

Estatística Descritiva no R

1. Resumos numéricos de dados
2. Diagrama de ramos e folhas
3. Distribuição de frequências e histogramas

Download do R

O download do R pode ser feito a partir do site

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

Download do R

A screenshot of a Google search results page for the query "r". The search bar shows "r". Below it, the navigation bar includes "Todas" (selected), Maps, Notícias, Imagens, Vídeos, Mais, Configurações, and Ferramentas. The search results indicate approximately 17,520,000,000 results found in 0,50 seconds. A red arrow points from the text "The R Project for Statistical Computing" to the first result.

The R Project for Statistical Computing
<https://www.r-project.org/> ▾ Traduzir esta página

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.

R: What is R?
R is a language and environment for statistical computing and ...

The R Foundation
The R Foundation is a not for profit organization working in the ...

R (linguagem de programação) – Wikipédia, a enclopédia livre
[https://pt.wikipedia.org/wiki/R_\(linguagem_de_programação\)](https://pt.wikipedia.org/wiki/R_(linguagem_de_programação)) ▾

R é um ambiente computacional e uma linguagem de programação que vem progressivamente se especializando em manipulação, análise e visualização ...

Criado por: Ross Ihaka e Robert Gentleman Última versão: 3.5.0 (23 de abril de 2018)

R Linguagem de programação

R é um ambiente computacional e uma linguagem que vem progressivamente se especializando em manipulação, análise e visualização gráfica de dados. Na atualidade é o ambiente computacional para essa finalidade. W

Influência de: Julia

Última versão: 3.5.0 (23 de abril de 2018)

Surgido em: 1993

Download do R



[Home]

Download

CRAN

R Project

About R

Logo

Contributors

What's New?

Reporting Bugs

Conferences

Search

Get Involved: Mailing Lists

Developer Pages

R Blog

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

- useR! 2020 will take place in St. Louis, Missouri, USA.
- R version 3.5.3 (Great Truth)** has been released on 2019-03-11.
- The R Foundation Conference Committee has released a [call for proposals](#) to host useR! 2020 in North America.

Download do R

CRAN Mirrors

The Comprehensive R Archive Network is available at the following URLs, please choose a location close to you. Some statistics on the status of the mirrors can be

If you want to host a new mirror at your institution, please have a look at the [CRAN Mirror HOWTO](#).

0-Cloud

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

Automatic redirection to servers worldwide, currently
Automatic redirection to servers worldwide, currently

Algeria

<https://cran.usthb.dz/>
<http://cran.usthb.dz/>

University of Science and Technology Houari Boumed
University of Science and Technology Houari Boumed

Argentina

<http://mirror.fcaglp.unlp.edu.ar/CRAN/>

Universidad Nacional de La Plata

Australia

<https://cran.csiro.au/>
<http://cran.csiro.au/>
<https://mirror.aarnet.edu.au/pub/CRAN/>
<https://cran.ms.unimelb.edu.au/>
<https://cran.curtin.edu.au/>

CSIRO
CSIRO
AARNET
School of Mathematics and Statistics, University of M
Curtin University of Technology

Austria

<https://cran.wu.ac.at/>
<http://cran.wu.ac.at/>

Wirtschaftsuniversität Wien
Wirtschaftsuniversität Wien

Belgium

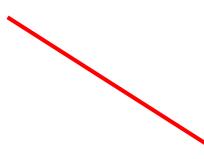
<https://www.freestatistics.org/cran/>
<http://www.freestatistics.org/cran/>
<https://lib.ugent.be/CRAN/>
<http://lib.ugent.be/CRAN/>

Patrick Wessa
Patrick Wessa
Ghent University Library
Ghent University Library

Brazil

<http://nbcgib.uesc.br/mirrors/cran/>
<https://cran-r.c3sl.ufpr.br/>
<http://cran-r.c3sl.ufpr.br/>
<https://cran.fiocruz.br/>
<http://cran.fiocruz.br/>
<https://vps.fmvz.usp.br/CRAN/>
<http://vps.fmvz.usp.br/CRAN/>
<https://brieger.esalq.usp.br/CRAN/>
<http://brieger.esalq.usp.br/CRAN/>

Computational Biology Center at Universidade Estadu
Universidade Federal do Parana
Universidade Federal do Parana
Oswaldo Cruz Foundation, Rio de Janeiro
Oswaldo Cruz Foundation, Rio de Janeiro
University of Sao Paulo, Sao Paulo
University of Sao Paulo, Sao Paulo
University of Sao Paulo, Piracicaba
University of Sao Paulo, Piracicaba



Download do R

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

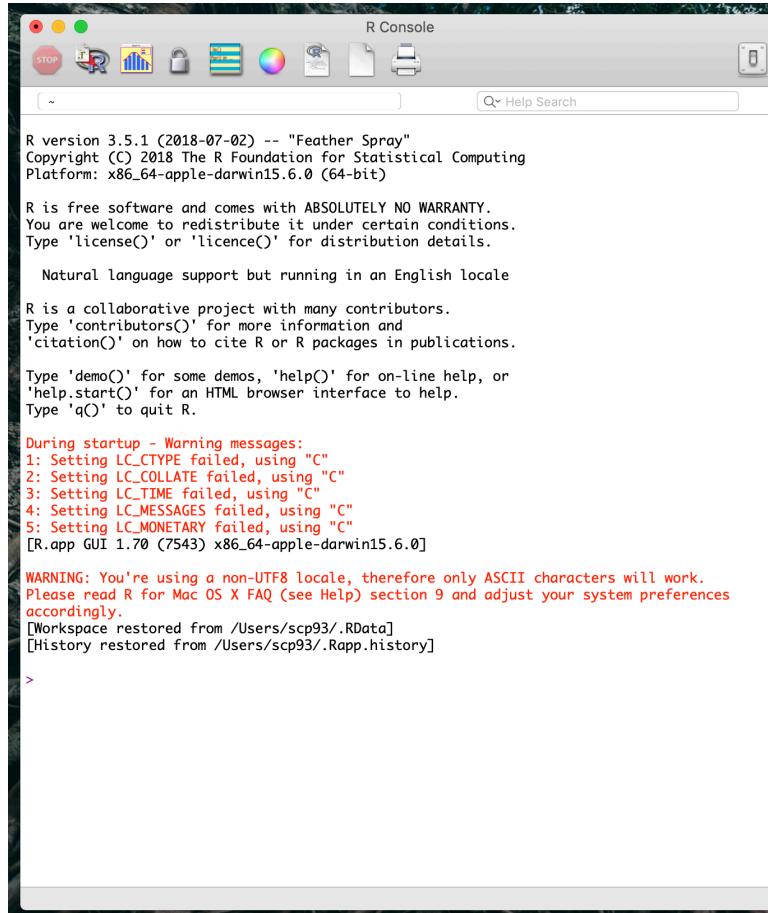
Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2019-03-11, Great Truth) [R-3.5.3.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 [available here](#). Please read about [new features and bug fixes](#) before filing corresponding feature requests or bug reports.
- Source code of older versions of R is [available here](#).
- Contributed extension [packages](#)

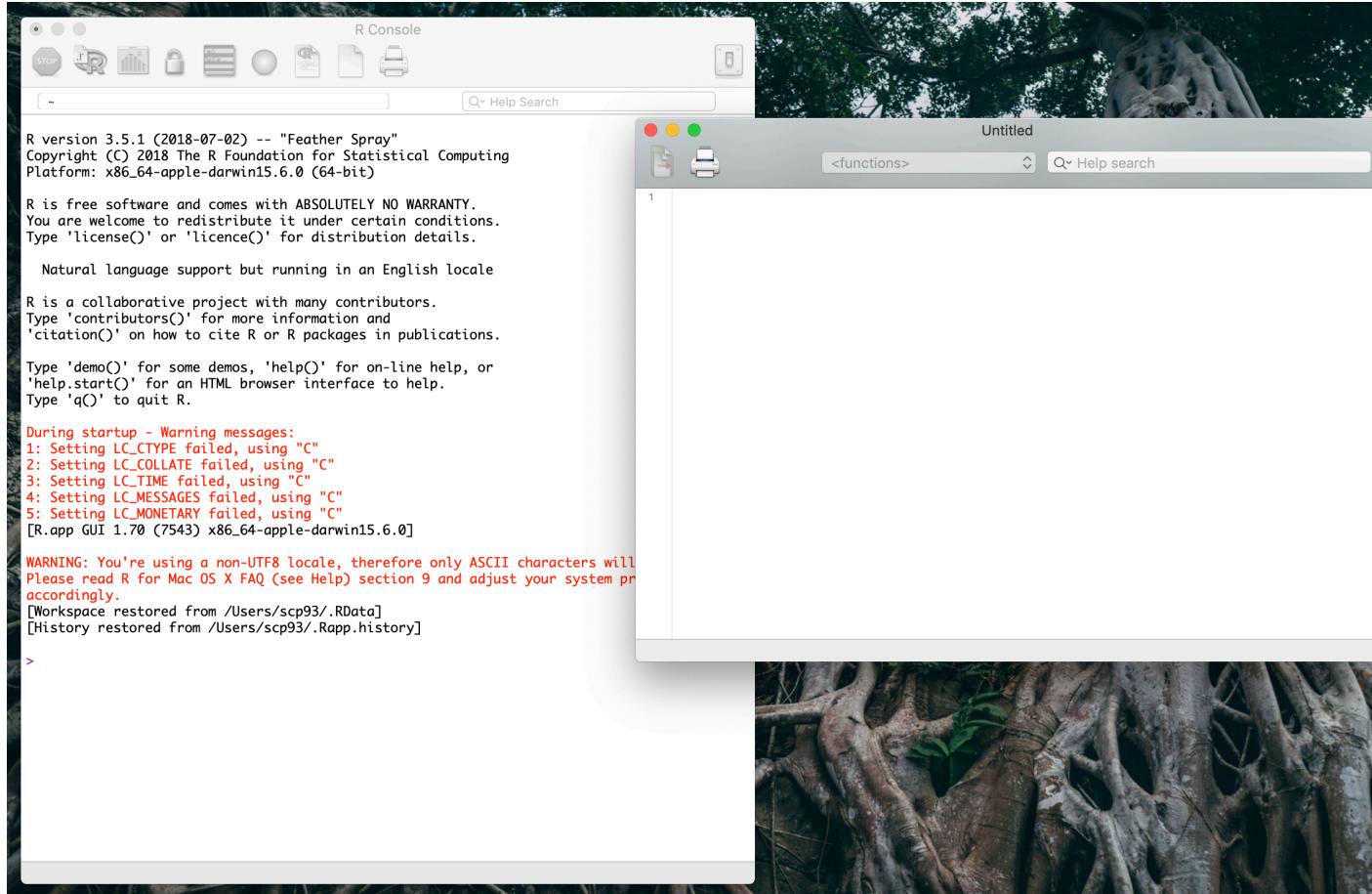
Questions About R

- 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.

