CAREERS

DATA SCIENCE

Industry allure

PhD holders with quantitative skills are landing posts at technology companies.

BY MONYA BAKER

Eli Bressert planned to spend his academic career in search of forming stars. He had completed a PhD in astronomy at the University of Exeter, UK, and had won a prestigious postdoctoral fellowship to study radio astronomy near Sydney, Australia. Citations of his papers and invitations for collaborations and conference talks were on the rise. He had no reason to want to work outside astronomy.

But a year into his studies in 2012, the grim reality of the academic job market began to make him nervous. "I sat down and calculated my odds," he recalls. "What was the chance of getting in at a good research institution in a place where my family would be happy?" He had already moved himself, his wife and their year-old son some 16,000 kilometres to Australia for his postdoc, and more transglobal moves for low pay and little stability did not appeal. Still, his research was going well, and he decided to carry on.

That same year, he and a colleague published a handbook on scientific programming, and he was recruited as an academic adviser to a start-up company that was creating software to help collaborators to co-author papers. Bressert loved the energy of the start-up and when he heard of a fellowship that groomed scientists for technology jobs in Silicon Valley, he applied — and was accepted.

He and his family moved again, this time 12,000 kilometres to Palo Alto, California. Today, he is head of data labs at Stitch Fix, a company in San Francisco, California, that creates predictive algorithms that help clients to choose clothes. He says that he loves his work evaluating computational methods in part because it offers more intellectual freedom and creativity than he had experienced in academia.

Bressert is hardly an anomaly — his company employs 20 PhD holders from disciplines as varied as astronomy, neuroscience and electrical engineering. Their biggest asset is rigorous ▶

thinking, says Eric Colson, Bressert's manager. PhD training means learning to formulate questions, test hypotheses and assess whether a solution is reliable. When it comes to modelling data, these qualities make PhD holders more sceptical than most, says Colson. "If it was perfect on the first try, a PhD's first response will be that it is too good to be true. PhDs have this patience and way of framing problems that MBAs don't have." Stitch Fix's PhD holders are just a few of the many young scientists, mainly in the United States, who have left the academic quagmire for jobs in industrial data science.

MAKE THE LEAP

Mathematicians and computer scientists are well-represented in the data-science field, but computing savvy and communication skills matter more than scientific speciality. Early-career researchers hoping to make the transition need to show that they can extract patterns from messy data and place those patterns in the context of commercial goals.

"It's important to remember that industry doesn't value insights. They value analyses that are actionable," says Michael Li, who is co-founder of The Data Incubator, a training course based in New York and Washington DC that prepares graduate students for jobs in data science. And academics skewer their chances by not knowing the ins and outs of industry, says Jake Klamka, who founded a similar training programme, Insight Data Science in Palo Alto. Otherwise qualified candidates can be dismissed as clueless for using the wrong word, such as the academic term 'study' instead of the industry argot 'experiment' or 'A/B test'.

Klamka found it hard to break into industry. He quit his PhD programme in particle physics at the University of Toronto, Canada, in 2010 and began developing tech tools in his kitchen. But although he had the expertise, he lacked knowledge of the industry. "I was 99.5% there in terms of skills," he says. "What I needed was guidance and mentorship." After a year of frustration, he headed to Silicon Valley, where he met software engineers and entrepreneurs who put him on the right track. And thanks in part to backing from the start-up incubator Y Combinator, based in Mountain View, California, he was able to launch his own company, Noteleaf.

Klamka knew that many of his friends in the physics community were interested in moving into industrial data science but were struggling, like he had, to break into industry. At the same time, his tech-community friends were

Trainees attend a Science to Data Science workshop in London.

PIVIGO ACADEMY

complaining that they had open positions but no one smart enough to fill them. So Klamka founded Insight Data Science to provide PhD holders with the training they need for a career in industrial data science. So far, everyone who has completed the 7-week programme has received job offers (see 'Learn the ropes').

JOB DESCRIPTIONS

Data-scientist jobs vary widely. Some require mainly tedious 'data munging', cleaning data and filling in gaps to make data sets suitable for relatively simple analysis. Some data scientists work as consultants on data applications; others craft new models and methodologies. Large firms such as LinkedIn, Google and Facebook, with their huge user bases and data sets, tend to support the most sophisticated data modelling.

Would-be data scientists should think broadly about their interests and where they can do what interests them, says Glenn Wong, who has a PhD in physics and is now vice-president at Recorded Future in Somerville, Massachusetts, which organizes web data to help clients deflect cyberattacks. "I don't mean 'how this snippet of DNA interacts with that snippet of DNA'," clarifies Wong, "but 'I like solving problems of a complex two-dimensional nature'. Or 'I like being surrounded by people who have wacky ideas and don't care about hierarchy.'"

Amy Heineike took a leave of absence from her PhD programme in computational social science to join a tech start-up based in San Francisco, California, that helps to advise and evaluate early-stage entrepreneurs. "The reason I was doing a PhD was to solve interesting problems, but we were already doing that," she says of her work at the firm. Several years out of academia, and now with stints at other start-ups under her belt, Heineike thinks that she has better opportunities to build ideas and implement them in industry because companies actually connect

---

LEARN THE ROPES

Find the data-science course to suit you

Many who plan to move into industry use their time at research institutions to burnish their skills and explore their options. Eli Bressert, head of data labs at Stitch Fix in San Francisco, California, recommends learning industry-favoured programming tools such as Python and R. For those who need to boost software skills, programmes such as Data Carpentry and Software Carpentry bring two-day courses to campuses across the world.

