



BIOC | Bioinformatics Competence Center

Bioinformatic Analysis of RNA-sequencing

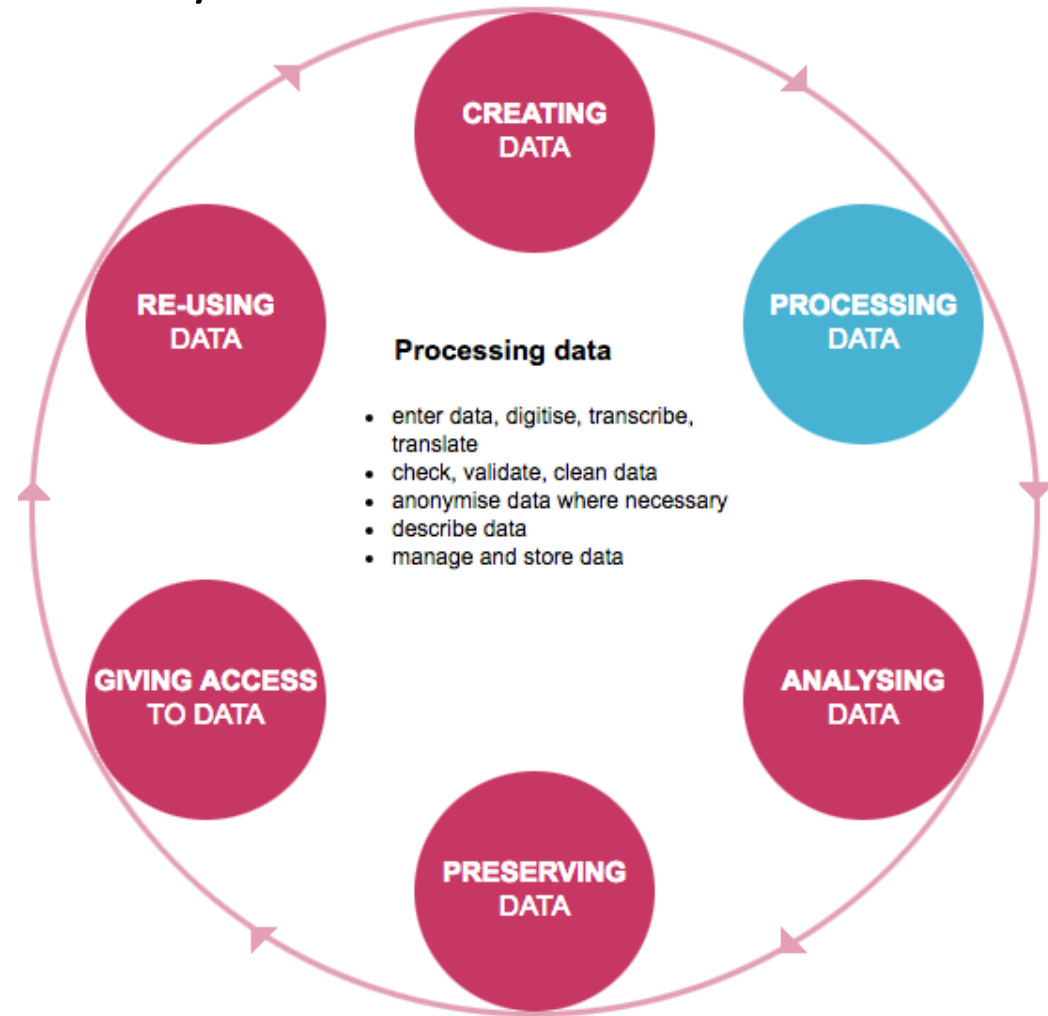
R refresh and Reproducible Research

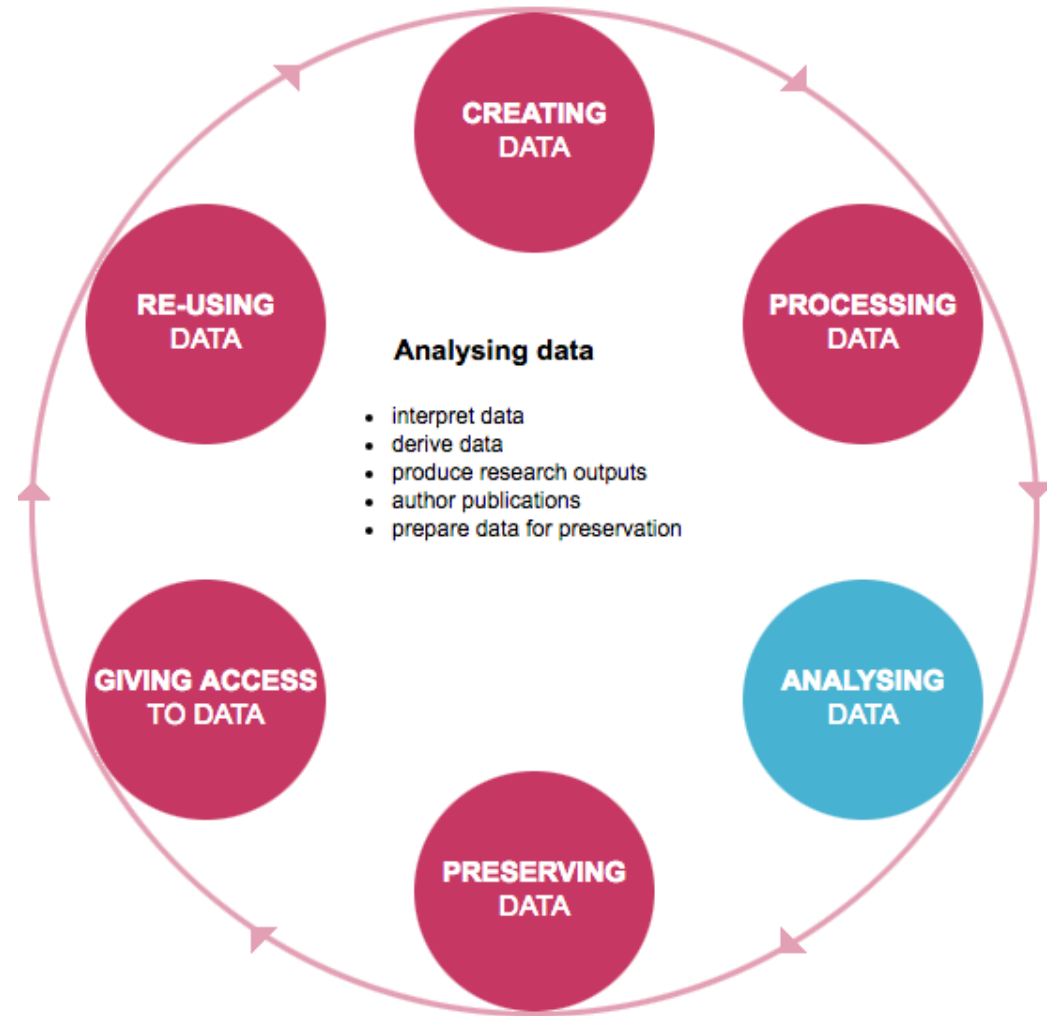
Allison Burns, Maxime Jan, Linda Mhalla,
Christian Iseli & Nicolas Guex
