

Syllabus

May 1st, 2025

Description

The generative AI applications of recent years have proved to be powerful tools promising to enhance the productivity of creative occupations previously relying exclusively on human intelligence. At the same time, however, generative AI has creative destruction potential that rivals the most impactful innovations since the Industrial Revolution. Among the most impacted fields is software engineering, where generative AI has significantly reduced the technical barriers to entering the field. Data scientists, engineers, and project managers are challenged to adapt and adopt new approaches, leveraging instead of fighting the novel status quo.

This course introduces fundamental computer science concepts and data science techniques adjusted to work in alignment with generative AI coding tools. It aims to equip students with the essential knowledge and skills to understand when and how to use generative tools in applications of computational methods to economics and data analysis. Previous experience with data science and programming languages can be helpful, but it is not required. The course offers a novel approach to introducing programming fundamentals relevant to data science in combination with generative AI tools. Further, it delves into more advanced programming, visualization, and data management topics, illustrating them using workflows with state-of-the-art generative AI tools. The course focuses on the following topics:

1. **Programming with Generative AI.** The course begins by concisely presenting the generative AI landscape. Then, it introduces students to programming concepts and practices with workflows involving intelligent coding companions. On a technical level, the introduction includes programming concepts such as data types, control structures, and functions. On a conceptual level, it highlights the benefits, as well as potential pitfalls, of using generative tools. The focus is on programming in R, a widely used language in data science, statistical learning, and computational economics.

2. **Data analysis.** The course covers techniques for data analysis, including exploratory data analysis, preprocessing, and feature engineering for machine learning using libraries such as [dplyr](#) and [TensorFlow](#). It showcases how coding tools can be used both in the design/brainstorming stage and during the data coding/implementation stage.
3. **Visualization.** Finally, the course introduces students to the basic concepts of graphics and visualization, including commonly used types of figures and visualizations, annotations, and labeling using the [ggplot2](#) library.

Prerequisites

The intended audience for the course is first-year postgraduate students in management and economics who want to specialize further in computational economics, data science, and economic applications of machine learning. Programming with R will be introduced from scratch, and no previous experience with the language is required. Experience with other languages, such as Python, can be helpful but is not a prerequisite. Instead, knowledge of basic statistics is a prerequisite. Knowledge of usual statistical measures of centrality and dispersion, histograms, and regression concepts introduced in typical statistics and econometrics courses of B.Sc. programs in economics or business is expected.

Learning Objectives

The course's first objective is to provide an introduction to programming with state-of-the-art workflows involving generative AI tools. The course guides students to explore programming with R using powerful libraries commonly used in data analysis, visualization, and statistical learning. On successful completion of the course, students will be able to

1. understand how and when to employ generative AI tools for implementing, debugging, and communicating coding content,
2. understand and apply programming concepts and practices in solving data science problems,
3. utilize standard aggregation and visualization techniques for economic data.

Second, the course aims to introduce students to working effectively on data science projects in a team. Upon completion, students will be able to

4. communicate technical data science ideas clearly and concisely.

Finally, the course aims to provide a high-level overview of the open-source R statistical software ecosystem. To this end, upon completion, students will be able to enrich and expand their knowledge of R statistical software and its libraries independently.

Key Concepts

To achieve the goals of this course, you must master the following key concepts:

1. Generative coding assistants.
2. Data filtering, transformation, and aggregation.
3. Data visualization.

Explanation of Assignments

The didactic approach of the course is to provide students with quizzes that test their understanding of basic concepts and group assignments that can stimulate working on technical subjects in collaboration with peers. The coding challenges and group assignments involve practical applications of the introduced concepts with economic and financial data use cases. The students are expected to coordinate and work on providing solutions in break-out sessions. Solutions should be developed independently across groups. The instructor will be available during the break-out sessions to advise and support the efforts of all groups.

Final Examination and Grading

The final grade for the course will be based on a portfolio examination consisting of the following partial assessments:

- a. two quizzes
- b. two group assignments
- c. a coding challenge
- d. an individual assignment

The partial assessments a. – c. are supplementary. To complete the course, participants must submit a minimum of three out of the five partial assessments (e.g., one quiz, one group assignment, and the coding challenge).

The partial assessment d. (individual assignment) is mandatory. The individual assignment must be submitted two weeks after the last class session, i.e. latest by June 15, 2025, End of Day. The link for submission will be available on MaxBrain.

The following grades will be used for grading the overall assessment of all elements of the portfolio:

Table 1: Grading Scale

| Course Grades | Description | |
|---------------|-----------------------------|--|
| 1.0 | very good (sehr gut) | An outstanding performance |
| 2.0 | good (gut) | A performance that is significantly above the average requirements |
| 3.0 | satisfactory (befriedigend) | A performance that meets average requirements |
| 4.0 | sufficient (ausreichend) | A performance that still meets the requirements despite its shortcomings |
| 5.0 | fail (nicht ausreichend) | A performance that no longer meets the requirements due to significant defects |

For the differentiated evaluation of the assignment performances, the grades can be raised or lowered by 0.3 to intermediate values; permissible grades are 1.0, 1.3, 1.7, 2.0, 2.3, 2.7, 3.0, 3.3, 3.7, 4.0, and 5.0.

