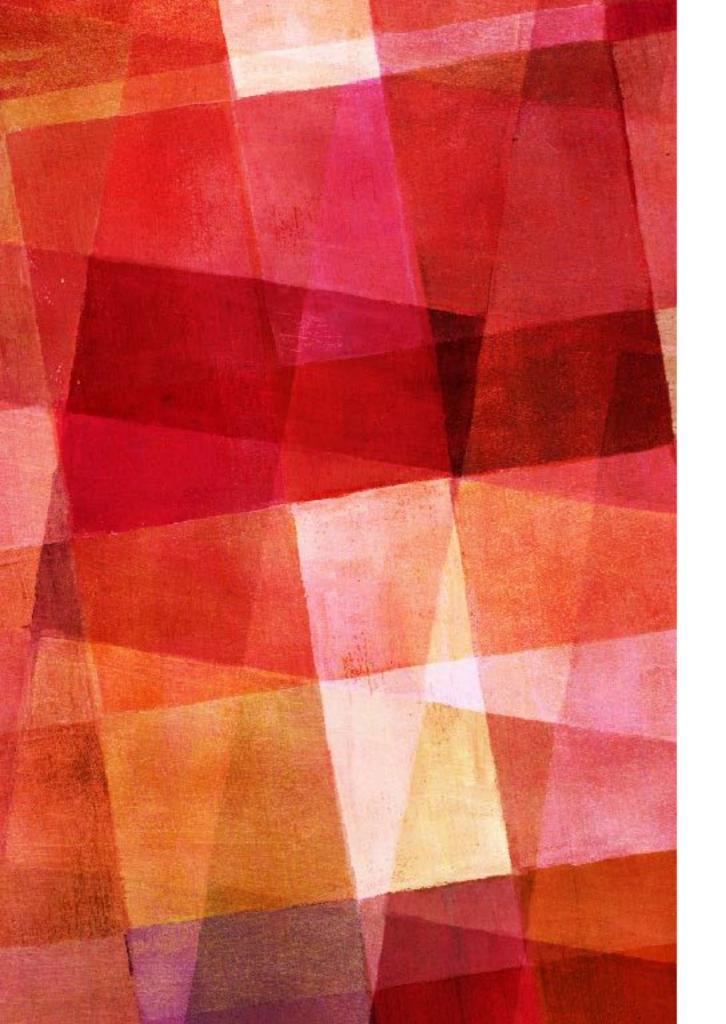
INTRODUÇÃO RSTUDIO

LINGUAGEM R

Plataforma ou ferramenta?



R ECOSYSTEM

- ➤ A plataforma R" é uma "ferramenta" gráfica para gerenciamento de dados estatísticos.
- ➤ R é plataforma de código aberto, mantido e desenvolvido por uma comunidade de desenvolvedores.
- ➤ O repositório de código R, bem c o m o o s "b i n á r i o s compilados" (software pronto para instalar) estão disponíveis em: http://cran.r-project.org
- R compreende um projeto centralizado com mais 8000 de "pacotes" complementares disponíveis livremente.





CRAN

Mirrors

What's new?

Task Views

Search

About R

R Homepage

The R Journal

Software

R Sources

R Binaries

Packages

Other

Documentation

Manuals

FAOs

Contributed

The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, Windows and Mac users most likely want one of these versions of R:

- · Download R for Linux
- Download R for (Mac) OS X
- · Download R for Windows

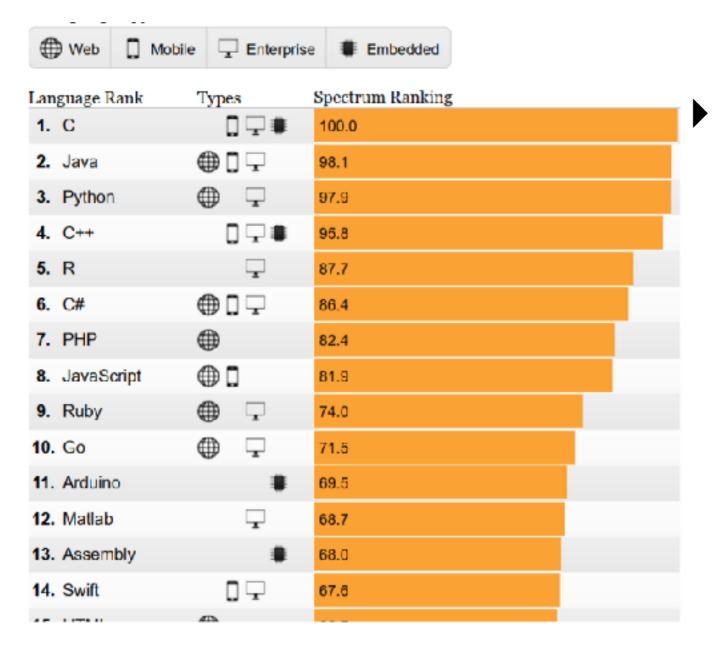
R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2013-09-25, Frisbee Sailing) R-3.0.2.tar.gz, read what's new in the latest version.
- Sources of R alpha and beta releases (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are <u>available here</u>. Please read about <u>new features and bug</u> <u>fixes</u> before filing corresponding feature requests or bug reports.
- Source code of older versions of R is available here.
- Contributed extension packages





SOFTWARE ESTATÍSTICOS

Os software de estatísticas mais populares (que não o R) e alguns de seus públicos:

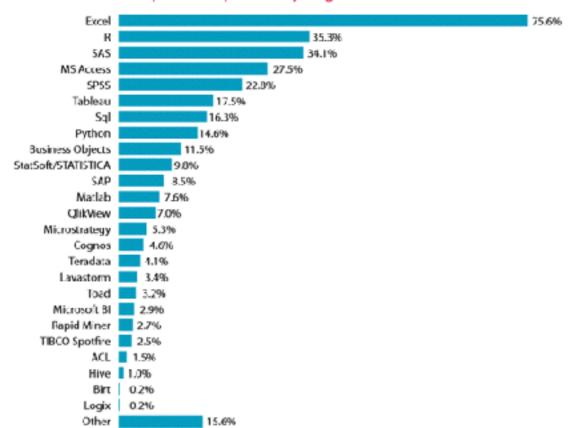
- ✓ SPSS: cientistas sociais
- ✓ Stata: cientistas sociais
- ✓ Mathematica / Matlab: engenheiros, matemáticos, informáticos e físicos
- √Python / NumPy: cientistas da computação, desenvolvedores web
- ✓ SAS: indústrias intensivas em dados (por exemplo, serviços financeiros)
- √ Excel: todos os tipos de organizações
- R é mais popular e usado por um número muito grande de analistas

Fonte: UMA ABORDAGEM PARA CRIAÇÃO DE VALOR EM DADOS ABERTOS PARA PEQUENAS E MÉDIAS EMPRESAS UTILIZANDO O ECOSSISTEMA R Albino, J. P.

