

## Estatística Computacional I

Lupércio França Bessegato  
Dep. de Estatística/UFJF



### Roteiro Geral



1. Programando em R
2. Gráficos em R
3. Preparação e limpeza de dados
4. Manipulação de dados
5. Tópicos especiais
6. Referências

Estatística Computacional I - 2020

2

## Gráficos em R

## Argumentos Padrão



## Argumentos Padrão



- Há muitos argumentos que são padrões:
  - ✓ São aceitos pelas funções de alto nível.
- Alguns argumentos padrão:
  - ✓ `col`: controla cores
  - ✓ `lty`: controla tipo de linha
  - ✓ `font`: controla fonte e texto
- Esses argumentos podem não estar nomeados especificamente
  - ✓ São aceitos e estão indicados por ...

Estatística Computacional I - 2020

82



- Interpretação desses argumentos pode variar de acordo com a função:
  - ✓ Exemplo: argumento `col`:

- `plot()`: afeta cor dos dados
  - Não interfere na cor dos eixos e dos rótulos
- `barplot()`: cor do preenchimento ou padrão usado com as barras

Estatística Computacional I - 2020

83



- Há argumentos padrão para controlar a aparência de eixos e rótulos.
  - ✓ `xlim`: amplitude eixo x
  - ✓ `ylim`: amplitude eixo y
  - ✓ `main`: título
  - ✓ `xlab`: rótulo eixo x
  - ✓ `ylab`: rótulo eixo y

Estatística Computacional I - 2020

84



- ✓ Alguns argumentos comando `plot()`:

- `lwd`: largura da linha
- `col`: cor da linha
- `lty`: tipo da linha
- `ylim`: amplitude do eixo y

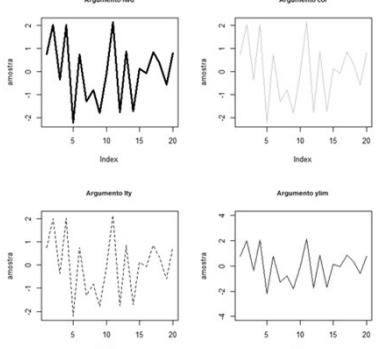
```
> # Argumentos em plot()
>
> # conjunto de dados
> # geração amostra
> set.seed(666)
> amostra <- rnorm(20)
> # Mudança de aspectos do gráfico
>
> par(mfrow = c(2, 2), cex.main = 0.8)
> # argumento lwd
> plot(amostra, type = "l", lwd = 3, main = "Argumento lwd")
> # argumento col
> plot(amostra, type = "l", col = "grey", main = "Argumento col")
> # argumento lty
> plot(amostra, type = "l", lty = "dashed", main = "Argumento lty")
> # argumento ylim
> plot(amostra, type = "l", ylim = c(-4, 4), main = "Argumento ylim")
```

Estatística Computacional I - 2020

85

✓ Mudança em aspectos de `plot()`:

- Resultado com argumentos diversos



86

Estatística Computacional I - 2020

## Plots de Múltiplas Variáveis

- Sistema gráfico tradicional tem várias funções para visualização de dados multidimensionais:
  - ✓ `persp()`: produz superfícies em 3D
  - ✓ `contour()` e `filled.contour()`: produz curvas de nível
  - ✓ `image()`: reticulado, com cor para representar 3<sup>a</sup> variável.
  - ✓ `symbol()`: símbolo representando 3<sup>a</sup> variável

87

Estatística Computacional I - 2020

• Plotando três variáveis:

✓ Funções: `persp()`, `contour()`, `image()` e `symbols()`.

```

> # gráficos com três variáveis
>
> help(volcano)
> par(mfrow=c(2, 2))
>
> ## Exagerando o relevo
> z <- 2 * volcano
> ## espaçamentos de 10 m (S to N)
> x <- 10 * (1:row(z))
> ## espaçamentos de 10 m (E to W)
> y <- 10 * (1:ncol(z))
>
> # n.1: gráfico em 3D
>
> par(mar = rep(0, 4))
> # comando persp() com argumento border = NA - não desenha grid
> persp(x, y, z, theta = 135, phi = 30, col = "light grey", scale = FALSE,
+       ltheta = -120, shade = 0.75, border = NA, box = FALSE)
> mtext("persp()", side = 3, line = -2, font = 2)
...

