

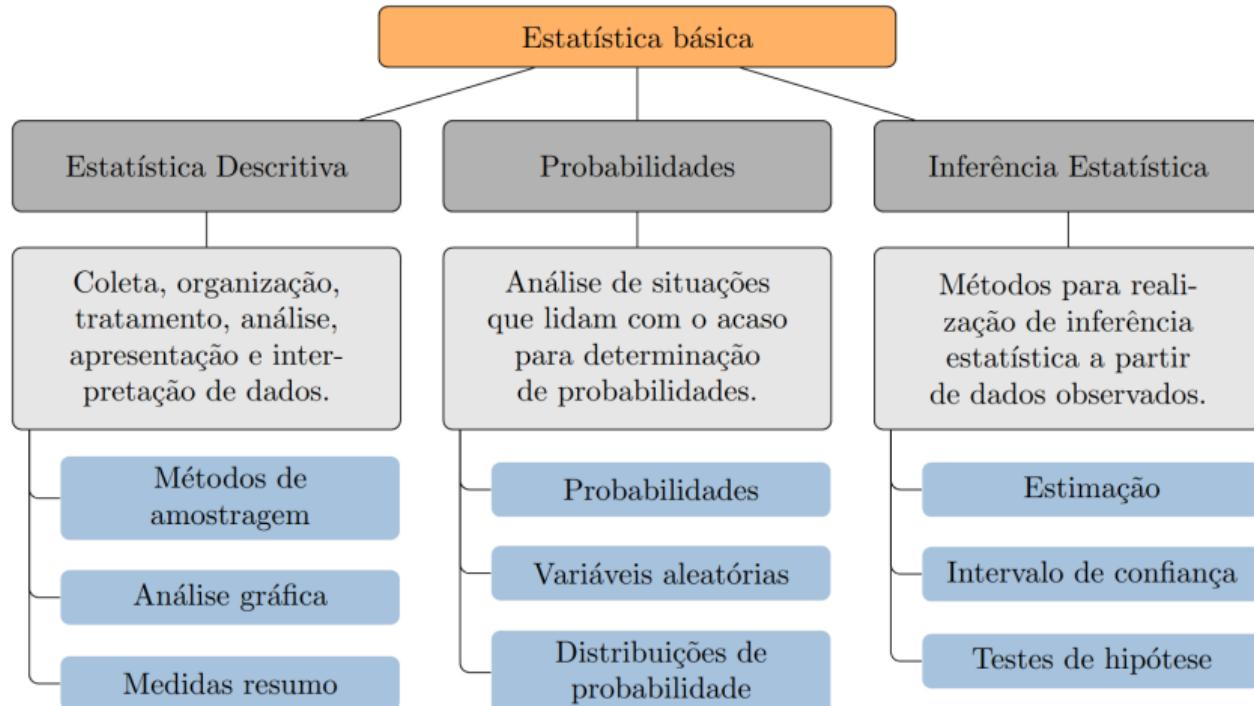
Análise exploratória de dados

Parte 1

Prof.: Eduardo Vargas Ferreira

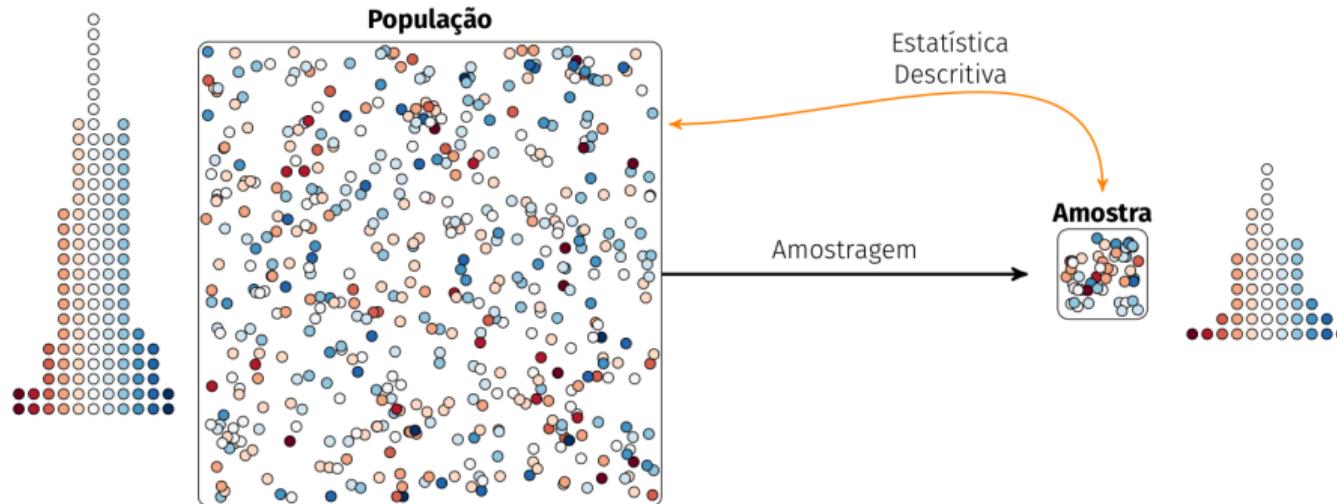


Resumo

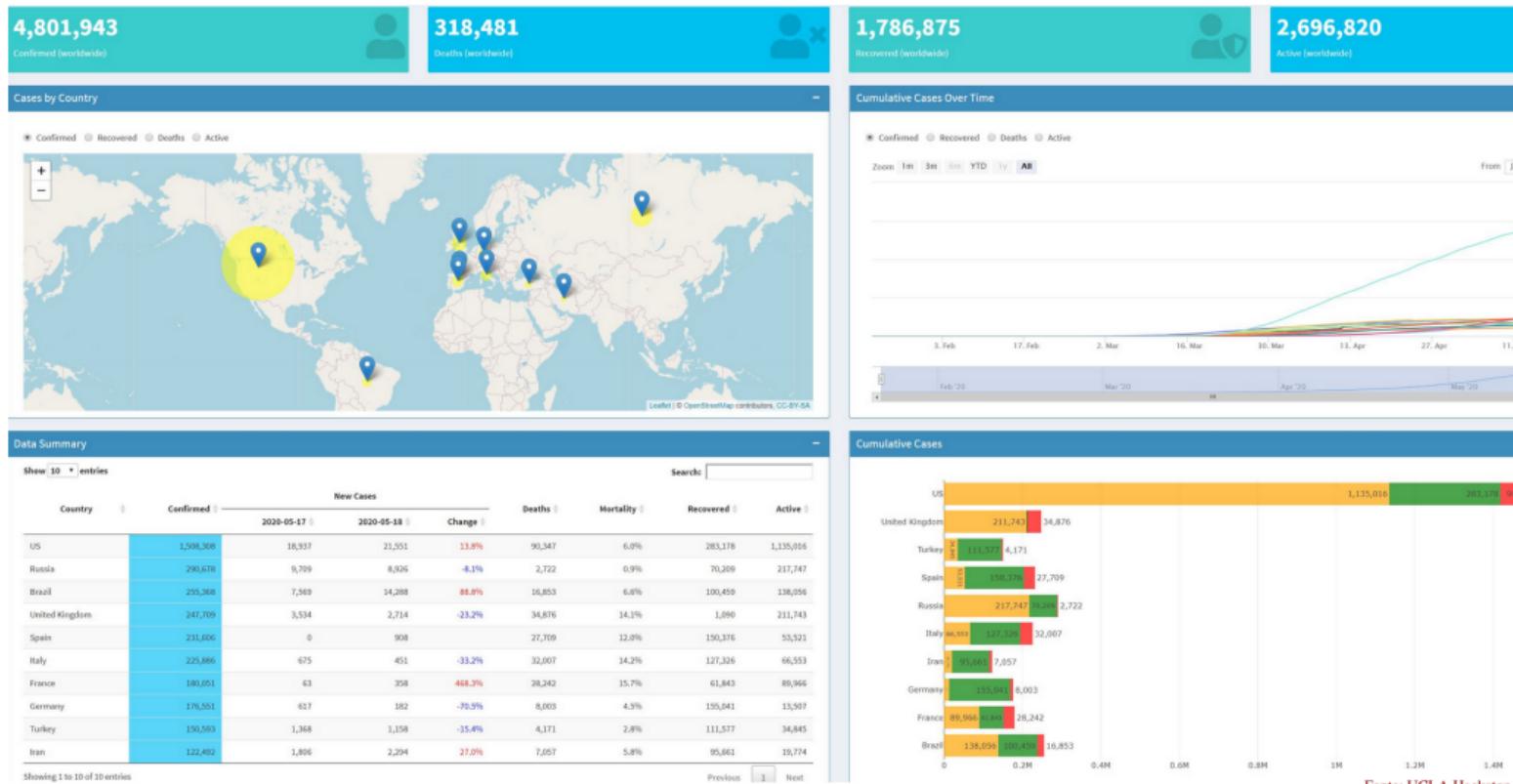


Papel da estatística descritiva

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



