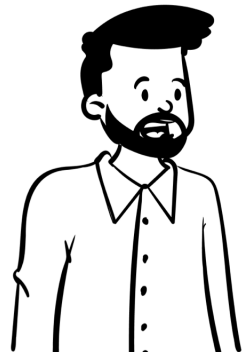
Communicating Data Visually



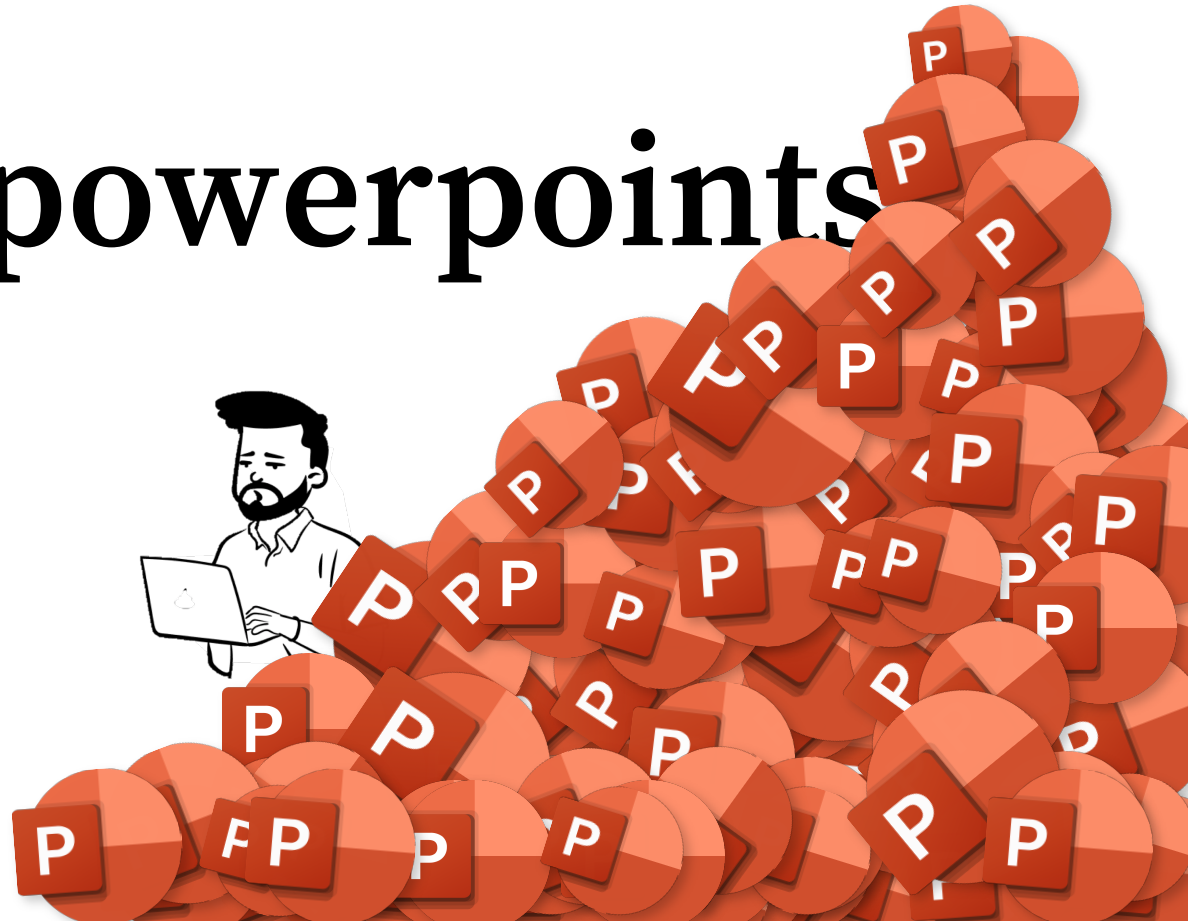
Matt Dzugan
@mattdzugan
mattdzugan.com
mattdzugan

👋 Hello ODSC, I’m Matt Dzugan

Director of Data Science at project44

@mattdzugan on twitter

I make a lot of powerpoints



Takeaways

Be intentional about creating the scope of a visual.
Is it meant to **explore** data or **explain** data?

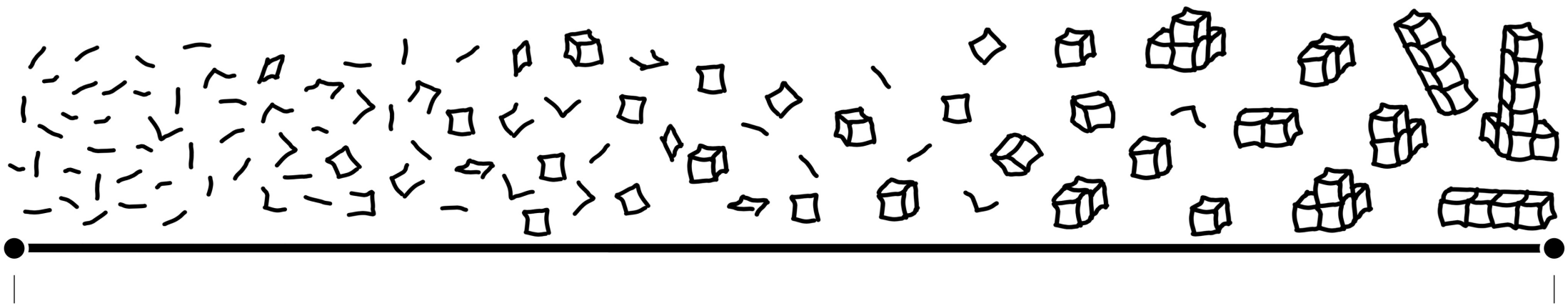
A list of **tools** of sources of **inspiration** for you to kickstart your journey into data visualization

Some concrete **DOs** and **DONTs** to help make your communication as professional as possible

A Mental Framework

There's no "right way" to create a good data visualization - but there are a few things to consider that will help you consistently succeed

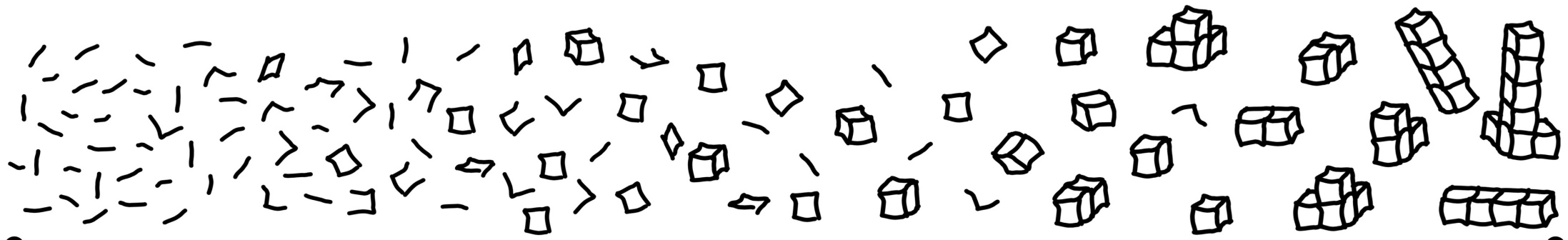
Data Communication Spectrum



Raw Data

Bottom-Line Summary

Data Communication Spectrum



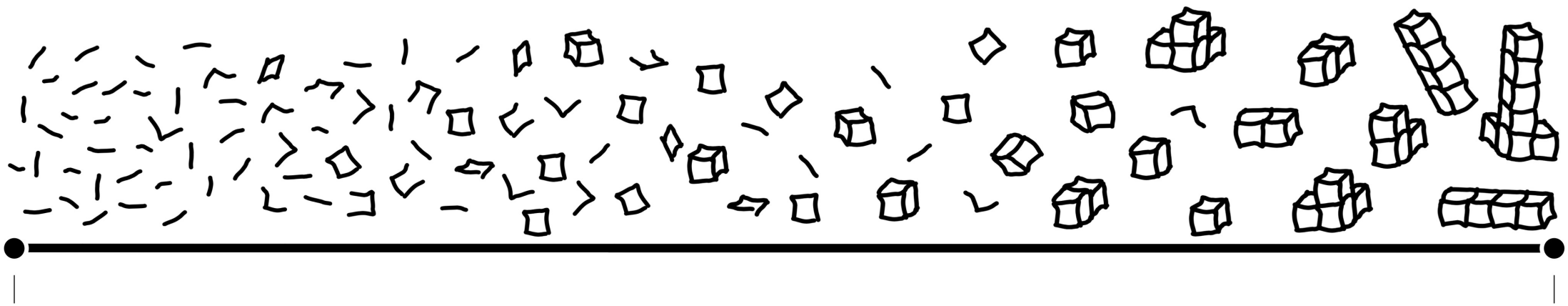
Raw Data

Bottom-Line Summary

Here are the results from the xyz model using the latest satellite measurements of the air column

