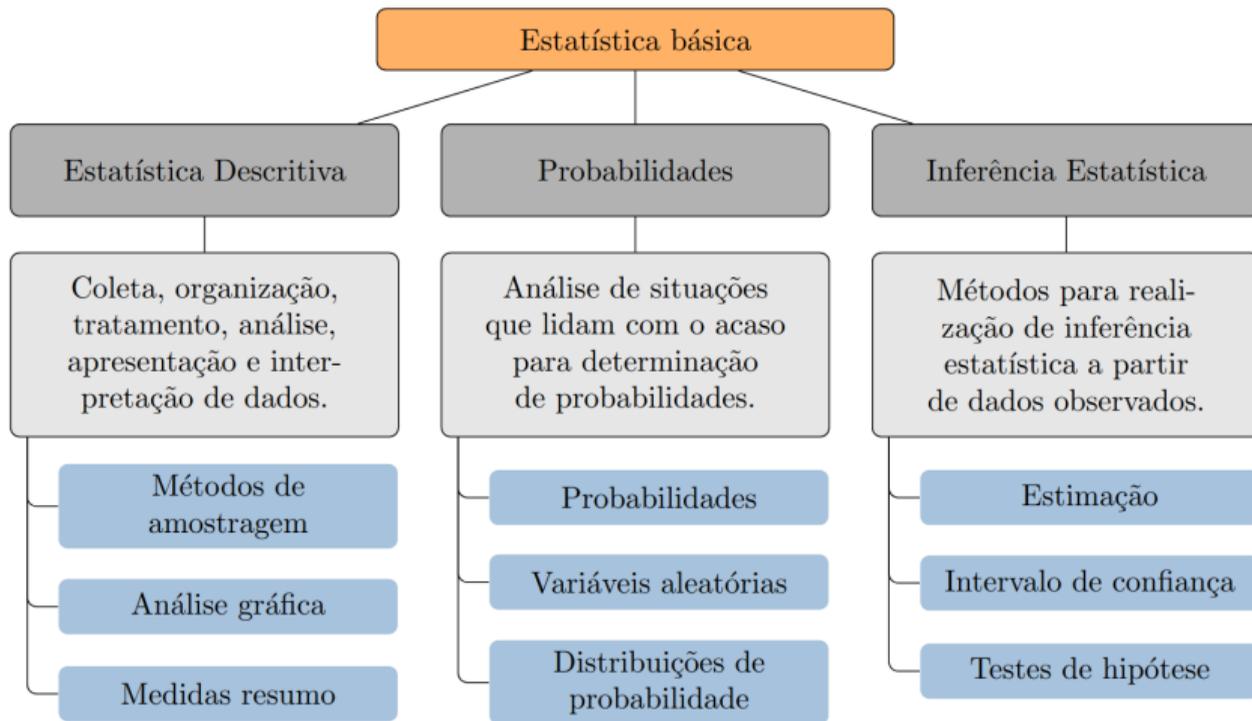


# Análise exploratória de dados

## Parte 1

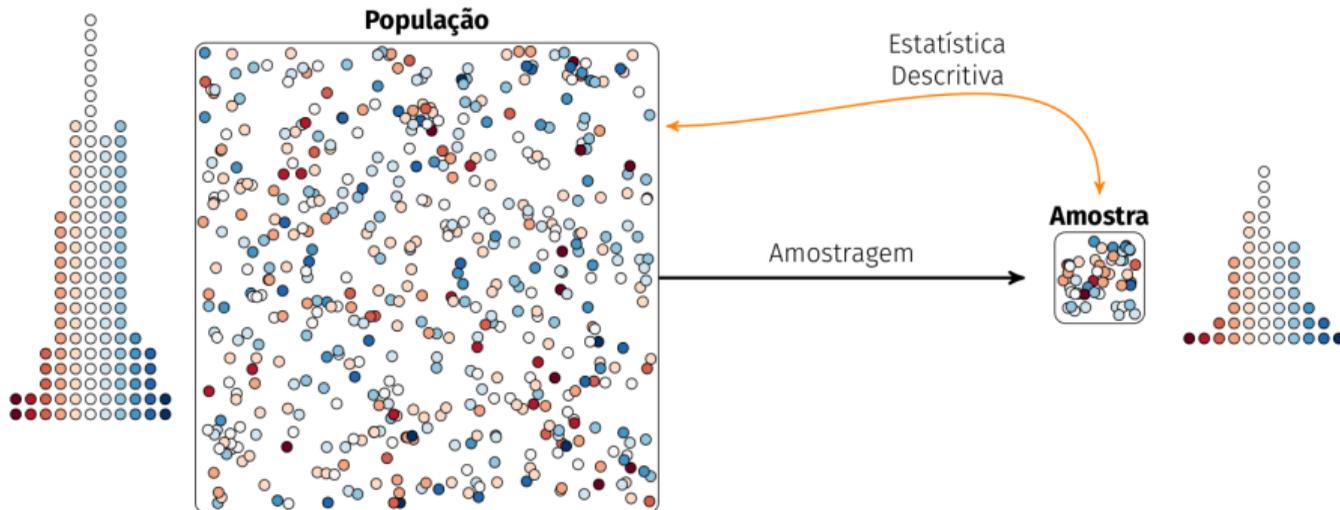
Prof.: Eduardo Vargas Ferreira





# Papel da estatística descritiva

- ▶ A **estatística descritiva** emprega métodos numéricos e gráficos para investigar padrões em um conjunto de dados apresentando-os de uma forma simplificada.



# Motivação

4,801,943

Confirmed (worldwide)



318,481

Deaths (worldwide)



1,786,875

Recovered (worldwide)



2,696,820

Active (worldwide)

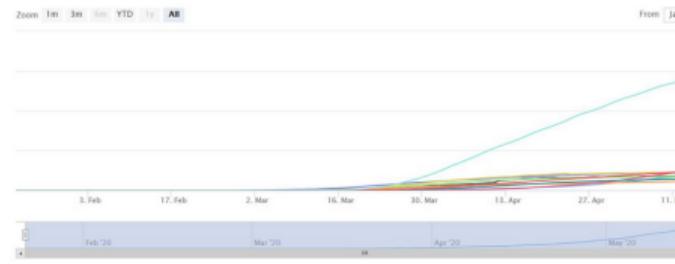
## Cases by Country

Confirmed Recovered Deaths Active



## Cumulative Cases Over Time

Confirmed Recovered Deaths Active



## Data Summary

Show 10 entries

Search:

| Country        | Confirmed | New Cases  |            |        | Deaths | Mortality | Recovered | Active    |
|----------------|-----------|------------|------------|--------|--------|-----------|-----------|-----------|
|                |           | 2020-05-17 | 2020-05-18 | Change |        |           |           |           |
