CRAFTING STORIES WITH DATA

Amer Tadmori
Mile High Data Day
09/14/17

Thank you so much for that introduction and thank you all very much for being here today. I know you had a couple choices of workshops to choose so I'm glad you've picked this one. I'm super excited to be here talking to you all about how to craft stories with data.

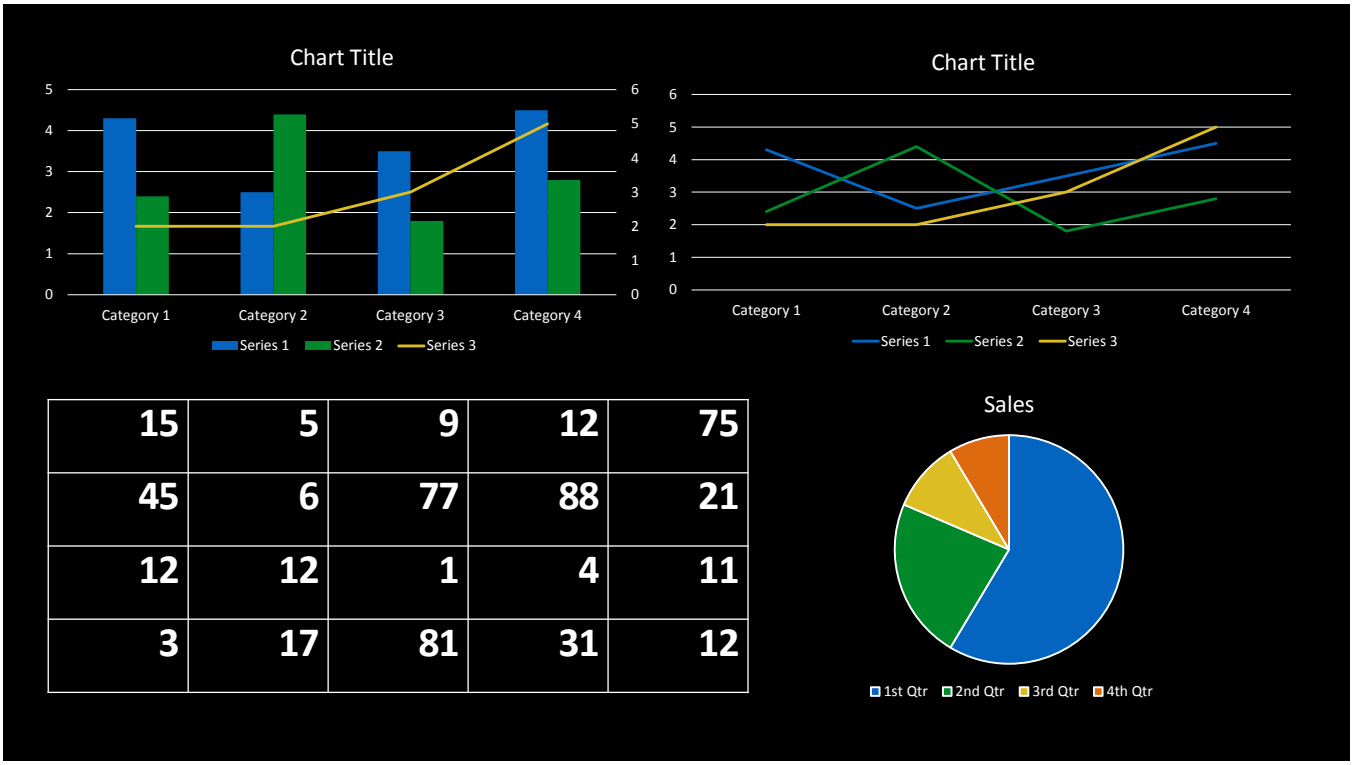
Why storytelling?

And I am very excited because this is a topic I'm very passionate about. But why? So why is storytelling so important when we're working with data? Well, it's because the numbers can't talk! The story is there in the numbers…they can reveal patterns, answer question but they really need you and your voice and your brain to get that out to where other people can understand them and take action on them. And ultimately, as a someone who analyzes data your goal is to have others take some sort of action based on what your work has uncovered.

“Numbers have an important story to tell. They rely on you to give them a clear and convincing voice” – Stephen Few

Storytelling will also give your work more impact. I’ll never forget one of my very first presentations for a big analysis. I was working on a food brand – a popular baking mix – and was giving a presentation on the results of their fall baking season promotion. Folks I had so much amazing data, organized into slides, I had been working on this for weeks. Get to the end of the presentation and the Chief Marketing Officer raises her hand and says “well okay so what? What do we do with this?” And let me tell you it was pretty rough getting called out in a meeting like that – but when I went back later and looked at it I realized I’d just presented a list if facts. Without tying them together. Without making them meaningful for the audience.

So along those lines storytelling is more than just telling others what you’ve discovered and what they should do about it, you use storytelling to engage the audience, and get them to WANT to take the actions that you’re asking them to take. Like Jennifer Newcomber mentioned this morning in the kickoff – we’re at the intersection of data and humanity! So by brining together a story you’re conveying that humanity piece of the data. Show of hands, how many of you have sat through presentations that go on like this….and this….and this….



