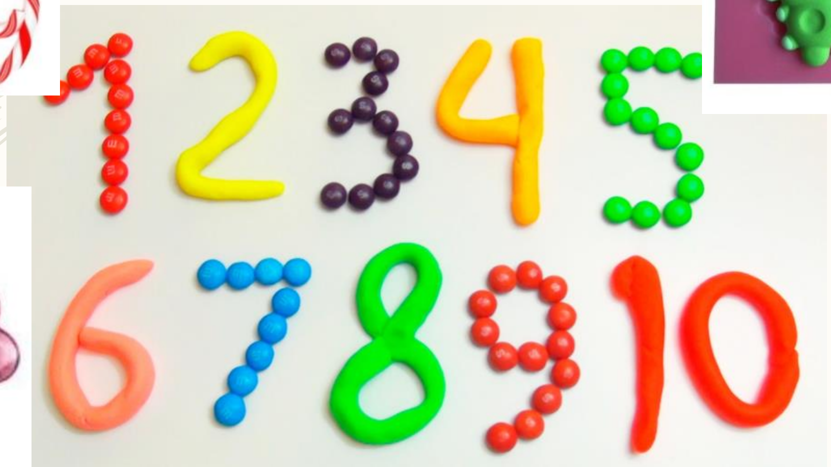
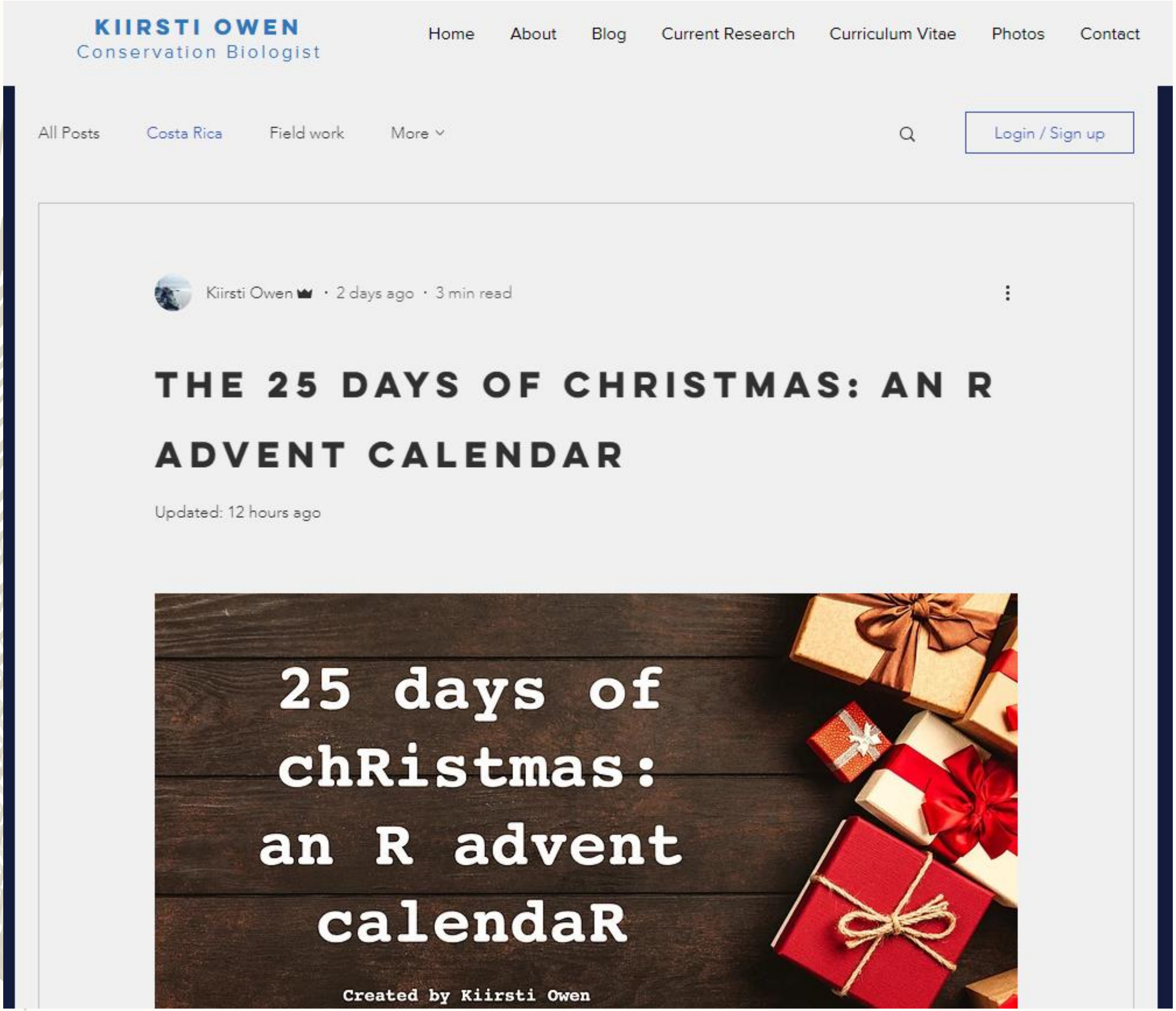
Aula 22 Goodies*



* Goodies related to animals, plants and numbers…



https://kiirstio.wixsite.com/kowen/post/the-25-days-of-christmas-an-r-advent-calendar

[MARMAM] PhD opportunities at the University of St Andrews

Reply | Reply All | Forward | ...

SH MARMAM <marmam-bounces@lists.uvic.ca> on behalf of Sonja Heir
To 'marmam@lists.uvic.ca'

Tue 11/26/2019 3:04 PM

The actual sender of this message is different than the normal sender. Click here to learn more.

ATT00001.txt
423 bytes

We have several exciting and funded* PhD opportunities available at the University of St Andrews, UK – please see the School of Biology website for details (PhD projects at the SOI/SMRU):
https://synergy.st-andrews.ac.uk/research/phd-study/phd-study-projects/phd-study-soi-projects/

Please note different deadlines & funding opportunities – some closing very soon!
01 December 2019 (*funded for students **worldwide**):

- **The seasonal ocean dynamics of the Amundsen Sea Embayment (using telemetry & oceanographic data from seal-born sensors)– Supervisor: Dr Lars Boehme**
- Impacts of fishing-induced changes in forage fish school structure on African penguin foraging – Supervisor: Prof Andy Brierley

13 December 2019 (*funded for UK/EU students):

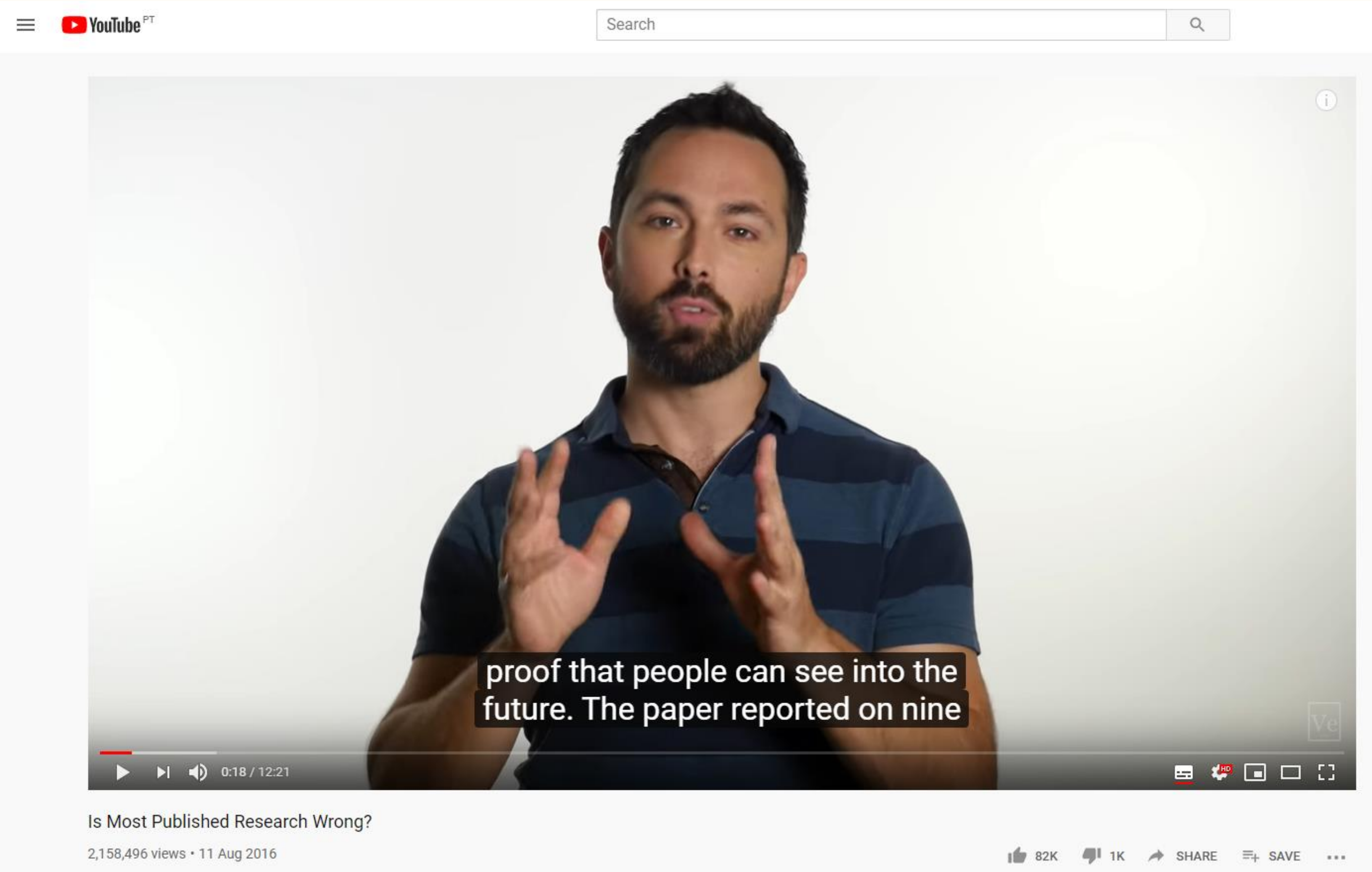
- A lab on a chip: using nano-plasmonics tongues for building miniaturized ecosystem sensors (SUPER DTP)- Supervisor: Dr Lars Boehme
-

06 January 2020 (*funded for UK/EU students)

- Killer whale predation of harbour seals in the coastal waters of Scotland: investigating the ecological drivers and consequences of an apex predator-prey interaction – Supervisors: Dr Gordon Hastie & Dr Saana Isojunno

