Análise Exploratória dos Dados

Danilo Sipoli Sanches

Departamento Acadêmico de Computação
Universidade Tecnológica Federal do Paraná
Cornélio Procópio

Resumo

1. Instâncias de dados, atributos e objetos;
2. Tipos de dados;
3. Qualidade dos dados;
4. Descrição estatística dos dados.

Instância de Dados e Atributos

Id.	Nome	Idade	Sexo	Peso	Manchas	Temp.	# Int.	Est.	Diagnóstico
4201	João	28	M	79	Concentradas	38,0	2	SP	Doente
3217	Maria	18	F	67	Inexistentes	39,5	4	MG	Doente
4039	Luiz	49	M	92	Espalhadas	38,0	2	RS	Saudável
1920	José	18	M	43	Inexistentes	38,5	8	MG	Doente
4340	Cláudia	21	F	52	Uniformes	37,6	1	PE	Saudável
2301	Ana	22	F	72	Inexistentes	38,0	3	RJ	Doente
1322	Marta	19	F	87	Espalhadas	39,0	6	AM	Doente
3027	Paulo	34	M	67	Uniformes	38,4	2	GO	Saudável

Tabela 1: Conjunto de dados hospital com seus atributos.

Tipo de Dados

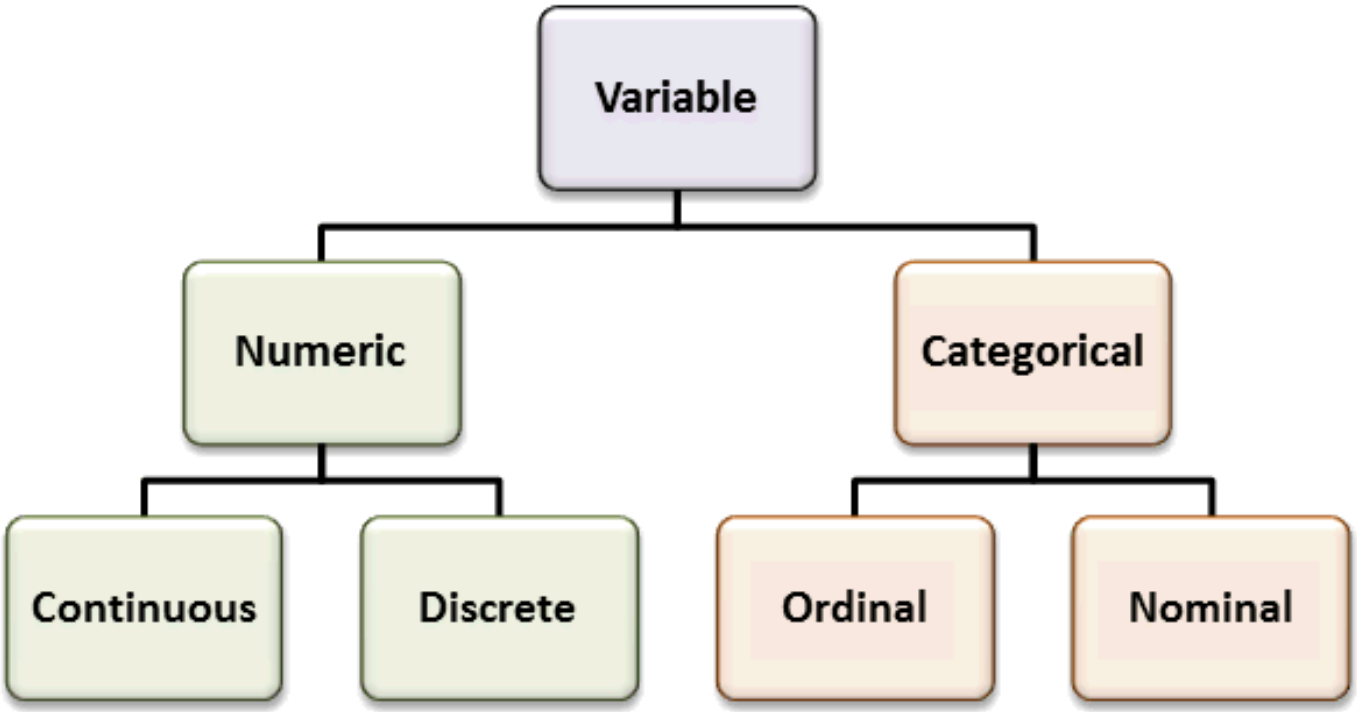


Figura: Estrutura dos Dados.

Tipo de Dados

O tipo define se o atributo representa quantidades, sendo então denominado quantitativo ou numérico, ou qualidades, quando é chamado qualitativo, simbólico ou categórico, pois os valores podem ser associados a categorias.