– summer 2023 –

Outlines

- Reproducible research
- Generate reports using R
- The basics of R
 - Variables
 - Manipulate data.frames and matrix
 - Functions
 - Read/write text and excel files
 - Plot data

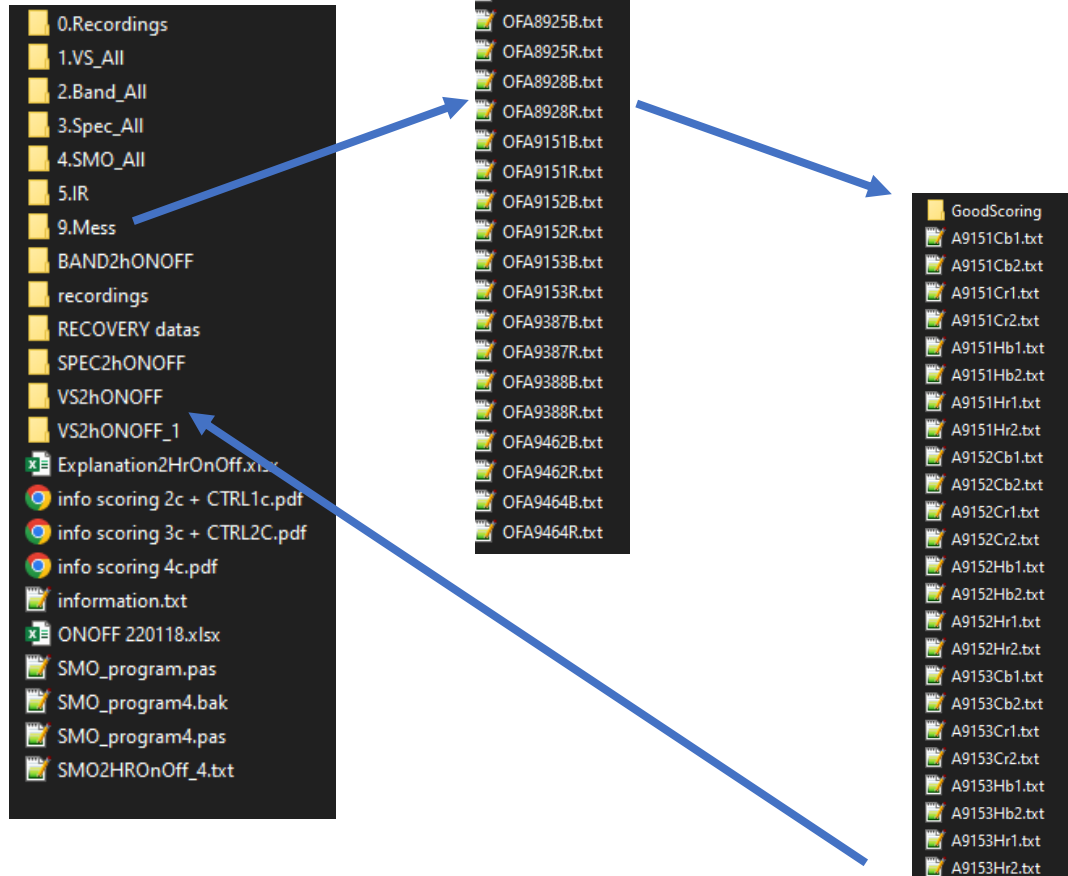
Reproducible research in data life-cycle





