



Data Science Academy

Big Data Analytics com R e Microsoft Azure Machine Learning Módulo 2



Data Science Academy



R Fundamentos Parte 1



Data Science Academy



Seja Bem-Vindo



Data Science Academy



Introdução



Data Science Academy

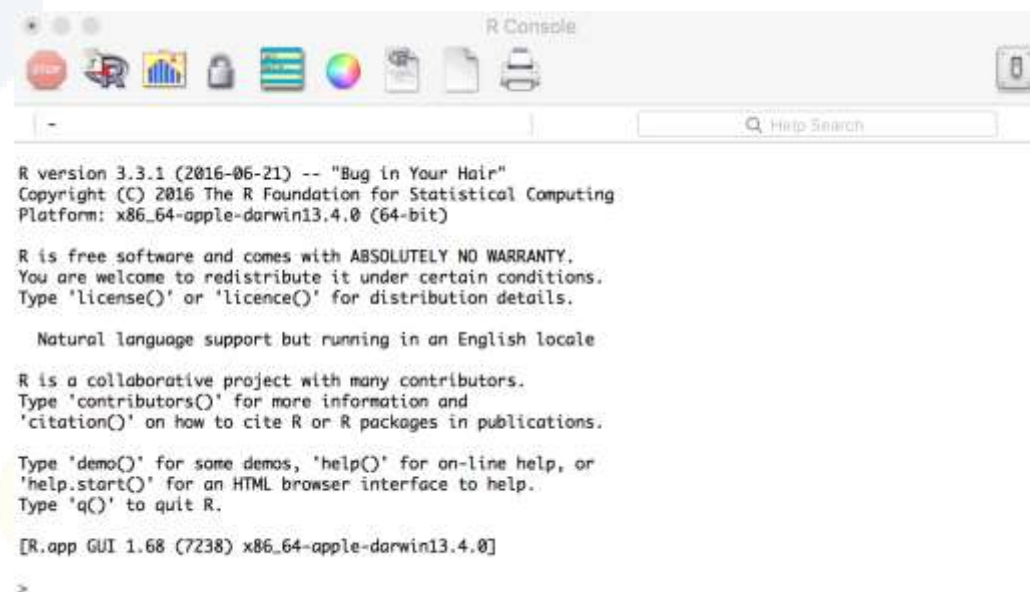
O que é a linguagem R?

- Linguagem de programação estatística
- Os objetos podem ser vetores, matrizes, funções, arquivos, conjuntos de dados, etc...
- A manipulação pode ser cálculo, entrada e saída de dados, análises estatísticas, gráficos, etc...
- Gratuita (open-source)
- Grande variedade de pacotes e bibliotecas (<https://cran.r-project.org>)
- Comunidade ativa
- Interativa



Data Science Academy

O que é o RStudio?



R Console

R version 3.3.1 (2016-06-21) -- "Bug in Your Hair"
Copyright (C) 2016 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin13.4.0 (64-bit)

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

Natural language support but running in an English locale

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

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

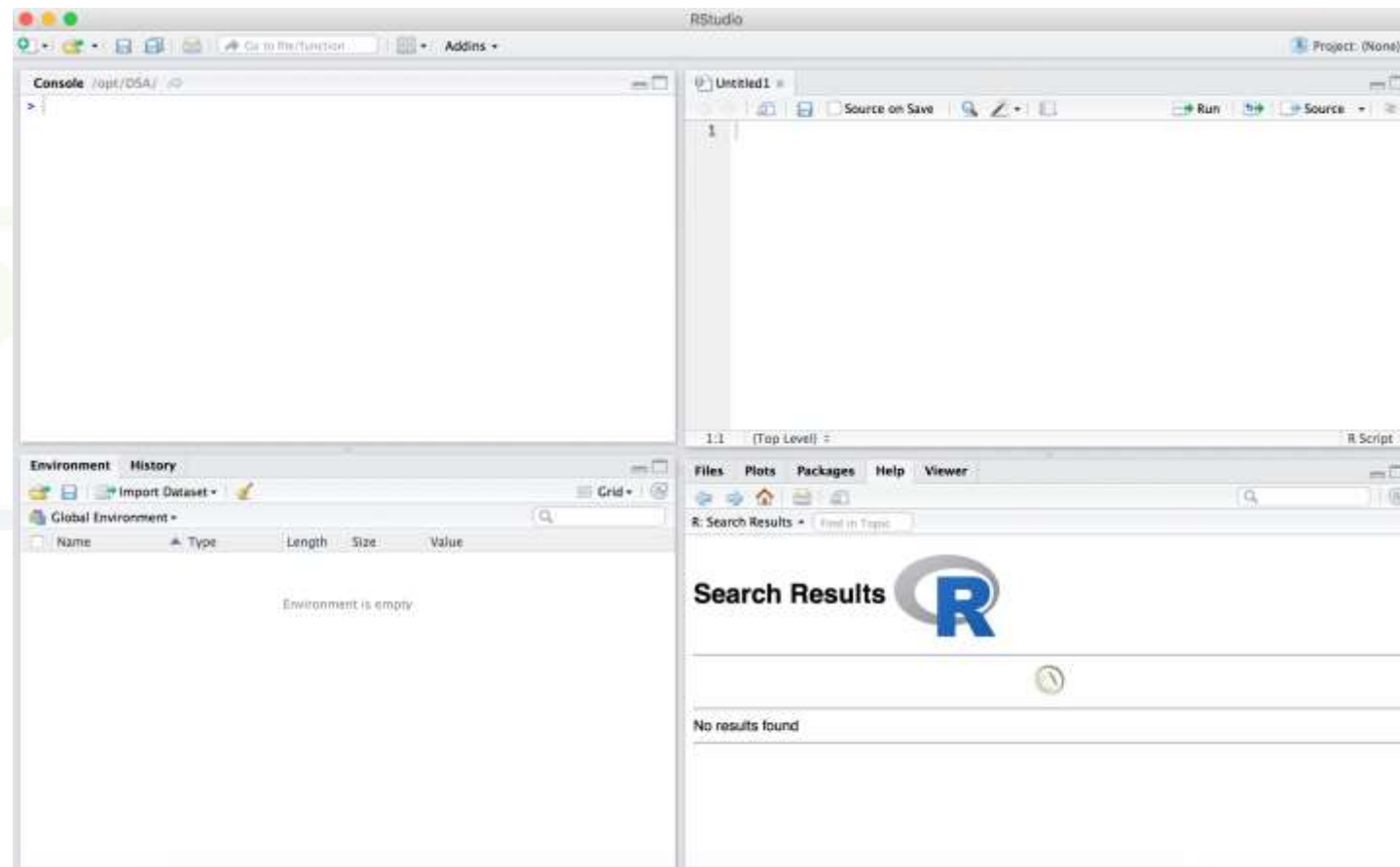
[R.app GUI 1.68 (7238) x86_64-apple-darwin13.4.0]

