Syllabus

Coding 1: Data Management and Analysis with R

- **Instructor:** Agoston Reguly

  Email: reguly_agoston@phd.ceu.edu
  Office hours: Monday 10:30-12:00 or by appointment

- **Credits:** 2 (4 ECTS)
- **Term:** Fall 2020-2021
- **Module:** MS in Business Analytics
- **Course level:** [MA/MS]
- **Prerequisites:** Mathematics and Informatics Pre-session for Business Analytics
- **Course drop:** Course can be dropped free of charge 24 hours after the first session. After this date drop is possible until the course is halfway over (late drop fee applies). No changes are allowed past that date.

1. COURSE DESCRIPTION

The course serves as an introduction to the R programming language and software environment for data exploration, data munging, data analysis, and data visualization. As such it complements Data Analysis 1 and 2 which focus on the statistical methods, while this class focuses on how to program these methods in R.

2. LEARNING OUTCOMES

- Produce meaningful descriptive statistics and informative graphs
- Become familiar with the R ecosystem
- Learn how to use R for the most common data tasks: loading, cleaning, transforming, summarizing, and visualizing data

Other outcomes. The course will also help develop skills in the following areas:

Learning Area	Learning Outcome
Critical Thinking	Evaluate and create proper program codes which are reusable.
Quantitative Reasoning	Reasonable code which optimizes cognitive and computational effort.
Technology Skills	Become familiar with R ecosystem, use of R Markdown.
Interpersonal Communication Skills	Understand coding and informatic language, be able to communicate on these topics both online and offline.
Management Knowledge and Skills	Be able to create such reports which serves as a pillar for the management's decision.
Cultural Sensitivity and Diversity	Appreciate the different practices and solutions in coding.
Ethics and Social Responsibility	Create easily readable and reliable codes which can be used and reproduced by others as well.

3. READING LIST

Required:

- Class materials will be hosted on github.com – link will be shared at the first class.

Optional readings:

- Hadley W. - Garrett G. [**HG**] (2017): R for Data Science. Online version here.
- Kieran H. [**KH**] (2019): Data Visualization. Online version here.
- James, Witten, Hastie & Tibshirani [**JWHT**] (2013): Introduction to Statistical Learning. Online version and supplementary codes are available here.

Databases. The CEU Library boasts a range of databases covering financial and company data, market and industry reports, global news and more. For a full list of databases visit the CEU Library.

- Refinitiv (Thomson Reuters) Eikon for Students + Datastream/Thomson ONE
  - Eikon: Platform used by finance practitioners including market traders to monitor and analyze financial information. Information, analytics and news on all major financial markets including real-time pricing data, financial research, global financial news and commentary, financial estimates, fundamentals analysis, visual analysis through charting. Import/export from Excel.
  - Datastream: Range of economic, securities and company financial data. Excel add-in.
  - Thomson ONE: Global overviews on 55,000 public companies, one million private companies. Reuters News, ownership, deals, private equity, key ratios, company filings, officers and directors. Investext analyst reports, active and historical research from 1,600 independent research firms, brokerages, investment banks.
- Standard & Poor's Capital IQ
  - Web and Excel-based platform combining deep global company information, credit ratings and research, and market research with powerful tools for risk assessments. Real-time and historical information on markets, industries, companies, transactions and people. Tearsheet data.
- Lexis Nexis Academic
  - Global database of news, business, legal and other sources. Full text of 350 newspapers, 300 magazines and journals, 600 newsletters. Wire services including Associated Press, Business Wire and PR Newswire. Company financial information, market research, industry reports.

4. TEACHING METHODS AND LEARNING ACTIVITIES

The course will involve a mix of

- Lectures to present basic ideas in R.
- Live-coding classes to show and practice coding in R.
- Seminar-type classroom to solve coding problems.
- Optional homework to practice through problems and deepen the student's knowledge.

5. ASSESSMENT

- Participation (38%): helping yourself and helping others. This will be based on your contribution to the online class forum, as well as any other means you took to help yourself figure out how to solve issues you had. This can include getting help from classmates, or searching online forums and email lists.
- Assignments (30%): You will have two assignments during the semester to guide you through the course requirements and to be able to do the take-home exam.
- Take-home exam (40%): At the end of the course you will have to carry out a complete data analysis task from scratch. Take-home exam will evaluate your code and your success in completing this task.

This sums to 108%, so you can lose 8% without it affecting your grade at all. There will be no final exam.

6. TECHNICAL/LAPTOP REQUIREMENT

You will need your laptop with RStudio and R installed in order to participate in class.

7. TOPIC OUTLINE AND SCHEDULE

Session	Topics	Readings
September 28	R ecosystem, basic syntax, R Markdown; how to get help	HG Ch 2,4,21,23,24
October 5	Tidyverse: loading, showing, filtering, and selecting data	HG Ch 3,8,9,10,11; KH Ch 2; JWHT Ch 2
October 12	Tidyverse: transforming and summarizing data	HG Ch 3, 5
October 19	ggplot2: data visualization	HG Ch 1,22; KH Ch 3
October 26	Sampling, simulating, bootstrapping	JWHT Ch 5
November 2	Hypothesis testing	-
November 11	Basic Regressions	JWHT Ch 3
November 18	Functions, loops, and friends	Ch 15, 17
November 25	Debugging, testing, and documenting	HG Ch 22, 24
December 2	Basic Spatial Data	KH Ch 7
December 9	Basic Time Series	-
December 16	Recap and next steps	-

Note that readings are covering more and sometimes less than what we are actually covering in class. All necessary information will be given at the seminars.

8. SHORT BIO OF THE INSTRUCTOR

Ágoston Reguly is a fifth-year Economics PhD student at the Department of Economics and Business. He has worked more than three years at the Hungarian Government Debt Management Agency as an expert analyst. His research topics are in econometrics, policy evaluation with machine learning techniques and optimal survey designs.