Posted on behalf of my colleagues.
Best wishes from Scotland,
Sonja

Dr Sonja Heinrich
Sea Mammal Research Unit
School of Biology
University of St Andrews
Scotland, UK



https://www.youtube.com/watch?v=42QuXLucH3Q nature

Subscribe Search Login

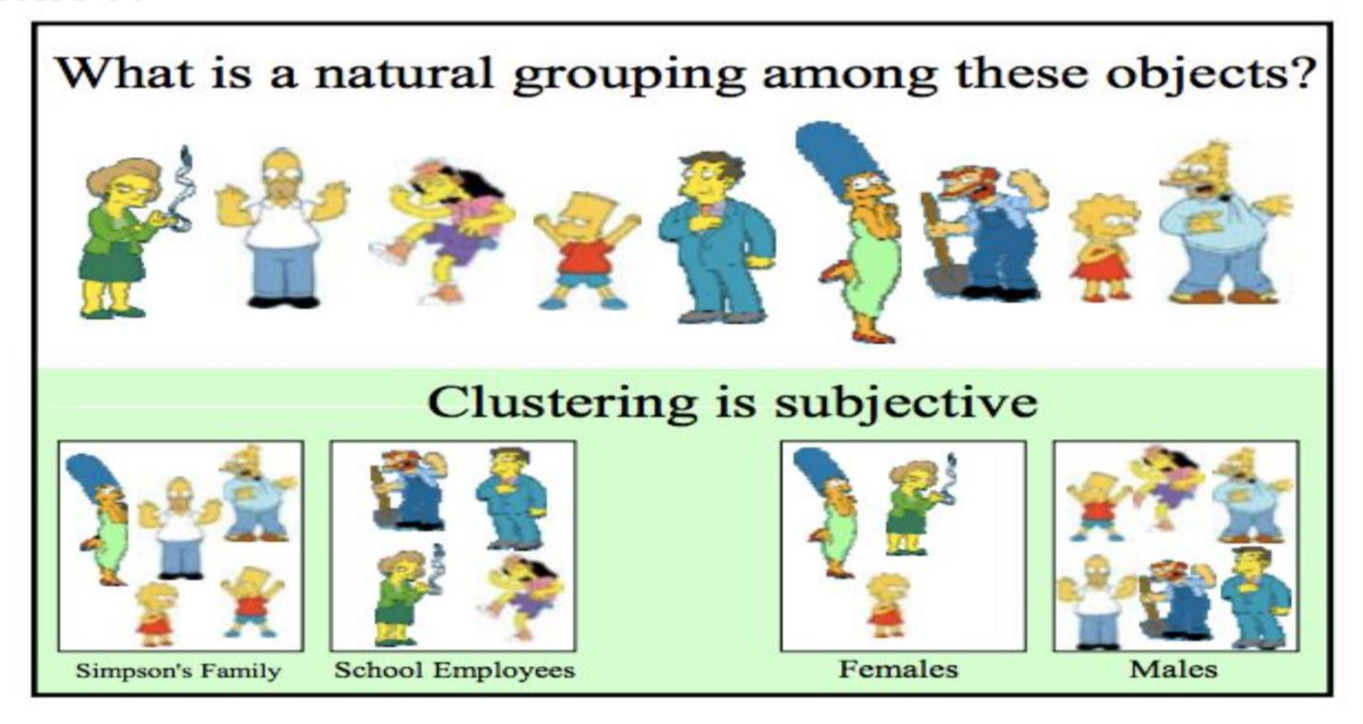
COMMENT · 27 NOVEMBER 2019

Climate tipping points — too risky to bet against

The growing threat of abrupt and irreversible climate changes must compel political and economic action on emissions.

Timothy M. Lenton, Johan Rockström, Owen Gaffney, Stefan Rahmstorf, Katherine Richardson, Will Steffen & Hans Joachim Schellnhuber

https://www.nature.com/articles/d41586-019-03595-0



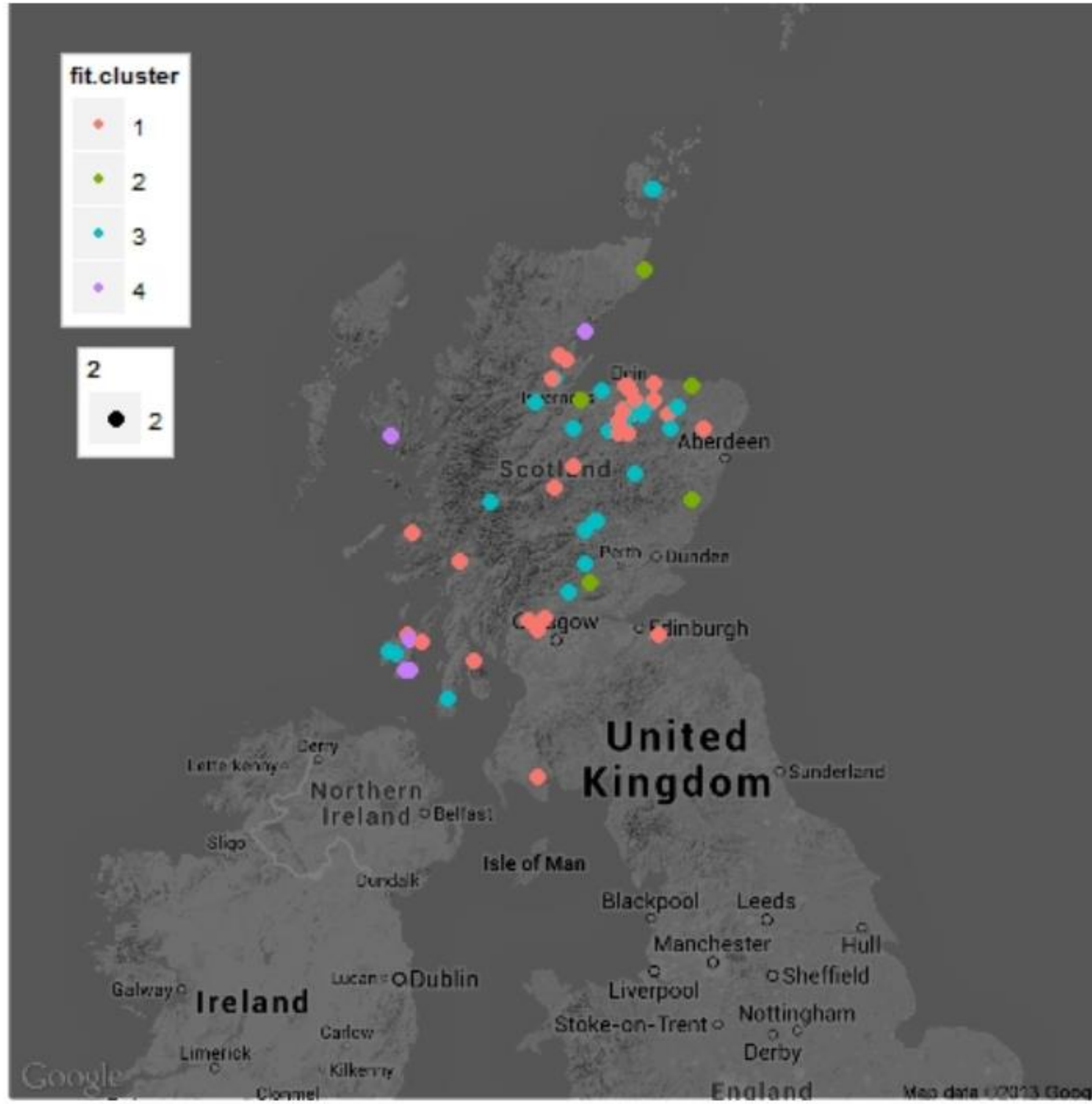
https://hackernoon.com/machine-learning-basics-its-your-cup-of-tea-af4baf060ace

December 30, 2013

K-means Clustering 86 Single Malt Scotch Whiskies

by Luba Gloukhov

The first time I had an Islay single malt, my mind was blown. In my first foray into the world of whiskies, I took the plunge into the smokiest, peatiest beast of them all — Laphroig. That same night, dreams of owning a smoker were replaced by the desire to roam the landscape of smoky single malts.



https://blog.revolutionanalytics.com/2013/12/k-means-clustering-86-single-malt-scotch-whiskies.html

A Classification of Pure Malt Scotch Whiskies

Francois-Joseph Lapointe; Pierre Legendre

Applied Statistics, Vol. 43, No. 1 (1994), 237-257.

Stable URL:
http://links.jstor.org/sici?sici=0035-9254%281994%2943%3A1%3C237%3AACOPMS%3E2.0.CO%3B2-%23

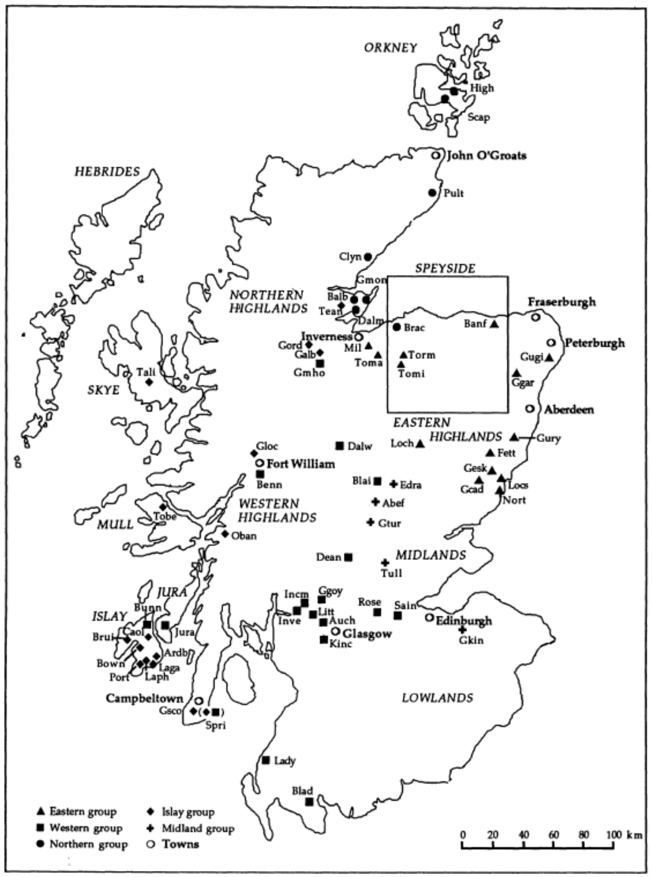


Fig. 2. Map of Scotland showing the positions of the Scotch distilleries, divided into 11 groups (symbols) in the regional classification of single-malt whiskies (Appendix B) (the six Speyside groups are deferred

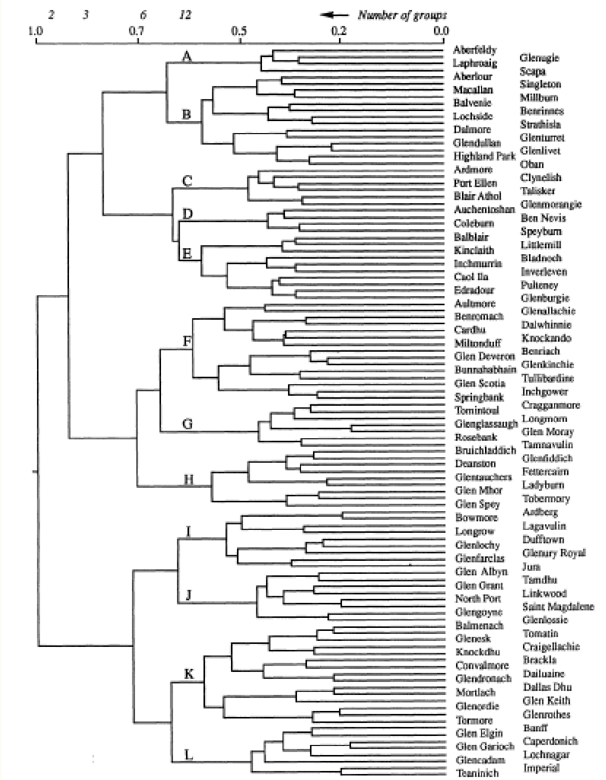


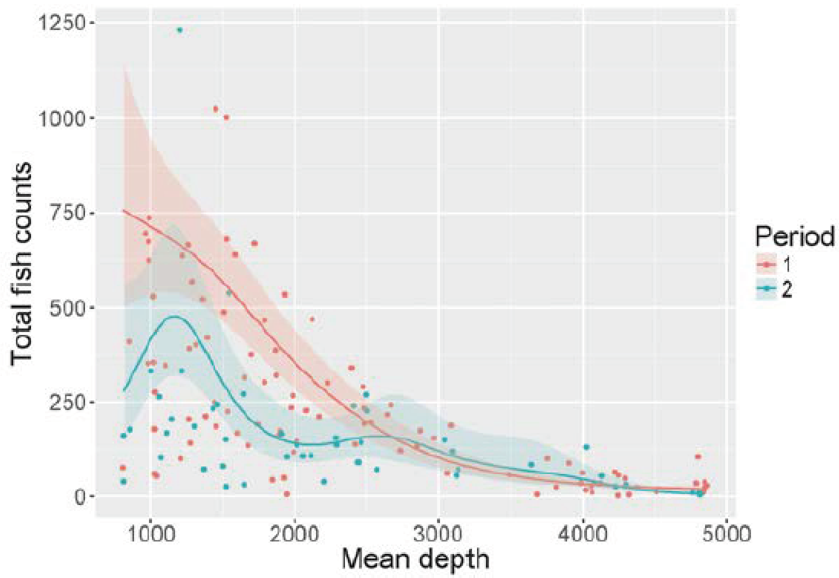
Fig. 1. Dendrogram representing the minimum variance hierarchical classification of single-malt Scotch whiskies: two scales are provided at the top of the graph—the number of groups formed by cutting the dendrogram vertically at the given points and the fusion distances of the hierarchical classification (represented by vertical segments in the dendrogram); the vertical order of the whiskies is partly arbitrary—swapping the branches of a dendrogram does not change the corresponding cophenetic matrix (the 12 groups detailed in Appendix A are labelled A–L here)

Data Exploration, Regression, GLM & GAM with introduction to R

Provided by: Highland Statistics Ltd

In cooperation with:

cE3c – eChanges, Faculty of Sciences, University of Lisbon, Portugal



We begin with an introduction to R and provide a protocol for data exploration to avoid common statistical problems. We will discuss how to detect outliers, deal with collinearity and transformations.