>



Data Science Academy

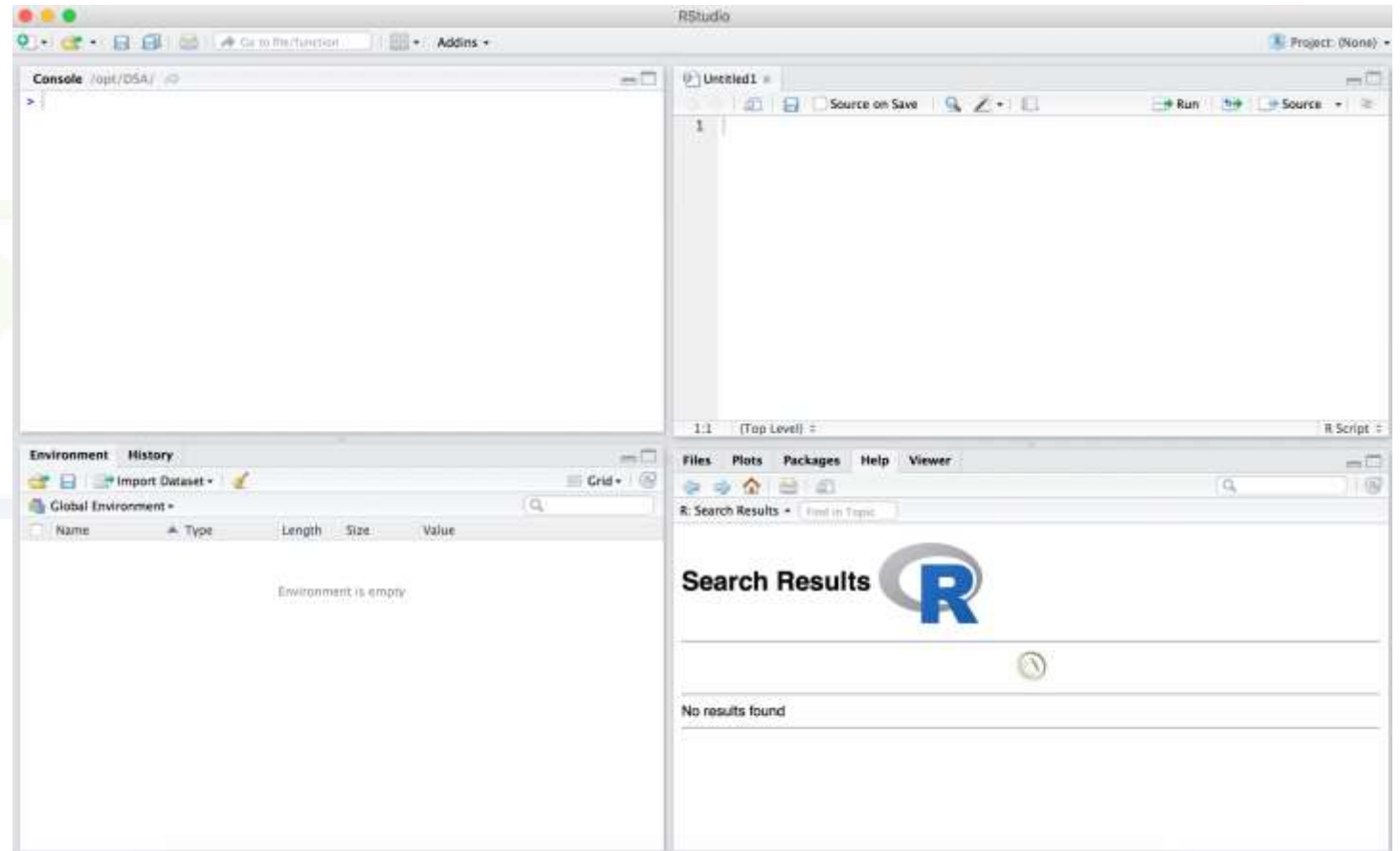
O que é o RStudio?




