

Gráficos em Estatística

Parte 1

Fundamentos

Estatística e o nosso tempo

Vivemos numa era em que a disponibilidade e o acesso a **dados** não tem precedentes na história da humanidade.

Por este motivo, a **Estatística** nunca esteve tanto em evidência.

Métodos estatísticos fornecem a base para a obtenção de dados relevantes, atuais, precisos e custo-efetivos.

Métodos estatísticos também movem a **extração de conhecimento** relevante dos dados para apoiar a tomada de decisões.

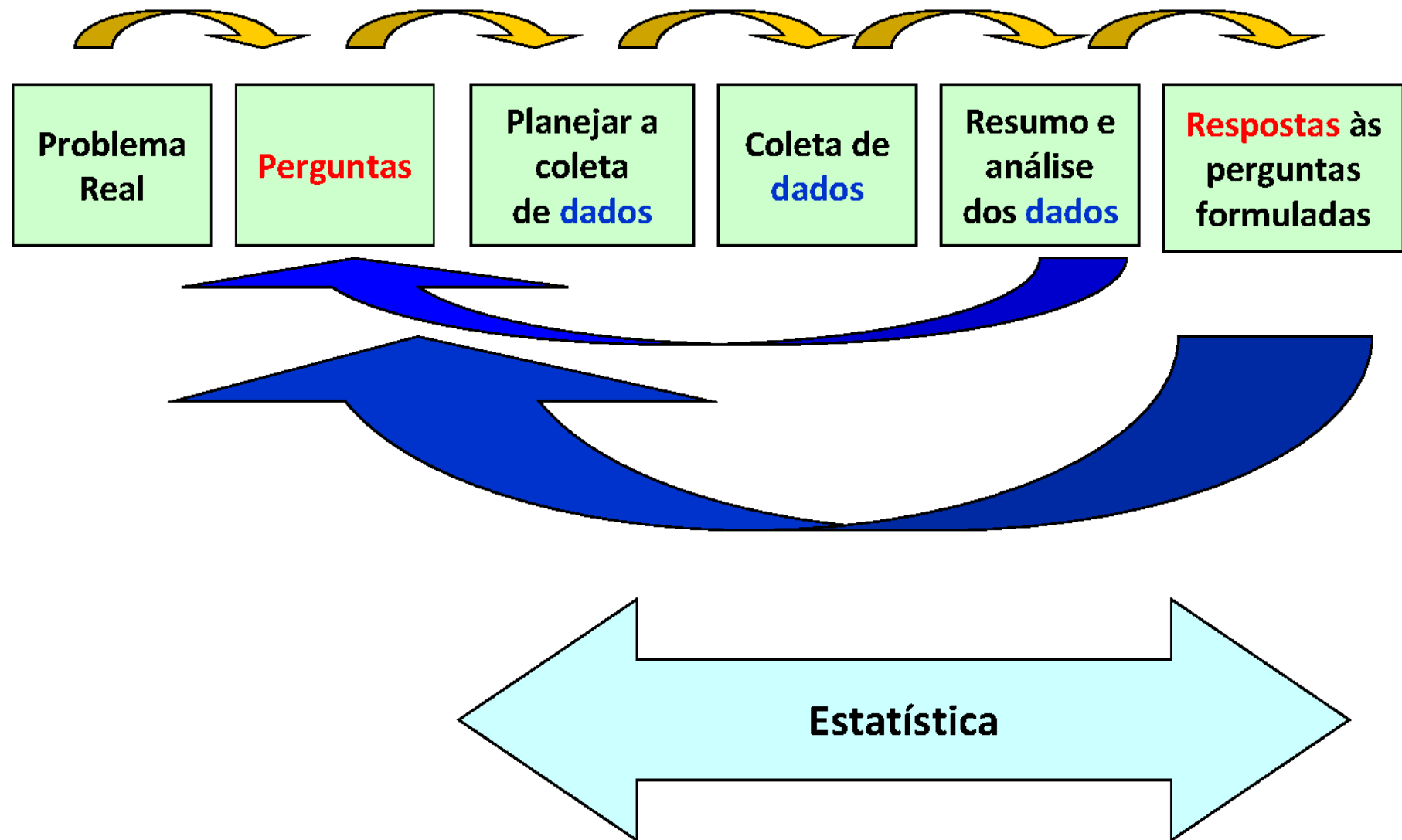
Estatística: o que é?

Estatística é a ciência que trata de como coletar, resumir, analisar, interpretar e apresentar dados sobre algum aspecto de interesse do mundo real, visando:

- Conhecer a realidade (responder perguntas), e
- Apoiar a tomada de decisões.

Estatística está na base dos processos de investigação e de geração de conhecimento empregados pela ciência moderna.

Processo de investigação (geração de conhecimento)



Estatística: como?

Como é que a **Estatística** oferece soluções para os problemas de **investigação e geração de conhecimento**?

Essencialmente, através:

- De cuidadoso **planejamento e realização** de operações de **coleta de dados e medidas** sobre os fenômenos de interesse (**registros administrativos, censos, pesquisas amostrais e experimentos**);
- Da **análise exploratória** dos dados coletados (**observações**);
- Da **formulação e ajuste** de **modelos estatísticos** para **descrever os dados de forma sintética**, e **facilitar a obtenção de respostas às perguntas de interesse** (**inferência**);
- Da **apresentação dos dados**, e de **resumos dos dados** e outros resultados em forma compreensível.

Estatística: componentes centrais

O planejamento de operações de coleta de dados é tratado em cursos de:

- **Planejamento de experimentos**;
- **Amostragem** (probabilística) e **Metodologia de Pesquisa**

que são as duas principais classes de estratégias disponíveis para **obtenção de observações e medidas** sobre fenômenos de interesse.

A **análise exploratória de dados** é tratada em cursos de '**Estatística Descritiva**' – p. ex. capítulos 2 e 3 do (Wild and Seber 2004).

A **formulação e ajuste de modelos aos dados** é tratada em cursos de '**Inferência**' e '**Modelagem Estatística**' – p. ex. capítulos 5, 6, e 10 a 12 do (Wild and Seber 2004).

A **apresentação de dados estatísticos** em tabelas e **gráficos**. Veja por exemplo (UNECE 2009a), (UNECE 2009b) e (Knafllic, 2015).

Dados: o que são

Em Estatística, **dados** são **coleções de números ou informações** aos quais se atribui um **significado**.

Considere os valores 1, 0, 2, 0, 4.

Assim apresentados, são apenas números.

Mas se você souber que são os **números de filhos tidos nascidos vivos** de 5 mulheres entrevistadas, viram **dados sobre a fecundidade** das mulheres.

Dados = Números + Significado.

Tipos de dados

Dados podem ser classificados segundo o **tipo** em quatro grupos:

Categóricos		Numéricos	
Nominais	Ordinais	Discretos	Contínuos

O tipo do dado influencia diretamente as operações que podemos fazer para:

- ✓ **Trabalhar** com os dados;
- ✓ **Resumir** os dados;
- ✓ **Interpretar** os dados;
- ✓ **Apresentar** os dados.

Dados categóricos nominais

Dados categóricos são aqueles que podemos alocar em categorias.

Dados categóricos são **nominais** quando não possuem ordenação lógica.

Exemples:

- 1. Sexo**
- 1 ○ Masculino 3 ○ Feminino.

2. Você tem conta bancária? 1 ☐ Sim 2 ☐ Não.

- ### 3. Pergunta do questionário da VII Encuesta de Presupuestos

Familiares 2011/12:

No leer al informante. Completar según su observación.

VP02. La vivienda principal, ¿posee algún negocio adosado?

1. Si. Especifique (a) _____

2. No

☐

Dados categóricos ordinais

Dados categóricos são **ordinais** quando as categorias podem ser ordenadas.

Ex.: Posição do time no último campeonato: 1º, 2º, 3º, etc.

Ex.: **Itens de Likert** – Geralmente consiste em apresentar uma frase ou afirmação, e pedir ao entrevistado para dizer se:

1. ☐ Discorda totalmente
2. ☐ Discorda parcialmente
3. ☐ Nem discorda nem concorda (indiferente / indeciso)
4. ☐ Concorda parcialmente
5. ☐ Concorda totalmente

Dados numéricos

Dados numéricos são aqueles que podemos medir ou contar, representando a resposta por números para os quais operações aritméticas fazem sentido, isto é, podemos somar, subtrair, etc.

Dados numéricos discretos.

Quantas contas bancárias você tem? (Número).

Dados numéricos contínuos.

Qual é seu peso? , Kg .

Qual é a área de sua residência? m².

Exercício 1.1: Dê um exemplo de cada tipo de dado numérico.

Resumindo

Adiante vamos estudar **como apresentar dados e seus resumos** na forma de **gráficos estatísticos**.

Para poder avançar nesse estudo é fundamental saber **identificar os tipos de dados** com que estamos lidando, em cada situação.

Objetivos (do curso)

1. Discutir os **princípios básicos** e algumas ideias para **apresentação gráfica de dados estatísticos**.
2. **Examinar bons e maus exemplos** de gráficos e figuras usados para resumir e apresentar dados estatísticos.
3. **Estimular leitura correta de gráficos**, aguçar seu sentido crítico em relação a gráficos mal feitos.
4. **Colaborar para elaboração de melhores gráficos** para resumo e apresentação de dados estatísticos.

Gráficos Estatísticos: Objetivos

1. Um gráfico deve mostrar um **resumo visual** para uma coleção de dados.
2. Esse resumo visual **deve substituir** o exame detalhado dos dados, ao menos numa primeira análise.
3. O gráfico deve “**contar uma história**” ou “**passar uma mensagem**” ao leitor.
4. Há muitas maneiras de apresentar (representar) dados estatísticos através de gráficos → escolher a maneira adequada será importante.

Gráficos Estatísticos: Eficácia

1. Se for **bem feito**, um gráfico pode **passar** uma **mensagem rapidamente**, e poupar o tempo que o leitor levaria para examinar os dados (O BOM).
2. Se for **mal feito**, um gráfico pode **enganar** a muitos, à exceção talvez do leitor ou analista mais atento, bem treinado e detalhista (O MAU).
3. Um gráfico é bom quando leva ao **correto entendimento da estória correta pelo leitor**.

Gráficos Estatísticos: Usos

Gráficos estatísticos podem servir a diferentes usos:

- A. **Exploração**, descoberta;
- B. **Interpretação**, inferência, diagnóstico;
- C. **Comunicação**, apresentação, destaque, clareza.

Gráficos Estatísticos: Princípios Gerais

1. **Mostre os dados.**
2. **Diga a verdade.** Os dados devem ser representados sem distorções, atenuações ou exageros.
3. **Siga um propósito definido:** exploração, interpretação, comunicação.
4. **Use o tipo de gráfico adequado** para cada conjunto de dados.

Gráficos Estatísticos: Princípios Gerais

1. Os dados devem destacar-se claramente do “fundo” da figura.
2. Evite o “lixo gráfico” (qualquer coisa que chame mais atenção do que os dados que se quer mostrar).
3. Inclua **rótulos e descrições** (metadados) adequados:
 - A. Título, conteúdo ou objetivo do gráfico;
 - B. O que cada eixo, fatia de torta, barra, etc. representa;
 - C. Escalas legíveis e adequadas;
 - D. Fonte dos dados usados para fazer o gráfico.

Tipos de Dados e de Gráficos

- Dados podem ser de dois tipos principais:
 - Dados categóricos;
 - Dados numéricos.
- O tipo dos dados condiciona de maneira central os tipos de gráficos disponíveis para sua representação.
- Um erro comum é usar gráficos inadequados ao tipo dos dados que se quer representar.

Gráficos em Estatística

Parte 2

Sumarização e apresentação de dados categóricos

Resumo de Dados Categóricos

Distribuição de Frequências (simples)

- Apresenta informação sobre categorias de uma variável;
- Lista **categorias** e **frequência** (número, contagem) **de respostas** por categoria (**distribuição de frequências**);
- Pode mostrar **frequências absolutas**, **proporções** (ou porcentagens) sobre o total, ou ambas.

Distribuição de Frequências Simples – Exemplo 2.1a

Tabela 2.1 - População residente em 2009, total nacional, segundo o sexo.

Sexo	População residente (1 000 pessoas)	Porcentagem
Total	191 796	100,0
Homens	93 356	48,7
Mulheres	98 439	51,3

Fonte: IBGE, Diretoria de Pesquisas, Coordenação de Trabalho e Rendimento, Pesquisa Nacional por Amostra de Domicílios 2008-2009.

Gráfico Para Frequências Simples – Exemplo 2.1b

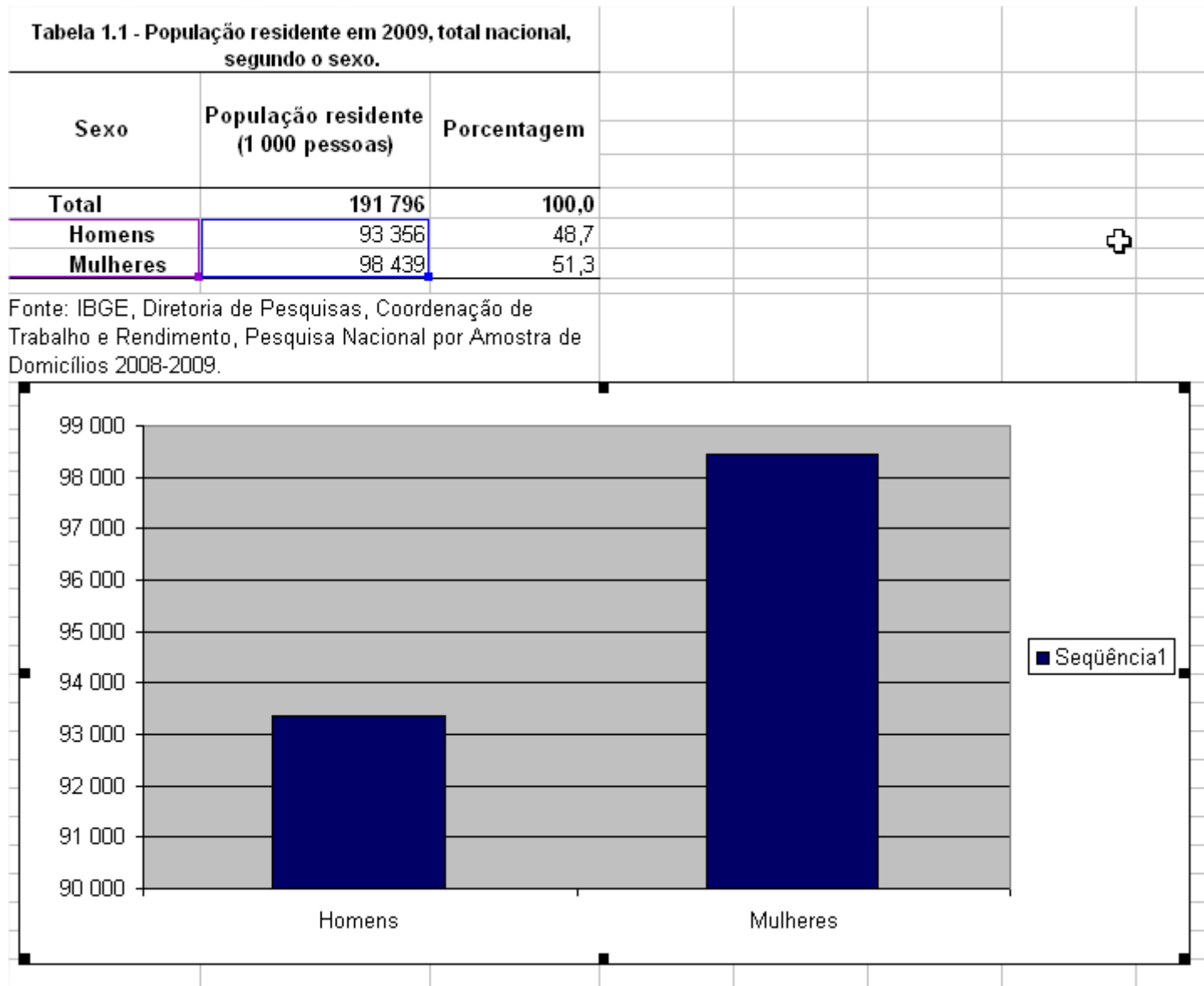
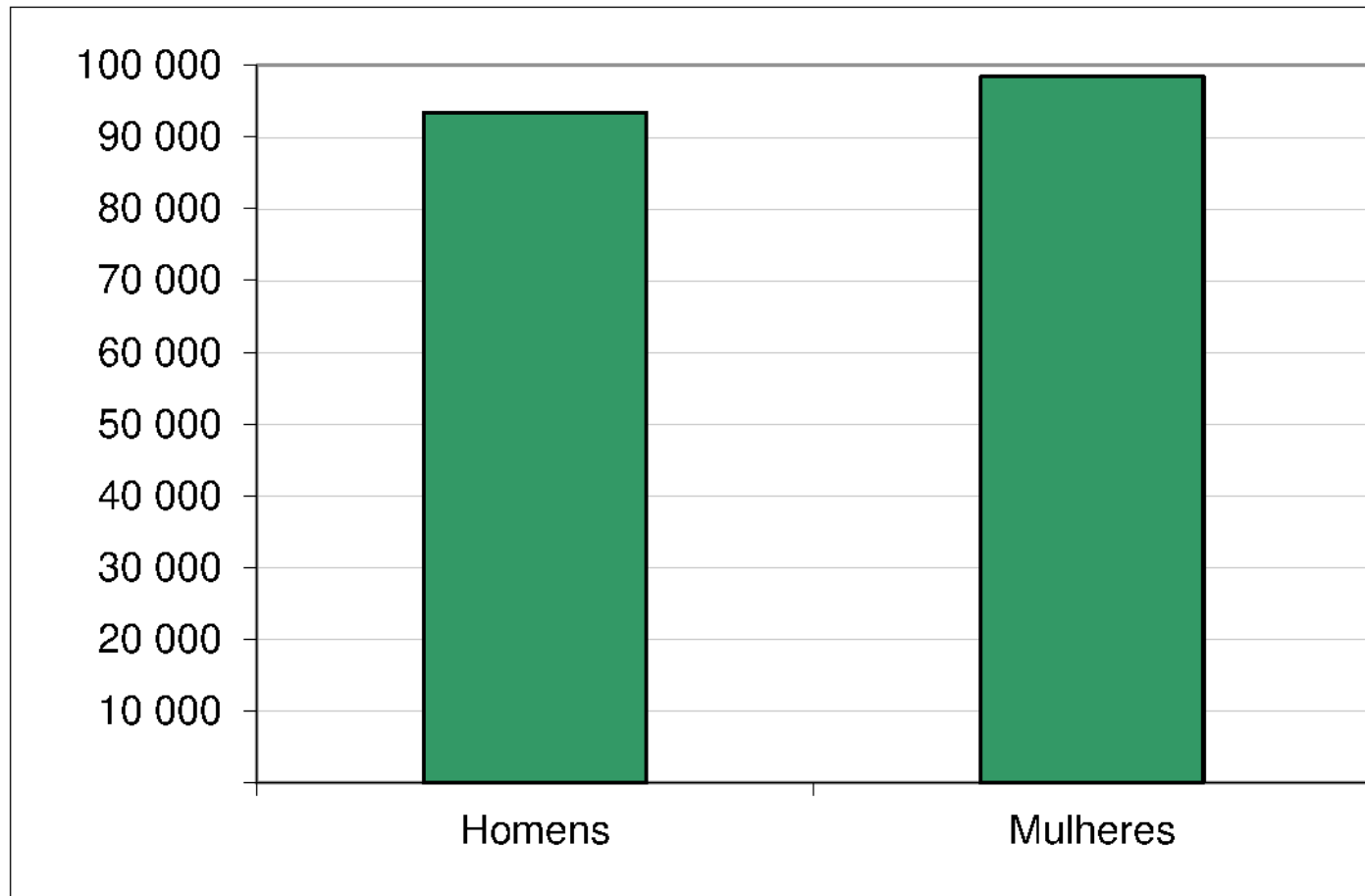


Gráfico Para Frequências Simples – Exemplo 2.1c

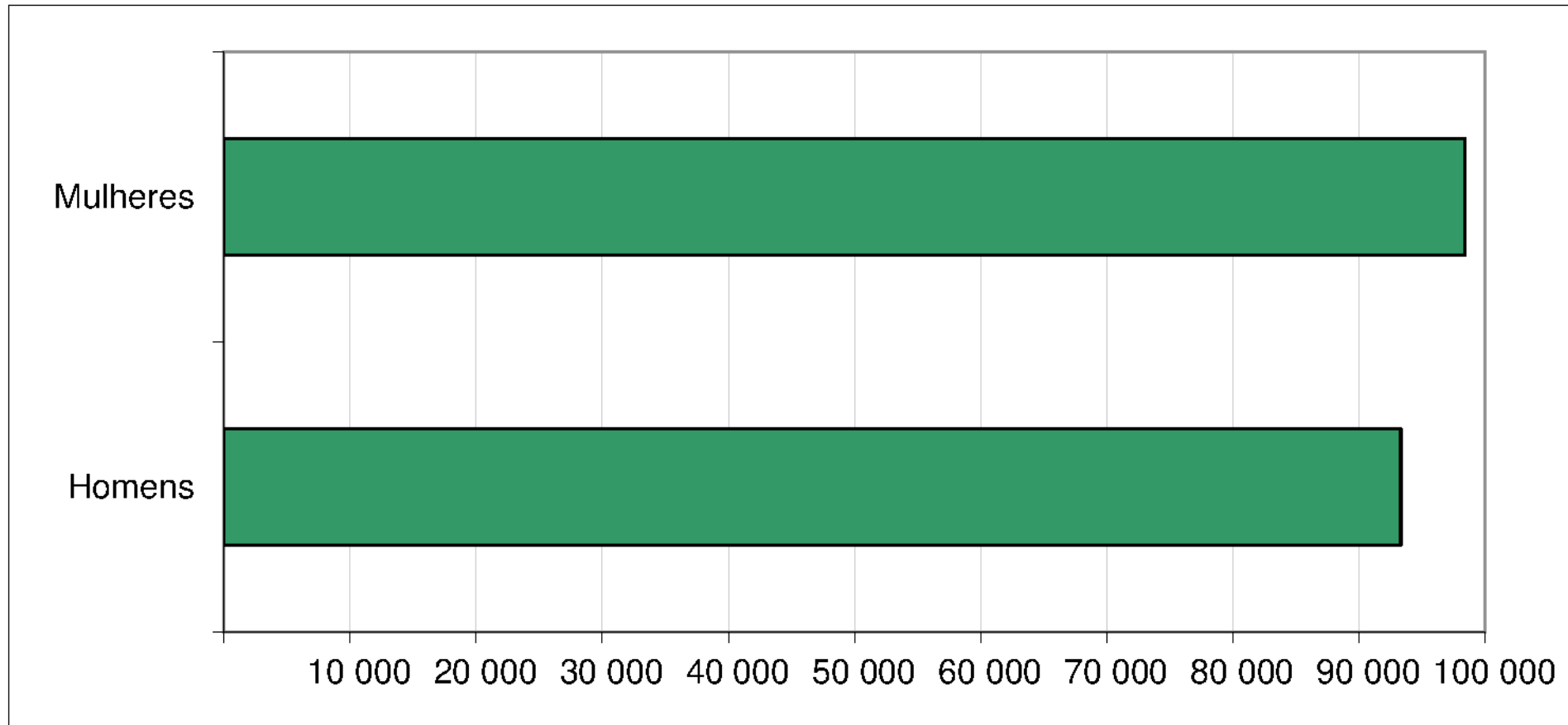
População residente (1 000 pessoas) em 2009 por Sexo do Morador



Fonte: PNAD 2009.

Gráfico Para Frequências Simples – Exemplo 2.1d

População residente (1 000 pessoas) em 2009 por Sexo do Morador



Fonte: PNAD 2009.

Gráficos para Dados Categóricos

Com **dados categóricos**, o que importa é revelar as **distribuições de frequências** (absolutas ou relativas).

Os **principais gráficos** para dados categóricos são:

- De torta / pizza / setores; em tempos recentes, variantes deste tais como o gráfico de rosca ou donut;
- De barras (ou colunas);
- De barras ou colunas justapostas.

Gráficos de Torta / Pizza / Setores

Servem para **representar** as **proporções** (frequências relativas) de casos / indivíduos em cada categoria.

Frequência relativa de cada categoria é representada pela **área de um setor num círculo** (fatia da pizza ou torta).

Usados para mostrar **como um todo se divide em partes**.

São úteis quando apenas uma **variável categórica** é medida / observada, **se houver poucas categorias** (< 6).

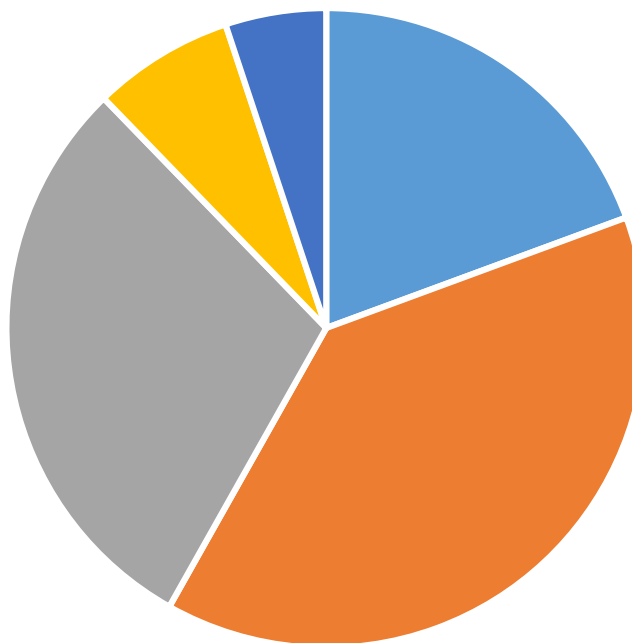
Mas podem levar a equívocos na interpretação dos dados.

Evite-os, se puder!

