

Visualização de dados

Unidade III

Cristiane Neri Nobre

Tabelas

Principais tipos de dados

Em um projeto de **ciência dos dados**, os dados coletados e produzidos podem ser organizados de diferentes maneiras:

- Tabela
- Gráficos

Tabelas

- Estruturas para organização e exibição de informação:
 - Dados organizados em **linhas e colunas**
 - Dados codificados como **texto**
- Apesar do uso de *grids*, o que caracteriza as tabelas é a organização dos dados em linhas e colunas
- Usadas há muito tempo

Quando usar tabelas?

- Documento será utilizado para **consulta** de valores **individuais**
- Será utilizada para **comparação** de valores individuais
- Valores **precisos** são necessários

Onde usamos tabelas?

- Relatórios
- Artigos
- Software

Organizando os dados

Como organizar os dados para que eles **contem uma história**?

- As subdivisões categóricas devem ser organizadas em **linhas ou colunas**?
- Os dados deveriam ser separados em **subgrupos** com quebras entre eles?
- Os dados deveriam ser **ordenados** de alguma forma particular?

O uso do espaço em branco

- A quantidade de espaços em branco pode ser manipulada para direcionamento dos olhos do usuário entre linhas e colunas (princípio da proximidade)
 - Se você deseja que o usuário avalie os dados coluna a coluna, privilegie o espaço entre colunas em detrimento do espaço entre linhas
 - Se você deseja que o usuário avalie os dados linha a linha, faça o oposto
- Se não há espaço suficiente, você deve usar outro método

Grids e linhas

- Podem ser usados para:
 - Delimitar linhas e colunas
 - Agrupar subconjuntos de dados
 - Ressaltar subconjuntos de dados

Grids e linhas

Product	Jan	Feb	Mar	Apr	May	Jun	Total
Product 01	93,993	84,773	88,833	95,838	93,874	83,994	541,305
Product 02	87,413	78,839	82,615	89,129	87,303	78,114	503,414
Product 03	90,036	81,204	85,093	91,803	89,922	80,458	518,516
Product 04	92,737	83,640	87,646	94,557	92,620	82,872	534,072
Product 05	83,733	75,520	79,137	85,377	83,627	74,826	482,220
Total	447,913	403,976	423,323	456,705	447,346	400,264	2,579,526

Product	Jan	Feb	Mar	Apr	May	Jun	Total
Product 01	93,993	84,773	88,833	95,838	93,874	83,994	541,305
Product 02	87,413	78,839	82,615	89,129	87,303	78,114	503,414
Product 03	90,036	81,204	85,093	91,803	89,922	80,458	518,516
Product 04	92,737	83,640	87,646	94,557	92,620	82,872	534,072
Product 05	83,733	75,520	79,137	85,377	83,627	74,826	482,220
Total	447,913	403,976	423,323	456,705	447,346	400,264	2,579,526

Grids e linhas

- Use com parcimônia as linhas de grades

Table 2: Range of parameter values.

Population initialization	Random
Representation	Binary
Crossover operator	Two Points
Crossover probability	70%, 75%, 80% and 85%
Mutation operator	One Point
Mutation probability	1%, 5% and 10%
Population size	100, 300 and 500
Number of generations	100, 300 and 500
Crossover selection method	Tournament (<i>size</i> = 2)
Composition of the new generation	Non-dominated individuals
Stopping criteria	Number of generations

Fonte: Do autor. Extraído de [1]

Cores e preenchimento

Quando espaço em branco não pode ser usado para delinear colunas ou linhas efetivamente, o uso de **cores suaves** para preenchimento do fundo é mais interessante que o uso de grids

Product	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Product 01	93,993	84,773	88,833	95,838	93,874	83,994	84,759	92,738	93,728	93,972	93,772	99,837
Product 02	87,413	78,839	82,615	89,129	87,303	78,114	78,826	86,246	87,167	87,394	87,208	92,848
Product 03	90,036	81,204	85,093	91,803	89,922	80,458	81,191	88,834	89,782	90,016	89,824	95,634
Product 04	92,737	83,640	87,646	94,557	92,620	82,872	83,626	91,499	92,476	92,716	92,519	98,503
Product 05	86,245	77,785	81,511	87,938	86,136	77,071	77,773	85,094	86,002	86,226	86,043	91,608
Product 06	88,833	80,119	83,956	90,576	88,720	79,383	80,106	87,647	88,582	88,813	88,624	94,356
Product 07	82,614	74,511	78,079	84,236	82,510	73,826	74,498	81,511	82,382	82,596	82,420	87,751
Product 08	85,093	76,746	80,421	86,763	84,985	76,041	76,733	83,957	84,853	85,074	84,893	90,384
Product 09	87,646	79,048	82,834	89,366	87,535	78,322	79,035	86,475	87,399	87,626	87,440	93,085
Product 10	90,275	81,420	85,319	92,047	90,161	80,672	81,406	89,070	90,021	90,255	90,063	95,888

