

LS 2026 NPFL112

! YOUR MAIN RESOURCES

- RStudio on Jupyter lab: <https://aic.ufal.mff.cuni.cz/jlab/>
- DataCamp: https://www.datacamp.com/users/sign_in
 - Starting Feb 19, 2026, you have premium access for six months (i.e. until Aug 18, 2026). Then your account falls back to the free tier.
- This website: <https://ufal.github.io/NPFL112/>

1. Feb 20

Goals

- access computer network in the lab (save your credentials)
- access the Jupyter Hub RStudio (save your credentials)
- access DataCamp (create your account with the same e-mail address with which you enrolled)
- Overview of
 - the course scope
 - grading requirements
- Know where to find presentations online:
 - <https://github.com/ufal/NPFL112?tab=readme-ov-file> or
 - Press **s** to display speaker view with ample fluent notes
- Know how to download presentations in other formats (pdf, html pages) from RStudio on Jupyter Hub
 - `SCRIPTS.NPFL112/pages`
 - `SCRIPTS.NPFL112/pdf`

Presentations

https://ufal.github.io/NPFL112/slides/01_Introduction.html static html pdf

https://ufal.github.io/NPFL112/slides/02_HowToRStudio.html

Activities

In R Studio

1. Log in at RStudio
2. In the Files tab (right bottom pane), create a new directory (folder) and call it - exactly! - `SUBMISSIONS.NPFL112` .
3. Explore the folders `SCRIPTS.NPFL112` , `DATA.NPFL112`
4. Download to your computer `~/SCRIPTS.NPFL112/pdf/04_NavigatingRStudioForProgramming.pdf`.
5. Log in at https://www.datacamp.com/users/sign_in (you should have received an invite from the DataCamp system Feb 19, around 9:30 PM). Sign up with the same e-mail address with which you have enrolled this course. If there is time left, start with your first home assignment.

Assignments

 Assignment 1: Deadline: Thu Feb 26, 9:00 AM

In RStudio, make sure that you have created your folder called `SUBMISSIONS.NPFL112`. Open the folder `HOMEWORK_ASSIGNMENTS`. Select (tick) the file `HW_001.R` and copy it into your `SUBMISSIONS.NPFL112` folder.

Open this file, read the instructions and proceed accordingly. Go through the entire file but do not spend more than 30 minutes with it. If you use AI, only adopt responses that you have fully understood. The goal is to make you *notice, learn, and internalize* something - or make your initial self-assessment.

 Assignment 2: Deadline: Fri Mar 6, 9:00, but aim at Fri Feb 27

<https://campus.datacamp.com/courses/free-introduction-to-r/chapter-1-intro-to-basics-1?ex=1>

2. Feb 27

Goals

1. Internalize the following concepts:
 1. Data types/classes (numeric, character, boolean)
 2. Data type coercion (what happens to your numeric vector when you blend in a non-digit, etc.)
 3. Data structures (vectors, data frames/tibbles)
 4. Functions and their arguments
 5. Working directory
2. Learn to invoke and read the built-in R Help
3. Open a Quarto (.qmd) file and run code chunks manually

Presentations

https://ufal.github.io/NPFL112/04_NavigatingRStudioForProgramming.html static html pdf

https://ufal.github.io/NPFL112/05_VariablesFunctions.html static html pdf

https://ufal.github.io/NPFL112/06_WorkingDirectory.html static html pdf

Activities

1. Lecture start: group work, exchange about HW_001.R: During your homework, which structures/patterns caught your eye?
2. Hands-on together: Copy the file `SCRIPTS.NPFL112/05_VariablesFunctions.qmd` to your home directory.
 1. Open it and run (execute) all code chunks. Watch what happens.
 2. When you are done with executing the chunks, find the “Render” button. Explore the rendering options. Render the file as an html file. It will appear in the same directory where you had the corresponding qmd file and will inherit its name. Open the resulting html file in a web browser.
 3. Open a new Quarto .qmd file. Render it by hitting the Render button.

