

Why spreadsheets are hard

Thinking about machine-reading human-generated spreadsheets today, and I think I've got a handle on why this is a problem.

- Data nerds think of data in its back-end sense, of “what form of data do I need to be able to analyse/ visualise this”. We normalise, we worry about consistency, we clean out the formatting.
- People used to creating spreadsheets for other people think of it more in the front-end sense, of “how do I make this data easily comprehensible to someone looking at this”.

Each has merits/demerits (e.g. reading normalised data and seeing patterns in it can be hard for a human; reading human-formatted data is hard for machines) and part of our work as data nerds is working out how to bridge that divide. Which is going to take work in both directions, but it's necessary and important work to do.