

2. Representação de variáveis qualitativas

2010

Tabela e quadro

Tabela. Forma de apresentação de dados com destaque para os elementos **numéricos**.

Em geral, é **aberta** nas laterais e **sem** linhas internas.

Elementos: **número** e **título** (O quê? Onde? Quando?), **cabeçalho**, **corpo**, **fontes** e **notas**.

Número e título

Tabela 2 - Consumo intermediário, a preços correntes, das Atividades Características do Turismo, segundo os setores de serviços - Brasil - 2003-2006

Cabeçalho da coluna indicadora

Cabeçalho das colunas numéricas

Setores de serviços	Consumo intermediário, a preços correntes, das Atividades Características do Turismo (1 000 000 R\$)			
	2003	2004	2005	2006
Total	60 195	61 728	68 148	75 774
Serviços de alojamento	2 729	2 844	3 251	3 797
Serviços de alimentação	25 775	26 579	29 212	34 424
Transporte ferroviário	20	20	23	23
Transporte rodoviário	9 514	10 356	11 143	12 211
Transporte aquaviário	118	177	180	150
Transporte aéreo	11 937	10 769	11 650	10 365
Serviços auxiliares dos transportes	3 427	4 191	4 429	4 990
Atividades de agências e organizadores de viagens	1 196	1 312	1 619	1 341
Aluguel de bens móveis	747	791	959	1 129
Atividades recreativas, culturais e desportivas	4 732	4 689	5 682	7 344

Fonte: IBGE, Diretoria de Pesquisas, Coordenação de Contas Nacionais. **Fonte**

Tabela e quadro

Quadro. Forma de apresentação de dados com destaque para os elementos **textuais**.

Em geral, é **fechado** nas laterais e **com** linhas internas.

Quadro 1. Classificação das pesquisas.

1. Quanto à Natureza das Variáveis	Qualitativas	Quantitativas	
2. Quanto ao Objetivo e Grau do Problema	Exploratória	Descritiva	Causal
3. Quanto ao Escopo (Amplitude e Profundidade)	Estudo de Caso	Estudo de Campo	Levantamento Amostral
5. Quanto ao Controle	Laboratório	Experimento de Campo	

2.1 Tabelas de frequencias

Dados: n observações de uma variável qualitativa x .

Existem k diferentes valores x_1, x_2, \dots, x_k , $1 \leq k \leq n$. Cada um deles ocorre f_j vezes, $j = 1, \dots, k$.

f_j é a **frequencia absoluta** (ou **frequencia**) de x_j , $j = 1, \dots, k$.

Frequencia relativa de x_j : $f_j^* = \frac{f_j}{n}$, $j = 1, \dots, k$.

A frequencia relativa pode ser expressa em %.

$$\sum_{j=1}^k f_j = n \quad \text{e} \quad \sum_{j=1}^k f_j^* = 1 \quad (100\%).$$

Tabela de frequencias: tabela com os valores de x_j e uma das ou ambas as frequencias f_j e f_j^* , $j = 1, \dots, k$.

Tabela 1. Descrição da tabela.

x	Frequencia	Frequencia relativa
x_1	f_1	f_1^*
x_2	f_2	f_2^*
...
x_k	f_k	f_k^*
Total	n	1 (100%)

Se a variável for nominal, a ordem das linhas é arbitrária.

As linhas podem ser ordenadas pelas frequencias.

Tabelas em R

```
> x
```

```
"SE" "SE" "NE" "NE" "S" "CO" "S" "NE" "SE" "SE" "SE" "NE" "SE" "SE" "NE"  
"SE" "SE" "SE" "NE" "NE" "S" "SE" "SE" "NE" "SE" "SE" "S" "S" "SE" "SE"  
"SE" "NE" "SE" "SE" "SE" "N" "S" "S" "NE" "NE" "S" "CO" "S" "N" "CO"  
"NE" "S" "SE" "S" "CO" "NE" "NE" "SE" "NE" "SE" "NE" "SE" "NE" "SE" "SE" "N"  
"S" "SE" "S" "CO" "S" "SE" "S" "SE" "SE" "SE" "NE" "NE" "SE" "NE" "S"  
"S" "S" "NE" "SE"
```

```
> (n = length(x))
```

```
80
```

```
> table(x)
```

```
CO  N  NE  S  SE      Valores de x em ordem alfabética.
```

```
5  3  21  18  33
```

```
> table(x) / n
```

```
CO      N      NE      S      SE
```

```
0.0625 0.0375 0.2625 0.2250 0.4125
```

```
> table(x) / n * 100
```

```
CO      N      NE      S      SE
```

```
6.25  3.75 26.25 22.50 41.25
```

