

# Data Communication

## Graphics that tell stories in an engaging way

Cédric Scherer // rstudio::conf // July 2022

# *Data Visualization*

---

is any graphical representation  
of information and data.

# Data Visualization

---

converts information into visual  
forms as quantifiable features.

# Data Visualization

---

helps to amplify cognition, gain insights,  
discover, explain, and make decisions.

# *Visualize Your Data*

---

# Visualize Your Data

“When Dmitry Kobak and Sergey Shpilkin [...] analysed the results, they found that **an unusually high number of turnout and vote-share results were multiples of five** (eg, 50%, 55%, 60%), a tell-tale sign of manipulation.”

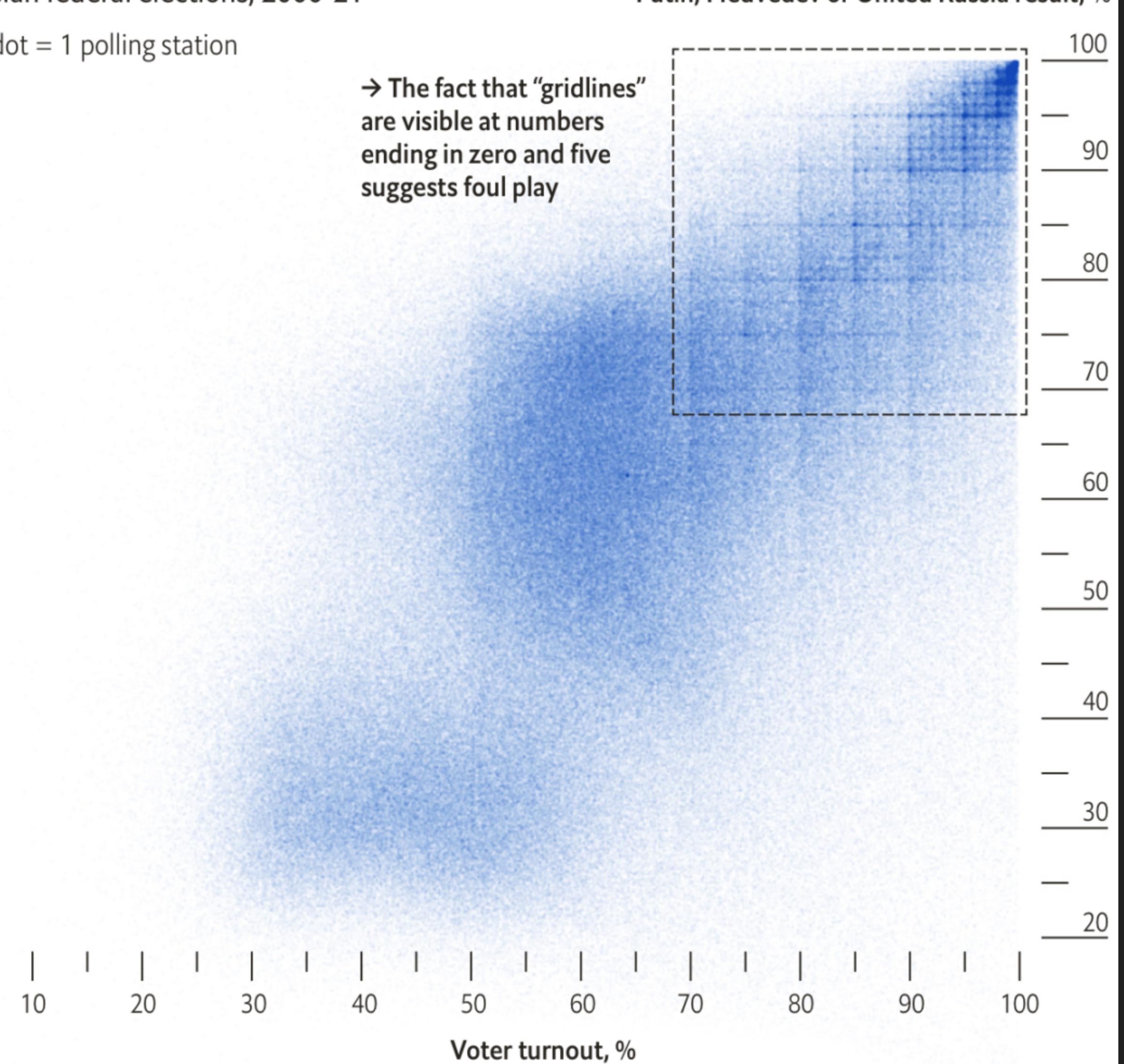
*“Russian elections once again had a suspiciously neat result”  
by The Economist*

## Fair and square?

Russian federal elections, 2000-21

● 1 dot = 1 polling station

→ The fact that “gridlines” are visible at numbers ending in zero and five suggests foul play

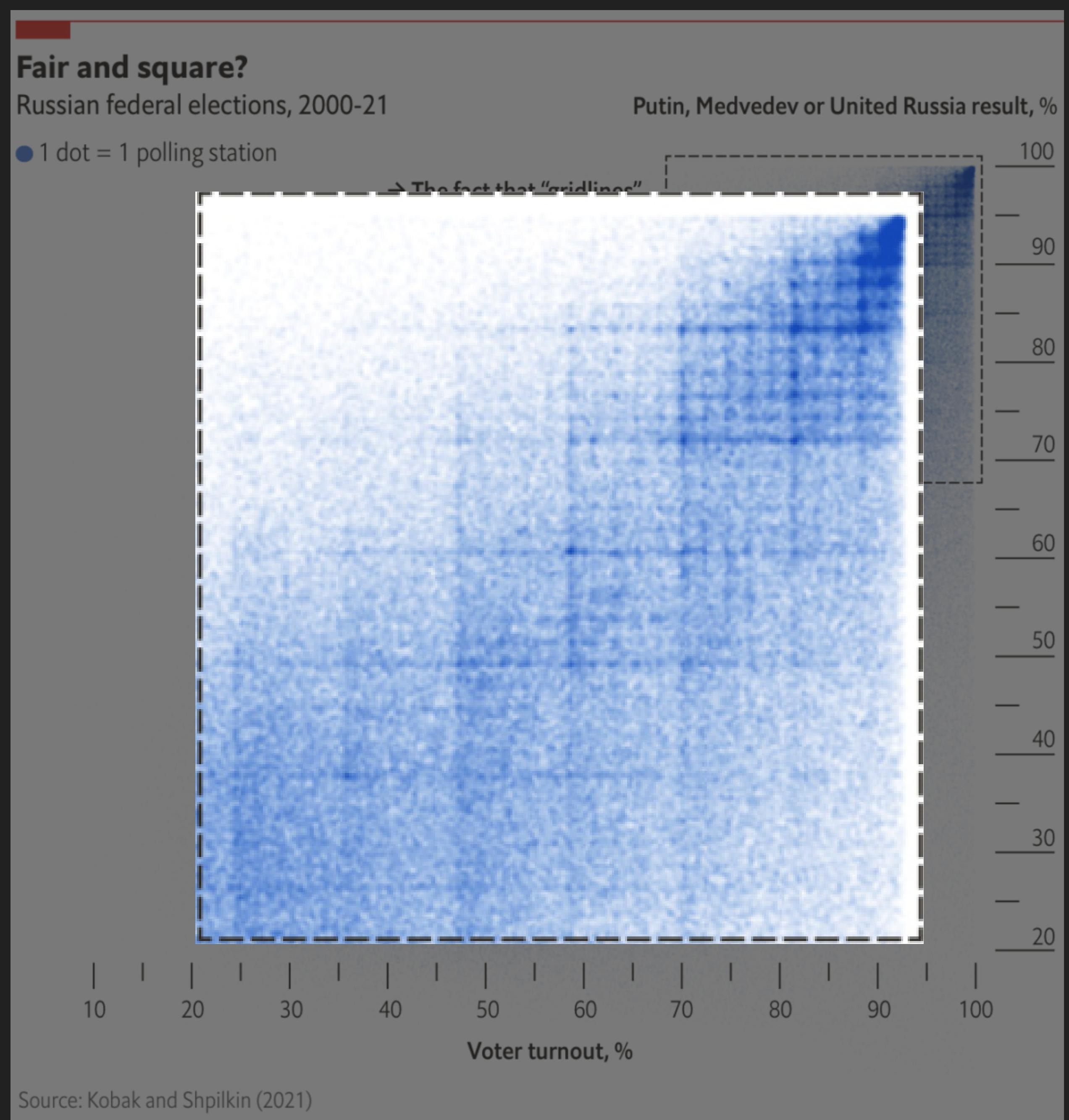


Source: Kobak and Shpilkin (2021)

# Visualize Your Data

“When Dmitry Kobak and Sergey Shpilkin [...] analysed the results, they found that **an unusually high number of turnout and vote-share results were multiples of five** (eg, 50%, 55%, 60%), a tell-tale sign of manipulation.”

*“Russian elections once again had a suspiciously neat result”  
by The Economist*

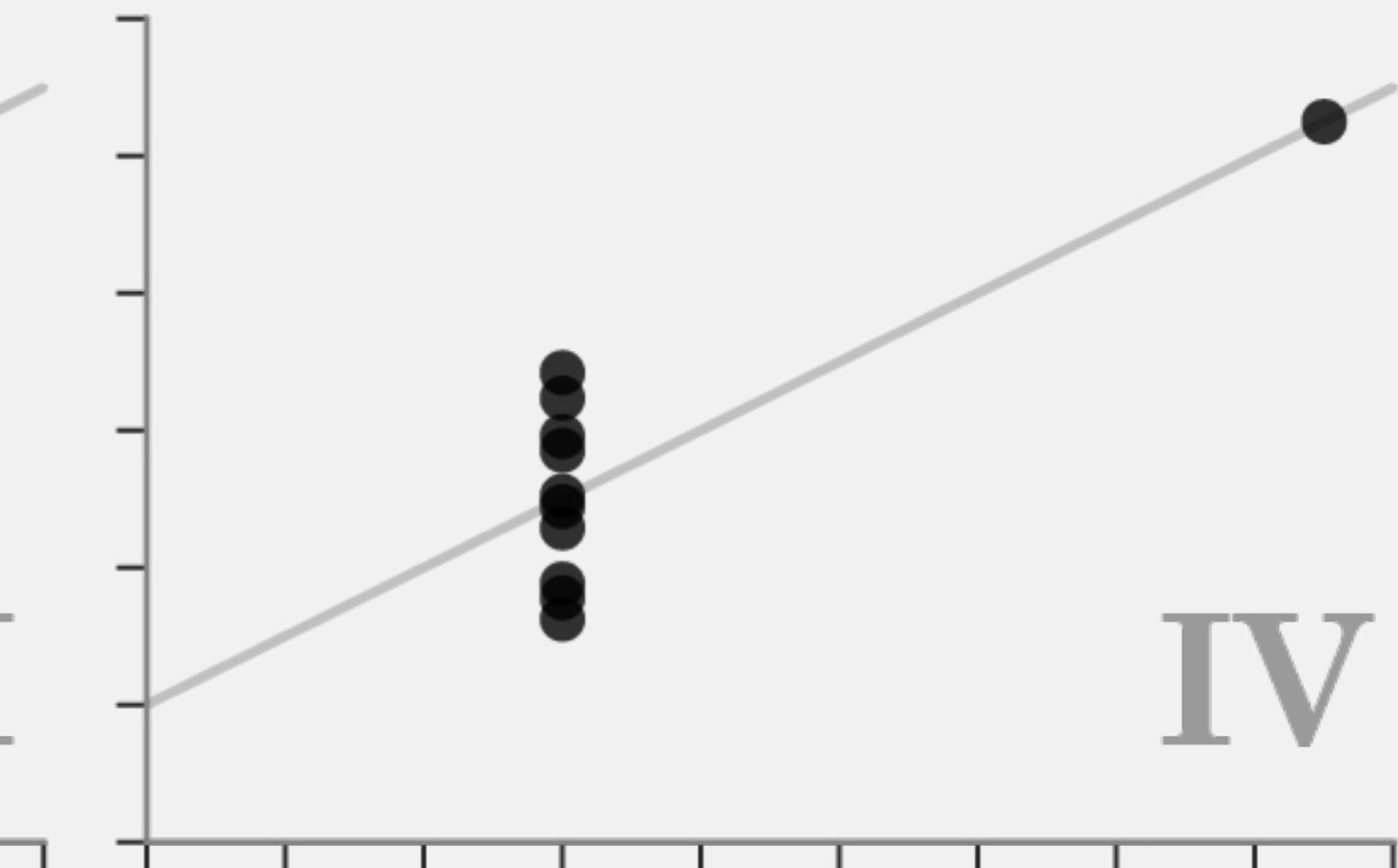
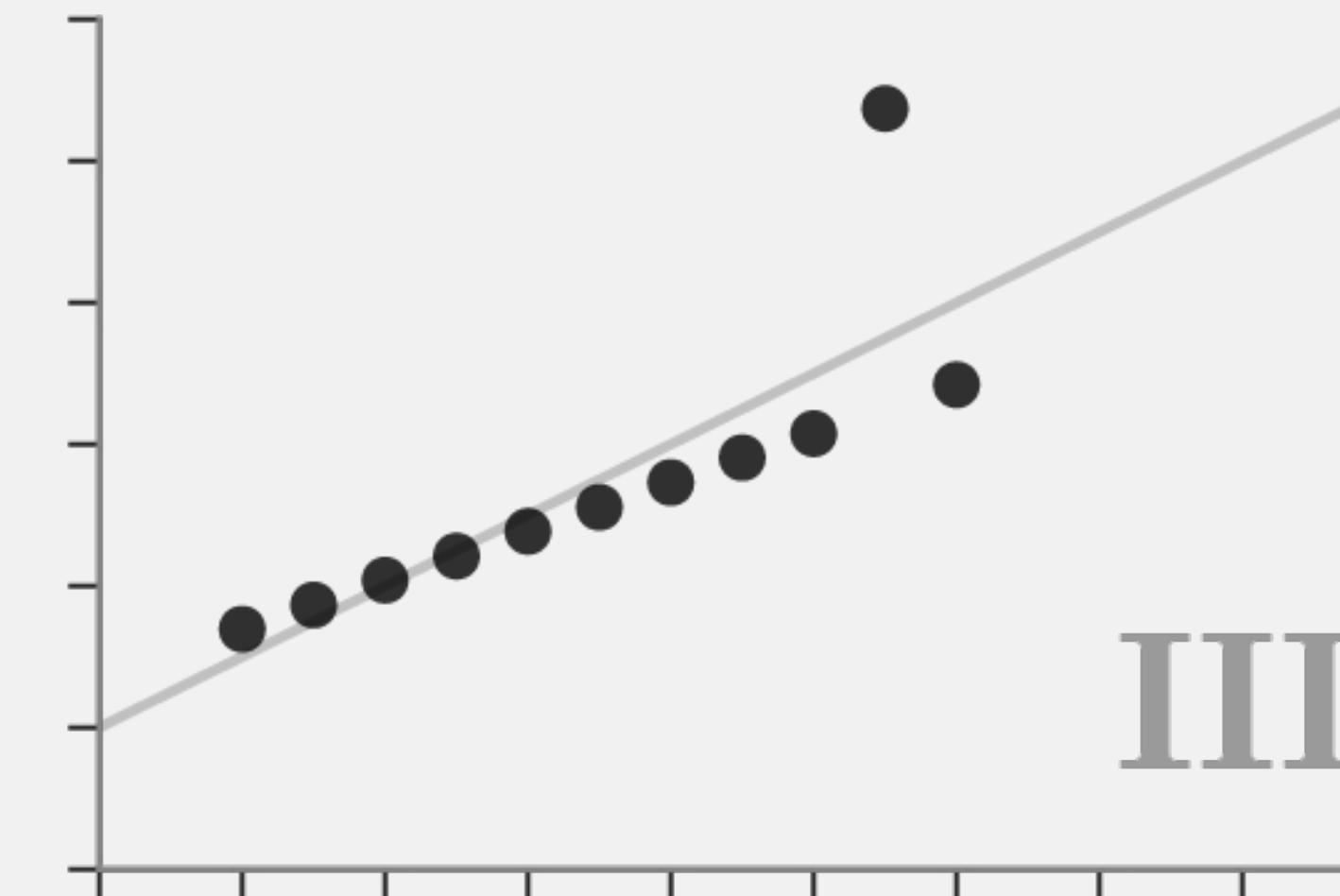
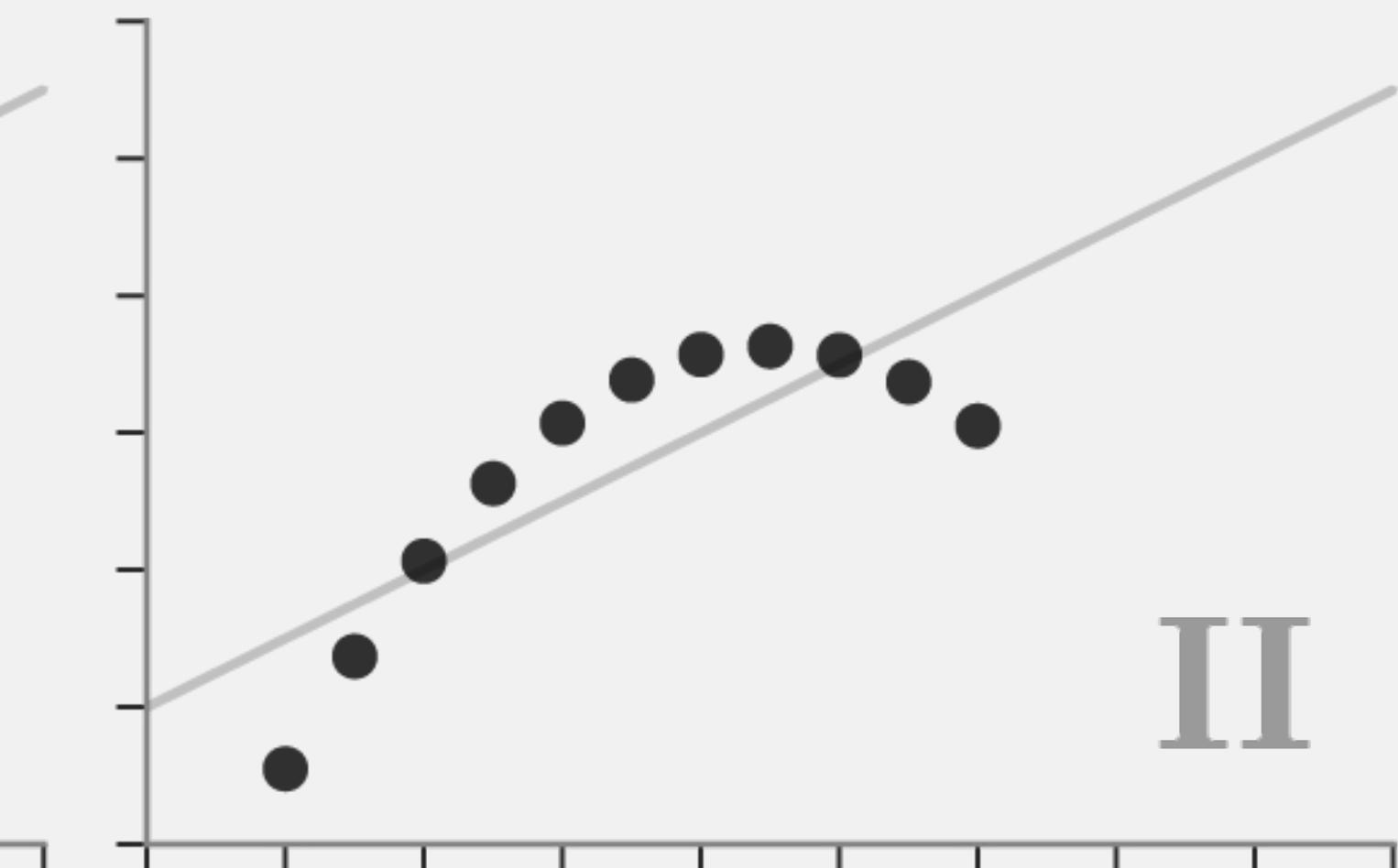
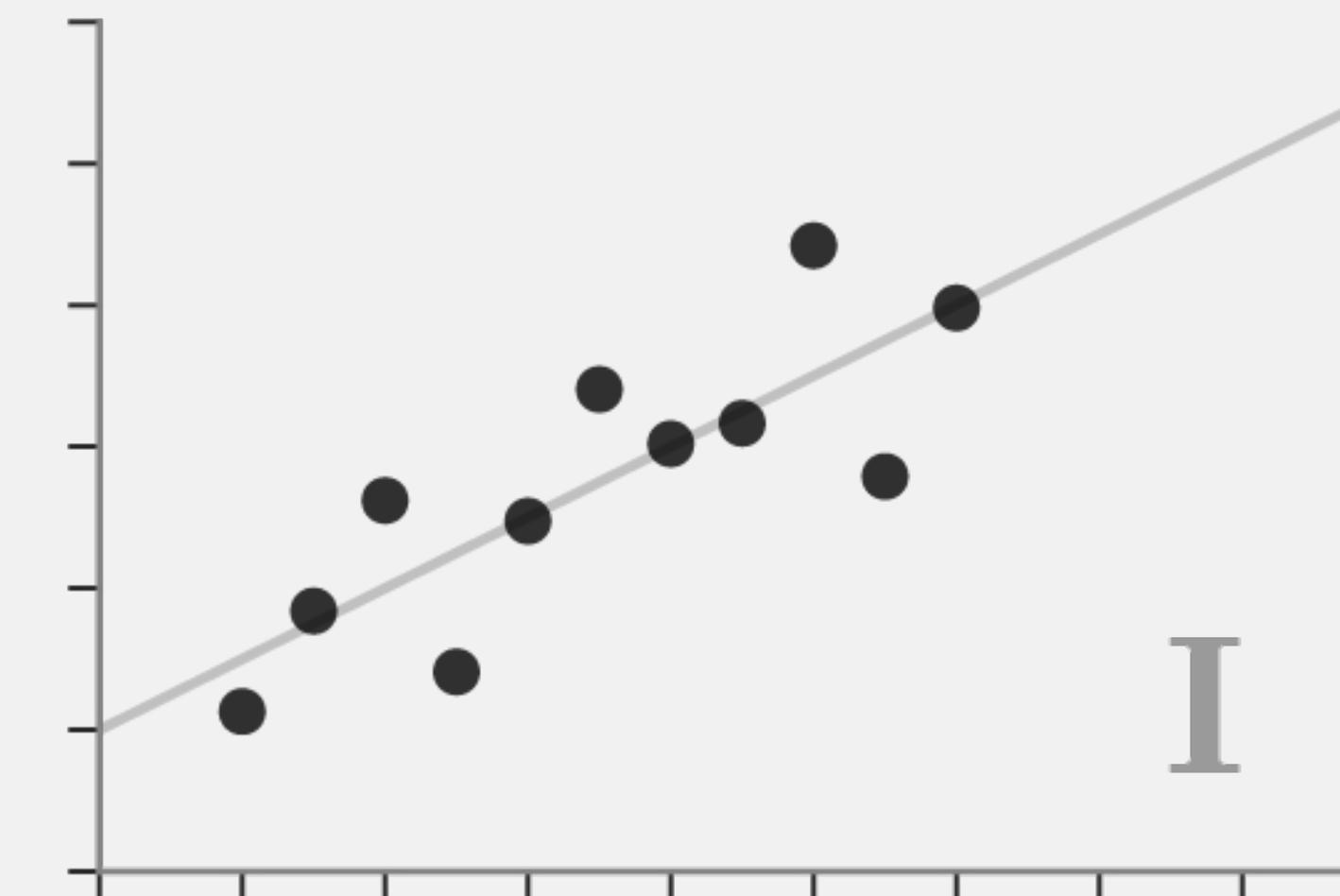


# Anscombe's Quartet

**each dataset has the same summary statistics**

mean, standard deviation, and correlation

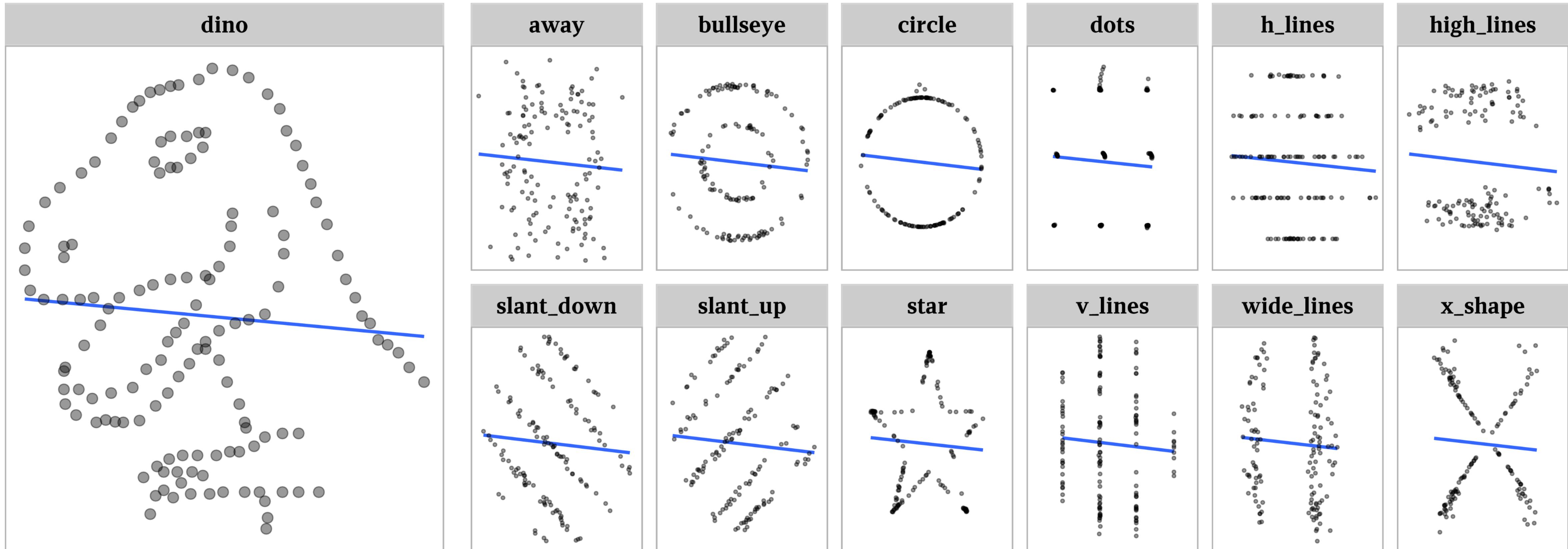
**but are visually distinct.**



*"Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing"*  
by Justin Matejka & George Fitzmaurice , ACM SIGCHI Conference on Human Factors in Computing Systems 2017

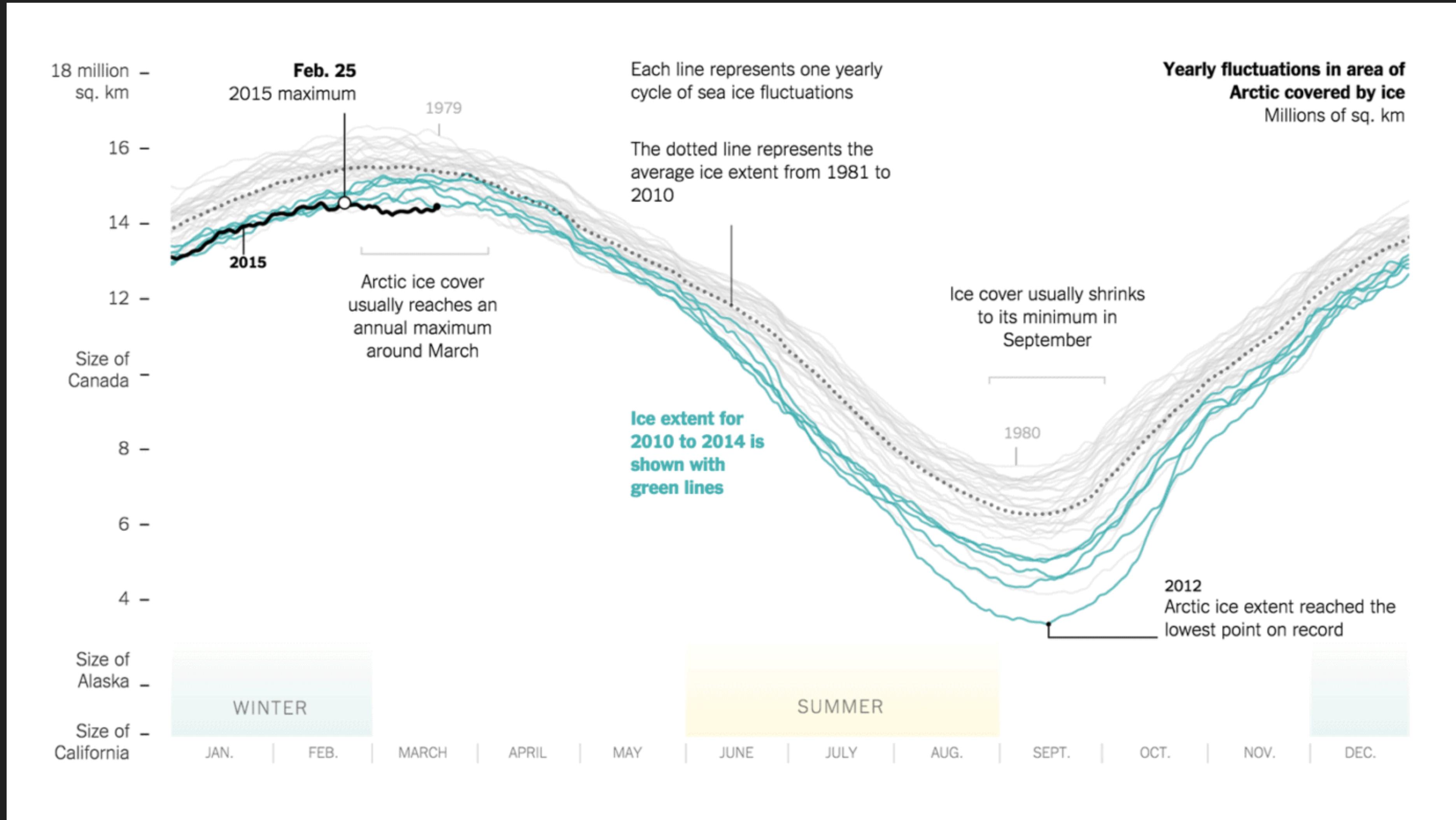
# Datasaurus Dozen

The Datasaurus by Alberto Cairo shows us why visualisation is important, not just summary statistics.

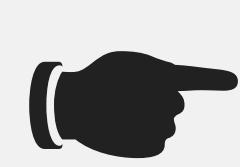


*"Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing"*  
by Justin Matejka & George Fitzmaurice , ACM SIGCHI Conference on Human Factors in Computing Systems 2017

# What makes it a good data visualization?



# What makes it a good data visualization



**INFORMATION** (integrity)

# What makes it a good data visualization

- 👉 **INFORMATION** (integrity)
- 👉 **STORY** (interestingness)

# What makes it a good data visualization

- 👉 **INFORMATION** (integrity)
- 👉 **STORY** (interestingness)
- 👉 **GOAL** (usefulness)

# What makes it a good data visualization

- ➔ **INFORMATION** (integrity)
- ➔ **STORY** (interestingness)
- ➔ **GOAL** (usefulness)
- ➔ **VISUAL FORM** (beauty)

# What Makes a Good Visualization?

explicit (implicit)



David McCandless  
InformationisBeautiful.net

taken from new book  
Knowledge is Beautiful

find out more  
[bit.ly/KIB\\_Books](http://bit.ly/KIB_Books)

# Information

---

**Understand your data and be accurate**

How to Steer Clear of Common Blunders When Working  
with Data and Presenting Analysis and Visualizations

# AVOIDING DATA PITFALLS



**BEN JONES**

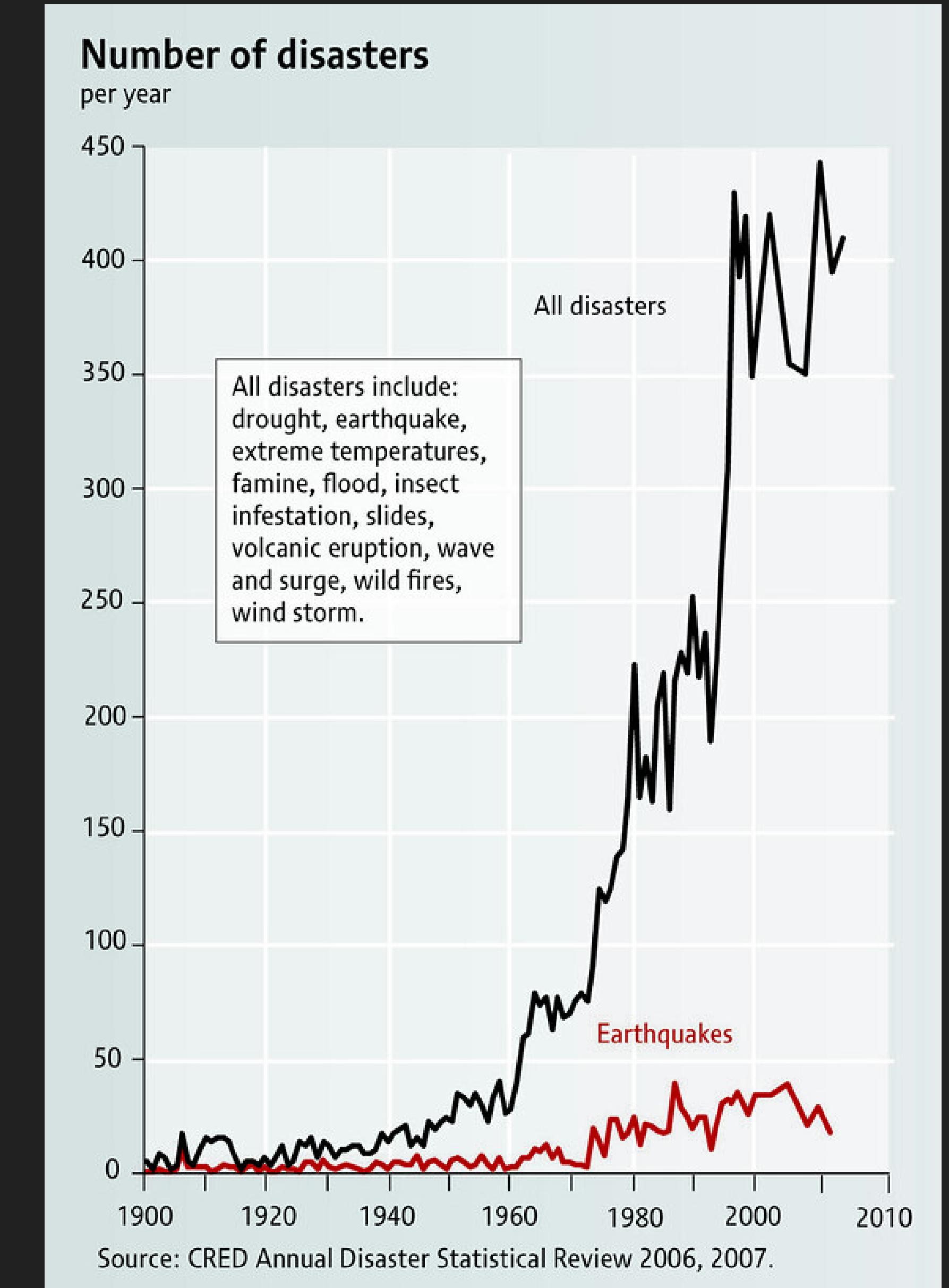
Founder and CEO, Data Literacy

**WILEY**

Our data is never a perfect  
reflection of the real world.

