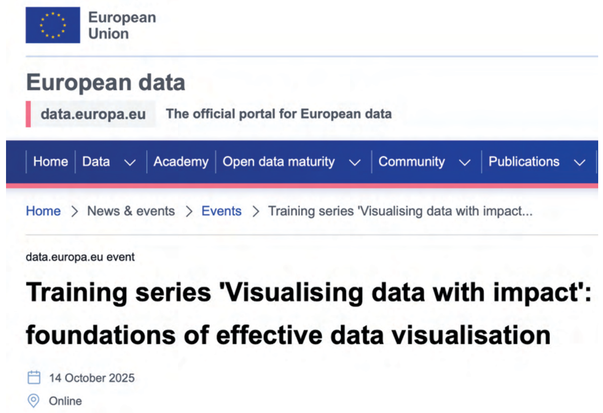
Designing With Integrity

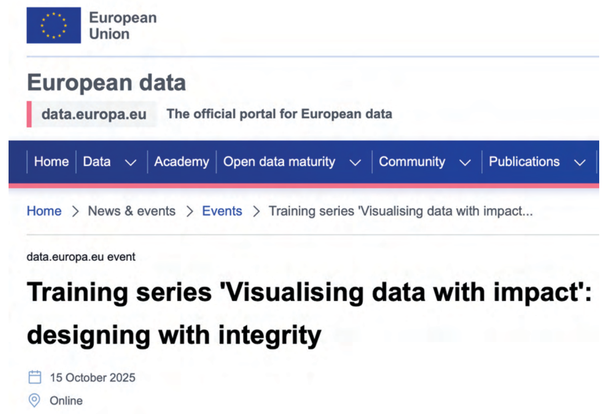
Alberto Cairo

OpenVisualizationAcademy.com



Episode ONE - October 14

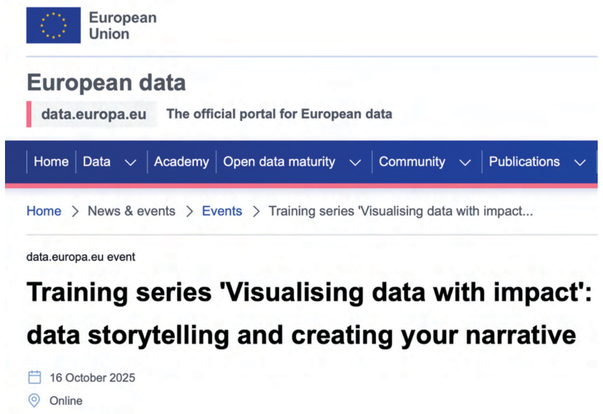
https://data.europa.eu/en/news-events/events/training-data-visualisation-session-1-foundations-effective-data-visualisation



Episode TWO - October 15

Today's webinar

https://data.europa.eu/en/news-events/events/data-visualisation-training-session-2-designing-integrity



Episode THREE - October 16

https://data.europa.eu/en/news-events/events/data-visualisation-training-session-3-data-storytelling-and-creating-your

EPISODE ONE: Foundations of effective data visualization
Defining information design and data visualization
Exploration, exposition, explanation, expression in visualization
Why do we visualize? The basics
The grammar of graphics
"Rules" versus decision-making

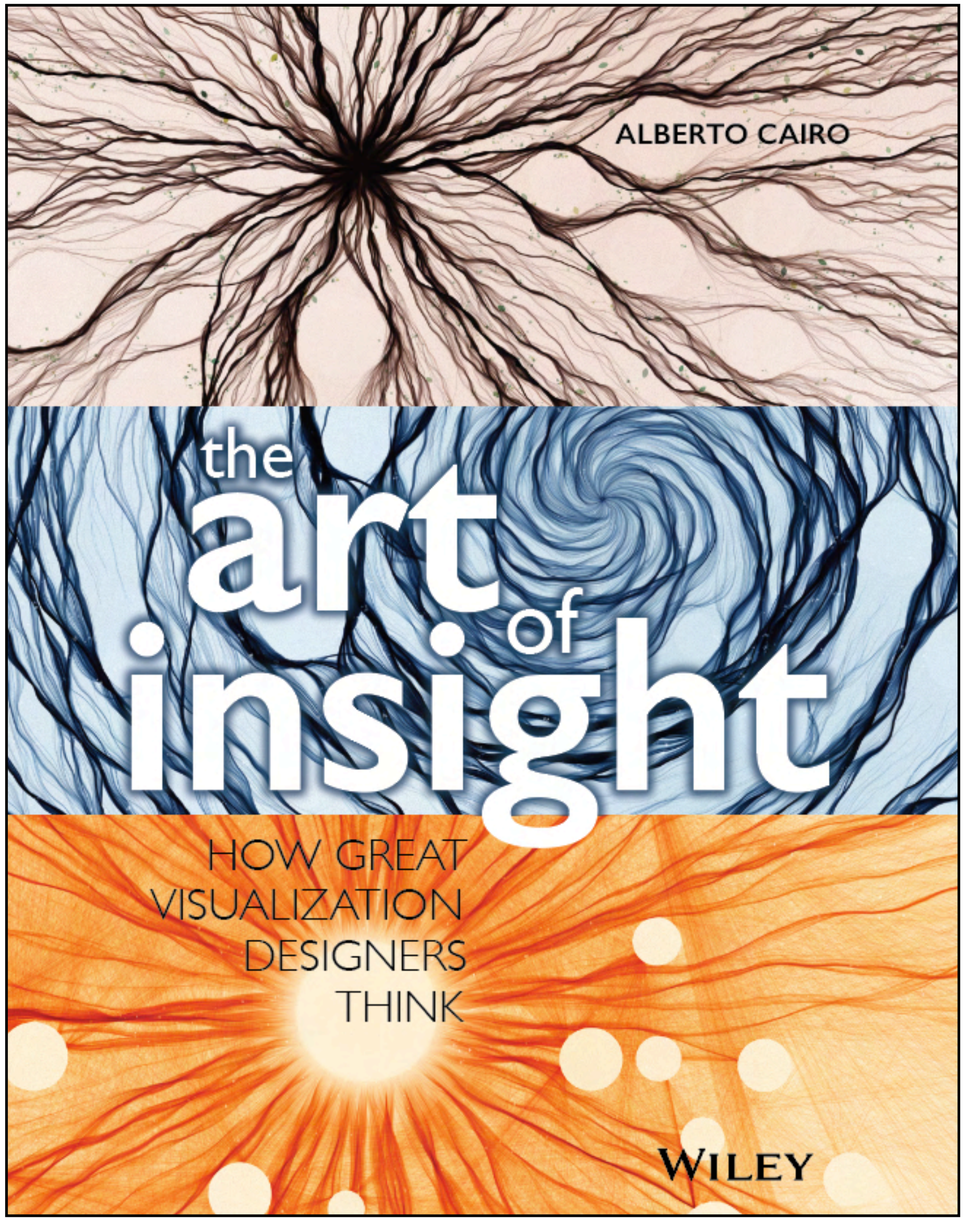
EPISODE TWO: Designing with integrity
Myths of visualization: From "A picture is worth a thousand words" to "the data should speak for itself"
The role of mental models
Why and when do charts "lie", and what to do about it?
A structured way for thinking about visualization and minimizing misunderstanding

EPISODE THREE: Data storytelling and creating your narrative
Structuring a layout
Building a narrative
Considering visual design
Creative visualization
What comes next in visualization?

EPISODE ONE: Foundations of effective data visualization
Defining information design and data visualization
Exploration, exposition, explanation, expression in visualization
Why do we visualize? The basics
The grammar of graphics
"Rules" versus decision-making

EPISODE TWO: Designing with integrity
Myths of visualization: From "A picture is worth a thousand words" to "the data should speak for itself"
The role of mental models
Why and when do charts "lie", and what to do about it?
A structured way for thinking about visualization and minimizing misunderstanding

EPISODE THREE: Data storytelling and creating your narrative
Structuring a layout
Building a narrative
Considering visual design
Creative visualization
What comes next in visualization?



Visualization design consists of **reasoning** about possible **choices** by considering the interplay between:

1. **Content:** The nature, origin, and limitations of the data.
2. **People:** Your audience.
3. **Intent:** The purpose(s) that we define.
4. **Constraints:** The limitations that we may face.
5. **Outcomes:** How the graphic is received.

Every design choice must be **deliberate.**

It is inevitably **subjective**, but it should never be **arbitrary**.

Myths about data visualization

“The data should speak for itself!”

“A picture is worth a thousand words”

“Show, don’t tell!”

…


RALEIGH
ATLANTA
WPTV FIRSTALERT WEATHER
11:00 PM
Hurricane Lee
15.4°N , 47.7°W
Max Winds: 80 mph
Moving: WNW at 14 mph
Pressure: 989 mb
All these considerations are interrelated
Mon 8:00 PM
145 mph
Sun 8:00 PM
150 mph
Sat 8:00 PM
155 mph
Sat 8:00 AM
150 mph
Fri 8:00 AM
120 mph
Thu 8:00 PM
100 mph
Thu 8:00 AM
85 mph
4
4
4
4
4
3
2
1
1
WEST PALM BEACH
CUBA
DOMINICAN REPUBLIC
GUADELOUPE

https://www.nytimes.com/interactive/2019/08/29/opinion/hurricane-dorian-forecast-map.html

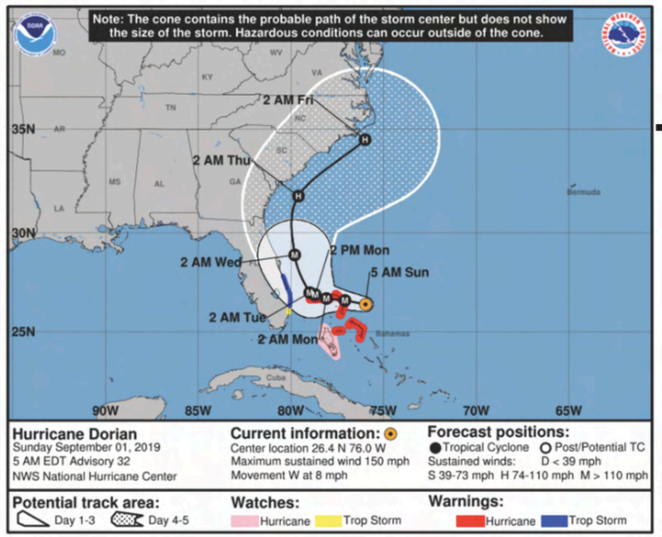
A designer might want to communicate something…

A designer might want to communicate something…

…but readers may interpret something different.

Traditional, unidirectional model of how graphical communication works

DESIGNER

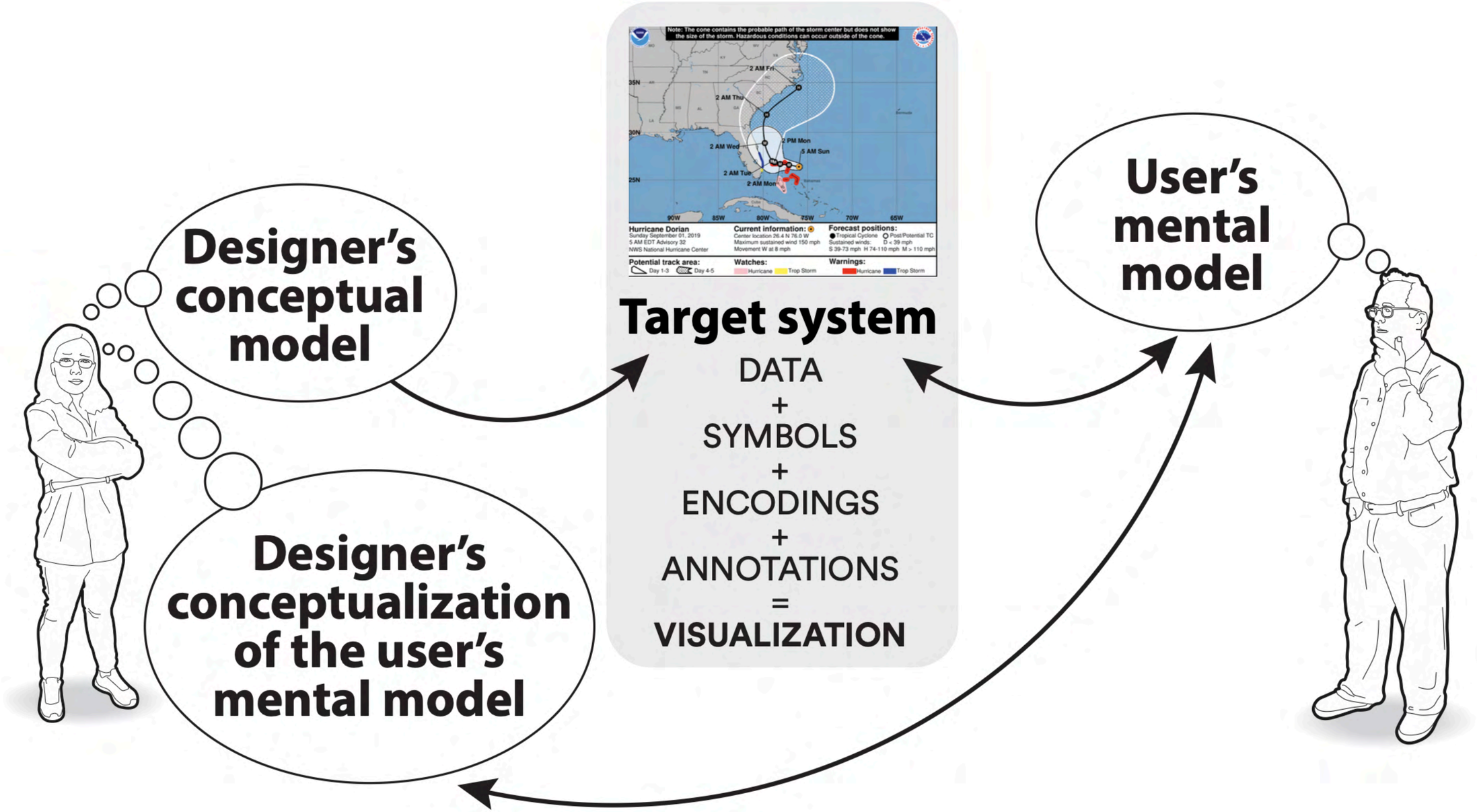


Visualization designed according to what the designer thinks will work best

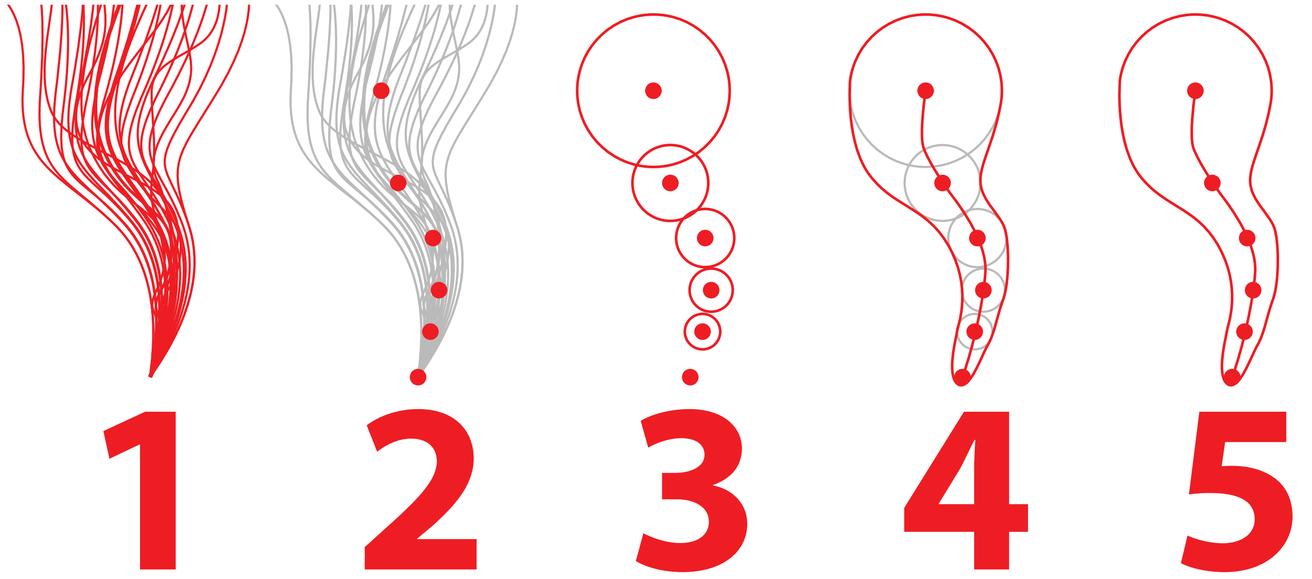
READER

This model is too simplistic

THE PSYCHOLOGY OF CHART COMPREHENSION, SIMPLIFIED

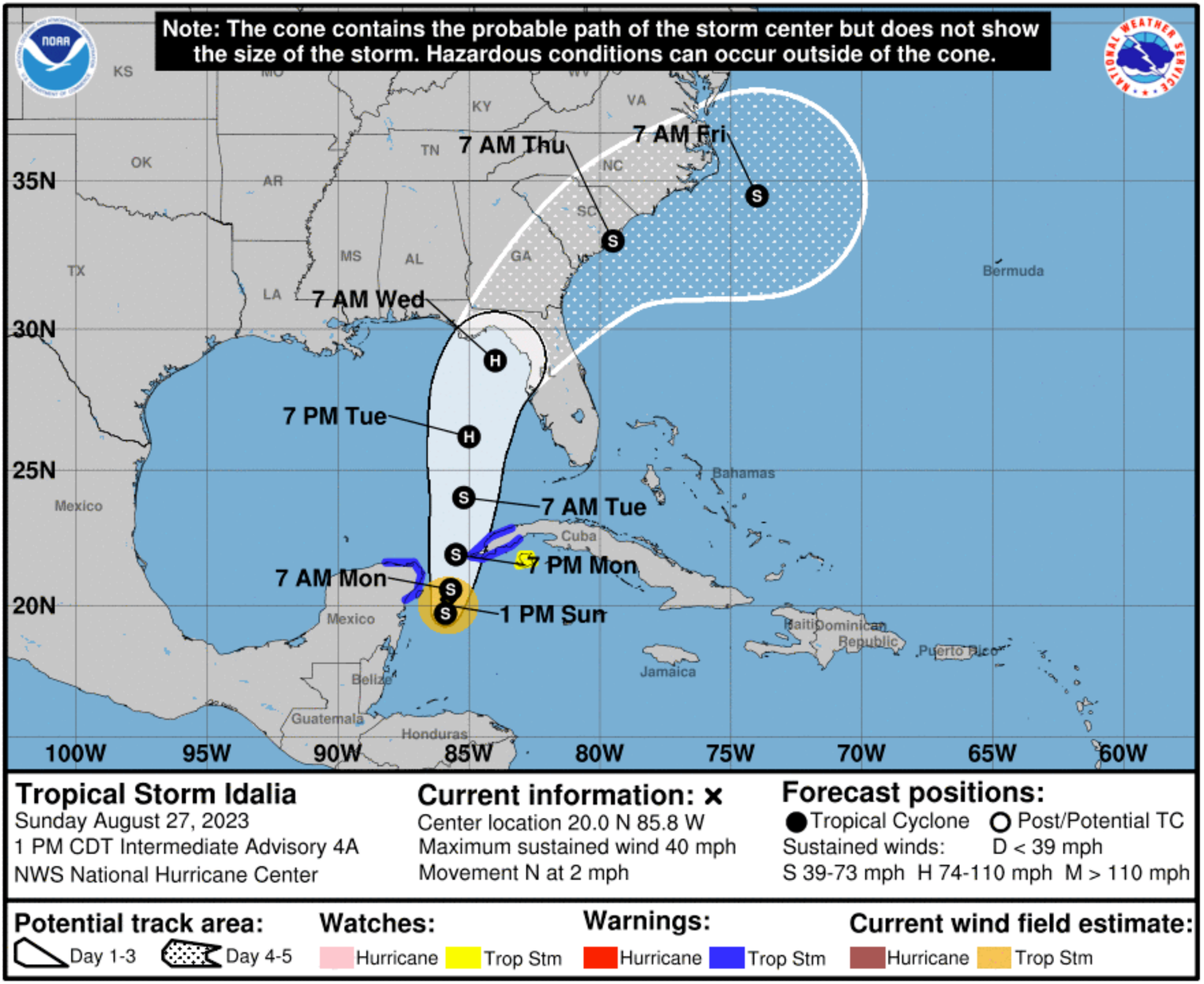


How to read the NHC cone of uncertainty



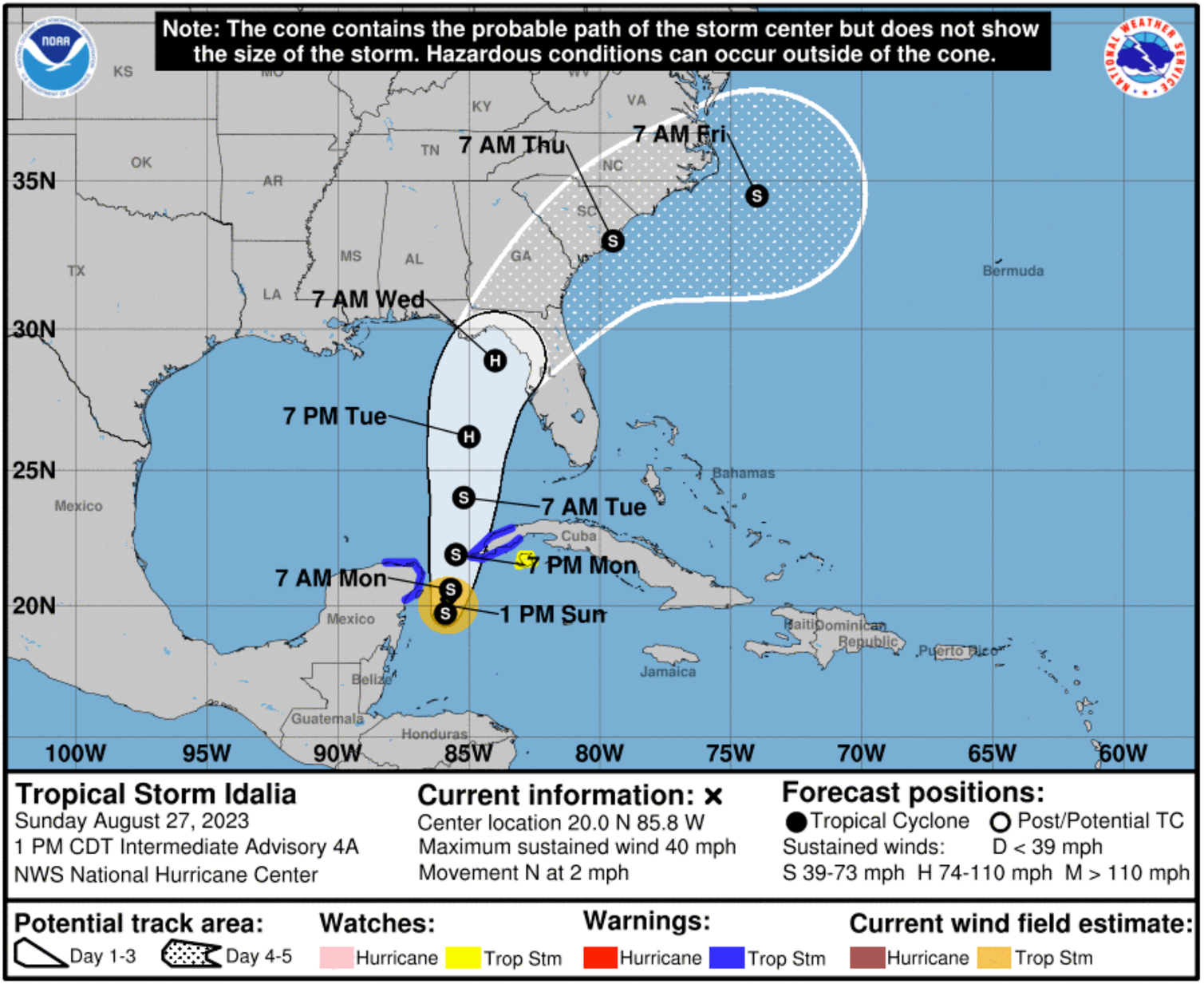
What do readers *really* need to know?

