

RICCHEZZE SPECIFICHE E FONDAMENTALI



Site A



Site C



Site B



Site D

INDICES DE RICHESSE SPÉCIFIQUE



- × **Richesse spécifique numérique:** Nombre d'espèces/nombre d'individus ou biomasse
- × **Densité d'espèces:**
Nombre d'espèces/m²

INDICES DE RICHESSE SPÉCIFIQUE (2)



- × Indice de diversité de **Margalef** D_{mg}
- × Indice de **Menhinick** D_{mn}

$$D_{Mg} = (S - 1) / \ln N$$

and Menhinick's index (Whittaker, 1977) D_{Mn}

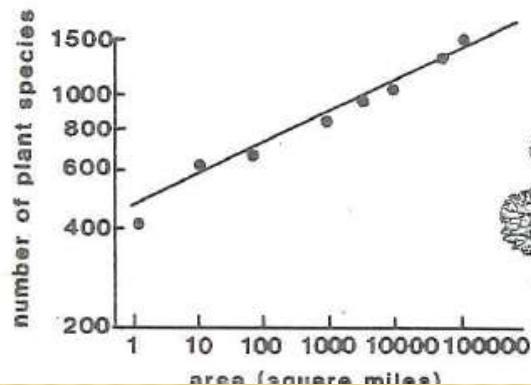
$$D_{Mn} = S / \sqrt{N}$$

INDICES DE RICHESSE SPÉCIFIQUE (1)

$$E(S) = \sum \left\{ 1 - \left[\frac{\binom{N-N_i}{n}}{\binom{N}{n}} \right] \right\}$$

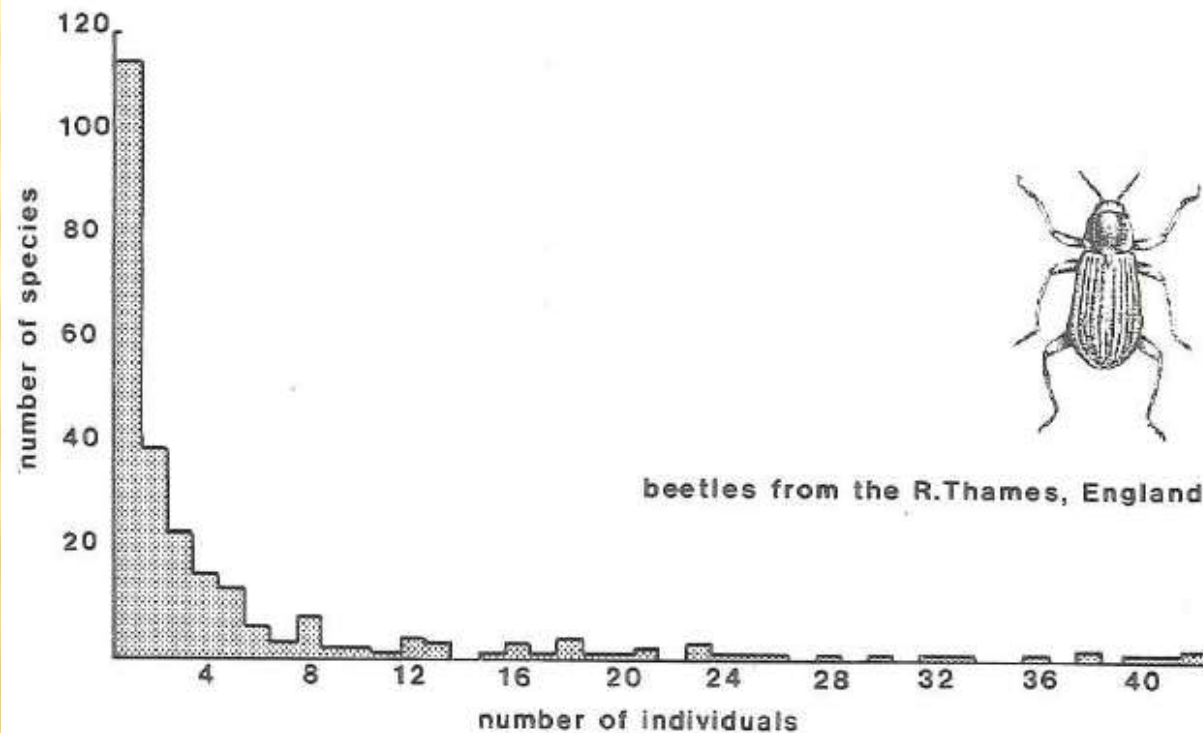
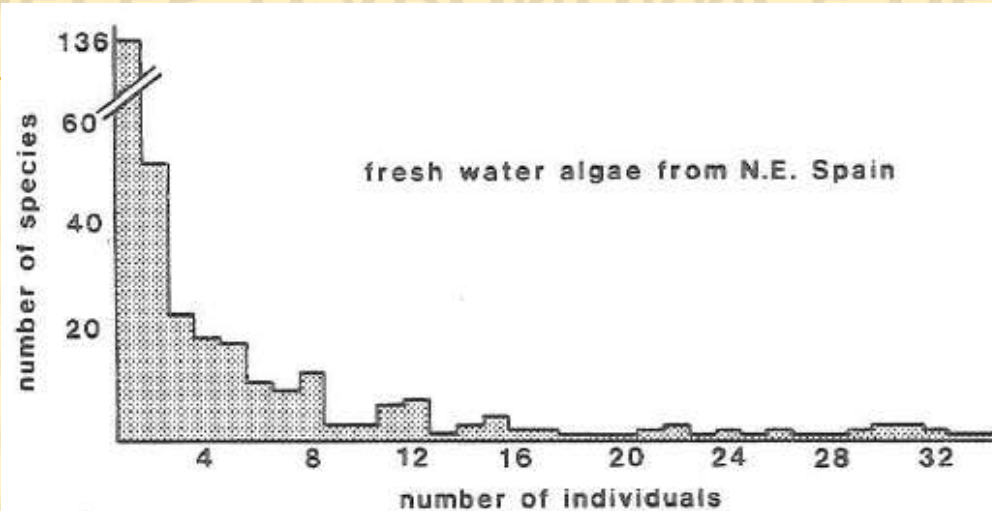
where $E(S)$ = expected number of species;
 n = standardized sample size;
 N = total number of individuals recorded;
 N_i = number of individuals in the i th species.

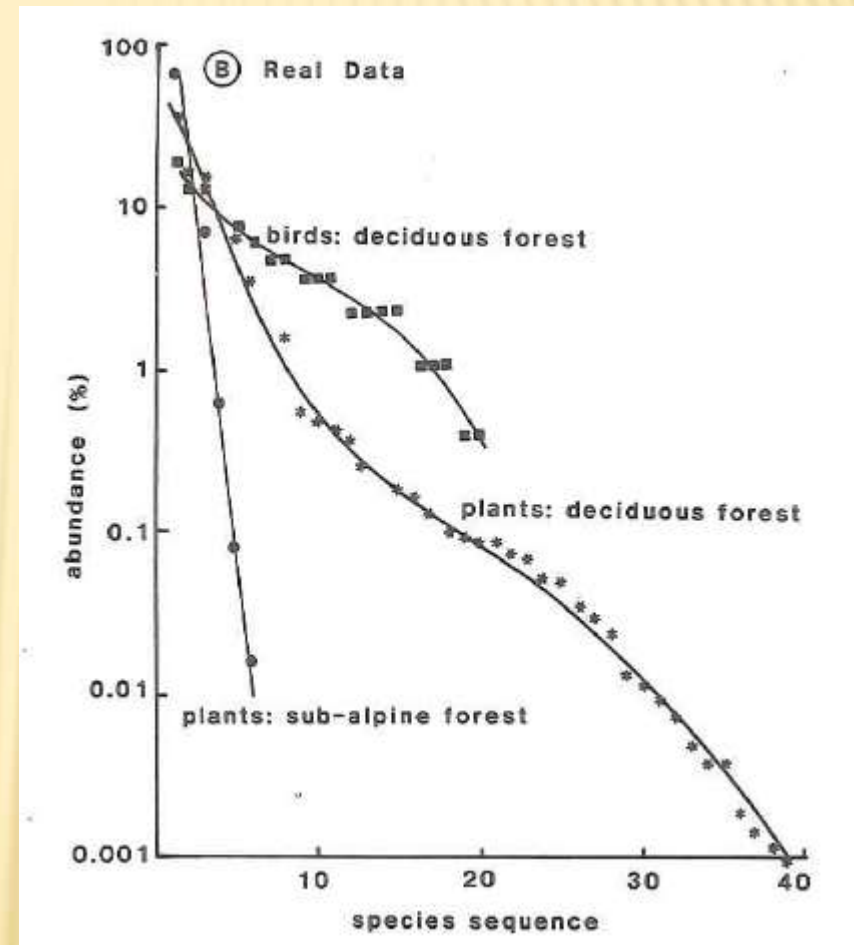
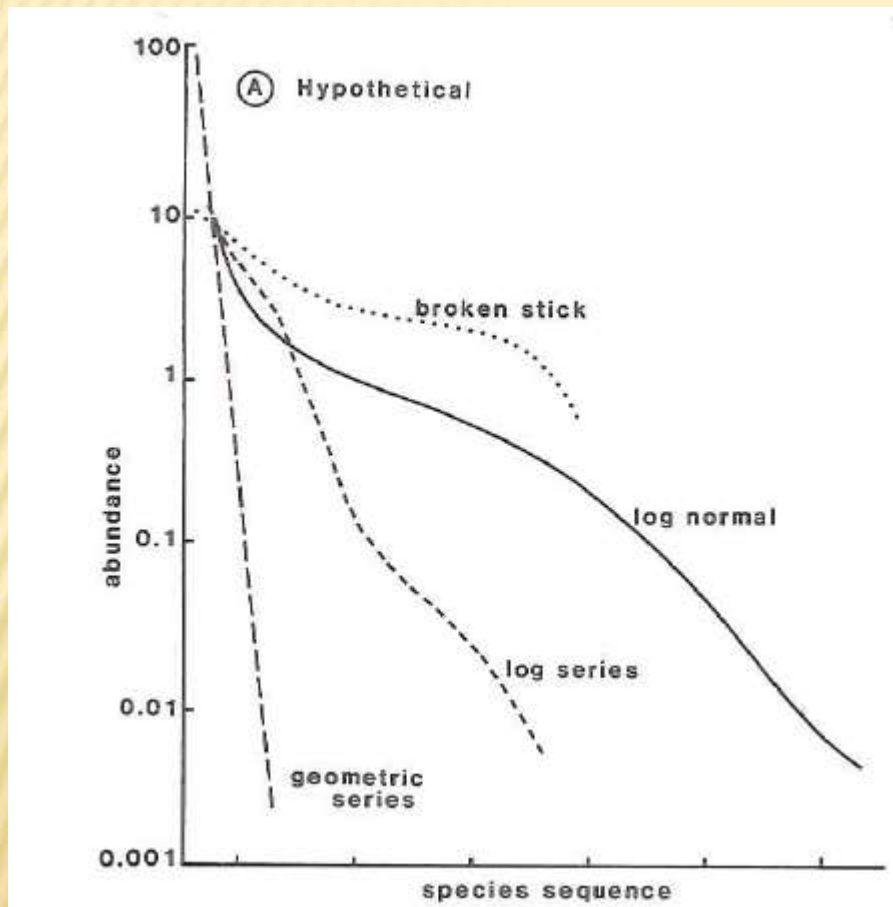
- ✗ **Raréfaction** (Sanders)
- ✗ Hurlbert (1971)



- ✗ **Richesse spécifique est associée à la taille de l'échantillon!**

MODÈLES D'ABONDANCE DES ESPÈCES



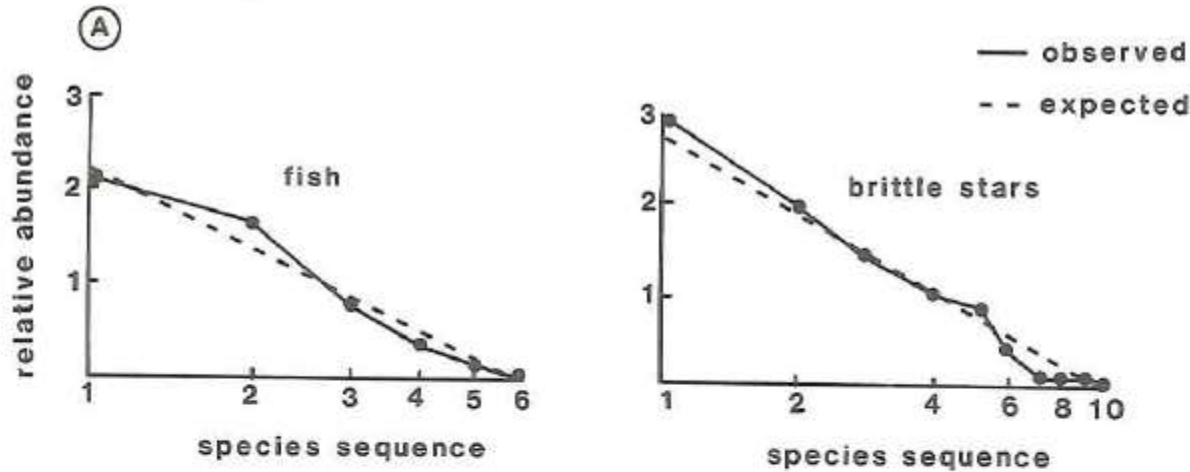


MÉTHODES GRAPHIQUES

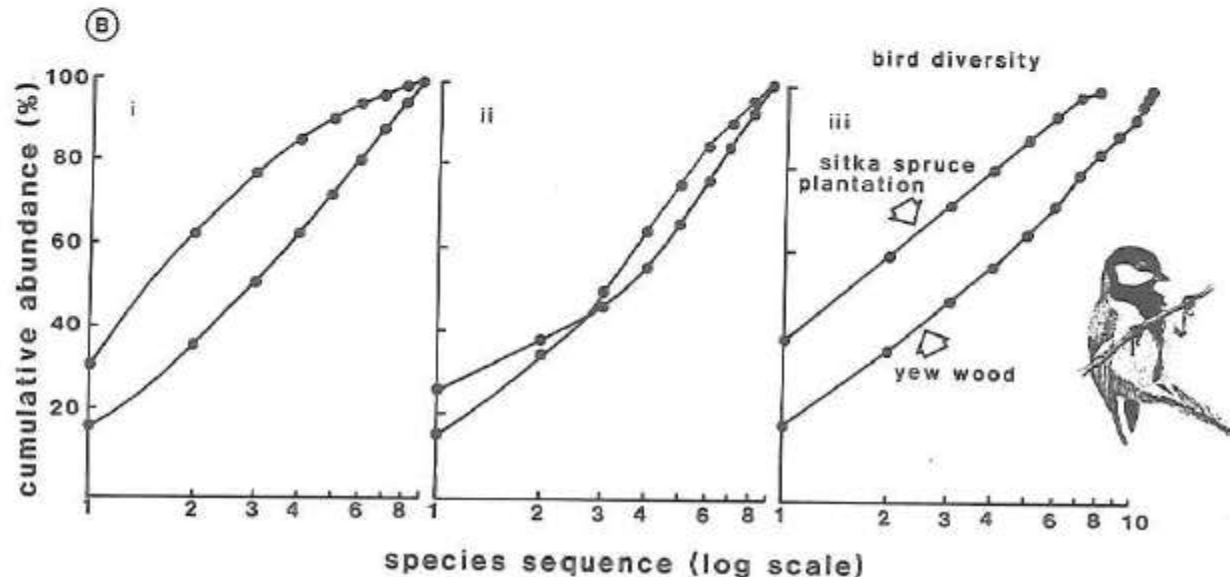
MÉTHODES GRAPHIQUES: ABONDANCE

- ✗ **Geometric series**: Rank/abondance are une manière de présenter les données.
- ✗ **Log series**: Distribution de fréquence où le nombre d'espèces est présenté contre le nombre d'individus/espèces
- ✗ **Log normal**: idem mais avec l'abscisse en échelle logarithmique.
- ✗ **Broken stick**: Rank/abondance mais où les rangs et non pas l'abondance est en log.

