Fish trait analyses

Jehaaaaaanne, Thibaaaaaault, Laureeeeeeent

Tue May 19 00:27:58 2020

Framework

This document presents the analyses done on the fish trait matrix.

Data

Abundances

DATRAS CGFS

Fish "traits"

Sources

Sources: - our own work on fishbase and IUCN - Beuhkof et al Pangea database and - benthos : xxx

Raw data

Missing values imputation

```
##                     nbid nbNA
## Species               84    0
## habitat                6    0
## feeding.mode           4    0
## tl                    71    0
## age.maturity          37    2
## growth.coefficient    77    0
## length.max            65    0
## age.max               37    0
## IUCN.status            6    0
```

```
##
##  iter imp variable
##   1   1  age.maturity
##   2   1  age.maturity
##   3   1  age.maturity
##   4   1  age.maturity
##   5   1  age.maturity
```

```
##                     nbid nbNA
## Species               84    0
## habitat                6    0
```

```
## feeding.mode          4    0
## tl                   71    0
## age.maturity         36    0
## growth.coefficient   77    0
## length.max           65    0
## age.max              37    0
## IUCN.status           6    0
```

Quantitative to qualitative data

```
## [1] 0 1 2 3 4 5
```

```
## [1]  0  2  3  4  5 14
```

```
## [1] 0.0 0.1 0.2 0.3 0.4 0.5 2.0
```

