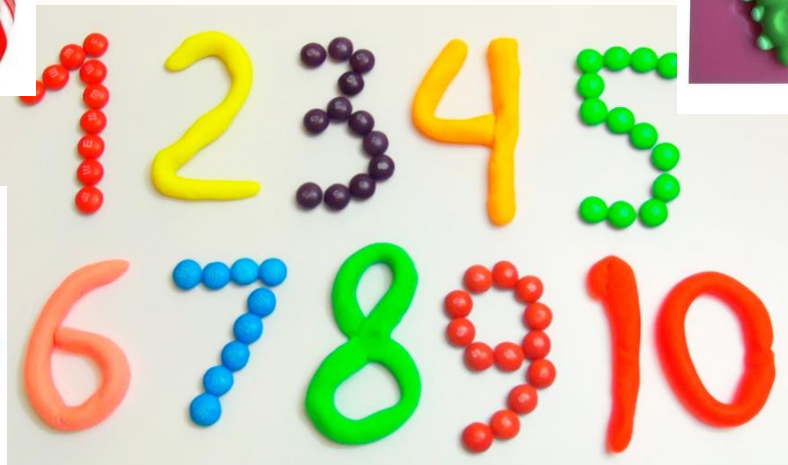


Goodies*



* Goodies related to animals, plants and numbers...

The cheat sheets below make it easy to learn about and use some of our favorite packages. From time to time, we will add new cheat sheets to the gallery. If you'd like us to drop you an email when we do, let us know by clicking the button to the right.

SUBSCRIBE TO CHEAT SHEET UPDATES HERE

- RStudio IDE
- Data Import
- Deep Learning with Keras
- R Markdown
- Data Transformation with dplyr
- Data Science in Spark with Sparklyr
- Shiny
- Data Visualization with ggplot2
- String manipulation with stringr
- Package Development
- Apply functions with purrr
- Dates and times with lubridate

Gestão de Páginas

- ▼ Ecologia Numérica
 - Ecologia Numérica(Tecnologias de Info
 - Teóricas
 - ▼ Práticas
 - ▶ Week1
 - PDFs
 - ▼ Outros Recursos
 - R Cheat Sheets

+ Criar

R Cheat Sheets

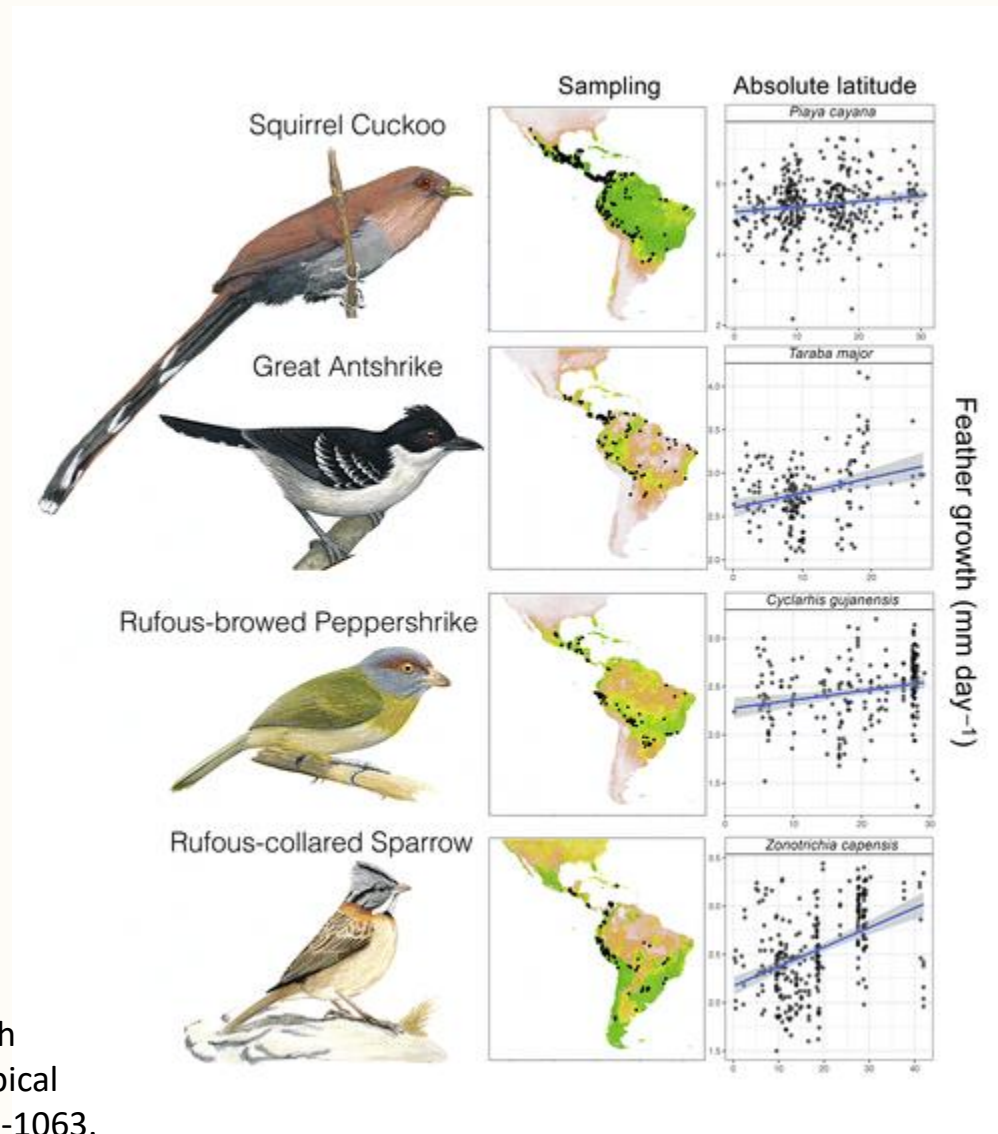
Página **Ficheiros 6** Permissões Link

Adicionar Ficheiro

#	Nome	Permissões	Operações
1	advancedR.pdf	Público	
2	base-r.pdf	Público	
3	cartography.pdf	Público	
4	data-import.pdf	Público	
5	rmarkdown-2.0 (1).pdf	Público	
6	rstudio-ide.pdf	Público	

September Content Alert for Auk and Condor - Fri 9/28/2018 3:01 AM

Uma imagem vale mais que mil palavras... Um bom sumário gráfico pode ser a forma mais eficaz de divulgar um trabalho!



Ryan S. Terrill (2018) Feather growth rate increases with latitude in four species of widespread resident Neotropical birds. *The Auk*: October 2018, Vol. 135, No. 4, pp. 1055-1063.

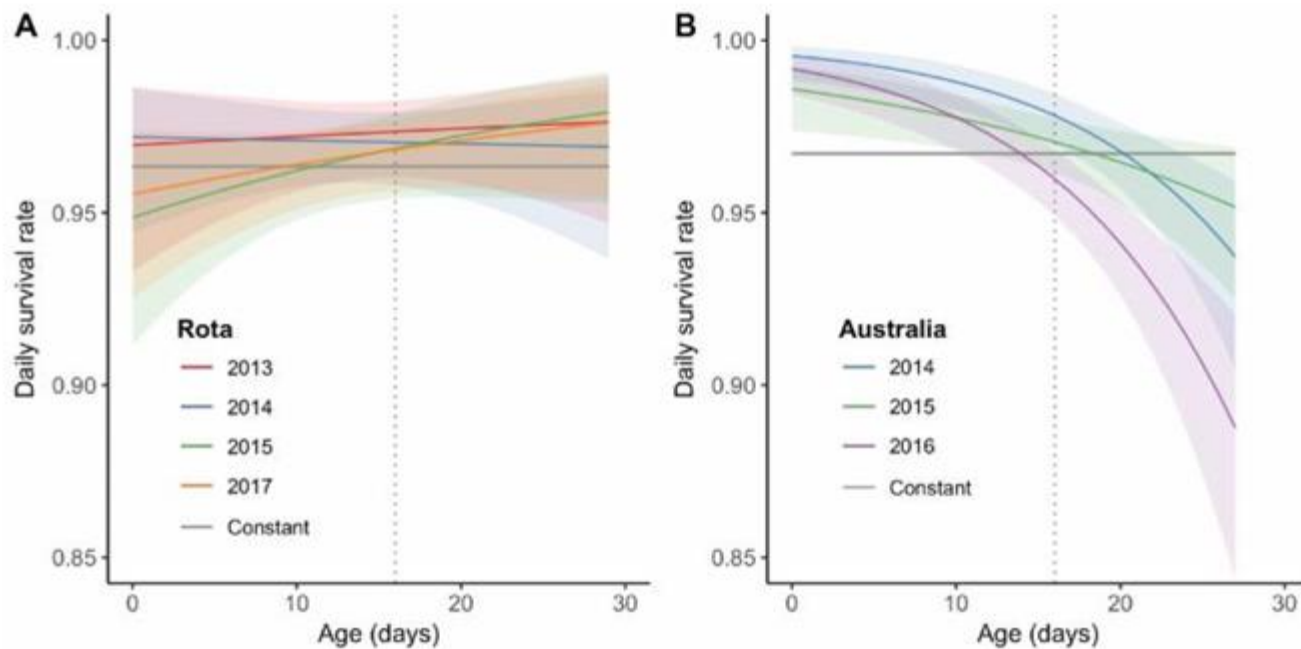
<https://doi.org/10.1642/AUK-17-176.1>

AUK Research Article

Variation in age-dependent nest predation between island and continental Rufous Fantail (*Rhipidura rufifrons*)

Lindsey Nietmann and Renee R. Ha

Published September 5, 2018

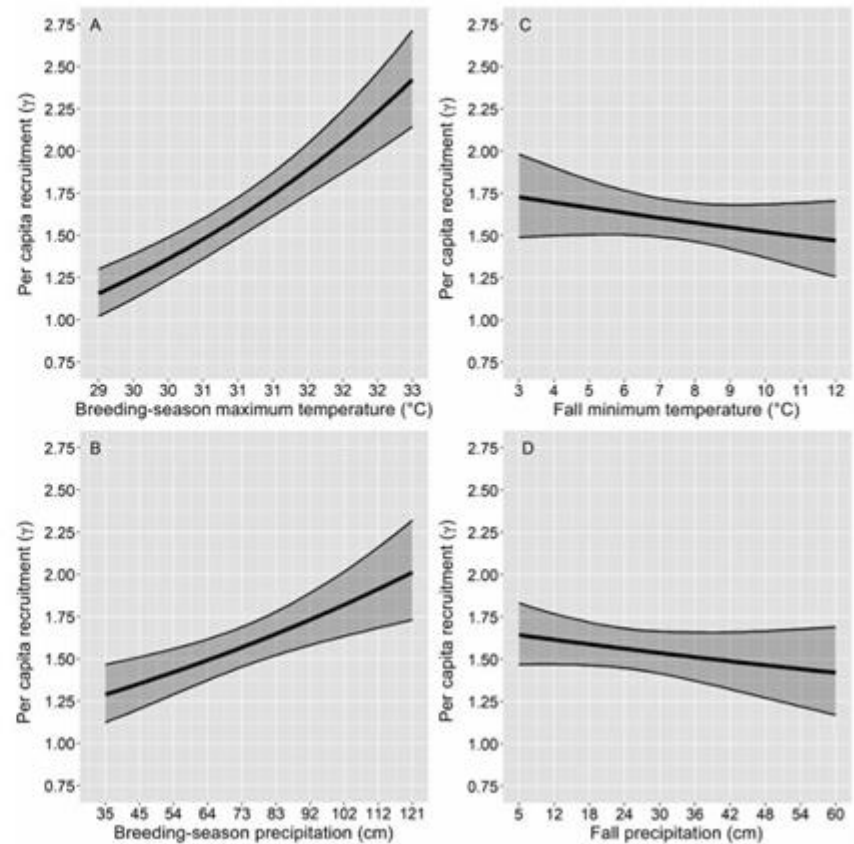


AUK Research Article

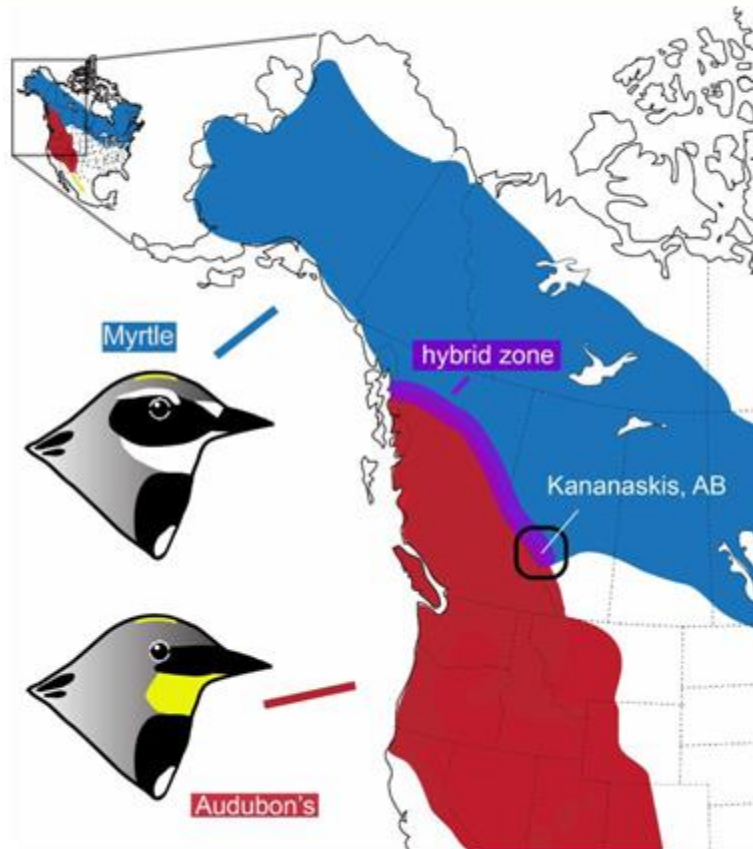
Factors influencing Northern Bobwhite recruitment, with implications for population growth

