the tidyverse and tidy data

What is the tidyverse?

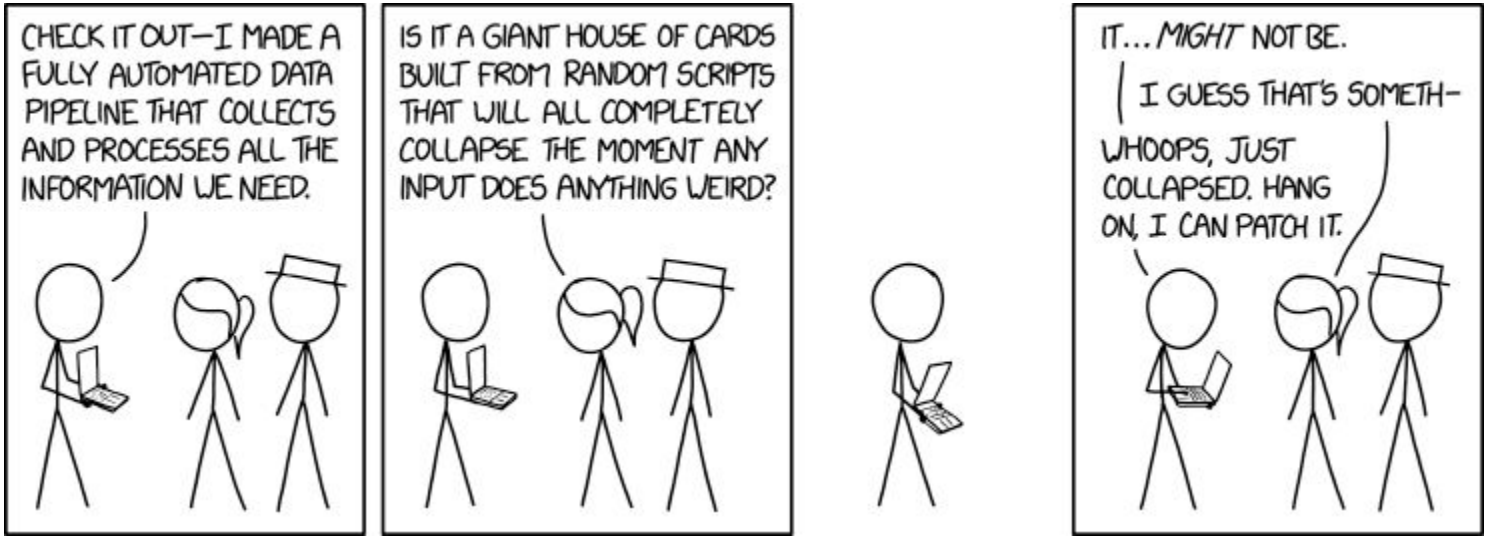
"The tidyverse is an opinionated collection of R packages designed for data science. All packages share an underlying design philosophy, grammar, and data structures."

- Keep everything simple
  - Use **existing** data structures instead of custom, aka, use tidy data
  - Functions should do one thing well
- Glue the simple things together; simple things put together are more powerful than one complex thing
- Design for humans

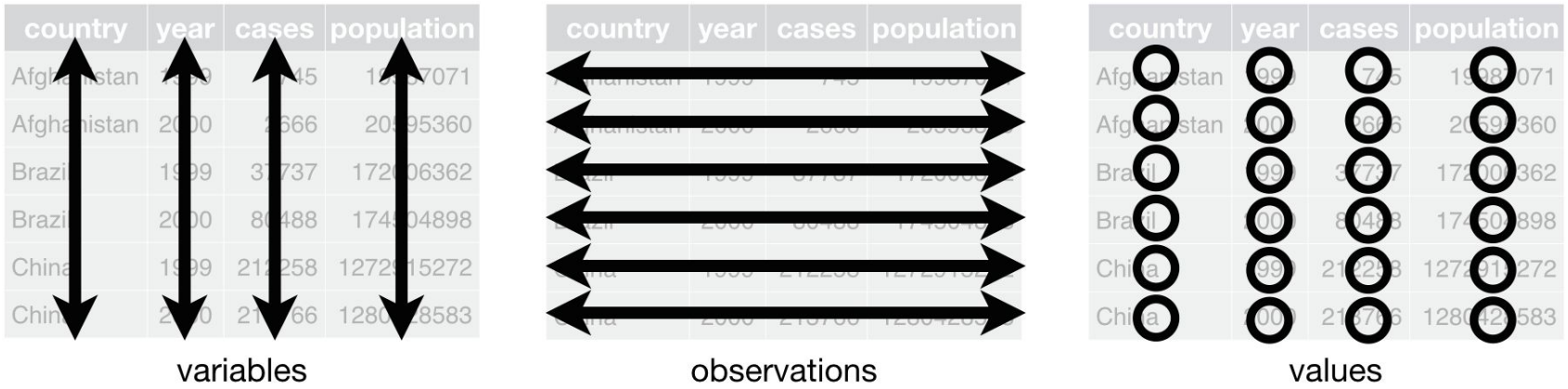
Data wrangling

Organizing your data into the form you want

- Everyone spends most of their time on wrangling ! It's hard!
- The tidyverse makes it much easier though
- focus on tidy data and `tidyr` (and a bit of `dplyr`) today, but most tidyverse packages have data wrangling applications



Tidy Data



1. Each variable is in a column.
2. Each observation is a row.
3. Each value is a cell.