→ Exercício 2.1

Exemplo 2.2a – Gráfico de Setores das Intenções de Voto

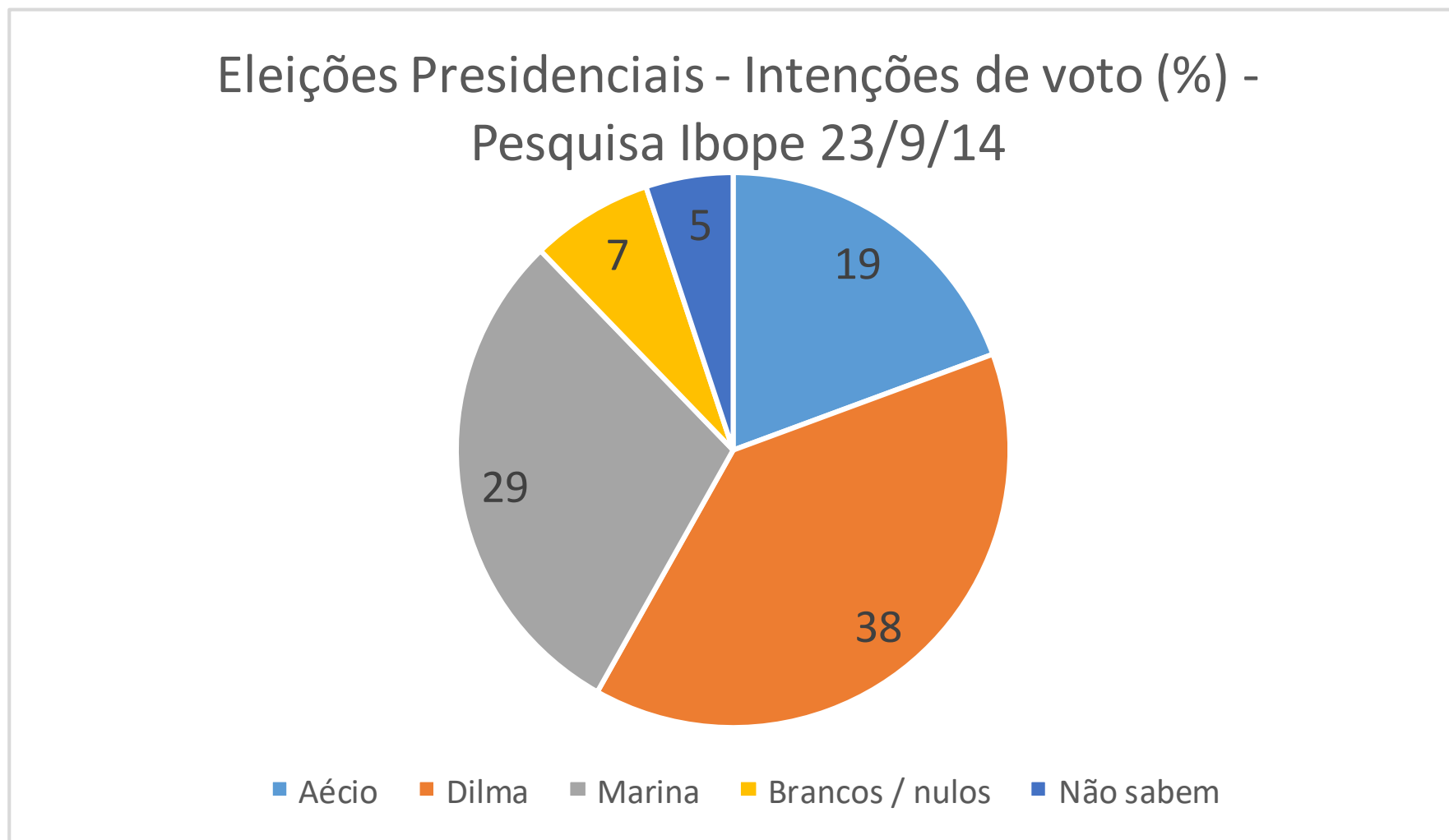
Eleições Presidenciais - Intenções de voto (%) -
Pesquisa Ibope 23/9/14



■ Aécio ■ Dilma ■ Marina ■ Brancos / nulos ■ Não sabem

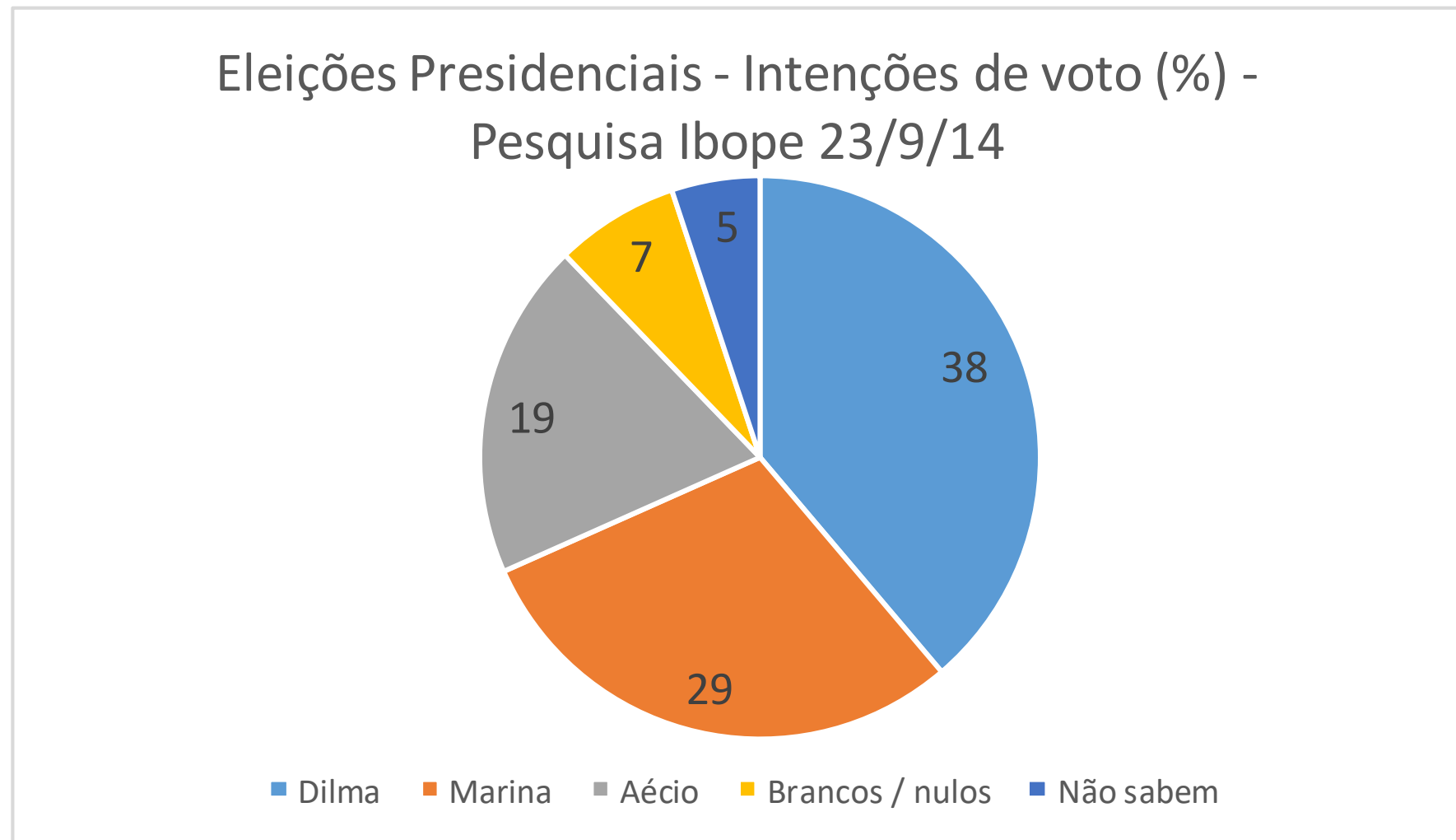
Fonte dos dados: O Globo. <http://infograficos.oglobo.globo.com/brasil/pesquisa-eleitoral-2014.html>

Exemplo 2.2b – Gráfico de Setores das Intenções de Voto



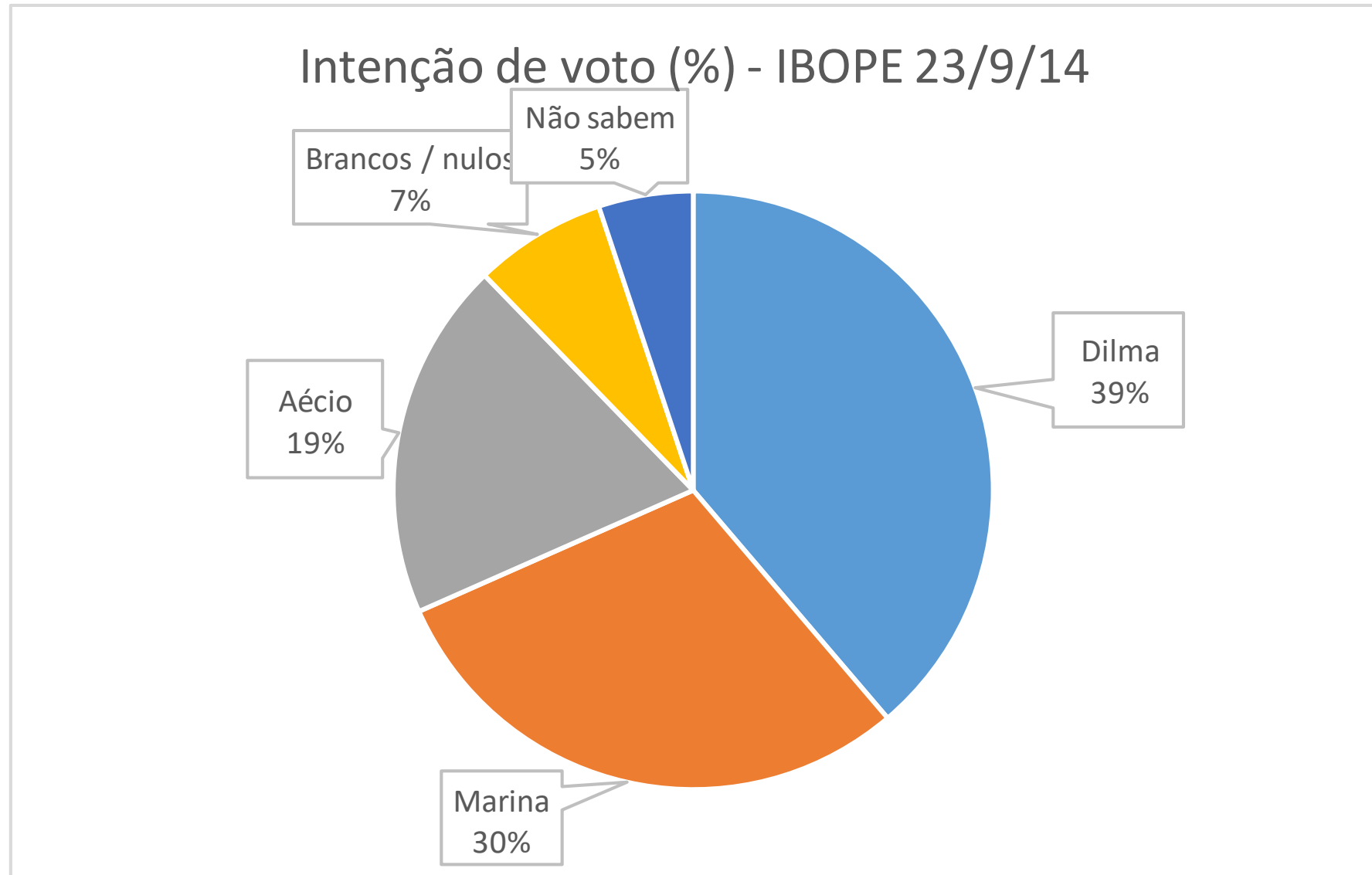
Fonte dos dados: O Globo. <http://infograficos.oglobo.globo.com/brasil/pesquisa-eleitoral-2014.html>

Exemplo 2.2c – Gráfico de Setores das Intenções de Voto



Sempre ordene categorias da maior frequência para a menor.

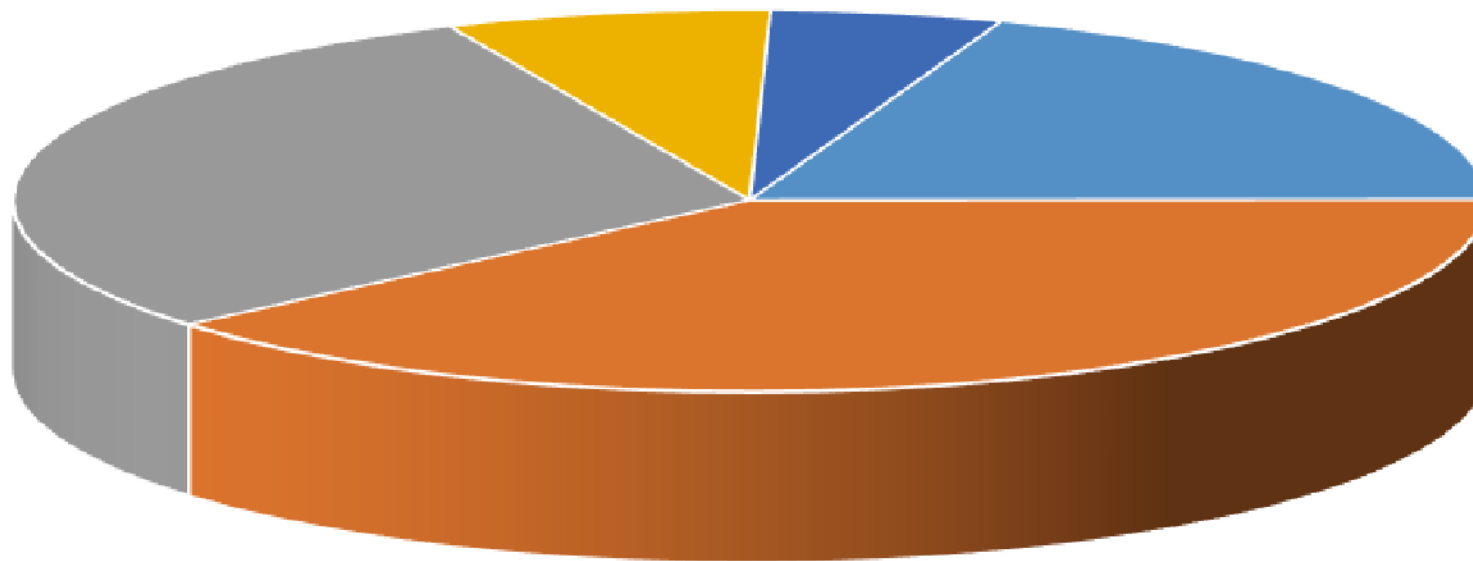
Exemplo 2.2d – Gráfico de Setores das Intenções de Voto



Forma com rótulos das categorias e valores juntos também OK.

Exemplo 2.2e – Gráfico de Setores das Intenções de Voto

Eleições Presidenciais - Intenções de voto (%) -
Pesquisa Ibope 23/9/14

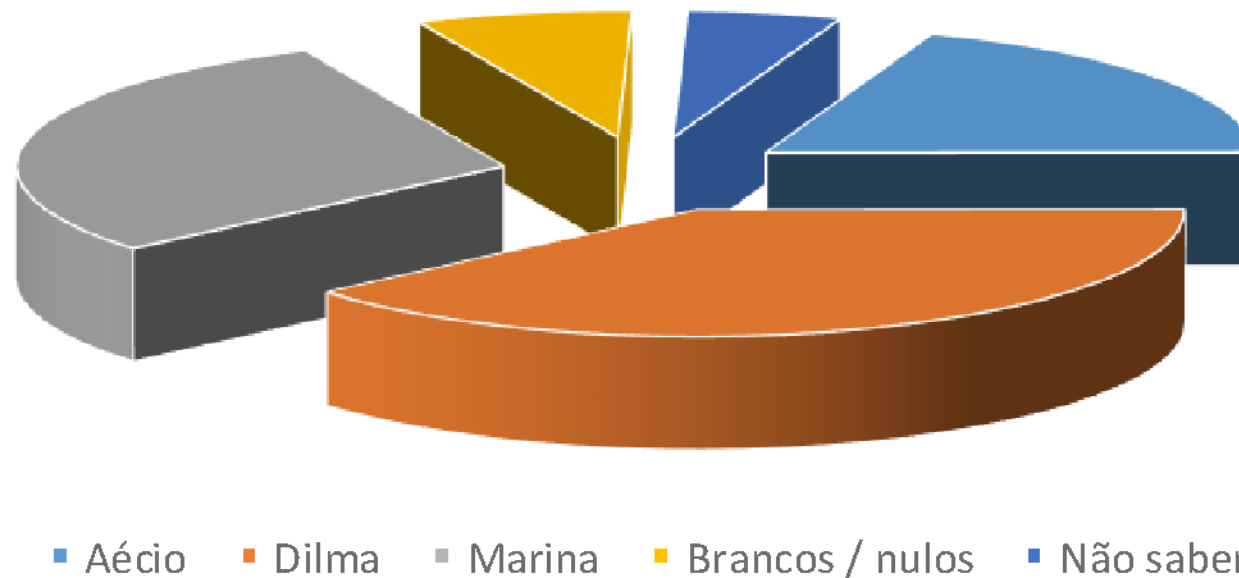


■ Aécio ■ Dilma ■ Marina ■ Brancos / nulos ■ Não sabem

Efeito 3D aplicado **distorce áreas** tornando difícil avaliar e comparar frequências. **Evite a todo custo!**

Exemplo 2.2f – Gráfico de Setores das Intenções de Voto

Eleições Presidenciais - Intenções de voto (%) -
Pesquisa Ibope 23/9/14



Separar fatias agrava problema, e contradiz filosofia do gráfico de mostrar como cada parte fica no todo. **Evite a todo custo!**

Gráficos de Torta / Pizza / Setores

Defeitos graves do gráfico no exemplo 2.2f:

1. O uso de efeitos 3D **distorce os dados** ao colocar as áreas ‘em perspectiva’, dificultando a apreensão dos valores pelo leitor.
2. Apresenta **lixo gráfico** ao colocar grande parte da área do gráfico para revelar as laterais das categorias, que não representam dados.
3. A **separação das fatias** da torta também colabora para tornar mais difícil a apreensão dos tamanhos relativos das frequências.
4. Ao mover os **rótulos** das categorias para uma **legenda**, também torna menos amigável a leitura do gráfico.

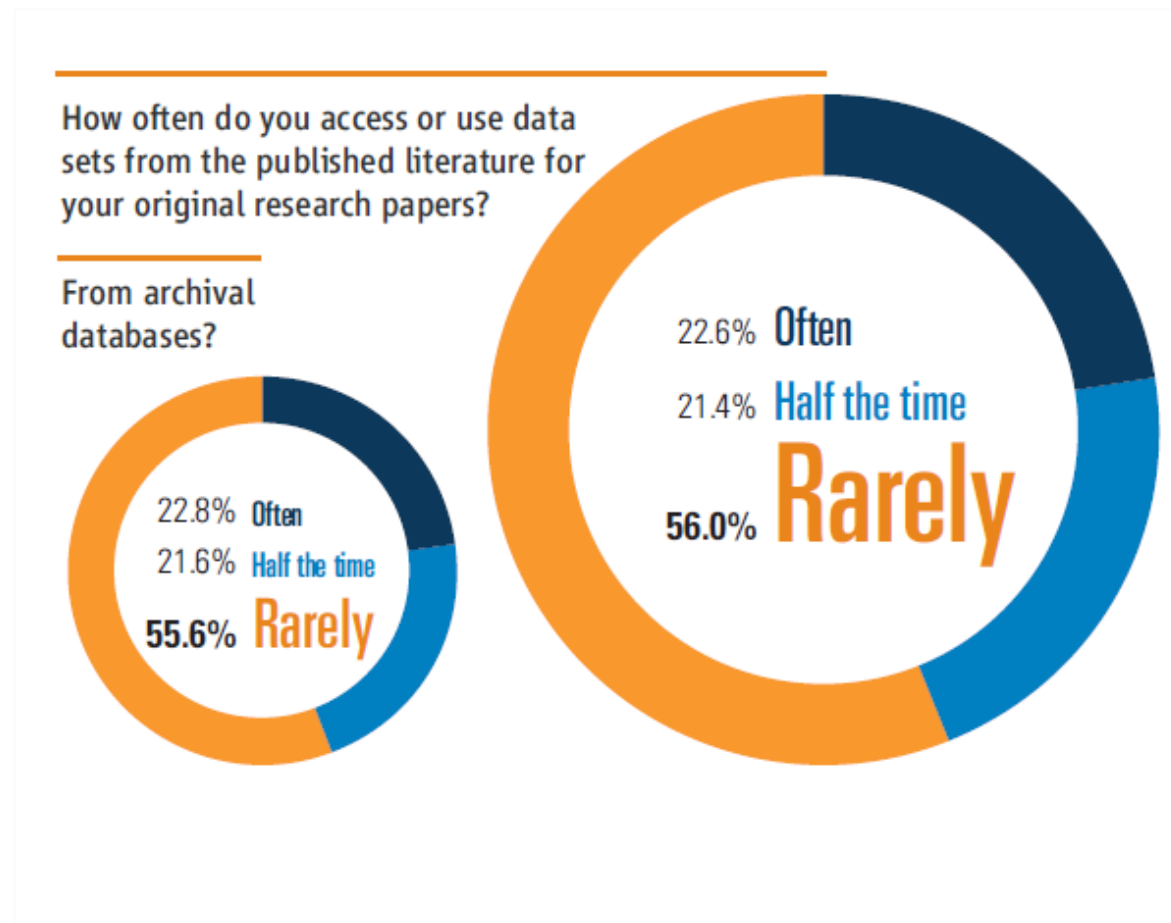
Exemplo 2.2g – Tabela de frequências das Intenções de Voto

Eleições Presidenciais - Intenções
de voto (%) - Pesquisa Ibope
23/9/14

Candidato	Intenções de voto (%)
Dilma	38
Marina	29
Aécio	19
Branco / nulos	7
Não sabem	5
Total	98

Se você precisar que o leitor 'leia' os números 'exatos' das porcentagens, é preferível apresentá-los numa tabela como a mostrada ao lado.

Exemplo 2.3 - Gráficos de Rosca ou Donut



Fonte: Revista *Science*, 11 de fevereiro de 2011, vol. 331, p. 692.
www.sciencemag.org

Gráficos para Variáveis Categóricas

Os gráficos **mais adequados** para revelar distribuições de variáveis categóricas são **gráficos de barras** ou **de colunas**.

Estes gráficos servem para apresentar a **distribuição de frequências** (absolutas ou relativas) de variáveis categóricas.

Num **gráfico de barras** as frequências são representadas por **barras horizontais**.

Num **gráfico de colunas** as frequências são representadas por **barras verticais**.

Gráficos para Variáveis Categóricas

A escolha entre um **gráfico de barras** ou um **gráfico de colunas** depende de:

- Número de categorias;
- Tamanho do espaço onde o gráfico vai ser apresentado.

Prefira sempre usar a **maior dimensão** do espaço onde o gráfico vai ser apresentado para representar a **escala numérica** das frequências.

Isto vai assegurar **melhor definição na escala** e melhor capacidade de comparar valores distintos de frequências.

Gráficos de Barras

Eixo vertical é usado para apresentar os **rótulos** da variável categórica.

Para cada categoria, uma **barra horizontal** é desenhada para representar a frequência (absoluta ou relativa) da categoria.

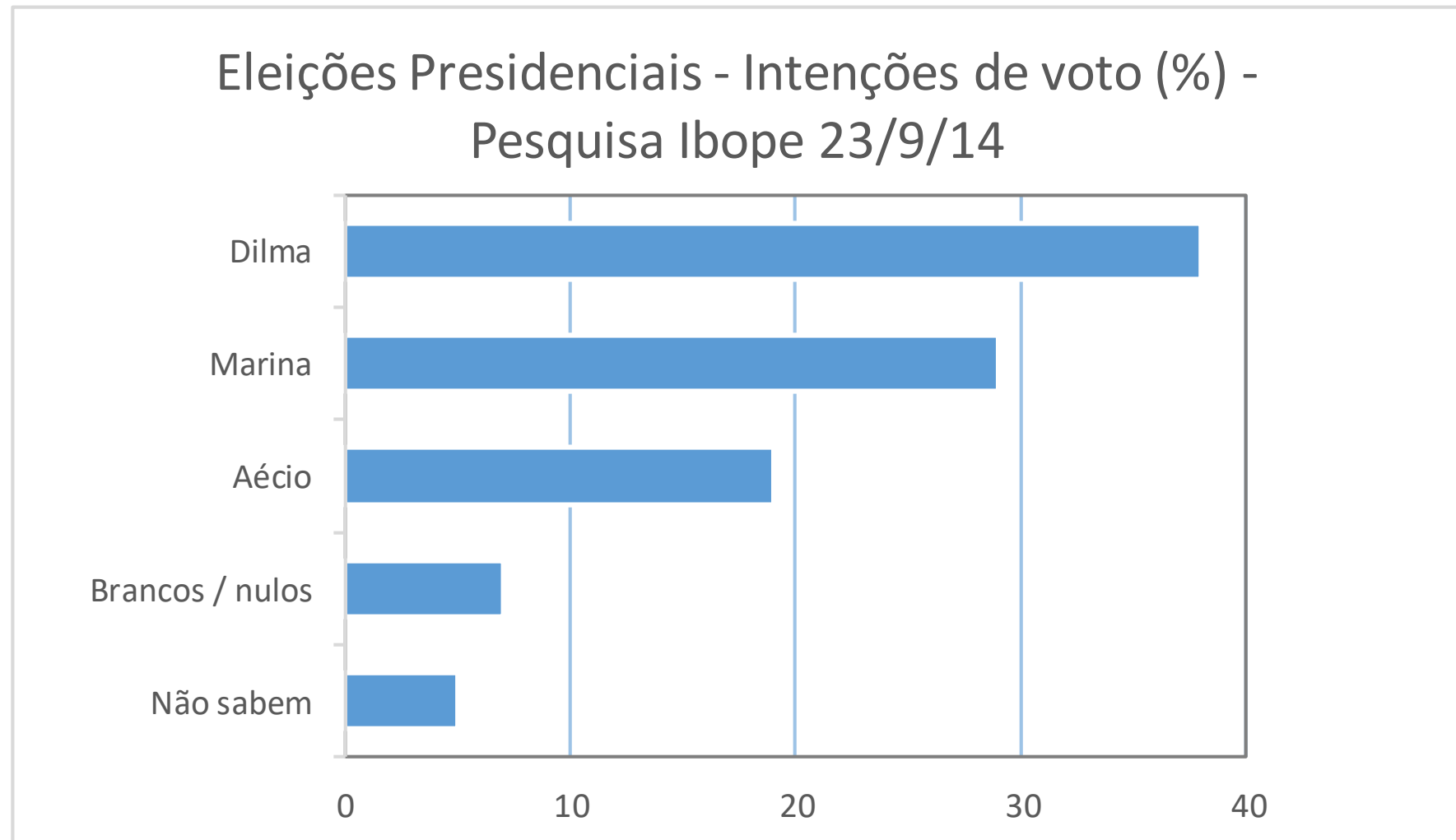
Uma decisão importante é **como ordenar as categorias**.

Como o gráfico é feito para revelar as frequências das categorias, a melhor opção é **apresentar as categorias em ordem de frequência**.

Esta opção só deve ser evitada quando for usar gráficos de barras para comparar frequências em dois períodos de tempo ou para duas populações distintas.

Nestes casos, pode ser mais indicado manter as categorias em outra ordem de interesse.

