Applied Data Science and Data Visualisation

Title	Applied Data Science and Data Visualisation
Semester	F2024
Master programme in	Chemical Biology / Environmental Science / Mathematical Bioscience / Molecular Health Science
Type of activity	Course
Teaching language	English
Study regulation	Read about the Master Programme and find the Study Regulations at [ruc.dk](ruc.dk) Læs mere om uddannelsen og find din studieordning på [ruc.dk](ruc.dk)

REGISTRATION AND STUDY ADMINISTRATIVE

Registration	Sign up for study activities at stads selvbetjening within the announced registration period, as you can see on the Studyadministration homepage. When signing up for study activities, please be aware of potential conflicts between study activities or exam dates. The planning of activities at Roskilde University is based on the recommended study programs which do not overlap. However, if you choose optional courses and/or study plans that goes beyond the recommended study programs, an overlap of lectures or exam dates may occur depending on which courses you choose.
Number of participants	
ECTS	5
Responsible for the activity	John Shorter (johnsh@ruc.dk)
Head of study	Lotte Jelsbak (ljelsbak@ruc.dk)
Teachers	
Study administration	INM Registration & Exams (inm-exams@ruc.dk)
Exam code(s)	U60176

ACADEMIC CONTENT

Overall objective	The overall objective of this course is to introduce the concept of data science and visualization of data to enable students within experimental sciences to design, perform, visualize, evaluate, interpret and communicate experiments where many parameters are measured and so called big data experiments ('omics' data). Furthermore, the aim is to provide students with the necessary methodological and data analysis skills to be able to evaluate validity and quality of methods and data related to analysis of large datasets.
Detailed description of content	The course consists of lectures combined with hands-on exercises, and projects where the students can work on their own data or other data from their own field. No previous programming experience is required, but students will be expected to learn basic programming (R and Rstudio) and statistical analysis during this course.
Course material and Reading list	No textbooks are needed, course material will be specified on moodle.
Overall plan and expected work effort	- Reading course material and problem solving at home: 38 hrs - Lectures: 8 hours - Hands-on exercises and problem solving in class: 24 hrs - Working on mini-projects and report writing: 65 hours **Total 135 hrs**
Format	
Evaluation and feedback	The course includes formative evaluation based on dialogue between the students and the teacher(s). Students are expected to provide constructive critique, feedback and viewpoints during the course if it is needed for the course to have better quality. Every other year at the end of the course, there will also be an evaluation through a questionnaire in SurveyXact. The Study Board will handle all evaluations along with any comments from the course responsible teacher. Furthermore, students can, in accordance with RUCs 'feel free to state your views' strategy through their representatives at the study board, send evaluations, comments or insights form the course to the study board during or after the course.
Programme	Most classes will begin with a short lecture / introduction to a new concept followed by time for discussion and work with programming exercises. Students will then work in pairs to analyze a new data set based on concepts covered in the introduction. Selected groups will then present their data analysis to the class. The students will write a report for each data set (usually a PowerPoint presentation or R markdown) where an emphasis should be on explaining the analyses used, the implication of the results, and on the visualization of the data.

These reports will be turned in at the end of the class, or before the next class, with the names of the two group members along with code used for the analysis and visualization.

The topic for mini-project is to present a visual and statistical analysis of an approved dataset. You will use what you learned during the semester to create an R script that goes step-by-step on an analysis, and you will present this analysis and script at the last class.

ASSESSMENT

Overall learning outcomes

After completing the course, the students will be able to:

- describe and explain the concepts of multivariable data processing and visualization
- handle multivariable data using relevant software such as python, R or using statistical software (e.g. SYSTAT, SPSS etc.)
- identify and extract relevant parameters from large data sets
- implement appropriate descriptive statistics on high complexity and big data
- describe and analyze the intrinsic structure of a large multivariable dataset using relevant methods, such as clustering methods, principal component analysis (PCA) or least-squares analyses (PLS)
- analyze multivate data using basic linear models with covariate adjustments, and interpret and discuss results these
- describe simple machine learning algorithms and explain their differences with regard to purpose of use, strengths and weaknesses, as well as use selected machine learning algorithms for tasks such as selection of the variable with the best predicting power, and interpret results from these.
- explain the results from these methods to both lay people and specialists
- be aware of the limitations of the chosen tests
- visualize the results in an informative and rigorous way.
- design complex experiments, including 'omics' experiments based on the methodological considerations of the ensuing data analysis
- write documents describing methodological considerations regarding the analysis of big ('omics') data
- communicate the knowledge and understanding gained from the course in a precise and scientific way.

