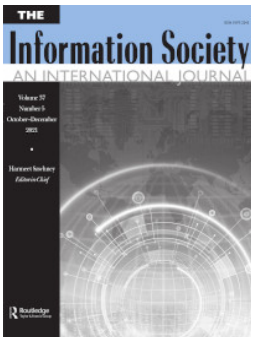
Routledge
Taylor & Francis Group



The Information Society

An International Journal

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/utis20

Data storytelling is not storytelling with data: A framework for storytelling in science communication and data journalism

Sorin Adam Matei & Lucas Hunter

THE INFORMATION SOCIETY
2021, VOL. 37, NO. 5, 312–322
https://doi.org/10.1080/01972243.2021.1951415

Routledge
Taylor & Francis Group

PERSPECTIVE

OPEN ACCESS
Check for updates

Data storytelling is not storytelling with data: A framework for storytelling in science communication and data journalism

Sorin Adam Matei and Lucas Hunter

Brian Lamb School of Communication, College of Liberal Arts, Purdue University, West Lafayette, Indiana, USA

ABSTRACT
Storytelling teaches. All good storytelling is good teaching. Storytelling is a form of teaching and learning because it asks the readers or listeners to replace well-established explanations that are considered facts with new, unexpected ones. We always learn new things when listening to good stories. Good stories do so by violating expectations and surprising the listener or the reader. Surprise breeds suspense, which generates engagement, which is a catalyst for learning – even truer for scientific and data-driven stories. In this paper we show how understanding the nature of good stories by focusing on the novelty they introduce and assumption they violate helps us to do effective scientific work and tell excellent data-based stories. We start by providing definitions of "story," "storytelling," and "good stories." We then outline a methodology for building stories and provide an illustrative example of an effective data-based story from the history of medicine.

ARTICLE HISTORY
Received 5 February 2021
Accepted 19 June 2021

KEYWORDS
Data story; data storytelling; expectation violation; learning; science communication; storytelling; teaching

Introduction

What is a story? Specifically, what makes a "good" story? Let us be direct: a story recounts an unexpected cause that creates an effect for which we already have an explanation. "Good," in this context, is not a mere subjective modifier. Good is not an issue of perception. A story is good because it satisfies a necessary and sufficient conditions; a good story always connects a cause to an effect (Schank 1991). Something leads to something else. Further, it is necessary the cause be unexpected. A good story neither states the obvious nor uses common sense as a universal yardstick (Dahlstrom 2014).

Let us consider the evolving story on scientific understanding of heart attacks. Until 1919, since most heart attacks struck the elderly, they were viewed as a natural way to die of old age. Even once it was discovered that heart attacks are the product of artery blockage, aging again was blamed. This was easily explainable in the era; even if the victim was young, the assumption was he or she prematurely aged. However, in 1919, a purported cause (aging) was challenged by a more credible one (illness). Heart attacks were discovered to be a distinct illness generated by abnormal thickening of arteries, independent of age. Yet, this only begged the question: What hardens blood vessels? It took two more decades to accept an earlier but largely ignored discovery, blood vessels are thickened by fat deposits. Even then, there was no clear understanding of how and why fat deposits accumulated, nor of how they lead to heart attacks. It took another couple of decades and a major seven-country[1] study directed by University of Minnesota researcher Ancel Keys to determine that a diet in rich fats (butter, lard, cheese) damaged the arteries and ultimately led to death. Some populations, such as the Japanese, ate very little fat and had very low blood cholesterol. The heart attack rate in Japan was correspondingly low. Other groups, such as the Finnish participants, ate foods rich in fat (butter, cheese, lard) at very high rates. They died of heart attacks at a rate 13 times higher than the Japanese. Strangely, Italian and Greek participants had high fat intake from olive oil yet had low levels of heart attacks (Goldstein and Brown 2015; Keys 1953, 1975) (Figure 1).

Further, the study found that when Japanese men ingested more fat, they also were more likely to get

CONTACT Sorin Matei smatei@purdue.edu College of Liberal Arts, Purdue University, 100 N. University Avenue, West Lafayette, IN 47907-2050, USA.

