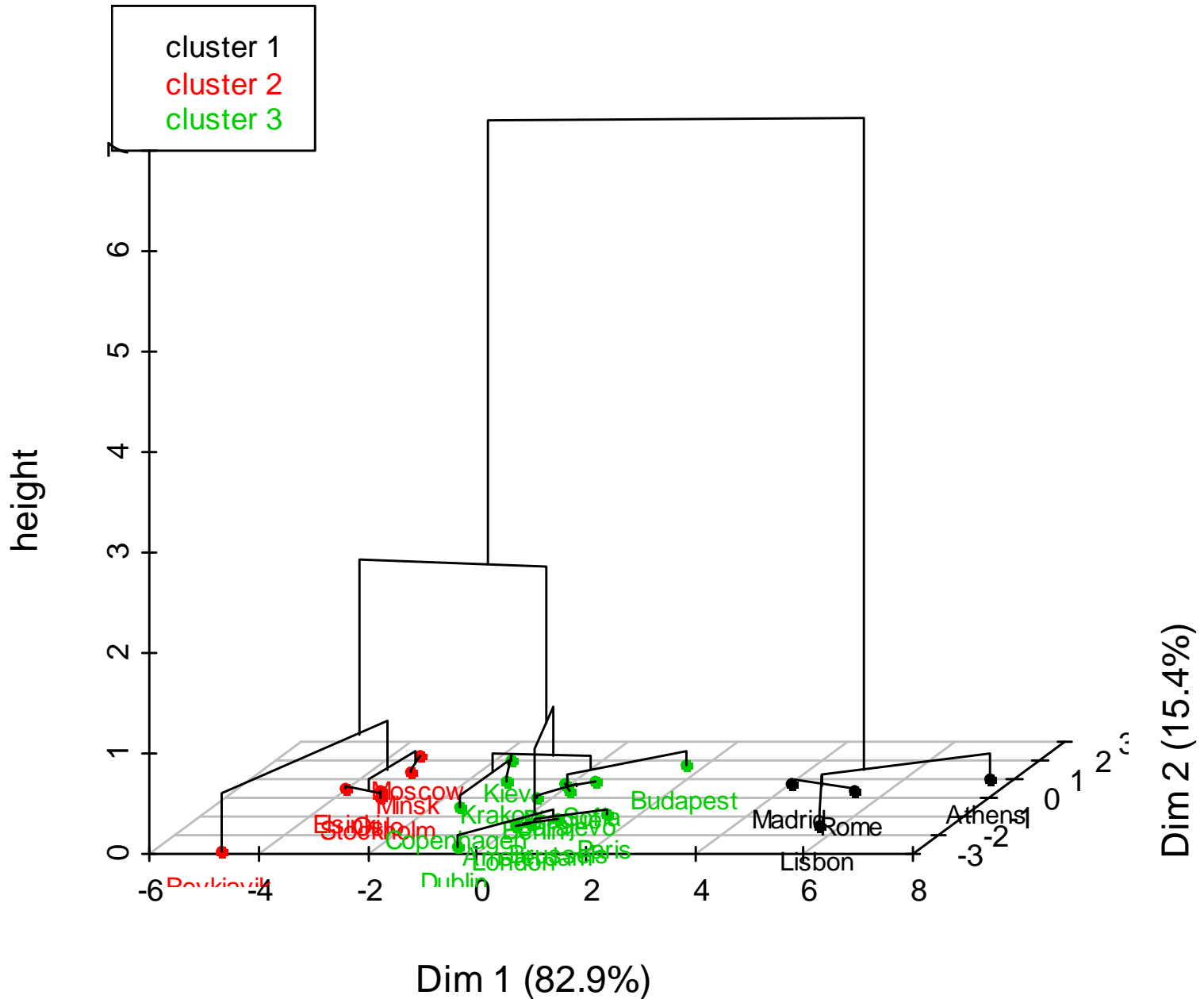


- `library(Rcmdr)`
- `library(FactoMineR)`
- `temperature <- read.table`  
(<http://factominer.free.fr/book/temperature.csv>,  
`header=TRUE, sep=";", dec=".", row.names=1`)
- `res.pca <- PCA(temperature[1:23,],  
quanti.sup=13:16, quali.sup=17, scale.unit=TRUE)`
- `res.hcpc <- HCPC(res.pca)`

# Factor map



# Hierarchical clustering on the factor map



# Construction d'un arbre hiérarchique

- On dispose d'un ensemble d'individus et des ressemblances entre individus mesurées par l'ACP
- On commence par regrouper les deux individus les plus proches, on constitue ainsi le premier nœud de l'arbre.
- On regroupe ensuite les deux éléments suivants les plus proches, le niveau d'agrégation est alors considéré comme plus élevé. On continue petit à petit jusqu'à tout agréger.
- La hauteur à laquelle on relie des éléments correspond à la ressemblance entre les éléments reliés.
- Un arbre peut être coupé pour faire apparaître une partition (un ensemble de classe), ce niveau de coupure est matérialisé par une ligne horizontale. Plus on coupe haut, moins le nombre de classes est important et plus la partition est grossière.

- Une partition est bonne si
  - À l'intérieur de chaque classe, la variabilité est faible, autrement dit si la variance des individus qui composent la classe est faible pour chaque variable
  - D'une classe à l'autre, la variabilité est grande, autrement dit si, pour chaque variable, la moyenne des individus qui composent une classe varie beaucoup d'une classe à l'autre
  - L'algorithme de Ward consiste à choisir à chaque pas le regroupement de classes tel que l'augmentation de l'inertie intra soit minimum

- `res.hcpc$desc.var$test.chi2`
- `res.hcpc$desc.var$category`
- `res.hcpc$desc.axes`
  
- `res.hcpc$data.clust`



# Régressions

- On dispose de plusieurs variables  $X_1, X_2, X_3$ , et  $Y$ , et on essaye de prédire les variations de  $Y$  à partir des variations de  $X_1, X_2, X_3$  .

- `library(questionr)`
- `data(hdv2003)`
- `data(rp99)`
- `d=hdv2003`
- `d2=rp99`

- `plot (d2$tx.chom,d2$proprio,pch=16,cex=0.7)`
- `M1=lm(d2$proprio~d2$tx.chom)`
- `summary(M1)`
- `abline(M1,col='red')`
  
- `plot (d2$cadres,d2$proprio,pch=16,cex=0.7)`
- `M2=lm(d2$proprio~d2$cadres)`
- `summary(M2)`
- `abline(M2,col='red')`
  
- `M3=lm(d2$proprio~d2$tx.chom+d2$cadres)`
- `summary(M3)`

# Régression logistique dichotomique

- `d$hard.rock2[d$hard.rock=='Non']=0`
- `d$hard.rock2[d$hard.rock=='Oui']=1`
  
- `d$sexe2[d$sexe=='Homme']=0`
- `d$sexe2[d$sexe=='Femme']=1`
  
- `M4=glm(d$hard.rock2~d$age + d$sexe2, family=binomial)`
- `summary(M4)`
- `odds.ratio(M4)`