



10 Kinds of Stories to Tell with Data

BY TOM DAVENPORT

For almost a decade I have heard that good quantitative analysts can “tell a story with data.” Narrative is—along with visual analytics—an important way to communicate analytical results to non-analytical people. Very few people would question the value of such stories, but just knowing that they work is not much help to anyone trying to master the art of analytical storytelling. What’s needed is a framework for understanding the different kinds of stories that data and analytics can tell. If you don’t know what kind of story you want to tell, you probably won’t tell a good one.

This insight came in an interview a couple of years ago with Joe Megibow, a leading analytics practitioner who was head of web analytics at Expedia and is now senior vice president of omnichannel e-commerce at American Eagle. We simultaneously realized that there are several different types of analytical stories, and that it might be useful to create a typology of them. I later created what I think is the first typology of analytical stories in my book (with Jinho Kim) *Keeping Up with the Quants* and, since its publication last year, I’ve refined the typology further. Practically speaking, there are four key dimensions that determine the type of story you can tell with data and analytics:

Time

Analytical stories can be about the past, present, or future. The most common type of analytical story is about the past—it’s a **reporting** story using descriptive analytics to tell what happened last week, month, quarter, or year. By the way, most visual analytics stories are also of this type. They’re not the most valuable form of story, but it’s undeniably useful to know what happened.

Stories about the present are most likely to involve some form of **survey**—an analysis of what people or objects are currently up to. It may actually involve survey research—asking people what they think about something. In some cases survey analysis involves a statistical model of what factors drive others. We might call those **explanatory survey** stories. In general there are lots of minor variations on the survey story. In the book we talk about social science surveys, surveys of cable TV viewers, and surveys of bombers in World War II.

Stories about the future are **predictions**; they use, of course, predictive analytics. They take data from the past (it’s hard to get data from the future!) to create a statistical model, which is then used to predict the future. Quants create prediction stories all the time—about what customers are likely to buy, about how likely it is for an event to happen, about future economic conditions. These types of prediction stories always involve assumptions (notably that the future will be like the past in some key respects) and probability. The good news is that we can specify the likelihood that the story will be true; wouldn’t that have been nice for fairy tales?

[continued]

ABOUT THE AUTHOR

Thomas H. Davenport is the President’s Distinguished Professor of IT and Management at Babson College, a research fellow at the MIT Center for Digital Business, co-founder of the International Institute for Analytics, and a senior advisor to Deloitte Analytics. He is author of the new book *Big Data at Work* and the best-selling *Competing on Analytics*.

Focus

Are you trying to tell a **what** story, a **why** story, or a **how to address the issue** story? (I am thankful to several Procter & Gamble executives for pointing this out). *What* stories are like reporting stories—they simply tell what happened. *Why* stories go into the underlying factors that caused the outcome. *How to address the issue stories* explore various ways to improve the situation identified in the what and the why stories. A really complete story may have all of these focus elements. P&G has made considerable progress in getting agreement on the *what* story quickly, and then spending more time and energy on the *why* and *how* topics. Key to doing that is having all parties involved in telling the story working off the same data.

Depth

There is also a depth dimension to analytical stories. When I spoke with Joe Megibow, then at Expedia, about this, he said that many of their stories were “CSI” projects—relatively small, ad hoc investigations to find out why something suboptimal was happening. One of his favorite examples involved discovering why some Irish customers were dropping online transactions when they got to the postal code input form. It turned out that some rural Irish locations don’t have postal codes. Just like on CSI—story solved in a short time.

The alternative I call “**Eureka**” stories, which involve long, analytically-driven searches for a solution to a complex problem. When you solve it, you want to yell, “Eureka!” My primary example in the book involved discovering the right way to refer and price potential buyers to real estate agents at Zillow. The project was core to the company’s business model and was worth a long story; after a few false starts and the use of several different analytical methods, they got it right. These types of stories are typically long, important, and expensive, so getting stakeholder buy-in is critical if you plan to reach the end of them.

Methods

Finally, there are different types of stories based on the analytical method used. Are you trying to tell, for example, a **correlation** story—in which the relationships among variables rose or fell at the same time—or a **causation** story, in which you’ll argue that one variable caused the other? In most cases, doing some sort of controlled experiment is really the only way to establish causation. People—particularly those in the media—tell bad stories all the time because they confuse causation with correlation.

These ten kinds of stories are not mutually exclusive. There are certainly other method-based stories, and probably other important dimensions as well. But knowing that there are at least ten ways to tell analytical stories is much more useful than knowing only that you should tell one. There are other important aspects of analytical storytelling as well, such as that the story told to businesspeople should generally begin with the result and recommended outcome. You can save the details of how you got there analytically for the footnote of your report or presentation. And terms like R2, coefficient, logistic, and heteroskedasticity should not be appearing in your public story at all!

The good news is that we can specify the likelihood that the story will be true; wouldn’t that have been nice for fairy tales?