MÉTHODES GRAPHIQUES: ABONDANCE



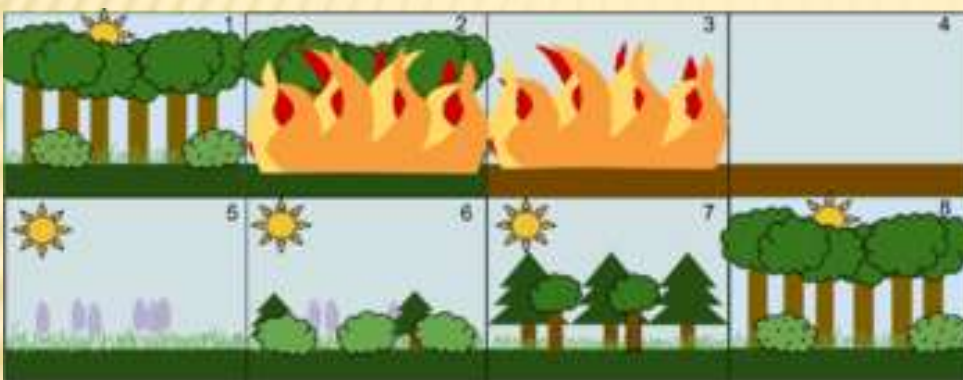
A. Graphes typiques du modèle « broken stick »



K-dominance plots (Platt et al. 1984)

SÉRIE GÉOMÉTRIQUE

- ✗ Situation où une espèce dominante va vider une proportion k d'une ressource limitée



Ce pattern d'abondance d'espèces est surtout trouvée dans les milieux pauvres ou dans les premiers stages de succession.  Log series

$$n_i = NC_k k(1 - k)^{i-1}$$

where n_i = the number of individuals in the i th species;

N = the total number of individuals;

$C_k = [1 - (1 - k)^s]^{-1}$ and is a constant which ensures

$$\text{Sum}(n_i) = N$$

LOG SERIES



The log series takes the form:

$$\alpha x, \frac{\alpha x^2}{2}, \frac{\alpha x^3}{3}, \dots, \frac{\alpha x^n}{n} \quad (2.5)$$

αx being the number of species predicted to have one individual, $\alpha x^2/2$ those with two and so on (Fisher *et al.*, 1943; Poole, 1974).

The total number of species, S , is obtained by adding all the terms series which reduces to the following equation

$$S = \alpha [-\ln(1 - x)]$$

x is estimated from the iterative solution of

$$S/N = (1 - x)/x [-\ln(1 - x)]$$

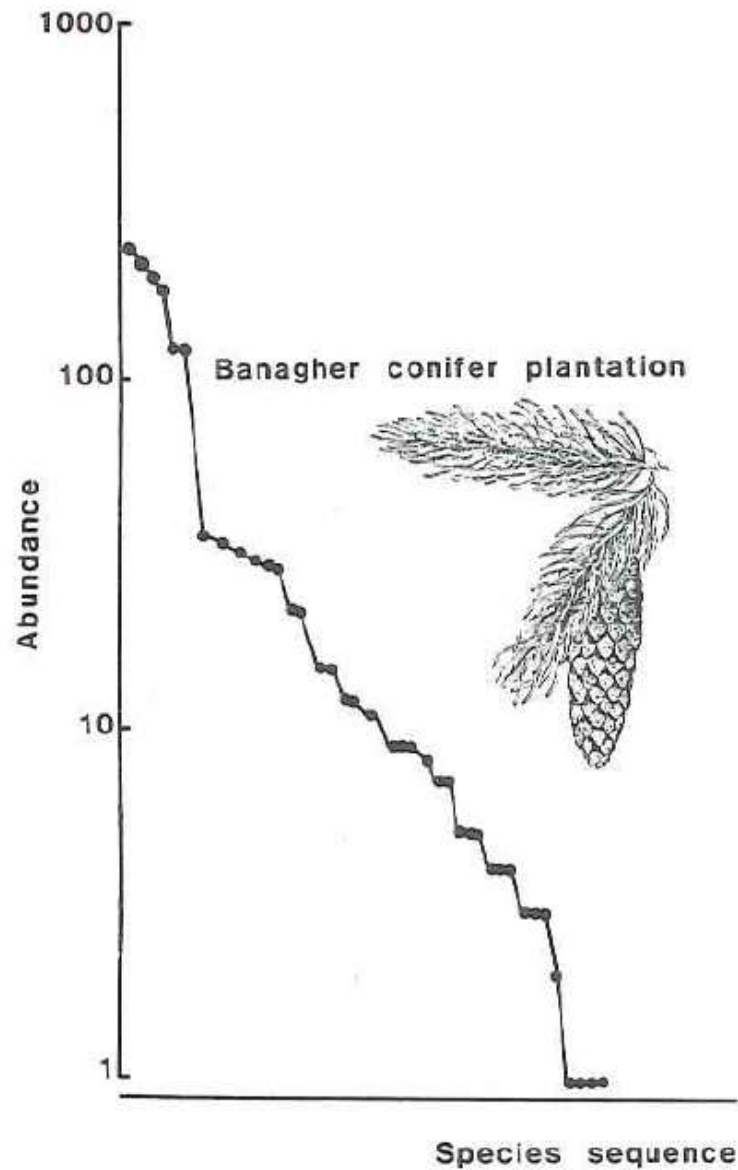
where N = the total number of individuals.

- ✗ **Fisher** a été le premier à l'utiliser pour modéliser la relation entre le nombre d'espèces et le nombre d'individus de ces espèces.
- ✗ Très utilisé en entomologie
- ✗ Lien étroit entre la série géométrique et log series
- ✗ Prévue dans les situations où les espèces arrivent dans un milieu non saturé à **intervalles aléatoires** et où ils occupent des fractions de la niche.

Intervalles réguliers = série géométrique.

Comme pour la série géométrique, un ou plusieurs facteurs dominant l'écologie d'un peuplement.

DIVERSITÉ DE LA VÉGÉTATION DANS UNE PLANTATION DE CONIFÈRES AU PAYS DE GALLES (U.K.)



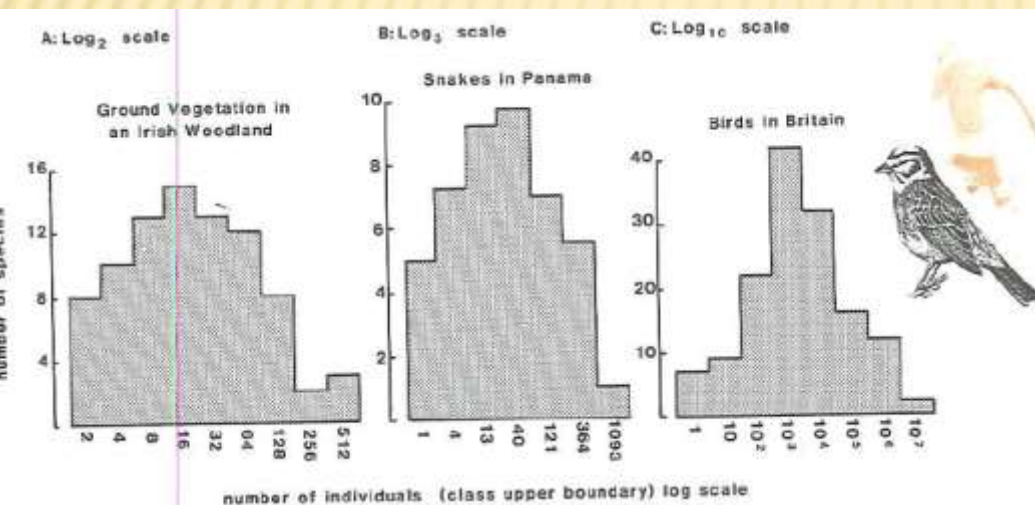
DISTRIBUTION LOG NORMAL

Preston (1948)

The distribution is usually written in the form:

$$S(R) = S_0 \exp(-a^2 R^2)$$

where $S(R)$ = the number of species in the R th octave (i.e. class) and left of the symmetrical curve;
 S_0 = the number of species in the modal octave;
 $a = (2\sigma^2)^{1/2}$ = the inverse width of the distribution.



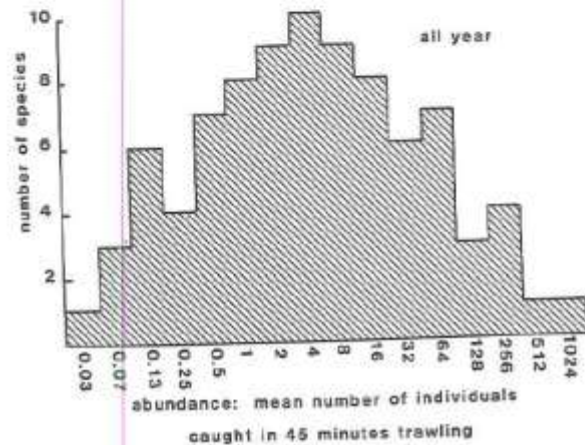
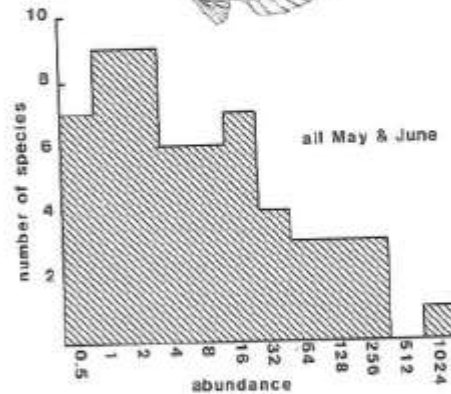
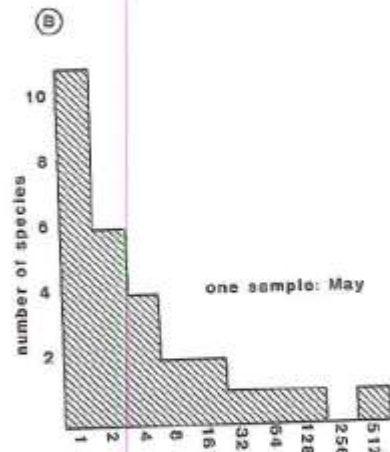
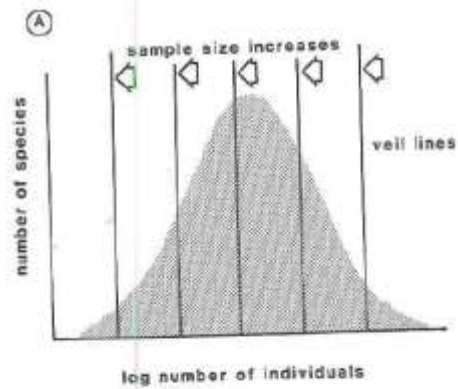
Peuplement « patchy »: Espèces rares (65%), Espèces intermédiaires (25%) et espèces abondantes (10%)

✗ La majorité des peuplements étudiés par les écologistes exhibent une distribution de l'abondance des espèces en log normal (peuplement large, mature et varié).

Distribution de la richesse aux USA et dans le U.K.