Over crowded charts

- Models typically are embedded in larger information systems that manage the flow of data from various sources into the model and handle the aggregation and reporting of model outcomes. Model calculations should be properly coordinated with the capabilities and requirements of information systems. Sound model risk management depends on substantial investment in supporting systems to ensure data and reporting integrity, together with controls and testing to ensure proper implementation of models, effective systems integration, and appropriate use.
- The use of models invariably presents model risk, which is the potential for adverse consequences from decisions based on incorrect or misused model outputs and reports. Model risk can lead to financial loss, poor business and strategic decision making, or damage to a bank's reputation. Model risk occurs primarily for two reasons:
- The model may have fundamental errors and may produce inaccurate outputs when viewed against the design objective and intended business uses. The mathematical calculation and quantification exercise underlying any model generally involves application of theory, choice of sample design and numerical routines, selection of inputs and estimation, and implementation in information systems. Errors can occur at any point from design through implementation. In addition, shortcuts, simplifications, or approximations used to manage complicated problems could compromise the integrity and reliability of outputs from those calculations. Finally, the quality of model outputs depends on the quality of input data and assumptions, and errors in inputs or incorrect assumptions will lead to inaccurate outputs.
- The model may be used incorrectly or inappropriately. Even a fundamentally sound model producing accurate outputs consistent with the design objective of the model may exhibit high model risk if it is misapplied or misused. Models by their nature are simplifications of reality, and real-world events may prove those simplifications inappropriate. This is even more of a concern if a model is used outside the environment for which it was designed. Banks may do this intentionally as they apply existing models to new products or markets, or inadvertently as market conditions or customer behavior changes. Decision makers need to understand the limitations of a model to avoid using it in ways that are not consistent with the original intent. Limitations come in part from weaknesses in the model due to its

Or this, the chart that's a ton of words

Score Range	Development Period		Validation Period		Dist. Difference	Proportion	Log of [Proportion]	Population Divergence
	Count	Percent A	Count	Percent B	B - A	B/A	Ln[B/A]	(B - A) * Ln(B/A)
816+	38,701	4.86%	66,569	4.57%	-0.29%	94.05%	-0.061	0.000177
804 - 815	38,789	4.87%	66,295	4.55%	-0.32%	93.45%	-0.068	0.000216
793 - 803	40,721	5.12%	66,574	4.57%	-0.54%	89.39%	-0.112	0.000608
782 - 792	40,725	5.12%	65,363	4.49%	-0.63%	87.76%	-0.131	0.000818
771 - 781	39,651	4.98%	62,829	4.32%	-0.67%	86.64%	-0.143	0.000954
760 - 770	37,991	4.77%	59,209	4.07%	-0.71%	85.22%	-0.160	0.001129
748 - 759	38,919	4.89%	60,849	4.18%	-0.71%	85.49%	-0.157	0.001112
734 - 747	42,167	5.30%	66,977	4.60%	-0.70%	86.85%	-0.141	0.000982
720 - 733	39,690	4.99%	64,160	4.41%	-0.58%	88.39%	-0.123	0.000714
705 - 719	40,024	5.03%	66,878	4.59%	-0.43%	91.37%	-0.090	0.000392
689 - 704	40,073	5.04%	69,966	4.81%	-0.23%	95.47%	-0.046	0.000106
673 - 688	38,195	4.80%	68,970	4.74%	-0.06%	98.74%	-0.013	0.000008
655 - 672	40,601	5.10%	76,668	5.27%	0.17%	103.25%	0.032	0.000053
636 - 654	39,706	4.99%	77,501	5.32%	0.34%	106.73%	0.065	0.000219
615 - 635	39,527	4.97%	78,447	5.39%	0.42%	108.52%	0.082	0.000346
591 - 614	39,692	4.99%	79,959	5.49%	0.51%	110.15%	0.097	0.000490
561 - 590	41,173	5.17%	85,277	5.86%	0.69%	113.25%	0.124	0.000853
524 - 560	39,855	5.01%	88,047	6.05%	1.04%	120.80%	0.189	0.001968
468 - 523	39,503	4.96%	93,282	6.41%	1.45%	129.12%	0.256	0.003694
<468	40,148	5.04%	91,657	6.30%	1.25%	124.83%	0.222	0.002779
Total	795,851	100.00%	1,455,477	100.00%	**Population Stability Index:**			0.017617

What about the big data table on a slide.

