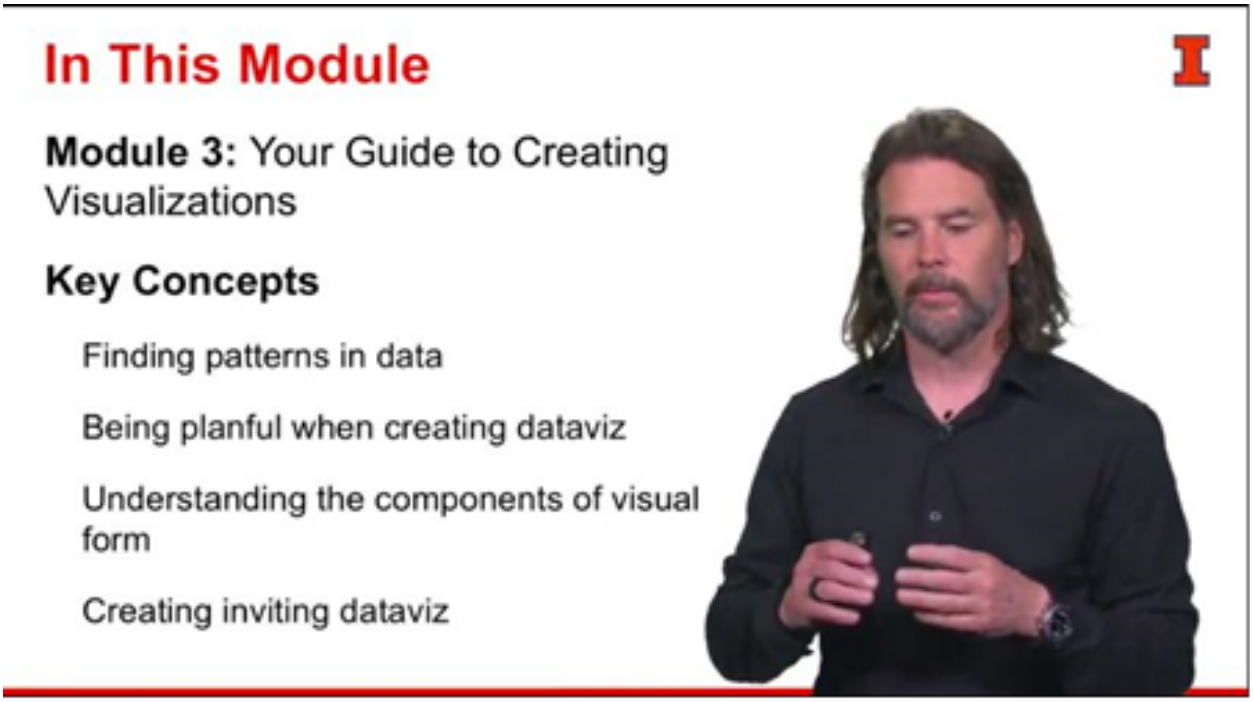
Module 3: Finding Your Data Story

Table of Contents

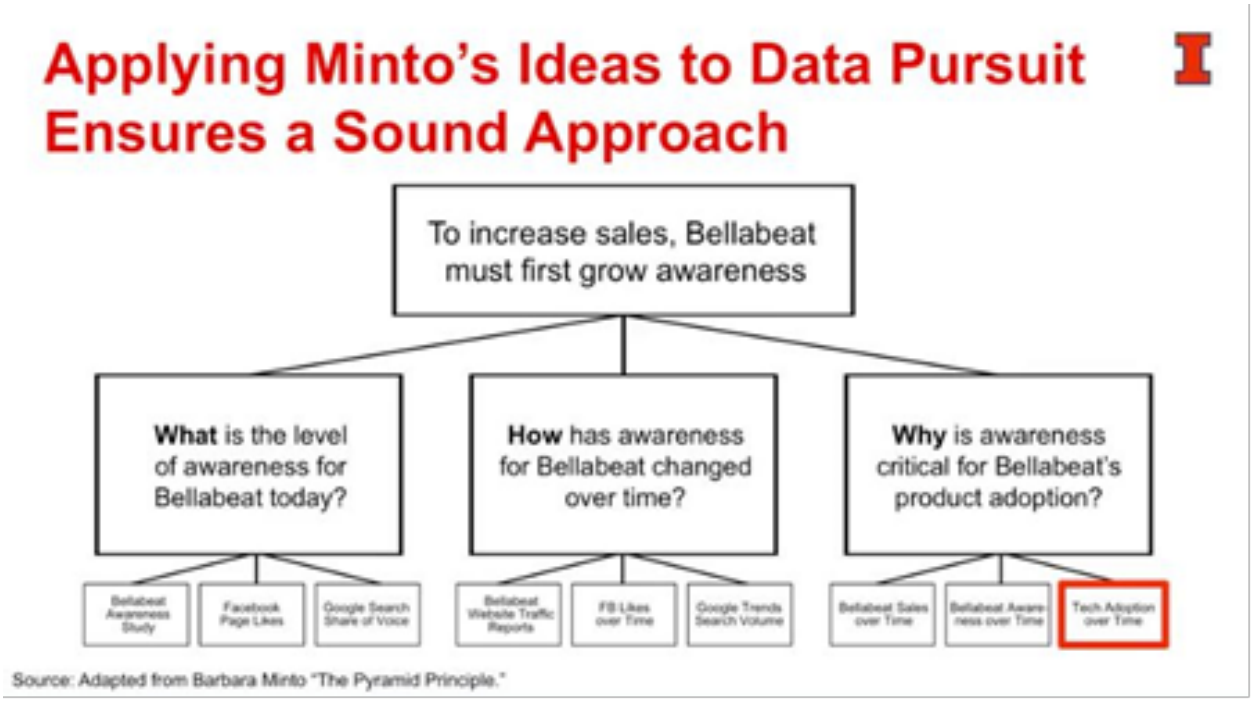
Lesson 3-0: Overview

Lesson 3-0.1: Overview



Module Three, you're guide to creating visualizations. So, now we have moved through many elements of the framework we're using to define good data visualization. In this module, we're going to really focus in on the visual form. To do so, we'll talk about a few key concepts, we'll talk about finding patterns in data, we'll talk about being planful and the way that you approach the creation of your data visualization, we'll talk about having the understanding of the components of visual form, what is good visual form, we'll introduce a framework that will do that. Then, we'll talk about creating inviting data visualization database. We'll give you a definition there.

In this module again, we're using this framework, which has guided us along the way and shown us the elements of a successful data visualization.In the next lesson, first lesson of this module, we will be focused in on the visual form. Actually, from here on out, we're going to focus on the visual form.

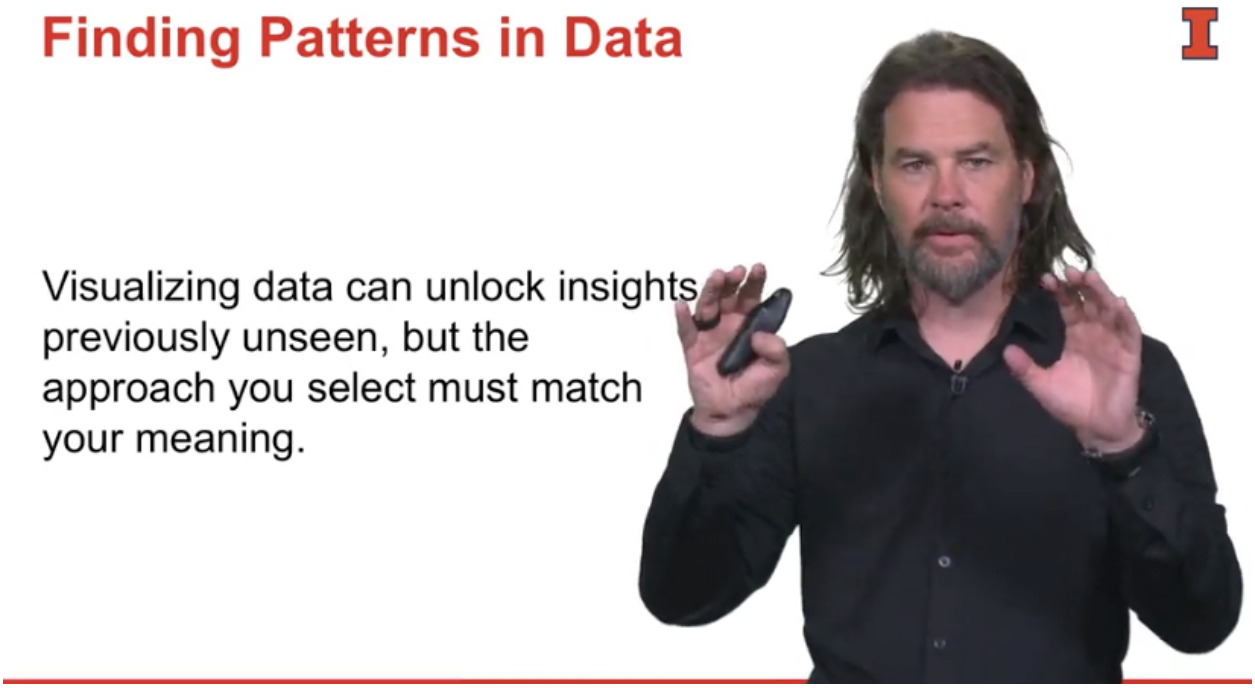


Remembering our Bellabeat case study, we have constructed an outline for a story. We identified an objective, we found some key questions that we know are missy and pretty robust and complete. For each of those key questions, we've identified a data source that will help us

answer that. Now, we're going to focus in on this last piece of data tech adoption over time, which will help us in our case study, answer the question, "Why is awareness critical to Bellabeat's product adoption?' Through this data and through the analysis, we helped show that adoption over time of technologies have led to success for those technologies and maybe even give us a sense for how quickly we should expect the adoption of Bellabeat's new technologies to take place across our market. Now, we have also found data. That's where we are, and as we move through this module, we'll start to analyze this data, seek some patterns from it, and then turn those visualizations into client-ready database.