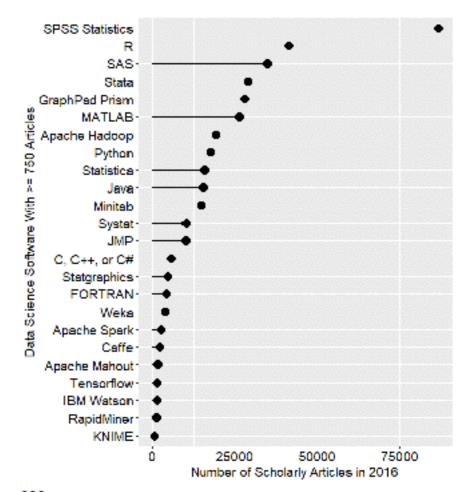
3o. Encontro Internacional de Dados, Tecnologia e Informação UNESP - Marília - 2016

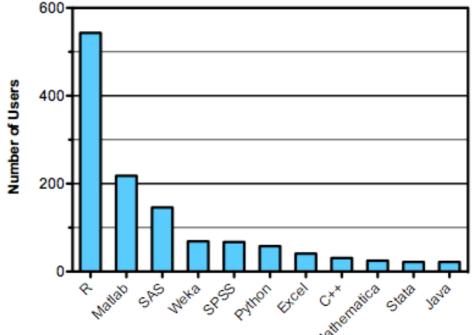
POPULARIDADE DO R

What self-service analytic tool are you currently using?



Fonte: The Popularity of Data Science Software Muenchen, R.A. http://r4stats.com/articles/popularity/





R GUI

Usuários — R — 94×20

R version 3.4.1 (2017-06-30) -- "Single Candle" Copyright (C) 2017 The R Foundation for Statistical Computing Platform: x86_64-apple-darwin16.6.0 (64-bit)

R é um software livre e vem sem GARANTIA ALGUMA. Você pode redistribuí-lo sob certas circunstâncias. Digite 'license()' ou 'licence()' para detalhes de distribuição.

R é um projeto colaborativo com muitos contribuidores. Digite 'contributors()' para obter mais informações e 'citation()' para saber como citar o R ou pacotes do R em publicações.

Digite 'demo()' para demonstrações, 'help()' para o sistema on-line de ajuda, ou 'help.start()' para abrir o sistema de ajuda em HTML no seu navegador. Digite 'q()' para sair do R.

> 2 + 4

Usuários estatísticos gostam de apontar e clicar

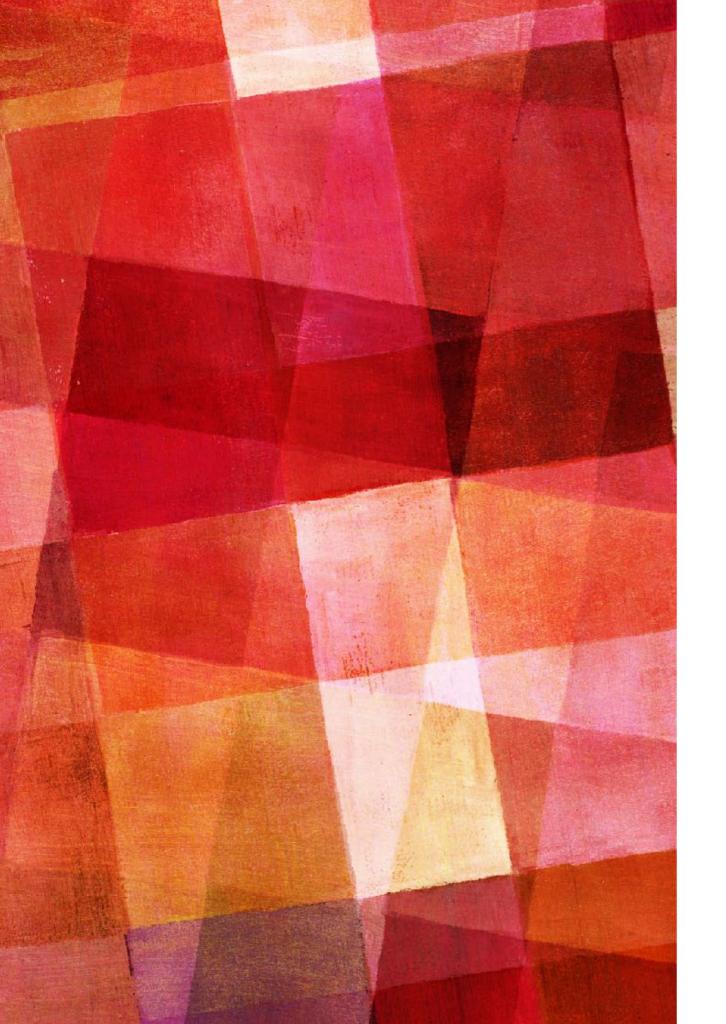
R é orientado a linha de comando; existem **GUIs** que podem ser carregadas como pacotes add-ons (*adicionais*);

O R-Studio é um ambiente de desenvolvimento integrado (IDE) para R, mas mais para desenvolvimento de codigo do que de análise estatística



R é gratuito, mas isso também não significa que não exista um mecanismo formal de suporte;

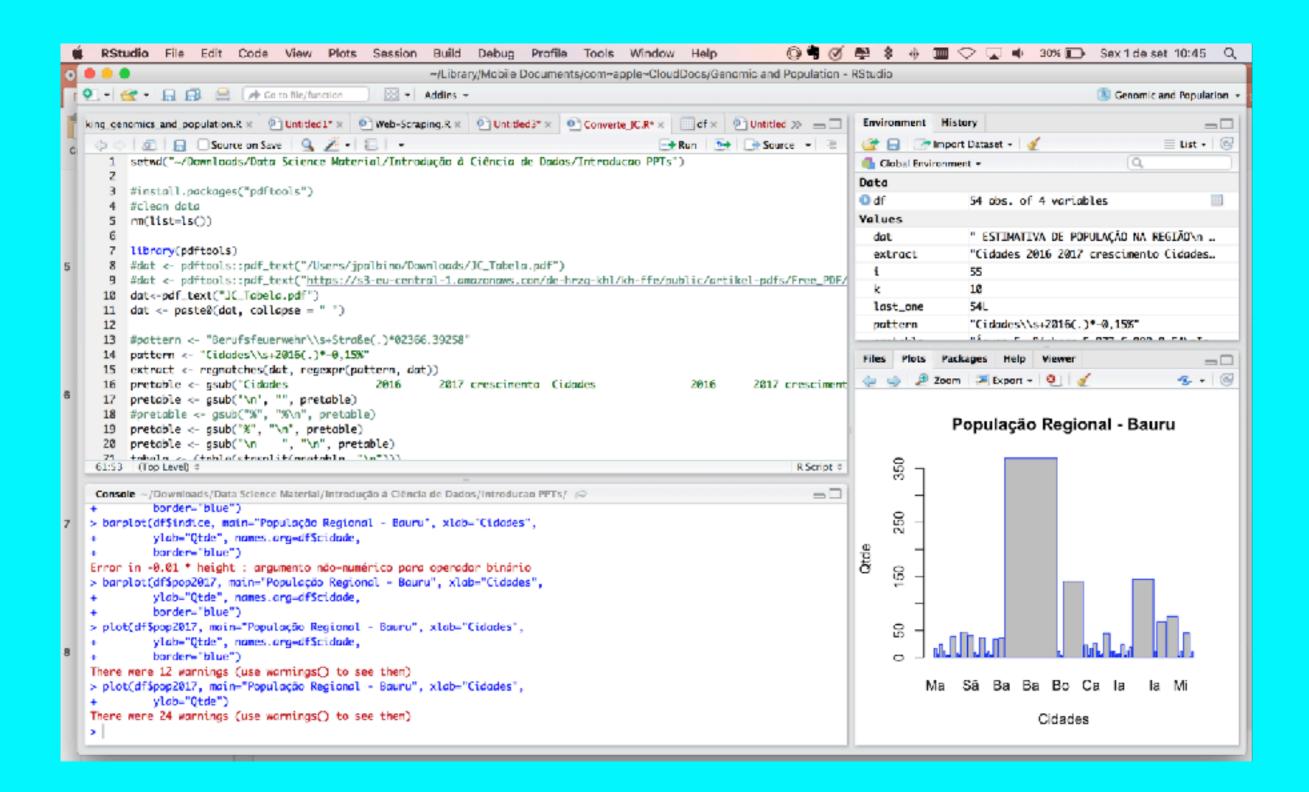
Muitas vezes, grandes organizações costumam contratam um fornecedor comercial



