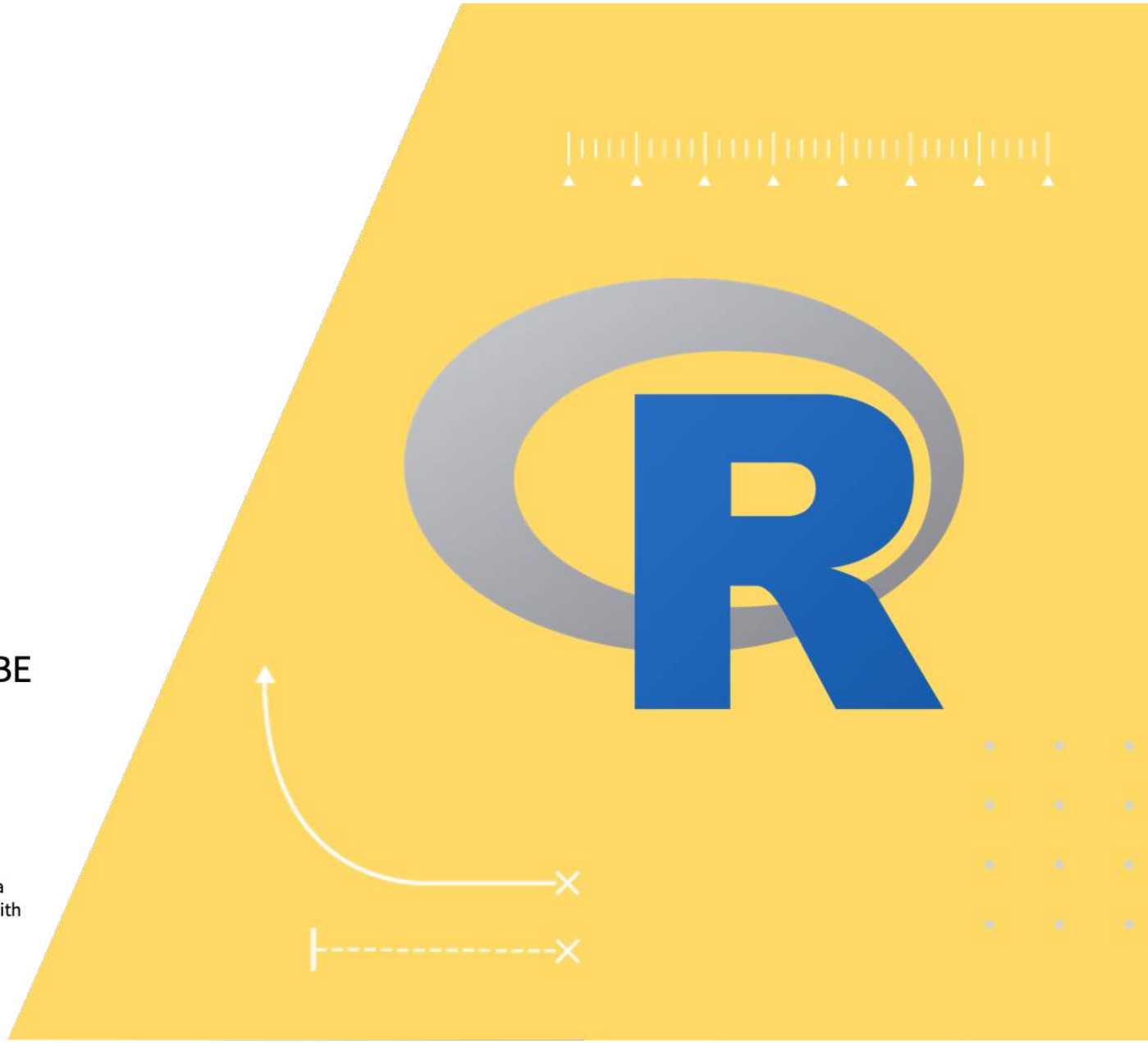
Preparado para: AT autoridade tributária e aduaneira

REFORM/SC2022/126
DELIVERABLE 4
MÓDULO 1
INTRODUÇÃO AO R

DESIGNING A NEW VALUATION MODEL
FOR RURAL PROPERTIES IN PORTUGAL

Formador: Luís Teles Morais | Nova SBE

Lisboa, 4 maio 2023



This project is carried out with funding by the European Union via the Structural Reform Support Programme and in cooperation with the Directorate General for Structural Reform Support of the European Commission

AARC NOVA SCHOOL OF BUSINESS & ECONOMICS LOBO VASQUES esri INNERLANDS

Programa

Módulos	Duração
Módulo 1 – Introdução ao R: - O que é o R? - Como instalar e configurar o R. - Sintaxe básica e comandos. - Tipos de dados, objetos e classes.	**4 Horas**
Módulo 2 – Gestão e tratamento de dados em R: - Carregar dados no R. - Perceber as estruturas de dados e *subsetting*. - Limpeza de dados: *missing values, outliers* e transformações - Juntar bases de dados	**8 Horas**
Módulo 3 – Estatística básica em R: - Estatísticas descritivas: medidas de dispersão central e variação. - Distribuições probabilísticas: variáveis discretas e contínuas. - Testes de hipóteses.	**8 Horas**

Módulos	Duração
Módulo 4 – Regressão Linear: - O modelo classico linear. - Estimação de parametros segundo o MMQ. - Testes de hipóteses: significância estatística e ajuste do modelo. - Modelo de regressão múltipla. - Testar as premissas: multicolinearidade, heteroscedasticidade e normalidade dos resíduos. - Critérios de seleção dos modelos.	**12 Horas**
Módulo 5 – O modelo: - Estrutura do modelo e premissas – Perceber o modelo (4 Hours). - Uso e tratamento dos dados (4 Hours). - Descrição do modelo (4 Hours). - Aplicação do modelo a cada piloto (12 Hours). - Aplicação autónoma do modelo a uma região (8 Hours).	**32 Horas**

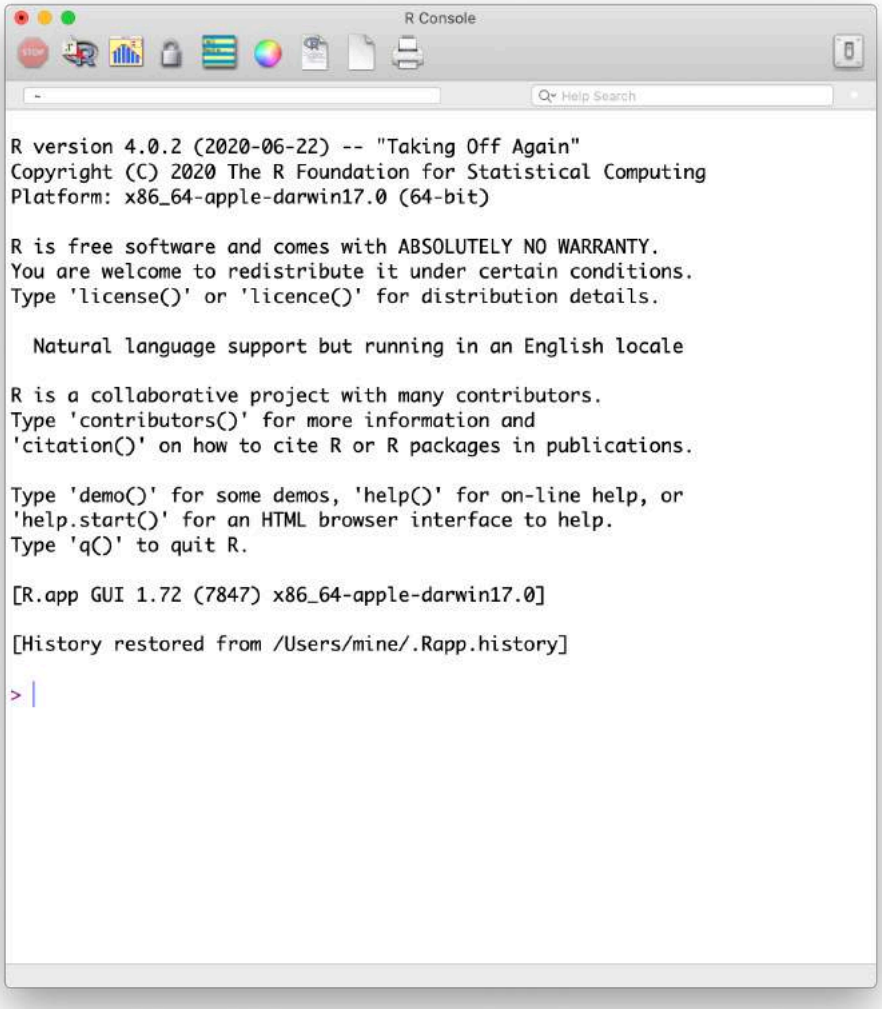
O que é o  ?

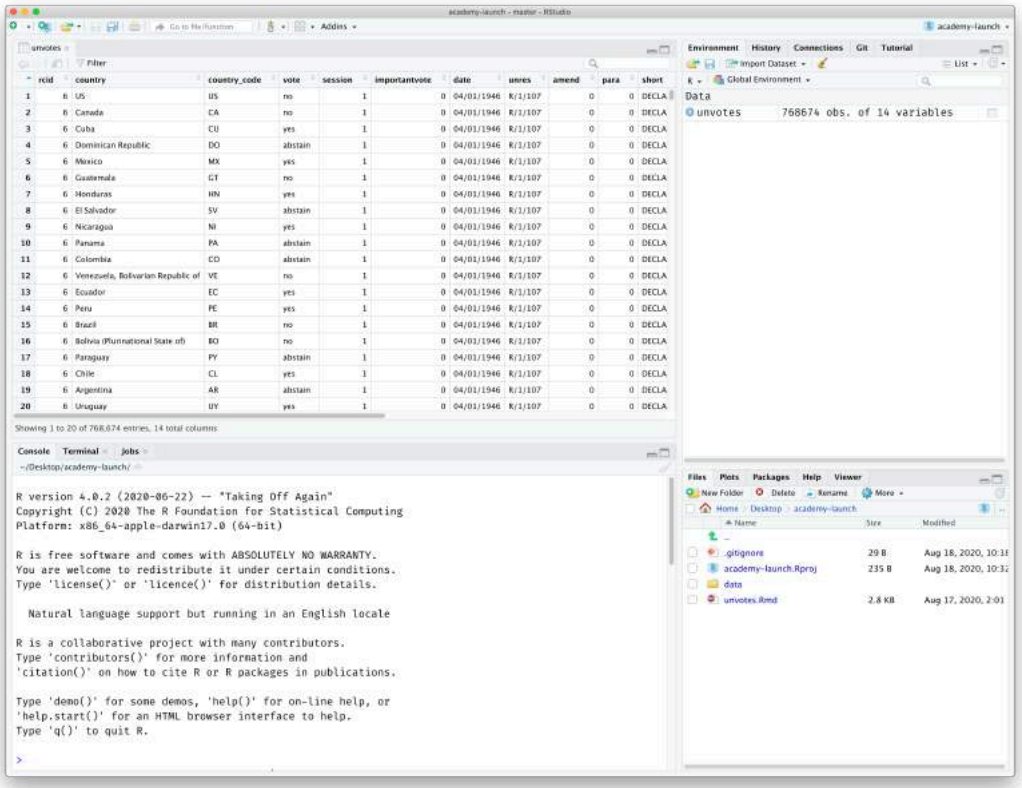
Linguagem de programação

Software de programação e edição (IDE, *Integrated Development Environment*)

Vamos escrever código R no RStudio

Linguagem de programação

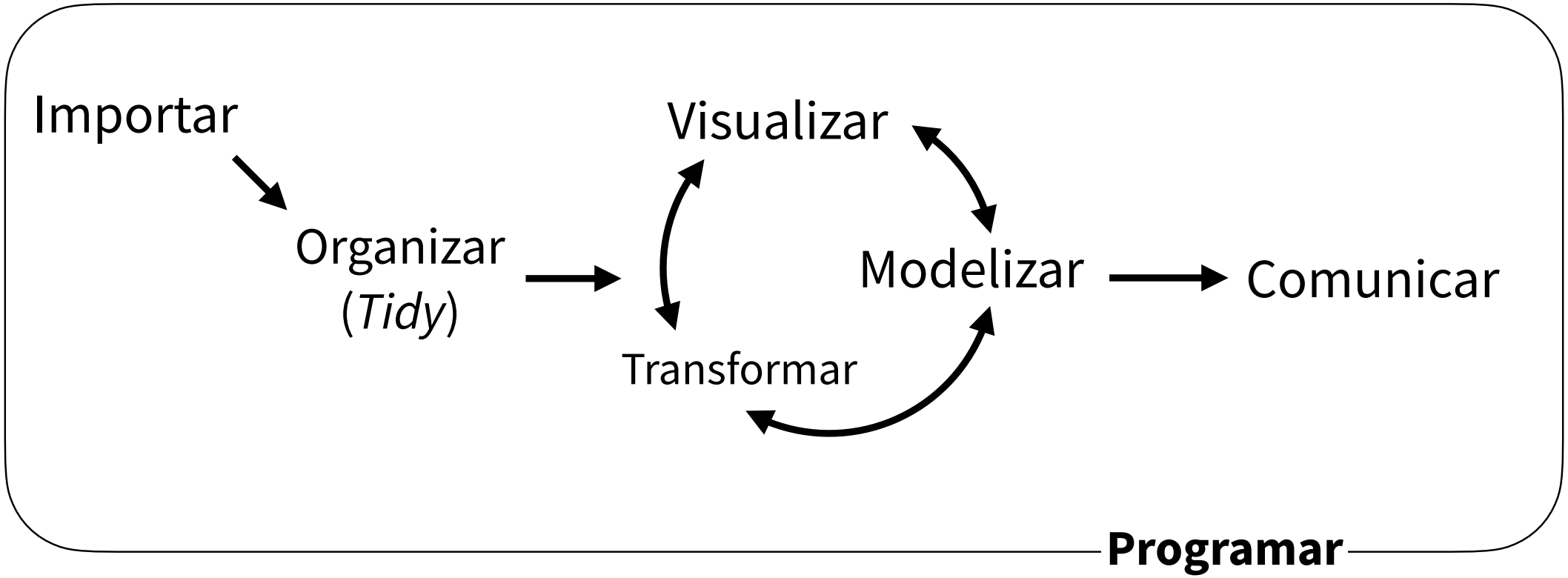




Software de programação e edição

Para que serve?

~~Análise estatística~~ Ciência de dados



Instalar e configurar

Instalação local de ambiente R

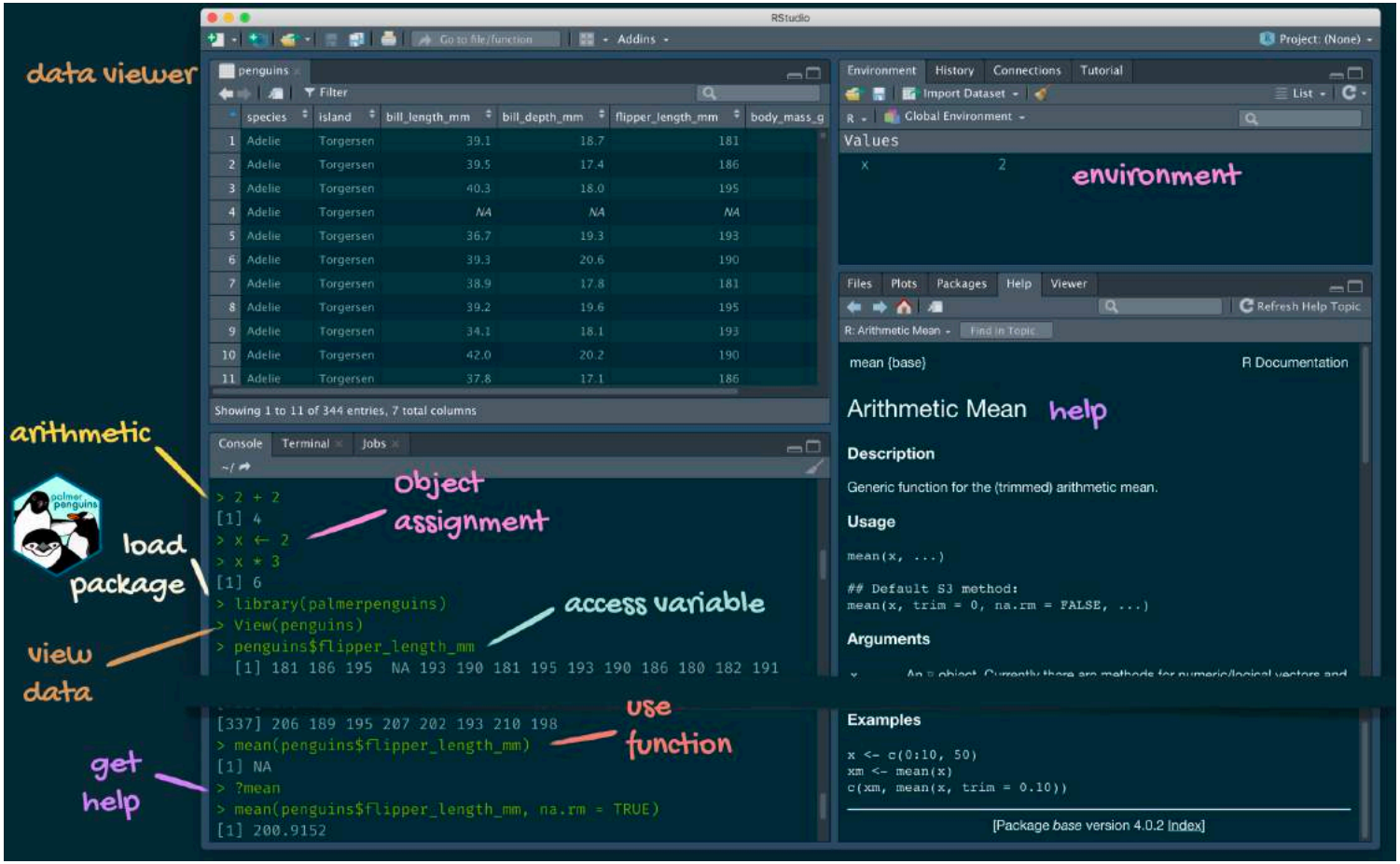
1. Versão recente do R (>=3.5.0)
   - cran.r-project.org
2. Versão recente do RStudio Desktop (>=1.1.456)
   - www.rstudio.com/download