Reproducible analysis starts with files and directory

Some project directory can be messy !



Avoid duplicating files !

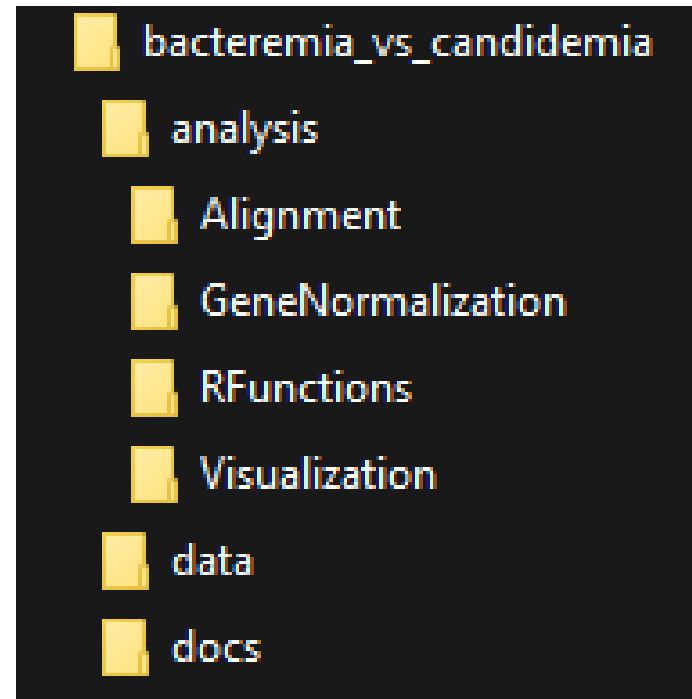
3 data files with identical names...

	A9151Cr2.txt	Taille : 63.2 Ko
	Modifié le : 18.01.2018 14:10	
	A9151Cr2.txt	Taille : 63.2 Ko
	Modifié le : 18.01.2018 14:10	
	A9151Cr2.txt	Taille : 63.2 Ko
	Modifié le : 18.01.2018 14:10	



Before analysing... try to be organized !

- One directory per project !
- Avoid duplicating data
- Avoid manual editing of RAW data
- Try to have some README files
- You can create 3 directories:
 - “Analysis” containing your code
 - “Data” containing RAW and/or processed data
 - “Docs” containing presentations, manuscript etc...



Think about files nomenclature

Differential expression of various glucocorticoid receptor mutants
during a time course experiment (5 conditions, 4 timepoints)

WT

G56R

G56R
heterozygote

R311Q

DNA binding domain deletion

Filenames we received -->

G1y
G6m
GhP21
GhP6m
GP10
GP21
Gplus
R1y
R6m
rd710
rd71y
rd721
rd76m
RP10
RP21
WT1y
WT6m
WTP10
WTP21

What could be rd710 and rd721 ?

sample	time	genotype	eye	days
WTP10	10days	WT	LE	10
WTP21	21days	WT	LE	21
WT6m	6months	WT	LE	183
WT1y	1year	WT	LE	365
GP10	10days	G56R_homo	LE	10
GP21	21days	G56R_homo	LE	21
G6m	6months	G56R_homo	LE	183
G1y	1year	G56R_homo	LE	365
RP10	10days	R311Q	LE	10
RP21	21days	R311Q	LE	21
R6m	6months	R311Q	LE	183
R1y	1year	R311Q	LE	365
rd710	10days	rd7	LE	10
rd721	21days	rd7	LE	21
rd76m	6months	rd7	LE	183
rd71y	1year	rd7	LE	365
Gplus	10days	G56R_het	LE	10
GhP21	21days	G56R_het	LE	21
GhP6m	6months	G56R_het	LE	183