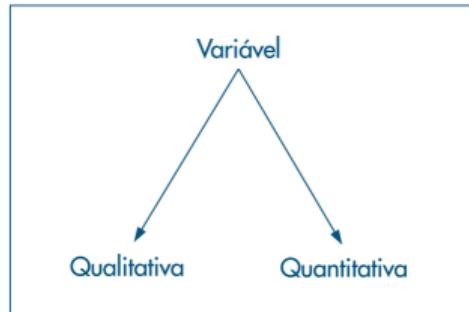
Motivação



O que é uma variável

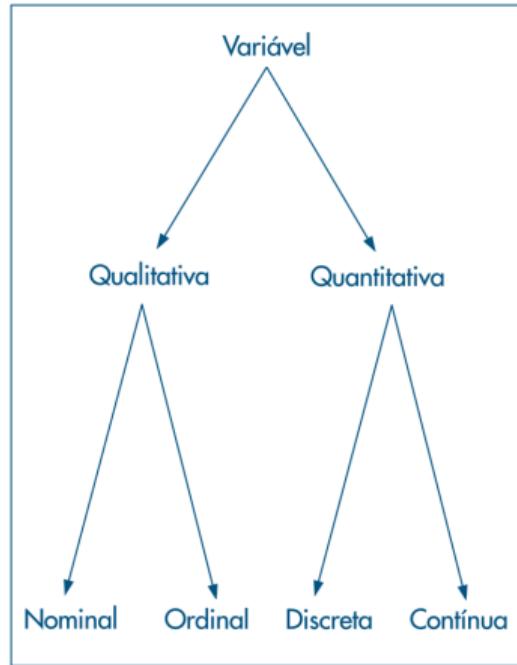
- **Variável:** é uma característica dos elementos de uma população ou de uma amostra, que pode assumir diferentes valores de interesse ao estudo.