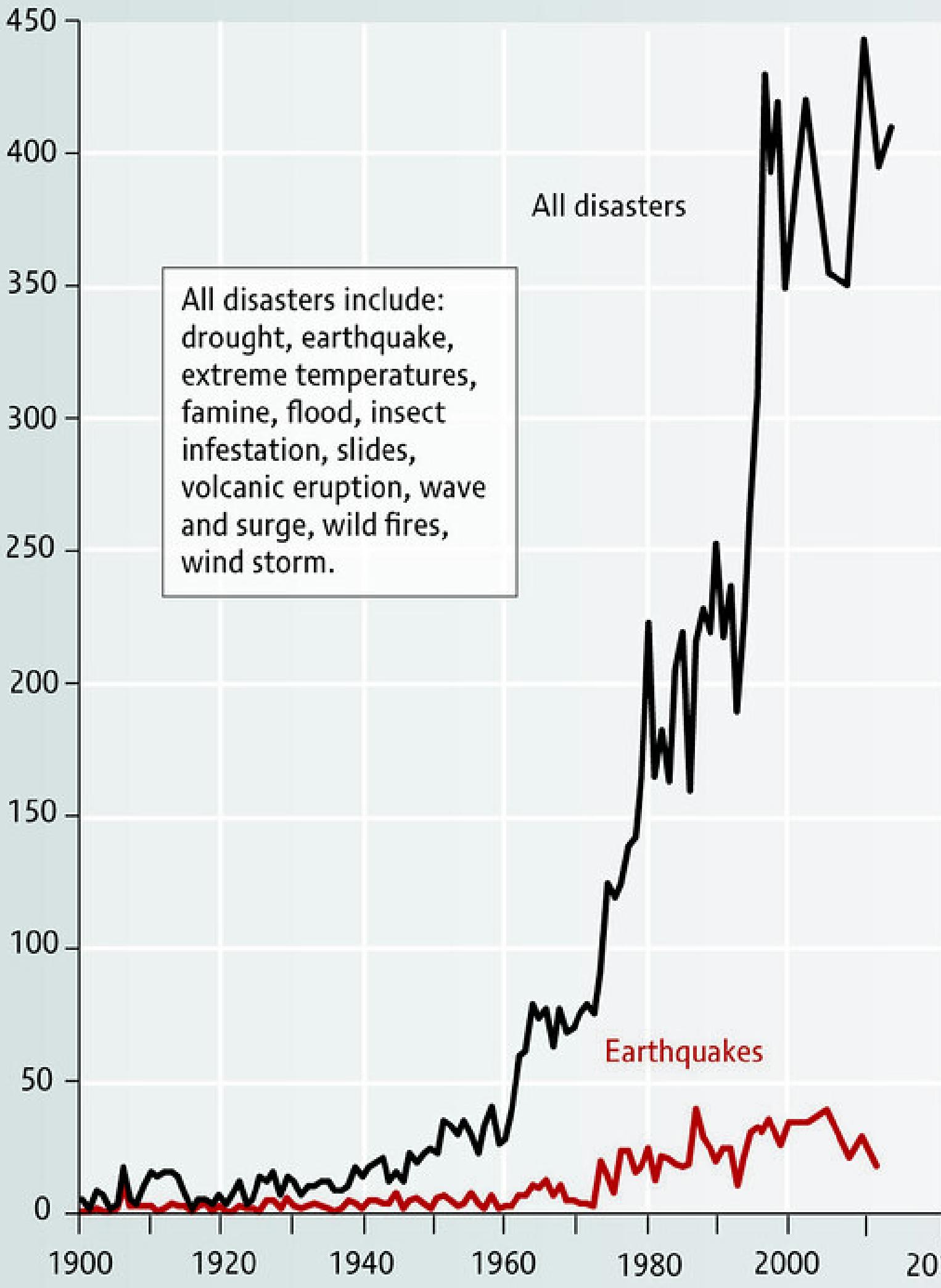
**Our data is never a perfect  
reflection of the real world.**

- **only a subset:** not crime but reported crime
- **collected by humans:** guesstimation, precision and errors
- **collected by machines:** precisions and errors



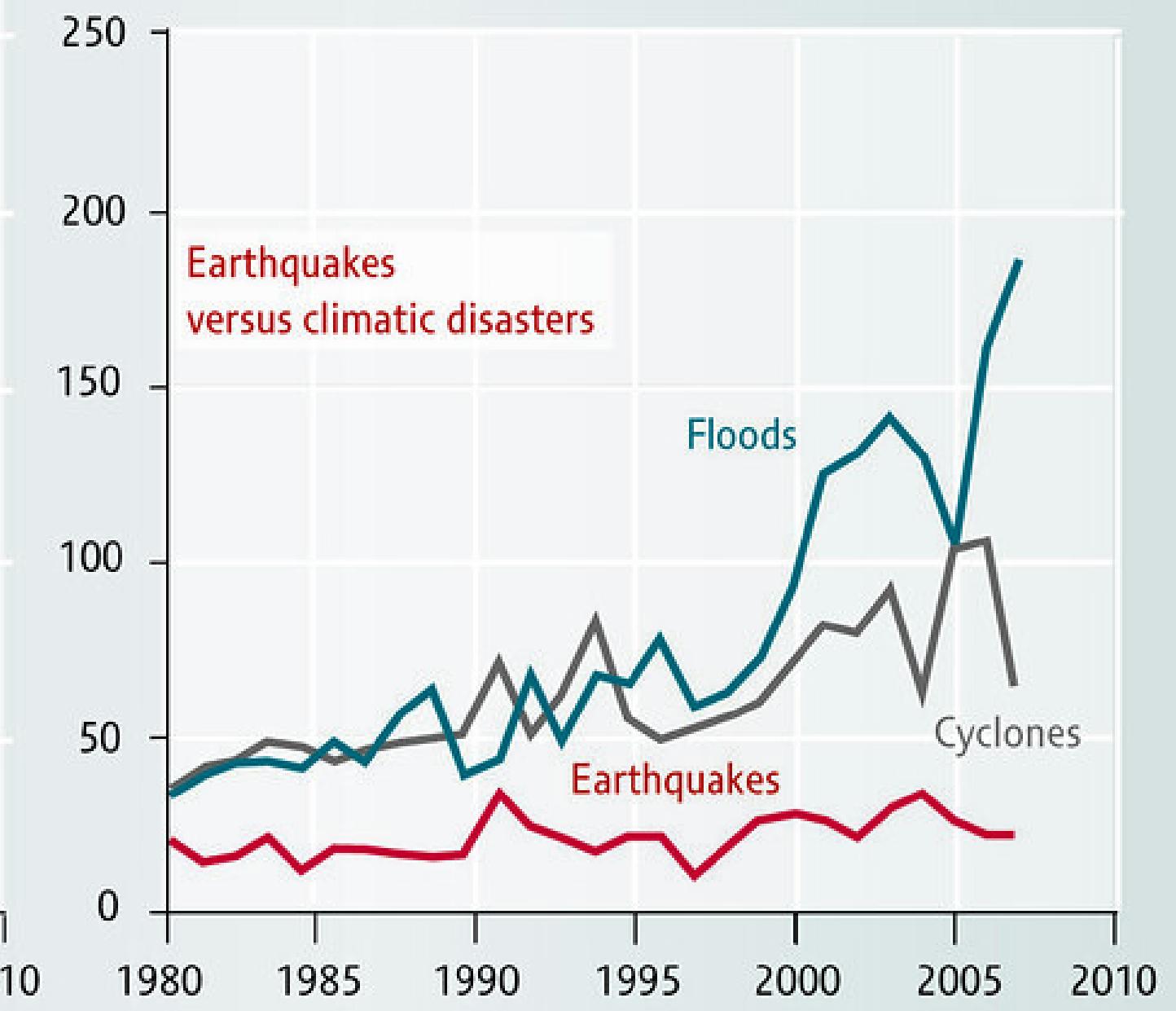
## Number of disasters

per year

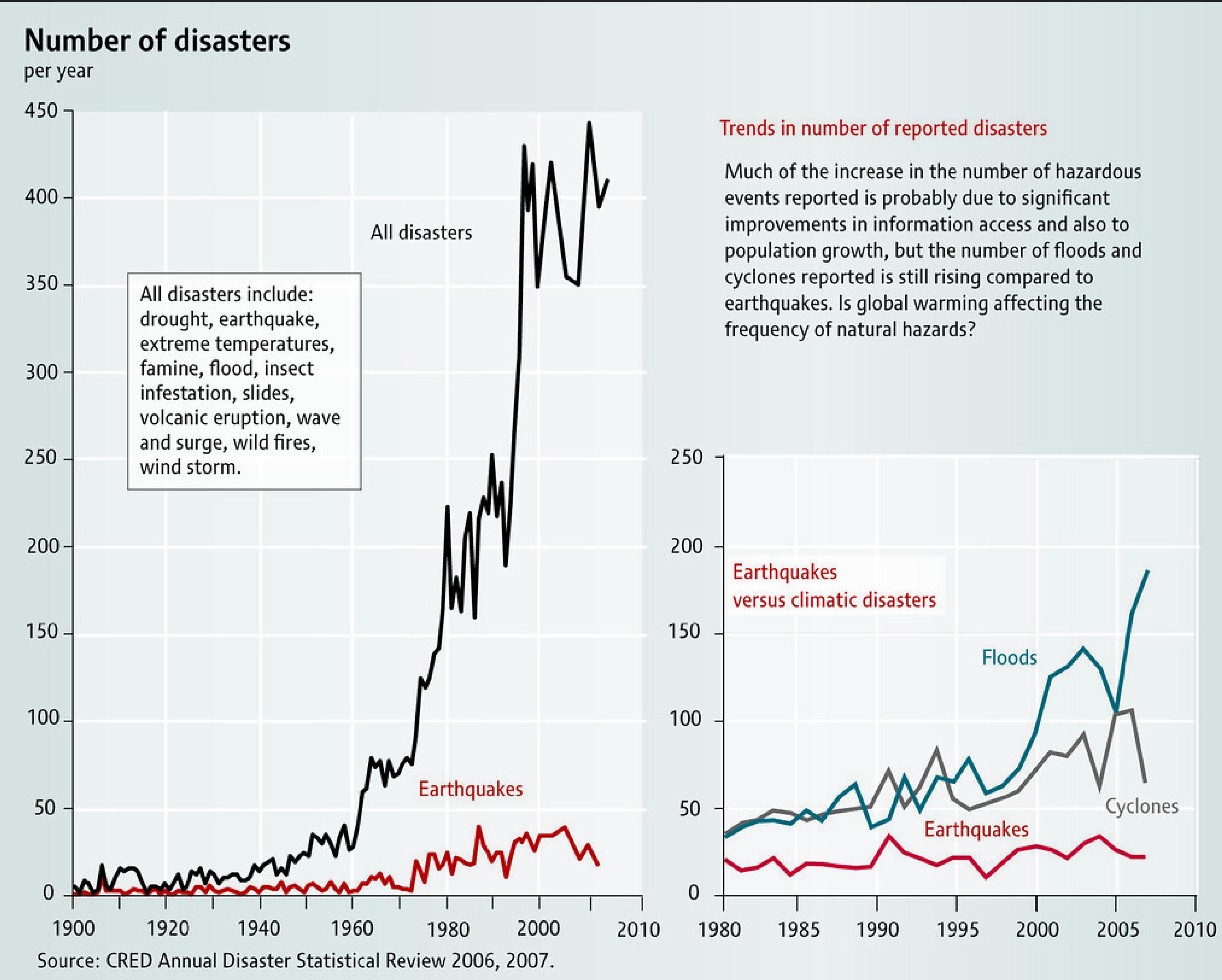


### Trends in number of reported disasters

Much of the increase in the number of hazardous events reported is probably due to significant improvements in information access and also to population growth, but the number of floods and cyclones reported is still rising compared to earthquakes. Is global warming affecting the frequency of natural hazards?

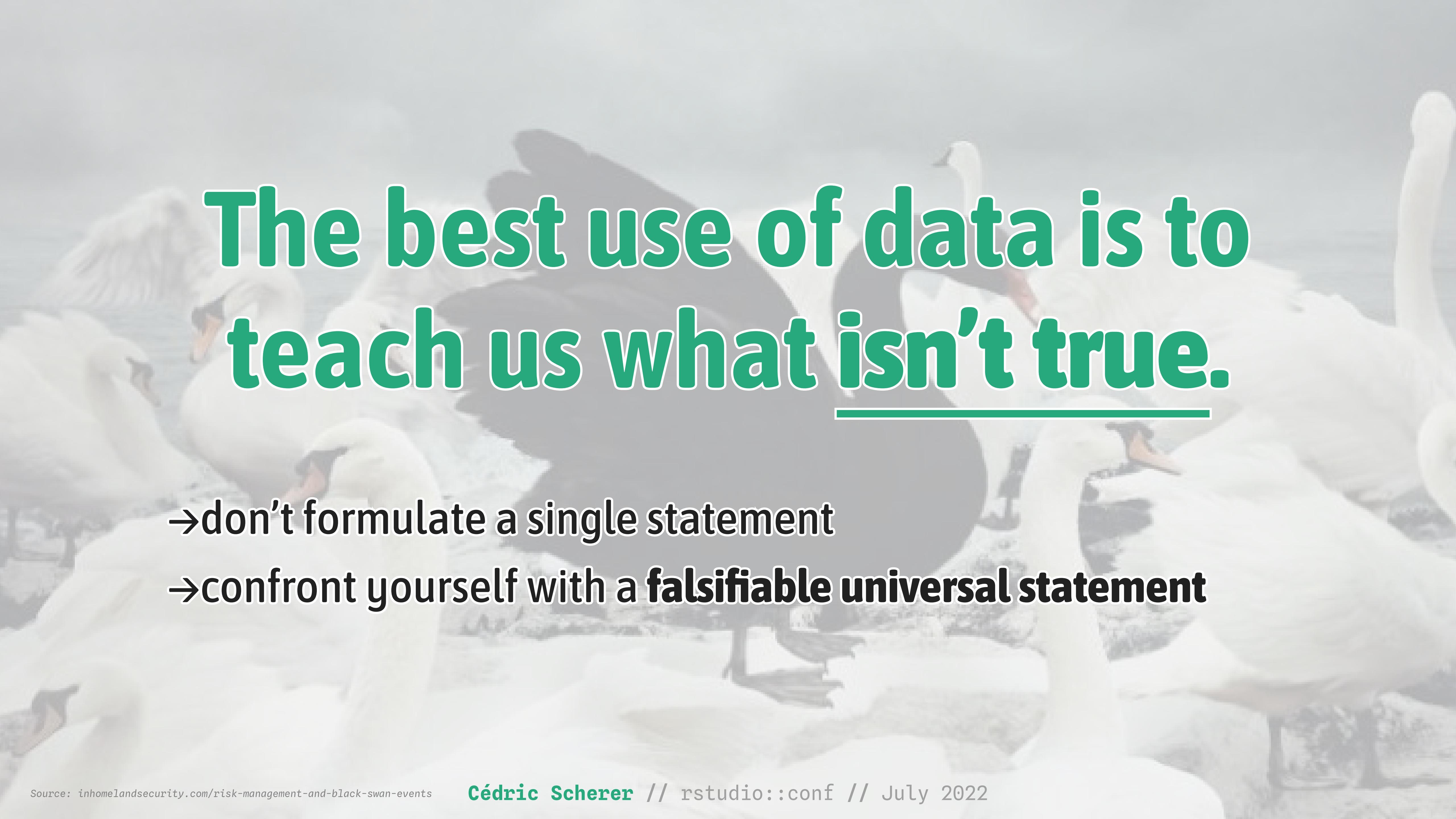


“Much of the increase of **hazardous events reported** is probably due to significant **improvements in information access**”



The best use of data is to  
teach us what isn't true.

---



The best use of data is to  
teach us what isn't true.

---

- don't formulate a single statement
- confront yourself with a **falsifiable universal statement**



The best use of data is to  
teach us what isn't true.

---

# Story

---

**Be clear about the message of your visualization**

# Who is my audience?

# Who is my audience?

Which story is **interesting** for them?

# Who is my audience?

Which story is **interesting** for them?

What are **relevant** details to include?

# Who is my audience?

Which story is **interesting** for them?

What are **relevant** details to include?

Which variables are **meaningful** to them?

# Who is my audience?

Which story is **interesting** for them?

What are **relevant** details to include?

Which variables are **meaningful** to them?

How will they **encounter** the visualization?

# Who is my audience?

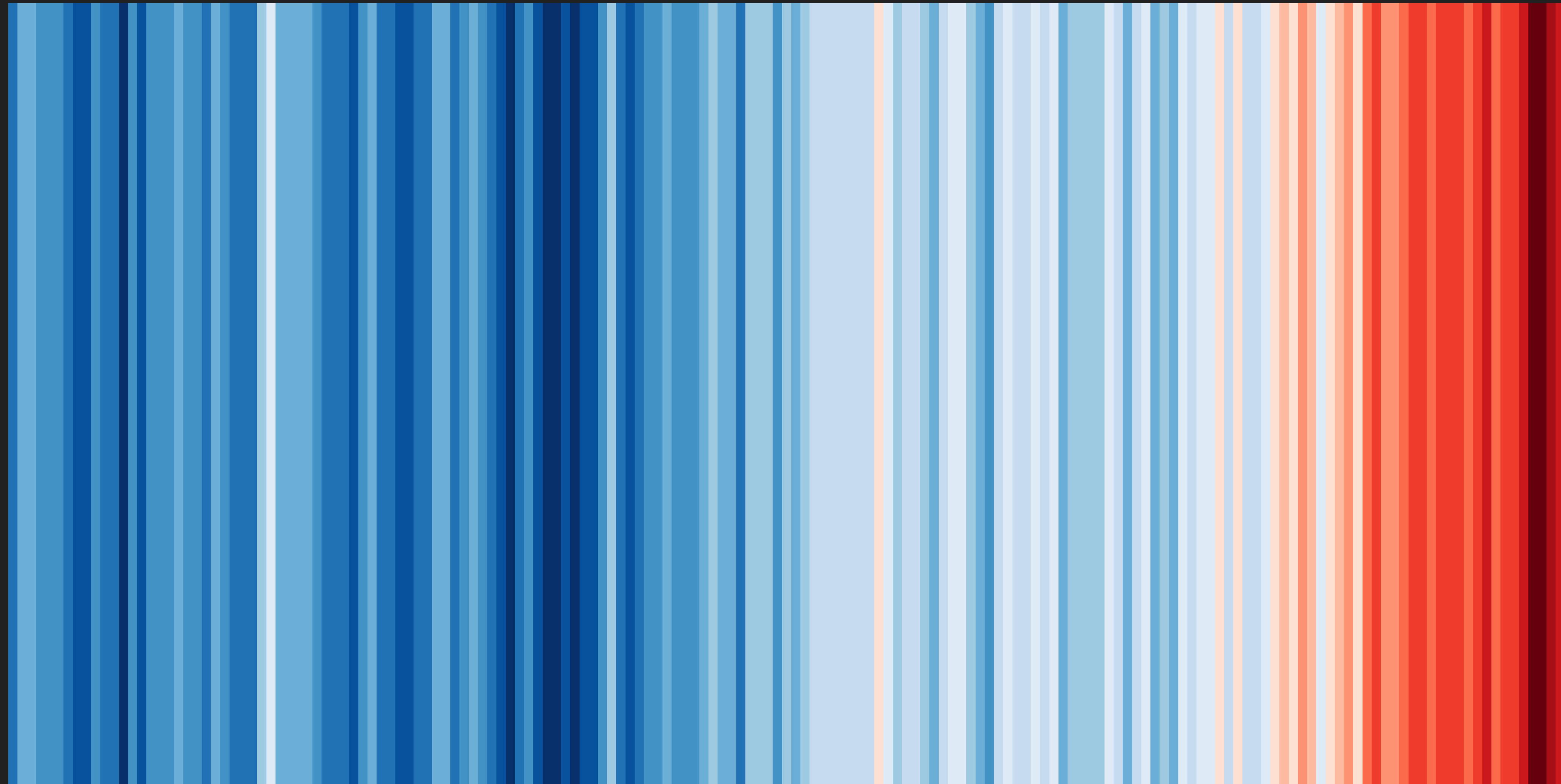
Which story is **interesting** for them?

What are **relevant** details to include?

Which variables are **meaningful** to them?

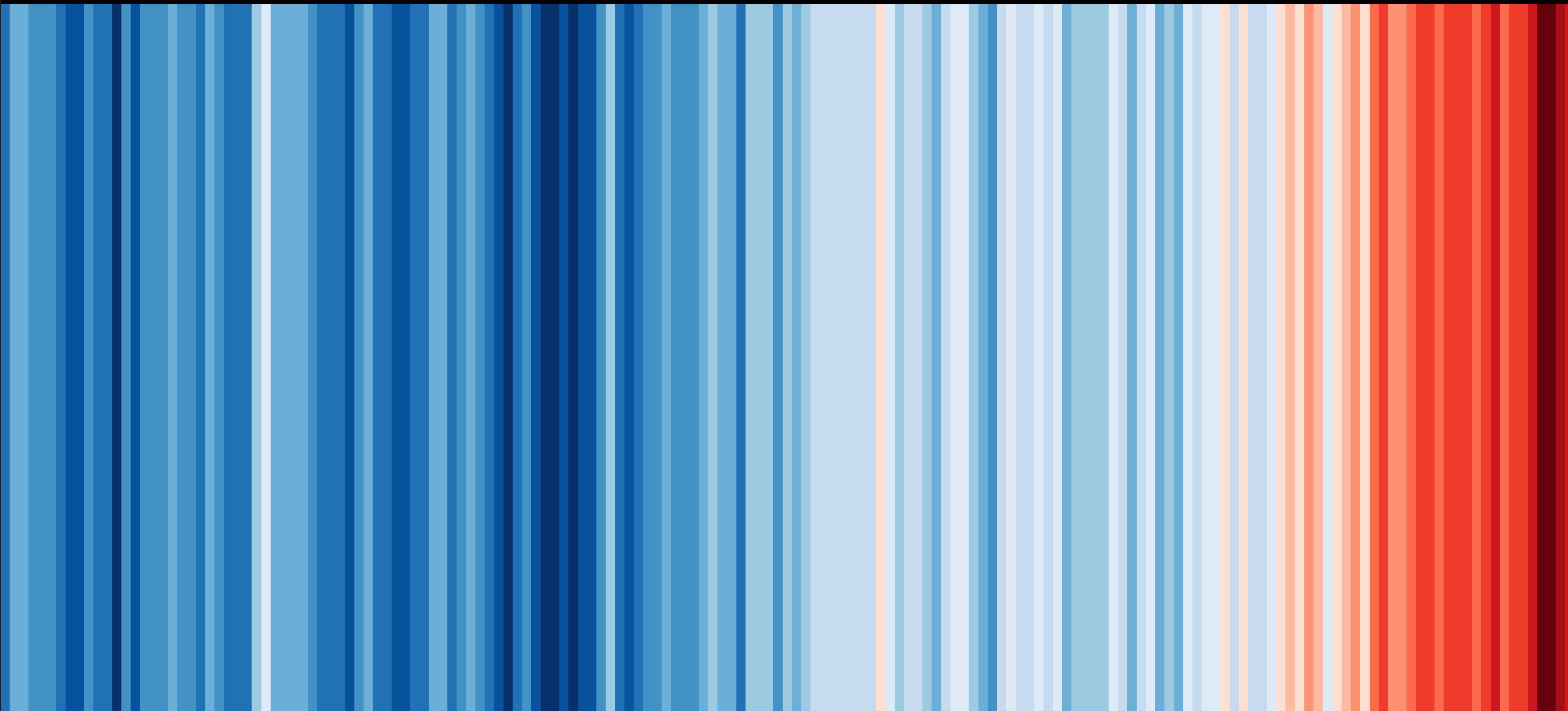
How will they **encounter** the visualization?

**Do I need a visualization at all??**



*Warming Stripes* by Ed Hawkins

# Global temperature change (1850-2019)



*Warming Stripes* by Ed Hawkins

## FAQ : Frequently asked questions

What are these graphics?



What do the graphics show?



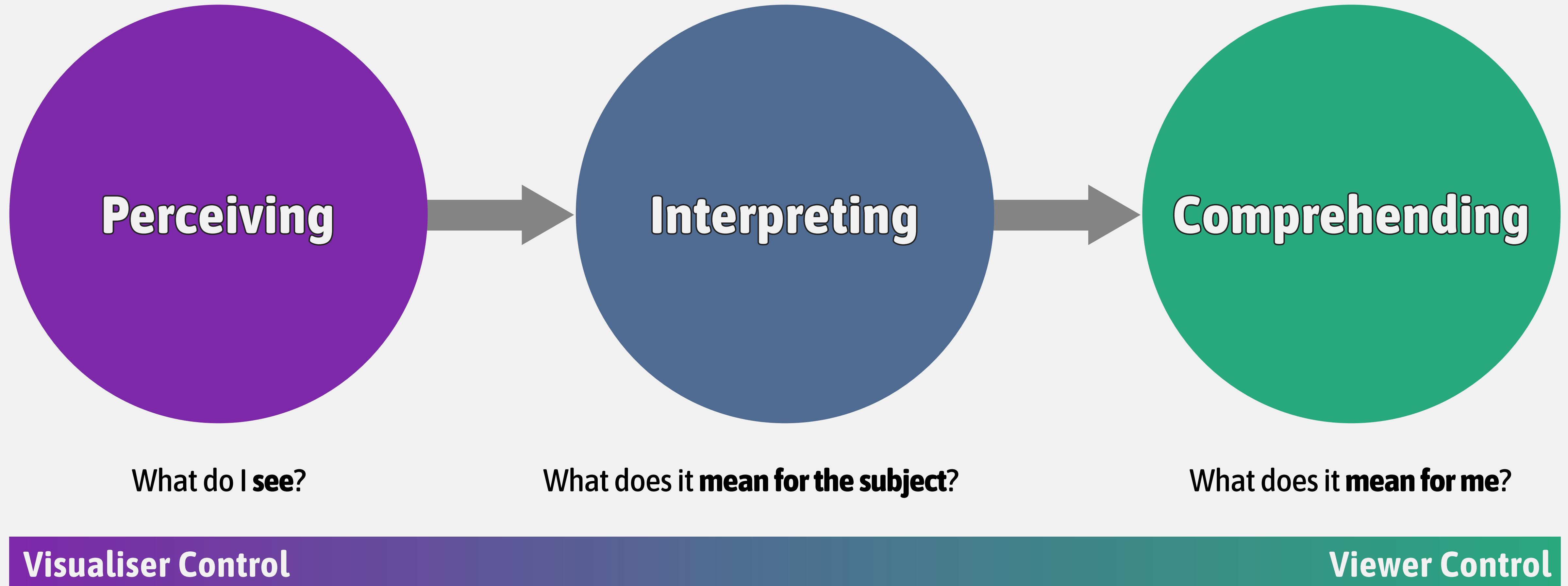
Why are there no numbers on the graphics?



» These graphics are specifically designed to be as simple as possible, and to start conversations about our warming world and the risks of climate change. There are numerous sources of information which provide more specific details about how temperatures have changed, so these graphics fill a gap and enable communication with minimal scientific knowledge required to understand their meaning.

These graphics are specifically designed to [...] start conversations about our warming world and the risks of climate change.

» These graphics are specifically designed to be as simple as possible, and to start conversations about our warming world and the risks of climate change. There are numerous sources of information which provide more specific details about how temperatures have changed, so these graphics fill a gap and enable communication with minimal scientific knowledge required to understand their meaning.

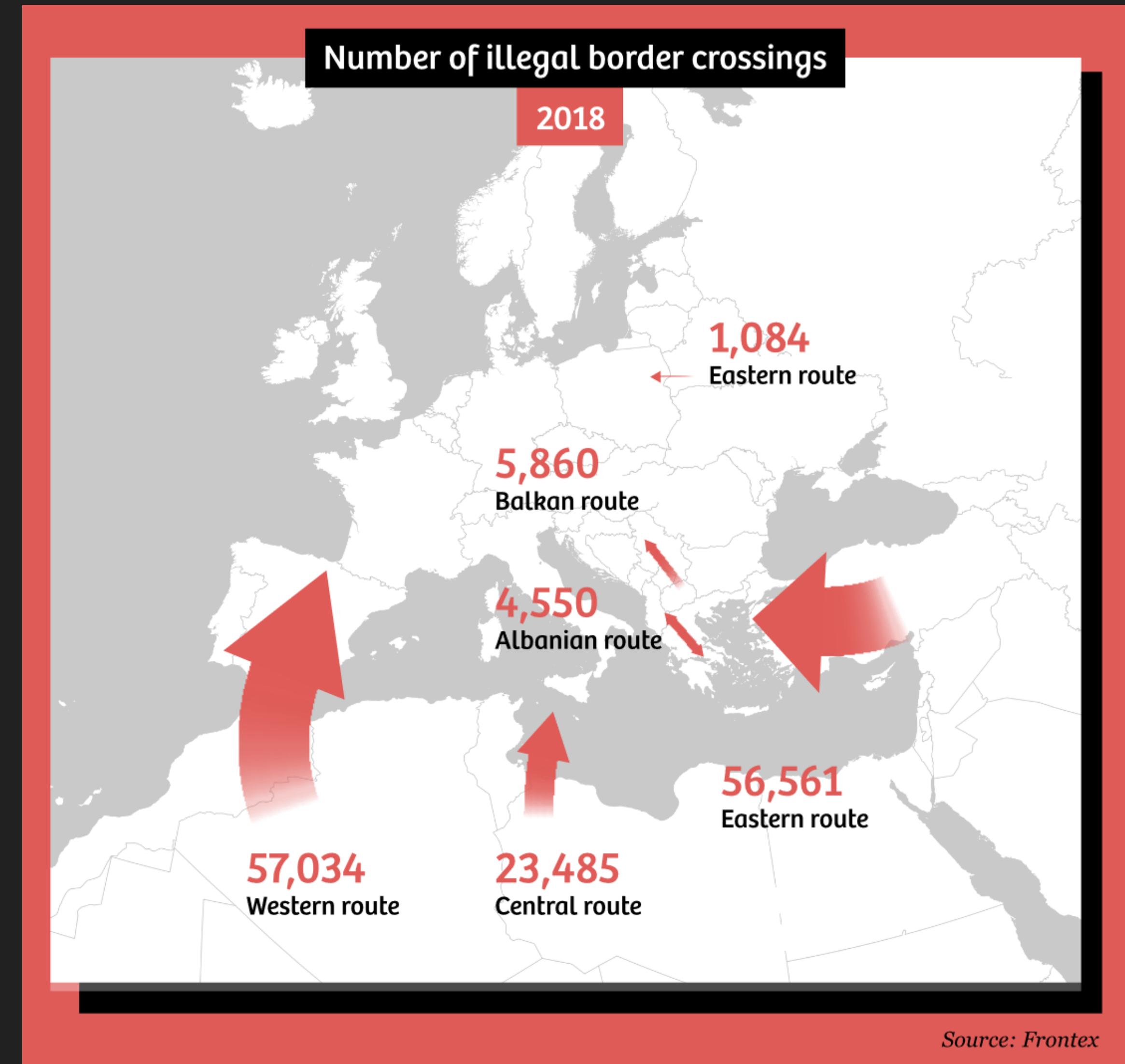


*Scheme by Andy Kirk*

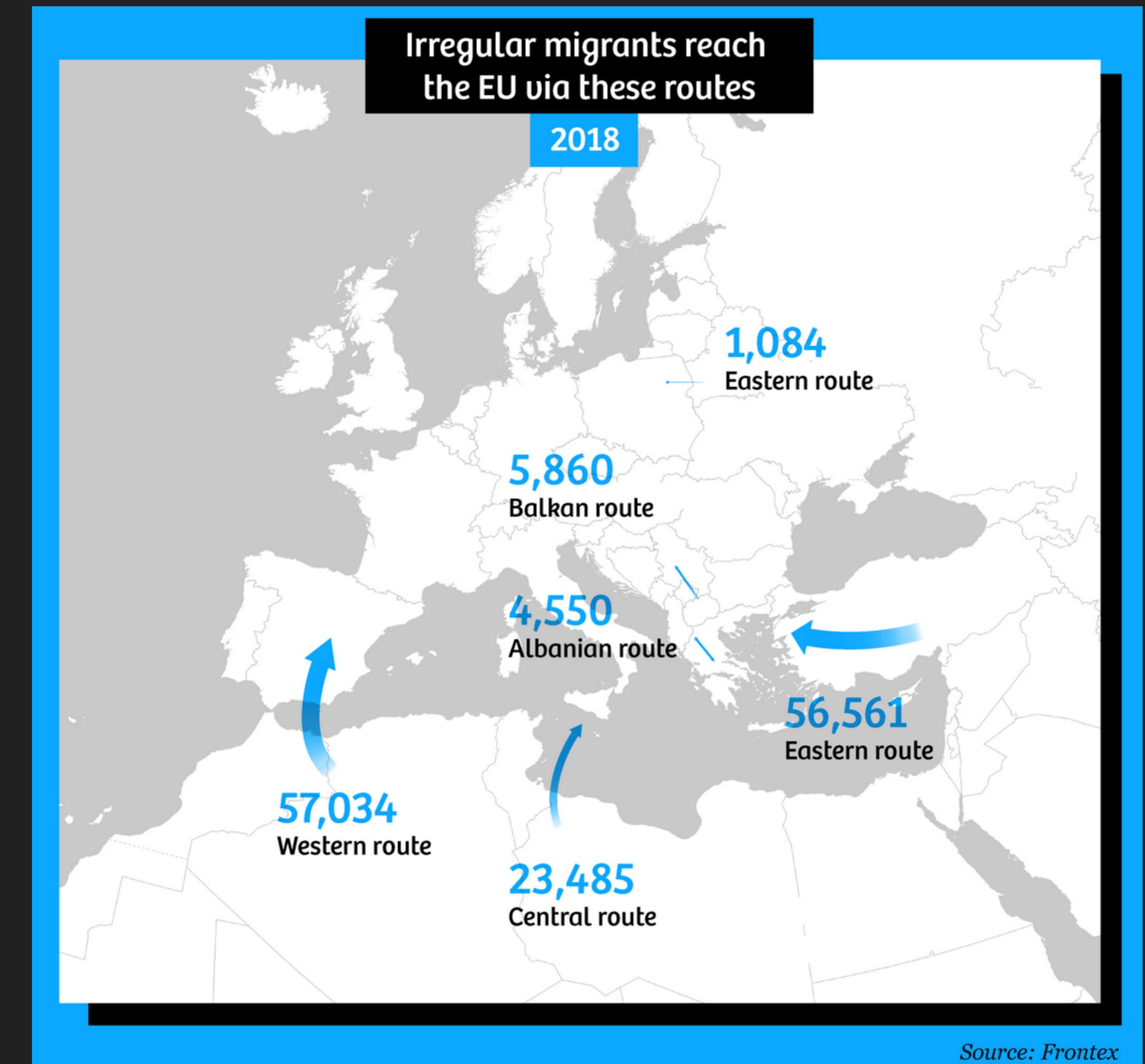
# Goal

---

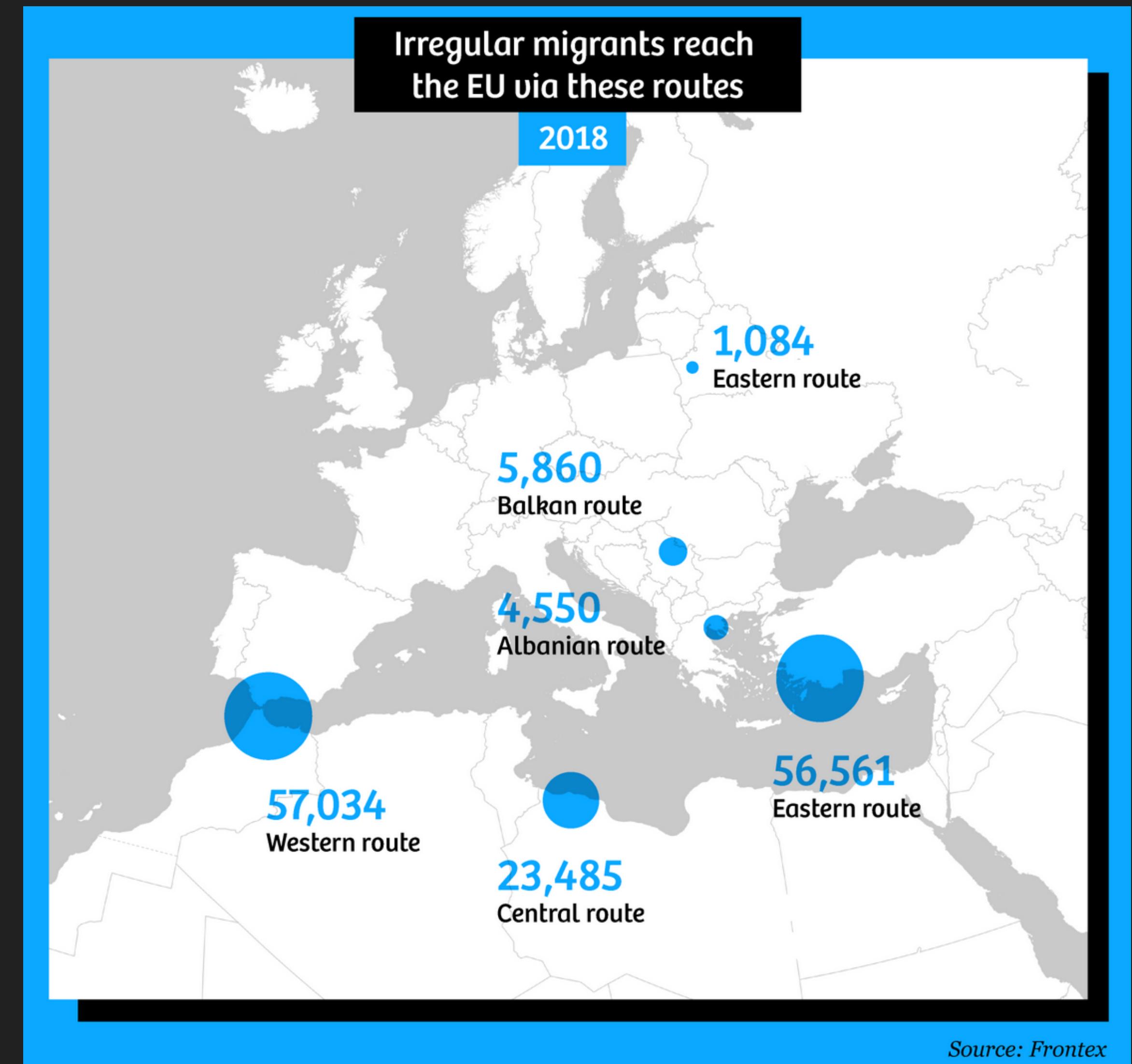
**Select charts that successfully transport your story**



