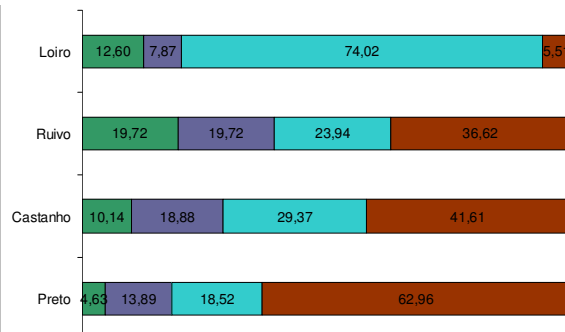


# Análise Descritiva de Dados

Parte I: Organizando e apresentando os dados de variáveis qualitativas



Tabelas e Gráficos

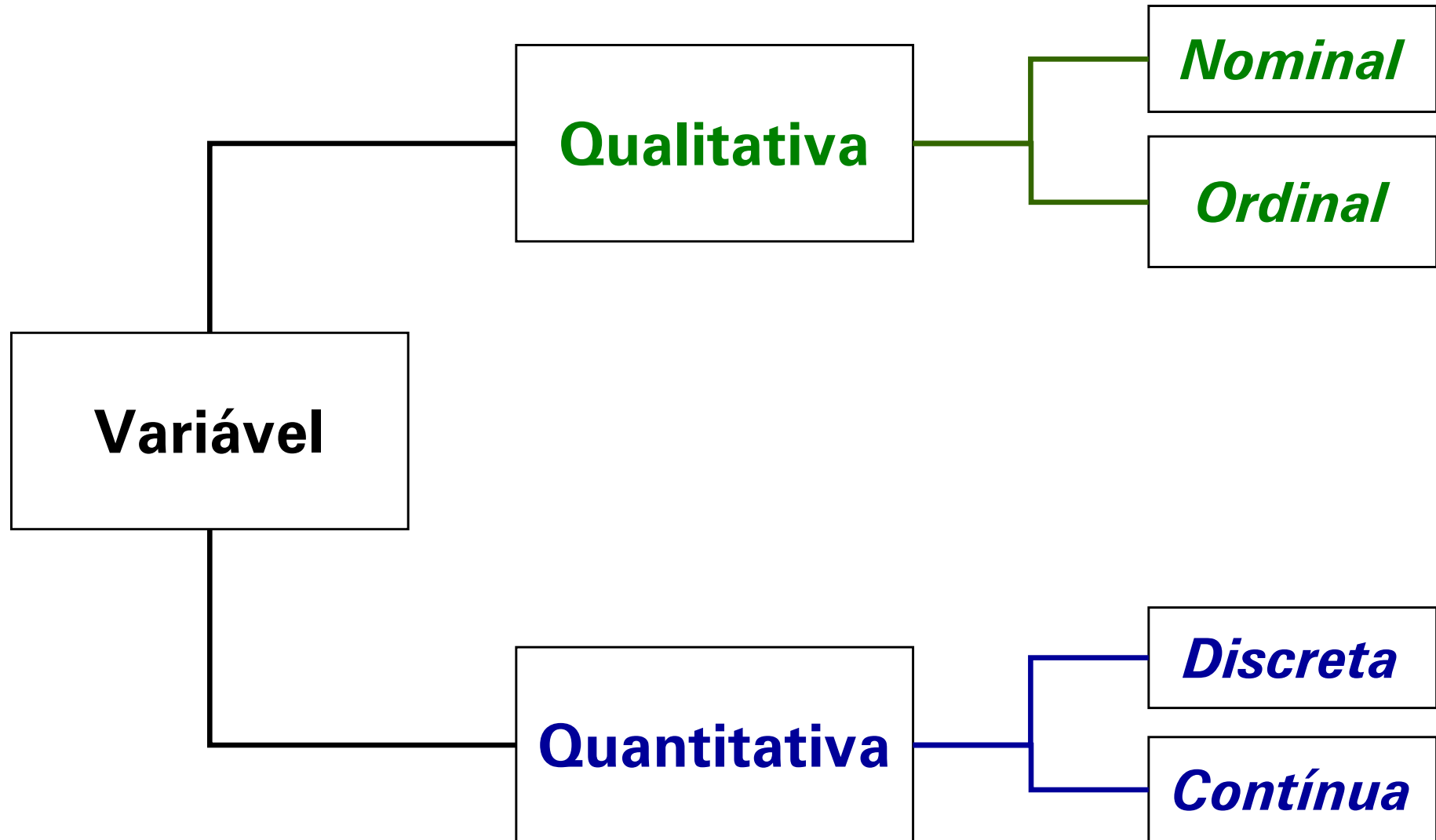
## Coletando os dados

PROJETO CICLAME	
• Identificação da planta:	ID-01
• Variedade:	<input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4
• Temperatura Dia:	<input type="checkbox"/> 14 °C <input type="checkbox"/> 16 °C <input type="checkbox"/> 18 °C <input type="checkbox"/> 20 °C <input type="checkbox"/> 26 °C
• Temperatura Noite:	<input type="checkbox"/> 14 °C <input type="checkbox"/> 16 °C <input type="checkbox"/> 18 °C <input type="checkbox"/> 20 °C <input type="checkbox"/> 26 °C
• Tipo de fertilizante:	<input type="checkbox"/> A <input type="checkbox"/> B <input type="checkbox"/> C <input type="checkbox"/> D
• Número de flores:	_____
• Área das folhas:	_____ cm <sup>2</sup>
• Ataque de pragas:	<input type="checkbox"/> Sim <input type="checkbox"/> Não
• Grau de coloração:	<input type="checkbox"/> fraco <input type="checkbox"/> médio <input type="checkbox"/> forte
• Data da coleta :	__/__/__
• Nome do funcionário responsável pela coleta:	_____