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



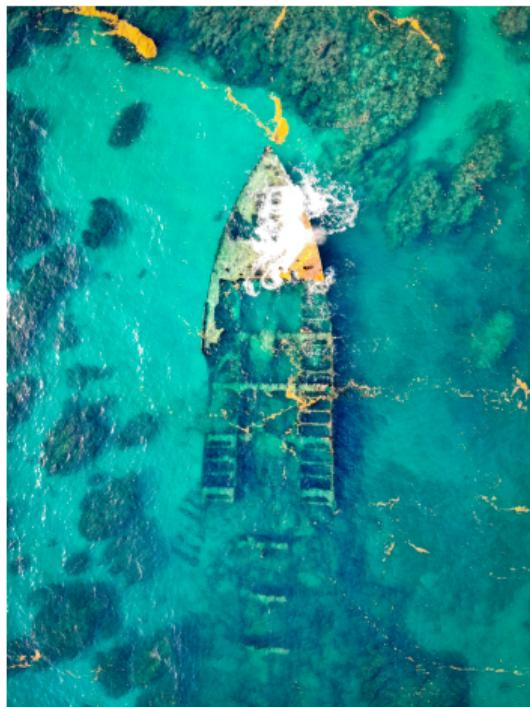
- Algumas variáveis apresentam uma qualidade:
- Região de procedência;
 - Estado civil;
 - Grau de instrução.
- Outras resultam de uma contagem ou mensuração:
- Número de filhos;
 - Salário;
 - Idade.