The possible path?

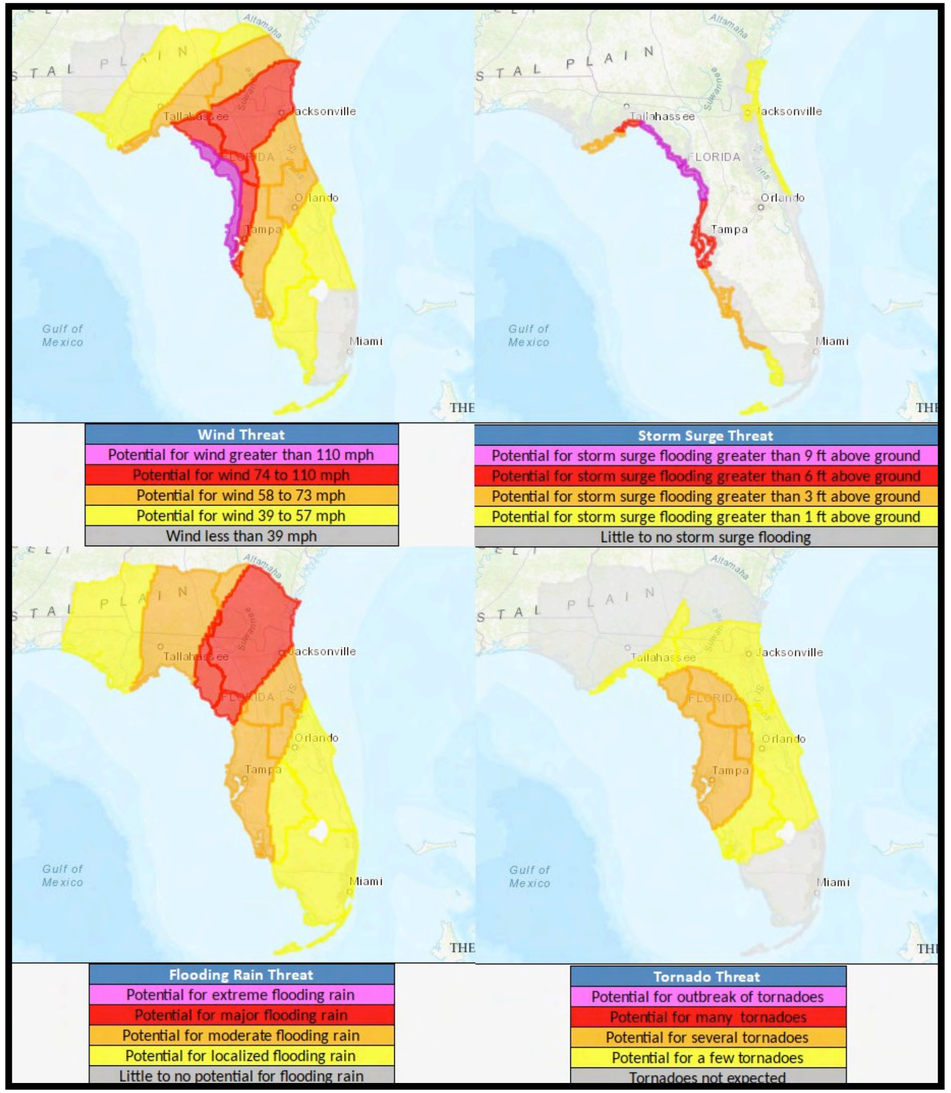


What do readers *really* need to know?

The possible path?



Or a hurricane's possible impacts (HTI)?

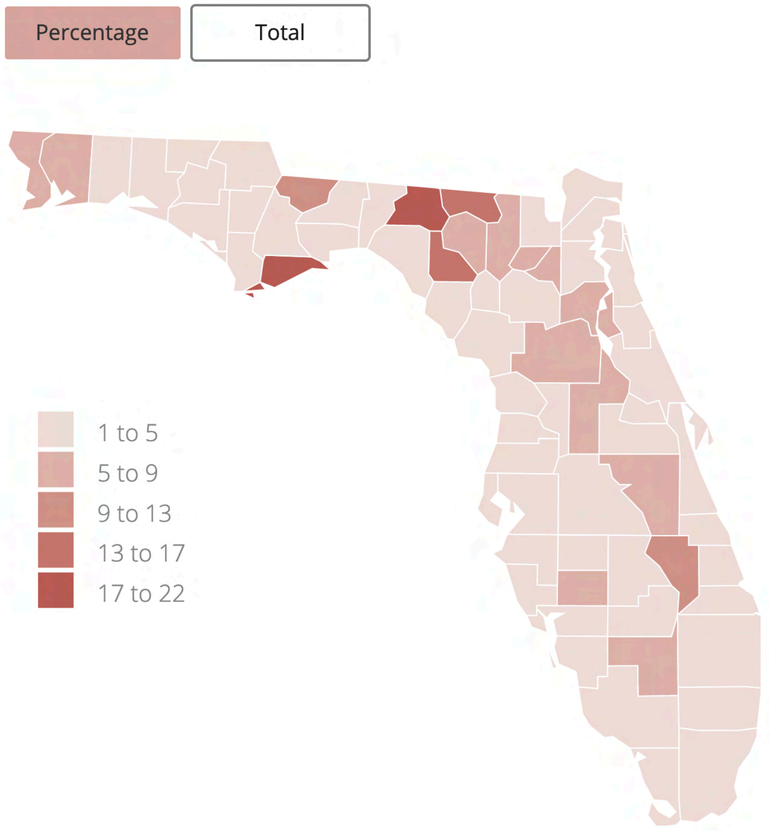


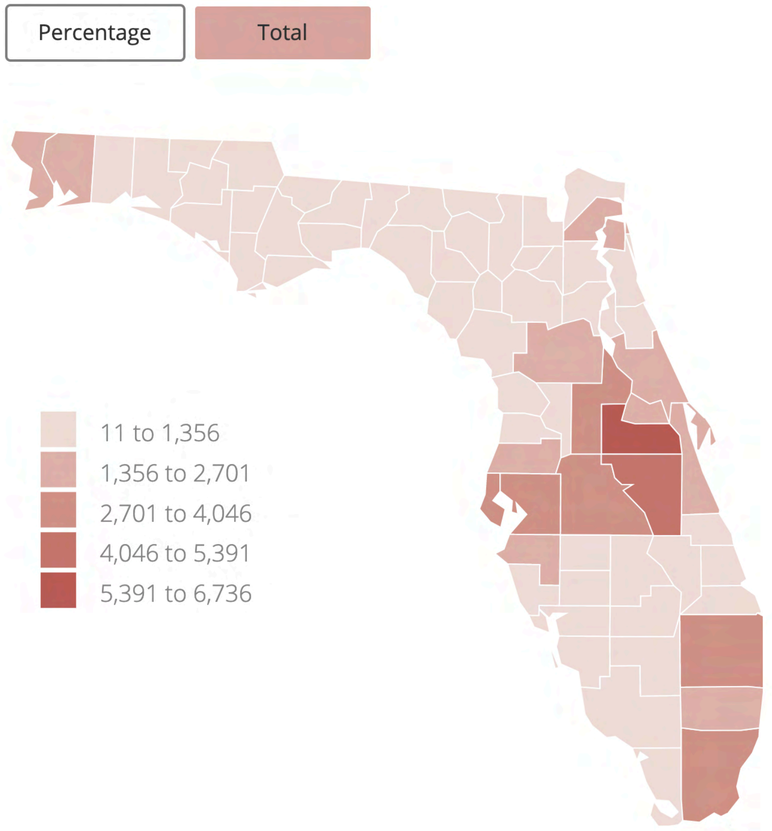
Charts can mislead even if they are well designed

Luís Melgar

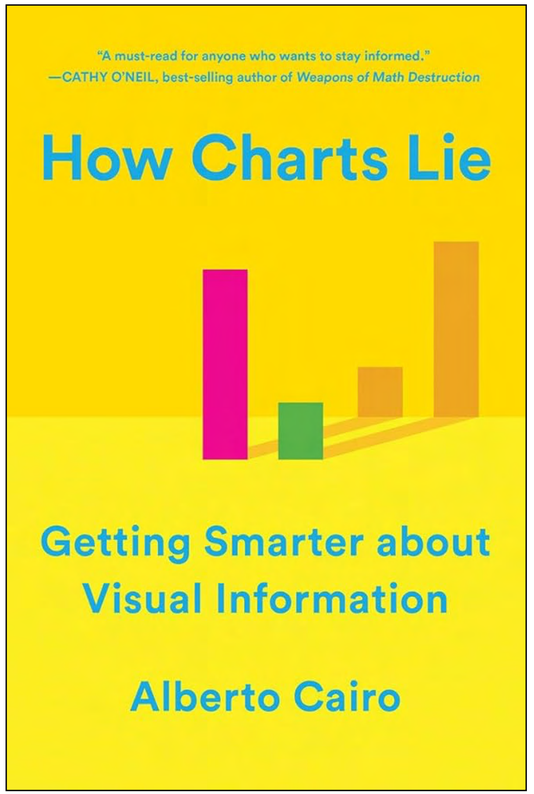
AT SCHOOL WITHOUT A ROOF

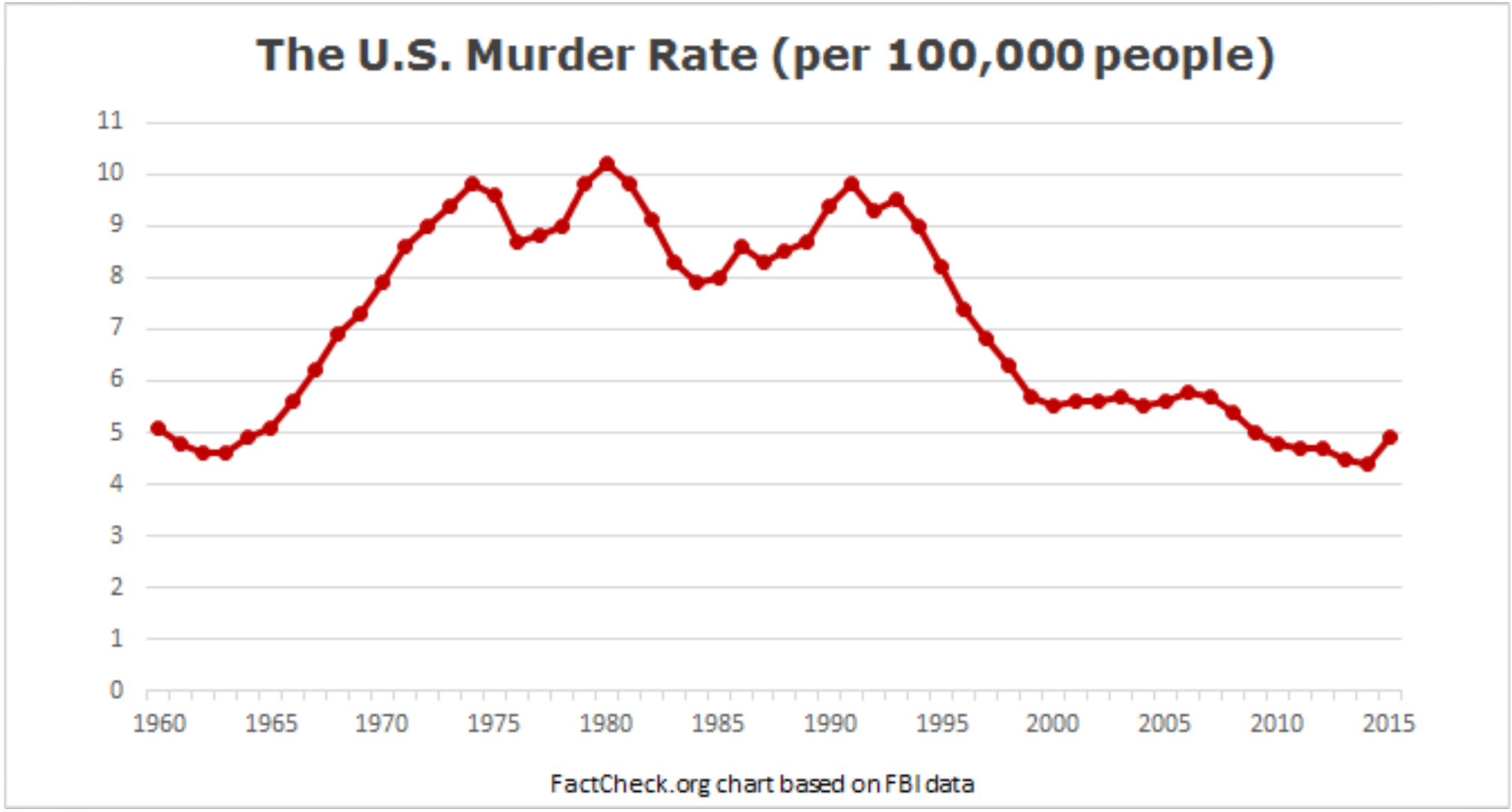
In Florida more than 71,000 students are homeless. During the last decade, this population rocketed as a result of the recession and how hard it has become for the poorest families to find affordable housing.



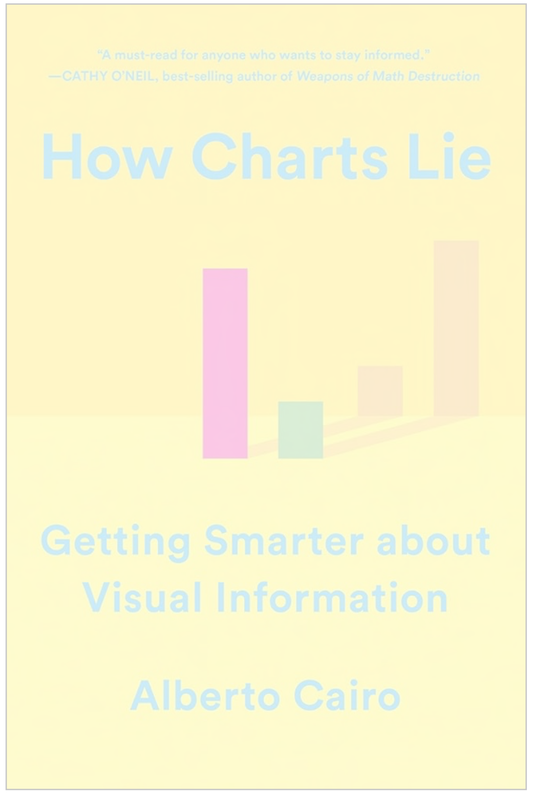


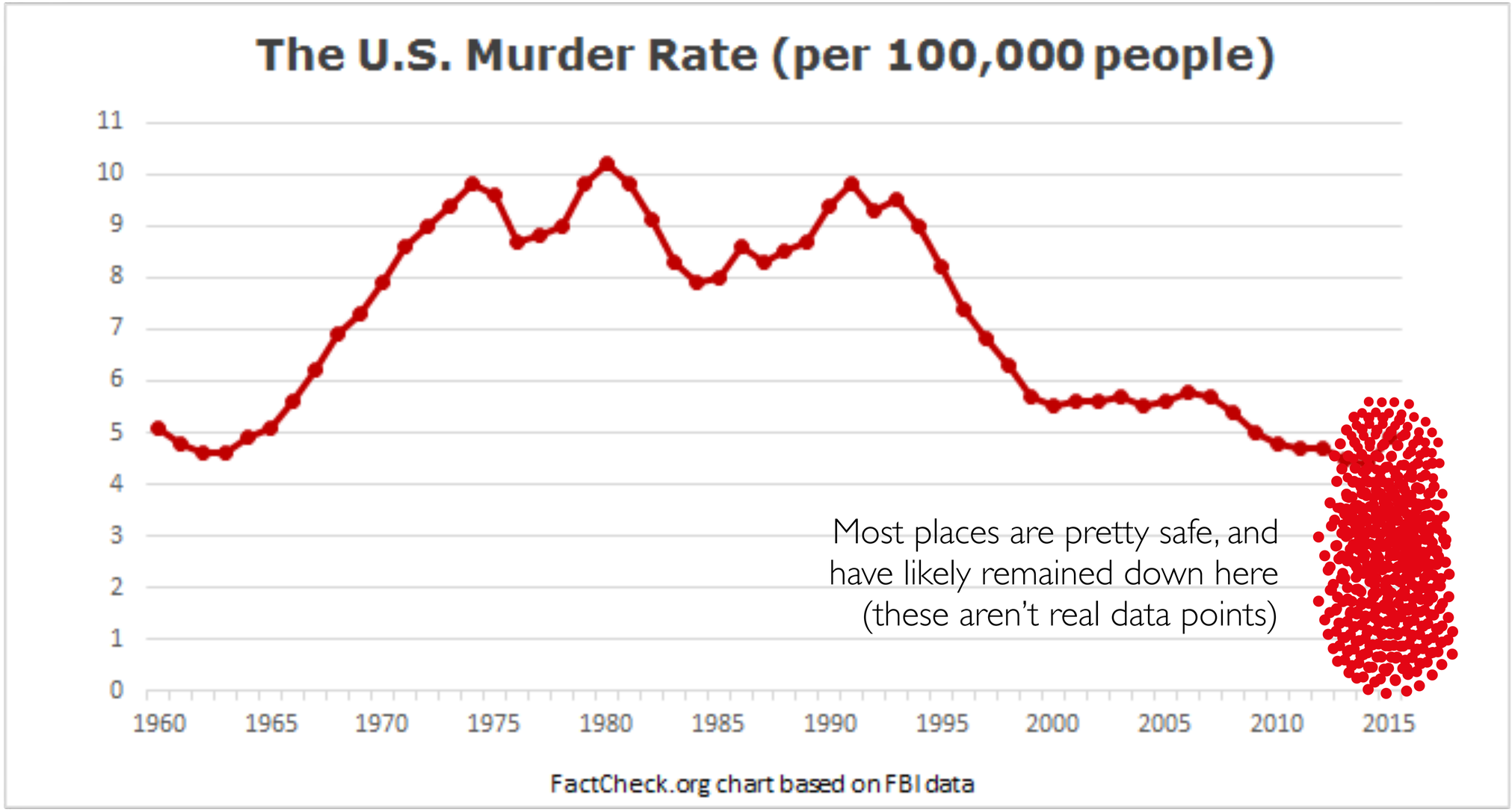
http://www.lmelgar.me/without-a-roof/


"A must-read for anyone who wants to stay informed."
—CATHY O'NEIL, best-selling author of Weapons of Math Destruction
How Charts Lie
Getting Smarter about Visual Information
Alberto Cairo

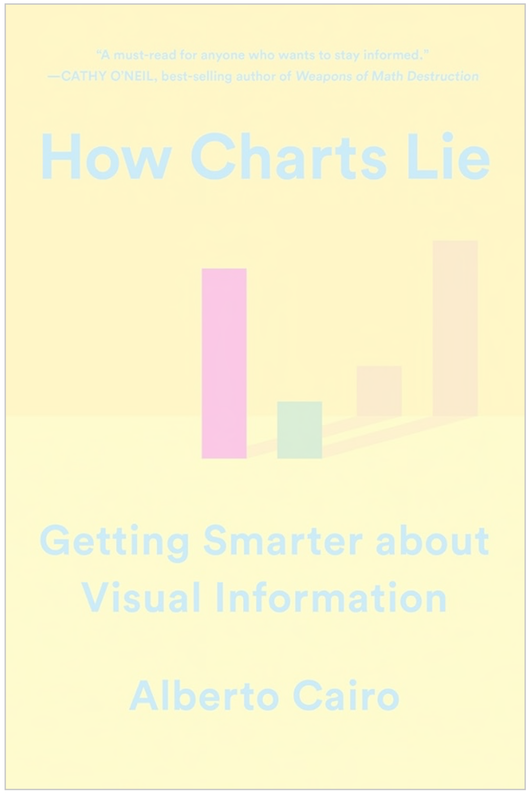

The U.S. Murder Rate (per 100,000 people)
11
10
9
8
7
6
5
4
3
2
1
0
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
FactCheck.org chart based on FBI data

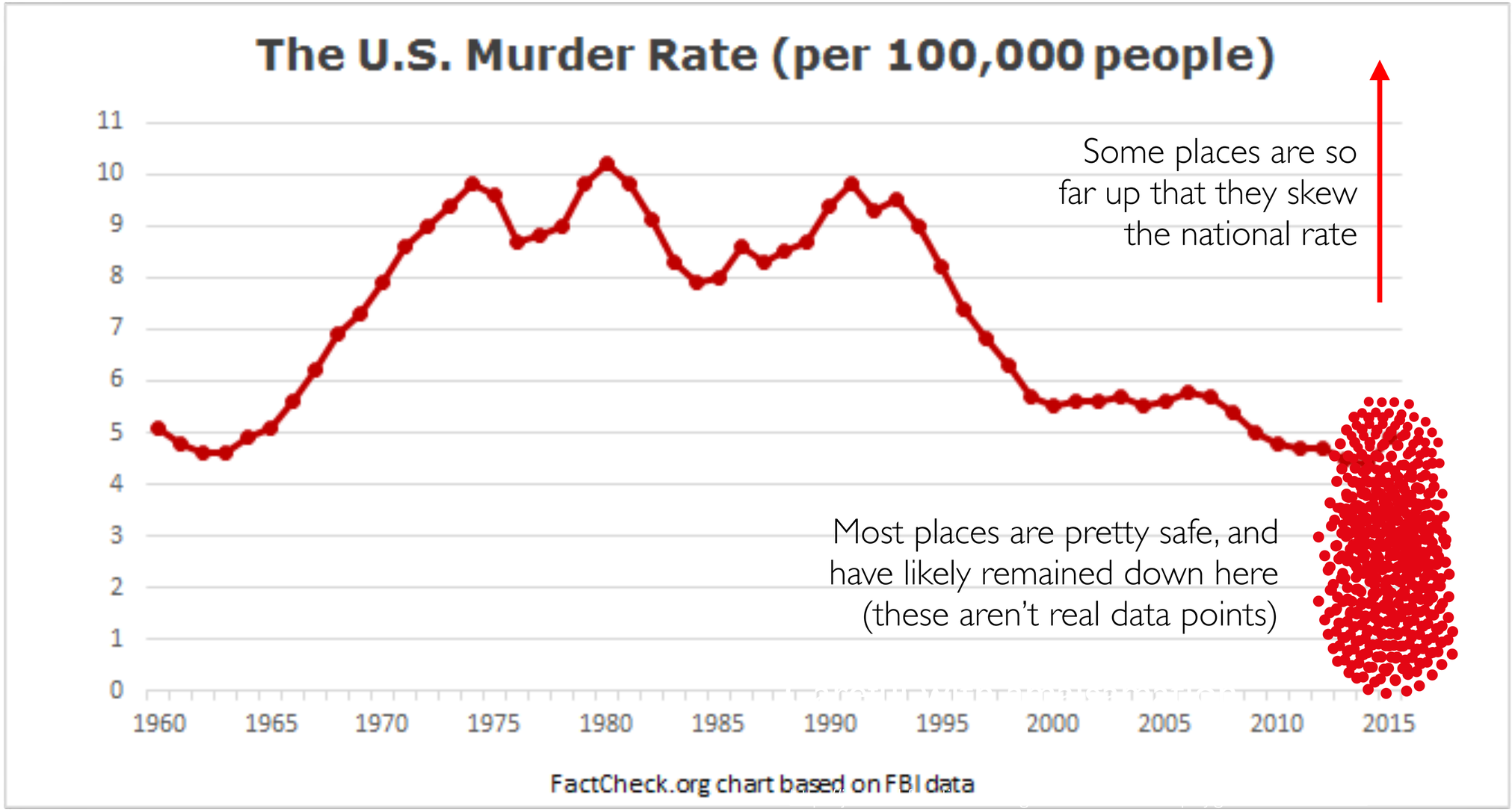
The danger of aggregating data too much, and presenting just averages and other statistical summaries

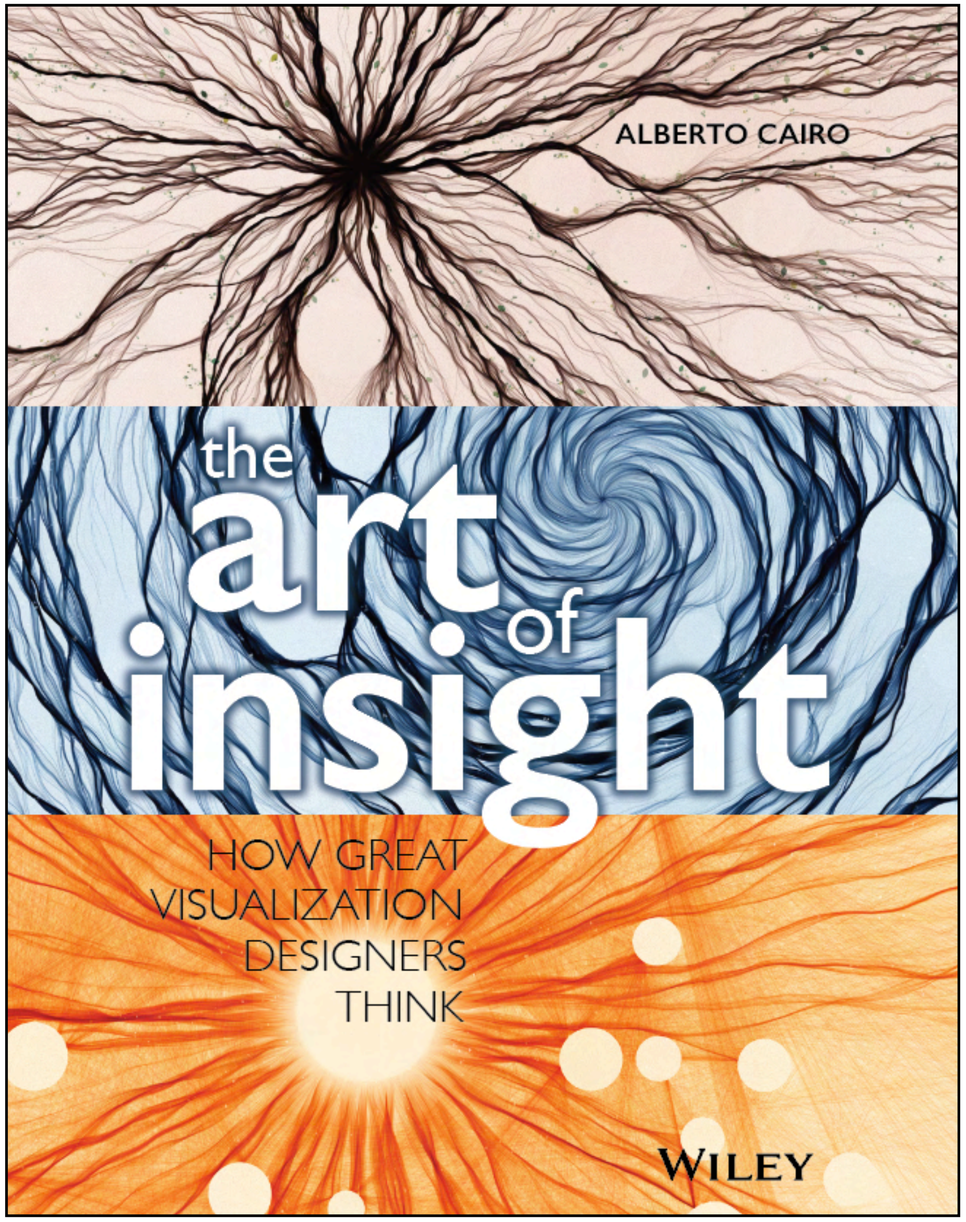




The danger of aggregating data too much, and presenting just averages and other statistical summaries







Visualization design consists of **reasoning** about possible **choices** by considering the interplay between:

1. **Content:** The nature, origin, and limitations of the data.
2. **People:** Your audience.
3. **Intent:** The purpose(s) that we define.
4. **Constraints:** The limitations that we may face.
5. **Outcomes:** How the graphic is received.

