MAT4376E Topics in Statistics
MAT5314E Topics in Probability and Statistics
Techniques of Data Analysis | Fall 2020 | Instructor: P. Boily

In October 2012, the *Harvard Business Review* published an article calling data science the **sexiest job of the 21st century**, a long cry from the business-as-usual practice of data geeks playing – at best – a supporting role in organizations. Today's data scientists are not just number-crunchers – as a combination of **data hacker**, **analyst**, **communicator**, and **trusted adviser**, they discover meaningful relationships in ever-growing masses of information, and play a leading role in the decision-making processes.

In groups or individually, you will apply specific data analysis tasks and investigate real-world datasets through the "Multiple I's" framework.

PRE-REQUISITES

Programming proficiency (R/Python/Matlab/ etc.); MAT2122, MAT2141, MAT2371, MAT2375 or MAT2377, and MAT3375 (or permission).

COURSE SCHEDULE

LEC: MON 08:30-10:00 (ZOOM)
LEC: WED 13:00-14:30 (ZOOM)
OFF: MON 10:00-11:30, THU 10:00-11:00 (ZOOM)

DELIVERABLES

Essay: Ethics in Data Science, 30-Sep (10%)

Projects: (90%)
1) Data Visualization, 15-Oct
2) Bayesian Data Analysis, 31-Oct
3) Queueing Systems, 15-Nov
4) Feature Selection & Data Reduction, 30-Nov
5) Anomaly Detection & Outlier Analysis, 21-Dec

MAT 4376E: all projects must be completed; only the best 4 will count towards your grade.
MAT 5314E: all projects have to be completed; all projects count towards your grade.

NOTES

Initiative and **independence** are rewarded: projects which stay solely within the suggested guidelines can score at most an 8/10 [A−].

Multiple I's Framework: intuition, initiative, innovation, interpretability, insights, integrity, independence, interactions.

SELECTED REFERENCES (extract)

Data Science Report Series
- A ggplot2 Primer
- A Soft Introduction to Bayesian Analysis
- Essentials of Queueing Methods
- Feature Selection and Data Reduction
- Anomaly Detection and Outlier Analysis

EXPECTATIONS

Data Science is not a spectator sport – I expect that you will spend on average 8-10 hours each week on this course.

Team work is crucial to insightful data analysis. You are encouraged to work in teams of 2 or 3, but it is not mandatory for you to do so. Be advised that the grade will be given to the whole group (independently of the quantity and quality of the work performed by each person).

You are expected to work on projects concurrently, and may have to use methods or concepts that have not been discussed in the lectures. More details will be provided in class.