# Organizando os dados

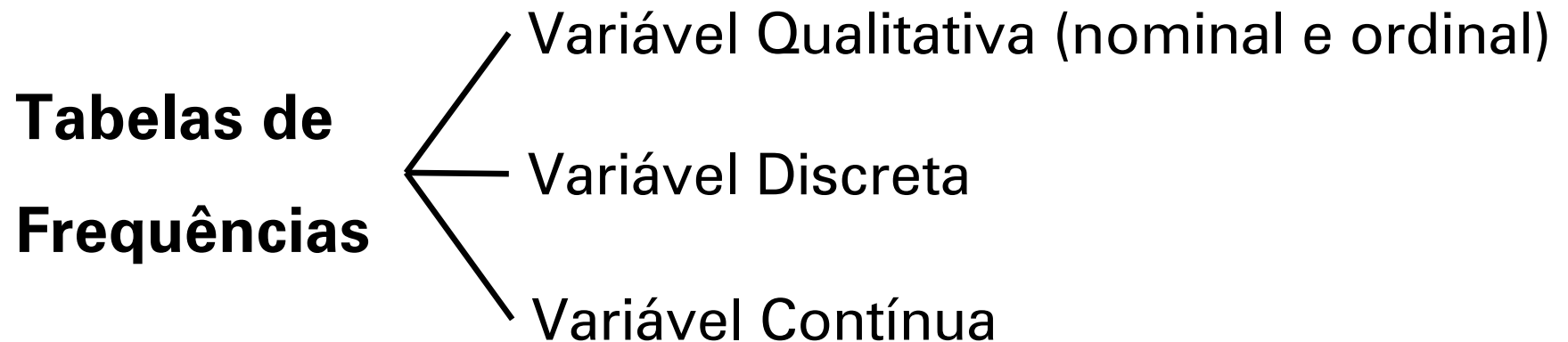
VARIÁVEIS →											
ELEMENTOS ↓		ID	Variedade	Temp. Dia	Temp. Noite	Fertilizante	Flores Número	Folhas Área	Pragas	Flores Coloração	Data
	1	01	1	14	14	A	12	14,3	Sim	Forte	03/03/2009
	2	02	1	14	14	A	14	12,2	Sim	Médio	03/03/2009
	3	03	1	14	14	A	10	18,4	Sim	Forte	05/03/2009
	4	04	1	20	14	B	4	10,1	Não	Fraço	05/03/2009
	5	05	1	20	14	B	9	11,4	Sim	Forte	04/03/2009
	6	06	2	20	14	B	10	12,4	Sim	Médio	04/03/2009
	...										
	1916	1916	3	14	20	C	11	13,4	Sim	Fraço	05/03/2009
	1917	1917	3	14	20	C	13	12,0	Sim	Fraço	03/03/2009
	1918	1918	4	20	20	D	8	14,9	Não	Médio	03/03/2009
	1919	1919	4	20	20	D	9	15,1	Não	Forte	03/03/2009
1920	1920	4	20	20	D	12	9,5	Não	Forte	04/03/2009	

# Classificando as variáveis



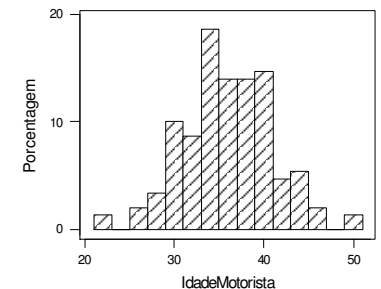
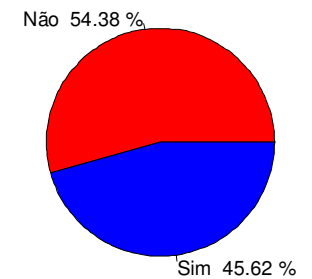
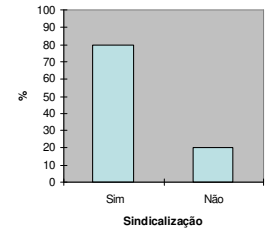
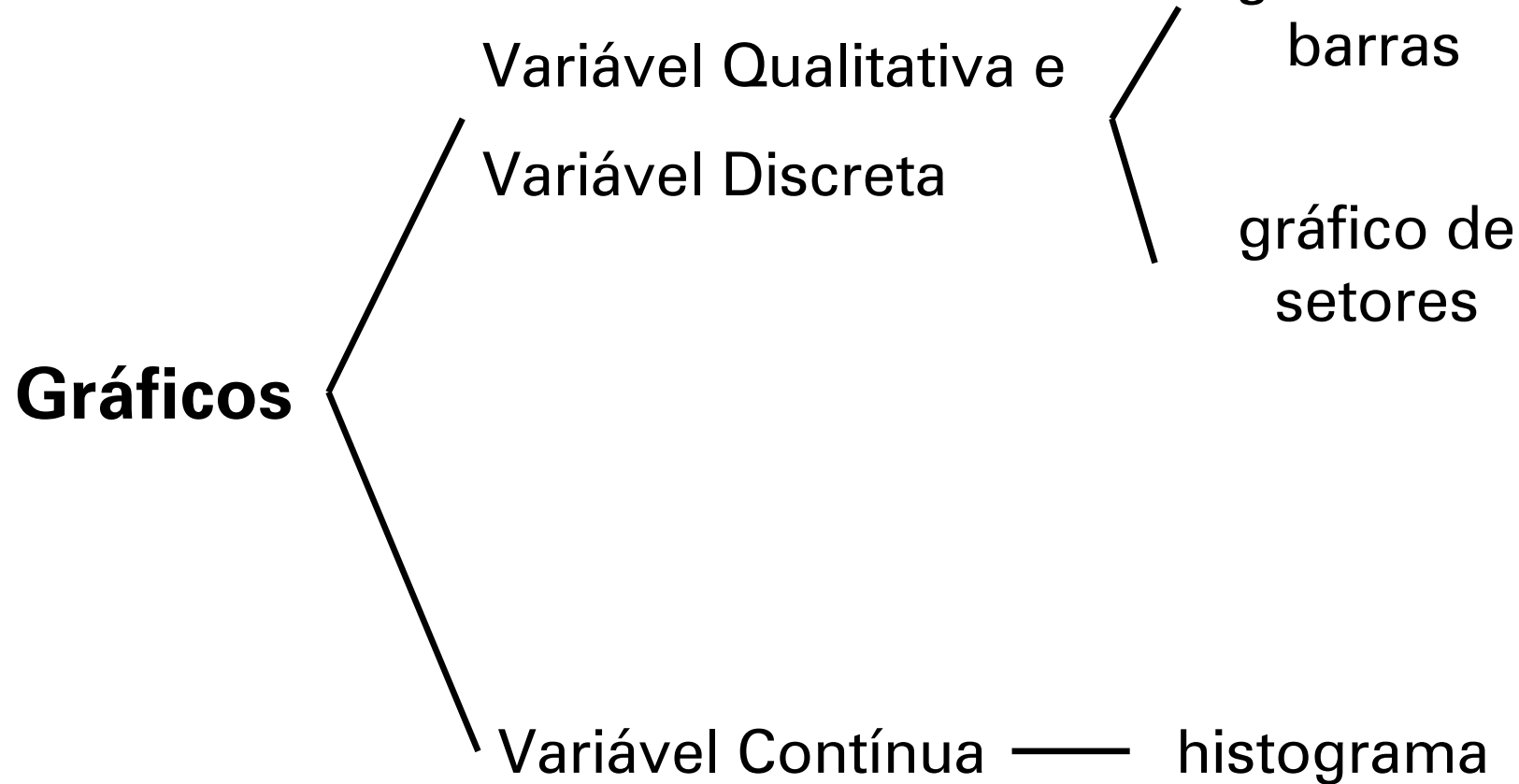
# Organização e Apresentação dos Dados

A forma de apresentação dos dados depende do tipo da variável que os gerou.



