

The Art of Insight

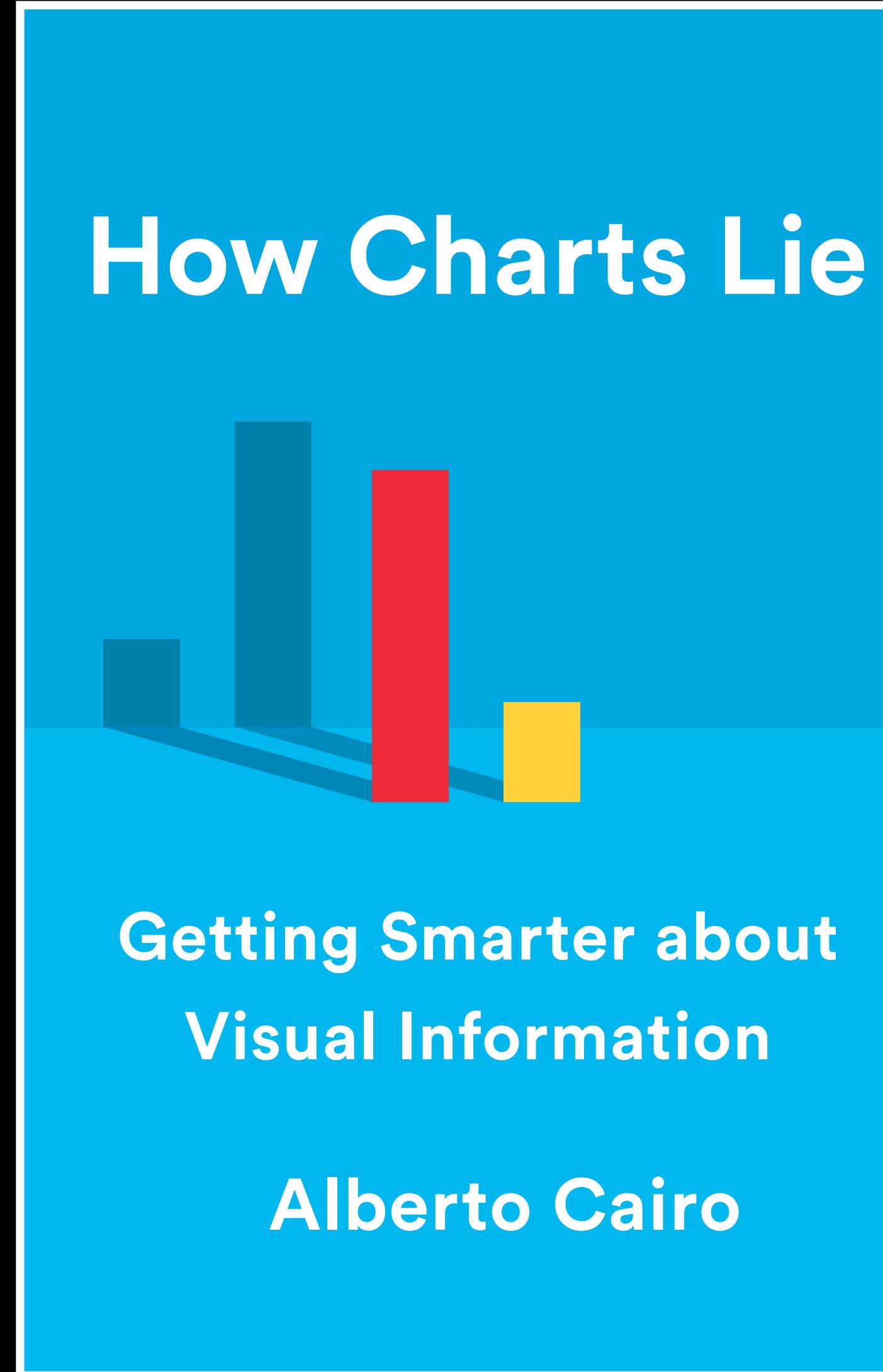
MAKING DECISIONS IN VISUALIZATION



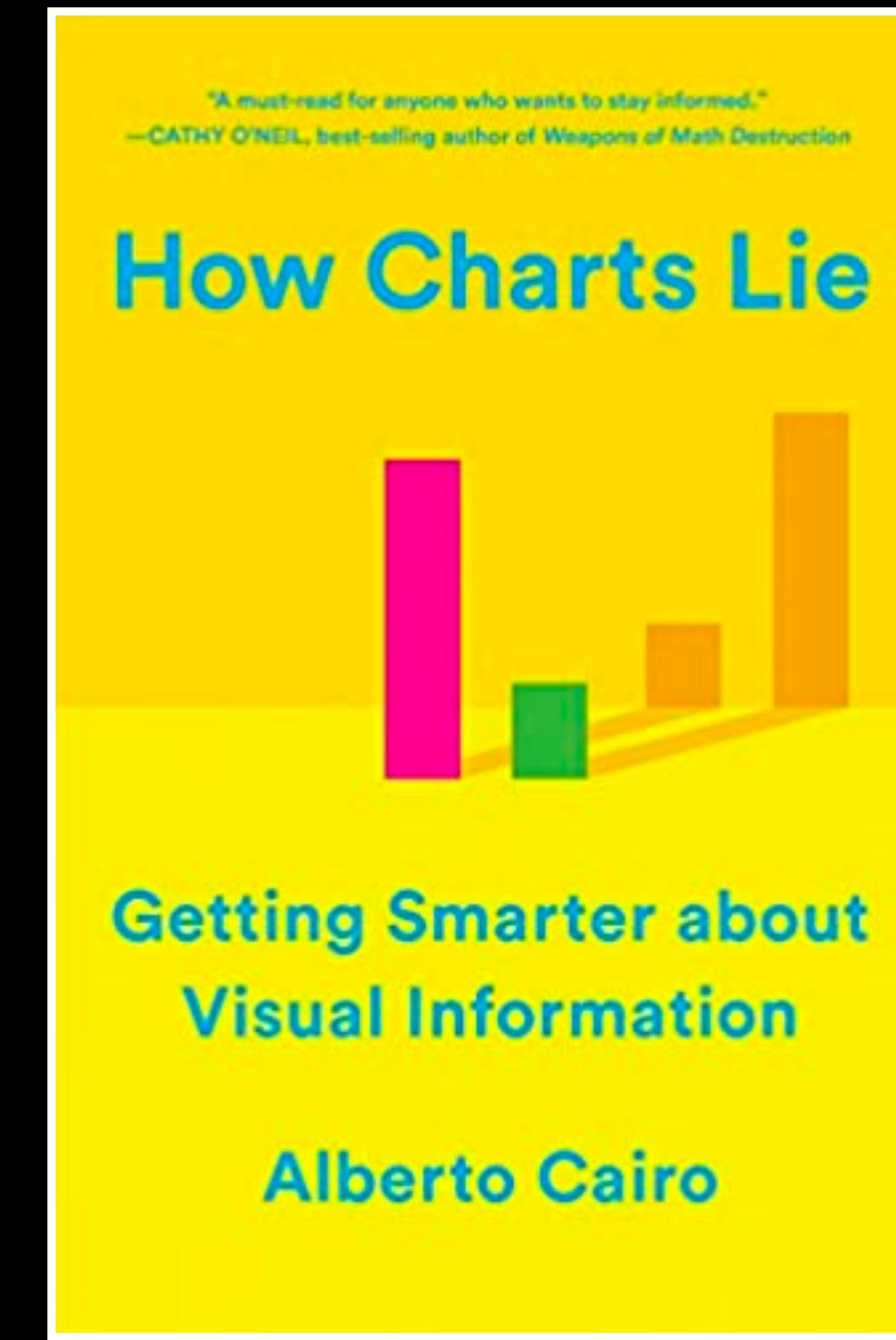
Alberto Cairo

NHS-R 2020

Primera edición
(Octubre 2019)



Edición en rústica
(Octubre 2020)



We are living through a Golden Age of visualization

More info

More info



Sales per countries



We are living through a Golden Age of visualization

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.

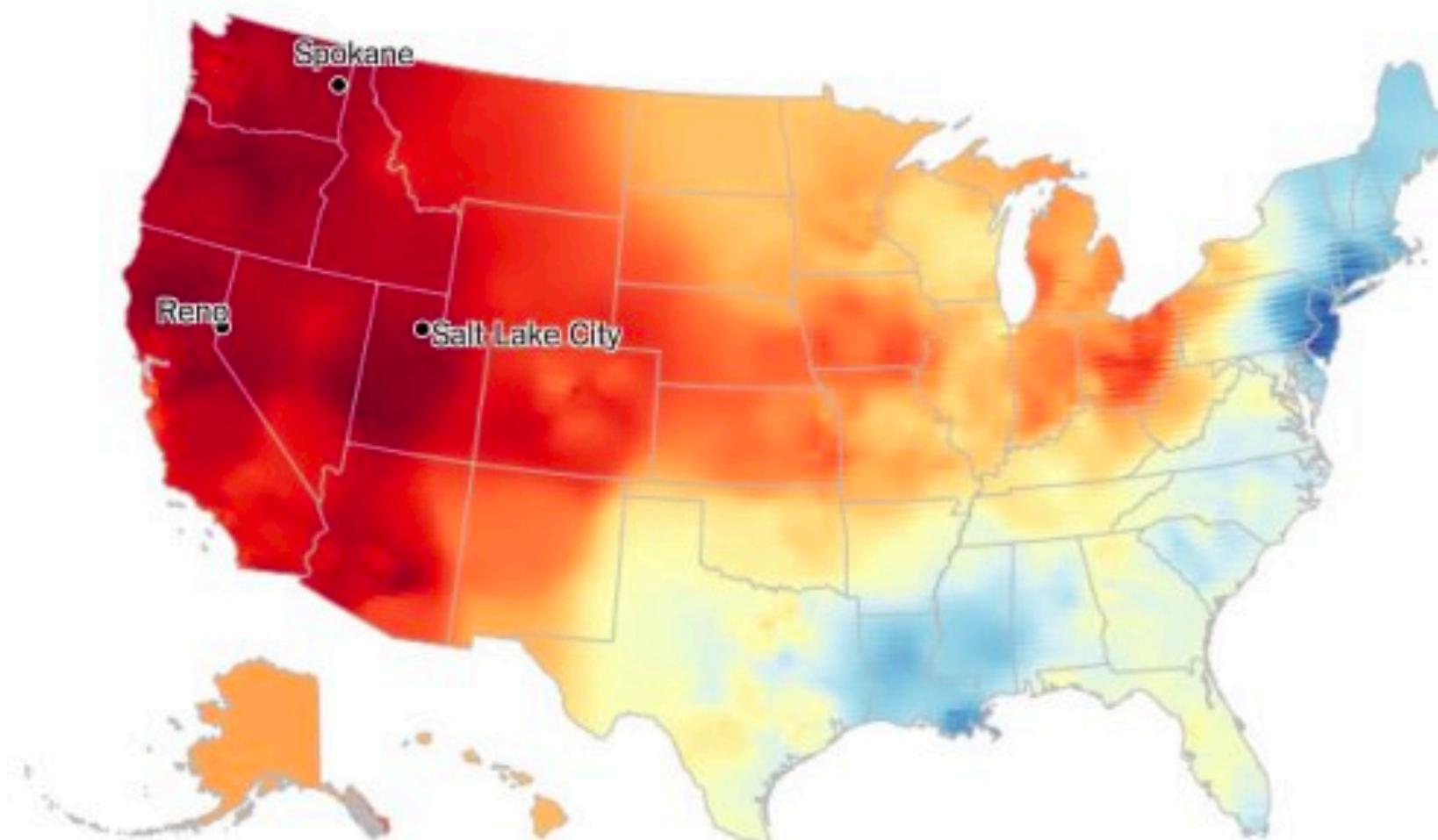
Your Map

See the pattern of your dialect in the map below. Three of the most similar cities are shown.

Least similar Most similar

Show least similar

SHARE YOUR MAP:

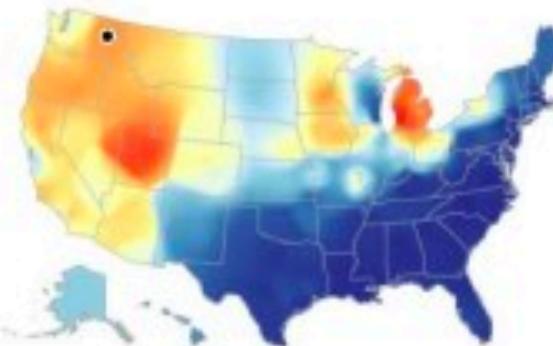


These maps show your most distinctive answer for each of these cities.

SALT LAKE CITY



SPOKANE



RENO



NYTimes' Dialect Quiz

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

We are living through a Golden Age of visualization

The Washington Post
Democracy Dies in Darkness

Sections 

Alberto Cairo To... 

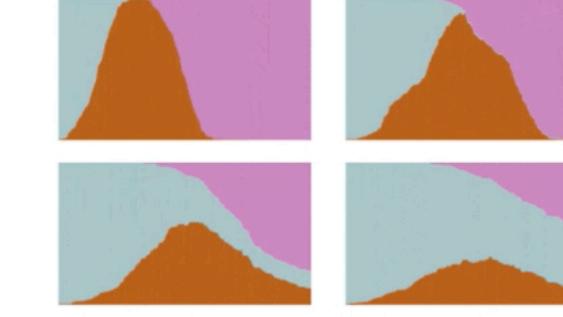








2.6k



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](#).

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

We are living through a Golden Age of visualization

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 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/>

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

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



Visuals are often ambiguous



Visuals are often ambiguous

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

1. “A picture is worth a thousand words”
2. “Visualization is intuitive”
3. “The data should speak for itself”
4. “Show, don’t tell!”
5. “Learn the rules of visualization (by reading
this book by [insert author’s name here])!”

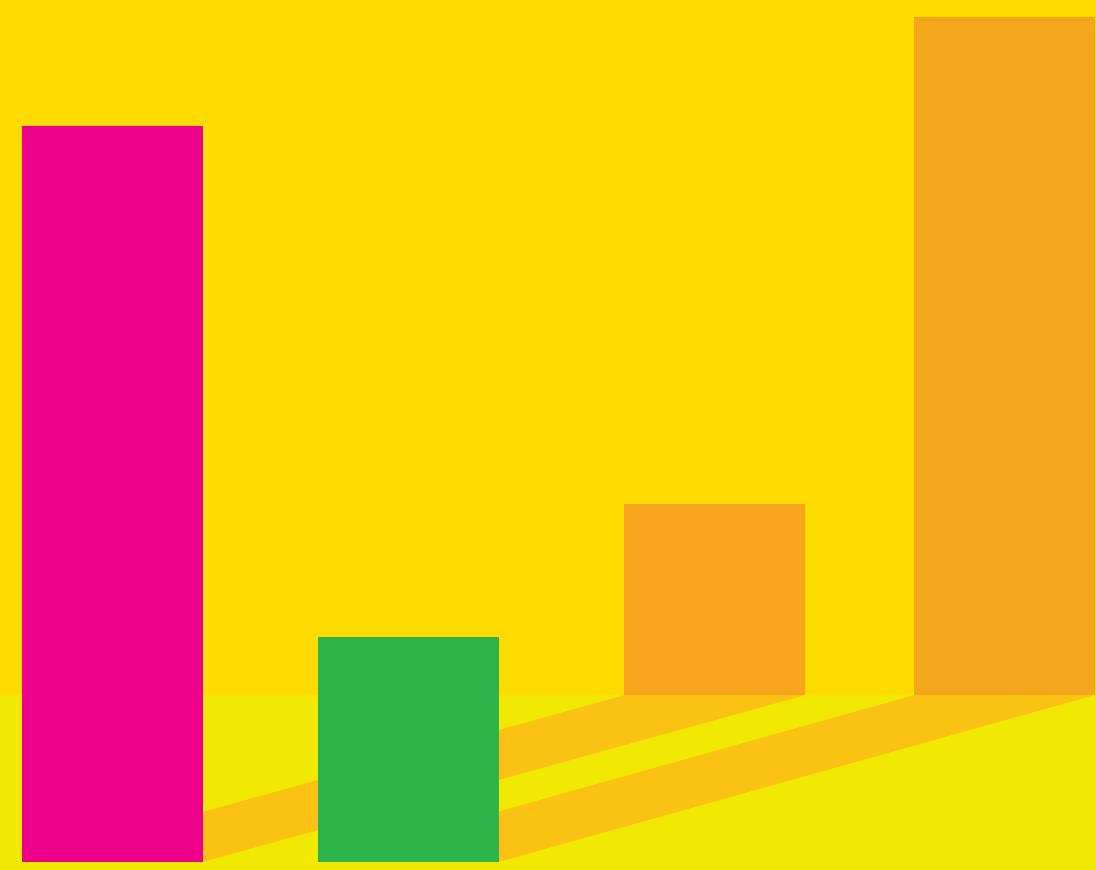
Visualizations can't be designed based just on our personal preferences—although these are important.

Visualization is similar to 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*

