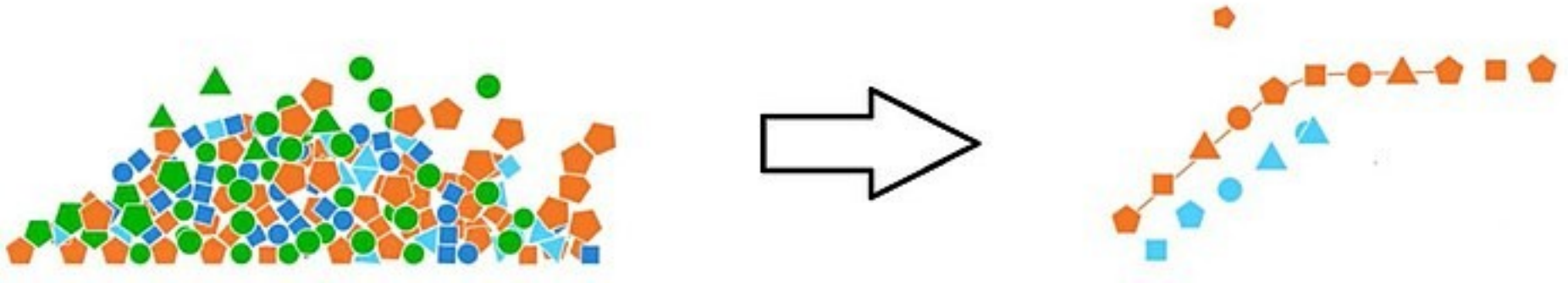
AGENDA

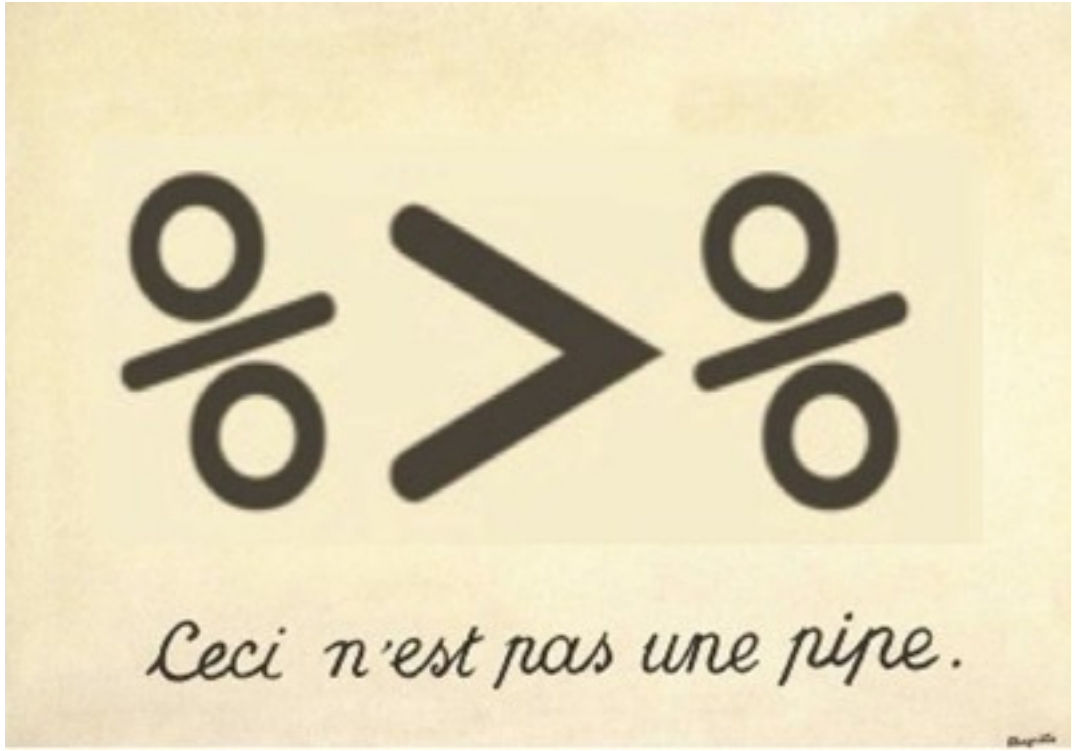
- What do we mean by data cleaning?
- Intro to “pipes”
- Exercises:
  - 1: Introduction to the “dplyr” package
  - 2: Introduction to the “data.table” package
  - 3: Introduction to the “tibble” package
  - 4: Introduction to the “tidyverse” collection of packages (optional)
  - 5: Isolating data with “dplyr” (optional)
  - 6: Deriving information with “dplyr” (optional)
- Connecting your Rstudio to Github through a git client

