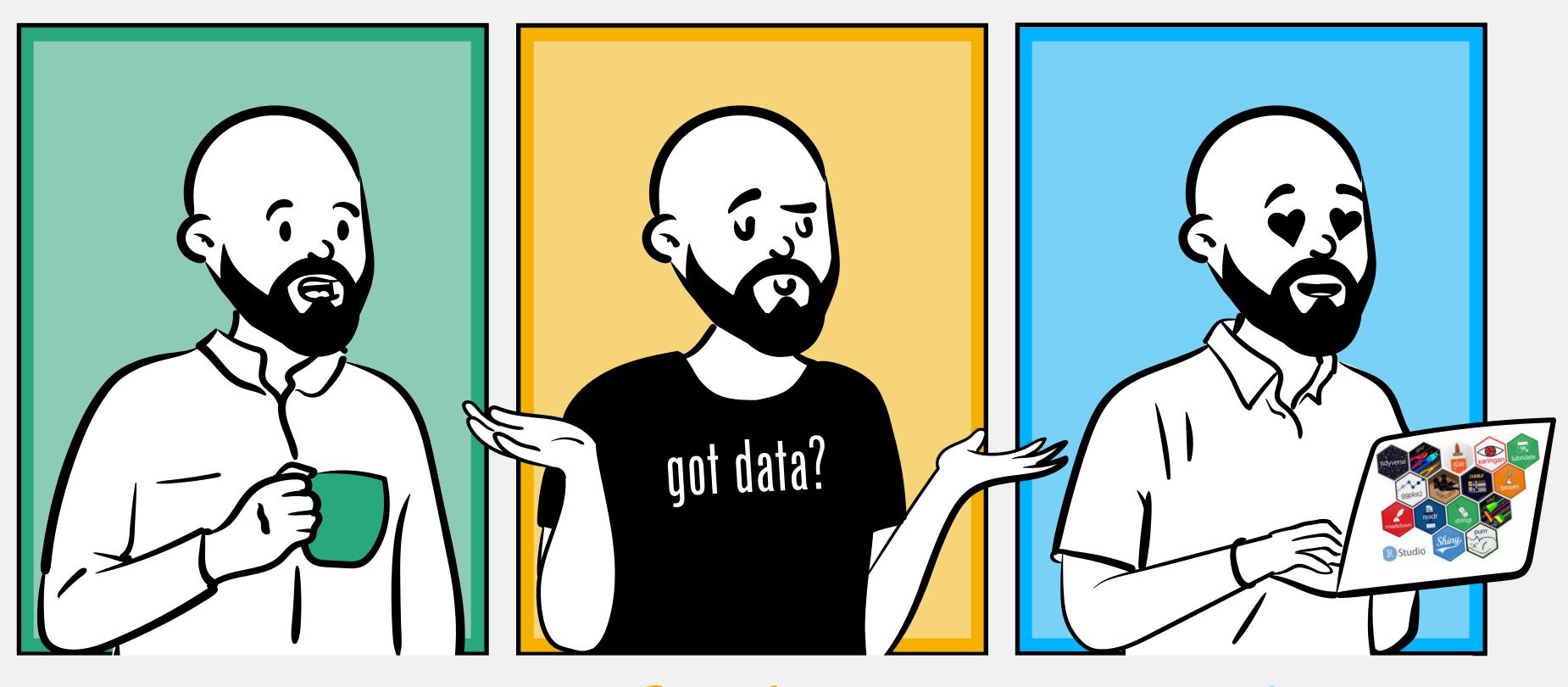


Cédric Scherer

Independent Data Visualization Specialist Computational Ecologist at IZW Berlin



Consulting

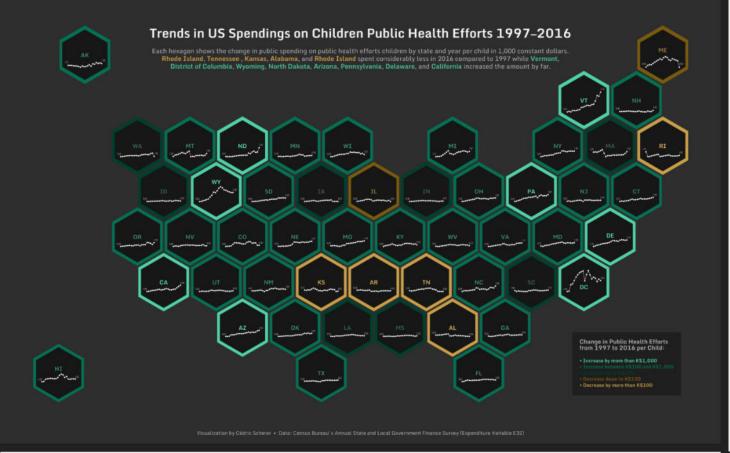
Coaching

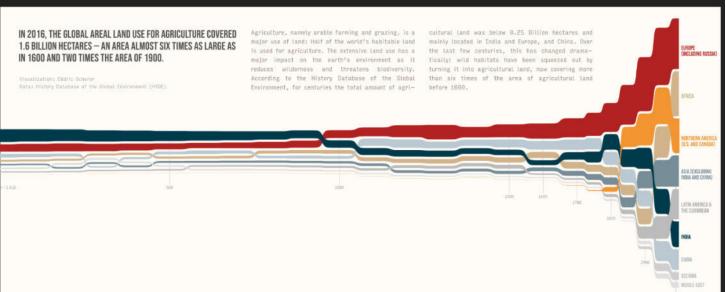
Coding

