TDS
Facility

Introduction to R for Life Sciences

João Lourenço, Tania Wyss & Nadine Fournier

Translational Data Science – Facility

SIB Swiss Institute of Bioinformatics

With slides from Diana Marek, Thomas Junier, Wandrille Duchemin, Leonore Wigger

From: First steps with R in Life Sciences

The Translational Data Science Facility

- Part of the **SIB Swiss Institute of Bioinformatics**
- Located at the AGORA Cancer Research Center in **Lausanne**
- Provides **statistics, bioinformatics and computational expertise** to molecular biology and applied research labs.
- Participates in fundamental and translational research by providing expertise in **data analysis** of single-cell and bulk multi-omics, spatial transcriptomics, flow cytometry, etc

For core facility service inquiry: nadine.fournier@sib.swiss
https://agora-cancer.ch/scientific-platforms/translational-data-science-facility/
https://www.sib.swiss/raphael-gottardo-group

Tell us about yourself !

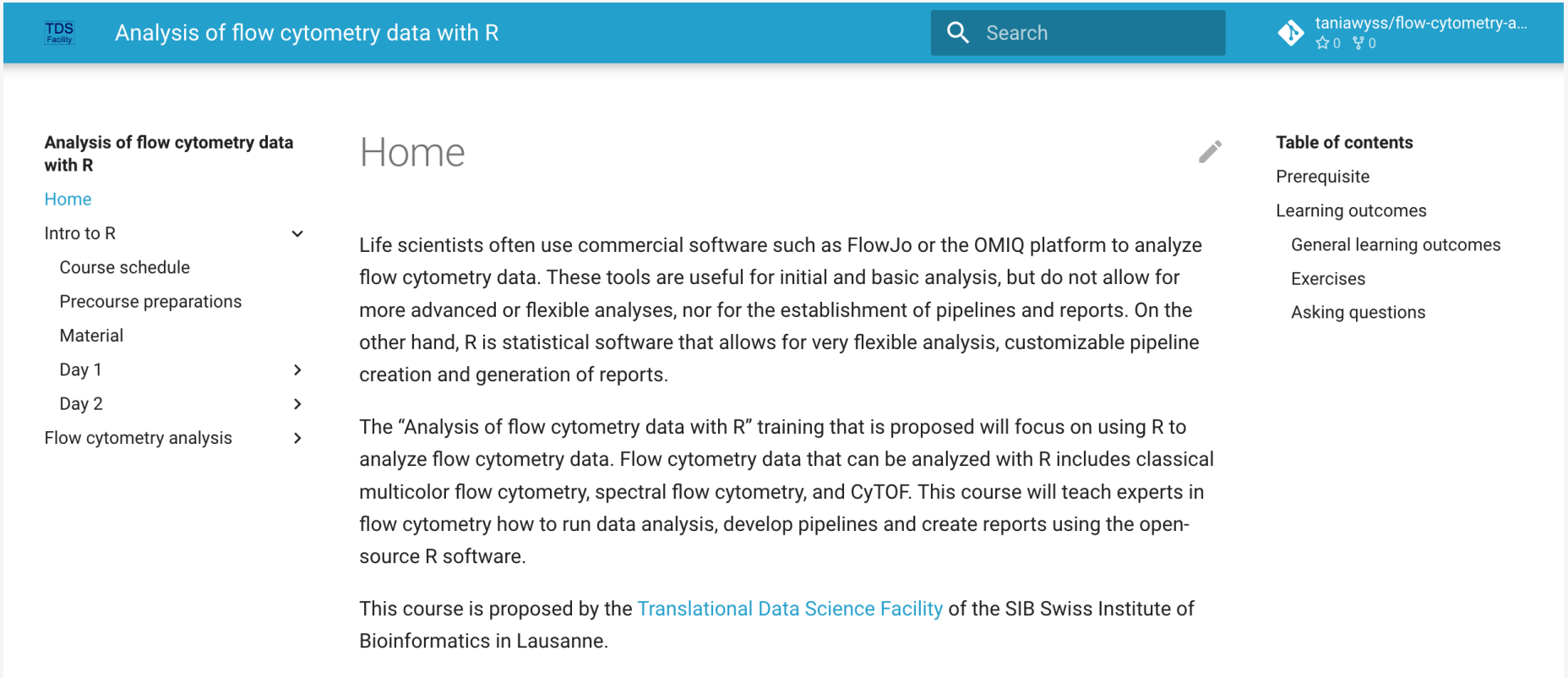
Share about yourself and your research, experience with programming, etc