Download do R



Download do R



Download do Rstudio

O download do R pode ser feito a partir do site

<https://www.rstudio.com>

Download do Rstudio

Google search results for "rstudio". A red arrow points to the first result.

RStudio: Open source and enterprise-ready professional software for ...
<https://www.rstudio.com/> ▾ Traduzir esta página
21 de fev de 2019 - RStudio provides popular open source and enterprise-ready professional software for the R statistical computing environment.

Resultados de rstudio.com

Download RStudio
Download the RStudio IDE or RStudio Server.

RStudio Desktop
Desktop. Run RStudio on your desktop. RStudio Desktop ...

RStudio – Wikipédia, a encyclopédia livre
<https://pt.wikipedia.org/wiki/RStudio> ▾
RStudio é um software livre de ambiente de desenvolvimento integrado para R, uma linguagem de programação para gráficos e cálculos estatísticos.

RStudio - Wikipedia



RStudio
Software

RStudio é um software livre de ambiente de desenvolvimento integrado para R, uma linguagem de programação para gráficos e cálculos estatísticos. [Wikipédia](#)

Linguagem: C++, Qt

Sistema operativo: Microsoft Windows, Mac OS X

Plataforma: x86, x64

Download do Rstudio

Google search results for "rstudio".

Aproximadamente 6.070.000 resultados (0,29 segundos)

RStudio: Open source and enterprise-ready professional software for ...
<https://www.rstudio.com/> ▾ Traduzir esta página
21 de fev de 2019 - RStudio provides popular open source and enterprise-ready professional software for the R statistical computing environment.

Resultados de rstudio.com

Download RStudio
Download the RStudio IDE or RStudio Server.

RStudio Desktop
Desktop. Run RStudio on your desktop. RStudio Desktop ...

RStudio – Wikipédia, a encyclopédia livre
<https://pt.wikipedia.org/wiki/RStudio> ▾
RStudio é um software livre de ambiente de desenvolvimento integrado para R, uma linguagem de programação para gráficos e cálculos estatísticos.

RStudio - Wikipedia



RStudio
Software

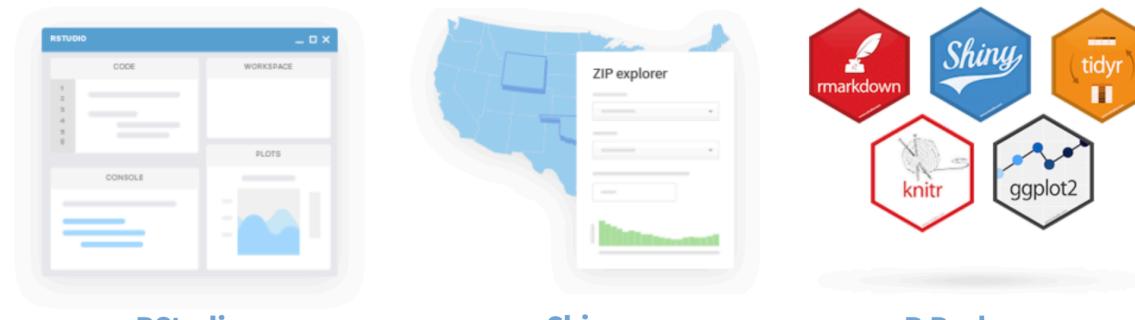
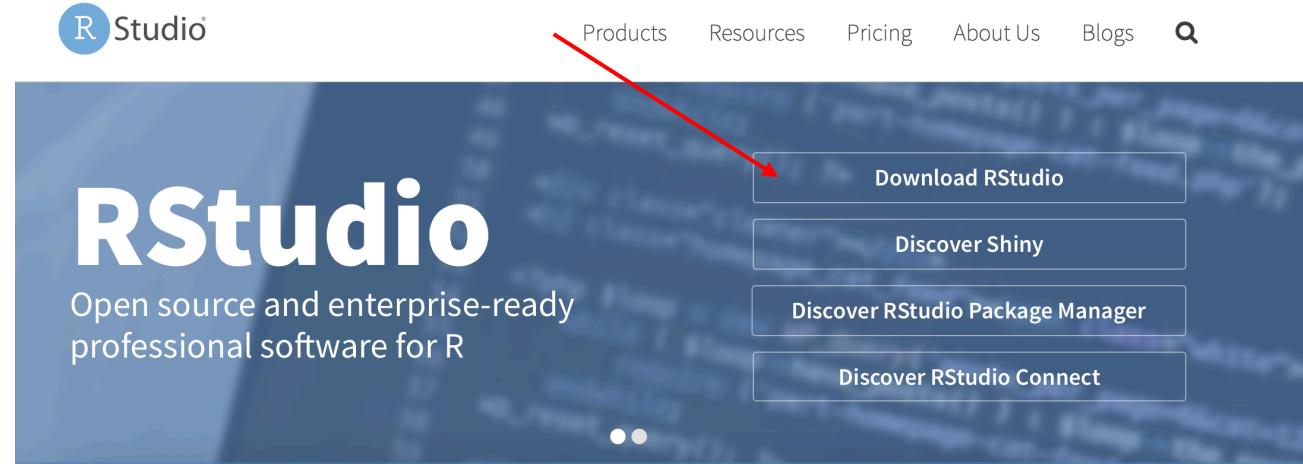
RStudio é um software livre de ambiente de desenvolvimento integrado para R, uma linguagem de programação para gráficos e cálculos estatísticos. [Wikipédia](#)

Linguagem: C++, Qt

Sistema operativo: Microsoft Windows, Mac OS X

Plataforma: x86, x64

Download do Rstudio



Download do Rstudio



Products Resources Pricing About Us Blogs

Choose Your Version of RStudio

RStudio is a set of integrated tools designed to help you be more productive with R. It includes a console, syntax-highlighting editor that supports direct code execution, and a variety of robust tools for plotting, viewing history, debugging and managing your workspace. [Learn More about RStudio features.](#)



RStudio Desktop
Open Source License

FREE

[DOWNLOAD](#)

[Learn More](#)

RStudio Desktop
Commercial License

\$995 per year

[BUY](#)

[Learn More](#)

RStudio Server
Open Source License

FREE

[DOWNLOAD](#)

[Learn More](#)

RStudio Server Pro
Commercial License

\$9,995 per year

[DOWNLOAD](#)

[Learn More](#)

RStudio Server Pro +
RStudio Connect
Commercial License

\$29,995 per
year

[TALK](#)

[Learn More](#)

Integrated
Tools for R



Priority
Support



Access via
Web Browser



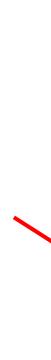
Enterprise
Security



Project
Sharing



Project
Sharing



Download do Rstudio



Products Resources Pricing About Us Blogs

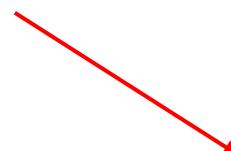
RStudio Desktop 1.1.463 — Release Notes

RStudio requires R 3.0.1+. If you don't already have R, download it [here](#).

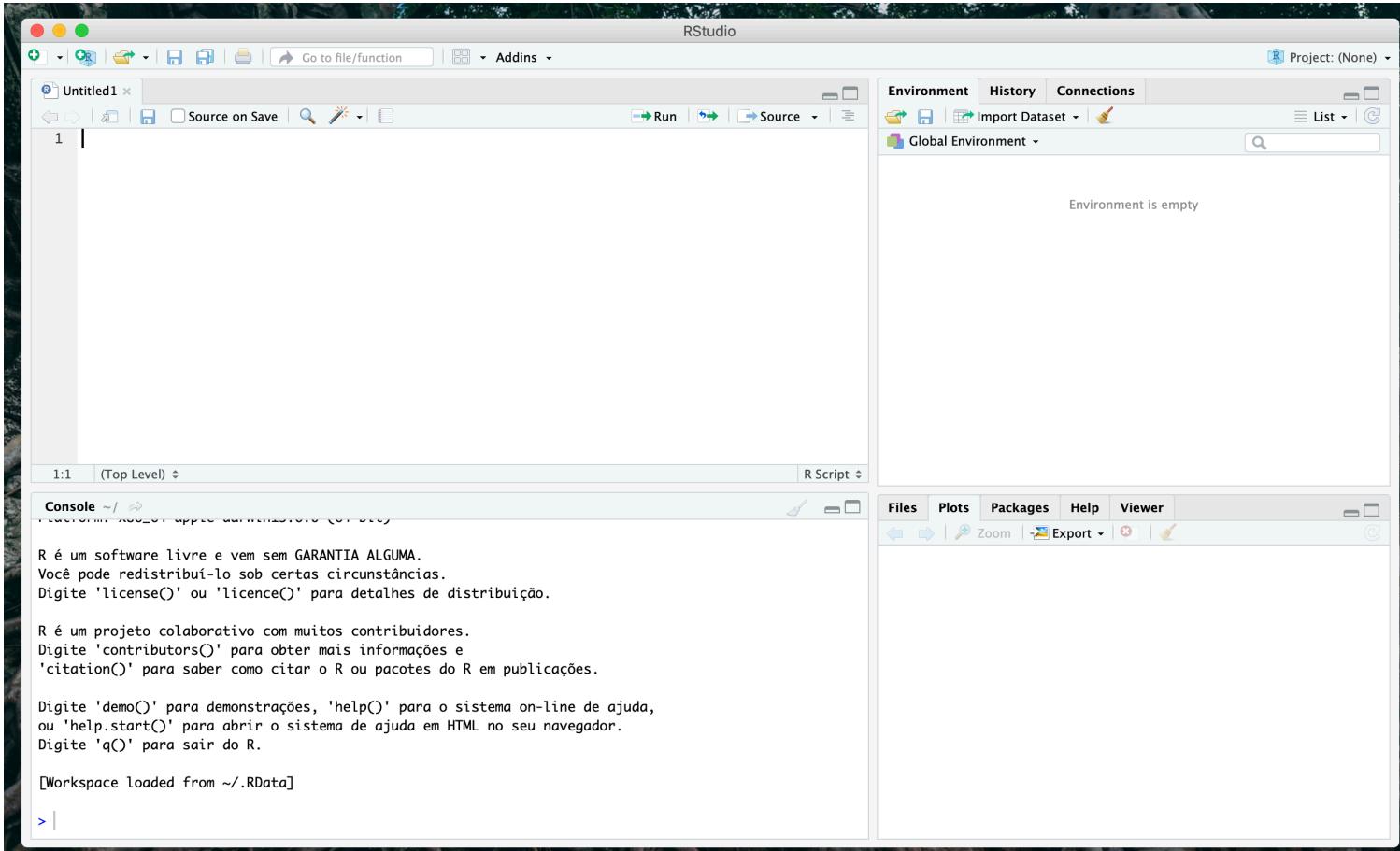
Linux users may need to [import RStudio's public code-signing key](#) prior to installation, depending on the operating system's security policy.