Every design choice must be **deliberate.**

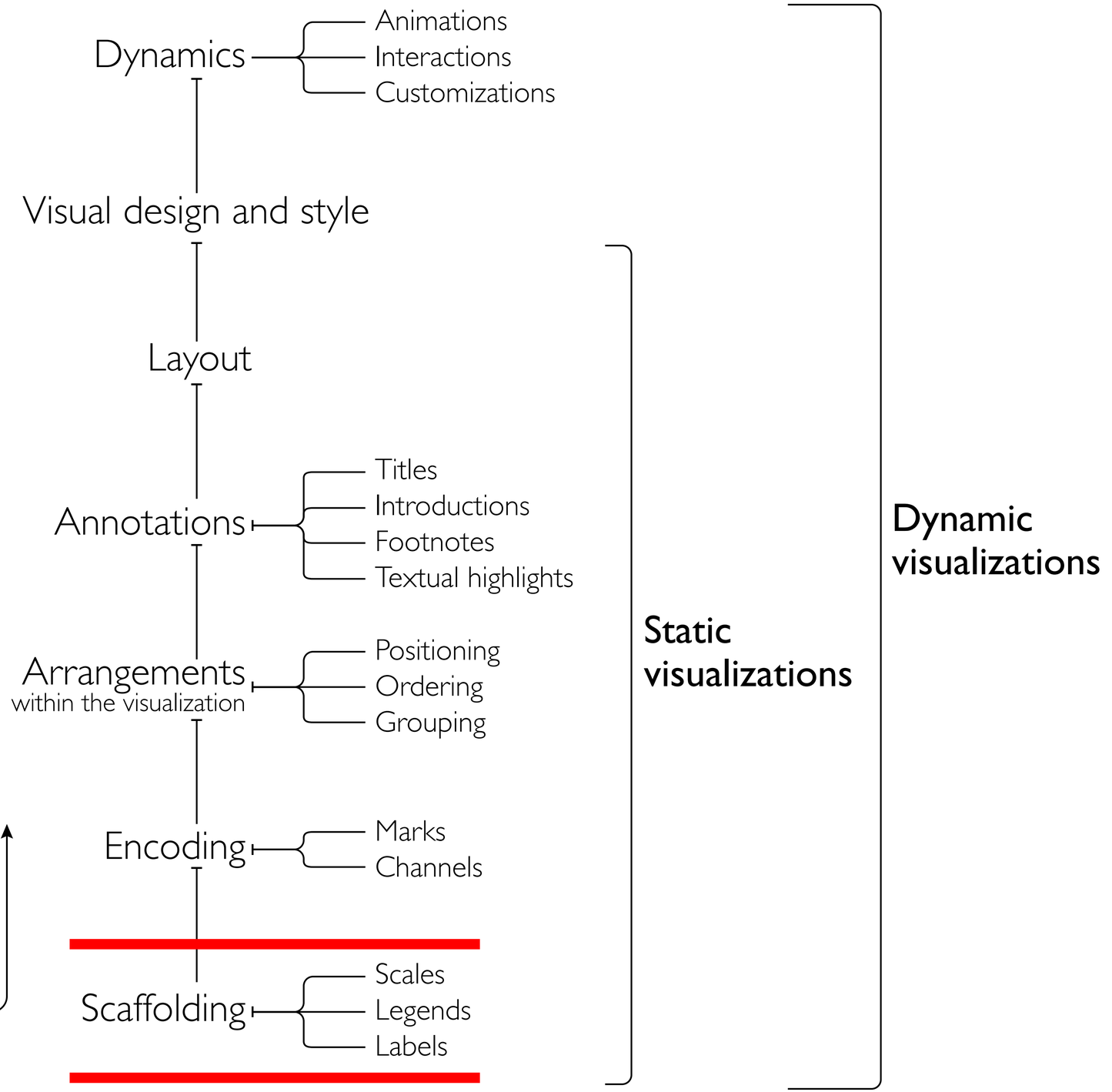
It is inevitably **subjective**, but it should never be **arbitrary**.

Reasoning about the components of a visualization

Visualization: Layers and elements to think about

Considerations about the information to be visualized —they influence everything else.

Read from the bottom-up



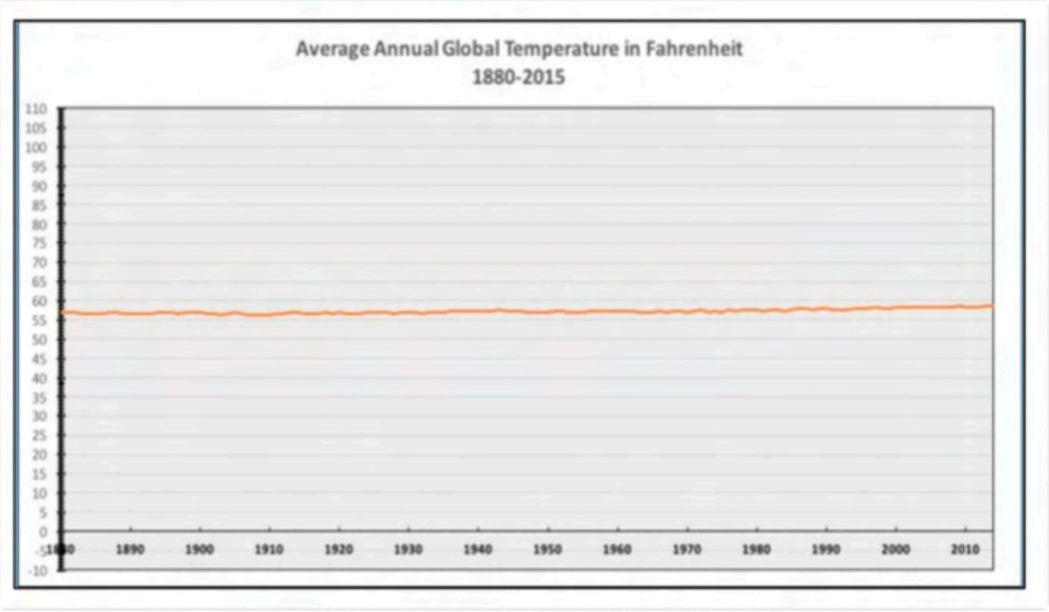
Decisions related to scales and legends can deeply affect our perception of the data

National Review ✓
@NRO

Follow

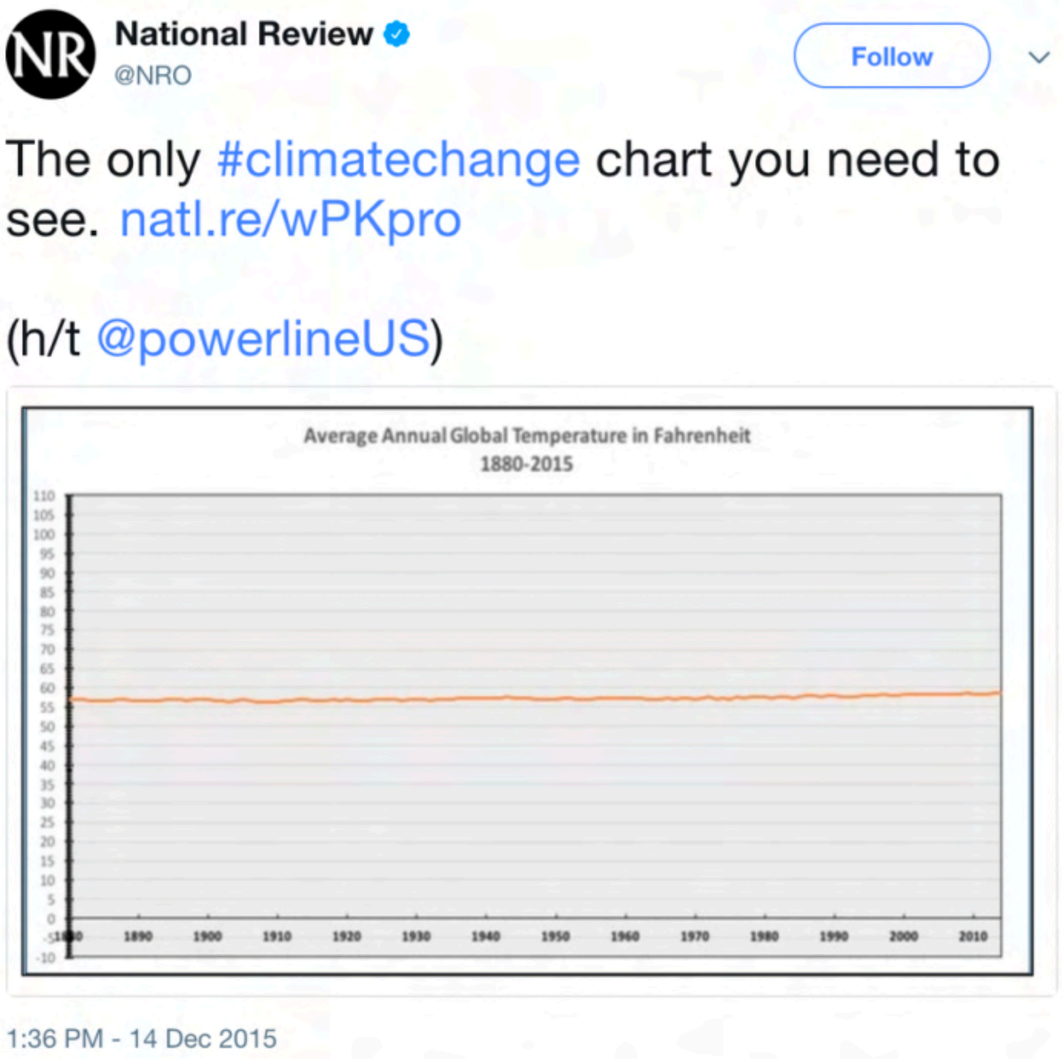
The only #climatechange chart you need to see. natl.re/wPKpro

(h/t @powerlineUS)

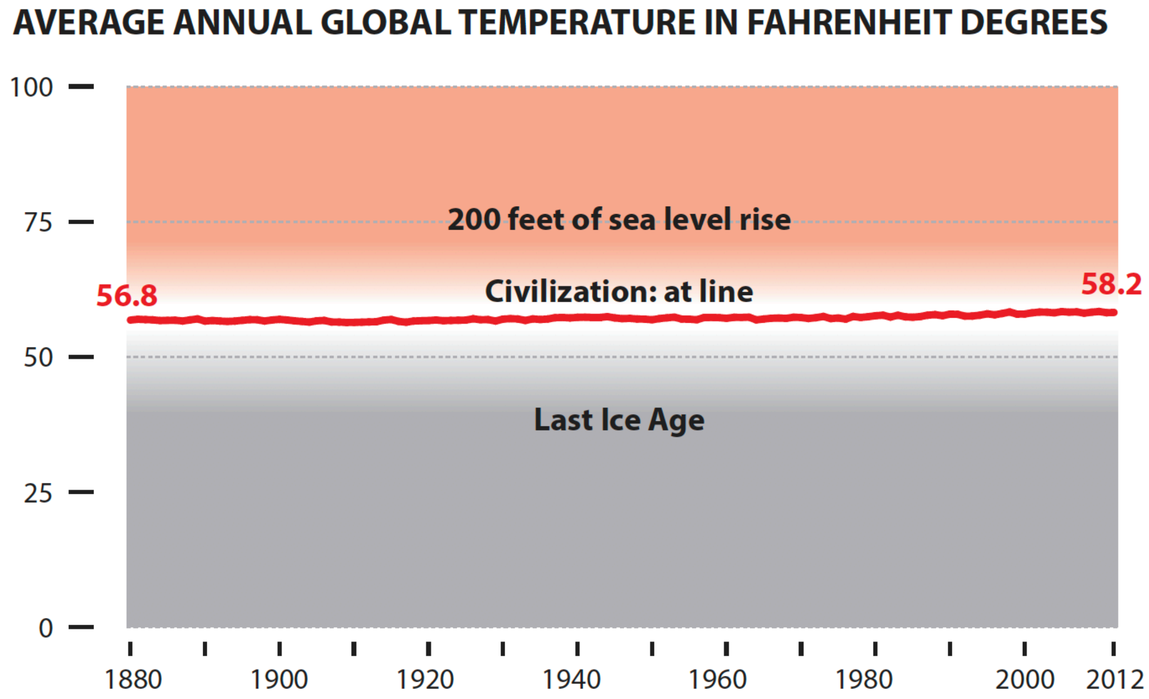


1:36 PM - 14 Dec 2015

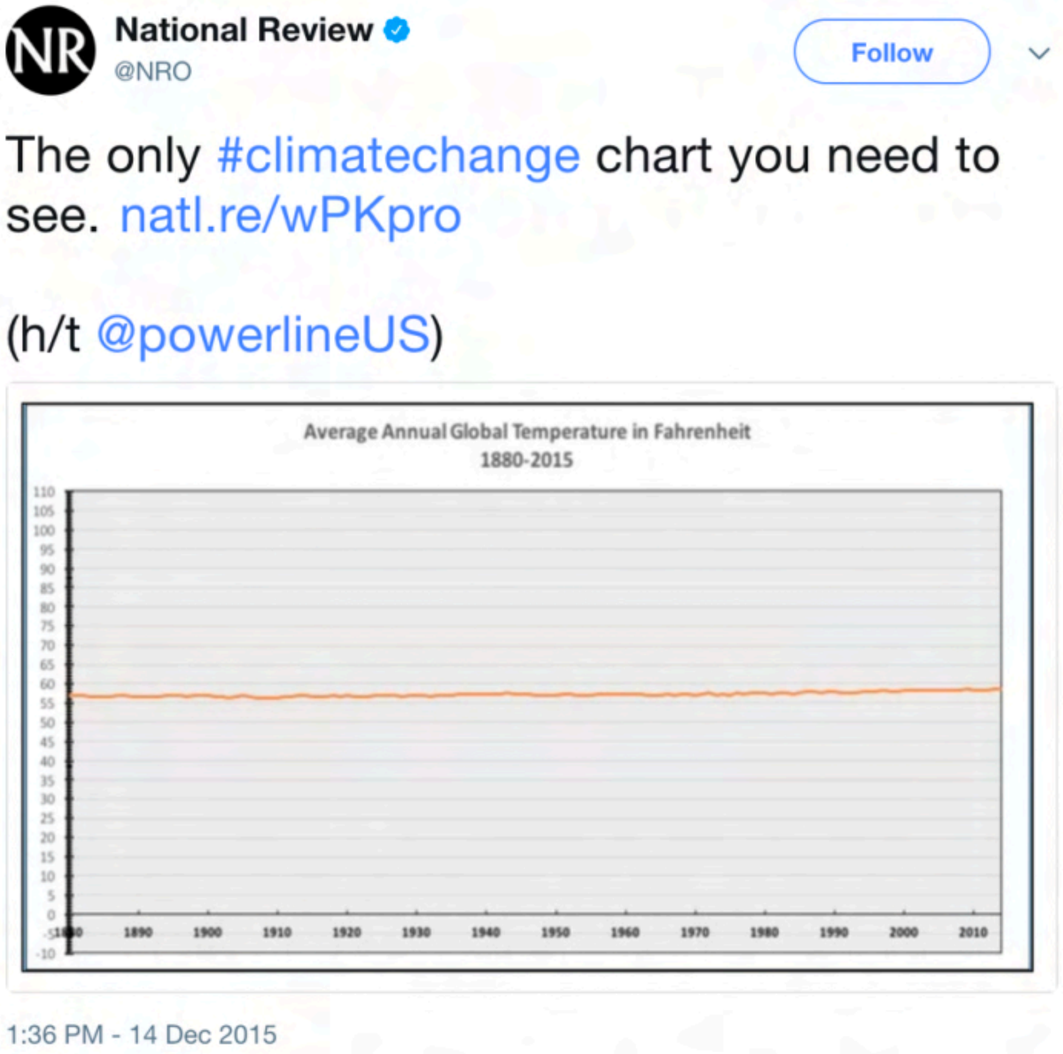
Decisions related to scales and legends can deeply affect our perception of the data

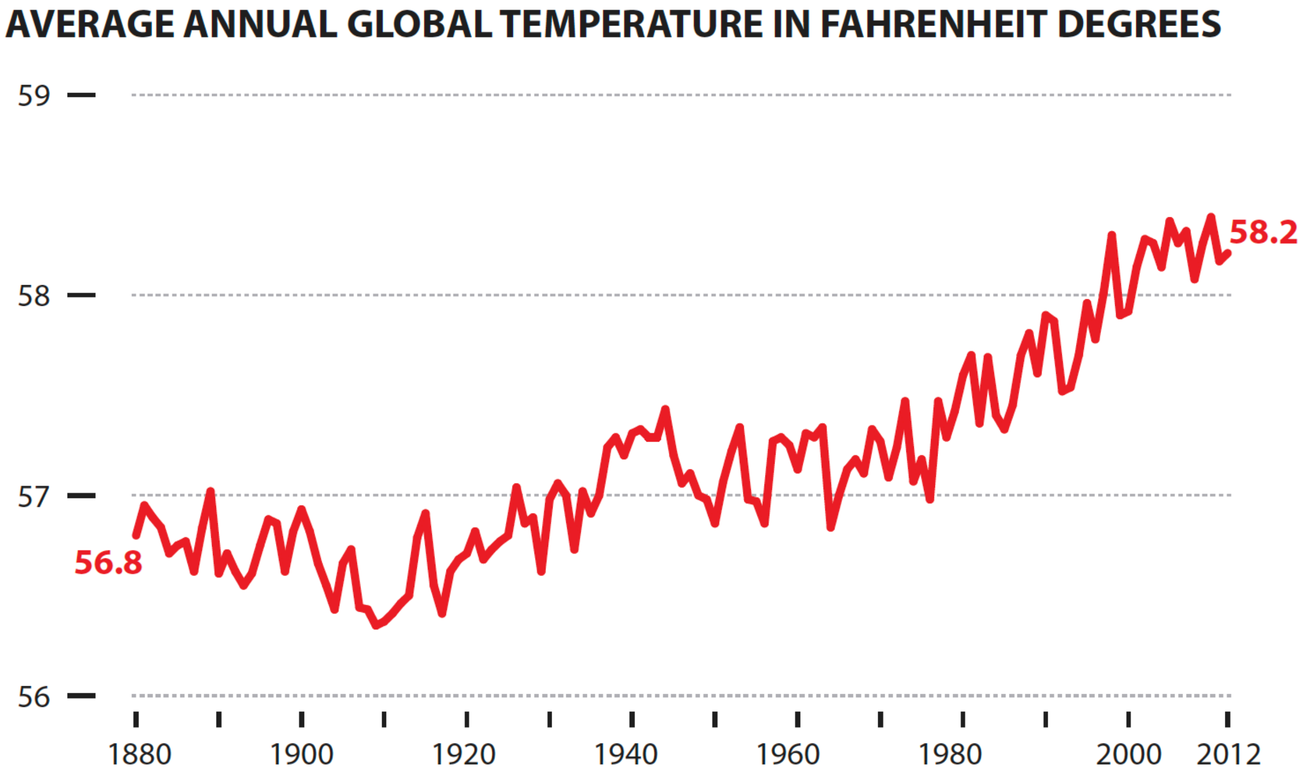


Richard Reiss, a research fellow at the City University of New York's Institute for Sustainable Cities added some tongue-in-cheek annotations to the original chart that reveal one of the many reasons why the choice of scales is so wrong:



Decisions related to scales and legends can deeply affect our perception of the data





A more reasonable Y-scale

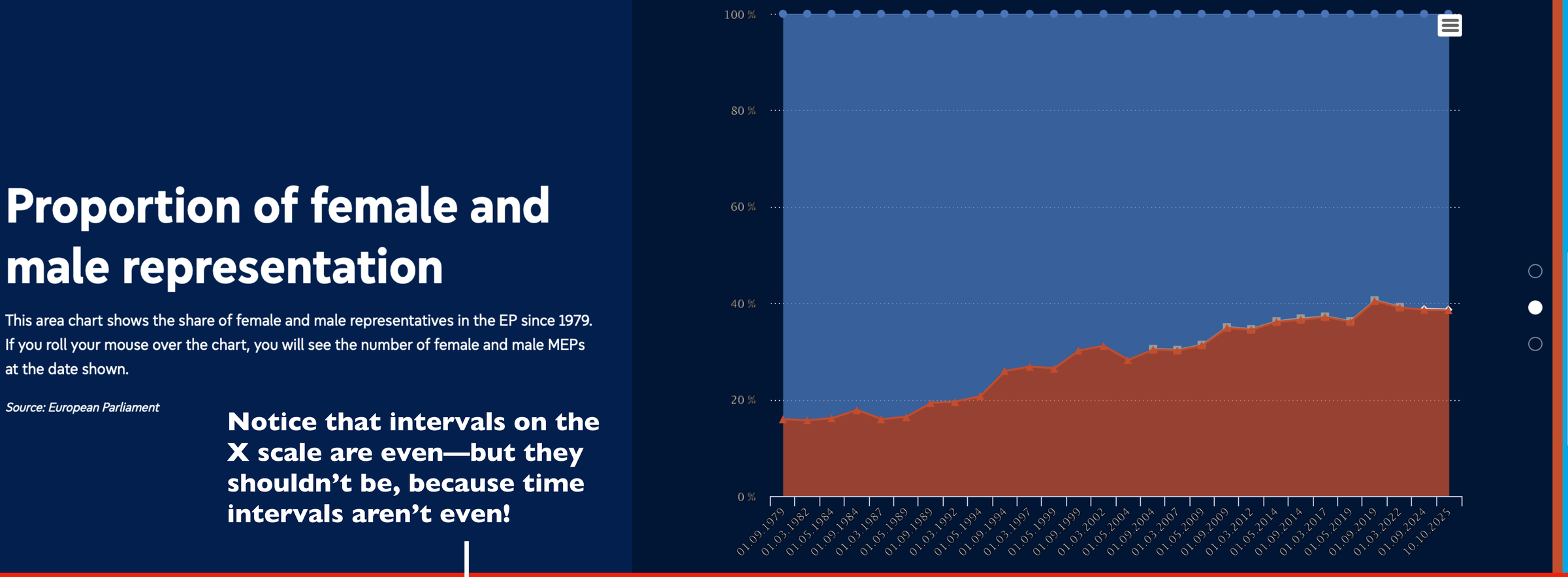
← Back to Homepage

Proportion of female and male representation

This area chart shows the share of female and male representatives in the EP since 1979. If you roll your mouse over the chart, you will see the number of female and male MEPs at the date shown.