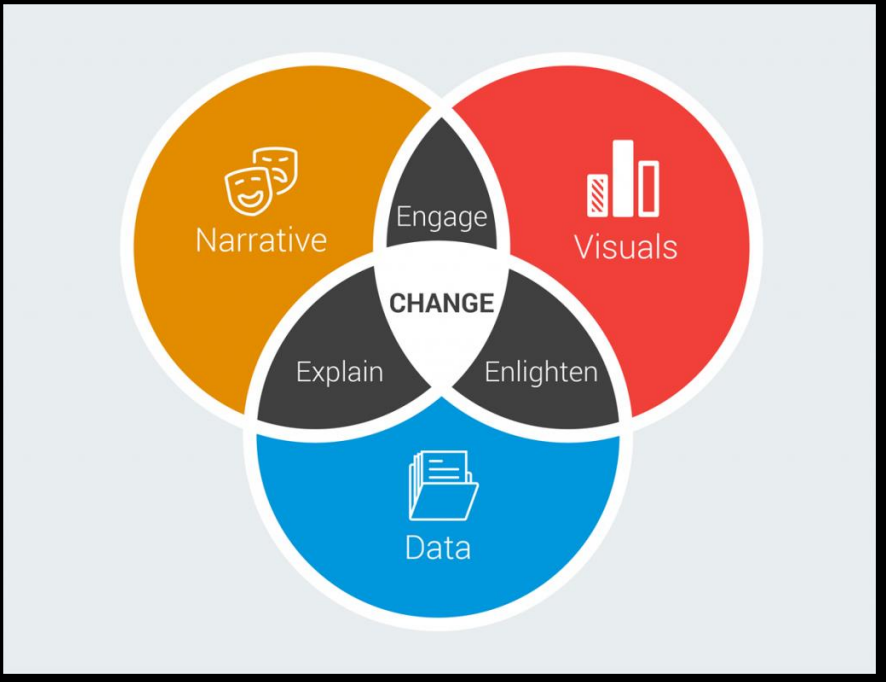
That last one gets me every time. Usually two things happen with this one:

1) The person presenting this chart will put it up and say something like “so I know this is a bit of an eye chart…” Well if you knew then why did you leave it like that haha.

2) They do this thing where they circle the one place they want you to look – so even though I’m showing all this extra data, ignore it and look over here at this tiny number in the lower right hand corner.

And that's a pretty frustrating experience for you as an audience member, because first of all you're super bored which is painful on it's own. But then you're also not getting the full value of the data, right? You want to leave a presentation with a sense that you've learned something, or that you have a recommendation of what to do next based on the analysis. What you don't want is a big stack of data

And if you're guilty of these don't feel bad, the reason I was able to quickly find so many examples is because I myself have done this in the past. Part of the learning process.

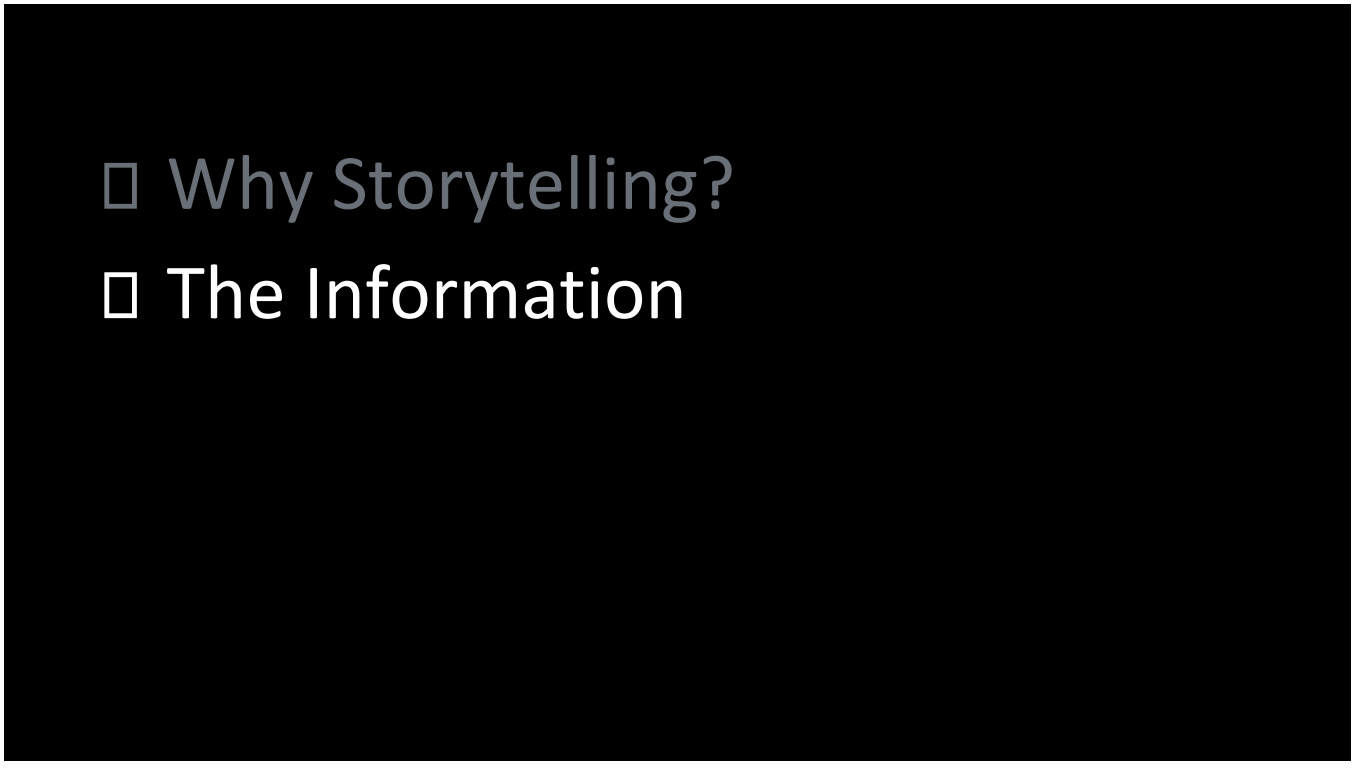


Source: https://www.forbes.com/sites/brentdykes/2016/03/31/data-storytelling-the-essential-data-science-skill-everyone-needs/#23a0307552ad