Installers for Supported Platforms

Installers	Size	Date	MD5
RStudio 1.1.463 - Windows Vista/7/8/10	85.8 MB	2018-10-29	58b3d796d8cf96fb8580c62f46ab64d4
RStudio 1.1.463 - Mac OS X 10.6+ (64-bit)	74.5 MB	2018-10-29	a79032ba4d7daaa86a8da01948278d94
RStudio 1.1.463 - Ubuntu 12.04-15.10/Debian 8 (32-bit)	89.3 MB	2018-10-29	8a6755fa9fae2bafce289df3358aaf63
RStudio 1.1.463 - Ubuntu 12.04-15.10/Debian 8 (64-bit)	97.4 MB	2018-10-29	bc50d6bd34926c1cc3ae4a209d67d649
RStudio 1.1.463 - Ubuntu 16.04+/Debian 9+ (64-bit)	65 MB	2018-10-29	cf659db18619cc78d1592fefaa7c753
RStudio 1.1.463 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (32-bit)	88.1 MB	2018-10-29	742f0bad60dfeaa3281576e14ad6699e
RStudio 1.1.463 - Fedora 19+/RedHat 7+/openSUSE 13.1+ (64-bit)	90.6 MB	2018-10-29	c7303067a0ca99deea7e427b856952d1



Download do Rstudio



Inserir de dados

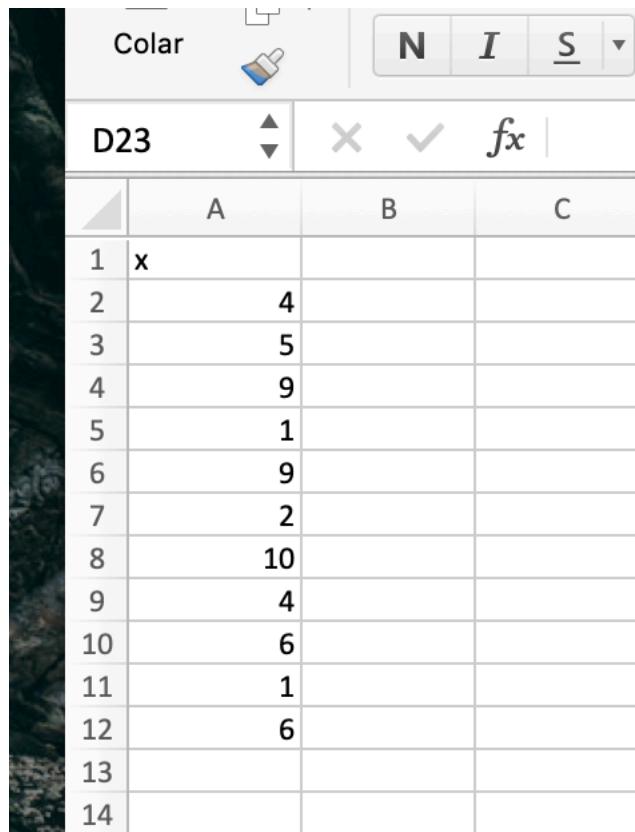
- Principais formas de entrar com dados no R
 1. Importar de algum arquivo externo
 2. Digitar os dados

Inserir de dados

```
# criar um vetor
x = c(4, 5, 9, 1, 9, 2, 10, 4, 6, 1, 6)

# ver o vetor criado
print(x)
x
```

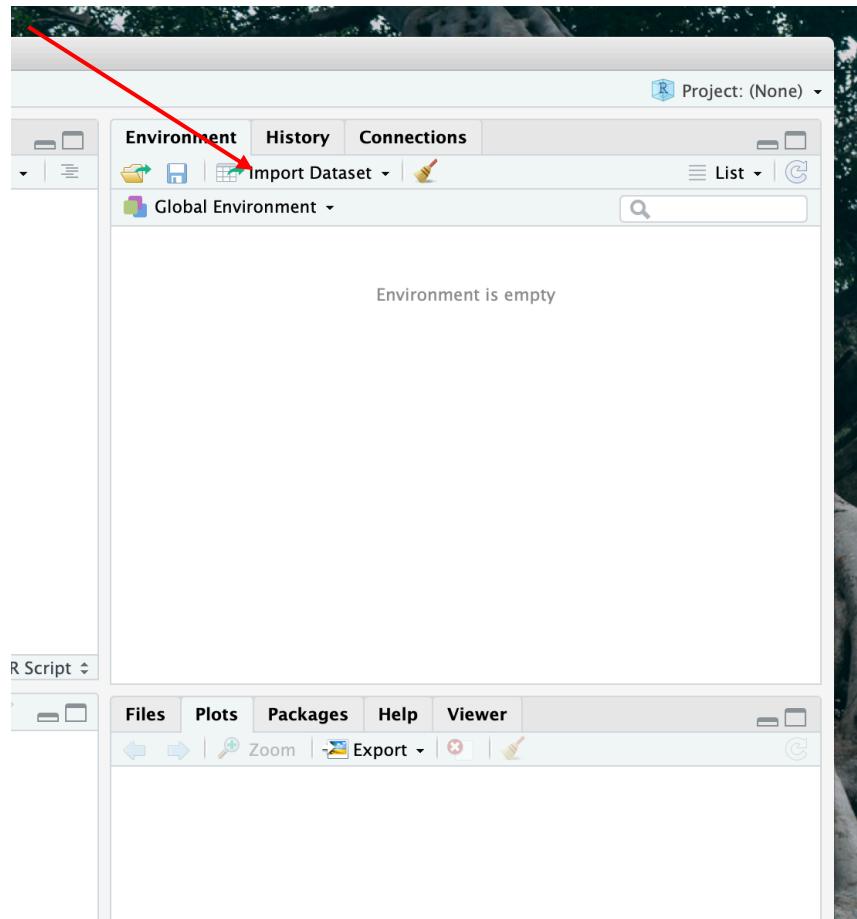
Inserir de dados



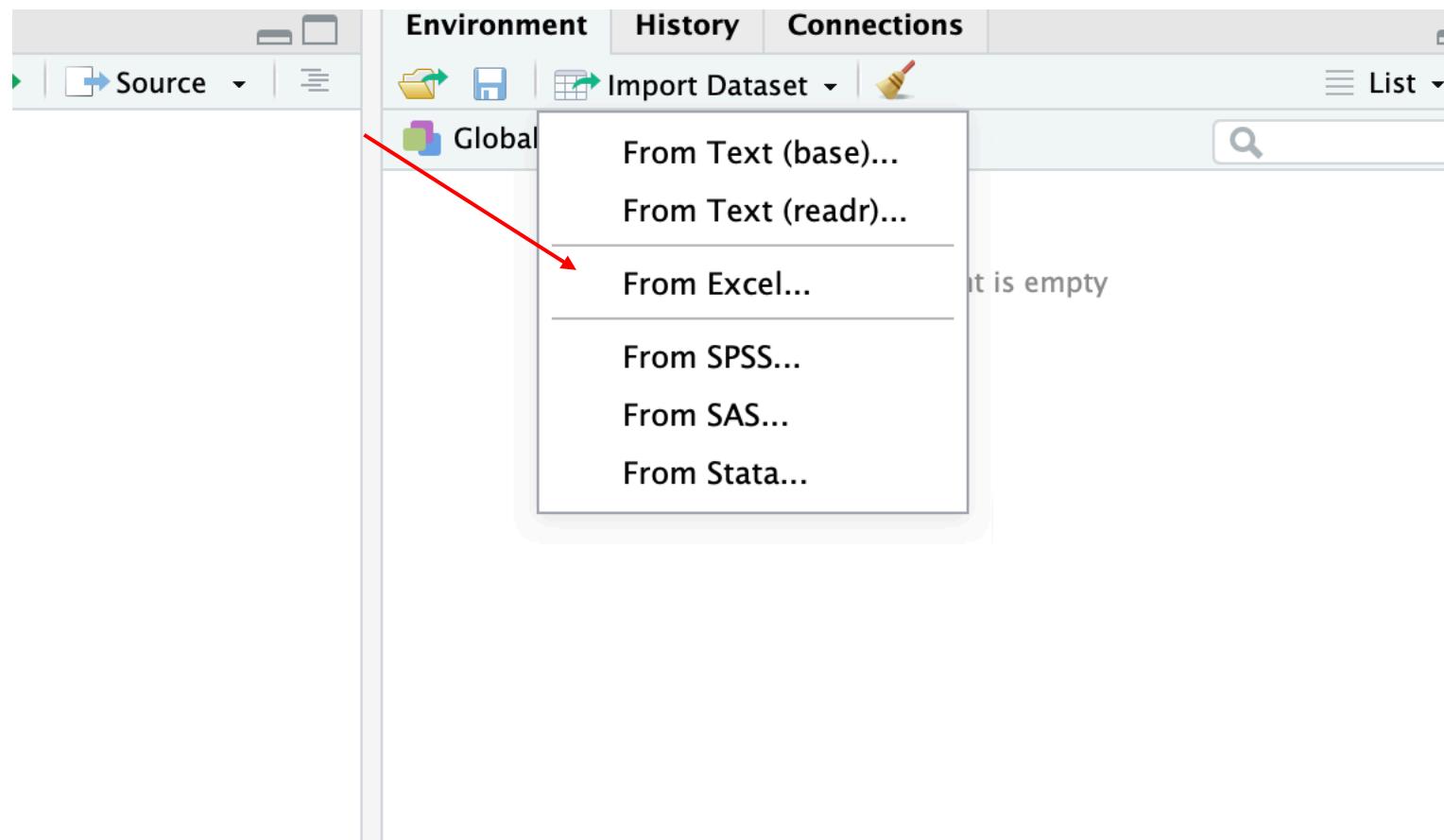
A screenshot of a Microsoft Excel spreadsheet. The ribbon bar at the top includes tabs for 'Colar' (Paste), 'N' (Normal), 'I' (Insert), and 'S' (Styles). The active cell is D23. The formula bar shows the address D23. Below the formula bar is a toolbar with icons for copy, paste, and other functions. The main area displays a table with columns A, B, and C and rows numbered 1 to 14. The data in column A is: Row 1: x; Rows 2-12: 4, 5, 9, 1, 9, 2, 10, 4, 6, 1, 6; Rows 13-14: empty.