Exemplo 2.2h - Gráfico de Barras das Intenções de Voto



Exemplo 2.2i - Gráfico de Barras das Intenções de Voto

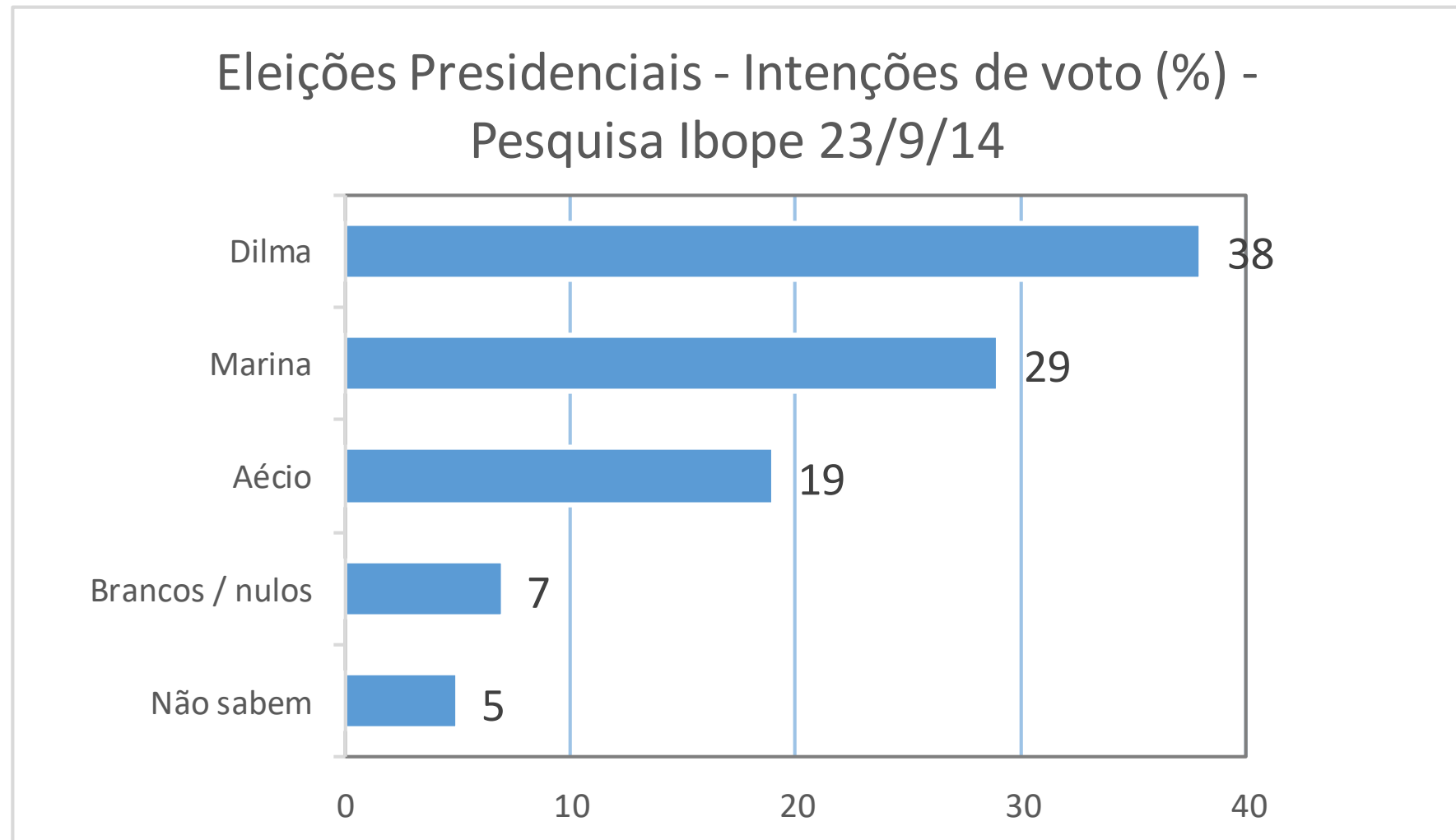


Gráfico de Barras Justapostas

Trata-se de uma variante do gráfico de barras que **ênfatiza como o todo se divide nas partes**.

Nesta variante as barras são ‘justapostas’, formando uma única grande barra, subdivida nas categorias.

Cores ou padrões diferentes são usados para diferenciar as barras das várias categorias.

Os rótulos das categorias podem ser apresentados junto das barras ou na forma de uma legenda.

Importante apresentar uma escala numérica para ‘leitura’ das frequências.

Para este tipo de gráfico, é mais comum usar **frequências relativas**.

Exemplo 2.2j - Gráfico de Barras Justapostas para as Intenções de Voto

Eleições Presidenciais - Intenções de Voto (%) - Pesquisa Ibope 23/09/2014

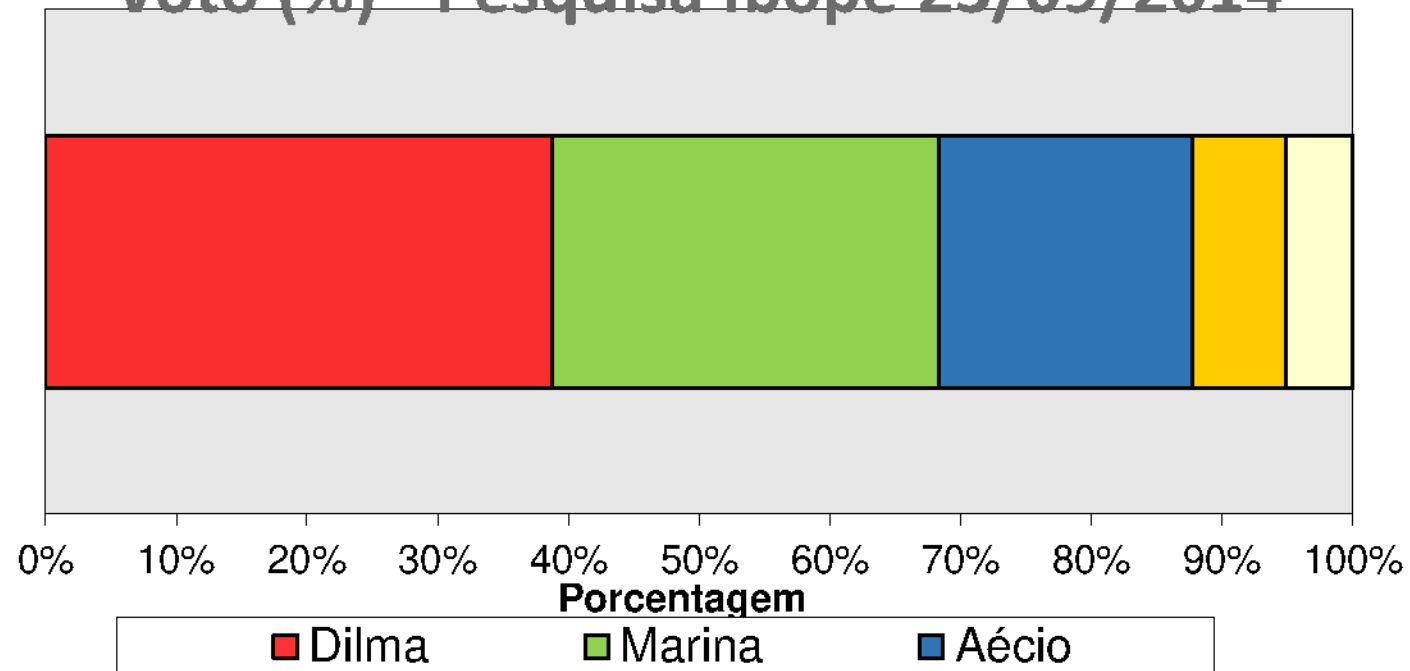
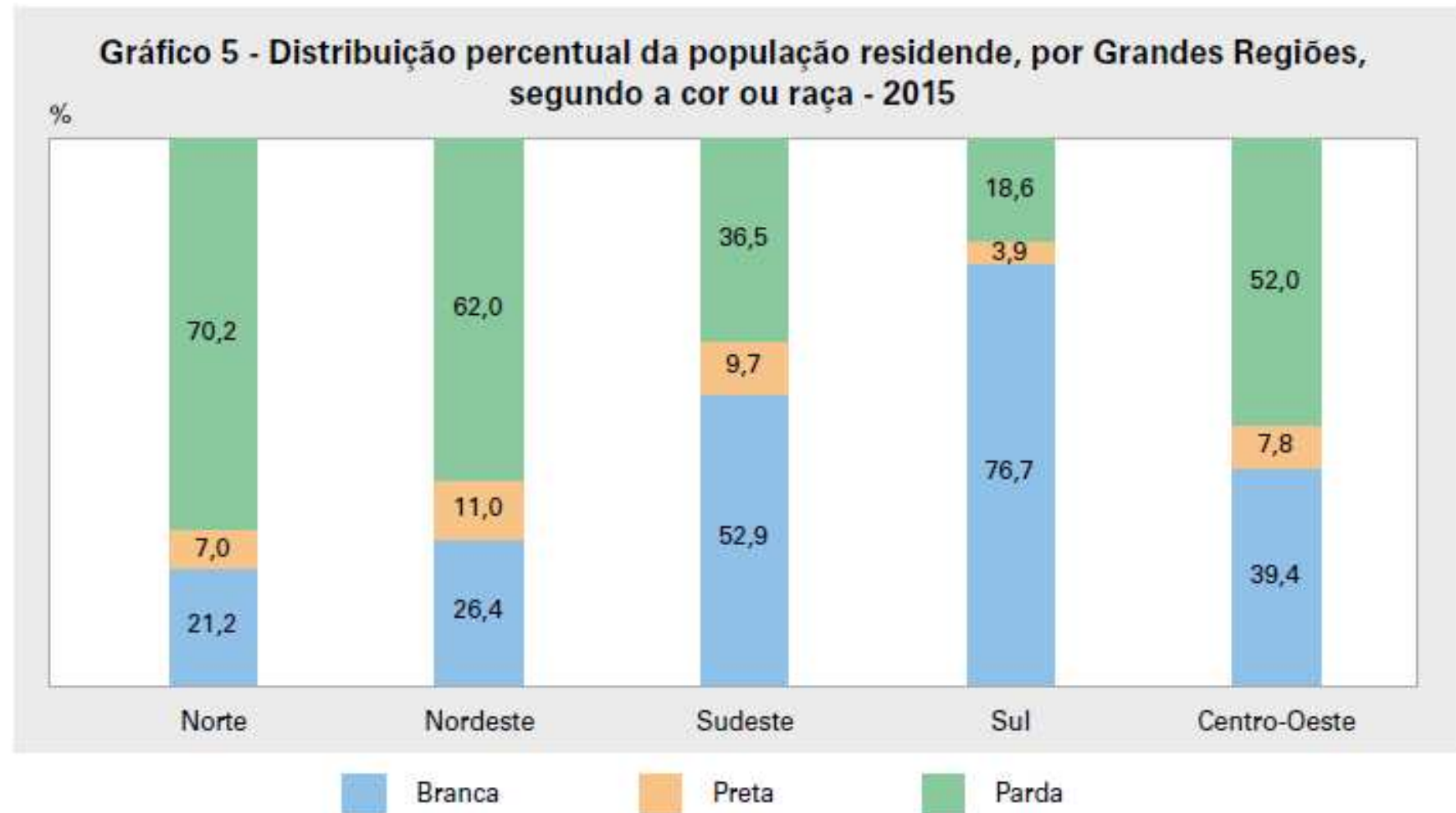


Gráfico de Barras Justapostas

Uma das vantagens do gráfico de barras justapostas é permitir **comparar distribuições** de uma mesma variável em dois (ou mais) períodos de tempo ou para populações distintas.

Os exemplos a seguir usam **gráficos de colunas justapostas**, mas as mesmas ideias seriam aplicáveis para gráficos de barras justapostas.

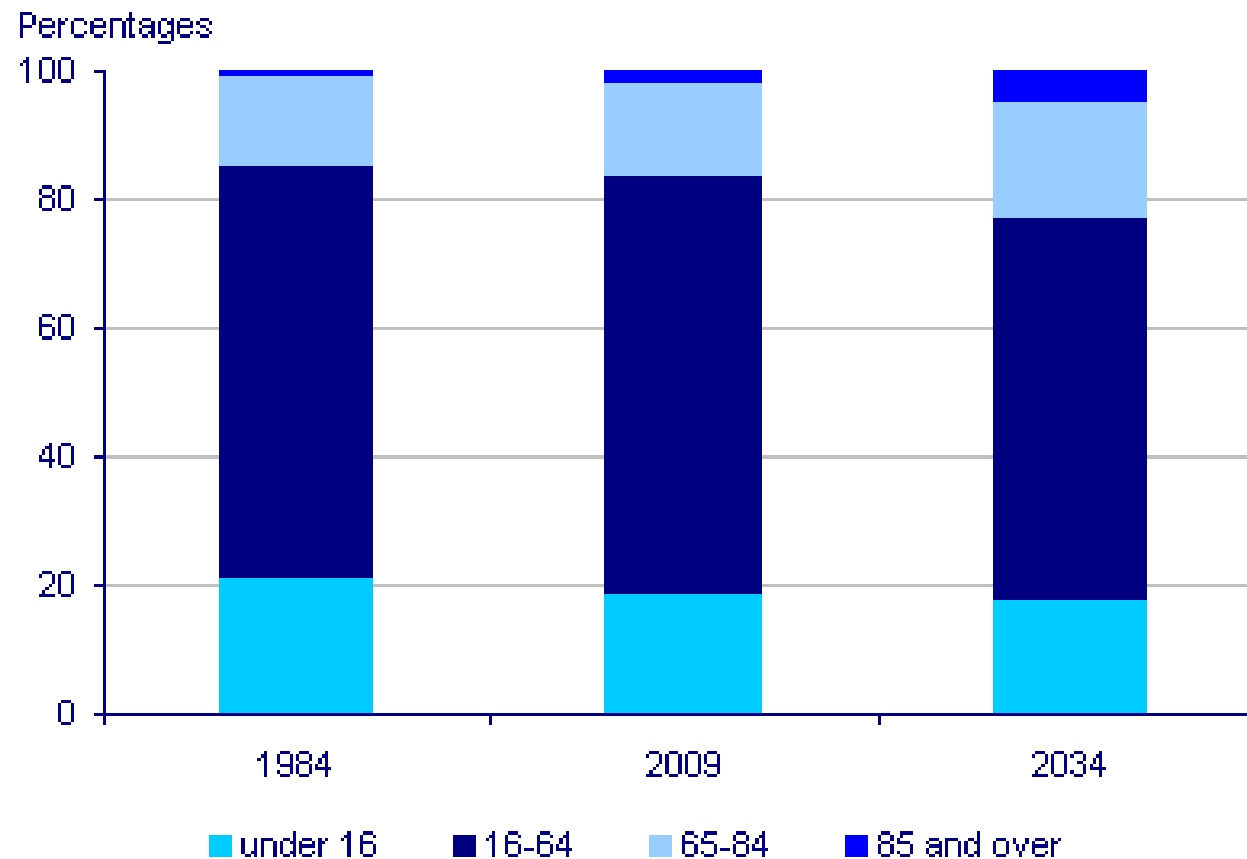
Exemplo 2.4 – Comparando Distribuições



Fonte: IBGE, Diretoria de Pesquisas, Coordenação de Trabalho e Rendimento, Pesquisa Nacional por Amostra de Domicílios 2015.

Como se poderia melhorar este gráfico?

Exemplo 2.5 - Envelhecimento populacional no Reino Unido

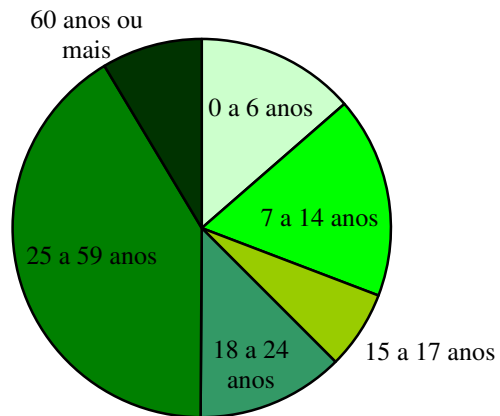


População por faixas de idade, Reino Unido, 1984, 2009 e 2034.

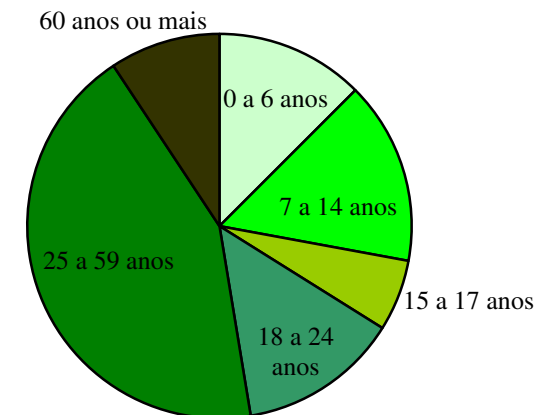
<http://www.statistics.gov.uk/cci/nugget.asp?id=949>

Exemplo 2.6 - Gráficos de Setores para Comparar Distribuições

Distribuição etária - 1997



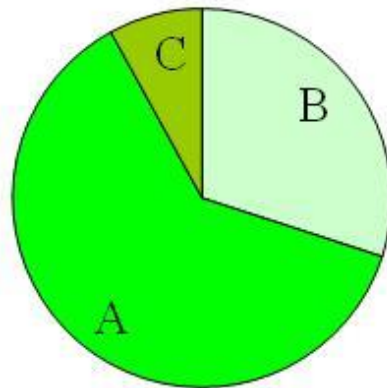
Distribuição etária - 2002



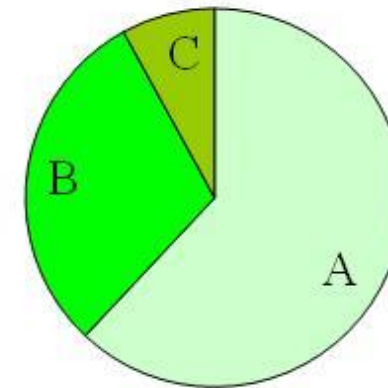
Fica evidente deste exemplo que gráficos de setores não servem para comparar distribuições como os gráficos de barras ou colunas justapostas.

Gráficos do Exercício 2.1

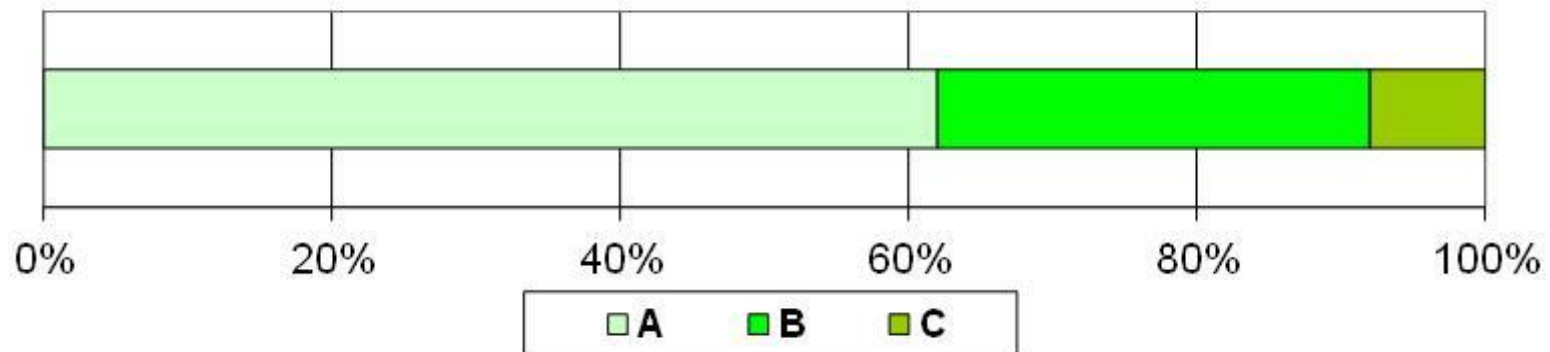
Distribuição de frequências das classes



Distribuição de frequências das classes



Distribuição de frequências das classes



Resumo sobre gráficos para distribuições de uma variável categórica nominal

1. Prefira gráficos de **barras** ou **colunas**, conforme a dimensão do espaço de apresentação.
2. Pessoas têm melhor capacidade de perceber e comparar tamanhos relativos em **áreas de retângulos** do que em áreas de círculos.
3. Portanto, se possível, **evite gráficos de setores**, pizza, torta ou rosca.
4. **Jamais aplique efeitos 3D** seja em gráficos de tortas ou em gráficos de barras ou colunas.
5. **Evite separar fatias** num gráfico de tortas, ou inventar formas alternativas.

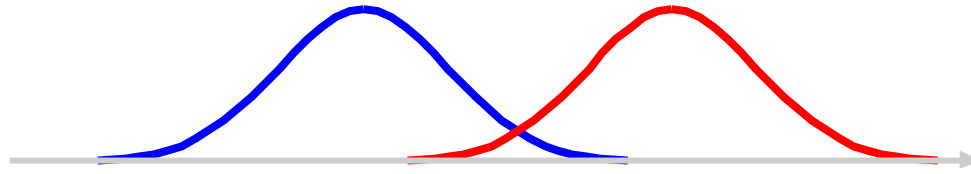
Gráficos em Estatística

Parte 3

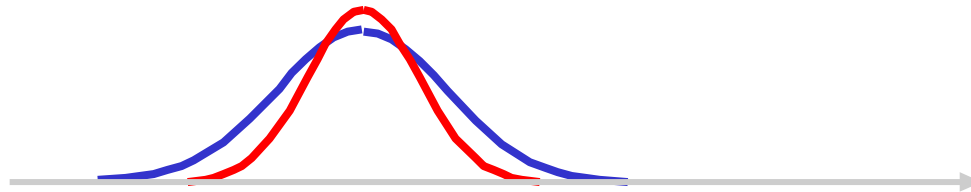
Gráficos para dados numéricos

Dados Numéricos: Aspectos a Destacar

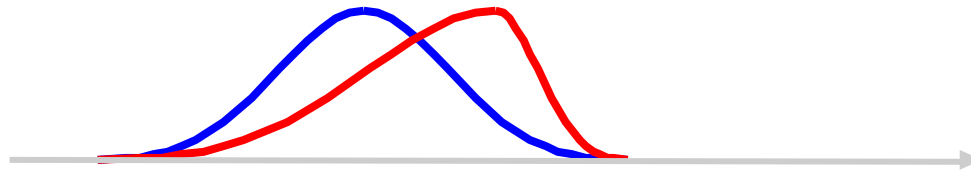
Posição,
localização,
centro



Variabilidade,
dispersão,
espalhamento



Forma (simetria
vs. assimetria)



Gráficos Para Dados Numéricos

Há muitas maneiras de apresentar **dados numéricos** em gráficos.

Com apenas **uma variável**, os principais gráficos são:

- Gráfico de barras ou colunas → dados discretos;
- Gráfico de pontos (*Dotplot*) → $n \leq 50$ valores;
- Ramo-e-folhas (*Stem-and-leaf*) → $n \leq 100$ valores;
- Gráfico de caixas (*Boxplot*) → n qualquer;
- Histograma → $n > 100$ valores.

Na sequência, vamos estudar cada um destes tipos de gráficos.

Dados Numéricos Discretos

Quando os dados da variável são **discretos**, a informação completa sobre os dados é resumida pela **distribuição de frequências**.

Portanto o tipo de gráfico adequado é **similar ao gráfico de colunas** (ou de barras) usado para variáveis categóricas nominais, **mas há uma diferença importante**.

Para dados numéricos discretos as **colunas ou barras** representando as frequências são dispostas ao longo do eixo dos valores, **sem espaços entre elas, e mantendo a ordem natural dos valores**.

Isto ocorre porque os valores cujas frequências se quer mostrar agora **são números**, que devem ser representados ao longo de um **eixo cartesiano**.

Exemplo 3.1 – Gols Marcados por Partida

Tabela 3.1 – Gols marcados por partida – Registro das primeiras 20 partidas Campeonato de futebol do estado do Rio de Janeiro 2010

Data	Mandante	Escore	Visitante	GolsTotal	GolsCasa	GolsFora
16/01/2010	Bangu	0x3	Boavista	3	0	3
16/01/2010	Olaria	2x2	Volta Redonda	4	2	2
16/01/2010	Madureira	2x1	América RJ	3	2	1
16/01/2010	Macaé	2x3	Botafogo	5	2	3
16/01/2010	Vasco	1x0	Tigres do Brasil	1	1	0
17/01/2010	Flamengo	3x2	Duque de Caxias	5	3	2
17/01/2010	Americano	0x3	Fluminense	3	0	3
17/01/2010	Friburguense	0x0	Resende	0	0	0
20/01/2010	Boavista	2x0	Americano	2	2	0
20/01/2010	Volta Redonda	1x3	Flamengo	4	1	3
20/01/2010	Fluminense	3x0	Bangu	3	3	0
20/01/2010	Resende	3x3	Macaé	6	3	3
20/01/2010	América RJ	1x2	Vasco	3	1	2
20/01/2010	Tigres do Brasil	2x1	Madureira	3	2	1
21/01/2010	Duque de Caxias	1x2	Olaria	3	1	2
21/01/2010	Botafogo	2x0	Friburguense	2	2	0
23/01/2010	Bangu	1x2	Flamengo	3	1	2
23/01/2010	Resende	1x2	Madureira	3	1	2
23/01/2010	Macaé	2x3	América RJ	5	2	3
24/01/2010	Fluminense	1x0	Volta Redonda	1	1	0

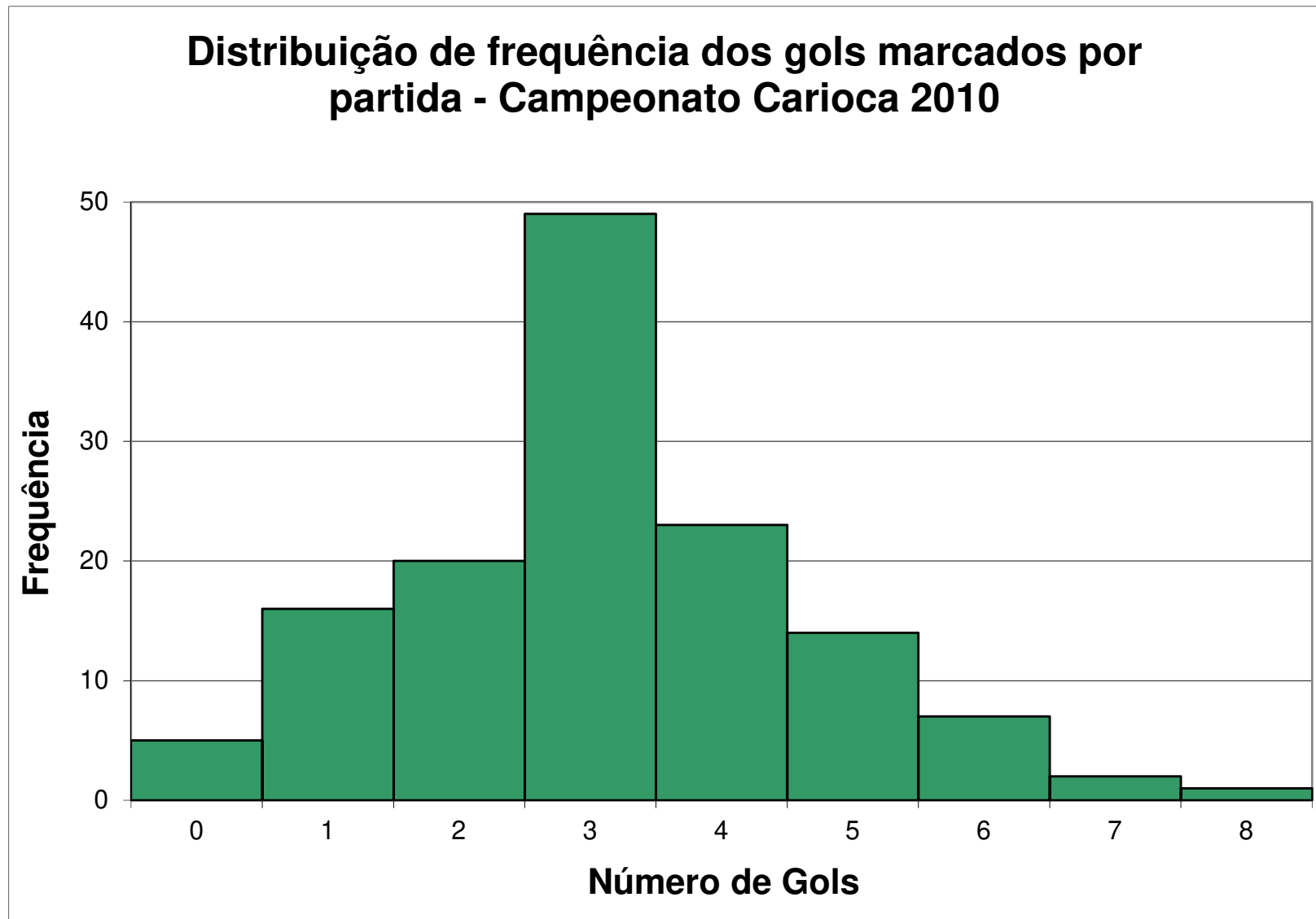
A distribuição de frequências do número de gols marcados por partida (nossa variável) é dada na tabela 3.2.