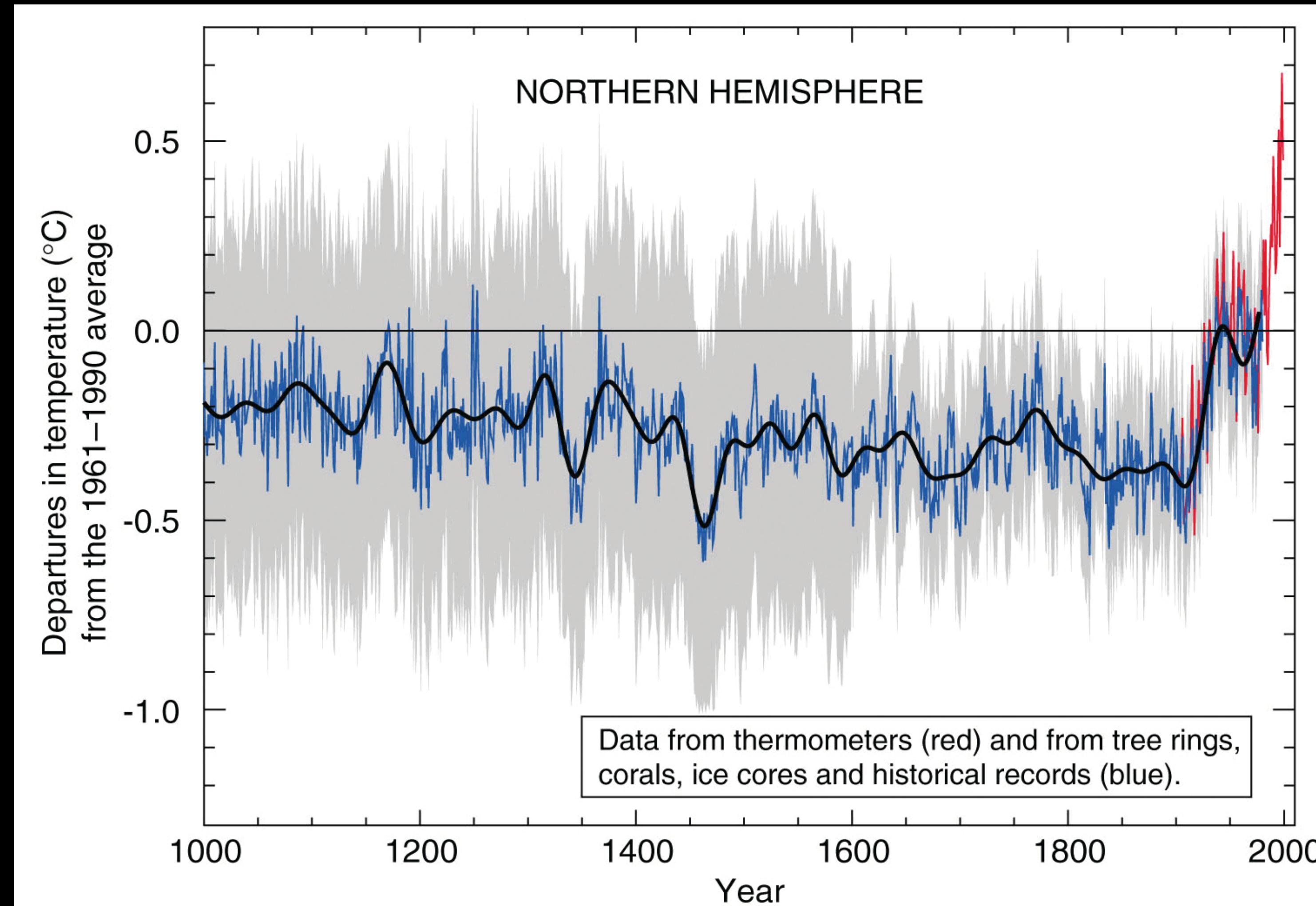


I. Why should my visualization exist?

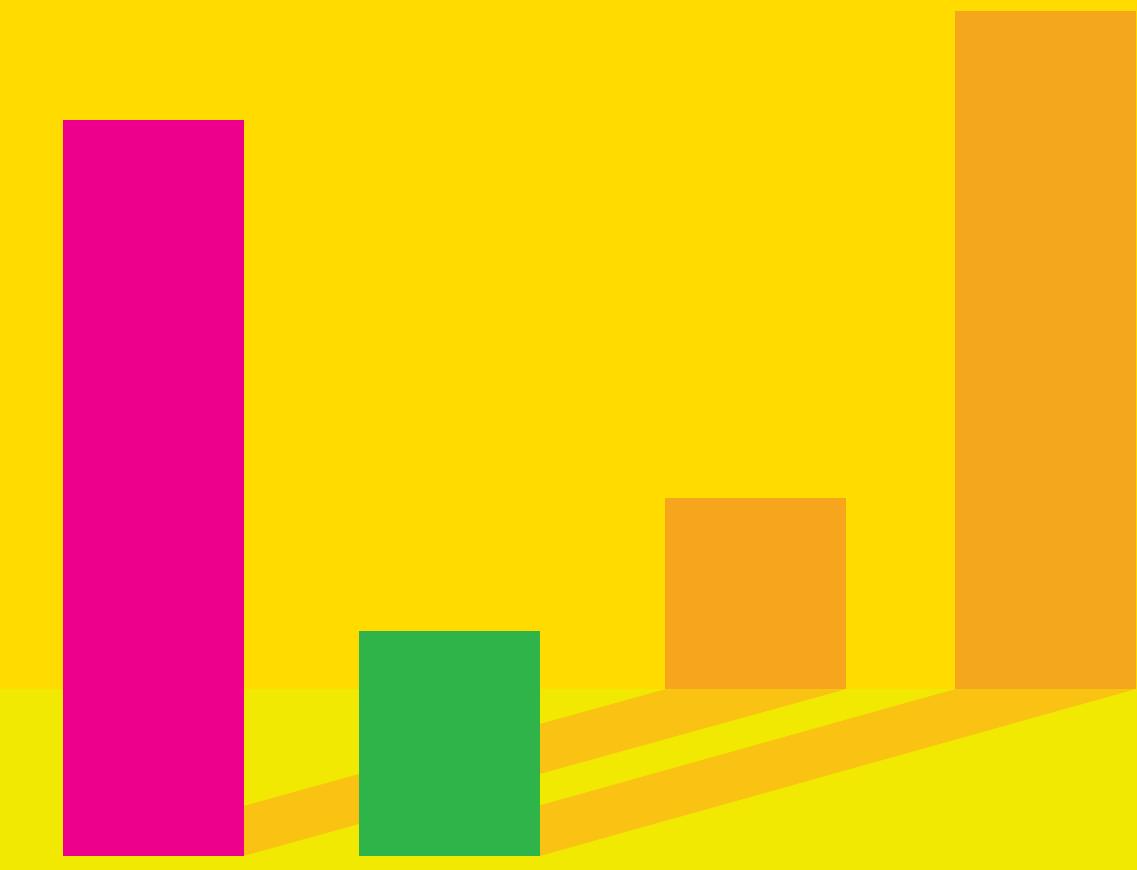
Detecting patterns

	A	B	C	D	E	F	G		A	B	C	D	E	F	G	H
1	YEAR	TEMP	YEAR	1 SIGMA	2 SIGMA			878	1876	-0.1891	1876	0.113228	0.226456	8.25297E-02	7.75207E-02	
2	1000	0.0659	1000	0.240346	0.480693	0.206137	0.123588	879	1877	-0.0140	1877	0.113228	0.226457	8.25299E-02	7.75209E-02	
3	1001	-0.1241	1001	0.240347	0.480694	0.206137	0.123589	880	1878	-0.0873	1878	0.113228	0.226457	8.25298E-02	7.75209E-02	
4	1002	-0.1208	1002	0.240346	0.480692	0.206136	0.123588	881	1879	-0.2959	1879	0.113229	0.226458	8.25302E-02	7.75212E-02	
5	1003	-0.1801	1003	0.240347	0.480694	0.206137	0.123589	882	1880	-0.2368	1880	0.113229	0.226457	8.25300E-02	7.75210E-02	
6	1004	-0.0711	1004	0.240347	0.480693	0.206137	0.123588	883	1881	-0.1977	1881	0.113229	0.226458	8.25302E-02	7.75212E-02	
7	1005	-0.1334	1005	0.240346	0.480692	0.206136	0.123588	884	1882	-0.2036	1882	0.113229	0.226457	8.25300E-02	7.75210E-02	
8	1006	-0.0644	1006	0.240346	0.480693	0.206137	0.123588	885	1883	-0.2489	1883	0.113228	0.226455	8.25293E-02	7.75204E-02	
9	1007	0.0042	1007	0.240347	0.480693	0.206137	0.123588	886	1884	-0.2125	1884	0.113229	0.226457	8.25301E-02	7.75211E-02	
10	1008	-0.1288	1008	0.240347	0.480693	0.206137	0.123588	887	1885	-0.1896	1885	0.113228	0.226457	8.25299E-02	7.75210E-02	
11	1009	-0.0296	1009	0.240347	0.480693	0.206137	0.123588	888	1886	-0.1084	1886	0.113228	0.226456	8.25298E-02	7.75208E-02	
12	1010	0.1187	1010	0.240347	0.480694	0.206137	0.123589	889	1887	-0.3265	1887	0.113228	0.226456	8.25296E-02	7.75206E-02	
13	1011	-0.1252	1011	0.240346	0.480692	0.206136	0.123588	890	1888	-0.1694	1888	0.113228	0.226457	8.25298E-02	7.75209E-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	7.75208E-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	7.75211E-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	7.75210E-02	
17	1015	-0.1146	1015	0.240346	0.480692	0.206136	0.123588	894	1892	-0.3186	1892	0.113228	0.226456	8.25295E-02	7.75205E-02	
18	1016	-0.1206	1016	0.240346	0.480692	0.206136	0.123588	895	1893	-0.3236	1893	0.113228	0.226456	8.25297E-02	7.75207E-02	
19	1017	-0.0815	1017	0.240347	0.480693	0.206137	0.123588	896	1894	-0.1970	1894	0.113228	0.226456	8.25295E-02	7.75205E-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	7.75207E-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	7.75208E-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	7.75208E-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	7.75211E-02	
24	1022	-0.0743	1022	0.240347	0.480693	0.206137	0.123588	901	1899	-0.3486	1899	0.113228	0.226456	8.25297E-02	7.75207E-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	7.75210E-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	7.75206E-02	

Detecting patterns



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



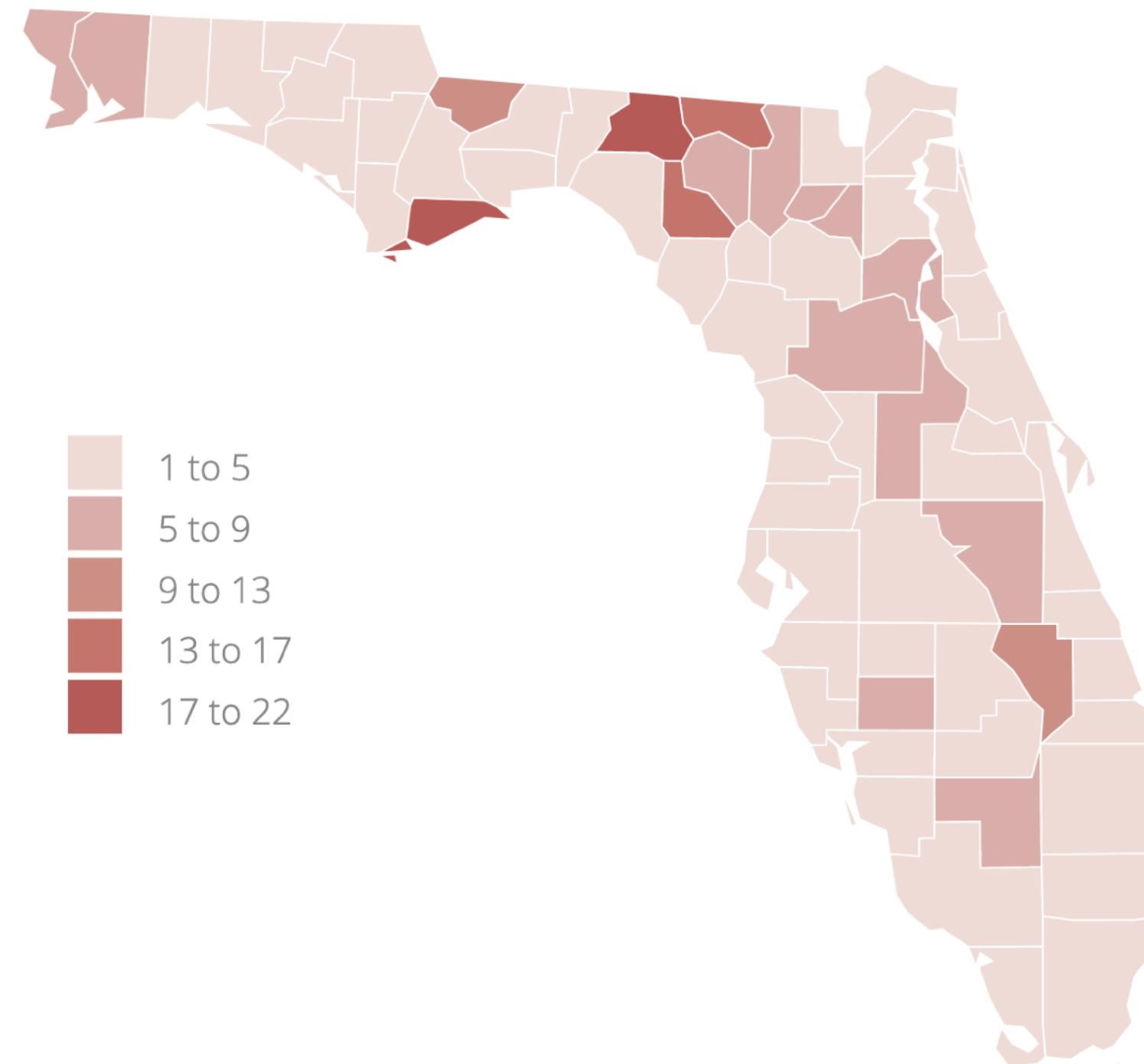
2. What to visualize?

AT SCHOOL WITHOUT A ROOF

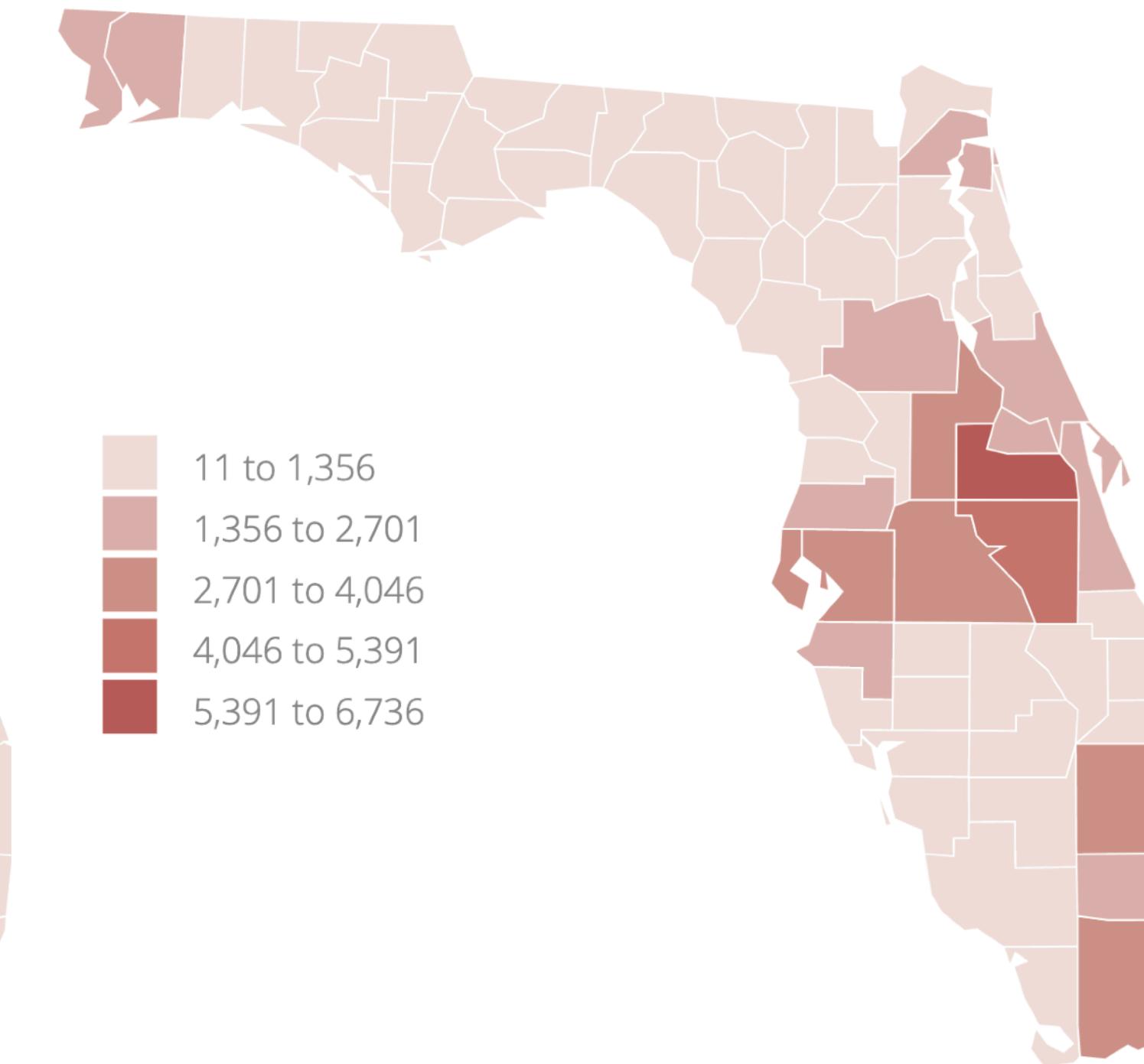
[http://
www.lmelgar.me/
without-a-roof/](http://www.lmelgar.me/without-a-roof/)

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.

Percentage Total



Percentage Total



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

Barcelona - 19 DIC 2014 - 15:51 EST

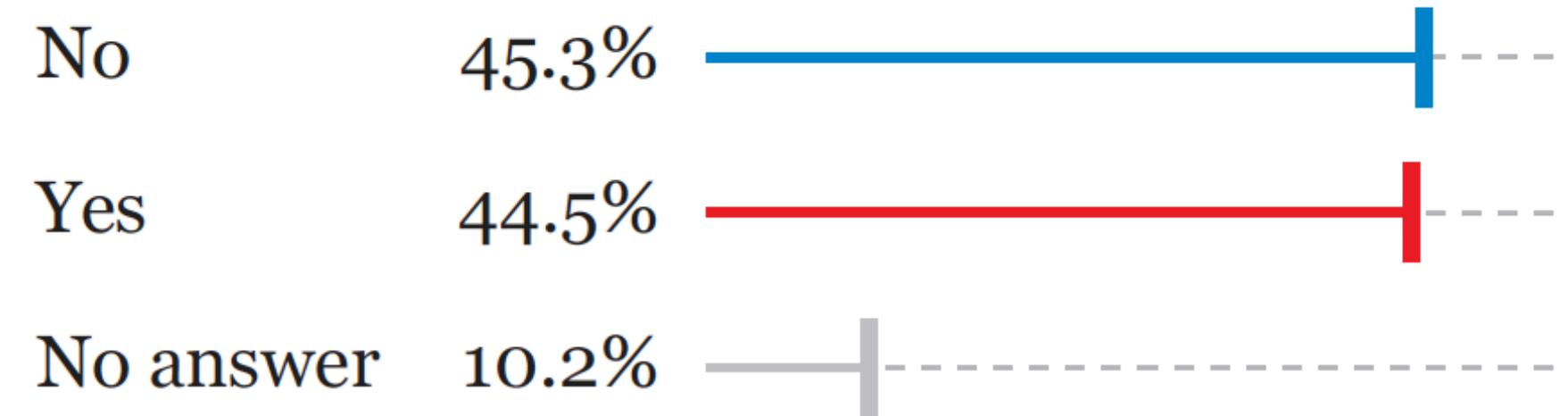


https://elpais.com/ccaa/2014/12/19/catalunya/1418984873_128596.html

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.

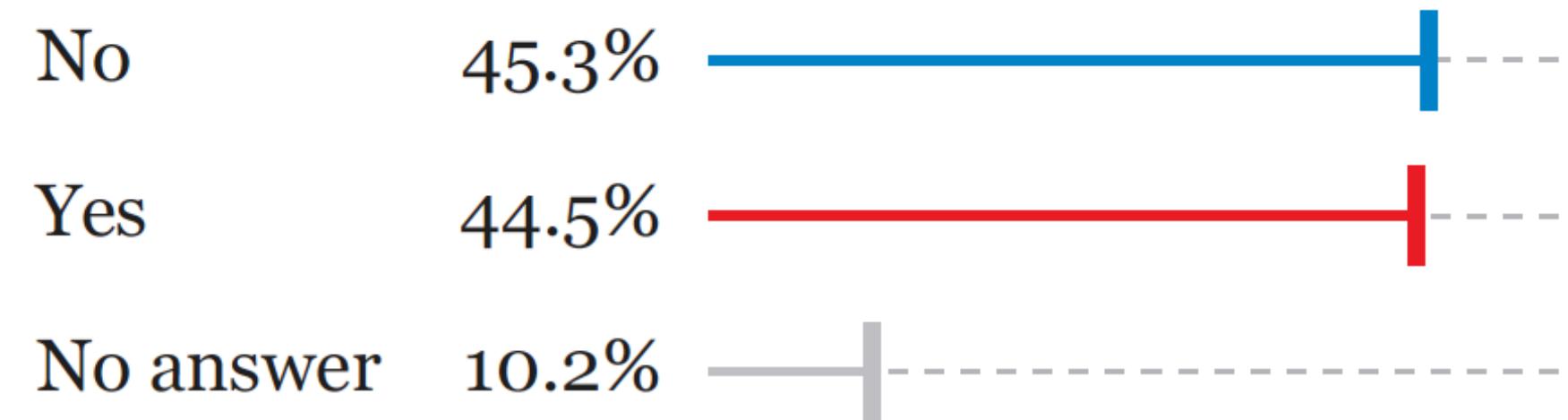
*Do you want Catalonia
to become an independent state?*



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?*



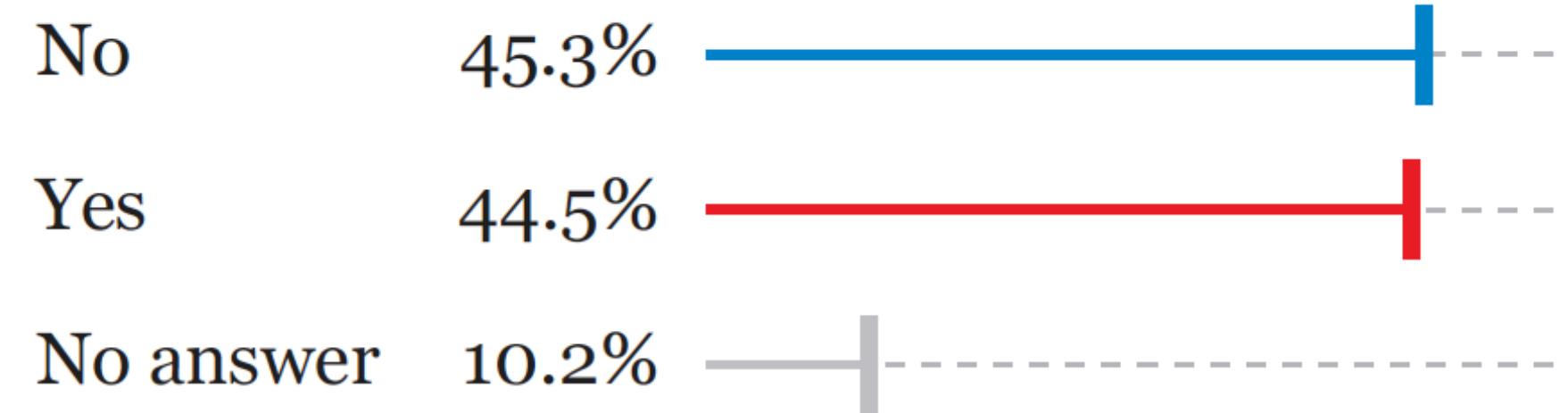
Margin of error: +/-2.95 at 95% confidence level

“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.

