Week 12: Visualization

12.1 Fundamentals of Visualization

In this lecture, we're looking at information visualization and how you use it as a data scientist. Let's start with why we visualize things. One, we want to convey findings to an audience, they help us reveal relationships in the data, help us understand things, they help tell a story, maybe to ourselves or others, drives a conversation that results in new insights, perhaps it raises new questions, and ultimately, it helps guide a decision. How does the data visualization process actually work? You can map this out as really five steps. Step one, you identify something you want to understand or perhaps make a decision about. Two, you devise a question. Three, you identify, gather, and clean the data needed to visualize. Four, you identify a type of visualization. And then, five, you generate that visualization itself. Let's look now at the types of visualizations you can use. First off, there's all sorts of visualizations that you can do. First, let's think of them dimensionally. You have one-dimensional visualizations, 2D, and 3D, or you can think of them more commonly by their appearance or maybe how you've seen them like in Microsoft Excel. You have charts and tables, you have maps, you have maybe network diagrams, you have time series, you have hierarchy diagrams, as well as, increasingly in the popular media, infographic. Now, there are so many that it can take an entire semester course to go through all these visualizations. That's why we have a class focused on information visualizations. You can really drill into these. But we're going to cover two key ones that we think are most valuable to you as a data scientist, data maps and time series. So what's a data map? Here's an example. In the 1850s, London was beset by cholera, and people thought it was spread through the air. It was incredibly deadly. People died. Now, Dr. John Snow was skeptical. And he started to talk to people and found that it actually came from a water pump. And the way that he did that is he plotted cases on a map of the neighborhood in London, making an illustration. It was easier to see the pattern in the data by visualizing it rather than just looking at a raw list of addresses. And this was really the first case of using data to inform public health. What was the question that he was trying to find? Where is cholera coming from? What's causing it? And what's the message with this? Well, in this instance, the message was or the takeaway was cholera actually comes from the water. It's a water pump. And if you look on this diagram, it's underneath the "B" in Broad Street. That's a data map. What are the elements of a data map? Well, it shows a snapshot of data at a specific date and time, and that's what we saw in the example of the cholera map. That was in 1854. It has incredible benefits. One, it's really easy to create, like making a map. You have addresses. You put them on the map. And it's also pretty accessible to a viewer, but it also has a lot of limitations. It's a single point in time so you can't really see trends. You can't see trends at all. You can tell a limited story, and you need to know a lot of the context in the background, as I just described with that data map of the cholera. Could you infer the story just from the image? No, not necessarily. The next example we look at is a time series, and this is a time series diagram of what's called "A Year of ESPN SportsCenter." What this depicts is the show SportsCenter on ESPN, which is broadcast every

night. What do people talk about? What is the show itself about? So what's the question? What sport is most talked about when on ESPN during the year? And you can see that in the visualization. It's baseball. It's basketball. It's a football. It's never a hockey stick. It's never a race car. So what's the message? The ebb and flow of sport-- early than late in the season with the playoffs. So running up to a season, we talk a lot about it. But then when we get in the middle of the season, we don't talk so much about it. You can see that dip in baseball, you can see it in basketball, and you can see it in football, as well. But then it comes back in the playoffs. So what are some of the characteristics of a time series visualization? Well, it depicts data over a period of time, and the greatest benefit is it really begins to help tell a story, like we just did with what people talk about during a season on ESPN. It changes over time, and it reveals trends in the data. Now, be careful. Those are trends rearward, rearward. They're not looking forward to it. It's not predictive, though people often see these sorts of things as being predictive. Those are two key types of data visualizations. We can now look at that in a different dimension. So there are various types of interactions of how a user can engage with a visualization. There's three that we identify. One is a static image, the second is a dynamic image that doesn't have any user input, and third is an interactive visualization with both user input and perhaps control. So a static image-- these are much of what we see. They're snapshots of a single visualization, maybe a data map or a time series. You as a user can't really interact with them. They don't update dynamically, the data's stuck, and this is what we often see in print-- The Economist, The New York Times, The Wall Street Journal, things like that. The next kind we look at is more of a dynamic style, and this here is a very famous one. This is from Google's Big Picture Research Group in 2012. Fernanda Viegas and Martin Wattenberg created this wind depiction. It takes recent public wind speed data and animates it on a map of the United States. Now, what's the question, and what's the message? Well, the question is, where is the wind blowing in the United States right now or when the data was gathered? Well, what's the message? That's kind of hard with this one. And if you drill into some of the things they say, it's almost a little bit environmental political. "Wind is an invisible, ancient source of energy that surrounds us, energy that powered the first explorations of the world, and that may be key to the future." That's what they list on the website. What's the finding? Take a moment and try it out. The third kind that we look at and we started to see increasingly in the popular media is a new type of interactive visualization. The one you see here comes from Phil Roth, which is a visualization of every Major League Baseball team's salaries over the past 15 years by player, as well as salaries into 2023. What you can do is you can select a year and then a team and a player and see how much money they got paid that year, say, this year, as well as into the future. So what's the question? There's an incredible number of questions you could have with this. How much did Alex Rodriguez get paid last year? How much might he get paid next year? What is the total payroll for the New York Yankees? But the broader questions that teams might be looking at is, how are Major League Baseball salaries changing with the impact of the new luxury tax and the collective bargaining agreement? That's a really specific question that people are looking at with this sort of visualization. And what's the message with this one? All sorts of messages, but one takeaway is looking at how teams are starting to shed long-term contracts, which might be part of the luxury tax. There's a lot that you can really look at with these interactive stories. You can see there's many different stories that you can tell. In

closing, we've talked about why we visualize, we've talked about the visualization process, we've touched on a couple of different types of visualizations, data maps and time series, and we focused on the new and rise of interactivity and what that offers you as a data scientist. Remember, think about what you want to convey and your story when making a visualization. Pick the appropriate visualization that you want, and just like any story, you need to experiment, try it out, see what works, and make it a way, perhaps, for others to tell their own stories.

