



Better learning of data science in a biology curriculum by using R, Rstudio, learnr & Github Classroom

G. Engels* & Ph. Grosjean

Numerical Ecology of Aquatic systems laboratory, COMPLEXYS and INFORTECH institutes,
University of Mons (UMONS), 6 avenue du champs de mars, 7000 Mons, Belgium
* Corresponding author : guyliann.engels@umons.ac.be – <https://guyliann.netlify.com>



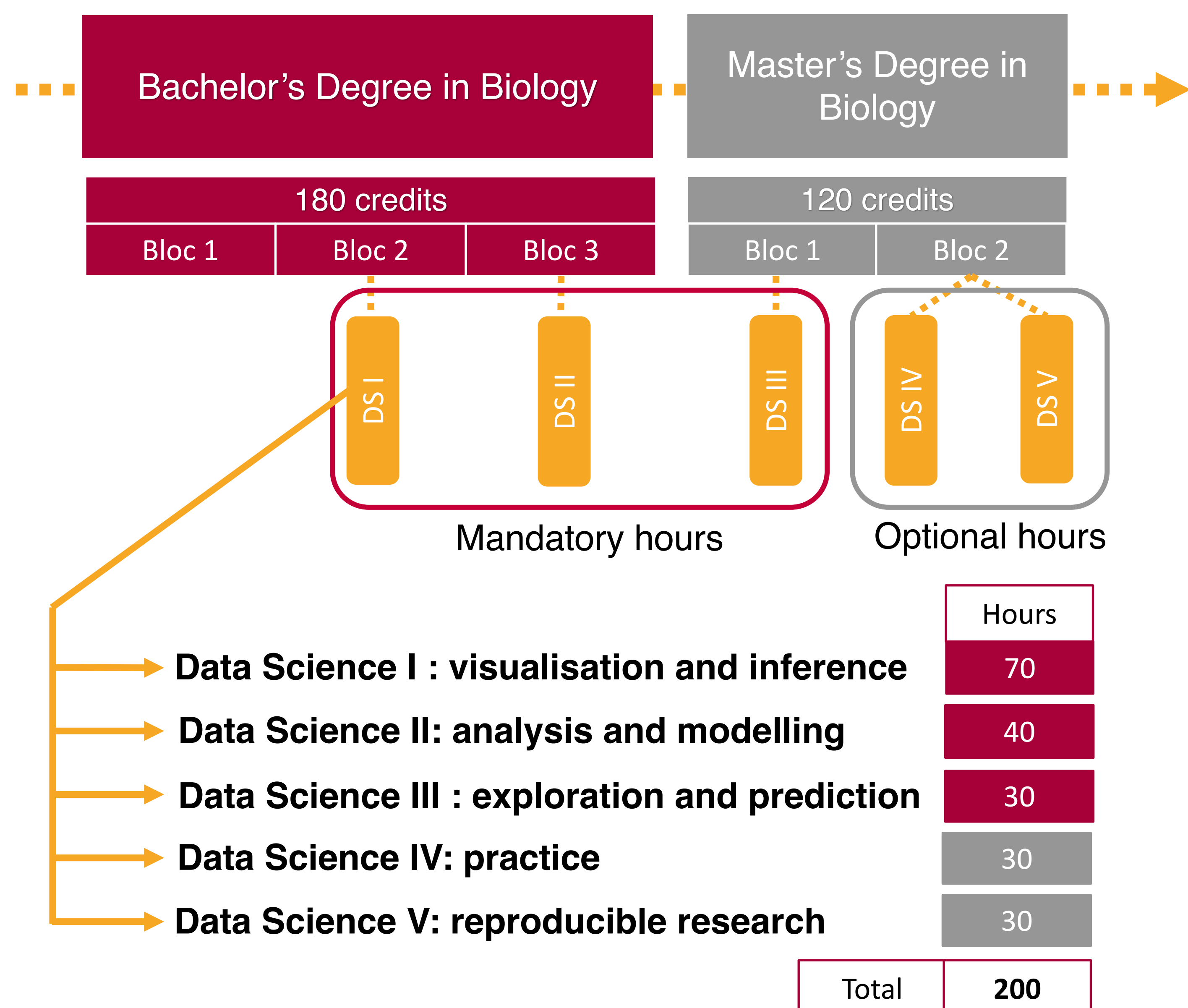
Introduction

In a biology curriculum at the University of Mons (UMONS, Belgium), the traditional biostatistics course are replaced by data science courses with modern and interactive tools. Furthermore, the teachers become guides and they help students to achieve self-study.