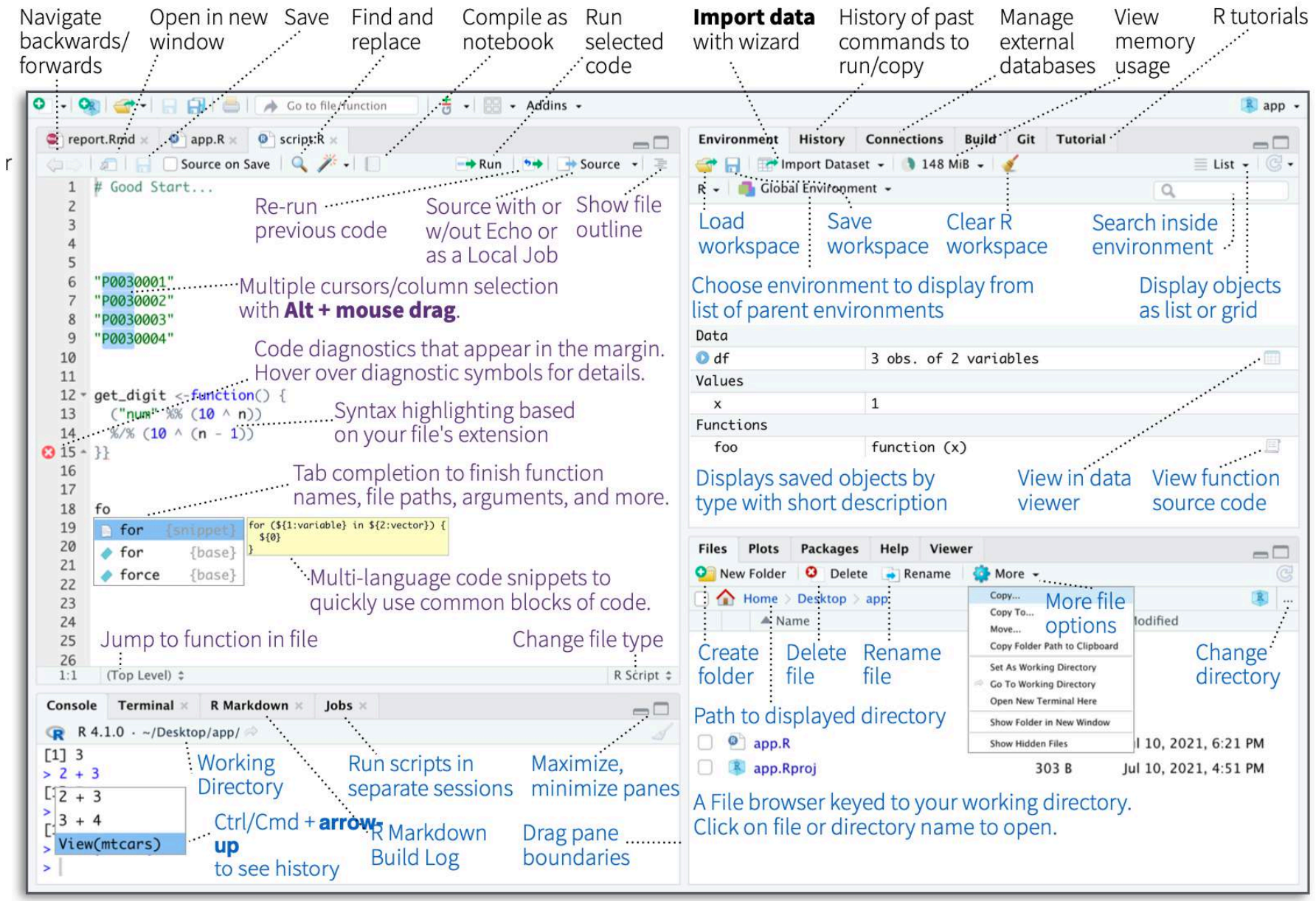
Vamos a isso

Aceda a este link para começar já

https://posit.cloud/content/5906356

O ambiente R Studio




Navigate backwards/ forwards
Open in new window
Save
Find and replace
Compile as notebook
Run selected code
Import data with wizard
History of past commands to run/copy
Manage external databases
View memory usage
R tutorials
Re-run previous code
Source with or w/out Echo or as a Local Job
Show file outline
Multiple cursors/column selection with Alt + mouse drag.
Code diagnostics that appear in the margin. Hover over diagnostic symbols for details.
Syntax highlighting based on your file's extension
Tab completion to finish function names, file paths, arguments, and more.
Multi-language code snippets to quickly use common blocks of code.
Jump to function in file
Change file type
Load workspace
Save workspace
Clear R workspace
Search inside environment
Choose environment to display from list of parent environments
Display objects as list or grid
Displays saved objects by type with short description
View in data viewer
View function source code
Create folder
Delete file
Rename file
More file options
Change directory
Path to displayed directory
A File browser keyed to your working directory. Click on file or directory name to open.
Working Directory
Run scripts in separate sessions
Maximize, minimize panes
Ctrl/Cmd + arrow-up to see history
R Markdown Build Log
Drag pane boundaries

Working directory

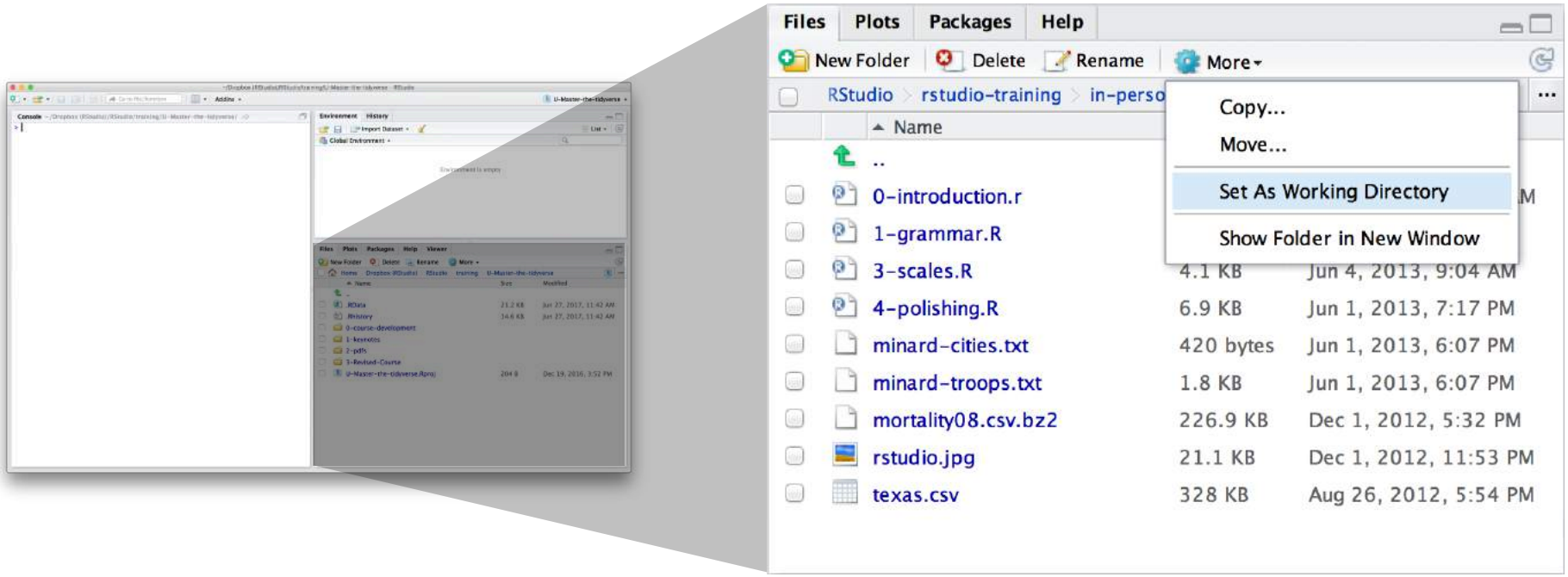
Uma sessão de R está sempre associada a uma pasta específica no computador (ou na *cloud*).

- > "working directory"
- Quaisquer ficheiros serão gravados pelo R nesta pasta
- Quando se pretenda carregar ficheiros, o R tentará encontrá-los nesta pasta

Alterar Working directory

Navegar para a pasta no painel **Files** e clicar:

More > Set As Working Directory



Sintaxe básica e comandos

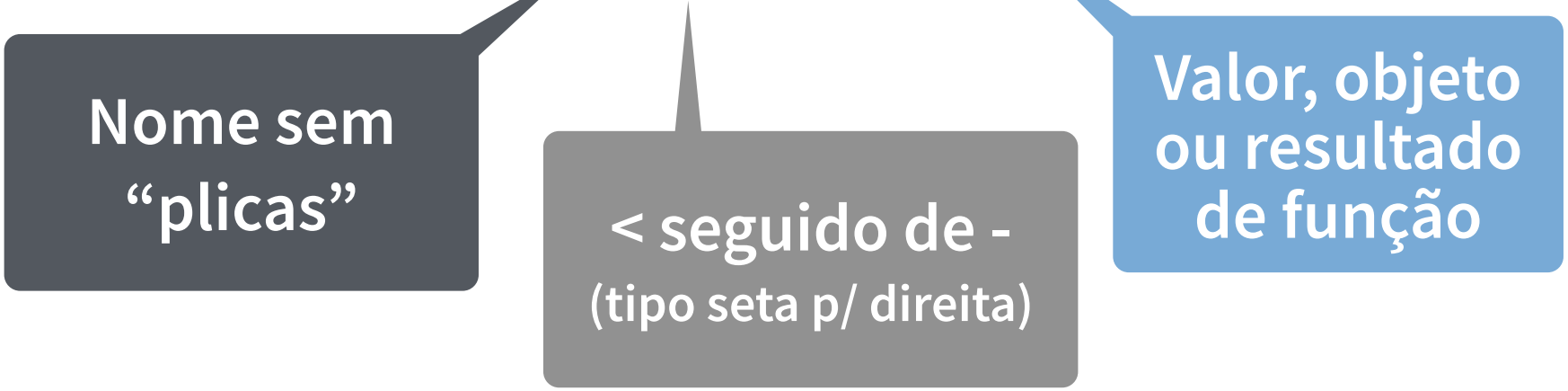
Básicos da linguagem R

Valores - `1, "Florida", "2010-01-25"`

Básicos da linguagem R

Valores - `1, "Florida", "2010-01-25"`

Objectos - `x <- 22/7`



Básicos da linguagem R

Valores - `1, "Florida", "2010-01-25"`

Objectos - `x <- c(22/7, 0.99, 3)`

c() é um objeto composto por vários valores (um vetor)

Básicos da linguagem R

Valores - `1, "Florida", "2010-01-25"`

Objectos - `x <- c(22/7, 0.99, 3)`

Funções - `round(x, digits = 3)`

Nome sem "plicas"

seguido de () para *correr* a função

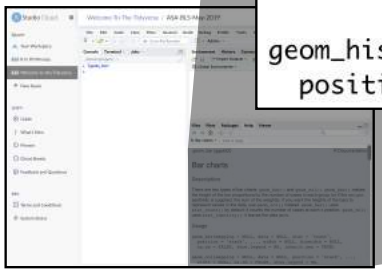
Argumentos da função: valores ou objectos + *opções* da função

A função mais importante em R

A função mais importante em R

```
?geom_freqpoly
```

? — **nome da função** (sem parênteses)



geom_freqpoly {ggplot2} R Documentation

Histograms and frequency polygons

Description

Visualise the distribution of a single continuous variable by dividing the x axis into bins and counting the number of observations in each bin. Histograms (geom_histogram()) display the counts with bars; frequency polygons (geom_freqpoly()) display the counts with lines. Frequency polygons are more suitable when you want to compare the distribution across the levels of a categorical variable.

Usage

```
geom_freqpoly(mapping = NULL, data = NULL, stat = "bin",
  position = "identity", ..., na.rm = FALSE, show.legend = NA,
  inherit.aes = TRUE)

geom_histogram(mapping = NULL, data = NULL, stat = "bin",
  position = "stack", ..., binwidth = NULL, bins = NULL,
```

Experimente

Porque é que este código não funciona?

```
num_golos <= 835
num_golos
```

Porque é que este código não funciona?

```
num_golos <- 835
num_golos
```

Tipos de dados em R

1. `numeric` (10.5, 55, 787)
2. `character` ou “string” ("ronaldo", "FALSE", "11.5")
3. `logical` ou *boolean* - (TRUE ou FALSE)
4. `integer` (1L, 55L, 100L)
5. `complex` (9 + 3i)

Só para confirmar

Quais dos seguintes objetos são **números**?

`1` `"1"` `“um"` um

Quais dos seguintes objetos são **números**?

1 "1" "one" one

números

Quais dos seguintes objetos são **números**?

1	"1"	"one"	one
número	**palavras (*strings*)**		

Quais dos seguintes objetos são **números**?

1	"1"	"one"	one
número	**palavras (strings)**		**objeto**

Experimente

Quais destes funcionam, depois de executar:
um <- 1

```
1 * 2
```

```
“1” * 2
```

```
“um” * 2
```

```
um * 2
```

Quais destes funcionam, depois de executar:

```
um <- 1
```

```
1 * 2
```

```
“1” * 2
```

```
“um” * 2
```

```
um * 2
```

O poder da consola

```
num_golos <- 835
```

```
1 / 200 * 30
#> [1] 0.15
```

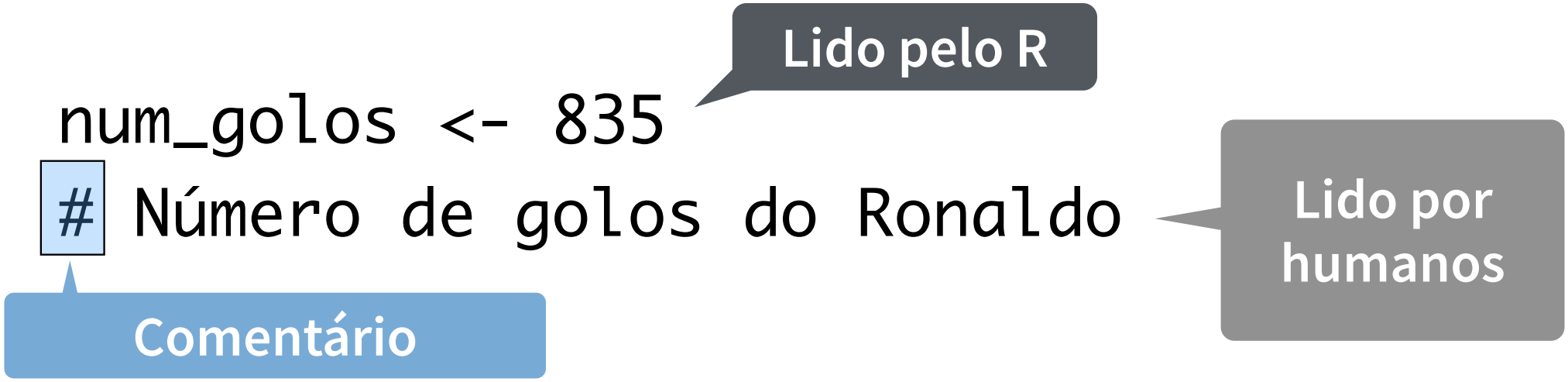
```
num_golos / 3
#> [1] 44.66667
```

```
log(2 ** 3)
#> [1] 1.791759
```

```
num_golos > 1000
#> [1] FALSE
```

```
num_golos <= 835
#> [1] TRUE
```

Lido por humanos



Lido por humanos

```
num_golos <- 835
# Número de golos do Ronaldo
```

Boas práticas

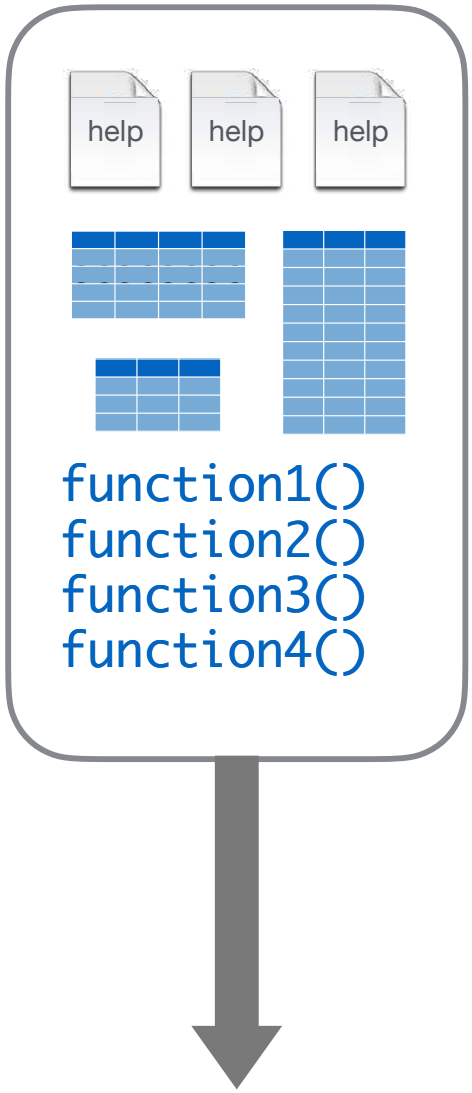
Lido por humanos

```
num_golos <- 835
# Número de golos do Por
```