12.2 Visualizations Through History

We're now going to look at an example of visualizations through history. We're going to look through a collection of examples from the publication The Economist, which has used visualizations in various forms since the 1840s. These were recently published in an online retrospective by The Economist, and it offers us a coherent perspective on the history of visualizations in popular media. The first one you'll see is from 1843, when The Economist first started. It's a table of British exports to the United States in 1841 and 1842. Now, it's very simple. What's the message here or the takeaway? Trade fell in half due to US duties and taxes. It's a very simple depiction of data. Fast forwarding then into the 1900s, The Economist published this image you see here. In 1901, railroad workers' rights in the United States were changing. What this showed was a geographical map of the US with lettered areas for various types of workers. Recall that many readers for The Economist may not be familiar with US geography, so this provided a way to show what types of workers were in certain parts of the country, as well as what does the United States actually look at. That was in 1901. Fast forwarding now into the 1920s-- in 1926, The Economist published this image, and basically what it shows is electrical production in Great Britain. It's a very simple time series, and it was drawn by hand. It's really hard to see everything that's compressed into that really, really small area, but it does depict how electrical production rose and fell in the 1920s. In the 1930s, you started to see The Economist experiment. What you can see here is a bar chart showing the relative bulk of goods passing through the port of London as compared to the landmark St. Paul's Cathedral, based on size. So you'll see that some things are much larger than St. Paul's Cathedral, which is great, if you know how big St. Paul's Cathedral is. But that may not be a reference point that people are familiar with. If you're in London, if you're an affluent reader of The Economist, you've probably been to St. Paul's and you can see that. This is one of the early infographics in popular media. Now, The Economist continued to do visualizations. Into the 1960s, and in 1967 in particular for this image, they depicted the global prospects for trade and other economic factors at a big macro level. So they would focus on a part of the world or an individual country or region and show what was going on there from a trade perspective. The 1970s, they started to experiment even more. And this is from 1975, and it's a time series visualization of interest rates and loans from the bank in London or a number of banks in London. And you can see how they rise and flow throughout the years. Turning now and fast forwarding into the 1990s, they go radical or crazy a little bit again. So this one in 1990 shows the relationship between interest rates and the strength of the US dollar in the 1970s and 1980s. And you can follow this around from the top left quadrant to the top right to the bottom right to

the bottom left. You can follow the spinning vortex around to see what is the relationship. That's what they're trying to do with this visualization here. Is there a pattern in the data that we can see? Fast forwarding now to the 2000s-- this one is from 2013, and it's a visualization of which country's riders won the annual Tour de France bicycle race. Now, if you look at this diagram, you can see that there were no races or no winners during World War I and World War II. If you look very closely, you'll see that there's no winner in 1999 through 2005. That's Lance Armstrong. They pulled that data out. Who knows if this visualization was political? Probably not-- but really what it's looking at is who's won the Tour de France through the years, what are the major life events or world events that have occurred, and what perhaps have been the controversies so far. In summary, what we've tried to do is look through these examples to show a recent history of visualization in a certain context, The Economist publication. Some closing thoughts-- each of these visualizations pushed the limits of technology at the time. Simply printing tables, introducing hand-drawn images and visualizations, that was really hard at the time, particularly in print media. Each of them also pushed the reader's ability to understand the data, and this took time. Now, that first one with the chart, that was pretty straightforward. But showing the map, showing a trend series diagram or a time series diagram with a bar chart, the other information visualization with St. Paul's Cathedral, those were hard. Now, granted, some of the visualizations worked, and others didn't. Some of them you probably were able to pick up on. Others were a little bit harder to use. But The Economist has stuck with it through all the years, continuing to use different innovative technologies to convey information visually, and they continue to do this today. Every single day, they publish a new visualization on their website.

12.3 Visualization Examples

12.3.1 Overview

Consider how different types of data are portrayed differently and which ones are more effective than others at informing or motivating their intended audience.

This element addresses the following learning objective of this course:

- LO5: Identify the audience and the most effective method to communicate a persuasive argument.

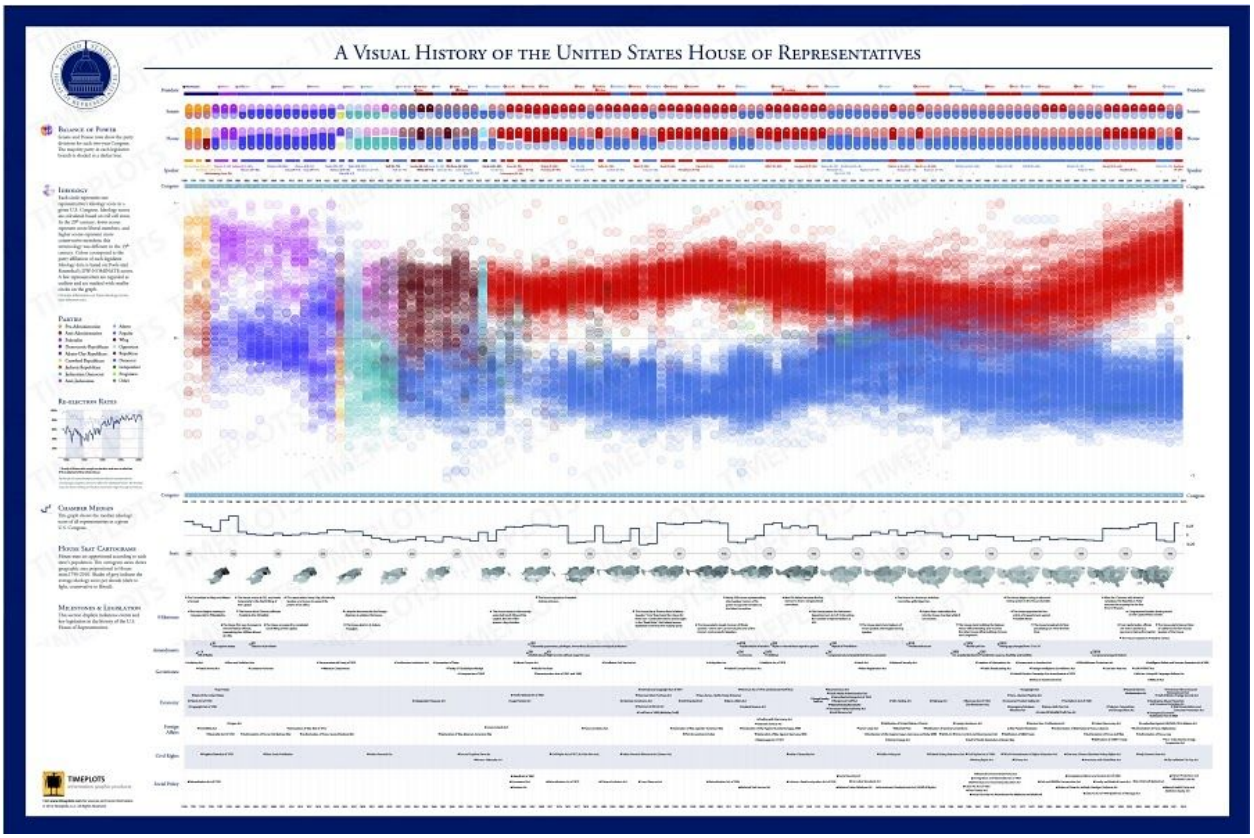
12.3.2 U.S. House of Representatives

The U.S. House of Representatives is made up of individuals with political beliefs that can be generally grouped into one of two parties: Republican or Democrat. This visualization depicts the political ideology of each of the 435 representatives since 1789, when the House of Representatives was founded. The number of representatives from each party rises and falls over time, with each party seeking to outnumber the other in order to push forward its political agenda.

12.3.3 Static Visualization: Visual History of the U.S. House of Representatives

While you look at the visualization think about the following questions:

1. Does it appear the parties are becoming more or less ideologically similar?
2. Why are there many different colors on the far left side, but only two on the right?
3. Does this help us better predict the future? If so, why? If not, why not?



12.3.4 Static Visualization: Your Contribution to the California Drought

Look at the (mostly) static graphics in the following *New York Times* link. Before you click the link, what do you think you will see? When you return from the page, ask yourself what was the main takeaway? What feelings did you experience when you read the page?

Your Contribution to the California Drought

12.3.5 Dynamic Visualization: Visualizing the World's Addiction to Plastic Bottles

Explore the visualization at the following link. Draw your attention to the dynamic page header.

Visualising the World's Addiction to Plastic Bottles

12.3.6 Dynamic Visualization: The Fall and Rise of U.S. Inequality, in Two Graphs

Explore the visualization at the following link. Press play on the second image.

The Fall And Rise Of U.S. Inequality, in Two Graphs

12.3.7 Interactive Visualization: City Populations

Explore the visualization at the following link.

Population Mountains

12.3.8 Interactive Visualization: Why Does the U.S. Lead the World in Incarceration?

Explore the visualization at the following link.

Why Does the U.S. Lead the World in Incarceration?

12.4 Why Do We Visualize Information?

Discussions Prompt
Spend 10 minutes on the following prompt:
Find two publicly available data visualizations: one that you find distinctly powerful and one that is weak or distracting. What makes them effective and poorly executed? Include the links in your response.

12.5 Storytelling With Data

12.5.1 Understand Context of Visualization Task

This element addresses the following learning objective of this course:

- LO5: Identify the audience and the most effective method to communicate a persuasive argument.

The ideas in this video and the ones that follow come from the Storytelling with Data text by Cole Nussbaumer Knaflic. To effectively communicate, we need to know the context around an ask. We can think of it as the who, what, and the how. Who is your audience? Be as specific as possible here.

What biases do they have that might help them support or might prevent their support of the idea? What do you need your audience to do or to know? What relevant background information is necessary?

And then how is, what data is available that will help persuade your audience? Here's an additional quick lesson from storytelling. We may be able to brute force persuade or convince an individual, but it's not as effective if you persuade them with the use of a narrative or a story. Also, make sure you have a problem that you need to solve. Otherwise, the data exercise is not important.

12.5.2 Minimize Audience Cognitive Load

This element addresses the following learning objective of this course:

- LO5: Identify the audience and the most effective method to communicate a persuasive argument.

Be careful about what you include in your report. Don't add information to the page just for the sake of it. Add information with a purpose. When our audience sees our work, we want to avoid the following. We don't want them to say what's going on? Or that looks fancy, but I have no idea what it is.

Our inclination might be to show as much as possible, to communicate how much work we've done, or how technologically savvy we are. But we don't need to go through all the details of our project.

Remember, the process of research is not the same as the process of presenting. When you present, spare the audience of all the unimportant research details. We might be tempted to include complicated visualizations just so people know how much work we did.

Here's one suggestion if you find yourself in that scenario. Maybe you've created a visualization that is accurate, but it's kind of messy. There are multiple axes. There's a ton of colors, and maybe it even moves. Maybe it's dynamic. You want to show to your manager how much work you've done, and you want to show your technical expertise.

One thing you could do is you send them to images or the two visualizations and ask for their guidance. Let's say something like, this one is technically accurate, but I think it might require too much time to explain it. So instead, I broke it down into two more approachable visualizations. What do you think?

Now, hopefully, your manager is like, yeah, that's too much. Pick the simple one. Here you get credit for your hard work, and you also end up displaying the simpler visualization in the final product.