You will want to wear a sweater outside tomorrow

Data Communication Spectrum



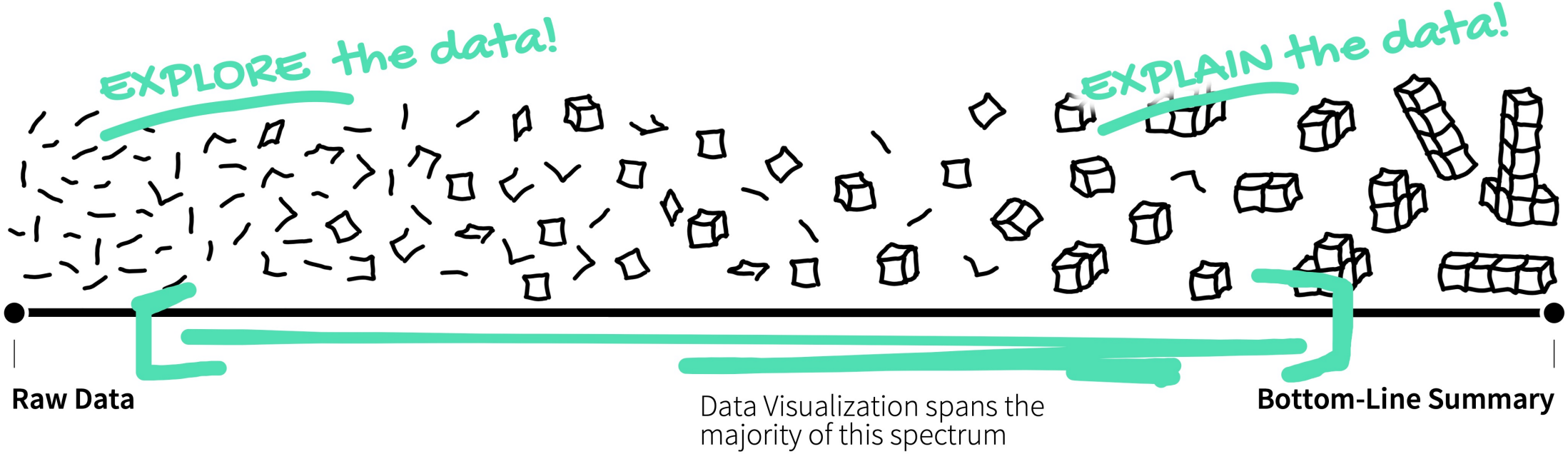
Raw Data

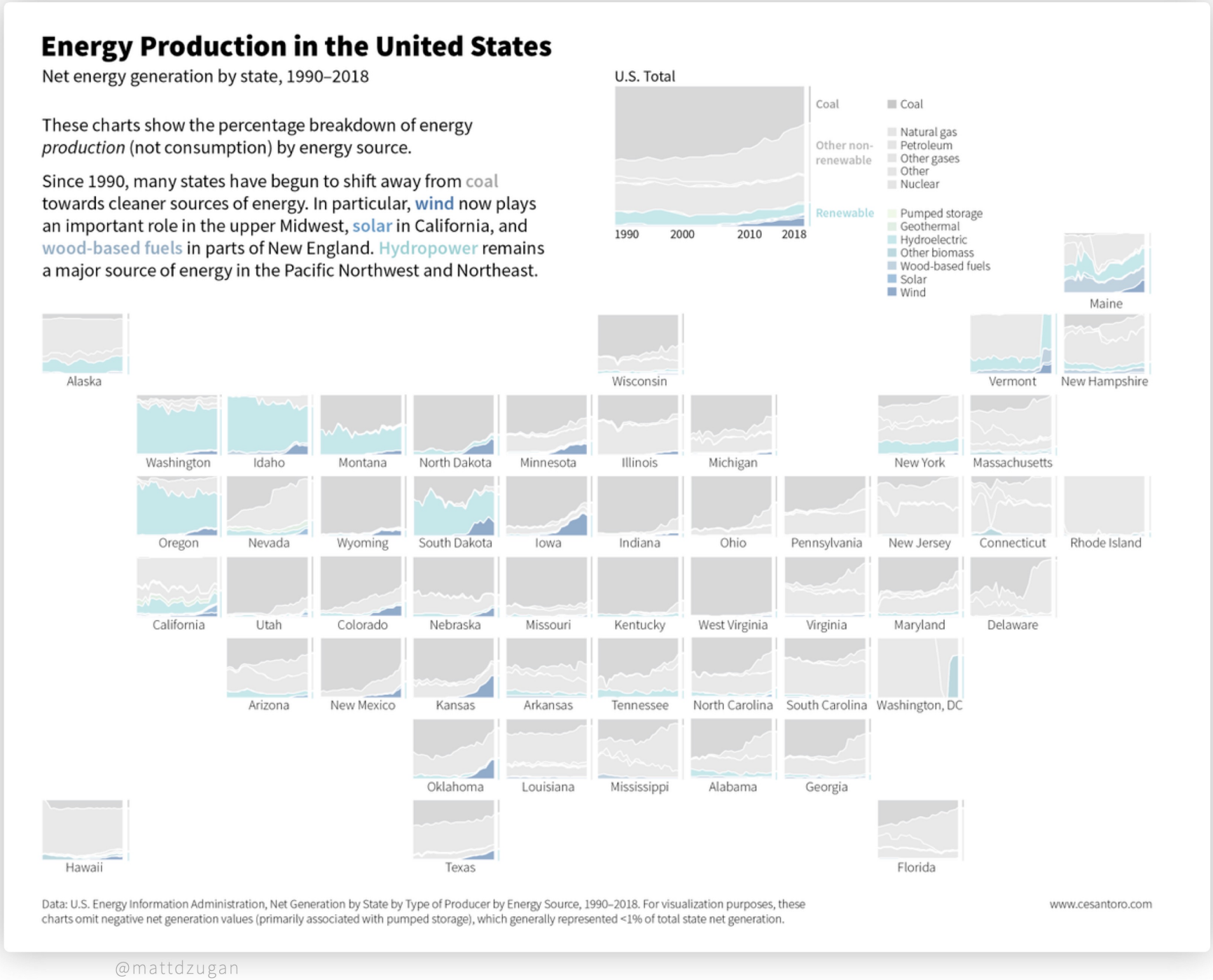
Lowest effort to prepare
Interpretation burden is on audience
Pure Facts, without bias
Multiple conclusions can be drawn

Bottom-Line Summary

Highest effort to prepare
Interpretation burden is on author
Author's biases may be present
Author controls how/which conclusions drawn

Data Communication Spectrum



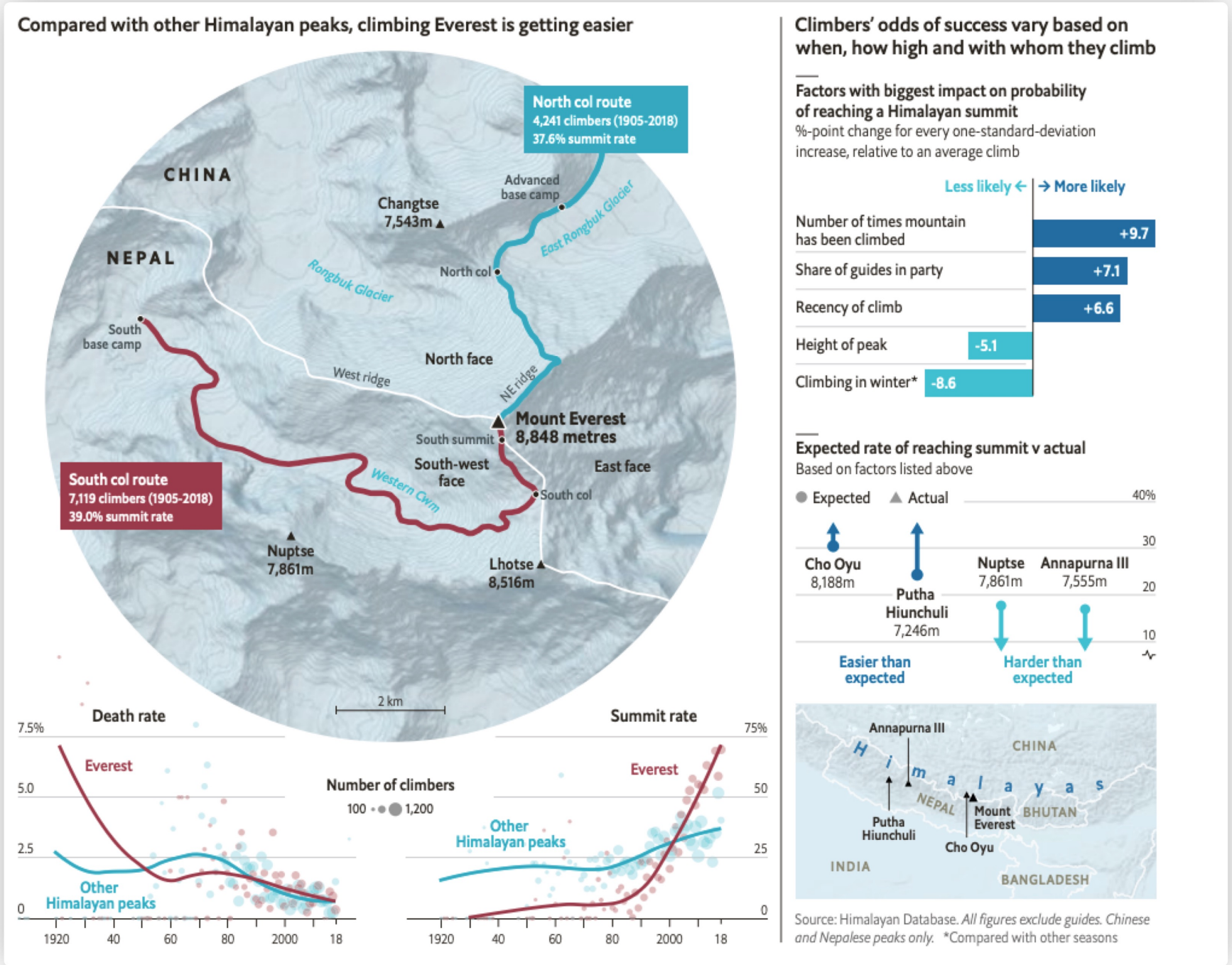


@mattdzugan

Explore

Notice how the layout allows you to quickly seek out information, in a way that is much more intuitive than looking at numbers in a table.

How many different “headlines” could you imagine walking away with from this graphic?