Make names self explanatory

→ G56R_homo.LE.10d

→ G56R_homo.LE.1y

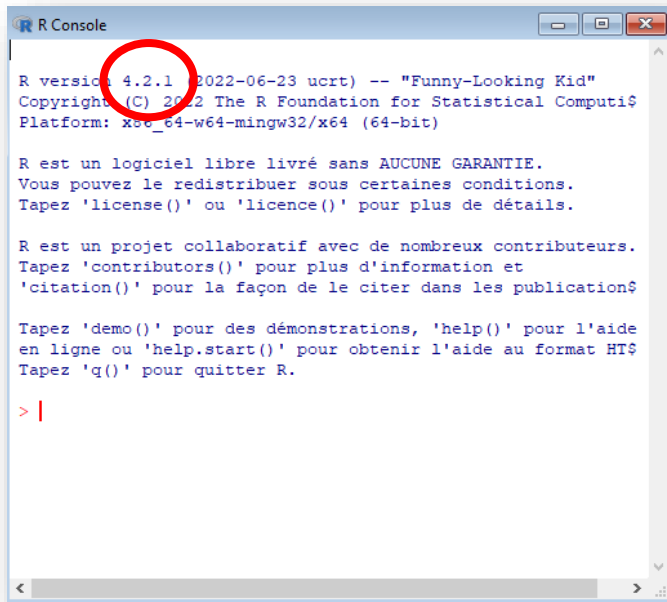
Sample nomenclature

- Be consistent
- Think about automated parsing with UNIX shell or R
- Avoid non alphanumeric characters: space “ ‘ [(- | # etc...
- Avoid starting with zeros (R will not like it)
- Use padding if necessary: ENSG00000135446 (keeps constant length)
- Plan for repeated experiments / sample acquisition

