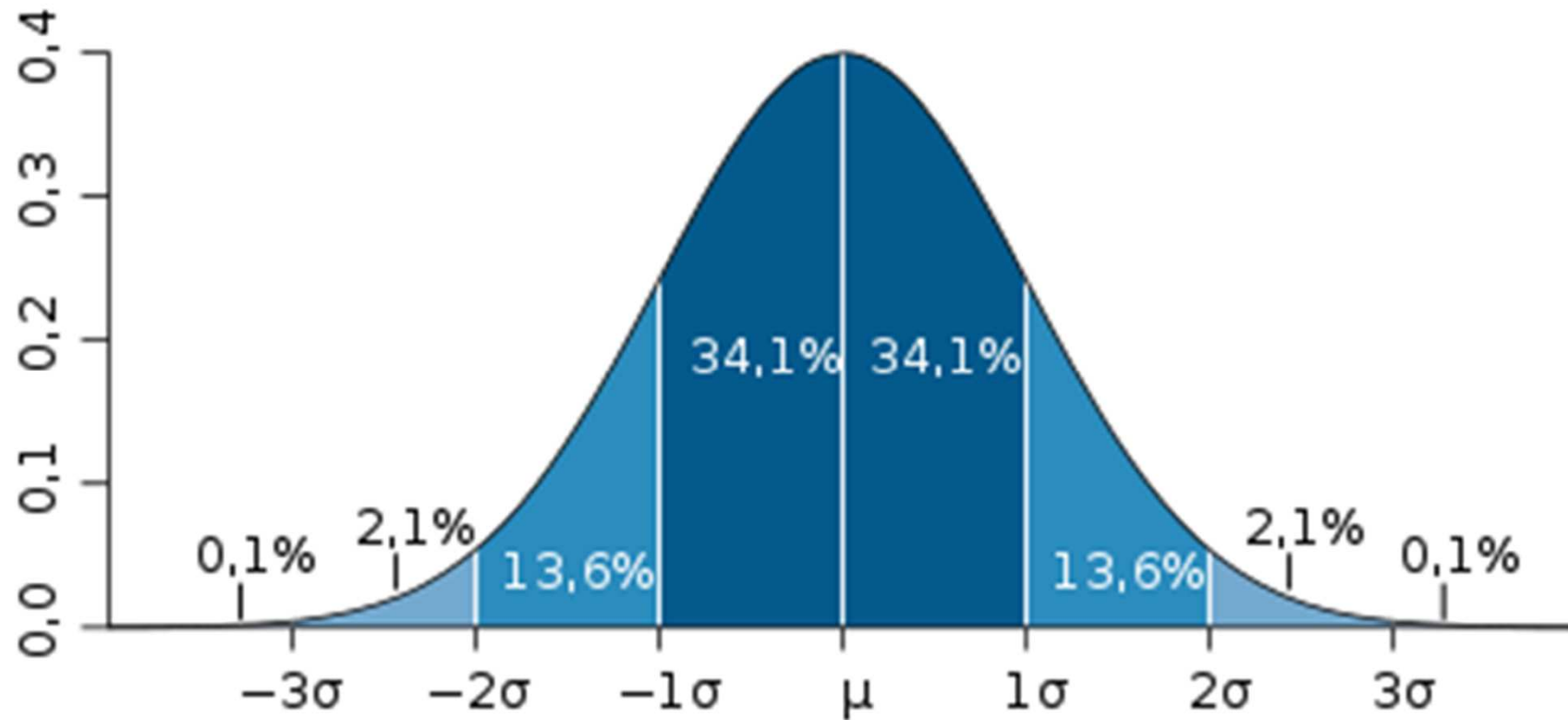


Importância prática do desvio padrão



Obs. Para uma distribuição normal.

Organização e representação dos dados

Outra forma de organizar e resumir a informação contida em dados observados é por meio de gráficos.

- Gráfico de pontos
- Gráfico de dispersão
- Gráfico de setores (pizza)
- Gráfico de barras ou colunas
- Histograma
- Gráfico de Pareto
- Box-plot

Elementos de um gráfico

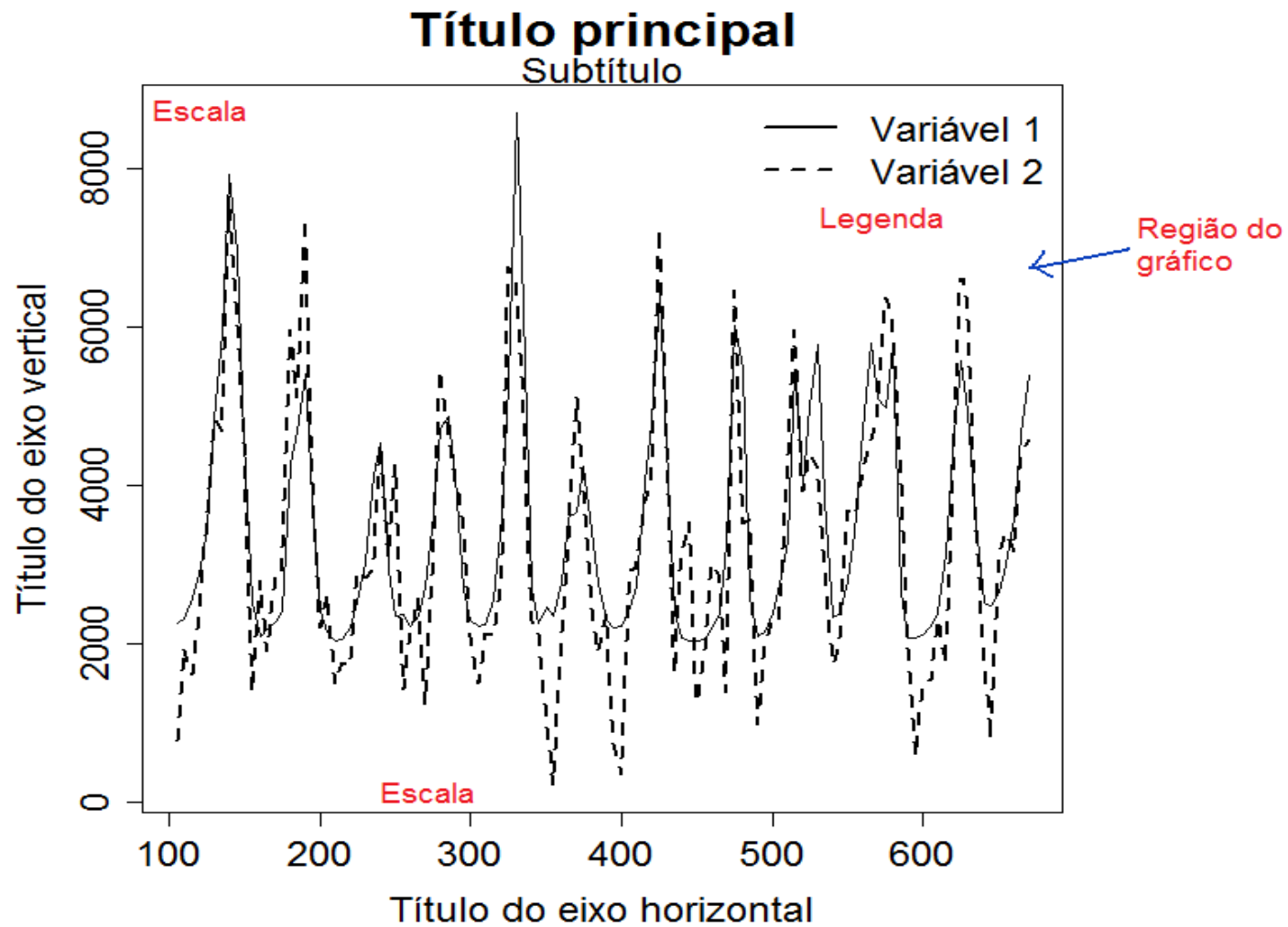
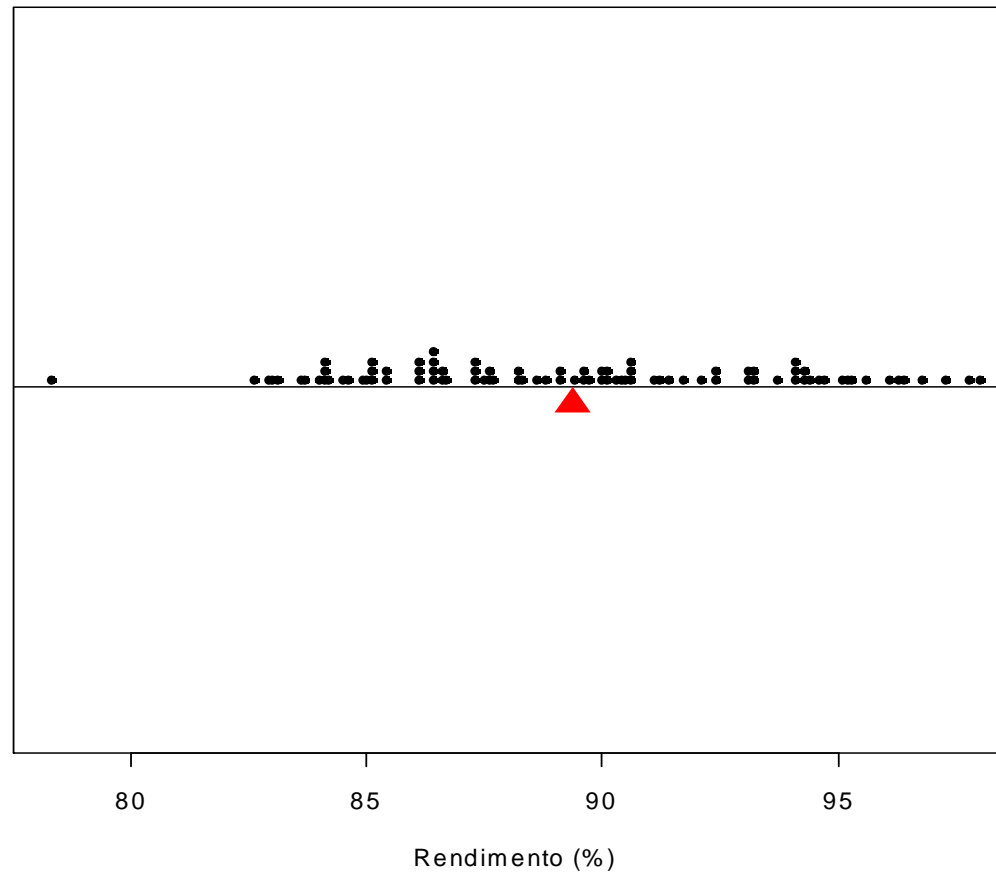