Form of examination

The course is passed through active, regular attendance and satisfactory participation.

Active participation is defined as:
The student must participate in course related activities (e.g. workshops, seminars, field excursions, process study groups, working conferences, supervision groups, feedback sessions).

Regular attendance is defined as:
- The student must be present for minimum 75 percent of the lessons.

Satisfactory participation is defined as:
- e.g. oral presentations (individually or in a group), peer reviews, mini projects, test, planning of a course session .

Assessment: Pass/Fail.

Form of Re-examination	Students that have not participated satisfactory must hand in renewed written products.
	Students that have only met the requirement of regular attendance between 50% and 75% must hand in an additional written product.

Type of examination in special cases	

Examination and assessment criteria	Participate actively is define as: The student must participate actively in lectures, discussion and problem solving classes. Students may be selected to present their report to the class at the end of a lecture. Active participation means students must work on data analysis during the class and present if they are called upon.
	Regular attendance is defined as: The student must be present for minimum of 75 percent of the lessons. This includes arriving on-time and staying until the end of the class.
	Satisfactory participation is defined as: The student must write and submit reports (usually a PowerPoint presentation) following every class. The student must work on a mini-project and present results at the final class.
	Assessment criteria in relation to satisfactory participation of the exercises and mini-project, students will be assessed by their ability to:
	- explain the analyses used - account for, how choice of analysis have implication on the results, the visualization of the data, and programming code for analysis and visualization - communicate the knowledge and understanding gained from the lesson in a precise way within the submitted reports

Exam code(s)	Exam code(s) : U60176

Course days:

Hold: 1

Applied Data Science and Visualisation (MHS, CB)

time	14-02-2024 12:15 til 14-02-2024 16:00
forberedelsesnorm	ikke valgt
forberedelsesnorm D-VIP	ikke valgt
location	22.1-009 - undervisningslokale (72)

Teacher	John Shorter (johnsh@ruc.dk)

Applied Data Science and Visualisation (MHS, CB)

time	21-02-2024 12:15 til 21-02-2024 16:00
forberedelsesnorm	ikke valgt
forberedelsesnorm D-VIP	ikke valgt
location	22.1-009 - undervisningslokale (72)
Teacher	John Shorter (johnsh@ruc.dk)

Applied Data Science and Visualisation (MHS, CB)

time	28-02-2024 12:15 til 28-02-2024 16:00
forberedelsesnorm	ikke valgt
forberedelsesnorm D-VIP	ikke valgt
location	22.1-009 - undervisningslokale (72)
Teacher	John Shorter (johnsh@ruc.dk)

Applied Data Science and Visualisation (MHS, CB)

time	06-03-2024 12:15 til 06-03-2024 16:00
forberedelsesnorm	ikke valgt
forberedelsesnorm D-VIP	ikke valgt
location	22.1-009 - undervisningslokale (72)
Teacher	John Shorter (johnsh@ruc.dk)

Applied Data Science and Visualisation (MHS, CB)

time	13-03-2024 12:15 til 13-03-2024 16:00
forberedelsesnorm	ikke valgt
forberedelsesnorm D-VIP	ikke valgt
location	22.1-009 - undervisningslokale (72)

| Teacher | John Shorter (johnsh@ruc.dk) |

Applied Data Science and Visualisation (MHS, CB)

time	20-03-2024 12:15 til 20-03-2024 16:00
forberedelsesnorm	ikke valgt
forberedelsesnorm D-VIP	ikke valgt
location	22.1-009 - undervisningslokale (72)
Teacher	John Shorter (johnsh@ruc.dk)

Applied Data Science and Visualisation (MHS, CB)

time	03-04-2024 12:15 til 03-04-2024 16:00
forberedelsesnorm	ikke valgt
forberedelsesnorm D-VIP	ikke valgt
location	22.1-009 - undervisningslokale (72)
Teacher	John Shorter (johnsh@ruc.dk)

Applied Data Science and Visualisation (MHS, CB)

time	10-04-2024 12:15 til 10-04-2024 16:00
forberedelsesnorm	ikke valgt
forberedelsesnorm D-VIP	ikke valgt
location	22.1-009 - undervisningslokale (72)
Teacher	John Shorter (johnsh@ruc.dk)

Applied Data Science and Visualisation - Hand-in of written products (reexam) (MHS, CB)

time	23-05-2024 10:00 til 23-05-2024 10:00
forberedelsesnorm	ikke valgt
forberedelsesnorm D-VIP	ikke valgt