**Distribuição de Frequências**

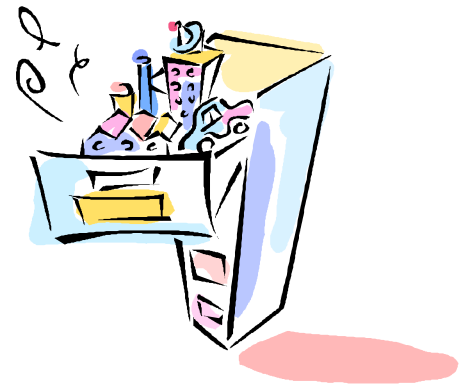
# Organização e Apresentação dos Dados



## Distribuição de Frequências

# Estudando a Distribuição de Frequências

Variáveis Qualitativas Nominais



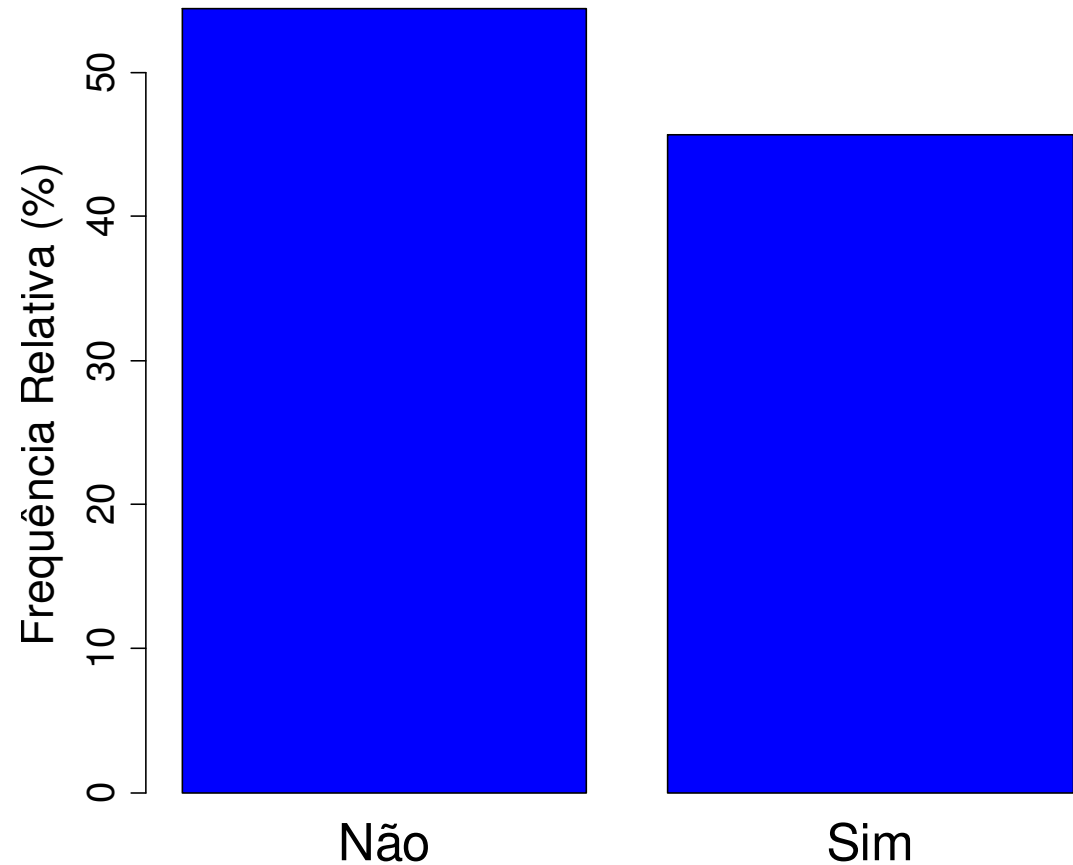
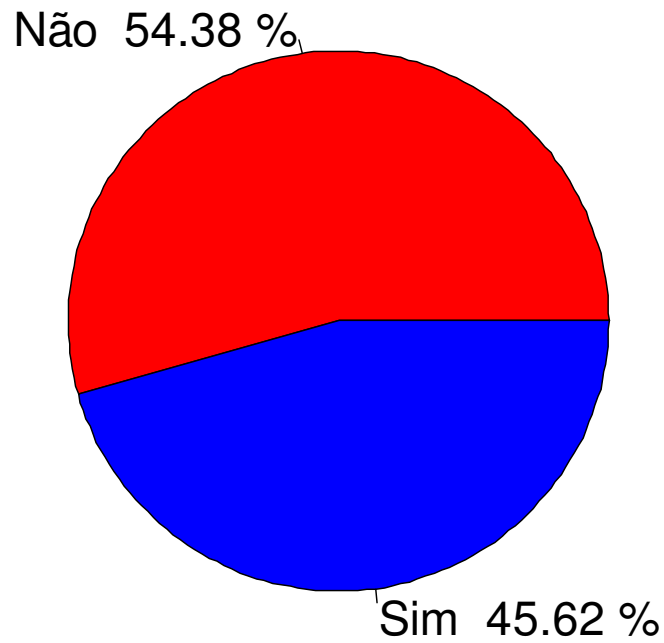
# Tabelas de Frequências para Variáveis Qualitativas Nominiais

**Tabela de distribuição de frequências das plantas da variedade 2 do ciclame segundo a ocorrência de praga**

Ocorrência de Pragas	Frequência Absoluta	Frequência Relativa (%)
Sim	219	45.63
Não	261	54.37
Total	480	100



