

Applied Data Science

B.A. International Management · Europa-Universität Flensburg · Fall Semester
2026/27 Prof. Dr. Claudius Gräbner-Radkowsch · Mondays 09:00–11:15 · MAD 099

2026-10-16

This seminar description is also available online. The current version, plus the material for every session, is on the **course website** (linked from the Moodle course); every page there can also be downloaded as a PDF. **Announcements, submissions** and the **question forum** run through the **Moodle course**.

1 Overall goal of the course

Businesses collect more data than they know what to do with. This course teaches you to do something useful with it: to take a raw, messy dataset, get it into shape, look at it properly, model what matters, and write it up so that a colleague can re-run your analysis and reach the same conclusion.

We do all of this in **R**, a free and open-source language that has become a standard tool in data-driven organisations.

The emphasis is on *doing*, not on mathematical theory. You will not derive estimators. You will build things that work, and learn to judge whether what you built is any good.

2 What this course assumes, and what it adds

You arrive from *Statistik I and II* having already covered — by hand and in R — descriptive statistics, correlation, confidence intervals, significance tests, regression including dummy variables, and basic data preparation.

This course does not repeat that. It adds four things:

1. **Making things visible.** Statistik I and II did not cover producing graphics in R. Two full sessions here do, and it is the skill you are most likely to need first in a job.
2. **Reproducible communication.** Quarto, Git and GitHub: work that someone else can re-run, check and build on.

3. **Prediction rather than explanation.** You know how to estimate a regression. You do not yet know what changes when the goal is to forecast something you have not seen — overfitting, train/test splits, and what to do when you have fifty candidate predictors and no theory.
4. **Machine learning.** What supervised and unsupervised learning actually are, and when each is the right tool.

3 Learning goals

Technical expertise

- Write R code to import, clean, reshape and summarise real business data
- Produce reproducible analysis reports with Quarto, in HTML and PDF
- Use Git and GitHub to version, track and share your work
- Build visualizations with `ggplot2` that answer a specific question
- Estimate, interpret and present regression models, and use them to predict
- Apply a supervised and an unsupervised machine learning method to a business problem

Methodological competence

- Identify the relevant data science steps for a business problem and plan a route to an answer
- Choose the visualization and the model that fit the question
- Distinguish supervised from unsupervised learning, and recognise which a problem calls for
- Tell the difference between describing a pattern and predicting a consequence
- Interpret results for an audience that does not know statistics
- Judge the quality and limits of a dataset and of an analysis

Social and personal skills

- Develop and implement solution strategies independently
- Present quantitative findings transparently and reproducibly
- Critically evaluate the data claims and visualizations you meet in professional life
- Use AI assistants as a productive tool while remaining responsible for the result

4 Schedule

Mondays 09:00–11:15 · Room **MAD 099** (Madrid building).

#	Date	Topic	Deliverable
1	14.09.2026	Welcome, tooling & the data science workflow	Report 1 issued
2	21.09.2026	R recap — the tidyverse way	
3	28.09.2026	Reproducible reports with Quarto	
–	05.10.2026	<i>No class</i> — use the time for the report and your R recap	
–	12.10.2026	<i>No class</i> — use the time for the report and your R recap	

#	Date	Topic	Deliverable
4	19.10.2026	Version control with Git and GitHub	Report 1 due Sun 18.10., 23:59 · debrief
5	26.10.2026	Data visualization I — the grammar of graphics	Report 2 issued
6	02.11.2026	Data visualization II — communication and deception	
7	09.11.2026	Working with AI in the data science workflow	Report 2 due Sun 08.11., 23:59
8	16.11.2026	From explanation to prediction	Report 2 debrief
9	23.11.2026	Choosing a prediction model: cross-validation, lasso and ridge	Report 3 issued
10	30.11.2026	Unsupervised learning — clustering for segmentation	
11	07.12.2026	Supervised learning — classification	Report 3 due Sun 06.12., 23:59
12	14.12.2026	Recap, synthesis and exam preparation	Report 3 debrief



The schedule is tentative

Topics and dates may shift as the course progresses. Changes are announced via Moodle, and the version on the course website is always the current one.

5 Sessions at a glance

The schedule above lists every topic. On the course website each session has its own page with the learning goals, the material and the take-home task.

Note that **slides, the live coding document and the exemplary solution are published after the session has taken place**, not before. Everything you need *for* a session is available beforehand.

6 Course information

6.1 Prerequisites

Successful completion of *Research Methods / Statistik I and II* is strongly recommended. Beyond that you need basic computer literacy and a willingness to be stuck for a while without giving up — that is what programming feels like at the start, for everyone.

You must bring **your own laptop**, on which you can install and run R, to every session and to the exam.

! You need to register at GitHub

From session 4 onwards we work with Git and GitHub, and reports 2 and 3 are submitted through it. Please create a free account at github.com — ideally before the course starts, and at the latest before session 4.

