

Apostila R

Adriana

21/05/2020

Apresentação

Estas notas se destinam a servir de apoio à utilização do software R nas disciplinas de Estatística Básica da Universidade Federal Rural do Rio de Janeiro. Nenhum tópico apresenta uma explicação exaustiva dos tópicos abordados. Recomenda-se a leitura das fontes consultadas para elaboração desse material, a saber: http://www.estatisticacomr.uff.br/?page_id=691

1. Apresentação do Software

O R é um programa gratuito e de código aberto. Nesse sentido, o programa apresenta uma extensa coleção de técnicas com possibilidades de atender demandas das mais distintas áreas. O seu caráter colaborativo também contribui para uma constante oferta de novos métodos implementados frequentemente. Maiores informações sobre o projeto, colaboradores e documentação podem ser encontradas na página oficial do projeto em: <http://www.r-project.org>.

2. Instalação

O R pode ser obtido gratuitamente no endereço <http://www.r-project.org>. Estão disponíveis versões para os sistemas operacionais Windows, MacOS X e Linux. Após acessar o site, deve-se escolher o CRAN (Comprehensive R Archive Network) para indicar o local de disponibilização do programa. No Brasil, atualmente, temos quatro opções: Universidade Federal do Paraná, Fundação Oswaldo Cruz, Universidade de São Paulo e USP Piracicaba.

3. Estrutura

Ao abrir o programa, será apresentada a interface do usuário conforme ilustra a Figura.

O símbolo > posicionado à direita do cursor é o prompt do R e indica que o programa está pronto para armazenar informações e operar comandos.

Com o intuito de facilitar e agilizar o uso do R, foram desenvolvidas outras interfaces de usuário. Dentre elas destaca-se o RStudio disponibilizado gratuitamente.



[Home]

Download

CRAN

R Project

About R

Logo

Contributors

What's New?

Reporting Bugs

Conferences

Search

Get Involved: Mailing Lists

Developer Pages

R Blog

R Foundation

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 4.0.1 \(See Things Now\) prerelease versions](#) will appear starting Wednesday 2020-05-27. Final release is scheduled for Saturday 2020-06-06.
- [R version 4.0.0 \(Arbor Day\)](#) has been released on 2020-04-24.
- [useR! 2020 in Saint Louis has been cancelled](#). The European hub planned in Munich will not be an in-person conference. Both organizing committees are working on the best course of action.
- [R version 3.6.3 \(Holding the Windsock\)](#) has been released on 2020-02-29.
- You can support the R Foundation with a renewable subscription as a [supporting member](#)

Figure 1: Etapas de Instalação - 1

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

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

Brazil
<https://nbcgib.uesc.br/mirrors/cran/>
<https://cran-rc3sl.ufpr.br/>
<https://cran.fiocruz.br/> 
<https://vps.fmvz.usp.br/CRAN/>
<https://brieger.esalq.usp.br/CRAN/>

Bulgaria
<https://ftp.uni-sofia.bg/CRAN/>

Wirtschaftsuniversität Wien

Patrick Wessa
Ghent University Library

Computational Biology Center at Universidade Estadual de Santa Cruz
Universidade Federal do Paraná
Oswaldo Cruz Foundation, Rio de Janeiro
University of São Paulo, São Paulo
University of São Paulo, Piracicaba

Sofia University 

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 to

- [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 links

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. If you do not know what this means, you probably do not want to do it!

- The latest release (2020-04-24, Arbor Day) [R-4.0.0.tar.gz](#), read [what's new](#) in the latest version.