Now, we don't want to think of our audiences as unsophisticated. At the same time, we don't want them to work too hard to understand our point. We want to avoid cluttered

visualizations, non-intuitive tables, graphics with multiple y-axes, non intentional use of colors, poor contrast of color and text in the background. There are some things to avoid.

The things I just listed can increase the cognitive load on our audience. And if we do this, we run the risk of losing the audience's attention. And if we lose their attention, then we lose the ability to communicate and the ability to persuade.

12.5.3 Preattentive Attributes

This element addresses the following learning objective of this course:

- LO5: Identify the audience and the most effective method to communicate a persuasive argument.

Preattentive attributes are the size, color, and position of items on the page. We can use these attributes to enable our audience to direct their attention to certain elements of our visualizations. Be mindful of color and the relative size of the text images and the position of items on the page. These things should be done with intent. If we do this well, we can draw the audience to the story before they realize that's what we want them to do.

12.5.4 Design Heuristics

This element addresses the following learning objective of this course:

- LO5: Identify the audience and the most effective method to communicate a persuasive argument.

Design choices should be intentional. If someone asked you what made you use a certain color, or if they ask you what's the added value of this image, you should have

an answer. Someone might say, hey, what's up with this graphic? What did you want to communicate with this? You should be able to clearly articulate what your intent was.

This advice applies to images, graphics, tables, charts, and even bullet points. If you don't have a clear intent, there's no way the audience will know what you're talking about. Said differently, your intent needs to be clear internally so your audience receives the information you mean to convey.

How do we get better at this? There's no single best way to kind of create the best graphical representation of an idea. It depends on the context, the persuasive task, the culture, the existing knowledge of the audience, et cetera. But one way we can improve our visual design skills is to keep track of things you see. This could be a bookmark in your web browser or a file on your computer.

When you see something that looks nice, file it away. Pause to think about why you like it and what makes it particularly persuasive or memorable. Also, I encourage you to save visualizations that are confusing, probably in a different folder, though. Think about what makes these visualizations particularly ineffective.

12.5.5 Accessibility

This element addresses the following learning objective of this course:

- LO5: Identify the audience and the most effective method to communicate a persuasive argument.

Remember that people have different vision abilities. Keep in mind the use of colors, the contrast between the foreground and the background, the size of text, line spacing, et cetera. We want to make things accessible to people because that's the right thing to do but also because we want people in your audience who you're trying to persuade to be able to see what you're showing them. If someone is having trouble seeing your presentation, the best case scenario is that they raise their hand and they tell you that they're having difficulty seeing what you're describing. And then you could clarify.

But the worst case scenario is that they don't understand what you're talking about, they don't ask for clarification. And then at that point, you failed at your persuasion task. The onus is on you. It's on you to make sure that your message lands. See if your company has a template that has been vetted. Or a quick search of the internet can also yield some helpful results on accessibility standards.

12.5.6 The Only Thing the Audience Sees

This element addresses the following learning objective of this course:

- LO5: Identify the audience and the most effective method to communicate a persuasive argument.

Your final communication steps are the only part of the entire analytical process that your audience sees. All the analysis, data cleaning, thinking, iteration that occurs before the project is presented to the audience, that's all important. But the only thing the audience sees is the communication step.

Here we're talking about the presentation, the report, the visualizations, and the story. So make sure you put a considerable amount of effort into this phase. In general, I recommend you expand how much time you spend in this critical communication phase. Also, I encourage you to think about the visualization earlier in the process. Think about how you're going to show the audience.

But even if you insist upon waiting to the end to think about communication, at least build out a little bit more time in your project planning for this important stage. Remember, this is the only stage of the research process that the audience sees. Develop it with care.

12.6 Minimize the Audience's Cognitive Load

Discussions Prompt

Spend five minutes on the following prompt:
What does it mean to minimize the audience's cognitive load?
Describe a situation where you viewed a presentation as a member of the audience and the cognitive demand was too much. What happened? If you could travel back in time, what recommendations would you give the presenter so they could more effectively convey their message?

12.7 Data Visualization Expert Interview

Interview with **Michael Porath**

Data Visualization Expert Interview
Michael Porath will join us to talk about visualization. He is the co-founder of Halftone, a San Francisco-based data visualization consultancy. Michael's company helps clients solve industry-specific problems by designing and building customized data visualization applications and information displays. Prior to founding Halftone, he taught courses on information visualization at UC Berkeley's School of Information, where he also earned a master's in Information Management and Systems (MIMS).

This element addresses the following learning objective of this course:

- LO5: Identify the audience and the most effective method to communicate a persuasive argument.

So everyone, I want to welcome Michael Porath here to join us. He is a graduate of the MIMS program here at UC Berkeley, and is going to chat with us about information visualization. Can you talk a little bit about how a company brings you into the process?

Maybe think of a specific project that you've worked on. And we don't need company names or anything like that, but can talk about what happens when you guys come into the room, or you come into a project? Can you take us through one of those?

Of course. Well, in fact right now we're actually just kicking off a new project with a company in the South Bay. They're sort of in the bigger marketing space, and they just have a lot of data to communicate and to show to their users in their product. And so they got us in to help them, one, to communicate it effectively, and also to make charts that are easy on eye and readable.

So how that works is they usually come in and they have to brief us pretty extensively on what it is that they want to achieve, because one of our goals is always to design for an objective. We never create visualizations in a vacuum or on a playing field. What we really want to do is have an objective and then visualize something to meet that objective, as with any other science or profession.

And so we had the kickoff day, or kickoff half-day, of a couple hours where they talked us through their use cases and their data, their clients, what's important to them, where they see challenges in their products and opportunities. And so that kind of gave us a pretty good solid basis of background information about the company, their users, and the information data we're dealing with. And then we went back to the drawing board in our office, and they provided us with example data sets from the real world. That's another thing that's really important to us, that we work with real data, not made up fake or mock-up data.

