

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 at N13 220
- **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 (40%):
 - a. Class participation (10%): attendance and active presence in classes.
 - b. Helping each other (15%):
 - i. Ask for help in slack if you are stuck with something!
 - ii. Help each other via commit to each other github account -- allow access to each other!
 - iii. It can be optional homework or the assignments.

- c. Demonstrate useful resources/materials (15%)
 - i. Reference your resource/material on slack channel
 - ii. Demonstrate it in class (1-3 min)
 - iii. Package with functions - 5%
 - iv. Alternative solution to problems - 2.5%
- Assignments (30%)
 - a. Task 1 (15%): Joint with DA 1 - teamwork on descriptive report for your dataset.
 - i. 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.)
 - b. Task 2 (15%): Carry out a simple regression analysis of a given subject – individual
 - i. Expected deadline: November 22 Sunday 23.55
- Take home examination (40%)
 - a. Carry out a complete data analysis project from scratch.
 - i. Joint project with Data Analysis 2.
 - ii. Deadlines:
 1. Choose your data and accepted by me - December 2 23.55
 2. Submit codes - TBA

This sums to 110%, so you can lose 10% 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	Introduction: Interface, type of input, ggplot, tryout of Rmarkdown	WG - Ch 1, 2, 4, 16
October 5	Github, dplyr, Data import and tidy approach with tibble: How to connect to github via R. Importing, exporting datasets with readr, data clean with dplyr	HG - 3, 7, 8, 9, https://happygitwithr.com/
October 12	Explanatory Data Analysis 1: summary statistics and graphs: descriptive functions, associations and ggplot more in depth	HG - 5, 22
October 19	Rmarkdown, hypothesis testing: Creating pdf/html/beamer, creating pretty tables and graphs in Rmarkdown, in-line report, Hypothesis testing: t-tests.	HG - Ch 21, 23
October 26	Explanatory Data Analysis 2: type of variables + advanced graphs, relational data, strings, factors, dates and times, PIPES	HG - 1, 10, 11, 12, 13, 14, 22
November 2	Functions, loops and friends: Writing functions and loops. Create bootstrap function	HG - Ch 15, 17
November 11	Univariate regression and simulating: lm function and its output. Simulating with sampling and confidence intervals. Predicting outcomes	JWHT - Ch 2, 3, 5
November 18	Univariate regressions II and error handling: Transformations, splines and other tricks, error handling for functions	-

November 25	Basic Spatial Data and visualization	-
December 2	Non-linear models: probit and logit: probit, logit models and their descriptives	-
December 9	Time series models: AR, ARMA, ARIMA, SARIMA, Philip-Perron test, GARCH models	-
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.