

# Week 1: Introduction

POP77001 Computer Programming for Social Scientists

Tom Paskhalis

12 September 2022

Module website: [bit.ly/POP77001](https://bit.ly/POP77001)

# Overview

- Module objectives
- Prerequisites and software
- Materials and books
- Module meetings
- Assessment and collaboration
- Weekly schedule



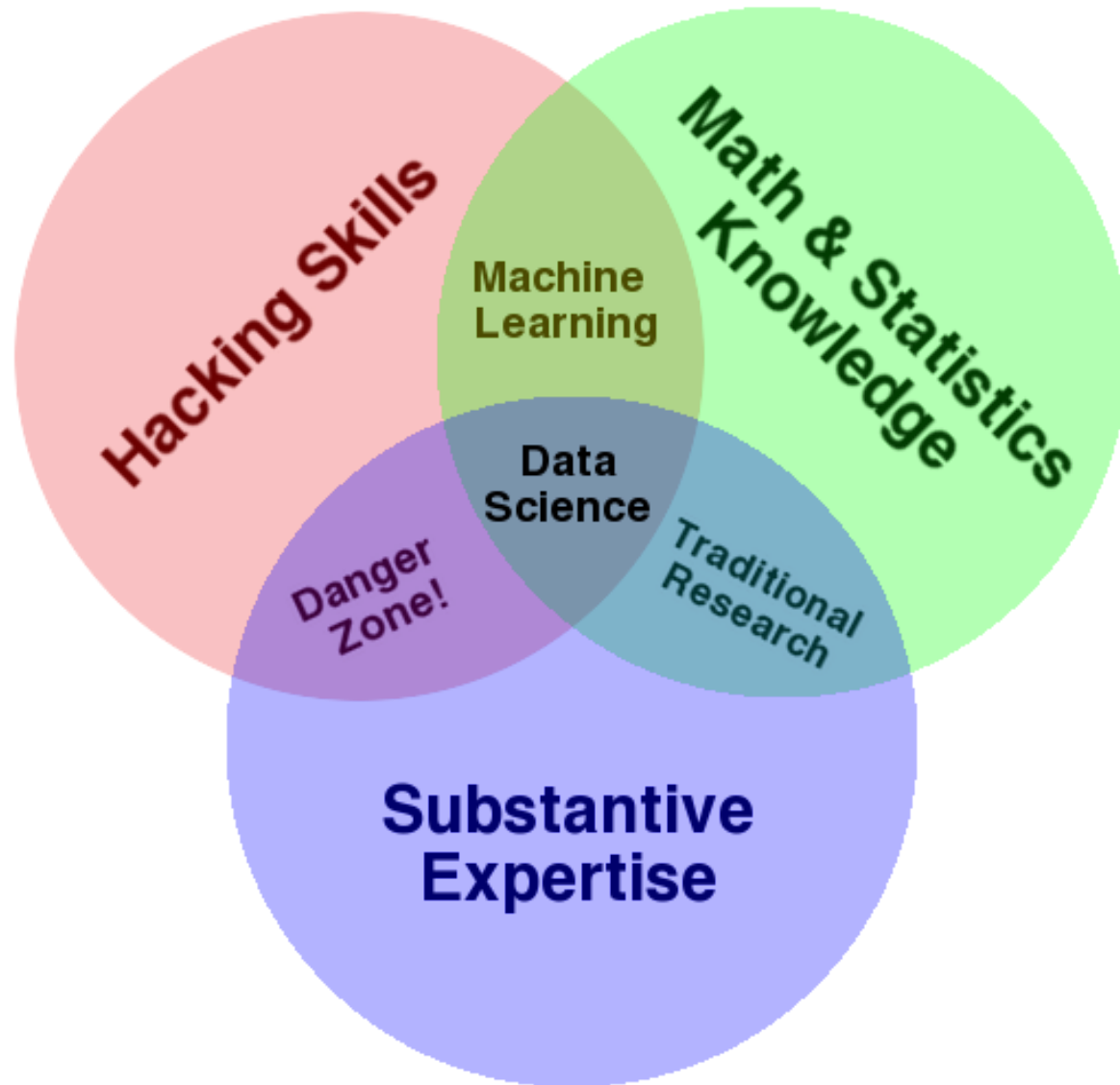
# Data Scientist: The Sexiest Job of the 21st Century

Meet the people who can coax treasure out of messy, unstructured data. by Thomas H. Davenport and D.J. Patil

From the Magazine (October 2012)



Source: [Harvard Business Review](#)



Source: [Drew Conway](#)

# AS SEEN BY USERS OF ...



Source: [Reddit](#)

# About me

- Assistant Professor in Political Science and Data Science, [Trinity College Dublin](#)
  - Before: Postdoctoral Fellow, [New York University](#)
  - PhD in Social Research Methods, [London School of Economics and Political Science](#)
- My research:
  - Political communication, social media, interest groups
  - Text analysis, machine learning, record linkage, data visualization
- Contact
  - [tom.paskhalis@tcd.ie](mailto:tom.paskhalis@tcd.ie)
  - [tom.paskhal.is](http://tom.paskhal.is)
  - [@tpaskhalis](#)

# Module Objectives

- Introduce the fundamentals of computer programming;
- Get familiar with R and Python programming languages;
- Develop understanding of core software design principles;
- Learn crucial data science techniques;
- Practice these concepts using social science examples.



