

# Introdução à Estatística

## Estrutura e tipos de dados

---

Fabio Cop ([fabiocopf@gmail.com](mailto:fabiocopf@gmail.com))

Instituto do Mar - UNIFESP

Última atualização em 18 de março de 2022



# Conteúdo da aula

---

1. Os dados por trás das imagens
  2. Tabulação de dados: alguns exemplos de dados estruturados
  3. Unidades amostrais e descritores (ou atributos)
  4. Tipos de variáveis e níveis de mensuração
-

# Os dados por trás das imagens

---

---

# Tabulação de dados: alguns exemplos de dados estruturados

## Programa das Nações Unidas para o Desenvolvimento (PNUD) - Brasil

IBGE	UF	Município	Região	População 2010	Porte	Capital	IDHM 2010	IDHM Renda 2010	IDHM Longevidade 2010	IDHM Educação 2010
292200	BA	Mucuri	Região Nordeste	36026	Pequeno II	NA	0.665	0.679	0.801	0.541
250053	PB	Alcantil	Região Nordeste	5239	Pequeno I	NA	0.578	0.550	0.765	0.458
261170	PE	Riacho Das Almas	Região Nordeste	19162	Pequeno I	NA	0.570	0.586	0.757	0.418
171855	TO	Riachinho	Região Norte	4191	Pequeno I	NA	0.572	0.503	0.760	0.490
314570	MG	Oliveira Fortes	Região Sudeste	2123	Pequeno I	NA	0.635	0.616	0.794	0.523
315790	MG	Santa Margarida	Região Sudeste	15011	Pequeno I	NA	0.610	0.635	0.785	0.455
350620	SP	Bento de Abreu	Região Sudeste	2674	Pequeno I	NA	0.744	0.706	0.820	0.712
351480	SP	Eldorado	Região Sudeste	14641	Pequeno I	NA	0.691	0.633	0.847	0.615
351565	SP	Fernão	Região Sudeste	1563	Pequeno I	NA	0.703	0.668	0.843	0.617
353990	SP	Poloni	Região Sudeste	5395	Pequeno I	NA	0.766	0.736	0.853	0.717
430820	RS	Flores da Cunha	Região Sul	27126	Pequeno II	NA	0.754	0.786	0.849	0.642
420190	SC	Aurora	Região Sul	5549	Pequeno I	NA	0.733	0.748	0.870	0.606
420950	SC	Laurentino	Região Sul	6004	Pequeno I	NA	0.749	0.742	0.855	0.662

# Tabulação de dados: alguns exemplos de dados estruturados

## Floresta Experimental de **Hubbard Brook (HBEF)** - EUA

Year	Treatment	Flow	Precipitation
1987	Deforested	790.47	1222.1
1977	Reference	843.75	1388.2
1967	Reference	780.76	1296.8
1962	Reference	699.29	1230.6
1982	Deforested	802.73	1088.2
1978	Reference	613.79	1085.7
1977	Deforested	966.25	1382.7
1985	Deforested	634.76	1128.4
1963	Reference	662.58	1151.7
1985	Reference	627.84	1137.2
1984	Reference	970.65	1403.5
1967	Deforested	1131.85	1315.1
1962	Deforested	777.31	1232.2



fonte: <https://hubbardbrook.org/>

# Tabulação de dados: alguns exemplos de dados estruturados

Passageiros do Titanic

Passageiro	Sobervivente	Classe	Nome	Sexo	Idade	SibSp	Parch	Ticket	Fare	Cabin	Embarked
836	1	1	Compton, Miss. Sara Rebecca	female	39	1	1	PC 17756	83.1583	E49	C
679	0	3	Goodwin, Mrs. Frederick (Augusta Tyler)	female	43	1	6	CA 2144	46.9000	NA	S
129	1	3	Peter, Miss. Anna	female	NA	1	1	2668	22.3583	F E69	C
509	0	3	Olsen, Mr. Henry Margido	male	28	0	0	C 4001	22525.0000	NA	S
471	0	3	Keefe, Mr. Arthur	male	NA	0	0	323592	7.2500	NA	S
299	1	1	Saalfeld, Mr. Adolphe	male	NA	0	0	19988	30.5000	C106	S
270	1	1	Bissette, Miss. Amelia	female	35	0	0	PC 17760	135.6333	C99	S
187	1	3	O'Brien, Mrs. Thomas (Johanna "Hannah" Godfrey)	female	NA	1	0	370365	15.5000	NA	Q
307	1	1	Fleming, Miss. Margaret	female	NA	0	0	17421	110.8833	NA	C
597	1	2	Leitch, Miss. Jessie Wills	female	NA	0	0	248727	33.0000	NA	S
277	0	3	Lindblom, Miss. Augusta Charlotta	female	45	0	0	347073	7.7500	NA	S
874	0	3	Vander Cruyssen, Mr. Victor	male	47	0	0	345765	9.0000	NA	S

