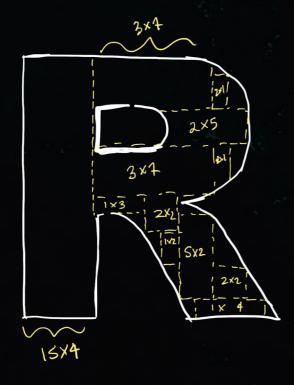
Introduction to R



Martín Amodeo

DBBF UNS IADO CONICET-UNS Bahía Blanca, Argentina

CIT, Changshu, China



内容 Content

- 对象类型 Types of *objects*
- 什么是数据框(dataframe)? ¿what is a dataframe?
- 整洁数据(Tidy data): 正确组织数据的规则 Tidy data. Rules for organizing your data properly

Objects





Types of Objects

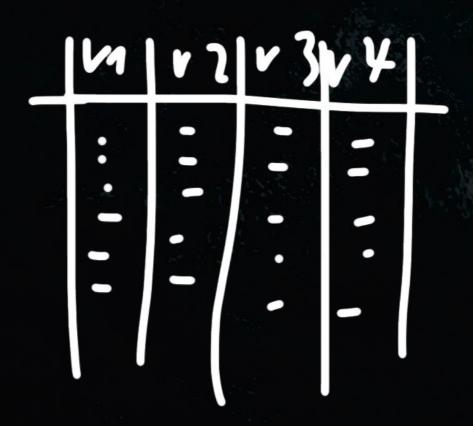
• Single Value

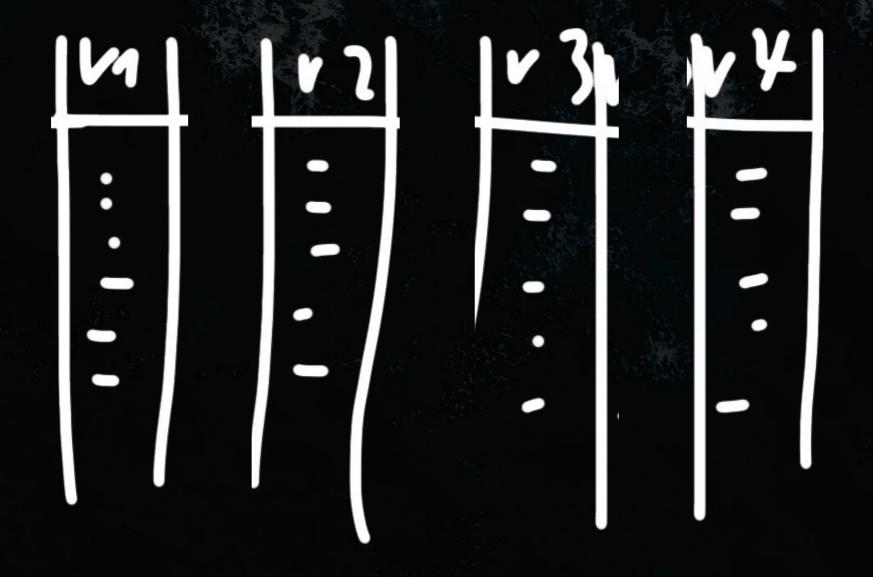
Vector

Dataframe

Types of Objects

• Data frame





Grouped Vectors

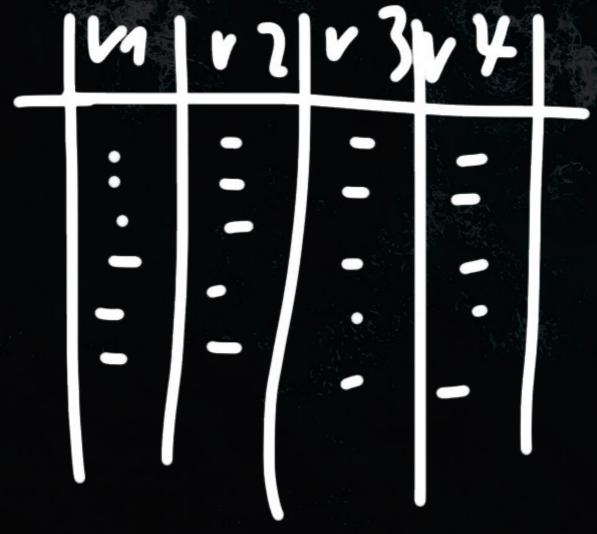


Variables

Observations

\$

data\$variable1 data\$variable2 data\$variable3



Tidy Dataframe

Simple, plain, complete

- Use the Rules for Names
- Unify labels
- Check spelling and writing
- No empty cells
- ONLY RAW DATA
- No extra information (average, graphs, etc)
- Missing Values as NA

