DIY Codebook

A skeptic is someone who maintains a consistently inquisitive attitude toward facts, opinions, or (especially) beliefs stated as facts. A skeptic asks questions when confronted with a claim that has been taken for granted. That's not to say a skeptic brow-beats someone for their beliefs, but rather that they set up reasonable experiments to test those beliefs.

Cathy O'Neil
On Being a Data Skeptic

Codebook:

- describes what's in a collection of data

It's easy to open a spreadsheet and make assumptions about what's there.

That will bite you in the ass.

Codebook:

- describes what's in a collection of data
- "codes": The name/number assigned to each variable

Codebook:

- describes what's in a collection of data

- "codes": The name/number assigned to each variable

- layout: describes how the data is organized

- What was the methodology used to gather this data?

(Who, What, Where, When, How)

- Has the methodology changed?

- Who provided the data?
- Was it all collected by the agency/company/nonprofit that provided it?

- Can you talk to the person who did the survey, compiled the data, retrieved the data for you?
- Ask them if you're doing it right.

- Ask your data stupid questions!
- How many records are in the data set?

- What information is unreliable?

- Who paid for it?

And in 2015, it will only get worse — because I'd bet the big PR firms have noticed the success of some of these smaller efforts and will try their hands at this new form of marketing. Don't be surprised when Kraft creates a map of which states consume the most macaroni and cheese, or Starbucks releases charts showing how pumpkin spice-related products lift the American economy each fall. **The wave of bullshit data is rising, and now it's our turn to figure out how not to get swept away.** Maybe Snopes sells life rafts?

Jacob Harris
The New York Times.