

# CURSO DE ANALISTA CRIMINAL

**Disciplina:** Estatística criminal

**Professor:** Ana Paula Gonçalves (UFMG),

Doutora em Sociologia



# Por que aprender estatística?

- Transformar dados e informações em evidências científicas;
- Sistematizar grande quantidade de informações;
- Buscar entender as generalizações;
- Contribui para a tomada de decisão;
- Pode resolver problemas do cotidiano.



# Qual livro usar para aprender estatística?

- Aquele que você melhor se adaptar.
- Algumas sugestões:

BARBETTA, Pedro Alberto. Estatística aplicada às Ciências Sociais. Florianópolis: EDUFSC,1994.

BUSSAB, Wilton; MORETTIN, Pedro A. Estatística básica. São Paulo: Saraiva, 2010.

FIELD, Andy. Descobrindo estatística usando SPSS. Porto Alegre: Artmed, 2009.

LEVIN, Jack; FOX, James A.; FORDE, David. Estatística para ciências humanas. São Paulo:Prentice Hall, 2004.

WILD, Christopher J.; SEBER, George A. F. Encontros com o acaso. Rio de Janeiro: LTC, 2004.

# Qual software usar?

## Comparação entre softwares estatísticos mais comuns

Software	Gratuito	Interface gráfica	Flexibilidade	Variedade de modelos e/ou módulos	limite de casos	Suporte técnico
Excel	não	sim	baixo	não	1.048.576 linhas por 16.384 colunas	Profissional
Libreoffice	sim	sim	baixo	não	1024 colunas e 1048576 linhas	Comunidade
Spss	não	sim	moderado	sim	mais de 2 milhões de registros e 250.000 variáveis.	Profissional e Comunidade
Jamovi	sim	sim	moderado	não	ilimitado	Comunidade extensa
Stata	não	não	Alta	sim	ilimitado	Profissional e Comunidade
R	sim	não	Alta	sim	Ilimitado	Comunidade e documentação extensa
Python	sim	não	Alta	sim	Ilimitado	Comunidade extensa

# Jamovi

- Gratuito;
  - Código aberto;
  - Interface gráfica amigável;
  - Realiza as principais análises estatísticas,
  - Possibilidade de inserir módulos específicos que são disponibilizados pela comunidade.
- 
- Há uma versão de instalação em **desktop**, para usar no computador, e uma versão para usar na **nuvem**.
  - Para realizar download do Jamovi, deve-se acessar o seguinte endereço eletrônico

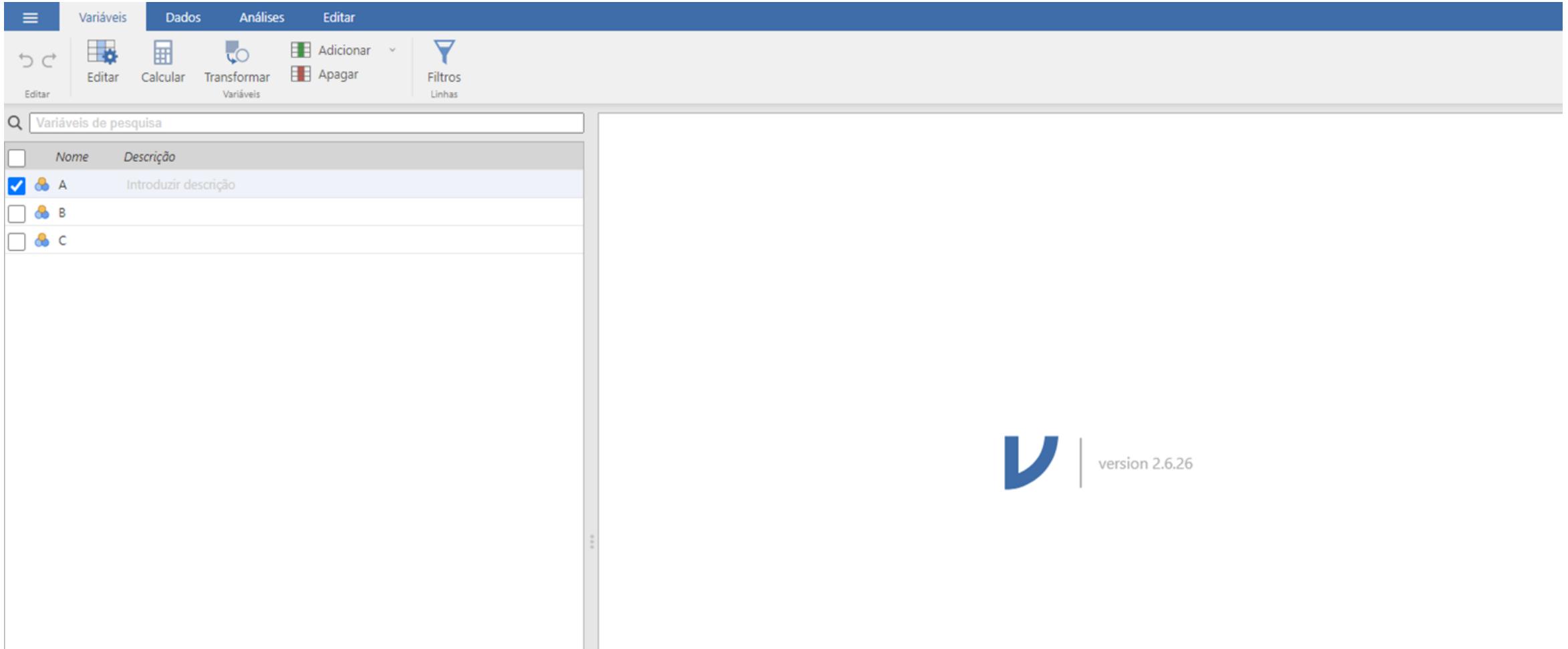
<https://www.jamovi.org/>.

# Jamovi

The screenshot shows the Jamovi software interface. The top navigation bar includes tabs for 'Variáveis', 'Dados' (selected), 'Análises' (highlighted in blue), and 'Editar'. Below the navigation bar are icons for various statistical analyses: Exploração, Testes t, ANOVA, Regressão, Frequências, and Fator. On the far right of the top bar are 'Módulos' and other standard window controls. The main area features a data grid with columns labeled A, B, and C. Column A has a single value in row 1. The y-axis on the left is numbered from 1 to 23. In the bottom right corner, there is a logo consisting of a stylized blue 'V' shape followed by the text 'version 2.6.26'.

version 2.6.26

# Jamovi



The screenshot shows the Jamovi software interface. The top navigation bar includes 'Variáveis', 'Dados', 'Análises', and 'Editar' tabs. Below the navigation bar are various icons for editing, calculating, transforming, adding, and deleting variables, along with filters and lines tools. A search bar labeled 'Variáveis de pesquisa' is present. The main area displays a list of variables:

	Nome	Descrição
<input checked="" type="checkbox"/>	A	Introduzir descrição
<input type="checkbox"/>	B	
<input type="checkbox"/>	C	

At the bottom right, there is a blue 'V' logo followed by the text 'version 2.6.26'.

# Jamovi

The screenshot shows the Jamovi software interface. The top navigation bar includes tabs for 'Variáveis' (Variables), 'Dados' (Data), 'Análises' (Analyses), and 'Editar'. The 'Dados' tab is currently selected. Below the navigation bar is a toolbar with icons for 'Colar' (Paste), 'Clipboard', 'Configurar' (Configure), 'Calcular' (Calculate), 'Transformar' (Transform), 'Pesos' (Weights), 'Adicionar' (Add), 'Apagar' (Delete), 'Filtros' (Filters), and 'Linhas' (Rows). The main area displays a data grid with three columns labeled A, B, and C. Column A has a single value 'A' in row 1. The data grid has a vertical row header on the left ranging from 1 to 33. The title bar at the top right reads 'jamovi - Untitled'.

**jamovi - Untitled**

Variáveis Dados Análises Editar

Colar Clipboard Configurar Calcular Transformar Pesos Adicionar Apagar Filtros Linhas

	A	B	C
1	A		
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			
25			
26			
27			
28			
29			
30			
31			
32			
33			

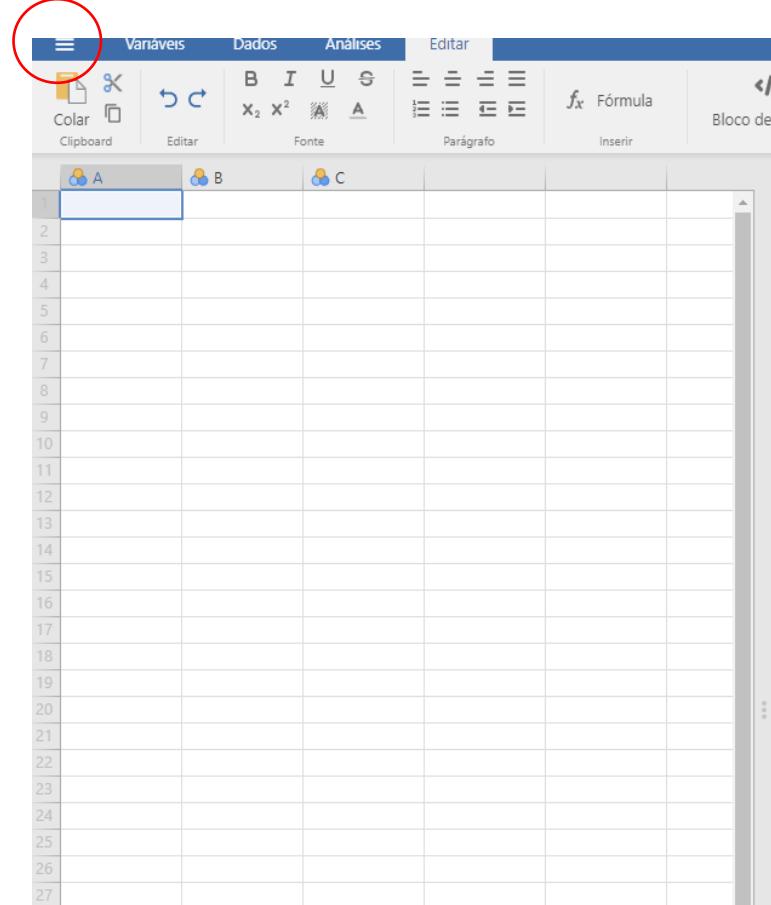


# Jamovi

The screenshot shows the Jamovi software interface. On the left is a spreadsheet view with columns labeled A, B, and C, and rows numbered 1 to 27. The first row (header) contains icons for 'Colar' (Paste), 'Clipboard', 'Editar' (Edit), 'Fonte' (Font), and 'Parágrafo' (Paragraph). The second row contains icons for 'B' (Bold), 'I' (Italic), 'U' (Underline), 'S' (Strikethrough), ' $x_1$ ' (Superscript), ' $x^2$ ' (Subscript), ' $A$ ' (Text), and ' $A$ ' (Text). The third row contains icons for 'Fórmula' (Formula), 'Inserir' (Insert), 'Bloco de Código' (Code Block), 'Cabeçalho' (Header), and 'Link'. The right side of the interface displays the 'Resultados' (Results) panel, which is currently empty. Below the results panel is a section titled 'Referências' (References) containing two entries:

- [1] The jamovi project (2024). *jamovi*. (Version 2.6) [Computer Software]. Retrieved from <https://www.jamovi.org>.
- [2] R Core Team (2024). *R: A Language and environment for statistical computing*. (Version 4.4) [Computer software]. Retrieved from <https://cran.r-project.org>. (R packages retrieved from CRAN snapshot 2024-08-07).

# Como importar um banco de dados



A screenshot of the 'Abrir' (Open) dialog in Jamovi. The dialog shows a list of files in the 'Documentos &gt; UFMG &gt; pesquisa vitimização' folder. The files listed are: 'Base Scol Completa Jan 2023 a Dez 2023\_hom\_dol.xlsx', 'SERIE HISTORICA(2020)\_202503121618.csv', 'SERIE HISTORICA(2021)\_202503121607.csv', and 'SERIE HISTORICA(2022)\_202503071042.csv'. A red box highlights the first file, 'Base Scol Completa Jan 2023 a Dez 2023\_hom\_dol.xlsx'. A text box with a red border contains the instruction: 'Para abrir um arquivo de Excel ele deve ter somente uma aba' (To open an Excel file, it must have only one sheet).

# Como importar um banco de dados

Screenshot of a data analysis software interface showing a table of imported data.

The interface includes a top navigation bar with tabs: Variáveis, Dados, Análises, and Editar. Below the navigation bar is a toolbar with icons for Colar (Clipboard), Editar, Configurar, Calcular, Transformar, Pesos, Adicionar (Add), Apagar (Delete), Filtros (Filters), and Linhas (Rows).

The main area displays a table with the following columns:

	anoBoletim	autoria_c...	Qtde_Aut...	Data do F...	MÊS	hora	Data Eluci...	municipio	bairro	F
1	2023	SIM	0	2023-01-02		1 08:10:00	2023-09-28	CURITIBA	ABRANCHES	R
2	2023	NÃO	0	2023-03-24		3 17:00:00		CURITIBA	ABRANCHES	R
3	2023	SIM	0	2023-11-13		11 08:05:00	2023-11-21	CURITIBA	ABRANCHES	R
4	2023	NÃO	0	2023-08-06		8 11:40:00		CURITIBA	ABRANCHES	R
5	2023	NÃO	0	2023-05-01		5 13:30:00		CURITIBA	ABRANCHES	R
6	2023	NÃO	0	2023-04-24		4 23:40:00		CURITIBA	ABRANCHES	R
7	2023	NÃO	0	2023-09-28		9 15:00:00		CURITIBA	ABRANCHES	R
8	2023	NÃO	0	2023-11-23		11 11:00:00		CURITIBA	ABRANCHES	R
9	2023	NÃO	0	2023-04-12		4 10:54:00		CURITIBA	AGUA VERDE	T
10	2023	NÃO	0	2023-03-23		3 11:00:00		CURITIBA	AGUA VERDE	R
11	2023	NÃO	0	2023-03-19		3 11:11:00		CURITIBA	AGUA VERDE	R
12	2023	NÃO	0	2023-12-18		12 09:53:00		CURITIBA	AGUA VERDE	R
13	2023	NÃO	0	2023-01-04		1 02:50:00		CURITIBA	AGUA VERDE	A
14	2023	NÃO	0	2023-01-17		1 15:40:00		CURITIBA	AGUA VERDE	R
15	2023	NÃO	0	2023-09-29		9 11:10:00		CURITIBA	AGUA VERDE	A
16	2023	SIM	0	2023-10-14		10 14:48:00	2023-10-14	CURITIBA	AGUA VERDE	A
17	2023	NÃO	0	2023-07-10		7 22:30:00		CURITIBA	AGUA VERDE	R
18	2023	NÃO	0	2023-07-07		7 03:30:00		CURITIBA	AGUA VERDE	A
19	2023	NÃO	0	2023-07-04		7 10:51:00		CURITIBA	AGUA VERDE	R
20	2023	NÃO	0	2023-06-05		6 10:40:00		CURITIBA	AGUA VERDE	A
21	2023	NÃO	0	2023-05-22		5 01:00:00		CURITIBA	AGUA VERDE	R
22	2023	NÃO	0	2023-01-23		1 09:20:00		CURITIBA	AGUA VERDE	R
23	2023	NÃO	0	2023-02-07		2 02:00:00		CURITIBA	AGUA VERDE	R
24	2023	NÃO	0	2023-03-12		3 08:30:00		CURITIBA	AGUA VERDE	R
25	2023	NÃO	0	2023-03-30		3 19:20:00		CURITIBA	AGUA VERDE	A
26	2023	NÃO	0	2023-11-23		11 05:21:00		CURITIBA	AGUA VERDE	T
27	2023	NÃO	0	2023-12-09		12 17:26:00		CURITIBA	AGUA VERDE	A
28	2023	NÃO	0	2023-08-03		8 23:30:00		CURITIBA	AHU	R
29	2023	NÃO	0	2023-01-17		1 04:30:00		CURITIBA	AHU	R
30	2023	NÃO	0	2023-05-25		5 21:08:00		CURITIBA	AHU	R
31	2023	NÃO	0	2023-03-29		3 13:00:00		CURITIBA	ALTO BOQUE...	R
32	2023	NÃO	0	2023-01-09		1 09:00:00		CURITIBA	ALTO BOQUE...	R
33	2023	NÃO	0	2023-02-19		2 20:23:00		CURITIBA	ALTO BOQUE...	R
34	2023	NÃO	0	2023-02-19		2 20:23:00		CURITIBA	ALTO BOQUE...	R
35	2023	NÃO	0	2023-02-14		2 21:15:00		CURITIBA	ALTO BOQUE...	R
36	2023	NÃO	0	2023-09-27		9 00:01:00		CURITIBA	ALTO BOQUE...	R
37	2023	SIM	0	2023-09-21		9 16:40:00	2023-10-16	CURITIBA	ALTO BOQUE...	R
38	2023	NÃO	0	2023-09-10		9 19:17:00		CURITIBA	ALTO BOQUE...	R

Ativar o Windows  
Acesse Configurações para ativar o W



version 2.6.26

# Salvando o banco de dados



Novo

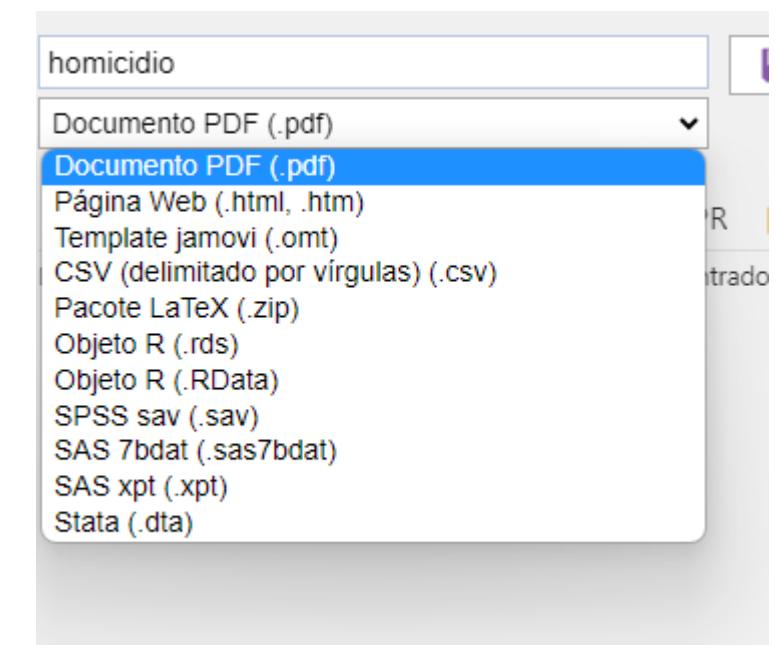
Abrir

Importar especial

Guardar

Guardar como

Exportar



# O que é uma variável de pesquisa?

- A variável é a característica de interesse que é medida em cada elemento da amostra ou população.
- A construção de uma variável está relacionada com a transformação da pergunta de pesquisa em variáveis de matrizes de dados.
- Uma variável também pode ser um dado já produzido que tem relação com a sua pergunta de pesquisa.
- A escolha das variáveis que serão utilizadas pelo pesquisador tem relação com a teoria e com a sua pergunta de pesquisa.

# Formas de mensurar uma variável

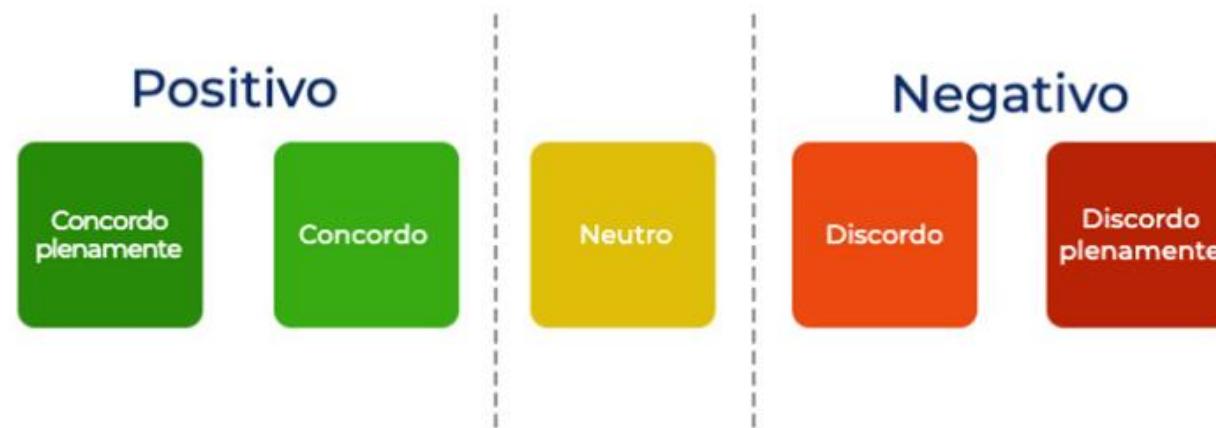
- **Qualitativas ou categóricas:** são as características que não possuem valores quantitativos, mas representam uma classificação dos indivíduos. Essas variáveis qualificam os indivíduos.
- **Nominal:**
  - ✓ É o nível de mensuração que só nomeia ou rotula os indivíduos.
  - ✓ Não existe ordenação entre as categorias.

