



Data Science Academy

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



Data Science Academy

www.datascienceacademy.com.br



R Fundamentos Parte 1



Data Science Academy

www.datascienceacademy.com.br



O que é a linguagem R?

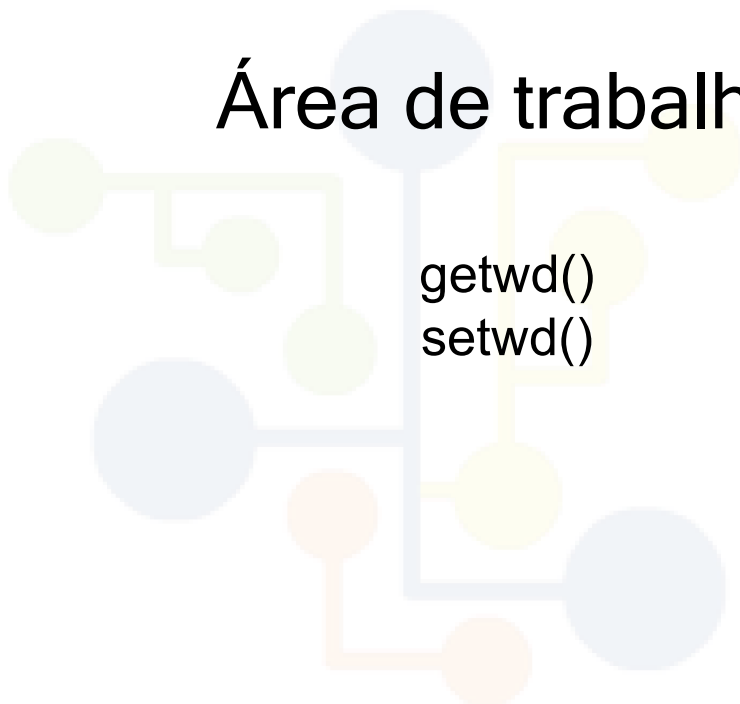
- Linguagem de programação estatística
- Os objetos podem ser vetores, matrizes, funções, arquivos, conjuntos de dados, etc...
- A manipulação pode ser cálculo, entrada e saída de dados, análises estatísticas, gráficos, etc...
- Gratuita (open-source)
- Grande variedade de pacotes e bibliotecas (<https://cran.r-project.org>)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



Data Science Academy




Área de trabalho



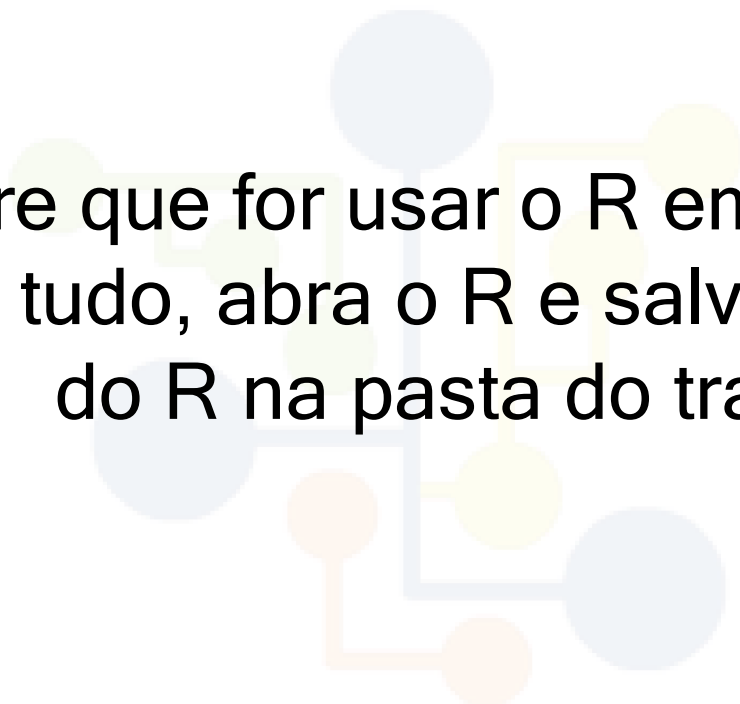
getwd()
setwd()



