

$\% > \%$

# Data management in the tidyverse

Hannes Rosenbusch & Michael Nunez

**Programming in Psychological Science**

*Why is data management  
important?*

## Multiple reasons:

- Data management important for collecting your own data
- Analyzing shared data
- Data collected by third-parties
- **"BIG"** data
- You will need to share your own data



**Poor data management is potentially very time consuming**



*Well-organized data*

# Making data *Tidy*

- ***Tidy Data*** is a paper by statistician / R developer Hadley Wickham
  - These ideas started the ***tidyverse*** packages for R
  - But the ideas of ***Tidy Data*** are universal to many data management programs
- Many data sets can either be transformed or summarized into ***Tidy Data***

# Properties of Tidy data

- Each *observation* in a row
- Each *variable* in a column
- Each experiment / observational unit in a data Table

## Other properties

- Data identifiers are the first columns (e.g., Participant ID)
- Fixed values by experimental design or data collection come next (e.g. experimental condition)
- Measured values are later columns



# Messy data in “long” format

AutoSave messy\_data.csv

File Home Insert Page Layout Formulas Data Review View Help

Clipboard Font Alignment Number Conditional Formatting

POSSIBLE DATA LOSS Some features might be lost if you save this workbook in the comma-delimited (.csv) format. To preserve these features, save it in an Excel file format.

	A	B	C	D	E	F	G
1	Index	ID	Script	Created	Subject	Name	Value
2	4.8E+07	155980	162422	28/09/2021 14:32	162574	In set (I)	TRUE
3	4.8E+07	155980	162422	28/09/2021 14:32	162574	Choice	I
4	4.8E+07	155980	162422	28/09/2021 14:32	162574	Response Time	1863
5	4.8E+07	155980	162422	28/09/2021 14:32	162574	In set (I)	TRUE
6	4.8E+07	155980	162422	28/09/2021 14:32	162574	Choice	I
7	4.8E+07	155980	162422	28/09/2021 14:32	162574	Response Time	1695
8	4.8E+07	155980	162422	28/09/2021 14:32	162574	In set (I)	TRUE
9	4.8E+07	155980	162422	28/09/2021 14:32	162574	Choice	s
10	4.8E+07	155980	162422	28/09/2021 14:32	162574	Response Time	1429
11	4.8E+07	155980	162422	28/09/2021 14:32	162574	In set (I)	TRUE
12	4.8E+07	155980	162422	28/09/2021 14:32	162574	Choice	s
13	4.8E+07	155980	162422	28/09/2021 14:32	162574	Response Time	2102
14	4.8E+07	155980	162422	28/09/2021 14:32	162574	In set (I)	TRUE
15	4.8E+07	155980	162422	28/09/2021 14:32	162574	Choice	s
16	4.8E+07	155980	162422	28/09/2021 14:32	162574	Response Time	1629
17	4.8E+07	155980	162422	28/09/2021 14:32	162574	In set (I)	TRUE
18	4.8E+07	155980	162422	28/09/2021 14:32	162574	Choice	I
19	4.8E+07	155980	162422	28/09/2021 14:32	162574	Response Time	1230
20	4.8E+07	155980	162422	28/09/2021 14:32	162574	In set (I)	TRUE
21	4.8E+07	155980	162422	28/09/2021 14:32	162574	Choice	s
22	4.8E+07	155980	162422	28/09/2021 14:32	162574	Response Time	2305
23	4.8E+07	155980	162422	28/09/2021 14:32	162574	In set (I)	TRUE
24	4.8E+07	155980	162422	28/09/2021 14:32	162574	Choice	s
25	4.8E+07	155980	162422	28/09/2021 14:32	162574	Response Time	1671
26	4.8E+07	155980	162422	28/09/2021 14:33	162574	In set (I)	TRUE

messy\_data

# Tidy-er data

AutoSave tidy\_data.csv - Saved

File Home Insert Page Layout Formulas Data Review View Help

Get Data From Text/CSV From Web From Table/Range Recent Sources Existing Connections Refresh All Queries & Connections Properties Edit Links Data Types Stocks (En... Geography Sort Filter Clear Reapply Advanced

Get & Transform Data Queries & Connections Sort & Filter

POSSIBLE DATA LOSS Some features might be lost if you save this workbook in the comma-delimited (.csv) format. To preserve these features, save it in an Excel file format.

	A	B	C	D	E	F	G	H
1	ID	Created	Subject	In set (I)	Choice	Accuracy	Response Time	
2	156046	28/09/2021 15:20	162643	FALSE	I	0	1318	
3	156046	28/09/2021 15:20	162643	FALSE	s	1	1660	
4	156046	28/09/2021 15:21	162643	TRUE	s	0	1093	
5	156046	28/09/2021 15:21	162643	FALSE	s	1	1132	
6	156046	28/09/2021 15:21	162643	TRUE	s	0	3520	
7	156046	28/09/2021 15:21	162643	TRUE	I	1	3112	
8	156046	28/09/2021 15:21	162643	TRUE	I	1	1222	
9	156046	28/09/2021 15:21	162643	TRUE	s	0	2209	
10	156046	28/09/2021 15:21	162643	FALSE	I	0	1828	
11	156046	28/09/2021 15:21	162643	FALSE	s	1	1286	
12	156052	28/09/2021 15:26	162651	TRUE	s	0	1608	
13	156052	28/09/2021 15:26	162651	TRUE	I	1	1295	
14	156052	28/09/2021 15:26	162651	FALSE	s	1	1032	
15	156052	28/09/2021 15:26	162651	TRUE	s	0	1819	
16	156052	28/09/2021 15:26	162651	TRUE	I	1	2265	
17	156052	28/09/2021 15:26	162651	TRUE	s	0	2405	
18	156052	28/09/2021 15:26	162651	FALSE	I	0	1218	
19	156052	28/09/2021 15:26	162651	FALSE	I	0	1737	
20	156052	28/09/2021 15:26	162651	FALSE	s	1	1168	
21	156052	28/09/2021 15:27	162651	TRUE	I	1	1335	
22								

Ready tidy\_data

Now see our .R script!