Course Specifications

Valid as from the academic year 2019-2020

Course

Lecturers in academic year 2019-2020
Verwaeren, Jan
LA26

Course offerings and teaching methods in academic year 2019-2020
A (semester 2) Dutch seminar: practical PC room classes
36.25 h
group work
2.5 h
lecture
15.0 h

Offered in the following programmes in 2019-2020
Bachelor of Science in Bioscience Engineering (main subject Agricultural Sciences)
5 A
Bachelor of Science in Bioscience Engineering (main subject Cell and Gene Biotechnology)
5 A
Bachelor of Science in Bioscience Engineering (main subject Chemistry and Food Technology)
5 A
Bachelor of Science in Bioscience Engineering (main subject Environmental Technology)
5 A
Bachelor of Science in Bioscience Engineering (main subject Forest and Nature Management)
5 A
Bachelor of Science in Bioscience Engineering (main subject Land and Water Management)
5 A
Joint Section Bachelor of Science in Bio-Engineering
5 A

Teaching languages
Dutch

Keywords
Data collection, data management, data exploration, data visualization, R, data wrangling

Position of the course
This course is part of the learning path ‘mathematics and data analysis’. In this course, students learn the fundamentals of the life-cycle of data in a scientific environment. Within each phase of this cycle, starting from planning, data-collection, and processing up to the (preliminary) analysis, preservation and publishing of data, students learn a number of established and emerging technologies that are popular in modern research. Special attention is given to the fact that data can emerge in multiple forms and formats such as small structured flat-file datasets, that collect observations of a small dedicated research project, and time series, up to unstructured text-data or network data. In a series of hands-on PC labs (in the software-environment R), students learn how to compose a data-collection plan, where publicly available data are augmented with new dedicated data. Students learn how to retrieve data (e.g. using an API or SQL) and learn how to check the integrity of the data, fuse data and perform a preliminary exploratory data analysis. This course also prepares the students for a follow-up course in statistical inference.

Contents
The life-cycle of data in a scientific environment is a common thread running through this course. In the plenary sessions, a (theoretical) overview is presented of the dominant scientific and technological aspects within each phase of the data life-cycle. During PC-labs, as selection of these techniques is elaborated upon and applied in

(Approved)
Access to this course unit via a credit contract is determined after successful competences assessment. This course unit cannot be taken via an exam contract.

Group work, lecture, seminar: practical PC room classes.

**Introduction to R**

a. Universal programming principles applied in R (data type, control flow, functions, IO)
b. Specific tools for data-manipulation and visualization in R

**Different forms of data and data formats: structured vs. unstructured, flat file data, text data, network data**

**Planning and data collection phase**

a. Inventorying available (data) sources and the use of ontologies
b. Introduction into experimental design (variable types, observational vs. experimental studies, measurement uncertainty, confounding)
c. Technologies for accessing (persistent) data sources (SQL and api’s)
d. Data traceability

**De (pre-) processing**

a. Data integrity and quality
b. Data fusion
c. Data wrangling

**Explorative data analysis and descriptive statistics**

a. Data visualization: plot types and ordination methods
b. Missing data
c. Introduction to model building and nearest-neighbour methods
d. Scientific communication: formatting of figures and tables

**Data preservation**

a. Introduction to (the use of) data bases and document management systems
b. Relational databases and SQL

**Data publishing**

a. Documenting data sets
b. Metadata and ontology
c. Data ownership

Initial competences

*Data science builds on certain learning outcomes of course units ‘Calculus, ‘Linear algebra’, and ‘Scientific computing’; or the learning outcomes have been achieved differently.*

Final competences

1. For a given research question, the student can make an inventory of the available data sources and propose an experimental design that can provide the additional data needed to answer the question.
2. The student is aware of the different forms in which data appear, are capable of performing basic integrity checks for the most important data forms/types and can select and apply a proper visualization method.
3. The student can query on-line data sources and fuse the data resulting from such a query.
4. The student can use R as a programming environment for data analysis.
5. The student can apply visualization and ordination methods to gain insight in datasets.
6. The student can apply the basic principles of model building.
7. The student can select a proper database technology for a given problem setting.
8. The student can apply data wrangling techniques on flat-file data.
9. The student can compose an format summarizing figures and tables that meet the quality standards required for scientific communication.

**Conditions for credit contract**

Access to this course unit via a credit contract is determined after successful competences assessment.

**Conditions for exam contract**

This course unit cannot be taken via an exam contract.

**Teaching methods**

Group work, lecture, seminar: practical PC room classes.

**Extra information on the teaching methods**

During the theoretical lectures, the fundamental concepts are discussed. The practical PC room classes exist of 10 hands-on practical sessions. In the group work, the students have to complete a real-life data collection and synthesis task.

**Learning materials and price**

Course notes (price indication 10 Euro), slides available on Minerva

**References**

Benjamin S. Baumer, Daniel T. Kaplan and Nicholas J. Horton. (2017) Modern Data
Course content-related study coaching

Students can make an appointment with the lecturer for asking questions related to the theoretical classes throughout the entire semester. Teaching assistants address questions w.r.t. the PC-labs and Minerva is used to provide on-line feedback if needed.

Evaluation methods

end-of-term evaluation and continuous assessment

Examination methods in case of periodic evaluation during the first examination period

Written examination with open questions, open book examination

Examination methods in case of periodic evaluation during the second examination period

Written examination with open questions, open book examination

Examination methods in case of permanent evaluation

Report

Possibilities of retake in case of permanent evaluation

Examination during the second examination period is possible in modified form

Extra information on the examination methods

The written exam (20% of total) evaluates the theoretical competences. Practical competences are evaluated during an open book PC-exam (75% of total). The written report (report of group-work) is evaluated and contributes to the final grade (5% of total).

Calculation of the examination mark

Students who eschew period aligned and/or non-period aligned evaluations for this course unit may be failed by the examiner.