Central Limit Theorem: large number of factors qui déterminent une variable

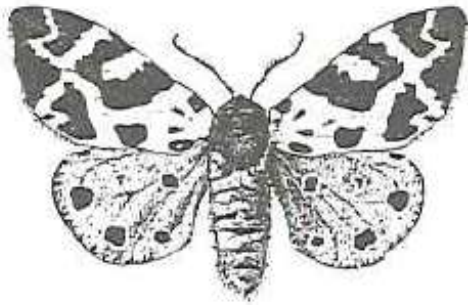
✗ One resource axis which is randomly and sequentially split (Pielou 1975)



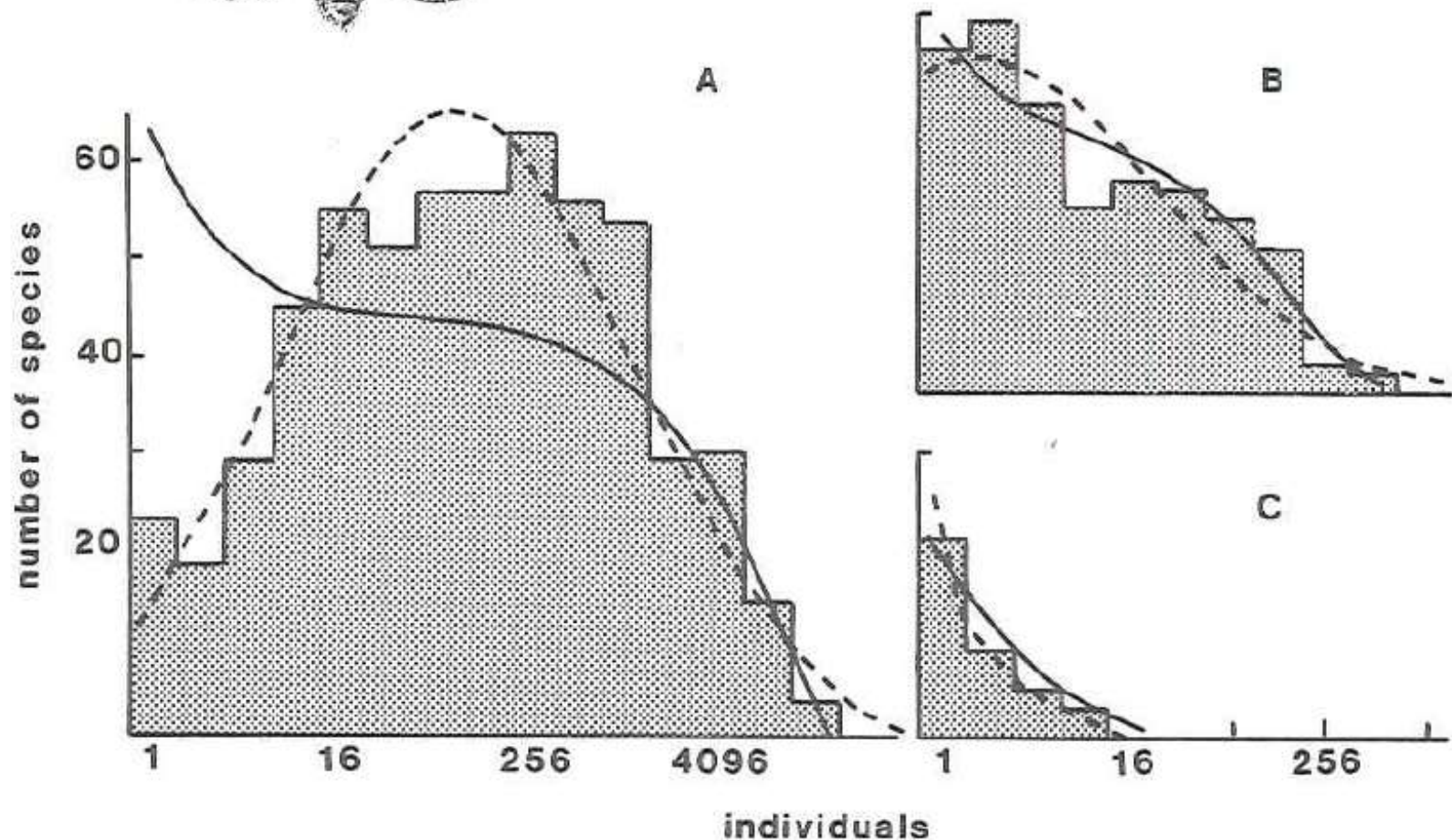
Problems with fitting a log normal (Pielou)



DISTRIBUTIONS LOG SERIES ET LOG NORMAL



— log series
--- log normal

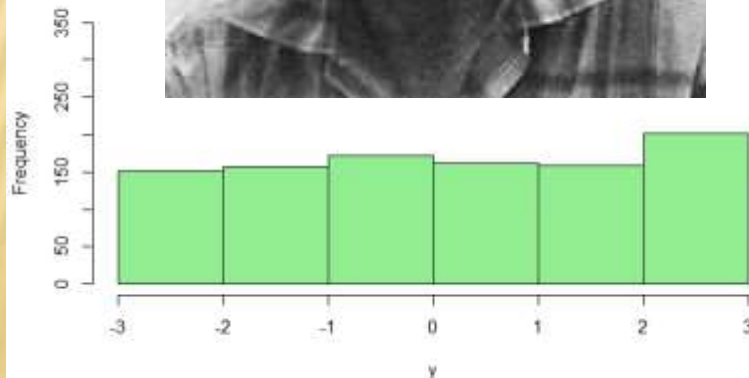
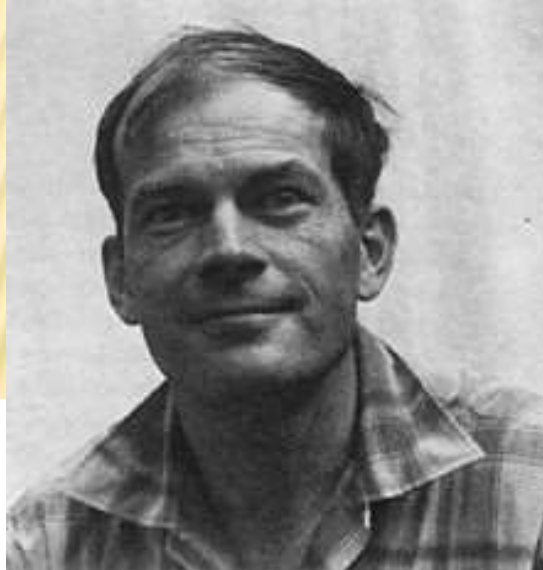


MODÈLE BROKEN STICK (BÂTON BRISÉ)

May (1975), after Webb (1974), expresses the model in terms of a standard species abundance distribution.

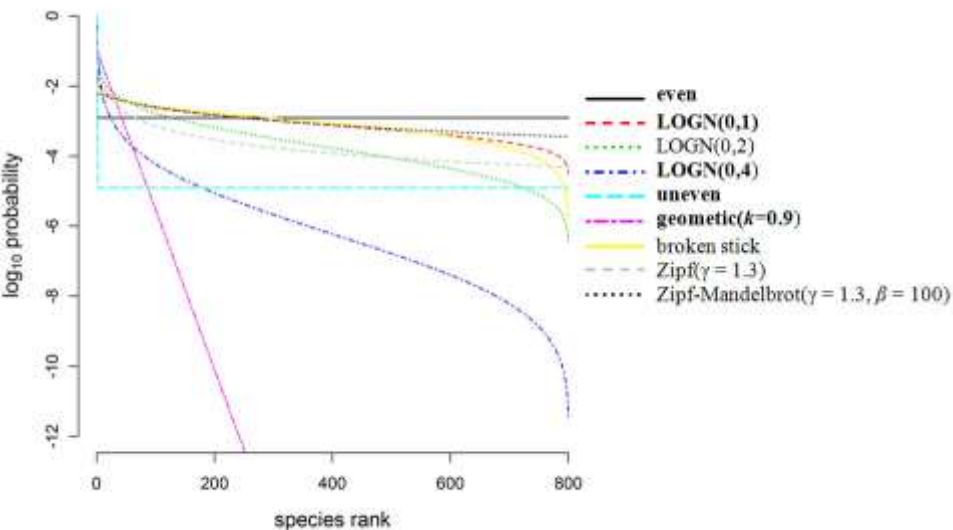
$$S(n) = [S(S-1)/N] (1 - n/N)^{S-2} \quad (2.14)$$

where $S(n)$ = the number of species in the abundance class with n individuals.

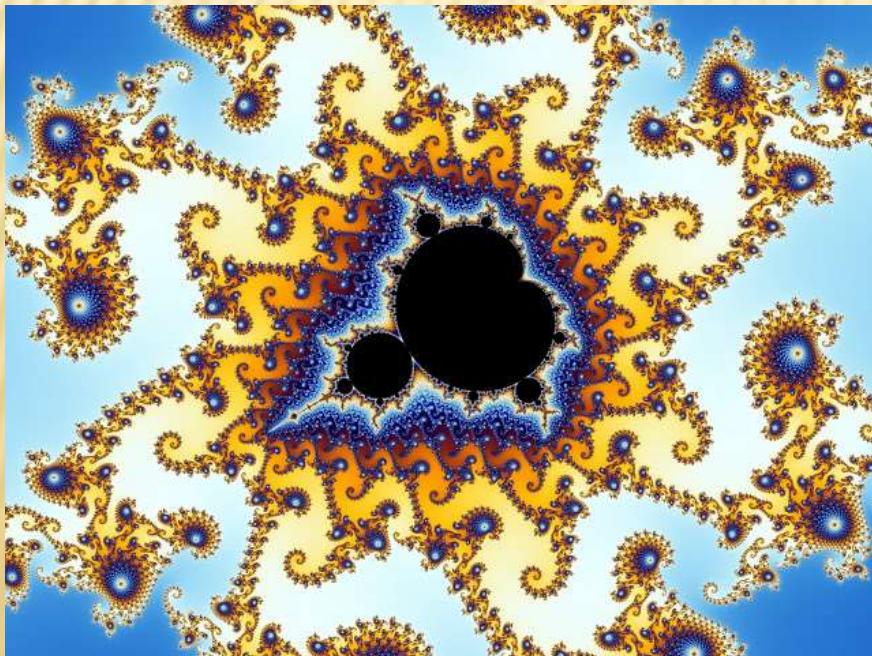


- ✗ « Random niche boundary »
- ✗ **MacArthur** (1957)
- ✗ Niche space (une seule ressource) = un bâton brisé de manière simultanée et aléatoire.
- ✗ Reflète un état plus ou moins équitable
~ **distribution uniforme**
« biologique »
- ✗ Indice d'un partage d'un facteur écologique important

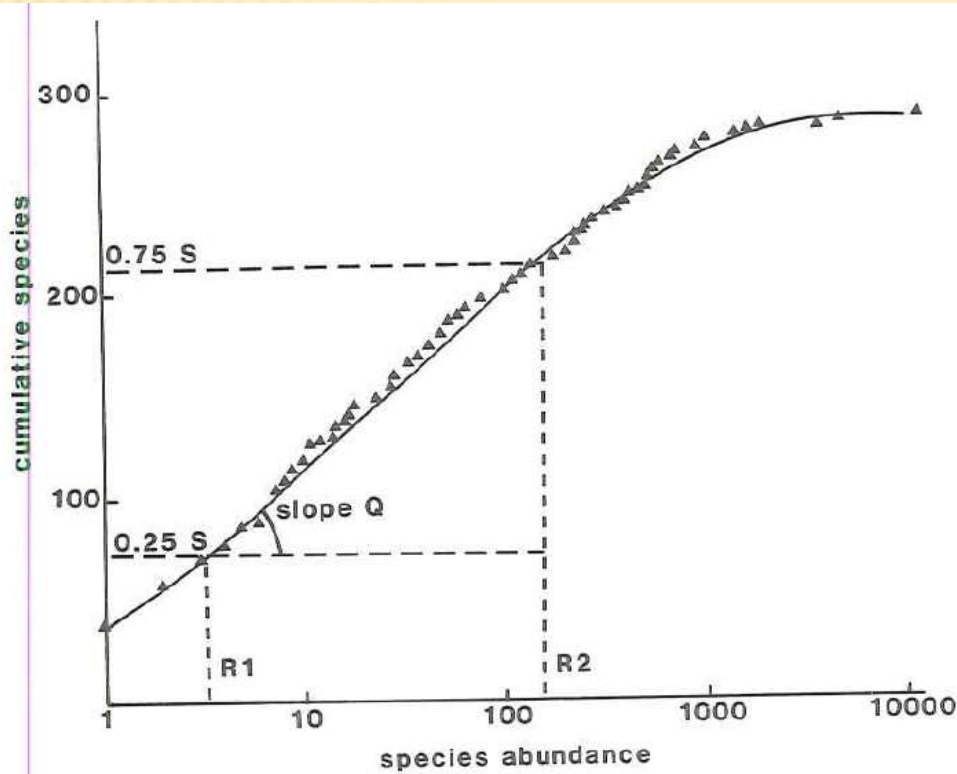
AUTRES MODÈLES



- ✗ 1) Zipf-Mandelbrot model (linguistics and information theory)
- ✗ 2) Dynamics model (Hughes)
- ✗ Truncated negative binomial distribution (Pielou): species abundance
- ✗ 4) Gamma distribution (échelle continue) (Pielou)



Q-STATISTICS



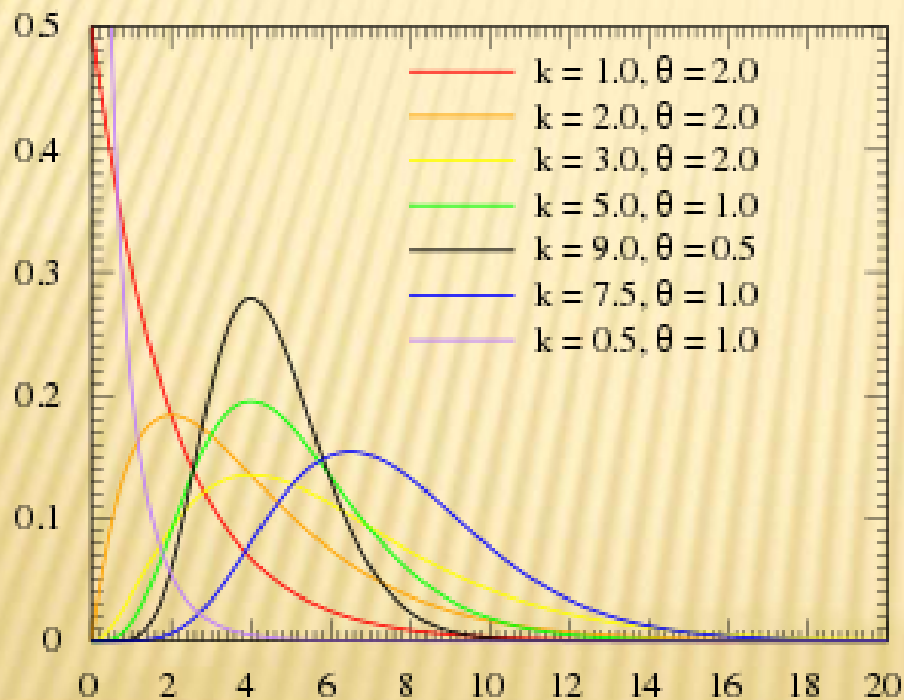
$$Q = \frac{\frac{1}{2}n_{R1} + \sum_{R1+1}^{R2-1} n_r + \frac{1}{2}n_{R2}}{\log(R2/R1)}$$

where n_r = the total number of species with abundance R ;
 S = the total number of species in the sample;
 $R1$ and $R2$ are the 25% and 75% quartiles;
 n_{R1} = the number of individuals in the class where $R1$ falls;
 n_{R2} = the number of individuals in the class where $R2$ falls.

