

Programação Estatística

Introdução ao R - Visualização de dados

Rachid Muleia, PhD in Statistics

Universidade Eduardo Mondlane
Faculdade de Ciências
Departamento de Matemática e Informática

2023-04-13

Gráficos suportados pelo R

- Gráfico circular
- Gráfico de barras
- Histograma
- Box plot
- Gráfico de linhas
- Diagrama de dispersão

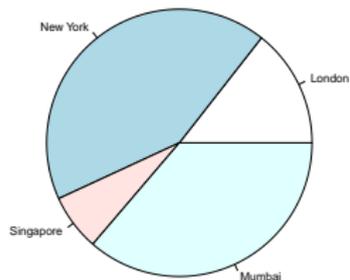
Gráfico circular

- O R dispõe de diversas livrarias para criação de gráficos
- O Diagrama ou gráfico circular representação gráfica para dados qualitativos
- O gráfico está dividido em sectores, onde cada sector representa uma fracção do total
- No R, o gráfico de circular pode ser criado usando a função `pie()`

Diagrama circular-Syntax

- Syntax para a criação de um diagrama circular - `pie(x, labels, radius, main, col, clockwise)`

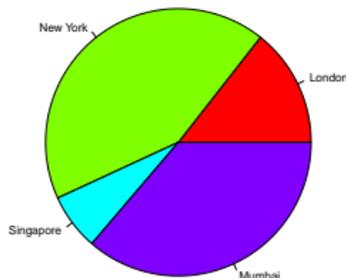
Exemplo



Digrama circular- alterando as cores

- a função `pie` permite alterar as cores do diagrama circular usando o argumento `col`

```
> x <- c(21, 62, 10, 53)
>
> labels <- c("London", "New York", "Singapore", "Mumbai")
>
> pie(x, labels, col = rainbow(length(x)) )
```



Digrama circular- legenda

■ No diagrama circular pode-se manipular a legenda

```
> x <- c(21, 62, 10, 53)
>
> labels <- c("London", "New York", "Singapore", "Mumbai")
>
> piepercent <- round(100*x/sum(x), 1) # percentagem para cada categoria
>
>
> pie(x, labels = piepercent, col =
+ rainbow(length(x)))
>
> legend("topright", c("Pune", "Nashik", "Aurangabad", "Mumbai"),
+ cex = 0.8, fill = rainbow(length(x)))
```

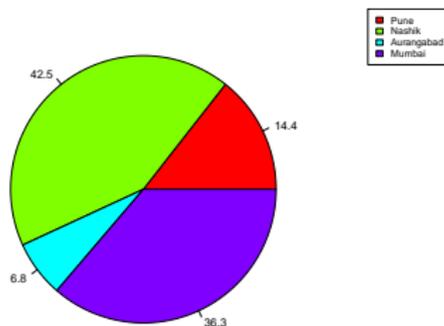


Gráfico de barras

- Um gráfico de barras representa dados em barras rectangulares com comprimento da barra proporcional ao valor da variável
- R usa a função `barplot()` para criar um gráfico de barras
- As barras podem ser verticais ou horizontais
- O R permite que cada barra apresente uma coloração

Gráfico de barras- Syntax

- `barplot(H, xlab, ylab, names.arg, col)`
 - `H` é um vector ou uma matriz de numéricos a ser usada para construir o gráfico de barras
 - `xlab` descrição do eixo das abcissas
 - `ylab` descrição do eixo das ordenadas
 - `names.arg` categorias a visualizar em cada
 - `col` recebe um vector de cores a aplicar nas barras

Gráfico de barras- exemplo

```
> # criar um gráfico de barras  
> H <- c(7,12,28,3,41)  
>  
> barplot(H)
```