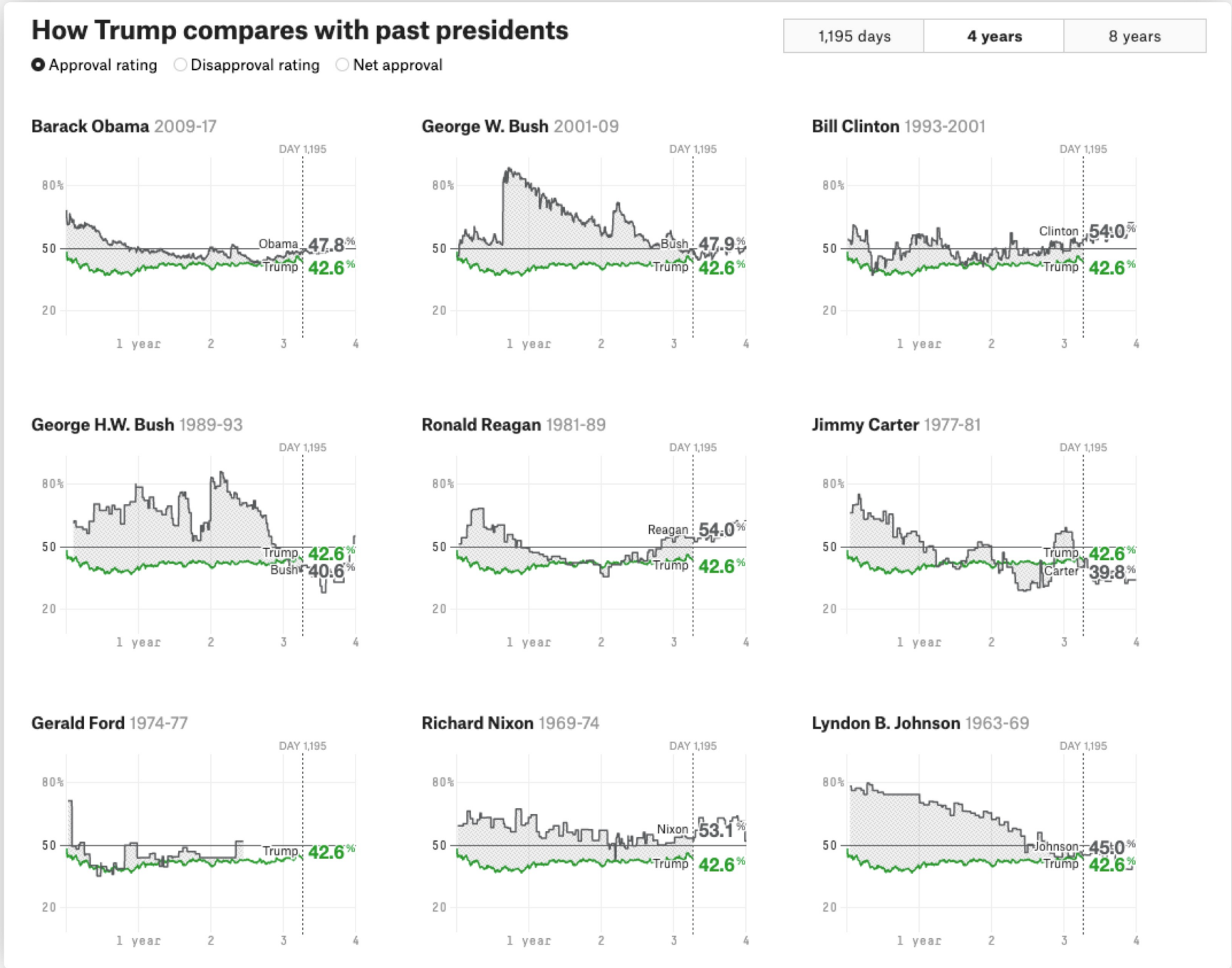


@mattdzugan

Explain

Notice how each of the 5 component-graphics all support the claim that Everest is getting easier to climb over time.

The author of the graphic has made an effort to convince you of this point. Are you skeptical at all? Why or why not?



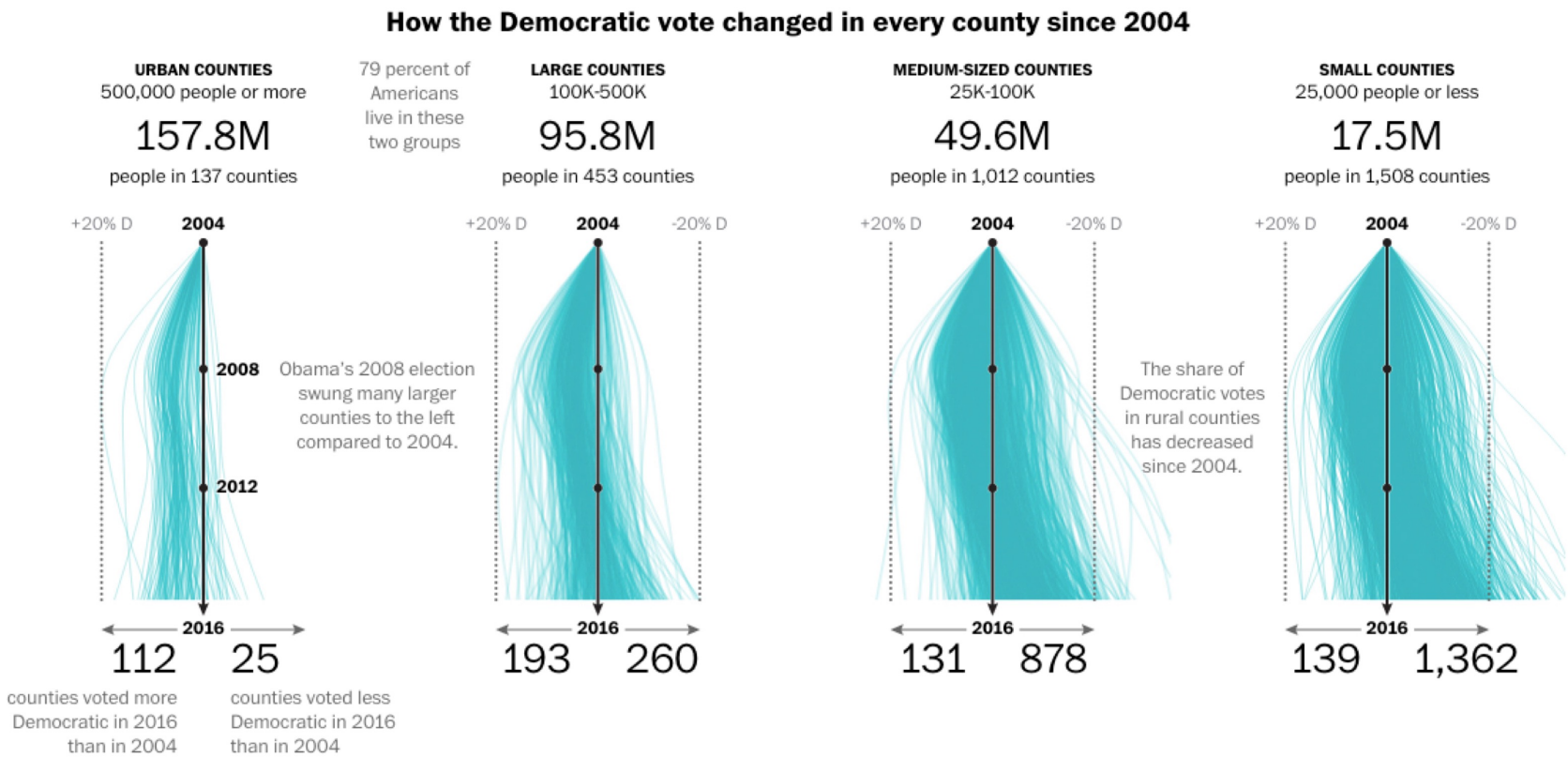
@mattdzugan

Explore

On one hand – this series of graphics offers a clear explanation that Trump's approval ratings are lower than his predecessors – but on the other hand it offers so much additional data that there are plenty of other relationships to explore.

Urban and rural America are becoming increasingly polarized

Since Obama's election in 2008, the trend of urban counties voting for Democrats and rural counties voting for Republicans has grown stronger.



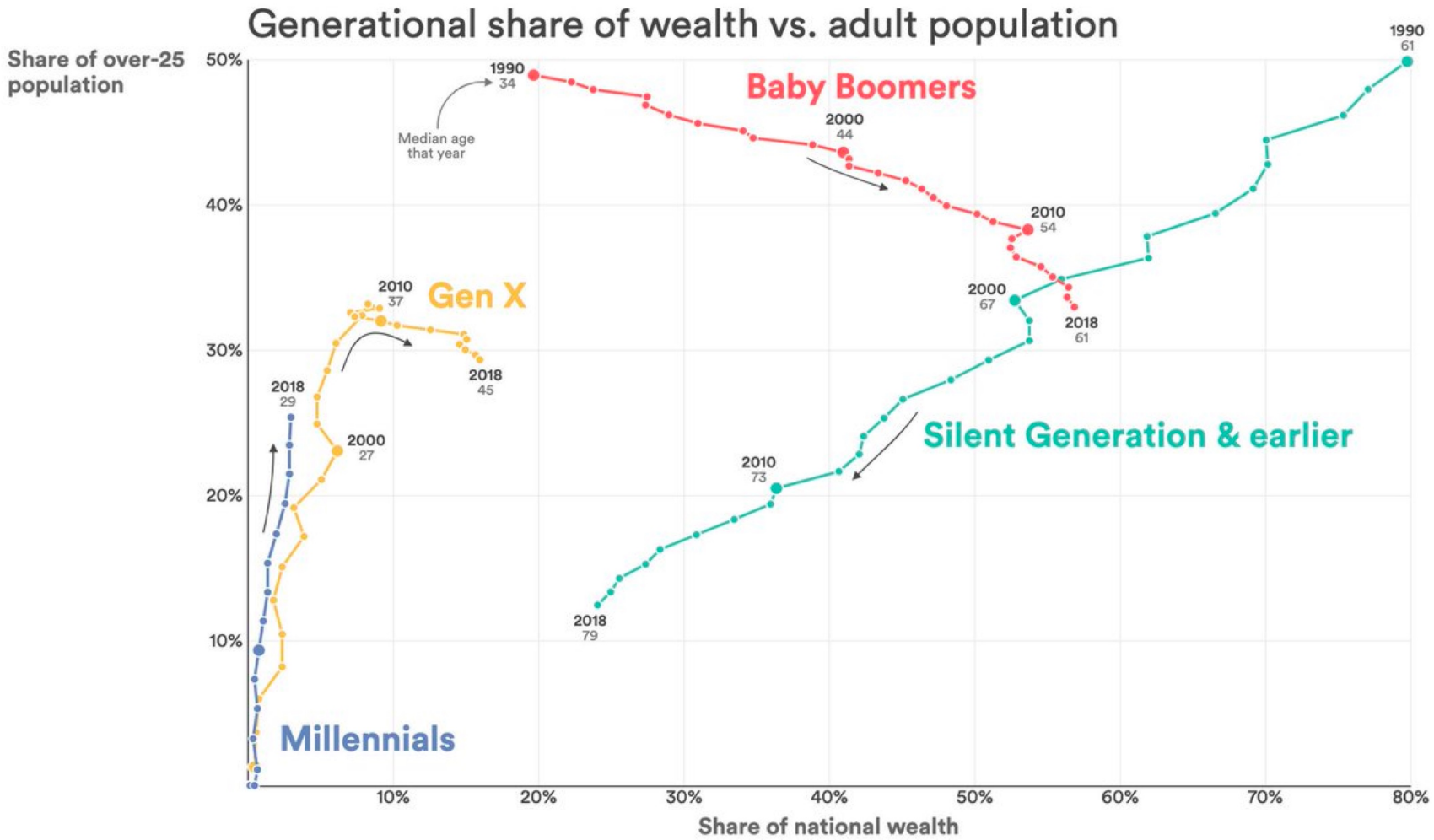
By Lazaro Gamio
Nov. 17, 2016

@mattdzugan

Explain

This visualization is an excellent example of being tailored to support one specific message. And it does a great job of showing the underlying data transparently while avoiding other junk.