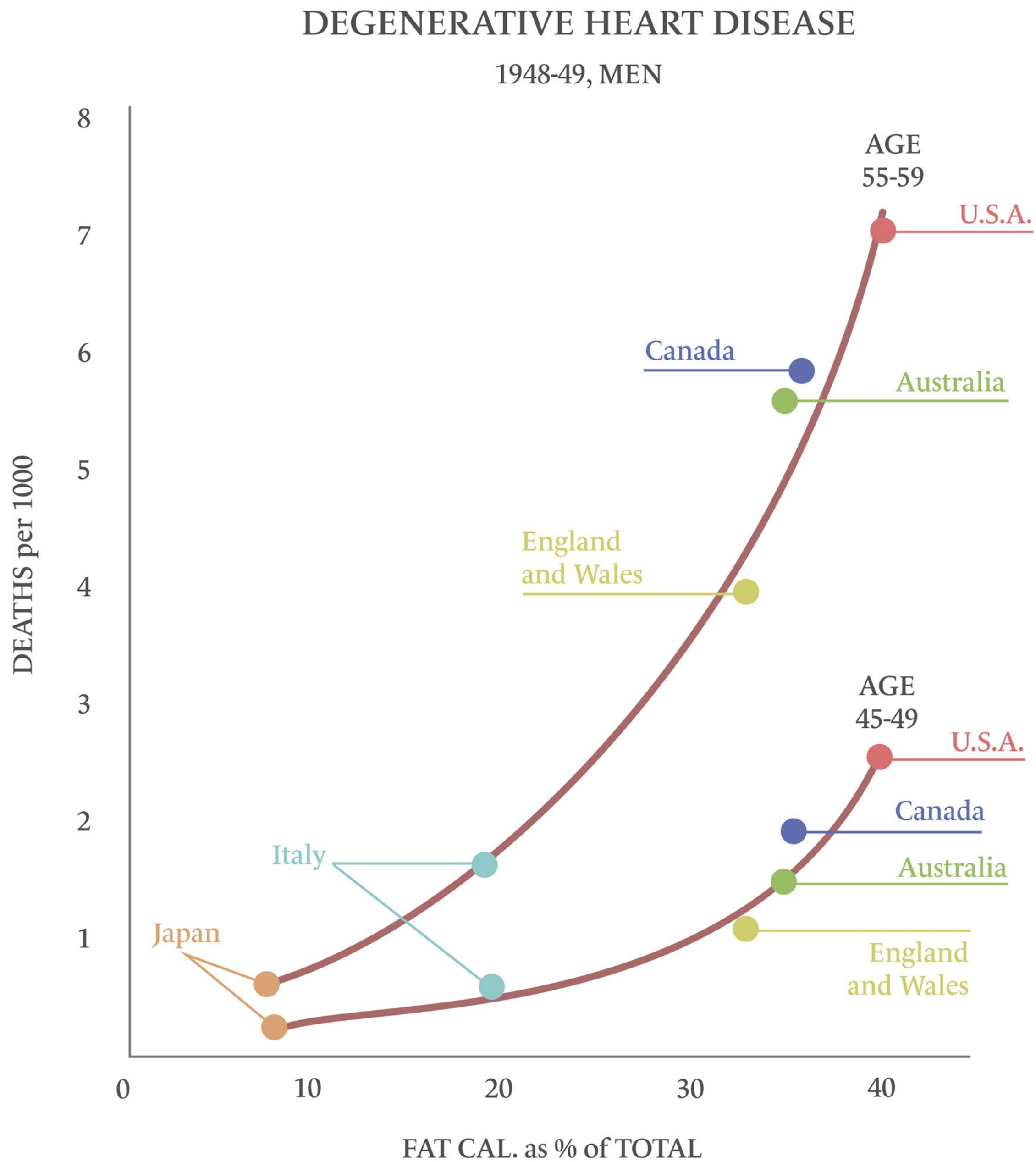


Figure 1. Degenerative heart disease, 1948-49, men; Reproduced after Keys (1953).

heart disease. Also, when East Asians moved to Western countries, changing their diet, their heart attacks increased as well. In the study diet and heart attacks were not only correlated, they were linked in time in a "before" and "after" sequence.

The narrative of heart disease makes a template data story (Dahlstrom 2014; Fadlallah et al. 2019). An old, common sense explanation, aging kills, was replaced by a newer, unexpected one: a good life, rich in calories, kills. The beauty of this relationship lies in its simplicity. An easy-to-read chart with proper contrasts and digestible measurements presents seemingly irrefutable evidence of a credible, casual correlation. Yet, as I explain below, when we return to this story to discuss its statistical grounding, even new explanations may not be sufficient. The relationship between cholesterol intake and heart attack rate is disputed by many (see Steinberg 2007 for an overview). The mechanisms by which fat attacks our arteries are not as intuitive as they seem. True causal effects require many studies over many decades to ascertain, such as the famous Framingham Heart Study (Goldstein and Brown 2015). In fact, behind these initial numbers there is an even deeper, more interesting numeric story which took decades to be uncovered: the Dr. Jekyll and Mr. Hyde story of "good" and "bad" cholesterol. Before we demonstrate the power of stories through that narrative, let us first study and reflect on why stories are more effective when they surprise and provide alternative explanations.

Good stories should violate our expectations because that makes us curious

The best stories employ surprise because surprise begets questions, which beget interest, which begets personal engagement, which leads to learning. All stories that use surprise work well, regardless of context or if they use numbers or not.

Imagine this as the beginning of an otherwise trivial story: A mail carrier rings the doorbell and a dog barks inside the house. This begins a story because an action caused an effect; one thing led to another. A story is, in the simplest terms, an action which causes something to occur. Narrative then contextualizes this action and reveals its outcome. Even basic English sentence structure reflects this: a subject takes action against an object, resulting in an effect. However, the mere presence of a causal chain does not entail a superior story. The reason is straightforward – nobody cares about a predictable chain-of-effects. A dog that barks at a mail carrier is a trivial fact. Any reader expects a dog to bark if a stranger rings the doorbell. Very few people listening to or reading this story will ask themselves: why did the dog bark and why did the mail carrier ring the bell? This is what mailmen and dogs do individually or when they encounter each other; it is the nature of their relationship. Although this story does connect a cause to an effect, it is not particularly "good" as we understand this adjective. A story which states an obvious relationship but does not use the relationship as a means to initiate a chain of events is a failed story.

For a story to be rhetorically effective, it must reveal the unexpected (Banister and Ryan 2001). By doing so, the story stirs our interest, which is a necessary condition of a "good" story. When stories violate our expectations, they intrigue. Then, unsure why the expectations have been subverted, the reader searches for an answer. Stories develop, each new question triggering our mind to find resolution. We follow this trail of questions until a new answer is revealed. To put it differently, the best stories are interesting because they keep the reader on their toes to the very end!

From the storyteller's perspective, this idea is simple logic. To get the attention of the audience, storytellers must disclose the mechanisms of the story. However, if they disclose to the audience what the audience already knows, no one will listen to the story. The storyteller then must feed the curiosity of the audience by revealing something the audience does not yet know. At the same time, storytellers cannot disclose everything from the very beginning, else the story has no suspense. The audience engages with the storyteller more deeply and asks more questions if the resolution is not quite clear. Then, when the answer is a bit different than what the audience expects the resulting surprise will sustain the attention of the audience, prolonging the story. This is a basic cinematic formula, which explains the success of a movie (Ely, Frankel, and Kamenica 2015).