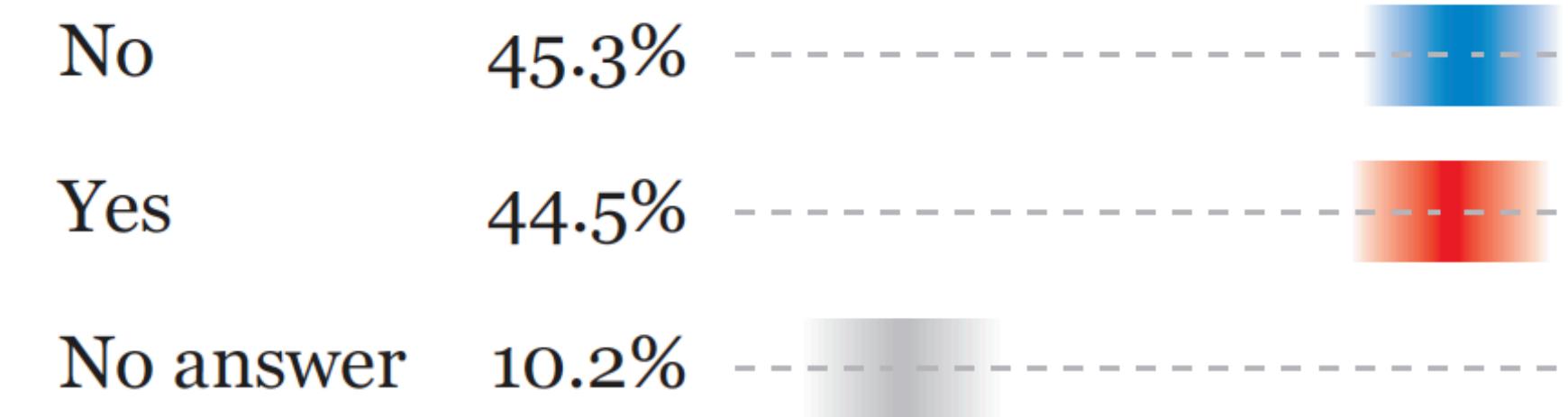
The latest poll by the Catalan executive's Opinion Studies Center (CEO) shows that 45.3 percent of citizens would vote no to the question: "Would you like Catalonia to become an independent state?" compared with 44.5 percent who would support the move.

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



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

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



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

"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"



3. How to explain it?



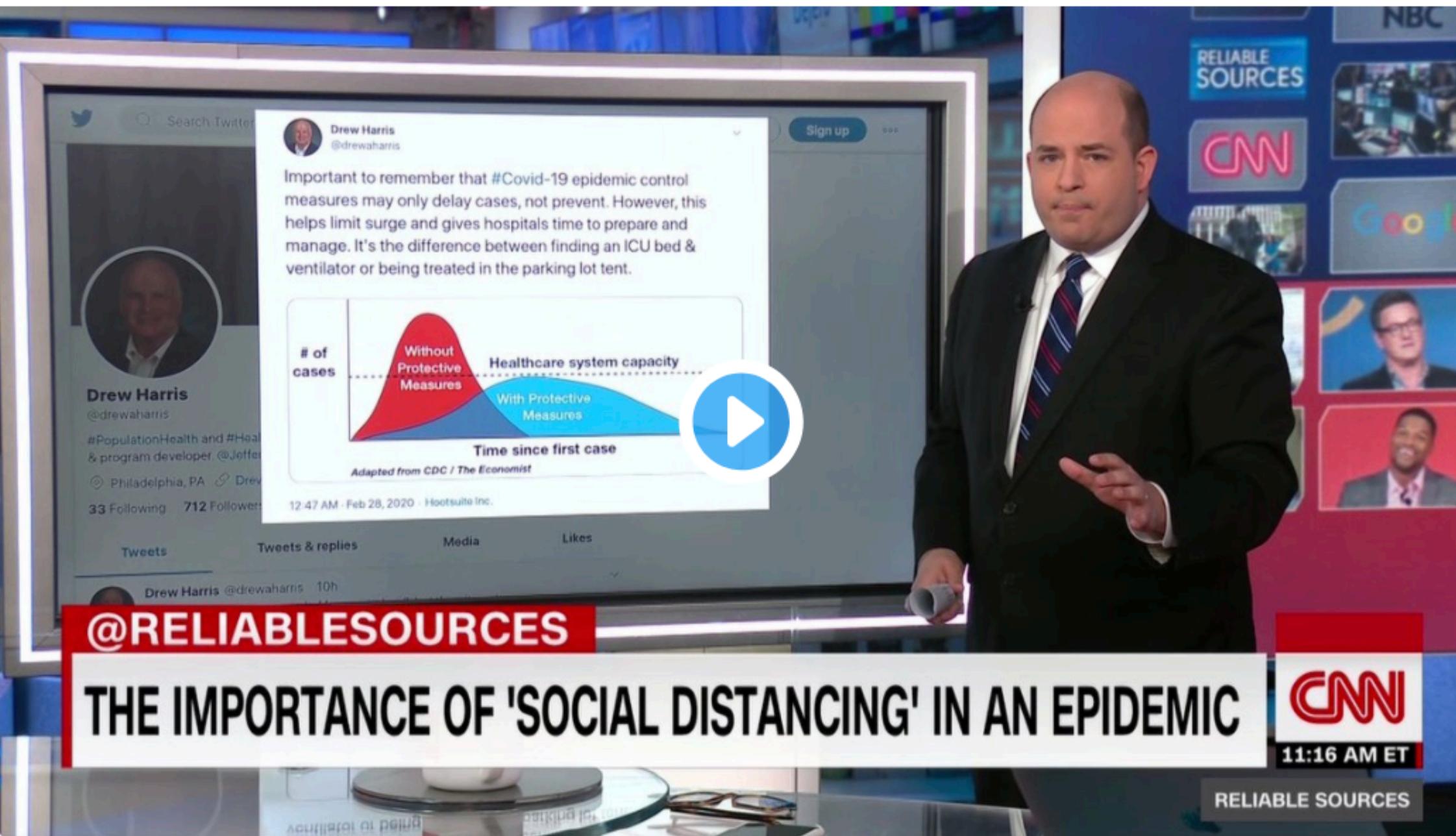
Alberto Cairo 
@AlbertoCairo



Good journalism isn't just showing charts. It's also about explaining them: [twitter.com/brianstelter/s...](https://twitter.com/brianstelter/status/123673377865658370)

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

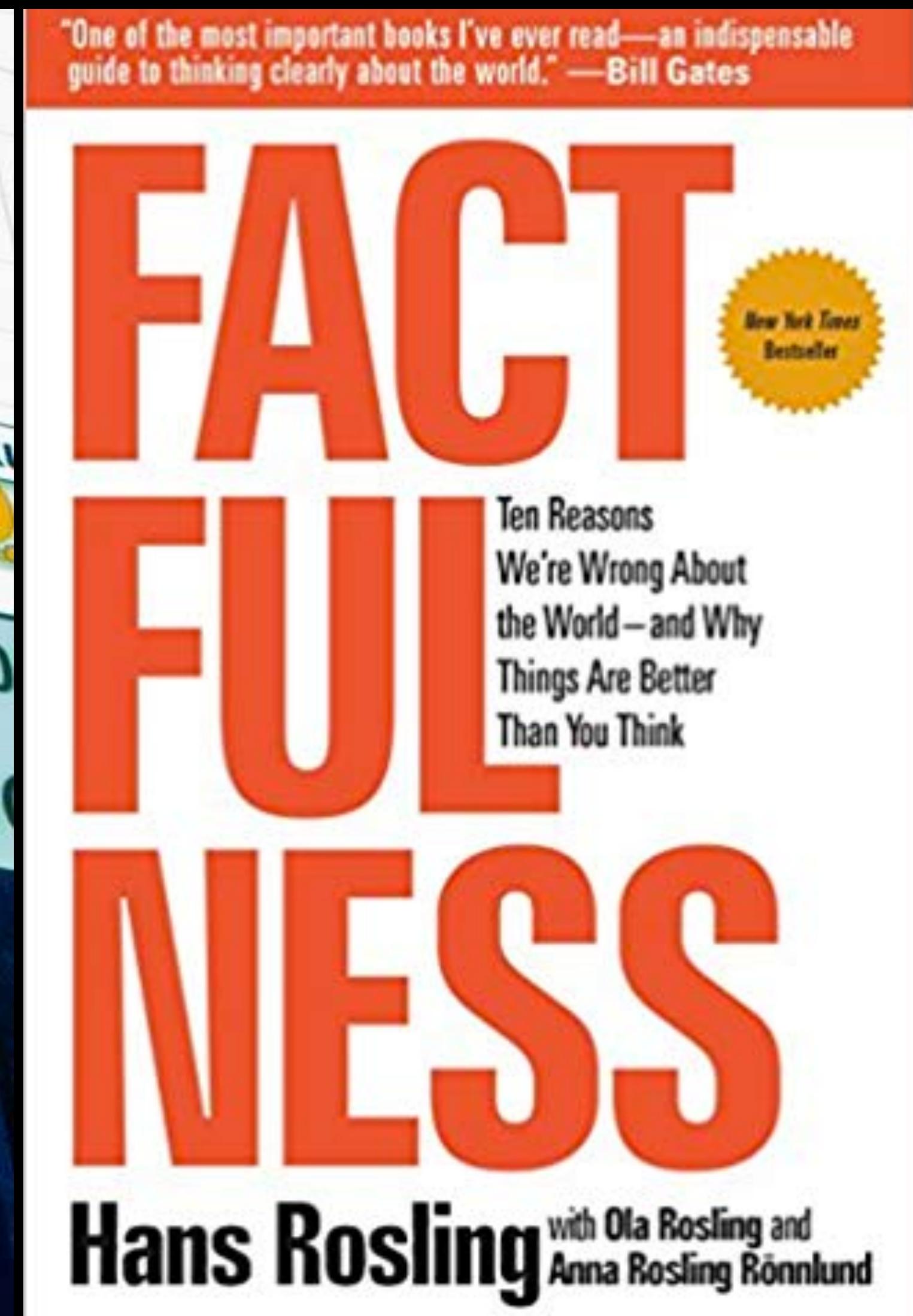
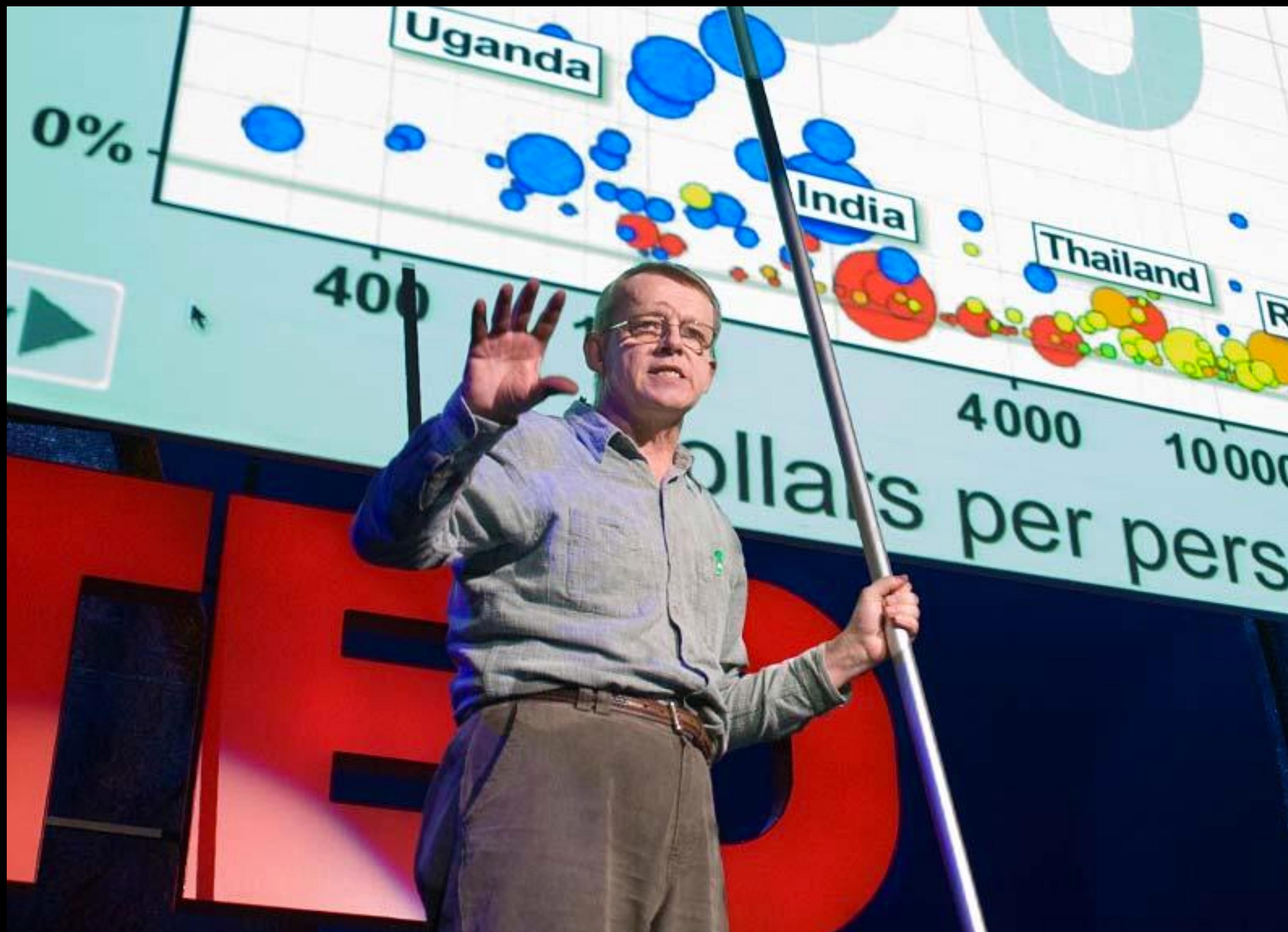


♡ 227 5:58 PM - Mar 8, 2020



<https://twitter.com/AlbertoCairo/status/123673377865658370>

Show AND tell



Hans Rosling, www.gapminder.org

BBC FOUR

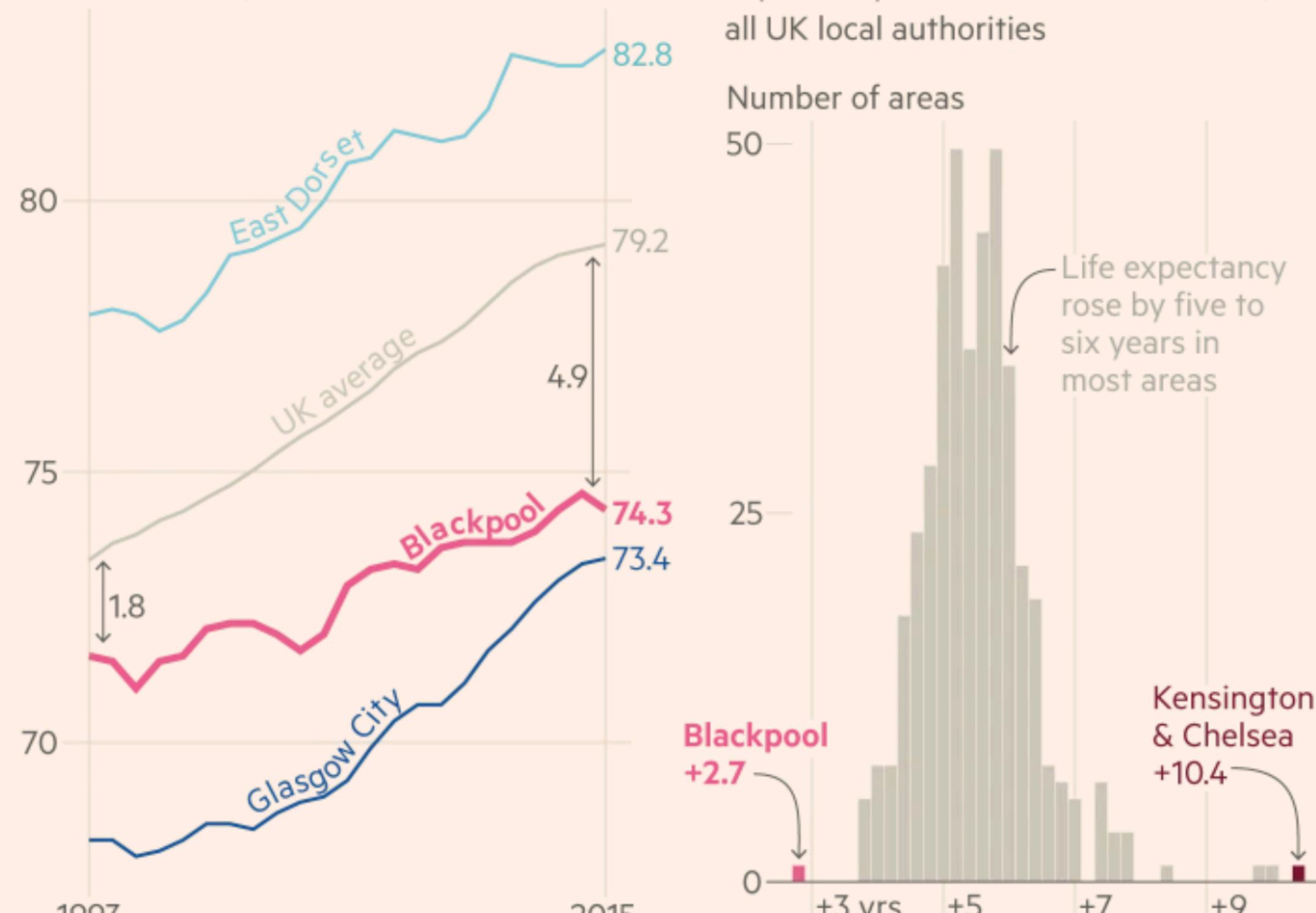
Show AND tell



Hans Rosling, *The Joy of Stats*

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

Male life expectancy at birth in selected local authorities, 1993-2015



Source: ONS

Graphic by John Burn-Murdoch / @jburnmurdoch

© FT

“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/>

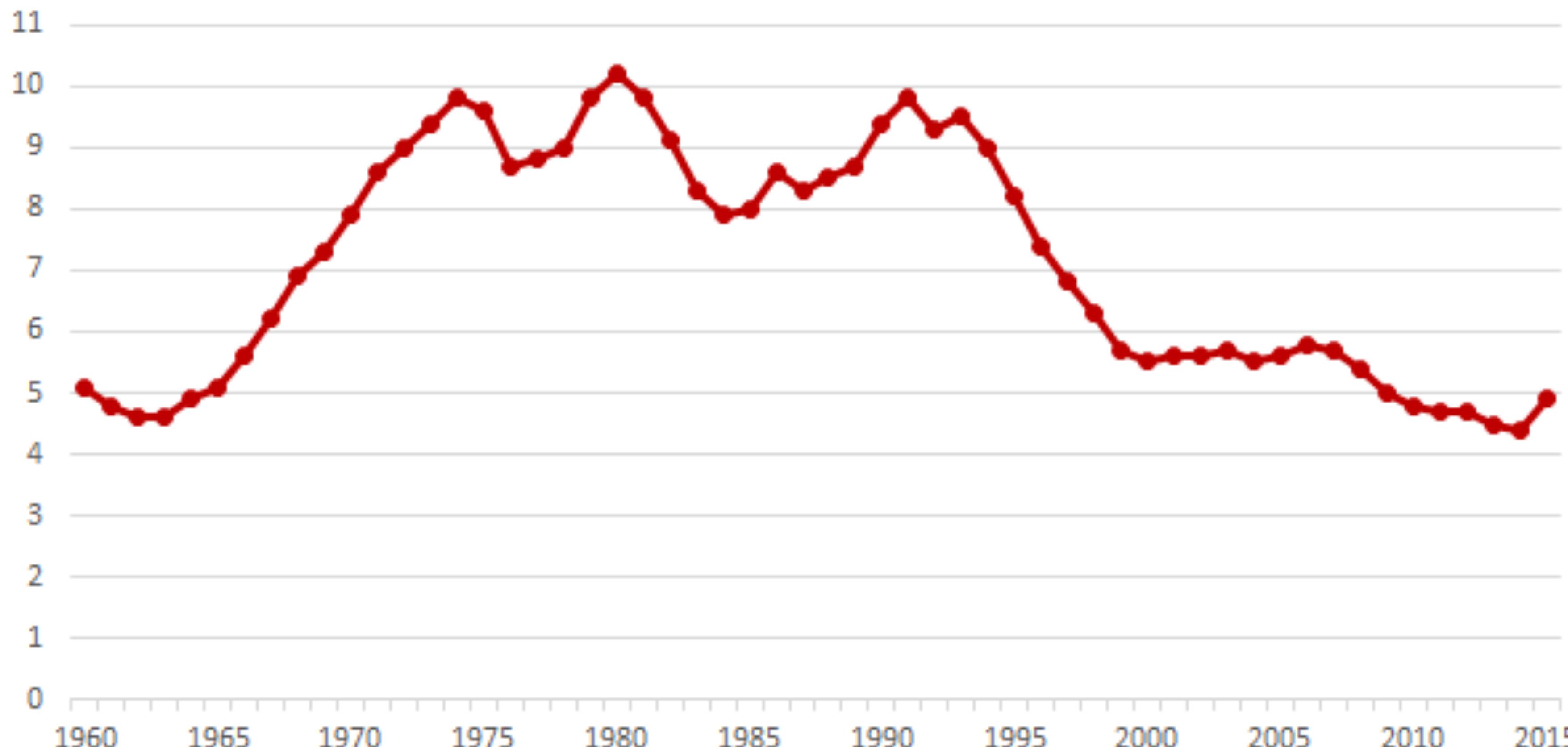
“Design secrets behind the FT’s best charts of the year”

