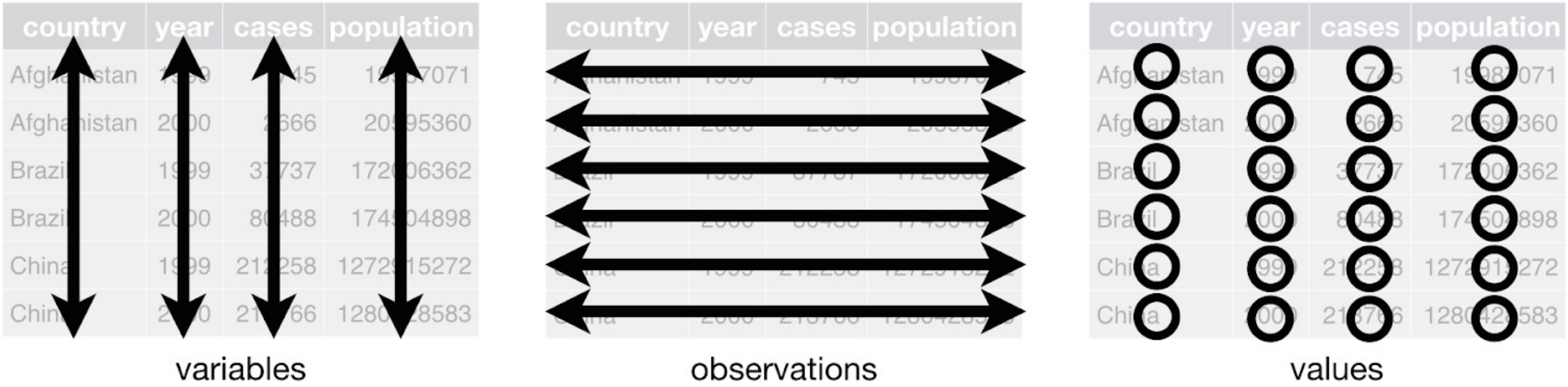
Introduction to Data Wrangling with `dplyr`

