Syllabus

Department of Economics and Business
CENTRAL EUROPEAN UNIVERSITY

Data Visualization 2: Practical Data Visualization with R

Instructor: Gergely Daroczi
Email: daroczig@ceu.edu
Office: by appointment

Credits: 1 US credit (2 ECTS credits)

Term: Winter 2022-2023

Course level: Master's

Prerequisites: Data Analysis 1: Exploration – Business Analytics track;
Data Analysis 2: Finding Patterns with Regressions – Business Analytics track

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

In this hands-on course, you will become more proficient with data munging and preparations for data visualization: mastering ggplot2, the R implementation of the Grammar of Graphics. Besides creating and customizing static plots, you will also learn about integrating these outputs in automated or ad-hoc reports, with a short introduction into interactive dashboards as well.

2. LEARNING OUTCOMES

Key outcomes: By the end of the course you will become familiar with exploring data and expressing insights via the most often used plotting tool of the R ecosystem.

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

Learning Area	Learning Outcome
Critical Thinking	Good and bad practices in data visualization, eg choosing chart type to effectively communicate data points and statistics; or identifying misleading or confusing aspects of visualizations.
Quantitative Reasoning	Getting familiar with the data transformations (filtering, aggregation, pivoting, joining tables) required for the visualizations.
Technology Skills	Programming skills in R, mainly using ggplot2 for static plots, but also getting familiar with reporting from R using Rmarkdown, and a quick introduction to dashboarding using Shiny.

Interpersonal Communication Skills	
Management Knowledge and Skills	
Cultural Sensitivity and Diversity	
Ethics and Social Responsibility	

3. READING LIST

Class materials will be available on GitHub.

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 coding sessions with demos and exercises.
Learning objectives will be achieved through actively taking part in the in-class exercises and solving homework and the final take-home assignment.

5. ASSESSMENT (including minimum pass requirement and grading

40% homework and 60% final project.

Grading Policy

Students shall not miss more than 1 day of classes, failing to do so will yield an administrative fail grade. To pass, students will need to get at least 50% of the homework AND at least 50% of the final project.

Grading will be based on the total score out of 100, in line with CEU's standard grading guidelines.

6. TECHNICAL/LAPTOP REQUIREMENT

Laptop with R, RStudio and git installed is required in the class and for the take-home assignments. Installation steps will be shared on GitHub.

7. TOPIC OUTLINE AND SCHEDULE

Session	Topics	Readings
1	Recap on ggplot2, MDS, scaling variables, Simpson's Paradox, introduction to data.table	Shared on GitHub.
2	Clustering, animated plots, Anscombe's quartett, geocoding and basic maps	Shared on GitHub.
3	Boxplot, data transformations, visualization themes, interactive plots	Shared on GitHub.

8. SHORT BIO OF THE INSTRUCTOR

Gergely Daroczi has a PhD in Sociology, 15 years of experience with R, founder of the Hungarian R meetup and main organizer of R conferences, authored a book on R and maintains a dozen of R packages, lived and worked in Hungary and USA at market research, fintech, adtech and healthtech companies as a data scientist and engineer both in individual contributor and management roles.