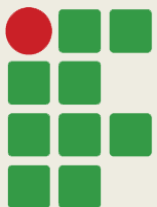
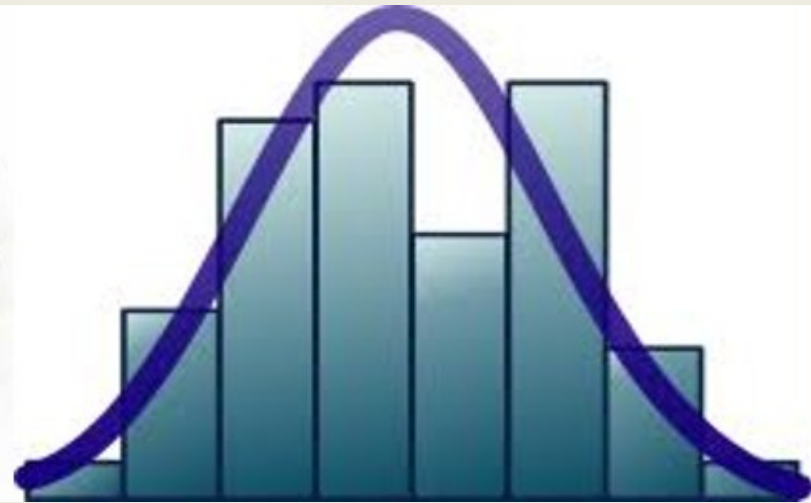


Probabilidade e Estatística



INSTITUTO FEDERAL
Catarinense
Campus Blumenau

Professor Jeovani Schmitt

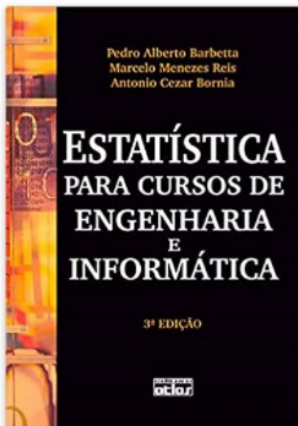


Probabilidade e Estatística

Aula 3

Análise Exploratória de Dados (AED)

- ✓ Dados e variáveis
- ✓ Distribuição de frequências: Tabelas e Gráficos
- ✓ Análise de variáveis **qualitativas**
- ✓ Atividade prática – Uso do R e Rstudio para fazer uma análise de dados
- ✓ Aspectos gerais do R



Capítulo 3

Análise Exploratória de Dados (AED)

- ✓ Descrever os dados;
- ✓ Conhecer algumas características do processo;
- ✓ Uso adequado de tabelas, gráficos e medidas.