| US             | 1,508,368 | 18,937     | 21,551     | 13.0%  | 90,347 | 6.0%      | 283,178   | 1,135,056 |
| Russia         | 290,678   | 9,709      | 8,926      | -8.1%  | 2,722  | 0.9%      | 70,209    | 217,747   |
| Brazil         | 255,368   | 7,569      | 14,288     | 88.6%  | 16,853 | 6.6%      | 100,459   | 138,056   |
| United Kingdom | 347,709   | 3,534      | 2,714      | -23.2% | 34,876 | 14.1%     | 1,090     | 211,743   |
| Spain          | 231,606   | 0          | 908        |        | 27,709 | 12.0%     | 150,376   | 55,521    |
| Italy          | 325,886   | 675        | 451        | -33.2% | 32,007 | 14.2%     | 127,326   | 66,553    |
| France         | 180,061   | 63         | 258        | 468.3% | 28,242 | 15.7%     | 61,843    | 80,966    |
| Germany        | 176,351   | 617        | 182        | -70.3% | 8,003  | 4.3%      | 155,041   | 13,507    |
| Turkey         | 130,593   | 1,368      | 1,158      | -15.4% | 4,171  | 2.8%      | 111,577   | 34,845    |
| Iran           | 122,492   | 1,806      | 2,204      | 27.0%  | 7,057  | 5.8%      | 95,861    | 19,774    |

