the truthful art

data, charts, and maps for communication

alberto cairo

The Truthful Art:
Data, Charts, and Maps for Communication

Alberto Cairo

New Riders

Part II: Truthful

3. The Truth Continuum

> To believe that truth is true is not due to daring, but due to humility: the honest man does not think he gets a veto over reality.
>
> **—John C. Wright.** *Transhuman and Subhuman: Essays on Science Fiction and Awful Truth*

I'll begin with an axiom:

Any visualization is a model.

Think of a locator map. A map is always a simplified depiction of a particular area. It doesn't look exactly like the area itself. Mapmakers remove needless features and emphasize the ones that matter to them in a process of rational and systematic abstraction. Road maps, for instance, highlight roads, cities, towns, and boundaries. Their function is to help you find your way around, not to display every single mountain, valley, or river.

The idea of model can be extended to any act of thinking and communication. We humans use models for perception, cognition, and reasoning because our limited brains are incapable of grasping reality in all its glorious complexity. Our senses and brain mediate our relationship with the world. Our vision doesn't consist of a high-resolution animation of what we have in front of our eyes. That's just a convenient illusion that our brains concoct.[1]

> [1] For more details about how and why the brain does this, see my previous book, *The Functional Art*. The stance I am adopting in this chapter is based on a philosophy of science called "model-dependent realism," outlined in Stephen Hawking's and Leonard Mlodinow's book *The Grand Design*: "There is no picture- or theory-independent concept of reality. Instead we will adopt a view that we will call model-dependent realism: the idea that a physical theory or world picture is a model (…) and a set of rules that connect the elements of the model to observations. We make models in science, but we also make them in everyday life. Model-dependent realism applies not only to scientific models but also to the conscious and subconscious mental models we all create in order to interpret and understand the everyday world. There is no way to remove the observer—us—from our perception of the world, which is created through our sensory processing and through the way we think and reason. Our perception—and hence the observations upon which our theories are based—is not direct, but rather is shaped by a kind of lens, the interpretive structure of our human brains."

A model is a sign—or a set of signs and their relations—that describes, explains, or predicts something about how nature works with a variable degree of accuracy. Good models abstract reality while keeping its essence at the same time.

Numerical thinking and communication are also based on models. Think of a statistic like the average. If I tell you that the average height of adult U.S. females is 63.8 inches, I am giving you a model that is intended to summarize *all* heights of *all* adult women in the United States. It's not a perfect model, just an approximation. It can be a fair one if a majority of female heights are close to this average.

If I showed you the entire data set of roughly 158 million women's height records, you wouldn't be able to understand anything. It would be too much information for a single human brain to compute. It is chimerical to think that we can make *perfect* models to observe, analyze, and represent reality. Designing incomplete but still informative ones is the most we can achieve.

Therefore, my axiom has an important coda:

The more adequately a model fits whatever it stands for without being needlessly complex, *and* the easier it is for its intended audience to interpret it correctly, the better it will be.

Some models, therefore, are better than others. "Better" means, in this context, more truthful, accurate, informative, and understandable.

Dubious Models

For **Figure 3.1**, I used data from the National Cable & Telecommunications Association (NCTA) mentioned in Chapter 2. The subtitle the NCTA originally wrote, which I'm reproducing verbatim, makes its message explicit, in case you had any doubts: broadband cable connections in the United States have improved quite a lot in the past few years. But have they?
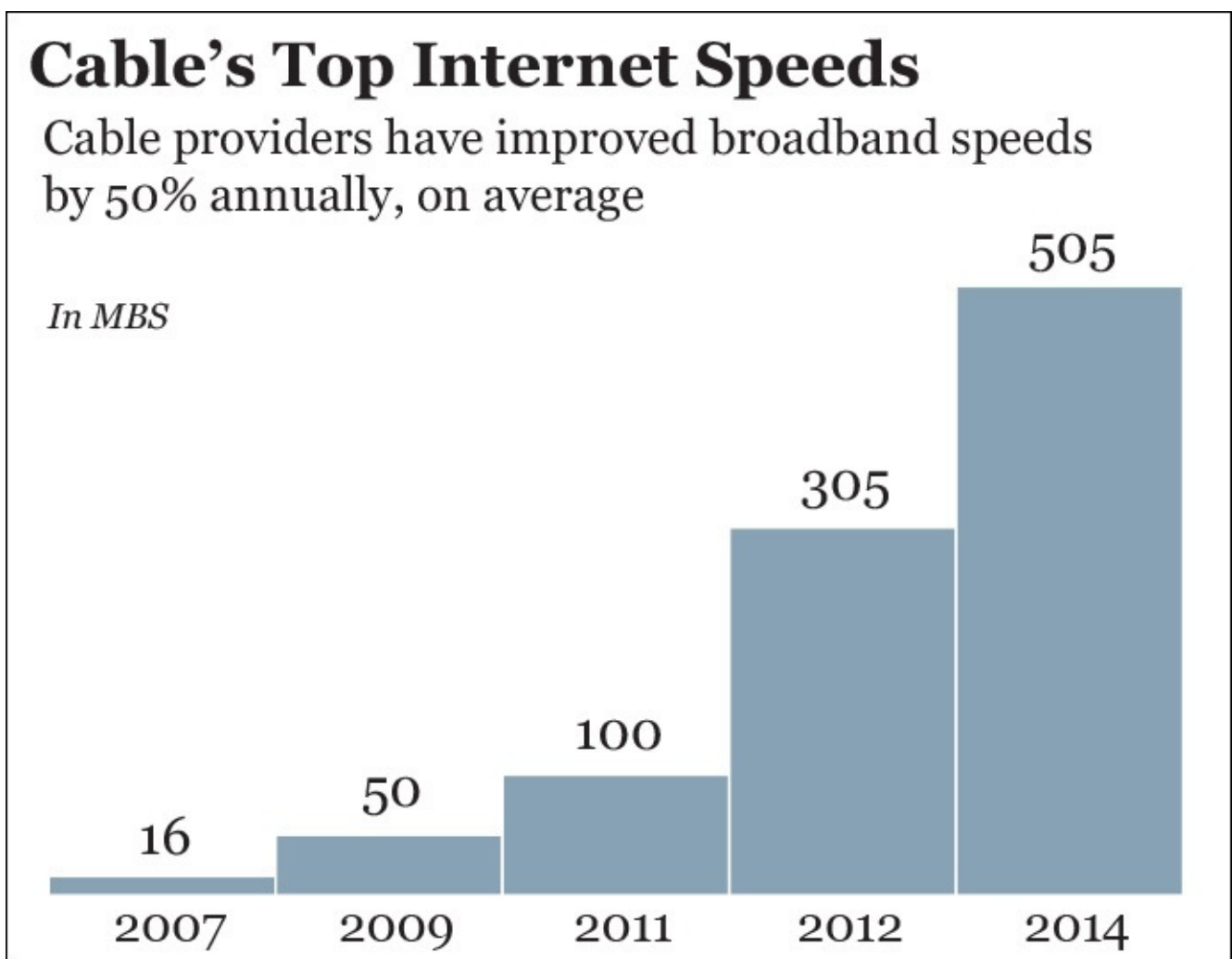
Figure 3.1 Misleading your audience may yield benefits in the short term. In the long term, however, it may destroy your credibility.

This isn't really a satisfactory model if we wish to understand the overall growth of Internet access speeds. This graphic is showing just top speeds! We are missing important information. How about minimum speeds, average speeds, and, more importantly, the number of people who have access to those top, average, and minimum speeds? That would make the model—the chart—more truthful. By the way, it'd also be necessary to

show all years, rather than what seem to be carefully picked ones.

<center>▎▎▊▊▍▎▊▎▎</center>

Some people devise bad visual models on purpose, to mislead their audience, but more often a faulty model is the result of a well-intentioned designer not paying proper attention to the data. **Figure 3.2** is a chart of traffic fatalities between 1975 and 2012. Consider it for a minute. I live in Florida, so I find it worrying that road deaths have increased so much, while most other states have improved noticeably. But is this what really happened?

Percent change in traffic fatalities 1975-2012
TOTAL FATALITIES

| | -75% | -50% | -25% | 0 | +25% |

DC
Massachusetts
Illinois
New York
Minnesota
Alaska
Michigan
Vermont
Iowa
New Jersey
Nebraska
Rhode Island
Wyoming
Washington
Oregon
Connecticut
Pennsylvania
Ohio
Idaho
New Mexico
Wisconsin
South Dakota
Indiana
California
Montana
New Hampshire
West Virginia
Maine
Maryland
Louisiana
Virginia
Missouri
Kansas
Utah
Colorado
North Carolina
Kentucky
Hawaii
Georgia
Tennessee
Delaware
Oklahoma
Alabama
Arkansas
Texas
North Dakota
South Carolina
Mississippi
Nevada
Florida
Arizona

Figure 3.2 Source: http://graphzoo.tumblr.com/post/85330752462/data-source-httpwwwwnhtsagov-code.

An analogy may come in handy: let's say you design a graphic about motor vehicle accidents that compares Chicago, Ill. (population of 2.7 million), to Lincoln, Neb. (population of 269,000). Would that be fair? Not entirely. The absolute numbers—the actual case count—are relevant, but you'd also need to apply some control to them, generating a *relative* variable. In the case of Figure 3.2, it would be the ratio between accidents and population or number of vehicles.

(Fatalities/Vehicles) × 100,000 = Rate of fatalities per 100,000 vehicles

We should apply that formula to both the 1975 and 2012 figures and then calculate the difference. It may well be that some states are experiencing a large drop in fatalities just because their populations have not changed that much and because roads and cars have become safer.