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.

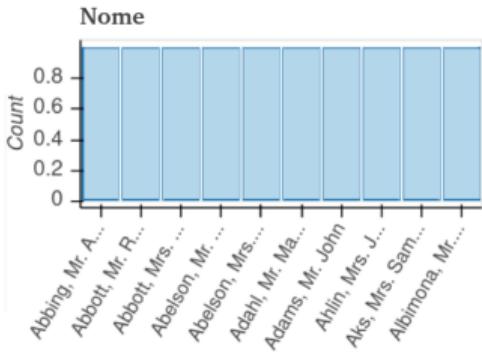
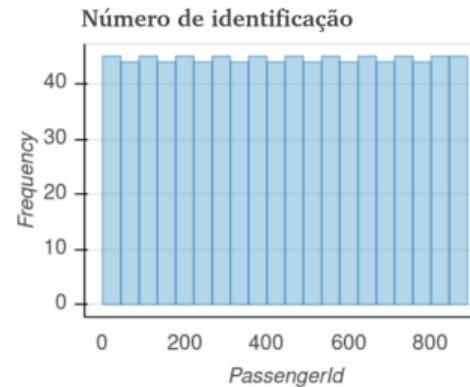
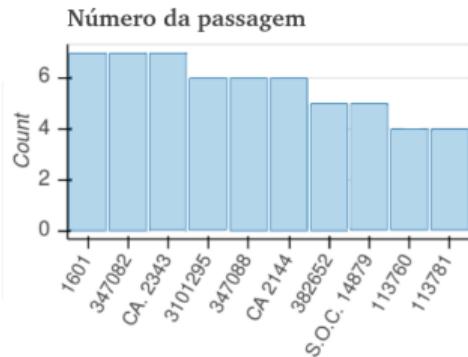
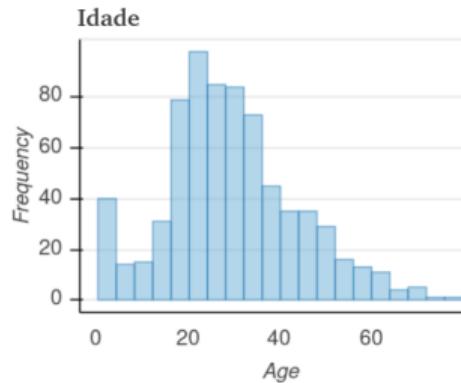
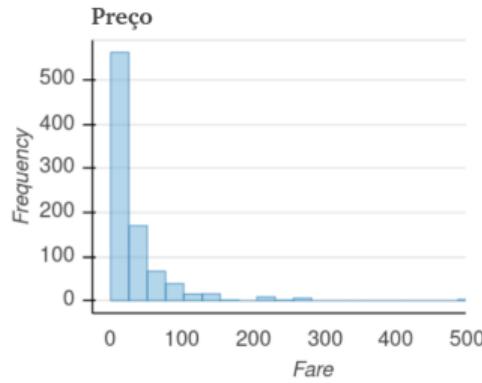
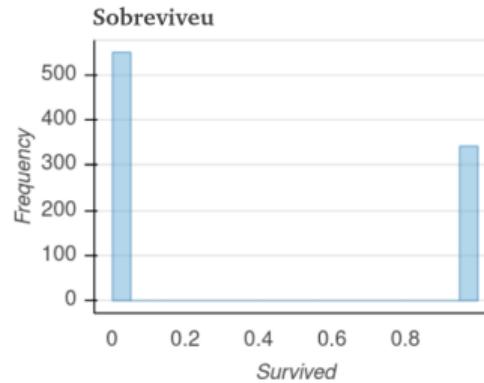
Exemplo: investigando o naufrágio do Titanic



