

# How to Tell a Story with Data

by Jim Stikeleather

APRIL 24, 2013

An excellent visualization, according to Edward Tufte, expresses “complex ideas communicated with clarity, precision and efficiency.” I would add that an excellent visualization also tells a story through the graphical depiction of statistical information. As I discussed in an earlier post, visualization in its educational or confirmational role is really a dynamic form of persuasion. Few forms of communication are as persuasive as a compelling narrative. To this end, the visualization needs to tell a story to the audience. Storytelling helps the viewer gain insight from the data. (For a great example, how much do you think steroids have influenced baseball?)

So how does a visual designer tell a story with a visualization? The analysis has to find the story that the data supports. Traditional journalism does this all the time, and journalists have become very good at storytelling with visualization via infographics. In that vein, here are some journalistic strategies on telling a good story that apply to data visualizations as well.

1. **Find the compelling narrative.** Along with giving an account of the facts and establishing the connections between them, don’t be boring. You are competing for the viewer’s time and attention, so make sure the narrative has a hook, momentum, or a captivating purpose. Finding the narrative structure will help you decide whether you actually have a story to tell. If you don’t, then perhaps this visualization should support exploratory data analysis (EDA) rather than convey information. However, for the designer of an

exploratory visualization it is still important to spark the viewers' imagination to encourage examining relationships among and facilitate interacting with the data - think gameification.

2. **Think about your audience.** What does the audience know about the topic? Is it meant for decision makers, general interested parties, or others? The visualization needs to be framed around the level of information the audience already has, correct and incorrect:
  - Novice: first exposure to the subject, but doesn't want oversimplification
  - Generalist: aware of the topic, but looking for an overview understanding and major themes
  - Managerial: in-depth, actionable understanding of intricacies and interrelationships with access to detail
  - Expert: more exploration and discovery and less storytelling with great detail
  - Executive: only has time to glean the significance and conclusions of weighted probabilities
3. **Be objective and offer balance.** A visualization should be devoid of bias. Even if it is arguing to influence, it should be based upon what the data says-not what you want it to say. Tufte found numerous charts that misled viewers about the underlying data, and created a formula to quantify such a misleading graphic called the "Lie Factor." The Lie Factor is equivalent to the size of the effect shown in the graphic, divided by the size of the effect in the data. Sometimes it is unintentional-a number that is three times bigger than another will be perceived nine times bigger if represented in 3D. There are simple ways to encourage objectivity: labeling to avoid ambiguity, have graphic dimensions match data dimensions, using standardized units, and keeping design elements from compromising the data. Balance can come from alternative representations (multiple clustering's; confidence intervals instead of lines; changing timelines; alternative color palettes and assignments; variable scaling) of the data in the same visualization. Maintaining objectivity and balance is not a trivial effort and is easily unintentionally

- violated. Viewers and decision makers will eventually sniff out inconsistencies which in turn will cause the designer to lose trust and credibility, no matter how good the story.
4. **Don't Censor.** Don't be selective about the data you include or exclude, unless you're confident you're giving your audience the best representation of what the data "says". This selectivity includes using discrete values when the data is continuous; how you deal with missing, outlier and out of range values; arbitrary temporal ranges; capped values, volumes, ranges, and intervals. Viewers will eventually figure that out and lose trust in the visualization (and any others you might produce).
  5. **Finally, Edit, Edit, Edit.** Also, take care to really try to explain the data, not just decorate it. Don't fall into "it looks cool" trap, when it might not be the best way explain the data. As journalists and writers know, if you are spending more time editing and improving your visualization than creating it, you are probably doing something right.

---

**Jim Stikeleather** serves as Executive Strategist, Innovation for Dell Services, the IT services arm of Dell, where he leads a team of information technology and business experts who identify, evaluate and assess the potential of new technologies, business models and processes to address evolving business, economic and social trends.

---

## Comments

Leave a Comment

POST

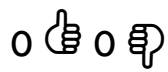
2 COMMENTS

---

Hayk Antonyan    a year ago

Thank you for this article.

REPLY



✓ **JOIN THE CONVERSATION**

---

#### POSTING GUIDELINES

We hope the conversations that take place on HBR.org will be energetic, constructive, and thought-provoking. To comment, readers must sign in or register. And to ensure the quality of the discussion, our moderating team will review all comments and may edit them for clarity, length, and relevance. Comments that are overly promotional, mean-spirited, or off-topic may be deleted per the moderators' judgment. All postings become the property of Harvard Business Publishing.