Figure 2: Etapas de Instalação - 2

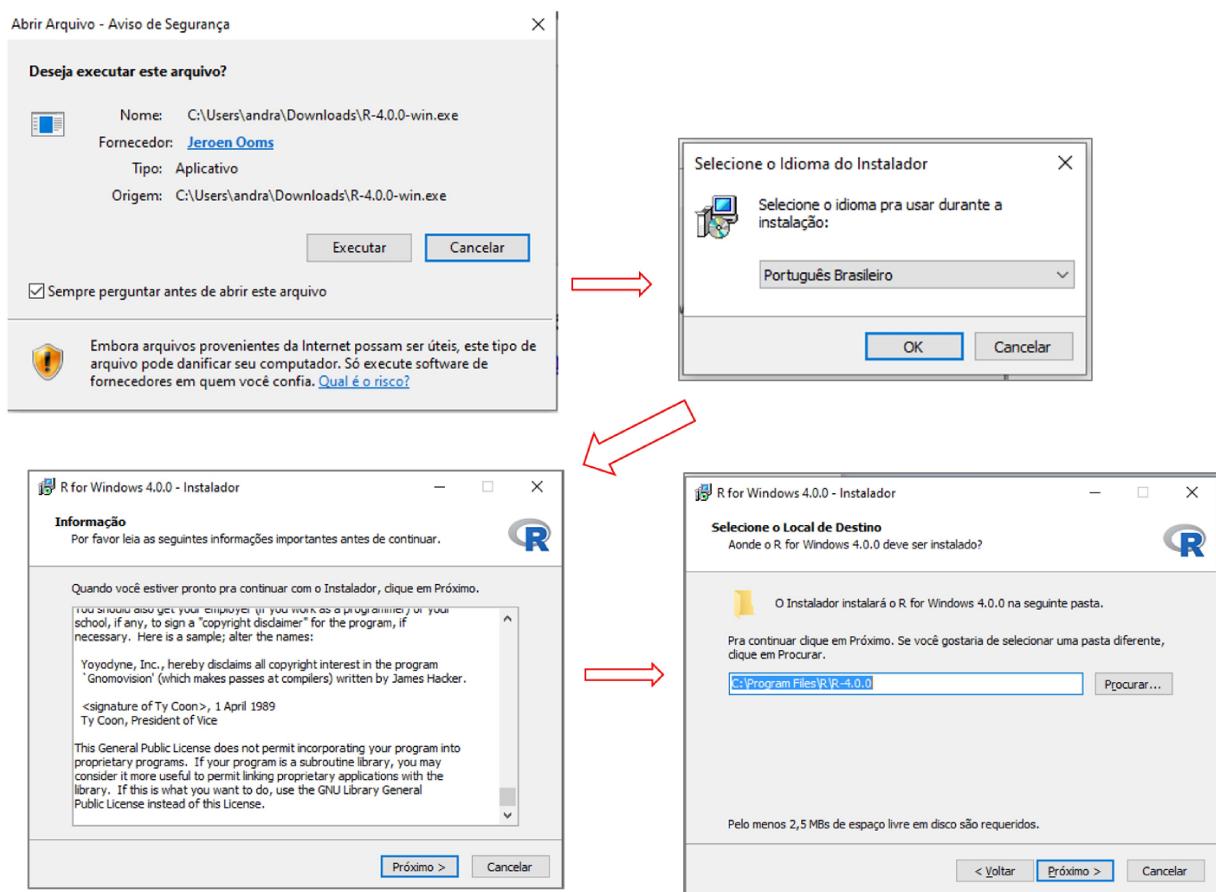


Figure 3: Etapas de Instalação - 3

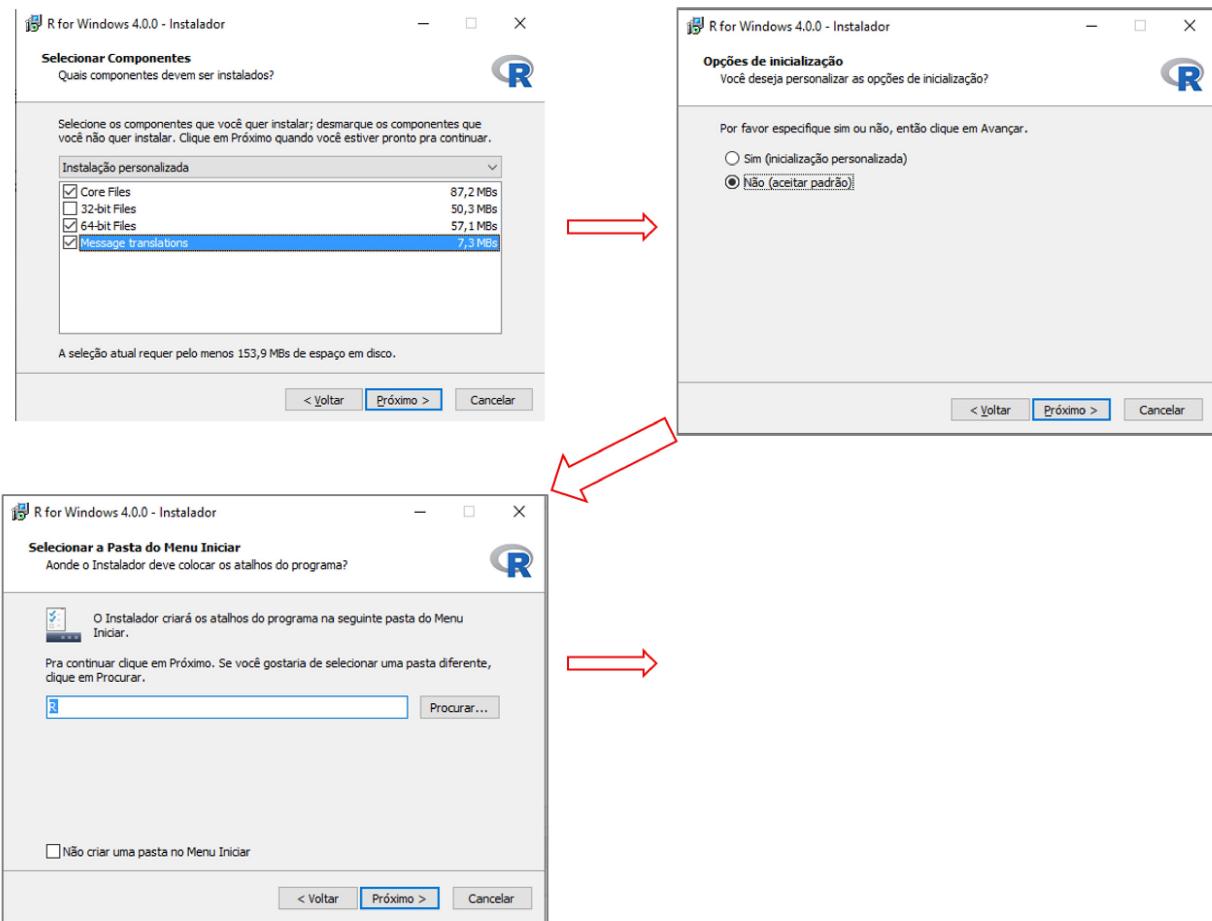


Figure 4: Etapas de Instalação - 4

The screenshot shows the official R Project website at www.r-project.org. The main navigation menu on the left includes links for Home, Download, CRAN (highlighted with a red arrow), R Project, About R, Logo, Contributors, What's New?, Reporting Bugs, Conferences, Search, Get Involved: Mailing Lists, Developer Pages, R Blog, and R Foundation. The main content area is titled "The R Project for Statistical Computing" and features sections for "Getting Started" and "News".

Figure 5: Interface do usuário do R no Windows

O RStudio é um programa que trabalha integrado ao R com destaque para seu editor de texto (*R script*) destinado à edição dos comandos e uma eficiente visualização de janelas.

Portanto, ao invés de digitar na interface do R, o usuário passará a operar na janela de edição de texto do RStudio e terá acesso a um conjunto de informações mobilizadas durante a sessão de operação do R.

Outro elemento fundamental na estrutura do R está relacionado com os comandos de execução de operações e técnicas no programa, denominado de funções. As funções são disponibilizadas no R através de pacotes (*package*), também chamados de bibliotecas (*library*).

