University of Gdańsk

Subject card

Subject name and code	Data Analysis and Visualisation in R, PG_00210490		
Field of study	Spatial Management		
Date of commencement of studies	October 2025	Academic year of realisation of subject	2026/2027
Education level	Master's studies	Subject group	Obligatory subject group in the field of study Optional subject group
Mode of study	full-time studies	Mode of delivery	at the university
Year of study	2	Language of instruction	English
Semester of study	3	ECTS credits	2.0
Learning profile	academic	Assessment form	credit
Conducting unit			
Name and surname of lecturer (lecturers)	Subject supervisor	dr Michał Marosz	
	Teachers		

Lesson types

Lesson type	Lecture	Tutorial	Laboratory	Project	Seminar	SUM
Number of study hours	30.0	0.0	0.0	0.0	0.0	30

E-learning hours included: 0.0

Learning activity and number of study hours

Learning activity	Participation in didactic classes included in study plan	Participation in consultation hours	Self-study	SUM
Number of study hours	30	0.0	20.0	50

Subject objectives

The aim of the course is to develop practical skills in importing, processing, analysing, interpreting and visualising data using the **R programming language** and selected packages from the **tidyverse** ecosystem. Students will learn the basic principles of working with data, preparing datasets for analysis, creating effective visualisations, and presenting analytical results in a clear, reproducible and methodologically sound way.

Learning outcomes

Course outcome	Subject outcome	Method of verification
	Is able to create and interpret visualisations and simple analytical reports in R, using the ggplot2 package, code, results, graphs and interpretative commentary to draw valid conclusions.	[SK2] presentation/project/paper/ report [SK4] test/exam - oral or written
	Can import, clean, transform and analyse data in R, including identifying missing data, checking variable types, performing aggregations, joining tables and carrying out basic exploratory data analysis.	[SK2] presentation/project/paper/ report [SK4] test/exam - oral or written
	Knows the basic components of the R working environment and understand the principles of organising a simple analytical project, comprising scripts, objects, data frames and a repeatable workflow.	[SK2] presentation/project/paper/ report [SK4] test/exam - oral or written

Subject contents	## Course content 1. **Introduction to R and RStudio** * basic elements of the working environment, * scripts, console and R projects, * installing and loading packages, * basic types of objects in R. 2. **Basics of working with data in R** * data frames and tibbles, * structure of datasets, * types of variables, * basic operations on objects and variables. 3. **Importing data from different sources** * importing CSV and TXT files, * importing spreadsheets, * using sample datasets available in R packages, * checking the correctness of imported data. 4. **Data cleaning and preparation for analysis** * identifying missing values, * checking and changing variable types, * detecting basic errors in data, * cleaning column names and variable values. 5. **Data transformation using tidyverse packages** * filtering observations, * selecting and renaming variables, * creating new variables, * grouping and aggregating data, * sorting data, * joining tables. 6. **Reshaping data** * data in wide and long formats, * reshaping data using `pivot_longer()` and `pivot_wider()`, * preparing data for analysis and visualisation. 7. **Basic exploratory data analysis** * describing variable distributions, * basic measures of descriptive statistics, * analysing relationships between variables, * comparing groups, * identifying outliers. 8. **Introduction to data visualisation in R** * the role of visualisation in data analysis, * selecting the appropriate type of chart for the data, * basic principles of clear and accurate visualisation. 9. **Creating charts using ggplot2** * structure of a chart in `ggplot2`, * scatter plots, line charts, bar charts and histograms, * boxplots and density plots, * mapping aesthetics, * grouping data in charts. 10. **Customising chart appearance** * axis scales, * colours and colour palettes, * labels, titles and legends, * graphical themes, * improving chart readability. 11. **Presenting variability and uncertainty in data** * confidence intervals, * error bars, * presenting dispersion and variability in data, * interpreting uncertainty in charts. 12. **Interpreting analytical and visual results** * formulating conclusions based on results, * identifying key relationships and patterns,

	* avoiding interpretation errors. 13. **Reproducibility of analyses** * organising an analytical project, * good practices in writing code, * documenting the stages of analysis, * basics of working with an analytical report. 14. **Preparing a simple analytical report** * combining code, results, charts and commentary, * structure of an analytical report, * presenting results in a clear and reproducible way. 15. **Independent work with documentation and problem solving** * using help in R, * interpreting error messages, * searching for solutions in documentation and R community resources, * good practices for further development of analytical skills.
Prerequisites and co-requisites	Ability to use a PC at a basic level. Knowledge of the basics of descriptive statistics.

Assessment methods and criteria

Subject passing criteria	Passing threshold	Percentage of the final grade
Test - practical use of R in data anlaysis and visualisation	51.0%	70.0%
Project - real data analysis (group work + presentation)	51.0%	30.0%

Recommended reading

Basic literature	## Basic reading list 1. Wickham, H., Çetinkaya-Rundel, M., Grolemund, G. **R for Data Science. Import, Tidy, Transform, Visualize, and Model Data**. 2nd edition. O'Reilly Media. A core textbook for learning data analysis in R using the `tidyverse` packages. 2. Wickham, H., Navarro, D., Pedersen, T. L. **ggplot2: Elegant Graphics for Data Analysis**. Springer. A key reference for creating data visualisations in R using the `ggplot2` package. 3. Chang, W. **R Graphics Cookbook. Practical Recipes for Visualizing Data**. 2nd edition. O'Reilly Media. A practical source of examples and solutions for data visualisation in R. 4. Healy, K. **Data Visualization: A Practical Introduction**. Princeton University Press. An introduction to the principles of effective, clear and responsible data visualisation using R and `ggplot2`. 5. R Core Team. **An Introduction to R**. The R Project for Statistical Computing. The official introductory manual to the R language, covering syntax, basic data structures, graphics and statistical analysis. 6. Grolemund, G. **Hands-On Programming with R**. O'Reilly Media. An accessible introduction to programming in R, especially useful for beginners.
Supplementary literature	1. Ismay, C., Kim, A. Y., Valdivia, M. **Statistical Inference via Data Science: A ModernDive into R and the Tidyverse**. 2nd edition. CRC Press. An introduction to data analysis, visualisation, regression and the basics of statistical inference using R and the `tidyverse`.
eResources addresses	Basic http://www.biecek.pl/R/ - R study tools shared by Przemysław Biecek https://education.rstudio.com/learn/ - RStudio-Education. Finding Your Way To R https://ggplot2-book.org/ - ggplot2: Elegant Graphics for Data Analysis https://r-graphics.org/ - R Graphics Cookbook, https://r4ds.hadley.nz/ -

Example issues/ example questions/ tasks being completed	## Sample practical test tasks
	
	### Task 1. Data import and initial data checking
	
	Import the file `students_results.csv`, which contains data on students' results from several tutorial groups.
	
	The file contains the following variables:
	
	* `student_id` – student identifier,
	* `group` – tutorial group,
	* `test_1` – result of the first test,
	* `test_2` – result of the second test,
	* `project` – project result,
	* `attendance` – number of attended classes.
	
	Complete the following steps:
	
	1. Import the data into R.
	2. Check the structure of the data.
	3. Check the number of observations and variables.
	4. Check whether there are any missing values in the dataset.
	5. Calculate basic descriptive statistics for numerical variables.
	
	--- 	
	
	### Task 2. Data cleaning and preparation
	
	In the `students_results.csv` dataset, some group names are recorded inconsistently, for example: `A`, `group A`, `Grupa A`, `B`, `group B`.
	
	Complete the following steps:
	
	1. Standardise the group names so that they have the form `A`, `B`, `C`.
	
	2. Replace any missing values in the `attendance` variable with `0`.
	
	3. Create a new variable `final_score`, calculated as a weighted average of the results:
	
	  * `test_1` – 30%,
	  * `test_2` – 30%,
	  * `project` – 40%.
	
	4. Create a new variable `passed`, which takes the value:
	
	  * `"yes"` if `final_score >= 50`,
	  * `"no"` if `final_score < 50`.
	
	5. Display the first 10 rows of the cleaned dataset.
	
	--- 	
	
	### Task 3. Grouping, aggregation and interpretation of results
	
	Using the prepared dataset, calculate a summary of results for each tutorial group.
	
	For each group, calculate:
	
	1. the number of students,
	2. the mean final score,
	3. the median final score,
	4. the minimum and maximum final score,
	5. the percentage of students who passed the course.
	
	Then:
	
	6. arrange the groups in descending order according to the mean final score,
	7. indicate the group with the highest mean final score,
	8. write a short interpretative comment consisting of 2–3 sentences.
	
	--- 	
	
	### Task 4. Data visualisation using ggplot2
	
	Using the `students_results.csv` dataset, prepare three charts:
	
	1. a histogram of the `final_score` variable,
	2. a boxplot showing the variation in `final_score` across groups,
	3. a bar chart showing the percentage of students who passed the course in each group.
	
	Each chart should include:
	
	* a title,
	* labelled axes,
	* a clear legend, if needed,
	* a readable graphical theme, for example `theme_minimal()`.

	Finally, choose one of the charts and write a short interpretation of the results shown. --- ## Task 5. Simple analytical report Prepare a short analytical report in a `.qmd` or `.Rmd` file using the `students_results.csv` dataset. The report should include: 1. a report title, 2. a short description of the analysed dataset, 3. code for importing the data, 4. a basic summary of the data, 5. a table with results aggregated by group, 6. at least two charts created with `ggplot2`, 7. a short interpretative comment on the results, 8. final conclusions in 3–5 sentences. The report should be possible to render as an HTML or PDF document.
Work placement	Not applicable

Document generated electronically. Does not require a seal or signature.