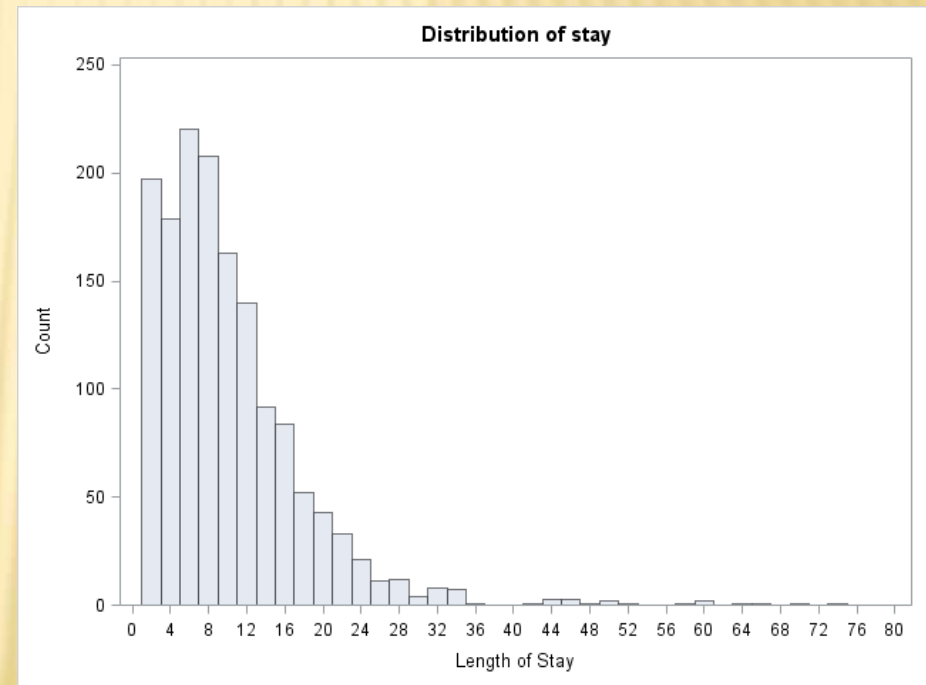
- ✗ Mesure de la diversité sans ajustement d'un modèle.
- ✗ Indice basé sur la mesure de la pente inter-quartile de la courbe de l'abondance cumulative des espèces.
- ✗ Pas de biais des espèces très rares ou très abondantes.

DISTRIBUTIONS

✕ Gamma



✕ Truncated negative binomial



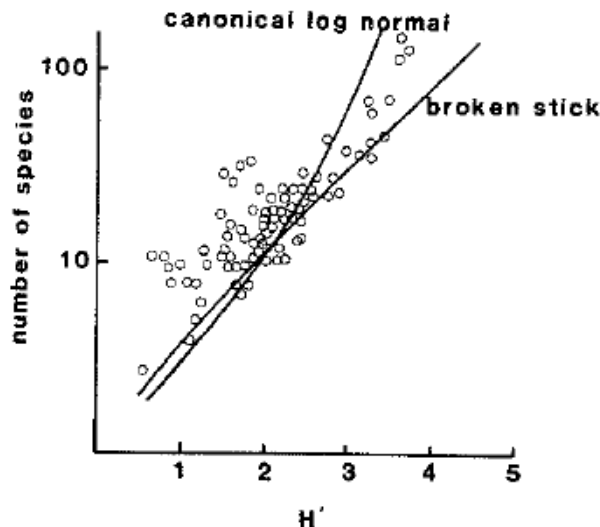
INDICES BASÉS SUR L'ABONDANCE PROPORTIONNELLE

Indice de Shannon

$$H' = -\sum p_i \ln p_i$$

Equitabilité

$$E = H'/H_{\max} = H'/\ln S$$



- ✗ Une alternative aux mesures de la diversité
- ✗ Indices non-parametriques (Southwood)
- ✗ *2 catégories d'indices:*
- ✗ A) Basés sur la théorie de Information
- ✗ **Indice de Shannon et Wiener**
- ✗ **Assume que les individus sont échantillonnés de manière aléatoire à partir d'une population infinie**

Species	Number of individuals
<i>Popamyia flava</i>	235
<i>Hydropsyche orris</i>	218
<i>Cheumatopsyche analis</i>	192
<i>Ocestis inconspicua</i>	87
<i>Hydropsyche betteni</i>	20
<i>Athripsodes transversus</i>	11
<i>Leptocella candida</i>	11
<i>Leptocella exquisita</i>	8
<i>Cheumatopsyche campyla</i>	7
<i>Polycentropus cinereus</i>	4
<i>Ocestus cinereus</i>	3
<i>Nyctiophylax vestitus</i>	2
<i>Cheumatopsyche aphanata</i>	2
<i>Neureclepsis crepuscularis</i>	1
<i>Triaenodes aba</i>	1
Shannon diversity $H' = 1.69$	
Brillouin diversity $HB = 1.65$	

(b) The Shannon index, unlike the Brillouin index, does not vary providing the number of species and their relative proportions remain constant.

	Number of individuals	
	Sample 1	Sample 2
	10	5
	10	5
	10	5
	10	5
	10	5
	10	5
	10	5
	10	5
	10	5
	10	5
	10	5
Shannon H'	2.30	2.30
Brillouin HB	2.13	2.01

INDICE DE BRILLOUIN

Lorsque l'échantillonnage aléatoire n'est pas assuré:

$$HB = \frac{\ln N! - \sum \ln n_i!}{N}$$

Idéal lorsqu'un peuplement est complètement échantillonné et que chaque individu est pris en compte

MESURES DE DOMINANCE

✗ Indice de Simpson (D)

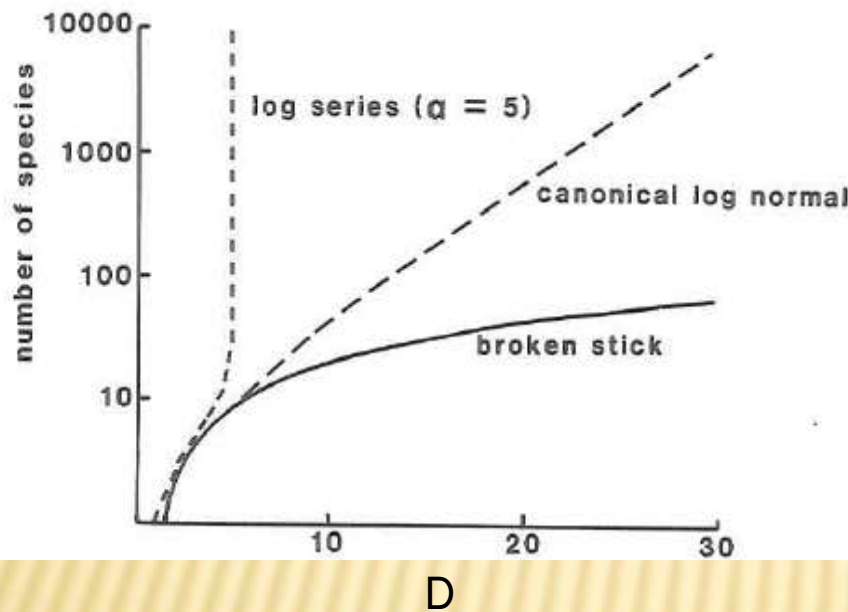
$$D = \sum p_i^2$$

$$D = \sum \left(\frac{n_i(n_i - 1)}{N(N - 1)} \right) \quad (2.27)$$

where n_i = the number of individuals in the i th species and N = the total number of individuals.

✗ Lorsque D augmente, la diversité diminue (souvent exprimé comme $1-D$ ou $1/D$)

✗ Biaisé vers les espèces abondantes et moins sensible à la richesse spécifique.

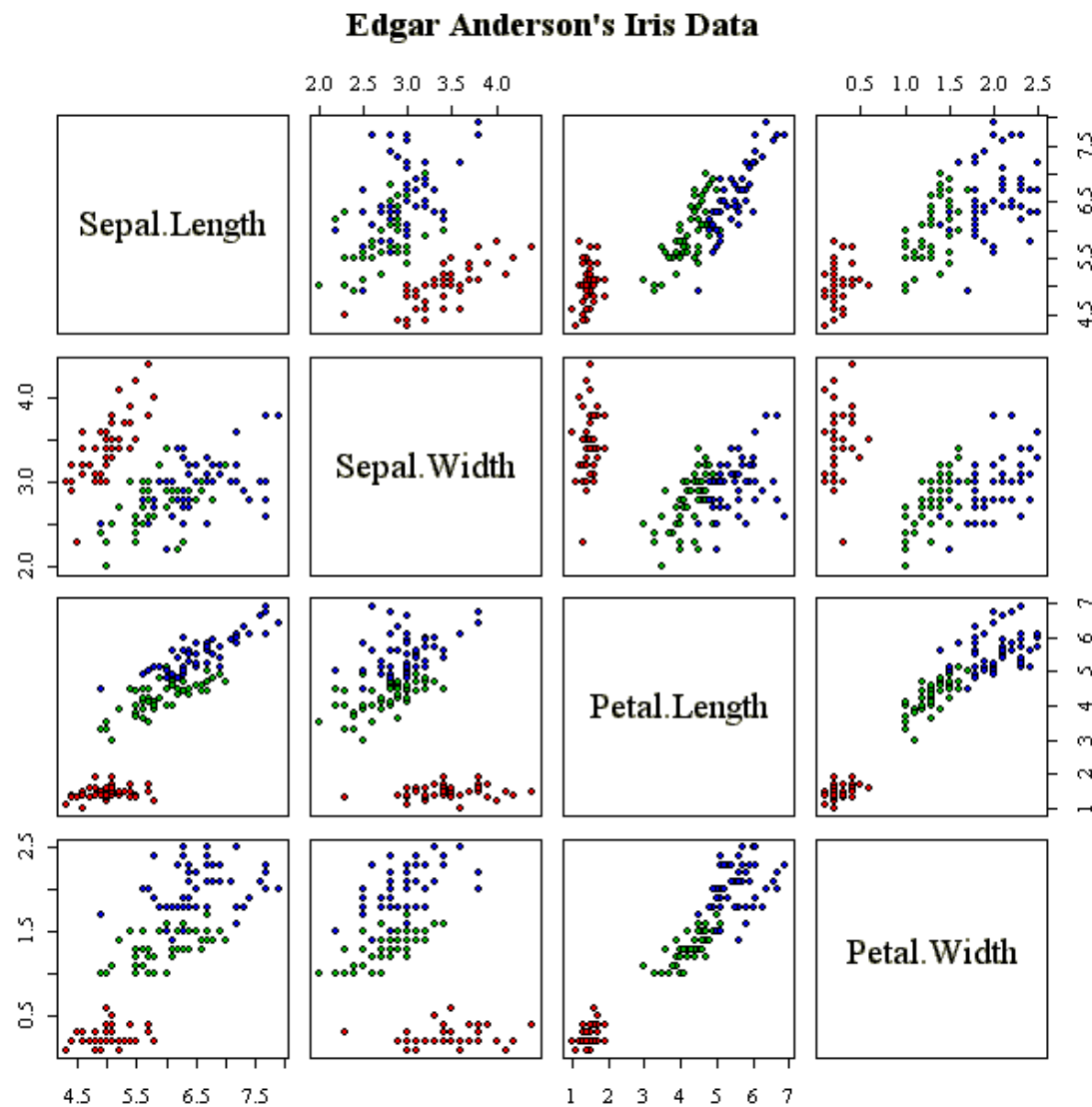


Indice de Berger-Parker

$$d = N_{\max} / N \quad (2.31)$$

where $N_{\max}$ = the number of individuals in the most abundant species

DEUXIÈME PARTIE: R



INTRODUCTION AU LOGICIEL R

The Comprehensive R Archive Network - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://streaming.stat.iastate.edu/CRAN/

Gmail: Email from Google x The Comprehensive R Archive Ne... x



CRAN

[Mirrors](#)

[What's new?](#)

[Task Views](#)

[Search](#)

About R

[R Homepage](#)

[The R Journal](#)

[Download R 2.10.1 for Windows](#) (32 megabytes)

[Installation and other instructions](#)

New features in this version: [Windows specific](#), [all platforms](#).

If you want to double-check that the package you have downloaded is correct, you will need a version of md5sum for windows: both [graphical](#) and [command line](#)

Frequently asked questions

- [How do I install R when using Windows Vista?](#)
- [How do I update packages in my previous version of R?](#)

C'EST QUOI R?