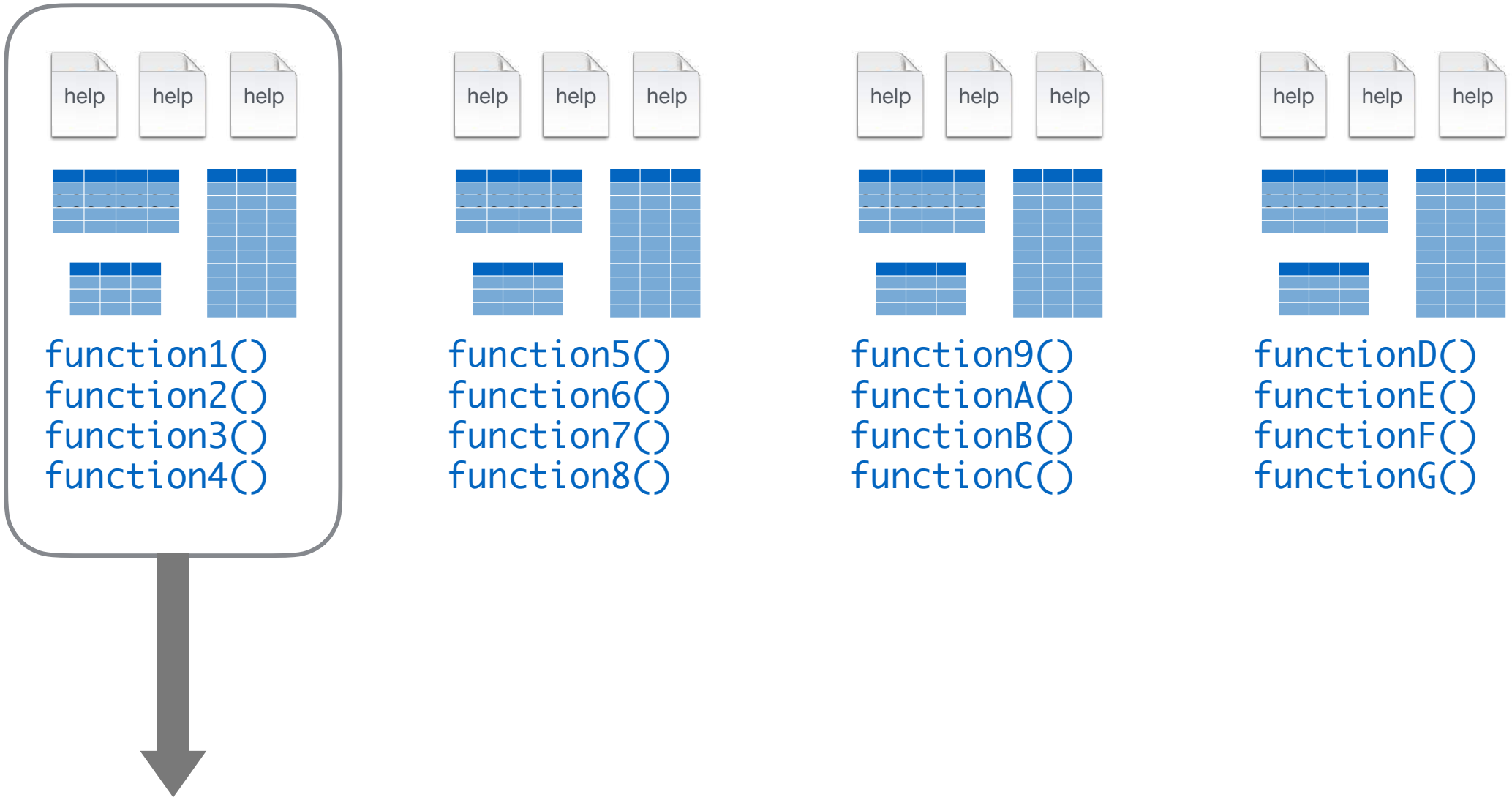
Boas práticas

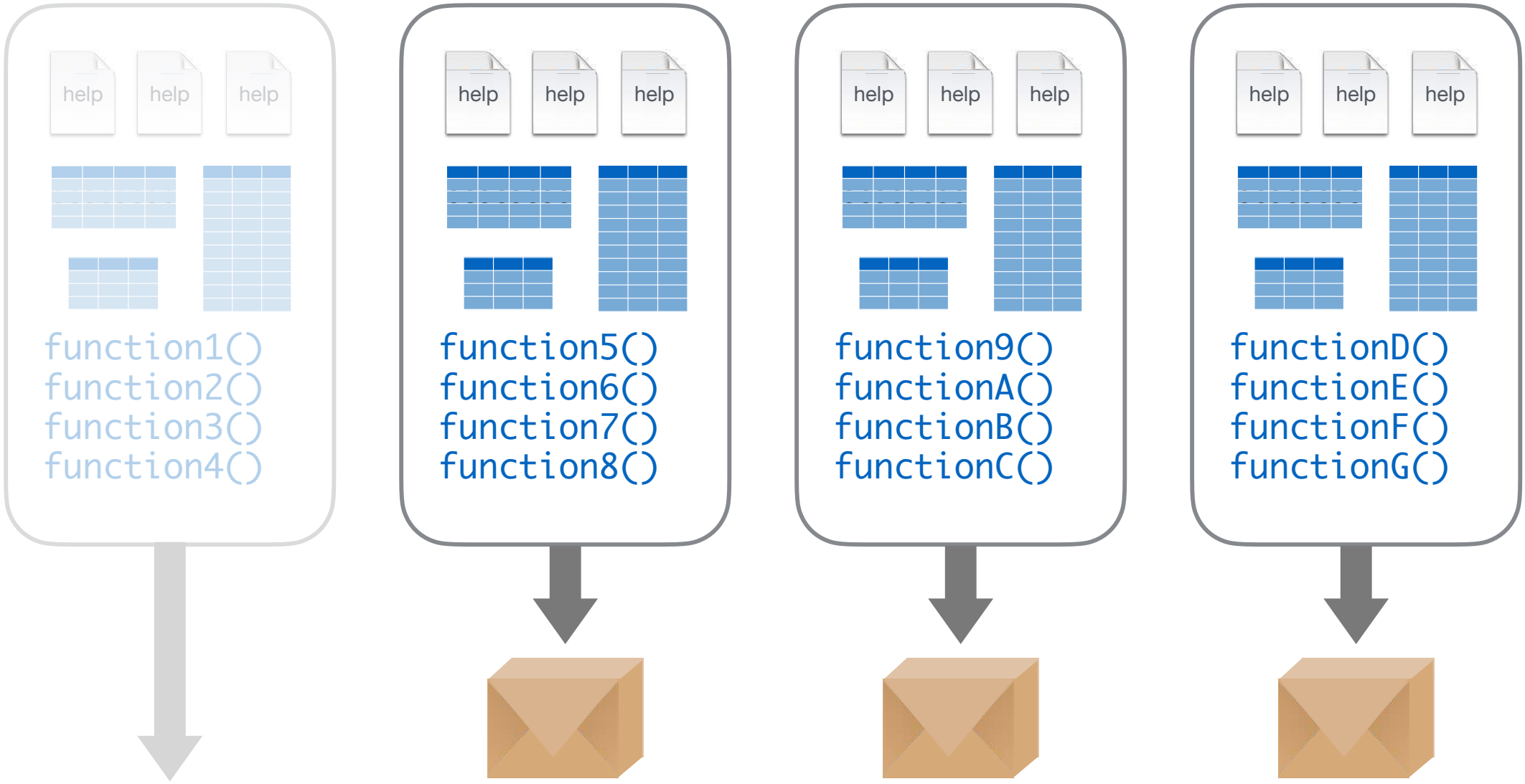
Más práticas:

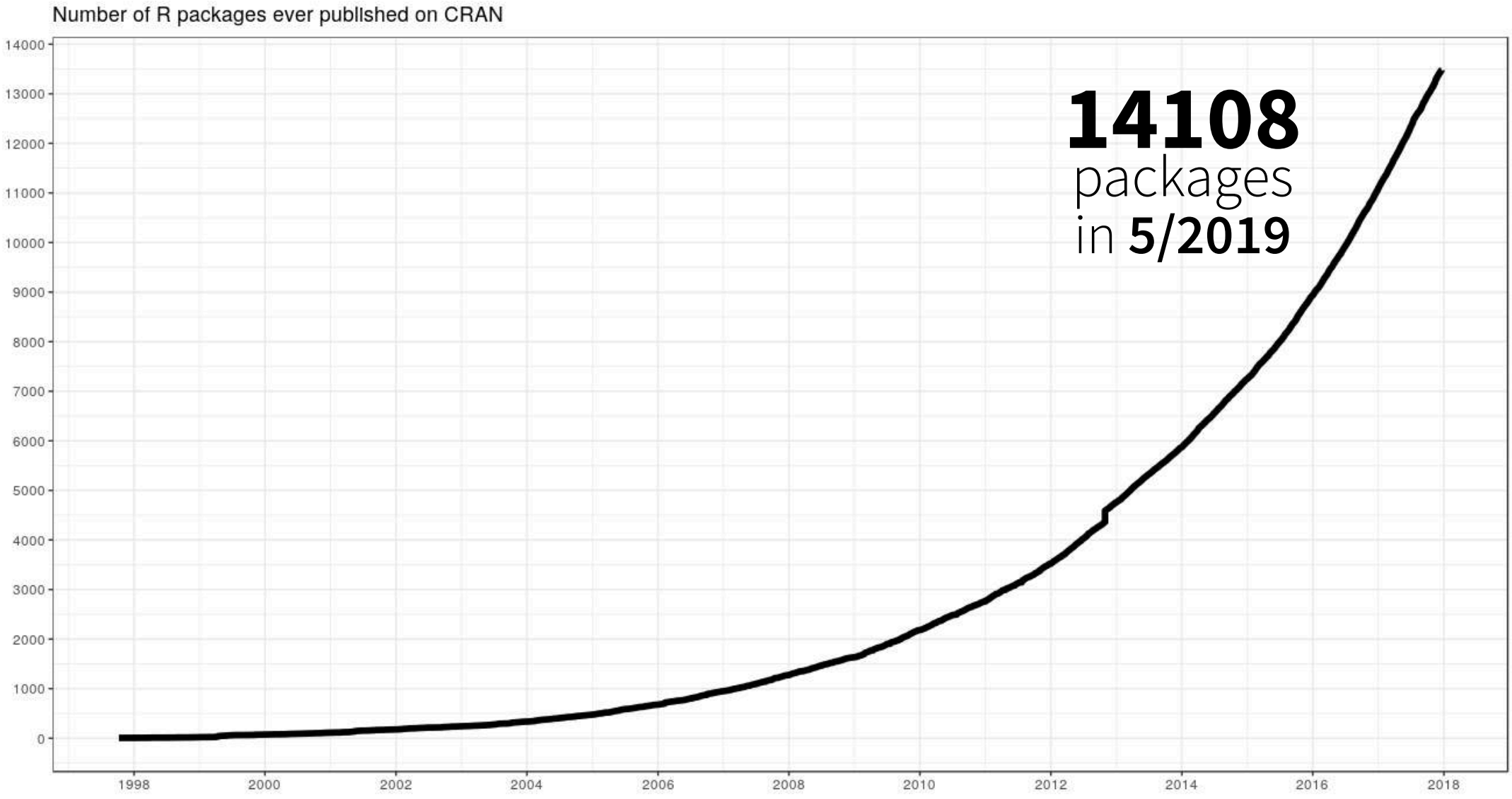
```
NúmeroDeGolos
Num.Gls
```



Base R


help
help
help
function1()
function2()
function3()
function4()
help
help
help
function5()
function6()
function7()
function8()
help
help
help
function9()
functionA()
functionB()
functionC()
help
help
help
functionD()
functionE()
functionF()
functionG()
Base R


help
help
help
function1()
function2()
function3()
function4()
help
help
help
function5()
function6()
function7()
function8()
help
help
help
function9()
functionA()
functionB()
functionC()
help
help
help
functionD()
functionE()
functionF()
functionG()
Base R
Pacotes R (**Packages**)