Date & Venue

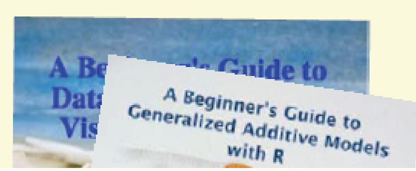
Date: 3 - 7 February 2020

Venue: PT Meeting Centre, R. Bojador 47, Parque das Nações, Lisbon, Portugal

Price: £500

Instructors: Dr. Alain Zuur
Dr. Elena Ieno

Authors of 11 books and providers of over 150 courses

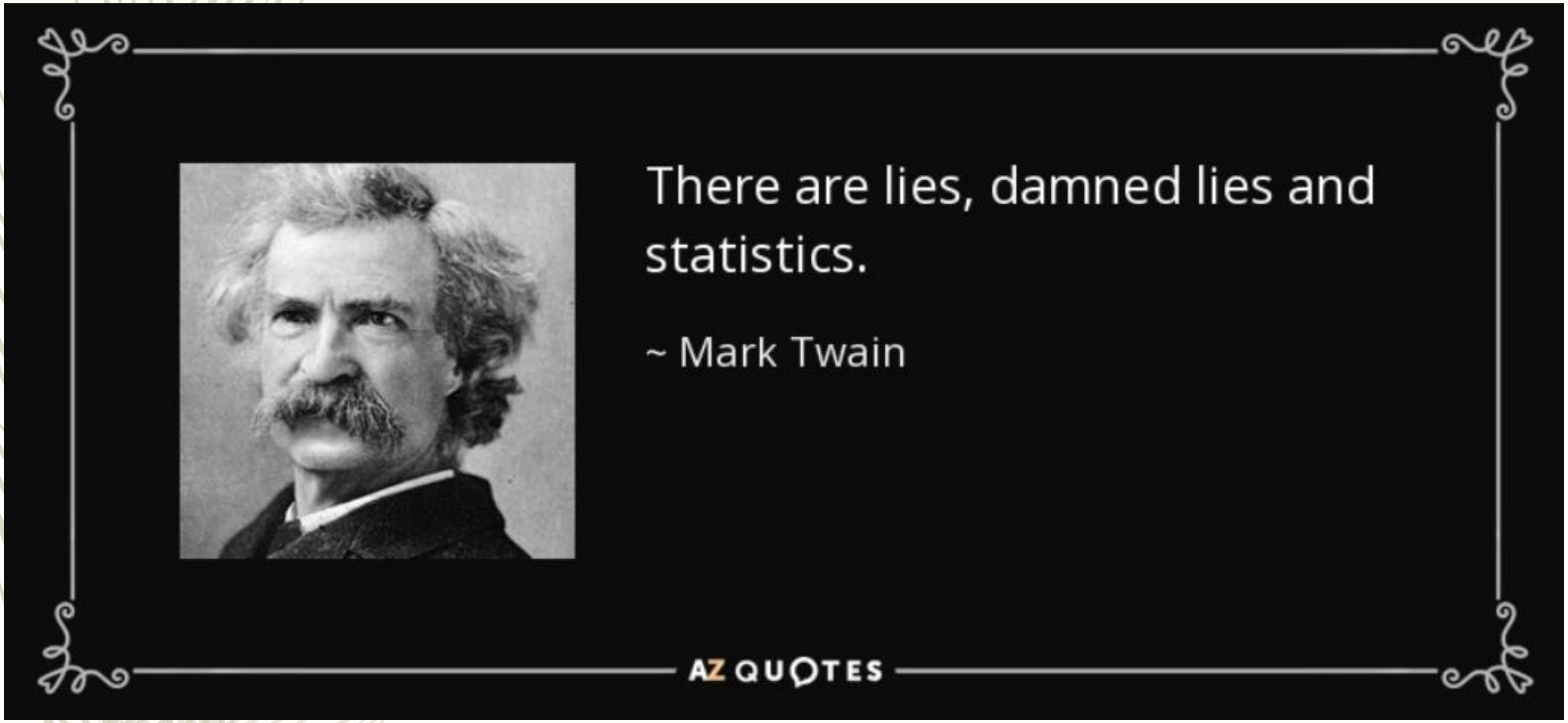


REGISTRATION

www.highstat.com

Dr Alain F Zuur
highstat@highstat.com
www.highstat.com

Ecologia Numérica - Aula Teórica 22 – 02-12-2019



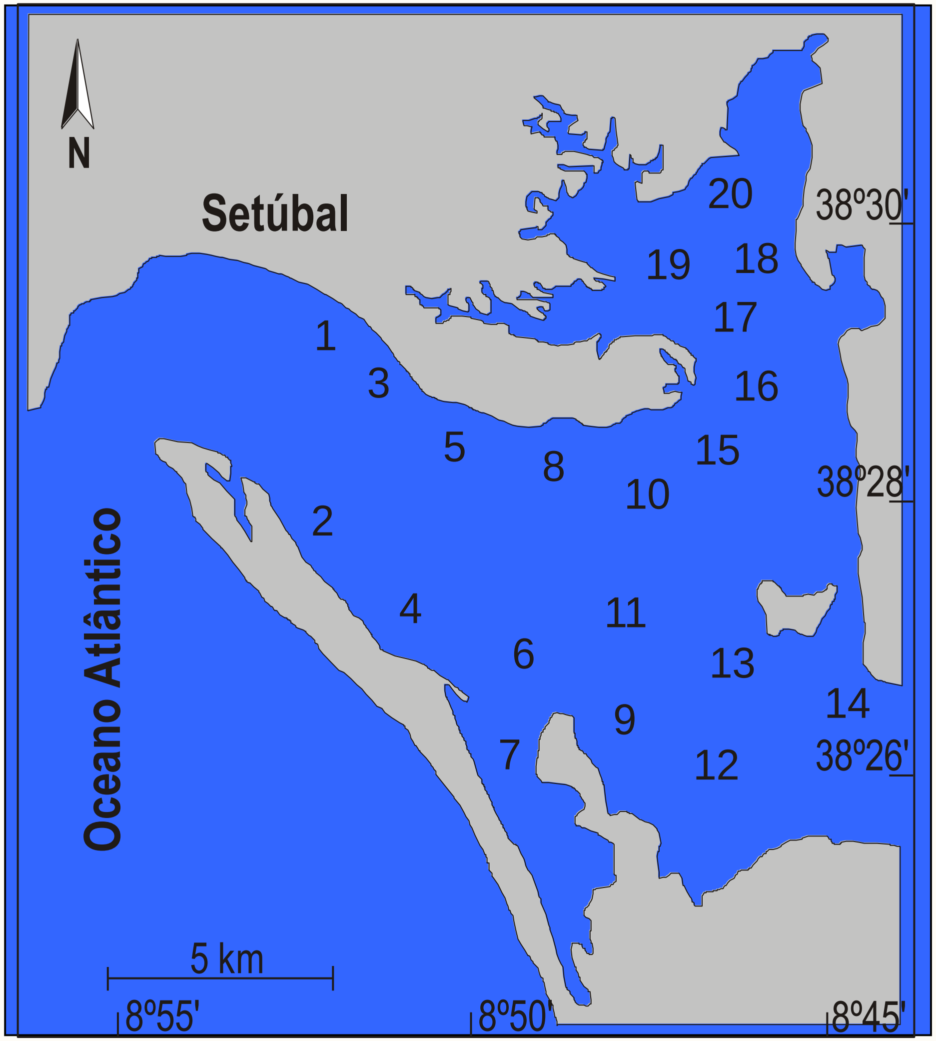
https://www.azquotes.com/quote/298634

Actually in itself a lie, because Mark Twain himself credited it to Disraeli (UK prime minister), but there’s no written record of the statement.

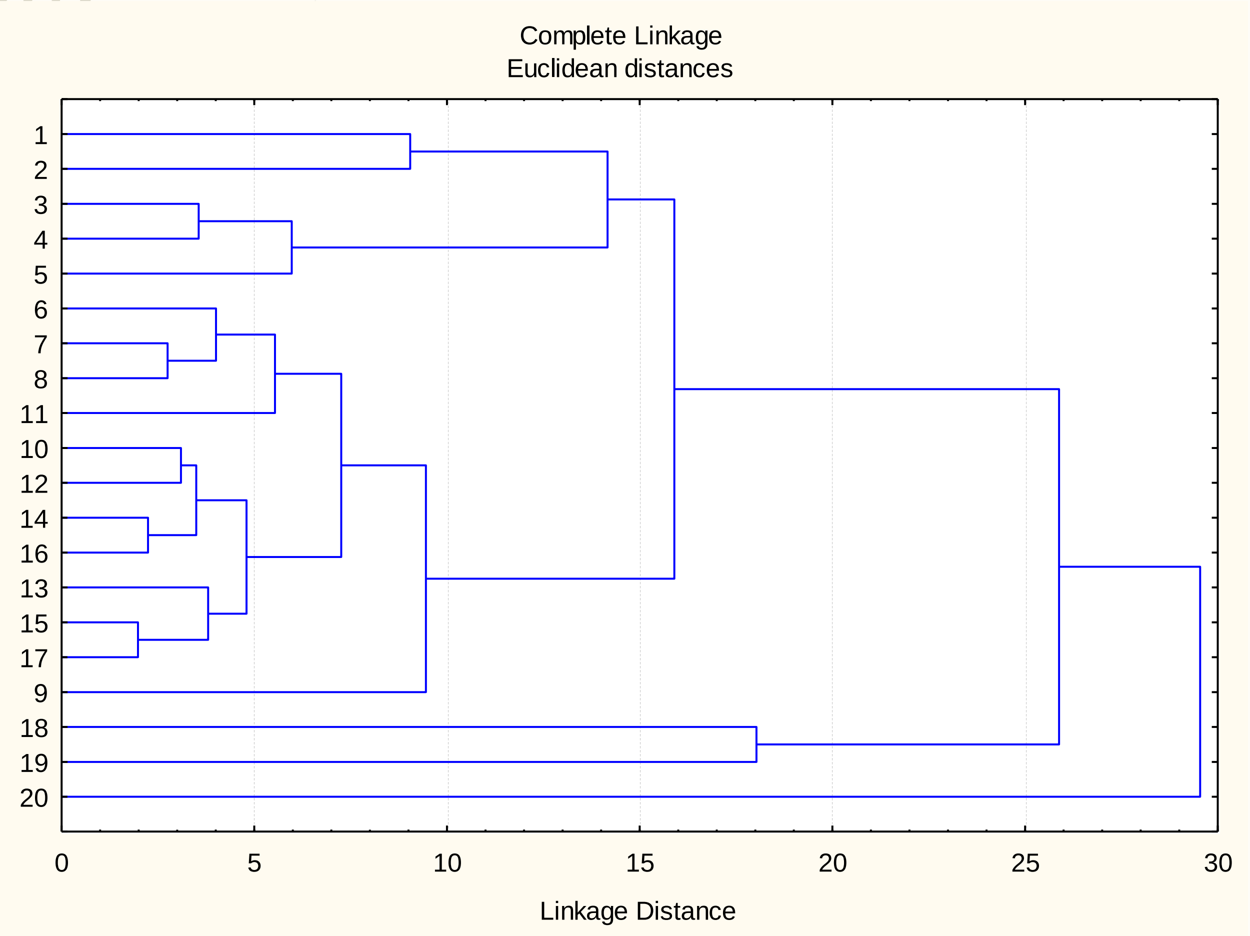
agrupamento

Exemplo:

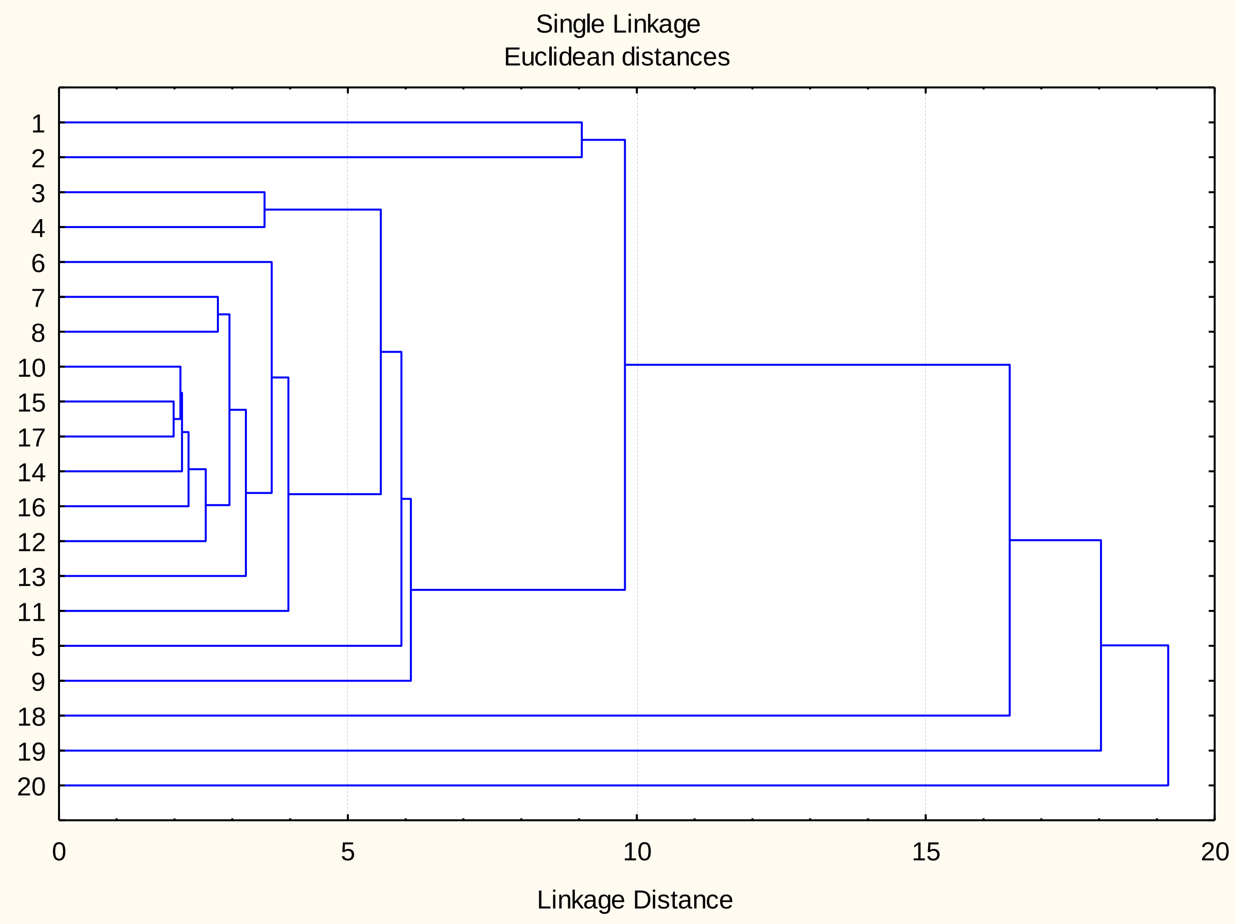
Dados de abundâncias de espécies de peixes em 20 estações de amostragem no estuário do Sado.



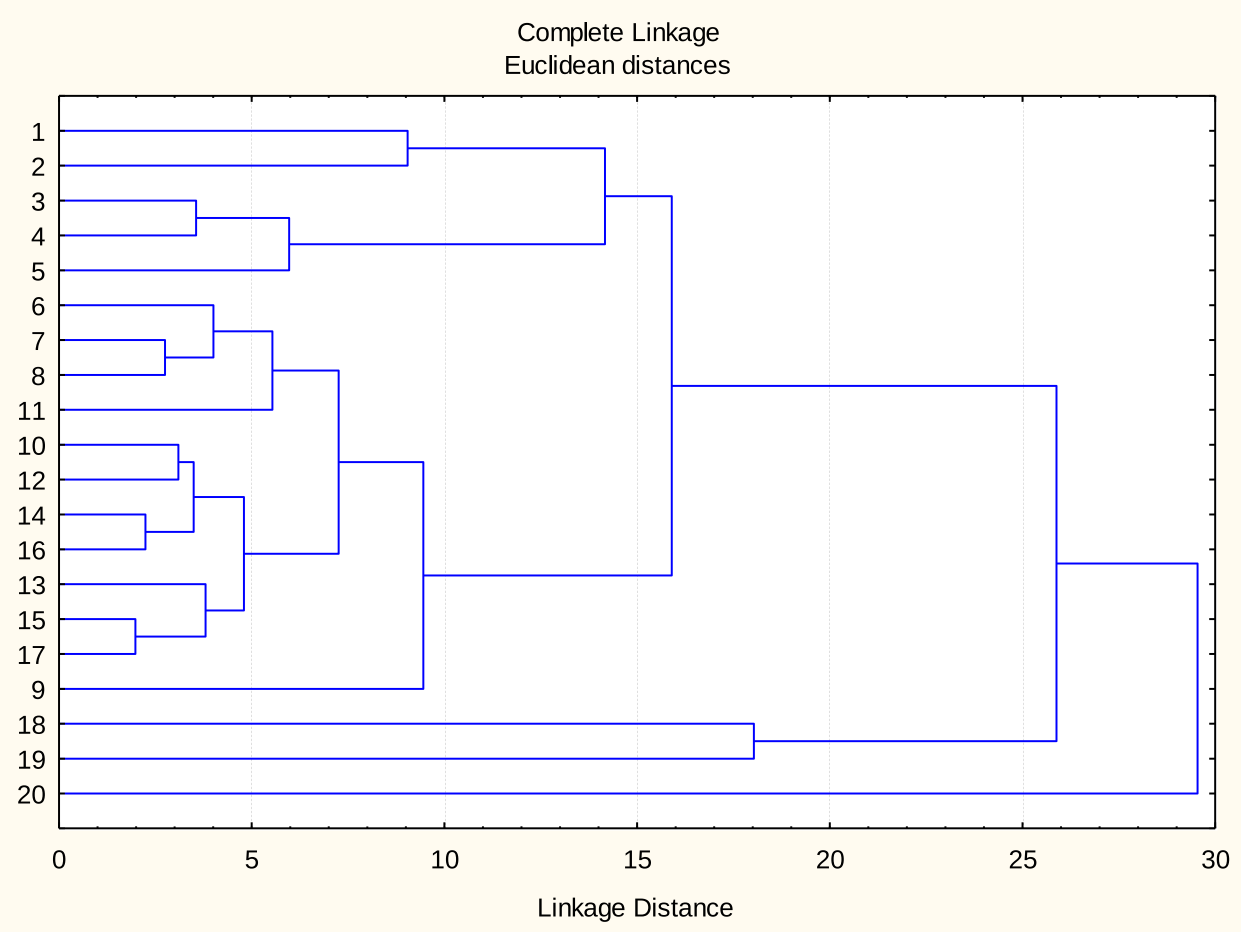
Resultados de uma análise classificativa



