

# manipulating data

adam okulicz-kozaryn

`adam.okulicz.kozaryn@gmail.com`

this version: Wednesday 13<sup>th</sup> September, 2023 16:01

# outline

misc

intuition

# outline

misc

intuition

## old ps comments

- keep it simple especially when learning new things!
- way easier to figure things out with a small and handy data
- say keep 5 vars and 50 obs
- so not only simplicity in code but also in data is good
  - later: large complex datasets and advanced code
- but do try to simplify, esp when learning and figuring it out
- if you have questions on my comments on your ps
- do ask for clarification!!
- i tend to be overly parsimonious
- always cite data!
  - at a minimum say where exactly it come from, ie the url
  - if ambiguous say which year, wave, version etc

# outline

misc

intuition

## general idea, intuition

- data management is mostly about:
- manipulating eg `gen`, `map`, `[:]`—today!
- (and related understanding, des sta/vis  
eg `describe`, `value_counts`—last week)
- (and `merge`—next week)
- pretty easy—no complicated code, no fancy things
  - but also lame, boring, and tedious
  - yet necessary!

## basic coding rules

- simplicity, clarity, efficiency:
  - drop everything that is not necessary
  - drop the clutter and be clean
- have “tight” code:
  - as few lines as possible that do as much as possible
- be lazy (copy from others, not 100% !)
  -
- more rules later