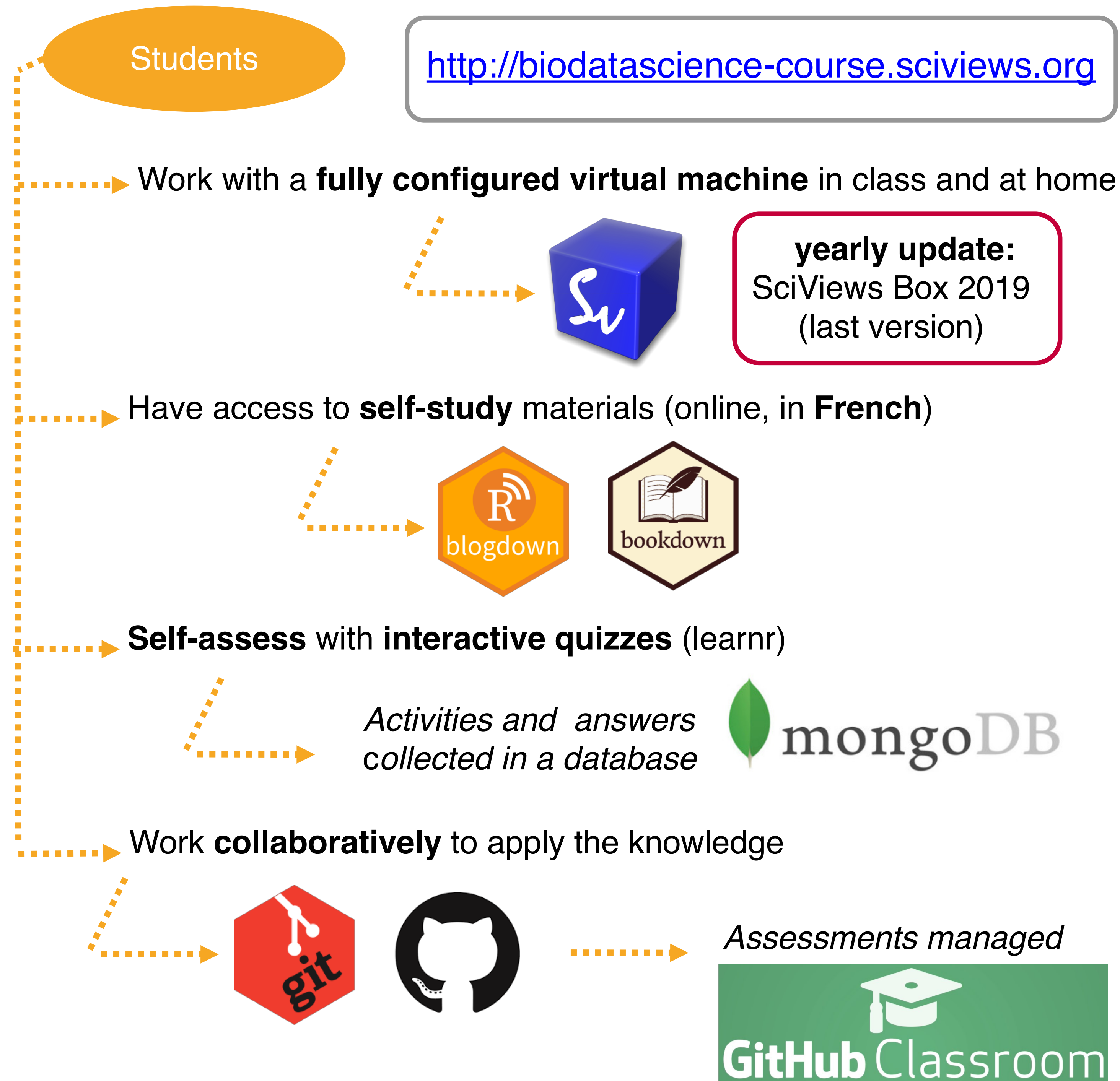
Biologists need effective and consistent tools for data analysis and processing (R), version manager (Git) and an interface (RStudio) for reproducible scientific report (R Markdown/R Notebook). However, students prefer a comprehensive software environment rather than installing all the components of the aforementioned tools on their personal computer.

