VISUALIZING DATA

Veracity, Accuracy, Accountability

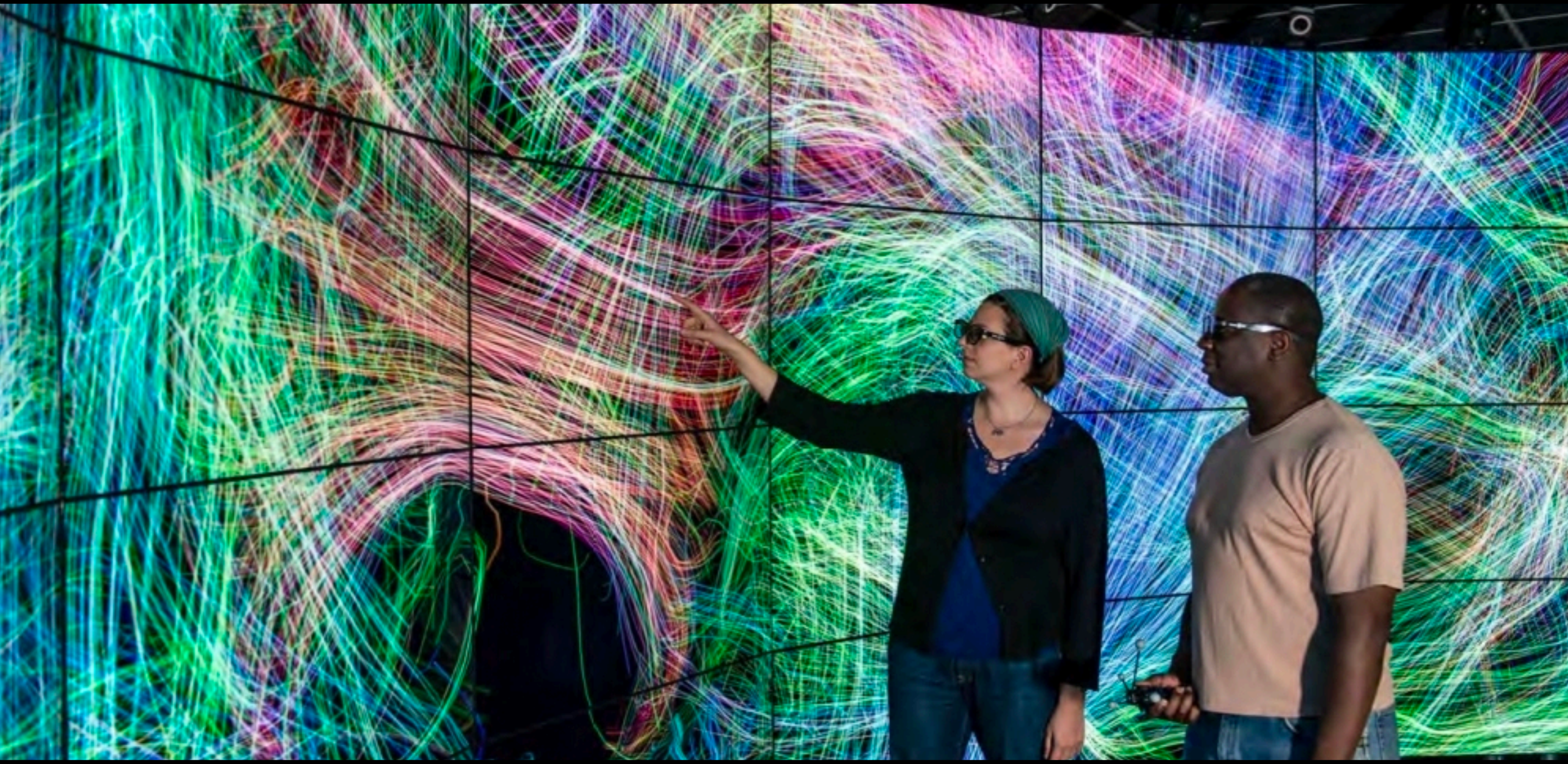
Alberto Cairo

Dialogues in Research Ethics

University of Miami, 2020

We are living through a Golden Age of visualization

We are living through a Golden Age of visualization



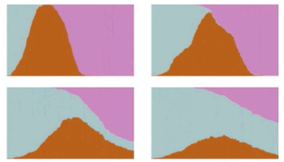
Graphic by the Electronic Visualization Laboratory (EVL) at the University of Illinois at Chicago (UIC)

The most-viewed piece *ever* published by The Washington Post online

Sections

The Washington Post
Democracy Dies in Darkness

Alberto Cairo To...



Health

Why outbreaks like coronavirus spread exponentially, and how to "flatten the curve"

By Harry Stevens March 14, 2020

PLEASE NOTE

The Washington Post is providing this story for free so that all readers have access to this important information about the coronavirus. For more free stories, sign up for our daily Coronavirus Updates newsletter.

2.6k

https://www.washingtonpost.com/graphics/2020/world/corona-simulator/

Common misconceptions when talking about visualization:

1. "A picture is worth a thousand words"
2. "Visualization is intuitive"
3. "The data should speak for itself"
4. "Show, don't tell!"

Visualizations can't be designed based just on our personal preferences—although these *are* important.

Visualization is a bit like writing: beyond some conventions and constraints regarding symbols, visual grammar, perception, and cognition, visualization **can't be based on *"rules"* that are set in stone**.

Instead, when designing visualizations, we need to be guided by **reasoned, justifiable choices**.

"Facts give us **reasons** […] when they count in favor of our having some belief or desire, or acting in some way."

Derek Parfit, *On What Matters*

Reasoning about visualization.
Key questions:

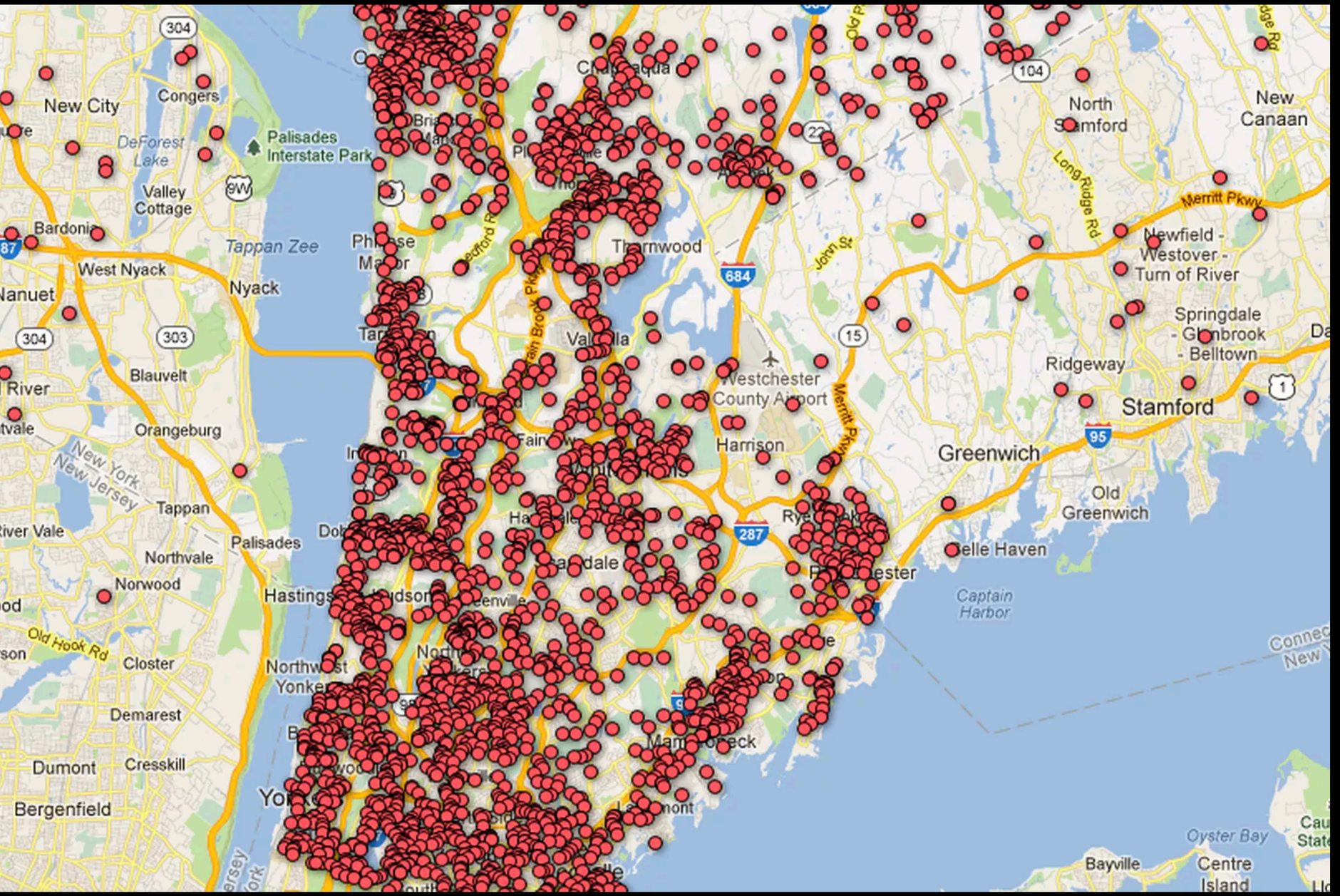
1. Why to visualize?
2. What to visualize?
3. Who to visualize for?
4. How much to visualize?
5. How to visualize it?
6. What style to use?

Note: this is just a tentative list of questions

1. Why should my visualization exist?

Do the potential benefits of designing my visualization outweigh the possible harm it might cause?

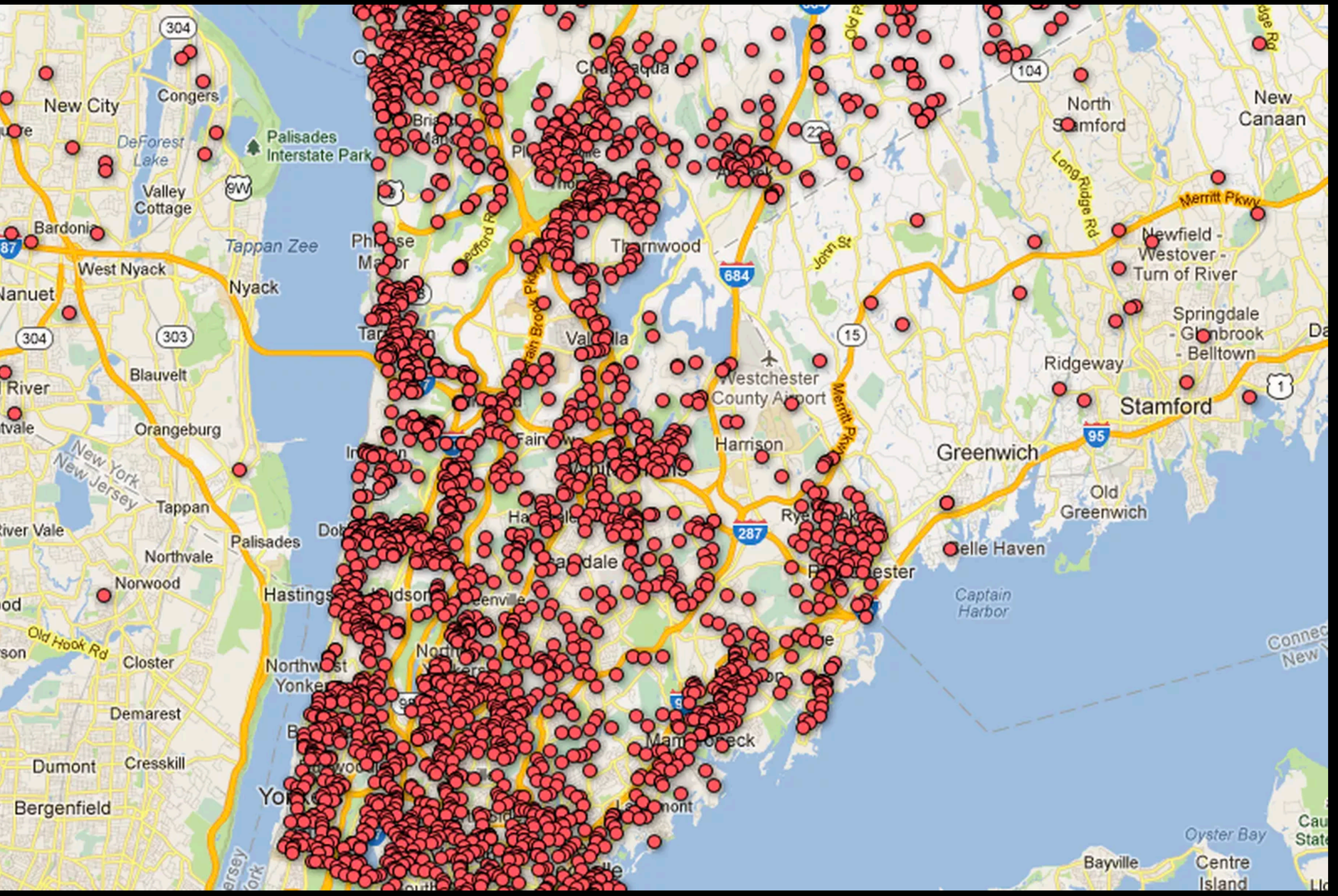
Published Dec. 23, 2012 (the Sandy Hook Elementary School shooting was on Dec. 14)