Thus far, we've determined an excellent story connects an unexpected cause to a known effect or a known cause with an unexpected effect. This relationship is the necessary condition for the best stories. When practiced, this relationship takes a precise shape.

Let us return to our mail carrier and dog story: The mail carrier rang the bell, the dog barked, and the mail carrier barked back. Now, this is quite a story, isn't it? It is not natural for mail carriers to bark back at dogs. Their job description might even forbid it (there is liability, OSHA training, and so on)!

More importantly, the minute the mail carrier muttered her fatuous bark, the story got legs. The mail carrier's bark violates not only your expectations, but your logic. You start questioning your own understanding of the situation. Reality seems to fall apart. To rationalize the situation, you begin asking questions. Does the mail carrier know the dog? Is he or she greeting the dog back? Is this his or her own pet? Yet, one's own dog should not bark at the owner; the dog should wag its tail and whimper. So, why did the dog do this? Perhaps the mail carrier is utterly frustrated with the many dogs who barked at the mail truck that day?

Storytelling = teaching + learning

The mail carrier's story violates what we know about the world as well as our understanding of cause and effect. This subversion sparks more questions, moving the story forward. A moving story develops new questions and links new unknown causes to effects. This pattern of question-and-answer is what makes a story "good."

Up to this point, we have discussed several features of a great, authentic story besides the most necessary one – there should be a relationship between unexpected causes and effects. An engaging story makes you ask questions, which fuels your imagination further, more deeply engaging the reader with the narrative. A reader, interested in a story, will follow the story to its end for the resolution. Then, when that resolution is used as a teaching point, the readers learns about both themselves and the world around them, changing themselves for the better. *The best stories excel because they induce a change in the audience through learning.*

Now, we have a compelling and complete examination of exemplary story mechanics. A skillfully told story has the ability to sustain itself, generate new sub-stories, and keep the readers engaged. Sequentially, the engaged reader connects an unexpected cause with a visible effect, they learn something new, and are compelled to act. In view of this, stories are part of the process of human discovery and growth. When we tell a good, interesting story, we discover new things or teach people new ideas that may be used practically. Stories are always teachable moments. They add value, reveal new possibilities, and encourage imagination of future scenarios and actions.

Science is a form of storytelling

Put another way, "good" stories speak about and lead to discovery. Likewise, the discovery process has the potential to create great narratives. Scientific storytelling, in turn, has a natural advantage over conventional storytelling because it consistently provides new, unexpected explanations. Since science grows by continuous subversion of our expectations, scientific endeavors steadily seed stories. By the nature of their practice, scientists can also be excellent storytellers. In fact, they are taught to think like storytellers. Yet, like Jourdain, the protagonist of Moliére's (1998) *Le Bourgeois Gentilhomme*, many scientists don't know it yet.

Meaningful science is, after all, the quest to provide unexpected answers to questions about observable (and often taken-for-granted) realities. Science is the enterprise of explaining why some known causal relationships are false and why we should replace those notions with new, unexpected causal relationships. Observable facts often violate common sense or previous experience, including established scientific knowledge. The latest theories are supposed to violate common sense or existing knowledge by reconsidering explanations for diverging observables. Furthermore, any existing theory needs to be contested and, if proven wrong, usurped by a new explanation at the heart of a new theory. The saga of science is one long, never-ending story. To see this in action, observe below the story of the relationship between cholesterol and heart attacks which, again, will become even more relevant when we will learn of the saga a story evolves into as it ages.

Science is not (nor should it be) in the business of ascertaining whether fact, or that which has been proven as true before, is still true or will be true for all time. The basic premise of science requires the current answer to any question be constantly checked against the latest facts and observations. This is why in science we never fully "prove" the validity of a hypothesis, we only "reject the known (null) hypothesis."

Why removing the Earth from the center of the universe violated common sense but made great science

Let us examine examples of science as a never-ending narrative, perpetually proposing unexpected solutions to known problems. In the process, we will better understand both storytelling as a *forma mentis* for the scientific mind and how science informs data storytelling.

Current theories about the movement of planets and stars utilize basic ideas dating to Copernicus, Kepler, Galileo, and Newton. The Sun is the center of the solar system. Earth and other planets revolve around the Sun as satellites revolve around the planets. Newton's formula for gravity (two bodies will attract each other directly proportional to their mass and inversely proportional with the square of the distance between them) mathematically explains both the shape of the orbits and the celestial bodies' speed of the movement within them.

However, this simple, direct, and intuitive model has not always been a self-evident truth. In fact, the simple statement that it is the Earth which moves, while the Sun in fixed, was unconsidered for many years and controversial once introduced. This is not only because some priests wanted to hide the truth from people to preserve religious dogmas; even long before there was organized religion, people believed the Earth was the center of the universe. To them, the notion the Sun revolved around the Earth was little more than common sense. "Just look up in the sky," they would say. "If you see the Sun, mark its position. An hour later, look again. What do you see? The Sun has moved!" According to their (and your) perception, the Sun moved around the Earth. The claim "Earth revolves around the Sun" violated verifiable observations. Thus, early astronomical theories, though not scientifically accurate, were justified by the senses. The moment Copernicus (and later, other

Renaissance scholars) challenged this long-held belief, a common understanding of our ancestors' beliefs were violated. There was substantial debate, which forced the proponents of the new theory to propose newer, more appropriate methods to prove their theories correct and human senses wrong.