For example – is there anything else (other than the headline) you can learn from this chart? Probably not. It's very focused.



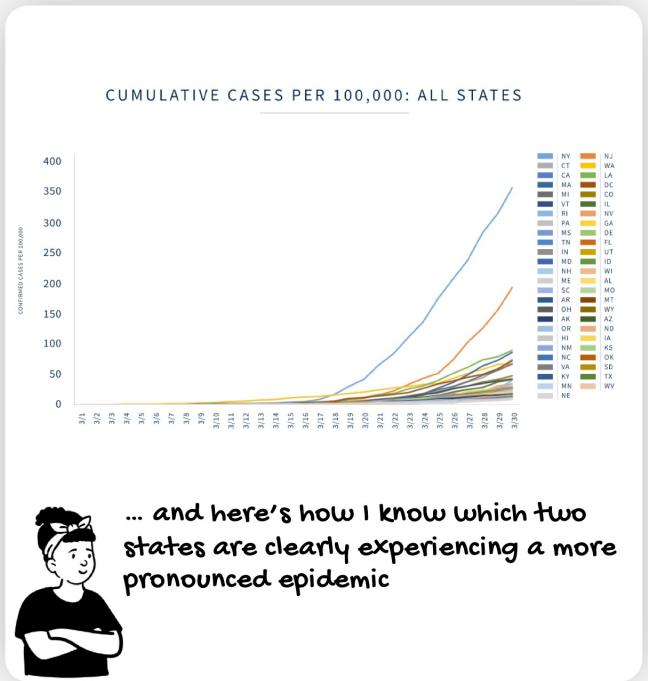
Explore

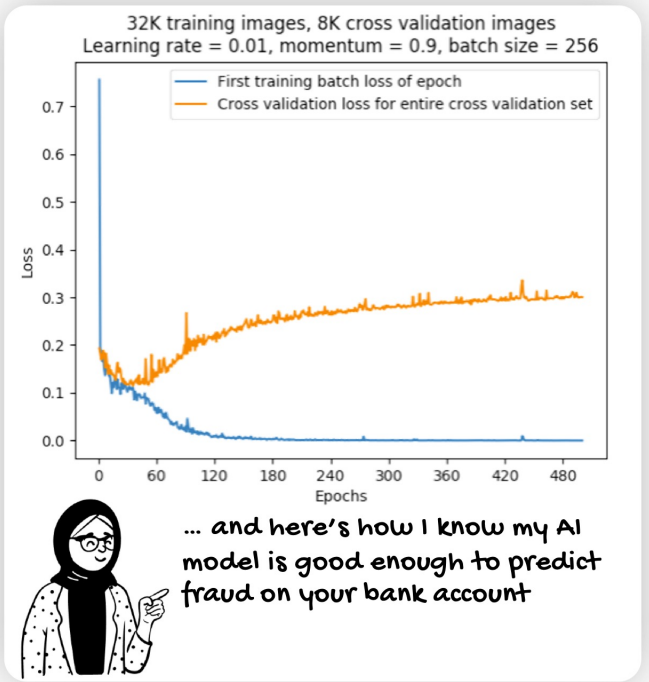
This graphic is a good example of something that takes a bit of mental investment.

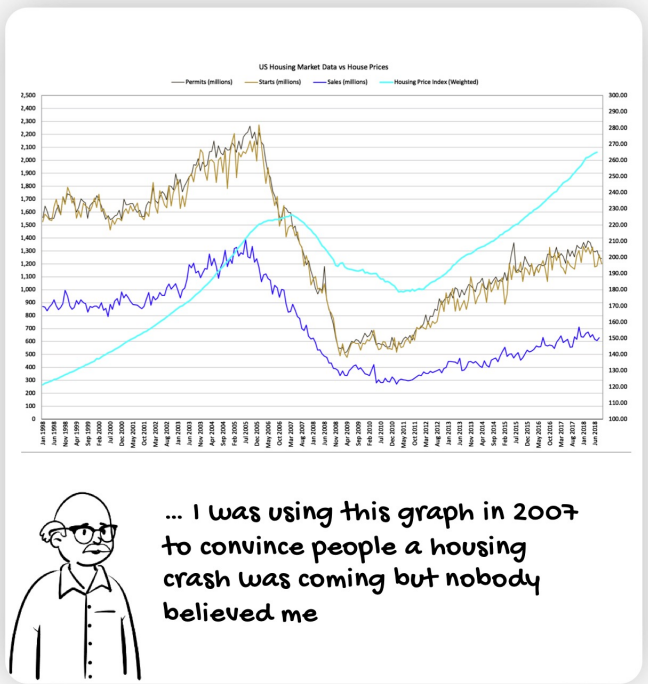
It's not immediately obvious "how to read it". But once you catch on – the number of conclusions you can reach is almost endless.

A Common Mistake…

A graphic used by an expert to **explore** their own data, is pasted directly into a document to **explain** a concept to a less-informed audience. Just because it makes sense to the creator, doesn't mean it offers the intended **explanation**.



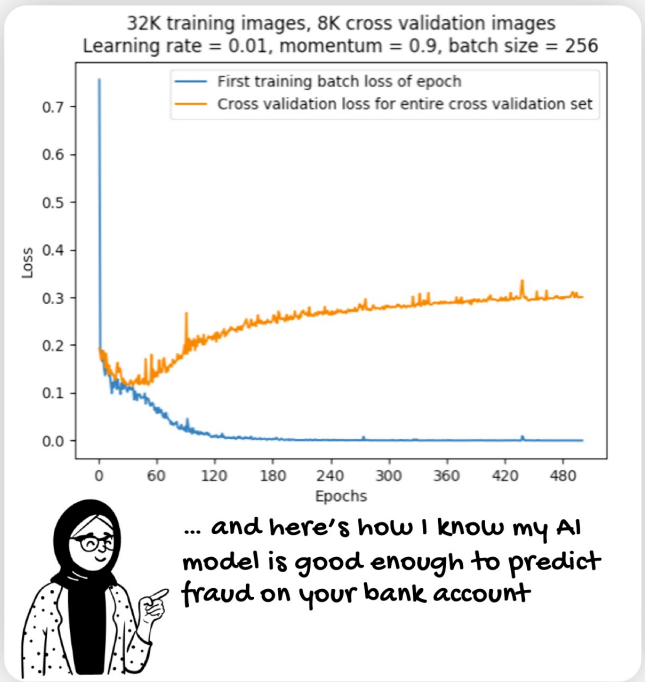




The context & background knowledge of these graphics' creators are necessary to understand the graphs

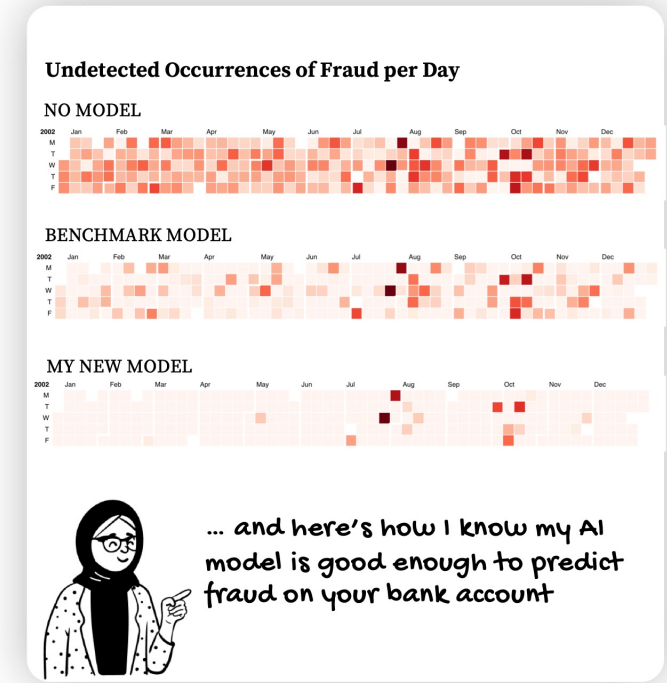
…A Potential Solution

Identify the core message being conveyed by your graphic. Something so simple it can be described in 10 words or less. Once this message is identified, try and design a graph that effectively conveys that simple message.



Take a step back.
What am I trying to explain?
"The model is good"
What's a clear way to show the model is working well?

Maybe a calendar of all undetected cases of fraud with 1) no model 2) a benchmark model & 3) my model



The graphic you use to discover the insight may not be the right graphic to communicate the insight

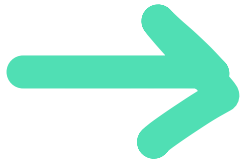
Design Considerations

Good visual communication is equal parts "form" and "function". Without both you either lose your audiences interest or convey no information.

Two Distinct Steps

1. Develop Your Story

- Who is the audience?
- What do they know? [current state]
- What do we want them to know? [desired state]
- What is the smallest set of info that takes them from [current] to [desired] state?



2. Create your Graphics

- Above all else, show the data
- Show context as needed
  - Remember audience's current state
- Highlight key points as needed
  - Remember audience's desired state
- Remove unnecessary Chart Junk

Develop Your Story Structure Before You Do the Project

I realize it sounds crazy
but I promise this is magic

Example for a ML Model

There's a problem in the world

Here's How Users Try to Solve it Today

We can Measure their Success (benchmark)

We have XYZ data available to us

Our approach to solving the problem

Our success metrics vs benchmark

Example for a ML Model

There's a problem in the world

Here's How Users Try to Solve it Today

We can Measure their Success (benchmark)

We have XYZ data available to us

Our approach to solving the problem

Our success metrics vs benchmark

Remember how YOU viewed the world
BEFORE you started your project?

That's probably where your audience is today!

