

Academic IT support for Data Science

Dr Simon Price

Advanced Computing Research Centre
IT Services, University of Bristol

What is "Data Science"?

Industry

- data analytics
- data engineering
- "big data" technologies
 - Hadoop stack (e.g. Spark)
 - NoSQL file stores & DBs

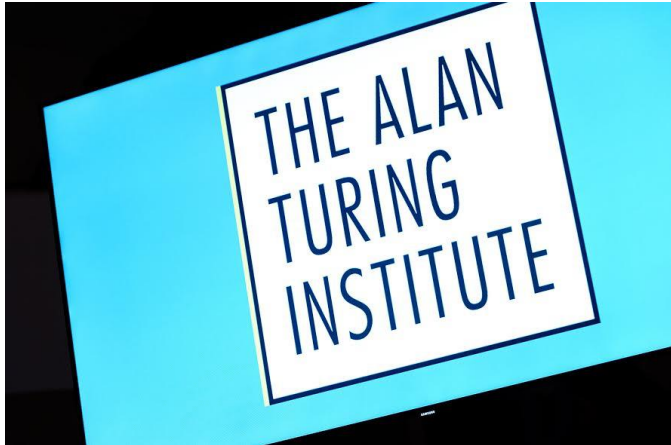
Academia

- research universities:
 - applied statistics
 - machine learning
 - computer science
- teaching universities:
 - same as Industry

Is Data Science new?

- 52% of data scientists on LinkedIn only earned the title in the past four years ([RJMetrics 2015](#))
- "50 Years of Data Science" ([Donoho 2015](#))

Major University Investment



- Alan Turing Institute (ATI), the UK's national centre for data science:
 - £52m (€67m) initial funding
 - Founding partners: British Library, Cambridge, Edinburgh, Oxford, UCL and Warwick (others may follow?)
- Many UK and US universities now have data science institutes
 - Rochester and Michigan have each invested similar sums to UK ATI



