Data Science in Medicine:
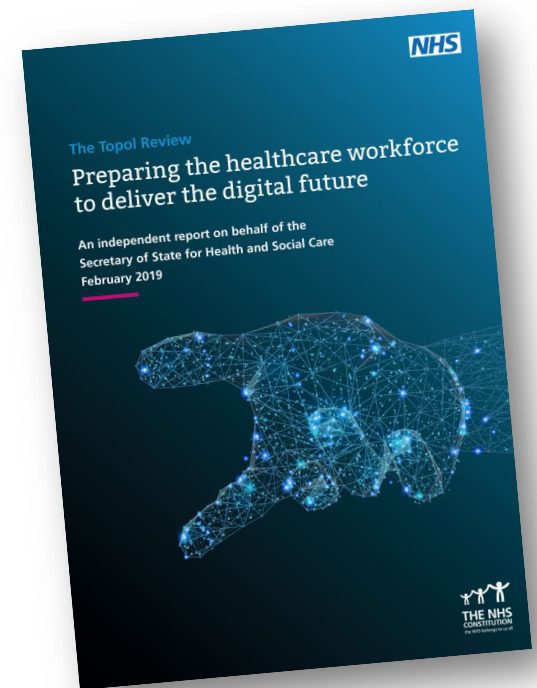
Lecture 1: Introduction

Dr Areti Manataki

Usher Institute

The University of Edinburgh

Why Data Science in Medicine

- Our aim: to equip you with the key foundations and data skills that are needed for the data-intensive medicine of the future

"Within 20 years, 90% of all jobs in the NHS will require some element of digital skills. Staff will need to be able to navigate a data-rich healthcare environment."

"Education providers should ensure genomics, data analytics and AI are prominent in undergraduate curricula for healthcare professionals."

The Topol Review

Preparing the healthcare workforce to deliver the digital future

An independent report on behalf of the Secretary of State for Health and Social Care February 2019

Why Data Science in Medicine

- Our aim: to equip you with the key foundations and data skills that are needed for the data-intensive medicine of the future

- In line with recommendations in the Topol Review and by the National Academy of Medicine

- You will need these skills during the intercalation year (Year 3) and in your SSC5a research project (Year 5).

Demystifying Health Data Science

Video available at: https://youtu.be/BojvZOzYeq4

Further info at https://www.hdruk.ac.uk/health-data-science/

Demystifying Health Data Science

- Health Data Science is a discipline that combines mathematics, computer science and health sciences to study different types of health problems using data.

- It provides the tools to manage and analyse very large amounts of different datasets across our healthcare systems.
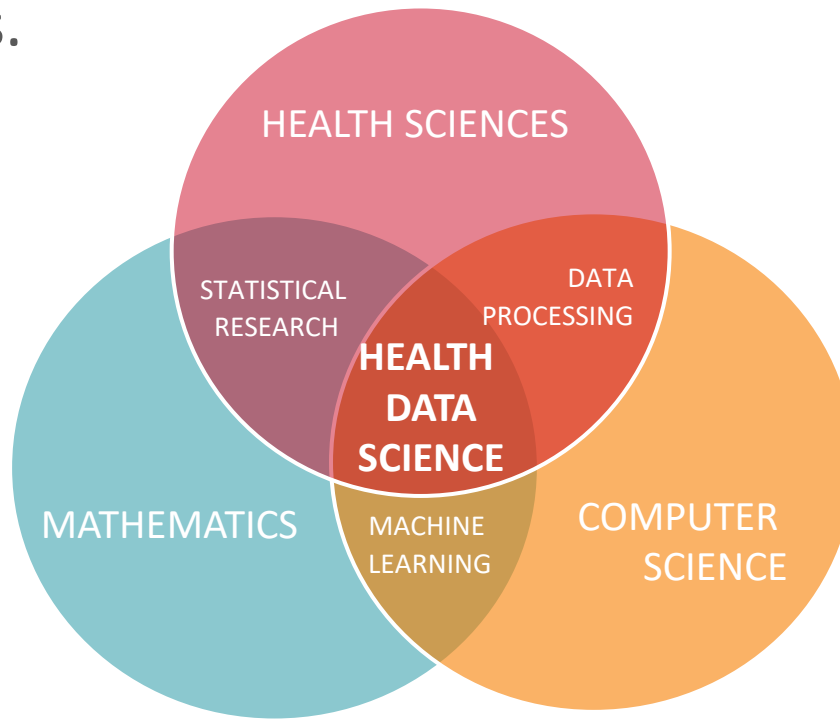
[Image adapted from Shelly, P. 2015. Data Science for the C-Suite]

Data is everywhere!