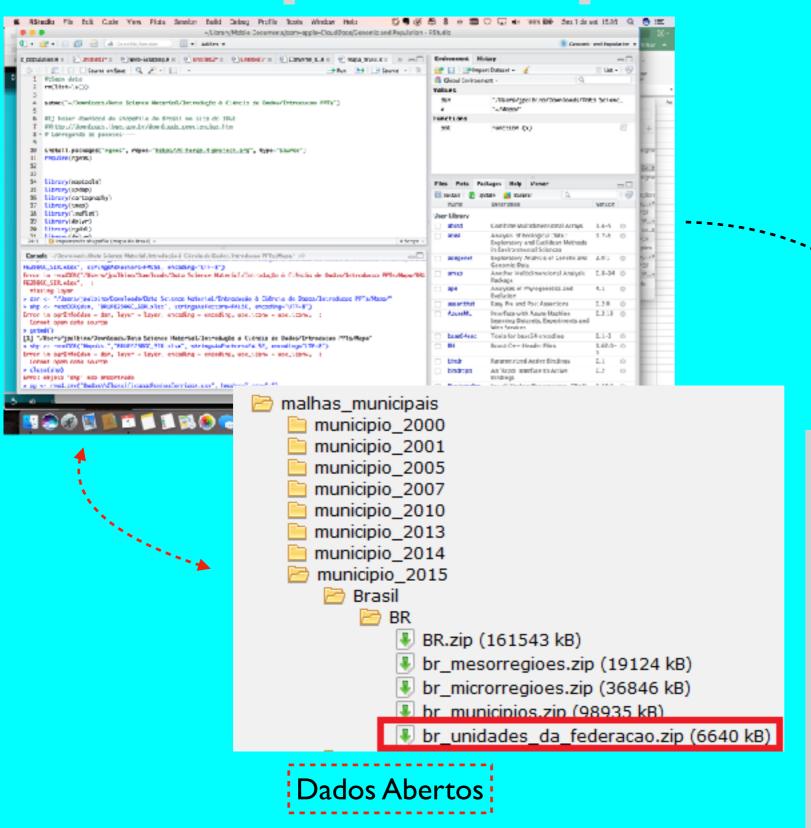
RSTUDIO

- ➤ RStudio permite que o usuário execute R em um ambiente mais amigável.
- ➤ É de código aberto (ou seja, gratuito) e disponível em http://www.rstudio.com/
- ➤ Para tutoriais e / ou recursos relacionados com a linguagem R, veja os seguintes links:
 - http://dss.princeton.edu/ training/ http:// libguides.princeton.edu/ dss

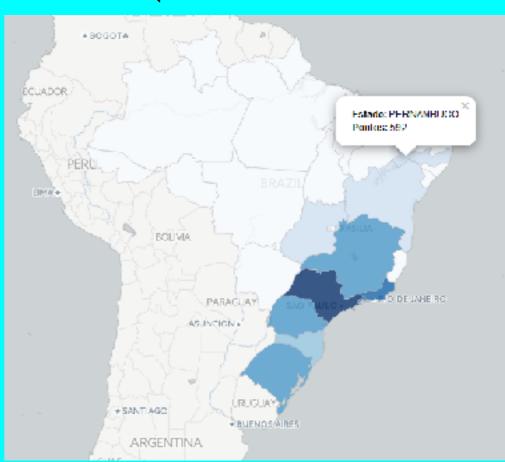
R Studio - R's IDE

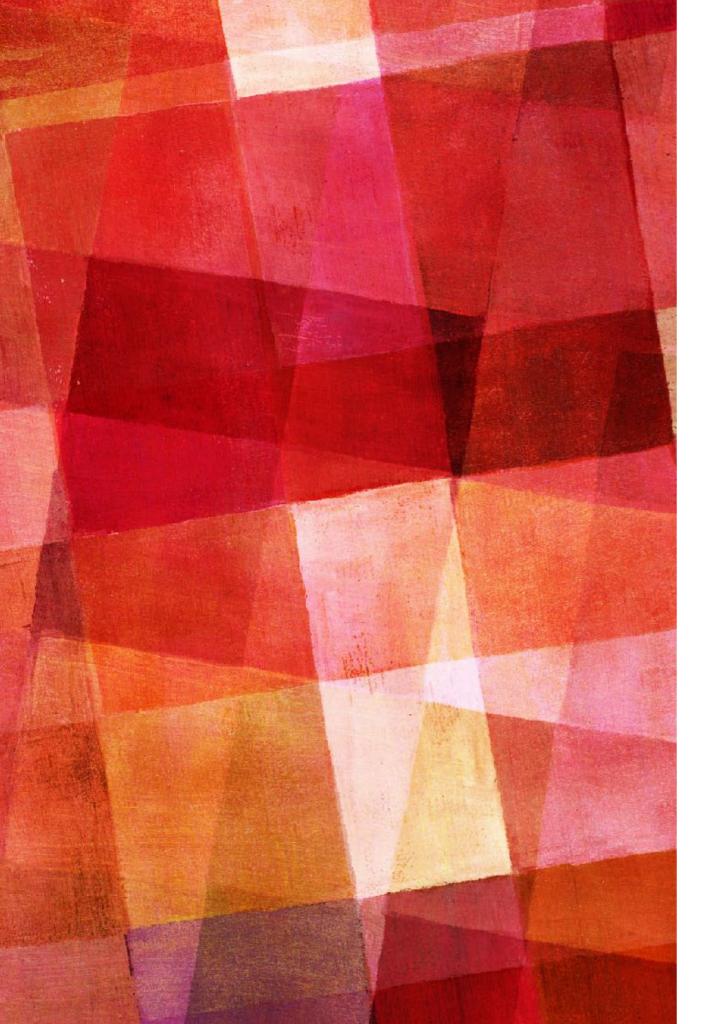


exemplo 2: mapa do brasil em R



Fonte: Criando um mapa do Brasil no R em 7 passos https://dataficacao.wordpress.com/2017/02/21/criando-mapa-brasil-r/