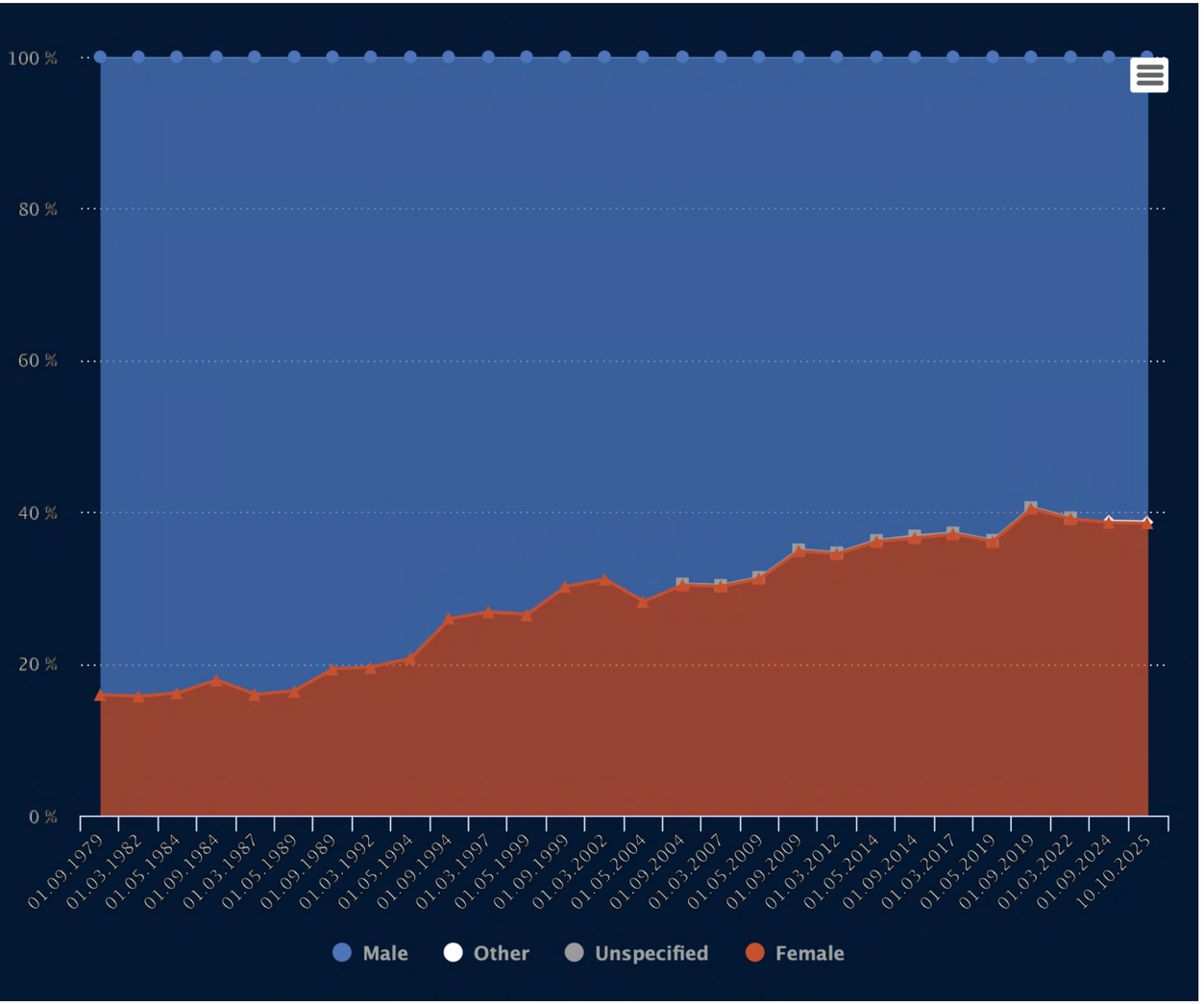
Source: European Parliament

Notice that intervals on the X scale are even—but they shouldn't be, because time intervals aren't even!

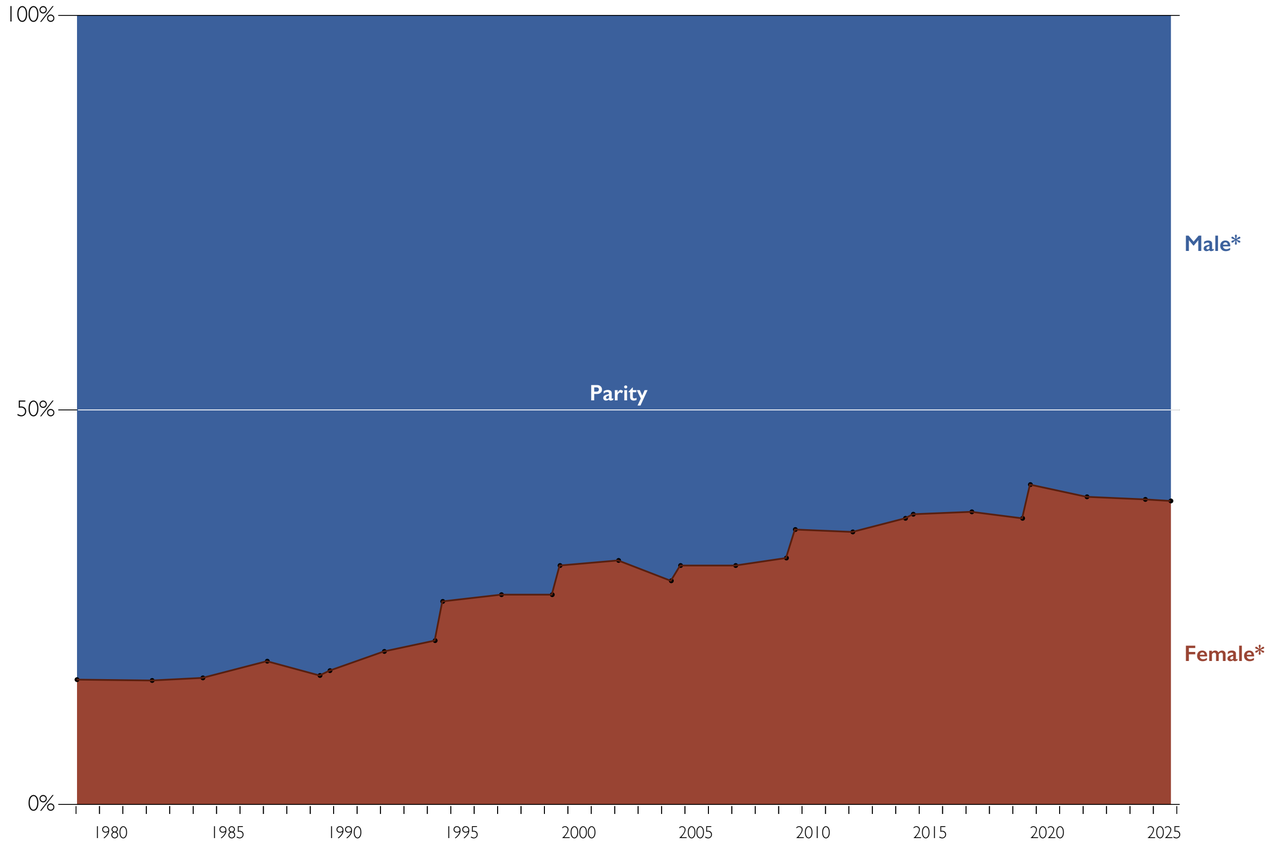


https://facts-and-figures.europarl.europa.eu/trends

Original

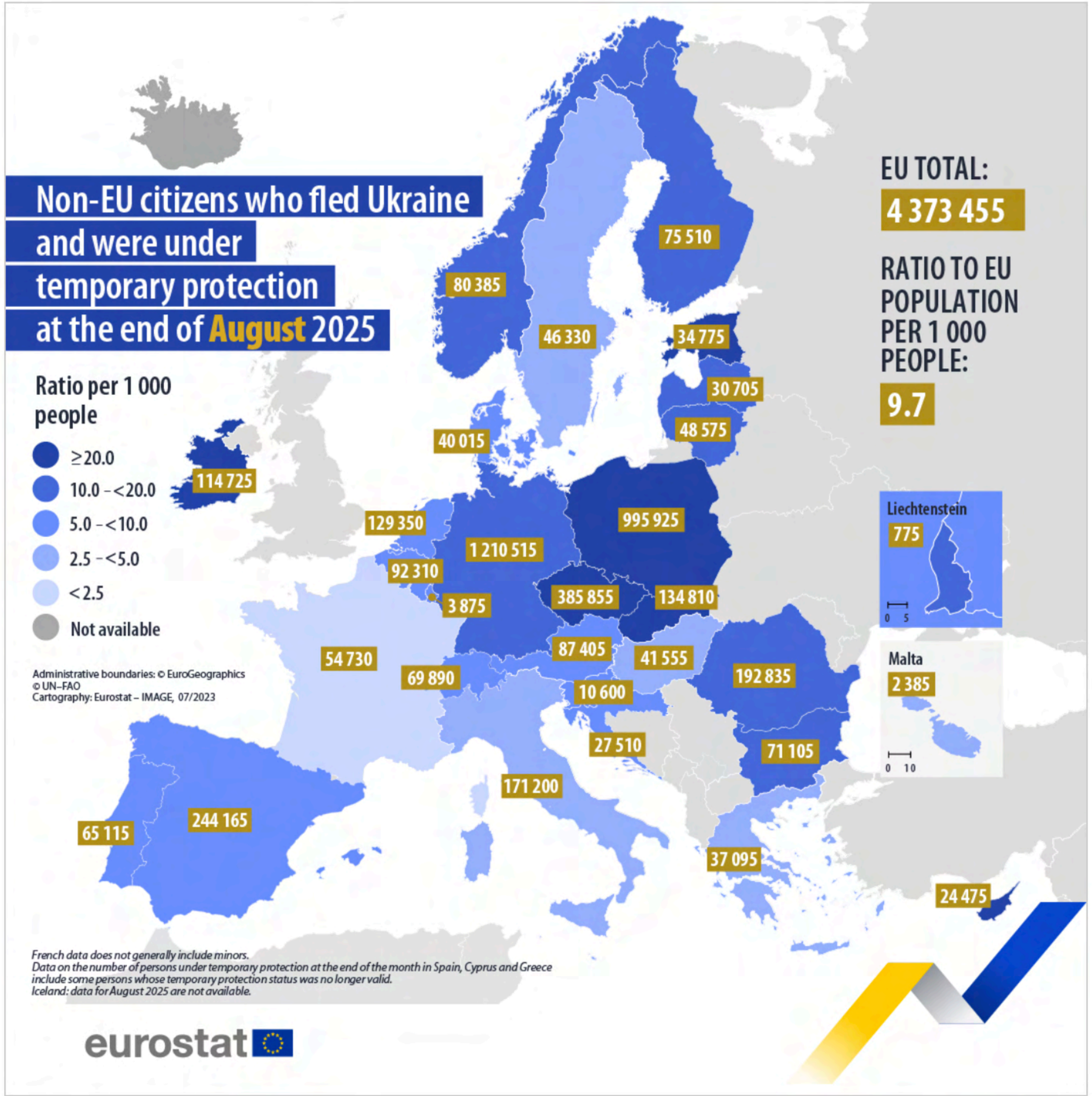


Alternative (warning: this is just a quick draft)



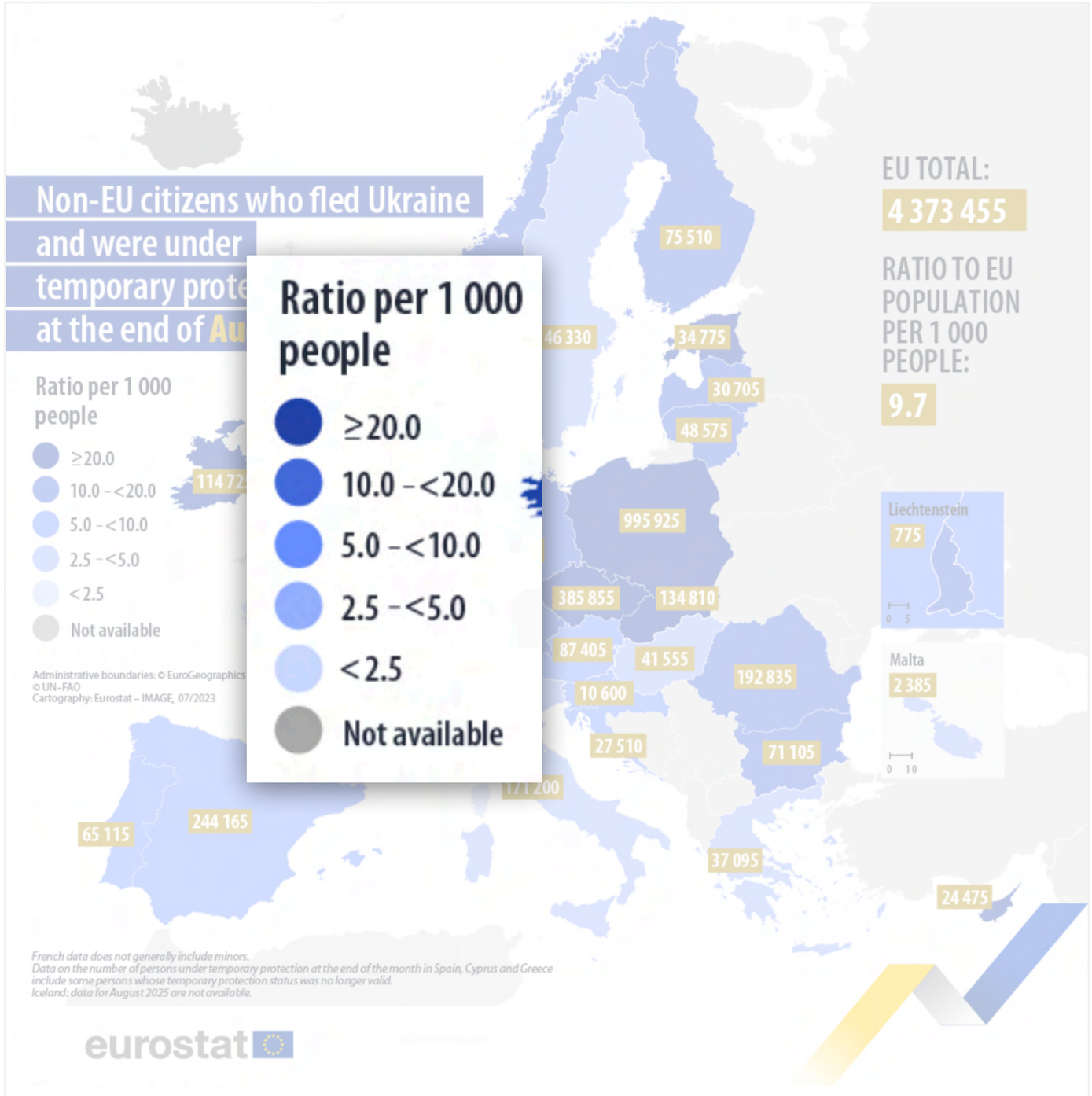
***Note about this quick-and-dirty chart makeover:** The original included the categories Other and Unspecified, corresponding to non-binary or transgender people. The percentages of these categories are too small to be represented accurately on a 0-100% scale. Therefore, an additional chart just for them would be needed.

https://facts-and-figures.europarl.europa.eu/trends



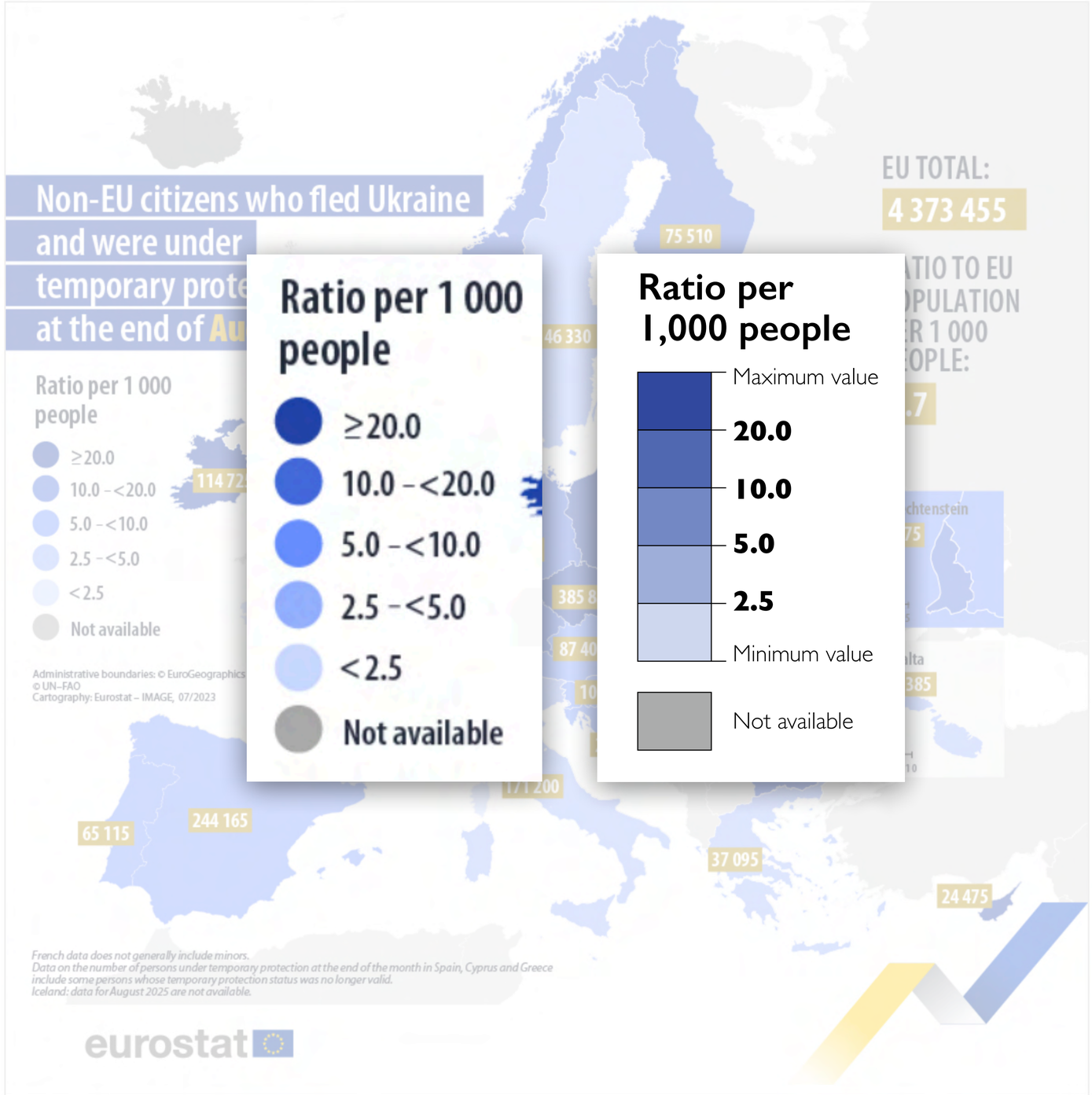
Map scales

https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Temporary_protection_for_persons_fleeing_Ukraine_-_monthly_statistics



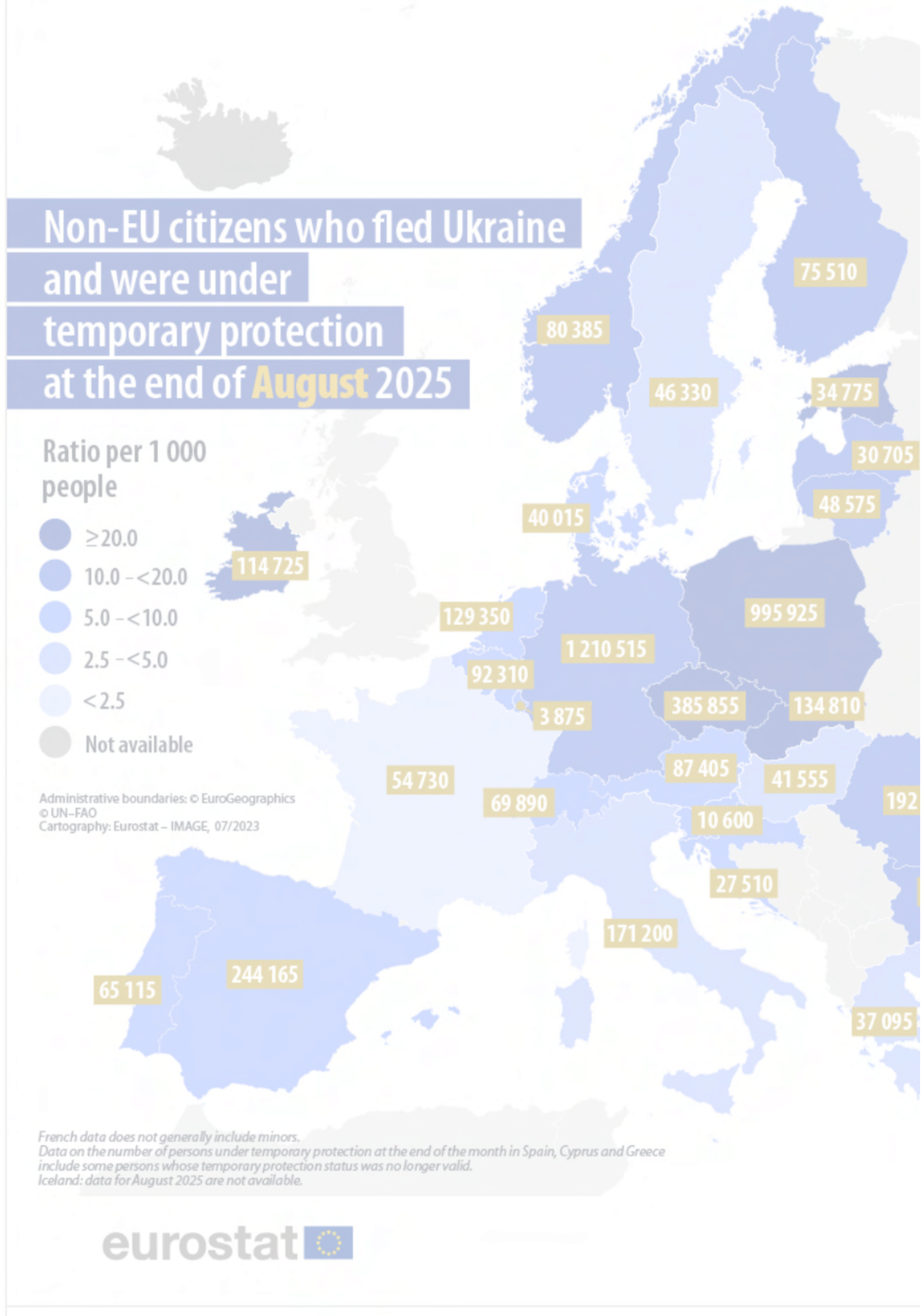
Map scales

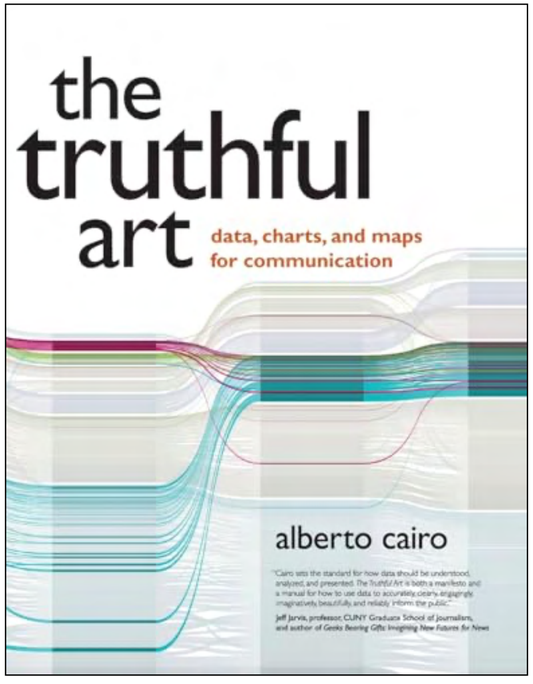
https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Temporary_protection_for_persons_fleeing_Ukraine_-_monthly_statistics

