Everyone should learn to code

Andrew (Ed) Edmondson

Alan Turing Institute

"I think it likely that in ten years' time every undergraduate programme will have to include some teaching in data science".

Professor Sir Adrian Smith FRS, Alan Turing Institute Director and Chief Executive, 2019

What is Data Science?

Data science is a “concept to unify statistics, data analysis, machine learning and their related methods” in order to “understand and analyze actual phenomena” with data.

Hayashi, Chikio (1 January 1998). “What is Data Science? Fundamental Concepts and a Heuristic Example”

What is Data Science?

"With this in mind, I would encourage you to think of data science not as a new domain of knowledge to learn, but a new set of skills that you can apply within your current area of expertise."

Preface to Python Data Science Handbook by Jake VanderPlas

What is Data Science?

Combining:

- Domain expertise
- Maths/stats
- Coding

. . . and it's not just for scientists, but all researchers.

What is Data Science?

Doing the job properly.

Coding Goals

- Understand what algorithms are, how they are implemented as programs on digital devices, and that programs execute by following precise and unambiguous instructions;
- Create and debug simple programs;
- Use logical reasoning to predict the behaviour of simple programs
- Use technology purposefully to create, organise, store, manipulate and retrieve digital content

Coding is not a specialist skill

- Understand what algorithms are, how they are implemented as programs on digital devices, and that programs execute by following precise and unambiguous instructions;
- Create and debug simple programs;
- Use logical reasoning to predict the behaviour of simple programs
- Use technology purposefully to create, organise, store, manipulate and retrieve digital content

National curriculum in computing: Key stage 1 (5-7 year olds).

Coding is for everyone!

- Gone are the days when coding was just for techies
- It's in the primary school curriculum
- => In 15 years' time most PhD students will have done it at primary school

Premise

- There will come a time when all new researchers will be able to code
- A researcher who codes will:
  - Do better research
  - Do quicker research
  - Do more reproducible research
  - Be able to supervise students who code
  - Publish more papers, books etc.
  - Progress faster in their career

Why does it matter?

- Store your data in Excel
- Manually edit it, make forumlae, make charts
- Include charts and results in paper

Reinhart, Rogoff... and Herndon: The student who caught out the profs

"Economists have been astonished to find that a famous academic paper often used to make the case for austerity cuts contains major errors." https://www.bbc.co.uk/news/magazine-22223190

If you did it in Python, R, MATLAB, . . .

- Code / data separation (=> don't edit the raw data)
- Use the code to test the data
- And tests to test the code
- Calculate results using code
- Generate figures using your code
- Get the right results

Conclusion

- Primary school => everyone **can** learn to code, *or at least learn what is possible*
- Data-based research done using code will be:
  - Better
  - Faster
- In the future all new researchers will be able to code (to some extent)
- Why not lead the way now?