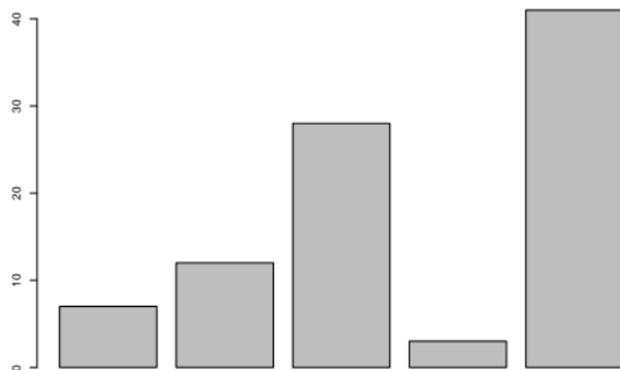


Gráfico de barras- atributos nas barras

```
> # criar um gráfico de barras
> H <- c(7,12,28,3,41)
> M <- c("Mar","Apr","May","Jun","Jul")
> barplot(H,names.arg = M,xlab = "Month",ylab = "Revenue",
+         col = "blue", main = "Revenue chart",border= "red")
```



Gráfico de barras- barras lado a lado ou sobrepostas

```
> data("mtcars")
> freq <- table(mtcars$am,mtcars$gear)
> freq

      3  4  5
0 15  4  0
1  0  8  5
> barplot(freq,beside = TRUE, col=c('blue','red'),legend.text = c('Automático', 'Manual'))
```

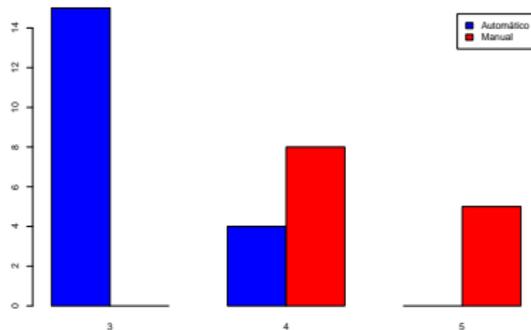
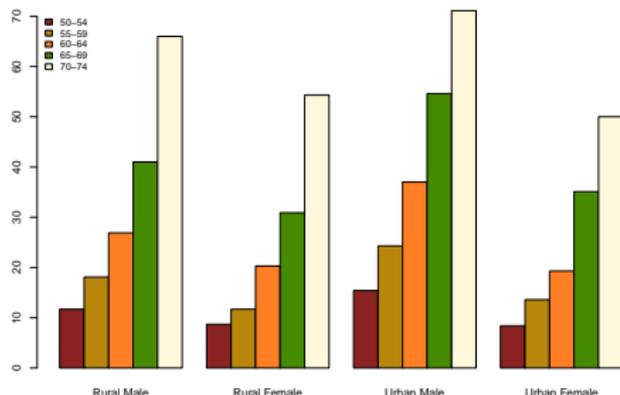


Gráfico de barras- barras lado a lado ou sobrepostas

```
> data("VADeaths")
> cores <- c("brown4", "darkgoldenrod", "chocolate1", "chartreuse4", "cornsilk") # definir as cores
> barplot(VADeaths, col = cores, beside = TRUE) # gráfico de barras
> # Acrescentar legenda
> legend("topleft", legend = rownames(VADeaths),
+       fill = cores, box.lty = 0, cex = 0.8) # Acrescentar legenda
```

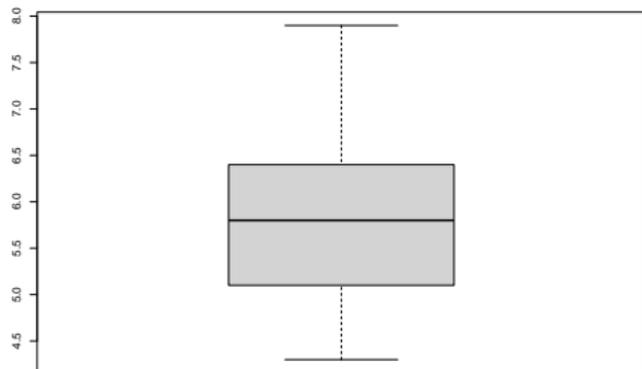


Box-plot

- Box-plot também é um gráfico concebido para variáveis contínuas.
- Permite conhecer a distribuição dos dados
- Identificar observações atípicas.
- Tem a vantagem de permitir comparações.
- Pode-se construir usando a função `boxplot()`