So ultimately then what do we want a story to do? We want to impact CHANGE. Just like Jennifer Newcomber talked about this morning this is ultimately the goal of our efforts. Trying to get funding, or trying to influence a policy change. To make this happen we need the relevant data, with clear visuals, and an engaging narrative. Easy right? No problem.

Take a look at this diagram – you'll see that while the goal is to be in that center intersection, being in the intersection of two of those is not a bad place to be. Oftentimes as data analysts or data scientists we really find ourselves in that intersection of data and visuals – so what we're doing is enlightening which is not all that bad. We're brining some insights to the fold which may be useful for others. For me, really the challenging part is weaving in that narrative.

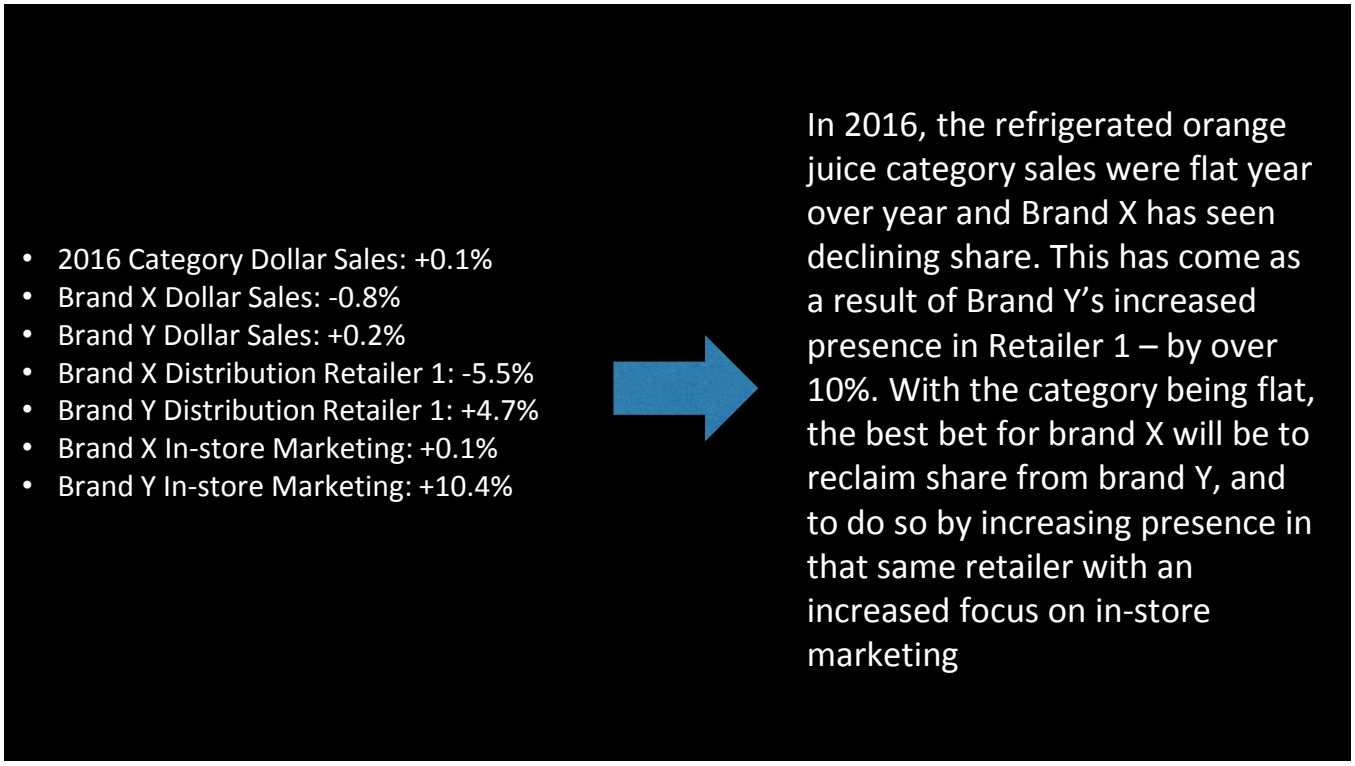
So with that let's talk about how we can get to that middle section there.



So starting with the data, or what I'm calling here Information, you'll likely have some sense of what you're looking for when you dive into the data, or what you're seeing as you go through the data. Your intuition is key here as I'm sure you've all gotten questions where you immediately start nodding your head saying yep I think I know what this answer might be and how I can get to it. And I would encourage you as you go along to keep writing and re-writing the story.