Mark D. McConnell, Adrian P. Monroe, Richard Chandler, William E. Palmer, Shane D. Wellendorf, L. Wes Burger, Jr., and James A. Martin

Published September 5, 2018



AUK Research Article



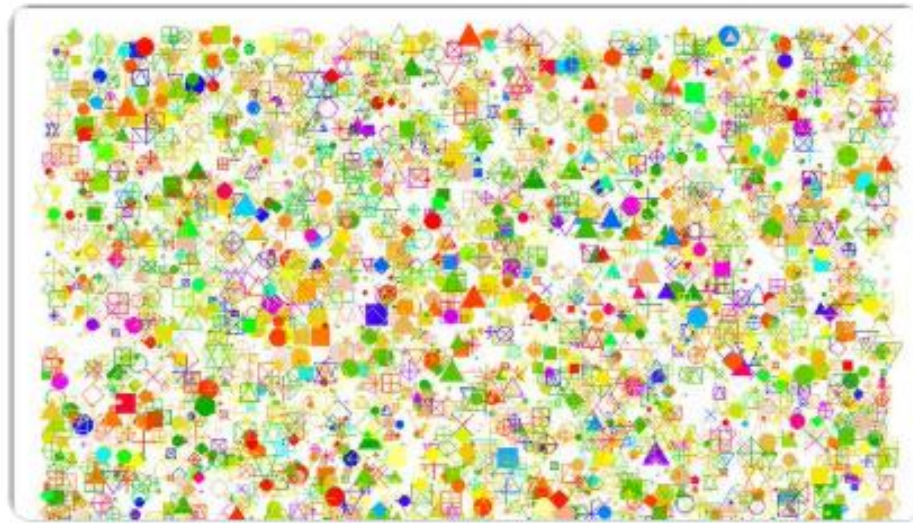
Similar hybrid composition among different age and sex classes in the Myrtle–Audubon's warbler hybrid zone

David P. L. Toews, Irby J. Lovette, Darren E. Irwin, and Alan Brelsford

Published September 26, 2018



Tiago André Marques @TiagoALOMarques · Oct 2
answer to "Can you change the type of point in a plot in R?"
[@accidental_aRt](#) #rstats



Remember

Question: Can you do * in a plot in R?

Answer: (most likely) yes!



Home



Explore



Notifications



Messages



Bookmarks



Lists



Profile



More

Tweet



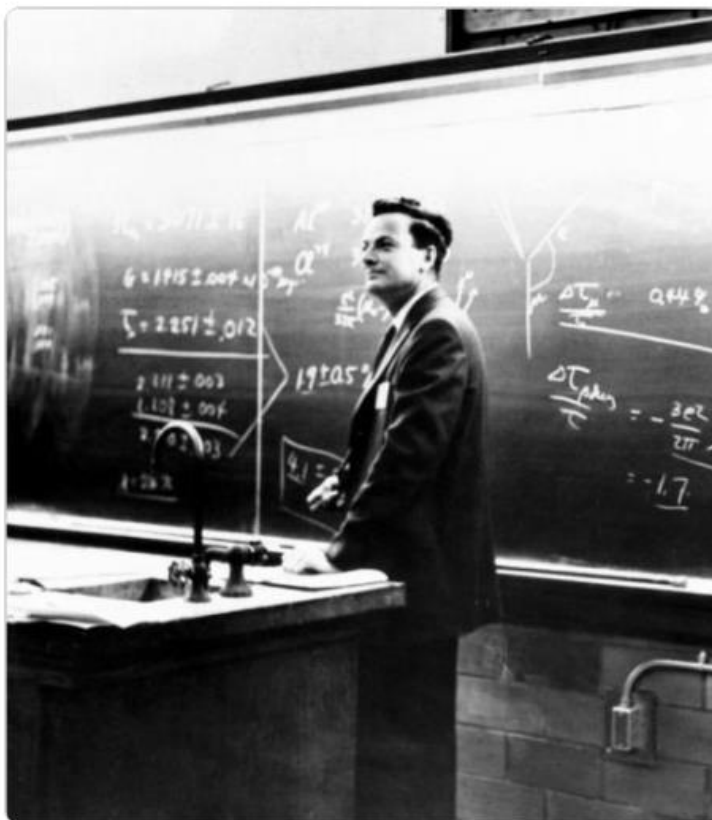
Tweet



Richard Feynman
@ProfFeynman

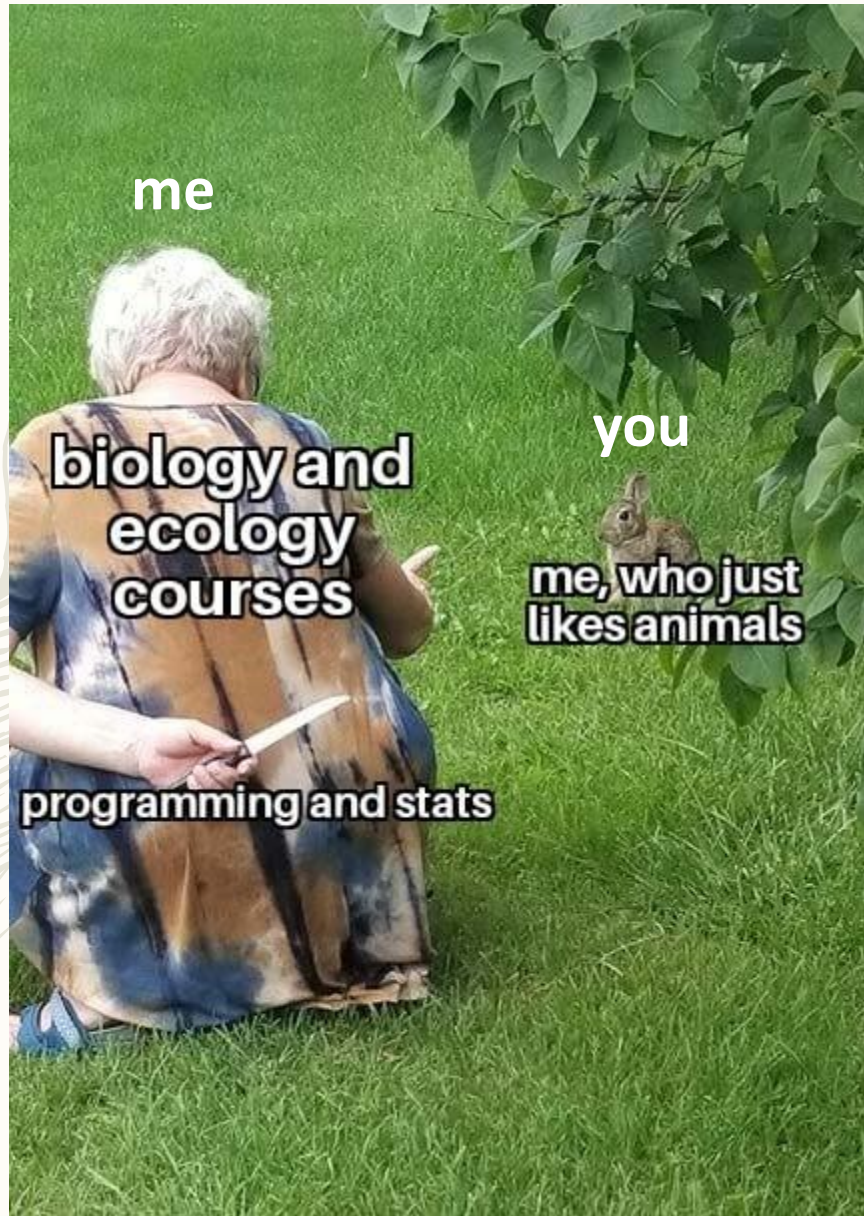
Teach your students

- to doubt,
- to think,
- to communicate,
- to question,
- to make mistakes,
- to learn from their mistakes, and most importantly
- have fun in their learning.