Showing 1 to 10 of 10 entries

Previous Next

## Cumulative Cases

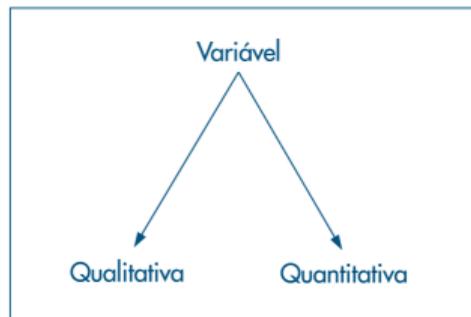


Fonte: UCLA Hackaton

# O que é uma variável

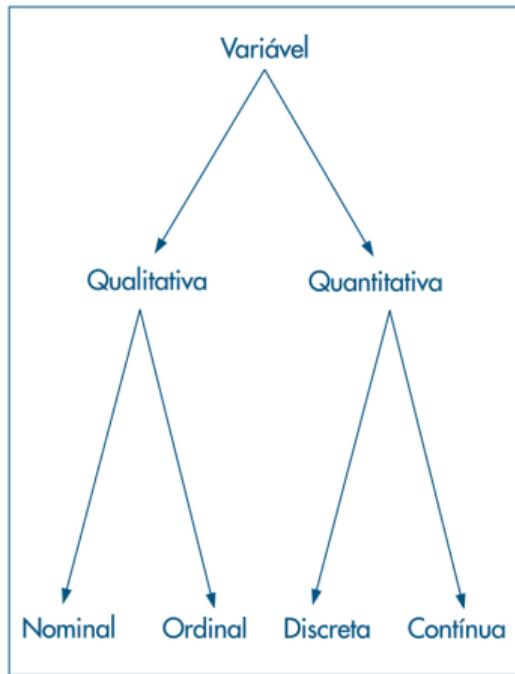
- ▶ **Variável:** é uma característica dos elementos de uma população ou de uma amostra, que pode assumir diferentes valores de interesse ao estudo.

| Variável              | Representação |
|-----------------------|---------------|
| Região de procedência | X             |
| Estado civil          | Y             |
| Grau de instrução     | Z             |
| Número de filhos      | S             |
| Salário               | U             |
| Idade                 | V             |



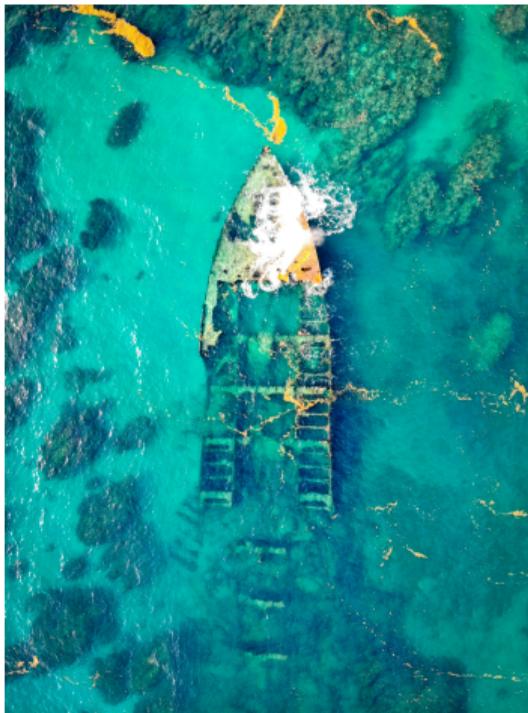
- ▶ Algumas variáveis apresentam uma qualidade:
  - Região de procedência;
  - Estado civil;
  - Grau de instrução.
- ▶ Outras resultam de uma contagem ou mensuração:
  - Número de filhos;
  - Salário;
  - Idade.

# Tipos de Variáveis



- ▶ Algumas variáveis apresentam uma qualidade:
  - Região de procedência;
  - Estado civil;
  - Grau de instrução.
- ▶ Outras resultam de uma contagem ou mensuração:
  - Número de filhos;
  - Salário;
  - Idade.