2021-07-07

Tidy Data

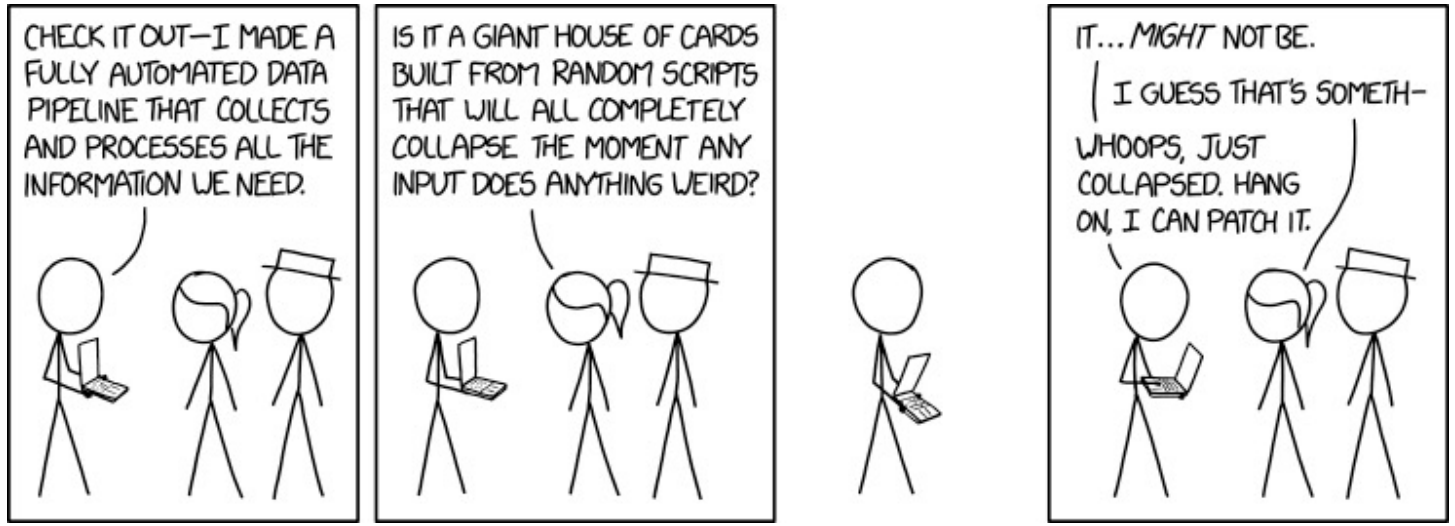


1. Each variable is in a column.
2. Each observation is a row.
3. Each value is a cell.

Data Wrangling

data wrangling – organizing your data into the form you want

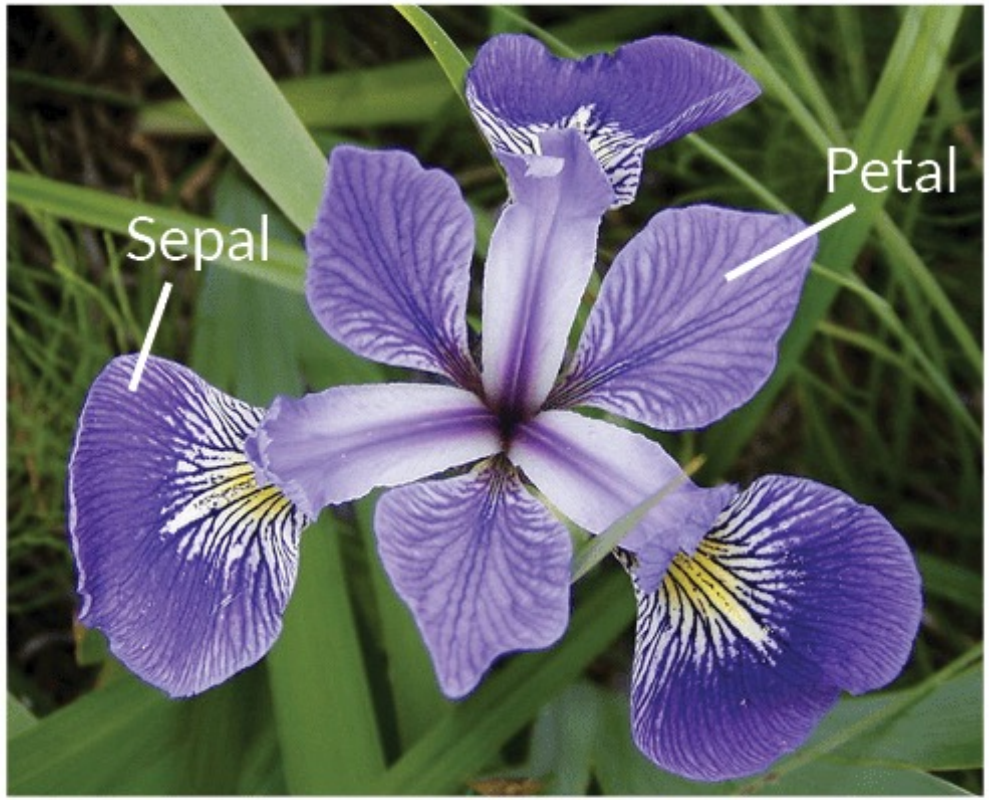
- Everyone spends most of their time on wrangling ! It's hard!
- The tidyverse makes it much easier though; that's it's primary purpose



dplyr

- mutate() adds new variables that are functions of existing variables
- select() picks variables based on their names.
- filter() picks cases based on their values.
- summarise() reduces multiple values down to a single summary.
- arrange() changes the ordering of the rows.
- You also need to know group_by() which allows you do any function by group.

The Data



Iris Versicolor

Iris Setosa

Iris Virginica

DEMO WITH RStudio

Resources

- R https://www.r-project.org/
- Bioconductor https://www.bioconductor.org/
- Tidyverse https://www.tidyverse.org/packages/

Where to Get Help

- RStudio Help Panel
- Google the function or error message
- Ask me or fellow students on Slack