

Programação Estatística

Introdução ao R - Iniciação

Rachid Muleia, PhD in Statistics

Universidade Eduardo Mondlane
Faculdade de Ciências
Departamento de Matemática e Informática

2023-02-24

- 1 Introdução
- 2 Organizar a sessão R
- 3 R como calculadora
- 4 Como obter ajuda
- 5 Directórios
- 6 Objectos

Introdução

Resultados de aprendizagem

- Entender os conceitos básicos de programação, como tipos de dados, estruturas de dados e indexação
- Saber aplicar funções básicas em R
- Saber importar, manipular e resumir dados em R
- Conceptualizar e criar estruturas de control para resolução de vários problemas
- Saber criar próprias funções “personalizadas”
- Saber criar gráficos para diferentes tipos de dados

O que é o R

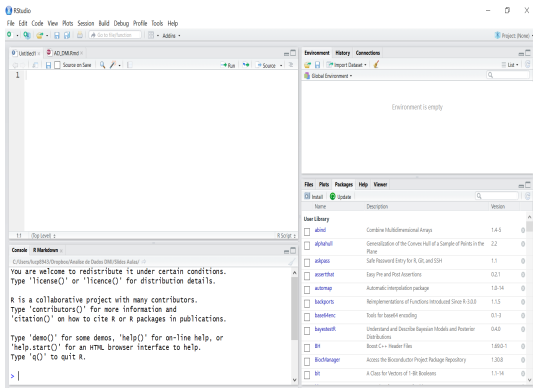
- R é um programa para computação estatística e visualização gráfica de dados
- O programa pode ser executado em várias plataformas, UNIX, Windows, MacOS
- Acesso: <https://www.r-project.org/>

Porque usar o R?

- É um programa gratuito
- Novos métodos estatísticos, geralmente, são implementados primeiro em R
- Tem uma vasta comunidade de usuários
- Muitos artigos científicos incluem o código R

RStudio

- O Rstudio é um ambiente de desenvolvimento integrado
- O RStudio é gratuito e pode funcionar e vários sistemas operativos
- Acesso: <https://posit.co/download/rstudio-desktop/>



Linguagens compiladas C vs interpretadas R

- C precisam de um programa completo para serem executadas
 - O programa é traduzido para um código de máquina
 - São mais rápidas
- R pode ser executado de forma interactiva
 - Os interpretadores passam por um programa linha por linha e executam cada comando
 - As linguagens interpretadas são mais flexíveis. Consequentemente, são lentas

Empresas que usam o R

Companies That Use R	Who Uses R at This Company?	What Does This Company Use R For?	Estimated Number of Employees
Accenture	R programming application developer, data scientist, full stack developer	Applications software development, statistical analysis, exploratory data analysis	<u>674,000</u>
Amazon	Senior data scientist, research scientist II, transportation analyst	Statistical analysis, exploratory data analysis, machine learning	<u>1,298,000</u>
Cognizant	Data engineer, data scientist, data analyst	Exploratory data analysis, statistical analysis	<u>289,500</u>
Deloitte Consulting	Data visualization developer, data scientist, backend developer	Machine learning, exploratory data analysis	<u>345,000</u>
Google	Data scientist, cloud AI engineer, analytical consultant	Exploratory data analysis, machine learning	<u>135,301</u>
HCL Technologies	Technical lead, data scientist	Machine learning	<u>169,000</u>
Infosys	Data science consultant, data science	Machine learning, exploratory data analysis	<u>260,000</u>
Meta Platforms	Data science manager, people data solutions analyst, data Scientist	Statistical analysis, exploratory data analysis	<u>58,604</u>
Novartis	Senior principal real-world evidence (RWE) research analyst, senior data scientist, exploratory data analyst	Exploratory data analysis, machine learning, statistical analysis	<u>105,794</u>
Snapdeal	Product analyst, data scientist	Machine learning, statistical analysis	<u>3,661</u>

Referências

- Peng, R.D. (2020). R Programming for Data Science.
<https://bookdown.org/rdpeng/rprogdatascience/>
- Phillips, N.D. (2018). YaRrr, The Pirate's Guide to R.
<https://bookdown.org/ndphillips/YaRrr/>
- Mahoney, M. (2019). Introduction to Data Exploration and Analysis with R.
<https://bookdown.org/mikemahoney218/IDEAR/>
- Golemund, G. and Wickham, H. (2019). R for Data Science.
<https://r4ds.had.co.nz/>
- Wickham, H. (2018). The tidyverse styleguide.
<https://style.tidyverse.org/>

