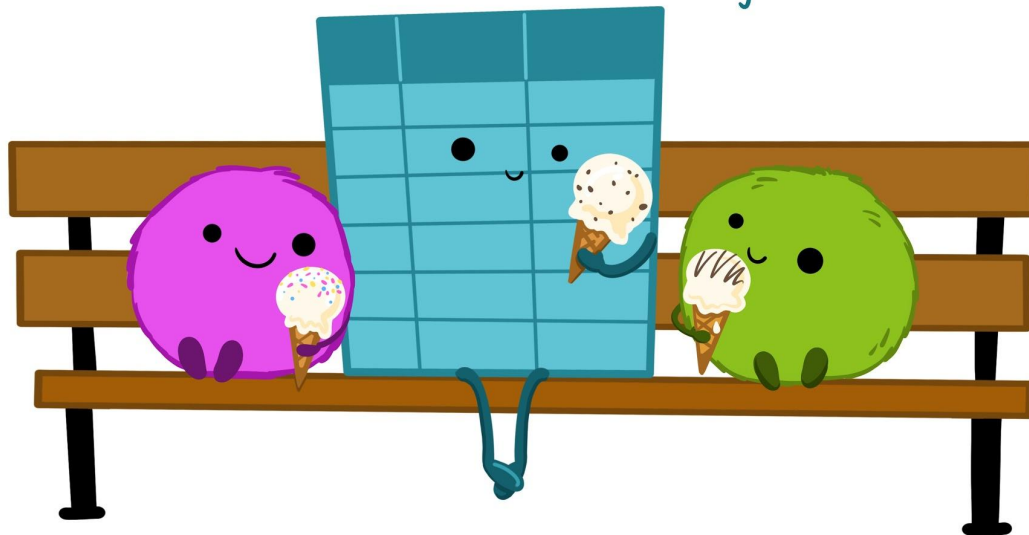


make friends with tidy data.



Tidy Data Illustrated Series

CC By Julie Lowndes & Allison Horst

Reuse: [These slides \(low res\)](#) • [Blog](#) • [Twitter](#) • [GitHub](#)

Please cite as: "Illustrations from the [Openscapes](#) blog [Tidy Data for reproducibility, efficiency, and collaboration](#) by Julia Lowndes and Allison Horst"

“**TIDY DATA** is a standard way of mapping the meaning of a dataset to its structure.”

—HADLEY WICKHAM

In tidy data:

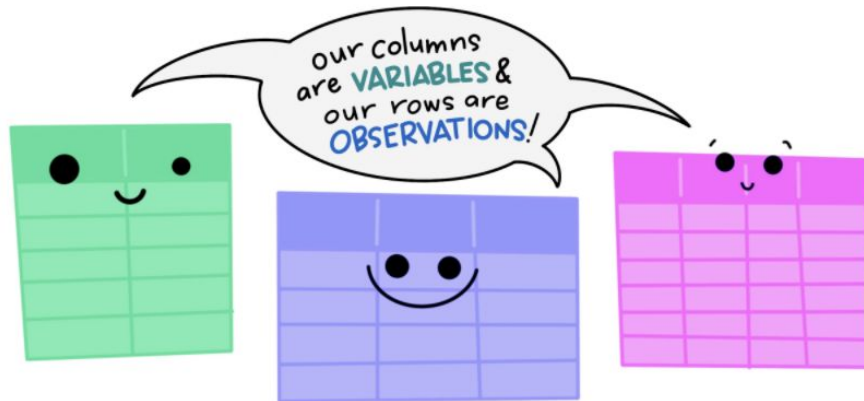
- each variable forms a column
- each observation forms a row
- each cell is a single measurement

each column a variable

id	name	color
1	floof	gray
2	max	black
3	cat	orange
4	donut	gray
5	merlin	black
6	panda	calico