Classificação das variáveis

QUALITATIVA

NOMINAL

cor dos olhos, marca

ORDINAL

classe social, escolaridade

QUANTITATIVA

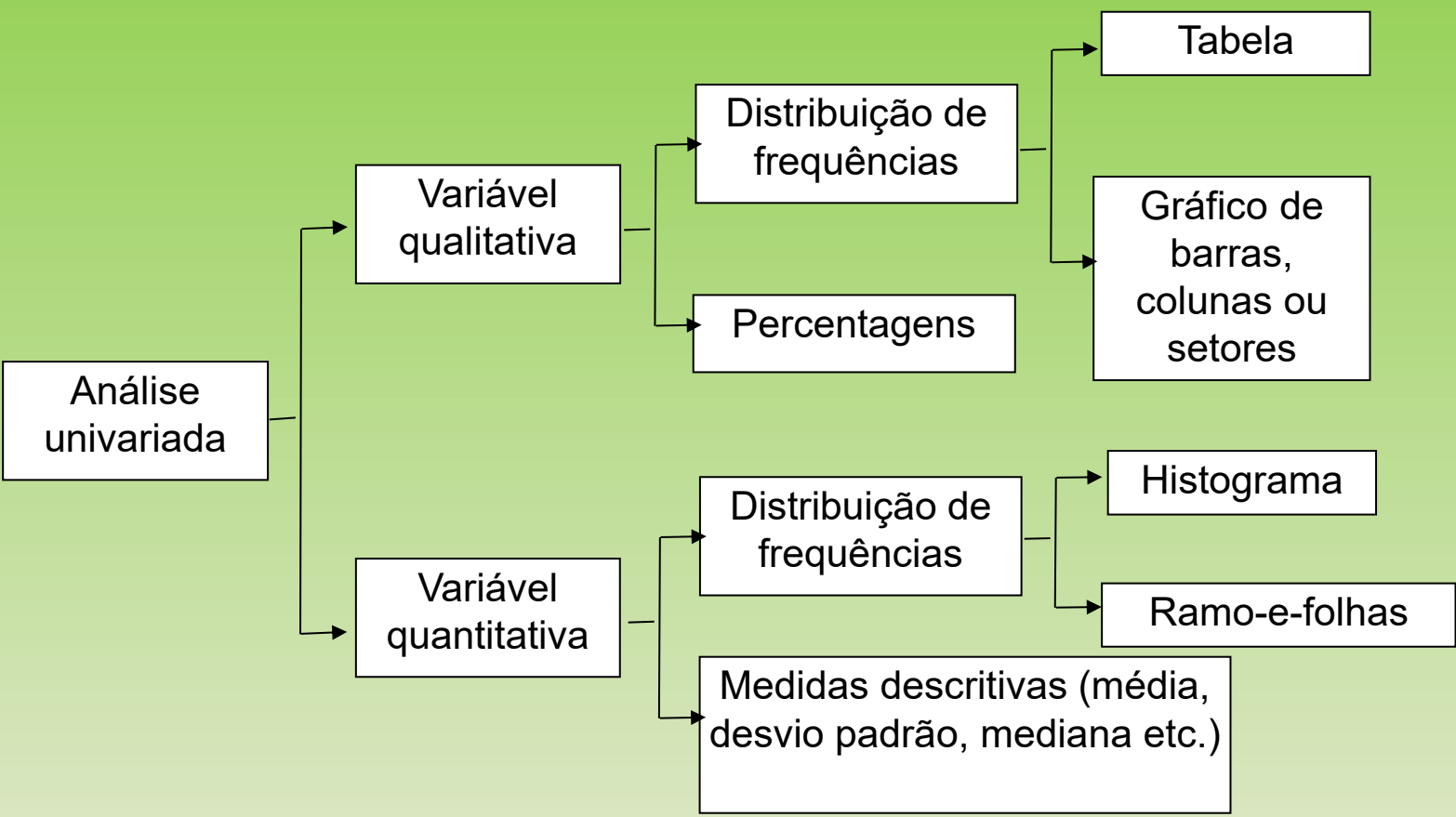
CONTÍNUA

peso, altura, salário, tempo

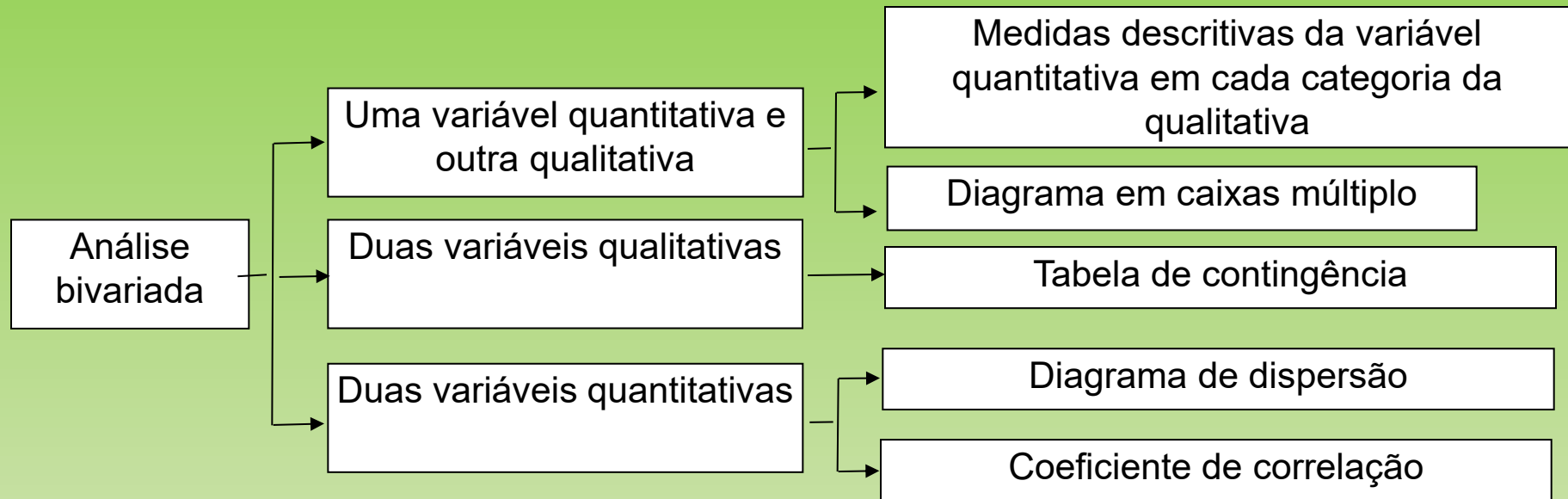
DISCRETA

nº de filhos, nº de carros

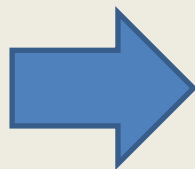
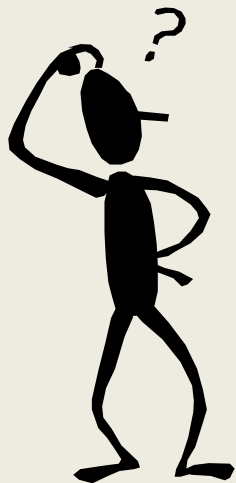
Esquema geral de uma Análise Exploratória de Dados (AED)



Esquema geral de uma Análise Exploratória de Dados (AED)



Como resumir, organizar e apresentar os dados?



Distribuição de frequências!

Distribuição de frequências

A **distribuição de frequências** consiste na organização dados de acordo com as **ocorrências** dos diferentes resultados observados.

Tabelas	Gráficos
<ul style="list-style-type: none">✓ frequência simples✓ frequência em classes✓ contingência ou dupla entrada ou ainda tabela cruzada	<ul style="list-style-type: none">✓ colunas✓ barras✓ setores (pizza)✓ diagrama de Pareto✓ linhas✓ histograma✓ pictograma✓ <i>box plot</i> (diagrama em caixas)

Tabelas

Frequência simples

Tabela 1 – Distribuição de frequências do provedor usado pelo visitante do site.

Provedor	Frequência	Porcentagem
A	10	25,0
B	17	42,5
C	7	17,5
D	6	15,0
Total	40	100,0

Tabela 4 - Satisfação dos alunos do terceiro em relação à aprendizagem no assunto Cônicas

satisfação	número de alunos	%
insatisfatória/ruim	16	17,4%
satisfatória/boa	62	67,4%
muito boa	14	15,2%
Total	92	100%

Fonte: Professor Jeovani

Frequência em classes

Tabela 1 - Valores de pressão arterial sistólica em 49 pessoas.

Academia VIDA - Blumenau - Março/2022.

Valores de pressão (mmHg)	número de pessoas	%
86 - 96	4	8,2%
96 - 106	12	24,5%
106 - 116	10	20,4%
116 - 126	11	22,4%
126 - 136	7	14,3%
136 - 146	4	8,2%
146 - 156	0	0,0%
156 - 166	1	2,0%
Total	49	100%

Tabelas

contingência ou dupla entrada (tabela cruzada)

Os alunos mais satisfeitos com sua aprendizagem são os que mais se dedicaram as atividades remotas?

Tabela 6 - Dedicção aos estudos x satisfação com a aprendizagem

dedicção aos estudos	satisfação com a aprendizagem			Total
	ruim	boa	muito boa	
Eu gostaria de ter me dedicado às atividades propostas mas não tive condições.	4	2	-	6
Eu poderia ter me dedicado mais às atividades propostas.	11	18	2	31
Quase sempre (na maioria das vezes) me dediquei às atividades propostas.	1	28	6	35
Sempre me dediquei às atividades propostas.	-	14	6	20
Total	16	62	14	92

fonte: Professor Jeovani

Elementos de uma tabela

Título
(O que, Onde, Quando) →

Cabeçalho →

Coluna indicadora →

rodapé →

Medalha	Número de Medalhas
Ouro	3
Prata	4
Bronze	8
Total	15

Fonte: <http://olimpiadas.uol.com.br/2008/quadro-de-medalhas/brasil.jhtm>

Corpo

Gráficos

A representação gráfica é usada para aumentar a legibilidade do resultado de uma pesquisa. Os gráficos devem ser autoexplicativos e de fácil compreensão.