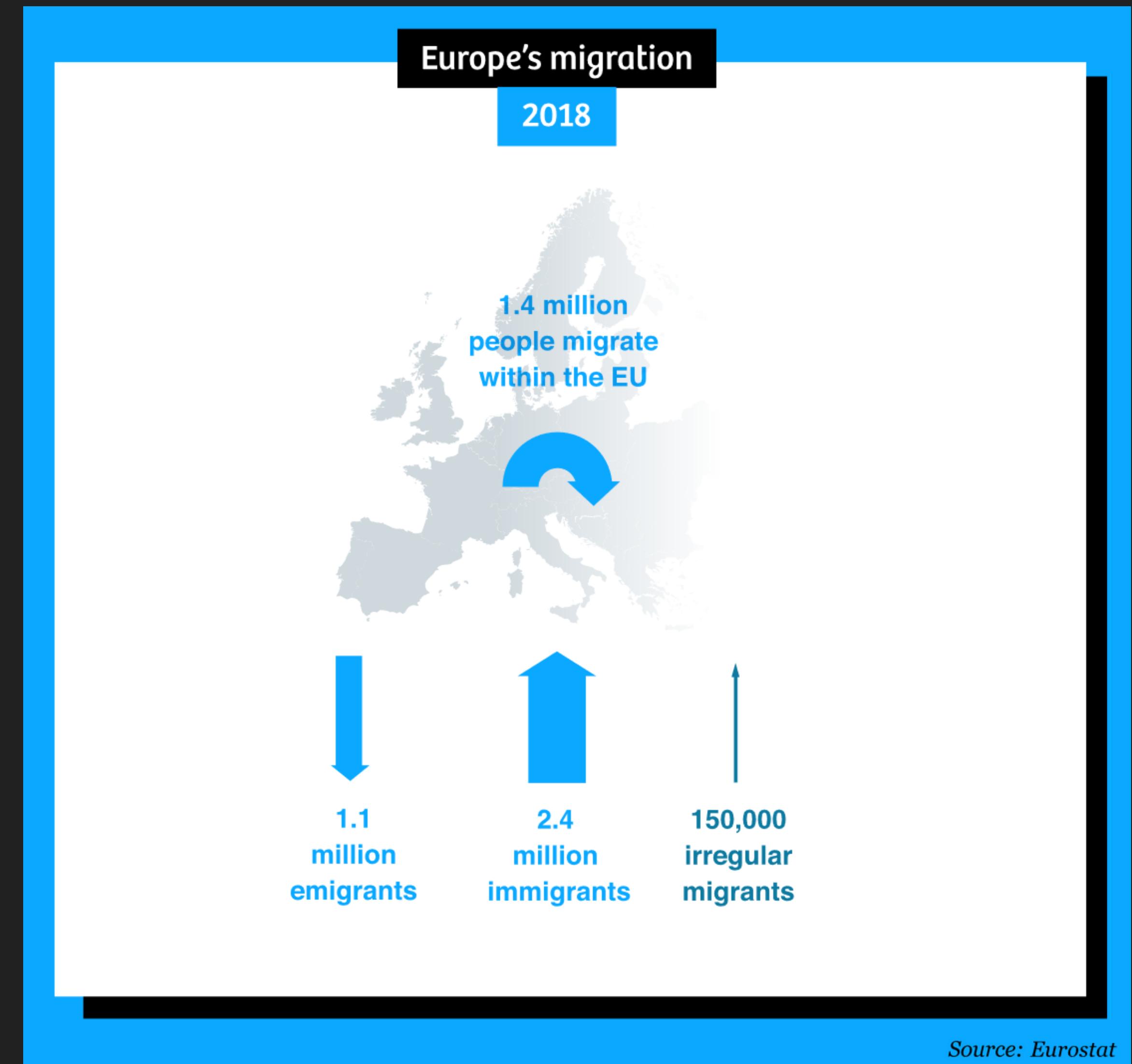
*“How maps in the media make us more negative about migrants” by Maite Vermeulen, Leon de Korte & Henk van Houtum*



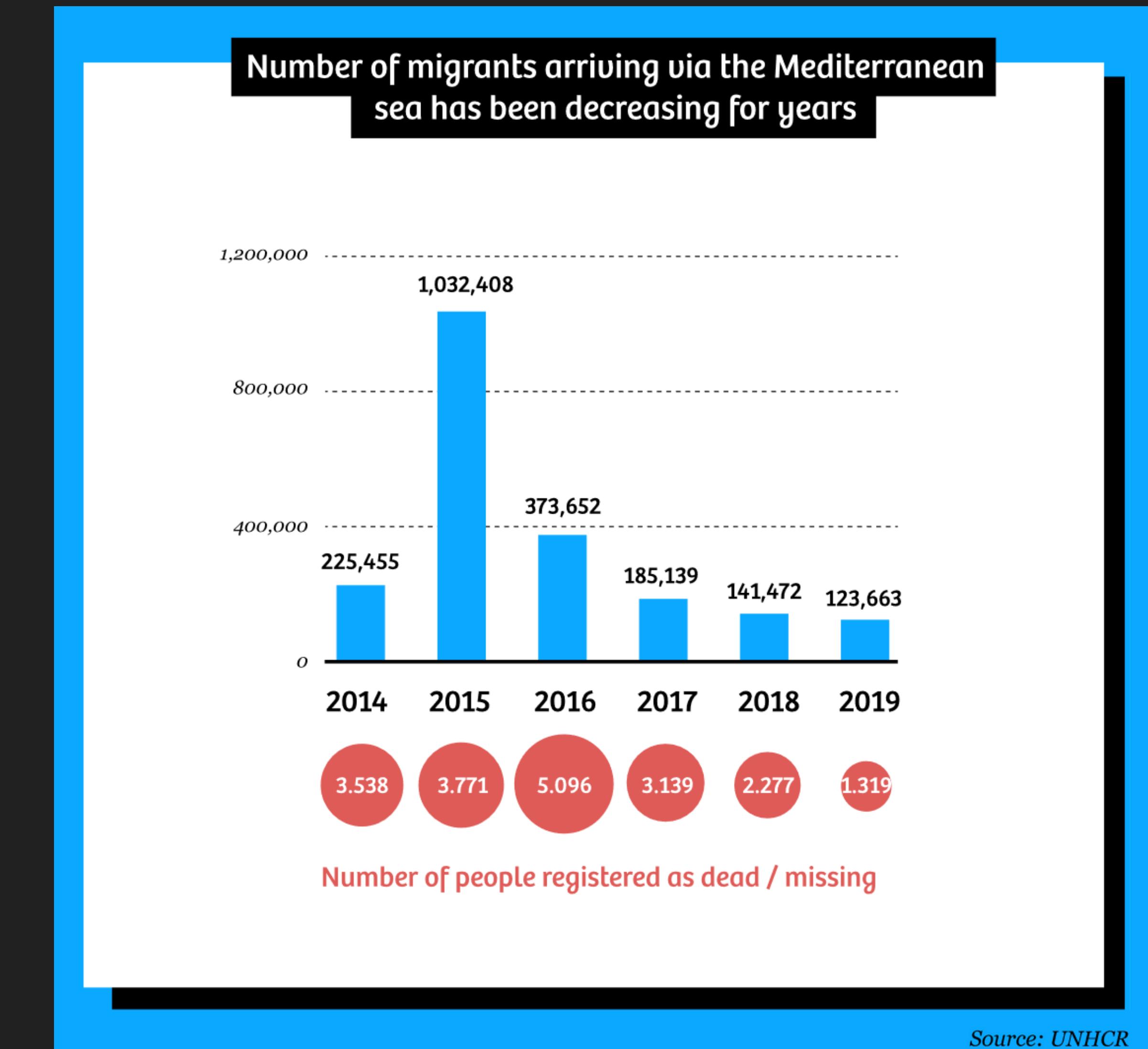
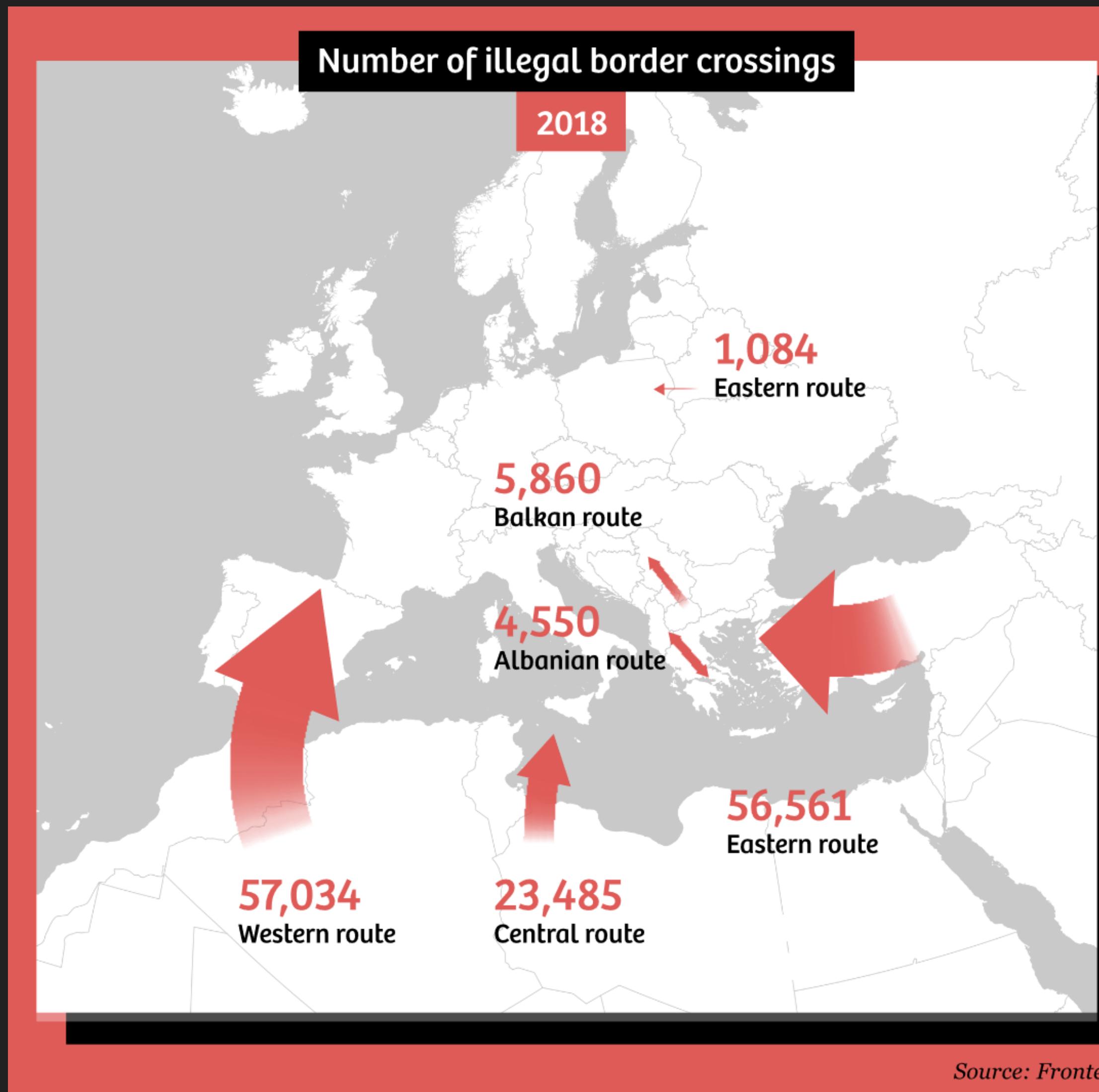
*“How maps in the media make us more negative about migrants” by Maite Vermeulen, Leon de Korte & Henk van Houtum*



*“How maps in the media make us more negative about migrants” by Maite Vermeulen, Leon de Korte & Henk van Houtum*



*“How maps in the media make us more negative about migrants” by Maite Vermeulen, Leon de Korte & Henk van Houtum*



*“How maps in the media make us more negative about migrants” by Maite Vermeulen, Leon de Korte & Henk van Houtum*

# Typology of Information Graphics

by Juuso Koponen & Jonatan Hildén, "Data Visualization Handbook" (2020), p. 25

**Is the information conceptual or measurable?**

👉 **Type of information:** depict conceptual information <> convert information into visual forms

# Typology of Information Graphics

by Juuso Koponen & Jonatan Hildén, "Data Visualization Handbook" (2020), p. 25

**Is the information conceptual or measurable?**

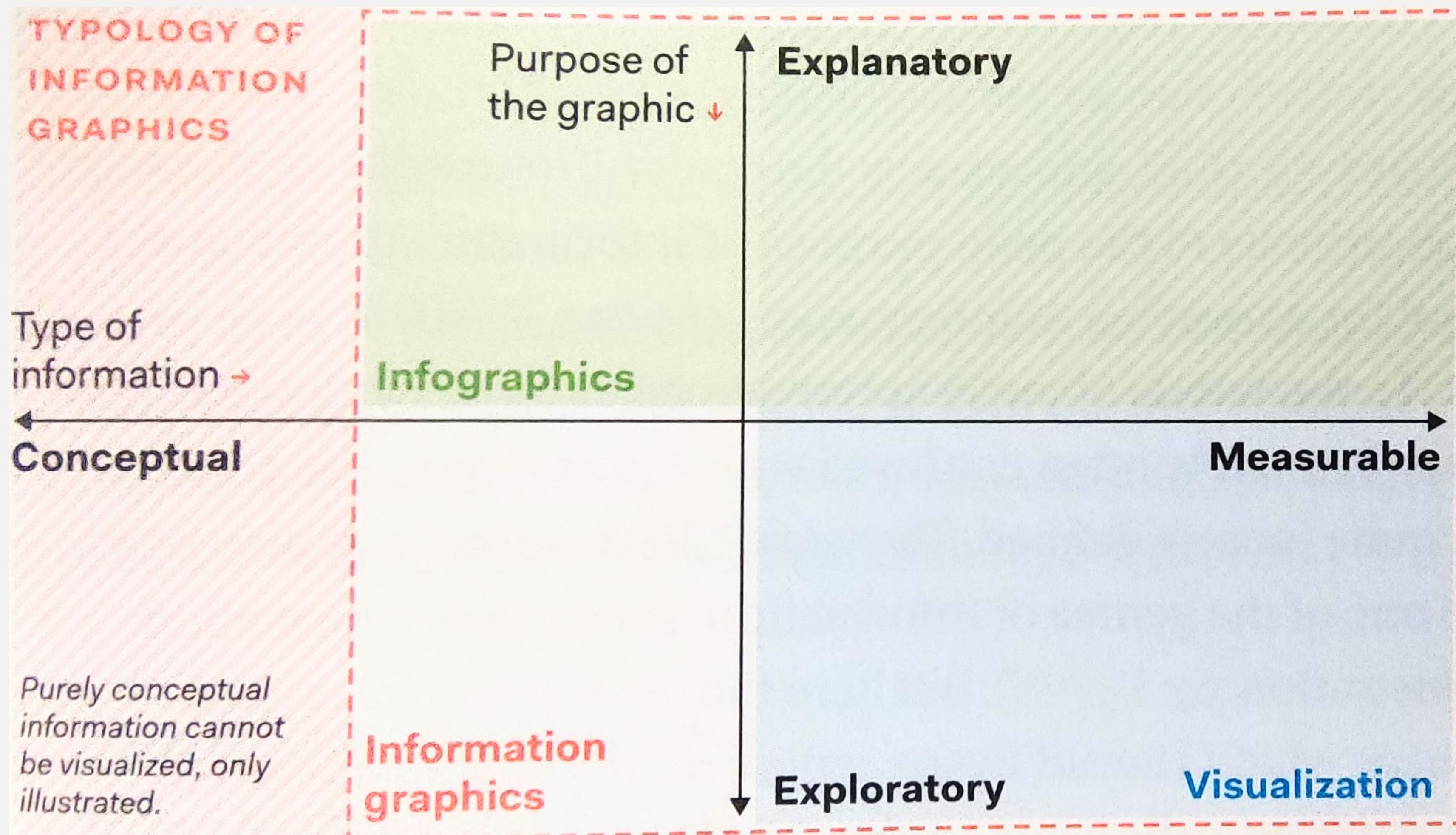
👉 **Type of information:** depict conceptual information <> convert information into visual forms

**Is the purpose to explore or to explain the information?**

👉 **Purpose of the graphic:** facilitate discovery <> communicate information

# Typology of Information Graphics

by Juuso Koponen & Jonatan Hildén, "Data Visualization Handbook" (2020), p. 25



“Visualizations can be designed and experienced in various ways, by people of various backgrounds, and in various circumstances. That's why **reflecting on the purpose of a visualization is paramount before we design it—or before we critique it.**”

Alberto Cairo

*Excerpt from the foreword to “Data Sketches” by Nadieh Bremer & Shirley Wu (CRC Press 2021)*

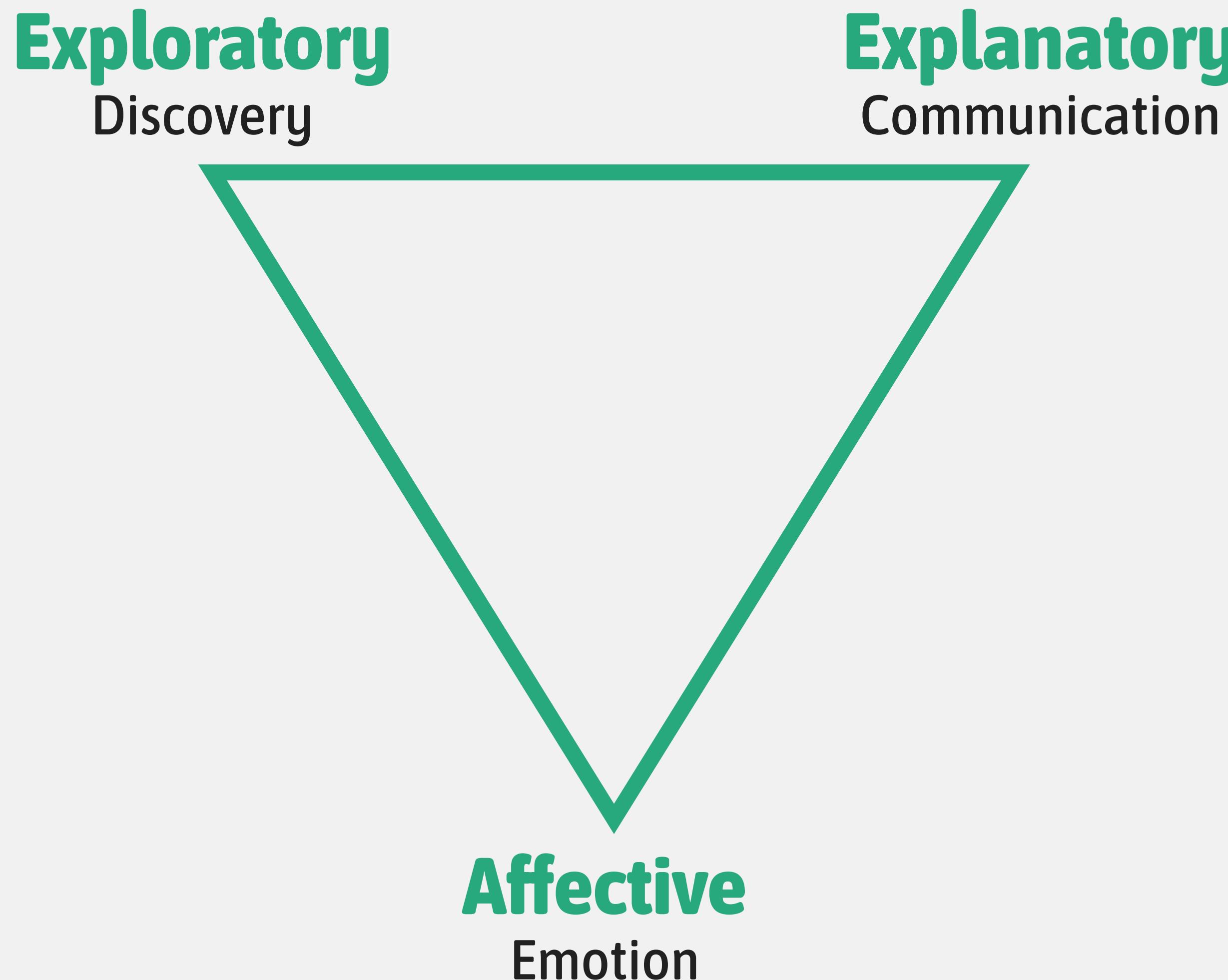
**“A common truism about information visualization is  
that it is primarily about ‘showing the data’. [...]”**

While this might be true for scientific (or financial, or many other) application fields, there are many good uses of visualization that go beyond a precise, “neutral” display of data.“

*Moritz Stefaner*

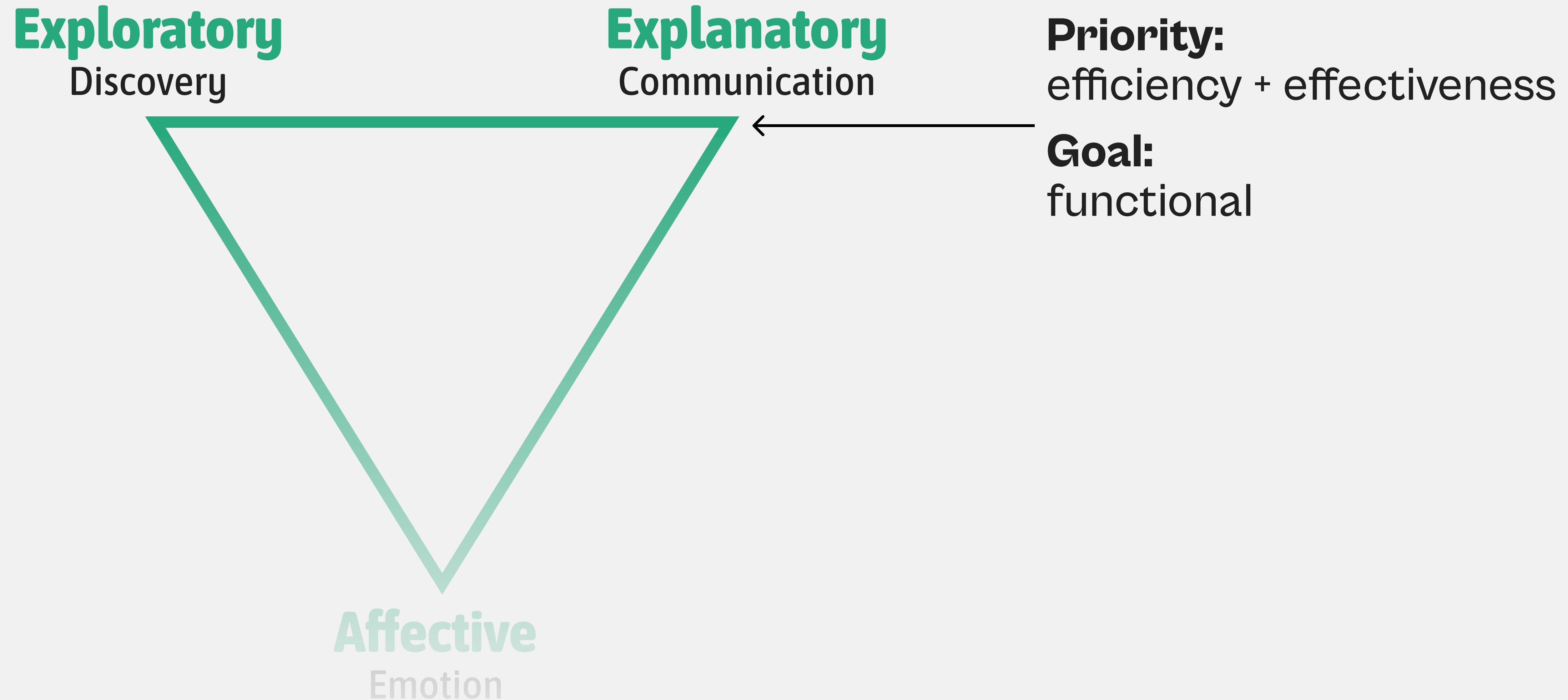
# The Vertices of Visualization

by Alberto Cairo, personal communication



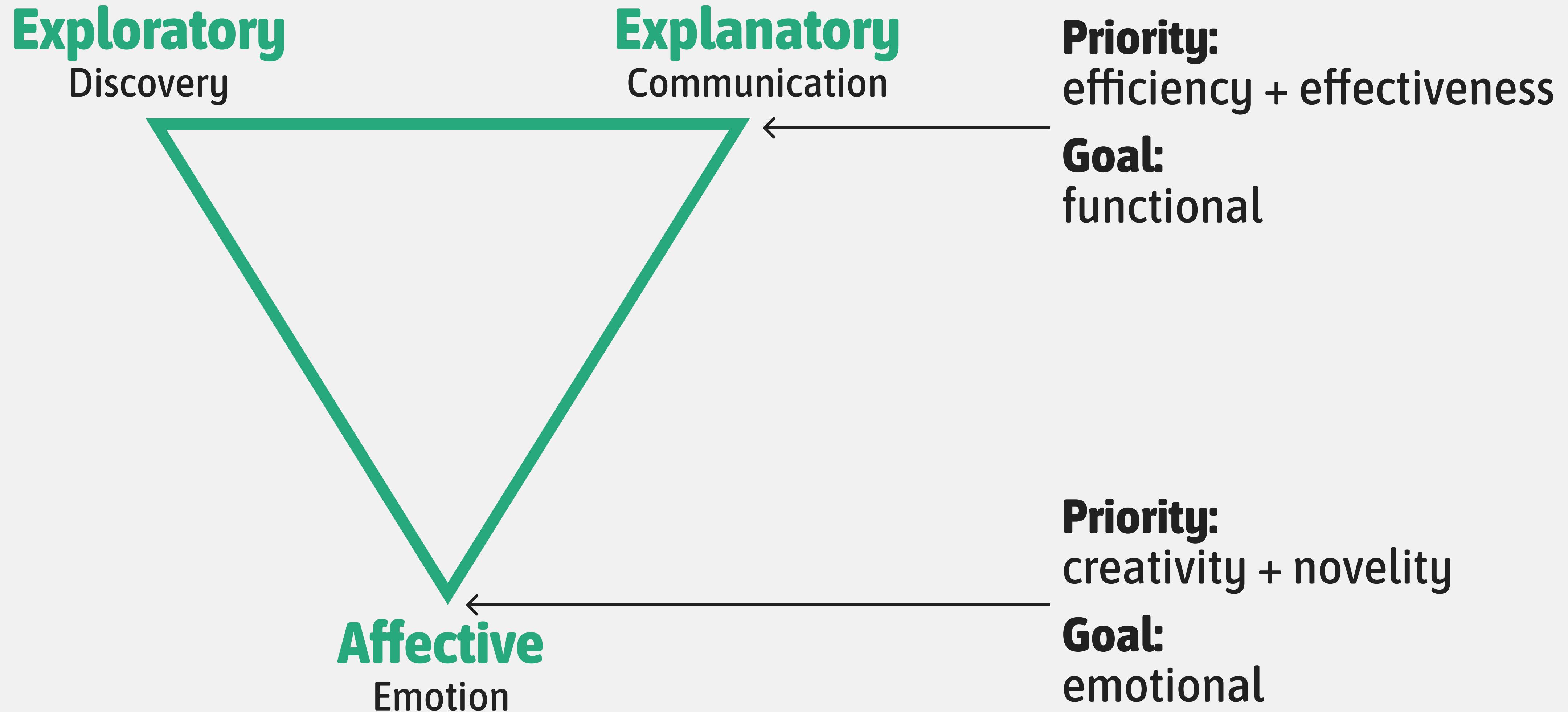
# The Vertices of Visualization

by Alberto Cairo, personal communication



# The Vertices of Visualization

by Alberto Cairo, personal communication



PERSPECTIVE

# Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm

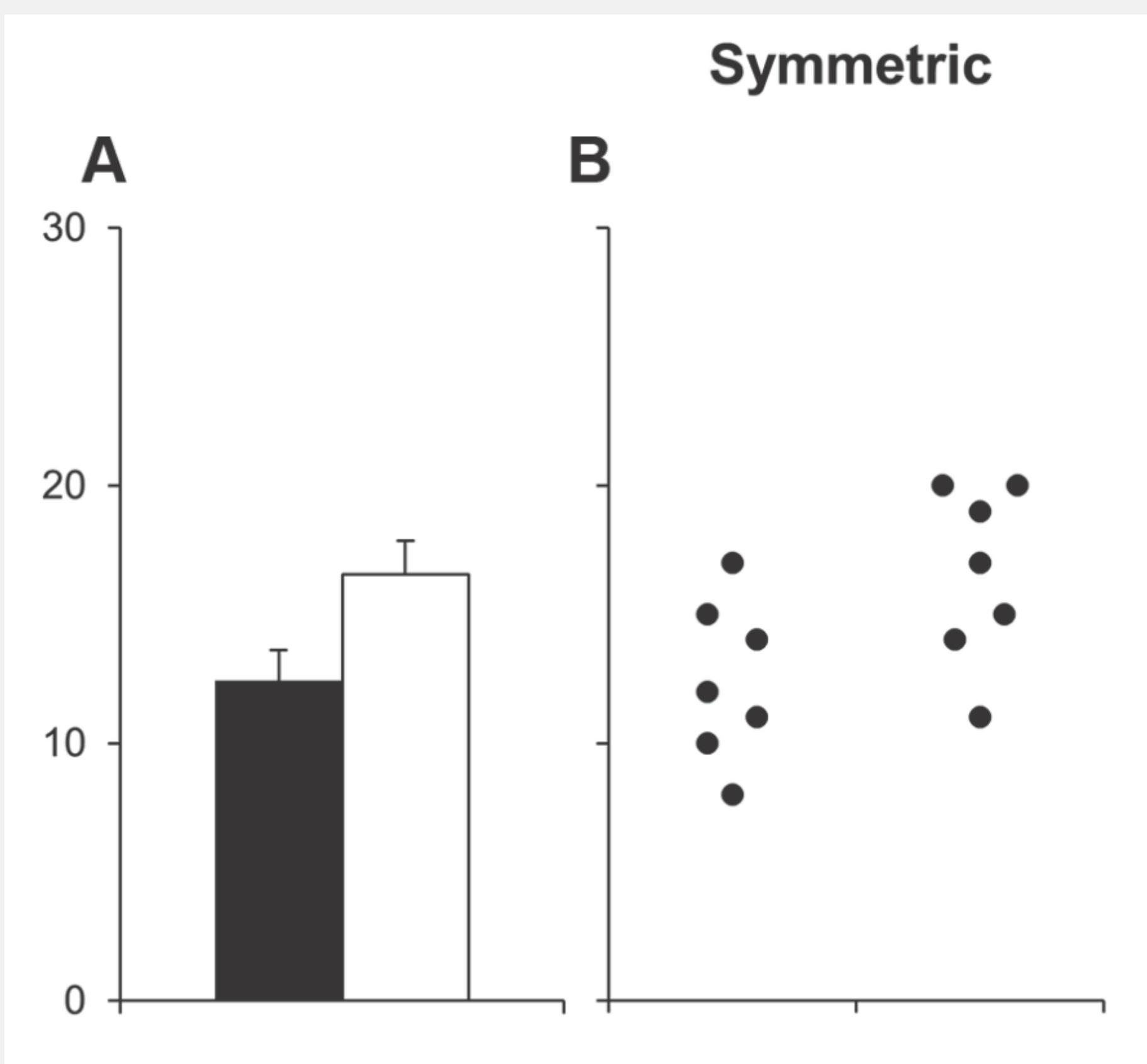
**Tracey L. Weissgerber<sup>1</sup>\*, Nataša M. Milic<sup>1,2</sup>, Stacey J. Winham<sup>3</sup>, Vesna D. Garovic<sup>1</sup>**

**1** Division of Nephrology & Hypertension, Mayo Clinic, Rochester, Minnesota, United States of America,  
**2** Department of Biostatistics, Medical Faculty, University of Belgrade, Belgrade, Serbia, **3** Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, Minnesota, United States of America