<https://www.ft.com/content/4743ce96-e4bf-11e7-97e2-916d4fbac0da>



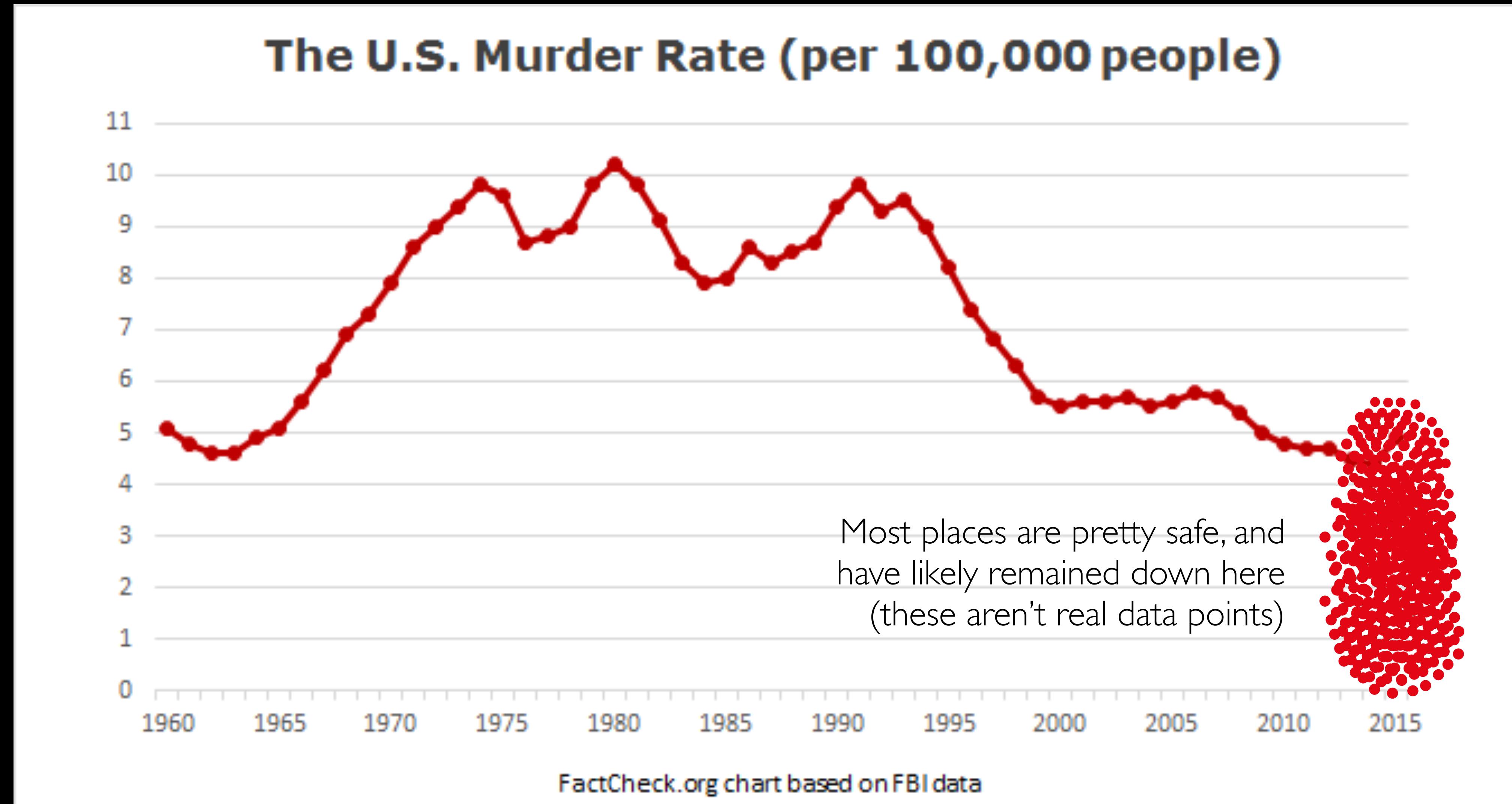
4. How much to visualize?

The U.S. Murder Rate (per 100,000 people)

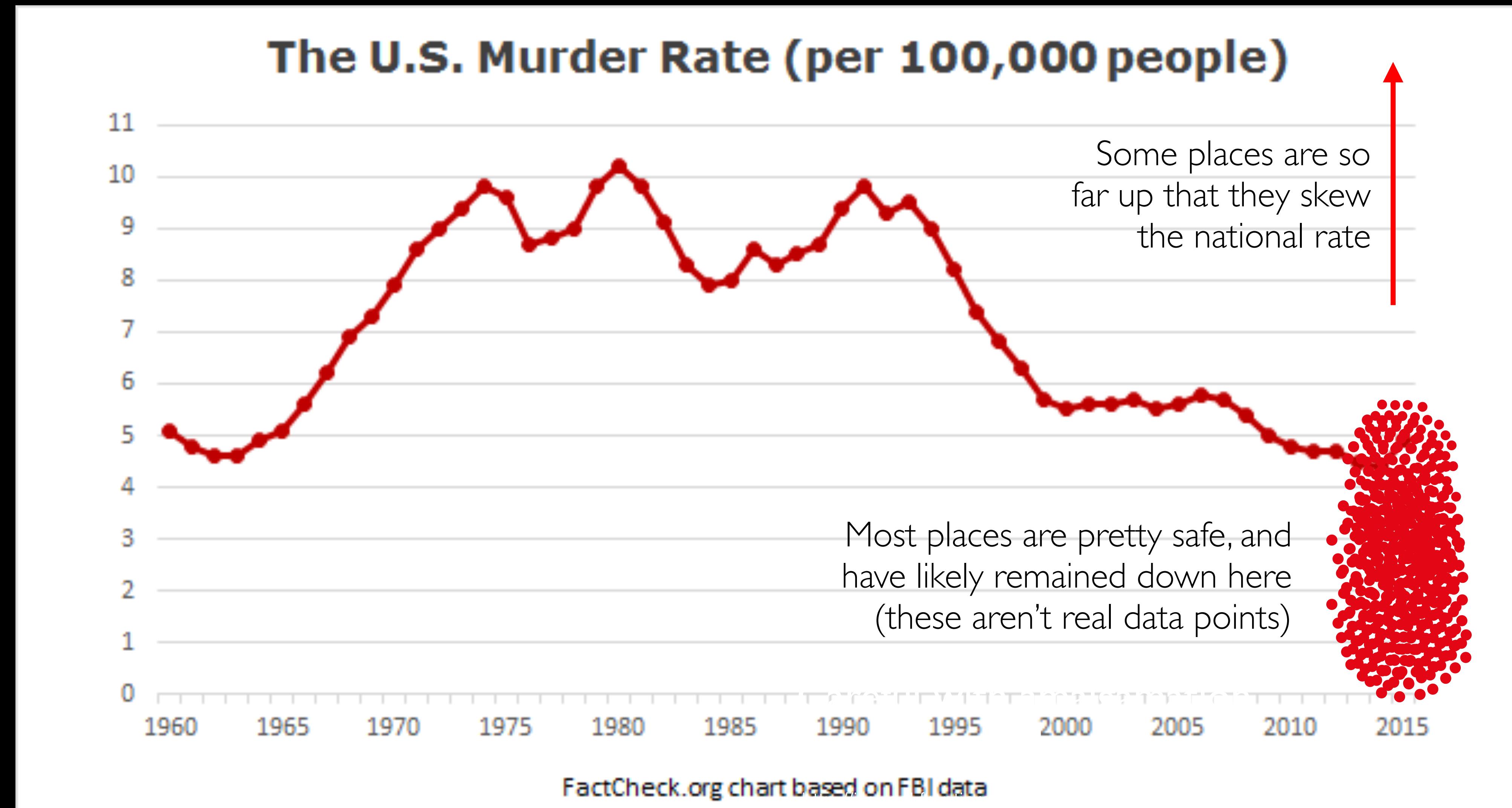


FactCheck.org chart based on FBI data

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?

Figure 2 - Main nationalities of arriving migrants – 2016

Greece

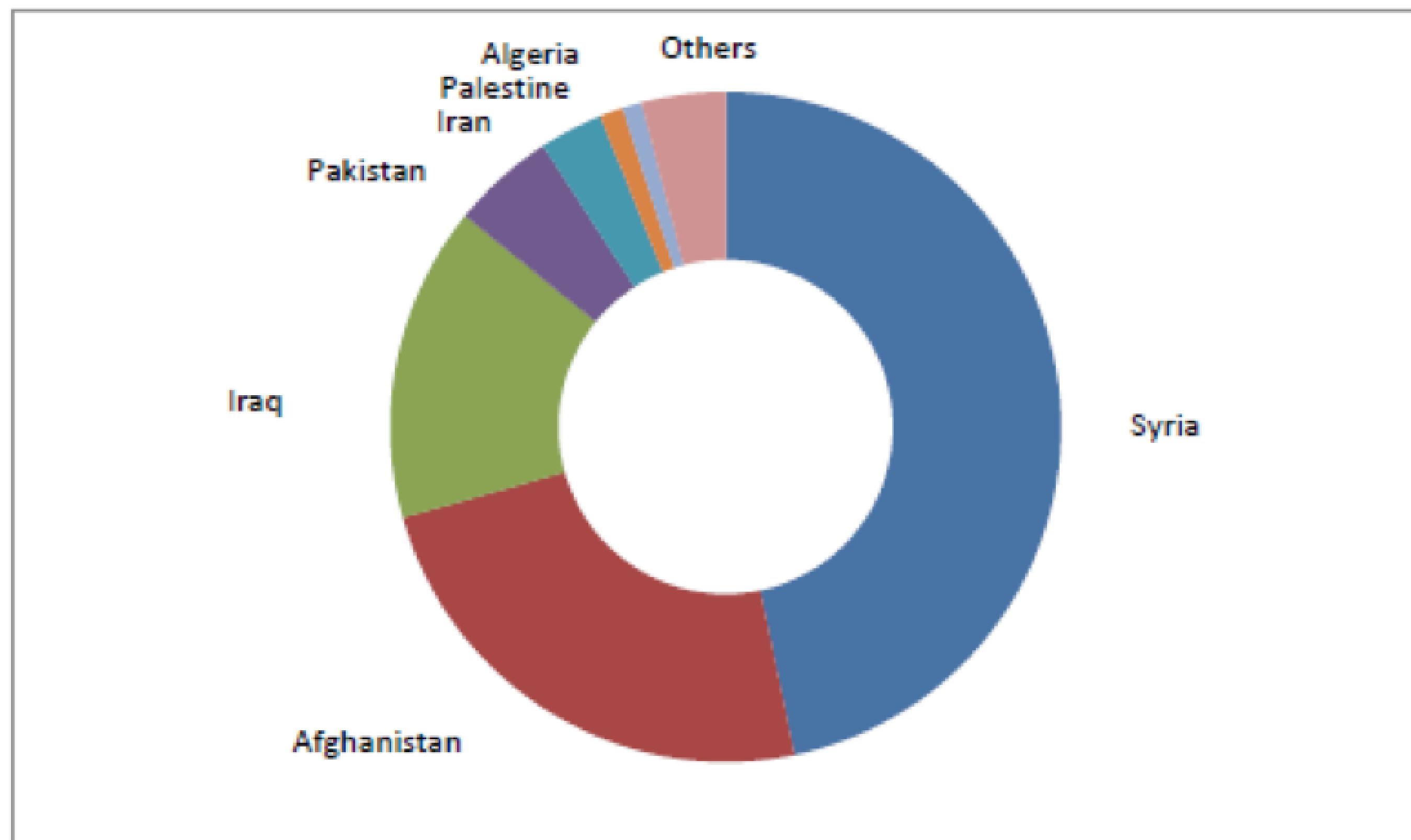
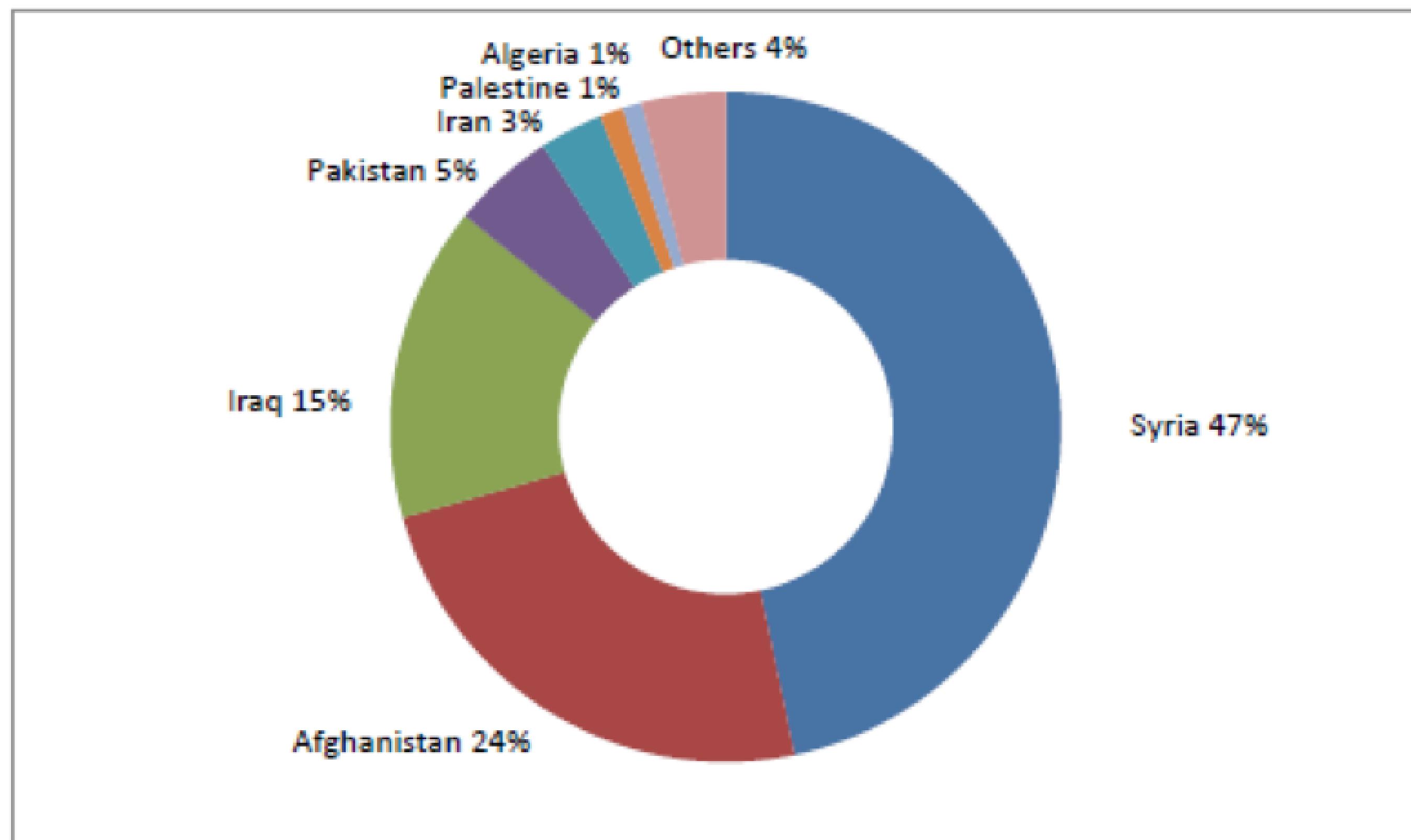
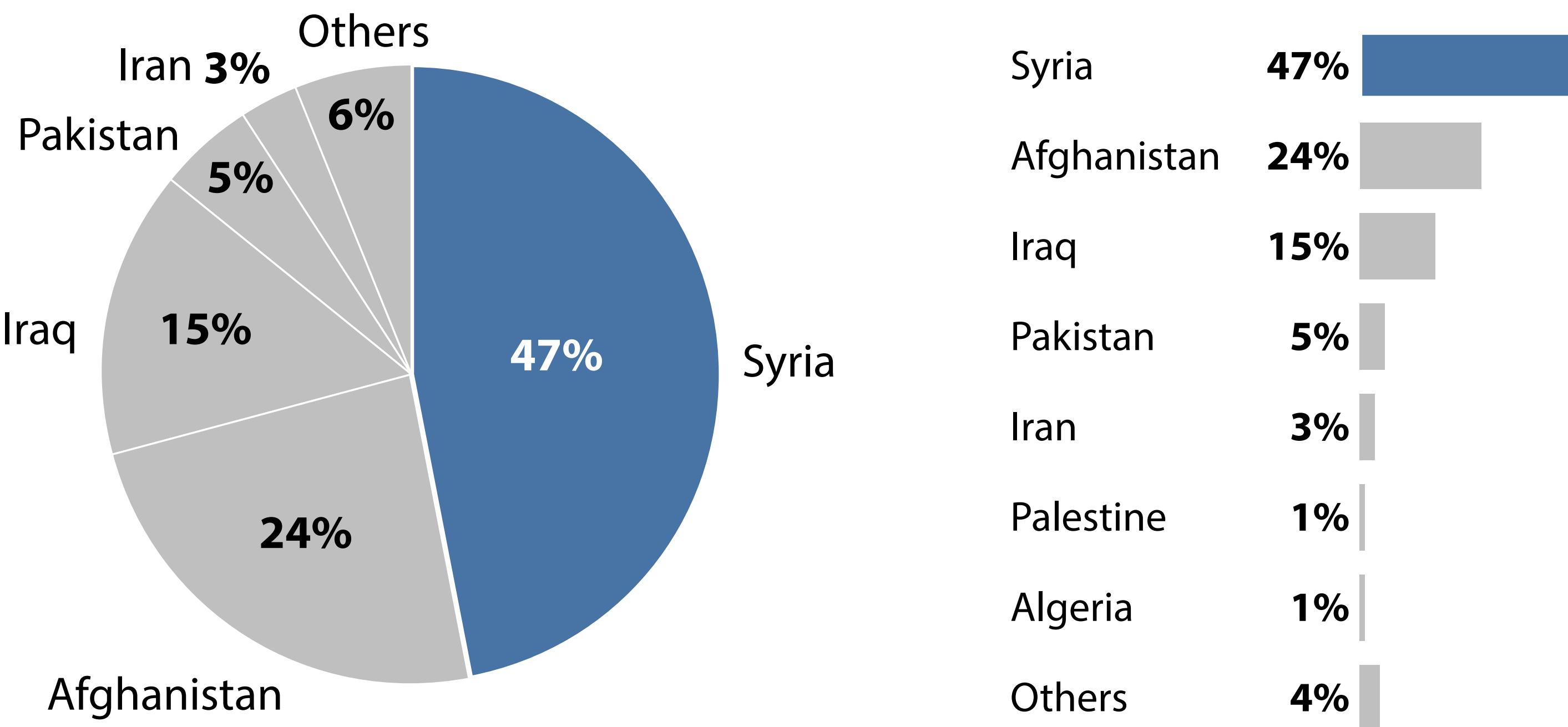
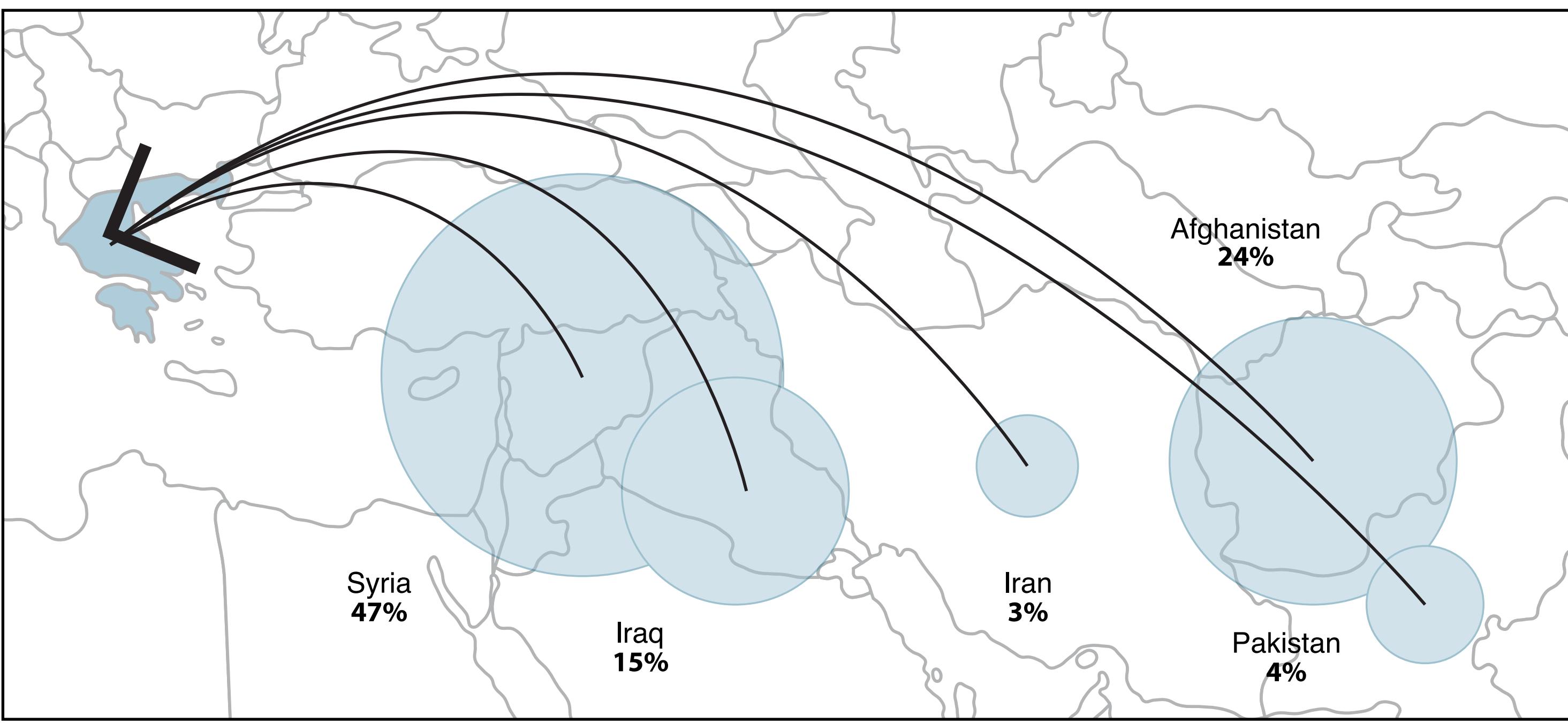


