COURSE DESCRIPTION CARD - SYLLABUS

Course name
Introduction to Programming in R [S1DSwB1>WdPwR]

Course

Field of study
Data Science in Business

Year/Semester
3/5

Area of study (specialization)
–

Profile of study
general academic

Level of study
first-cycle

Course offered in
Polish

Form of study
full-time

Requirements
compulsory

Number of hours

Lecture
30

Laboratory classes
0

Other
0

Tutorials
30

Projects/seminars
0

Number of credit points

5,00

Coordinators

dr hab. inż. Yevhen Revtiuk
yevhen.revtiuk@put.poznan.pl

Lecturers

Prerequisites

Basic knowledge of computer operation. Ability to install and run software. Basic file and directory operations (e.g., saving, moving, copying). Familiarity with basic mathematical operations. Understanding of concepts such as variables, functions, and data arrays. Logical thinking and problem- solving skills. Ability to analyze simple problems and structure them effectively.

Course objective

The objective of the course is to introduce students to the fundamentals of programming in R, with a particular focus on data analysis and visualization. Students will learn the basics of R syntax, data handling, and the use of R packages for solving computational and statistical problems. The course will prepare participants for independent work with R in the context of engineering, scientific disciplines, and data analysis.

Course-related learning outcomes

Knowledge:
Characterizes basic data analysis methods in R, including operations on data structures, programming, and result visualization [DSB1_W01].

Describes data processing concepts in R, including data importing, cleaning, and transformation [DSB1_W02].
Explains statistical methods and exploratory data analysis techniques available in R packages and their application in business analysis [DSB1_W03].
Presents data visualization techniques and tools for creating charts using ggplot2 and other libraries [DSB1_W07].
Characterizes basic econometric and statistical models implemented in R, including linear regression and correlation analysis [DSB1_W09].

Skills:
Designs and implements data analyses in R, utilizing operations on vectors, data frames, and lists [DSB1_U03].
Analyzes and visualizes data from various sources, applying statistical and graphical functions available in R [DSB1_U04].
Formulates analytical problem specifications and selects appropriate methods for data exploration and analysis in R [DSB1_U05].
Justifies the choice of analytical tools and methods in the R environment and evaluates the effectiveness of applied solutions [DSB1_U11].
Develops programming skills in R by analyzing documentation and utilizing available educational resources [DSB1_U15].

Social competences:
Critically analyzes their own programming and statistical skills, striving for development and updates in the context of new data analysis methods [DSB1_K01].
Utilizes current scientific advancements in data analysis and R programming, considering their practical applications in business [DSB1_K02].

Methods for verifying learning outcomes and assessment criteria

Learning outcomes presented above are verified as follows:

Final Project Assessment (50%)
Description:
Students prepare a final project in which they conduct a comprehensive data analysis using R. The project must include data acquisition, preprocessing (data cleaning), statistical analysis, and presentation of results in the form of visualizations and a report. A key advantage will be the ability to apply appropriate R libraries and functions depending on the analytical problem.
Assessment Criteria:
• Quality and accuracy of data analysis: Correctness of analytical methods, effectiveness of data cleaning and preparation, proper selection of analytical techniques based on data type.
• Programming skills in R: Knowledge of syntax, use of functions and R packages in the context of problem-solving, code structure, readability, and efficiency.
• Originality and creativity: Innovative approach to data analysis, creative use of programming tools.
• Clarity of the report and communication of results: Clear presentation of results, transparency of analysis, justification of applied methods and visualizations, accuracy, and coherence of the report.
Practical Exercises and Homework Assignments (50%)
Description:
Regular homework assignments and practical exercises during classes, focusing on solving programming problems and data analysis in R. Tasks will cover both basic data operations (loading, cleaning) and more advanced topics such as statistical analysis and data visualization.
Assessment Criteria:
• Accuracy of task execution: Correctness of solutions, proper implementation of algorithms, accuracy of calculations.
• Application of appropriate R functions: Ability to select and apply the right tools and functions for each task (e.g., dplyr, ggplot2, tidyr packages).
• Timeliness and independence: Timely submission of assignments and independent problem-solving skills.
• Presentation of results: Ability to present analysis results in the form of charts, tables, or reports, tailored to the audience (e.g., charts for analysis vs. charts for presentation).