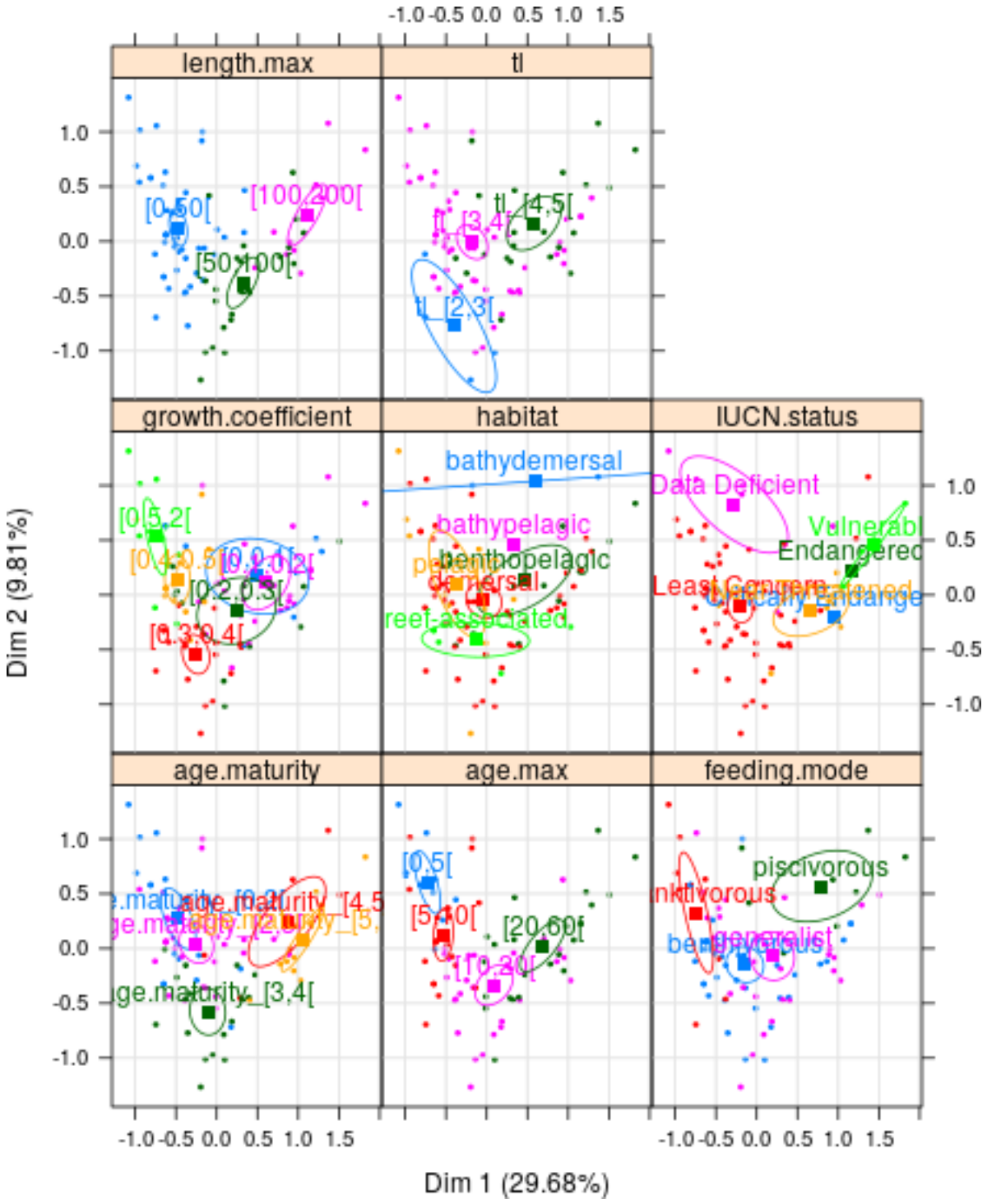
```
## [1]   0  50 100 200
```

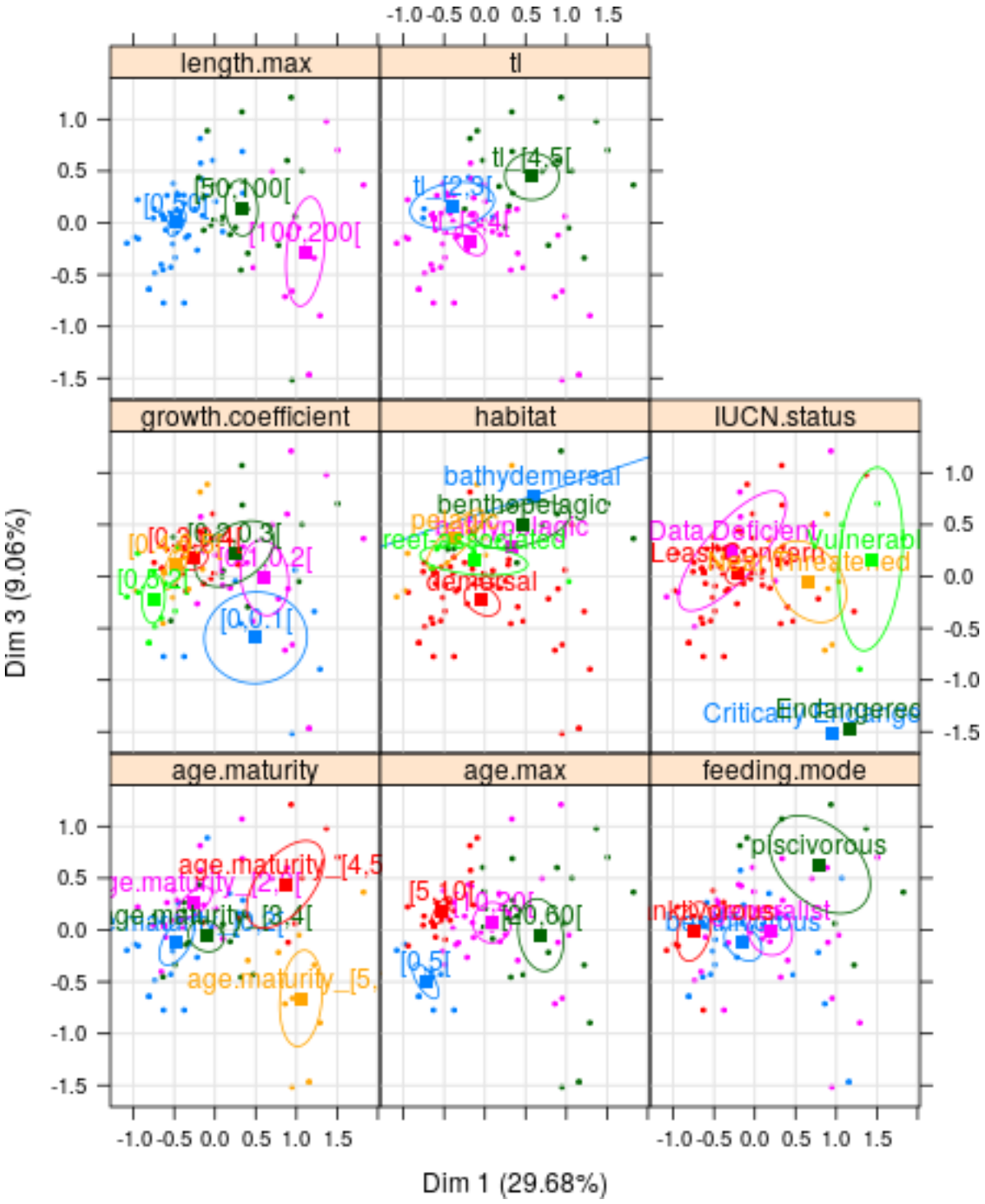
```
## [1]  0  5 10 20 60
```

```
##                    nbid nbNA
## Species              84    0
## habitat               6    0
## feeding.mode          4    0
## tl                    3    0
## age.maturity          5    0
## growth.coefficient    6    0
## length.max            3    0
## age.max               4    0
## IUCN.status           6    0
```

Multiple correspondance analysis

Multiplle correspondance analyses on the trait matrix using Burt tables.


