

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 <u>or</u> summarized into *Tidy Data*

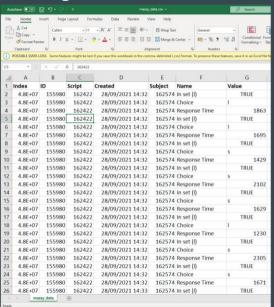
Properties of Tidy data

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

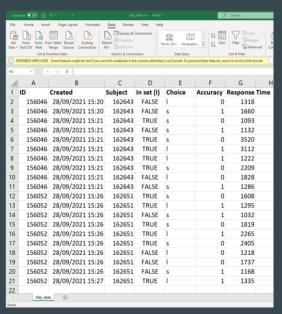
Other properties

- 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



Tidy-er data



1

Now see our .R script!