Transcript for
Why Data Should Be Over-the-Counter
By Jenny Grant Rankin, Ph.D.
TEDxTUM, October 24, 2015

http://tedxtalks.ted.com/video/Why-data-should-be-over-the-cou

Let's talk about heroes.

It's 1854.

Within a few square blocks of London, people start dying mysteriously. Within just 10 days 500 people die, and no one knows why.

Meet Jon Snow. The *real* John Snow.

He's a doctor who convinces London officials to remove the handle from a local water pump.

See, Snow plotted the deaths' locations on a map.

He made a small black mark where every death took place.

This allowed officials to see that the closer you lived to the water pump, the higher your household's death rate.

Snow had identified the worst cholera epidemic to ever hit the nation – and the disease was being distributed through a single water pump. Snow ended this epidemic using data visualization. Snow became a hero because he could *effectively* communicate data.

Now it's 1855. Britain and France are battling Russia in the Crimean War.

55% of the nearly 2 million soldiers die. But the guns and bombs aren't what's killing them.

Meet Florence Nightingale. She's a nurse, a statistician, and a social reformer.

Nightingale convinces Britain to take funds away from weapons, and spend that money on better healthcare instead - right in the middle of the war.

She does this by visualizing the mortality data. Diagraming the preventable disease in blue-gray, Britain can see that preventable disease comprises the vast majority of deaths.

Nightingale gets her hospital and better healthcare conditions, and the soldiers'' death rate drops to a fraction of what it had been. Nightingale was a hero because she could *effectively* communicate data.

We've just talked about some heroes in data visualization. But data has not always lead to success.

Imagine it's 1986.

Seven crewmembers board the Challenger space shuttle.

There have been concerns about the booster rocket's O rings functioning properly. The O ring is a seal that prevents hot gases from making contact with other parts of the shuttle.

To determine if it would be safe to launch, the space shuttle engineers and managers view charts concerning O ring damage and launch temperatures. They look at this diagram…

…and they also look at this graph. They view this data projected on a screen, just like you are viewing. But do they get the information they need? Because, for me, these aren't very clear.

Many of the engineers still have concerns, but the ultimate conclusion is to launch the shuttle anyway on that cold January day.

The Challenger explodes within 73 seconds, killing all crewmembers. But it didn't have to happen.

A series of data design experts have since *redrawn* those charts used by Challenger engineers.

First, we can take away all the unnecessary lines and notes that cluttered the graph. …and we can make it more clear we're showing the number of failed launches, such as those where the shuttle might explode, and we're arranging those failures by *temperature*, since the day's temperature is a factor that impacts O-ring failure.

Yet the information still tells us very little. So we're going to add something that was previously missing from the graph.

Notice all the failed launches are on the graph: cases where there were 1, 2, or 3 failures at each launch temperature.

But what about the successful launches? What about data on the 17 cases where there were 0 problems with the shuttle's launch? That data was missing from this graph shown to engineers and managers on the eve of the Challenger's launch.

So let's add it. Here, shown in green, are the cases where there was no O ring failure, and the shuttle launched properly.

What can you *instantly* notice about all 17 successful launches? They all occurred when the launch temperature was at least 66 degrees. There were a few outliers, but the significant trend was that warmer days were in the safe zone, whereas a launch on a day colder than 66 degrees would probably result in O ring failure. This improved graph shows that the colder the day, the greater the danger.

When the Challenger launched, the day was 31 degrees cold – so cold, it doesn't even fit on the graph. But as we can see, O ring failure was highly likely at that temperature.

Remember those unclear graphs that were used? Many speculate the Challenger disaster might never have happened if people were fluent in communicating their high stakes information.

So we've looked at data visualization done right, and we've looked at what can happen when data is misunderstood. This brings us to the most vital differentiator between the two.

Design. Design lets us communicate data so it can be understood. Sometimes this involves visualizing data, but there are less-visual aspects to data communication, as well.

A few years ago I conducted a quantitative study of how high stakes data is communicated to educators.

We educators use student data to inform how we help kids. When teachers understand the information, this works great. We can help students in ways never possible without today's data technology.

For example, we might learn this student is struggling with literary analysis, isn't trying very hard, and has attendance problems. That information can help us really target this student's specific needs.

But what if we misunderstood the data? What if *word* analysis is really at the root of the student's reading challenges? What if the student is actually trying hard, but is challenged by a learning disability? What if health problems are the real reason this student keeps missing school? Even if I use multiple measures and my own insight as an educator, if I routinely misunderstand data, I can treat problems that aren't really there, and overlook other problems that are. I could fail to give a student what the student really needs.

