# **MAKING DECISIONS IN VISUALIZATION**



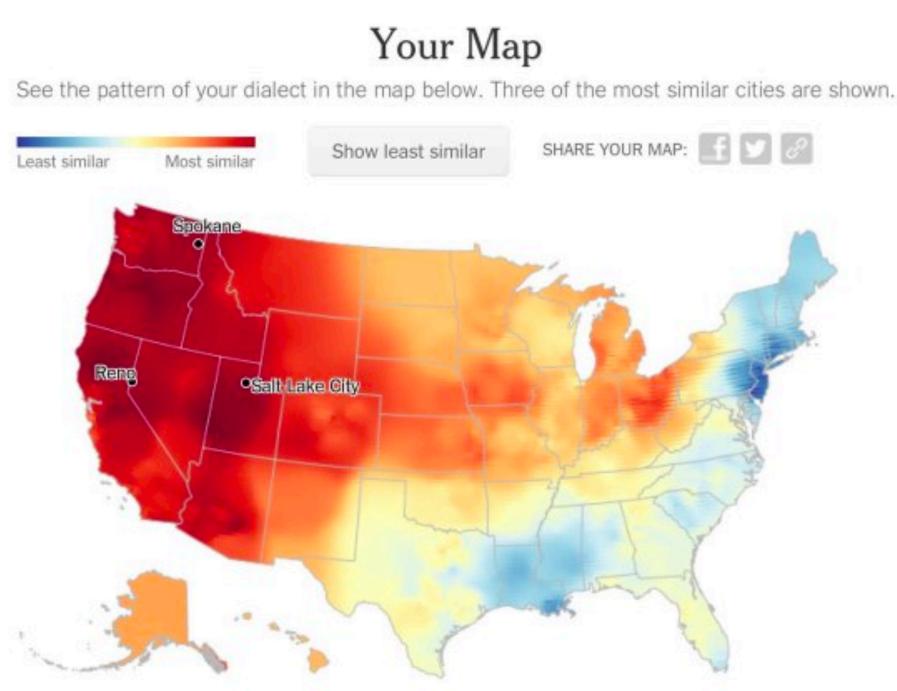
# Alberto Cairo

Coda.Br 2020

# We are living through a golden age of information graphics

## How Y'all, Youse and You Guys Talk

What does the way you speak say about where you're from? Answer all the questions below to see your personal dialect map.



These maps show your most distinctive answer for each of these cities.

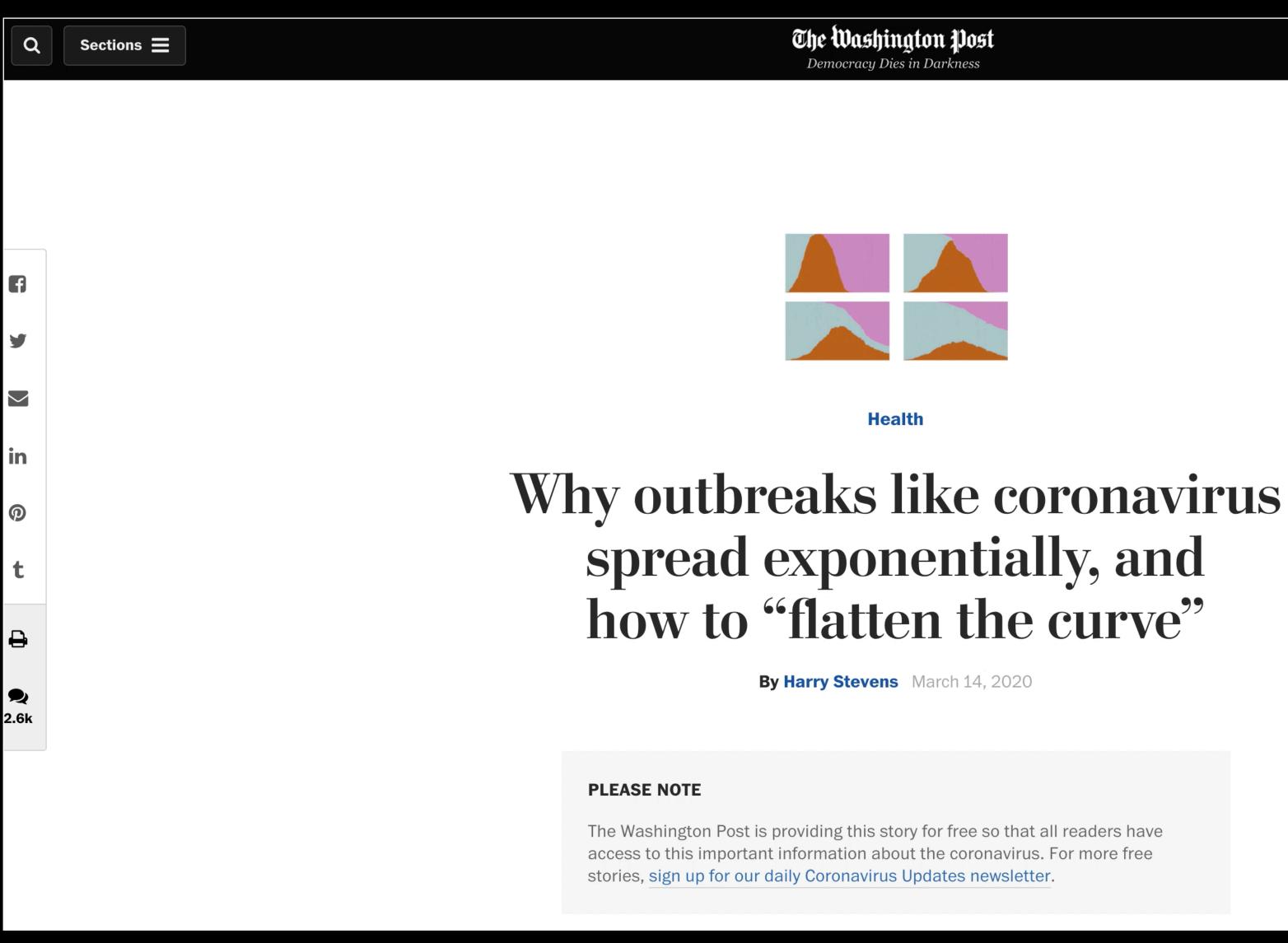






# NYTimes' Dialect Quiz

https://www.nytimes.com/interactive/2014/ upshot/dialect-quiz-map.html



Alberto Cairo • University of Miami • <u>www.thefunctionalart.com</u> • Twitter: @albertocairo

Alberto Cairo To... 👤

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

set the record for converting readers to subscribers.

https://www.washingtonpost.com/pr/2020/06/26/washington-post-expand-graphics-design-teams-with-14-new-positions/

Alberto Cairo • University of Miami • <u>www.thefunctionalart.com</u> • Twitter: @albertocairo

The Post's visual journalism, which involves staff throughout the newsroom, has attracted large audiences and contributed to record subscriber growth. Six of the seven most visited stories in The Washington Post's history have been graphics, including the coronavirus simulator that became the most visited article in The Post's history, with more than three times as many visits as the second. It also includes this year's Democratic candidate quiz, which

# At the same time, we're still dealing with some misconceptions and myths:

2. "Visualization is intuitive" 3. "'The data should speak for itself" 4. "Show, don't tell!"

I. "A picture is worth a thousand words"

5. "Learn the rules of visualization (by reading this book by [insert author's name here])!"

Alberto Cairo • University of Miami • <u>www.thefunctionalart.com</u> • Twitter: @albertocairo

## Visuals are often ambiguous



Alberto Cairo • University of Miami • <u>www.thefunctionalart.com</u> • Twitter: @albertocairo

## Visuals are often ambiguous



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."

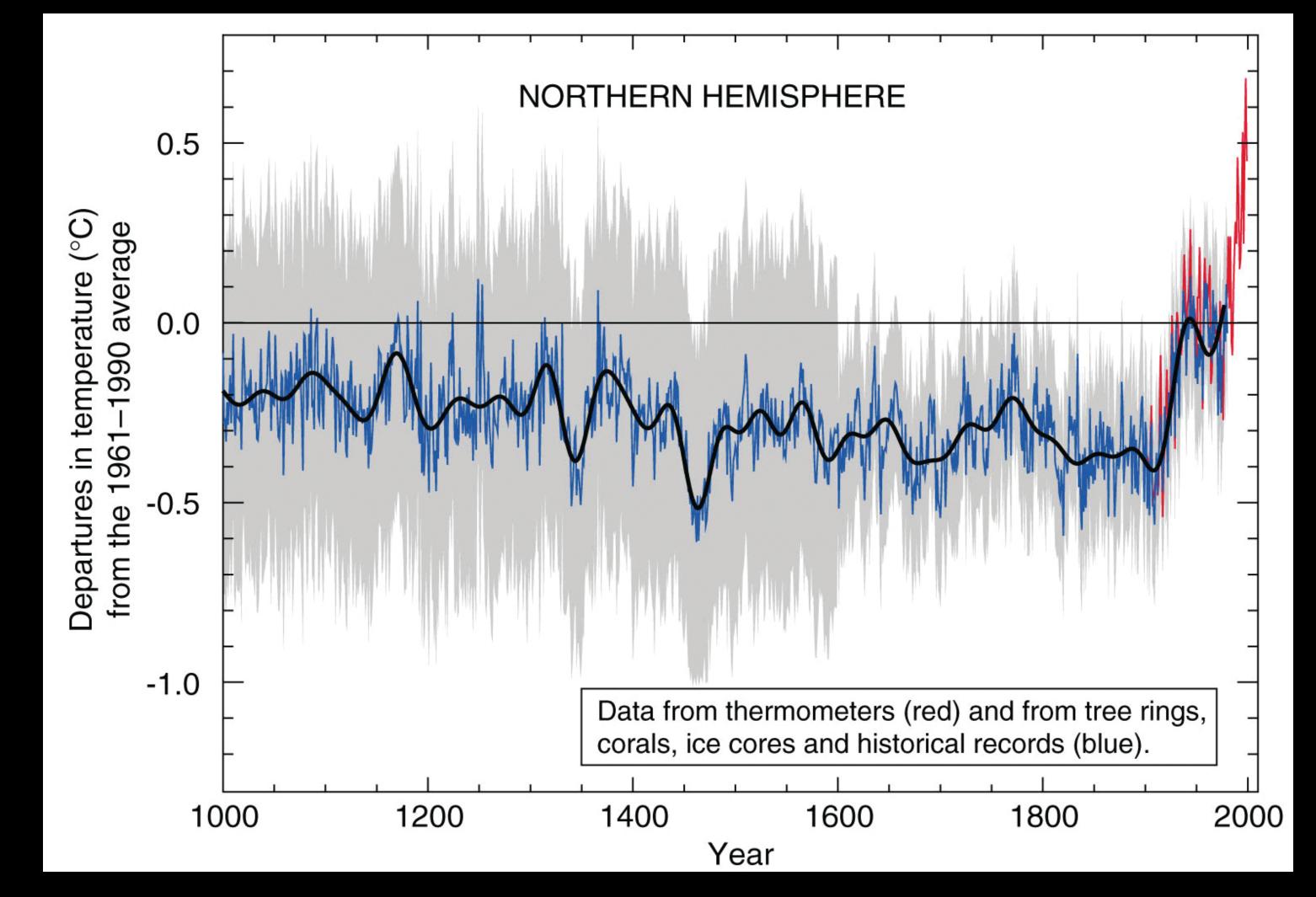
Alberto Cairo • University of Miami • <u>www.thefunctionalart.com</u> • Twitter: @albertocairo

Derek Parfit, On What Matters

# I.Why should my visualization exist? Do the potential benefits of designing my visualization outweigh the possible harm it might cause?

