FCT/Unesp – Presidente Prudente Departamento de Matemática e Computação

Fundamentos sobre Dados Parte 1

Prof. Danilo Medeiros Eler danilo.eler@unesp.br





Sumário

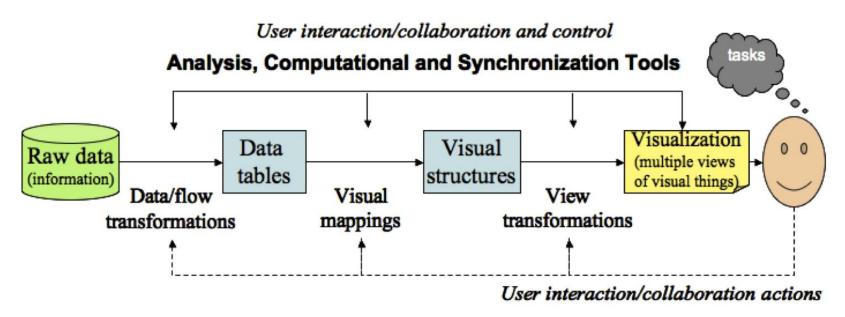
- Tipos de Dados
- Estrutura dentro e entre instâncias
- Processamento dos dados





Processo de Visualização

 Primeiramente, vamos retomar o pipeline de visualização utilizado pela maioria dos sistemas







- De uma forma simples, cada observação ou variável de uma instância de dado representa uma única informação
- Essas informações podem ser categorizadas como sendo ordinal (numérica) ou nominal (não numérico)
 - Elas podem se dividir em subcategorias





- Ordinal
 - Dados com valores numéricos
 - Subcategorias
 - Binários: assume dois valores (e.g., 0 e 1)
 - Discreto: um único valor inteiro ou um subconjunto específico (e.g., 2, 4, 6)
 - Contínuo: representa valores reais





Ordinal

Contínuo, discreto e binário

Identificador	Salário	Idade	Peso	Nível	Aprovado
P1	1500,89	30	87,6	1	0
P2	789,52	48	74,6	2	0
P3	1000,00	28	70,5	2	1
P4	589,36	39	90,5	3	1





Nominal

- Dados com valores não numéricos
- Subcategorias
 - Categórico: um valor de um conjunto finito de possibilidades (e.g., vermelho, verde, azul)
 - Ranqueado: uma variável categórica com alguma ordem (e.g., pequeno, médio, grande)
 - Arbitrário: uma variável com infinitas possibilidades de valores, sem uma relação de ordem específica (e.g., endereços)





Nominal

Arbitrário, categórico, ranqueado

ID	Nome	Olhos	Camiseta	Escolaridade
P1	José da Silva	Azul	Р	Graduado
P2	Augusto dos Santos	Castanho	Р	Ensino Médio
P3	Roberto Carlos	Verde	G	Graduado
P4	André Silveira	Castanho	M	Especialista
P5	Rafaela Andrade	Preto	GG	Graduado





- Uma outra maneira de categorizar os dados é utilizando o conceito matemático de escala
- Nesse caso, os atributos que definem as variáveis podem ser categorizados como
 - Relação de Ordem
 - Métrica de Distância
 - Existência de Zero Absoluto





- Relação de ordem: permite que os dados possam ser ordenados
 - Por definição, todas ordinais e variáveis nominais ranqueadas

Identificador	Salário	ldade	Peso	Nível	Aprovado
P1	1500,89	30	87,6	1	0
P2	789,52	48	74,6	2	0
P3	1000,00	28	70,5	2	1
P4	589,36	39	90,5	3	1





- Relação de ordem: permite que os dados possam ser ordenados
 - Por definição, todas ordinais e variáveis nominais ranqueadas

ID	Nome	Olhos	Camiseta	Escolaridade
P1	José da Silva	Azul	Р	Graduado
P2	Augusto dos Santos	Castanho	Р	Ensino Médio
P3	Roberto Carlos	Verde	G	Graduado
P4	André Silveira	Castanho	M	Especialista
P5	Rafaela Andrade	Preto	GG	Graduado





- Métrica de distância: permite calcular distâncias entre os registros (instâncias)
 - Está presente em todas variáveis ordinais, mas geralmente não é encontrada nas variáveis nominais

Identificador	Salário	Idade	Peso	Nível	Aprovado
P1	1500,89	30	87,6	1	0
P2	789,52	48	74,6	2	0
P3	1000,00	28	70,5	2	1
P4	589,36	39	90,5	3	1





- Métrica de distância: permite calcular distâncias entre os registros (instâncias)
 - Está presente em todas variáveis ordinais, mas geralmente não é encontrada nas variáveis nominais

ID	Nome	Olhos	Camiseta	Escolaridade
P1	José da Silva	Azul	Р	Graduado
P2	Augusto dos Santos	Castanho	Р	Ensino Médio
P3	Roberto Carlos	Verde	G	Graduado
P4	André Silveira	Castanho	M	Especialista
P5	Rafaela Andrade	Preto	GG	Graduado





- Existência de zero absoluto: permite fixar um menor valor. É útil para diferenciar tipos de variáveis ordinais
 - O peso possui um zero absoluto; já o saldo bancário não possui





Referências

- Ward, M., Grinstein, G. G., Keim, D. Interactive data visualization foundations, techniques, and applications. Natick, Mass., A K Peters, 2010.
 - Capítulo 2





Referências

- Aulas de visualização da wiki.icmc.usp.br
 - Prof. Dr. Fernando Paulovich (ICMC/USP)
 - Profa. Dra. Maria Cristina Ferreira de Oliveira (ICMC/USP)
 - Profa. Dra. Rosane Minghim (ICMC/USP)