length.max
tl
growth.coefficient
habitat
IUCN.status
age.maturity
age.max
feeding.mode
[0,50[
[100,200[
[50,100[
tl_[3,4[
tl_[4,5[
tl_[2,3[
[0,5,2[
[0,2,0,3[
[0,3,0,4[
bathydemersal
bathypelagic
benthopelagic
demersal
reef-associated
Data Deficient
Vulnerabl
Endangered
Least Concern
Critically Endange
[0,5[
[5,10[
[20,60[
[10,20[
piscivorous
inktivorous
generalist
Dim 1 (29.68%)
Dim 2 (9.81%)



With three axes 48.55% of total variance mapped. Quick analyses: axe one for big fishes with intermediate growth parameters (size, trophic level, age, age maturity) versus small one with low life expectancy, low trophic level and co. Axe two and three not so clear, but help to distinguish the axe one properties.

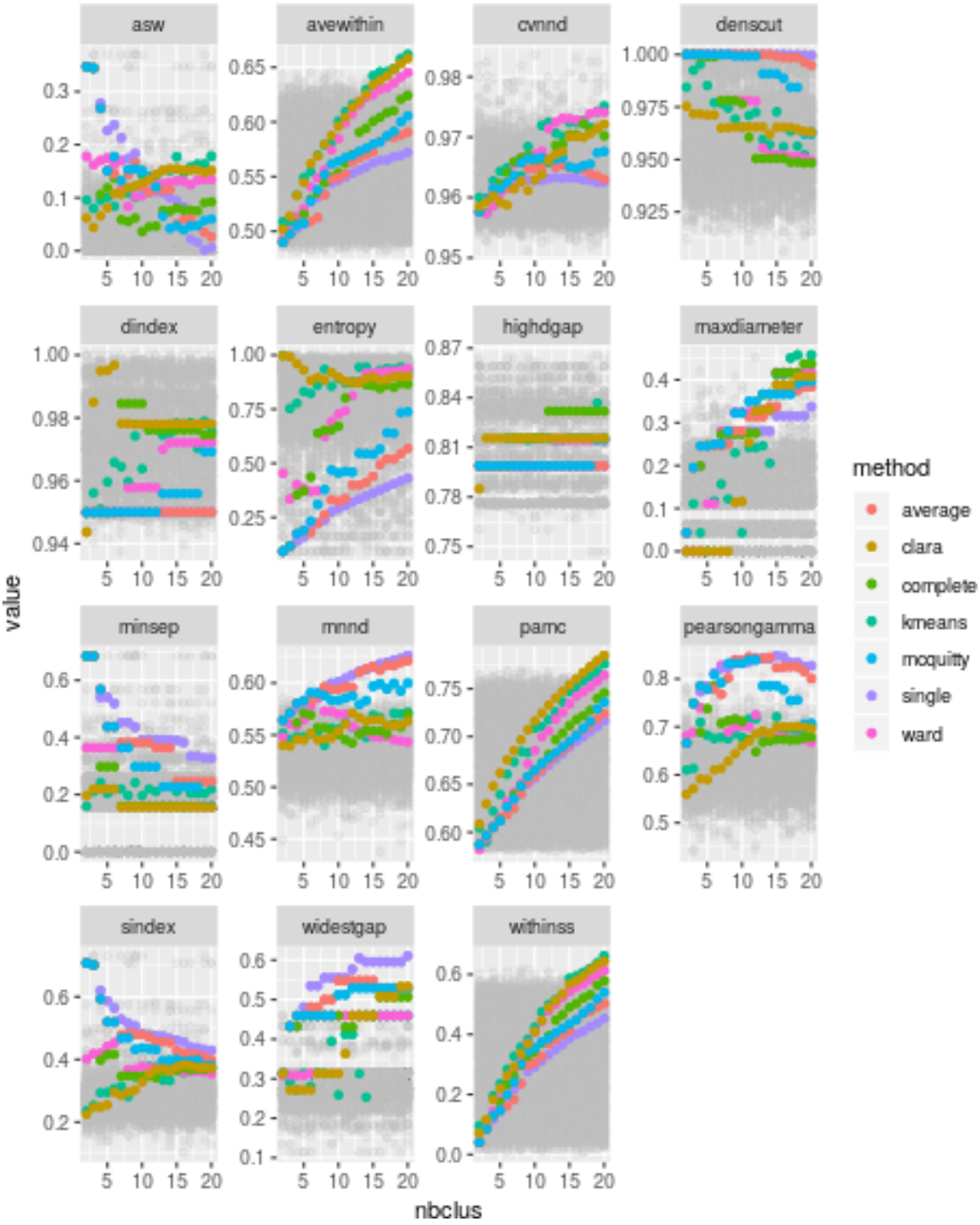
Groups identification in the multiple correspondance analysis spaces