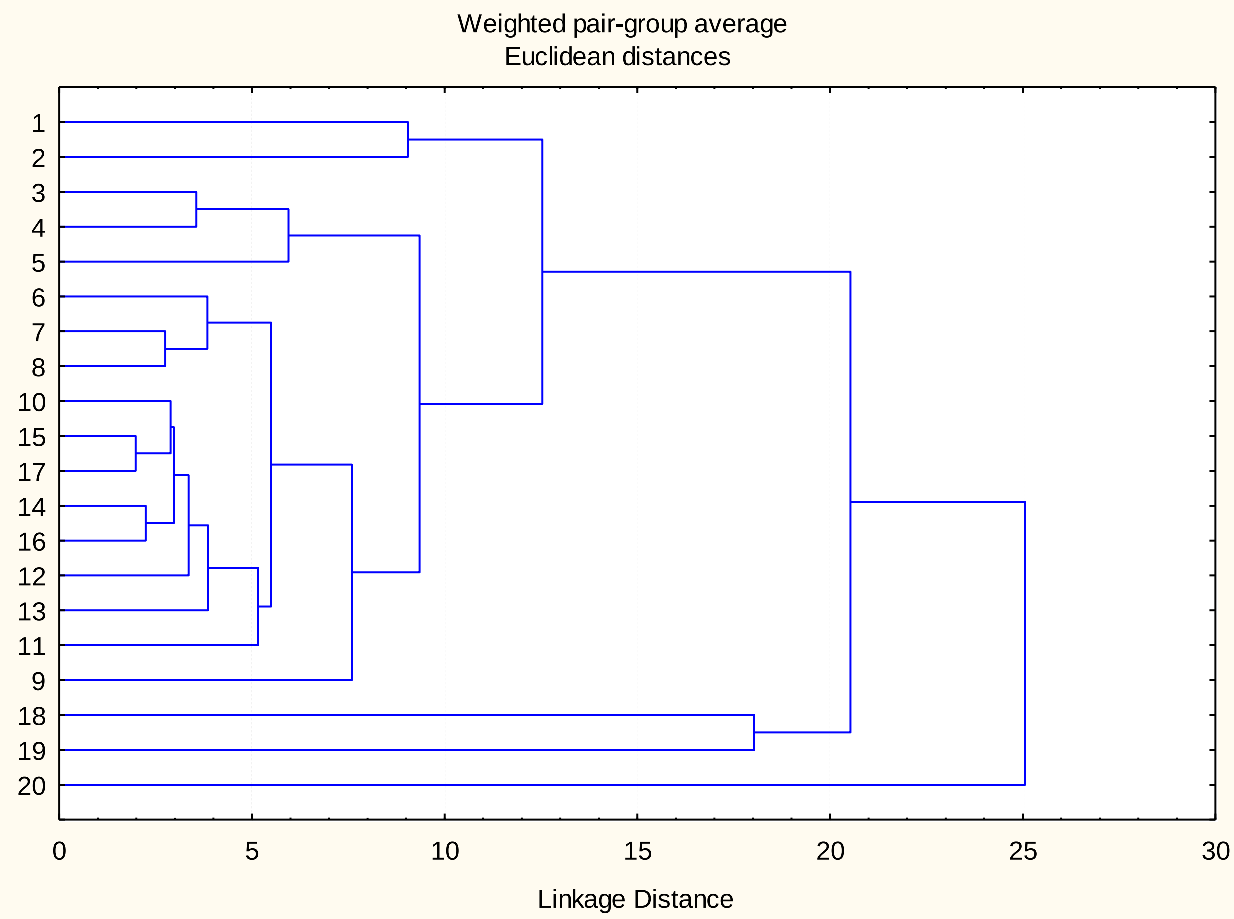
agrupamento



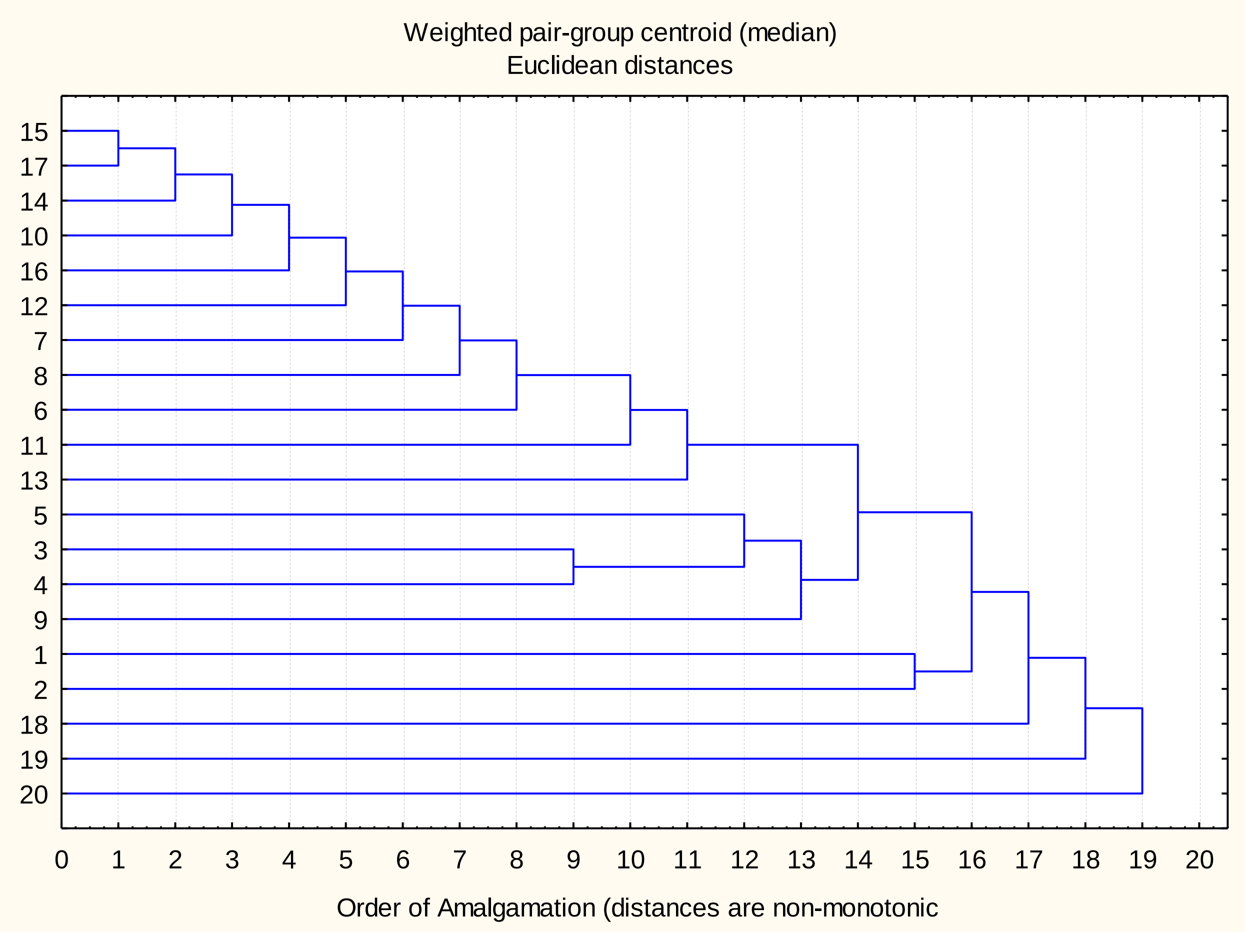
agrupamento



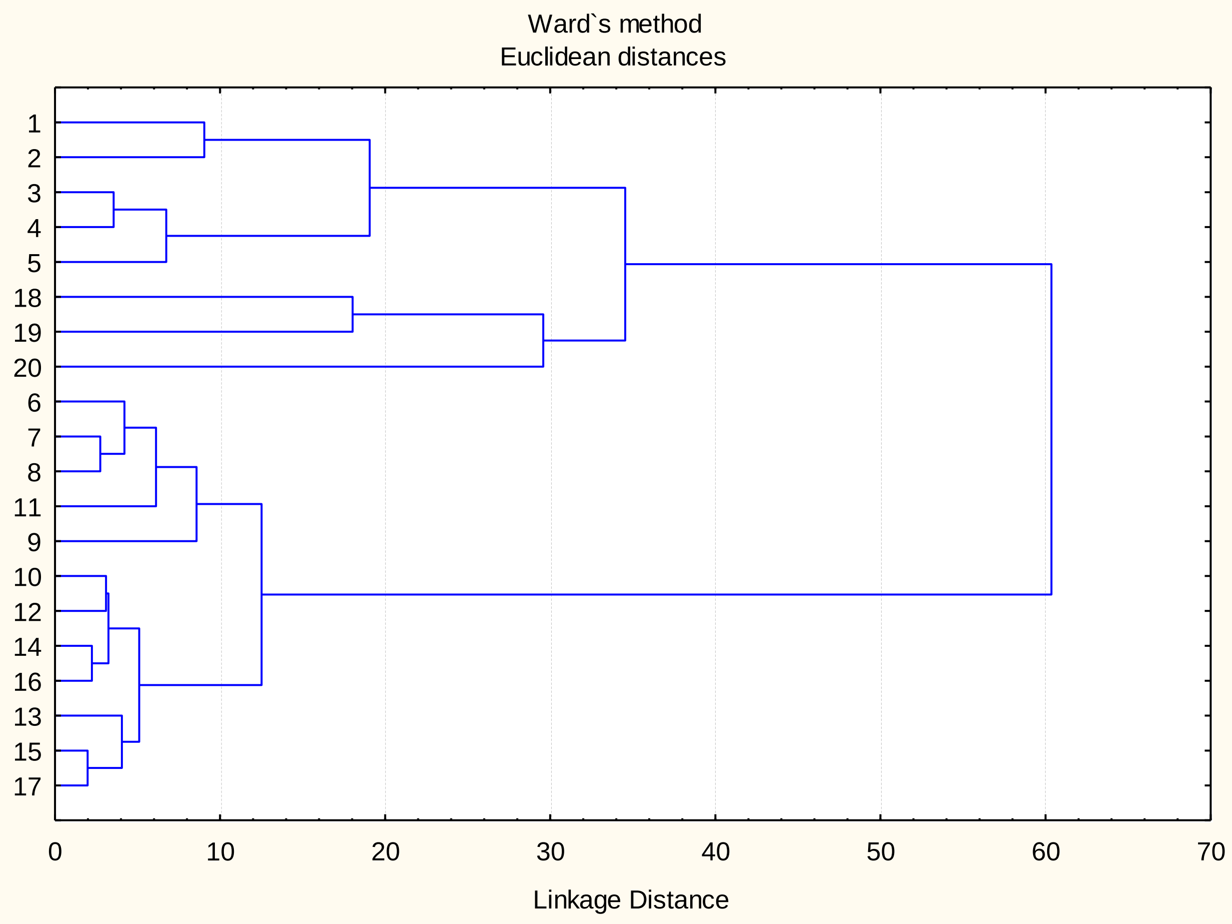
agrupamento



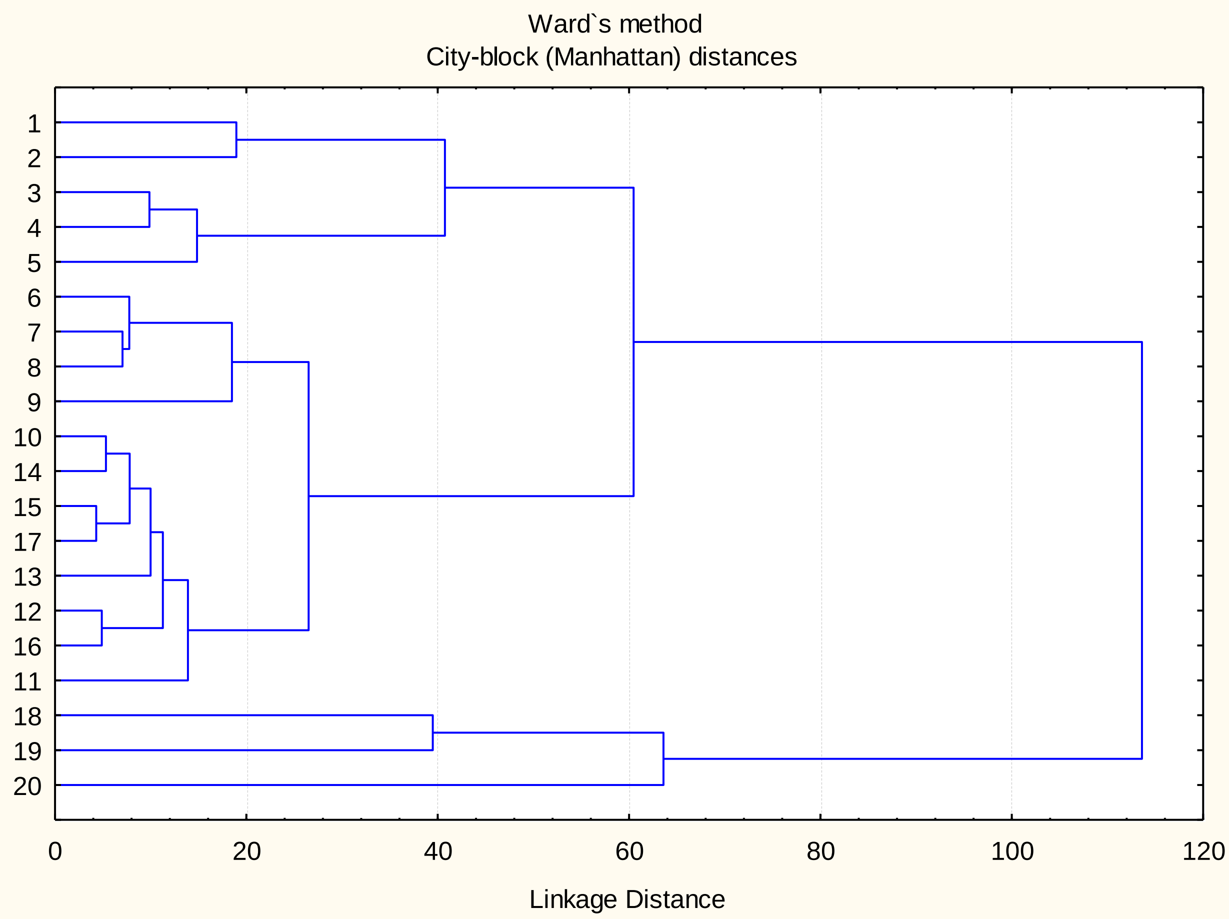
agrupamento



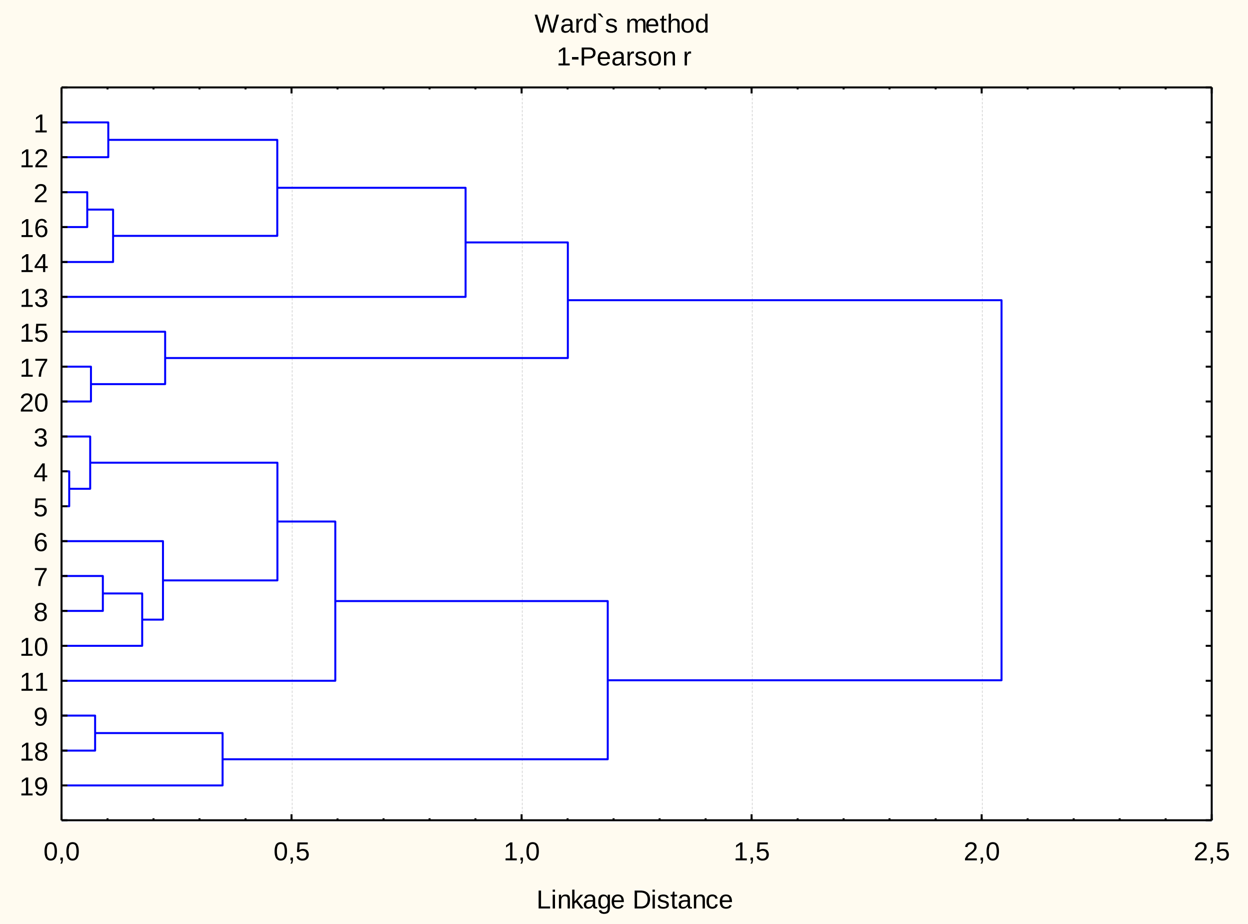
agrupamento



agrupamento



agrupamento

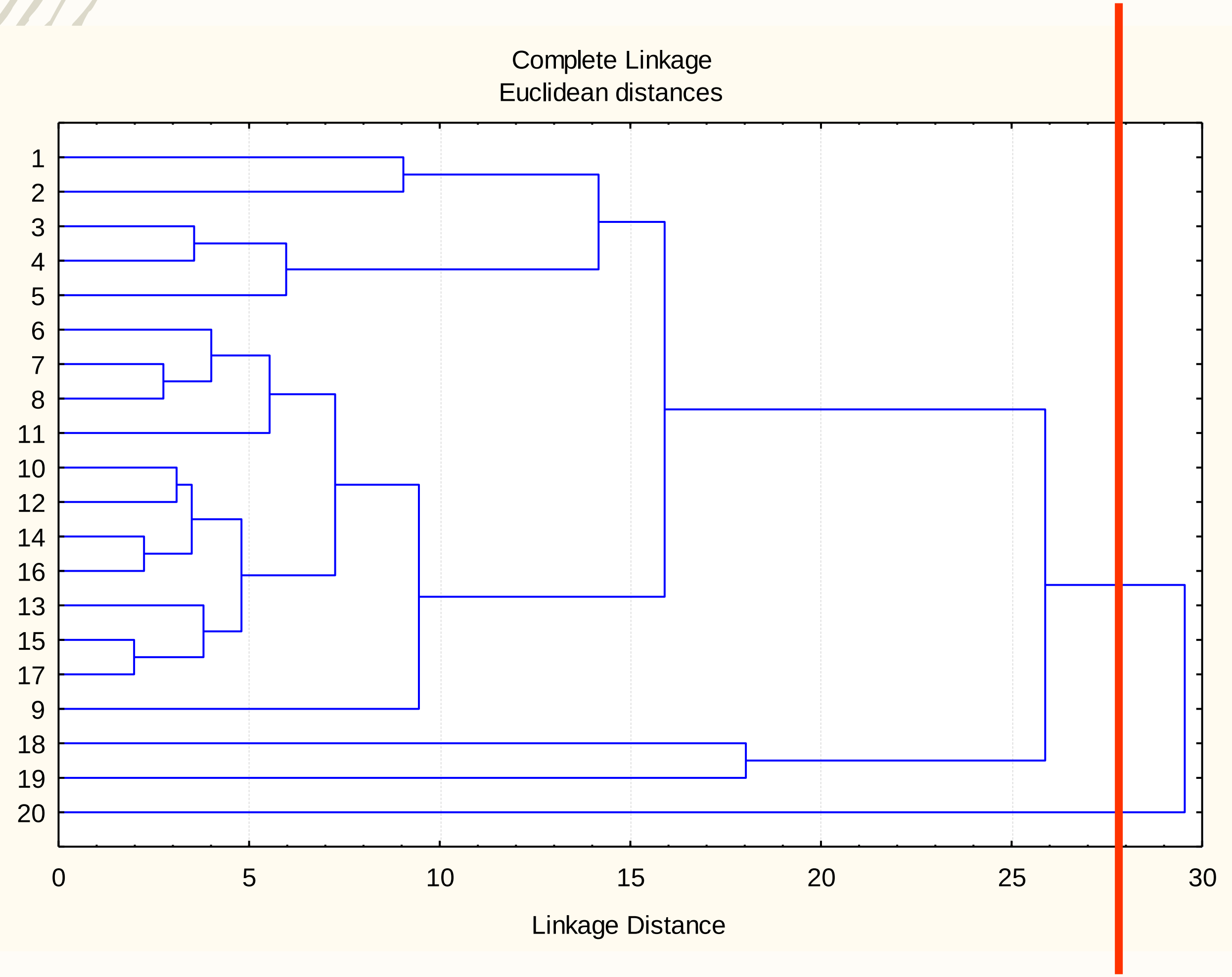


Número de grupos e interpretação dos dendrogramas

- Não existe nenhuma regra para a selecção do nº de grupos a considerar;
- Devemos procurar que os grupos sejam bem diferenciados;
- A interpretação é feita com recurso aos atributos dos elementos constituintes dos vários grupos, quer numa abordagem exploratória quer confirmatória;
- Recurso a estatísticas descritivas.