	A	B	C
1	x		
2	4		
3	5		
4	9		
5	1		
6	9		
7	2		
8	10		
9	4		
10	6		
11	1		
12	6		
13			
14			

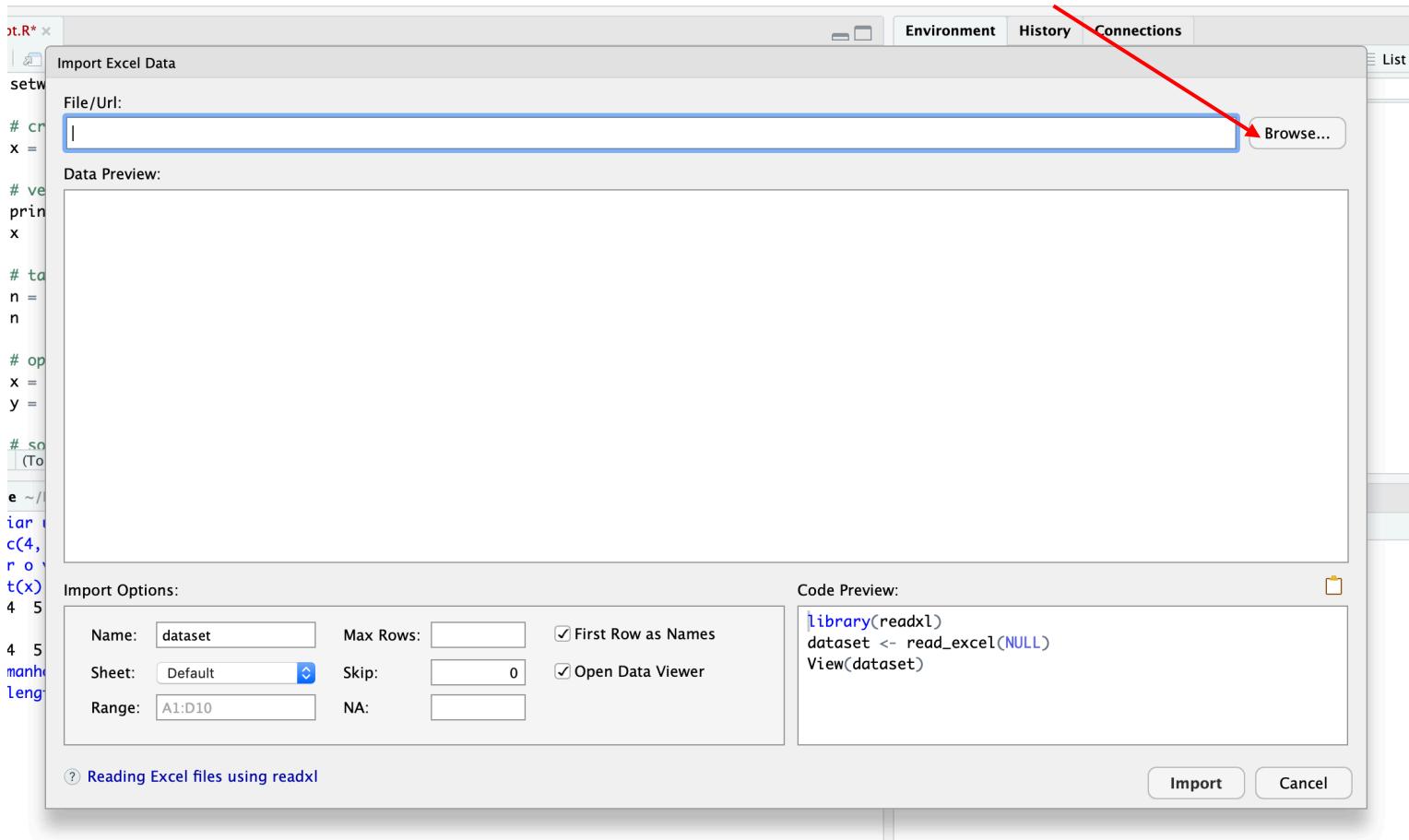
Inserir de dados



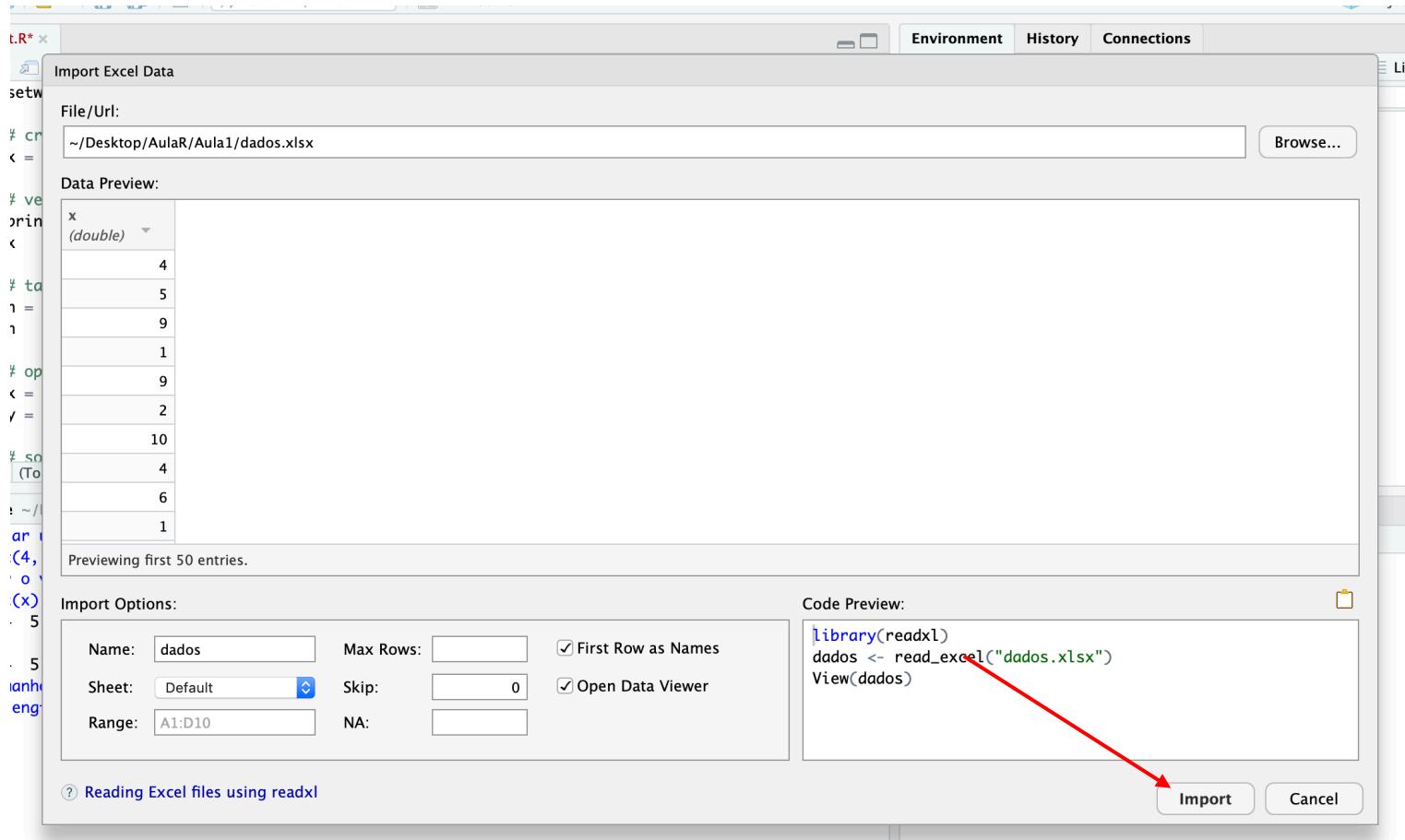
Inserir de dados



Inserir de dados



Inserir de dados



Inserir de dados

A screenshot of the RStudio interface illustrating the process of inserting data into a dataset. A red arrow points from the 'script.R*' tab in the top-left corner to the 'dados' tab in the top-right corner of the Data View panel.

The Data View panel shows a grid of data with 11 rows and 2 columns, labeled 'x'. The data is:

	x
1	4
2	5
3	9
4	1
5	9
6	2
7	10
8	4
9	6
10	1
11	6

The Global Environment panel on the right shows the 'dados' dataset with 11 observations and 1 variable.

The Console at the bottom shows the R code used to create the vector:

```
> # criar um vetor  
> x = c(4, 5, 9, 1, 9, 2, 10, 4, 6, 1, 6)  
> # ver o vetor criado
```

Resumos numéricos de dados

Média da amostra

Se as n observações em uma amostra forem denotadas por x_1, x_2, \dots, x_n , então a **média amostral** será dada por

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n} \sum x_i$$

Resumos numéricos de dados

Média da amostra no R

```
# criar um vetor  
x = c(4, 5, 9, 1, 9, 2, 10, 4, 6, 1, 6)  
  
# tamanho do vetor  
n = length(x)  
  
# calcula a media  
sum(x)/n
```

Resumos numéricos de dados

Média da amostra no R

```
# calcula a media  
mean(x)
```

Resumos numéricos de dados

Variância da amostra

Se x_1, x_2, \dots, x_n for uma amostra de n observações, então a **variância amostral** será dada por

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n - 1} = \frac{1}{n - 1} \sum (x_i - \bar{x})^2$$

Resumos numéricos de dados

Variância da amostra no R

```
# ----- criar um vetor -----
x = c(4, 5, 9, 1, 9, 2, 10, 4, 6, 1, 6)

# ----- calcula variância -----
xbar = mean(x)
y = (x-xbar)^2
sum(y)/(n-1)
```

Resumos numéricos de dados

Variância da amostra no R

```
# funcao  
var(x)
```

Resumos numéricos de dados

Amplitude da amostra

Se x_1, x_2, \dots, x_n for uma amostra de n observações, então a **amplitude da amostra** será dada por

$$r = \max(x_i) - \min(x_i)$$

Resumos numéricos de dados

Amplitude da amostra no R

```
# ----- calcula a amplitude -----
max(x)-min(x)

# funcao
range(x)
```

Resumos numéricos de dados

Função summary

Função genérica usada para produzir resumos dos dados ou resultados de funções.

```
summary(x)
  Min. 1st Qu. Median   Mean 3rd Qu.   Max.
1.000  3.000  5.000  5.182  7.500 10.000
```

Exemplo

The screenshot shows the Kaggle website interface for the "Titanic: Machine Learning from Disaster" competition. At the top, there is a navigation bar with the Kaggle logo, a search bar, and links for Competitions, Datasets, Kernels, Discussion, Learn, and more. A user profile icon and a notification bell are also visible.

The main content area features a large image of the Titanic ship at night. Above the image, it says "Getting Started Prediction Competition". Below the image, the competition title is "Titanic: Machine Learning from Disaster" and the subtitle is "Start here! Predict survival on the Titanic and get familiar with ML basics".

Below the title, there is a "Kaggle · 10,183 teams · Ongoing" badge. At the bottom of the main section, there is a navigation bar with tabs: Overview (which is underlined), Data, Kernels, Discussion, Leaderboard, Rules, Team, My Submissions, and Submit Predictions (which is highlighted with a blue background).

Under the "Overview" tab, there is a sidebar with links: Description, Evaluation, Tutorials, and Frequently Asked. To the right of the sidebar, there is a "Start here if..." section with the text: "You're new to data science and machine learning, or looking for a simple intro to the Kaggle prediction competitions." There is also a "Competition Description" link.