Photo by National Cancer Institute, Unsplash

Photo by Scott Graham, Unsplash

Course material

1. Website
https://taniawyss.github.io/intro-to-R/



2. Ask us questions!

Outline & Schedule

Morning

01

Introduction to R and the RStudio environment, working with scripts files
Exercises
(9:00 – 10:30)
10:30 – 10:50 Coffee break

02

Syntax, data types and structures, importing data
Exercises
(10:50 -12:00)
12:00 – 13:00 Lunch break

Outline & Schedule

Afternoon

03

Graphics
Exercises
(13:00 – 15:30)

04

15:30 -15:50 Coffee break

04

Statistics
Exercises
(15:50 – 16:50)

16:50 - 17:00 Feedback and end of day

Course Content

R is vast and can't be learned in one day. The scope of this course is to:

- Give you a basic understanding of concepts behind R
- Allow you to import and manipulate data in R
- Show you how to create your first plots

This course is only the first step in your R journey!

01 Introduction to R and the RStudio environment

What is R ?

- R is a programming language and an environment for statistical computation and graphics.

  - A simple development environment with a console and a text editor

  - Facilities for data import, manipulation and storage
  - Functions for calculations on vectors and matrices
  - Large collections of data analysis tools
  - Graphical tools

https://www.r-project.org/

R's user community

- Group of core developers who maintain and upgrade the basic R installation. New version every 6 months.

- Anyone can contribute with add-on packages which provide additional functionality (thousands of such packages available) and help for each function.