The greatest (scientific) story ever told: Or how the apple pulled the Earth toward it

Galileo – and later, Newton – continued the story of (what we now call) "the heliocentric theory of the solar system," or an explanation for why and how the planets move around the Sun. The final contribution was the theory of gravity, which redefined mass and weight and proposed a new cause which made objects fall to the ground. Until Galileo, a simple theory formulated by Aristotle (and no more elaborate than what our senses told us) claimed that all objects have a natural place in the universe (Darling 2007). Weight was the expression of their attraction to that pre-determined place. Elements such as water and Earth (including rocks, metals, and other minerals) belonged in the center of the universe. Given the geocentric philosophy of the time, the center of the universe was the center of the Earth. Fire, vapors, or hot air belonged to the world above Earth, and they sought to rise and escape the Earth to be reunited with their pre-destined location (Grant 1996).

The Earth itself was not considered the source of the object's need to return to the center of the universe. Earth was reasoned to be a sphere around that center. Similarly, the planets and the Sun were part of more distant spheres of the universe. They had their own natural state, which was to revolve around the world. On a large scale, however, there was no unified theory to explain why the Sun moves around the Earth, why some objects fall to the ground, or why fire reaches toward to the heavens. Every object had its own intrinsic characteristics, explained by its "natural" state. To the thinkers of the time, these observations contained logic. Our senses certified every single proposition in the Aristotelian philosophy of nature, which explains its extraordinary longevity.

A fundamental flaw in the Aristotelian model of the macrophysical world was pointed out by Galileo, whose experiments indicated, contrary to popular belief, there might be a force (rather than a "natural state" of the objects) responsible for the phenomenon known as "weight" (Grant 1996). This force is a function of the object's mass. However, it was Newton who defied all intuitive perceptions of macro-reality. He affirmed, in a stroke of genius expressed mathematically with the famous formula,

$$F = G\frac{m_1 m_2}{r^2}$$

that all objects are attracted to each other by a mysterious gravitational force. This is proportional to their objects' mass ("m_1," "m_2") and manifests itself as a force of attraction ("F"; "G" = gravitational constant; "r" = distance) of other objects toward it, and is reciprocated toward those same objects.

Newton's claim was nothing short of stupefying. He alleged the proverbial apple which fell on his head did so because the Earth attracted the apple, *and* because the apple attracted the Earth, a borderline heresy. The fact that the apple ended up moving toward the Earth, and not vice-versa, was due to a discrepancy between the masses of the two objects, not because of some intrinsic natural difference. Yet, Newton's claim was so counterintuitive it demanded new questions. How far can gravity reach? How does gravity travel? What is it? How does it compare to other forces, such as magnetism or electrostatic attraction? These are still open questions in the scientific narrative as we cannot yet fully explain if or how gravity can be integrated with the other three fundamental forces of the universe (weak, strong, and electromagnetic). Thus, Newton's story maintains very long legs as its journey has not yet ended.

The incomplete nature of Newton's theory highlights something else; even the very best stories provide temporary answers. In fact, the next theory (story) of gravity is already in development. Although it has not yet been experimentally demonstrated, string theory claims our concept of gravity and electromagnetism (among everything else in the universe, including fundamental particles) are mere figments of reality, conceptualized by our current mathematical models. These supposedly-material realities are the best current way to express a deeper reality, which might be organized as a collection of strings. According to this theory, our universe with all its intricacies can be seen to be arising from the vibrations of these strings whose interferences and combinations create all visible, measurable objects, and effects: particles, forces, radiation, mass, and so on. For now, this theory remains merely hypothetical, as no one can identify and measure "strings" in their "habitat." They are too small, too multidimensional to be captured with our current instruments. A new, even more surprising theory than string theory may lurk beyond our present observational capacities, waiting to capture our imagination, test our reason, and add a new chapter to the narrative which predates Aristotle (Davis 1971).

How many ways can you violate common sense and get away with it?

To recap, good stories violate expectations and, by doing so, fuel their own narrative progress. Scientific research is naturally disposed to storytelling because it is (or should be) in the business of breaching expectations. How does story subversion work in scientific practice, though?

Murray Davis, a sociologist and organizational science scholar, offers a high-level introduction to violation of expectations as a scientific method. He works thorough taxonomy of subverting strategies capbable to create strong research programs with great potential for storytelling (Davis 1971). He proposes effective research and storytelling strategies can be identified in two domains, and within each domain lie several paths of development.

The first domain is that of single phenomenon. When we observe individual entities (people, molecules, messages, organizations, etc.), we hold to pre-conceived models of them. These models work with simple – yet effective – stipulations: organization, composition, level of granularity, uniqueness, stability, functionality, and evaluation of the observable facts. The job of a scientist is to propose new explanations which violate these predispositions. Davis summarizes this strategy in a simple heuristic list of contrarian positions a researcher can adapt to generate interesting research.

According to Davis, you can propose an effective program of research (and consequently generate great stories) if you can show with respect to:

Organization

- what appears disorganized is in fact organized or vice-versa.

Composition

- what appears to be heterogeneous phenomena are in fact part of the same class or vice-versa.

Abstraction

- what appears to be individual phenomenon is, in reality, the expression of a group of system process or vice-versa.

Generalization

- what appears to be general or universal is, in fact, local or particular or vice-versa.

Stabilization

- what appears to be stable and unchanging is, in fact, fleeting and liable to transform or vice-versa.

Function

- what appears to not work as predicted can, in fact, work well by other criteria or vice-versa.

Evaluation

- what appears to be a negative phenomenon may, in fact, be a positive one or vice-versa.

The method proposed by Davis for building research agendas is relevant not only for individual phenomenon, but also classes or groups of phenomena.

Just like with individual phenomenon, you can propose an effective program of research of classes or groups of phenomena if you can identify in classes or groups:

Correlation or co-variation

- unrelated phenomena are in fact correlated or vice-versa.

Co-existence

- phenomena that appear to coexist, repel each other or vice-versa.

Opposition

- similar phenomena are in fact opposed to each other or vice-versa.

Causal direction

- effects are, in fact, causes or vice-versa.