Graph by https://besteamperu.org

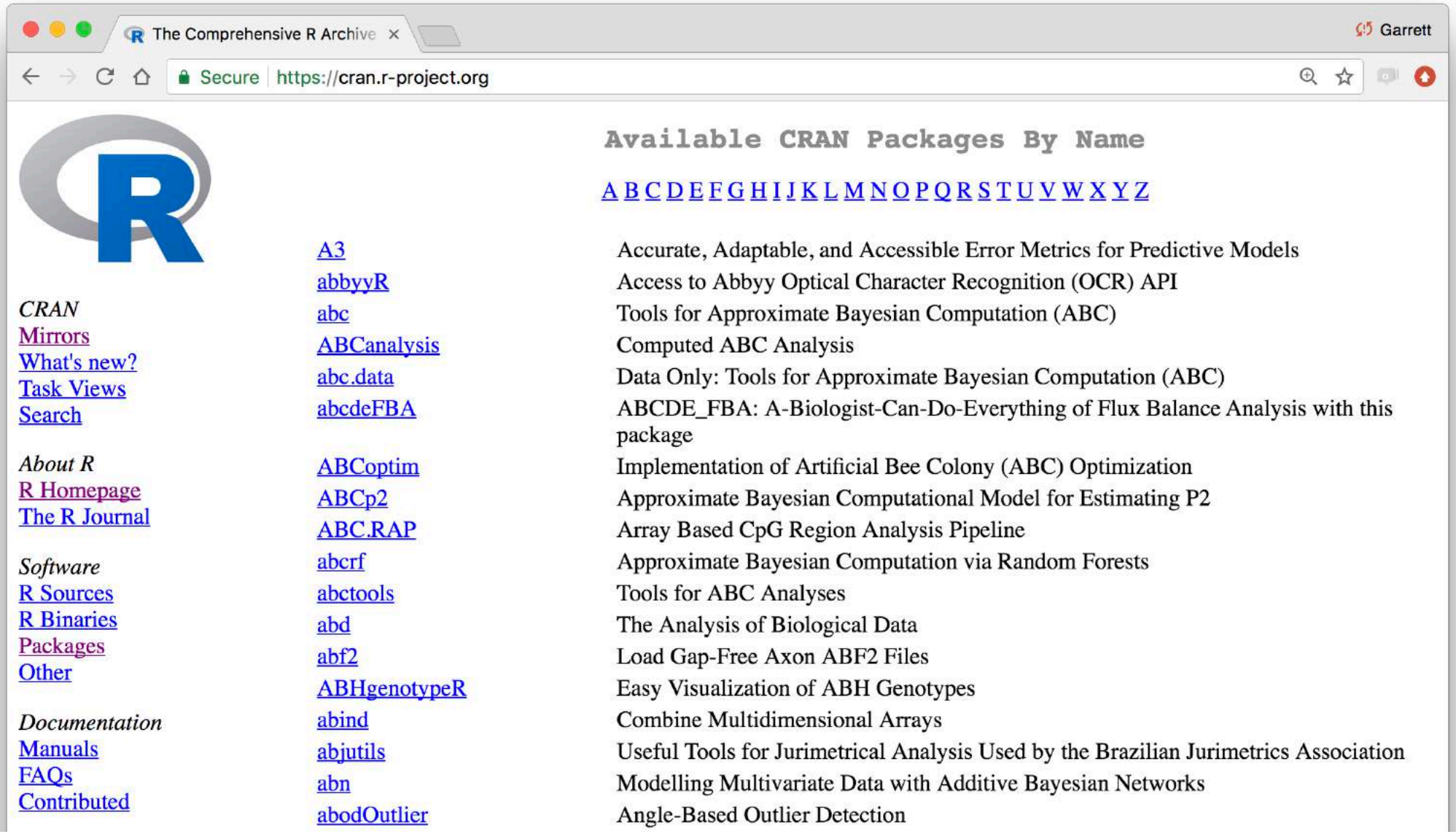
Como usar pacotes R

1

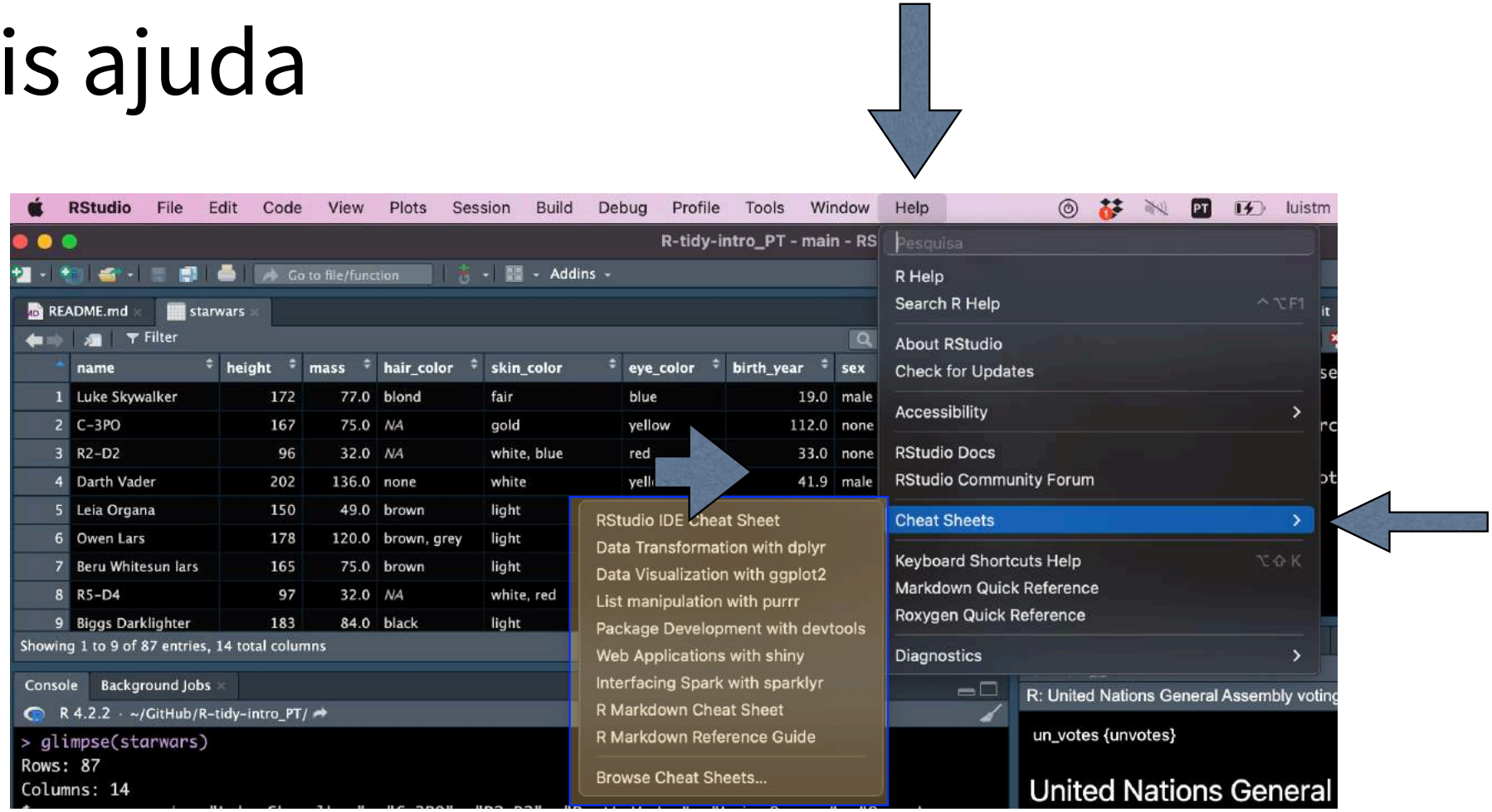
```
install.packages("foo")
```

Descarrega e instala pacotes

1 x por máquina

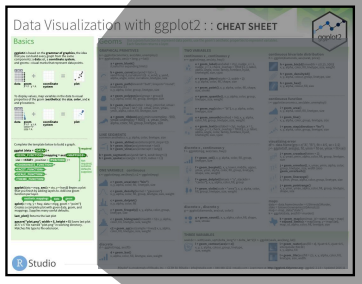

The Comprehensive R Archive
Garrett
Secure
https://cran.r-project.org
Available CRAN Packages By Name
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
CRAN
Mirrors
What's new?
Task Views
Search
About R
R Homepage
The R Journal
Software
R Sources
R Binaries
Packages
Other
Documentation
Manuals
FAQs
Contributed
A3 Accurate, Adaptable, and Accessible Error Metrics for Predictive Models
abbyyR Access to Abbyy Optical Character Recognition (OCR) API
abc Tools for Approximate Bayesian Computation (ABC)
ABCanalysis Computed ABC Analysis
abc.data Data Only: Tools for Approximate Bayesian Computation (ABC)
abcdeFBA ABCDE_FBA: A-Biologist-Can-Do-Everything of Flux Balance Analysis with this package
ABCoptim Implementation of Artificial Bee Colony (ABC) Optimization
ABCp2 Approximate Bayesian Computational Model for Estimating P2
ABC.RAP Array Based CpG Region Analysis Pipeline
abcrf Approximate Bayesian Computation via Random Forests
abctools Tools for ABC Analyses
abd The Analysis of Biological Data
abf2 Load Gap-Free Axon ABF2 Files
ABHgenotypeR Easy Visualization of ABH Genotypes
abind Combine Multidimensional Arrays
abjutils Useful Tools for Jurimetrical Analysis Used by the Brazilian Jurimetrics Association
abn Modelling Multivariate Data with Additive Bayesian Networks
abodOutlier Angle-Based Outlier Detection

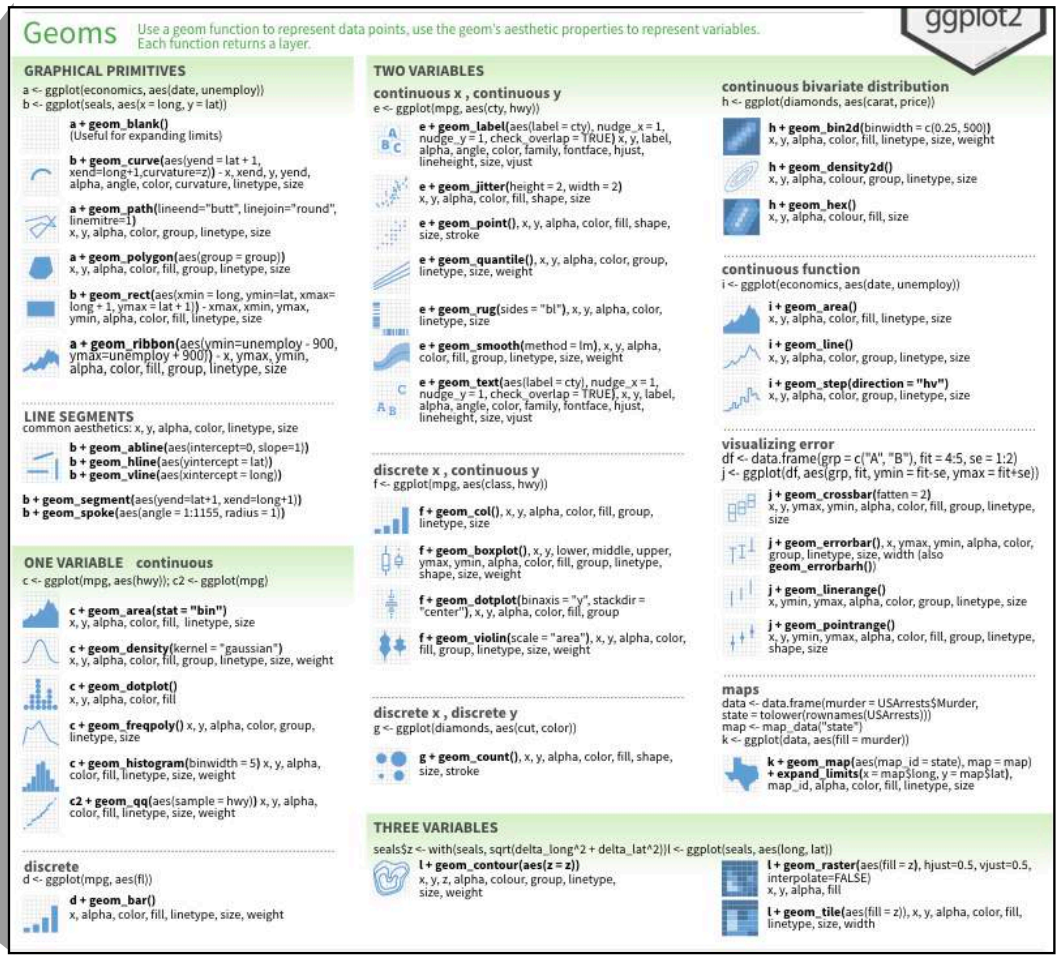
Mais ajuda


RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help
R-tidy-intro_PT - main - RS
Go to file/function
Addins
README.md
starwars
Filter
name height mass hair_color skin_color eye_color birth_year sex
1 Luke Skywalker 172 77.0 blond fair blue 19.0 male
2 C-3PO 167 75.0 NA gold yellow 112.0 none
3 R2-D2 96 32.0 NA white, blue red 33.0 none
4 Darth Vader 202 136.0 none white 41.9 male
5 Leia Organa 150 49.0 brown light
6 Owen Lars 178 120.0 brown, grey light
7 Beru Whitesun lars 165 75.0 brown light
8 R5-D4 97 32.0 NA white, red
9 Biggs Darklighter 183 84.0 black light
Showing 1 to 9 of 87 entries, 14 total columns
Console Background Jobs
R 4.2.2 · ~/GitHub/R-tidy-intro_PT/
> glimpse(starwars)
Rows: 87
Columns: 14
Pesquisa
R Help
Search R Help
About RStudio
Check for Updates
Accessibility
RStudio Docs
RStudio Community Forum
Cheat Sheets
Keyboard Shortcuts Help
Markdown Quick Reference
Roxygen Quick Reference
Diagnostics
RStudio IDE Cheat Sheet
Data Transformation with dplyr
Data Visualization with ggplot2
List manipulation with purrr
Package Development with devtools
Web Applications with shiny
Interfacing Spark with sparklyr
R Markdown Cheat Sheet
R Markdown Reference Guide
Browse Cheat Sheets...
R: United Nations General Assembly voting
un_votes {unvotes}
United Nations General

Sem batota

Cheat sheets



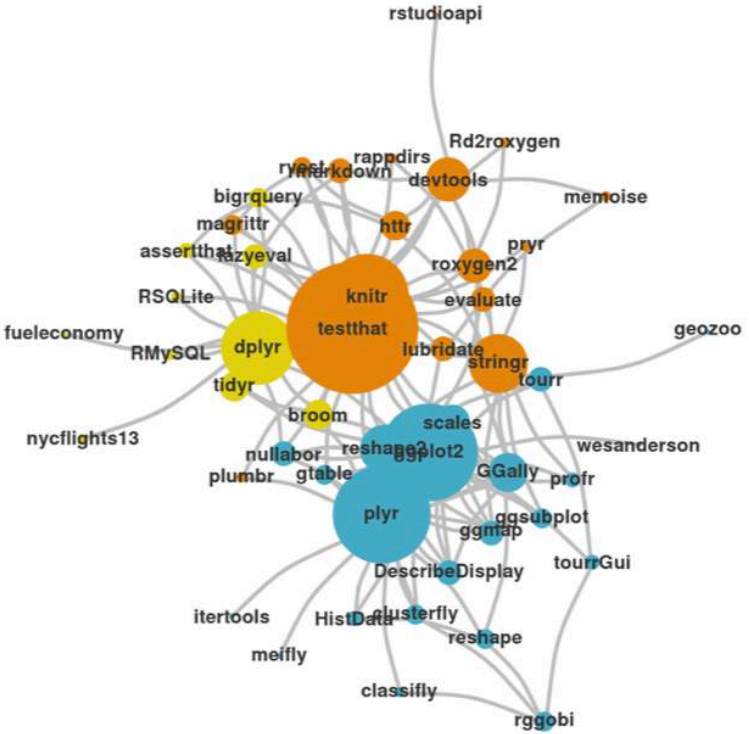



tidyverse

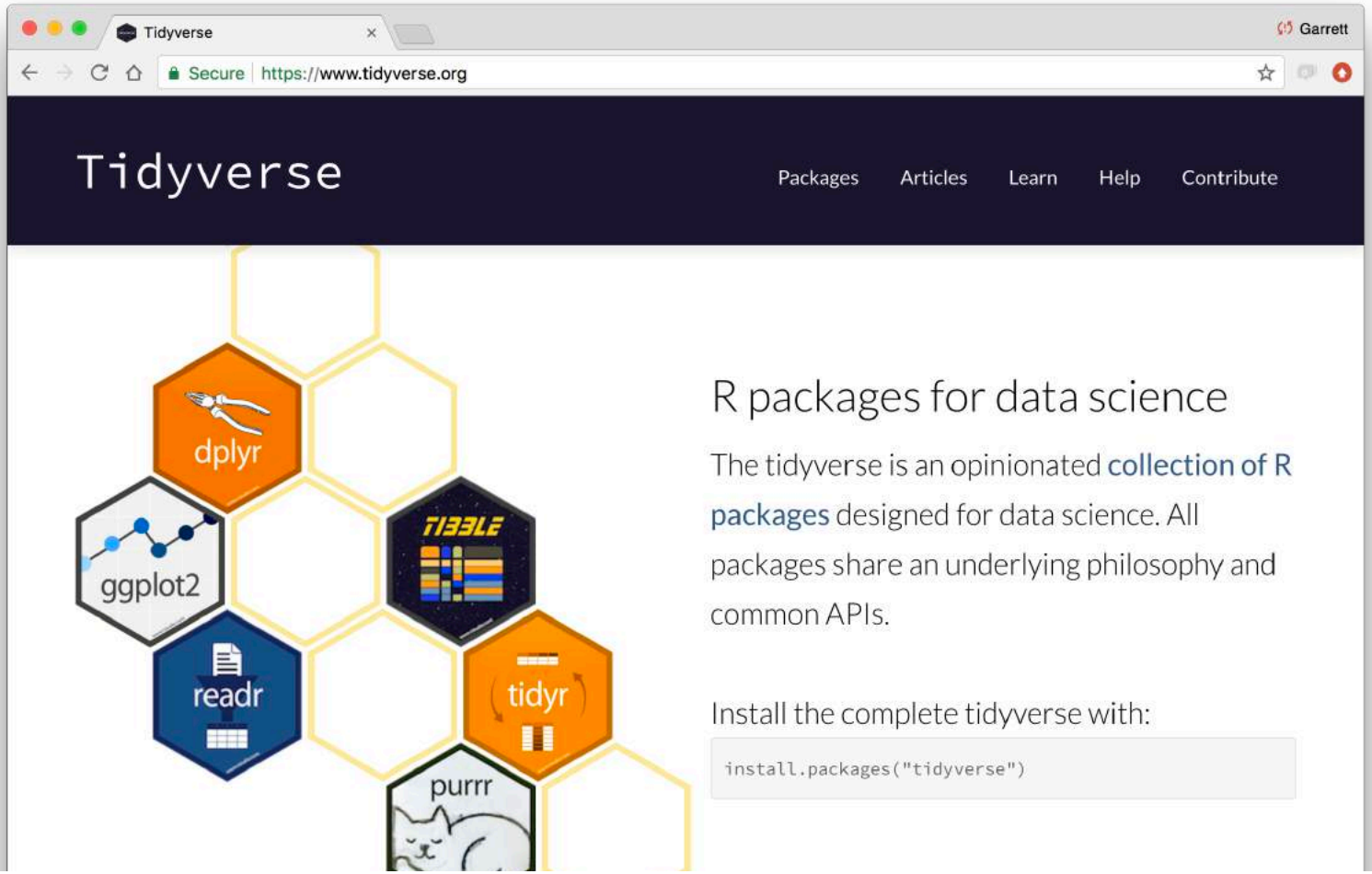
Bem-vindos ao Tidyverse

Bem-vindos ao Tidyverse

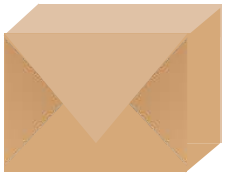
- Uma coleção de pacotes R modernos, com uma filosofia comum e que funcionam em conjunto
- **Tudo o que precisa para trabalhar com dados em R**



tidyverse.org



tidyverse



Um pacote que serve de atalho para instalar e carregar todos os componentes do **tidyverse**

```
install.packages("tidyverse")
```

```
install.packages("tidyverse")
```

equivale a

```
install.packages("ggplot2")
install.packages("dplyr")
install.packages("tidyr")
install.packages("readr")
install.packages("purrr")
install.packages("tibble")
install.packages("hms")
install.packages("stringr")
install.packages("lubridate")
install.packages("forcats")
install.packages("DBI")
install.packages("haven")
install.packages("httr")
install.packages("jsonlite")
install.packages("readxl")
install.packages("rvest")
install.packages("xml2")
install.packages("modelr")
install.packages("broom")
```


Já instalado no RStudio Cloud

Como usar pacotes R

1

```
install.packages("foo")
```

Descarrega e instala pacotes

1 x por máquina

2

```
library("foo")
```

Carrega pacote

1 x por sessão de R

```
install.packages("tidyverse")
```

equivale a

```
install.packages("ggplot2")
install.packages("dplyr")
install.packages("tidyr")
install.packages("readr")
install.packages("purrr")
install.packages("tibble")
install.packages("stringr")
install.packages("forcats")
install.packages("lubridate")
install.packages("hms")
install.packages("DBI")
install.packages("haven")
install.packages("httr")
install.packages("jsonlite")
install.packages("readxl")
install.packages("rvest")
install.packages("xml2")
install.packages("modelr")
install.packages("broom")
```

```
library("tidyverse")
```

equivale a

```
library("ggplot2")
library("dplyr")
library("tidyr")
library("readr")
library("purrr")
library("tibble")
library("stringr")
library("forcats")
```

O que entendemos por dados?