Figura 1. Descrição do gráfico.

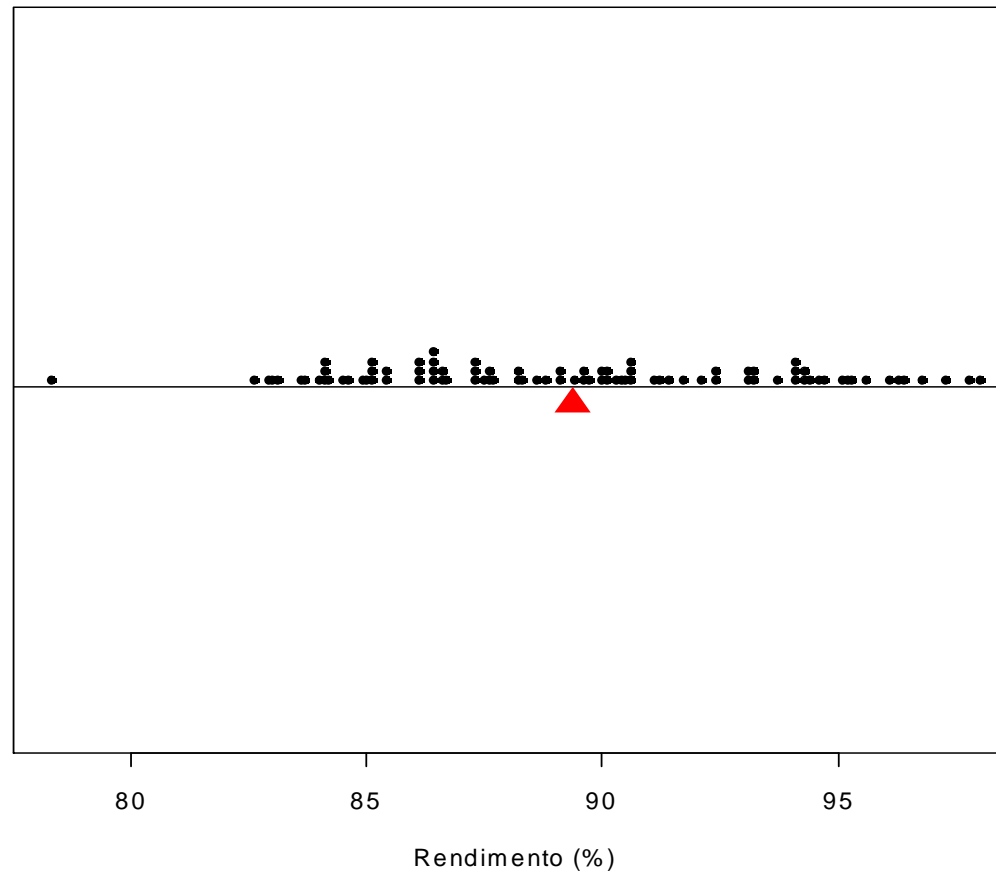
Gráfico de pontos

Cada ponto representa uma observação com valores na resta. O triângulo representa a média dos dados. Valores iguais são dispostos verticalmente no mesmo ponto.



Exemplo em R (Gráfico de pontos)

```
> stripchart(dados, xlab="Rendimento (%)", pch= 20, method = "stack")  
> abline(h = 0.98)  
> points(mean(dados), 0.93, pch = 17, col = "red", cex = 2)
```



Representação gráfica adequadas

Variáveis qualitativas:

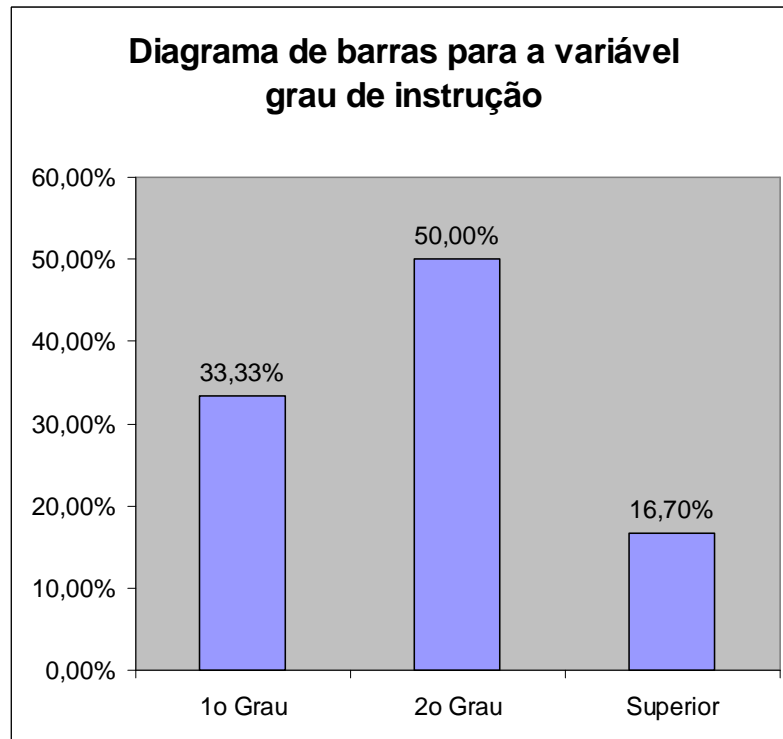
- Gráfico de barras
- Gráfico de Pareto
- Gráfico de setores (pizza)

Variáveis quantitativas:

- Se estiverem organizados em intervalos de classe, valem os mesmos gráficos das variáveis qualitativas, além do histograma.
- Diagrama em caixas (Box plot).
- Para dados temporais: gráfico de linha.

Gráfico de colunas ou barras

Retângulos verticais (ou horizontais) espaçados com alturas (ou bases) iguais às frequências dos valores da variável.



Grau de instrução

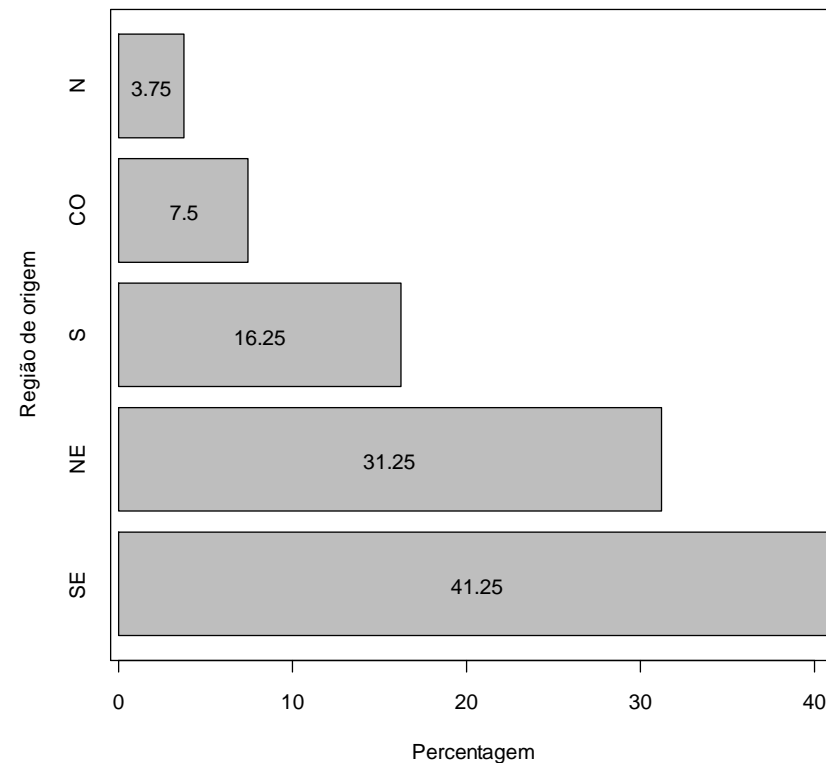


Gráfico de Pareto

Gráfico de barras com os valores da variável em ordem decrescente de frequências e com as **frequências relativas acumuladas** no segundo eixo vertical.