Is telling good stories as simple as saying "it ain't so"? is it even ethical?

While this recipe-like presentation of Davis's method seems simple and effective enough, it should justifiably raise the question of its validity. Further, as scientists, we have a responsibility to question the morality of using the contrarian method. Is crafting a research strategy as simple as upturning the alleged certainties of today? Is not this an invitation to a fishing expedition or worse – to unethical attempts to prove all we know as false (and, in most cases, the opposite is true) only to garner plaudits?

If one were to think about the cultural prevalence of astrology and other pseudosciences or the anti-vaccination movement, one could say questioning science for the sake of turning facts on their head is not only deceiving, but may result in real harm. However, one should note that neither Davis nor this paper claims scientific research or satisfying stories should start with the unwarranted contestation of known facts, theories, and explanations solely for the sake of contrarianism. Solid research (and, thus, the data stories derived from it) starts with the observation of notable and persistent inconsistencies between facts and known theories which explain them. In effect, we advise nothing more or less than what any research methods textbook would advise: if a known theory does not fit the facts, change the theory, not the facts. In this respect, Davis's work and this paper situate themselves in the mainstream of the scientific research tradition.

Data storytelling is not storytelling with data

Everything thus far is relevant to both scientists and communication practitioners interested in scientific stories. As we mentioned in the introduction, one of our motivations for defining "good" stories is to empower practical communication of data related stories. This simple intuition can be traced to the first thing learned in many intro-to-journalism and mass communication classes: "If the dog bites the mail carrier, it is not a story; if the mailman bites the dog, THAT is a story!" In the interest of full disclosure, the example used above was chosen in comradery with this famous dictum, and with the first author's former job at the BBC World Service, where he used to work during the early 1990s.

The argument unfolded thus far may not stun scientists either. Even if they were not aware of how close their enterprise is to storytelling, they should've been asked – at least once – the unavoidable "what are you telling me that I do not know?" question. The goal of this essay was to lay the foundation for a renewed conversation about the function of storytelling in scientific, data grounded environments. In light of the definition of a "good" story, we can boldly proclaim data stories need to contravene assumptions and provide new explanations for known facts. What lingers is the unaccounted for: How should data be purposed to violate previously held assumptions?

That is to say, data storytelling is a type of narrative that upends the public's assumptions about a *causal relationship*. Furthermore, data storytelling exchanges one causal explanation with another, using the rules of causal inference specific to quantitative analysis. The causal relationship, rigorously speaking, might be hard to prove. It is not easy to show that one data series causes another since mere correlation is not causation. Typically, we need to make the case by tracking variables across time, since in causal analysis the cause should precede the effect in time. This means data storytelling may be a hard trade to practice. But who claimed data storytelling is such a simple, speedy, cheap feat? Those who imagine that adding a bit of data to a narrative converts it magically into "data storytelling" have missed the point; reliable data stories should be close or even indistinguishable from groundbreaking science. For all causal claims made with data, the criteria of causality must be met: there should be a lag between cause and effect, third factors should be controlled for, and most importantly effects should be large and significant. Of course, having set the bar so high, many might say that this definition of data storytelling is too exigent. Some might say it creates a widening gap between quantitative analysis and storytelling – sending the former to an ivory tower while keeping the latter in a realm of anecdotes and fables. Far from it! We need more data, rigor, and accuracy in data-driven storytelling, yet data stories must follow the rules of data analysis. May the gods of statistics be mollified: the more radically a story violates common assumptions, the better the story.

Data storytelling at its finest: Surprise the audience and then surprise it again!

We began our story of high-grade data storytelling by recounting how physicians and biologists saw blood cholesterol as inherently responsible for cardio-vascular disease. Further, they proposed that ingestion of high amounts of heavy fat leads to high cholesterol levels in the blood. By 1961, the American Heart Association was advising doctors to tell patients affected-by or predisposed-to heart disease to switch to a low-fat diet, doubly so if that diet included "heavy" fats, such as those found in lard or butter (Kritchevsky 1998).

Yet, scientists soon observed that not all people who had high blood cholesterol levels were equally prone to a heart attack (Lloyd-Jones et al. 2003). A large study including thousands of patients over more than three decades, conducted in Framingham, Massachusetts, found that the risk of coronary and heart disease depended more on the composition of your cholesterol than simply on its level (Mahmood et al. 2014). Further research revealed cholesterol in the blood can be of two types: heavier (HDL) and lighter (LDL) (Castelli 1988). More importantly, LDL cholesterol, which is more abundant and tends to vary with diet and exercise regimen, sticks to blood vessel walls more frequently. About one third of cholesterol is made of another, high density blood fat, HDL, which is considered healthy. HDL molecules tend to bind to those of LDL and carry them away, preventing LDL from sticking to arteries. Think of HDL as a "garbage collector."

LDL and HDL are not present in the blood in a fixed ratio. For the same level of cholesterol, there could be more or less LDL correspondingly low or high HDL. Those in the High HDL/Low LDL situation will have heart attacks less frequently. Those with High LDL and Low HDL, will be heart attack prone. The ratio between the two types of cholesterol depends on many things including diet. Nuts, fish, and olive oil lower LDL and increase HDL levels. In conclusion, health depends on a ratio of cholesterol levels, not an absolute value of cholesterol levels, and the ratio can be moderated by diet and exercise.

The following figures emphasize the core idea behind this data story, which rejects the broad assumption "cholesterol is bad," and proposes a new one in two parts: (1) you can have high cholesterol and still live, and (2) yet you need to lower "bad cholesterol" and increase "good cholesterol" if you really want to live a long and healthy life.