Data Science Academy

O que é o RStudio?

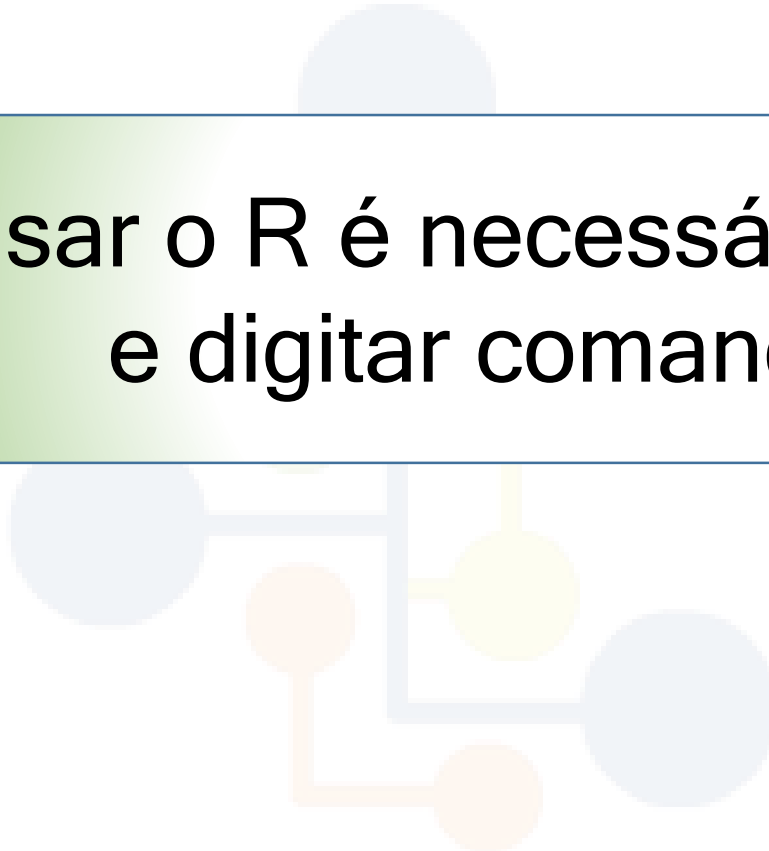
- IDE (Integrated Development Environment)
- Gratuito
- Disponível para os principais sistemas operacionais




Data Science Academy



Para usar o R é necessário conhecer
e digitar comandos



Data Science Academy

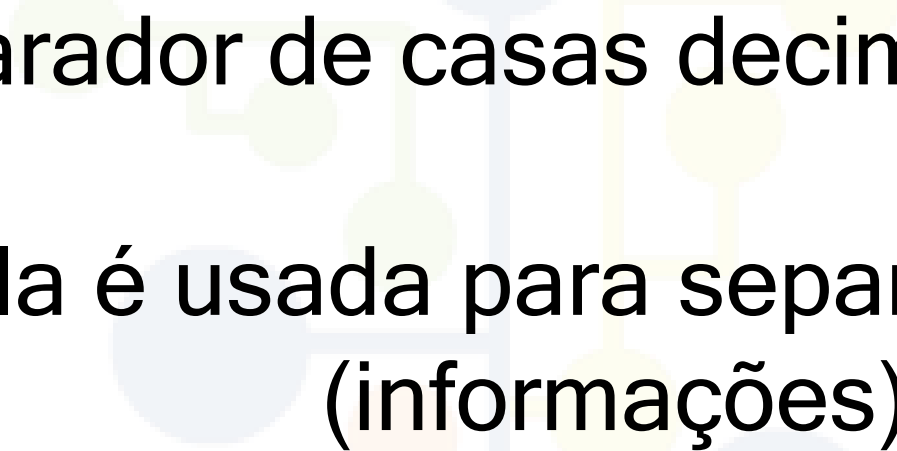



O R é case-sensitive, isto é, ele diferencia letras maiúsculas de minúsculas.

Portanto A é diferente de a.



Data Science Academy



O separador de casas decimais é ponto "."

A vírgula é usada para separar argumentos
(informações)



Data Science Academy



Evite usar acentos



Data Science Academy



Workspace (Área de Trabalho)