Example for a ML Model

There's a problem in the world

Here's How Users Try to Solve it Today

We can Measure their Success (benchmark)

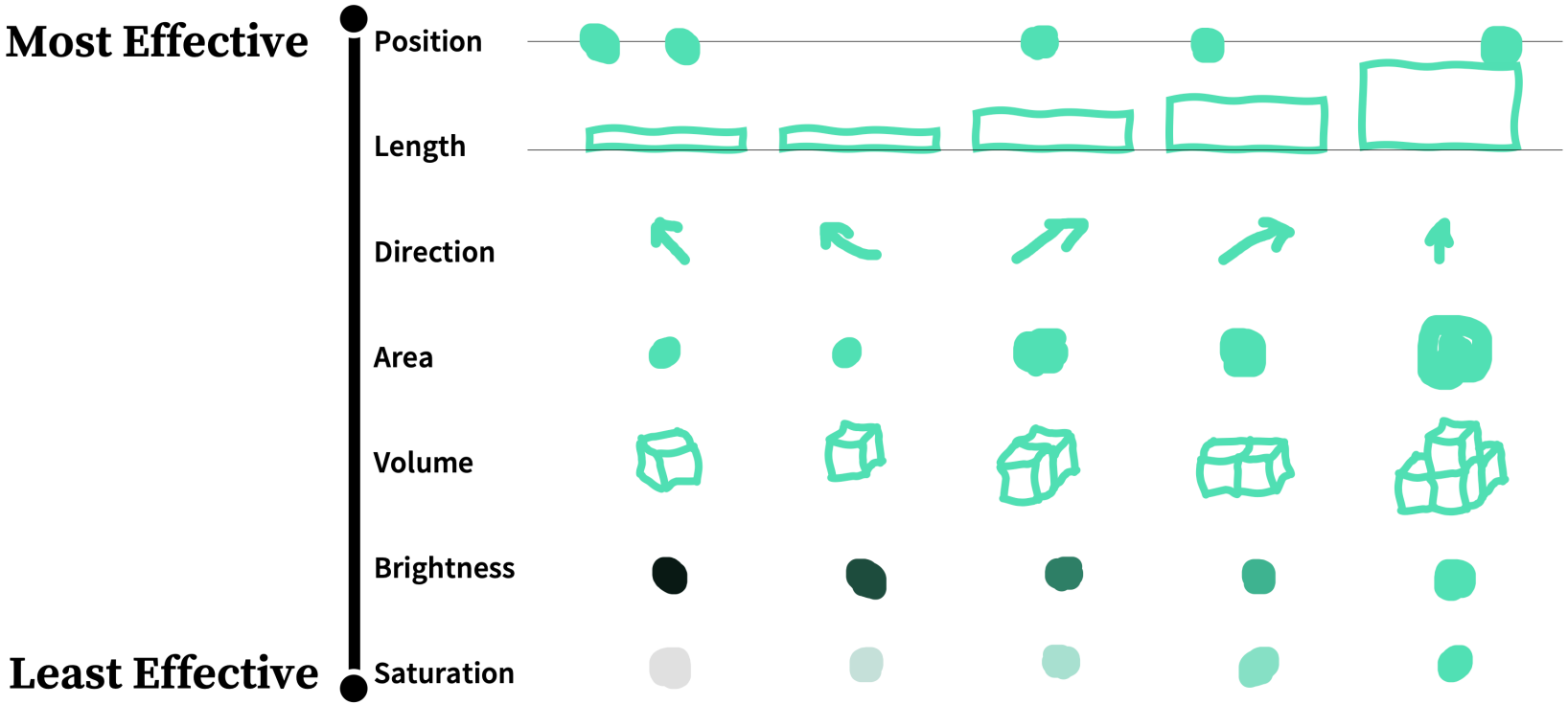
We have XYZ data available to us

Our approach to solving the problem

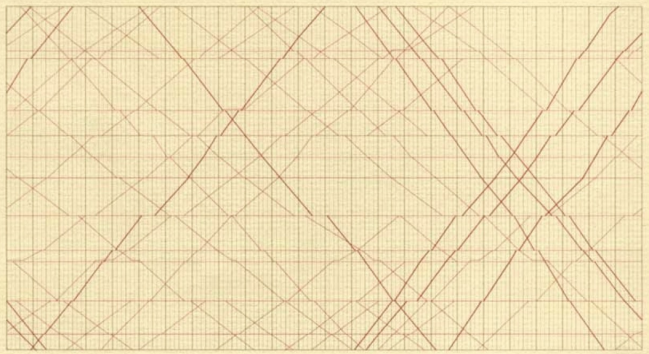
Our success metrics vs benchmark

What To Consider

Studies have identified the easiest ways for people to understand differences in quantitative data, shown here on a scale from most effective to least.



The results of surveys tracking audiences' understandings of charts and graphics are clear: Not all visual elements are created equal

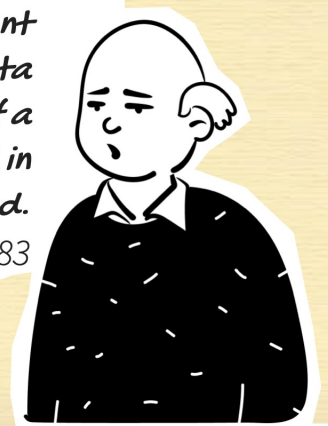


The Visual Display of Quantitative Information

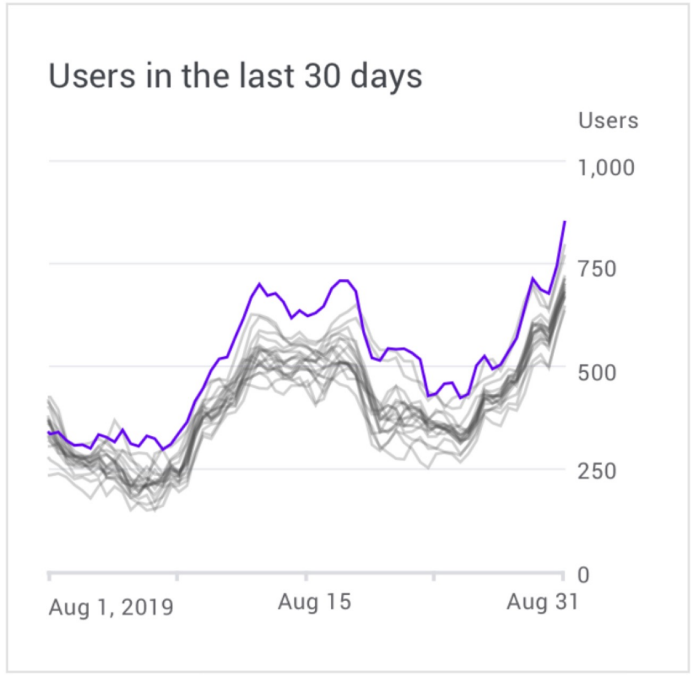
EDWARD R. TUFTE

Above all else show the data
Tufte, 1983

A large share of ink on a graphic should present data-information, the ink changing as the data change. Data-ink is the non-erasable core of a graphic, the non-redundant ink arranged in response to variation in the numbers represented.
Tufte, 1983

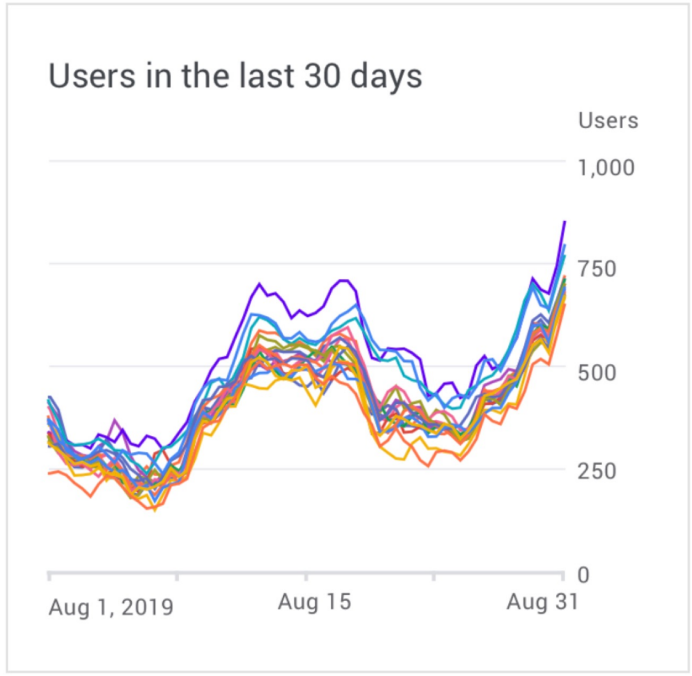


Highlight with Color, don't Identify



Do

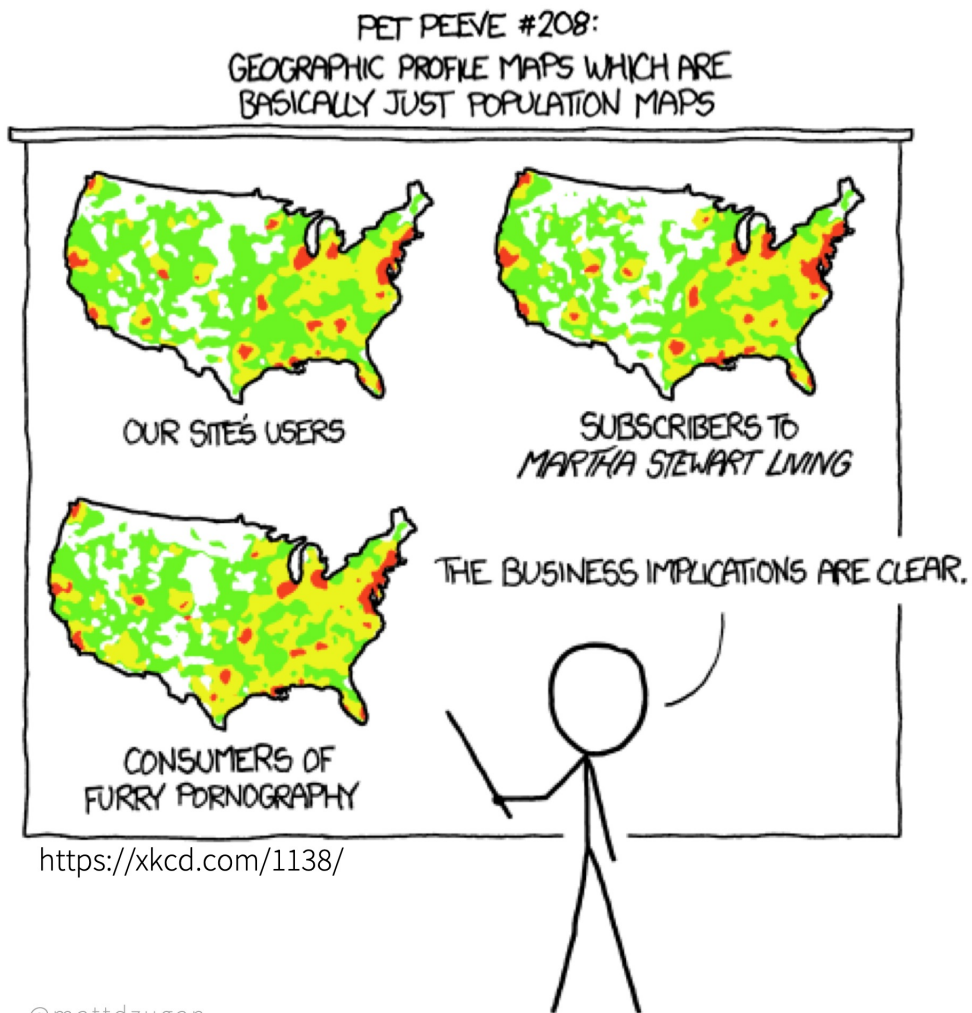
Use a combination of color highlights and neutral colors to provide contrast and emphasis.



Caution

Many colors in a single chart can hinder focus.

Transform your data before Plotting



https://xkcd.com/1138/

More often than not, normalizing your dataset by population ("per capita") or area is going to be helpful before mapping it

FiveThirtyEight

One thing that bothers me is large numbers presented without context. We're always seeing things like, "This canal project will require 1.15 million tons of concrete." It's presented as if it should mean something to us, as if numbers are inherently informative. So we feel like if we don't understand it, it's our fault.

Randall Munroe - XKCD

FiveThirtyEight

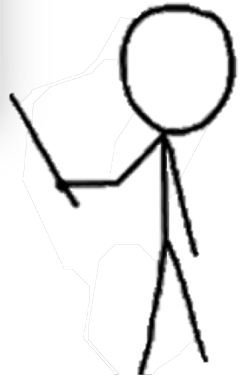
But I have only a vague idea of what one ton of concrete looks like. I have no idea what to think of a million tons. Is that a lot? It's clearly supposed to sound like a lot, because it has the word "million" in it. But on the other hand, "The Adventures of Pluto Nash" made $7 million at the box office, and it was one of the biggest flops in movie history.

Randall Munroe - XKCD

FiveThirtyEight

It can be more useful to look for context. Is concrete a surprisingly large share of the project's budget? Is the project going to consume more concrete than the rest of the state combined? Will this project use up a large share of the world's concrete? Or is this just easy, space-filling trivia? A good rule of thumb might be, "If I added a zero to this number, would the sentence containing it mean something different to me?" If the answer is "no," maybe the number has no business being in the sentence in the first place.

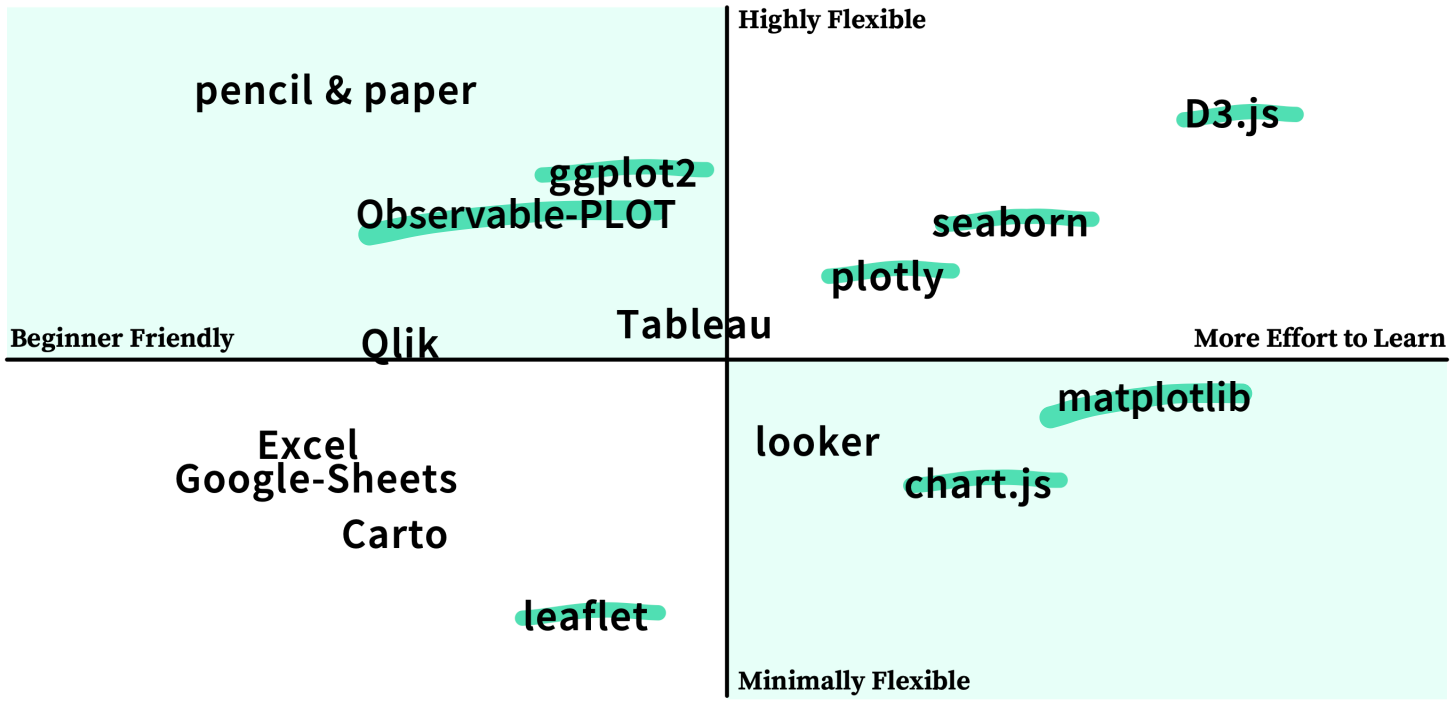
Randall Munroe - XKCD



Tools & Software

Sometimes you need to make a graphic as quickly as possible. Other times it's worth taking a moment to create something customized. There's tools out there for everything.

Landscape of Tools

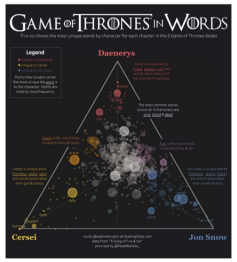


Coding Skills Required

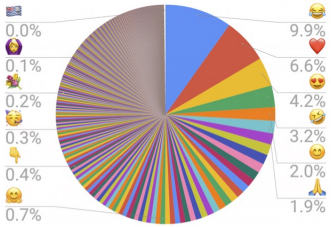
You might also think about a tool's ability to create interactive visuals, or its ability to handle #bigdata

Avoid “Tool-Worship”

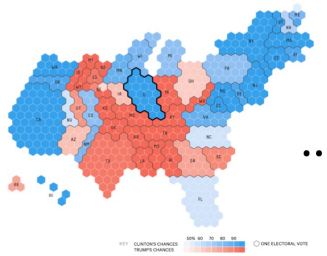
Tableau can do



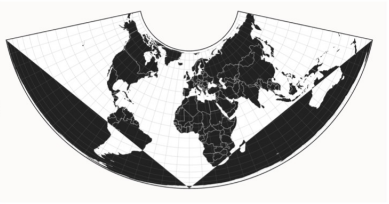
...but it can also do



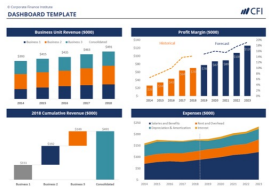
D3.js can do



...but it can also do



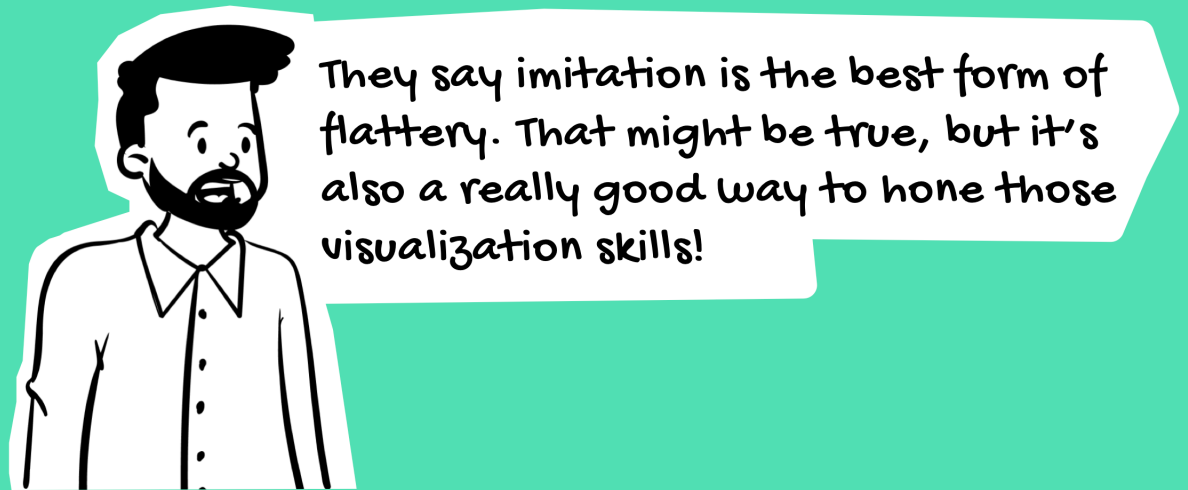
Excel can do



...but it can also do

A shiny new pair of basketball shoes isn't going to make me as good as Michael Jordon. Also, Michael Jordon could beat me even if he was wearing ski boots.

Learning by Example



Inspiration

Twitter
Hashtags like #dataviz, #MakeoverMonday and #TidyTuesday & posts from the best of the best

Big Name Newspapers
Washington Post, New York Times, The Economist all publish phenomenal Data Visualizations

Boutique and Niche Publications
Sites like pudding.cool, flowingdata.com, distill.pub, fivethirtyeight.com lead the way in viz innovation

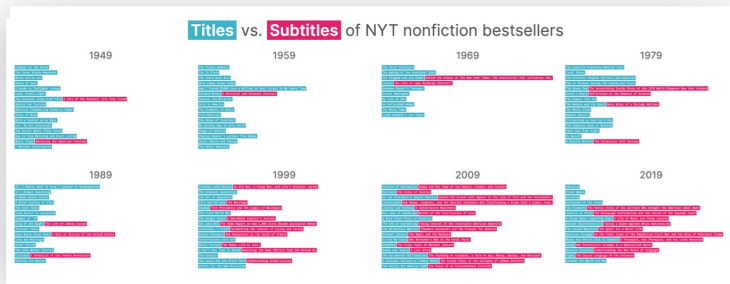
reddit.com/r/dataisbeautiful
Tons of great and diverse content from folks around the world including a monthly contest!