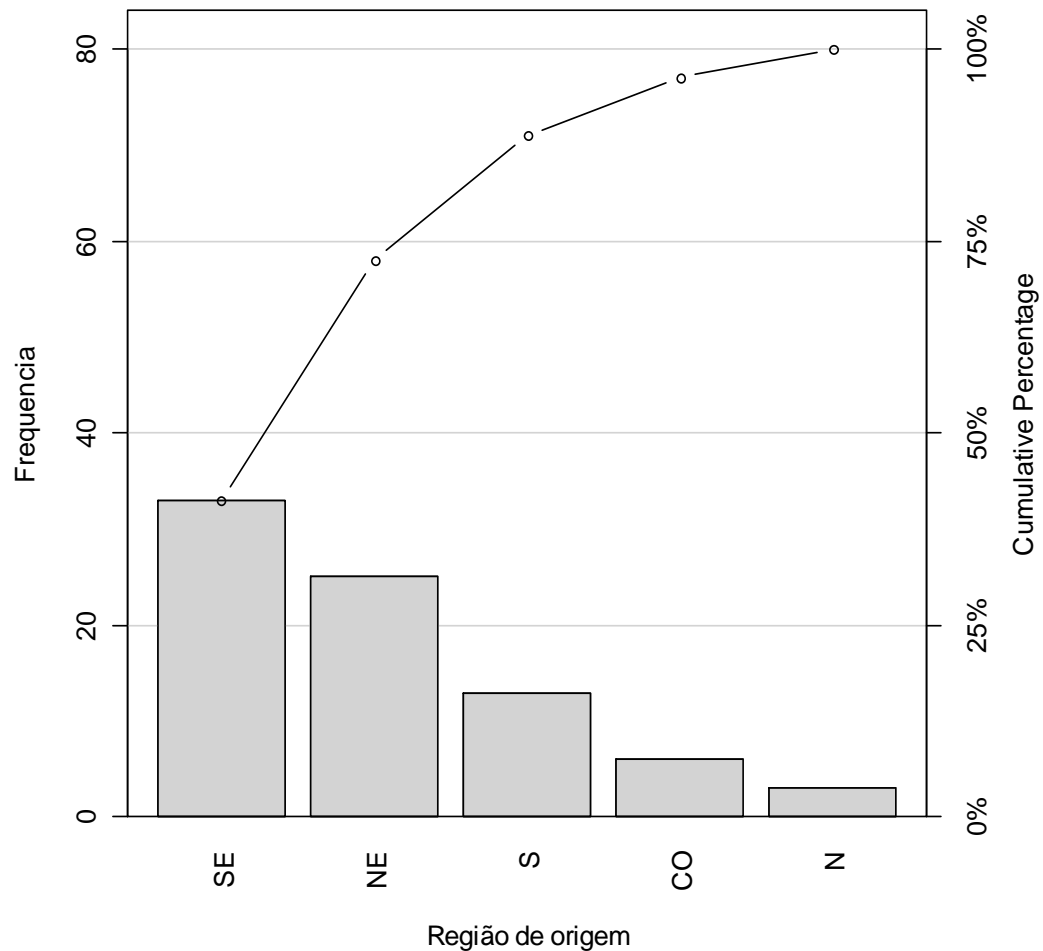


Gráfico de setores (pizza)

Gráfico **circular** utilizado para destacar a composição das partes de um todo. O ângulo central de cada setor é proporcional à frequência representada (usualmente em %).

Diagrama circular para a variável grau de instrução

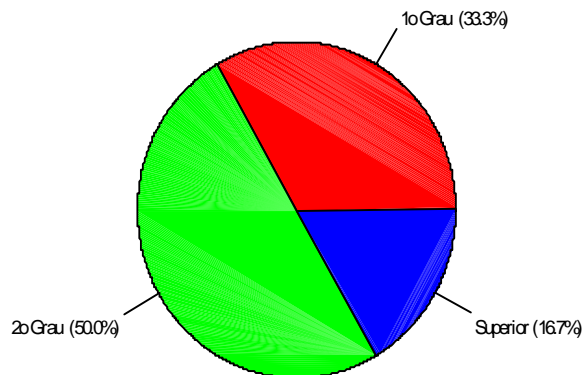
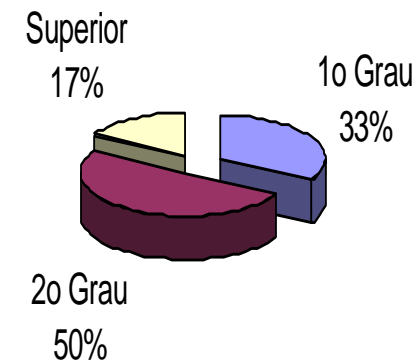
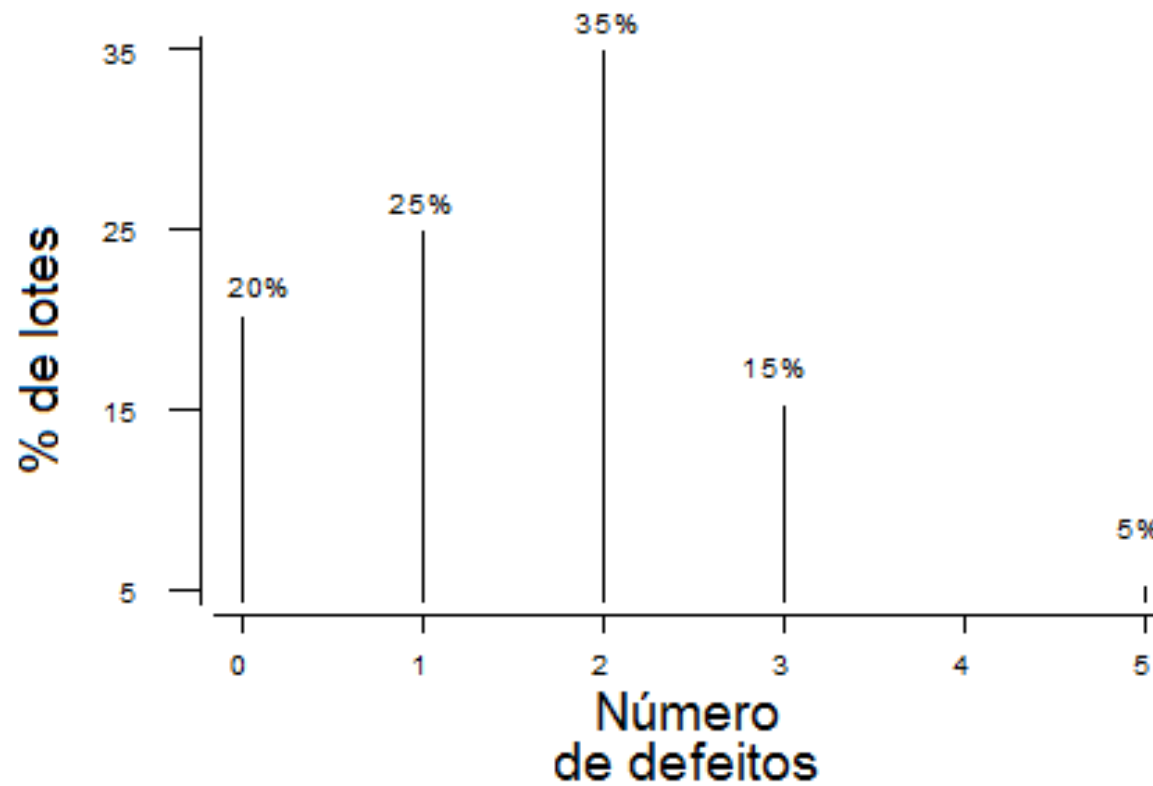