# Detecting patterns

9 🔴 🔵					<b>I</b>	Vorkbook1								Workbook1		
2 🛅		R 🖹 🔏 🔞 • (	on - Σ - Aμ,	• 搔 🐼 🔄 🖷						🔏 🔓 🖺 🞻 👩 • 🕤	× <b>∑</b> • ∰u	• 🐨 • 🕢 🖻 🖡		TOTADOUR I		
A Ho			SmartArt Formu	ulas Data Revi				t te			artArt Form		eview			
E	fit	Font		Alignment	Numbe			F I	idit :	Font		Alignment	Numb		Format	t
<b>i</b> - I	Fill 🔻 Calibri (B			abc 🔻 📰 🖓 Wrap	Text - General	▼ <b></b>	Normal	•	Fill Cal		A•  ≡  ≡		ap Text 🔻 General	<b>▼</b>	Normal Bad	
Paste (	✓ Clear ▼ B I	<u>U</u>			terge 🔻 🥞 🔻 % 🦻	Conditional	Neutral	Paste	Clear • B				Merge 🔻 贒 🔻 % 🦻	Conditional	Neutral	culation
He		(• fx B	С	D	E	F	G	H	6 <b>;</b> 8	<b>B</b>	6	D	E	E	G	Ц
1	A /EAR				L SIGMA	F	U	878	1876	-0.1891	1876	0.113228	0.226456	► 8.25297E-02		
-	000	0.0659	1000	0.240346	0.480693	0.206137	0.123588	879	1878	-0.0140	1878	0.113228	0.226458	8.25297E-02 8.25299E-02		
3								880	1878							
4	1001	-0.1241	1001	0.240347	0.480694	0.206137	0.123589	881		-0.0873	1878	0.113228	0.226457	8.25298E-02		
5	1002	-0.1208	1002	0.240346	0.480692	0.206136	0.123588	882	1879	-0.2959	1879	0.113229	0.226458	8.25302E-02		
5 6	1003	-0.1801	1003	0.240347	0.480694	0.206137	0.123589	883	1880	-0.2368	1880	0.113229	0.226457	8.25300E-02		
7	1004	-0.0711	1004	0.240347	0.480693	0.206137	0.123588	884	1881	-0.1977	1881	0.113229	0.226458	8.25302E-02		
8	1005	-0.1334	1005	0.240346	0.480692	0.206136	0.123588	885	1882	-0.2036	1882	0.113229	0.226457	8.25300E-02		
9	1006	-0.0644	1006	0.240346	0.480693	0.206137	0.123588	886	1883	-0.2489	1883	0.113228	0.226455	8.25293E-02		
<b>10</b>	1007	0.0042	1007	0.240347	0.480693	0.206137	0.123588	887	1884	-0.2125	1884	0.113229	0.226457	8.25301E-02		
11	1008	-0.1288	1008	0.240347	0.480693	0.206137	0.123588	888	1885	-0.1896	1885	0.113228	0.226457	8.25299E-02		
12	1009	-0.0296	1009	0.240347	0.480693	0.206137	0.123588	889	1886	-0.1084	1886	0.113228	0.226456	8.25298E-02		
12	1010	0.1187	1010	0.240347	0.480694	0.206137	0.123589	890	1887	-0.3265	1887	0.113228	0.226456	8.25296E-02		
	1011	-0.1252	1011	0.240346	0.480692	0.206136	0.123588		1888	-0.1694	1888	0.113228	0.226457	8.25298E-02		
14	1012	-0.1634	1012	0.240347	0.480694	0.206137	0.123588	891	1889	-0.1339	1889	0.113228	0.226456	8.25298E-02		
15	1013	-0.0791	1013	0.240347	0.480693	0.206137	0.123588	892	1890	-0.3107	1890	0.113229	0.226457	8.25301E-02		
16	1014	-0.1120	1014	0.240347	0.480693	0.206137	0.123588	893	1891	-0.1754	1891	0.113229	0.226457	8.25300E-02		
17	1015	-0.1146	1015	0.240346	0.480692	0.206136	0.123588	_	1892	-0.3186	1892	0.113228	0.226456	8.25295E-02		
18	1016	-0.1206	1016	0.240346	0.480692	0.206136	0.123588	000	1893	-0.3236	1893	0.113228	0.226456	8.25297E-02		
19	1017	-0.0815	1017	0.240347	0.480693	0.206137	0.123588		1894	-0.1970	1894	0.113228	0.226456	8.25295E-02		
20	1018	-0.2031	1018	0.240346	0.480693	0.206137	0.123588	897	1895	-0.1578	1895	0.113228	0.226456	8.25297E-02		
21	1019	0.0305	1019	0.240347	0.480693	0.206137	0.123588	898	1896	-0.0804	1896	0.113228	0.226456	8.25298E-02		
22	1020	0.1098	1020	0.240347	0.480694	0.206137	0.123589	899	1897	-0.0537	1897	0.113228	0.226456	8.25298E-02		
23	1021	0.0244	1021	0.240347	0.480693	0.206137	0.123588	900	1898	-0.2195	1898	0.113229	0.226457	8.25301E-02	2 7.75211	E-02
24	1022	-0.0743	1022	0.240347	0.480693	0.206137	0.123588		1899	-0.3486	1899	0.113228	0.226456	8.25297E-02	2 7.75207	/E-02
25	1023	-0.0323	1023	0.240347	0.480693	0.206137	0.123588	902	1900	-0.1253	1900	0.113229	0.226457	8.25300E-02	2. 7.75210	)E-02
26	1024	-0.0434	1024	0.240346	0.480693	0.206137	0.123588	903	1901	-0.1575	1901	0.113228	0.226456	8.25296E-02	7.75206	5E-02

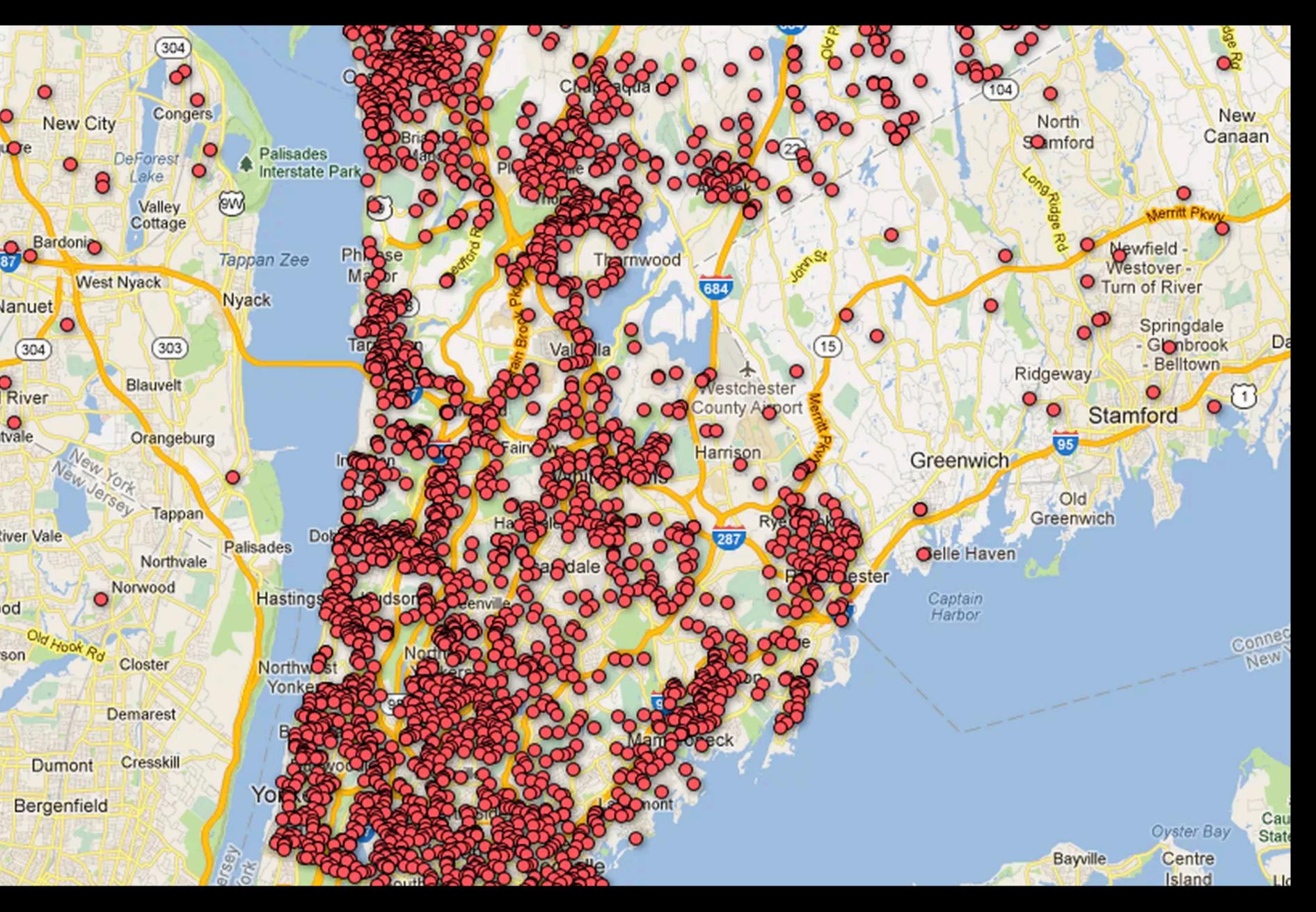


# Michael E. Mann, Raymond S. Bradley, and Malcolm K. Hughes Intergovernmental Panel on Climate Change (IPCC), Third Report, 2001

Alberto Cairo • University of Miami • <u>www.thefunctionalart.com</u> • Twitter: @albertocairo