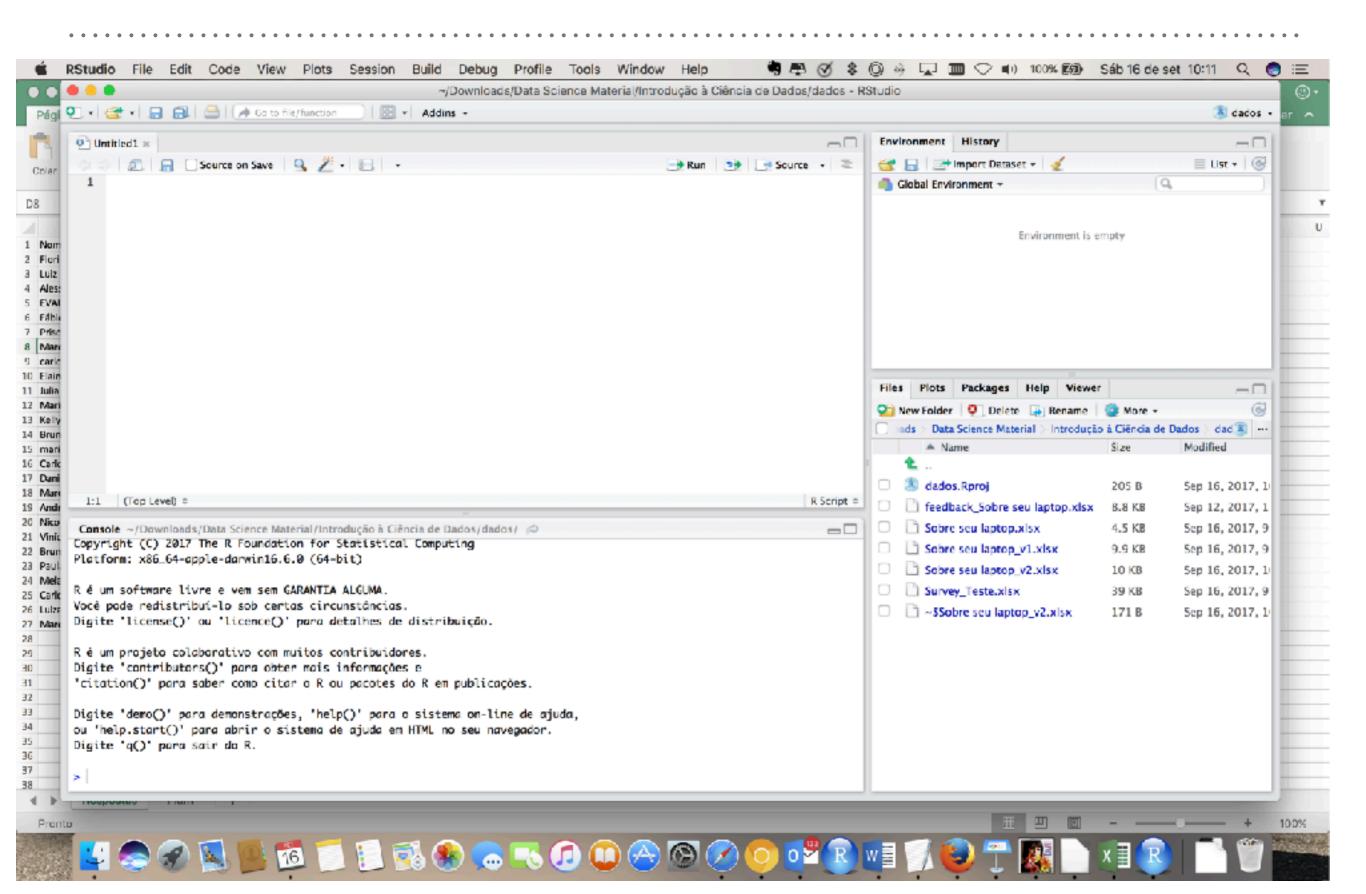




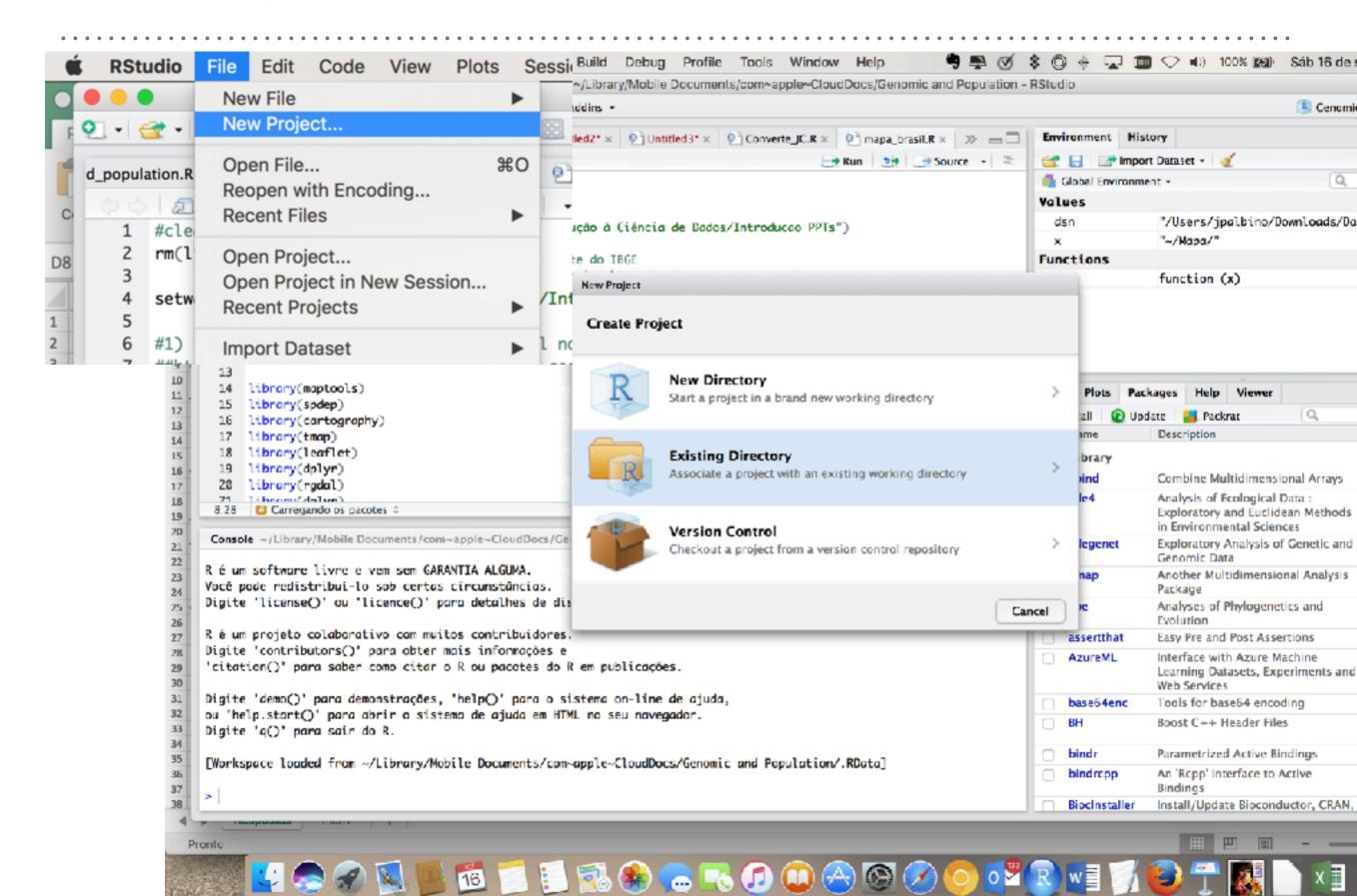
INICIANDO UM PROJETO

Passo-a-passo para iniciar um projeto

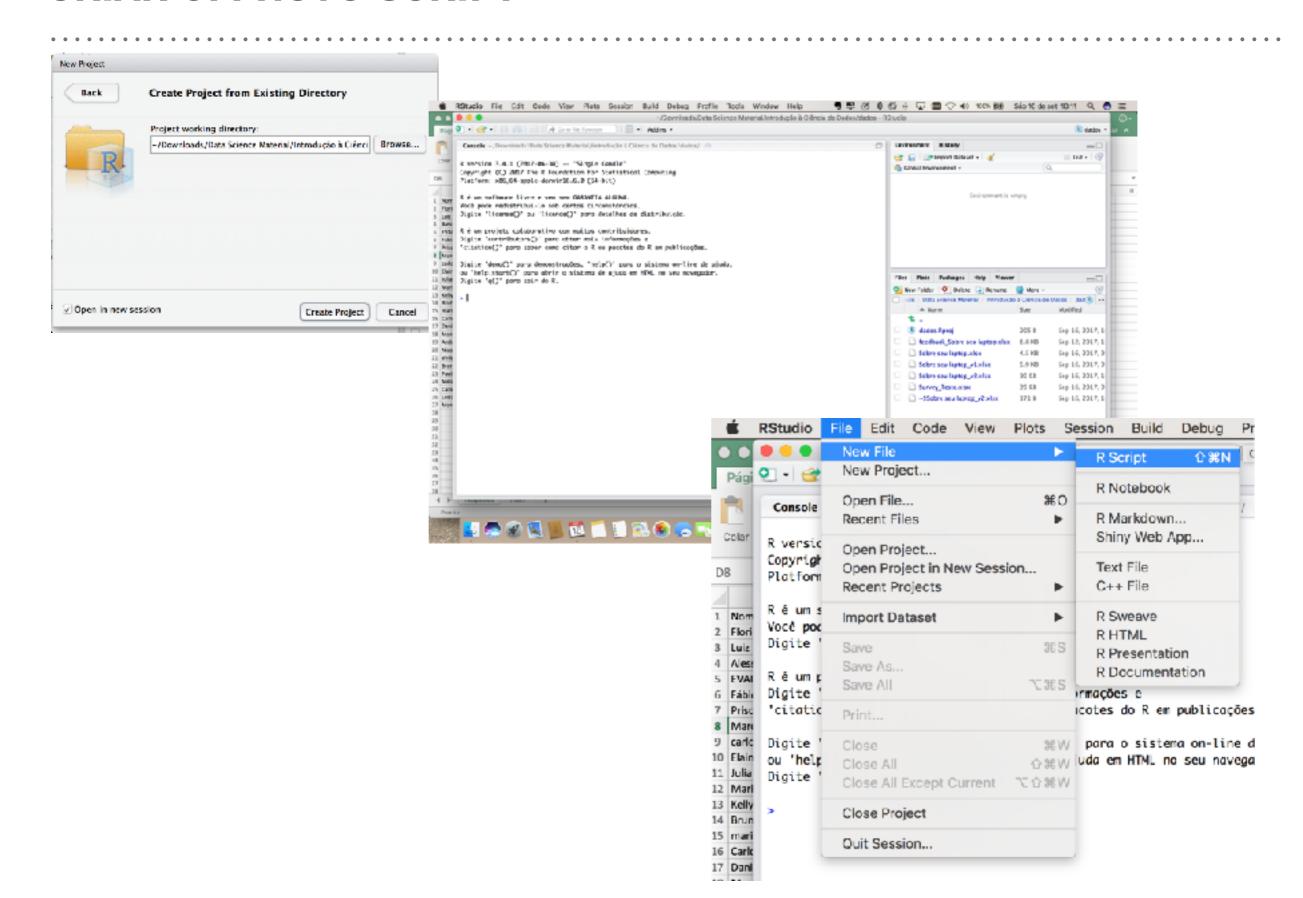
RSTUDIO - TELA INICIAL