Diagrama circular para a variável grau de instrução



Representação gráfica



Histograma

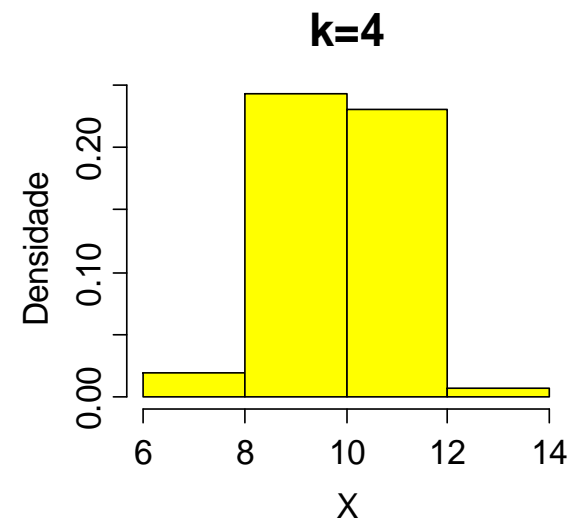
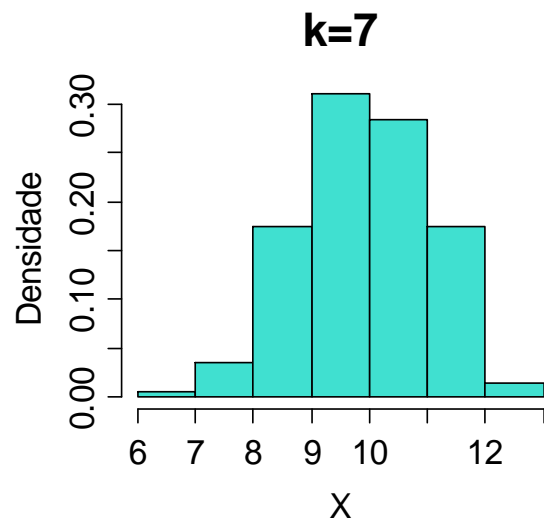
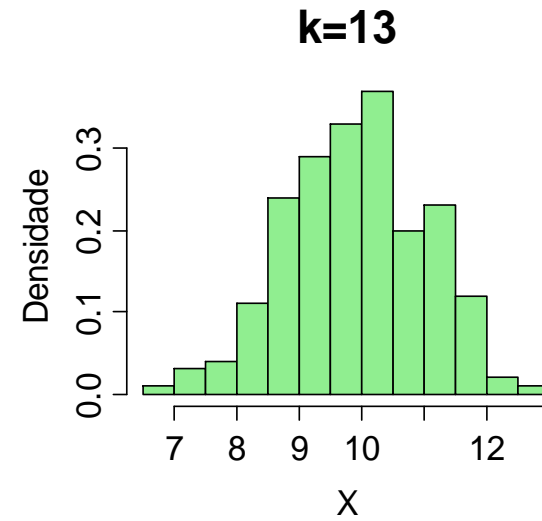
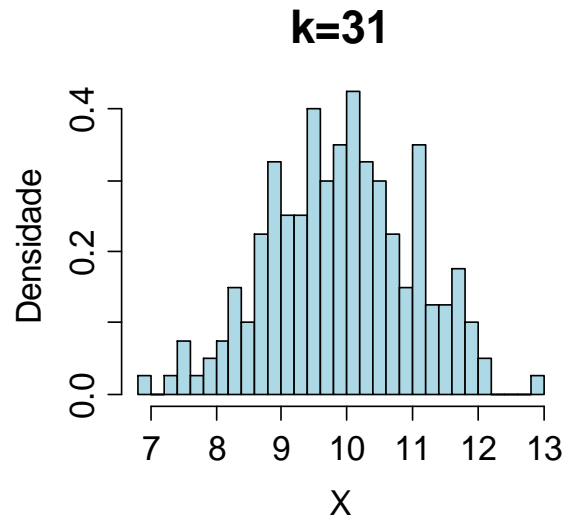
Gráfico de **barras adjacentes** com **bases** iguais às **amplitudes** das classes e **alturas** iguais às **densidades**.

Nota 1: As barras são juntas, pois um intervalo de classe começa quando termina o outro.

Nota 2: Escolha do número de classes (geralmente, $5 \leq k \leq 15$).

Nota 3: Na construção do histograma, quanto **maior** for o **n**, **melhor**.

Histograma



Exemplo

Variável: **viscosidade** (em u.v.) de um líquido a uma certa temperatura.

```
13.9 14.9 15.9 15.8 14.8 15.1 15.8 15.0 15.1 14.6 14.7 16.6 13.6 15.9 13.1
15.2 14.7 16.0 15.6 17.4 15.3 14.2 15.9 15.1 15.9 16.1 16.2 13.8 14.6 16.0
15.8 15.5 16.5 17.1 15.3 15.5 17.8 15.4 15.4 14.6
```

Amostra **ordenada**:

```
13.1 13.6 13.8 13.9 14.2 14.6 14.6 14.6 14.7 14.7 14.8 14.9 15.0 15.1 15.1
15.1 15.2 15.3 15.3 15.4 15.4 15.5 15.5 15.6 15.8 15.8 15.8 15.9 15.9 15.9
15.9 16.0 16.0 16.1 16.2 16.5 16.6 17.1 17.4 17.8
```

$n = 40$

Min.	Median	Mean	Max.
13.10	15.40	15.39	17.80

Exemplo. Histograma.

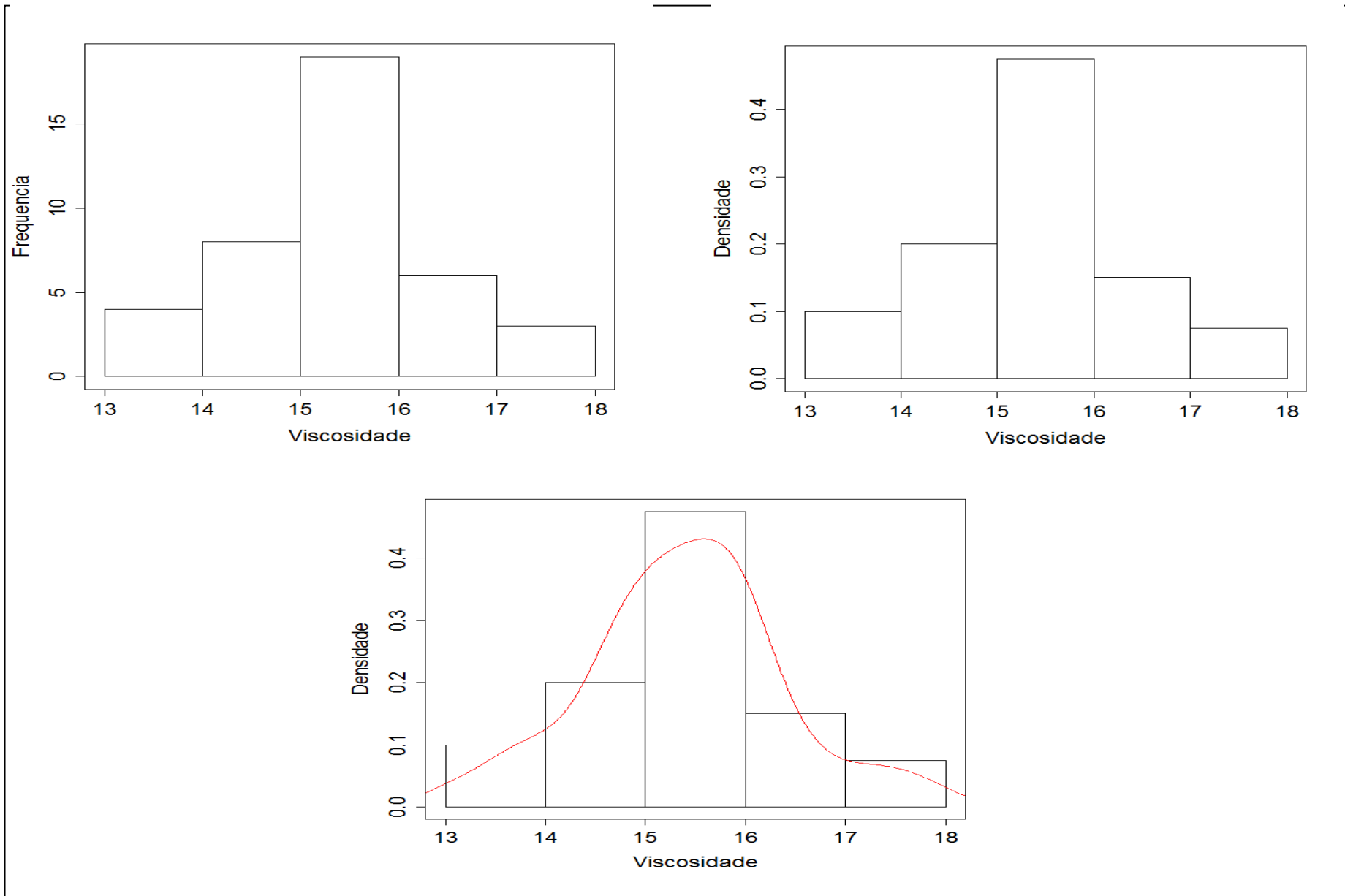
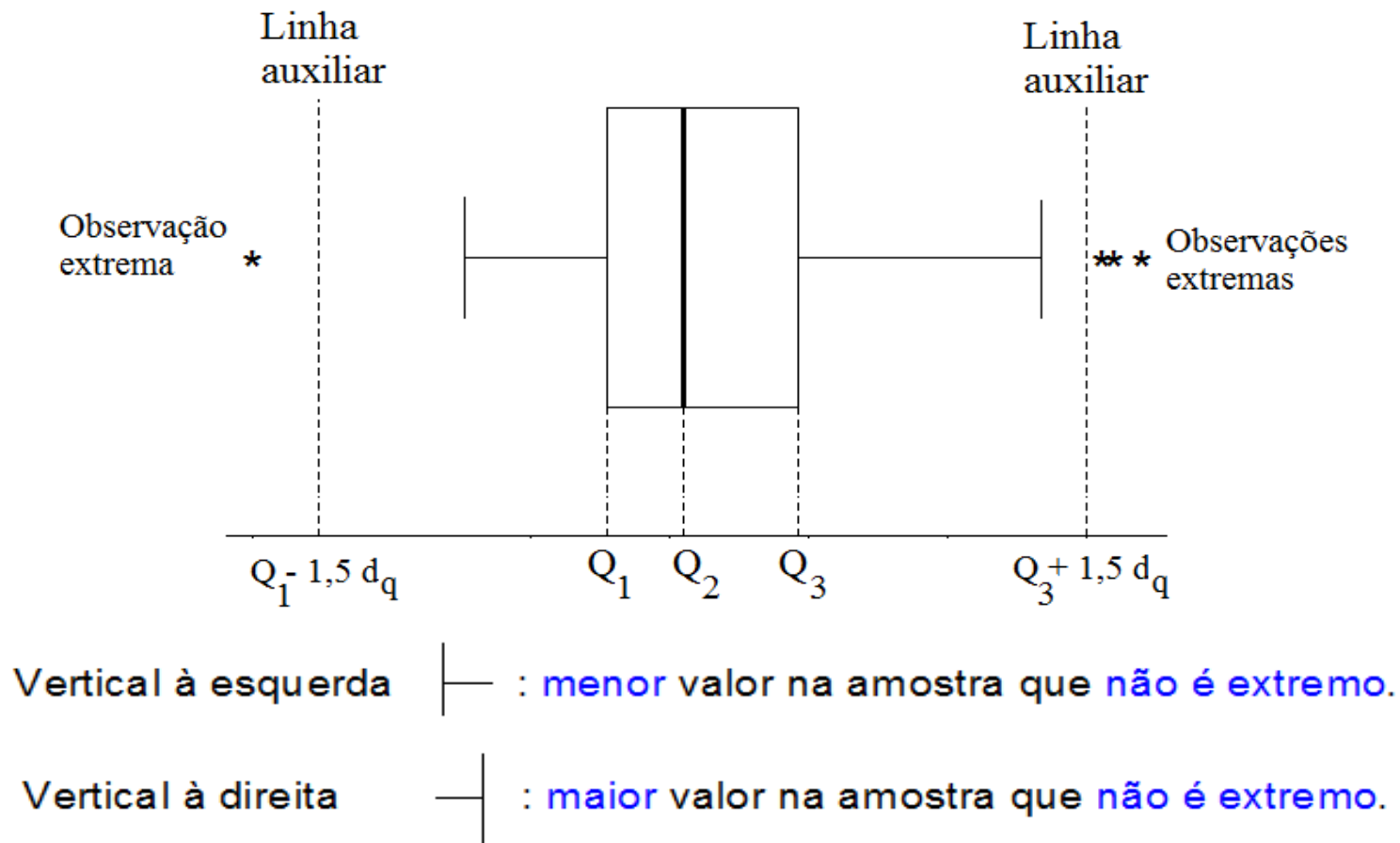