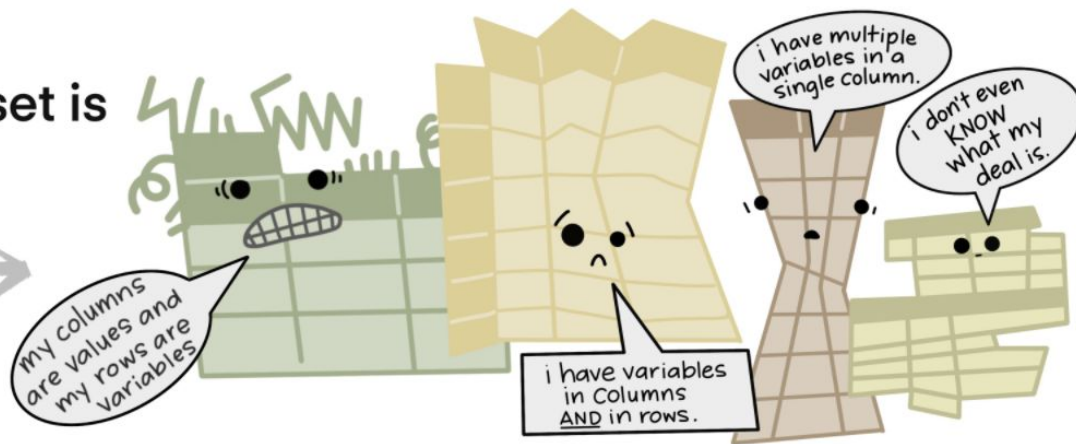
each row an observation

The standard structure of tidy data means that
"tidy datasets are all alike..."

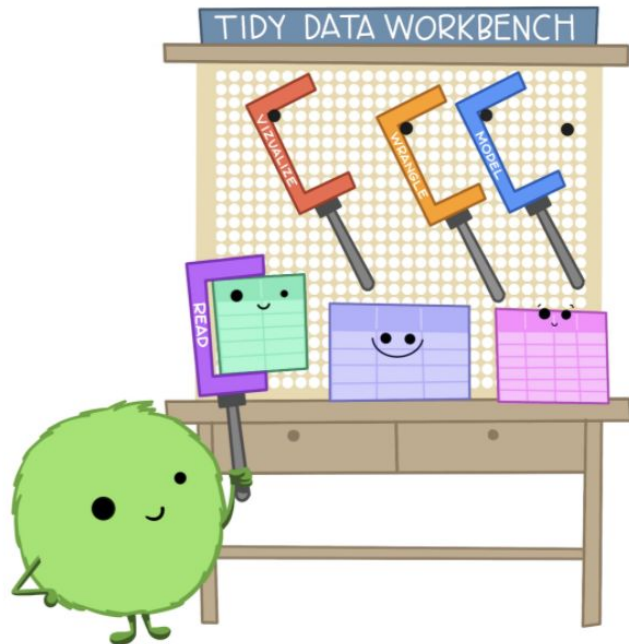


"...but every messy dataset is
messy in its own way."