Lesson 3-1 Finding Patterns in Data

Lesson 3-1 Finding Patterns in Data



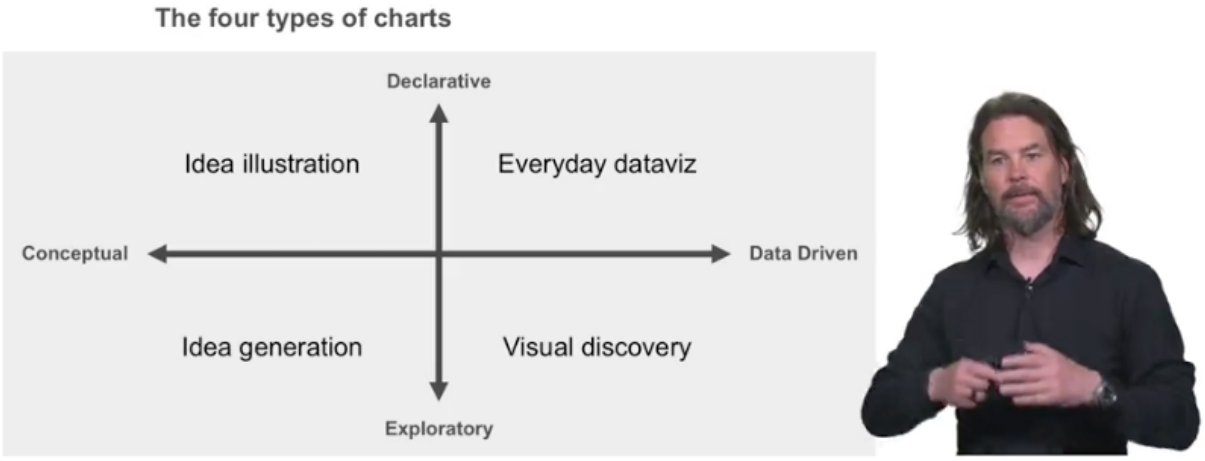
Finding patterns in data. Now, that we have data, we want to start to identify the stories that we see and the best way to do that is to visualize that data and see what patterns jump out. There will be things that come to us through visualization that we wouldn't be able to see when the data is in tabular form. But before we dive in, let's get an understanding of the different types of charts that we will create.

Analyzing Data Effectively Begins with Understanding

Source: Adapted from Scott Berinato, "Good Charts: The HBR Guide to Making Smarter, More Persuasive Data Visualizations."

A framework is laid out very well by Berinato and he asks a couple of very important questions to assess where you are on your communication journey and then what sort of chart you should use. The two questions are, are you at a point where you are conceptual or data-driven. Do you have data or not? Then, the second is are you at a point or stage where you can be declarative with a message, you have something that you want to tell and communicate or you're still exploring and looking for that data? Depending on how you answer those two questions determines where you are on your journey, and therefore, the kind of effort that you need to put in and the type of chart that you need to create

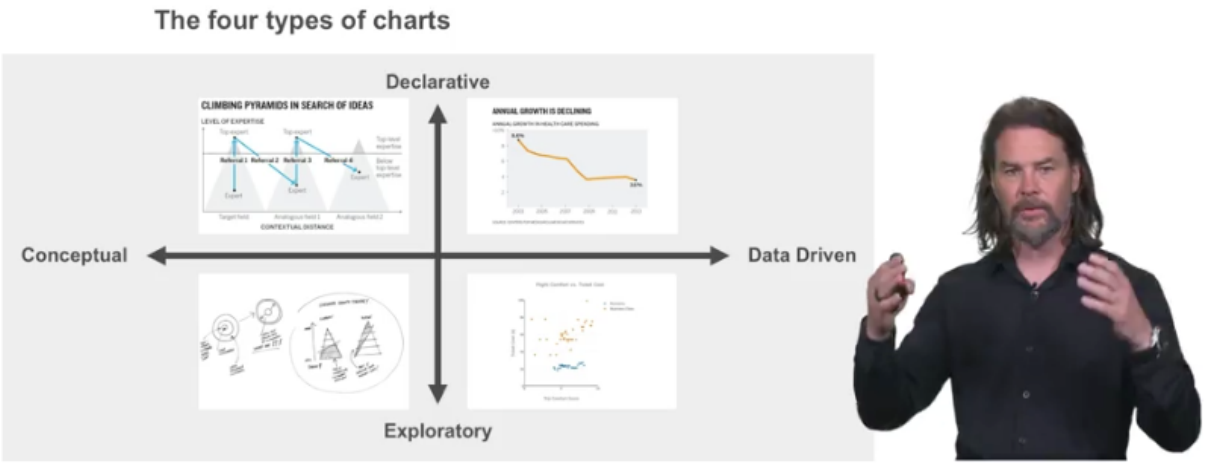
Each quadrant requires different forms of visualization.



Source: Adapted from Scott Berinato, "Good Charts: The HBR Guide to Making Smarter, More Persuasive Data Visualizations."

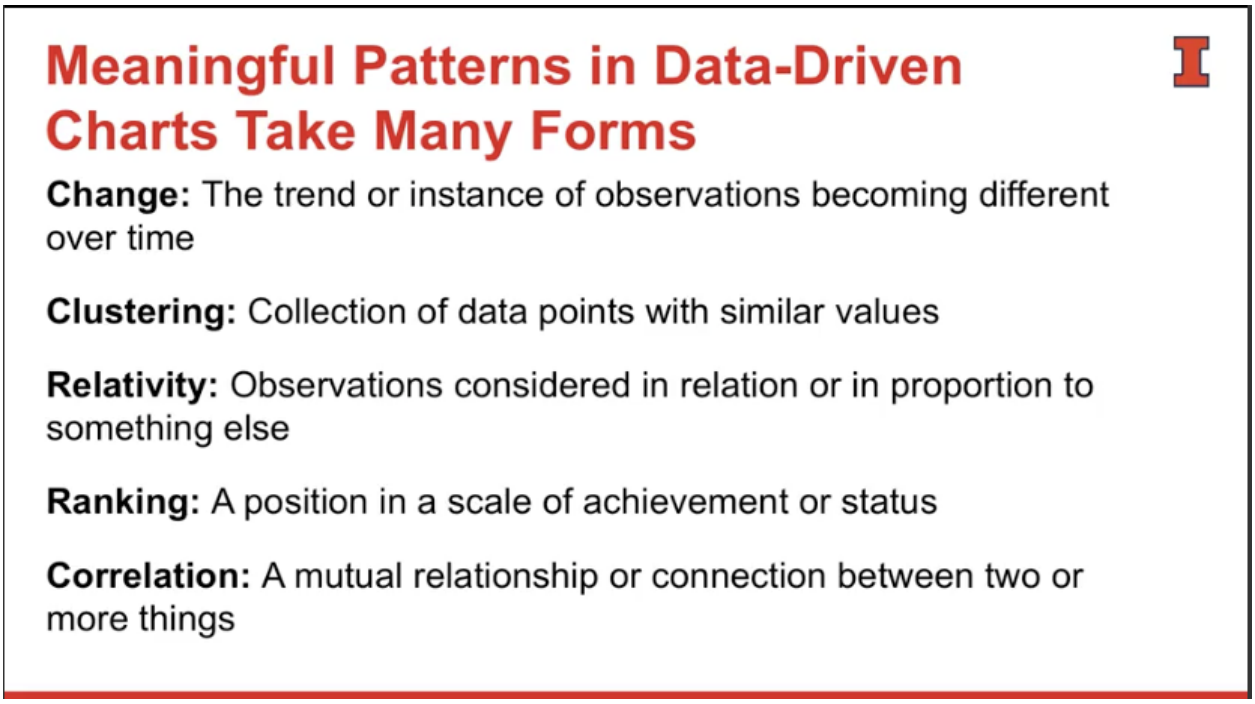
So, those two questions then translate well into this two-by-two. At each phase, there is a different type of chart that you will be creating. If you are declarative but yet conceptual, so you are trying to tell the story through not data but through concept, you are creating what Berinato would call idea illustration. Idea generation is conceptual, still exploratory, still trying to figure out how we put a concept visually in front of an audience. Visual discovery is where we are being exploratory with a dataset. So we're looking into that data trying to find our story and everyday database what I might call client ready database, is a declarative statement based on data. We have a story we want to tell, we want to get that in front of a client and audience.

Data storytelling is best done when charts are involved.



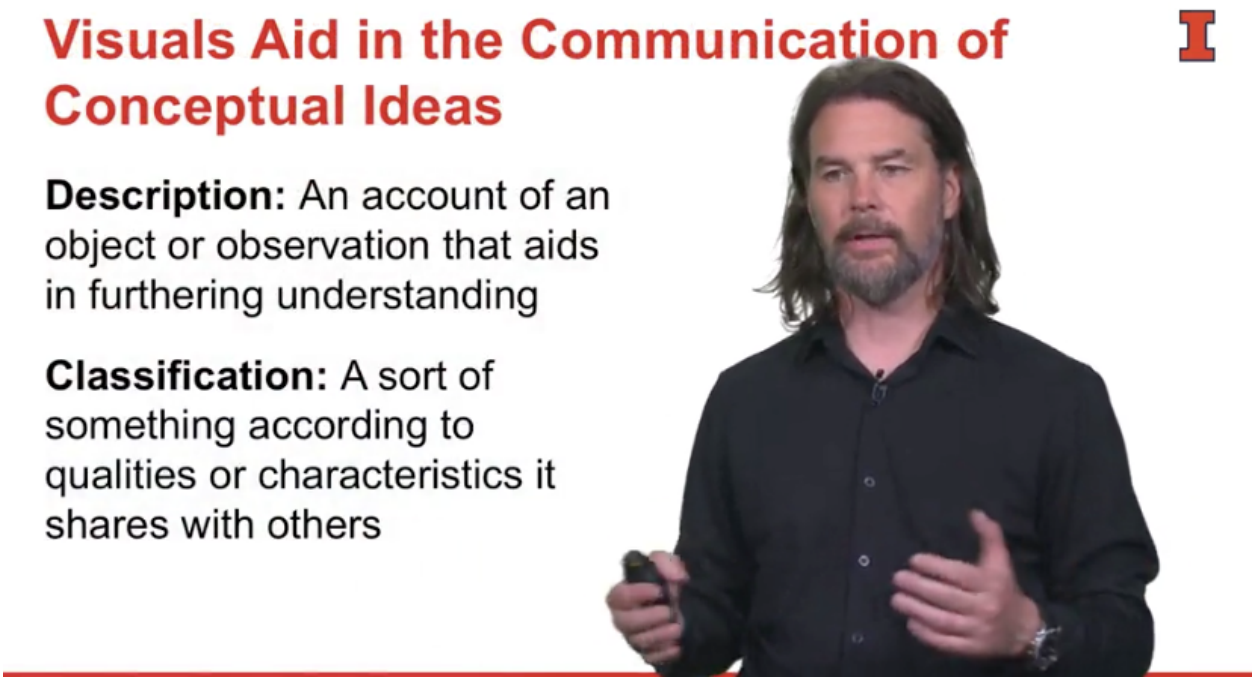
Source: Adapted from Scott Berinato, "Good Charts: The HBR Guide to Making Smarter, More Persuasive Data Visualizations."

Visual discovery is where we are being exploratory with a dataset. So we're looking into that data trying to find our story and everyday database what I might call client ready database, is a declarative statement based on data. We have a story we want to tell, we want to get that in front of a client and audience. At each of those phases, there's a different type of chart we'll create. Conceptual declarative is something that we want to put forward, so we do need to put some effort into creating a conceptual chart like that. Conceptual exploratory, we don't really know what's going to stick, this is in a room with your your teammates brainstorming and drawing ideas on a white board. The things that are more important for us in this course are on the right side of this framework. This is where we are being data-driven, and we will really be talking about two different phases. We're either in exploratory mode, we have data, we don't know what the story is yet, or we are in declarative mode with our data. We have our story set, we want to communicate that story, and that's going to have great implications on the amount of time that we spend polishing our database. Where we are today, is simply looking for patterns. So, there is no need to polish these visuals. What we're going to create is work product is for our eyes only, it is allowing us to test the hypothesis that we have or answer the questions that we've asked looking for patterns.