Diagrama de ramo e folhas

Um **diagrama de ramo e folhas** é uma boa maneira de obter uma apresentação visual informativa sobre o conjunto de dados x_1, x_2, \dots, x_n , em que cada número x_i consiste em, no mínimo, dois dígitos. Para construir um diagrama de ramo e folhas, uso as seguintes etapas:

1. Divida cada número x_i em duas partes: um **ramo**, consistindo em um ou mais dígitos iniciais, e uma **folha**, consistindo nos dígitos restantes.
2. Liste os valores do ramo em uma coluna vertical.
3. Ao lado do ramo, registre a folha para cada observação.
4. Escreva as unidades para os ramos e folhas no gráfico

Diagrama de ramo e folhas

Exemplo:

```
> print(x)
[1] 57 57 60 65 72 74 77 78 81 93
> stem(x)

The decimal point is 1 digit(s) to the right of the |

  5 | 77
  6 | 05
  7 | 2478
  8 | 1
  9 | 3
```

Diagrama de ramo e folhas

Exemplo:

```
> print(x)
[1] 701 701 703 706 711 713 714 715 717 725
> stem(x)

The decimal point is 1 digit(s) to the right of the |

 70 | 113
 70 | 6
 71 | 134
 71 | 57
 72 |
 72 | 5
```

Diagrama de ramo e folhas

No R, utilizar a função `stem`

`stem {graphics}`

R Documentation

Stem-and-Leaf Plots

Description

`stem` produces a stem-and-leaf plot of the values in `x`. The parameter `scale` can be used to expand the scale of the plot. A value of `scale = 2` will cause the plot to be roughly twice as long as the default.

Diagrama de ramo e folhas

No R, utilizar a função stem

```
# ----- Cap 6.2 -----
# -- diagrama de ramo e folhas
set.seed(1)
x = rbinom(40, 120, .5)
stem(x)
```

Diagrama de ramo e folhas

No R, utilizar a função stem

```
> # ----- Cap 6.2 -----
> # -- diagrama de ramo e folhas
> set.seed(1)
> x = rbinom(40, 120, .5)
> stem(x)
```

The decimal point is 1 digit(s) to the right of the |

```
4 | 7
5 | 13
5 | 566788899
6 | 001112222223334444
6 | 5556669
7 | 1
7 | 5
```

Exemplo



Distribuições de frequências e histograma

Uma **distribuição de frequência** é um resumo mais compacto dos dados, em relação ao diagrama de ramo e folhas. Para construir a distribuição de frequência, temos que dividir a faixa de dados em intervalos, geralmente chamados de **intervalos de classe**.

O número intervalos depende do número de observações e da dispesão dos dados. Geralmente, achamos que 5 a 20 intervalos são satisfatórios na maioria dos casos e que o número de intervalos deve crescer com n . Por isso, na prática trabalha-se bem se o **número de intervalos de classe** for aproximadamente igual a raiz quadrada do número de observações.

Distribuições de frequências e histograma

No R

```
# -- Distribuicao de frequencia

# dados
set.seed(1)
x = sort(rbinom(100, 150, .5))

# tamanho da amostra
n = length(x)

# numero de classe
k = floor(sqrt(n))

# criacao dos intervalos de classe
classe = seq(from= min(x), to = max(x)+.1, length.out = k+1)

# cria tabela de frequencia
cortes = cut(x, breaks = classe, right = F)
table(cortes)
```

Exemplo



It's been 84 years...

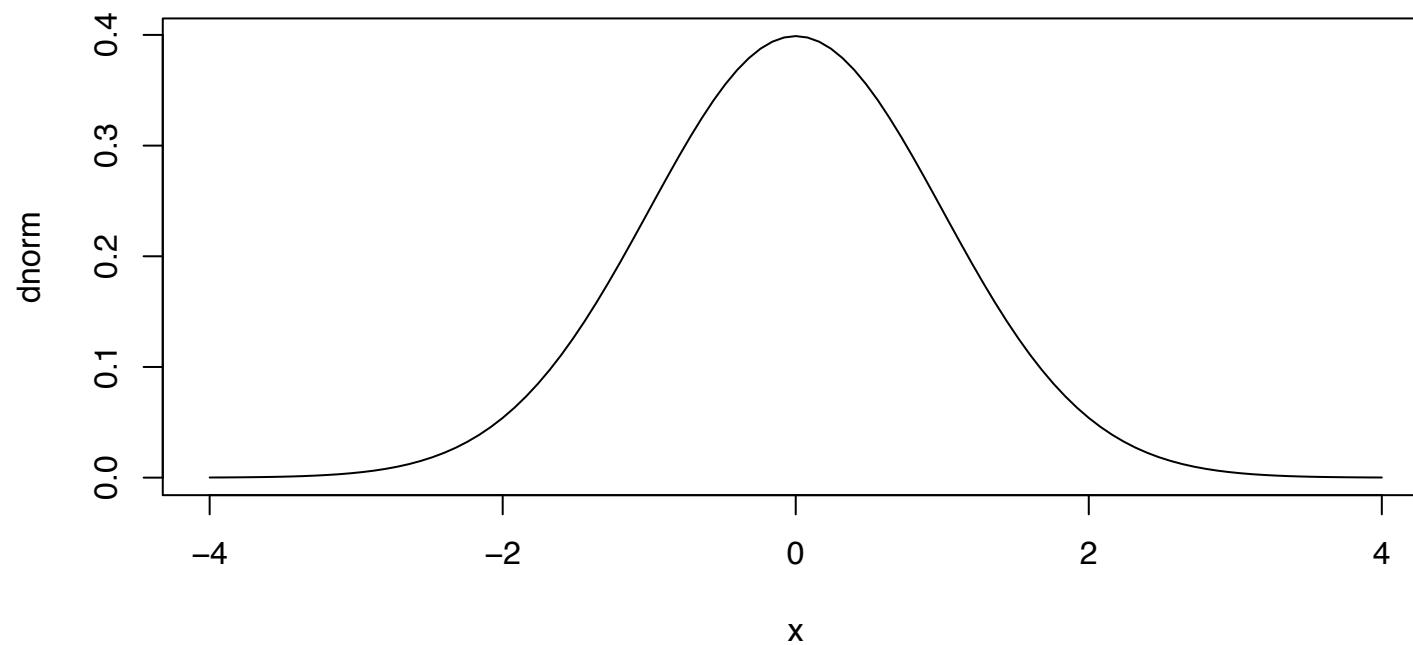
Distribuições de frequências e histograma

O **histograma** é uma disposição visual da distribuição de frequências. Os estágios para costruir um histograma são dados por:

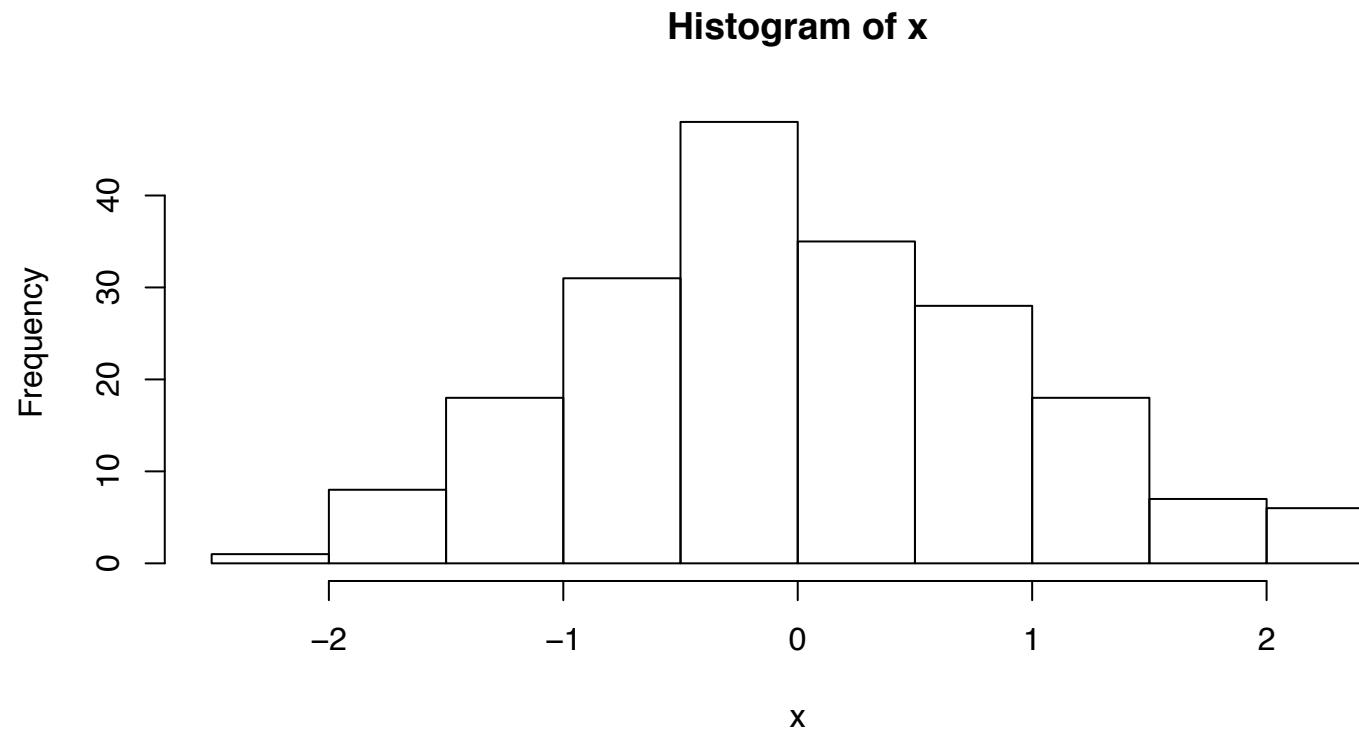
1. Marque os limites dos intervalos de classe em um eixo horizontal.
2. Marque e nomeie o eixo vertical com as frequências ou com as frequências relativas.
3. Acima de cada célula, desenhe um retangulo em que a altura seja igual à frequência (ou a frequências relativas) correspondente àquela célula.