Step 1: Have a story to tell

Once I feel like I have a sense of a story, I'll write it out. BEFORE I've put together any slides, or any formal visualizations of the data, I start with a story. That story is based on the data, and will ultimately help you decide what exactly to put into your presentation. And when I first learned this it sort of blew my mind, because I'd never been taught this before. My process used to be analyze the data, put together 80 slides, then reorganize, tweak, weave those slides into a story, ultimately keeping 20-30 of them and putting the rest in the appendix. I had it all backwards!



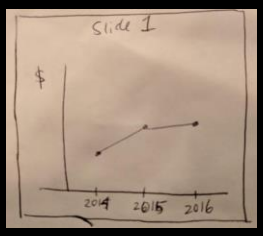
Here's an example. On the left I've got a list of data points. Now, these data points together give a lot of relevant information, but where we want to get to is the story in the right. If I were to ask a colleague week from now, or a month from now, what those data points on the left were, it's very unlikely that they'd remember. The story on the right however, forms a narrative, or a short story, that you'll take with you going forward.

As a quick example of that, how many of you listening to the panel this morning – and heard Quinn give his talk? Just shout it out loud when Quinn was talking about his experience surveying Westminster – what percentage are having to make a decision between paying rent and paying for necessities?

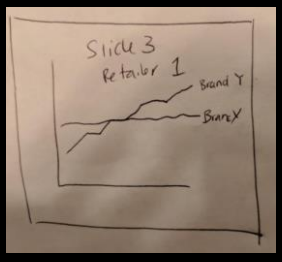
That blurb on the write is what I'll have written down before I've made a single slide or visualization.

An important aspect of this story will be your audience – who is this story being communicated to, and what do you want them to take away from it? Presenting to the head of marketing is going to be much different than presenting to fellow analysts. One will want more high level, while the other will be very interested in the detail. So keep that in mind when formulating this initial story

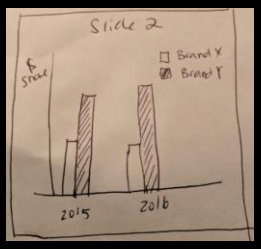
In 2016, the refrigerated orange juice category sales were flat year over year



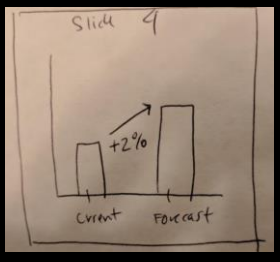
This has come as a result of Brand Y's increased presence in Retailer 1.



Brand X has seen declining share

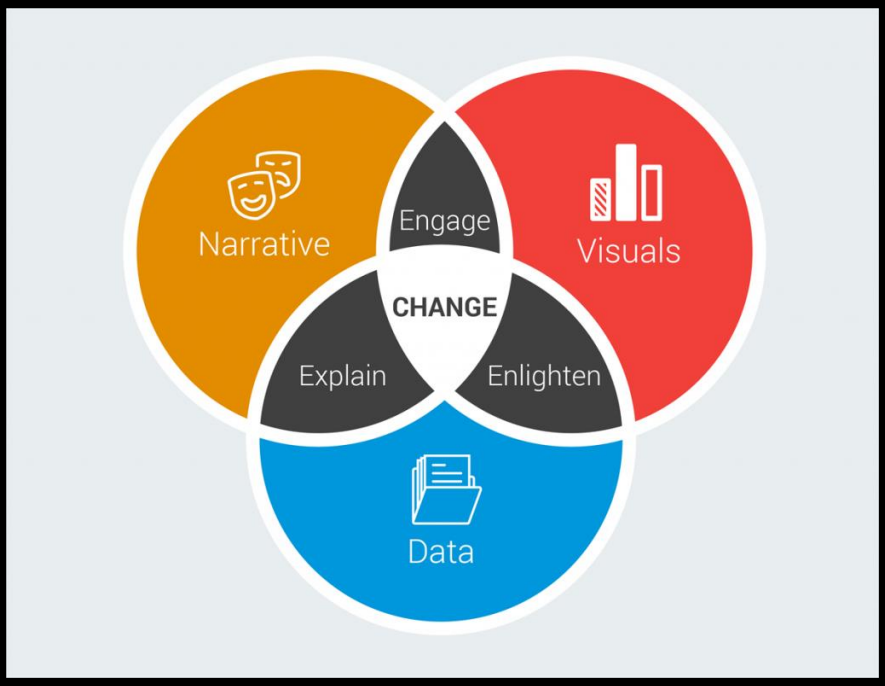


The best bet for brand X will be to reclaim share from brand Y: increase presence in retailer with an increased focus on in-store marketing



Now that you have this written out, what it gives you are the titles to your slides. Now you can storyboard! So all we've done now is taken our story and broken it up. Along with your storyboard, you can start to hand sketch your visualizations. It will allow you to focus on what's important to show on each slide based on the headline that is a part of the broader story.

You can come back and review this step when you're nearing completion of your presentation. Next time you're putting one together, take all the headlines and pasted them into a word document one right after another. Does it read like a story? Good way to go back and check.

