MSc in Data and Computational Science – Introduction

Dr Lennon Ó Náraigh

University College Dublin

School of Mathematics and Statistics

6th September 2024

What is it all about?

- Joint master's degree in Statistics and Applied and Computational Mathematics
- Aims to give graduates sought-after skills in Data Analytics industries
- Also, aims to give graduates skills that can be used in academic research (e.g. Ph.D.), either in Statistics or Applied and Computational Mathematics

Twin-track nature of degree (academic/industry) exemplified by split in third trimester – Stream 1 ('traditional' dissertation under supervision of an academic), and Stream 2 (mini project, taught modules).

It is not computer science...

The programme is heavily focused on **computational science** (i.e. algorithms, scientific computing, parallel computing for simulation, and mathematical modelling) and statistics (both theoretical and computational).

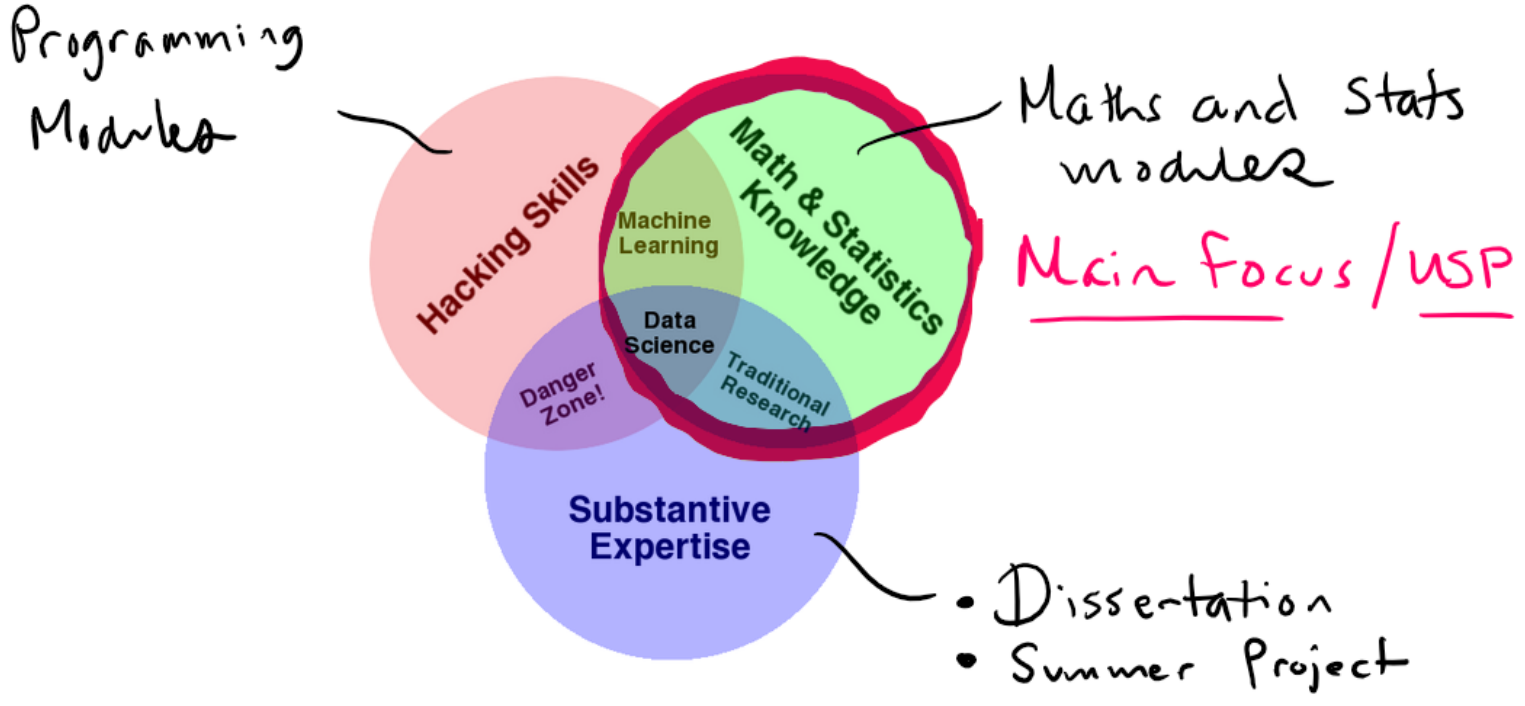
It is not computer science...

The programme is heavily focused on **computational science** (i.e. algorithms, scientific computing, parallel computing for simulation, and mathematical modelling) and statistics (both theoretical and computational).

... but you will learn a lot of programming:

- C, R, Matlab, Python
- Option modules in High-Performance Computing (MPI) with the Irish Centre for High-End Computing (ICHEC)

Rather, the programme will equip you with the skills at the heart of Data Science



http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram

Why?

> “8 out of 10 people working in data science think data science is all about Machine Learning... With the advent of Automated Machine Learning and Deep learning... the whole machine learning has just come down to one big piece of automated code. Which means many data scientists would become redundant and lose jobs in near future whose main everyday task is building machine learning models.”

`https://www.linkedin.com/pulse/` (July 17th 2017)

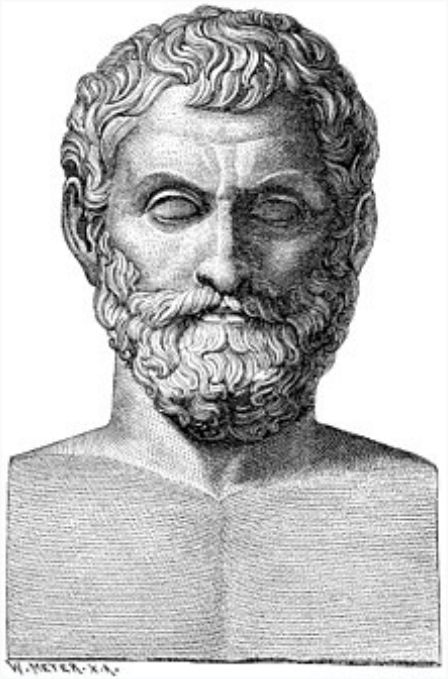
Why?

> So what should one learn? Learn things which help you understand 'Cause' and 'Effect' relationship between things... Learn the hard stuff – learn **Statistics especially Bayesian**... **Learn anything which gives you the ability to link an effect and the probable reason causing it** to help you take good decisions.

This degree will teach you these things!

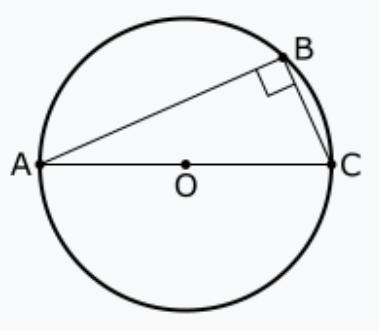
The Ideal Graduate...

Thales of Miletus



Posthumous portrait of Thales by Wilhelm Meyer, based on a bust from the 4th century

Born	c. 626/623 BC Miletus, Ionian League (modern-day Balat, Didim, Aydın, Turkey)
Died	c. 548/545 BC (aged c. 78)
Era	Pre-Socratic philosophy
Region	Western philosophy



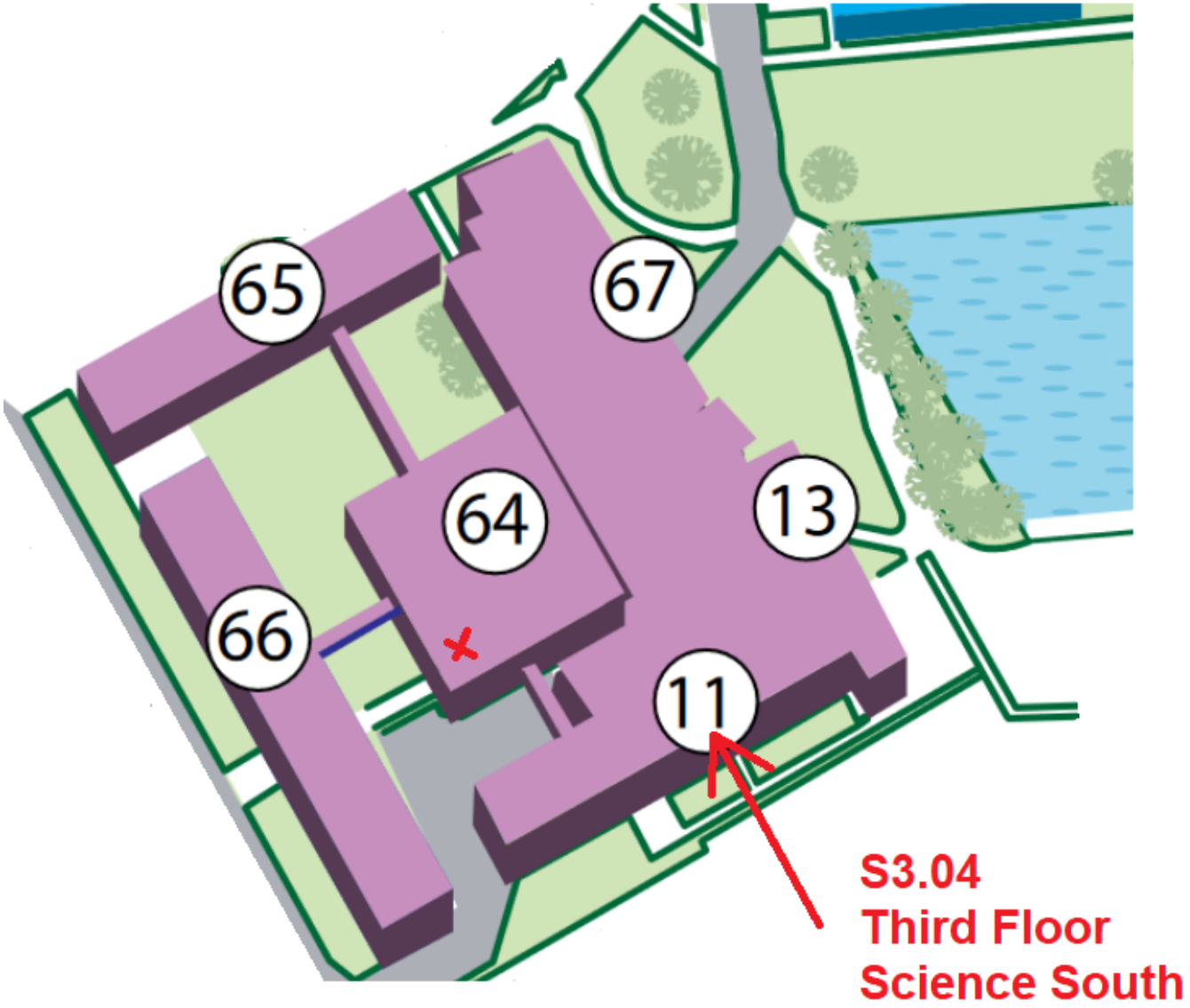
Thales' theorem: if $\overline{AC}$ is a diameter and B is a point on the diameter's circle, the angle $\angle ABC$ is a right angle.

Thales also:

- Predicted a good harvest;
- Secured an option on the use of olive presses for that year;
- Rented out the olive presses to other people at harvest time.

Used **data** and **mathematics** to make a business insight (arbitrage). First Data Scientist?

For these reasons, the Programme is based in the School of Mathematics and Statistics.



Structures: 90 Credits, Including 45 Credits of Core Modules

Taught MSc: 90 Credits (ECTS). Includes 45 Credits of **Core Modules**:

Core Modules

				Credit Allocation			
Trimester	**Module ID**	**Title**	**Credits**	**Autumn**	**Spring**	**Summer**	**Inactive Module?**
Autumn	MATH40550	Applied Matrix Theory	5.00				
Autumn	STAT20230	Modern Regression Analysis	5.00				
Autumn	STAT30340	Data Programming with R	5.00				
Autumn	STAT40800	Data Prog with Python (online)	5.00				
Autumn	STAT41040	Principles of Prob & Stats	5.00				
Spring	ACM40990	Optimisation in ML	5.00				
Spring	ACM41000	Uncertainty Quantification	5.00				
Spring	STAT40150	Multivariate Analysis	5.00				
Spring	STAT40850	Bayesian Analysis (online)	5.00				

All students must take 15 Credits of Option Modules in Autumn / Spring

All students must take 15 Credits of **Option Modules** from the following list:

Option Modules

				Credit Allocation			
Trimester	Module ID	Title	Credits	Autumn	Spring	Summer	Inactive Module?
A)3OF: (Option modules) Students must take 15 credits							
Autumn	ACM40290	Numerical Algorithms	5.00				
Autumn	ACM40660	Scientific Programming (ICHEC)	5.00				
Autumn	STAT40400	Monte Carlo Inference	5.00				
Spring	ACM40640	High Performance Comp. (ICHEC)	5.00				
Spring	STAT30250	Advanced Predictive Analytics	5.00				
Spring	STAT30270	Statistical Machine Lrng	5.00				
Spring	STAT40970	Machine Learning & AI (online)	5.00				

Some Tips

- Aim for balanced trimesters where possible (6+6 is the ideal)
- Option modules from ICHEC (Irish Centre for High-End Computing) were very popular last year - **more practical**.
- If taking STAT 40970 (Machine Learning and AI (online)) it is recommended to take STAT 30270 (Statistical Machine Learning) concurrently - **more theoretical**.
- STAT 40400 (Monte Carlo Inference) is another module which is **more theoretical**.
- Finding it hard to narrow down options? Can **audit** a maximum of one module per trimester (with programme director signoff).

Warning

- Your registration is your responsibility.
- Make sure you remain compliant – 90 credits total.

Remaining 30 Credits: Summer Programme

Two Streams:

- Stream 1 – 30-credit full-time dissertation.
  - Small groups
  - Academic supervisor
  - Dissertation at the end.
- Stream 2 – 30 credits of project work and taught modules:
  - 'Mini-project' (ACM 40960, 15 ECTS)
  - Flexible timetable
  - Choice of online taught modules (15 ECTS)

Summer Programme – Continued

B)1OF: (Stream 1 Core) Dissertation **Students complete a dissertation under academic supervision**							
Summer	ACM40910	Research Project II	30.00				
C)1OF: (Stream 2 core) **Students on Stream 2 must complete this core module** 'Mini Project' - Research Experience							
Summer	ACM40960	Projects in Maths Modelling	15.00				
D)3OF: (Stream 2 options) Wide range of online modules **Students on Stream 2 must take 15 credits from this option list**							
Summer	STAT40780	Data Prog with C (online)	5.00				
Summer	STAT40810	Stochastic Models (online)	5.00				
Summer	STAT40830	Adv Data Prog with R (online)	5.00				
Summer	STAT40840	Data Prog with SAS (online)	5.00				
Summer	STAT40950	Adv Bayesian Analysis (online)	5.00				
Summer	STAT40960	Stat Network Analysis (online)	5.00				