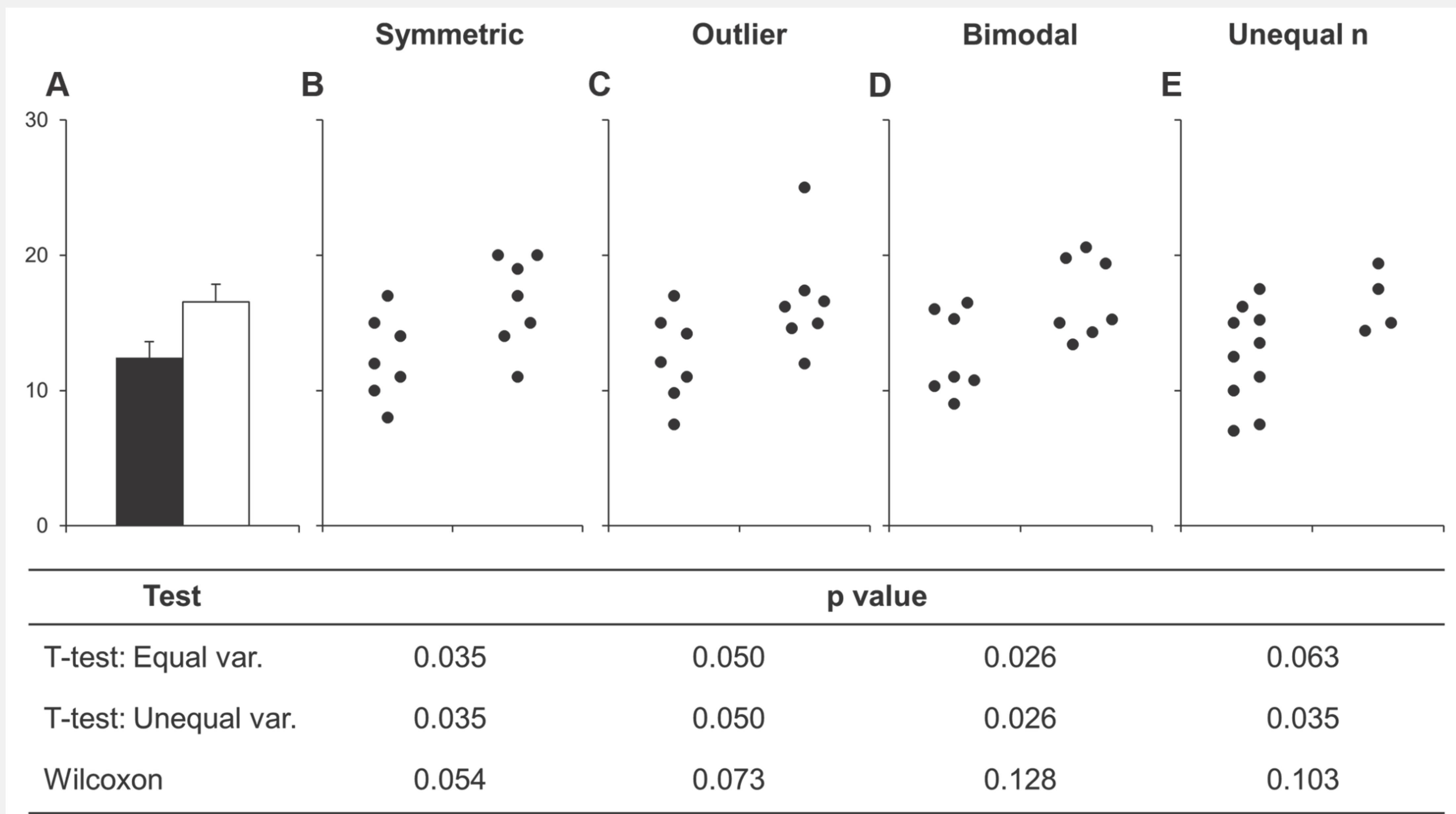
\* [weissgerber.tracey@mayo.edu](mailto:weissgerber.tracey@mayo.edu)

*Weissgerber et al. (2015) PLoS Biology*

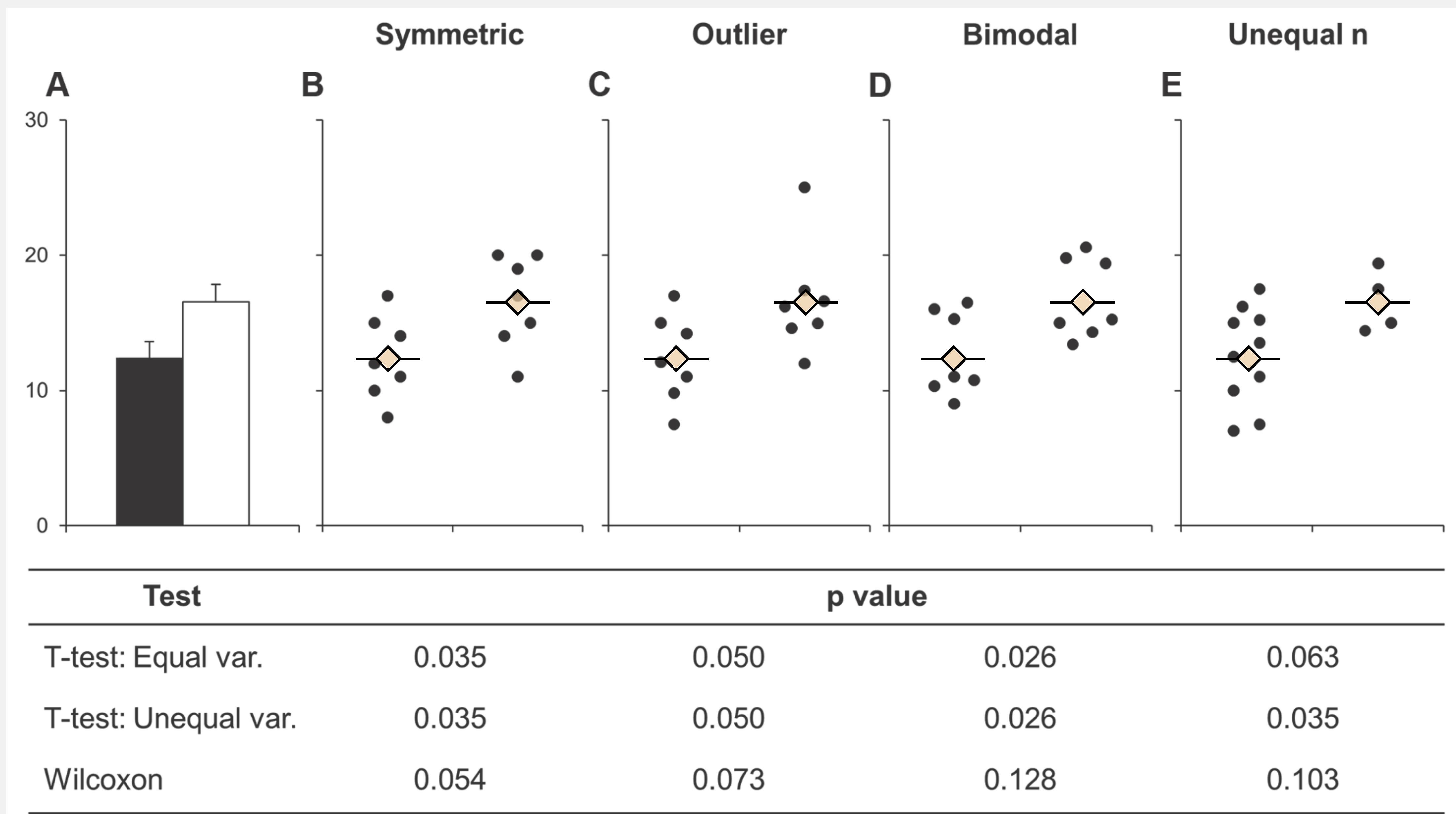
## Symmetric



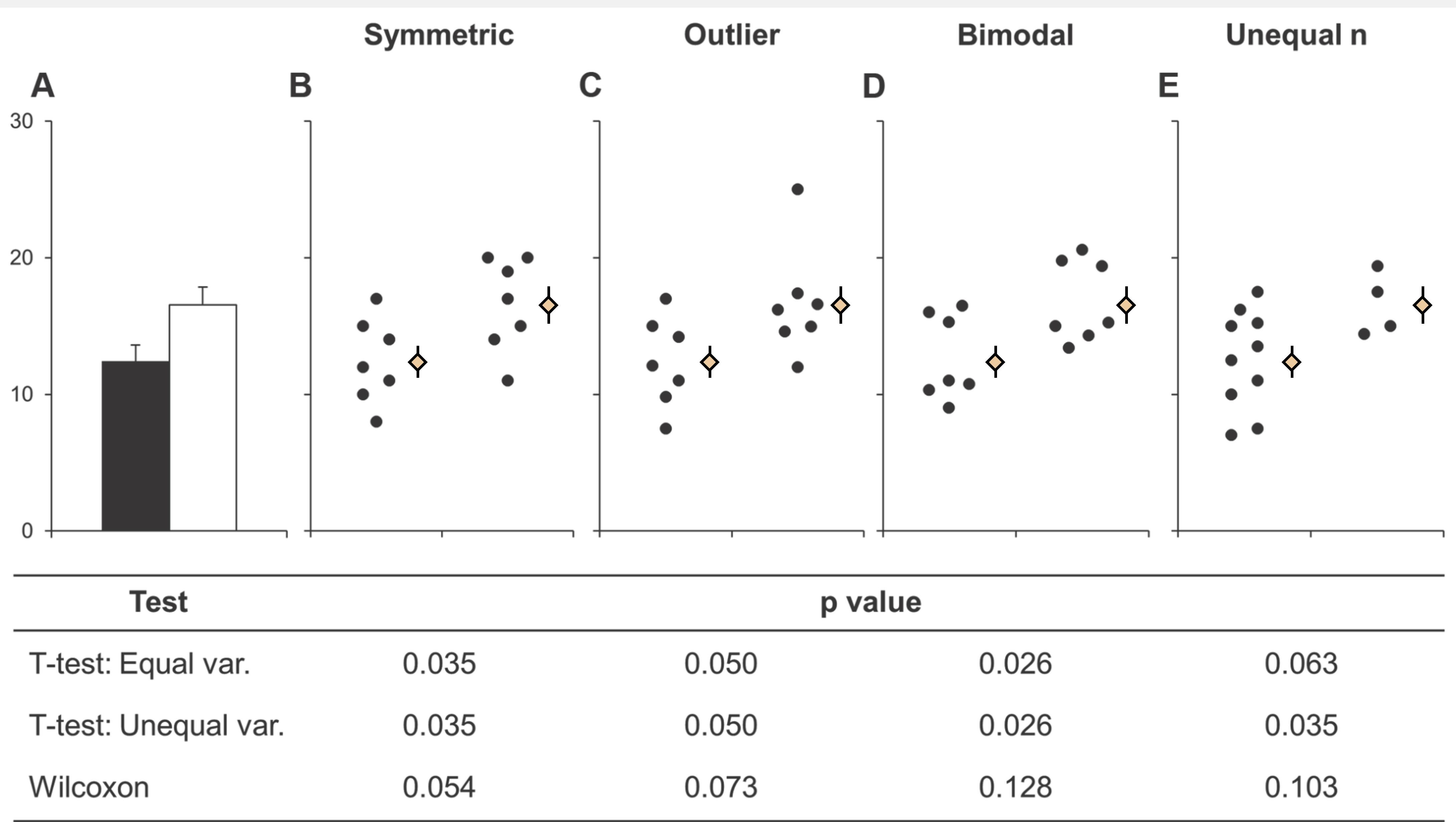
Weissgerber et al. (2015) PLoS Biology



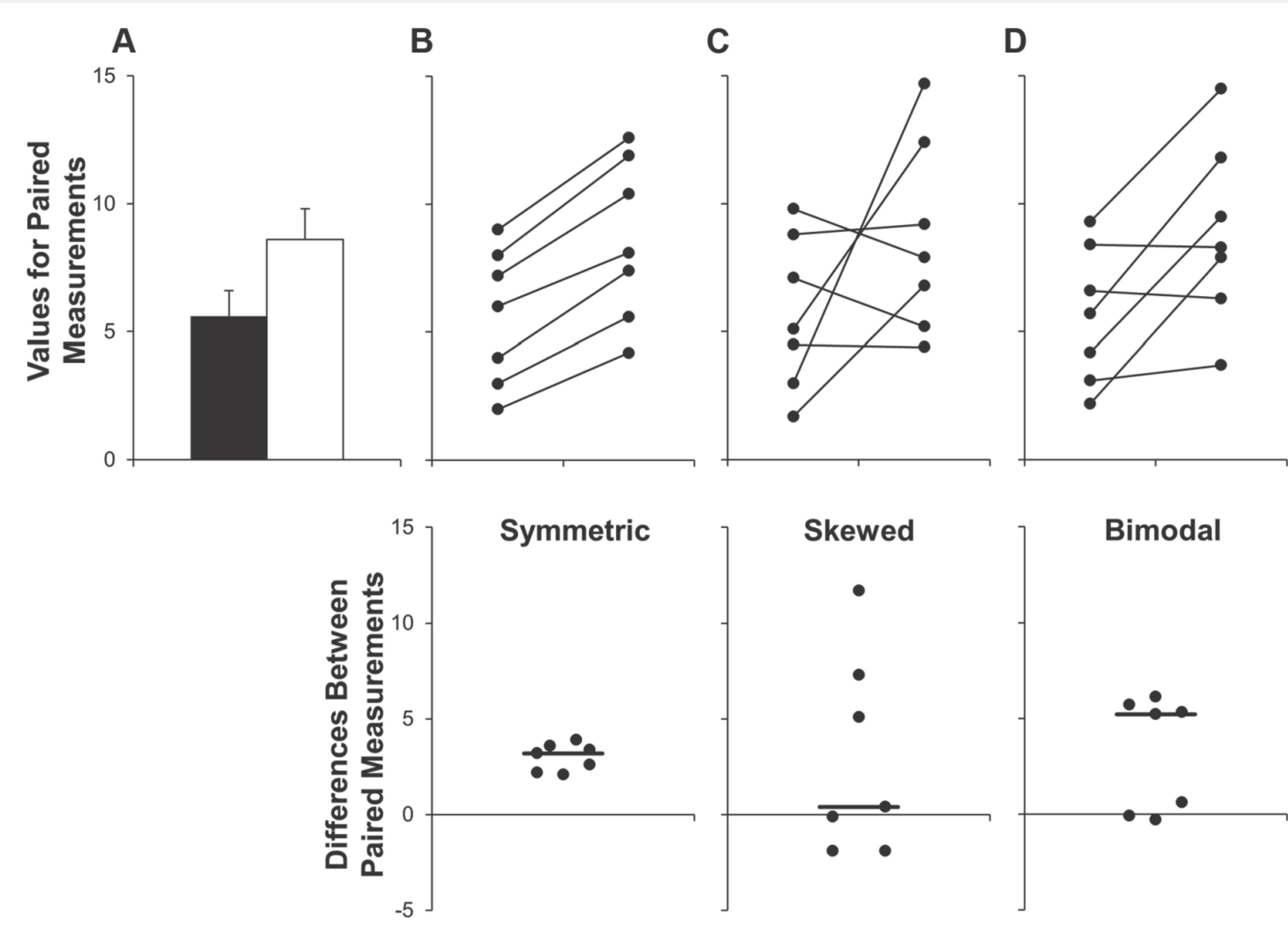
Weissgerber et al. (2015) PLoS Biology



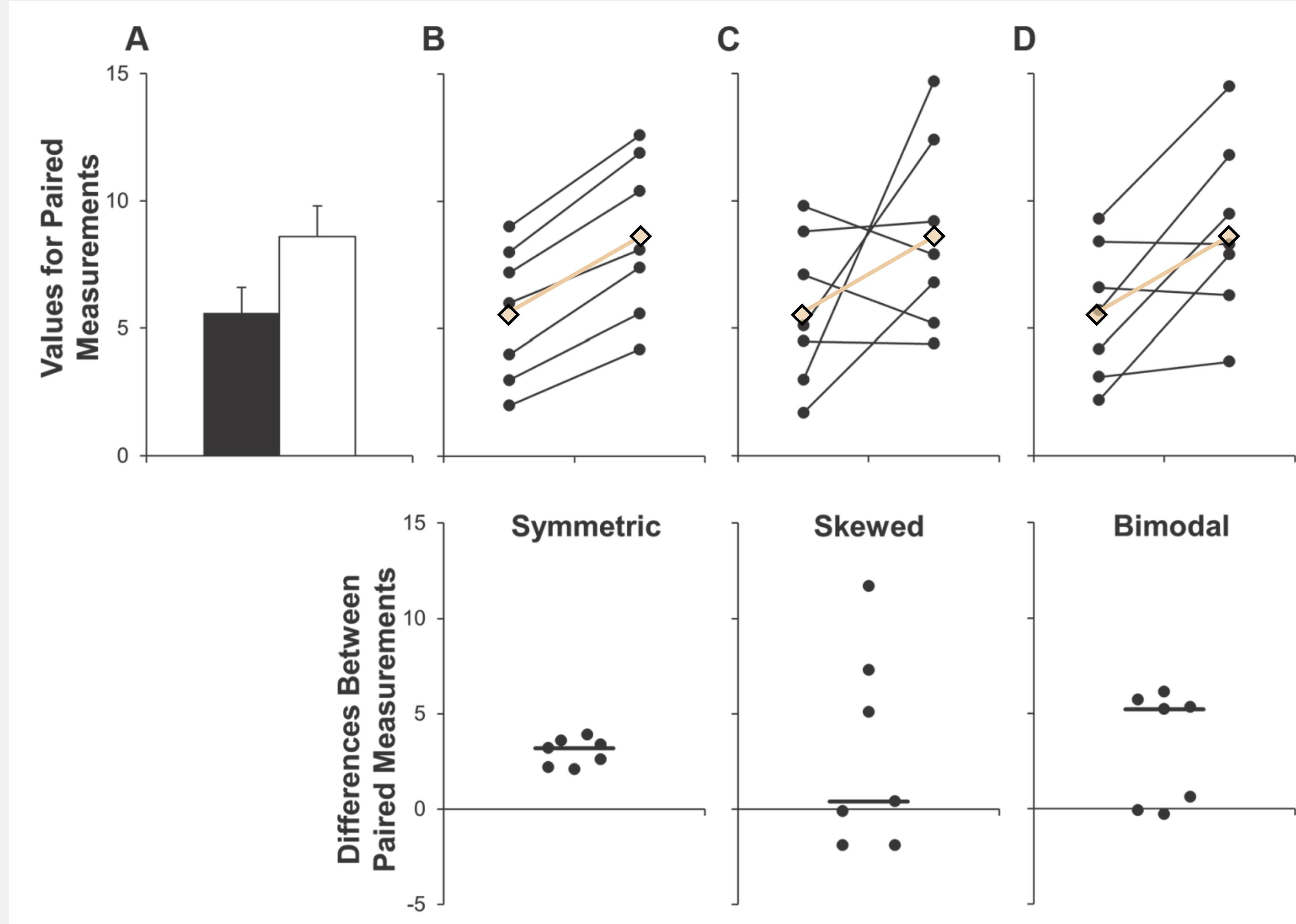
Modified from Weissgerber et al. (2015) PLoS Biology



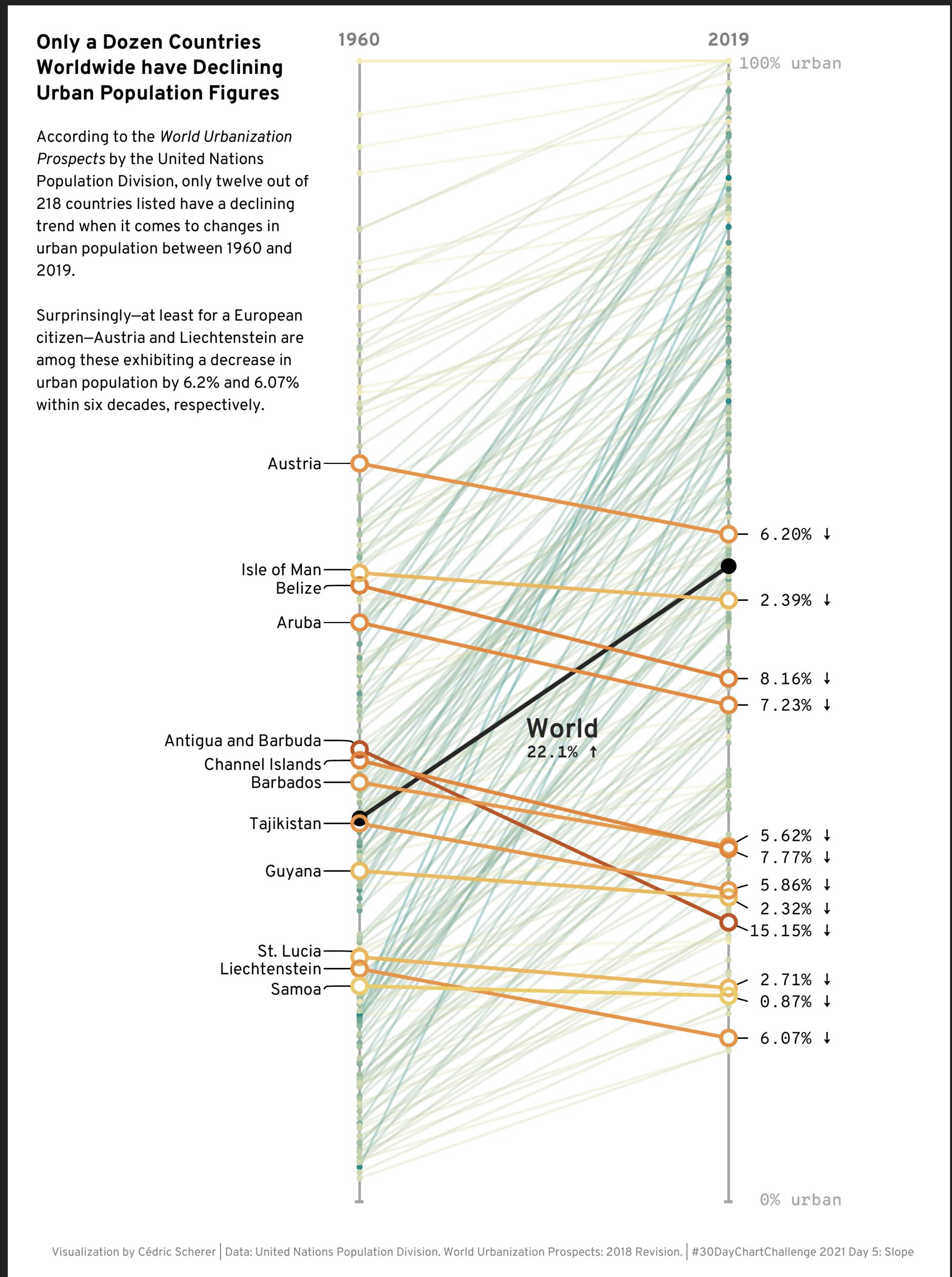
Modified from Weissgerber et al. (2015) PLoS Biology



Weissgerber et al. (2015) PLoS Biology



Weissgerber et al. (2015) PLoS Biology



# Not my cup of coffee...

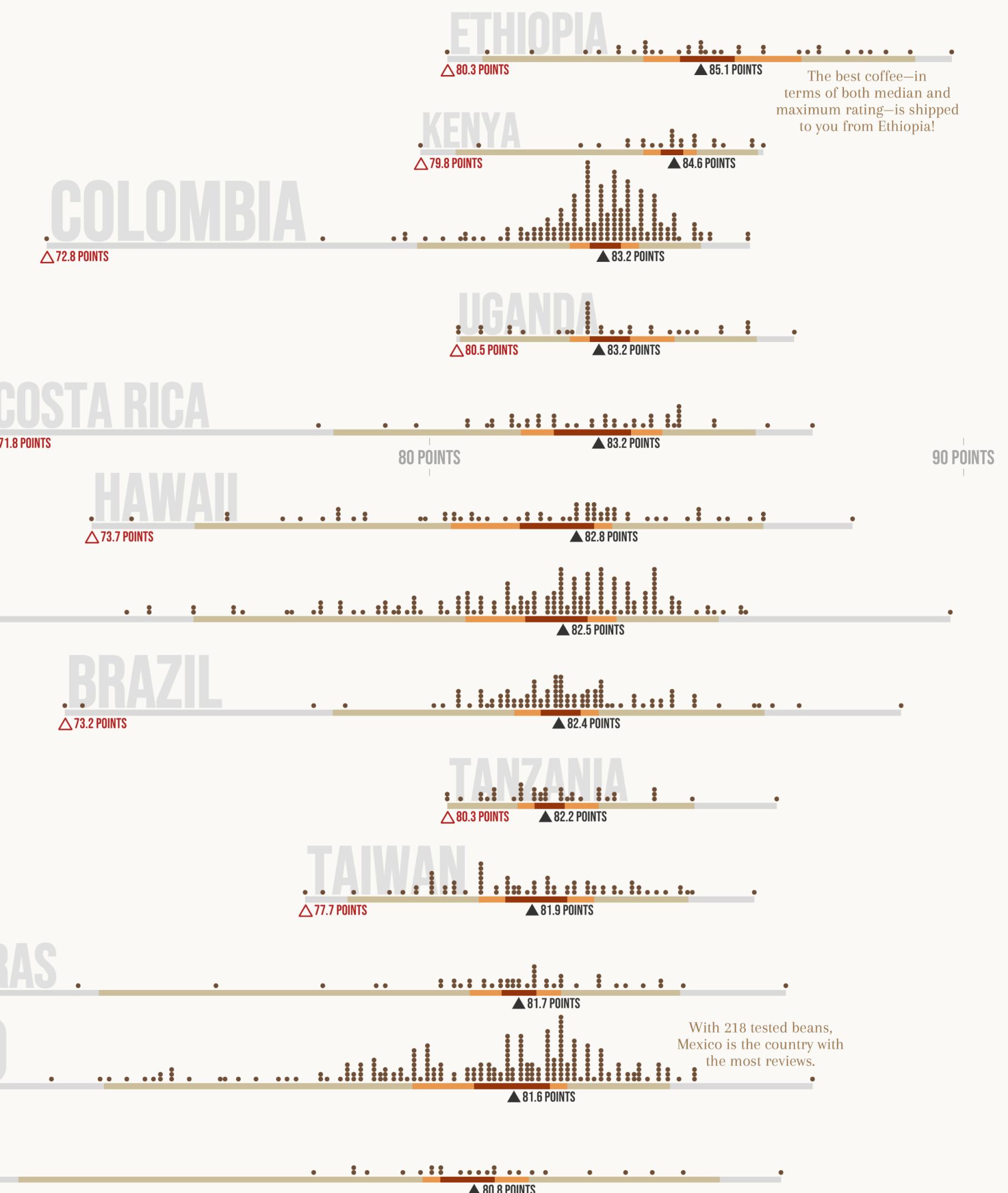
Each dot depicts one coffee bean rated by Coffee Quality Institute's trained reviewers. In addition, the multiple interval stripes show where 25%, 50%, 95%, and 100% of the beans fall along the rating gradient from 0 to 100 points. The rated coffee beans range from 59.8 points (Guatemala) to 89.9 (Ethiopia). Only countries of origin with 25 or more tested beans are shown. The red empty triangle marks the minimum rating, the black filled triangle indicates each country's median score.

Visualization by Cédric Scherer



The coffee bean with the lowest rating has its origin in Guatemala.

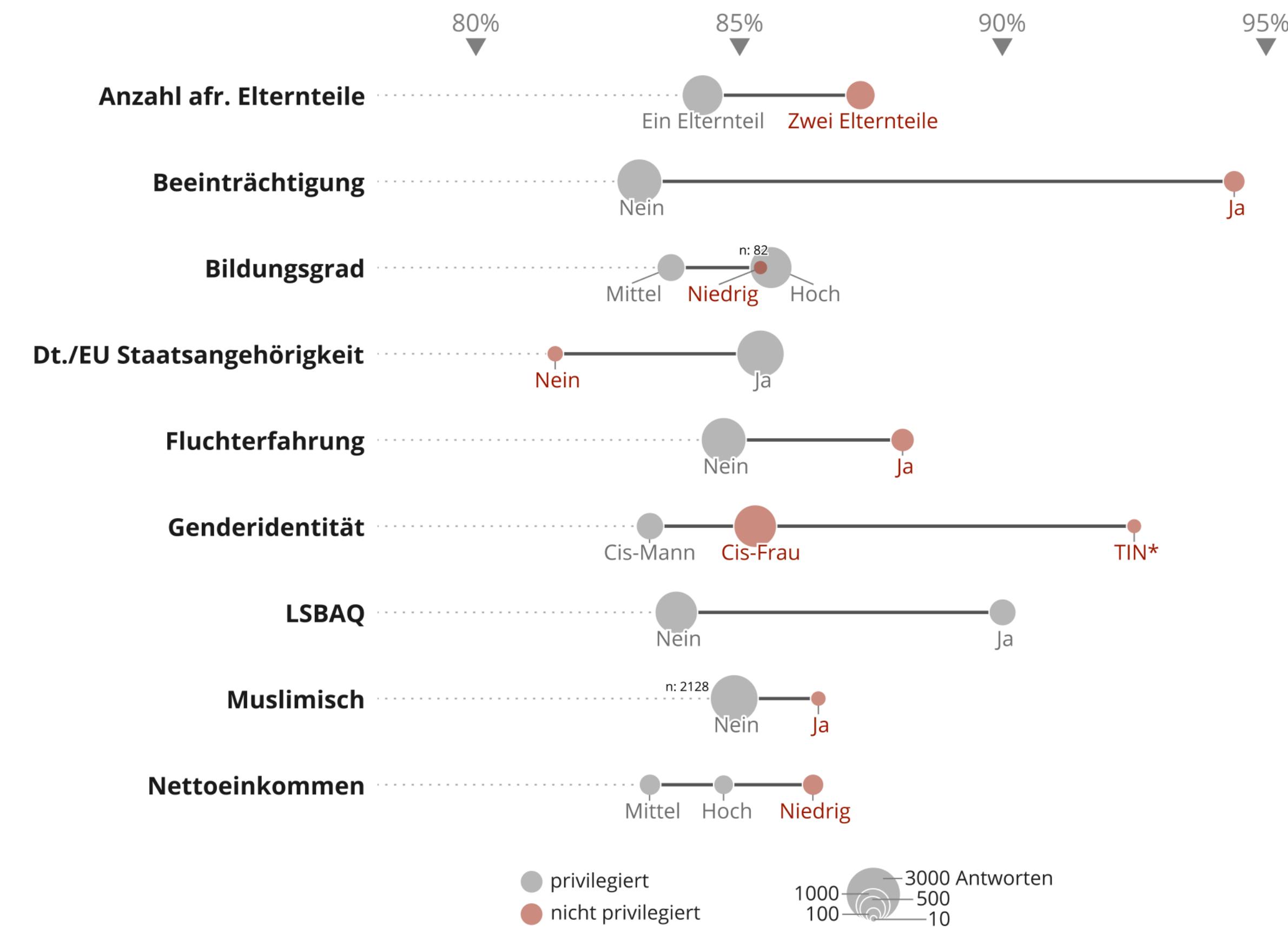
59.8 POINTS



*"Not my cup of coffee", #TidyTuesday Contribution*

Cédric Scherer // rstudio::conf // July 2022

## Häufigkeit von Diskriminierungserfahrungen entlang ausgewählter Vielfaltsdimensionen im Lebensbereich „Medien und Internet“



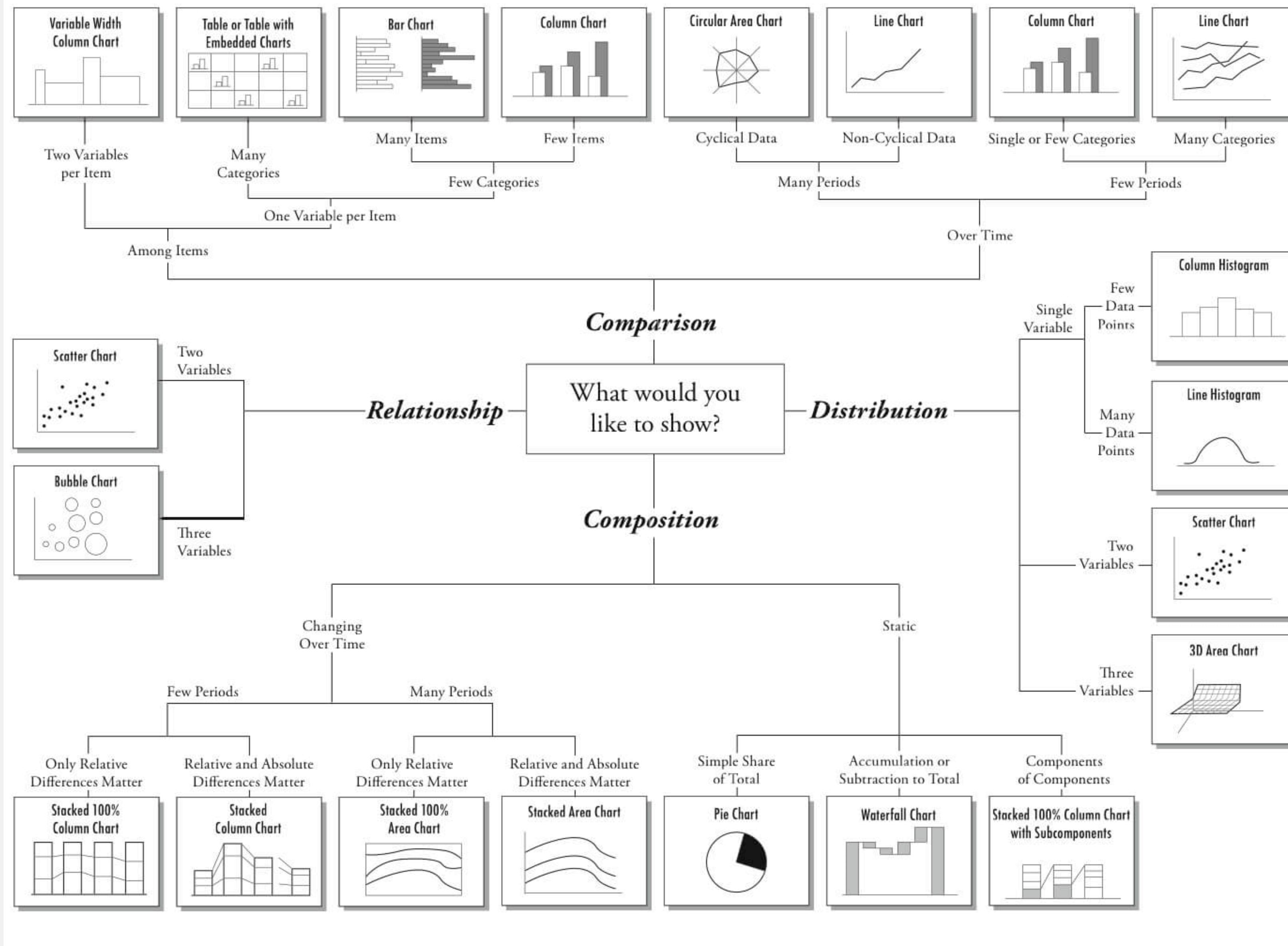
**Lesebeispiel:** LSBAQ-Befragte des Afrozensus geben im Vergleich mit heterosexuellen Afrozensus-Befragten häufiger an, im Lebensbereich „Medien und Internet“ in den letzten zwei Jahren Diskriminierung erlebt zu haben.

