Data manipulation

Objective

- Use git branches in the "real world"
- Reproducibly clean, summarize, and organize dataframes using tidyverse packages
- Understand how R stores different data types (data frames, vector types, missing data)
- Know what the tidyverse is, how it differs from base R, and the philosophy behind using it here.
- Use the pipe to chain operations together
- Use dplyr functions to subset data (select, filter; logic and select fxns including where, ==, %in%, !) and manipulate data (mutate; lubridate; split-apply-combine)
- Use tidyr functions to reshape data (pivot_wider and pivot_longer)

Lesson outline

- Review from last week
- Warm-up: create a branch for today's work
- Slides/discussion: using R for reproducible data analysis
  - Why use R?
  - What is the tidyverse and why use it?
  - Install dplyr and tidyr
- Live coding: How R thinks about data
  - Data carpentry R ecology revamp episode #2
  - Data frames
  - Vectors and data types
  - Missing data
- Live coding: dplyr and tidyr
  - Data carpentry R ecology revamp episode #3
  - Recent DC + R lesson
  - Chaining lines together with the pipe
    - `%>%`, `|>`
  - Subsetting and filtering data
    - incl. selection and `pick` https://dplyr.tidyverse.org/reference/dplyr_tidy_select.html
  - Adding columns
  - Split-apply-combine
  - Reshaping
- Live coding: advanced tidyverse topics
  - Options: across; dates; advanced joins; others?
- Live coding/discussion: getting help
  - `reprex`
- Live coding: practice modify-add-commit cycle

- Homework: None

Installation & materials

1. Slides
2. Install R packages ‘dplyr’, ‘tidyr’, ‘readr’
3. Data carpentry R ecology revamp episode #2
4. Data carpentry R ecology revamp episode #3