Diagrama em caixa (*box plot*)

Representação dos dados por meio de um **retângulo** construído com os **quartis**. Fornece informação sobre a variabilidade ($d_q = Q_3 - Q_1$) e valores extremos.



Exemplo. Variável viscosidade.

1º quartil (Q1) = 14,775. Em R: `quantile(dados, 0.25)`

Mediana (Md ou Q2) = 15,4. Em R: `quantile(dados, 0.5)`

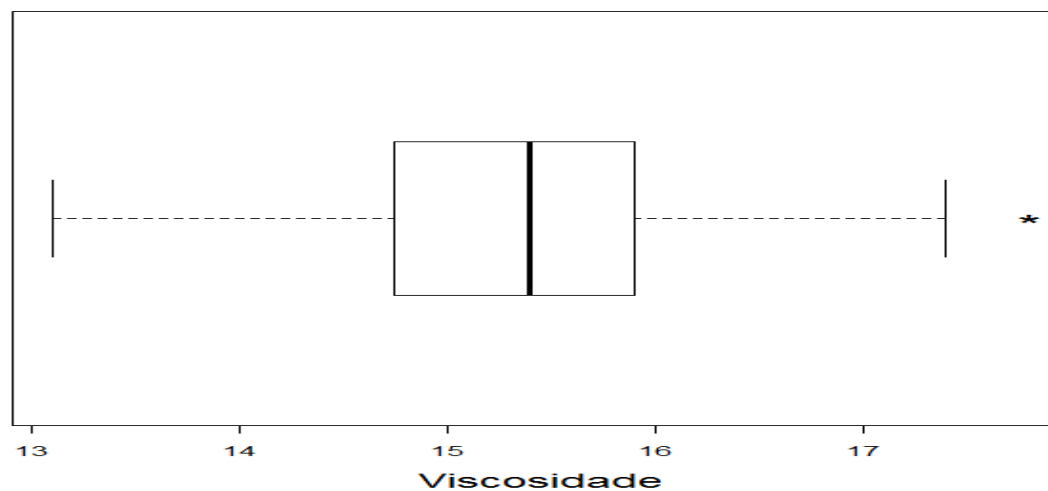
3º quartil (Q3) = 15,9. Em R: `quantile(dados, 0.75)`

d_q = intervalo interquartil = $Q3 - Q1 = 1,125$.

Lnhas auxiliares passam por $Q1 - 1,5d_q = 13,0875$ e

$Q3 + 1,5d_q = 17,5875$.

```
> boxplot(viscosidade, xlab = "Viscosidade", horizontal = TRUE)
```



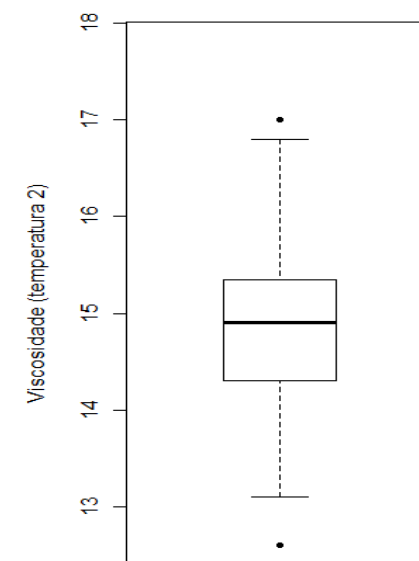
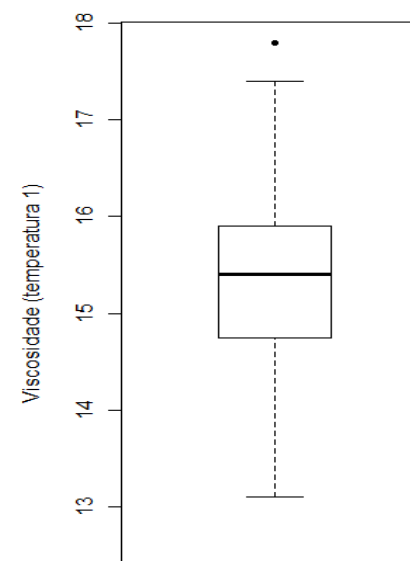
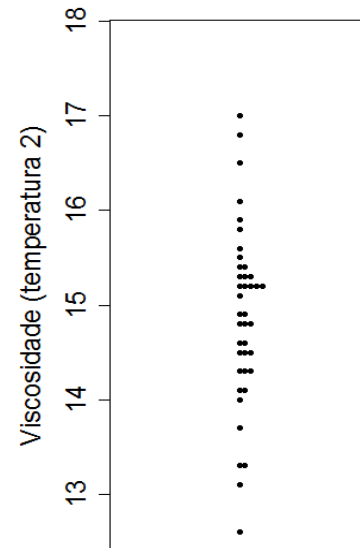
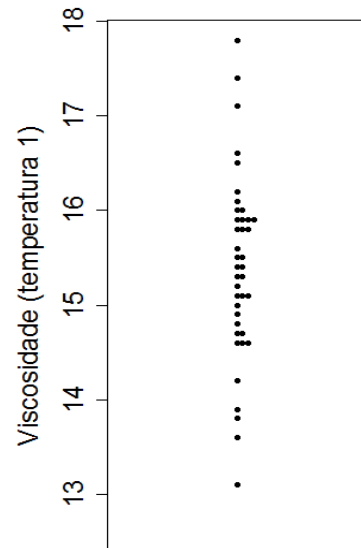
Exemplo. Variável viscosidade medida em duas temperaturas.

Temperatura 1 (lâmina 50).