# Detecting patterns

## 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."

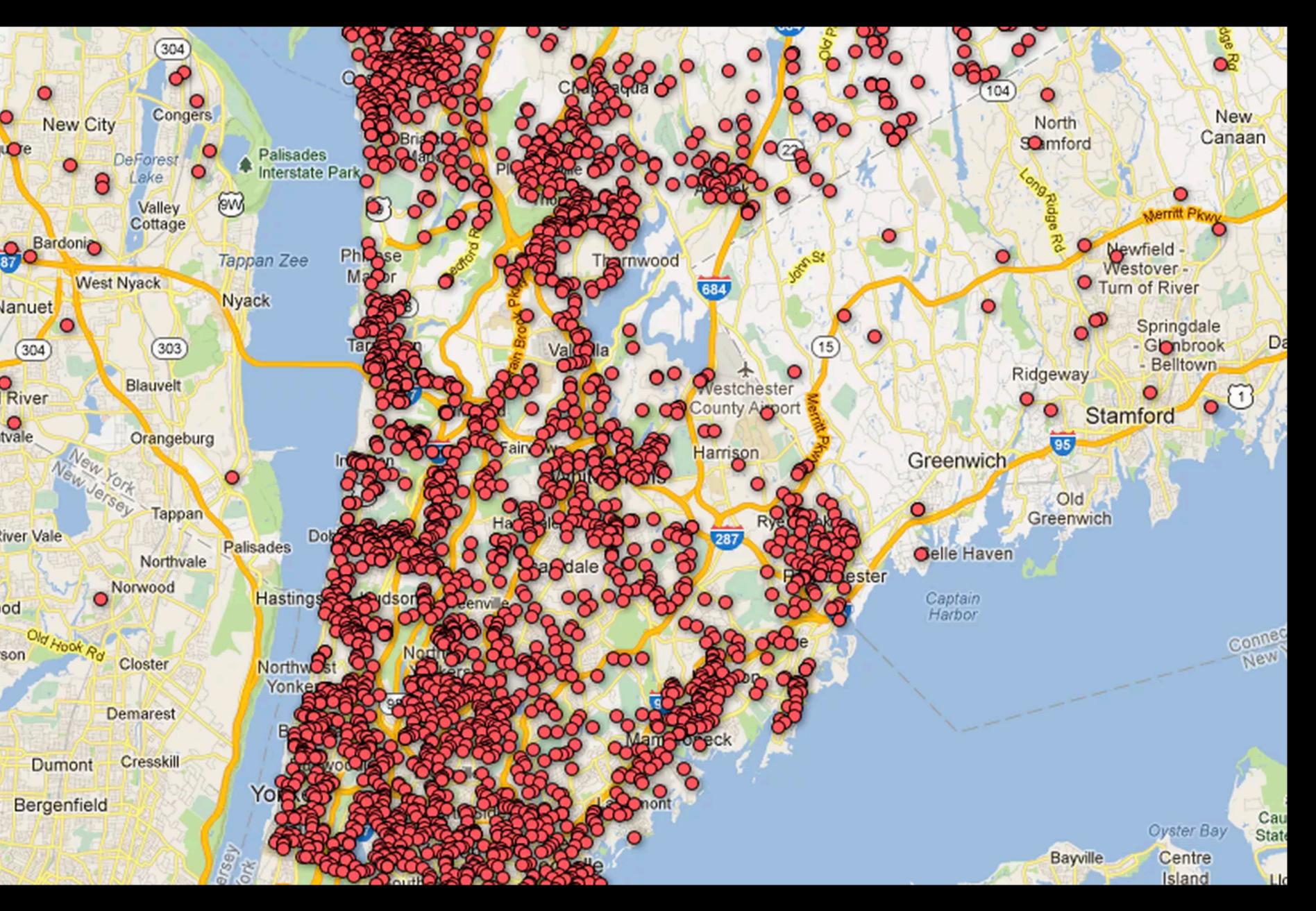
https://www.theverge.com/ 2/12/25/3802960/new-<u>york-newspaper-posts-map-</u> with-names-addresses-ofgun-owners







## 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)

River

lanue

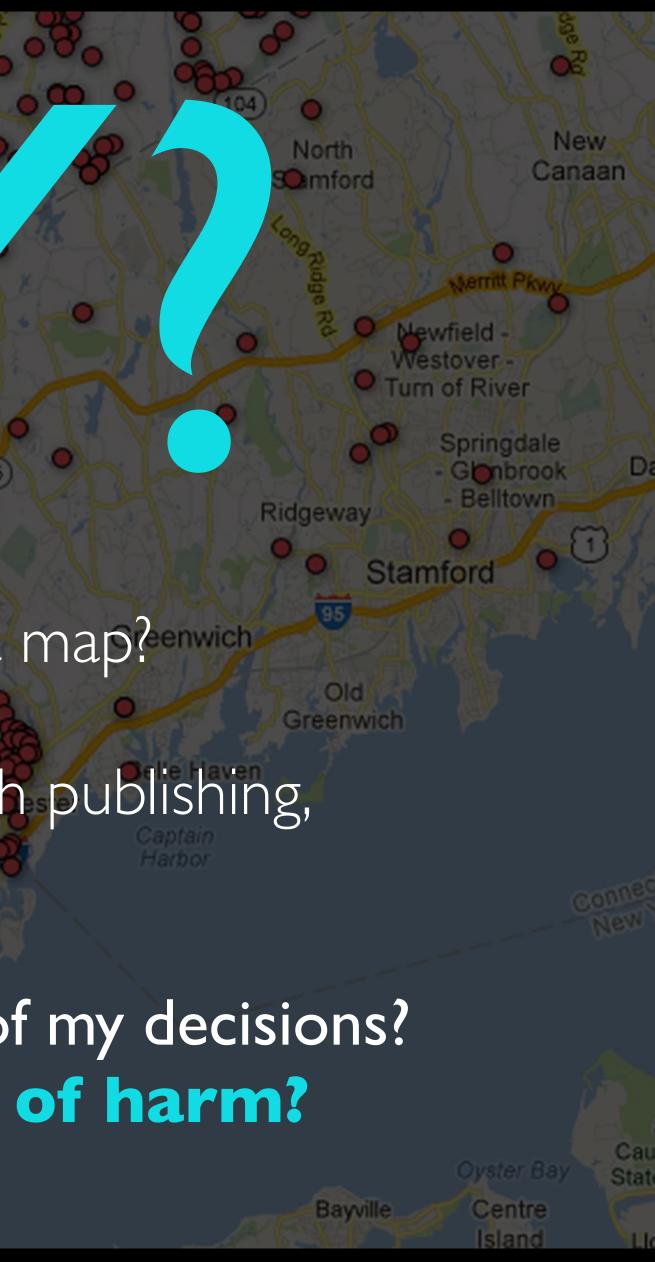
iver Vale

od Or Why should this data be made public? Why should it be made public through a map? Why should it be *this type* of map? Even if we decided that this data is worth publishing, wouldn't a different map be better?

state F

Nyack

What are the potential consequences of my decisions? Are the benefits worth the risk of harm?

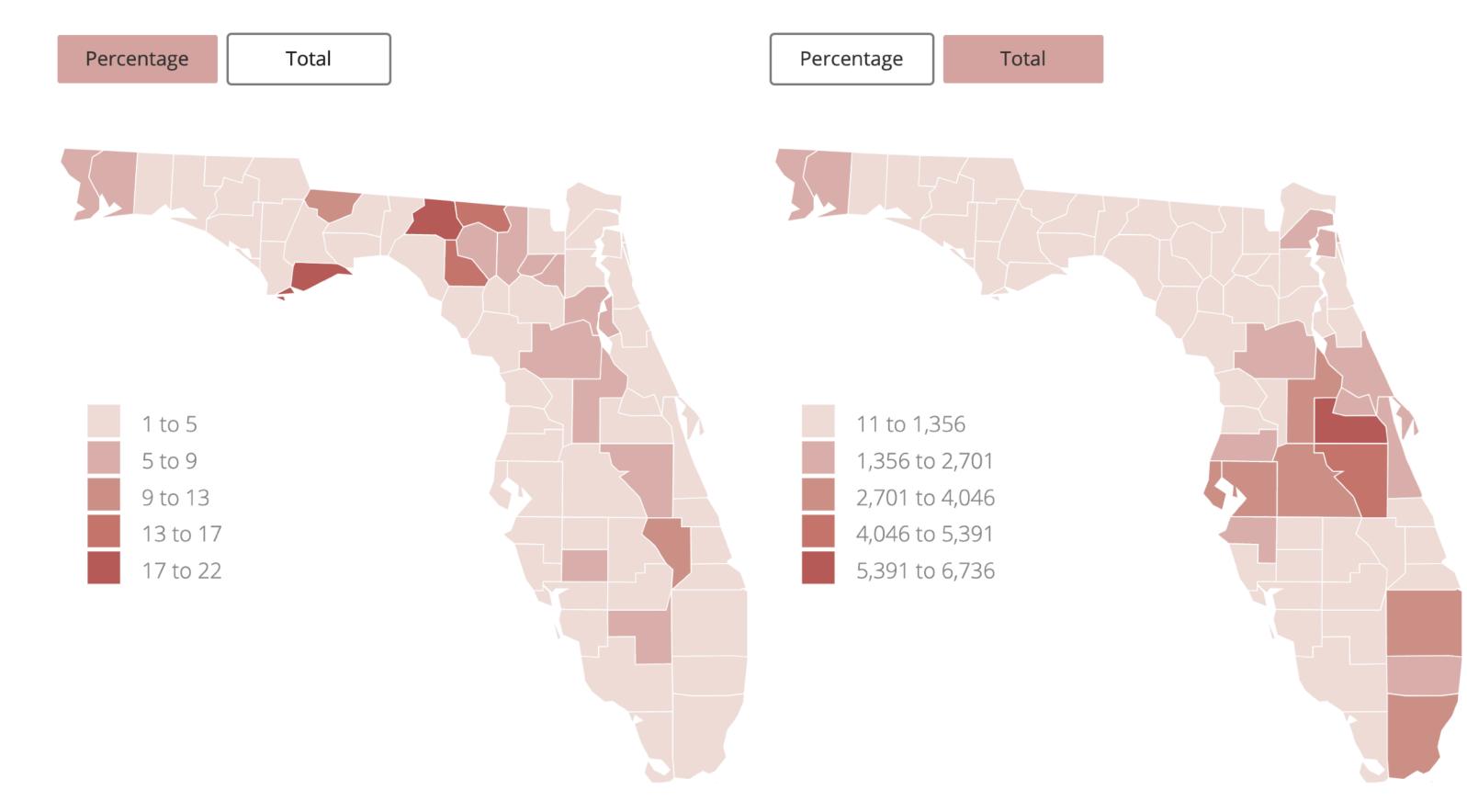


# 2. What to visualize?

Do I understand my data, its limitations, uncertainty, or glitches? What or who is being measured (*o not being measured*,) and why?

# ATSCHOOL 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.





# $\equiv \mathbf{ELPAIS}$

ANDALUCÍA CATALUÑA C. VALENCIANA GALICIA MADRID PAÍS VASCO MÁS COMUNIDADES TITULARES »

### DATOS DEL CENTRO DE ESTUDIOS DE OPINIÓN >

# El no a la independencia de Cataluña gana al sí por primera vez desde 2012

El 'CIS catalán' constata que el apoyo a la secesión cae un 9% en los últimos dos meses



PERE RÍOS

Barceicha - 19 DIC 2014 - 15:51 EST

https://elpais.com/ccaa/2014/12/19/catalunya/1418984873\_128596.html

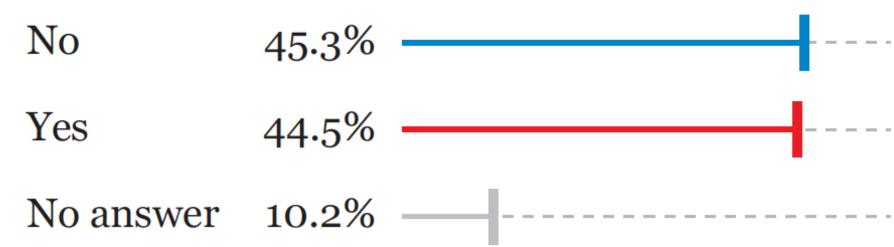




For the first time since Catalan leader Artur Mas began his ongoing independence drive in 2012, a survey shows that a majority in the region would reject secession if a referendum were held now.

The latest poll by the Catalan executive's Opinion Studies Center (CEO) shows that 45.3 percent of citizens would vote no to the question: "Would you like Catalonia to become an independent state?" compared with 44.5 percent who would support the move.

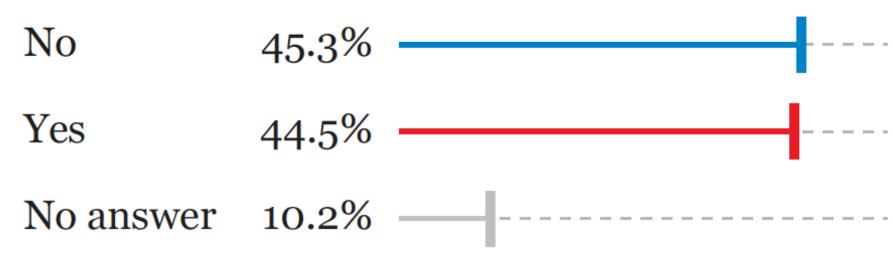
### Do you want Catalonia to become an independent state?