We can go even further. An exploration of this topic could take into account not just the number of vehicles but also how many miles those vehicles are driven, as commute distances in the United States vary a lot. You can see the results in **Figure 3.3**. I highlighted Florida in both charts. Also, notice North Dakota. I wonder if those figures are related to the oil boom in the late 2000s. According to the Pew Research Center, North Dakota added nearly 100,000 workers between 2009 and 2014 alone, and commutes are long.[2]

2 "How North Dakota's 'man rush' compares with past population booms." http://www.pewresearch.org/fact-tank/2014/07/16/how-north-dakotas-man-rush-compares-with-past-population-booms/

Change in traffic fatalities 1975-2012
PER 100,000 VEHICLES

-50	-25	0

Wyoming
New Mexico
Alaska
Vermont
Nevada
Montana
Idaho
Arizona
South Carolina
Georgia
Oregon
Louisiana
North Carolina
Texas
Tennessee
South Dakota
West Virginia
Wisconsin
Delaware
Illinois
Washington
District of Columbia
Missouri
Arkansas
New Hampshire
Massachusetts
Utah
Minnesota
Iowa
Indiana
New York
Maine
Nebraska
Virginia
Colorado
Michigan
FLORIDA
Hawaii
Kentucky
Alabama
California
New Jersey
Pennsylvania
Maryland
Oklahoma
Ohio
Mississippi
Rhode Island
Kansas
Connecticut
North Dakota

Change in traffic fatalities 2000-2012
PER MILLION ANNUAL VEHICLE-MILES*

-0.012	-0.006	0	0.006

Mississippi
Alaska
District of Columbia
Idaho
Utah
Nevada
Louisiana
Arizona
FLORIDA
Montana
Colorado
South Dakota
Arkansas
South Carolina
Tennessee
Wyoming
North Carolina
Missouri
Minnesota
Illinois
New Mexico
Nebraska
Alabama
Michigan
Washington
West Virginia
Georgia
Connecticut
Wisconsin
Iowa
California
Kansas
Hawaii
Texas
New Jersey
Ohio
Virginia
Maryland
Oregon
Indiana
Delaware
New York
New Hampshire
Massachusetts
Kentucky
Pennsylvania
Rhode Island
Vermont
Maine
Oklahoma
North Dakota

*Notice that on this chart I am using a different time range. I only found data going back to 2000

Figure 3.3 When we apply some controls, the results are quite different.

To make our visualization even better, we may want to apply more controls, like the effect of traffic regulations. For instance, seat belt laws in the United States began to be adopted in the mid-1980s. By the end of that decade, many states still didn't mandate seat belts for adults. New Hampshire still doesn't. Besides, in 16 states you won't get a ticket for not using your seat belt unless you're pulled over for another infraction. According to the Centers for Disease Control and Prevention, "Seat belts reduce serious crash-related injuries and deaths by about half."[3]

[3] I am taking this statistic at face value, but you may want to check it: "Seat Belts: Get the Facts." http://www.cdc.gov/motorvehiclesafety/seatbelts/facts.html

In 2010, *Wired* magazine proclaimed "The Web Is Dead. Long Live the Internet" on its front cover.[4] The point of the story was that Web browsers were in decline as a means of accessing data from the Internet. Other technologies, such as apps or video streaming, were on the rise.

[4] "The Web Is Dead. Long Live the Internet." http://www.wired.com/2010/08/ff_webrip/all/

Readers of that article were greeted by a stacked area chart similar to the first one in **Figure 3.4**. Web browsers were born in the early 1990s, saw their peak in 2000, when they accounted for more than 50 percent of Internet traffic, and declined later. In 2010, just 23 percent of the data downloaded from the Internet were accessed through a browser.

Internet Traffic in the U.S.

67% Other means of accessing data from the Internet

23% Web browsers

1990 — 2000 — 2010

Internet Traffic in the U.S.

7 — *Exabytes (1 exabyte = 10 million terabytes)*

Other means of accessing data from the Internet

Web browsers

1990 — 2000 — 2010
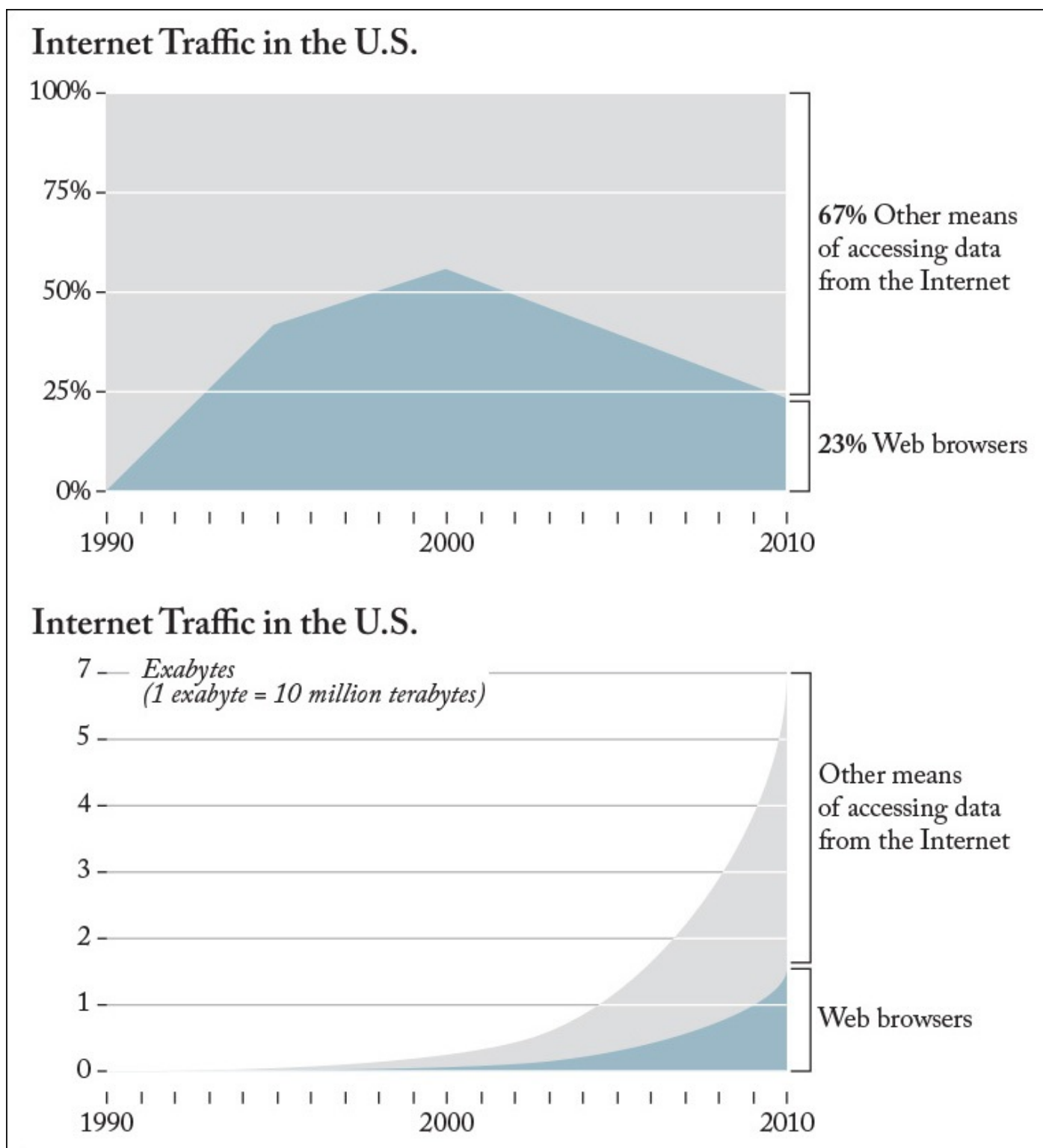
Figure 3.4 Two versions of the same story. (Sources: *Wired* magazine and [BoingBoing.net](#).)

Many of the problems with this chart—and with the story as a whole—were spotted by **Rob Beschizza**, managing editor of BoingBoing.net.[5] In the mid-90s, the number of Internet users was counted in *millions*. By 2010, it was counted in *billions*. Moreover, the kind of content people download has changed quite a lot. In the '90s, we mostly consumed text and low-resolution pictures. In 2010, video and file-sharing services were widespread. Many more people with much better connections were downloading much larger files. Those facts are reflected in the second chart in Figure 3.4, which is based on the one that Beschizza designed. The Web is hardly dying, after all.

5 "Is the web really dead?" http://boingboing.net/2010/08/17/is-the-web-really-de.html

The Mind, That Clumsy Modeler

A visualization is a model that serves as a conduit between a mental model in the designer's brain and a mental model inside the audience's brains.[6] Blunders, therefore, can originate on the designer's side but also on the reader's side. I've learned this the hard way, by being misled by wonderful and accurate visualizations numerous times.

6 Designer and audience can be the same individual, of course. In some cases, you design a visualization to improve your mental models, to understand something better.

Every year I travel to Kiev, Ukraine, to teach a one-week introduction to visualization workshop. As I write this, Ukraine is suffering the consequences of a chain of destabilizing events: a pro-Russian president was expelled by popular protests, a pro-Western government was democratically elected, Russia then annexed the region of Crimea, and, as of this writing, Russia's president, Vladimir Putin, continues supporting separatist militia groups in the eastern part of the country.[7]

7 BBC News has a good recap of the crisis at http://www.bbc.com/news/world-europe-25182823.

None of this had happened in 2012, when I flew to Kiev for my regular gig. Before my trip I had been in touch with **Anatoly Bondarenko**, a programmer, journalist, and visualization designer who works for an online media organization called Texty (http://texty.org.ua).

I met Anatoly in person, and he showed me his work. I was enthralled by one map in particular (**Figure 3.5**), which displayed the results of the 2012 parliamentary elections. Orange circles correspond to districts won by pro-Western parties. Blue circles identify districts where pro-Russian parties got a majority of the vote. In both cases, the intensity of the color is proportional to the percentage of the vote obtained by the winning parties. The size of each circle is proportional to the size of the voting population.

Обрано 33,646 з 33,646 виборчих дільниць.