Product	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Product 01	93,993	84,773	88,833	95,838	93,874	83,994	84,759	92,738	93,728	93,972	93,772	99,837
Product 02	557,378	502,704	526,780	568,319	556,673	498,084	502,621	549,936	555,807	557,254	556,068	592,033
Product 03	574,100	517,785	542,583	585,369	573,373	513,027	517,699	566,434	572,481	573,972	572,750	609,794
Product 04	304,273	274,426	287,569	310,246	303,888	271,904	274,381	300,210	303,415	304,205	303,557	323,191
Product 05	30,427	27,443	28,757	31,025	30,389	27,190	27,438	30,021	30,342	30,420	30,356	32,319
Product 06	305,186	275,249	288,432	311,176	304,799	272,720	275,204	301,111	304,325	305,118	304,468	324,161
Product 07	6,104	5,505	5,769	6,224	6,096	5,454	5,504	6,022	6,087	6,102	6,089	6,483
Product 08	61,220	55,215	57,859	62,422	61,143	54,708	55,206	60,403	61,048	61,207	61,076	65,027
Product 09	50,813	45,828	48,023	51,810	50,748	45,407	45,821	50,134	50,670	50,801	50,893	53,972
Product 10	52,337	47,203	49,464	53,365	52,271	46,770	47,196	51,638	52,190	52,326	52,214	55,591

Sequenciamento das colunas

Se há um relacionamento hierárquico entre conjuntos de subdivisões categóricas, as mesmas devem ser sequenciadas da esquerda para a direita

- Valores quantitativos derivados de outros valores quantitativos presentes na tabela, devem ser colocados à direita das colunas das quais derivam

Demographic questions

QN*	Question	N	%
1	Gender		
	Male	280	54.6
	Female	232	45.2
	Other	1	0.2
2	Age group		
	29 or less	344	67.1
	30 to 47	129	25.1
	48 or more	40	7.8
3	Education		
	Elementary or high school	53	10.3
	University graduate	340	66.3
	Masters or doctorate	120	23.4
4	Field of study		
	Exact Sciences - Engineering	242	47.2
	Biological Sciences - Health	58	11.3
	Sciences		
	Applied Social Sciences - Humanities - Linguistic, Arts or Letters	185	36.1
	Other	28	5.4

*QN: Question number in the questionnaire

Fonte: Do autor.

Sequenciamento das colunas

QN	Options	Facebook		Instagram		Twitter		Linkedin		Snapchat	
		N	%	N	%	N	%	N	%	N	%
5	A	0	0.0	200	39.0	262	51.1	245	47.76	368	71.7
	B	120	23.4	139	27.1	175	34.1	233	45.4	74	14.4
	C	160	31.2	90	17.5	31	6.0	21	4.1	36	7.0
	D	124	24.2	48	9.4	27	5.3	8	1.6	18	3.5
	E	56	10.9	24	4.7	9	1.5	2	0.4	10	1.9
	F	53	10.3	12	2.3	9	1.7	4	0.8	7	1.4

Fonte: Do autor.

Sequenciamento das colunas

Colunas contendo valores que serão comparados pelos usuários devem ser colocadas lado a lado

Method / Class	Approach	Hydrolase	Isomerase	Lyase	Ligase	Oxidoreductase	Transferase	Average
InfoGainAttributeEval	FS (Filter)	88.13	87.58	71.02	76.01	78.26	92.36	82.23
OneRAttributeEval		89.83	86.72	74.42	78.20	77.95	93.24	83.39
ReliefAttributeSelect		88.63	86.53	70.93	70.02	75.82	91.99	80.65
RandomSearch		89.19	86.42	71.63	68.54	75.96	92.44	80.70
CAE(SVM)	FS (Wrapper)	88.55	86.92	71.38	71.66	77.83	92.31	81.44
CAE(RF)		90.06	87.28	73.57	70.18	75.07	92.57	81.45
GA(KNN)		89.19	78.84	77.80	78.10	92.52	92.91	84.89
Factor Analysis	FE/FT	88.30	68.90	62.10	81.90	87.60	89.80	79.77
Average		88.99	83.65	71.61	74.33	80.13	92.20	81.82