# Gráficos para visualizar a Distribuição de Frequências de Variáveis Qualitativas Nominais

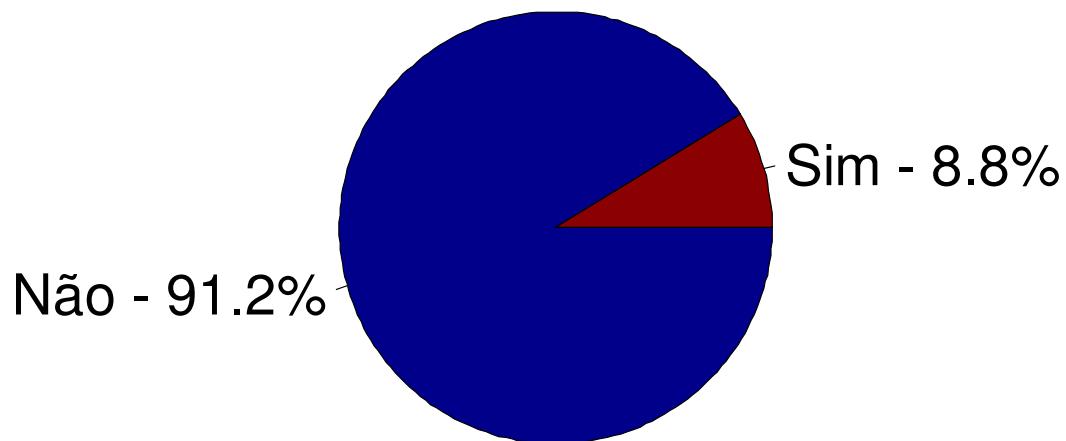
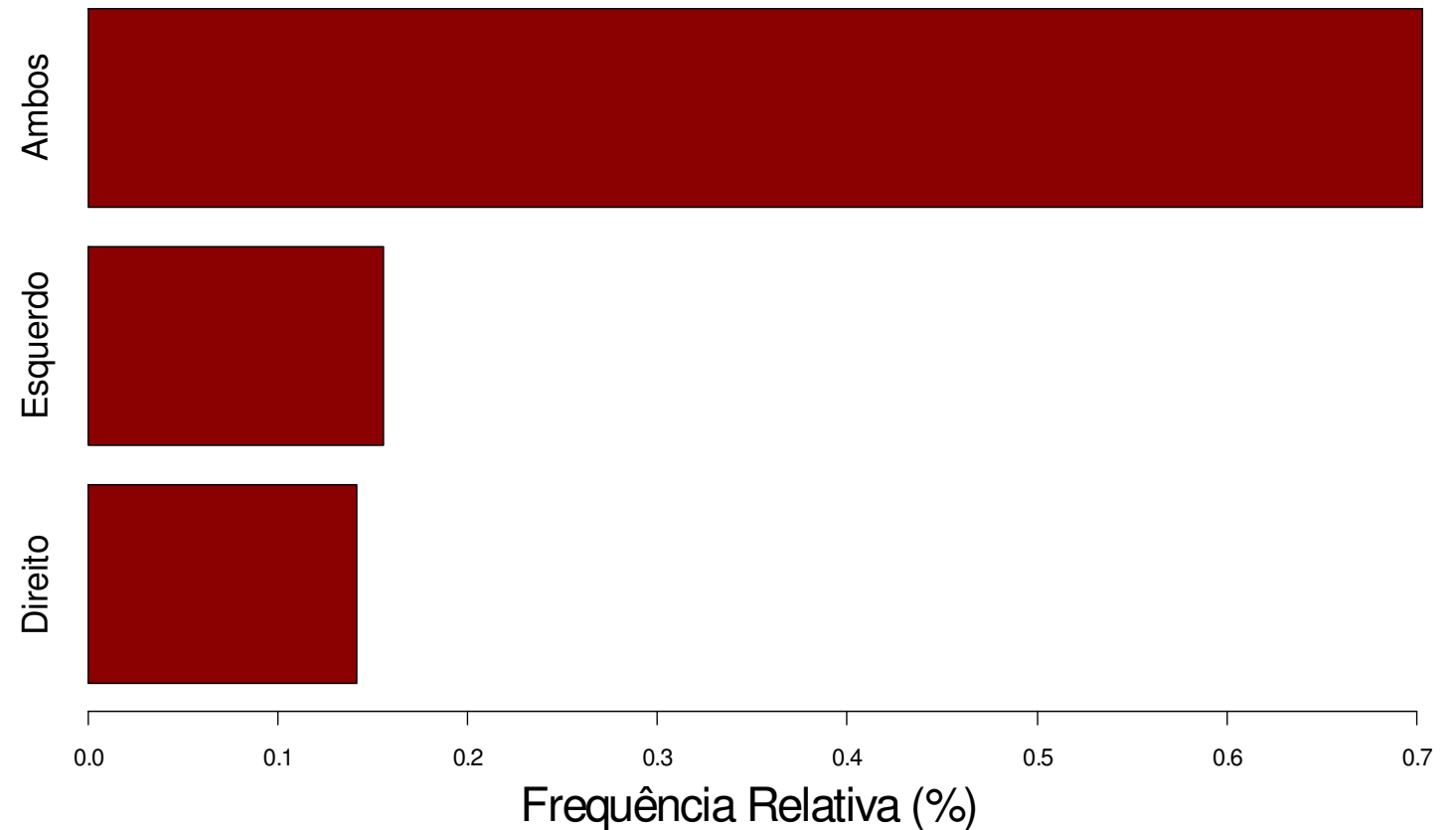


**Gráfico de Setores / Gráfico de Barras**

**Exemplo 2:**  
Pesquisa sobre  
Saúde e  
Nutrição na  
população  
americana  
(NHANES  
2007-2008)

[CDC, 2008]

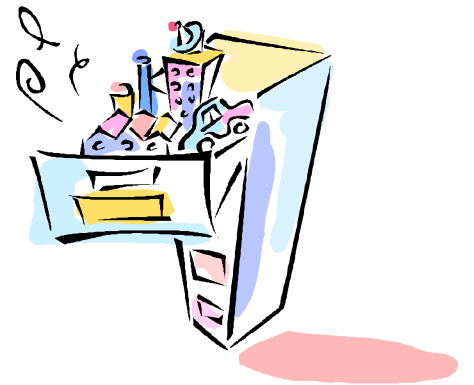
Qual dos seus olhos já foi operado?



Você já teve algum de seus  
olhos operados ?

# Estudando a Distribuição de Frequências

Variáveis Qualitativas Ordinais



# Tabelas de Frequências para Variáveis Qualitativas Ordinais

**Tabela de distribuição de freqüências das plantas da variedade 3 do ciclame segundo a coloração das flores**

Coloração	Freqüência Absoluta	Freqüência Relativa (%)	Freqüência Absoluta Acumulada	Freqüência Relativa (%) Acumulada
Fraco	150	31.25	150	31.25
Média	158	32.92	308	64.17
Forte	172	35.83	480	100
Total	480	100	----	----

**A ordem natural das categorias deve ser respeitada**

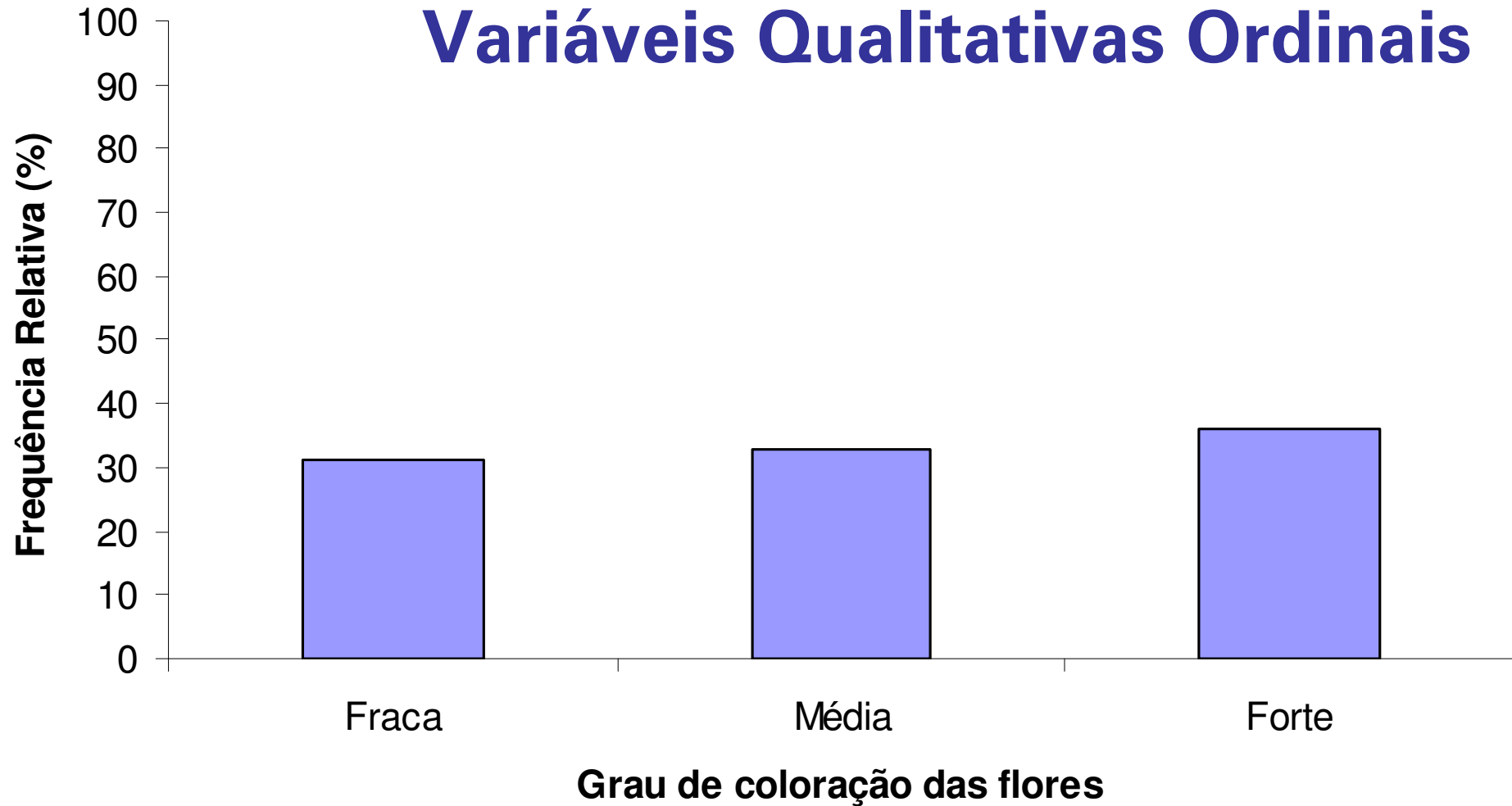
# Tabelas de Frequências para Variáveis Qualitativas Ordinais