~72 мандатів для ПР
~31 мандатів для КПУ
~61 мандатів для "Батьківщини"
~33 мандатів для УДАРу
~25 мандатів для Свободи

Недійсні

0.00 0.01 0.02 0.03 0.04 0.05 0.06 0.07

Явка з 16-00 до 20-00

0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40
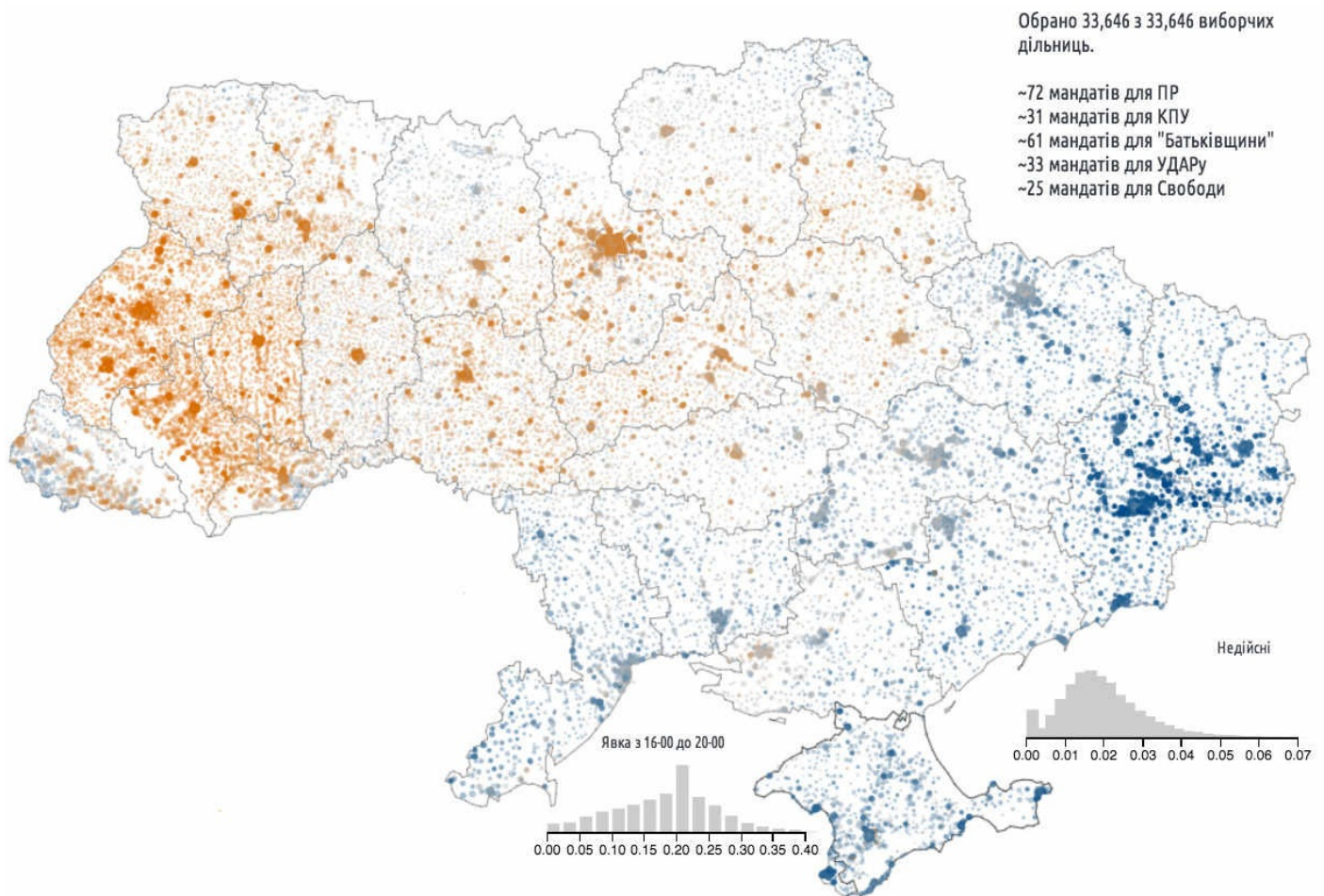
Figure **3.5** Results of the Ukrainian parliamentary elections, 2012. (Visualization by Texty: http://texty.org.ua/mod/datavis/apps/elections2012/.)

You don't need to understand a word of Ukrainian to immediately reach a striking insight: Ukraine, before the crisis began, was a fundamentally divided country. The west was (and still is) mostly pro-Western, and the east and south were mostly pro-Russian.

A year after my visit, Ukrainians in the western part of the country began protesting against President Viktor Yanukovych after he rejected an agreement with the European Union and proposed to strengthen ties with Russia. Yanukovych fled to Russia a few months later, when the protests turned into a full-blown uprising.

When the protests exploded, Western media began publishing maps like the ones in **Figure 3.6**. If you compare them to Figure 3.5, which had had such an impact on me one year earlier, you'll perceive an almost perfect overlap: pro-Western vote corresponds with protests and with fewer people who have Russian as their primary language. What a conspicuous pattern!

Figure 3.6 Ukraine seems to be a completely divided country.

I immediately shot an e-mail to Anatoly asking, "Do you remember the map that you showed me when I was in Kiev? It explains *everything* that is going on right now in your country! It's so prescient! Ukraine is clearly two completely different countries!"

A few hours later, Anatoly replied. His suggestion, which I am not reproducing verbatim, became a motto that I share with my students every semester: **"It's more complicated than that."** I usually add: "And if it's really more complicated than that, then that complexity, which is crucial for understanding the story, needs to be shown in the visualization." **Good visualizations shouldn't over-simplify information. They need to clarify it. In many cases, clarifying a subject requires *increasing* the amount of information, not *reducing* it**.[8]

> [8] The difference between simplifying and clarifying was suggested to me by famous explanation graphics designer Nigel Holmes (http://www.nigelholmes.com).

What could have been added to my mental model was the results of polls like the one that Anatoly attached to his e-mail (**Figure 3.7**). Voting patterns may reveal a sharp ideological divide in Ukraine, but when you ask people if they want to strengthen ties with the European Union or with Russia, the portrait you get is much more interesting. In the east, the most pro-Russian region of Ukraine, just 51 percent of people were in favor of a trade agreement with Russia. In the south, where Crimea is, people's preferences are split almost equally.
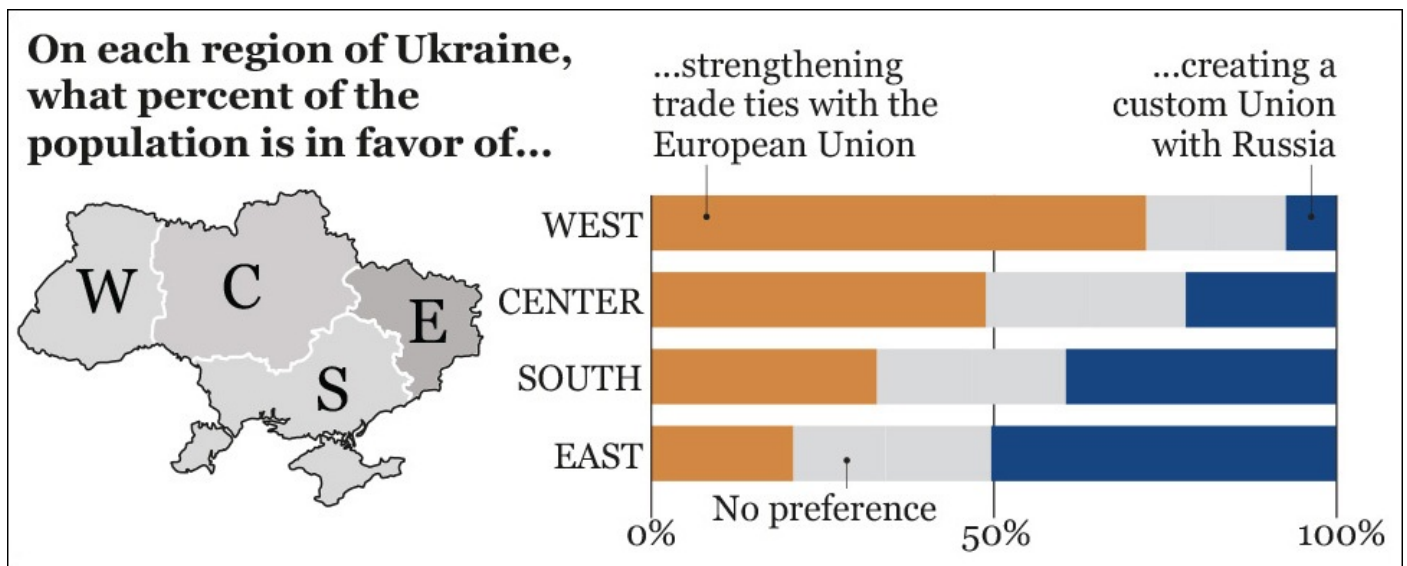
Figure 3.7 What Ukrainians want.

According to Anatoly, Western media were not presenting a nuanced account of the situation. Everyone in Ukraine knows that their country is contradictory and messy—is there any country that isn't?—and that the "Westernized people here, Russianized people there" narrative is a gross mischaracterization. The divide exists, but it's not as clear-cut as it seems. When Ukrainians see Texty's election results map, they know how to put it in context. They see beyond the visualization because they come to it with previously acquired knowledge.

That doesn't happen with readers in other parts of the world. When they see Texty's map, they may create faulty mental models, like I did. They jump to conclusions. **What you design is never exactly what your audience ends up interpreting, so reducing the chances for misinterpretation becomes crucial.** In a case like this, adding a textual explanation to the map can really help. The map would still show the divide. Text would warn people against making far-fetched inferences.

Why Are We So Often Mistaken?

In the past decade, books about how prone to error human reasoning is have proliferated wildly.[9] The portrait of the human mind that those books outline is humbling. Here's the scenario that they describe:

[9] After writing this line, I looked at the shelves in my home office and saw Daniel Kahneman's *Thinking Fast and Slow*; Michael Shermer's *The Believing Brain*; Christopher Chabris' and Daniel Simons' *The Invisible Gorilla*; Carol Travis' and Elliot Aronson's *Mistakes Were Made (but not by me);* Robert Kurzban's *Why Everyone (Else) Is a Hypocrite*; Dean Buonomano's *Brain Bugs*; Will Storr's *The Unpersuadables*; Mahzarin R. Banaji's and Anthony G. Greenwald's *Blindspot*; and David Eagleman's *Incognito.* And there are more in my office at the University of Miami.

1. I detect interesting patterns, regardless of whether or not they are real. I'll call this **the patternicity bug**.

2. I immediately come up with a coherent explanation for those patterns. This is **the storytelling bug**.

3. I start seeing all further information I receive, even the one that conflicts with my explanation, in a way that confirms it. I refuse to give my explanation up, no matter

what. This is **the confirmation bug**.

A caution is pertinent at this point. These are indeed bugs or biases, but they play a key role in our survival. Many authors have pointed out that the human mind didn't evolve to discover the truth, but to help us survive in a world where we needed to make quick, intuitive, life-or-death decisions even when information was punily incomplete. The consequence is that we have all inherited what we could call the "Faint noise behind those bushes > Possible predator > Run or prepare to defend yourself" algorithm.

Snap judgments and intuitions are still a crucial component of reasoning. Risk expert **Gerd Gigerenzer** has written that, "Intuition is unconscious intelligence based on personal experience and smart rules of thumb."[10] To develop good models of reality—descriptive, explanatory, predictive, and so on—we use both intuition and deliberate thinking. Even the most hard-nosed scientific theories begin as gut feelings that are thoroughly tested later.

[10] Gigerenzer has written several books about risk and uncertainty. My favorite one, from which this quote is extracted, is *Risk Savvy: How to Make Good Decisions* (2014).

In the face of patchy evidence, which may make testing impossible, making guesses can be a good option if we have the domain-specific knowledge to do so with accuracy. For instance, after 20 years of designing visualizations, I have developed a gut feeling for when a graphic will work. However, if experiments—even unscientific ones, where I simply show a graphic to friends—show that any of my intuitive design choices is wrong, I need to be prepared to discard it.

Our challenge is that **we snap-judge whether we have the necessary knowledge for it or not**. This is why I created a simplistic mental model of Ukraine after casually comparing several maps of the country: I lacked the domain-specific information to make good inferences from the data, but *I made them anyway*. The very faculties that aided our survival in the past are the ones that can lead us to mistakes in the modern world. As visualization designers and data communicators, it is of utmost importance to be aware of this challenge.

Mind Bug 1: Patternicity

The first mental bug in our list is our astonishing capacity for detecting patterns, visual and otherwise. This is the very faculty that makes visualization such a powerful tool: transform tons of numbers into a chart or a data map and, suddenly, you'll see those numbers under a completely different light.

However, many patterns that your eyes and brain detect in data are the result of pure coincidences and noise. Author **Michael Shermer** calls our tendency to perceive patterns, even when there's nothing meaningful in front of us, **patternicity**. Other scientists call it **apophenia**. Start tossing a dice, and if you get the same number three or four times in a row, you will automatically begin to suspect that there's something wrong with it. It may be amiss, but all may be normal, too. Randomness makes that result possible. Besides, randomness rarely looks truly random to us.

See **Figure 3.8**. It's not very elegant, I know, but don't worry about that. Those are unemployment rates in nine fictional countries between 2010 and 2015. If you just take a

quick look at them, you may not notice anything, but stare at them for 30 seconds, and you'll likely begin seeing patterns. Do you notice that some data points repeat regularly over time, that some highs and lows tend to appear in the same years in some of these countries? The more you scrutinize this array of charts, the more you'll get to see.
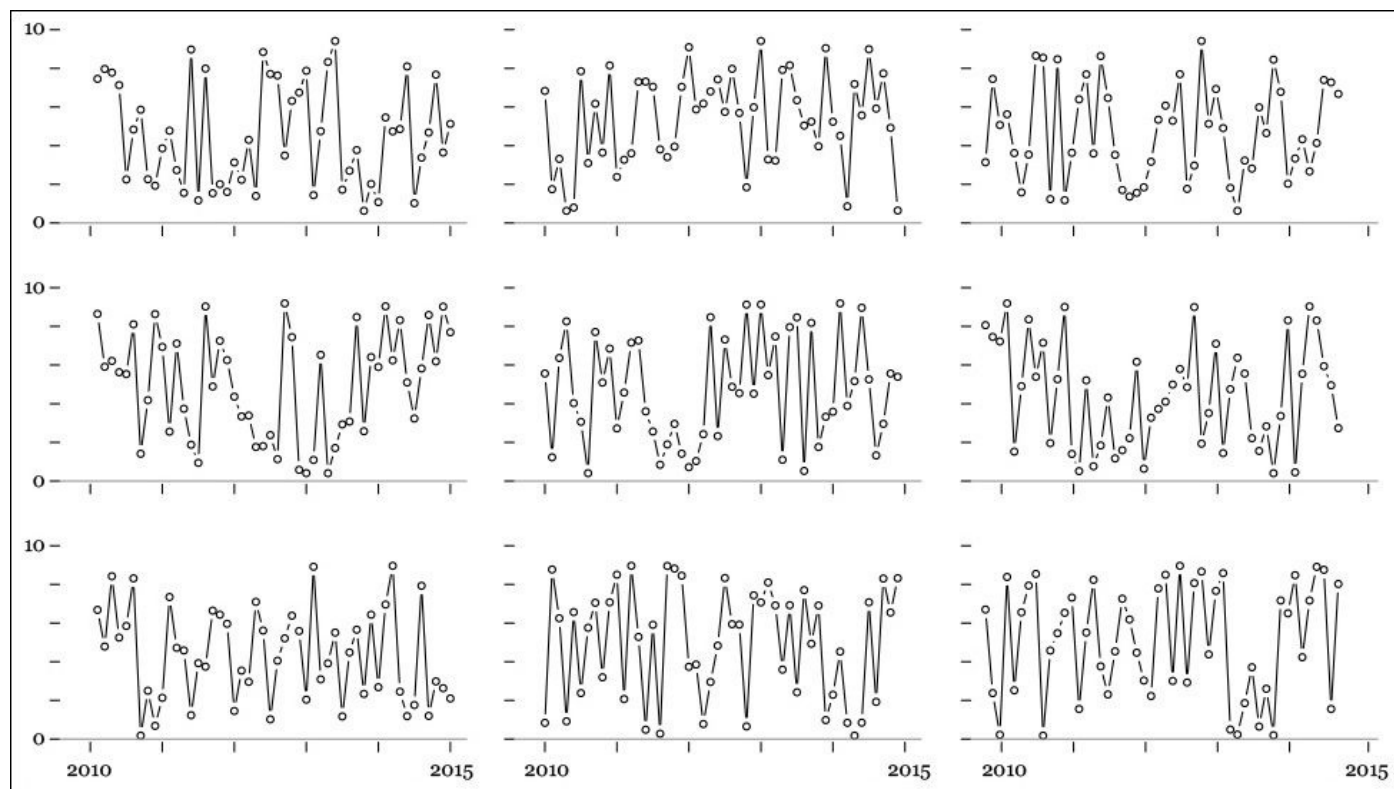
Figure 3.8 Charts designed with random numbers.

Well, all those charts are completely random. I wrote a four-line script that generated 50 numbers between 1 and 10, ran it nine times, and designed the charts. Even if I *know* that they are based on meaningless data, my brain still *wants to see* something interesting in these charts. It whispers, "There are so many coincidences here…. It's impossible that they have appeared just as a result of nature playing dice, isn't it?"[11]

11 Discussing the weird world of quantum mechanics, Albert Einstein claimed that God—a.k.a. "nature," as Einstein was an agnostic—doesn't play dice with reality. According to Stephen Hawking, "Einstein was very unhappy about this apparent randomness in nature (…) He seemed to have felt that the uncertainty was only provisional: but that there was an underlying reality in which particles would have well-defined positions and speeds, and would evolve according to deterministic laws, in the spirit of Laplace."
http://www.hawking.org.uk/does-god-play-dice.html.

Randomness and uncertainty in data may render many news stories and visualizations (think of charts showing tiny changes in the unemployment rate or in the stock market) completely meaningless. Still, people keep making them because of our natural inclination for storytelling.

Mind Bug 2: Storytelling

Once we detect patterns, it is only natural that we'll try to find a cause-effect explanation for them. My reaction to comparing Texty's map of Ukraine to others in Western media was to jump to conclusions, to fill in the gaps, and to frame the results as a story—a made-up one. We humans feel patterns first, we build narratives based on them, and then we look for ways of justifying the rationality of our narrative.

Essayist **Will Storr** defines "story" this way in his book *The Unpersuadables* (2014):

> At its most basic level, a story is a description of something happening that contains some form of sensation, or drama. It is, in other words, an explanation of cause and effect that is soaked in emotion (…) We are natural-born storytellers who have a propension to believe our own tales.

We humans love exciting stories. We see a single event and we transform it into a general rule. We stereotype and generalize. We read about an Asian math genius, and we infer that all Asians are good at math. We see two events in sequential order, and we automatically infer a causal relationship between them. Many among us still believe that cold weather causes colds, when the truth is that what makes cold transmission more likely is our habits: when it's cold outside, we tend to stay indoors more, closer to people who may be already sick.

It happens to me, to you, to all of us. Being aware of this mental bug is paramount for visualization designers, but we're generally oblivious to it, with dire consequences. In January 2014, infographics designer **Raj Kamal** wrote an article in which he explained his creative process:

> It's the "message" that decides the presentation. The numbers, visual, or text or a combination of these are to only support the way of putting the message across. This also changes the way one conceptualizes a graphic. The thought starts with the message and then gets into putting other related information together to support it instead of starting with the data and thinking of what to make of it (…) The advantage of taking this route is also that you are not just restricted by topics or numbers or just presenting "news." You can go a step further and air your "views," too, to make a point.[12]

12 "Everyday Visuals as News." http://visualoop.com/16740/everyday-visuals-as-news-views-and-graphics. Kamal is a very talented Indian designer. He has recently recognized that those words don't really describe how infographics should be made.

Proceeding this way is a recipe for disaster, but it is one that is too common in the news industry. A managing editor comes up with a headline—"We're going to show how raising the minimum wage increases unemployment"—and then asks her reporters to *just* look for data to support it. It is the flaw, I believe, behind *Wired*'s story about the death of Web browsers, discussed before.

Storytelling can be a potent tool for communicating effectively, but it is dangerous if it blinds us toward evidence that should compel us to tweak or discard our models. This is why the modern rise of uncritical advocacy and activism, as well as opinionated "journalism," worries me so much.

After we've grown enamored of our beautiful models, it's difficult to get rid of them. We all have a very sensitive mental trigger that reacts when we're challenged by contradicting information. If it could speak, that trigger might say, "I am a reasonable, well-read brain that thinks carefully about everything, weighing all evidence available. How do you even dare to *suggest* that I'm wrong?" This tunnel-vision effect is the result of cognitive dissonance, and the best way we humans have developed to cope with it is the confirmation bug, also called **confirmation bias**.

Mind Bug 3: Confirmation

Once a good story takes over our understanding of something, we'll attach to it like leeches to warm, plump flesh. An attack on our beliefs will be seen as personal. Even if we are presented with information that renders our beliefs worthless, we'll try to avoid looking at it, or we'll twist it in a way that confirms them. We humans try to reduce dissonance no matter what.[13] To do it, we can selectively search just for evidence that backs our thoughts, or we may interpret any evidence, old and new, in a way that also achieves that goal.

[13] The literature on cognitive dissonance and confirmation bias is also abundant. *Mistakes Were Made (but not by me),* by Carol Travis and Elliot Aronson, is a good primer. The examples mentioned in this section come from this book.

Psychologists have observed these effects so many times that it's appalling that the general public isn't more aware of them. In a classic 2003 study, Stanford University social psychologist **Geoffey Cohen** presented welfare policies to groups of liberals and conservatives.[14] Liberal people endorsed conservative policies if they were presented as coming from the Democratic Party. The opposite was also true: conservatives favored liberal policies if they were told that they had been proposed by the Republican Party.