For example: {sample}_{condition}_{time}_{replicate}

Reproducible research with R

 ,  Studio® and **R Markdown**



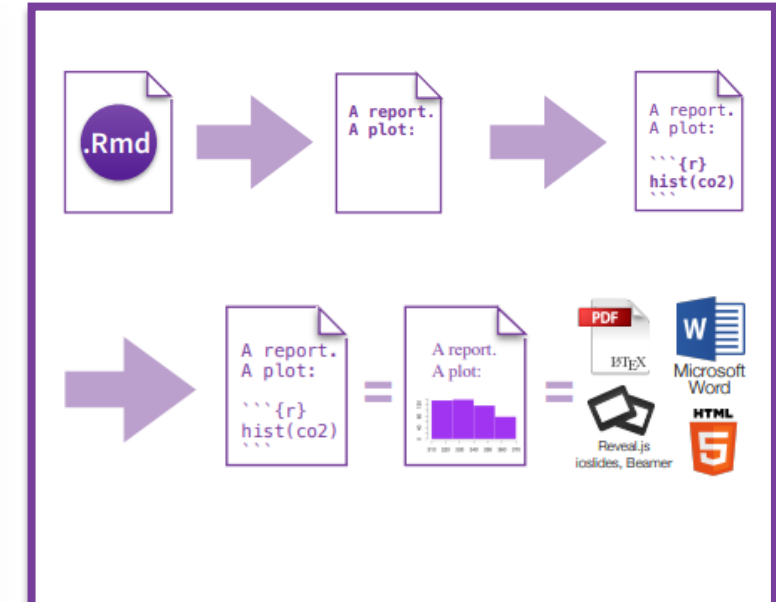
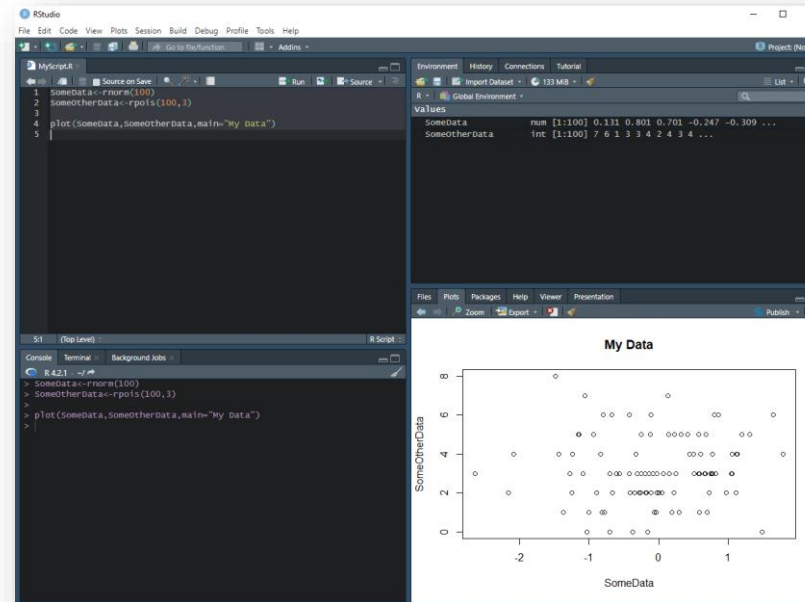
```
R version 4.2.1 (2022-06-23 ucrt) -- "Funny-Looking Kid"
Copyright (C) 2022 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

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.

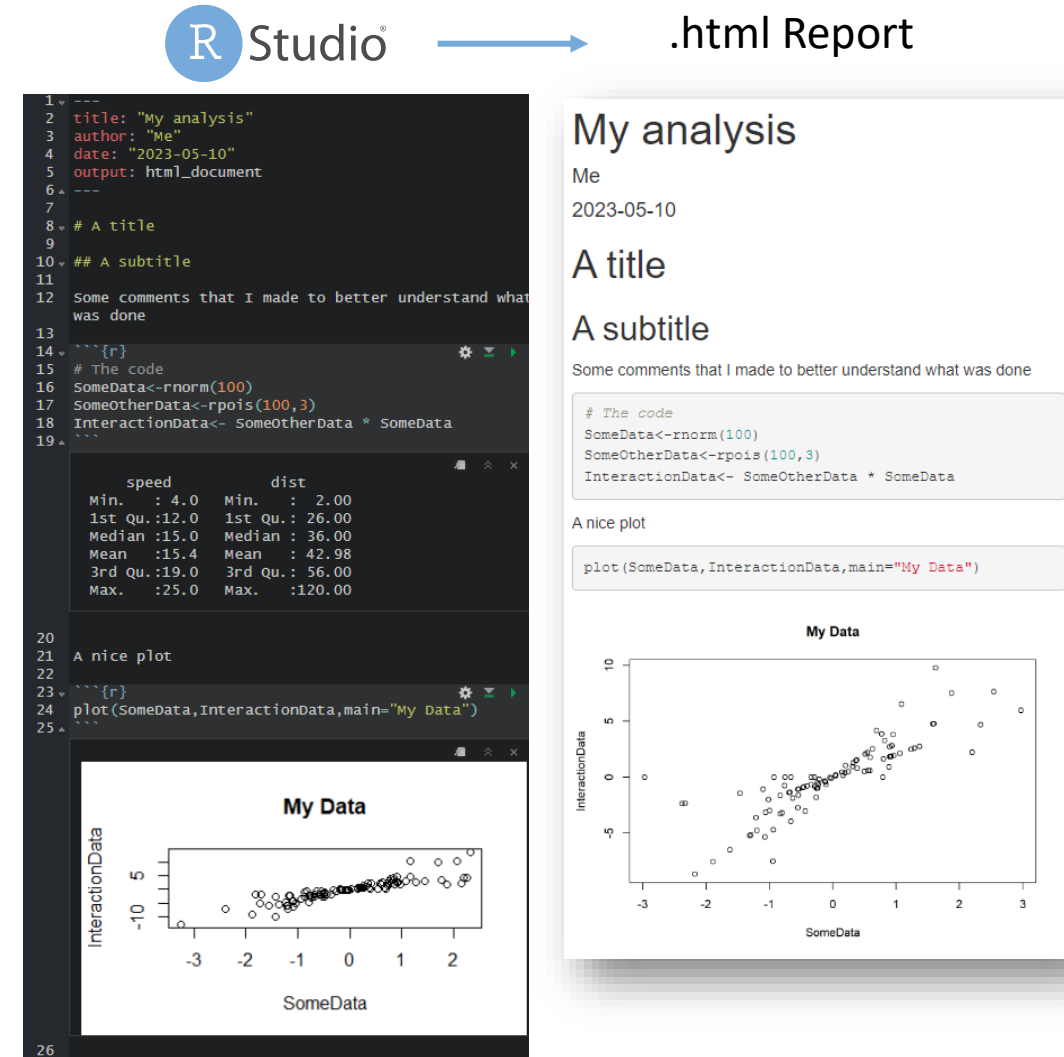
Tapez 'demo()' pour des démonstrations, 'help()' pour l'aide
en ligne ou 'help.start()' pour obtenir l'aide au format HTML.
Tapez 'q()' pour quitter R.

> |
```