Glenn Wong, vice-president of the cybersecurity company Recorded Future in Somerville, Massachusetts, took Harvard Business School seminars when he was a physics PhD student at Harvard University in Cambridge, Massachusetts. The seminars later helped him to cruise through interviews at management-consulting companies.

As a postdoctoral student in synthetic biology at the Massachusetts Institute of Technology in Cambridge, Joy Tharathorn Rimchala, now a data scientist at financial-software company Intuit in Mountain View, California, was uncertain about leaving her academic career until she began auditing a computer-science course. "This is when I decided that data science is cool; at least as cool as my PhD," she says.

Rimchala and Bressert both moved into industry through a programme offered by Insight Data Science in Palo Alto, California. (Last year, a parallel programme opened in New York City and one in Boston will launch in July.) Course attendees work in teams to develop data-driven web applications, and also meet with data scientists at tech firms. The course is free: its costs are met by tech companies, which pay to hire fellows.

A similar initiative is Science to Data Science in London, which offers a 5-week workshop for about 85 students, who pay a £360 (US$540) fee that covers accommodation. After a week and a half of course work, small teams work with mentors from local companies to build practical tools with the companies' own data. Most of last year's fellows returned to their laboratories after finishing the inaugural programme in September, but 75% now have data-science jobs in industry, says co-founder Kim Nilsson, who has a PhD in astrophysics.

Another option is the free, 7-week Data Incubator course, which is based in New York and Washington DC and opens in San Francisco in summer 2015. Finally, the 12-week NYC Data Science Academy programme in New York City, which launched this year, costs $16,000 including coursework on tools such as R, Hadoop and Python. All programmes have more applicants than places. **M.B.**

ACADEMIC DATA DRIVE
Universities create data–science hubs

Academic science, not just industry, has a growing need for data scientists. A US$58-million effort launched last year aims to fill this gap by creating data-science hubs at the University of Washington in Seattle, the University of California, Berkeley (UCB), and New York University. The universities, along with the Gordon and Betty Moore Foundation in Palo Alto, California, and the Alfred P. Sloan Foundation in New York City, are co-funding the hubs. Grants from the Moore Foundation will be given to investigators to develop and refine data-use tools.

Karthik Ram, an assistant researcher at UCB's newly created Berkeley Institute for Data Science, is one of the first beneficiaries. His career advancement depends on his contributions of open-source code and efforts to make data more reproducible, rather than on the conventional criteria for tenure-track posts, such as publication and citation records.

Moore Foundation programme manager Chris Mentzel describes Ram and his colleagues as pioneers in a field that is gaining momentum. "We are trying to create homes for these types of researchers," he says. **M.B.**

---

with the people who use the products.

But PhD graduates have to be comfortable with abandoning quests for ever-greater accuracy in favour of commercial goals. Once a data model is working, academics might focus on sophisticated tweaks to improve accuracy and account for outliers. "But in industry, you'd be saying, 'How do I build this into the software; how do I make sure that it won't crash?'" says Heineike. "You have to go the distance for what users really want, and that's something you don't necessarily have time for in academia."

Some hiring managers worry that a desire to craft increasingly accurate models can lead academicians into an unproductive morass. John Baker, who founded a consultancy for data-science services called Datakin in Boston, Massachusetts, recalls an astrophysicist nicknamed 'Dark Matter' by his colleagues because his zeal for perfecting data models meant that he never completed his projects.

David Freeman, head of security data science at the networking firm LinkedIn in Mountain View, says that it is possible to weed out those with such tendencies during interviews. When asked to describe their accomplishments, the most-promising candidates focus more on codes they have implemented than papers they have published. Portfolios developed independently or at boot camps are another good sign of an industry fit, says Baker. "You can tell who is really academic and who really has potential by their projects."

Will Cukierski got noticed this way. He earned his PhD at Rutgers University in New Brunswick, New Jersey, where he taught computers to recognize telltale pathologies in cancerous tissues. But at night, he worked on a challenge from streaming-media provider Netflix: a US$1-million prize to anyone who could best its own movie-recommendation algorithms. He didn't win, but he caught the bug and started to spend his free time on similar contests hosted by the data-science company Kaggle, based in San Francisco. In 2012, company executives contacted him — they had noticed his entries and thought that he could earn a spot on their team. He started there as a data scientist a week after he defended his PhD.

For many PhD holders, the key to success is to find a company whose product or service fascinates them, says Sebastian Gutierrez, author of *Data Scientists at Work*. "You need someone who is excited enough about the business that they actually care that they need to meet quarterly budgets and goals."

Posts for data scientists are starting to emerge in academia (see 'Academic data drive'), but many find the industry environment more appealing. "In industry I can use 20% of the time to achieve 80% of the goal, instead of vice versa," says Shani Offen, formerly a research professor in neuroscience at New York University and now a data scientist at the question-answering site About.com, based in New York. Tommy Guy, a data scientist at the tech giant Microsoft in Bellevue, Washington, likes being rewarded for getting the right answer, no matter what it is. For instance, he can use data analysis to conclude that a proposed new feature would be unpopular with users and argue to dump it, saving the company a considerable sum and earning accolades. Conversely, he says, academia rarely rewards negative results.

Freeman likes the pace at LinkedIn. He recalls doing cutting-edge research in his postdoctoral work at Stanford University in California. "But the thing I was working on would not be seen in actual use for 20 years, if ever. I was looking for something with more immediate impact." And there's nothing like constant deadlines to focus the mind. ■

Monya Baker *writes and edits for* Nature Careers.