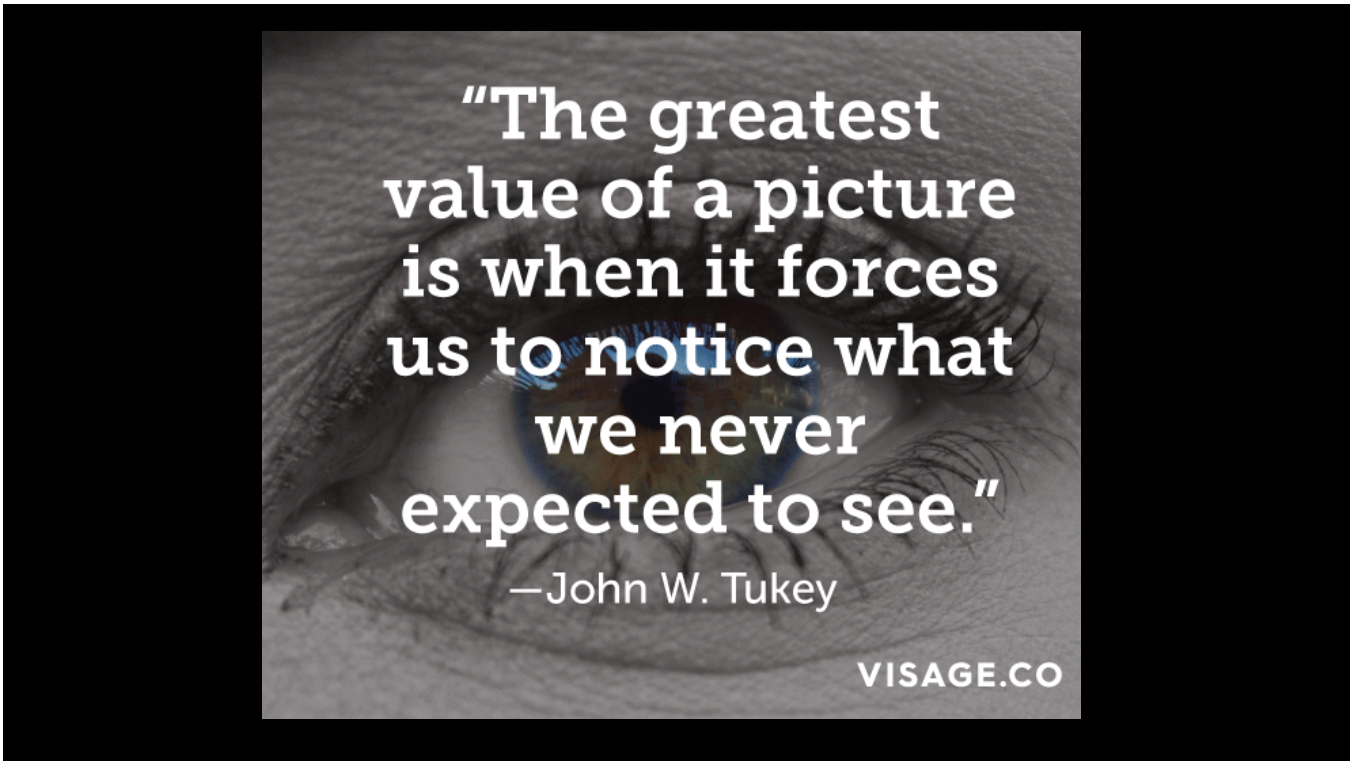


Source: https://www.forbes.com/sites/brentdykes/2016/03/31/data-storytelling-the-essential-data-science-skill-everyone-needs/#23a0307552ad

Now we're getting somewhere! Thinking back to our earlier diagram at this point we're right in that area between data and narrative. We've got the narrative and the relevant data, and we're starting to get into the visuals piece. As you can see just by reading the story, you've got a pretty good explanation of what's going on with this analysis and what the results are. We can now focus on the visuals to bring this explanation to life.

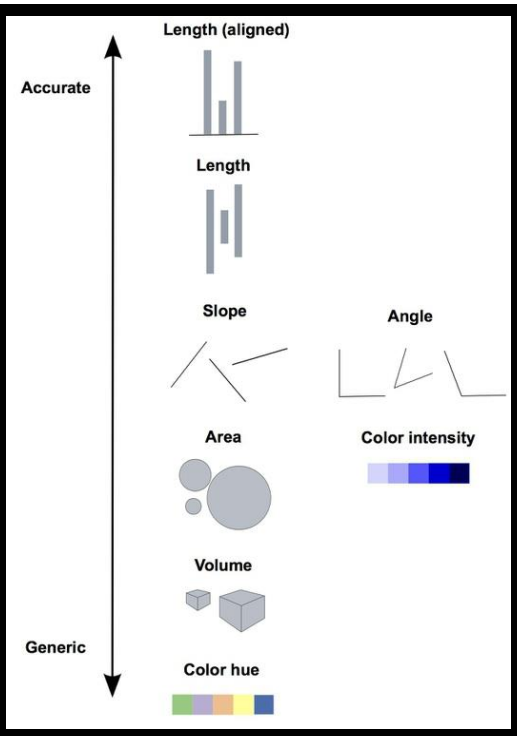
- Why Storytelling?
- The Information
- The Visualization