Tabelas em R

```
> sort(table(x) / n) * 100
```

```
   N    CO    S    NE    SE
3.75 6.25 22.50 26.25 41.25
```

```
> freqa = table(x)
```

```
> freqr = table(x) / n * 100
```

```
> sort(freqr, decreasing = TRUE)
```

```
   SE    NE    S    CO    N
41.25 26.25 22.50  6.25  3.75
```

```
> freqr[freqr > 10]
```

```
   NE    S    SE
26.25 22.50 41.25
```

```
> sum(freqr[c("CO", "SE")])
```

```
[1] 47.5
```

```
> which.max(freqr)
```

```
SE
```

```
5
```

```
> sort(table(x), decreasing = TRUE) / n * 100
```

```
   SE    NE    S    CO    N
41.25 26.25 22.50  6.25  3.75
```

```
> names(freqr)
```

```
[1] "CO" "N"  "NE" "S"  "SE"
```

```
> freqr[2]
```

```
   N
3.75
```

```
> freqr["NE"]
```

```
26.25
```

```
> freqr[setdiff(names(freqr), c("N", "S"))]
```

```
   CO    NE    SE
6.25 26.25 41.25
```

```
> freqr[which.max(freqr)]
```

```
41.25
```

2.2 Gráficos

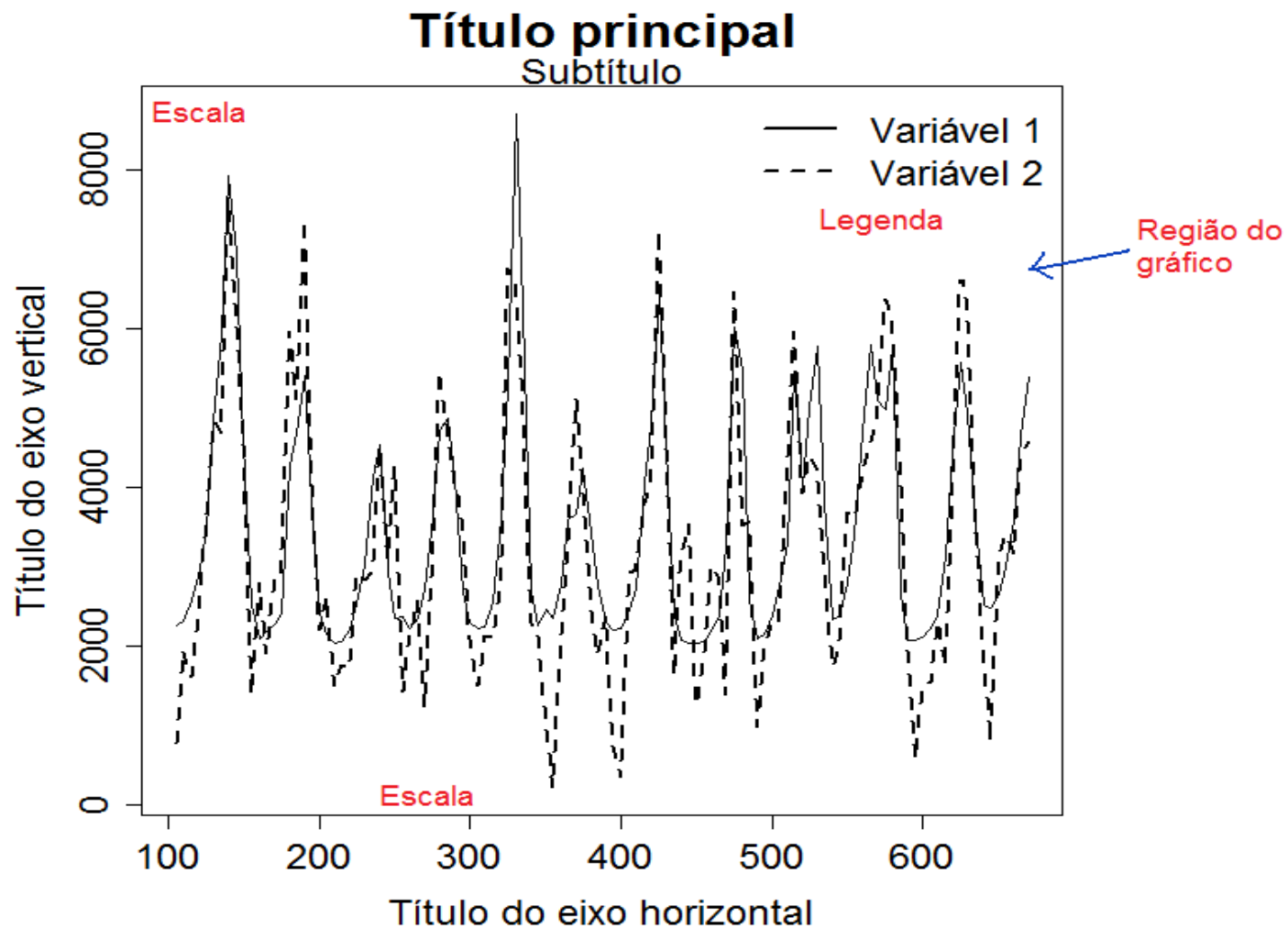


Figura 1. Descrição do gráfico.

Gráfico de barras (*bar plot*)

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

```
> freqa = table(x)
```

```
> freqr = sort(table(x)) / n * 100
```

```
> barplot(freqa)
```

```
> posb = barplot(freqr, xlab = "Região de origem", ylab = "Porcentagem")
```

```
> text(posb, freqr / 2, freqr)
```

```
> box()
```