Exemplo 3.1 – Gols Marcados por Partida

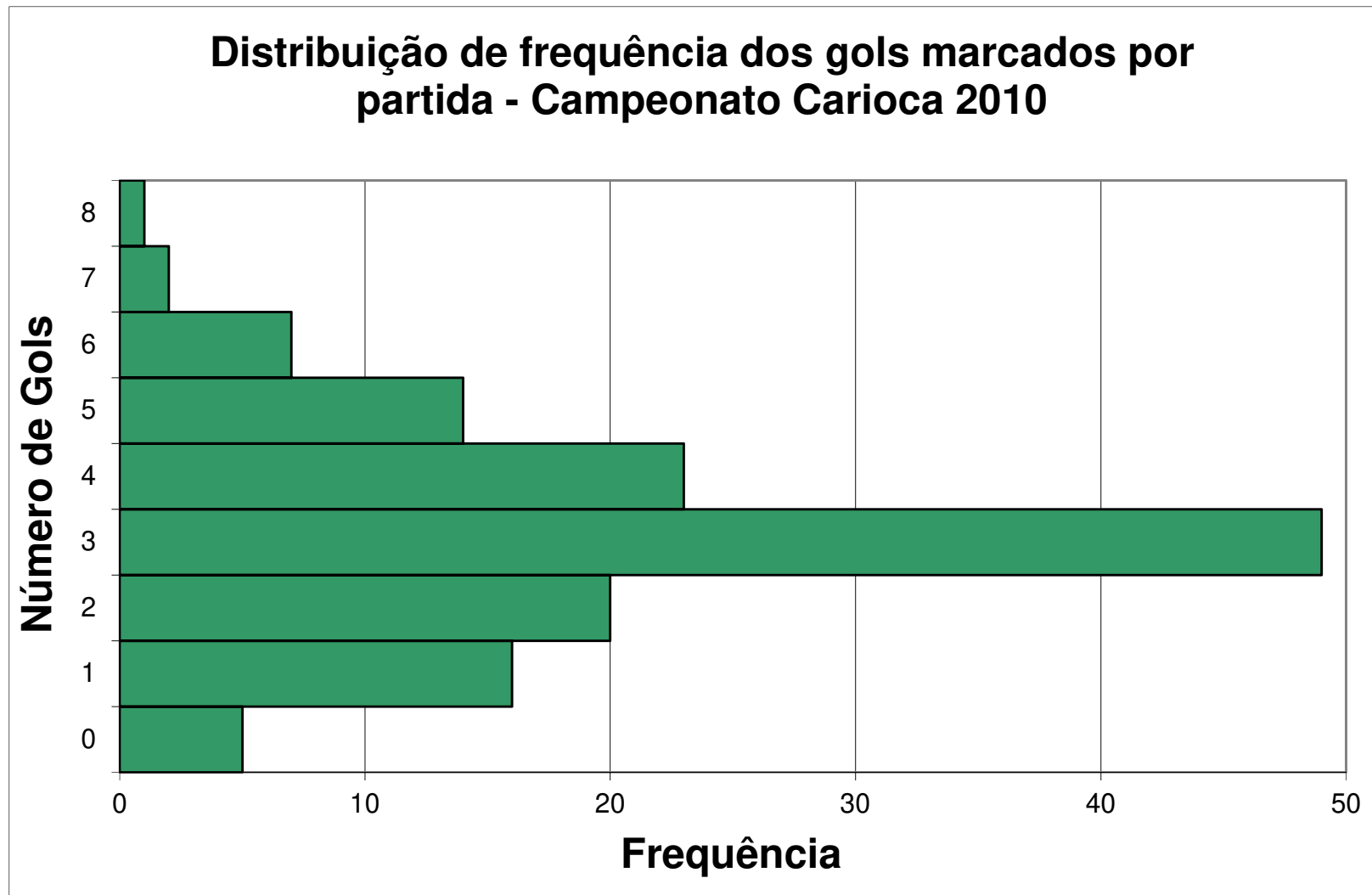
Tabela 3.2 – Distribuição de frequências dos gols marcados por partida no campeonato de futebol do estado do Rio de Janeiro 2010

Gols	Frequência	Frequência Relativa (%)
0	5	3,6
1	16	11,7
2	20	14,6
3	49	35,8
4	23	16,8
5	14	10,2
6	7	5,1
7	2	1,5
8	1	0,7
Total	137	100,0

Exemplo 3.1 – Gols Marcados por Partida



Exemplo 3.1 – Gols Marcados por Partida



Dados Numéricos ‘Contínuos’

Quando os dados da variável **são contínuos**, o conjunto de valores observados tende a ser numeroso.

Isto implica que obter a **distribuição de frequências** para resumir a informação sobre os dados **não é eficiente**.

Por este motivo, **para dados numéricos contínuos**, é comum **usar resumos numéricos** das propriedades da distribuição.

Além disto, é também conveniente usar alguns **tipos de gráficos alternativos** que funcionam bem quando os dados são contínuos.

O primeiro destes tipos de gráficos é chamado de **Gráfico de Pontos** ou **‘Dotplot’** em inglês.

Gráfico de pontos ('Dotplot')

Neste gráfico, **cada valor** da variável é **representado por um ponto, marcado ao longo de um eixo** (geralmente horizontal) onde está a escala das medidas da variável.

Este tipo de gráfico **funciona bem quando não há muitos valores** a apresentar ($n \leq 50$ é ideal).

Se os dados forem muito numerosos, evite usar este tipo de gráfico.

Este gráfico não está disponível 'diretamente' no Excel, embora seja possível construí-lo com algum trabalho.

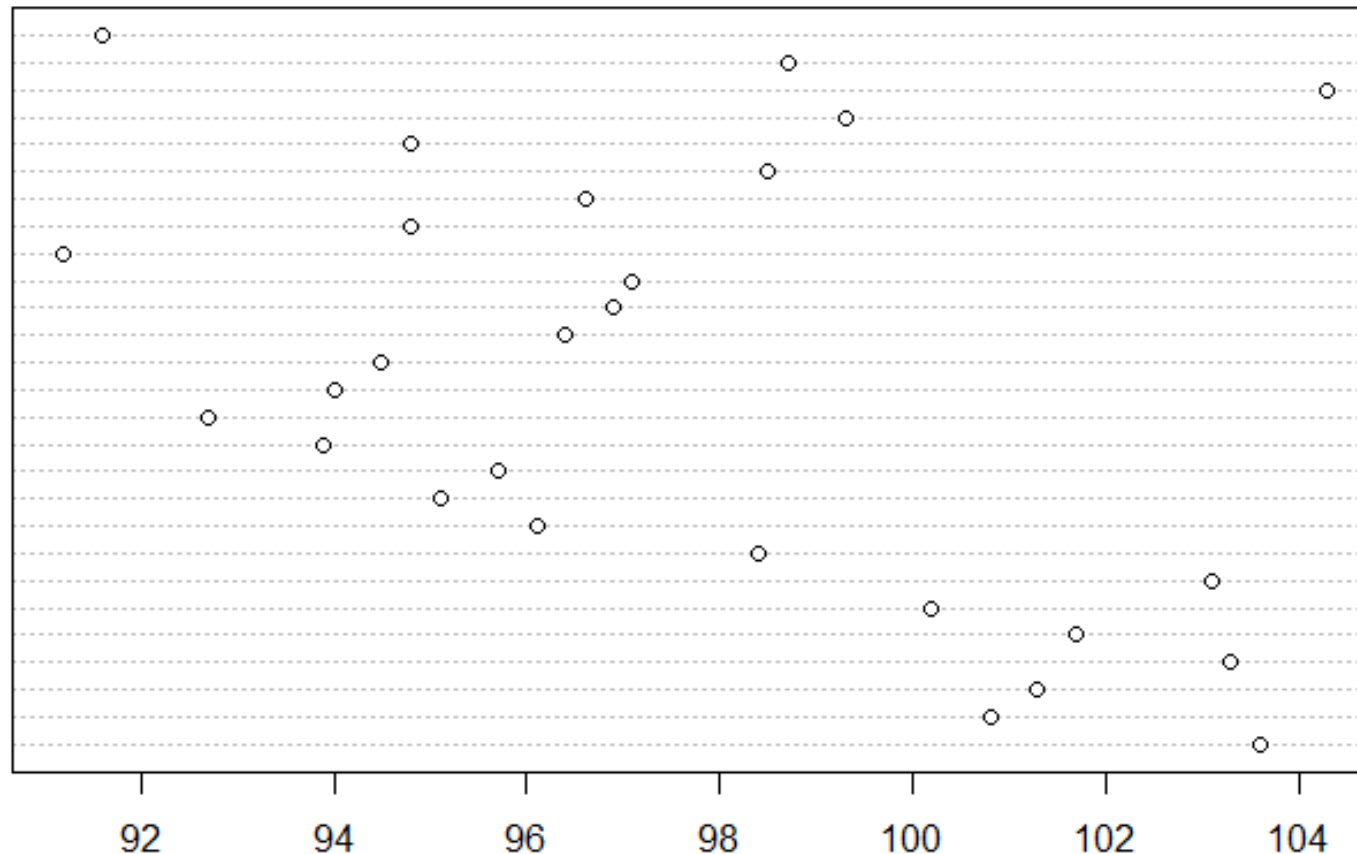
No sistema R, a função para obter este gráfico é '**dotchart**', disponível na biblioteca '**graphics**'.

Exemplo 3.2 – Razão de sexo por UF – Censo 2010

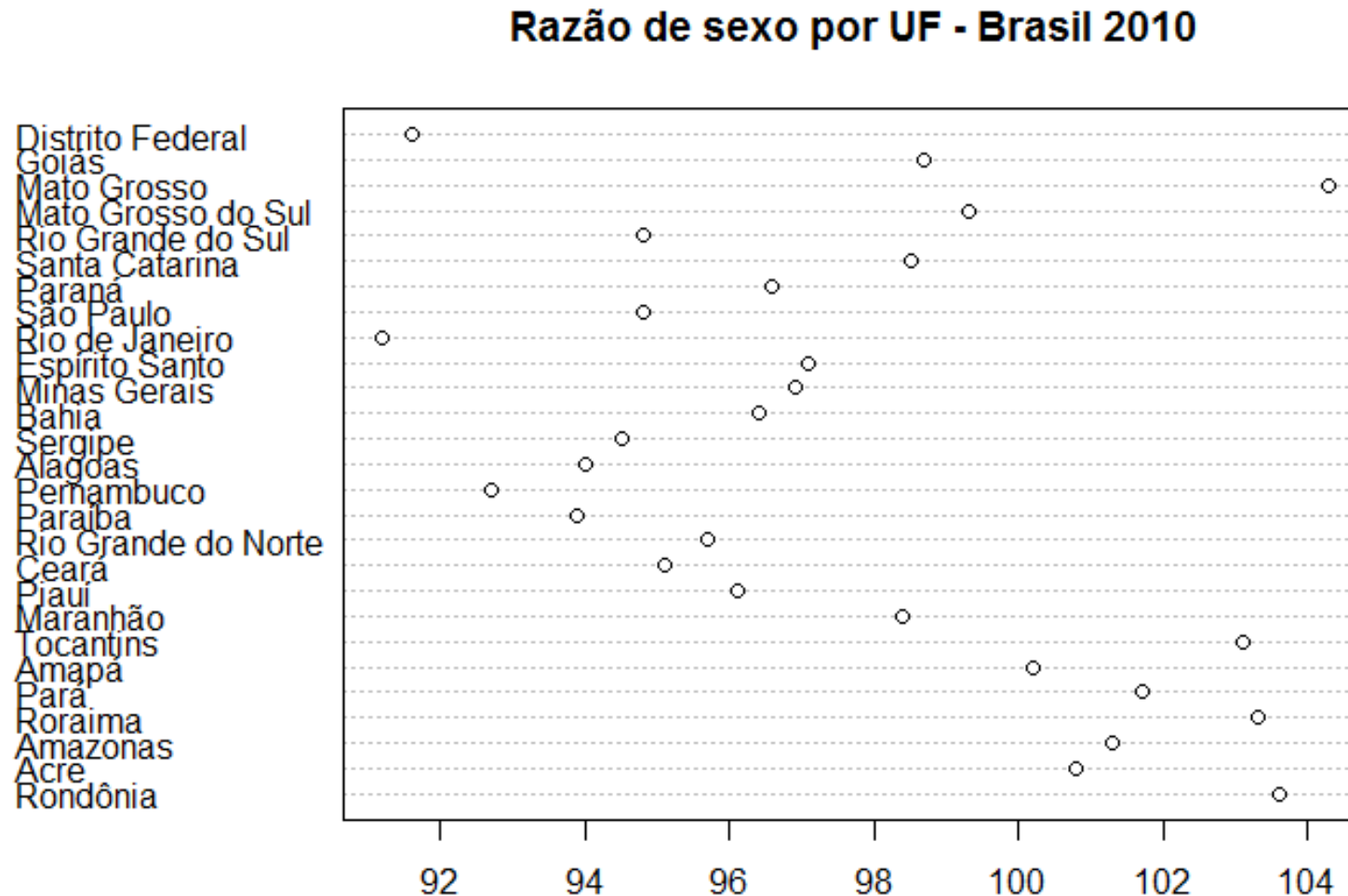
Unidade da Federação	Razao de Sexo	Unidade da Federação	Razao de Sexo
Rondônia	103,6	Alagoas	94,0
Acre	100,8	Sergipe	94,5
Amazonas	101,3	Bahia	96,4
Roraima	103,3	Minas Gerais	96,9
Pará	101,7	Espírito Santo	97,1
Amapá	100,2	Rio de Janeiro	91,2
Tocantins	103,1	São Paulo	94,8
Maranhão	98,4	Paraná	96,6
Piauí	96,1	Santa Catarina	98,5
Ceará	95,1	Rio Grande do Sul	94,8
Rio Grande do Norte	95,7	Mato Grosso do Sul	99,3
Paraíba	93,9	Mato Grosso	104,3
Pernambuco	92,7	Goiás	98,7
		Distrito Federal	91,6

Exemplo 3.2a – Razão de sexo por UF – Censo 2010

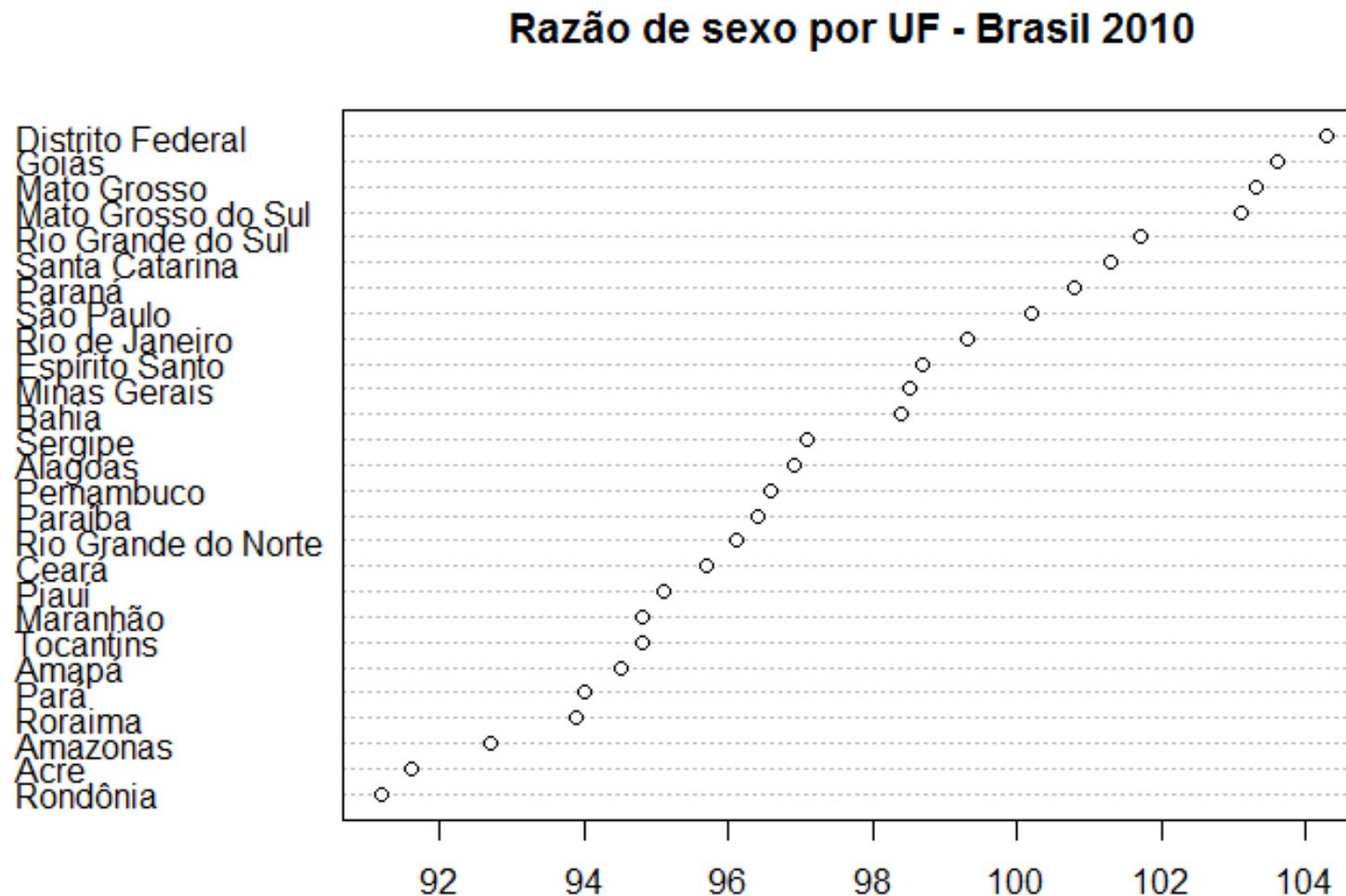
Razão de sexo por UF - Brasil 2010



Exemplo 3.2b – Razão de sexo por UF – Censo 2010



Exemplo 3.2c – Razão de sexo por UF – Censo 2010



Ramo e folhas

Ramo e folhas é um gráfico **fácil e rápido** que ajuda a **organizar** uma lista de números, ao mesmo tempo em que se obtém uma ideia da **forma de sua distribuição**.

Este gráfico é **adequado quando o número de dados a resumir não é muito grande** ($n \leq 100$).

Trata-se, na verdade, de uma versão ‘econômica’ de um **histograma**.

A maneira mais rápida de descrever um ramo e folhas é fazer um.

O **exemplo 3.3** apresenta uma tabela com as probabilidades de morrer antes de completar 5 anos para crianças do sexo masculino, estimadas para os países das Américas, pela UNSD em 1999.

Exemplo 3.3 – Probabilidade de morrer antes de completar 5 anos por país – crianças do sexo masculino - 1999

País	Prob. de Morrer Antes dos 5 Anos - Homens (Por Mil)	País	Prob. de Morrer Antes dos 5 Anos - Homens (Por Mil)
Antigua e Barbuda	25	Guatemala	54
Argentina	24	Guiana	77
Bahamas	14	Haiti	111
Barbados	13	Honduras	45
Belize	37	Jamaica	17
Bolívia	88	México	31
Brasil	49	Nicarágua	49
Canadá	6	Panamá	28
Chile	12	Paraguai	35
Colômbia	29	Peru	53
Costa Rica	18	Saint Kitts and Nevis	25
Cuba	9	Santa Lucia	20
Dominica	14	Saint Vincent and the Grenadines	20
República Dominicana	55	Suriname	29
Equador	41	Trinidad e Tobago	16
El Salvador	40	Estados Unidos da América	9
Granada	25	Uruguai	19
		Venezuela	26

Exemplo 3.3a – Ramo e folhas para a probabilidade de morrer (por mil) antes de completar 5 anos por país – crianças do sexo masculino - 1999

Ramo	Folhas									
0	6	9	9							
1	2	3	4	4	6	7	8	9		
2	0	0	4	5	5	5	6	8	9	9
3	1	5	7							
4	0	1	5	9	9					
5	3	4	5							
6										
7	7									
8	8									
9										
10										
11	1									

Exemplo 3.3 – ‘Leitura’ do gráfico

1. **Concentração** de valores entre 10 e 30.
2. Distribuição dos valores é **assimétrica**, com cauda mais longa à direita.
3. Maior valor (Haiti) é quase 20 vezes maior que o menor valor (Canadá).
4. Dois grupos de **valores atípicos** (Guiana e Bolívia + Haiti).
5. Taxa do Brasil está no **quarto superior** da distribuição.

Todos estes fatos poderiam fazer parte da **estória estatística** a ser contada com base nos dados representados neste gráfico.

Ramo e folhas – Resumindo os passos

Passo 1: criação dos ramos.

- Divida o intervalo de dados em partes iguais, para criar os **ramos**.
- A ideia geral é usar os **primeiros dígitos** de cada número como ramo.
- Coloque os ramos “**empilhados**” numa coluna vertical.

Passo 2: adição das folhas.

- Cada **folha** representa uma **observação**.
- O **próximo dígito** (após o primeiro) no número é usado como folha.
- Se o número tiver mais algarismos, estes são simplesmente **descartados**.
- Geralmente, faz-se **ordenação das folhas** dentro de cada ramo.

Ramo e folhas

Como se pode ver, os **ramos** fazem o papel das **classes de valores**.

As **folhas** são usadas para indicar a **frequência de observações** em cada uma das classes (como num **histograma**).

Além disso, no **ramo e folhas** é possível **recompor os dados originais** (sem identificar os países, claro), pois a **junção do ramo com a folha retorna um valor observado** da medida representada.

O exemplo 3.3 ilustra bem porque este gráfico não é adequado quando há muitos dados.

Gráficos tipo ramo e folhas são facilmente elaborados usando **sistemas estatísticos**, mas não estão disponíveis no Excel.

O comando para obter estes gráficos no sistema R é '**stem**', disponível na biblioteca '**graphics**'.

Exemplo 3.3a – Ramo e folhas obtido no R

```
> stem(Ex.3.3$Taxa, scale=2)
```

The decimal point is 1 digit(s) to the right of the |

```
0 | 699
1 | 23446789
2 | 0045556899
3 | 157
4 | 01599
5 | 345
6 |
7 | 7
8 | 8
9 |
10 |
11 | 1
```

Gráfico de Caixas (Boxplot)

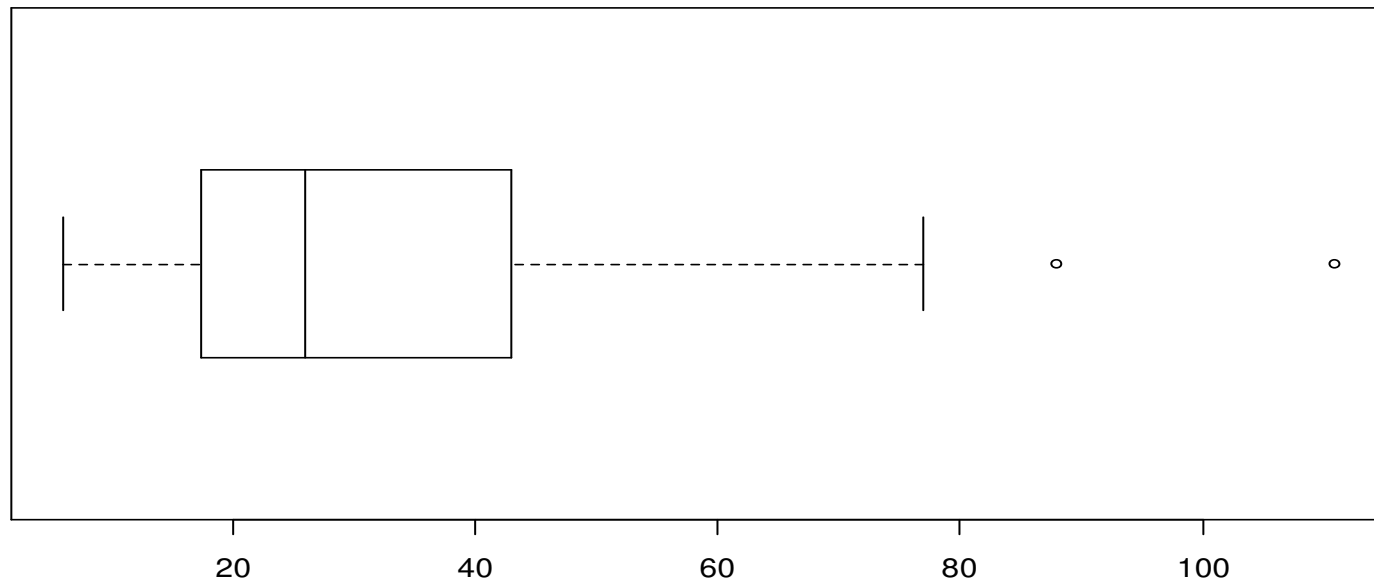
Representação gráfica do **esquema de cinco números**, com alguns **detalhes adicionais**.

