

GRÁFICOS



Gráficos

São representações visuais dos dados (ou de sua distribuição de frequências), que permitem uma rápida interpretação.

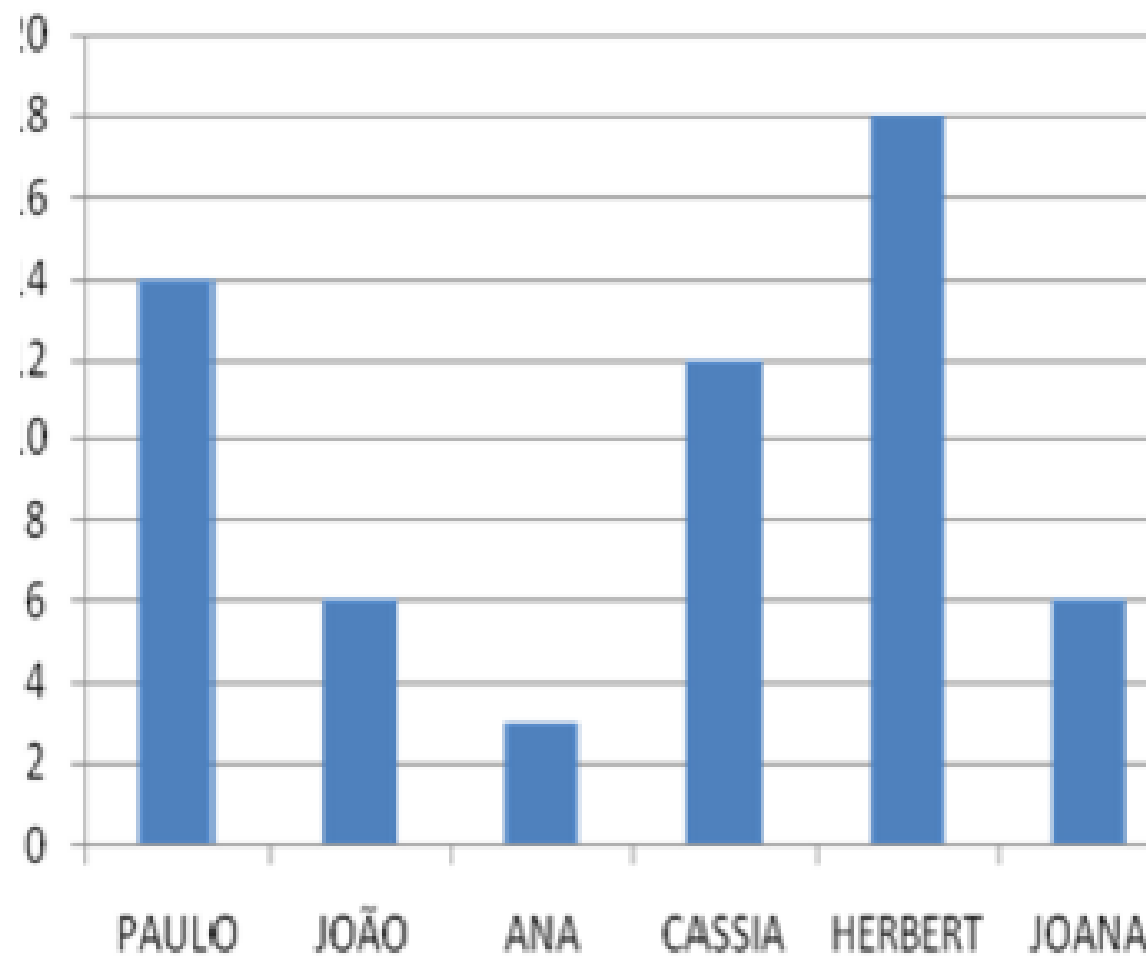
- **PARA VARIÁVEIS QUALITATIVAS (categóricas): Setor e Barras/Colunas**
 - **PARA VARIÁVEIS QUANTITATIVAS: Histograma, boxplot, dispersão, linha (além do de Setor e de Barras/Colunas também)**
-

Gráfico de barras/colunas

Para variáveis quantitativas discretas e qualitativas.

Nesse tipo de gráfico, o eixo vertical representa a frequência absoluta/percentual/relativa, e o eixo horizontal representa as classes da variável. Sendo assim, cada coluna representa a frequência de uma determinada classe.

Vendas de Animais



Escola de Ensino Fundamental	Matrículas de professores	Matrícula de alunos
Privada	1751	25280
Pública Estadual	1186	21328
Pública Municipal	947	18432
Pública Federal	29	280

Fonte: www.ibge.com.br

- Exemplo:

```
prof<-c(1751,1186,947,29)
```

Vamos nomear cada barra:

```
names(prof)<-c("privada","estadual","municipal","federal")
```

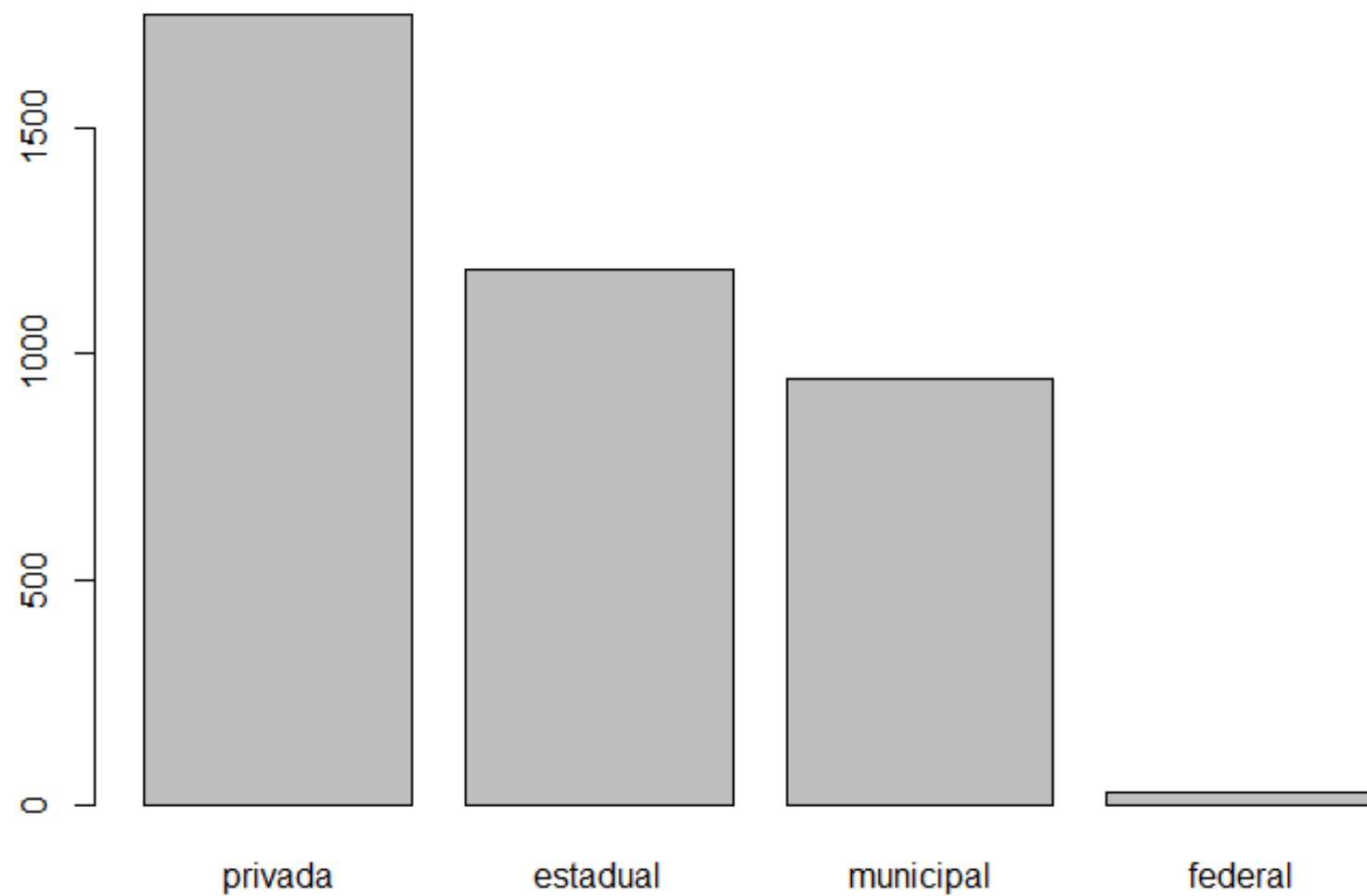
```
barplot(prof) #Fazendo o gráfico
```

Colocando título no gráfico:

```
barplot(prof, main="Distribuição de professores do ensino  
fundamental em Niterói, 2009")
```

Gráfico de
barras/colunas

Distribuição de professores do ensino fundamental em Niterói, 2009



Exemplo

Gráfico de barras/colunas

- Parâmetros que são compartilhados pela maioria das funções que fazem gráficos:

main: título do gráfico

xlab: rótulo do eixo x

ylab: rótulo do eixo y

xlim: limite inicial e final do eixo x, deve ser apresentado como vetor com o valor mínimo e máximo, *exemplo* `c(0,10)`

ylim: limite inicial e final do eixo y, deve ser apresentado como vetor com o valor mínimo e máximo, *exemplo:* `c(-10,-1)`

col: cores que devem ser usadas, pode ser um único valor tipo *character* ou um vetor com o nome das cores, *exemplo:* `c("red", "blue", "gray")`

Gráfico de barras/colunas

Alterações no gráfico:

- **mudar o tamanho da fonte do título, acrescente o argumento `cex.main=valor`**

barplot(prof, main="Professores de escolas de ensino fundamental em 2009 no município de Niterói", cex.main=1.2)

- **mudar o tamanho dos números do eixo y utilize o argumento `cex.axis=valor`.**

barplot(prof, cex.axis=0.9)

- **Colocar nome nos eixos x e y:**

barplot(prof, main="Distribuição de professores do ensino fundamental em Niterói, 2009", ylab="matrículas", xlab="escolas")

- **mudar o tamanho da fonte dos nomes dos eixos, utilize o argumento `cex.lab=valor`.**

barplot(prof, main="Distribuição de professores do ensino fundamental em Niterói, 2009", ylab="matrículas", xlab="escolas", cex.lab=0.9)

Gráfico de barras/colunas

Alterações no gráfico:

- **É possível incluir subtítulo no gráfico incluindo o argumento sub = “texto”.**

```
barplot(prof, names.arg=names(prof), main="Distribuição de professores do ensino fundamental em Niterói,  
2009", ylab="matrículas", xlab="escolas", sub="fonte:www.ibge.com.br")
```

- **Para preencher as barras com cores uniformes diferentes do padrão estabelecido, usamos o comando col = “nome da cor” ou col = valor, este valor representa o número da cor.**

```
barplot(prof,col=c("blue","red","green","orange"))
```

```
barplot(prof,col=c(1,2,3,4))
```

- **Podemos excluir a linha das bordas das barras dessa forma:**

```
barplot(prof,border=NA)
```