Distribuições de frequências e histograma

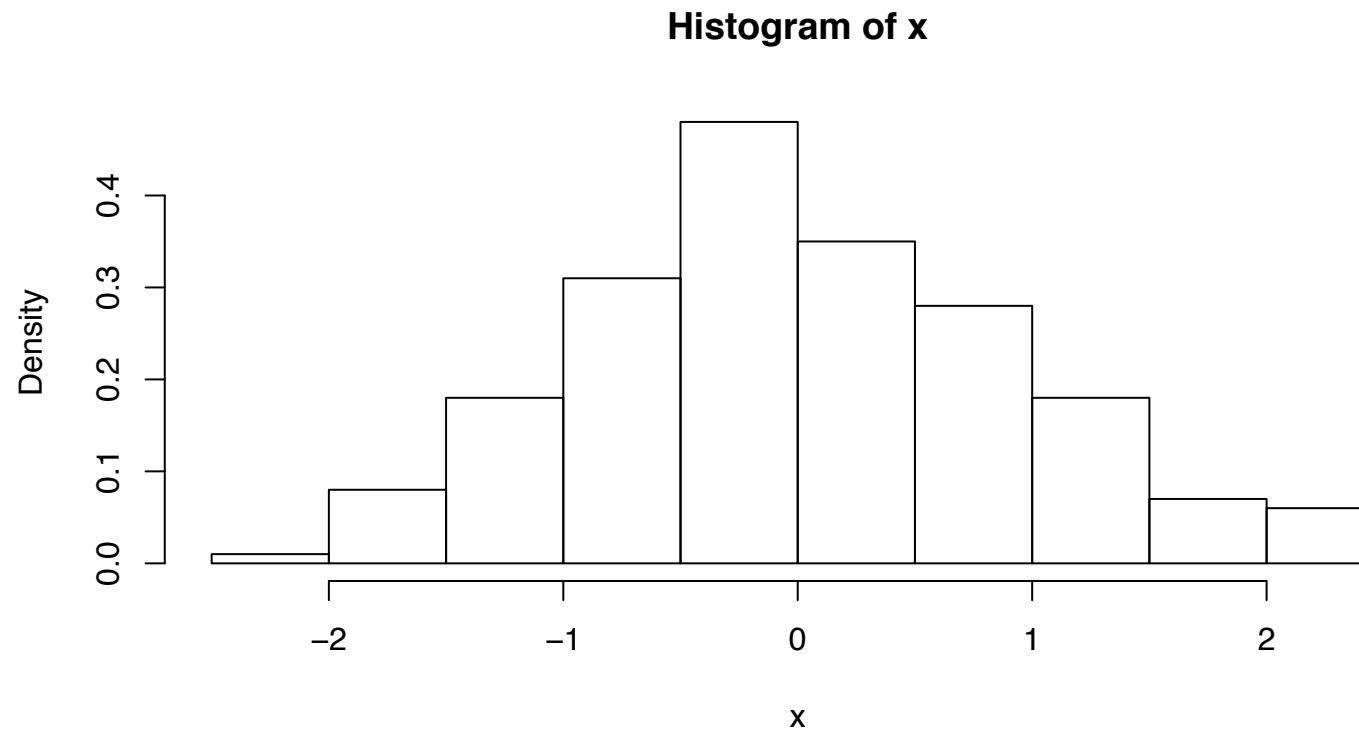
Se a população tem distribuição normal



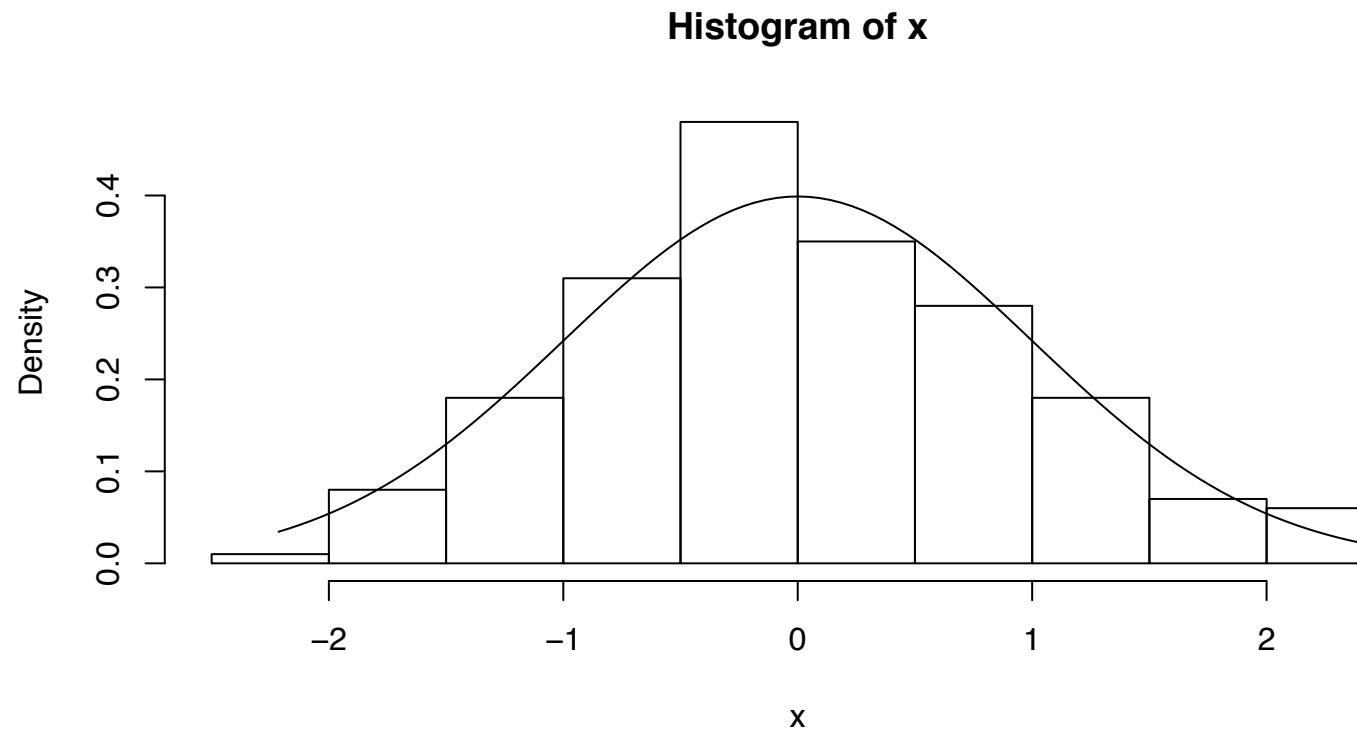
Distribuições de frequências e histograma



