UNIVERSITAT DE
BARCELONA

Cluster analysis

Píndoles d’estadística avançada
STeL (Març 2021)
Sessió 3

PROF. S. CIVIT

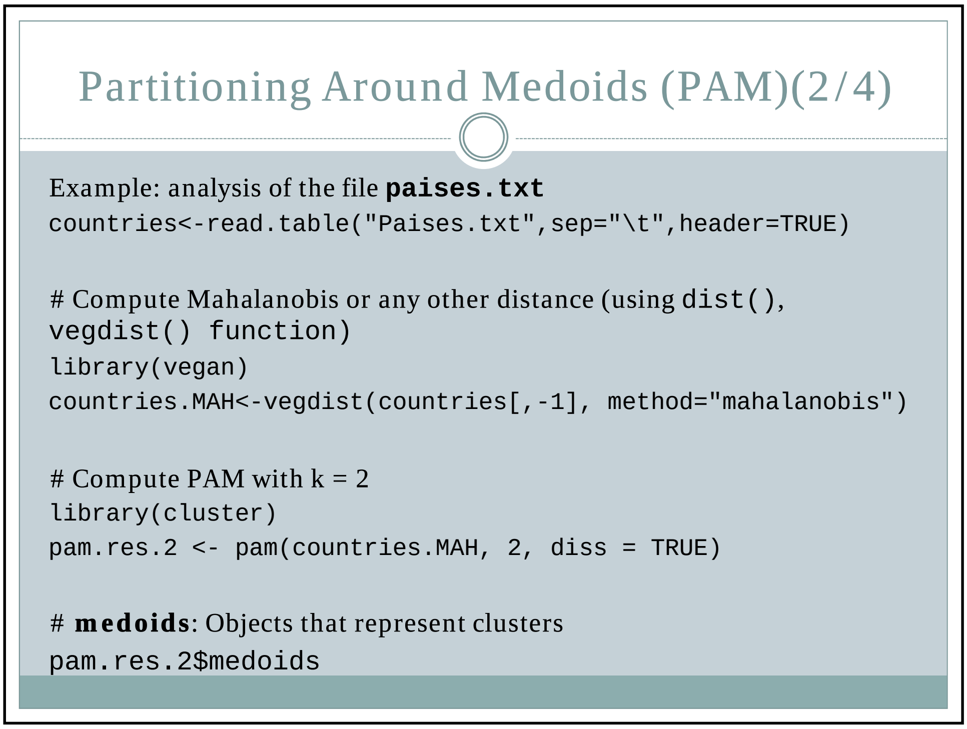
Partitioning Around Medoids (PAM)(1/4)

# Function `pam()`, simplified format:
`pam(x, k)`

# **x**: possible values includes:

- **Numeric data matrix or numeric data frame**: each row corresponds to an observation, and each column corresponds to a variable.
- **Dissimilarity matrix**: in this case **x** is typically the output of **daisy()** or **dist()**

# **k**: The number of clusters



Partitioning Around Medoids (PAM)(2/4)

Example: analysis of the file **`paises.txt`**

```
countries<-read.table("Paises.txt",sep="\t",header=TRUE)
```

# Compute Mahalanobis or any other distance (using `dist()`, `vegdist() function)`

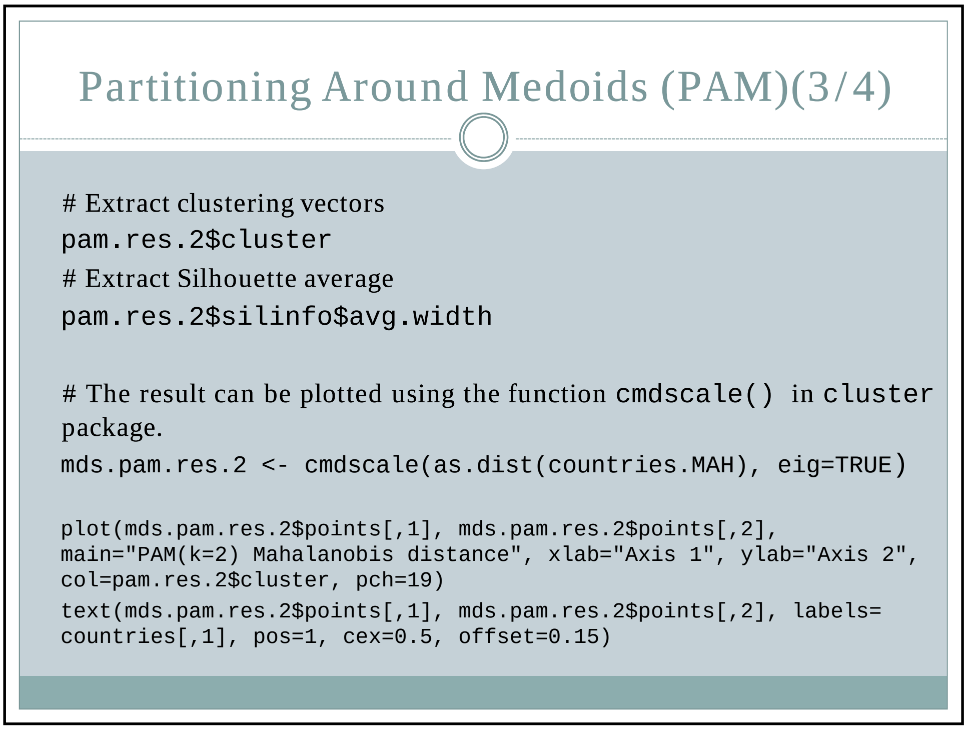
```
library(vegan)
countries.MAH<-vegdist(countries[,-1], method="mahalanobis")
```