Exemplos:

- qualitativos: (pequeno, médio, grande) e (matemática, física, química)
- quantitativos: (23, 45, 12)

Atributos quantitativos

Os valores de um atributo quantitativo podem tanto ser ordenados quanto utilizados em operações aritméticas. Valores quantitativos podem ser contínuos e discretos.

Tipo de Dados

Classificação dos tipos de atributos da Tabela 1.

Atributo	Classificação
Id.	Qualitativo
Nome	Qualitativo
Idade	Quantitativo discreto
Sexo	Qualitativo
Peso	Quantitativo contínuo
Manchas	Qualitativo
Temp.	Quantitativo contínuo
#Int.	Quantitativo discreto
Est.	Qualitativo
Diagnóstico	Qualitativo

Figura: Tipo dos atributos do conjunto hospital.

Exploração dos Dados

Informações úteis podem ser extraída de um conjunto de dados.

Estatística Descritiva

Uma das formas mais simples de explorar um conjunto de dados é a extração de medidas da estatística descritiva.

- Frequência;
- Localização ou tendência central (ex: a média);
- Dispersão ou espalhamento (ex: desvio padrão);
- Distribuição.

Estatística Descritiva - Medidas

https://anovabr.github.io/mqt/estatistica-descritiva.html

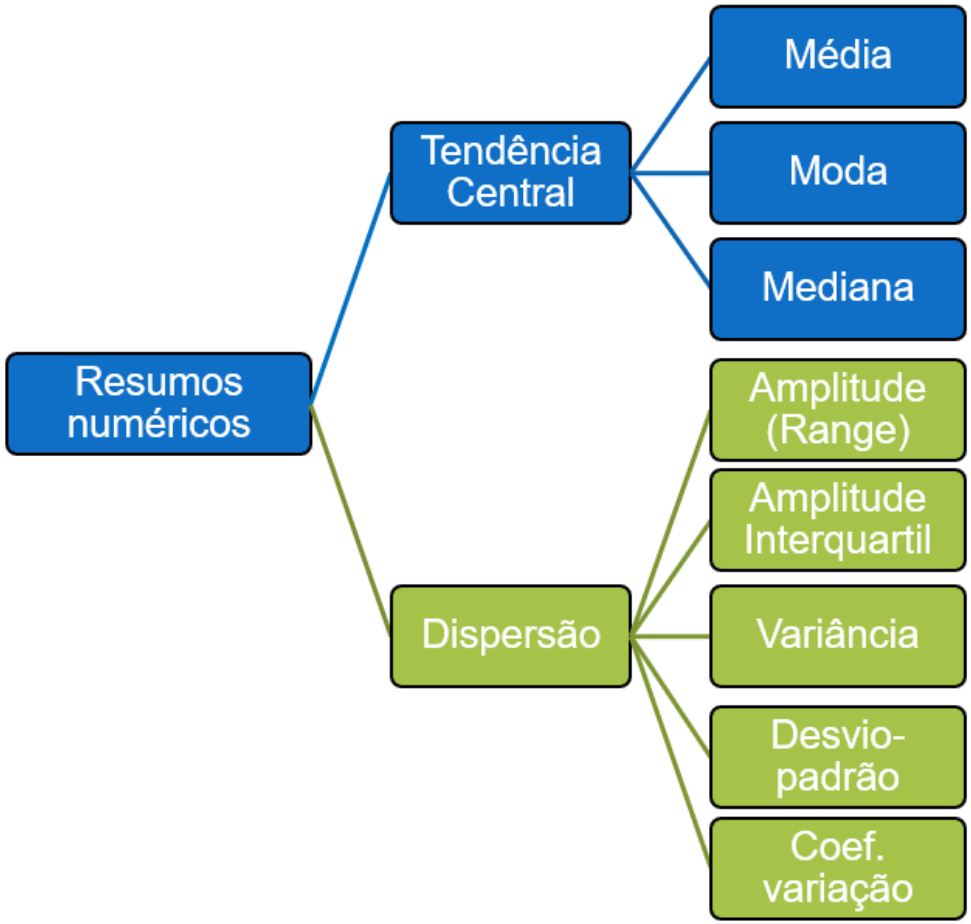
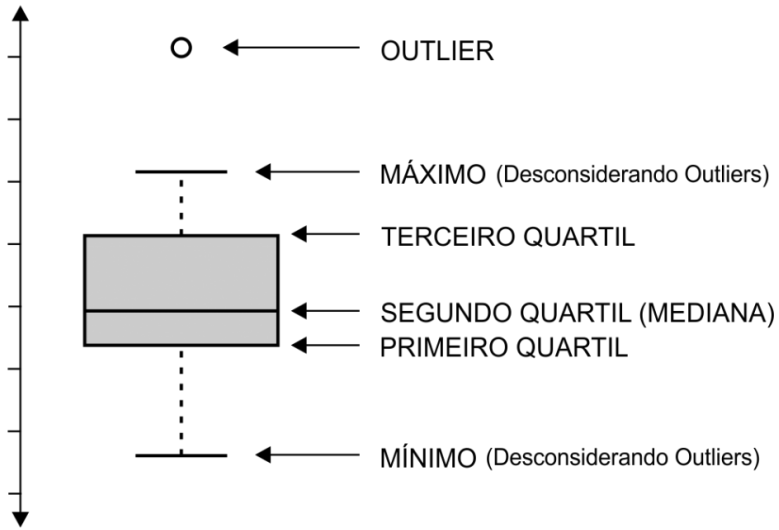


