



# Introdução à Estatística

## Estrutura e tipos de dados

---

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

Instituto do Mar - UNIFESP

Última atualização em 05 de outubro de 2021

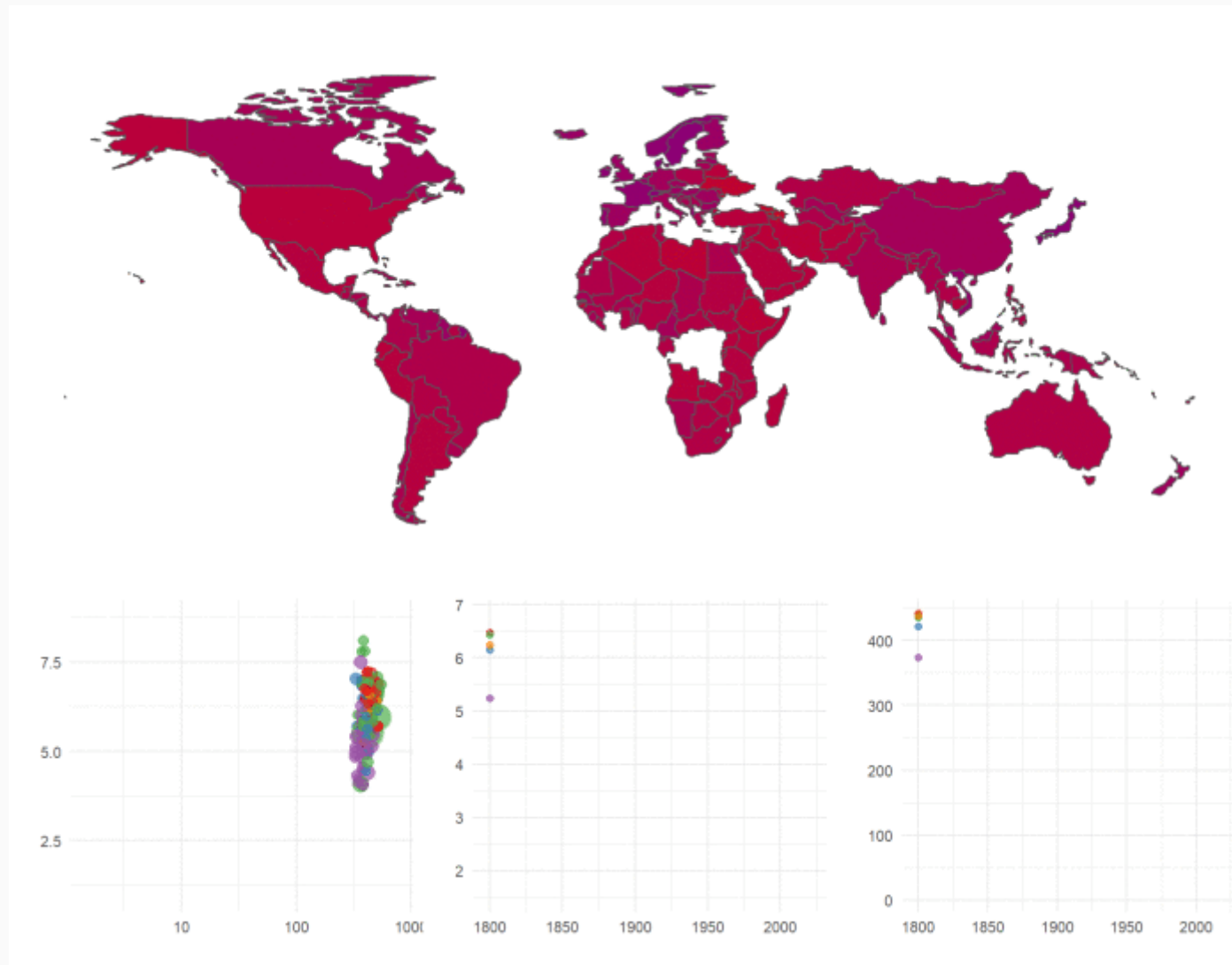


# Conteúdo da aula

---

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
510310	MT	Cocalinho	Região Centro-Oeste	5490	Pequeno I	NA	0.660	0.674	0.807	0.528
230930	CE	Nova Russas	Região Nordeste	30965	Pequeno II	NA	0.614	0.585	0.754	0.524
250300	PB	Caaporã	Região Nordeste	20362	Pequeno II	NA	0.602	0.565	0.757	0.509
250435	PB	Caturité	Região Nordeste	4543	Pequeno I	NA	0.623	0.617	0.782	0.502
251700	PB	Umbuzeiro	Região Nordeste	9298	Pequeno I	NA	0.584	0.545	0.712	0.514
240320	RN	Doutor Severiano	Região Nordeste	6492	Pequeno I	NA	0.621	0.567	0.747	0.566
320480	ES	São José do Calçado	Região Sudeste	10408	Pequeno I	NA	0.688	0.669	0.814	0.599
313620	MG	João Monlevade	Região Sudeste	73610	Médio	NA	0.758	0.724	0.849	0.708
350280	SP	Araçatuba	Região Sudeste	181579	Grande	NA	0.788	0.782	0.841	0.744
350730	SP	Boracéia	Região Sudeste	4268	Pequeno I	NA	0.754	0.723	0.866	0.685
