HOCHSCHULE OSNABRÜCK
UNIVERSITY OF APPLIED SCIENCES

Syllabus

Modern Data Analytics with R

Instructor: Prof. Dr. Nicolas Meseth

Module number: 44B0586

Module description: https://www.hs-osnabrueck.de/module/44b0586/

Course Description

The module introduces the basics of modern data analytics with R. The students acquire the tools to perform the main steps in the exploratory data analysis cycle proposed by [1]. This includes loading data from various sources, transforming data, visualizing data, and communicating data in an appropriate format.

Learning Objectives

- **Data Analysis Pitfalls**: Know and identify common challenges in data analysis and learn strategies to mitigate them.
  - Potential problems at the source of the data
  - False outcomes due to inappropriate data analysis methods or wrong application
  - Unclear or misleading messages through bad communication and visualization

- **Exploratory Data Analysis (EDA) Process**: Recall the essential steps involved in conducting an exploratory data analysis.

- **Conducting EDA with R & Tidyverse**: Apply R and the Tidyverse tools to perform exploratory analysis on unfamiliar data sets.
  - **Loading Data from Various Sources**: Import and manage data from diverse sources using R and Tidyverse tools, adapting techniques to accommodate different data formats and structures.
  - **Data Transformation with dplyr**: Apply dplyr functions within the Tidyverse for effective data transformation tasks.

- **Visualization Decision-Making**: Analyze data scenarios to determine suitable visualization types.
- **Chart Creation with R & ggplot2**: Design and implement various chart types using R and ggplot2.
- **Communication of Results with Quarto**: Effectively communicate and present data analysis results, encompassing the creation of clear, well-structured, and visually appealing reports and presentations.

Learning Content

- Loading data with `readr` and `readxl`
  - Loading data from various sources (CSV, URL, Excel, R-Data)
  - Changing column metadata (name, data type)
  - Checking data after loading (data types, missing values, data ranges)
  - Tidying data (long form, atomic values, no summary rows)

- Data transformation with `dplyr`
  - Picking columns with `select`
  - Reducing rows with `filter`
  - Sorting rows with `arrange`
  - Adding new columns and changing existing ones with `mutate` or `transmute`
  - Aggregate rows with `group_by` and `summarize`

- Data visualization with `ggplot2`
  - The grammar of graphics
  - Simple plots for trends & developments, amounts & proportions, distributions, and associations:
    - Line chart
    - Bar chart
    - Histogram, box plot, ridgeline plot
    - Scatter plot, heat map
  - Tuning a visualization

- Communicating findings
  - Blending text and code in reusable reports with Quarto
    - Options for hiding code, results, warnings etc.

    - Integration of plots
    - Using inline-R for data values within text
- Cross-references
- Table of contents and bibliography
- Output formats (HTML, Word, PDF)
- Parameters for dynamic reports

Course Schedule

Required Text and Material

Wickham, Hadley, et al. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. 2nd edition, O'Reilly Media, Inc, 2023.

Wilke, C. *Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures*. First edition, O'Reilly Media, 2019.

Huntington-Klein, Nick. *The Effect: An Introduction to Research Design and Causality*. CRC Press, Taylor & Francis Group, 2022.

Assignments

Attendance and Participation

Attendance and participation are expected. If less than 3 students show up for a live session, the session will not take place.

Academic Integrity Policy