

DS4B: The Course [DS4B Session 1a]

[Cross-posted from [LinkedIn](#)]

I've designed and taught 4 university courses in the past 3 years: ICT for development and social change (DSC, with Eric from BlueRidge), coding for DSC, data science coding for DSC and data science for DSC (with Stefan from Sumall). The overarching theme of each of them was better access and understanding of technical tools for people who work in areas where internet is unavailable or unreliable, and where proprietary tools are prohibitively expensive.

The tagline for the non-coding courses was the "anti-bullsh*t courses": a place where students could learn what words like 'Drupal' and 'classifier' mean before they're trying to assess a technical proposal with those words in, but also a place to learn how to interact with data scientists and development teams, to learn their processes, language and needs, and ask the 'dumb questions' (even though there are no dumb questions) in a safe space.

Throw a rock on the internet and you'll hit a data science course. There are many of these, many covered by the open data science masters, and many places to learn the details of machine learning, R, business data science et al. What there isn't so much of on the internet is courses on how to specify a data science problem, how to interact with the people coming off those machine learning courses, or how to do data science work when you're days from the nearest stable internet connection. The latest course, Data Science for Development and Social Change, was designed for that. It changed a bit from its inception: we found ourselves heavily oversubscribed with students from not just International Development, but also Journalism, Medicine and other departments, and adjusted from a 12-person hands-on intensive lab format (e.g. the coding course included D3 down to the CSS, HTML, SVG and Javascript level) to a 30+ person lab lecture with ipython notebooks.

The slideset shared here is from the latest inception of the lab part of that course (the non-lab half showed social data science in different contexts): a set of sessions that started as a weekly one-hour chat between myself and the Quito ThoughtWorks team, and grew into a multi-city weekly data science session.

The way to approach these slides is to read the slide notes in each session's powerpoint file (you'll probably have to download the files to see this: I'll slowly upload the notes if needed), then try the exercises in the ipython notebooks for that session.

The [first session](#) is an introduction, a 'why are we doing this', with a bunch of downloads and setups (student- and developer- tested!) as an exercise. I'll write about that soon. The downloads and setups are integral to these sessions (to leave someone with enough tools, notes and information on their machines to be able to do basic data science, and understand what a data scientist does all day, even where there's no hope of stable Internet (yet)), but though the learning

OverCognition

Journeys through development data.

<http://overcognition.com>

is much better if you try the examples, the slides can be mostly followed without them.