# *How to be...*
A DATA SCIENTIST

Suzy Moat uses data from Google, Wikipedia and social networks to measure and even anticipate human behaviour. Carinya Sharples finds out that she can even predict a riot

When the riots hit England in the summer of 2011, what had been confined to Tottenham spread like wildfire across London and then cities around England at random – or so it initially seemed to data scientist Suzy Moat, Assistant Professor of Behavioural Science at Warwick Business School. "We can do a pretty good job of modelling how the riots spread within London by assuming that if a riot had just occurred, it would become more likely that another riot would soon occur nearby," she explains. "However, this model is not very good at explaining how rioting began outside of London in cities such as Manchester and Birmingham."

Suzy and her team turned to Twitter for help. "In our Twitter-based model, we assume that if a riot occurs in one location, and we can see lots of communication between that location and a second location, then the risk of rioting might spread to this second location. We find that this model is much better at predicting the spread of the riots through England as a whole, suggesting that patterns of how information spreads might help us better understand patterns of how behaviour spreads." And if the police had been able to see in real-time which areas were communicating, they might have been better able to anticipate which city the riots would spread to next and target their resources more effectively.

If that all sounds a bit like something out of the film *Minority Report* to you, you're not far off. But instead of strange bald women who lie in tanks and see into the future, it's data whizzes who are providing the insight into past, present and future human behaviour, with the help of Google search data, Wikipedia entries, Tweets and other online chatter. The riots map is just one example that Suzy Moat broke down in a fascinating talk at TEDxZurich, entitled Decoding Our Digital Traces, but there are countless other areas worth exploring. In fact, Suzy says the most frustrating part is that she and her team have far more ideas than they have time to try out. So what does it take to be a data scientist? We asked Suzy to share her journey – and some top tips on how to get to where she is now.
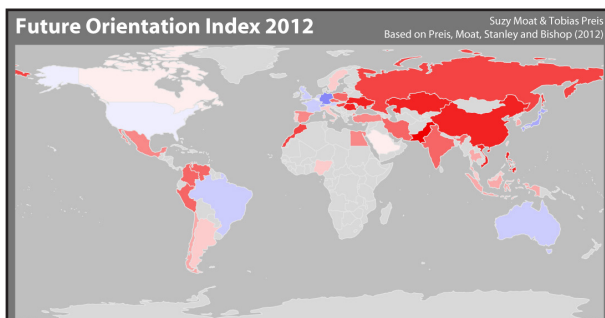
YOU'RE ASSISTANT PROFESSOR OF BEHAVIOURAL SCIENCE AT WARWICK BUSINESS SCHOOL. WHAT DOES THAT INVOLVE?
I'm a university academic, which is a job with many different sides. Most importantly, I'm a data scientist and researcher. I write computer programs, which analyse data from sites such as Google, Wikipedia, Twitter and Flickr to find out whether we can use data from the internet to measure and even predict what humans do in the real world. I write reports on this work, many of which are free to download online. I also spend a lot of time travelling around the world to present these findings to other

Future Orientation Index 2012
Suzy Moat & Tobias Preis
Based on Preis, Moat, Stanley and Bishop (2012)

scientists, as well as people from government, business and the general public.

HOW DID YOU GET TO WHERE YOU ARE NOW?
I'm lucky as I've studied and worked in a number of countries, including Germany, Switzerland and Belgium, and I have a visiting position in the US too. I started off by studying computer science as an undergraduate, which I loved – I'd been programming ever since my Dad taught me to code on a ZX Spectrum when I was a kid. I'm a bit obsessed with languages too and was fascinated by how people communicate, so I took a Masters and PhD in psychology and linguistics. When I finished my PhD, the amount of data on human behaviour and communication was really exploding, due to our increasing use of technology and the internet. I got a sequence of post-doctoral researcher jobs looking at to what extent we can use this data to make predictions of how humans behave, including one collaboration with the Metropolitan Police.

