

What is a “data state of mind”? And how you can develop it.

By: MaryJo Webster

Journalists just learning how to analyze and visualize data always ask me how to incorporate their new skills into their everyday work.

I’ve learned over the years that, in addition to learning how to punch the buttons in their software of choice and understanding some basic analysis concepts, they also need to train their brain to have a “data state of mind.”

To me, that’s the biggest key to making data a staple in your reporting. And this is something you can start doing right away, even if you barely know your way around a spreadsheet.

“Data state of mind” is a phrase that I’ve sort of stolen, and slightly adapted, from the “documents state of mind” that great investigative journalists Donald Bartlett and James Steele regularly promoted at Investigative Reporters and Editors conferences in years past.

The modern twist is that instead of constantly being on the lookout for documents (i.e. pieces of paper) that will support or prove your key point, you should be looking for data stored in computers. Usually that’s going to be data that you analyze in order to quantify or measure something or otherwise prove your point.

The common thread between the two concepts is that you’re not relying solely (or even primarily) on people to tell you what’s going on. Documents and data tend to have a better memory and be more honest.

You’ll also usually get a much better story.

Let me give you a few examples of a “data state of mind” in action, and then I’ll give you some tips on how to train your brain.

- A colleague of mine got a press release a few years ago about an increase in children being treated in U.S. hospitals for medication overdoses, so he decided to see if that held true in Minnesota. Instead of the usual approach of calling experts, he first turned to the state’s death certificate database to find poisoning-related deaths. He didn’t see the trend he was looking for, but found an even better one. It wasn’t primarily children dying from over-the-counter medication overdoses, it was young adults. When he started calling state and federal experts, they told him this was a topic that had just started hitting their radar. It was too new for a press release.
- A while back, a suburb of St. Paul, Minnesota, was considering a new law that would set prohibitions on where high-level sex offenders could live. The reporter had done a simple,

traditional story saying the council was going to discuss this at an upcoming meeting. I suggested that we try to measure what might happen if they passed this proposed law. The big question: Would there be anywhere for the sex offenders to live in this small suburb? Using mapping software, we plotted locations of the places the sex offenders would have to stay a certain distance away from under this new law – schools, churches and parks being the big ones – and added buffers that represented the prohibited areas. The resulting map showed the prohibited areas covered almost the entire city. A few days after our story ran, the council dropped the proposal.

- Another reporter wanted to know how often people were criminally charged (and convicted) for killing a pedestrian in a traffic accident. It didn't take long for him to learn that nobody in Minnesota tracked this. There weren't reports out there. There wasn't even a database. The so-called experts had no answer to his question. But the pieces of data existed. The state compiles a crash database of traffic accidents, including all that result in injury or death; however, it doesn't have any names or very much identifying information about the crash. There is also a death certificate database that includes date of injury/death and very detailed information about where and how the person died. Since pedestrian deaths are so rare, it was easy to match the two datasets to find the victims. The crash data also included the official case number with the police agency that handled the crash, which led him to reports with victims' names and made it possible to search court records to look for charges and/or convictions. In the end, he essentially built his own database and told a one-of-a-kind story. (And the answer is that the drivers are rarely charged)

Hopefully you noticed that all three of those examples don't rise to the level of a "project" and all came out of beat reporting. All were done within a matter of days, maybe weeks. All of them also benefited from the fact that somebody in the newsroom knew that these databases existed, which helped launch the idea in the first place.

And most importantly, all of them had something they were trying to measure: Are more young people dying from medication overdoses? How much of the city would be left for sex offenders to live in? How often are people convicted for killing pedestrians?

Let's turn to those tips for developing a "data state of mind."

The first place to start is to frame your story ideas as questions, not statements. So for example, instead of saying you want to do a story about "unsafe bridges," flip it to a question like this: "What percentage of bridges in the state are unsafe?" This frames your story into something quantifiable and gives you a bit of a roadmap about the data you need (something that measures bridge safety), the universe of data you need (all the bridges in the state) and what analysis you'll need to do (percentage of the total).

Another very basic thing you can start doing, even if your data skills are not very far along, is to think about data in the same way you think about your human sources. After all, both types of sources have a lot in common: Both can answer questions, raise questions, and point you in the right direction (or wrong direction). If you misstate your question, both are apt to give you an unexpected answer. Both can only tell you about things they know about. Some sources are good as tipsters, but

not for officially quoting; others are your go-to sources over and over, and a few are immensely flawed.