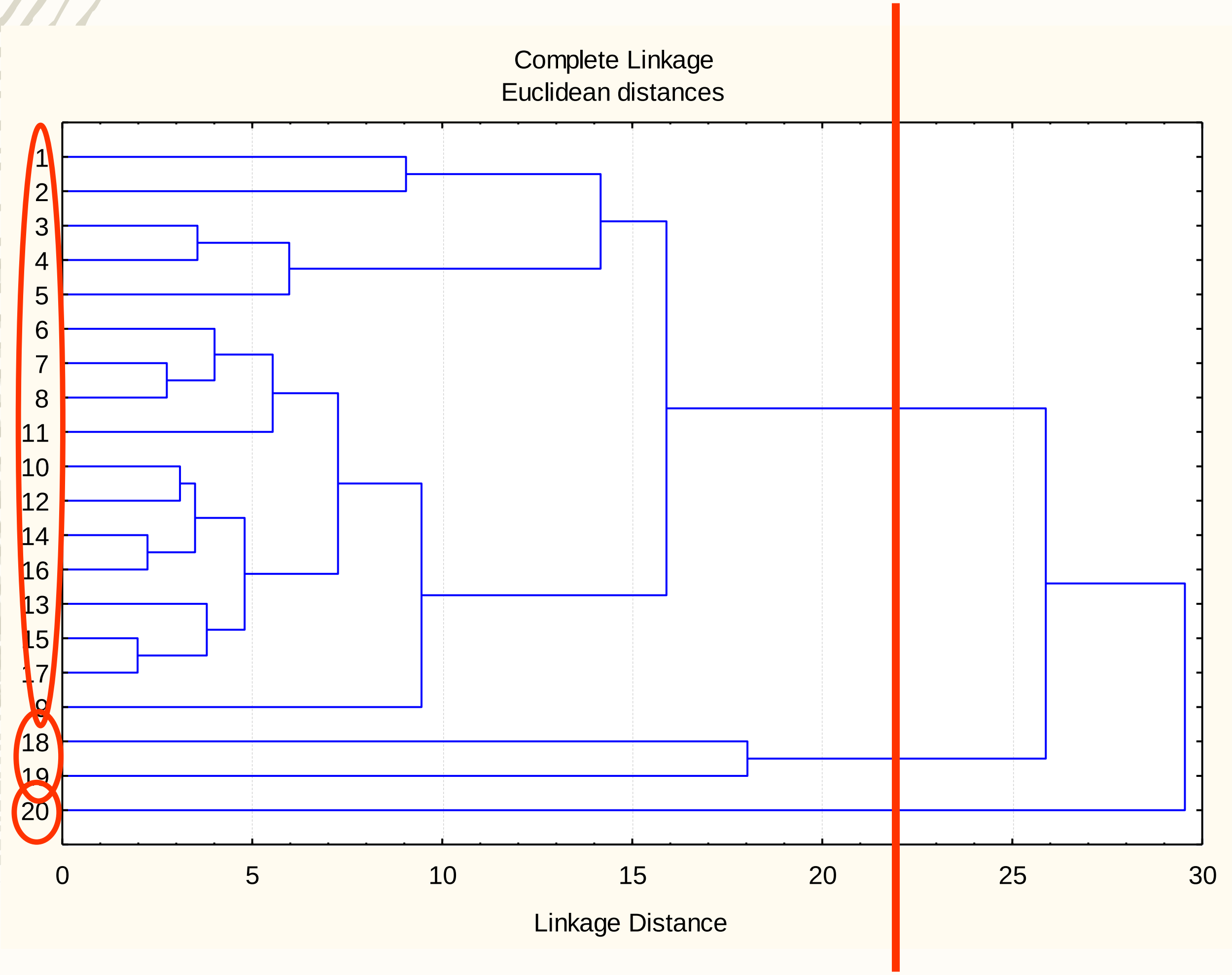
agrupamento

Número de grupos



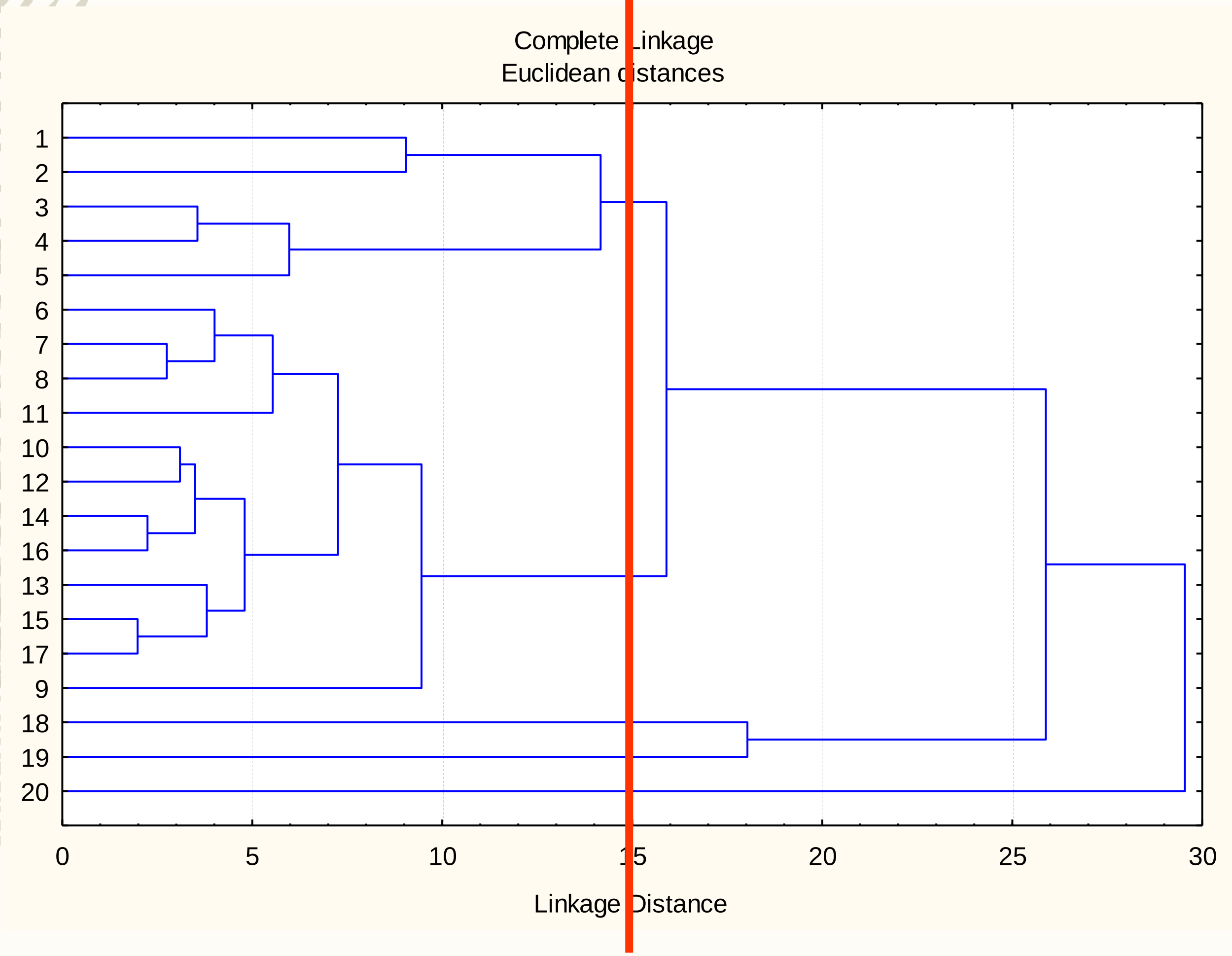
agrupamento

Número de grupos

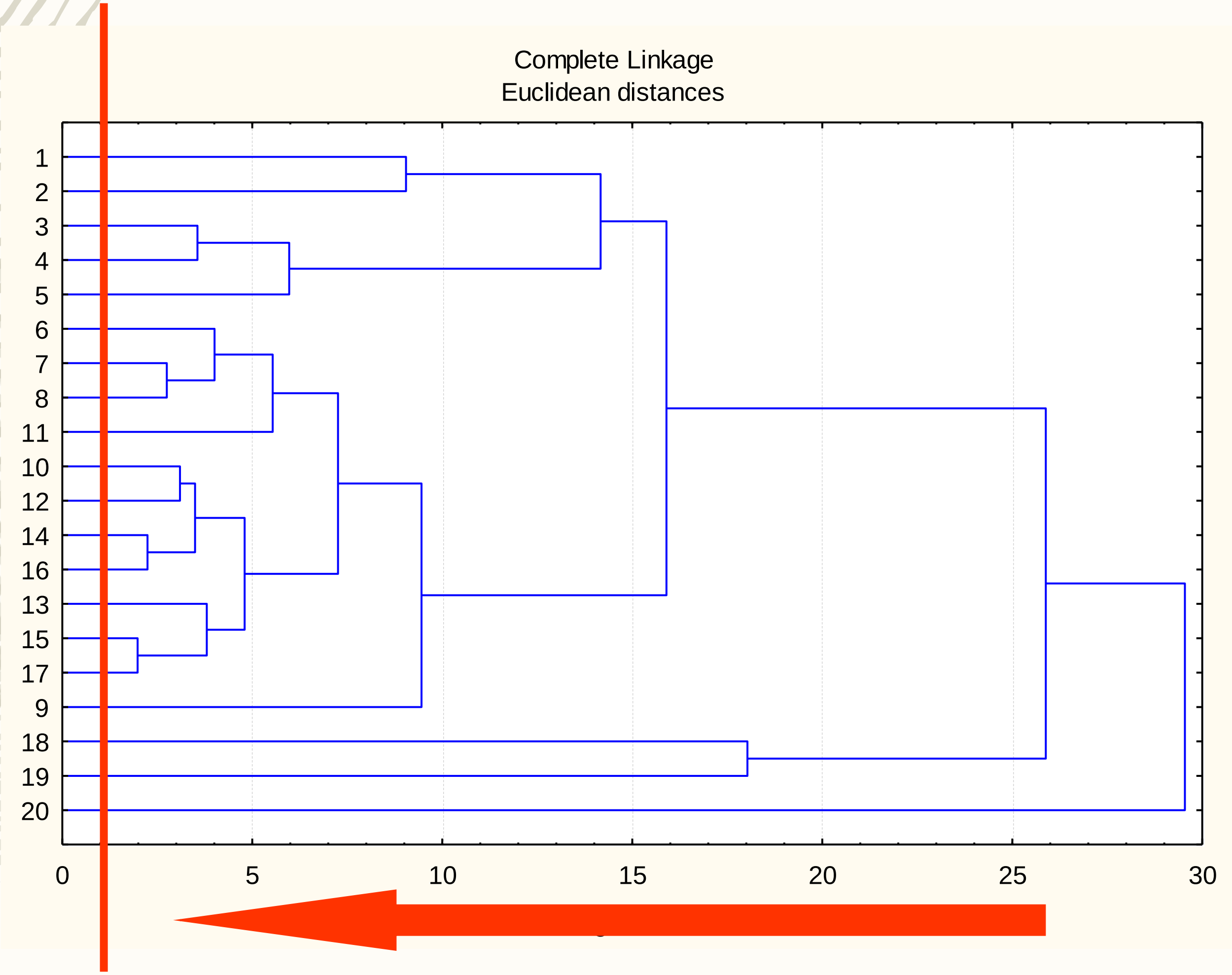


agrupamento

Número de grupos



N.º de grupos


Complete Linkage
Euclidean distances
1
2
3
4
5
6
7
8
11
10
12
14
16
13
15
17
9
18
19
20
0
5
10
15
20
25
30

Clustering steps

1. Olhar (com olhos de ver!) para os dados
2. Transformar os dados?
3. Escolher a distância (distância vs. associação)
4. Escolher o método de agrupamento
5. Validar e interpretar os resultados

Variáveis binárias
(presença ausência)

Talvêz o mais famoso método de clustering…

UPGMA

```
hclust(distancias,method="average")
```

Table 4.1 The four methods of average clustering. The names in quotes are the corresponding arguments of function **`hclust()`**

	Arithmetic average	Centroid clustering
Equal weights	Unweighted Pair-Group Method using arithmetic Averages (UPGMA) "`average`"	Unweighted Pair-Group Method using Centroids (UPGMC) "`centroid`"
Unequal weights	Weighted Pair-Group Method using arithmetic Averages (WPGMA) "`mcquitty`"	Weighted Pair-Group Method using Centroids (WPGMC) "`median`"

agrupamento

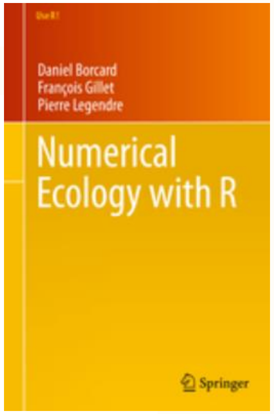
used, consider *fuzzy partitions*, in which membership is continuous (between 0 and 1). Depending on the clustering model, the result can be a single partition or a series of hierarchically nested partitions. Clustering is *not* a typical statistical method in that it does not test any hypothesis. Clustering helps bring out some features hidden in the data; it is the user who decides if these structures are interesting and worth interpreting in ecological terms.

Note that most clustering methods are computed from association matrices, which stresses the importance of the choice of an appropriate association coefficient.

4.7 Interpreting and Comparing Hierarchical Clustering Results

4.7.1 Introduction

Remember that clustering is a heuristic procedure, not a statistical test. The choices of an association coefficient and a clustering method influence the result. This stresses the importance of choosing a method that is consistent with the aims of the analysis. The

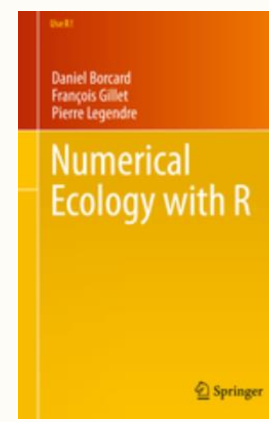


Validação dos grupos

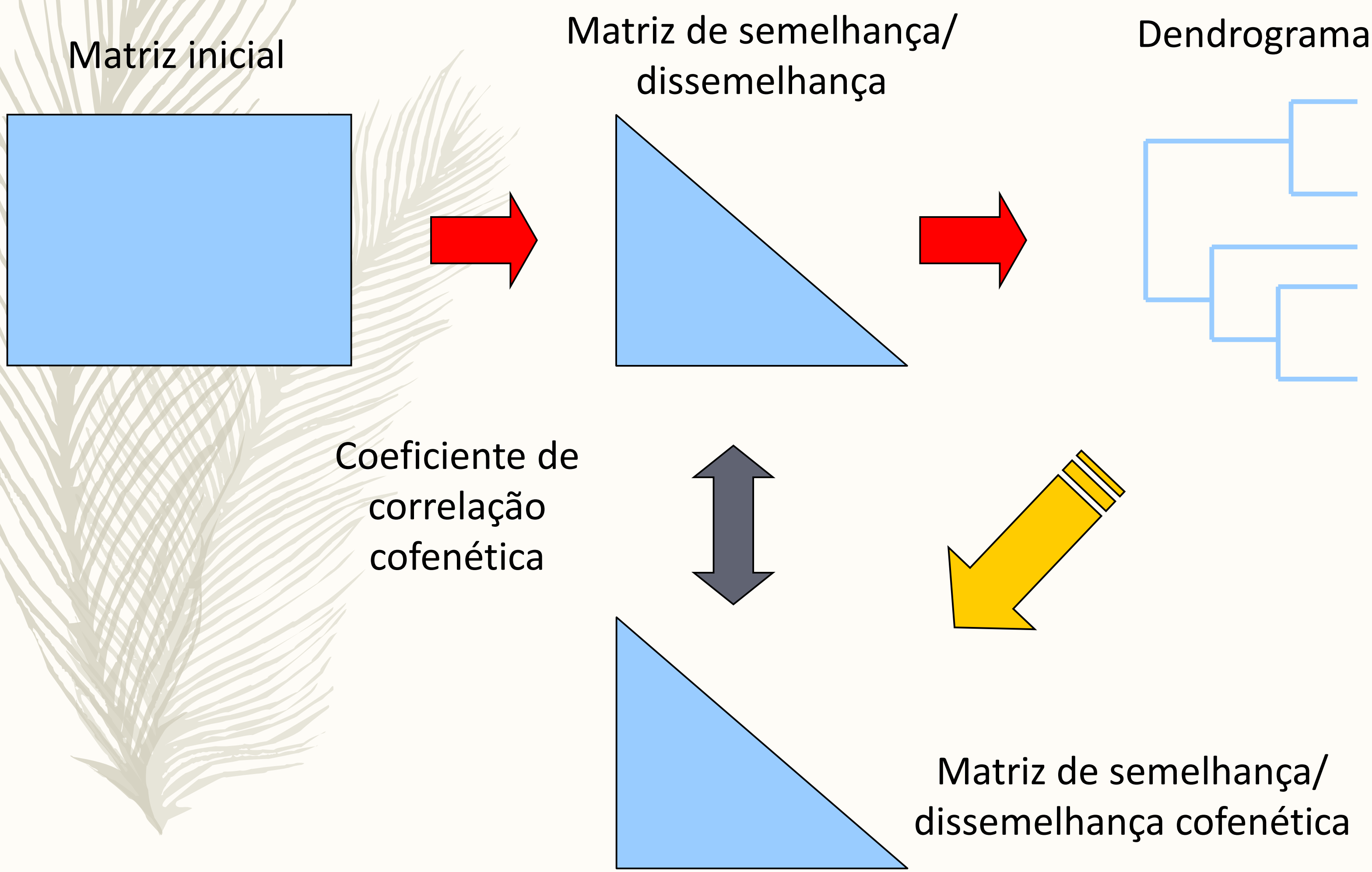
- É um procedimento importante, embora muitas vezes negligenciado;
- A principal metodologia consiste em determinar uma medida de concordância entre o resultado final (dendrograma) e a matriz de semelhança/dissemelhança inicial;
- Coeficiente de correlação cofenética.

4.7.2 *Cophenetic Correlation*

The cophenetic distance between two objects in a dendrogram is the distance where the two objects become members of the same group. Locate any two objects, start from one, and "climb up the tree" to the first node leading down to the second object: the level of that node along the distance scale is the cophenetic distance between the two objects. A cophenetic matrix is a matrix representing the cophenetic distances among all pairs of objects. A Pearson's r correlation, called the *cophenetic correlation* in this context, can be computed between the original dissimilarity matrix and the cophenetic matrix. The method with the highest cophenetic correlation may be seen as the one that produced the best clustering model for the distance matrix.