Permite visualização rápida do **centro**, **dispersão** e **(as)simetria** dos dados, mais eventual presença de **valores atípicos**.

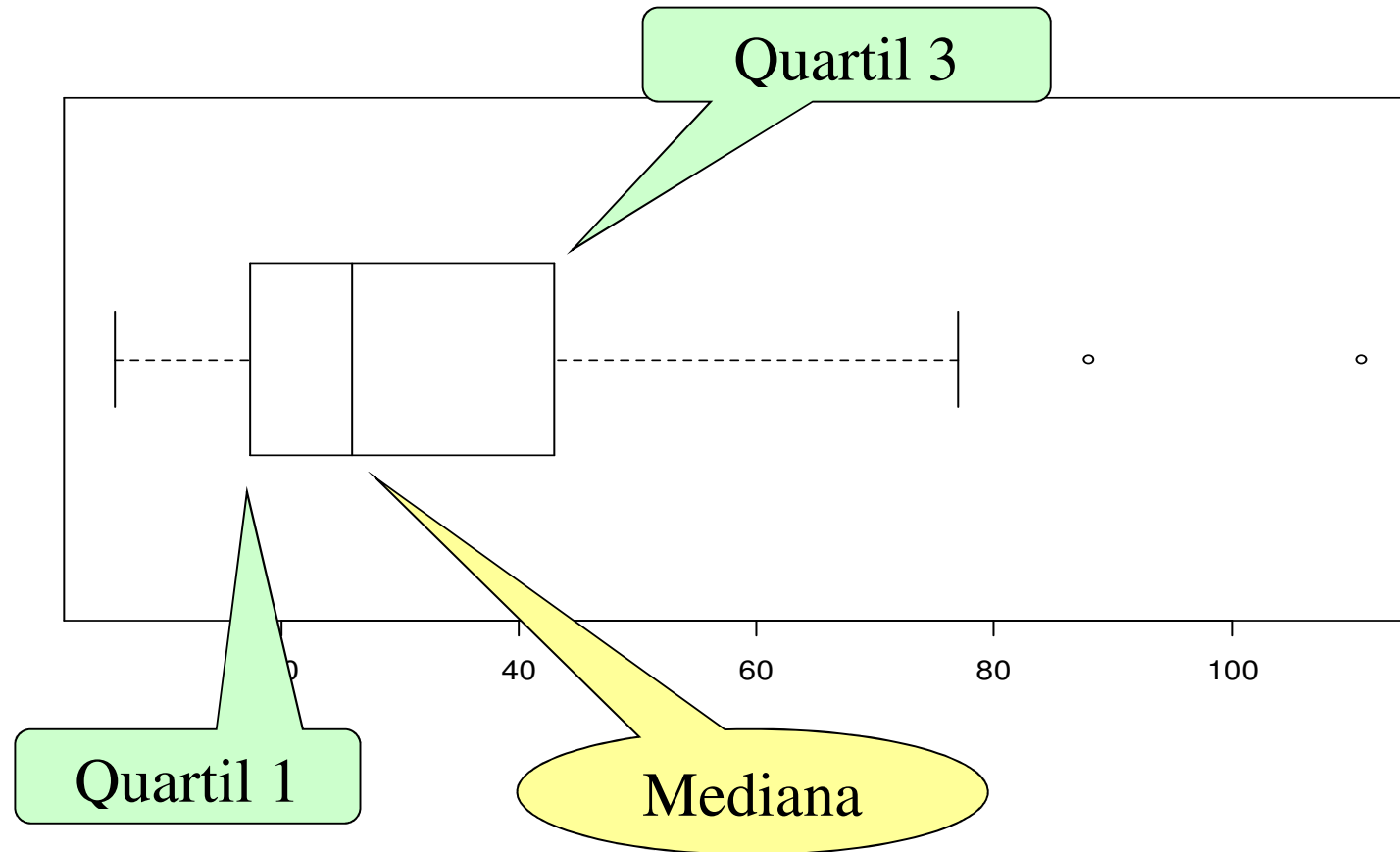
Muito útil para **comparar dados de vários grupos**.

Popularizado por **John Tukey** a partir dos anos 1970 em seu livro '*Exploratory Data Analysis*'.

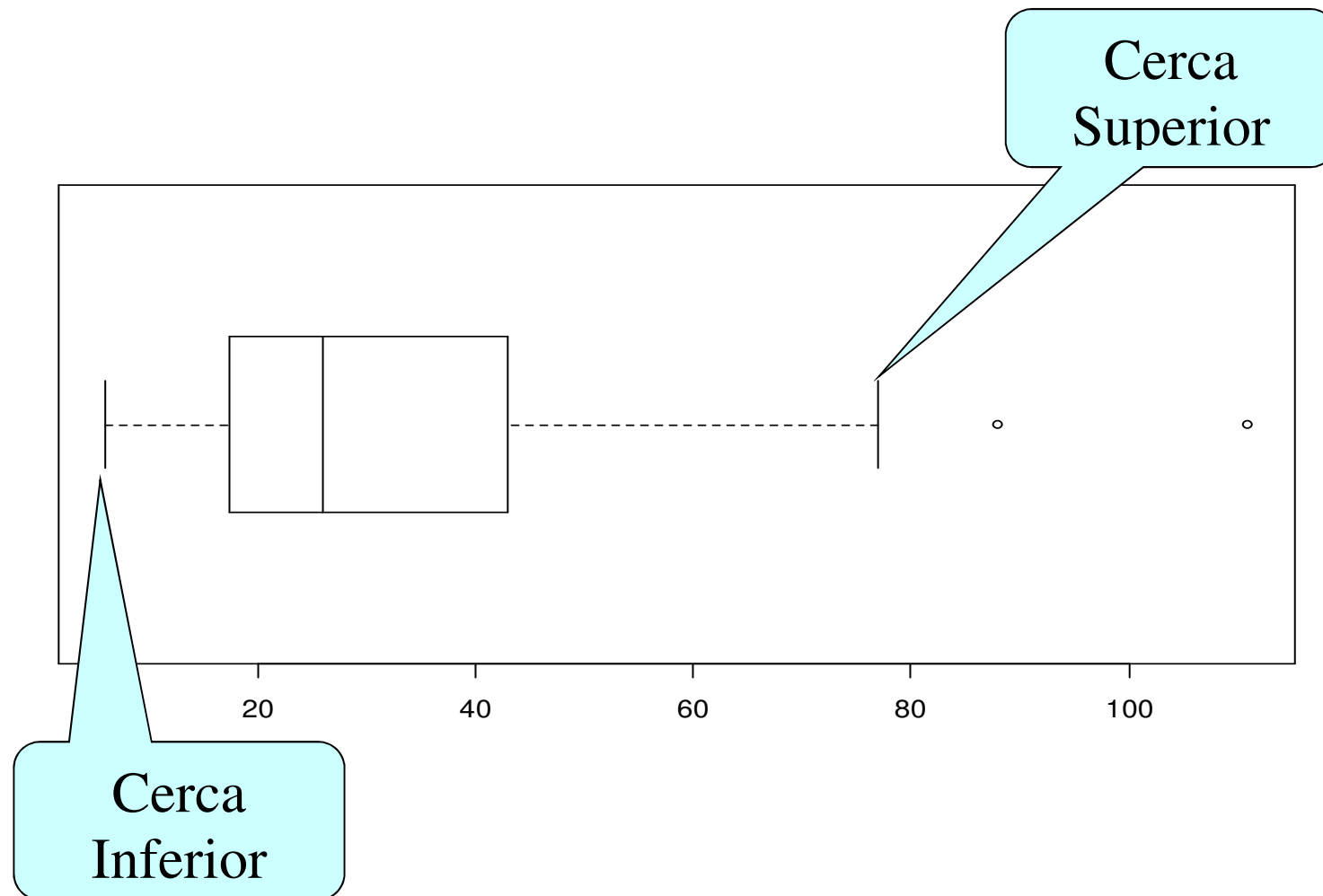
Exemplo 3.3b – Gráfico de Caixas para a probabilidade de morrer antes de completar 5 anos por país – crianças do sexo masculino - 1999



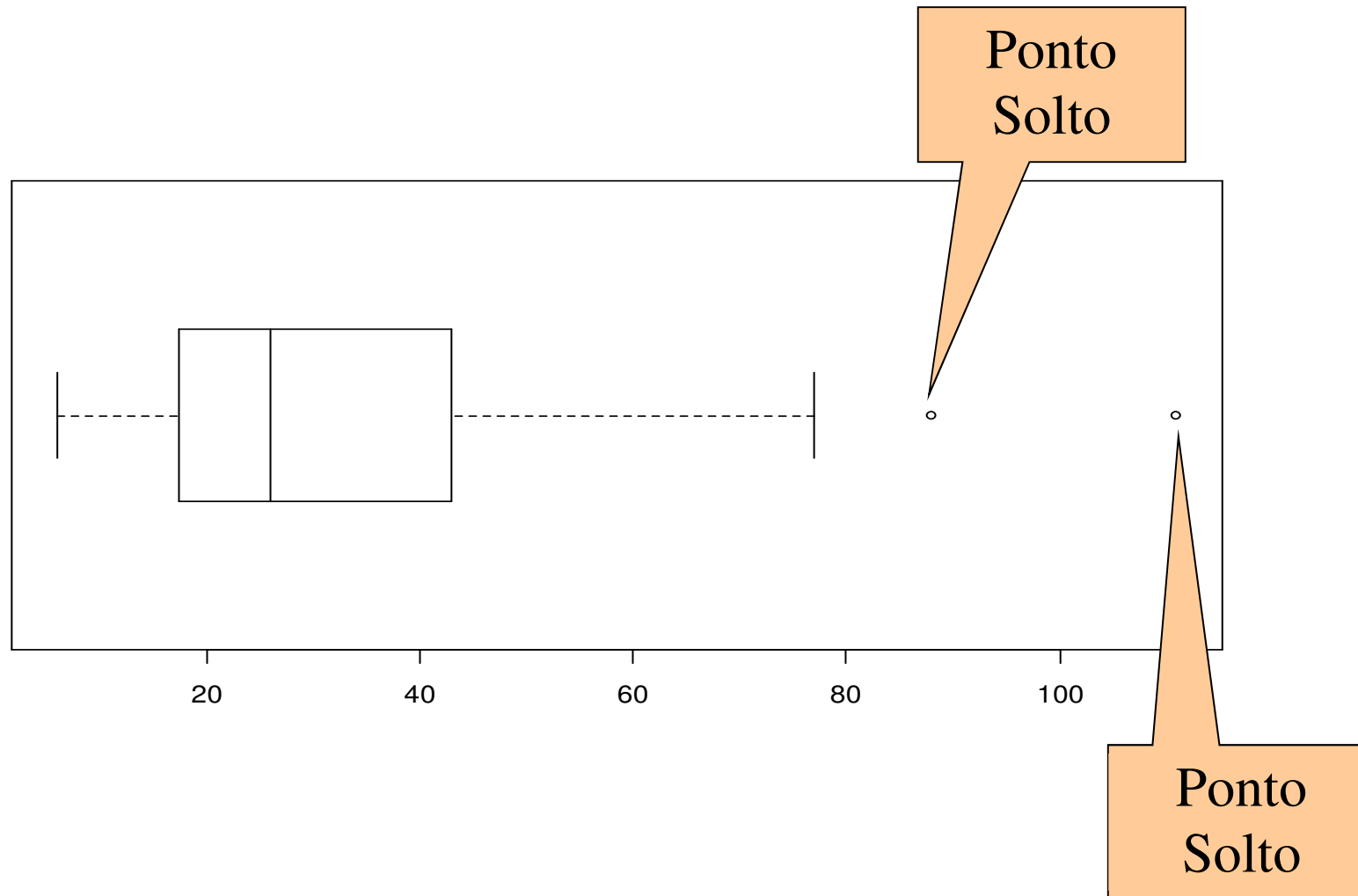
Exemplo 3.3b – Gráfico de Caixas para a probabilidade de morrer antes de completar 5 anos por país – crianças do sexo masculino - 1999



Exemplo 3.3b – Gráfico de Caixas para a probabilidade de morrer antes de completar 5 anos por país – crianças do sexo masculino - 1999



Exemplo 3.3b – Gráfico de Caixas para a probabilidade de morrer antes de completar 5 anos por país – crianças do sexo masculino - 1999



Exemplo 3.3b – Gráfico de Caixas para a probabilidade de morrer antes de completar 5 anos por país – crianças do sexo masculino - 1999

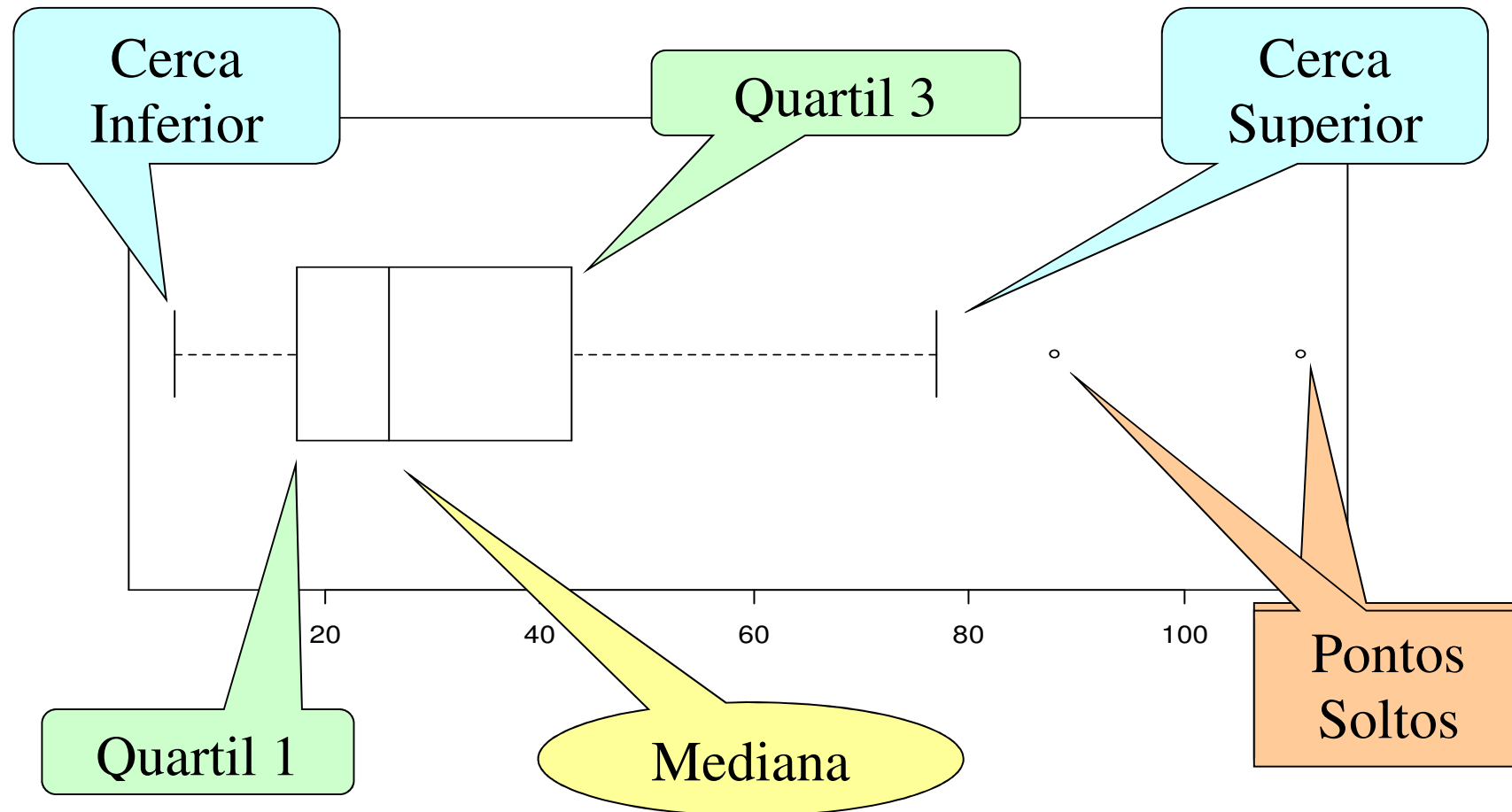


Gráfico de Caixas (Boxplot)

O comprimento da caixa é igual à amplitude interquartílica ou intervalo interquartílico.

- Comprimento da Caixa = $Q3 - Q1$

As **cercas inferior e superior** são desenhadas para representar um intervalo dentro do qual os valores são considerados ‘**típicos**’.

Valores (pontos) fora das cercas ocorrem com frequência baixa
→ são considerados **atípicos**.

- Cerca Superior = $\text{Min}[Q3 + 1,5 \times (Q3 - Q1); \text{Max.}]$
- Cerca Inferior = $\text{Max}[Q1 - 1,5 \times (Q3 - Q1); \text{Min.}]$

Construindo Gráficos de Caixas

1. Desenhe uma **caixa** com os extremos nos **quartis inferior e superior**.
2. Desenhe uma **linha** na caixa onde fica a **mediana**.
3. Calcule a **largura da caixa**, isto é, a **amplitude interquartílica**.
4. Desenhe as **cercas inferior e superior**, fora da caixa, com afastamento igual a 1,5 vezes a amplitude interquartílica do quartil mais próximo. Se o mínimo ou máximo ocorrer antes da cerca, pare aí.
5. **Destaque pontos** que caiam **fora das cercas**.

Gráfico de Caixas (Boxplot)

Este gráfico não é facilmente elaborado com Excel.

Mas está disponível em vários sistemas estatísticos.

No sistema R, o comando para elaborar este gráfico é '**boxplot**', disponível na biblioteca '**graphics**'.

Interpretando Gráficos de Caixas

Boxplots **dividem** os **dados** em quatro **quartos**.

Dados **abaixo da** caixa formam o **primeiro quarto**.

Dados na **primeira metade da caixa**, o **segundo quarto**, e assim por diante.

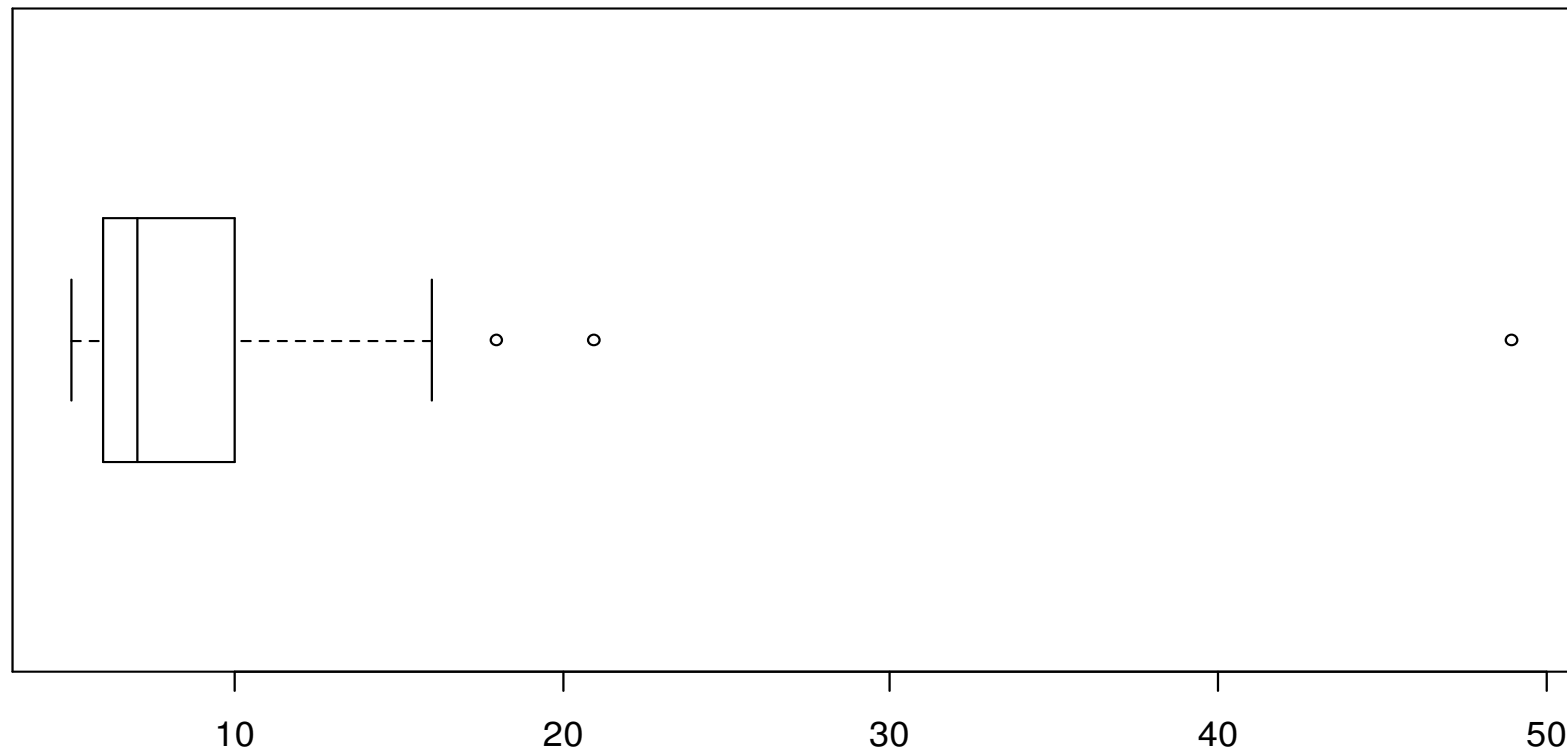
A **posição dos quartis** permite **localizar** os dados no eixo dos valores.

O **afastamento dos quartis**, e o **comprimento das cercas** permitem examinar a **dispersão dos dados**.

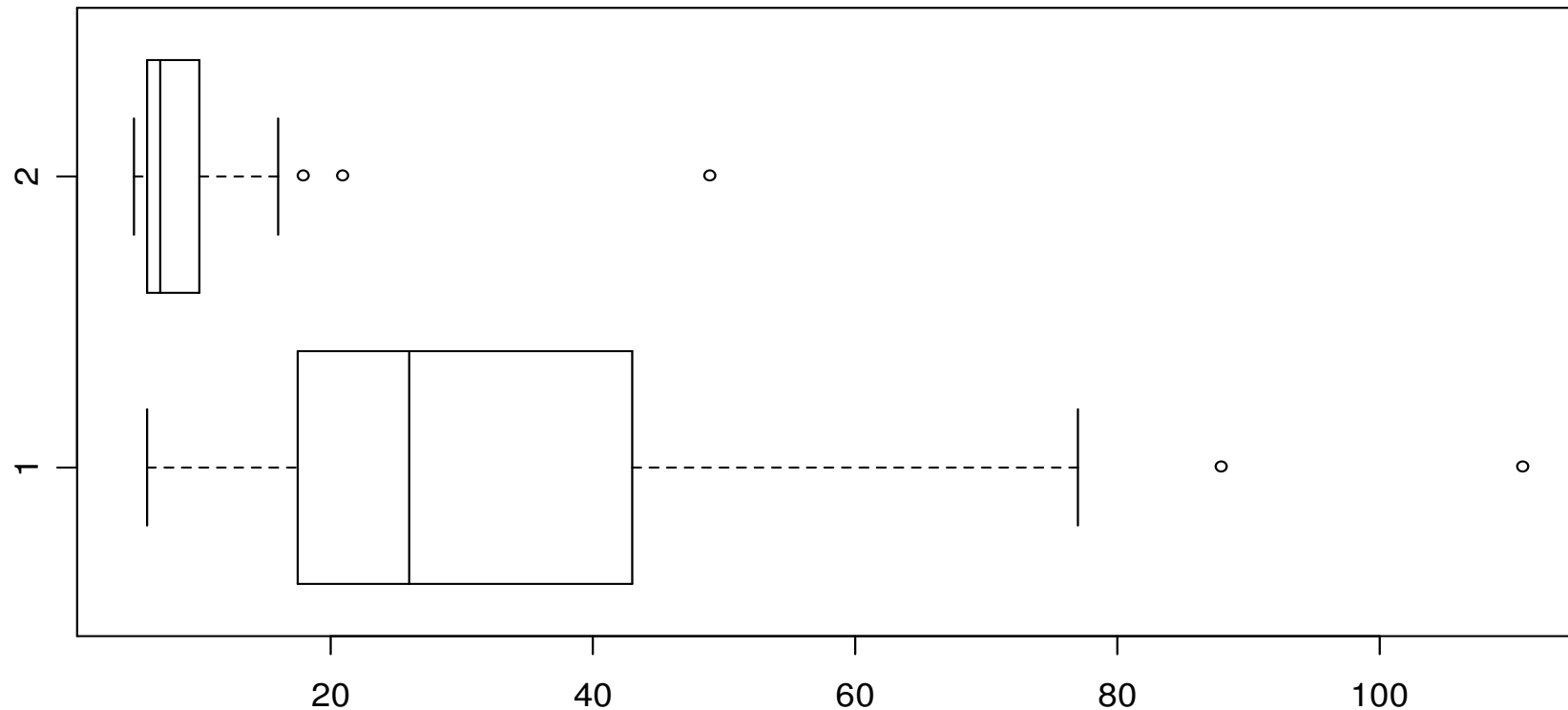
A **distância** entre os limites dos quartos à esquerda e à direita da mediana permite verificar a **simetria** (ou **assimetria**) dos dados.

Dados atípicos são facilmente identificados.