Data Science Academy

Área de trabalho

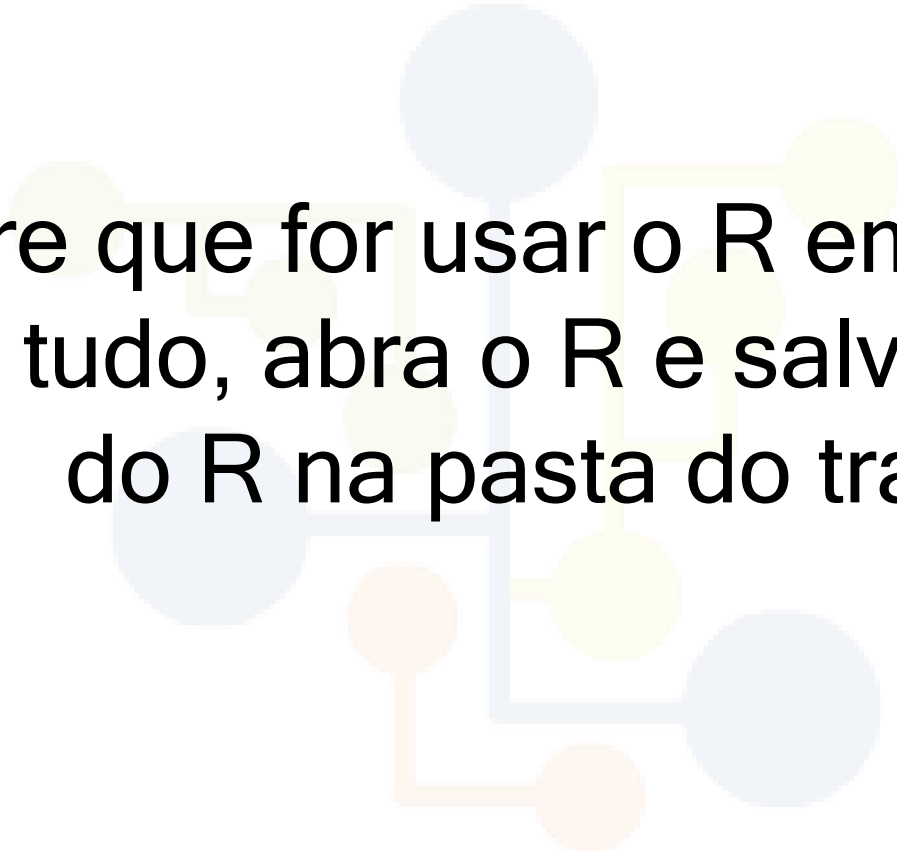
getwd()
setwd()



Data Science Academy



Sempre que for usar o R em um trabalho,
antes de tudo, abra o R e salve um workspace
do R na pasta do trabalho



Data Science Academy

Pacotes
(packages)

CRAN
Comprehensive R
Archive Network

<https://www.r-project.org>



[Home]

Download

CRAN

R Project

About R

Logo

Contributors

What's New?

Mailing Lists

Bug Tracking

Development Site

Conferences

Search

R Foundation

Foundation

Board

Members

Donors

Donate

Documentation

Manuals

FAQs

The R Journal

Books

Certification

Other

The R Project for Statistical Computing

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To **download R**, please choose your preferred [CRAN mirror](#).

If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

News

- **R version 3.3.1 (Bug in Your Hair)** has been released on Tuesday 2016-06-21.
- **R version 3.2.5 (Very, Very Secure Dishes)** has been released on 2016-04-14. This is a rebadging of the quick-fix release 3.2.4-revised.
- **Notice XQuartz users (Mac OS X)** A security issue has been detected with the Sparkle update mechanism used by XQuartz. Avoid updating over insecure channels.
- The **R Logo** is available for download in high-resolution PNG or SVG formats.
- **useR! 2016**, will take place at Stanford University, CA, USA, June 27 - June 30, 2016.
- **The R Journal Volume 7/2** is available.
- **R version 3.2.3 (Wooden Christmas-Tree)** has been released on 2015-12-10.
- **R version 3.1.3 (Smooth Sidewalk)** has been released on 2015-03-09.



Data Science Academy




Lista de Pacotes R

<https://cloud.r-project.org>

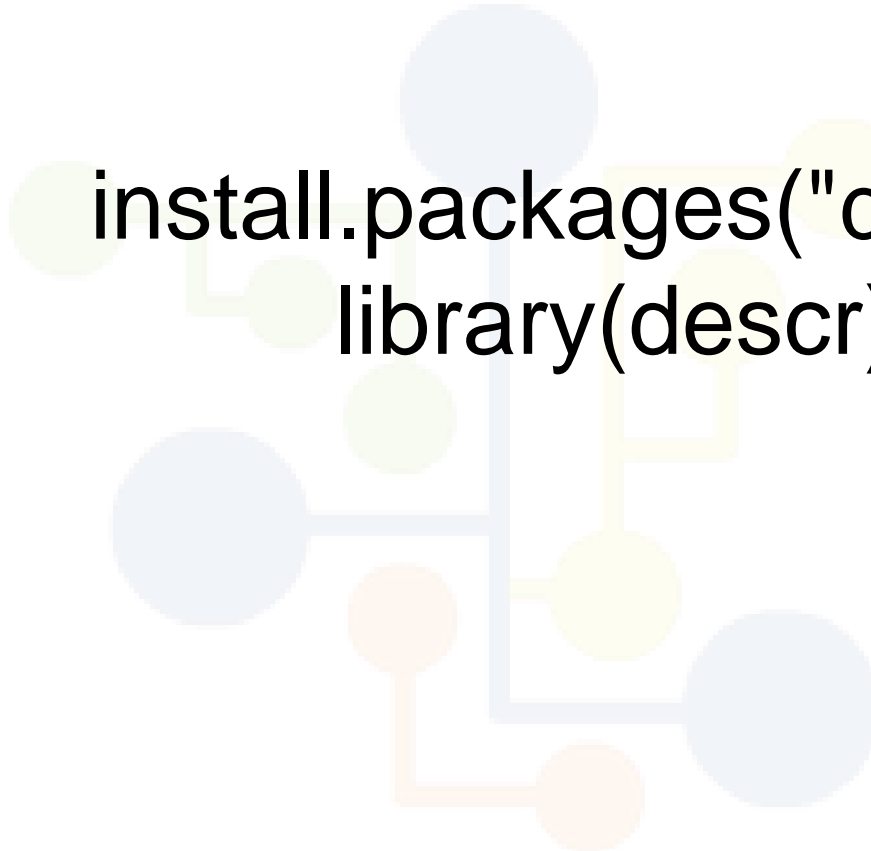
(Clique na opção Packages no menu)



