

The only data visualization guide you'll ever need (in 5 principles)

By Jordan Goldmeier

When I first started teaching data visualization in 2009, I spent a ton of time explaining principles of Gestalt Psychology and Preattentive Attributes. I held up Stephen Few and Edward Tufte as data visualization experts who knew how to rid the world of ineffective charts.

And, perhaps back in 2009, such emphasis on visualization principles was what the world really needed to hear. Tableau was only just becoming a thing and Power BI hadn't yet been released. At the time, Microsoft was touting how easily you could ruin a chart by adding a third dimension. And, I really hated infographics.

But since then, times changed, and I've eased up.

A decade a later, a set of new principles is required.

I've been teaching this stuff in some form or another for the last decade. A pattern has emerged: folks would come to the podium afterward and ask me *how would you visualize such-and-such?*

And I realized something: yes, the principles of data visualization were (and are) still important to teach because they form the foundation for how to design charts and graphs. But, in my pursuit to help analysts become owners of their data products, I've uncovered the following additional problems:

- These principles expose that many analysts feel like they're in over their heads. With this new information, they feel even more lost.
- Data visualization in most organizations is driven by upper management who want their analysts to be "data-vis experts" but then often overrule them. These analysts wonder how they could convince their bosses to do things the right way.
- Many analyst are being tasked to create infographics, which deviate from the tried-and-true visualization principles.

Having fielded hundreds of questions from analysts who work in industries spanning government, banking, retail, fashion and more I truly think more needs to be said about data visualization if only to help my fellow analysts who work in the trenches.

My 5 Principles of Communicating Data

As a result, I've come up with the following principles of communicating data.

1. What we present must always be the natural extension of the underlying problem.
2. Visual complexity often results when we don't know our audience.
3. Visual simplification often results when we don't trust our audience.
4. Proper data visualization is about trade-offs.
5. You will always win if you respect your audience.

Infographics vs. Data graphics

Before moving forward, I want to draw a distinction between infographics and data graphics.

Less Complex

Infographics



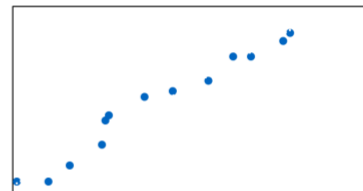
■ 1st Qtr ■ 2nd Qtr

- Analysis is complete.
- Charts support a preexisting narrative.

High-Level Audience

More Complex

Data Graphics



- Exploratory Analysis
- Charts used to help identify results.

Subject Matter Experts

I don't present either of these as discrete categories. Rather, both exist on a continuum, heavily mediated by personal preference, company culture, and audience. At their core, however, we can think of the two categories like so:

- **Data graphics** are for when we are investigating data, looking to uncover some story.
- **Infographics** are for when we already have the story but we want graphics to support the message.

Data graphics look to follow data visualization principles more rigorously. Activities that would generate data graphics include exploratory analysis and research. We should expect these charts from engineering teams, data scientists, and statisticians looking to communicate the results of their activities with confidence. The operative word in this definition is *investigation*.

Infographics, on the other hand, serve an already completed analysis. These charts and graphics exist to support and assist in delivering a preexisting *narrative*. Activities that generate infographics include end-of-year reports, impact reports for non profits, and keynote presentations. We should expect these charts from business analysts, marketing analysts, and (some) c-suite executives. Infographics don't need to follow data visualization rules to a tee, though they should *never* confuse the message being delivered (more on that in a second).

A personal inventory - which do you like more? Infographics or data graphics?

To make the most of the principles I'm about to go through, you need to do a personal inventory.

Where do you personally fall on this continuum?

What types of graphics do you enjoy creating?

Knowing the answers to these questions will help you best make use of the following principles.