Register with your **university email address** so you can also claim the [GitHub Student Developer Pack](#), which is free for students.

6.2 Course details

- **Module:** W1 11, *Applied Data Science* (205304100), Elective 1
- **Credits:** 5 LP (150 hours total)
- **Contact time:** 27 hours (12×2.25 hours)
- **Self-study:** 123 hours
- **Time and place:** Mondays, 09:00–11:15, room **MAD 099**
- **Language:** English
- **Moodle:** [course 17916](#), enrolment password AdvData26

i On time requirements

This course covers a lot of content. I expect you will need to invest as much time as the 5 ECTS credits that this course is worth. This equates to 150 hours in total, of which only 27 are contact hours.

6.3 Software

All software is free of charge:

- **R** (4.4), **RStudio** (2024.04), **Quarto** (1.5)
- **Git** and a **GitHub** account, from session 4 — set up *before* session 4, following the tutorial linked on [Getting Started](#)

Installation instructions are on the [Getting Started](#) page.

6.4 Data

Most datasets come from Békés and Kézdi (2021), *Data Analysis for Business, Economics and Policy*, and are licensed [CC BY-NC-SA 4.0](#). Every session page states where its data came from, what a row represents, and what was changed.

6.5 Session format

The sessions usually follow a similar rhythm:

Time	Block
09:00–09:15	Debrief of the take-home task (when there is one)
09:15–09:45	Input lecture
09:45–10:10	Live coding — I code, you watch and ask questions
10:10–10:20	Break
10:20–10:55	In-session exercise
10:55–11:10	Debrief of the exercise
11:10–11:15	Intro to the take-home task (when there is one)

Bring your laptop, charged, to every session.

7 Assessment

7.1 Preliminary examinations: three Quarto reports

Over the semester you write **three reports** in Quarto, each analysing a small business problem with the tools covered up to that point.

- Written **individually**
- Graded **pass/fail** only
- **All three must be passed** to be admitted to the final exam
- Submitted as a rendered HTML or PDF document **together with everything needed to re-run it** — for Report 1 a zip of your project folder, for Reports 2 and 3 the GitHub repository. The analysis must reproduce.
- **One revision in total** across the three reports (see below)

	Topic	Issued	Due	Submitted via
Report 1	Barcelona hotel market	Session 3, 28.09.	Sun 18.10., 23:59	Moodle
Report 2	announced in session 6	Session 6, 02.11.	Sun 08.11., 23:59	Classroom 50 (GitHub)
Report 3	announced in session 10	Session 10, 30.11.	Sun 06.12., 23:59	Classroom 50 (GitHub)

Deadlines and rules are collected on the [Reports](#) page. The **task sheets themselves are not on this website**: Report 1's is on Moodle, and those of Reports 2 and 3 come in the Classroom 50 repository whose link is on the Reports page.

! Revisions: one in total, not one per report

If a report does not pass, you may revise and resubmit it. Across the whole semester you have **one revision in total** — so if you use it on Report 1, Reports 2 and 3 have to pass first time. The revised version is due one week after you receive the feedback. This is deliberate. A revision is there to catch a report that went wrong, not to serve as a first draft.

i On the GitHub submissions

Reports 2 and 3 run through GitHub, which we set up in session 4 with a throwaway practice assignment first. **Moodle remains available as a fallback for every report.** If the tooling misbehaves, submit through Moodle and tell me — a technical problem will never cost you a submission.

7.2 Final exam

- **120 minutes**, written
- **Open book:** all course material and free online resources are permitted, including free AI assistants
- **Not permitted:** paid tools and any form of collaboration
- Written on your own laptop, and handed in through **Classroom 50**, exactly like Reports 2 and 3 — see the [Classroom 50 manual](#)
- The only graded component of the module

The exam is applied: you are given data and a question, and you answer it. It is not a test of what you can recite.

8 Working with AI

You may use free AI assistants throughout this course, including in the exam. This is deliberate — using them well is part of the job you are training for.

Two conditions apply to everything you hand in:

1. **You are responsible for every line.** If you cannot explain what a piece of code does, or why a result means what you claim, it does not count as your work.
2. **Declare it.** Each report ends with a short note on how you used AI.

Session 7 is devoted to doing this properly, including the ways in which AI assistants confidently get statistics wrong.

9 Contact

Prof. Dr. Claudius Gräbner-Radkowsch Europa-Universität Flensburg · Internationales Institut für Management und ökonomische Bildung (IIM) · Abteilung Plurale Ökonomik

claudius.graebner@uni-flensburg.de · claudius-graebner.com

For questions on content or logistics, please use the Moodle forum — that way everyone benefits from the answer. An appointment during my office hours can be booked [here](#).

References

Békés, G. and Kézdi, G. (2021) *Data analysis for business, economics, and policy*, Cambridge: Cambridge University Press, available at <https://gabors-data-analysis.com/>.