**Tabela de distribuição de frequências das plantas da variedade 3 do ciclame segundo a coloração das flores**

Coloração	Frequência Absoluta	Frequência Relativa (%)	Frequência Absoluta Acumulada	Frequência Relativa (%) Acumulada
Fraco	150	31.25	150	31.25
Média	158	32.92	308	64.17
Forte	172	35.83	480	100
Total	480	100	----	----

**As colunas de frequências acumuladas só fazem sentido quando as categorias da variável possuem uma ordem natural**

# Gráficos para visualizar a Distribuição de Frequências de Variáveis Qualitativas Ordinais



**Gráfico de Barras (respeitando a ordem das categorias)**

## Exemplo 3: estudo das condições de saúde de crianças de um município brasileiro

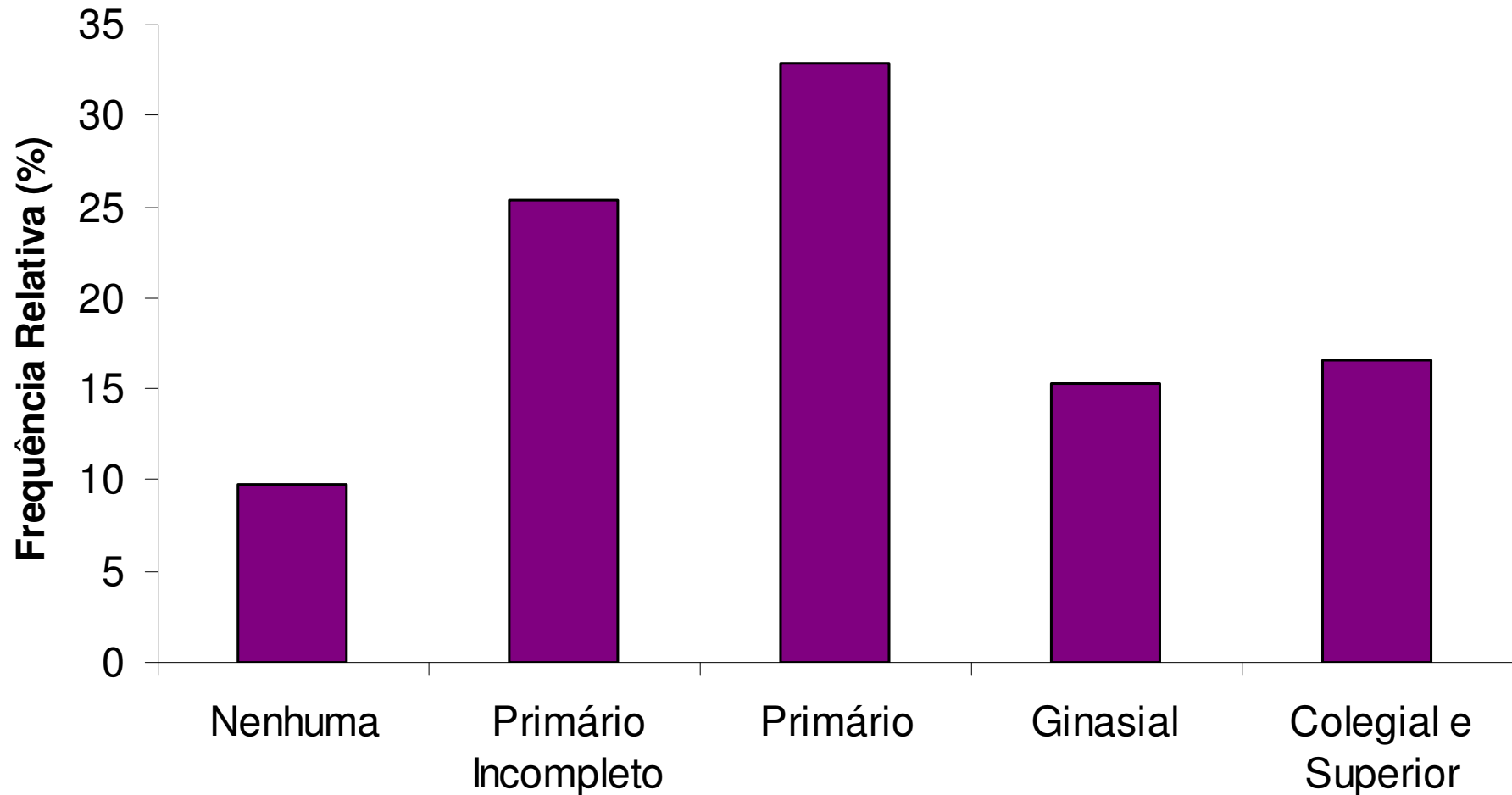
Distribuição de freqüências das crianças entrevistadas segundo escolaridade do chefe de família

Escolaridade do chefe da família	Freqüência Absoluta	Freqüência Relativa (%)	Freqüência Relativa Acumulada (%)
Nenhuma	99	9.8	9.8
Primário Incompleto	256	25.4	35.2
Primário	331	32.9	68.1
Ginasial	154	15.3	83.4
Colegial e Superior	167	16.6	100
<b>Total</b>	<b>1007</b>	<b>100</b>	<b>-----</b>

# Exemplo 3 : estudo das condições de saúde de crianças de um município brasileiro

[Monteiro e Benício, 1987 ]