```

88

Estatística Computacional I - 2020

✓ Continuação do código:

```

> # n.2: curvas de nível
> par(mar = rep(0.5, 4))
> # curvas de nível de Volcano
> contour(x, y, z, asp = 1, labcex=0.35, axes=FALSE)
> rect(0, 0, 870, 620)
> mtext("contour()", side = 3, line = -1.5, font = 2)
>
> # n.3: gráfico de calor
> # gráfico de calor de Volcano
> image(x, y, z, asp = 1, col = grey(0.5 + 1:12/24), xlab = "",
+       ylab = "", axes = FALSE)
> rect(min(x) - 5, min(y) - 5, max(x) + 5, max(y) + 5)
> mtext("image()", side = 3, line = -1.5, font = 2)
>
> # n.4: gráfico de bolhas
> help(trees)
> dim(trees)
[1] 31 3
> attach(trees)
> par(mar = c(3, 3, 2, 0.5))
> # Girth é diâmetro, em polegadas
> symbols(Height, Volume, circles = Girth/24, inches = FALSE,
+          main="", xlab="", ylab="", bg = grey(Girth/max(Girth)))
> mtext("symbols()", side = 3, line = 0.5, font = 2)
> detach(trees)

```

89

Estatística Computacional I - 2020

✓ Gráficos com três variáveis:

✓ Compare as informações dos 3 primeiros

90

Estatística Computacional I - 2020

- Funções para gráficos de conjuntos de dados com mais de três variáveis:
  - ✓ pairs(): matriz de scatterplots
    - Plota cada variável contra todas as restantes
  - ✓ stars(): gráfico estrela, para variáveis contínuas
  - ✓ mosaicplot(): gráfico mosaico para variáveis categóricas

91

Estatística Computacional I - 2020

• Plotando dados multivariados:

✓ Funções:

- pairs(), stars() e mosaicplot().

```
> # gráficos com mais de três variáveis
> ## Multiplos plots usando argumento oma
> ## IMPORTANTE: não é a melhor para plotagem múltipla
>
> # n.1: matriz de scatterplots
> par(cex = 0.6)
> pairs(iris[1:4], oma = c(18, 4, 4, 4),
+       panel=function(x, y, ...){
+         points(x, y, lwd=0.1,
+         pch = ".") })
>
> # gráficos n.2 e n.3
> par(cex = 1)
> par(new = TRUE)
> par(omi = c(0, 0, 4.7, 0), mfrow = c(1, 2), mfg = c(1, 1), xpd = NA)
>
> # n.2: gráfico estrela
> par(mar = c(1, 1, 0, 1))
> palette(grey(0.5 + 1:8/24))
> stars(mtcars[1:8, 1:7], len = 0.8, cex = 0.5,
+       draw.segments = TRUE, xpd = NA)
...

```

92

Estatística Computacional I - 2020

– Continuação:

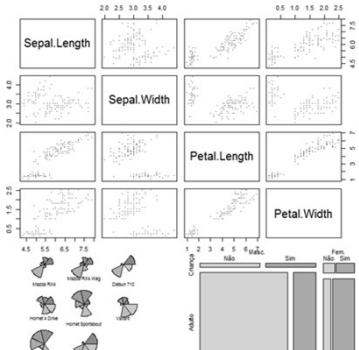
```
> # n.3: mosaic plot
>
> par(mar = c(0, 1, 1, 1))
>
> # Adicionando linhas vazias a variáveis do conjunto de dados
> # para tentar nomear rótulos em tamanho pequeno
> # Não é a melhor maneira
> niveis <- dimnames(Titanic)
> niveis$Sex <- c("Masc.\n\n", "Fem.\n\n")
> niveis$Survived <- c("Não\n", "Sim\n")
> niveis$Age <- c("Criança\n", "Adulto\n")
> dimnames(Titanic) <- niveis
> mosaicplot(~ Sex + Age + Survived, data = Titanic, off = rep(5, 3),
+            ylab = "", main = "",
+            color = c("light grey", "dark grey"))