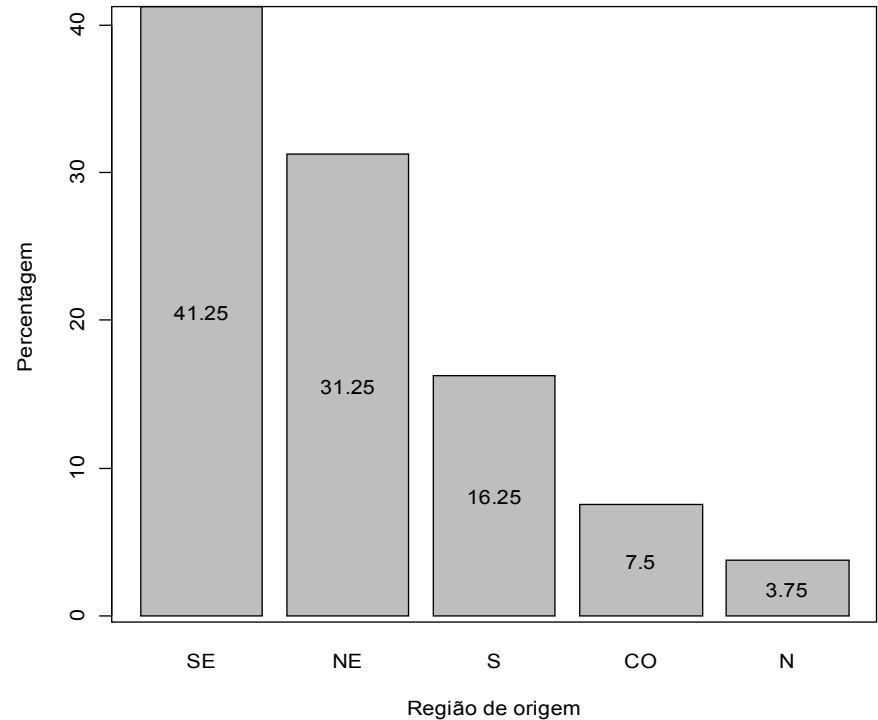
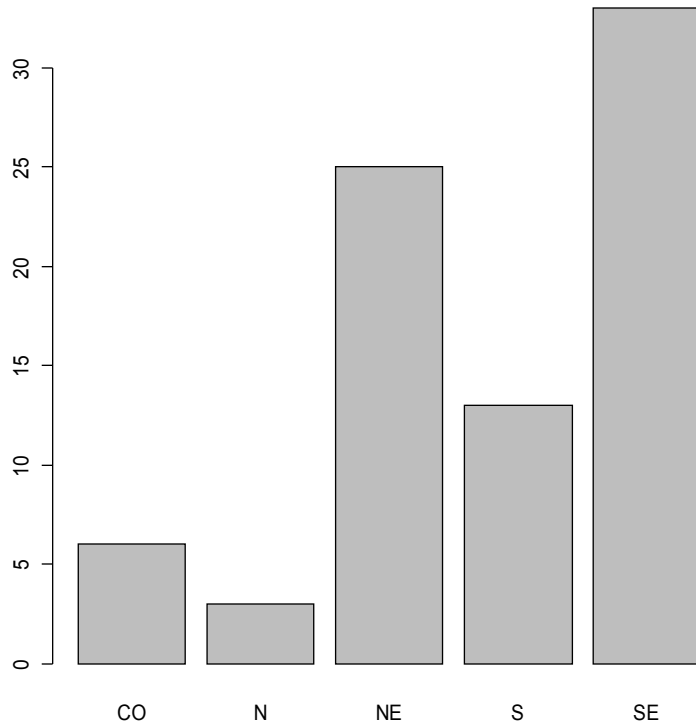
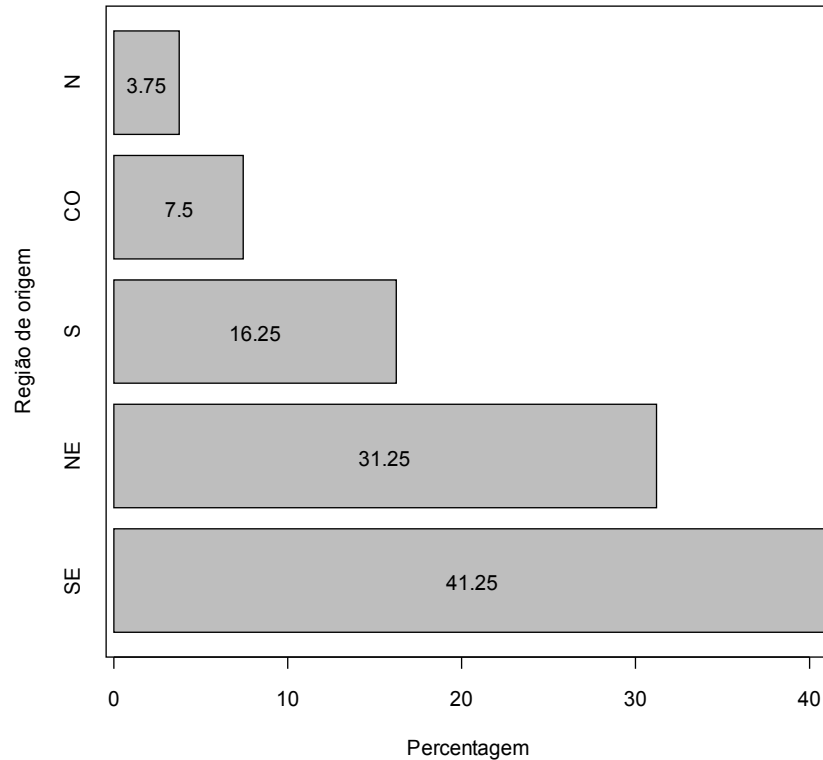


Gráfico de barras

```
> posb = barplot(freqr, ylab = "Região de origem", xlab = "Percentagem",  
horiz = TRUE)
```

```
> text(freqr / 2, posb, freqr)
```

```
> box()
```



```
> posb = barplot(freqr, xlab = "Região de origem", ylab = "Percentagem",  
angle = 45, density = 15, col = "black")
```

```
> text(posb, freqr / 2, freqr, col = "red")
```

```
> box()
```