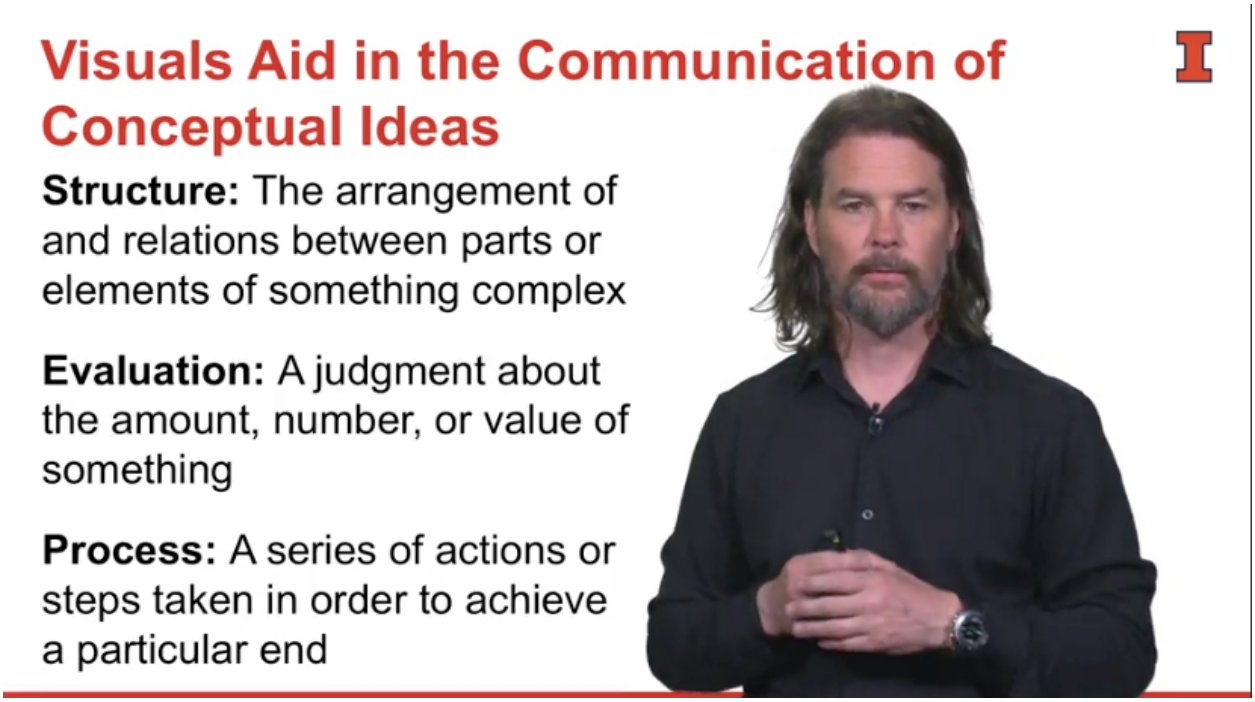


When we do that, there are five different types of patterns that we will be looking for. We will look for change, either change over time and trend or sudden change. We will look for clustering, collection of data points that are similar to one one and yet different from others. We will look for relativity, how two different data points relate to one another. Ranking, what is best, what is worst, what is at the top of the scale, what's the lowest of the scale and and everything in between. Then correlation between datasets. How does one set of data influence or impact another. Those are the patterns we should be seeking because each of them reveal a story. Now, when we are looking for those patterns, there are certain visual techniques that will reveal them.

When we do that, there are five different types of patterns that we will be looking for. We will look for change, either change over time and trend or sudden change. We will look for clustering, collection of data points that are similar to one one and yet different from others. We will look for relativity, how two different data points relate to one another. Ranking, what is best, what is worst, what is at the top of the scale, what's the lowest of the scale and and everything in between. Then correlation between datasets. How does one set of data influence or impact another. Those are the patterns we should be seeking because each of them reveal a story. Now, when we are looking for those patterns, there are certain visual techniques that will reveal them. For instance, if we're looking at change over time, a line graph is a fine way to identify that. A pie chart might be great at depicting relativity, it is not going to reveal for us change. So, putting a dataset that we have into a pie chart will not tell us or answer any question that we have relative to change. It's important for us as an analyst to identify the techniques that we'll use to answer the questions that we have and ensure that they are in alignment there. These categories of change and this idea that certain visual forms fit with each discrete category, is not unique to having data.

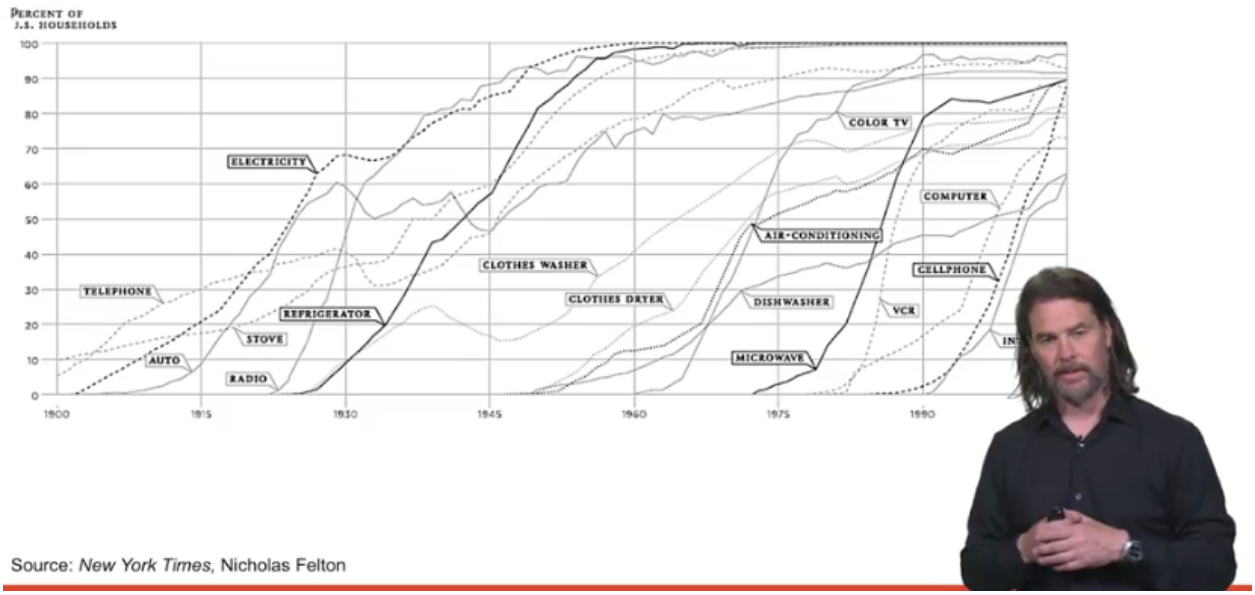


This works on conceptual ideas, as well. When we're dealing with conceptual ideas, we're looking for either a description, a classification, structure, evaluation, some process, and for each of those different conceptual ideas, there is a visual form that works, and then a number of others that wouldn't for that particular conceptual form.



Again, for an analyst, if we are revealing or trying to declare conceptual ideas, ensuring that we're using the right visual technique, the right visual pneumonic, to make that that point is important. So, let's see this idea come to life, we'll use the Bellabeat case study that we've been

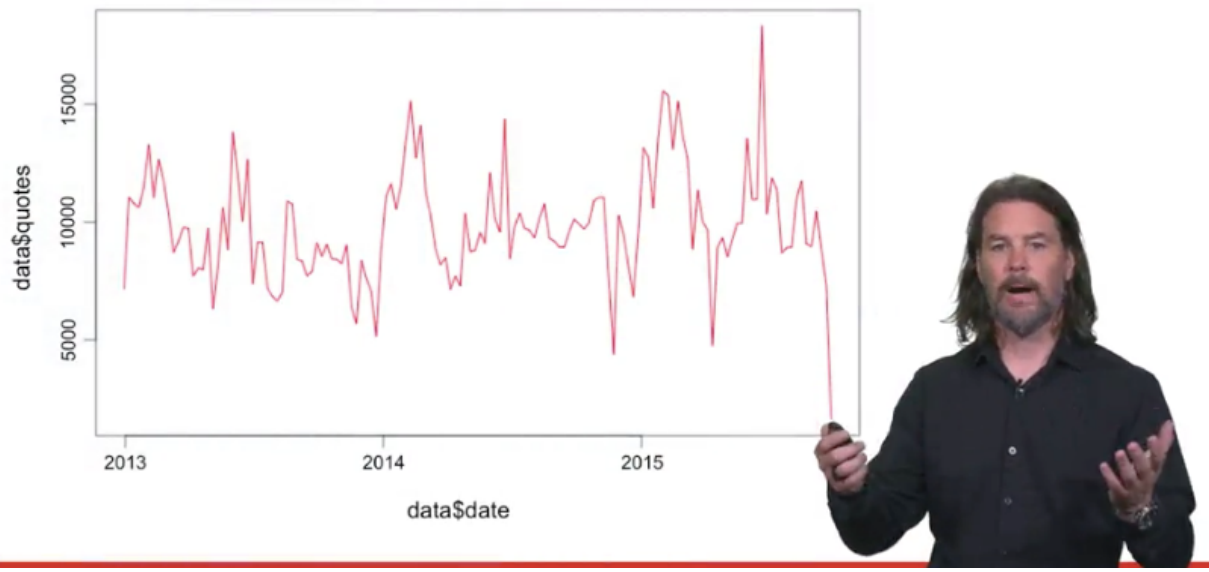
tracing. Again, we're at a point we have data and we're able to start to assess the the adoption of technologies over time as we seek to answer the question of why awareness is important for Bellabeat? If we were to take this data and throw it into a visualization, one of the visual technique that we may use that would work well would be a line chart.