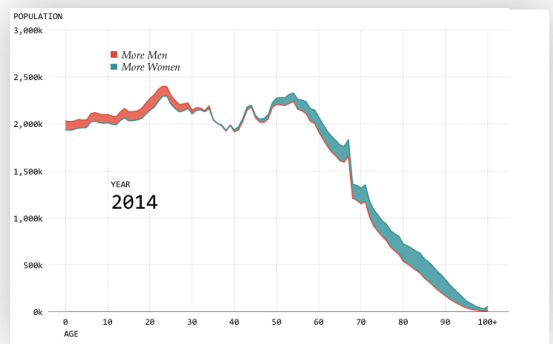
Data Visualization Society
The “DVS” aims to collect and establish best practices and foster a community

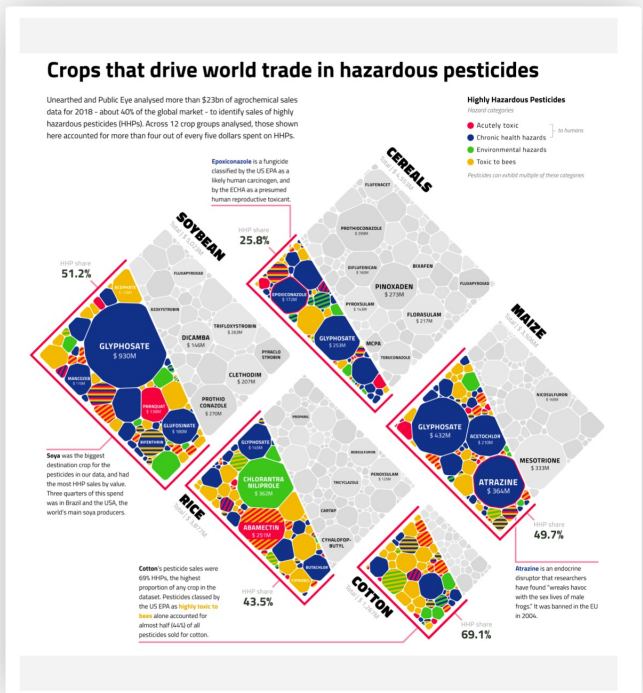
Twitter

Here's an initial list of folks to get you started

@duncangeere
@AlliTorban
@alangwilson
@AmyCesal
@yuliakrolik
@veltman
@jburnmurdoch
@adamrpearce
@W_R_Chase
@billshander
@NadiehBremer
@infowetrust
@darenw
@MollzMP
@thomasp85
@DataToViz
@Elijah_Meeks
@lisacrost
@flowingdata
@AlbertoCairo
@codenberg
@micahstubbs
@mbostock