Ex: Sexo, cor ou raça, partidos políticos, meio de transporte usado para ir ao trabalho.

# Formas de mensurar uma variável

- **Ordinal:**

- ✓ É o nível de mensuração que nomeia e rotula os indivíduos, mas há uma ordenação entre as categorias.  
Ex: Escolaridade ( 8º anos, 9º ano, ....), Estágio de uma doença ( inicial, intermediário, terminal), intervalo de renda por salários mínimos ( até 1 SM, de 1 SM até 5 SM), Escalas.
- ✓ Escala Likert: Usada para entender o comportamento e preferência dos indivíduos.



# Formas de mensurar uma variável

- **Quantitativas:**
  - ✓ São as características que podem ser medidas em uma escala numérica.
  - ✓ Sua ordenação faz sentido.
- **Intervalar**
  - ✓ São variáveis numéricas cujo zero não representa ausência de valor.  
Ex: Temperatura em graus celsius, Posição em km que o indivíduo inicia um viagem em uma rodovia.
- **Razão:**
  - ✓ São variáveis numéricas cuja quantificação está relacionada com o zero como ausência de valor (zero absoluto).  
Ex: peso, altura, salário em reais

# Formas de mensurar uma variável

## ATENÇÃO

- Uma variável originalmente quantitativa pode ser coletada de forma qualitativa.
  - ✓ Idade em anos completo = quantitativa
  - ✓ Faixa etária = qualitativa
  - ✓ Salário em reais = quantitativa
  - ✓ Faixa de salários mínimos = qualitativa
- Nem sempre uma variável representada por números é quantitativa.  
Ex: número do telefone de uma pessoa, número da casa, número de sua identidade.
- Importante: caso você observem variáveis qualitativas representadas por números é só a codificação das respostas.  
Ex: 1 = Feminino e 2 = Masculino; 1 = sim e 2 = não

# Formas de mensurar uma variável

População	Variável nominal	Variável ordinal	Variável escalar ou de razão
Estudantes	sexo	Grau de escolaridade	Anos de Estudo
Chocolate	Tipo	Qualidade do produto	Preço
Domicílios	ter saneamento	Tipo de coleta de esgoto	Quantidade de moradores

# Banco de dados para as aulas

- Banco de dados 1 – Homicídios  
Informações sobre as vítimas de homicídio em 2023
- Banco de dados 2 - Municípios  
Quantidade de crimes por municípios no Paraná em 2024 e variáveis de contexto

# Categorizando as variáveis no Jamovi

- O Jamovi define as variáveis a seguinte forma:
  1. Nominal,
  2. Categórica,
  3. Contínua ( variáveis de razão e intervalar).
  4. ID para o banco de dados ter uma chave primária.

The screenshot displays the Jamovi software interface. The top navigation bar includes tabs for 'Variáveis', 'Dados' (selected), 'Análises', and 'Editar'. Below the toolbar, there are various data management tools like 'Colar' (Paste), 'Editar' (Edit), 'Configurar' (highlighted with a red circle), 'Calcular', 'Transformar', 'Pesos', 'Adicionar', 'Apagar', 'Filtros', 'Adicionar', 'Apagar', and 'Linhas'. The main workspace is titled 'jamovi - homicidio.xlsx' and shows the 'VARIÁVEL DE DADOS' configuration for the variable 'Entidade'. The 'Entidade' field is highlighted with a blue border. The configuration panel includes fields for 'Descrição', 'Tipo de medida' (set to 'ID'), 'Tipo de Dados' (set to 'Texto'), and 'Valores omissos'. To the right, a 'Níveis' (Levels) section lists 14 entries, each labeled 'NI'. At the bottom, a data grid shows 14 rows of data with columns for 'Entidade', 'Natureza', 'IP', 'Natureza...', 'aisp', and 'dhmc'. The bottom right corner of the interface shows the Jamovi logo and 'version 2.6.26'.

Entidade	Natureza	IP	Natureza...	aisp	dhmc
DIVISAO DE ...	HOMICIDIO ...	IP 828/2023		1	NI
DIVISAO DE ...	MORTE A AP...	IP 81445/2023		1	NI
DIVISAO DE ...	HOMICIDIO ...	IP 301831/20...		1	NI
DIVISAO DE ...	SUICIDIO	IP 206206/20...		1	NI
DIVISAO DE ...	HOMICIDIO ...	IP 115528/20...		1	NI
DIVISAO DE ...	MORTE A AP...	IP 109870/20...		1	NI
DIVISAO DE ...	MORTE A AP...			1	NI
DIVISAO DE ...	MORTE A AP...			1	NI
DIVISAO DE ...	SUICIDIO	IP 99386/2023		1	NI
DIVISAO DE ...	MORTE A AP...	IP 80215/2023		1	NI
DIVISAO DE ...	MORTE A AP...	IP 74587/2023		1	NI
DIVISAO DE ...	SUICIDIO	IP 335256/20...		1	NI
DIVISAO DE ...	SUICIDIO	IP 2749/2023		1	NI
DIVISAO DE ...	ACIDENTE DE...	IP 26287/2023		1	NI

# Categorizando as variáveis no Jamovi

## Banco de dados 1

VARIÁVEL DE DADOS

Entidade

texto descrevendo a variável

Tipo de medida: Nominal

Níveis:

- AO DE HOMICÍDIOS E
- RITO POLICIAL METRO
- GACIA DE DELITOS DE
- 3.º DELEGACIA REGIONAL I

Inserir código das não respostas , não se aplica e vazio.

Para variáveis ordinais. É possível escrever o rótulos dos números e ajustar os níveis hierárquicos

Tipos de Dados

Valores omissos

Reter os níveis não utilizados nas análises

The screenshot shows the 'VARIÁVEL DE DADOS' (Variable Properties) dialog in Jamovi. The variable 'Entidade' is selected. The 'Tipo de medida' (Type of measurement) is set to 'Nominal'. The 'Níveis' (Levels) dropdown menu is open, displaying categorical values: 'AO DE HOMICÍDIOS E', 'RITO POLICIAL METRO', 'GACIA DE DELITOS DE', and '3.º DELEGACIA REGIONAL I'. A red box highlights this list with the text 'Para variáveis ordinais. É possível escrever o rótulos dos números e ajustar os níveis hierárquicos'. A red arrow points from the 'Valores omissos' (Missing values) section to the 'Níveis' dropdown. Another red box contains the text 'Inserir código das não respostas , não se aplica e vazio.' A red arrow points from this text to the 'Valores omissos' section. At the bottom, there is a toggle switch for 'Reter os níveis não utilizados nas análises' (Retain unused levels in analyses).

	Entidade	Natureza	IP	Natureza ...	aisp	dhmc
1	DIVISAO DE ...	HOMICIDIO ...	IP 828/2023		1	NI
2	DIVISAO DE ...	MORTE A AP...	IP 81445/2023		1	NI
3	DIVISAO DE ...	HOMICIDIO ...	IP 301831/20...		1	NI
4	DIVISAO DE ...	SUICIDIO	IP 206206/20...		1	NI
5	DIVISAO DE ...	HOMICIDIO ...	IP 115528/20...		1	NI
5	DIVISAO DE ...	MORTE A AP...	IP 109870/20...		1	NI
7	DIVISAO DE ...	MORTE A AP...			1	NI
8	DIVISAO DE ...	MORTE A AP...			1	NI
9	DIVISAO DE ...	SUICIDIO	IP 99386/2023		1	NI
0	DIVISAO DE ...	MORTE A AP...	IP 80215/2023		1	NI
1	DIVISAO DE ...	MORTE A AP...	IP 74587/2023		1	NI



VE

# Categorizando as variáveis no Jamovi

## Banco de dados 2

- 1) Abrir o banco de dados 2 – municípios e verificar se as variáveis estão categorizadas corretamente.
- 2) Ajustar as variáveis necessárias

# Adicionando variáveis

The screenshot shows the jamovi software interface for analyzing a dataset named "homicidio.xlsx". The top navigation bar includes tabs for Variáveis, Dados, Análises, and Editar. The Dados tab is active. The toolbar contains various icons for clipboard operations (Colar, Clipboard), configuration (Configurar), calculation (Calcular), transformation (Transformar), weights (Pesos), and variable management (Adicionar, Apagar). The "Adicionar" icon is circled in red.

The main workspace displays a "VARIÁVEL DE DADOS" configuration panel for a variable named "Idade". It includes fields for "Descrição", "Tipo de medida" (set to "Continua"), "Tipo de Dados" (set to "Inteiro"), and "Valores omissos". To the right is a "Níveis" panel with up and down arrows for reordering levels, and a plus sign (+) to add new levels. A toggle switch at the bottom left enables/disables levels not used in analyses.

At the bottom, a data grid lists rows from 1 to 14, with columns for "Nome\_vitimado", "Apelido", "Y", and "Id". A context menu is open over the "Id" column, with the "Adicionar variável" option highlighted and circled in red. This menu also includes options like "Cortar", "Copiar", "Colar", "Apagar variável", "Configurar...", and "Transformar...".

The bottom status bar indicates the dataset has 6484 rows, 0 filters applied, 0 rows deleted, and 0 cells edited.

A blue logo and the text "version 2.6.26" are visible in the bottom right corner.

Pronto	Filtros 0	Contagem em linha 6484	Filtrado 0	Apagado 0	Adicionado 0	Cells edited 0
1 IS DE...	HE-MAN					
2 ARLO...						
3 KUBIS						
4 A SIL...						
5 N CEZ...						
6 OSO ...						
7 QUI...						
8 ANO ...						
9 KAZ...						
10 SCHE...						
11 O L...						
12 HAS...						
13 TEIX...						
14 D BA...						

# Calculando variáveis

- 1) Uso da variável calculada no banco de dados 2
- 2) Somar as variáveis estelionato + estelionato e fraude eletrônica

The screenshot shows a software interface for managing variables in a dataset. On the left, a context menu is open with the following options:

- Cortar
- Copiar
- Colar
- Adicionar variável >** (highlighted)
- Apagar variável
- Configurar...
- Transformar...

Under the "Adicionar variável" option, there are three main categories:

- Variável de Dados**
  - Inserir
  - Acrescentar
- Variável calculada**
  - Inserir
  - Acrescentar
- Variável transformada**
  - Inserir
  - Acrescentar

The "Variável calculada" section is currently active, showing a dialog box titled "VARIÁVEL CALCULADA". Inside the dialog, the variable name "Estelionato total" is entered in the "Nome" field, and the formula `= ESTELIONATO+`ESTELIONATO FRAUDE ELETRONICA`` is entered in the "Fórmula" field. To the right of the formula, there are two dropdown menus: "Funções" (Functions) and "Variáveis" (Variables). The "Funções" menu is set to "Matemática" and contains the following functions: DESCUMPRIR MEDIDAS PROTETIVAS, DESOBEDIENCIA, DIFAMACAO, EXP, LN, LOG10. The "Variáveis" menu lists several variables: ABS, DROGAS PARA O CONSUMO PESSOAL, ESTELIONATO, ESTELIONATO FRAUDE ELETRONICA, and DIFAMACAO.

# Calculando variáveis

- 1) Uso da variável calculada no banco de dados 2
- 2) Somar as variáveis estelionato + estelionato e fraude eletrônica

as	Estelionat..
0	70
0	24
0	95
45	1372
0	18
0	86
0	166
0	73
0	71
0	29
0	164
0	25
0	174
0	19
0	185
0	38
1	1195
0	1638
0	187
0	6
0	117
21	2303
0	5
0	0

# Calculando variáveis

Exercício:

- 1) Uso da variável calculada no banco de dados 2
- 2) Somar as variáveis furto qualificado e furto simples para criar a variável furto total
- 3) Calcular a quantidade de furto total por habitante para os municípios do Paraná

# Transformando variáveis

Queremos saber se há ou não favelas nos municípios do Paraná

## Atenção:

1. Nomes nas fórmulas precisam vir dentro de “”
2. As transformações são usadas para variáveis quantitativas

The image shows a software interface for transforming variables. On the left, a context menu is open over a variable named "Qnt Favelas". The menu options are: Cortar (Cut), Copiar (Copy), Colar (Paste), Adicionar variável (Add variable), Apagar variável (Delete variable), Configurar... (Configure...), and Transformar... (Transform...). The "Transformar..." option is highlighted with a blue selection bar.

On the right, a dialog box titled "VARIÁVEL TRANSFORMADA" (Transformed Variable) is displayed. It has a title bar "Favelas". Below it is a "Descrição" (Description) field. Under "Variável fonte" (Source variable), there is a dropdown menu showing "Qnt Favelas" and a "Nenhum" button with an "Editar..." (Edit...) button next to it. A dropdown menu is open under "Nenhum", showing the options "Nenhum" and "Criar nova transformação..." (Create new transformation...), with "Criar nova transformação..." highlighted by a blue selection bar.

# Transformando variáveis

● TRANSFORMAR usado por 1

Transformar 1

Descrição Variável sufixo

+ Adicionar condição de recodificação

`fx se $source == 0 usar 'não'` `X`

`fx caso contrário usar 'sim'`

Tipo de medida Nominal 

Qnt Favelas	Oa Favelas	Estelionat...
0	não	70
0	não	24
0	não	95
45	sim	1372
0	não	18
0	não	86
0	não	166
0	não	73
0	não	71
0	não	29
0	não	164
0	não	25
0	não	174
0	não	19
0	não	185
0	não	38
1	sim	1195
0	não	1638
0	não	187
0	não	6
0	não	117
21	sim	2303
0	não	5

# Transformando variáveis

Exercício:

Transformar a variável quantidade de casos de perseguição no município para uma variável se teve ou não perseguição no município em 2024.

# População e Amostra

- Amostragem busca obter informações sobre o todo, baseando-se no resultado de uma parte.
- Usamos esta ideia nosso dia-a-dia:
  - ✓ Quando provamos o feijão para saber se o sal está no ponto
  - ✓ Quando fazemos exame de sangue

## Pesquisa censitária

- Entrevista todo o público alvo.
- Requer mais recursos financeiros e recursos humanos

X

## Pesquisa Amostral

- Entrevista uma parte do todo.
- O resultado se dá a partir de estimativas.
- Requer uma logística menor.
- É mais barata.

# População e Amostra

- **População ou público alvo:**

- ✓ É o conjunto de todos os elementos para os quais desejamos que as conclusões da pesquisa sejam válidas.
- ✓ É a agregação dos elementos a partir dos quais será extraída uma amostra.

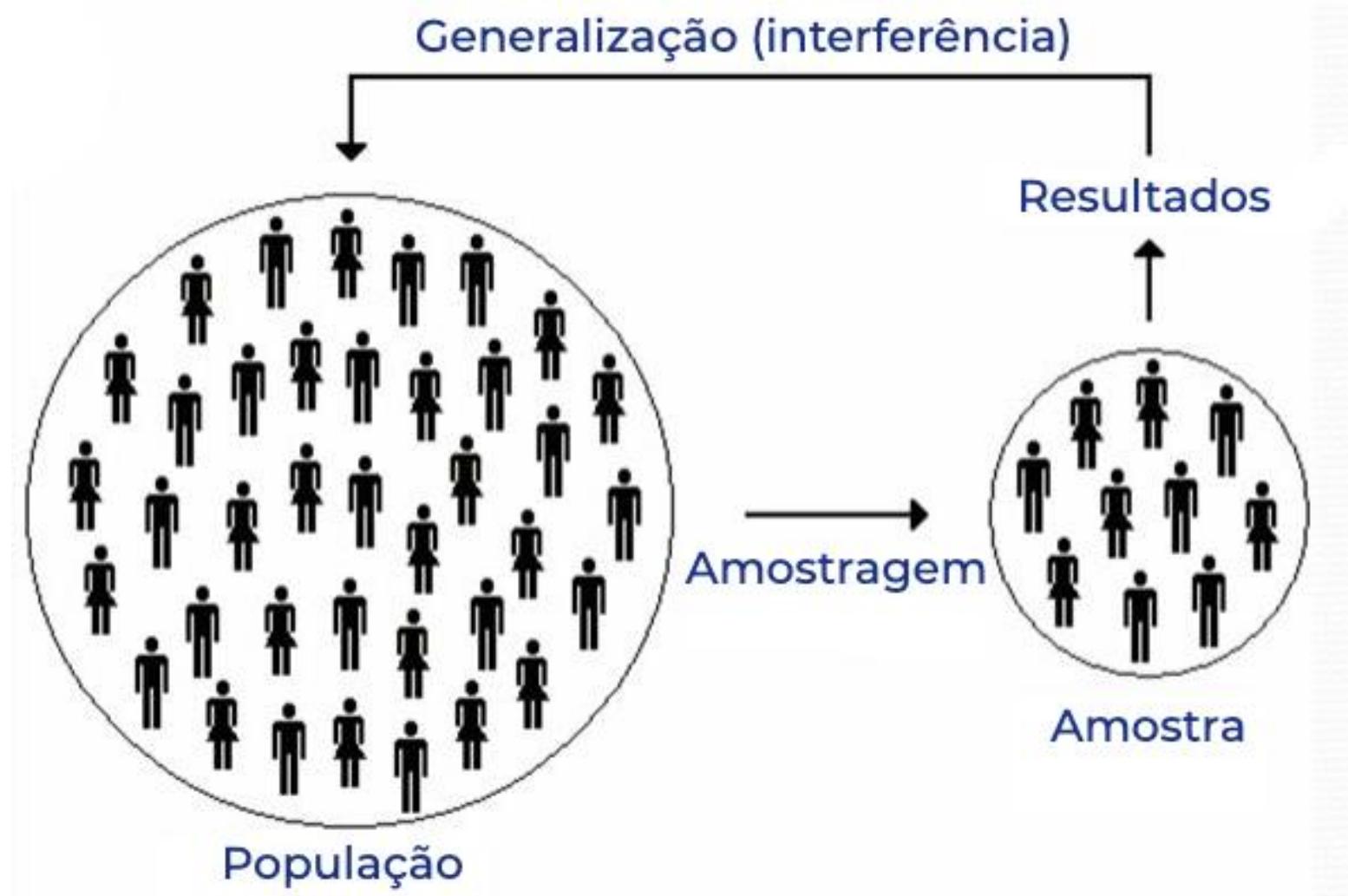
Exemplo: Se uma prefeitura está interessada em estimar características da população que utiliza determinado equipamento urbano público a partir de uma amostra, a população de referência é a totalidade de indivíduos que frequentam o equipamento urbano em certo período de tempo.



# População e Amostra

- Amostra: É um recorte de elementos extraídos da população.
- Unidade de Amostragem: É a unidade a ser selecionada para se chegar aos elementos da população.  
Ex: indivíduos, salas de aula, bairros.....
- População Alvo: População para qual se gostaria de obter a informação.
- População de Pesquisa: População a ser realmente coberta pela pesquisa (amostrada).