3:34 AM · Oct 6, 2019 · Twitter for Android

3.8K Retweets 10.6K Likes



me

biology and
ecology
courses

programming and stats

you

me, who just
likes animals



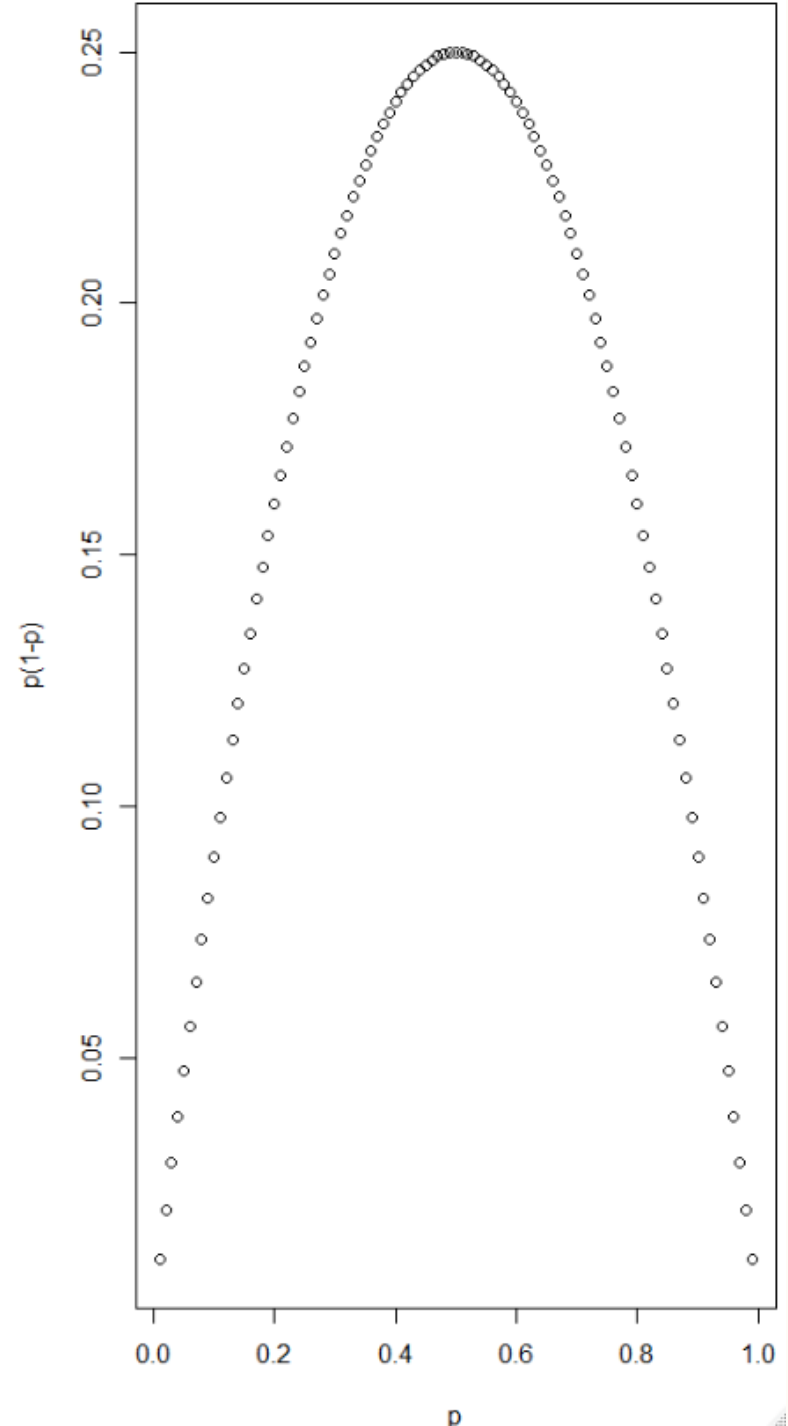
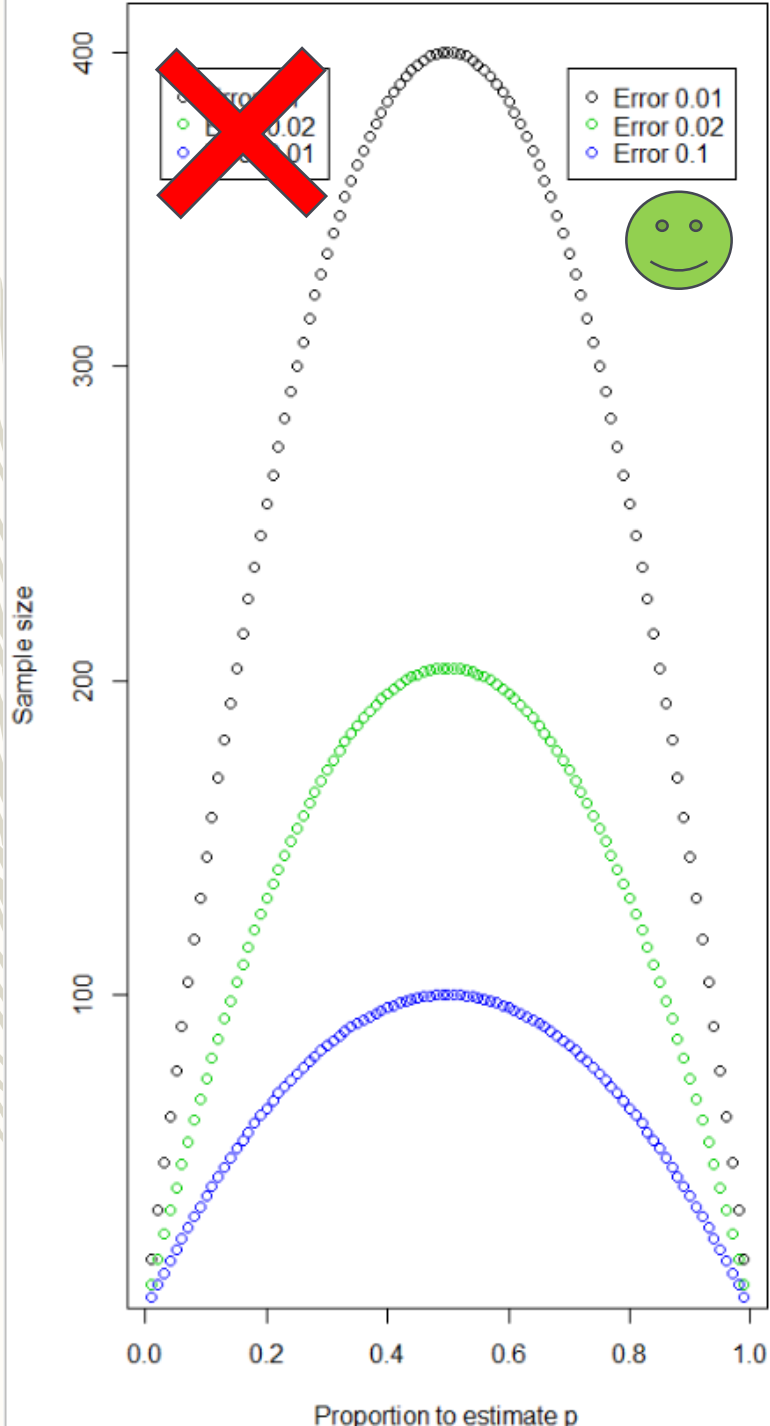


There is no data that can be displayed in a pie chart, that cannot be displayed BETTER in some other type of chart.

— *John Tukey* —

AZ QUOTES

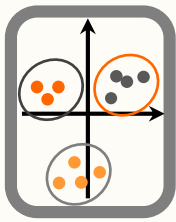
<https://www.azquotes.com/quote/727634>



Código para a figura do slide anterior

$$n = \frac{t_{\alpha}^2 \hat{p}(1 - \hat{p})}{d^2}$$

```
#criar função que calcula n – ver slide da aula anterior
get.n=function(talfa,p,d){
  n=talfa^2*p*(1-p)/d^2
  return(n)
}
#grafico do tamanho da amostra para vários erros
ps=seq(0.01,0.99,by=0.01)
par(mfrow=c(1,2),mar=c(4,4,0.5,0.5))
plot(ps,get.n(2,ps,0.01),ylab="Sample size",xlab="Proportion to estimate p")
points(ps,get.n(2,ps,0.02),col=3)
points(ps,get.n(2,ps,0.1),col=4)
#legenda certa
legend("topright",legend=paste0("Error ",c(0.01,0.02,0.1)),inset=0.05,col=c(1,3,4),pch=1)
#legenda errada
legend("topleft",legend=paste0("Error ",c(0.1,0.02,0.01)),inset=0.05,col=c(1,3,4),pch=1)
plot(ps,ps*(1-ps),xlab="p",ylab="p(1-p)")
```



amostragem

delineamento experimental

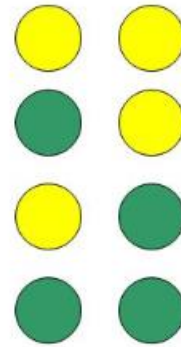
Delineamento experimental

Balancing e blocking:

- Devem ser utilizadas unidades experimentais homogêneas (por vezes é difícil, em particular em experiências fora do laboratório).
- Deve usar-se a informação acerca de variáveis relacionadas.
- Deve usar-se um número grande de replicados.
- Deve usar-se *designs* experimentais mais eficientes: *balancing* e *blocking* devem ser parte integrante dos designs.

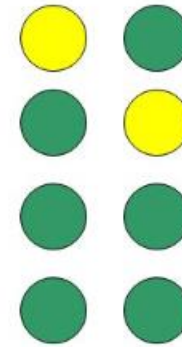
Balancing

In a balanced experimental design, all treatments have equal sample size



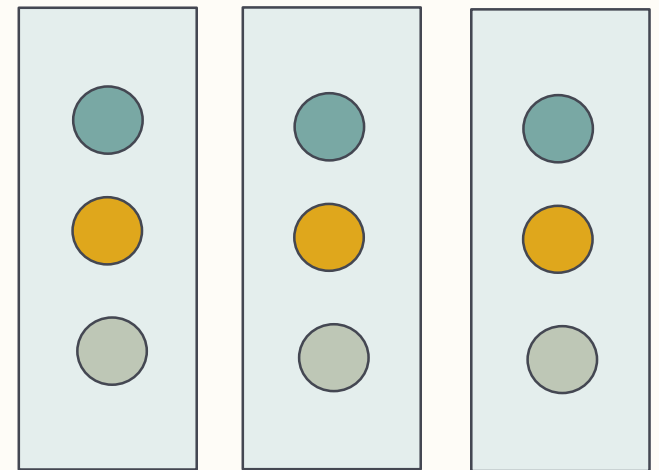
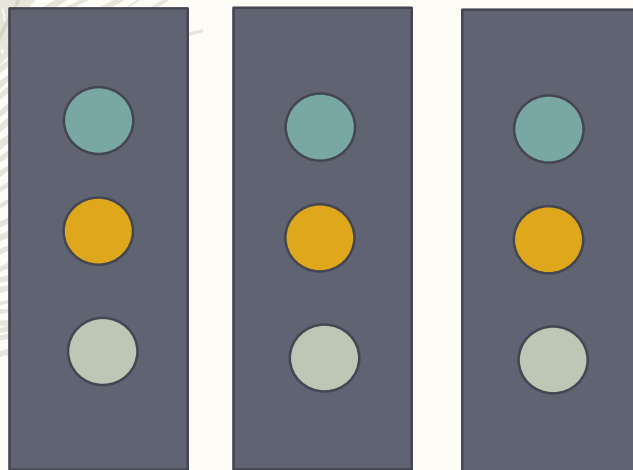
Balanced

Better than



Unbalanced

Blocking – fazer replicados dentro de unidades de amostragem homogêneas entre si (exemplos de blocos “naturais”: ninhada, árvore, tanque, poça, etc)



A questão fundamental tem a ver com as fontes de variação.

$$\text{Observações} = \text{Tratamento} + \text{Erro}$$

+ replicados

variabilidade

blocos

$$\text{Observações} = \text{Tratamento} + \text{Bloco} + \text{Erro}$$

Tudo o que não conseguirmos explicar como fonte de variação acaba na componente do erro. Se o erro for grande, não conseguimos fazer inferências – nada é diferente de nada, o erro é demasiado grande para tirar outras conclusões.

Quando parte da variação é explicada, por um factor, por um bloco, ou por um tratamento, a componente de variância associada ao erro diminui, e ficamos mais próximos de fazer inferências interessantes.



planeamento de estudos

Quais os aspectos a ter em conta no planeamento de estudos em ecologia?

Exemplos



planeamento de estudos

1. Pretende comparar a abundância de uma espécie de ave em diferentes habitats.





planeamento de estudos

2. Pretende comparar a abundância e a dieta duma espécie de mamíferos em diferentes regiões.





planeamento de estudos

3. Pretende estudar a relação entre a densidade dum gastrópode que vive na zona entre-marés e a distância ao zero hidrográfico, em praias com diferentes graus de hidrodinamismo.





planeamento de estudos

4. Pretende caracterizar as capturas duma pescaria artesanal ao longo do ano e obter estimativas globais para toda essa frota.





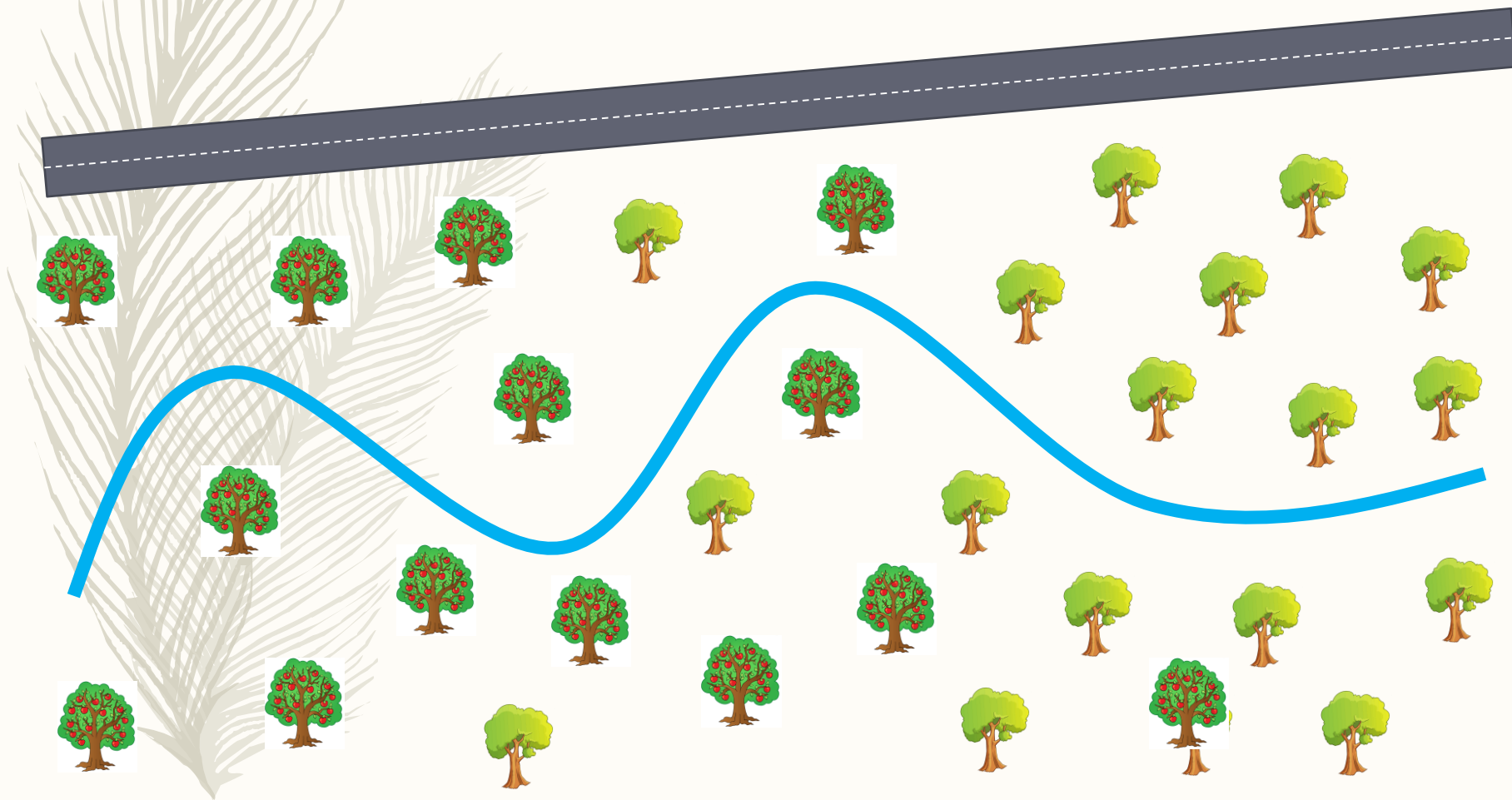
planeamento de estudos

5. Pretende avaliar o efeito de dois contaminantes (Hg e Pb) no crescimento de uma espécie de crustáceo. Para o efeito vai realizar uma experiência em laboratório.

