

Aplicações Computacionais em Exploração e Análise de Dados:

Introdução ao Uso do Software R

Lupércio França Bessegato
Dep. de Estatística/UFJF

Apresentação



- Lupércio F. Bessegato
 - √ Professor do Departamento de Estatística/UFJF
 - √ Membro permanente do corpo docente do Mestrado Acadêmico de Administração/UFJF
 - √ Linhas de pesquisa e extensão em Métodos Estatísticos Aplicados à Gestão e à Engenharia
 - √ Site: http://www.ufjf.br/lupercio_bessegato/
 - √ E-mail: lupercio.bessegato@ufjf.edu.br

Introdução ao Uso do R - 2016

2

Introdução ao Uso do Software R

Roteiro Geral



1. Visão geral da linguagem R
2. Descrição de dados
3. Visualização gráfica
4. Análise exploratória de dados
5. Referências

Introdução ao Uso do R - 2016

4

Visão Geral da Linguagem R

Introdução



- Ferramenta de análise estatística de dados com muitos recursos e possibilidades;
- Constantemente atualizado e revisado incluindo novas técnicas e métodos;
- Programa de domínio público:
√ <http://cran.r-project.org>

Download do Programa



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:

- Linux
- Mac OS X
- Windows

R-3.3.2 for Windows (32-bit)

Source Code for R

This directory contains the source code for R. See [here](#) for a 64-bit version.

Note: CRAN does not host source code for R on Windows.

If you want to double-check that the package you have downloaded exactly matches the package distributed by R, you can use the `install.packages()` function and `sum()` for windows: both [graphical](#) and [command line versions](#) are available.

Subdirectories:

- [base](#)
- [contrib](#)

Frequently asked questions

- Does R run under my version of Windows?
- How do I update packages in my previous version of R?
- Should I run 32-bit or 64-bit R?

Please see the [R FAQ](#) for general information about R and the [R Windows FAQ](#) for Windows-specific information.

Comandos



- Para solicitar uma tarefa do R podemos digitar uma linha de comando no console

```
R Console
R version 2.11.0 (2010-04-22)
Copyright (C) 2010 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

R é um software livre e vem sem GARANTIA ALGUMA.
Você pode redistribuí-lo sob certas circunstâncias.
Digite 'license()' ou 'licence()' para detalhes de distribuição.

R é um projeto colaborativo com muitos contribuidores.
Digite 'contributors()' para obter mais informações e
'citation()' para saber como citar o R ou pacotes do R em publica
Digite 'demo()' para demonstrações, 'help()' para o sistema on-li
ou 'help.start()' para abrir o sistema de ajuda em HTML no seu na
Digite 'q()' para sair do R.

> sqrt(4)
[1] 2
> |
```



- Todas as funções do R devem ser digitadas em letras **minúsculas**
 √ O R é sensível a letras maiúsculas e minúsculas.
- Todas as palavras-chaves do R estão em letras minúsculas
- R usa um ponto “.” em vez de virgula “,” quando há números com casas decimais.

Alguns Comandos



- Soma → +
- Subtração → -
- Divisão → /
- Multiplicação → *
- Potenciação → ** ou ^
- Raiz quadrada de n → \sqrt{x}
- Logaritmo natural → $\log(x)$
- Logaritmo decimal → $\log_{10}(x)$

- Consulte Table of useful R commands

√ http://www.ufjf.br/lupercio_bessegato

Table of Useful R commands

Command	Purpose	Command	Purpose
help()	Obtain documentation for a given R command	plot()	Produce a scatterplot
example()	View some examples on the use of a command	qqplot()	Lattice command for producing a qqplot
c(), read()	Enter data manually as a vector in R	lm()	Determine the least square regression line
seq()	Make arithmetic progression vector	anova()	Analysis of variance (see use on results of lm())
rpo()	Make vector of repeated values	predict()	Obtain predicted values from linear model
data()	Load (data into a data.frame) built-in dataset	grow()	grow (observed - predicted) for a model fit to data
View()	View dataset in a spreadsheet-type format	residuals()	residuals from a linear model
write()	Display internal structure of an R object	sample()	take a sample from a vector of data
read.csv(), read.table()	Load into a data.frame an existing data file	replicate()	repeat some process a set number of times
library(), require()	Make available an R add-on package	resample()	produce resampling of values for input vector
dim()	See dimensions (if of matrix) of data.frame	ecdf()	build empirical cumulative distribution function
length()	Give length of a vector	dnorm(), etc.	tools for normal distributions
is()	Identify memory contents	dpois(), etc.	tools for poisson distributions
rm()	Remove an item from memory	pnorm(), etc.	tools for normal distributions
names()	List names of variables in a data.frame	ppois(), etc.	tools for poisson distributions
hist()	Command for producing a histogram	ks.test(), etc.	hypothesis test and confidence interval for 1 proportion
histgram()	Lattice command for producing a histogram	prop.test()	inference for 1 proportion using normal approx.
plot()	Make a stem plot	chisq.test()	perform a chi-square test
table()	List all values of a variable with frequencies	fisher.test()	Fisher test for contingency table
xtabs()	Cross tabulation tables using formulae	t.test()	student t test for inference on population mean
mosaicplot()	Make a mosaic plot	qqnorm(), qqline()	tools for checking normality
cut()	Group values of a variable into larger bins	addmargins()	add marginal sums to an existing table
mean(), median()	Identify "center" of distribution	prop.table()	compute proportions from a contingency table
apply()	apply function to a volume table by factors	par()	query and set graphical settings
summary()	Display 5-number summary and more	power.t.test()	power calculations for 1- and 2-sample t
var(), sd()	Find variance, sd of values in a vector	anova()	compute analysis of variance table for fitted model
sort()	Arrange all values in a vector		
quantile()	Find the position of a quantile in a dataset		
barplot()	Produce a bar graph		
barchart()	Lattice command for producing bar graphs		
boxplot()	Produce a boxplot		
boxplot()	Lattice command for producing boxplots		

Help



- Help no R é bastante completo e autosuficiente.
- Se não conhecemos o comando de alguma técnica podemos chamar a função **help**:
 √ `help(palavra-chave)` ou `?palavra-chave`

Alguns Comandos para Ajuda

- Sempre procure ajuda no Google!

```

help.start()      # ajuda geral
help(hist)       # ajuda sobre a função hist
?hist            # Idem
apropos("test")  # lista todas as funções contendo a string test
example(hist)    # mostra um exemplo da função hist

# procura por hist em manuais de ajuda e mailing lists
RSiteSearch("hist")

# vignettes de pacotes instalados e em uso
vignette()       # mostra vignettes disponíveis
vignette("foo")  # mostra vignette específica

# ajuda sobre conjunto de dados de pacotes instalados e em uso
help(datasetname) # para conjunto de dados datasetname
help(faithful)    # para conjunto de dados faithful
    
```

13

Introdução ao Uso do R - 2016

Help na Web

- R Project
 ✓ <http://www.r-project.org>
- Tutorial de Introdução ao R
 ✓ <http://ecologia.ib.usp.br/bie5782/doku.php?id=start>
- R Graphical Manual
 ✓ http://www.imsbio.co.jp/RGM/R_image_list?page=1&sort=-

14

Introdução ao Uso do R - 2016

Armazenando dados

- Tipos de dados:
 - ✓ Numéricos
 - ✓ Caracteres: compostos por letras ou palavras.
 - ✓ Lógicos:
- Quando os dados são armazenados, eles são chamados de objetos.
- Para armazenar um objeto basta utilizar o símbolo “< - ” ou “= ”.