- **Colocar as colunas na horizontal:**

```
barplot(prof, horiz=T)
```


Gráfico de barras/colunas

- Também podemos criar o gráfico de barras de duas variáveis, um ao lado do outro, na mesma janela gráfica. Para isso, vamos imaginar que os dados estejam armazenados em um objeto do tipo matriz, conforme comando a seguir. Estes dados são os do exemplo:

```
alunosprof<- matrix(c(1751,1186,947,29,25280,21328,18432,280), nrow=4, ncol=2)
```

- Para colocarmos nome nas linhas e colunas da matriz, utilizamos o argumento `dim.names`:

```
alunosprof<- matrix(c(1751,1186,947,29,25280,21328,18432,280), nrow=4, ncol=2, dimnames=list(c("privada",  
"estadual", "municipal", "federal"), c("professores", "alunos")))
```

- Agora usamos o comando `barplot` para fazer o gráfico:

```
barplot(alunosprof)
```

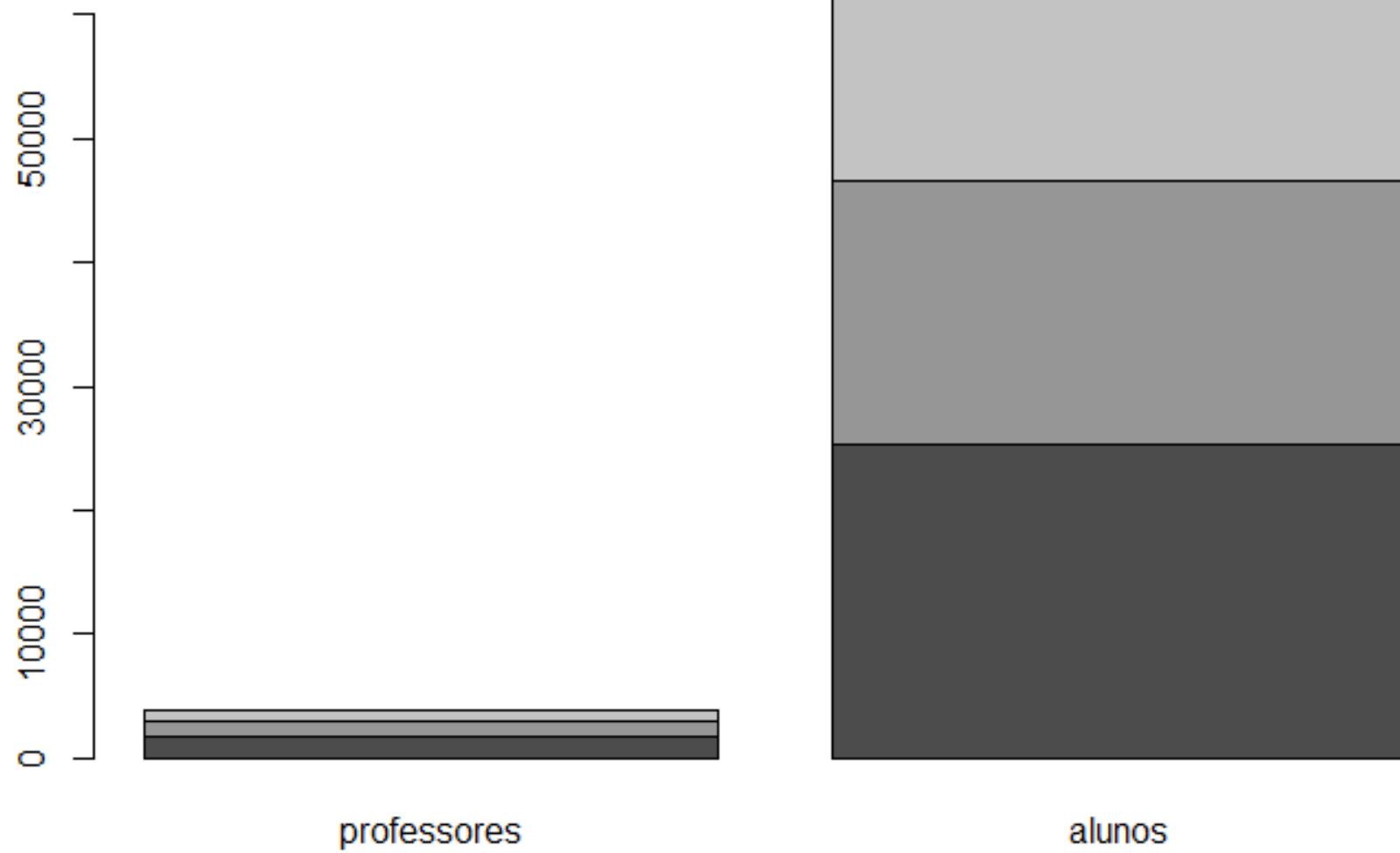


Gráfico de barras/colunas

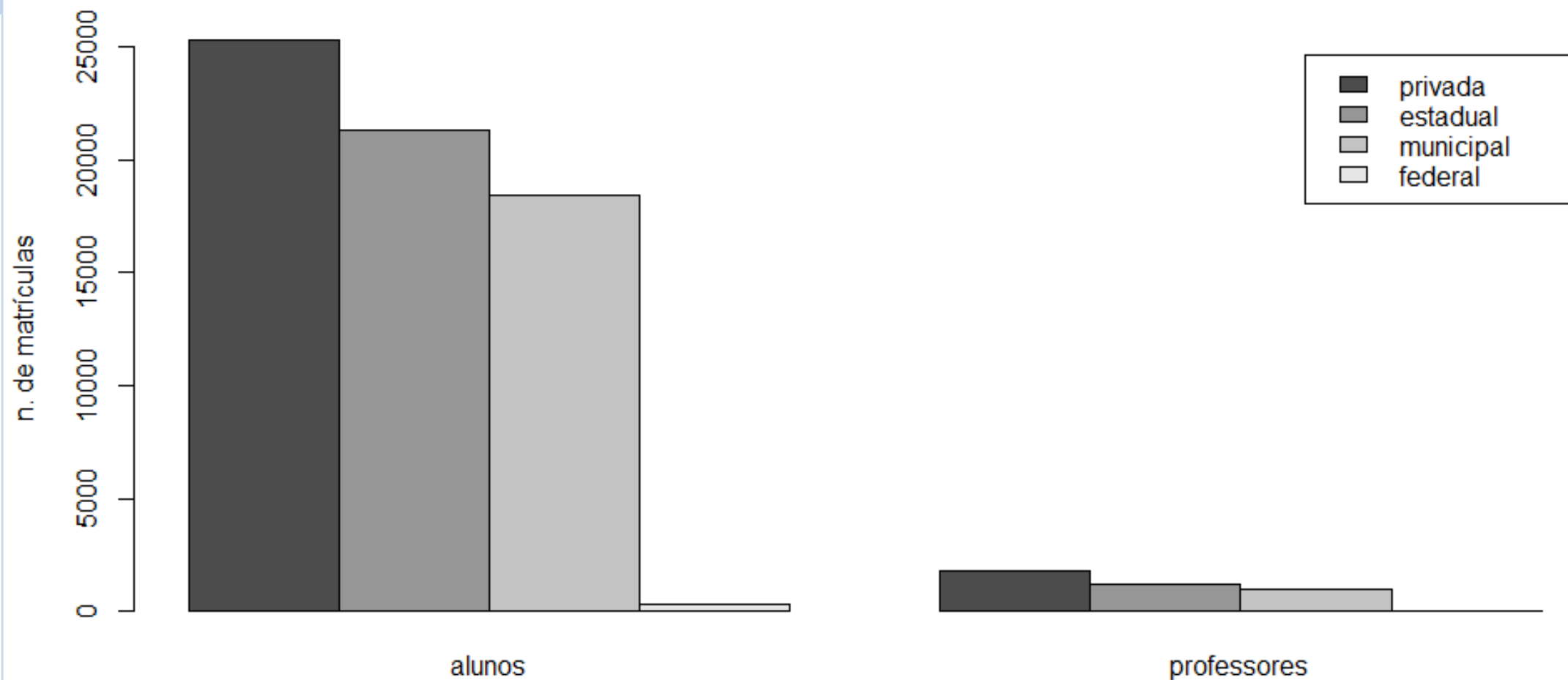
- O default desse comando é colocar todas as categorias uma em cima da outra, acumulando os resultados. Mas podemos usar o seguinte argumento para obtermos **barras lado a lado**:

```
barplot(alunosprof, beside=TRUE)
```

- Para finalizarmos a apresentação do gráfico, adicionaremos **legenda, título, nomes nos eixos e subtítulo citando a fonte, inverteremos a ordem dos dados**, colocando primeiro a sequência de alunos e depois a de professores resultando numa disposição mais apresentável do ponto de vista estético.

```
barplot(alunosprof[,2:1],beside=TRUE,legend.text=rownames(alunosprof), main="Distribuição de matrícula de  
alunos e professores do ensino fundamental em Niterói, 2009",ylab="n. de  
matrículas",xlab="matrícula",sub="fonte:www.ibge.com.br")
```

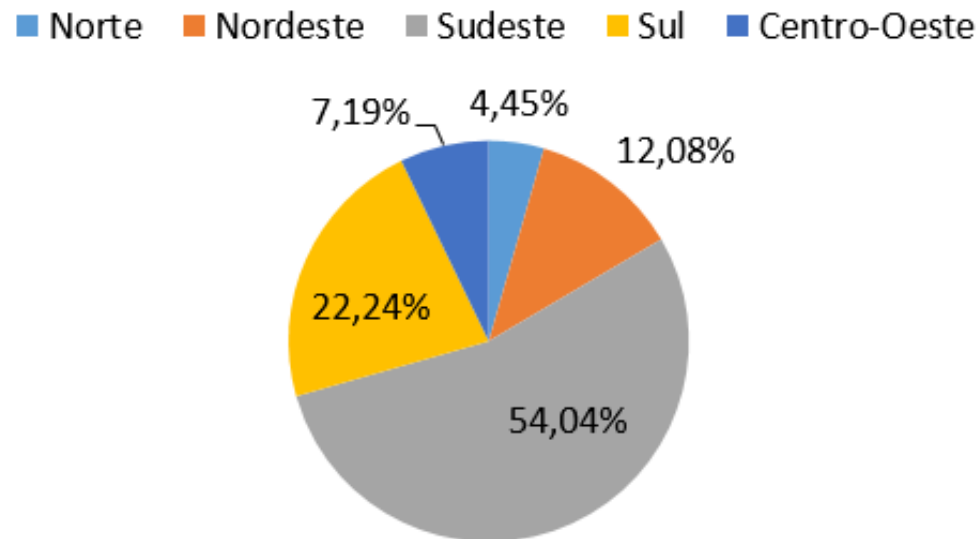
Distribuição de matrícula de alunos e professores do ensino fundamental em Niterói, 2009



matrícula
fonte: www.ibge.com.br

Gráfico de setores (pizza)

Figura - Acidentes de trabalho por regiões no Brasil.



Fonte: AEPS (2015) – Adaptado

- Para variáveis quantitativas discretas e qualitativas.
- Mais conhecido como gráfico de pizza, ele é formado por um círculo de raio qualquer que representa o todo, e divide-se esse círculo em setores (fatias), um para cada categoria, que serão proporcionais à frequência daquela categoria.
- A circunferência é dividida em setores, com ângulos centrais proporcionais às frequências (ocorrências) das classes, por exemplo : 100% corresponde à 360°, assim como 50% à 180°.
- Indicado quando se tem poucas classes.