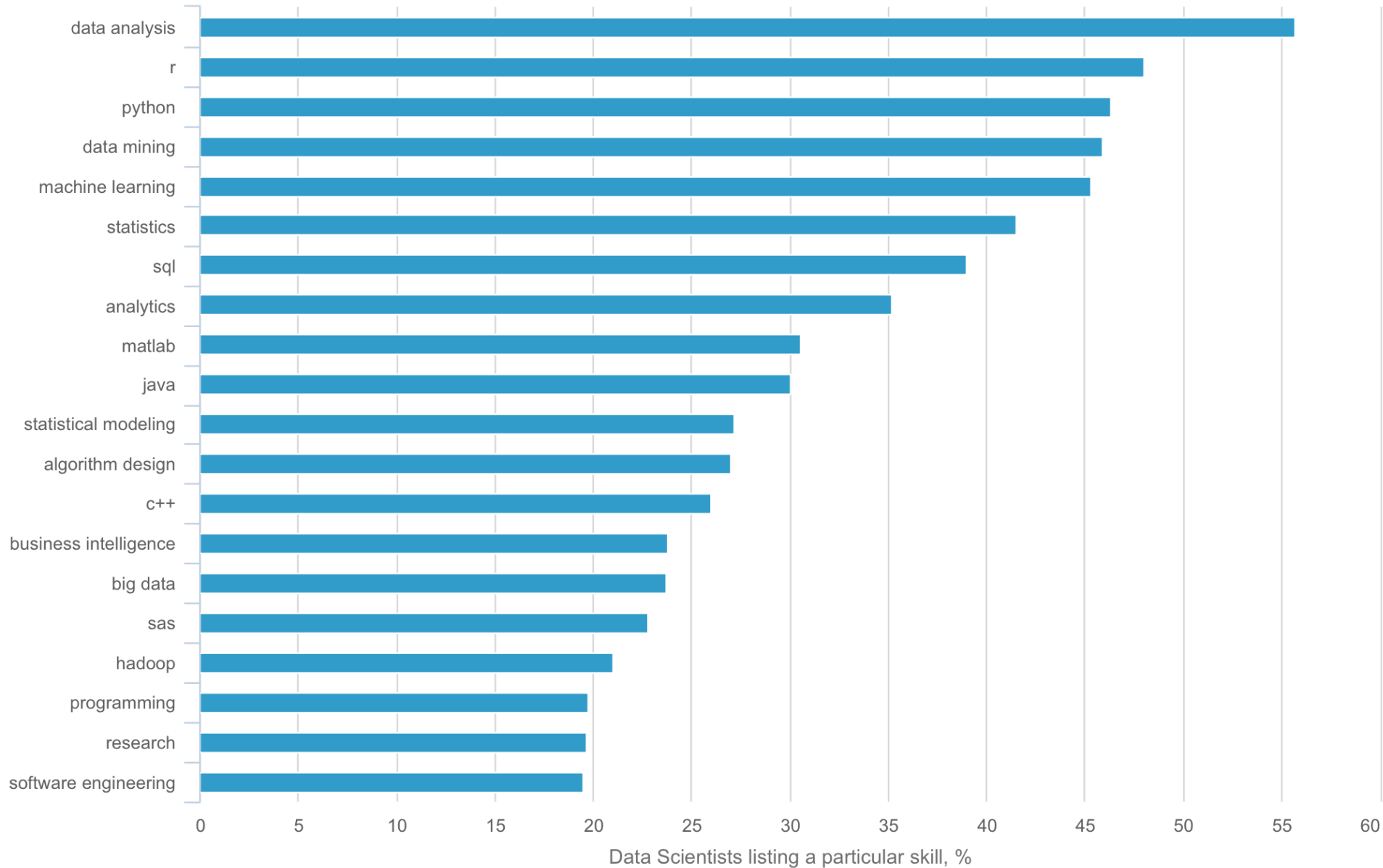
GOERGEN INSTITUTE FOR DATA SCIENCE
AT THE UNIVERSITY OF ROCHESTER



Academic IT and Data Science


- **Is data science good news for IT Services?**
 - 21% data science courses delivered online ([Swanstrom 2016](#)). Good fit with existing enterprise IT support for TEL.
 - Data science research depends heavily on IT but requires more than traditional enterprise IT.
 - Challenges arise from multidisciplinary nature of data science; there is a skills gap in IT **and** in research groups.

Top 20 Data Science Skills



Source: [JRMETRICS](#) analysis of 254,000 skill records of self-declared Data Scientists on LinkedIn, June 2015.

Bristol IT staff skills survey

- **Aim:** to better inform Research IT support
- 20 questions (mix of multiple choice and free text)
- Ran in August 2014 via  **bos** onlinesurveys.ac.uk
- Response rate: 124 / 193 (64%)

Page 3: Hardware skills

- 6 Leave the row blank where you have no active knowledge or experience.

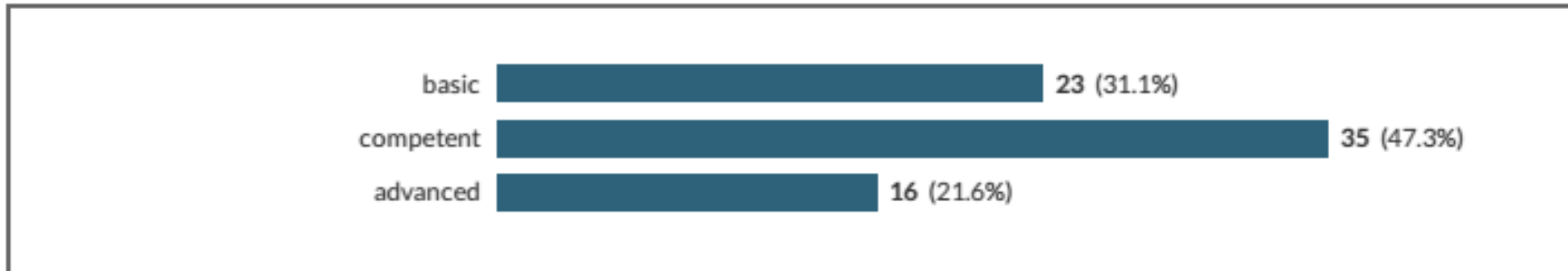
Having trouble with the format of this question? [View in tableless mode](#)

	Skill level			How often do you use your knowledge?		
	basic	competent	advanced	frequently	occasionally	not for 18 months
Computer Server Hardware	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Local Clusters	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Computing for Data Acquisition	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Computing for Instrument/Apparatus Control	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Specialist Desktop Hardware for Research	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