Assignments

🔥 Assignment 3. Deadline: Fri Mar 6, 9:00 (always entire chapters!)

<https://campus.datacamp.com/courses/free-introduction-to-r/chapter-2-vectors-2?ex=1>

<https://campus.datacamp.com/courses/free-introduction-to-r/chapter-5-data-frames?ex=1>

<https://campus.datacamp.com/courses/free-introduction-to-r/chapter-4-factors-4?ex=1>

3. Mar 6

Goals

1. Internalize the following concepts:
 1. Vector recycling
 2. Vector element vs. vector position index
 3. logical operators (>, <, >=, <=, !=, ==, &, |)
2. Know how to:
 1. Extract elements from a vector (by position or by a condition expressed by a logical operator).
 2. Extract values from data frames by rows and columns in base R.
 3. Read a tabular file (.csv) into a data frame object.
 4. Write a data frame object into a tabular file (.csv)
3. Recognize transformations of tabular data (no coding): filtering rows, selecting columns, aggregations, aggregations in groups

Presentations

Finish https://ufal.github.io/NPFL112/05_VariablesFunctions.html (from slide *No Coercion with errors*).

i Activity to new topic

table transformations, see below

https://ufal.github.io/NPFL112/07_Exploring_dataframes.html static html pdf

Activities

Look at the (printed) tables and keep them ready at your hands. The tables are bits of gapminder data, also file `clips_handouts.pdf` in `SCRIPTS.NPFL112` and https://ufal.github.io/NPFL112/clips_handouts.pdf. Each sheet contains a pair of tables. What do you need to do to the table on the left to obtain the table you see to the right? Try your luck in a multiple choice test here: https://quest.ms.mff.cuni.cz/class-quiz/quiz/NPFL112_02_01_clips

and check out whether you can describe most of the transformations - in your own words, no scientific jargon, no coding. Note the difficult ones and ask about them after the exercise. Duration: 10 - 15 minutes.

Assignments

 Assignment 4: Deadline: Fri Mar 20

<https://app.datacamp.com/learn/courses/introduction-to-the-tidyverse?cf-exp=comm--ai-native-cdp-layout%3Aon>

4. Mar 13

Goals

- Understand these `ggplot2` concepts:
 - Aesthetic scales and their mapping to variables
 - Mapping inheritance (where to write the aesthetic mappings)
 - *Geom/geometry* in `ggplot2`.
- Learn about the `ggplot2` cheatsheet: <https://rstudio.github.io/cheatsheets/data-visualization.pdf> (also in RStudio Help!)

Presentations

https://ufal.github.io/NPFL112/08_DiversePlots.html static html pdf

Activities

Mentimeter WarmUp: <https://www.menti.com/>

Teacher's GUI: <https://www.mentimeter.com>, presentation *Dplyr and ggplot2 for beginners*. Needs login and the presentation must be launched.

Your first data report in Quarto: tracking the world's billionaires

1. Open a new Quarto (.qmd) file (a Document, but Presentation would be fine, too).
2. In the dialog window, give the document the title “Billionaires Investigation” and fill out your name.
3. Save the file into SUBMISSIONS.NPFL112 under the name `my_first_data_report`.
4. Check out the YAML header. Add today's date to it with this line: `date: today`
5. Hit the **Render** button (top of the pane). If you did it correctly, RStudio's bottom right pane expands with the **Viewer** tab fronted, showing your rendered html file. Switch to the **Files** tab and find the resulting html file that **Viewer** is showing you. Can you see the date, too?
6. Get back in the top-left pane and start editing your Quarto document. Have a look at the template in both the **Source** and the **Visual** editor mode (toggle these options in the top left corner). Particularly inspect the Running Code section and hit the green arrow/triangle to run the code inside. Once you have read the text in the template, you can erase it.
7. Insert a code chunk with `Ctrl + Alt + i` or with the green `+C` button in the top right corner of your pane.
8. Type into the chunk the code to mount/activate the necessary libraries `dplyr`, `readr`, and `ggplot2`.