Gráfico de setores (pizza)

```
amostra<-read.csv("amostra.csv", header=TRUE, sep=";")
```

```
pie(table(amostra$EstCivil))
```

```
pie(prop.table(table(amostra$EstCivil)))
```

Para colocar as porcentagens utilizamos o argumento **labels**:

```
rotulos<-paste(round((prop.table(table(amostra$EstCivil))*100),2), "%")
```

```
pie(prop.table(table(amostra$EstCivil))*100, main="Estado Civil", labels=rotulos, col=c("pink",  
"blue", "purple"))
```

Gráfico de setores (pizza)

E para acrescentar legenda:

```
legend("topright", pch=15, col=c("pink","purple","blue"), legend=c("casado","divorciado","separado"))
```

- A função *legend()* recebe como argumentos “x=” e “y=”, que irão informar as coordenadas para que a legenda seja plotada. Contudo, “x=” e “y=” podem ser substituídos pelas palavras: “topright”, “topleft”, “bottomright” e “bottomleft”. Essas palavras irão determinar a posição da legenda.
- Além disso, devemos informar as cores e qual *character* desejamos utilizar para representar as cores na legenda. Essas escolhas são feitas por meio dos argumentos já conhecidos: “col=” e “pch=”, respectivamente.
- Por fim, devemos informar o texto que será adicionado para cada *character* da legenda. Isso é feito por meio do uso do argumento “legend=”.

Estado Civil

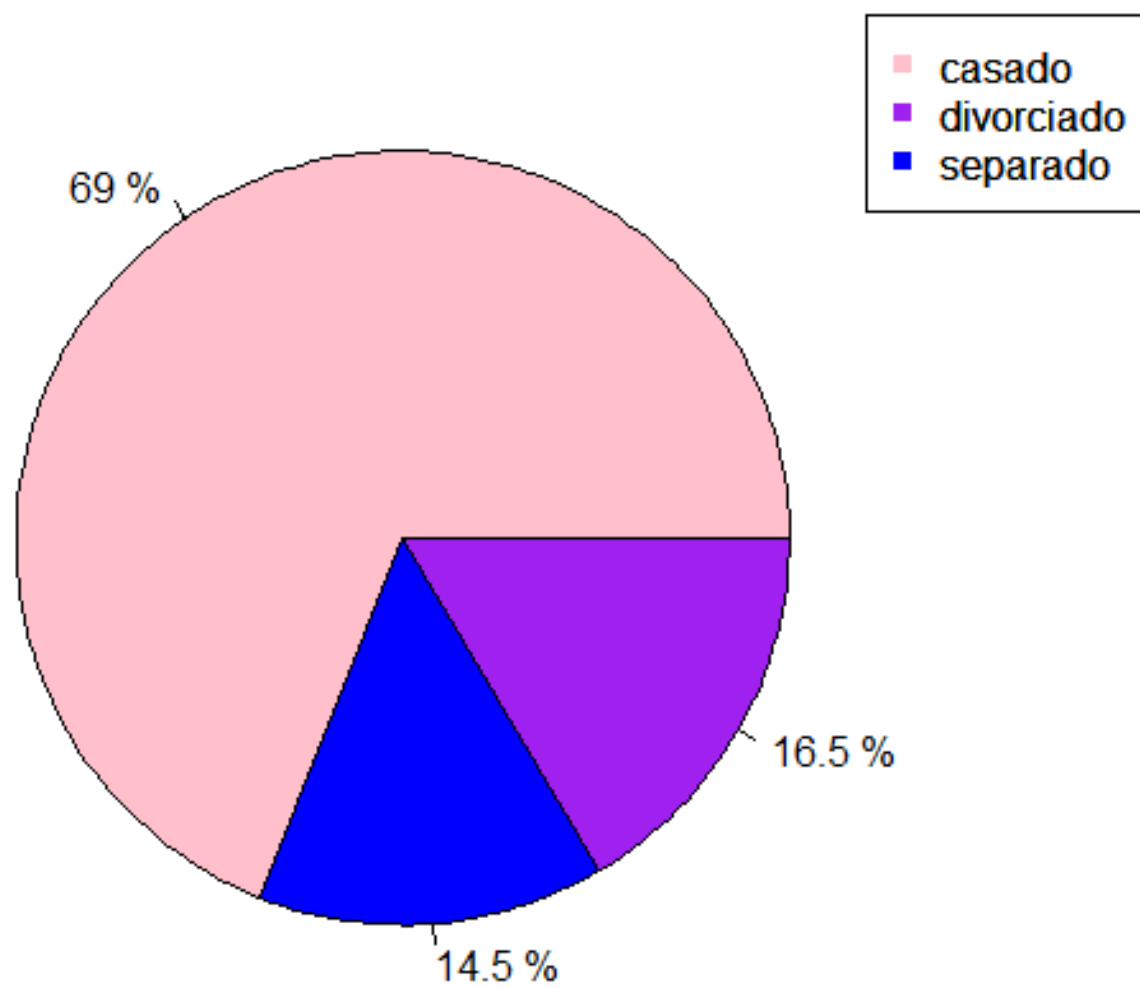


Gráfico de linhas

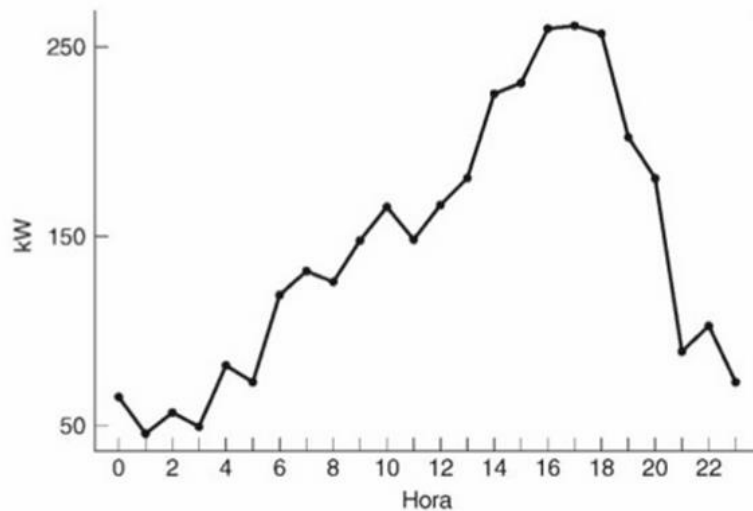


Figura 8-12 Dados da demanda horária de quilowatt (kW) em um dia de pico de verão para um edifício de escritórios.

Geralmente utilizado para representar uma série temporal (variável discreta).

- É composto por dois eixos, um vertical e outro horizontal, e uma linha que mostra a evolução de um fenômeno ou o processo no decorrer de um período. Os pontos indicam a frequência da classe correspondente e eles são interligados entre si.
- Usualmente, com gráficos temporais o tempo é representado no eixo horizontal, sendo o eixo vertical calibrado para acomodar a amplitude dos valores representados nos resultados observacionais.

Gráfico de linhas

Vamos considerar o seguinte conjunto de dados:

```
mes<-1:12
```

```
temperatura<-c(-5.44,-5.17,0.11,6.89,12.67,17.94,20.44,19.5,15.67,9.72,4.06,-2.56)
```

```
plot(mes,temperatura,type='b')
```