[14] Geoffrey L. Cohen, "Party Over Policy: The Dominating Impact of Group Influence on Political Beliefs." https://ed.stanford.edu/sites/default/files/party_over_policy.pdf

When asked about why they were in favor or against the policies, both liberals and conservatives said that they had carefully analyzed the evidence. All of them were blind to their own tendency to self-deception, but were ready to attribute self-deception to others.

Similar studies have been done about opinions on gun control, Israeli-Palestinian negotiations, supernatural beliefs, and so on. A very worrying result from some of them is that, in religiously or politically contentious issues, such as climate change, more and better information may not lead to better understanding, but to more polarization. A study titled "The Tragedy of the Risk-Perception Commons: Culture Conflict, Rationality Conflict, and Climate Change" is worth quoting extensively:

> The principal reason people disagree about climate change science is not that it has been communicated to them in forms they cannot understand. Rather, it is that positions on climate change convey values—communal concern versus individual self-reliance; prudent self-abnegation versus the heroic pursuit of reward; humility versus ingenuity; harmony with nature versus mastery over it —that divide them along cultural lines. Merely amplifying or improving the clarity of information on climate change science won't generate public consensus if risk communicators fail to take heed of the cues that determine what climate change risk perceptions express about the cultural commitments of those who form them.

> In fact, such inattention can deepen polarization. Citizens who hold hierarchical and individualistic values discount scientific information about climate change in part because they associate the issue with antagonism to commerce and industry. (…) Individuals are prone to interpret challenges to beliefs that predominate with their cultural community as assaults on the

competence of those whom they trust and look to for guidance. That implication—which naturally provokes resistance—is likely to be strengthened when communicators with a recognizable cultural identity stridently accuse those who disagree with them of lacking intelligence or integrity.[15]

15 Available online at https://www.law.upenn.edu/live/files/296-kahan-tragedy-of-the-riskperception1pdf

Think about this the next time you feel tempted to call someone you disagree with an idiot during a discussion in social media. **The way we present information matters as much as the soundness of the information itself.**

Do you want to get an even better example of dissonance reduction and confirmation bias? Think of your own media consumption. My political opinions are those of a centrist in economic issues and those of a secular liberal in socio-cultural ones. Can you make a guess of which newspapers and weekly magazines I've read on a regular basis for years? I bet that many of you immediately said *The New York Times* and the *New Yorker* magazine.

I don't read those publications just because they are high-quality journalistic products. I also enjoy them because I can read them with my unconscious ideological outrage alarm on standby mode. That's much harder to do when I read quality conservative publications, such as *The Weekly Standard* or *The American Spectator*. But I do read these, too, consciously silencing the annoying little imp who whines, "This is bonkers," before evaluating the arguments.[16] It is hard work, but it's work that needs to be done. My mind, as yours, as anyone's, needs to be disciplined. Left to its own will, when contradicted, it becomes a toddler prone to vociferous tantrums.

16 Notice the "quality" adjective. I avoided suggesting cable networks of any ideological stripe on purpose.

Exposing yourself to contradicting evidence isn't enough. You also need to use tools and methods to evaluate it, as not all opinions and interpretations of evidence and models of reality are equally valuable. The books mentioned in this chapter can help a lot with that, but I'd like to at least share with you two quotes and a graphic that I like to bring to my classes when discussing the many challenges we face when designing visualizations.

Here are the quotes:

> "The first principle is that you must not fool yourself—and you are the easiest person to fool." Richard Feynman.[17]

17 Caltech commencement address, 1974. http://tinyurl.com/h9v3fyp

> "(…) The great tragedy of Science—the slaying of a beautiful hypothesis by an ugly fact." Thomas Henry Huxley.[18]

18 "Biogenesis and Abiogenesis," 1870. http://aleph0.clarku.edu/huxley/CE8/B-Ab.html

Both Feynman and Huxley were referring to science, but I believe that neither was talking *just* about science. They were talking about life in general. The first principle in life is that you must not fool yourself, and a great tragedy in life is that beautiful ideas must be slayed by ugly facts, whenever they appear.

And now, the graphic, which I call **the truth continuum**.

Truth Is Neither Absolute, Nor Relative

Let's go back to the beginning of the chapter, when I wrote that every visualization is a model and that the quality of any model is higher the better that model fits the reality it stands for, without being needlessly complicated. Remember that a model is an abstraction that describes, explains, or predicts something about the workings of nature.[19]

[19] Some of my thoughts in this section have been greatly influenced by David Deutsch's book *The Beginning of Infinity*. See "To learn more" at the end of the chapter.

For the next mental exercise, forget that this book is about data and visualization. Imagine that I use the word "model" to refer to *any* kind of model, from mere opinions to scientific theories, or to the different ways to communicate them, from text to visualizations.

Unless you're a pathological liar or a very particular kind of journalist or strategic communicator, whenever you create a model, you want it to be as close to the truth as possible. Let's suppose that we create a continuum and that we put our model in the middle, like this:

How can we make this model move further to the right? By applying rigorous thinking tools such as logic, statistics, experiments, and so on. More and better information begets better models.[20] A model solidly grounded on these methods is likely to be closer to being true than to being false. I say "likely" because in this little mental exercise I'm assuming that we don't know what "absolutely true" really means. We can't. We are humans, remember?

[20] If we are able to interpret it appropriately, needless to say.

Nonetheless, almost 400 years after Sir Francis Bacon created empirical and experimental science, we have collected enough evidence to know that these methods work. They never give us a perfect understanding of reality but, based on their inherent self-correcting nature (good theories are inevitably killed by *better* theories), they do give us a series of better *approximations*.

Notice that the scale in the diagram is truncated. To understand why, I will refer you to the metaphor of the Island of Knowledge in the prologue. Here you have it again: our island expands and expands, eating away the Sea of Mystery, but the Shoreline of Wonder will never touch the coveted horizon that we long for. For the same reason, on this linear diagram, we'll never know for sure how far we still are from the left or right ends.

It's possible to visualize a comparison between two models that describe, explain, or predict the same reality: one based on applying rigorous methods and another one that is the product of pure guesswork. The diagram would look like this:
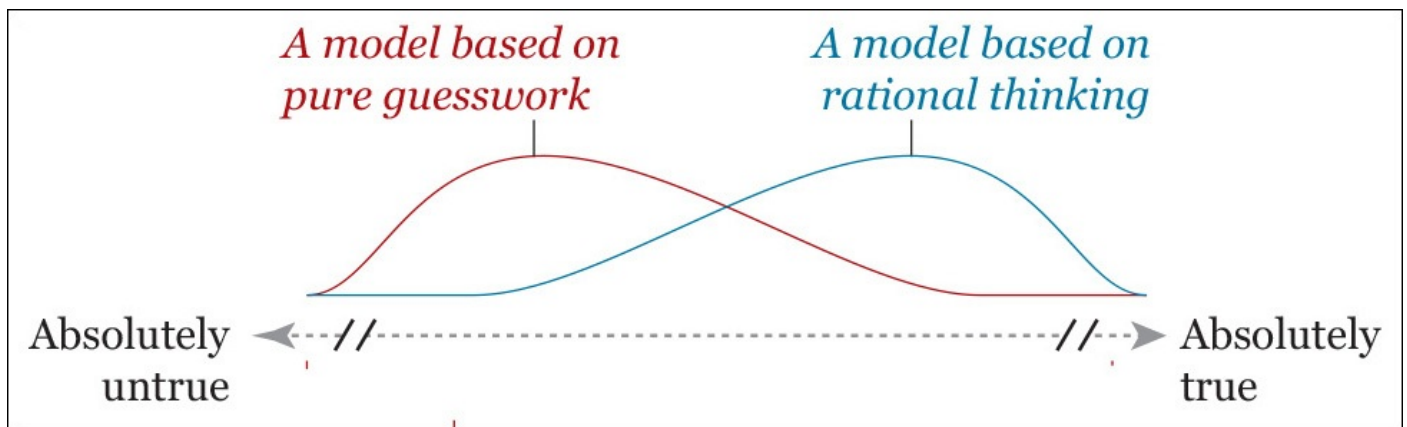
Some clarification is necessary. First of all, the models aren't represented by dots anymore, but by lines with a smooth bump. Here's what this means: when you devise a model, it's never possible to know exactly where it lies in the continuum. All you know is that evidence-based reasoning may move you closer to the right-most end. That's what the bump represents: the higher the curve, the larger the likelihood of the model being in that point in the continuum. Still, there's a possibility that, no matter how rigorous you have been, your model is still inadequate. That's why the blue line extends all the way to the left end.

The bump of the red line is on the left because there are more ways of getting things wrong than right when your only thinking strategy is wacky guesswork. The red line extends all the way to the right end of the continuum because everyone can have a lucky day and hit truth by pure chance.

But it is not true, in science and in other realms of rational inquiry, that there may be competing explanations for the same reality, for the same event, phenomenon, and so on? Absolutely. How do you then decide which one is better?

A good evaluation of the tools and methods used to reach them (logic, statistics, experimentation, and so on) can help in choosing an explanation. However, just to continue with our idea of a truth continuum, it may happen that *multiple good models* exist at the same point on the scale. If they are all based on sound reasoning, *they will all be provisionally true* until more evidence is collected and analyzed. They are all true in the sense that they are equally rigorous, and equally effective, precise, and accurate at describing or explaining a reality, or at making predictions about it.

We can visualize this. See the diagram below, but instead of dots, imagine that you have brackets and curves, like in the previous one. I didn't use them here because the picture got very messy.
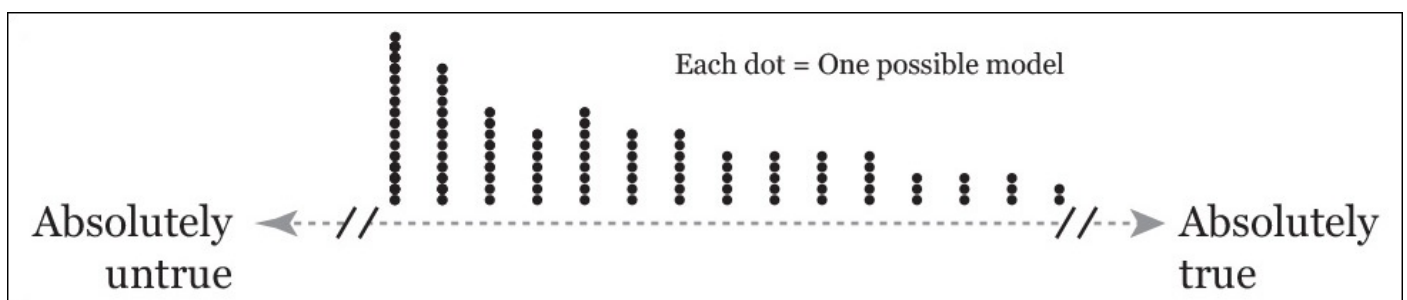
There are many more dots on one side than on the other because there are many more ways of screwing up than of being right about anything.