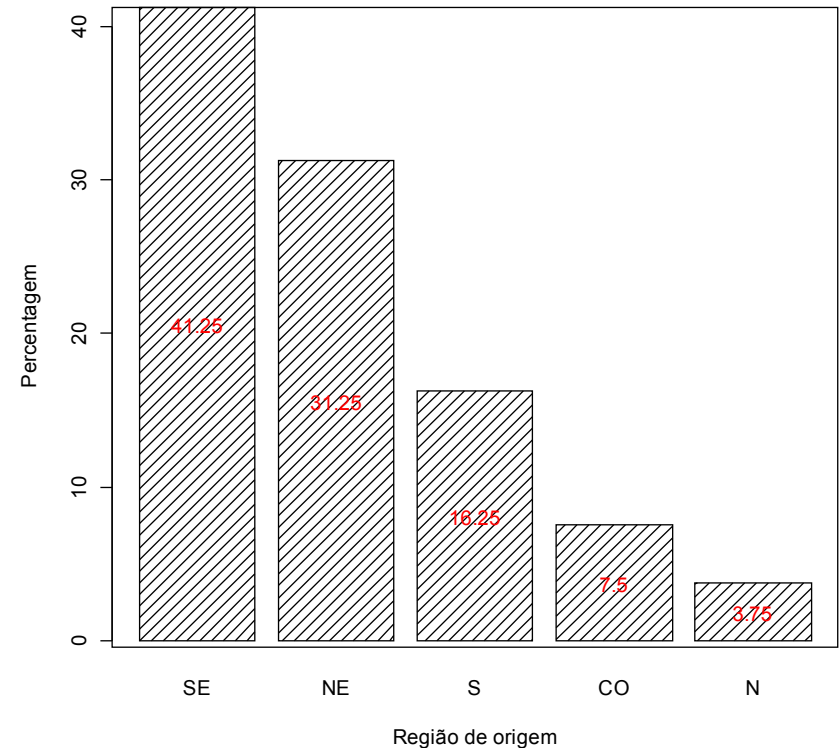
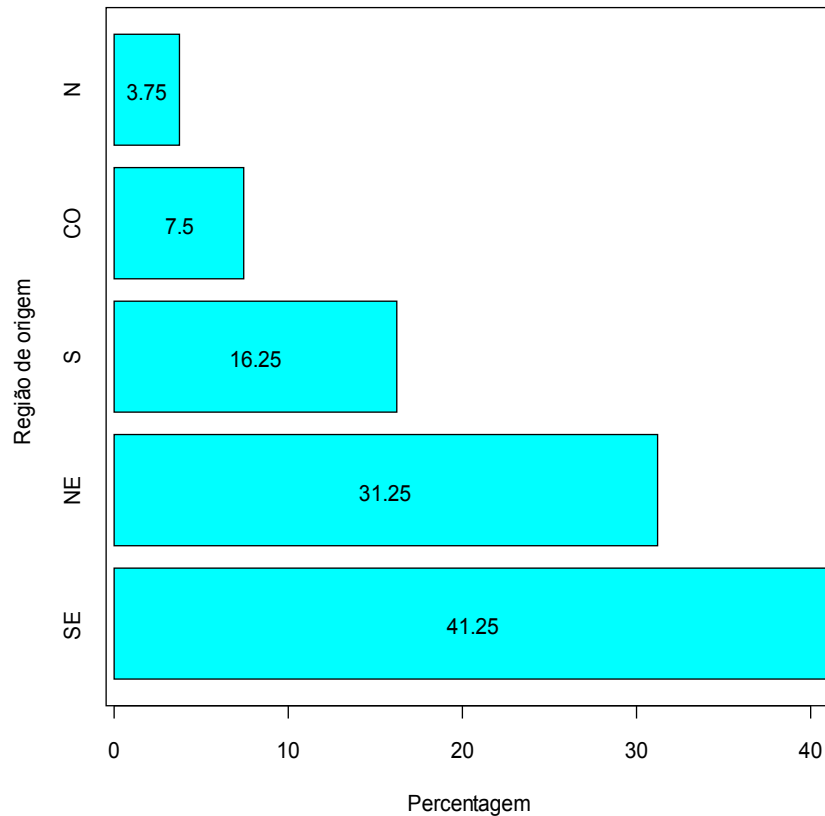