type = "p" para pontos,

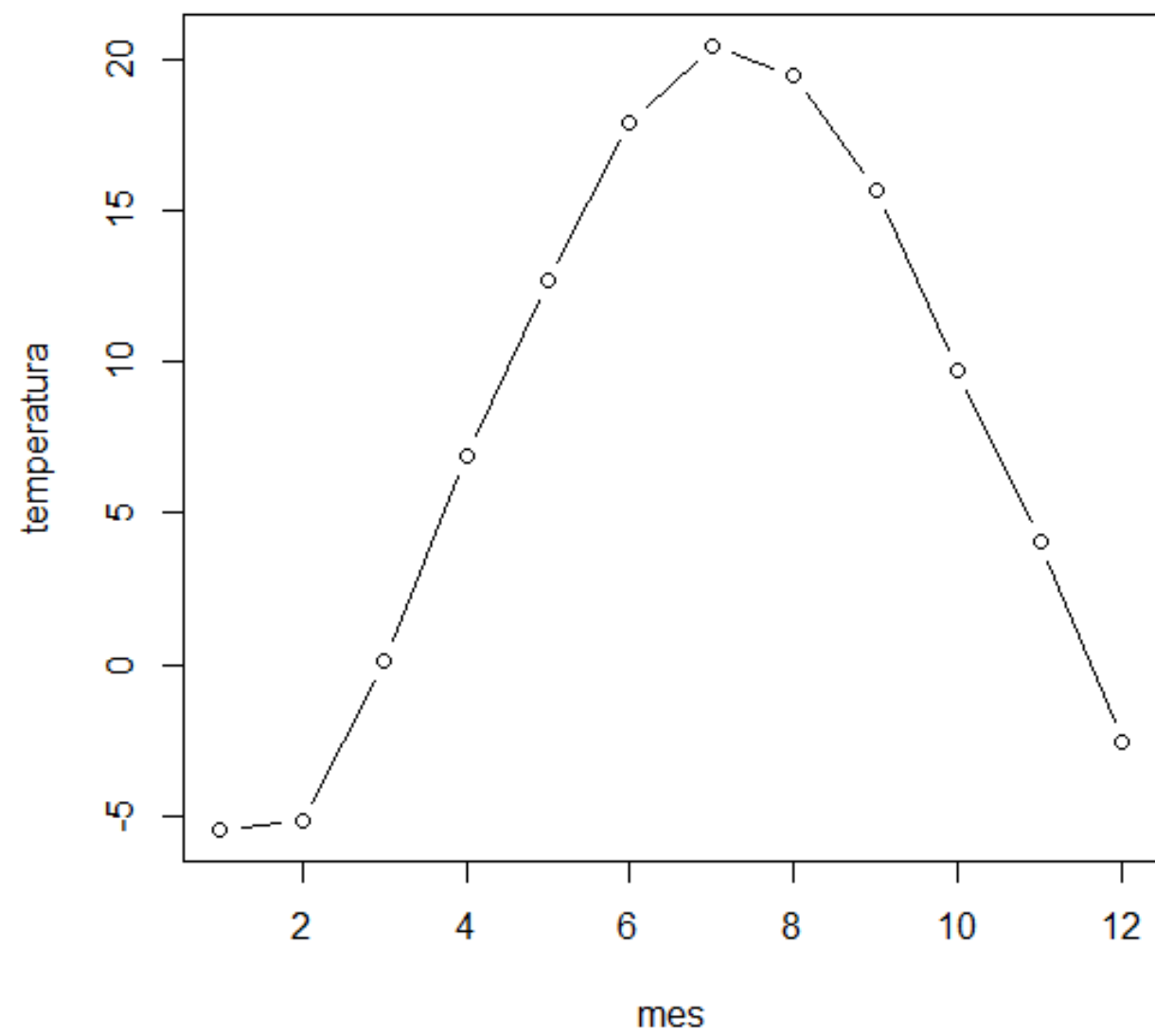
"l" para linhas,

"b" para pontos e linhas,

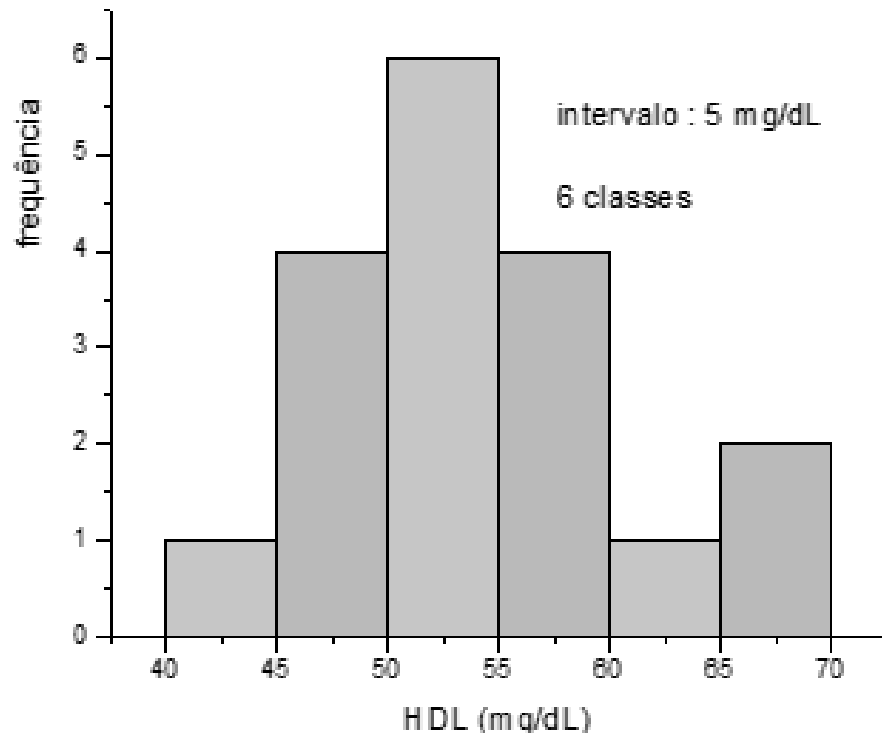
"c" para linhas descontínuas nos pontos,

"o" para pontos sobre as linhas,

"n" para nenhum gráfico, apenas a janela.



Histograma

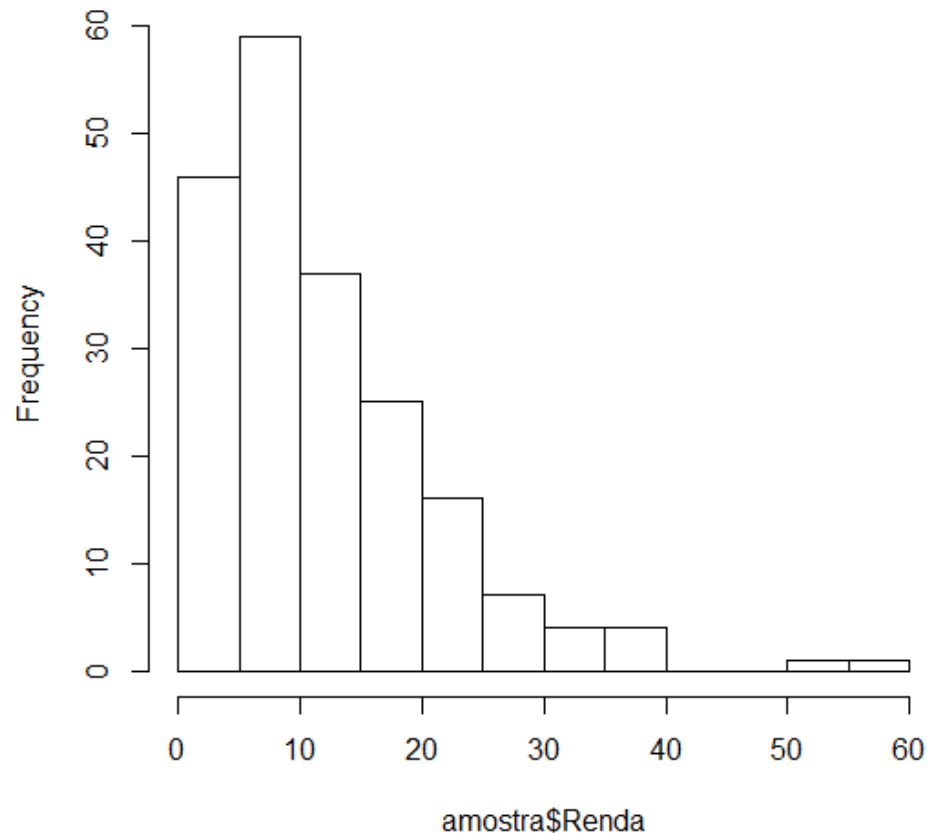


Utilizado para representar a distribuição de uma variável aleatória contínua.

O histograma é um tipo de gráfico de colunas justapostas (unidas), no qual o eixo horizontal representa uma variável quantitativa contínua, e o eixo vertical simboliza a frequência (absoluta, relativa ou percentual) do intervalo correspondente.

Histograma

Histogram of amostra\$Renda



`hist(amostra$Renda)`

- O número de classes pode ser alterado com o comando *breaks*.

OBS.: Nesse caso são aceitos um simples número indicando quantas classes devem ser formadas. Internamente, há uma função de tolerância que pode não responder ao valor informado. Dependendo da configuração dos dados, o R, automaticamente, faz um ajuste diferente do solicitado.

Histograma

- As classes podem ser explicitamente definidas pelo usuário de várias maneiras:

```
hist(amostra$Renda, br=c(0,15,30,45,60, max(amostra$Renda)))
```

```
hist(amostra$Renda, br=seq(0,60,len=8))
```

- O número de classes é importante na construção do histograma, pois isso afeta a forma do gráfico e por sua vez, a maneira como os dados ou informações serão interpretadas. O número de classes, k , pode-se ser definido pelo método de Sturges, através da seguinte expressão:

$$k = 1 + 3,3 \cdot \log(n)$$

Histograma

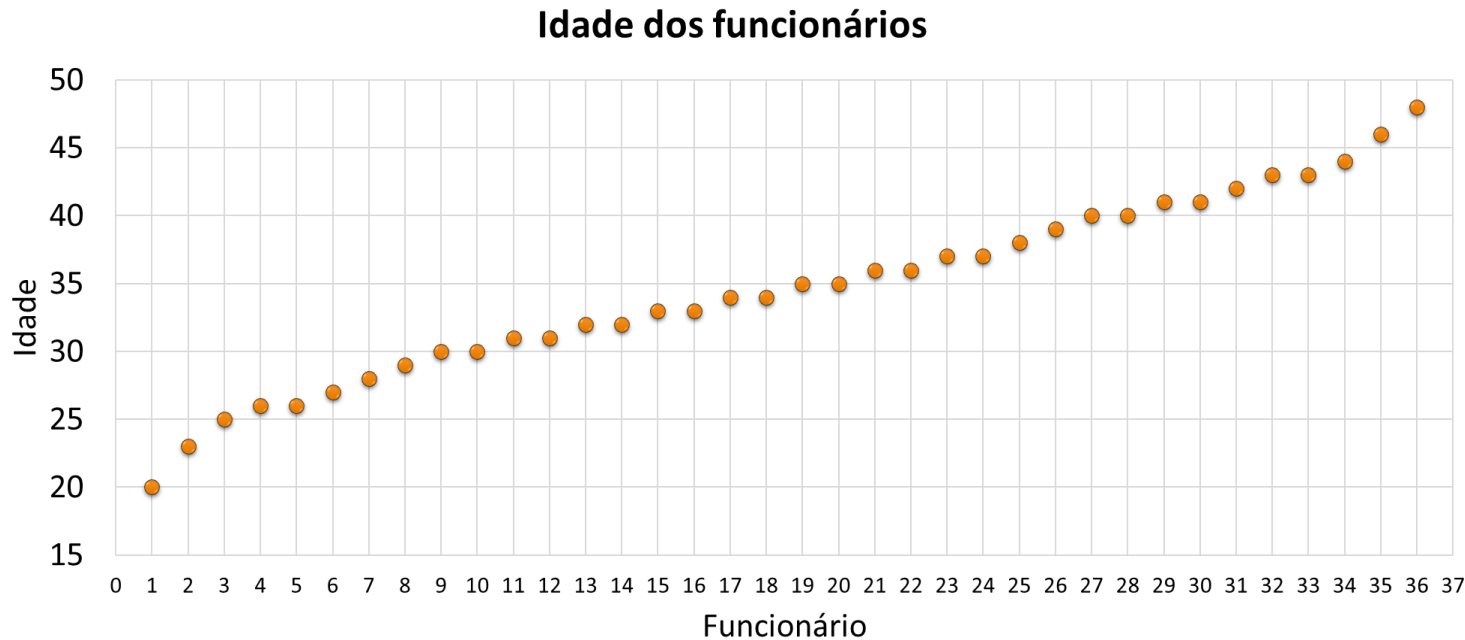
- Podemos colocar nosso gráfico em frequência relativa:

hist(amostra\$Renda, freq=FALSE)

- Colocar linhas de sombreamento:

hist(amostra\$Renda, density=30)

Gráfico de dispersão ou Scatter plots



- Gráficos que mostram a dispersão de dados são úteis para identificar muitas características. Além da dispersão, outliers, tendências entre outros aspectos, podem ser explorados com gráficos de dispersão ou Scatter plots.

Gráfico de dispersão ou Scatter plots

1. Univariado

```
plot(amostra$Renda)
```

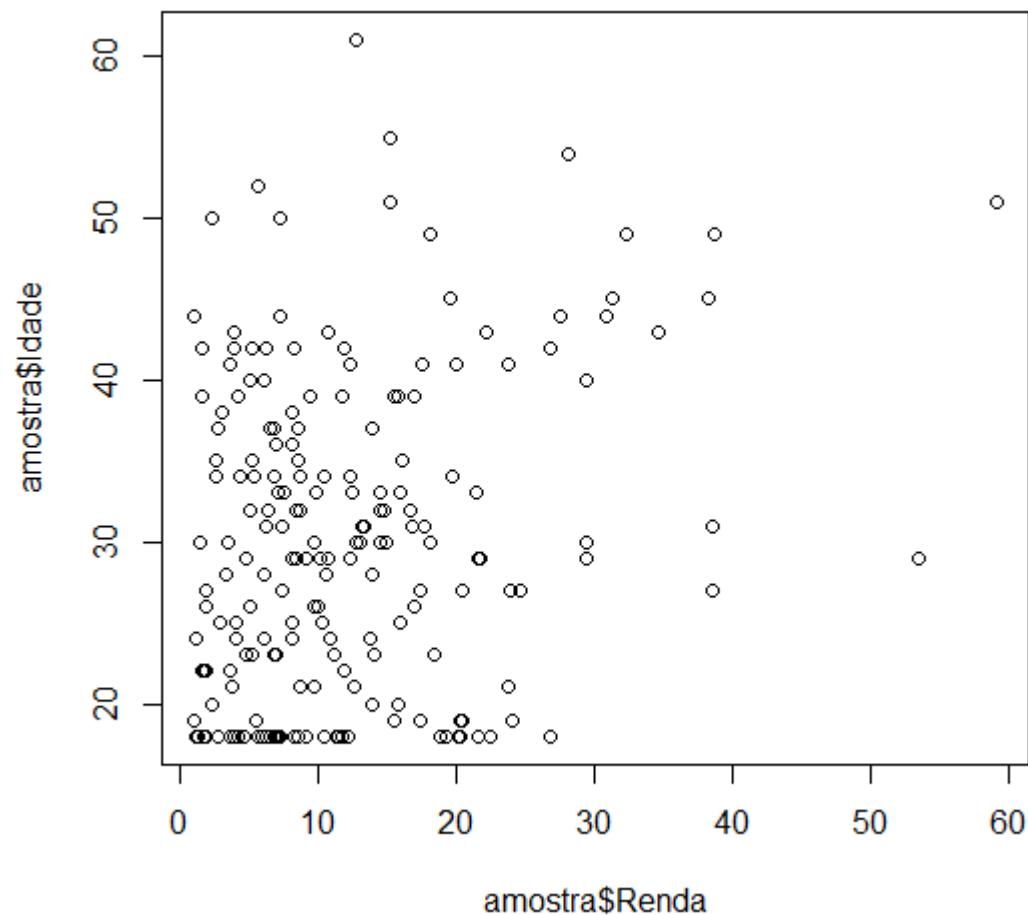
2. Bivariado

```
plot(amostra$Renda, amostra$Idade)
```

O gráfico de dispersão é a 1ª etapa da **Análise de Correlação**: Área da estatística que analisa o **comportamento conjunto** de duas **variáveis quantitativas** e verifica se existe algum tipo de **relação** entre elas.