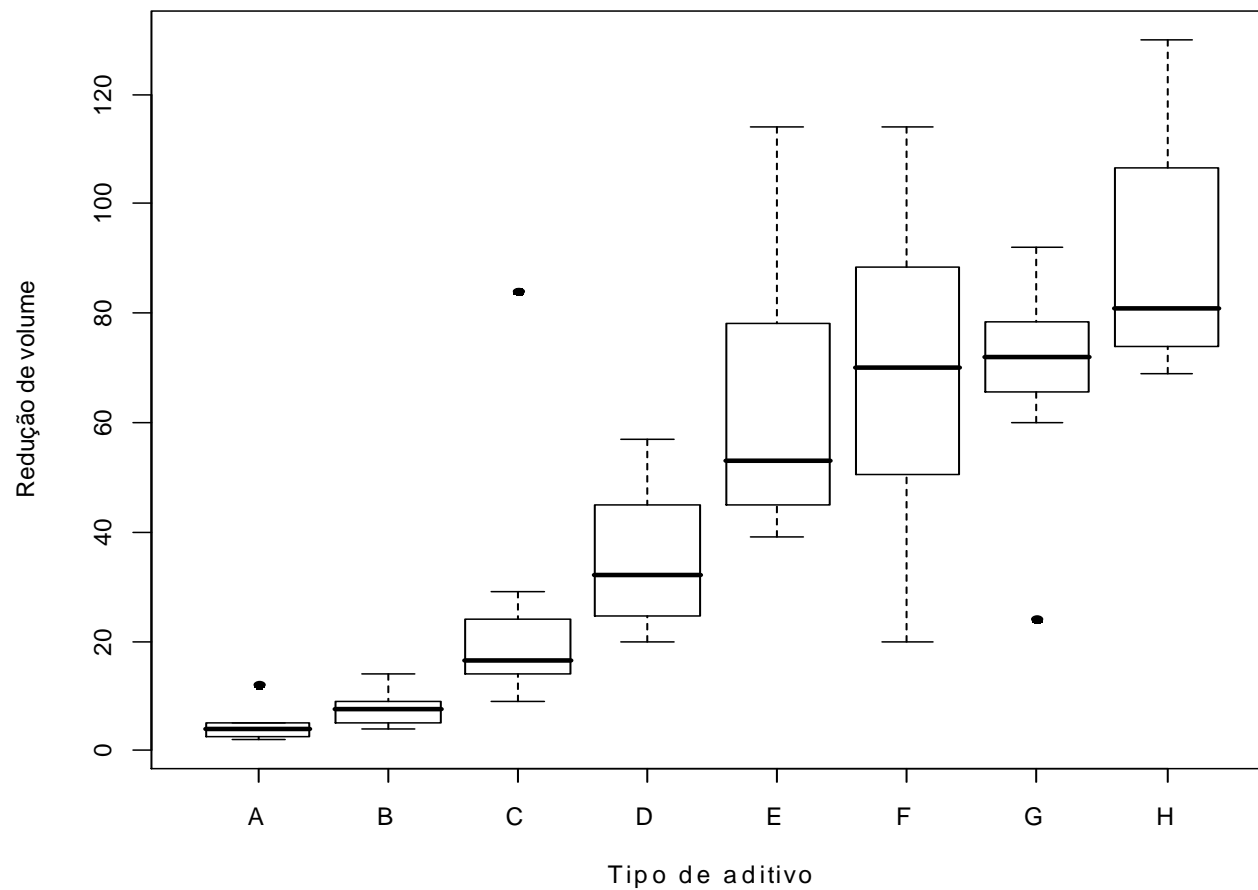
13.9 14.9 15.9 15.8 14.8 15.1 15.8 15.0 15.1 14.6 14.7 16.6 13.6 15.9 13.1
15.2 14.7 16.0 15.6 17.4 15.3 14.2 15.9 15.1 15.9 16.1 16.2 13.8 14.6 16.0
15.8 15.5 16.5 17.1 15.3 15.5 17.8 15.4 15.4 14.6

Temperatura 2 (n = 40).

13.3 14.5 15.3 15.3 14.3 14.8 15.2 14.5 14.6 14.1 14.3 16.1 13.1 15.5 12.6
14.6 14.3 15.4 15.2 16.8 14.9 13.7 15.2 14.5 15.3 15.6 15.8 13.3 14.1 15.4
15.2 15.2 15.9 16.5 14.8 15.1 17.0 14.9 14.8 14.0



Exemplo de *box plot*



Análise exploratória. Redução *versus* tipo. Variabilidade. Simetria. Valores extremos.

Gráfico de linha



O Estado de S. Paulo, 28/2/2010.

Associação entre variáveis quantitativas

$(x_1, y_1), \dots, (x_n, y_n)$: amostra **bivariada**.

Representação gráfica: gráfico de dispersão (*scatter plot*)

Medida de associação: **coeficiente de correlação linear de Pearson**.

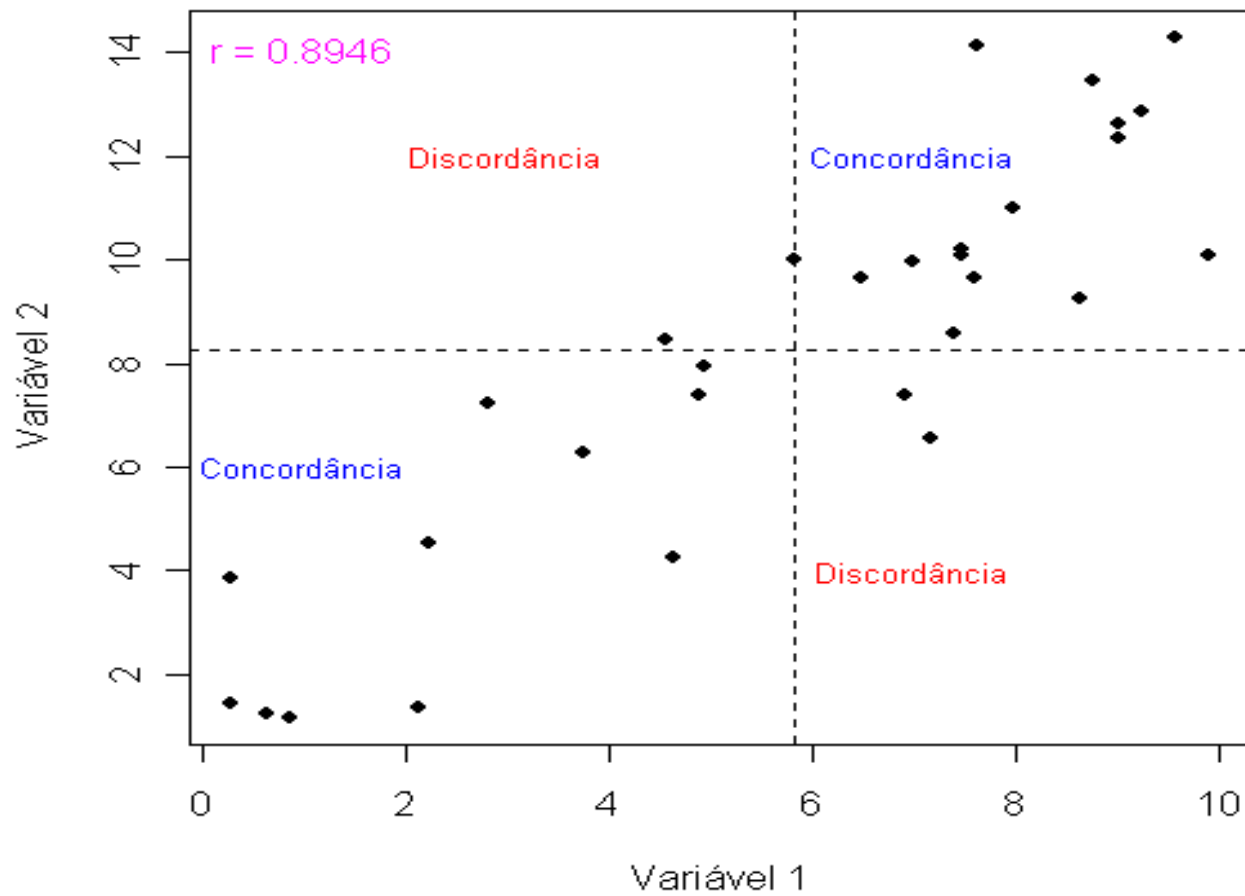
$$r = \frac{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{s_x s_y}$$

Numerador: **covariância** entre x e y.