**Distribuição de frequências das crianças entrevistadas segundo escolaridade do chefe da família**

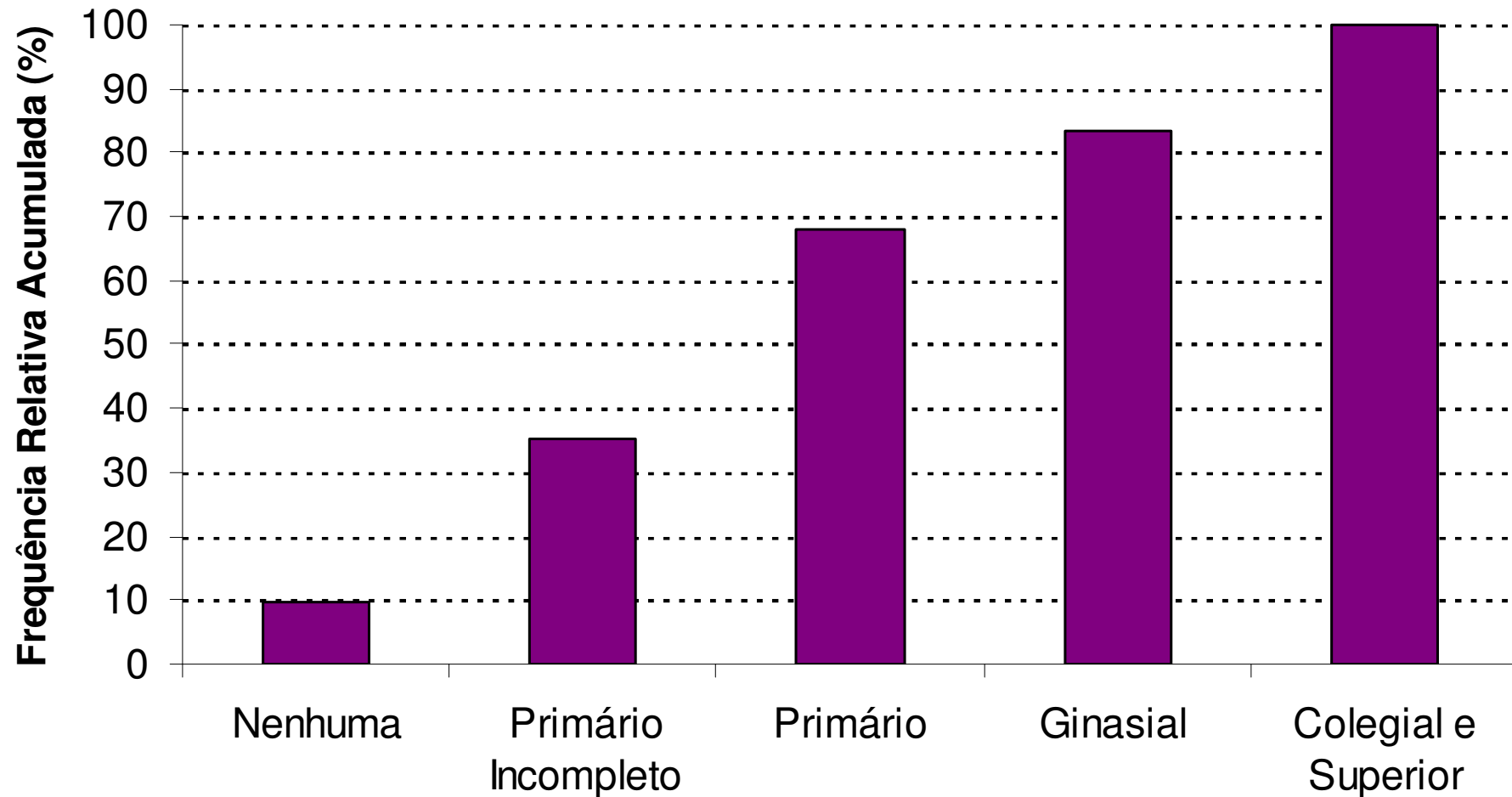




# Exemplo 3: estudo das condições de saúde de crianças de um município brasileiro

[Monteiro e Benício, 1987 ]

**Distribuição de frequências acumuladas das crianças entrevistadas segundo escolaridade do chefe da família**



## Exemplo 4: Pesquisa sobre Saúde e Nutrição (NHANES 2007-2008)

[CDC, 2008]

**Tabela de distribuição de freqüências da qualidade de audição  
(NHANES 2007- 2008)**

Qualidade de Audição	Freqüência Absoluta	Freqüência Relativa (%)	Freqüência Absoluta Acumulada	Freqüência Relativa (%) Acumulada
Excelente	4610	47.72	4610	47.72
Boa	3418	35.38	8028	83.10
Pouco problemática	1022	10.58	9050	93.68
Moderadamente problemática	381	3.94	9431	97.62
Muito problemática	211	2.18	9642	99.80
Surdo	19	0.20	9661	100
Total	9661	100	----	----

Como você classifica sua audição?

S

MMP

MP

PP

B

E

E - Excelente

B - Boa

PP - Pouco problemática

MP - Moderadamente Problemática

MMP - Muito Problemática

S - Surdo

0.0

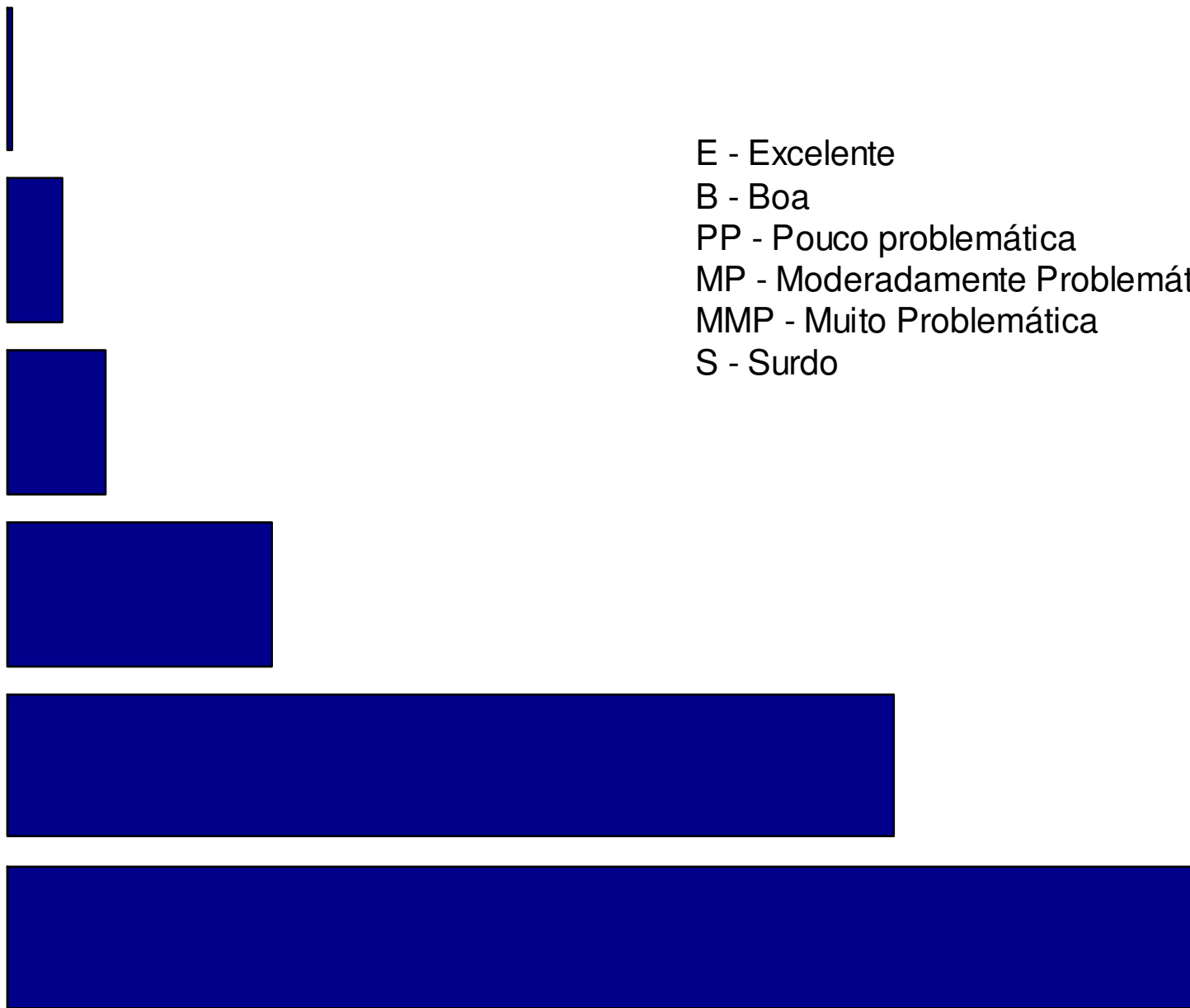
0.1

0.2

0.3

0.4

Frequência Relativa (%)