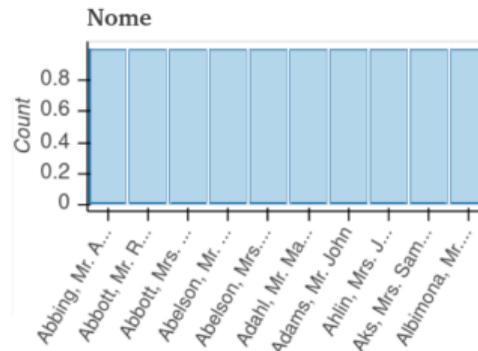
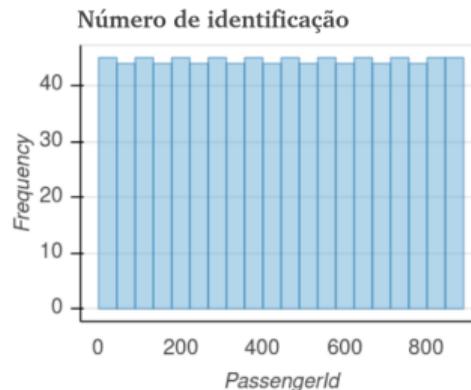
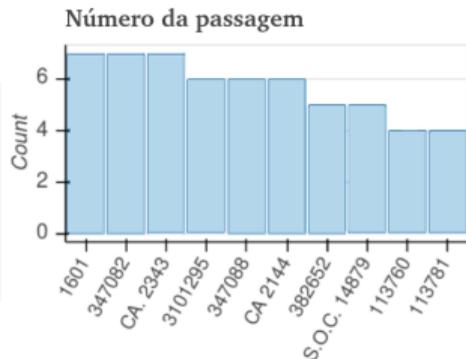
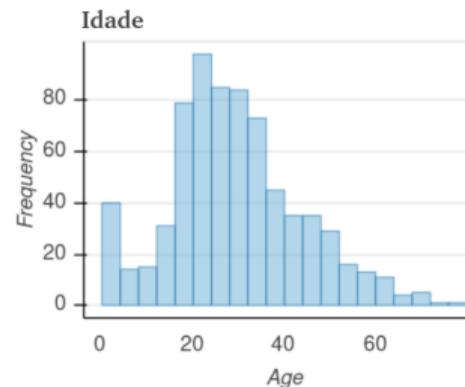
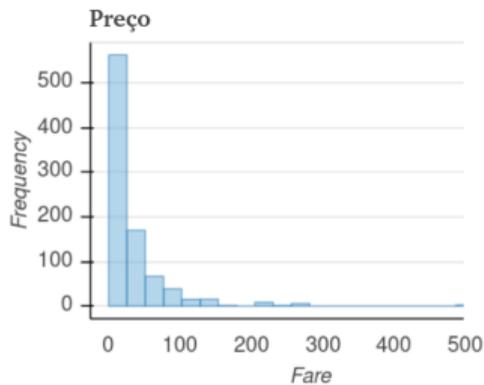
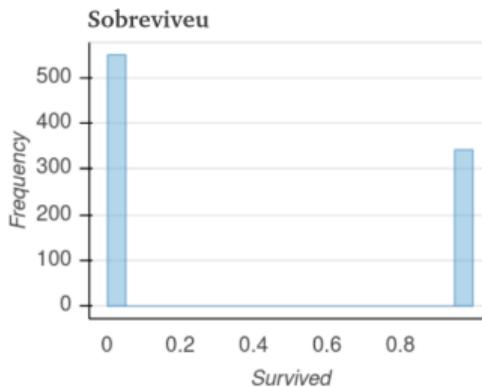
# Exemplo: investigando o naufrágio do Titanic