- Genomic data

- Electronic patient records

- Medical images

- Clinical notes

- Wearable devices

- Social media posts

- Bank card transactions

...and much much more

Medicine is now a data-intensive discipline

- **Healthcare Analytics**: diagnosis, treatment effectiveness, hospital management, etc.

- **Precision Medicine**: By integrating different types of data for every patient we can take a more personalised approach to therapies (tailoring them to suit each individual).

- **Stratified Healthcare**: The ability to characterise individuals much more precisely allows us to identify key differences across human populations and to act accordingly in healthcare provision.

Opportunities brought by Health Data Science

Data and its analysis is revolutionising how medicine is understood, how biomedical research is conducted and how healthcare is delivered.

- Better understand disease
- Earlier and improved diagnosis
- Prevention of disease
- Enhanced predictions (e.g. in epidemiology)
- Safer and more effective treatments
- More effective integrated care pathways
- Driving clinical research
- Precision medicine

#datasaveslives

- 100,000 Genomes Project:
  - Aimed at sequencing 100,000 genomes from around 85,000 NHS patients affected by a rare disease, or cancer.
  - Combining genomic sequence data with medical records has created a ground-breaking research resource.
  - To date, actionable findings have been found for 1 in 4/1 in 5 rare disease patients, and around 50% of cancer cases contain the potential for a therapy or a clinical trial.

- Breakthroughs in medical imaging:
  - Machine Learning technology has been used on thousands of eye scans to identify signs of eye disease and recommend how patients should be referred for care.

nature medicine

Article | Published: 13 August 2018

Clinically applicable deep learning for diagnosis and referral in retinal disease

Jeffrey De Fauw, Joseph R. Ledsam, [...] Olaf Ronneberger ✉

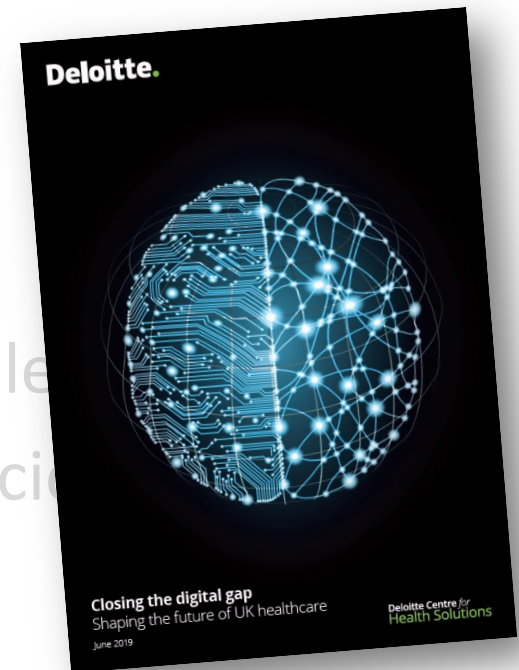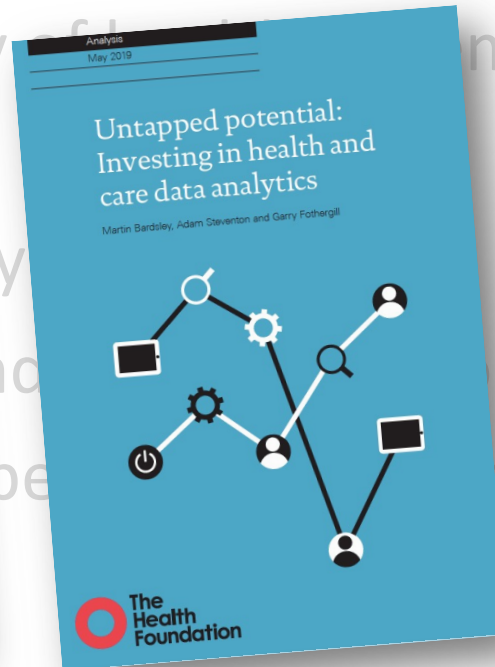*Nature Medicine* **24**, 1342–1350 (2018) | Download Citation ⬇

Abstract

The volume and complexity of diagnostic imaging is increasing at a pace faster than the availability of human expertise to interpret it. Artificial intelligence has shown great promise in classifying two-dimensional photographs of some common diseases and typically relies on databases

Challenges for Health Data Science

- Digital maturity of health systems
- Data quality
- Data complexity
- Data sharing and linkage in a complex environment
- Seamlessly embedding AI & Data Science in clinical practice
- NHS culture
- Data ethics
- Training the medics of the future

Challenges for Health Data Science

- Training the medics of the future

Module Overview

Data Science in Medicine

How can we represent and interpret medical data?

Hands-on, practical experience
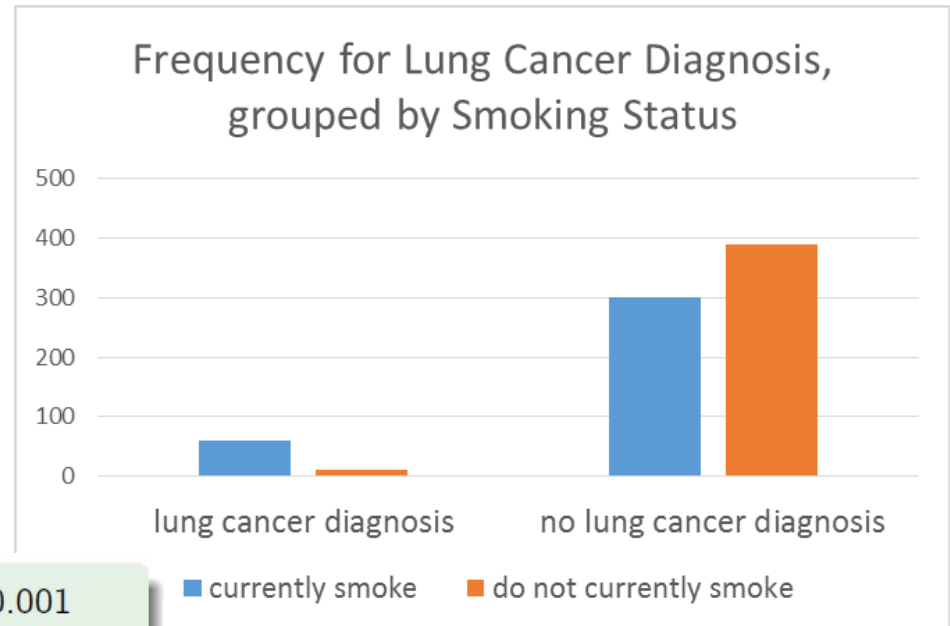
Data Science in Medicine – in a nutshell

How can we represent and interpret medical data?

Hands-on, practical experience

Topics covered:

- Statistical analysis of biomedical data

- Relational databases for medicine and healthcare

- Medical ontologies and graph data

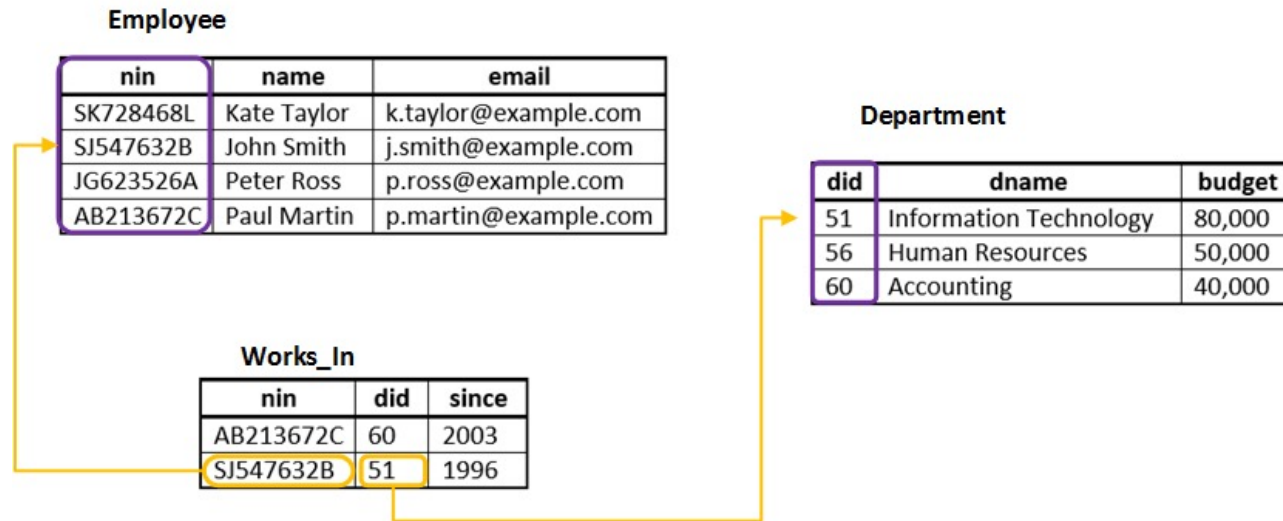- Epidemiology

Data Science in Medicine – in a nutshell

Frequency for Lung Cancer Diagnosis, grouped by Smoking Status

| ρ	p = 0.10	p = 0.05	p = 0.01	p = 0.001
N = 7	0.669	0.754	0.875	0.951
N = 8	0.621	0.707	0.834	0.925
N = 9	0.582	0.666	0.798	0.898
N = 10	0.549	0.632	0.765	0.872

ine and healthcare

data

- Statistical analysis of biomedical data (L2, L3, L5, L6)
- Epidemiology

Data Science in Medicine – in a nutshell

Employee

nin	name	email
SK728468L	Kate Taylor	k.taylor@example.com
SJ547632B	John Smith	j.smith@example.com
JG623526A	Peter Ross	p.ross@example.com
AB213672C	Paul Martin	p.martin@example.com

Department

did	dname	budget
51	Information Technology	80,000
56	Human Resources	50,000
60	Accounting	40,000

Works_In

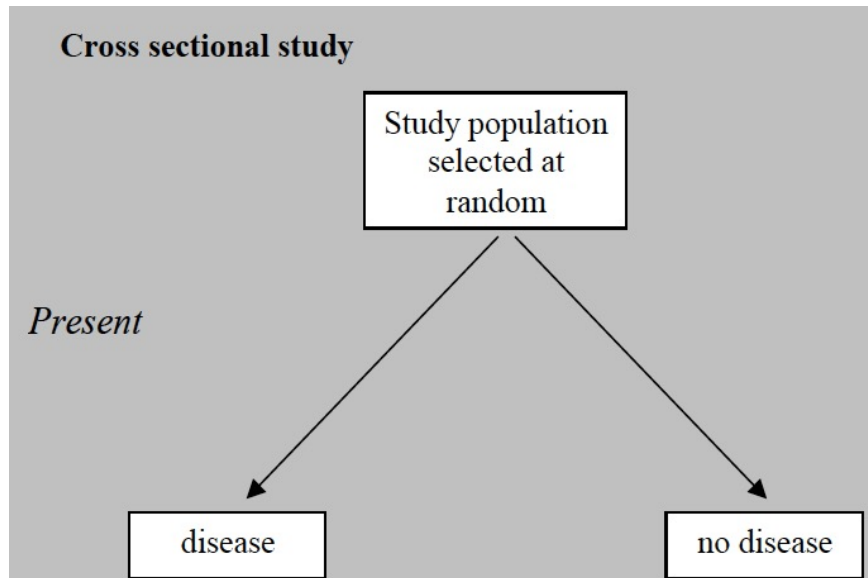| nin	did	since
AB213672C	60	2003
SJ547632B	51	1996

- Statistical analysis of biomedical data

- Relational databases for medicine and healthcare (L7, L9)