351890	SP	Guzolândia	Região Sudeste	4754	Pequeno I	NA	0.697	0.676	0.808	0.621
430865	RS	Garruchos	Região Sul	3234	Pequeno I	NA	0.671	0.657	0.833	0.553
421775	SC	Sul Brasil	Região Sul	2766	Pequeno I	NA	0.707	0.713	0.832	0.595

fonte: <https://www.br.undp.org>

# Tabulação de dados: alguns exemplos de dados estruturados

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

Year	Treatment	Flow	Precipitation
1976	Deforested	1142.59	1511.4
1984	Deforested	1000.54	1396.8
1984	Reference	970.65	1403.5
1959	Reference	918.23	1479.1
1980	Reference	548.28	1101.1
1981	Deforested	1129.09	1631.5
1959	Deforested	1012.05	1482.6
1973	Deforested	1585.73	1804.0
1962	Reference	699.29	1230.6
1971	Reference	676.19	1173.4
1963	Deforested	773.64	1138.6
1966	Deforested	1189.34	1222.3
1985	Deforested	634.76	1128.4



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)

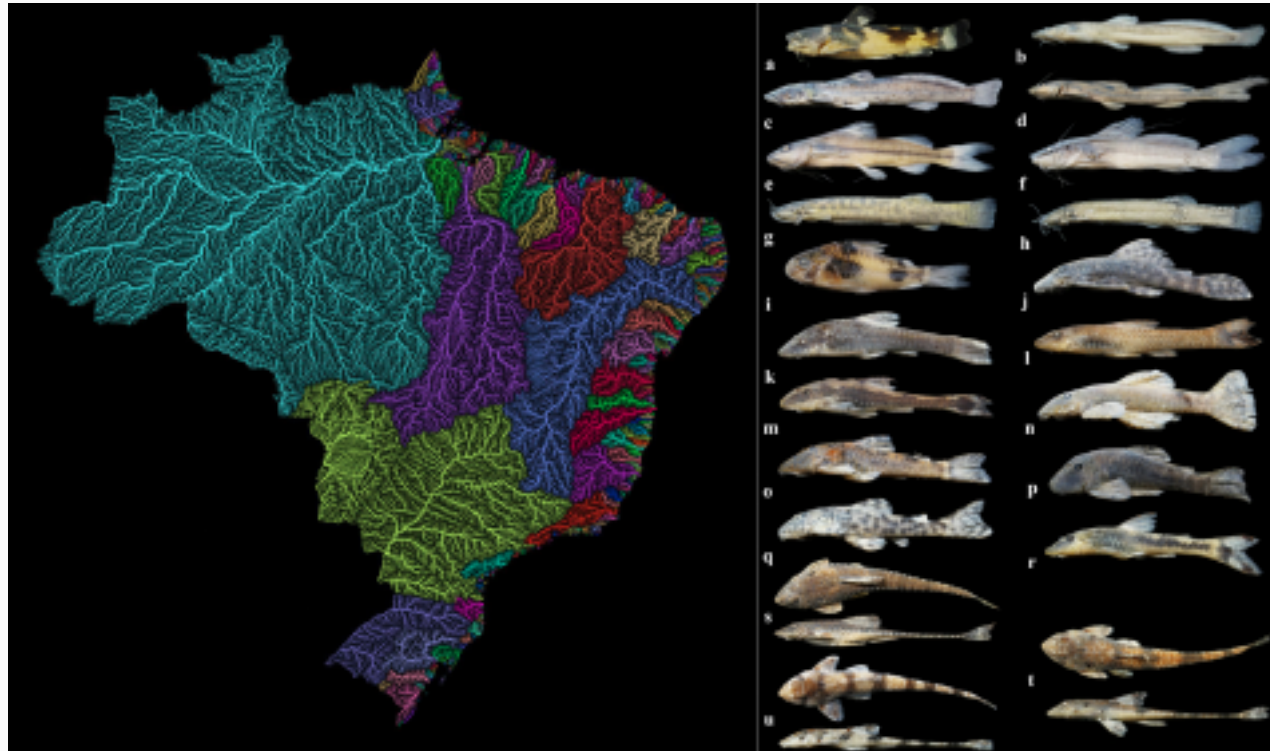
Em conjuntos de dados estruturados, cada linha representa uma **unidade amostral** e cada coluna representa uma **variável** que descreve uma característica da unidade, i.e. um atributo desta unidade.

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