"Where are the gun permits in your neighborhood?" That's the question posed by **The Journal News**, a New York newspaper that published a Google map on Sunday that shows the names and addresses of pistol or revolver permits in Westchester and Rockland counties."

https://www.theverge.com/2012/12/25/3802960/new-york-newspaper-posts-map-with-names-addresses-of-gun-owners

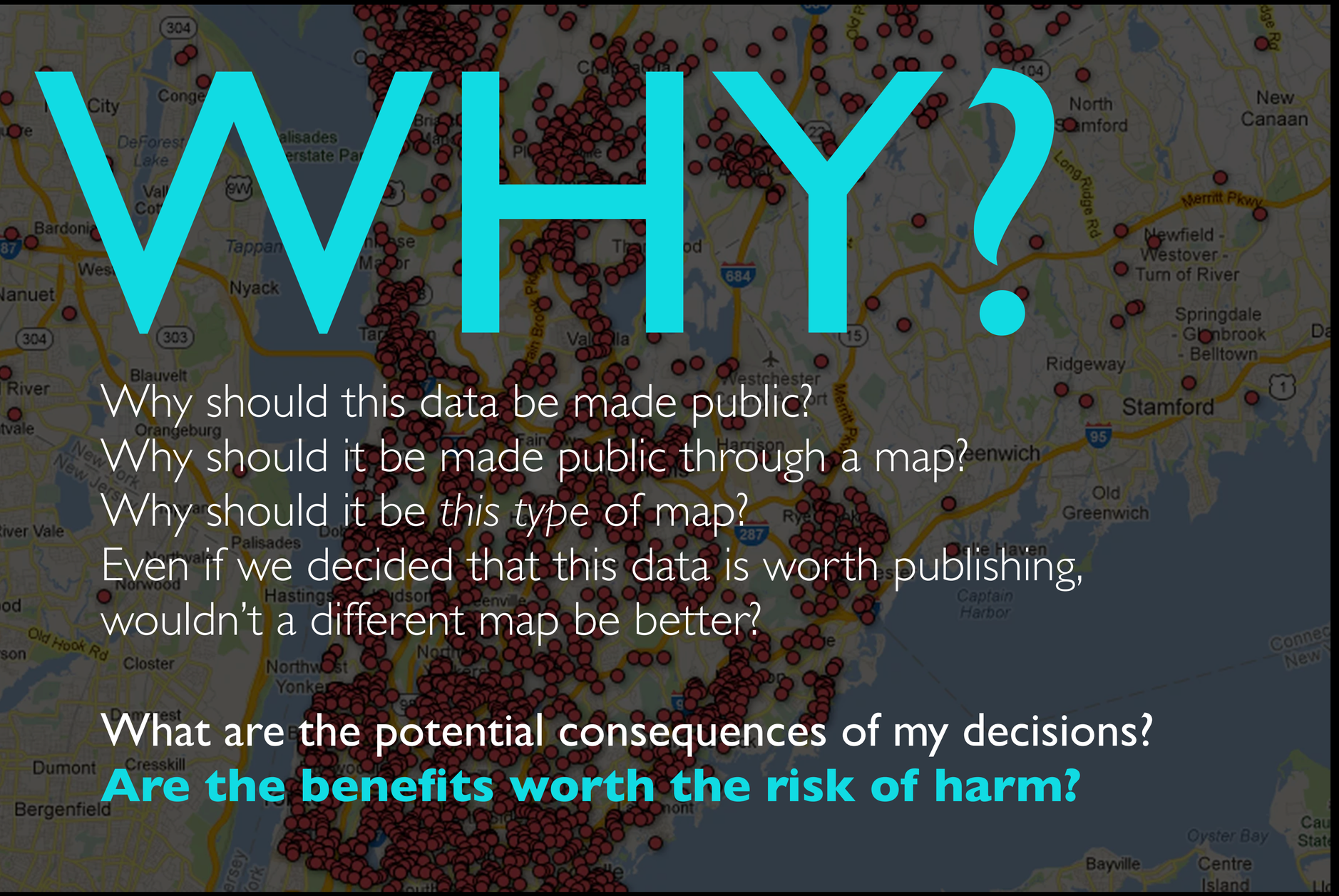
Published Dec. 23, 2012 (the Sandy Hook Elementary School shooting was on Dec. 14)



"We felt sharing information about gun permits in our area was important in the aftermath of the Newtown shootings."

Janet Hasson, president and publisher of the Journal News Media Group

Published Dec. 23, 2012 (the Sandy Hook Elementary School shooting was on Dec. 14)



2. What to visualize?

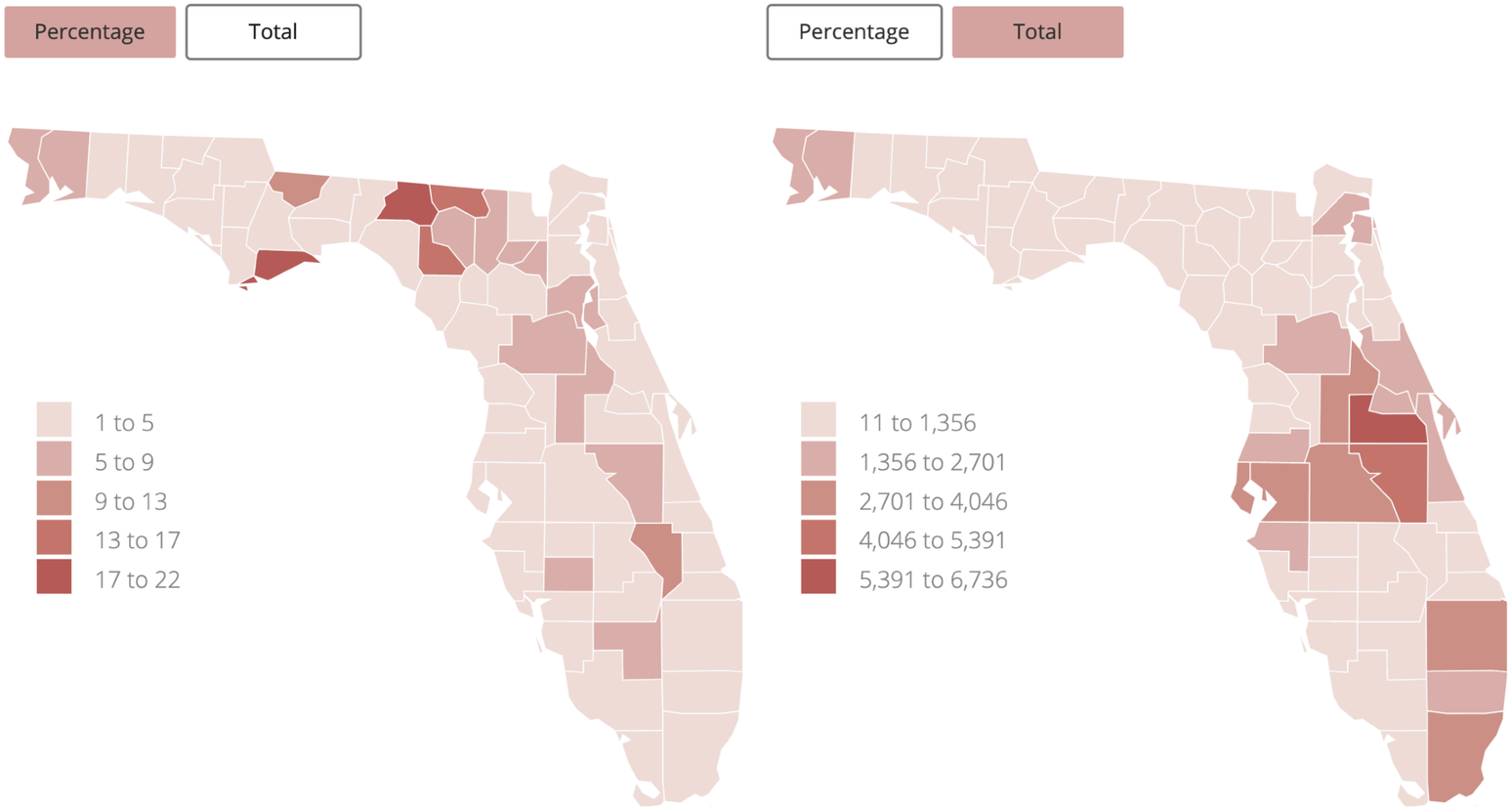
Do I understand my data, its limitations, uncertainty, or glitches?

What or who is being measured (*o not being measured*,) and why?

AT SCHOOL WITHOUT A ROOF

http://www.lmelgar.me/without-a-roof/

In Florida more than 71,000 students are homeless. During the last decade, this population rocketed as a result of the recession and how hard it has become for the poorest families to find affordable housing.



Kansas is the nation's porn capital, according to Pornhub

Blue states watch more porn. But what's the matter with Kansas?

According to Pornhub Insights, Kansas leads the nation in porn pageviews per capita at roughly 194. They don't specify what interval this is over (monthly, weekly, etc), but the state-by-state comparison is nonetheless interesting.

Plotting Obama vote share in 2012 versus porn consumption, it looks like blue states consume more porn per capita than red ones. Aside from Kansas - a clear outlier - and Georgia, the remaining top ten per-capita porn consumers are all blue. Similarly, New Mexico and Maine are the only blue states in the bottom ten per-capita porn consumers.

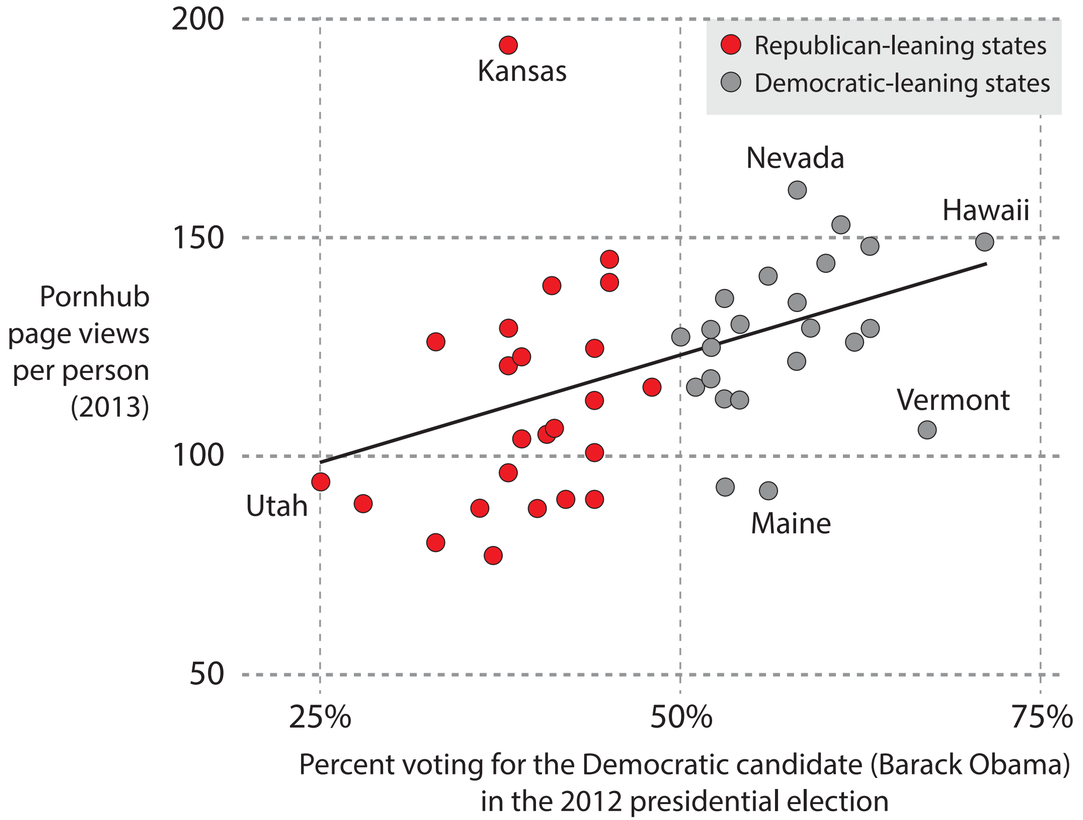
https://wonkviz.tumblr.com/post/82488570278/kansas-is-the-nations-porn-capital-according-to

Kansas is the nation's porn capital, according to Pornhub

Blue states watch more porn. But what's the matter with Kansas?

According to Pornhub Insights, Kansas leads the nation in porn pageviews per capita at roughly 194. They don't specify what interval this is over (monthly, weekly, etc), but the state-by-state comparison is nonetheless interesting.