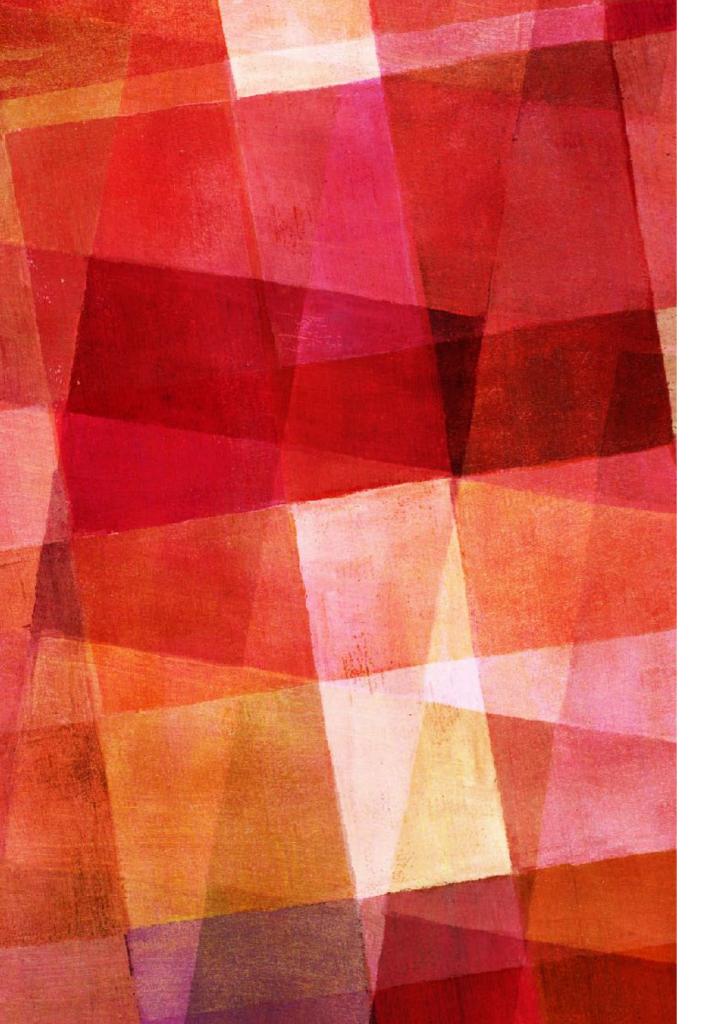


NOVO PROJETO



CRIAR UM NOVO SCRIPT





CARACTERÍSTICAS DA LINGUAGEM

➤ R é uma linguagem "orientada a objetos"

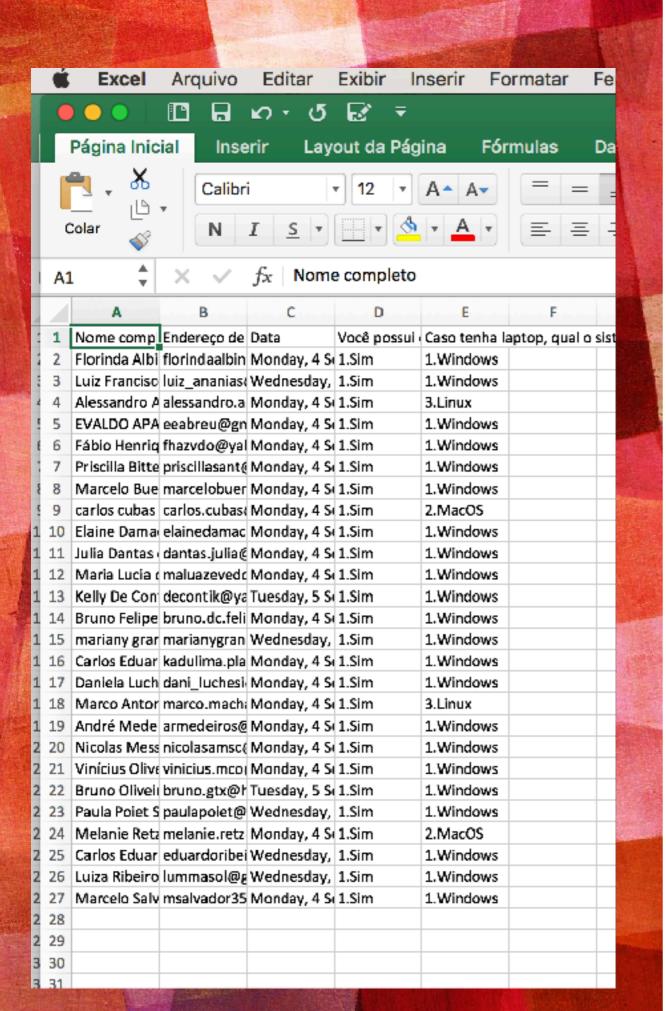
e.x., MyModel <- lm(wt \sim ht, data = mydata)