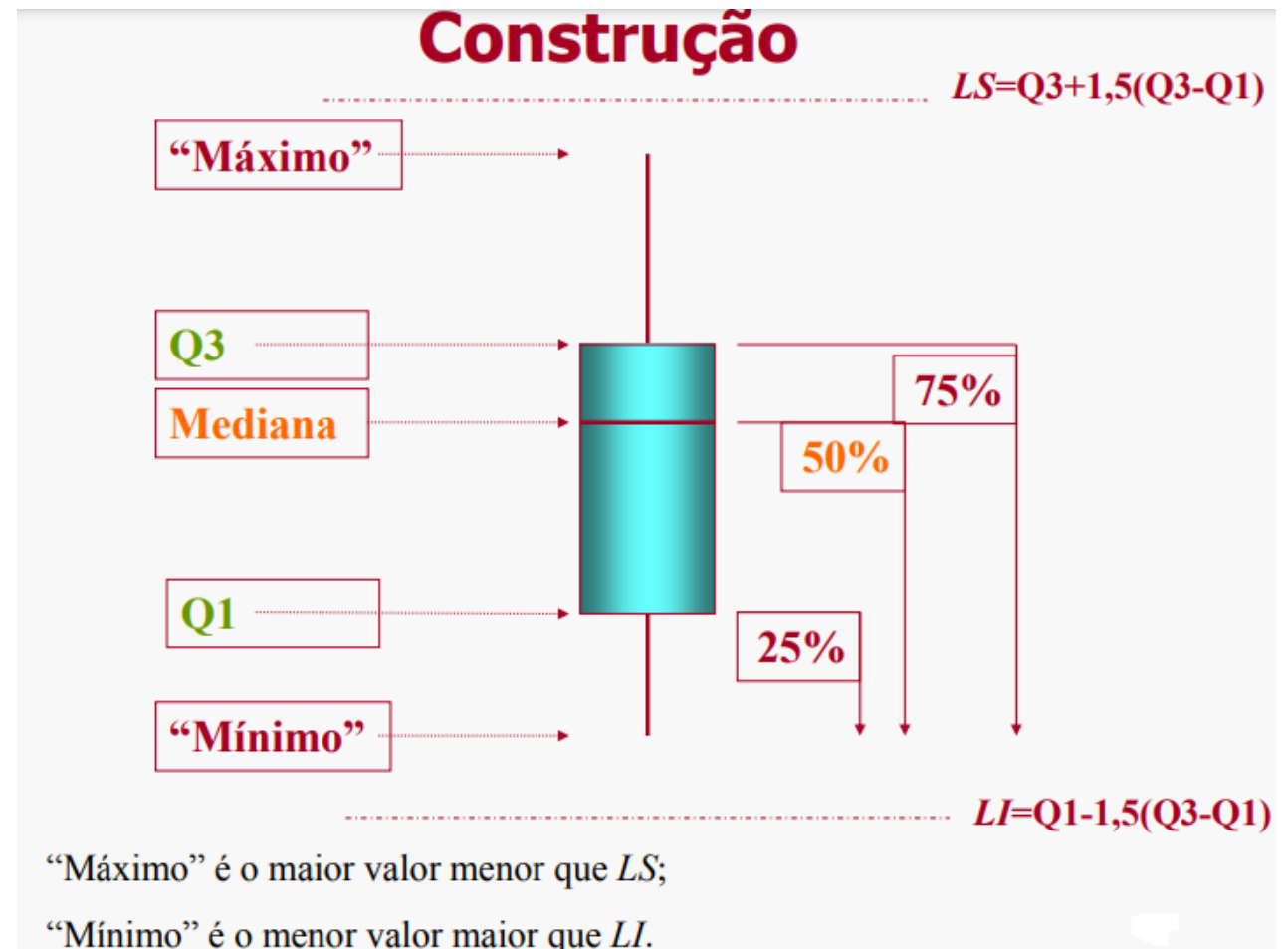
Verifica-se visualmente a existência de associação entre as variáveis a partir do **diagrama de dispersão**.

Gráfico de dispersão ou Scatter plots

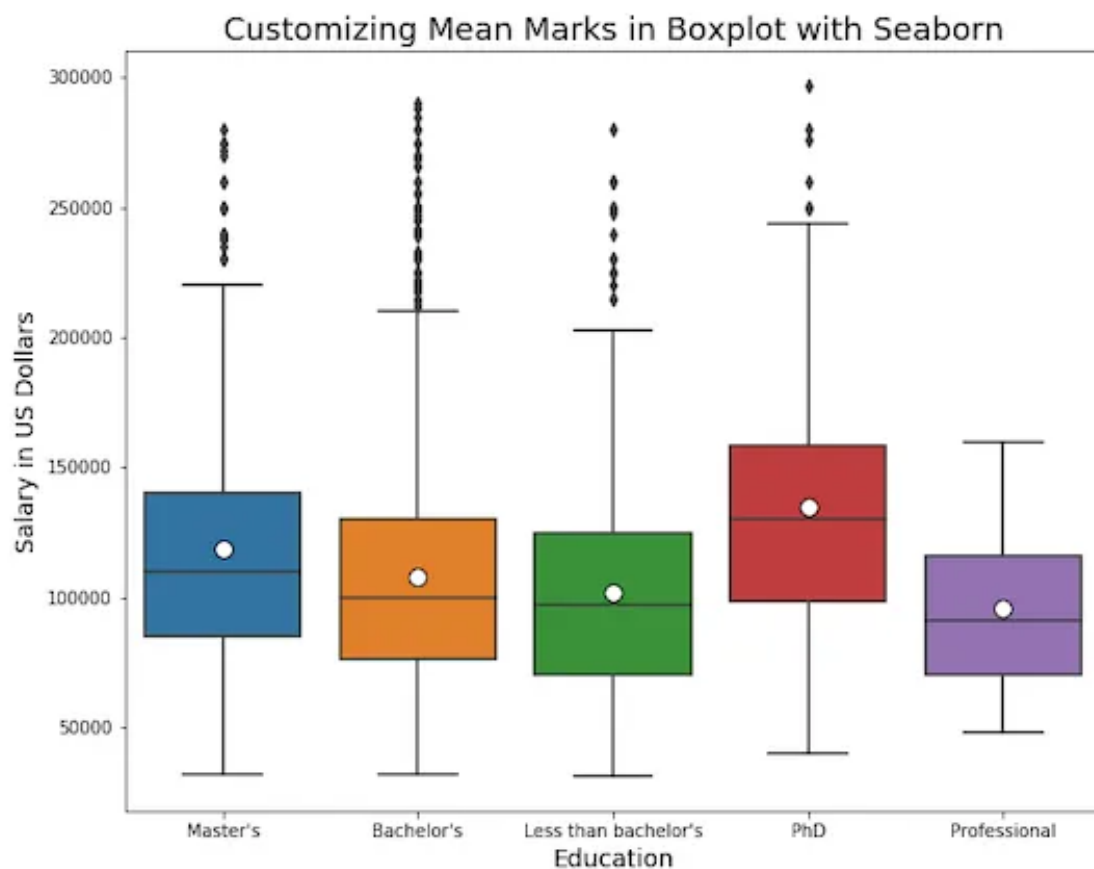


BoxPlot

- O boxplot (gráfico de caixa) é um gráfico utilizado para avaliar a distribuição dos dados.
- Fornece diferentes informações do conjunto de dados:
- **LOCALIZAÇÃO - DISPERSÃO – ASSIMETRIA – OUTLIERS**
- Ele é formado pelo primeiro e terceiro quartil e pela mediana.
- As hastes inferiores e superiores se estendem, respectivamente, até o valor mínimo e o valor máximo que estejam dentro dos limites calculados.
- Os pontos fora destes limites são considerados valores discrepantes (*outliers*) e são denotados por asterisco (*).



Interpretação do BoxPlot



O Retângulo é chamado de **Intervalo Interquartilico** e contém 50% dos valores do conjunto de dados.

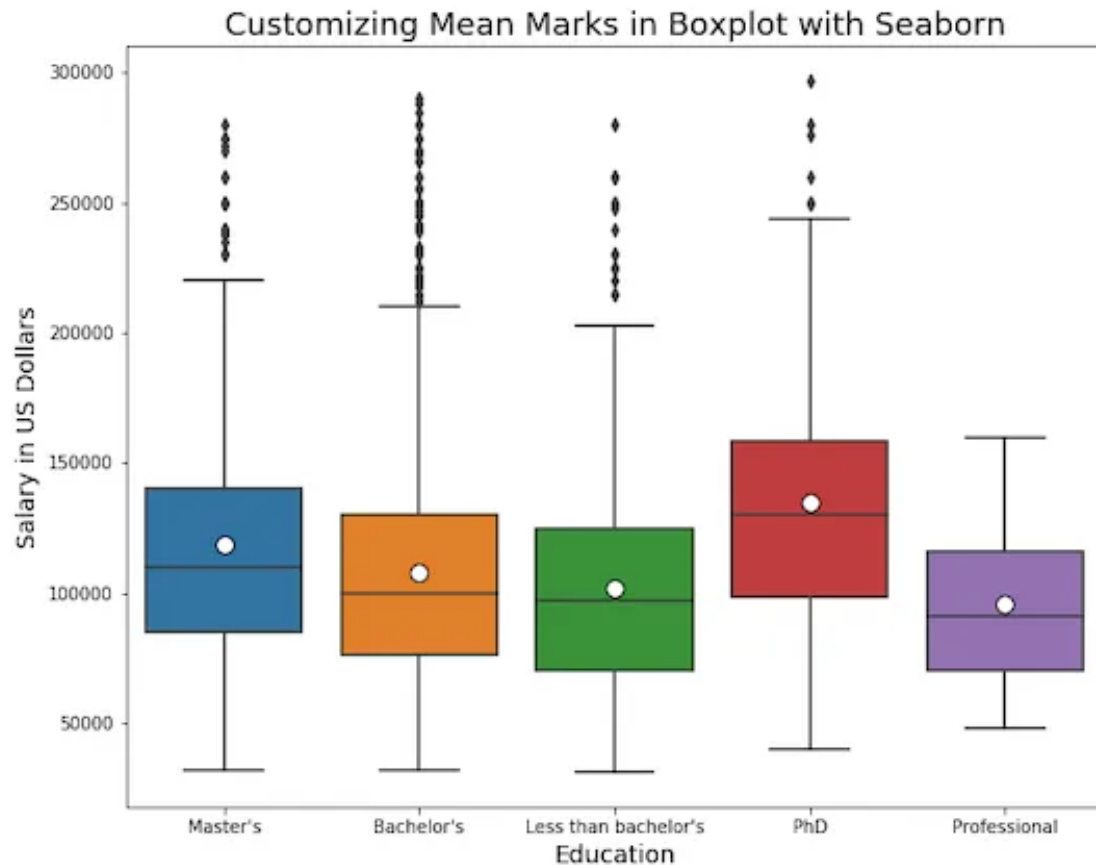
A posição da linha mediana no retângulo informa sobre a assimetria da distribuição.