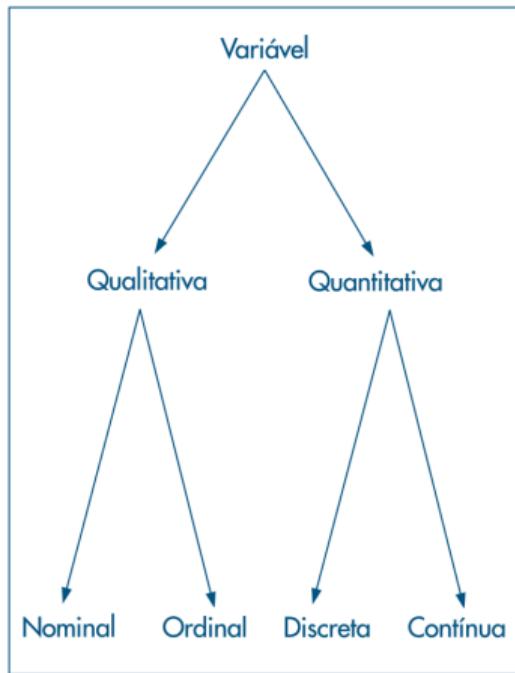
## Dataset Statistics

|                            |          |
|----------------------------|----------|
| Number of Variables        | 12       |
| Number of Rows             | 891      |
| Missing Cells              | 866      |
| Missing Cells (%)          | 8.1%     |
| Duplicate Rows             | 0        |
| Duplicate Rows (%)         | 0.0%     |
| Total Size in Memory       | 315.0 KB |
| Average Row Size in Memory | 362.1 B  |
| Numerical Variables        | 7        |
| Categorical Variables      | 5        |

# Exemplo: investigando o naufrágio do Titanic



# Variável quantitativa



**Discreta:** os valores formam um conjunto enumerável:

- Número de bactérias por litro de leite;
- Número de curtidas;
- Quantidade de acidentes no mês.

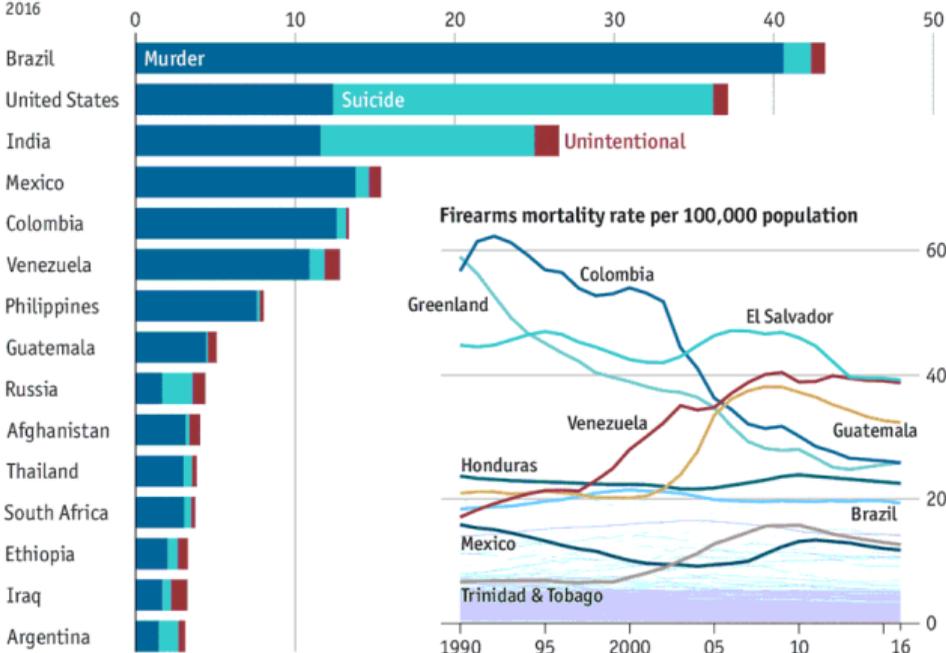
**Contínua:** os valores pertencem à um intervalo:

- Pressão arterial;
- Tempo de acesso à plataforma;
- Gasto com alimentação.

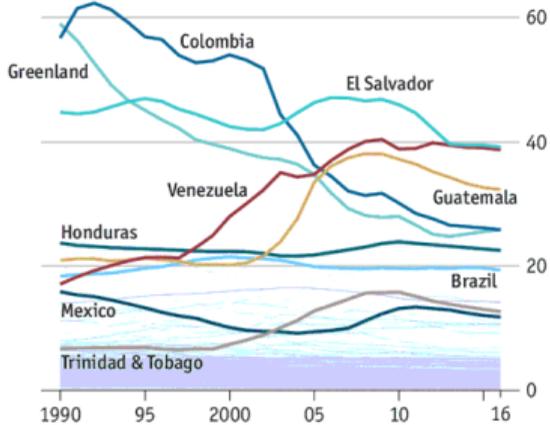
# Exemplo: mortes com armas de fogo

## Shooting range

Firearms deaths by incident type, '000



Firearms mortality rate per 100,000 population



Source: "Global mortality from firearms, 1990-2016" by Mohsen Naghavi et al., JAMA, 2018

economist.com

# Qualitativa e quantitativa andam juntas

- 1) A variável idade, medida em anos completos, é **quantitativa (contínua)**.