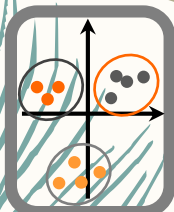
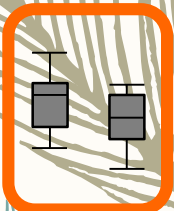
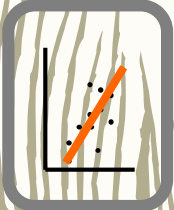
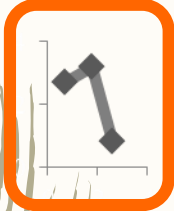


Tarefa (grupos de 3 ou 4):

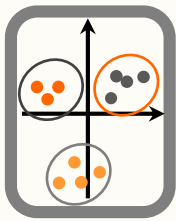
Defina um protocolo de amostragem para recolher um conjunto de dados que permita responder à seguinte pergunta: será que os insectos tem diferentes níveis de actividade durante o dia, e esses níveis de actividade dependem de quão próximos estão de um rio? O esquema abaixo representa a área de estudo. Discuta as potenciais implicações e constrangimentos, justificando as escolhas que fizer. Enviar a resolução ao docente por e-mail.



Ecologia Numérica

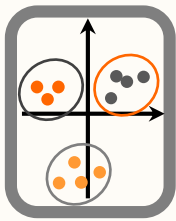


análise
exploratória
de dados



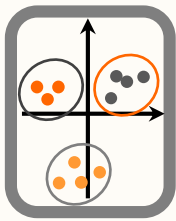
análise exploratória de dados

- Qual o objectivo da análise exploratória de dados?
- Qual a sua importância?
- Que indicadores podem ser utilizados e que informação transmitem?
- O que deverá ser diagnosticado numa análise exploratória de dados?



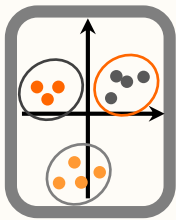
análise exploratória de dados

- Antes de qualquer procedimento analítico mais elaborado deve proceder-se à análise exploratória dos dados
- Este tipo de análise permite-nos obter um maior conhecimento sobre os conjuntos de dados e identificar aspectos importantes par a selecção dos procedimentos a efectuar seguidamente



análise exploratória de dados

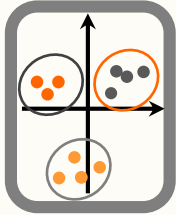
- Geralmente baseada em estatísticas descritivas e representações gráficas
- As estatísticas descritivas mais frequentemente utilizadas são medidas de tendência central e de dispersão dos dados



Medidas de tendência central



- Média: `mean(x)`
- Mediana: `median(x)`
- Moda: `mode(x)`
- Quantis: `quantile(x, probs)`
- Ponto médio da amplitude: `diff(range(x))/2`

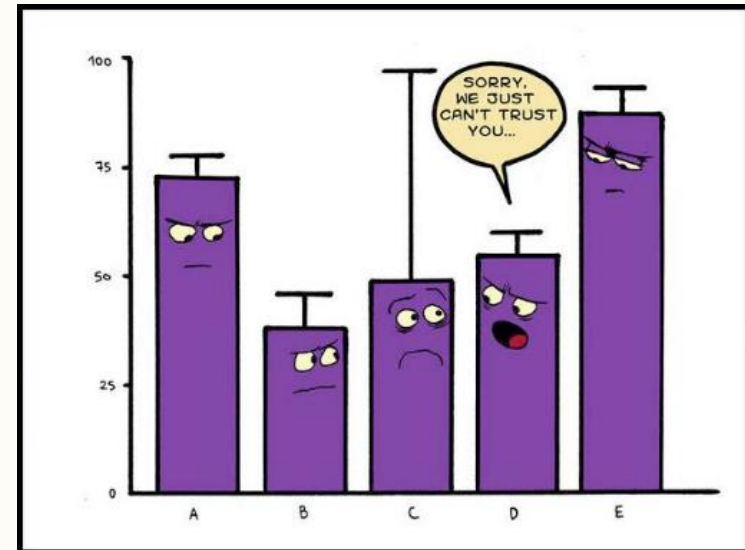


análise exploratória de dados

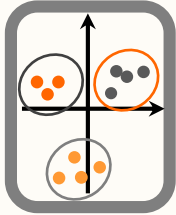
Medidas de dispersão

- Amplitude: $\text{diff}(\text{range}(x))$
- Amplitude interquantil
- Variância: $\text{var}(x)$
- Desvio-padrão: $\text{stdv}(x)$
- Coeficiente de variação: $\text{stdv}(x) / \text{mean}(x)$

?



```
diff(quantile(x, probs=c(0.25, 0.75)))  
quantile(x, 0.75) - quantile(x, 0.25)  
summary(x)[5] - summary(x)[2]
```



análise exploratória de dados

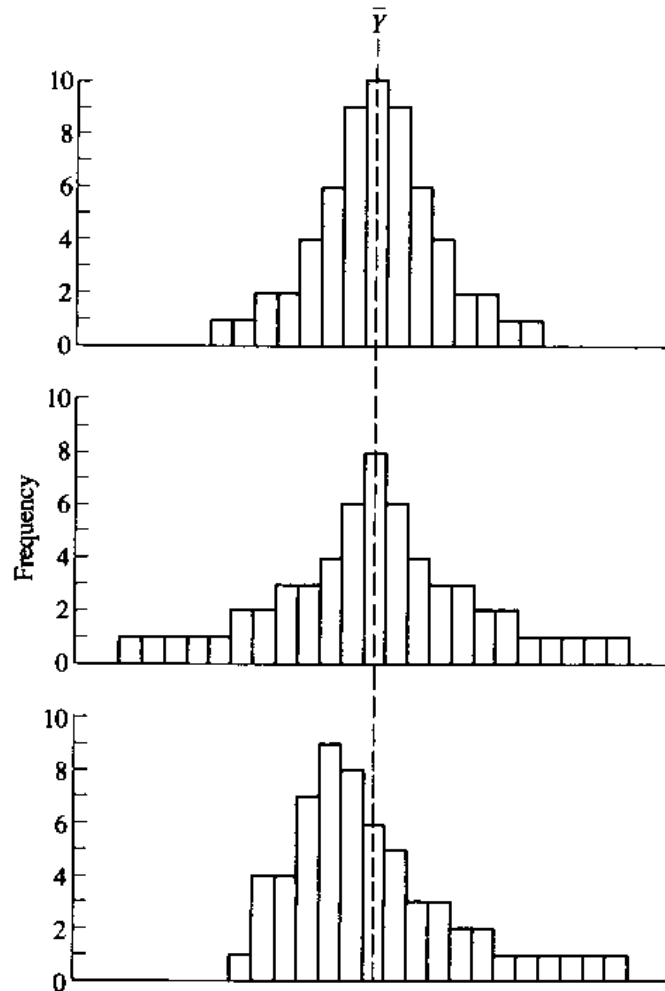
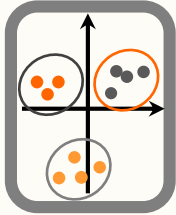


FIGURE 4.2 Three frequency distributions with identical means and sample sizes but different dispersion patterns.



análise exploratória de dados

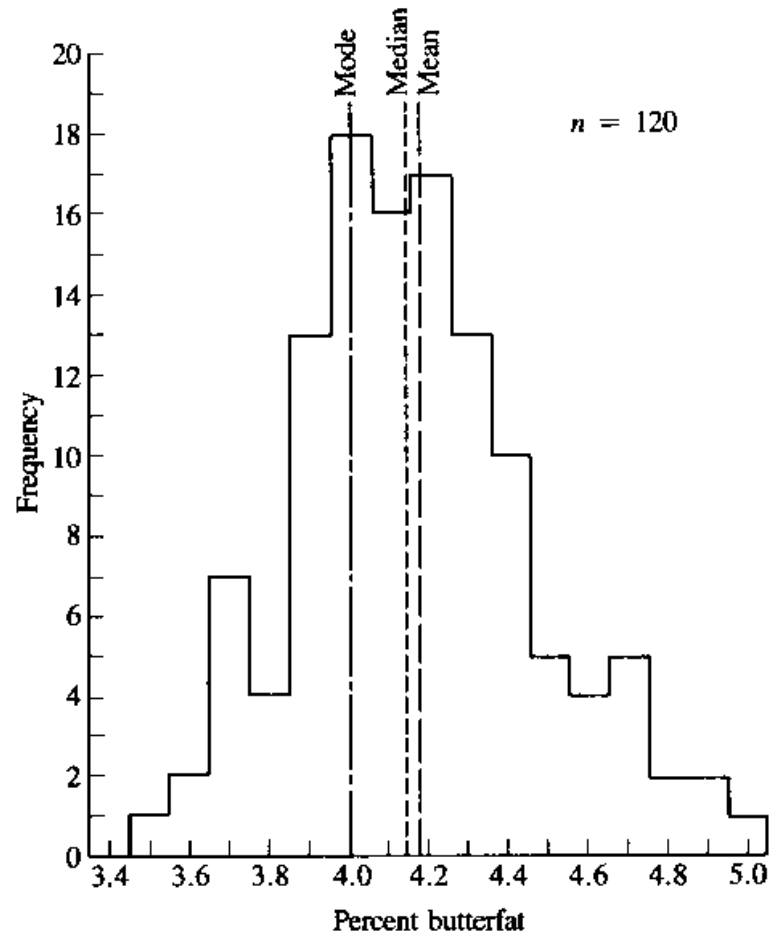
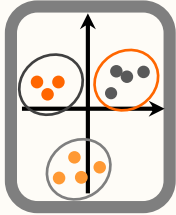


FIGURE 4.1 An asymmetrical frequency distribution (skewed to the right) showing location of the mean, median, and mode. Percent butterfat in 120 samples of milk (from a Canadian cattle breeder's record book).