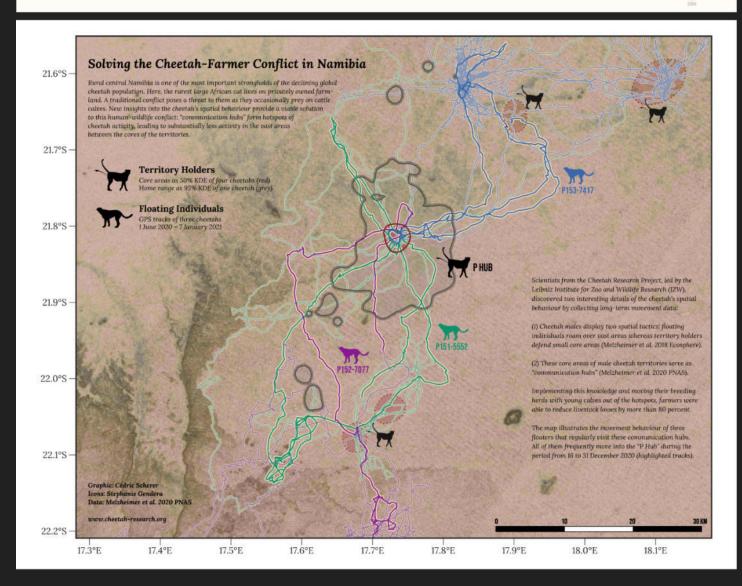


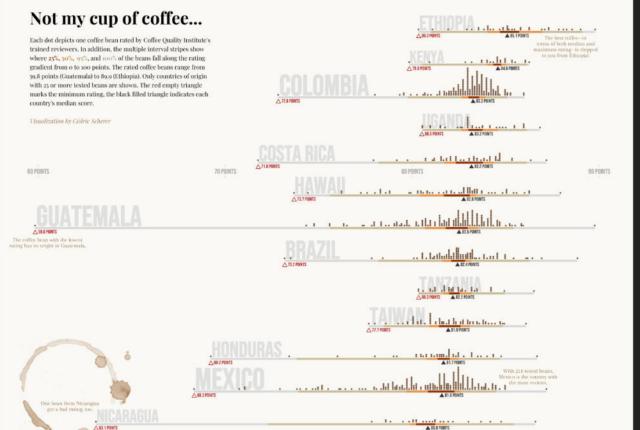




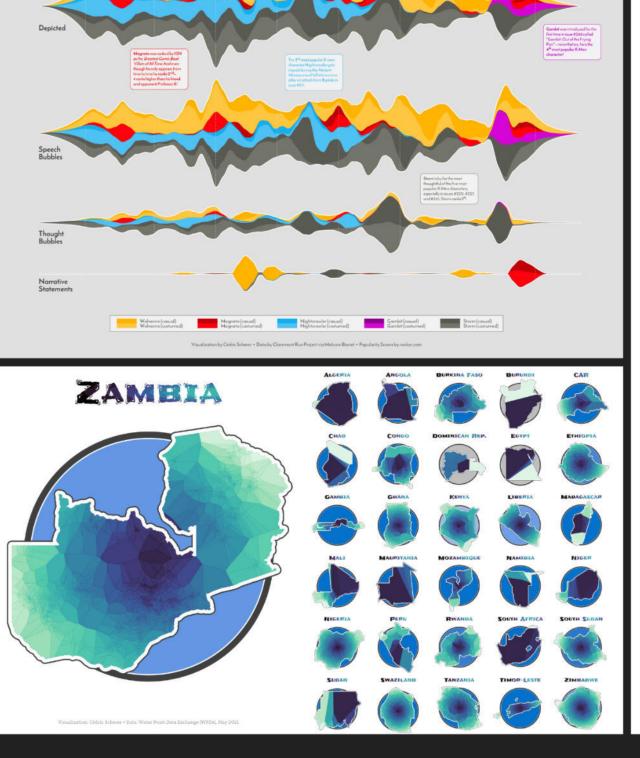








Appearance of the Five Most Popular X-Men Characters in Chris Claremont's Comics





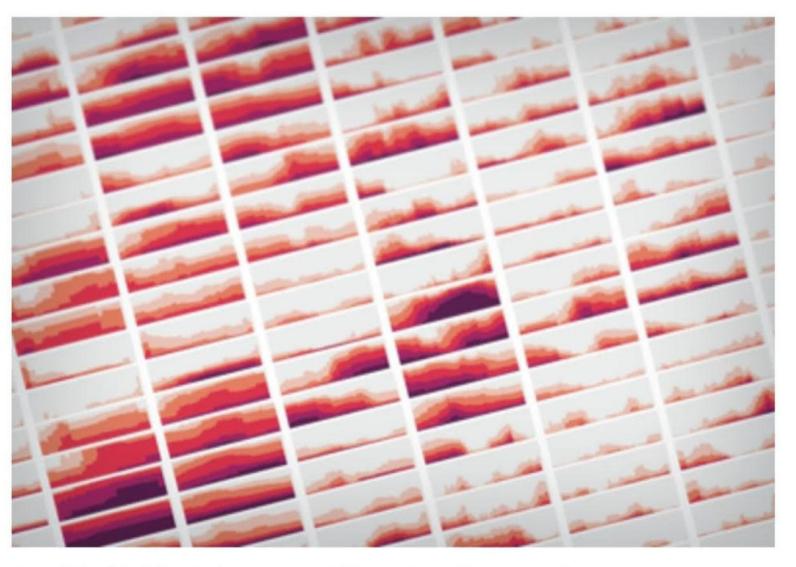


CLIMATE CHANGE

Climate Change Drives **Escalating Drought**

The past two decades have seen some of the most extreme dry periods in U.S. history

By Clara Moskowitz, Cédric Scherer, Georgios Karamanis | Scientific American November 2021 Issue