Summer Programme – Stream 1 Dissertation

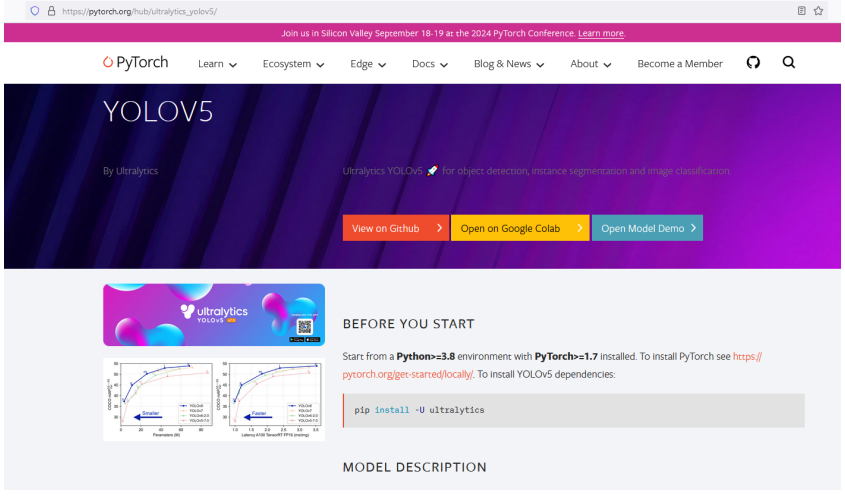
- By default, all students are registered to Stream 2 (taught modules / mini-projects).
- During the autumn trimester, interested students will be invited to apply for a dissertation under the supervision of a staff member in lieu of the taught modules.
- Dissertation students to work full-time on their project for the duration of the summer trimester.
- Available projects based on supervisors, the number of positions in Stream 1 may therefore be limited (application process).
- Applications will be reviewed by a committee consisting of university faculty, and will be assessed on the basis of first-trimester and the student's suitability for one of the available research projects.
- Application results early in spring trimester, dissertation students will have their registration changed at that point.

Summer Programme – Stream 2

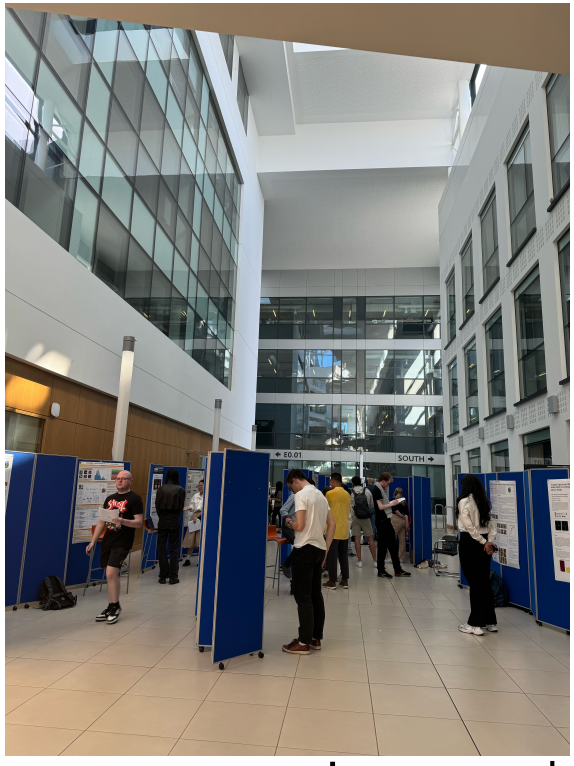
- Industry-relevant taught modules (online)
- Industry-relevant project experience, assessing group work, written communication skills, etc.
- Opportunity to talk about 'Master's project' in interviews
- **We will have a 're-orientation session' in January where we can talk about summer choices again.**

ACM 40960 Projects in Mathematical Modelling

Finding ‘Waldo’ using a Convolutional Neural Network



The YOLOv5 package



Poster presentation at end of module

Teaching and Learning

- Lectures are primarily face-to-face.
- Discuss problems / issues early with module coordinator first and programme director later.

Module substitution

- You might have taken the equivalent of one or two of the core modules before – discuss with academic directors to find substitutes.
- Students are invited to look in the first instance at option modules not already selected.
- After that, students should look to other appropriate stage-4 modules in the School of Mathematics and Statistics.
- Liaise with Programme Director (me):

  `onaraigh@maths.ucd.ie`

- Final registration change to be carried out by programme administrator Natalia Zadorozhnyaya (`dataandcomp@ucd.ie`).
- Natalia can also help with other technical registration issues.

Contacts Again

- Programme director: me (onaraigh@maths.ucd.ie).
- Deputy director: Dr Wagner Barreto-Souza (wagner.barreto-souza@ucd.ie)
- Natalia Zadorozhnyaya (school office, Science South) is the Taught Graduate Administrator and will help with registration issues (dataandcomp@ucd.ie).
- These slides, and a detailed information document, are available on my personal website:

  https://maths.ucd.ie/ ∼ onaraigh/data_comp_sci.html