COURSE DESCRIPTION CARD - SYLLABUS

Course name
Data Visualization in R [S1DSwB1>WDwR]

Course

Field of study
Data Science in Business

Year/Semester
3/6

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

4,00

Coordinators

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

Lecturers

Prerequisites

The course assumes that students have the following skills and knowledge: 1. Basic Knowledge of Mathematics and Statistics: Understanding basic statistical concepts such as mean, median, standard deviation, probability distributions, correlation, and regression will be helpful for data analysis. 2. Computer Literacy: Basic computer skills, including familiarity with operating systems, installing software, and using essential office tools. 3. Basic Knowledge of R Programming: Students should have a basic understanding of R, including working with data, using fundamental functions and packages (e.g., dplyr, ggplot2, tidyr). The course expects students to be able to import and manipulate data, perform basic analysis, and create simple visualizations using R. 4. Understanding of Data Concepts: Ability to work with different data types, such as numerical, categorical, and time-series data, as well as understanding basic data quality concepts (e.g., missing values, outliers).

Course objective

The objective of the course "Data Visualization in R" is to introduce students to basic data visualization methods using the R programming language. Students will learn to create various types of charts and visualizations that will help with data interpretation and support decision-making processes. The course aims to develop data analysis skills through visualization, familiarize students with tools and packages in R, such as ggplot2 and plotly, and implement best practices in presenting analysis results. Upon completion, students will be able to independently create clear and effective charts tailored to the characteristics of the data and the needs of the audience.

Course-related learning outcomes

Knowledge:
Characterizes data exploration and visualization techniques in R, including packages like ggplot2, plotly, lattice, and base R graphics [DSB1_W02].
Describes methods for preparing data for visualization, including data cleaning, transformation, and aggregation [DSB1_W03].
Explains the principles of visual perception and color theory in the context of designing effective data visualizations [DSB1_W07].

Skills:
Selects appropriate data visualization methods in R, adapting them to the type of data and analysis objectives [DSB1_U02].
Analyzes and visualizes data from various sources using different types of charts, such as scatter plots, bar charts, histograms, box plots, and heat maps [DSB1_U04].
Designs and implements interactive data visualizations using advanced techniques like faceting, multidimensional visualization, and geospatial visualization [DSB1_U08].
Uses machine learning tools to analyze relationships and predict trends based on data visualizations [DSB1_U09].

Social competences:
Critically analyzes their own skills in data visualization, striving for continuous improvement and alignment with current standards [DSB1_K01].
Engages in initiatives related to data analysis and visualization, promoting best practices in visualization for business analysis and science [DSB1_K03].

Methods for verifying learning outcomes and assessment criteria

Learning outcomes presented above are verified as follows:

The evaluation for the course is divided into two main components:
1. Final Project (50%)
o Description: Students will prepare a final project that includes creating a comprehensive data visualization based on a real dataset. The project must involve data analysis, selection of appropriate visualization tools, and presentation of results.
o Evaluation Criteria:
Quality and correctness of visualizations (appropriate charts for the data, readability, and aesthetics).
Ability to apply visualization techniques in data analysis.
Originality, creativity in choosing tools and methods.
Clarity of the report, justification for chart selection, and clear communication of results.
2. Practical Exercises and Homework (50%)
o Description: Regular homework assignments and practical exercises during the classes, focusing on creating data visualizations using R. The tasks include both basic charts and more advanced analyses involving multiple variables.
o Evaluation Criteria:
Accuracy of task completion and correctness of the created charts.
Use of appropriate visualization techniques depending on the data.
Timeliness and independence in completing tasks.
Ability to adjust charts for different audiences (e.g., presentation charts vs. publication charts).