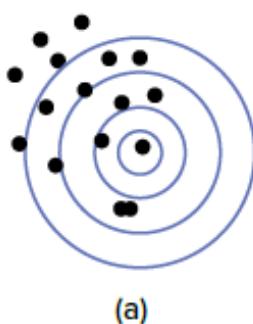
# População e amostra



# População e Amostra

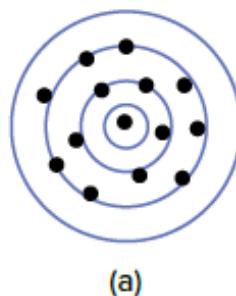
- **Viés de Seleção**

- ✓ Vício: Mede a proximidade de cada observação do valor alvo que se procura atingir.
- ✓ Precisão: Mede a proximidade de cada observação da média de todas as observações.
- ✓ A distribuição amostral precisa ter alta precisão e alta exatidão (não viciado)



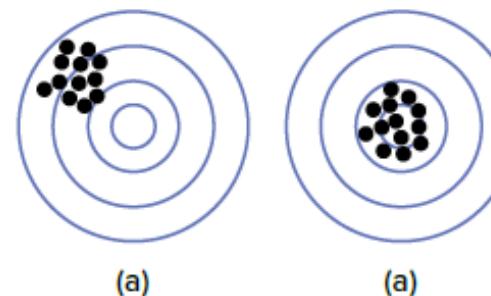
(a)

Baixa precisão  
Baixa exatidão



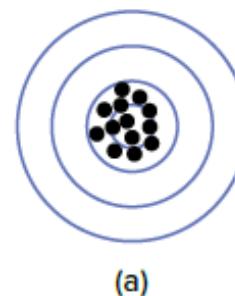
(a)

Baixa precisão  
Alta exatidão



(a)

Alta precisão  
Baixa exatidão



(a)

Alta precisão  
Alta exatidão

Representação da precisão e exatidão em medidas experimentais (Fonte: Adaptado de Silva Neto, 2012)

# População e Amostra

- **Margem de erro:**

✓ A possível variação das respostas da população com relação à amostra.

Ex: margem de erro de 5%. As estimativas obtidas podem ser 5% para mais ou para menos

- **Intervalo de confiança:**

✓ A probabilidade da amostra representar com precisão a população em questão.

✓ Representa a probabilidade de uma pesquisa ter os mesmos resultados se for aplicada com um outro grupo de pessoas, dentro do mesmo perfil de amostra e com a mesma margem de erro.

Ex: Intervalo de confiança de 95%. Se a pesquisa for repetida 100 vezes a chance de obter o mesmo resultado é de 95 vezes.

# População e Amostra

## Notação matemática:

- Para um intervalo com nível de confiança  $(1 - \alpha)\%$ :

Coeficiente de Confiança =  $1 - \alpha$

$\alpha$  é chamado de nível de significância, ou margem de erro.

Ex: 95% de confiança  $\Rightarrow 1 - \alpha = 0.95 \Rightarrow \alpha = 0.05$



# Tipos de amostra

- **Amostra probabilística:**

- ✓ Quando temos o cadastros de toda a população.
- ✓ É realizado um sorteio aleatório da população da amostra.

- **Amostra não probabilística:**

- ✓ Temos informações sobre o público alvo, mas não o cadastro.
- ✓ Não é possível realizar um sorteio aleatório da amostra.
- ✓ Não temos informações sobre o público alvo.

# Tipos de amostra

- **Aleatória simples:**

- ✓ É aquela em que cada elemento da população tem a mesma chance de ser selecionado para a amostra.
- ✓ Cada elemento é sorteado aleatoriamente sem reposição.
- ✓ A ordem dos elementos é considerada.

- **Sistemática**

- ✓ O pesquisador decide escolher os casos amostrados de acordo com um dado intervalo.

Exemplo:

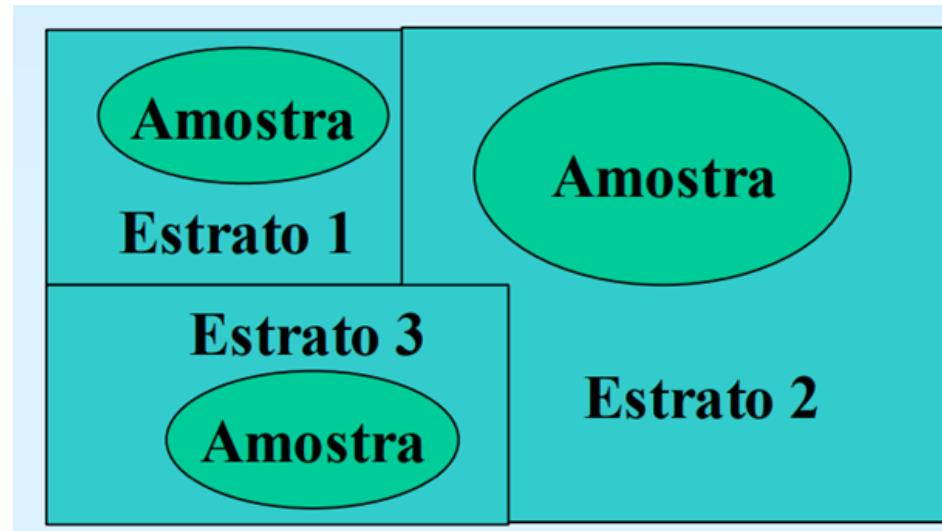
Contamos com uma população de 100 elementos e queremos uma amostra com 30% do total:

Selecionamos um de cada três elementos em ordem. O 3º elemento, depois o 6º, depois o 9º, e assim por diante.

- Virtualmente idênticos aos resultados obtidos através de procedimentos de amostragem aleatória simples.

# Tipos de amostra

- **Estratificada**
  - Usada quando há subconjuntos homogêneos da população, mas heterogêneos se comparados uns aos outros em termos dos interesses da pesquisa.
  - Os subconjuntos devem ser mutuamente exclusivos ( Ex: homem, mulher).
  - A amostra aleatória é usada em cada estrato.
  - São feitas estimativas para cada um dos estratos.



# Tipos de amostra

- **Conglomerado:**

- ✓ Os elementos da população são reunidos em grupos e, por sua vez, alguns destes são sorteados para compor a amostra.
- ✓ Não necessita do cadastro de todo o público alvo.

Ex: Público alvo de um estudo são os alunos da rede pública do município Z. Só temos o cadastro das escolas e das turmas. Sorteamos as escolas e posteriormente as turmas. Todos os alunos das turmas sorteadas serão entrevistados.

Nesse caso há 2 estágios na amostra por conglomerados.

- As formas de amostragem podem ser combinadas!

# Tipos de amostra

- **Amostra não probabilística por cotas:**
  - ✓ Usada quando há informações do público alvo ,mas não há um cadastro.
  - ✓ Assemelha-se com a amostragem estratificada.
  - ✓ A população é vista de forma segregada , dividida em diversos subgrupos.
  - ✓ Seleciona-se uma cota de cada subgrupo, proporcional ao seu tamanho.
  - ✓ Para compensar a falta de aleatoriedade na seleção, costuma-se dividir a população em um grande número de subgrupos.

# Tipos de amostra

- **Amostragem não probabilística intencional ou por julgamento:**
  - ✓ Selecionam-se os casos a partir de conhecimentos prévios a respeito dos elementos da população.
  - ✓ Nesses casos o pesquisador julga que os elementos selecionados possuem características típicas representativas da população que lhe interessa, podendo essa população de interesse ser mais ou menos coincidente com população geral ou absolutamente desviante
  - ✓ Não é possível fazer inferência.

# Tipos de amostra

- **Bola de neve:**

- ✓ Não temos o cadastro nem informações precisas sobre o público alvo.
- ✓ O pesquisador deve ao menos ter um conjunto inicial de indivíduos a serem pesquisados.
- ✓ A partir daí, pede-se aos primeiros entrevistados que indiquem novos nomes para compor a amostra.
- ✓ A amostra vai assim crescendo como uma bola de neve à medida que os sujeitos da investigação vão indicando novos nomes ao investigador.
- ✓ Utilizadas em estudos exploratórios
- ✓ Não é possível fazer inferência.

# Indicadores Sociais

- ✓ Um indicador social, pode ser entendido como uma definição operacional ou parte de uma definição operacional de qualquer um dos conceitos centrais à geração de um sistema de informação descritivo de um sistema social” (COSTA, [1975), 2017).
- ✓ Indicador social é uma medida, em geral, quantitativa, dotada de significado social substantivo, usado para substituir, quantificar ou operacionalizar um conceito social abstrato (JANNUZZI, 2001).
- ✓ Pode ser considerado o elo de ligação entre a Teoria social e a Evidência empírica dos fenômenos sociais observados (JANNUZZI, 2001).
- ✓ “É um recurso metodológico, empiricamente referido, que informa algo sobre um aspecto da realidade social ou sobre mudanças que estão se processando na mesma”.

# Fontes de dados para construção de Indicadores

- Pesquisas censitárias e amostrais
  - Registros administrativos.
- 
- Repositórios que compilam as informações
  - Site dos órgãos produtores de informações

# Censo Demográfico

- Os censos demográficos são pesquisas estatísticas cujo levantamento consiste na visita a todos os domicílios de um país e constituem a única fonte de referência para o conhecimento das condições de vida da população em todos os municípios e em seus recortes territoriais internos distritos, bairros e localidades, rurais ou urbanos.
- O Censo Demográfico permite, tal como objetivado pelo IBGE, permite conhecer profundamente a população residente no Brasil, subsidiando a implementação e a aprimoramento de políticas públicas.

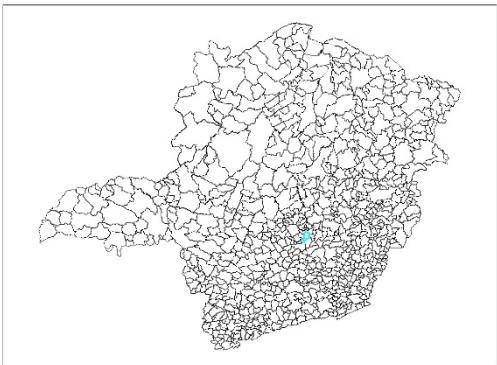


# Censo Demográfico

Estados



Municípios



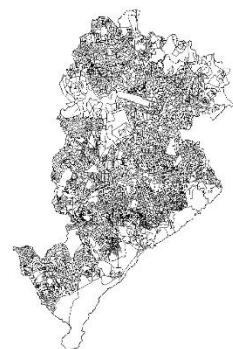
Áreas de  
ponderação\*



Bairros



Setores  
censitários\*\*



# PNAD Contínua

- Visa acompanhar as flutuações trimestrais e a evolução, no curto, médio e longo prazos, da força de trabalho, e outras informações necessárias para o estudo do desenvolvimento socioeconômico do País.
- A PNAD Contínua foi implantada, experimentalmente, em outubro de 2011 e, a partir de janeiro de 2012, em caráter definitivo, em todo o Território Nacional.
- É uma pesquisa amostral.
- **Área de abrangência:** Brasil, Grandes Regiões, Unidades da Federação, Regiões Metropolitanas que contêm Municípios das Capitais, Região Integrada de Desenvolvimento - RIDE Grande Teresina, e Municípios das Capitais.
- **Unidade de análise:** Morador e o domicílio.

# PNAD Contínua

- **Periodicidade:**

- ✓ Mensal - Conjunto restrito de indicadores relacionados à força de trabalho e somente para o nível geográfico de Brasil;
- ✓ Trimestral - Conjunto de indicadores relacionados à força de trabalho para todos os níveis de divulgação da pesquisa;
- ✓ Anual - Demais temas permanentes da pesquisa e indicadores complementares à força de trabalho; e
- ✓ Variável - Outros temas ou tópicos dos temas permanentes a serem pesquisados com maior periodicidade ou ocasionalmente.

# PNAD Contínua

- **Suplemento violência:**
- 2021: Vitimização, com perguntas relacionadas a roubo, furto e sensação de segurança;
- 2009: Características da vitimização e o acesso à justiça no Brasil.

# PNAD Contínua

- **Vantagens:**
  - ✓ Abrangência de temas: renda, trabalho, educação, habitação, etc.
  - ✓ Mais barata por ser uma pesquisa amostral.
  - ✓ Possibilidade de gerenciamento dos dados, exemplo: criação de novas variáveis
- **Limitações:**
  - ✓ É uma entrevista. Portanto, o entrevistado não precisa dizer a verdade.
  - ✓ Problemas de subdeclaração da renda, principalmente as mais elevadas.
  - ✓ Não considera o valor monetário da produção para próprio consumo e próprio uso.

# Sidra - IBGE

- **SIDRA - Sistema IBGE de Recuperação Automática**

<https://sidra.ibge.gov.br/home/pmc/brasil>

- ✓ Compila em tabelas informações de diversas pesquisas feitas pelo IBGE
- ✓ Censo
- ✓ PNAD
- ✓ Contas nacionais e PIB
- ✓ Pesquisa de Orçamento Familiar – POF
- ✓ Pesquisa Nacional de Saneamento Básico - PNSB

# Sidra - IBGE

The screenshot shows the SIDRA website interface. At the top, there's a navigation bar with links for PESQUISAS, ACERVO, TERRITÓRIO, CONTATO, AJUDA, and a user profile icon. A search bar labeled 'Filtro' is also present. Below the navigation, there are two main sections: 'PPM' (Pesquisa da Pecuária Municipal) and 'Pesquisa Nacional por Amostra de Domicílios Contínua - Acesso à televisão e posse de telefones para uso pessoal 2021'. A sidebar on the left lists various survey categories like Indicadores, População, Economia, and Meio Ambiente. The main content area displays a table titled 'Características Gerais' with five rows of data. At the bottom, there's another section titled 'Deficiência:'.

Número	Nome	Período	Território
200	População residente, por sexo, situação e grupos de idade - Amostra - Características Gerais da População	1970, 1980, 1991, 2000, 2010	BR, GR, UF, ME, MI, MU, RM, RS, RD, AU
2093	População residente por cor ou raça, sexo, situação do domicílio e grupos de idade - Amostra - Características Gerais da População	2000, 2010	BR, GR, UF, ME, MI, MU, RM, RS, RD, AU
3460	Pessoas de 15 anos ou mais de idade, por grupos de idade e sexo, segundo o nível de instrução e a cor ou raça - Características Gerais da População	2010	BR, GR, UF, ME, MI, MU, RM, RS, RD, AU
3461	Pessoas de 10 anos ou mais de idade, ocupadas na semana de referência, por cor ou raça, segundo o sexo, a posição na ocupação e a categoria do emprego no trabalho principal - Características Gerais da População	2010	BR, GR, UF, ME, MI, MU, RM, RS, RD, AU
3462	Pessoas residentes em domicílios particulares, exclusive as cuja condição no domicílio era pensionista, empregado(a) doméstico(a) ou parente do(a) empregado(a) doméstico(a), por cor ou raça, segundo o sexo e as classes de rendimento nominal mensal domiciliar per capita - Características Gerais da População	2010	BR, GR, UF, ME, MI, MU, RM, RS, RD, AU

# IBGE Cidades

IBGE Cidades: <https://cidades.ibge.gov.br/>

Compilado de informações sobre o Brasil, UFs e municípios brasileiros.



# DataSUS

- Os Sistemas de Informação em Saúde (SIS) são ferramentas para auxiliar os gestores a processarem, adquirirem e analisarem os dados.
- Os SIS fornecem indicadores em três grandes áreas:
  1. determinantes de saúde;
  2. indicadores sobre o sistema e serviço de saúde; e
  3. indicadores de situação de saúde

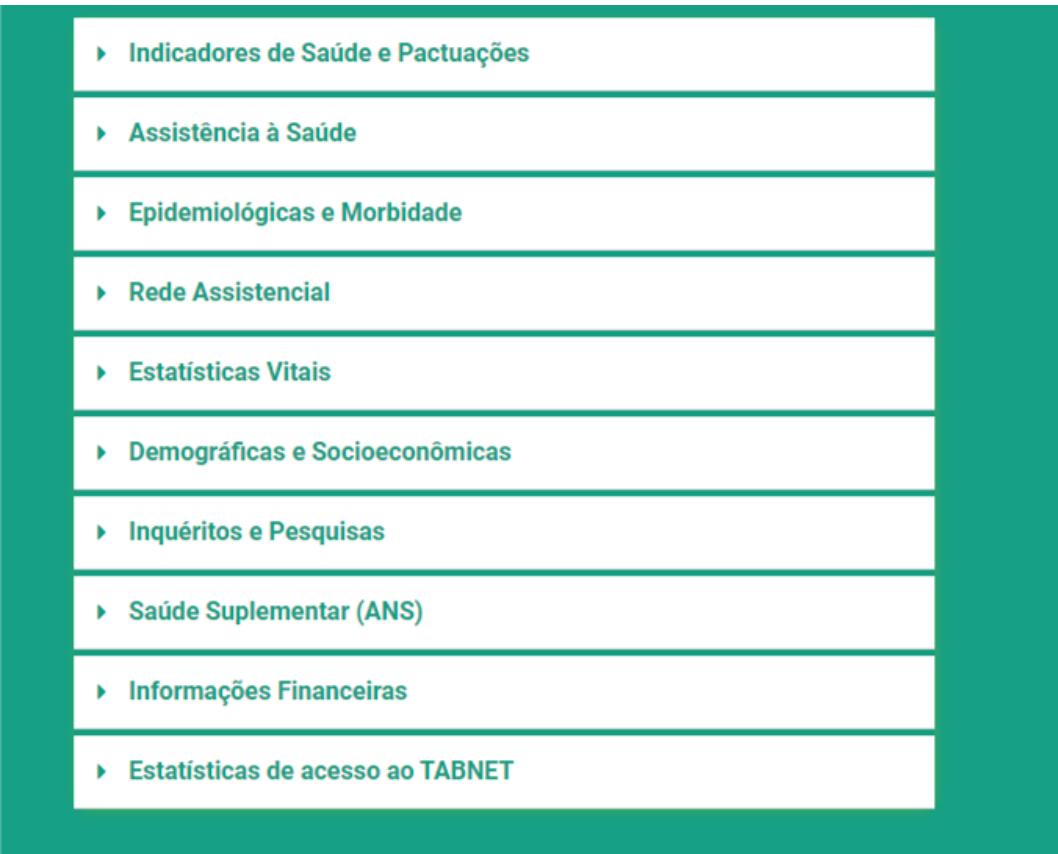
# DataSUS

- Para análises na área de segurança pública:
- Morbidade são muito importantes, pois conseguem mensurar, por exemplo, mortes violentas, mortes por acidentes e autoextermínio.
- Sistema de Informação de Agravos de Notificação (Sinan):
  - Violência interpessoal e a violência autoprovocada (2009)
  - Violência doméstica e sexual (2011)

# DataSUS

- Tabnet:

<https://datasus.saude.gov.br/informacoes-de-saude-tabnet/>



# DataSUS

## Indicadores de Saúde e Pactuações

## Assistência à Saúde

## Epidemiológicas e Morbidade

## Rede Assistencial

## Estatísticas Vitais

- Nascidos Vivos – desde 1994
- Mortalidade – desde 1996 pela CID-10
- Dados Preliminares de 2021
- Painéis de monitoramento (SVS)
- Correção e redistribuição de óbitos segundo a Pesquisa de Busca Ativa
- Mortalidade – 1979 a 1995, pela CID-9
- Câncer (sítio do Inca)

## Demográficas e Socioeconômicas

## Inquéritos e Pesquisas

## NASCIDOS VIVOS - MATO GROSSO

### Linha

Município  
Região de Saúde (CIR)  
Região de Saúde/Município  
Macrorregião de Saúde  
Divisão administrativa estadual

### Coluna

Não ativa  
Região de Saúde (CIR)  
Macrorregião de Saúde  
Divisão administrativa estadual

### Conteúdo

Nascim p/resid.mãe  
Nascim p/ocorrênc

## PERÍODOS DISPONÍVEIS

1999
1998
1997
1996
1995
1994

# DataSUS

## ► Epidemiológicas e Morbidade

- Morbidade Hospitalar do SUS (SIH/SUS)
- Casos de Aids – Desde 1980 (SINAN)
- Casos de Hanseníase – Desde 2001 (SINAN)
- Casos de Tuberculose – Desde 2001 (SINAN)
- Doenças e Agravos de Notificação – 2007 em diante (SINAN)
- Doenças e Agravos de Notificação – 2001 a 2006 (SINAN)
- Notificações de casos suspeitos de SCZ – desde 2015
- Programa de Controle da Esquistossomose (PCE)
- Estado Nutricional (SISVAN)
- Hipertensão e Diabetes (HIPERDIA)
- Câncer de colo de útero e de mama (SISCOLO/SISMAMA)
- Sistema de Informação do Câncer – SISCAN (colo do útero e mama)
- Tempo até o inicio do tratamento oncológico – PAINEL – oncologia

## Doenças e Agravos de Notificação - 2007 em diante (SINAN)

- Opção selecionada: Doenças e Agravos de Notificação – 2007 em diante (SINAN)
- Acidente de Trabalho
  - Acidente por Animais Peçonhentos
  - Acidente de Trabalho com Exposição a Material Biológico
  - Botulismo
  - Câncer Relacionado ao Trabalho
  - Cólera
  - Coqueluche
  - Dermatose Relacionada ao Trabalho
  - Toxoplasmose Gestacional
  - Transtorno Mental Relacionado ao Trabalho
  - Varicela
  - Violência Interpessoal / Autoprovocada
  - Zika Vírus

# Instituto Paranaense de Desenvolvimento Econômico e Social (IPARDES)

- Pesquisas voltadas para o entendimento da sociedade, da economia e do meio ambiente.
- No campo da **segurança pública**, os dados gerados nos dois primeiros eixos fornecem **informações contextuais sobre a população**, contribuindo para a compreensão das dinâmicas criminais.