# Para compreender melhor ...

Caderno de Exercícios

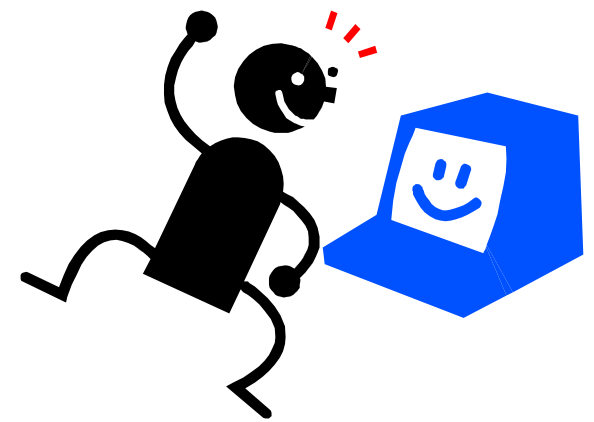
Exercícios da **Seção 2**

Exercícios 2.1, 2.2 e 2.3



[www.minha.ufmg.br](http://www.minha.ufmg.br)

# Praticando Estatística



<http://cran.r-project.org/>

Introdução ao ambiente de  
computação estatística **R**



## O que é o R ?

R é um programa de computador **gratuito** que possibilita o uso de diversos métodos estatísticos na análises de dados.

## Onde consigo o R ?

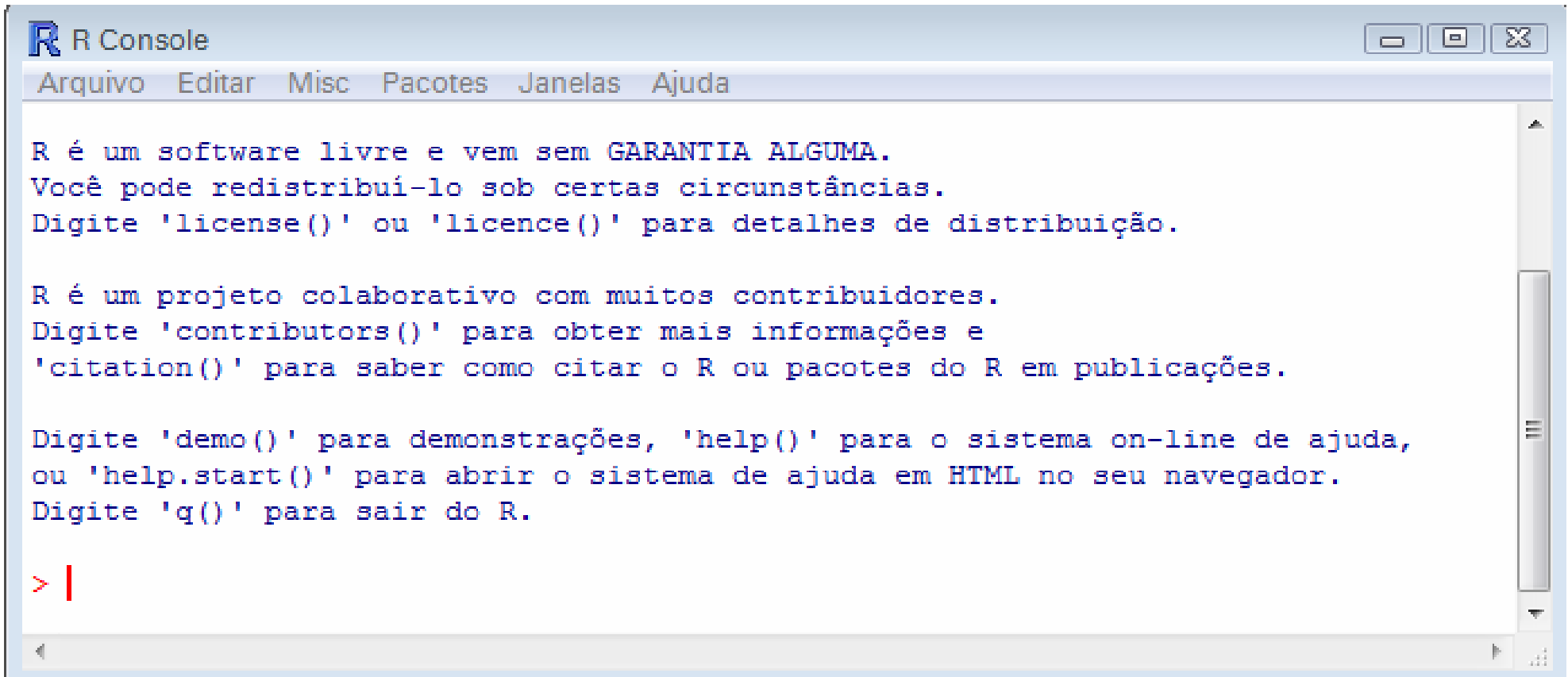
A última versão do R (para *Windows*®) pode obtida em <http://cran.r-project.org/bin/windows/base/R-2.12.0-win.exe>

## Como instalo o R ?

Clicar duas vezes no arquivo executável e seguir instruções.  
(Para mais instruções, consultar a apostila do **R no Moodle**)



# Como é o R ?

A screenshot of the R Console window. The title bar reads 'R Console' and includes standard window controls (minimize, maximize, close). The menu bar contains 'Arquivo', 'Editar', 'Misc', 'Pacotes', 'Janelas', and 'Ajuda'. The main text area displays the following information in a monospaced font:

R é um software livre e vem sem GARANTIA ALGUMA.  
Você pode redistribuí-lo sob certas circunstâncias.  
Digite 'license()' ou 'licence()' para detalhes de distribuição.

R é um projeto colaborativo com muitos contribuidores.  
Digite 'contributors()' para obter mais informações e  
'citation()' para saber como citar o R ou pacotes do R em publicações.