depois: hist(MyModel\$residuals)

- ➤ Observação:
 - ✓ Não há necessidade de criar a interação ou a variável lob (bp) fora do comando lm ().
 - $ex. Im(wt \sim ht*age + log(bp), data = mydata)$
 - ✓ faz a regressão de wt em ht e age; realiza a interação ht-por-age; e depois calcula o log (bp).

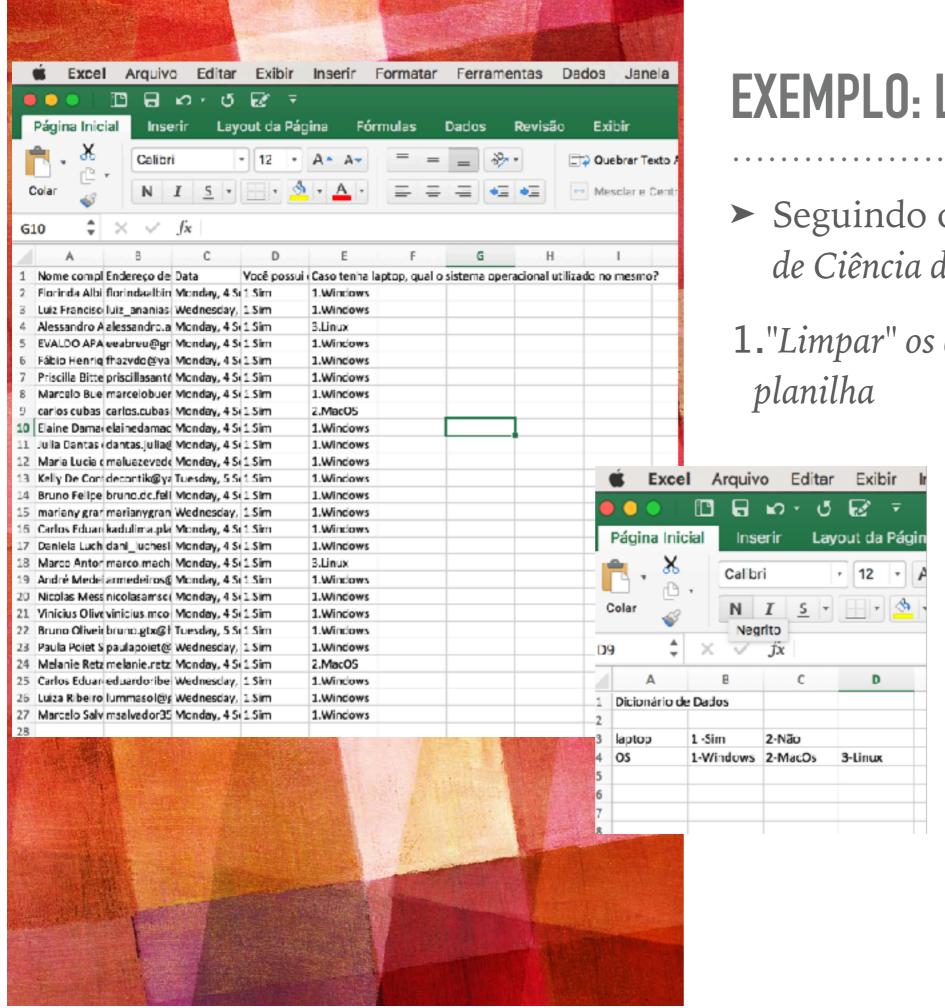
fluxo de ciência de dados Avaliar o experimento Dados "crus" Teste A/B Atualizar/ "límpeza" dos dados Aprimorar o modelo "Exploração" dos dados (Estatística básica. Desenvolvimento Aprender plotagem de gráficos, o modelo do "modelo" tendências, míneração de padrões) (Seleção de variáveis, seleção de atributos ou seleção de subconjunto de variáveis)

Domínío do Conhecimento Conhecimento especializado



EXEMPLO: LER ARQUIVO .XLS

- > Seguindo o processo do Fluxo de Ciência de Dados
- 1.Obter os dados: dados "crus" planilha da pesquisa do Moodle