Figure 2 - Main nationalities of arriving migrants – 2016

Greece

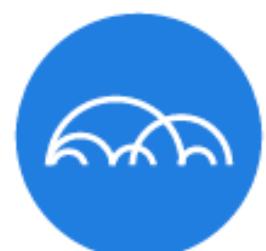




The Data Visualisation Catalogue

About • Suggest • Shop • Resources

Search by Function



Arc Diagram



Area Graph



Bar Chart



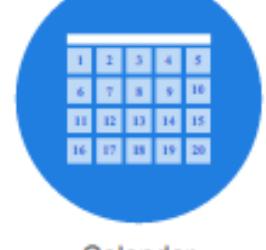
Box & Whisker Plot



Brainstorm



Bubble Chart



Calendar



Chord Diagram



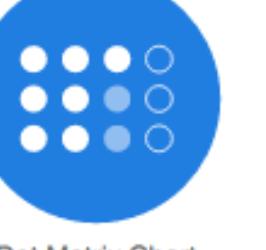
Choropleth Map



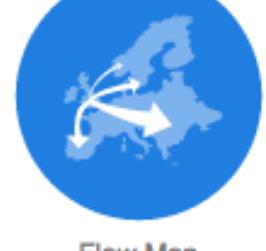
Circle Packing



Donut Chart



Dot Matrix Chart



Flow Map



Histogram

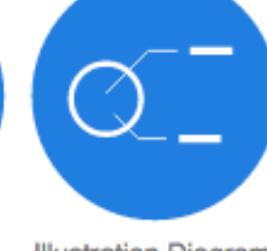
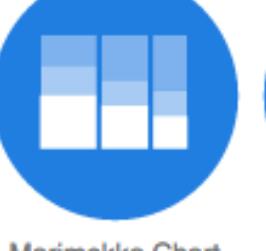


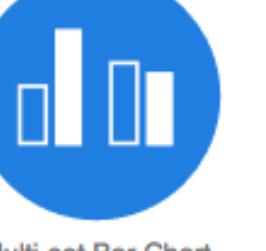
Illustration Diagram



Line Graph



Marimekko Chart



Multi-set Bar Chart



Nightingale Rose Chart



Non-ribbon Chord Diagram



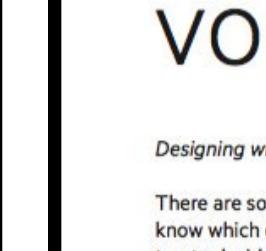
Parallel Sets



Pictogram Chart



Pie Chart



Proportional Area Chart



Radar Chart



Radial Bar Chart



Sankey Diagram



Scatterplot



Deviation

Emphasise variations (+/-) from a fixed reference point. Typically the reference point is zero but can also be a target or a long-term average. Can also be used to show sentiment (positive/negative/neutral).

Example FT uses Trade surplus/deficit, climate change

Diverging bar A simple standard bar chart that can handle both negative and positive magnitude values.

Diverging stacked bar Perfect for presenting survey results which involve sentiment (eg. disagree/agree).

Spine chart Spine chart is a single value chart with multiple components (eg. Male/Female).

Surplus/deficit filled line The shaded area of this chart allows a balance to be shown – either against a baseline or between two series.

Correlation

Show the relationship between two or more variables. Be mindful that unless you tell them otherwise, many readers will assume the relationships you show them to be causal (ie. one causes the other).

Example FT uses Inflation & unemployment, income & life expectancy

Scatterplot The standard way to show the relationship between two continuous variables, each with its own axis.

Line + column A good way of showing the relationship between an amount (columns) and a rate (line).

Connected scatterplot Usually used to show how a relationship between 2 variables has changed over time.

Bubble Like a standard bar chart with additional detail by sizing the circles according to a third variable.

XY heatmap A good way of placing the patterns between 2 categories of data, less good at showing fine differences in amounts.

Ranking

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

Example FT uses Wealth, deprivation, league tables, constituency election results

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

Violin plot Similar to a box plot but shows the full probability density function for complex distributions (data that cannot be summarised with a simple average).

Dot strip plot Similar to a pie chart but for showing the size and sex breakdown of a population distribution; effectively back-to-back histograms.

Dot plot Good for showing how ranks have changed over time or vary between categories.

Lollipop chart Lollipops draw more attention to the data value than standard bar charts, but can also show rank and value effectively.

Distribution

Show values in a dataset and how often they occur. The shape (or 'view') 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 (geographic) distribution

Histogram The standard way to show a statistical distribution - keep the gaps between columns small to reflect the 'shape' of the data.

Boxplot Summarises multiple distributions by showing the median (centre) and range of the data.

Violin plot Similar to a box plot but shows the full probability density function for complex distributions (data that cannot be summarised with a simple average).

Population pyramid A standard way for showing the age and sex breakdown of a population distribution; effectively back-to-back histograms.

Slope Perfect for showing how ranks have changed over time or vary between categories.

Dot strip plot Similar to a pie chart but for showing the size and sex breakdown of a population distribution; effectively back-to-back histograms.

Lollipop chart Lollipops draw more attention to the data value than standard bar charts, but can also show rank and value effectively.

Change over Time

Give emphasis to changing trends. These can be short (one-day) movements or extended series (several decades or centuries). Choosing the right time period is important to provide suitable context for the reader.

Example FT uses Share price movements, economic time series

Line The standard way to show a changing time series - keep the gaps between columns small to reflect the 'shape' of the data.

Column Columns work well for showing change over time – but usually best with one series of data at a time.

Line + column Columns work well for showing change over time – but usually best with one series of data at a time.

Pie A common way of showing part-to-whole relationships over time within an amount (coloured) and a rate (grey).

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

Dot plot Similar to a pie chart – but the centre can be a good way of making space to include more information about the data (eg. total).

Slope Good for showing individual values in a distribution, can be a problem when too many values have the same value.

Barcode plot Like dot strip plots, good for displaying all the data in a single chart, but keep in mind when highlighting individual values.

Cumulative curve A good way of showing how unequal a distribution is; x axis is frequency, y axis is always a measure.

Part-to-whole

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.

Example FT uses Fiscal budgets, company structures, national election results

Stacked column A simple way of showing part-to-whole relationships but can be difficult to read with many components.

Proportional stacked bar A good way of showing the size and proportion of data at the same time – as long as the data are not too complicated.

Bar See above. Good when the data are not time series and labels have long category names.

Paired column As per standard stacked bars, but allows for multiple series. Can become tricky to read with more than 2 series.

Dot plot See above.

Dot matrix Similar to a pie chart – but the centre can be a good way of making space to include more information about the data (eg. total).

Treemap Similar to a pie chart – but the centre can be a good way of showing the size and proportion of data at the same time – as long as there are many small segments.

Proportional stacked bar A good way of showing part-to-whole relationships – problem is that each area is sized according to a particular value.

Area chart Use with care – these are good at showing changes to totals, but sometimes changes across multiple categories can be very difficult.

Barcode plot Like dot strip plots, good for displaying all the data in a single chart, but keep in mind when highlighting individual values.

Fan chart (projections) Good for showing the uncertainty in future projections – usually this grows the further forward to projections.

Connected scatterplot A good way of showing changing data for two variables when there is a repeating pattern of progression.

Calender heatmap A great way of showing temporal patterns (daily, weekly, monthly) and the presence of showing precision in quantity.

Magnitude

Show size comparisons. These can be relative (size being able to see larger/larger) or absolute (need to see fine differences). Usually these show a ratio (for example, barrels, dollars or people) rather than a calculated rate or per cent.

Example FT uses Commodity production, market capitalisation

Basic choropleth (rate/ratio) The standard approach for plotting data on a map – should always be rates rather than totals and use a sensible base geography.

Waterfall Designed to show the sequencing of data through a flow process, typically budgets, which can include +/− components.

Chord A complex but effective diagram which can illustrate 2-way flows (and net winner) in a matrix.

Network Used for showing the strength and interconnectedness of relationships of varying types.

Spatial

Used only when precise locations or 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.

Flow

Show size comparisons. These can be relative (size being able to see larger/larger) or absolute (need to see fine differences). Usually these show a ratio (for example, barrels, dollars or people) rather than a calculated rate or per cent.

Visual vocabulary

Designing with data

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; Ben Bern; Li Fausto; Graham Parry; Billy Ehrenberg; Paul McCullum; Martin Stalbe
Inspired by the Graphic Contourism by Jon Schwabish and Severine Ribecca

ft.com/vocabulary



<http://www.datavizcatalogue.com/>

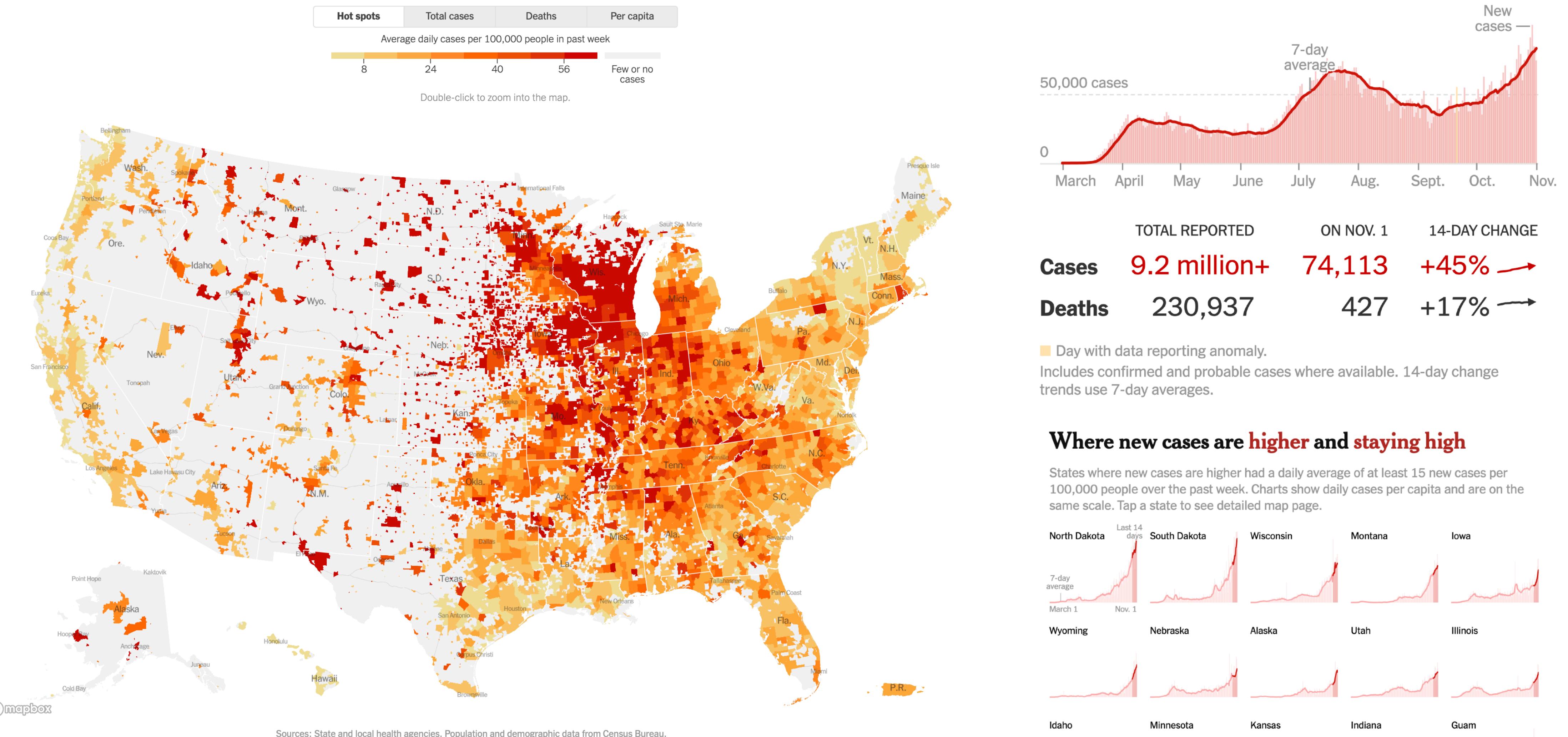
<https://github.com/ft-interactive/chart-doctor/blob/master/visual-vocabulary/Visual-vocabulary.pdf>



6. What style to use?

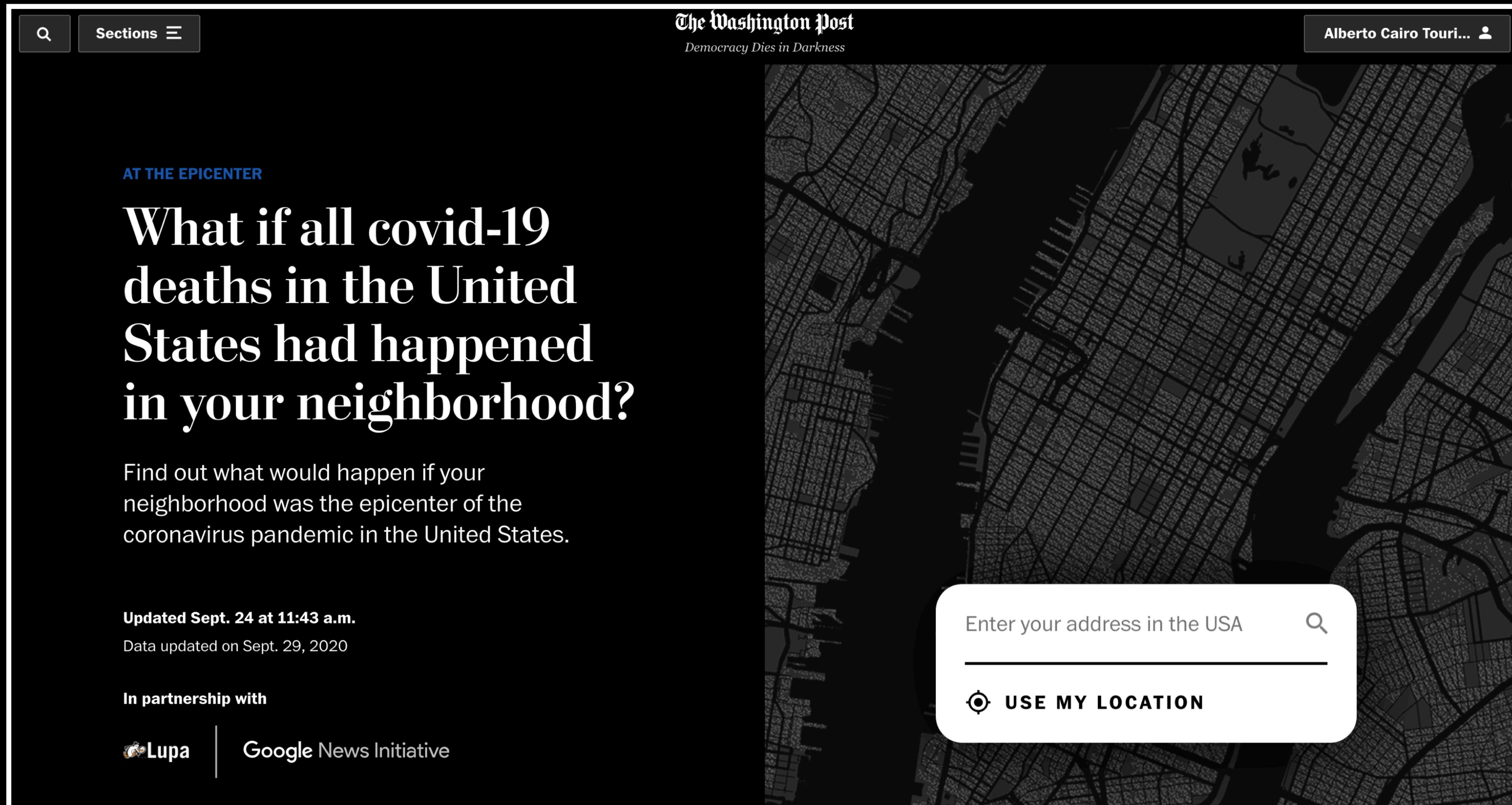
Standard visualizations

Appropriate for graphics we use all the time

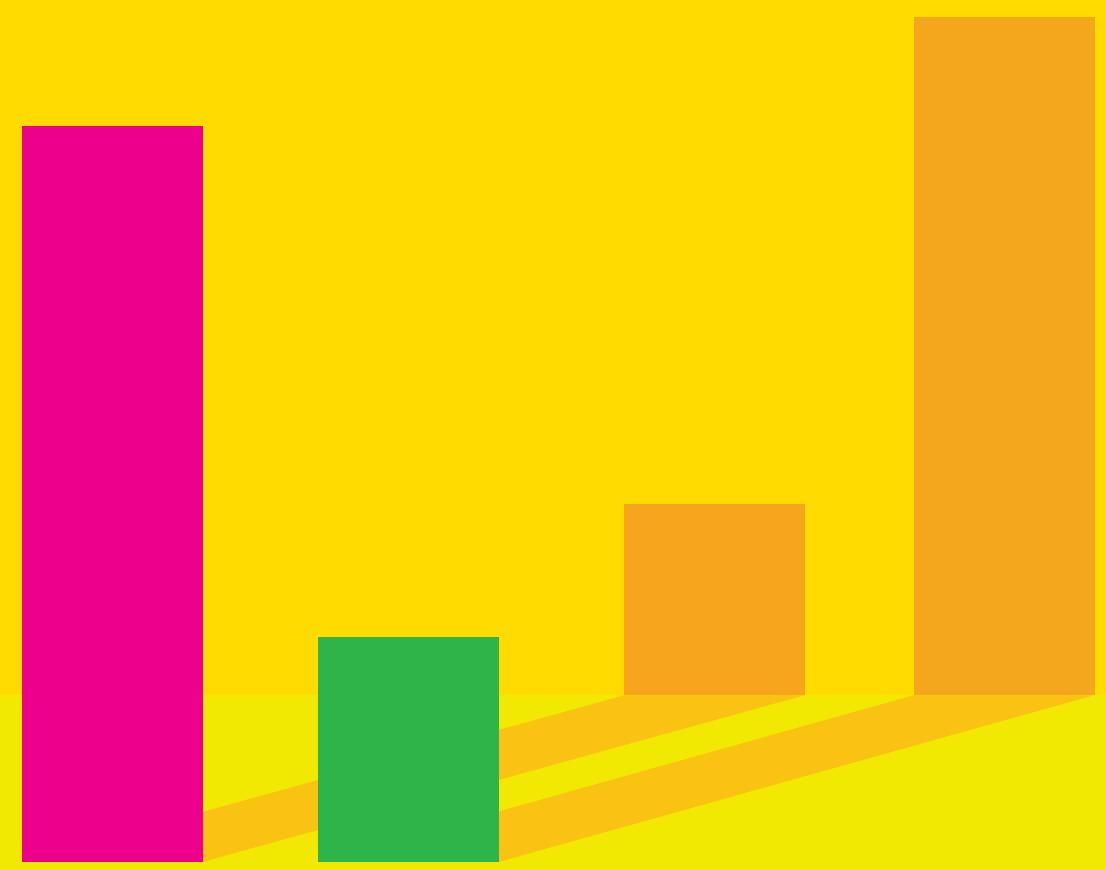


<https://www.nytimes.com/interactive/2020/us/coronavirus-us-cases.html>

Unusual or unique visualizations



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



Where to go from here?
(resources)

