STATISTICAL MODELS FOR CLINICAL AND HEALTH DATA

V. Monbet

[1] Université de Rennes/UFR Mathématiques

Objectives

- Learn about statistical methods and models for clinical and health data
- Develop your ability in formulating a problem (from a medical description to a statistical question)
- Develop your ability in choosing a method/model for a specific (statistic) problem and use it to reply the clinical/health question
- Learn how to provide a critical analysis of the results.
- Develop your ability to read and understand biological, medical and health science articles.

- When ?
  Thuesday, from 3pm to 6 :30pm
  9 lectures + practice (If you have a laptop, please bring it with you.)
- Where ?
  Campus de Beaulieu, bat 2A

- Evaluation
  Labs + projects

- Softwares/material : R (and Python), github, google scholar, etc

Context

Clinical studies

- What are the features characterizing patients vs controls ?
  ex : + ref.
- What is the efficiency a treatment ?
  ex : + ref.
- predict survival of patients with heart failure from serum creatinine and ejection fraction alone [Chicco and Jurman, 2020]

(Public) Health questions

- Self-poisoning by E-cigarette and E-liquids : National Reports to French Poison Control Centers from July 2019 to December 2020 : VIGIlance and VAPE : The VIGIVAPE Study [Franchitto et al., 2024]
- Biological embodiment of educational attainment and future risk of breast cancer : findings from a French prospective cohort [Berger et al., 2025]

Classical Methods

Content of the course

1. Introduction, resaech reproductibility, good practices
2. Statistical tests
3. Multivariate linear regression
4. ANOVA
5. Logistic regression
6. Multinomial and ordinal regression
7. Poisson regression (?)

with possibly variable selection for various datasets [1] and data quality [2].

8. Variable selection methods
9. Repeated data, curves, missing data

Theoretical (lectures + articles readings) and pratical aspects (Python or R).

1. ex : longitudinal data
2. missing data may occur

Data (examples)

- Clinical data
  A (often small) number of patients (including controls) are observed.
  Typically, two groups : patients vs controls, treated vs control, ...
  Data are obtained from humans or animal models

- (Public) Health data
  Cohorts

Data (sources)

Clinical data

- UCI
- kaggle

(Public) Health data

- Cohorts : Gazelle, British, ...
- OCDE

Research should be reproducible

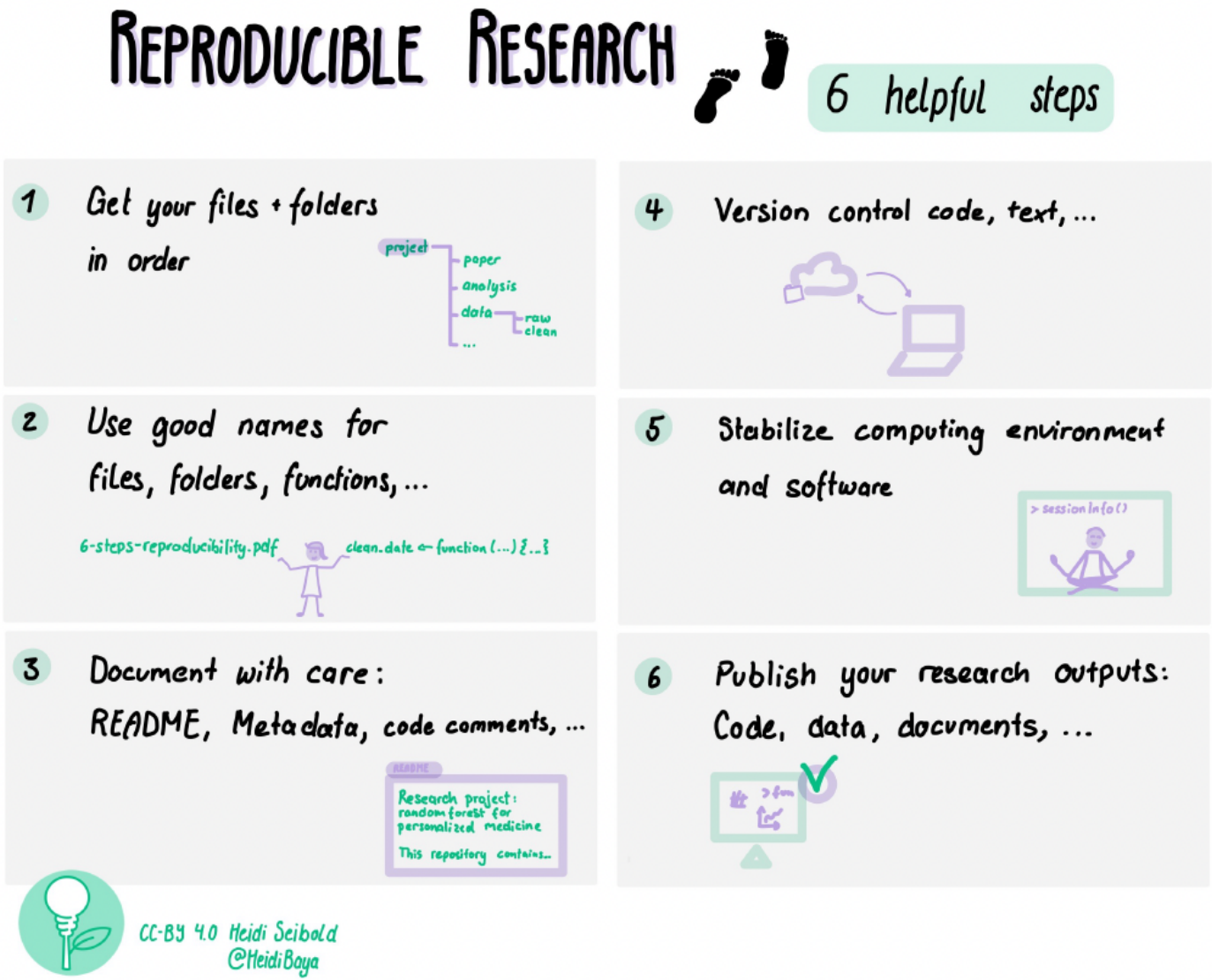
.