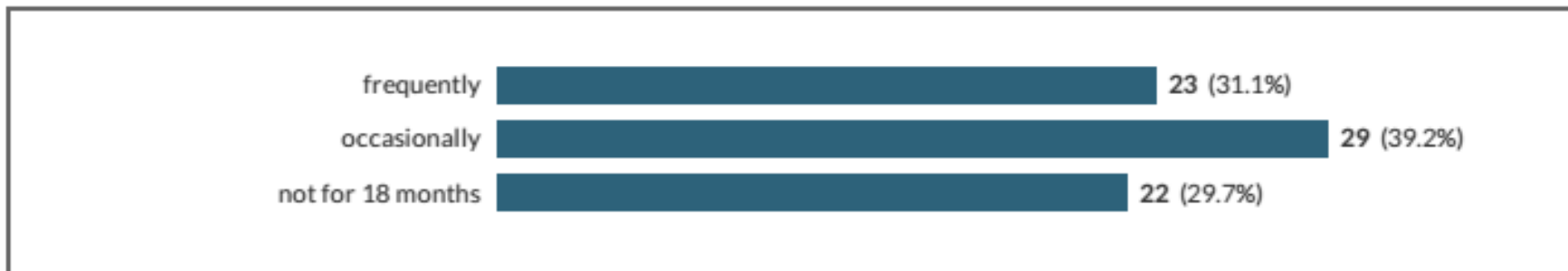
Bristol IT staff skills survey results

6.1 Computer Server Hardware

6.1.a Computer Server Hardware - Skill level



6.1.b Computer Server Hardware - How often do you use your knowledge?

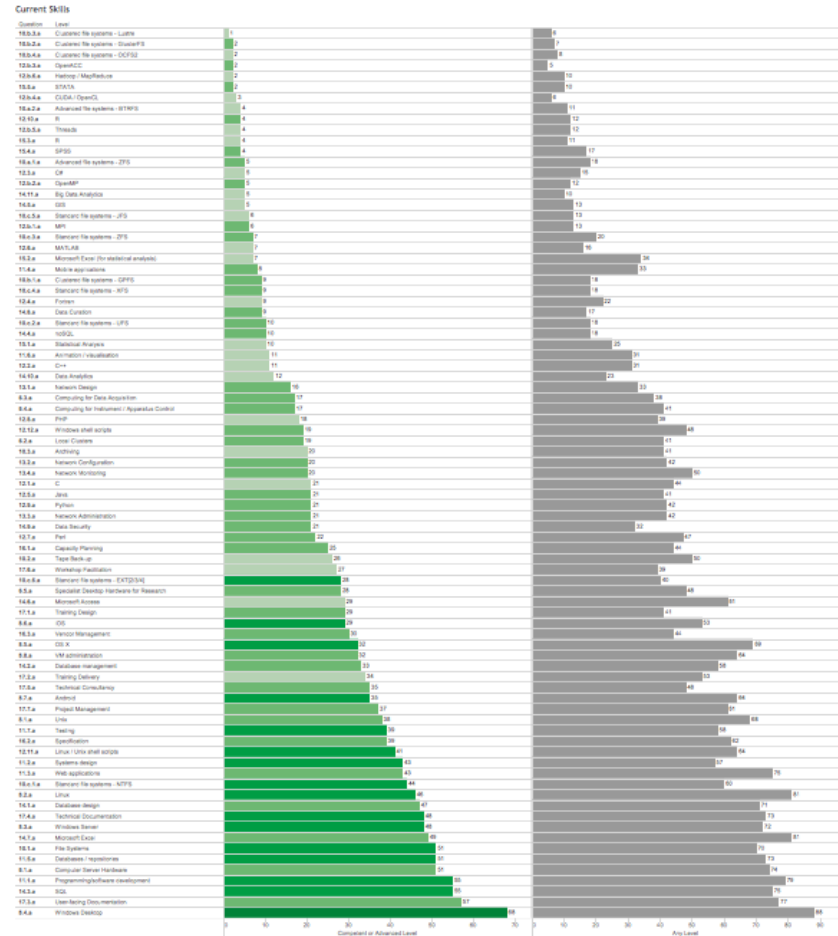


Online survey designed and run with [BOS](#)

Bristol IT staff skills survey results

What are the rarest skills (at Competent or Advanced levels)