análise exploratória de dados

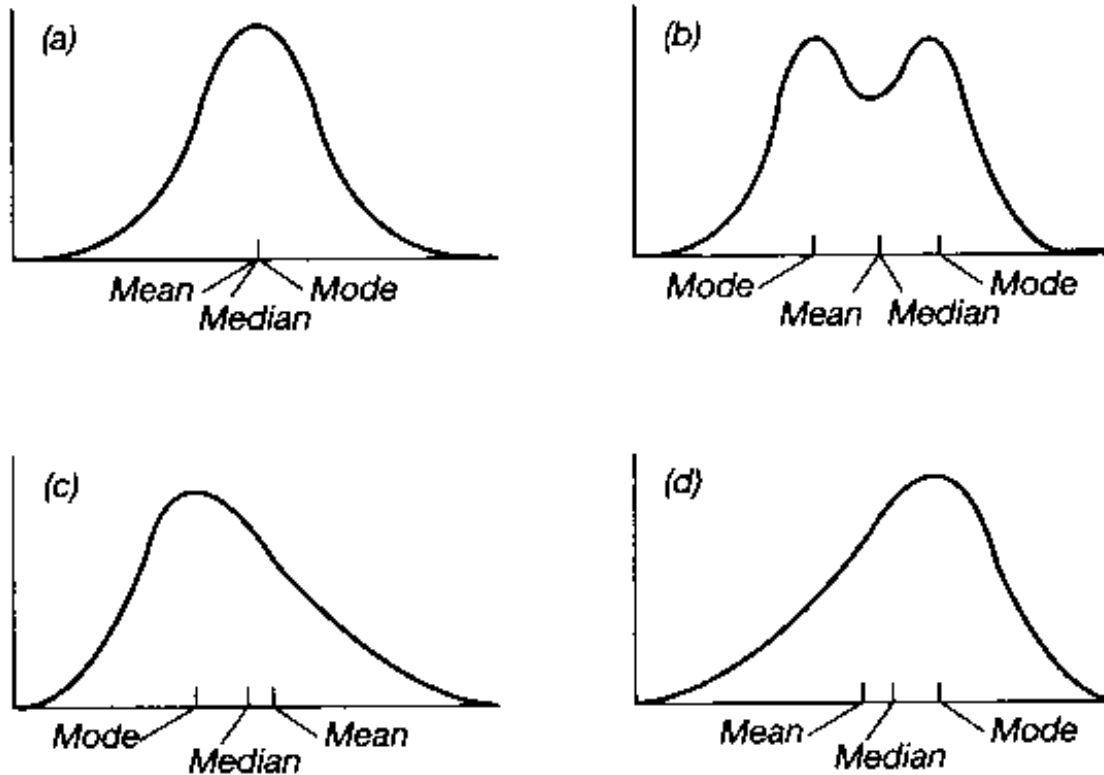
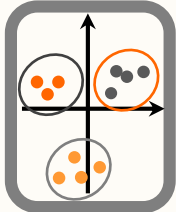


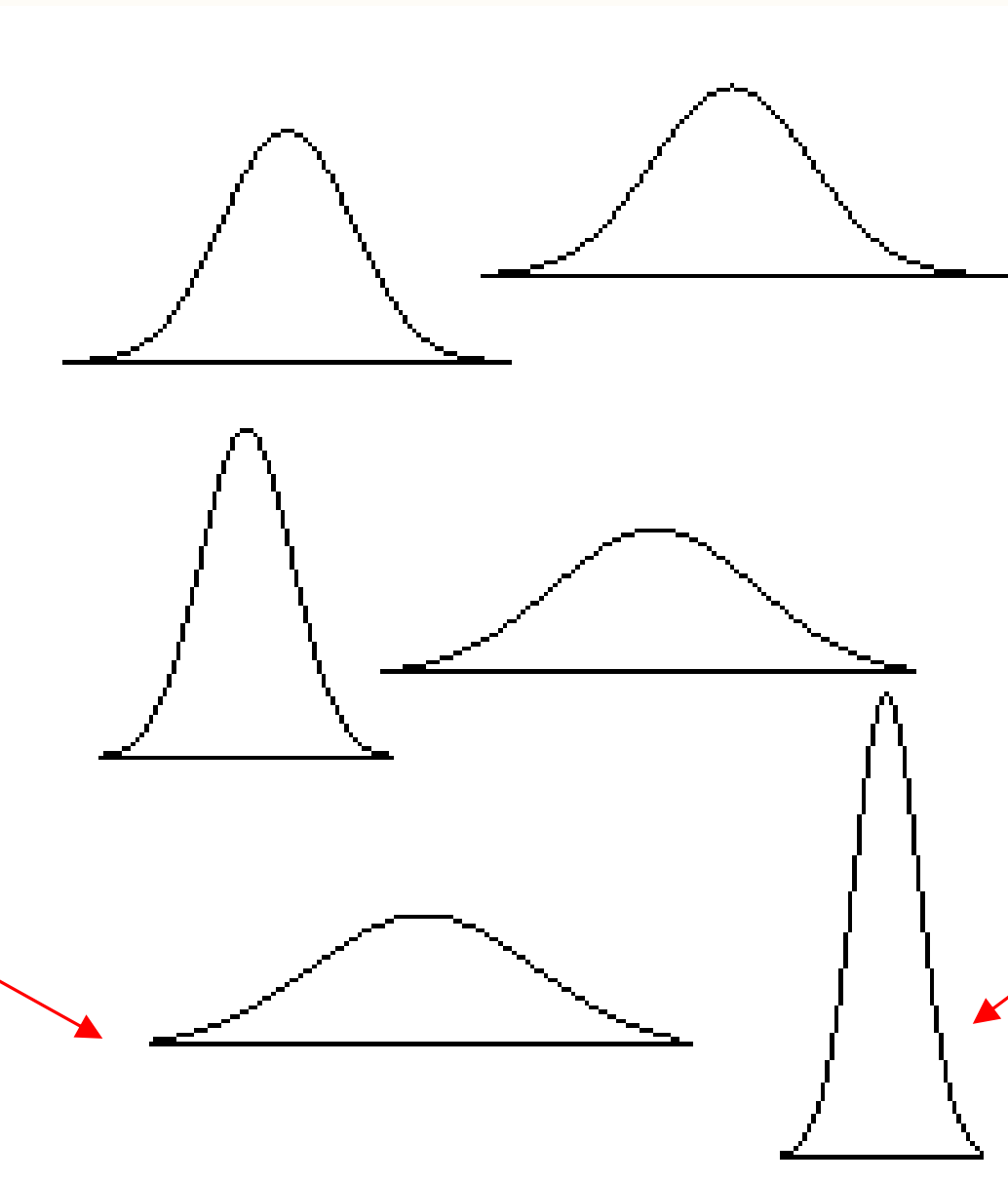
Figure 3.2 Frequency distributions showing measures of central tendency. Values of the variable are along the abscissa (horizontal axis), and the frequencies are along the ordinate (vertical axis). Distributions (a) and (b) are symmetrical, (c) is positively skewed, and (d) is negatively skewed. Distributions (a), (c), and (d) are unimodal, and distribution (b) is bimodal. In a unimodal asymmetric distribution, the median lies about one-third the distance between the mean and the mode.[†]



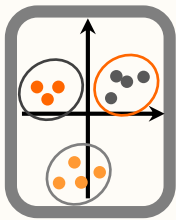
análise exploratória de dados



platicúrtica



leptocúrtica



Análises gráficas

- Histogramas
- Diagramas de caule e folhas
- Gráficos de caixas e bigodes
- Gráficos de sectores circulares
- Gráficos de barras
- Gráficos de linhas
- Gráficos de dispersão

Nunca se fazem
demasiados gráficos
(numa análise
exploratória de dados!)

Um recurso muito bom quando estamos à procura “daquele” gráfico para impressionar!

Classes de gráficos

Distribution



Violin



Density



Histogram



Boxplot



Ridgeline

Correlation



Scatter



Heatmap



Correlogram



Bubble



Connected scatter



Density 2d

Ranking



Barplot



Spider / Radar



Wordcloud



Parallel



Lollipop



Part of a whole



Grouped and Stacked barplot



Treemap



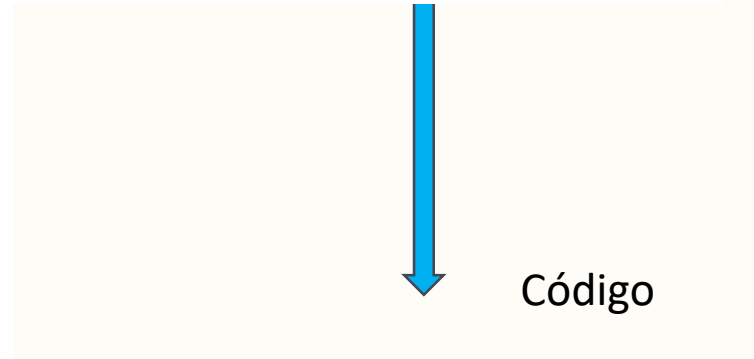
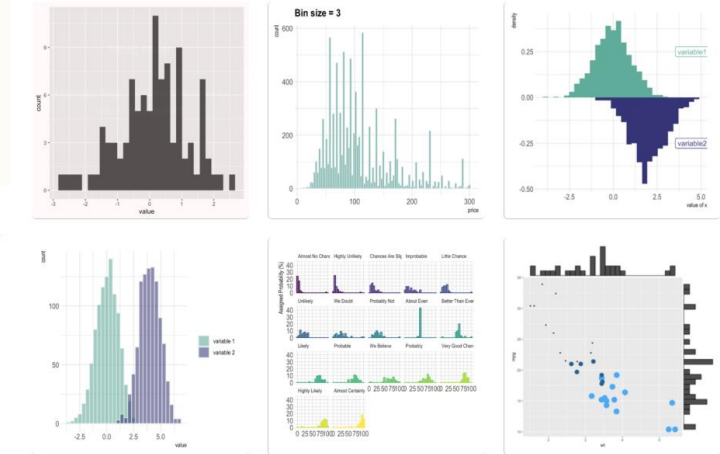
Doughnut



Pie chart



Dendrogram



Código

Basic histogram with `geom_histogram`

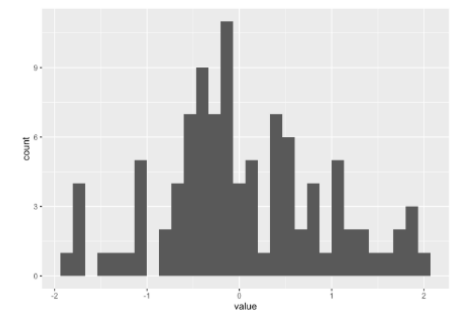
It is relatively straightforward to build a histogram with `ggplot2` thanks to the `geom_histogram()` function. Only one numeric variable is needed in the input. Note that a warning message is triggered with this code: we need to take care of the bin width as explained in the next section.

```
# library
library(ggplot2)

# dataset:
data=data.frame(value=rnorm(100))

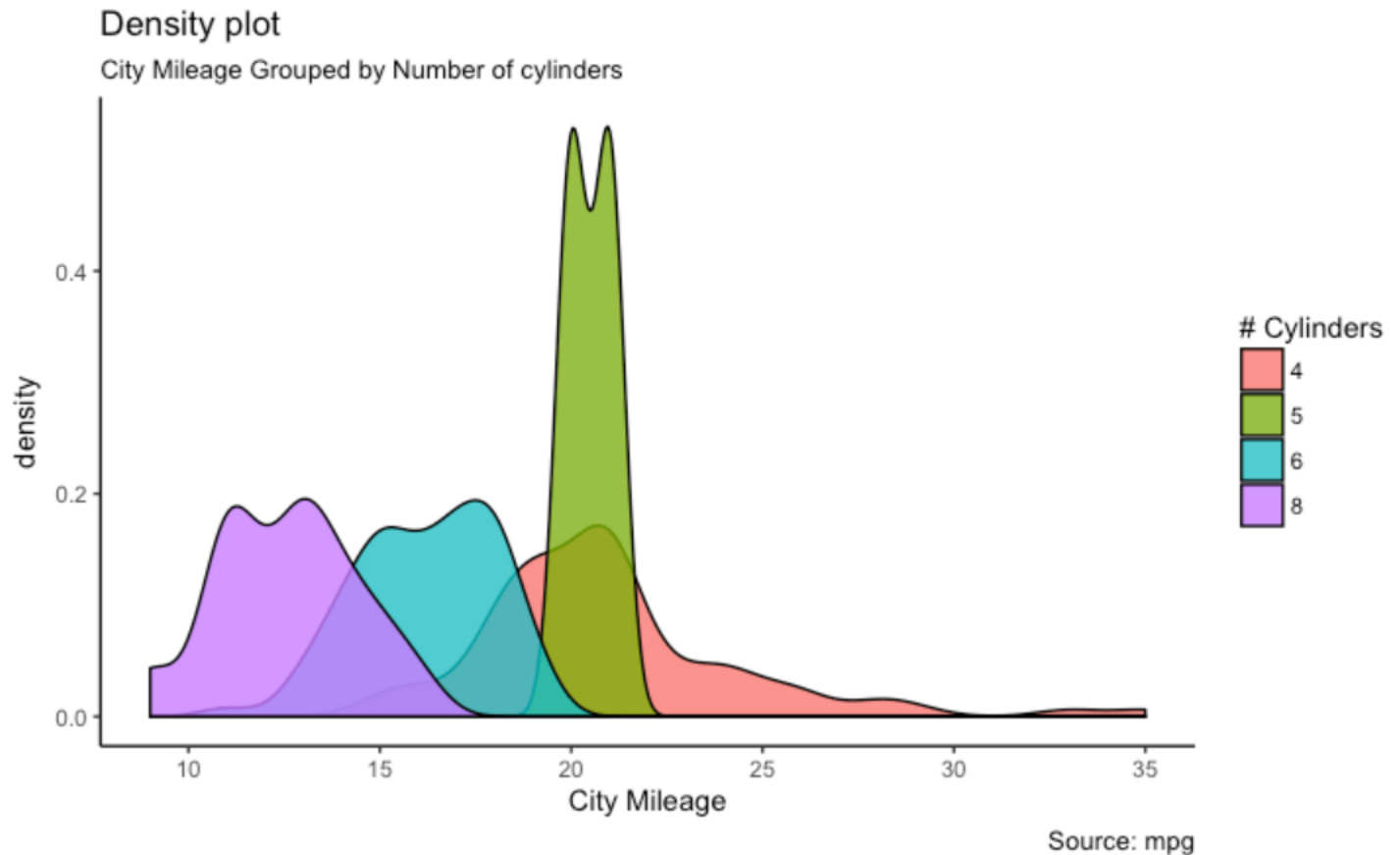
# basic histogram
p <- ggplot(data, aes(x=value)) +
  geom_histogram()

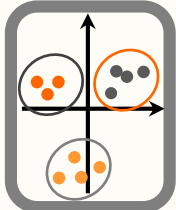
#p
```