PROGRAMMING LANGUAGES

3

- Analysis/Data mining:

- > R language
- > Python
- > SQL
- > (Perl)
- > (Octave)

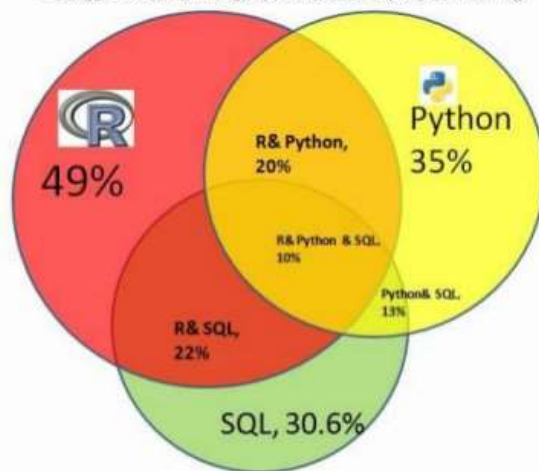
- Big Data (Hadoop)

- > Java (!)
- > Python

- Visualization

- > JavaScript

KDNuggets 2014 Poll: Languages used for Analytics/Data Mining



@dspadawan

- × Langage de programmation
- × Calcul
- × Statistiques
- × Graphes
- × Gratuit
- × Large User Group
- × **Pas facile à apprendre**
- × **Statistica, Minitab, SPSS**

HISTORIQUE

- ✖ R est basé sur le langage de programmation S, qui a été développé par **John Chambers** et d'autres au **Bell Laboratories** en **1976**.
- ✖ Au début des années **1990s**, **Ross Ihaka** et **Robert Gentleman** (University of Auckland, Nelle Zélande) ont expérimenté avec le langage et appelé le produit R.
- ✖ Depuis **1997**, R est développé par le R Development Core Team

INTRODUCTION

- ✗ $> 2 + 2$
- ✗ [1] 4
- ✗ $> \text{Log}(2)$
- ✗ [1] 0.6931472
- ✗ $> \text{Log}_{10}(2)$
- ✗ [1] 0.30103
- ✗ $> 2 + 2w$
- ✗ Error: syntax error in « $2 + 2w$ »

UN EXEMPLE DE CODE

```
set.seed(653) # Set seed in order to provide reproducibility

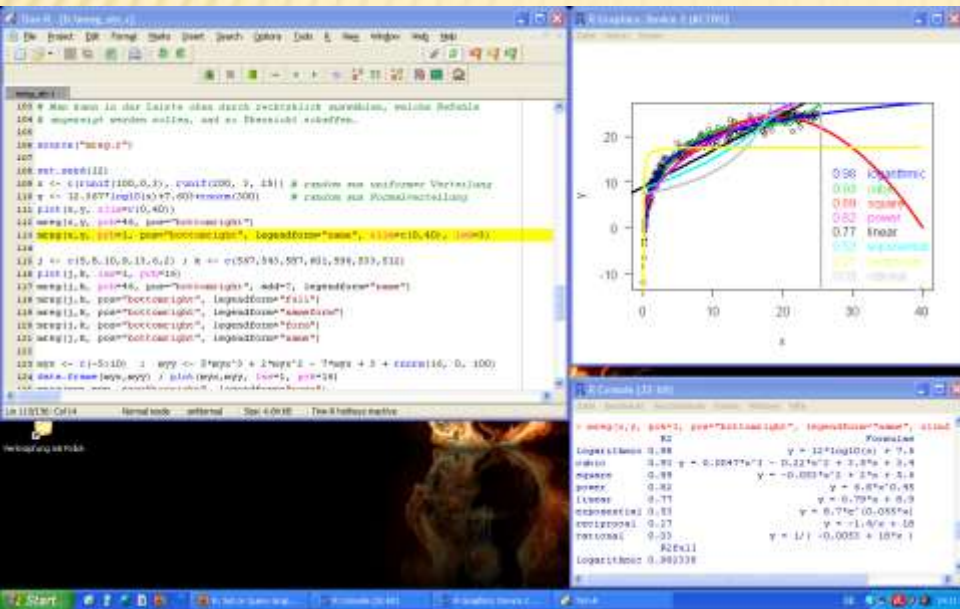
# Create example data
N <- 10000 # Sample size of 10000
y <- rnorm(N) # y without any missing values
x <- 0.5 * y + rnorm(N) # x correlated with y

# Create missings according to the MCAR response mechanism
MCAR_missings <- rbinom(N, 1, 0.25) == 1 # 25% of Y are set to miss

# Missing values according to the MAR response mechanism
x_normalized <- (x - min(x)) / (max(x) - min(x)) # Normalize x to
x_normalized <- x_normalized^2 # x_normalized to the power of 2 in
MAR_missings <- rbinom(N, 1, x_normalized) == 1 # Use x_normalized
```

EDITEURS

- ✗ Word: **déconseillé**
- ✗ R text editor
- ✗ Notepad
- ✗ **Tinn-R**
- ✗ RWindEdt



HELP/AIDE

- × **> ?boxplot**
- × **> help (boxplot)**
- × Sinon:
- × click Help -> Html help
- × Utilisez le Search Engine & Keywords
- × Google
- × R newsgroups
- × www.r-project.org

PACKAGES/LIBRAIRIES

- ✗ Une librairie est une collection de fonctions pour effectuer certaines tâches.
- ✗ *Pour voir si vous avez un package:*
- ✗ Packages -> Load package; sélectionnez un package et cliquez OK
- ✗ Certains Packages sont inclus dans l'installation de base
- ✗ > library(MASS)
- ✗ *Si le package n'est pas installé:*
- ✗ > Packages -> set the CRAN mirror and select a server puis
- ✗ Packages click Install package

DOSSIER DE TRAVAIL

- ✗ > `setwd(file = "C:\\Samraoui\\")`
- ✗ Il faut éviter les symboles dans les noms: &, *
- ✗ Ou
- ✗ > `setwd(file = "C:/Samraoui/")`
- ✗ R n'accepte pas ö, é, û, etc.
- ✗ *Pour connaître le dossier courant*
- ✗ `getwd()`

AJUSTER LA CONSOLE ET QUITTER R

- ✖ *Pour ajuster la police/couleur de fond:*
- ✖ > Edit -< GUI preferences
- ✖ **Quitter R**
- ✖ q()
- ✖ Or q(save = « no »)
- ✖ *La sauvegarde dans le workspace:*
- ✖ File -< Save Workspace
- ✖ > save.image("C:\\Users\\Sam\\Documents\\Test2019")
- ✖ File -< Load Workspace

ENLEVER LES VARIABLES

- ✗ *Pour refaire une analyse avec d'autres données: 3 options*
- ✗ 1) Quitter R et le redémarrer:
- ✗ 2) Misc-< Remove all objects
- ✗ `> rm(list=ls(all=TRUE))`
- ✗ *Autres options utiles:*
- ✗ Sous l'onglet Edit, vous pouvez choisir « Select all »

UN EXEMPLE DE « EDIT ET SELECT ALL »

```
x R version 3.6.0 (2019-04-26) -- "Planting of a Tree"
x Copyright (C) 2019 The R Foundation for Statistical Computing
x Platform: x86_64-w64-mingw32/x64 (64-bit)

x R is free software and comes with ABSOLUTELY NO WARRANTY.
x You are welcome to redistribute it under certain conditions.
x Type 'license()' or 'licence()' for distribution details.

x   Natural language support but running in an English locale

x R is a collaborative project with many contributors.
x Type 'contributors()' for more information and
x 'citation()' on how to cite R or R packages in publications.

x Type 'demo()' for some demos, 'help()' for on-line help, or
x 'help.start()' for an HTML browser interface to help.
x Type 'q()' to quit R.

x Registered S3 methods overwritten by 'ggplot2':
x   method      from
x   [.quosures  rlang
x   c.quosures  rlang
x   print.quosures rlang
x [Previously saved workspace restored]

x > log(10)
x [1] 2.302585
x > log10(10)
x [1] 1
x > save.image("C:\\Users\\Sam\\Documents\\Test2019")
x > rm(list=ls(all=TRUE))
x >
```

ENTRÉE D'UN JEU DE DONNÉES RÉDUIT

- ✖ Wing1 <- 59
- ✖ > Wing2 <- 55
- ✖ > Wing3 <- 53.5
- ✖ > Wing4 <- 55
- ✖ > Wing5 <- 52.5
- ✖ >
- ✖ > sqrt(Wing1)
- ✖ [1] 7.681146
- ✖ > 2*Wing1
- ✖ [1] 118
- ✖ > (Wing1+Wing2+Wing3+Wing4+Wing5)/5
- ✖ [1] 55

- ✖ *Vous devez sauvegarder l'information si vous voulez réutiliser une variable*
- ✖ > SQ.wing1 <- sqrt(Wing1)
- ✖ > SQ.wing1
- ✖ [1] 7.681146
- ✖ > Mul.W1 <- 2*Wing1
- ✖ > Mul.W1
- ✖ [1] 118
- ✖ > Av <- (Wing1+Wing2+Wing3+Wing4+Wing5)/5
- ✖ > Av
- ✖ [1] 55

THE C FUNCTION (CONCATENATING)

- ✖ *Quatre variables morphométriques (8 observations) d'une espèce d'oiseau*
- ✖ `Wingcrd <- c(59, 55, 53.5, 55, 52.5, 57.5, 53, 55)`
- ✖ `Wingcrd[1]`
- ✖ `Wingcrd[1:5]`
- ✖ `Wingcrd[-2]`

- ✖ `S.win <- sum(Wingcrd)`
- ✖ `S.win`
- ✖ **440.5**

THE C FUNCTION (CONCATENATING): SUITE

- ✗ Tarsus <- c(22.3, 19.7, 20.8, 20.3, 20.8, 21.5, 20.6, 21.5)
- ✗ Head <- c(31.2, 30.4, 30.6, 30.3, 30.3, 30.8, 32.5, **NA**)
- ✗ Wt <- c(9.5, 13.8, 14.8, 15.2, 15.5, 15.6, 15.6, 15.7)

- ✗ **sum(Head, na.rm = TRUE)**
- ✗ **216.1**
- ✗ *Essayez: sum(Head)*

COMBINAISON DE VARIABLES

- ✖ `> BirdData <- c(Wingcrd, Tarsus, Head, Wt)`
- ✖ `> BirdData`
- ✖ Vecteur de 32 observations/mesures

- ✖ `Id <- c(1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2,`
- ✖ `2, 3, 3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 4, 4)`

- ✖ `> rep(c(1, 2, 3, 4), each = 8)`
- ✖ `[1] 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4`

- ✖ `> rep(1 : 4, each = 8)`
- ✖ `[1] 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4`

- ✖ `> a <- seq(from = 1, to = 4, by = 1)`
- ✖ `> a`
- ✖ `[1] 1 2 3 4`
- ✖ `> a <- seq(from = 1, to = 4, by = 1)`
- ✖ `> rep(a, each = 8)`
- ✖ `[1] 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4`

- ✖ Pourquoi toutes ces complications? Cela dépend des fonctions dans R

COMBINAISON DE VARIABLES (SUITE)

```
x > VarNames <- c("Wingcrd", "Tarsus", "Head", "Wt")
x > Id2 <- rep(VarNames, each = 8)
x > > Id2
x [1] "Wingcrd" "Wingcrd" "Wingcrd" "Wingcrd" "Wingcrd" "Wingcrd" "Wingcrd"
x [8] "Wingcrd" "Tarsus" "Tarsus" "Tarsus" "Tarsus" "Tarsus" "Tarsus"
x [15] "Tarsus" "Tarsus" "Head" "Head" "Head" "Head" "Head"
x [22] "Head" "Head" "Head" "Wt" "Wt" "Wt" "Wt"
x [29] "Wt" "Wt" "Wt" "Wt"
x >
x > Z <- cbind(Wingcrd, Tarsus, Head, Wt)
x > Z
x      Wingcrd Tarsus Head  Wt
x [1,]  59.0   22.3 31.2  9.5
x [2,]  55.0   19.7 30.4 13.8
x [3,]  53.5   20.8 30.6 14.8
x [4,]  55.0   20.3 30.3 15.2
x [5,]  52.5   20.8 30.3 15.5
x [6,]  57.5   21.5 30.8 15.6
x [7,]  53.0   20.6 32.5 15.6
x [8,]  55.0   21.5  NA 15.7
```

COMBINAISON DE VARIABLES (SUITE)

```
× > n <- dim(Z)
× > n
× [1] 8  4
× > nrow <- dim(Z)[1]
× > nrow
× [1] 8

× > Z2 <- rbind(Wingcrd, Tarsus, Head, Wt)
× > Z2
×      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
× Wingcrd 59.0 55.0 53.5 55.0 52.5 57.5 53.0 55.0
× Tarsus   22.3 19.7 20.8 20.3 20.8 21.5 20.6 21.5
× Head     31.2 30.4 30.6 30.3 30.3 30.8 32.5  NA
× Wt       9.5 13.8 14.8 15.2 15.5 15.6 15.6 15.7
```

VECTEURS ET MATRICES

```
x > W <- vector(length = 8)
x > W[1] <- 59
x > W[2] <- 55
x > W[3] <- 53.5
x > W[4] <- 55
x > W[5] <- 52.5
x > W[6] <- 57.5
x > W[7] <- 53
x > W[8] <- 55
x >
x >
x > Dmat <- matrix(nrow = 8, ncol = 4)
x > Dmat
x   [,1] [,2] [,3] [,4]
x [1,] NA  NA  NA  NA
x [2,] NA  NA  NA  NA
x [3,] NA  NA  NA  NA
x [4,] NA  NA  NA  NA
x [5,] NA  NA  NA  NA
x [6,] NA  NA  NA  NA
x [7,] NA  NA  NA  NA
x [8,] NA  NA  NA  NA
x >
```


VECTEURS ET MATRICES (SUITE)

```
x > Dmat[, 1] <- c(59, 55, 53.5, 55, 52.5, 57.5, 53, 55)
x > Dmat[, 2] <- c(22.3, 19.7, 20.8, 20.3, 20.8, 21.5,
x +      20.6, 21.5)
x > Dmat[, 3] <- c(31.2, 30.4, 30.6, 30.3, 30.3, 30.8,
x +      32.5, NA)
x > Dmat[, 4] <- c(9.5, 13.8, 14.8, 15.2, 15.5, 15.6,
x +      15.6, 15.7)
x >
x > Dmat
x   [,1] [,2] [,3] [,4]
x [1,] 59.0 22.3 31.2  9.5
x [2,] 55.0 19.7 30.4 13.8
x [3,] 53.5 20.8 30.6 14.8
x [4,] 55.0 20.3 30.3 15.2
x [5,] 52.5 20.8 30.3 15.5
x [6,] 57.5 21.5 30.8 15.6
x [7,] 53.0 20.6 32.5 15.6
x [8,] 55.0 21.5  NA 15.7
```

VECTEURS ET MATRICES (SUITE)

✗ `> colnames(Dmat) <- c("Wingcrd", "Tarsus", "Head", "Wt")`

✗ `> Dmat`

✗ Wingcrd Tarsus Head Wt

✗ [1,] 59.0 22.3 31.2 9.5

✗ [2,] 55.0 19.7 30.4 13.8

✗ [3,] 53.5 20.8 30.6 14.8

✗ [4,] 55.0 20.3 30.3 15.2

✗ [5,] 52.5 20.8 30.3 15.5

✗ [6,] 57.5 21.5 30.8 15.6

✗ [7,] 53.0 20.6 32.5 15.6

✗ [8,] 55.0 21.5 NA 15.7

VECTEURS ET MATRICES (SUITE)

- ✖ *Si les données existent sous formes de variables (vecteurs: Wingcrd, Tarsus, etc.):*
- ✖ `> Dmat2 <- as.matrix(cbind(Wingcrd,Tarsus,Head, Wt))`
- ✖ `> Dmat2`
- ✖