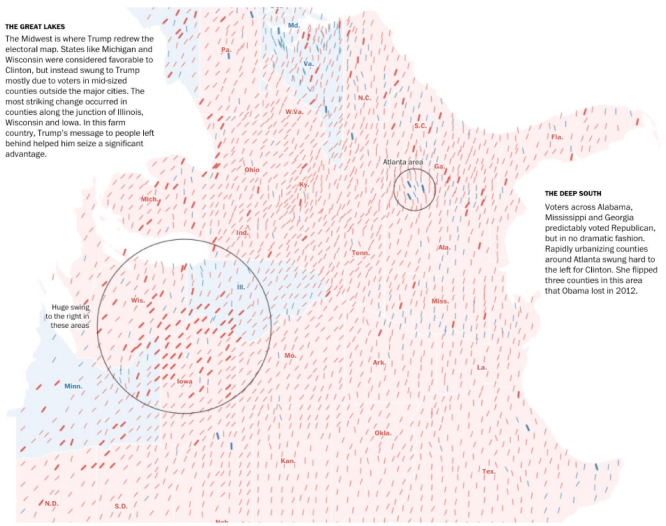




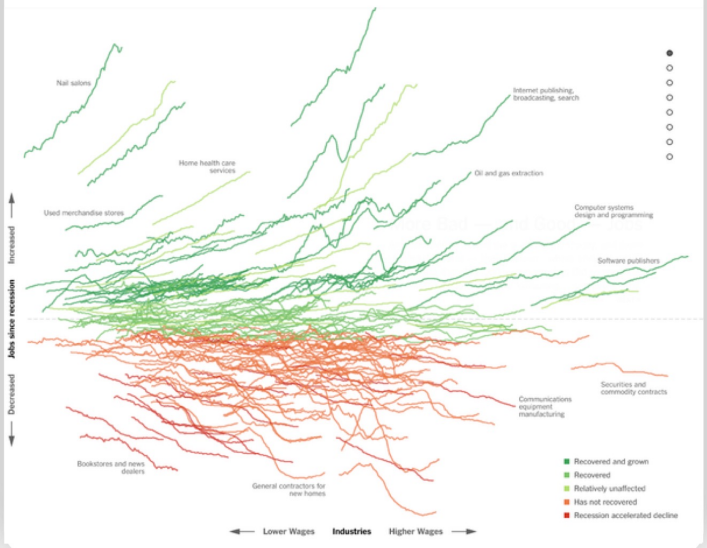


Newspapers

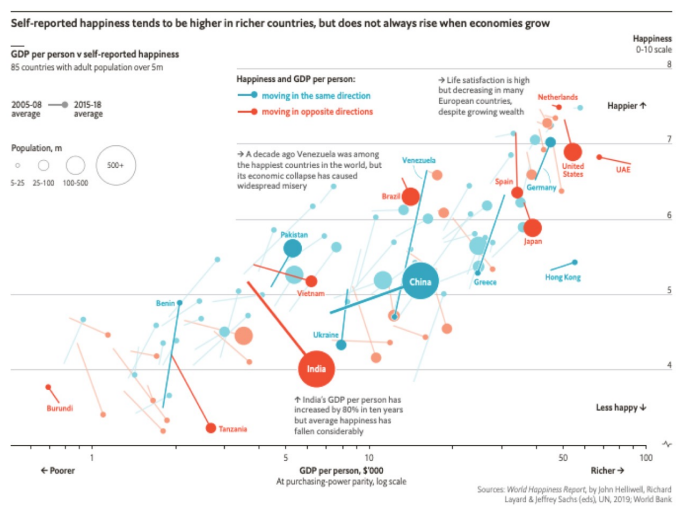
Washington Post

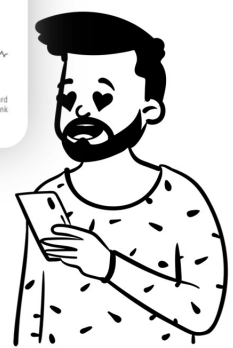


New York Times

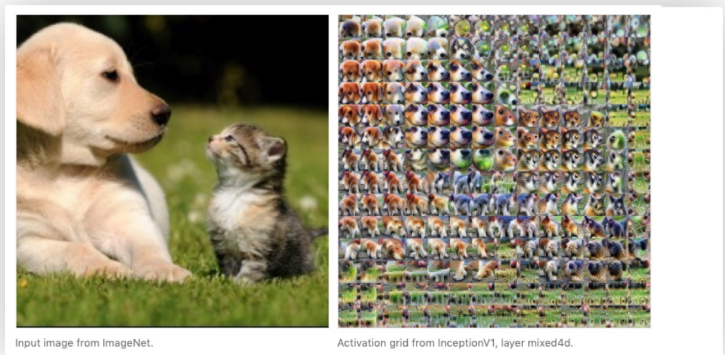


The Economist

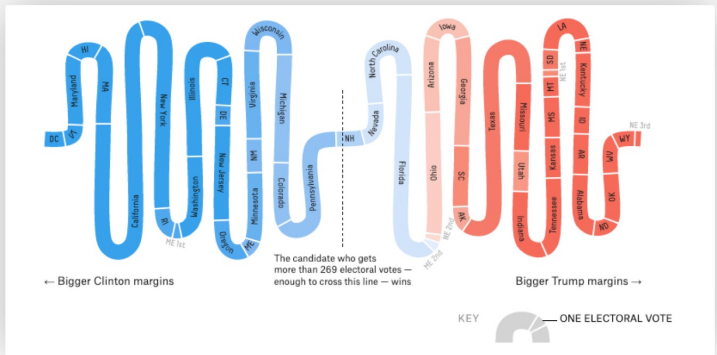




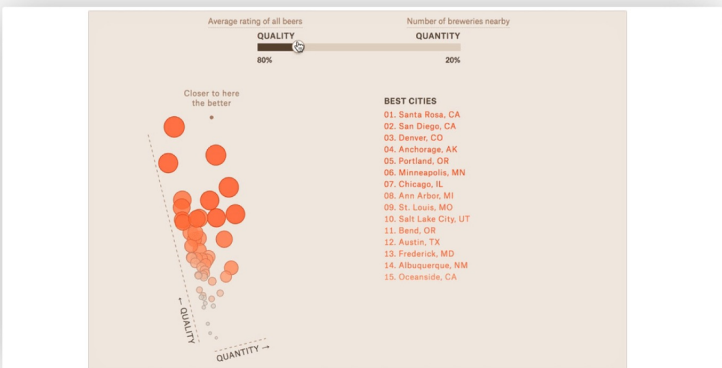
Niche Publications



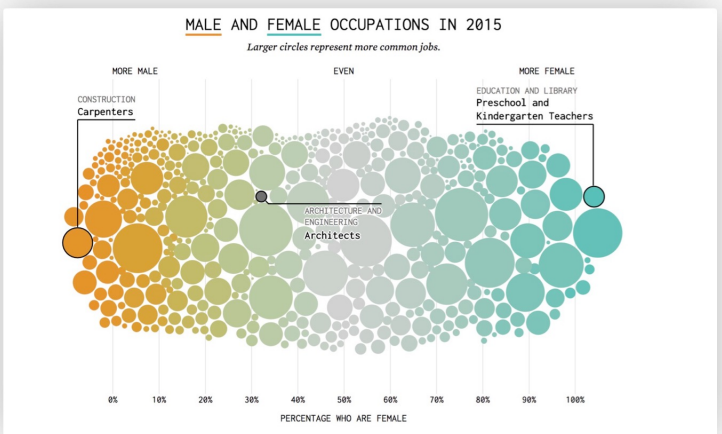
https://distill.pub



https://fivethirtyeight.com



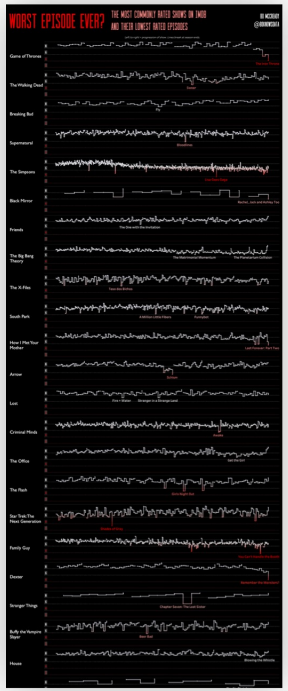
https://pudding.cool



https://flowingdata.com/

Reddit

https://www.reddit.com/r/dataisbeautiful/



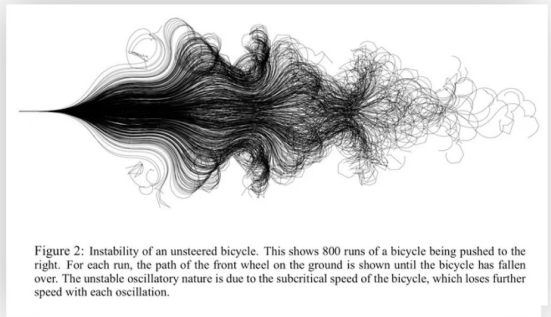
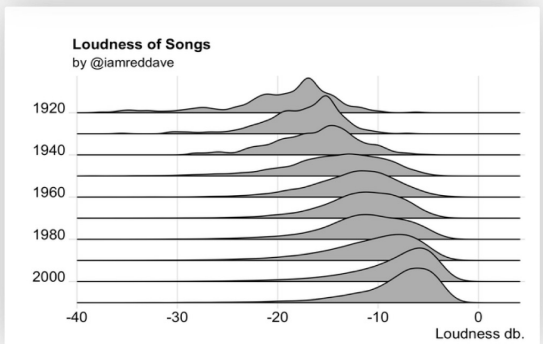
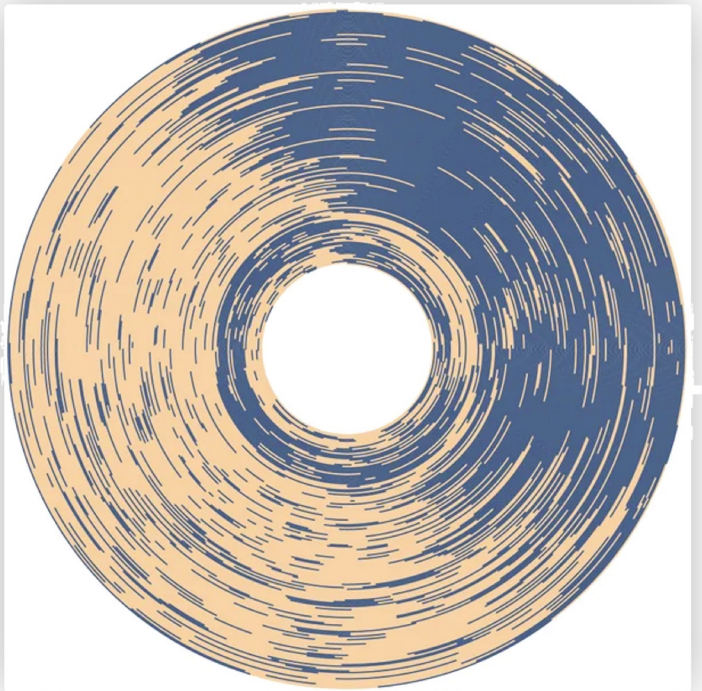


Figure 2: Instability of an unsteered bicycle. This shows 800 runs of a bicycle being pushed to the right. For each run, the path of the front wheel on the ground is shown until the bicycle has fallen over. The unstable oscillatory nature is due to the subcritical speed of the bicycle, which loses further speed with each oscillation.





Data Viz Society

I Learned Data Viz in a Year, and You Can Too

How I went from making simple charts to running workshops, and what I learned along the way

Evelina Judeikytė
Mar 24 · 7 min read ★

How Self-Employed Data Visualization Designers Make a Living

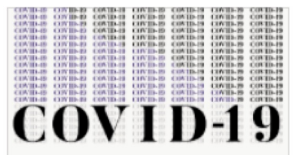
Interviewing four designers about how they started—and made—their careers in data visualization

Jane Zhang
Mar 31 · 19 min read ★

How John Burn-Murdoch's Influential Dataviz Helped the World Understand Coronavirus

An interview with the Financial Times data-journalist about his experience visualizing the COVID-19 pandemic

Jason Forrest
Apr 14 · 16 min read



Ten Considerations Before you Create another Chart about COVID-19

To sum it up—#vizresponsibly; which may mean not publishing your visualizations in the public domain at all.

Amanda Makulec
Mar 11 · 9 min read

The Hungarian Statistician Behind Three Volumes of Visualization Masterpieces

Revisiting the life and work of Lajos Illyefalvi (1881–1944), the great data chronicler of Budapest

Attila Bátorfy
Apr 9 · 7 min read ★

DATA VISUALIZATION SOCIETY

ABOUT PROGRAMS RESOURCES JOBS PUBLICATION CHALLENGES JOIN

The **Data Visualization Society** aims to collect and establish best practices, fostering a community that supports members as they grow and develop data visualization skills. As data grows more central to our personal and professional lives, the practice of data visualization becomes ever more important.

JOIN THE COMMUNITY!

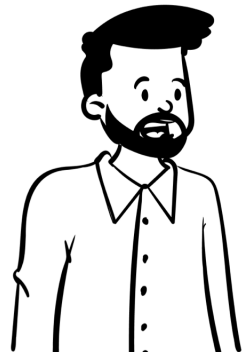
BECOME A MEMBER

The graphic you use to discover the insight may not be the right graphic to communicate the insight

First develop your story, then create effective graphics to tell that story

Communicating is both art and science. Learn from those who you feel do it well!

Thank You!!!

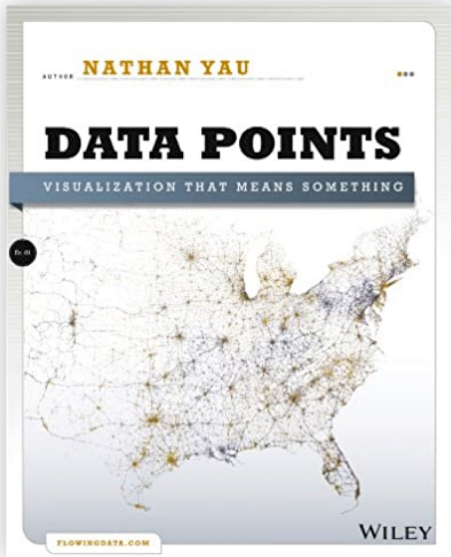


Matt Dzugan

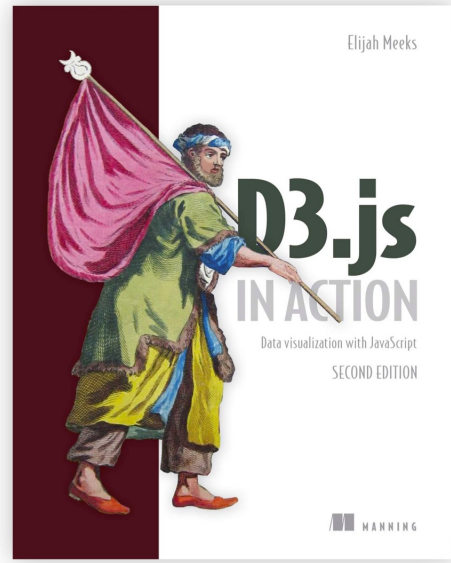
@mattdzugan
mattdzugan.com
mattdzugan

Further Reading

Targeted for a large audience, Cairo's book is meant to inform its readers how to spot BS and how to think critically about data and statistics in the media.



This book is like a much more exhaustive version of this presentation, it talks about a few core ideas and that illustrates them through LOTS of examples.



Essentially a textbook for learning D3, the field's powerhouse library for creating data-viz. This is the library used by all the Pros. If you want to make stuff that you see in the New York Times – this is the book to read.

This is a free web-book that is less about visualization than it is communication, especially around AI and data science. https://linguafranca.standardnotation.ai

References

Communicating Data Visually

https://www.openpeeps.com/
https://www.nytimes.com/interactive/2020/us/coronavirus-us-cases.html

A Mental Framework

https://twitter.com/ClaireESantoro/status/1223247039330443264?s=19
https://infographics.economist.com/2019/AChristmasGiftForYou/AYearInGraphicDetail.pdf
https://projects.fivethirtyeight.com/trump-approval-ratings/
https://www.washingtonpost.com/graphics/politics/2016-election/urban-rural-vote-swing/
https://twitter.com/veltman/status/1201567824168308739
https://twitter.com/Scavino45/status/1245112211829334018
https://towardsdatascience.com/my-first-kaggle-competition-9d56d4773607
https://medium.com/swlh/predicting-the-housing-market-is-easier-than-you-think-45239a366dc1

Design Considerations

https://www.knowablemagazine.org/article/mind/2019/science-data-visualization
https://www.edwardtufte.com/tufte/posters
https://en.wikipedia.org/wiki/Charles_Joseph_Minard
https://en.wikipedia.org/wiki/Edward_Tufte
https://material.io/design/communication/data-visualization.html
https://xkcd.com/1138/
https://fivethirtyeight.com/features/xkcd-randall-munroe-qanda-what-if/

Tools & Software

https://public.tableau.com/en-us/gallery/game-thrones-words?gallery=votd
https://viz.wtf/post/189495004858/the-most-frequent-emoji-the-unicode-consortium
https://projects.fivethirtyeight.com/2016-election-forecast/
https://bl.ocks.org/mbostock/5735770

Learning by Example

https://twitter.com/veltman/status/1162029263739953153
https://twitter.com/flowingdata/status/1052593303554650112
https://twitter.com/NadiehBremer
https://www.washingtonpost.com/graphics/politics/2016-election/election-results-from-coast-to-coast/
https://www.nytimes.com/interactive/2014/06/05/upshot/how-the-recession-reshaped-the-economy-in-255-charts.html
https://infographics.economist.com/2019/AChristmasGiftForYou/AYearInGraphicDetail.pdf
https://distill.pub/2019/activation-atlas/
https://pudding.cool/2017/04/beer/
https://projects.fivethirtyeight.com/2016-election-forecast/
https://flowingdata.com/
https://www.reddit.com/r/dataisbeautiful/comments/fqqzki/worst_episode_ever_the_most_commonly_rated_shows/
https://www.reddit.com/r/dataisbeautiful/comments/88q0d1/songs_have_gotten_louder_over_time_oc/
https://www.reddit.com/r/dataisbeautiful/comments/5l39mu/my_daughters_sleeping_patterns_for_the_first_4/
https://www.reddit.com/r/dataisbeautiful/comments/7s8hck/paths_of_800_unmanned_bicycles_being_pushed_until/
https://www.datavisualizationsociety.com/
https://medium.com/nightingale