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

www.datascienceacademy.com.br



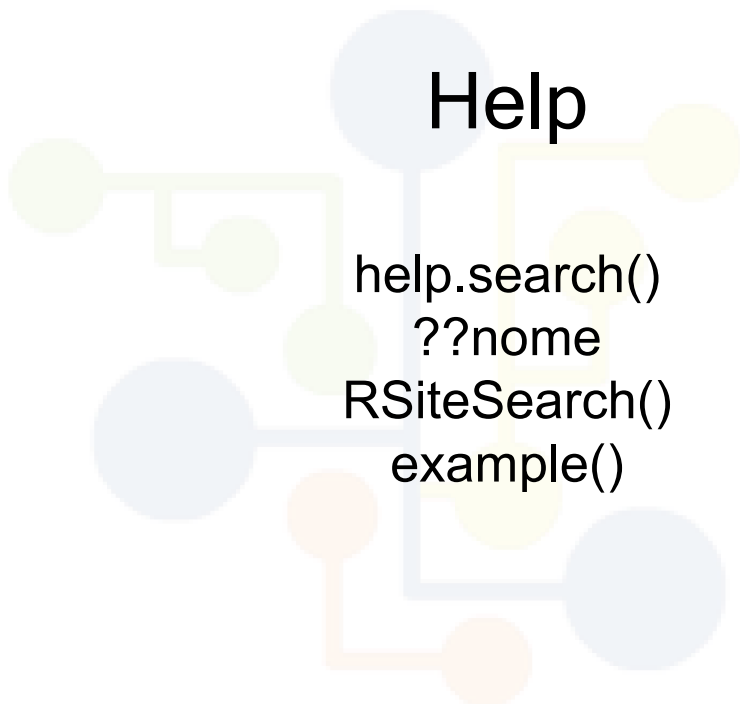
```
install.packages("descr")  
library(descr)
```



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



Histórico

função()

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

quit() ou q()



Data Science Academy



Operadores



Data Science Academy

www.datascienceacademy.com.br

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



Data Science Academy

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

Data Science Academy

www.datascienceacademy.com.br



Valores faltantes e especiais

NA - Not Available

NaN - Not a Number

Inf e -Inf



Data Science Academy



Tipos de Objetos



Data Science Academy

www.datascienceacademy.com.br



Vetores

Conjunto de elementos de uma mesma natureza



Data Science Academy



Matrizes

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



Data Science Academy



Array

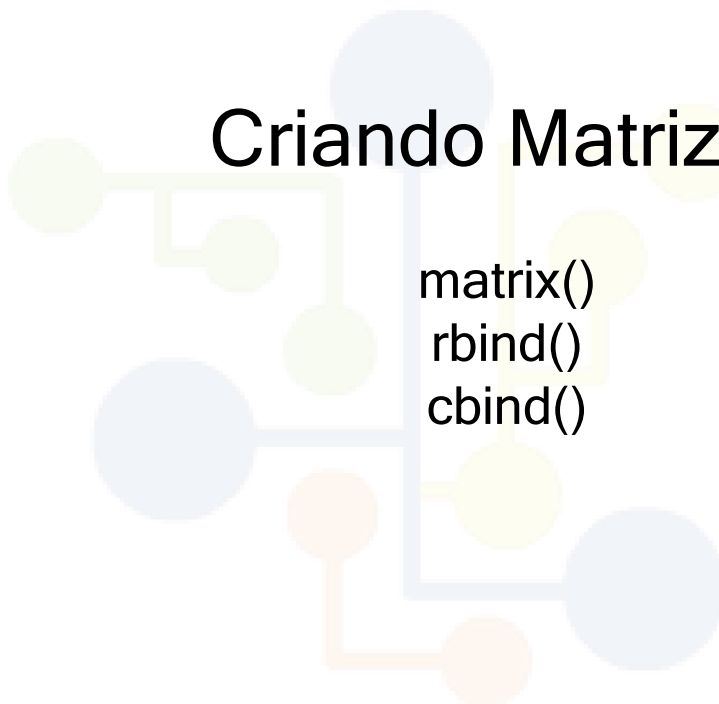
Generaliza o conceito de Matriz. Enquanto na matriz os elementos são organizados em 2 dimensões (linhas e colunas), em um array os elementos podem ser organizados em um número arbitrário de dimensões



Data Science Academy



Criando Matrizes



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



Data Science Academy



Criando Arrays

`array()`



Data Science Academy



Listas

Generalizações de vetores, representam uma coleção de objetos



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