—HADLEY WICKHAM

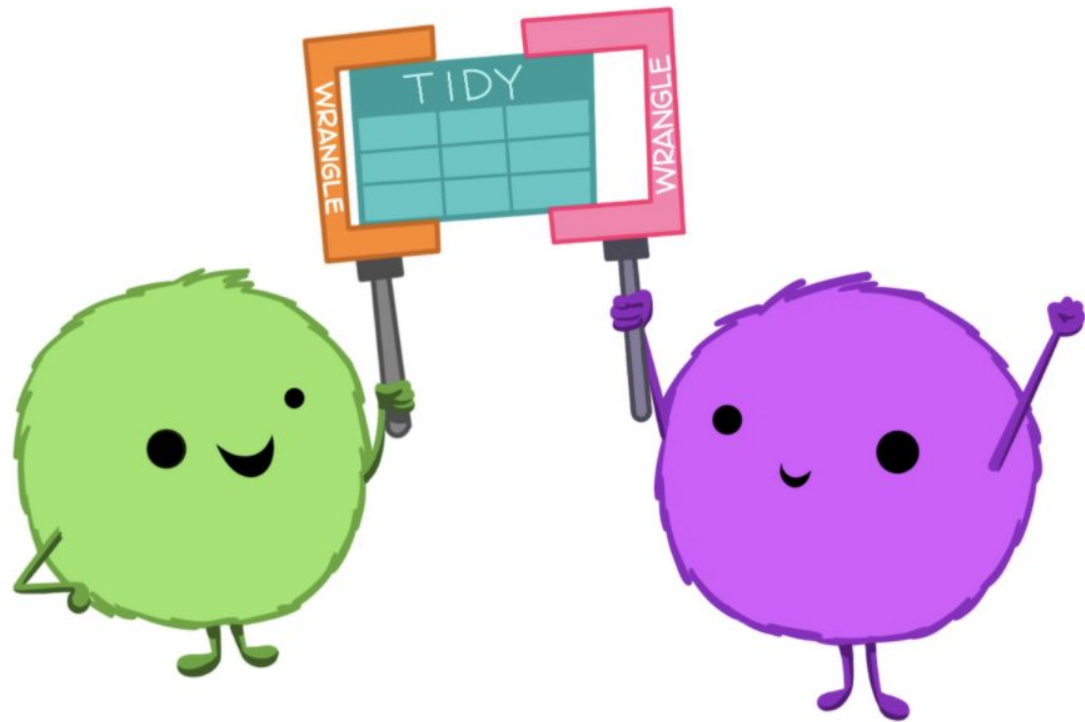


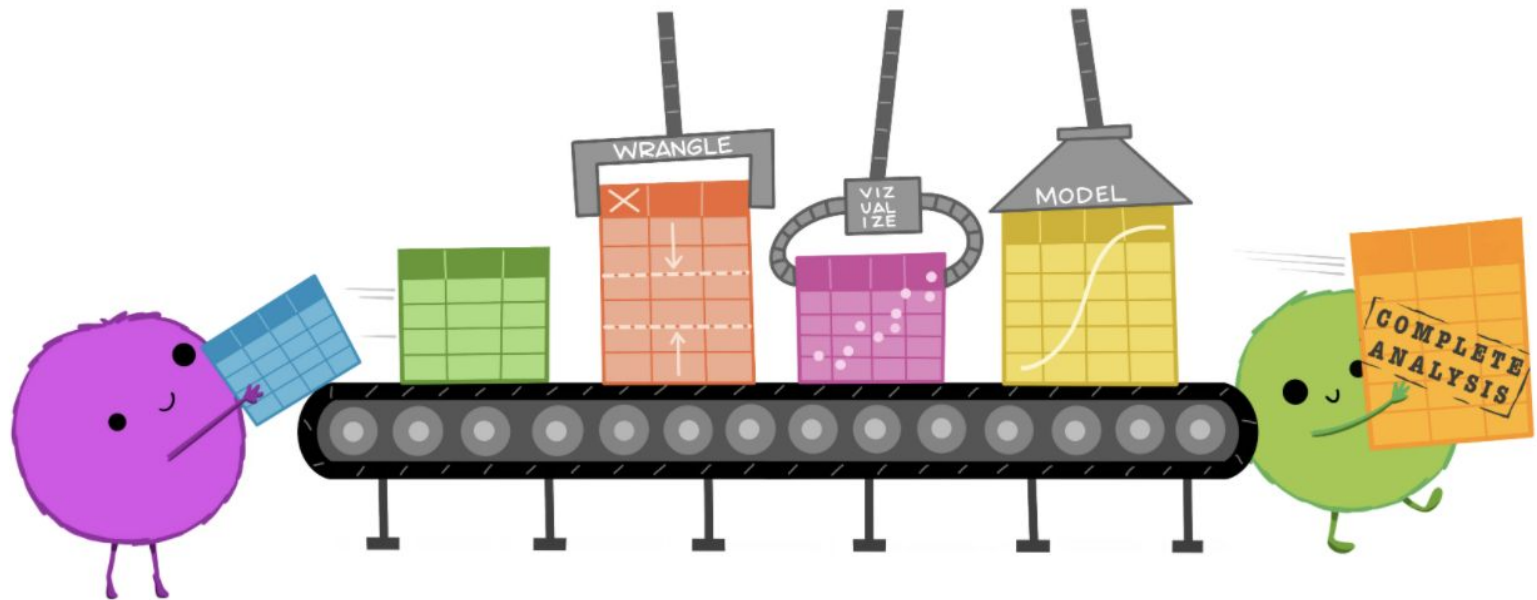
When working with tidy data, we can use the same tools in similar ways for different datasets...

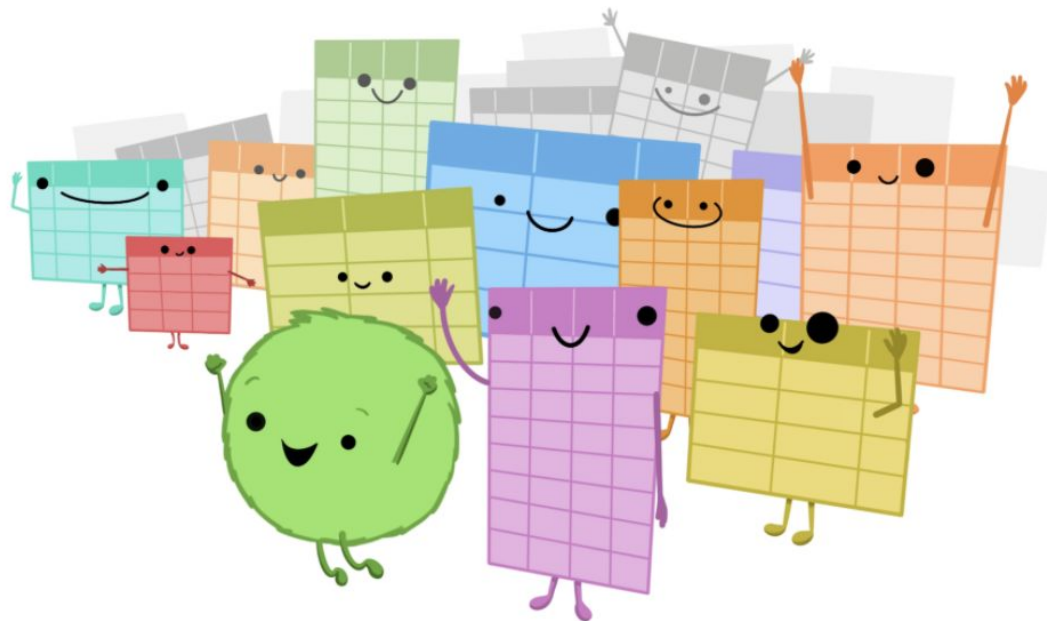
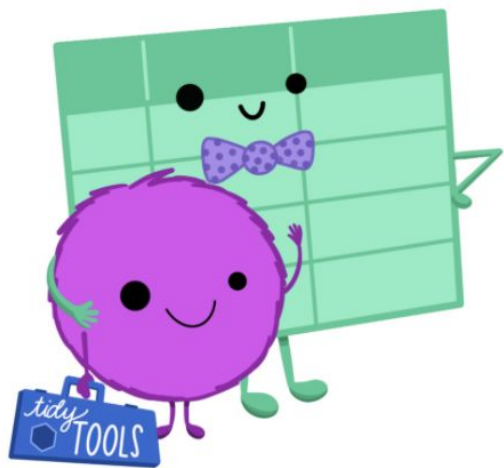