Data Science Academy



```
install.packages("descr")  
library(descr)
```



Data Science Academy



Help



`help.search()`

`??nome`

`RSiteSearch()`

`example()`



Data Science Academy



Demonstrações

Algumas funções em R, possuem demonstrações de uso

`demo()`



Data Science Academy



Sessão

Informações sobre a sessão

`sessionInfo()`



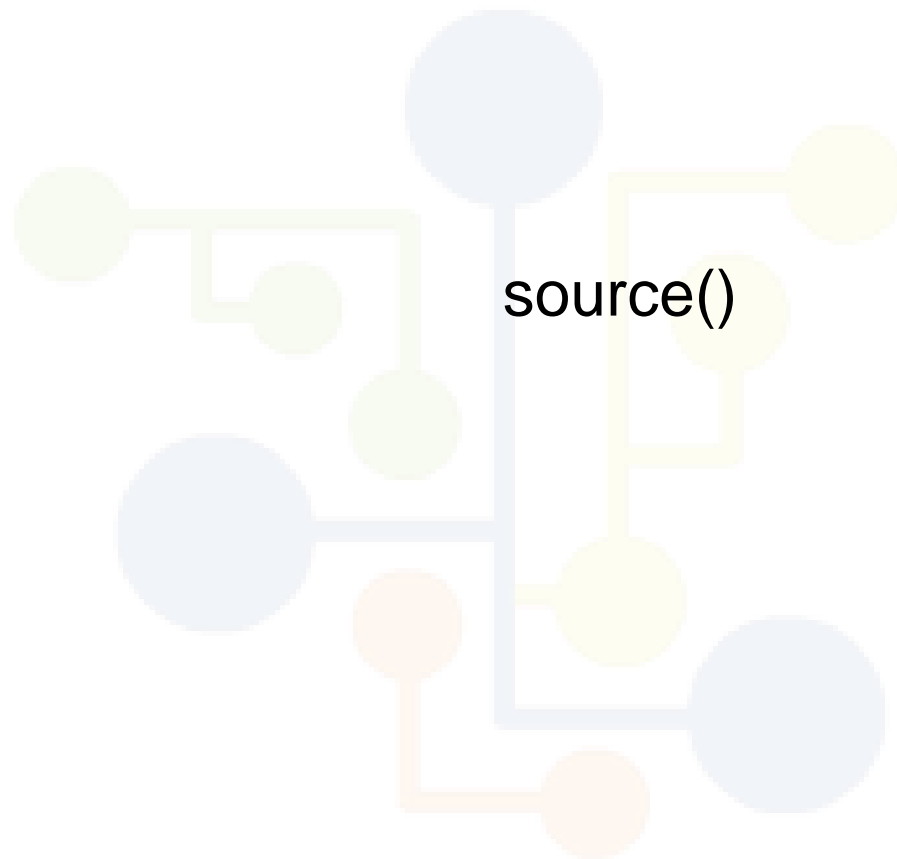
Data Science Academy

Imprimir na Tela

`print()`



Data Science Academy



Data Science Academy



Data Science Academy



R Fundamentos Parte 1



Data Science Academy



Explorando o RStudio



Data Science Academy

- Interface do R (Console R)
- IDE (Integrated Development Environment) - RStudio
- Configuração da área de trabalho do RStudio
- Customização da visualização
- Criação de scripts
- Execução de código
- Construção de gráficos
- Consulta da documentação
- Instalação de pacotes
- Autocomplete





Por que você deve usar scripts R?

script.R



Data Science Academy



Histórico

função()

Objetos são salvos em .RData
Histórico de comandos em .Rhistory

quit() ou q()



Data Science Academy



Operadores



Data Science Academy

Operadores Matemáticos

Operador	Operação
+	Soma
-	Subtração
*	Multiplicação
/	Divisão
^	Potência
%%	Módulo

Use parênteses para separar partes dos cálculos



Data Science Academy

Operadores Relacionais

Operador	Operação
>	Maior
<	Menor
>=	Maior ou igual
<=	Menor ou igual
==	Igual (comparação)
!=	Diferente



Operadores Lógicos

Operador	Operação
&	and
	or
!	not
TRUE (ou 1)	Valor booleano verdadeiro(1)
FALSE (ou 0)	Valor booleano falso (0)






Tipos de Dados



Data Science Academy



Numérico
1, 5.9, -134

Caractere

Tipos de
Dados

Complexo

Lógico



Numérico

Caractere
"A", "bolha"

Tipos de
Dados

Complexo

Lógico



Numérico

Caractere

Tipos de
Dados

Complexo

$$z = x + yi$$

Lógico



Numérico

Caractere

Tipos de
Dados

Complexo

Lógico
True/False

Data Science Academy



Valores faltantes e especiais

NA - Not Available

NaN - Not a Number

Inf e -Inf



Data Science Academy



Variáveis



Data Science Academy



Variáveis

nome_da_variável = valor



Data Science Academy



Variáveis

Matemática
Programação
Estatística



Data Science Academy



Palavras Reservadas



FALSE Inf NA NaN NULL TRUE
break else for function if in next repeat
while



Data Science Academy



Tipos de Objetos



Data Science Academy



O que são objetos?