Source: *New York Times*, Nicholas Felton

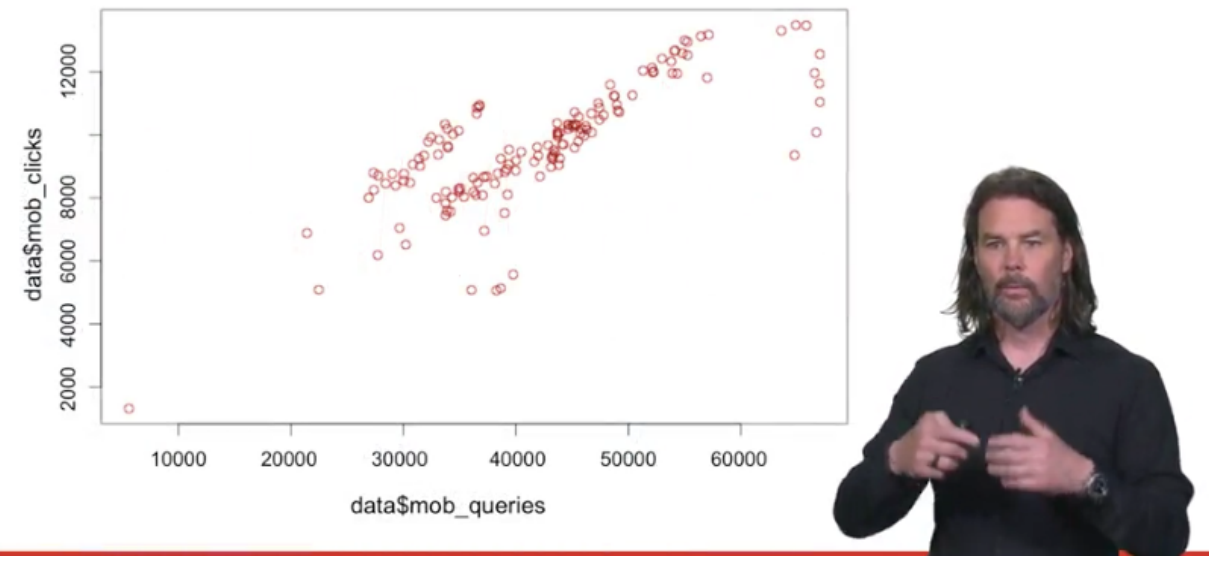
We know from the previous conversation that that aligns. That line chart would show adoption over time and from this point, can start to pick out the interesting stories that will either answer our key question or allow us to know that we need to go back to the drawing board. Now, to do this at this point, one tool that I find useful is R. I talked about R before. But, R does a great job of moving an enormous amount of data and visualizing it really efficiently.

Plot—Time Series



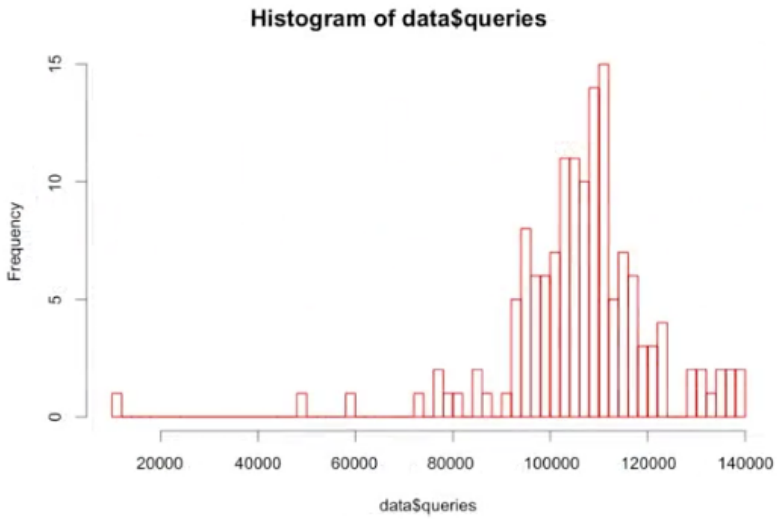
You can create box, time-series plots

Plot—Relationship



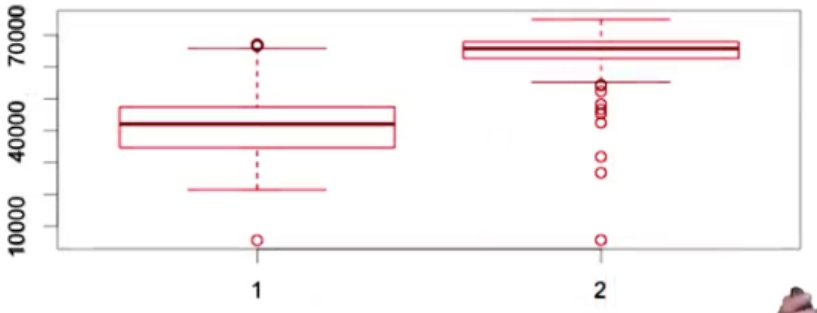
time-series plots or relationship-plots,

Histograms



histograms,

Boxplot



boxplots.

Bonus: Package Gpairs (the Power of R)

Even using the power of R and the community of developers, you can find scripts and new packages like Gpairs, you see described here, that do things that other programs are not able to. In this case, do tremendous correlation across a wide variety of data in one single visual, so that you as an analyst know where your story is and where it is not.

Use Visualization to Reveal Patterns and Stories in Data

Understand the analysis situation you face: Conceptual or data-driven? Declarative or exploratory?

Choose the visualization technique that will most effectively illuminate the pattern you seek.

Use a tool that efficiently creates the visualization you need.

Calibrate the amount of chart polish you apply to where you are on your communication journey.

These things are important to us because the most effective and efficient way for us to get at data and get at data stories is to visualize that data and start to see these patterns. To do that, we do need to understand where we are, right? We need to understand if we're dealing in

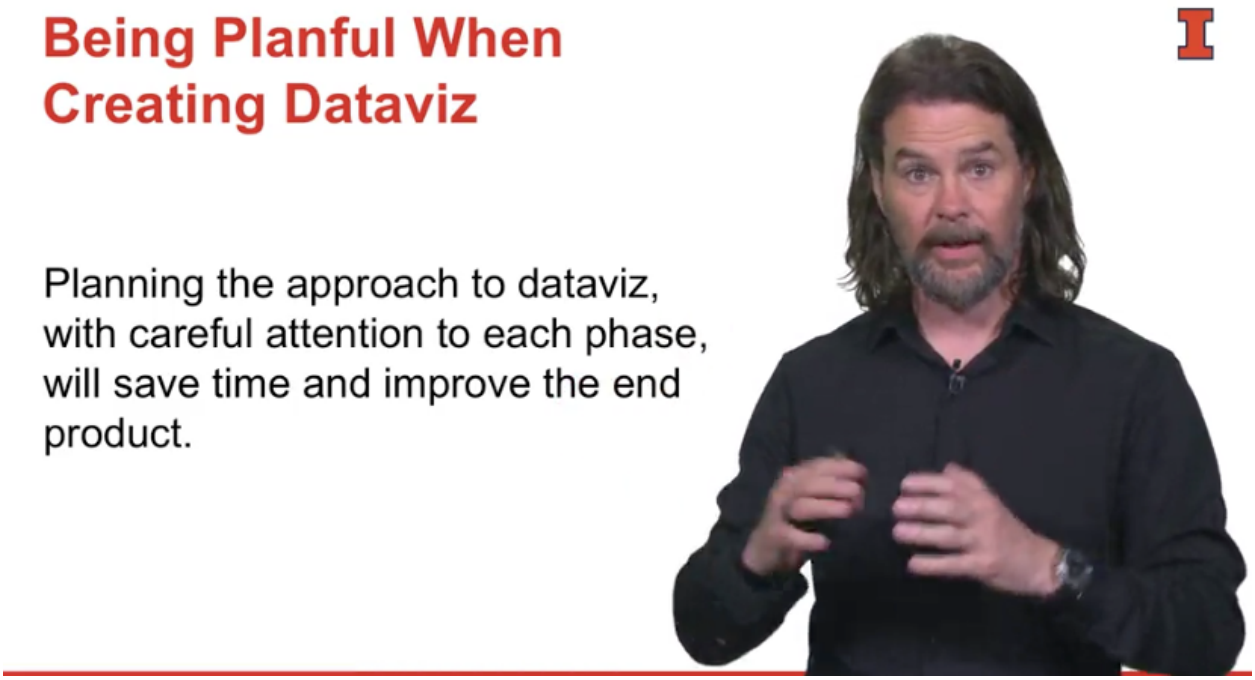
conceptual matters or a data-driven matters. We need to know are we in a state of declaring a story or we still exploring? The answers to those questions dictate which type of chart we use. The visual technique that we apply to seek either to reveal or ultimately to communicate that pattern needs to fit with that pattern. So, as an analyst getting familiar with the techniques that you can use and when to apply them appropriately, is important. The tool that we choose should provide the visual technique very efficiently. At this point, efficiency is very important. We don't want to fumble around with a hard to use tool to get at the visualization we want to create. Learn those tools that do work for you quickly so that you can move on to the next phase of your evaluation in your analysis. Finally, importantly, calibrate the amount of polish you have to where you are on the communication journey. When we are exploratory with the dataset, it doesn't matter if the visual looks great yet. We're not going to show that visual to our clients, to our stakeholders. It's for our eyes only. Once we start to migrate into what Berinato calls every day databases or client ready database, we do start to apply that polish, utilizing pre-attentive atributtes, doing the things that are going to make that message stick quickly, but at this point, is not relevant, not necessary and would only serve to slow us down.

References

Felton, N. (2008). *How Americans Spend Their Money* [Online image]. Retrieved from https://goo.gl/vNNdYH

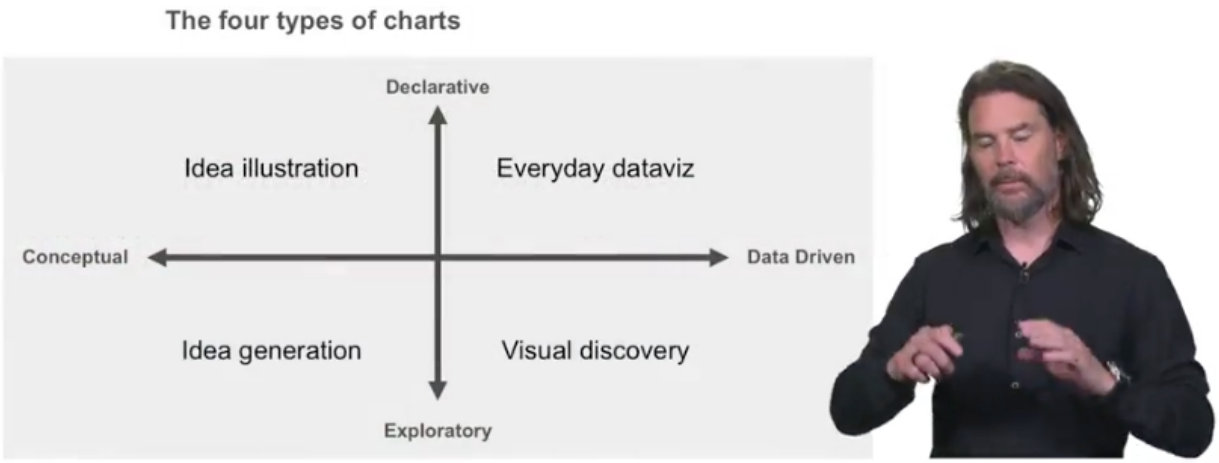
Lesson 3-2: Being Planful When Creating Dataviz

Lesson 3-2.1: Being Planful When Creating Dataviz



Being Planful When Creating Dataviz. So, now we're starting to migrate into this everyday dataviz or this client-ready dataviz, and when we do this, our approach is going to change. We do need to start putting more attention to detail, we need to apply more polish to this visualization because this is what we're going to put in front of our stakeholders. So, the kind of effort is going to be notched up a little bit on this visual.