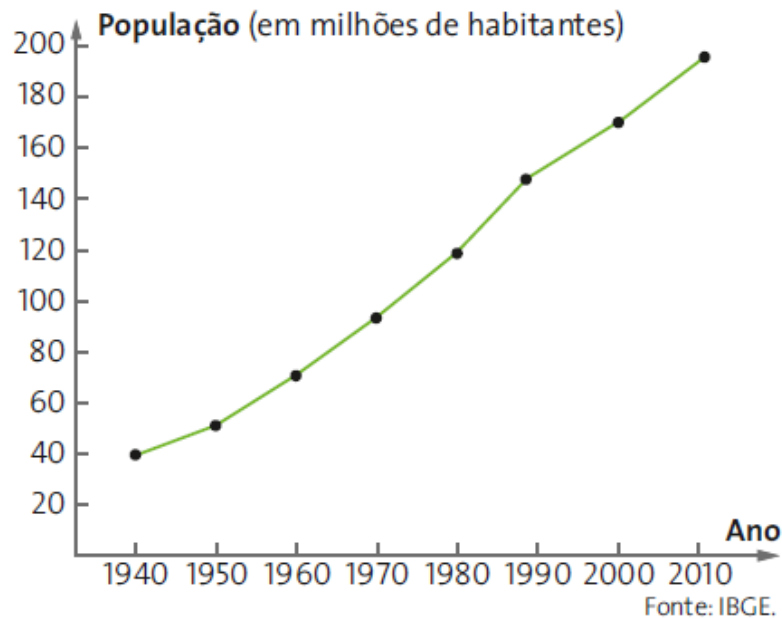


- Ter um título, onde se destaca o fato, o local e o tempo;
- Colocar a fonte de obtenção dos dados, caso não seja o próprio autor que tenha feito a coleta.

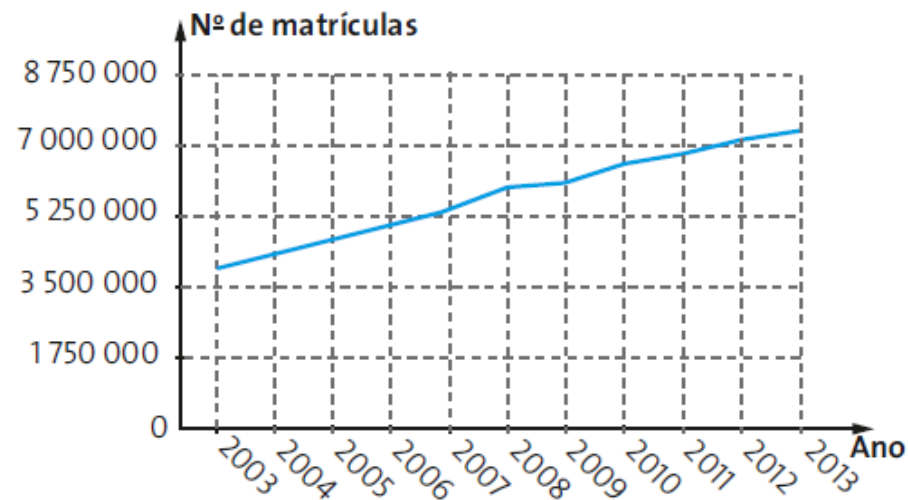
Tipos de Gráficos

Linhas

 Crescimento da população brasileira (1940-2010)



 Número de alunos matriculados no Ensino Superior no Brasil




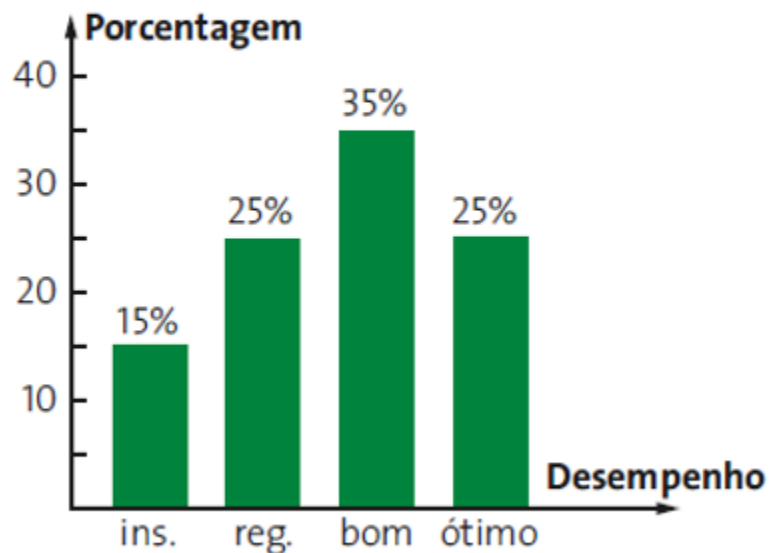
Fonte: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Disponível em: <<http://painel.mec.gov.br/painel.php?modulo=principal/detalhamentoIndicador&acao=A&detalhes=pais&indid=439>>. Acesso em: 11 fev. 2016.