Programme content

The course program includes the following topics:
1. Reminder of Basic Statistical Concepts
o Creating various statistics for later visualization.
2. Introduction to R and Visualization Tools
o Basic programming principles in R.
o Installation and configuration of R and RStudio.
o Discussion of packages: ggplot2, plotly, lattice, base R graphics.
3. Basics of Data Visualization with ggplot2
o The concept of "Grammar of Graphics."
o Structure of plots in ggplot2 (aes(), geom_, theme_).
o Creating scatter plots, bar plots, and line plots.
4. Data Processing and Preparation for Visualization
o Data cleaning and transformation (using packages like dplyr, tidyr).
o Data merging and filtering.
o Formatting data for plots.
5. Advanced Graphs and Visualization Techniques
o Histograms, box plots, heat maps.
o Multidimensional plots and faceting.
o Correlation matrix visualization.
o Visualizing data on maps using leaflet.
6. Best Practices in Data Visualization
o Visual perception and color theory.
o How to avoid mistakes and manipulation in visualization.
7. Case Studies and Student Projects
o Visualizing real-world data (e.g., financial, weather, social data).
o Creating reports and data presentations.
o Overview of tools supporting data analysis (Markdown, R Notebooks).

Course topics

The course program includes the following topics:
1. Discussion on the use of data visualization in R for data science, with examples from statistics, management, and marketing - this part is conducted during lectures.
2. Introduction to Data Visualization and R Programming
o The importance of data visualization in analysis and presentation of results.
o Overview of available data visualization tools in R.
o Installation and configuration of R packages (e.g., ggplot2, plotly).
o Basic data operations in R: importing, processing, and cleaning data.
3. Basic Types of Charts
o Scatter plots, histograms, bar charts.
o Differences between charts and when to use them.
o Creating basic plots in ggplot2.
4. Line and Time-Series Charts
o Creating line charts for time-series data analysis.
o Using the ggplot2 package for drawing time-series plots.
o Manipulating axes and labels on charts.
5. Visualization of Categorical Data
o Bar charts and pie charts.
o Comparing different types of categorical charts and their applications.
o Working with table data and grouped charts.
6. Visualization of Relationships Between Variables
o Box plots and scatter plots.
o Using charts to analyze relationships between variables.
o Grouped data charts (e.g., charts with colors representing grouping variables).
7. Working with Maps and Geospatial Data
o Visualizing geospatial data in R: maps, geospatial charts.
o Visualizing spatial data: locations, boundaries, and geographical analysis.
8. Aesthetics and Customization of Charts
o Personalizing charts: colors, labels, titles, and legends.
o Customizing charts for different audiences: publication charts vs. presentation charts.

o Discussing principles of good chart design (e.g., Tufte's principles).
9. Examples of Data Analysis and Presentation of Results
o Case analysis: how to choose the appropriate chart type based on data.
o Creating reports with data analysis results.
o Presenting visual results: how to effectively communicate data through charts.
10. Course Summary and Final Project
o Discussion of final projects: independently creating data visualizations.
o Summing up the knowledge and skills gained.
o Presentation of student projects, analyzing the results.

Teaching methods

The teaching methodology for the course includes:
1. Interactive Lectures:
o Short introductory lectures on data visualization in R, covering key concepts, tools, and principles of good chart design.
o Examples of the application of visualization in various fields (e.g., economic analysis, social sciences, statistics, management) will be presented, along with a discussion on best practices.
o Lectures will be interactive, encouraging students to ask questions and discuss examples in real-time.
2. Practical Exercises (Laboratory):
o Classes based on solving problems, where students practically apply R and packages for creating charts and analyzing data.
o Students will be given data to analyze and create appropriate visualizations based on specific tasks.
o The exercises will be conducted in workshop format, allowing individual work, with the instructor providing support as needed.
3. Individual Consultations:
o Students will have the opportunity to attend individual consultations with the instructor to discuss difficulties related to the material or tasks. The consultations will cover both theoretical and practical aspects of programming in R.

Bibliography

Basic:
1. Hadley Wickham, Garrett Grolemund - R for Data Science (2017)
2. Kieran Healy - Data Visualization: A Practical Introduction (2018)
3. Claus O. Wilke - Fundamentals of Data Visualization (2019)

Additional:
1. Hadley Wickham - ggplot2: Elegant Graphics for Data Analysis (2016),
2. Nathan Yau - Data Points: Visualization That Means Something (2013)
3. Fernando Miguez - Data Visualization with R (2022)
4. David McCandless - Information is Beautiful (2009)

Breakdown of average student's workload

	Hours	ECTS
Total workload	100	4,00
Classes requiring direct contact with the teacher	60	2,50
Student's own work (literature studies, preparation for laboratory classes/ tutorials, preparation for tests/exam, project preparation)	40	1,50