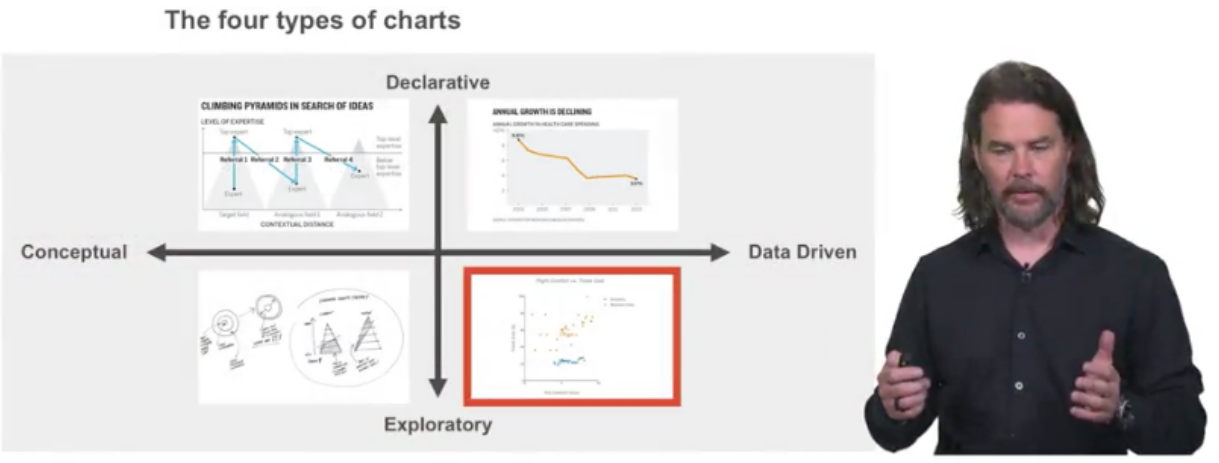
Each Quadrant Requires Different Forms of Visualization



Source: Adapted from Scott Berinato, "Good Charts: The HBR Guide to Making Smarter, More Persuasive Data Visualizations."

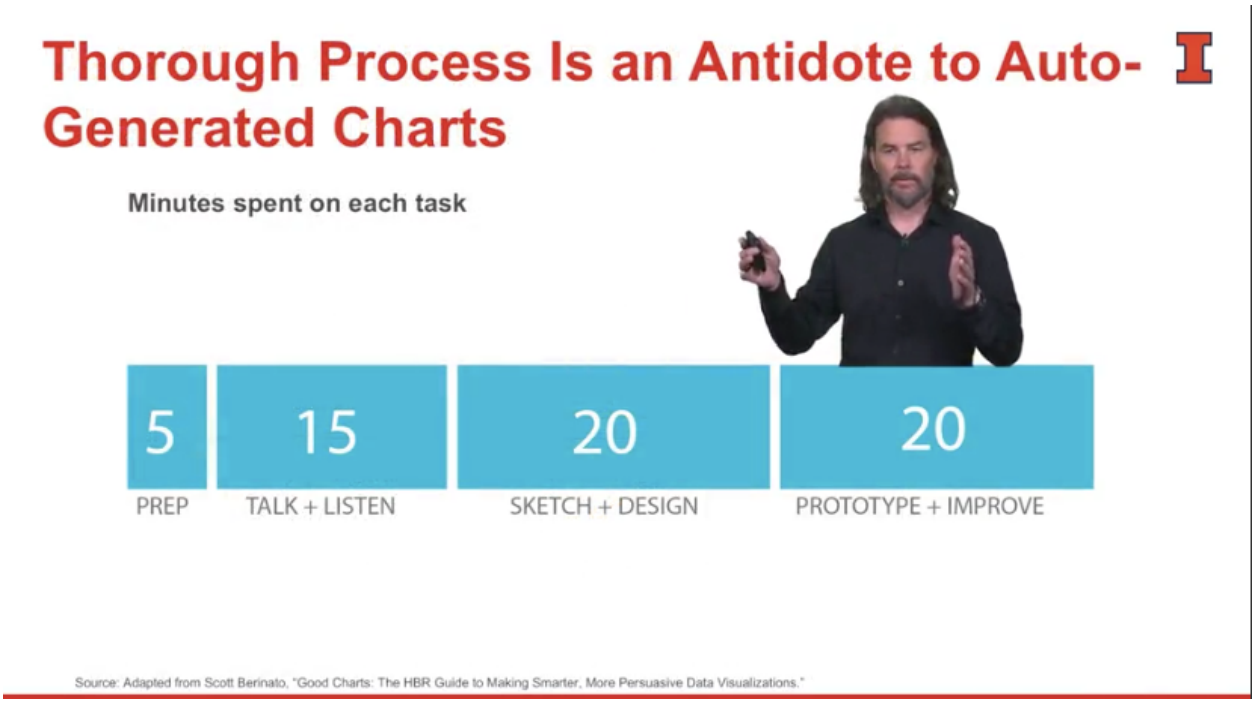
Again, we are moving from a point of visual discovery to everyday database,

Data Storytelling Is Best Done When Charts Are Involved

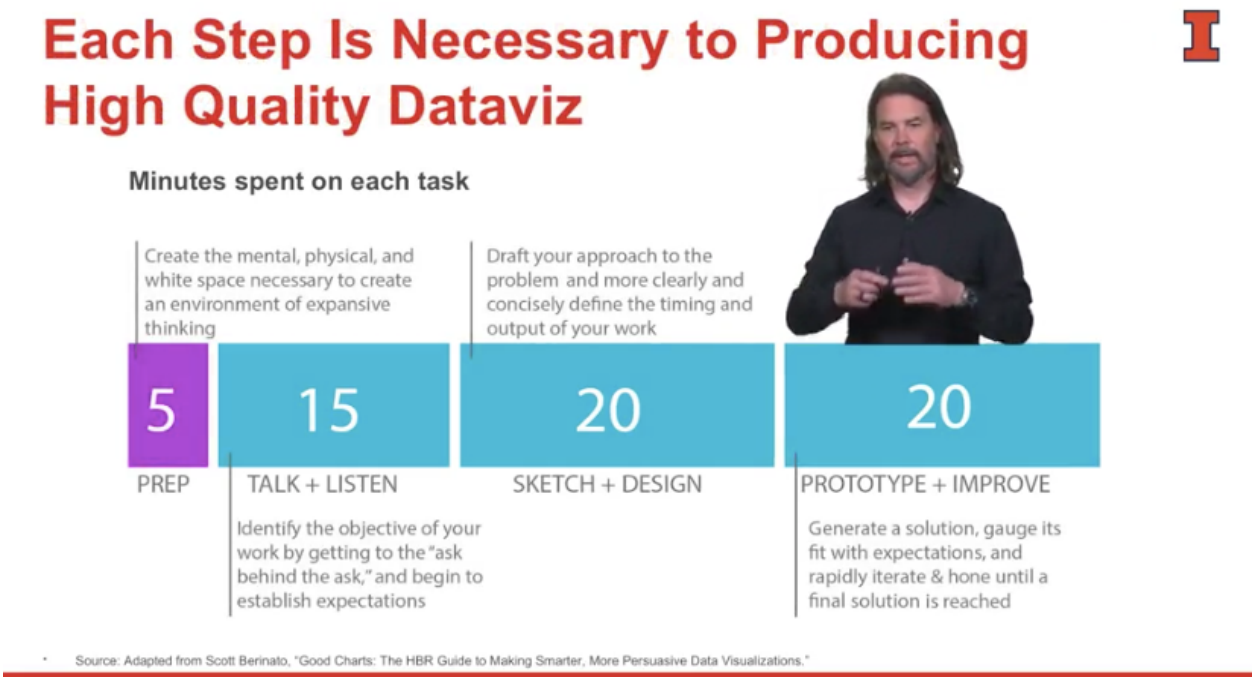


Source: Adapted from Scott Berinato, "Good Charts: The HBR Guide to Making Smarter, More Persuasive Data Visualizations."

and Bernado lays out I think a very solid plan on how we might approach that transition.

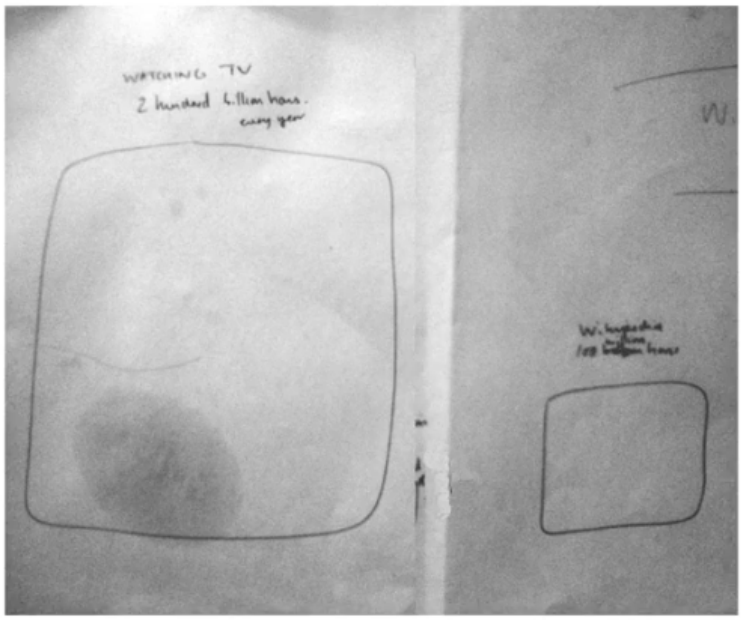


What he says is that if he had an hour to spend creating an everyday database, he would break it up in this way. He would have five minutes of prep where he would actually create



the mental physical white space that he would need to work appropriately. He would then talk and listen for 15 minutes. So, gaining further context, understanding who the audience is that he is presenting to and informing him, in forming the ideas that he'll need to achieve with that everyday database. The second two-thirds of the time are spent actually on the imagery and creating the visual. Half of that time 20 minutes is spent actually drafting by hand kind of an old

world traditional concept but in this case it works very well. Prototyping and improving then is the final one. So, that first five minutes really is used just to clear your head and get yourself ready for work. The talking and listening to the following 15 minutes. The purpose there is to get an understanding of what it is you need to achieve. You've got this great insight, you think you've got a good story, validate that. Work with your colleagues to ensure that it is on mark that it is on path. Once you have that, start then towards the actual development. The sketching piece is a bit of a lost art and I think too often, we take this idea that we have and we rush into our applications and start using our computer to generate graphics. Well, if we just took a moment and with pen and paper, visualize some ideas, we would no longer be limited by our understanding of the the package, the application, or the limitations of the application itself. We could just design where we're relying only on our creativity.



Source: David McCandless

This is something that the the great data visualization artists do. David McCandless certainly does. I shared this image earlier where he with pen and paper drafts every data visualization that he creates. It's a good practice. It's a good habit to build and will make you more efficient and ultimately better at communicating with data.

