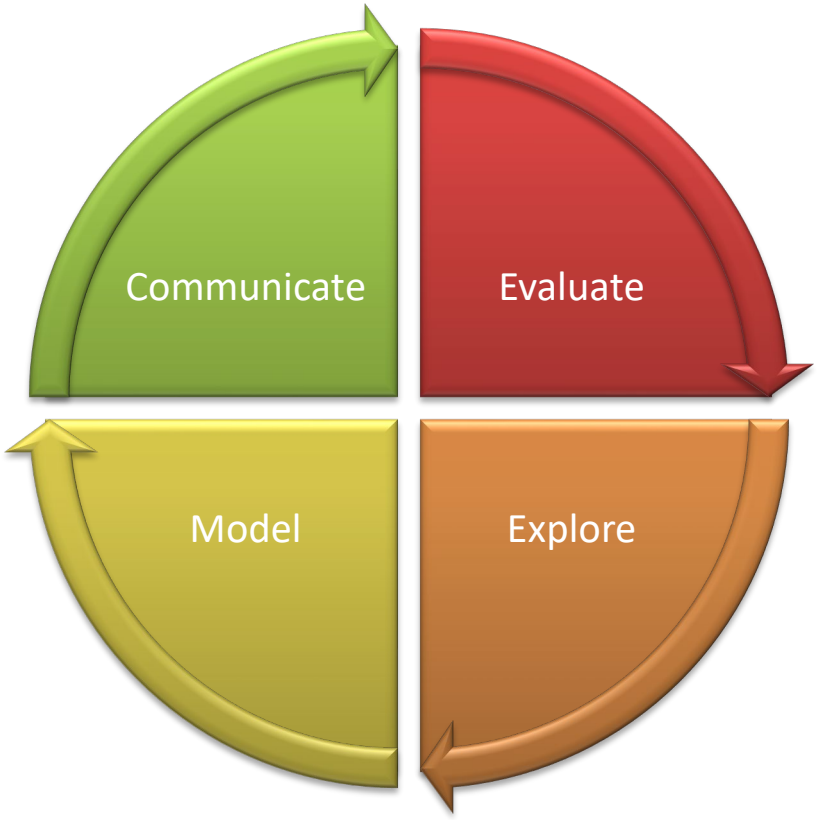
MATH9102: Fundamentals of Data Analysis

Lecturer: Dr. Deirdre Lawless
5 ECTS
100% Continuous Assessment

What is the module about?



- Equip learners with core principles and tools to analyse complex data
- Critically evaluate data quality and suitability
- Use numerical, graphical, and computational techniques
- Build and evaluate predictive models to uncover patterns and trends
- Present clear, evidence-based insights

How will I learn?

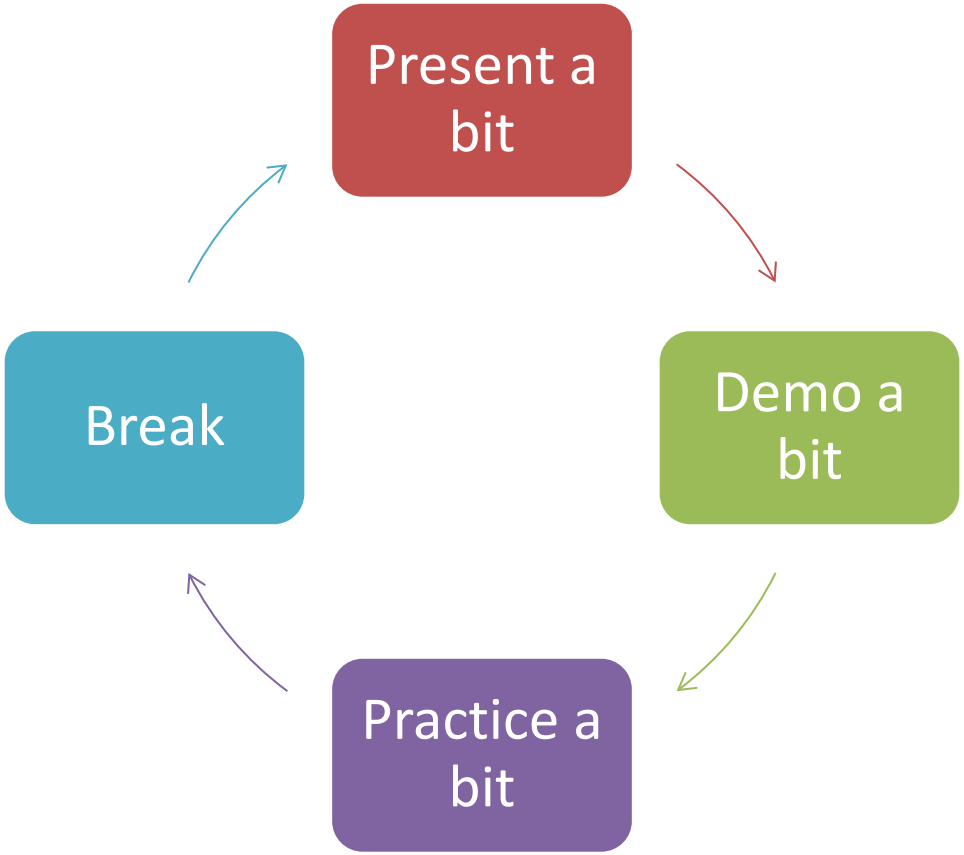
Interactive *lectorials* (lecture + demo +practice)

Hands-on R programming

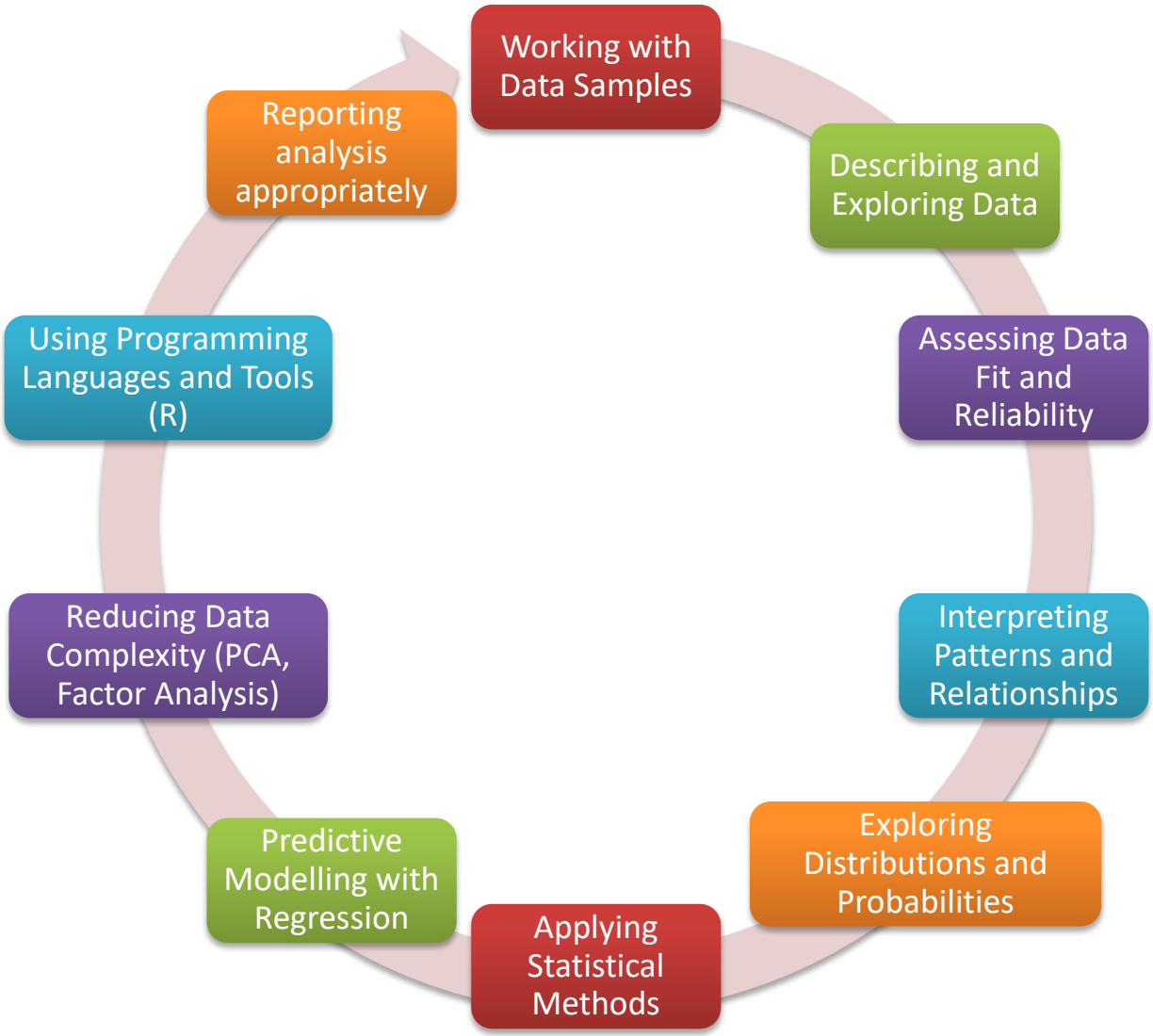
Real-world data used during examples

Assessment through authentic project work

How class sessions will run

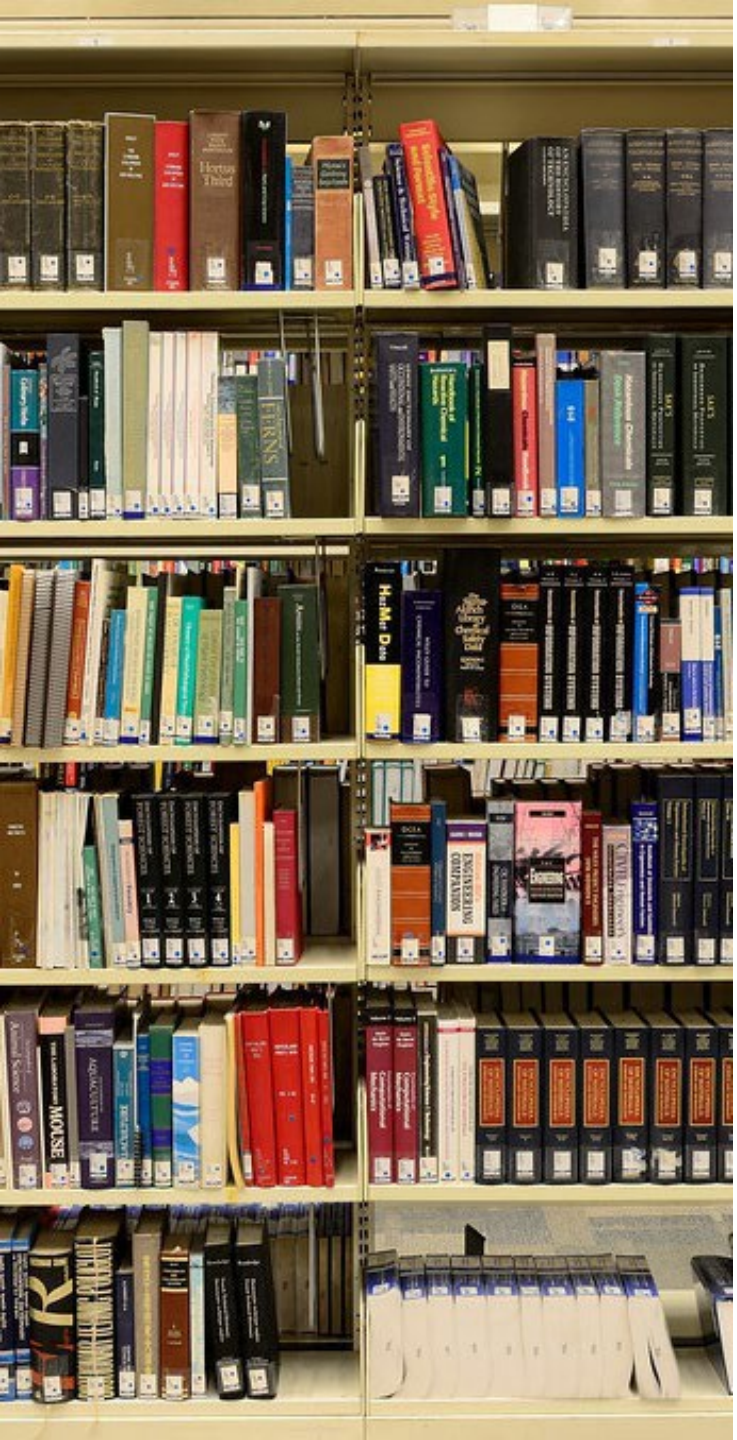


Module Content



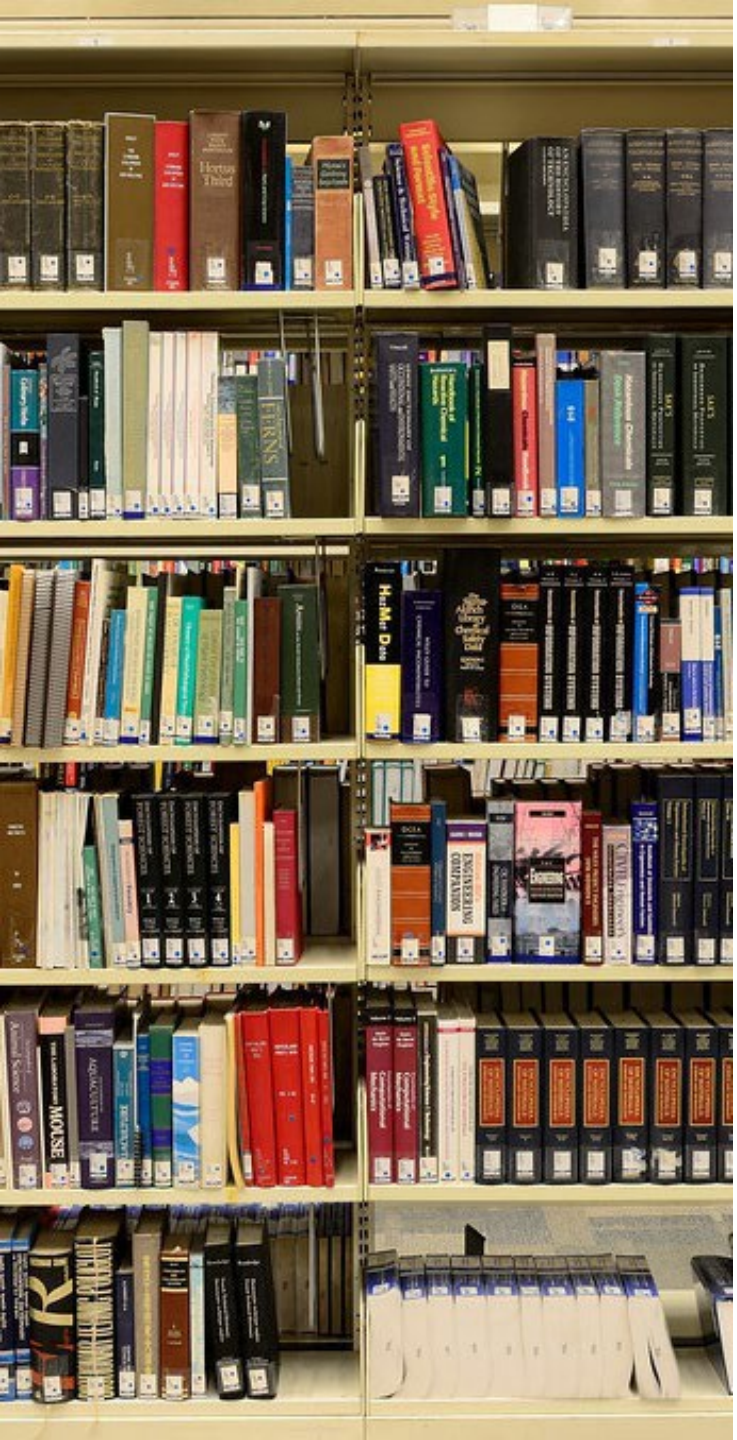
Module Learning Outcomes

1. Critically evaluate fundamental data analysis concepts.
2. Apply advanced numerical and graphical techniques to explore and summarize datasets.
3. Assess the suitability of datasets for use in data science workflows and justify methodological choices based on these assessments.
4. Select and apply appropriate methods for exploring data and identifying meaningful patterns or relationships.
5. Interpret and synthesize analytical outcomes to extract meaningful insights and support evidence-informed decision-making.
6. Design, implement, and critically assess models for predicting outcomes based on data patterns.
7. Apply and evaluate data reduction techniques to enhance interpretability and modelling performance.
8. Communicate insights clearly and effectively through well-structured reports that demonstrate analytical depth, clarity of interpretation, and actionable recommendations.
9. Demonstrate proficiency in using appropriate programming tools and languages to develop, implement, and communicate end-to-end data science workflows, including data preparation, analysis, modelling, and results presentation.



Locating Module Material

While waiting for Brightspace to be setup, material can be located at this link: https://tinyurl.com/FundDA-2025



Locating Module Material

Within Brightspace module:

- Material for Lectures/Labs
  - Organised by week number
  - Lecture notes (pdf) (generally available in advance of class)
  - Lab exercise and solutions (solutions to exercises will be published after the classes are completed)
  - Generally, will be available in advance of classes
  - Please Note:
  - Presentations used during class sessions are intended **to be a supplement to attending class not a replacement**
- Continuous Assessment
  - Links to specs, submission boxes, rubrics etc
- Datasets Used
- Useful Resources

While waiting for Brightspace to be setup, material can be located at this link: https://tinyurl.com/FundDA-2025

Practical Component

nners

rting point will serve all beginners, but here are 6 ways to begin learning R.

l , **RStudio, and R packages like the tidyverse.** These three installation steps are often

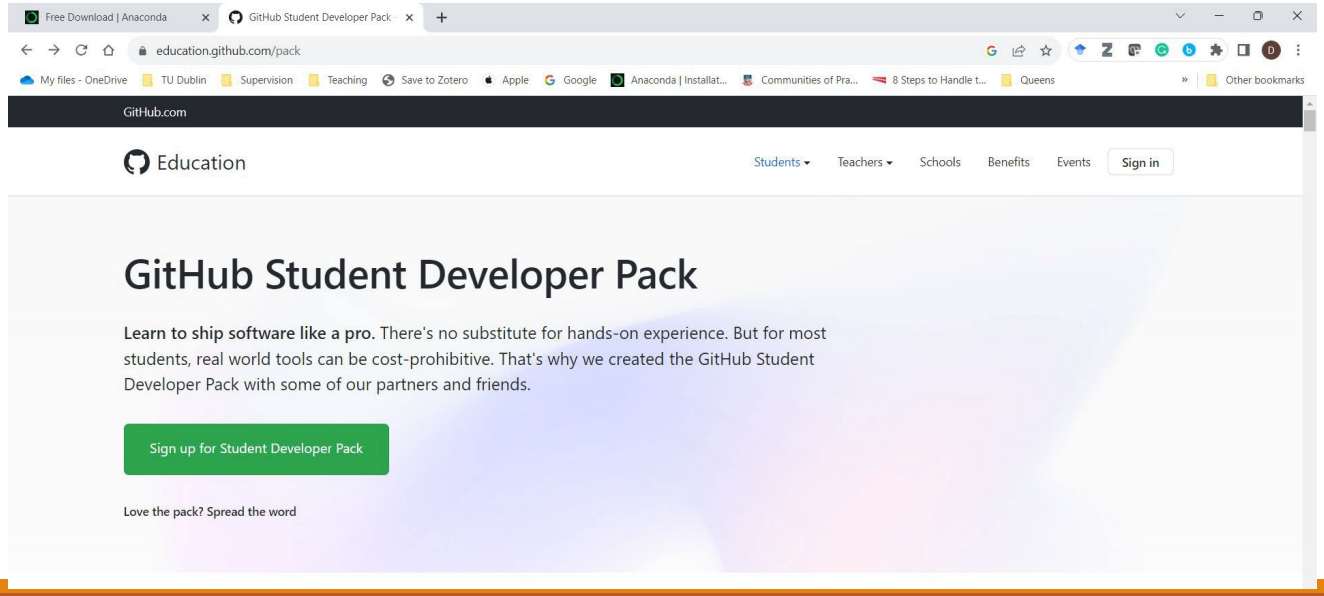
R and Rstudio will be used during classes

Getting started with R and RStudio: https://education.rstudio.com/learn/beginner/

Practical Component

Get yourself set up with a GitHub Student Developer Pack - https://education.github.com/pack

- Lots of free tools
- Lots of free resources
- Extra time for using Codespaces for free



Code Provided to Support Learning

This is a module in how to conduct a data analysis

- We will be using R as a tool

This is not a module in how to use R

- That should be covered in Working with Data

This is not a module about visualisation

- Fancy graphs are covered in Data Visualisation

Code Provided to Support Learning

You will learn how to:

- Design an analysis
- Describe your data (using appropriate visuals)
- Conduct relevant statistical tests
- Build simple predictive models
- Evaluate the findings in the context of your analysis
- Express your findings appropriately

You will be provided with R code snippets which implement some of the above

Datasets

A range of datasets will be used throughout the module to illustrate concepts

You will be provided with these datasets or details of where to find and download them as needed

Some Books

- Peter Bruce,Peter C. Bruce,Andrew Bruce,Peter Gedeck. (2020), Practical Statistics for Data Scientists, 2nd. O'Reilly Media, p.0, [ISBN: 978-1492072942].
- https://datapot.vn/wp-content/uploads/2023/12/datapot.vn-Practical-Statistics-for-Data-Scientists.pdf?srsltid=AfmBOopXEOEs6uamO_ogfXsNiiAurw8LY3A_uV-TQHVRL4hORSwvVTiC

- Hadley Wickham,Mine Çetinkaya-Rundel,Garrett Grolemund. (2023), R for Data Science, 2nd. O'Reilly Media, p.0, [ISBN: 978-1492097402].
- https://digitallibrary.tsu.ge/book/2019/september/books/R-for-Data-Science.pdf

- Mike McGrath. (2023), R for Data Analysis in Easy Steps, 5th. In Easy Steps, p.0, [ISBN: 978-1840789980].
- Multiple copies in the Grangegorman library (Park House)

Assessment (100% CA)

Phase ONE (40% of module marks):

- You will choose a dataset
  - You will apply your learning from the module to construct an appropriate statistical description of selected concepts in this dataset and construct an appropriate report.
    - There will be requirements w.r.t the types of statistical variable you should use.
  - You will then conduct an initial statistical analysis of a selected concepts represented in this dataset and construct an appropriate report.
    - There will be requirements w.r.t the types of statistical tests you should use.
    - Put together a plan for phase TWO.
  - You will be required to submit your report, your R code, your dataset.

Note: You can and are recommended to use the same dataset for the Working with Data and Data Visualization modules.

Assessment (100% CA)

Phase TWO (60% of module marks):

- You will build, analyse and report on:
  - A linear regression model (using the same dataset as phase one)
  - A logistic regression model (using the same dataset as phase one)
  - Build a second linear/logistic model and compare to the model previously built (using the same dataset as phase one)
  - Dimension reduction – using a second dataset provided to you.

How to succeed in the module

Focus on learning the process of conducting a data analysis:

- What are you trying to discover or show?
  - Figure out a question you are trying to answer/theory you are trying to test
- What data do you need to collect?
- Once you have data, how do you describe the data you have?
  - You need to explain this to whoever will be the consumer of your work
- What analysis should you conduct?
  - You need to know the types of statistical tests and models you need and how to explain the outcomes to your consumer
- How do you interpret your analysis?
  - You need to know how to interpret the outcomes of the analysis and present these to your consumer
- How will you present your findings?

How to succeed in the module

Work steadily through the material

- Keep up
- Make use of the lab time allocated to work on CA

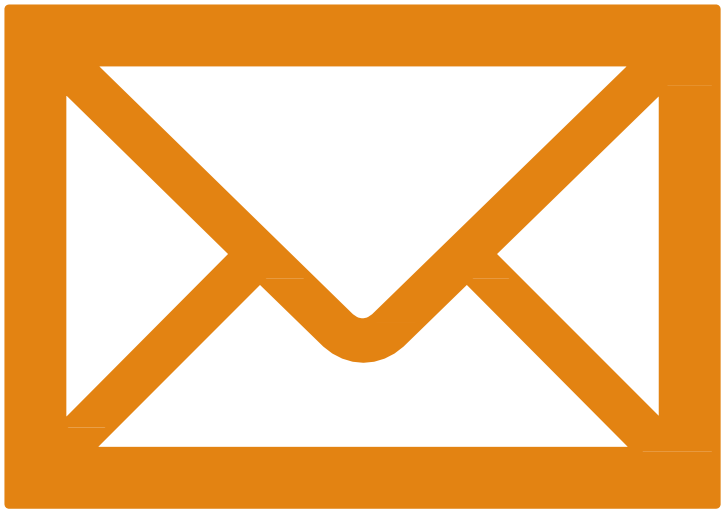
Make your own notes

- On the topics we cover in class
- On how to use the software
  - Make comments in your scripts/output and save it somewhere

Don’t be afraid to ask questions

- Of me, of each other, on the web…

Keep going…



Contact

- **In person:**
  - During class.
- **Email:**
  - deirdre.lawless@tudublin.ie