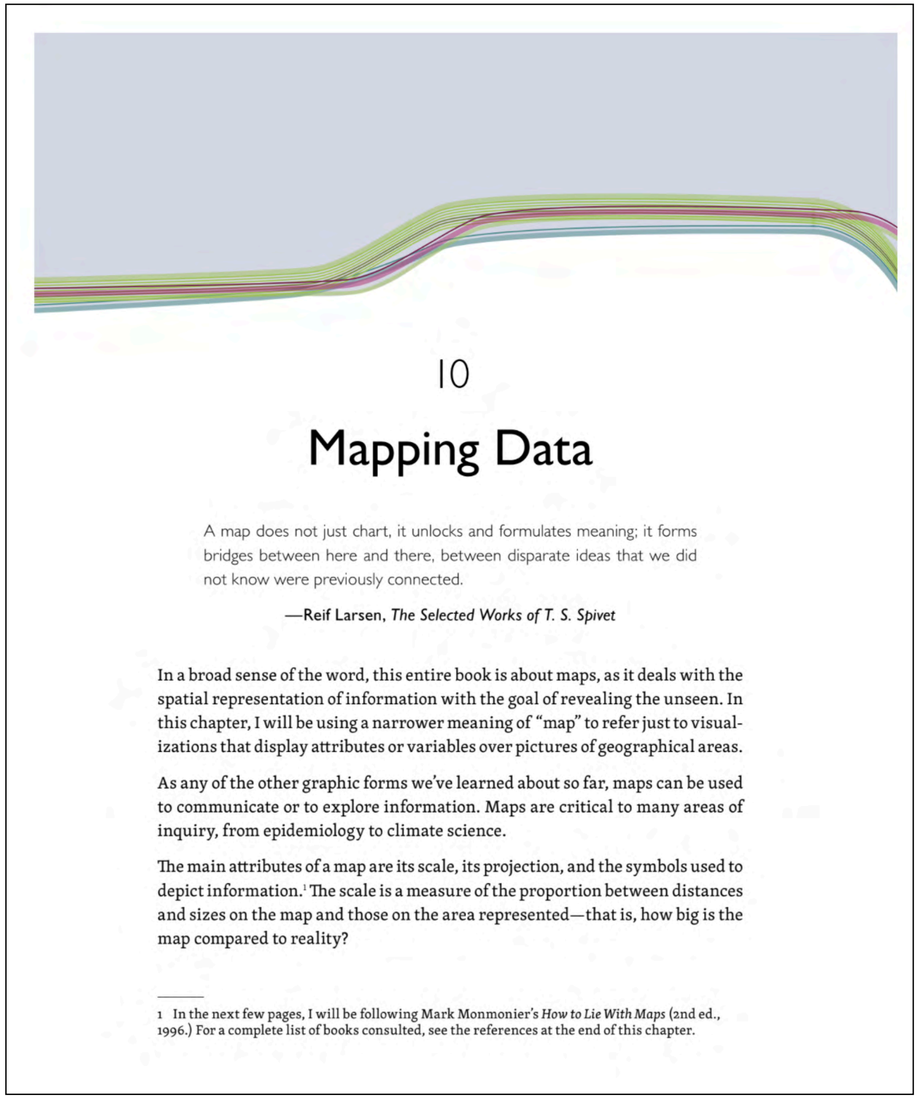


Map scales

https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Temporary_protection_for_persons_fleeing_Ukraine_-_monthly_statistics







10

Mapping Data

A map does not just chart, it unlocks and formulates meaning; it forms bridges between here and there, between disparate ideas that we did not know were previously connected.

—Reif Larsen, *The Selected Works of T. S. Spivet*

In a broad sense of the word, this entire book is about maps, as it deals with the spatial representation of information with the goal of revealing the unseen. In this chapter, I will be using a narrower meaning of "map" to refer just to visualizations that display attributes or variables over pictures of geographical areas.

As any of the other graphic forms we've learned about so far, maps can be used to communicate or to explore information. Maps are critical to many areas of inquiry, from epidemiology to climate science.

The main attributes of a map are its scale, its projection, and the symbols used to depict information.[1] The scale is a measure of the proportion between distances and sizes on the map and those on the area represented—that is, how big is the map compared to reality?

1 In the next few pages, I will be following Mark Monmonier's *How to Lie With Maps* (2nd ed., 1996.) For a complete list of books consulted, see the references at the end of this chapter.

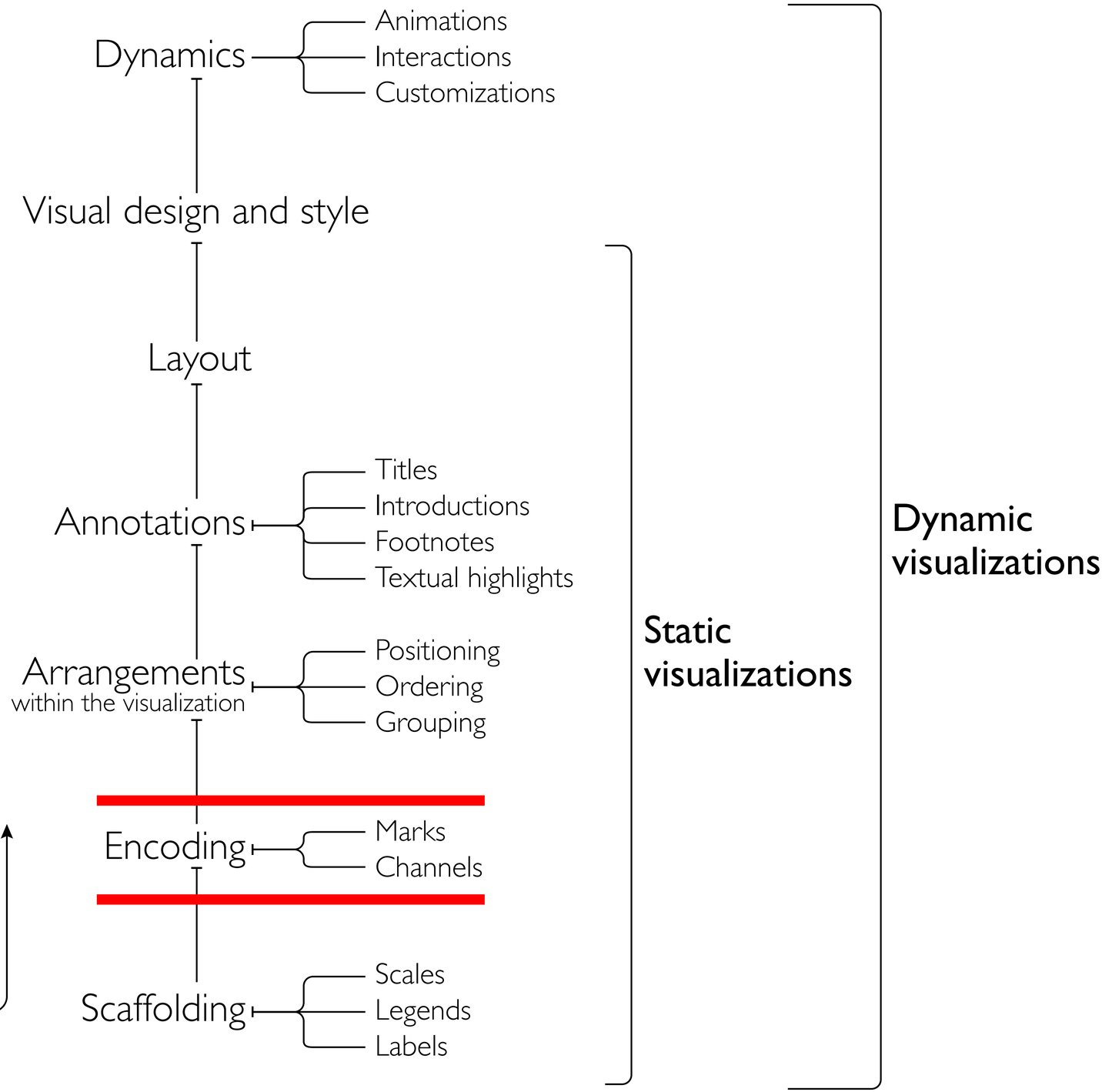
Chapter 10 of *The Truthful Art*

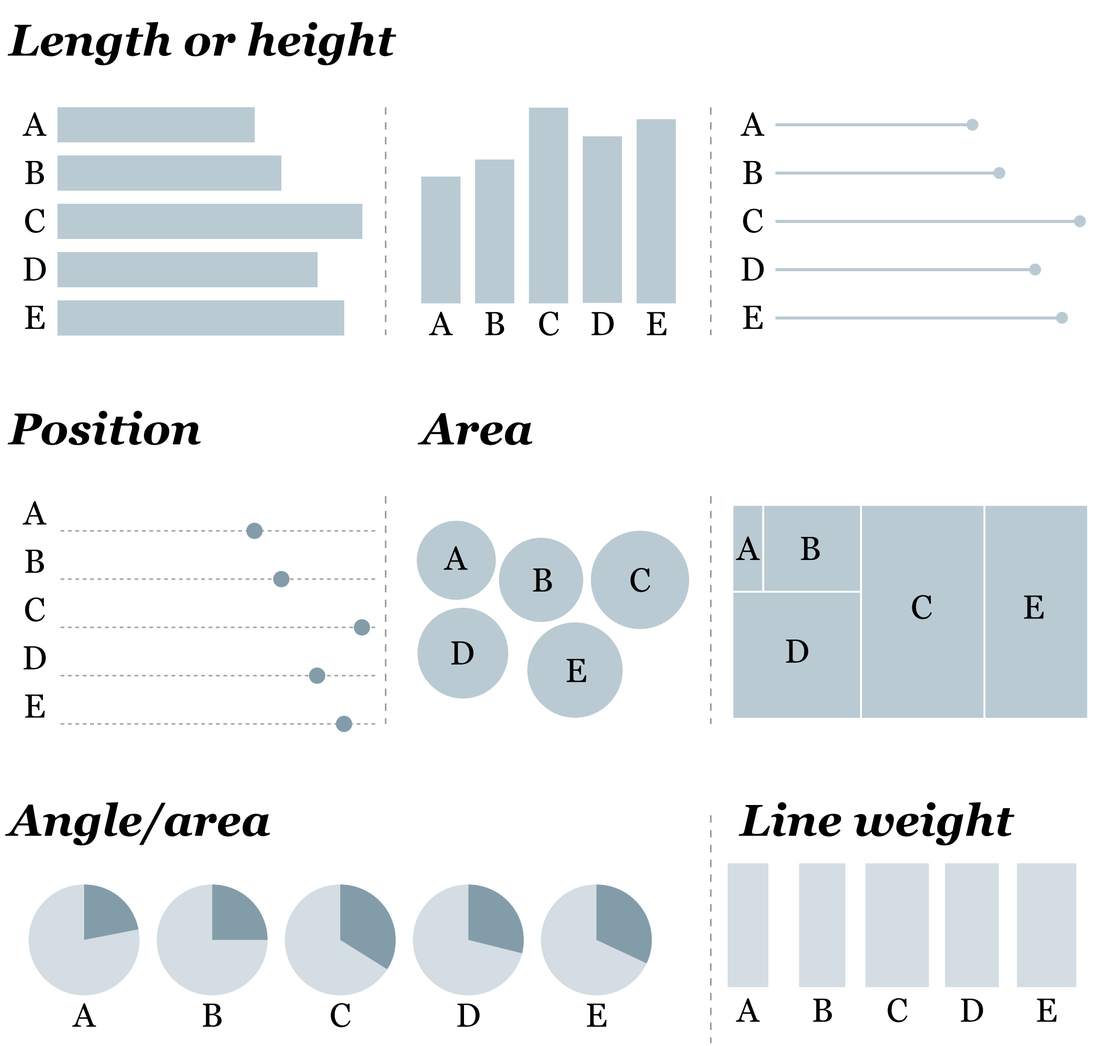
https://drive.google.com/file/d/1Ap4ehrBDXBSDPcnPXCTeg9cVH1U9kOeX/

Visualization: Layers and elements to think about

Considerations about the information to be visualized —they influence everything else.

Read from the bottom-up





Figures represented
in all these graphics:
22%, 25%, 34%, 29%, 32%

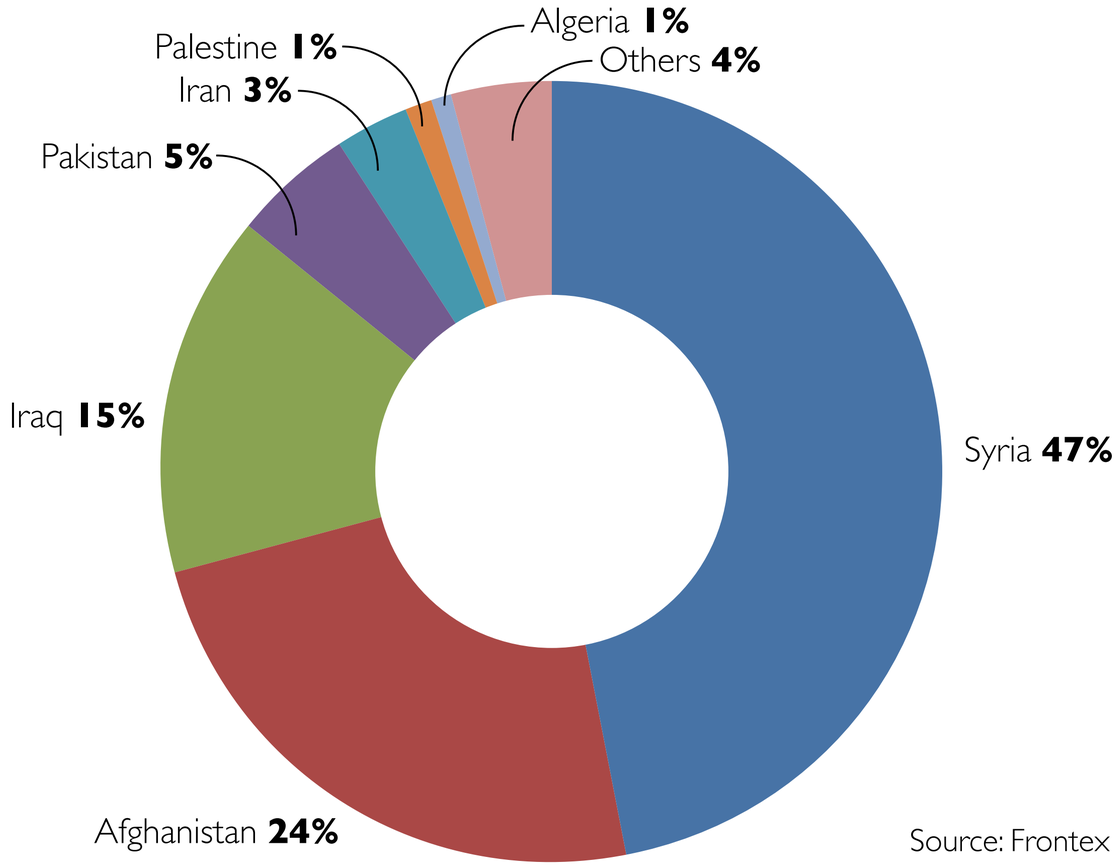
Encoding

Data visualization consists of mapping data onto attributes of objects—commonly abstract shapes, called **“marks”**.

These attributes that vary with the data are called **“visual channels”.**

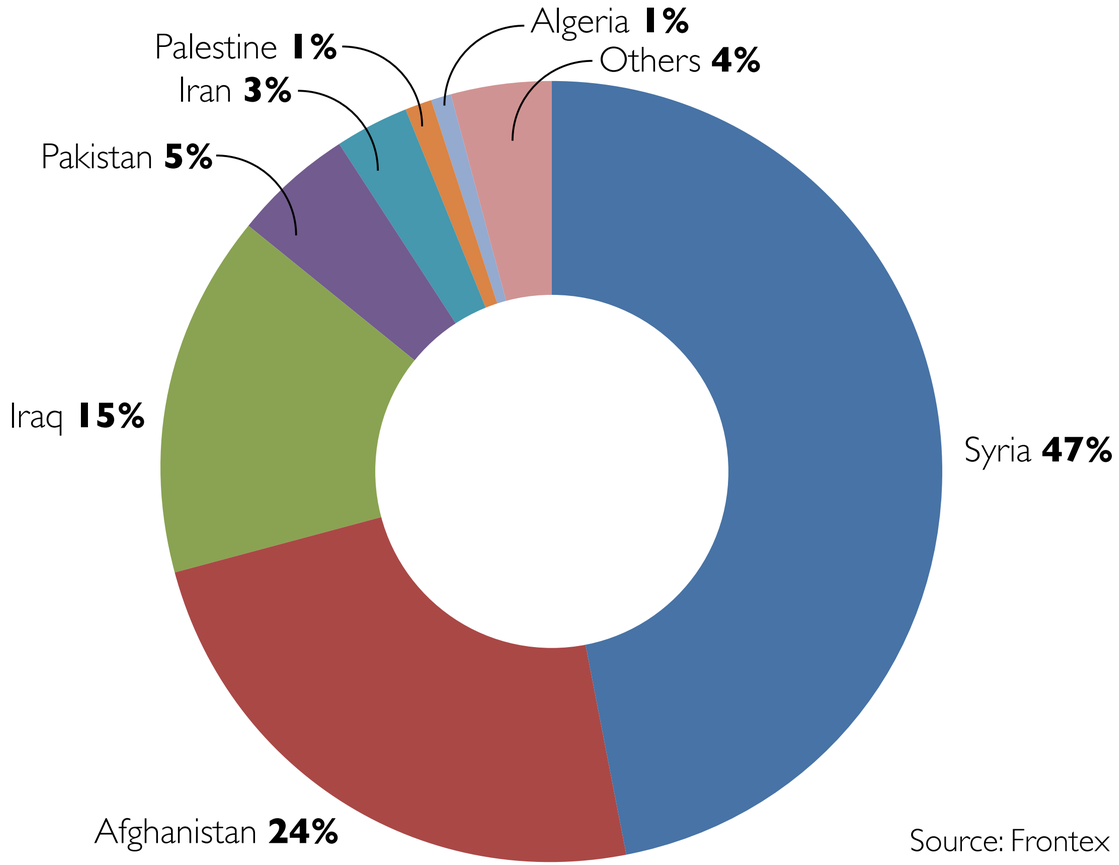
https://web.cse.ohio-state.edu/~shen.94/Melbourne/Slides/TamaraChp5.pdf