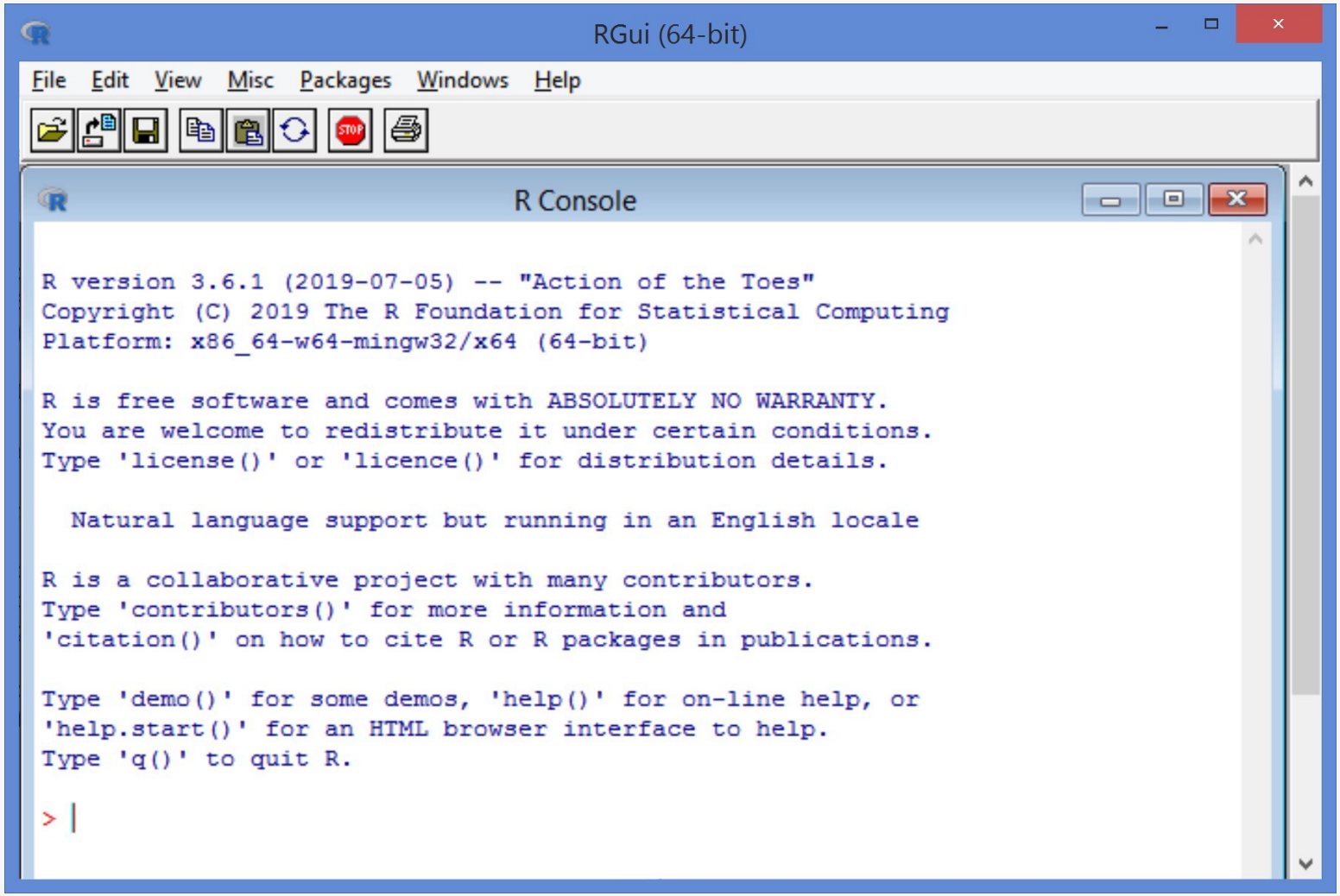
- Online help

  - in user group forums, *eg*:
  https://stat.ethz.ch/mailman/listinfo/r-help
  http://stackoverflow.com/questions/tagged/r

  - in countless online tutorials, books, blogs

RGui (R Graphical user interface)

- Together with the programming language, a (minimal) graphical user interface is installed.



R Combined with RStudio

https://posit.co/products/open-source/rstudio/

RStudio is an integrated development environment (IDE), designed to help you be more productive with R

It includes:

- A console
- A syntax-highlighting editor that supports direct code execution
- Tools for viewing the workspace and the history
- A file explorer, a package explorer, plot and help display areas

We suggest RStudio as a more powerful, more comfortable alternative to the RGUI

RStudio interface

RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help 93% Ven. 10:56 Diana Marek

~/TrainingSIB/Courses_2017/First_Steps_R_Oct2017 - RStudio

Go to file/function Addins First_Steps_R_Oct2017

first_script.R test.R examples.R mice_study.R

Source on Save Run Source

```
1  #### My first script #####
2  #### October 2017   #####
3
4
5  # list workspace
6  ls()
7
8  # Reset R'brain
9  rm(list=ls())
10
11 # check wd
12 getwd()
13
14 #set wd
15 setwd("/Users/dmarek/EducationSIB/Courses_2016/First_Steps_R_June2016/R_intro_course")
16
17 # confirm wd
18 getwd()
19
20 #load packages if needed (to do every time you launch your R session)
21 #library("boot")
22 #library("lattice")
```

2:26 October 2017 R Script

Editor
(scripts)

Console Terminal

~/TrainingSIB/Courses_2017/First_Steps_R_Oct2017/

```
R est un logiciel libre livré sans AUCUNE GARANTIE.
Vous pouvez le redistribuer sous certaines conditions.
Tapez 'license()' ou 'licence()' pour plus de détails.

R est un projet collaboratif avec de nombreux contributeurs.
Tapez 'contributors()' pour plus d'information et
'citation()' pour la façon de le citer dans les publications.

Tapez 'demo()' pour des démonstrations, 'help()' pour l'aide
en ligne ou 'help.start()' pour obtenir l'aide au format HTML.
Tapez 'q()' pour quitter R.

[Workspace loaded from ~/TrainingSIB/Courses_2017/First_Steps_R_Oct2017/.RData]

>
```