For the first time since Catalan leader Artur Mas began his ongoing independence drive in 2012, a survey shows that a majority in the region would reject secession if a referendum were held now.

The latest poll by the Catalan executive's Opinion Studies Center (CEO) shows that 45.3 percent of citizens would vote no to the question: "Would you like Catalonia to become an independent state?" compared with 44.5 percent who would support the move.

### Do you want Catalonia to become an independent state?



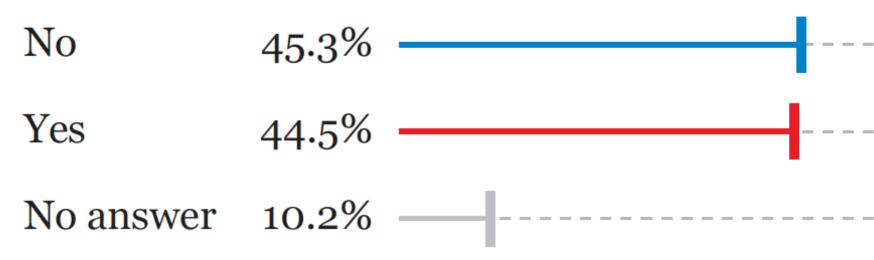
Margin of error: +/-2.95 at 95% confidence level

"The margin of error of the poll is 2.95, a relevant fact considering the tight difference between the YES and the NO to the independence of Catalonia"

For the first time since Catalan leader Artur Mas began his ongoing independence drive in 2012, a survey shows that a majority in the region would reject secession if a referendum were held now.

The latest poll by the Catalan executive's Opinion Studies Center (CEO) shows that 45.3 percent of citizens would vote no to the question: "Would you like Catalonia to become an independent state?" compared with 44.5 percent who would support the move.

### Do you want Catalonia to become an independent state?



Margin of error: +/-2.95 at 95% confidence level

"The margin of error of the poll is 2.95, a relevant fact considering the tight difference between the YES and the NO to the independence of Catalonia"

### Do you want Catalonia to become an independent state?

 No	45.3%
 Yes	44.5%
 No answer	10.2%

The probability of the tiny difference between the "No" and the "Yes" being just due to random chance is very high

# Disclosing limitations and uncertainty

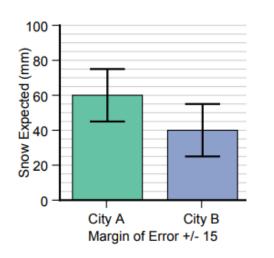
**Uncertainty and graphicacy** How should statisticians, journalists, and designers reveal uncertainty in graphics for public consumption?

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

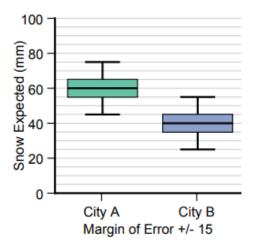
Alberto Cairo • University of Miami • <u>www.thefunctionalart.com</u> • Twitter: @albertocairo

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

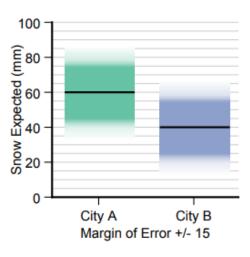
Michael Correll Student Member, IEEE, and Michael Gleicher Member, IEEE

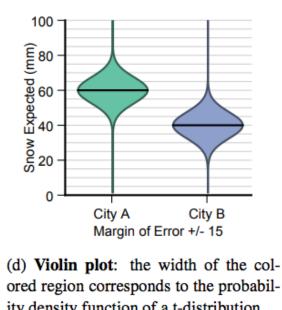


(a) **Bar chart** with error bars: the height of the bars encodes the sample the 95% t-confidence interval, the box is mean, and the whiskers encode a 95% t- a 50% t-confidence interval confidence interval.



(b) **Modified box plot**: The whiskers are



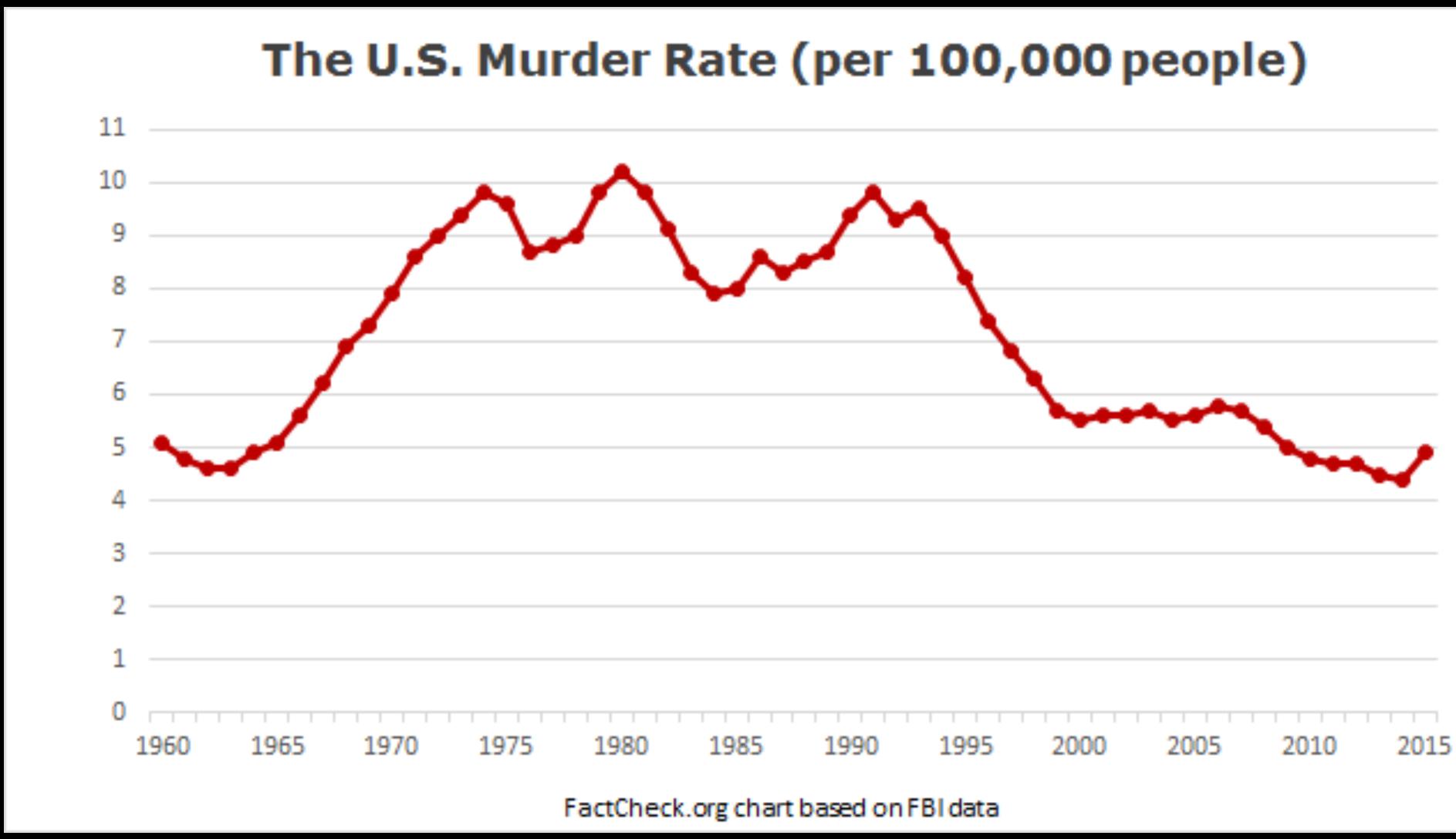


(c) **Gradient plot**: the transparency of the colored region corresponds to the cumulative density function of a t- ity density function of a t-distribution. distribution.

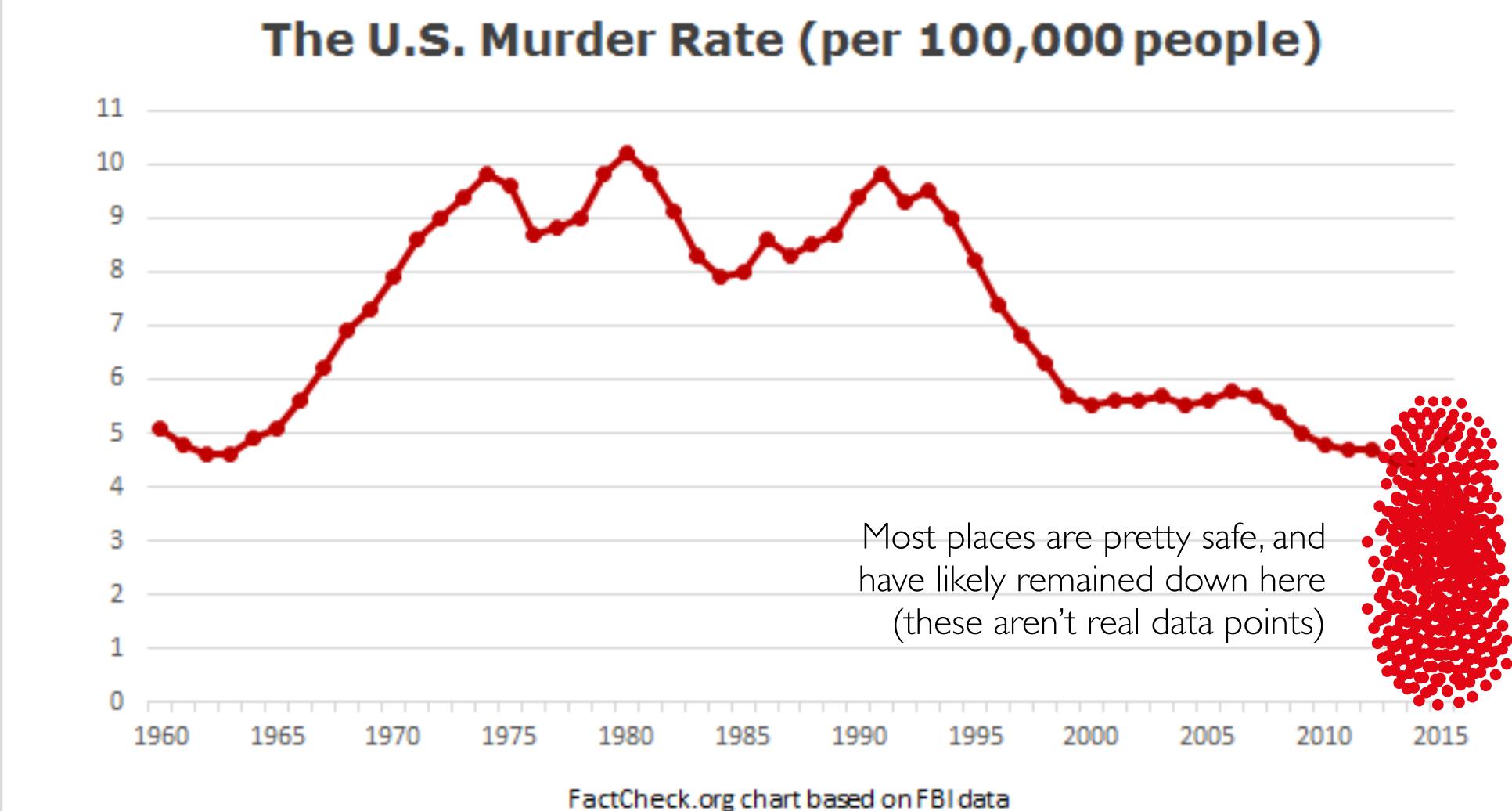
https://graphics.cs.wisc.edu/Papers/2014/CG14/Preprint.pdf