I should also point out my favorite way that they are different: Data will let you sit there and ask it questions for hours on end and not kick you out.

You can start by figuring out what data is on your beat and then spending some time with that data. Think of it like you're taking a source out for coffee. Your editor insists that you need to get to know your sources, right? Then it's worth whatever time it takes.

Here are a few ways to find data on your beat:

- When interviewing sources, listen for any numbers or vague references to numbers (i.e. they tell you that "crime is up"). Ask where they got that information and be sure to get a specific enough answer so that you know how to get that data yourself.
- Ask the agency you cover if they keep a list of all their databases. Look for forms that the public might fill out (applying for a permit, registering a gun, requesting a housing code variance, etc) and then ask the agency what happens to the information from those forms.
- Figure out what things the agency is required to track. Of course, money will always be at the top of that list. But agencies always have some core function; ask them to explain to you how their tracking system works. Scour whatever reports the agency publishes for summary tables and references to the sources they used.

Knowing all this really comes in handy when story ideas start bubbling up. If you know an agency has a database about something, you're far more likely to think of an idea where it could be useful.

The next thing you can start doing is tuning your radar to pick up on opportunities to use data. There are certain story types that almost always benefit from data.

The most obvious are trend stories.

For example, let's say you're hearing a lot of concern about a rise in burglaries. A traditional reporter would call up the police department's spokesperson and ask them for some quotes, maybe a few summary numbers. A "data state of mind" reporter would file a public records request for all the incidents categorized as burglaries over some recent time period (perhaps 6 months?) and then also ask for comparable data from the same 6-month span in the previous year (or multiple years). He or she would also interview the police media contact person, but the questions would try to focus in on things like whether or not the department has done its own analysis (if so, what did they find?), and other details that might help inform the reporter's own analysis. A lovely perk to getting the raw data is that you'll also get addresses, and perhaps names, of the people whose homes were burglarized.

Other good data opportunities:

- Anytime a government agency has created a program to do something, go back a year or more later and measure whether it has succeeded.

- Breaking news stories: How often has this occurred in the past? This is really useful for putting a big news event in context.
- The rumor or myth that is always circulating. Can you figure out whether it's true? I think these can be gold mines because we often just take them for granted. What if you could prove that it's not true? That would be a popular story, for sure.
- I also think it would be worth trolling back through stories you did a year or more ago to either find missed opportunities or stories that are ready for a follow up. Is there a broader or deeper story that needs to be done?
- Finally, take a moment to realize that you are probably already asking for "data" but just not the right stuff. Most reporters routinely ask for what is considered summary data, and also get reams of it via press releases, reports and websites.

This might be a table showing the total number of crimes reported each of the past few years. Or it might be one big number, such as the estimated number of homeless people in the city. Or maybe two big numbers showing year-over-year change.

When you find yourself asking a source to give you summary numbers for a story, stop and think whether it would be better to get the data that those numbers are based on.

For example, if you are writing a story about how overtime has caused the police department's budget to go through the roof, think about whether detailed data showing how much overtime each officer used might be more insightful than merely the big dollar values from one year to the next. Or what if you knew how much overtime was generated each week or each pay period and could use that to pinpoint a particular event that resulted in a lot of overtime?

I also find a lot of reporters will ask for "a list" of something and not realize that they're missing out on opportunities. One reporter asked for a list of all the car accidents where distracted driving was involved. She got a spreadsheet with two fields: a date and the county where it occurred. Nothing else. But if she had asked for all the fields of information that were tracked for each accident she would've gotten dozens of fields that tell you really interesting related pieces like age of the driver, whether anyone was hurt or killed, whether other cars were involved, and much more.

The more data you have, both in terms of rows and columns, the more questions you can ask it and the more answers you'll get.

In my experience, I've found that reporters who adapt this kind of "data state of mind" are far more likely to keep using, and expanding, their data skills. Coming up with great story ideas that require data is a remarkable motivator.

Original URL:

http://datadrivenjournalism.net/news_and_analysis/what_is_a_data_state_of_mind_and_how_you_c_an_develop_it