Distribuições de frequências e histograma

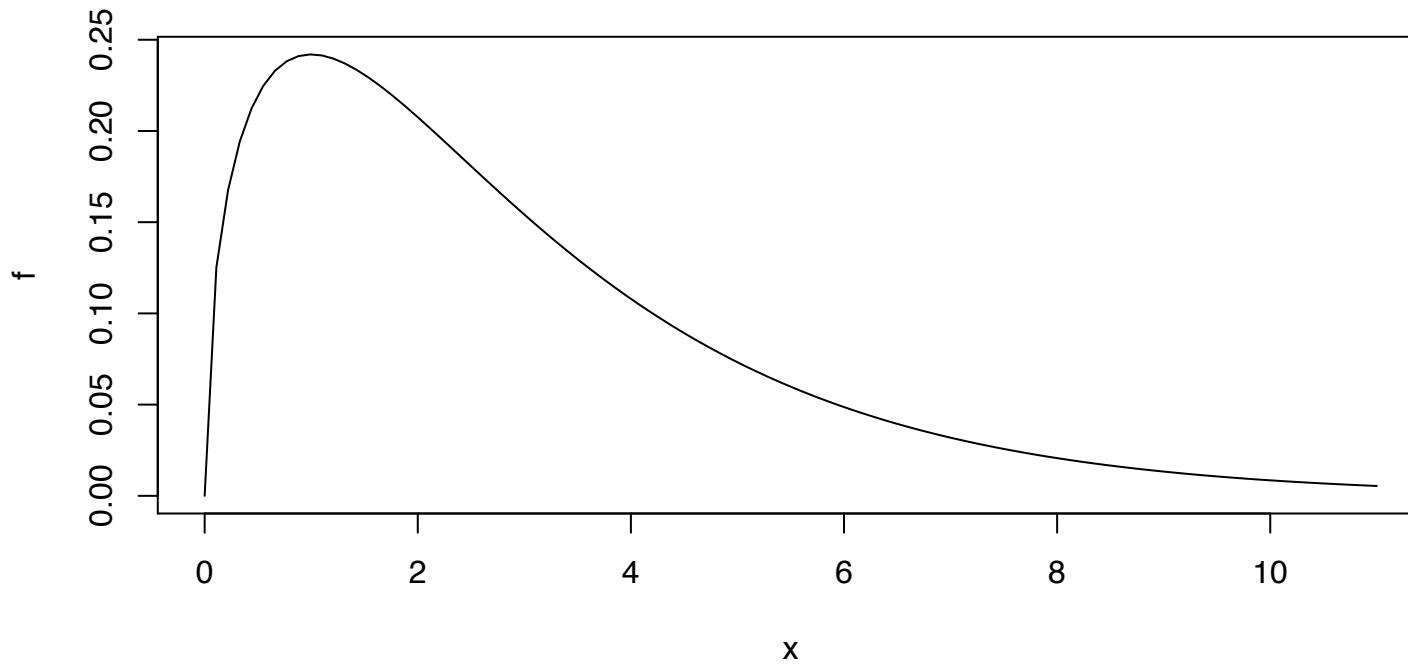


Distribuições de frequências e histograma

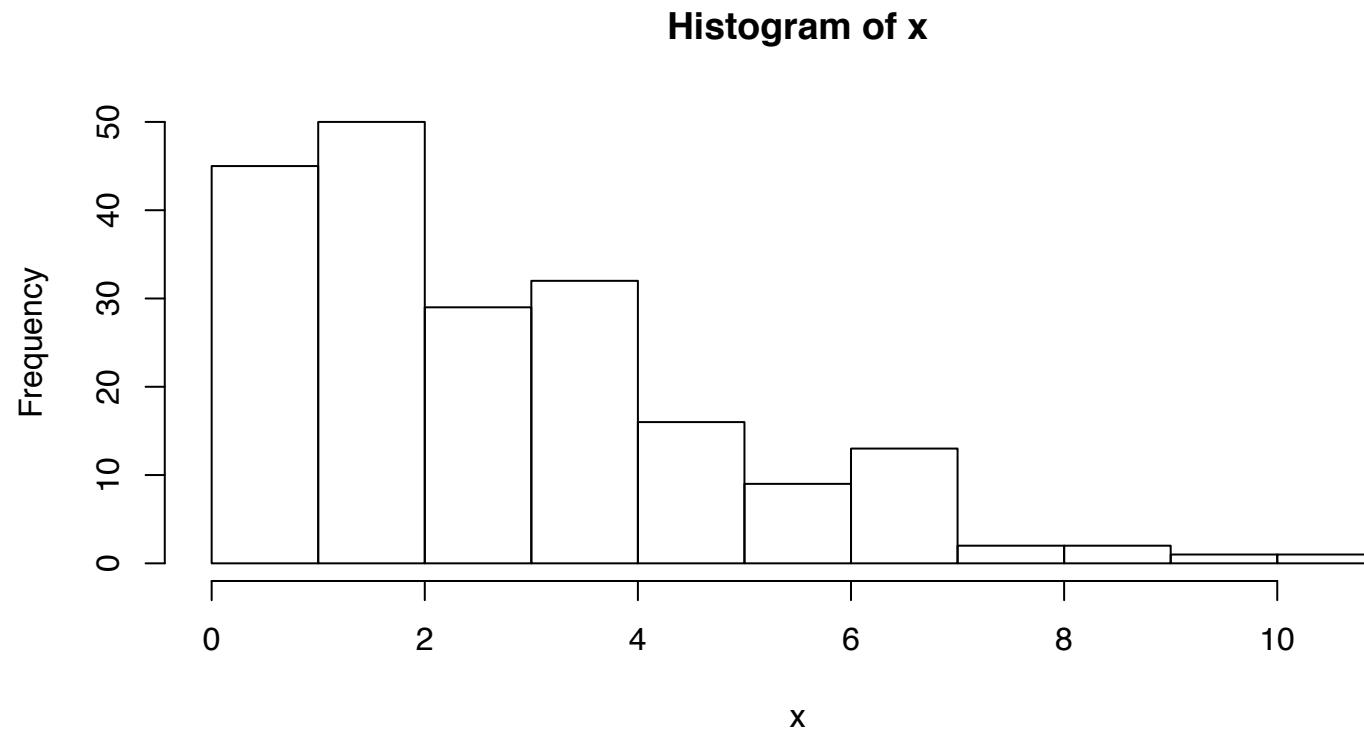


Distribuições de frequências e histograma

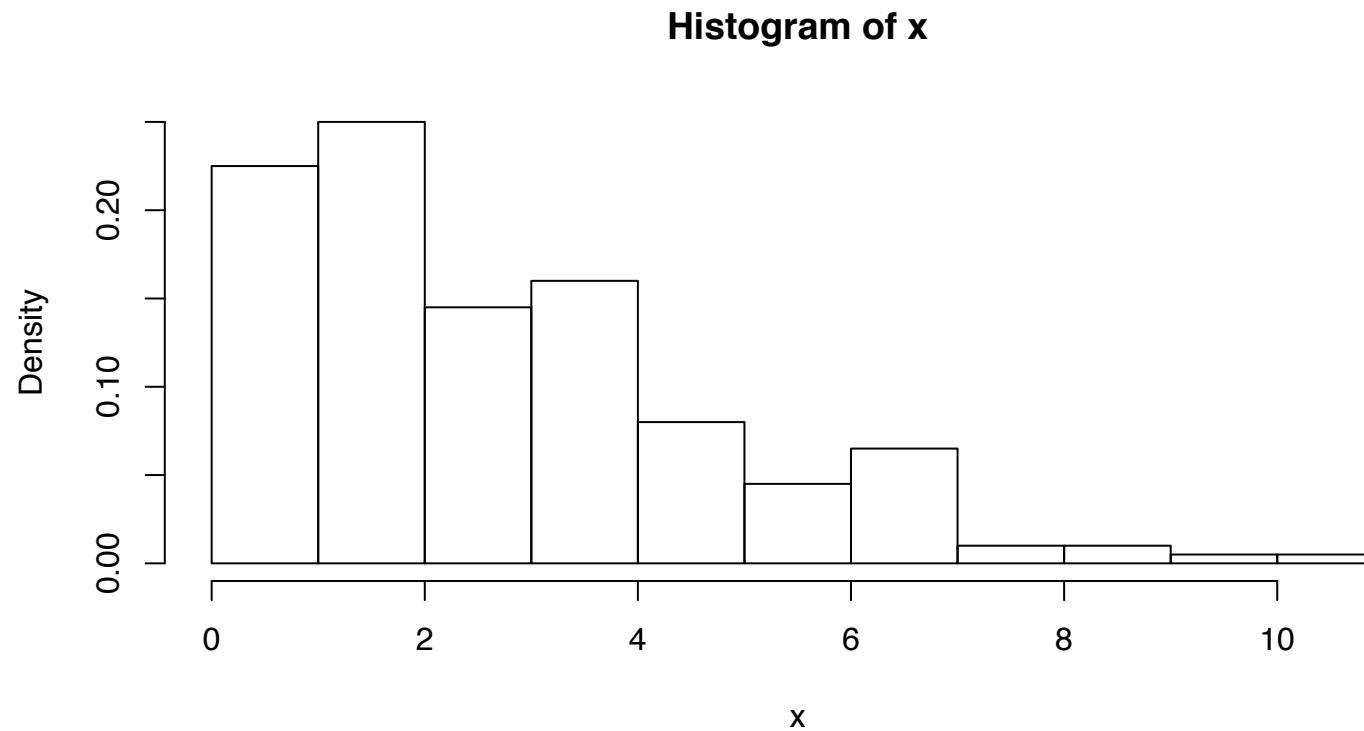
Se a população tem distribuição chi-quadrado



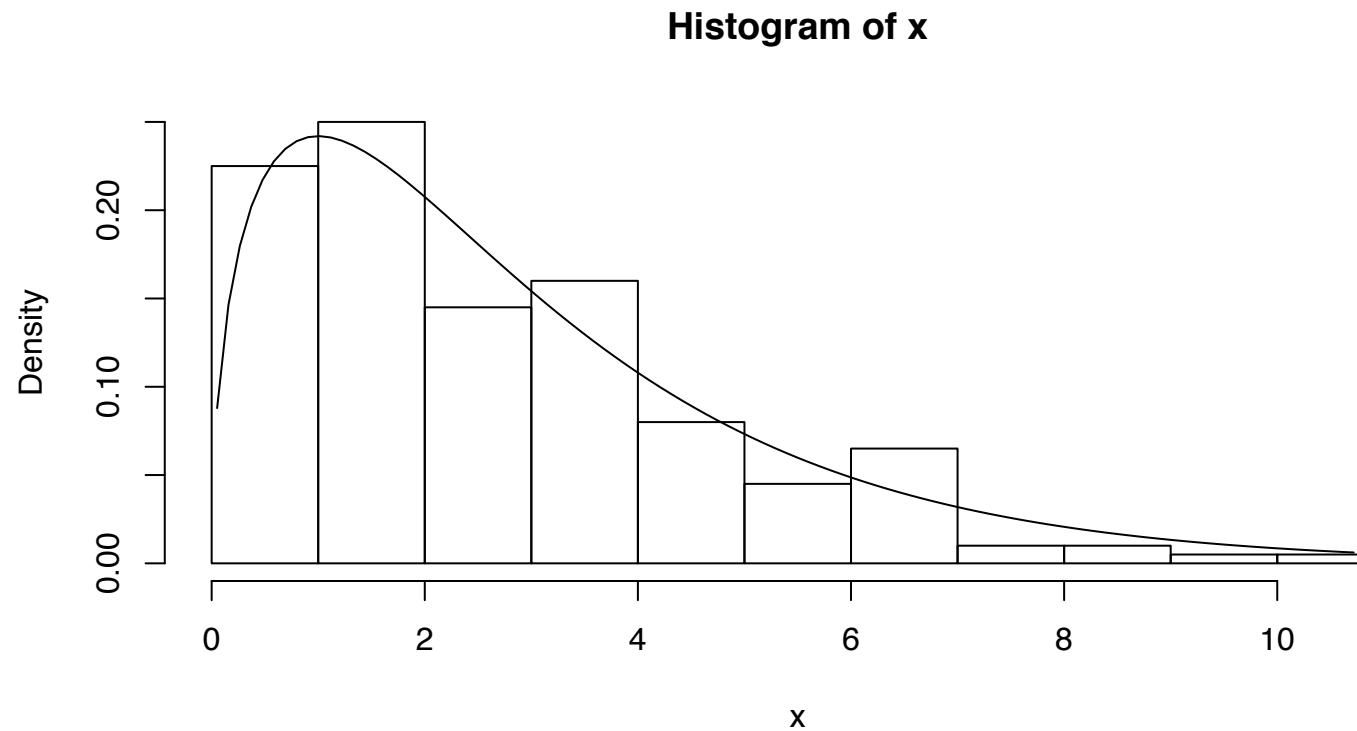
Distribuições de frequências e histograma