Dados

infopédia
DICIONÁRIOS PORTO EDITORA

nome masculino

1. cada um dos elementos conhecidos de um problema
2. base para a formação de um juízo ou cálculo
3. informação
4. INFORMÁTICA informação que pode ser aceite, armazenada, tratada ou fornecida pelo computador

Dados ⟶ Dataset

(Conjunto de dados)

- Cada coluna é uma **variável**
- Cada linha é uma **observação**

Dados ⟶ Dataset

(Conjunto de dados)

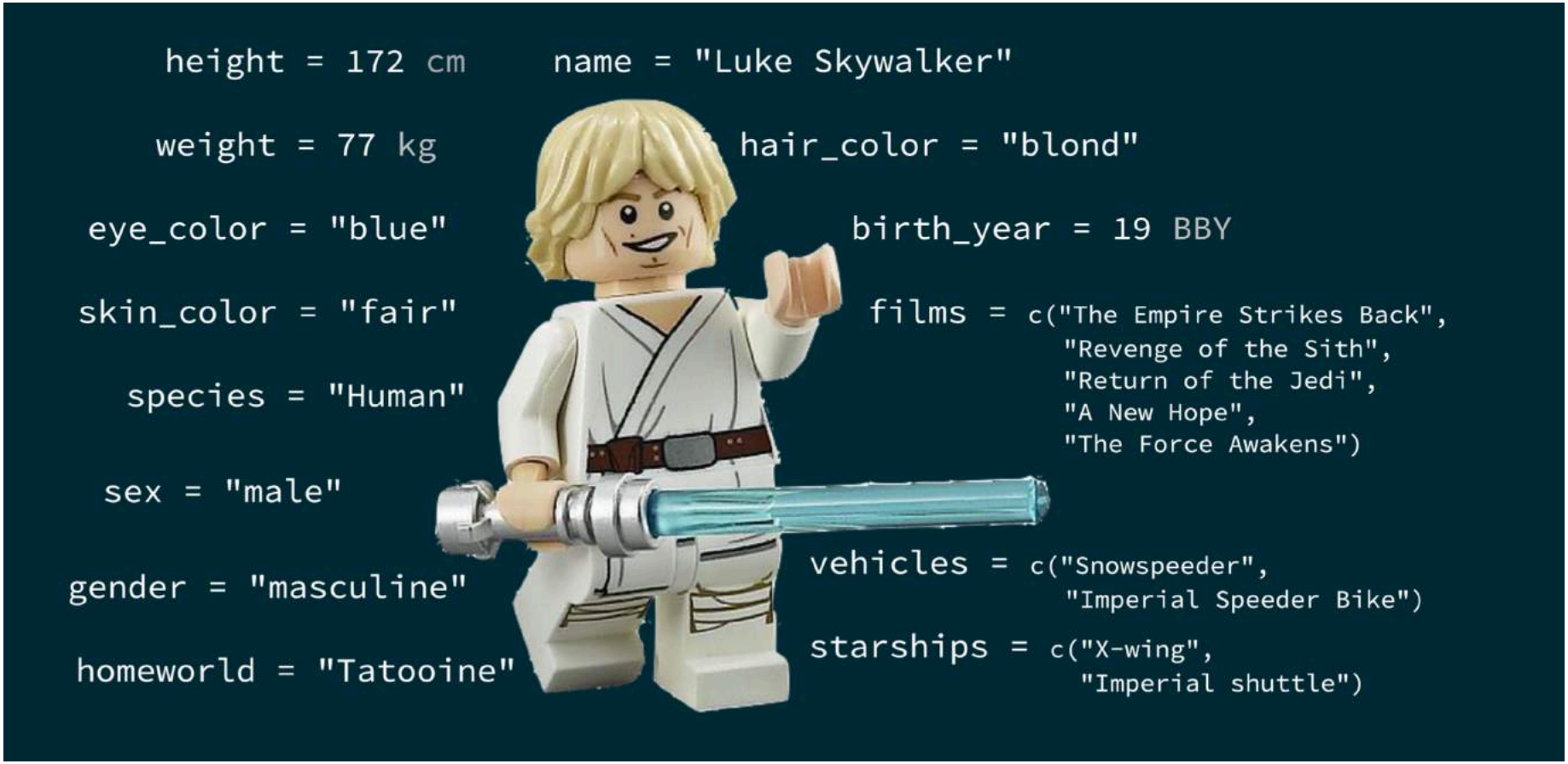
- Cada coluna é uma **variável**
- Cada linha é uma **observação**

Ex.: `starwars`

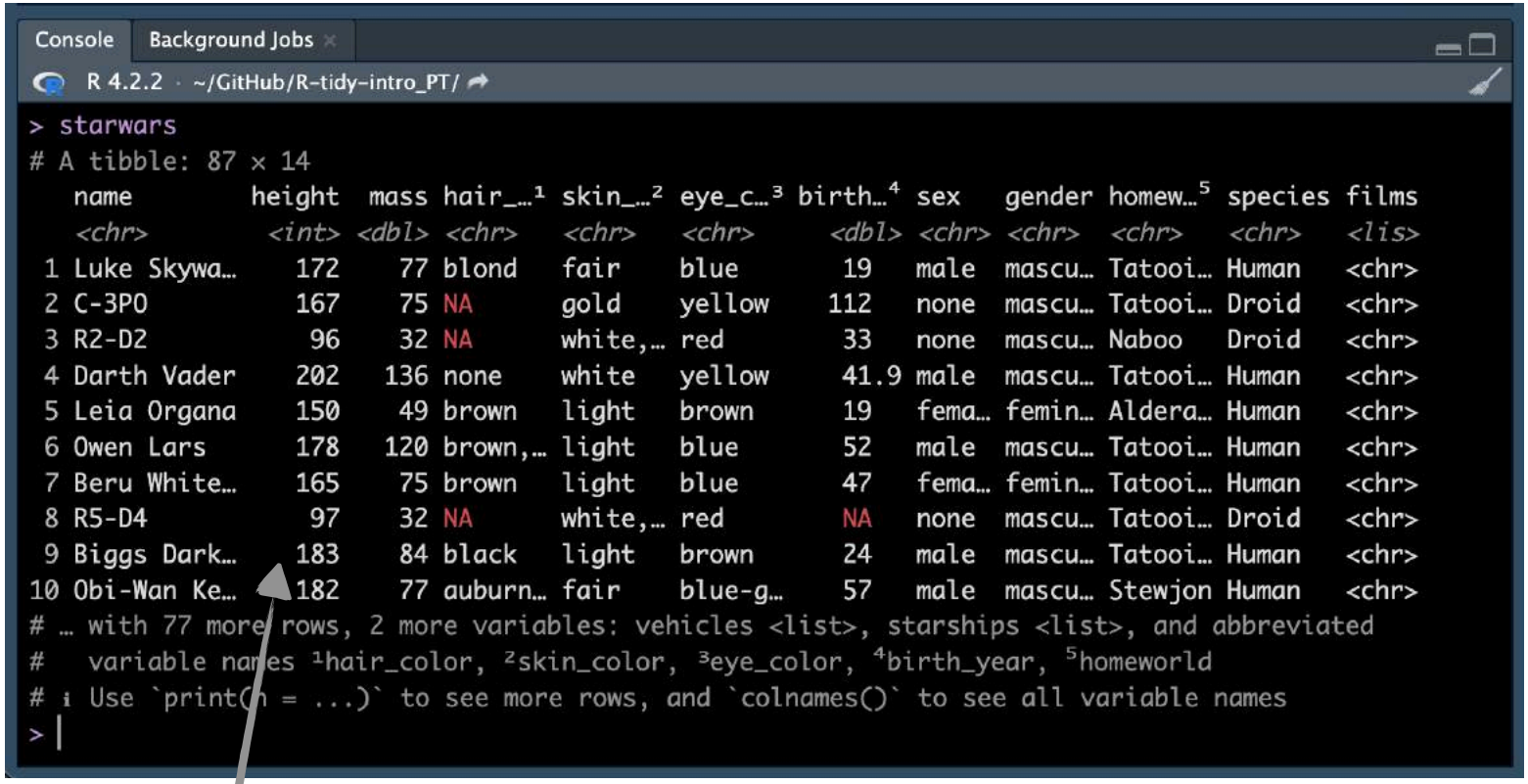
```
## # A tibble: 87 × 14
##   name    height  mass hair_color skin_color eye_color birth_year
##   <chr>    <int> <dbl> <chr>      <chr>      <chr>           <dbl>
## 1 Luke S…    172    77 blond      fair       blue             19
## 2 C-3PO      167    75 <NA>       gold       yellow          112
## 3 R2-D2       96    32 <NA>       white, bl… red              33
## 4 Darth …    202   136 none       white      yellow           41.9
## 5 Leia O…    150    49 brown      light      brown            19
## 6 Owen L…    178   120 brown, gr… light      blue             52
## # … with 81 more rows, and 7 more variables: sex <chr>,
## #   gender <chr>, homeworld <chr>, species <chr>, films <list>,
## #   vehicles <list>, starships <list>
```

- R: data.frame ou tibble

- Uma **observação**

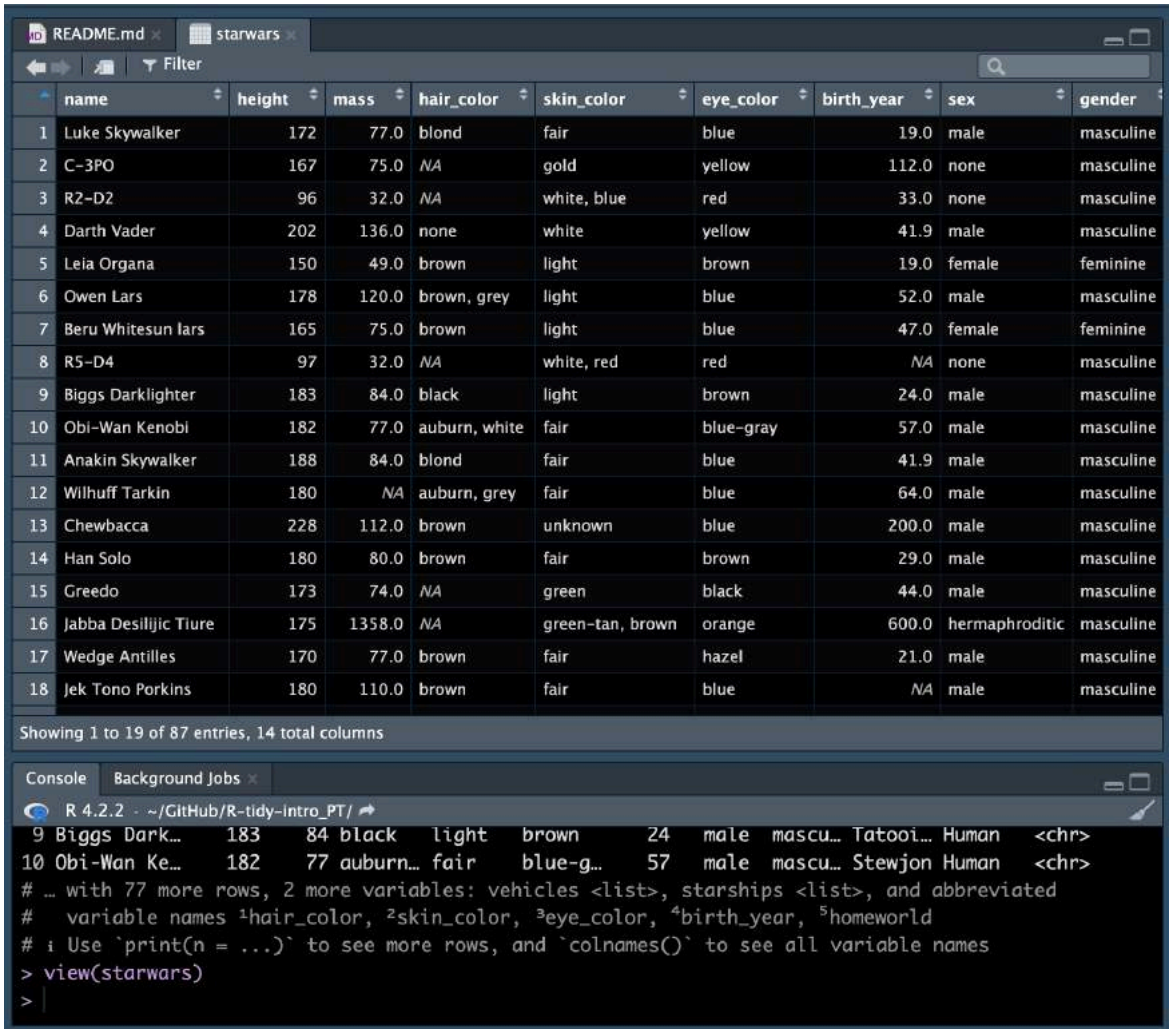


- Dar uma olhadela: `starwars`



- Uma **variável** `starwars$height`

- Dar uma olhadela: `view(starwars)`



README.md | starwars

Filter

	name	height	mass	hair_color	skin_color	eye_color	birth_year	sex	gender
1	Luke Skywalker	172	77.0	blond	fair	blue	19.0	male	masculine
2	C-3PO	167	75.0	*NA*	gold	yellow	112.0	none	masculine
3	R2-D2	96	32.0	*NA*	white, blue	red	33.0	none	masculine
4	Darth Vader	202	136.0	none	white	yellow	41.9	male	masculine
5	Leia Organa	150	49.0	brown	light	brown	19.0	female	feminine
6	Owen Lars	178	120.0	brown, grey	light	blue	52.0	male	masculine
7	Beru Whitesun lars	165	75.0	brown	light	blue	47.0	female	feminine
8	R5-D4	97	32.0	*NA*	white, red	red	*NA*	none	masculine
9	Biggs Darklighter	183	84.0	black	light	brown	24.0	male	masculine
10	Obi-Wan Kenobi	182	77.0	auburn, white	fair	blue-gray	57.0	male	masculine
11	Anakin Skywalker	188	84.0	blond	fair	blue	41.9	male	masculine
12	Wilhuff Tarkin	180	*NA*	auburn, grey	fair	blue	64.0	male	masculine
13	Chewbacca	228	112.0	brown	unknown	blue	200.0	male	masculine
14	Han Solo	180	80.0	brown	fair	brown	29.0	male	masculine
15	Greedo	173	74.0	*NA*	green	black	44.0	male	masculine
16	Jabba Desilijic Tiure	175	1358.0	*NA*	green-tan, brown	orange	600.0	hermaphroditic	masculine
17	Wedge Antilles	170	77.0	brown	fair	hazel	21.0	male	masculine
18	Jek Tono Porkins	180	110.0	brown	fair	blue	*NA*	male	masculine

Showing 1 to 19 of 87 entries, 14 total columns

Console | Background Jobs

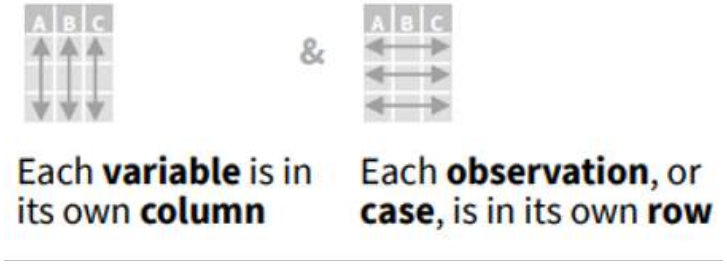
R 4.2.2 · ~/GitHub/R-tidy-intro_PT/

```
 9 Biggs Dark…    183    84 black   light   brown      24   male  mascu… Tatooi… Human   <chr>
10 Obi-Wan Ke…    182    77 auburn… fair    blue-g…    57   male  mascu… Stewjon Human   <chr>
# … with 77 more rows, 2 more variables: vehicles <list>, starships <list>, and abbreviated
#   variable names ¹hair_color, ²skin_color, ³eye_color, ⁴birth_year, ⁵homeworld
# ℹ Use `print(n = ...)` to see more rows, and `colnames()` to see all variable names
> view(starwars)
>
```

Dados ⟶ Dataset

(Conjunto de dados)

- Cada coluna é uma **variável**
- Cada linha é uma **observação**

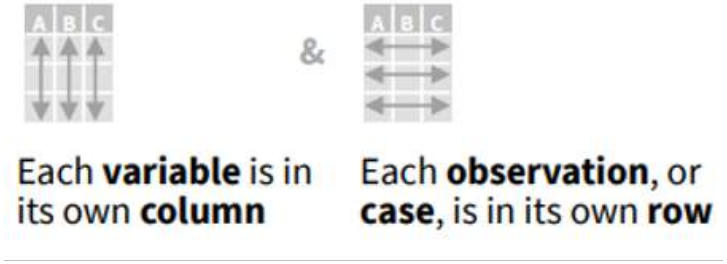


Sempre?

Dados ⟶ Dataset

(Conjunto de dados)

- Cada coluna é uma **variável**
- Cada linha é uma **observação**



Sempre?

Nem sempre. Só com dados bem arrumados (***tidy***)

Exemplos de dados não-tidy

TABELAS DE RETENÇÃO NA FONTE PARA O CONTINENTE - 2023 Semestre 1

TABELA I - TRABALHO DEPENDENTE

NÃO CASADO

Remuneração Mensal Euros		Número de dependentes					
		0	1	2	3	4	5 ou mais
Até	762.00	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Até	790.00	2.0%	0.4%	0.0%	0.0%	0.0%	0.0%
Até	812.00	4.0%	0.7%	0.0%	0.0%	0.0%	0.0%
Até	863.00	7.0%	4.4%	0.9%	0.0%	0.0%	0.0%
Até	964.00	9.3%	6.6%	3.4%	0.0%	0.0%	0.0%
Até	1 051.00	11.2%	7.8%	5.6%	1.3%	0.0%	0.0%
Até	1 113.00	12.0%	8.7%	6.4%	3.2%	0.0%	0.0%
Até	1 194.00	13.0%	10.6%	8.2%	5.0%	2.6%	0.1%
Até	1 280.00	14.0%	11.7%	9.2%	6.0%	3.5%	1.1%
Até	1 380.00	15.1%	12.7%	10.4%	6.9%	4.5%	2.1%
Até	1 466.00	16.2%	13.8%	11.4%	8.0%	6.5%	4.0%
Até	1 609.00	17.2%	14.8%	12.3%	10.0%	7.5%	5.0%

Exemplos de dados não-tidy

LISTA DAS TAXAS IMI POR MUNICÍPIO DO DISTRITO BEJA PARA O ANO 2021

Código Municipio	Município	Prédios Urbanos Avaliados nos Termos do CIMI	Prédios Rústicos	Taxas por freguesia	Dedução fixa por agregado
0201	ALJUSTREL	0,3000 %	0,80 %		
0202	ALMODOVAR	0,3000 %	0,80 %		+Info
0203	ALVITO	0,3000 %	0,80 %		+Info
0204	BARRANCOS	-	-	+Info	+Info
0205	BEJA	0,3200 %	0,80 %		+Info
0206	CASTRO VERDE	0,3000 %	0,80 %		
0207	CUBA	0,3000 %	0,80 %		
0208	FERREIRA DO ALENTEJO	0,3750 %	0,80 %		+Info
0209	MERTOLA	0,3750 %	0,80 %		+Info

Exemplos de dados não-tidy

	A	AA	AB	AC	AD	AE	AF	AG	AH
1	Estimated HIV Prevalence% - (Ages 15-49)	2004	2005	2006	2007	2008	2009	2010	2011
2	Abkhazia								
3	Afghanistan						0.06	0.06	0.06
4	Akrotiri and Dhekelia								
5	Albania								
6	Algeria	0.1	0.1	0.1	0.1	0.1			
7	American Samoa								
8	Andorra								
9	Angola	1.9	1.9	1.9	1.9	2	2.1	2.1	2.1
10	Anguilla								
11	Antigua and Barbuda								
12	Argentina	0.4	0.4	0.4	0.4	0.5	0.4	0.4	0.4
13	Armenia	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.2
14	Aruba								
15	Australia	0.1	0.1	0.1	0.1	0.1	0.2	0.2	0.2
16	Austria	0.2	0.2	0.2	0.3	0.3	0.3	0.4	0.4
17	Azerbaijan	0.06	0.06	0.06	0.1	0.1	0.1	0.1	0.1
18	Bahamas	3	3	3	3.1	3.1	2.9	2.8	2.8