# Compute PAM with k = 2

```
library(cluster)
pam.res.2 <- pam(countries.MAH, 2, diss = TRUE)
```

# **medoids**: Objects that represent clusters

```
pam.res.2$medoids
```



Partitioning Around Medoids (PAM)(3/4)

# Extract clustering vectors

```
pam.res.2$cluster
```

# Extract Silhouette average

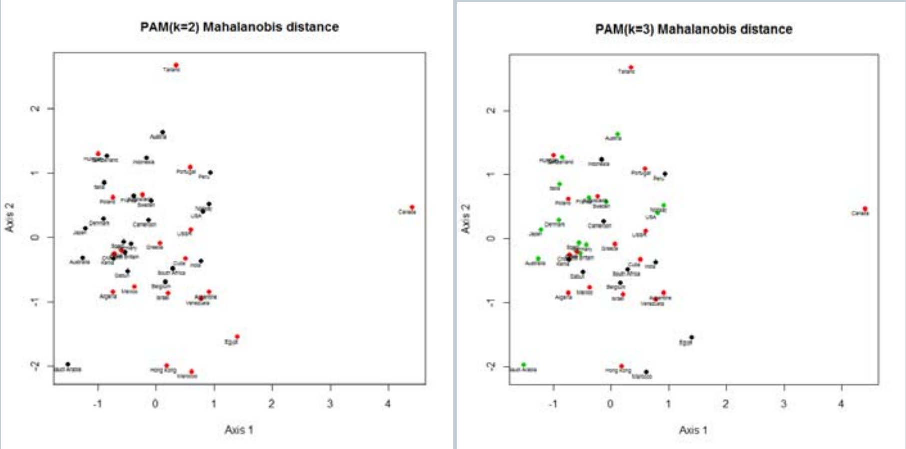
```
pam.res.2$silinfo$avg.width
```

# The result can be plotted using the function `cmdscale()` in `cluster` package.

```
mds.pam.res.2 <- cmdscale(as.dist(countries.MAH), eig=TRUE)

plot(mds.pam.res.2$points[,1], mds.pam.res.2$points[,2],
main="PAM(k=2) Mahalanobis distance", xlab="Axis 1", ylab="Axis 2",
col=pam.res.2$cluster, pch=19)
text(mds.pam.res.2$points[,1], mds.pam.res.2$points[,2], labels=
countries[,1], pos=1, cex=0.5, offset=0.15)
```

Partitioning Around Medoids (PAM)(4/4)



Fuzzy-K means (1/4)