Tipos de Gráficos


Colunas e Barras

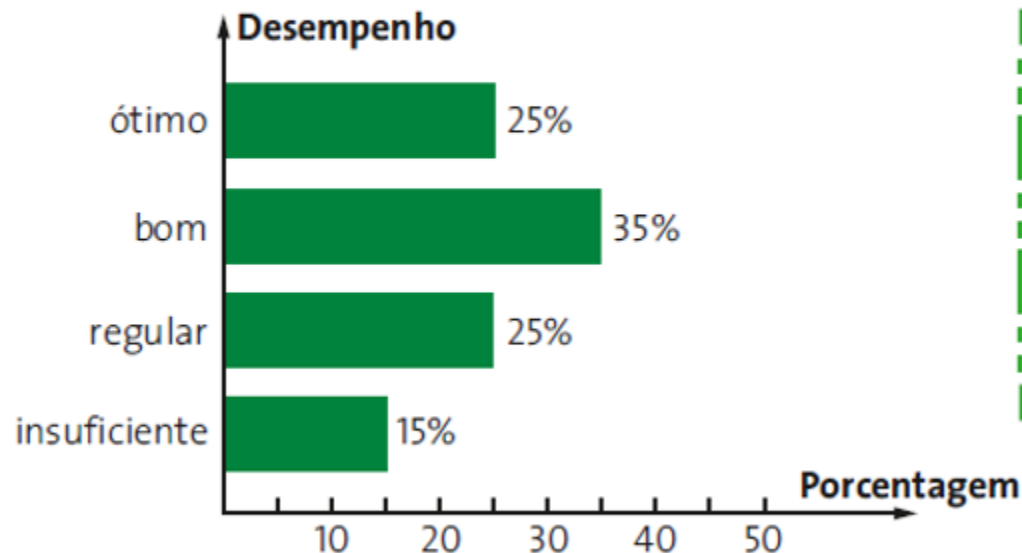
Barras e colunas - São retângulos de mesma base, verticais ou horizontais, com comprimentos proporcionais aos respectivos dados.

 Desempenho em Química dos alunos de uma classe



Fonte: Dados fictícios.

 Desempenho em Química dos alunos de uma classe



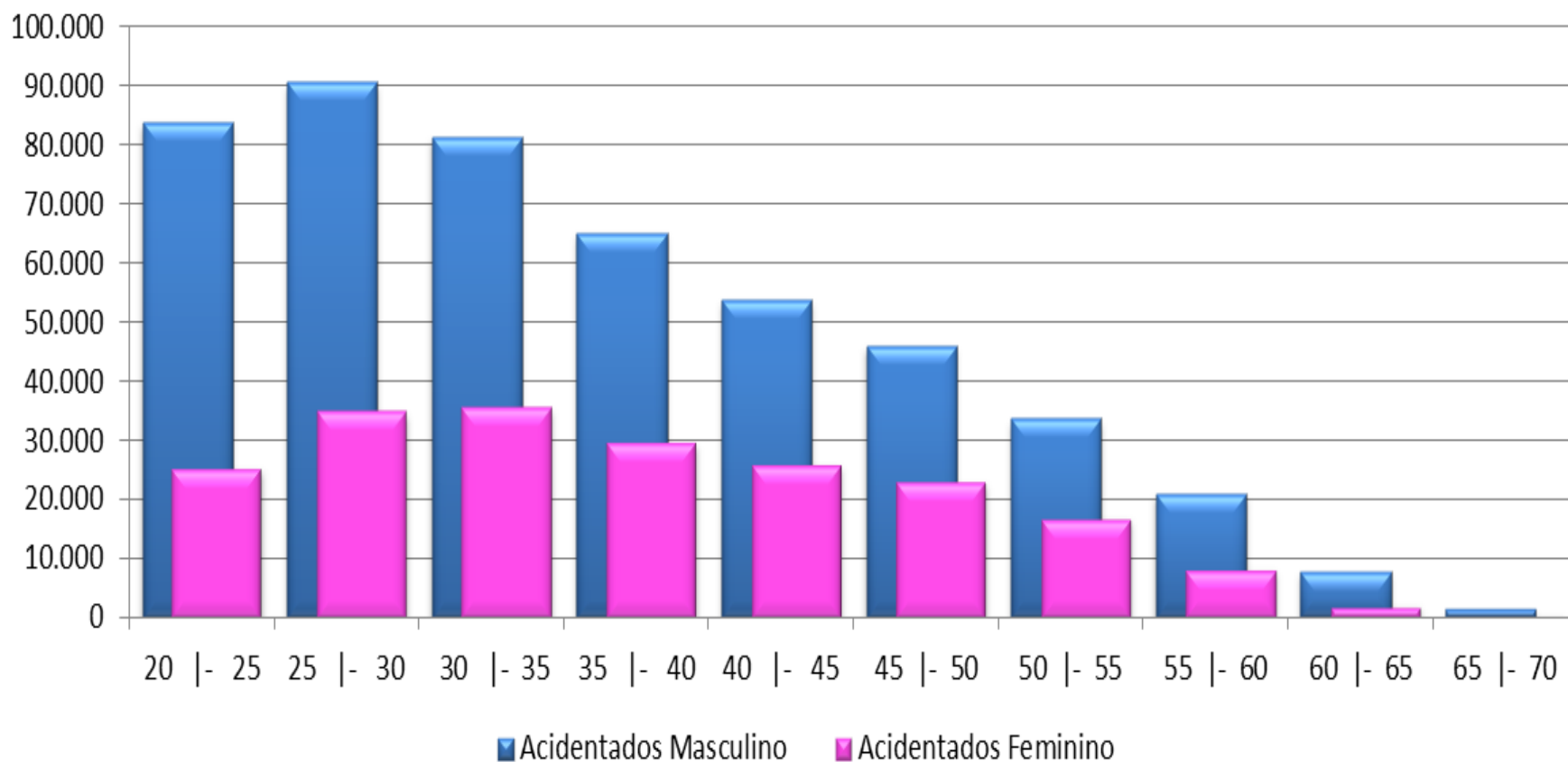
Fonte: Dados fictícios.

Tipos de Gráficos

Barras ou colunas múltiplas

Quando se deseja comparar dois ou mais fenômenos ao mesmo tempo, ao longo de um período ou categorias.

