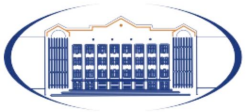
Data Analytics in R
Session 1

Maria Kunevich

Course instructor - *Maria Kunevich*

(msk41@tlu.ee)

Education:



SYKTYVKAR STATE UNIVERSITY

MA (specialist) - Education (Teaching Foreign Languages)

Exchange: Education
Minor: Linguistics

MPhil in Theoretical and Applied Linguistics

PhD candidate in Theoretical and Applied Linguistics

Master in Human-Computer Interaction

Research interests: language acquisition, interaction between learners and systems, data science techniques (experimental data analysis, data visualisation, machine learning, NLP)

What about you?

Miro board link:

https://miro.com/welcomeonboard/SUw3RHpXWjBpUDF2V0dwTWhkOFVVUnlUTnZ4Qm1oVGtSTVN0SmNCUmJyODhydU9hUzA3VUpNZVZHRnNBenhqVHwzMDc0NDU3MzY2NDE0OTE3Njc4?share_link_id=308470562965

Course information

This course provides an ***introduction*** to R-language and basic concepts in data analysis.

Three main parts:

- R environment and ecosystem
- Data project management: organisation of the working environment, data preparation, cleaning, visualization
- Data analysis: descriptive and inferential statistics, modelling
- Extra part: R as a programming language: writing your own functions and packages
- Extra part: communicating results - R Markdown

Course information

Learning outcomes:

At the end of this course students will be able to:

- use R for basic data project handling
- create and modify R datasets
- create figures and plots in R
- use R for basic descriptive analysis
- perform and interpret basic statistical tests (inferential statistics)
- perform and interpret basic statistical models
- communicate the results of their work/projects through RStudio environment and R Markdown

Topics for the course

- R and RStudio
- Basics of R-language
- Creating and accessing objects in R
- Reading and writing datasets
- Manipulating data
- Tidyverse - R packages for data science
- Managing data using the dplyr package
- Descriptive statistics in R (central tendency, spread, distribution).
- Visual data representation
- Statistical tests in R. Statistical models in R (Linear regression, Analysis of variance)

General information

How this course will work:

- **Weekly in-person sessions** for lectures and hands-on practice
- Active participation is highly encouraged
- The deadlines for all assignments are **strict deadlines** (usually Wednesday 23:59)
- Penalty is **-20%** from the **max grade** for the assignment for each extra late day
- The exact schedule with the assignments will be provided
- There is **no exam**

Grading of the course

- Max 100 points
- **10 homework assignments** (tutorials and quizzes), 5 points each (50 points)
- **1 paper summary** (20 points)
- **Project** completed individually or in a team of two (30 points)
- Grading: A starts from 91, B from 81, etc
- In order to pass the course, the student must have at least 51 point (grade E) in total and get at least **50%** from r**egular exercises** and a **project**

Grading of the course

- **10 homework assignments** for regular practice (DataCamp, tutorials on GitHub)
- Feedback for the tasks, using the data from the task as an illustration for data analysis
- **Paper review** - learn to analyse and provide an in-depth report on someone's work
- A template will be provided, try to find influential papers in your field that use R for data analysis
- Up to two pages, summarise the main ideas, findings, analyse what approaches to data analysis were taken, what are the advantages and limitations of such approaches
- **Project** - learn to use R for your own needs, project requirements will be provided
- You will present intermediate progress at the **end of November**
- Final presentation - in **December**

Course: materials and communication

Course webpage:

https://github.com/Maria-13/DataAnalytics_R

Communication:

https://github.com/Maria-13/DataAnalytics_R/discussions