Now,



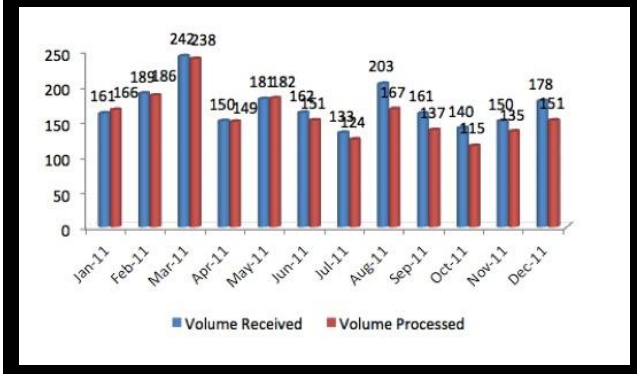
This probably goes without saying for this group but visualization is a very important part of the story. People are very visual, so being able to see information in graphical format will help increase their understanding… when done the right way!

That is, first and foremost the right way means that you present the data objectively and without bias. It is far too common these days to see misleading visuals coming our way so we will be determined not to fall into that crowd of lying visualizations. I think the closing keynote this afternoon is actually about the misuse of data so I won't dwell too much on it here, but there's definitely enough of that out there to fill a whole presentation. I also promise not to use this time as a rant against pie charts.



Source: http://paldhous.github.io/ucb/2016/dataviz/week2.html

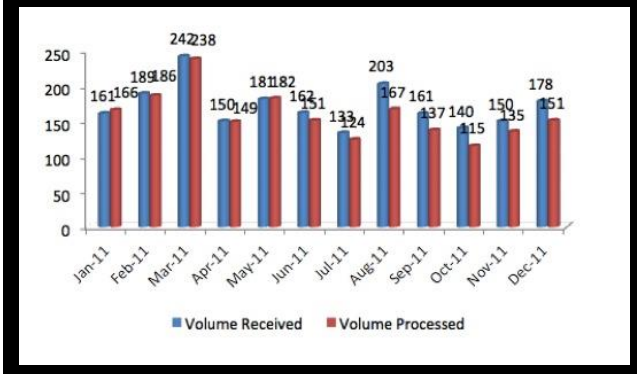
So what are some good things we can do when it comes to visualizing data? Well we can pick the right visualization and to do that we start by understanding the idea of visual encodings. A visualization is essentially mapping data to a graphic. When we do that there are multiple ways we can do that – but humans are much better and perceiving differences in some encodings vs. others. Things like position relative to others, and length are very accurate. It's why we can look at an x-y graph and tell differences almost immediately. Same with length, looking at a bar graph humans are very good at differentiating between lengths. As we go down this scale we get less and less accurate. A bar that's 3 units long is 2 longer than a bar that's one unit long – but how much bigger is blue than red? That's why we consider it more generic.

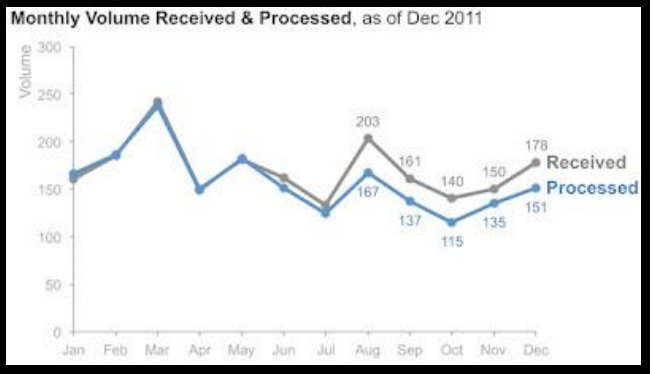


Source: http://www.storytellingwithdata.com/blog/2012/06/drawing-attention-with-data-labels

The other thing I'll say about creating visualizations is that simpler is better! It's very tempting to show every data point on the slide. But stop and think – hey, do I NEED these labels, or these data points, or these colors? Here's an example of a visualization that's trying to tell us SOMETHING. Again this is another example of something I used to present. I thought it was important, there's a great point in here! After you stare at this for a while, what you may start to notice is that the height difference between be blue and red bars increases a big as we go through the year. The story that's being told here is that there was an unexpected spike in volume received in August, and the team has been behind ever since. They need more resources to deal with the backlog.

This slide is confusing enough where even if you were presenting it person, and made that claim, people would be listening less to you and more focused on trying to read the chart. The chart is there to support you, not the other way around. How might we improve this chart?