Os comandos básicos (funções) se encontram no pacote *base*. Há uma considerável oferta de pacotes com conteúdos diversos desenvolvidos pelos colaboradores do sistema. Para acessá-los, é preciso realizar a instalação do pacote de interesse no R, pois diferentemente do pacote *base*, os demais não são disponibilizados na instalação do R.

A instalação é realizada com o uso do comando > `install.package()`. Uma vez que o pacote é instalado, deve-se fazer o seu carregamento na memória do programa com o uso do comando > `library()`.

```
install.package ("agricolae")
library("agricolae")
```

4. Operações Matemáticas Básicas

```
#Raiz
sqrt(100)

## [1] 10

#Exponencial
exp(0)

## [1] 1

#log na base e
log(1)

## [1] 0

#log na base 10
log10(1)

## [1] 0
```

Operadores Lógicos

Trata-se de operados binários para realização de testes entre dois objetos.

rstudio.com/products/rstudio/download/#download

RStudio Desktop 1.3.959 - [Release Notes](#)

- 1.** Install R. RStudio requires [R 3.0.1+](#).
- 2.** Download RStudio Desktop. Recommended for your system:

 [DOWNLOAD RSTUDIO FOR WINDOWS](#)
1.3.959 | 171.41MB

Requires Windows 10/8/7 (64-bit)

All Installers

Linux users may need to [import RStudio's public code-signing key](#) prior to installation.