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

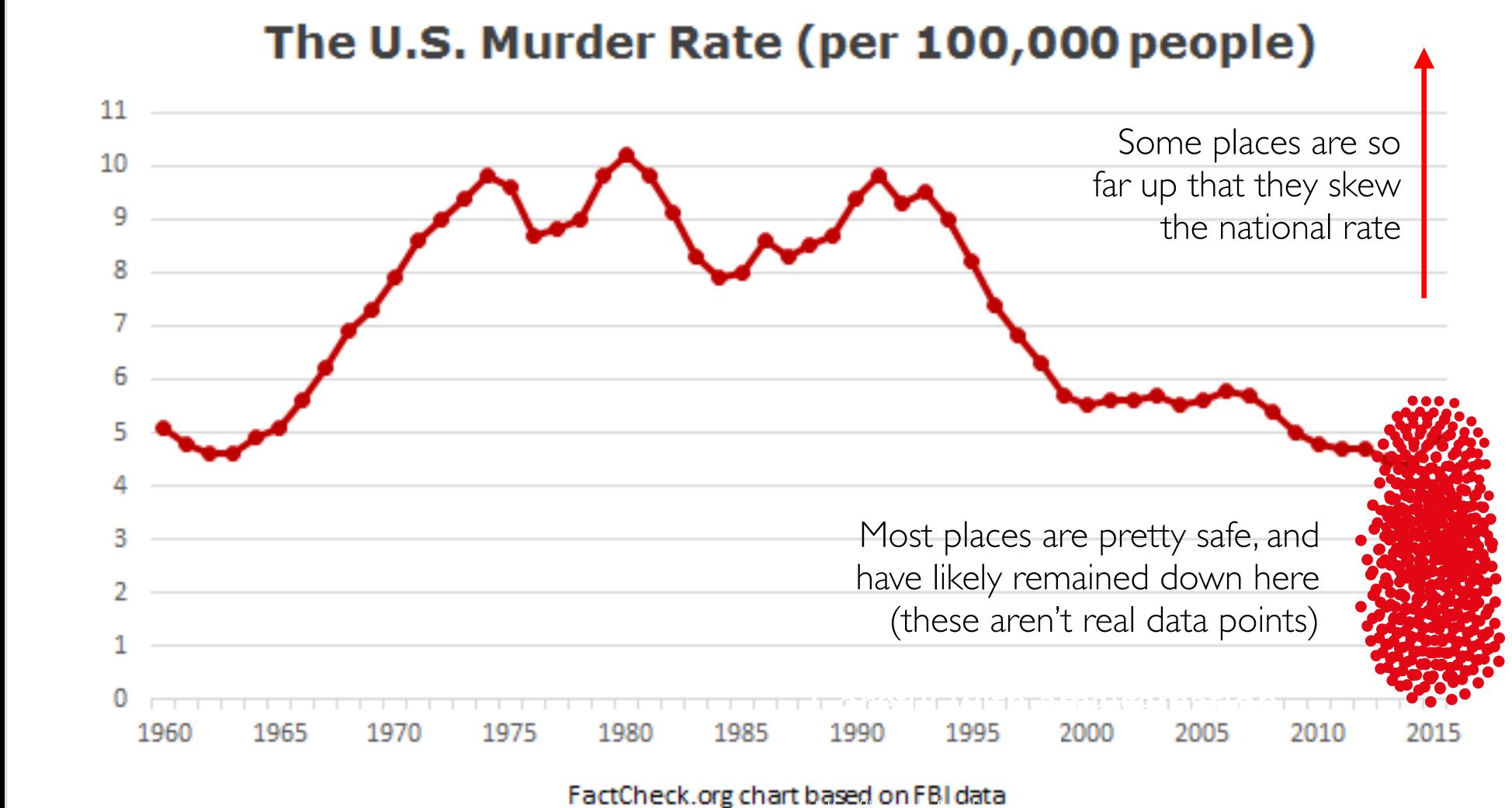
# **3. How much to visualize?** Am I showing too little? Am I showing too much?



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



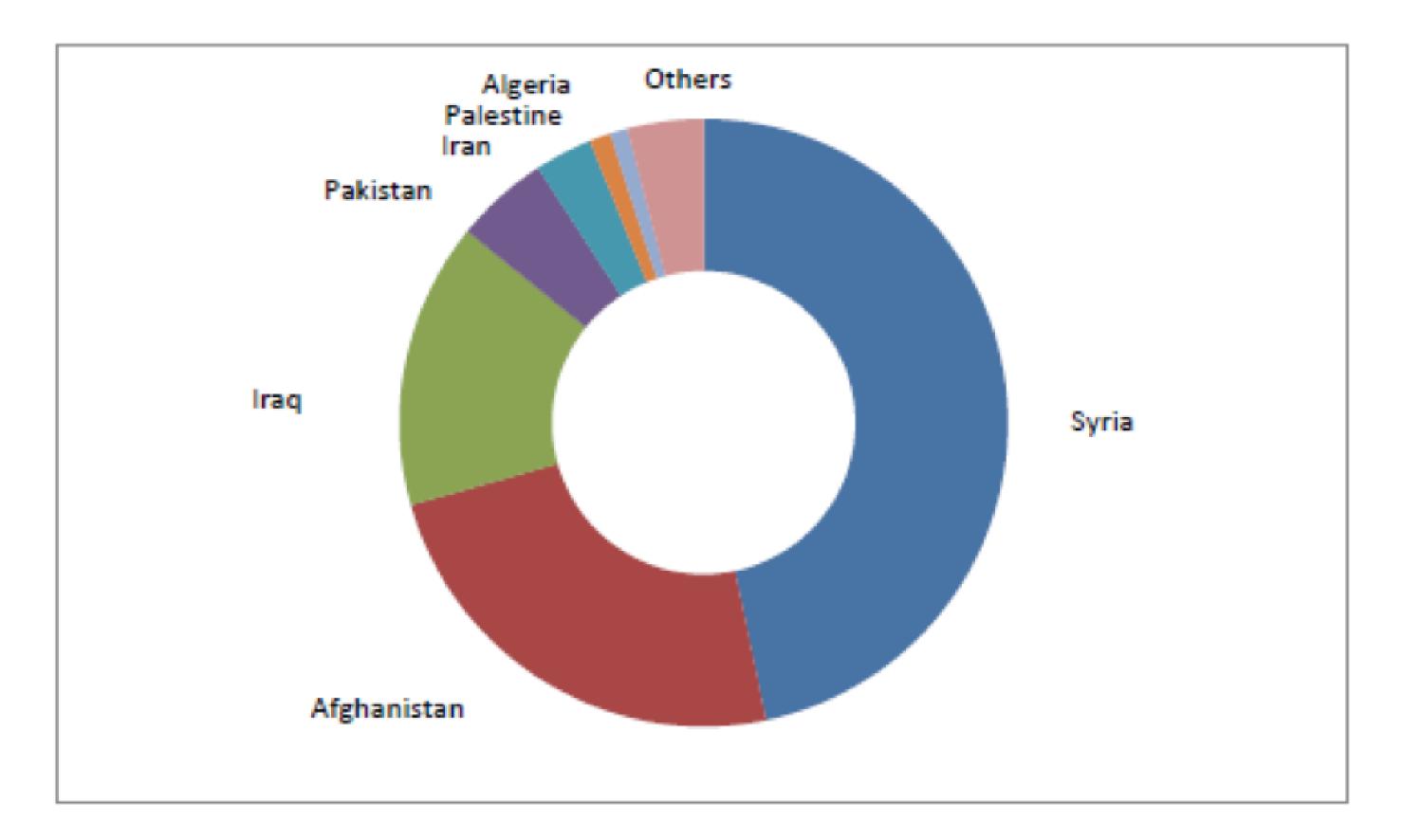
## The danger of aggregating data too much, and presenting just averages and other statistical summaries



# **4. How to visualize it?** What types of charts or maps should I use? What is the best way to organize the visualization?

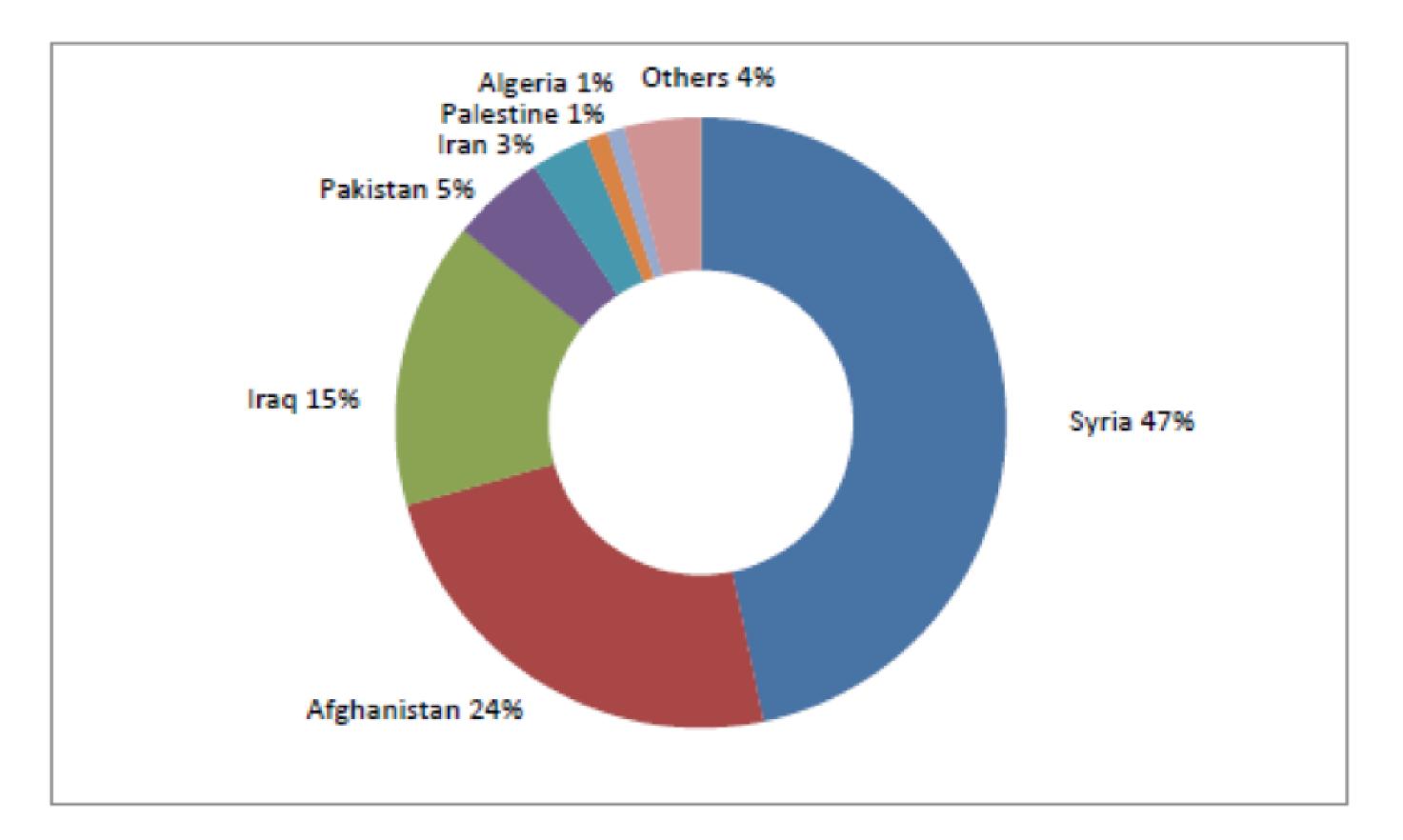
#### Figure 2 - Main nationalities of arriving migrants – 2016