No R há por default muitos muitos muitos recursos gráficos para explorar

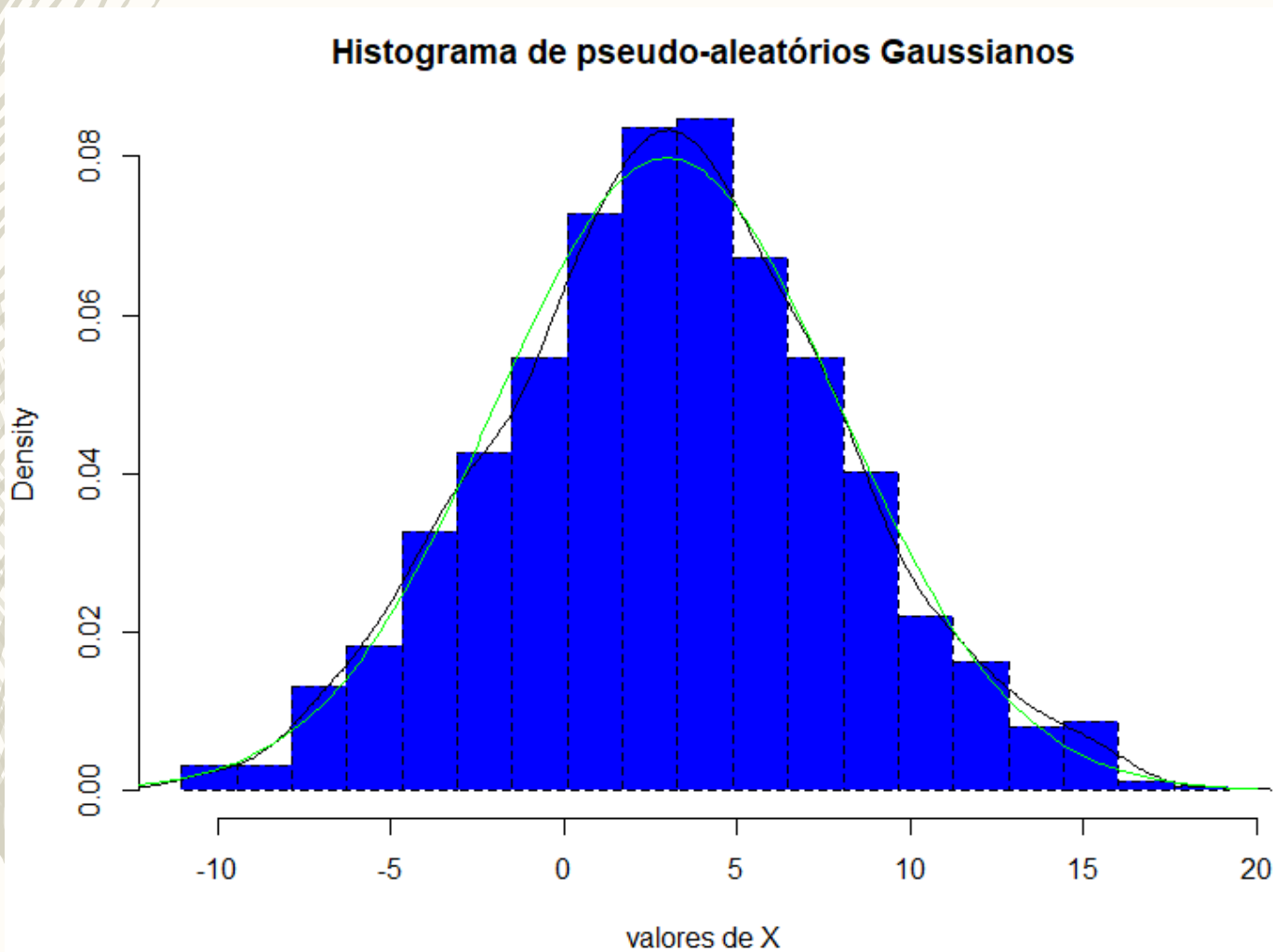
Mas um aparte de que não falaremos, o [ggplot2](#) pode ser o futuro





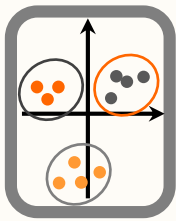
análise exploratória de dados

Histogramas



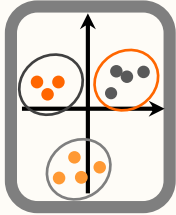
Código para fazer a figura do slide anterior!

```
```{r}
#so that my plot is equal in any computer
set.seed(123)
#set graphical windowdetails
par(mfrow=c(1,1),mar=c(4,4,2.5,0.5))
#generate 1000 random numbers from a Gaussian distribution with mean 3 and sd 5
media=3;dp=5
dados=rnorm(1000,mean=media,sd=dp)
#get limits for histogram bins
mybr=seq(min(dados),max(dados),length=20)
#define a title
titulo="Histograma de pseudo-aleatórios Gaussianos"
hist(dados,breaks=mybr,col="blue",lty=2,xlab="valores de X",main=titulo,freq=FALSE)
#juntar um kernel smooth
lines(density(dados))
#e a verdadeira densidade
xs=seq(-20,20,by=0.1)
lines(xs,dnorm(xs,media,dp),col="green")
```
```



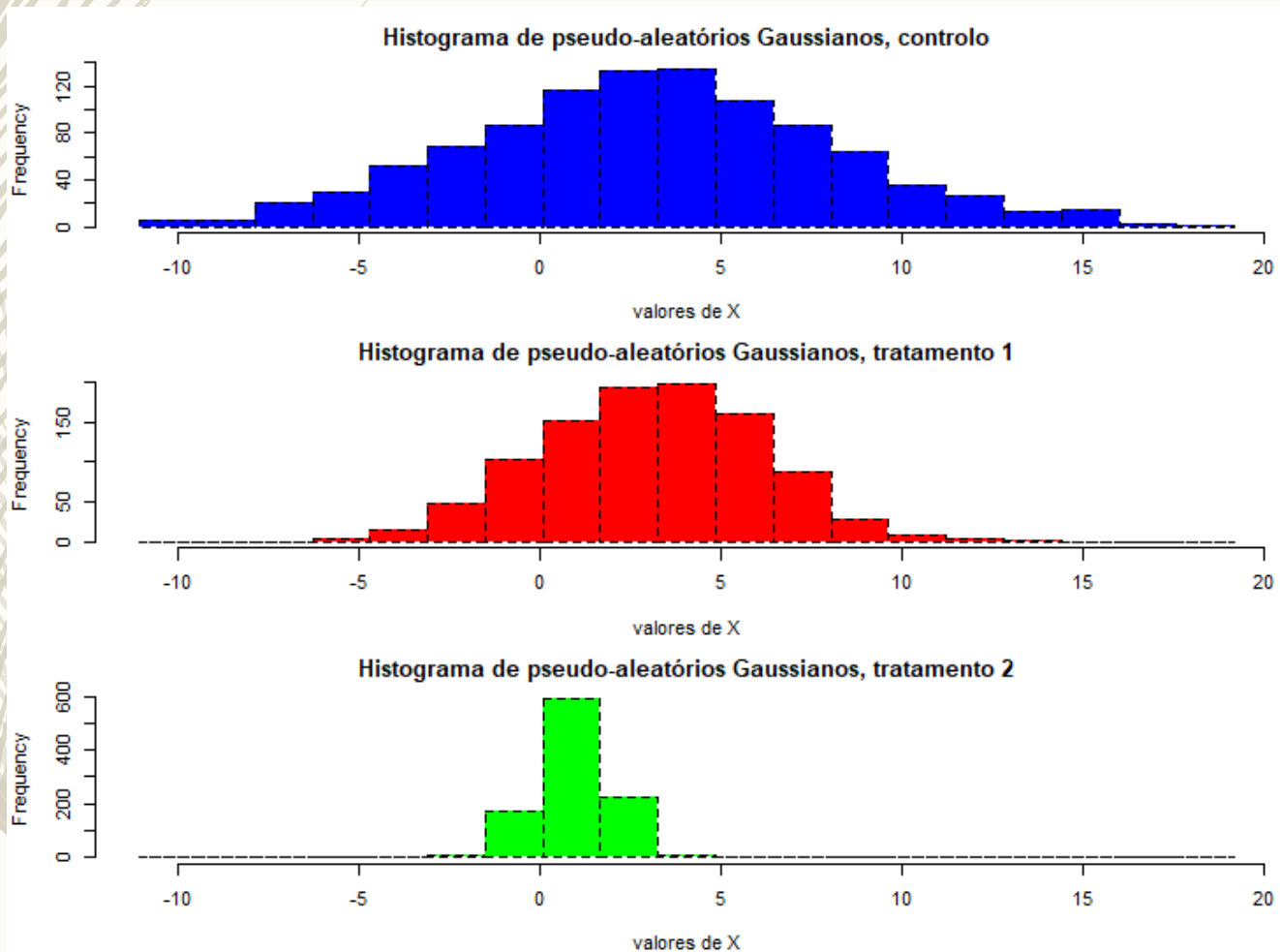
Histogramas

- Permitem avaliar:
 - a forma da distribuição
 - simetria
 - dispersão
- N^o classes pode alterar a forma
- N^o de observações deve ser elevado



análise exploratória de dados

Histogramas



? hist

hist {graphics}

R Documentation

Histograms

Description

The generic function `hist` computes a histogram of the given data values. If `plot = TRUE`, the resulting object of [class](#) "histogram" is plotted by [plot.histogram](#), before it is returned.