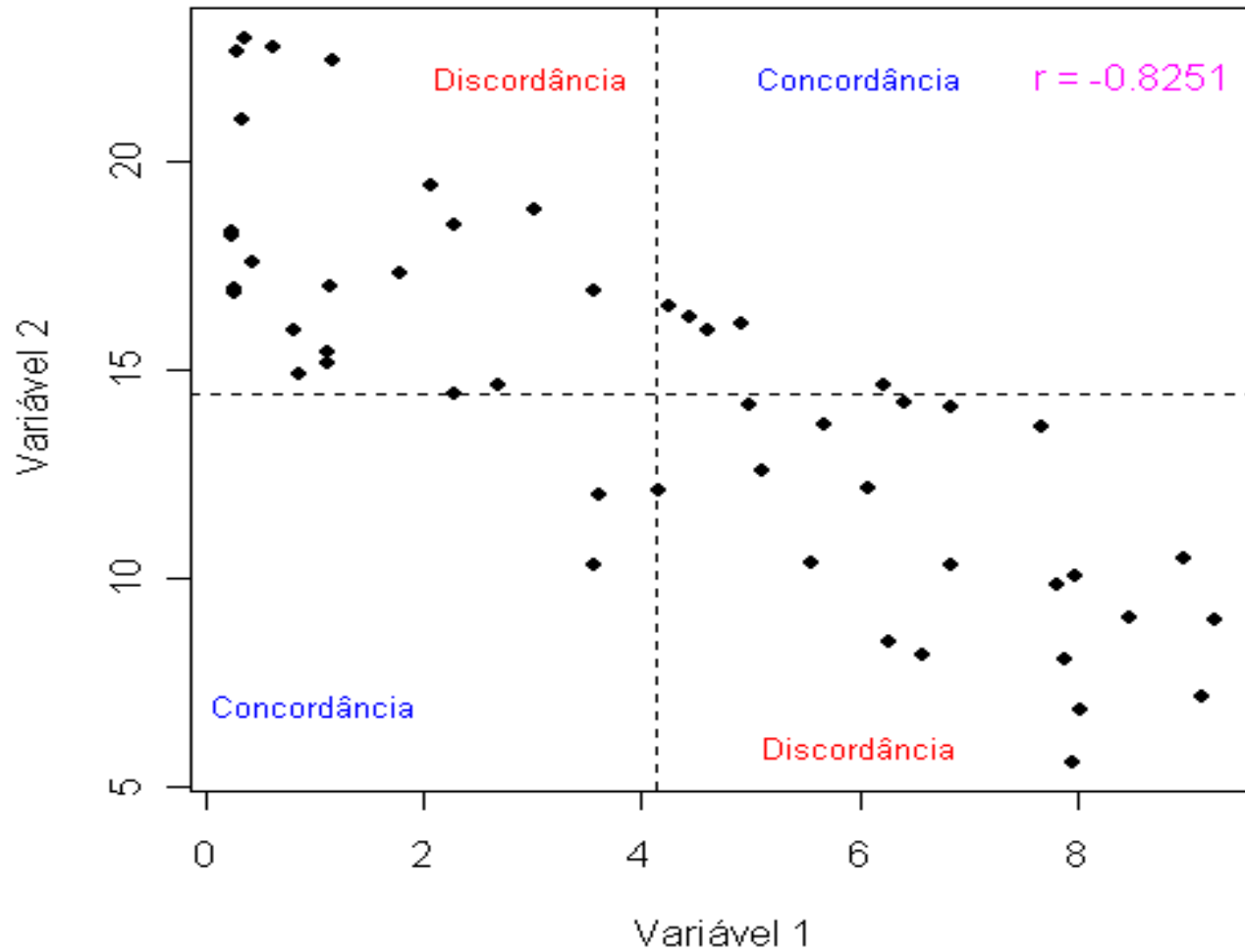
Propriedades: (1) $-1 \leq r \leq 1$ e

(2) $|r| = 1$ se, e somente se, a relação entre x e y for linear ($y = a + bx$, $b \neq 0$ e o sinal de r é o sinal de b).

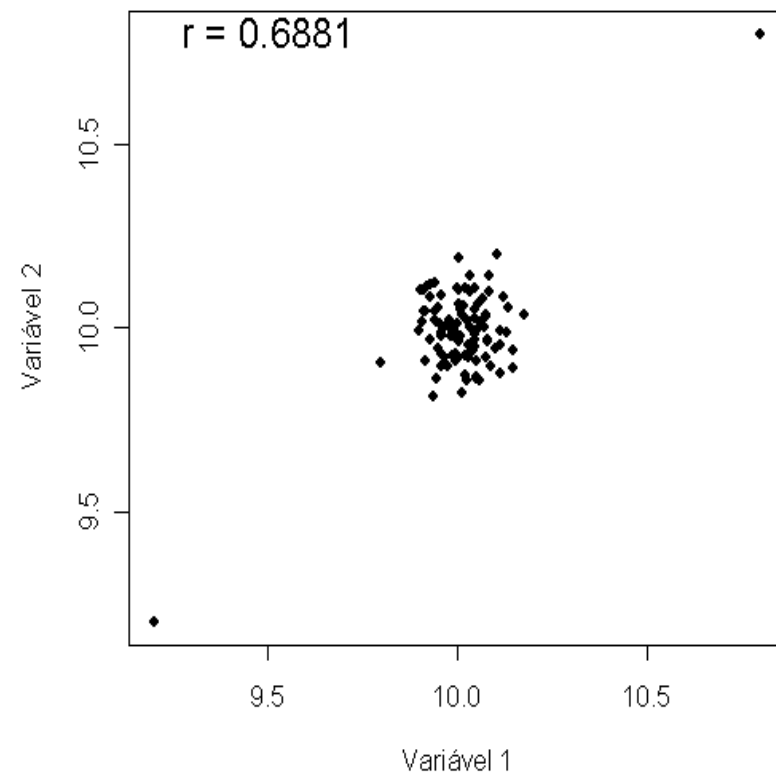
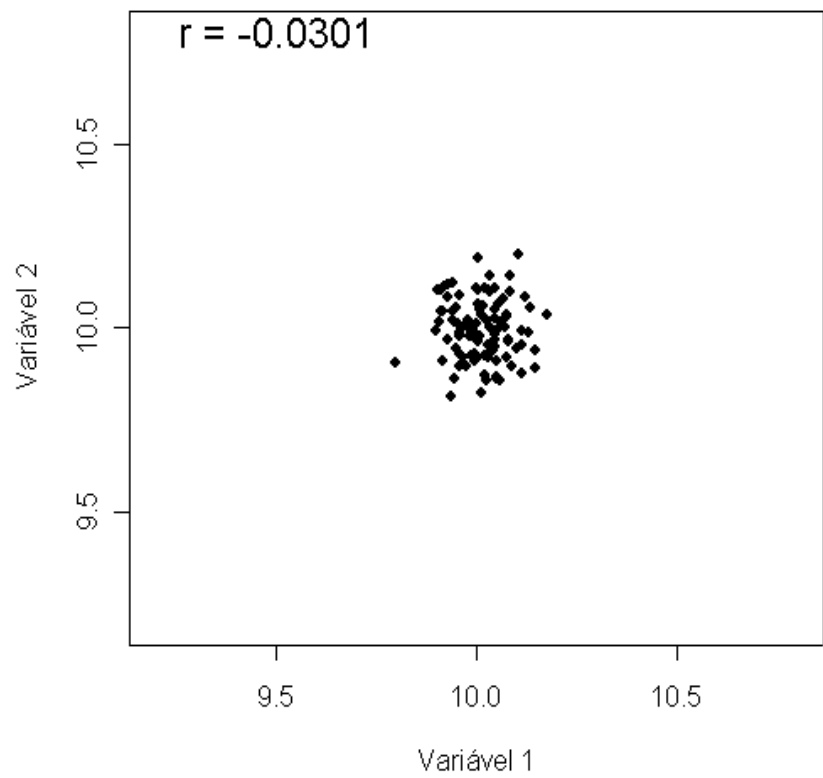
Associação entre variáveis quantitativas



Associação entre variáveis quantitativas

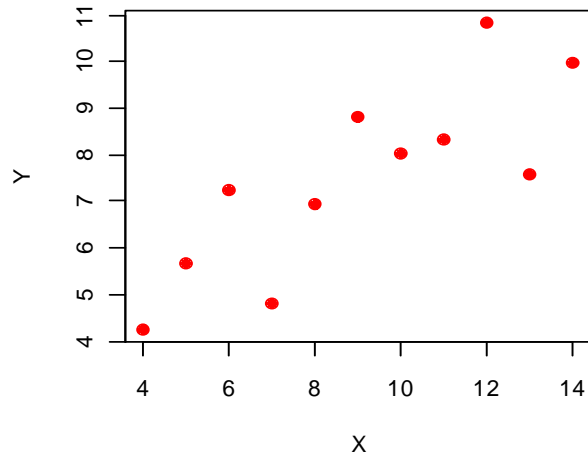


Associação entre variáveis quantitativas

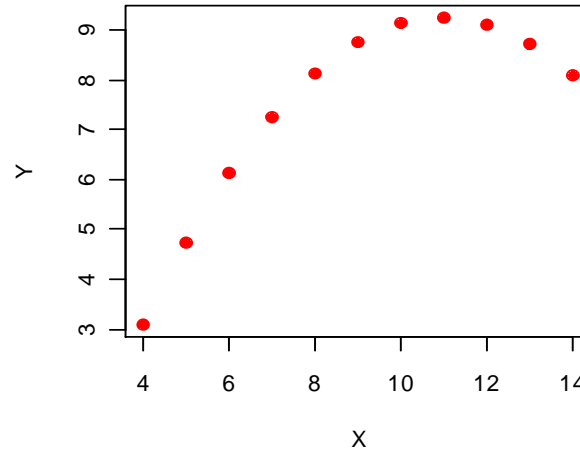


Associação entre variáveis quantitativas

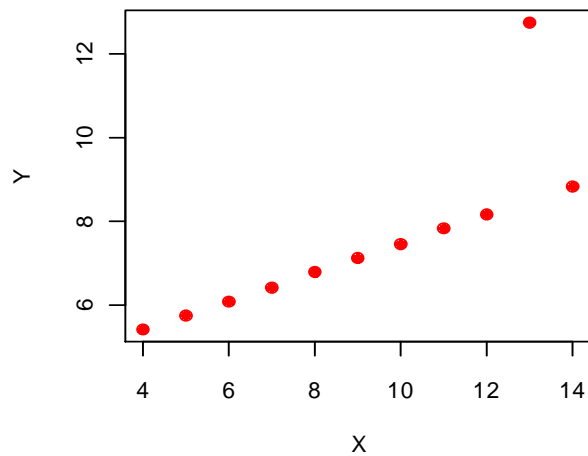
Exemplo 1



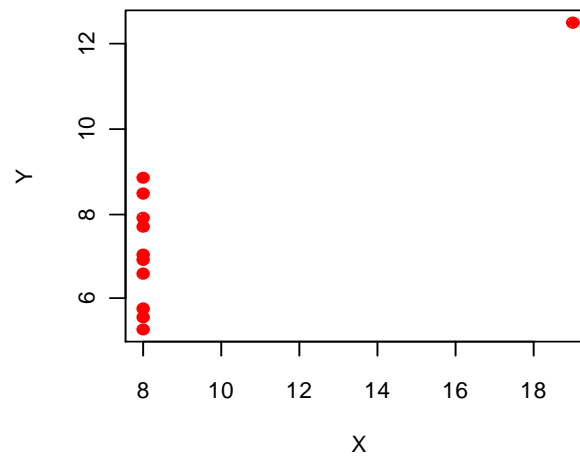
Exemplo 2



Exemplo 3



Exemplo 4



Correlações:

Exemplo 1:
0,8164

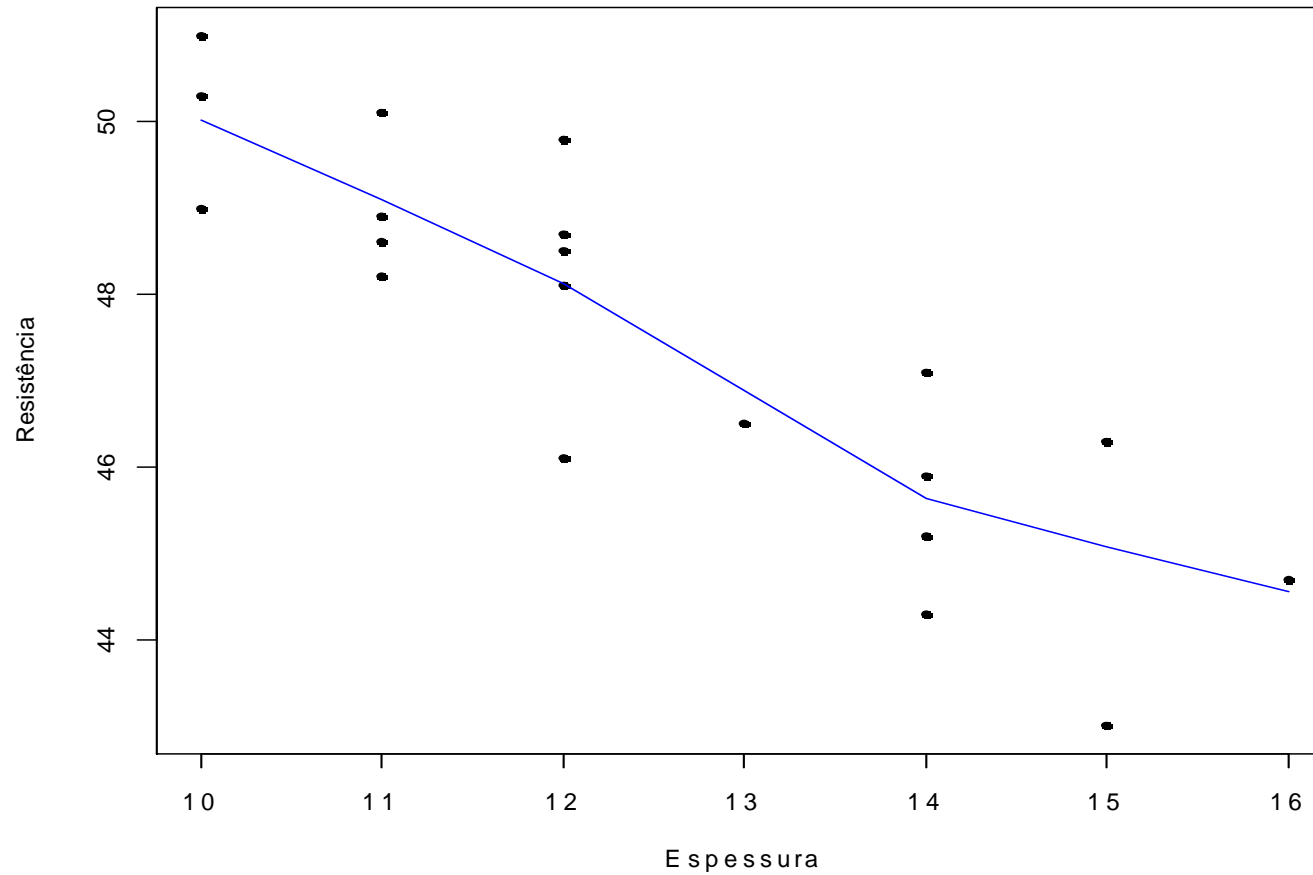
Exemplo 2:
0,8162

Exemplo 3:
0,8163

Exemplo 4:
0,8165

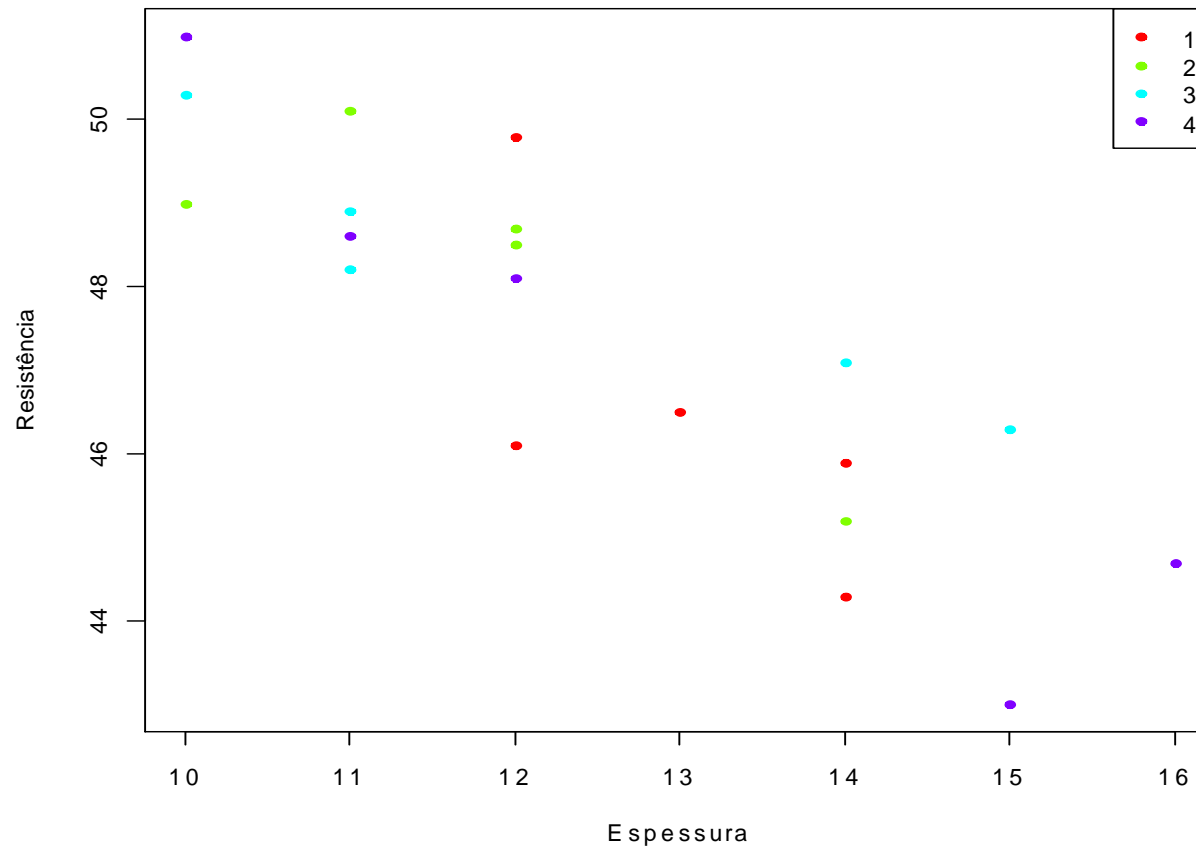
Exemplo em R.

```
> plot(espessura, resistencia, xlab = "Espessura", ylab =  
"Resistência", pch = 20)  
> lines(lowess(espessura, resistencia), col = "blue")
```



Exemplo em R.

```
> cores = rainbow(length(levels(cola)))  
> plot(espessura, resistencia, xlab = "Espessura", ylab =  
"Resistência", pch = 20, col = cores[cola])  
> legend("topright", levels(cola), pch = 20, col = cores)
```



Um exemplo em R

Rendimento (em %) de 90 bateladas de um substrato de cerâmica no qual um revestimento metálico foi aplicado.

```
> dados = scan("dados2-11-Mont.txt")
```

```
Read 90 items
```

```
> summary(dados)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
78.30	86.10	89.25	89.38	93.10	98.00

```
> sd(dados)
```

```
[1] 4.315905
```

```
> quantile(dados, c(0.1, 0.4, 0.7, 0.9))
```

10%	40%	70%	90%
84.10	87.60	91.82	95.21