Gráfico de barras

```
> posb = barplot(freqr, ylab = "Região de  
origem", xlab = "Percentagem", horiz =  
TRUE, col = "cyan")
```

```
> text(freqr / 2, posb, freqr)
```

```
> box()
```



```
> mcores = rainbow(length(freqr))
```

```
> posb = barplot(freqr, xlab = "Região  
de origem", ylab = "Percentagem",  
col = mcores)
```

```
> text(posb, freqr / 2, freqr)
```

```
> box()
```

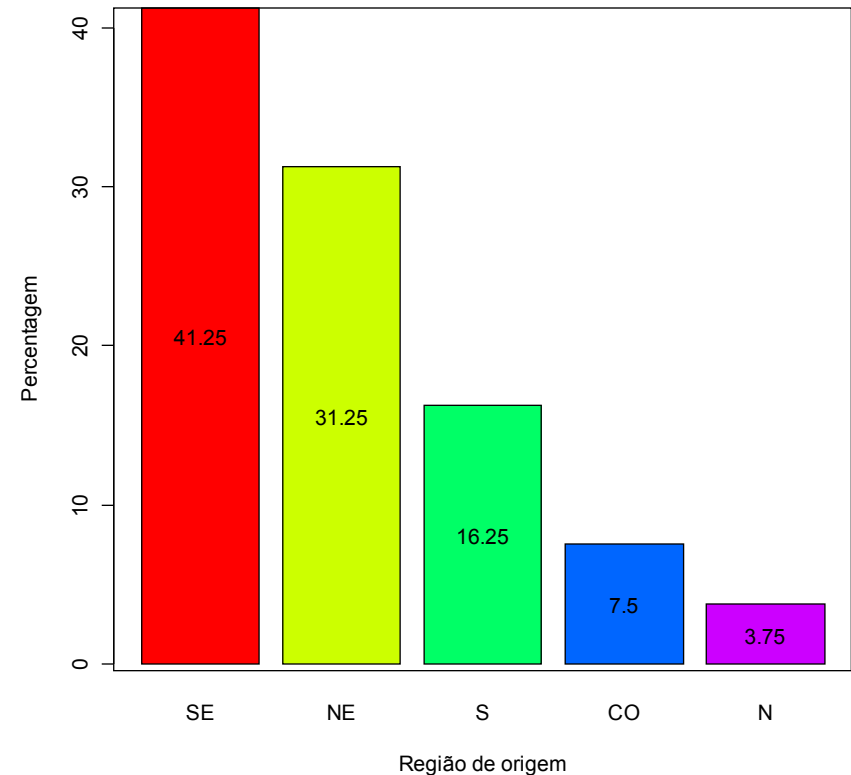


Gráfico de Pareto (*Pareto plot* ou *Pareto chart*)

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

Frequencia acumulada: $F_1 = f_1, F_2 = F_1 + f_2, \dots, F_k = n.$

Frequencia relativa acumulada: $F_1^* = f_1^*, F_2^* = F_1^* + f_2^*, \dots, F_k^* = 1.$

Frequências relativas acumuladas podem ser expressas em %.

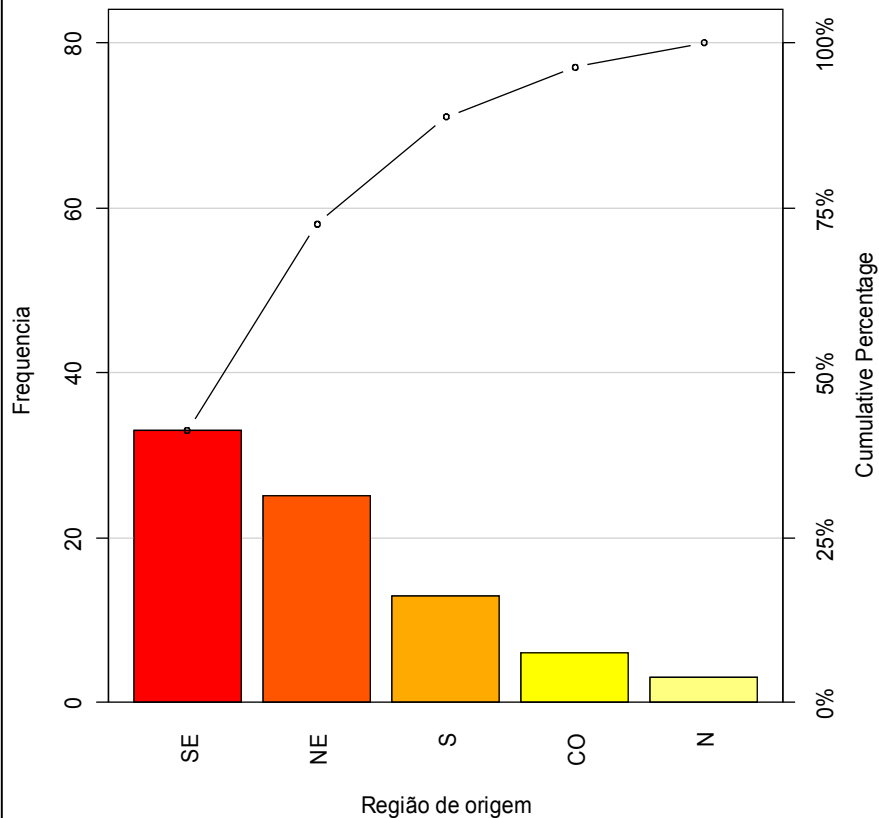
Obs. Se a variável for nominal, a ordem dos valores é arbitrária e deve ter interpretação.