Distribuições de frequências e histograma

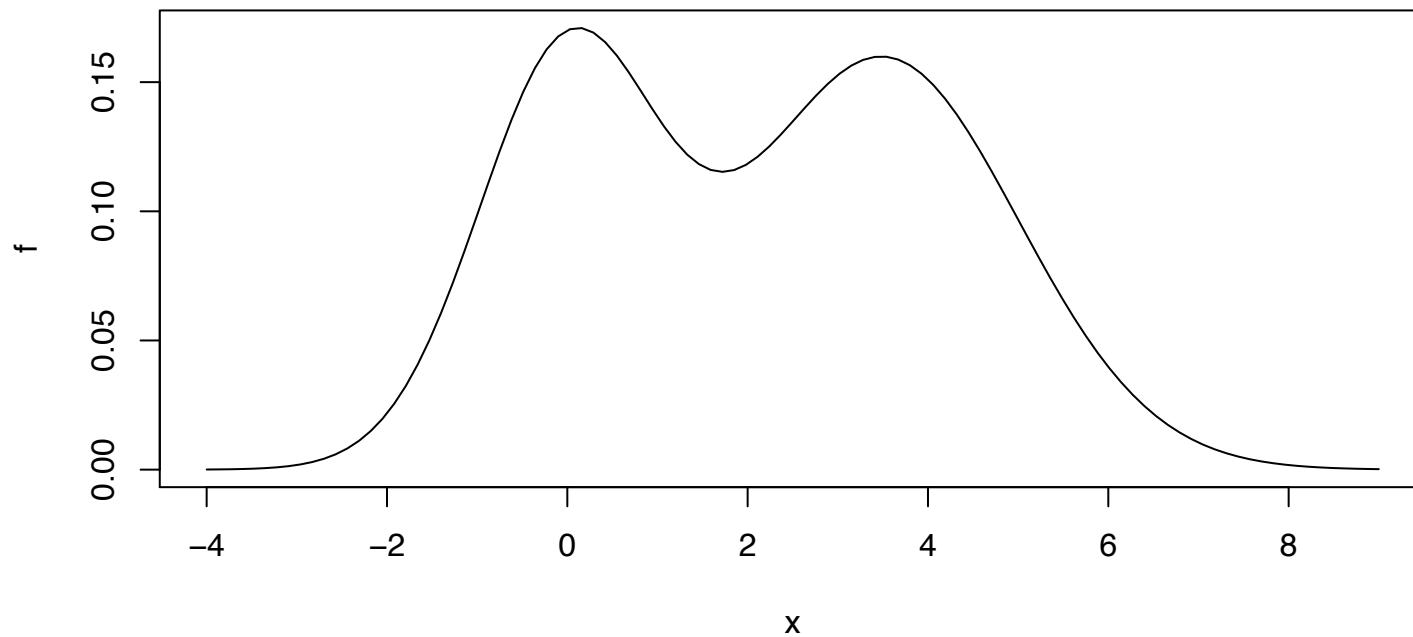


Distribuições de frequências e histograma

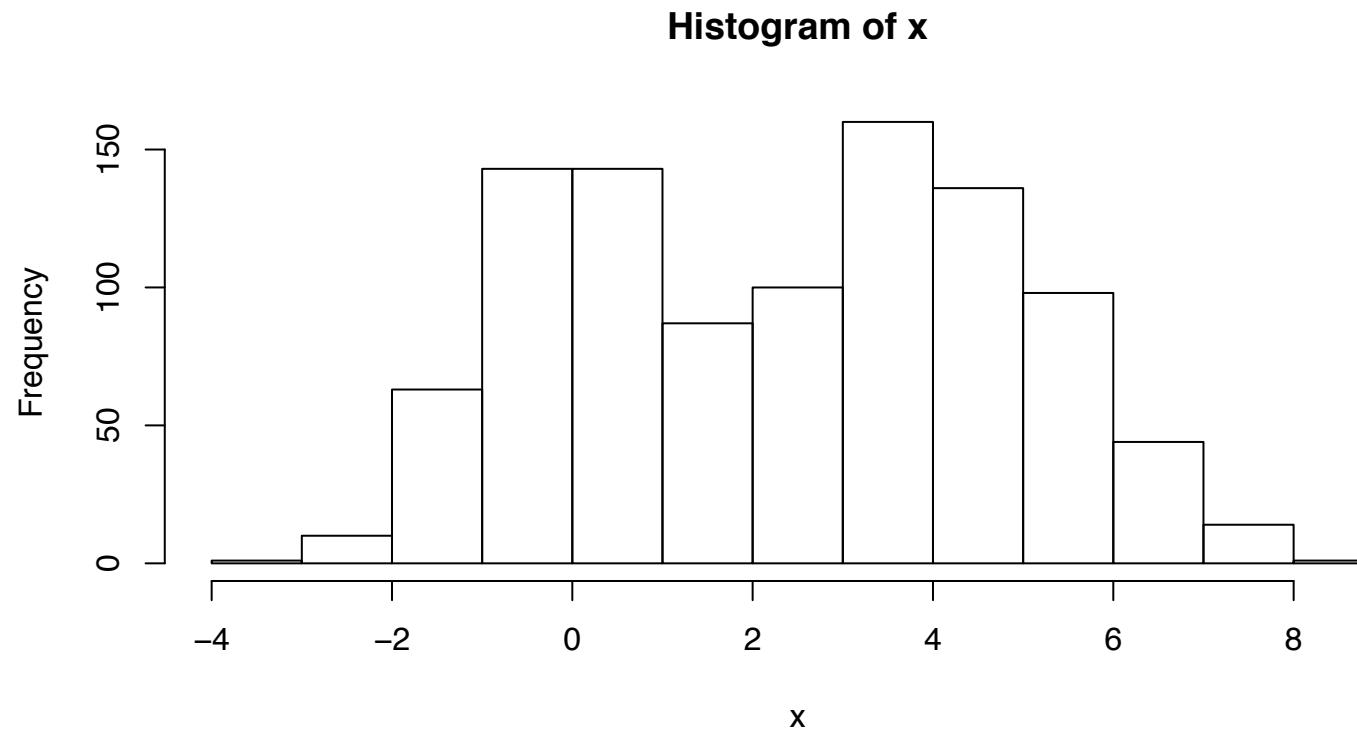


Distribuições de frequências e histograma

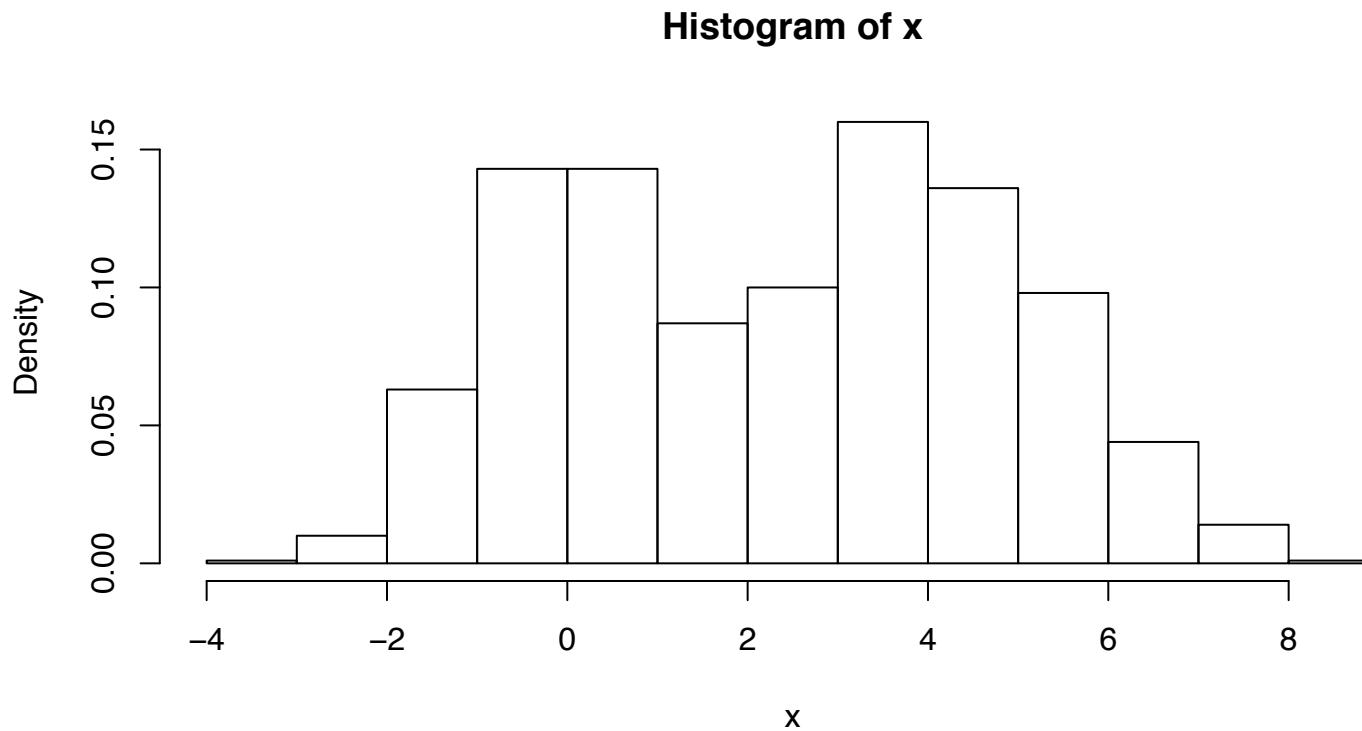
Se a população tem distribuição



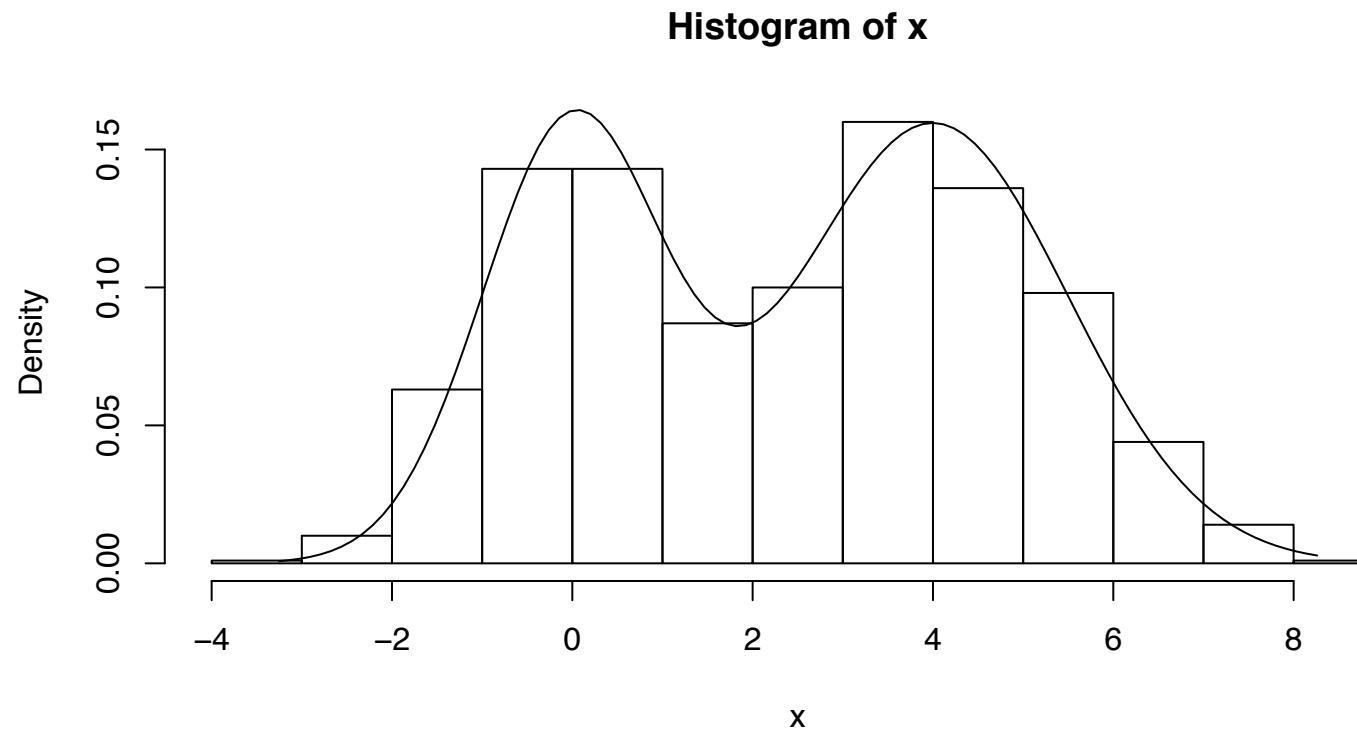
Distribuições de frequências e histograma



Distribuições de frequências e histograma



Distribuições de frequências e histograma



Distribuições de frequências e histograma

No R usar a função `hist`

`hist {graphics}`

R Documentation

Histograms

Description

The generic function `hist` computes a histogram of the given data values. If `plot = TRUE`, the resulting object of [class "histogram"](#) is plotted by [plot.histogram](#), before it is returned.

Usage

Distribuições de frequências e histograma

Para histograma com frequência:

```
# histograma para frequencia  
hist(x)
```

Distribuições de frequências e histograma

Para histograma com frequência relativa:

```
# histograma para frequencia relativa  
hist(x, probability = T)
```

Exemplo



QUESTION:

