

Data Science

COMP5122M

Introduction

Roy Ruddle

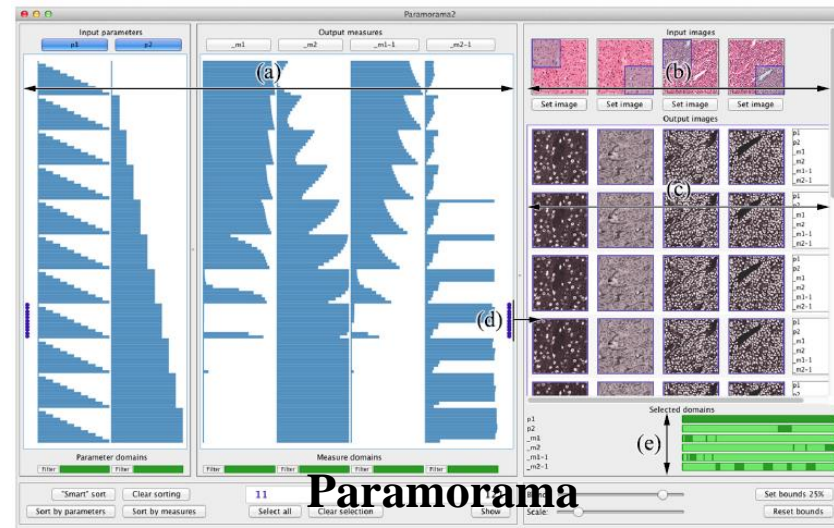
Professor of Computing

Director of Research Technology, Leeds Institute for Data Analytics

Fellow of the Alan Turing Institute

Roy Ruddle

- Background: Engineering, computing & psychology
- Visualization research
 - Presentation vs. New insights vs. Pipeline design
 - My trademarks
 - Fundamentals → applications → real-world usage
 - Interfaces & User evaluation
 - Ultra-high definition (Powerwall) displays



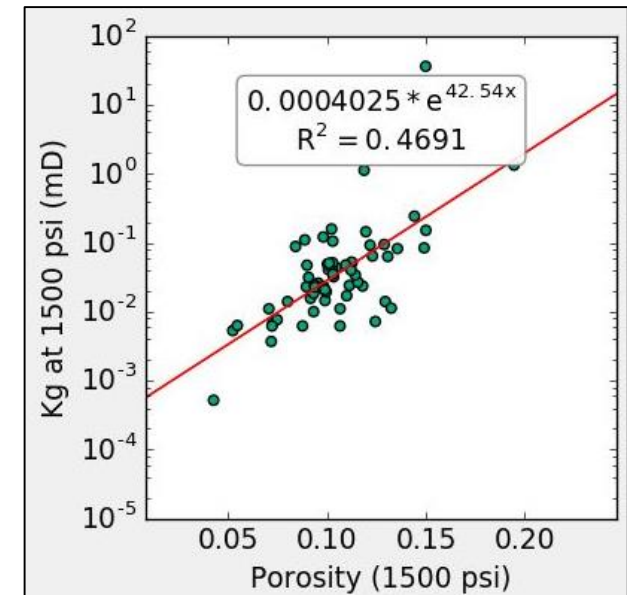
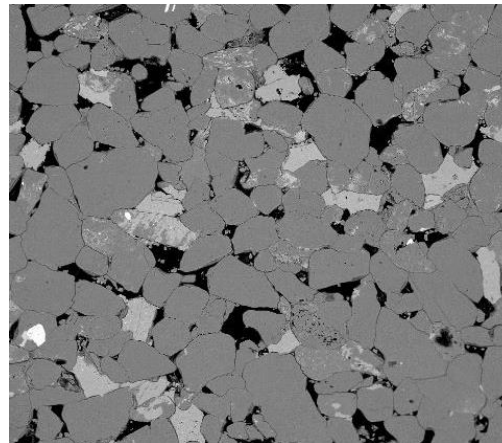
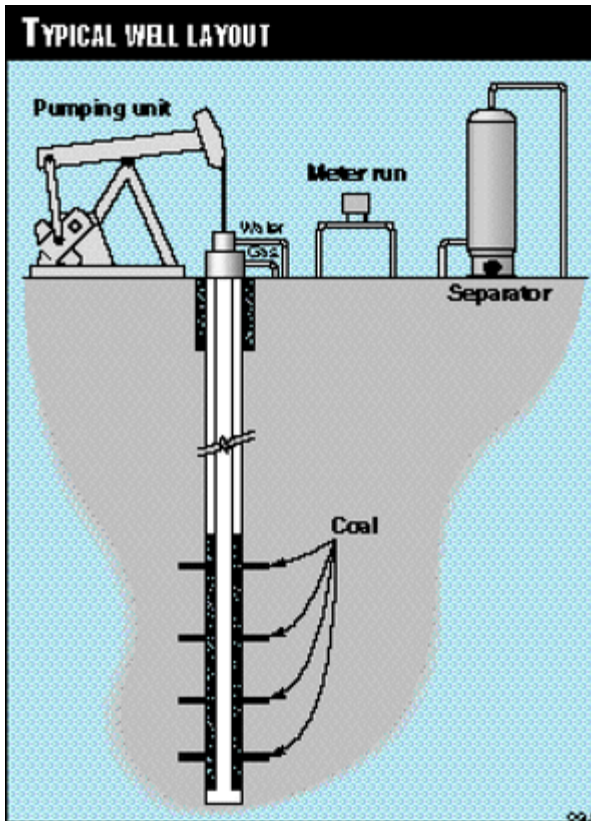
Leeds Virtual Microscope

- **Research (2006-17) → launched as product**
 - **Roche uPath Enterprise Software (2019)**



PETMiner

- Research (2013-17) → spin-off (Petriva Ltd; 2017)



QuantiCode

- **2016-2020; EPSRC EP/N013980/1; £980k**
- **Develop novel data mining and visualization tools and techniques**
 - For quantitative and coded longitudinal data
 - E.g., Retail & health records
- **Partners include**
 - NHS Digital, Leed City Council, Bradford Institute for Health Research, and Sainsbury's

Ruddle & Hall. (2019). QualDash: Using miniature visualizations of descriptive statistics to investigate the quality of electronic health records. Proc. HEALTHINF.

<https://raruddle.files.wordpress.com/2019/01/ruddle-healthinf-2019.pdf>

QualDash

- **2017-2020; NIHR; £880k**
- **Aim**
 - **Designing and evaluating an interactive dashboard to improve quality of care**
- **Deployed in 5 hospitals**
 - **For 2 National Clinical Audits**
 - **Paediatric intensive care**
 - **Myocardial Ischaemia (heart attack management)**

Elshehaly, et al. (in press). QualDash: Adaptable generation of visualisation dashboards for healthcare quality improvement. IEEE TVCG.

<https://raruddle.files.wordpress.com/2020/09/elshehaly-ieee-tvcg-2021-1.pdf>

Leeds Institute for Data Analytics

- **Opened 2015** <https://lida.leeds.ac.uk/>
 - 2000 m²
 - £50m research portfolio
 - Staff from 5+ faculties, and external organisations
- **Specialisms**
 - Medical bioinformatics
 - Consumer data
- **Infrastructure**
 - Safe rooms (confidential data)
 - Storage & high-speed network
 - Visualization workstations & Powerwalls

The Alan Turing Institute

- The UK's national institute for data science and artificial intelligence
- Opened 2015
- Physically located in London
- 13 partner universities
 - Birmingham, Bristol, Cambridge, Edinburgh, Exeter, Leeds, Manchester, Newcastle, Oxford, Queen Mary, Southampton, UCL, and Warwick

Data science module introduction

- Teaching format
- Topics
- Practicals
- Assessment
- Getting ahead

Module aims are to learn how to:

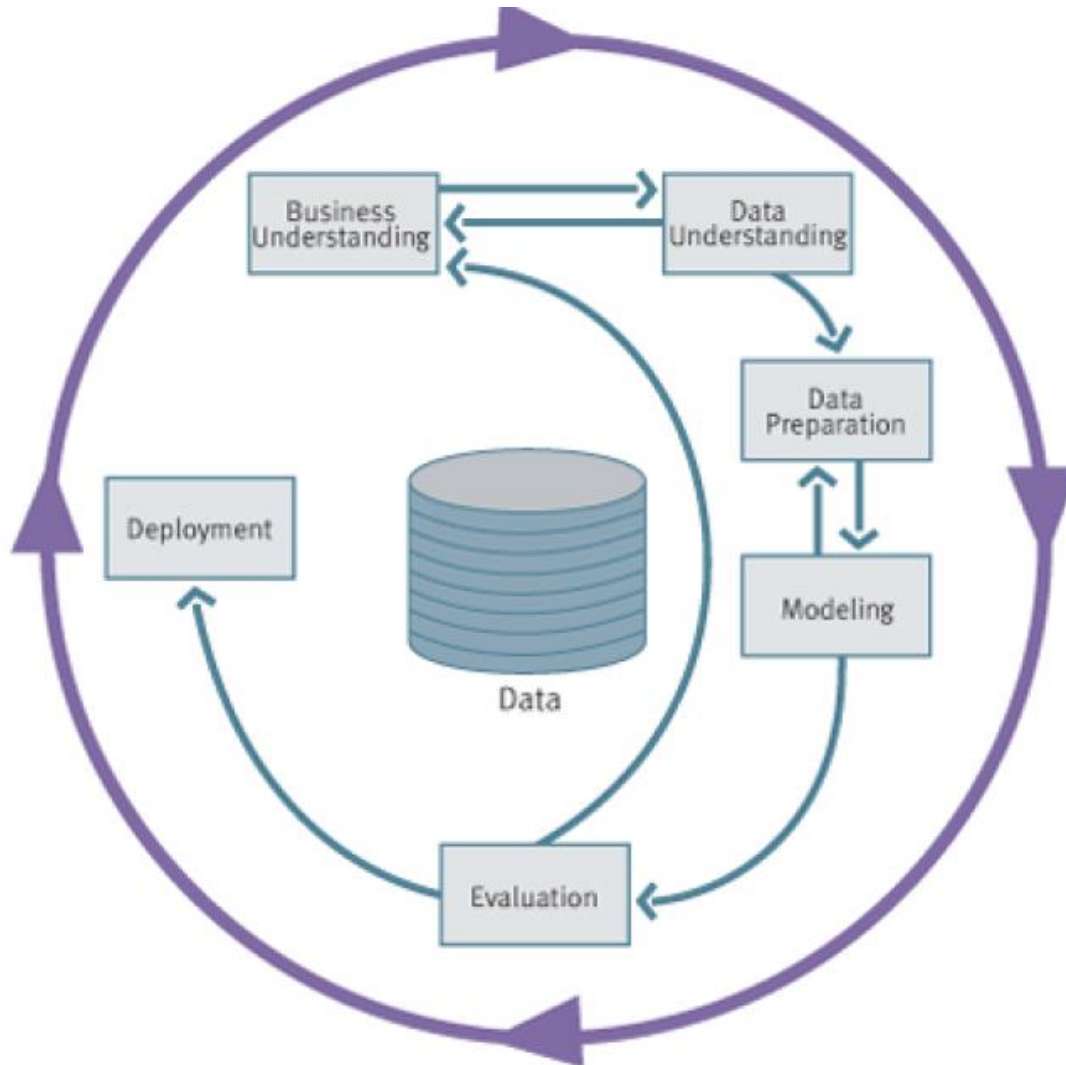
- Understand the work of a data scientist
- Think analytically
- Understand ethics & data governance issues
- Apply problem-solving skills to effectively analyse data and communicate findings

Teaching format

- **Each week**
 - **Lecture (live on Zoom & recorded)**
 - Type questions into chat or raise your hand
 - **Practical (ditto)**
- **Private study**
 - **Provost & Fawcett. 2013. Data Science for Business: What you need to know about data mining and data-analytic thinking. ISBN-13: 978-1449361327.**
 - **And other resources (see Minerva)**
- **Discussion: Microsoft Teams**

Topics

- Each lasts ≈ 2 weeks



CRISP-DM: Cross-Industry Standard Process for Data Mining (Shearer, 2000).

Practicals

- **Analysis software**
 - Pandas (Python), R ... or anything else!
- **Visualization software**
 - Matplotlib, etc.
 - Tableau
 - Great for info vis, and for employability
 - Free licence under Tableau for Teaching program

Assessment

- **Individual**
 - **Data quality issues (5%) & peer feedback (5%)**
 - **Ethics (20%)**
 - **Online test (40%)**
- **Groups (max. 4 people)**
 - **Analysis case study (30%)**

Getting ahead

- **Tableau tutorial (see practicals)**
- **Data Science for Business textbook**
- **See Minerva announcements for details about the self-study you're expected to do for each topic**

Introduce yourself

- Before the end of today, please complete the “Survey: Introduce yourself”
 - Only 3 questions
 - Under “Assessment” on Minerva

Questions?