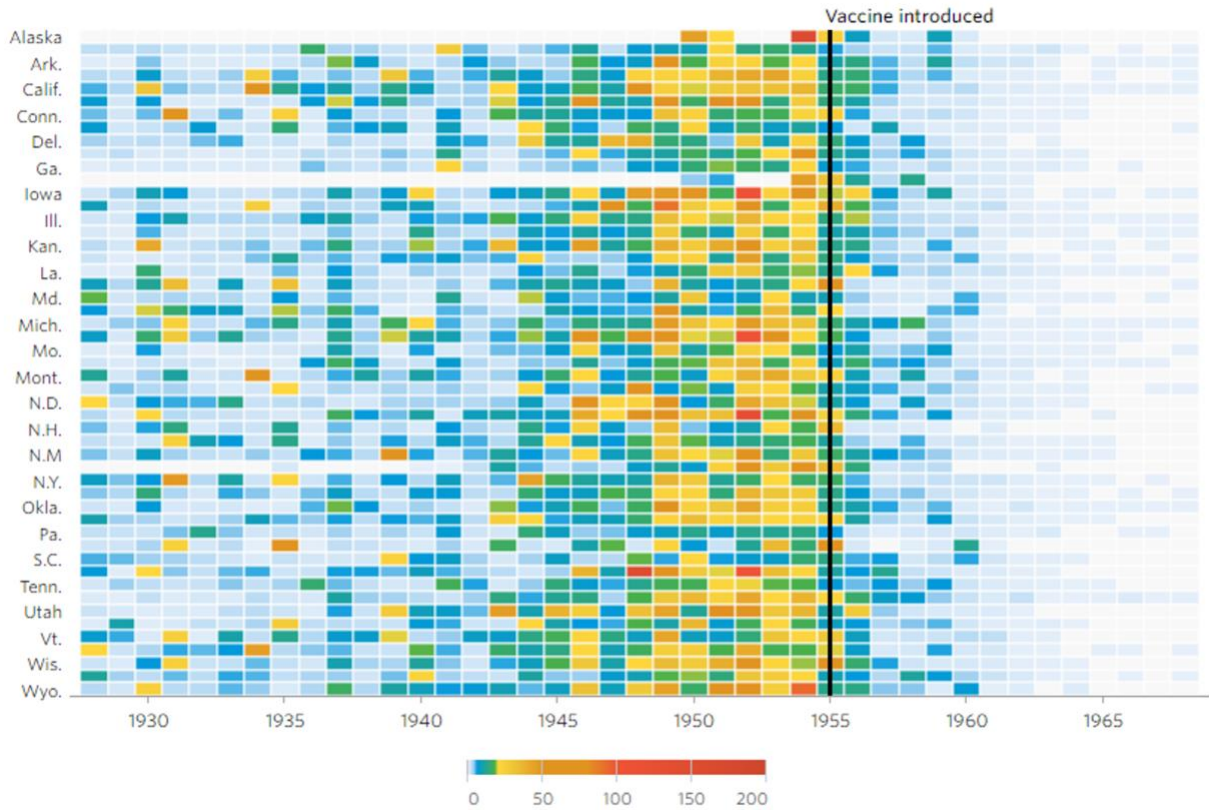
Principle 1 - What we present must always be the natural extension of the underlying problem.

No matter how you choose (or are asked) to present data, what you present must be the natural extension of the underlying problem. To illustrate what I'm talking about, let's compare two charts that often serve as fodder for this discussion.

The problem: demonstrating the frequency of polio cases in the United States before and after the vaccine was introduced.

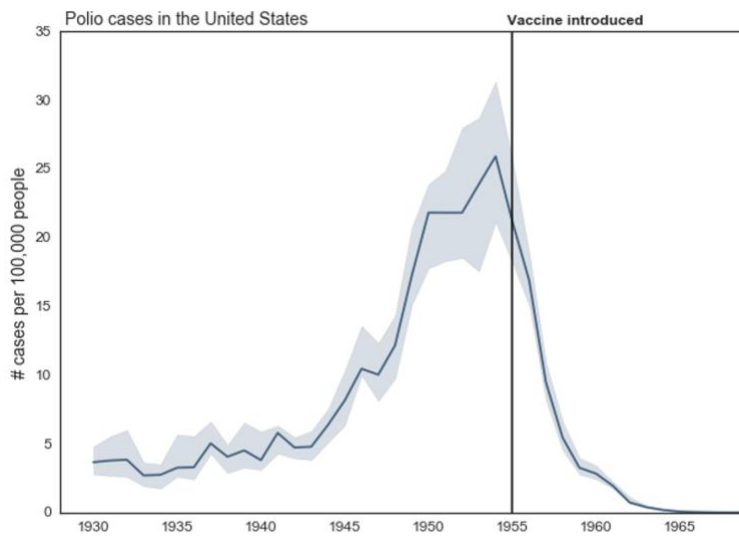
The Wall Street Journal developed the following chart:

Polio



Battling Infectious Diseases in the 20th Century: The Impact of Vaccines, Wall Street Journal

And the Royal Statistical Society created this one:



Data source: Project TYCHO (tycho.pitt.edu) | Author: Randy Olson (randalolson.com / @randal_olson)

The former chart could be considered more of an infographic. Many people prefer this graphic to the second one as it provides more information and appears more interesting to look at. And, personally, I don't hate it.