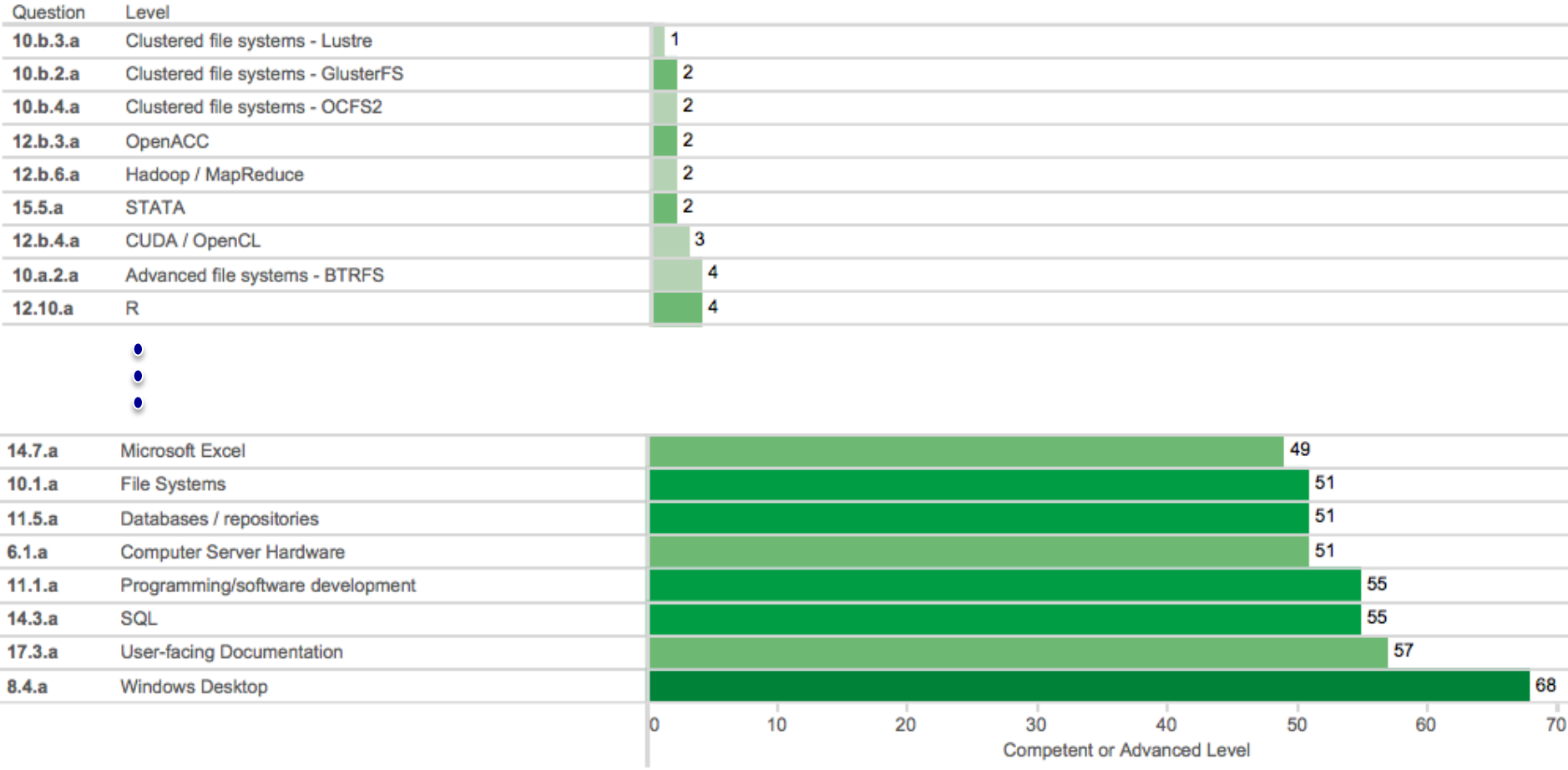
...and for each skill, what percentage have experience more recent than 18 months?