```

93

Estatística Computacional I - 2020

✓ Plotando dados multivariados:



✓ Todos os gráficos foram efetuados no mesmo plot

94

Estatística Computacional I - 2020

## Plots Especializados

- Sistema gráfico tradicional (e pacotes associados) contém muitas funções para produzir gráficos que são apropriados para um particular tipo de dado ou técnica de análise

95

Estatística Computacional I - 2020

- Algumas funções:
  - ✓ `dotchart()`:
    - gráfico de pontos desenvolvido por Cleveland
  - ✓ `sunflowerplot()`:
    - Utilizada para plotar valores empurrados
  - ✓ `plot()`:
    - Produz diagramas para objeto `dendrogram`.

96

Estatística Computacional I - 2020

- Plotando com gráficos especializados:
  - ✓ Funções:
    - `dotchart()`, `sunflowerplot()` e `plot()`.

```
> # gráficos especializados (Fig. 2.8)
>
>
> par(mfrow = c(2, 2))
>
> # n.1: gráfico de pontos de Cleveland
>
> par(mar=c(7, 0, 3, 1), mex = 0.7)
> # dotplots
> dotchart(t(VADeaths[1:3,]), xlim = c(0, 40), cex = 0.6,
+           labels = c("RM", "RF", "UM", "UF"))
>
> # n.2: gráfico de pontos de Cleveland
>
> par(mar=c(5, 3, 2, 1), mex = 1)
> sunflowerplot(x = sort(round(rnorm(100))), y = round(rnorm(100),0),
+                xlab = "", ylab = "", xlim = c(-3, 3), ylim = c(-3, 3), cex =
0.5,
+                size = 1/12, seg.lwd = 1, seg.col = "grey", axes = FALSE)
> axis(1, at = seq(-3, 3, 3))
> axis(2, at = seq(-3, 3, 3))
...

```

97

Estatística Computacional I - 2020

- Continuação:

```

> # agrupamento hierárquico
>
> hc <- hclust(dist(USArrests), "ave")
> dend1 <- as.dendrogram(hc)
> dend2 <- cut(dend1, h = 70)
>
> # n.3: variação de dendrograma
>
> par(mar = c(1, 0, 2, 5.5), cex = 0.7)
> # dend2$lower não é um dendrograma mas uma lista objetos. :
> plot(dend2$lower[[3]], horiz = TRUE, type = "tr", axes=FALSE, cex=0.8)
>
> # n.4: dendrograma
>
> par(mar = c(6, 0, 2, 0))
> # bordas em diferentes tipos e cores:
> plot(dend2$lower[[2]], edgePar = list(col = c("black", "grey")),
+ edge.root = TRUE, axes = FALSE, cex = 0.8)

```

98

Estatística Computacional I - 2020

✓ Plots especializados:

✓ Visualizações de dados especiais e de técnicas de análise

99

Estatística Computacional I - 2020

- Outras funções especializadas:
  - ✓ Coplot()
  - ✓ Hexbin(){hexbin}:
  - ✓ Assocplot():
- E muitas outras mais!!!

100

Estatística Computacional I - 2020

## Gráficos Interativos

- Sistema gráfico tradicional produz de gráficos estáticos
  - ✓ Interação limitada com saída gráfica:
    - locator():
      - Usuário clica no plot e obtém coordenadas do ponto
    - identify():
      - Pode ser usada para adicionar rótulos nos dados

101

Estatística Computacional I - 2020



- Vários pacotes fornecem recursos interativos:
  - ✓ `tcltk`:
    - Funcionalidades para construir componentes GUI (*Graphical User Interfaces*)
  - ✓ `Rggobi`:
    - Conectam o R com outros softwares de aplicação gráfica
  - ✓ Outros:
    - `gWidget`.



Estatística Computacional I - 2020

102

## Referências



## Bibliografia Recomendada



- ALBERT, J.; RIZZO, M. *R by Example*. Springer, 2012.
- CHRISTIAN, N. *Basic Programming*, Lecture Notes
- DALGAARD, P. *Introductory statistics with R*. Springer, 2008.
- MURRELL, P. *R Graphics*. Chapman & Hall, 2006.

Estatística Computacional I - 2020

245