# Instituto Paranaense de Desenvolvimento Econômico e Social (IPARDES)

- Caderno Estatístico Municipal, produzido pela instituição, reúne, em um único lugar, dados sobre os municípios paranaenses obtidos a partir de diversas fontes oficiais.

- Para acessar os cadernos municipais:

<https://www.ipardes.pr.gov.br/Pagina/Caderno-Estatistico-Municipal>

- Para acessar o perfil dos municípios ( informação sobre um município):

<https://www.ipardes.pr.gov.br/Pagina/Perfil-dos-municipios>

# Instituto Paranaense de Desenvolvimento Econômico e Social (IPARDES)

## Caderno Estatístico Municipal

Seleção/Capa | Caracterização do Território | Agropecuária 1 | Agropecuária 2 | Comunicações | Cultura | **Demografia** (red circle) | Domicílios | Educação >

< >

Selecionar o município ou estado:

Estado do Paraná

- Abatiá
- Adrianópolis
- Agudos do Sul
- Almirante Tamandaré
- Altamira do Paraná
- Alto Paraíso
- Alto Paraná
- Alto Piquiri
- Altônia
- Alvorada do Sul
- Amaporã
- Ampére
- Anahy
- Andirá
- Ângulo
- Antonina
- Antônio Olinto
- Apucarana
- Arapongas

IPARDES INSTITUTO PARANAENSE DE  
DESENVOLVIMENTO ECONÔMICO E SOCIAL

Estado do Paraná

DEMOSGRAFIA

POPULAÇÃO ESTIMADA

INFORMAÇÃO	2018	2019	2020	2021	2024
População Estimada - Total	11.348.937	11.433.957	11.516.840	11.597.484	11.824.665

FONTE: IBGE

POPULAÇÃO CENSITÁRIA, SEGUNDO FAIXAS ETÁRIAS

FAIXAS ETÁRIAS	2000	2010	2022
Menores de 1 ano	172.991	144.631	130.474
De 1 a 4 anos	713.282	569.406	567.659
De 5 a 9 anos	924.411	768.392	759.916
De 10 a 14 anos	936.446	909.071	737.248
De 15 a 19 anos	949.680	928.631	761.026
De 20 a 24 anos	871.046	901.332	846.638
De 25 a 29 anos	795.601	880.232	889.031
De 30 a 34 anos	779.980	836.377	874.321
De 35 a 39 anos	731.167	787.370	873.773
De 40 a 44 anos	619.958	766.648	873.530
De 45 a 49 anos	517.022	700.184	779.549
De 50 a 54 anos	415.988	592.618	756.703

# Instituto Paranaense de Desenvolvimento Econômico e Social (IPARDES)

## Perfil dos Municípios

**IPARDES**  
INSTITUTO PARANAENSE DE  
DESENVOLVIMENTO ECONÔMICO E SOCIAL

**PERFIL DO MUN**

TERRITÓRIO E AUTORIDADE ELEITA				
Desmembrado de				
Data de Instalação				
Data de Comemoração (Aniversário)				
Altitude da Sede (IBGE) (m)				
Distância à Capital (SEIL) (km)				
Autoridade Eleita (TRE)	Eduardo Pimentel Slaviero			
ÁREA TERRITORIAL E DEMOGRÁFICA	FONTE	DATA	MUNICÍPIO	ESTADO
Área Territorial (km <sup>2</sup> )	IAT	2024	:	:
Densidade Demográfica (habitantes/km <sup>2</sup> )	IPARDES	2024	:	:
Grau de Urbanização (%)	IBGE	2022	:	:
População Estimada	IBGE	2024	:	:
População Censitária	IBGE	2022	:	:
População Censitária Urbana	IBGE	2022	:	10.179.847
População Censitária Rural	IBGE	2022	:	1.264.533
População - Contagem (1)	IBGE	2007	:	10.284.503
Taxa de Crescimento Geométrico Populacional (%)	IBGE	2022	:	0,76

Selezione o Município:

Curitiba

Pesquisar

- Abatiá
- Adrianópolis
- Agudos do Sul
- Almirante Tamandaré
- Altamira do Paraná
- Alto Paraíso
- Alto Paranaíba



Site IPARDES->  
Estatísticas -> perfil  
dos municípios

# Exercícios

- 1) Buscar no Sidra a população do Paraná (UF) e a população de Curitiba nos dados do Censo Demográfico 2022.
- 2) Buscar a quantidade de domicílios que tiveram vítimas de furtos de veículos no Brasil em 2021.
- 3) Buscar a quantidade de mortes no Paraná em 2022.
- 3) Buscar no Ipardes a população do Paraná (UF) e a população de mais um município de livre escolha.

# Exercícios

1) Buscar no Sidra a população do Paraná (UF) e a população de Curitiba nos dados do Censo Demográfico 2022.

The screenshot shows the Sidra System interface with the following details:

- Filtro:** A search bar at the top left.
- Indicadores:** A section on the left under "População" containing:
  - Censo Demográfico - Censo-Demográfico (highlighted)
  - Estatísticas do Registro Civil - Registro-Civil
  - Pesquisa Nacional por Amostra de Domicílios Contínua Anual - PNADCA
  - Pesquisa Nacional por Amostra de Domicílios - PNAD
  - Estimativas de População - EstimaPop
  - Projeções da População
  - Pesquisa de Orçamentos Familiares - POF
  - Pesquisa Nacional de Saúde - PNS
  - Pesquisa Assistência Médico-Sanitária - AMS
  - Pesquisa Nacional de Saneamento Básico - PNSB
- Demográfico 2022:** A dropdown menu at the top center.
- Demográfico 2010, Demográfico 2000, Contagem 2007, Contagem:** Other dropdown menus at the top center.
- Geral:** A section under "Resultados do Universo" containing:
  - Início
  - Primeiros Resultados do Universo
  - População e Domicílios (highlighted)
  - Quilombolas
  - Indígenas
  - Resultados preliminares da Amostra
- Resultados do Universo:** A section containing:
  - População por Idade e Sexo
  - População por Cor ou Raça
  - Características dos Domicílios
  - Quilombolas e Indígenas por Sexo e Idade
  - Quilombolas: Alfabetização e Características dos Domicílios
  - População e Domicílios: Situação Urbana ou Rural
- Outras opções:** A vertical column on the right containing:
  - Alfabetização
  - Registros de Nascimentos
  - Tipos de Domicílios
    - Indígenas: Alfabetização, Registros de Nascimentos e Características dos Domicílios
  - Composição Domiciliar e Óbitos Informados
  - Favelas e Comunidades Urbanas
  - Indígenas: Principais

# Exercícios

1) Buscar no Sidra a população do Paraná (UF) e a população de Curitiba nos dados do Censo Demográfico 2022.

População				Período	Território
	Número	Nome			
	4709	População residente, Variação absoluta de população residente e Taxa de crescimento geométrico		2022	BR, GR, UF, IN, IM, CU, MU
	4714	População Residente, Área territorial e Densidade demográfica		2022	BR, GR, UF, IN, IM, CU, MU

Níveis territoriais

- »  Brasil [0/1]
- »  Grande Região [0/5]
- »  Unidade da Federação [1/27]
- »  Região Geográfica Intermediária [0/133]
- »  Região Geográfica Imediata [0/510]
- »  Concentração Urbana [0/185]
- »  Município [0/5570]

Início / Unidade da Federação

- 32. Espírito Santo
- 33. Rio de Janeiro
- 35. São Paulo
- 41. Paraná
- 42. Santa Catarina
- 43. Rio Grande do Sul
- 50. Mato Grosso do Sul
- 51. Mato Grosso
- 52. Goiás
- 53. Distrito Federal

# Exercícios

1) Buscar no Sidra a população do Paraná (UF) e a população de Curitiba nos dados do Censo Demográfico 2022.

The screenshot shows the Sidra System interface. On the left, there's a sidebar for 'Níveis territoriais' with options like Brasil, Grande Região, Unidade da Federação, etc., and 'Município [5569/5570]' which is selected. The main search bar has 'curiti' entered. Below it, under 'Inicio / Município', there are two entries: '4106902. Curitiba (PR)' with a checked checkbox, and '4204806. Curitibanos (SC)'. At the top right, there are buttons for 'Layout', 'Ordenar', and 'Funções'. To the right, a table titled 'Tabela 4714 - População Residente, Área territorial e Densidade demográfica' is displayed. It has three sections: 'Variável - População residente (Pessoas)', 'Ano - 2022', and 'Unidade da Federação e Município'. The data rows show 'Paraná' with a population of 11.444.380 and 'Curitiba (PR)' with a population of 1.773.718. A note at the bottom says 'Fonte: IBGE - Censo Demográfico'.

Variável - População residente (Pessoas)	
Ano - 2022	
Unidade da Federação e Município	
Paraná	11.444.380
Curitiba (PR)	1.773.718

Fonte: IBGE - Censo Demográfico

# Exercícios

2) Buscar a quantidade de domicílios que tiveram vítimas de furtos de veículos no Brasil em 2021.

**PESQUISAS** ▾ **ACERVO** **TERRITÓRIO** **CONTATO** **AJUDA**

Filtro

Indicadores

População

Censo Demográfico - Censo-Demográfico

Estatísticas do Registro Civil - Registro-Civil

Pesquisa Nacional por Amostra de Domicílios Contínua Anual - PNADCA

Pesquisa Nacional por Amostra de Domicílios - PNAD

Estimativas de População - EstimaPop

Projeções da População

Pesquisa de Orçamentos Familiares - POF

Pesquisa Nacional de Saúde - PNS

Pesquisa Assistência Médico-Sanitária - AMS

Pesquisa Nacional de Saneamento Básico - PNSB

## PNADC/A - Tabelas com os indicadores

- Características gerais dos moradores
- Características gerais dos domicílios
- Outras formas de trabalho - Afazeres domésticos
- Outras formas de trabalho - Afazeres domésticos e Cuidados de pessoas
- Outras formas de trabalho - Cuidados de pessoas
- Outras formas de trabalho - Próprio consumo
- Outras formas de trabalho - Trabalho voluntário
- Outras formas de trabalho - Outras formas de trabalho - total
- Educação
- Turismo
- Características adicionais do mercado de trabalho
- Tecnologia de informação e comunicação
- Rendimento de todas as fontes
- Sensação de segurança
- Furtos e roubos

# Exercícios

2) Buscar a quantidade de domicílios que tiveram vítimas de furtos de veículos no Brasil em 2021.

	8510	Domicílios em que pelo menos um morador foi vítima de furto nos últimos 12 meses, por tipo de furto	2021	BR, GR, UF, RM, MU, RD
--	------	---	------	------------------------

Domicílios em que pelo menos um morador foi vítima de furto nos últimos 12 meses (Mil unidades)  
 Coeficiente de variação - Domicílios em que pelo menos um morador foi vítima de furto nos últimos 12 meses (%): <1 de 1> casas decimais  
 Percentual de domicílios em que pelo menos um morador foi vítima de furto nos últimos 12 meses (%): <1 de 1> casas decimais  
 Coeficiente de variação - Percentual de domicílios em que pelo menos um morador foi vítima de furto nos últimos 12 meses (%): <1 de 1> casas decimais

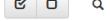
Tipo de furto [1/6]



Notas

1. Totalização de categorias não disponível para a variável "Percentual de domicílios em que pelo menos um morador foi vítima de furto nos últimos 12 meses"  
2. Totalização de categorias não disponível para a variável "Coeficiente de variação - Percentual de domicílios em que pelo menos um morador foi vítima de furto nos últimos 12 meses"

Ano [1/1]



Unidade Territorial [2/81]

Níveis territoriais

>  Brasil [1/1]  
>  Grande Região [0/5]  
>  Unidade da Federação [1/27]  
>  Região Metropolitana [0/20]  
>  Município [0/27]

Inicio



Ativar o W  
Acesse Config

**Tabela 8510 - Domicílios em que pelo menos um morador foi vítima de furto nos últimos 12 meses, por tipo de furto**

**Variável - Domicílios em que pelo menos um morador foi vítima de furto nos últimos 12 meses (Mil unidades)**

**Ano - 2021**

**Tipo de furto - Carro**

**Brasil e Unidade da Federação**

**Brasil**

**184**

**Paraná**

**-**

**Fonte: IBGE - Pesquisa Nacional por Amostra de Domicílios Contínua Anual - 4º trimestre**

# Exercícios

## 3) Buscar a quantidade de mortes no Paraná em 2022.

Opção selecionada: Mortalidade – desde 1996 pela CID-10

- Mortalidade geral
- Mortalidade de residentes no exterior
- Óbitos por causas evitáveis - 0 a 4 anos
- Óbitos por causas evitáveis - 5 a 74 anos
- Óbitos infantis
- Óbitos de mulheres em idade fértil e óbitos maternos
- Óbitos por causas externas
- Óbitos fetais
- Óbitos por causa múltipla
- Óbitos por causas mal definidas e inespecíficas
- Óbitos por causas pouco úteis
- Completude dos campos de dados da mãe em óbitos fetais e menores de 1 ano
- Completude dos campos de dados de óbitos de mulheres em idade fértil
- Completude dos campos de dados de óbitos por causas externas

Abrangência Geográfica:

Paraná

# Exercícios

3) Buscar a quantidade de mortes no Paraná em 2022.

## ► MORTALIDADE - PARANÁ

Linha	Coluna	Conteúdo
Município Região de Saúde (CIR) Região de Saúde/Município Macrorregião de Saúde	Não ativa Região de Saúde (CIR) Macrorregião de Saúde Divisão administrativa estadual	Óbitos p/Residência Óbitos p/Ocorrência

## ► PERÍODOS DISPONÍVEIS

- 2023
- 2022
- 2021
- 2020
- 2019
- 2018

## ► SELEÇÕES DISPONÍVEIS

- [+] Município
- [+] Região de Saúde (CIR)
- [+] Macrorregião de Saúde
- [+] Divisão administrativa estadual
- [+] Microrregião IBGE
- [+] Região Metropolitana - RIDE
- [+] Capítulo CID-10
- [+] Causa CID-10

# Exercícios

## 3) Buscar a quantidade de mortes no Paraná em 2022.

### ► MORTALIDADE - PARANÁ

Óbitos p/Residênc segundo Município  
Período: 2022

Município	Óbitos p/Residênc
TOTAL	89.808
MUNICIPIO IGNORADO - PR	124
410010 ABATIA	76
410020 ADRIANOPOLIS	69
410030 AGUDOS DO SUL	107
410040 ALMIRANTE TAMANDARE	779
410045 ALTAMIRA DO PARANA	32
410050 ALTONIA	200
410060 ALTO PARANA	117
410070 ALTO PIQUIRI	99
410080 ALVORADA DO SUL	115
410090 AMAPORA	40
410100 AMPERE	151
410105 ANAHY	29
410110 ANDIRA	222
410115 ANGULO	31
410120 ANTONINA	192
410130 ANTONIO OLINTO	73
410140 APUCARANA	1.144
410150 ARAPONGAS	976
410160 ARAPOTI	224
410165 ARAPUA	35
410170 ARARUNA	125
410180 ARAUCARIA	917
410185 ARIRANHA DO IVAI	29
410190 ASSAI	168

Ativar o Windows

Acesse Configurações para ativar o Windows.

29

168

# Análise descritiva

- **Frequência absoluta e Frequência relativa**
- ✓ **Frequência absoluta** é o número de vezes que um número aparece dentro de um conjunto de dados.
- ✓ **Frequência relativa ou percentual de respostas** é a quantidade de vezes que essa resposta aparece em relação ao total de respostas.
- ✓ **Frequência acumulada** é a soma das frequências de cada linha da tabela. Nesse caso, a última linha deve conter o valor total do conjunto de casos.

# Análise descritiva

Tabela 1 - Pessoas de 15 anos ou mais de idade que foram vítimas de roubo nos últimos 12 meses (por mil pessoas) nas capitais da região Sul

Estado	Frequência absoluta	Frequência relativa	Frequência acumulada
Curitiba (PR)	15	36%	36%
Florianópolis (SC)	0	0%	36%
Porto Alegre (RS)	27	64%	100%
<b>Total</b>	<b>42</b>	<b>100%</b>	

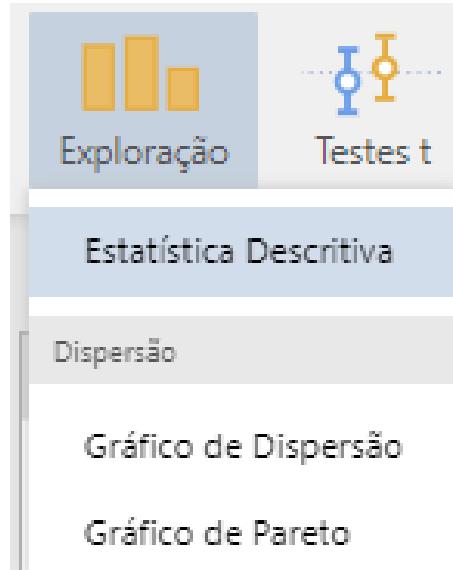
Fonte: IBGE – Pesquisa Nacional por Amostra de Domicílios Contínua Anual – 4º trimestre.

# Análise descritiva

- Exercício 1 :

No banco de dados município fazer a tabela de frequência para a variável construída se tem ou não favela.