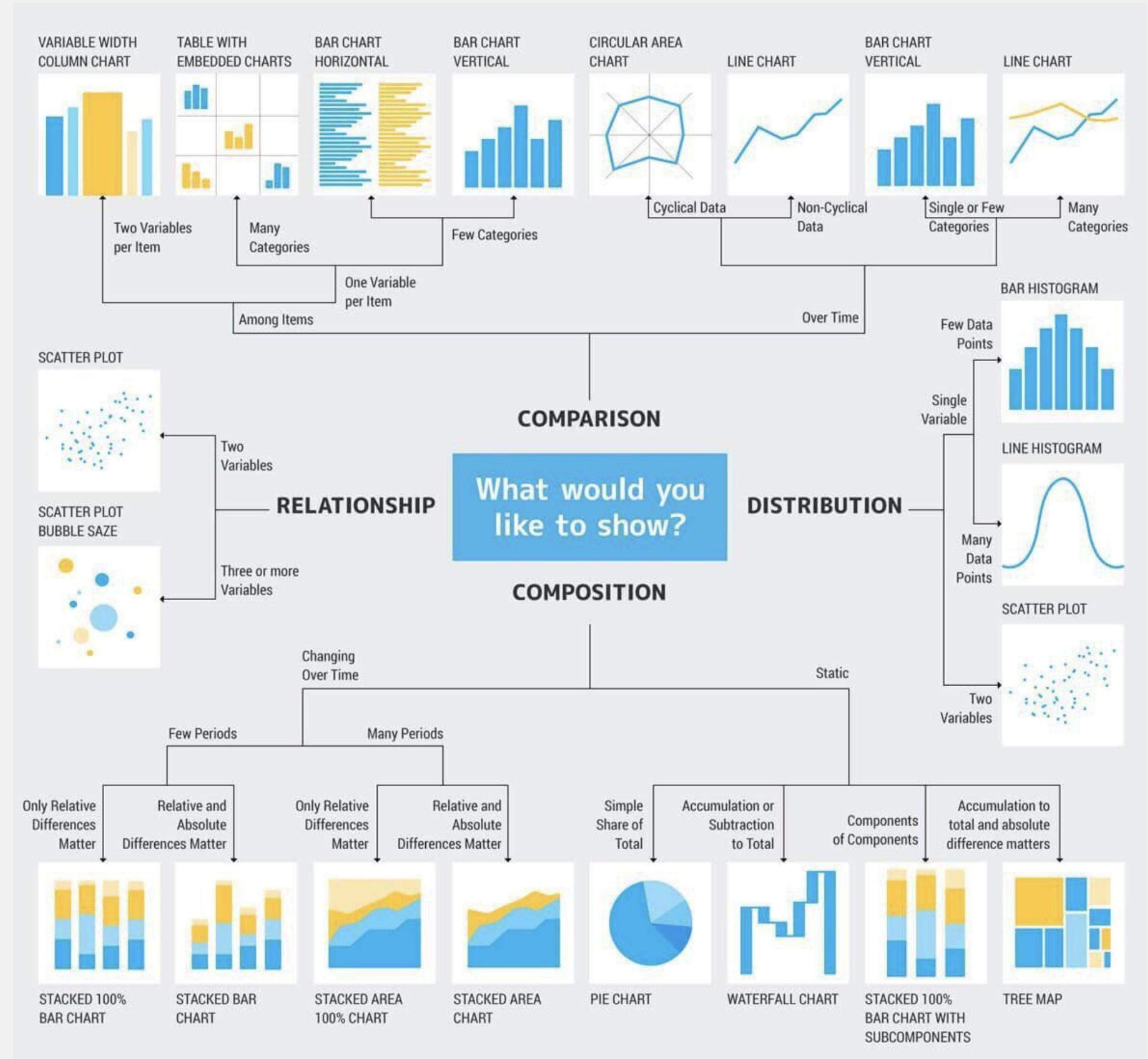
Quelle: Abb. 46 in Aikins, M A; Bremberger, T; Aikins, J K; Gyamerah, D; Yıldırım-Caliman, D (2021): Afrozensus 2020 | Datenteam: Reiber, L; Vivanco, J | Design: Scherer, C  
Lizenz: CC-BY-NC by EOTO & CFE | afrozensus.de

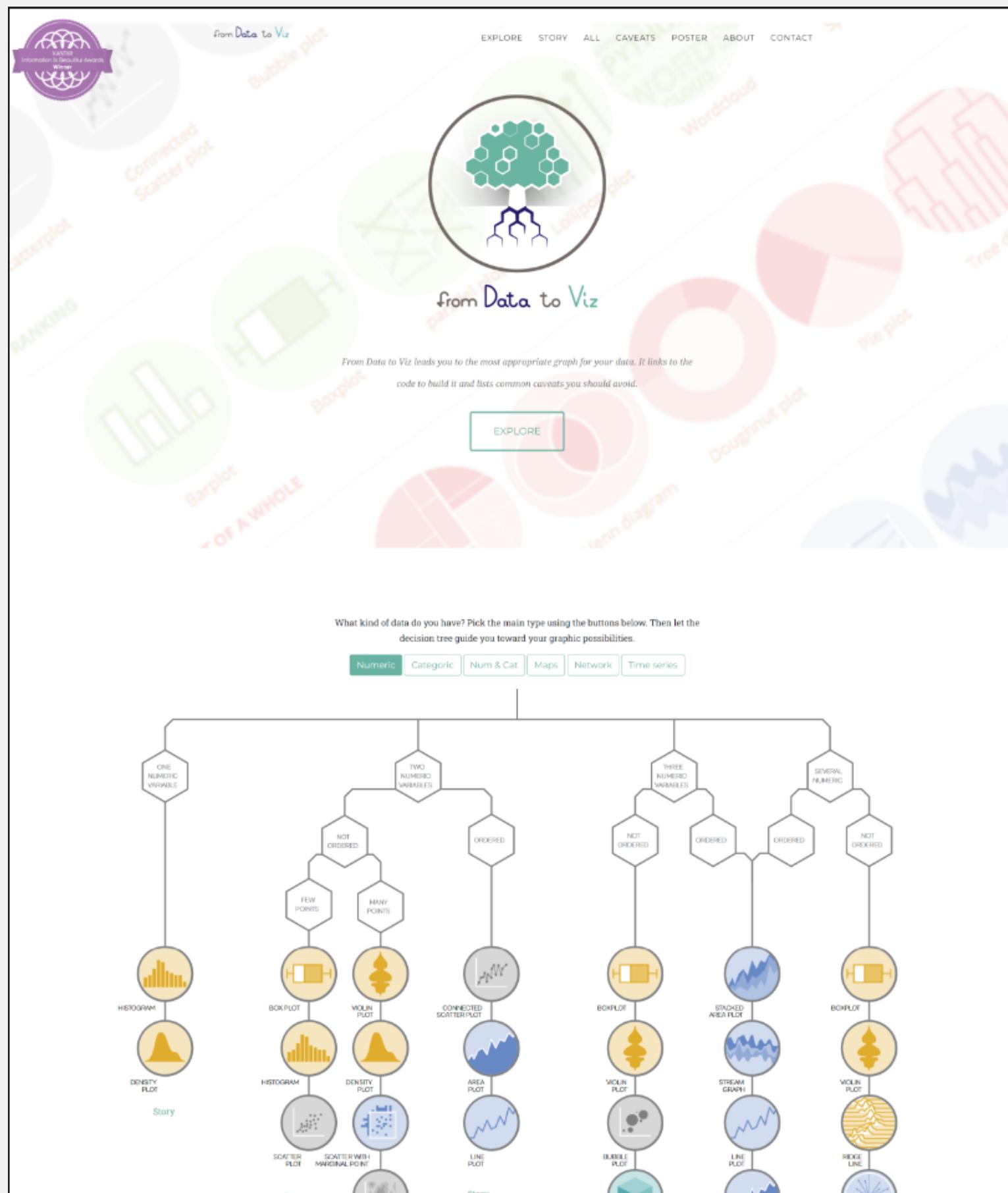
**Abb. 46 “Afrozensus 2020” by Citizens For Europe & EOTO e.V.**

# Chart Suggestions—A Thought-Starter

www.ExtremePresentation.com  
© 2009 A. Abela — a.v.abela@gmail.com







[data-to-viz.com](http://data-to-viz.com)



[datavizproject.com](http://datavizproject.com)



[visualizationuniverse.com](http://visualizationuniverse.com)



# from Data to Viz

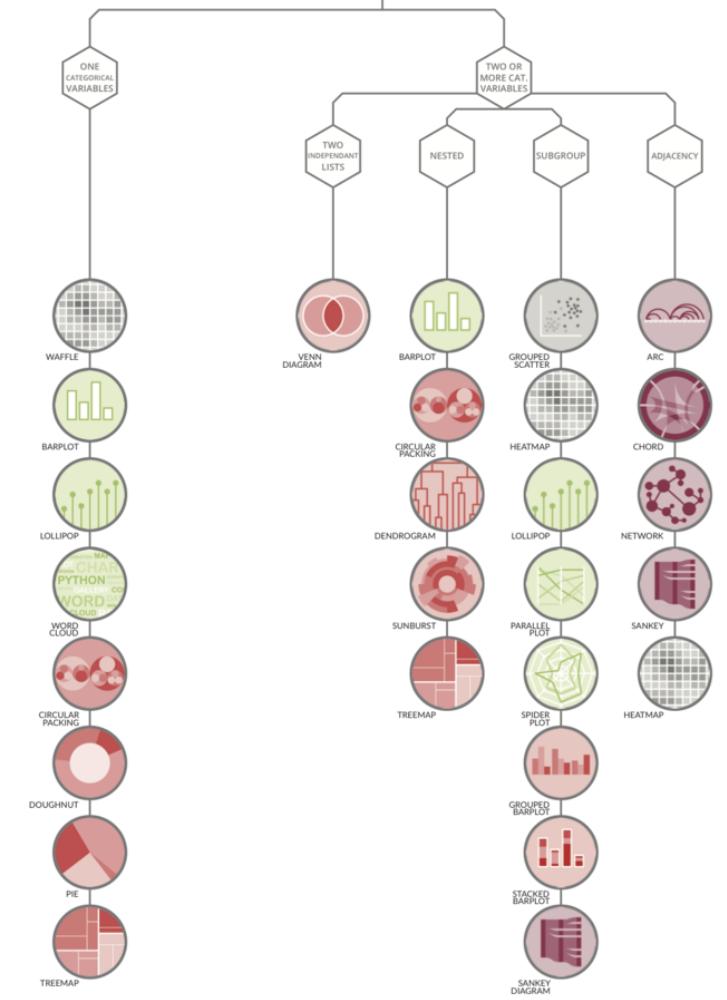
'From Data to Viz' is a classification of chart types based on input data format. It will help you find the perfect chart in three simple steps :

- 1 Identify what type of data you have.
- 2 Go to the corresponding decision tree and follow it down to a set of possible charts.
- 3 Choose the chart from the set that will suit your data and your needs best.

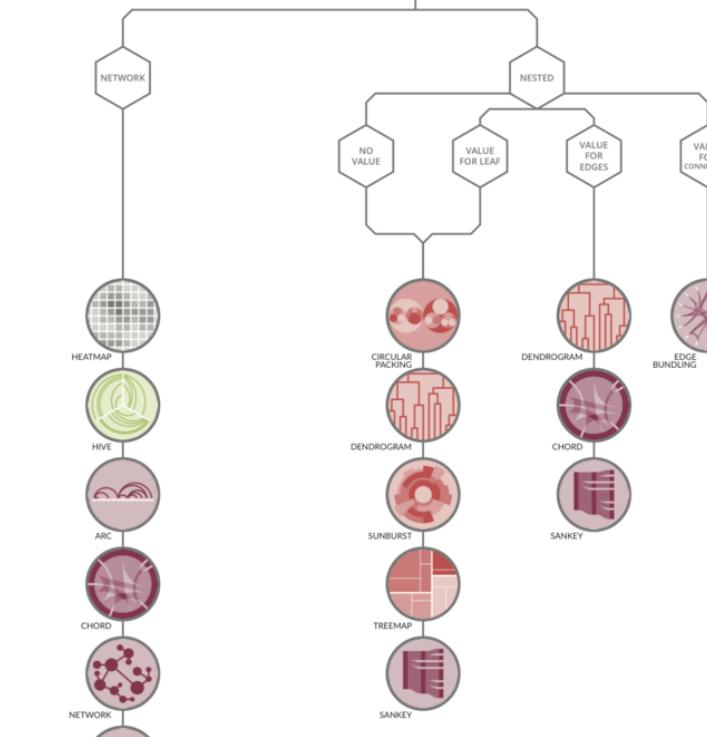
Dataviz is a world with endless possibilities and this project does not claim to be exhaustive. However it should provide you with a good starting point. For an interactive version and much more, visit:

[data-to-viz.com](http://data-to-viz.com)

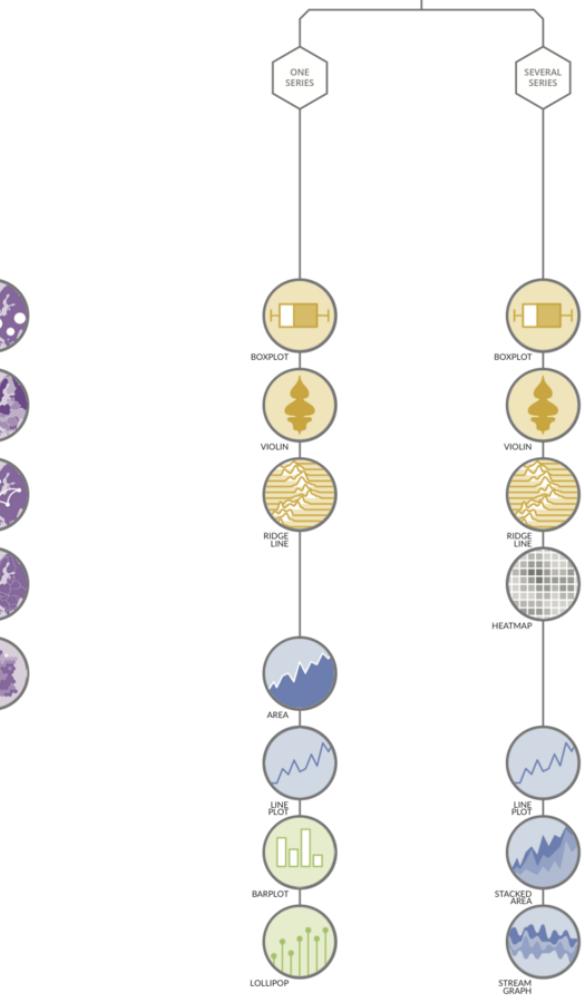
## CATEGORIC



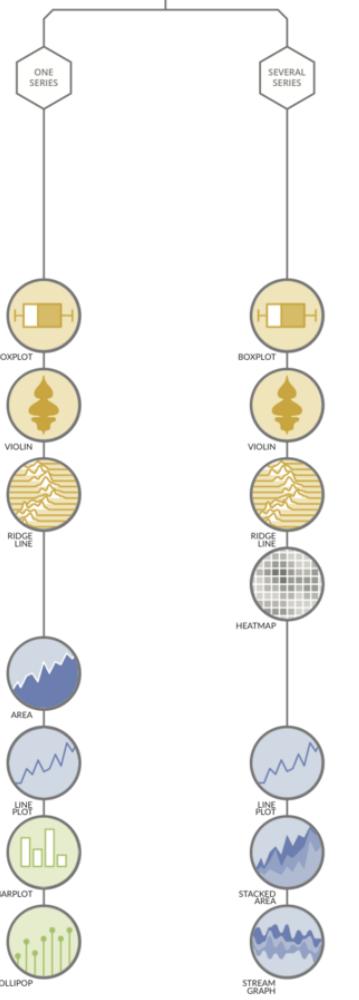
## RELATIONAL



## MAP



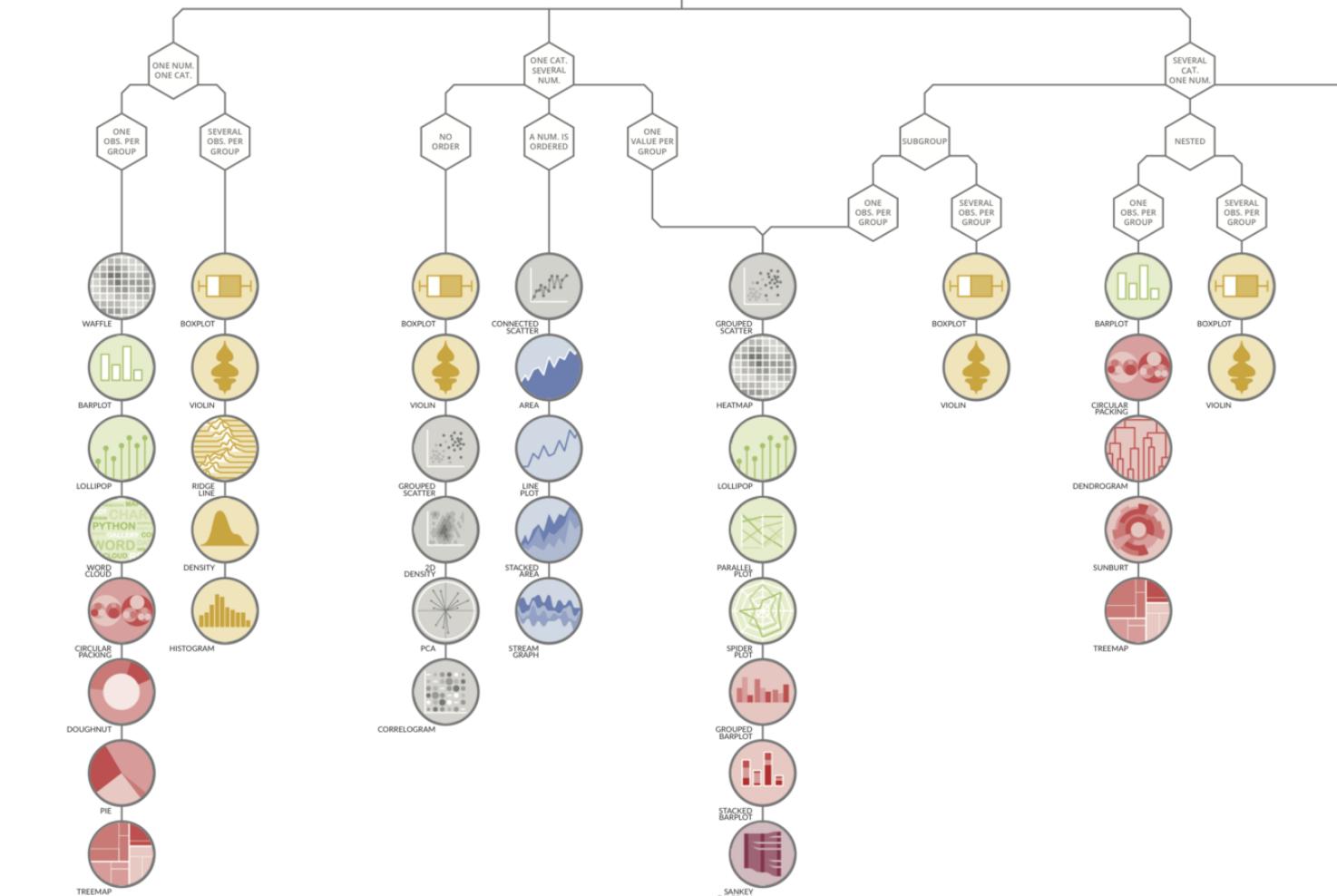
## TIME SERIES



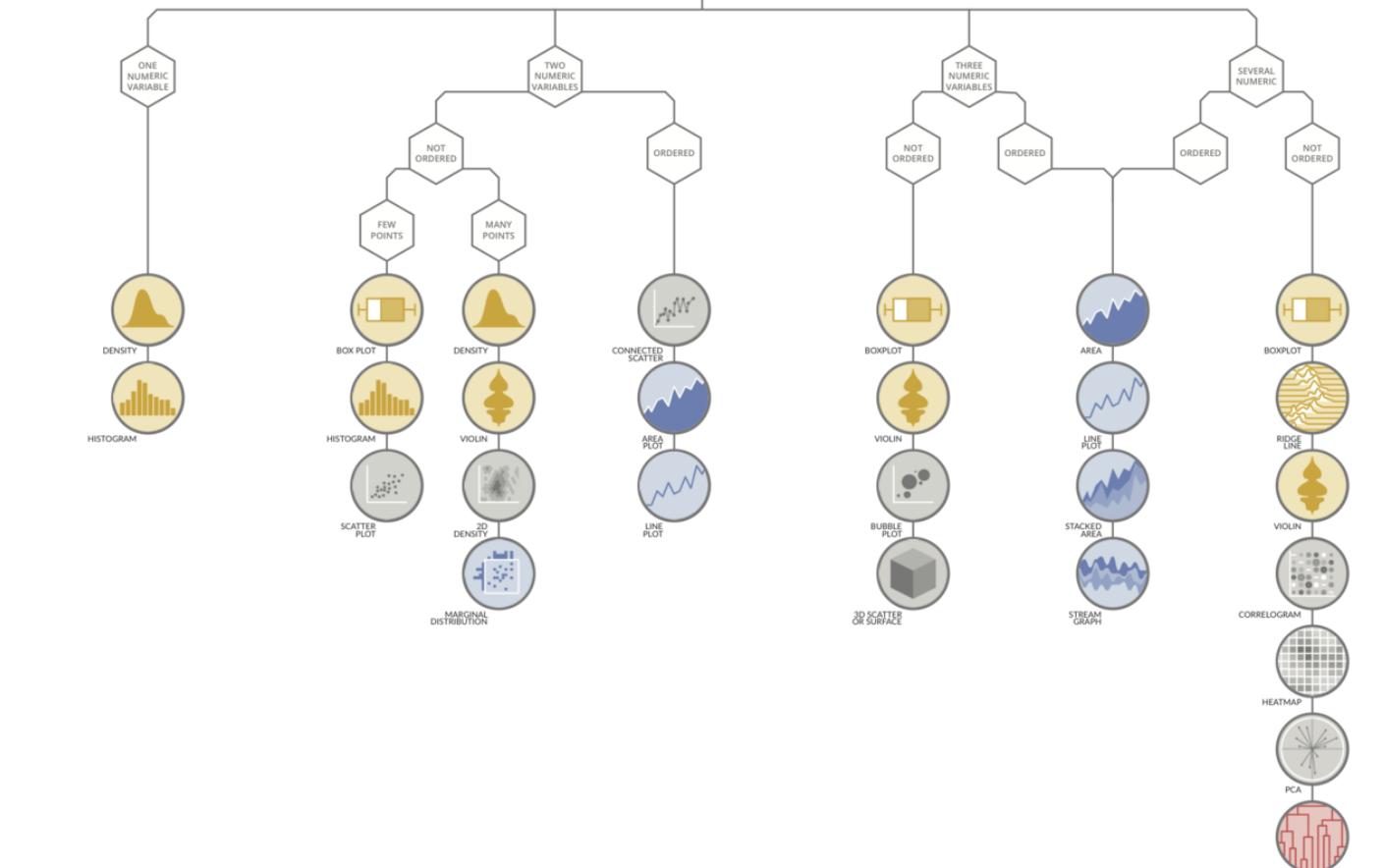
### WHAT DO YOU WANT TO SHOW ?

- |                   |             |
|-------------------|-------------|
| ● Distribution    | ● Evolution |
| ● Correlation     | ● Maps      |
| ● Ranking         | ● Flow      |
| ● Part of a whole |             |

## CATEGORIC AND NUMERIC

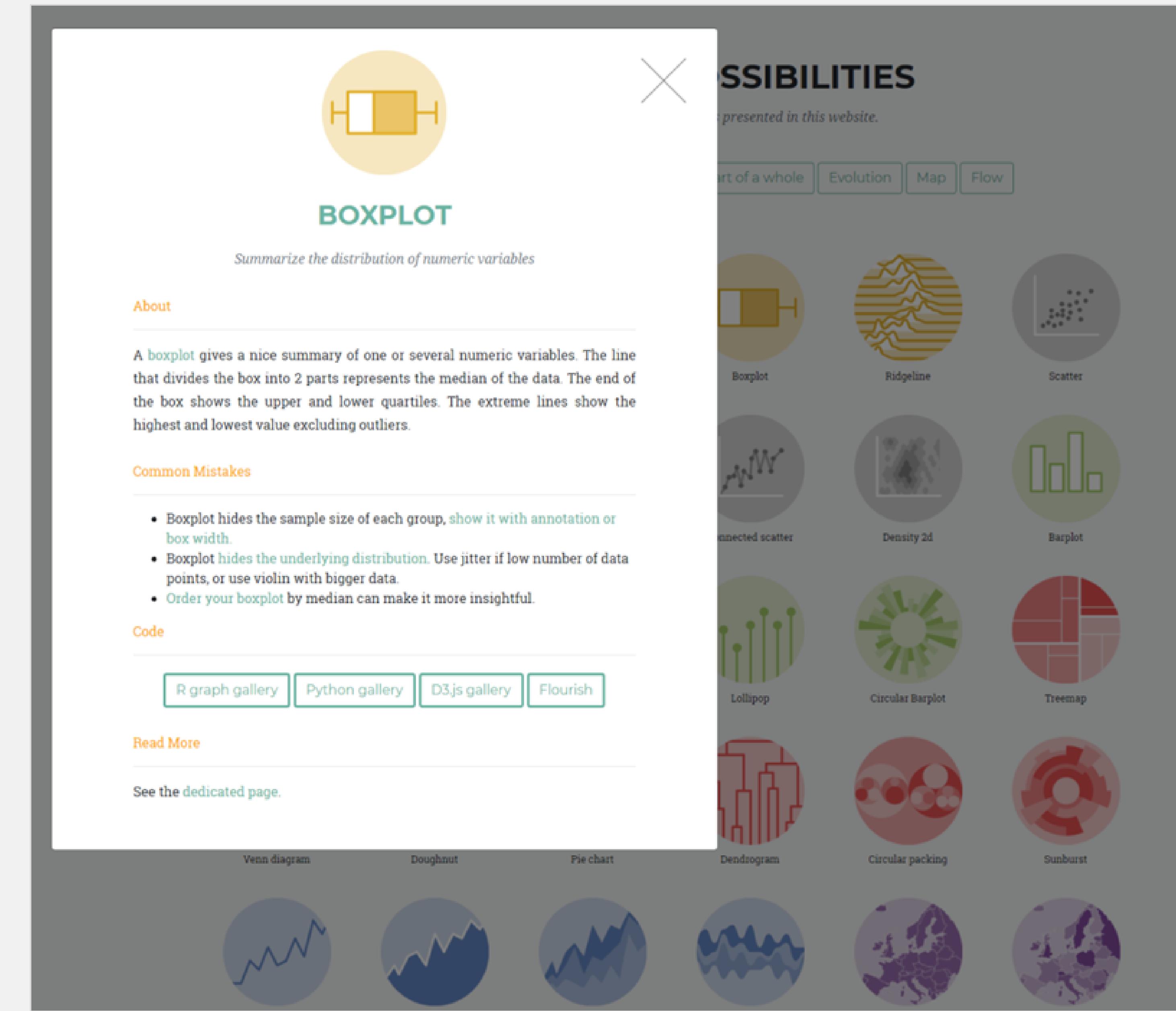


## NUMERIC



[data-to-viz.com](http://data-to-viz.com)

Cédric Scherer // rstudio::conf // July 2022



**POSSIBILITIES**

presented in this website.

[Part of a whole](#) [Evolution](#) [Map](#) [Flow](#)

Boxplot Ridgeline Scatter

Connected scatter Density 2d Barplot

Lollipop Circular Barplot Treemap

Dendrogram Circular packing Sunburst

Venn diagram Doughnut Pie chart

**BOXPLOT**

*Summarize the distribution of numeric variables*

[About](#)

A boxplot gives a nice summary of one or several numeric variables. The line that divides the box into 2 parts represents the median of the data. The end of the box shows the upper and lower quartiles. The extreme lines show the highest and lowest value excluding outliers.

[Common Mistakes](#)

- Boxplot hides the sample size of each group, [show it with annotation or box width](#).
- Boxplot hides the underlying distribution. Use jitter if low number of data points, or use violin with bigger data.
- [Order your boxplot by median](#) can make it more insightful.

[Code](#)

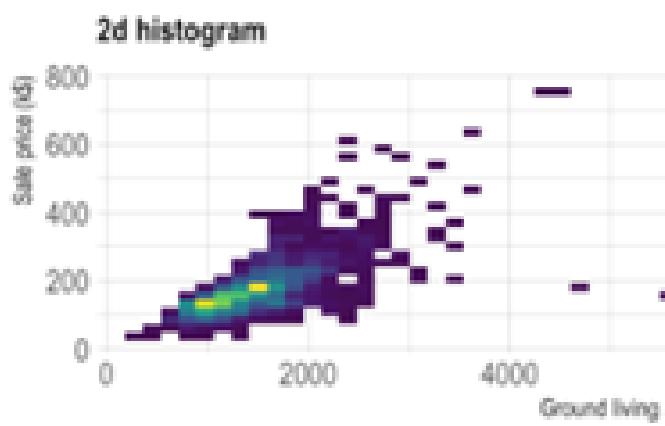
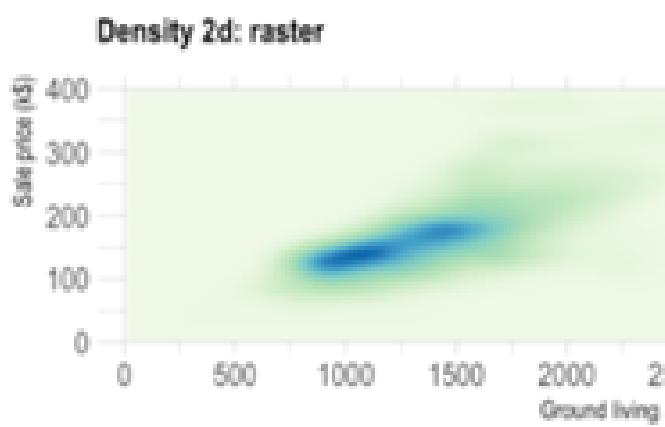
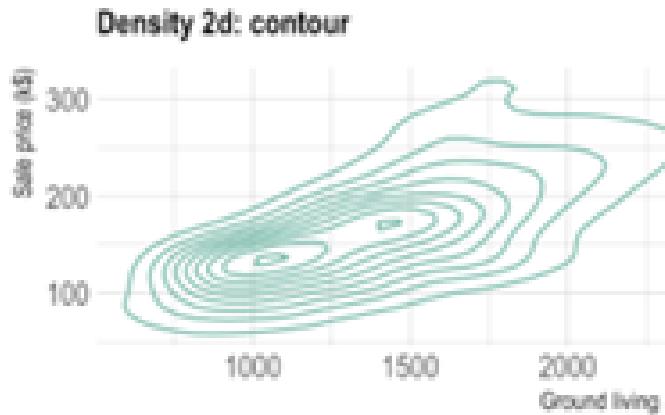
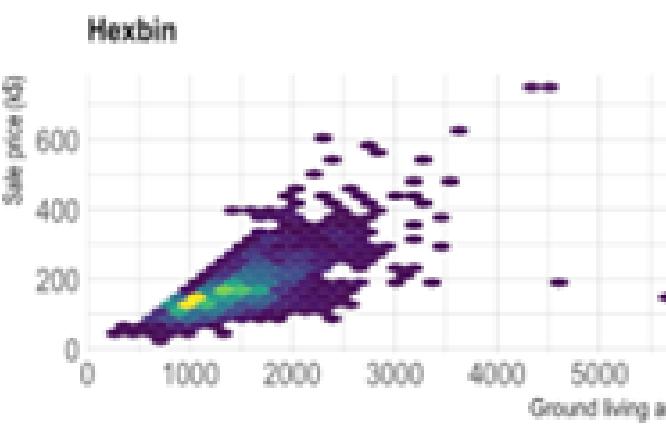
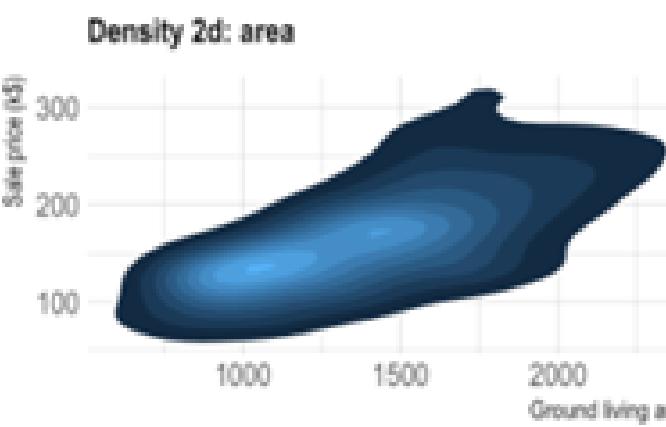
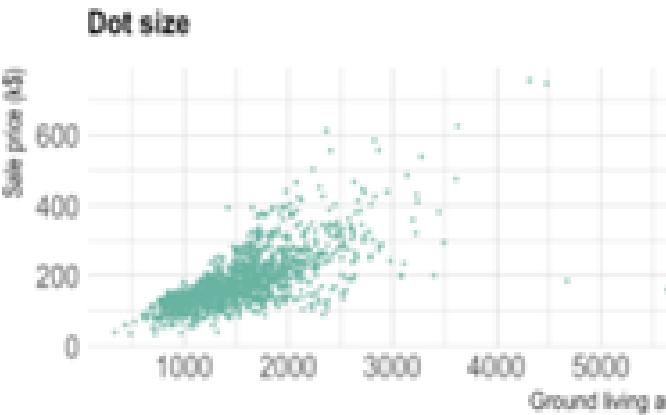
[R graph gallery](#) [Python gallery](#) [D3.js gallery](#) [Flourish](#)

[Read More](#)

See the [dedicated page](#).