Figure 2 shows the risk of developing heart disease represented as a "risk sheet." This is a two-dimensional chart, showing how LDL and HDL variables interact with each other to determine heart disease risk. The sheet changes color corresponding to risk level as the combinations, which are not linear, vary. The color

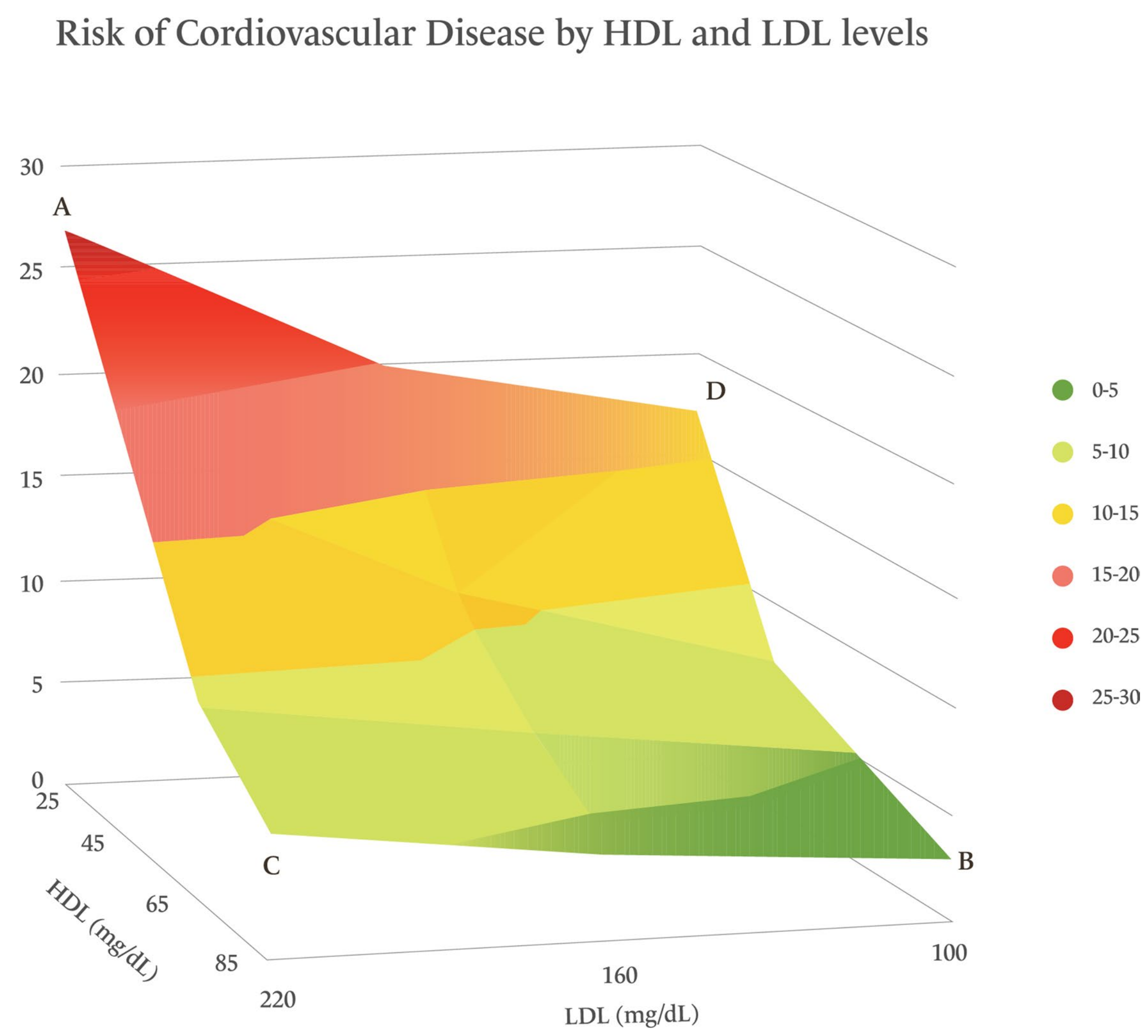


Figure 2. Risk of cardiovascular disease by HDL and LDL levels. Chart recreated from data presented in Castelli (1988).

of the sheet represents the level of risk: red is high, green is low. The values range from minimum risk of 0-5%, colored deep green, to a maximum of 25-30%, colored maroon, risk to develop heart disease over a period of 10 years. The subjects studied are men between 50 and 70, who were generally more susceptible to a heart attack at the time of the study (1960s-1980s). These data were reported in 1988 by Dr. William P. Castelli, a co-author of the famous Framingham Heart Study (Castelli 1988).

The horizontal axis, which runs from left to right, measures the level of "bad" LDL cholesterol, from 100 to 220 mg/dL (milligram per deciliter). The depth axis, front to back, measures the levels of "good" HDL cholesterol (25 to 85 milligram per deciliter).

The highest risk is where the upper left corner of the sheet (A) is anchored on the back wall of the chart. At this level HDL is the lowest (35) and the LDL value is the highest (220). The greenest area is in the lower right corner (B), where the opposite is true: LDL is the lowest (100) and HDL is the highest (85). That's where you want to be.

Now, let us look at the lower left corner of the sheet (C), the one in the foreground, pale green, and let us follow the edge of the sheet up, as it rises, turning orange, pale red, red, and dark red at the upper corner (A). This is the true surprise that lies in wait. As we go higher from corner C to A, the chance of developing heart disease over ten years skyrockets from about 5% (on in twenty) to over 25% (one in four chances). Along this path, the level of LDL, bad cholesterol is high and constant. What changes is the level of HDL, which drops. Lowered HDL in the body is the equivalent of a lowered shield in medieval battle.

Despite that, the more pressing point is if you add up the level of HDL and LDL, which gives you an approximation of your level of total cholesterol, at various levels of risk, you realize you can have a rather low chance to develop coronary disease even with a high cholesterol level, which in corner C is over 300. The opposite is true as well! At corner D, the cholesterol level (125) is significantly under that considered harmful by doctors (over 200), yet your risk will be higher than at corner C, which is 50% higher than normal.