15

Introdução ao Uso do R - 2016

Tipos de Objetos

- Variável
- Vetor
- Matriz
- Data Frame
- Array
- Lista
- Fatores

16

Introdução ao Uso do R - 2016

Comando de Atribuição – Variável



- Atribuição de valor à variável x (ou qualquer outro ‘nome’)
 - √ $x <- 4$ ou $x = 4$
- Atribuição de valores às variáveis $x1$ e $x2$:
 - √ $x1 = -4$
 - √ $x2 = 6$
- Atribua à variável z o valor da soma de $x1$ e $x2$
- Evite usar acentos ou ç na denominação das variáveis

Introdução ao Uso do R - 2016

17

Nomes Inválidos



```
24e<-4 #O nome do objeto começa com números
12<-2 #O nome do objeto é um número
e*2<- #O nome do objeto contém o símbolo da multiplicação
x->5 #A seta está para o lado errado
```

Introdução ao Uso do R - 2016

18

Operações com as Variáveis



- Calcular:

$$x^{x1} + \frac{z}{x2}$$

```
> x^x1+z/x2
[1] 2.916667
```

Introdução ao Uso do R - 2016

19

Apagar Variável



- Comando `rm(variavel)`
 - √ Cuidado! Não tem undo
 - √ `rm(x,x1,x2,z)` ou `rm(x); rm(x1); rm(x2); rm(z)`
- Apagar tudo!
 - √ `rm(list=ls(all=TRUE))`.

Introdução ao Uso do R - 2016

20

Limpeza do Console



- Para limpar o console usa-se CTRL + L.
-

Vetores



- Coleção unidimensional de dados pontuais de mesmo tipo (números, strings, lógico)
- Comando para armazenar dados em vetor
√ `c(dados)` # concatenar

```
xNum <- c(1,2,5.3,6,-2,4) # vetor numérico  
xCar <- c("um","dois","três") # vetor de caracteres  
xLog <- c(TRUE,TRUE,TRUE,FALSE,TRUE,FALSE) # vetor lógico
```

Vetores – Indexação



- Elementos particulares de uma estrutura de dados

```
xNum[2] # retorna o 2º elemento do vetor  
xCar[c(1,3)] # retorno o 1º e o 3º elementos do vetor
```

Diretório de Trabalho



- Comando para verificar o diretório de trabalho que o R está usando:
√ `getwd()`
- Ideal sempre deixar scripts e dados de trabalho no mesmo diretório!
- Comando para mudar o R para seu diretório
√ `setwd("caminho_ate_diretorio")`
√ Ex.: "D:/Lupercio/Documents"

Mudança Diretório – Barra de Ferramentas



- Sugestão:
 - ✓ Sempre mude para seu diretório de trabalho quando iniciar a sessão em R
 - ✓ Guarde nele seus dados, gráficos, scripts, etc



Introdução ao Uso do R - 2016

25

Outros Comandos Úteis



- Verifique se mudança funcionou com o comando `getwd()`:
- Verifique se seu diretório de trabalho está vazio:
 - ✓ `dir()`
- Verifique se o workspace está vazio:
 - ✓ `ls()`

Introdução ao Uso do R - 2016

26

Uso de Script



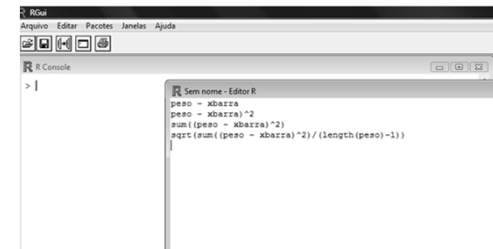
- Facilita para:
 - ✓ Correção ou expansão de comandos
 - ✓ Repetição de comandos
 - ✓ Armazenamento de resultados



Introdução ao Uso do R - 2016

27

Compilação do Script



- Usa-se a tecla F5 para compilar:
 - ✓ A linha em que se encontra o cursor (no script)
 - ✓ As linhas selecionadas (no script)
- Resultado compilação no console

Introdução ao Uso do R - 2016

28

Vantagens



- Facilidade para corrigir os comandos ou valores
- Possibilidade de armazenar todos os resultados
- Repetição dos passos corretos de toda a sessão de trabalho
- Trabalho 'limpo'
 - √ Para limpar o console usa-se CTRL + L

Introdução ao Uso do R - 2016

29

Vetores Numéricos



- Comando concatenar: c(dados)
 - √ Criar o vetor peso com os valores 60, 72, 57, 90, 95, 72

```
> peso<- c(60, 72, 57, 90, 95, 72)
> peso
[1] 60 72 57 90 95 72
> |
```

- √ Criar o vetor altura com os valores 1,75; 1,80; 1,65; 1,90; 1,74; 1,91

```
> altura= c(1.75, 1.80, 1.65, 1.90, 1.74, 1.91)
> altura
[1] 1.75 1.80 1.65 1.90 1.74 1.91
> |
```

Introdução ao Uso do R - 2016

30

Outras Funções para Criar Vetores



- Função seq
 - √ Lista sequência de números que quiser, no intervalo que quiser
 - √ `seq(from=1, to=1, length.out = NULL, by = ((to - from)/(length.out - 1)), along.with = NULL)`
 - from, to: início e fim da sequência
 - by: incremento da sequência
 - length.out: tamanho desejado da sequência
 - along.with: comprimento do objeto do argumento

Introdução ao Uso do R - 2016

31

Outras Funções para Criar Vetores




- Função rep
 - √ Lista números repetidos, quantos números quiser com quantas repetições quiser
 - √ `rep(x, times = 1, length.out = NA, each = 1)`
 - x: um vetor ou um fator ou uma lista
 - times: vetor com o nº de vezes de repetições de cada elemento
 - length.out: tamanho desejado da sequência
 - each: cada elemento de x é repetido each vezes
 - along.with: comprimento do objeto do argumento