Data science in a biology curriculum

The biology curriculum includes 5 courses on 4 consecutive years. The complete formation includes **200 hours** of in-class practical work.

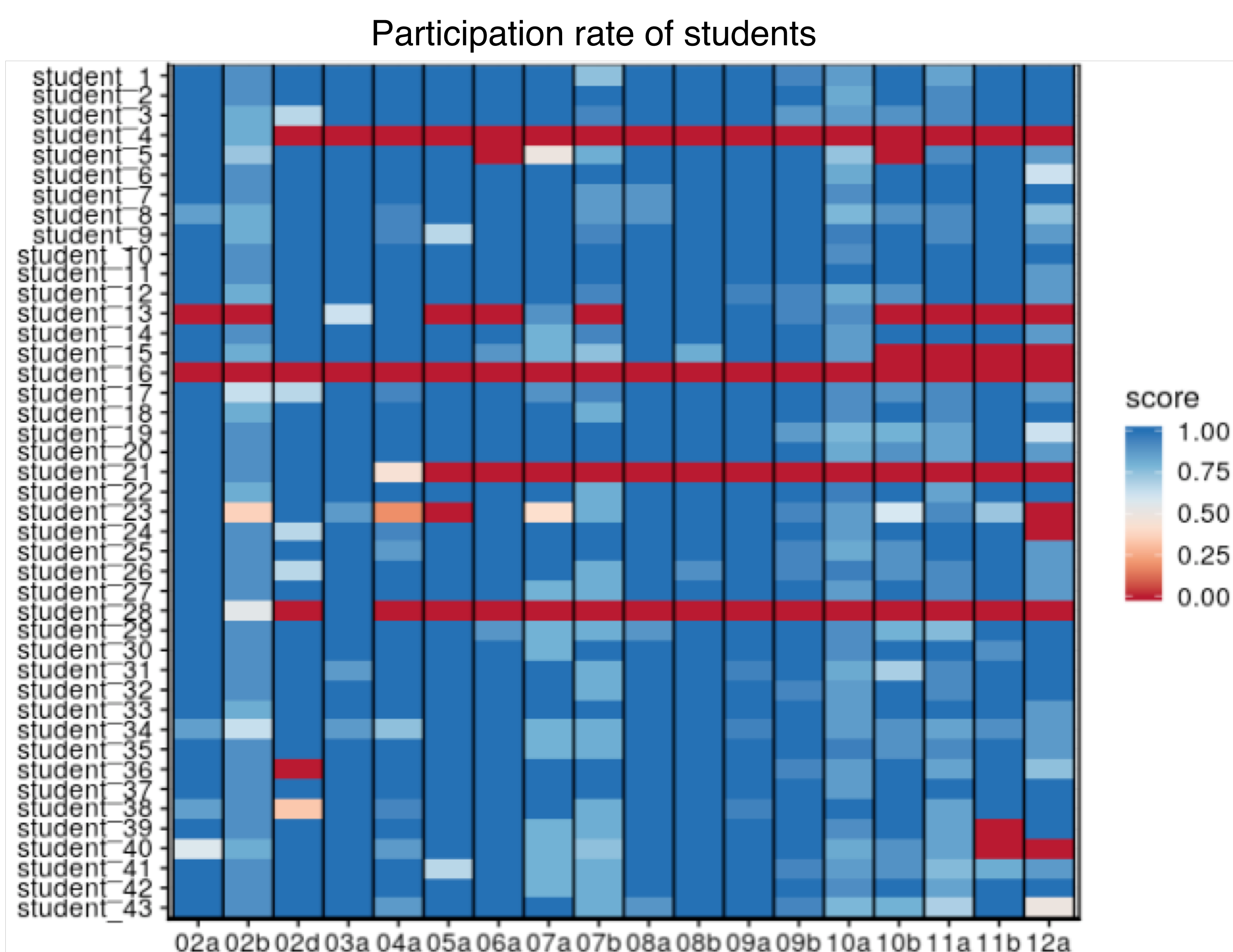


Pedagogical approach & tools



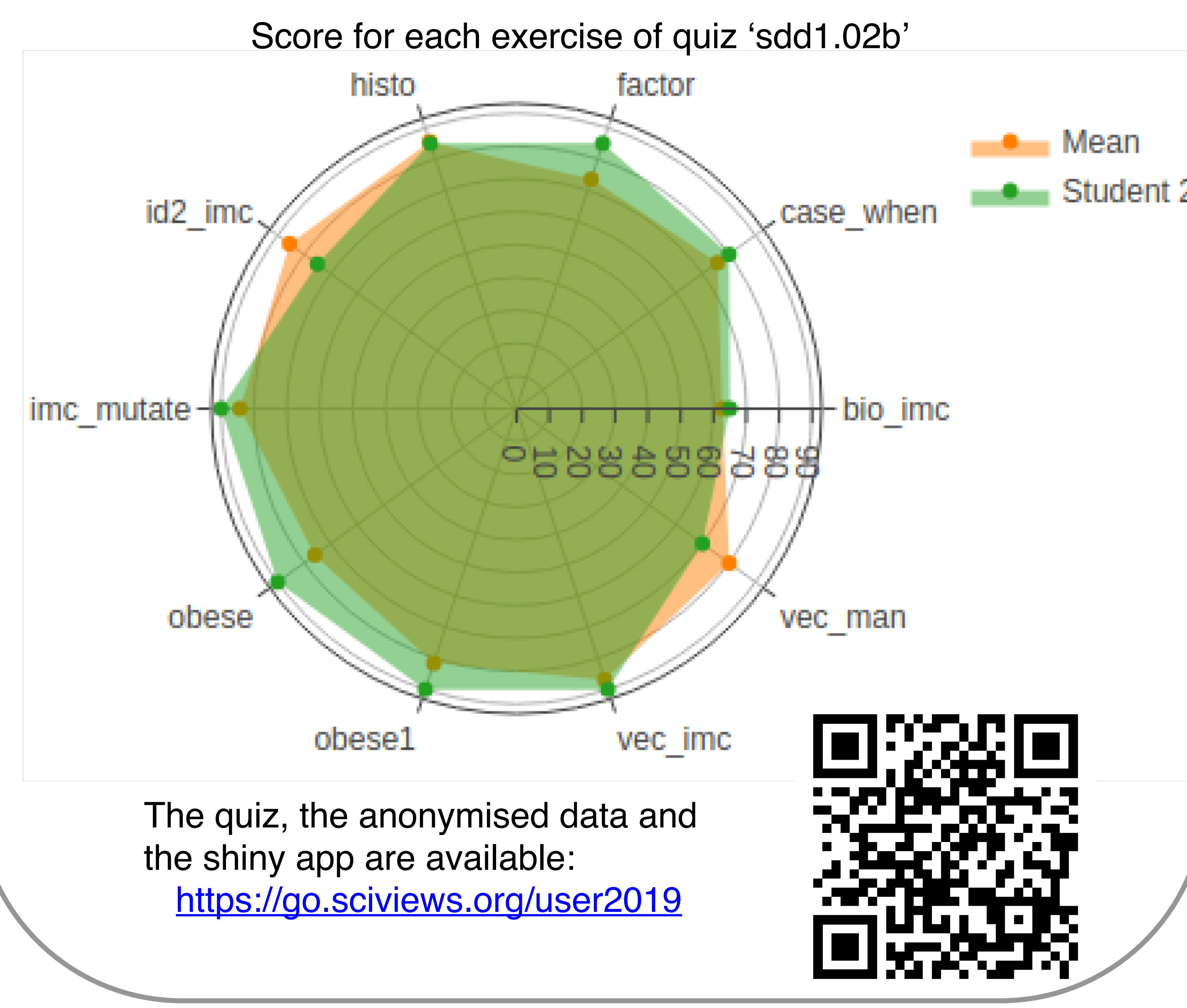
Attendance of students in data sciences I

The figure below shows the participation rate of students during their self-assessment for each quiz (score is the ratio of the submitted responses on the number of questions). Similar interactive quizzes (learnr) and the 187 GitHub repositories created by the students (GitHub Classroom) are used for summative assessments.



Diagnostic tools

We have collected more than **41 000 entries** with **18 interactive quizzes** and **43 students**. A shiny app is developed to reveal the quality of learning and to improve the quizzes. For instance, the figure shows that the student 2 is better than average for quiz 'sdd1.02b'.



Conclusion

This pedagogical approach shows a particularly large participation rate and excellent overall result. We never got such a high success rate in the previous traditional biostatistics courses. The students are clearly more motivated to use a computer.

Continuous and self learning is the key to acquire data science skills. This is possible thanks to the SciViews Box usable both in class and at home and to the self-evaluation quizzes. Student activity monitoring in a database provides useful and detailed information on the learning process.