NB: Short session today! Available on email

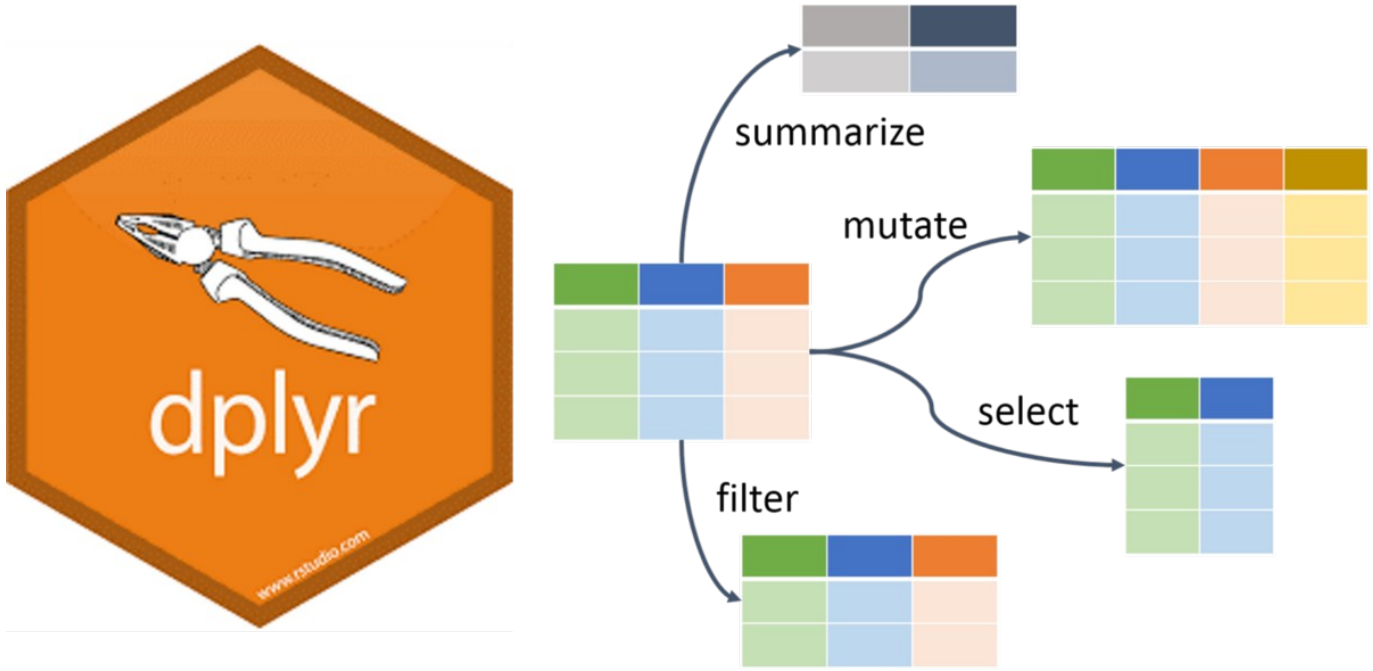
DATA CLEANING/WRANGLING/MANIPULATION



PIPES

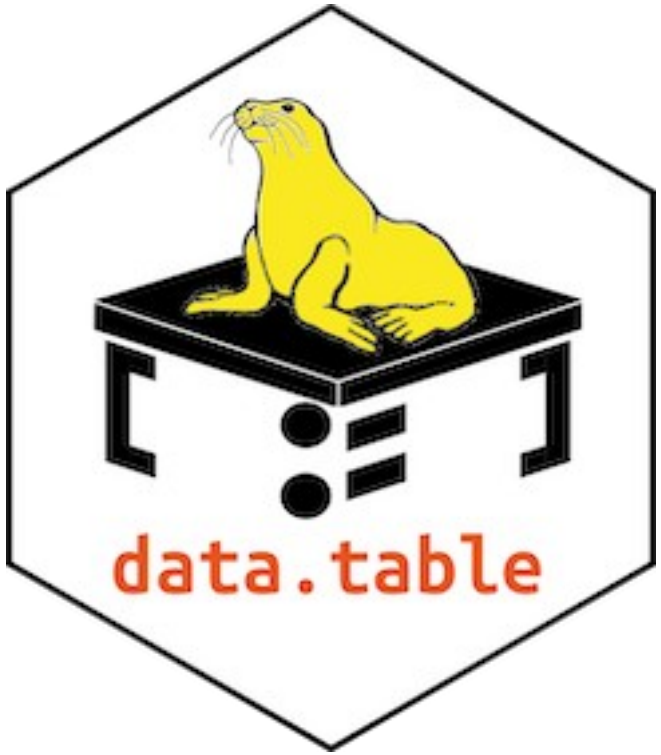


EXERCISE1: DPLYR



https://www.datacamp.com/courses/data-manipulation-with-dplyr-in-r

EXERCISE 2: DATA.TABLE

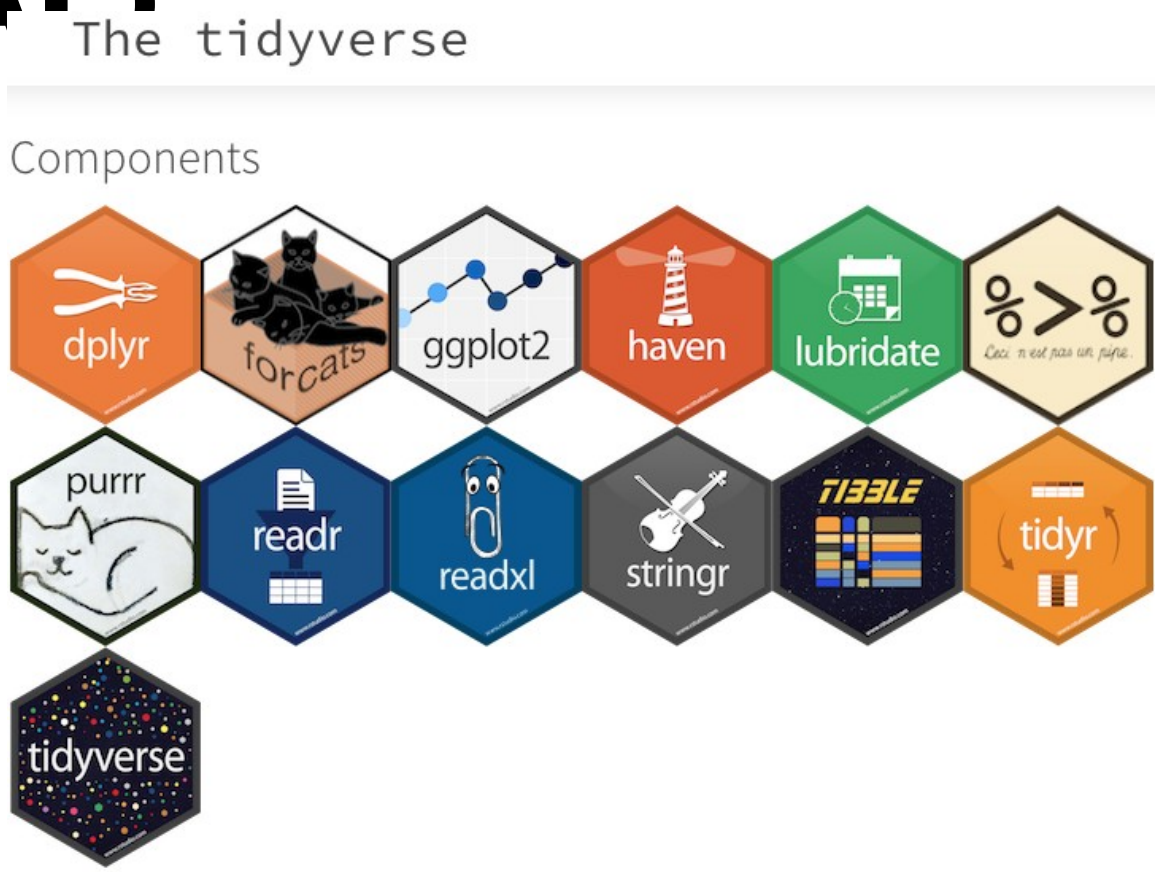


https://campus.datacamp.com/courses/data-manipulation-with-datatable-in-r

EXERCISE 3: TIBBLE

https://rstudio.cloud/learn/primers/2.1

EXERCISE 4: TIDYVERSE (OPTIONAL)



https://campus.datacamp.com/courses/working-with-data-in-the-tidyverse

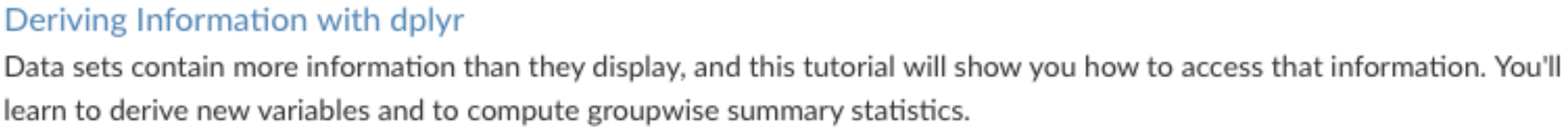
EXERCISE 5: ISOLATING DATA WITH DPLYR (OPTIONAL)

Isolating Data with dplyr

Master three simple functions for finding, and extracting, the data in your data set. Here you will learn to select variables, filter observations, and arrange values. Here, you will also meet R's pipe operator, %>%.

https://rstudio.cloud/learn/primers/2.2

EXERCISE 6: DERIVING INFORMATION WITH DPLYR (OPTIONAL)



https://rstudio.cloud/learn/primers/2.3

GITHUB

Chapter 12 Connect RStudio to Git and GitHub

Here we verify that RStudio can issue Git commands on your behalf. Assuming that you've gotten local Git to talk to GitHub, this means you'll also be able to pull from and push to GitHub from RStudio.

https://happygitwithr.com/rstudio-git-github.html

Chapter 6, 7, 8, 9 and 12

NEXT TIME

- Date: 30.09.2021
- Before class: finish tutorial on Datacamp/Rstudio.cloud primers + connect your Rstudio to Github
- Topic: Data visualization
- Exercise: Visualizations in own R scripts.

DEPARTMENT OF MANAGEMENT
AARHUS UNIVERSITY