Box-plot

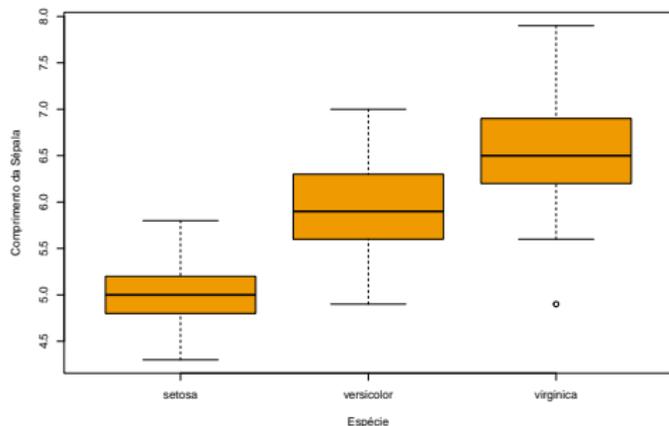
```
> boxplot(iris$Sepal.Length)
```



Box-plot

Vamos comparar o comprimento das sépalas entre as espécies.

```
> boxplot(iris$Sepal.Length~iris$Species,  
+         xlab='Espécie', ylab='Comprimento da Sépala',col='orange2')
```

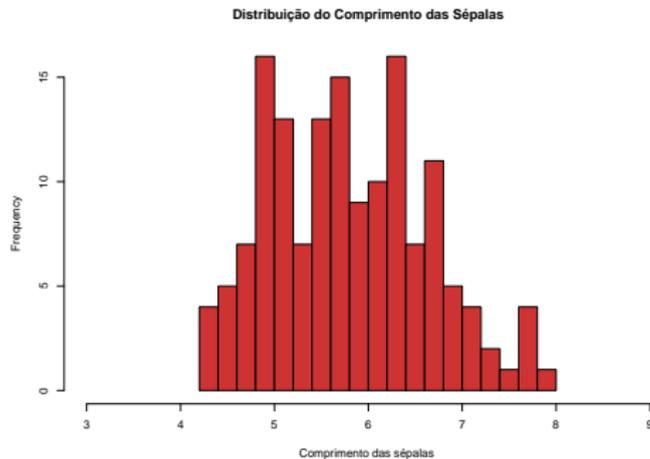


Histograma

- O histograma representa as frequências de valores de uma variável agrupados em intervalos.
- Este é um gráfico para variáveis de natureza contínua (numéricas).
- Serve para verificar a distribuição dos dados (Simétricos/Assimétricos)
- Permite identificar observações atípicas (outliers)
- Pode-se construir usando a função `hist()`.

Histograma

```
> hist(iris$Sepal.Length, col = 'brown3', breaks = 20, xlim = c(3, 9),  
+      xlab = 'Comprimento das sépalas ', main = 'Distribuição do Comprimento das Sépalas')
```



Histograma + densidade

Para adicionar uma curva densidade ao histograma iremos usar a função `density()` e a função `lines()`

```
> hist(iris$Sepal.Length, col='grey',breaks = 20,xlim = c(3, 9),  
+      xlab = 'Comprimento das sépalas ', main='Distribuição do Comprimento das Sépalas',prob=TRUE)  
> lines(density(iris$Sepal.Length),col='red')
```

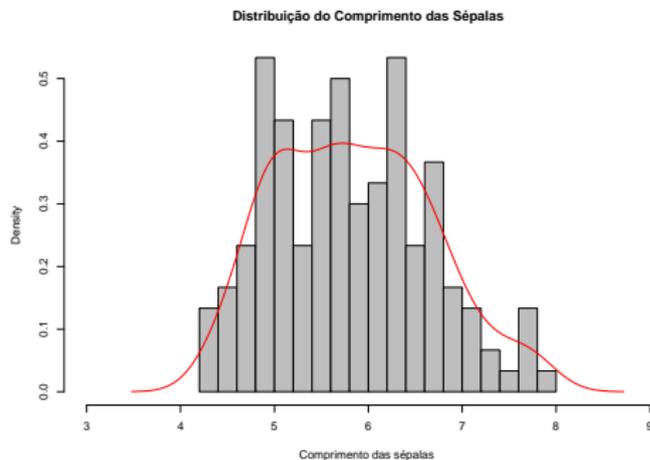


Diagrama de dispersão

- O diagrama de dispersão serve para mostrar a relação entre duas variáveis numéricas.
- Para a sua concepção vamos usar a função `plot()`.

```
> plot(iris$Sepal.Length,iris$Petal.Width)
```