Two definitions from the American Statistical Association :

- **Reproducibility** : A study is reproducible if you can take the original data and the computer code used to analyze the data and reproduce all of the numerical findings from the study.

- **Replicability** : This is the act of repeating an entire study, independently of the original investigator without the use of original data (but generally using the same methods).

6 Steps Towards Reproducible Research

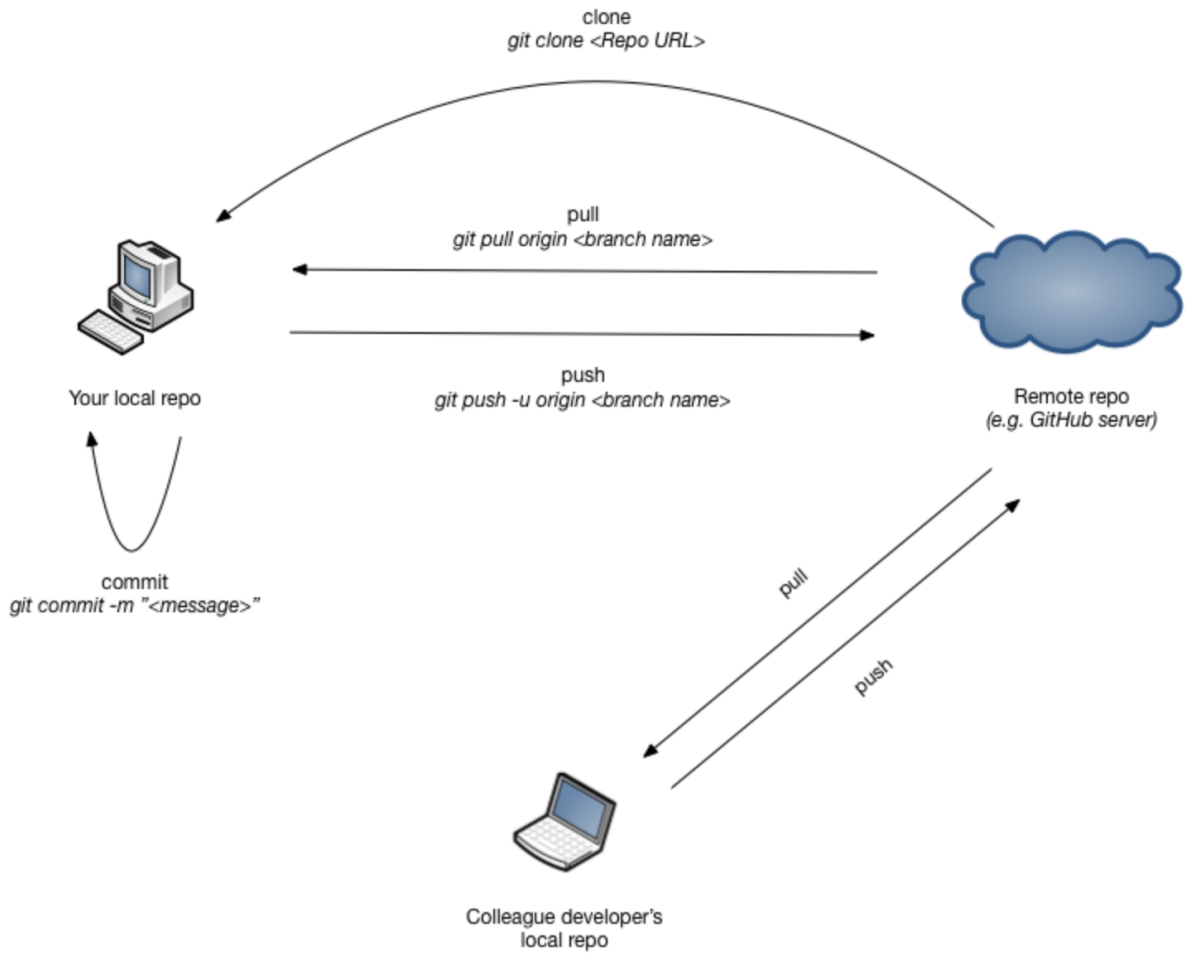


Source [Seibold, 2023]

Why should I know about Git

Some major benefits of using Git are

- Keep an archive of every version of your project
- All you and your co-authors to work at the same time
- You can easily see what changes were made and by whom
- Allows you to contribute to open source projects
- Allows you to make your project open source so others can contribute to your project



What is comming next ?

- Practice on Git
- Combine : create a github (for David's data ?), start to build a descriptive statistics table (Python or R)
- Papers reading

References


Berger, E., Dudouet, R., Dossus, L., Baglietto, L., Gelot, A., Boutron-Ruault, M.-C., Severi, G., Castagné, R., and Delpierre, C. (2025).
Biological embodiment of educational attainment and future risk of breast cancer : findings from a french prospective cohort.
BMJ open, 15(2) :e087537.


Chicco, D. and Jurman, G. (2020).
Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone.
BMC medical informatics and decision making, 20(1) :16.


Franchitto, N., Bloch, J., Solal, C., Group, F. P. R., and Pélissier, F. (2024).
Self-poisoning by e-cigarette and e-liquids : national reports to french poison control centers from july 2019 to december 2020 : Vigilance and vape : the vigivape study.
Nicotine and tobacco research, 26(3) :281–288.


Seibold, H. (2023).
Practical steps towards reproducible research.