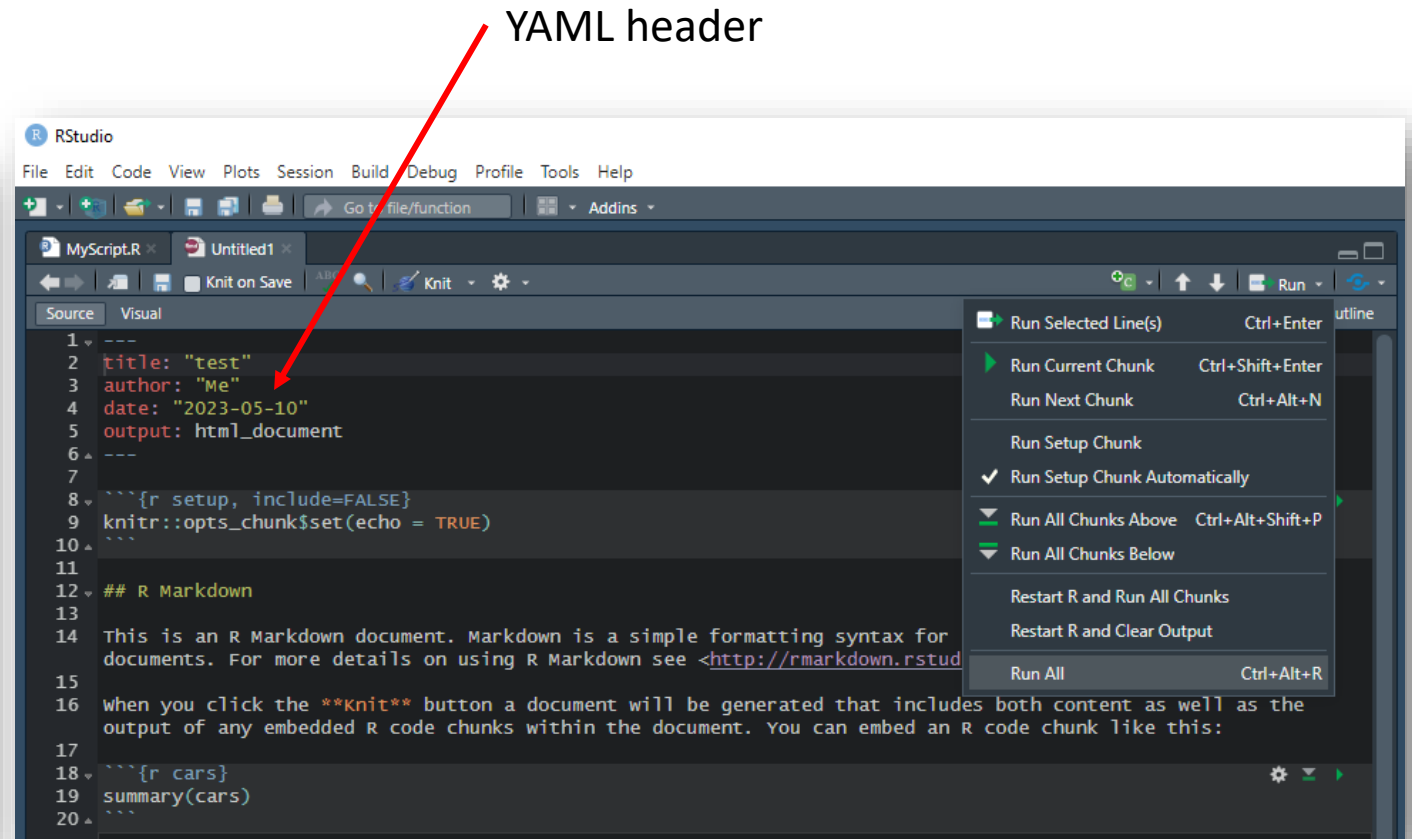
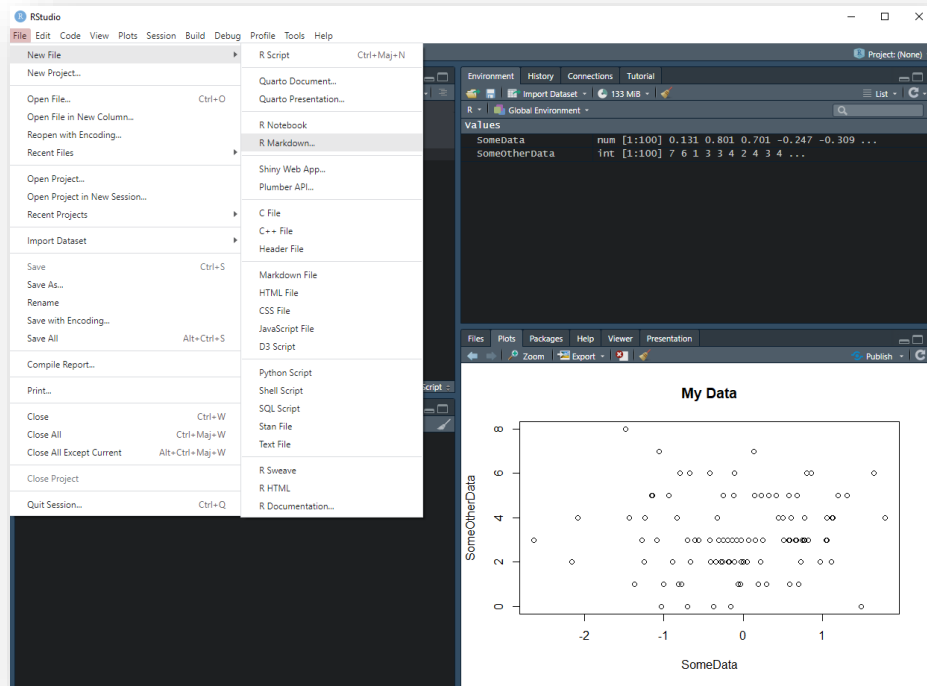