No gráfico de Pareto os valores são ordenados pelas suas frequências.

Gráfico de Pareto

```
> library(qcc)
```

```
> pareto.chart(freqa, main = "", xlab =  
"Região de origem", ylab = "Frequencia")
```



```
> pareto.chart(freqa, main = "", xlab =  
"Região de origem", ylab = "Frequencia",  
col = "lightgray")
```

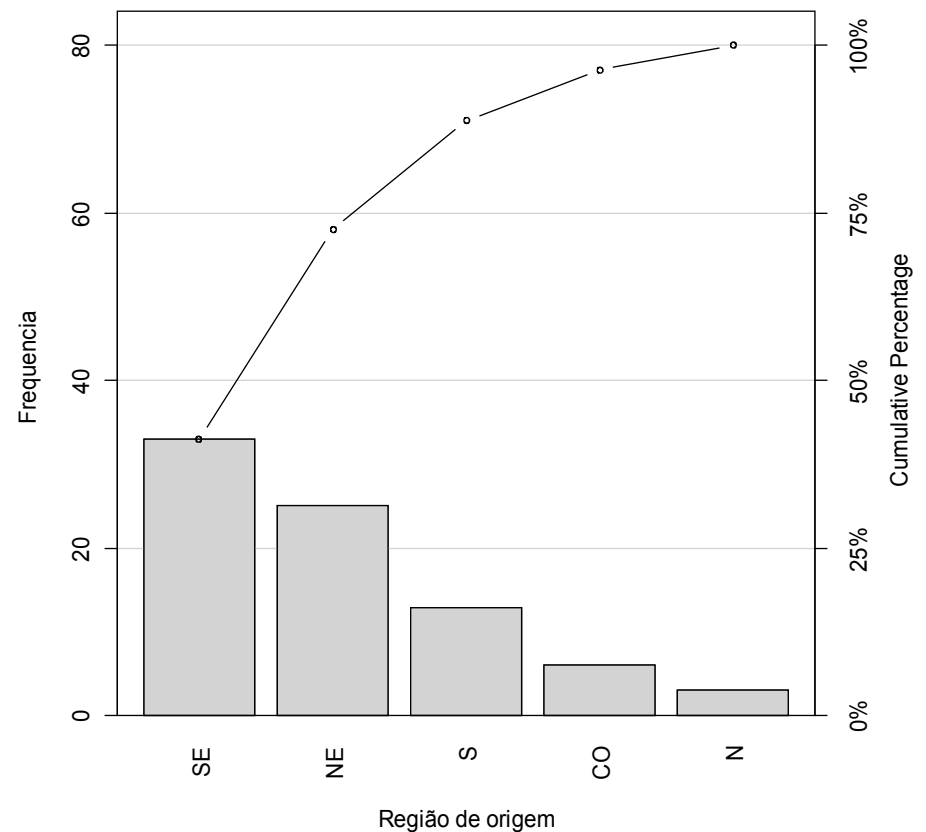
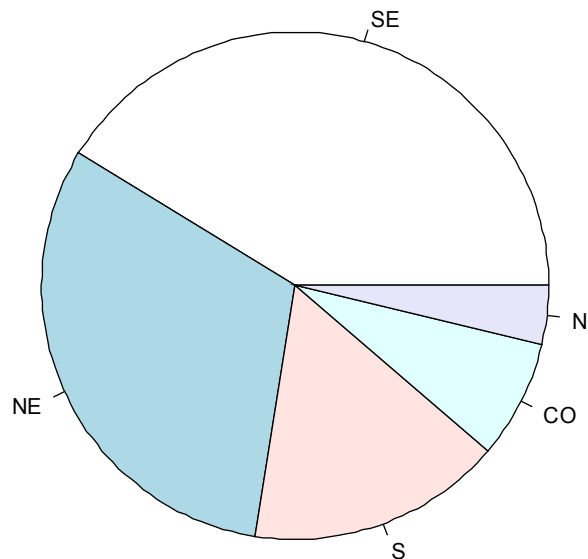