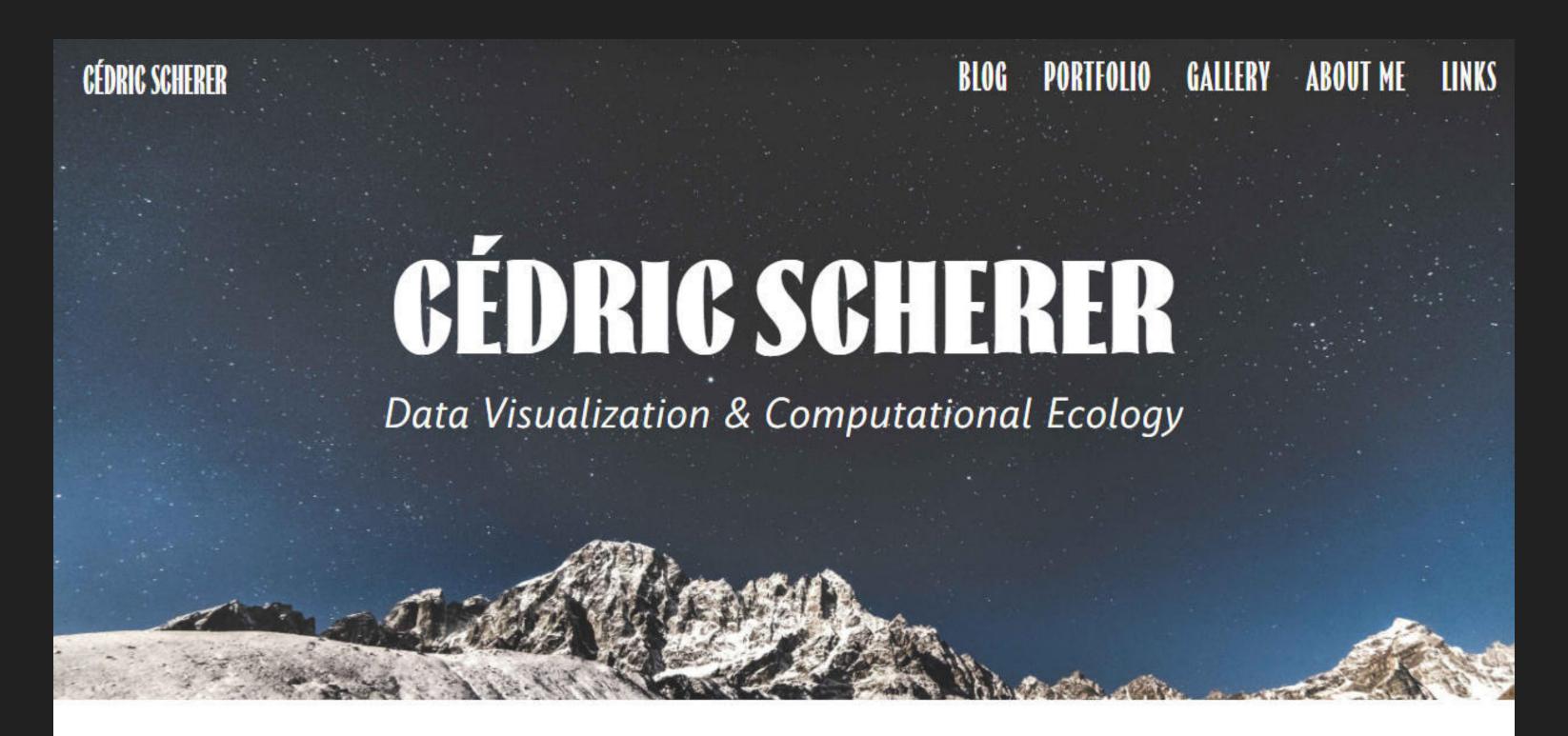


Credit: Cédric Scherer and Georgios Karamanis





* cedricscherer.com



The World's Countries Colored by Their First Letter

While preparing the mapping section for a Pearson-O'Reilly training, I got the idea to visualize the first letter of each country. And got especially curious about how much landmass each letter covers. Turns out: A, C and R are covering the largest areas!

Posted by Cédric · Friday, August 27, 2021

A Quick How-to on Labelling Bar Graphs in ggplot2

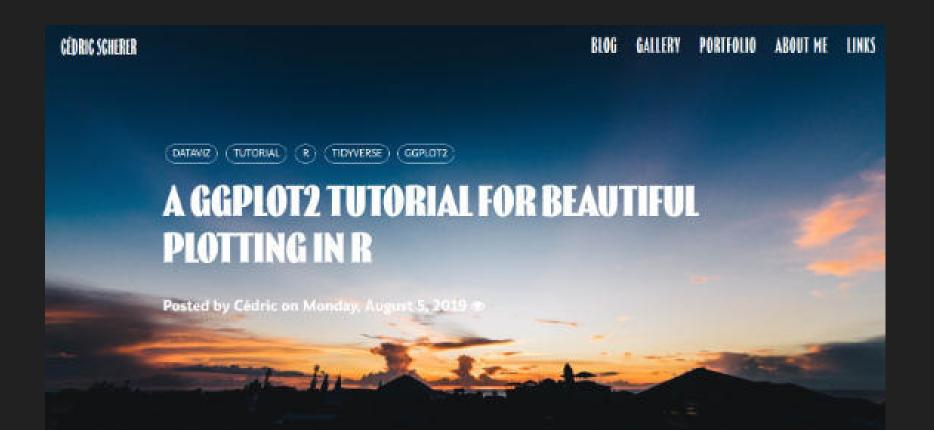
Bar charts are likely the most common chart type out there and come in several



Always coding. Passionate about design. Worried about nature.

Proud dad.

* cedricscherer.com



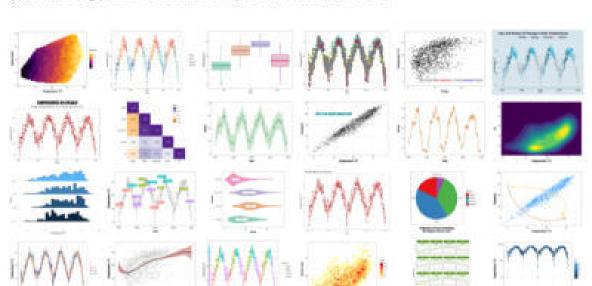
Last update: 2021-02-09

Introductory Words

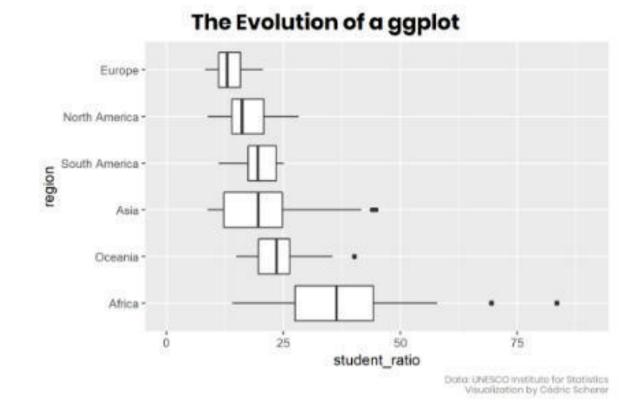
I don't care, just show me the content!

Back in 2016, I had to prepare my PhD introductory talk and I started using [ggplot2] to visualize my data. I never liked the syntax and style of base plots in R, so I was quickly in love with ggplot. Especially useful was its faceting utility. But because I was short on time, I plotted these figures by trial and error and with the help of lots of googling. The resource I came always back to was a blog entry called Beautiful plotting in R: A ggplot2 cheatsheet by Zev Ross, updated last in January 2016. After giving the talk which contained some decent plots thanks to the blog post, I decided to go through this tutorial step-by-step. I learned so much from it and directly started modifying the codes and over the time I added additional code snippets, chart types and resources.

Since the blog entry by Zev Ross was not updated for some years and step by step this became a unique version of a tutorial, I decided to host the updated version on my GitHub. Now it finds its proper place on this homepage! (Plus I added a ton of other updates—just to name a few: The fantastic {patchwork}. [ggtest] and [ggforce] packages. How to deal with custom fonts and colors. A collection of R packages tailored to create interactive charts. And several other chart types including pie charts because everyone looooves pie charts!)







- Aim of this Tutorial
- Data Preparation
- The Default Boxplot
- Sort Your Data!
- V Let Your Plot Shine-Get Rid of the Default Settings
- . The Choice of the Chart Type
- More Geoms, More Fun, More Info!
- Add Text Boxes to Let The Plot Speak for Itself
- Bonus: Add a Tile Map as Legend
- A The Final Evolved Visualization
- E Complete Code for Final Plot
- Post Scriptum: Mean versus Median

Aim of this Tutorial

In this series of blog posts, I aim to show you how to turn a default ggplot into a plot that visualizes information in an appealing and easily understandable way. The goal of each blog post is to provide a step-by-step tutorial explaining how my visualization have evolved from a typical

ist die grafische Darstellung von Informationen und Daten.





wandelt Daten in visuelle Formen als quantifizierbare Merkmale um.





hilft Einsichten zu gewinnen, zu entdecken, erklären und zu entscheiden.