Programme content

1. Introduction to R
• Brief history and applications of R.
• Installation and configuration of R and RStudio.
• Basic operations in the R console.
2. Basic Syntax of R
• Data types (numeric, logical, character).
• Arithmetic, logical, and assignment operators.
• Variables and basic operations on them.
3. Data Structures in R
• Vectors and operations on vectors.
• Matrices and arrays.
• Data frames - basic operations, filtering, sorting.
• Lists - structure and applications.
4. Working with Data in R
• Importing and exporting data (CSV, Excel, databases).
• Basic operations on datasets (merging, filtering, aggregation).
• Handling missing values and data cleaning.
5. Programming in R
• Conditional statements (if, else).
• Loops (for, while) and alternative solutions (apply, lapply, sapply).
• Defining custom functions.
6. Data Visualization
• Basic plots in R (histograms, boxplots, scatter plots).
• Creating visualizations using the ggplot2 package.
• Customizing plots and exporting results.
7. Basics of Data Analysis and Statistics in R
• Basic descriptive statistics (mean, median, standard deviation).
• Using statistical functions and significance tests.
• Correlation analysis and linear regression.
8. Mini Final Project
• Independent problem-solving in data analysis using R.
• Presentation of results in the form of a report or presentation.

Course topics

1. Introduction to R and the RStudio Environment
• Installation and configuration of R and RStudio.
• First steps in R - basic operations in the console.
• Basic syntax of the R language.
2. Data Types and Structures in R
• Basic data types (numeric, logical, character).
• Vectors, matrices, and arrays - creation and operations.
• Data frames - filtering, sorting, and modifications.
• Lists and their applications.
3. Basic Data Operations
• Importing and exporting data (CSV, Excel, databases).
• Data cleaning and transformation.
4. Programming in R
• Conditional statements (if, else).
• Loops (for, while) and alternative iteration methods (apply, lapply, sapply).
• Creating custom functions.
5. Data Visualization in R
• Creating basic plots (histograms, scatter plots, boxplots).
• Advanced data visualization with the ggplot2 package.
• Customizing plots and exporting results.
6. Basic Statistics and Data Analysis
• Descriptive statistics (mean, median, variance, standard deviation).
• Correlation analysis and statistical tests.
• Linear regression in R.

7. Advanced Data Analysis (Optional for More Advanced Groups)
• Clustering and cluster analysis.
• Introduction to multiple regression and predictive modeling.
• Large-scale data processing.
8. Integrating Essential Statistical Concepts
• Throughout the course, fundamental statistical concepts necessary for data analysis in R will be introduced.

Teaching methods

Theoretical Lectures
Key topics related to programming in R will be covered during lectures, including language syntax, basic data types, data structures, data manipulation, and fundamentals of statistical analysis. The lectures will introduce students to the theoretical aspects of working with R and present principles that will be applied in practical exercises.
Practical Exercises
Practical classes will focus on hands-on work with R, where students will independently complete exercises related to data analysis, algorithm implementation, and result visualization. Both individual and group assignments will provide opportunities to apply theoretical knowledge in practice.
Students will continuously apply discussed concepts by solving specific data processing problems in R.
Computer Labs
In these sessions, students will work on their laptops, directly engaging with R. Each student will have access to the necessary software to independently solve problems and complete assignments.
Instructors will be available for assistance and consultation.
The goal of the labs is to help students gain practical programming skills in R and learn how to efficiently work with data.
Individual Consultations
Students will have the opportunity to meet with the instructor for individual consultations to discuss difficulties related to the material or assignments. Consultations will cover both theoretical concepts and practical aspects of programming in R.
Case Studies and Real Data Analysis
Real-world data examples will be analyzed in class, allowing students to apply their skills to solve practical programming challenges. Case studies will reinforce learning and increase motivation by demonstrating real-world applications of R programming.
Presentations and Discussions
Students will present the results of their projects or solutions to assignments, improving their ability to communicate results and defend their approaches. They will also participate in discussions on selected topics related to programming and data analysis in R.

Bibliography

Basic:
1. Crawley, M. J. (2013). The R Book. Wiley.
2. Kabacoff, R. I. (2015). R in Action: Data Analysis and Graphics with R. Manning.
3. Wickham, H., & Grolemund, G. (2017). R for Data Science. O'Reilly Media.
4. Zuur, A. F., Ieno, E. N., & Meesters, E. H. W. G. (2009). A Beginner's Guide to R. Springer.

Additional:
1. Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer.
2. Matloff, N. (2011). The Art of R Programming: A Tour of Statistical Software Design. No Starch Press.
3. Xie, Y. (2016). R Markdown: The Definitive Guide. CRC Press.
4. Venables, W. N., & Ripley, B. D. (2002). Modern Applied Statistics with S. Springer.
5. Peng, R. D. (2016). R Programming for Data Science. Leanpub.

Breakdown of average student's workload

	Hours	ECTS
Total workload	125	5,00
Classes requiring direct contact with the teacher	62	2,50
Student's own work (literature studies, preparation for laboratory classes/ tutorials, preparation for tests/exam, project preparation)	63	2,50