RStudio 1.2 requires a 64-bit operating system. If you are on a 32 bit system, you can

Figure 6: Download RStudio - 1

← → 🔍 rstudio.com/products/rstudio/download/

Choose Your Version

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's new solution for every professional data science team. RStudio Team includes RStudio Server Pro, RStudio Connect and RStudio Package Manager.

[LEARN MORE](#)

RStudio Desktop	RStudio Desktop	RStudio Server	RStudio Server Pro
Open Source License	Commercial License	Open Source License	Commercial License
Free	\$995 /year	Free	\$4,975 /year (5 Named Users)
DOWNLOAD	BUY	DOWNLOAD	BUY
Learn more	Learn more	Learn more	Evaluation Learn more

Figure 7: Download RStudio - 2

The Comprehensive R Archive Network

Austria <https://cran.wu.ac.at/>
 Belgium <https://www.freestatistics.org/cran/>
 Brazil <https://nbcgilb.uesc.br/mirrors/cran/>
<https://cran.csl.ufrj.br/> 
<https://cran.fiocruz.br/> 
<https://vps.fmvz.usp.br/CRAN/>
 Bulgaria <https://ftp.uni-sofia.bg/CRAN/>

Wirtschaftsuniversität Wien
 Patrick Wessa
 Ghent University Library
 Computational Biology Center at Universidade Estadual de Santa Cruz
 Universidade Federal do Paraná
 Oswaldo Cruz Foundation, Rio de Janeiro
 University of São Paulo, São Paulo
 University of São Paulo, Piracicaba

Sofia University

Download and Install R
 Precompiled binary distributions of the base system and contributed packages, Windows and Mac users most likely want to:

- [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 links.
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. If you do not know what this means, you probably do not want to do it!

- The latest release (2020-04-24, Arbor Day) [R-4.0.0.tar.gz](#), read [what's new](#) in the latest version.

Figure 8: Interface do usuário no RStudio

Operadores	Descrição
<code>==</code>	Igual a
<code>!=</code>	Diferente de
<code><</code>	Menor que
<code>></code>	Maior que
<code><=</code>	Menor ou igual a
<code>>=</code>	Maior ou igual a
<code>x y</code>	x OU y
<code>x & y</code>	x E y

5. Armazenamento de dados

Existem quatro tipos básicos de dados no R, os numéricos, os caracteres, os lógicos e os complexos. Os mais utilizados são os numéricos, dados compostos por números, e os de caracteres, compostos por letras ou palavras.

As informações imputadas no R serão armazenadas na memória do programa e passam a ser denominadas de *objetos*. Para criar um *objeto*, basta associar à informação de interesse o símbolo `<-` ou `=` antecedido pelo seu respectivo nome. No caso de *objetos* com dados de caracteres, a informação a ser armazenada deve vir entre aspas.

```
#Objeto com dado numérico
x<-10
x

## [1] 10

#Objeto com dado de caracteres
y <- "bola"
y

## [1] "bola"
```

Algumas regras devem ser seguidas na hora de nomear um objeto. - O nome do objeto precisa começar com uma letra, não pode ser um número, não pode conter símbolos referentes a funções ou nome de funções e a “seta” `<-` deve ser estar sempre apontada para o nome do objeto. - Cabe destacar que o R reconhece a diferença entre letras minúsculas e maiúsculas.

6. Arquivos R

O R utiliza um diretório do computador para gravar, ler, importar e exportar arquivos. Para saber o diretório em uso, digite: `> getwd()`

Para alterar o diretório de trabalho, digite: `> setwd("C:/Endereço")`

Os principais tipos específicos de arquivos são: o `*.RData` e o `*.Rhistory`.

O arquivo com extensão `*.RData` representa a área de trabalho na qual são salvos os objetos criados durante uma sessão do R. Para salvar uma área de trabalho basta acessar o menu suspenso do R em: **Arquivo -> Salvar**.

O arquivo com extensão `*.Rhistory`, denominado de histórico, armazena todos os comandos utilizados em uma sessão R. Os arquivos de histórico apresentam o formato de texto e por isso podem ser visualizados como um arquivo `*.txt`. Para salvar um histórico basta acessar o menu suspenso do R em: **Arquivo -> Salvar**

Outra maneira de salvar uma sessão de trabalho no R é acessando o menu suspenso do R em: **Arquivo -> Salvar em arquivo**. Os objetos e comandos serão salvos em um arquivo `*.txt`.

Caso o R esteja em uso a partir do *RStudio* basta salvar o script, um arquivo que contém todas os comandos digitados.