Tidy Data

NA

Missing Data

- Lost Sample
- Lost Number
- Sample not measured
- Experimental error

		Α	В	С	D	Е	F	G	Н		J	K	L	М	N	0	Р	Q	R	S
- }																		POC-ST CAS	W.C	PM NORTH
- 1		_	\Rightarrow	teros figeramentos filitik														POC	186.0	17 68.50
-8			druing14	10-60 un	Numero defibros 24	850.23	et modad - differences de blancos (RCF-859). 890 24 871 86	NO 15	Social filtro (sal) 296-44	lotal fitre (md) pup	50074	\$000 rd 0.80	Chitrone 14 0.04	0.04	Paravaranteno 0.03	0.0	0 14	0.81	29.4	5 5.31
- 7		_		10-60 um	24	271.19 211.06	271.86 200.86	125.44 226.87	958.86 2528.48	252	2.51	2.61	0.12	0.12	9.31	0.0				+
- 4				0-115 un	12	884.70	88430	107.51	2596.52	250	0.91	0.93	0.12	0.12	0.01	0.0		0.2	17.5	258
- #4				0-115 un	12 12	220.14	21034 22014	-28,82 1416-92	15982.81	15.99	1941	1931	2.15	14	1,071	0.00	97.6			
-12	,	_		435 km	22	854.35 859.88	220.14 864.75 859.88	158.27 182.47	1814.26	1.81	0.42		0.05	0.05	0.00			0.31	68.3	8 8.25
- 17				415 um	22	228.67	22867	992.64	12150.00	12.15	10.88	10.884	2.86	1.86	0.641	0.09	52.6	1		
- 17																		PGC-97		SEM DICHES 1)
-17																		2.8	111.0	104
_13	\Box		tardi-14			#70.50	atese	129.70	1896-10	140	1.896	1.496	0.17	0.13	0.08	0.00	2 6.8	POC IN A	6 MC	994 (net-1) 5 159
- 25			ar ar	10-60 um	59	420.86	870 SG 856 R4	197.09	212158	2.12	7.177 71.787	2.122	0.17	0.17	0.121	0.00		0.1	103	159
34				0-185 un	59	430.86 218.12 151.69 881.48	256.94 278.12 251.69 181.48	197.09 1766.17 278.18 25.15	21787.05 1545.45 593.49	2129 2129 155 059	21.383 1.545	2 122 21 383 1 545 0 593	0.27 2.72 0.19 0.07	0.19	0.121 1.282 0.091	0.00		0.1	12.6	6 1.19
				60-115 um 60-115 um	58	88148 21117	181.48 211.12	1864.18	12728.68	12.73	12,729	0.593	1.59	0.03	0.750	0.06	1 523	2		
- 71	•	-	$\overline{}$	H35un H35un	58	353.66 377.62	211.17 151.66 177.62	218.86	2681.35 1022.04	2.64 1.02		2 5 8 2 1 6 2 2	0.14	0.11		0.00			203	18.0
-29				435 un	51	218.85	214.85	905.41	10480.49	10.48	10.480	10.480	131	16	0.631	0.05	51.6			
- 41																		POC-sil		SPM b (mes-1)
- 33																		3.4	1203	19950
- 39	\vdash		prist .	tino (70-60um putina/putosano)	310	126.55	175.88	246.29	4492.21	4.49	4.492	4.492	0.56	0.56	0.366	0.03	223	POC INT.	6 MC	990 (met-1) 22 2.08
37				teo (30-60um outre/outower)	110	350.42 191.83	850.42	229.45 2825.87	2750.01 81477.81	2.76 81.48	2,350	2,360 81,423	0.85	0.45		0.01				- 120
37				Moro (60-200um outina/outokano)	- 115	880.15	880.15	180.17	2914.88	291	2.915	2.915	0,00 0,00	0,00	0.171	0.01	0,0	0.6	55.6	6 285
-31				Moro (60-200um outrea/outokano)	92	196.00	196.10 196.10	2502.89	20044.88	20.04	20,044	20,044	251	16	1.18		9 99.7			
43				New (NEED are patting/patterand) New (NEED are patting/patterand)	100	155.49	355.49	-6.61 204.27	2212.85	721	2,218	7211	0.38	0.29	0.181	0.01	10.9	- 0.5	55.0	3 1.45
7:				MKD (2000 um outina/outorano)	100	21859	21859	1494.01	14148.76	14:15	14:149	14:149	1.77	16	0.811	0.03	69.7			
- 47																				
- 49	•																			
- 48																				
- 30																				
- 52																				
- 37		-		Mos Representos RR	Numero defiltros	dhan an a		rac		obstance of	drawn Madel Brown M.	de anno Marion De como	de mer bederde eres bit							=
- 30	3		dnay16	ID-SC MIN	24	Chen net-1	Pararacatoro 146	0.45	29.48	5.11	10.47	4.95	0.70							9 domber
-33				10-60 MH	24	0.12 0.12 0.14 0.14	4.72 12.46				22.55 59.24 44.56	4.95 16.03 42.25 21.03 19.15	0.70 2.25 5.95 2.97 2.70							
- 21					MAN DANKE	0.16	12/48 6:21 5:65	0.45	29.48	5.81	40.50	21.03 19.15	2.92							
- 5					Si .	1,71205080757	1.26	0.00	0.00	0.00	23.41	11.03	156							
8 5								POC	LANC .	sinting, a	chana Market Statement I	dien/YOC NIVAN	duno himfu out t							
- 8		_		0-185 un	12	0.12	459	0.11	17.91	251	54.15	25.61 0.00	4.61							
6				60-115 un	17	14	11				0.00	0.00	4 61 0 00 0 00 1 54 2 66							9 domber
- 88					MANAGO Desarro	0.04	152 255	0.21	17.91	258	18.05 81.25	854 1438	154 256							
- 70)				Si tac o	1,71205080757	158	0.00	0.00	0.00		855	154							
17								POC												