fonte: <https://www.kaggle.com/c/titanic/>

# Tabulação de dados: alguns exemplos de dados estruturados

Penguins do Arquipélago de Palmer, Antartica																
studyName	Sample Number	Species	Region	Island	Stage	Individual ID	Clutch Completion	Date Egg	Culmen Length (mm)	Culmen Depth (mm)	Flipper Length (mm)	Body Mass (g)	Sex	Delta 15 N (o/oo)	Delta 13 C (o/oo)	Comments
PAL0910	54	Chinstrap penguin (Pygoscelis antarctica)	Anvers	Dream	Adult, 1 Egg Stage	N92A2	Yes	2009-11-23	50.7	19.7	203	4050	MALE	9.93727	-24.59066	NA
PAL0910	111	Gentoo penguin (Pygoscelis papua)	Anvers	Biscoe	Adult, 1 Egg Stage	N32A1	Yes	2009-11-20	50.5	15.2	216	5000	FEMALE	8.24691	-26.18466	NA
PAL0910	53	Chinstrap penguin (Pygoscelis antarctica)	Anvers	Dream	Adult, 1 Egg Stage	N92A1	Yes	2009-11-23	45.7	17.3	193	3600	FEMALE	9.41500	-24.80500	NA
PAL0809	79	Adelie Penguin (Pygoscelis adeliae)	Anvers	Torgersen	Adult, 1 Egg Stage	N38A1	Yes	2008-11-09	36.2	16.1	187	3550	FEMALE	7.82381	-26.51382	NA
PAL0809	61	Gentoo penguin (Pygoscelis papua)	Anvers	Biscoe	Adult, 1 Egg Stage	N18A1	Yes	2008-11-03	45.3	13.8	208	4200	FEMALE	8.65015	-26.79053	NA