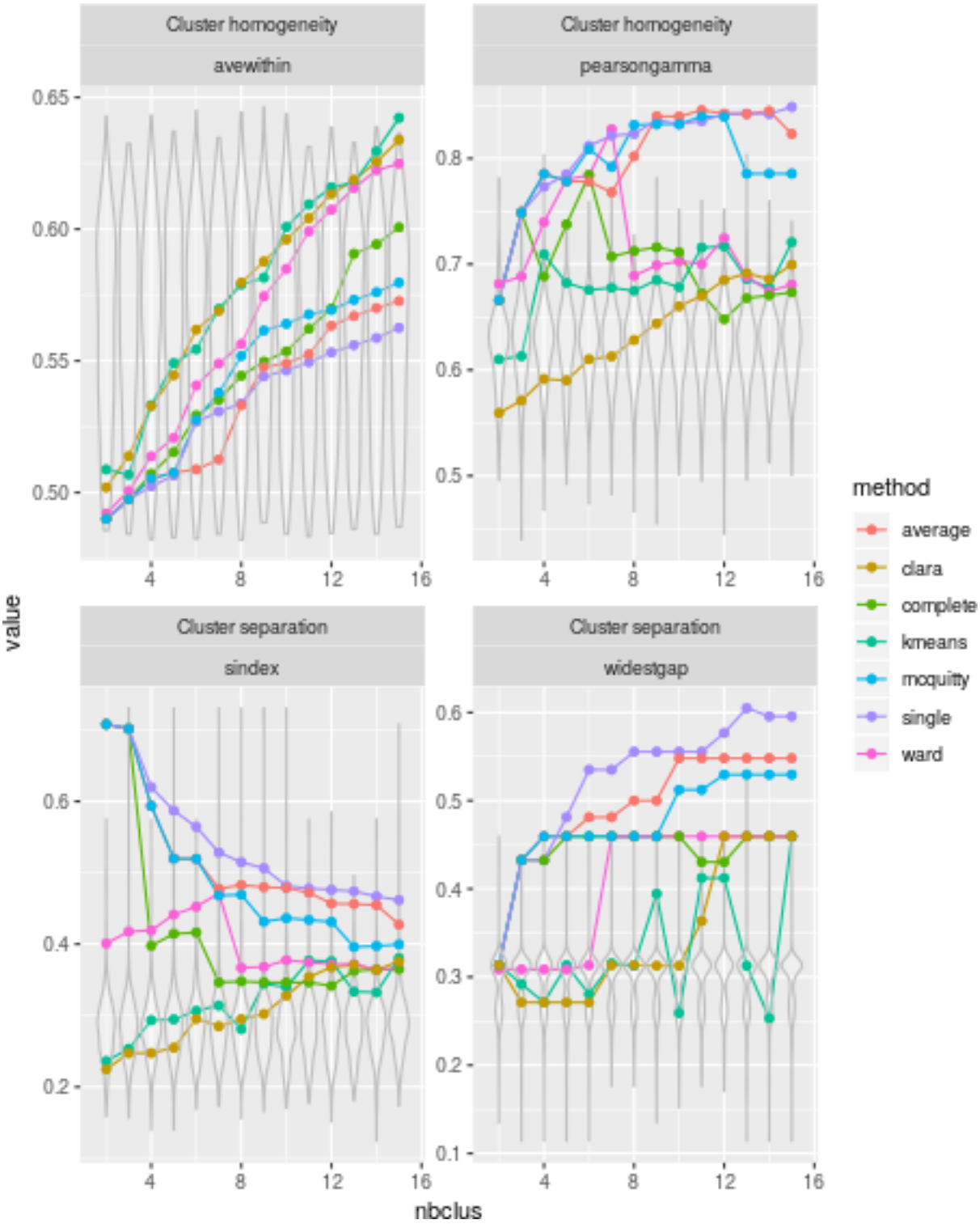
The MCA results are used to identify group of similar traits caracteristics in the MCA space. First methods test using Hennig approach and the fpc packages, then group with the best method. Methods tested with 2 to 20 groups:

- 4 random classification (cf "stupid" classifiers in Hennig),
- hierarchical clustering with ward, single, complete, average, mcquitty criterions,
- kmeans,
- clara methods (large dataset computation methods).