Plotting Obama vote share in 2012 versus porn consumption, it looks like blue states consume more porn per capita than red ones. Aside from Kansas - a clear outlier - and Georgia, the remaining top ten per-capita porn consumers are all blue. Similarly, New Mexico and Maine are the only blue states in the bottom ten per-capita porn consumers.



https://wonkviz.tumblr.com/post/82488570278/kansas-is-the-nations-porn-capital-according-to

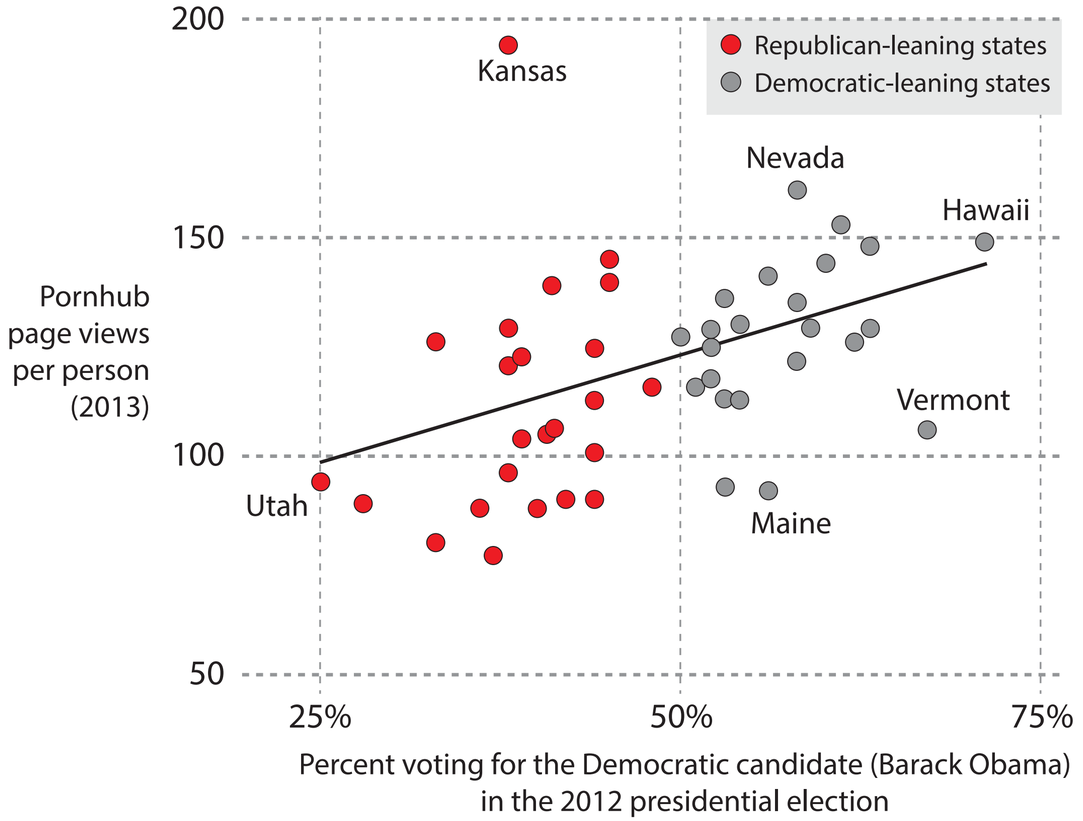
Kansas is the nation's porn capital, according to Pornhub

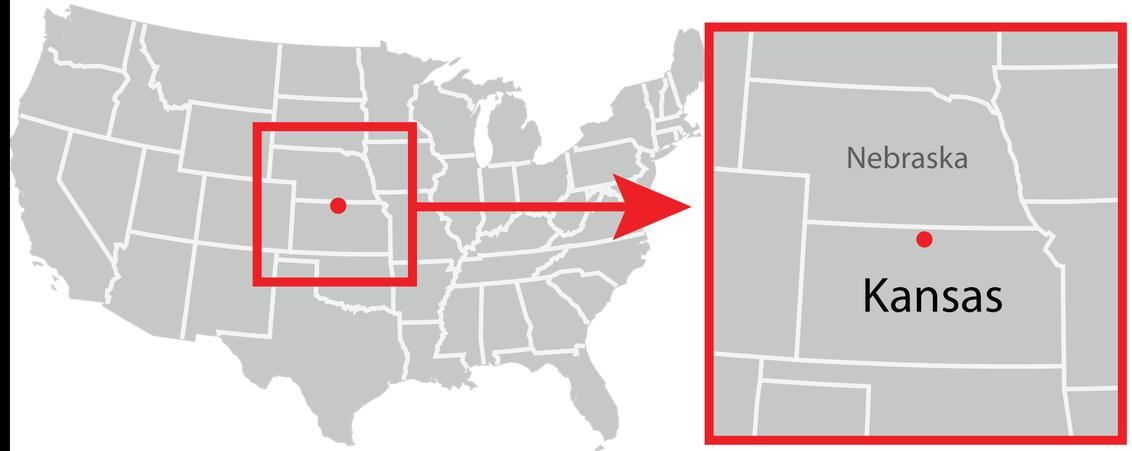
Blue states watch more porn. But what's the matter with Kansas?

According to Pornhub Insights, Kansas leads the nation in porn pageviews per capita at roughly 194. They don't specify what interval this is over (monthly, weekly, etc), but the state-by-state comparison is nonetheless interesting.

Plotting Obama vote share in 2012 versus porn consumption, it looks like blue states consume more porn per capita than red ones. Aside from Kansas - a clear outlier - and Georgia, the remaining top ten per-capita porn consumers are all blue. Similarly, New Mexico and Maine are the only blue states in the bottom ten per-capita porn consumers.

https://wonkviz.tumblr.com/post/82488570278/kansas-is-the-nations-porn-capital-according-to





● Geographic center of the contiguous United States

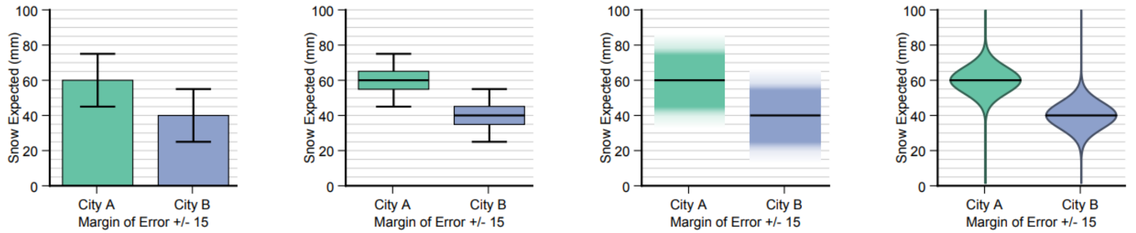
Disclosing limitations and uncertainty

Uncertainty and graphicacy
How should statisticians, journalists, and designers reveal uncertainty in graphics for public consumption?

https://ec.europa.eu/eurostat/cros/powerfromstatistics/OR/PfS-OutlookReport-Cairo.pdf

Error Bars Considered Harmful:
Exploring Alternate Encodings for Mean and Error

Michael Correll *Student Member, IEEE*, and Michael Gleicher *Member, IEEE*



(a) **Bar chart** with error bars: the height of the bars encodes the sample mean, and the whiskers encode a 95% t-confidence interval.

(b) **Modified box plot**: The whiskers are the 95% t-confidence interval, the box is a 50% t-confidence interval.

(c) **Gradient plot**: the transparency of the colored region corresponds to the cumulative density function of a t-distribution.

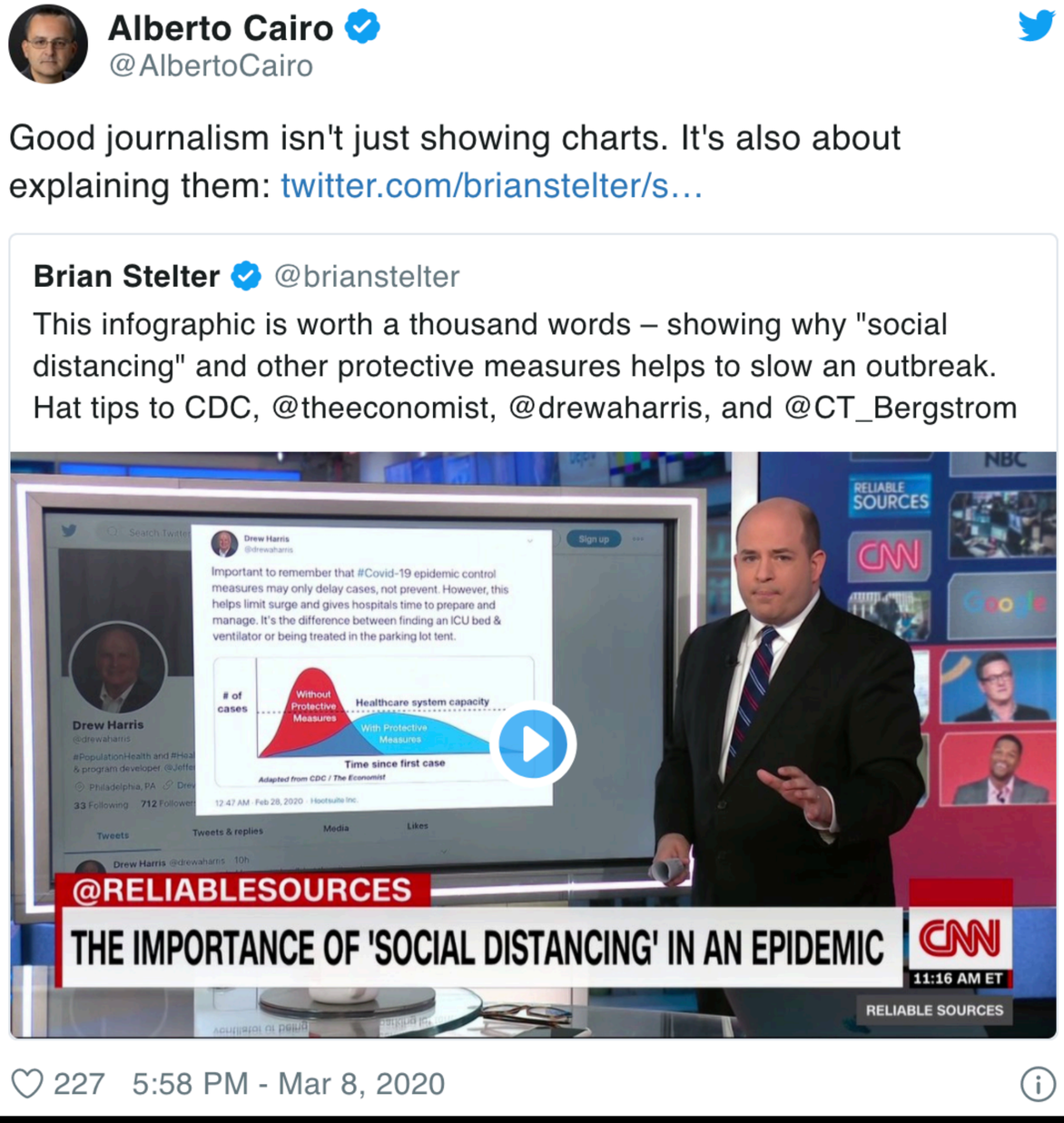
(d) **Violin plot**: the width of the colored region corresponds to the probability density function of a t-distribution.

https://graphics.cs.wisc.edu/Papers/2014/CG14/Preprint.pdf

Collection of papers about visualizing uncertainty:
https://www.dropbox.com/sh/jk4ginxyai6ylqu/AABvqdyT1hJtyFN9nKNHyX9Ba?dl=0

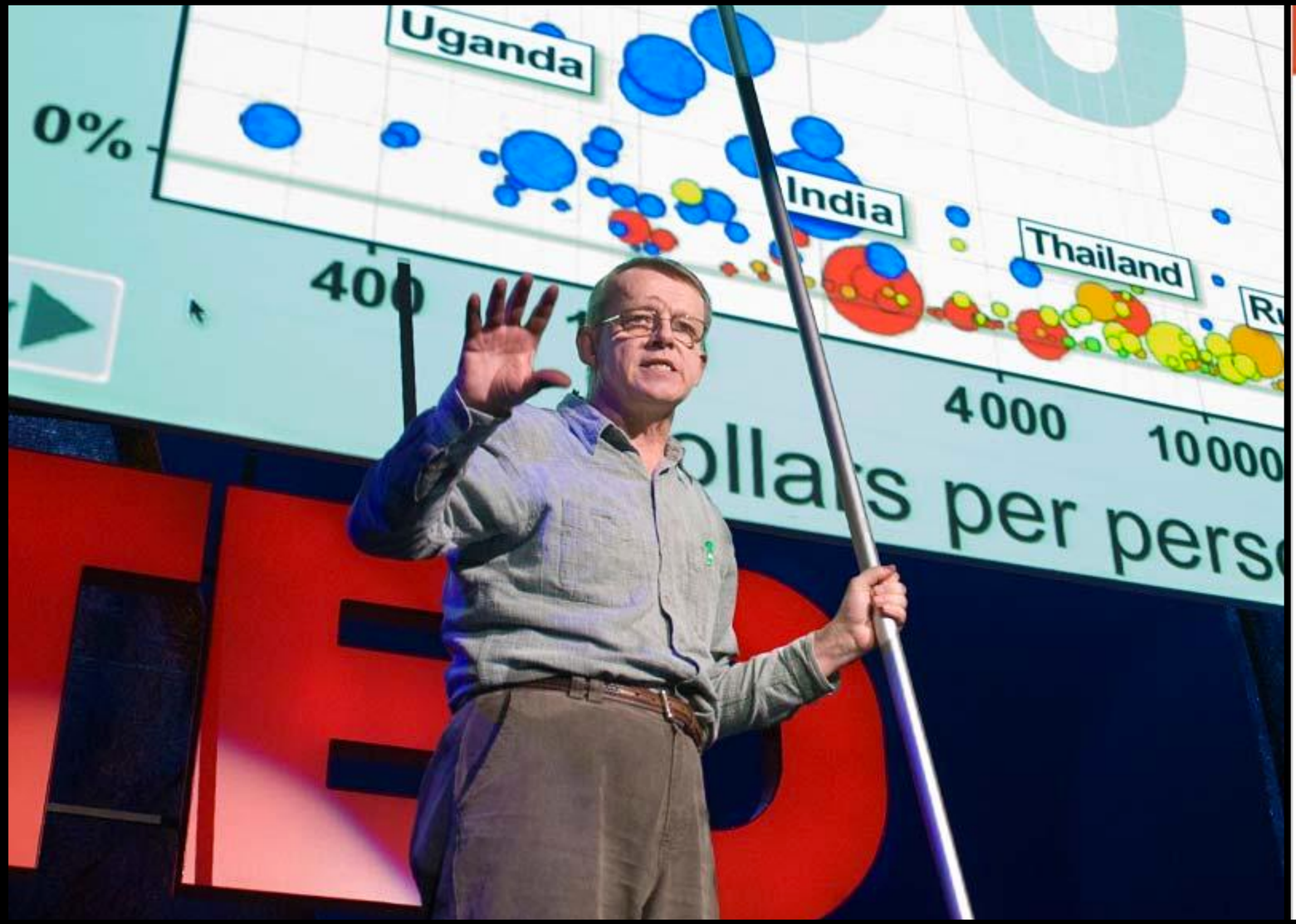
3. Who to visualize for?