And so we just sifted through many large data sets just to find interesting factors, again to get a sense of what the data is and to see where the opportunities are to visualize. Because not everything is interesting to visualize or necessary. So we design usually for analysis, for data exploration, and for that we're just throwing all the data on the canvas and then trying to see what's there.

So that's the second step. And then the third step, third and end step, will be to iterate on that process. So we do that over, and over, and over, and over, until we find something that's interesting, relevant, and that works on a larger scale.

Got it. And what I would imagine is maybe feeds back into what maybe one of their objectives was and what they were trying to do. Does that seem about right?

Can you repeat it again?

Sure. So it sounded like you'd taken that second step, you took the data and you joked about putting it up on the wall and just looking at what you have. And I think that sounds like a key part of doing visualization itself is just looking at data in new and interesting sorts of ways.

And you mentioned finding a nugget, or finding something interesting and following that path. And what it sounds like is, do you loop that back to then what their original objectives were? What they were trying to find in the data.

Absolutely. Yeah, absolutely. They are the subject matter experts, even though we're briefed on the content. And so they still know best how to meet their objectives.

And so we always run those sort of prototypes, or just very rough visualizations, we run that by our clients, get their feedback, try to see if it meets their objectives, and then we go back to the drawing board and make another iteration. And we slowly but steadily sort of converge on something that meets their objective and uses their own data, and something that wasn't maybe necessarily visible when we started out looking at the data. So it's usually a process, and it's iterative, so you don't quite get to the right answer right away. It's really sort of digging for the right gold nuggets in the data. And that especially is true with larger data sets, with big data types.

Who do you think, or what are some of the great visualizations out there, whether it's from history, or from ones that you see today, or outlets. Where should we look for inspiration and guidance on how to make these?

Great question. It's really-- we're kind of lucky right now because there's so much out there, really. We're talking business, again nonprofits. There's a ton of blogs out there that deal with visualizations, both in an academic sense if that's of interest, but also in a very applicable sort of professional way.

There's blogs about how to visualize well and effectively with Excel. There's tons of blogs out there about other more specific techniques. Thinking about specific visualizations, one that I always like to mention is Crimespotting that was created a couple of years ago by a design studio here in a city called Stamen. And they used police reports to show where in Oakland and San Francisco there are a lot of crimes happening.

So things like robberies, things like theft, or even car theft. And you really see, when you look at the data-- and it's shown on a map, and it's shown by category-- and you can really see, if you know the city of San Francisco a little bit, how along the Mission corridor, along Mission street, just crime peaks incredibly. And so it's a great analysis tool, but it's also almost ironically or cynically fun to look at, because you can relate to it.

Everyone sort of walked through the city and you see certain things, and that reappears on the map. And it's all based on very generic police reports, but it's made available and accessible through visualizing it. And I think that's really helpful.

What are your thoughts on, with visualization there's definitely a sense of art and design with that. And how does one go about learning to do that? How do you learn those sorts of skills?

I think a lot of it is just practice, repetition, and sort of training your eye to see it right. And there's books out there to teach you that. But I always think of design and visualizations as sort of fitting in the 80/20 rule.

You can get 80% just by learning the very bare principles of what works in visualization and how our perception really works when seeing things on a display, or seeing things in front of us. And so if you just sort of study that, you get a lot of bang for your buck. And then the remaining 20%, those are maybe really the visualizations that stand out

from the masses and that are really beautifully crafted. But I think specifically in a business setting, if people know how to create effective visualizations, what mistakes to avoid, what design principles to follow, they can get really, really far. And that makes a huge difference compared to default visualizations that Excel spits out, or other visualizations that are being used in a business or business intelligence big data setting.

Interesting. Are there any perils of visualization and ways that they can be used maybe manipulatively?

Yeah, definitely. I think every data set can be visualized in 100 different ways or more. And there are examples where the same data set is taken, visualized in one way, and then the same data set is visualized in another way by a different organization, usually on the other side of the political spectrum, and it looks completely different. And it conveys a completely different message.

I remember this one visualization from the New York Times focusing on unemployment rates. And they used exactly that to show how you can deceive and manipulate with data, showing the same data set from the perspective of Democrats, from the perspective of conservatives, and then from a quote, unquote "neutral perspective" being the New York Times, although they all fall on the spectrum somewhere. But they can really show by just toggling between the conservative point of view or talking point, and the Democratic talking point, and then the neutral talking point, how the same display changes completely.

And those, we're talking things like cutting off an often axis so that something that's really steep is all of a sudden becoming super flat. Or we're talking about things like coloring things that are not important in bright red so that they really stand out. So really simple techniques. And with great power comes great responsibility. If you know the right tools, it's all on the designer, all on the analyst use the tools to right in order to stay ethically correct.

Is there anything else you think our students should know about visualization? At this point, and coming into this graduate program, learning about visualization, maybe taking

future courses in it, anything else that we haven't covered in our class that you think would be a good nugget for them to know?

I think one important thing is everyone is talking about big data and visualization of big data. I think even small data, and we're talking about a spreadsheet of 20 rows, that can be visualized really well, really effectively, or horribly. So even there, there's so much potential to do it right.

I would really say start there. And if people can do that well, that puts them on a level that is 80% there already. So you don't necessarily have to look out for the big, complex, really intriguing things. There's a lot of potential with just simple data sets, small data sets, small spreadsheets to visualize correctly, better, and more effectively.

Great, that's perfect. That's excellent, a wonderful way to wrap it up. So I just want to thank Michael for being with us today. This has been a great interview learning about his experiences. We'll have more links to his content and so forth posted on the site. So thank you very much for joining us.