Data Science Academy

O que são objetos?

- Vetores
- Matrizes
- Listas
- Dataframes
- Funções





Vetores



Data Science Academy



Vetores

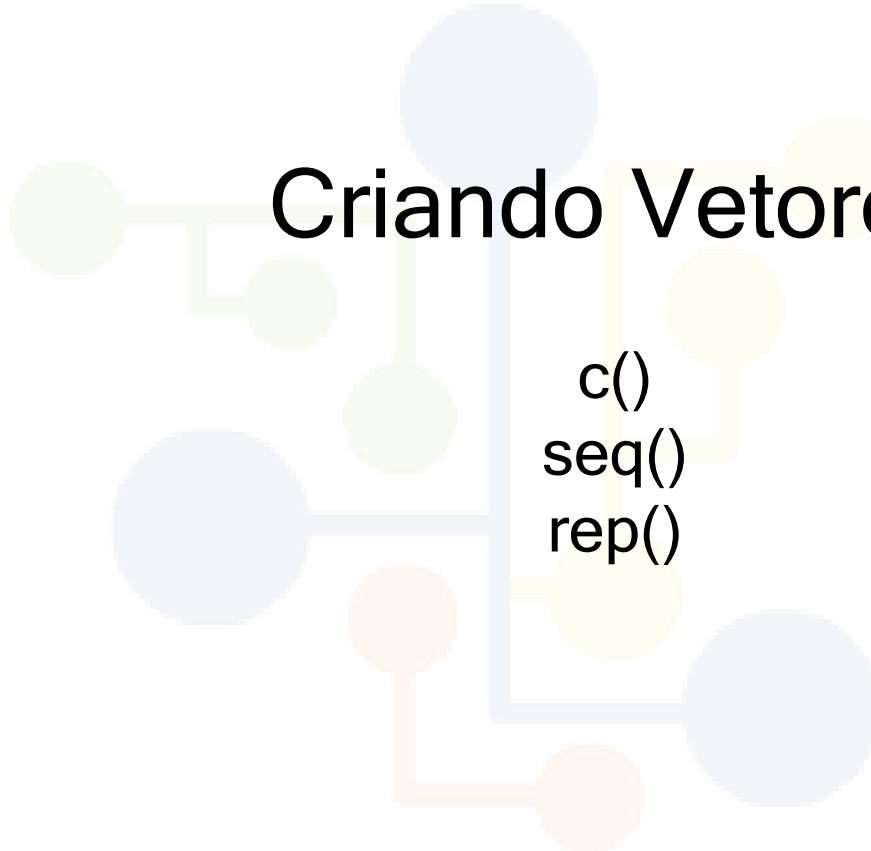
Conjunto de elementos de uma mesma natureza



Data Science Academy



Criando Vetores



`c()`
`seq()`
`rep()`



Data Science Academy



Matrizes



Data Science Academy



Matrizes

Conjuntos de elementos de uma mesma natureza, organizados em linhas e colunas



Data Science Academy



Array

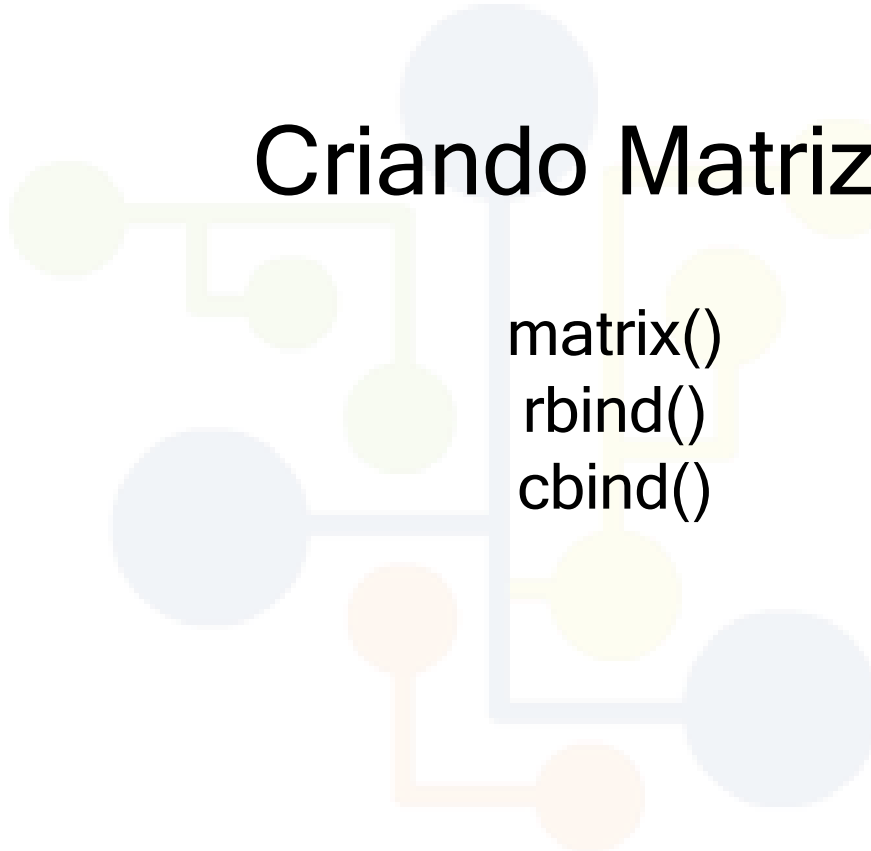
Generaliza o conceito de Matriz. Enquanto na matriz os elementos são organizados em 2 dimensões (linhas e colunas), em um array os elementos podem ser organizados em um número arbitrário de dimensões