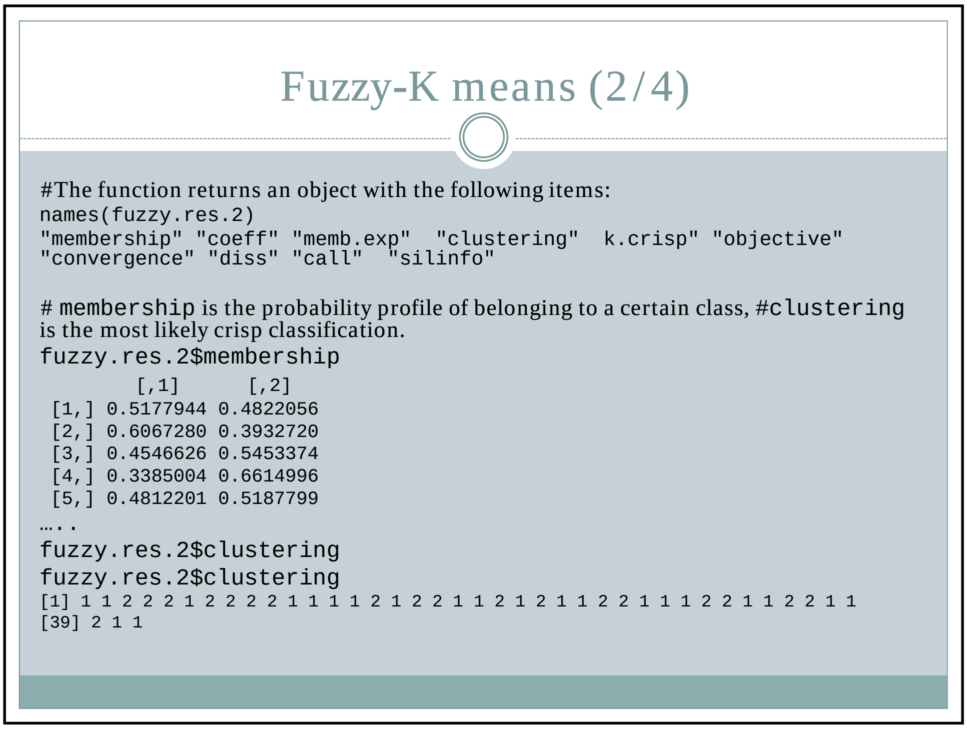
#Fuzzy classification is similar to K-means clustering in finding the optimal classification for a given number of classes, but the produced classification is a probability profile of class membership. The fuzzy clustering is provided by function `fanny` in package `cluster`.

# Fuzzy clustering defaults to Euclidean metric, but current versions can accept any dissimilarities.

#use lower "membership exponent" (`memb.exp`): the default 2 gives complete fuzziness, lower values give crisper classifications.

# Function `fanny()`, simplified format:

```
library(cluster)
fuzzy.res.2 <- fanny(countries.MAH, 2, diss = TRUE,
memb.exp=1.2)
```



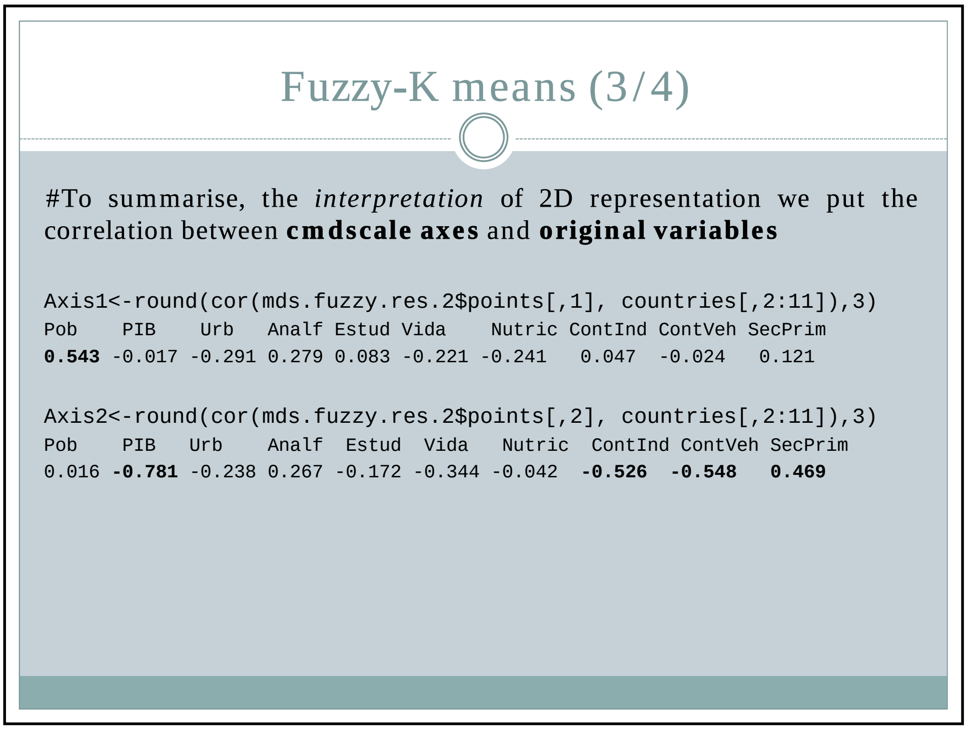
Fuzzy-K means (2/4)

#The function returns an object with the following items:

```
names(fuzzy.res.2)
"membership" "coeff" "memb.exp"  "clustering"  k.crisp" "objective"
"convergence" "diss" "call"  "silinfo"
```