- Medical ontologies and graph data
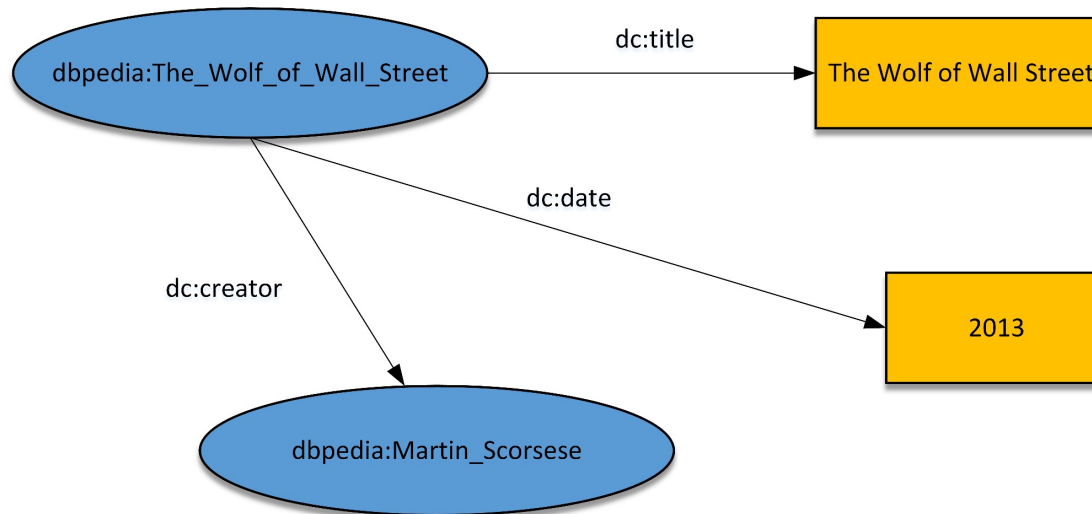
- Epidemiology

Data Science in Medicine – in a nutshell

- Statistical analysis of biomedical data
- Relational databases for medicine and healthcare
- Medical ontologies and graph data (L10)
- Epidemiology

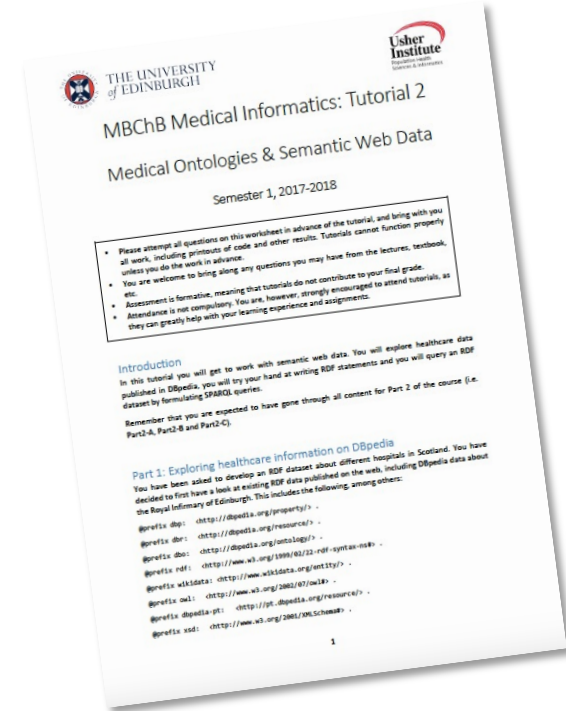# Data Science in Medicine – in a nutshell

- Statistical analysis of biomedical data

- Relational databases for medicine and healthcare

- Medical ontologies and graph data

- Epidemiology (L4, L8, L11-L13)

Data Science in Medicine – delivery
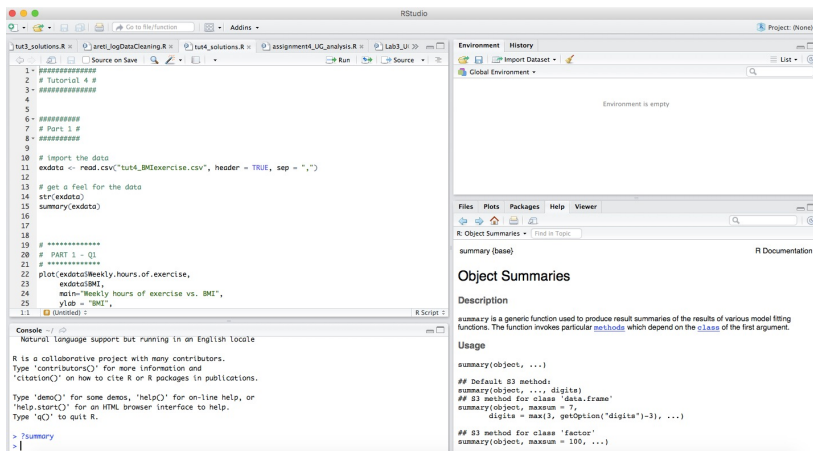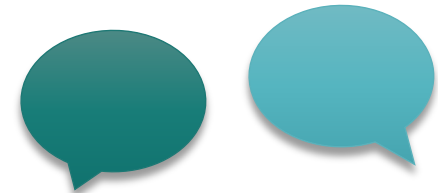
Online pre-recorded video lectures