Data Science Academy



Criando Matrizes



`matrix()`
`rbind()`
`cbind()`



Data Science Academy



Criando Arrays

`array()`



Data Science Academy



Listas



Data Science Academy



Listas

Generalizações de vetores, representam uma coleção de objetos



Data Science Academy



Criando Listas

`list()`



Data Science Academy



Strings




Data Science Academy



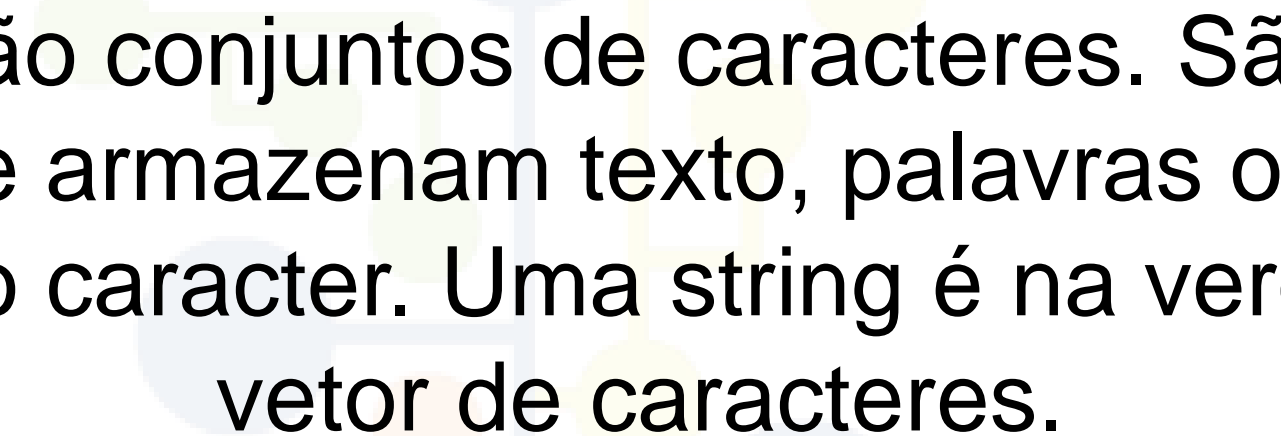
Strings



Data Science Academy



Strings são conjuntos de caracteres. São objetos em R que armazenam texto, palavras ou mesmo um único caracter. Uma string é na verdade um vetor de caracteres.





DataFrames



Data Science Academy



DataFrames

Similar a matrizes, porém cada coluna pode possuir elementos de natureza diferente



Data Science Academy



Criando DataFrames

`data.frame()`



Data Science Academy

Criando DataFrames

`read.xls()` - Excel

`read.mtp()` - Minitab

`read.spss()` - SPSS

`read.table()` - Arquivos txt

`read.csv()` - Arquivos csv

`read.delim()` - Leitura de arquivos delimitados



Data Science Academy



Estruturas de Controle



Data Science Academy

Condicionais If-Else


```
if(condição){  
    conjunto de tarefas}  
else  
    {outro conjunto de tarefas}  
ifelse(condição, tarefa1, tarefa2)
```