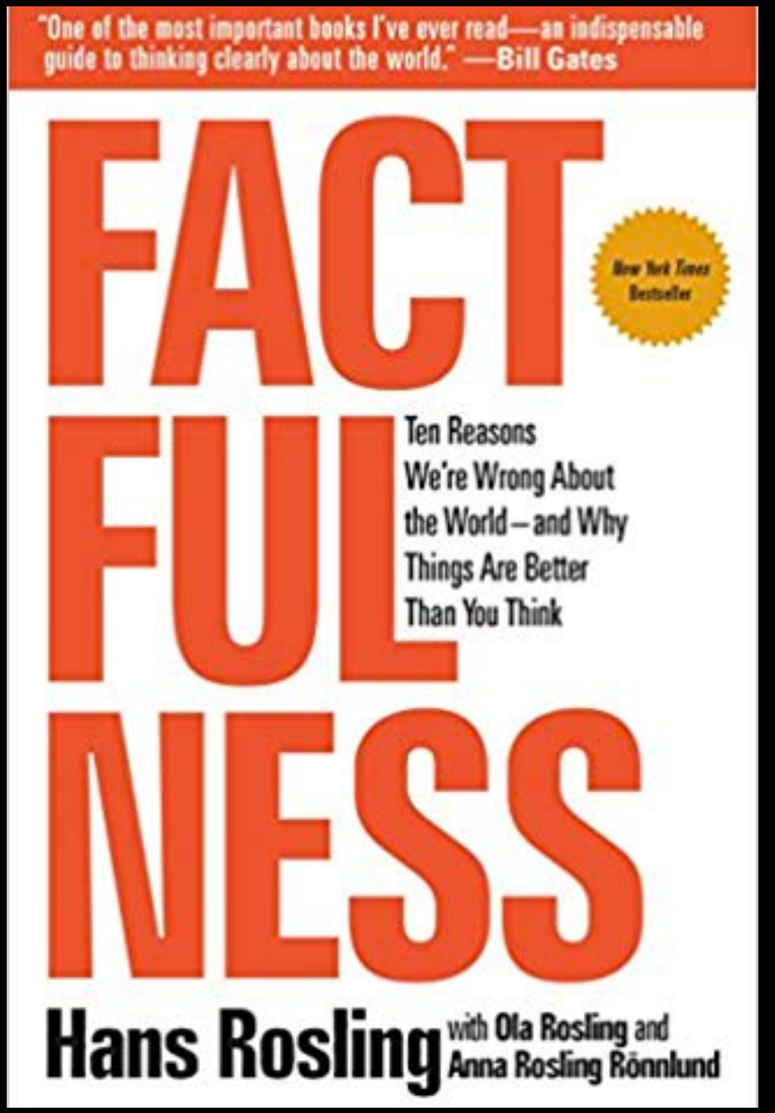
Have I thought about how my intended audience
will access my graphic? Will they understand it? Can I explain it better?



https://twitter.com/AlbertoCairo/status/1236773377865658370

Show AND tell



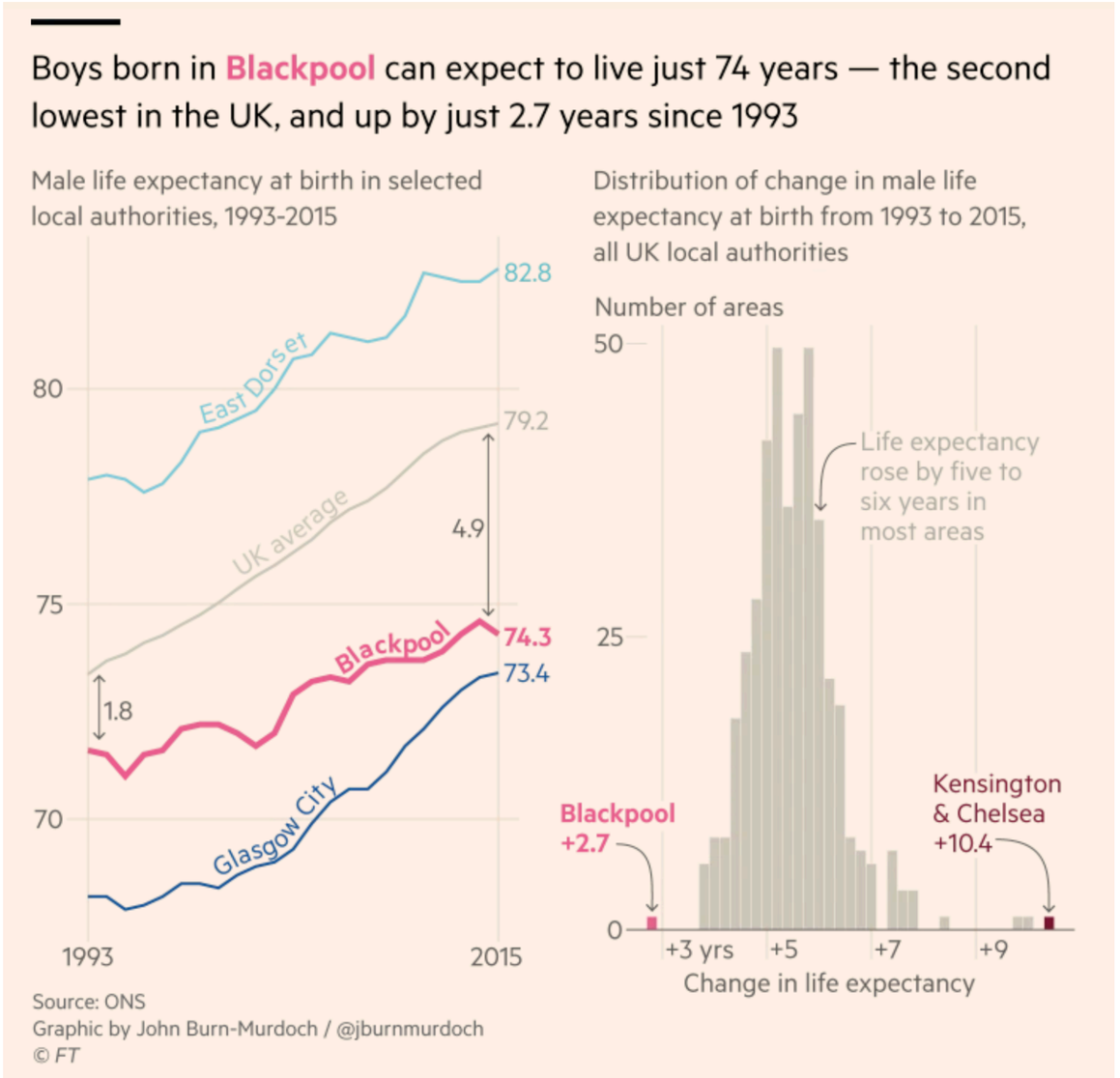


Hans Rosling, www.gapminder.org

Show AND tell

Hans Rosling, *The Joy of Stats*

Alberto Cairo • University of Miami • www.thefunctionalart.com • Twitter: @albertocairo



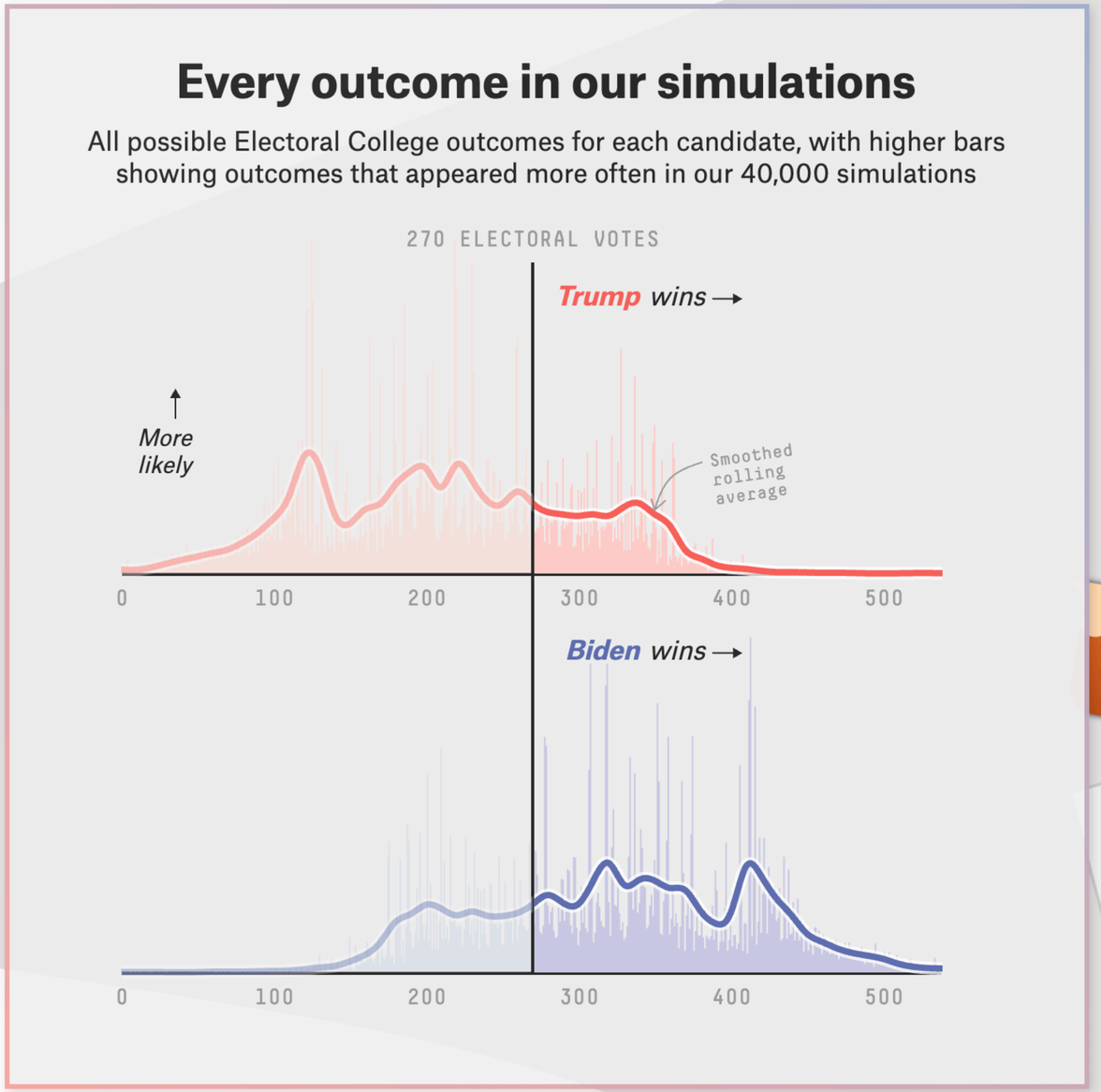
"Design secrets behind the FT's best charts of the year"
https://www.ft.com/content/4743ce96-e4bf-11e7-97e2-916d4fbac0da

"I and my colleagues here at the FT, we really do think one of the most valuable things we can do as data visualization practitioners is add this expert annotation layer."

