

rbtl - Data wrangling with dplyr

Lars Schöbitz

Global Health Engineering - ETH Zurich

2022-05-12

Today

1. Notes for the exam
2. Homework Assignment 11
3. Data wrangling with dplyr
 - Live Coding Exercise
4. Data wrangling with dplyr
 - Programming Exercise
5. Homework Assignment 12

Learning Objective

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

Notes for exam

Notes for exam - practise!

- tempting to copy/paste (especially from others)
- practise as much as you can
- read instructions carefully
- identify instructions are phrased

Notes for exam - levels of difficulty

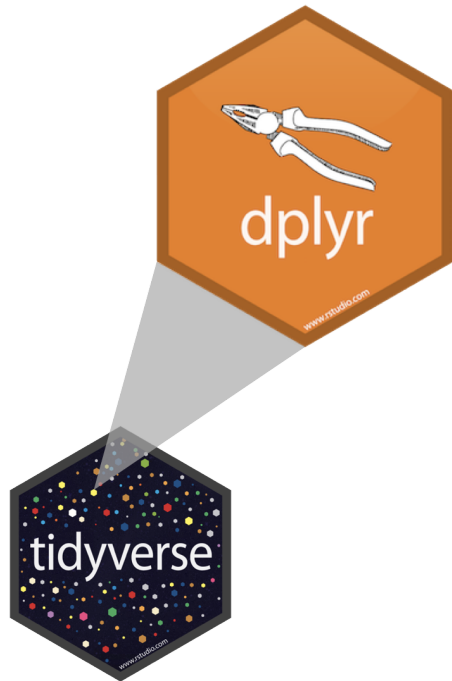
- fill in the blanks
- detailed instructions with named functions
- basic instructions with analysis goals

Homework Assignment 11

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 - Star Wars Characters

Rows: 87

Columns: 14

```
$ name      <chr> "Luke Skywalker", "C-3PO", "R2-D2", "Darth V...
$ height    <int> 172, 167, 96, 202, 150, 178, 165, 97, 183, 1...
$ mass      <dbl> 77.0, 75.0, 32.0, 136.0, 49.0, 120.0, 75.0, ...
$ hair_color <chr> "blond", NA, NA, "none", "brown", "brown, gr...
$ skin_color <chr> "fair", "gold", "white, blue", "white", "lig...
$ eye_color  <chr> "blue", "yellow", "red", "yellow", "brown", ...
$ birth_year <dbl> 19.0, 112.0, 33.0, 41.9, 19.0, 52.0, 47.0, N...
$ sex        <chr> "male", "none", "none", "male", "female", "m...
$ gender     <chr> "masculine", "masculine", "masculine", "masc...
$ homeworld  <chr> "Tatooine", "Tatooine", "Naboo", "Tatooine",...
$ species    <chr> "Human", "Droid", "Droid", "Human", "Human",...
$ films      <list> <"The Empire Strikes Back", "Revenge of the...
$ vehicles   <list> <"Snowspeeder", "Imperial Speeder Bike">, <...
```

Live Coding Exercise

ae-12-data-transformation-dplyr

1. Head over to the GitHub Organisation for the course.
2. Find the repo for week 11 that has your GitHub username.
3. Clone the repo with your username to the RStudio Cloud.
4. Open the file: `ae-12a-dplyr.qmd`
5. Use your Sticky Notes to let me know when you are ready.

Break One



15:00

Photo by [Blake Wisz](#)

Break Two



10:00

Photo by [Blake Wisz](#)

Homework Assignment

Submission

- All details in assignment week 12
- Due: Wednesday, 19th May at 23:59 (2 points)

Evaluation

- 5 mins
- anonymous
- after each lecture

kutt.it/rbtl-eval

Programming

ae-12-data-transformation-dplyr

1. Open the file: `ae-12b-dplyr.qmd`
2. Work through the exercises
3. Finalise as part of your homework

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](#).