At this point you may be feeling tempted to send me an e-mail to ask why I bore you with this philosophical discussion. Here's the reason: **the way we think about theories and opinions being more or less true is identical to the way we can think of truer or untruer visualizations**. And the strategies that we apply to make our opinions truer are similar to the strategies that we can apply to make better visualizations.

On the morning of August 15, 2013, my breakfast was ruined by an alarming headline: "Study finds more than a quarter of journalism grads wish they'd chosen another career." It belonged to a story on the website of the Poynter Institute, one of the most distinguished United States journalism education institutions.

I immediately clicked the link and started reading.[21] I have three appointments in the University of Miami, and one of them is in the department of journalism. Was I going to lose one-third of my students in the near future?

21 Here's the story: http://www.poynter.org/news/mediawire/221280/more-than-one-quarter-of-journalists-wish-theyd-chosen-another-career/. An important caveat: the headline mentions journalism grads, but the data represent journalism and mass communication grads. Mass communication comprises also marketing, advertisement, public relations, etc.

The story began with this data point:

"About 28 percent of journalism grads wish they'd chosen another field, the annual survey of grads by the University of Georgia's Grady College says."

A percentage isn't bad per se, as I am sure that the people who conducted the survey and the person who wrote the story know their math. But is this percentage truthful enough? A single figure alone is rarely meaningful, so, at least for now, until we think more about the evidence we have, let's put this story to the left on the truth continuum.

How can we create a more truthful model? An obvious strategy would be to explore the data set further. **When doing a visualization, or the analysis that precedes it, always ask yourself: compared to what, to whom, to when, to where…?**[22]

22 Here's Edward Tufte in E*nvisioning Information* (1990): "At the heart of quantitative reasoning is a single question: Compared to what?"

In our case, we can begin with the "when" factor, as it's present in the survey. The first graphic in **Figure 3.9** shows the variation of our percentage since 1999, which is minimal, despite dramatic changes in the job market. This is something worth analyzing.
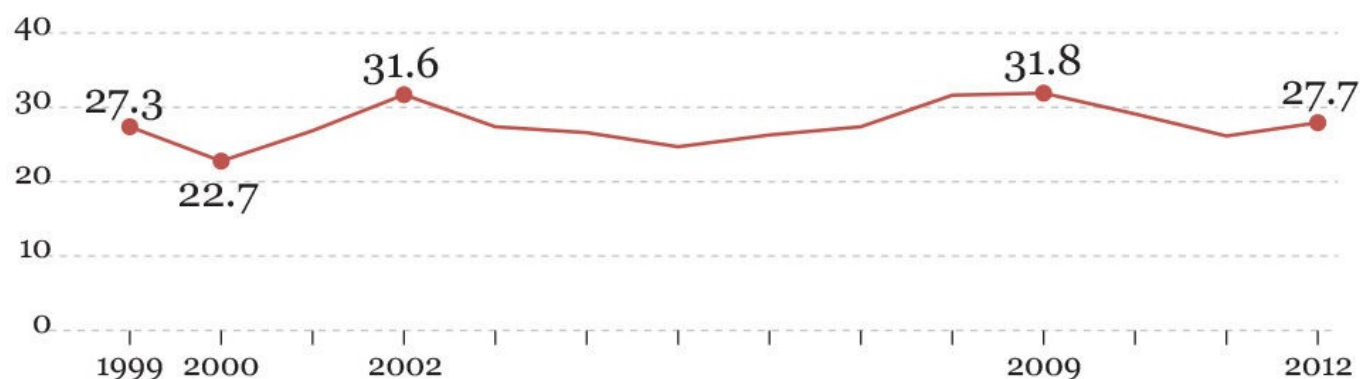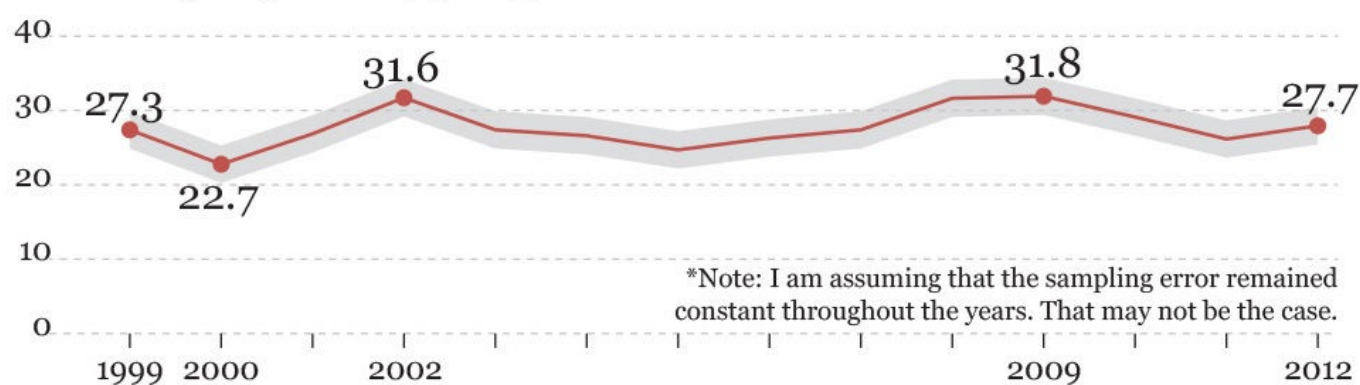
Figure 3.9 How many journalism and mass communication grads say they aren't happy with their choice of major? I am assuming that the error was the same in all years, which may not be the case.

We could go even further and reveal the sampling error, something done in the second chart in that same figure. Any study based on sampling a population won't give you a precise number, but a range.[23] The sampling error here is 2.3 percentage points. That means that if the graphic shows a value of 30, what you're actually saying is that you are reasonably confident that the actual value in that moment in time was between 27.7 percent (30 minus 2.3) and 32.3 percent (30 plus 2.3). We have made our model/story/graphic a bit better. Let's move it closer to the right.

[23] We'll talk about what "error" is in the next chapters.

Absolutely untrue ←·//······································//·→ Absolutely true

Model 1
28% wished that they'd
chosen a different field

Model 2 Model 1 + Change & error
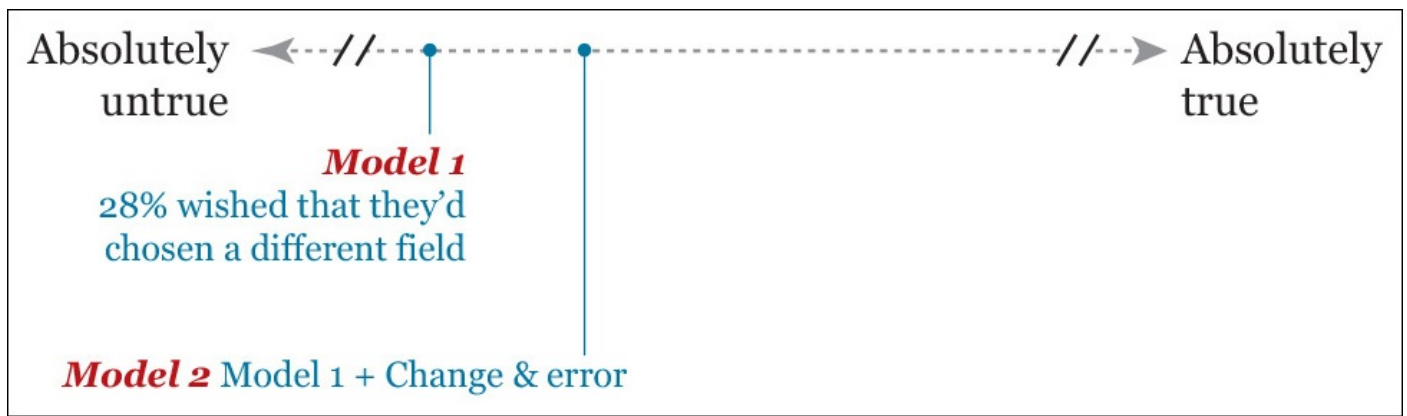
Now, if you were to give an even more accurate assessment of how badly journalism and mass communication grads feel about their career choice, what else would you need to look into?

I'd say that we'd need to compare them to other grads. I wonder, for instance, how many philosophy majors—to choose a major at random, ahem—now regret not double-majoring in computer science, or vice versa. The study I am using as a source itself recognizes that not comparing to other disciplines is a shortcoming.

Imagine that we can survey other grads, talk about it in the story, and display it in our visualization. The result would likely stand a bit further to the right.

Absolutely untrue ←·//······································//·→ Absolutely true

Model 1
28% wished that they'd
chosen a different field

Model 3
Model 2 +
a comparison
with other grads

Model 2 Model 1 + Change & error
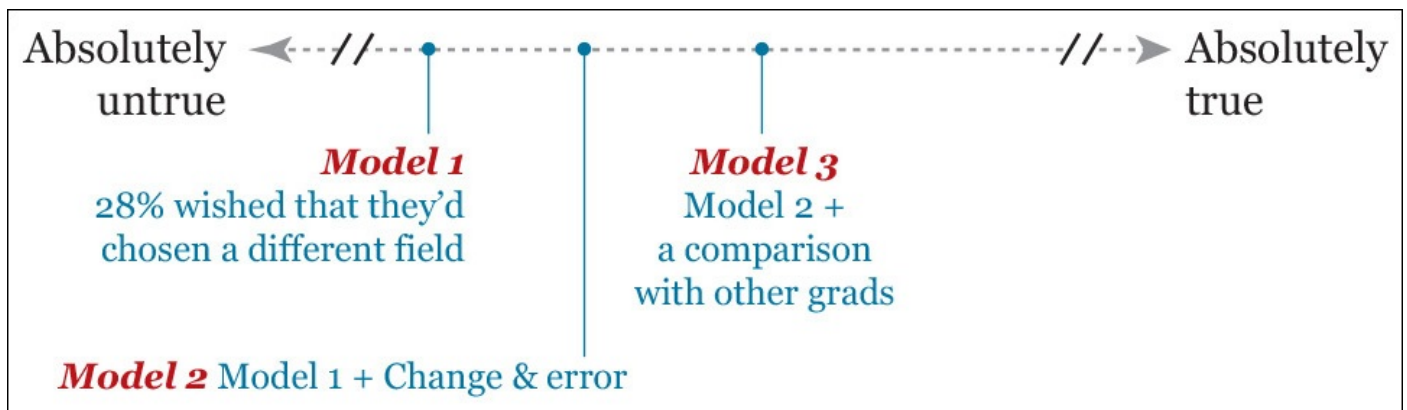
What I've done so far is to increase **depth**, not just in my informal analysis of the data behind the story. All those multiple levels of depth should be revealed and explained to readers if we want them to generate a mental model that is similar to ours.