This comes from the quick reading of the Hennig paper found on ArXiv (fpc paper, the metric agregation and the cluster strategy and selection to bee data - yes bee). To select the methods and the number of cluster, differents metrics are plotted and interpreted. Two type of metrics: metrics to assess cluster homogeneity and metrics to assess cluster separation. The choice of the metrics has to be in line with the cluster objectives (here to reduce the dimension of the species number in DATRAS, according to some traits similarity). Choice of the measures link to the aim of the clustering. Some random notes taken during the reading of the Hennig's papers:

- cluster homogeneity:
  - average within-cluster dissimilarities or average distance with cluster (avewithin): individuals within cluster should be functionally more similar than individuals in different cluster (smaller values indicates better clustering quality),
  - correlation between distances and a 0-1 vector where 0 means same cluster, 1 means different clusters (pearsongamma): group of species should represent the functional distance in the MCA space (youhouu, headache incoming), it is a representation of dissimilarity structure by clustering (a guess: higher is better?). Measure of the quality of the cluster representatiion "in some way".
- cluster seperation
  - separation index, based on the distance for every point to closest point not in the same cluster (sindex): individuals should be functionaly separated from other species (smaller values are better)
  - widest within-cluster gap (widestgap): assess connectivity (no gap in cluster : smaller value are better)

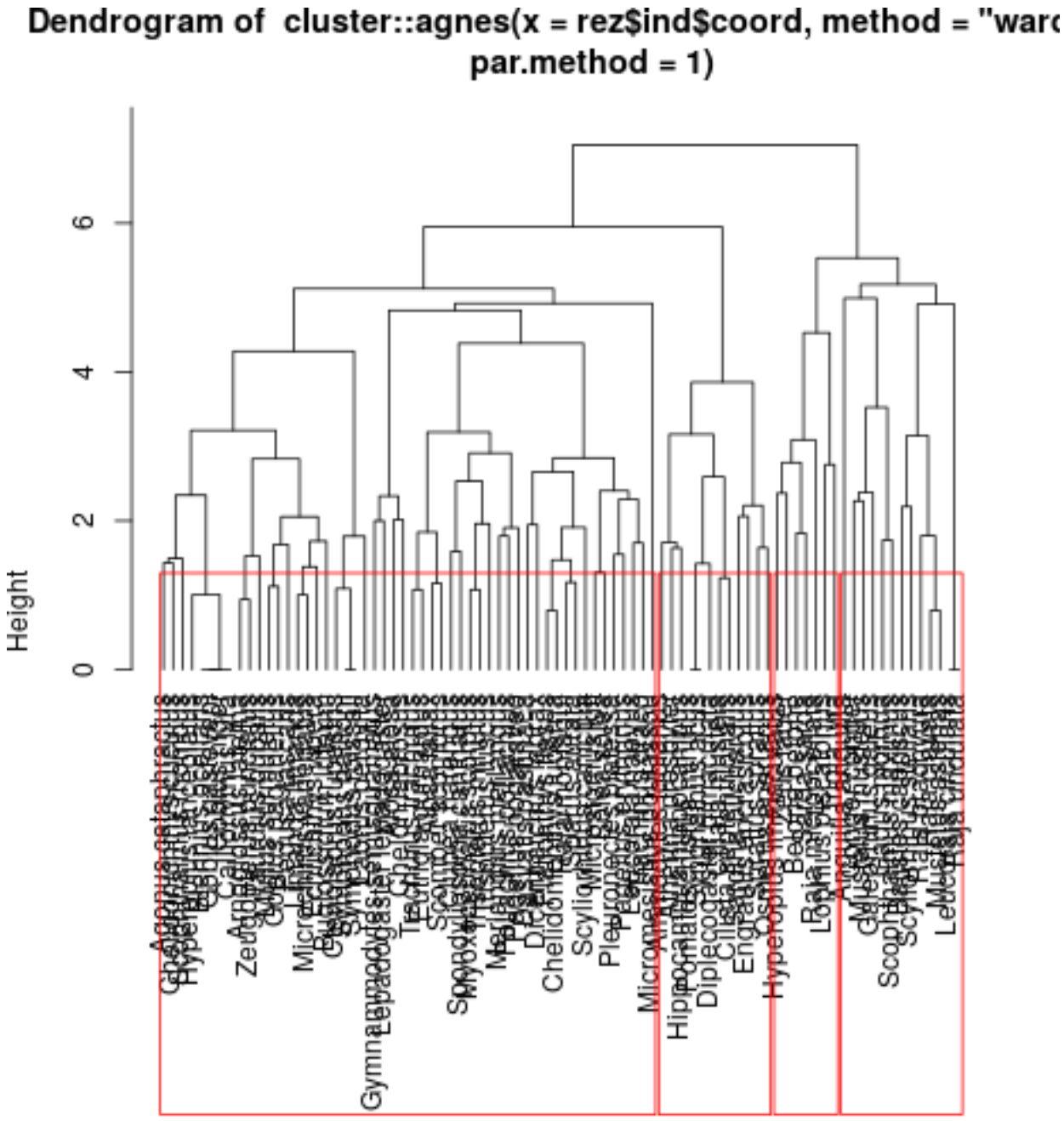

asw
avewithin
cvnnd
denscut
dindex
entropy
highdgap
maxdiameter
minsep
mnnd
pamc
pearsongamma
sindex
widestgap
withinss
value
nbclus
method
average
clara
complete
kmeans
mcquitty
single
ward