Introdução ao Uso do R - 2016

32

Exemplos



```


> seq(1,100,1) #Sequência de números de 1 até 100, com intervalo de 1 número
[1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
[19] 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36
[37] 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54
[55] 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72
[73] 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90
[91] 91 92 93 94 95 96 97 98 99 100

> seq(1,100,2) #Sequência de números de 1 até 100, com intervalo de 2 números
[1] 1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49
[26] 51 53 55 57 59 61 63 65 67 69 71 73 75 77 79 81 83 85 87 89 91 93 95 97 99
    
```

33

Introdução ao Uso do R - 2016

Exemplos



```

> seq(1,100,10) #Sequência de números de 1 até 100, com intervalo de 10
[1] 1 11 21 31 41 51 61 71 81 91


> seq(1,20,1) #Sequência crescente de 1 até 20
[1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

> seq(20,1,-1) #Sequência decrescente de 20 até 1
[1] 20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1
    
```

34

Introdução ao Uso do R - 2016

Exemplos



```

> rep(1,10) #Número 1 repetido 10 vezes
[1] 1 1 1 1 1 1 1 1 1 1

> rep(1:5,2) #Sequência de 1 até 5, repetida 2 vezes
[1] 1 2 3 4 5 1 2 3 4 5


> rep(c(1,4),3) #Números 1 e 4 repetidos 3 vezes, alternadamente
[1] 1 4 1 4 1 4

> c(rep(1,3),rep(4,3)) #Números 1 e 4 repetidos 3 vezes, sem alternar
[1] 1 1 1 4 4 4
    
```

35

Introdução ao Uso do R - 2016

Alguns Vetores do R



- LETTERS
- letters
- month.name
- month.abb

36

Introdução ao Uso do R - 2016

Algumas Estatísticas Descritivas



- Média:
 $\sqrt{\text{mean}(\text{dados})}$
- Desvio-padrão:
 $\sqrt{\text{sd}(\text{dados})}$
- Variância
 $\sqrt{\text{var}(\text{dados})}$
- Mediana
 $\sqrt{\text{median}(\text{dados})}$

Introdução ao Uso do R - 2016

37

- Estatísticas descritivas dos vetores *peso* e *altura*



```
# calcula estatísticas descritivas dos vetores peso e altura

mean(altura)      # media do vetor altura
sd(altura)        # desvio padrão do vetor altura
var(altura)       # variância do vetor altura
sd(altura)^2
median(altura)    # mediana do vetor altura
sum(altura)       # soma dos elementos do vetor altura
length(altura)   # quantidade de elementos do vetor altura
```

\sqrt{E} o coeficiente de variação?

Introdução ao Uso do R - 2016

38

Aritmética de Vetores



- O R pode manipular vetores como simples objetos
- Cálculo do Índice de Massa Corporal

$$IMC = \frac{\textit{peso}}{\textit{altura}^2}$$

```
> imc<-peso/altura^2
> imc
[1] 19.59184 22.22222 20.93664 24.93075 31.37799 19.73630
> |
```

Introdução ao Uso do R - 2016

39

- Podemos digitar uma sequência de comandos em um editor de textos (p.ex. o Bloco de Notas, e depois colar no console para execução das tarefas solicitadas



```
sum(peso) ; length(peso) ; xbarra=sum(peso)/length(peso)
```

Introdução ao Uso do R - 2016

40

Vetores Lógicos



- Podem tomar os valores TRUE ou FALSE (ou NA)
 $\sqrt{c(T,T,F,T)}$
- Expressões de Relação

```
>
> 1m<25
[1] FALSE FALSE FALSE TRUE FALSE
>
> |
```

Operadores Lógicos



Símbolo	Função
<	Menor que
>	Maior que
<=	Menor que ou igual a
>=	Maior que ou igual a
= =	Igual a
! =	Não igual a
&	E (para combinar expressões)
	Ou (para combinar expressões)
!	Não (para combinar expressões)

Matrizes



- Todas as colunas tem de ter o mesmo comprimento e conter o mesmo tipo de dados (numérico, caracter, etc.)
- `matrix(vector, nrow=r, ncol=c, byrow=F, dimnames=list(char_vector_rownames, char_vector_colnames))`
 $\sqrt{\text{byrow=TRUE: matriz deve ser montada pelas linhas.}}$
 $\sqrt{\text{dimnames: nomes para linhas e colunas (opcional)}}$

Exemplos




- Criação de matrizes

```
# gera matriz numérica 5 x 4
y <- matrix(1:20, nrow=5, ncol=4)

# outro exemplo
celulas <- c(1, 26, 24, 68)
rnomes <- c("R1", "R2")
cnomes <- c("C1", "C2")
minha.matriz <- matrix(celulas, nrow=2, ncol=2, byrow=TRUE,
  dimnames=list(rnomes, cnomes))
```

Exemplos




- Matrizes a partir de combinação de objetos

```

x <- c(11, 12, 13)           # Cria um vetor `x` com 3 valores
y <- c(55, 33, 12)         # Cria um vetor `y` com 3 valores
rbind(x, y)                # Combina os vetores em linhas e
                           # cria uma matriz 2 x 3
cbind(x, y)                # Combina os vetores em colunas e
                           # cria uma matriz 3 x 2
biometria <- cbind(peso,altura,imc) # cria matriz de dados
biometria
    
```

Introdução ao Uso do R - 2016 45

- Atribuindo nomes às linhas e colunas




```

> x<-matrix(1:12,nrow=3,byrow=F)
> rownames(x)<-c('Primeira','Segunda','Terceira')
> colnames(x)<-LETTERS[1:4]
> x
      A B C D
Primeira 1 4 7 10
Segunda  2 5 8 11
Terceira 3 6 9 12
> |
    
```

Introdução ao Uso do R - 2016 46

Desconstrução de Matriz




- Podemos desconstruir uma matriz aplicando a função `c`
- √ Combina todas as colunas em um vetor

```

c(minha.matriz)           # Combina todas as colunas em um vetor
    
```

Introdução ao Uso do R - 2016 47

Matrizes – Indexação



- Identificação de linhas, colunas ou elementos usando subscritos.

```

# Cria matriz com 20 números aleatórios de normal padrão
x <- matrix(rnorm(20), ncol = 4)

x[,4]                    # 4ª coluna da matriz
x[3,]                    # 3ª linha da matriz
x[2:4,1:3]               # linhas 2,3 e 4 das colunas 1,2 e 3
X[c(1,3), c(2, 4)]      # linhas 1 e 3 das colunas 2 e 4
    
```

Introdução ao Uso do R - 2016 48

Arrays



- Arrays são similares a matrizes, mas podem ter mais de duas dimensões.