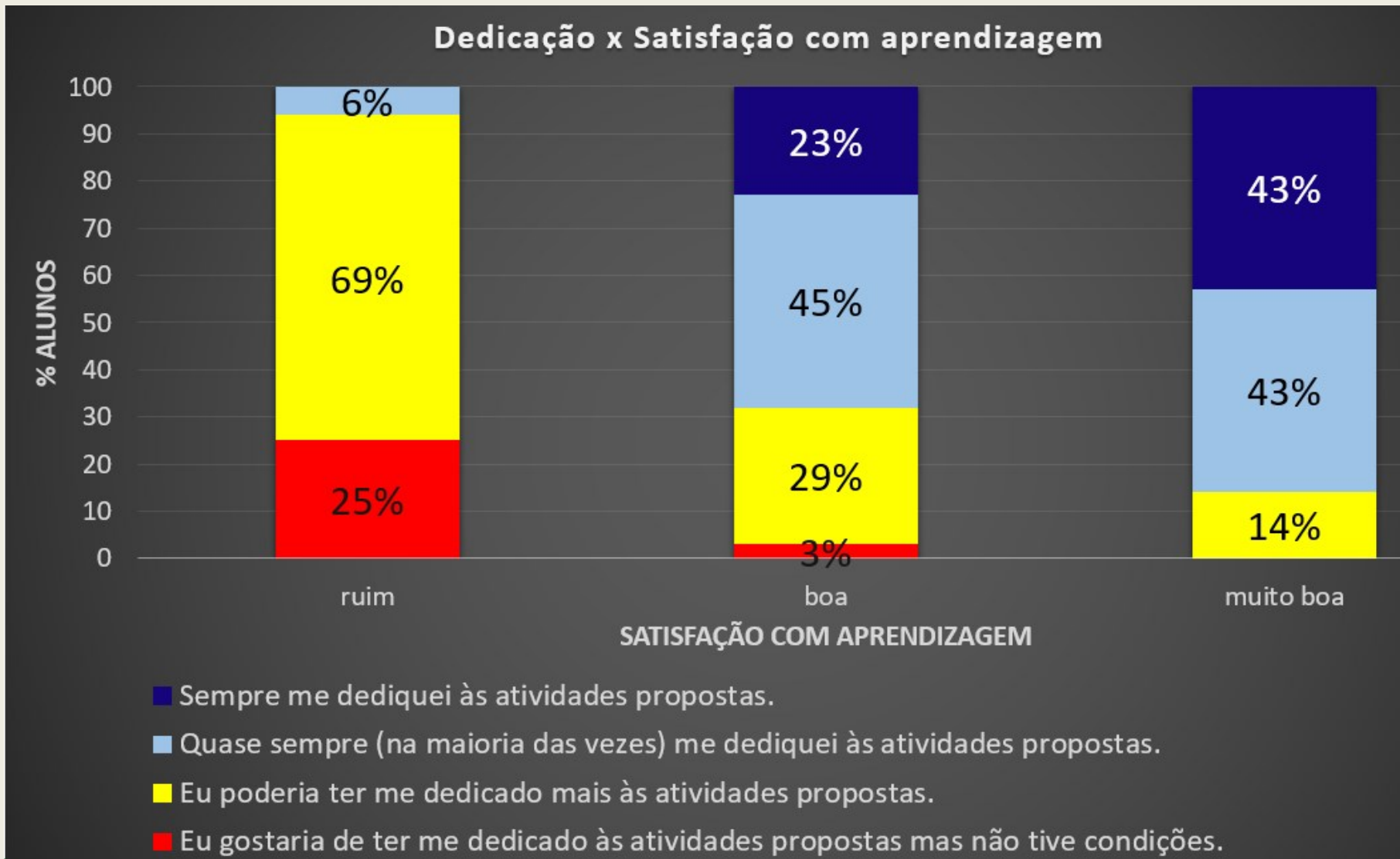
Acidentes de trabalho por idade e sexo, no Brasil, em 2011



Tipos de Gráficos

colunas empilhadas

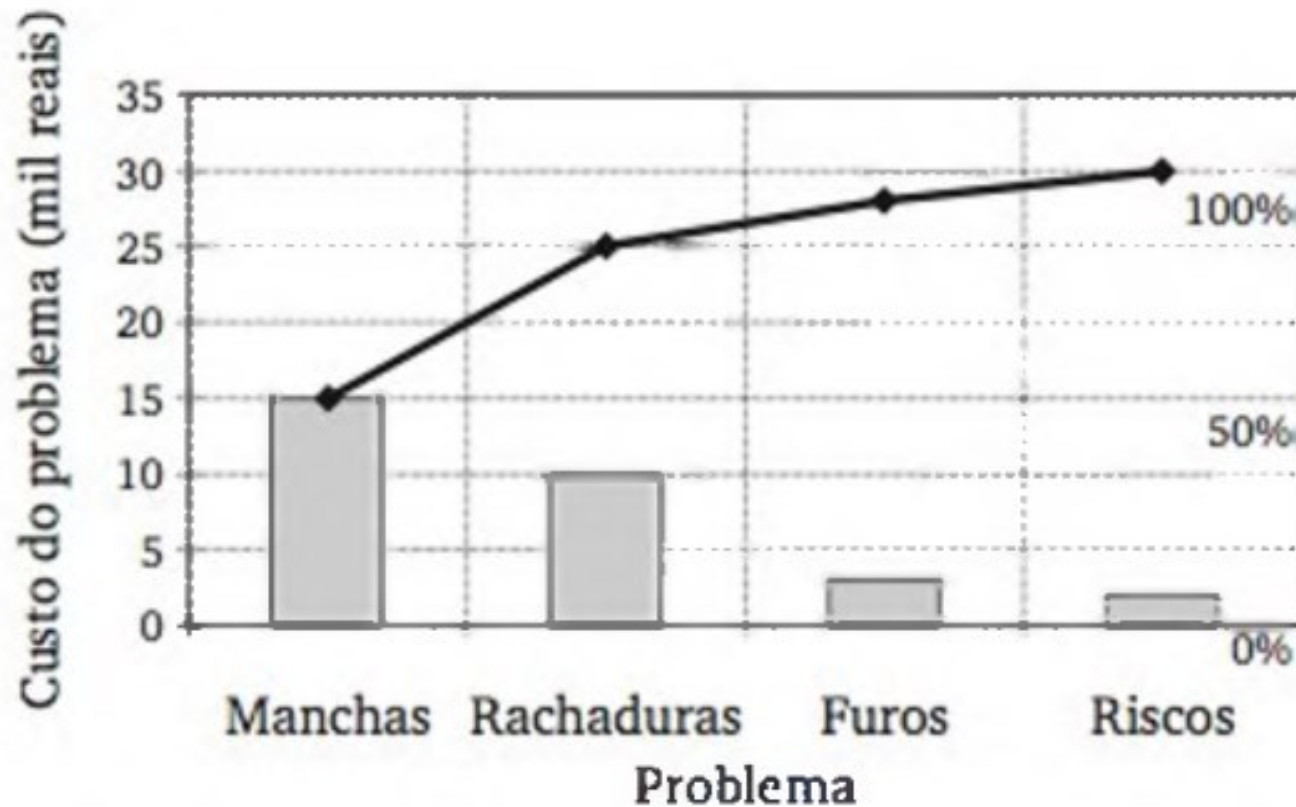
Os alunos mais satisfeitos com sua aprendizagem são os que mais se dedicaram as atividades remotas?