	Wingcrd	Tarsus	Head	Wt
[1,]	59.0	22.3	31.2	9.5
[2,]	55.0	19.7	30.4	13.8
[3,]	53.5	20.8	30.6	14.8
[4,]	55.0	20.3	30.3	15.2
[5,]	52.5	20.8	30.3	15.5
[6,]	57.5	21.5	30.8	15.6
[7,]	53.0	20.6	32.5	15.6
[8,]	55.0	21.5	NA	15.7

NOTION DE DATAFRAME

- ✗ *Jusqu'à présent nous avons utilisé c, cbind, rbind, vector, et matrix pour combiner les données. Une autre option est data.frame*
- ✗ `> Dfrm <- data.frame(WC = Wingcrd, TS = Tarsus,`
- ✗ `+ HD = Head, W = Wt)`
- ✗ `> Dfrm`
- ✗

	WC	TS	HD	W
1	59.0	22.3	31.2	9.5
2	55.0	19.7	30.4	13.8
3	53.5	20.8	30.6	14.8
4	55.0	20.3	30.3	15.2
5	52.5	20.8	30.3	15.5
6	57.5	21.5	30.8	15.6
7	53.0	20.6	32.5	15.6
8	55.0	21.5	NA	15.7
- ✗ 1 59.0 22.3 31.2 9.5
- ✗ 2 55.0 19.7 30.4 13.8
- ✗ 3 53.5 20.8 30.6 14.8
- ✗ 4 55.0 20.3 30.3 15.2
- ✗ 5 52.5 20.8 30.3 15.5
- ✗ 6 57.5 21.5 30.8 15.6
- ✗ 7 53.0 20.6 32.5 15.6
- ✗ 8 55.0 21.5 NA 15.7

L'AVANTAGE DES DATAFRAMES

- ✗ Vous pouvez changer les données et combiner les variables numériques, characters et facteurs dans un dataframe:
- ✗ `> Dfrm2 <- data.frame(WC = Wingcrd, TS = Tarsus,`
- ✗ `+ HD = Head, W = Wt, Wsq = sqrt(Wt))`
- ✗ `> Dfrm2`
- ✗

	WC	TS	HD	W	Wsq
1	59.0	22.3	31.2	9.5	3.082207
2	55.0	19.7	30.4	13.8	3.714835
3	53.5	20.8	30.6	14.8	3.847077
4	55.0	20.3	30.3	15.2	3.898718
5	52.5	20.8	30.3	15.5	3.937004
6	57.5	21.5	30.8	15.6	3.949684
7	53.0	20.6	32.5	15.6	3.949684
8	55.0	21.5	NA	15.7	3.962323

- ✗
- ✗ Les objets `Wt` et `Dfrm$W` sont distincts!
- ✗ Notez que ni `cbind` ni `matrix` ne peuvent combiner différents types de données.

LISTES

```
x > AllData <- list(BirdData, Id, Z, VarNames)
x > AllData
x [[1]]
x [1] 59.0 55.0 53.5 55.0 52.5 57.5 53.0 55.0 22.3 19.7 20.8 20.3 20.8 21.5 20.6
x [16] 21.5 31.2 30.4 30.6 30.3 30.3 30.8 32.5 NA 9.5 13.8 14.8 15.2 15.5 15.6
x [31] 15.6 15.7

x [[2]]
x [1] "Wingcrd" "Wingcrd" "Wingcrd" "Wingcrd" "Wingcrd" "Wingcrd" "Wingcrd"
x [8] "Wingcrd" "Tarsus" "Tarsus" "Tarsus" "Tarsus" "Tarsus" "Tarsus"
x [15] "Tarsus" "Tarsus" "Head" "Head" "Head" "Head" "Head"
x [22] "Head" "Head" "Head" "Wt" "Wt" "Wt" "Wt"
x [29] "Wt" "Wt" "Wt" "Wt"

x [[3]]
x Wingcrd Tarsus Head Wt
x [1,] 59.0 22.3 31.2 9.5
x [2,] 55.0 19.7 30.4 13.8
x [3,] 53.5 20.8 30.6 14.8
x [4,] 55.0 20.3 30.3 15.2
x [5,] 52.5 20.8 30.3 15.5
x [6,] 57.5 21.5 30.8 15.6
x [7,] 53.0 20.6 32.5 15.6
x [8,] 55.0 21.5 NA 15.7

x [[4]]
x [1] "Wingcrd" "Tarsus" "Head" "Wt"
```


QUEL INTÉRÊT DES LISTES?

- ✗ La plupart des fonctions dans R (modélisations, tests statistiques, etc.) sont produites sous forme de listes.

- ✗ `> AllData[3]`

- ✗ `[[1]]`

- ✗

	Wingcrd	Tarsus	Head	Wt
--	---------	--------	------	----

- ✗

[1,]	59.0	22.3	31.2	9.5
------	------	------	------	-----

- ✗

[2,]	55.0	19.7	30.4	13.8
------	------	------	------	------

- ✗

[3,]	53.5	20.8	30.6	14.8
------	------	------	------	------

- ✗

[4,]	55.0	20.3	30.3	15.2
------	------	------	------	------

- ✗

[5,]	52.5	20.8	30.3	15.5
------	------	------	------	------

- ✗

[6,]	57.5	21.5	30.8	15.6
------	------	------	------	------

- ✗

[7,]	53.0	20.6	32.5	15.6
------	------	------	------	------

- ✗

[8,]	55.0	21.5	NA	15.7
------	------	------	----	------

IMPORTATION DES DONNÉES

- ✖ `setwd("C:\\Samraoui\\")`
- ✖ `Chard <- read.table(file = "Chardon.txt", header = TRUE)`
- ✖ `Chard2 <- read.table(file =
« Chardon2.txt", header = FALSE, dec = " . ")`
- ✖ `Chard3 <- read.table(file =
« Chardon2.txt", header = FALSE, dec = " , ")`
- ✖ Il existe également `read.csv` pour ce type de fichiers

MERGEZ ET EXPORTEZ DES DONNÉES

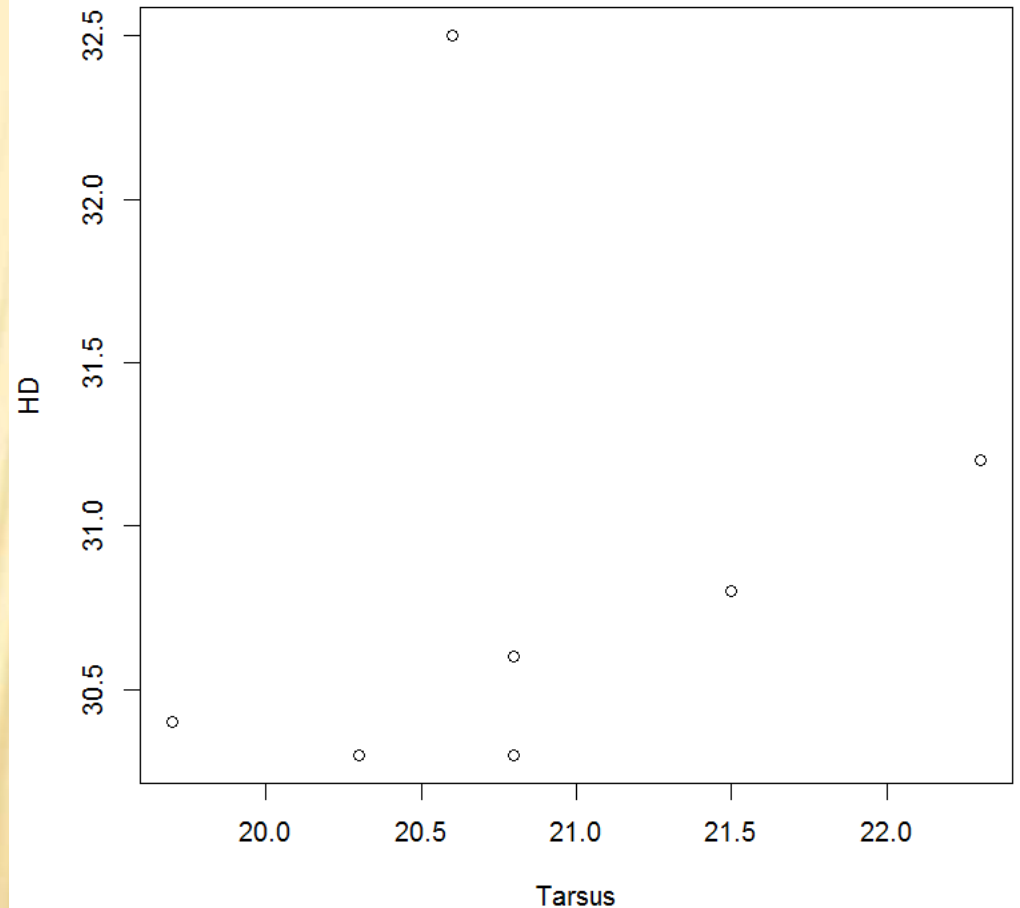
- ✗ `>Sq1<-read.table(file="C:/Biologie/squid1.txt",header=T)`
- ✗ `>Sq2<-read.table(file="C:/Biologie/squid2.txt",header=T)`
- ✗ `>ZMerged<-merge(Sq1,Sq2,by="Sample")`
- ✗ `>write.table(ZMerged, file = "C:/Biologie/MergedSquid.txt",`
- ✗ `sep = " ",quote=FALSE,append=FALSE,`
- ✗ `na="NA")`

LA PROCHAINE ÉTAPE...

- × > attach(Dfrm2)
- × > hist(HD)
- × Équivalent à hist(Dfrm2\$HD)
- × > boxplot(WC)
- × > sum(Wt)
- × [1] 115.7
- × > mean(Wt)
- × [1] 14.4625