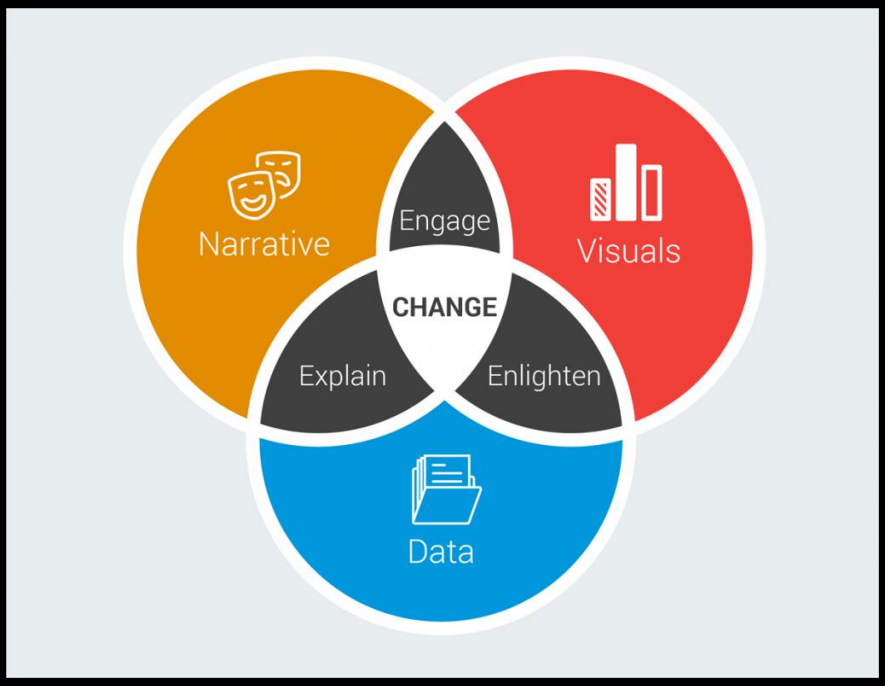




Source: http://www.storytellingwithdata.com/blog/2012/06/drawing-attention-with-data-labels

You'll see quite a few changes have been made. Notably, it's been changed to a line chart – lines are very good at showing something happening over time. Also a lot of the labels have been taken away. If the key point of this chart is to show how far behind the team has been since August, show the data from August on! I don't need to know the exact in/out volume in Jan, Feb, just that they were about equal those months so no one was falling behind.

So to reiterate – you don't want your visualizations to be a distraction while you present, you want them to be your support!



Source: https://www.forbes.com/sites/brentdykes/2016/03/31/data-storytelling-the-essential-data-science-skill-everyone-needs/#23a0307552ad

So that's it, right? We've got the relevant information with the engaging narrative, and the clear visuals! And sure, if this is an analysis you're emailing to someone then yep, this covers it. But what about when you have to stand in front of a room and actually tell the story…out loud…to other people?! This is where we get to REALLY make this come alive!

- Why Storytelling?
- The Information
- The Visualization
- The Presentation

This is where you get the audience to FEEL things about the analysis, and about their need to take action. An impactful presentation can hold their interest as you're taking through your story that you've carefully crafted. Think of this as the difference between you reading a book silently to yourself, vs. listening to it on audiobook being read by the author. The audience will get the sentiment straight from the source! The very best way to do that…

You’re a person… talking to other people

is to think of it in this context. This is my less cheesy way of saying “be authentic” I’m always amazed when I can talk people one on one, or even in a group of friends, and they seem totally normal – only when they go up to present all of a sudden they turn into Data Robot 5000 or a ball of nerves. They feel the need to change for some reason. And I think this is where having that story really helps out! When you’re no longer compelled to list out data points, and cover off on every minute piece of information, you’ve gone from someone giving a presentation to someone telling a story to their friends at a party.

This is where I think

Practice…out loud!

And once you start feeling more comfortable with that first part, you'll be able to clean up or refine your presentation. And that will come with practice – too few people practice their presentations out loud before they give them. A lot of times it's a time crunch, where you're finishing the presentation minutes before the meeting then you rush in there only remembering the most recent slide you just worked. Give yourself time to talk through the presentation.

When I say practice I don't mean doing that thing where you sit at your desk and flip through the slides and think to yourself okay on this slide I'm going to talk about this and next I talk about that etc. NOPE doesn't count! I mean actually telling the story out loud. Ideally to a colleague or a friend and getting feedback. First couple times you might even go so far as to write a script. When you say things out loud you'll start to notice things that don't sound natural. While practicing this presentation I'm giving you today I read a note out loud and I immediately said oh wow that's going to sound out of place. So it's important to say things out loud

Get comfortable with silence

Now even if you follow JUST the first item I mentioned about talking to other people, you'll be in pretty good shape from a presentation standpoint. Add to that some practice and now hey you're starting to get really good. If you want to really knock some socks off, start to get comfortable with silence. What I mean by that is, feel comfortable speaking slowly and letting some words hang with nothing after them. It's natural for us to want to fill the silence with words like um, uhh, kind of - and honestly lots of fine presenters still use those. But if you listen to a REALLY good communicator, you'll notice that they will say something and then let it land. And there's something interesting about those pauses, because you'll see people focused. Or you might physically see them lean in just slightly.

And that's when you'll know that you have their attention, and that's the way you can really tell a great story.

CONNECT & LEARN MORE!

Amer Tadmori
https://www.linkedin.com/in/amertadmori/
amer.tadmori@gmail.com

General Assembly

Upcoming Workshops

SQL Bootcamp: October 21st & 22nd
Data Visualization & Intro to Tableau: Nov 7th
Intro to Data Analytics: November 13th and December 4th
Crafting Stories with Data: December 5th
https://generalassemb.ly/education?format=classes-workshops