# `membership` is the probability profile of belonging to a certain class, #`clustering` is the most likely crisp classification.

```
fuzzy.res.2$membership
          [,1]      [,2]
 [1,] 0.5177944 0.4822056
 [2,] 0.6067280 0.3932720
 [3,] 0.4546626 0.5453374
 [4,] 0.3385004 0.6614996
 [5,] 0.4812201 0.5187799
…..
fuzzy.res.2$clustering
fuzzy.res.2$clustering
[1] 1 1 2 2 2 1 2 2 2 2 1 1 1 1 2 1 2 2 1 1 2 1 2 1 1 2 2 1 1 1 2 2 1 1 2 2 1 1
[39] 2 1 1
```

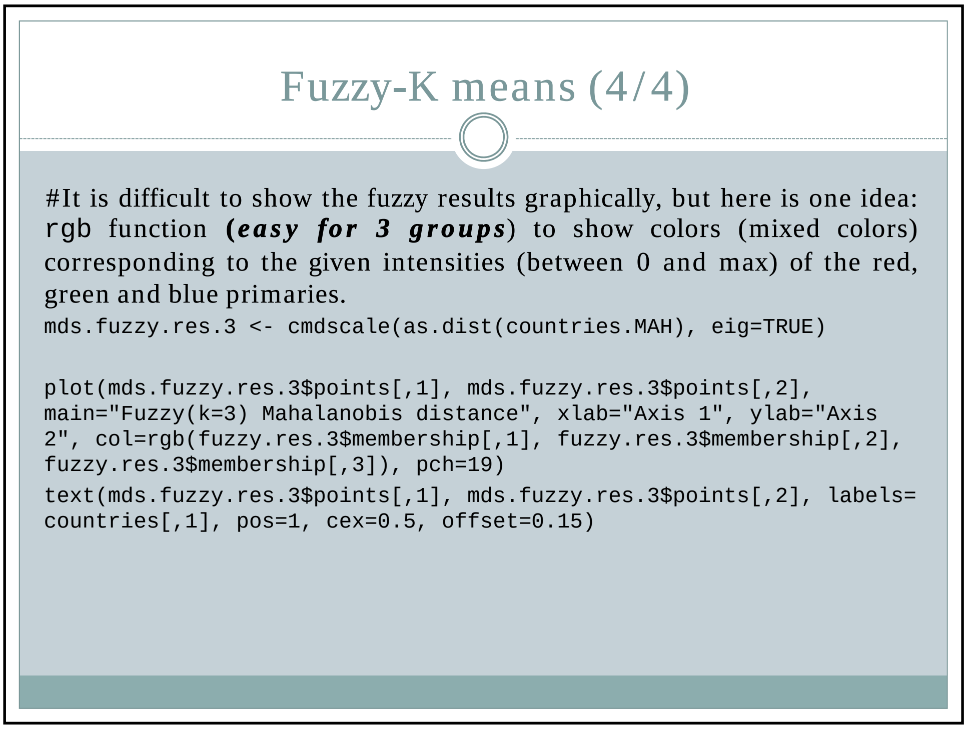


Fuzzy-K means (3/4)

#To summarise, the *interpretation* of 2D representation we put the correlation between **cmdscale axes** and **original variables**

```
Axis1<-round(cor(mds.fuzzy.res.2$points[,1], countries[,2:11]),3)
Pob    PIB    Urb    Analf Estud Vida    Nutric ContInd ContVeh SecPrim
0.543 -0.017 -0.291 0.279 0.083 -0.221 -0.241   0.047  -0.024   0.121

Axis2<-round(cor(mds.fuzzy.res.2$points[,2], countries[,2:11]),3)
Pob    PIB   Urb    Analf  Estud  Vida   Nutric  ContInd ContVeh SecPrim
0.016 -0.781 -0.238 0.267 -0.172 -0.344 -0.042  -0.526  -0.548   0.469
```



Fuzzy-K means (4/4)

#It is difficult to show the fuzzy results graphically, but here is one idea: `rgb` function (***easy for 3 groups***) to show colors (mixed colors) corresponding to the given intensities (between 0 and max) of the red, green and blue primaries.

```
mds.fuzzy.res.3 <- cmdscale(as.dist(countries.MAH), eig=TRUE)

plot(mds.fuzzy.res.3$points[,1], mds.fuzzy.res.3$points[,2],
main="Fuzzy(k=3) Mahalanobis distance", xlab="Axis 1", ylab="Axis
2", col=rgb(fuzzy.res.3$membership[,1], fuzzy.res.3$membership[,2],
fuzzy.res.3$membership[,3]), pch=19)
text(mds.fuzzy.res.3$points[,1], mds.fuzzy.res.3$points[,2], labels=
countries[,1], pos=1, cex=0.5, offset=0.15)
```