IE 48A Sp. Topics in Essentials of Data Analytics

Syllabus

This course has a simple objective. You are expected to perform an end-to-end reproducible data analysis from raw data to the communication of model outputs. Warning: Course load is heavy.

Course Webpage TBA

Office Hours   by appointment

Contact   Email: berk.orbay@boun.edu.tr

Reference Materials

- *An Introduction to Data Science* (2013) version 3   by Stanton

- *Business Analytics for Managers* (2011)   by Jank

- Udacity Course – Data Analysis with R:
  https://classroom.udacity.com/courses/ud651

- Big Data @ Work: Dispelling the Myths, Uncovering the Opportunities   by Davenport

- Introduction to Statistical Learning   by Gareth et al.

- R for Data Science   by Wickham

- DataCamp

More recommended references on online platforms (Course Page).

Course Expectations

- Proven ability to manipulate data sets and creating summary tables

- Proven ability to visualize data with the proper choice of tools (e.g. histogram, scatterplot, pie charts)

- Proven ability to code with R and related packages (e.g. tidyverse) or equivalent in Python/Julia

- Proven ability to perform input/output operations

- Proven ability to perform reproducible research

- Proven ability to apply basic data mining algorithms (e.g. regression, logistic regression) and interpret the output

- Proven ability to communicate findings of analyzed data in a coherent and understandable way

Course Format

The lectures will be formatted as readings and video tutorials; these will be available on DataCamp. You will be responsible for going over these class materials on your own, prior to our class meetings. The R files presentations/notes will also be available on course webpage prior to (or after) each class, when necessary.

During class meetings, there will be very little lecturing (no more than 45 minutes). Instead, you will tackle different problems individually or in groups under my supervision using R and other software. Although the class's main language is R (heavily recommended); Python (and perhaps Julia) is also welcomed.

You will be submitting your work on Progress Journals; GitHub Pages micro websites displaying all your work. All your work including your data should be open to public. Otherwise it will not be graded. (see examples on https://mef-bda503.github.io/)

Final Exams: The exam will be 120-180 minutes, it will be primarily problem solving with written and R components. It will require extensive use of R, and it will be "open notes". You may not share books, materials, computers or electronic devices during an exam. You may not use any electronic communication tools such as email, IM, cell phones, pagers, etc. (May change to take-home final)

Group Project: There will be one group project. For this project, all group members will receive marks respective to their contribution. In general, students may freely communicate within their group and between groups. Collaboration is endorsed. Some group projects can be united under a major project. **Data theme of this year is Turkey (details will be on course webpage).**

Students are responsible for forming and managing their groups. It is expected that students will manage their groups so that everyone performs a fair share of the work, and that all perspectives are heard and considered.

Grading

Final course grades will be based on:

Attendance, Participation and Assignments	35%
Group Project	35%

Final Examination	35%

No late assignments will be accepted.

TENTATIVE SCHEDULE

Week 1

Introduction to Exploratory Data Analysis (EDA)
Basics of Base R

Week 2

Introduction to tidyverse (dplyr)
Introduction to RMarkdown

Week 3

Tidyverse cont'd (ggplot)
Functions

Week 4

dplyr joins
Shiny

Week 5

Package development
Cloud deployment

Week 6

Machine Learning
Project Presentations