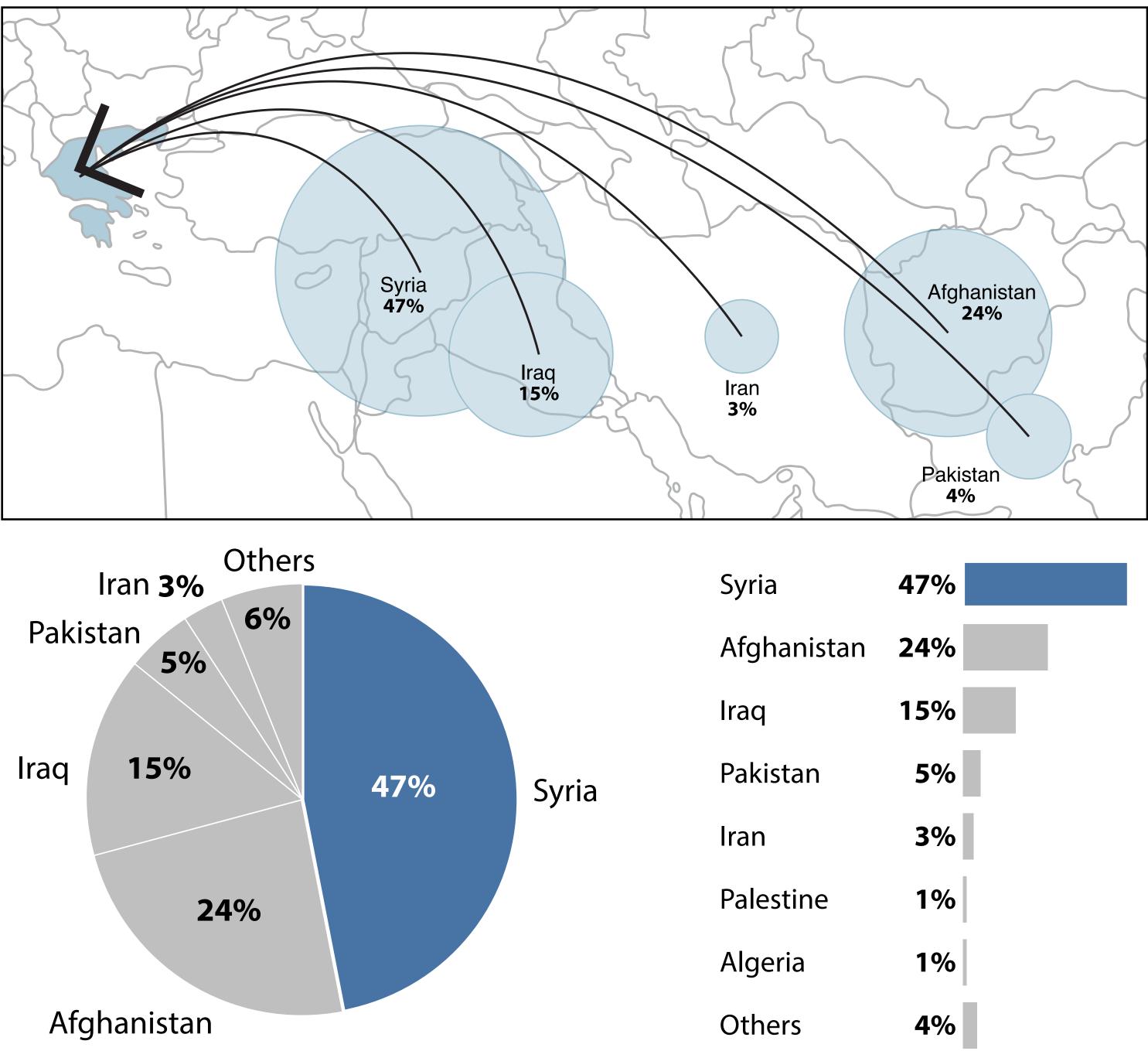
#### Greece



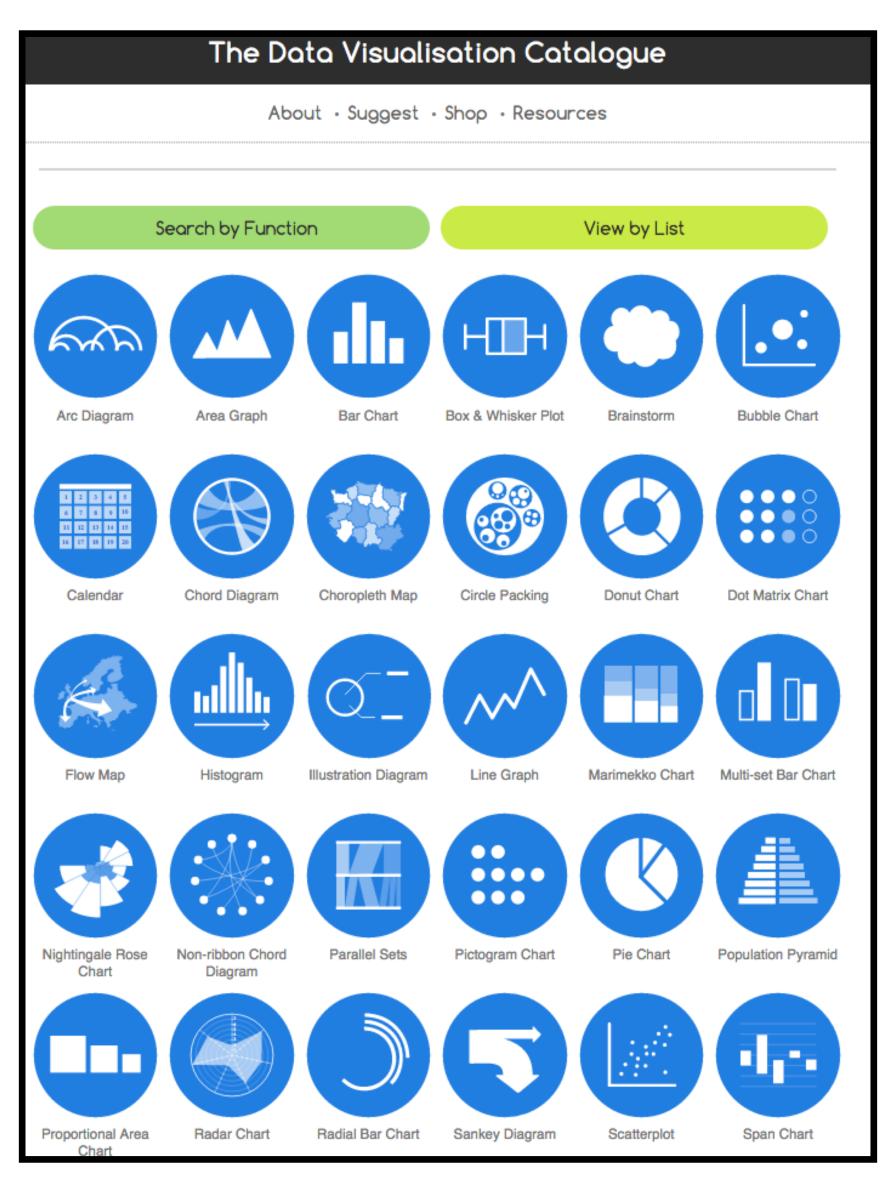
#### Figure 2 - Main nationalities of arriving migrants – 2016

#### Greece

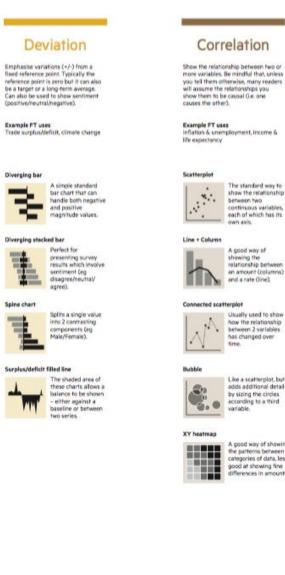




	Syria	47%
	Afghanistan	24%
	Iraq	15%
Syria	Pakistan	5%
	Iran	3%
	Palestine	1%
	Algeria	1%
	Others	4%



#### http://www.datavizcatalogue.com/



### Visual vocabulary

Designing with data

**Diverging bar** 

7

ł

The second

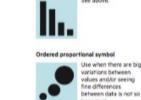
There are so many ways to visualise data - how do we know which one to pick? Use the categories across the top to decide which data relationship is most important in your story, then look at the different types of chart within the category to form some initial ideas about what might work best. This list is not meant to be exhaustive, nor a wizard, but is a useful starting point for making informative and meaningful data visualisations.

FT graphic: Alan Smith; Chris Campbell; Ian Bott; Liz Faunce; Graham Parrish; Billy Ehrenberg; Paul McCallum; Martin Stabe Inspired by the Graphic Continuum by Jon Schwabish and Severino Ribecci

ft.com/vocabulary

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

The standard way to show the relationship between two continuous variables, each of which has its



Dot strip plot





how ranks have changed over time or vary between categories. Lollipop chart Lolipops draw more attention to the data value than standard bar/column and can also show rank and value affectiony.



#### Ranking

Use where an item's position in an ordered list is more important than absolute or relative value. Don't be afraid to highlight the points of interest.

Example FT uses Wealth, deprivation, league tables,

Standard bar cham display the ranks of values much more easily when sorted into order.

See above

on a strip are a space-efficient method of laying out ranks across multiple categories.

Ordered bar

F

Ordered column

Show values in a dataset and how often they occur. The shape for Sken of a distribution can be a memorable way of highlighting the lack of uniformity or equality in the data.

Example FT uses Income distribution, population (age/sex) distribution

Distribution

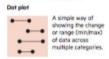








Dot strip plot distribution, can be a problem when too many dots have the same value.





A good way of showing how unequal a distribution is: y axis is always cumulative frequency, x axis is always a measure











• 1

Fan chart (projections)

Use to show the uncertainty in future projections - usually this grows the further forward to projection.

A good way of showing changing data for two variable whenever there is a relatively clear patter

atterns (daily, week onthly) – at the pense of showing Priestley timeline Great when date and

Change over Time

Give emphasis to changing trends. These can be short (btra-dey) movements or extended series traversing docades or conturies: Choosing the correct time period is important to provide suitable contex for the reader.

Example FT uses Share price movements, economic time series

The standard way to show a changing time series. If data are irregular, consider markers to represent data periors.

for showing change over time - but usually best with only one series of date at a time.

A good way of showing the relationship over time between an amount columns) and a rate dine).

Usually focused on day-to-day activity, these charts show opening/closing and hy/low points of each day.

changing data as long as the data can be simplified into 2 or 3 points without missing a key part of story.

Use with care – mean are good at showing changes to total, but seeing change in components can be very difficult.

Stock price



varying size across multiple categories





Show how a single entity can be broken down into its component elements. If the reader's interest is solely in the size of the components consider a magnitude-type chart instead. consider a instead.















