

Wut Iz Data Science?

[Cross-posted from ICanHazDataScience]



Hello. You're probably here because you're curious about data science. Or how and why it's relevant to [human development](#) - i.e. not software development, but work that "gives everyone the chance to lead full lives". Which is a very very broad concept, that includes the work done by big agencies like the UN, down to communities and individuals just trying to make a difference, or journalists writing stories about where the gaps and problems are. You might be here because you're one of these people, and want to know how this new "data science" and "big data" will change your work; you might be a volunteer coder building tools to help them, or someone I haven't anticipated yet. Whoever you are - welcome, and I hope this can be at least a little bit useful.

So. What is data science, and what's special about it when it's done to help human development?

If you look up "data science", you'll get several good papers on it:

- O'Reilly Radar "[What is data science](#)", @mikeloukides 2010
- Quora "[What is data science](#)", various, including @medriscoll, 2010
- Dataists, "[A taxonomy of data science](#)", @hmason and @chrishwiggins, 2010

And the distinct impression (if you don't read the articles above) that a) it's all about statistics and

coding, and b) there are a *lot* of pretty pictures involved.

Let's stop for a moment. Yes, there are lots of pictures. And yes, there's also statistics and coding. But that's not the heart of data science. I've also been a systems engineer for a long time now, and from what I can tell from data scientists' work and my own forays into it, is that a data scientist is one of those rare people who works at the boundary between engineering and business. What they effectively are, are translators - between the world of number and text crunching and the insights that drive business, progress, politics, stories. That's why there are so many pictures: it's not just because pictures are pretty, contain lots of information and draw in more people than numbers or words do (just look at any introductory course on giving a good presentation for some of the theory behind this), but it's because what's being created here is insight. Insight based not only on the numbers that a data scientist is given to work with, but all the other information available to them in the universe (internet, internal data stores and others) or obtained as part of their work (through FOIAs, direct investigation or discussion).

There have already been studies into the types of personality who make good data scientists (especially by people who, with the current shortage, are keen to develop their own), but we could also define this in terms of the company that they typically keep. The translator-personality goes some way to explaining the types of people that data scientists seem to hang out with - data geeks, architects, systems designers et al. The data gathering urge explains some of their connections to open data types, makers, and their recent interest in areas like the internet of things.