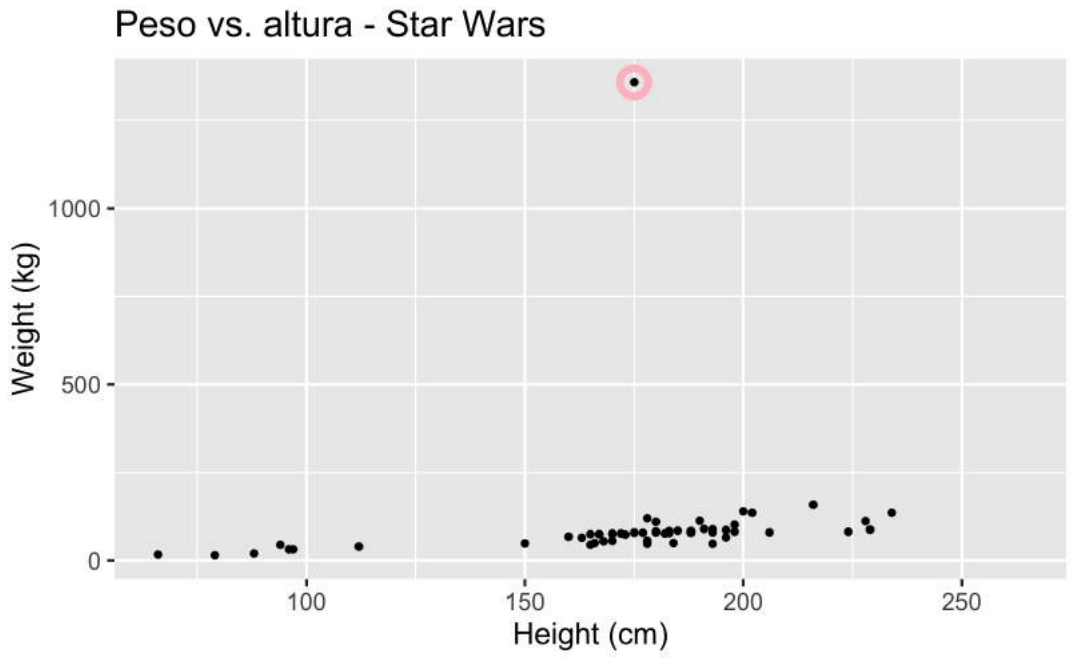
Visualização de dados no R

Análise exploratória de dados

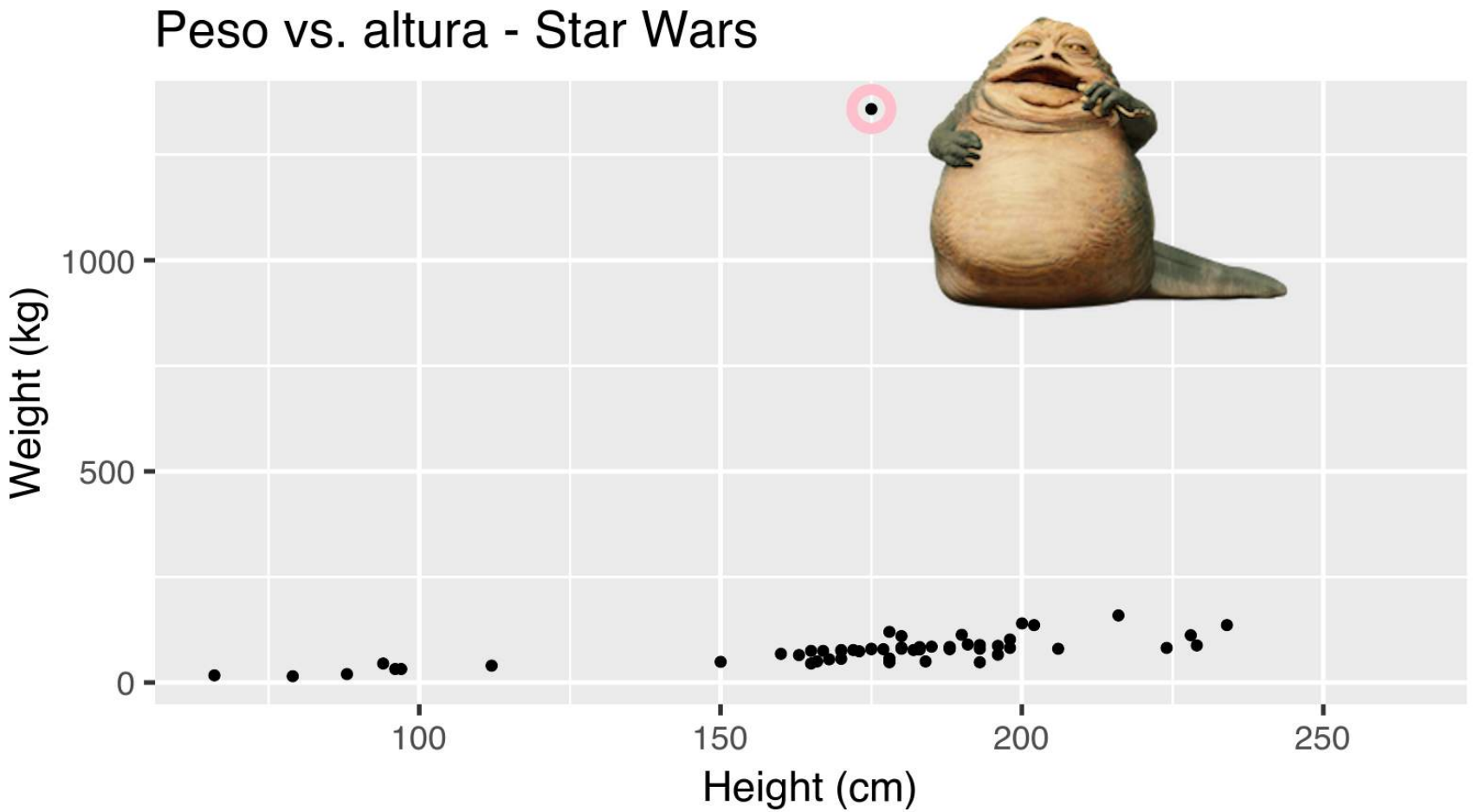
- Análise a conjuntos de dados de forma a sumarizar e entender as suas principais características
- Muitas vezes, é mais fácil fazê-lo visualmente com gráficos (agora)
- Mas também pode ser necessitar de manipulação e transformação dos dados, e cálculo de estatísticas básicas (próximas sessões)

Peso vs. altura

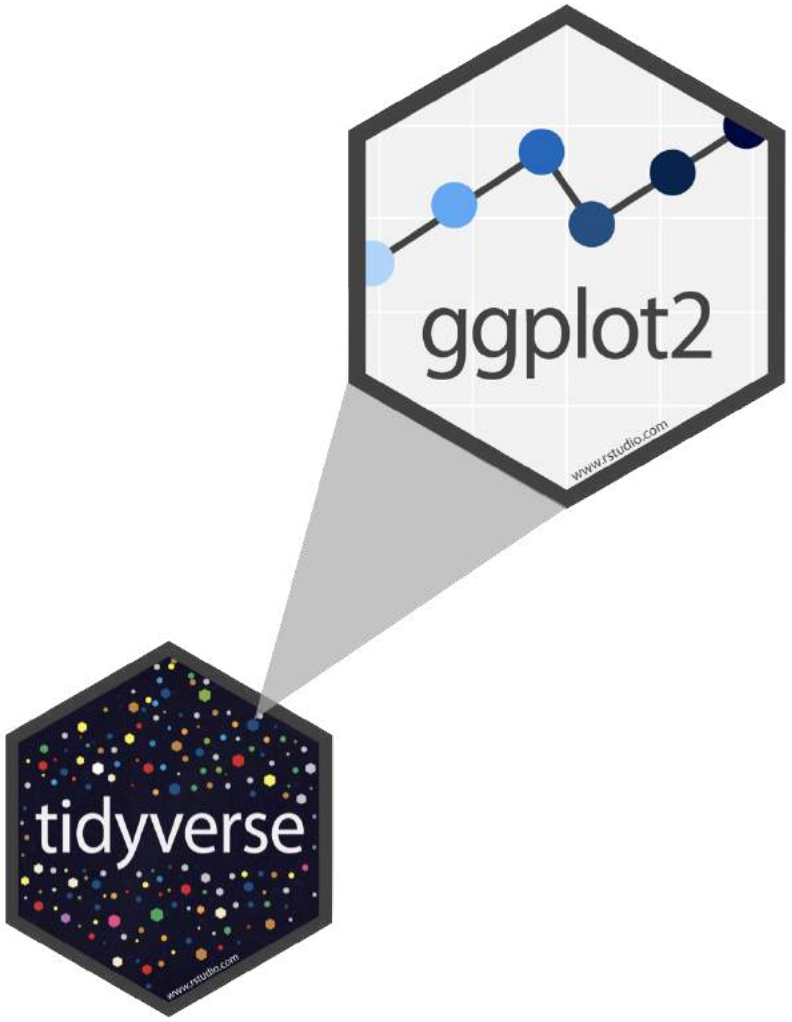
Qual é a relação peso vs. altura entre as personagens do Star Wars? Haverá outras variáveis que possam ajudar a compreender esta relação?



Jabba!


Peso vs. altura - Star Wars
Weight (kg)
1000
500
0
100
150
200
250
Height (cm)

Visualização de dados no tidyverse = ggplot

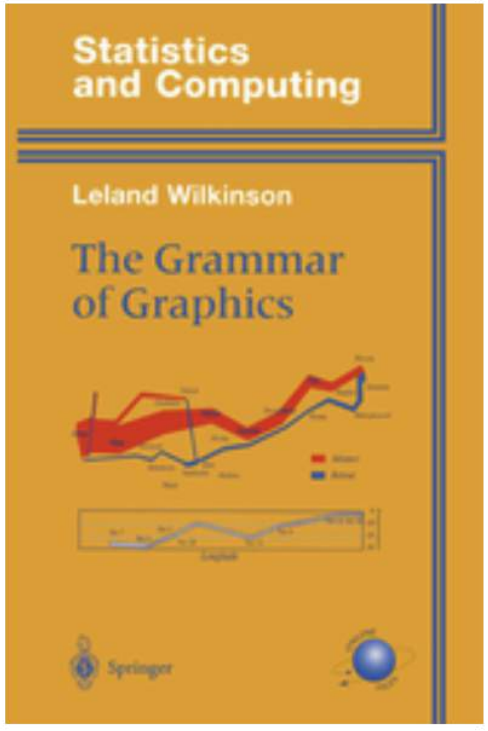


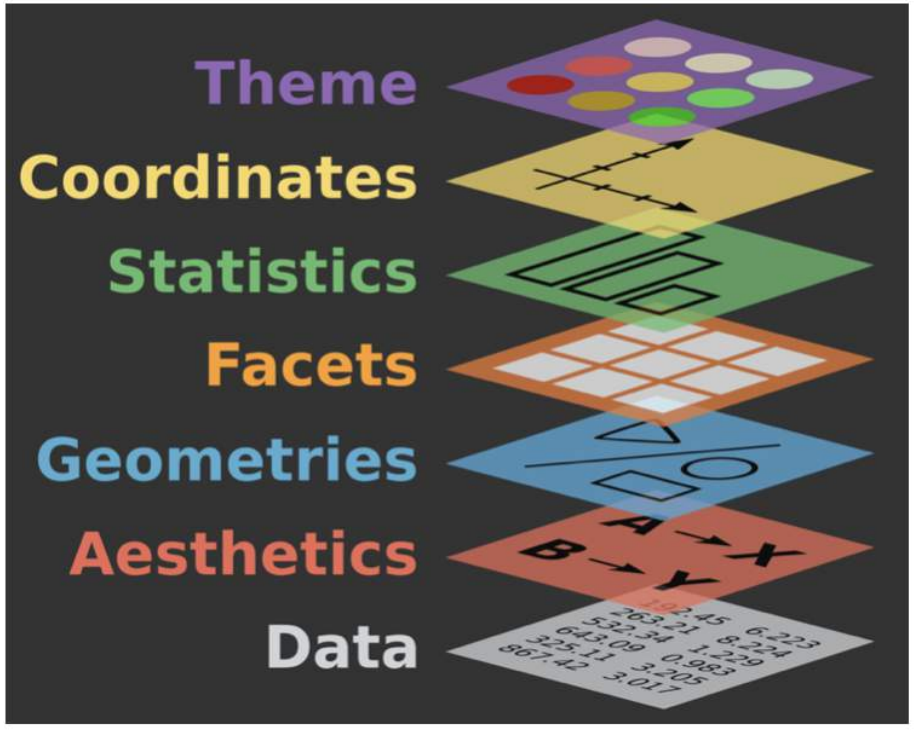
- gg = Gramática dos Gráficos (*Grammar of Graphics*)
- Inspirado pelo livro **Grammar of Graphics** de Leland Wilkinson
- Estrutura básica do código

```
ggplot(data = [dataset],
       mapping = aes(x = [x-variable],
                     y = [y-variable])) +
   geom_xxx() +
   other options
```

Gramática dos Gráficos

Descrição geral dos vários componentes que **qualquer** gráfico tem

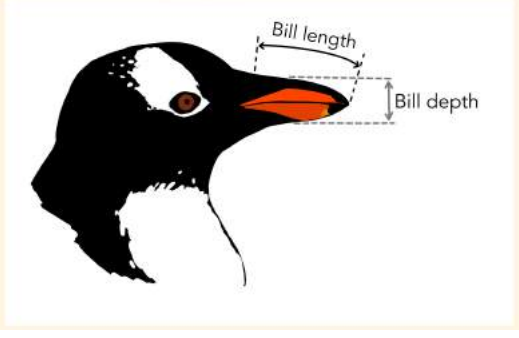




Source: BloggoType

Dados: Pinguins de Palmer

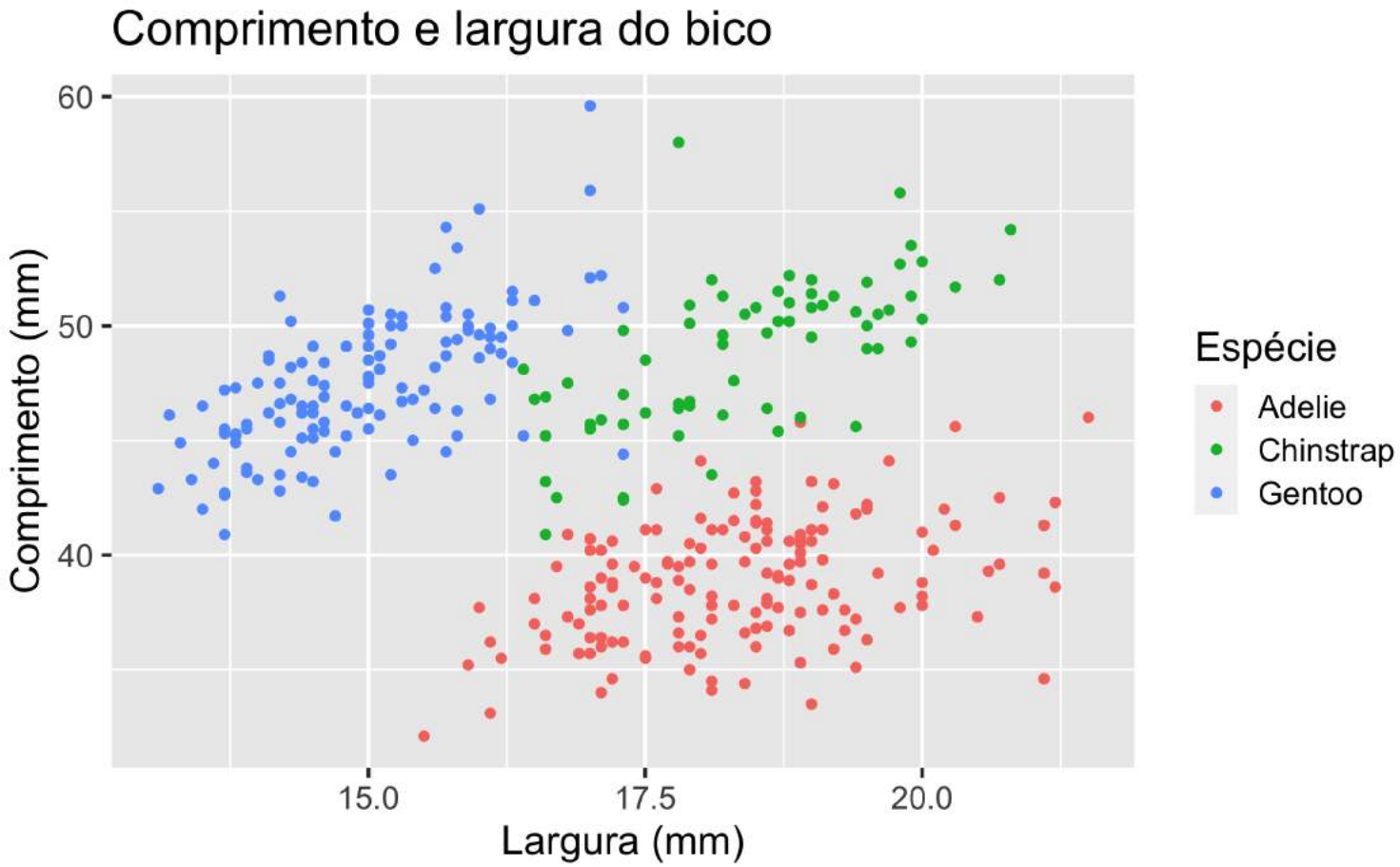
Medições feitas a diferentes espécies de pinguins numa ilha do Archipelago Palmer: dimensões (barbatanas, bico, peso) e sexo.




```
library(palmerpenguins)
glimpse(penguins)
```

```
## Rows: 344
## Columns: 8
## $ species           <fct> Adelie, Adelie, Adelie, Adelie, Adeli
## $ island            <fct> Torgersen, Torgersen, Torgersen, Torg
## $ bill_length_mm    <dbl> 39.1, 39.5, 40.3, NA, 36.7, 39.3, 38.
## $ bill_depth_mm     <dbl> 18.7, 17.4, 18.0, NA, 19.3, 20.6, 17.
## $ flipper_length_mm <int> 181, 186, 195, NA, 193, 190, 181, 195
## $ body_mass_g       <int> 3750, 3800, 3250, NA, 3450, 3650, 362
## $ sex               <fct> male, female, female, NA, female, mal
## $ year              <int> 2007, 2007, 2007, 2007, 2007, 2007, 2
```