Console
(or terminal)

Environment History Connections

Import Dataset List

Global Environment

Data	
mice_data	50 obs. of 3 variables
mice_weight_HFD	29 obs. of 3 variables
mids	num [1:2, 1] 0.7 1.9
Values	
mean_weight_diet	num [1:2(1d)] 28.7 37.1
mean_weight_genotype	num [1:2(1d)] 33.7 33.4
n_weight_diet	int [1:2(1d)] 21 29
n_weight_genotype	int [1:2(1d)] 24 26
sd_weight_diet	num [1:2(1d)] 2.61 5
sd_weight_genotype	num [1:2(1d)] 4.69 6.92

Workspace
(Environment and History)

Files Plots Packages Help Viewer

New Folder Delete Rename More

Home > TrainingSIB > Courses_2017 > First_Steps_R_Oct2017 > course_datasets

Name	Size	Modified
..		
class.csv	402 B	Mar 6, 2016, 6:33 PM
etubiol.csv	2.4 KB	Mar 7, 2016, 12:13 PM
mammals_survey.csv	1.2 MB	Jan 30, 2017, 7:41 AM
melanoma_data.txt	5.9 KB	Jan 30, 2017, 7:41 AM
mice_data.csv	632 B	Feb 28, 2016, 4:15 PM
my_data_frame.csv	262 B	Mar 2, 2015, 4:49 PM
pigs.csv	206 B	Mar 6, 2016, 6:32 PM
smoker.csv	4.6 KB	Feb 26, 2015, 12:23 PM
snp.csv	618 B	Mar 6, 2016, 9:19 PM

File explorer, plots, packages, help

Console: The Command Line

```
~/TrainingSIB/Courses_2017/First_Steps_R_Oct2017/

R est un logiciel libre livré sans AUCUNE GARANTIE.
Vous pouvez le redistribuer sous certaines conditions.
Tapez 'license()' ou 'licence()' pour plus de détails.

R est un projet collaboratif avec de nombreux contributeurs.
Tapez 'contributors()' pour plus d'information et
'citation()' pour la façon de le citer dans les publications.

Tapez 'demo()' pour des démonstrations, 'help()' pour l'aide
en ligne ou 'help.start()' pour obtenir l'aide au format HTML.
Tapez 'q()' pour quitter R.

[Workspace loaded from ~/TrainingSIB/Courses_2017/First_Steps_R_Oct2017/.RData]

>
```

The prompt ">" indicates that R is waiting for you to type a command

Try it out...

Type the following at the command prompt:

Simple calculations

```
> 1 + 1
```

Assign values to a variable names

```
> x  <- 128.5
```

Display content of variables

```
> x
```

Pre-defined functions

```
> abs(-11)
```

After each command, hit the return key.

This causes R to execute it.

Note the assignment operator <- with which we can keep values in the memory, by assigning a value and a name to a variable and store it in the session's memory.
We can use either <- or = to assign values to an object

Stick to one for consistency.

