Loops and Iteration
Why?

- calculate summary statistics for several variables
- modify multiple columns at once
- filter based on many columns
- expand date columns into components
- read many files
- save multiple plots

R has a style where you don't manually loop; instead, you give a function to another function, and R loops for you
Across()

- select which columns to apply something to → .cols
- choose what function to apply → .fns
- optionally control new column names → .names
- You use across() mainly inside:
- summarize()
- mutate()
- sometimes filter() (via helpers)

Selecting columns: .cols
You choose the columns just like in select()

- a:d
- starts_with("a")
- where(is.numeric)
- everything()
  Ex:
  summarize(df, across(where(is.numeric), median))

Naming new columns: .names

- The default name is {column}_{function}.
  - Override it:
    - across(a:d, mean, .names = "mean_{.col}")
- *This matters most in mutate() where overwriting vs duplicating columns is a choice.

Filtering across columns with dplyr:

- if_any() = keep rows where any selected columns satisfy condition
- if_all() = keep rows where all selected columns satisfy condition
  Ex:
  filter(df, if_any(a:d, is.na))

Keys:

- across() does the same to many columns
- purrr::map() does the same to many elements/files/objects
  - https://rstudio.github.io/cheatsheets/purrr.pdf