Usage

```
hist(x, ...)
```

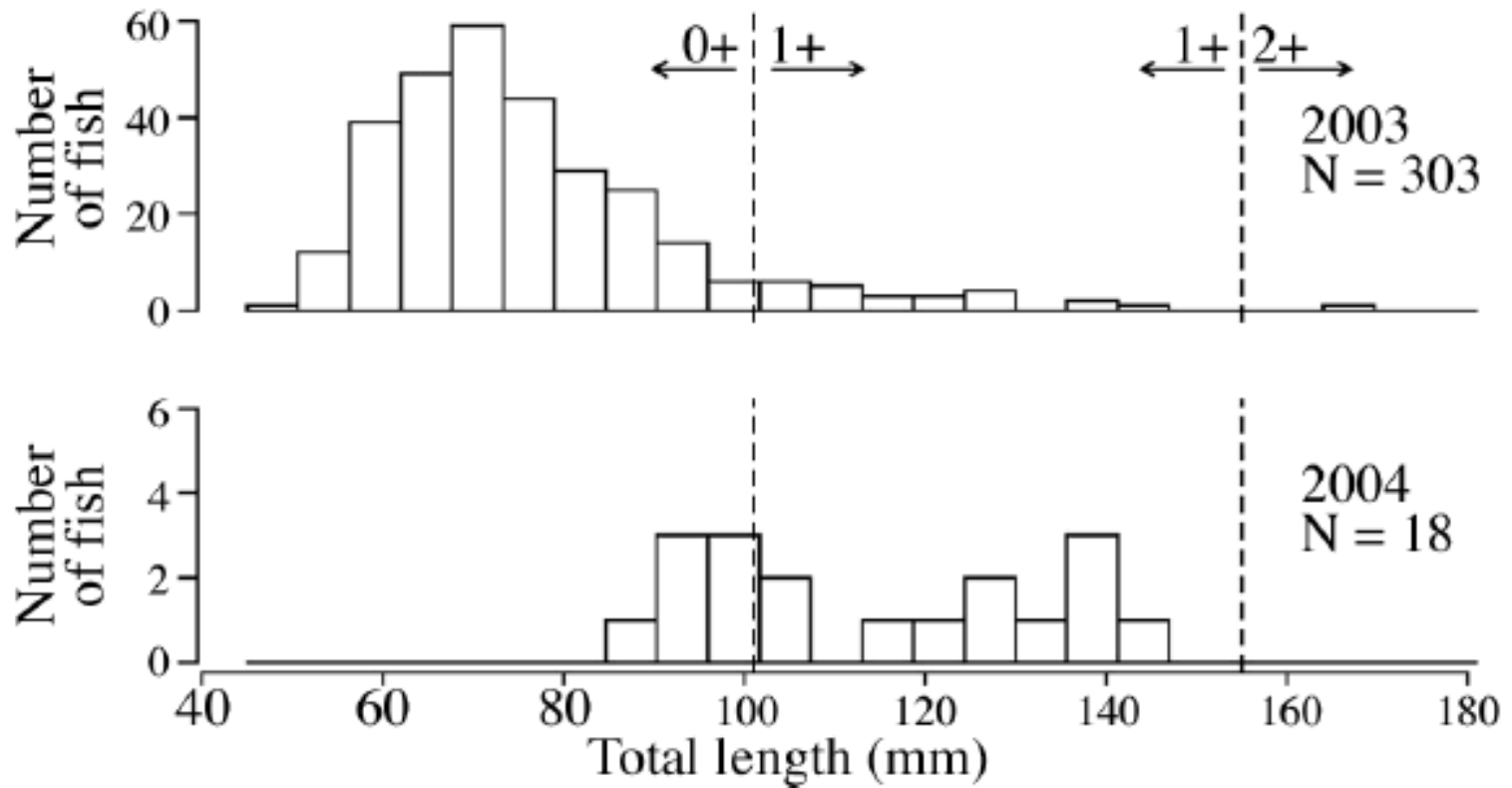
```
## Default S3 method:
```

```
hist(x, breaks = "Sturges",  
     freq = NULL, probability = !freq,  
     include.lowest = TRUE, right = TRUE,  
     density = NULL, angle = 45, col = NULL, border = NULL,  
     main = paste("Histogram of" , xname),  
     xlim = range(breaks), ylim = NULL,  
     xlab = xname, ylab,  
     axes = TRUE, plot = TRUE, labels = FALSE,  
     nclass = NULL, warn.unused = TRUE, ...)
```

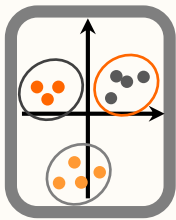
Arguments

`x` a vector of values for which the histogram is desired.

`breaks` one of:

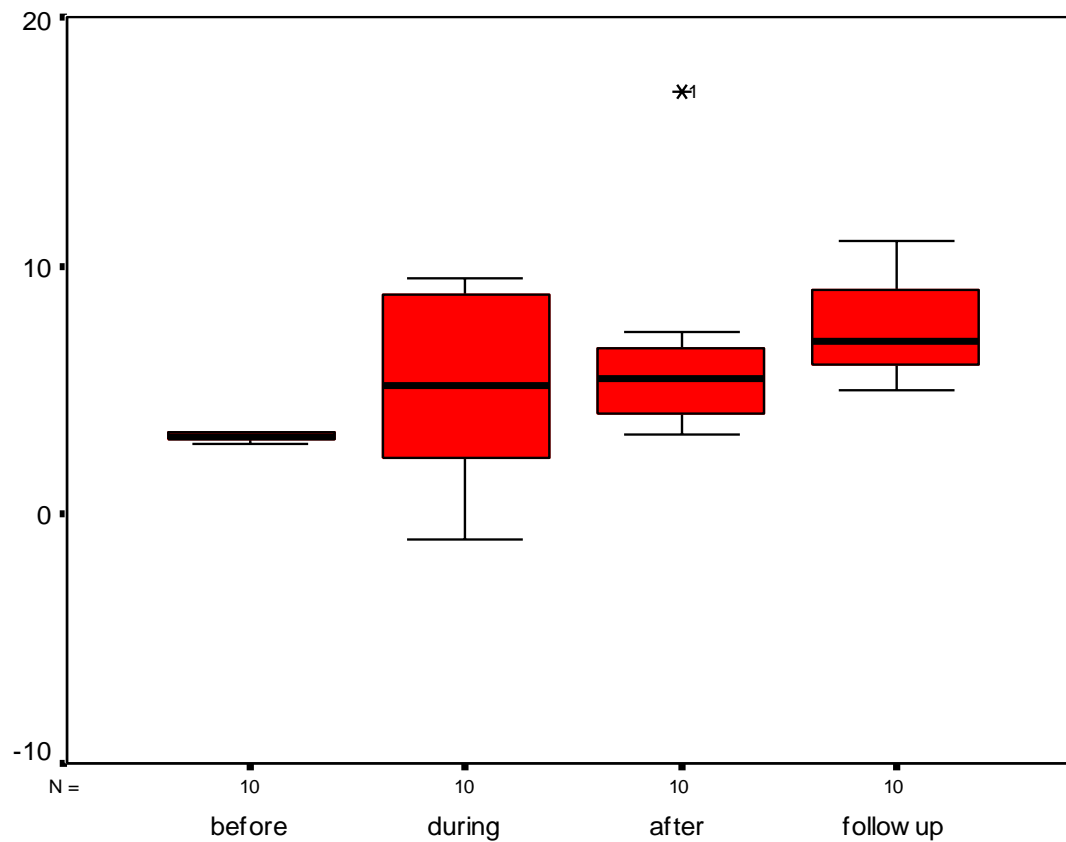


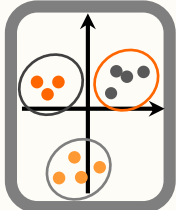
Ribeiro, F.; Chaves, M.; Marques, T. & Moreira Da Costa, L. 2006 First record of *Ameiurus melas* (Siluriformes, Ictaluridae) in the Alqueva reservoir, Guadiana basin (Portugal) *Cybium* **30**: 283-284



análise exploratória de dados

Box-plots

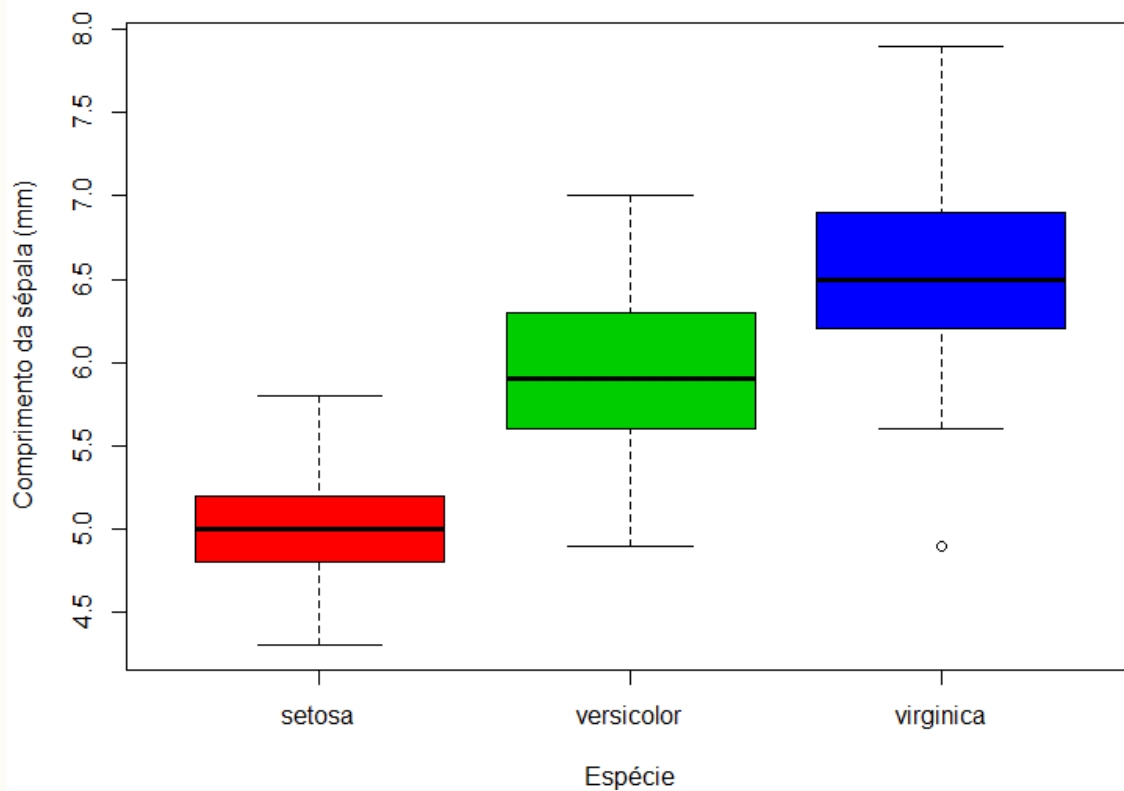




análise exploratória de dados

```
{r}  
data(iris)  
boxplot(iris$Sepal.Length~iris$Species,  
xlab="Espécie",  
ylab="Comprimento da sépala (mm)",col=2:4)
```