Tipos de Gráficos

Diagrama de Pareto

É uma das ferramentas dos programas de qualidade. As categorias são ordenadas decrescentemente pelas frequências observadas.



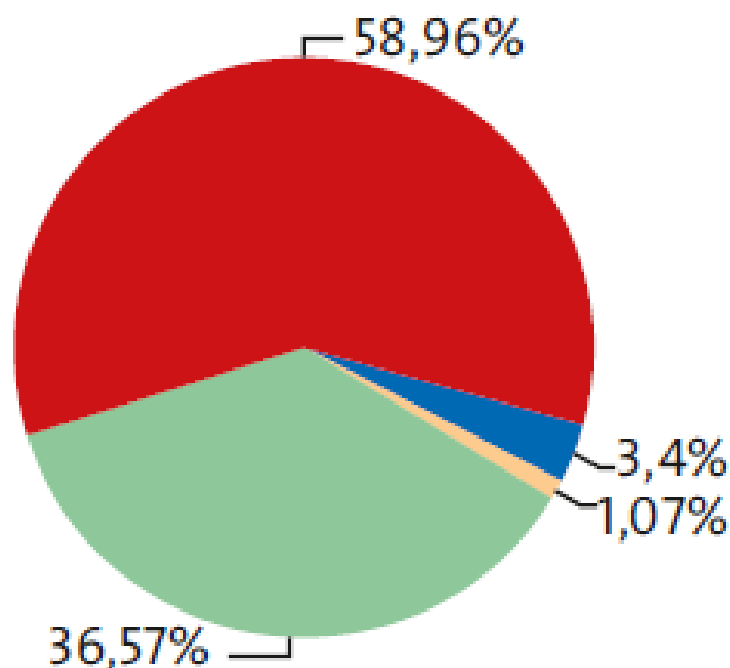
Tipos de Gráficos

Setores (ou pizza)



Bolsas de pós-graduação concedidas no país em 2013

- Pós-Doutorado
- Doutorado
- Mestrado
- Mestrado Profissional



Banco de imagens/Arquivo da editora

Fonte: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes). Disponível em: <<http://painel.mec.gov.br/painel.php?modulo=principal/detalhamentoIndicador&acao=A&detalhes=pais&in-did=114>>. Acesso em: 10 maio 2016.

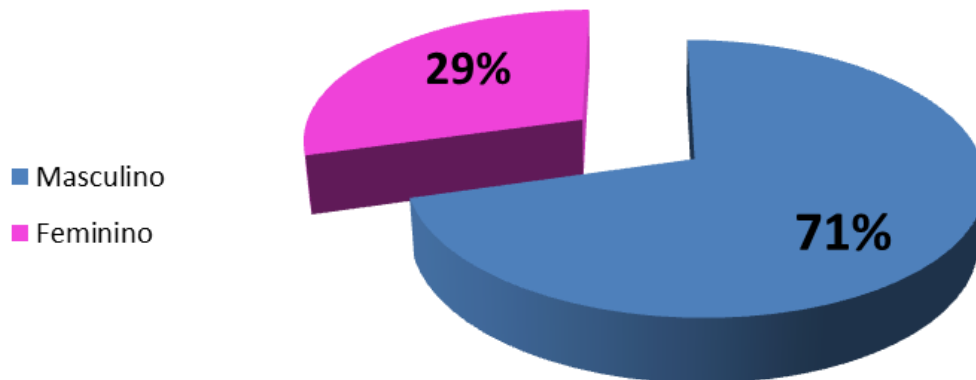
Tipos de Gráficos

Setores (pizza)

Cada setor tem um ângulo de abertura proporcional aos respectivos dados. O tamanho total da amostra corresponde a 360° do gráfico.

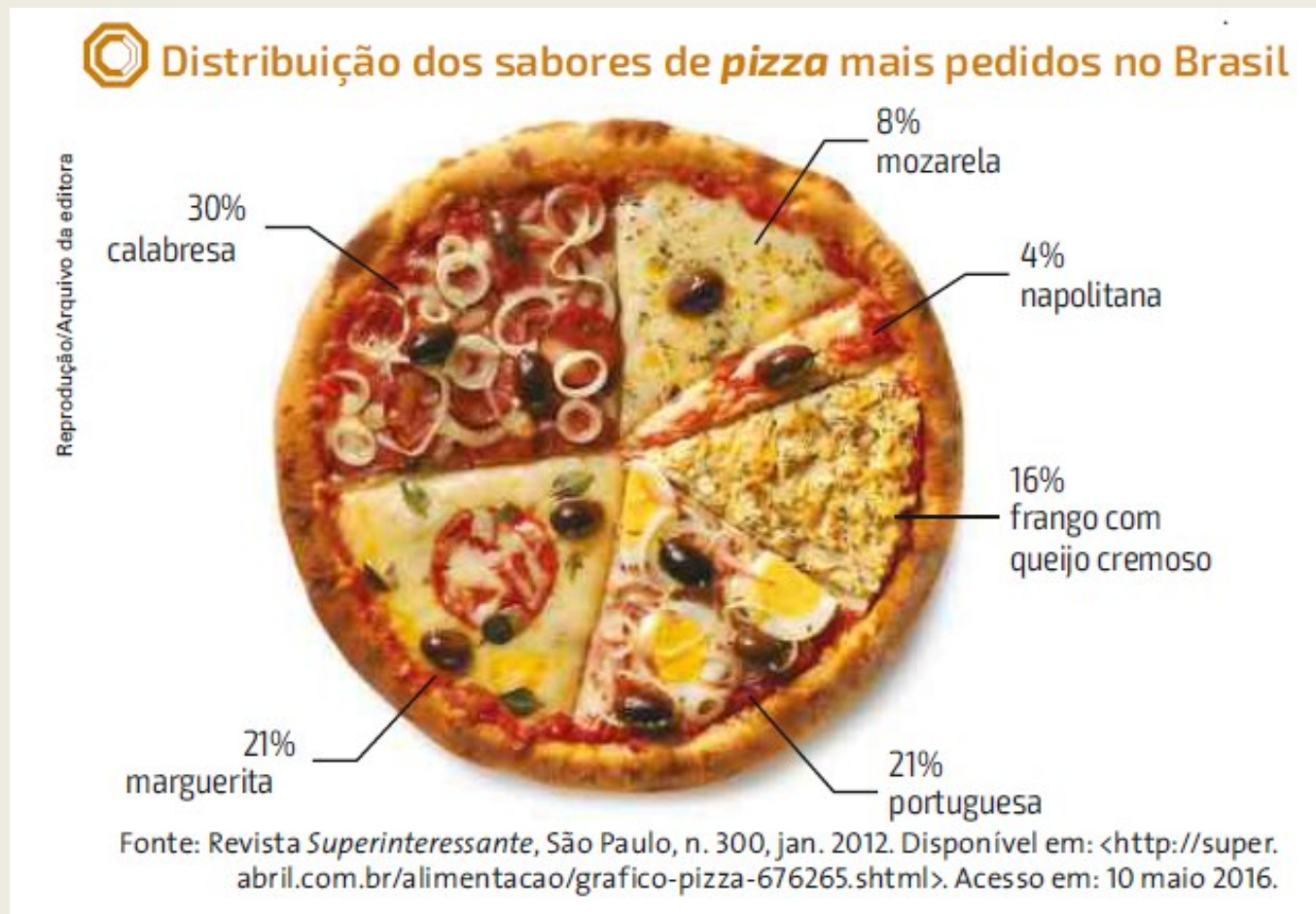
$$\frac{\text{Total da amostra}}{\text{cada frequência}} = \frac{360^\circ}{x^\circ}$$