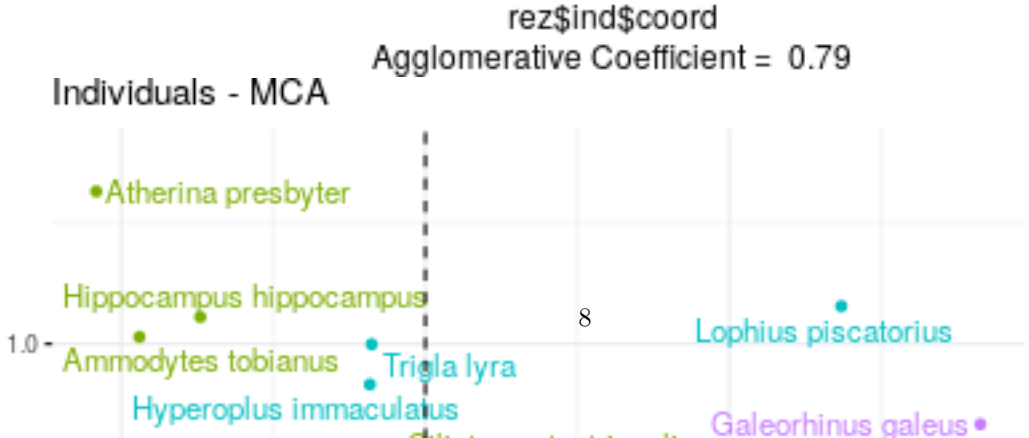


Results :