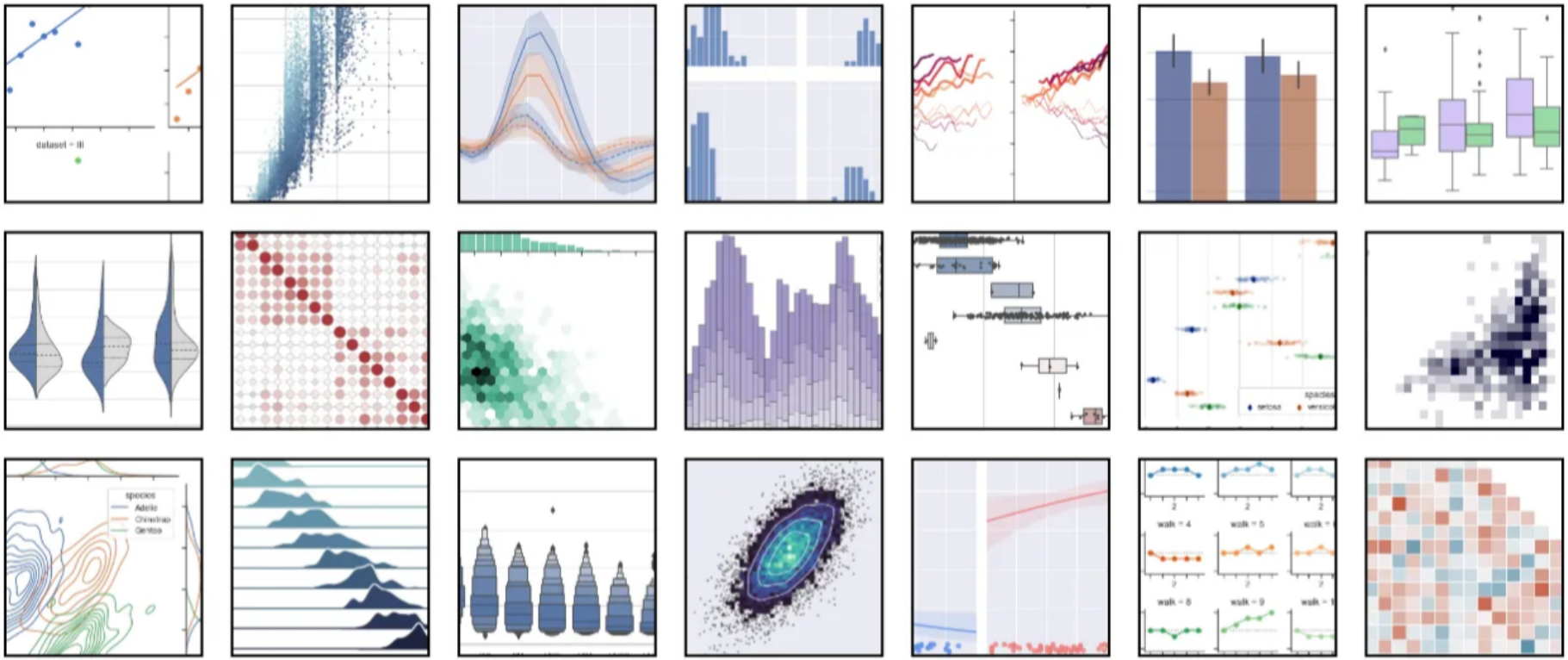
Gráfico de Caixas - Box Plot

- É uma ferramenta gráfica que permite visualizar e analisar a distribuição dos dados e suas variações através de quartis e seus limites;
- Ele permite a análise de dados por meio da presença de outliers (valores discrepantes de um conjunto de dados).



Bibliotecas gráficas em Python

fonte: https://kent.medium.com/



Matplotlib vs Seaborn

fonte:
dataexpertise.in/mastering-matplotlib-data-visualization/