But is increasing depth enough? Not always. We also need to think of **breadth**. So far, I have toyed with just one variable, the percentage of grads who claim that they wish they had chosen another career. But aren't there other important factors that should become part of our model?

In **Figure 3.10**, I have summarized just a few, such as median annual wages. I've added the median for all occupations and a few other jobs in media (just a portion of journalism grads end up being reporters and editors). Many other variables could also be evaluated, including the number of news organizations, which has been in decline for years, versus the relative health of the marketing and public relations industry, which also employs journalists.

Percent change in employment, projected 2012-2022

Total, all occupations	+11%	
Broadcast news analysts	-2%	
Reporters and correspondents	-14%	
		0

Median annual wages, May 2012. In US $

Total, all occupations	34,750
Broadcast news analysts	55,380
Reporters and correspondents	34,750
Public relations specialists	54,170
Advertising sales agents	46,290
Graphic designers	44,150
Web developers	62,500

SOURCE: U.S. Bureau of Labor Statistics, Occupational Employment Statistics

Figure 3.10 Comparison of several occupations in media.

If we reach this point, we're ready to update our truth continuum diagram again.

Absolutely untrue ◄--//-------•----------•----------•--------//--► Absolutely true

Model 1
28% wished that they'd chosen a different field

Model 3
Model 2 + a comparison with other grads

Models that, besides all previous information, take into account variables that may lead grads to say that they wish they'd chosen a different major

Model 2 Model 1 + Change & error
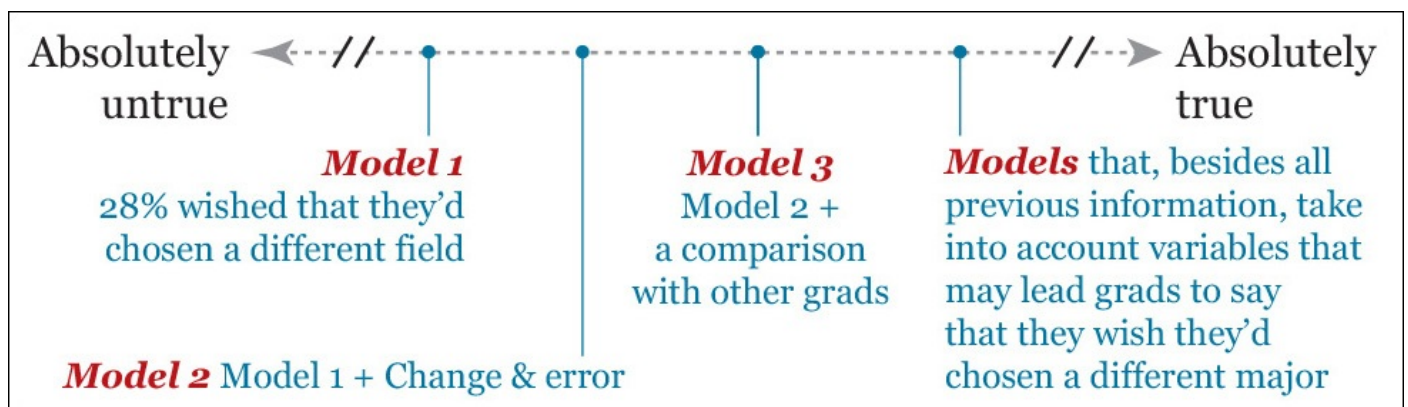
Let's do a quick recap:

1. Job and wage prospects for journalism and mass communication grads who wish to be reporters and achieve that goal aren't great, to say the least. The situation may be a bit better for those few who land jobs as broadcast analysts or news graphic designers or Web developers, and even better for those who work in strategic communication.

2. The number of large- and medium-sized news organizations in the United States has been in decline in the past few years. This is due to shrinking circulations and audiences, as well as diminishing ad revenue. Salaries in other industries that hire journalism grads (marketing, for instance) are better, but not enormously different.

3. In spite of all that, the percentage of journalism grads who say that they wish they had chosen another career has changed very little since 1999. It's just 0.4 percentage points higher in 2012 than it was in 1999, and the sampling error of this survey is 2.3

percentage points.

My tentative conclusion is that, if we refer specifically to journalists, the headline and the angle of the story could adopt a positive tone. Instead of "More than a quarter of journalism grads wish they'd chosen another career," we could say, "Even if job prospects for journalists have worsened substantially and they may worsen even further in the future, the percentage of grads who wish they'd chosen another career hasn't changed at all in more than a decade." Is my model, the analysis, and the presentation of that analysis perfect? It isn't—far from it, as I haven't paid a lot of attention to those grads who didn't study journalism, but other areas of mass communication. That may change my headline quite a lot. More work to do!

<center>▌▌▌▌▌▌▌▌</center>

I am a visual person who is unable to think without scribbling little diagrams, so let me show you what I drew when I was writing the previous few pages (**Figure 3.11**). I'm aware that it's a bit obscure. Remember that I designed it to clarify my own messy ideas, so it's a visual model to improve just my own mental model. It led me to the key takeaway of this chapter:
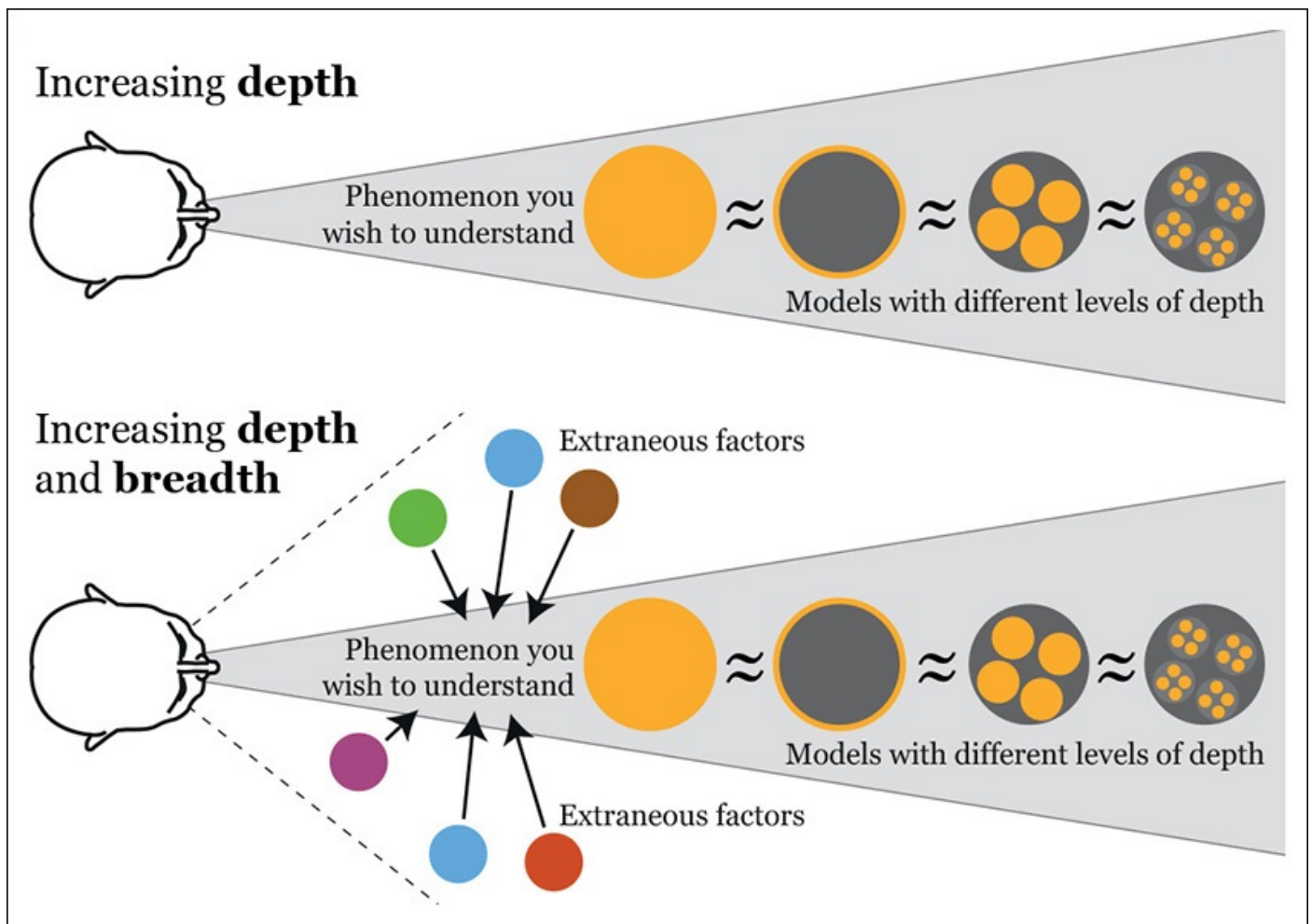
Figure 3.11 Levels of depth and breadth.

Don't rush to write a headline or an entire story or to design a visualization immediately after you find an interesting pattern, data point, or fact. Stop and think. Look for other sources and for people who can help you escape from tunnel vision and confirmation bias. Explore your information at multiple levels of depth and

breadth, looking for extraneous factors that may help explain your findings. Only *then* can you make a decision about what to say, and how to say it, and about what amount of detail you need to show to be true to the data.

The last portion of that paragraph addresses some objections that I've found repeatedly when presenting these ideas to journalists and designers. The main objection goes like this: "I buy the idea of increasing depth and breadth when analyzing my information, but when it comes to presenting the results, **I need to *simplify*.** People have short attention spans! Doesn't your suggestion force us to increase the complexity of visualizations endlessly?"

No, it doesn't.

I am very aware that some inescapable constraints may get in the way of creating a great visualization or story: the space available for it, the time that you have to produce it, and the time that you guess engaged readers will spend reading it.

I'm also aware that we can't present all the information we've collected—at least, not all at once. We need to show a summary first, but this summary needs to accurately reflect reality, and it cannot be *the only* thing we show. We should let people explore as many layers of depth and breadth as is appropriate and reasonable, given the time and space constraints mentioned.