7. Estrutura de Dados

No R, os dados contidos em um objeto podem assumir distintas estruturas. As mais utilizadas são os vetores, as matrizes e os data-frame.

7.1. Vetor

O vetor é a forma mais simples armazenamento de dados. Pode ser um vetor numérico ou de caracteres. A função `c()` é utilizada na criação do vetor.

```
#Vetor com dado numérico
x<-c (10,20,30)
x

## [1] 10 20 30

#Vetor de caracteres
y <- c("João", "Ana", "Pedro Henrique")
y

## [1] "João"          "Ana"           "Pedro Henrique"
```

No caso de um vetor com caracteres representar uma variável categórica, é possível associar os atributos de formato de fator e o nome das categorias ao vetor com as funções `fator()` e `levels()`.

```
w <- factor(c("Masculino","Feminino","Masculino"))
levels(w)

## [1] "Feminino"  "Masculino"
```

7.1.1. Manipulação de vetores

Função `seq()` - lista uma sequência de números no intervalo desejado

```
# Sequência de 1 a 50, com intervalo igual a 1
seq(1,50,1)

## [1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
## [24] 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46
## [47] 47 48 49 50

# Sequência de 1 a 50, com intervalo igual a 5
seq(1,50,5)

## [1] 1 6 11 16 21 26 31 36 41 46
```

A função `rep()` lista números repetidos, quantos números quiser com quantas repetições quiser

```
# Repetição do número 1 três vezes
rep(1,3)

## [1] 1 1 1

# Repetição dos números de 1 a 5 duas vezes
rep(1:5,2)

## [1] 1 2 3 4 5 1 2 3 4 5

# Repetição dos números 0 e 1, repetidos 10 vezes, alternadamente.
rep(c(0,1),10)

## [1] 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1
```

```
# Repetição dos números 0 e 1, repetidos 10 vezes, sem alternar.
c(rep(0,10),rep(1,10))

## [1] 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1
```

```
# Repetição de Sim 10 vezes
rep("Sim",10)

## [1] "Sim" "Sim" "Sim" "Sim" "Sim" "Sim" "Sim" "Sim" "Sim"
```

Seleção de elementos dentro de um vetor. Para isso são utilizados colchetes com a indicação das posições ou os próprios elementos.

```
# Seleciona o primeiro elemento do vetor x
x<- c(10,20,30,40,50)
x[1]

## [1] 10

# Seleciona os elementos na posição 1,3,5
x[c(1,3,5)]
```

```
## [1] 10 30 50
```

```

#Seleciona os elementos que não estão nas posições 2 e 4
x<- c(10,20,30,40,50)
x[-c(2,4)]

## [1] 10 30 50
#Seleciona elemento igual a 20 dentro do vetor x
x[x==20]

## [1] 20
# Seleciona o mínimo entre os elementos do vetor x
x[x==min(x)]

## [1] 10

```

7.1.2. Operações com Vetores

Existe a possibilidade de somar, subtrair, dividir ou multiplicar dois vetores (numéricos).

Cada valor de uma posição de um vetor é somado, subtraído, dividido ou multiplicado pelo valor da posição correspondente do outro vetor.

Atenção: a soma de vetores não pode ser confundida com o somatório de vetores.

7.2 Matrizes

No R uma Matriz é um conjunto de vetores numéricos. Podemos criar uma matriz a partir de um ou mais vetores.

```

#Matriz criada a partir do vetor que varia de 1 a 9
#A matriz é composta pelo preenchimento sucessivo das colunas.
matrix(c(1:9),nrow = 3)

```

```

##      [,1] [,2] [,3]
## [1,]     1     4     7
## [2,]     2     5     8
## [3,]     3     6     9
#A matriz é composta pelo preenchimento sucessivo das linhas.
matrix(c(1:9),nrow = 3,byrow = TRUE)

```

```

##      [,1] [,2] [,3]
## [1,]     1     2     3
## [2,]     4     5     6
## [3,]     7     8     9

```

```

#Matriz organizada por linhas
x <- c(1,2,3)
y <- c(10,20,30)
rbind(x,y)

```

```

##      [,1] [,2] [,3]
## x     1     2     3
## y    10    20    30
#Matriz organizada por colunas
x <- c(1,2,3)
y <- c(10,20,30)
cbind(x,y)

```

```

##      x   y
## [1,] 1 10
## [2,] 2 20
## [3,] 3 30

```

```

x<-matrix(seq(1:9),ncol = 3)
x

```

```

##      [,1] [,2] [,3]
## [1,]     1     4     7
## [2,]     2     5     8
## [3,]     3     6     9