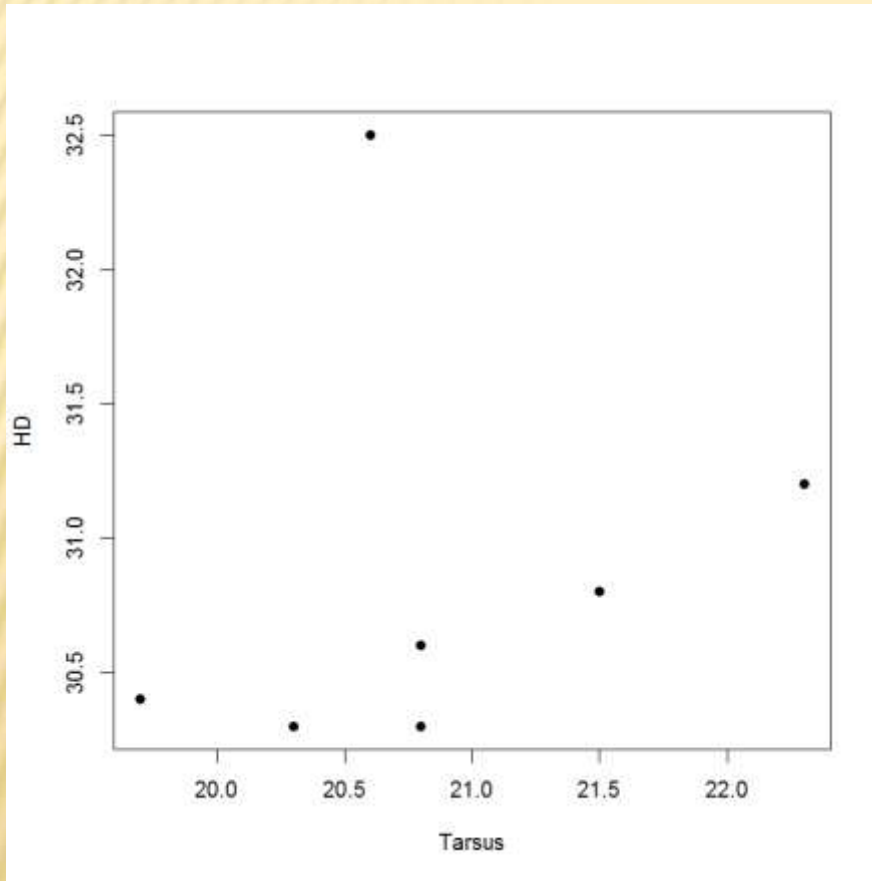
PLOT 1

- ✗ `plot(HD,Wt)`
- ✗ `plot(Tarsus,HD)`

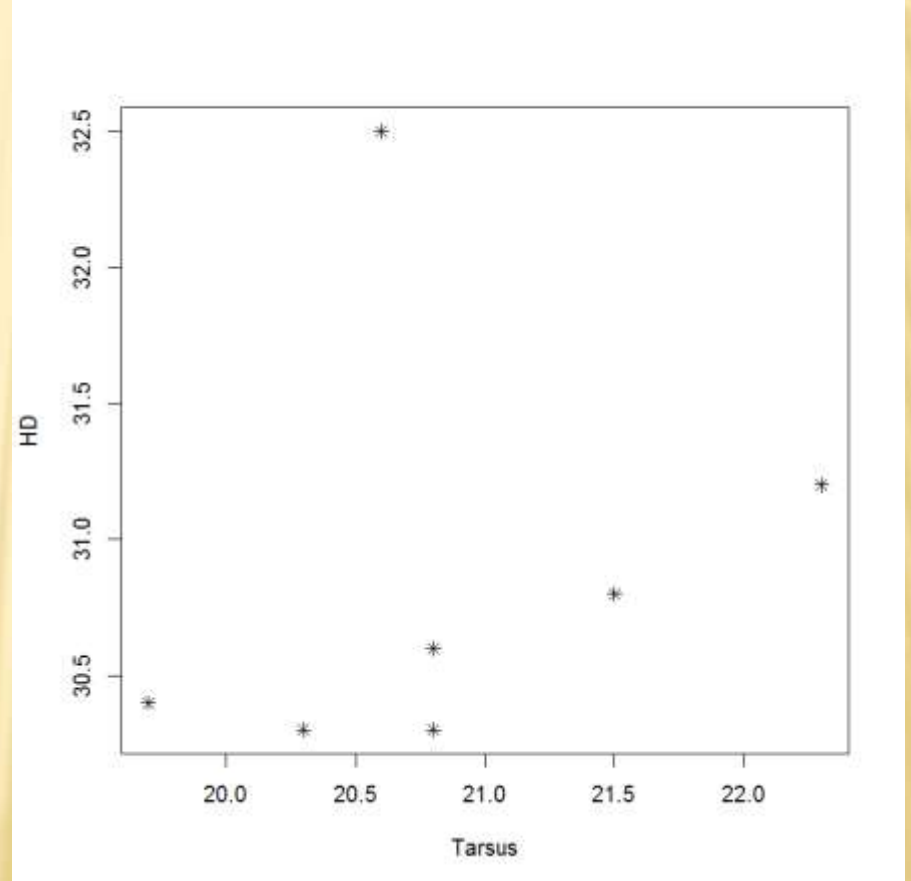


CHANGEMENT DES SYMBOLES DES GRAPHES

```
>plot(Tarsus,HD, pch = 16)
```

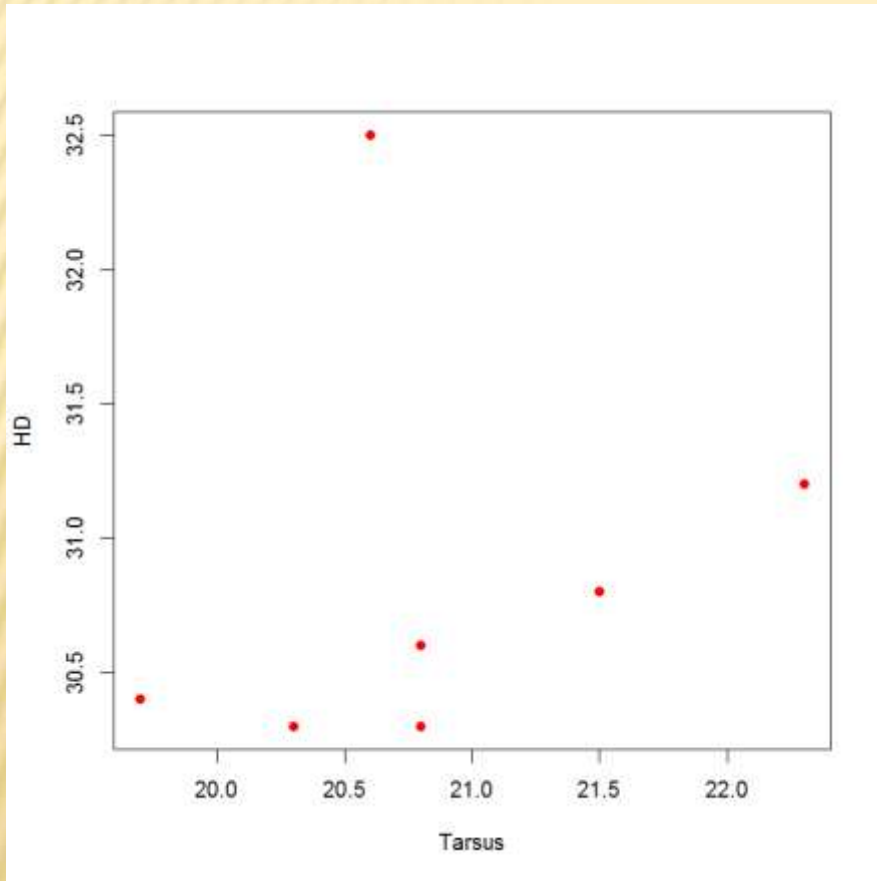


```
> plot(Tarsus,HD, pch = 8)
```

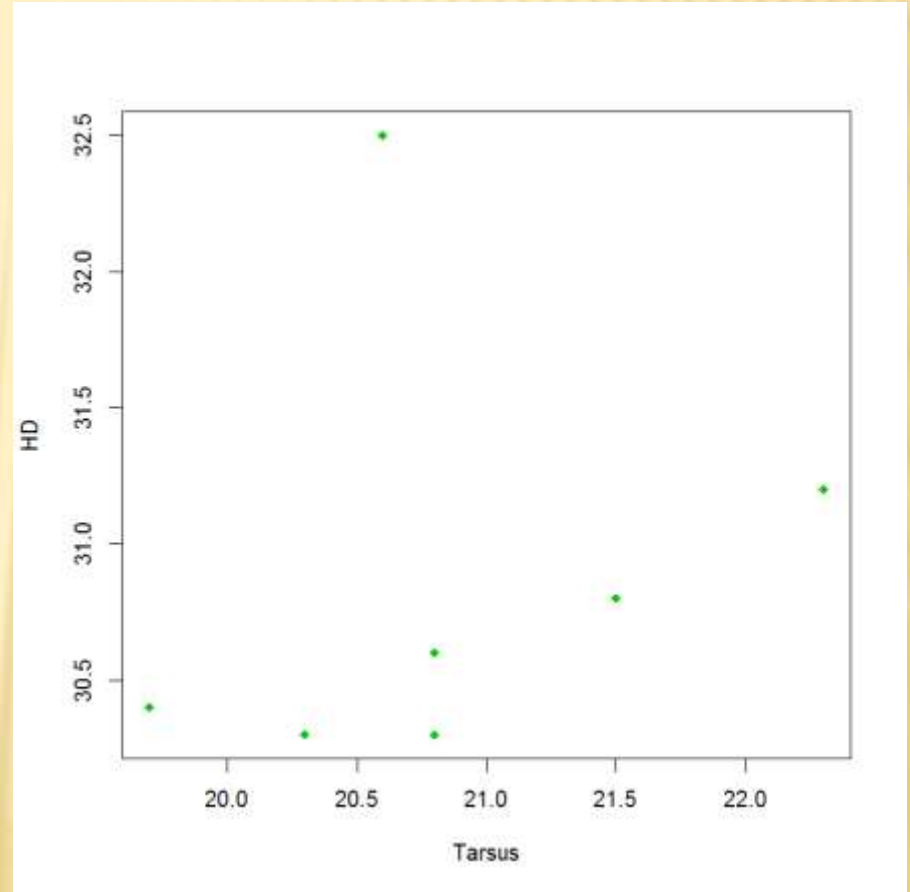


CHANGEMENT DES COULEURS DES GRAPHES

```
> plot(Tarsus,HD, col = 2, pch = 16)
```

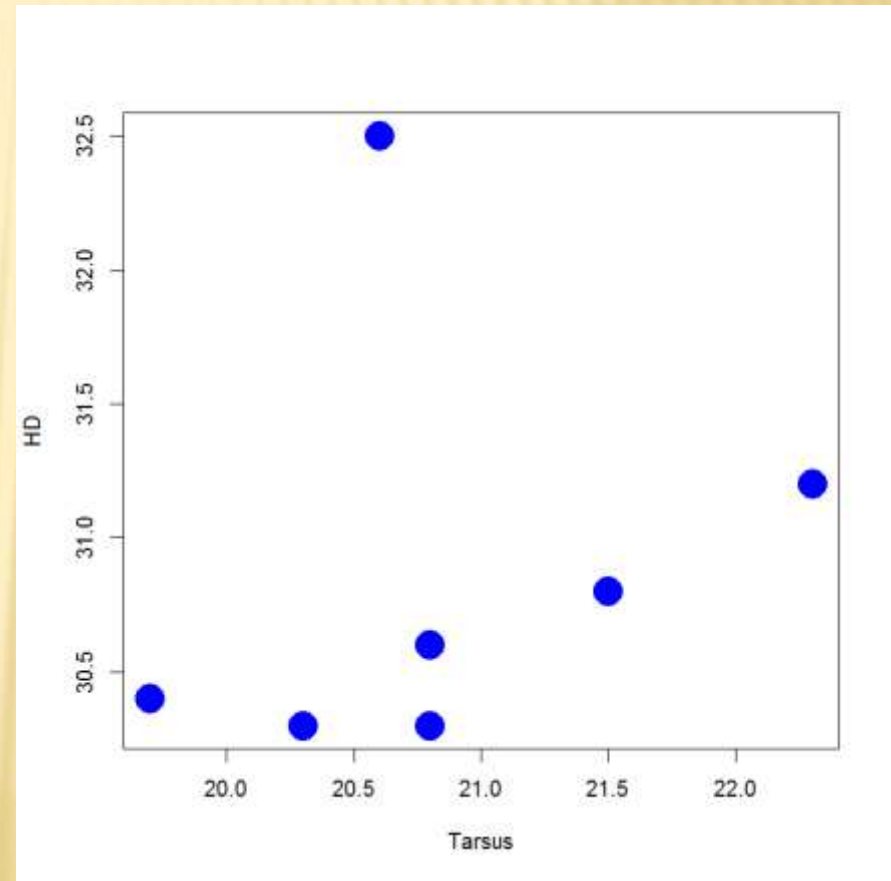
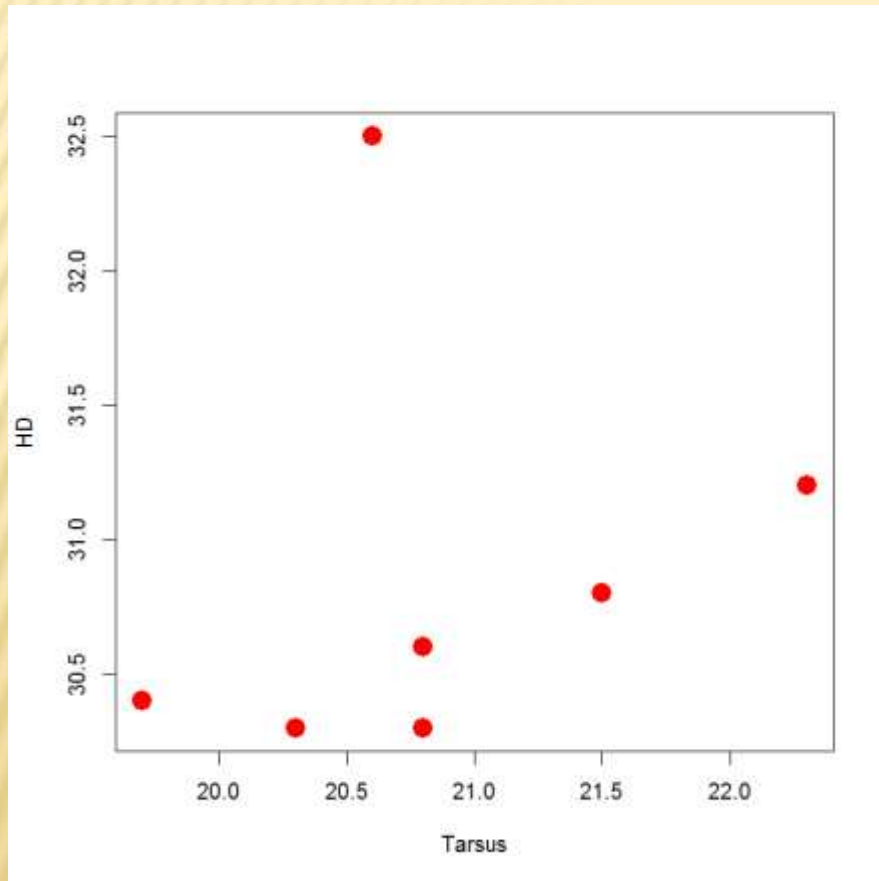


```
> plot(Tarsus,HD, col = 3, pch = 18)
```



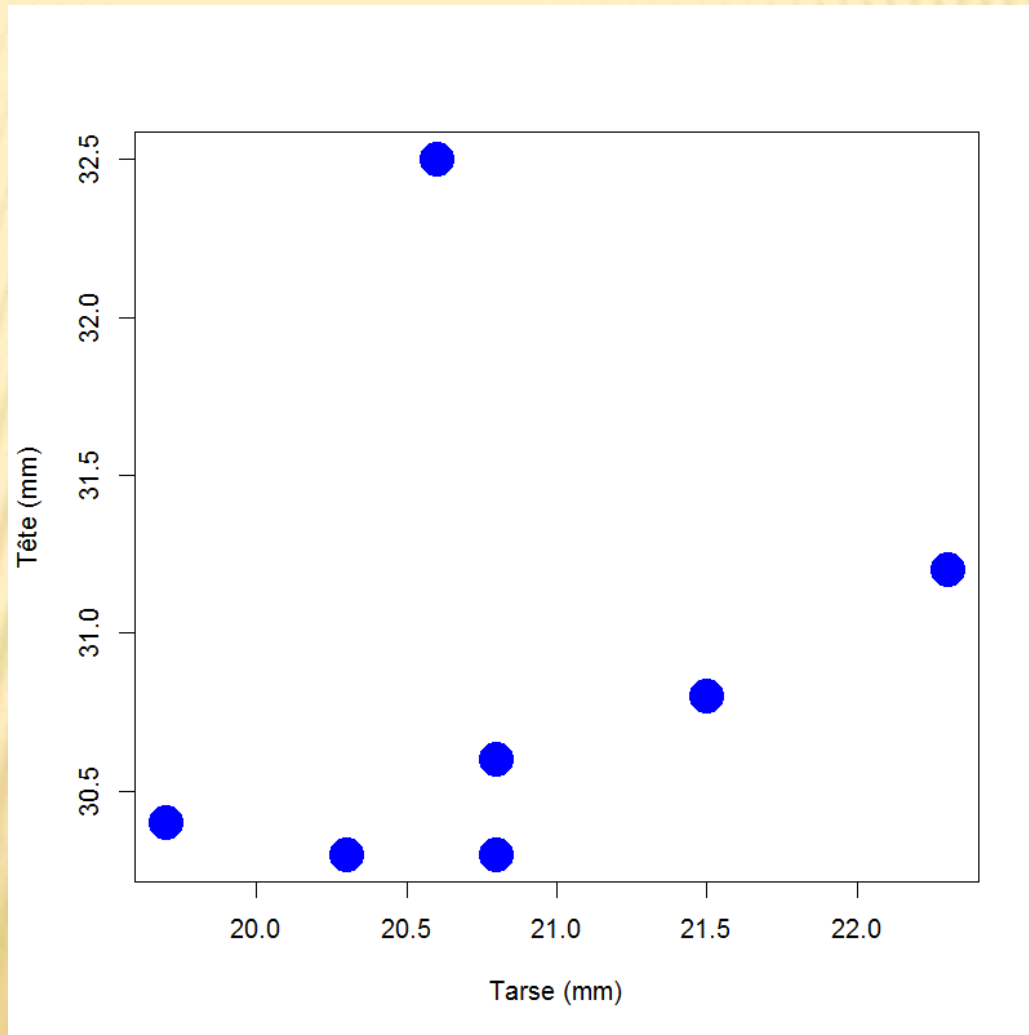
CHANGEMENT DE LA TAILLE DES SYMBOLES

> plot(Tarsus,HD, col = 2, pch = 19, cex = 2) > plot(Tarsus,HD, col = 4, pch = 19, cex = 3)