- Mas, se for informada apenas a faixa etária (0 a 5 anos, 6 a 10 anos, etc...), é **qualitativa (ordinal)**.

# Qualitativa e quantitativa andam juntas

2) O peso dos lutadores de boxe é uma variável **quantitativa (contínua)**.



- Mas, se o classificarmos por categorias (peso-pena, peso-leve, etc.) será **qualitativa (ordinal)**.

# Qualitativa e quantitativa andam juntas

3) Qual tipo de cimento escolher? **Isso depende da dilatação, resistência, rendimento, etc.**



# Qualitativa e quantitativa andam juntas

- 4) Quem é o melhor aluno? **Isso depende das notas nas avaliações, participação, etc.**



# Qualitativa e quantitativa andam juntas

- 5) Qual a rota mais adequada? **Isso depende da distância, segurança, trânsito, etc.**



# Exemplo: informações socioeconômicas dos funcionários

**Tabela:** informações socioeconômicas de 20 funcionários da empresa.

| ID | Estado civil | Grau de instrução  | Nº de filhos | Salário* | Idade |       | Região de procedência |
|----|--------------|--------------------|--------------|----------|-------|-------|-----------------------|
|    |              |                    |              |          | anos  | meses |                       |
| 1  | solteiro     | ensino fundamental | -            | 4,00     | 26    | 03    | interior              |
| 2  | casado       | ensino fundamental | 1            | 4,56     | 32    | 10    | interior              |
| 3  | casado       | ensino fundamental | 2            | 5,25     | 36    | 05    | capital               |
| 4  | solteiro     | ensino médio       | -            | 5,73     | 20    | 10    | capital               |
| 5  | solteiro     | superior           | 3            | 6,26     | 40    | 07    | interior              |
| 6  | casado       | ensino fundamental | -            | 6,66     | 28    | 00    | outra                 |
| 7  | solteiro     | ensino fundamental | -            | 6,86     | 41    | 00    | interior              |
| 8  | solteiro     | ensino fundamental | 2            | 7,39     | 43    | 04    | capital               |
| 9  | casado       | ensino médio       | 1            | 7,59     | 34    | 10    | outra                 |
| 10 | solteiro     | ensino médio       | 1            | 7,44     | 23    | 06    | interior              |
| 11 | casado       | ensino médio       | 2            | 8,12     | 33    | 10    | interior              |
| 12 | solteiro     | ensino fundamental | 5            | 8,46     | 27    | 06    | interior              |
| 13 | solteiro     | ensino médio       | 3            | 8,74     | 37    | 11    | interior              |
| 14 | casado       | ensino médio       | 4            | 8,95     | 44    | 10    | interior              |
| 15 | casado       | ensino médio       | -            | 9,13     | 30    | 11    | capital               |
| 16 | solteiro     | superior           | 4            | 9,35     | 38    | 06    | capital               |
| 17 | casado       | superior           | 4            | 9,77     | 31    | 10    | interior              |
| 18 | casado       | superior           | -            | 9,80     | 39    | 08    | outra                 |
| 19 | solteiro     | ensino fundamental | 3            | 10,53    | 25    | 10    | interior              |
| 20 | solteiro     | ensino médio       | 2            | 10,76    | 37    | 08    | capital               |

\* número de salários mínimos

**Fonte:** dados hipotéticos

# Referências

- ▶ Bussab, WO; Morettin, PA. Estatística Básica. São Paulo: Editora Saraiva, 2006 (5ª Edição).
- ▶ Magalhães, MN; Lima, ACP. Noções de Probabilidade e Estatística. São Paulo: EDUSP, 2008.