Uma distribuição simétrica teria a mediana no centro do retângulo.

- Se a mediana é próxima de Q1, então, os dados são **positivamente assimétricos**.

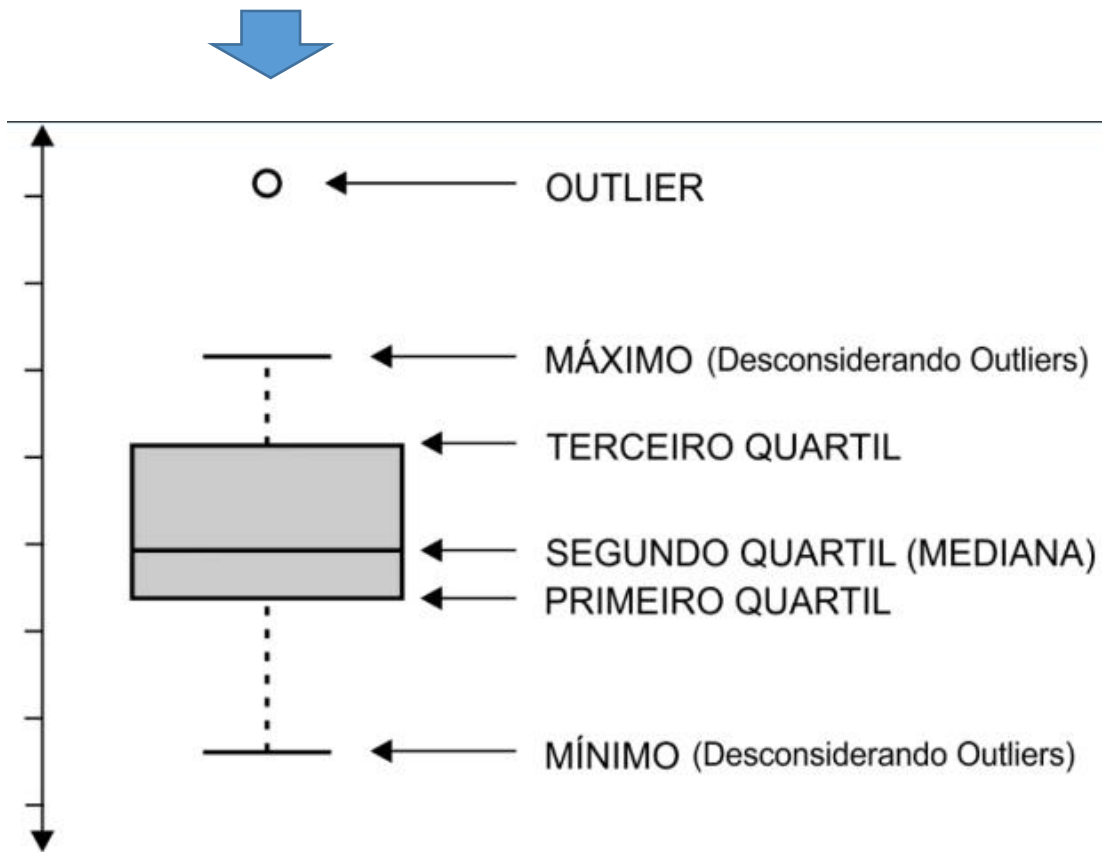
- Se a mediana é próxima de Q3 os dados são **negativamente assimétricos**.

Interpretação do BoxPlot



- O tamanho do retângulo indica sobre a variabilidade dos dados, ou seja:
 - Caixas mais achatadas indica uma distribuição mais homogênea, com pouca variabilidade em torno da média (valores parecidos entre si).
 - Caixas mais largas indicam uma distribuição mais heterogênea, ou seja, muita variabilidade em torno da média (valores diferentes entre si).

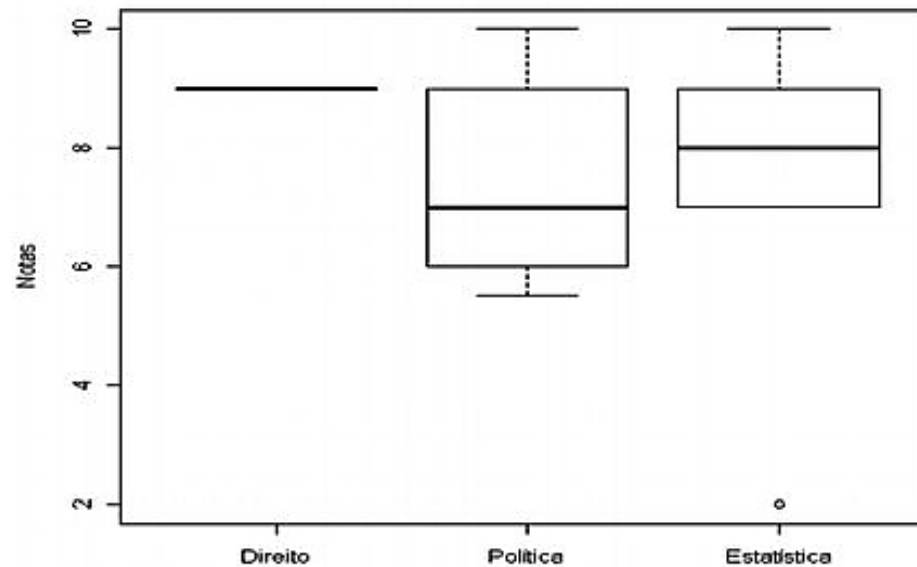
BoxPlot



- Os pontos ou asteriscos fora das “linhas” desenhadas são chamados de **OUTLIERS**.
- Em estatística, outlier significa valor aberrante ou valor atípico, é uma observação que apresenta um grande afastamento das demais ou que é inconsistente.
- A existência de outliers implica, tipicamente, em prejuízos a interpretação dos resultados dos testes estatísticos aplicados às amostras.
- **Causas** do aparecimento de outliers:
 - Leitura, anotação ou transição incorreta dos dados.
 - Erro na execução do experimento ou na tomada da medida.
 - Variabilidade inerente dos elementos da população.

INTERPRETAÇÃO DE GRÁFICOS

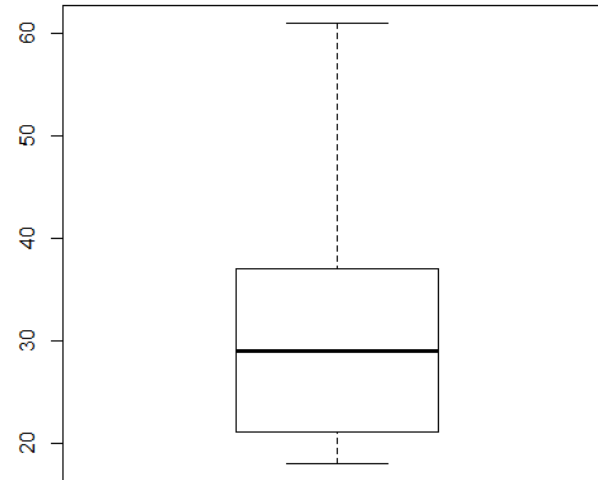
Compare e indique as diferenças existentes entre as distribuições das variáveis notas dos cursos de Direito, Política e Estatística, utilizando o “boxplot” e as medidas calculadas em:



Por que em Direito o gráfico é uma reta? Quem tem maior variabilidade e o por qual razão? Porque a linha da mediana em Política não está centralizada? Existe algum dado discrepante?

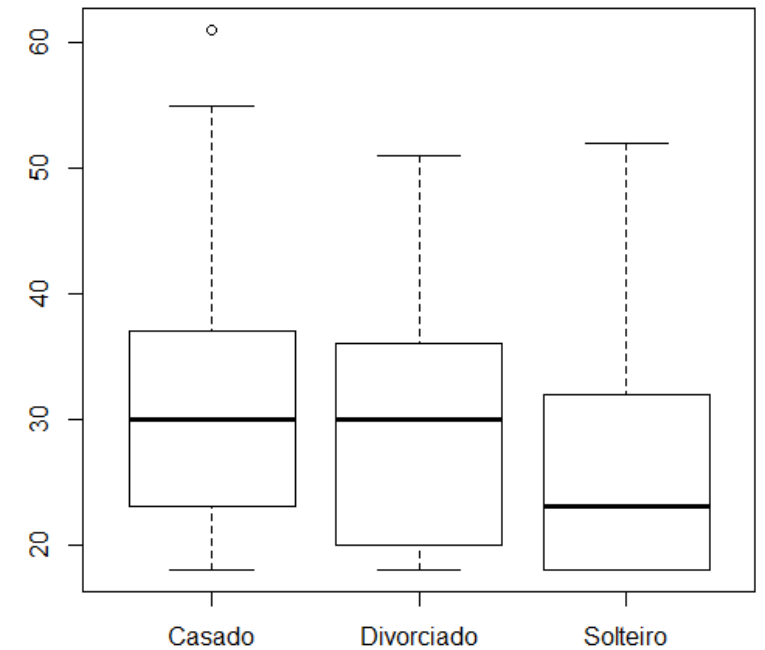
Boxplot

`boxplot(amostra$Idade)`



Boxplot Estratificado:

`boxplot(amostra$Idade~amostra$EstCivil)`

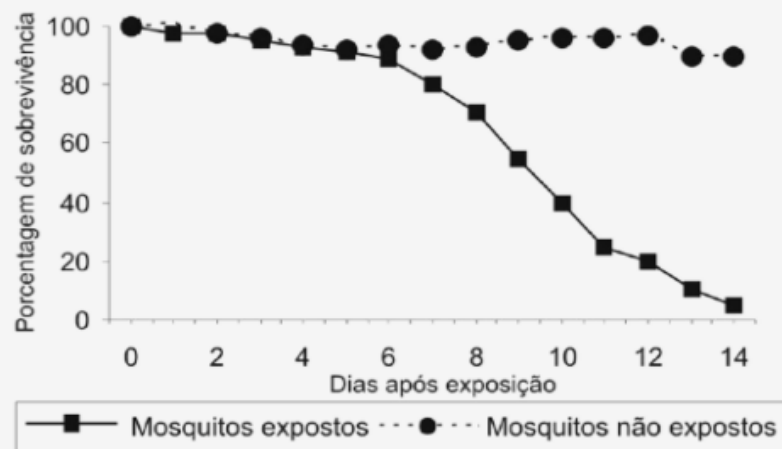


INTERPRETAÇÃO DE GRÁFICOS

- A primeira exigência ao deparar com um gráfico parece óbvia, mas nem sempre é verificada com o cuidado necessário:
 - ***IDENTIFICAR EXATAMENTE O QUE REPRESENTAM O EIXO HORIZONTAL E O EIXO VERTICAL, BEM COMO AS UNIDADES DE CADA GRANDEZA***
 - A característica de um gráfico cartesiano é basicamente mostrar como uma grandeza varia em função de outra.
 - Interpretar significa que o gráfico pode nos dar mais informações além daquelas exibidas à primeira vista....
-

INTERPRETAÇÃO DE GRÁFICOS

Foram publicados recentemente, trabalhos relatando o uso de fungos como controle biológico de mosquitos transmissores da malária. Observou-se o percentual de sobrevivência dos mosquitos *Anopheles* sp. após exposição ou não a superfícies cobertas com fungos sabidamente pesticidas, ao longo de duas semanas. Os dados obtidos estão presentes no gráfico abaixo.