Acidentes trabalhistas por sexo, no Brasil, em 2011



Tipos de Gráficos

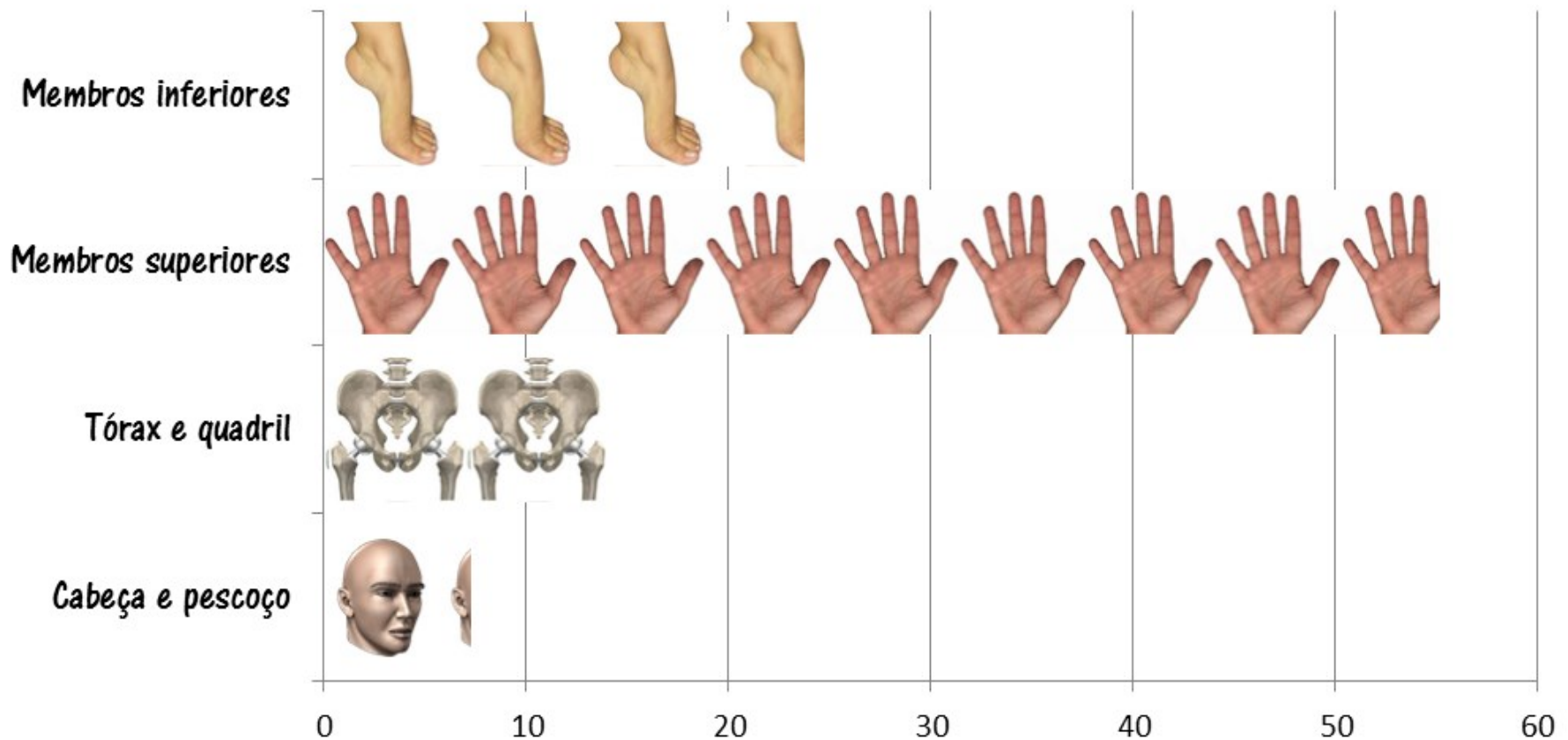
Pictograma Usam imagens que guardam relação com o assunto exposto! Desperta atenção e curiosidade!