- 7 clusters! 4 to 5 should be better

Final temporary results and Representation for interpretation

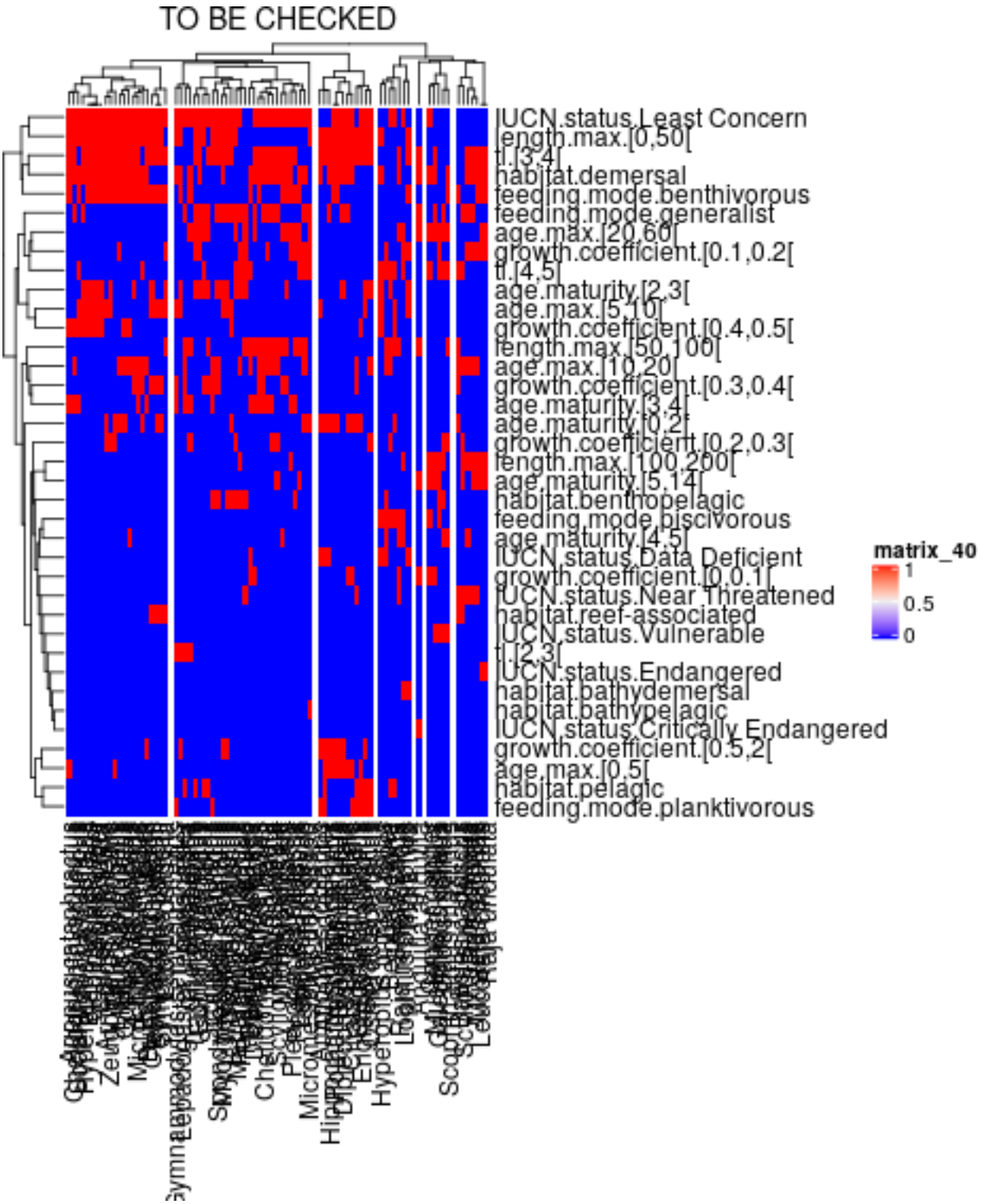
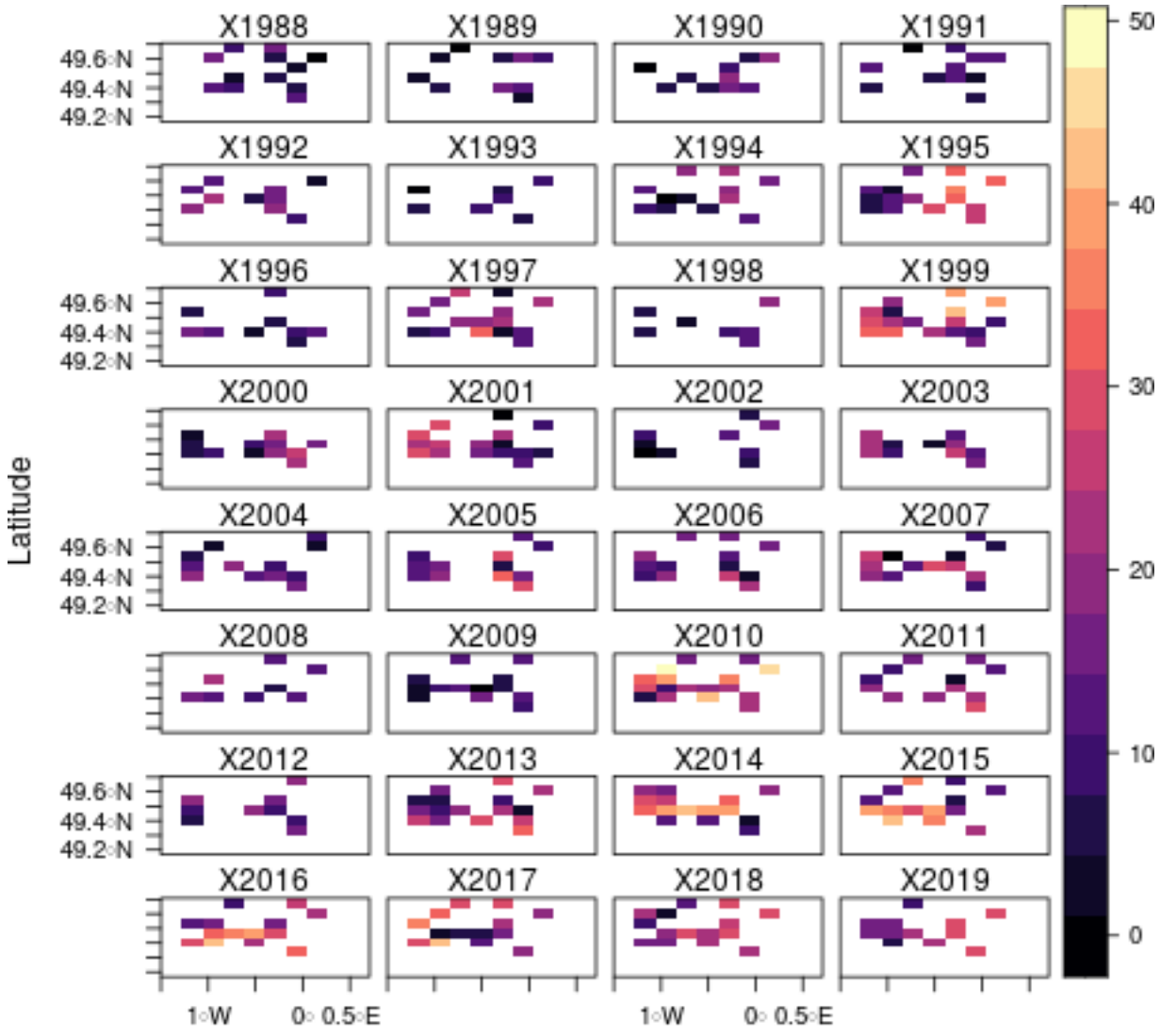
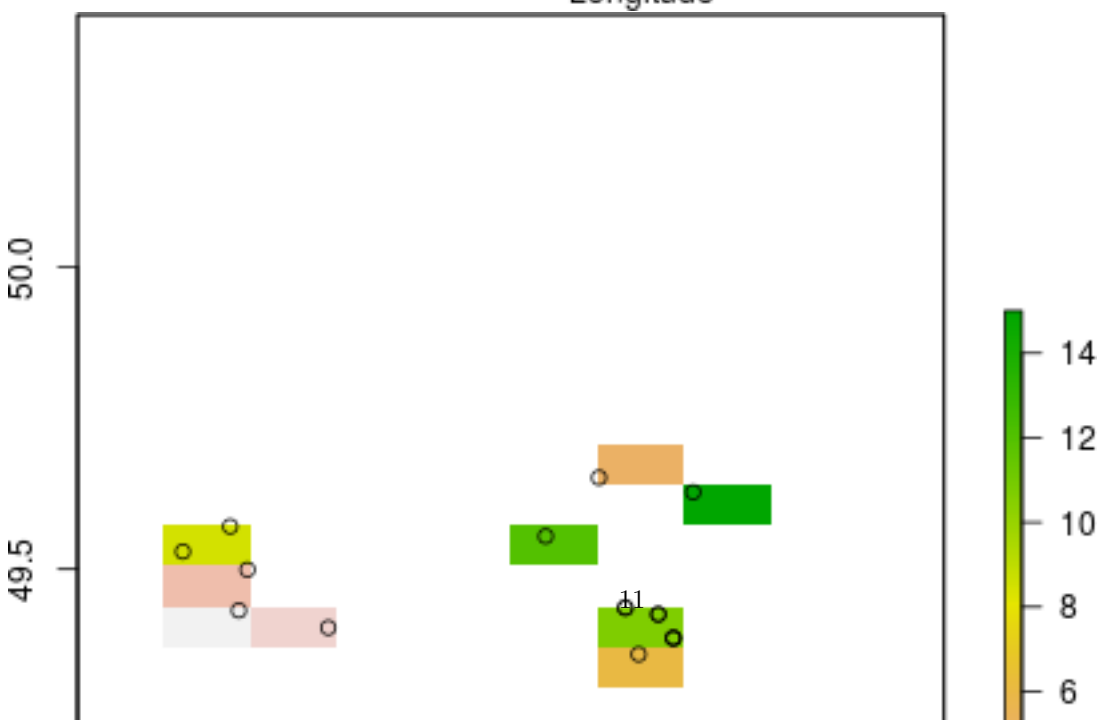




