Welcome to the Master of Data Science!

Prof. Howard Bondell

School of Mathematics and Statistics
University of Melbourne
howard.bondell@unimelb.edu.au

20 February 2018

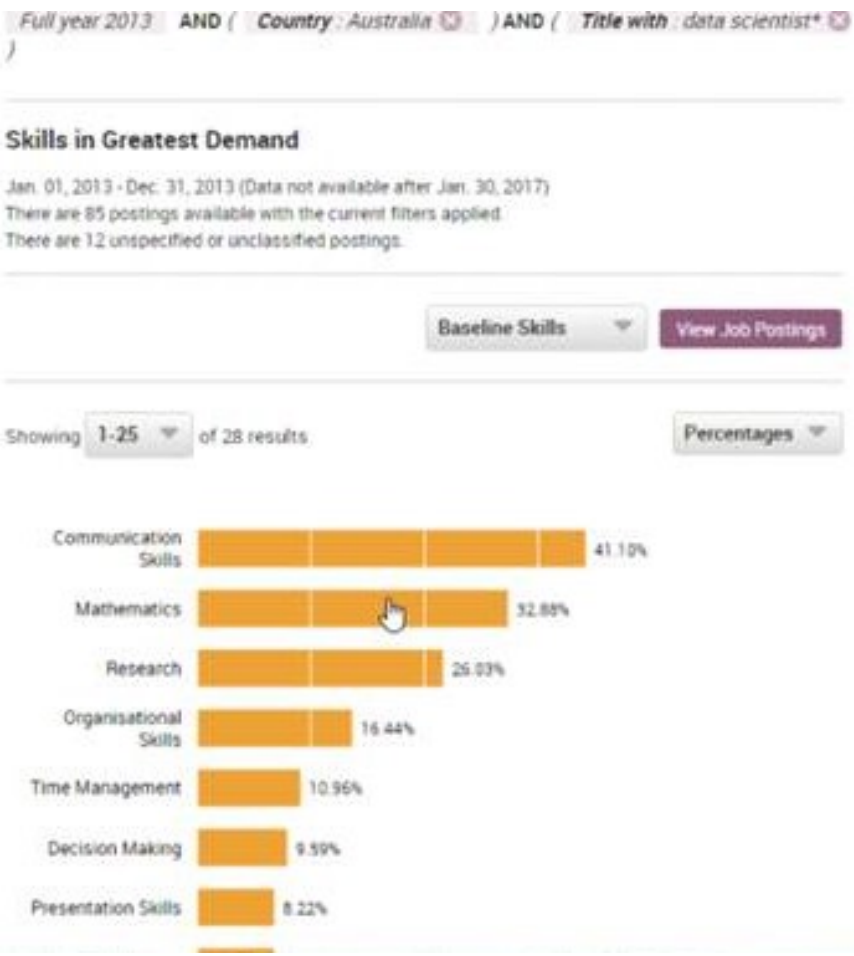
Some background:

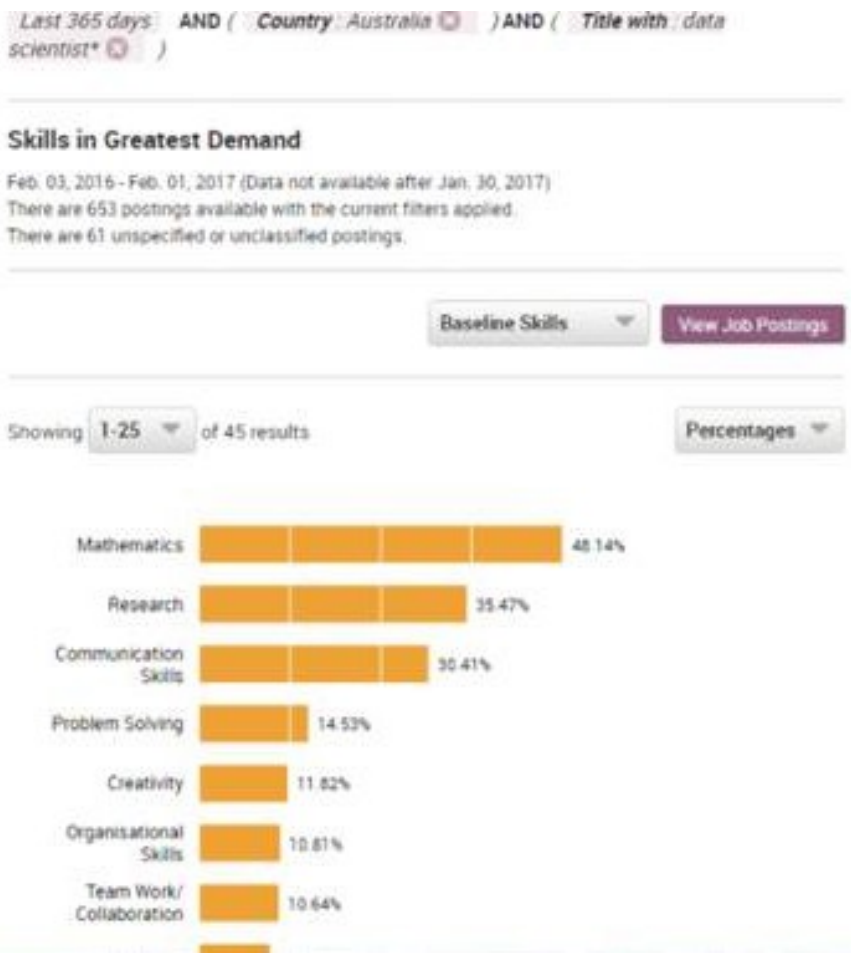
Data collated by Burning Glass Technology in 2015