Tipos de Gráficos

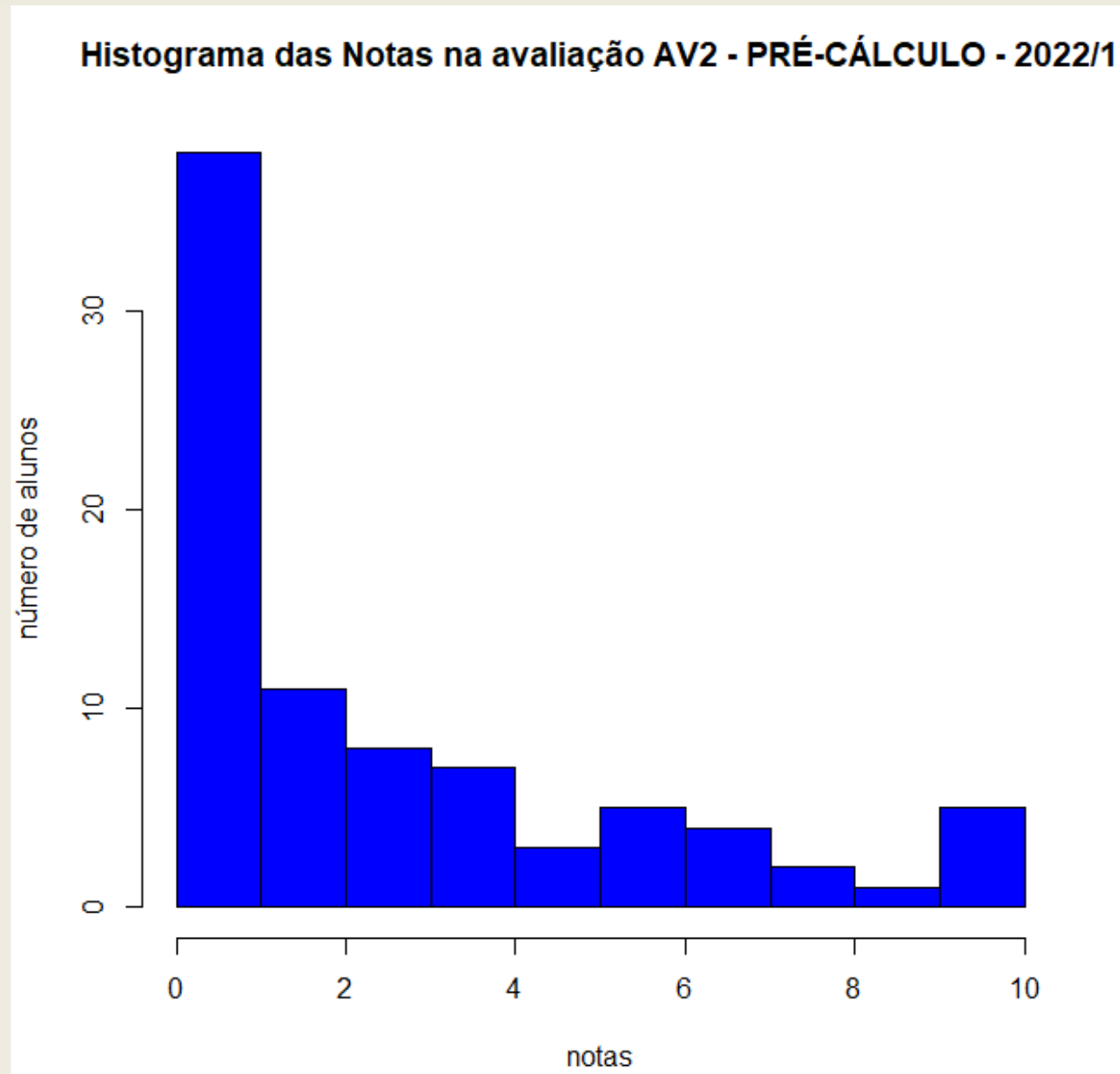
Pictograma

Partes do corpo mais afetadas por AT em 2011, no Brasil



Tipos de Gráficos

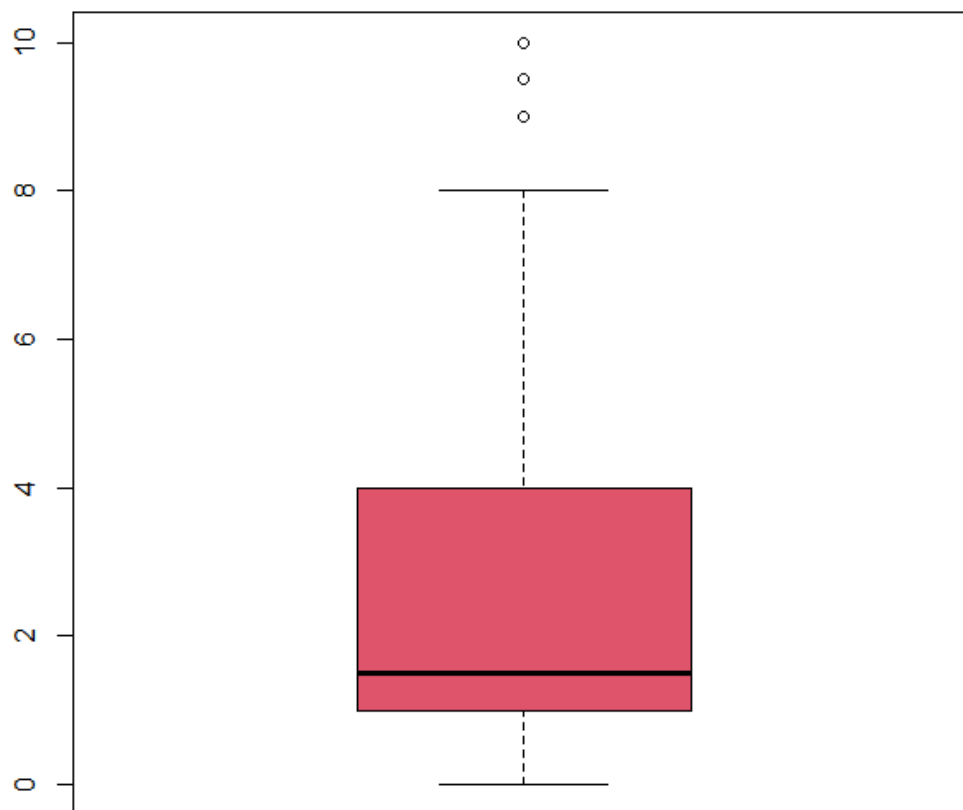
Histograma



Tipos de Gráficos

Box Plot (Diagrama em caixas)

Notas na avaliação AV2 - PRÉ-CÁLCULO - 2022/1



Permite avaliar facilmente os valores típicos, a assimetria da distribuição, a dispersão e os dados discrepantes (*outliers*) de conjuntos de dados referentes a variáveis quantitativas.