No grupo exposto aos fungos, o período em que houve 50% de sobrevivência ocorreu entre os dias:

a) 2 e 4

b) 4 e 6

c) 6 e 8

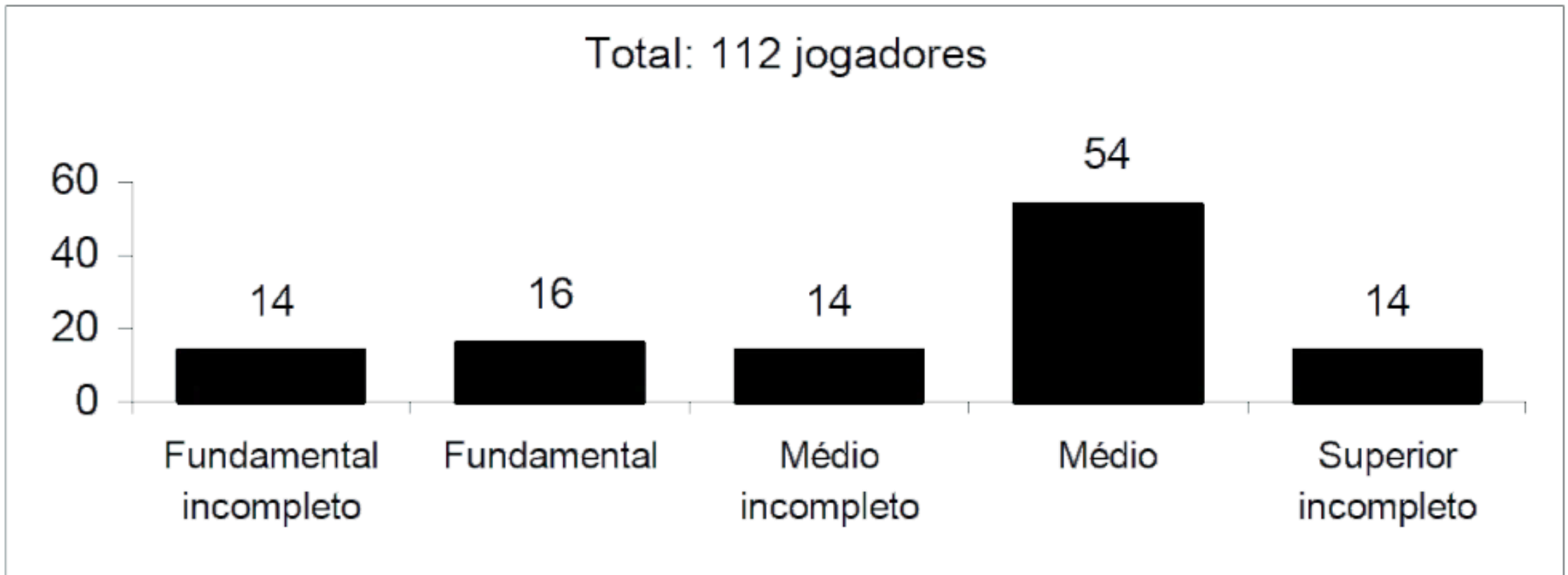
d) 8 e 10

e) 10 e 12

INTERPRETAÇÃO DE GRÁFICOS

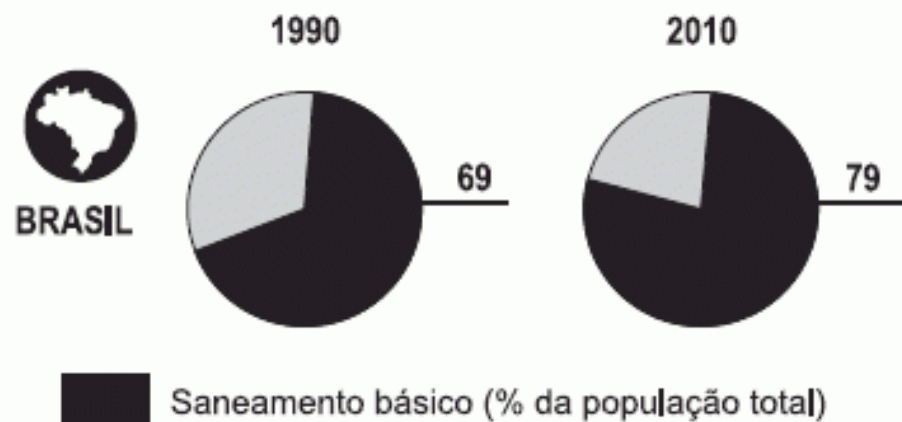
A escolaridade dos jogadores de futebol nos grandes centros é maior do que se imagina, como mostra a pesquisa abaixo, realizada com os jogadores profissionais dos quatro principais clubes de futebol do Rio de Janeiro.

De acordo com esses dados, determine o percentual dos jogadores dos quatro clubes que concluíram o Ensino Médio.



A força da água limpa

As novas tecnologias e o empenho dos organismos públicos, associados aos interesses e boas práticas da iniciativa privada, ampliaram a rede de esgotos.



Fontes: Unep e Unicef

Revista Veja. São Paulo: Abril. 13 jun. 2012. p. 97. Adaptado.

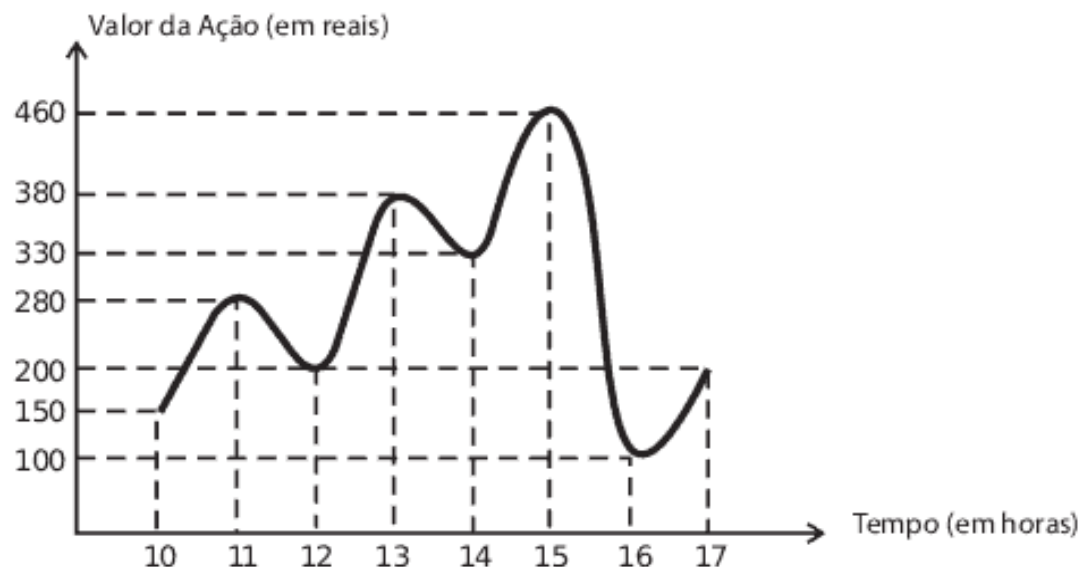
Considere que, em 1990, a população brasileira era de 145 milhões de habitantes e, em 2010, de 190 milhões.

Com base nos percentuais apresentados na reportagem, o número de habitantes, no Brasil, que contam com saneamento básico aumentou, de 1990 para 2010, em, aproximadamente,

INTERPRETAÇÃO DE GRÁFICOS

- a) 65 milhões
- b) 50 milhões
- c) 45 milhões
- d) 25 milhões
- e) 10 milhões

Investidor	Hora da Compra	Hora da Venda
1	10:00	15:00
2	10:00	17:00
3	13:00	15:00
4	15:00	16:00
5	16:00	17:00



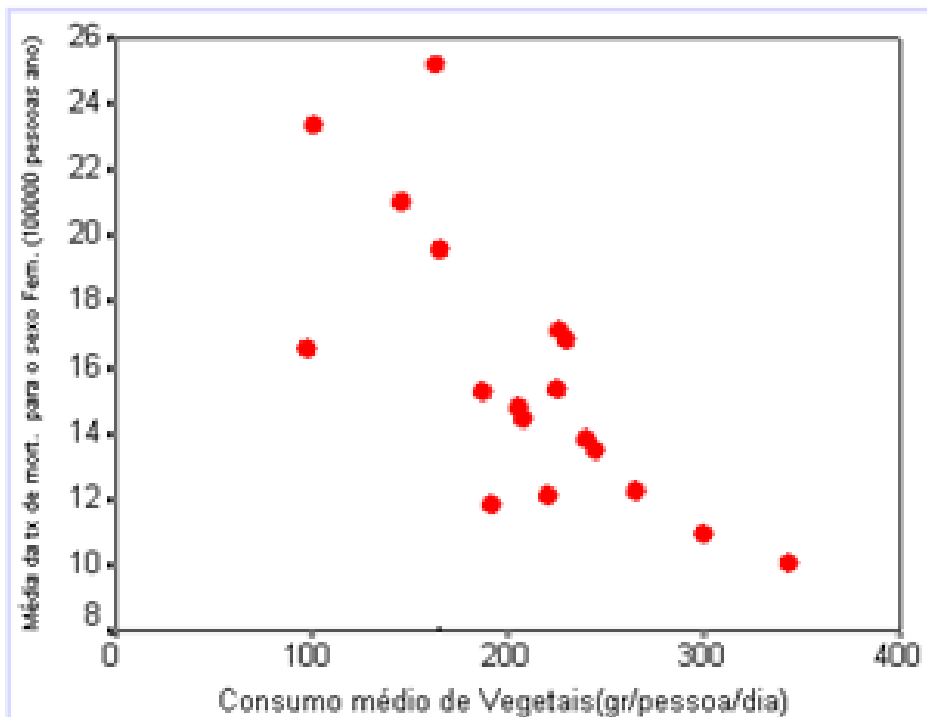
INTERPRETAÇÃO DE GRÁFICOS

- O gráfico fornece os valores das ações da empresa XPN, no período das 10 às 17 horas, num dia em que elas oscilaram acentuadamente em curtos intervalos de tempo.
- Neste dia, cinco investidores compraram e venderam o mesmo volume de ações, porém em horários diferentes, de acordo com a tabela.
- Com relação ao capital adquirido na compra e venda das ações, determine investidor que fez o melhor negócio.