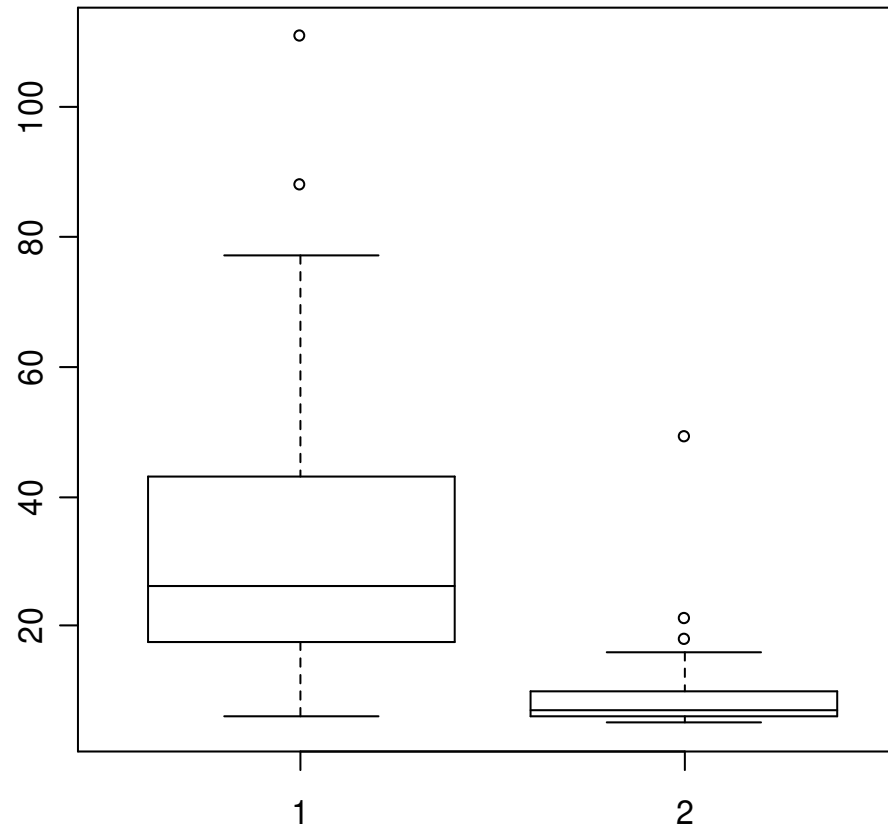
Exemplo 3.4 – Gráfico de Caixas para a probabilidade de morrer antes de completar 5 anos por país – crianças do sexo masculino – 1999 – Países da Europa Ocidental



Exemplo 3.5a – Gráficos de Caixas para a comparar a probabilidade de morrer antes de completar 5 anos por país – crianças do sexo masculino – 1999 – Países das Américas (1) e da Europa Ocidental (2)



Exemplo 3.5b – Gráficos de Caixas para a comparar a probabilidade de morrer antes de completar 5 anos por país – crianças do sexo masculino – 1999 – Países das Américas (1) e da Europa Ocidental (2)



Histograma

Gráfico **parecido** com o ramo e folhas.

Ideia é **revelar a distribuição de frequência dos dados**, mas **suprimindo detalhes** que não ajudem a **‘enxergar o todo’**.

Para **grandes conjuntos de dados**, o **histograma é melhor que ramo e folhas**, porque não lista cada valor.

Histograma **não funciona tão bem** quando há **poucos dados** ($N < 100$). Nestes casos prefira o ramo e folhas.

Histograma: Como Elaborar

Passo 1: criação das classes.

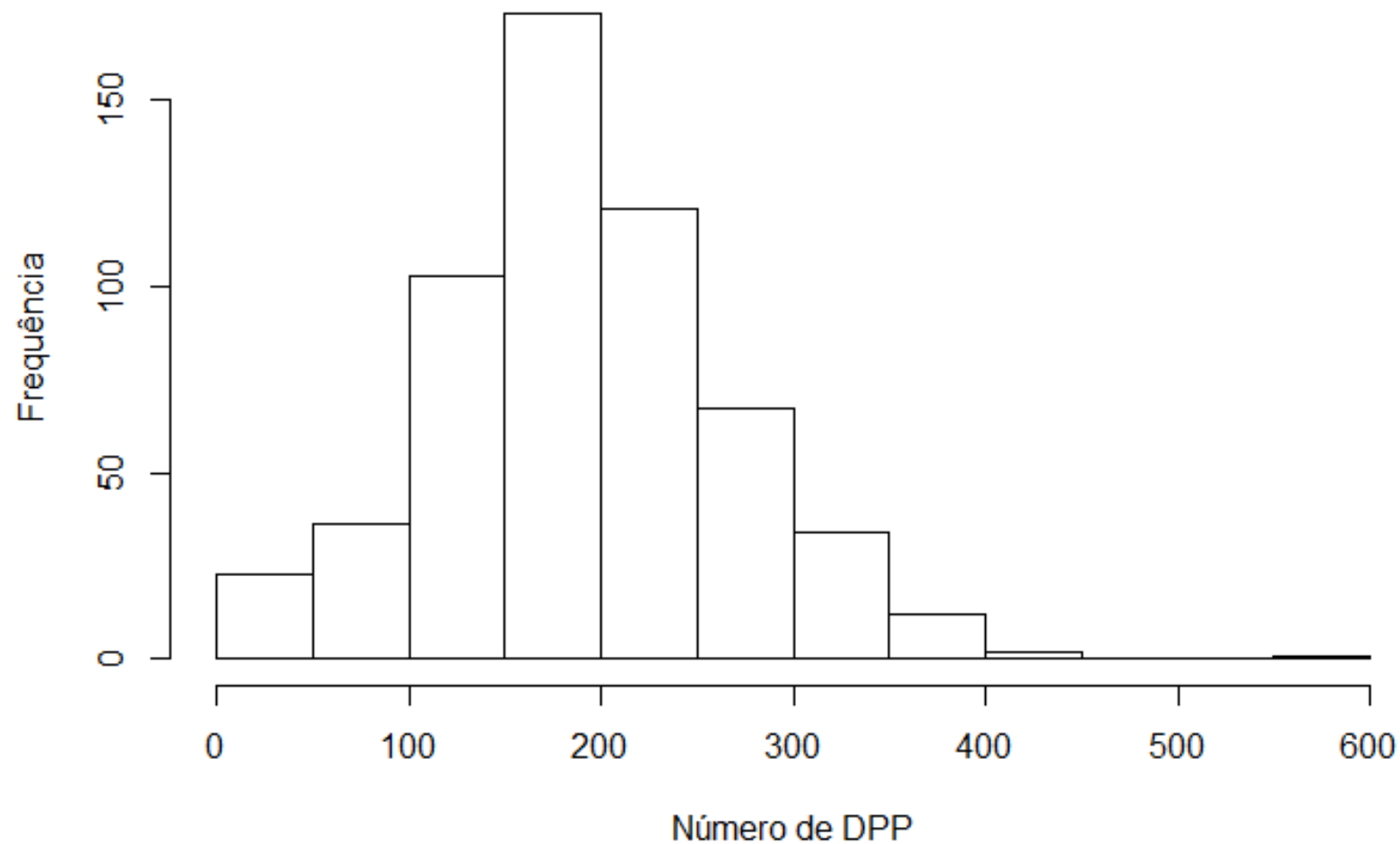
- **Divida** o intervalo de dados **em partes iguais**, para criar as **classes**.
- A ideia geral é similar à usada para definição dos **ramos**.

Passo 2: contagem das frequências.

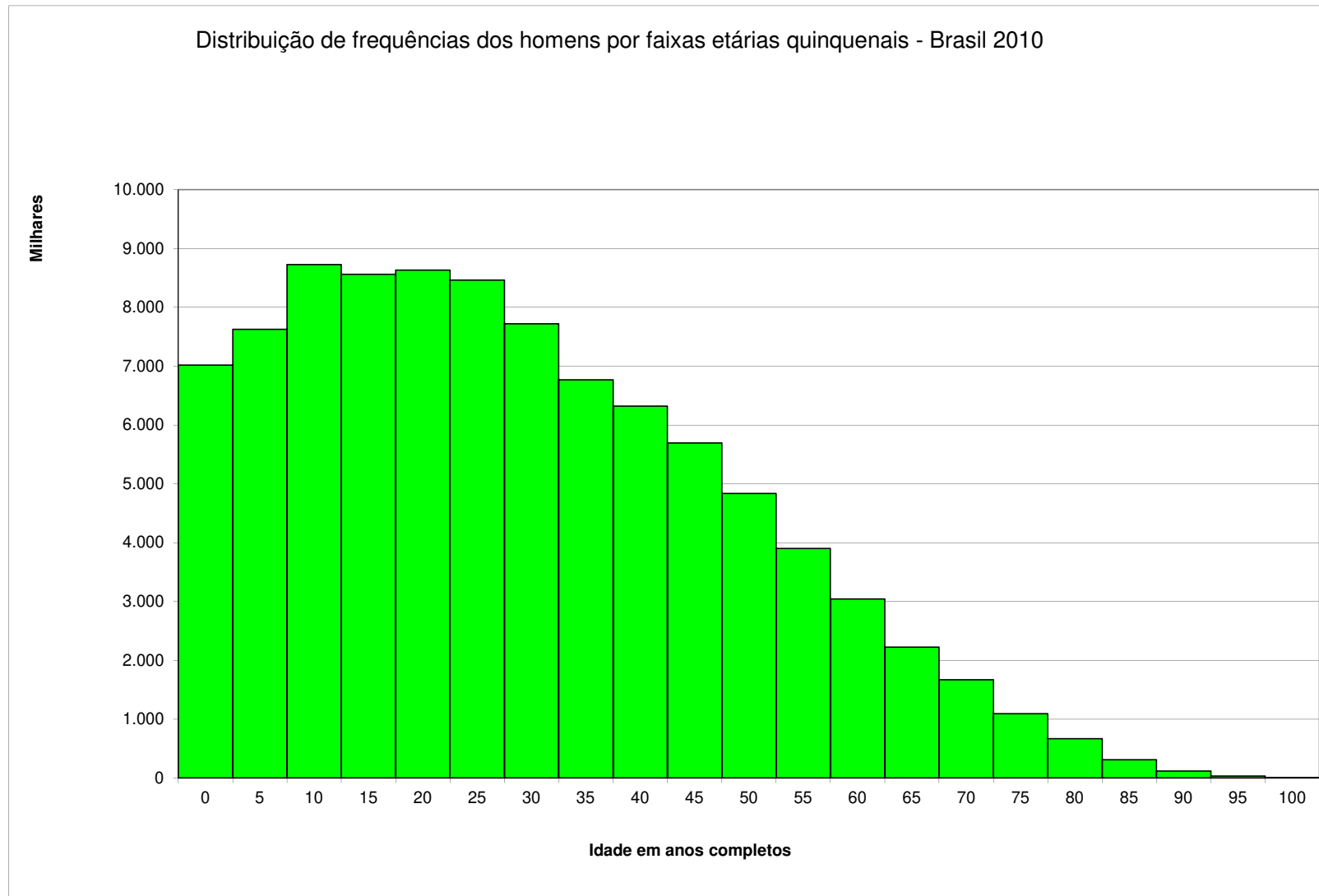
- Ao invés de registrar cada valor nas folhas, simplesmente **conte** quantos valores “caem” em cada classe.
- Em seguida, desenhe um **gráfico de colunas** (ou de barras), onde a altura de cada coluna é a frequência da classe que representa.
- Como já indicado, as **colunas não devem ser separadas por espaços**.
- Também pode ser usada a frequência relativa ou proporção de valores na classe.

Exemplo 3.6 – Distribuição do número de domicílios particulares permanentes por setores censitários - Vitória – 2010

Histograma do número de domicílios particulares permanentes

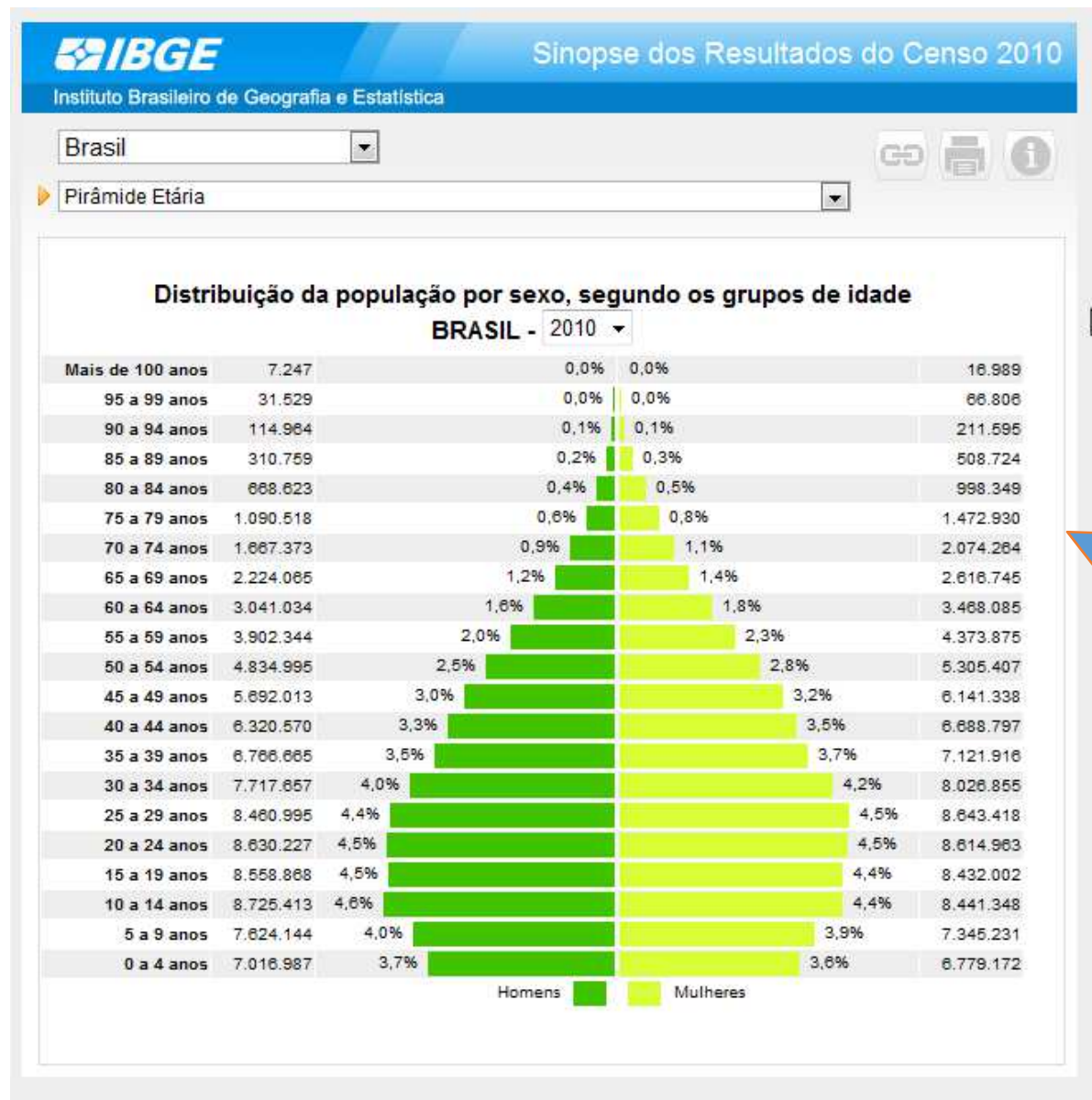


Exemplo 3.7a – Distribuição da população masculina por idades – 2010



Fonte: IBGE – Sinopse do Censo Demográfico 2010.

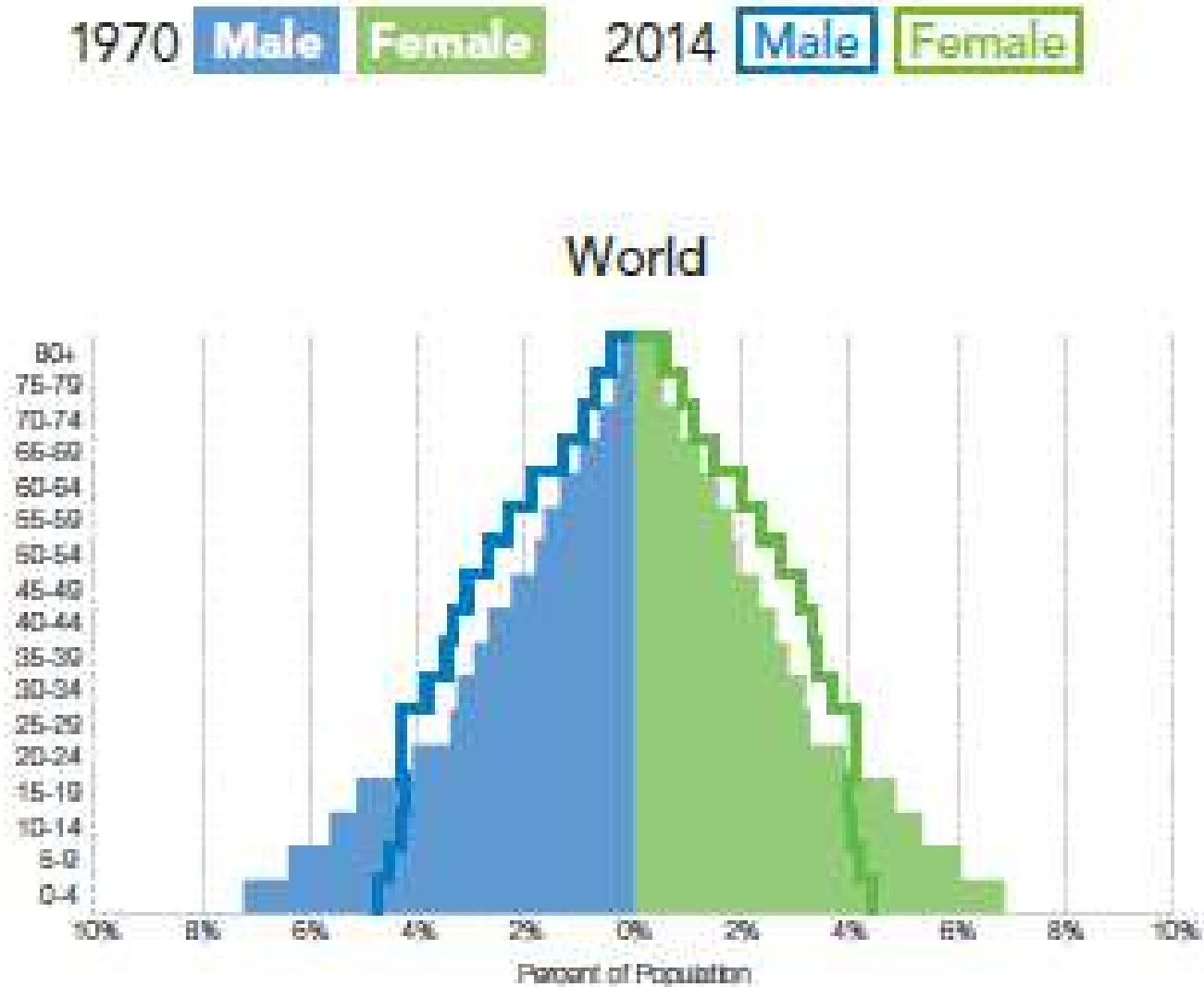
Exemplo 3.7b – Distribuição da população por idades – 2010



Capturado em 01/05/2011 de
<http://www.ibge.gov.br/home/default.php>

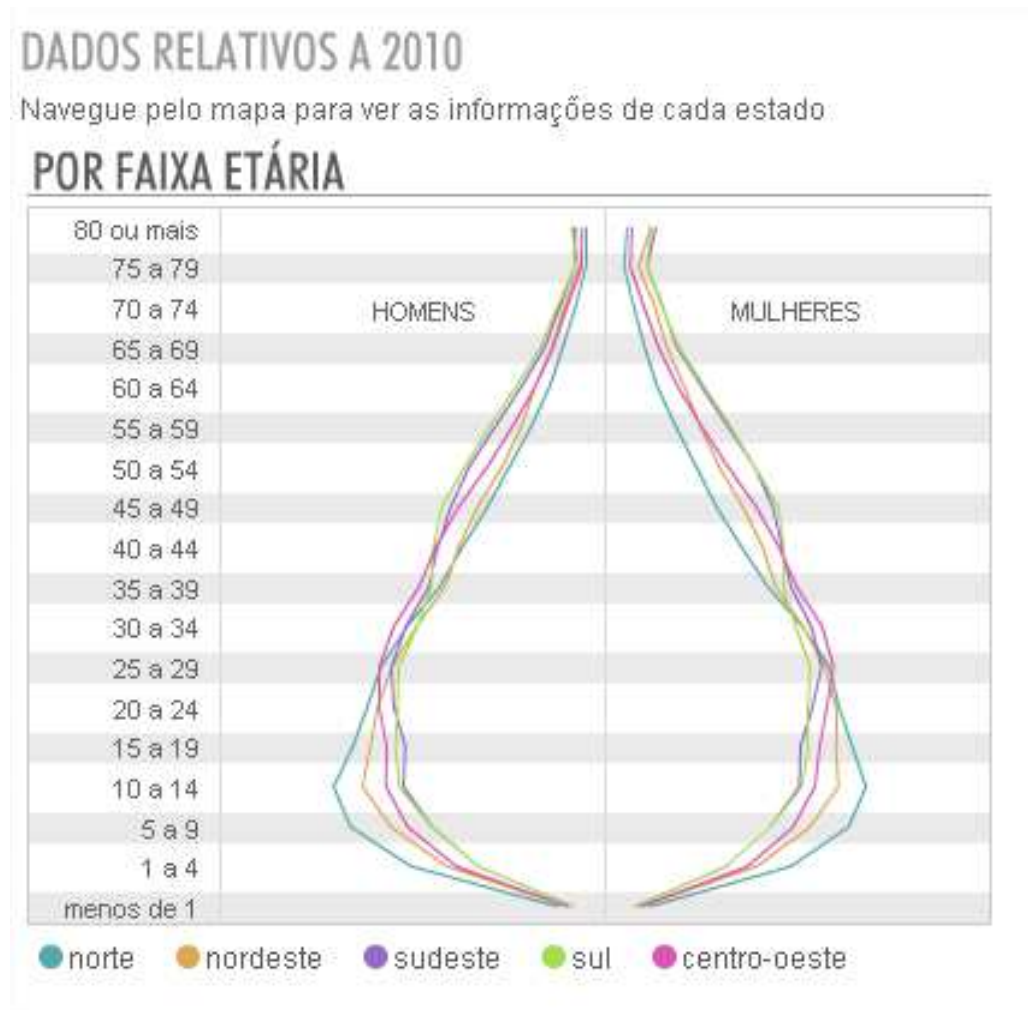
Tabela combinada
com gráfico...
Excesso de detalhes
tornando a 'leitura'
do gráfico difícil.
Falta escala.

Exemplo 3.8 – Pirâmides etárias sobrepostas para comparação



Fonte: Population Reference Bureau, World Data Sheet 2014.

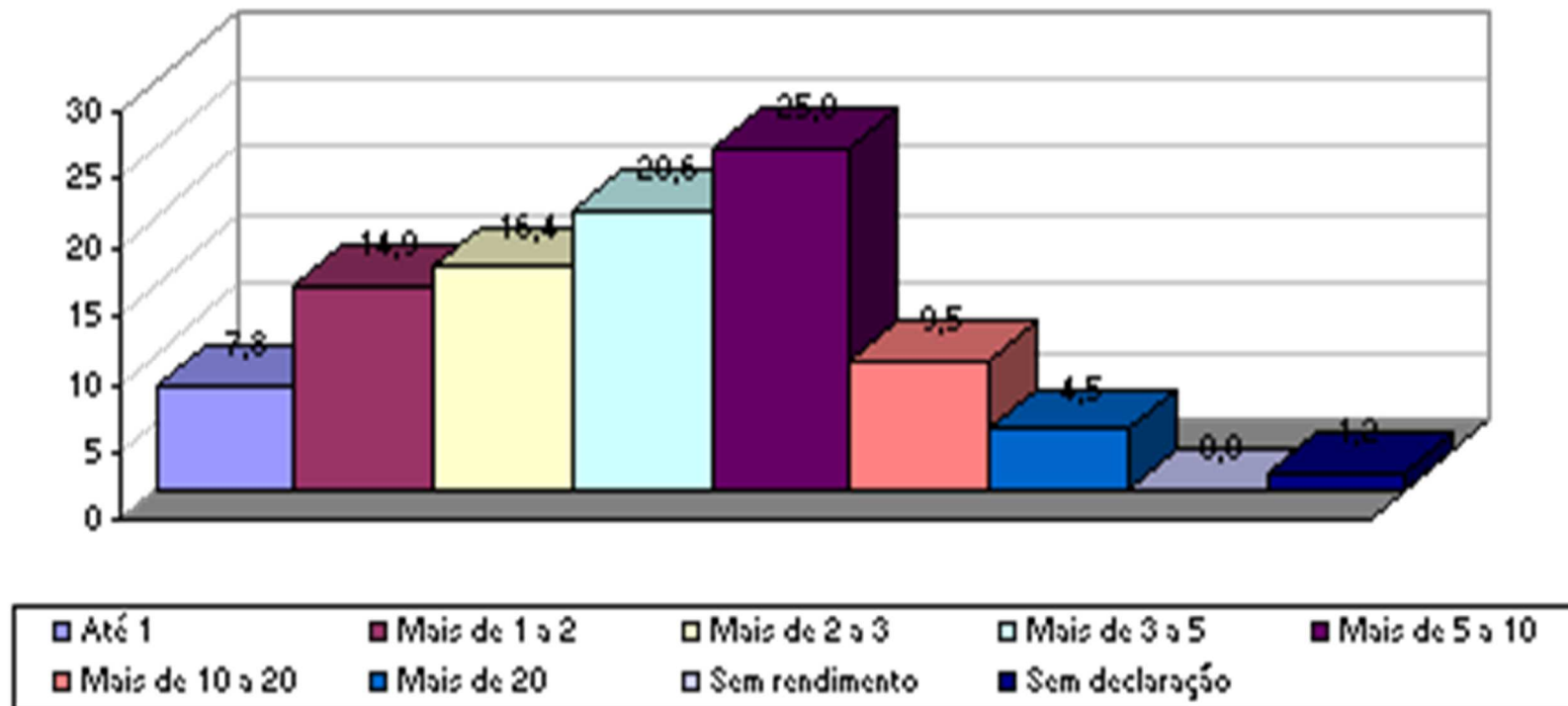
Exemplo 3.9– Distribuição da população por idades – 2010



Fonte: Capturado de <http://oglobo.globo.com/> em 29/04/2011 – 11:05.

Exemplo 3.10 Distribuição de Rendimentos – PNAD 1999

**Distribuição dos empregados, militares e estatutários, segundo as classes de rendimento do trabalho principal, em salários mínimos
Brasil - 1999**



Exemplo 3.10 Distribuição de Rendimentos – PNAD 1999

Defeitos do gráfico:

1. Uso de efeitos 3D causam distorção das frequências relativas;
2. Tipo de gráfico errado: gráfico de colunas em vez de histograma;
3. Altura das colunas usada para representar frequência relativa, mas classes de renda de amplitudes distintas → área das colunas deveria representar a frequência relativa;
4. Gráfico sugere que a classe de rendimento ‘modal’ é ‘Mais de 5 a 10’, mas este efeito decorre da maior amplitude deste intervalo.
5. Uso de cores distintas para cada coluna sugere que pessoas pertencem a diferentes categorias, mas a variável subjacente é numérica contínua.

Resumo sobre gráficos para distribuições de uma variável numérica

1. Determine **tipo da variável**, se discreta ou contínua.
2. Determine **número de observações** que precisam ser resumidas (n).
3. **Escolha o tipo** de gráfico que vai elaborar.
4. **Faça o gráfico** usando ferramenta adequada.
5. **Interprete o gráfico** para ‘contar a história’ revelada pelos dados.
6. **Nunca** use efeitos 3D.
7. Cuidado com as **escalas**.