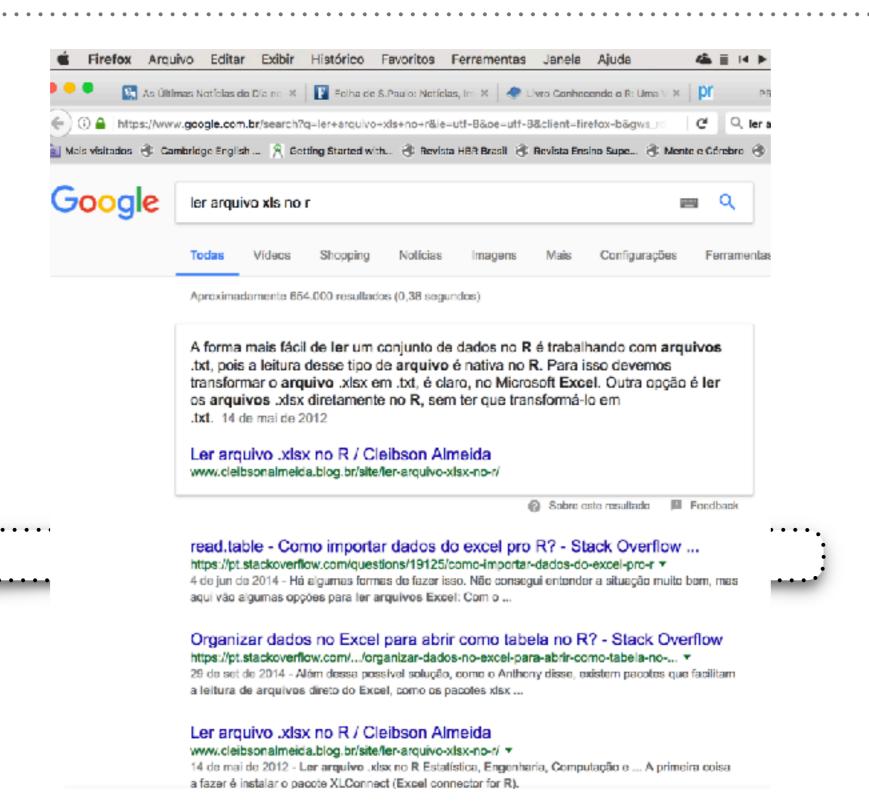
EXEMPLO: LER ARQUIVO .XLS

 Seguindo o processo do Fluxo de Ciência de Dados

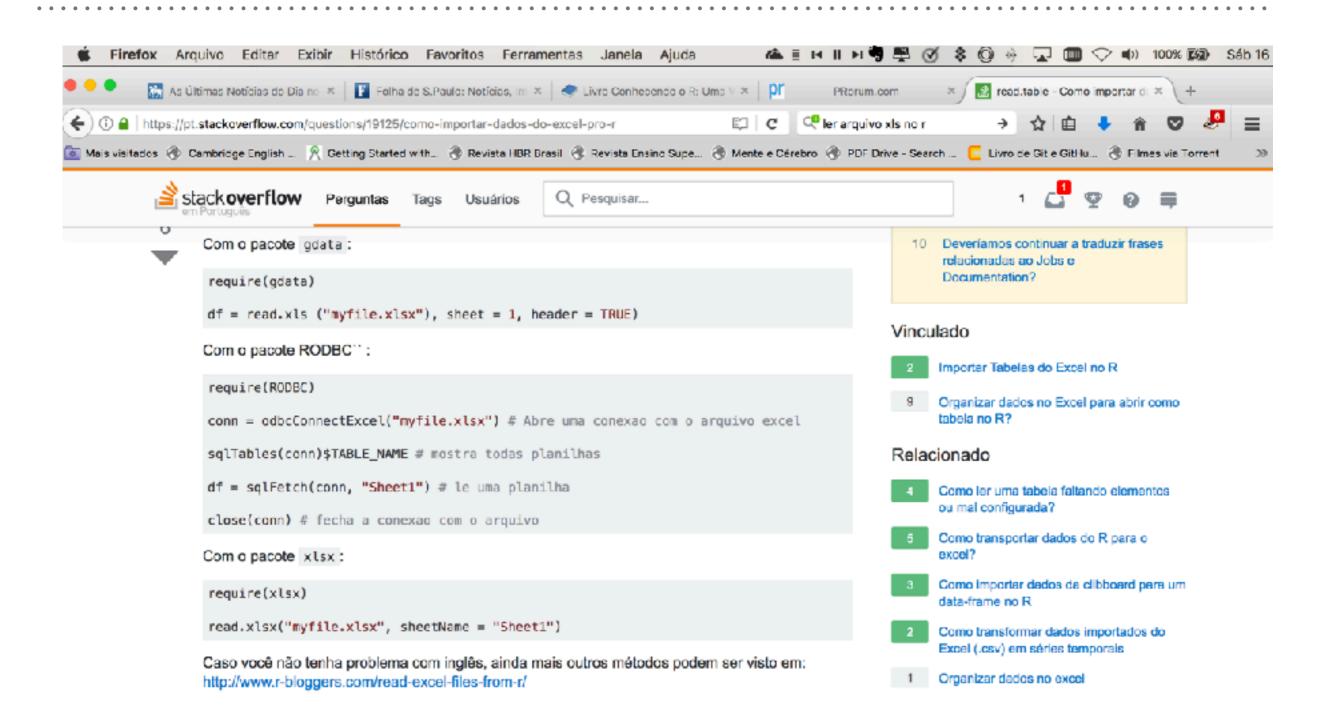
1."Limpar" os dados - formatar

- (Excel Arquivo Editar E	nidix	Ins
•	👓 🕛 🖪 အေး က ေပ 🗟	₹ ₹	
Ε,	Página Inicial Inserir Layout	do Bá	alaa
ı,	agilia Illicial Illiserii Layout	ua Pa	gilla
ď	Calibri 🔻	12 -	A
-		1	
С	olar N I S v	1 0	y w
F7	‡ × √ fx		
\overline{A}	Α	В	С
1	Nome completo	laptop :	50
2	Florinda Albino	1	1
3	Luiz Francisco Ananias Junior	1	1
4	Alessandro Ap. Antonio	1	3
5	EVALDO APARECIDO DE ABREU	1	1
6	Fábio Henrique de Azevedo	1.	1
7	Priscilla Bittencourt	1	1
8	Marcelo Bueno	1.	1
9	carlos cubas	1	2
10	Elaine Damaceno	1	1
11	Julia Dantas de Oliveira Penteado	1	1
12	Maria Lucia de Azevedo	1	1
13	Kelly De Conti Rodrigues	1.	1
14	Bruno Felipe	1	1
15	mariany granato	1.	1
16	Carlos Eduardo de Lima	1	1
1.7	Daniela Luchesi	1.	1
18	Marco Antonio Machado	1	3
19	André Medeiros	1	1
20	Nicolas Messias	1	1
21	Vinicius Oliveira	1	1
22	Bruno Oliveira	1	1
23	Paula Polet Sampedro	1	1
24	Melanie Retz Godoy dos Santos Zwicker	1.	2
25	Carlos Eduardo da Trindade Ribeiro	1	1
26	Luiza Ribeiro Mattar	1	1
27	Marcelo Salvador Celestino	1	1
28			

MODO MAIS FÁCIL: DR. GOOGLE



DR. GOOGLE



DR.GOOGLE #2



https://www.r-bloggers.com/read-excel-files-from-r/

CÓDIGO

```
# Localizando o arquivo .xls
setwd("/Users/jpalbino/Downloads/Data Science Material/Introdução à
Ciência de Dados/dados")
# No R, para utilizar determinadas funções, é preciso "carregar" um "pacote"
# Neste caso, carregamos o "pacote" xlxs diretamente do CRAN-R através do
comando library()
# Caso o mesmo ainda não esteja transferido o comando é:
## install.packages("xlsx")
library(xlsx)
df <- read.xlsx("sobre seu laptop v2.xlsx", sheetName = "respostas")
```

SAÍDAS

> library(xlsx)

11

Carregando pacotes exigidos: xlsx

Carregando pacotes exigidos: rJava

Carregando pacotes exigidos: xlsxjars

> read.xlsx("sobre_seu_laptop_v2.xlsx", sheetName = "respostas")

 $Error\ in\ read.xlsx("sobre_seu_laptop_v2.xlsx",\ sheetName = "respostas"):$

1 1

Cannot find the sheet you requested in the file!