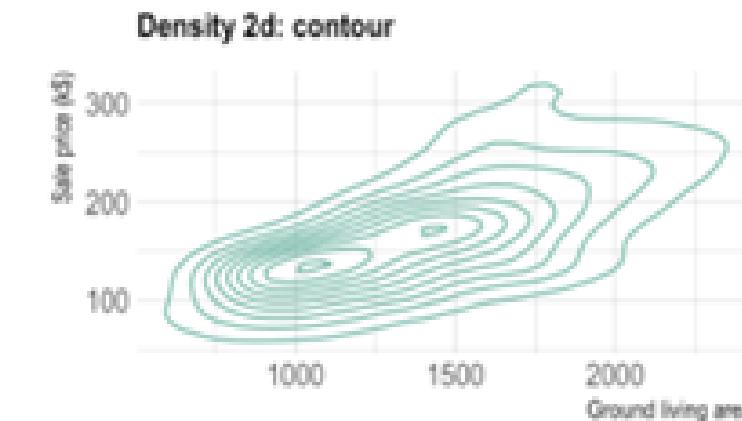
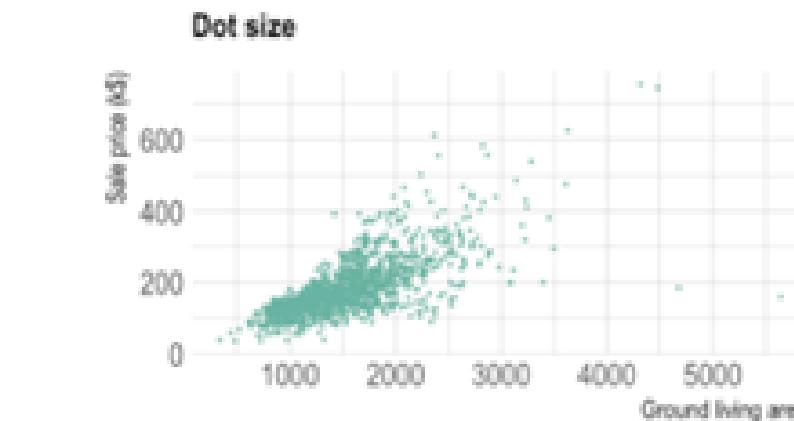
# Overplotting

The most common pitfall with scatterplot is overplotting: when the sample size gets big, dots are plotted on top of each other what makes the chart unreadable. There are several work around to avoid this issue as describe in this [specific post](#). Here is a summary of the different offered techniques:



CODE

```
# code for all graphics:  
p <- data %>%  
  ggplot( aes(x=grLivArea, y=salePrice/1000)) +  
    theme_ipsum() +  
    theme(  
      plot.title = element_text(size=12)  
    ) +  
    ylab("Sale price (k$)") +  
    xlab("Ground living area")  
  
# Reduce dot size  
p1 <- p + geom_point(color="#69b3a2", alpha=0.8, size=0.2) + ggtitle("Dot size")  
  
# Use density estimate  
p2 <- p + geom_density2d(color="#69b3a2") + ggtitle("Density 2d: contour")  
  
# Use density estimate (area)  
p3 <- p + stat_density_2d(aes(fill = ..level..), geom = "polygon") + ggtitle("Density 2d: area") + theme(legend.position="none")  
  
# With raster  
p4 <- p +  
  stat_density_2d(aes(fill = ..density..), geom = "raster", contour = FALSE) +  
  scale_fill_distiller(palette=4, direction=1) +  
  scale_x_continuous(expand = c(0, 0)) +  
  scale_y_continuous(expand = c(0, 0)) +  
  theme(  
    legend.position="none"  
  ) +  
  ggtitle("Density 2d: raster") +  
  xlim(0,2500) +  
  ylim(0,400)  
  
# Hexbin  
p5 <- p + geom_hex() +  
  scale_fill_viridis() +  
  theme(legend.position="none") +  
  ggtitle("Hexbin")  
  
# 2d histogram  
p6 <- p + geom_bin2d() +  
  scale_fill_viridis() +  
  theme(legend.position="none") +  
  ggtitle("2d histogram")  
  
p1 + p2 + p3 + p4 + p5 + p6 + plot_layout(ncol = 2)
```



## Going further

You can learn more about each type of graphic presented in this story in the dedicated

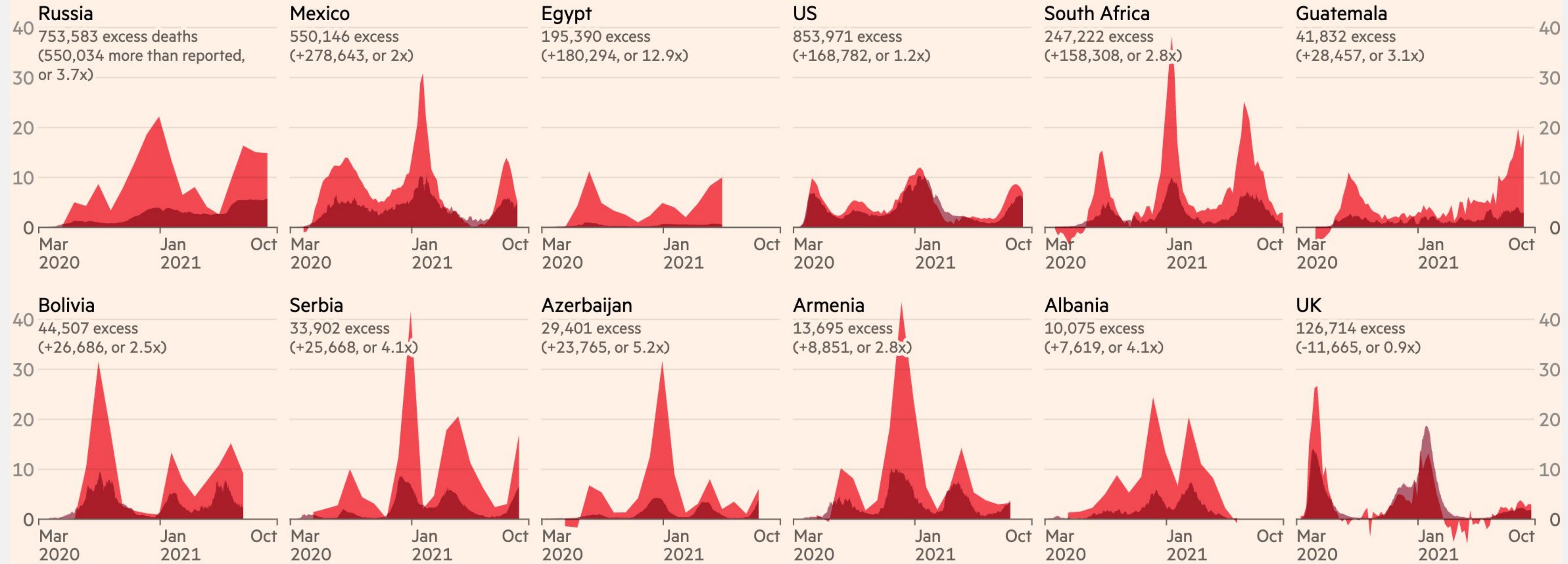
[data-to-viz.com](http://data-to-viz.com)

Cédric Scherer // rstudio::conf // July 2022

# The Power of Small Multiples

Russia has recorded more than 753,000 excess deaths during the pandemic, almost four times the official Covid death toll provided by state agencies

Daily **excess deaths** vs **reported deaths**, per million people

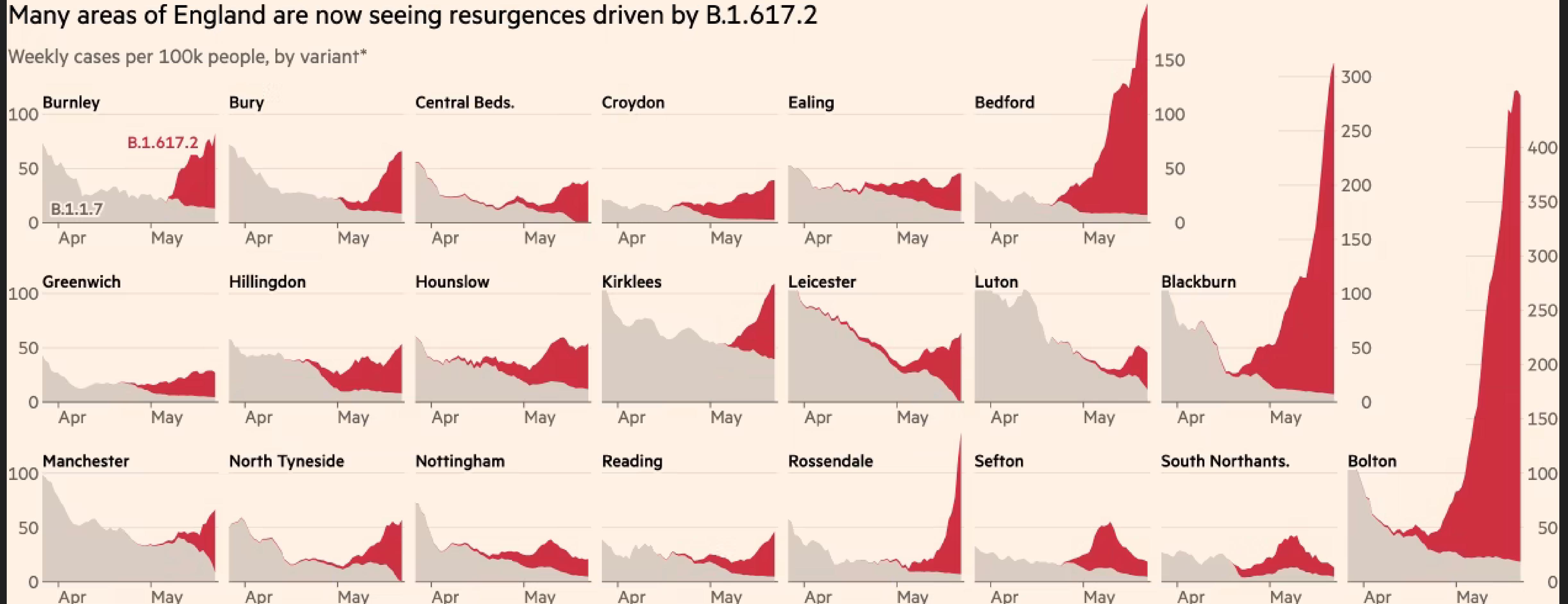


Source: Johns Hopkins CSSE; FT analysis of national mortality data and Karlinsky & Kobak's World Mortality Dataset  
© FT

**"Russia's excess mortality soars since start of Covid pandemic"** by John Burn-Murdoch (Financial Times)

## Many areas of England are now seeing resurgences driven by B.1.617.2

Weekly cases per 100k people, by variant\*



\*Based on applying proportions of sequenced samples to total cases

Sources: FT analysis of data from the Sanger Institute and UK government Covid-19 dashboard

© FT

***"UK virus cases hit 6-week high but vaccines diminish threat"*** by John Burn-Murdoch (Financial Times)

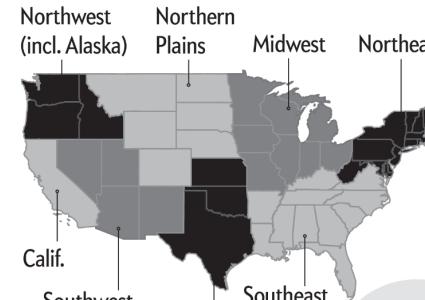
## GRAPHIC SCIENCE

Text by Clara Moskowitz | Graphic by Cédric Scherer and Georgios Karamanis

# Escalating Drought

Climate change is intensifying periods of extreme dryness, particularly in the U.S. West

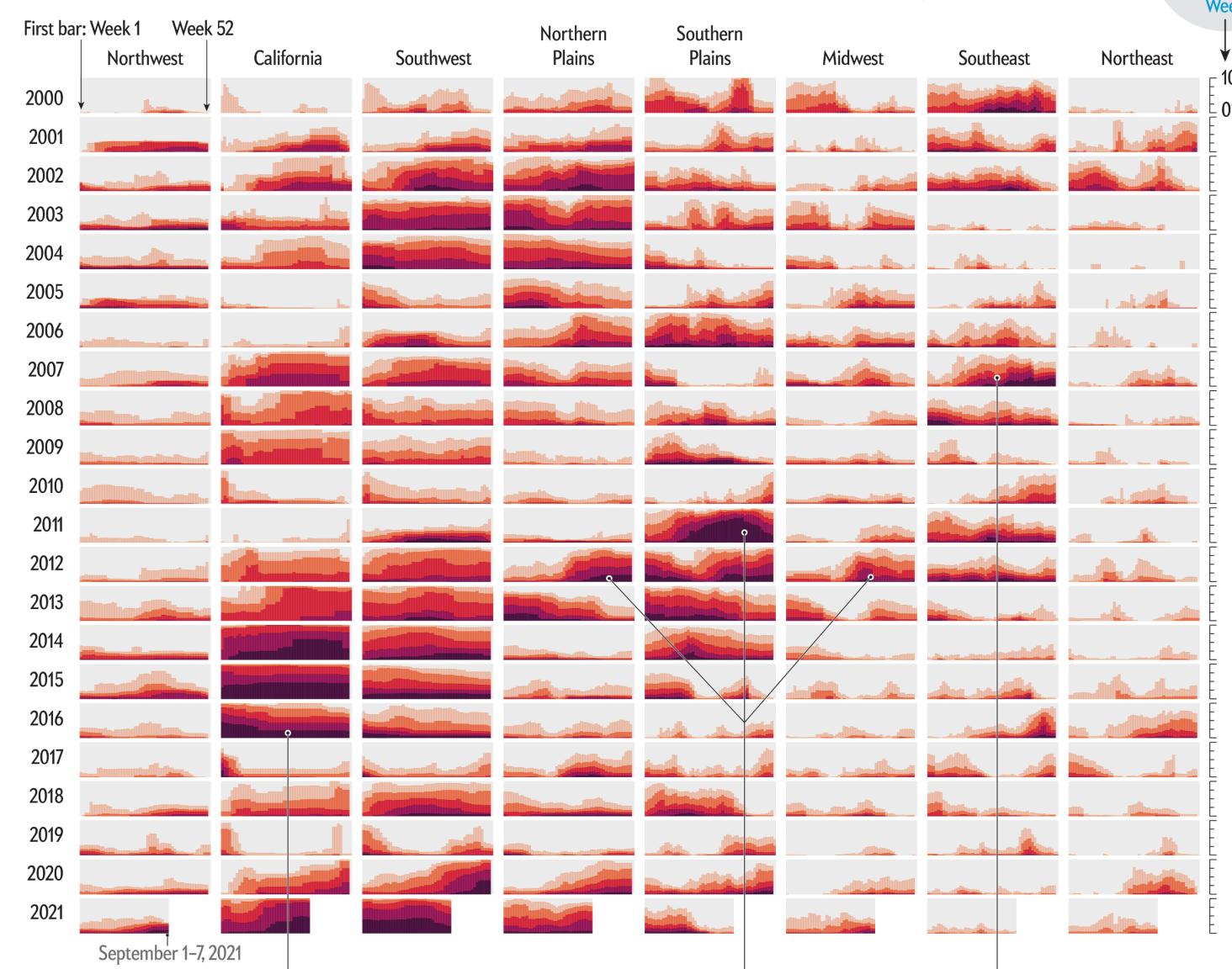
For more than 20 years the National Drought Mitigation Center (NDMC) has been monitoring dozens of indices of drought around the country, including satellite measurements of evaporation and color in vegetation, soil-moisture sensors, rainfall estimates, and river and streamflow levels. Although the agency's weekly assessments have identified periods of exceptional drought before, lately dryness has been ramping up. "The changing climate is definitely contributing to more natural disasters, drought being one of them," says Brian Fuchs, a climatologist who oversees the weekly report at the NDMC. "We're seeing more frequent and high-intensity episodes. This year some of these areas in the West have been in drought more than they have been without drought."



Percent of Region That Experienced Drought Each Week

### Drought Extent and Intensity by Region over Time

Category: Abnormally Dry Moderate Drought Severe Drought Extreme Drought Exceptional Drought

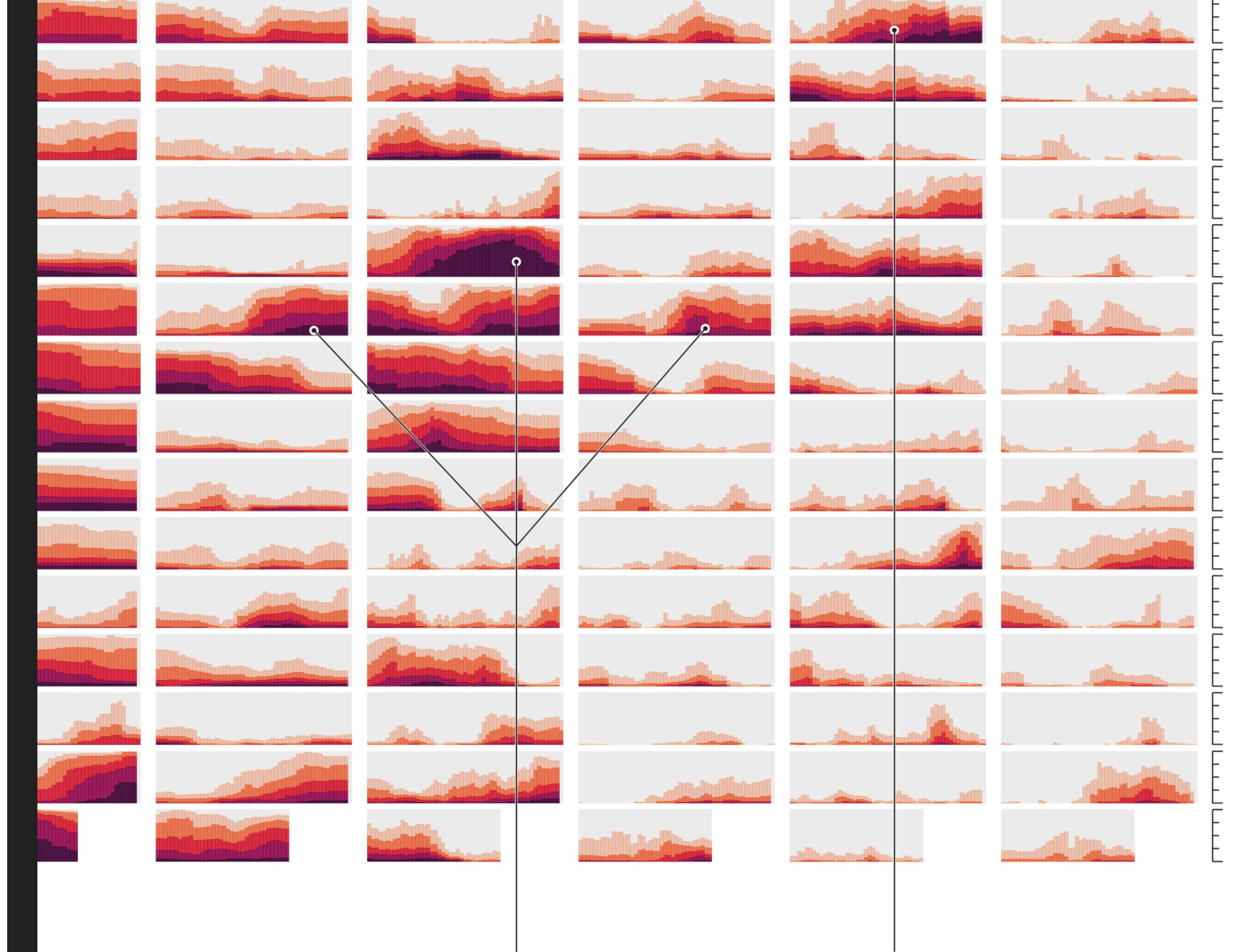


California experienced its hottest drought in recorded history from 2012 to 2016. A warming climate makes the atmosphere thirstier, which increases evaporation and boosts drought.

A drought that originated in the Southern Plains in 2011 eventually spread to the Midwest and Northern Plains when the moisture coming in from the Gulf of Mexico was absorbed by the parched South before it could reach the North.

The Southeast's driest year to date was 2007, when only 31.85 inches of rain fell in Atlanta, 62 percent of its average yearly rainfall.

Source: U.S. Drought Monitor; jointly produced by the National Drought Mitigation Center at the University of Nebraska-Lincoln, U.S. Department of Agriculture, and National Oceanic and Atmospheric Administration (data)



74 Scientific American, November 2021

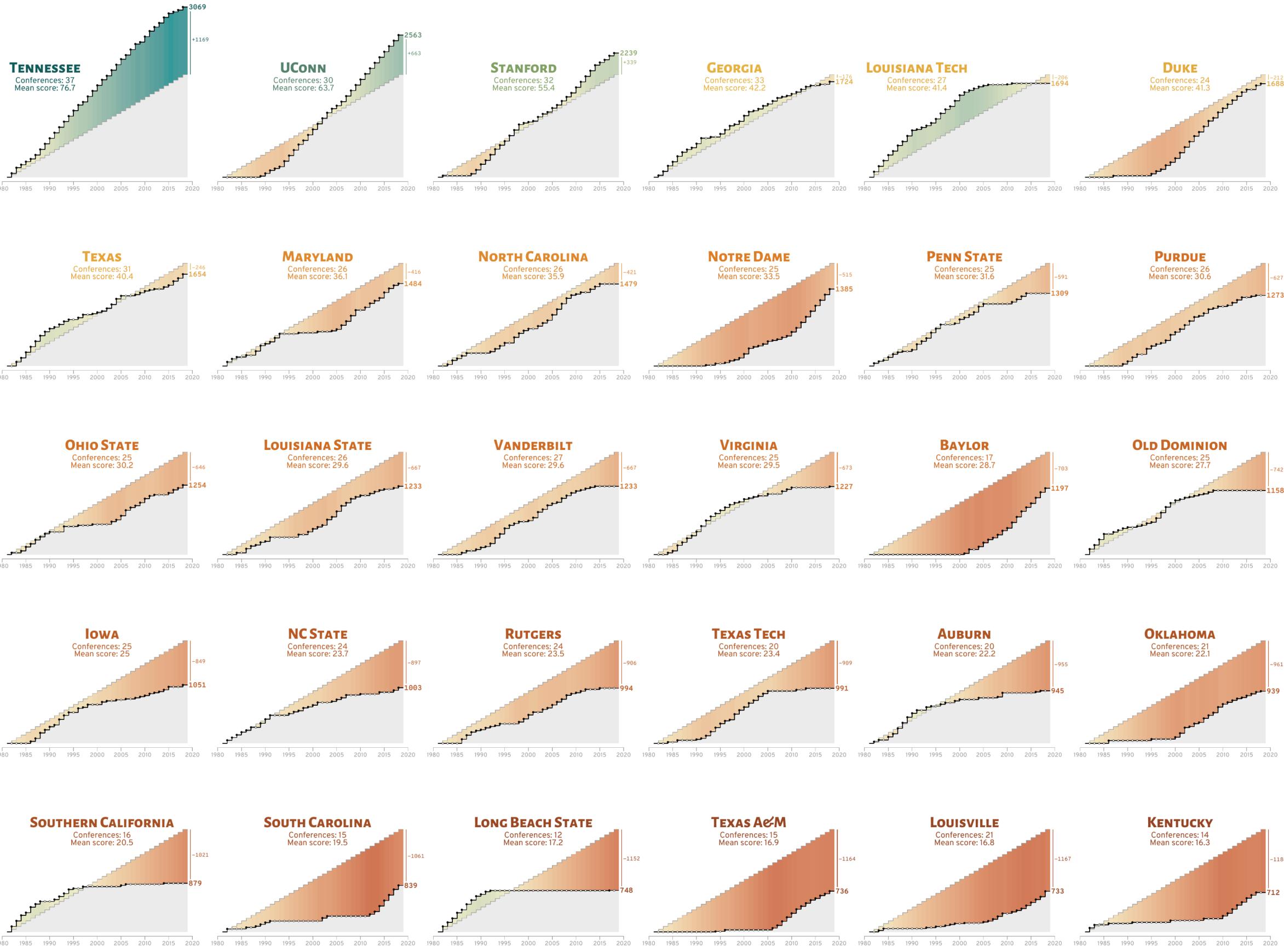
**"Escalating Drought"**, together with Georgios Karamanis for Scientific American, Issue Nov 2021

Cédric Scherer // rstudio::conf // July 2022

## — THE RISE & FALL OF WOMEN'S COLLEGE BASKETBALL DYNASTIES —

A number of teams that were the titans of the early NCAA women's basketball tournament have struggled in recent decades. And in their place, a new ruling class of schools has emerged to become the defining programs of the modern age. FiveThirtyEight estimated the team strength over time based on NCAA Tournament seeds as a proxy in the absence of game-level data. To measure this, FiveThirtyEight awarded "seed points" in proportion to a given seed number's expected wins in the tournament, calibrated to a 100-point scale where the No. 1 seed gets 100 points, No. 2 gets 70 points, and so forth.

The visualization shows the cumulative sum of awarded seed points on a 100-point scale from the very first women's NCAA basketball tournaments in 1982 until 2018 in comparison to a hypothetical team that participated in all of the 37 conferences and gained half of the points each time (grey line). The curves highlight the fall of yesterday's women's basketball powerhouses such as **Louisiana Tech**, **Long Beach State**, **Southern California**, and **Old Dominion** that have been very good throughout the history of the women's tournament but have experienced big drop-offs in seed points over the last years. At the same time, schools such as **UConn**, **Stanford**, **Notre Dame**, **Baylor**, and **Duke** started slow but picked up steam into the present day. Some teams, such as **Tennessee**, have been relatively consistent throughout the NCAA era gathering always more seed points than an average team. Shown are the top 30 college teams that participated in at least ten conferences between 1982 and 2018, sorted by the cumulative sum of seed points.



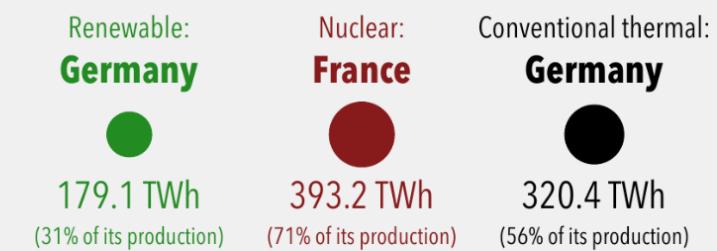
VISUALIZATION BY CÉDRIC SCHERER · DATA BY FIVETHIRTYEIGHT

*"The Rise and Fall of Women's College Basketball Dynasties"*, #TidyTuesday Contribution

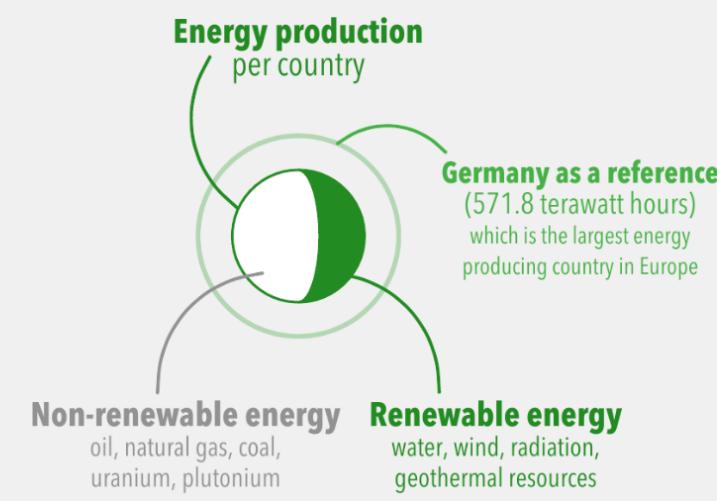
Cédric Scherer // rstudio::conf // July 2022

## How European countries generated electricity in 2018

Germany is the largest energy producing country in Europe. It generates the most renewable and conventional thermal energy, representing 31% and 56% of its overall production respectively. France is the second largest energy European producer and by far the largest nuclear energy provider: 71% of its production is based on nuclear fission to generate heat.

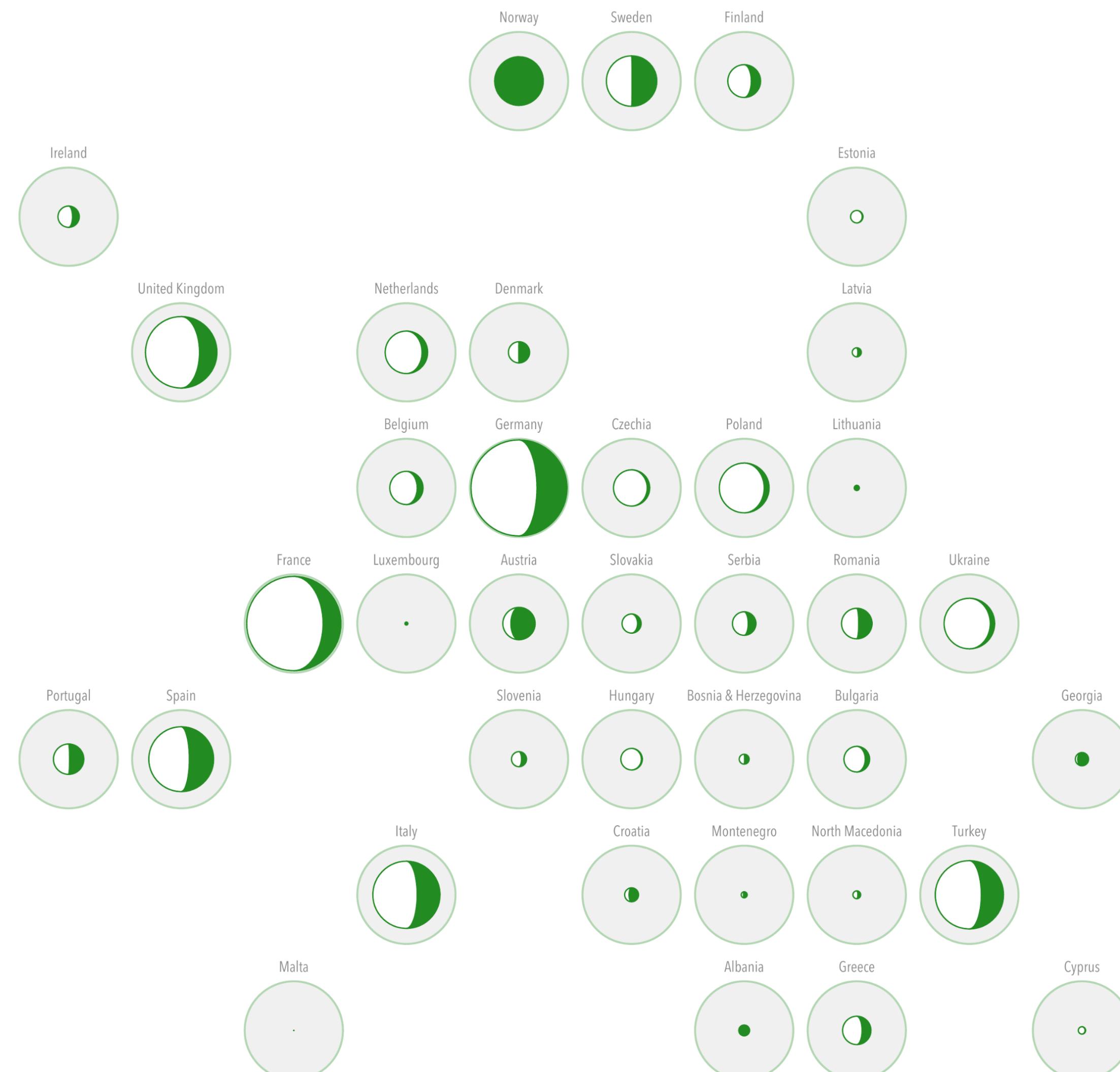


Renewable energy is energy that comes from resources that are naturally replenished such as sunlight, wind, water, and geothermal heat. Unlike fossil fuels, such as oil, natural gas and coal, or nuclear power sources such as uranium and plutonium, renewable energy regenerates naturally in a short period of time.

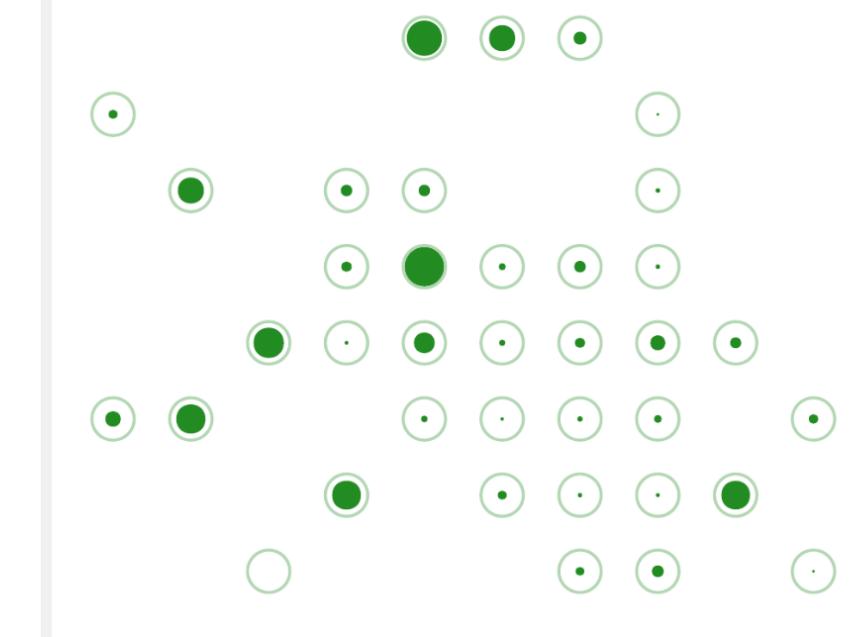


Norway had an electricity production almost entirely made up of renewable energy (98%). This makes Norway the second largest producer of this energy type in Europe. Interestingly, most of the renewable energy is produced by hydro power that take up 95% and only 3% by wind. In contrast, twelve European countries were reported to produce less than 20% of their energy with renewable resources: Malta (0%), Hungary (5%), Estonia (6%), Czechia (7%), Cyprus (9%), Ukraine (9%), Poland (10%), Netherlands (13%), Bulgaria (17%), Belgium (18%), Slovakia (19%), and France (19%).

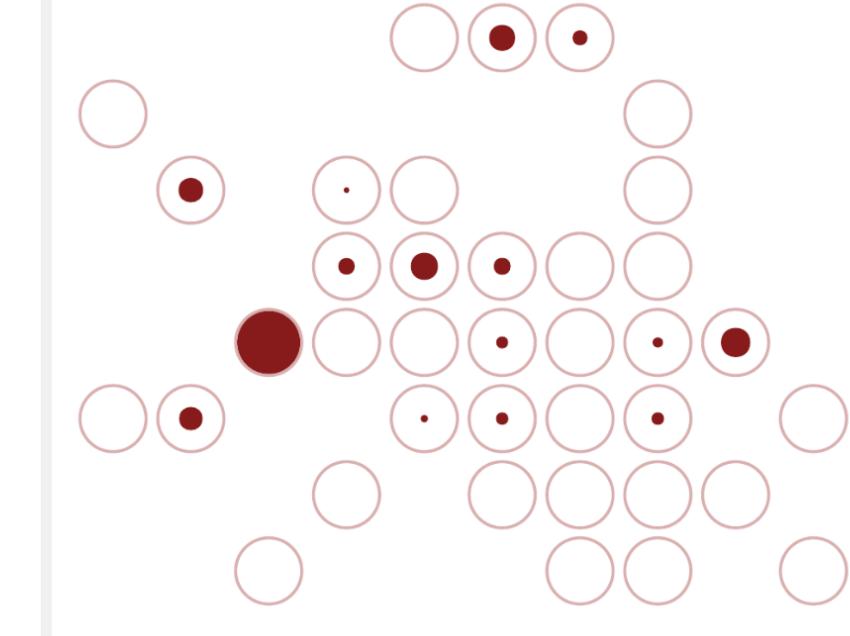
Note: Energy production is mapped to the area of the circles.  
Visualization by Cédric Scherer • Data by Eurostat



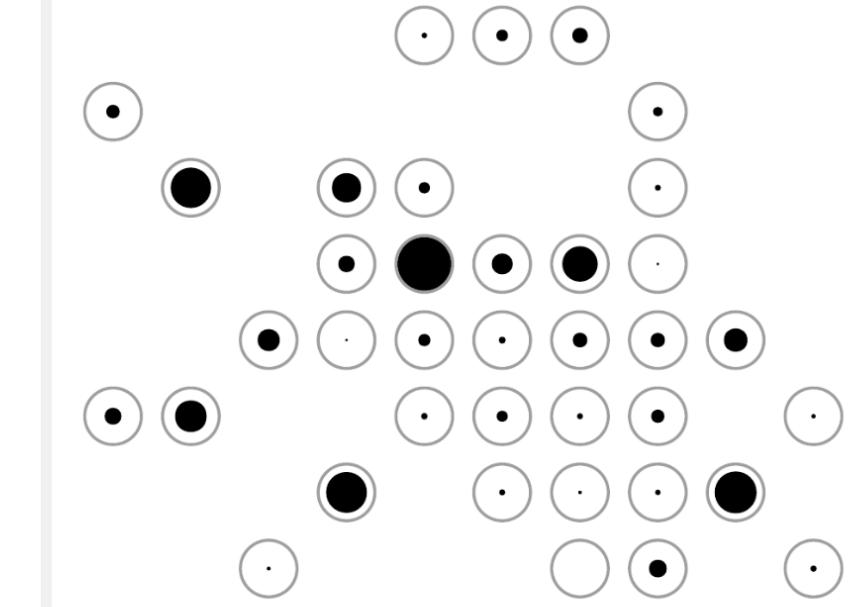
### Renewable energy



### Nuclear energy

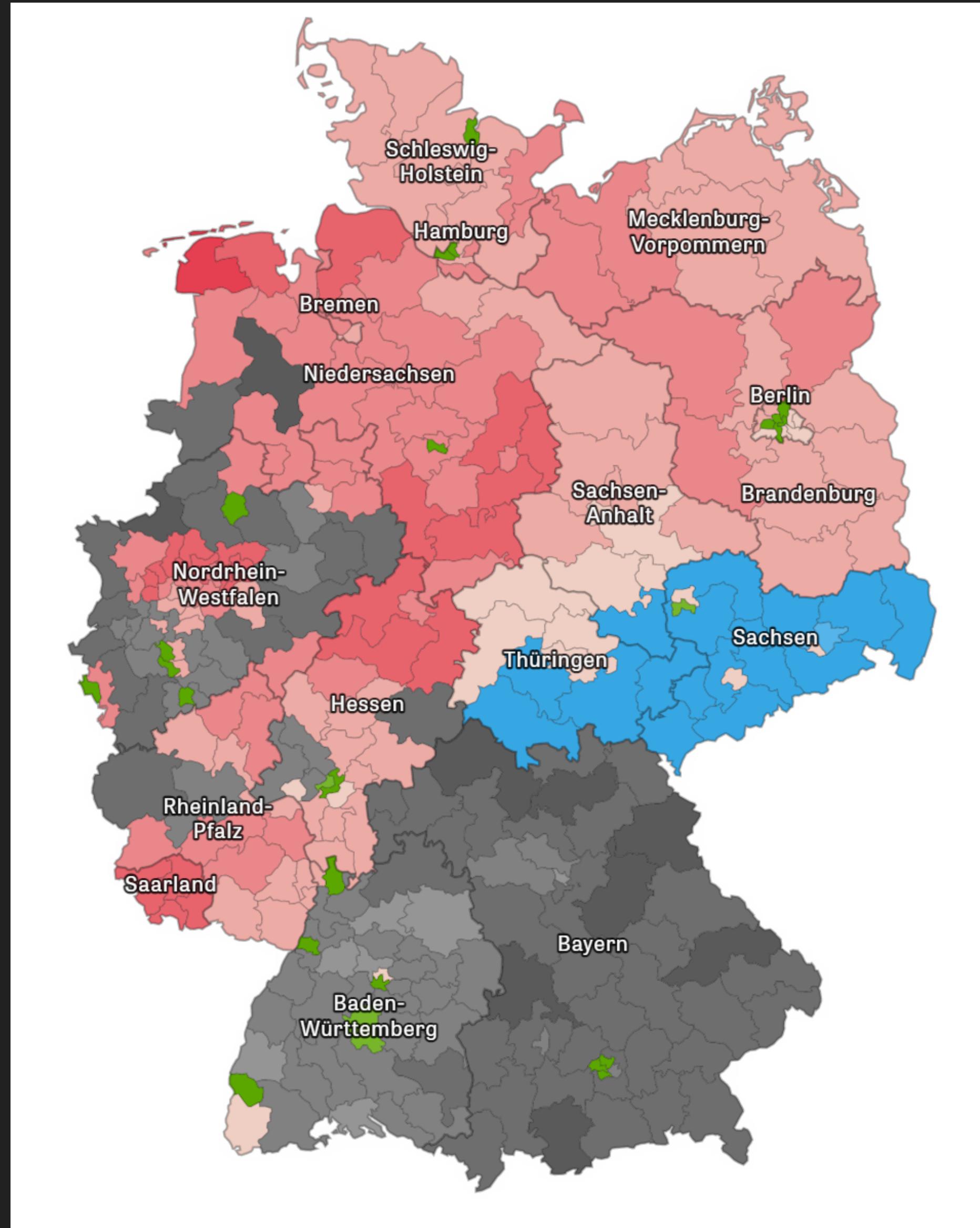


### Conventional thermal energy



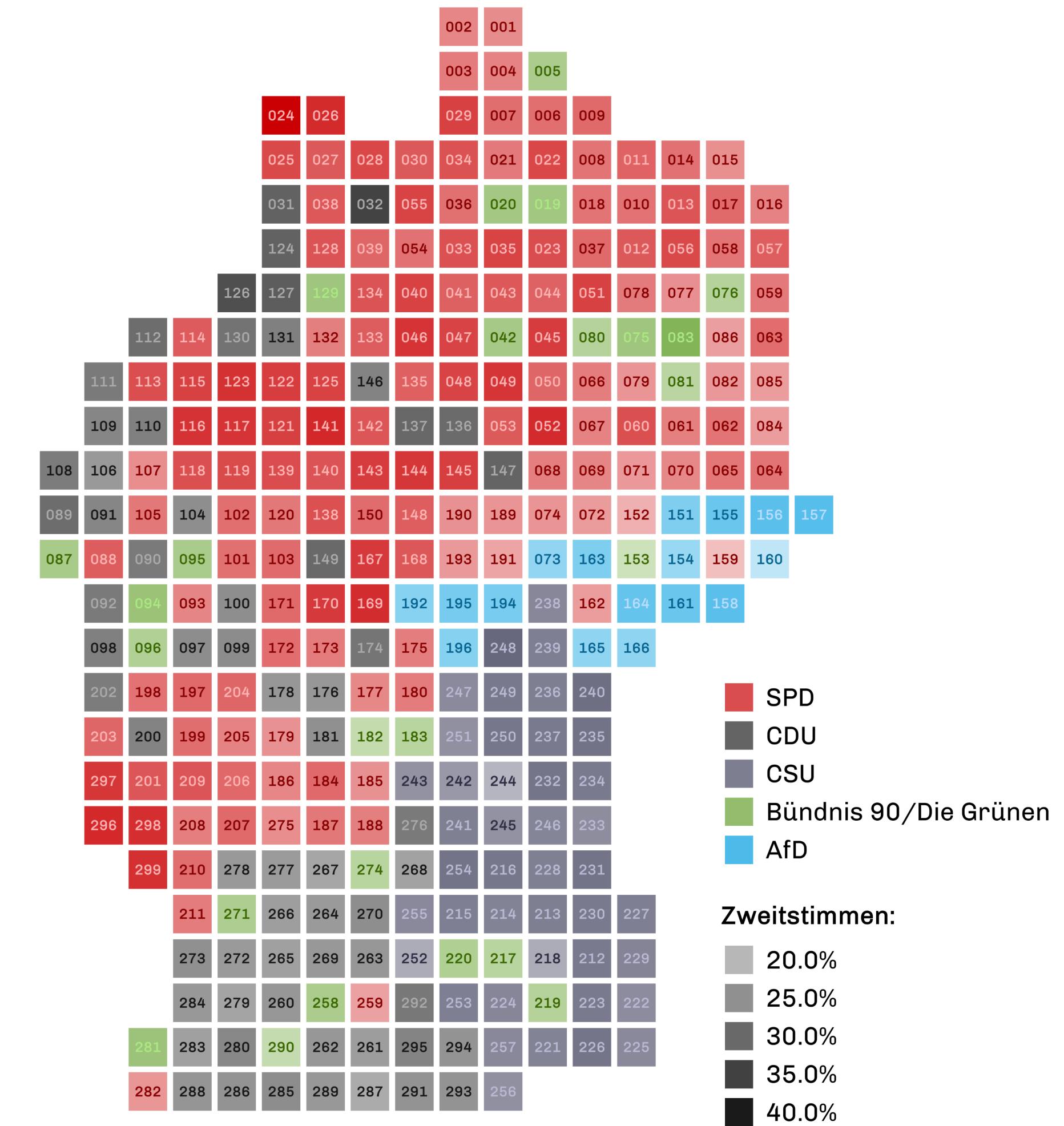
**“How European countries generated electricity in 2018”, #TidyTuesday Contribution**

Cédric Scherer // rstudio::conf // July 2022



## Ergebnisse der Bundestagswahl 2021

Die stärksten Parteien nach Prozent der Zweitstimmen.