Gráficos em Estatística

Parte 4

Gráficos Para Duas Variáveis Numéricas

Gráficos – Objetivos e Mensagens

Nas sessões anteriores vimos gráficos que podem ser úteis para:

- **Apresentar distribuições de frequência de variáveis;**
- **Mostrar como um todo se divide em partes;**
- **Comparar distribuições de uma variável para populações distintas.**

Nesta sessão, nosso foco será para gráficos que servem para:

- **Mostrar variações ao longo do tempo (ou do espaço);**
- **Mostrar correlações ou associações entre variáveis.**

Gráficos Para Dados Numéricos (2 ou + Variáveis)

Gráficos úteis quando há duas ou mais **variáveis numéricas** incluem:

- **Gráfico de linhas;**
- **Diagrama de dispersão.**

O **gráfico de linhas** é indicado quando uma das variáveis representa o **tempo** (ou uma direção), e se pretende revelar o **movimento dos dados ao longo do tempo** (ou desta direção).

O **diagrama de dispersão** é para explorar ou revelar **relações entre variáveis**.

Quando há mais que duas variáveis envolvidas, geralmente é preciso usar **combinações desses gráficos**.

Gráficos de Linhas

São adequados para **mostrar evolução de dados numéricos ao longo do tempo.**

Muito usados para apresentar **dados de séries temporais** ou **resultados de pesquisas repetidas.**

As séries de dados representadas neste tipo de gráfico são formadas por pares de valores: (t, x_t) com $t=1, 2, \dots, T$.

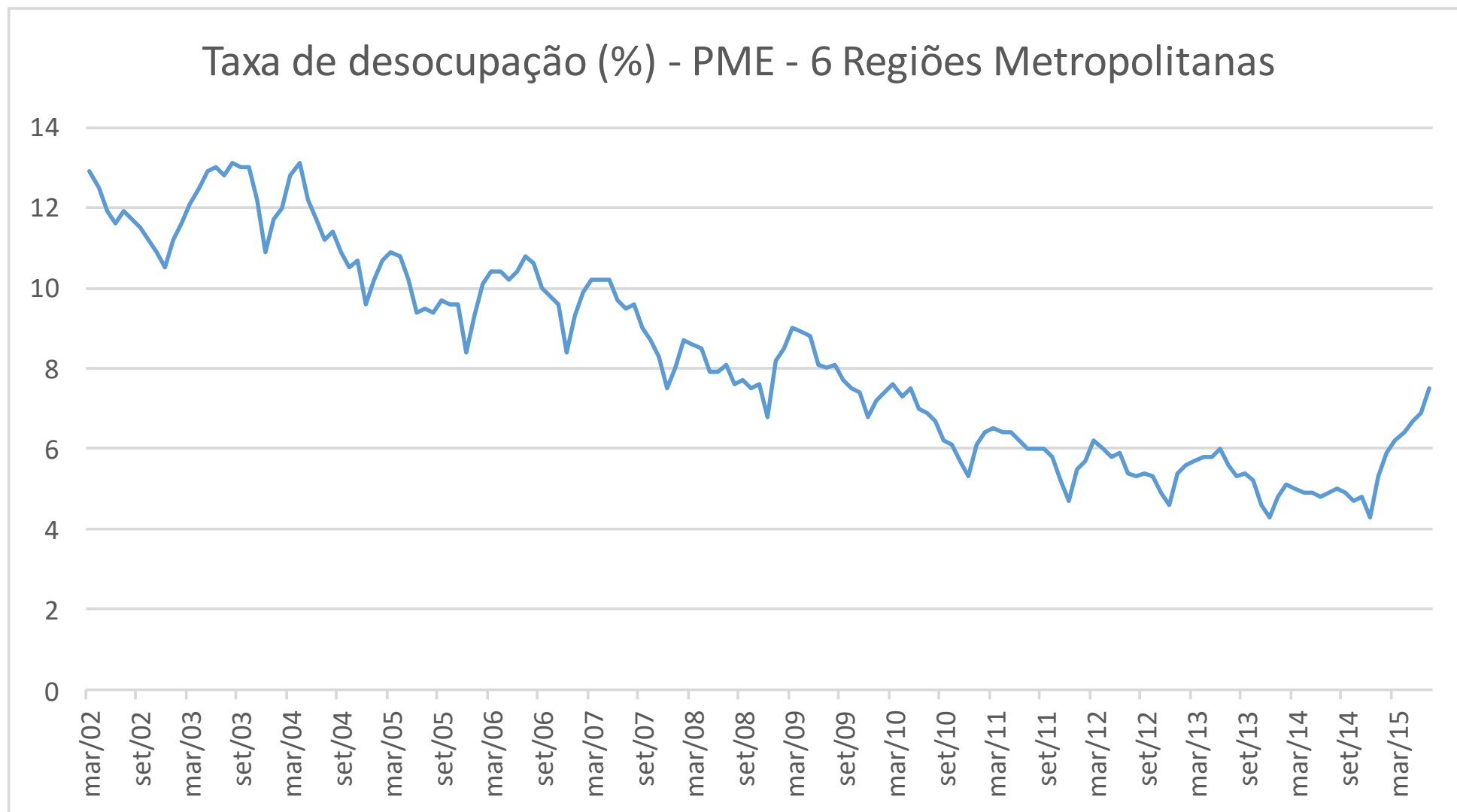
- t representa o **período** em que foi observada a variável de interesse;
- x_t representa o **valor da medida** obtida no tempo t .

Usa-se **um eixo para o tempo (t)**, e **outro para os dados (x_t)**.

Geralmente o eixo do tempo está na direção horizontal.

Os **pontos** representando as medidas referentes a cada período **são ligados por linhas.**

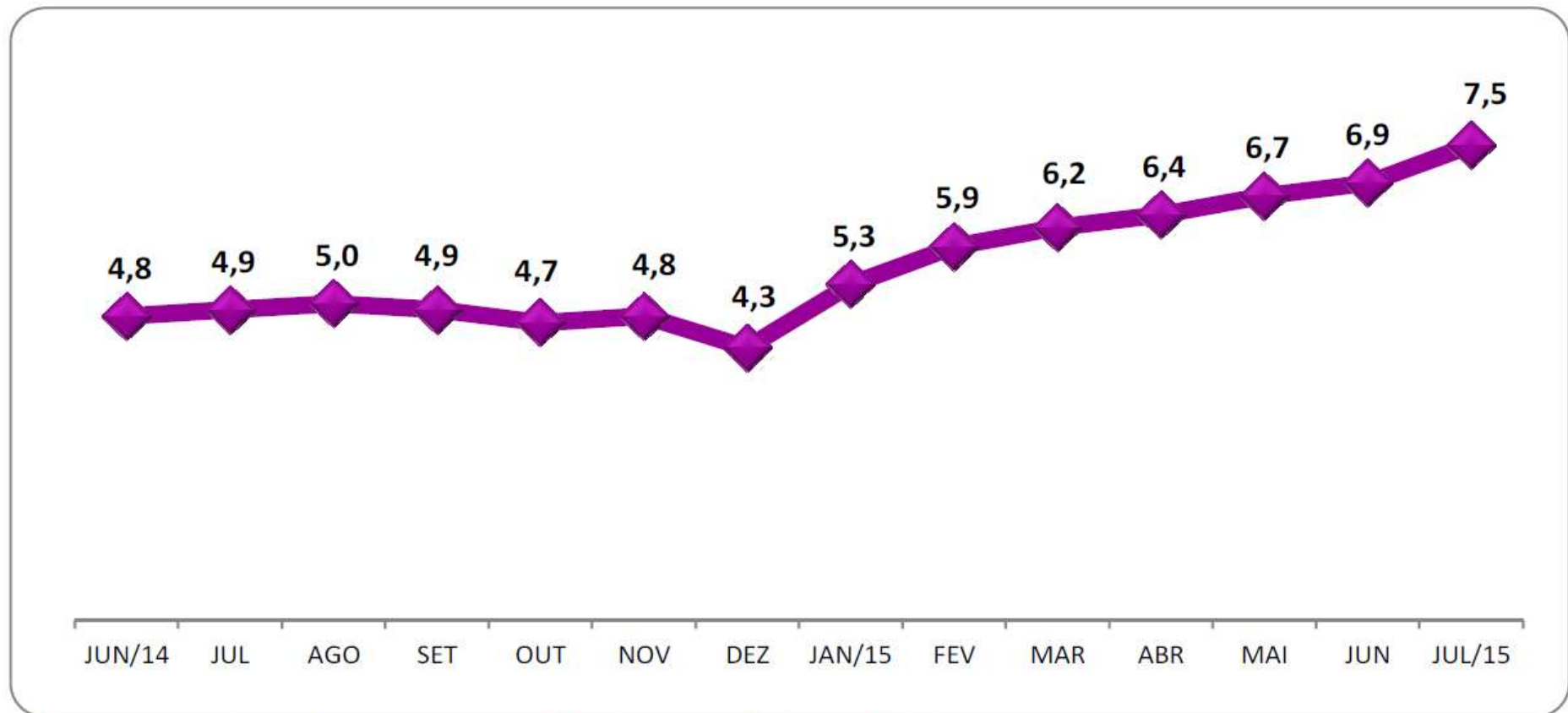
Exemplo 4.1a – Taxa de desocupação



Fonte: Elaboração própria, usando dados de IBGE, Pesquisa Mensal de Emprego.

Exemplo 4.1b – Taxa de desocupação

O gráfico a seguir mostra a evolução da taxa de desocupação, de JUNHO de 2014 a JULHO de 2015, no total das seis regiões metropolitanas abrangidas pela pesquisa



FONTE: IBGE, Diretoria de Pesquisas, Coordenação de Trabalho e Rendimento, Pesquisa Mensal de Emprego.

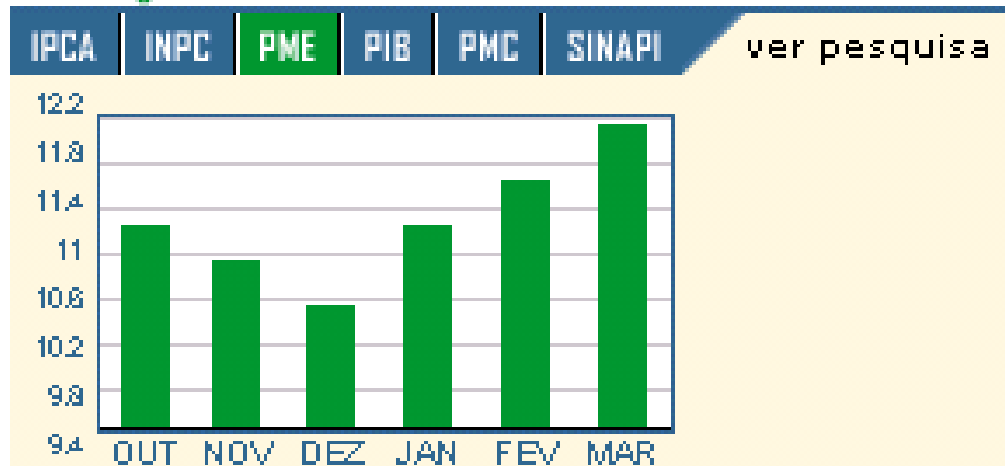
Que problemas há neste gráfico?

Gráficos para representar evolução no tempo

1. Os gráficos adequados para este fim são os **gráficos de linhas**.
2. Mas algumas vezes a '**criatividade**' entra em cena e outros gráficos são usados para este fim.
3. Veja os **maus exemplos** a seguir.

Exemplo 4.2a – Taxa de Desocupação

Variação dos indicadores



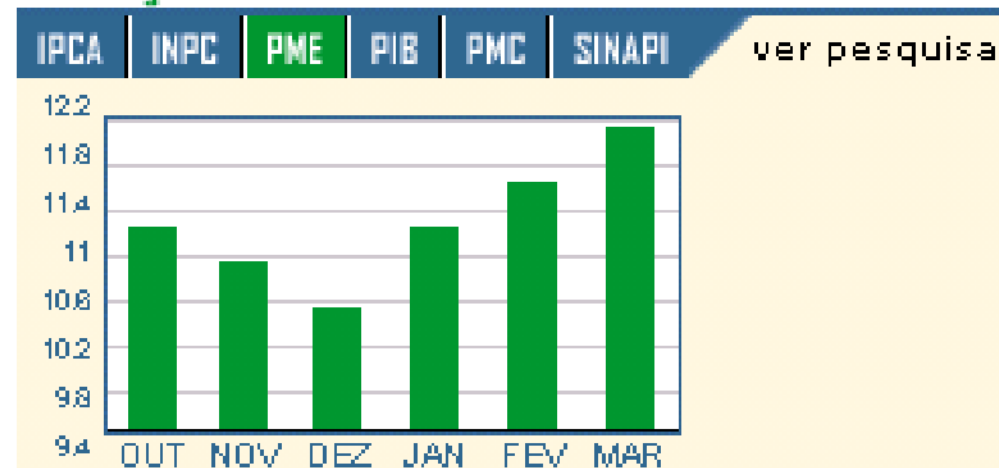
A PME - Pesquisa Mensal de Emprego fornece indicadores de mercado de trabalho, acompanhando a dinâmica conjuntural de emprego e desemprego. O gráfico mostra a taxa de desemprego.

Gráfico já usado na página do IBGE na internet.

<http://www.ibge.gov.br/home/default.php>

Exemplo 4.2a – Taxa de Desocupação

Variação dos indicadores



A PME - Pesquisa Mensal de Emprego fornece indicadores de mercado de trabalho, acompanhando a dinâmica conjuntural de emprego e desemprego. O gráfico mostra a taxa de desemprego.

Principais defeitos do gráfico

1. Tipo inadequado para mostrar evolução no tempo.
2. Escala não iniciada em zero, num gráfico de colunas.
3. Eixo do tempo tem informação incompleta (não dá para saber o ano).
4. Texto longo explica objetivo da pesquisa e não do gráfico.
5. Unidade de medida (%) não explicitada nem no eixo vertical nem na descrição do gráfico.
6. Formatação do eixo vertical tem variação no número de casas decimais dos valores na escala.
7. Título genérico (Variação dos indicadores) é inadequado. A série mostrada não é de variações.

Gráfico já usado na página do IBGE na internet.

<http://www.ibge.gov.br/home/default.php>

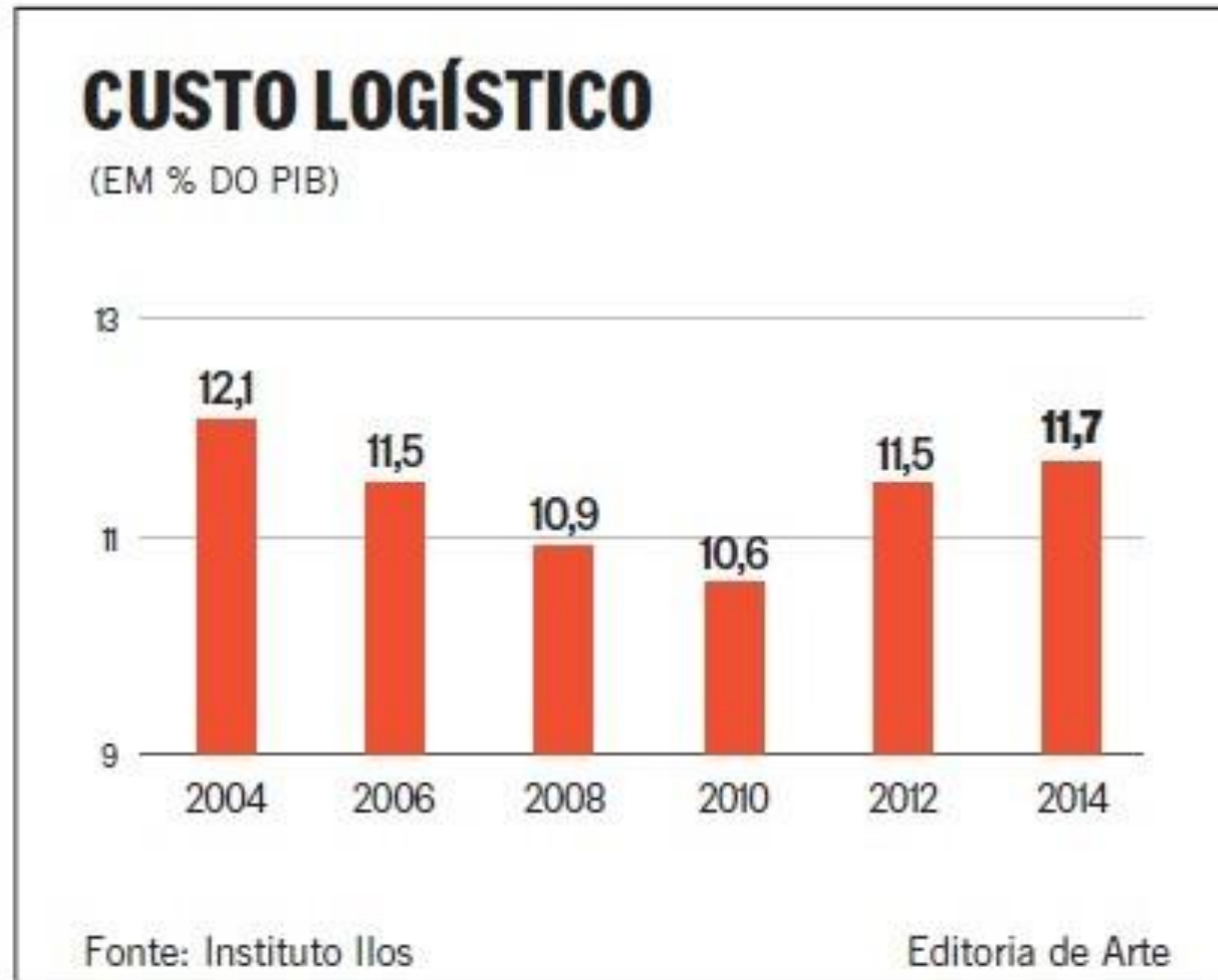
Exemplo 4.2b – Taxa de Desocupação



Gráfico atualmente usado na página do IBGE na internet.

<http://www.ibge.gov.br/home/default.php>

Exemplo 4.3 – Evolução do custo logístico no Brasil



<http://blogs.oglobo.globo.com/miriam-leitao/post/carga-pesada.html>

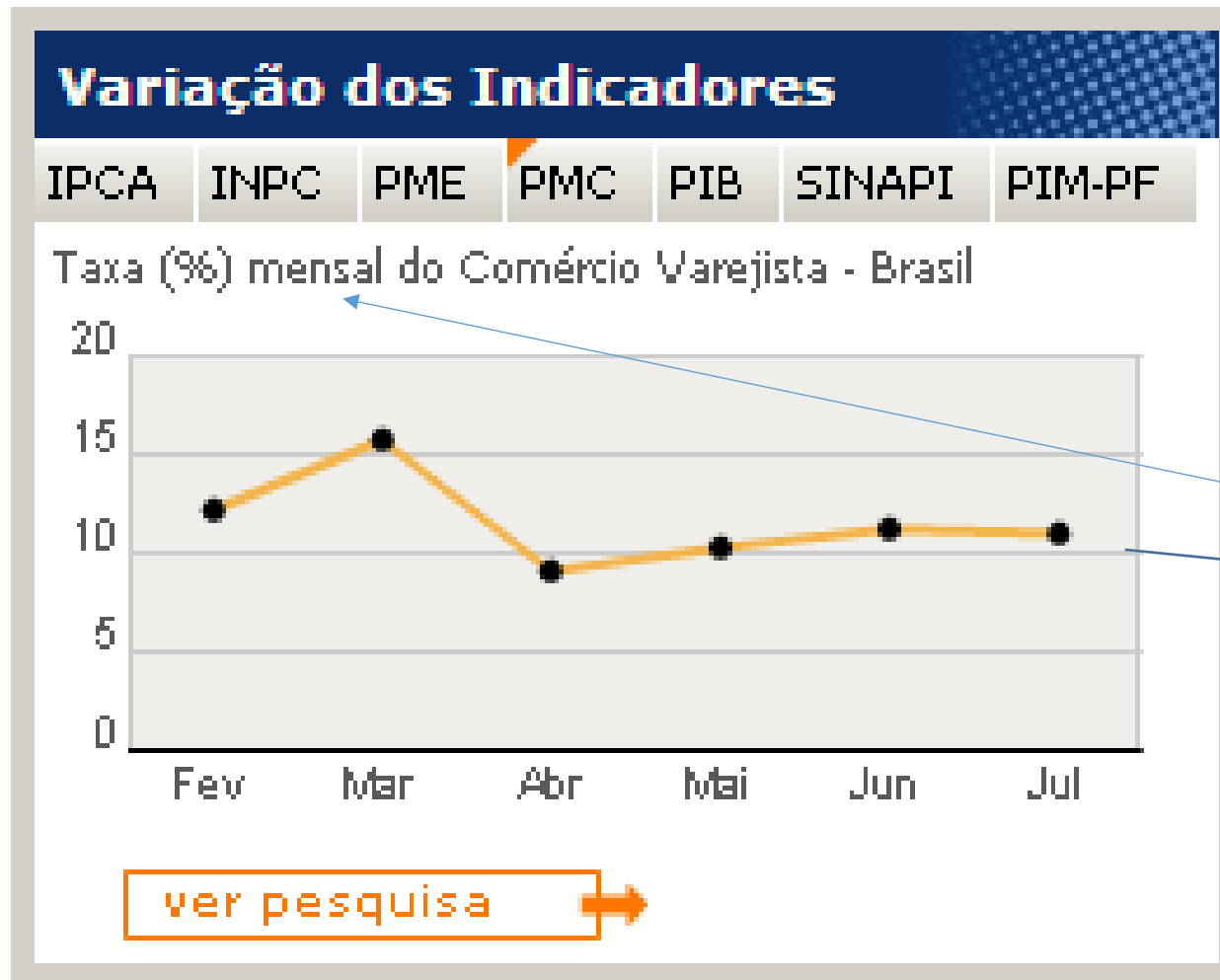
Capturado em 20/09/15.

Exemplo 4.4 – Inflação anual no Brasil



https://www.google.com.br/search?q=globonews+grafico+infla%C3%A7%C3%A3o&client=safari&hl=pt&prmd=niv&source=lnms&tbn=isch&sa=X&ved=0CAkQ_AUoA moVChMI6d7074-DyAIVzJKQCh3FIA4A#imgsrc=_UscwrkV84-qEM%3A

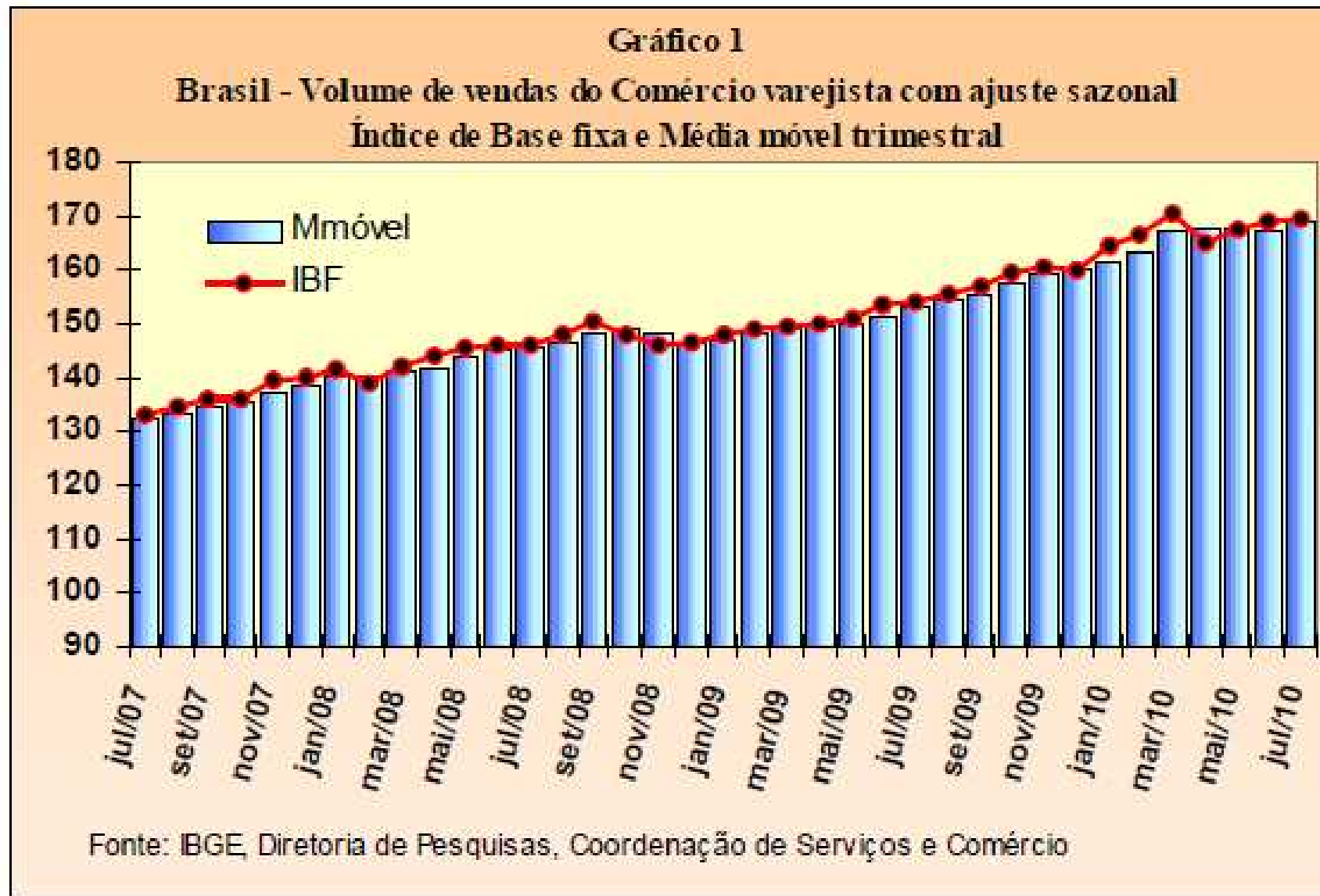
Exemplo 4.5 – Variação das vendas no comércio



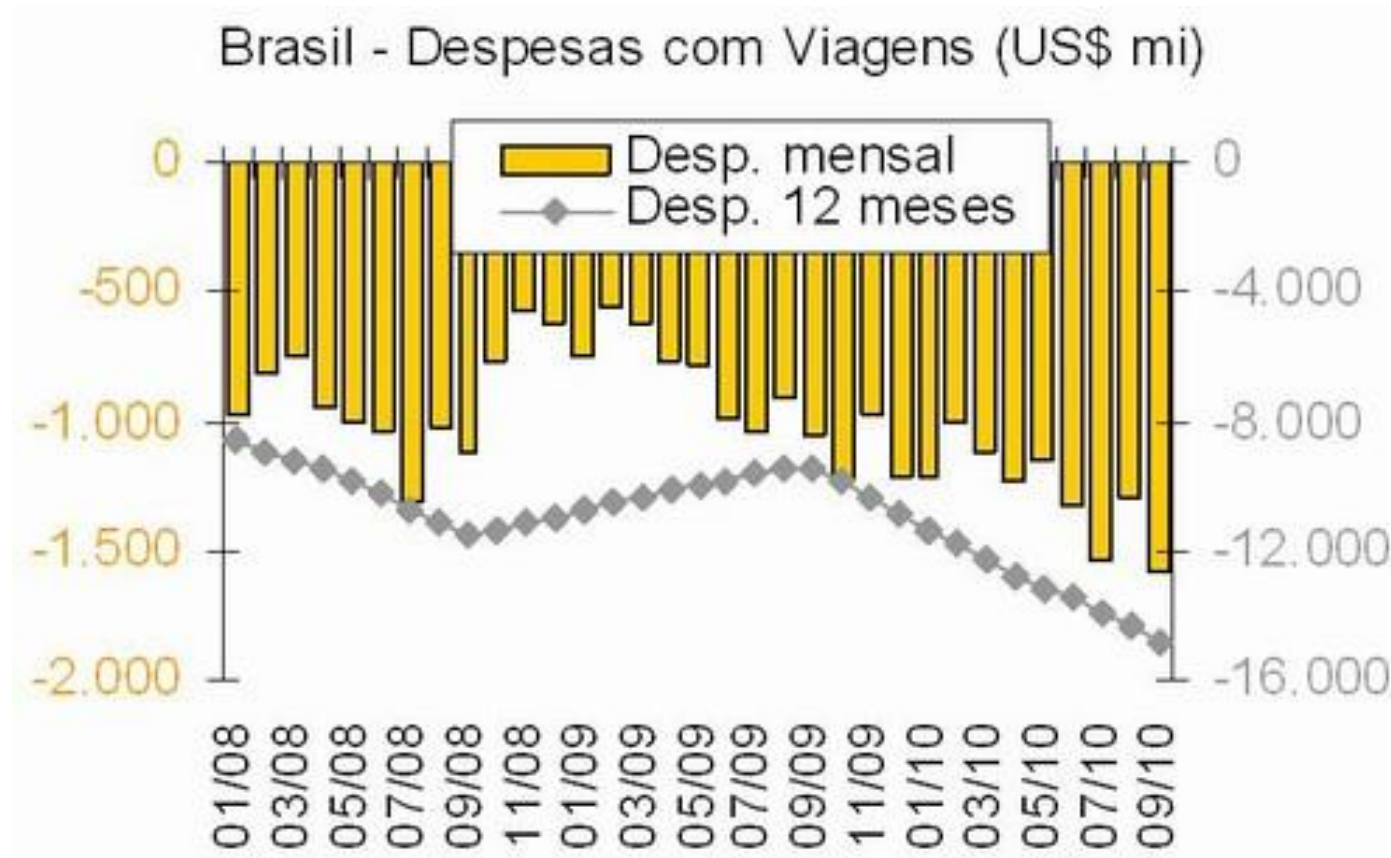
Série mostra variações anuais, não mensais.

