

Data analytics: the skills need in STEM

Conference report

Organised in partnership with
the Royal Statistical Society

Held on 16 November 2016

THE
ROYAL
SOCIETY

ROYAL
STATISTICAL
SOCIETY
DATA | EVIDENCE | DECISIONS

Introduction: the importance of data

Data are now being recorded at a volume and velocity that can no longer be effectively analysed through traditional approaches. Consequently, the field of data analytics is growing rapidly in importance across many sectors of industry, government and academia.

For industry, data analytics can make companies more innovative and productive by understanding their markets, opportunities and risks, allowing them to use predictive decision making to reduce costs. Many companies are now beginning to see the value data can add across all their business operations and markets, from research and development to manufacturing to sales. It is also opening up new approaches in research, where the ability to collect, interrogate and analyse data at scale will lead to many future breakthroughs, eg in biomedical science. However, to meet the potential of data analytics to drive new research and innovations and increase productivity, it will be necessary to train and develop a skilled data science workforce.

On 16 November 2016, The Royal Society hosted a conference entitled *Data analytics: the skills need in STEM*, organised in partnership with the Royal Statistical Society. Featuring speakers across multiple sectors, this conference explored the need for data analytics skills in the scientific workforce. The industry speakers gave a clear articulation of their requirements, outlining the gaps and challenges that exist and suggesting possible solutions.

This document is not a verbatim record, but summarises the discussions that took place during the day and the key points raised. As such, it reflects the views and opinions of the speakers, and not necessarily those of the Royal Society or the Royal Statistical Society. The conference was chaired by Dr Robert Hercock, BT, and member of the Royal Society's Science, Industry and Translation committee. A full programme with a list of speakers is available on our website: royalsociety.org/science-events-and-lectures/2016/11/data-skills-workshop



Dr Robert Hercock, BT

Issues around supply and demand for data scientists in the UK

- Data analytics is a relatively new field. Many companies are only beginning to develop capabilities, while educational systems and recognised career paths are not yet fully established.
- Quantifying supply and demand for data skills is not trivial, particularly as there are multiple definitions of the term “data science” and no standard occupational classification for “data scientist” using the Office for National Statistics database.
- Top end estimates suggest that 58,000 new data science jobs are being created each year.¹
- In many sectors, demand is outstripping supply. For example, GSK reported 25% of their data science vacancies have remained vacant over the past 18 months, while McLaren Applied Technologies also noted difficulties attracting applicants.
- The need for programming and coding skills means data scientists often come from computer science backgrounds, but also from other disciplines with strong quantitative skills, from mathematics to biology.
- Employers are not articulating to the education system the roles and opportunities that exist for data scientists within their organisations, meaning educators and students do not see potential career opportunities.
- Research has suggested that the profession is male-dominated, with only 1 in 5 senior tech roles held by a woman and 1 in 10 tech teams having no women working in them at all.²
- Many data analysts working in the UK come from the EU, so the impacts of Brexit and any potential restrictions on mobility need to be monitored.



Dr Caroline Hargrove, McLaren Applied Technologies (left) and Adrian Mardell, Jaguar Land Rover.

1. Count us in: Quantitative skills for a new generation. The British Academy, 2015. www.britac.ac.uk/sites/default/files/Count-Us-In-Full-Report_0.pdf (accessed January 2017)

2. Women in Technology Survey 2016, Mortimer Spinks. Downloadable from www.mortimerspinks.com/media-centre/research/launching_our_5th_annual_women_in_tech_survey.cfm (accessed January 2017)

What technical skills are needed?

- Expert data analysts need deep mathematical and statistical rigour aligned with strong technical skills in coding, modelling and visualisation.
- The growth in size of many datasets beyond the power of single unit processing requires the ability to work on distributed systems using programming languages; confidence working with large amounts of semi-structured, heterogeneous data; and the understanding of machine learning techniques and the underpinning Bayesian statistics.
- As data science moves into new domains, analysts need to be able to develop new computational methods and understand how to collect new datasets.
- Where profit margins are small, eg in online advertising, longstanding, established algorithms which are cheaper to operate are still commonly used. Therefore, data analysts need to be able to apply methods and tools appropriate to the given context.

The wider skills requirements

- Many big data problems require interdisciplinary approaches, and there is a demand for people who can work confidently across disciplines.
- General communication skills and the ability to translate the results of data analysis into a business context are important.
- Being able to connect statistical metrics with commercial metrics, understood by business managers, eg turnover or sales, is essential in demonstrating the benefits of data analytics' predictive power to an organisation, so that it can become embedded in business planning
- Creativity, a willingness to be explorative and an aptitude for making and applying hypotheses are needed when working in new domains.
- Understanding of the societal and governance aspects is necessary to ensure work meets legal and professional standards, and mistakes and controversies are avoided.
- It is unrealistic to expect individual data scientists to hold all the technical and wider skills listed above (the so-called “unicorn” or, in German, *eierlegende Wollmilchsau* – the egg-laying, wool-growing, milk-producing pig). Organisations need to consider how they can build expertise by recruiting talent into multi-disciplinary teams.

Tools

- A number of tools already exist to help data scientists increase their productivity and two areas of development were discussed at the meeting:
 - Automating data tasks using machine learning and AI
 - Integrating human and machine decision making through visual analytics
- Tools will augment, rather than replace, the work of humans, through “human-in-the-loop” development.
- As end-users will operate without review, tools must be accessible and problems arising from inconsistencies, bad technology, biases and unsound results need to be avoided.