NO Tidy Data

Empty cells, special characters

	KANDELL KEN	1. 1. 1. 1.		(8)	and the	
	Α	В	С	D	E	F
1		n.obs	vel.max	v el. media	dir	
2	2013-06-30 10:50:00	612	45.6	27.7	ЙM	
3	2013-07-07 10:50:00		33.3	19.1	ЙM	
4	2013-08-08 10:50:00	514	\sim	30.5	ЙM	
5	2013-08-17 10:50:00	508	NA	33.7	NW	
6	2013-09-25 10:50:00	567	50.6	3.3	NW	
7	2013-10-20 10:50:00	539	66.2	21.9	NW	
8	2013-11-21 10:50:00	430	93.1	38.4	NA	
9	2013-12-07 10:50:00	584	52.4	24.7	NA	
10	2013-12-23 10:50:00		53.5	18.1	NW	
11	2014-04-21 10:50:00	539	26.8	14.4	NW	
12	2014-06-17 10:50:00	454	NA	NA	ΝA	
13	2014-06-24 10:50:00	394	\sim	NA	ΝA	
14	2014-07-10 10:50:00	654	\sim	NA	ΝA	
15	2014-07-26 10:50:00	568	NA	NA	NA	
16	2014-08-04 10:50:00	641	42.4	32.9	NW	
17	2014-08-20 10:50:00	125	54.4	16.6	NW	
18	2014-08-27 10:50:00	109	16.5	7.4	NW	
19	2014-09-12 10:50:00	72	49.5	17.5	NW	
20	2014-11-15 10:50:00	554	41.3	7.4	NW	
21	2014-11-24 10:50:00	206	45.4	NA	NA	
22	2014-12-01 10:50:00	622	40.5	16.6	NW	
23						
24						
25						
26						
~-						

Tidy Data

Simple, plain, complete, use of NA

Praxelis Data