GREECE - Main nationalities of arriving migrants in 2016



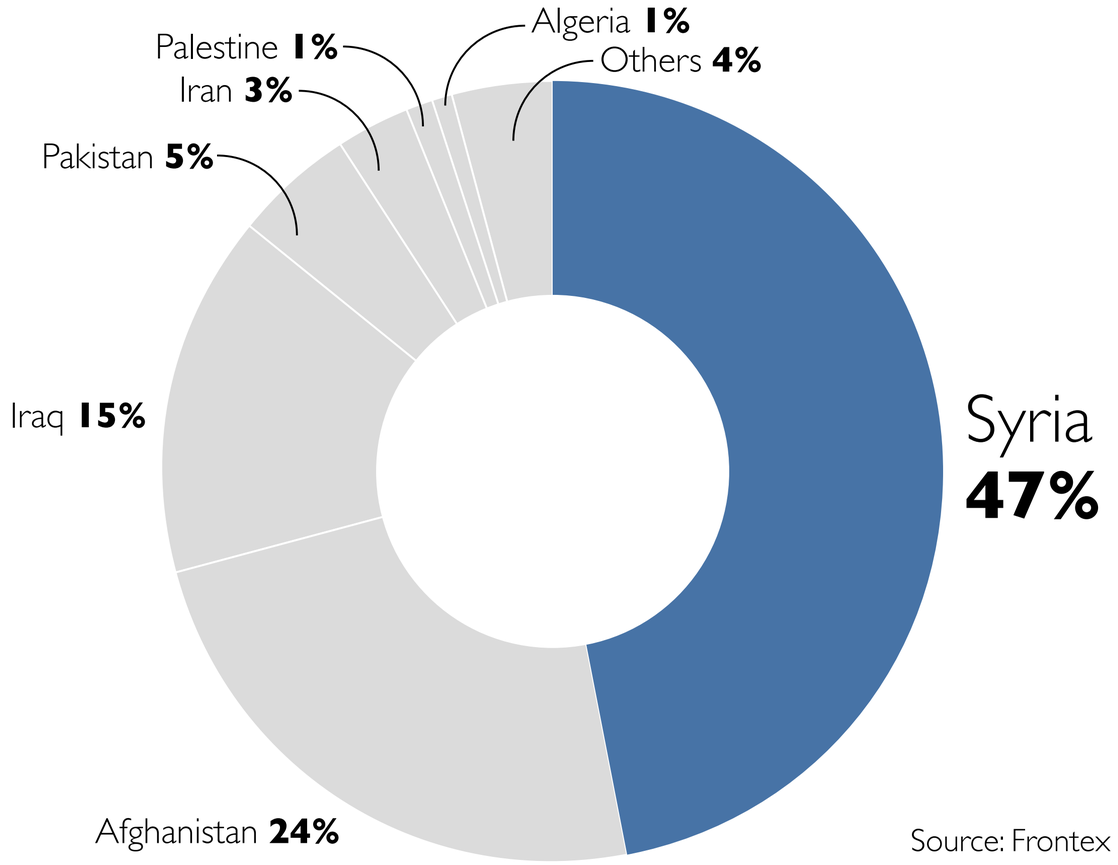
Source: Frontex

GREECE - Main nationalities of arriving migrants in 2016

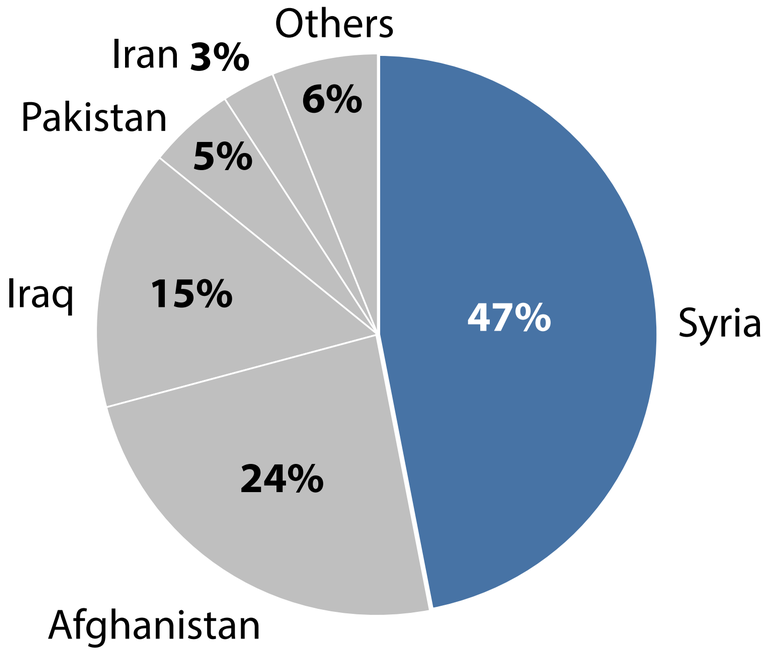


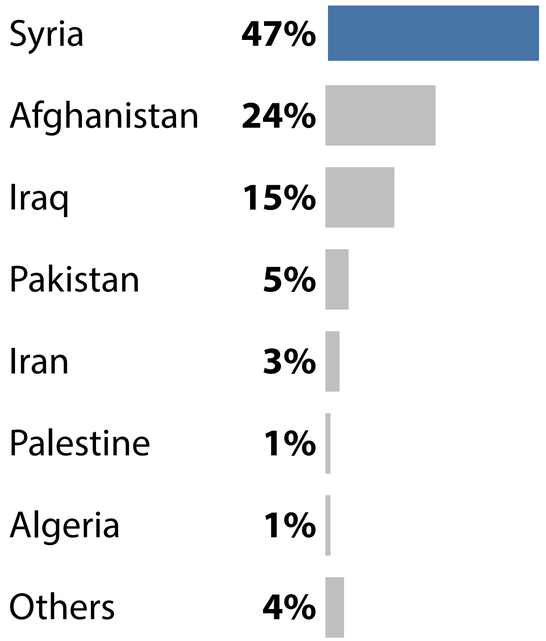
Source: Frontex

GREECE - Main nationalities of arriving migrants in 2016



Source: Frontex

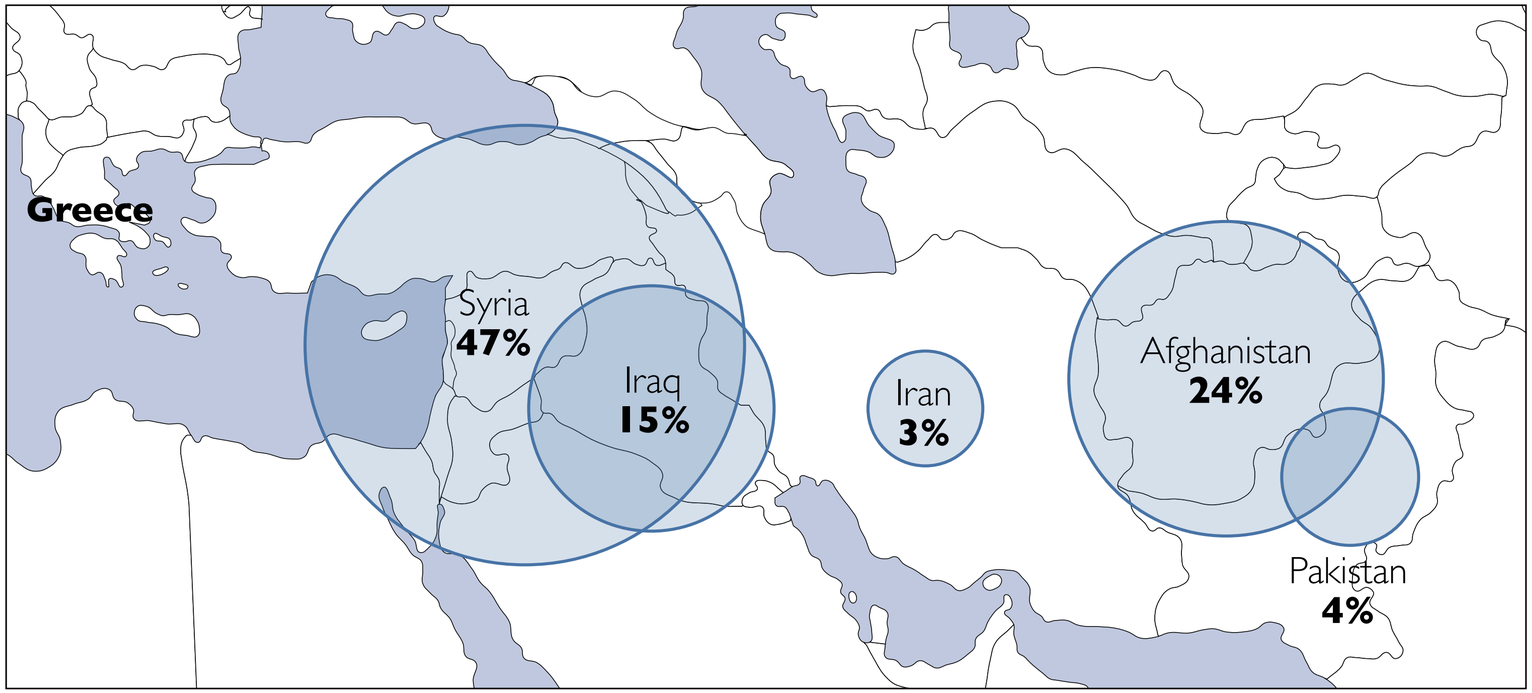

Others
Iran 3%
Pakistan
5%
Iraq
15%
47%
Syria
24%
Afghanistan
6%

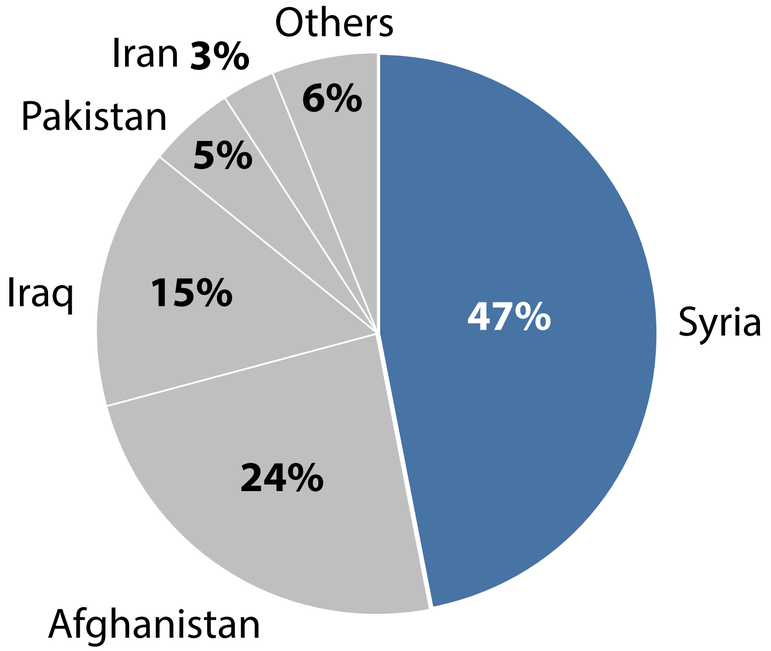

Syria 47%
Afghanistan 24%
Iraq 15%
Pakistan 5%
Iran 3%
Palestine 1%
Algeria 1%
Others 4%

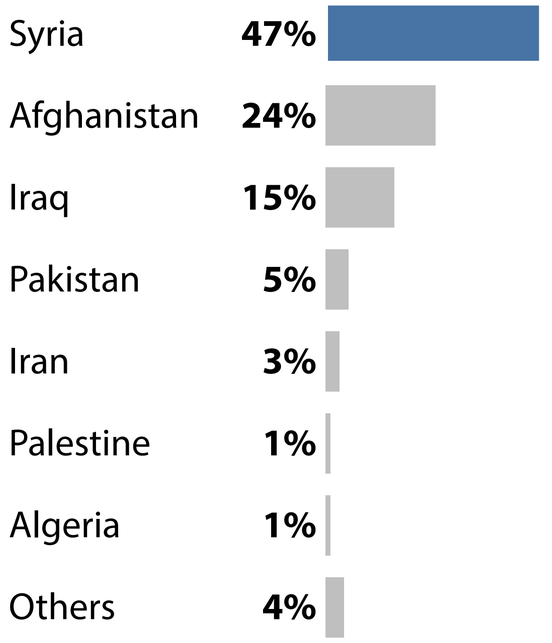
None of these charts is *right* or *wrong*, *good* or *bad* per se.

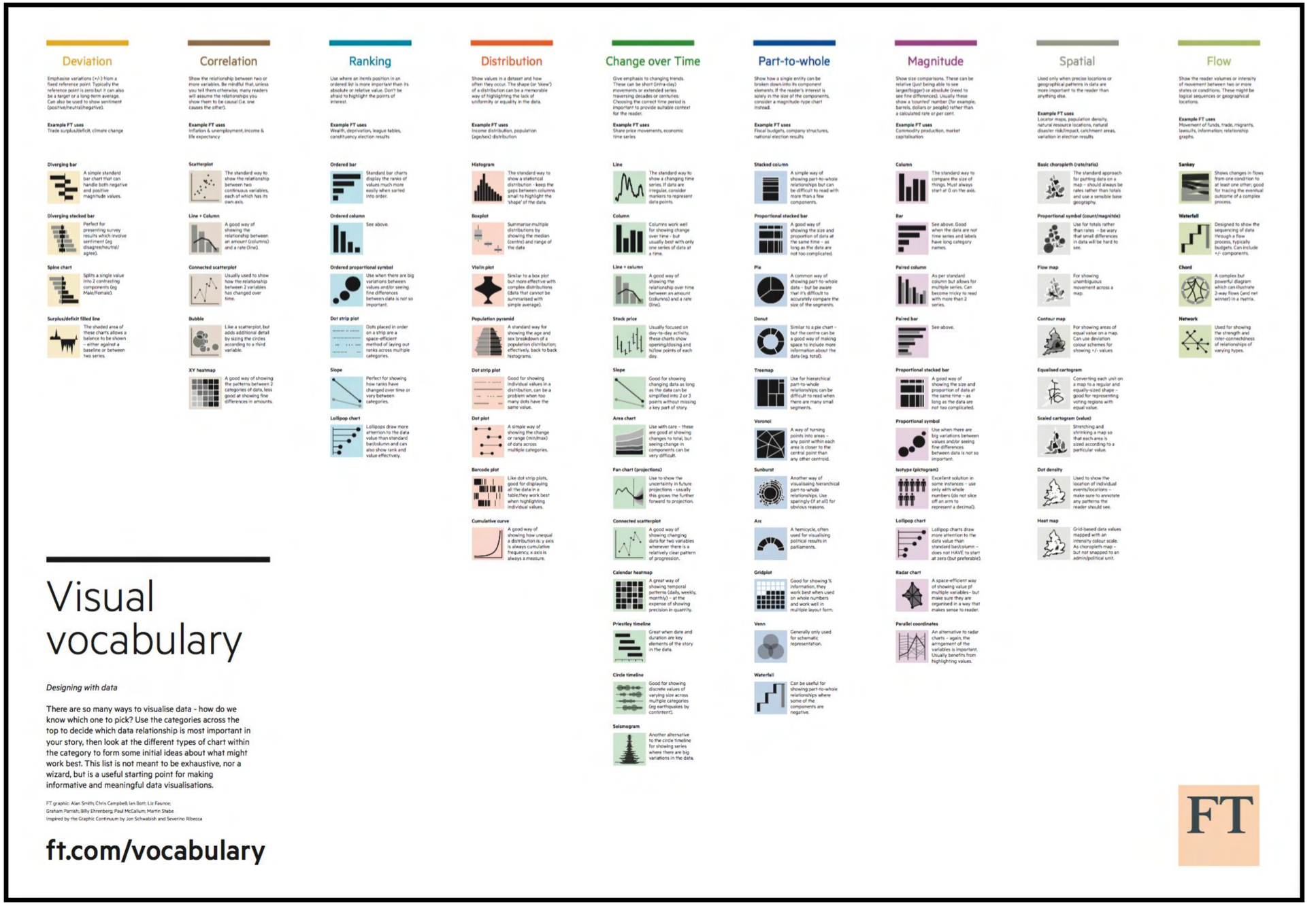
They are just ***better*** or ***worse*** depending on our **intent**:

What it is that we want to communicate, what we want our reader to **be able to see in the data.**



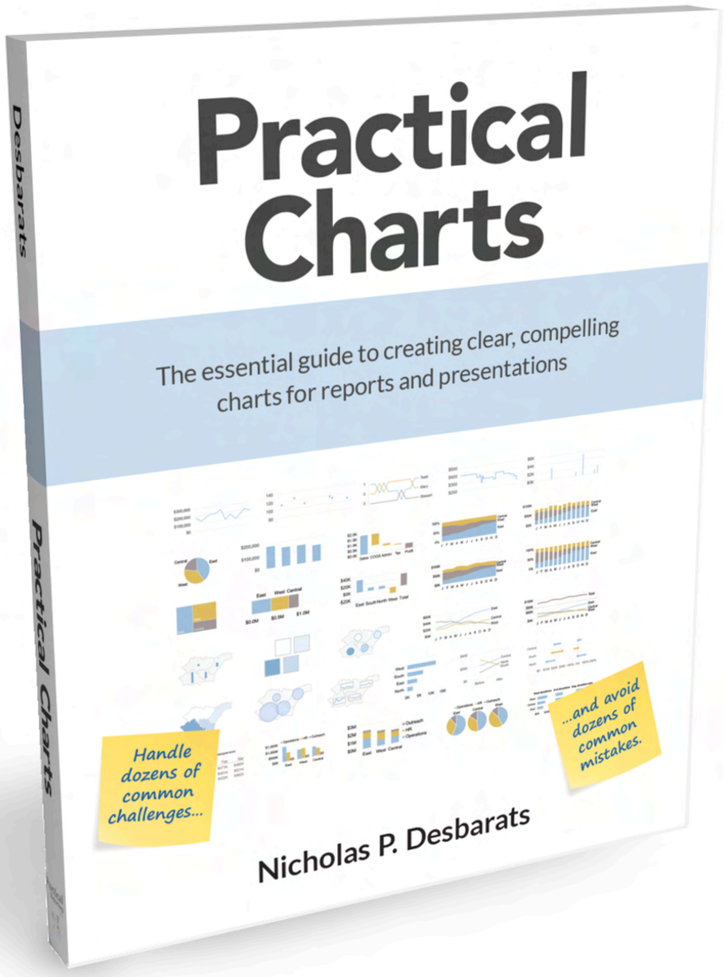


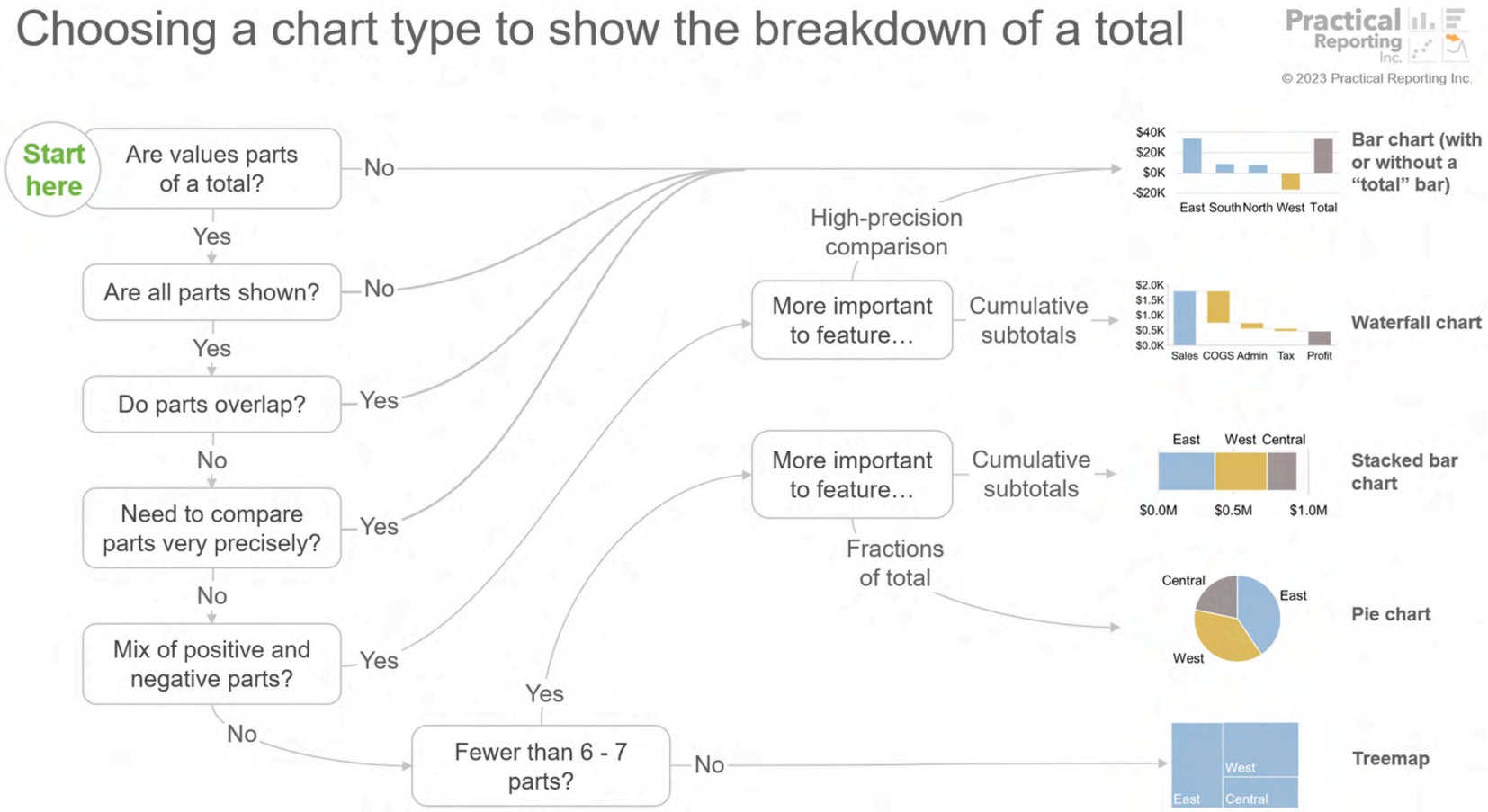




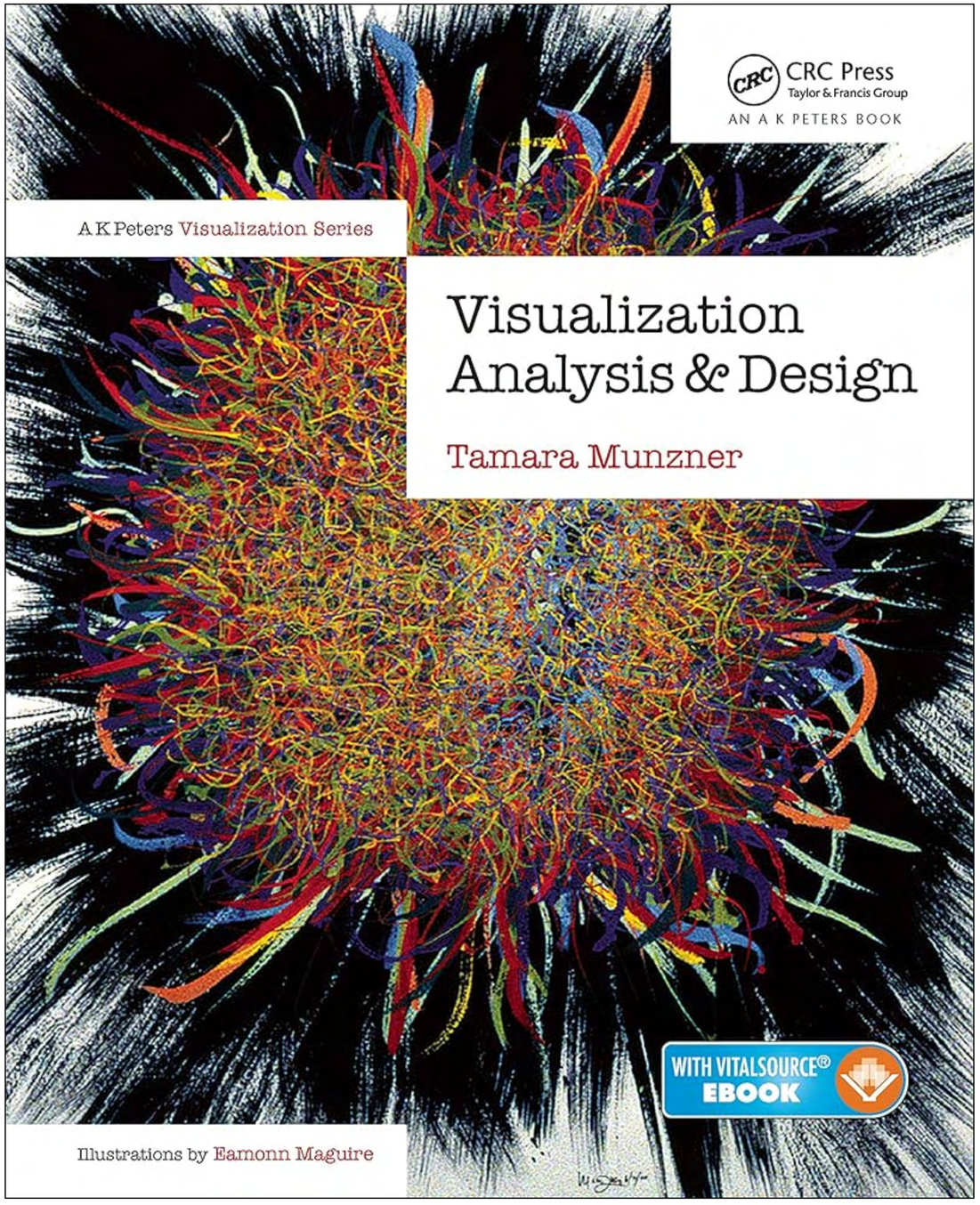
https://journalismcourses.org/wp-content/uploads/2020/07/Visual-vocabulary.pdf

http://www.datavizcatalogue.com/

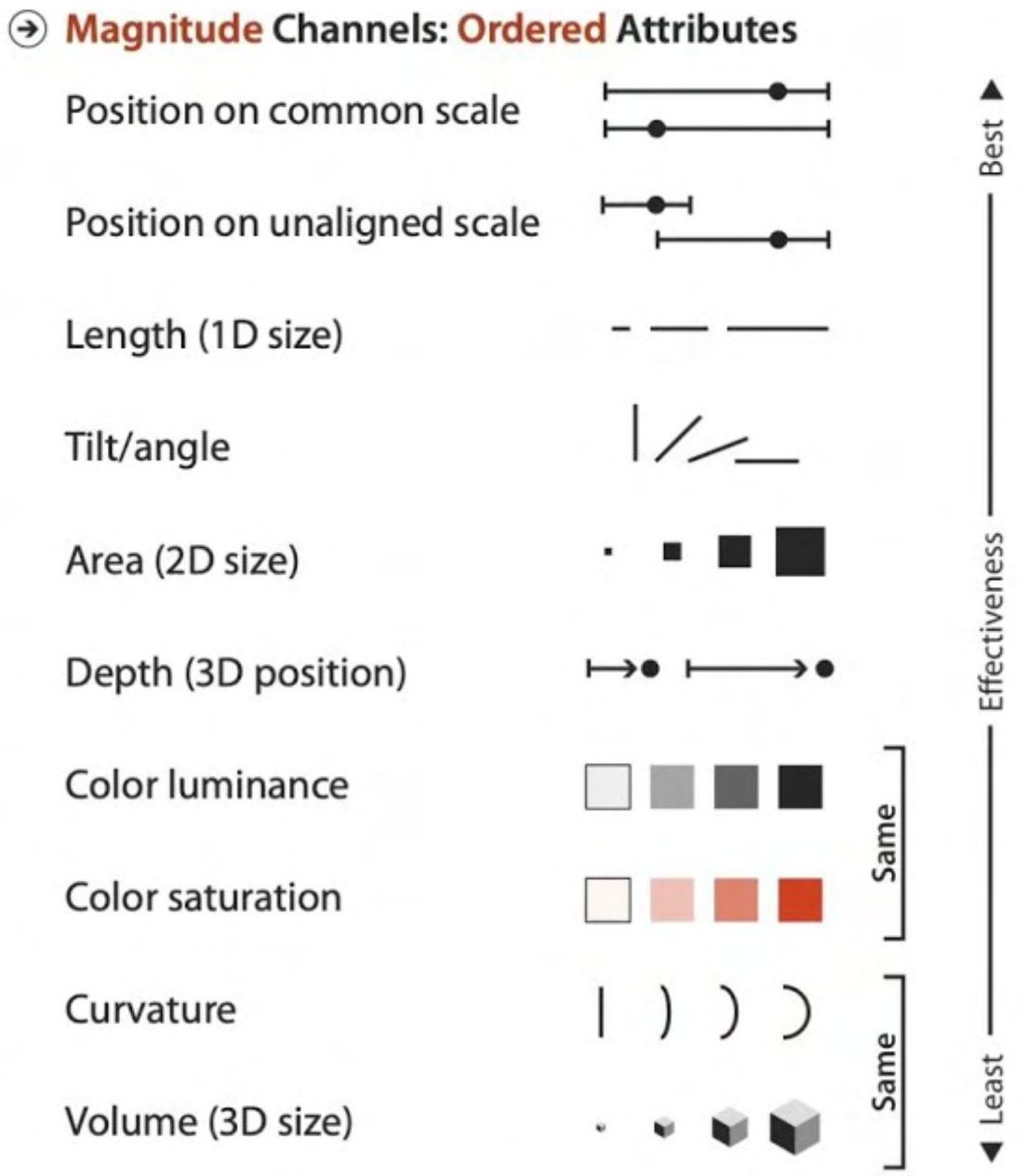


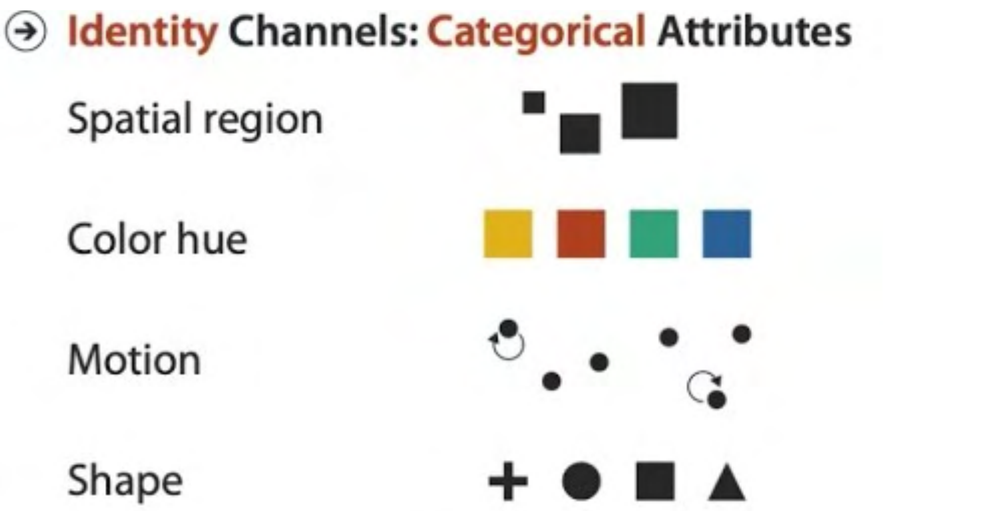


https://www.practicalreporting.com/practical-charts-book


CRC Press
Taylor & Francis Group
AN A K PETERS BOOK
A K Peters Visualization Series
Visualization Analysis & Design
Tamara Munzner
WITH VITALSOURCE® EBOOK
Illustrations by Eamonn Maguire

Encodings/channels hierarchy from Tamara Munzner's *Visualization Analysis and Design* (2014)





- **expressiveness**
  - match channel and data characteristics
- **effectiveness**
  - channels differ in accuracy of perception

- Hereditary cancer syndromes are the exception

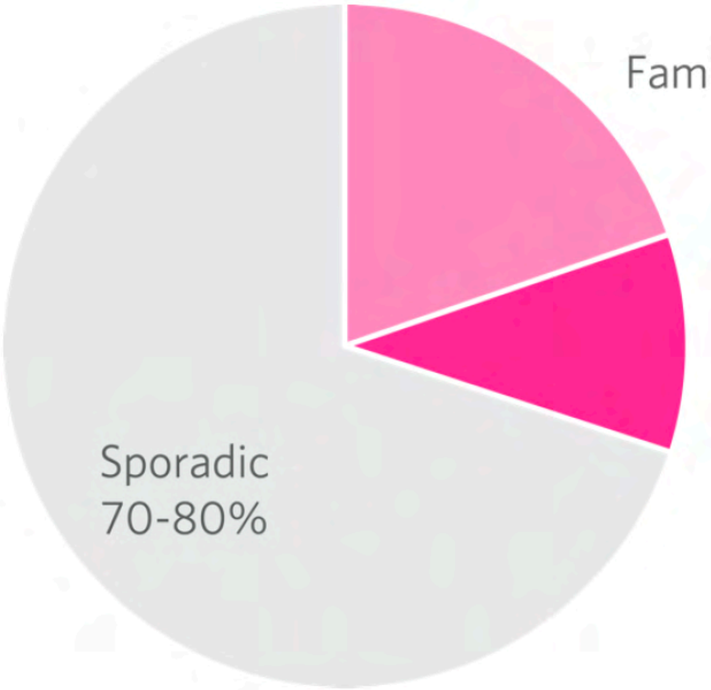
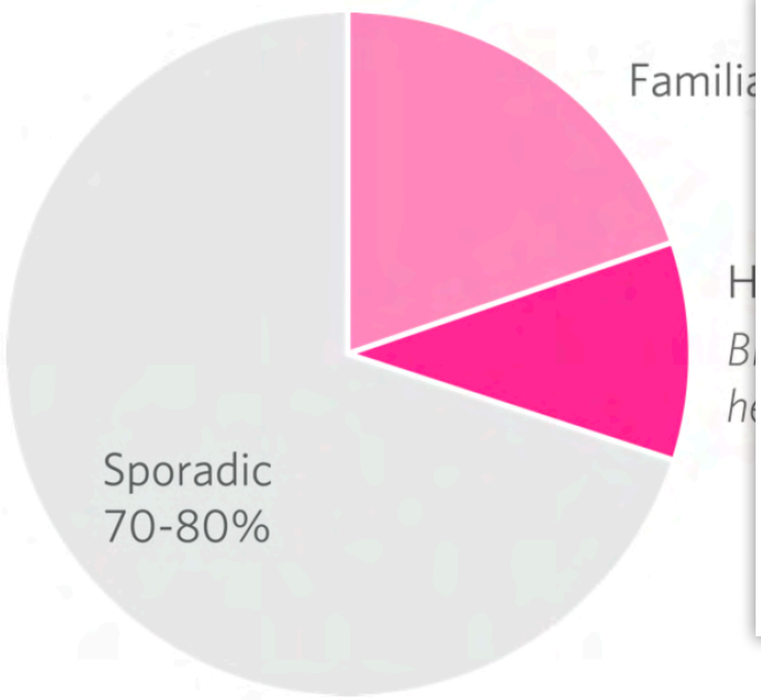


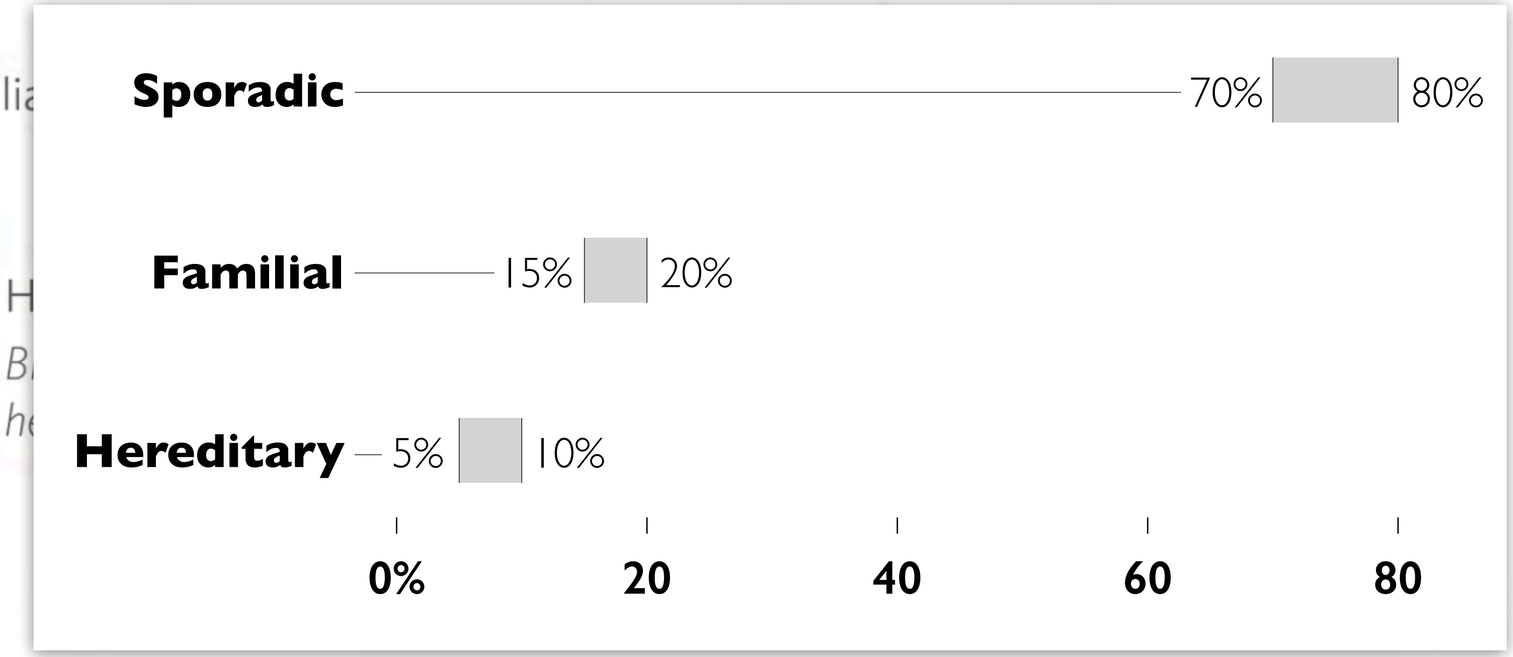
Chart by Ceres Fernández

A pie chart "expresses" a total and its parts

Ceres Fernández

- Hereditary cancer syndromes are the exception



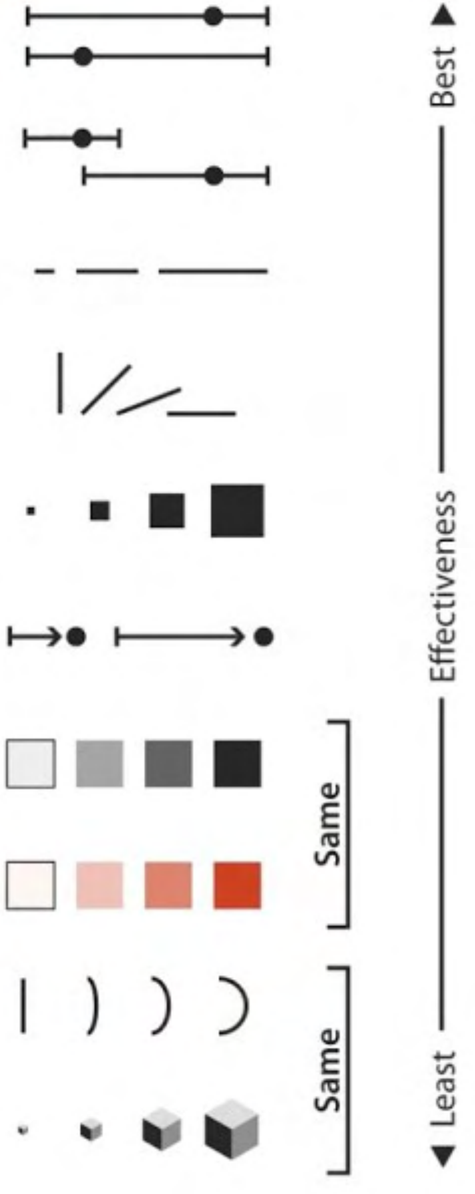


This chart "expresses" ranges

Encodings/channels hierarchy from Tamara Munzner's *Visualization Analysis and Design* (2014)

Magnitude Channels: Ordered Attributes

- Position on common scale
- Position on unaligned scale
- Length (1D size)
- Tilt/angle
- Area (2D size)
- Depth (3D position)
- Color luminance
- Color saturation
- Curvature
- Volume (3D size)



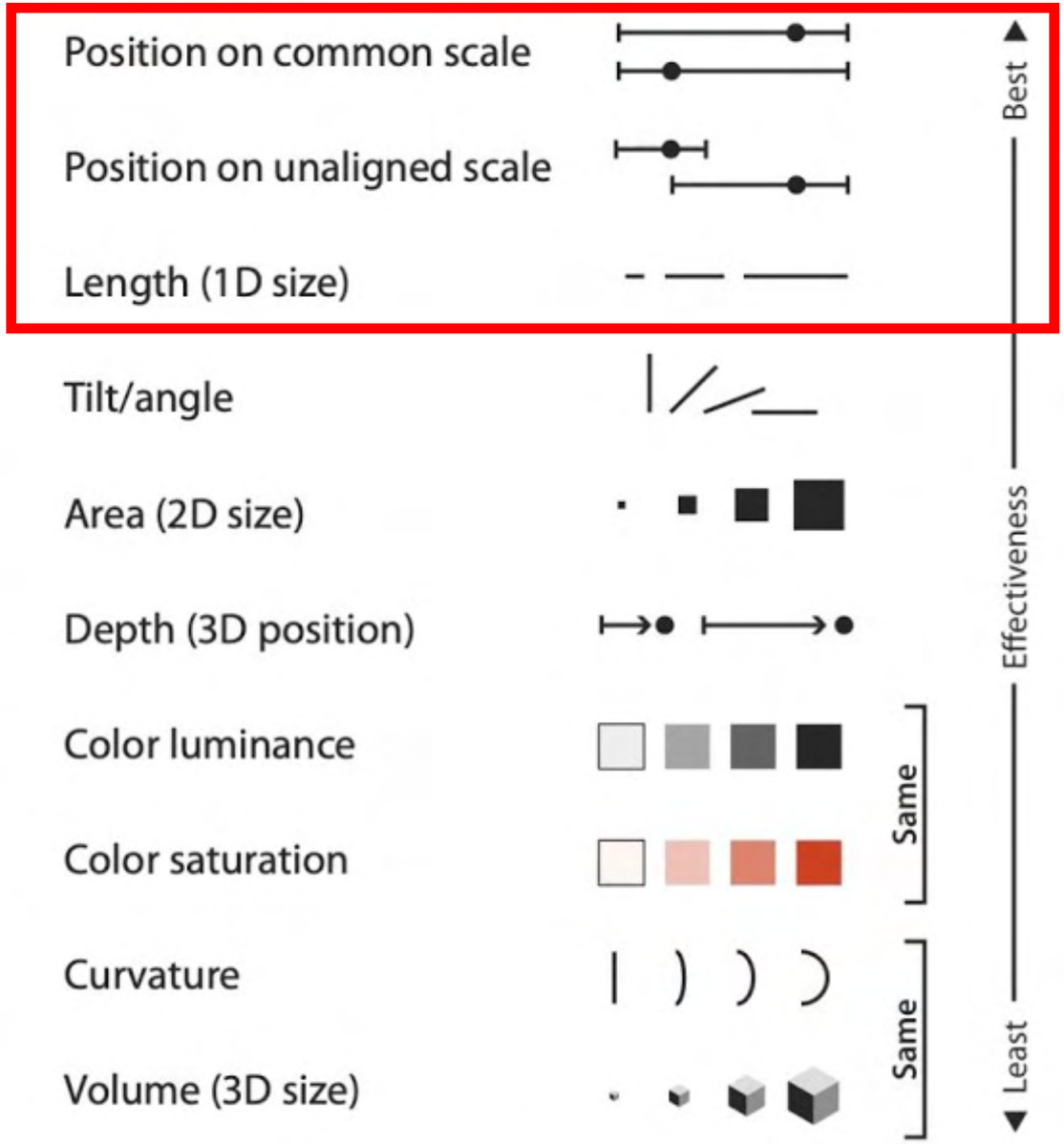
Identity Channels: Categorical Attributes

- Spatial region
- Color hue
- Motion
- Shape

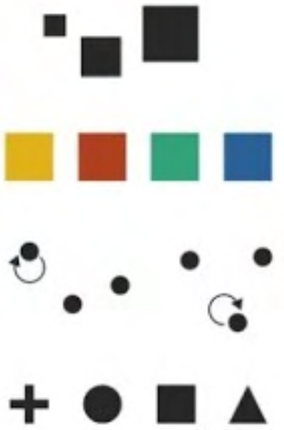
- expressiveness
  - match channel and data characteristics
- **effectiveness**
  - **channels differ in accuracy of perception**

Encodings/channels hierarchy from Tamara Munzner's *Visualization Analysis and Design* (2014)

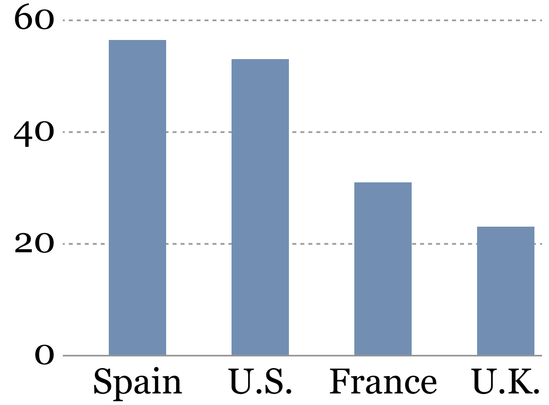
Magnitude Channels: Ordered Attributes



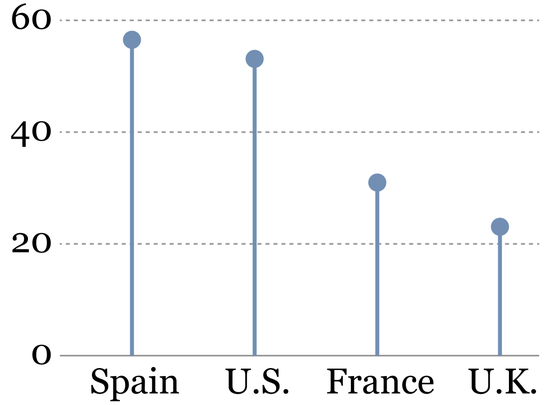
Identity Channels: Categorical Attributes



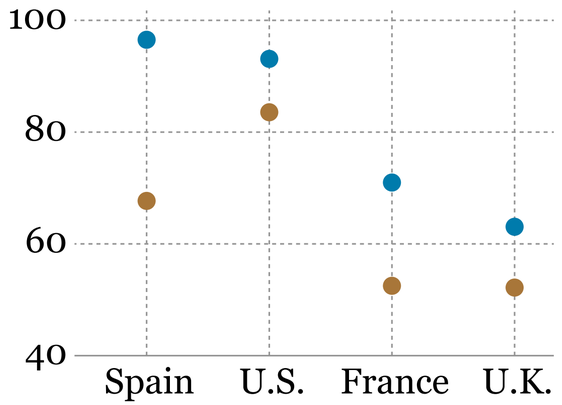
- expressiveness
  - match channel and data characteristics
- **effectiveness**
  - **channels differ in accuracy of perception**



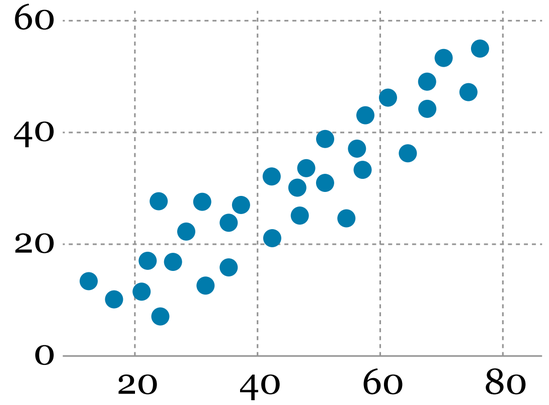
BAR CHART



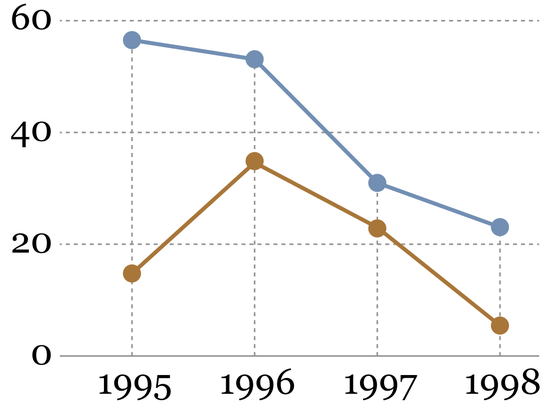
LOLLIPOP CHART



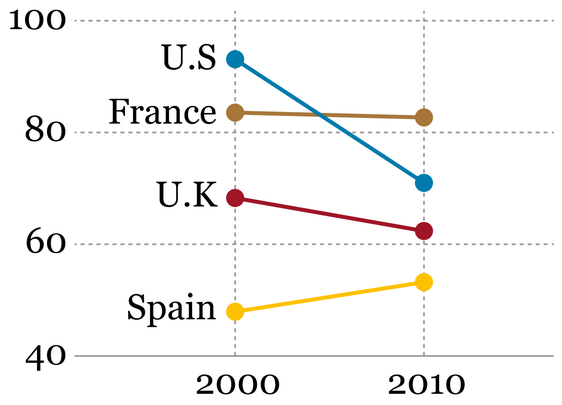
DOT CHART
(or dot plot)



SCATTER CHART
(or scatter plot)



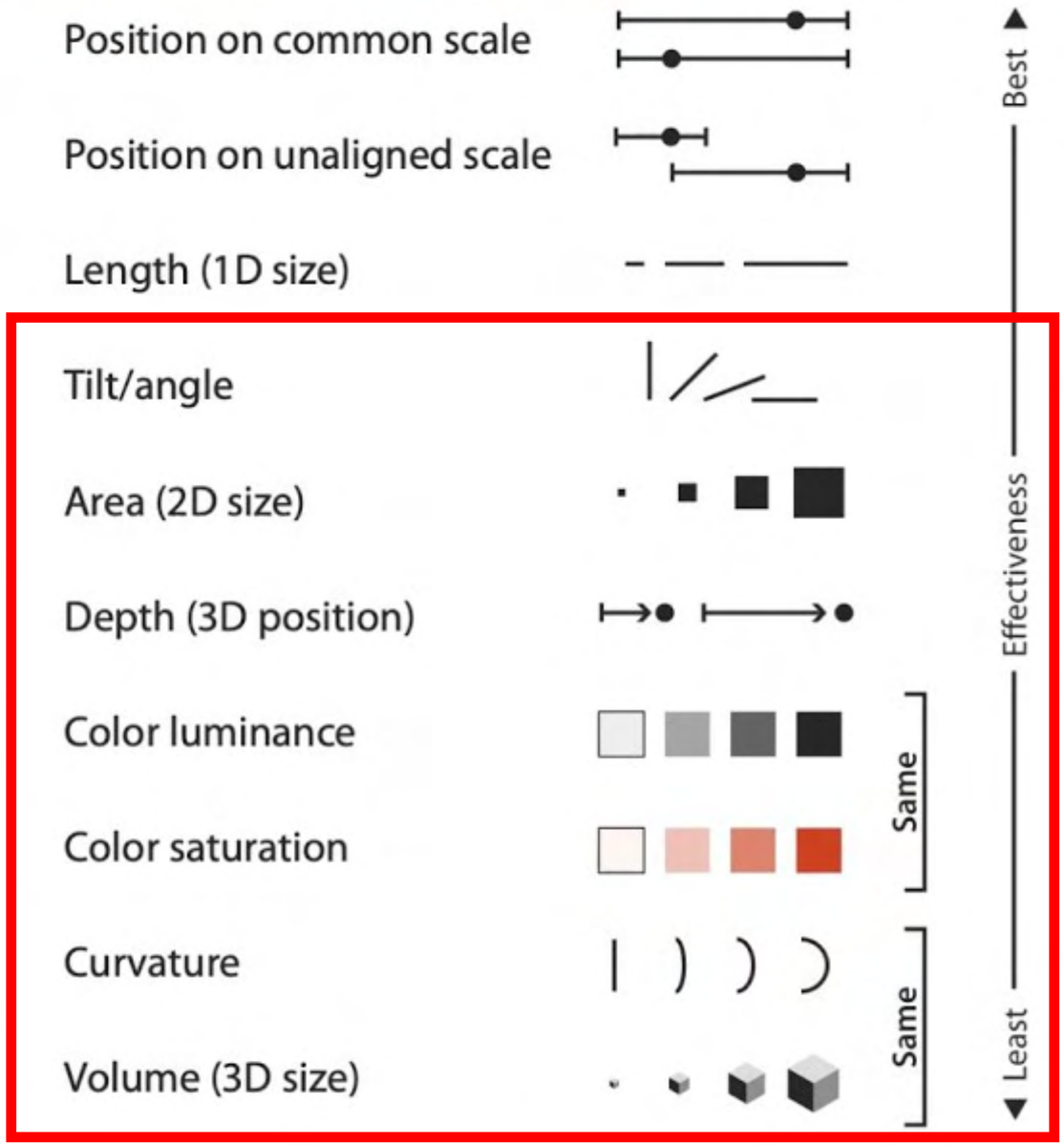
LINE CHART
(or time series chart)