Introdução ao Uso do R - 2016

49

Fatores



- Estrutura de dados para variável categórica:
 - √ Há análises que o R precisar distinguir códigos categóricos dos numéricos
 - √ Ex.: média de variável categórica

```
# variável sexo com 20 componentes "masc" e
# 30 "fem"
sexo <- c(rep("masc",20), rep("fem", 30))
sexo <- factor(sexo)
# armazena sexo como 20 1's and 30 2's e associa
# internamente 1=fem, 2=masc (alfabeticamente)
# R agora trata sexo com uma variável nominal
summary(sexo)
```

Introdução ao Uso do R - 2016

50

Variável Categórica Codificada Numericamente



```
# variável codificando nível de dor de 5 pacientes
dor <- c(0, 3, 2, 2, 1) # cria o vetor dor
fdor <- factor(dor, level = 0:3) # Transforma em categórica

# muda nome dos níveis do fator
levels(fdor) <- c("nenhum", "leve", "medio", "grave")
fdor
levels(fdor) # extrai os nomes dos níveis do fator
as.numeric(fdor) # extrai códigos numéricos dos níveis do fator
```

Introdução ao Uso do R - 2016

51

Fator Ordenado



- Usado para representação de variável ordinal

```
# Cria string denominada nivel (nível de negócios).
nivel <- c("baixo", "medio", "baixo", "baixo", "baixo", "baixo",
"medio", "baixo", "medio", "medio", "medio", "medio",
"alto", "alto", "baixo", "medio", "medio", "baixo", "alto")
is.factor(nivel) # verifica se nivel é fator
is.character(nivel) # verifica se nivel é string
fnivel <- factor(nivel) # transforma nivel em fator
is.factor(fnivel) # verifica se nivel é fator
levels(fnivel) # extrai os nomes dos níveis do fator

# Criando fator com a ordem correta dos níveis
fnivel.ord <- factor(nivel, levels = c("baixo", "medio", "alto"))
levels(fnivel.ord)
# ordenando fator
fnivel.ord2 <- ordered(fnivel, levels = c("baixo", "medio", "alto"))
levels(fnivel.ord2)
```

Introdução ao Uso do R - 2016

52

Listas



- Coleção ordenada de objetos (componentes)
 - √ Permite que você reúna objetos de diferentes tipos sob o mesmo nome.

```
# Exemplo de uma lista com 4 componentes
# uma string, um vetor numérico, uma matriz e um escalar
w <- list(nome = "Thiago", numeros=peso, matriz=biometria, idade=62)
w
w$matriz

# Criação de lista contendo duas listas
v <- c(lista1,lista2)
```

Introdução ao Uso do R - 2016

53

Listas – Indexação



- Identificação dos objetos de uma lista usando `[[]]`.

```
minhalista[[2]] # 2º componente da lista
minhalista[["nome_dele"]] # componente denominado nome_dele na lista
```

Introdução ao Uso do R - 2016

54

Data Frames



- É mais geral que uma matriz
 - √ Colunas diferentes de mesmo comprimento podem ter diferentes tipos de dados (numérico, caracter, lógico, fator, etc.)

```
d <- c(1,2,3,4)
e <- c("vermelho", "branco", "vermelho", NA)
f <- c(TRUE,TRUE,TRUE,FALSE)
meubanco <- data.frame(d,e,f)
names(meubanco) <- c("ID","Cor","verificou") # nome das variáveis
```

Introdução ao Uso do R - 2016

55

Data Frames – Indexação



- Há várias maneiras para identificar os elementos de um data frame

```
meubanco[3:5] # colunas 3,4 e 5 do data frame
meubanco[c("ID", "verificou")] # columns ID e do data frame
meubanco$Cor # variable verificou do data frame
```

Introdução ao Uso do R - 2016

56

Exemplo – Consumo de Energia



- Consumo de energia pré-menstrual:
√ `consumo.pre = c(5260, 5470, 5640, 6180, 6390)`
- Consumo de energia pós-menstrual:
√ `consumo.pos = c(3910, 4220, 3885, 5160, 5645)`
- Conjunto de dados (Data frame):
√ `dados.altman = data.frame(consumo.pre, consumo.pos)`
- Tamanho do conjunto de dados:
√ `dim(dados.altman)`

Introdução ao Uso do R - 2016

57

```
> consumo.pre = c(5260, 5470, 5640, 6180, 6390)
> consumo.pos = c(3910, 4220, 3885, 5160, 5645)
> dados.altman = data.frame(consumo.pre, consumo.pos)
> dados.altman
  consumo.pre consumo.pos
1      5260      3910
2      5470      4220
3      5640      3885
4      6180      5160
5      6390      5645
> |
```



- Dados são emparelhados
- Variáveis acessíveis usando notação \$
√ `dados.altman$consumo.pre`

Introdução ao Uso do R - 2016

58

Indexação



- Consumo pré-menstrual mulher #5
√ `consumo.pre[5]`
- Consumo pré-menstrual mulheres #1, 3, 5
√ `consumo.pre[c(1, 3, 5)]`
- Consumo pré-menstrual 3 primeiras mulheres
√ `consumo.pre[1:3]`

Introdução ao Uso do R - 2016

59

- Consumo pré-menstrual todas exceto mulheres #3, 5 e 7


√ `consumo.pre[-c(1, 3, 5)]`

```
> consumo.pre[5]
[1] 6390
> consumo.pre[c(3, 5, 7)]
[1] 5640 6390 NA
> consumo.pre[1:3]
[1] 5260 5470 5640
> consumo.pre[-c(1, 3, 5)]
[1] 5470 6180
> impar=c(1, 3, 5)
> consumo.pre[-impar]
[1] 5470 6180
> dados.altman$consumo.pre[5]
[1] 6390
> |
```



Introdução ao Uso do R - 2016

60



- Pode-se indexar pelo do Data frame
 √ dados.altman\$consumo.pre[5]
- Todas as medidas da mulher #5:
 √ dados.altman[5,]
- Consumos de energia pós-menstrual 3 primeiras mulheres
 √ dados.altman[1:3,2]

61

Introdução ao Uso do R - 2016




```

> dados.altman$consumo.pre[5]
[1] 6390
>
> dados.altman[5,]
consumo.pre consumo.pos
5          6390          5645
>
>
> dados.altman[1:3,2]
[1] 3910 4220 3885
>
> |
    
```

62

Introdução ao Uso do R - 2016

Seleção Condicional



```


consumo.pos[consumo.pre>6000]

consumo.pos[consumo.pre>6000 & consumo.pre<=6500]

consumo.pre>6000 & consumo.pre<=6500
    
```

63

Introdução ao Uso do R - 2016



```

>
> consumo.pos[consumo.pre>6000]
[1] 5160 5645
>
>
> consumo.pos[consumo.pre>6000 & consumo.pos <=6500]
[1] 5160 5645
>
>
> consumo.pre>6000 & consumo.pos <=6500
[1] FALSE FALSE FALSE TRUE TRUE
>
> condicao=consumo.pre>6000 & consumo.pos <=6500
>
> consumo.pos[condicao]
[1] 5160 5645
>
> |
    
```