Fonte: <http://www.ibge.gov.br/home/default.php>

Exemplo 4.6 – Variação no Volume de Vendas no Varejo



Exemplo 4.7 - Emprego e Câmbio Explicam Aumento de Gastos no Exterior



Fonte:

<http://oglobo.globo.com/economia/miriam/default.asp?ch=S&a=&pagAtual=2> Enviado
por Valéria Maniero - 25.10.2010 - 16h12m

Exemplo 4.8 – Número de Chamadas Recebidas



Fonte: <http://www.mds.gov.br/falemds>

Gráficos de Linhas – Problemas Mais Comuns

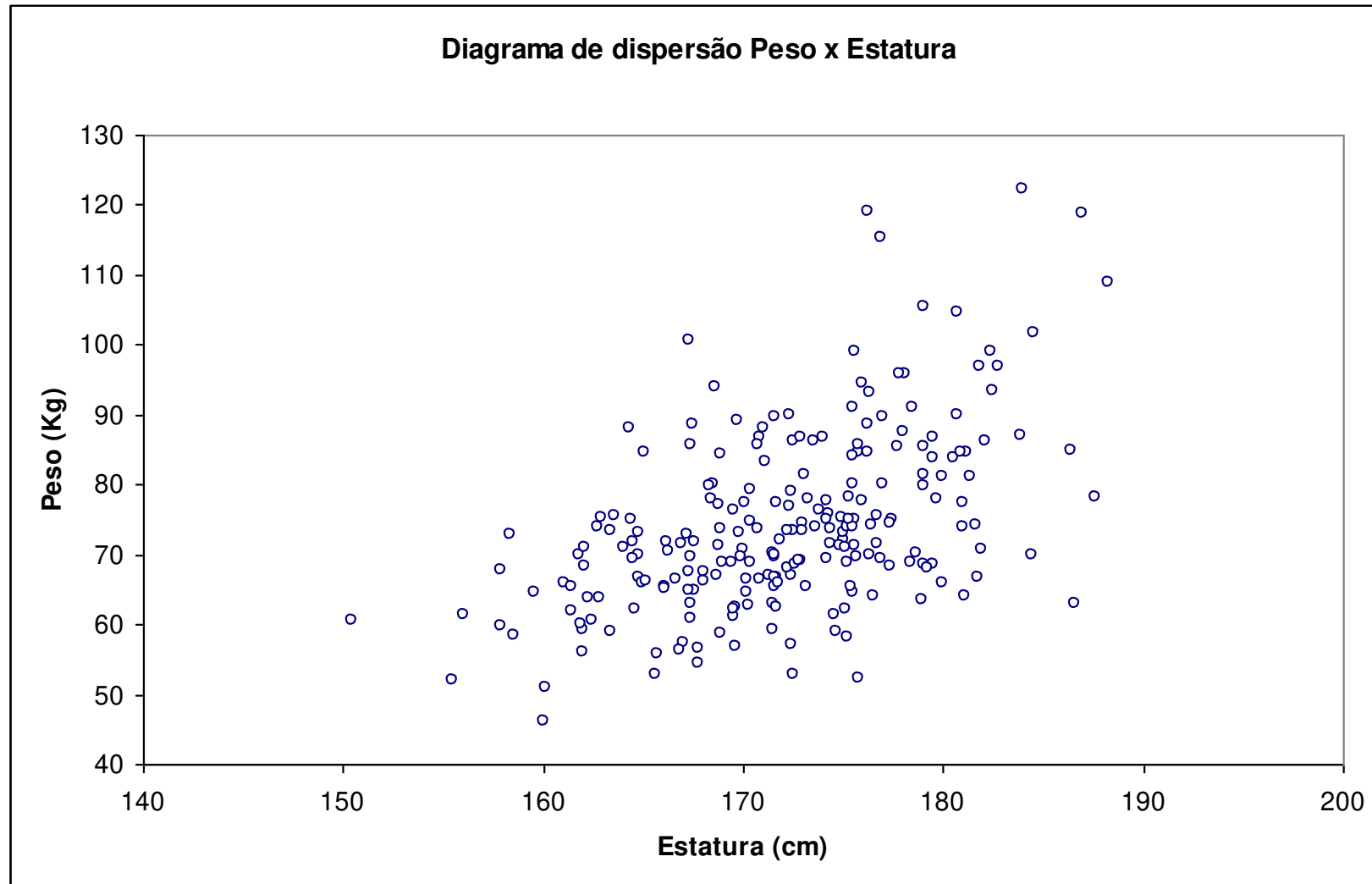
1. Escolha equivocada do tipo de gráfico.
2. Escolha de símbolos que não pontos e linhas.
3. Excesso de informação num só gráfico.

Mantenha a simplicidade e use corretamente o gráfico de linhas para mostrar dados que evoluem no tempo.

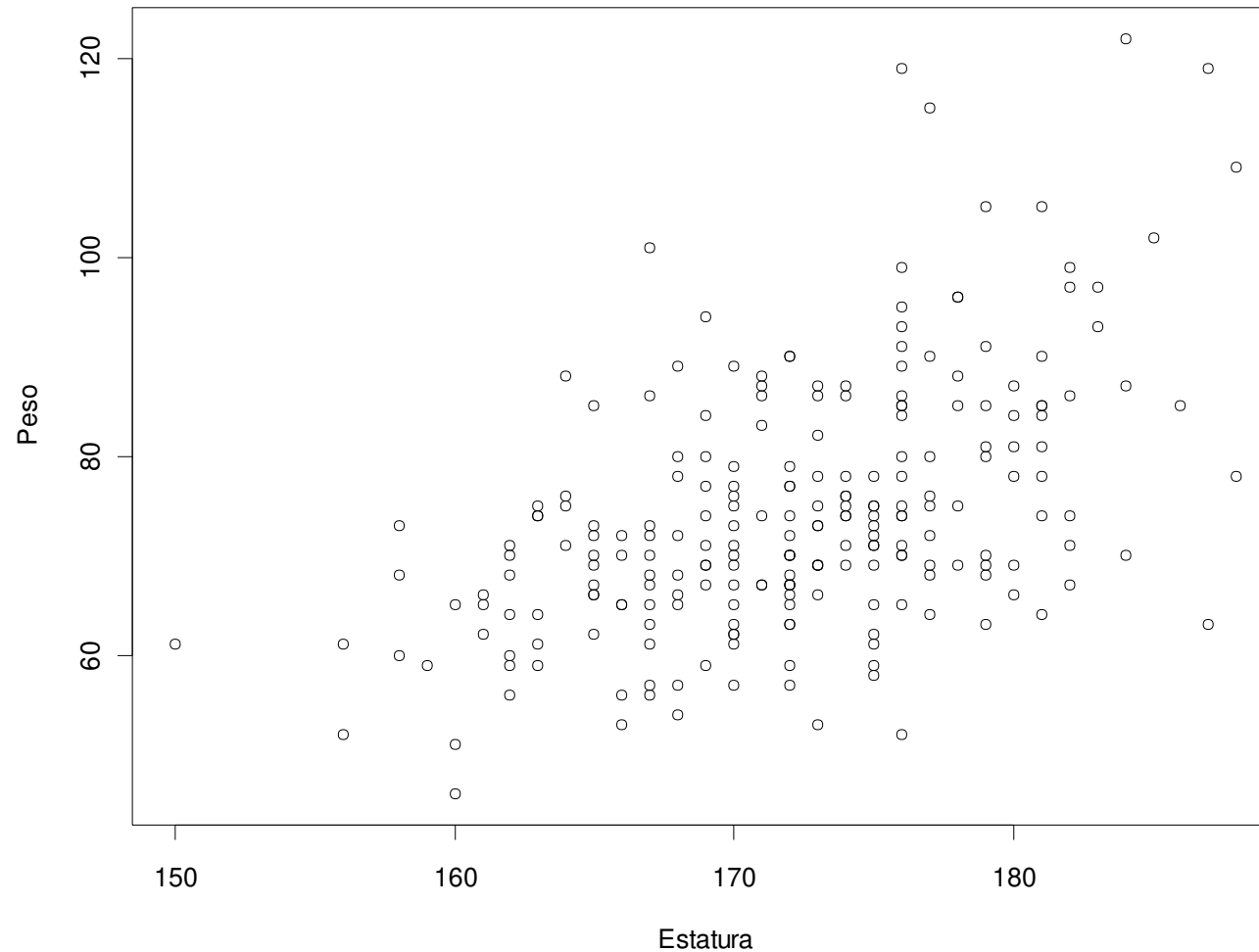
Gráficos para Explorar Relações entre Variáveis

1. Suponha que você tem dados representando **duas medidas numéricas**, **Estatura** (medida em cm) e **Peso** (medido em Kg), para amostra de 231 homens adultos.
2. O gráfico padrão para **revelar associação** entre estas medidas é o **diagrama de dispersão** (*scatter plot*).
3. Represente uma variável ao longo do eixo horizontal (x) e outra ao longo do eixo vertical (y). **Cada ponto representa um indivíduo** (observação).
4. As coordenadas de cada ponto devem ser marcadas conforme as escalas nos eixos x e y, respectivamente.
5. Os pontos **não devem ser ligados** ou conectados por linhas.

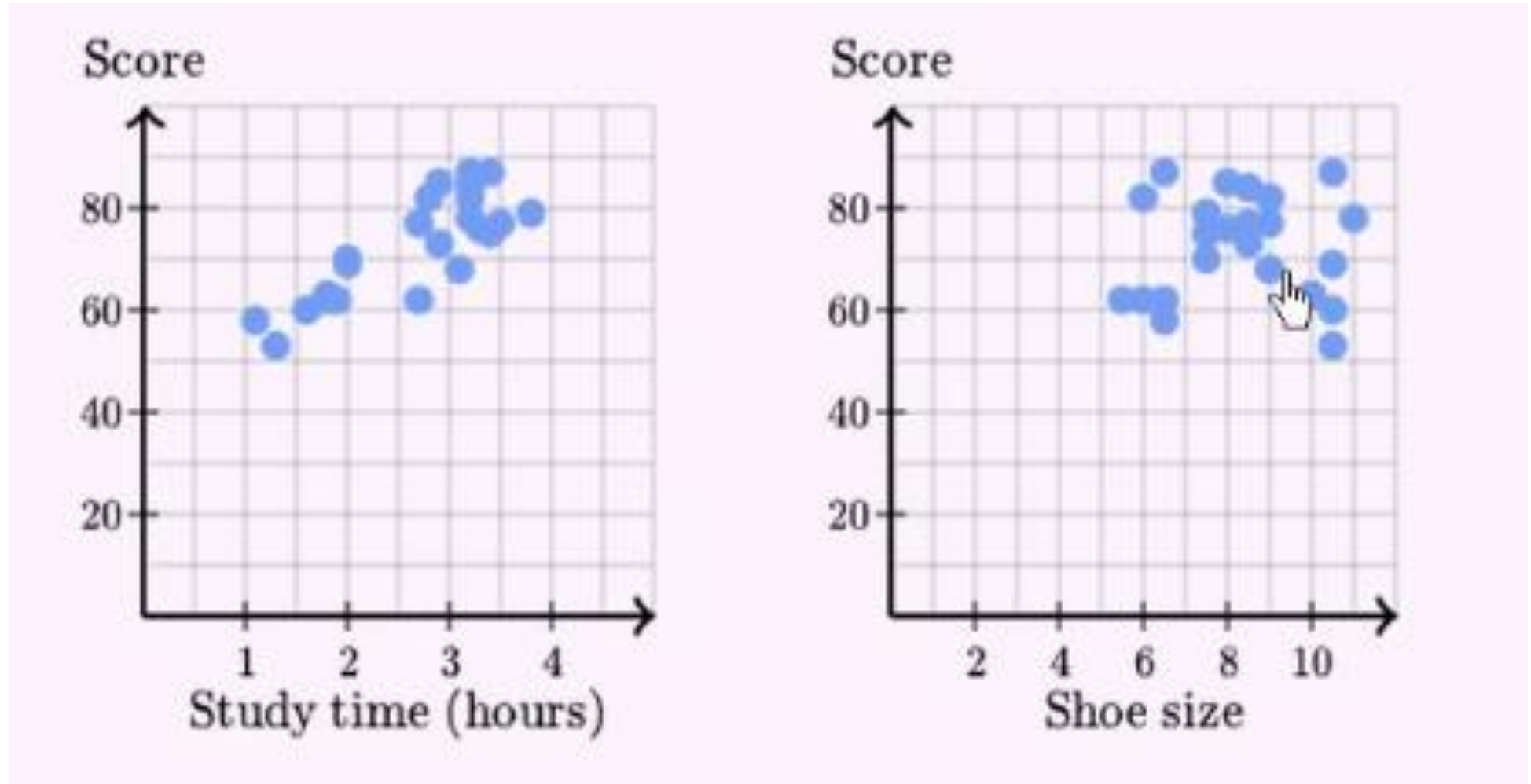
Exemplo 4.9a – Diagrama de Dispersão Peso × Estatura Obtido no Excel Após Alguma Edição



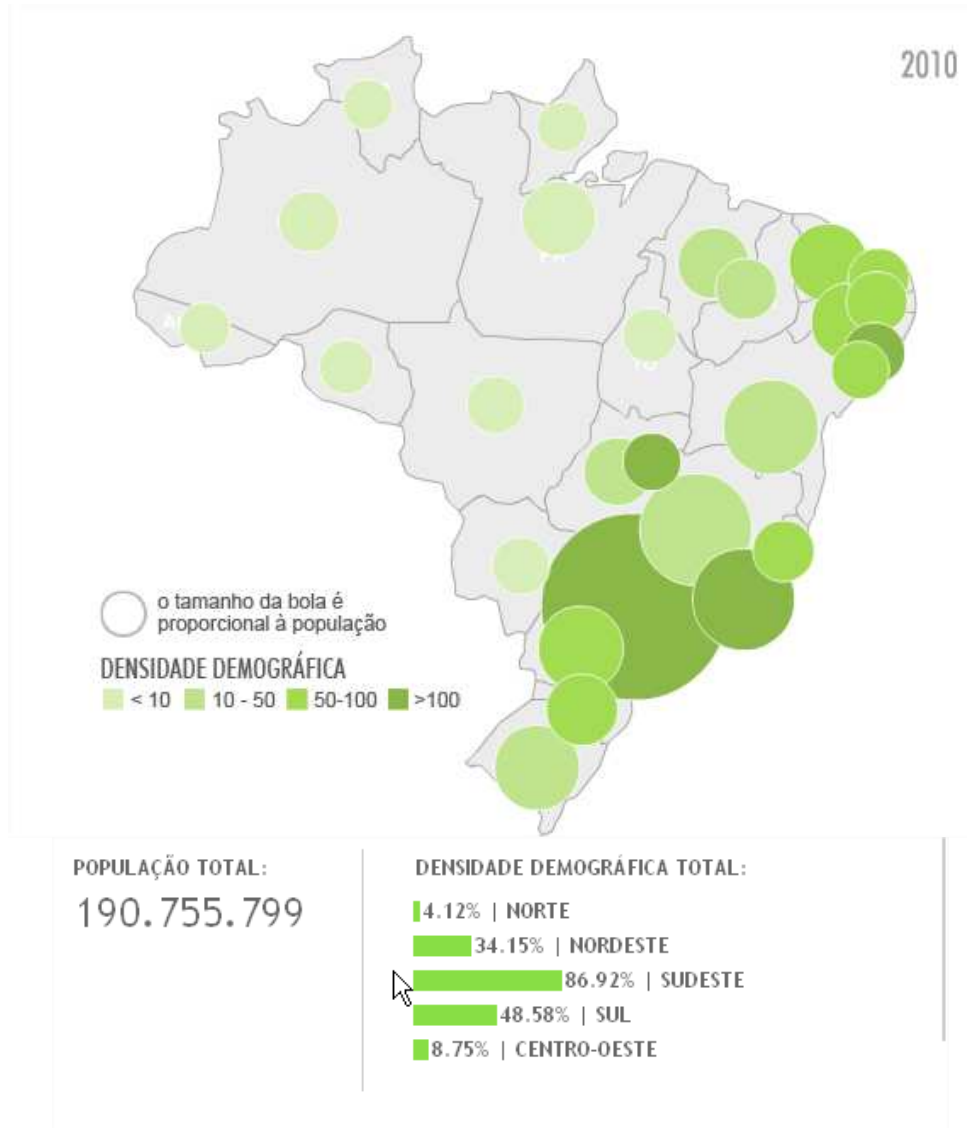
Exemplo 4.9b – Diagrama de Dispersão Peso × Estatura Obtido no R - `plot(Estatura, Peso)`



Exemplo 4.10 – Dois diagramas de dispersão envolvendo a mesma ‘resposta’



Exemplo 4.11 Distribuição da Densidade Demográfica por UF e Regiões – Censo 2010

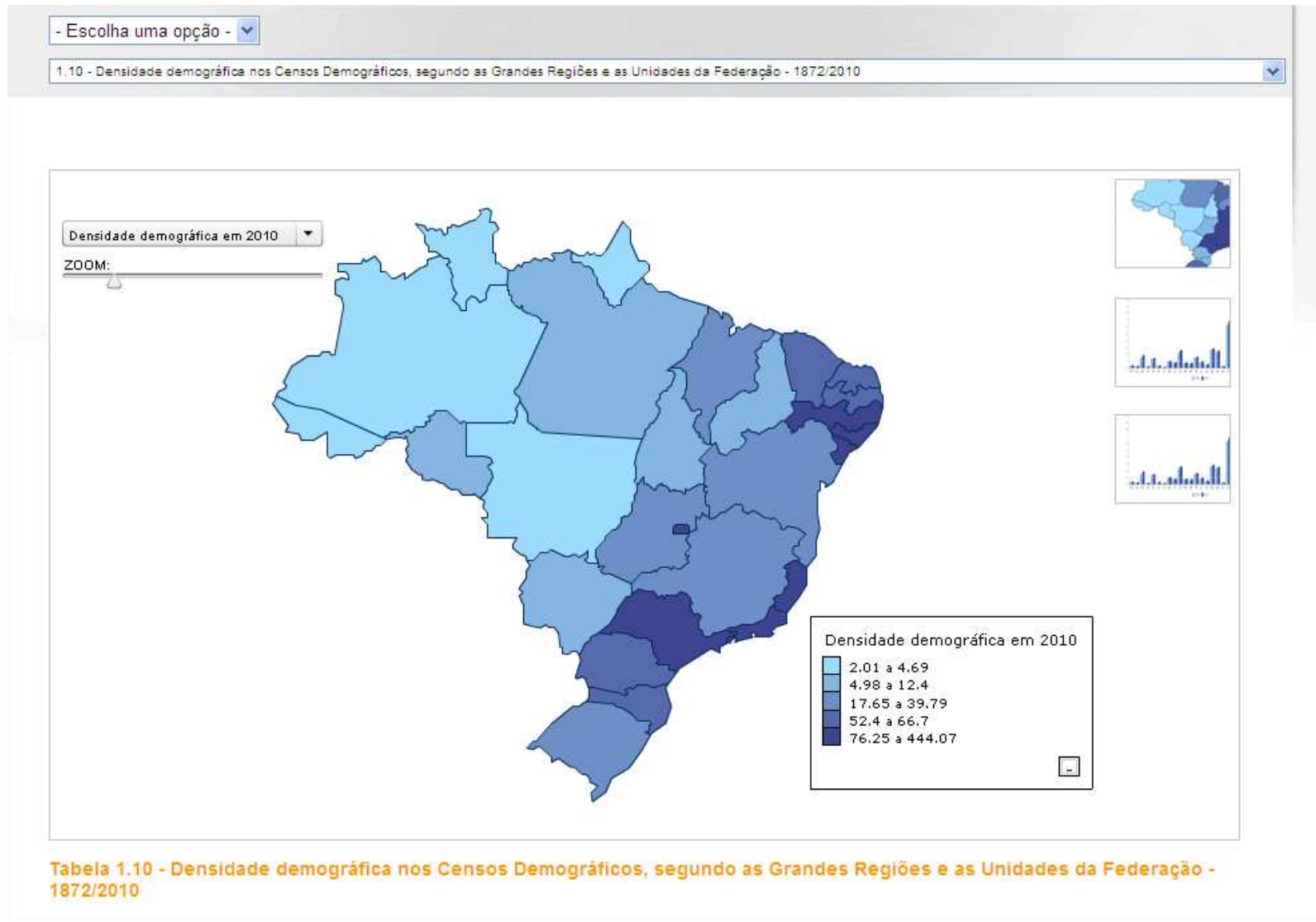


Fonte: Capturado de
<http://oglobo.globo.com/>
em 29/04/2011 – 11:05.

Exemplo 4.11 Defeitos dos gráficos

1. Unidade de medida da densidade demográfica não informada.
2. Valores da densidade demográfica (hab/km²) são apresentados em % no gráfico de barras!
3. Gráfico de barras não tem escala.
4. Nomes das categorias no gráfico de barras apresentados no 'lugar' errado (à direita das barras, em vez de à sua esquerda).
5. No mapa, a principal variável representada deveria ser a densidade demográfica, mas a população dos estados toma precedência.
6. Mapas para a densidade demográfica deveriam colorir cada estado com as faixas de densidade indicadas na legenda.
7. Excesso de informação no gráfico prejudica sua leitura e apreensão.

Exemplo 4.12 – Densidade demográfica por UF – Censo 2010



Fonte: Capturado de <http://www.censo2010.ibge.gov.br/sinopse/index.php?dados=10&uf=00> em 09/05/2011 às 18:06.

Problemas Mais Comuns em Gráficos

1. **Tipo de gráfico inadequado.**
2. **Eixos sem rótulos ou descrição.**
3. **Eixos / escalas não iniciadas em zero.**
4. **Variações não explicadas nos rótulos ou na escala dos eixos.**
5. **Unidades de medida incorretas.**
6. **Áreas desproporcionais à frequência dos dados.**
7. **Excesso de informação ou lixo gráfico.**

Esta lista pode servir para você verificar se não cometeu nenhum destes erros no(s) seu(s) gráfico(s).

Alguns Bons Exemplos para Motivar

Envelhecimento populacional no Reino Unido:

<http://www.neighbourhood.statistics.gov.uk/HTMLDocs/dvc5/agemap.html>

Gráfico premiado pela Royal Statistical Society.

Envelhecimento populacional em alguns países

<http://visualization.geblogs.com/visualization/aging/>

Envelhecimento populacional no Brasil

<http://economia.ig.com.br/brasil+tem+15+anos+para+colher+bonus+da+forca+de+trabalho/n1237814998689.html> Hans

Rosling e a Fundação Gapminder

<http://www.gapminder.org/>

Um Roteiro Para Analisar Gráficos Estatísticos

1. A **mensagem** ou estória do gráfico está evidente, é **clara**, faz sentido?
2. O **título** ou objetivo do gráfico é **claro**?
3. É indicada a **fonte dos dados**, na figura ou no texto que a acompanha?
4. As informações foram obtidas de uma **fonte confiável**?
5. Há **rótulos adequados** para identificar os eixos, etc.?
6. Os **começam em zero** ou não?
7. As **escalas** dos eixos **são constantes**?
8. Existem “**quebras**” nos eixos que sejam **difíceis de notar**?
9. Se forem **dados financeiros**, foram ajustados para inflação ou câmbio?
10. Há **lixo no gráfico**, atrapalhando a sua “leitura”?

Resumindo

Cada gráfico deve contar uma história ou passar uma mensagem ao leitor.

Um gráfico é tão bom quanto o correto entendimento da história correta pelo leitor.

Se você tem várias histórias para contar, use vários gráficos.

“Keep it as simple as possible, but not simpler” - Albert Einstein.

Referências

Cairo, A. (2016). *The Truthful art: data, charts, and maps for communication*. New York: NEW RIDES.

Cleveland, W. S. (1993). *Visualizing Data*. New Jersey: Hobart Press.

Few, S. (2009). *Now you see it: Simple Visualization Techniques for Quantitative Analysis*. California: Analytics Press.

Knafllic, C. N. (2015). *Storytelling with data: a data visualization guide for business professionals*. Canada: Wiley.

UNECE (2009a). *Making data meaningful – Part 1 – a guide to writing stories about numbers*.

UNECE (2009b). *Making data meaningful – Part 2 – a guide to presenting statistics*.

Wild, C.J. e Seber, G.A.F. (2004). *Encontros com o acaso: um primeiro curso de análise de dados e inferência*. Rio de Janeiro: LTC Editora.