Each Step is Necessary to Producing High Quality Dataviz

Minutes spent on each task

PREP	TALK + LISTEN	SKETCH + DESIGN	PROTOTYPE + IMPROVE
5	15	20	20

Create the mental, physical, and white space necessary to create an environemnt of expansive thinking

Idenitfy the objective of your work by getting to the "ask behind the asl" and begin to establish expectations

Draft your approach to the problem and more clearly and concisely define the timing and output of your work

Generate a solution, gauge its fit with expectations, and rapidly iterate & hine until a final solve is reached

Source: Adapted from Scott Berinato, "Good Charts: The HBR Guide to Making Smarter, More Persuasive Data Visualizations."

After that is done, 20 minutes, the final one third of this hour is spent prototyping and improving the visual. The idea here being that the first visual you create shouldn't feel like the last visual. It is a starting point. You can begin then in your application to improve it over time and there are techniques that you can apply to make it incrementally better as you go. That is important. Too often we feel like we're putting all this effort into a final database. We should think of it more as a fluid process and iterative process indeed.

Dutifully Approaching Dataviz Creation Produces Better Results

Transitioning visualizations from "work product" to "client ready" requires attention to detail.

Planning the approach to dataviz creation helps ensure each element of the process is sound.

Sketching visualizations will ultimately save time and effort.

Producing a final dataviz is a study in iteration, not a "one-and-done" experience.

So, being planful at this point is important as we transition from this work product output into client ready database, this requires more attention to detail. Planning our approach out and being beautiful about that really does ensure that each element of that process is sound and we'll build ultimately to a great data visualization. Sketching pen to paper is a under-represented art that should not, should be something that you build as a habit of your own. Will save you time, will save you effort ultimately and most importantly will allow you to open up your creativity as you think about the way that you might effectively visualize the data and the data store that you want to tell. Finally, producing that final database, it is a study in iteration. It is not a one-and-done experience. You should not feel like the first database that you're going to crank out is the right one. Put out a database, work with it, apply some of the tools that we'll talk about later on in the course to make it better and better and really treat it like a living, breathing, iterative, data visualization.

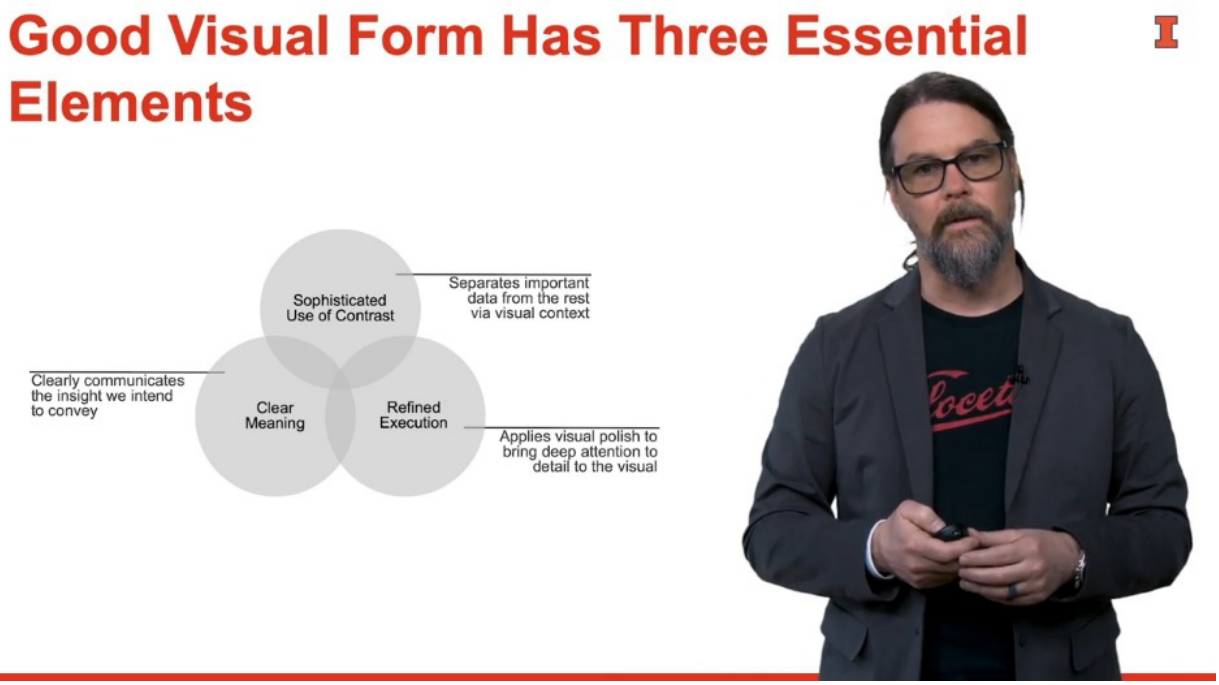
Lesson 3-3: Understanding The Components of Visual Form

Lesson 3-3.1: Understanding The Components of Visual Form

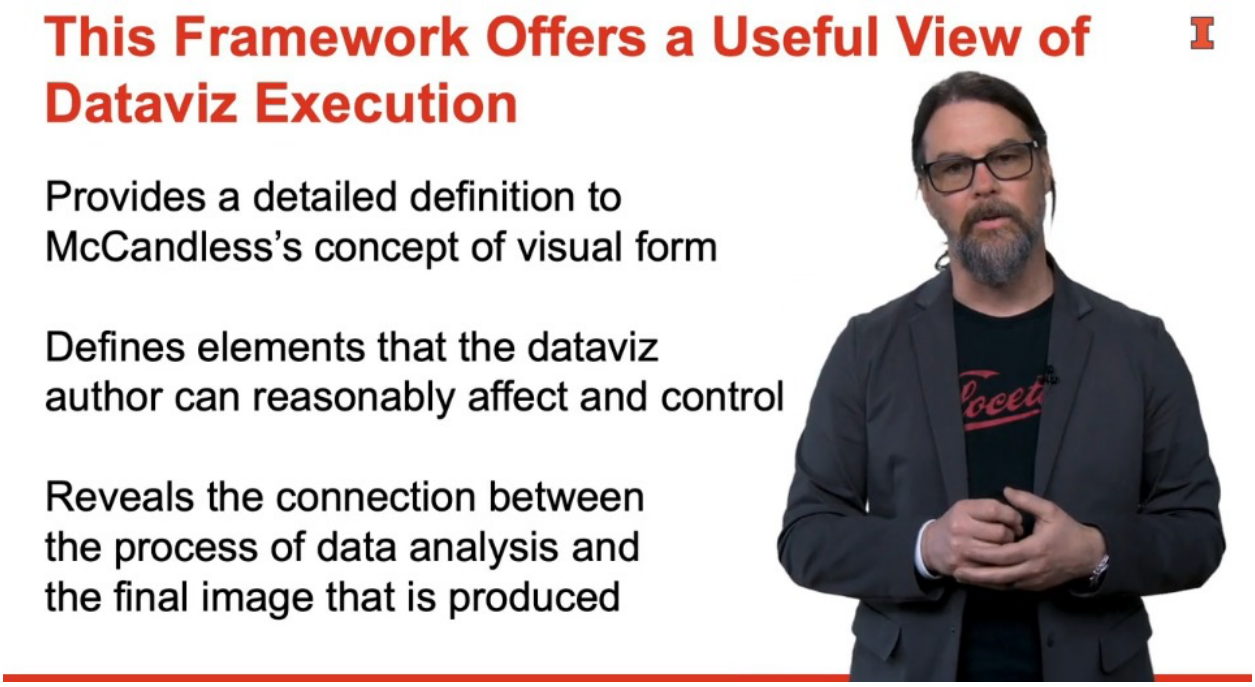
Understanding the Components of Visual form. Up into this point, we've been using a framework that has displayed to us all the different elements that need to go into a data visualization for it to be successful. But we really haven't delved too far into what good visual form really means or really used any kind of definition of that. That's what we're going to talk about right now.

So here is this framework that we've been using. Again, every single element, vitally important. The visualization only works if they are all present and all successful. This idea of the visual form, the actual picture that we're creating as we communicate with our audience is very important. But to this point, we don't really know what makes for good visual form and what doesn't. So I hope to provide a little bit of a definition there.

As we think about what makes good visual form, there's a really a three part framework that helps us answer that question. Good visual form has three different elements. It has very clear meaning, it has sophisticated use of contrast, and it has refined execution.



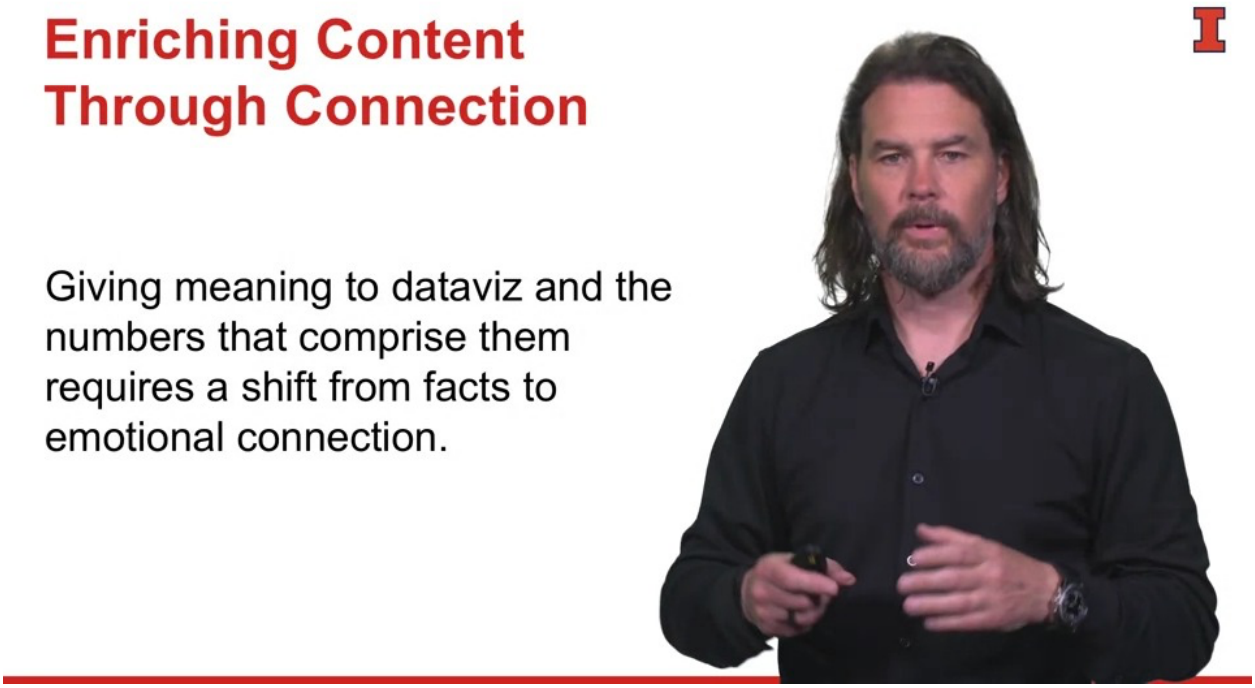
Clear meaning clearly communicates the intended insight, that insight that we have worked so hard to pull out of our data. Sophisticated use of contrast draws our audience's attention to the elements of that chart that we want them to see and keeps them away from those distracting elements that we don't want them to spend much time on. Finally, this idea of refined execution puts a lot of polish on the visual. It gives great attention to detail. It also serves to keep our audience focused on what is important and not to be distracted by those elements that aren't as important on a page.



