

Tidy Data Quiz

Name

KEY

SLUNetID

1) [5 pts] True or false. According to this article, the following are properties of 'tidy data':

- ☒ T F Each variable forms a column
☐ T ☒ F Each observation forms a column
☐ T ☒ F Each variable forms a row
☒ T F Each observation forms a row
☒ T F Each type of observational unit forms a table

Hint: We've talked about variables as 'labels' and observations as 'individuals'

2) [1 pt] The following tables all present the same information. Which of the following tables are considered 'tidy' according to this article? Circle the correct answer.

	treatmenta	treatmentb
John Smith	—	2
Jane Doe	16	11
Mary Johnson	3	1

(a)

	John Smith	Jane Doe	Mary Johnson
treatmenta	—	16	3
treatmentb	2	11	1

(b)

person	treatment	result
John Smith	a	—
Jane Doe	a	16
Mary Johnson	a	3
John Smith	b	2
Jane Doe	b	11
Mary Johnson	b	1

(c)

[2 pts] Explain what makes the table you selected 'tidy' compared to the other tables.

Each variable is in a column + rows consist of one + only one distinct measurement.

flip over

3) [2 pts] The following is a list of common properties seen in messy datasets:

- (a) Column headers are values, not variable names.
- (b) Multiple variables are stored in one column.
- (c) Variables are stored in both rows and columns.
- (d) Multiple types of observational units are stored in the same table.
- (e) A single observational unit is stored in multiple tables.

Each of the following tables has (at least) one of the above problems. Put the letter for one problem seen in each of the tables below (if the table has more than one problem, you only need to list one).

religion	<\$10k	\$10-20k	\$20-30k
Agnostic	27	34	60
Atheist	12	27	37
Buddhist	27	21	30
Catholic	418	617	732
Don't know/refused	15	14	15
Evangelical Prot	575	869	1064
Hindu	1	9	7
Historically Black Prot	228	244	236
Jehovah's Witness	20	27	24
Jewish	19	19	25

Information on income and religion from the Pew Forum

Messy data property: a

country	year	m014	m1524	m2534	m3544	m4554	m5564	m65	mu	f014
AD	2000	0	0	1	0	0	0	0	—	—
AE	2000	2	4	4	6	5	12	10	—	3
AF	2000	52	228	183	149	129	94	80	—	93
AG	2000	0	0	0	0	0	0	1	—	1
AL	2000	2	19	21	14	24	19	16	—	3
AM	2000	2	152	130	131	63	26	21	—	1
AN	2000	0	0	1	2	0	0	0	—	0
AO	2000	186	999	1003	912	482	312	194	—	247
AR	2000	97	278	594	402	419	368	330	—	121
AS	2000	—	—	—	—	1	1	—	—	—

Information on the number of individuals of certain ages and genders by country and year.

Messy data property: a, b, d

~~There is also an argument that costs be made for e~~