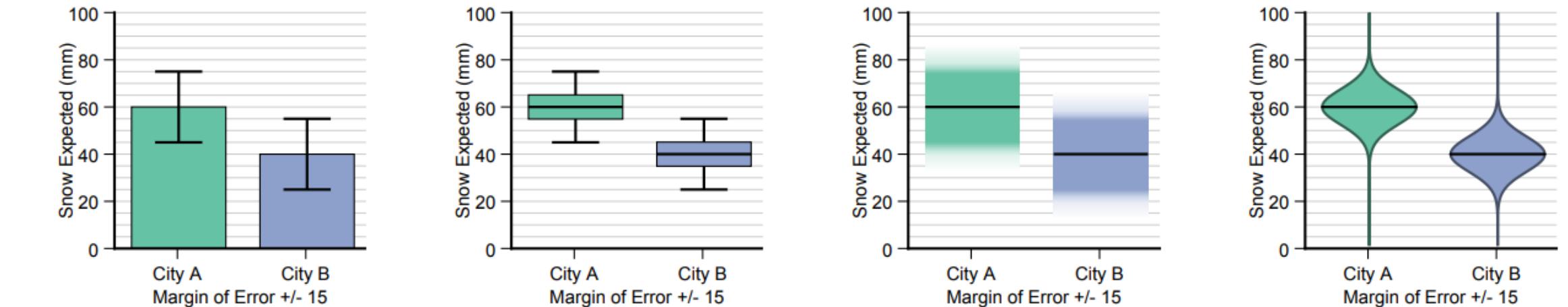
Disclosing limitations and uncertainty

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.

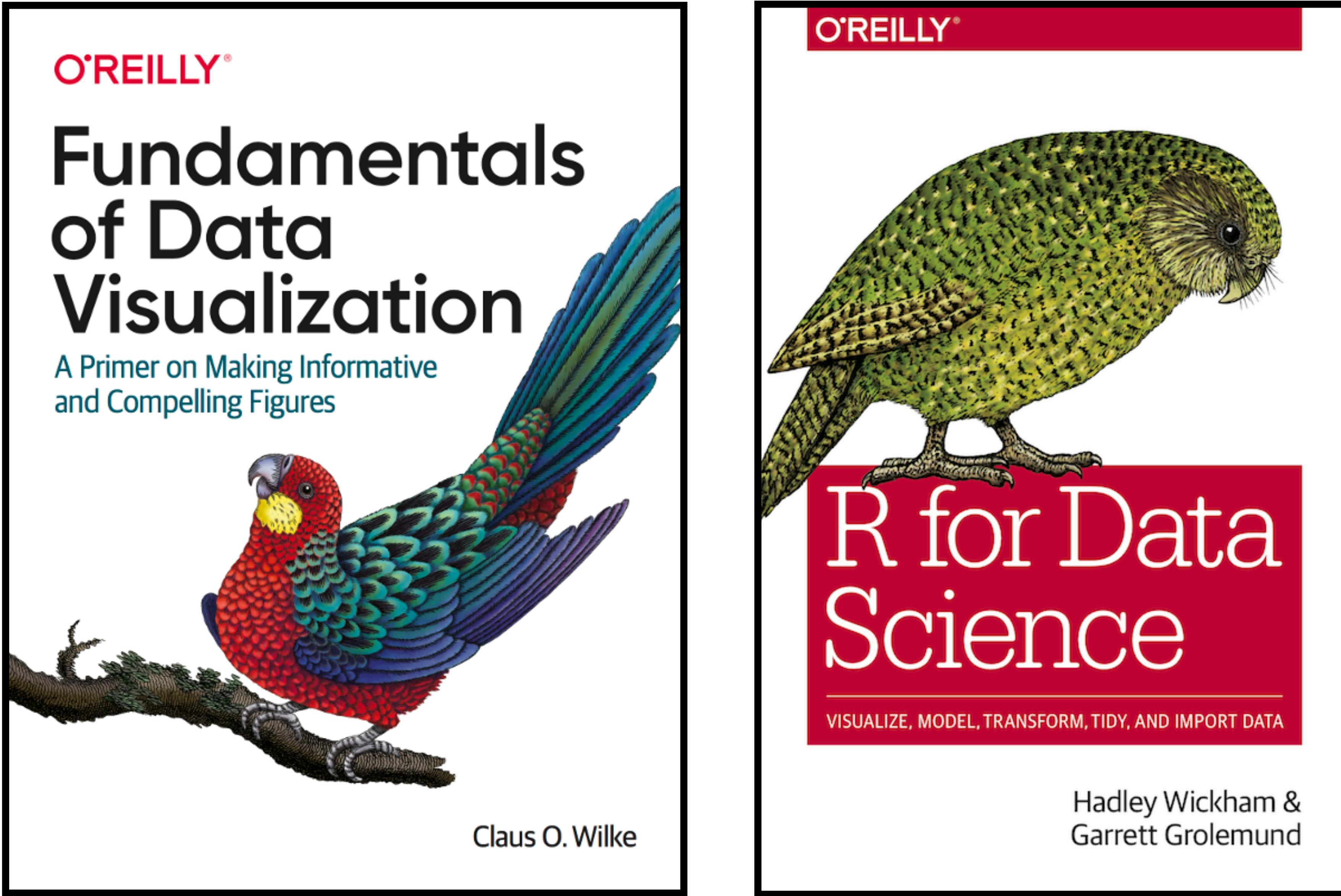
(d) **Violin plot:** the width of the colored region corresponds to the probability density function of a t-distribution.

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

Collection of papers about visualizing uncertainty:

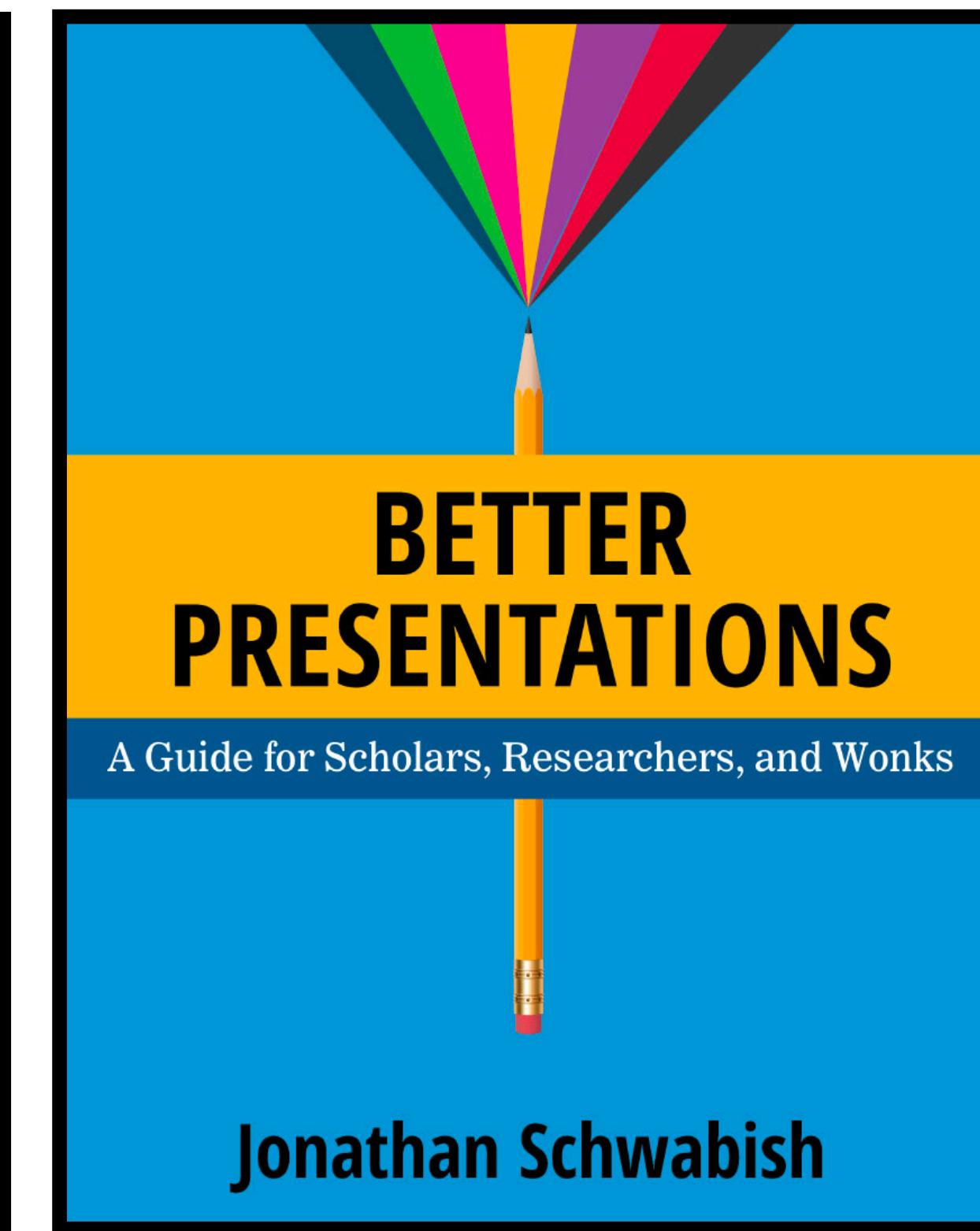
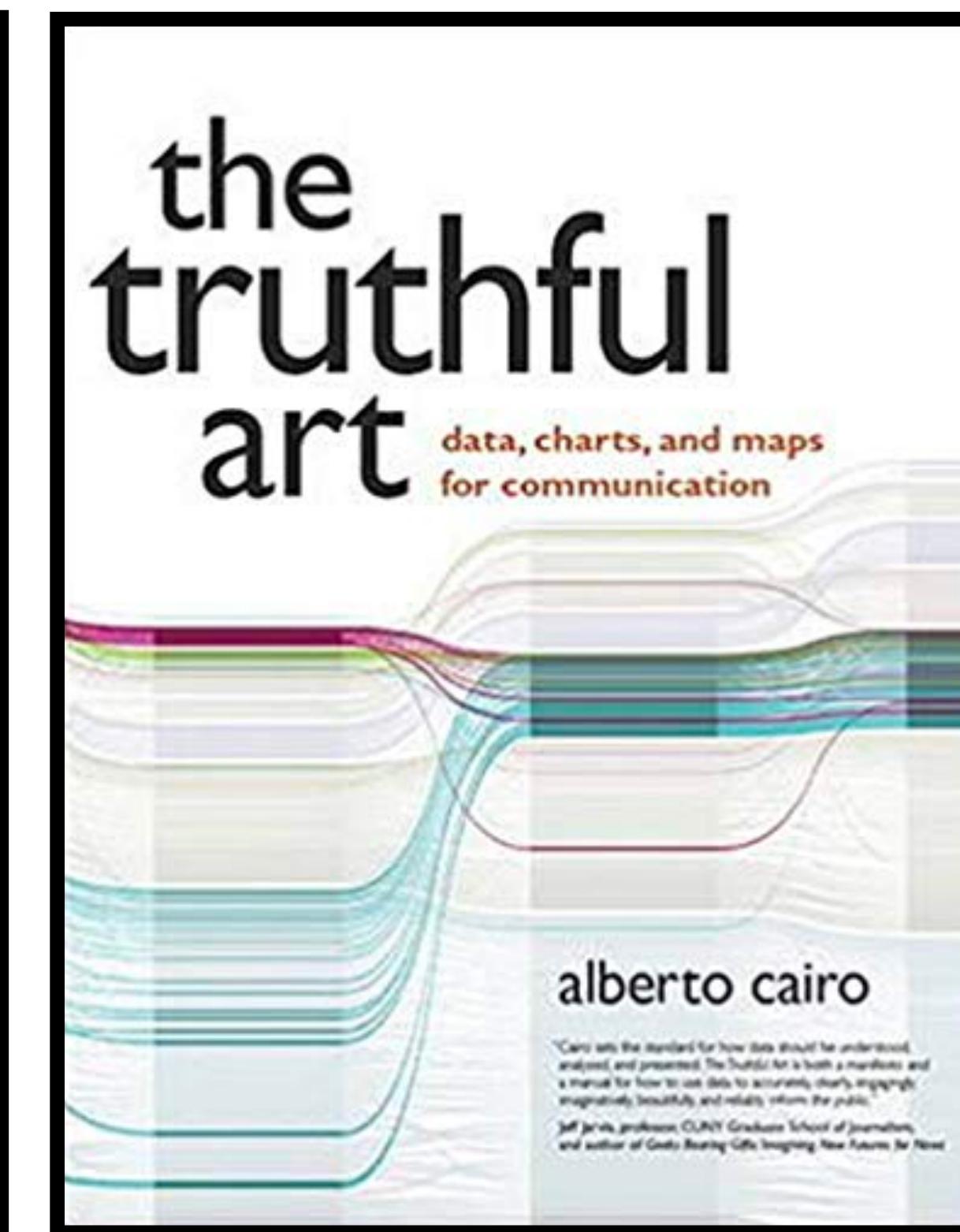
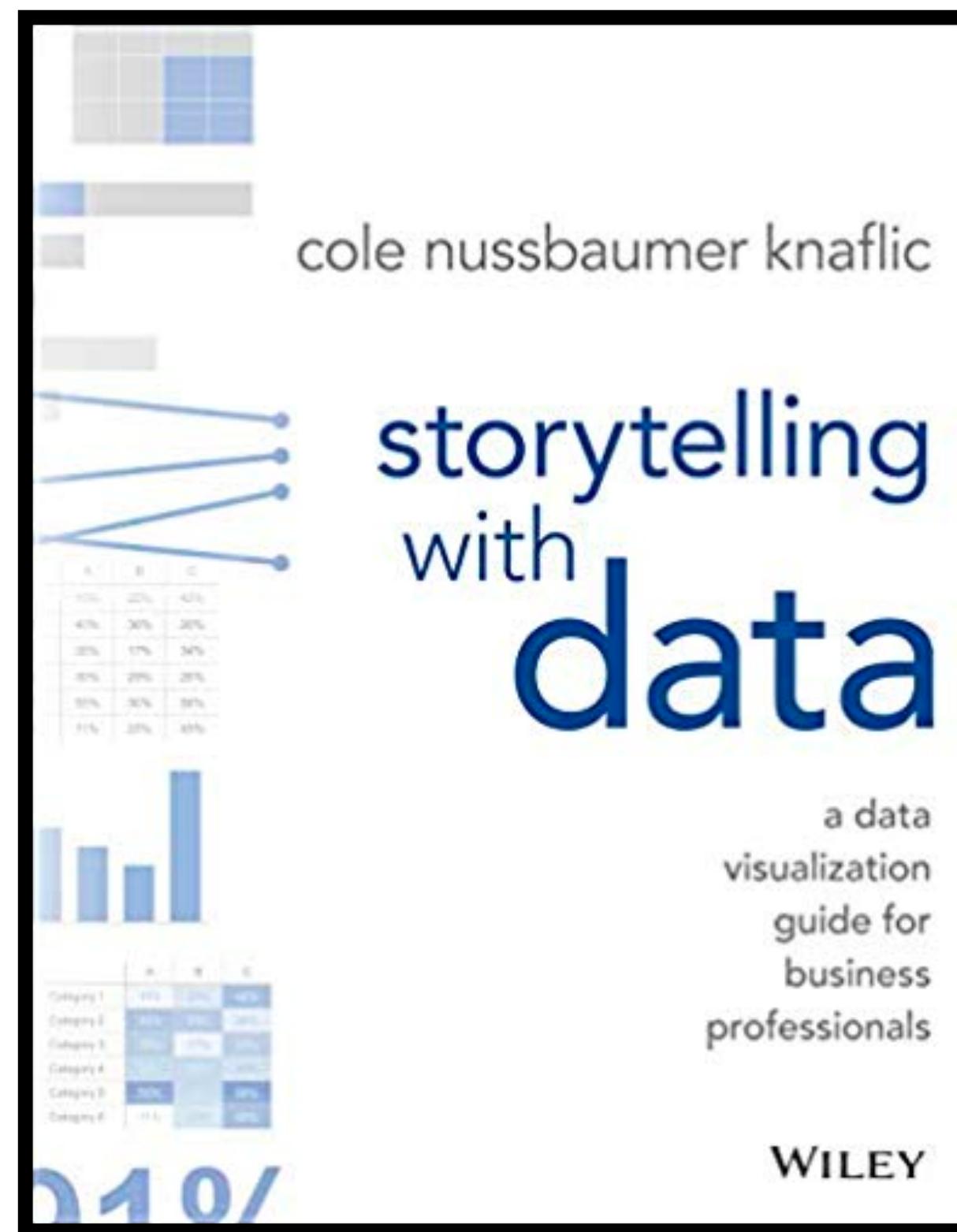
<https://www.dropbox.com/sh/jk4ginxyai6ylqu/AABvqdyTlhJtyFN9nKNHyX9Ba?dl=0>