# Análise descritiva



The dialog box displays a list of variables and settings:

- Variables list:
  - Parda
  - Sem instrução e fundamental incom.
  - Superior completo
  - Via pavimentada - Existe
  - Existência de iluminação pública - ...
  - Existência de ponto de ônibus / van
  - BD
  - Qnt Favelas
  - Estacionamento total (highlighted in blue)
- Buttons:
  - (move selected variable to the right)
  - Separar por (separate by) → (move selected variable to the right)
- Output settings:
  - Estatística Descritiva
  - Variáveis em colunas (Variables in columns)
  - Tabelas de Frequências (Tables of Frequencies)

# Análise descritiva

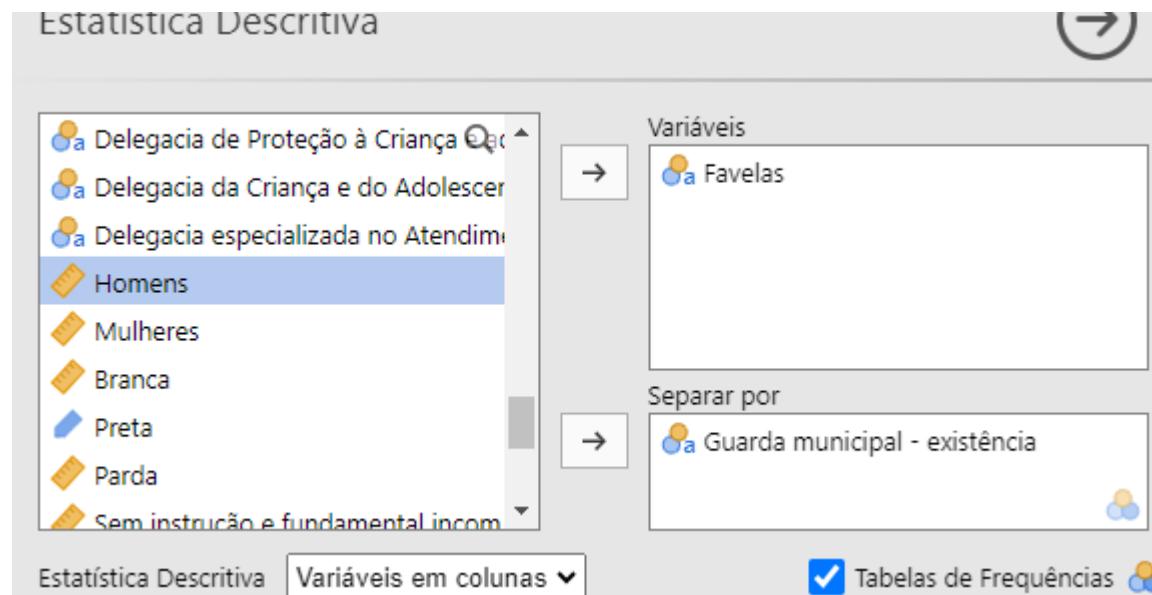
## Frequências

Frequências de Favelas

Favelas	Contadores	% do Total	% acumulada
não	358	89.7 %	89.7 %
sim	41	10.3 %	100.0 %

# Análise descritiva

- Tabela cruzada: analisar as informações de duas ou mais variáveis
- Exemplo: Ter ou não favelas com ter ou não guarda civil no município



# Análise descritiva

- Tabela cruzada: analisar as informações de duas ou mais variáveis  
Exemplo: Ter ou não favelas com ter ou não guarda civil no município

## Frequências

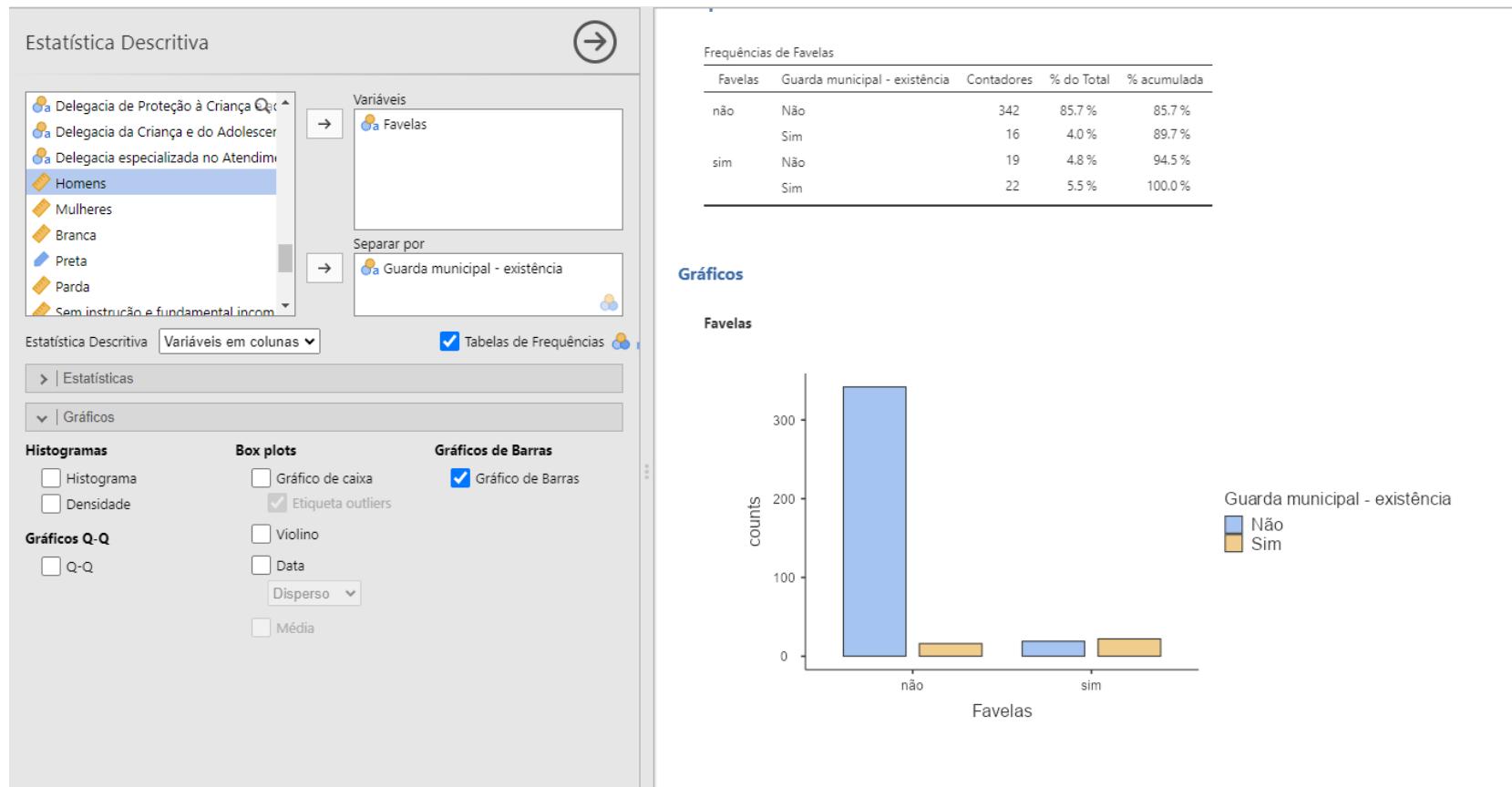
Frequências de Favelas

Favelas	Guarda municipal - existência	Contadores	% do Total	% acumulada
não	Não	342	85.7 %	85.7 %
	Sim	16	4.0 %	89.7 %
sim	Não	19	4.8 %	94.5 %
	Sim	22	5.5 %	100.0 %

# Análise descritiva

Gráficos para ter ou não favelas:

1) Pela estatística descritiva usar o gráfico de barras



# Análise descritiva

Gráficos para ter ou não favelas:

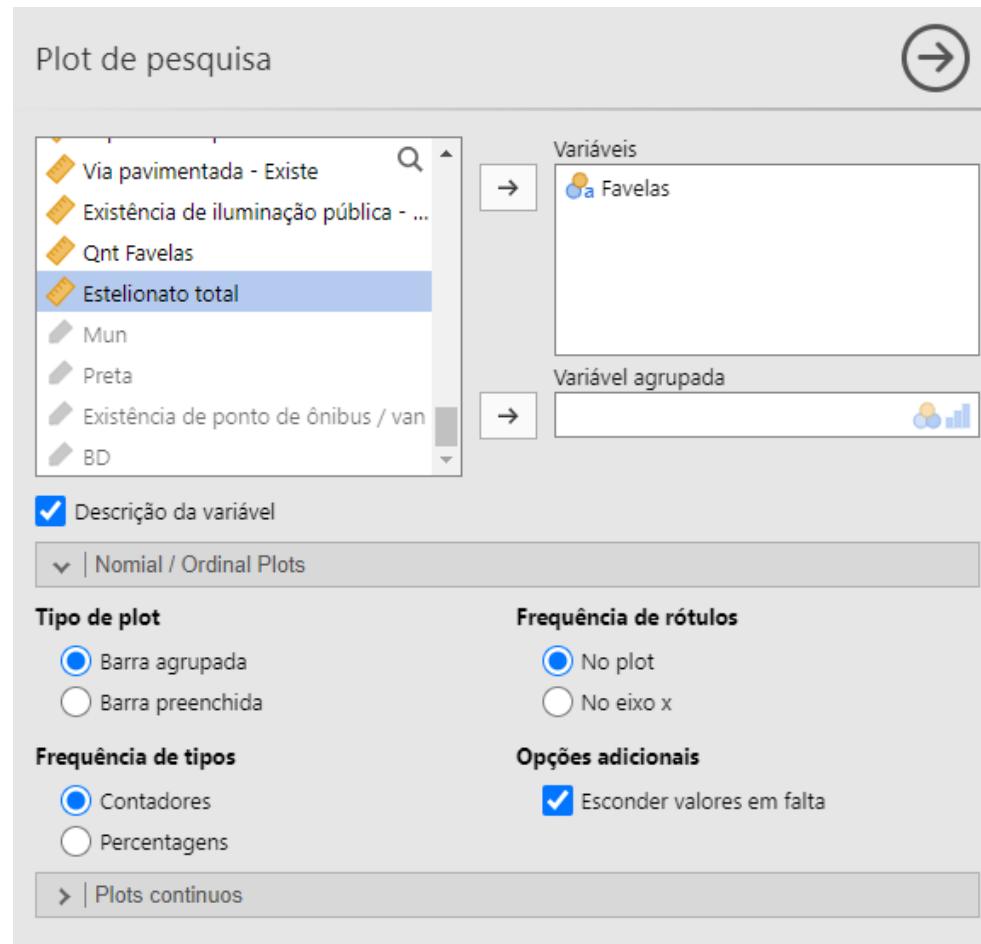
2) Instalar o módulo surveymv

The screenshot shows the RStudio interface. At the top, there's a navigation bar with tabs: 'Variáveis', 'Dados', 'Análises', 'Editar', and 'Módulos'. The 'Módulos' tab is highlighted with a red circle. Below the navigation bar, there are several icons for different analysis types: Exploração, Testes t, ANOVA, Regressão, Frequências, Fator, and jYS. The main area shows a package manager with the 'Disponível' tab selected. A search bar at the top of the manager has 'survey' typed into it, with a yellow highlight. Two packages are listed: 'surveymv - Survey Plots 0.7.2' by Ravi Selker, which is described as generating summary plots for survey data and is marked as 'INSTALADO'; and 'OneHotEncoding - Converts categorical variables into binary columns. 1.0.0' by Imad Chakri.

# Análise descritiva

Gráficos para ter ou não favelas:

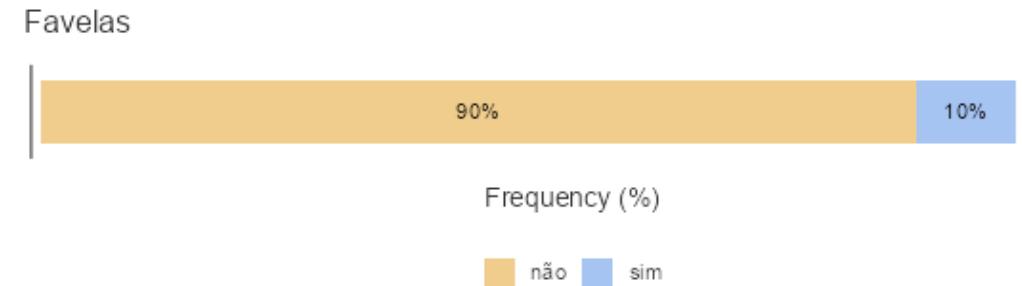
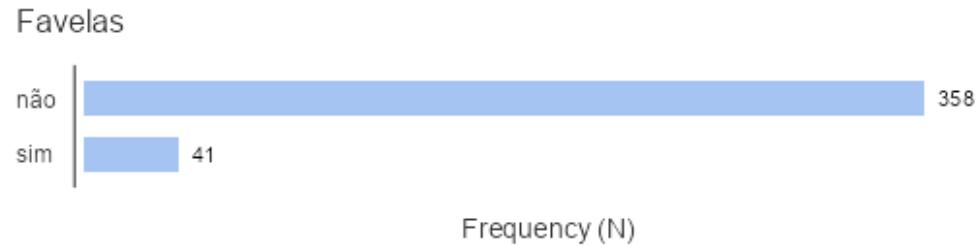
2) Instalar o módulo surveymv



# Análise descritiva

Gráficos para ter ou não favelas:

2) Instalar o módulo surveymv



# Análise descritiva

## Gráficos para ter ou não favelas e existência da guarda municipal:

## 2) Instalar o módulo surveymv

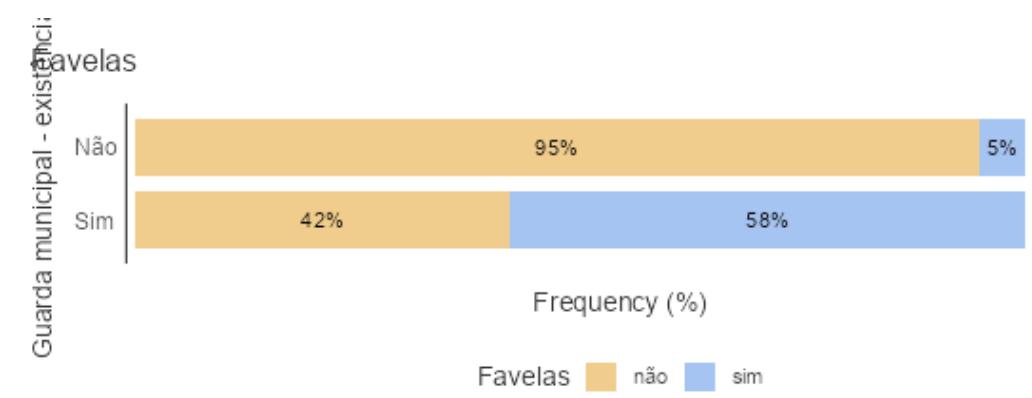
## Plot de pesquisa

Descrição da variável

▼ | Nominal / Ordinal Plots

<b>Tipo de plot</b>	<b>Frequência de rótulos</b>
<input type="radio"/> Barra agrupada	<input checked="" type="radio"/> No plot
<input checked="" type="radio"/> Barra preenchida	<input type="radio"/> No eixo x
<b>Frequência de tipos</b>	<b>Opções adicionais</b>
<input type="radio"/> Contadores	<input checked="" type="checkbox"/> Esconder valores em falta
<input checked="" type="radio"/> Percentagens	

> | Plots continuos



# Análise descritiva

Exercício:

2) No banco de dados município, fazer as frequências e gráficos para as variáveis conselho municipal de segurança e guarda municipal.

1. Frequência das variáveis individuais
2. Tabela cruzada
3. Gráficos individuais e gráficos cruzados

# Análise descritiva

- **Proporção:**
- Representa a relação entre uma parte de um conjunto e o todo. Por isso, todas as unidades do numerador devem estar incluídas no denominador.

$$P = \frac{a}{a+b}$$

# Análise descritiva

- **Exercício:**
- Qual a proporção de furtos qualificados em relação ao furto total em Curitiba em 2024?

$$P = \frac{9307}{45106} = 0,206 \text{ ou } 20\% \text{ ou seja, } 20 \text{ a cada } 100 \text{ furtos foram furtos qualificados}$$

# Análise descritiva

- **Exercício:**
- Qual a proporção de furtos qualificados em relação ao furto total em Curitiba em 2024?

No Jamovi

- 1) Criar uma coluna furto total somando furto qualificado e furto simples
- 2) Criar a variável proporção de furtos por município
- 3) Fazer um filtro para Curitiba

# Análise descritiva

- **Exercício:**
- Qual a proporção de furtos qualificados em relação ao furto total em Curitiba em 2024?

No Jamovi

- 1) Criar uma coluna furto total somando furto qualificado e furto simples



# Análise descritiva

- **Exercício:**
- Qual a proporção de furtos qualificados em relação ao furto total em Curitiba em 2024?

No Jamovi

- 2) Criar a variável proporção de furtos por município

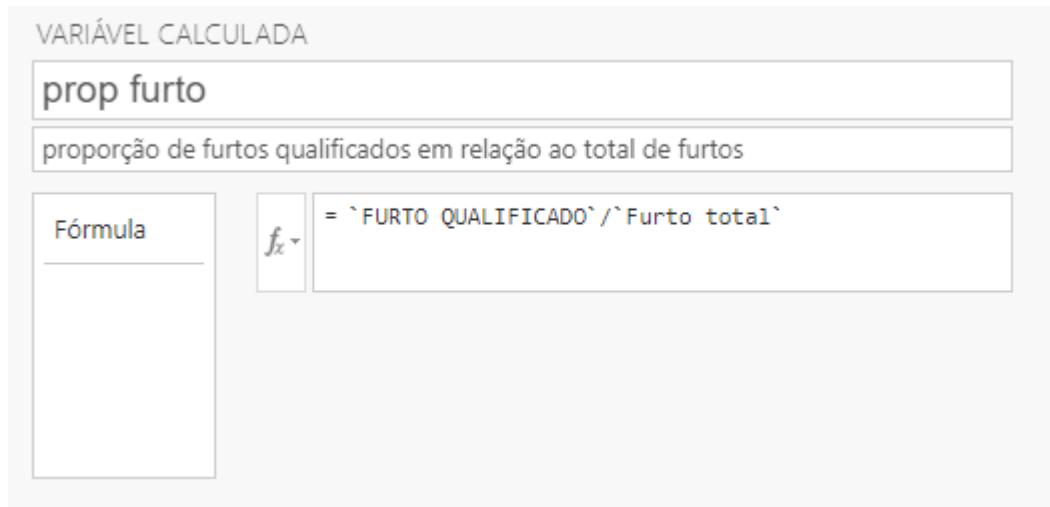
VARIÁVEL CALCULADA

prop furto

proporção de furtos qualificados em relação ao total de furtos

Fórmula

$f_x = \text{`FURTO QUALIFICADO`} / \text{`Furto total'}$



# Análise descritiva

- **Exercício:**
- Qual a proporção de furtos qualificados em relação ao furto total em Curitiba em 2024?

No Jamovi

3) Fazer um filtro para Curitiba

The screenshot shows the Jamovi software interface. At the top, there's a toolbar with tabs for 'Dados' (Data), 'Análises' (Analyses), and 'Editar' (Edit). Below the toolbar are various icons for data management: 'Configurar' (Configure), 'Calcular' (Calculate), 'Transformar' (Transform), 'Variáveis' (Variables), 'Pesos' (Weights), 'Adicionar' (Add), 'Apagar' (Delete), 'Filtros' (Filters), 'Linhas' (Rows), and 'Apagar' (Delete again). A red circle highlights the 'activo' (active) switch for the first filter in the 'FILTROS EM LINHA' (Line Filters) panel. The 'Filtro 1' (Filter 1) is set to 'Mun=="Curitiba"'. Below the filters, a data table is displayed with columns: 'Filtro 1', 'ano', 'cidade', 'CodMun', 'ADQUIRI...', 'AMEACA', and 'APR...'. The data rows show various entries, with the last row having a green checkmark in the 'Filtro 1' column, indicating it matches the filter.

Filtro 1	ano	cidade	CodMun	ADQUIRI...	AMEACA	APR...
✗	2024		4106506	11	263	
✗	2024		4106555	5	35	
✗	2024		4106571	1	95	
✗	2024		4106605	56	316	
✗	2024		4106704	1	42	
✗	2024		4106803	2	184	
✗	2024		4106852	2	24	
✓	2024		4106902	1958	15591	
✗	2024		4107009	7	155	

# Análise descritiva

- **Taxa:**

- ✓ Taxa é a variação de uma medida y em função da variação de uma medida x, em um determinado tempo.
- ✓ Em geral, a taxa está associada com a rapidez de mudança de fenômenos.
- ✓ As taxas para os dados de violência são multiplicados por 100.000 habitantes, mas podemos usar outras escalas como 1.000 habitantes, 10.000 e etc.

# Análise descritiva

- **Taxa:**

Exemplo:

Tomando como exemplo a taxa de roubos em Curitiba em 2023 por 1.000 habitantes, o cálculo seria:

$$T = \frac{\text{número de roubos em Curitiba no ano de 2023}}{\text{população de Curitiba no ano de 2023}} \times 1.000$$

# Análise descritiva

- **Exercício:**
- Qual a taxa de mortalidade no Paraná em 2022?