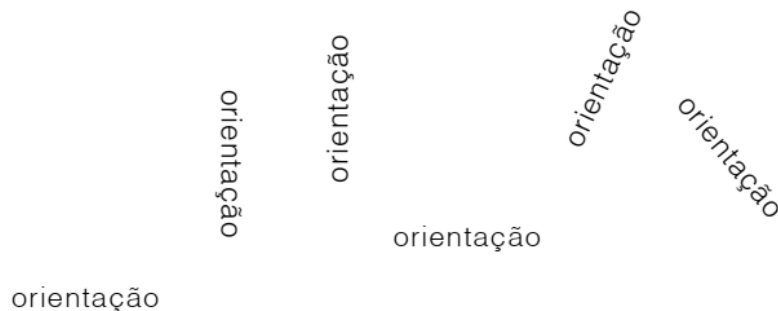
Fonte: Do autor

Sequenciamento dos dados

- Números e datas têm ordenações naturais bastante significativas
- Quando se trata de nomes, a ordenação alfabética nos auxilia a encontrar valores desejados em grandes tabelas
- Porém, em alguns casos pode-se criar outra ordenação que faça mais sentido

Orientação do texto

Wigdor e Balakrishnan (2005) descobriram que a leitura do texto inclinado a 45° em qualquer direção era, em média 52% mais lenta. Para inclinação a 90° em qualquer direção, era 205% mais lenta.



Alinhamento do texto

Números que representem valores quantitativos devem ser alinhados à direita

Table 4. Evaluation of the clusters obtained by DBSCAN.

<i>MinPts</i>	ϵ	Clusters	Silhouette index	Noise
2500	0.01	2	0.2197	60.73%
2500	0.05	4	0.4811	18.16%
2500	0.18	2	0.7926	2.76%
2500	0.26	2	0.7925	1.83%
5000	0.04	2	0.3801	34.17%
5000	0.13	3	0.8396	5.03%
5000	0.19	2	0.7928	2.70%
5000	0.26	2	0.7925	1.85%
5000	0.46	1	N/A	0.38%
6000	0.08	1	N/A	27.07%
6000	0.13	2	0.7558	14.41%
6000	0.19	2	0.7892	3.16%
6000	0.25	2	0.7924	2.04%
7000	0.04	1	N/A	44.41%
7000	0.08	1	N/A	27.08%
7000	0.15	2	0.7560	14.28%
7000	0.20	2	0.7553	12.32%
7000	0.52	1	N/A	0.35%

Fonte: Do autor

Alinhamento do texto

Datas devem ser **alinhadas à esquerda** usando um formato que mantenha **constantes o número de dígitos**

Date	Date	Date	Date
12/17/02	12/17/02	12/17/02	12/17/02
1/2/03	01/02/03	1/2/03	01/02/03
1/17/03	01/17/03	1/17/03	01/17/03
2/9/03	02/09/03	2/9/03	02/09/03
10/29/03	10/29/03	10/29/03	10/29/03
12/1/03	12/01/03	12/1/03	12/01/03
1/1/03	01/01/03	1/1/03	01/01/03

Uso de cores

Você pode usar cores para ressaltar valores importantes

Tabela

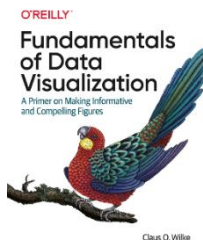
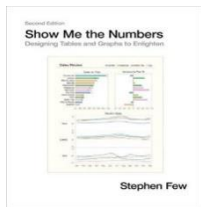
	A	B	C
Categoria 1	15%	22%	42%
Categoria 2	40%	36%	20%
Categoria 3	35%	17%	34%
Categoria 4	30%	29%	26%
Categoria 5	55%	30%	58%
Categoria 6	11%	25%	49%

Mapa de Calor

BAIXO-ALTO

	A	B	C
Categoria 1	15%	22%	42%
Categoria 2	40%	36%	20%
Categoria 3	35%	17%	34%
Categoria 4	30%	29%	26%
Categoria 5	55%	30%	58%
Categoria 6	11%	25%	49%

Leitura recomendada



[1] Jandre et al. Analysis of school performance of children and adolescents with Attention-Deficit/Hyperactivity Disorder - a dimensionality reduction approach. In Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2021) - Volume 5: HEALTHINF, pages 155-165 ISBN: 978-989-758-490-9 ISSN: 2184-4305 2021

[2] Song H, Schwarz N. If It's Hard to Read, It's Hard to Do: Processing Fluency Affects Effort Prediction and Motivation. Psychological Science. 2008;19(10):986-988. doi:10.1111/j.1467-9280.2008.02189.x

[3] Wigdor D., Balakrishnan R. (2005) Empirical Investigation into the Effect of Orientation on Text Readability in Tabletop Displays. In: Gellersen H., Schmidt K., Beaudouin-Lafon M., Mackay W. (eds) ECSCW 2005. Springer, Dordrecht. https://doi.org/10.1007/1-4020-4023-7_11

Links:

<https://homepages.dcc.ufmg.br/~raquelcm/index.php?alias=visualizacao2011>