64

Introdução ao Uso do R - 2016

Ordenação



- Comando sort: ordena os elementos do vetor
 - √ Default é crescente
 - √ `sort(consumo.pos)`

- Ordenação decrescente

√ `sort(consumo.pos, decreasing=TRUE)`

```
>
> consumo.pos
[1] 3910 4220 3885 5160 5645
>
>
> sort(consumo.pos)
[1] 3885 3910 4220 5160 5645
>
> sort(consumo.pos, decreasing=TRUE)
[1] 5645 5160 4220 3910 3885
>
> |
```

Introdução ao Uso do R - 2016

65

- Comando order



- √ Fornece número de ordem de cada elemento do vetor
- √ `order(consumo.pos)`

```
>
> order(consumo.pos)
[1] 3 1 2 4 5
>
>
> ordem<-order(consumo.pos)
>
>
> consumo.pos[ordem]
[1] 3885 3910 4220 5160 5645
>
>
> consumo.pre[ordem]
[1] 5640 5260 5470 6180 6390
>
> |
```

Introdução ao Uso do R - 2016

66

Funções Úteis



- Utilizem os comandos em objetos criados

```
length(object) # number of elements or components
str(object)    # structure of an object
class(object)  # class or type of an object
names(object)  # names

c(object,object,...) # combine objects into a vector
cbind(object, object, ...) # combine objects as columns
rbind(object, object, ...) # combine objects as rows

object        # prints the object

ls()          # list current objects
rm(object)    # delete an object

newobject <- edit(object) # edit copy and save as newobject
fix(object)    # edit in place
```

Introdução ao Uso do R - 2016

67

Criação de Data-frame




- Maneira diferente de criar data-frame
 - √ `edit(data.frame())`
 - √ `dados.novo <- edit(data.frame())`
- É aberta janela com planilha e pode-se digitar informações do banco de dados.
- Para nomear a 1ª. variável, clica-se em “var1”
 - √ Janela para nomear a variável e seu tipo (numérica ou caractere).

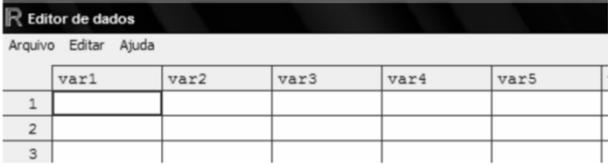
Introdução ao Uso do R - 2016

68

Exemplo




```
edit(data.frame())
```



Introdução ao Uso do R - 2016 69


Carregando Pacotes



- Pacotes:
 - √ Conjuntos de funções específicas do R
 - √ No repositório do R está armazenada uma quantidade muito grande de pacotes que tem funções para um certo conjunto de tarefas
 - √ Para usar um pacote:
 - baixar o pacote (download) do repositório
 - carregar o pacote na sua área de trabalho.

Introdução ao Uso do R - 2016 70


Pacotes




- Quais pacotes estão disponíveis na sua instalação de R?
 - √ `library()`
- Interface hipertexto de ajuda:
 - √ `help.start()`
 - √ Escolher o link “Packages”
 - √ Clique no nome de um dos pacotes
 - Lista todos os objetos que este pacote contém.

Introdução ao Uso do R - 2016 71

• Página do R Project




Package Index 

Packages in C:\Program Files\R\R-3.3.1\library

ABCDEFGHIJKLMNOPQRSTUVWXYZ

<p>abind abundant acepack ade4 asbio asserthat base base64enc BB BH BHH2 BiasedUn BioGenerics BiocInstaller</p>	<p>Combine Multidimensional Arrays Abundant regression and high-dimensional principal fitted components ace() and avas() for selecting regression transformations Analysis of Ecological Data : Exploratory and Euclidean Methods in Environmental Sciences A Collection of Statistical Tools for Biologists Easy pre and post assertions. The R Base Package Tools for base64 encoding Solving and Optimizing Large-Scale Nonlinear Systems Boost C++ Header Files Useful Functions for Box, Hunter and Hunter II Biased Urn Model Distributions S4 generic functions for Bioconductor Install/Update Bioconductor, CRAN, and github Packages</p>
--	---


Introdução ao Uso do R - 2016 72



- Quais pacotes estão carregados na sua sessão?
`√ search()`
- Instalação de pacote direto do R
`√ install.packages("vegan")`

73

Introdução ao Uso do R - 2016



Exemplo

- Geração de amostra aleatória:

```
# 15 números aleatórias de uma distribuição normal, com média 1 e
# desvio-padrão=3


x1 <- rnorm(n = 15, mean = 1, sd = 3)
hist(x1)      # histograma de x1
truehist(x1)  # outro tipo de histograma
```

- Uso de função do pacote MASS:

```
search()
library(MASS)      # carrega pacote MASS
search()
truehist(x1)
help(package=MASS) # ajuda sobre o pacote
```

74

Introdução ao Uso do R - 2016




Demos

- Oferece demonstrações de como usar funções

```
# todos os demos (pacotes carregados)
demo()
# todos demos (pacotes instalados)
demo(package = .packages(all.available = TRUE))
demo(plotmath)      # demo sobre anotações em gráfico
demo(graphics)      # demos sobre tipos de gráficos
```

75

Introdução ao Uso do R - 2016




Help

- Bom site para verificar outros tipos de ajuda
<http://stackoverflow.com/questions/15289995/how-to-get-help-in-r>

76

Introdução ao Uso do R - 2016

Conjuntos de Dados do R




- R traz vários conjuntos de dados internos, que são geralmente usados em demos ou exemplos
- Comando para ver a lista dos conjuntos de dados carregados:
 $\sqrt{\text{data}()}$

77

Introdução ao Uso do R - 2016

Carregamento Built-in Data Set



- Carregamento do conjunto de dados:
 $\sqrt{\text{mtcars}}$: Motor Trend Car Road Tests

```
# Conjunto de dados: mtcars: Motor Trend Car Road Tests
data(mtcars)      # carregamento
head(mtcars, 8)   # Print das primeiras 8 linhas
help(mtcars)      # informações sobre o banco
```


- Manipulação do conjunto de dados

```
nrow(mtcars)      # Número de linhas (observações)
ncol(mtcars)      # Número de colunas (linhas)
str(mtcars)       # estrutura do objeto
```