Sunburst 

11



00 Waterfall

Ч	showing part-to-w relationships when some of the
	components are negative.

















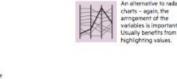


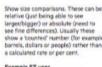












Magnitude

Example FT uses





Paired colum

Paired bar

F

Proportional stacked ba

lltu



mpare the size of ngs. Must always rt at 0 on the axis

umn but allows Itiple series. Car

ee above





於

Converting each unit o a map to a regular and equally-sized shape – good for representing voting regions with equal value

ed cartogra	im (value)
Å	Stretching and shrinking a map so that each area is sized according to a particular value.

Used to show the location of individual events/locations - make sure to annotate any parterns the reader should see.



leat map	
3	Grid-based d mapped with intensity colo As choroplet but not snap

iata value n an our scale. Ih map -iped to an

Isotype (pictog	ram)
ttttt ttt	Excellent solution in some instances – use only with whole numbers (do not sice off an arm to represent a decimal).
Lollipop chart	
-	Lollipop charts draw more attention to the data value than

	represent a decimal).
rt	
•	Lollipop charts draw more attention to the data value than standard bar/column – does not HAVE to star at zero (but preferable

showing value pf ultiple variables - b ake sure they are partised in a way th

chart	
	Lollipop charts dra more attention to data value than standard bar/colur does not HAVE to

	represent a dec
Lollipop chart	
	Lollipop charts- more attention data value than standard bar/co does not HAVE at zero (but pre

Radar char

Parallel coordinate

ttttt ttt	Excellent solution i some instances – u only with whole numbers (do not sl off an arm to represent a decima
Lollipop chart	
=	Lollipop charts dra more attention to t data value than

only with whole numbers (do not sil
off an arm to represent a decimal
Lollpop charts draw













Spatial

Used only when precise locations geographical patterns in data are more important to the reader than anything else.

Example PT uses Locator maps, population density, natural resource locations, natural disaster risk/impact, catchment are variation in election results

**Basic choropleth (rate/ratio)** 

The standard approach for putting data on a map – should always be rates rather than totals and use a sensible base

Use for totals rather than rates - be wary that small differences in data will be hard to see.

















wing the size and portion of data at a same time - as ing as the data at **Proportional symbol** •••

#### Flow

Show the reader volumes or intensity of movement between two or more states or conditions. These might be logical sequences or geographical locations.

Example FT uses Movement of funds, trade, migrants, lawsuits, information; relationship graphs.

Sankey



hows changes in flows rom one condition to t least one other; good or tracing the eventual utcome of a complex



Designed to show the sequencing of data through a flow process, typically budgets. Can include a flow process, typically



A complex but powerful diagram which can illustrate 2-way flows Cand net winner) in a matrix.



Used for showing the strength and inter-connectdness of relationships of varying types.





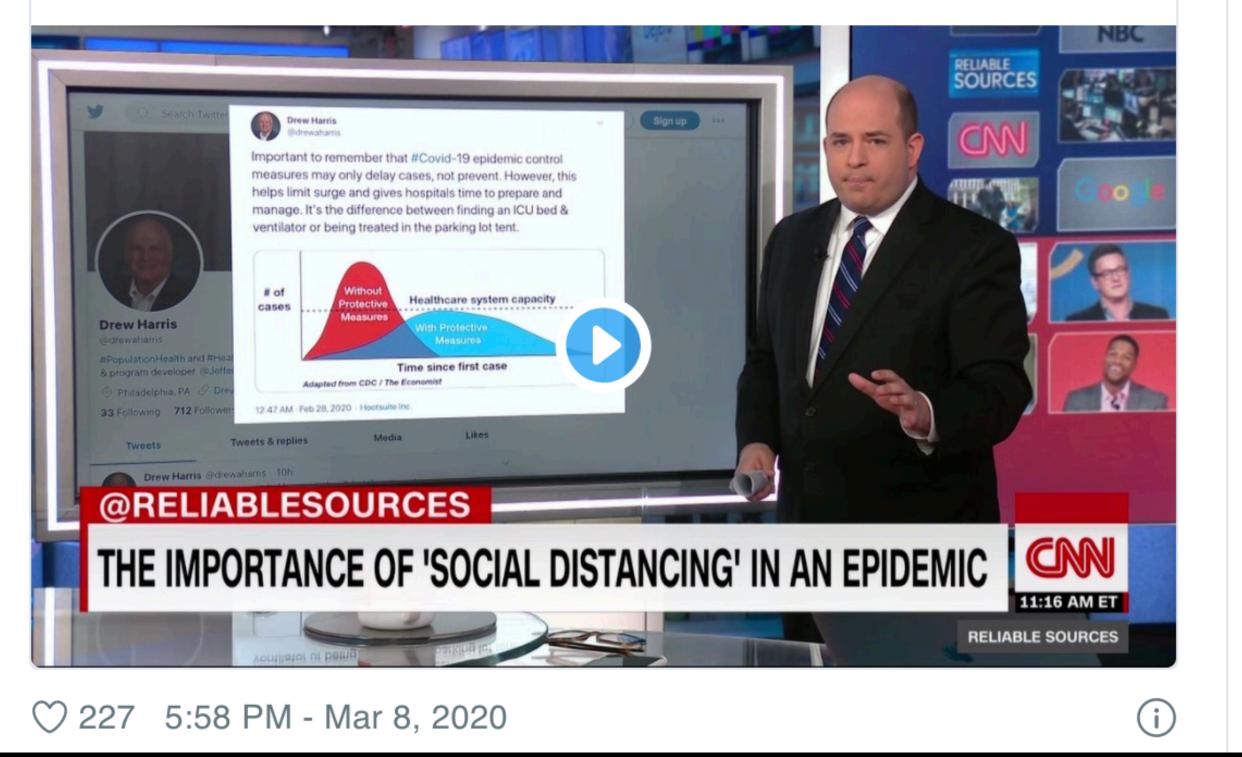


Alberto Cairo 🤡

Good journalism isn't just showing charts. It's also about explaining them: twitter.com/brianstelter/s...

Brian Stelter 🤣 @brianstelter

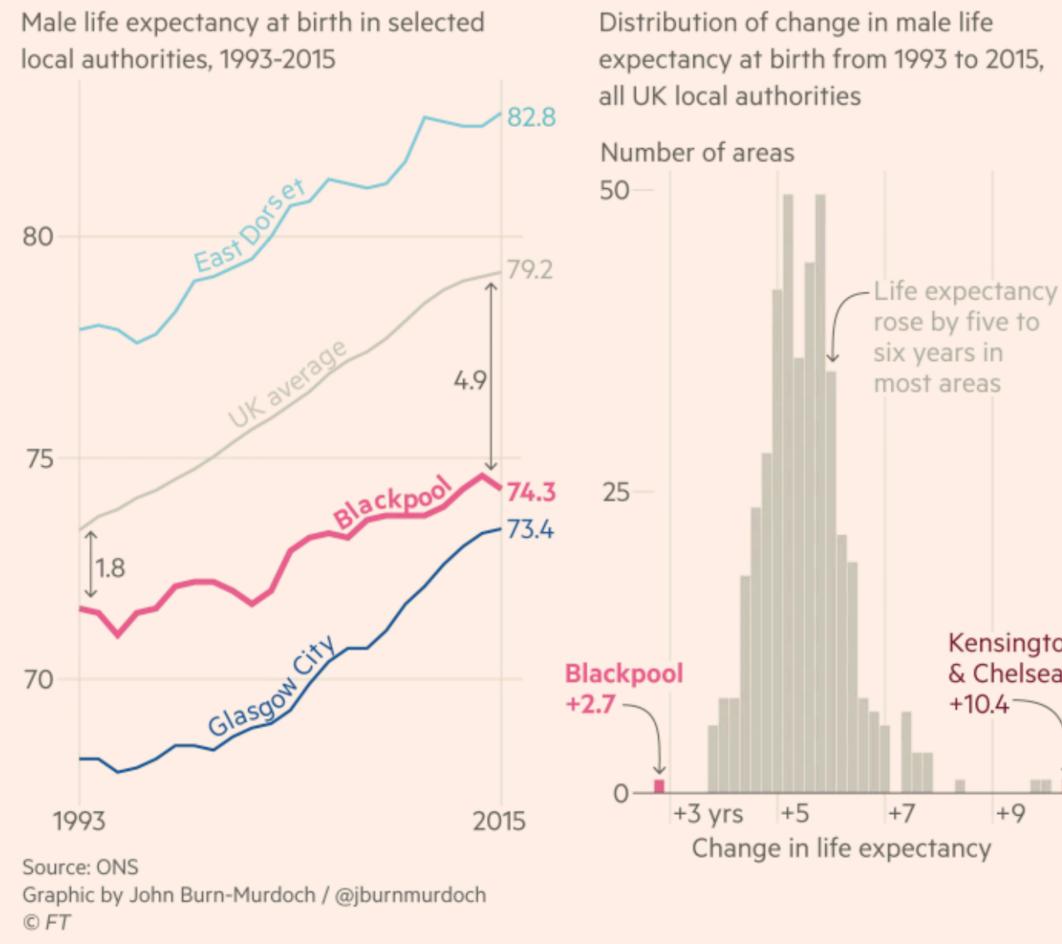
This infographic is worth a thousand words – showing why "social distancing" and other protective measures helps to slow an outbreak. Hat tips to CDC, @theeconomist, @drewaharris, and @CT\_Bergstrom



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



#### Boys born in **Blackpool** can expect to live just 74 years — the second lowest in the UK, and up by just 2.7 years since 1993