Plot Code


Comprimento e largura do bico
Comprimento (mm)
60
50
40
15.0
17.5
20.0
Largura (mm)
Espécie
Adelie
Chinstrap
Gentoo

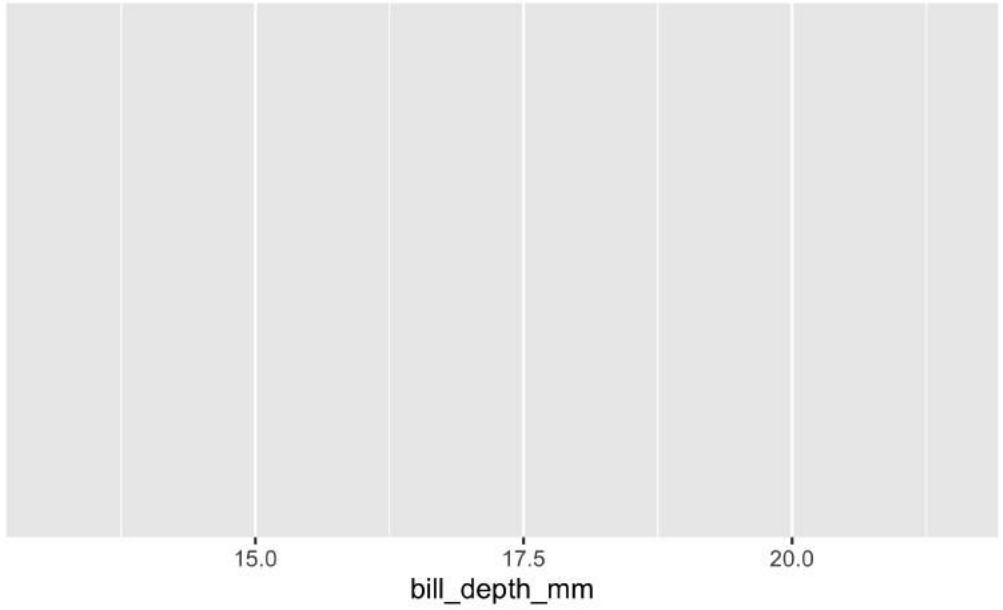
Construção de um gráfico no ggplot2

Começamos com o data frame `penguins`

```
ggplot(data = penguins)
```

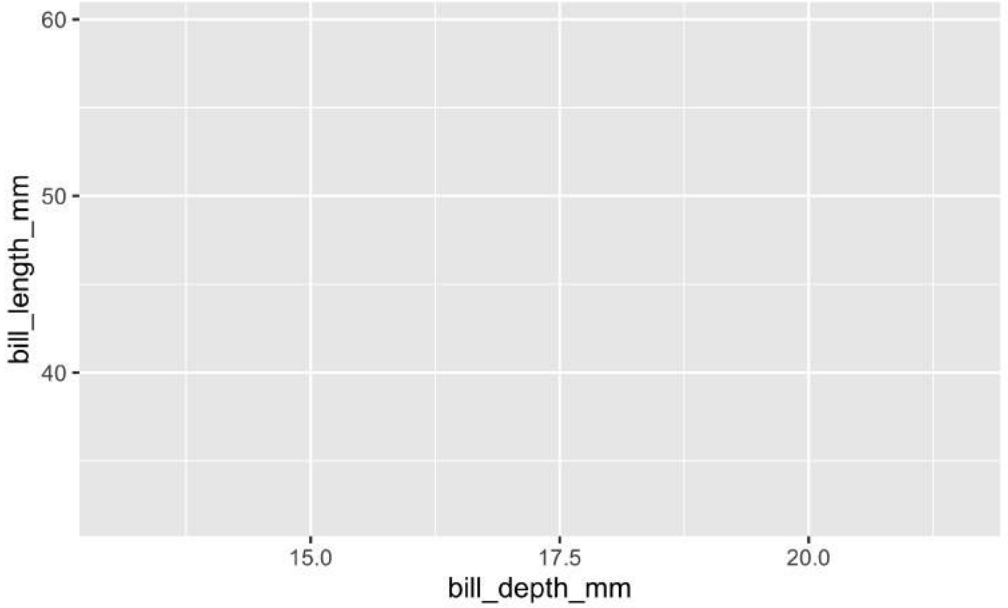

> Começamos com o data frame `penguins`, **mapeamos a largura do bico ao eixo dos x**

```
ggplot(data = penguins,
       mapping = aes(x = bill_depth_mm))
```



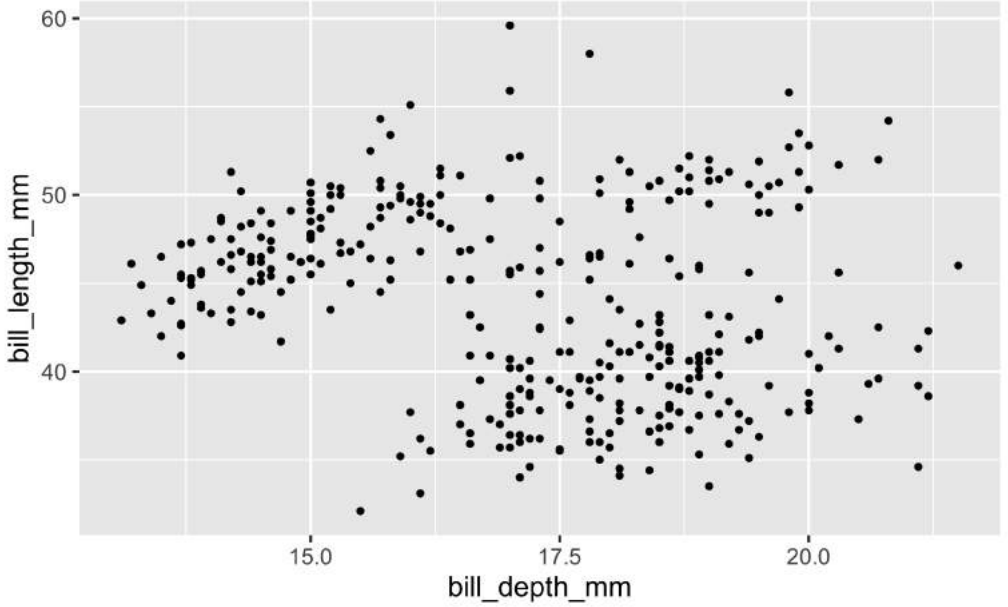
> Começamos com o data frame `penguins`, mapeamos a largura do bico ao eixo dos x **e o comprimento ao eixo dos y.**

```
ggplot(data = penguins,
       mapping = aes(x = bill_depth_mm,
                     y = bill_length_mm))
```



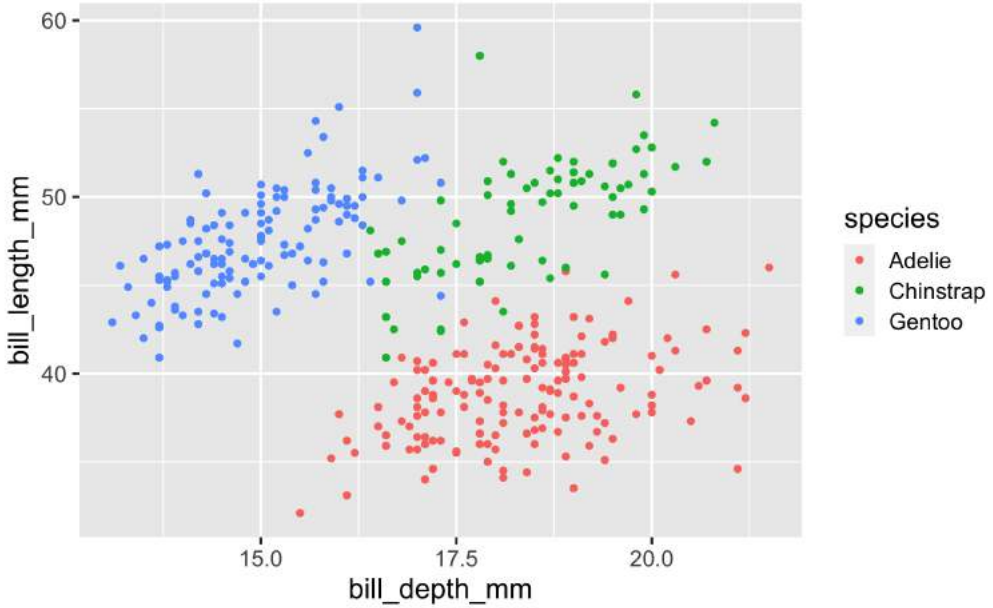
> Começamos com o data frame `penguins`, mapeamos a largura do bico ao eixo dos x e o comprimento ao eixo dos y. **Representamos cada observação por um ponto**

```
ggplot(data = penguins,
       mapping = aes(x = bill_depth_mm,
                     y = bill_length_mm))
  geom_point()
```



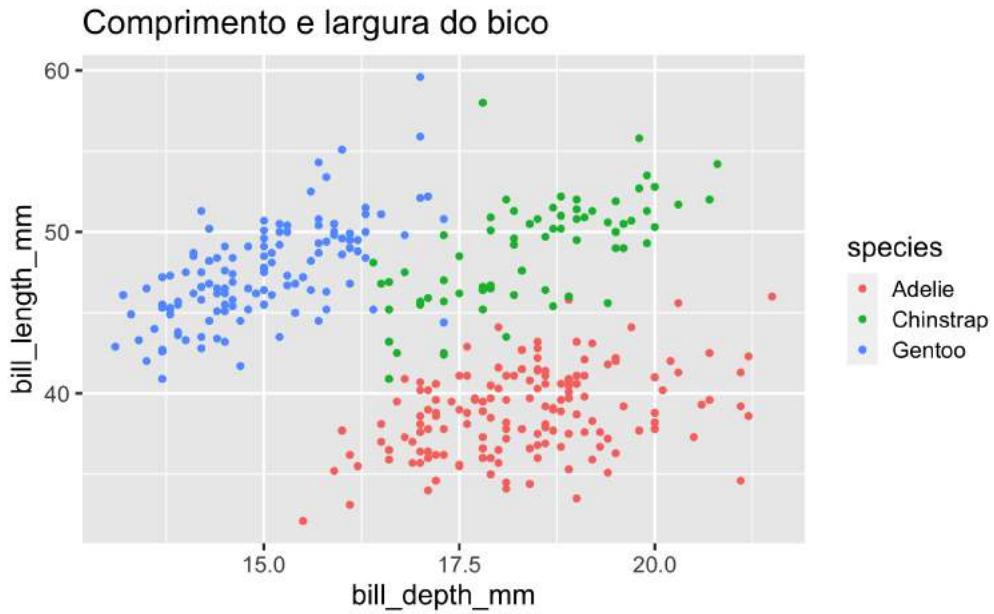
> Começamos com o data frame `penguins`, mapeamos a largura do bico ao eixo dos x e o comprimento ao eixo dos y. Representamos cada observação por um ponto **e mapeamos a espécie à cor dos pontos.**

```
ggplot(data = penguins,
       mapping = aes(x = bill_depth_mm,
                     y = bill_length_mm,
                     colour = species)) +
  geom_point()
```



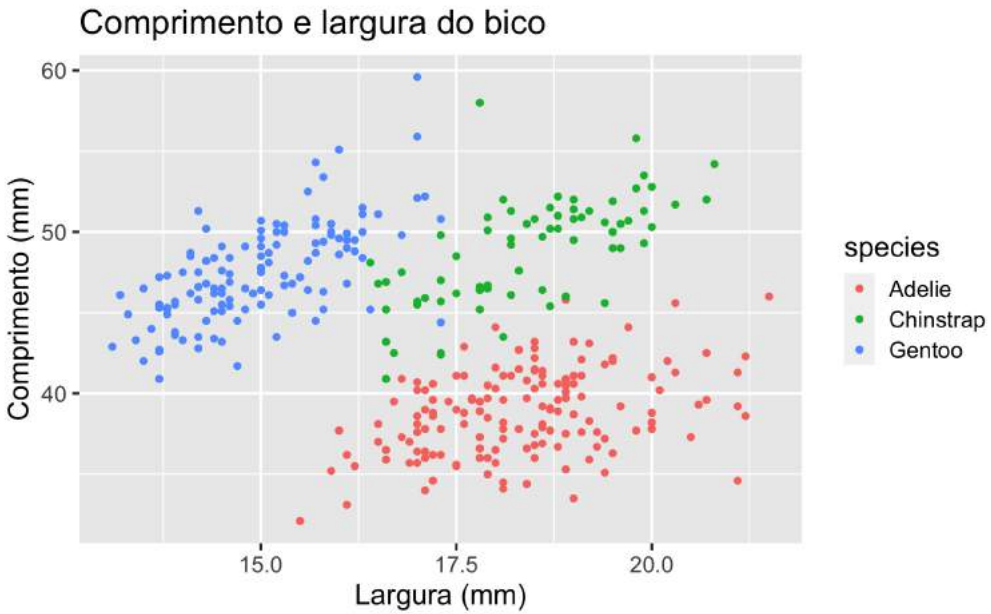
> Começamos com o data frame `penguins`, mapeamos a largura do bico ao eixo dos x e o comprimento ao eixo dos y. Representamos cada observação por um ponto e mapeamos a espécie à cor dos pontos. **Damos um título ao gráfico,**

```
ggplot(data = penguins,
       mapping = aes(x = bill_depth_mm,
                     y = bill_length_mm,
                     colour = species)) +
  geom_point() +
  labs(title = "Comprimento e largura do b
```



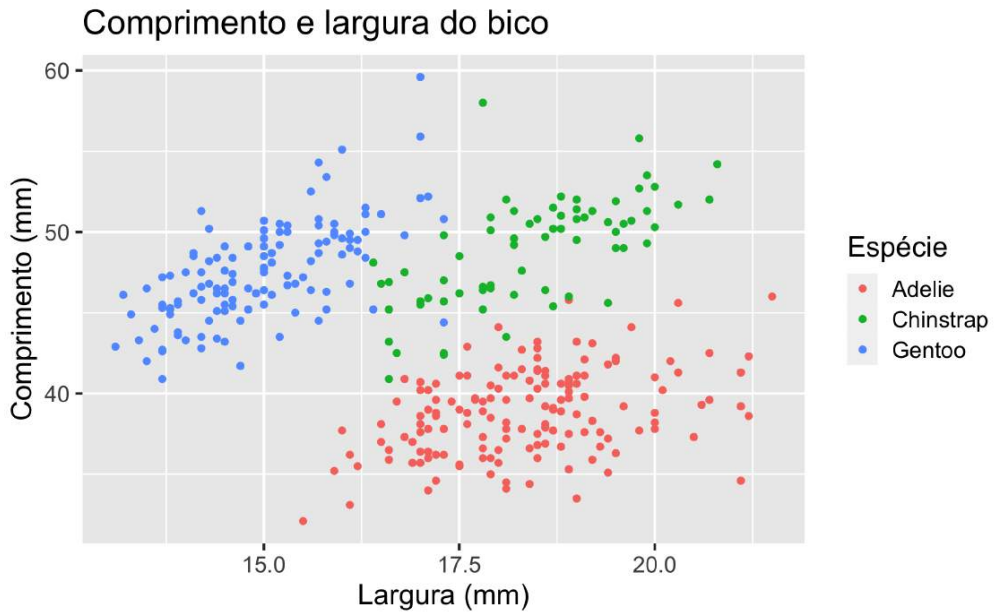
> Começamos com o data frame `penguins`, mapeamos a largura do bico ao eixo dos x e o comprimento ao eixo dos y. Representamos cada observação por um ponto e mapeamos a espécie à cor dos pontos. Damos um título ao gráfico, **alteramos as legendas dos eixos**

```
ggplot(data = penguins,
       mapping = aes(x = bill_depth_mm,
                     y = bill_length_mm,
                     colour = species)) +
  geom_point() +
  labs(title = "Comprimento e largura do b
       x = "Largura (mm)", y = "Compriment
```



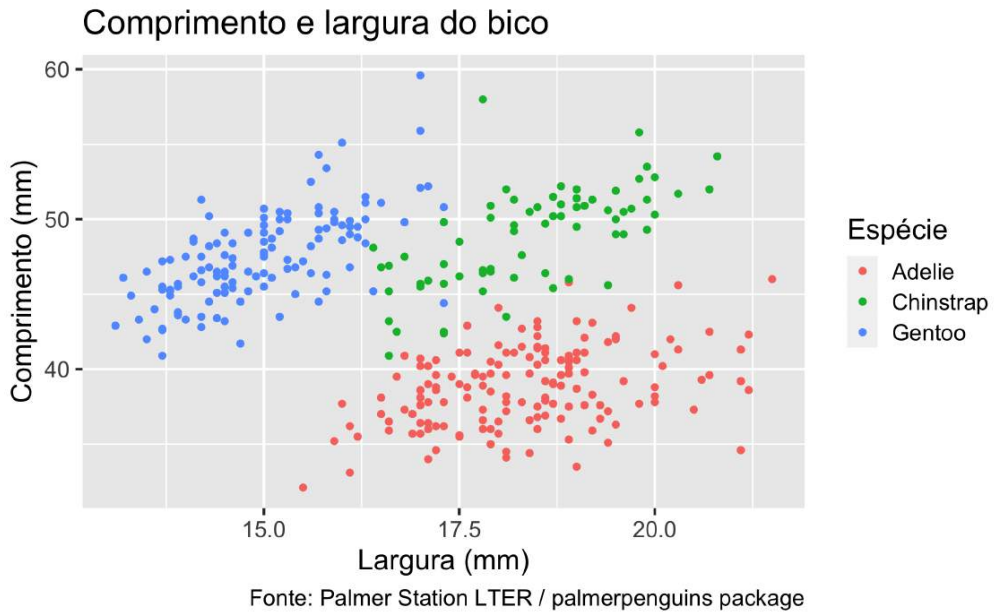
> Começamos com o data frame `penguins`, mapeamos a largura do bico ao eixo dos x e o comprimento ao eixo dos y. Representamos cada observação por um ponto e mapeamos a espécie à cor dos pontos. Damos um título ao gráfico, alteramos as legendas dos eixos **e o título da legenda das cores.**

```
ggplot(data = penguins,
       mapping = aes(x = bill_depth_mm,
                     y = bill_length_mm,
                     colour = species)) +
  geom_point() +
  labs(title = "Comprimento e largura do b
       x = "Largura (mm)", y = "Compriment
       colour = "Espécie")
```

