Introduction: Fundamentals of Data Analysis

ian.mcloughlin@atu.ie

Winter 22/23

An introduction to the fundamentals of data analysis, including the acquisition, cleaning, and exploration of data sets.

Learning Outcomes

On completion of this module the learner will/should be able to:

1. Source and investigate sets of data.
2. Programmatically explore and visualize data.
3. Apply basic mathematical data analysis techniques to data sets.
4. Write programs to automate basic data analysis techniques.

Assessment

Assessment is in the form of a portfolio[1]. There is no final exam.

[1] The portfolio will be in the form of a GitHub repository demonstrating your work throughout the semester.

20% Presentation of portfolio

40% Theory element of portfolio

40% Practical element of portfolio

Delivery

- This is a semester long module. Realistically, we will have ten uninterrupted teaching weeks[2].
- There are many ideas about how lecturers should deliver modules. Some suggest a top-down, structured approach where topics are clearly defined ahead of time. Others suggest involving students in decisions, letting content evolve during the semester. Let's not be idealistic about it: we'll have an initial plan and tailor it during the semester.
- It is worth discussing in the Moodle forums what you as a class would like to work on. Just keep in mind that everyone will want something different. Also, remember that there is one lecturer and dozens of students in each of several modules. Time is limited, we will have to be careful about scope creep.

[2] Each semester typically has thirteen teaching weeks but some of those weeks have public holidays and other interruptions.

Lectures and Practicals

- Traditionally, lectures covered theory and students applied the theory in practicals. That can feel a bit contrived and artificial, especially in computing where practice often comes before theory.

- We won't make a clear distinction between lectures and practicals where possible. Rather, we will focus on topics, projects, and problems.
- Notes will be in the form of Jupyter notebooks and practical work will be indicated in those notebooks.

Topics

We will start with a plan to cover these five topics.

Randomness : Data and stochasticism.

Information: Shannon entropy.

Bias: Cognitive and statistical.

Outliers: Identifying outliers in data.

Cleansing: Fixing anomalies and inconsistencies.

Advice

- Everyone procrastinates, you need a strategy to compensate. You will be less stressed if work regularly, a bit every week.
- Review the marking scheme regularly and work to it.
- Be able to demonstrate your work. This is easier for practical work, you often have code and the like. Theoretical work can be demonstrated through writing, images, plots, and diagrams.
- You will have to grapple with the uncertainty of making your own content and design decisions. That can be difficult.

Policies

- In April 2022, GMIT merged with IT Sligo and LyIT to become ATU, the Atlantic Technological University.
- Although the merger has happened, it will take a couple of years for our systems and policies to fully merge.
- During this time, we will continue to use GMIT's policies where an ATU policy has not yet superseded them.
- That means the GMIT Quality Assurance Framework [3].

GMIT is now ATU.

[3] GMIT. Quality Assurance Framework. https://www.gmit.ie/general/quality-assurance-framework