When designing a visualization, we'll always need to make an informed choice about the amount of information that is needed for the audience to understand its messages well. Then, we can weigh that estimate against the time that we have to create our visualization and the room we have for it on the page or on screen.

This compromise reflects something mentioned at the beginning of this chapter: **it's unrealistic to pretend that we can create a *perfect* model. But we can certainly come up with a *good enough* one**.

Finally, unless there's a very good reason not to, we must disclose our sources, data, and the methods used to analyze them and to design our visualizations. Organizations like ProPublica are doing it already (**Figure 3.12**).

Tobacco Debt
How Cash From Big Tobacco Went From Boon to Burden

The Millions New York Counties Coulda Got

By Lena Groeger and Cezary Podkul, ProPublica, Oct. 23, 2014

Tweet 68   Like 0

In 1999, New York counties had a choice to make. They had just been promised annual payments from tobacco companies as part of a national settlement to reimburse them for smoking-related health care costs. Like winning the lottery, they co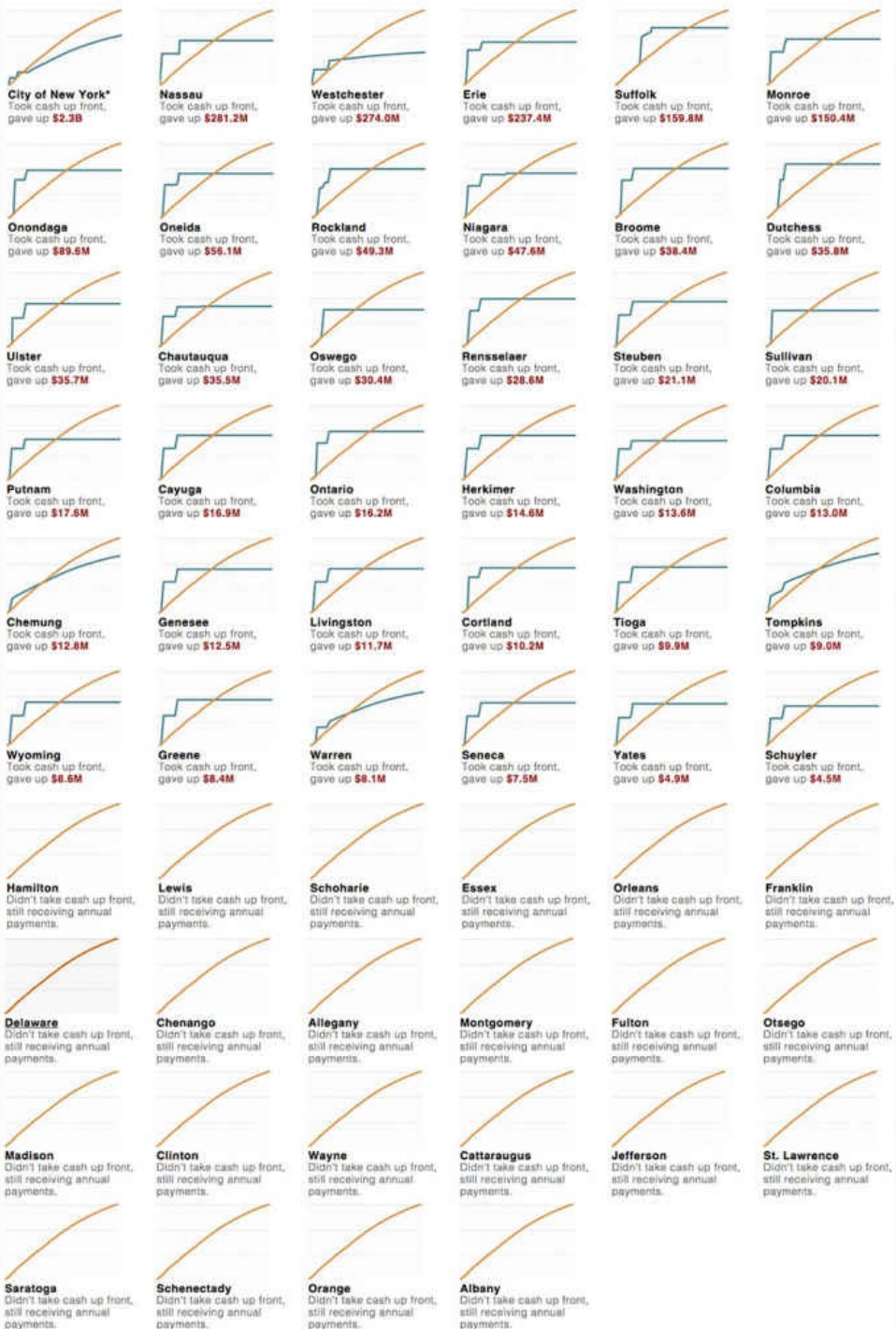uld either get small payments indefinitely -- or take a lump sum immediately by entering into "securitization" deals. Counties knew that these deals would mean less money in the long run, but bankers said they offered protection in case the payments shrank or went away. Now the cost is clear: millions pledged to investors that counties could have kept for themselves. Related Story »

Select a county to see how its deals are shaping up: Lookup a county ▾

City of New York*
Took cash up front,
gave up $2.3B

Nassau
Took cash up front,
gave up $281.2M

Westchester
Took cash up front,
gave up $274.0M

Erie
Took cash up front,
gave up $237.4M

Suffolk
Took cash up front,
gave up $159.8M

Monroe
Took cash up front,
gave up $150.4M

Onondaga
Took cash up front,
gave up $89.5M

Oneida
Took cash up front,
gave up $56.1M

Rockland
Took cash up front,
gave up $49.3M

Niagara
Took cash up front,
gave up $47.6M

Broome
Took cash up front,
gave up $38.4M

Dutchess
Took cash up front,
gave up $35.8M

Ulster
Took cash up front,
gave up $35.7M

Chautauqua
Took cash up front,
gave up $35.5M

Oswego
Took cash up front,
gave up $30.4M

Rensselaer
Took cash up front,
gave up $28.6M

Steuben
Took cash up front,
gave up $21.1M

Sullivan
Took cash up front,
gave up $20.1M

Putnam
Took cash up front,
gave up $17.6M

Cayuga
Took cash up front,
gave up $16.9M

Ontario
Took cash up front,
gave up $16.2M

Herkimer
Took cash up front,
gave up $14.6M

Washington
Took cash up front,
gave up $13.6M

Columbia
Took cash up front,
gave up $13.0M

Chemung
Took cash up front,
gave up $12.8M

Genesee
Took cash up front,
gave up $12.5M

Livingston
Took cash up front,
gave up $11.7M

Cortland
Took cash up front,
gave up $10.2M

Tioga
Took cash up front,
gave up $9.9M

Tompkins
Took cash up front,
gave up $9.0M

Wyoming
Took cash up front,
gave up $8.6M

Greene
Took cash up front,
gave up $8.4M

Warren
Took cash up front,
gave up $8.1M

Seneca
Took cash up front,
gave up $7.5M

Yates
Took cash up front,
gave up $4.9M

Schuyler
Took cash up front,
gave up $4.5M

Hamilton
Didn't take cash up front,
still receiving annual
payments.

Lewis
Didn't take cash up front,
still receiving annual
payments.

Schoharie
Didn't take cash up front,
still receiving annual
payments.

Essex
Didn't take cash up front,
still receiving annual
payments.

Orleans
Didn't take cash up front,
still receiving annual
payments.

Franklin
Didn't take cash up front,
still receiving annual
payments.

Delaware
Didn't take cash up front,
still receiving annual
payments.

Chenango
Didn't take cash up front,
still receiving annual
payments.

Allegany
Didn't take cash up front,
still receiving annual
payments.

Montgomery
Didn't take cash up front,
still receiving annual
payments.

Fulton
Didn't take cash up front,
still receiving annual
payments.

Otsego
Didn't take cash up front,
still receiving annual
payments.

Madison
Didn't take cash up front,
still receiving annual
payments.

Clinton
Didn't take cash up front,
still receiving annual
payments.

Wayne
Didn't take cash up front,
still receiving annual
payments.

Cattaraugus
Didn't take cash up front,
still receiving annual
payments.

Jefferson
Didn't take cash up front,
still receiving annual
payments.

St. Lawrence
Didn't take cash up front,
still receiving annual
payments.

Saratoga
Didn't take cash up front,
still receiving annual
payments.

Schenectady
Didn't take cash up front,
still receiving annual
payments.

Orange
Didn't take cash up front,
still receiving annual
payments.

Albany
Didn't take cash up front,
still receiving annual
payments.

Sources: See our methodology.

* City of New York includes the 5 counties of Bronx, Kings, New York, Queens and Richmond.

Figure 3.12 Visit the visualization: http://projects.propublica.org/graphics/ny-millions and note the "see our methodology" button at the bottom, which will take you to an in-depth discussion about how the data were gathered and analyzed.

▌▌▌▌▌▌▌▌▌

The Laws of Simplicity, by John Maeda, is one of those books that everyone reads and many misinterpret. This is one of its most famous passages:

Simplicity is about subtracting the obvious and adding the meaningful.

Many people memorize the first half and conveniently forget the second. **Simplicity isn't just about *reduction*. It can** (and should) **also be about *augmentation*. It consists of removing what isn't relevant from our models but also of bringing in those elements that are essential to making those models truer**.

The Skills of the Educated Person

In his book *Mapping It Out,* cartographer Mark Monmonier, whom we'll find again in the chapter about maps, outlined the main skills that any educated person should cultivate. They are:

- **Literacy**, or fluency in written expression and understanding of texts.

- **Articulacy,** or fluency in oral communication.

- **Numeracy**, or fluency in analyzing, summarizing, and presenting data.

- **Graphicacy**, or fluency in interpreting and using visuals.

If you've reached this page, you probably understand why I completely agree with him, and you're ready for the next phase of our trip: understanding data.

To Learn More

- Deutsch, David. *The Beginning of Infinity: Explanations That Transform the World*. New York: Viking, 2011. This is the book that influenced me the most when writing this chapter.

- Godfrey-Smith, Peter. *Theory and Reality: An Introduction to the Philosophy of Science*. Chicago: University of Chicago Press, 2003. A concise and profound introduction to the subject.

- Shermer, Michael. *The Believing Brain: From Ghosts and Gods to Politics and Conspiracies—How We Construct Beliefs and Reinforce Them as Truths*. New York: Times Books, 2011. My favorite introduction to brain bugs—or features.