Bristol IT staff skills scarcity



Current Skills



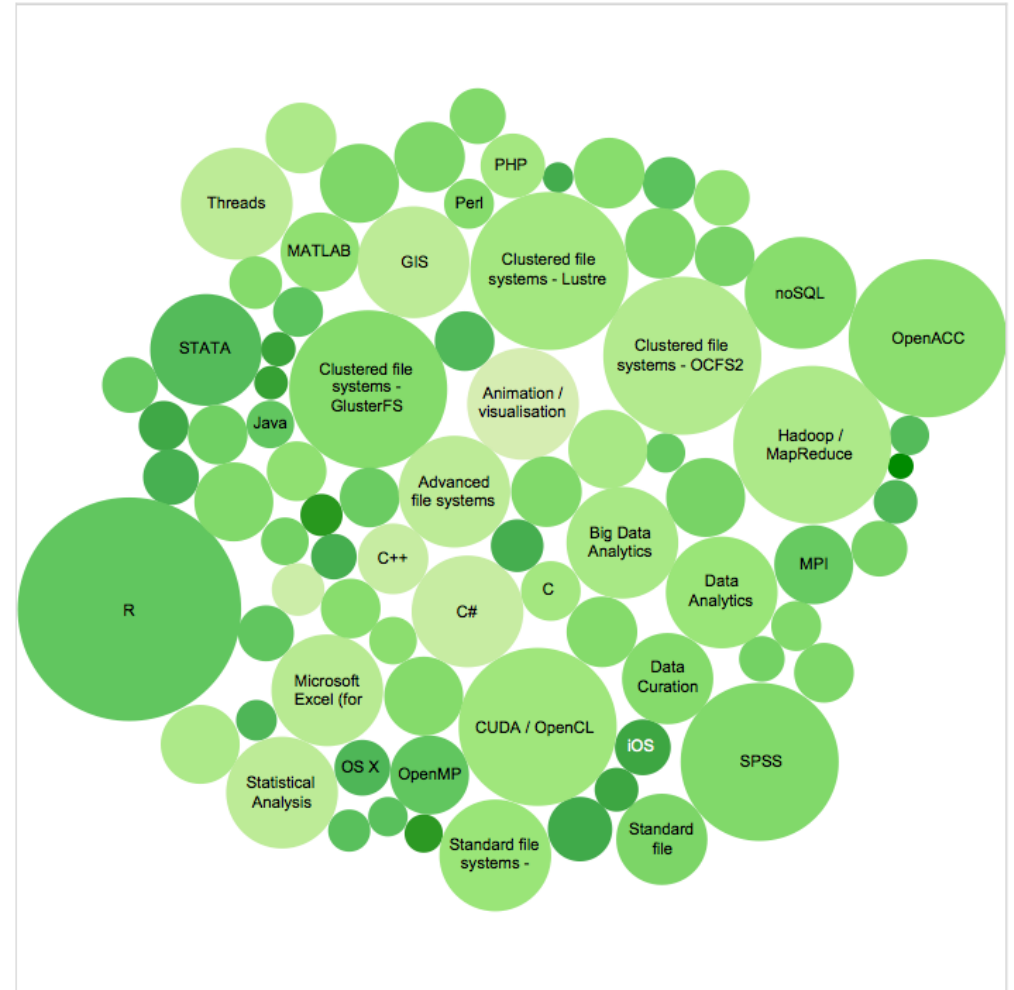
Bristol IT Services rarest skills

What are the rarest skills at Advanced level

...and for each skill, how many people have current experience?



Rare Current Expertise



Bridging the Skills Gap

- Data science research needs both **engineering** and **analytical** skills at *advanced* levels.
 - Many IT staff have the engineering skills but few have the *advanced analytical* skills.
 - Many researchers have the analytical skills but few have the *advanced engineering* skills.
- Facilitating data science research requires IT to contribute (mainly) engineering expertise:
 - IT staff **costed in** to research proposals as part of research team;
 - training researchers and promoting best engineering practice;
 - as well as the usual "free-at-point-of-use" IT advice support.

Co-designing the Bristol Data Institute

Staffing options:

- Shared posts between Institute and IT.
- IT staff supplied via chargeable research IT facility at day rate.
- IT staff seconded into research groups at full economic cost.
- Widening Institute remit to involve all data-related professional services staff (e.g. corporate data, research data management, education).

Chosen model:



- Academic and admin staffing only; no IT staff in Institute.
- Experimenting with a mix of IT staff at facility day rate and, for larger projects, seconding into research groups.
- IT membership of Governance.

Questions?

Dr Simon Price
simonprice.info