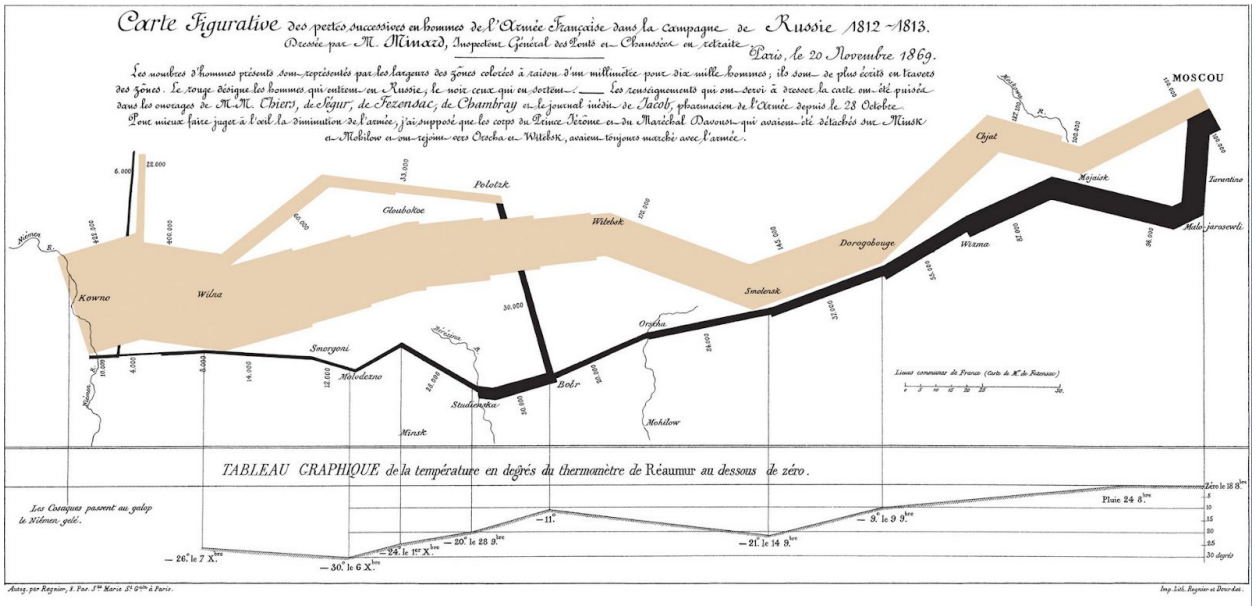
You're welcome.

Thanks.

12.8 Visualization Discussion: Napoleon's March

12.8.1 Map of Napoleon's March

In the preceding discussion, Steve refers to Edward Tufte and Minard's map of Napoleon's march below. For a larger image, right-click and download the file.



This element addresses the following learning objective of this course:

- LO5: Identify the audience and the most effective method to communicate a persuasive argument.

12.8.2 Visualization Discussion: Map of Napoleon's March

Welcome again. We have Sandra from the School of Information, Elaine from the School of Information. And we're going to spend some time talking about some key elements of some really interesting information visualizations. Now before we get started, I should say that information is all over the place now, right? I mean, pick up a copy of Business Week these days, and it's basically an information visualization magazine. But we want to go back and-- just for the sake of discussion-- kind of, untangle some of this stuff-- a few classic examples of information visualizations that continue to make a difference in the way people think about this stuff.

So before we get started as online students, take a look at this-- the classic information visualization from sometime in the 1800s that anyone who's ever read an Edward Tufte book knows by heart. What's so special about it? What is so-- why is this the iconic thing that we all think of?

So I really love this visualization, because it's so simple, and yet you get-- on the first glance you kind of understand what's going on. We're looking at the progression of Napoleon's army into Russia, and then also on the way back, and all the-- like, the

dwindling numbers. But as you look closer at the visualization, you realize there's a lot more than just numbers.

We're looking at-- I think there's dates. We're definitely looking at temperature. We're looking at where they were, and how-- where they were, where-- when-- as numbers started to decrease. And as you keep looking at these, it just, it's-- there's so many details there that you don't see at the first glance.

Hmm-- Elaine, what's your reaction? I mean--

I think it's remarkable that this is so old. And you know, you would think that you wouldn't necessarily have to be literate to understand this. And you could just look at it. And if someone kind of explained what different things meant, you could follow along and still come away with just as strong of an impression of it as if you had read all of the different tags and numbers.

So anything strikes you as, sort of, strange, or buried assumptions that this, sort of, really portrays? I mean, for me, I'll say this is a graph showing hundreds of thousands of deaths. And it feels a little cold actually for portraying something that is human.

Yeah, it definitely feels very much removed from the entire event.

Yes.

When I say it's fascinating, it's fascinating from an information visualization standpoint, not necessarily from this was a great situation, because it obviously wasn't.

I think it definitely makes the assumption that if-- as people disappeared from the troops that are portrayed here, that they died. Versus maybe a lot of people were defecting. I assume that they died, because it was very cold. But what if people just ran away. And so there are some elements missing, but I guess the main takeaway that the army was getting smaller and smaller remains.

So I want to pose a question to the online student with this very classic visualization. Let's say that you are being called into a situation where you needed to explain to someone who didn't know the history why Napoleon lost this war. And this is the one visualization you had. What would you want to add to it to make that explanation more powerful?

12.8.3 Discussion Prompt

What would you add to your explanation, using this visualization, of why Napoleon lost the war?

12.9 Visualization Discussion: Afghanistan Stability/COIN Dynamics

12.9.1 Afghanistan Stability/COIN Dynamics

In the next segment, Steve refers to the Afghanistan Stability/COIN Dynamics visualization that was featured in the *New York Times*.

Please download the Afghanistan Stability/COIN Dynamics visualization here.

This element addresses the following learning objective of this course:

- LO5: Identify the audience and the most effective method to communicate a persuasive argument.

12.9.2 Visualization Discussion: Afghanistan Stability/COIN Dynamics

So let's turn to you folks. If you could add one thing to this map in order to help people understand what happened and why Napoleon lost this war, what would you do?

It would be kind of interesting if you could add some indication of Napoleon's sort of favor at home. So as more French people were dying or leaving the army, how did that affect his position in power in France? What did this do to-- did this ultimately lead to his downfall?

Yeah.

I would think that the person who made this visualization, he didn't have at his disposal computer interaction. And so putting this on a computer display and maybe being able to roll over certain portions of it and display a narrative of what was going on would really add even more of a personal touch to it.

Because in many ways, this is the story of the birth of modern Europe, actually. I mean, Napoleon loses this war, french power begins to dissipate in Europe, and that leaves Germany and Russia as the kind of principal competitors with the British for hegemony in Europe. And that's sort of the story of the next really 100 years.

I mean, it starts right there, and this is where the map or the visualization kind of ends. And that feels a little bit like a kind of-- God, I wish there was something more here. Well, probably today, we'd have better ways to do that. Well, anything else you want to say about this visualization, about why it's so iconic or how it would be done differently today?

The first time I saw this image, I thought it had been created on a computer.

Ah, interesting.

And then I was just very amazed that the person did this by hand. I think that that's quite remarkable. And you know, and again, appealing to a wider audience during that time I think is very important. mean,

As a static image, I think it is-- you could easily see this as something that was done now with modern technology. I mean, the way it focuses on conveying one point of information to start with is, I think, a very classic thing. I think it's fundamental to a good information visualization creation.

Not trying to overload the agenda.

Right.

As I was just trying to do.

[LAUGH]

I mean, we have lots of technology at our disposal to create really interesting things, but that doesn't mean we necessarily need to use it all the time.

Interesting. Interesting. Well, let's turn to another example of a more contemporary information visualization, which is less about data and more about causal relationships, or an attempt to portray causal relationships.

So online students, take a look at this next visualization. You may recognize this. It was widely reported in The New York Times. I believe it was 2010. And actually, widely derided. People made fun of this almost mercilessly. And so let's try to kind of unravel, what is going on here? What do you think the people who drew this intended?

So my first reaction is, I have no idea what's going on in this thing. It's very obviously trying to explain that there is a lot-- there is a lot going on in this situation. It's not a very good interpretation, I think, of it. A lot of the points are grouped and they're given labels,

but the labels cover a lot of it, so you actually can't see what's going on. And I kind of feel like the person who created this either had no idea what they were doing or were intentionally trying to confuse the viewer.

Interesting. Interesting. Probably more like the former than the latter. But Elaine, you've actually spent a fair amount of time in the federal bureaucracy, and this map was created for the US government, or at least the Department of Defense. What's your reaction to this emotionally, and then from your experience?

It almost seems to me like they wanted to create this so they wouldn't get a lot of questions. Because sometimes when you present information and it's almost too easy to understand, people can nail a lot of holes in what you're saying and really take down your point, versus this image feels very unapproachable. I don't know how to respond to it.

I can't even-- with the larger letters, so what I'm assuming are the main ideas, I don't even know how they relate to each other, let alone all of the sub areas and small phrases. And I don't even know how those relate to the larger ideas. That is something that's very basic that I think should be in any visualization. But here, I just don't know what's going on.

So I want to pose a question to the online student about this diagram. Presumably, it strikes anyone as really trying to portray an extraordinarily complex situation. And let's say you were handed this, and actually didn't know a lot about the war in Afghanistan, but were handed this visualization and asked to improve it. What is the question that you would want to pose to the person who handed you the visualization?

12.9.3 Discussion Prompt

If you were asked to improve the Afghanistan Stability/COIN Dynamics visualization, what would you ask of the person handing it to you?

12.10 Visualization Discussion: Salary vs. Performance

12.10.1 Salary vs. Performance

In the next element, Steve refers to the Salary vs. Performance visualization.

Please view the Salary vs. Performance visualization below:

Salary vs. Performance

This element addresses the following learning objective of this course:

- LO5: Identify the audience and the most effective method to communicate a persuasive argument.

12.10.2 Visualization Discussion: Salary vs. Performance

Any thoughts about what else you would want to know about the person who is trying to use this for some purpose in order to help make it better?

I mean, I think Elaine brings up a really good point of if we could separate the different groups and show how those are related to each other as opposed to simply overlaying the text and assuming that you can understand what's going on, I think that would definitely-- it would be a good step to starting to improve it.

I think there's no sense of which idea is more important. So I would assume, even if you're coming at it from a military point of view, the Department of Defense has some idea of what they can either enact more policies around or what might be more important to them. But everything has an equal weight on this map, assuming that the colors are equally seen by all people.

Mm-hmm. Interesting. There's an apocryphal story you may know that floats around Washington that says this map was put up in a briefing to the commander of the forces in Afghanistan at the time, General Stanley McChrystal. And he looked at it and he said, when we understand this visualization, then we will have won the war. And that's probably the response actually that you don't want to get. And yet, it's kind of a courageous effort to take the complexity of what really was going on on the ground here and portray it for people in a way that they could at least see it, if not talk about it systematically.

Right. I do think that this image is strong, and making sure everyone in the room knew that this was complex. So it is easy to focus on one idea at a time and think that you can tease that apart. But here you understand that you have to take it in total. And so if

that's the main idea that the person who created this image wanted to impart on the viewers, I think they were successful.

Would have been good to know a little bit more about the cognitive tolerance of the people in the room. Like, how much complexity are these people used to dealing with? And how much do they like to deal with? Because presumably, people have different preferences around that. And boy, you'd sort of like to know, wouldn't you? Before you put something like this up on the board?

I think it's also easy to assume actually that when you have-- when you're used to seeing visualizations that are for sort of consumer use almost, it's easy to assume that it's really easy to make a clear visualization. Very obviously, this is a complicated situation with so many different consequences, it takes time to think of the best way to convey your information.

