Data Reshaping

Pandas: Reshaping

Tidy Data

"It is often said that 80% of data analysis is spent on the process of cleaning and preparing the data." (Dasu and Johnson, 2003)

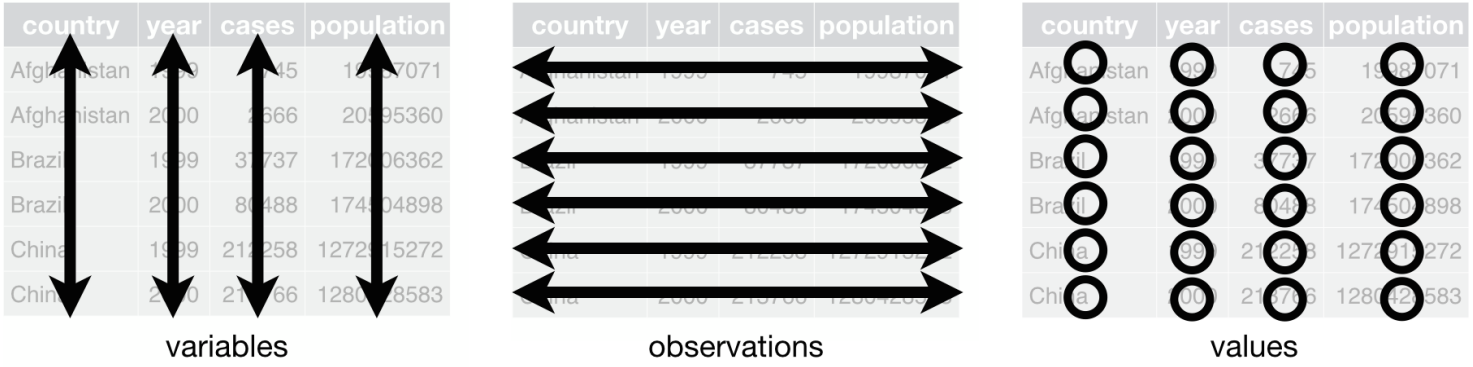
Tidy Data

- Makes data suitable for many tasks
- Data easier to manipulate
- Data easier to model
- Data easier to visualize
- New data is added as new observations

Pandas: Reshaping

Tidy Data

1. Each variable must have its own column.
2. Each observation must have its own row.
3. Each value must have its own cell.



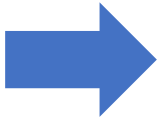
Source: http://r4ds.had.co.nz/tidy-data.html

Pandas: Reshaping

Tidy Data: Example

Wide Data

Student	Math	Sport	Art
Stuart	2	3	4
Bob	3	1	2
Kevin	3	2	1



Tidy Data

Student	Subject	Grade
Stuart	Math	2
Bob	Math	3
Kevin	Math	3
Stuart	Sport	3
Bob	Sport	1
Kevin	Sport	2
Stuart	Art	4
Bob	Art	2
Kevin	Art	1