Books available online

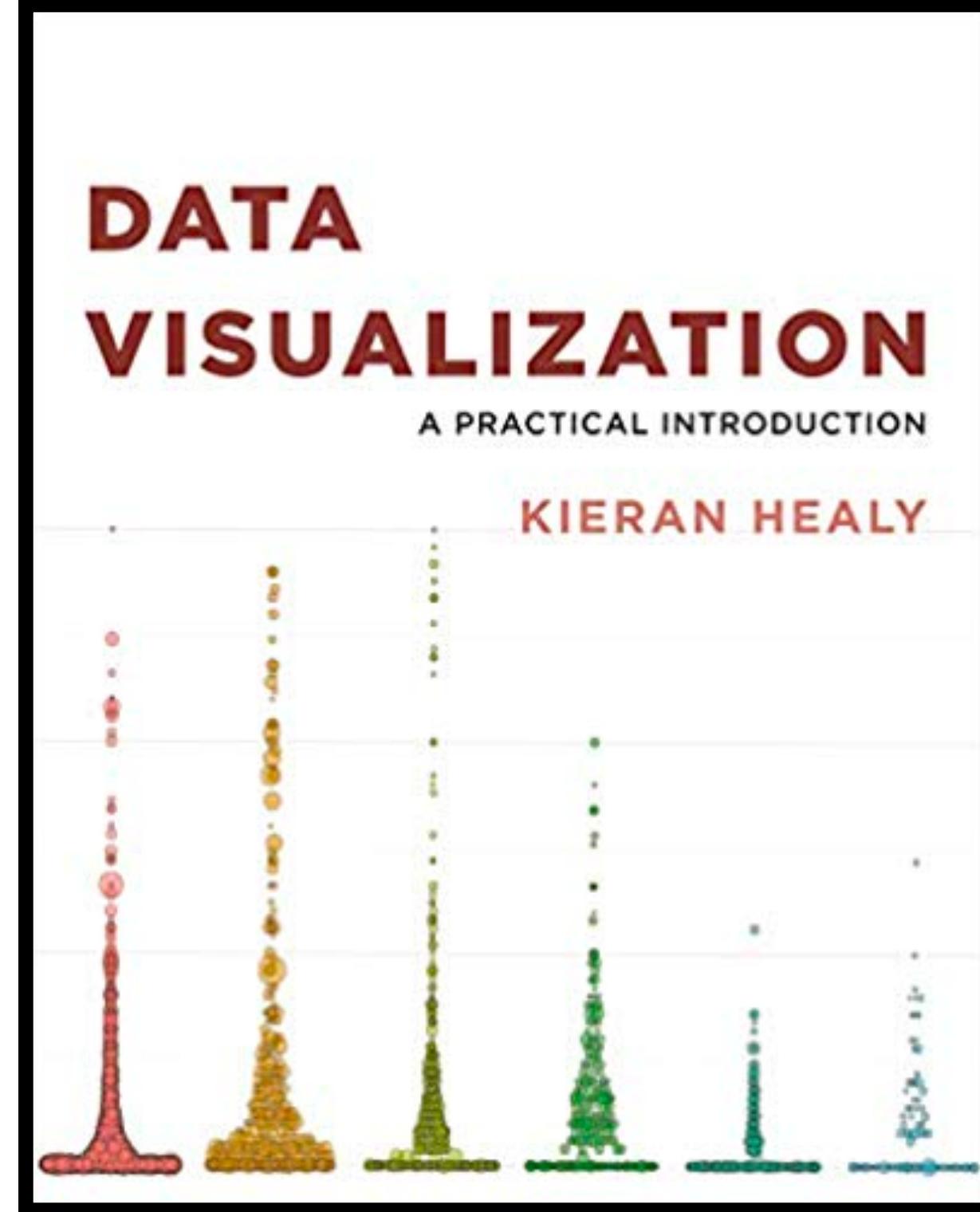
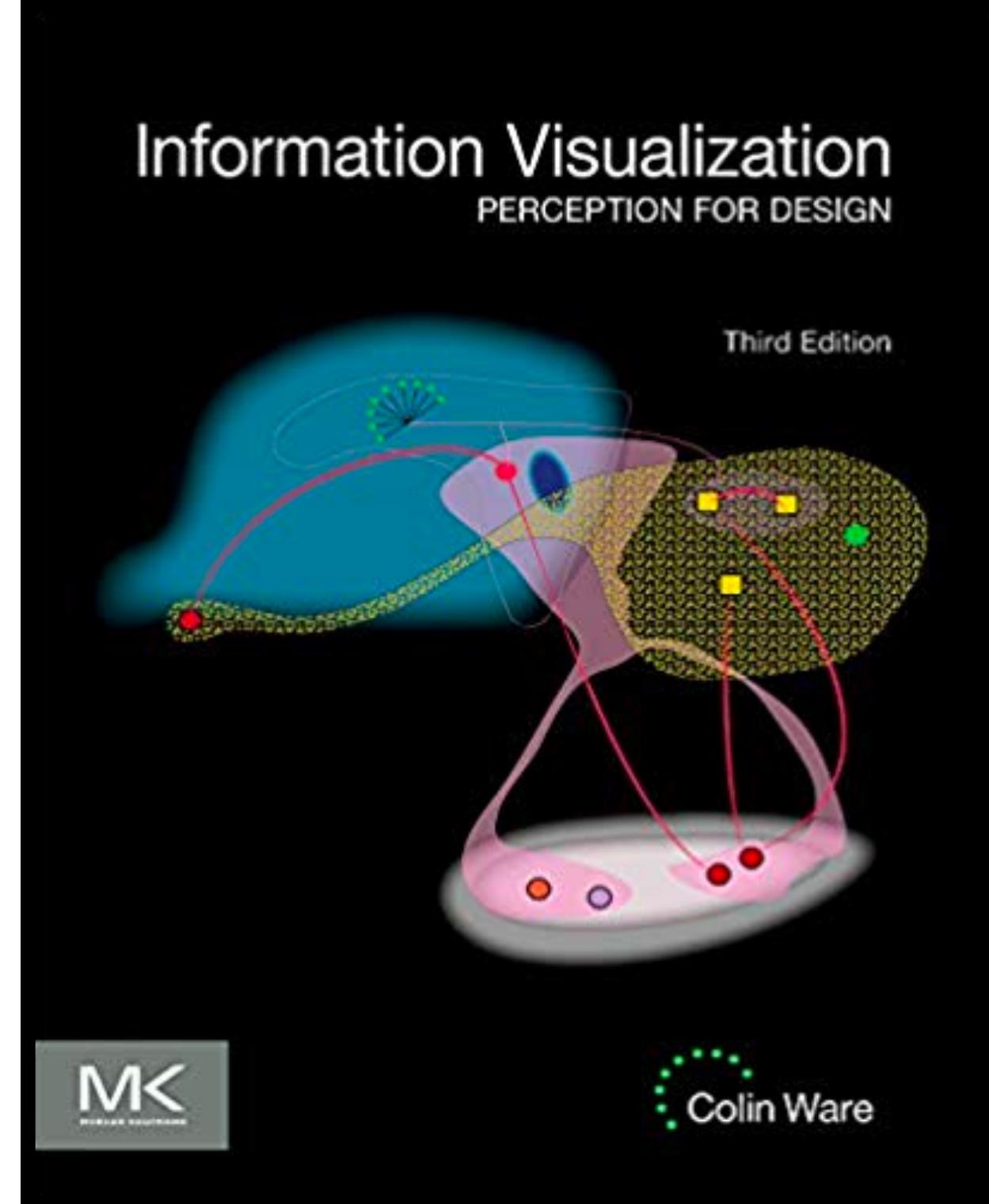
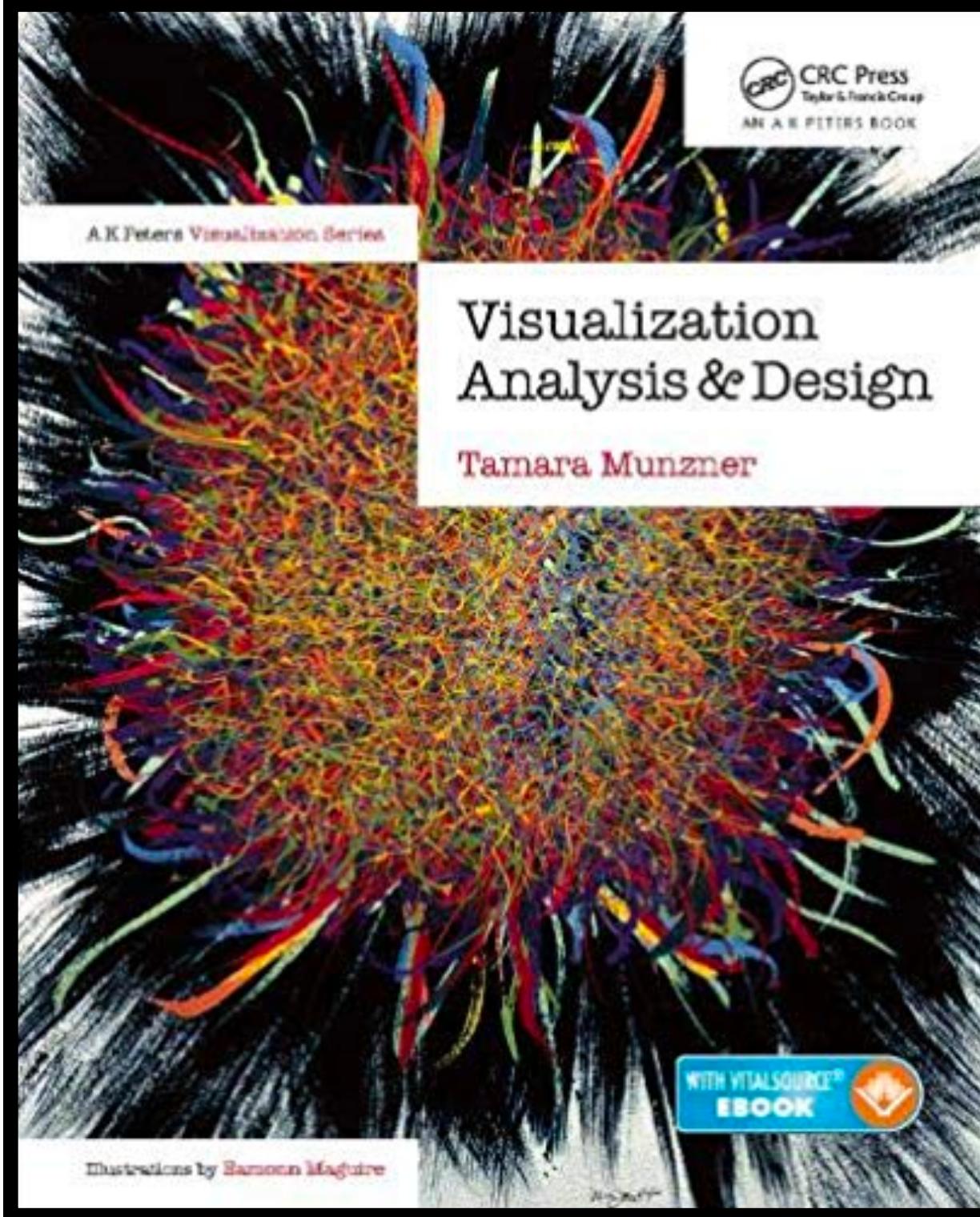


<https://serialmentor.com/dataviz/>

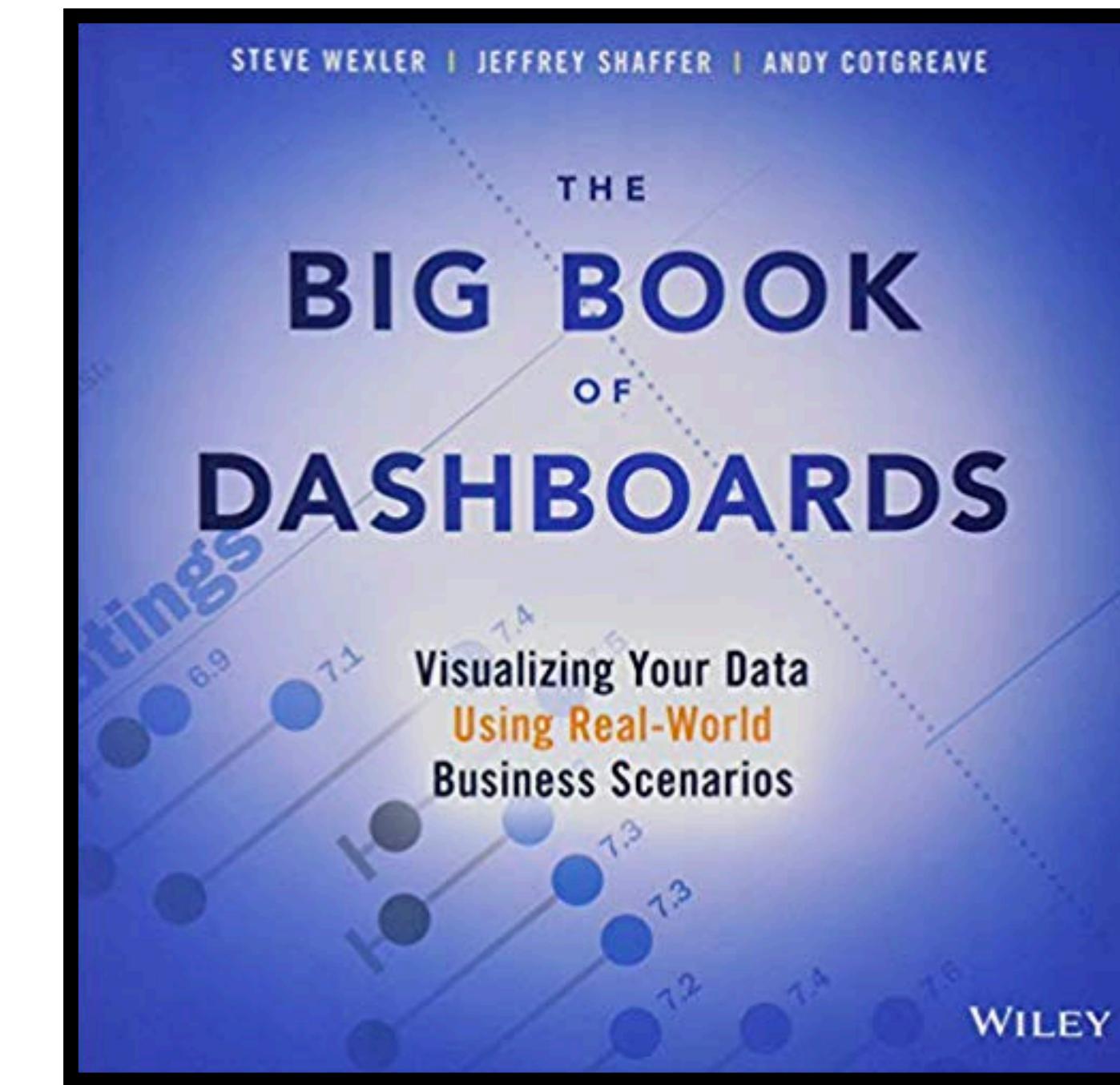
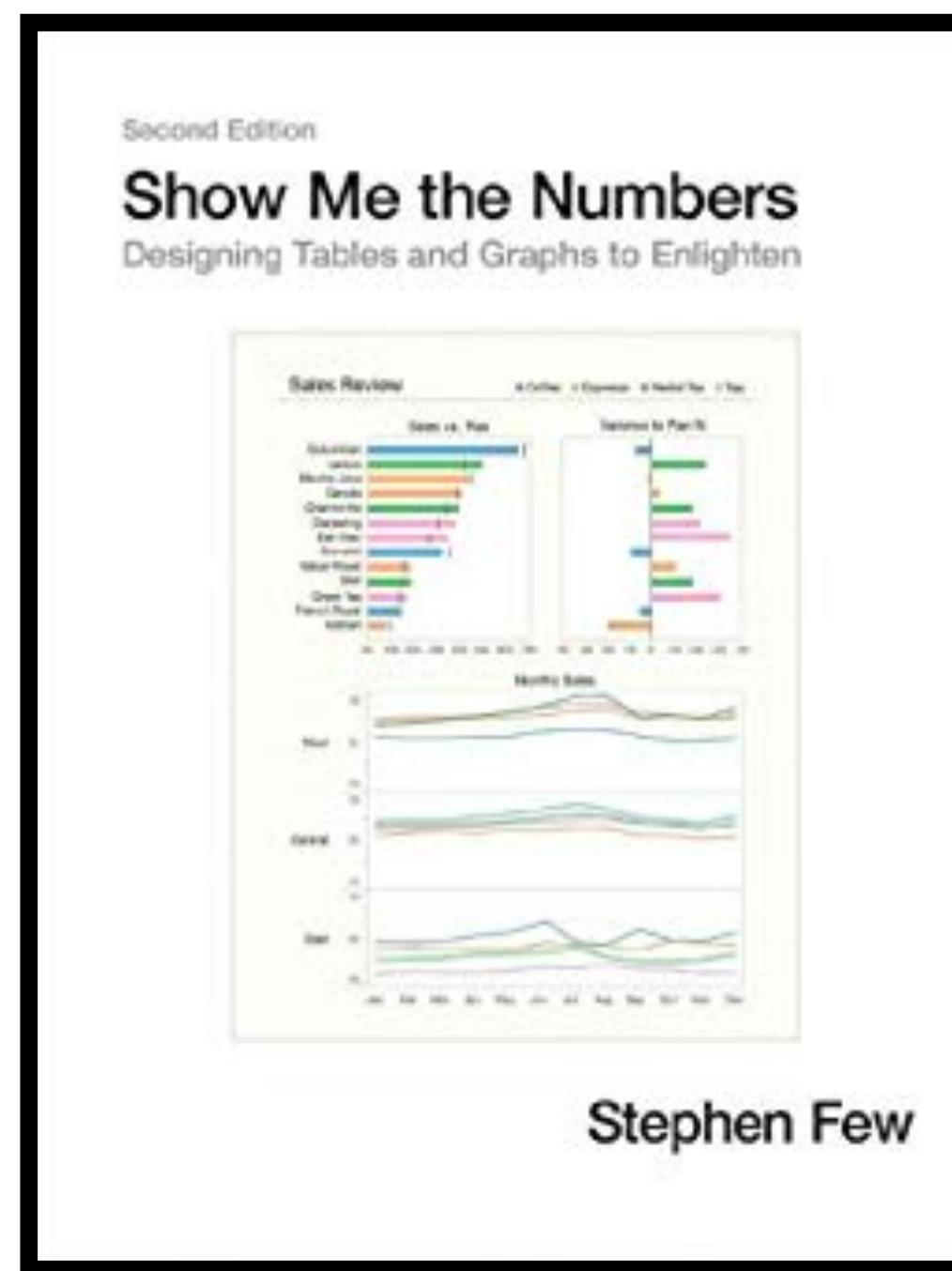
<https://r4ds.had.co.nz/>