YOUR BACKGROUND IS IN COMPUTER SCIENCE AND PSYCHOLOGY. WHY DO THESE AREAS COMPLEMENT EACH OTHER?
Previously, our only way of measuring how humans behave was to put them in an experiment, or ask them to write down answers for a survey. Now, we increasingly rely on networked computer systems and smart cards to support our everyday activities, and everything we do generates data: buying bread at the supermarket, taking a ride on the Tube or calling a friend for a chat. This offers lots of intriguing new opportunities for us to find repeating patterns in what people do, which we can use to make better predictions of how humans will behave in the future. Such predictions could help us better anticipate where crimes might occur, or how illness might spread.

WHAT KIND OF DATA PREDICTIONS HAVE YOU MADE?
My collaborator Tobias Preis and I are fascinated by the idea that data on how we communicate online might help us better measure and even predict what humans do in the real world. In one study, with our collaborators Steven Bishop

Courtesy TEDxZurich

and Gene Stanley, we measured how often internet users searched for the next year – for example, 2015 – and how often they searched for the previous year – for example, 2013. We found that internet users from countries where people are richer tend to Google for more information about the future than internet users in countries where people are poorer.

WHAT PROJECTS ARE YOU CURRENTLY WORKING ON?
Recently, we started looking at photos that people upload to Flickr. We looked at what photos were taken around Hurricane Sandy, a huge hurricane that crashed into New York and New Jersey in late 2012. We gathered data on all the photos tagged with "Hurricane", "Sandy" and "Hurricane Sandy" and found that the worse the hurricane was in a given hour, the more photos were taken. However, this work has involved processing the text that people post and search for online. We're excited about what might be possible if we process the actual pictures instead. Our excellent students Merve Alanyali and Chanuki Seresinhe are cooking up some fantastic work on this at the moment.

SUZY'S FIVE THINGS YOU NEED TO BE A DATA SCIENTIST

1. ABILITY TO PROGRAM
"You need to be able to get hold of the data you are interested in analysing, whether this is data on who's friends with who on Twitter, where Boris bikes are being used in London right now, or how many complaints of loud music were made in New York last week. To do this, you need to be able to program. While a lot of this data is free online, it would generally be impossible – never mind extremely boring! – to click on everything you need, so you need to write code to do this for you."

2. KNOWLEDGE OF STATISTICS
"You need to be able to work out what patterns are in your data. People often repeat patterns of behaviour – for example, it's fairly easy to predict at what times the Tube will be packed. However, some repeating patterns of behaviour are rather more hidden than this. For example, there's a famous case of a supermarket in the US finding patterns in people's shopping that they used to predict when people were going to have a baby. To find these patterns – and work out whether they're reliable or not – you need to learn statistics."

3. PEOPLE SKILLS
"You need to be able to find questions that people want an answer to, and communicate the answer to them once you've got it. So it's not enough to be highly technically skilled – you need to be able to relate to other people as well. If you can tell a good story in written or spoken form, this is a very useful skill. Knowing a lot about a particular problem area – like crime or financial trading – is also extremely helpful."

4. A DEGREE
"To become a data scientist in business or at university, you ideally need to get an undergraduate degree, which will help you learn and demonstrate your skills in programming, statistics and communication. There are a few new courses emerging specifically in data science. Otherwise, it's a good move to pick a technical subject such as computer science, statistics, physics or maths, and do as many optional modules as you can in social science subjects such as psychology and economics along the way. If you want to stay at university and become an academic, the first step in an academic career is to get a PhD."

5. CREATIVITY AND DETERMINATION
"A slight obsessive streak is definitely useful for getting things done; however, research is also a very creative job. You pick what you work on, so you need to be able to come up with good ideas and then follow them through, just like an artist does. There's some performing and storytelling involved too if you want to do a good job of teaching or giving presentations of your work – otherwise your audience and students will fall asleep! Finally, you need to be very determined to get what you want and definitely not be too fazed by rejection. Applications for research money and attempts to publish your work get rejected all the time in academia – you just have to keep trying."