Further insight can be derived from this chart by comparing the risk level of the lower right corner (the greenest one, B), which is rather low (under 5%), with that of the upper right one (the pale red one, D), which is over 15%. The risk at point D is several times that of point B. Yet, the total cholesterol level decreases. If we add the HDL and LDL levels for each corner, we get 185 for the greenest corner (B) (100 LDL + 85 HDL) and 125 (100 LDL + 25 HDL) for the pale red one (D). Thus, total cholesterol at corner B (low risk) is greater than corner D (high risk). Despite conventional wisdom, a lower level of total cholesterol, as is the case for point D, can in fact hide a much higher risk of heart attack. Therefore, it is not just the level of cholesterol that matters, but how the different levels of good and bad cholesterol are mixed. An old explanation, "cholesterol is bad," is replaced by a new one: a lack of good cholesterol is bad for you.

For an ordinary person, the implication of the story is tremendous. Nothing works better than the fear of the things we do not see. Being healthy is not about being disciplined in a blind way, but about being smart about our choices.

In terms of data storytelling, a simpler visualization may still present the breadth of our point. Let us distill our data story into the simplest chart we can (Figure 3).

The horizontal axis indicates levels of both good and bad cholesterol without numeric values: low, medium, and high. The vertical axis shows risk of heart disease. The lines are labeled LDL (blue) and HDL (orange). What matters here is the slope of each line. The steeper the slope, the greater the impact of the represented cholesterol on the risk factor, while the direction suggests the positive or negative relationship. It is obvious that the steeper line is the orange one, suggesting HDL has a greater impact on your risk for heart disease and in a good way. As the level of HDL increases, your risk of coronary disease decreases. At the same time, as the LDL decreases, your chances of getting heart disease decreases, too, but at a much slower rate. Lowering your LDL from high to low, decreases your risk from about 15 to 10 (less than 50%). Increasing your HDL from low to high, decreases your risk from 20 to under 5, which is more than four times. The proportional difference in the effect is more than eight times (over 400%/less than 50%).

Conclusions

A narrative in which an expected cause is connected to a known fact is a trivial story, neither a good story nor one worthy of telling. Because the audience did not learn anything new from it or have its assumptions challenged, everyone's time was wasted. It did not make the audience pause, take stock, analyze, speculate, lead to a change of mind, or even force the audience to refute the story. A story worth telling challenges

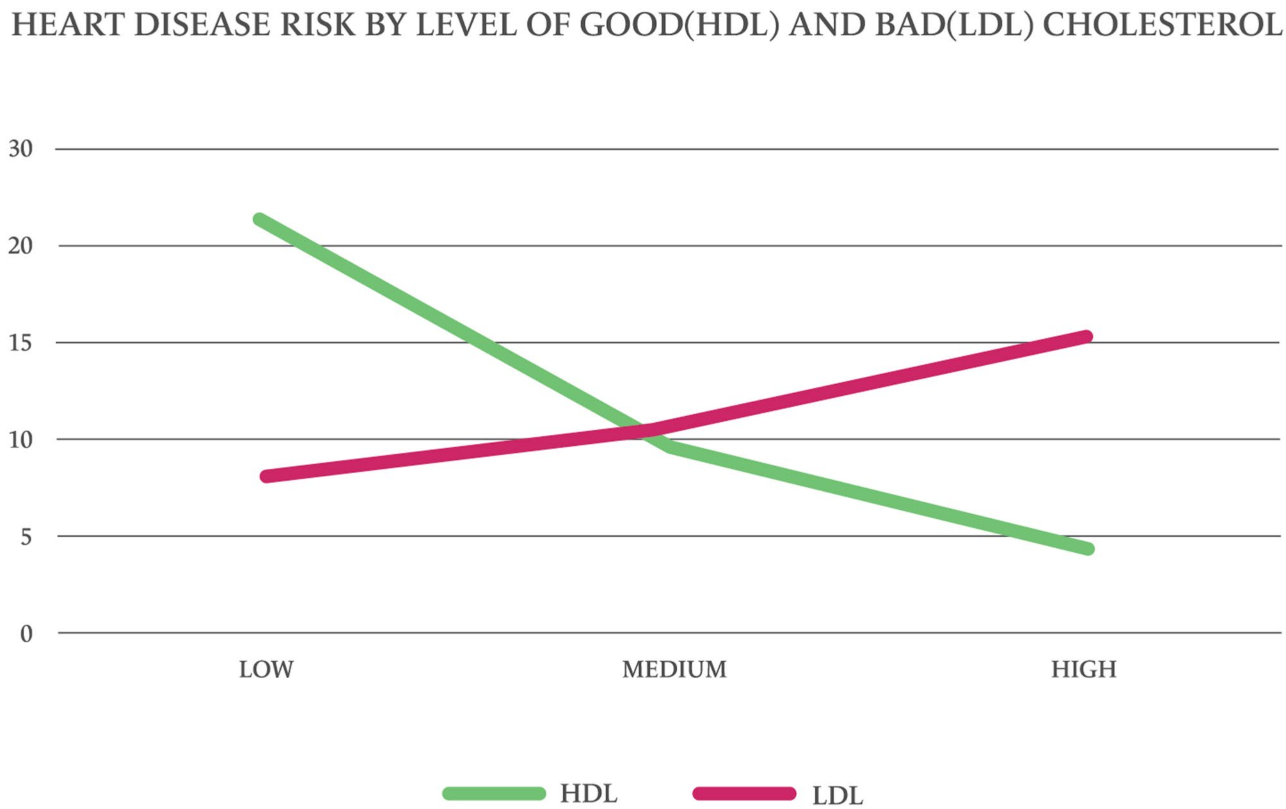


Figure 3. Heart disease risk by level of good (HDL) and bad (LDL) cholesterol.

and changes the reader. Ordinary stories leave the audience static emotionally and intellectually.