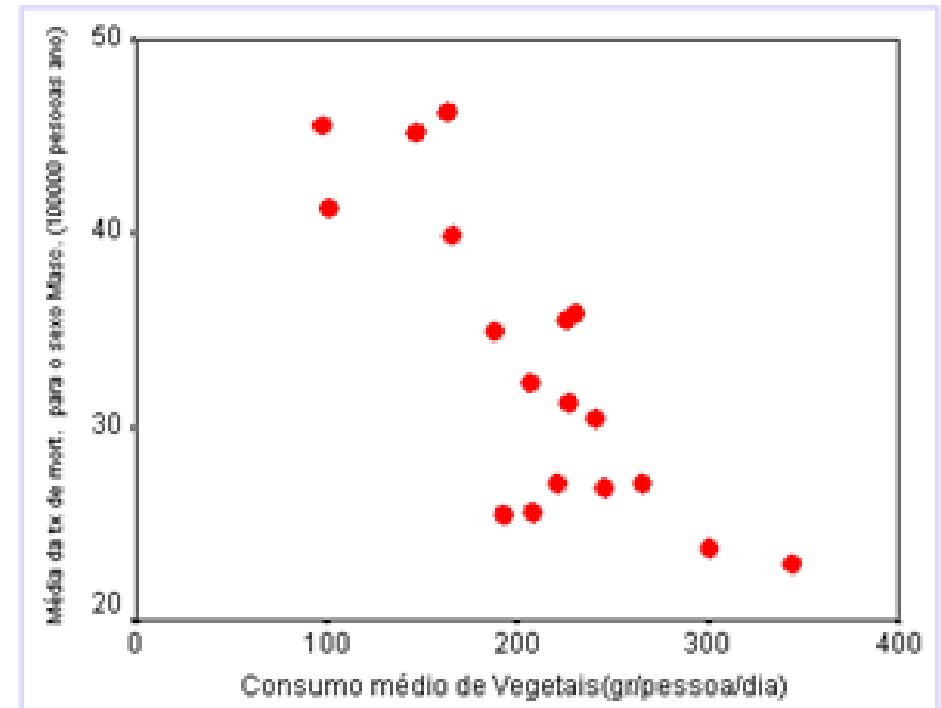
INTERPRETAÇÃO DE GRÁFICOS

Considere um estudo onde o objetivo era analisar a relação entre o consumo de vegetais com a mortalidade por câncer no estômago, feminino e masculino, na região de PP. Uma forma simples de visualizar uma possível relação entre a quantidade de vegetais consumida e a taxa de mortalidade é utilizar um diagrama de dispersão para estas duas variáveis.

Interprete o que tais resultados indicam no estudo.



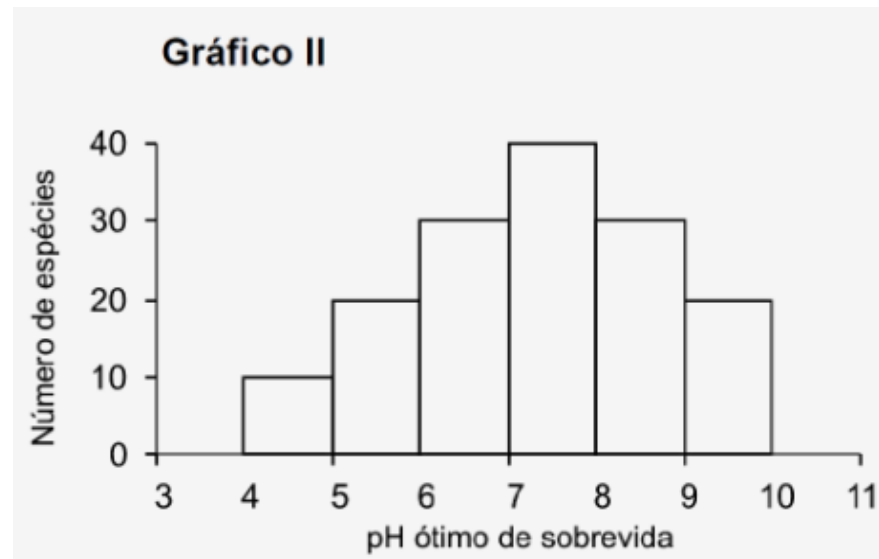
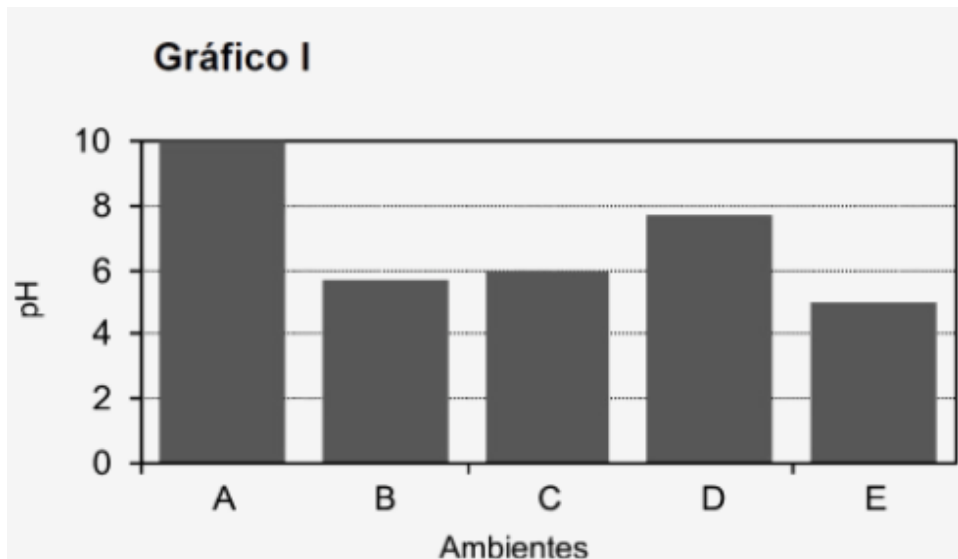
Masculino



Feminino

INTERPRETAÇÃO DE GRÁFICOS

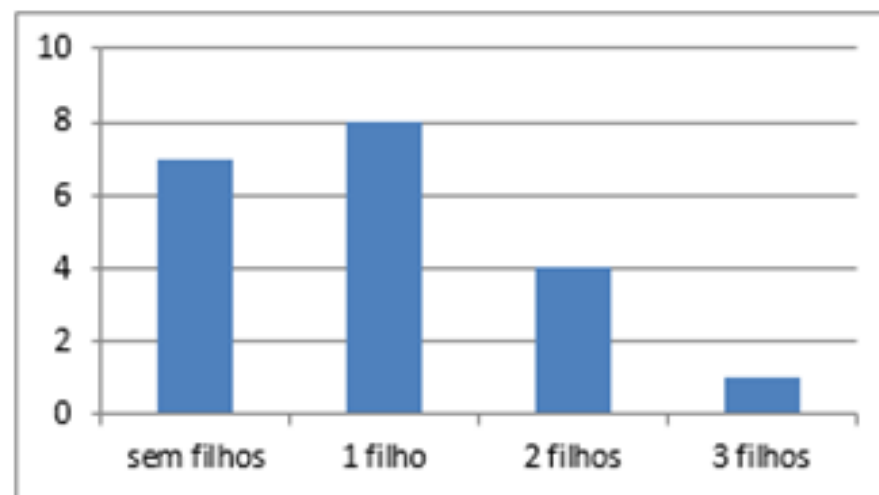
Um estudo caracterizou 5 ambientes aquáticos, nomeados de A a E, em uma região, medindo parâmetros físico-químicos de cada um deles, incluindo o pH nos ambientes. O Gráfico I representa os valores de pH dos 5 ambientes. Utilizando o gráfico II, que representa a distribuição estatística de espécies em diferentes faixas de pH, pode-se esperar um maior número de espécies no ambiente:



- a) A
- b) B
- c) C
- d) D
- e) E

INTERPRETAÇÃO DE GRÁFICOS

As formandas do Ensino Médio do Colégio Atlas, turma de 2006, resolveram se reencontrar para uma reunião comemorativa 10 anos após a formatura. Algumas haviam se casado e tido filhos nesse período e outras não. A distribuição das ex-alunas de acordo com o número de filhos que tiveram é mostrada no gráfico de barras abaixo:



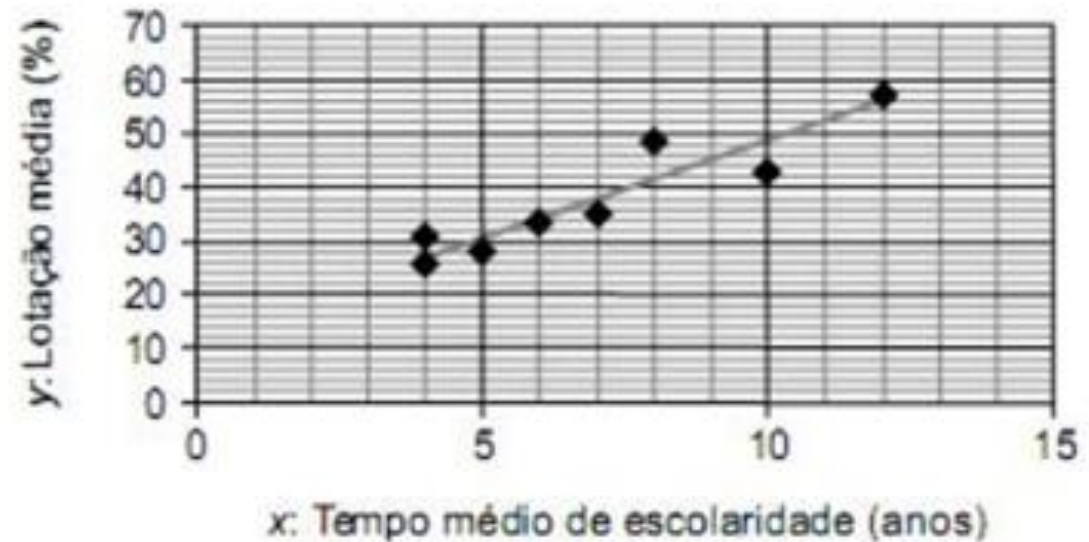
A porcentagem de ex-alunas que tiveram filhos nesse período é igual à:

- a) 55% b) 40% c) 20% d) 35% e) 65%

INTERPRETAÇÃO DE GRÁFICOS

- O diagrama de dispersão a seguir resulta de uma pesquisa com amostras aleatórias para comparar níveis de escolaridade (em anos de estudo) com a lotação média (em percentual) das salas de teatro das localidades de onde as amostras aleatórias foram colhidas.

Anos de Escolaridade x lotação de salas de teatro



INTERPRETAÇÃO DE GRÁFICOS

CONCLUSÕES:

- **Centro dos dados** (a média ou mediana)
- **Amplitude dos dados** (máximo – mínimo)
- **Variabilidade dos dados**
- **Simetria ou assimetria** do conjunto de dados e a presença ou não de outliers.
- **Valores discrepantes (outliers)**