Reproduce analysis with Rmarkdown

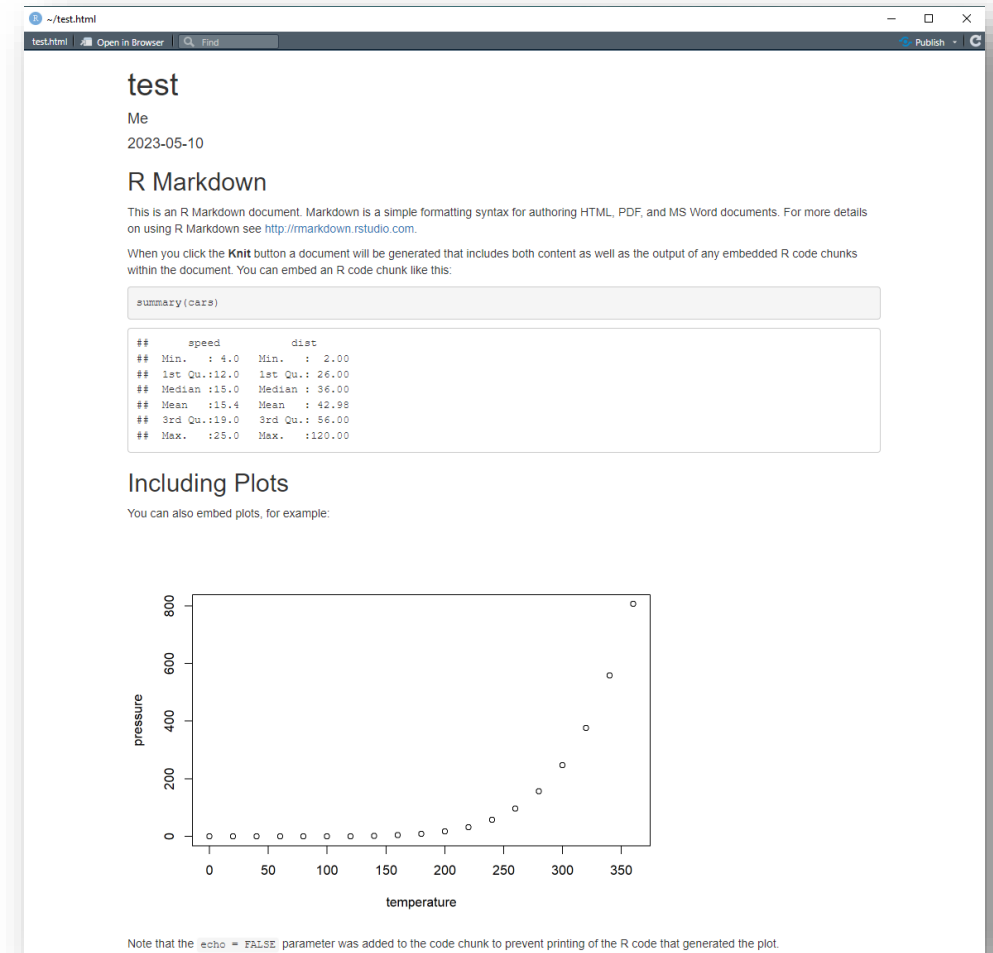
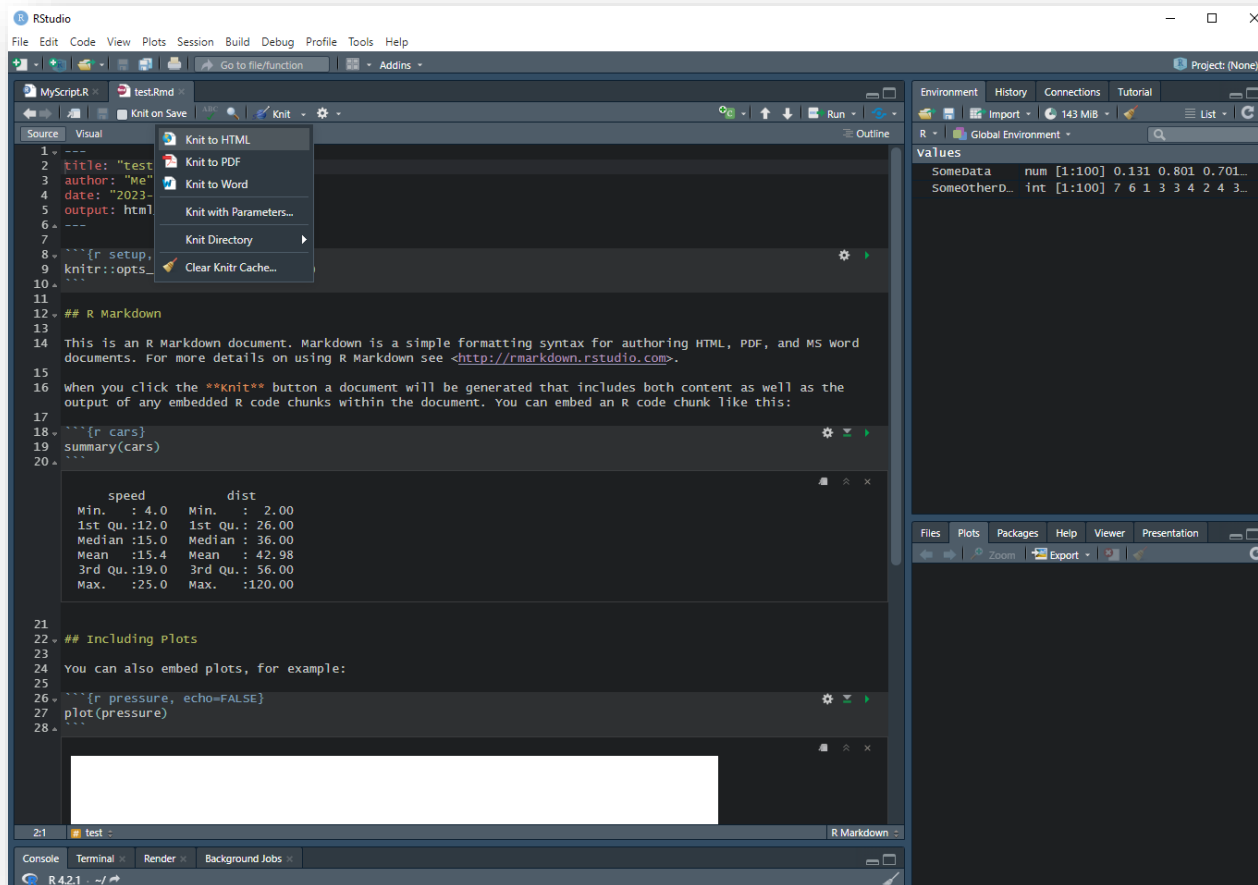
- Share reports with your PI/colleagues
- Remember what you have done
- View code, figures
- Navigate through analyses using tabs



Can everyone create an .Rmd file and run the default file ?



Can everyone create an .Rmd file, run the default file and create a report ?



Exercise 1

Have a look at [CheatSheets pdf](#)

- Create a Rmarkdown with the following information:
 - Title: "Exercise 1"
 - Date: Date when the Rmarkdown was generate (automatically change if it is runned another day)
 - report as a html document
 - Create necessary title and subtitle
- Create a simple chunk of code with a simple addition
- Load iris dataset using the following function: **data(iris)**
- Display your iris data.frame, writing **iris** in a chunk.
 - Add **df_print: paged** in your YAML header in the html document option and display the iris data.frame again.
 - Show rows 6 by 6 by using the **rows.print** options in chunk option
- Show the boxplot of the iris dataset using **boxplot(iris)**
- Write the mean Sepal length, embedded in a text
- Show Sepal length, width and Petal length, width using tabs and the function **plot()**, e.g. **plot(iris\$Sepal.Width)**
- Add **sessionInfo()** at the end of the document