```

```

#Localização da 1ªcoluna
x[,1]

```

```

## [1] 1 2 3

```

```

#Localização da 1ªlinha
x[2,]

```

```

## [1] 2 5 8

```

```

#Localização do elemento x23
x[2,3]

```

```

## [1] 8

```

Operações com Matrizes

```

#Matriz x
x <- matrix(seq(1:9),ncol=3);x

```

```

##      [,1] [,2] [,3]
## [1,]     1     4     7
## [2,]     2     5     8
## [3,]     3     6     9

```

```

#Matriz y
y <- matrix(rep(c(1:3),3),ncol=3);y

```

```

##      [,1] [,2] [,3]
## [1,]     1     1     1
## [2,]     2     2     2
## [3,]     3     3     3

```

Transposta de x

```
#Transposta de x
xt <- t(x);xt

##      [,1] [,2] [,3]
## [1,]     1     2     3
## [2,]     4     5     6
## [3,]     7     8     9

#Multiplicação de elemento por elemento
x*y

##      [,1] [,2] [,3]
## [1,]     1     4     7
## [2,]     4    10    16
## [3,]     9    18    27

#Multiplicação entre as matrizes
x%*%y

##      [,1] [,2] [,3]
## [1,]    30    30    30
## [2,]    36    36    36
## [3,]    42    42    42
```

7.3. Data-frame

O data-frame é uma estrutura de dados semelhante a uma matriz, com a diferença que pode armazenar caracteres além de números.

As informações provenientes de uma pesquisa, geralmente, seguem a estrutura de um data-frame no qual as colunas representam as variáveis de pesquisas e as linhas os indivíduos/unidades experimentais.

```
nome<-c("Júlia", "João","Isabela","Gustavo")
idade <- c(20,18,22,21)
peso <- c(50,63,60,80)
sexo <- factor(c("F","M","F","M"))
dados <- data.frame(nome,idade,peso,sexo);dados

##      nome idade peso sexo
## 1 Júlia    20   50    F
## 2 João     18   63    M
## 3 Isabela  22   60    F
## 4 Gustavo  21   80    M

#Selecionar indivíduos com idade superior a 20 anos
dados1 <- dados[dados$idade>20,];dados1

##      nome idade peso sexo
## 3 Isabela  22   60    F
## 4 Gustavo  21   80    M
```

```

#Selecionar indivíuos do sexo Feminino
dados[dados$sexo=="F",]

##      nome idade peso sexo
## 1   Júlia    20   50   F
## 3 Isabel   22   60   F

#Selecionar indivíudos com peso diferente de 60
dados[dados$peso!=60,]

##      nome idade peso sexo
## 1   Júlia    20   50   F
## 2   João     18   63   M
## 4 Gustavo  21   80   M

```

```

#Seccionar dados apenas de Júlia
dados[1,]

##      nome idade peso sexo
## 1   Júlia    20   50   F

#Ordenar ordem de acordo com a coluna de idade
dados[order(dados$idade),]

##      nome idade peso sexo
## 2   João     18   63   M
## 1   Júlia    20   50   F
## 4 Gustavo  21   80   M
## 3 Isabel   22   60   F

#Ordenar ordem decrescente de acordo com a coluna de idade
dados[rev(order(dados$idade)),]

##      nome idade peso sexo
## 3 Isabel   22   60   F
## 4 Gustavo  21   80   M
## 1   Júlia    20   50   F
## 2   João     18   63   M

```

8. Importação de conjunto de dados para o R

O R possui uma série de comandos para a importação de bancos de dados. Os comandos de leitura de bancos de dados seguem a seguinte estrutura:

```
função("local do arquivo/nome do arquivo", argumentos do comando)
```

Argumentos: `dec = " "`, #Especificar separador decimal. Pode ser `.` ou `,`, `sep = " , "`, # Especificar separador das coluna. Pode ser `;ou`, `header= TRUE` # Especificar se primeira linha é cabeçalho, caso contrário use `FALSE`

```

# Formato .csv
read.csv("c:/data.csv", sep = ",", dec = ".", header = TRUE)

```

```
# Formato .csv
read.csv2("c:/data.csv", sep = ";", dec = ",", header = TRUE)

# Formato .txt
read.table("C:/data.txt", sep = " ", header= TRUE)

# Formato .xlsx
dados <- read.xlsx(file="C:/data.xlsx", header=TRUE)
```