Introduction

Ágoston Reguly
with
Mihály Orsós

Coding 1: Data management and analysis with R

2020

This course

This course introduces to R statistical programming.

The aim of this course is to teach you how to carry out a complete, reproducible data analysis project.

Agenda

We are going to cover:

1. How R and coding in general works.
2. How to use version control in your work.
3. Importing and exporting data.
4. Data wrangling (management, cleaning, set ready to analyse)
5. Visualization and carry out explanatory data analysis.
6. Creating your own functions, loops and simulations.
7. How to implement Data Analysis theory in practice.
8. How to create nice reports and presentations.
9. Some advanced analysis tool which might not covered in DA 2: spatial and financial time-series econometrics.

Office hours

Instructor: Ágoston Reguly (reguly_agoston@phd.ceu.edu)

- *Office hours*: Monday 10:30-12:00 or by appointment at N13 220.
- Weekends - I am not checking my email.

Teaching Assistant: Mihály Orsós (orsosm@ceu.edu)

- Here to help with any technical difficulties - great knowledge in R
- Instructor/mentor for R (Coding Practice with R)
- Later he will teach Coding 2: Web scraping with R

Course Material

The core course material is what we cover during the classes.

- Codes are going to be uploaded to github:
  - https://github.com/CEU-Economics-and-Business/ECBS-5208-Coding-1-Business-Analytics
  - Share your github repo with us! @regulyagoston and @misrori
- We are going to cover the tidyverse approach.
  - Note that there are many possible solutions in coding!

There is a great book by Garrett Grolemund and Hadley Wickham:

- R for data science
  - You can look up and enrich your knowledge there!
  - I will give optional homework from there! It is highly recommended to practice through these exercises!

Other additional materials

As we progress I will refer to other sources as well.

In general there are two additional great books:

- Kieran Healy (2018) Data Visualization
- Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani: An Introduction to Statistical Learning with Applications in R.

Assessment I

Participation (40%)

- Class participation (10%): attendance and active presence in classes.
- Helping each other (15%):
  - Ask for help in slack if you are stuck with something!
  - Help each other via commit to each other github account – allow access to each other!
  - It can be optional homework or the assignments.
- Demonstrate useful resources/materials (15%)
  - Reference your resource/material on slack channel
  - Demonstrate it in class (1-3 min)
  - Package with functions - 5%
  - Alternative solution to problems - 2.5%

Assessment II

- Assignments (30%)
  - Task 1 (15%): Joint with DA 1 - teamwork on descriptive report for your dataset.
    - Deadline: 25 October Sunday 23.55 - on a github repo (can decide you create a new one or put it into one member's repo.)
  - Task 2 (15%): Carry out a simple regression analysis of a given subject - individually
    - More details when we get closer in time.
    - Expected deadline: November 22 Sunday 23.55 -
- Take home examination (40%)
  - Carry out a complete data analysis project from scratch.
  - Joint project with Data Analysis 2.
  - Deadlines:
    - Choose your data and accepted by me - December 2 23.55
    - Submit codes - TBA

The points sums to 110%, so you can lose 10% without it affecting your grade at all. There will be no paper based exam.

Grading policy

- To pass, you will need to get at least 50% of the overall grade AND at least 50% on the take-home exam.
- Lectures - 75% attendance is required (8/12)
  - In case of online participation: write me with the reason unless there is a departmental change.