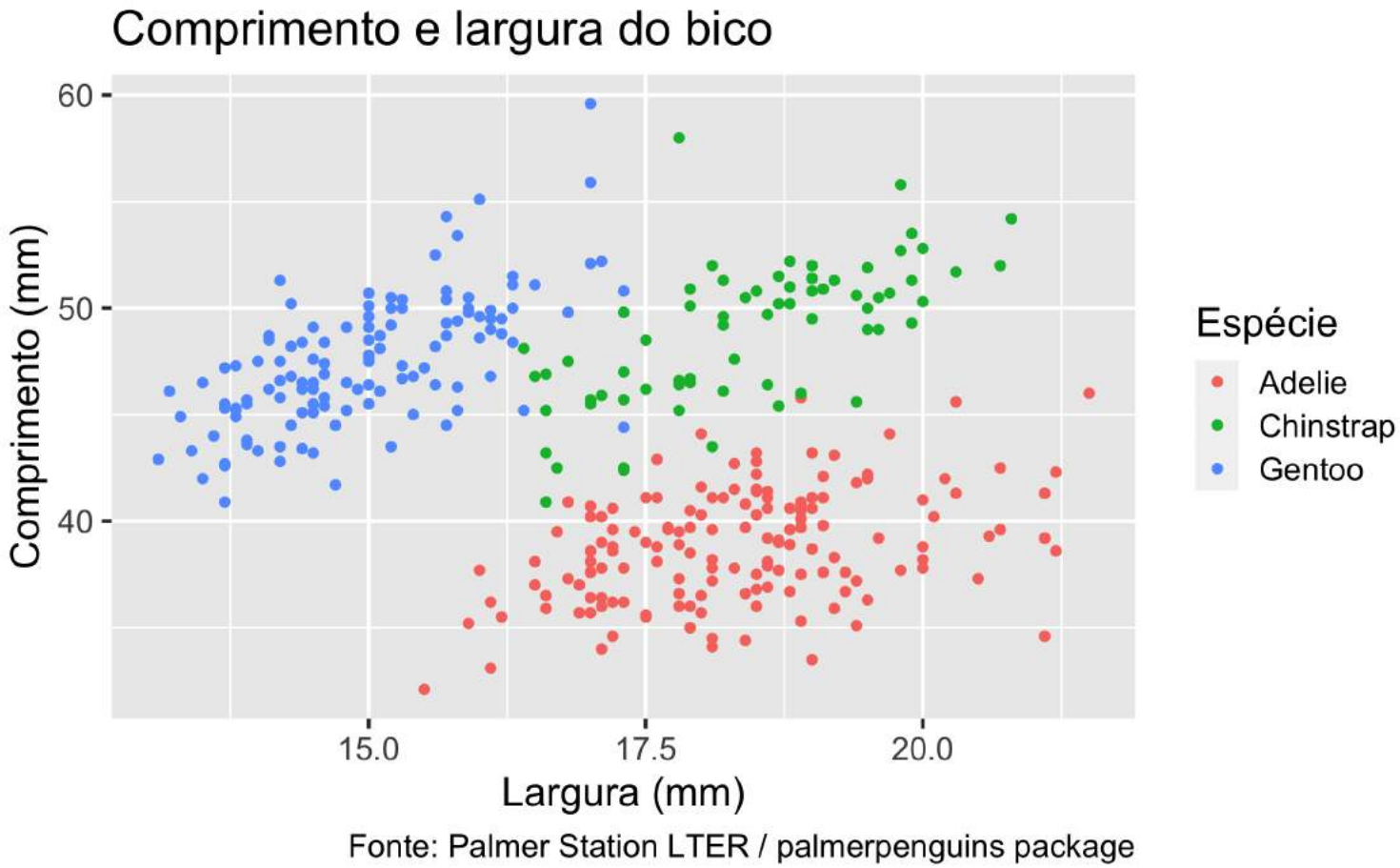


> Começamos com o data frame `penguins` mapeamos a largura do bico ao eixo dos x e o comprimento ao eixo dos y. Representamos cada observação por um ponto e mapeamos a espécie à cor dos pontos. Damos um título ao gráfico, alteramos as legendas dos eixos e o título da legenda das cores. **Acrescentamos uma nota com a fonte dos dados.**

```
ggplot(data = penguins,
       mapping = aes(x = bill_depth_mm,
                     y = bill_length_mm,
                     colour = species)) +
  geom_point() +
  labs(title = "Comprimento e largura do b
       x = "Largura (mm)", y = "Compriment
       colour = "Espécie",
       caption = "Fonte: Palmer Station LT
```



Plot Code Narrative



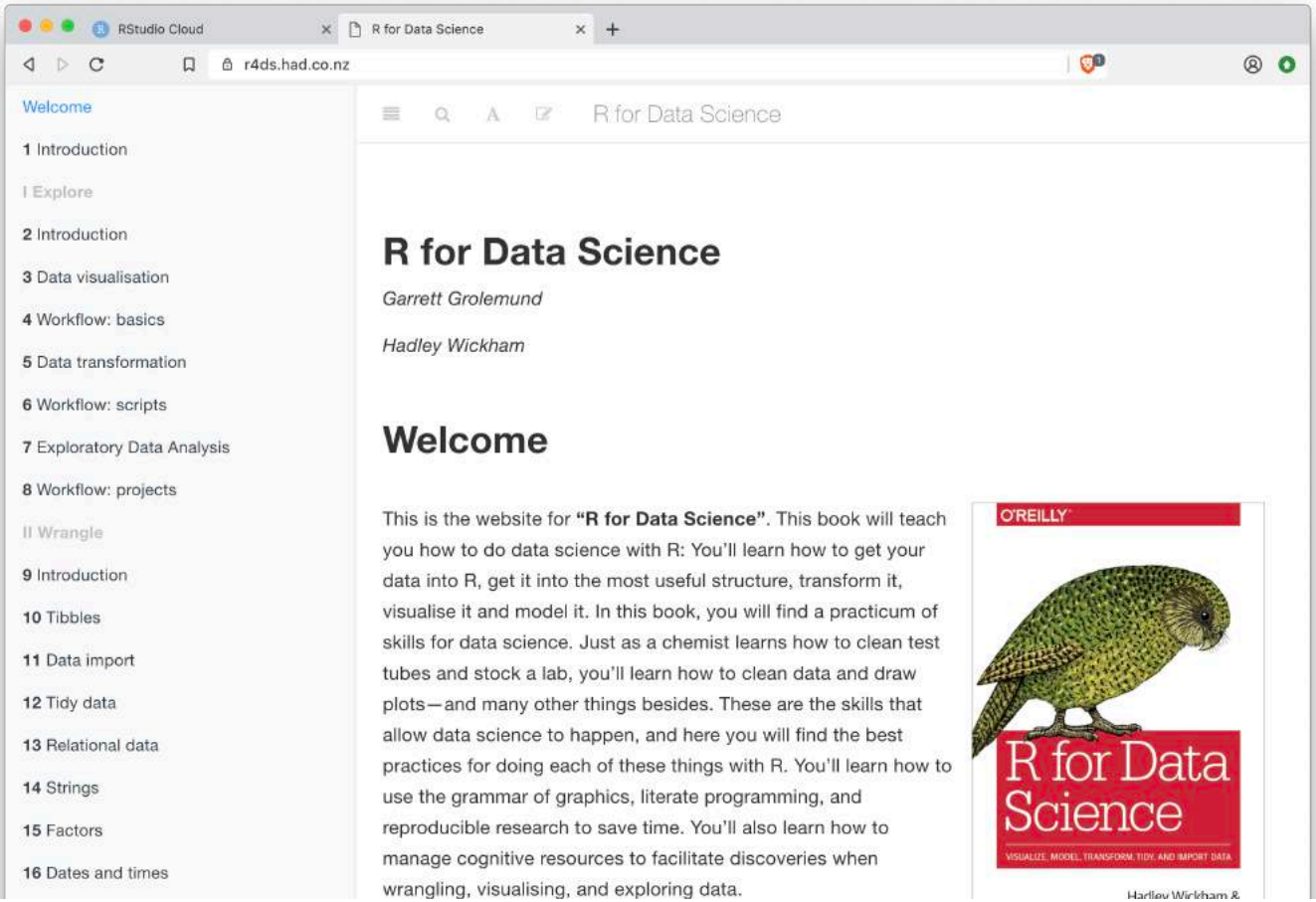
Dica: podemos omitir os nomes dos primeiros argumentos em ggplot

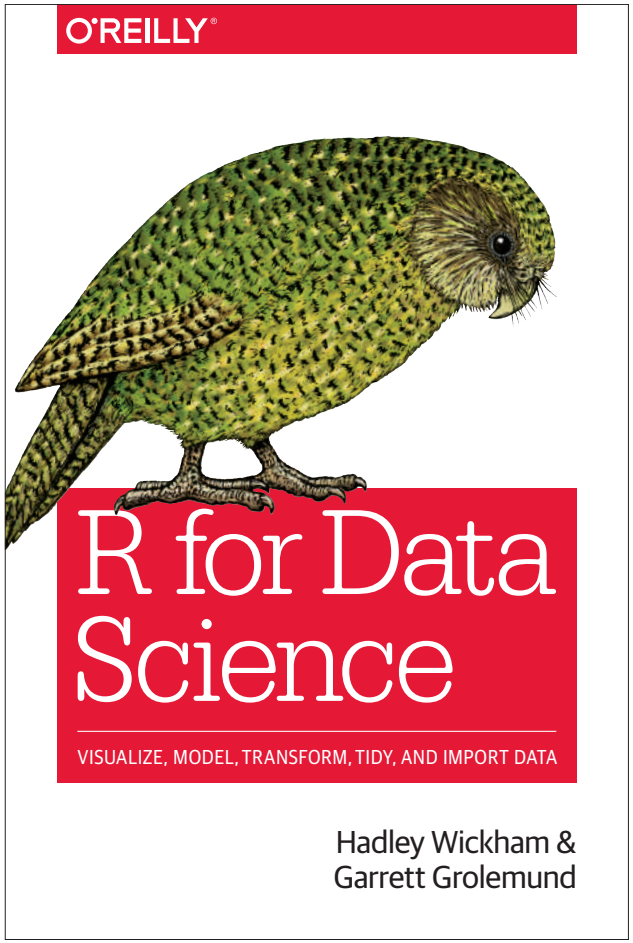
```
ggplot(data = penguins,
       mapping = aes(x = bill_depth_mm,
                     y = bill_length_mm,
                     colour = species)) +
  geom_point()
```

```
ggplot(penguins,
       aes(x = bill_depth_mm,
           y = bill_length_mm,
           colour = species)) +
  geom_point()
```

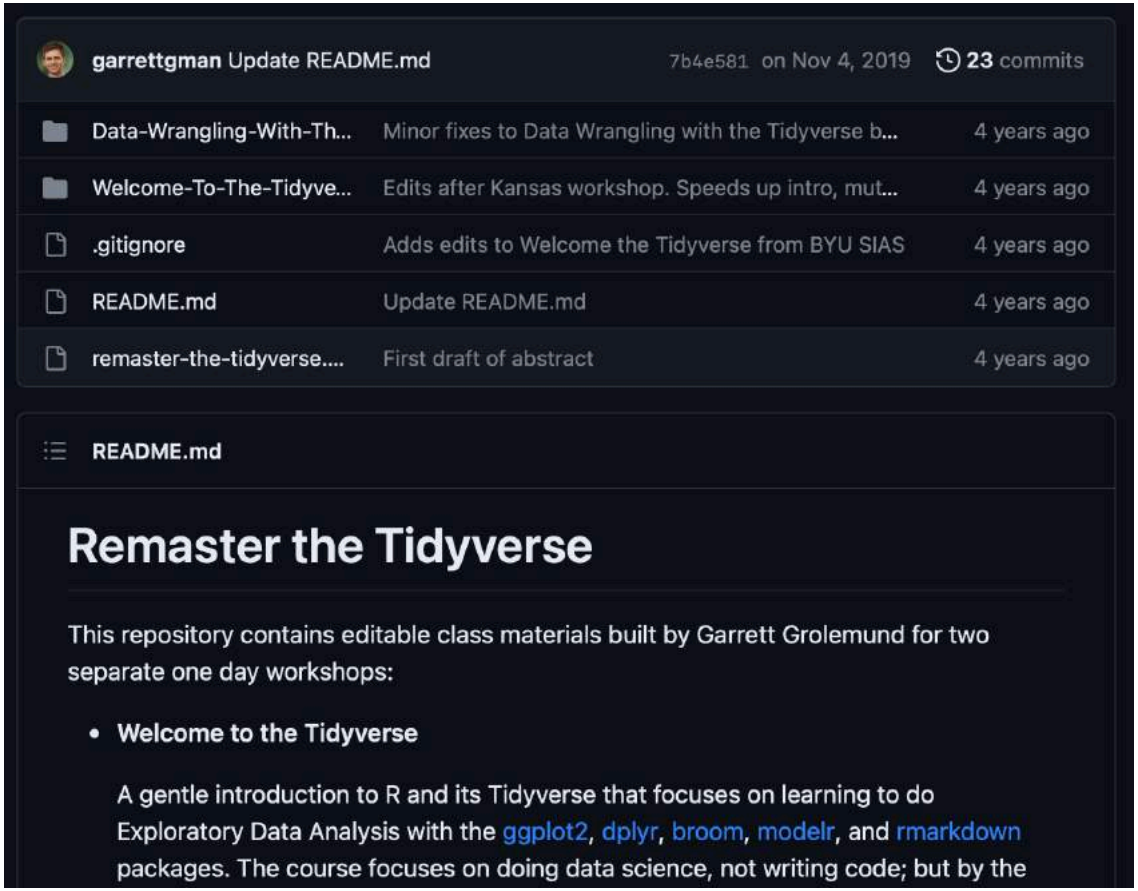
Fontes

r4ds.had.co.nz

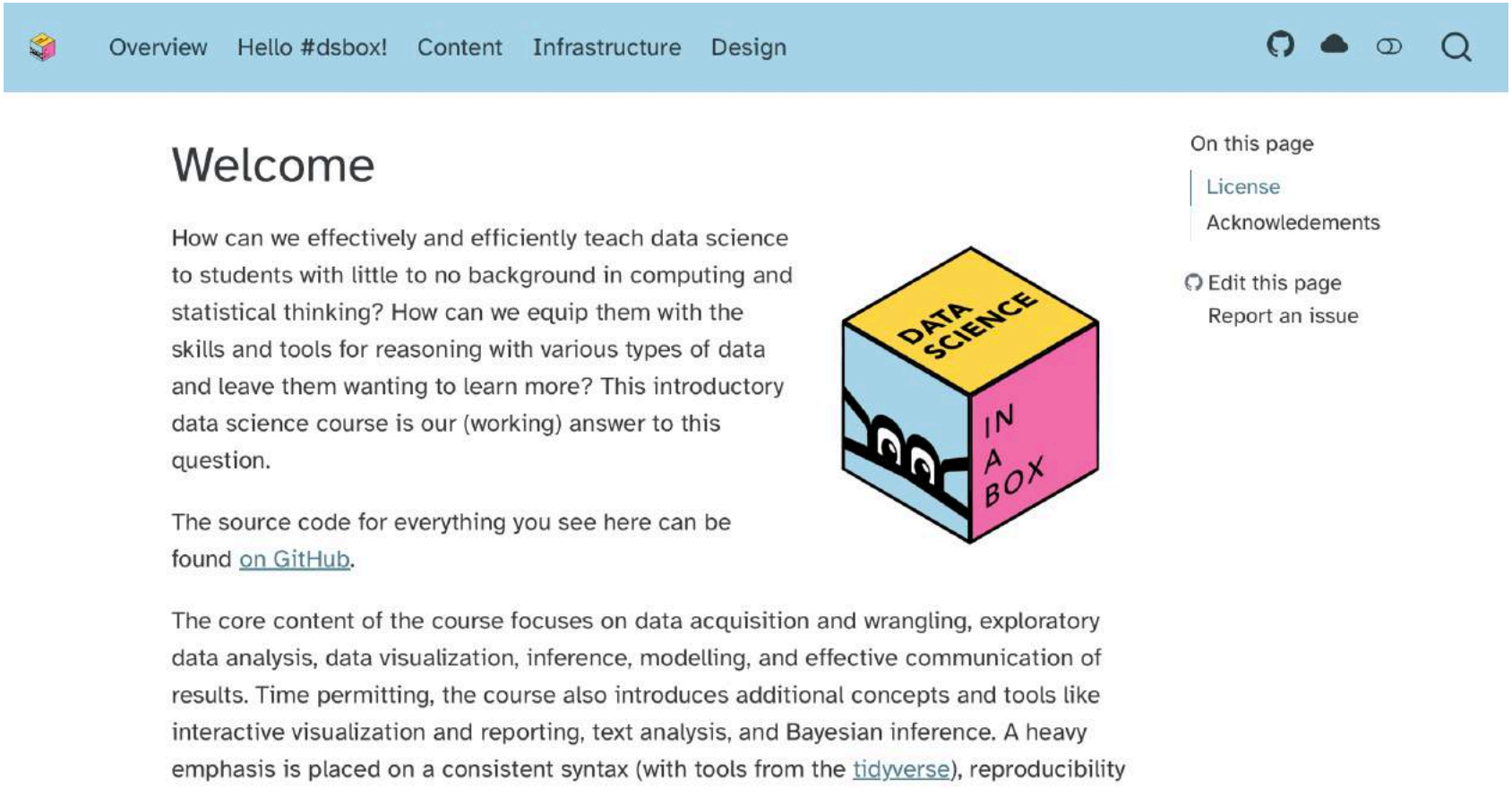




https://github.com/rstudio-education/remaster-the-tidyverse



https://datasciencebox.org/



Obrigado e até amanhã!

luis.morais@novasbe.pt