...but working with untidy data often means reinventing the wheel with one-time approaches that are hard to iterate or reuse.

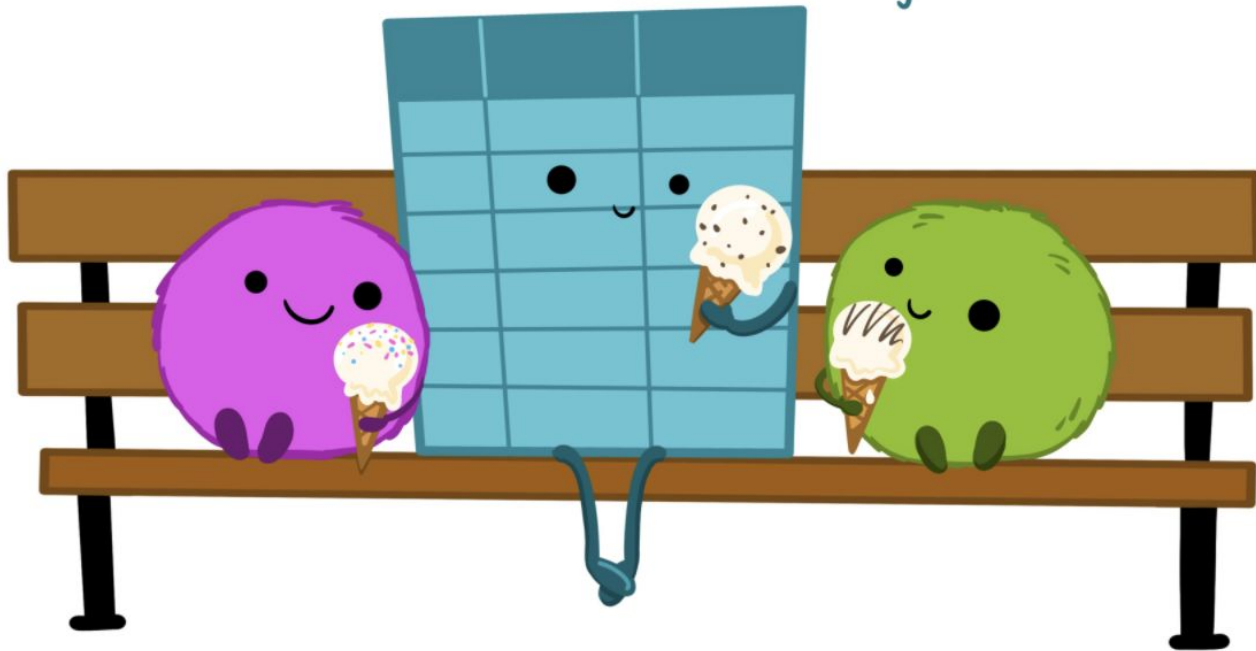








make friends with tidy data.



Learn more about tidy data:

Wickham (2014). Tidy Data. Journal of Statistical Software. <http://jstatsoft.org/v59/i10>

“Informal and code-heavy version”: <https://tidyr.tidyverse.org/articles/tidy-data>

Broman & Woo (2018). Data Organization in Spreadsheets. <https://peerj.com/preprints/3183/>

Grolemund & Wickham (2016). R for Data Science: Ch 12 <https://r4ds.had.co.nz/>