

Data Transformation with dplyr

CVEN 5837 - Summer 2022

Lars Schöbitz

<https://cven5837-ss22.github.io/website/>

<https://cven5837-ss22.github.io/website/>

Data Organisation in Spreadsheets

Think, Pair, Share

Questions

1. Why should you not leave a blank cell in a spreadsheet used for data collection?
2. Which of the 12 rules for data organization was the least comprehensible to you?

- **Think** for 2 minutes
- **Pair** with in break-out rooms for 4 minutes
- **Share** your answer with the class

02:00

<https://cven5837-ss22.github.io/website/>



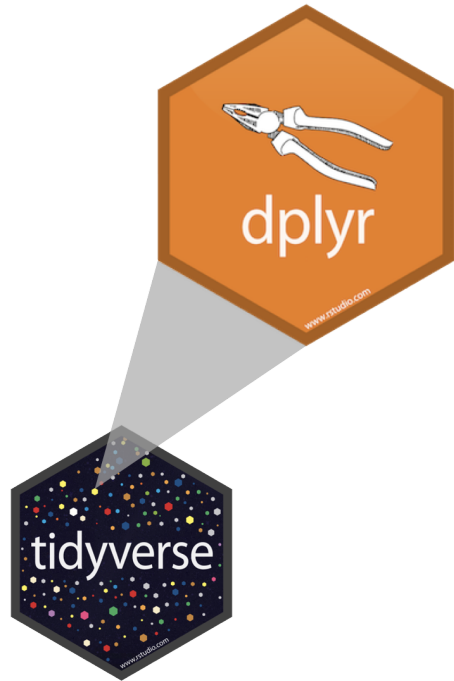
Learning Objectives (for this week)

1. Learners can apply ten functions from the dplyr R Package to generate a subset of data for use in a table or plot

Data wrangling with dplyr

A grammar of data wrangling...

... based on the concepts of functions as verbs that manipulate data frames



- **select**: pick columns by name
- **arrange**: reorder rows
- **slice**: chooses rows based on location
- **filter**: pick rows matching criteria
- **relocate**: changes the order of the columns
- **mutate**: add new variables
- **summarise**: reduce variables to values
- **group_by**: for grouped operations
- ... (many more)

dplyr rules

Rules of **dplyr** functions:

- First argument is always a data frame
- Subsequent arguments say what to do with that data frame
- Always return a data frame
- Don't modify in place

Live Coding Exercise: SDG 6.2.1

live-03a-data-transformation

1. Head over to rstudio.cloud
2. Open the workspace for the course (cven5837-ss22)
3. Open “Projects”
4. Open the “course-materials” project
5. Follow along with me

Break



10:00

<https://cven5837-ss22.github.io/website/>

Photo by [Blake Wisz](#)



Mortenson Center
in Global Engineering
UNIVERSITY OF COLORADO BOULDER

Pair Programming Exercise

Pair Programming Exercises

- Two learners work together in a break out session
- One person (the driver) shares the screen and does the typing
- The other person (the navigator) offers comments and suggestions
- Roles get switched

hw-03a-data-visualiation

1. Head over to rstudio.cloud
2. Open the workspace for the course (cven5837-ss22)
3. Open “Projects”
4. Open the “course-materials” project

Homework week 3

Homework late submission

- ~~25%~~ 10% each day that a homework is submitted late

Homework due dates

- All material on [course website](#)
- Homework assignment due: Friday, 22th July
- Learning reflection due: Monday, 25th July

Thanks! 🌻

Slides created via revealjs and Quarto:

<https://quarto.org/docs/presentations/revealjs/> Access slides as [PDF on GitHub](#)

All material is licensed under [Creative Commons Attribution Share Alike 4.0 International](#).