Grafik: Cédric Scherer & Ansgar Wolsing • Daten: DIE ZEIT

Left: Choropleth Map by Die Zeit | Right: Tile Grid Map by Cédric Scherer & Ansgar Wolsing

# *Visual Form*

---

**Follow design rules and data visualization principles**

# What is good DataViz design?

# What is good DataViz design?

- Clean layout — “less is more”

# What is good DataViz design? *often but not necessarily!*

- Clean layout — “less is more”

# What is good DataViz design?

*often but not necessarily!*

- Clean layout — “less is more”
- Use direct annotations to ease readability + interpretability

# What is good DataViz design? *often but not necessarily!*

- Clean layout — “less is more”
- Use direct annotations to ease readability + interpretability
- Make use of hierarchy to guide the reader

# What is good DataViz design? *often but not necessarily!*

- Clean layout — “less is more”
- Use direct annotations to ease readability + interpretability
- Make use of hierarchy to guide the reader
- Consistent use of colors, spacing, typefaces, and weights

# What is good DataViz design?

*often but not necessarily!*

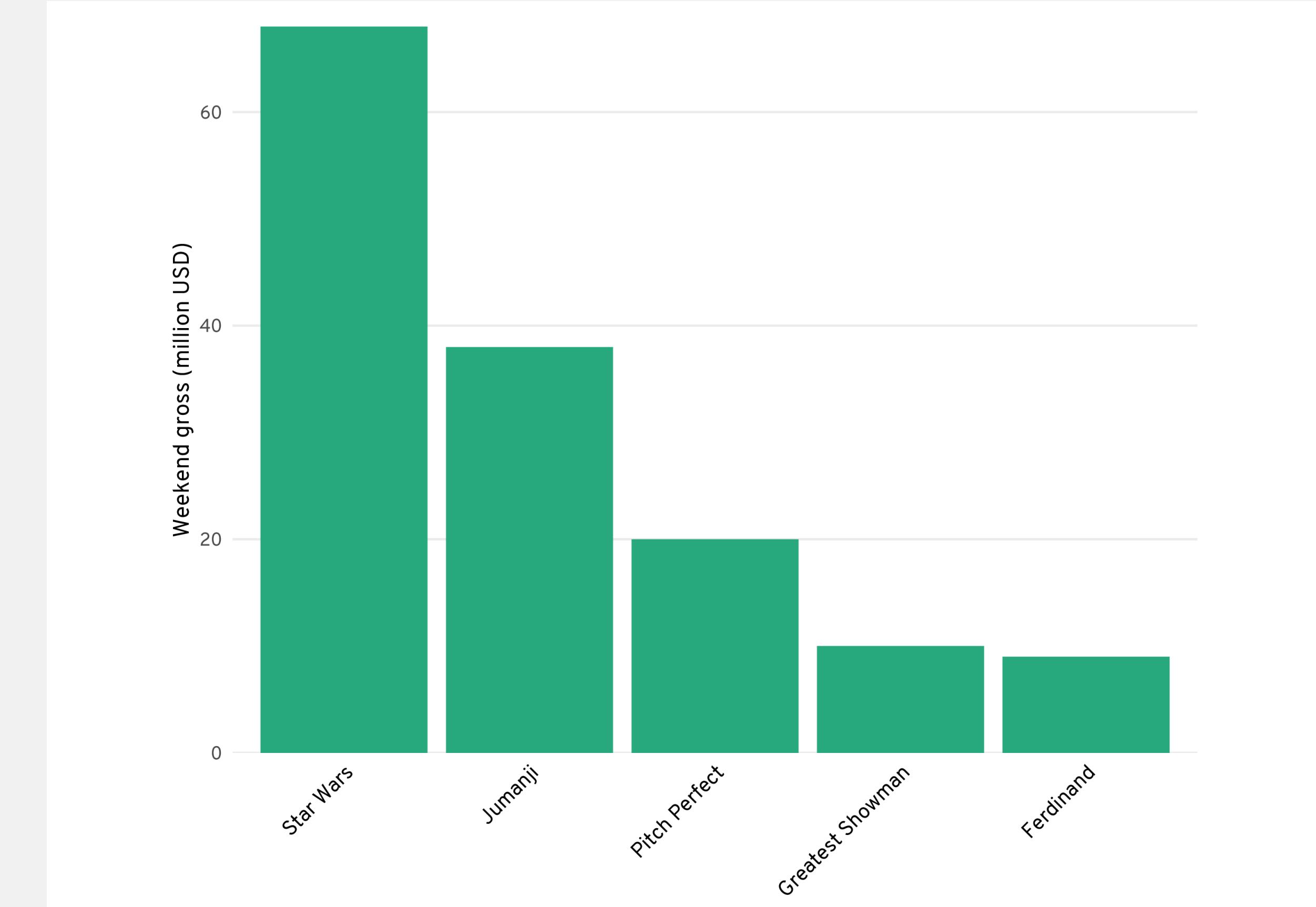
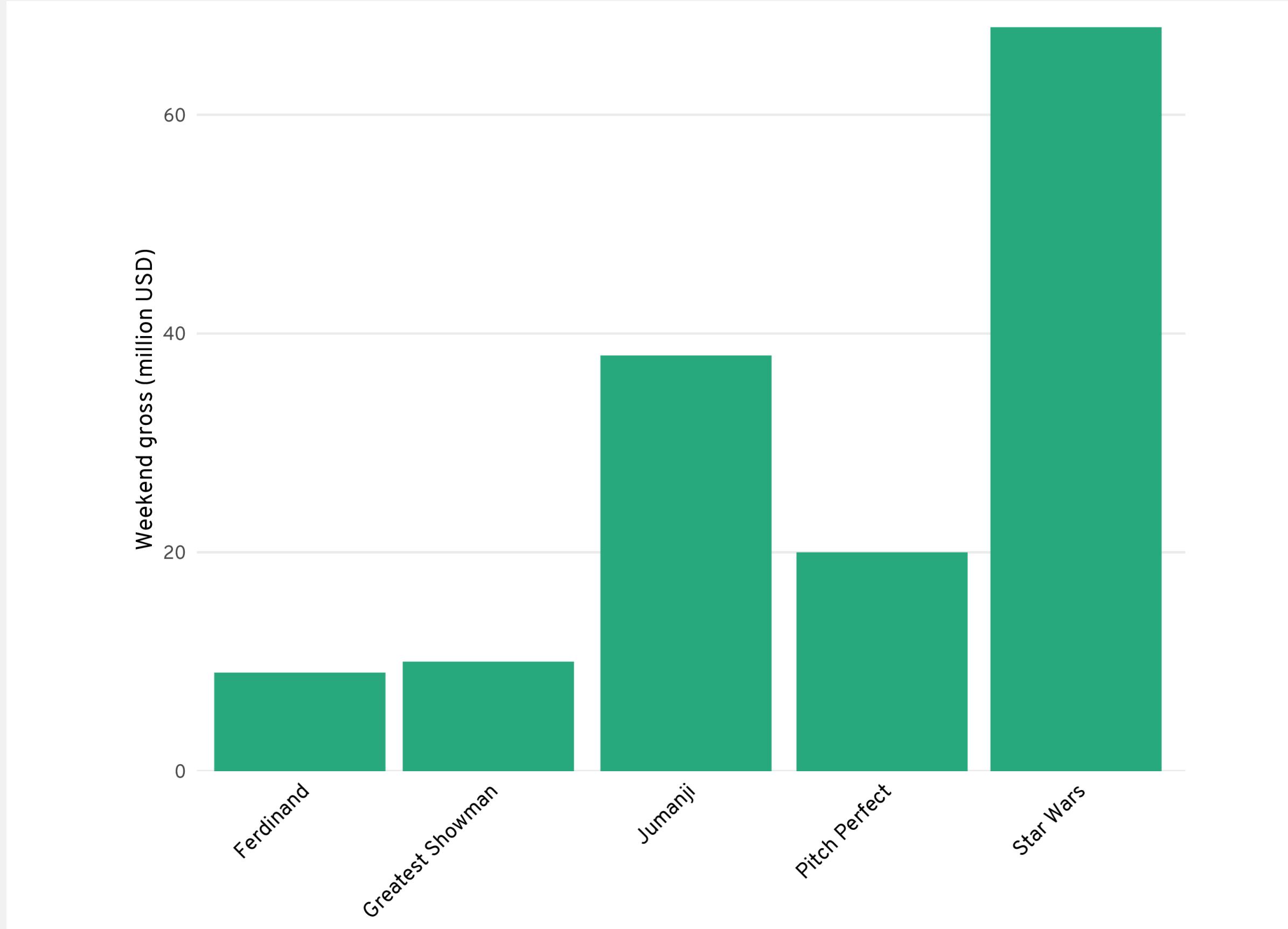
- Clean layout — “less is more”
- Use direct annotations to ease readability + interpretability
- Make use of hierarchy to guide the reader
- Consistent use of colors, spacing, typefaces, and weights
- Use colors wisely and make sure they work for colorblind persons

# What is good DataViz design?

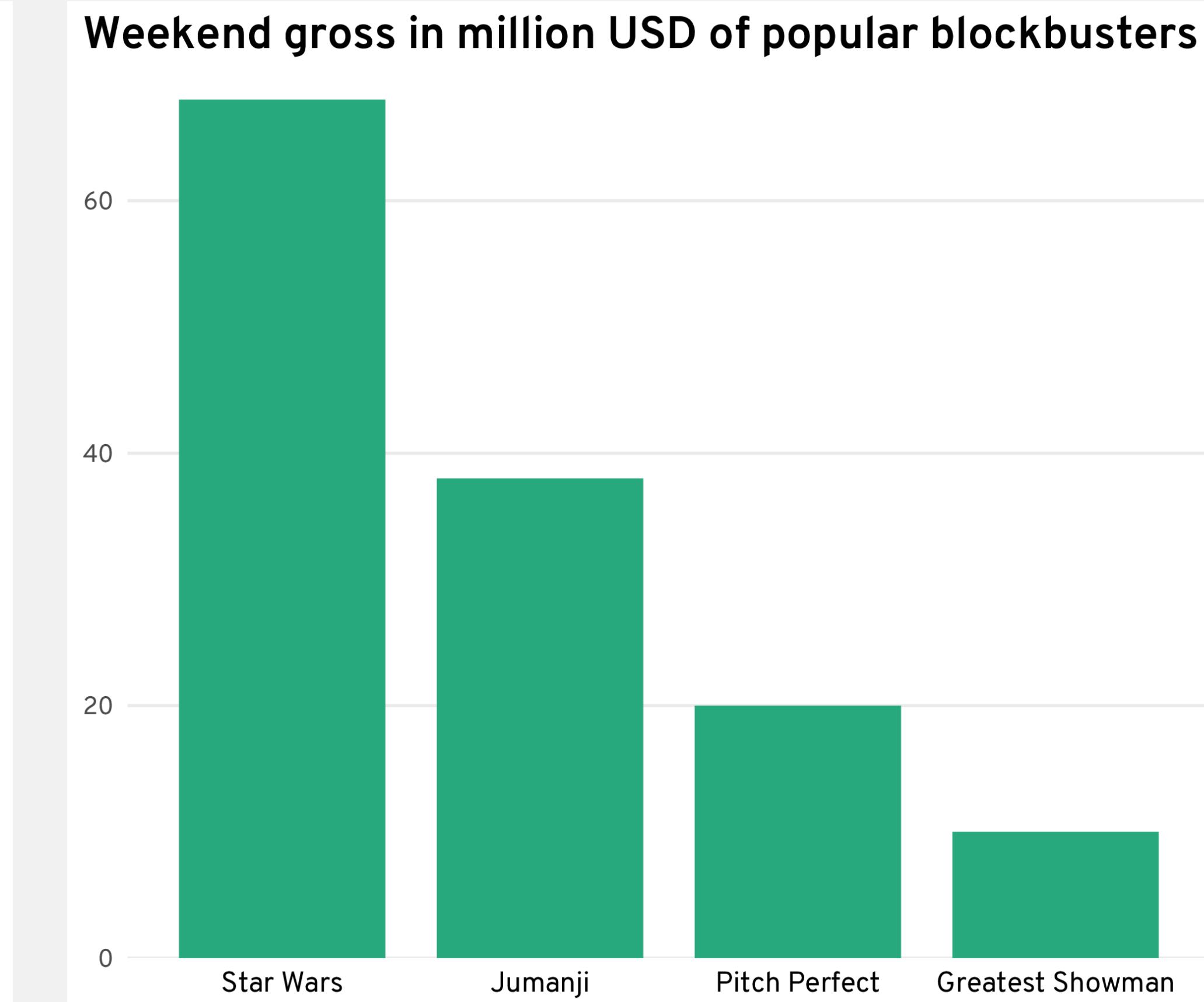
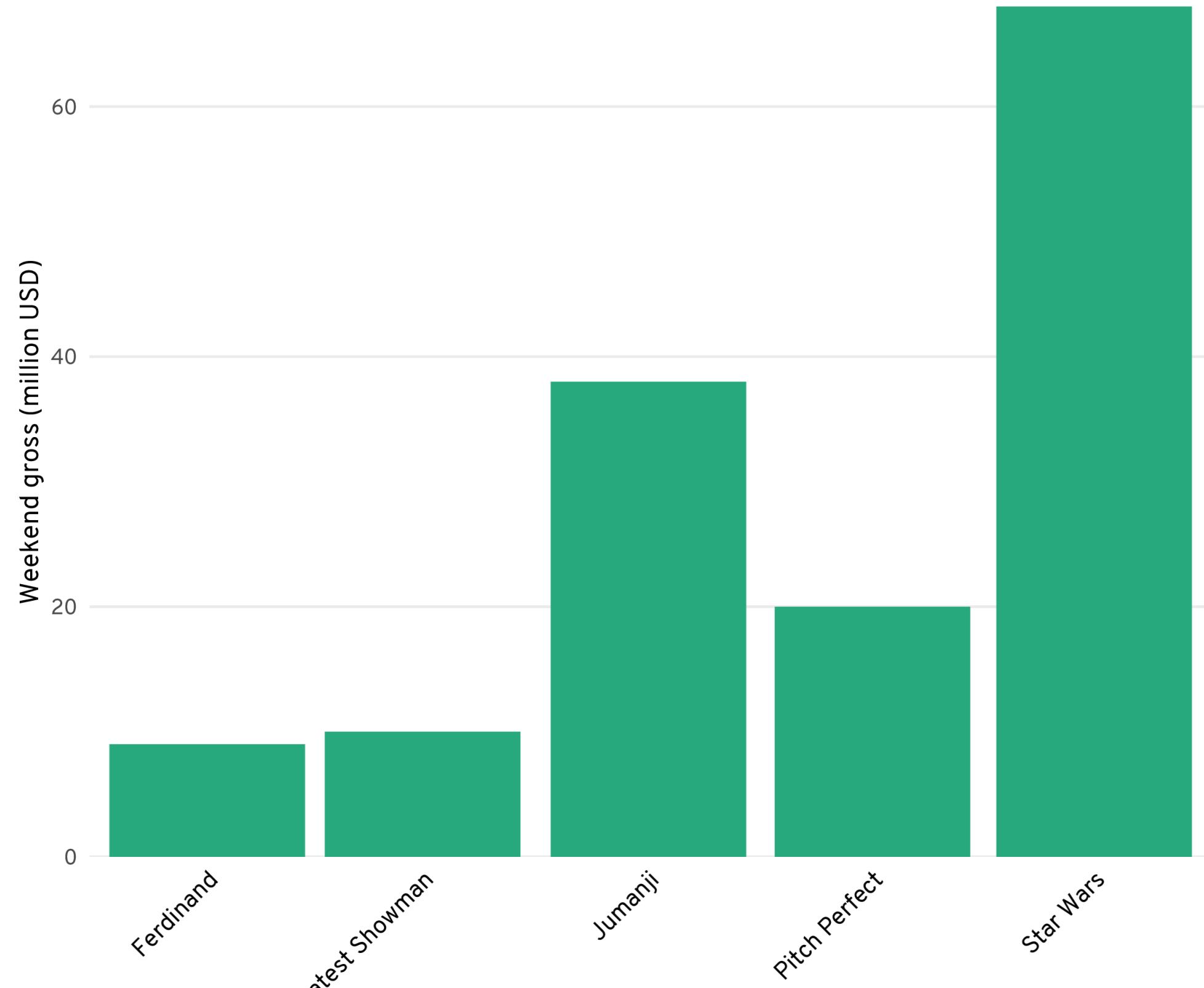
*often but not necessarily!*

- Clean layout — “less is more”
- Use direct annotations to ease readability + interpretability
- Make use of hierarchy to guide the reader
- Consistent use of colors, spacing, typefaces, and weights
- Use colors wisely and make sure they work for colorblind persons
- Most important information should receive the main attention

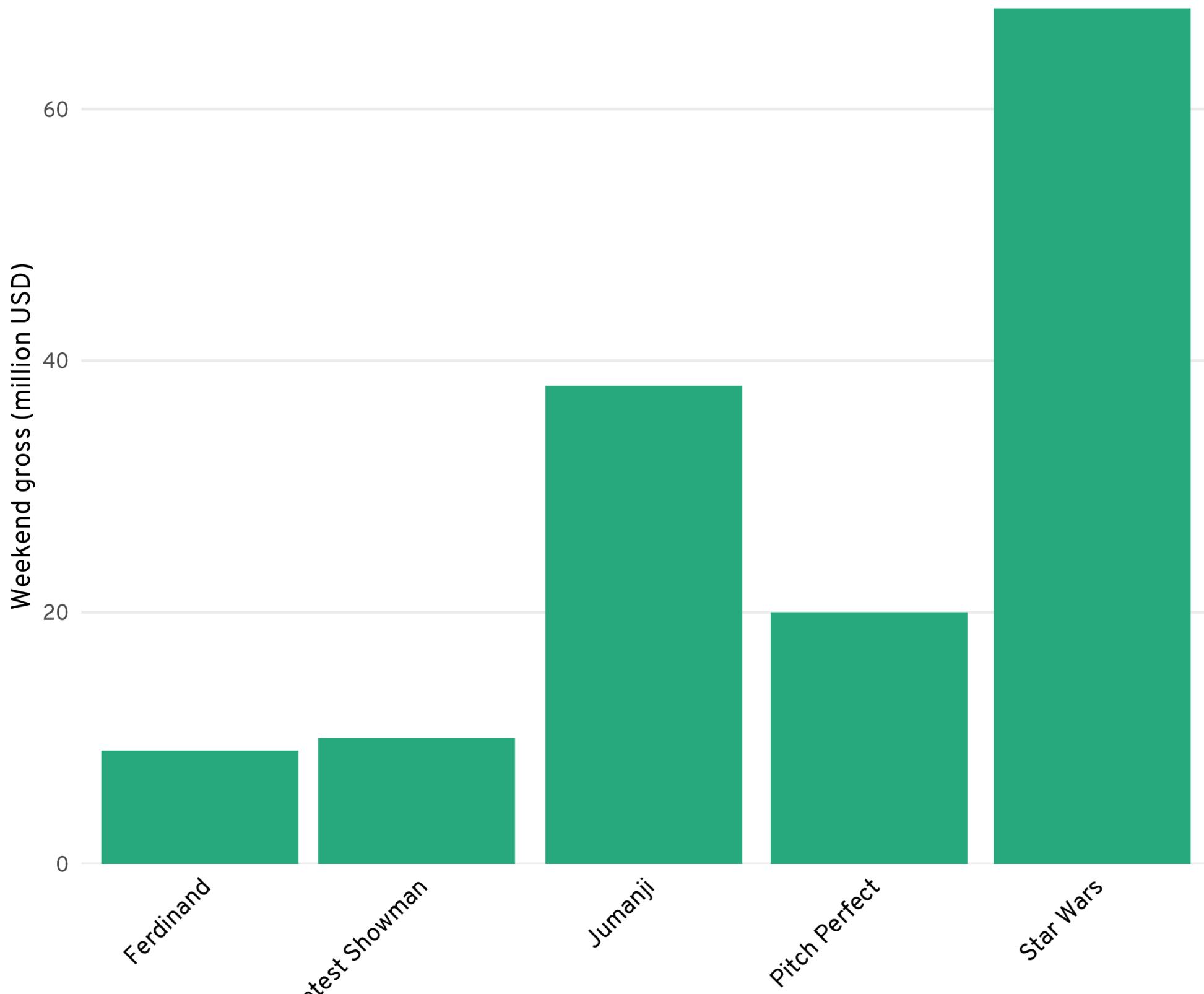
# Order your data



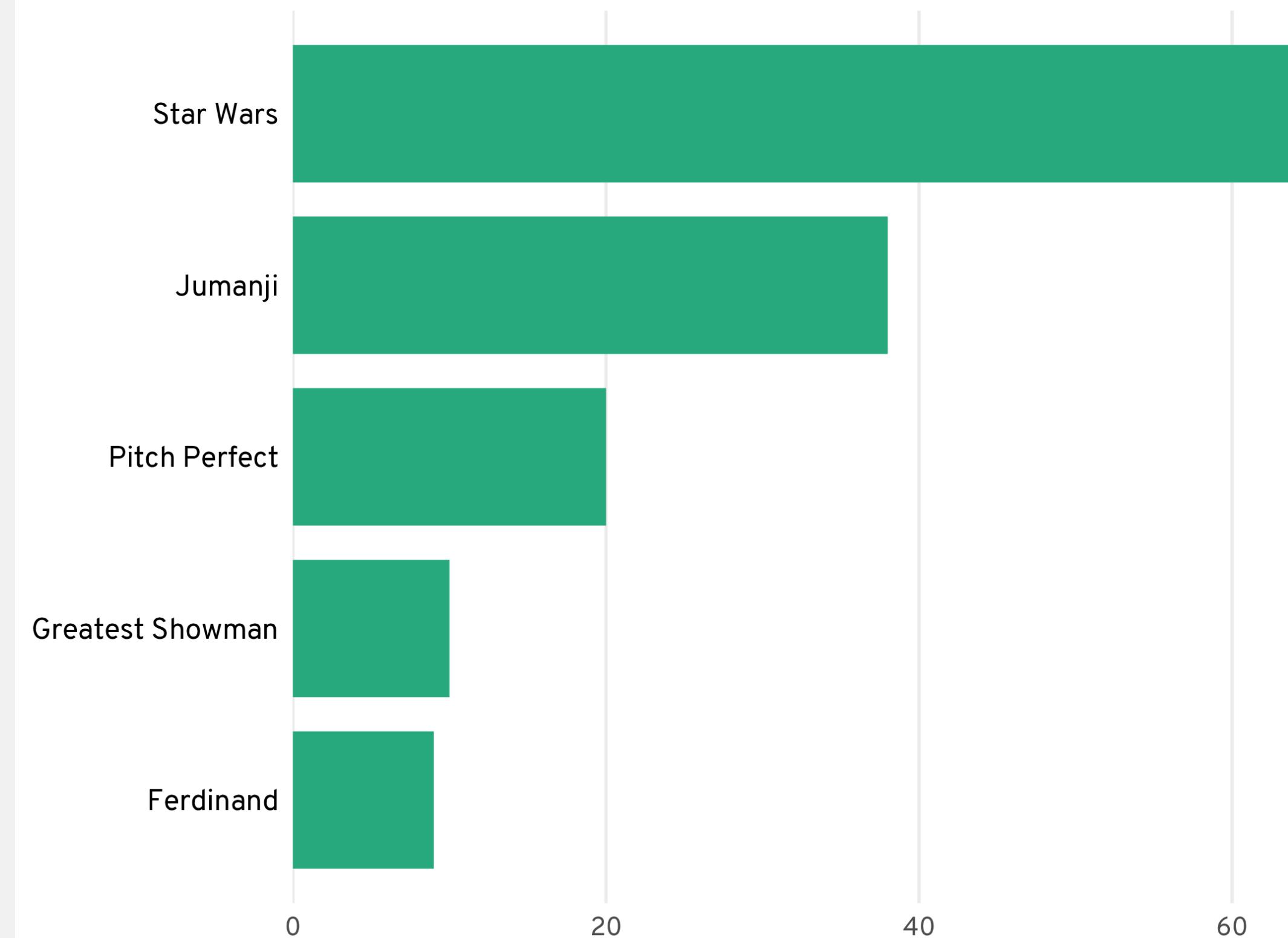
# Don't rotate your text



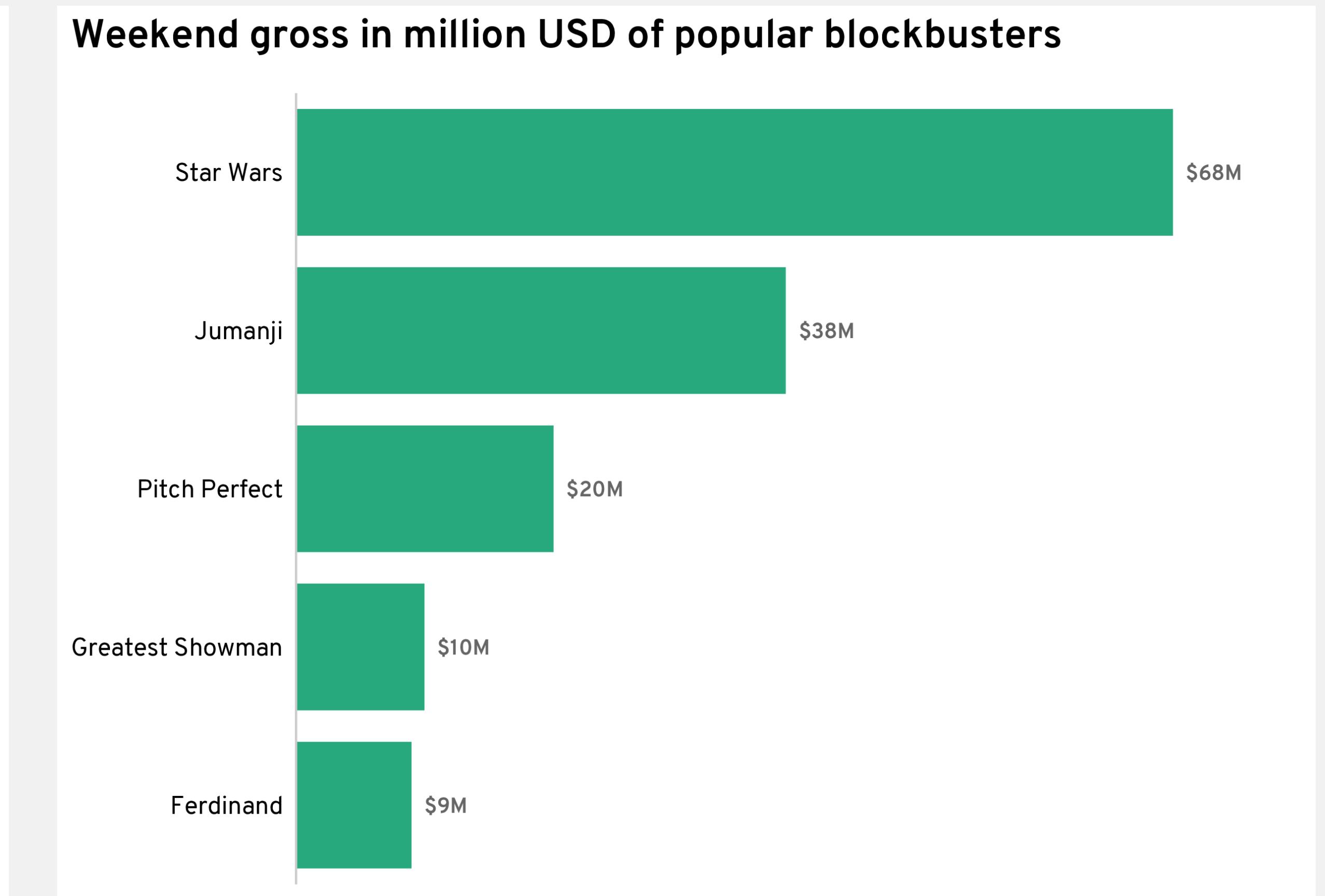
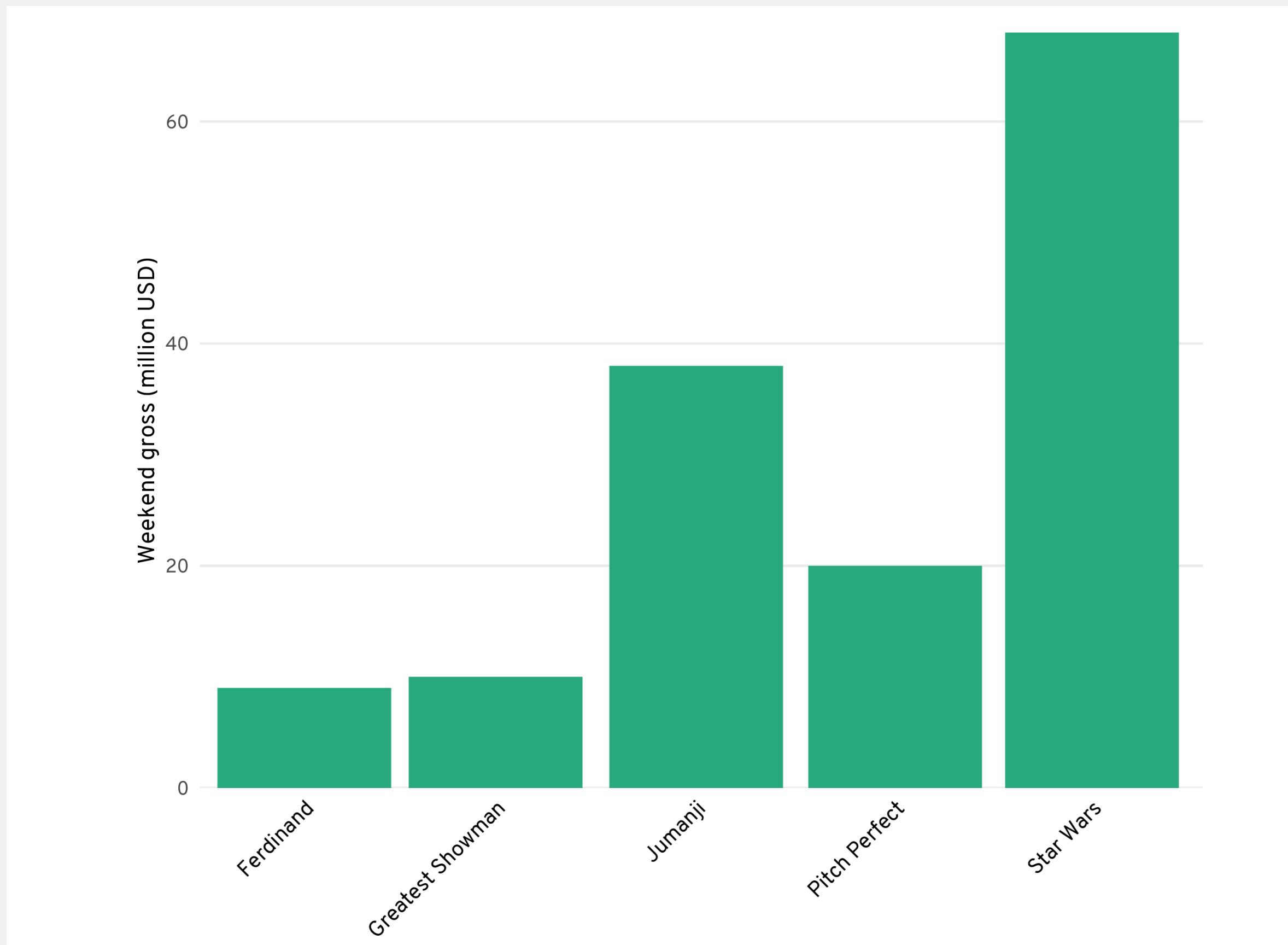
# Don't rotate your text