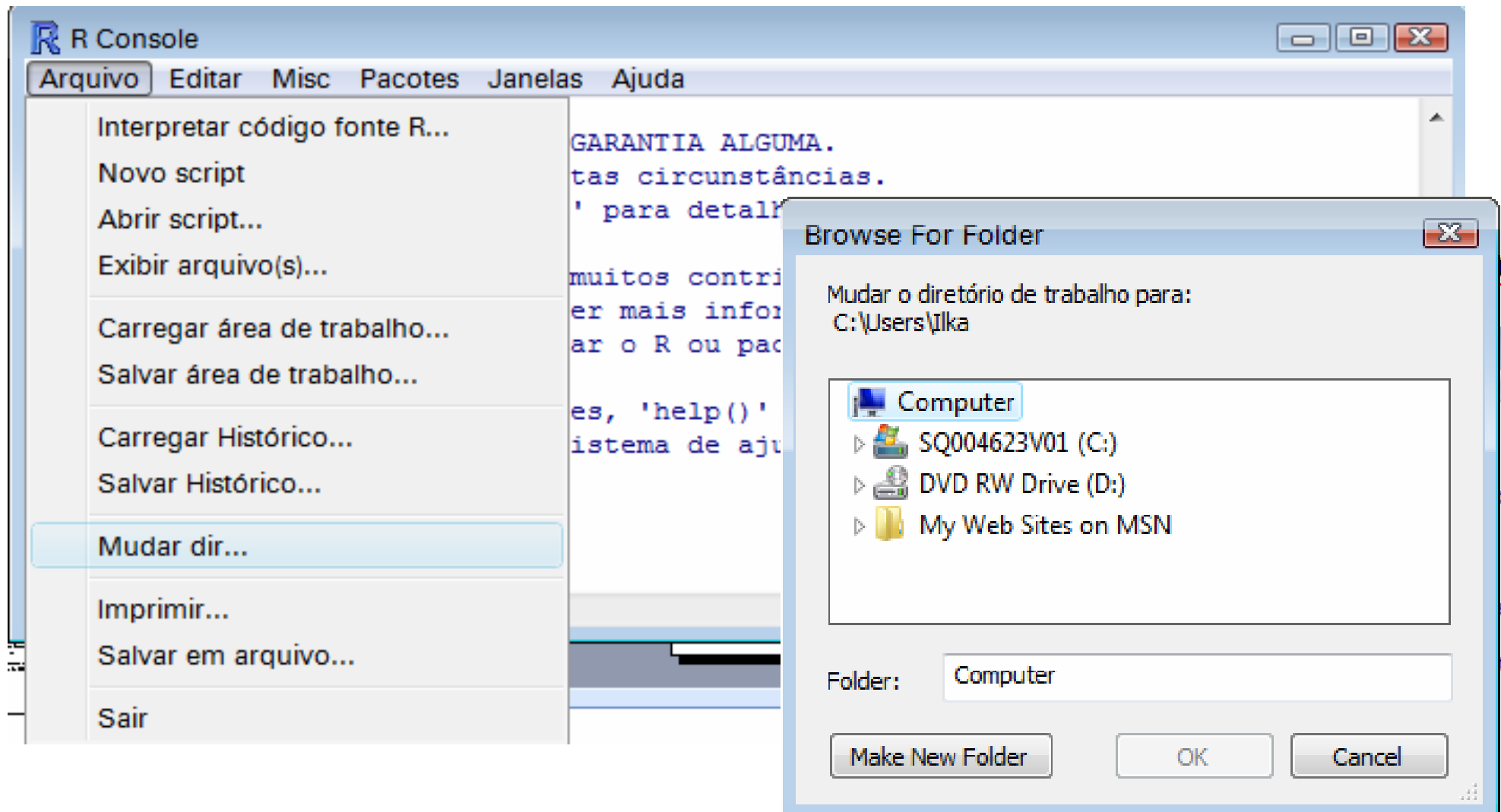
Digite 'demo()' para demonstrações, 'help()' para o sistema on-line de ajuda,  
ou 'help.start()' para abrir o sistema de ajuda em HTML no seu navegador.  
Digite 'q()' para sair do R.

> |



# Como começar a usar o R ?

## Passo 1: indicar o diretório de trabalho







# Como começar a usar o R ?

## Passo 2: ler o conjunto de dados

```
> Ciclame <- read.table("Ciclame.txt", header=TRUE)
```

```
> Ciclame[1:4, ]      #Vendo as quatro primeiras linhas
```

	Variedade	Regime	Dia	Noite	Fertilizante	Flores	Coloracao	Area	Pragas	
1		2	5	14	20	1	12	forte	10.4	1
2		2	5	14	20	1	8	fraca	7.5	0
3		2	5	14	20	1	8	fraca	7.7	0
4		2	5	14	20	1	6	forte	5.8	1

```
> Ciclame$Flores      #Vendo os dados da variável "Flores"
```

```
> Ciclame$Flores[1:5] #Vendo os 5 primeiros dados da  
#variável "Flores"
```

# Como fazer gráficos e tabelas no R ?



## Passo 3: construir tabela de frequências

```
tabela <- table(Ciclame$Pragas)
```

```
names(tabela) <- c("Não", "Sim") # 0 – Não ; 1 – Sim
```

```
tabela
```

```
  Não  Sim
```

```
985 935
```

## Passo 4: fazer gráfico de setores ("torta")

```
pie(tabela)
```

# Como fazer gráficos e tabelas no R ?



**Passo 5:** 1- fazer gráfico de barras (verticais)

```
barplot(tabela, col="green",  
        ylab="Frequência Absoluta")
```

2 - fazer gráfico de barras (horizontais)

```
barplot(tabela, col="green",  
        ylab="Frequência Absoluta", horiz=TRUE)
```

# Como fazer gráficos e tabelas no R ?



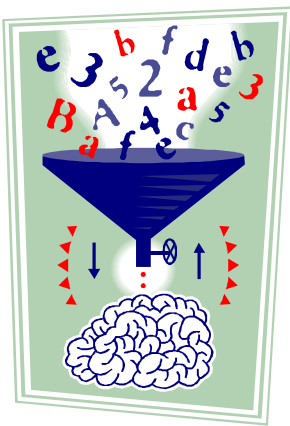
**Passo 6:** se você já tem as frequências das classes

```
Cirurgia.olhos <- c(89, 98, 443, 6542)
```

```
# Frequências Absolutas das classes
```

```
names(Cirurgia.olhos) <- c("Direito", "Esquerdo",  
"Ambos", "Nenhum") #Nomes das classes
```

```
barplot(Cirurgia.olhos, xlab="Frequência Relativa (%)",  
main="Qual dos seus olhos já foi operado?",  
col="darkred", horiz=T)
```



# Fim da segunda aula

Análise Descritiva de Dados:  
Organizando e Apresentando Dados de  
Variáveis Qualitativas

# Próxima Aula

Análise Descritiva de Dados:  
Organizando e Apresentando Dados de  
Variáveis Quantitativas

# Referências Bibliográficas

Center for Disease Control - CDC (2008). *National Health and Nutrition Examination Survey* (NHANES 2007-2008). Disponível em:  
[http://www.cdc.gov/nchs/nhanes/nhanes2007-2008/quex07\\_08.htm](http://www.cdc.gov/nchs/nhanes/nhanes2007-2008/quex07_08.htm)

MONTEIRO, C. A.; BENÍCIO, M. H. D'A. (1987) *Estudo das condições de saúde das crianças do Município de São Paulo, SP (Brasil), 1984/1985*. VI. Doença respiratória. *Rev.Saúde públ., S. Paulo*, 21:380-86

REIS, E.A.; REIS, I.A. (2001) *Análise Descritiva de Dados- Tabelas e Gráficos*, Relatório Técnico do Departamento de Estatística da UFMG. Disponível em: <http://www.est.ufmg.br>