Data Science Notes

Talks 2

Hilary’ Mason’s talk 2

Books 3

Thinking with Data 3

# Talks

## [Hilary’ Mason](https://www.youtube.com/watch?v=R-9fuo-Ri8E)

Successfully using data:

* + What problem are we really solving?
  + What does success look like? What is your error metric? How do we know when we’ve won?
  + What’s the first thing we’ll do with our data once we get it? Know what you want to do with it, don’t’ just drown in it

# Books

## Thinking with Data

Chapter 1 Scoping: why before how

* Picking the right techniques has to be secondary to asking the right questions.
* we need to give ourselves space to think.
* The secret is to have structure that you can think through, rather than working in a vacuum. Structure keeps us from doing the first things to cross our minds. Structure gives us room to think through all the aspects of a problem.
* Our first place to find structure is in creating the scope for a data problem. A scope is the outline of a story about why we are working on a problem (and about how we expect that story to end).
* When a problem is well-scoped, we will be able to easily converse about or write out our thoughts on each.
* A mnemonic for these four areas is CoNVO: *co*ntext, *n*eed, *v*ision, *o*utcome.
* We should be able to hold a conversation with an intelligent stranger about the project, and afterward he should understand (at a high level)
* **Context**
* Who are the people with an interest in the results of this project? What are they generally trying to achieve? What work, generally, is the project going to be furthering?
* Example: “This department in a large company handles marketing for a shoe manufacturer with a large online presence. The department’s goal is to convince new customers to try its shoes and to convince existing customers to return again. The final decision maker is the VP of Marketing.”
* Contexts emerge from understanding **who we are working** with and **why they are doing what they are doing.**
* We learn the context from **talking to people**, and continuing to talk to them until we understand what **their long-term goals** are.
* The work we do should further the mission espoused in the contexts.
* **Needs**
* What are the specific needs that could be fixed by intelligently using data? These needs are meaningful to the organization.
* **The need is to solve the problem that having the model will solve.**
* **When we correctly explain a need, we are clearly laying out what it is that could be improved by better knowledge.**
* What will this spreadsheet teach us? What will the tool let us know? What will we be able to do after making this graph that we could not do before?
* We want to reduce the amount of illegal grease dumping in the sewers. Where might we look to find the perpetrators?
* Again, **writing** is a big help here. write down what we think the need is
* A need is not merely an open question or problem, but the final result/action desired. Eg:
* Rather than: “The manager wants to know where users drop out on the way to buying something”
* Say: “The manager wants more users to finish their purchases. How do we encourage that?”
* This version is action oriented and invites more open ended possibilities
* The need is NOT the model – the business problem needing to be solved/achieved
* **A data science need is a problem that can be solved with knowledge, not a lack of a particular tool.**
* **Vision**
* The ideas we will be able to come up with will mostly be variations on things that we have seen before.
* acquire a good mental library of examples: read widely and experimenting with new ideas, talk to people about the problems they’ve solved,
* two main tactics:
* mock ups
* low grade / detail idealization of the the form of the final results … could be graph or sentences…but it presents the type of knowledge we hope to get out of the project
* argument sketches
* tells us roughly what we need to do to be convincing at all. It is a loose outline of the statements that will make our work relevant and correct.
* A lot like what we do when writing a grant.. give us a sense of logic behind the solution