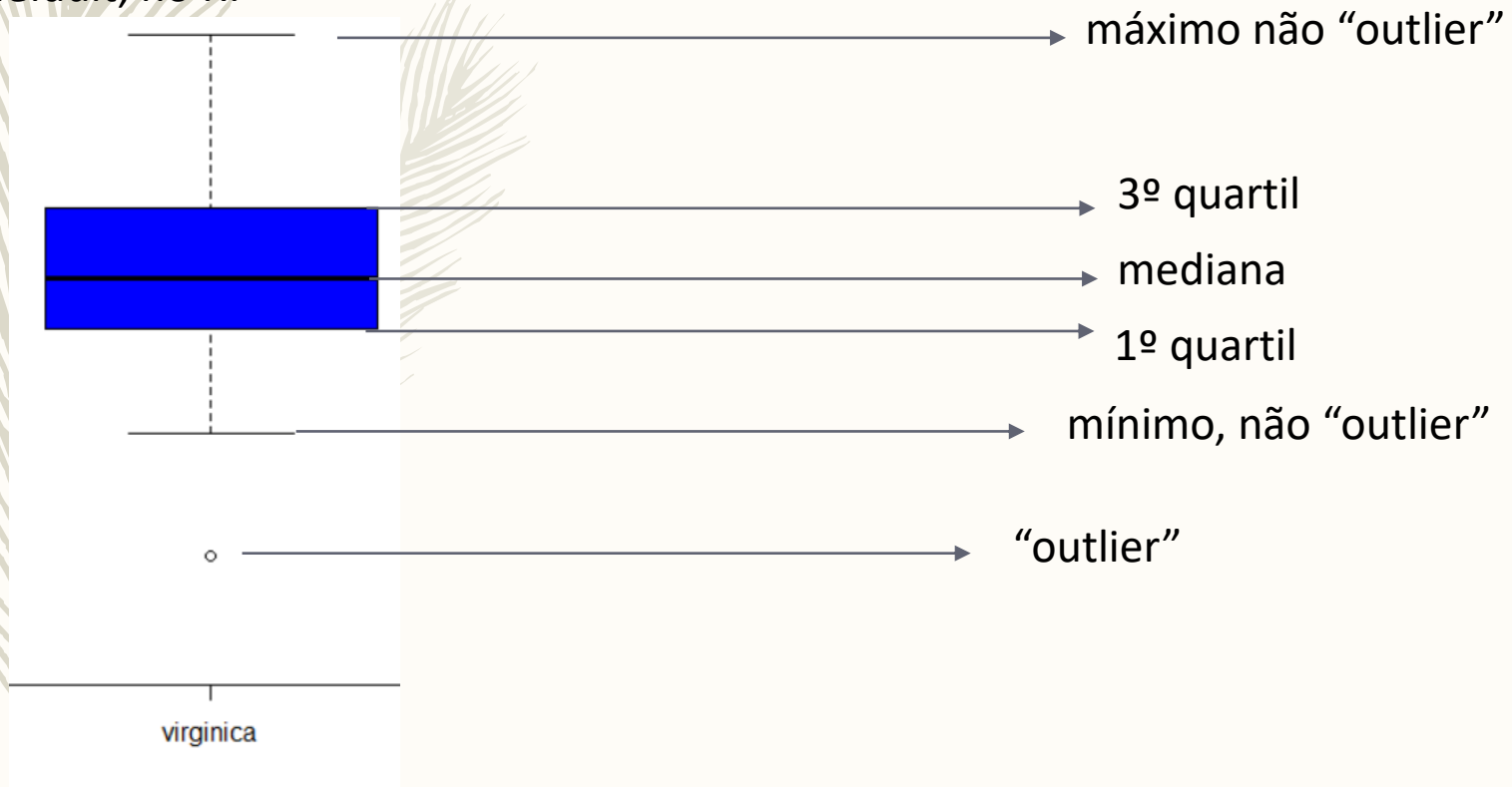
Box-plots



No R, função `boxplot`

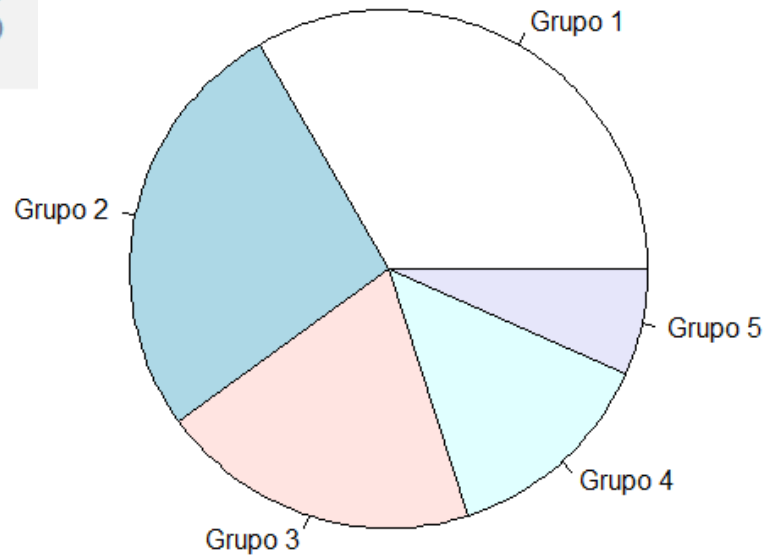
Para perceber e controlar o que é cada elemento, ver função `boxplot.stats`

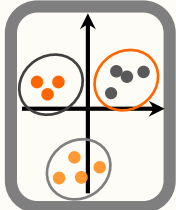
Por default, no R:



Pie charts, queijos ou gráficos de sectores circulares

```
plot({r}  
par(mfrow=c(1,1),mar=c(4,4,2.5,0.5))  
pie(5:1,labels=paste0("Grupo ",1:5))
```

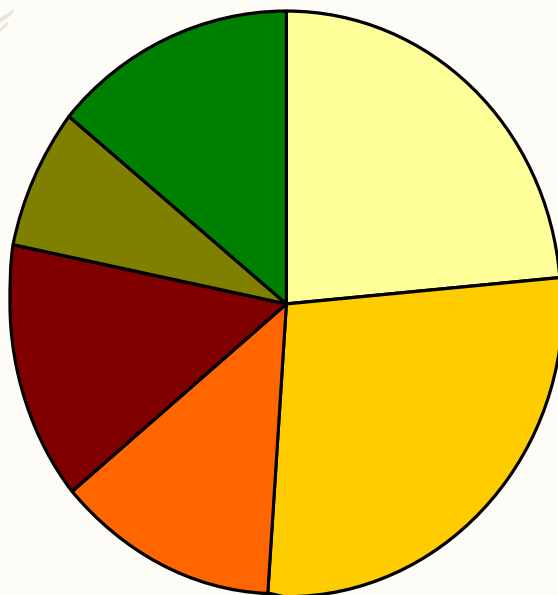


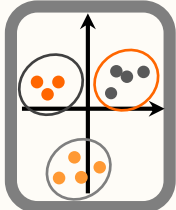


análise exploratória de dados

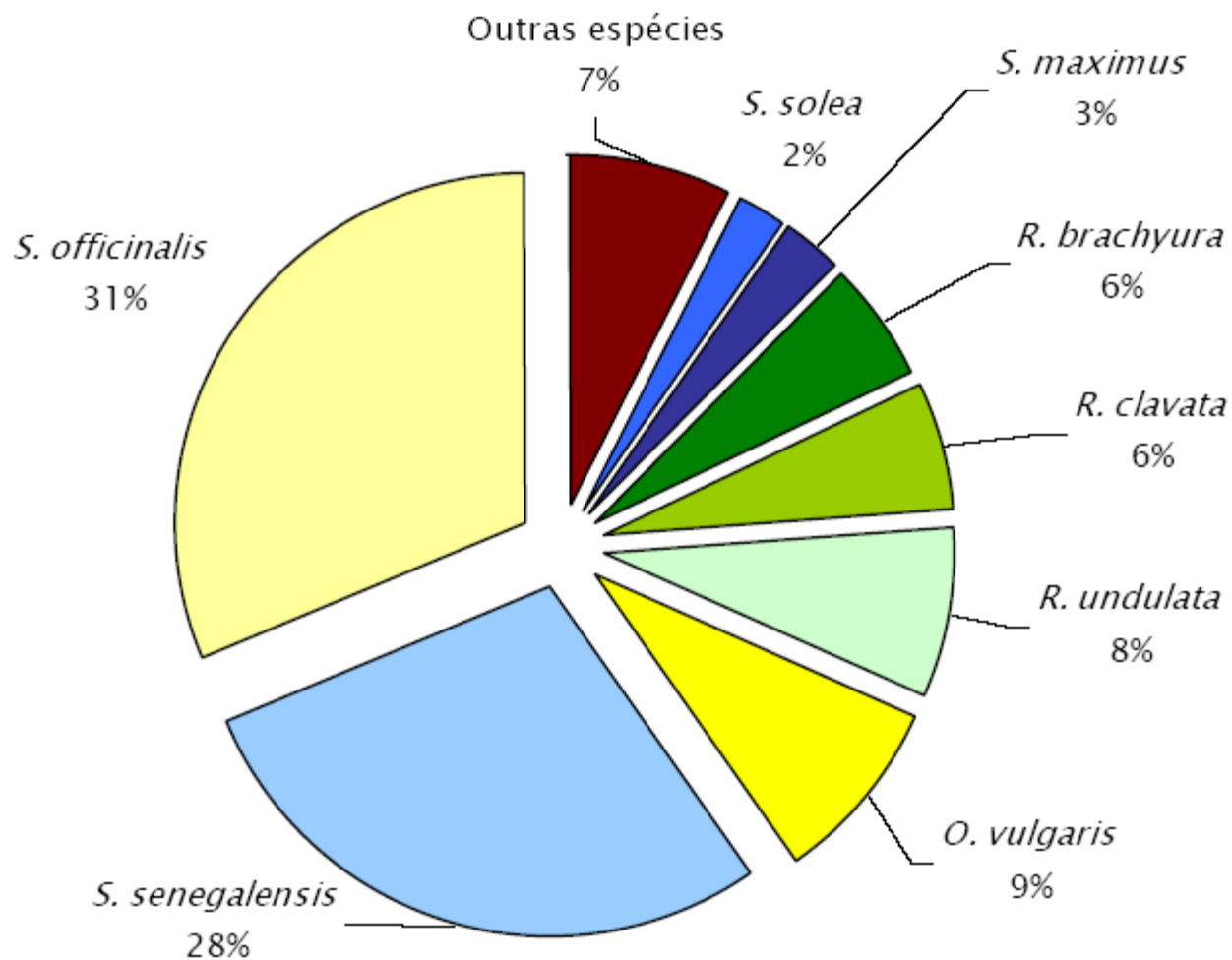
Gráficos de sector circular

IN

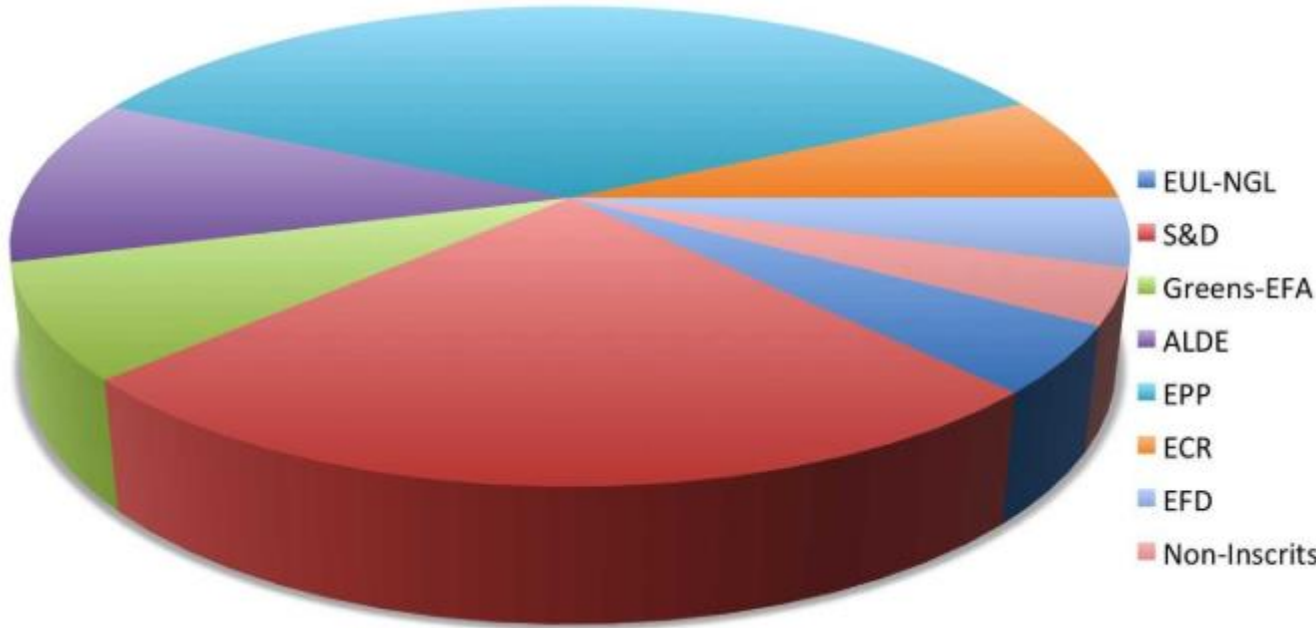




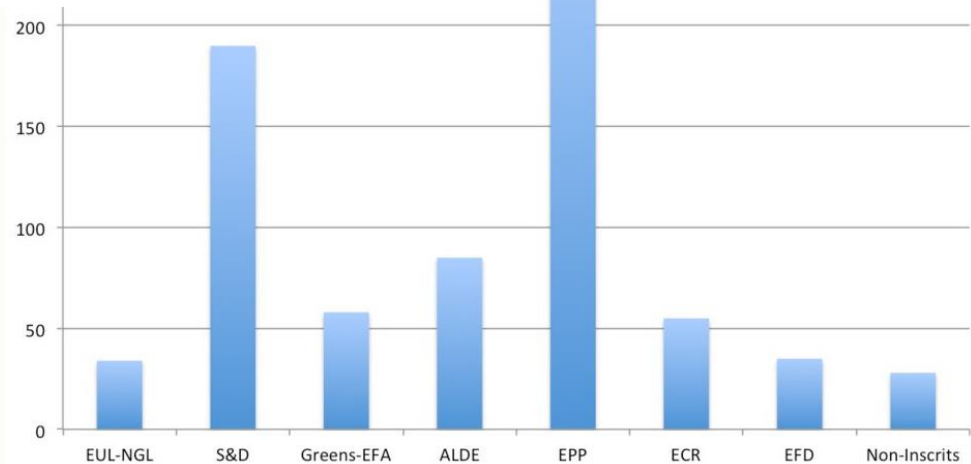
Gráficos de sector circular



Deceiving pie charts...



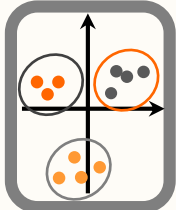
Government Party Breakdown



Pie charts, queijos ou gráficos de sectores circulares



The only good pie chart ever !



análise exploratória de dados

Gráficos de barras

Não é a parte de cima de uma caixa com bigodes!

Representa a media e o erro padrão associado!!

