Data Science Ethics [DS4B Session 1e]

This is what I usually refer to as the "Fear of God" section of the course...

Ethics

Most university research projects involving people (aka “human subjects”) have to write and adhere to an ethics statement, and adhere to an overarching ethics framework, e.g. “The University has an ethical commitment to minimize the risks to research subjects and to ensure that individuals who participate in research projects conducted under its auspices… do so voluntarily and with an informed understanding of what their involvement will mean”. Development data scientists are not generally subject to ethics reviews, but that doesn’t mean we shouldn’t also ask ourselves the hard questions about what we’re doing with our work, and the people that it might affect.

At a minimum, if you make data public, you have a responsibility, to the **best of your knowledge**, skills, and advice, to **do no harm to the people connected to that data**. Data science projects can be very powerful (especially if we’ve designed them well), development data science can affect many people, and we need to be mindful of who we’re affecting and the risks we might unintentionally cause them with our work. With that power comes responsibility, and a sometimes-difficult balance between making data available to people who can do good with it, and protecting that data’s subjects, sources, and managers.

I start by asking these two questions:

- Could this work increase risk to anyone?
- How will I respect privacy and security?

Risk

Risk is defined as “The probability of something happening multiplied by the resulting cost or benefit if it does”. There are three parts to this: cost/benefit (what might happen), probability (how likely that is to happen) and subject (who the risk is to). For example:

- Risk of: physical, legal, reputational, privacy harm
- Likelihood (e.g. low, medium, high)
- Risk to: data subjects, collectors, processors, releasers, users

So, basically, make a list of what bad things might happen, who to, how likely these bad things are, and what you should be doing to prevent or mitigate them, up to and including stopping the project or making sure that anyone at risk is aware of that risk and can consent to be subject to it. It doesn't have to be a down-to-the-tiniest-thing detailed list, but you do have to think about who could be harmed by your work and how.

And this isn't a one-time thing. New risks can occur as new data becomes available, or the original environment around your project changes: you need to be thinking about potential risks right from the planning phase of your project through to when you're sharing data or insights (and beyond, if stale data or old results in changed contexts could also cause harm).

Risks can be obvious (e.g. try not to get your sources or subjects targeted by making them easy to find), but they can also be subtle. Some of the more subtle ones include:

- Accidentally propagating inbuilt prejudices against minorities (see Cathy O'Neill's excellent work on risk-based sentencing models and upcoming book)
- Basing results on datasets with inbuilt biases (which is most datasets, which is why you should always be cynical about what is and isn't in your data: see e.g. Kate Crawford's work on hidden biases and AI's White Guy Problem)
- Explaining results using warped representations (e.g. hiding US city populations in a choropleth instead of using a cartogram)
- Falling foul of Simpson's Paradox, where two seemingly related datasets either have a common cause (e.g. XKCD's cross-correlation of a population map) or are accidentally related (and you weren't cynical enough about this – see Spurious Correlations for lots of these)
- Reporting critically important numbers without being cynical about their meaning (e.g. How bad data fed the Ebola epidemic)
- Accidentally making data subjects easy to identify.

You don't have to make this list on your own. Groups including the Responsible Data Forum have been doing great work on data risk (and I've been lucky to be part of that hivemind too): try reading those groups' articles to kickstart your thinking about your project's potential risks.

Privacy and Security

That last risk on the list (accidentally making data subjects easy to identify), is important. Privacy and security is a complex topic, but at a minimum you should be thinking about PII (Personally identifiable information). PII is this: "any data that could potentially identify a specific individual. Any information that can be used to distinguish one person from another and can be used for de-anonymizing anonymous data can be considered PII."

When you think about PII risk, think beyond name, address, phone number, social security number. Think about things like these (which can all accidentally release PII):

- Unique identifiers: Names, addresses, phone numbers etc
- Locations: lat/long, GIS traces, locality (e.g. home + work as an identifier)
- Members of small populations (e.g. there may be only a small number people fitting this profile in the given geographical area)

- Untranslated text (e.g. text that your team can't read and understand)
- Codes (e.g. “41”, especially if you don't have a codebook telling you what they mean)
- Slang terms
- Can be combined with other datasets to produce PII (aka reidentification)

There is much literature on how easy reidentification can be easy from very small amounts of data, so you will have to think of it also in terms of risk: how likely is it that someone will want to reindentify, how easy is it for them to do this, and what can you do to make it harder.

But Don't Panic

Don't panic. When you first start thinking about risk in data science, the usual reaction is an “OMG I can’t do anything without causing harm” one. That’s not the right answer either. Be aware of the risks. Take a deep breath, and remember that risk is cost times probability, and that if you give them the chance, people will often make a surprising but informed decision about their own risks.

Sometimes, the only answer is to walk away from the project. If it isn't, and risk is high, consider mitigations like releasing data and results to a smaller group of people (e.g. academics, direct responders, people in your organisation, data subjects) with the caveat that once you release, you no longer have control of that data and have implicitly trusted other people to do the right thing too. Consider releasing data at a different granularity, e.g. use town/district instead of street, and/or a subset or sample of the data 'rows' and 'columns'. Look in places like the RDF website to see what other people have done as mitigations.

Be aware of ethics. Be as honest and cynical about data and results as you can. And don’t start a development data science project without doing a basic risks check.

[Image by Steve Calcott, licensed under cc-by-nc 2.0]