Our Data Capability – united we stand

**Blog: Published on ONS Intranet, 18 July, 2017**

Last week we had the 22nd annual Methodology Symposium. This blog is the summary of my presentation. The day was a great success and I want to thank all those who attended and presented. I was the last to speak so decided to tell a few stories and introduce a few of my data heroes. I have a growing list but chose the following people because they presented at TED, which means you can easily listen to them too.

But before I start, I want to remind you of Plato’s Allegory of the Cave.



Image: Wikimedia Commons

In the allegory, Plato describes prisoners who are shackled to a wall, unable to move their heads. Behind them are puppeteers and a source of light. All they can see in front of them are shadows and these shadows are their Truth. They know of nothing else. They may describe a shadow of a book as a book, but it is an inadequate representation of the real thing. It is only when they are ‘released’ – or educated – that they are able to see the light, and hence the truth.

The modern day is no different as the following picture beautifully highlights.



Our digital world has become our truth despite it being a poor representation of the world around us. Are our statistics, data and KPIs actually our shadows? Do they really represent our economy, society and environment adequately?

My data heroes



Image. Ted.com

My first data hero is Anne Milgram, Attorney General of New Jersey, 2007 to 2010. I admire her because she isn’t a data scientist or an analyst but because she used data to solve a problem. When she became District Attorney in New Jersey she asked a pertinent question which goes to the heart of the purpose of having a criminal justice service in the first place. She asked whether the system is making people safer. It turned out that there was no clear answer to the question. Various measures showed that people were incarcerated for low level crime or waiting for their day in court while more serious criminals were walking free. She observed that the officers working in the system – police officers, lawyers and judges were making the best decisions they could based on the individual cases in front of them and the law but the sum of those decisions wasn’t leading to a safer society.

Anne went on to develop an assessment tool that gave judges more information about the likelihood of offenders posing a risk. The assessments doesn’t replace judgment, it simply provides judges with more information to make better decisions.

My second data hero is Dave Troy from Baltimore.



Image: Peoplemaps.org

I don’t quite know what Dave’s official title is but his website is full of very interesting data. This particular example is a new way to map cities. It isn’t a geographical or a demographic map. Rather, it maps the social relationships of people living in a city, who they talk to about their interests and who they don’t talk to, on-line of course. He argues, convincingly in my opinion, that mapping networks in this way tells us more about diversity than our current proxies such as ethnicity and religion.

Dave mapped many cities and it is interesting to see how different they are. Disciplines such as computational sociology are interested in studying what makes a ‘healthy network’ which could revolutionise our approach to solving some of our more pertinent social issues.



Image: Stanford University

My third hero is Russ Altman. He is a biochemist and has studied an area that is very close to my heart. He demonstrated something which I suppose is obvious yet governed by trial and error. Medicines undergo a huge amount of rigorous testing, randomized control trials and retesting before they are marketed but they are not tested in combination with other medicines. In a specific example, he studied the effect of combining common blood pressure and depression medications on glucose levels and found that there is an increased risk which can lead to diabetes. Without this knowledge, the solution becomes the addition of a third medication.

Russ’s study was in the USA and was reliant on research databases and scraped data to find out if people who were taking these medicines were also searching symptoms of diabetes. That’s interesting, but we have the NHS! Imagine what could be done if we mined prescribing data. I’m especially interested in dementia and my hypothesis is that at least some of it can be explained by medication and treated by changing it.



Image: YouTube

My fourth and final hero for this blog is James Glattfelder. James is a Physicist and he studies complexity. After the financial crash he studied the complex network of board membership and power (measured by share ownership) in transatlantic companies and found that around 700 players controlled about 80% of the network. His work offers new means to understanding how stable or unstable financial systems are.

When I was working for Manchester City Council we took inspiration from these (and other) data leaders and applied their techniques to our Troubled Families programme. We first set up a database that linked and matched data from 18 different sources with no common key. The complex web of interrelationships between entities (people, places, services, interventions) highlighted that the situation is far more complex than it seemed. The system was used operationally by our workers on the front line as well as for analytics and insight. As the dataset grew over time we could apply data mining techniques. One of the most interesting was the use of cluster analysis which highlighted similarities between groupings of families that could respond differently to different interventions. This meant that instead of focusing on an over-simplistic, binary and therefore meaningless question – “does this work?” – we were able to demonstrate which services work for which groups, ensuring that we didn’t throw the baby out with the bathwater and continued to develop new services for those who needed them most. (The Deputy Chief Executive, Geoff Little, was an invaluable champion of this work. He is my hero too.)

So what does this mean for analysts and methodologists? To get closer to the truth we must embrace complexity and develop new tools that enable us to link data and analyse it. There are no single variables that really explain our society and our economy and there are certainly no silver bullets to solving our problems but data science offers us the tools to generate better statistics for better decisions by judges and social workers as well as politicians and regulators who make decisions that impact us all.

Sarah.