Individuals - MCA
Atherina presbyter
Hippocampus hippocampus
Ammodytes tobianus
Trigla lyra
Hyperoplus immaculatus
Lophius piscatorius
Galeorhinus galeus
Engraulis encrasicolus
Ciliata septentrionalis
Pomatoschistus minutus
Zeus faber
Syngnathus acus
Diplecogaster bimaculata
Micromesistius poutassou
Conger conger
Sprattus sprattus
Ciliata mustela
Sarda sarda
Gadus morhua
Trisopterus minutus
Hyperoplus lanceolatus
Myoxocephalus scorpius
Trisopterus luscus
Mustelus mustelus
Leucoraja circularis
Taurulus bubalis
Callionymus lyra
Trachurus trachurus
Merlangius merlangus
Raja undulata
Gobius niger
Gobius gasteveni
Eutrigla gurnardus
Pollachius pollachius
Raja microocellata
Arnoglossus laterna
Pegusa lascaris
Scophthalmus maximus
Blennius ocellaris
Agonus cataphractus
Belone belone
Raja clavata
Gobius paganellus
Dasyatis pastinaca
Mustelus asterias
Lepadogaster lepadogaster
Pagellus bogaraveo
Leucoraja naevus
Symphodus roissali
Scophthalmus rhombus
Alosa fallax
Anguilla anguilla
Sardina pilchardus
Buglossidium luteum
Echiichthys vipera
Trachinus draco
Symphodus bailloni
Spondyliosoma cantharus
Pleuronectes platessa
Ctenolabrus rupestris
Scomber scombrus
Pagellus erythrinus
Clupea harengus
Microstomus kitt
Dicentrarchus labrax
Limanda limanda
Solea solea
Raja montagui
Scyliorhinus canicula
Gymnammodytes semisquamatus
Balistes capriscus
Microchirus variegatus
Sparus aurata
Chelidonichthys lucerna
Platichthys flesus
Chelon labrosus
Alosa alosa
Dim2 (9.8%)
Dim1 (29.7%)
1.0
0.5
0.0
-0.5
-1.0
-1
0
1
Groups
1
2
3
4
5
6
7

Heatmap

Nice plot but impossible to include in the markdown for now...



Rasteriiiiiiize