💡 Call all *tidyverse* libraries simultaneously

You can as well load all tidyverse libraries simultaneously by calling `library(tidyverse)`.

- In the same chunk, create a variable called `df` (to remember that it is a “data frame”) by reading in the file with the following path: `~/DATA.NPFL112/billionnaires_combined.tsv`. It is a `.tsv` file (tab-separate values), so you need the `read_tsv` function from the `readr` library.

 Understand the file-reading functions of the `readr` library

There are dedicated functions for specific formats, such as the U.S. `.csv`, the European `.csv`, and `.tsv`. All these formats capture tables in plain text by a convention of separating columns with a specific delimiter character (comma, semicolon, or tabulator). They all are derived from the `read_delim` function. You can try it out with this very file. Mind to set the `delimiter` argument to `tabulator`. Check out the other arguments of these functions in their documentation!

- Explore the data frame `df`. Create another code chunk and print first ten rows of this data frame.
- Overwrite `df` so that it will only contain data from the year 2022.
- Create yet another code chunk. In this code chunk, write the code to print a histogram of the wealth distribution among the billionaires (column `daily_income`). Experiment with the `binwidth` argument of `geom_histogram`.
- Now create a new code chunk that prints a scatterplot with `age` on the x-axis and `daily_income` on the y-axis. Can you observe any relation between age and daily income? E.g. do senior billionaires appear to earn more per day than the younger ones?
- Above each chunk with a plot add a heading formatted as Heading 2 and a short text describing for your audience what the given plot demonstrates.
- Render your document without changing anything in the YAML header.
- Now replace `html` with `pdf` in the YAML header and hit the Render button again. If you get an error message, install a package called `tinytex` and try again without explicitly calling/mounting/activating it. RStudio ought to do it by itself. If you do not get a pdf file, give up. Make a note about this failure in your file and re-render it as html. That always works.

Assignments

Assignment 5: Deadline: Mar 27

Finish the billionaires exercise from this lecture and place it into your SUBMISSIONS.NPFL112 folder.

Assignment 6: Deadline: April 8, noon

Do [this exercise](#) and place the resulting file (R or .qmd from a Quarto file) into your SUBMISSIONS.NPFL112 folder. If you produce a .qmd file (you are encouraged to!), please also add a rendered file (html or pdf).

5. Mar 20

Goals

Wrap up operations on a single table with `dplyr`.

Presentations

https://ufal.github.io/NPFL112/09_Aggregations_with_dplyr.html static html pdf

https://ufal.github.io/NPFL112/11_Computations_mutate_with_dplyr.html static html pdf

Activities

Look again at the [worksheets with data frames](#). Do not take them out of their plastic sleeves. From the label sheet, pick for each data frame sheet a label that encodes the transformation of the left table into the right table.

Assignments

 Assignment 6: Deadline: Mar 27

<https://app.datacamp.com/learn/courses/introduction-to-data-visualization-with-ggplot2?cf-exp=comm--ai-native-cdp-layout%3Aon>

6. Mar 27

Double lecture (Vaclav missing)

Goals

Presentations

Activities

Group work on paper: interpret someone else's code in greatest possible detail. Hazard guesses! The code has one issue in it. Will you find it? Compare your results with other groups. In-person lecture: code in hardcopy. Here is a [link to the code in pdf](#).

Assignments

 Assignment : Deadline:

ddd

Good Friday Apr 4

Goals

Presentations

Activities

Assignments

 Assignment : Deadline:

ddd

7. Apr 10

Deadline for HW02 was April 8, noon.

Goals

Presentations

Activities

Assignments

 Assignment : Deadline:

ddd

8. Apr 17

Goals

Presentations

Activities

Assignments

 Assignment : Deadline:

ddd

9. Apr 24

Goals

Presentations

Activities

Assignments

 Assignment : Deadline:

ddd

May 1, May 8: two-week gap

Goals

Presentations

Activities

Assignments

 Assignment : Deadline:

ddd

10. May 15

Goals

Presentations

Activities

Assignments