Organizar a sessão R

Sessão R - como estruturar

- Primeiro inicializar o RStudio.
- Abra um script novo ou pré-existente no editor de texto ou Rstudio (extensão .R)
- Especificar o directorio de trabalho usando o RStudio no campo session ou `setwd('directorio')`. Use `getwd()` para verificar
- Carregar todas as livrarias/pacotes
 - `install.packages("nome_da_livraria")`
 - chamar para o ambiente de trabalho
`library(nome_da_livraria)`
- Sempre comente o seu código usando #
- Os outputs serão gravados no seu directório de trabalho (caso não especifique outro directório)

R como calculadora

Operações básicas

```
> 1+2 # adição
> 4-1 # subtração
> 2*4 # multiplicação
> 8/4 # divisão
> sqrt(4) # raiz quadrada
```

****ATENÇÃO****: inteiro vs resto da divisão

- `5 %/ 3` # inteiro da divisão ---> 1
- `5 %% 3` # resto da divisão ---> 2

Funções matemáticas importantes

```
> exp(1)
> exp(log(5))
> sin(pi/2)
> cos(pi/2)
> max(4,2,5,1)
> min(4,2,5,1)
> sum(4,2,5,1)
> prod(4,2,5,1)
> sqrt(16)
> factorial(4)      #factorial de 4 ---> 4*3*2*1
> choose(5,2)
```

■
$$\frac{n!}{k!(n-k)!}$$

Mais funções

`log10()`, `log2()`, `tan()`, `asin()`, `acos()`, `atan()`, `sinh()`, `cosh()`, `tanh()`,
`asinh()`, `acosh()`, `atanh()`, `abs()`, `round()`, `floor()`, `ceiling()`, `trunc()`,
`signif()`...

- Visitar a documentação do R
- Construir funções personalizadas

Operadores lógicos

```
==      # igualdade-comparação  
<       # menor  
>       # maior  
<=     # menor ou igaul  
>=     # maior ou igual
```


Como obter ajuda

Documentação - buscando ajuda

Um dos recursos mais importantes que você precisará ao iniciar seu caminho para se tornar um usuário do R é obter ajuda.

```
help(max) #ajuda para a função max
?max #igual a help(max)
help("exp")
help.start()
help.search("max") # lista todos os comandos que podem estar
??max # igual a help.search(max)
example(exp) #exemplo de como usar a função 'exp'
demo('graphics') # apresenta uma demonstração mais holística
```

Comunidade R

The RStudio Community is in the process of transitioning. Soon, we will introduce a new Posit Community portal. It will serve as an umbrella to a wide set of community activities. This forum will then continue as Posit Forum.

To learn more about why RStudio became Posit PBC, check out our recent announcement: [RStudio is Now Posit](#).

If you catch any issues with these updates, please let us know in [#rstats](#).



2. Stack Overflow - R

This section of Stack Overflow is for asking questions about R programming language.



stackoverflow.com/ [+ Follow](#)

f 517.4K · w 230.6K · i 30 posts / day · o Sep 2008

Get Email Updates 

OR

3. Reddit - The Statistical Computing with R



San Francisco, California, US

This subreddit is for discussing R stories, questions, and news from the Reddit community.

reddit.com/r/rstats [+ Follow](#)

f 1.7M · w 1.1M · i 5 posts / day · o Oct 2009

Directórios

Directórios e ficheiros

- `getwd()`: informa o actual directório de trabalho
- `setwd()`: especificar novo directório de trabalho
- `list.files()` ou `dir()`: lista todos os ficheiros que estão no directório de trabalho
- `dir.create()`: cria um novo directório (pasta), dentro do seu directório de trabalho
- `file.create()`: cria um novo ficheiro
- `file.rename()`: permite renomear um dado ficheiro
- `file.copy()`: cria uma copia de um determinado ficheiro

Objetos

Objectos/variáveis

- Objectos podem ser criados usando o operador `<-` (operador de atribuição)
 - `x <- 2`
 - Um objecto pode ter um nome arbitrário
 - Alternativamente, pode se usar o operador `=`
- Um objecto pode armazenar qualquer expressão
- Toda expressão que não é armazenado num objecto é impressas

Tipos de objectos

- Numeric
 - armazena valores de ponto flutuante (números reais)
- Boolean
 - um valor lógico (TRUE or FALSE)
- Strings (character)
 - sequencia de caracteres

O tipo determinado automaticamente quando variável é criada usando `<-`