That said, from a data visualization standpoint, the chart could also give the impression that certain “hot areas” have something in common. Yet, the order the data is presented in is arbitrary: the designer made the choice to present the data alphabetically but there is no underlying ordinality at play here. In addition, it uses multiple colors to encode one single continuous variable rather than, say, a single shade of that color. (We would expect multiple colors like this to be used in a divergent series.) Finally, because it skips state labels, it doesn't really allow you to look up information for a specific state (an easy fix if we wanted).

The latter chart focuses on the larger narrative, placing stronger emphasis on the range of cases and its underlying trend. Moreover, it uses two colors: a lighter blue to show the range, and a darker blue to show the average trend. Finally, it's largely devoid of what Edward Tufte calls “Chart Junk,” defined as any ink on a chart that doesn't specifically contribute to the overall message.

But it also feels unapproachable to certain audiences. And, if you wanted to understand true state-by-state results, you wouldn't be able to.

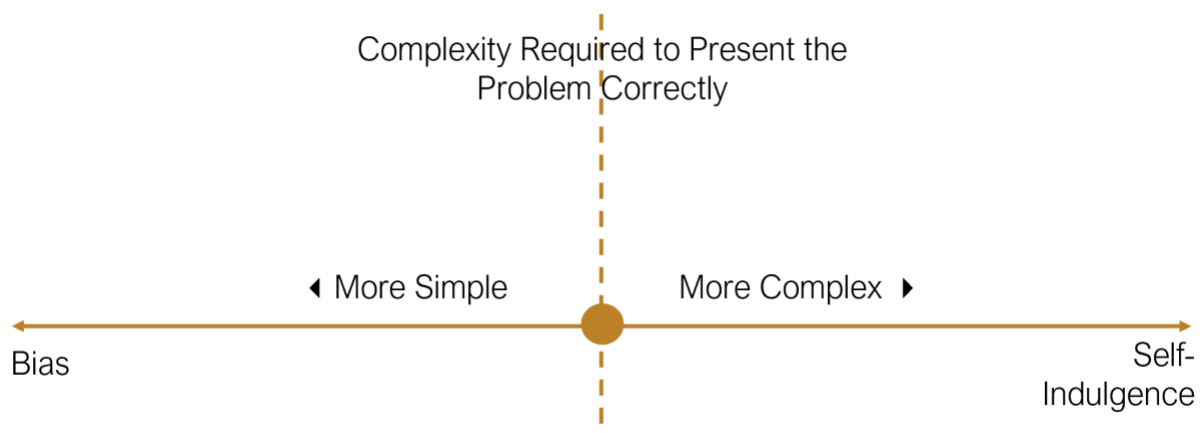
So, which is better?

~_(ツ)_/~

It all depends on you and what you feel confident presenting. That being said, both demonstrate two ends of a set of visual choices we can make. The WSJ journal's version is more visually complex, the Royal Statistical Society's version is more visually simple.

Problem complexity vs. visual complexity

Let's assume there is a correct level of visual complexity and simplicity that communicates the problem correctly depending upon you and your audience.



Most of us will have a personal preference for how to present data whether we tend toward the complex or the simple. It's this tendency that causes a lot of stress among analysts, especially when what we're being asked to produce is out of line with personal preference. So it's worth thinking about the personal inventory from above. Which side of the spectrum do you feel yourself most naturally pulled?

When we make things more complex than they need to be it's the result of a self-indulgence (at least, in my case). That is to say, I am putting myself before the audience. Because I don't really know them, I am making things as complicated as possible so that I can field any question asked – and so they think I'm smart.

On the other hand, when we make things that are too simple, it's because we have might have some type of bias. The most common bias is that our audience isn't smart enough to understand what we're presenting.

Naturally, the next two principles flow from this discussion:

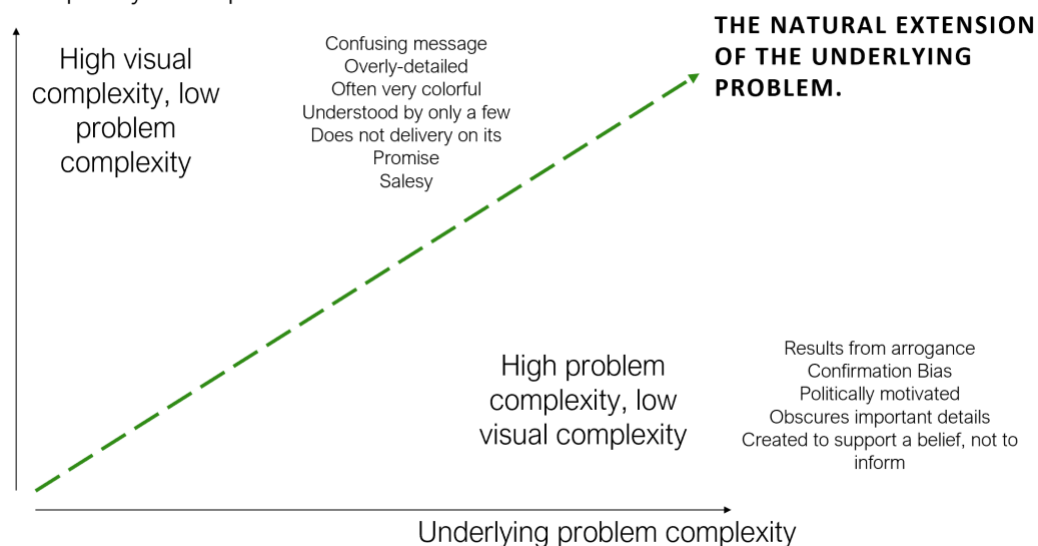
Principle 2 - Visual complexity often results when we don't know our audience.

and...

Principle 3 - Visual simplification often results when we don't trust our audience.

Continuing on this theme, I believe the right answer is to take our data visualizations and align them to the following model.

Complexity of the presentation.



On the X-Axis, we model increasing problem complexity; on the Y, we model increasing visual complexity. Note, again, the underlying problem is not something I propose as being static. Rather, it's a function of personal preference, culture, and the question you're truly trying to solve.

In any case, when the visual complexity is high, but the problem complexity is low, the results are often overly detailed and colorful. Think vendor sales material.

Likewise, when the problem complexity is high but the visual complexity is low, there's usually something being obscured deliberately so that you won't ask questions. When I say politically motivated above, I don't just mean federal politics, I also mean office politics.

Most importantly....

Principle 4 - Proper data visualization is about trade-offs.

To really understand if your data visualization gets the job done, it's worth trying to identify where your design lands on the problem complexity vs. visual complexity model.

Think about this way. I am implying the natural extension of the underlying problem can be fit to some linear model. But remember: *all models are wrong and only some are useful*. Expect your data visualization to have some error from the line as your personal preference creates noise. It's far more important to understand where you are with respect to the line than it is to create something that fits to the model exactly.

Practice thinking economically about your data visualizations. You might not always be able to show exactly what you want the way you want, but you can often get close. Take note of what you're leaving out, because that information can be delivered to management should you try to encourage them to think differently.

Can't decide the best way to visual your data? Present all of your designs. Sometimes two different designs aren't in conflict: they're just telling a different story. Remember, there's no right answer here. There's only trade offs. (To be fair, there are plenty of wrong answers.)

Principle 5 - You will always win if you respect your audience.

I think the over emphasis on data visualization has created a league of analysts who think they're smarter than everyone. I think you can learn a lot if you understand where your audience is coming from.

Personally, I hate the idea that we're told to present information to leadership like they're in the fifth grade. Platitudes like "if you can't explain it simply, you don't understand it," disrespect

your audience, your craft, and Albert Einstein (to whom the quote is often erroneously attributed).

Matching problem complexity with visual complexity is a continuous negotiation between your personal preference, your company culture, and the problem you're trying to solve. (I know I'm repeating myself here, but it's worth repeating these points!)

Conclusion

Data visualization is as much an art as it is a science. Unlike a linear system which might have an optimal solution, data visualization is heavily contextual. Rather than seeking to become an expert you should look to create an atmosphere in which you are most comfortable presenting your results. Confidence in what you do will be the single most important factor in driving change in your organization. I present these lessons to you so that you should find the confidence in yourself.

Rather than resent management for not seeing it your way, take control and accountability for what you to produce. Stand up for what you believe is right, and you'll make a difference. I promise you.

You won't win every battle trying to change the minds of those around you (and don't expect to). But simply having an opinion, a desire to improve, and basic respect for others will make you stand out in the crowd.

I encourage you to think like an analyst in all of this: improvements follow naturally from iteration. To become better at data visualization, use the tools I've provided you. (And then let me know what improvements you have!) Create a working group in your company to solicit feedback. And remember you're not looking to make it perfect each time, rather you're looking to see if your visualization gets you to the point where you can iterate into future improvements.

I wish you the best of luck. I know you'll do well.