78

Introdução ao Uso do R - 2016

- Colunas como vetores




```
mtcars$mpg        # acesso à variável mpg
mtcars$disp       # acesso à variável disp
pairs(mtcars, main = "Conjunto mtcars")
plot(mpg, disp)
# acesso às variáveis sem especificar banco
attach(mtcars)
plot(mpg, disp)   # plot das variáveis mpg e disp
objeto <- lm(disp ~ mpg) # regressão linear
objeto           # resultados
summary(objeto)  # ajuste do modelo de regressão
abline(lm(disp ~ mpg)) # plot da reta de regressão
# estratificação do plot pela variável cyl
coplot(mpg ~ disp | as.factor(cyl), data = mtcars,
       panel = panel.smooth, rows = 1)
detach(mtcars)   # interromper acesso (final do trabalho)
```

79

Introdução ao Uso do R - 2016

Exportação de Bancos de Dados



- Dados exportados para outras extensões

```
# para arquivo texto (delimitado por Tab)
# formato do R
write.table(mtcars, "mtcars.txt", sep="\t")
# muda ponto decimal
write.table(mtcars, "mtcars.txt", dec = ",", sep="\t")

# para arquivo csv (Comma Separated Value)
write.csv(mtcars, "mtcars.csv") # formato do R
write.csv2(mtcars, "mtcars.csv") # formato brasileiro

# para planilha Excel
library(xlsx) # requer versão recente do R
write.xlsx(mtcars, "mtcars.xlsx")
To SPSS

# write out text datafile and
# an SPSS program to read it
```

Exportação de Bancos de Dados



- Dados exportados para outras extensões

```
# para arquivo do SPSS
# escreva arquivo de texto e um programa SPSS para sua leitura
library(foreign)
write.foreign(mtcars, "mtcars.txt", "mtcars.sps", package="SPSS")

# Para arquivo SAS
# escreva arquivo de texto e um programa SAS para sua leitura
library(foreign)
write.foreign(mtcars, "mtcars.txt", "mtcars.sas", package="SAS")

# Para arquivo Stata
# exportação de data frame para formato binário do Stata
library(foreign)
write.dta(mtcars, "mtcars.dta")
```

Introdução ao Uso do R - 2016

81

Comando sink



- Desvia as saídas do R para uma conexão (e desliga a conexão).

```
# Grava saída de uma análise
sink('analise-saida.txt')
set.seed(12345) # semente aleatoria
x <- rnorm(10, 10, 1)
y <- rnorm(10, 11, 1)
# alguma saída
cat(sprintf("x tem %d elementos:\n", length(x)))
print(x)
cat("y =", y, "\n")
cat("=====\n")
cat("Teste-t entre x e y\n")
cat("=====\n")
t.test(x,y)
# para o arquivo
sink()
```

Introdução ao Uso do R - 2016

82

Importação de Dados



- Importação para o R de dados disponíveis em formato eletrônico
- Comandos
 - √ read.table
 - √ read.csv
 - √ read.csv2
 - √ outros

Introdução ao Uso do R - 2016

83

Comando read.table



- Importação de dados em formato texto (arquivo do tipo ASCII)

```
# arquivo sem cabeçalho
ex01 <- read.table("gam01.txt")
head(ex01)
# arquivo com cabeçalho na 1a. linha
ex02 <- read.table("exemplo02.txt", head=T)
head(ex02)
# arquivo com campos separados por : e decimais, por vírgula
ex03 <- read.table("dadosfic.csv", head=T, sep=":", dec=",")
head(ex03)
# leitura direta pela web
ex04 <- read.table("http://www.leg.ufpr.br/~paulojus/dados/gam01.txt")
head(ex04)
# leitura de arquivo com informações em suas 1as. linhas
teste.data <- read.table("test2.txt", skip=4, header=TRUE, sep="\t")
head(teste.data)
```

Introdução ao Uso do R - 2016

84

Comando read.csv



- Importação de dados em formato csv (Comma Separated Value)

```
# arquivo csv formato inglês
aereas.data <- read.csv("AirPassengers.csv", header=TRUE)
head(aereas.data)
# arquivo csv - comando read.table
aereas.data2 <- read.table("AirPassengers.csv", header=TRUE, sep=",")
head(aereas.data2)
# arquivo csv - gravado em formato brasileiro
solo <- read.csv("solo.csv", header = TRUE, dec = ".", sep = ";")
head(solo)
# comando read.csv2 (leitura direta de csv em formato brasileiro)
solo2 <- read.csv2("solo.csv", header = TRUE)
head(solo2)
```

Introdução ao Uso do R - 2016

85

Importação do Excel



- Importação direta de planilhas com extensão xlsx

```
# lê a primeira guia da planilha meuexcel.xlsx
# nomes das variáveis na primeira linha
library(xlsx)
dados <- read.xlsx("meuexcel.xlsx", 1)

# lê a guia na planilha chamada minhaguia
dados <- read.xlsx("meuexcel.xlsx", sheetName = "minhaguia")
```

Introdução ao Uso do R - 2016

86

Importação do SPSS



- Importação direta de arquivos com extensão sav

```
# exportar os dados no SPSS
get file='meusdados.sav'.
export outfile='meusdados.por'.

# no R
library(Hmisc)
dados <- spss.get("meusdados.por", use.value.labels=TRUE)
# última opção converte níveis em fatores no R
```

Introdução ao Uso do R - 2016

87

Importação do Stata



- Importação direta de arquivos com extensão dta

```
# importação do arquivo do Stata
library(foreign)
dados <- read.dta("meusdados.dta")
```

Introdução ao Uso do R - 2016

88

Importação do SAS



- Importação direta de arquivos com extensão xpt

```
# exportar os dados no SAS
libname out xport 'c:/mydata.xpt';
data out.mydata;
set sasuser.mydata;
run;

# in R
library(Hmisc)
dados <- sasxport.get("c:/mydata.xpt")
# variáveis de caracteres são convertidas para fatores no R
```

Resumos Numéricos



- Principais medidas resumo para exploração de conjunto de dados:

√ Medidas de posição:

- Média
- Mediana

√ Medidas de dispersão:

- Desvio-padrão (variância)
- Distância interquartilica

Funções Mais Usadas



Símbolo	Função
<code>sum(x)</code>	Soma dos elementos de x
<code>prod(x)</code>	Produtório dos elementos de x
<code>max(x)</code>	Elemento máximo de x
<code>min(x)</code>	Elemento mínimo de x
<code>range(x)</code>	Elementos máximo e mínimo de x
<code>length(x)</code>	Quantidade de elementos do vetor x
<code>mean(x)</code>	Média dos elementos de x
<code>median(x)</code>	Mediana dos elementos de x
<code>var(x)</code>	Variância dos elementos de x
<code>sd(x)</code>	Desvio padrão dos elementos de x
<code>quantile(x, p)</code>	Quantil dos elementos de x , correspondente a p
<code>corr(x, y)</code>	Correlação entre os elementos de x e y

Exemplo




- Conjunto de dados de turma de alunos com as variáveis:

- √ Sexo
- √ Peso
- √ altura

```
# Carregando e conhecendo o banco

Dados <- read.csv2(file="turma.csv")
head(dados)
dim(dados) # tamanho do conjunto de dados
attach(dados)
head(Altura)
head(Peso)
head(Sexo)
is.factor(Sexo) # verifica se categórica está como fator
```

Função length(x)




- Calcula quantidade de elementos de vetor
- Verifica a quantidade de variáveis:

```
length(Peso)           # Calcula o tamanho da amostra
# usada no conjunto de dados
length(dados)         # Informa quantidade de variáveis
```

Introdução ao Uso do R - 2016 93

Valores Extremos




- Funções
 - √ `min(x)`: determina o menor valor da variável
 - √ `max(x)`: determina o maior valor da variável
 - √ `range(x)`: determina o menor e o maior valor da variável

```
min(Peso)              # Menor peso observado
max(Peso)              # Maior peso observado
range(Peso)            # Menor e maior peso observado (vetor)
```

Introdução ao Uso do R - 2016 94

Outras Funções



- Soma e produto:
 - √ `sum(x)`: soma todos os elementos de `x`
 - √ `prod(x)`: multiplica todos os elementos de `x`.


```
sum(Peso)              # soma todos os pesos observados
prod(Peso)             # multiplica todos os pesos observados
sum(Peso)/length(Peso) # cálculo do peso médio
```

- Média
 - √ `mean(x)`: médias dos elementos de `x`

```
mean(Peso)             # média dos pesos observados
mean(Altura)          # média das alturas observada
```

Introdução ao Uso do R - 2016 95

Função Aplicada a Grupos da Variável



- Determinação da média de alguns valores da variável.
 - √ Aplicando diretamente o comando `mean`


```
mean(Peso[Sexo=="F"]) # média dos pesos das alunas
mean(Peso[Sexo=="M"]) # média dos pesos das alunas
```

- √ Comando `tapply`:
 - Aplica função a cada a cada grupo de valores dado por uma combinação única dos níveis de certos fatores.

```
tapply(Peso, Sexo, FUN = mean) # Calcula peso médio por Sexo
```

Introdução ao Uso do R - 2016 96

Mediana




- `median(x)`: calcula mediana da variável observada.

<code>median(Peso)</code>	# mediana dos pesos de todos os alunos
<code>median(Altura)</code>	# mediana das alturas de todos os alunos
<code>median(Peso[Sexo=="M"])</code>	# mediana dos pesos dos alunos

Introdução ao Uso do R - 2016 98

Dispersão




- $\sqrt{\text{var}(x)}$: variância dos elementos de x
- $\sqrt{\text{sd}(x)}$: desvio padrão dos elementos de x .

<code>var(Peso)</code>	# variância do peso de todos os alunos
<code>sd(Peso)</code>	# desvio padrão do peso de todos os alunos
<code>var(Peso[Sexo=="F"])</code>	# variância do peso das alunas
<code>sd(Altura[Sexo=="M"])</code>	# desvio padrão da altura dos alunos

Introdução ao Uso do R - 2016 99

Quantis




- `quantile(x, p)`: determina quantil, onde x é a variável observada e p é uma probabilidade.

<code>quantile(Peso, 0.7)</code>	# Percentil 70 dos pesos
<code>quantile(Peso, c(0.25, 0.75))</code>	# 1° e 3° quartis dos pesos
<code>quantile(Peso[Sexo=="F"], 0.7)</code>	# Percentil 70 das alunas
<code>quantile(Peso, 0.5)</code>	# mediana de todos os pesos

Introdução ao Uso do R - 2016 100

Correlação




- Relação linear entre duas variáveis quantitativas
 - $\sqrt{\text{corr}(x, y)}$: coeficiente de correlação linear entre as variáveis x e y .

<code>cor(Peso, Altura)</code>	# correlação linear entre peso e altura
<code>cor.test(Peso, Altura)</code>	# teste de significância da correlação

- Opções do comando:
 - $\sqrt{\text{cor}(x, y, \text{method}='pearson')}$: default
 - $\sqrt{\text{cor}(x, y, \text{method}='spearman')}$
 - $\sqrt{\text{cor.test}(x, y, \text{method}='pearson')}$: default
 - $\sqrt{\text{cor.test}(x, y, \text{method}='spearman')}$

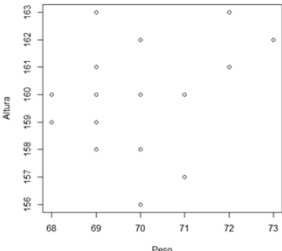
Introdução ao Uso do R - 2016 101

Gráfico de dispersão




√ `plot(x,y)`: gráfico da relação das variáveis quantitativas x e y .

```
plot(Peso, Altura) # gráfico de dispersão entre Peso e Altura
```



Introdução ao Uso do R - 2016 102

Resumo de dos Dados




- Resumo de 5 números e média
 - √ `summary(x)`: fornece o mínimo, 1º quartil, Mediana, 3º quartil, máximo e média dos elementos de x .
 - √ Usada apenas com variáveis quantitativas

```
summary(Peso) # resumos da variável Peso
summary(Altura) # resumos da variável Altura
```

Introdução ao Uso do R - 2016 103

Tabelas




- Resumo da frequência dos níveis de variável categórico (ou variável discreta).
- `table(x)`:

```
table(Sexo) # tabela de contingência de Sexo
prop.table(table(Sexo)) # tabela de frequência relativa
```

Introdução ao Uso do R - 2016 104

Tabela de Frequência – Variável Contínua




- Não há comando específico no R. É necessário construí-la:
 - √ Exemplo com o conjunto de dados `faithful`.

```
duracao <- faithful$eruptions
range(duracao)
# sequencia para intervalo dos dados (aproximado)
breaks <- seq(1.5, 5.5, by=0.5)
# aloca elementos em sub-intervalos de tamanho 0.5
duracao.cut <- cut(duracao, breaks, right=FALSE)
# calcula a frequência de erupções em cada sub-intervalo
duracao.freq <- table(duracao.cut)
# tabela com os resultados
cbind(duracao.freq)
```

Introdução ao Uso do R - 2016 105

Histograma



- Visualizando a variável duracao:


```

hist(duracao)
hist(duracao, label = T)      # histograma com frequências

duracao.hist <- hist(duracao) # cria objeto com o histograma
str(duracao.hist)            # estrutura do objeto histograma
# limites dos sub-intervalos do histograma
duracao.hist$breaks
# frequência de valores em cada ub-intervalo
duracao.hist$counts
    
```

Introdução ao Uso do R - 2016 106

Geração de um Gráfico Aleatório



- Geração de 50 pontos ao acaso entre 0 e 2:

```

x <- runif(50, 0, 2)
y <- runif(50, 0, 2)
    
```