Machine learning

- Machine learning and AI can free up data analysts’ time by automating laborious and time-consuming tasks, eg data wrangling, or undertaking ones currently beyond their capacity. There is great opportunity for analysts to develop and shape machine learning algorithms to their advantage
- Tools such as the University of Cambridge’s Automatic Statistician³ will empower human users to choose the best model to interpret their data and allow them to concentrate on analysis, interpretation and application of the findings.
- The lack of interpretability, ie the ability of machines to explain their decision making, presents difficulties when things “go wrong.”

Visual analytics

- Visual analytics supports analytical reasoning by building interactive visual interfaces to help domain experts make predictions and decisions based upon data, and understand the uncertainties associated with it.
- Visual analytics tools, such as those being developed in BT and McLaren Applied Technologies, must be intuitive, powerful and adaptive, and relate results to the business context.



Dr Ling Ge, Imperial College London

3. www.automaticstatistician.com/index (accessed January 2017)

Solutions

Encouraging people into data science

- Employers need to better articulate and promote what data science does in their organisations. This includes reaching out to different STEM disciplines and communicating with all stages of the education system.
- Career advice and guidance needs to highlight the multiple paths into data science careers, supported by outreach programmes, visible role models, training and mentoring.
- Young professionals want to work in an inclusive way and company leadership needs to adapt to create an environment that is compelling for everyone. Overcoming myths and perceptions about data science will make roles more attractive and encourage a more diverse workforce.
- Businesses should see the trend for programmers and software engineers to share information across communities as an opportunity, not a threat, and integral to their workforce's professional development.

Education and training

- A recent survey of the UK's digital workforce has suggested that the majority of training is either done in-house by companies or self-taught.⁴
- Online training courses, competitions, MOOCs and modules are increasingly used, and the best allow participants to work on problems with a business context.
- As argued by the Royal Statistical Society, there is a clear need for the educational system, including Universities, to grow the underlying pipeline of trained data scientists with the necessary depth of quantitative skills.⁵

- Universities should work with industry to ensure they provide graduates with the appropriate statistical and technical education. Industry fellows working in academic institutions can input into curriculum development and help ideas cross between sectors and disciplines.
- As recommended in the recent Shadbolt and Wakeham reviews into science degree accreditation and graduate employability, work experience during degrees greatly improves employment outcomes.^{6,7}
- However, more needs to be done to encourage and support employers to take on graduate placements, particularly outside London and the South-East.
- Universities can also offer modules, professional development and conversion courses in data analytics to PhD students, post-doctoral and academic staff, and industry professionals, such as those offered by Imperial College London's Data Science Institute.
- At later career stages, opportunities to gain real business experience are important in helping people transition into data science careers, eg Pivigo's S2DS training programme that brings together teams of PhD graduates to work with companies on data analytics projects.

4. TechNation 2016: Transforming UK Industries, TechCityUK/NESTA, 2016. www.techcityuk.com/wp-content/uploads/2016/02/Tech-Nation-2016_FINAL-ONLINE-1.pdf (accessed January 2017)

5. Education and skills for the data economy, Royal Statistical Society, 2016. www.rss.org.uk/RSS/Influencing_Change/Data_manifesto/Education_and_skills_for_the_data_economy/RSS/Influencing_Change/Data_democracy_sub/Education_and_skills_for_the_data_economy.aspx (accessed January 2017)

6. Shadbolt Review of Computer Sciences Degree Accreditation and Graduate Employability, 2016 www.gov.uk/government/uploads/system/uploads/attachment_data/file/518575/ind-16-5-shadbolt-review-computer-science-graduate-employability.pdf (accessed January 2017)

7. Wakeham Review of STEM Degree Provision and Graduate Employability, 2016 www.gov.uk/government/uploads/system/uploads/attachment_data/file/518582/ind-16-6-wakeham-review-stem-graduate-employability.pdf (accessed January 2017)

Developing an organisational capability

- Business leadership teams need to understand what data analytics can do for their organisation, how it can be implemented and what “good” looks like in terms of results and recruitment.
- Data analytics will involve clear problem definition, standard procedures for data preparation and storage, and engaging end-users early in the process, rather than parachuting decisions on them.
- Upskilling the workforce can involve training existing employees, as Jaguar Land Rover is currently doing.
- Alternatively, driving change can be achieved through recruitment of people with higher qualifications at entry level (eg preference for PhD students over graduates) or experienced data scientists into senior roles within the organisation, approaches that GSK are currently using.
- Capability can be developed by creating centres of excellence where employees can access advice, such as Volkswagen’s Data:Lab, and by embedding data scientists within operational teams.



Sir Mark Walport FMedSci FRS, Government Chief Scientific Adviser



Dr Ralf Herbrich, Amazon



The Royal Society is a self-governing Fellowship of many of the world's most distinguished scientists drawn from all areas of science, engineering, and medicine. The Society's fundamental purpose, as it has been since its foundation in 1660, is to recognise, promote, and support excellence in science and to encourage the development and use of science for the benefit of humanity.

The Society's strategic priorities emphasise its commitment to the highest quality science, to curiosity-driven research, and to the development and use of science for the benefit of society.

These priorities are:

- Promoting science and its benefits
- Recognising excellence in science
- Supporting outstanding science
- Providing scientific advice for policy
- Fostering international and global cooperation
- Education and public engagement

For further information

The Royal Society
6 – 9 Carlton House Terrace
London SW1Y 5AG

T +44 20 7451 2500

W royalsociety.org

Registered Charity No 207043

Issued: February 2017 DES4780