Data Science Academy



Loop For



```
for(i in 1:N){  
conjunto de tarefas}
```



Data Science Academy



Loop For



```
for(i in 1:N){  
conjunto de tarefas}
```

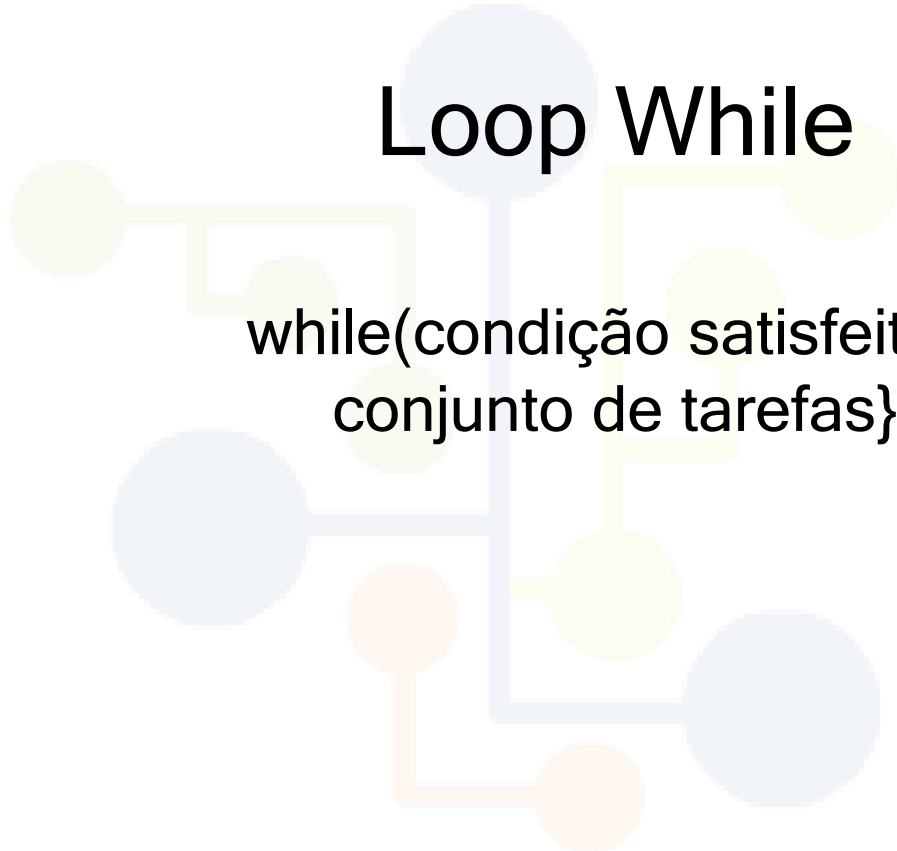
```
for(i in 1:10){  
print('Isso é um loop')}
```



Data Science Academy



Loop While



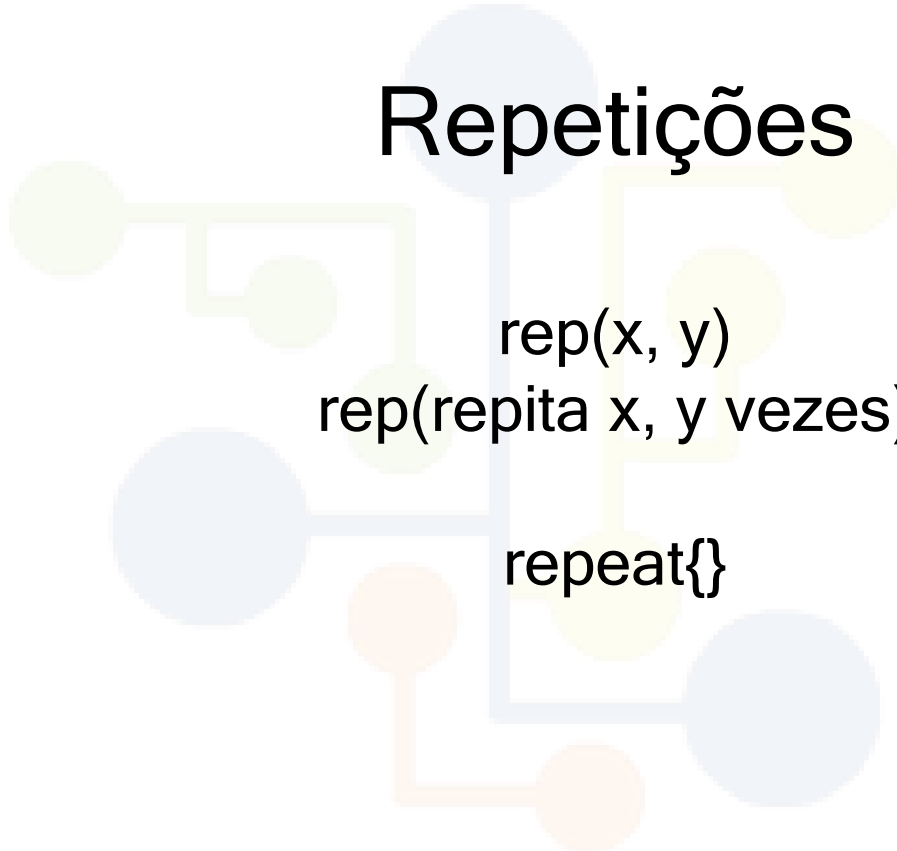
```
while(condição satisfeita){  
conjunto de tarefas}
```



Data Science Academy



Repetições



`rep(x, y)`
`rep(repita x, y vezes)`

`repeat{}`



Data Science Academy



O que nós vimos até aqui?





Data Science Academy



Cansei por hoje...



Data Science Academy

- 
- 
- Tivemos uma visão geral do RStudio
 - Operadores Aritméticos, Relacionais e Lógicos
 - Tipos de Dados
 - Variáveis
 - Tipos de Objetos
 - Vetores
 - Matrizes
 - Listas
 - Strings
 - DataFrames
 - Estruturas de Controle





Obrigado!



Data Science Academy

Curta Nossas Páginas nas Redes Sociais

E fique sabendo das novidades em Data Science, Big Data, Internet das Coisas e muito mais...



www.facebook.com/dsacademybr



twitter.com/dsacademybr



www.linkedin.com/company/data-science-academy

www.datascienceacademy.com.br



Data Science Academy



www.datascienceacademy.com.br

www.datascienceacademy.com.br



Data Science Academy