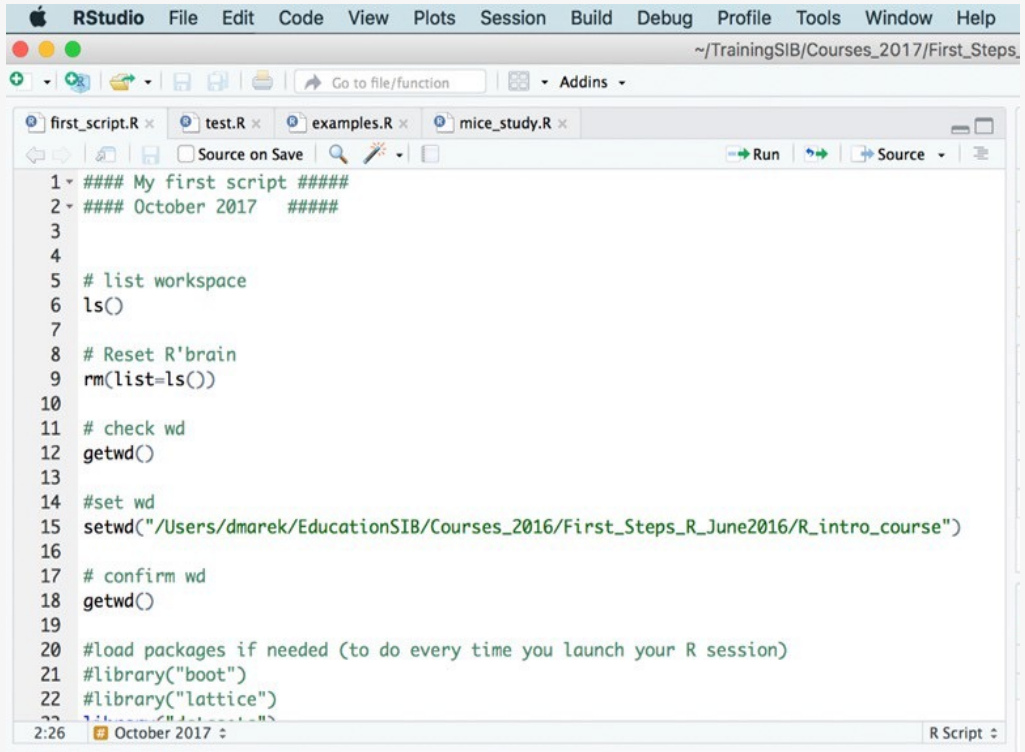
02 Working with script files

Editor: Write code to a script file

A script is a file that contains commands to be executed in succession.

Write your code into a script and save it

- to have documentation later of what you did
- to be able to re-use the code and create variations
- for easy execution



Notice the syntax highlighting

Create a new script and type code

- Create a new script using File > New File > R script. **Don't forget to save your script often.**
- By default, scripts are saved to the working directory.
- Files can be saved to other locations (File -> Save As...)
- Start **Typing code** at the top of the script

```
# My first command:
2 + 3
```

- **Notice the syntax highlighting**
- **Comments** : "#" at the beginning of a line or before a command: helping text ; everything that follows is ignored by the during executing ; R does not support multi-line comments

Send Code From a Script to the Console

Run individual lines, one by one:

- In RStudio: put the cursor anywhere in a line, hit

Ctrl + enter (Windows)

Cmd + return (Mac)

or click the "Run" button

Tip: Run part of a line or multiple lines: Highlight the code, then proceed as above

Save, close and open scripts

- **Save a script:** File > Save or
- **Close and open a script:** File > Close and File > Open File

Tips:

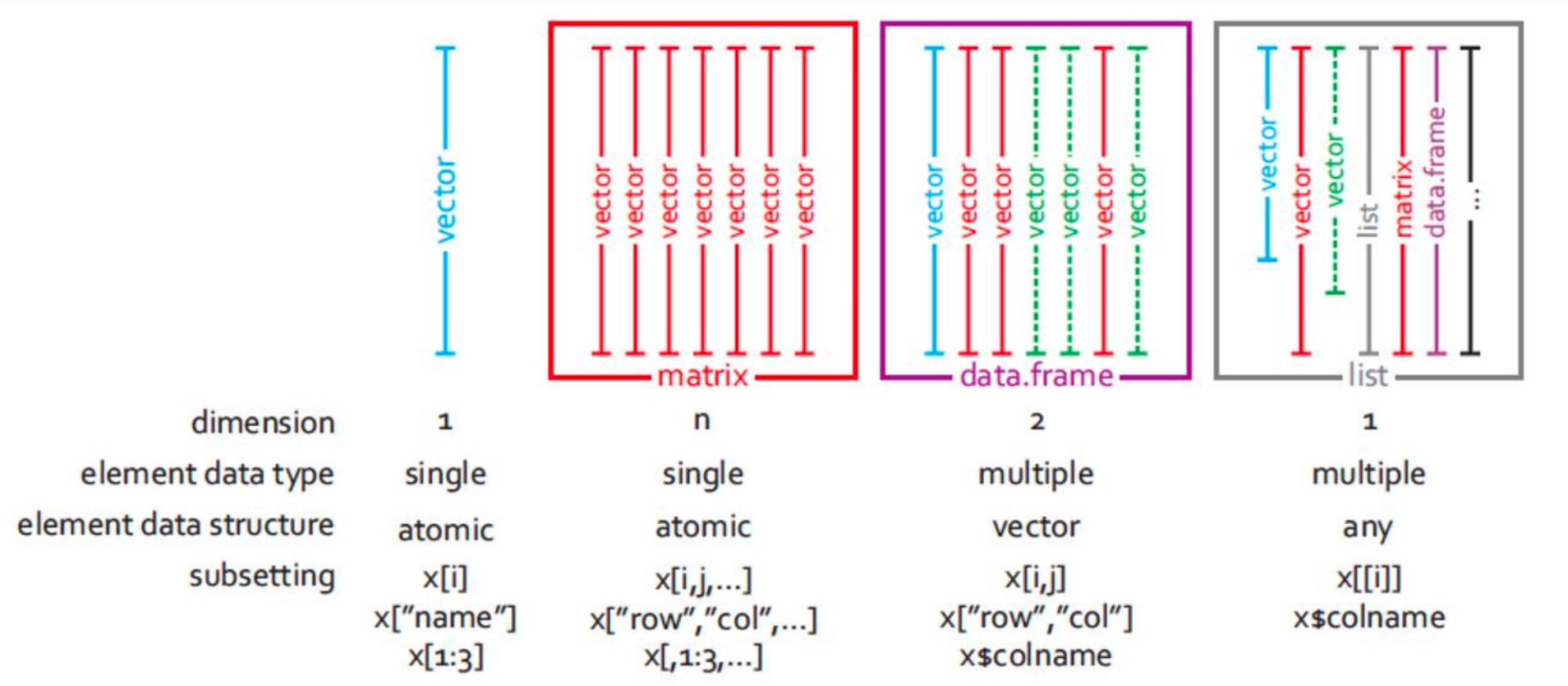
- Most of your code should be developed and saved in scripts.
  - You can execute individual lines of code interactively while you are writing it.
  - You can run the entire script once it is ready and debugged.

We continue working on the provided script

Download and open it

02 Syntax, data types and structures, importing data

Common object classes



Example of a well-formated dataset

	A	B	C	D
1	Sample_ID	Age	Sex	Disease
2	M417	71	male	Healthy
3	M244	73	female	Tumor
4	M255	60	male	Healthy
5	M229	75	male	Tumor
6	M420	68	female	Healthy
7	M368	73	male	Healthy
8	M403	68	male	Tumor
9	M230	56	male	Tumor
10	M370	84	male	Tumor
11	M406	69	male	Tumor
12	M245	70	male	Tumor
13	M409	NA	female	Tumor
14	M395AR_dm	67	male	Tumor
15	PB	57	male	Healthy
16	M318	62	male	Healthy
17	M423	72	female	Tumor
18	M398_DMOS	61	female	Tumor
19	M233	74	male	Tumor
20	M381	57	male	Healthy
21	M408	65	male	Tumor
22	M402	68	male	Healthy

Example of a spreadsheet in Excel

- A header line with variable names (4 variables, 1 in each column)
- No blank spaces in variable names (use _ instead)
- Variable names do not contain symbols other than _
- One observation per row
- No comments or other content around the data table
- Indicate missing values with NA

Additional learning and practicing

Wandrille Duchemin's First Steps with R in Life Sciences (2 days):
It includes more on statistics!
https://github.com/sib-swiss/first-steps-with-R-training/tree/master

Introduction to statistics with R (3 days), for R beginners also:
https://sib-swiss.github.io/Introduction-to-statistics-with-R/day1/

Introduction to R for Cancer Scientists
https://bioinformatics-core-shared-training.github.io/r-intro/index.html

Glittr.org

Glittr

Git repositories with
bioinformatics training material