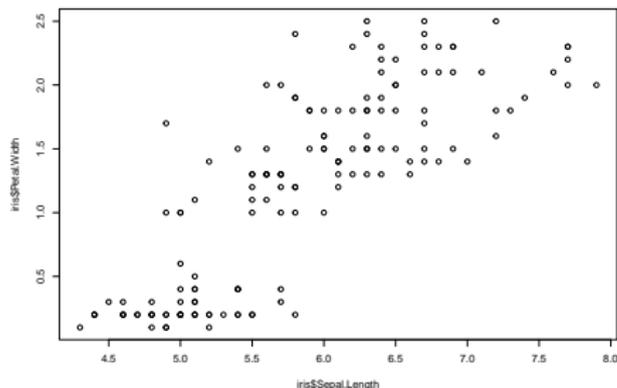


Diagrama de dispersão

A relação entre o comprimento das sépalas e a largura das pétalas pode ser analisada para as três espécies.

```
> plot(iris$Sepal.Length,iris$Petal.Width, col=iris$Species,pch = 16)  
> legend('topleft', legend =c('setosa','versicolor','virginica'),  
+       col =c('black','indianred1','green'),pch=16)
```

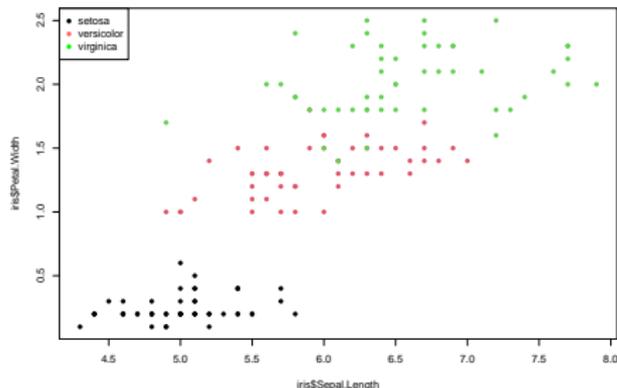


Gráfico de linhas

- Um gráfico de linhas é um gráfico que conecta uma série de pontos desenhando segmentos de linha entre eles.
- Gráficos de linha são geralmente usados para identificar o tendências em dados.
- A função `plot()` em R é usada para criar a linha gráfico.

Gráfico de linhas

A sintaxe básica para criar um gráfico de linha em R é -
`plot(v, tipo, col, xlab, ylab)`

A seguir está a descrição dos parâmetros usados -
- `v` é um vector contendo os valores numéricos.
- `tipo` assume o valor "p" para desenhar apenas os pontos, "l" para desenhar apenas as linhas e "o" para desenhar os dois pontos e linhas.
- `xlab` é o rótulo do eixo x.
- `ylab` é o rótulo para o eixo y.
- `main` é o título do gráfico.
- `col` é usado para dar cores aos pontos e linhas.

Gráfico de linhas

```
> # Criar grafico de linha  
> v <- c(7,12,28,3,41)  
>  
> plot(v,type = "o")
```

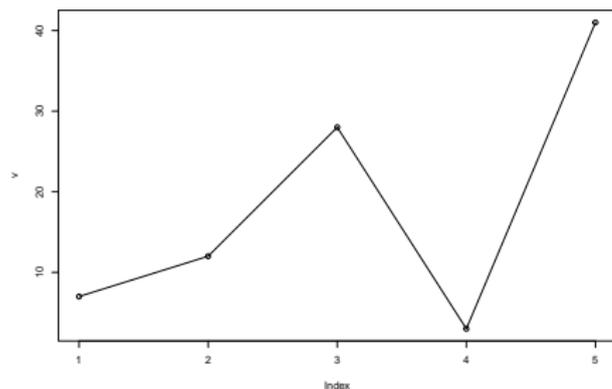


Gráfico de linhas (mais de uma linha)

```
> # Criar grafico de linha  
>  
> v <- c(7,12,28,3,41)  
> t <- c(14,7,6,19,3)  
> plot(v,type = "o",col = "red", xlab = "Month", ylab = "Rain  
+ fall", main = "Rain fall chart")  
> lines(t, type = "o", col = "blue")
```