- 1) Usar o número de mortes no Paraná em 2022 extraído do DATASUS
- 2) Usar a população do Paraná extraído do Sidra em 2022
- 3) Fazer o cálculo

# Análise descritiva

- **Exercício:**
- Qual a taxa de mortalidade no Paraná em 2022 por 1.000 habitantes?

$$T = \frac{\text{número mortes no Paraná em 2022}}{\text{população do Paraná em 2022}} \times 1.000$$

$$T = \frac{89.808}{11.444.380} \times 1.000 = 78,4$$

# Análise descritiva

- **Exercício:**
- Qual a taxa de mortalidade no Paraná em 2022 por 10.000 e 100.000 habitantes?

$$T = \frac{\text{número mortes no Paraná em 2022}}{\text{população do Paraná em 2022}} \times 10.000$$

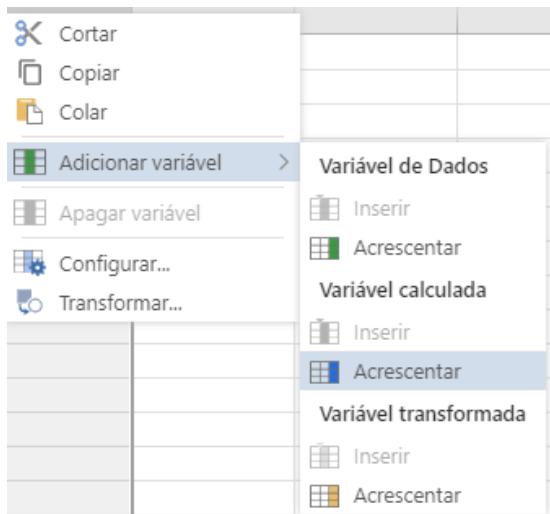
$$T = \frac{\text{número mortes no Paraná em 2022}}{\text{população do Paraná em 2022}} \times 100.000$$

$$T = \frac{89.808}{11.444.380} \times 1.000 = 784,77$$

$$T = \frac{89.808}{11.444.380} \times 100.000 = 7847,7$$

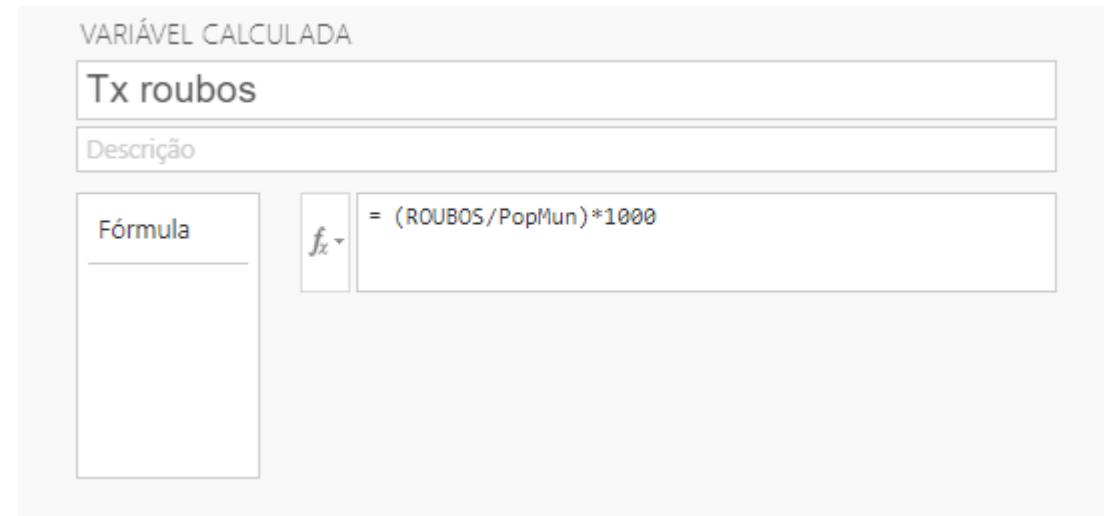
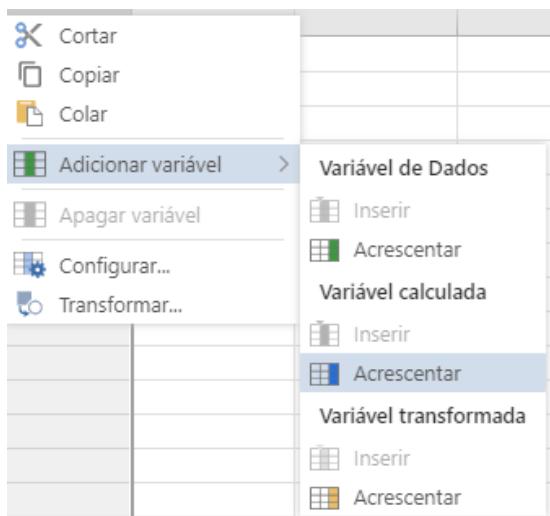
# Análise descritiva

- **Exercício:**
- Usar o banco de dados município para criar uma variável com a taxa de roubos nos municípios do Paraná em 2024 e analisar.
- 1) Adicionar uma nova variável calculada



# Análise descritiva

- **Exercício:**
- Usar o banco de dados município para criar uma variável com a taxa de roubos por mil habitantes nos municípios do Paraná em 2024 e analisar.
- 1) Adicionar uma nova variável calculada



# Estatística Descritiva

- **Medidas de Tendência central:** Moda, Média e Mediana.

Quando falamos em tendência central, estamos tentando descrever o valor do meio ou os valores típicos de uma distribuição.

- Onde os dados estão concentrados?
- Quais valores parecem ser típicos?

# Estatística Descritiva

- A média ou  $\bar{x}$  é a soma dos valores dividida pelo número de valores observados

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

sendo n número de observações  $x_1, x_2, \dots, x_n$ ,

A média é uma medida de centro sensível à influência das observações extremas.

# Estatística Descritiva

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

- Dada a nota final de 10 alunos do curso de métodos de pesquisa quantitativa:

10 6 5 7 7 8 4 10 7 8

Qual a média da distribuição?

$$\bar{x} = \frac{10+6+5+7+7+8+4+10+7+8}{10} = 7,2$$

# Estatística Descritiva

- Média para dados agrupados

$$\bar{X} = \frac{\sum_{i=1}^k f_i \cdot \bar{x}_i}{n}$$

$f_i$ = Frequência Absoluta em cada classe

$x_i$ = Ponto médio de cada intervalo de classe

$k$  = Número de classes

# Estatística Descritiva

Qual é a média de idade dos 30 primos de uma família?

Idade	Frequência
6-10	5
11-15	8
16-20	17
Total	30

$$\bar{x} = \frac{\sum_{i=1}^k f_i \cdot \bar{x}_i}{n}$$

Ponto médio
6 7 <b>8</b> 9 10
11 12 <b>13</b> 14 15
16 17 <b>18</b> 19 20

$$\bar{x} = \frac{(5*8)+(8*13)+(17*18)}{30} = 15$$

# Estatística Descritiva

- Média ponderada:

$$Mp = \frac{(N_1 * P_1) + (N_2 * P_2) + \dots + (N_x * P_x)}{P_1 + P_2 + \dots + P_x}$$

Onde,

N= número de observações

P = peso

# Estatística Descritiva

$$Mp = \frac{(N1*P1)+(N2*P2) \dots (Nx*Px)}{P1+P2+\dots+Px}$$

Calcule a média final do aluno:

Nota	Peso
80	30
20	30
70	40

$$Mp = \frac{(80*30)+(20*30)+(70*40)}{30+30+40} = 58$$

# Estatística Descritiva

- **Mediana (Md) :**

- ✓ É o ponto do meio de uma distribuição, o número tal que a metade (50%) das observações são menores do que ele e metade (50%) maiores.
- ✓ A mediana é especialmente útil quando existem valores extremos.
- ✓ Os valores para a variável em estudo têm distribuição de frequências assimétrica (das ferramentas gráficas)

# Estatística Descritiva

- ✓ Quando o grupo de valores é ímpar, a mediana é o termo central da distribuição. Sua posição pode ser calculada através da seguinte fórmula:

$$Md = \frac{n + 1}{2}$$

- ✓ Quando o grupos de valores é par, a mediana é a média aritmética dos termos centrais. Sua posição pode ser calculada através da seguinte fórmula:

$$Md = \frac{n}{2} \text{ e } \frac{n}{2} + 1$$

# Estatística Descritiva

$$Md = \frac{n}{2} \text{ e } \frac{n}{2} + 1$$

- Dada a nota final de 10 alunos do curso de métodos de pesquisa quantitativa:

10 6 5 7 7 8 4 10 7 8

Qual a posição da mediana? Qual é a mediana?

4 5 6 7 7 7 8 8 10 10

$$\text{Posição} = \frac{n}{2} = \frac{10}{2} = 5^{\text{o}} \text{ termo} ; = \frac{n}{2} + 1 = \frac{10}{2} + 1 = 6^{\text{o}} \text{ termo}$$

$$Md = 7+7/2 = 7$$

# Estatística Descritiva

$$Md = \frac{n+1}{2}$$

- Dada a nota final de 11 alunos do curso de métodos de pesquisa quantitativa:

10 6 5 7 7 8 4 10 7 8 3

Qual a posição da mediana? Qual é a mediana?

3 4 5 6 7 7 7 8 8 10 10

$$\text{Posição} = \frac{n+1}{2} = \frac{11+1}{2} = 6^{\text{o}} \text{ termo ;}$$

■ ■ ■ ■ ■

# Estatística Descritiva

- **Moda**
  - ✓ É o valor com maior frequência de ocorrência nos dados.
  - ✓ A moda é fácil de definir mas não é fácil de ser calculada, porque requer a tabulação da frequência de ocorrência de todos os valores distintos dos dados.
  - ✓ É a única medida de tendência central útil para dados qualitativos.
  - ✓ Para os dados agrupados a moda será a classe modal.

# Estatística Descritiva

Dada a nota final de 10 alunos do curso de métodos de pesquisa quantitativa:

10 6 5 7 7 8 4 10 7 8

Qual é o valor da moda?

Moda =7

Se a nota dos 10 alunos fosse:

10 8 5 7 7 8 4 10 7 8

Moda = 7 e 8 Distribuição Bimodal

# Estatística Descritiva

- **Moda X Média X Mediana**

Em cinco testes realizados, um consumidor interessado em três modelos de carro compacto obteve a seguinte relação de consumo em quilômetros por litro de gasolina.

	Teste 1	Teste 2	Teste 3	Teste 4	Teste 5
Carro A	28	32	28	30	34
Carro B	31	29	31	29	31
Carro C	29	32	28	32	30

Qual será a medida de tendência central usada pelo fabricante para fazer uma boa propaganda de seus carros?

Carro A: Média

Carro B: Mediana ou Moda

Carro C: Moda

# Estatística Descritiva

- Exercício:
- Usando o banco de dados 1
- 1) Qual é a média, moda e variância da idade das vítimas de homicídio?

Estatística Descritiva

Variáveis

Idade

Tendência Central

Média

Mediana

Moda

Soma

Estatística Descritiva

	Idade
N	6426
Omissos	58
Média	39.4
Mediana	37.0
Moda	26.0

## Estatística Descritiva

Estatística Descritiva	
	Idade
N	6426
Omissos	58
Média	39.4
Mediana	37.0
Moda	26.0

# Estatística Descritiva

- **Medidas de Dispersão:** Amplitude, Desvio Médio, Variância, Desvio Padrão e Coeficiente de Variação.
  - Quanta variação há nesses dados?
  - Quão dispersos estão os valores?
  - Há valores incomuns?

# Estatística Descritiva

## Amplitude

Maior Observação ( valor máximo) – Menor Observação ( valor mínimo)

- É fortemente afetada pela presença de valores atípicos nos extremos dos dados.

Dada a nota final de 10 alunos do curso de métodos de pesquisa quantitativa:

10 6 5 7 7 8 4 10 7 8

Qual a amplitude da distribuição?

Amplitude =  $10 - 4 = 6$

# Estatística Descritiva

## Desvio Médio

É a média aritmética dos valores absolutos dos desvios.

Os desvios são obtidos pela diferença entre cada valor observado e a média de seus valores.

$$DM = \frac{\sum |x_i - \bar{x}|}{n}$$

Onde,

$x_i$  = observação

$\bar{x}$  = média

n= total de observação

Quando maior o valor encontrado, maior o desvio da distribuição em relação a média.

# Estatística Descritiva

$$DM = \frac{\sum |x_i - \bar{x}|}{n}$$

Dada a nota final de 10 alunos do curso de métodos de pesquisa quantitativa:

10 6 5 7 7 8 4 10 7 8

Dada a nota final de 11 alunos do curso de métodos de pesquisa quantitativa:

10 6 5 7 7 8 4 10 7 8 3

Qual o desvio médio das duas distribuições? Qual teve maior variabilidade?

DM1= 1,44

DM2= 1,68

# Estatística Descritiva

A variância é uma medida de dispersão que mostra o quanto distante cada valor desse conjunto está do valor central (médio).

Variância é a média aritmética dos desvios quadrados.

$$\sigma^2 = \frac{\sum(x_i - \mu)^2}{N}, \text{ onde...}$$

$\sigma^2$ : variância populacional;  
 $x_i$ : cada observação do conjunto de dados populacional;  
 $\mu$ : média da população;  
 $N$ : número de observações.

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}, \text{ onde...}$$

$s^2$ : variância da amostra;  
 $x_i$ : cada observação do conjunto amostral;  
 $\bar{x}$ : média da amostra;  
 $n$ : número de observações da amostra.

# Estatística Descritiva

## • Desvio Padrão

- É a raiz quadrada da variância.

$$D_p = \sqrt{var}$$

- Desvios-padrão somente podem ser comparados para conjuntos de dados medidos nas mesmas unidades.
- Um grande desvio padrão indica que os pontos dos dados estão espalhados longe da média e um pequeno desvio padrão indica que os pontos dos dados estão agrupados perto da média.

# Estatística Descritiva

Dada a nota final de 10 alunos do curso de métodos de pesquisa quantitativa:

10 6 5 7 7 8 4 10 7 8

Dada a nota final de 11 alunos do curso de métodos de pesquisa quantitativa:

10 6 5 7 7 8 4 10 7 8 3

Qual o desvio padrão das duas distribuições? Qual distribuição teve maior variabilidade?

D<sub>p1</sub>=1,93

D<sub>p2</sub>=2,22

# Estatística Descritiva

- **Coeficiente de Variação (CV)** é usado para comparar a variabilidade de conjuntos de dados com diferentes unidades de medida Expresso como porcentagem para descrever o desvio-padrão em relação à média.

$$CV = \frac{100 \cdot s}{\bar{x}}$$

Onde:

S = Desvio padrão

$\bar{x}$  = média

Quanto menor o coeficiente de variação, mais homogênea é a variável naquele conjunto de dados.

Variável	$\bar{x}$	DP	CV
Idade	35,6 anos	5,08 anos	0,143 (14,3%)
Tempo de profissão	6,5 anos	2,98 anos	0,458 (45,8%)
Salário	537,52 reais	25,34 reais	0,047 (4,7%)

# Estatística Descritiva

Exercício:

1) Calcular o máximo, mínimo, amplitude, variância e desvio padrão da idade das vítimas

▼ | Estatísticas

**Dimensão da Amostra**

N  Omissos

**Valores Percentílicos**

Pontos de Corte para  grupos iguais  
 Percentis

**Dispersão**

Desvio-padrão  Mínimo  
 Variância  Máximo  
 Amplitude  AIQ

**Dispersão média**

Erro-padrão da Média  
 Intervalo de Confiança para a Média  %

**Tendência Central**

Média  
 Mediana  
 Moda  
 Soma

**Distribuição**

Assimetria  
 Curtose

**Normalidade**

Shapiro-Wilk

**Anómalos**

Mais extremo  Valores

## Estatística Descritiva

Estatística Descritiva	
	Idade
N	6426
Omissos	58
Desvio-padrão	17.5
Variância	305
Amplitude	106
Mínimo	0
Máximo	106

# Estatística Descritiva

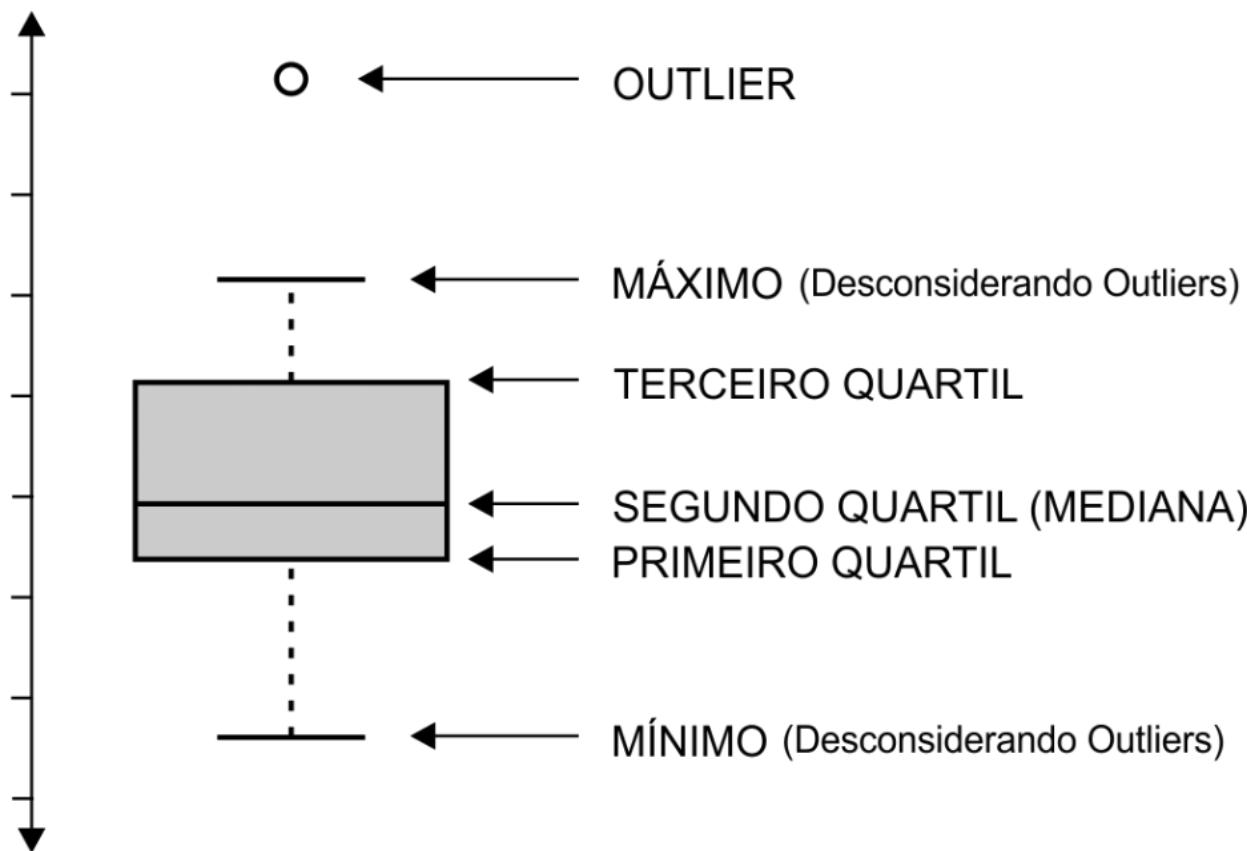
- Exercício usando o Banco de dados 2:
- 2) Qual é a média, moda e variância do crime de perseguição e violência doméstica? Compare.
- 2.1) Calcule o máximo, mínimo, amplitude, variância e desvio padrão para perseguição e violência doméstica
- 2.2) Calcule as medidas de tendência central somente para 2 municípios e compare

# Box Plot

- **Boxplot ou gráfico de caixa:**
- Usado para analisar a distribuição dos dados.
- É possível analisar a posição, dispersão, simetria, caudas e valores discrepantes (outliers) do conjunto de dados
- Este gráfico é formado pelo primeiro e terceiro quartil e pela mediana da distribuição .
- As hastas inferiores e superiores se estendem, respectivamente, do quartil inferior até o menor valor não inferior ao limite inferior e do quartil superior até o maior valor não superior ao limite superior

# Box Plot

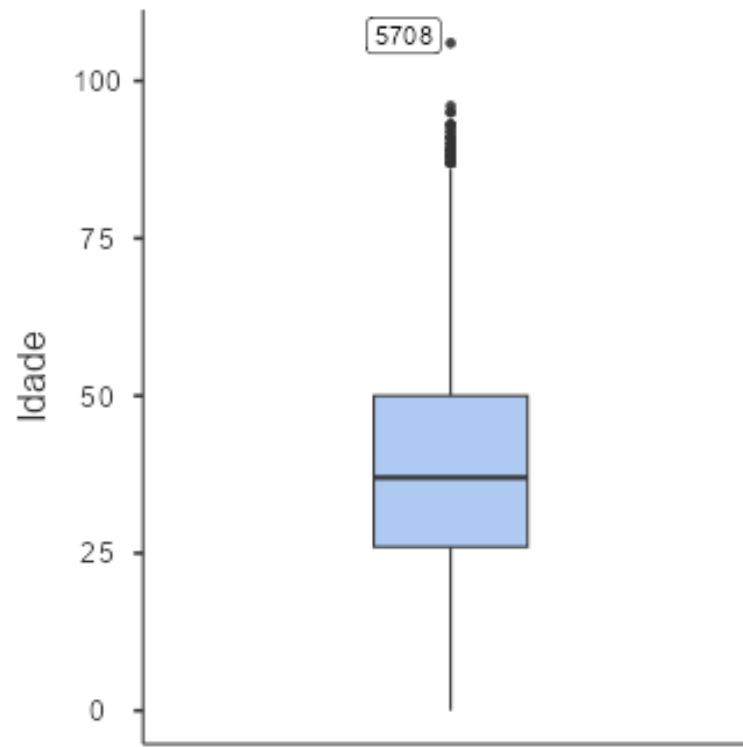
- Boxplot ou gráfico de caixa:



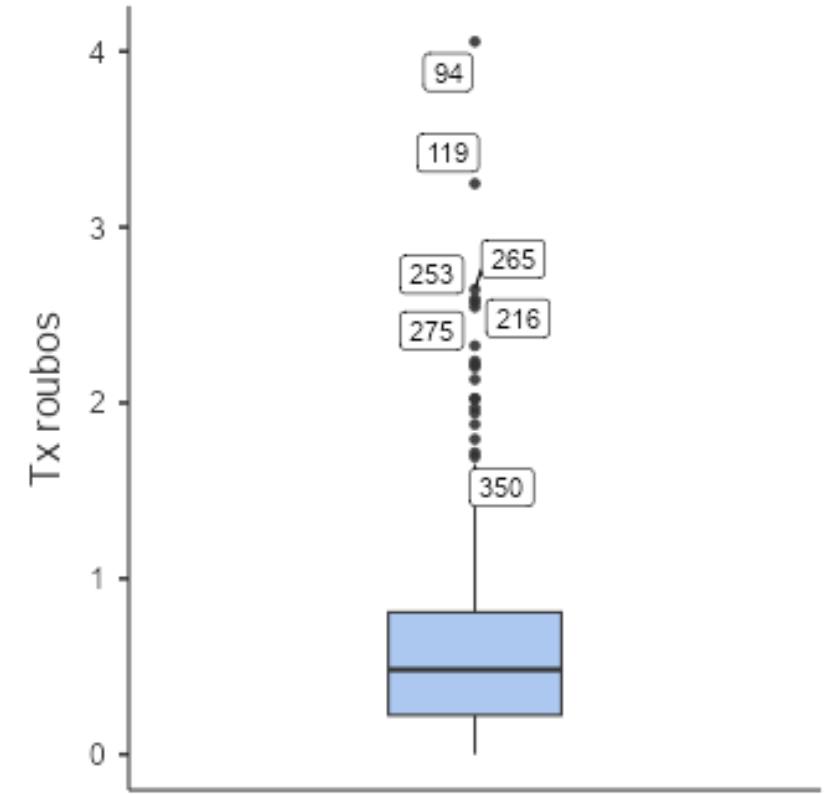
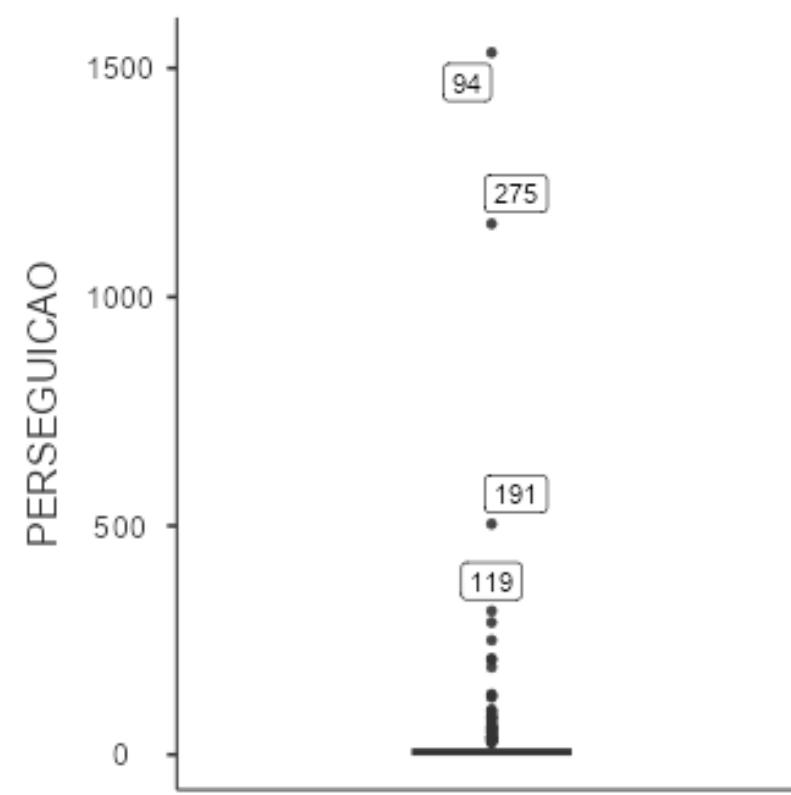
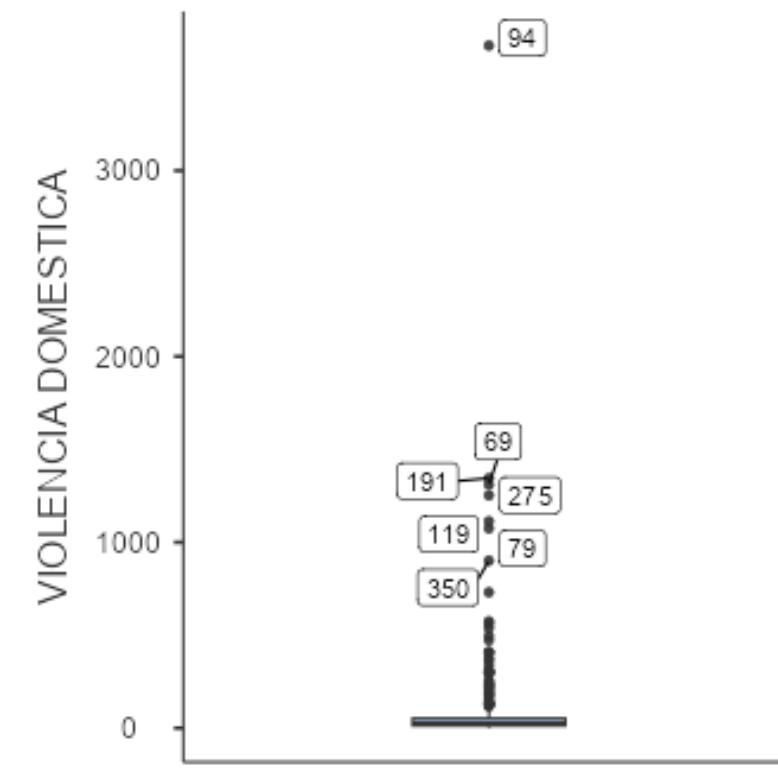
# Box Plot

- **Exercícios:**
  1. Fazer um box plot para idade das vítimas.
  2. Fazer um box plot para perseguição e violência doméstica usando o banco de dados 2.
  3. Fazer um box plot da taxa de roubos usando o banco de dados 2

# Box Plot



# Box Plot



# Organização e visualização dos dados

- Qual é a melhor forma de representar os dados ?
- Existe uma forma única para todos os dados ?
- Que recursos adicionais eu posso utilizar para o tratamento dos dados ?



# Organização e visualização dos dados

- Aqueles que trabalham com a estatística devem procurar organizar, explorar e resumir os dados de uma maneira sucinta.
- Seus métodos podem ser visual (diagramas e gráficos) ou numérico (medidas estatísticas ou tabelas).
- Um conjunto de dados univariados pode ser analisado em termos de duas características, quais sejam: tendência central e dispersão.
- Uma boa análise exploratória pode resolver bastante do seu problema de pesquisa!

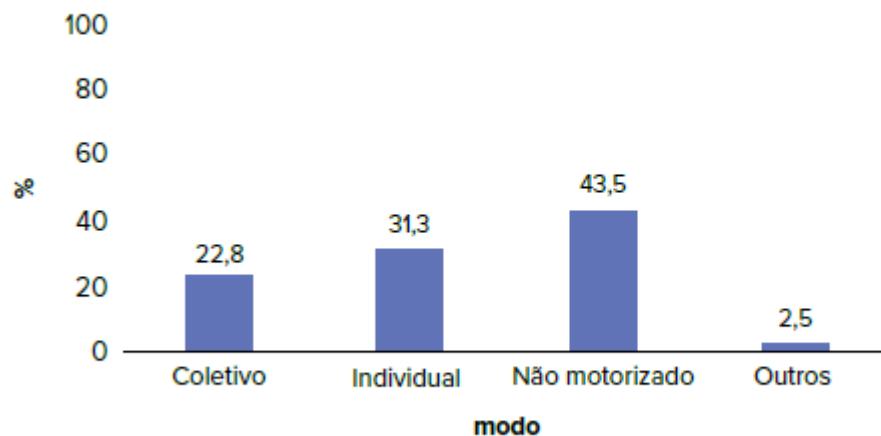
# Organização e visualização dos dados

- **Gráfico de barras:**

- ✓ Dados do tipo categórico são tipicamente representados usando-se o gráfico de barras.
- ✓ Cada barra representa uma categoria ou um atributo.
- ✓ O tamanho de cada barra reflete a frequência de cada categoria.
- ✓ Gráficos de barras verticais são mais comuns, mas os gráficos de barras horizontais podem ser úteis quando os rótulos dos eixos são longos ou quando têm muitas categorias

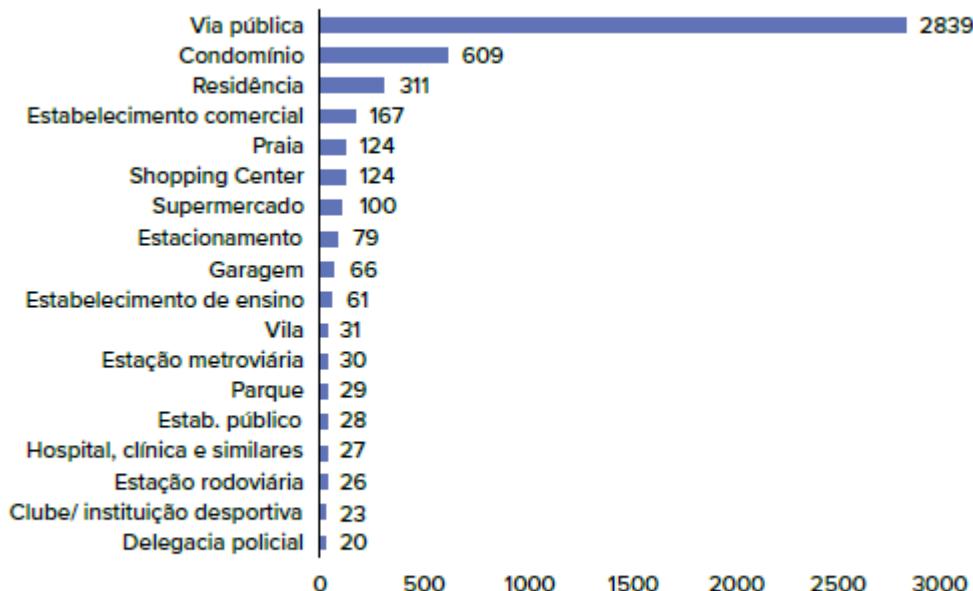
# Organização e visualização dos dados

Gráfico 2 – Modos de transportes usados pelos idosos na RMBH em 2012



Fonte: BHTRANS, 2012.

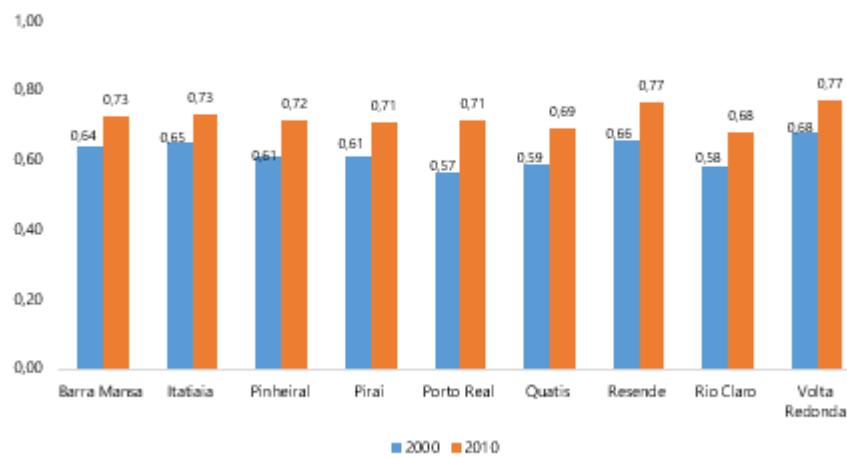
Gráfico 3 – Local de ocorrência de frutos de bicicleta no Rio de Janeiro em 2017



Fonte: ISP, 2017.

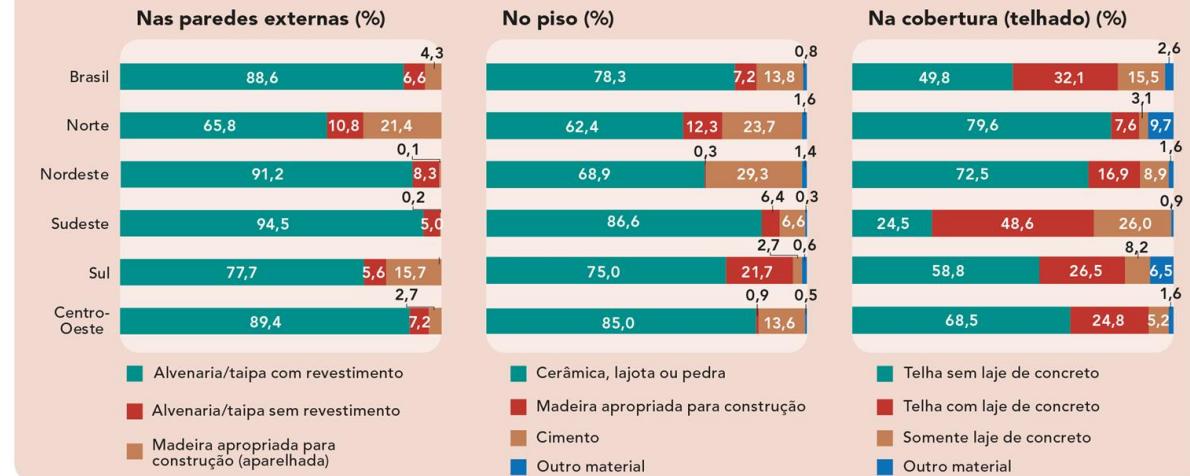
# Organização e visualização dos dados

IDH-M dos municípios do Vale do Paraíba Fluminense de 2000 – 2010



Fonte: PNUD, 2013.

Domicílios, por material predominante, segundo as Grandes Regiões



Fonte: IBGE, Diretoria de Pesquisas, Coordenação de Trabalho e Rendimento, Pesquisa Nacional por Amostra de Domicílios Contínua 2019.

Nota: Domicílios particulares permanentes.

# Organização e visualização dos dados

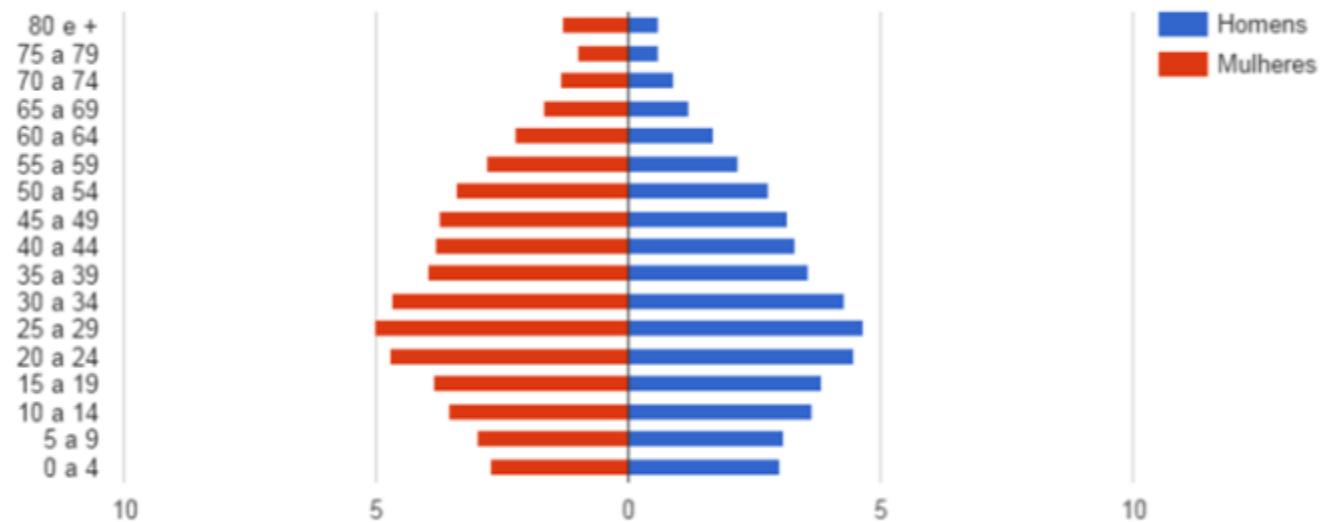
- **Pirâmide etária:**

Pirâmide etária: usada para classificar a população de uma determinada localidade conforme as faixas de idade, dividindo-as por sexo

- ✓ É um gráfico de barras superpostas que se concentram em torno de um eixo.
- ✓ As barras inferiores representam a população mais jovem e as barras superiores representam a população mais velha
- ✓ Do lado direito do eixo, sempre se quantifica a população feminina e, do lado esquerdo, a população masculina.

# Organização e visualização dos dados

Pirâmide etária da população de BH em 2010



Fonte: IBGE, 2011

# Organização e visualização dos dados

## ➤ Gráfico de Pizza:

- O único uso correto de um gráfico de pizza é retratar dados que representam partes de um todo (por exemplo, percentuais de divisão do mercado).
- São ineficientes quando têm fatias demais.
- Passam informação sobre a frequência relativa dos vários grupos melhor que uma tabela de frequências, se houver poucas categorias.
- Para construção é estabelecida uma relação entre um ângulo, em graus e a frequência observada em cada categoria:

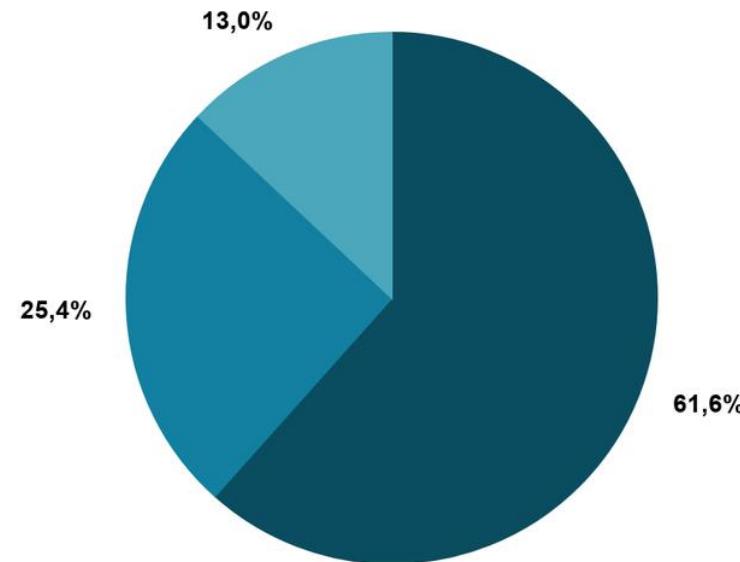
$$\frac{360^\circ}{total} = \frac{\hat{ângulo}}{frequênci\text{a da categoria}}$$

# Organização e visualização dos dados

## Resultado das negociações salariais coletivas

De janeiro a abril de 2021, em relação à inflação\*

■ Abaixo ■ Igual ■ Acima



\*Medida pelo INPC

Fonte: Salariômetro/Fipe

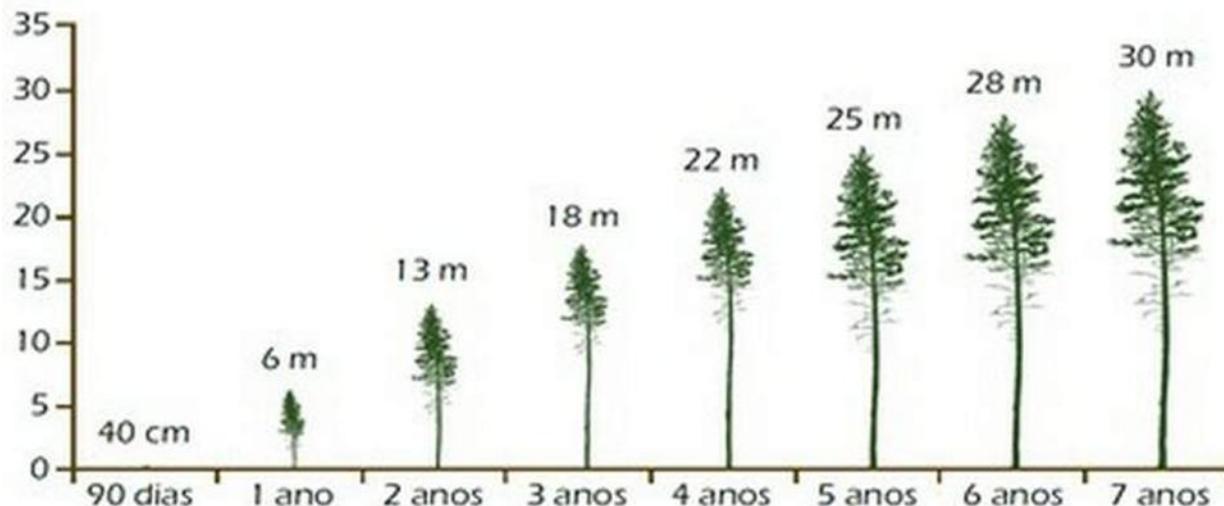
BBC

# Organização e visualização dos dados

## ➤ Pictograma:

- É um representação visual em que os valores dos dados são trocados por figuras ou símbolos para agradar visualmente uma audiência geral.