Gráfico de setores (*pie chart*)

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 %).

> `pie(freqr)`



> `pie(freqr, col = mcores)`

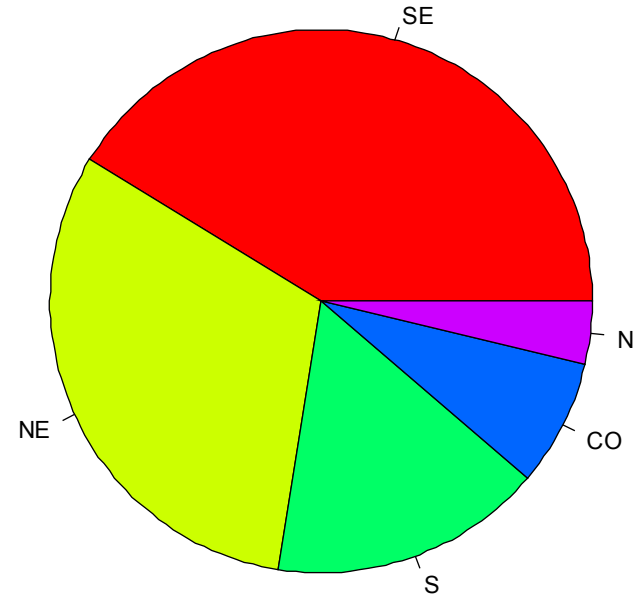
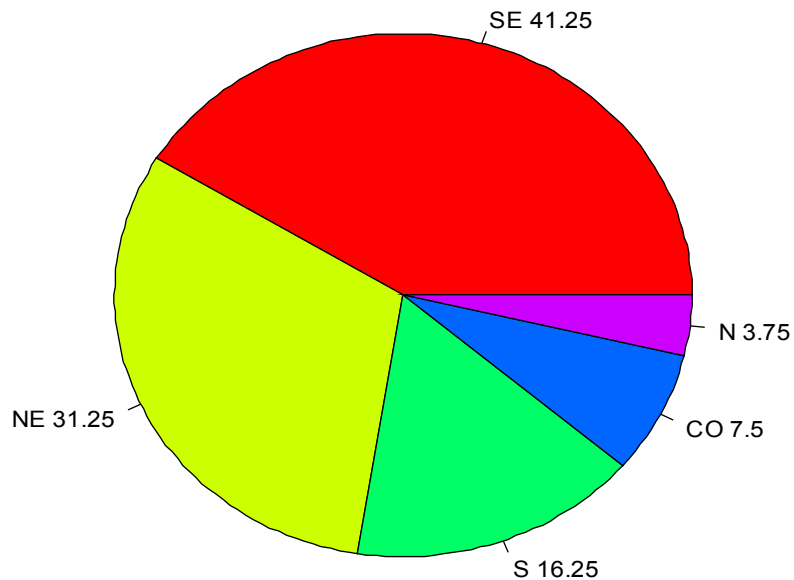


Gráfico de setores

```
> pie(freqr, col = mcores, labels  
= paste(names(freqr), freqr))
```



```
> pie(freqr, labels =  
paste(names(freqr), freqr), col =  
gray(seq(0.4, 1.0, length =  
length(freqr))))
```