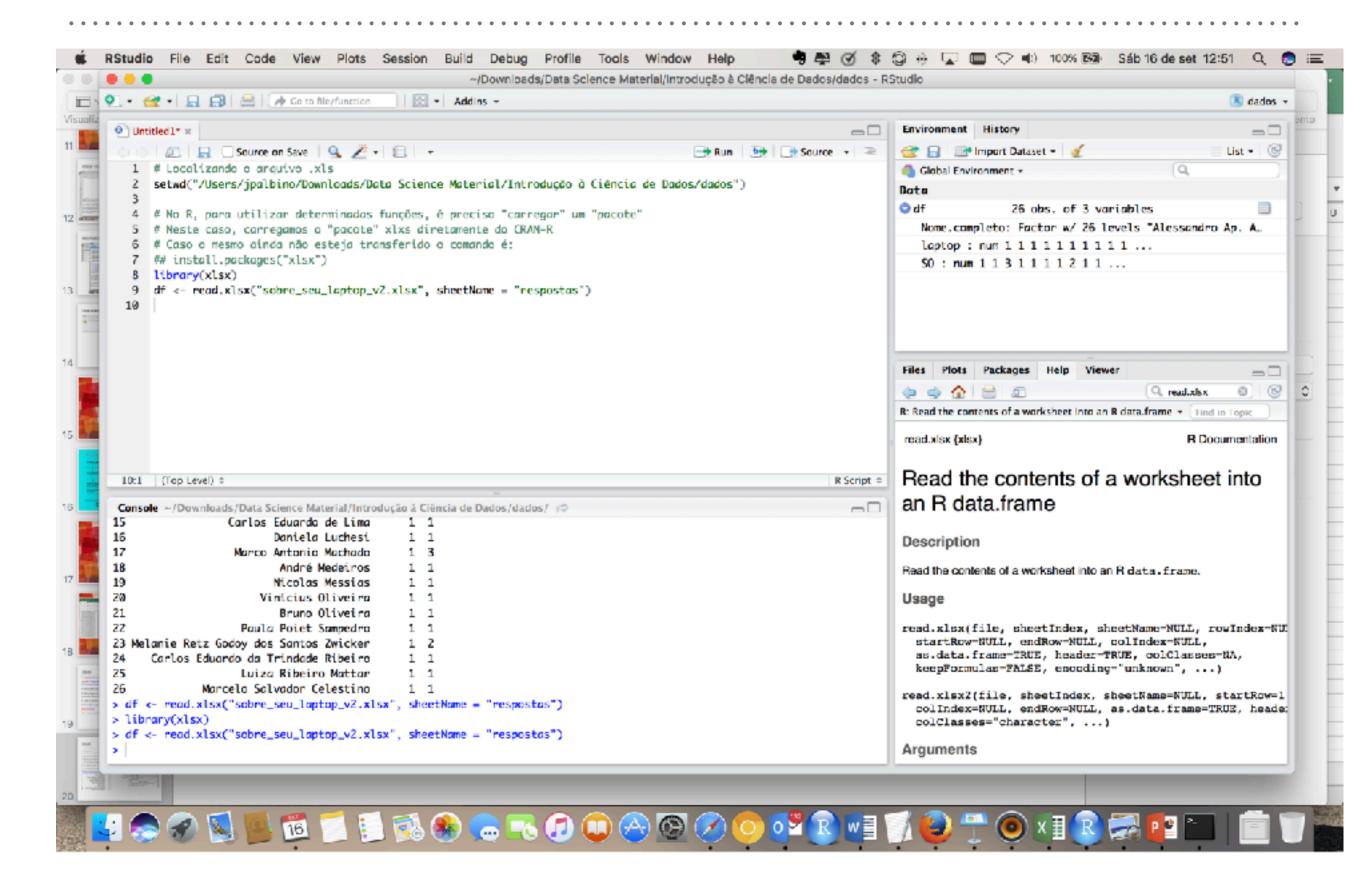
Maria Lucia de Azevedo

```
- In ( 100% [2] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3] ■ 100% [3 ■ 100% [3] ■ 100% [3] ■ 100% [3 ■ 100% [3] ■ 100% [3 ■ 100% [3] ■ 100% [3 ■ 100% [3] ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100% [3 ■ 100%
                                                                                                                                                                                                                                                                                                                                                                                                                             Sáb 16 de set 12:49
   Console ~/Downloads/Data Science Material/Introdução à Ciência de Dados/dados/
                                                                                                                                                                                                                                                                   de Dados/dados - RStudio
> read.xlsx("sobre_seu_laptop_v2.xlsx", sheetName = "respostas")
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    dados •
                                                                                                      Nome.completo laptop SO
                                                                                               Florinda Albino
                                                                                                                                                                                                                                                                                                                                                           History
                                                                                                                                                                                                                                                                                                   \neg \Box
                                                                                                                                                                                                                                                                                                                          Environment
                                           Luiz Francisco Ananias Junior
                                                                                                                                                                                                                                                                    ♦ Source → 🖃
                                                                                                                                                                                                                                                                                                                                       | Import Dataset -

    List → 

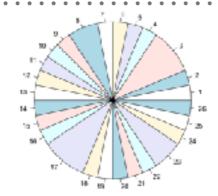
 3
                                                                    Alessandro Ap. Antonio
                                                                                                                                                                                                                                                                                                                          Global Environment •
                                                         EVALDO APARECIDO DE ABREU
                                                                                                                                                                                                                                                                                                                         Data
 5
                                                         Fábio Henrique de Azevedo
                                                                                                                                                                                                                                                                                                                         o df
                                                                                                                                                                                                                                                                                                                                                                             26 obs. of 3 variables
                                                                        Priscilla Bittencourt
                                                                                                                                                                                                                                                                                                                                   Nome.completo: Factor w/ 26 levels "Alessandro Ap. A..
 7
                                                                                                      Marcelo Bueno
                                                                                                                                                                              1 1
                                                                                                                                                                                                                                                                                                                                   laptop : num 1 1 1 1 1 1 1 1 1 1 ...
 8
                                                                                                          carlos cubas
                                                                                                                                                                                                                                                                                                                                    50 : num 1 1 3 1 1 1 1 2 1 1 ...
 9
                                                                                               Elaine Damaceno
                                                                                                                                                                              1 1
 10
                           Julia Dantas de Oliveira Penteado
                                                                                                                                                                              1 1
```

AMBIENTE DO RSTUDIO



ALGUNS GRÁFICOS

Alguns gráficos básicos # Tipo pizza pie(df\$SO)



Para melhorar a visualização, usaremos a função table() para subdividir o gráfico em fatores.

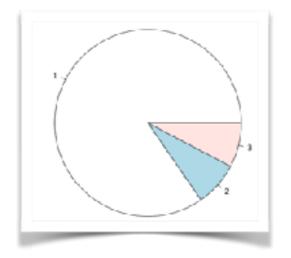
table() usa fatores de classificação cruzada (classificação de acordo com mais de um atributo ao mesmo tempo) para construir uma

tabela de contingência (tabela estatística que mostra as freqüências dos dados, classificados de acordo com duas variáveis:

as linhas indicam uma variável e as colunas indicam outra variável), das contagens em cada combinação de níveis de fatores.

> table(df\$SO)

pie(table(df\$SO))



pie(table(df\$SO), labels = c("1 = Windows", "2 = MacOs", "3 = Linux"), edges = 200, radius = 0.8, density = NULL, angle = 45, col = NULL, border = NULL, lty = NULL, main = "Sistemas Operacionais")

Sistemas Operacionais

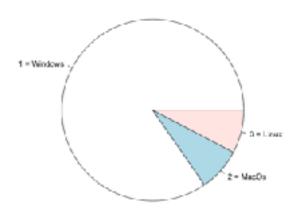


GRÁFICO MELHORADO

> table(df\$SO)

```
# | 1 2 3 | (fatores, onde 1 = Windows, 2 = MacOs e 3 = Linux)
# | 22 2 2 | (frequências relativas dos dados/fatores)
```

lbls <- c("Windows", "MacOs", "Linux") pct <- round(table(df\$SO)/sum(table(df\$SO))*100, digits=1) #> pct

```
lbls <- paste(lbls, pct) # add percents to labels
lbls <- paste(lbls,"%",sep="") # ad % to labels
pie(table(df$SO), labels = lbls, edges = 200, radius = 0.8,
    clockwise = TRUE, density = 60, angle = 45, col = c("purple", "green3","cyan"),
    border = NULL, lty = NULL,
    main = "Sistemas Operacionais Utilizados pelos Alunos")
```

Sistemas Operacionais Utilizados pelos Alunos