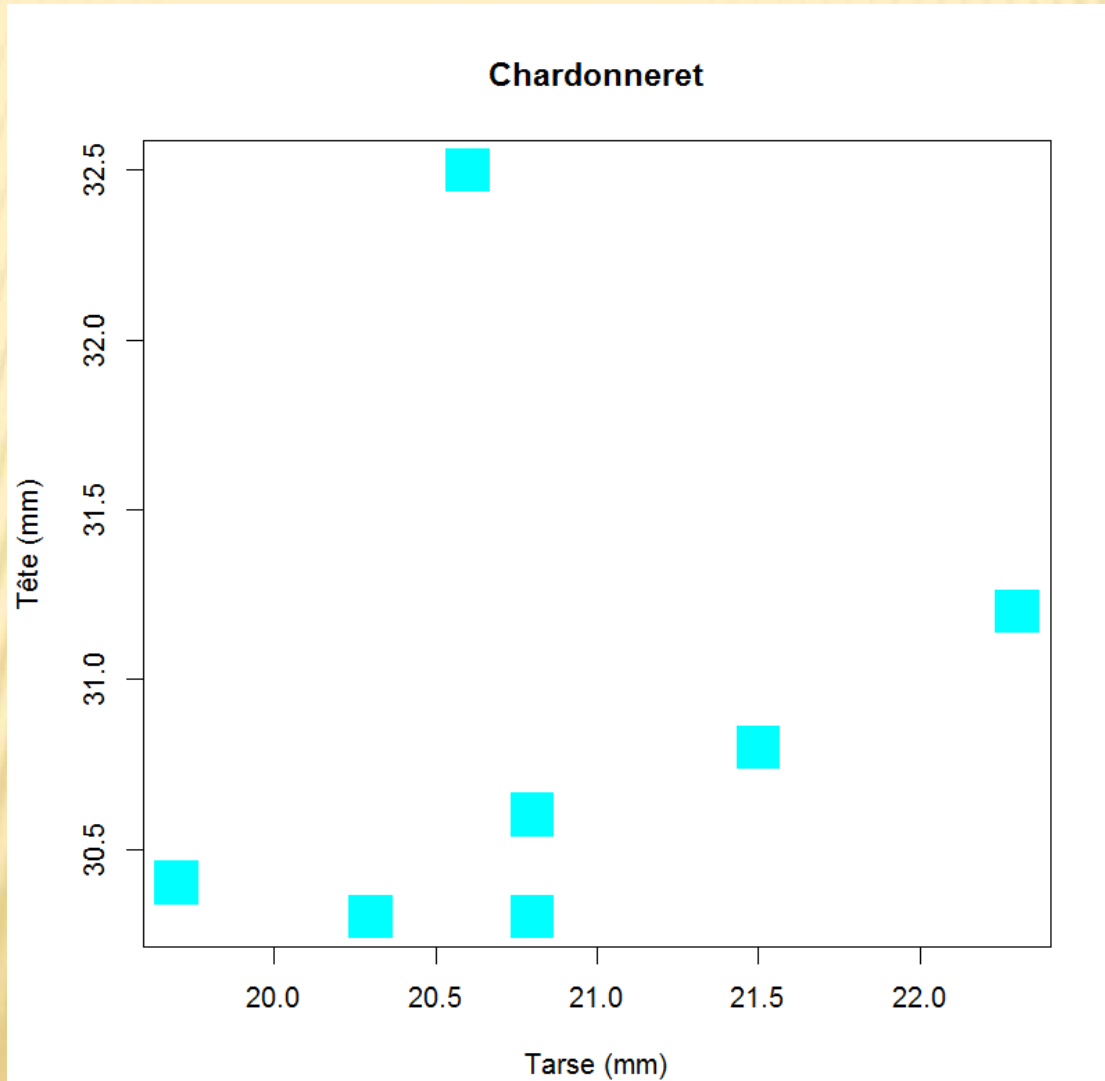
LÉGENDES DES AXES

```
> plot(Tarsus,HD, col = 4, pch = 19, cex = 3, xlab = "Tarse (mm)", ylab = "Tête (mm)")
```



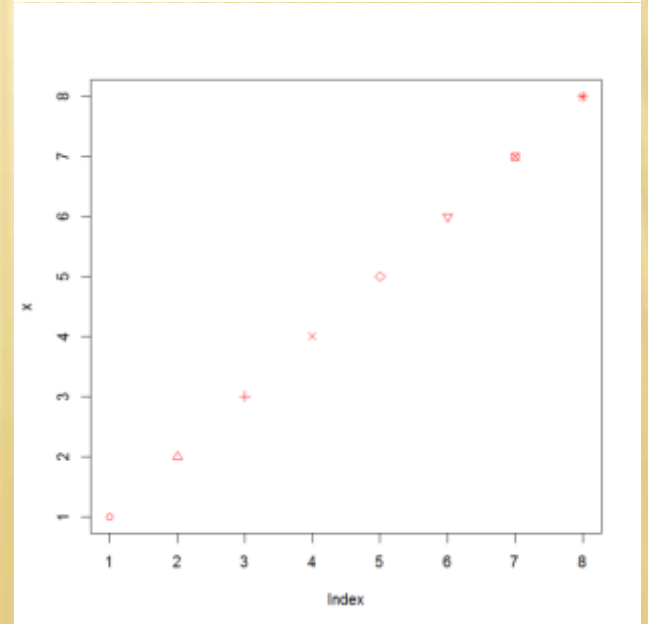
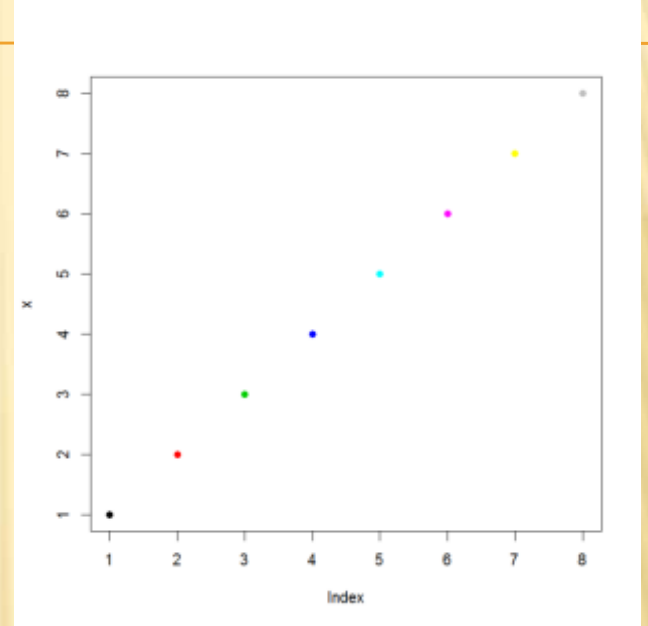
TITRE DU GRAPHE

```
> plot(Tarsus,HD, col = 5, pch = 15, cex = 3.5, main = "Chardonneret", xlab = "Tarse (mm)", ylab = "Tête (mm)")
```

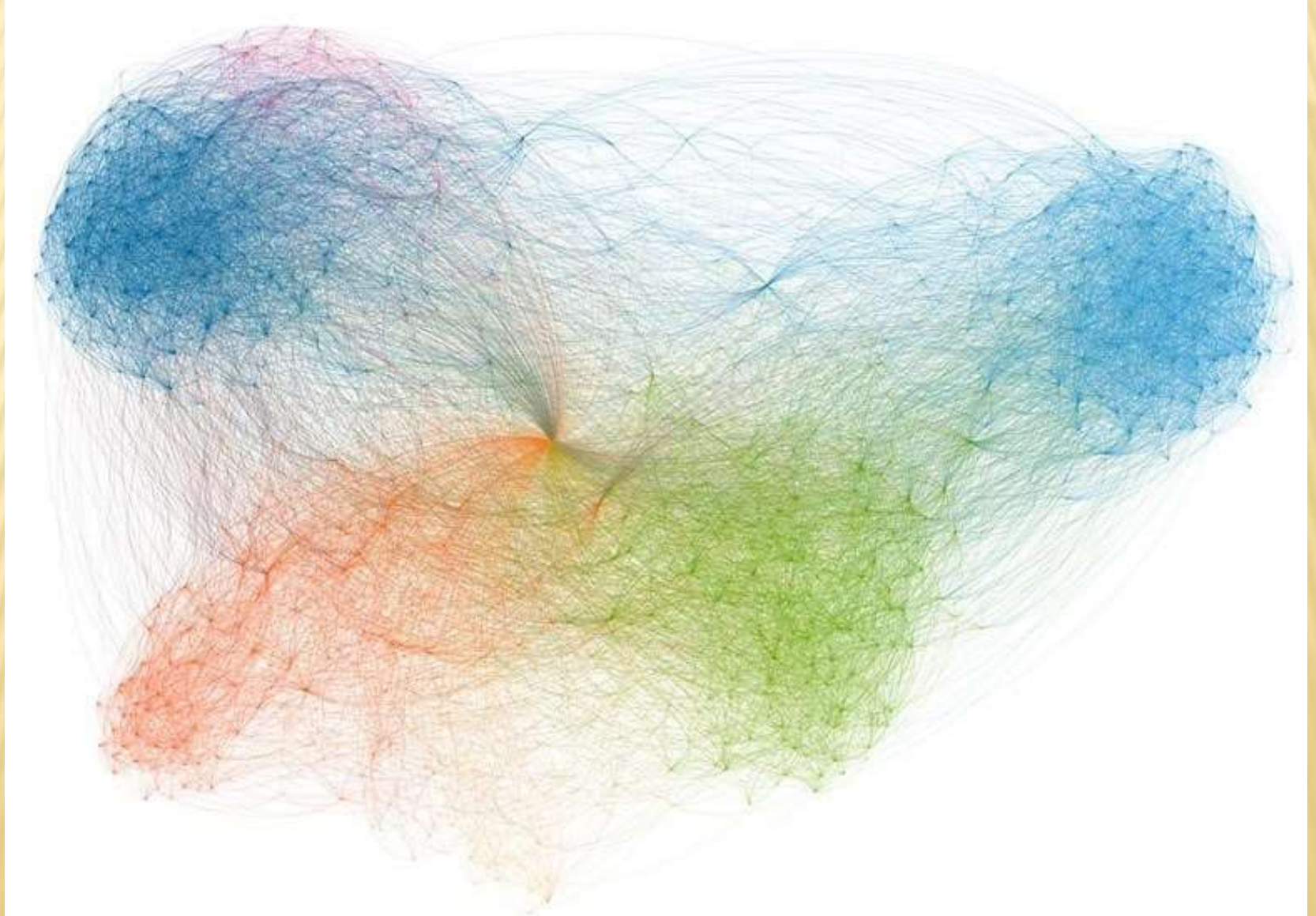


EN CAS D'OUBLI...

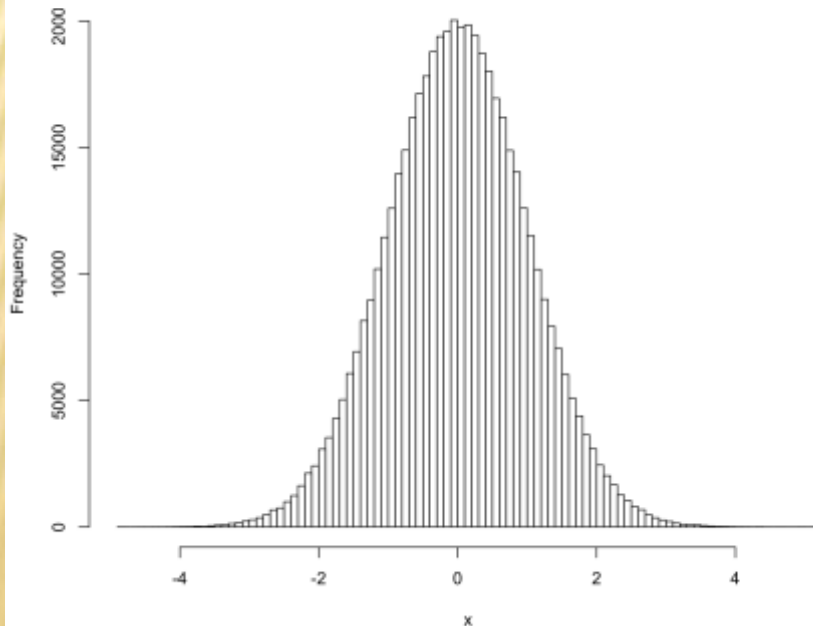
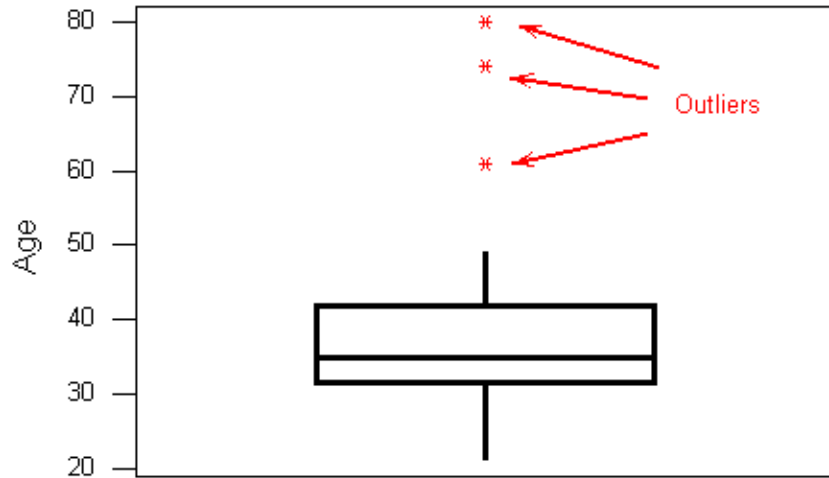
- ✖ `x <- 1:8`
- ✖ `> plot(x, col = x)`
- ✖ `> plot(x, col = x, pch = 19)`
- ✖ `> plot(x, col = 2, pch = x)`



TROISIÈME PARTIE: EXPLORATION DES DONNÉES ET MODÉLISATION



EXPLORATION DES DONNÉES: PLAN



- ✗ Outliers
- ✗ Normalité
- ✗ Homogénéité
- ✗ Collinéarité