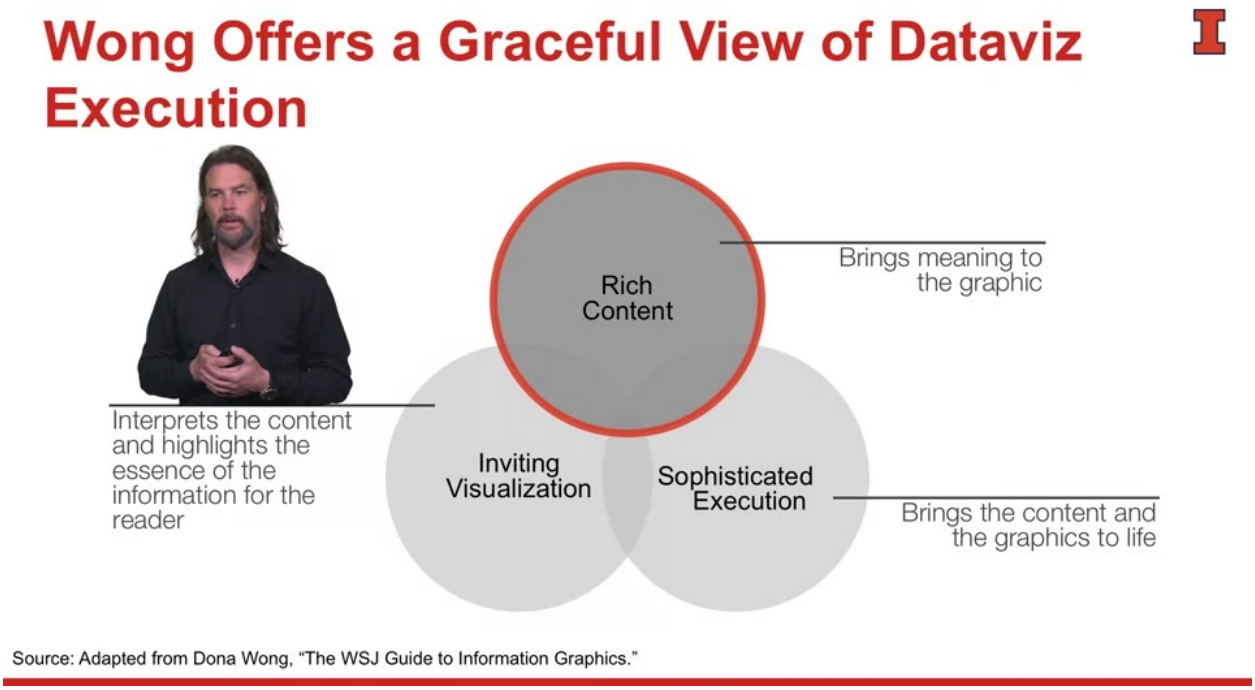
This framework is important for a number of reasons, and I like how it ties back to the original five-step McCandless framework on visual form. It does that tie very directly. We are taking this element of McCandless's framework, that being visual form, and we are defining what makes for good visual form inside of that framework. It also then reveals that connection between the process of data analysis, data collection, setting a goal, all those important elements that we've been talking to up until this point, and ties it directly to the image that we will ultimately produce.

Lecture 3-4: Enriching Content Through Connection

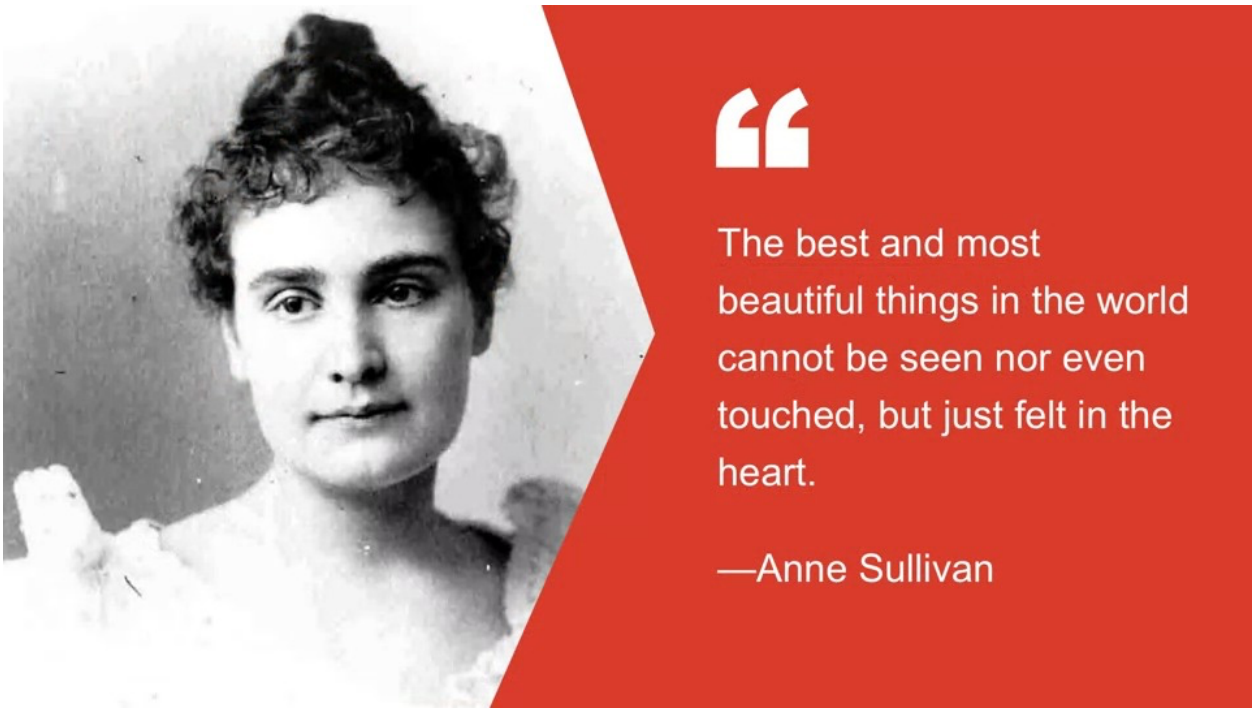
Lesson 3-4.1: Enriching Content Through Connection



Enriching content through connections. So let's take a look at how we bring rich content or ensure rich content to our data story and our data visualization. Much of this as we said, we'll be covered in our process of collecting data, moving towards an objective, building a story. If we're doing all of that correct, then we will have a very good opportunity to present rich content. But there are some things that we can keep in mind that will help us ensure the richness of the visuals that we're presenting.

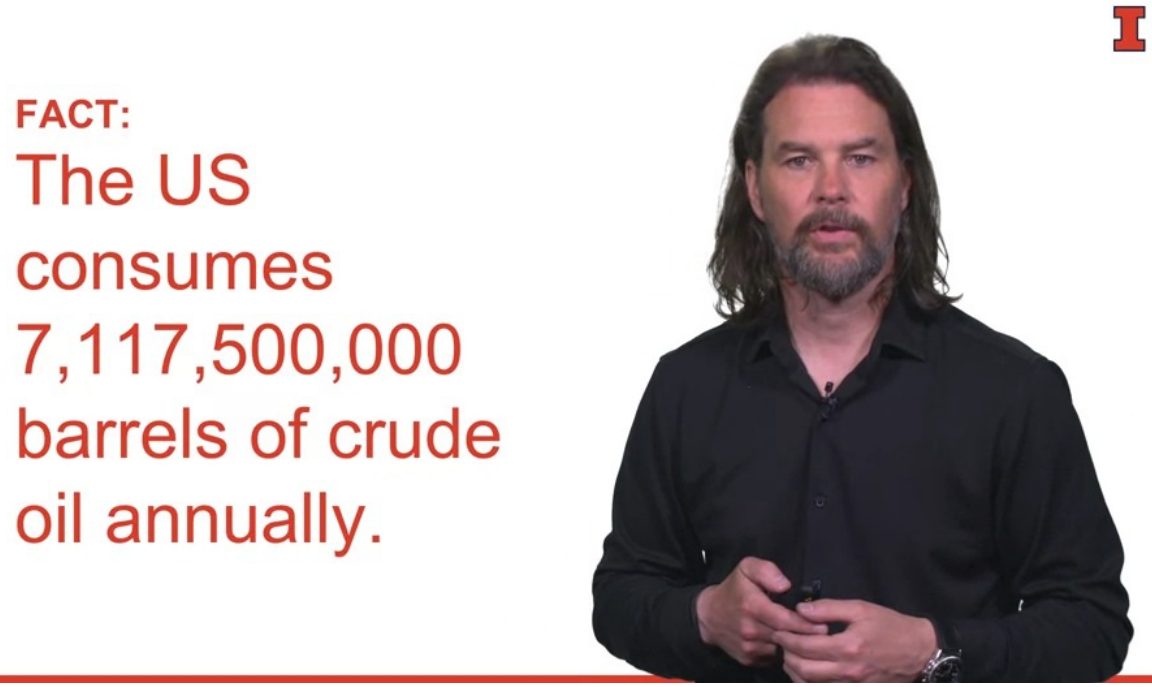


These are all as as we know elements of good visual form using the framework that Donna Wong has provided, and we are focusing here on developing rich content.