One last thought about this, and I remember my reaction when I first saw this was, this is where we get to when we're getting to a stage in a new way of conveying information that represents almost a kind of visualization fatigue. Like, this is what's going to make people sort of want to do something else. Do you get that same sense?

I think so. I think when you come-- when an image is presented to you, kind of you might be excited by, oh, what is this trying to impart? And if you're having to work for the information, the more you're having to work, the more I lose interest.

Well, let's turn to one last visualization to get a feel for one of the places where information visualization can be particularly fun and exciting. The stakes are also really large, and that's, of course, in the sports world. Online student, take a look at this, what appears to be a really simple portrayal of the relationship between salaries and performance in Major League Baseball.

So when we look at this information visualization-- and as you can see, if you play around with it, it's dynamic. You can move from year to year. That's kind of nice. What's your reaction to this one as baseball fans, I understand?

[LAUGH]

So full disclosure, I am not a big baseball fan. But my first reaction when I saw this visualization was, what is going on?

[LAUGH]

Until you actually read the text that supports the interactive visualization, you don't really understand what's going on. There's one thing to say what the title is, but you actually need to play around with it and actually see it. So this does have some really interesting interactive components to it, but it takes what is a really interesting, fairly simple idea and makes it really complex.

Yeah. Yeah.

I really had to work with this image. I know nothing about baseball, and so when I was going through, I was trying to figure out what the red lines mean? What do the blue lines mean? Obviously, I understand slope and what that implies. But looking at this, it took me quite a bit of time-- especially I thought the ability to make this proportional, you could change it. I thought that was really useful, but when I didn't understand what was going on, it just seemed like another layer I had to, like, understand before I could get to what the maker of this image wanted me to understand.

So who do you think this image was actually designed for?

I'm assuming baseball fans.

[LAUGH]

If you-- and you could go right to your team. I think that was also-- once I went to teams I recognized, I had a little bit of an easier time. You go to the New York Yankees, and you know that they're a really well-known team, and it makes a little bit more sense. But when I looked at this as, I don't have a home team that I really follow, and I had to look at all of them at once. So perhaps if you were really into baseball, this would be a much easier image to understand.

I think if you are in sort of the higher up position in the baseball world-- so if you are considering salaries, if you're considering, how do I manage my team, this could be potentially really useful information. So for example, if you can tell that some team is paying less than yours, and yet they're doing much better, maybe we need to consider what's going on in our own organization.

So it says, dig here.

Right.

Yeah. Interesting. Interesting.

I'm really intrigued by what would happen if you did this with, say, nonprofits or places where you're giving donations. Because here, I don't pay for any professional team, so it really doesn't matter.

Well, you do. You just don't know about it.

[LAUGH]

That's true. But if I were giving money to an organization and I saw them compared to other organizations and how they were spending money, that would really change my spending habits.

Really interesting. Any other examples of how you could do something like this better in another realm? Because I mean, it's a clear question that would be of interest to almost anyone. Like, what are we getting for our money?

I would really-- I would not want to know what, say, if I were working at a corporation, what other people were making. But it would make the point potentially if you were looking at a very large organization that paid their CEOs quite a bit. That would really be interesting to see--

CEO pay versus performance of stock or something like that.

Right.

Yeah. Something that everybody is concerned about if you're an investor. Yeah. Any other thoughts about that, Sandra?

Actually, the only thought I have about this visualization was that it would have been so much easier if they just put the numbers.

Oh, interesting.

[LAUGH]

Interesting.

If they put the numbers and moved those as the rank instead of putting the lines that just don't add anything.

So let's pose this question to the online student. Here we have a simple concept. How good-- what am I getting for my money? Pay versus performance? And we have a really-- what appears to be an overly complicated visualization. So could you sketch out

for us just a really simple notion of how to do this better? Just very intuitively, how would you do it better?

12.10.3 Discussion Prompt

Please outline how you would make this visualization better.

12.10.4 Visualization Wrap-Up

This element addresses the following learning objective of this course:

- LO5: Identify the audience and the most effective method to communicate a persuasive argument.

So let's wrap up with just a little bit of a reflection on the general principles. Some of my colleagues who think about information visualization from a very technical standpoint will argue that there are no general principles. That we're just learning as we go and that we have to sort of feel our way through this really fast evolving field.

Other people who are in the cognitive science behavioral space will say, well, there are some general principles of how we ought to present information to people. And they are cross-cultural and they go across sectors and it doesn't matter if you're talking about a grocery store or a baseball team. What's your perspective on that question?

So my background is actually in cognitive science. We need to understand the information that we convey, how we convey it. What does it actually-- how does that affect who is viewing it? For the baseball example, there's actually an implicit expectation of cultural norms.

So the downward slope is red and the upward slope is blue. This is definitely a Western red is bad, other colors blue, green, those are good. It's really difficult to separate ourselves from those expectations. So it's really important to understand that we need to be aware of these things.

We can't just throw data at people.

And so for cognitive scientists, there's a really great arbitrage opportunity, intellectual arbitrage opportunity here.

Yeah. I mean, if you know how to manipulate what you're showing you can definitely manipulate your viewers.

I'm in favor of intellectual arbitrage. How does it look from your perspective, Emily?

I'm always reminded of, and perhaps I'm dating myself, in elementary school, when the teachers would have an overhead projector. And they would hide most of the clear slide they were putting up. And then they would unveil it over time and so you just saw a shadow.

And it seems to me that a lot of these either images that we interact with on the web or these visualizations would be much more helpful if they presented one thing and then you got used to that. And they added on the different layers.

Interesting. So we've forgotten what we knew about the overhead projector. Yeah. Great story. Well, thank you so much for working with us on this. Thank you. Thank you.

12.11 Critique a Visualization

Discussions Prompt

Please spend five minutes on the following prompt.

As you view this GIF, think about what changes were made to the visualization: Food Calories

In a sentence or two, describe how this figure was improved. If applicable, in a sentence or two, describe how this figure was made worse.