A satisfying story is a means of discovery. It drives the audience to ask the storyteller, present or virtual, more questions. A compelling story sparks public imagination and draws people to think. An adept story compels a person to adopt an attitude. Further, data stories, at their best, push the audience from simply believing, to knowing with a degree of confidence. The audience is moved toward what Plato defined as "justified true belief," a workable definition of knowledge (Chappell 2013).

When relying on data to make their point, stories which use a new theoretical model (one that better fits the facts while violating previous theories) have a higher chance to capture public attention. Thus, it is the job of the professional communicator, especially data scientists or journalists, to seek explanations which violate assumptions and known theories in a *justified* manner. In doing so, communicators need to follow the rules of sound inferential reasoning. First, numbers that explain should be directly correlated to changes in the observable facts. In other words, there should always be two variables: a cause and an effect. Second, causal factors need to occur before the effects they generate. Essentially: A cause *then* an effect. And third, all factors that might be responsible for both the cause and the effect should be accounted for and eliminated as a possible cause.

Opposing assumptions and determining a strong causal mechanism does not guarantee success, though. As already mentioned, stories which defy both our theoretical assumptions and the way we think about the facts will always have a hard time entering the zeitgeist. Often, until the validity of new facts is established, the story will sound too incredible to believe. Also, what works for the public might not work for experts. There are things that general audiences may not know, just as there are factors the experts over-evaluate.

Often, simple things – trivial statistical procedures (like regressions) used to recognize faces in pictures, for instance – trigger a shrug from many computer scientists while causing ordinary people to shake their heads in amazement. An intriguing finding is always interesting to someone, somewhere, and its impact needs to be judged accordingly. However, one thing should always be remembered: no one ever changed the world and got famous by forcing open doors.

Note

1. Informal exploratory studies were conducted in Italy, Spain, South Africa, and Japan. More formal pilot studies were conducted in Finland, Italy, and Greece.

Funding

Purdue University Integrated Data Science Initiative.

References

Banister, F., and C. Ryan. 2001. Developing science concepts through story-telling. *School Science Review* 83:75–84.

Castelli, W. P. 1988. Cholesterol and lipids in the risk of coronary artery disease: The Framingham Heart Study. *The Canadian Journal of Cardiology* 4 (Suppl A):5A–10A.

Chappell, T. 2013. Plato on knowledge in the *Theaetetus*. *The Stanford Encyclopedia of Philosophy*. Accessed June 16, 2021. https://plato.stanford.edu/archives/win2013/entries/plato-theaetetus/

Dahlstrom, M. F. 2014. Using narratives and storytelling to communicate science with nonexpert audiences. *Proceedings of the National Academy of Sciences* 111 (Supplement_4):13614–20. doi: 10.1073/pnas.1320645111.

Darling, D. 2007. *Gravity's arc: The story of gravity from Aristotle to Einstein and beyond*. New York: Wiley.

Davis, M. S. 1971. That's interesting! Towards a phenomenology of sociology and a sociology of phenomenology. *Philosophy of the Social Sciences* 1 (2):309–44. doi: 10.1177/004839317100100211.

Ely, J., A. Frankel, and E. Kamenica. 2015. Suspense and surprise. *Journal of Political Economy* 123 (1):215–60. doi: 10.1086/677350.

Fadlallah, R., F. El-Jardali, M. Nomier, N. Hemadi, K. Arif, E. V. Langlois, and E. A. Akl. 2019. Using narratives to impact health policy-making: A systematic review. *Health Research Policy and Systems* 17 (1):26. doi: 10.1186/s12961-019-0423-4.

Goldstein, J. L., and M. S. Brown. 2015. A century of cholesterol and coronaries: From plaques to genes to statins. *Cell* 161 (1):161–72. doi: 10.1016/j.cell.2015.01.036.

Grant, E. 1996. *The foundations of modern science in the Middle Ages: Their religious, institutional and intellectual contexts*. Cambridge, UK: Cambridge University Press.

Keys, A. 1953. Atherosclerosis: A problem in newer public health. *Journal of the Mount Sinai Hospital, New York* 20 (2):118–39.

Keys, A. 1975. Coronary heart disease-the global picture. *Atherosclerosis* 22 (2):149–92. doi: 10.1016/0021-9150(75)90001-5.

Kritchevsky, D. 1998. History of recommendations to the public about dietary fat. *The Journal of Nutrition* 128 (2 Suppl):449S–52S. doi: 10.1093/jn/128.2.449S.

Lloyd-Jones, D. M., P. W. F. Wilson, M. G. Larson, E. Leip, A. Beiser, R. B. D'Agostino, J. I. Cleeman, and D. Levy. 2003. Lifetime risk of coronary heart disease by cholesterol levels at selected ages. *Archives of Internal Medicine* 163 (16):1966–72. doi: 10.1001/archinte.163.16.1966.

Mahmood, S. S., D. Levy, R. S. Vasan, and T. J. Wang. 2014. The Framingham Heart Study and the epidemiology of cardiovascular diseases: A historical perspective. *The Lancet* 383 (9921):999–1008. doi: 10.1016/S0140-6736(13)61752-3.

Molière. 1998. *Tartuffe and The Bourgeois Gentleman* (bilingual edition, ed., trans. S. Appelbaum). Mineola, NY: Dover Publications.

Schank, R. C. 1991. *Tell me a story: A new look at real and artificial memory*. New York: Atheneum.

Steinberg, D. 2007. *The cholesterol wars: The skeptics vs. the preponderance of evidence*. New York: Academic Press.