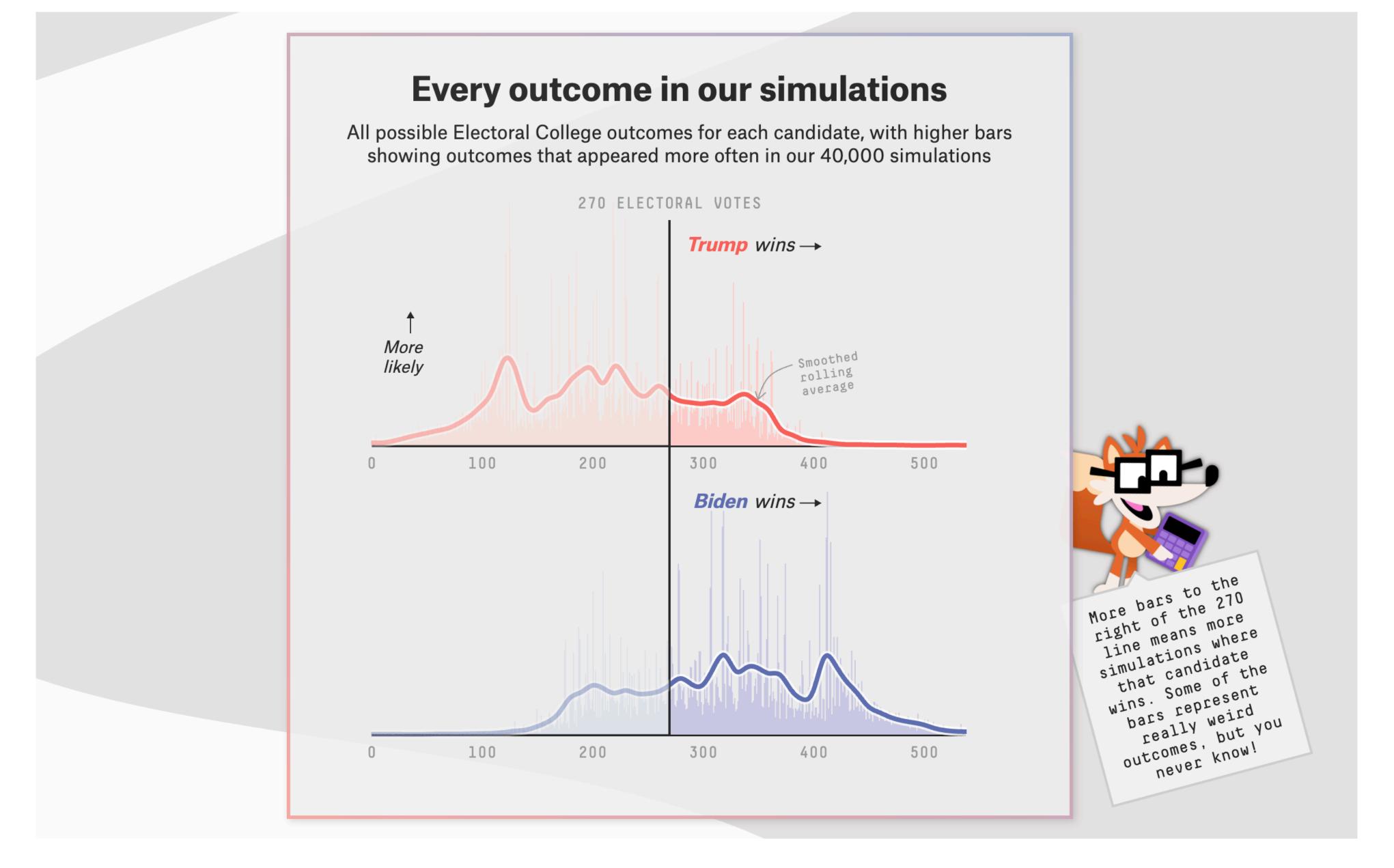


#### "Design secrets behind the FT's best charts of the year" https://www.ft.com/content/4743ce96-e4bf-11e7-97e2-916d4fbac0da

Kensington & Chelsea +9

"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/

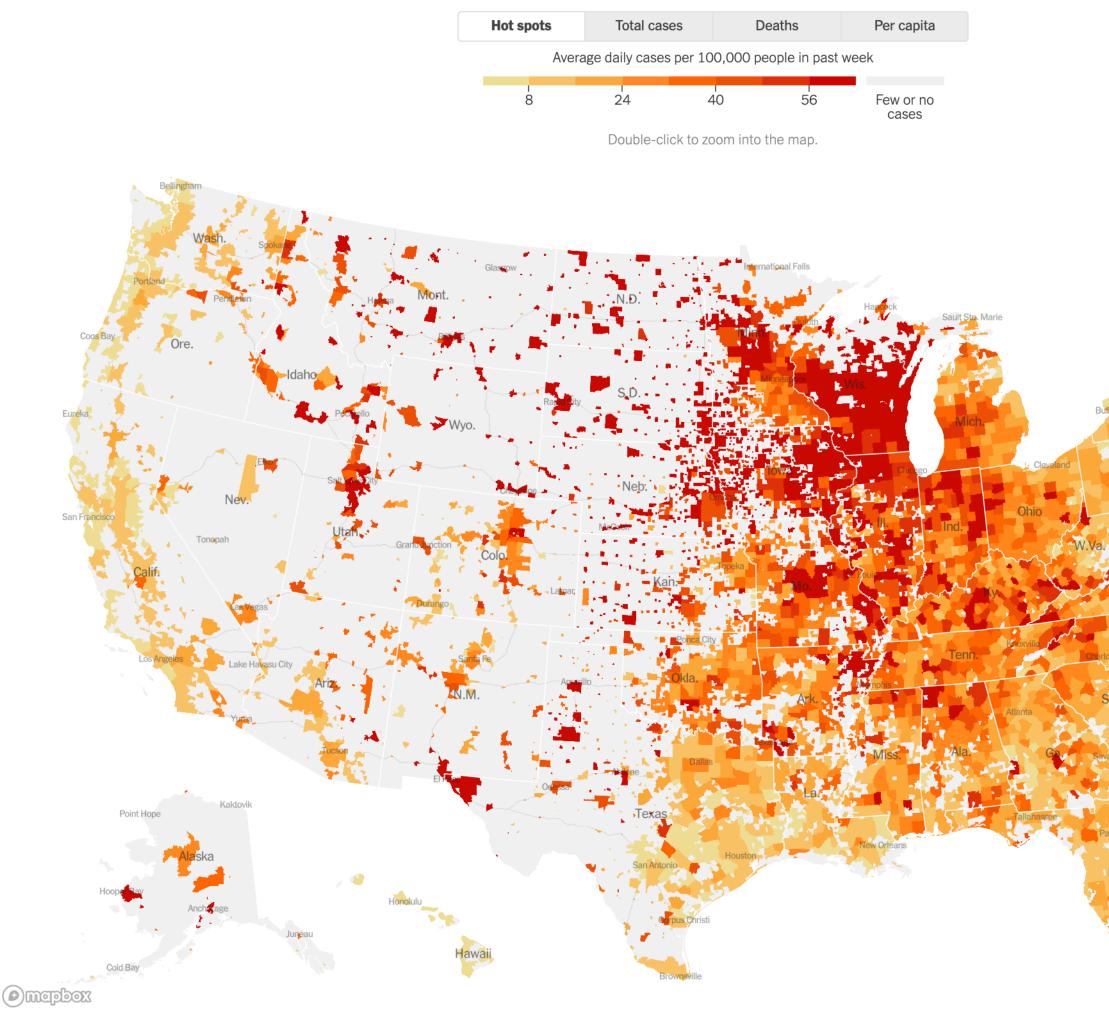
episode-155-john-burn-murdoch/



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

# 5. What 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



Sources: State and local health agencies. Population and demographic data from Census Bureau.

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

Alberto Cairo • University of Miami • <u>www.thefunctionalart.com</u> • Twitter: @albertocairo

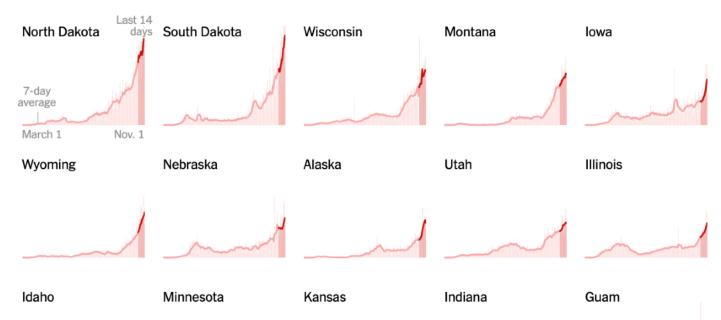


Day with data reporting anomaly.

Includes confirmed and probable cases where available. 14-day change trends use 7-day averages.

#### Where new cases are higher and staying high

States where new cases are higher had a daily average of at least 15 new cases per 100,000 people over the past week. Charts show daily cases per capita and are on the same scale. Tap a state to see detailed map page.



# Fully customized style:

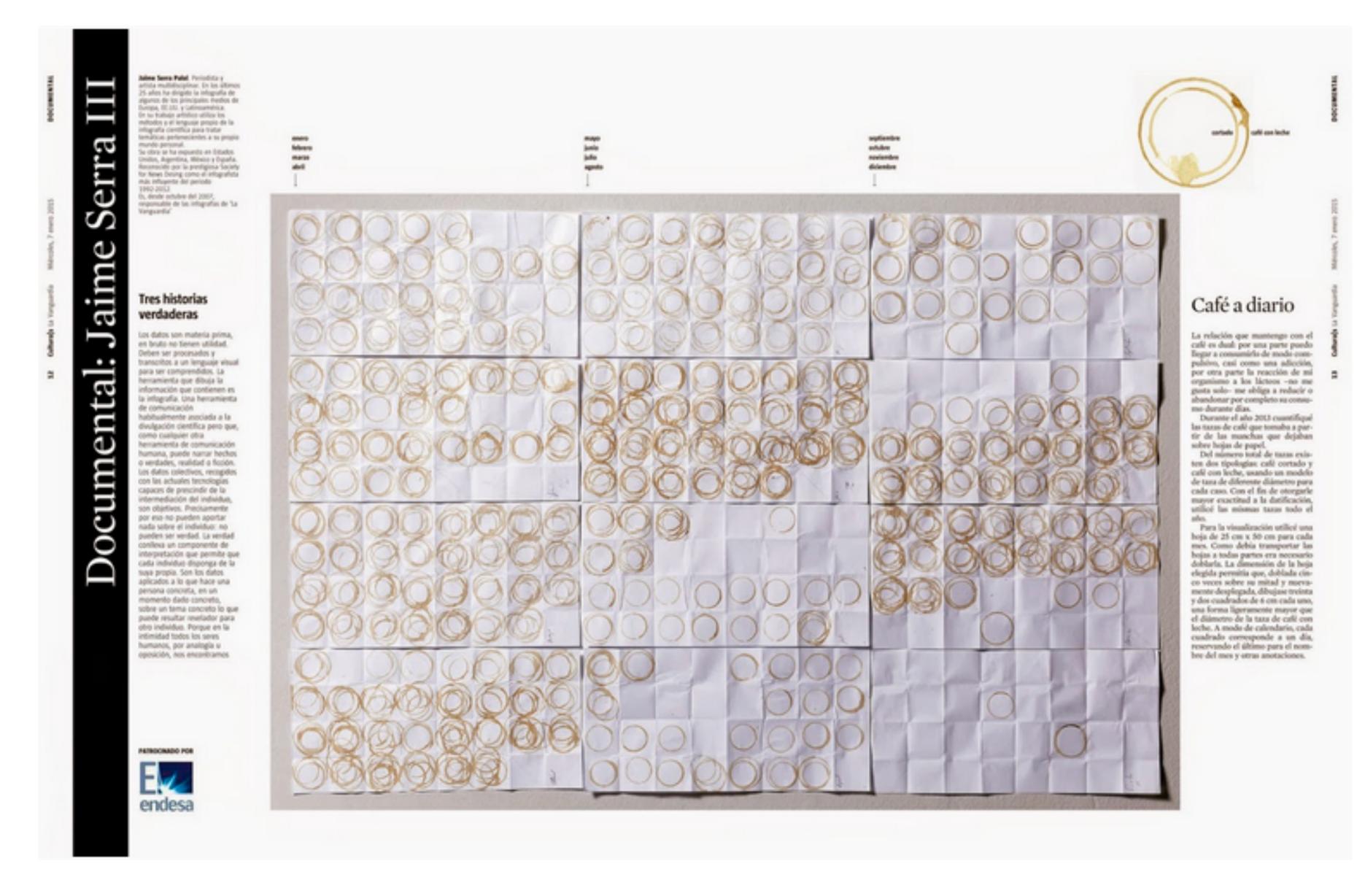


## https://jaimeserra-archivos.blogspot.com/

Alberto Cairo • University of Miami • <u>www.thefunctionalart.com</u> • Twitter: @albertocairo

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

### Fully customized style: Appropriate for one-time use when we want to provoke curiosity, surprise —or simply a smile



### Unusual or unique visualizations

**Google** News Initiative

#### NO EPICENTRO

## E se todos os mortos por Covid-19 no Brasil fossem seus vizinhos?

Descubra o que aconteceria com a sua vizinhança caso o epicentro da epidemia de Covid-19 no Brasil fosse sua casa.

Publicado em 24 de julho de 2020. Dados atualizados em 1 de novembro de 2020.

## http://piaui.folha.uol.com.br/lupa/epicentro/



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.



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

# The End.