- Gráfico dos 50 pontos com título, subtítulo, rótulos eixos x e y:

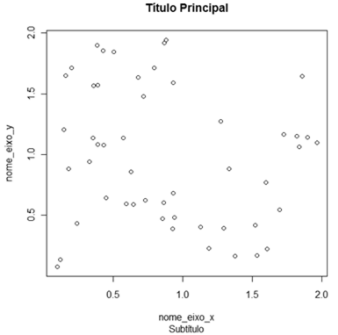
```

plot(x, y, main = "Título Principal", sub = "Subtítulo",
      xlab = "nome_eixo_x", ylab = "nome_eixo_y")
    
```

Introdução ao Uso do R - 2016 107

Gráfico Gerado






√ O gráfico de cada um será diferente
 – Se rodar de novo o resultado também será outro

Introdução ao Uso do R - 2016 108

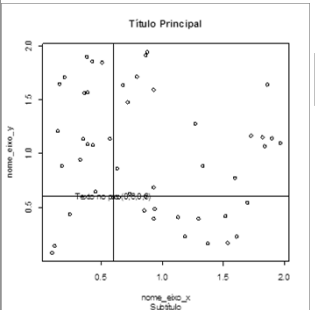
Adição de Dados



- Adicionando texto e linhas ao gráfico

```

text(0.6, 0.6, "texto no pto (0.6,0.6)")
# linhas pelo ponto (0.6, 0.6)
abline(h = 0.6, v = 0.6)
    
```



- `abline(a, b)` plota a reta $y=a+bx$

Introdução ao Uso do R - 2016 109

Coordenadas das Margens



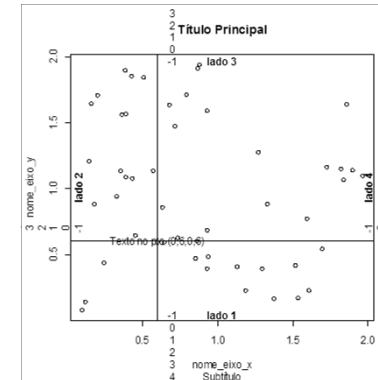
- Coordenadas das margens através função `mtext`

```
# coordenadas da margem
mtext(-1:4,side=1,at=0.7,line=-1:4)# coordenadas da margem
# loop para as coordenadas das margens
for(lado in 1:4) mtext(-1:4, side = lado, at = 0.7, line = -1:4)
# lado das margens
mtext(paste("lado", 1:4), side = 1:4, line = -1, font = 2)
```

Introdução ao Uso do R - 2016

110

Gráfico c/ Coordenadas Margens



- Layout de um gráfico padrão

Introdução ao Uso do R - 2016

111

Construindo um Gráfico por Partes



- Permite controle fino de cada elemento do gráfico

√ Desenha-se primeiro o gráfico sem os elemento

```
plot(x,y,type="n",xlab="",ylab="",axes=F) # plota-se nada!
```

√ Os elementos serão adicionados subsequentemente

Introdução ao Uso do R - 2016

112

Montagem do Gráfico



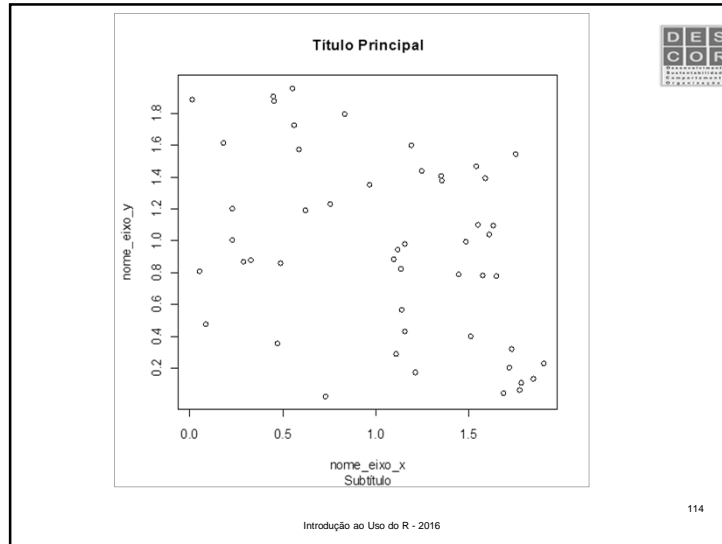
- O gráfico pode ser montado executando cada comando por vez

√ Verifique o que acontecerá

```
points(x, y) # plota os pontos do gráfico
axis(1) # plota o eixo x
axis(2, at = seq(0.2, 1.8, 0.2)) # plota o eixo y
box() # caixa do gráfico
# Título, sub-título, nomes dos eixos
title(main = "Título Principal", sub = "Subtítulo",
xlab = "nome_eixo_x", ylab = "nome_eixo_y")
```

Introdução ao Uso do R - 2016

113



Histogramas

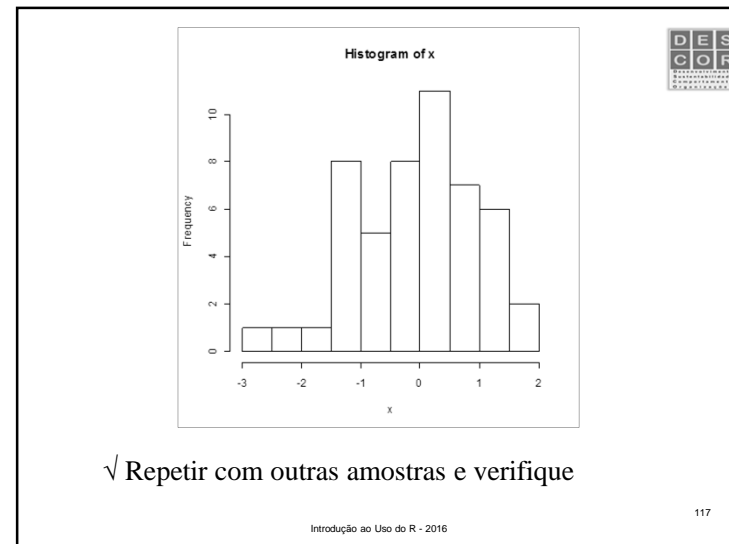
Construção de Histograma

- Geração de uma amostra aleatória com distribuição de frequências com simetria:
`x <- rnorm(50)`
- Construção do histograma (default)
`hist(x)`

DES
COR

Introdução ao Uso do R - 2016

116



Referências

Bibliografia Recomendada



- ALBERT, J.; RIZZO, M. *R by Example*. Springer, 2012.
- CHAPMAN, C.; FEIT, E. M. *R for marketing research and analytics*. Springer, 2015.
- DALGAARD, P. *Introductory statistics with R*. Springer, 2008.