John Burn-Murdoch

Financial Times

https://policyviz.com/podcast/episode-155-john-burn-murdoch/



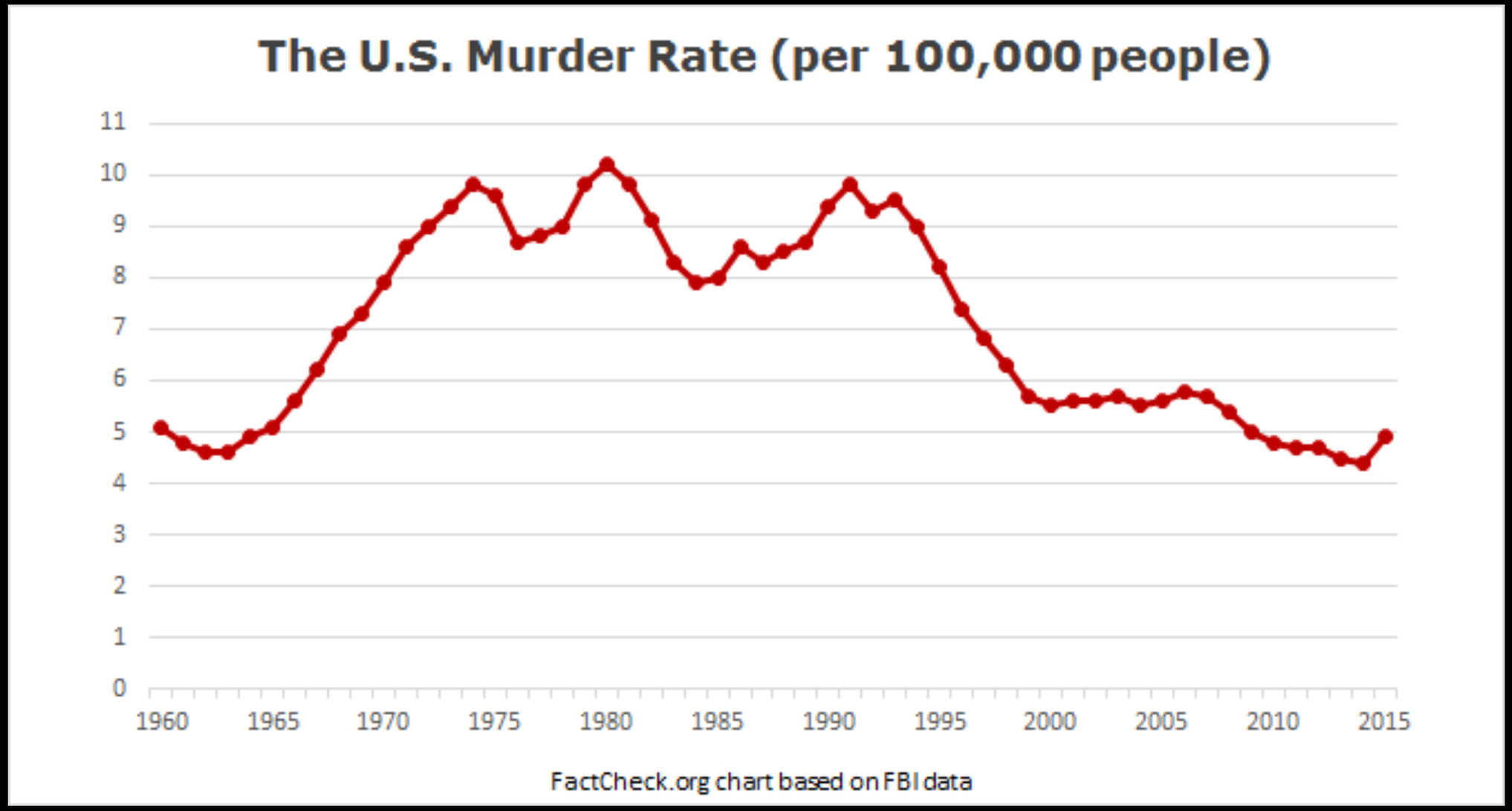
More bars to the right of the 270 line means more simulations where that candidate wins. Some of the bars represent really weird outcomes, but you never know!

https://projects.fivethirtyeight.com/2020-election-forecast/

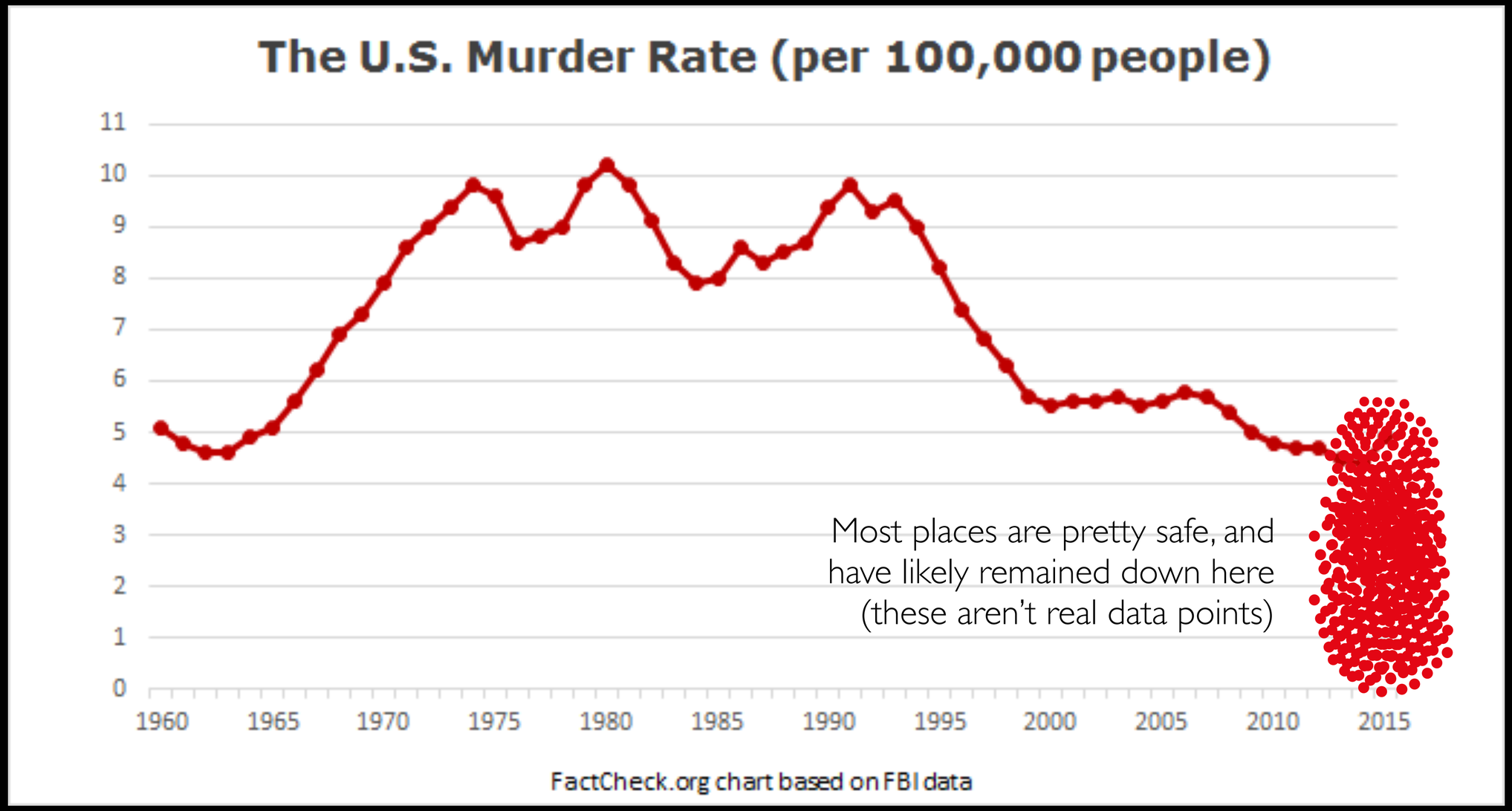
4. How much to visualize?

Am I showing too little?

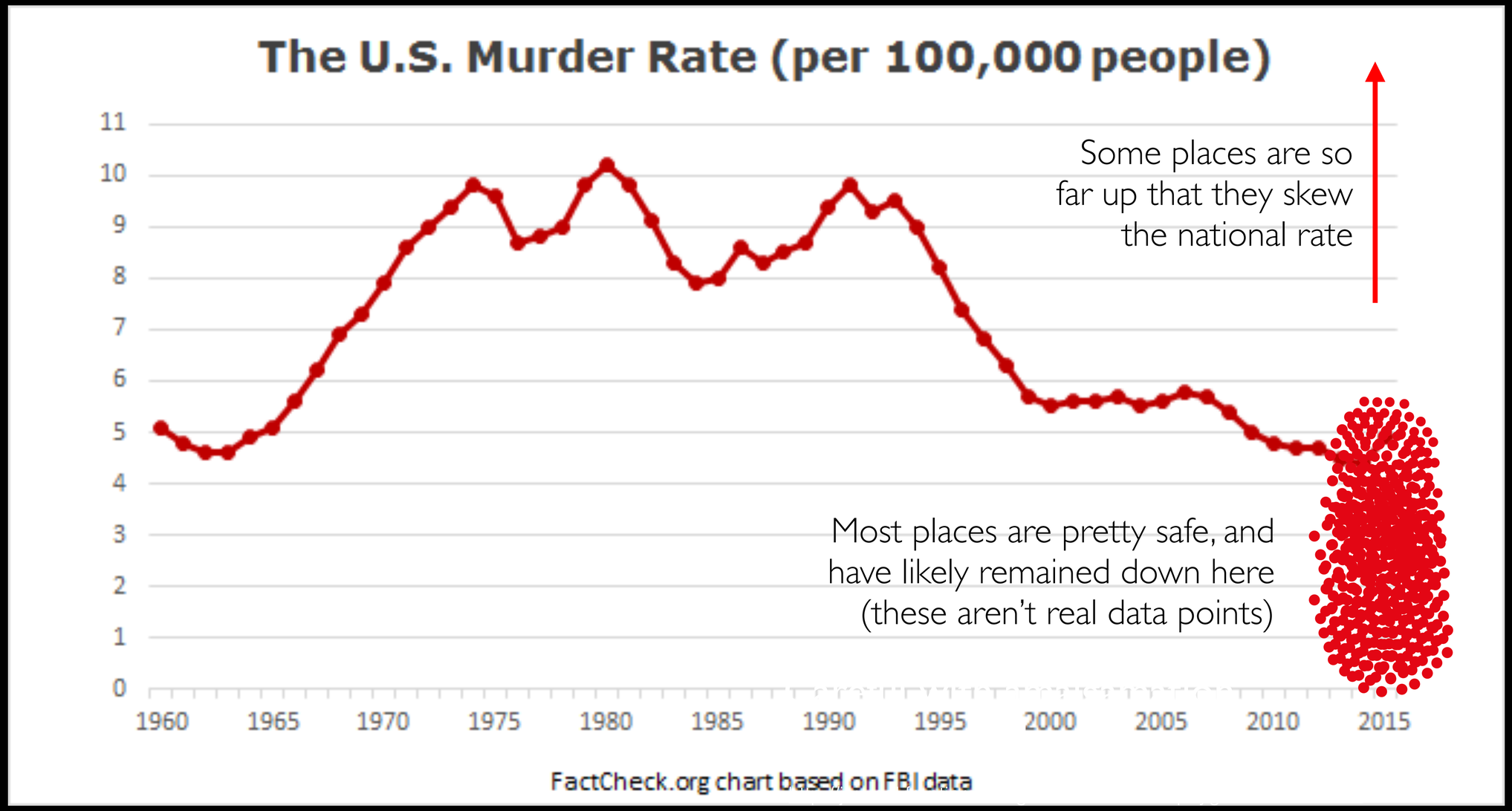
Am I showing too much?


The U.S. Murder Rate (per 100,000 people)
11
10
9
8
7
6
5
4
3
2
1
0
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
FactCheck.org chart based on FBI data

The danger of aggregating data too much, and presenting just averages and other statistical summaries



The danger of aggregating data too much, and presenting just averages and other statistical summaries



5. How to visualize it?

What types of charts or maps should I use?

What is the best way to organize the visualization?

Figure 2 - Main nationalities of arriving migrants – 2016

Greece

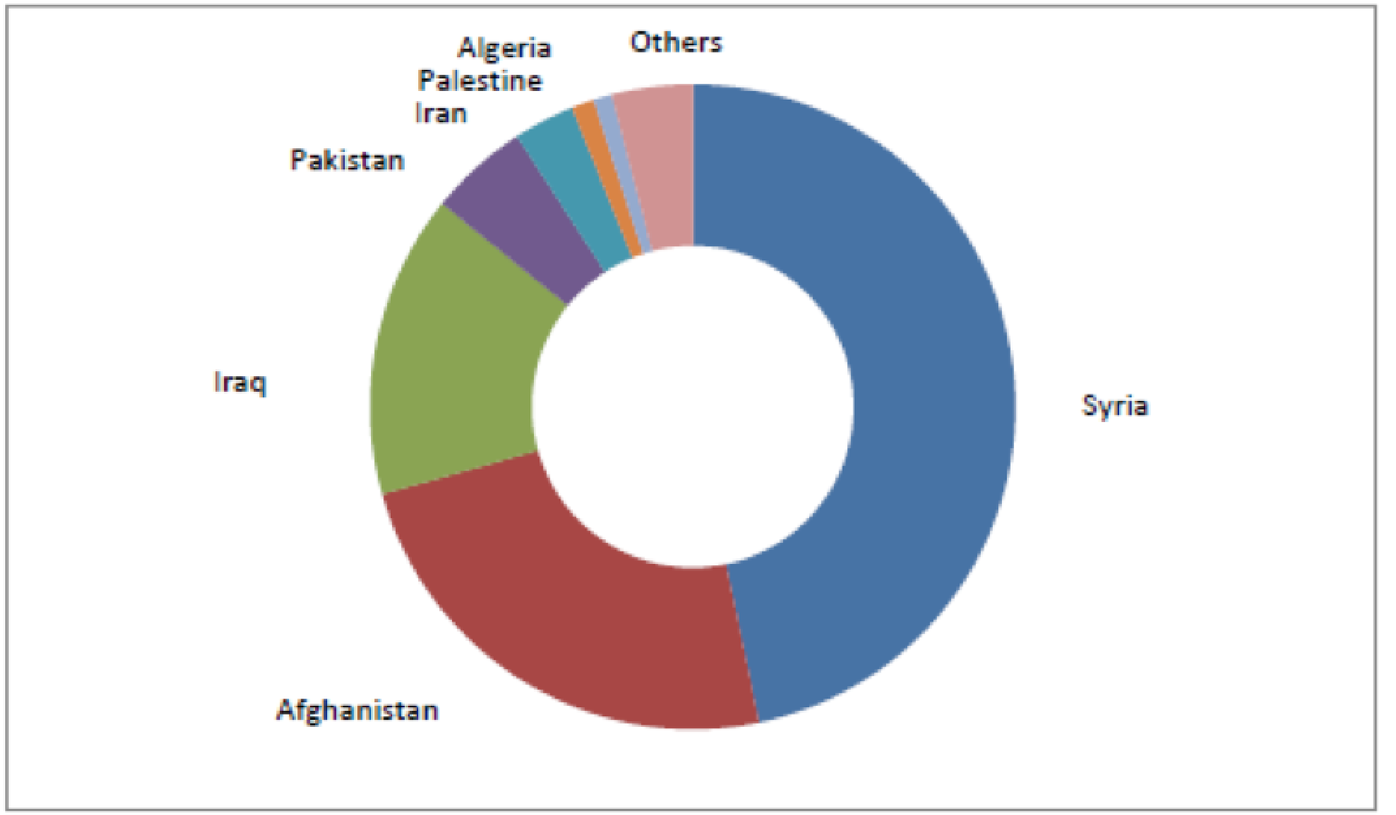
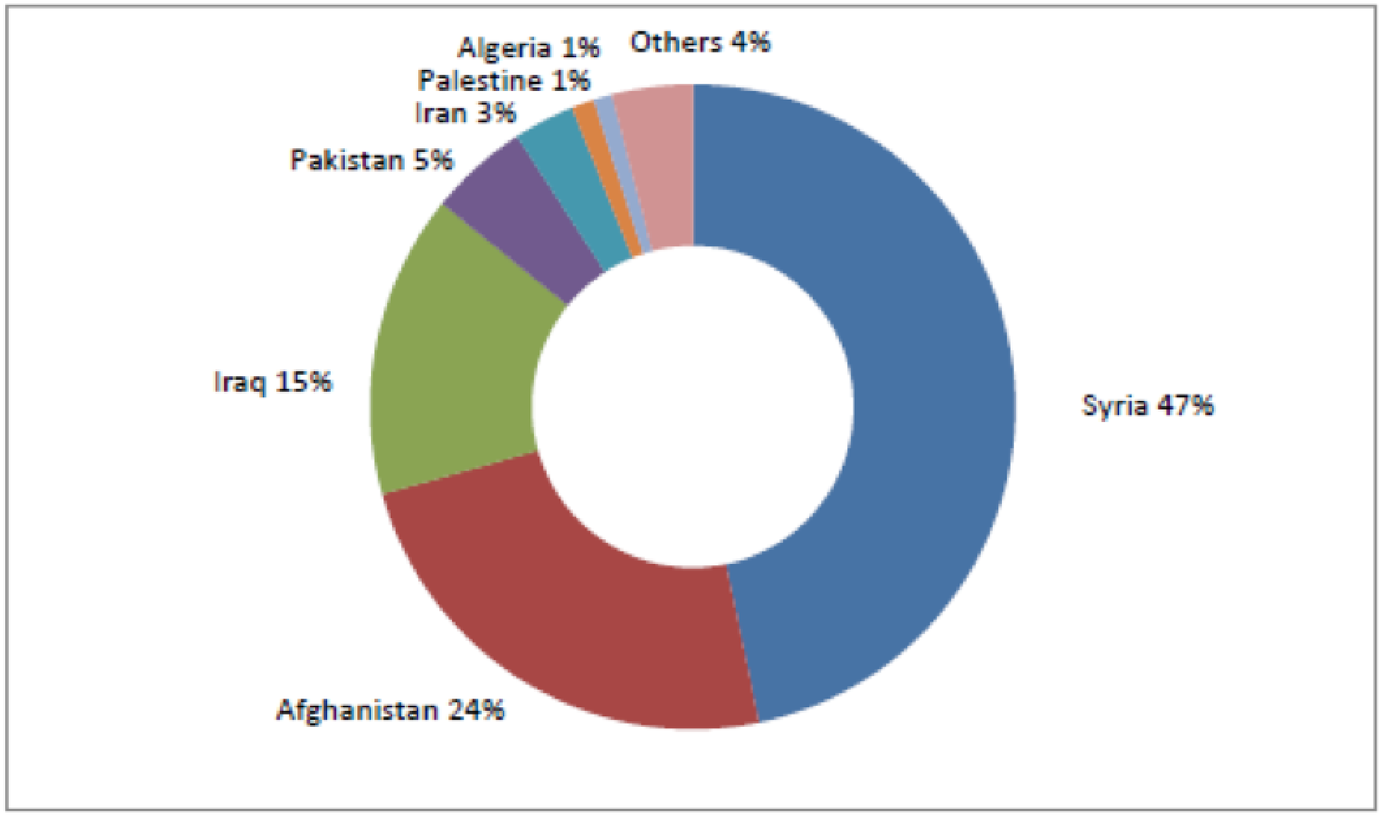
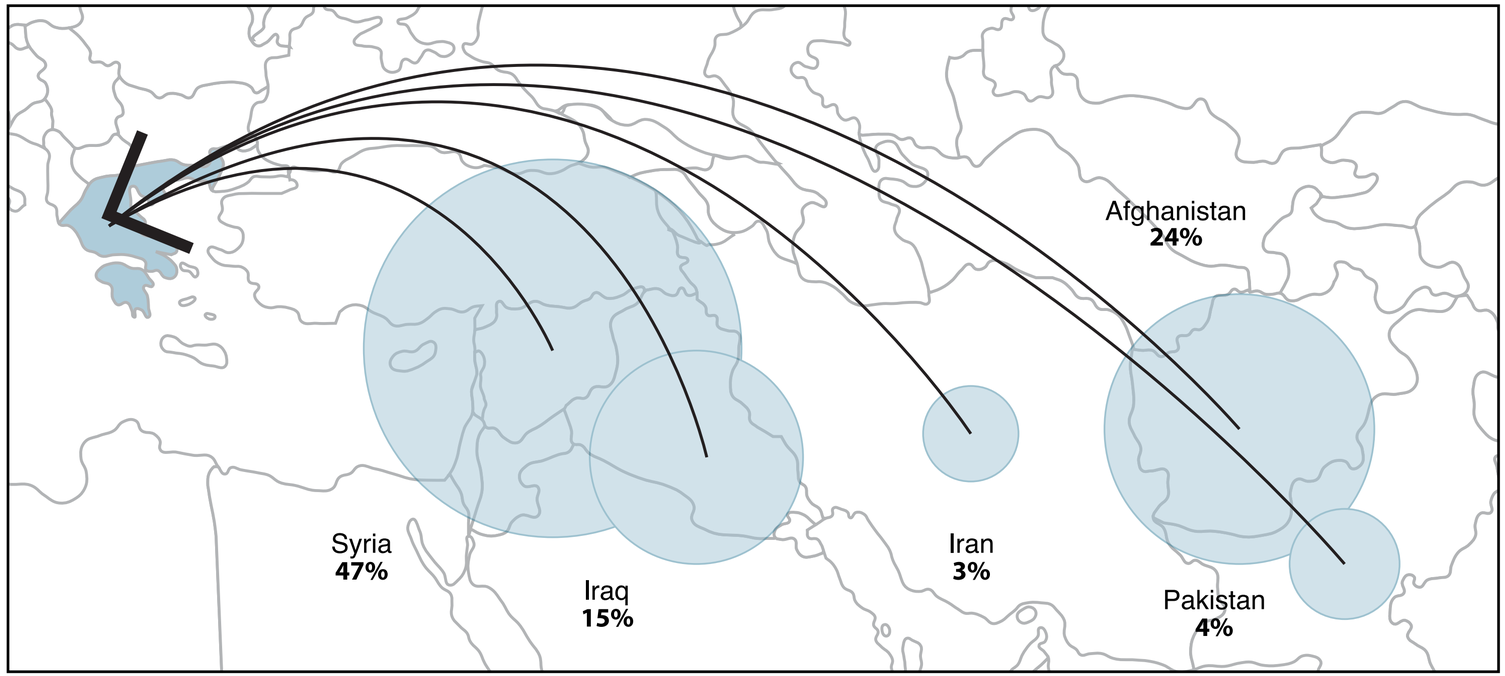
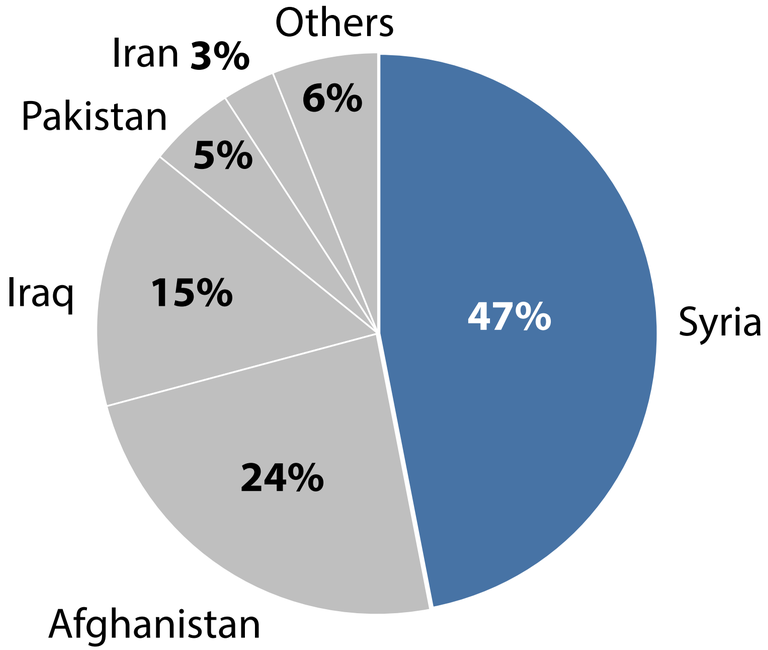


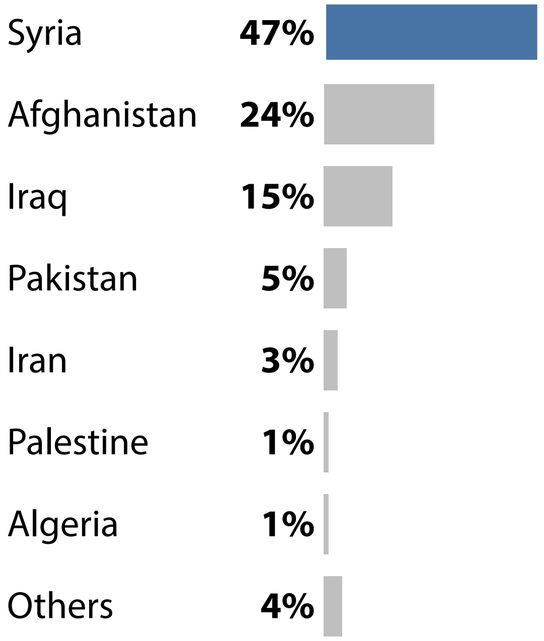
Figure 2 - Main nationalities of arriving migrants – 2016

Greece

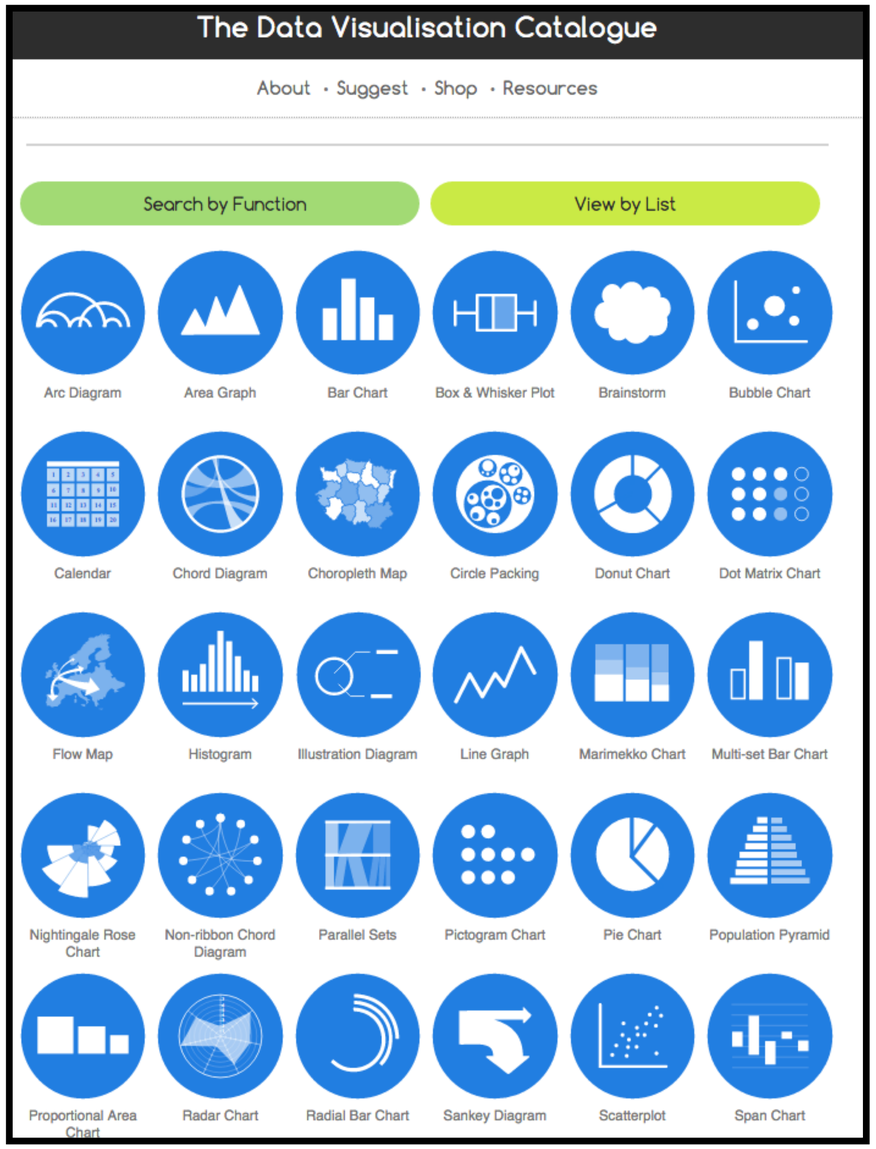


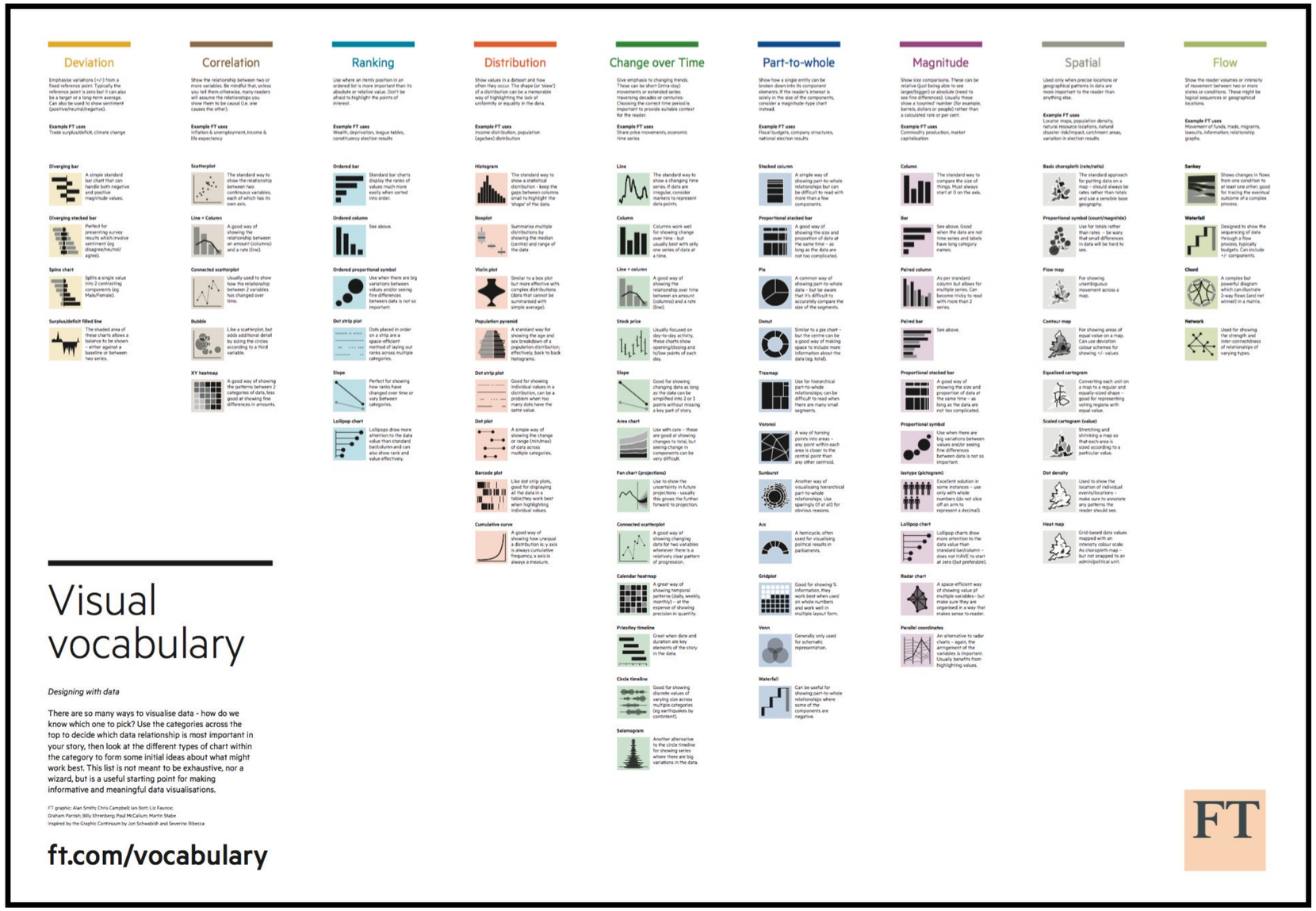






Alberto Cairo • University of Miami • www.thefunctionalart.com • Twitter: @albertocairo

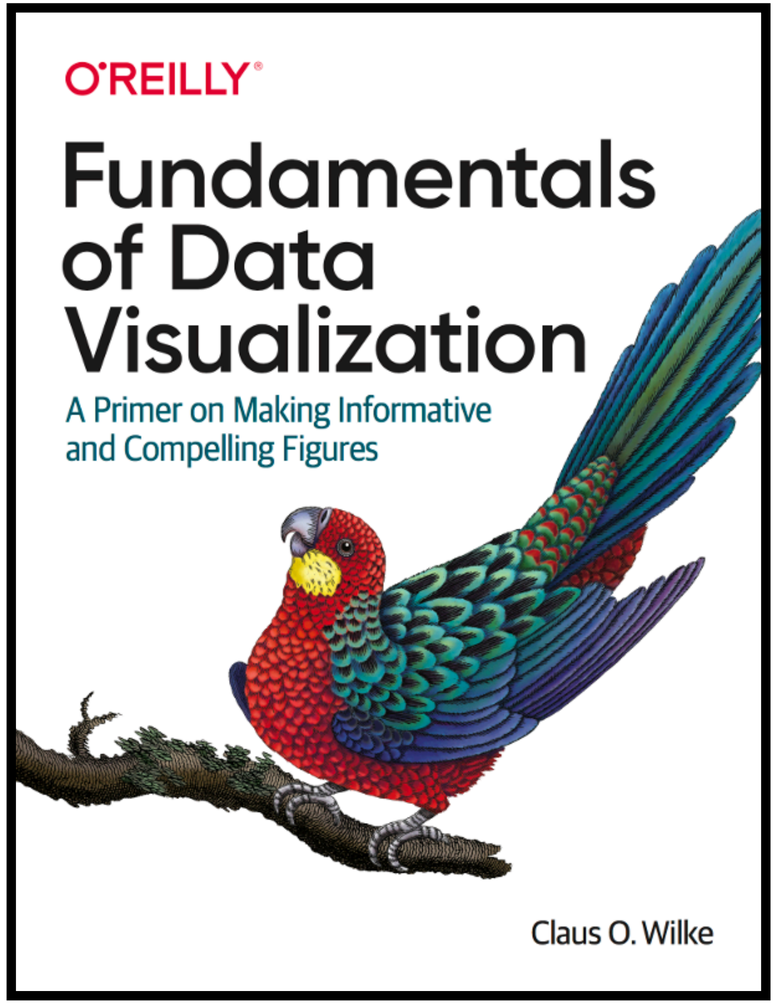


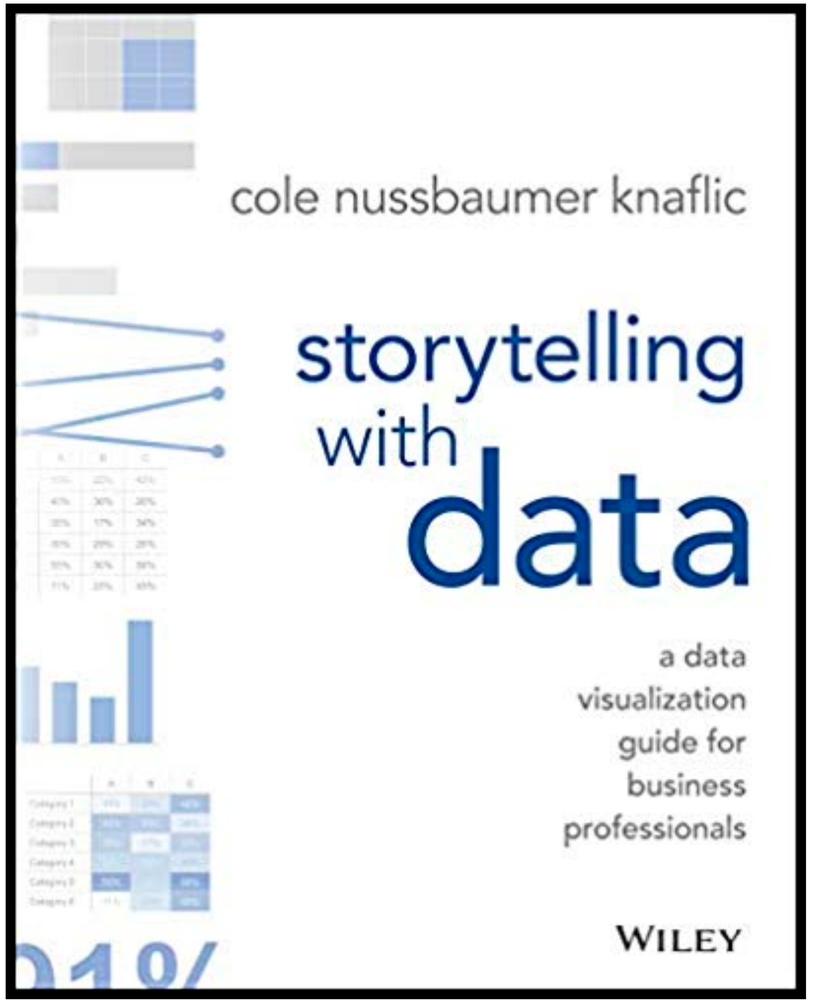


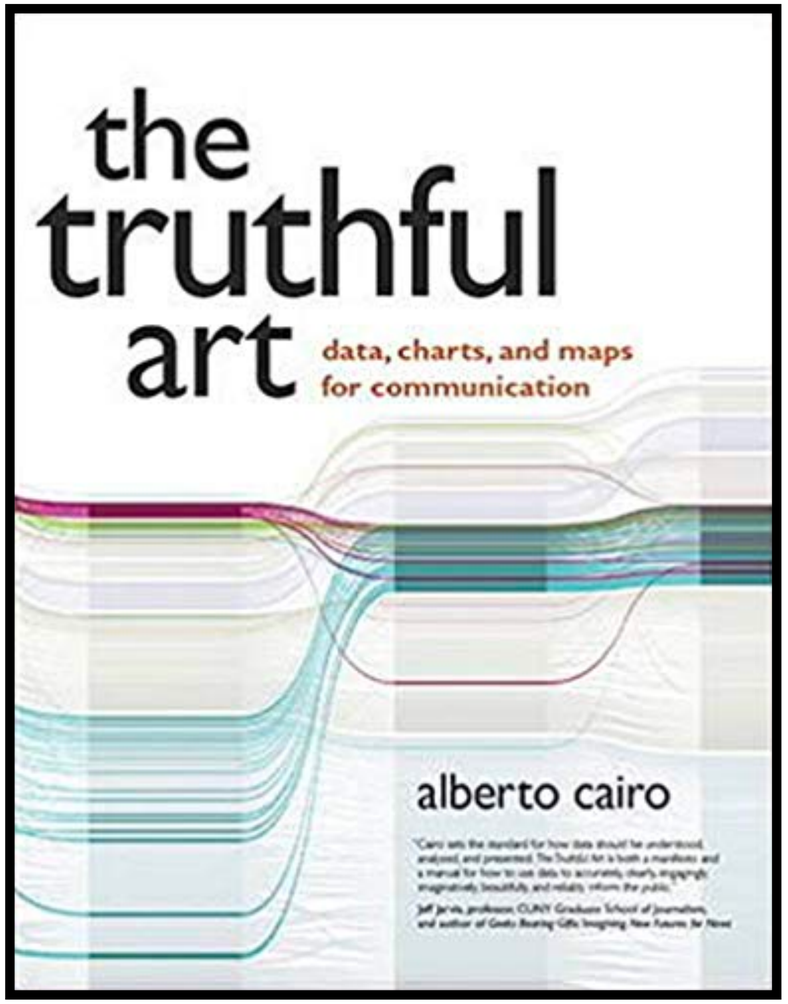
http://www.datavizcatalogue.com/

https://github.com/ft-interactive/chart-doctor/blob/master/visual-vocabulary/Visual-vocabulary.pdf

Books to make design choices







Draft available online:
https://serialmentor.com/dataviz/

6. What visual style to use?

Not all visualizations need to be minimalist.