fonte: <https://allisonhorst.github.io/palmerpenguins/>



# Unidades amostrais e descritores (ou atributos)

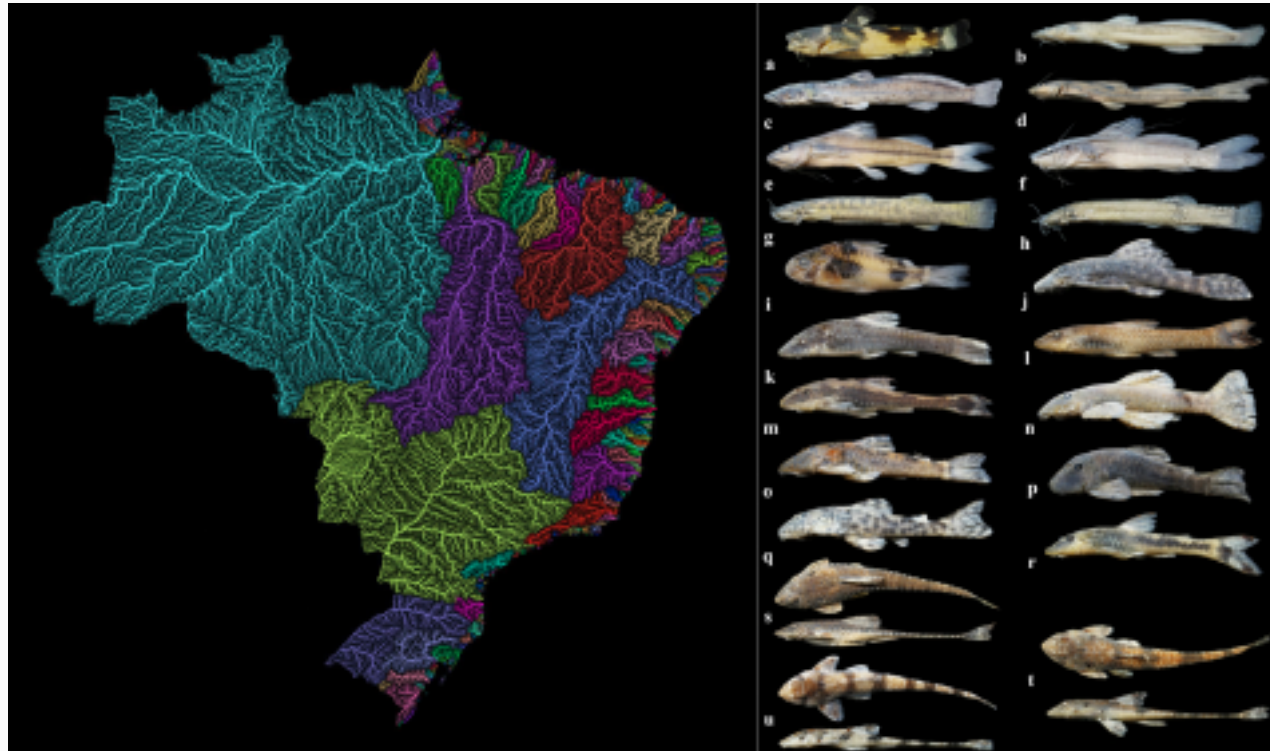
Em conjuntos de dados estruturados, cada linha representa uma **unidade amostral** e cada coluna representa uma **variável** que descreve uma característica da unidade, i.e. um atributo desta unidade.

ID	Descritor 1	Descritor 2	Descritor 3	Descritor 4	Descritor 5	Descritor 6	Descritor 7
Unidade amostral 1							
Unidade amostral 2							
Unidade amostral 3							
Unidade amostral 4							
Unidade amostral 5							
Unidade amostral 6							
Unidade amostral 7							
Unidade amostral 8							
Unidade amostral 9							
Unidade amostral 10							

# Tipos de variáveis e níveis de mensuração

## Variáveis Qualitativas

- *Categorias não-ordenadas*: Refere-se a atributos que **não-podem** ser ordenados, por exemplo a nomenclatura das bacias hidrográficas brasileiras, ou grupos taxônicos.



# Tipos de variáveis e níveis de mensuração

## Variáveis Qualitativas

- *Categorias ordenadas*: Refere-se a atributos que podem ser ordenados, embora **não seja possível** quantificar a magnitude das diferenças entre níveis. Considere por exemplo, a classificação do estágio de regeneração de uma área florestal.



# Tipos de variáveis e níveis de mensuração

## Variáveis Quantitativas

- *Discretas*: Variáveis **enumeráveis** como a contagem de barcos que saem para pescar em um determinado dia, o número de peixes de um cardume.





# Tipos de variáveis e níveis de mensuração

## Variáveis Quantitativas

- *Contínuas*: Variáveis **não enumeráveis**, i.e. podem assumir infinitos valores, sem intervalos vazios. Considere a vazão em  $m^3/seg$  que verte de uma cachoeira, o volume de chuva em um determinado dia, altura da maré ou a velocidade do vento. O limite de precisão para estas medidas é somente aquele que podemos mensurar com os aparelhos disponíveis.



# Tipos de variáveis e níveis de mensuração

Quais informações as variáveis carregam? Quais ações são possíveis sobre essas variáveis?

Tipos de variáveis	Níveis de mensuração	Agrupar itens	Ordenar itens	Gerar intervalos proporcionais	Partir de um ponto de origem	Exemplos
Qualitativa não-ordenada	Nominal					Grupos taxônicos
Qualitativa ordenada	Ordinal					Estágio de regeneração ambiental
Quantitativa Discreta/Contínua	Intervalar					Temperatura, Dados padronizados
Quantitativa Discreta/Contínua	Razão					Volume de chuva, Número de ovos