

Markup Languages and Reproducible Programming in Statistics

General information

- Course: [Markup languages and reproducible programming in statistics \(202000010\)](#)
- Course coordinator: Gerko Vink (g.vink@uu.nl)
- Teachers: Gerko Vink (g.vink@uu.nl) & Hanne Oberman (h.i.oberman@uu.nl)
- Schedule: [MyTimetable](#)

Course description

This course gives an overview of the state-of-the-art in statistical markup, reproducible programming and scientific digital representation. Students will get to know the professional field of statistical markup and its innovations and challenges. It consists of meetings in which students will learn about markup languages (LaTeX and Markdown), learn efficient programming with R Markdown, experience developing Shiny web apps, get to know version control with Git and will create and maintain their own data archive repository and personal (business card) page through GitHub. Combining these lectures, the students get acquainted with different viewpoints on marking up statistical manuscripts, areas of innovation, and challenges that people face when working with, analyzing and reporting (simulated) data. Knowledge obtained from this course will help students face multidimensional problems during their professional career.

Syllabus

The course website serves as syllabus. Background information, e.g. about course aims, can be found on [Osiris](#). If you really want to have a PDF version, you can download it [here](#).

2023

Markup languages and reproducible programming in statistics (202000010)

General information

Course ID
202000010

Course type
Course

Credits
2.5 EC

Category / Level
M (Master)

Instruction language
English

Offered by
Utrecht University - Faculty of Social Sciences

Is mandatory for

- M&S for the Behavioural, Biomedical and Social Sciences (MBBM)

Course enrolment

Enrolment periods

- **Semester I**
 - Timeslots
Timeslot not applicable
 - Enrolment period
 - Start courses
4 September 2023

Course description

Course goals

TESTING AND COURSE AIMS

1. Developing and publishing a reproducible research archive that contains reproducible code, data and a typeset manuscript following a markup language;
2. Developing and publishing a personal repository page;

- a. Students develop fundamental knowledge and understanding in the state of the art in statistical markup languages and reproducible programming (Knowledge and Understanding)
- b. They can determine the most effective markup strategies to address a typesetting problem (Applying)
- c. They can efficiently organize a reproducible programming process (Applying)
- d. They can produce repositories up to the standards of international programming and coding conventions and initiatives (Communication)
- e. They can produce publications up to the typesetting standards of international peer- reviewed journals (Communication)

Assignment

Students will individually choose one statistical topic and work on a Markup manuscript about this topic. Students will need to perform calculations and program code for this script. All work for the student needs to be combined in an easy understandable and insightful data archive and materials portfolio and will need to be posted on a personal GitHub repository. This portfolio will be graded on 1) Quality of the markup language and coding skills, 2) Quality of the data archive and 3) Quality of the online repository.

Details about the exact grading of the portfolio can be found in the course manual.

In order to pass the course, the final grade must be 5.5 or higher, your contribution to the course should be sufficient and all assignments and practical assignments should be handed in and/or passed. Otherwise, additional work is required concerning the assignments and/or exercises you have failed.

After taking this course students can understand innovations in statistical markup, statistical simulation and reproducible research. Students are also able to approach challenges from different professional viewpoints. They have gained experience in marking up a professional manuscript and designing a state-of-the-art statistical archive in an open source repository.

Content

This course gives an overview of the state-of-the-art in statistical markup, reproducible programming and scientific digital representation. Students will get to know the professional field of statistical markup and its innovations and challenges. It consists of meetings in which students will learn about markup languages (LaTeX and Markdown), learn efficient programming with R Markdown, experience developing Shiny web apps, get to know version control with Git and will create and maintain their own data archive repository and personal (business card) page through GitHub. Combining these lectures, the students get acquainted with different viewpoints on marking up statistical manuscripts, areas of innovation, and challenges that people face when working with, analyzing and reporting (simulated) data. Knowledge obtained from this course will help students face multidimensional problems during their professional career.

Note that for external parties, costs for participation may be involved.

Students will need their own laptop computer. Students should have experience in programming with R and should be familiar with the IDE RStudio.

Prerequisites and entry requirements

Assumed previous knowledge

Statistical programming with R

Entrance requirements course enrolment

You must meet the following requirements

- Enrolled for a degree programme of faculty Faculty of Social Sciences
- Completed all course modules listed below
 - MSBBSS04 Computational inference with R (201300004)

Materials

Required material

- Internet pages
The course page <https://www.gerkovink.com/markup> holds, at the start of each iteration of the course, all relevant information and the course materials. Exercises, issues and questions are submitted through the course GitHub page. No other channels of dissemination (like e.g. blackboard) are used.

Instructional modes

Instructional modes

- Lecture/seminar
Attendance requirement
Yes

Tests

Tests

- Portfolio
Test weight
100

Minimum grade
5.5

Knowledge and skills development

Lecturers

Course contact

- dr. G. Vink

Lecturer

- dr. G. Vink

Teaching team

Instructors



I am a statistician masquerading as a dark data scientist with a passion for educating people. I aim to be at the cutting edge of both teaching and research and have an interest in new developments concerning the presentation of data, results and knowledge. I have a specific interest for problems where not all data is available (some information is left in the dark), information privacy, computational evaluation and programming. I am based in the Netherlands and work as an associate professor at Utrecht University (Utrecht, Netherlands).

I live my life with a wife and two kids, a surfboard and a pair of hiking shoes. You can reach me via e-mail at g.vink@uu.nl

Hanne Oberman is my (see above) PhD student. Hanne is a key figure in Utrecht's Open Science Community and has a knack for new developments in creating and maintaining robust software. Hanne is a contributor and developer to multiple R-packages, some of which are in the top 10 packages developed at Utrecht University. Oh, and Hanne is much too polite, which is why I wrote this text on her behalf. You can reach Hanne at h.i.oberman@uu.nl



Course support

Support

We will not use the course page on Blackboard.

Instead, all course materials can be found here on the course website, www.gerkovink.com/markup.

For handing in assignments, we will use [the course repository on GitHub](#).

Please post your content-related questions as [an issue in the course repository](#).

Questions regarding personal circumstances may be shared through e-mail with the teaching team.

Markup Languages and Reproducible Programming in Statistics

Week	Date	Focus	Location
1	13 Sep	Introduction to Markup Languages; and LaTeX	BOL 1062
2	27 Sep	Markdown in (most) Flavors	BOL 1025
3	01 Nov	Version Control and Development Flow	SGG 128
4	08 Nov	Reproducible Research Repositories	BOL 1025
5	22 Nov	Developer Portfolio	BOL 1025
6	06 Dec	Packages, Code Robustness and Unit Testing	BOL 1025

Project description

This course is graded by means of a portfolio. Students collect evidence of their ability to:

1. Use markup languages for scientific manuscript writing
2. Produce a reproducible repository that conforms to rigorous open science and computing standards
3. Present this portfolio using markup languages
4. Develop, test, maintain and host this portfolio.