Communication and Advice on Academic Matters

I am available for advice on academic matters via email at p [dot] karapanagiotis [at] rug [dot] nl. If you wish to meet with me in person, please set up appointments ahead of time to make sure I am available. Appointments can be arranged by email.

Reading Material

Textbook The most important source for the course is [RDS] (Wickham, Getinkaya-Rundel, and Golemund 2023). It can be found online in an accessible [e-book](#) format. It gives a modern introduction to R, emphasizing the most prevalent data science activities.

Course Blog The course discussion topics are available online. The blog contains slides, handouts, and interactive components (figures, videos, dashboards) that can be used when studying. Last but not least, the blog has indices for all the course concepts, learning activities, and applications, with links to the corresponding material section to which they are introduced.

Recommended Supplementary Resources

Textbooks Two additional more advanced books are [ISLR] (Gareth et al. 2014) and [DLR] (Chollet, Kalinowski, and Allaire 2022). The first is an intermediate-level book with a comprehensive hands-on approach to statistical learning with R. The second is more

advanced and requires some programming experience and a more in-depth understanding of the R programming language. It covers deep learning for classification, computer vision, and time series applications using [Keras](#) and [TensorFlow](#) in R.

Online Resources Helpful online resource for prompt engineering and generative AI can be found at OpenAI’s [Cookbook](#) and Microsoft Cloud Advocates’ [Generative AI for Beginners \(Version 3\)](#).

Course Schedule

The course schedule is a general guide of activities for the course. The schedule is subject to change. The instructor will notify participants of specific schedule changes as needed. Pre-reading assignments should be completed before each corresponding session.

Table 2: Organization

| Date | Topics | Reading | Assignment |
|----------------|--|---|--|
| May 9th, 2025 | Sections (1) and (2) | RDS chap. 1-8, 28, 29 | Quiz 1 |
| May 10th, 2025 | Sections (3) and (4. Part I) | <ul style="list-style-type: none"> • RDS chap. 9-15 • ggplot2 • dplyr | Group Assignment 1 |
| May 30th, 2025 | Sections (4. Part II), (5), and (6 Part I) | <ul style="list-style-type: none"> • RDS chap. 16-24 • ISLR chap. 3 • dplyr • neermatch | <ul style="list-style-type: none"> • Quiz 2 • Coding Challenge |
| May 31st, 2025 | Sections (6 Part II), (7), and (8) | <ul style="list-style-type: none"> • RDS chap. 25, 26, 27 • markets | Group Assignment 2 |

In more detail, the course has the following structure:

1. Preamble (3 hours)
 - a. What is generative AI?
 - b. How do generative tools impact the development process?
 - c. Introduction to fundamental programming concepts and practices.
 - d. Overview of the R programming language.
 - e. Why R?
 - f. A preview of the R ecosystem. What can one do with R?
 - g. Setting up the development environment.
2. An Overview of Data Science Activities with Generative Tool Workflows (4 hours)
 - a. Brainstorming and co-designing with generative tools.

- b. Visualization.
 - c. Transformation.
 - d. Programming.
 - e. Data management.
 - f. Analysis and presentation.
 - g. Quiz 1: Generative AI and data science operations.
3. Data Visualization (4 hours)
- a. Overview of the [ggplot2](#) library.
 - b. Basic plotting with [ggplot2](#).
 - c. Aesthetics.
 - d. Coordinate systems and geometric transformations.
 - e. Annotations and Labels.
 - f. Best practices for data visualization with generative tools.
 - g. Group Assignment 1: Data visualization.
4. Data Analysis and Transformation (6 hours)
- a. Introduction to data transformations.
 - b. Data types.
 - c. Overview of the [dplyr](#) library.
 - d. Data manipulation and transformation.
 - e. Aggregation and grouping operations.
 - f. Merging and joining data.
 - g. Best practices in data analysis with generative tools.
 - h. Quiz 2: Data analysis and transformation.
5. Working with Large Datasets (4 hours)
- a. Databases.
 - b. Reading and writing data.
 - c. Working with unstructured data and generative tools.
 - d. Web scraping.
 - e. Merging and linking data.
 - f. Coding challenge.
6. Econometrics (3 hours)
- a. Formulas.
 - b. Regressions and Estimations.
 - c. Market models.
7. Programming (4 hours)
- a. Control structures (if-else, loops).
 - b. Functions and libraries.

- c. Lists and vectors.
 - d. Input-output operations.
 - e. Best practices in R programming.
 - f. Group Assignment 2: Programming with code companions.
8. Epilogue (1 hour)
- a. Where to go from here?
 - b. A preview of statistical and machine learning with R.

References

- Chollet, F., T. Kalinowski, and J. J. Allaire. 2022. *Deep Learning with R, Second Edition*. 2nd ed. Manning.
- Gareth, James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2014. *An Introduction to Statistical Learning: With Applications in R*. Springer Publishing Company, Incorporated.
- Wickham, Hadley, Mine Getinkaya-Rundel, and Garrett Golemund. 2023. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. 2nd ed. O'Reilly Media. <https://r4ds.hadley.nz/>.