### Ciclo do crescimento do eucalipto



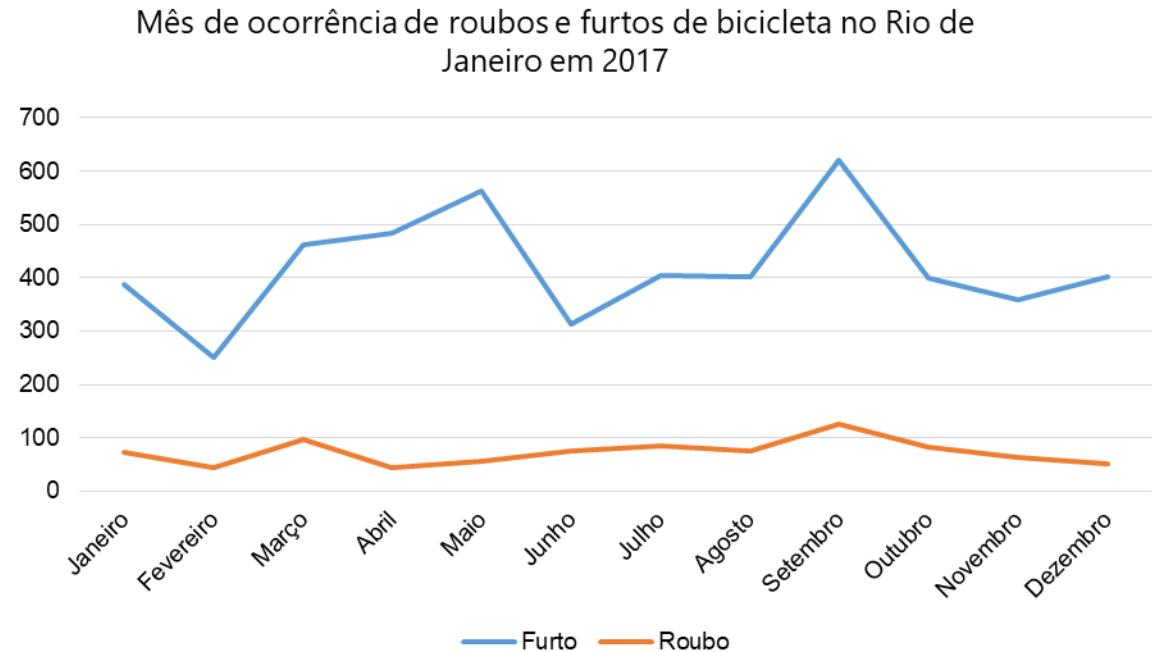
Fonte: IBGE, Diretoria de Pesquisas, Coordenação de Agropecuária, Pesquisa da Pecuária Municipal 2017.

Fonte do gráfico: IBGE, Produção da Pecuária Municipal 2017, pág. 4.

# Organização e visualização dos dados

## ➤ Gráficos temporais:

- Uso para as variáveis que são medidas em intervalos ao longo do tempo.
- Nosso interesse principal reside na mudança ao longo do tempo.
- Sempre coloque o tempo na escala horizontal de seu gráfico e a variável que está medindo na escala vertical.



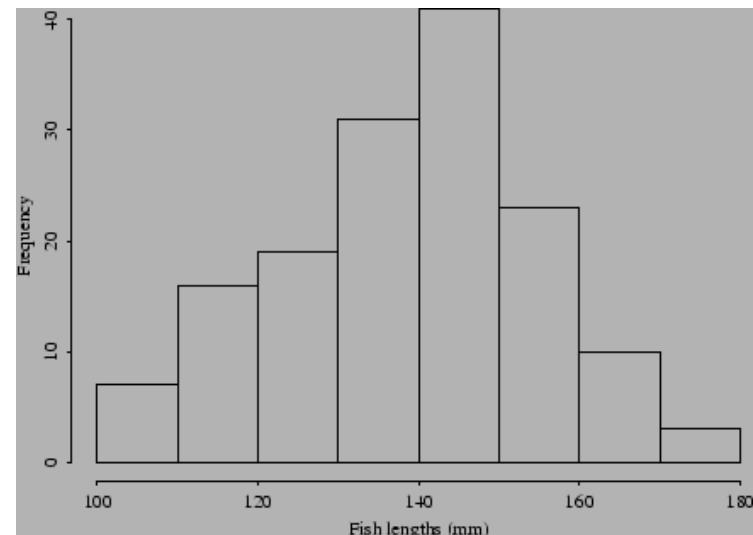
Fonte: ISP, 2017

# Organização e visualização dos dados

- **Histograma:**

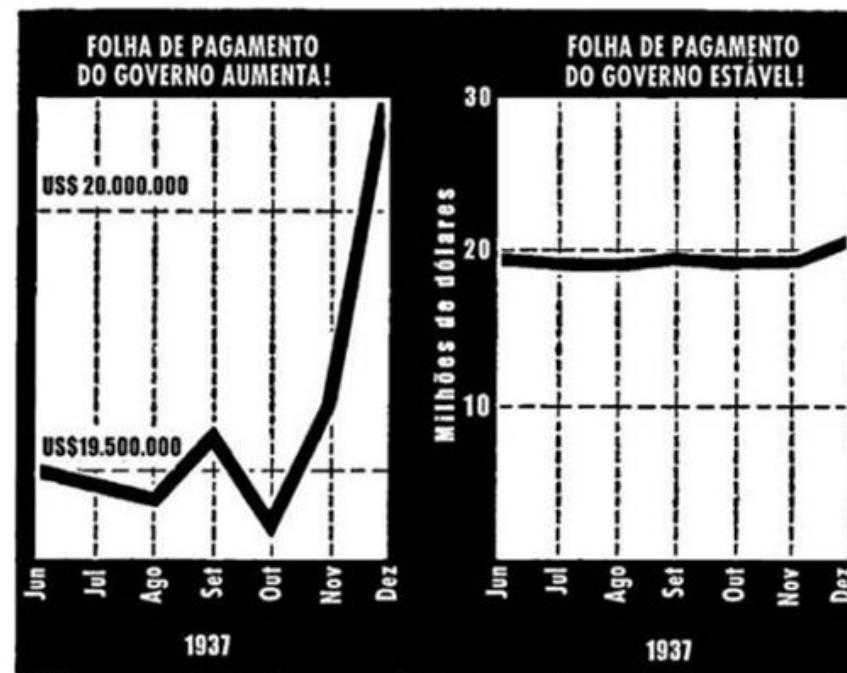
- ✓ Mostra a distribuição de frequência.
- ✓ A escala horizontal representa classes de valores de dados e a escala vertical representa as frequências
- ✓ As alturas das barras correspondem aos valores das frequências, e as barras são desenhadas adjacentes umas às outras.

Comprimento do peixe (mm)	Frequência
100-109	7
110-119	16
120-129	19
130-139	31
140-149	41
150-159	23
160-169	10
170-179	3



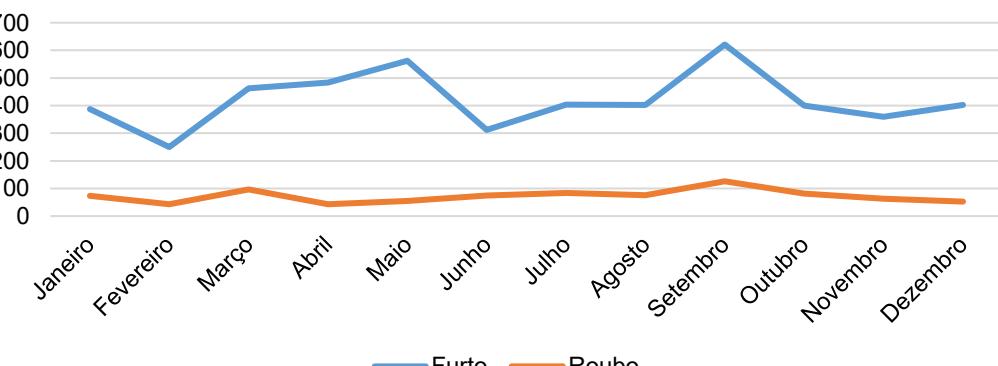
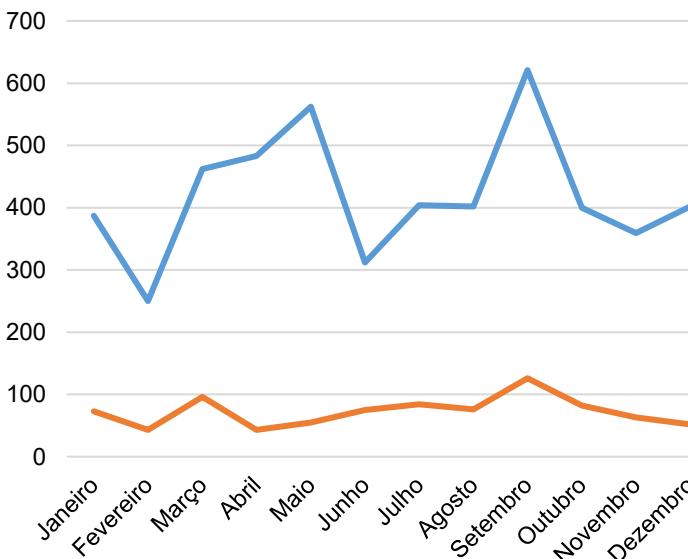
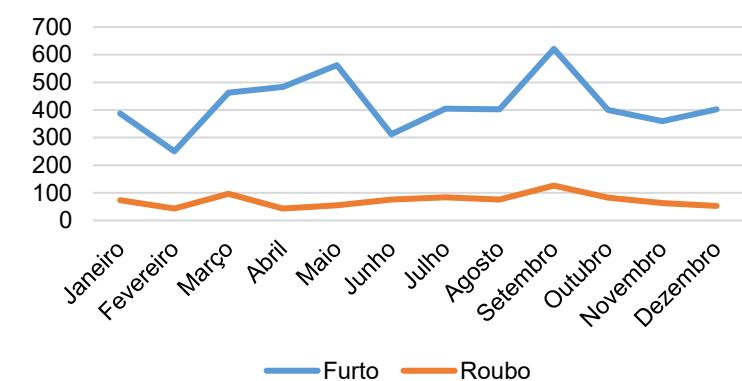
# Organização e visualização dos dados

- Os gráficos sempre precisam ter seus eixos para facilitar a interpretação.
- Sem as “demarcações” nos eixos, o leitor não pode identificar valores individuais dos dados.



# Organização e visualização dos dados

Cuidado ao ajustar os gráficos para não distorcê-los.



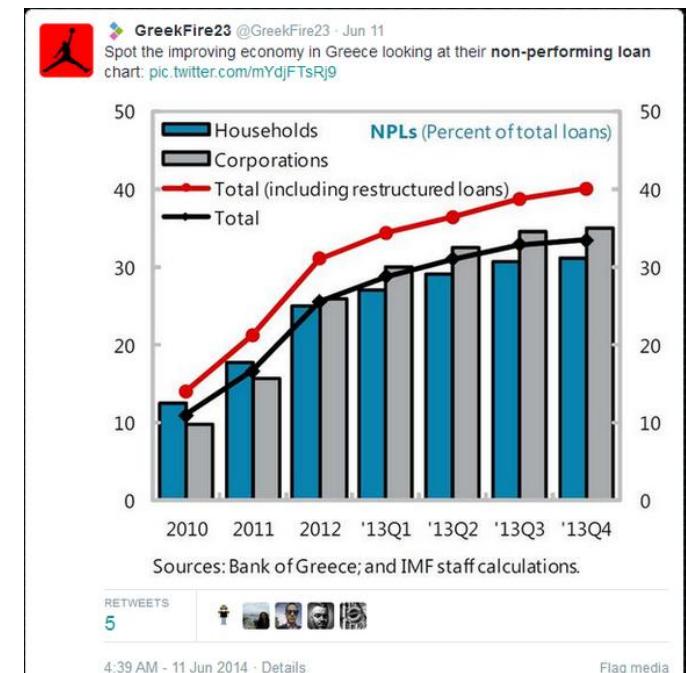
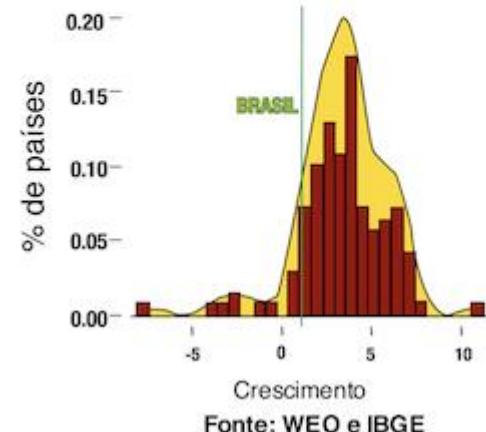
# Organização e visualização dos dados

- Gráficos com muitas informações podem confundir o leitor.
- Gráfico de pizza comumente é usada para variáveis com poucas categorias

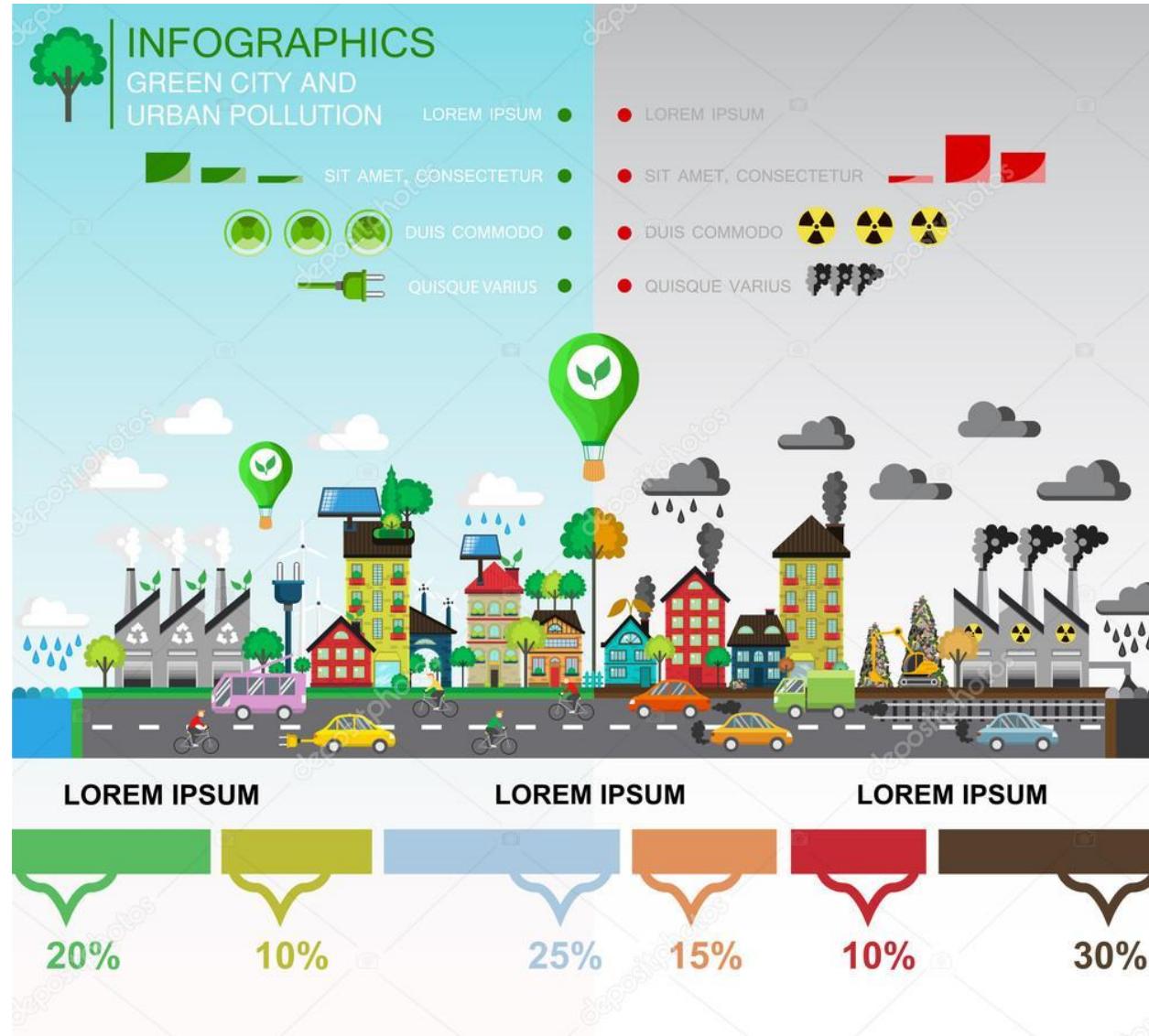
Repartição da oferta interna de energia, em 2018



## CRESCIMENTO DAS ECONOMIAS

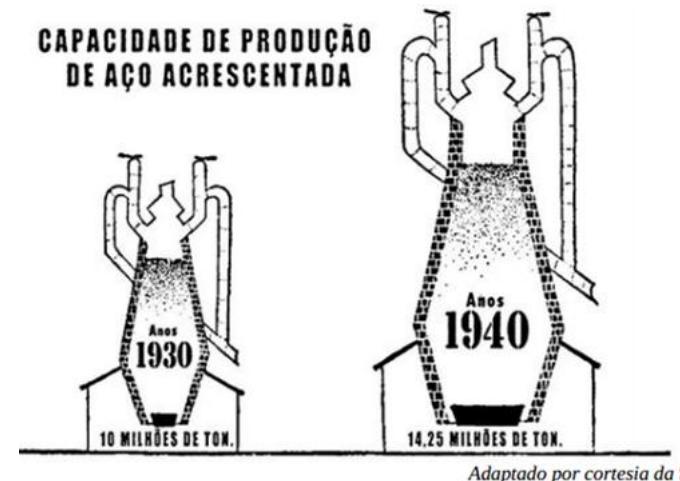


# Organização e visualização dos dados



# Organização e visualização dos dados

- Não distorça a área do gráfico.
- Não troque as barras do gráfico de barras por imagens.
- Como a altura da imagem aumenta, o seu comprimento também aumenta, distorcendo a área.



# Exercício Final

**Em grupos de 4 ou 5 usando o banco de dados 2 – Municípios realizar uma análise exploratória dos dados completa**

- 1) Escolher variáveis quantitativas e qualitativas;
- 2) Fazer tabelas de frequência para as informações qualitativas;
- 3) Fazer uma tabela cruzada com as variáveis qualitativas;
- 4) Fazer os Gráficos;
- 5) Calcular as medidas de tendência central e de dispersão para as variáveis quantitativas;
- 6) Fazer o box plot e outras gráficos que ajudem nas análises das variáveis quantitativas;
- 7) Organizar os dados, escrever ao menos 2 parágrafos de análise e apresentar para a turma.