Face to face tutorials

Face to face labs

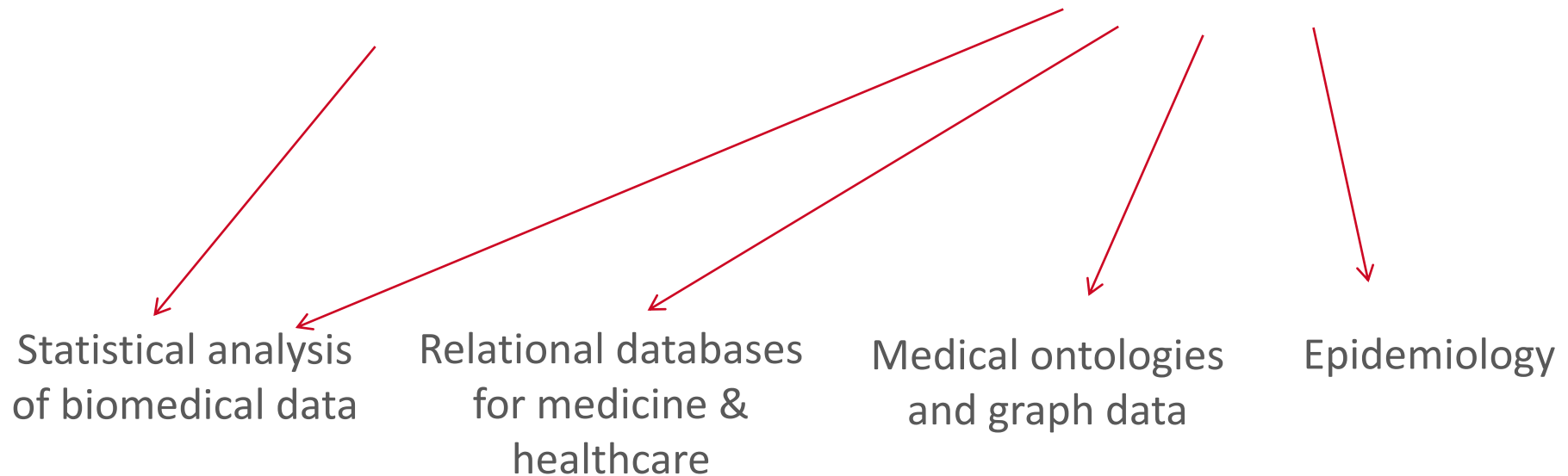Online discussion forum

Data Science in Medicine – assessment

2 practical assignments   +   final exam

Statistical analysis of biomedical data

Relational databases for medicine & healthcare

Medical ontologies and graph data

Epidemiology

Data Science in Medicine – FAQs

- **Q: Why are we taking Data Science in Medicine?**
  A: Medicine is now a data-intensive discipline. This module will equip you with the basic knowledge and skills needed to understand and work with data throughout your career.

- **Q: Are tutorials and labs compulsory?**
  A: Attending tutorials and labs is compulsory. If you have good reasons to miss a session, you should let your year coordinator, tutor and lab demonstrator know in advance to arrange a recap.

- **Q: What am I expected to do before a tutorial?**
  A: You should go through the corresponding lecture material and attempt all tutorial questions.

- **Q: What am I expected to do before a lab?**
  A: You should go through the corresponding lecture material.

Data Science in Medicine – more info

- Visit the course website

- Post your question in the discussion forum

- Contact the course organiser A.Manataki@ed.ac.uk and/or the Year 2 coordinator Stuart.Mallen@ed.ac.uk

Conclusions

- There is an unprecedented volume of data in biomedicine and healthcare.

- Exciting opportunities: improved diagnostics, learning health system, precision medicine

- It's important to develop key data science skills for your future career!

This can be fun!