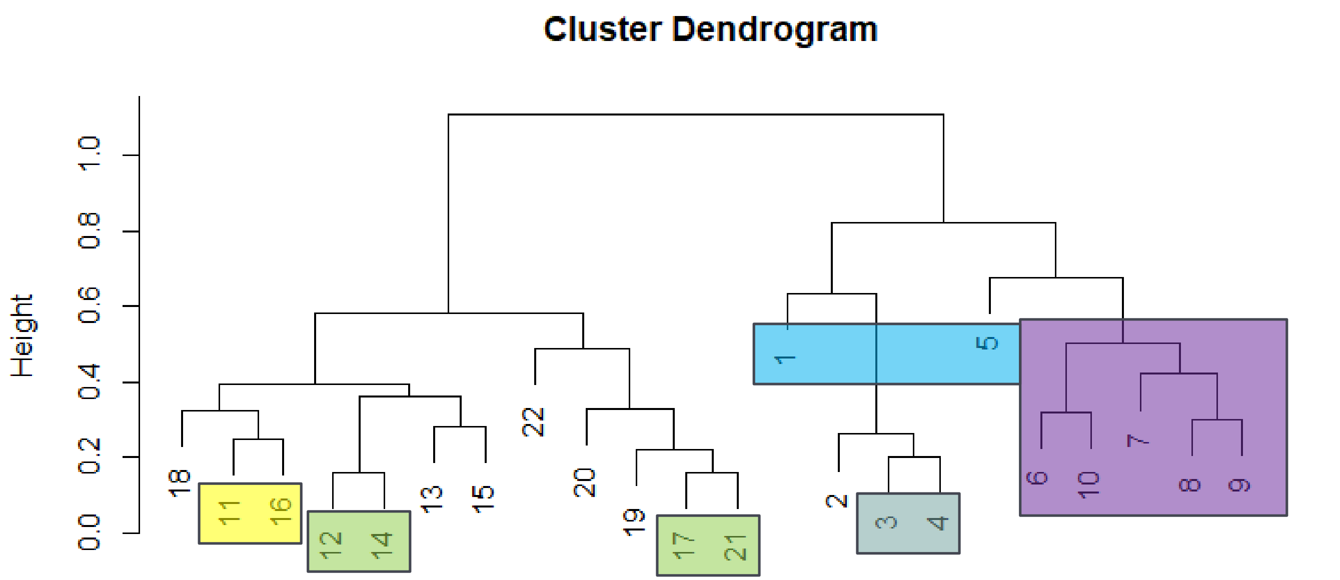
Validação dos grupos



```
habitats <- read.csv("DataTP9habitats123.csv", sep=";")
library(cluster)
teste<-hclust(dist(habitats[,-1],method="manhattan"),method="average")
par(mfrow=c(1,1))
plot(teste)
```

A primeira coluna são os *labels* dos sitios

```
> round((dist(habitats[,-1],method="manhattan")),2)
      1    2    3    4    5    6    7    8    9   10   11   12   13   14   15   16   17   18   19   20   21
2  0.70
3  0.60 0.24
4  0.60 0.28 0.20
5  0.82 0.76 0.64 0.62
6  0.90 0.72 0.58 0.48 0.64
7  0.89 1.03 0.87 0.81 0.69 0.59
8  0.96 0.98 0.86 0.78 0.68 0.48 0.47
9  0.96 1.10 0.92 0.84 0.72 0.56 0.37 0.30
10 1.04 0.86 0.68 0.60 0.66 0.32 0.53 0.44 0.42
11 1.19 1.19 1.09 1.03 0.77 0.75 0.82 0.69 0.61 0.57
12 1.45 1.33 1.29 1.25 0.81 0.97 1.00 0.85 0.73 0.67 0.42
13 1.29 1.29 1.19 1.13 0.69 0.81 0.92 0.61 0.67 0.63 0.34 0.30
14 1.53 1.41 1.41 1.35 0.93 1.05 1.08 0.85 0.83 0.77 0.48 0.16 0.32
15 1.57 1.41 1.45 1.37 0.89 1.07 1.20 0.83 0.91 0.89 0.58 0.42 0.28 0.40
16 1.42 1.34 1.32 1.26 0.82 0.94 0.97 0.88 0.78 0.68 0.25 0.23 0.29 0.27 0.41
17 1.54 1.42 1.42 1.36 1.20 1.06 1.11 0.86 0.84 0.80 0.61 0.61 0.53 0.51 0.37 0.54
18 1.45 1.33 1.33 1.27 1.07 0.97 1.00 0.79 0.75 0.69 0.34 0.42 0.44 0.38 0.48 0.31 0.33
19 1.51 1.39 1.39 1.33 1.07 1.03 1.06 0.83 0.81 0.75 0.42 0.42 0.40 0.32 0.30 0.35 0.19 0.26
20 1.46 1.36 1.38 1.32 1.26 1.04 1.03 0.82 0.86 0.84 0.75 0.71 0.65 0.59 0.57 0.72 0.30 0.61 0.41
21 1.56 1.40 1.44 1.36 1.14 1.06 1.19 0.84 0.90 0.88 0.65 0.65 0.51 0.55 0.31 0.60 0.16 0.45 0.25 0.28
22 1.63 1.69 1.65 1.67 1.65 1.45 1.44 1.21 1.21 1.21 0.94 1.00 0.98 0.86 0.82 0.87 0.45 0.62 0.58 0.41 0.51
```




```
> round(cophenetic(teste),2)
      1    2    3    4    5    6    7    8    9   10   11   12   13   14   15   16   17   18   19   20   21
2  0.63
3  0.63 0.26
4  0.63 0.26 0.20
5  0.82 0.82 0.82 0.82
6  0.82 0.82 0.82 0.82 0.68
7  0.82 0.82 0.82 0.82 0.68 0.50
8  0.82 0.82 0.82 0.82 0.68 0.50 0.42
9  0.82 0.82 0.82 0.82 0.68 0.50 0.42 0.30
10 0.82 0.82 0.82 0.82 0.68 0.32 0.50 0.50 0.50
11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11
12 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 0.40
13 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 0.40 0.36
14 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 0.40 0.16 0.36
15 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 0.40 0.36 0.28 0.36
16 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 0.25 0.40 0.40 0.40 0.40
17 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 0.58 0.58 0.58 0.58 0.58 0.58
18 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 0.32 0.40 0.40 0.40 0.40 0.32 0.58
19 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 0.58 0.58 0.58 0.58 0.58 0.58 0.22 0.58
20 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 0.58 0.58 0.58 0.58 0.58 0.58 0.33 0.58 0.33
21 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 0.58 0.58 0.58 0.58 0.58 0.58 0.16 0.58 0.22 0.33
22 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 0.58 0.58 0.58 0.58 0.58 0.58 0.49 0.58 0.49 0.49 0.49
```

```
> round(cophenetic(teste),2)
      1    2    3    4    5    6    7    8    9   10   11   12   13   14   15   16   17   18   19   20   21
2  0.63
3  0.63 0.26
4  0.63 0.26 0.20
5  0.82 0.82 0.82 0.82
6  0.82 0.82 0.82 0.82 0.68
7  0.82 0.82 0.82 0.82 0.68 0.50
8  0.82 0.82 0.82 0.82 0.68 0.50 0.42
9  0.82 0.82 0.82 0.82 0.68 0.50 0.42 0.30
10 0.82 0.82 0.82 0.82 0.68 0.32 0.50 0.50 0.50
11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11
12 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 0.40
13 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 0.40 0.36
14 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 0.40 0.16 0.36
15 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 0.40 0.36 0.28 0.36
16 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 0.25 0.40 0.40 0.40 0.40
17 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 0.58 0.58 0.58 0.58 0.58 0.58
18 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 0.32 0.40 0.40 0.40 0.40 0.32 0.58
19 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 0.58 0.58 0.58 0.58 0.58 0.58 0.22 0.58
20 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 0.58 0.58 0.58 0.58 0.58 0.58 0.33 0.58 0.33
21 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 0.58 0.58 0.58 0.58 0.58 0.58 0.16 0.58 0.22 0.33
22 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 1.11 0.58 0.58 0.58 0.58 0.58 0.58 0.49 0.58 0.49 0.49 0.49
```

matriz de distâncias originais

matriz de distâncias cofenéticas

```
> cor(as.numeric(dist(habitats[,-1],method="manhattan")),as.numeric(cophenetic(teste)))
[1] 0.7985337
```

Coeficiente de correlação cofenética

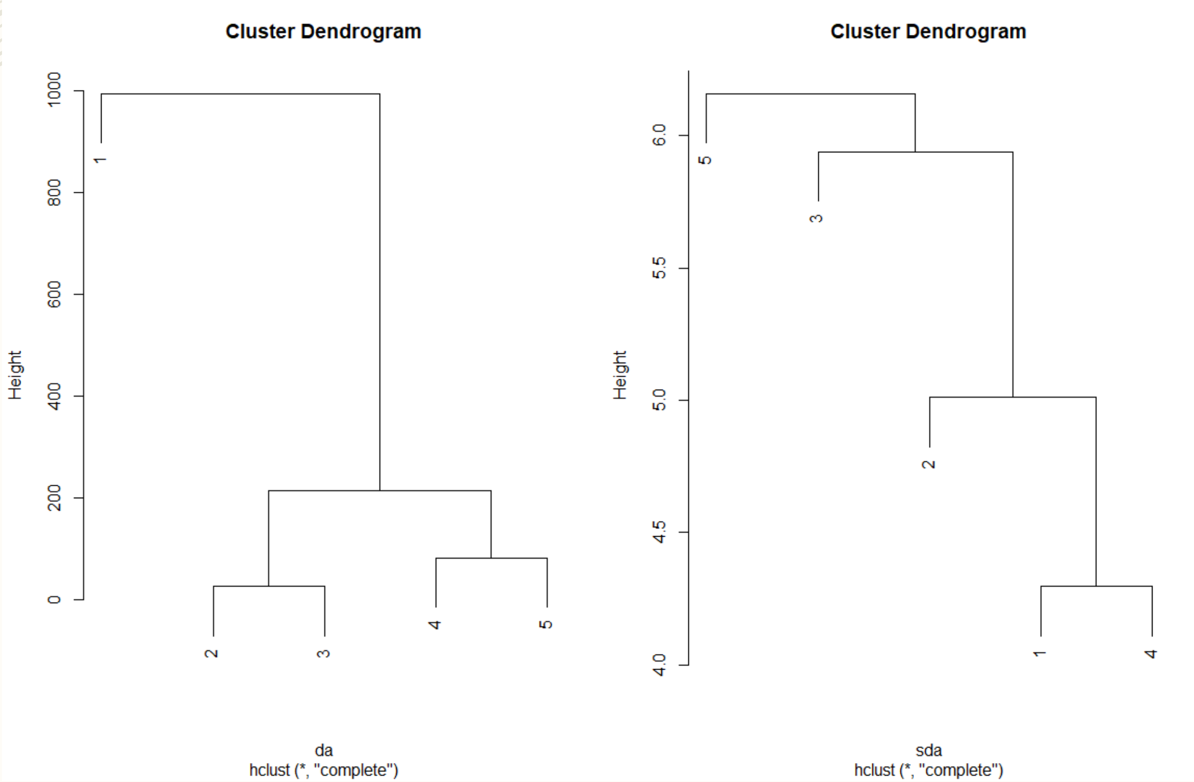
https://stats.stackexchange.com/questions/149852/validate-dendrogram-in-cluster-analysis-what-is-the-meaning-of-cophenetic-corre

```
set.seed(12345)
abund=matrix(rpois(75,lambda=8),ncol=15,nrow=5)
#make 1 species really abundant
abund[,1]=c(1000,200,200,10,10)
#and one with the inverse pattern, but less abundant
abund[,2]=c(10,10,200,200,1000)/10
#get distance matrix
da=dist(abund)
hcdaC=hclust(da,method="complete")
#get distance matrix over scaled data
sda=dist(scale(abund))
hcsdaC=hclust(sda,method="complete")
par(mfrow=c(1,2))
plot(hcdaC)
plot(hcsdaC)
```

Transformar os dados pode ser fundamental!

unscaled

scaled



```
> sda
         1        2        3        4
2 5.010198
3 5.675741 5.938724
4 4.296223 4.906653 4.662499
5 5.926676 6.032184 6.158771 5.805220
```

Cluster Dendrogram

Cluster Dendrogram

sda
hclust (*, "single")

sda
hclust (*, "complete")

```
> sda
         1        2        3        4
2 5.010198                           
3 5.675741 5.938724                  
4 4.296223 4.906653 4.662499         
5 5.926676 6.032184 6.158771 5.805220
```