# Module Materials

- Course website: [bit.ly/POP77001](https://bit.ly/POP77001)
- GitHub repository: [github.com/ASDS-TCD/POP77001\\_Computer\\_Programming\\_2022](https://github.com/ASDS-TCD/POP77001_Computer_Programming_2022)

# Books

- Guttag, John. 2021 *Introduction to Computation and Programming Using Python: With Application to Computational Modeling and Understanding Data*. 3rd ed. Cambridge, MA: The MIT Press
- Matloff, Norman. 2011. *The Art of R Programming: A Tour of Statistical Software Design*. San Francisco, CA: No Starch Press.
- McKinney, Wes. 2017. *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. 2nd ed. Sebastopol, CA: O'Reilly Media
- Roger D. Peng. 2016. [\*R Programming for Data Science\*](#). Leanpub.
- Wickham, Hadley, and Garrett Grolemund. 2017. [\*R for Data Science: Import, Tidy, Transform, Visualize, and Model Data\*](#). Sebastopol, CA: O'Reilly Media.
- Wickham, Hadley. 2019. [\*Advanced R\*](#). 2nd ed. Boca Raton, FL: Chapman and Hall/CRC.

# Additional Online Materials

- [Git Book](#)
- [The Hitchhiker's Guide to Python](#)
- [Python For You and Me](#)
- [Python Wikibook](#)
- [Python 3 Documentation](#)
- [R Documentation](#)
- [R Inferno](#)

# Prerequisites and Software

- Introductory module - no formal prerequisites
- Laptop with Windows/Mac/Linux OS (no Chrome books)
- Required software:
  - **Jupyter** - web-based interactive computational environment
  - **Python** (version 3+) - versatile programming language
  - **R** (version 4+) - statistical programming language
- Additional software:
  - **Git** - version control system
  - **GitHub** - git-based online platform for code hosting
  - **RStudio** - integrated development environment for R
  - **Spyder** - integrated development environment for Python
  - **Visual Studio Code** - feature-rich text editor






# Module Meetings

- 11 two-hour lectures
  - Monday 14:00 in PX 201 [7-9 Leinster Street South](#)
- 11 two-hour tutorials
  - Wednesday 14:00 in PX 201 [7-9 Leinster Street South](#)
- No lecture/tutorial in Week 7
- Office hours:
  - Thursday 11:00 - 13:00 [online](#)

# Assessment

- Participation (10 %)
  - Tutorial attendance
- 4 assignments (40%)
  - Bi-weekly programming exercises
  - Due by 23:59 on Friday of weeks 3, 5, 9 and 11 on Blackboard
- Final project (50%)
  - Final R/Python project demonstrating familiarity with programming concepts and ability to communicate results
  - Due by 23:59 on Friday, 16 December 2022

# Assessment criteria

1.  Code exists
2.  Code runs and does what it has to do
3.  Code is legible (meaningful naming, comments)
4.  Code is modular (no redundancies, use of abstractions)
5.  Code is optimized (no needless loops, runs fast)

Marks at Trinity: <https://www.tcd.ie/academicregistry/exams/student-guide/>

# Plagiarism

- Plagiarising computer code is as serious as plagiarising text (see [Google LLC v. Oracle America, Inc.](#))
- All submitted programming assignments and final project should be done individually;
- You may discuss general approaches to solutions with your peers;
- But do not share or view each others code;
- You can use online resources but give credit in the comments.



# Module Outline

Week	Language	Topic	Assignment Due
1	-	Introduction to Computation	
2	R	R Basics	
3	R	Control Flow in R	Assignment 1
4	R	Functions in R	
5	R	Debugging and Testing in R	Assignment 2
6	R	Data Wrangling in R	
7	-	-	
8	Python	Fundamentals of Python Programming I	
9	Python	Fundamentals of Python Programming II	Assignment 3
10	Python	Data Wrangling in Python	
11	Python, R	Complexity and Performance	Assignment 4
12	Python, R	Web Scraping	

# Next

- Introduction to Computation