Multiple studies show: nearly all teachers use data to inform what they do for students, but educators draw correct conclusions from less than half of the data they view.

But we can't blame educators. They're smart, well-educated, and caring! *I'm* an educator!

Rather, if intelligent people misunderstand the information presented to them, they are no more to blame than the Challenger space shuttle managers who overlooked problems that were poorly charted.

What I found in my research is that high stakes data is often communicated in poorly *designed* ways, which leads to misunderstanding and poor decision-making.
I wanted to change this. But how?

During my research my daughter got sick. Just a flu, nothing too serious. She didn't need a doctor, but she did need medicine.

I went to the pharmacy and looked at over-the-counter medicine. It was easy to understand what the different products were for and to find what my daughter needed. Because when something is "over-the-counter," it is easy to understand and use without the help of an expert. There are key components in place to be sure the product is used appropriately and does what it's supposed to do.

First off, the medicine's *content* has to have ingredients that work. For example, a pill containing nothing but sugar and food coloring could not have cured my daughter's flu.

Effective labeling tells you how to use the product. It has directions for use, ingredients, warnings, and more.

Appropriate packaging helps you understand the product's nature. For example, if a label had an image of a child on it, yet the medicine was intended only for adults, that could be dangerously misleading.

Supplemental documentation is often tucked inside the package. This gives added explanations and details that can't fit on the label.

And an online help system helps you understand the product. Over 50 million users per year use WebMD, an online help tool where people can search for information by topic or keyword. It would be negligent to sell medicine without these over-the-counter components.

In these ways, when something is "over-the-counter," it is easy to understand and used without the help of an expert.

This is exactly what we need for high stakes data.

So I studied how vital a data report's *content* is to the data's value.

…and how data *labels* can best function, such as with better titles and report footers that offer warnings and guidance to stop viewers from misunderstanding the data that is displayed.

I studied the best ways to *package* or *display* data so the design encourages proper analyses.

…and how supplemental documentation can walk viewers through how to understand and use a data display.

I studied how a help system can best be added to a data system, so users are helped in using the technology but also in using its data. It turns out, I was onto something.

So I summarized over 300 studies and other expert sources from a wide variety of fields, and I published this as a set of standards for improved data reporting – a way of displaying data so it is "over-the-counter" and thus easy for people to understand and use.

There are standards for how best to *label* data, such as limiting a report footer to 328 characters…

How best to *package* or *display* data, such as using 2-dimentional graphs rather than 3-D…

…and how to best handle every "over-the-counter" aspect we've talked about that can help data make more sense to viewers. But I wanted more evidence.

In a quantitative study involving 211 educator participants, I tested a sampling of these standards.

With just one change to the way data was communicated, people's understanding of that data more than doubled.

Applying other Over-the-Counter Data Standards tripled participants' understanding of that data.

And other standards caused participants to understand the data 4 times as often. I tested these standards in isolation to isolate variables, but together we could conceivably make it so there's nearly no chance data can be misunderstood. We *can make* important data easy to understand.

And various groups are currently using these standards to help students, who are affected by their teachers' use of data.

Metro Nashville Public Schools in the state of Tennessee is one example of a school district applying these data reporting standards, and this single district's presentation of data affects 82,000 students.

The State of South Dakota is another example, and this single state's added supports for data use affect approximately 150,000 students.

Illuminate Education is an example of a data systems company that applies these standards, and this single company's data systems affect 5 million students throughout the U.S. and in other countries, as well.

I wrote these free standards for the reporting of education data, but the standards hold implications and potential for reporting any form of data when it is imperative that data be understood. But this is not a finite solution. Researchers should continually be examining how best we can share data with others.

We know how design can impact the communication of data in good or bad ways.

And we know there are research-based standards for *how* best to make data "over-the-counter" when we display it for others. So how does this all come together? What is the next step needed to ensure all the powerful data out there is presented in ways we can understand?

The key is building people's ability to leverage design in the best ways possible.

If we merely glance at data displays, we see a wide variety of design at play. Design is like a language with its own vocabulary. Technology allows us to use this language more than ever before, and we need to know the options at our disposal and the guidelines for effective communication.

We can't cut the arts and design from students' educations, or we rob them of a universal language. When we're fluent in visualizing data and making data "over-the-counter" for those who view it, we can communicate important messages across global boundaries, across cultural and language barriers, and across other obstacles thrown in our paths.

If our data matters, then it matters that we communicate that data in over-the-counter ways so it can be understood. Because, like medicine, data really does impact lives.

Thank you.