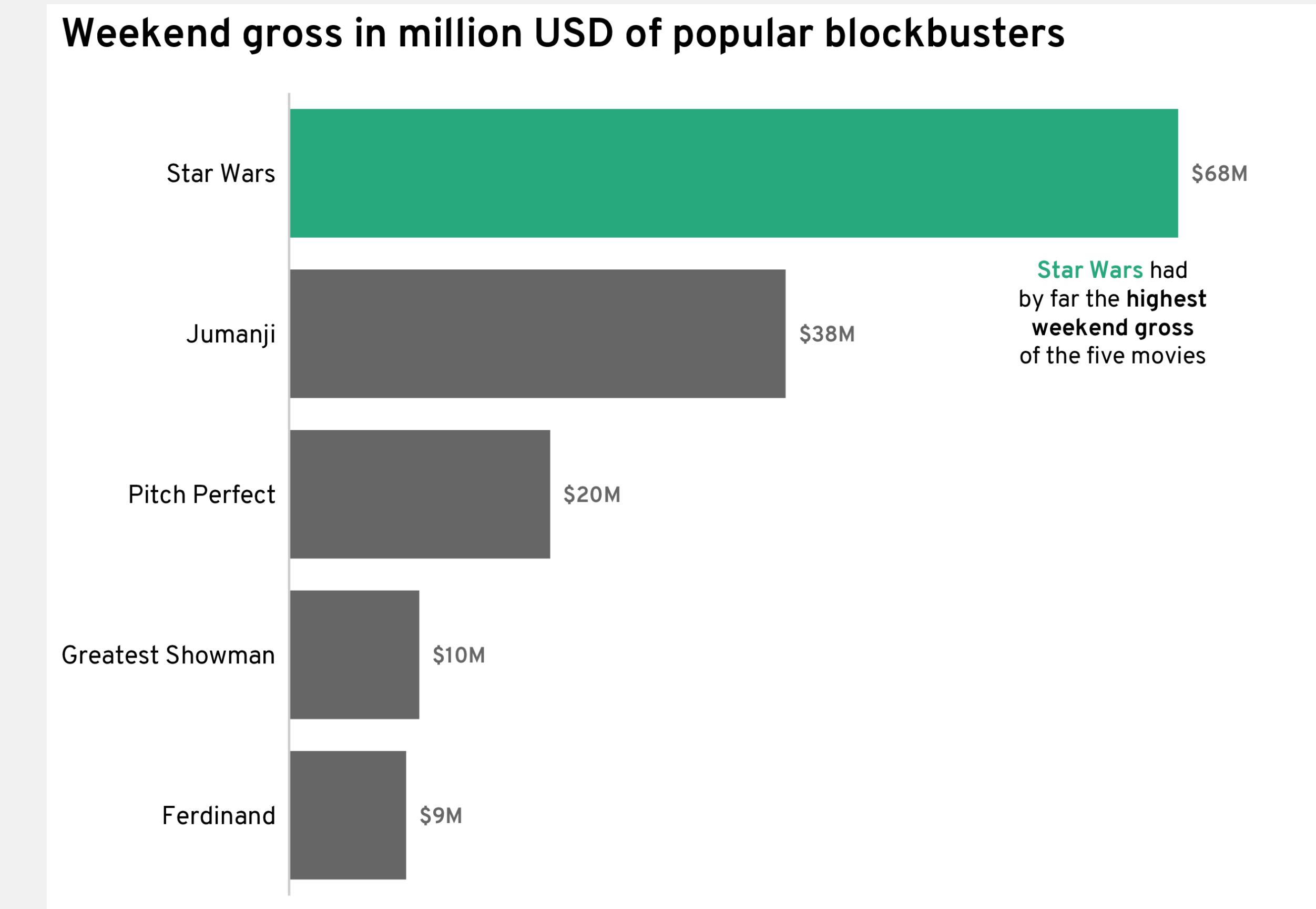
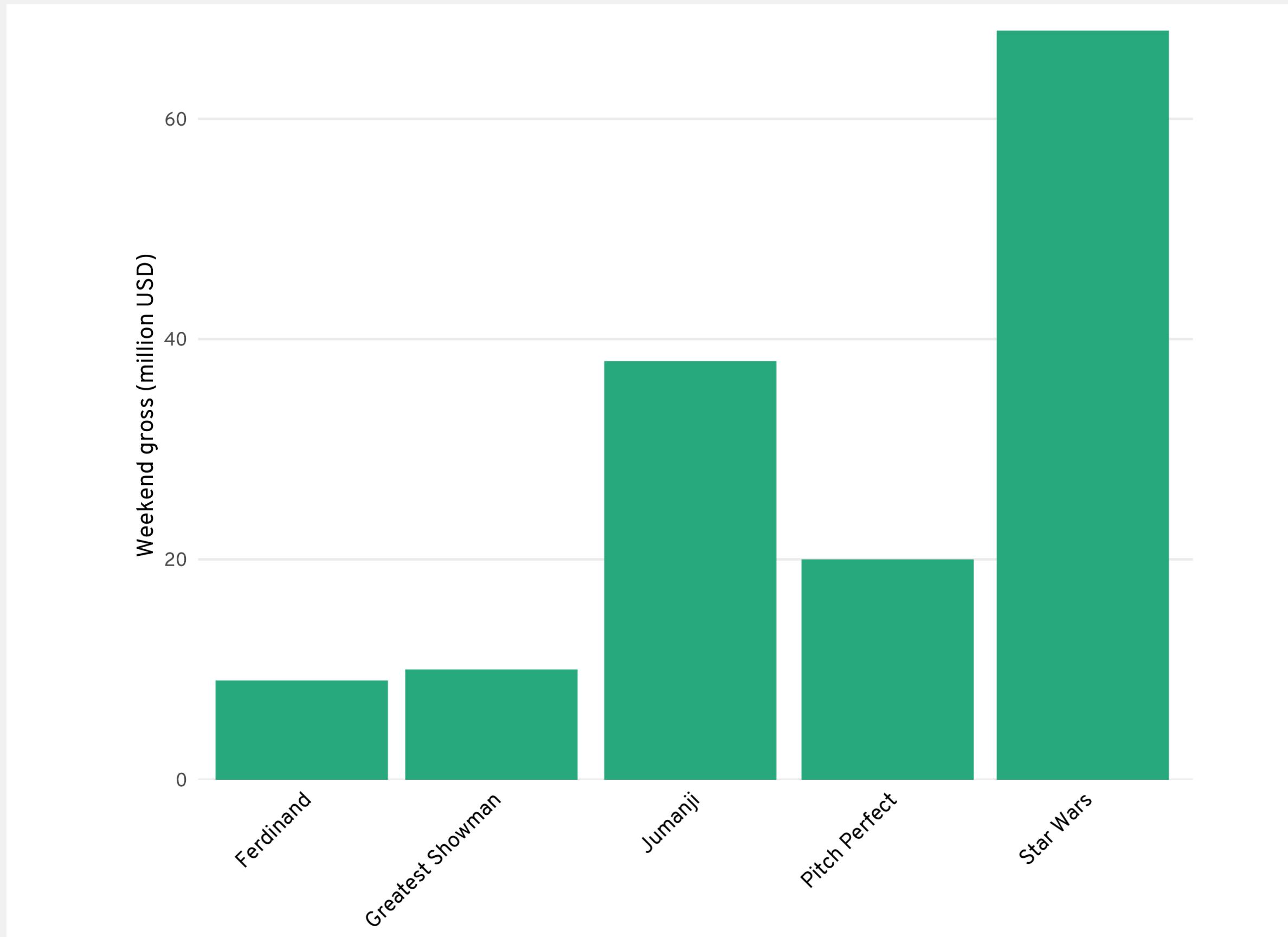
Weekend gross in million USD of popular blockbusters



# Add direct labels



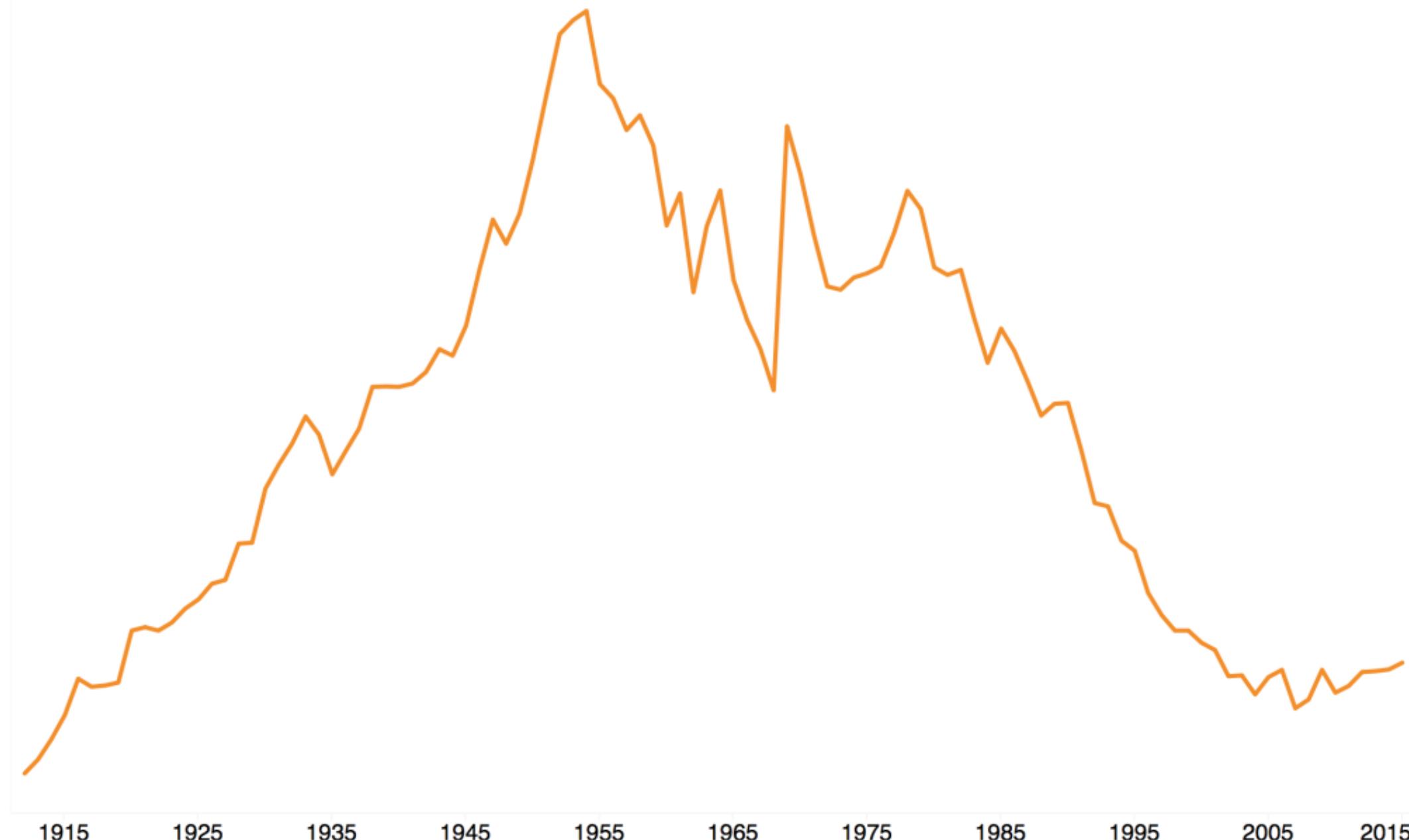
# Use colors + annotations wisely



# The Power of Annotations

Rise and Fall of the name **Neil** in the USA  
Births 1912-2015

Source: data.gov



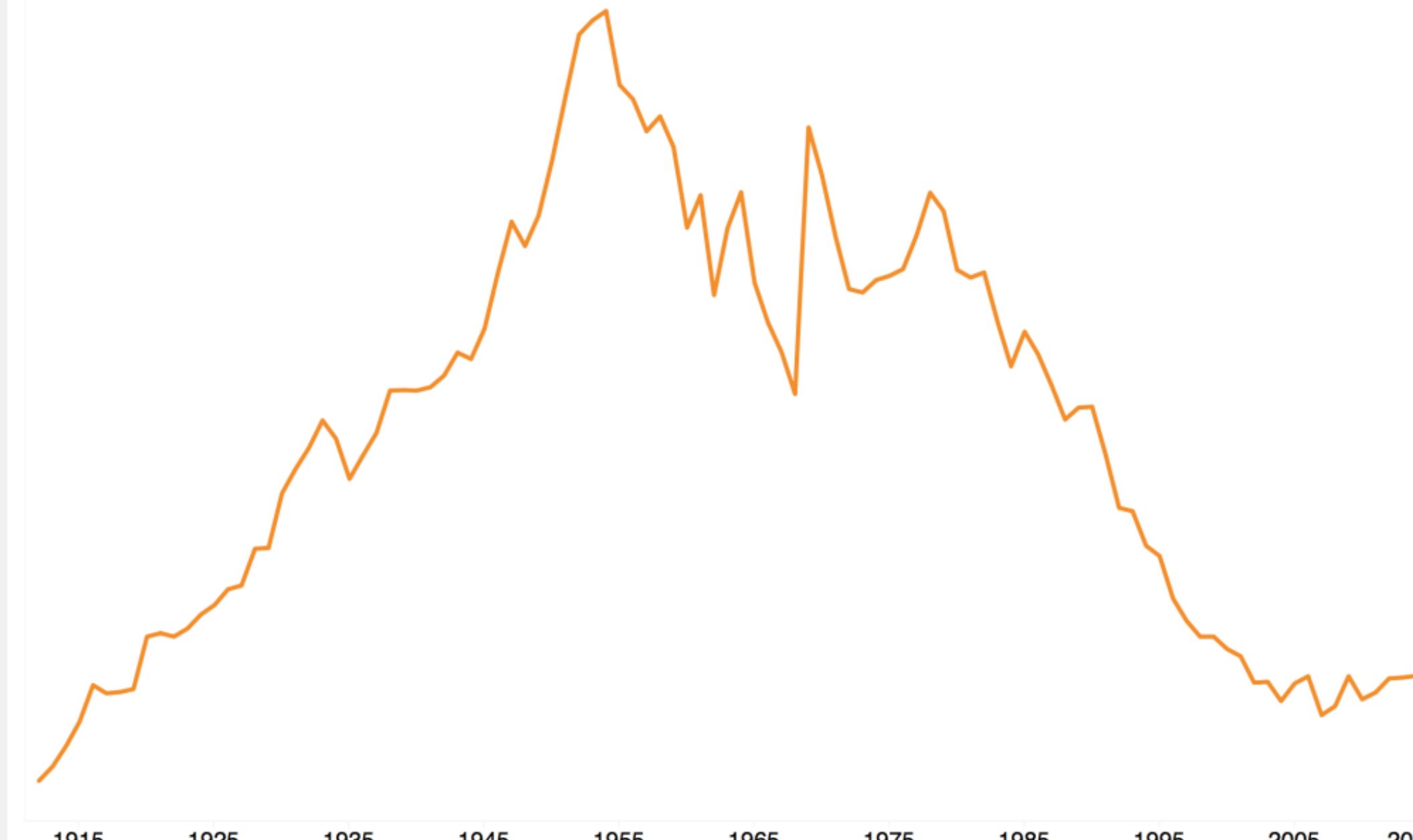
Visualisation: [@theneilrichards](#)

"Is white space always your friend?" by Neil Richards

# The Power of Annotations

Rise and Fall of the name **Neil** in the USA  
Births 1912-2015

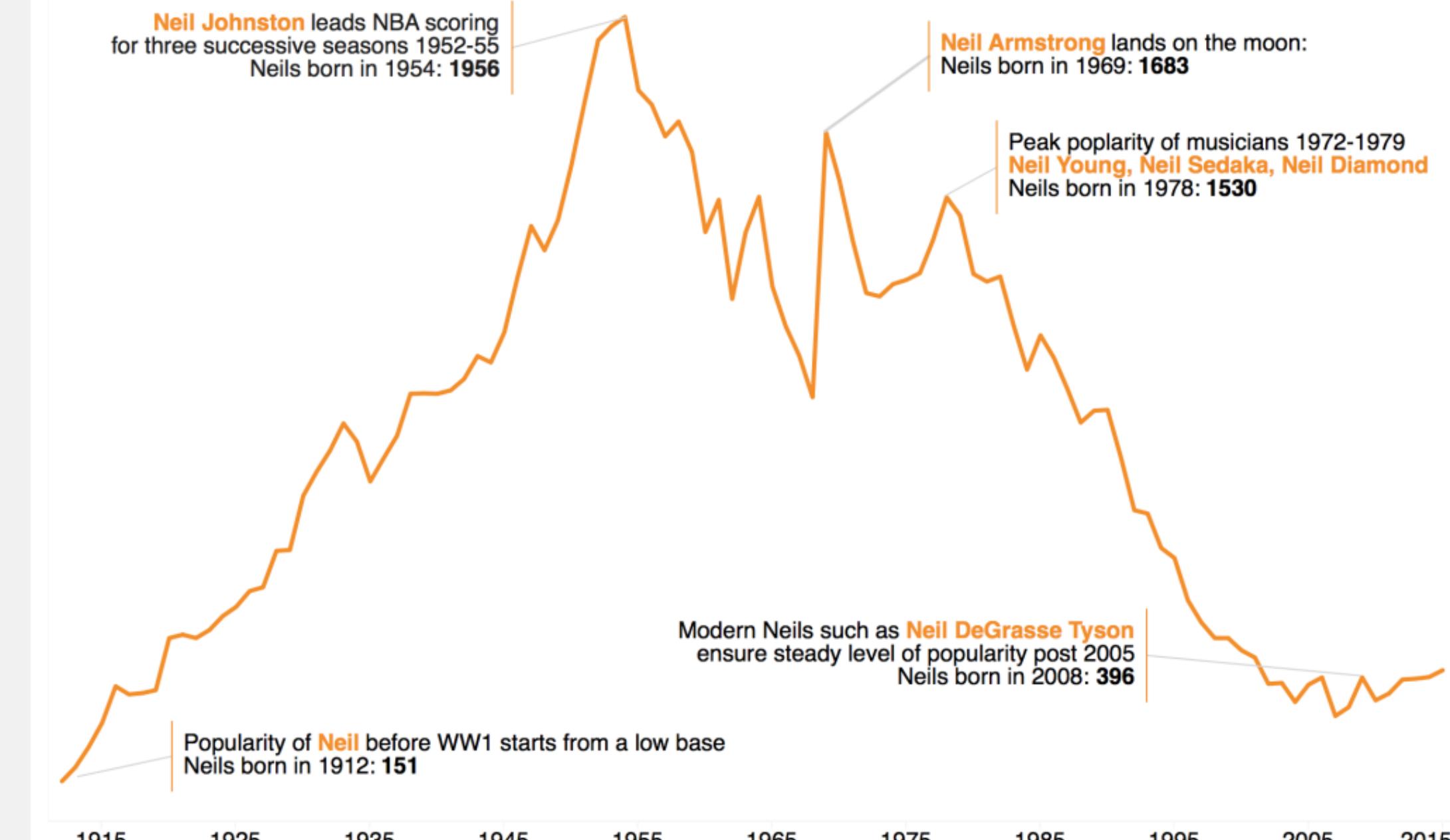
Source: data.gov



Visualisation: [@theneilrichards](#)

Rise and Fall of the name **Neil** in the USA  
Births 1912-2015

Source: data.gov



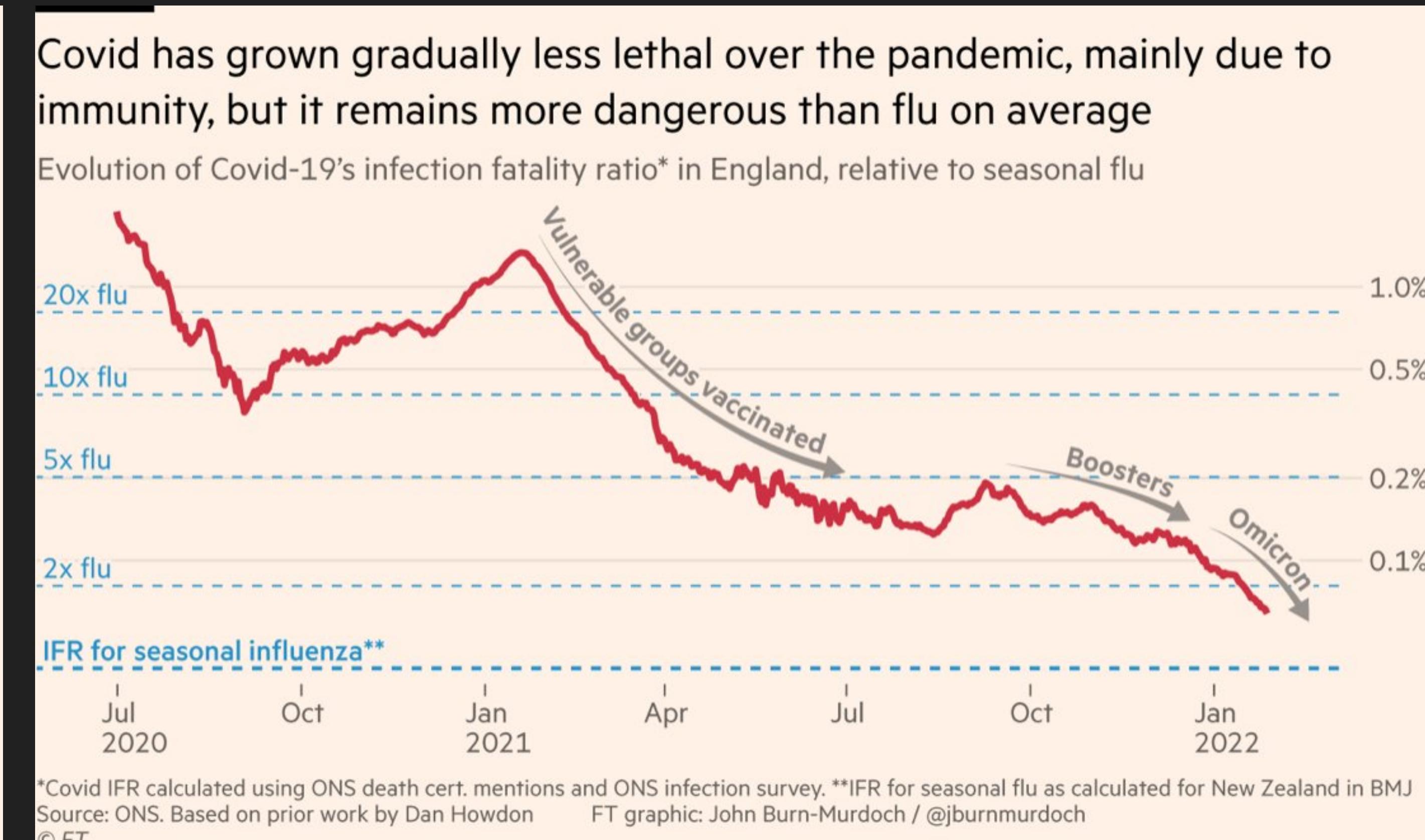
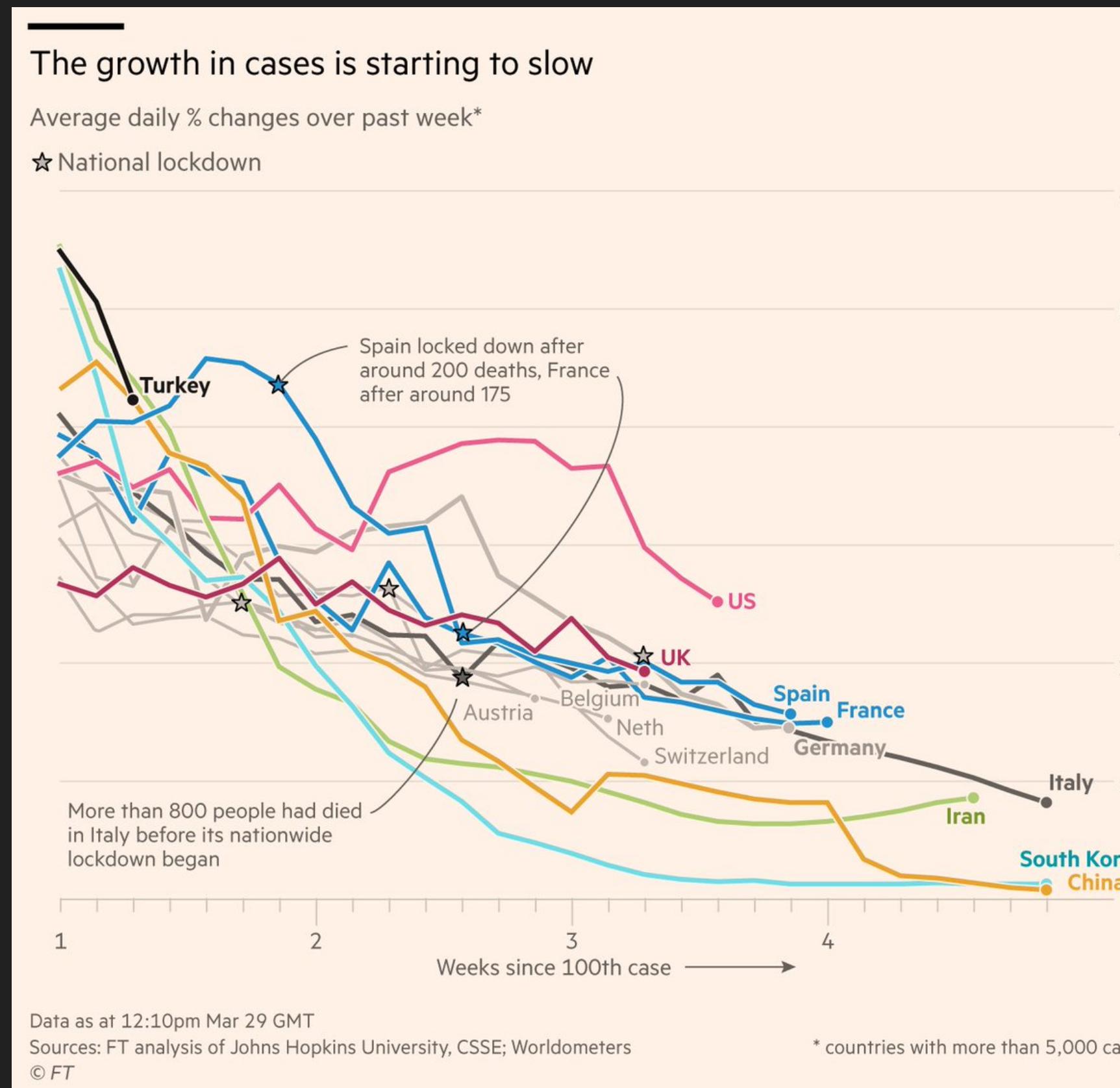
Visualisation: [@theneilrichards](#)

#SWDChallenge

"Is white space always your friend?" by Neil Richards

**“The key thing we do is to add a title to the chart, as an entry point and to explain what is going on. Text and other annotations add enormous value for non-chart people.”**

~ John Burn-Murdoch, Financial Times



Exports and Imports to and from DENMARK & NORWAY from 1700 to 1780.



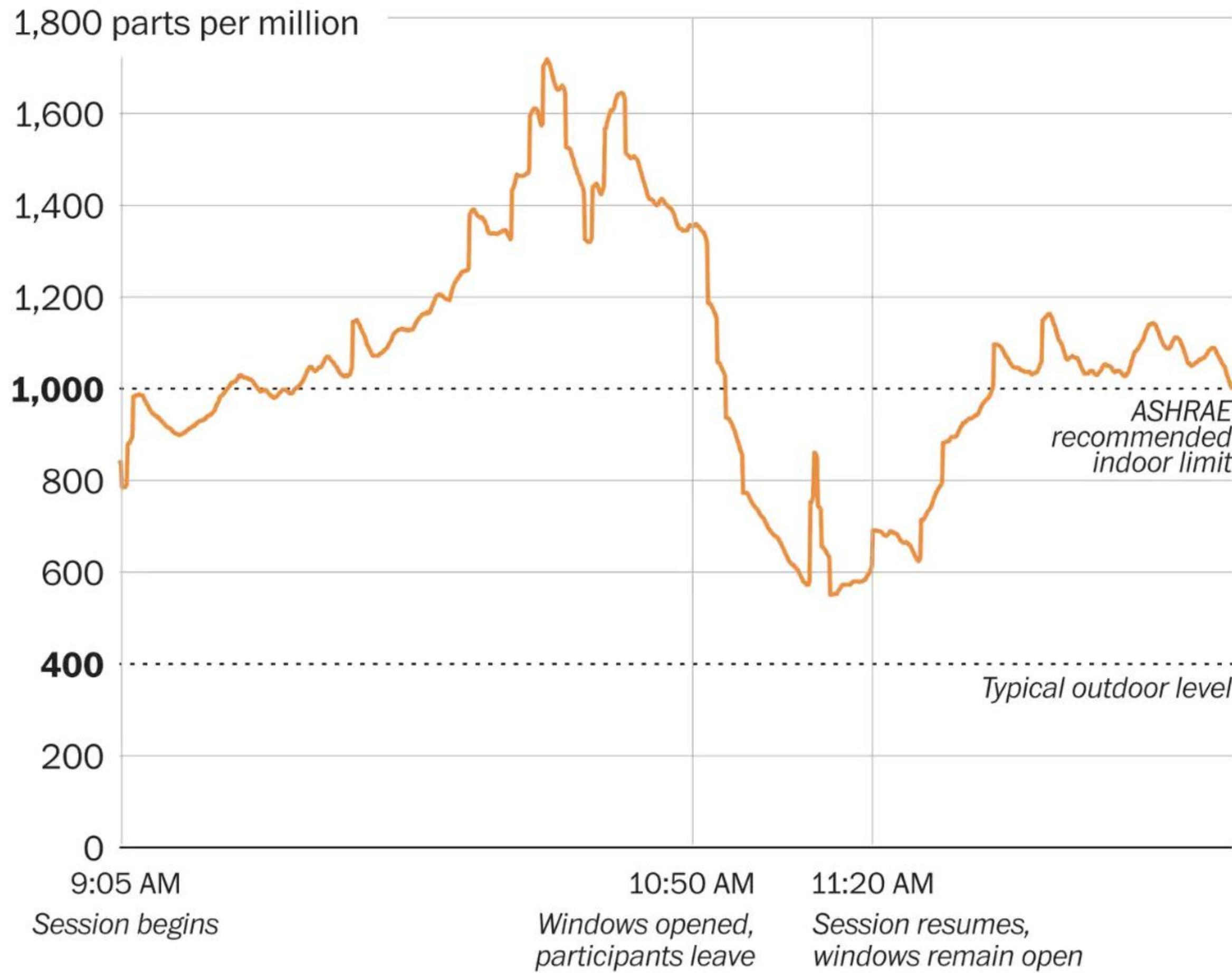
Annotated time-series chart by William Playfair from "The Commercial and Political Atlas and Statistical Breviary" (1786)

# *Wrap Up*

---

# Clearing the air

CO<sub>2</sub> levels in an occupied conference room on June 4, 2019



Source: Adam Ginsburg

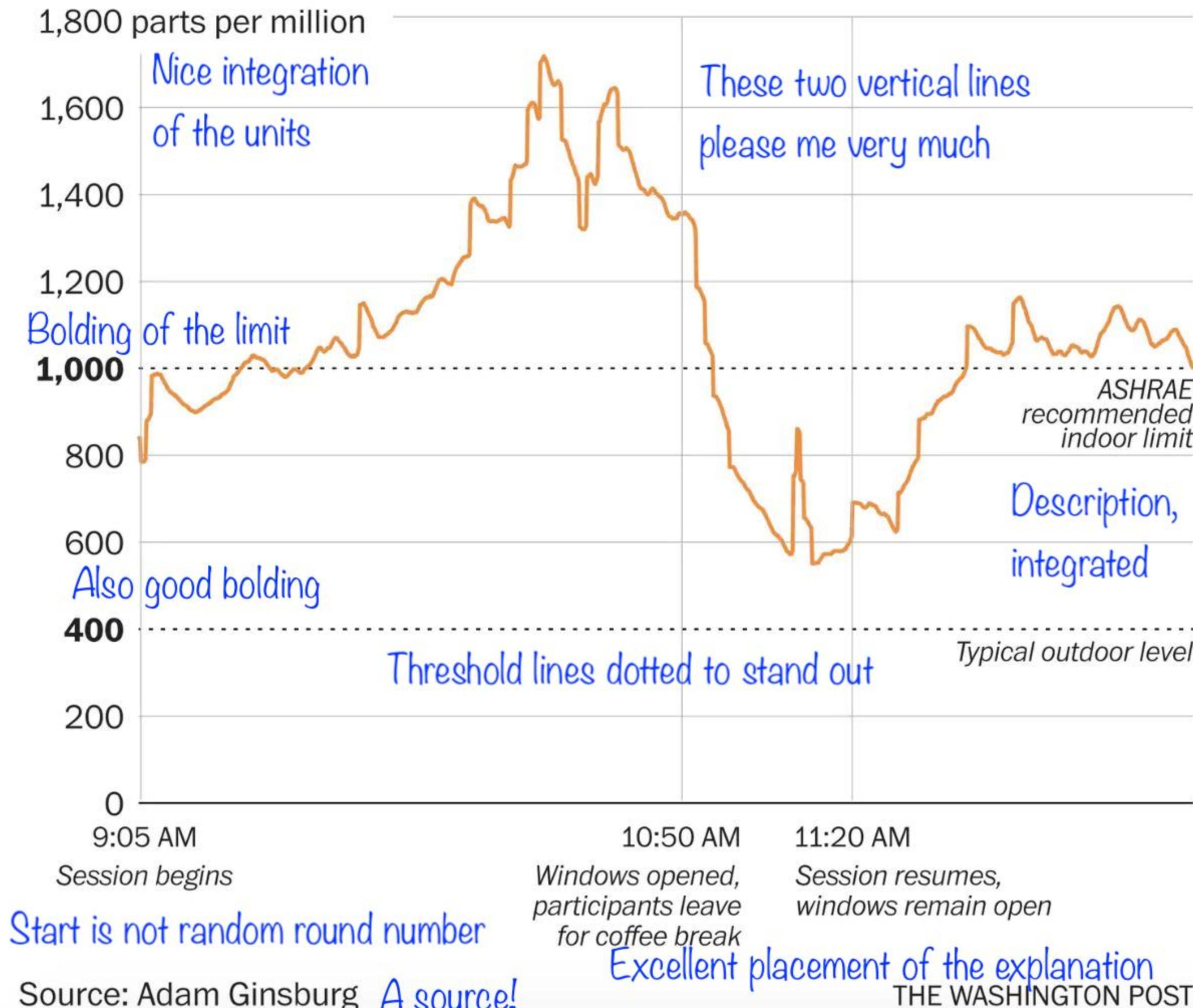
THE WASHINGTON POST

“Clearing the Air” by Adam Ginsburg (Washington Post)

# Clearing the air

Fun and helpful title

CO<sub>2</sub> levels in an occupied conference room on June 4, 2019  
Units and metho in a subtitle, NOT in vertical text on the side



Notes by Francis Gagnon (Voilà)

# Information .....

Understand your data and be accurate.

# Story .....

Be clear about the message of your visualization.

# Goal .....

Select charts that successfully transport your story.

# Visual Form .....

Follow design rules and data visualization principles.

# Your Turn!

We form groups and each group gets a number between 1 and 10.

- Open the image file(s) with the according number in the folder  
**exercises/4-1-data-communication**
- Discuss the visualization with regard to the 4 levels of dataviz design.
  - Overall, do you think it is a good or a bad visualization?
  - What are details you like?
  - How could one improve the chart?
  - Is there another (potentially better) way to tell the story?
    - Sketch it (and think about how you could build it with ggplot2)