Not all visualizations need to be flashy and innovative, either.

Standard visualizations

Appropriate for graphics we use all the time

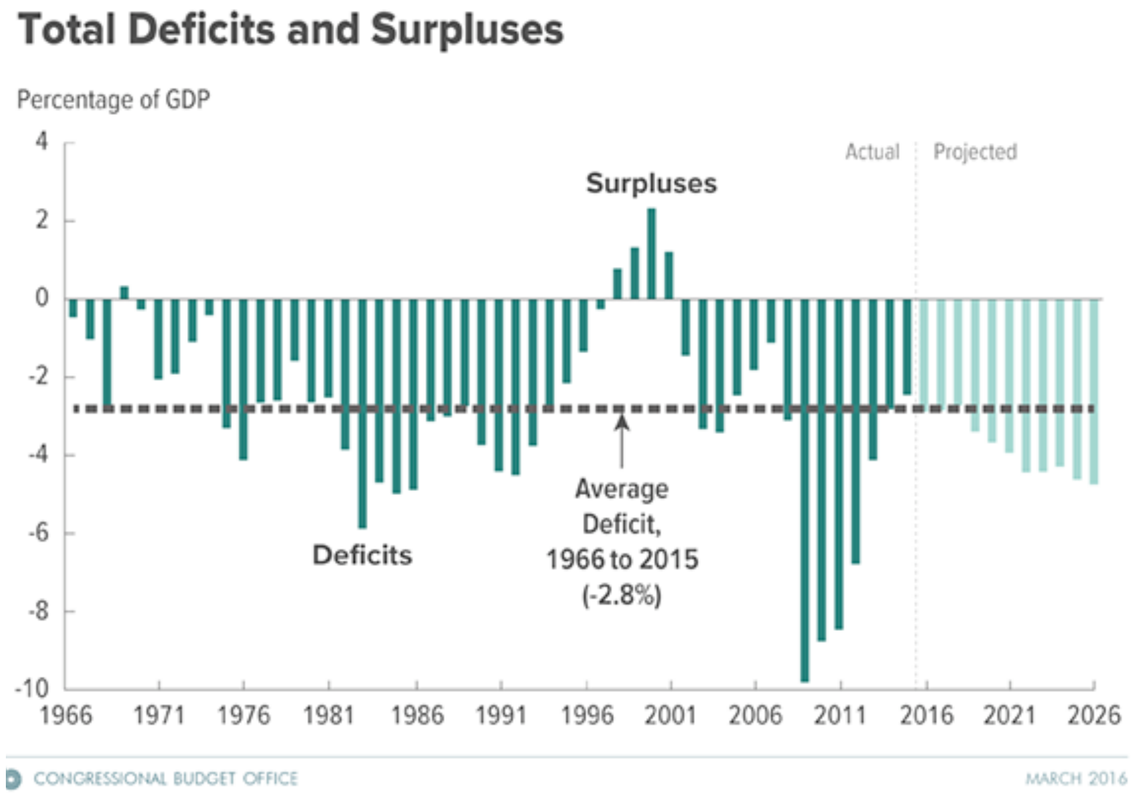
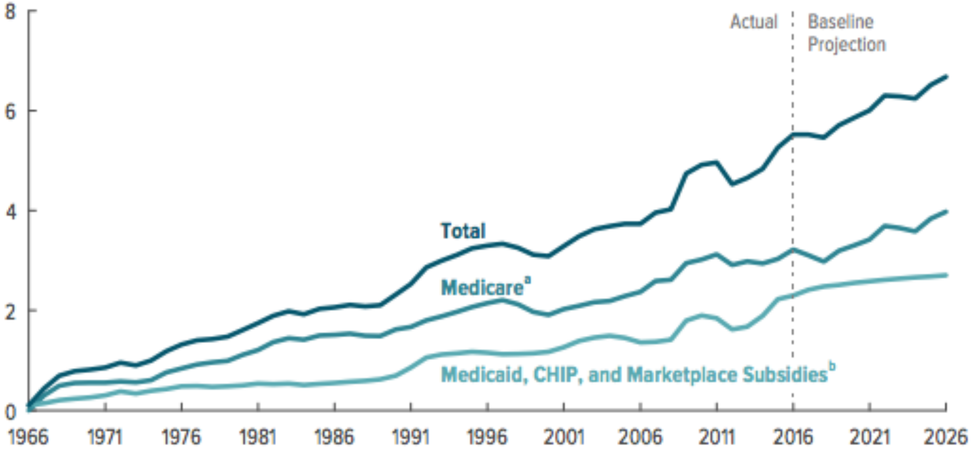


Figure 5-1.

Federal Spending on the Major Health Care Programs, by Category

Percentage of Gross Domestic Product



Over the next decade, federal spending on health care is projected to take up an increasing share of gross domestic product.

Source: Congressional Budget Office (as of August 2016).

CHIP = Children's Health Insurance Program.

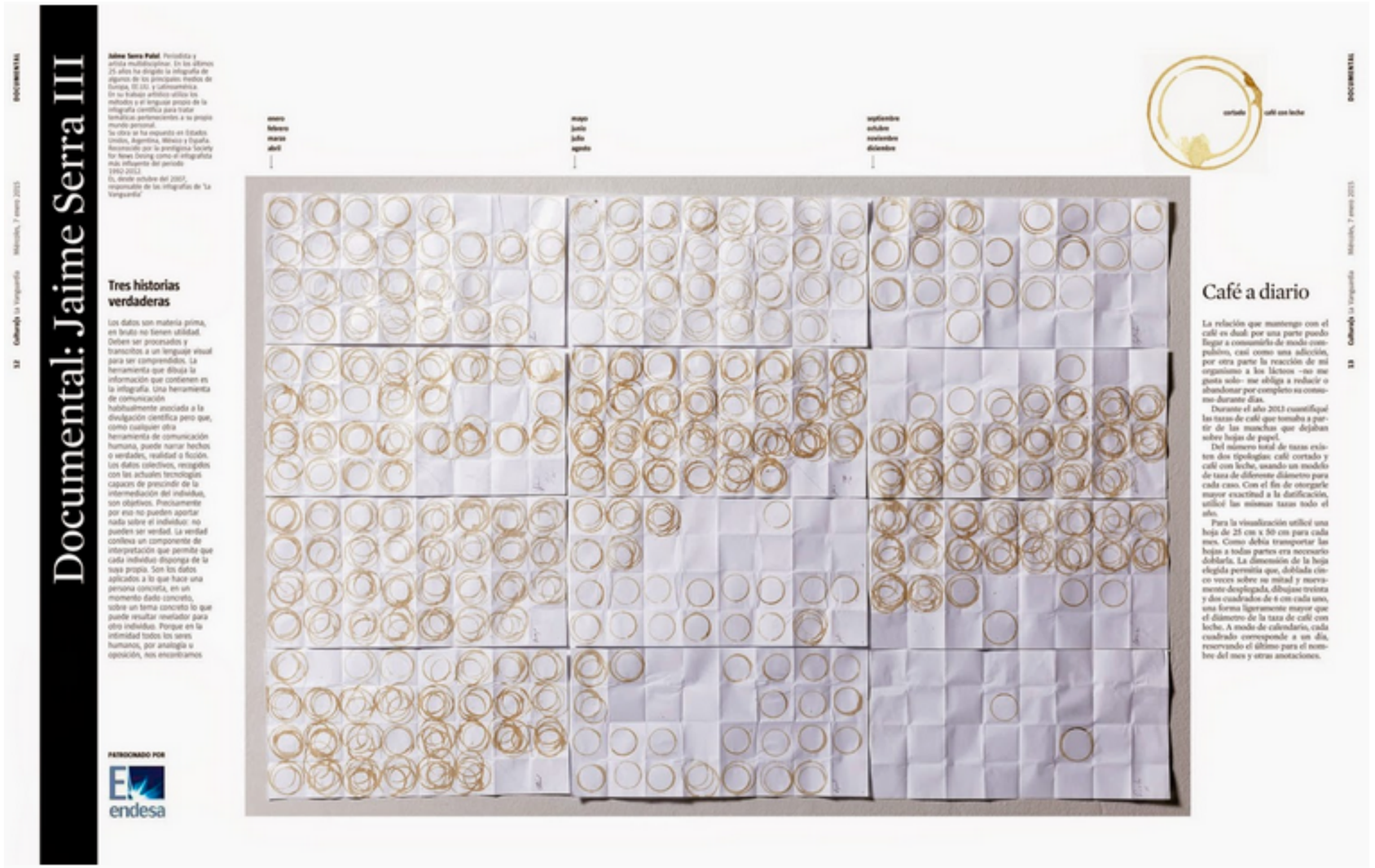
a. Net Medicare spending (includes offsetting receipts from premium payments by beneficiaries, recoveries of overpayments made to providers, and amounts paid by states from savings on Medicaid's prescription drug costs).

b. Spending to subsidize health insurance purchased in the marketplaces established under the Affordable Care Act and provided through the Basic Health Program and spending to stabilize premiums for health insurance purchased by individual people and small employers.

https://jaimeserra-archivos.blogspot.com/

Fully customized visualizations:

Appropriate for one-time use when we want to provoke curiosity, surprise —or simply a smile

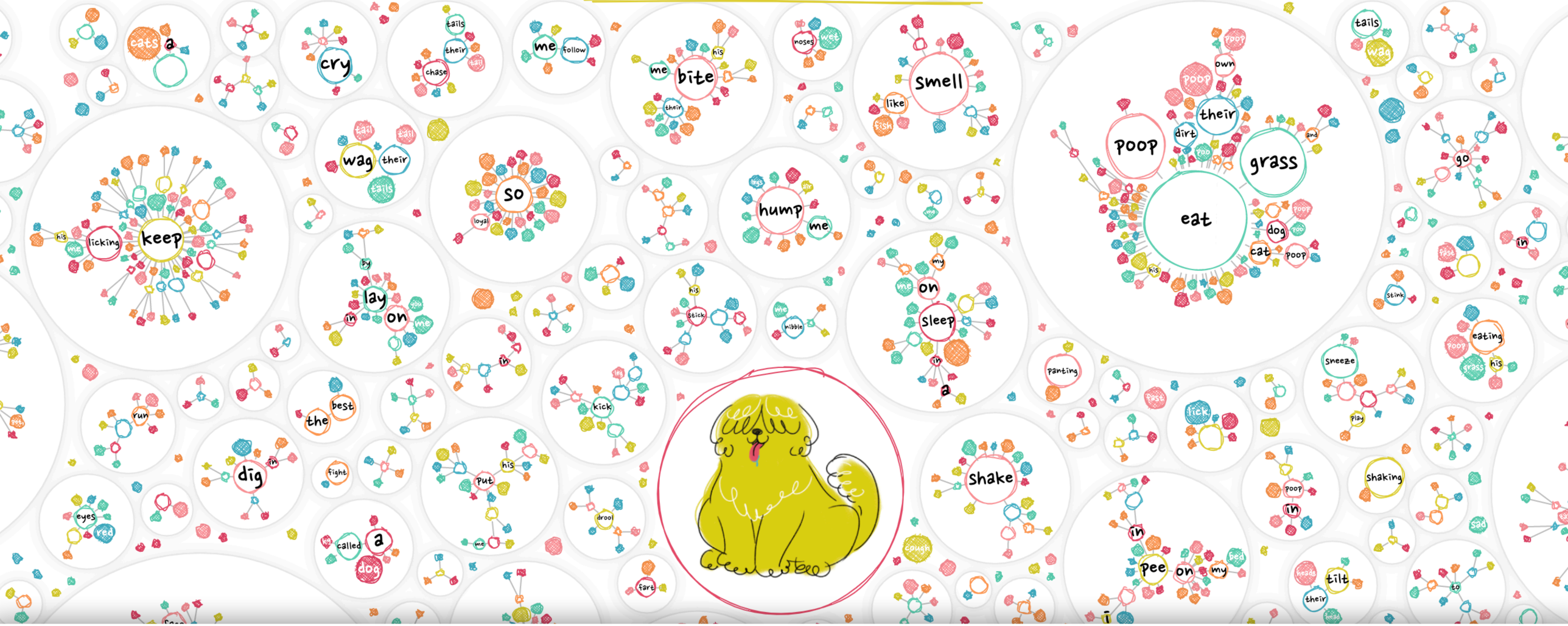


http://visualoop.com/28792/portfolio-of-the-week-jaime-serra

Making data feel “warmer”

https://www.washingtonpost.com/graphics/2020/national/coronavirus-deaths-neighborhood/

The purpose of visualization isn't visualization per se. The purpose of visualization is to help people **make sense of the world** through a combination of visuals and words.



The End.

www.thefunctionalart.com , www.albertocairo.com , alberto.cairo@gmail.com