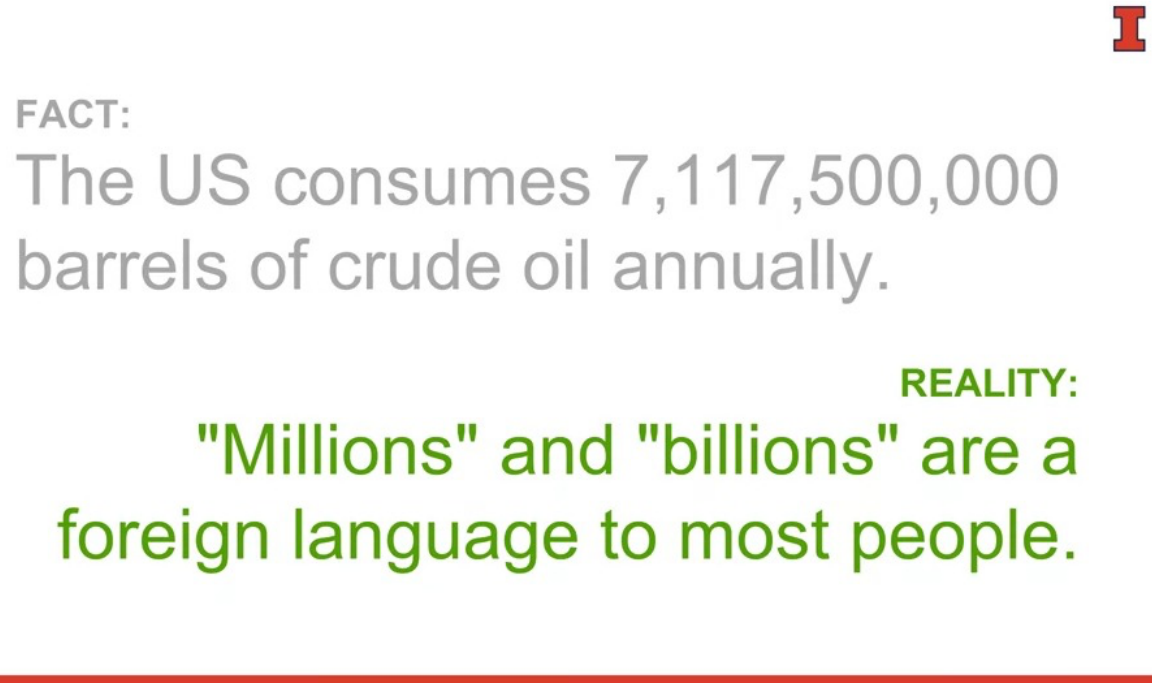


The importance of this I think was really summed up well by Anne Sullivan with this famous quote. The best, the most beautiful things in the world cannot be seen or even touched, but just felt in the heart. We should keep this quote in mind as we are presenting our data story because throwing facts and just figures at an audience will not make an emotional connection with them. It will not have the kind of impact that we want our story to have. So we can do that though by

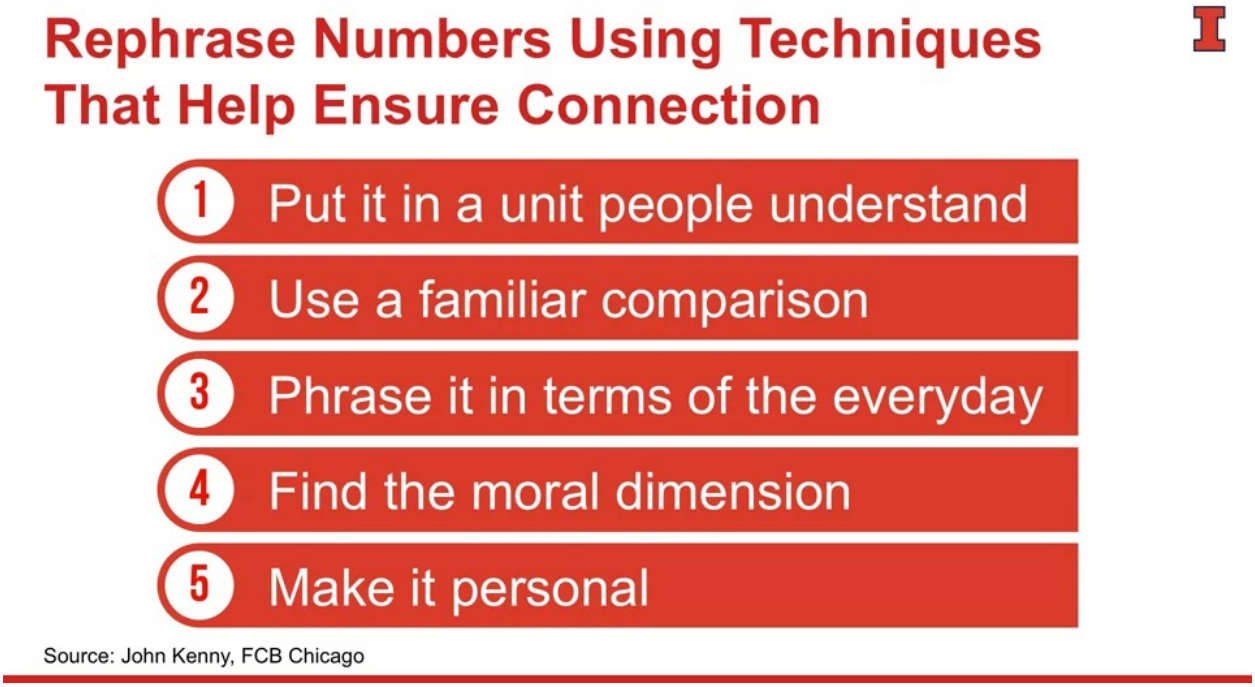
developing rich content. And the way that we phrase numbers here will be important and help determine whether we're making that rich emotional connection with our audience or not.



Let's take an example. Here's a figure U.S consumes 7,117,500,000 barrels of crude oil annually. This is purely fact. It seems like a great amount of oil, but really we don't have any context here. We just don't know.



The reality is that this idea of millions and billions, these are foreign concepts to most people throwing out. That figure will tend to just go right over their head. Sure it seems big, but we don't really have any emotional connection to hang on to.



For this reason, John Kenney my friend and head planner at FCB in Chicago, develop these five different ways to frame or position numbers to ensure that you are making an emotional connection to ensure that we are using rich content. These techniques will get you there. Let's take a look at each of them.

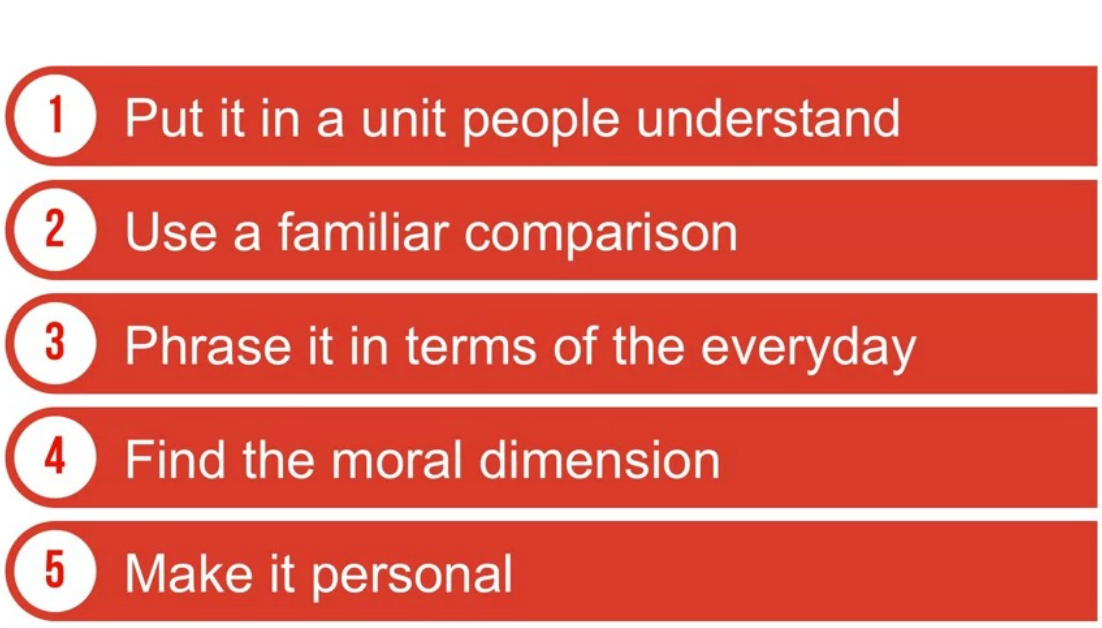
The first is putting your unit, your number in some unit that people understand. So, in this case we've taken that idea of 7.1 billion barrels of crude oil and we've rephrased it in this way, Americans use enough oil each day to make 36 billion plastic water bottles. Now, we're using billion here as well. But we've tied to an item that consumers understand. No doubt as you are watching this video, you probably have a water bottle near you to imagine 36 billion of those puts that into immediate context and is very understandable.

Another technique is to use a familiar comparison like distance each day. Americans use enough oil for 39 trips to the sun and back. We see the sun every day at least every good day. And to know how far away that is, we might not understand it in miles. But we do see that it is a long ways away to think that we as a country use enough oil to make 39 round trips is an enormous amount of oil that is putting it in much better context. Held on to the comparison that we understand the distance to the sun.

Another type of familiar distance can familiar comparison can be in time and just 2 minutes American use one million gallons of oil. Again, we don't really understand millions and billions, but we know it's big. And we know that 2 minutes is a time frame that we can understand and put context around. So to say that we use a million of anything in 2 minutes would express to us that it is big because that familiar comparison of time makes sense to the audience.

We can phrase this number in the terms of the everyday Americans use 48 times more oil than water every day. We don't necessarily see the oil we're burning. We know when we ride our cars were using oil. We know when we heat our homes, we're using oil. But we don't actually see that happening. We see the water we're drinking. We see the water we're putting on our yards. We, we understand the use of water to know that nearly 50 times more oil is consumed than the water that we see consumed. That is saying something and that for an audience puts that that number in the terms of the everyday.

Another approach. One that can be sort of high risk, high reward is the idea to find the moral dimension. In this case we've phrased that 7.1 billion in this way. In less than four decades, the world's finite oil supply will be gone forever. That can have a very emotional connection. It can also be very polarizing this technique. So you have to practice this appropriately and understand where your audience is coming from, so that you don't run the risk of offending or disassociating or turning your audience off yet. This can be a very strong and effective technique in phrasing numbers in an emotional sense. Finally, we can make those numbers personal. In this case an example would play out in this way in a year. A typical American family uses 70 barrels of oil. This is what it means to me, 70 barrels. I don't really understand how much that is, but I know these barrels are large. I've seen photos of them. I've seen, I've seen them on television. To say that I and my family will use 70 of them seems like a lot of consumption much more than I would have expected from myself. I have just taken this number 7.2 billion and I've made it personal. I brought it down to the personal level so that I understand directly my impact.



Source: John Kenny, FCB Chicago

Using one of those five techniques when we are talking about numbers, making those numbers that we're throwing around more relatable, more impactful, will help ensure that we are indeed using rich content. And making an emotional connection when we're talking to our audience.

Enrich Your Content by Connecting to Your Audience

Rich content brings meaning to a graphic for the audience.

Numbers alone rarely have the impact carried by numbers and emotion.

Data + goal + story approach serves to bring together requisite elements of rich content.

Emotion can be introduced through Kenny's techniques that make numbers more accessible and understandable.

This emotional connection is important for us because we're attempting to either influence our audience or connect with our audience through a recommendation. We have whatever it is we're trying to do to achieve an objective. So making that connection is indeed important. In fact it brings meaning to the graphic that we're displaying. Numbers alone rarely do this. They rarely carry the impact and the emotion that we want. Using those two things together. Numbers plus, emotion will have far greater impact on our audience. The process we've gone through of collecting data, identifying a goal, putting together a story. This approach will help ensure that we are building rich content into our story for sure. If we have deep robust data, if we have a clear goal, if we've got a good story, much of the elements are there. Yet using the techniques that Kenny introduced to us will ensure that we're making the best use of emotion and accessibility when presenting these numbers.