The number of jobs being advertised for Data Science increased 717% from 2012 - 2015.

Source: Fiona Simpson, Careers & Industry Consultant, Science Faculty

Change in demand for data science from 2013 (left) to 2017 (right):





Areas of active research in data science

- Extreme values and rare events
- Functional data analysis
- Design and optimisation of experiments
- Spatial statistics and time series analysis
- Bioinformatics and statistical genetics
- Machine learning / Artificial Intelligence
- Text data analysis
- Network data analysis
- ...

Example 1: Recommender Systems

- Netflix, Amazon, etc., take consumer ratings and use this information to recommend movies or products to the consumer

- How do they make recommendations?
- Goal: Predict rating that user would give to each item
- Matrix Completion
- Recommend items with high predicted ratings

Users	Movie1	Movie2	Movie3	Movie4	Movie5	Movie6	···
User1	?	?	4	?	1	?	···
User2	2	5	2	?	?	2	···
User3	?	?	5	3	2	4	···
User4	1	?	?	4	?	?	···
User5	2	3	?	?	?	?	···
···	···	···	···	···	···	···	···

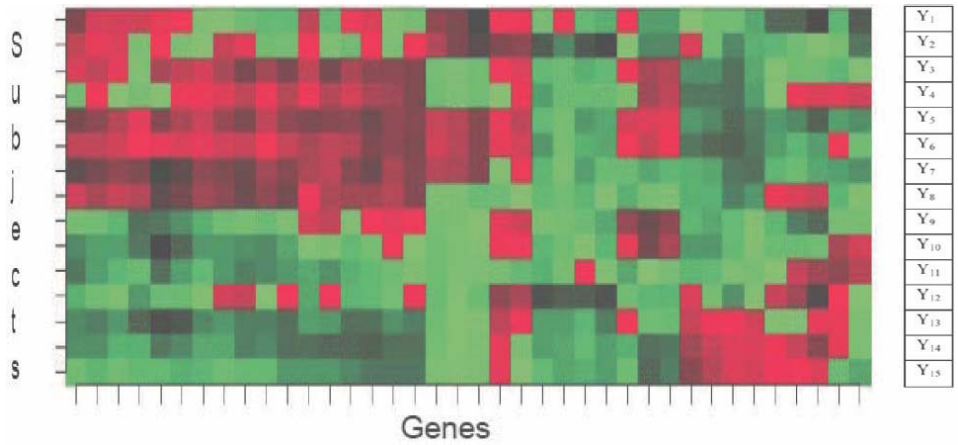
Techniques Used

- Low-rank matrix approximations
- Similarity weighting
- Regression modelling
- Large-scale programming - millions of *predictions*

Example 2: Genomics Disease Association

- Genomic data often yields measurements of thousands, or millions, of genes

- Goal: Discover genes having association with response, such as disease status
- Often just a small number have association
- Finding needles in haystack



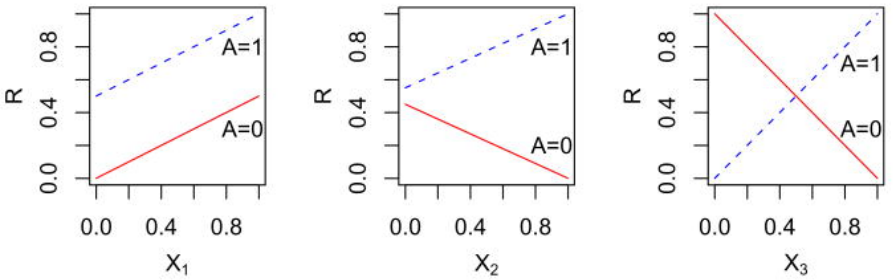
Techniques Used

- Regression modelling
- Regularisation
- Variable selection
- Sparsity
- Large-scale programming - millions of *predictors*

Example 3: Personalised / Individualised Treatments

- Individuals can respond differently to treatments, there is not always one treatment best for all patients

- Goal: Assign treatment that would work best for that particular patient
- Using characteristics such as genetics, family history, etc., find subgroups
- Could be large number of potential variables, most not useful for decision-making



(a) No interaction (b) Non-qualitative interaction (c) Qualitative Interaction

Techniques Used

- Regression modelling
- Regularisation
- Variable selection (on interactions)
- Classification