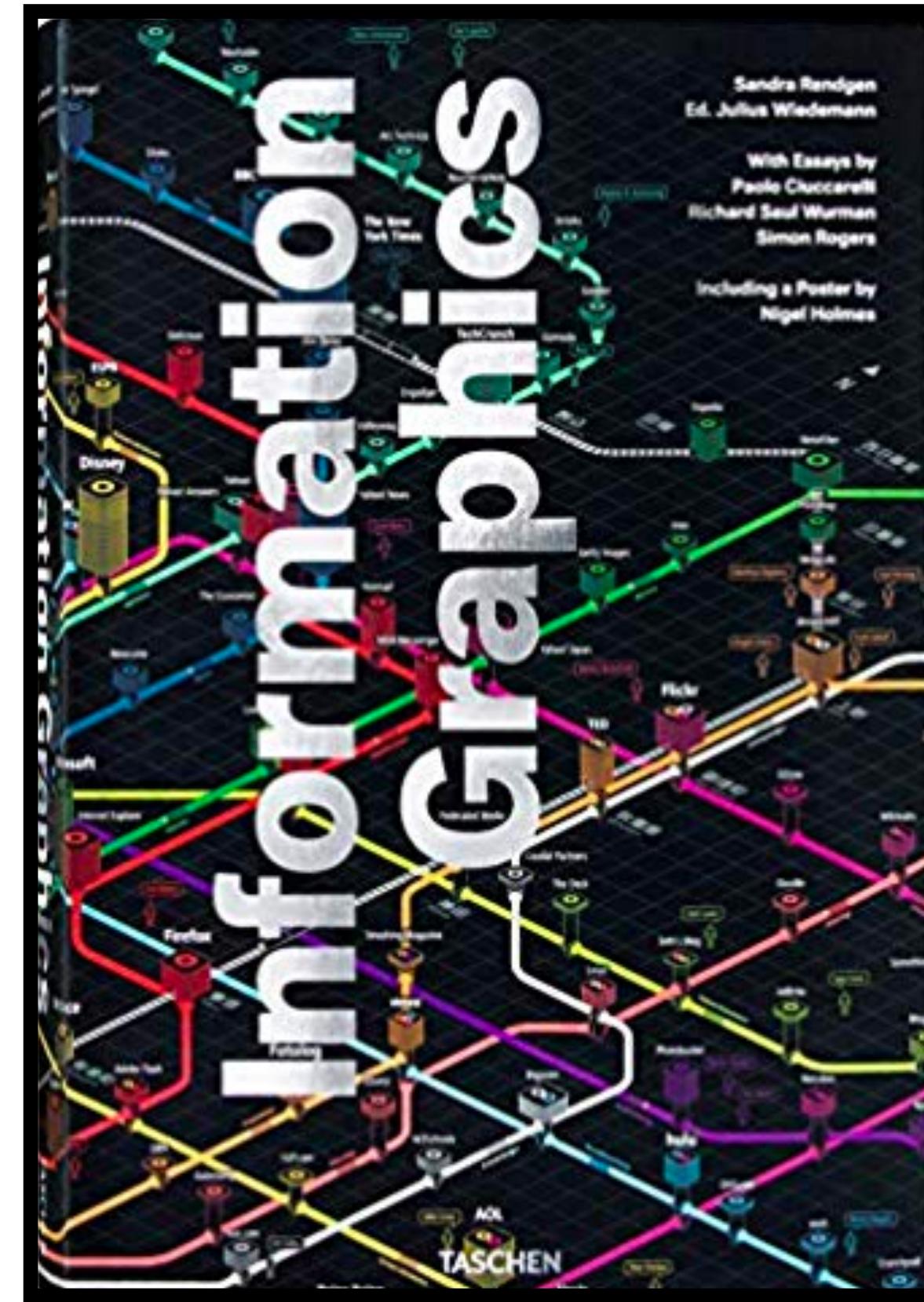
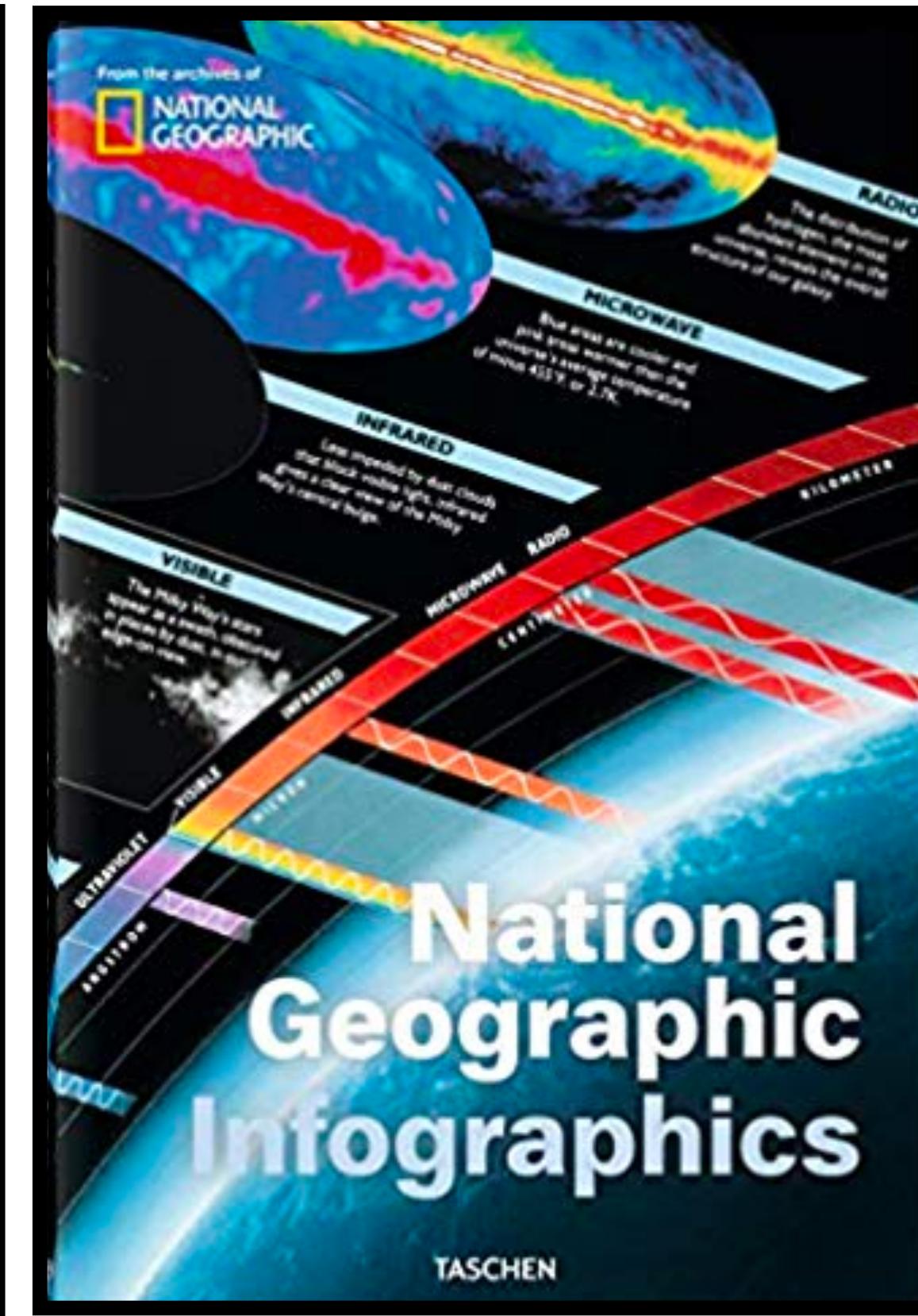
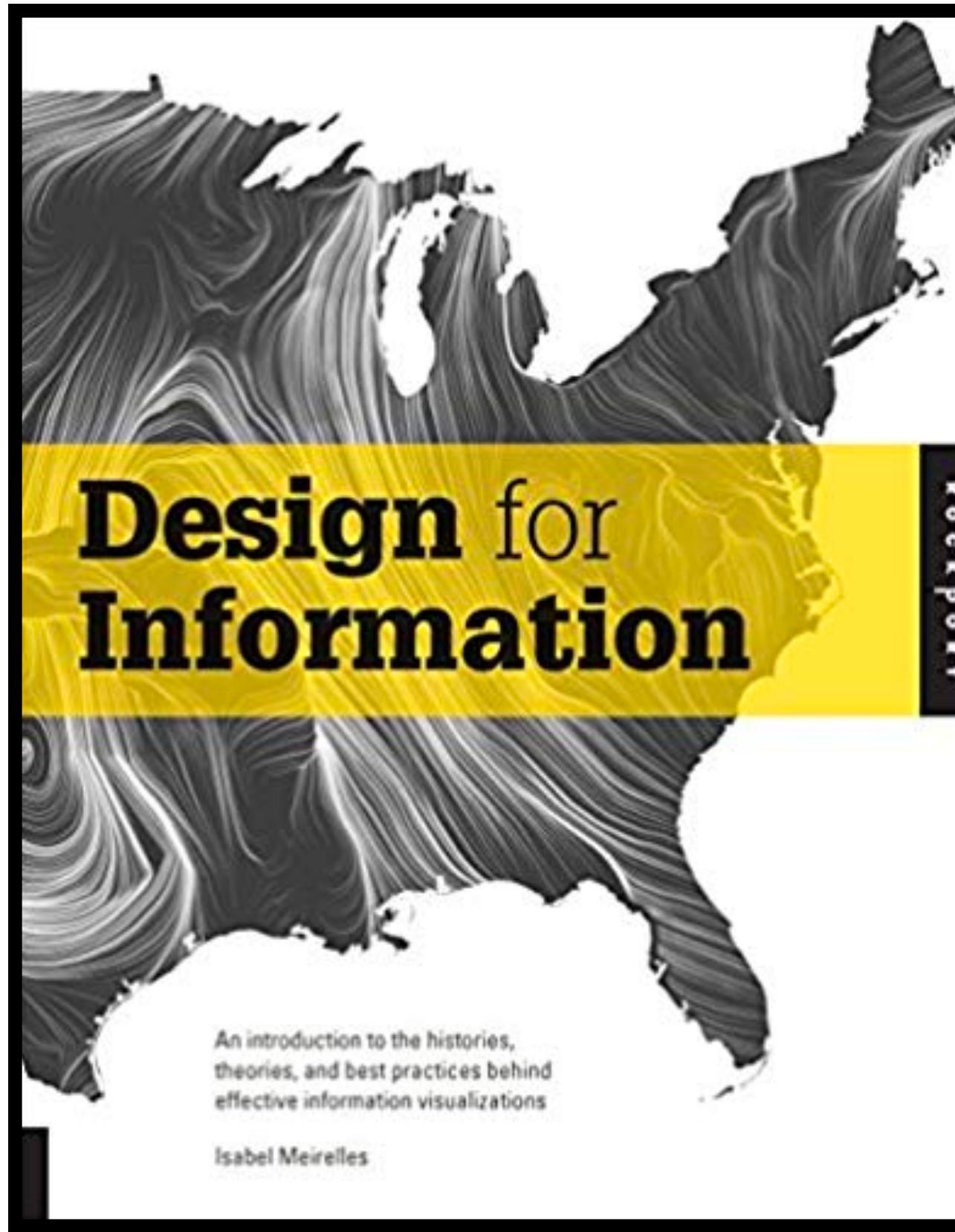
Fundamentals



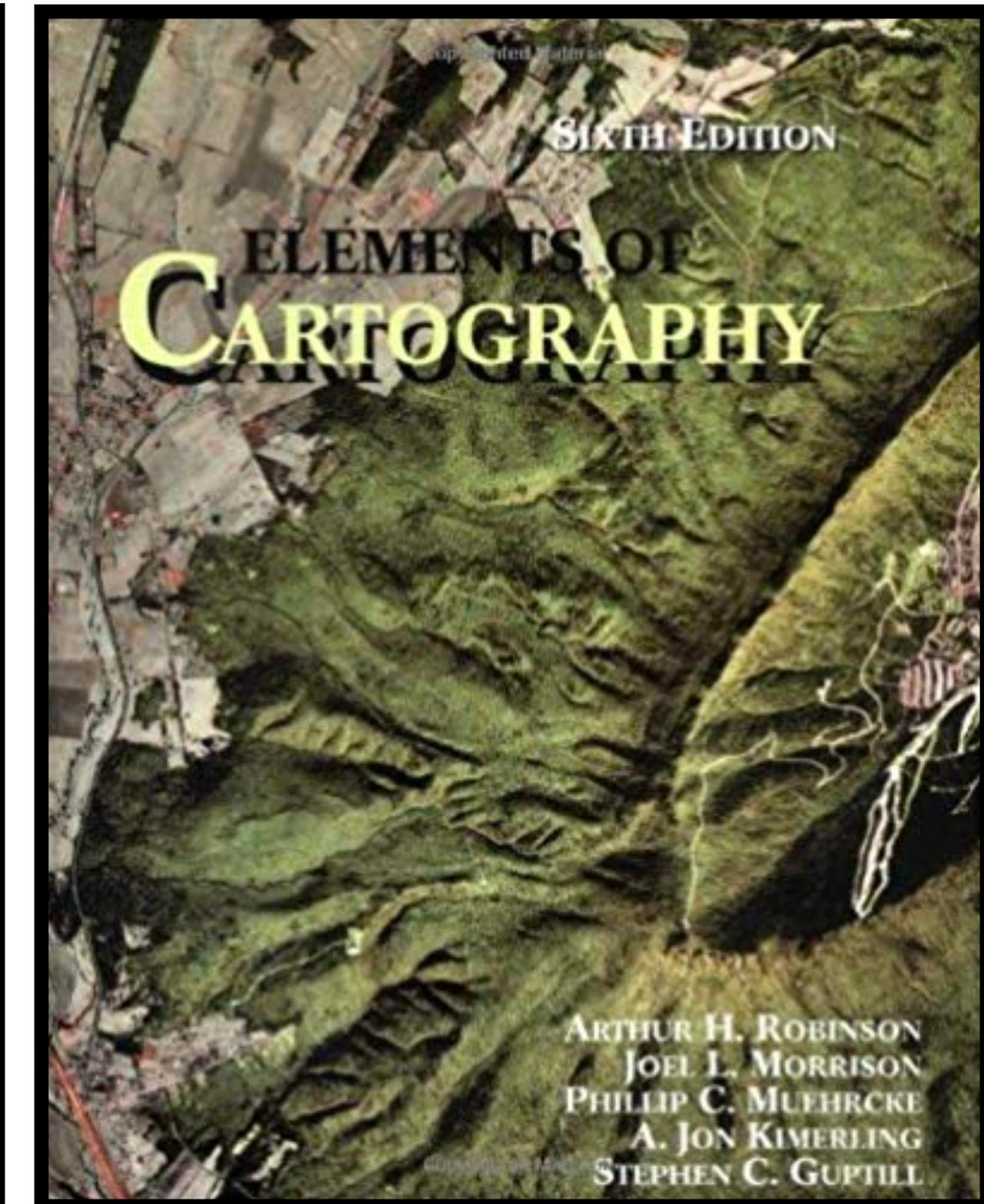
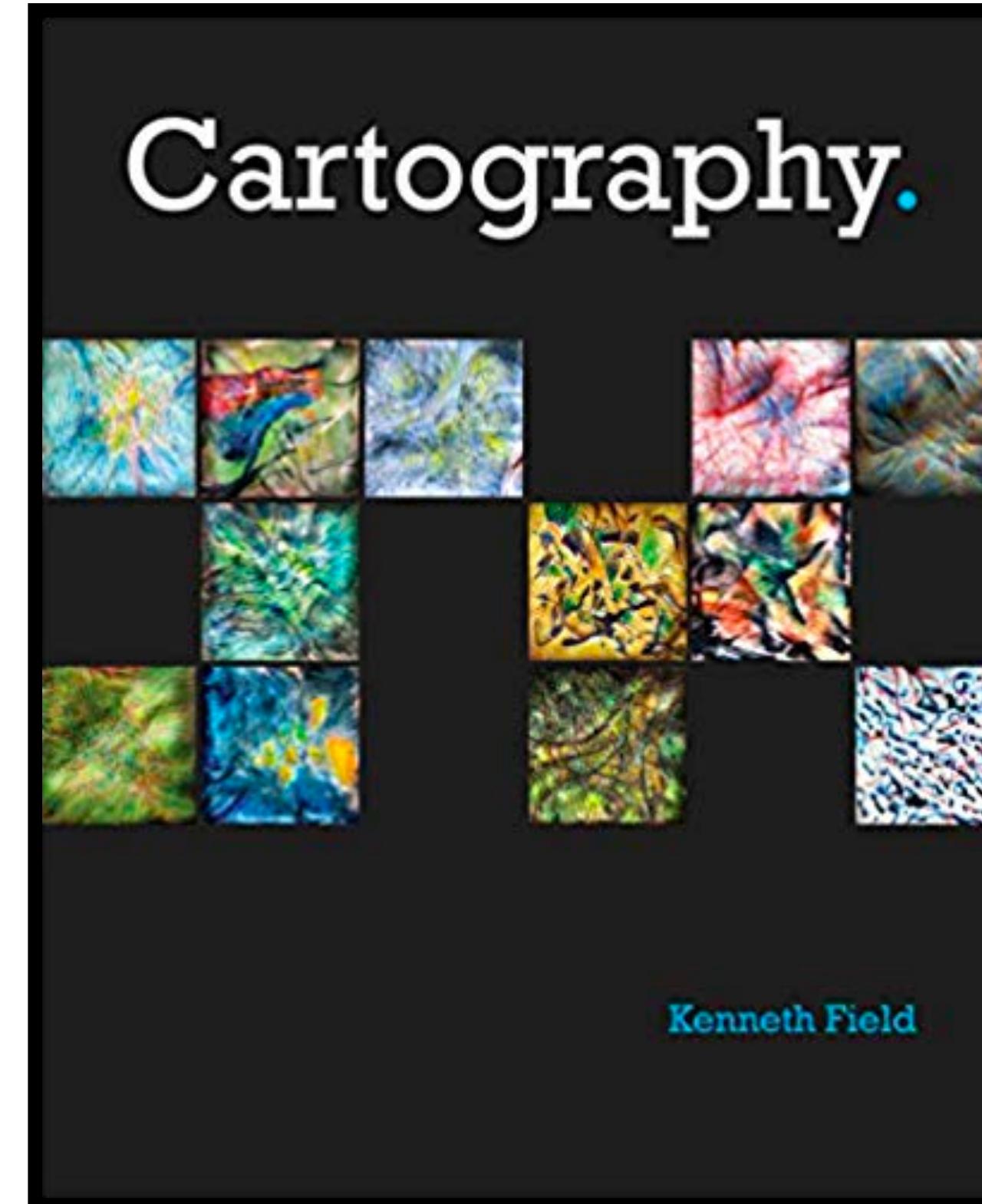
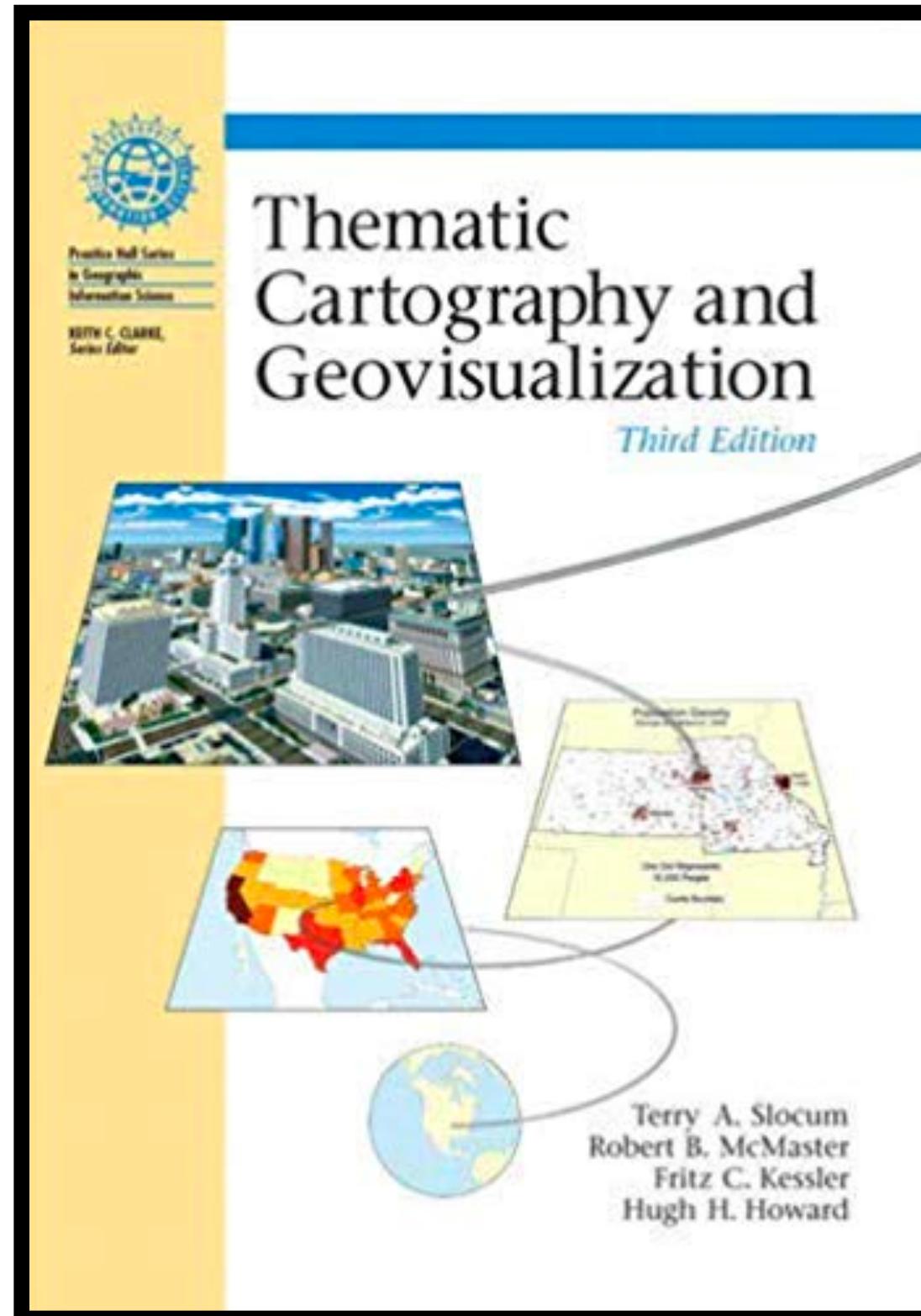
Exploratory/scientific visualization



Business analytics



History, inspiration, and examples



Mapping



The End.

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