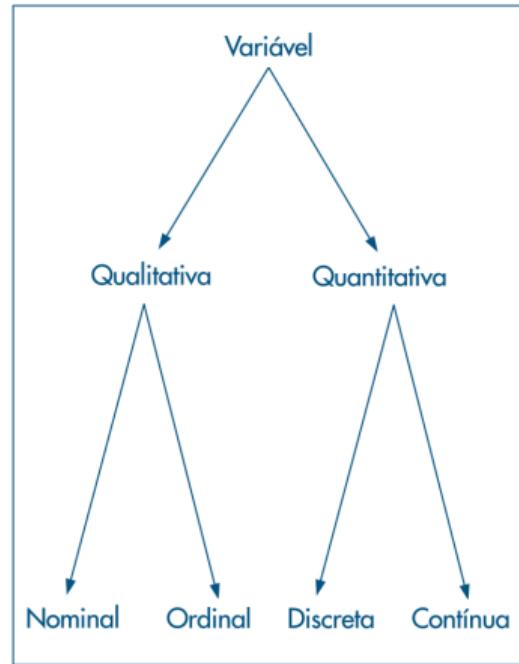
Dataset Statistics

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.

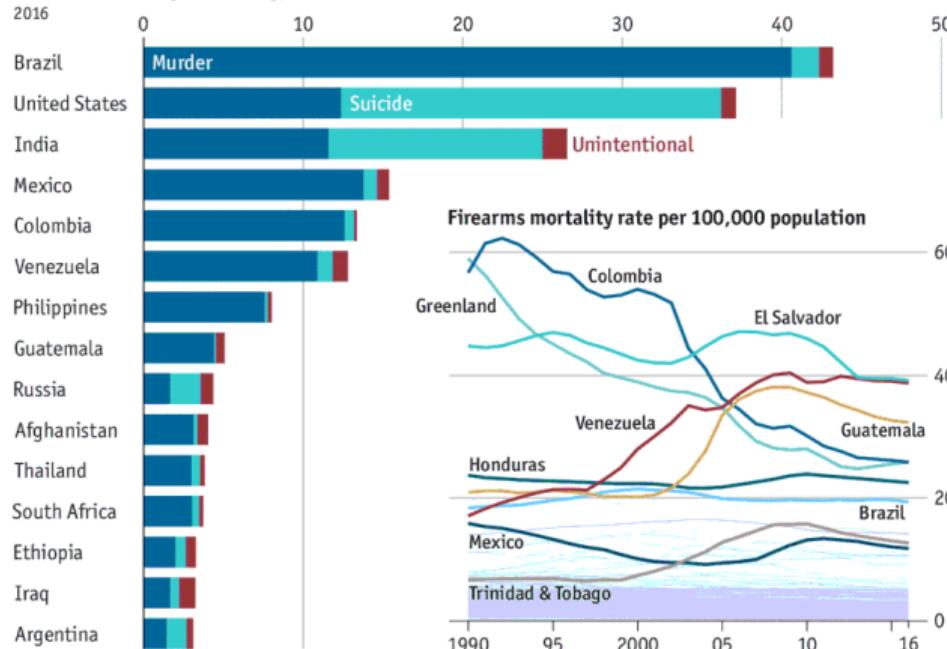
Contínua: os valores pertencem à um intervalo:

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

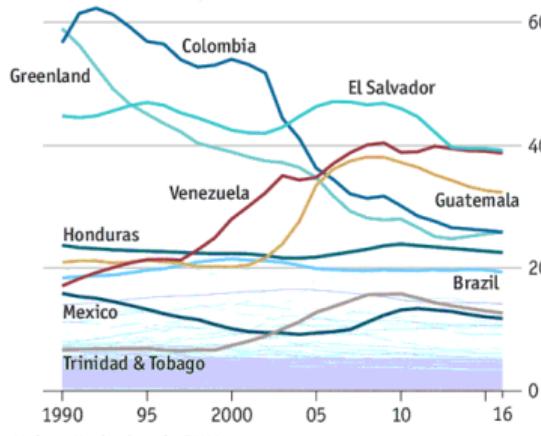
Exemplo: mortes com armas de fogo

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 socioeconómicas 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 (5^a Edição).
- Magalhães, MN; Lima, ACP. Noções de Probabilidade e Estatística. São Paulo: EDUSP, 2008.

