MAT4376E Topics in Statistics
Techniques of Data Analysis
Summer 2023 | 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 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 this **online course**, you and your **teammates** will apply specific data analysis techniques 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).

ONLINE COURSE SCHEDULE

LEC: TUE 11:30-13:00 (**Zoom**)
LEC: THU 11:30-13:00 (**Zoom**)
OFF: by appointment (**Zoom**, **Slack**)

DELIVERABLES

Projects:
1) Data Visualization, 28-May
2) Bayesian Data Analysis, 18-Jun
3) Queueing Systems, 09-Jul
4) Anomaly Detection & Outlier Analysis, 31-Jul

All projects are weighed equally.

NOTES

Initiative and **independence** (& other "multiple I's") are rewarded: projects which stay solely within the suggested guidelines can score at most a B+.

Engagement is mandatory in this course. Students who do not attend and participate in at least 18 of the 24 of the sessions will be docked 10 marks at the end of term.

You may need to use methods or concepts that have not been discussed in the lectures. More details will be provided in class.

EXPECTATIONS

You are expected to spend 8-10 hours per week on this course.

Teamwork is crucial to insightful data analysis; in this course, all projects are done in teams of 3 or 4 (which can change from one project to another). Be advised that a grade is given to the whole group (independently of the quantity and quality of the work performed by each person).

One objective is to learn to navigate tight deadlines, and to plan your analysis/reporting accordingly (**10 page limit, no exception**).

Do not wait until the last minute before starting work on your projects.

There may be times when you are unable to deliver the projects by the deadline due to reasons outside your control. You are requested to inform me (and to submit the work you have already completed) as soon as you become aware of such a situation (within reason) so that we can discuss alternatives.