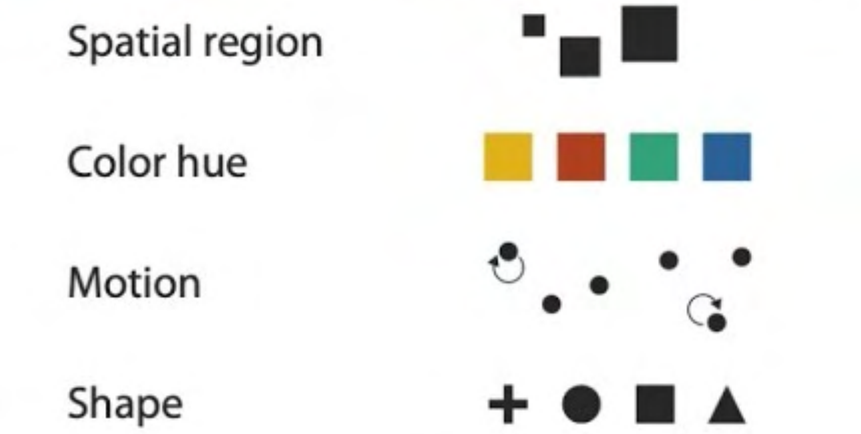


SLOPE CHART

Height, length, position are appropriate encodings
to help readers make accurate estimates

Encodings/channels hierarchy from Tamara Munzner's *Visualization Analysis and Design* (2014)

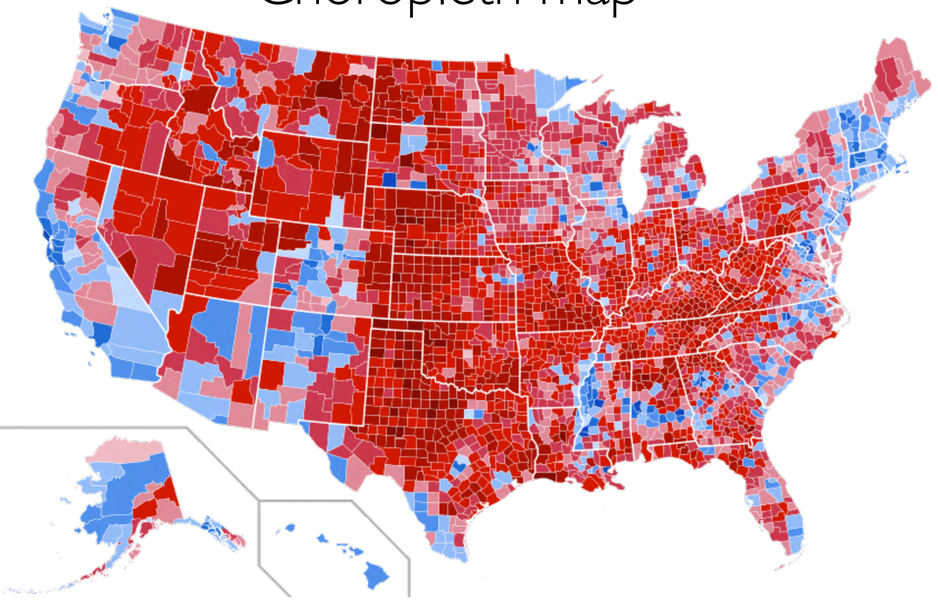




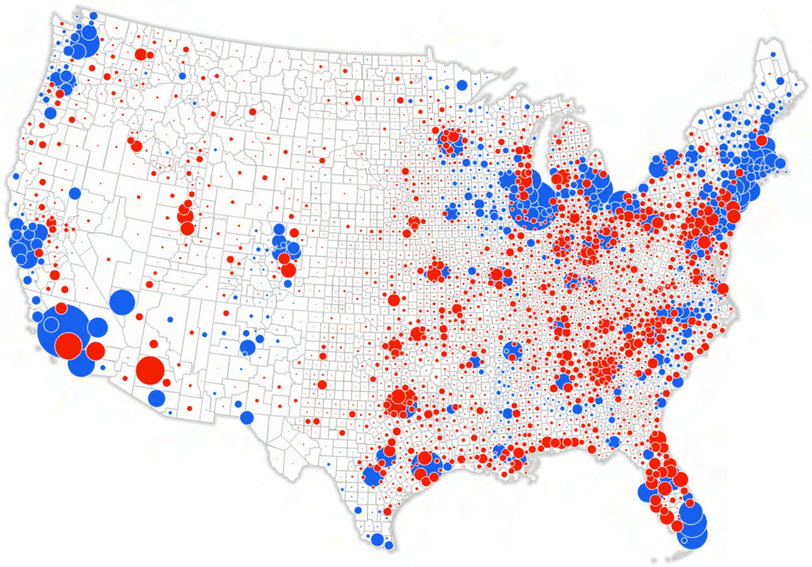
- expressiveness
  - match channel and data characteristics
- effectiveness
  - channels differ in accuracy of perception

A visual channel can be "effective" in the non-technical sense of the term because, in many cases, high accuracy of perception isn't the goal.

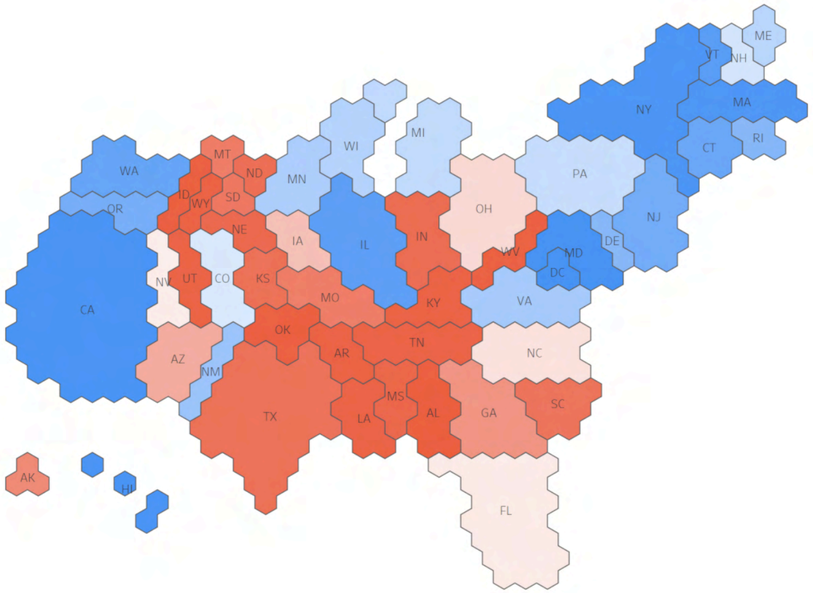
Choropleth map

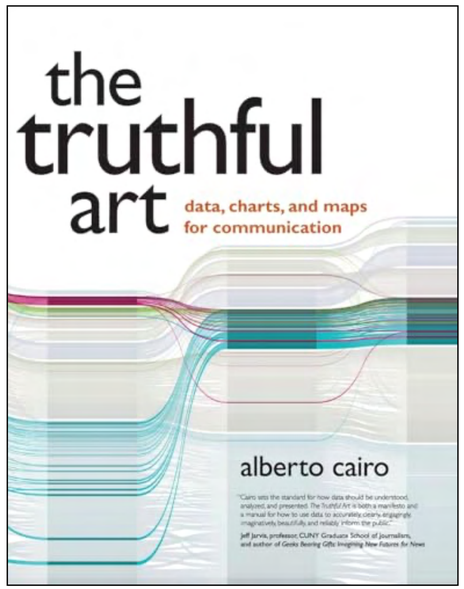


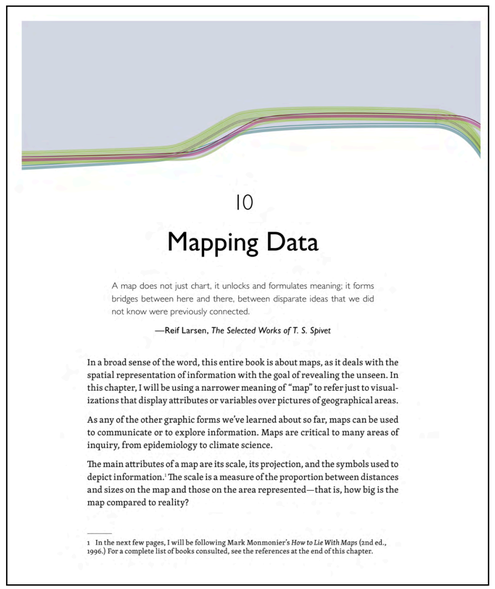
Proportional symbol map



Cartogram







10

Mapping Data

A map does not just chart, it unlocks and formulates meaning; it forms bridges between here and there, between disparate ideas that we did not know were previously connected.

—Reif Larsen, *The Selected Works of T. S. Spivet*

In a broad sense of the word, this entire book is about maps, as it deals with the spatial representation of information with the goal of revealing the unseen. In this chapter, I will be using a narrower meaning of "map" to refer just to visualizations that display attributes or variables over pictures of geographical areas.

As any of the other graphic forms we've learned about so far, maps can be used to communicate or to explore information. Maps are critical to many areas of inquiry, from epidemiology to climate science.

The main attributes of a map are its scale, its projection, and the symbols used to depict information.[1] The scale is a measure of the proportion between distances and sizes on the map and those on the area represented—that is, how big is the map compared to reality?

1 In the next few pages, I will be following Mark Monmonier's *How to Lie With Maps* (2nd ed., 1996.) For a complete list of books consulted, see the references at the end of this chapter.

Chapter 10 of *The Truthful Art*

https://drive.google.com/file/d/1Ap4ehrBDXBSDPcnPXCTeg9cVH1U9kOeX/

Truck Sales Slip, Tripping Up Chrysler

Over the past few years, Chrysler executives said they were following the lead of Toyota and Honda, focusing on vehicles that met the needs of their customers. But as American consumers turned away from large trucks and S.U.V.'s in 2006, Chrysler continued to churn out big vehicles, which are now sitting unsold at dealerships across the country.

READING THE CHART

Boxes are scaled proportionally according to number of cars sold in 2006

25,000

100,000

Change in sales from 2005 to 2006

Below | Above industry average

No 2005 sales

−10% −2.6 0 +10 +100

Many of these vehicles were introduced in 2005.

SALES CHANGE '05 TO '06

◄TRUCKS, VANS, S.U.V.'S | CARS►

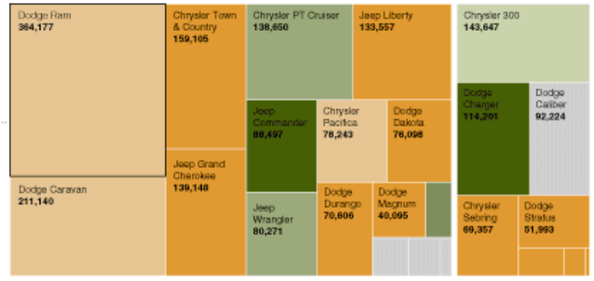
Chrysler Group **−7.0%**

Trucks/vans/S.U.V.'s 1.6 million

Cars 0.5 million

Pickups, minivans and S.U.V.'s made up 76 percent of Chrysler's sales, which left it vulnerable when consumers shifted to cars.

Dodge Ram



General Motors **−8.7%**

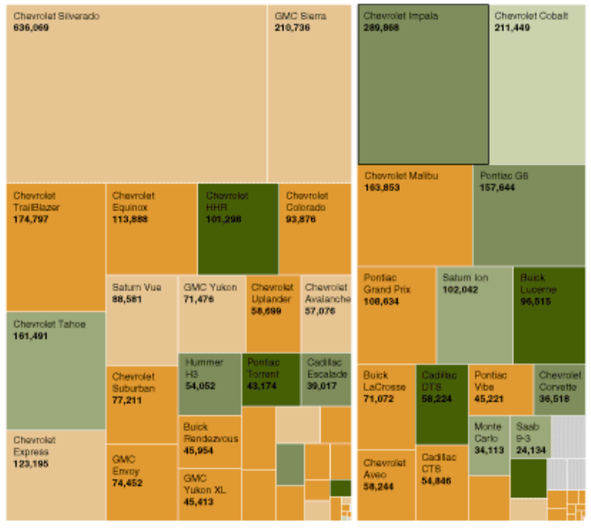
Trucks/vans/S.U.V.'s 2.5 million

Cars 1.6 million

G.M. introduced new versions of its large S.U.V.'s in late 2005, hoping they would bolster sales. Instead, sales of big vehicles were hurt when gas prices climbed. One of the few standouts was the Chevrolet HHR, new in 2005.

Chevrolet Impala

The Chevrolet Impala, with or without flashing lights, did well in 2006, when a redesign came out.



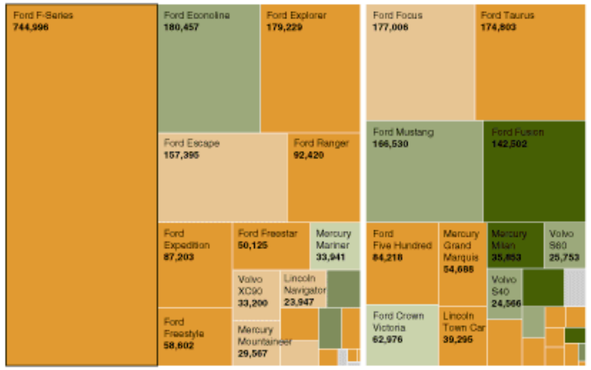
Ford **−8.3%**

Trucks/vans/S.U.V.'s 1.8 million

Cars 1.1 million

Even the country's best-selling vehicles, the F-Series, slumped in 2006, with sales dropping 13 percent. One of Ford's bright spots was the new Fusion sedan, which made its debut in late 2005 and sold well in its first full year.

Ford F-Series



◄TRUCKS, VANS, S.U.V.'S | CARS►

Toyota **+12.5%**

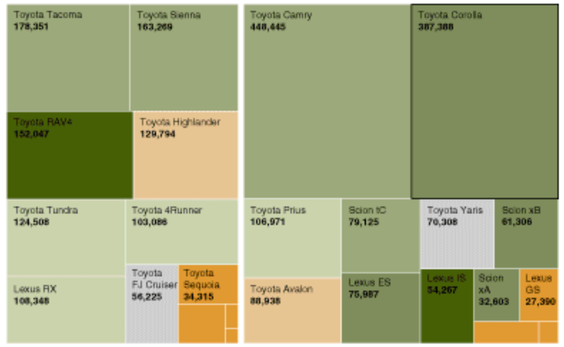
Trucks/vans/S.U.V.'s 1.1 million

Cars 1.5 million

Toyota rolled out a new version of the Camry, and once again it was the country's best-selling car.

Toyota Corolla

Corolla sales also jumped, along with gas prices. Toyota could not escape the decline in sales of supersized S.U.V.'s like its Sequoia.



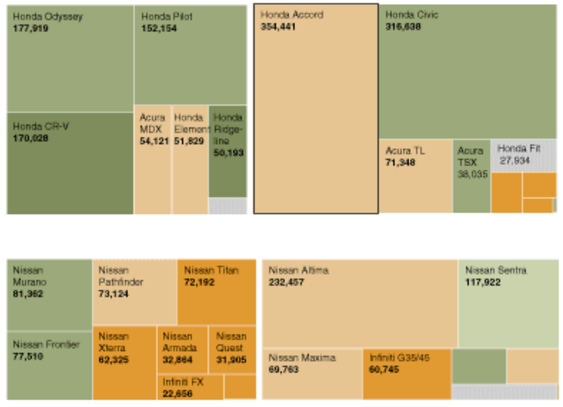
Honda **+3.2%**

Trucks/vans/S.U.V.'s 0.7 million

Cars 0.8 million

Like the Corolla, the small Honda Civic did well. But the Accord stalled. Buyers, it seems, are waiting for the new version to be released this year.

Honda Accord



Nissan **−5.3%**

Trucks/vans/S.U.V.'s 0.5 million

Cars 0.6 million

BMW **+2.1%**

Trucks/vans/S.U.V.'s 0.1 million

Cars 0.3 million

Mercedes-Benz **+10.3%**

Trucks/vans/S.U.V.'s 0.1 million

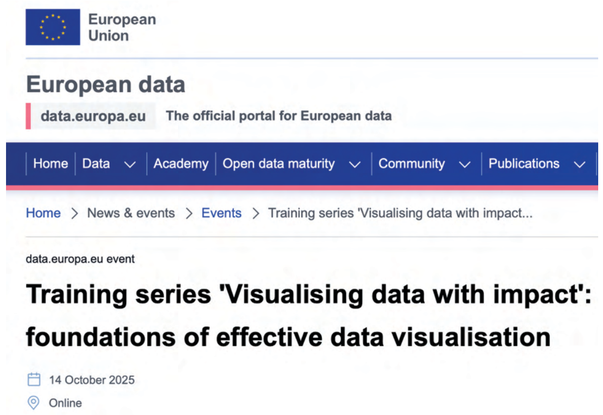
Cars 0.2 million

Mercedes-Benz, owned by DaimlerChrysler, had a comeback in 2006, thanks to a new version of its flagship S-Class. BMW sales were helped by a new version of its 3 Series sport sedan.

Sources: Ward's AutoInfoBank; Edmunds

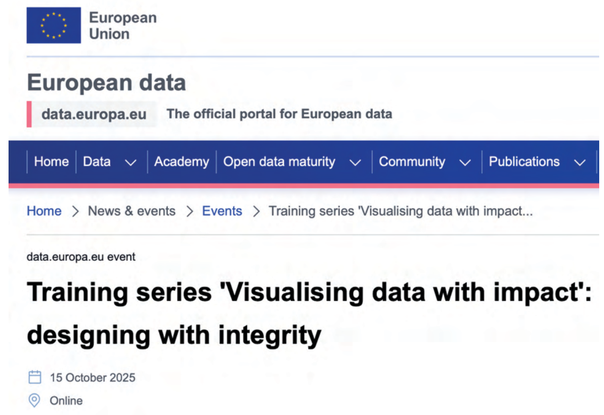
Amanda Cox and Hannah Fairfield/ The New York Times

Treemap by *The New York Times*



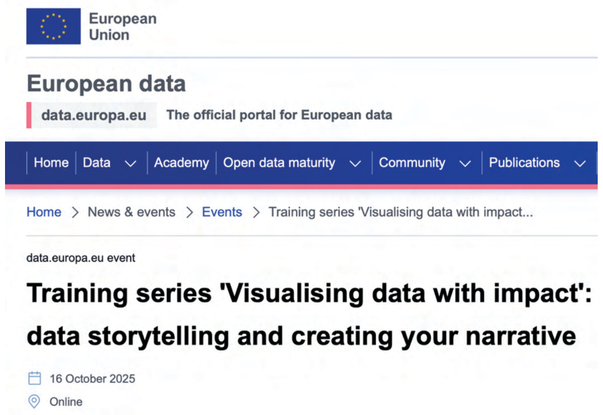
Episode ONE - October 14

https://data.europa.eu/en/news-events/events/training-data-visualisation-session-1-foundations-effective-data-visualisation



Episode TWO - October 15

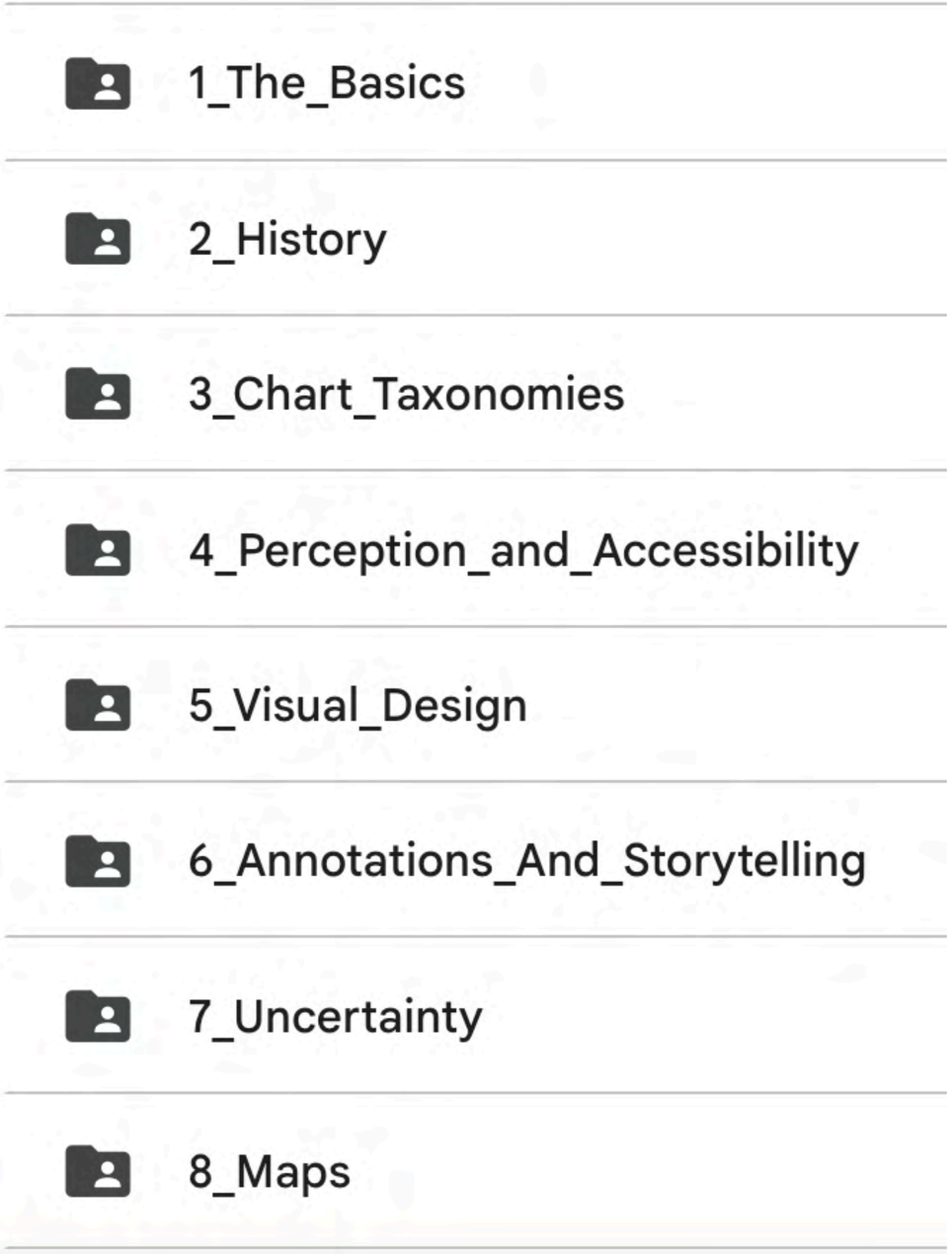
https://data.europa.eu/en/news-events/events/data-visualisation-training-session-2-designing-integrity



Episode THREE - October 16

Tomorrow's webinar

https://data.europa.eu/en/news-events/events/data-visualisation-training-session-3-data-storytelling-and-creating-your



Extra readings

https://tinyurl.com/mr2st327

Designing With Integrity

Thank you!

Alberto Cairo

OpenVisualizationAcademy.com