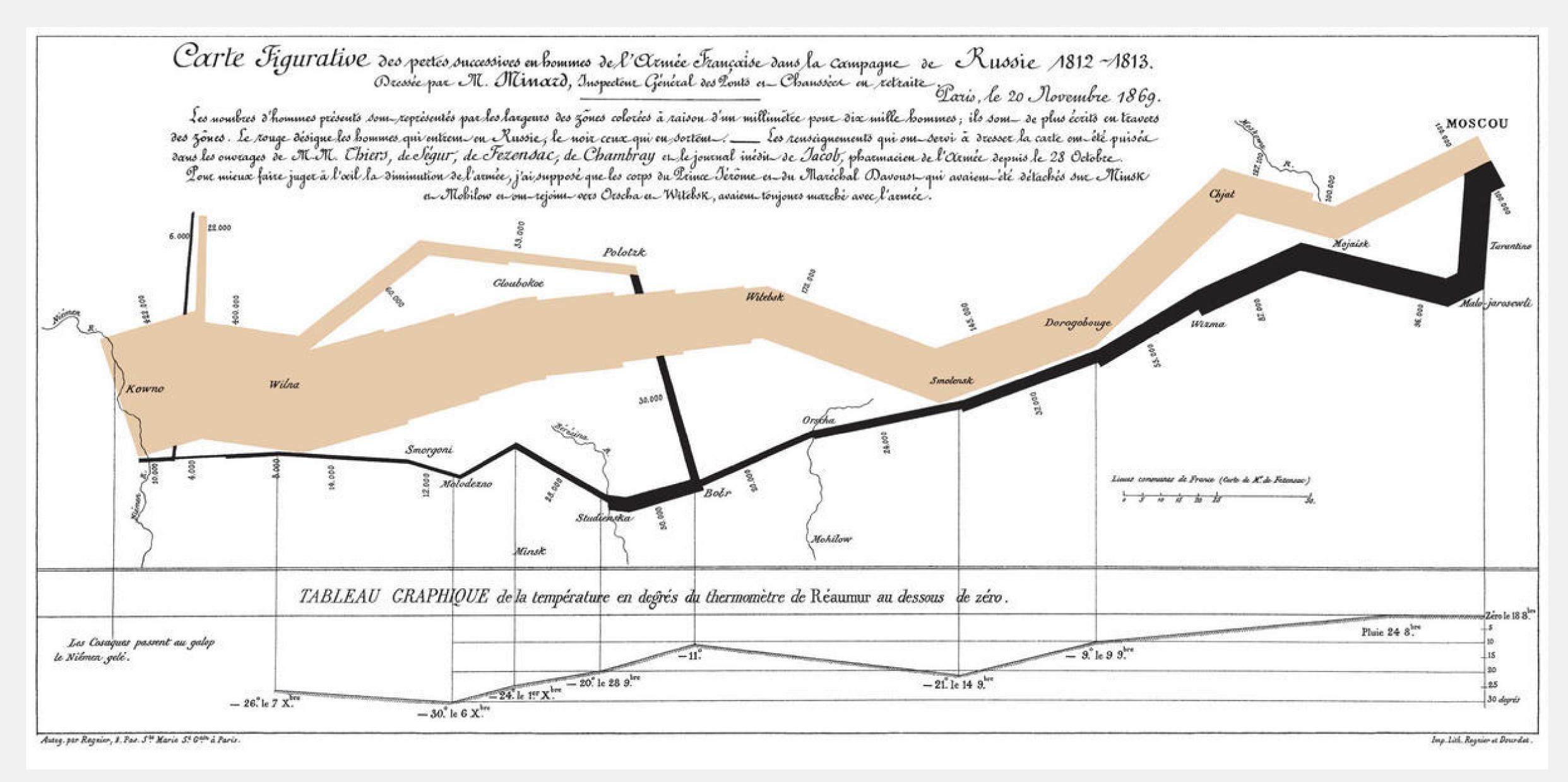
ist einerseits Kunst und andererseits Wissenschaft.



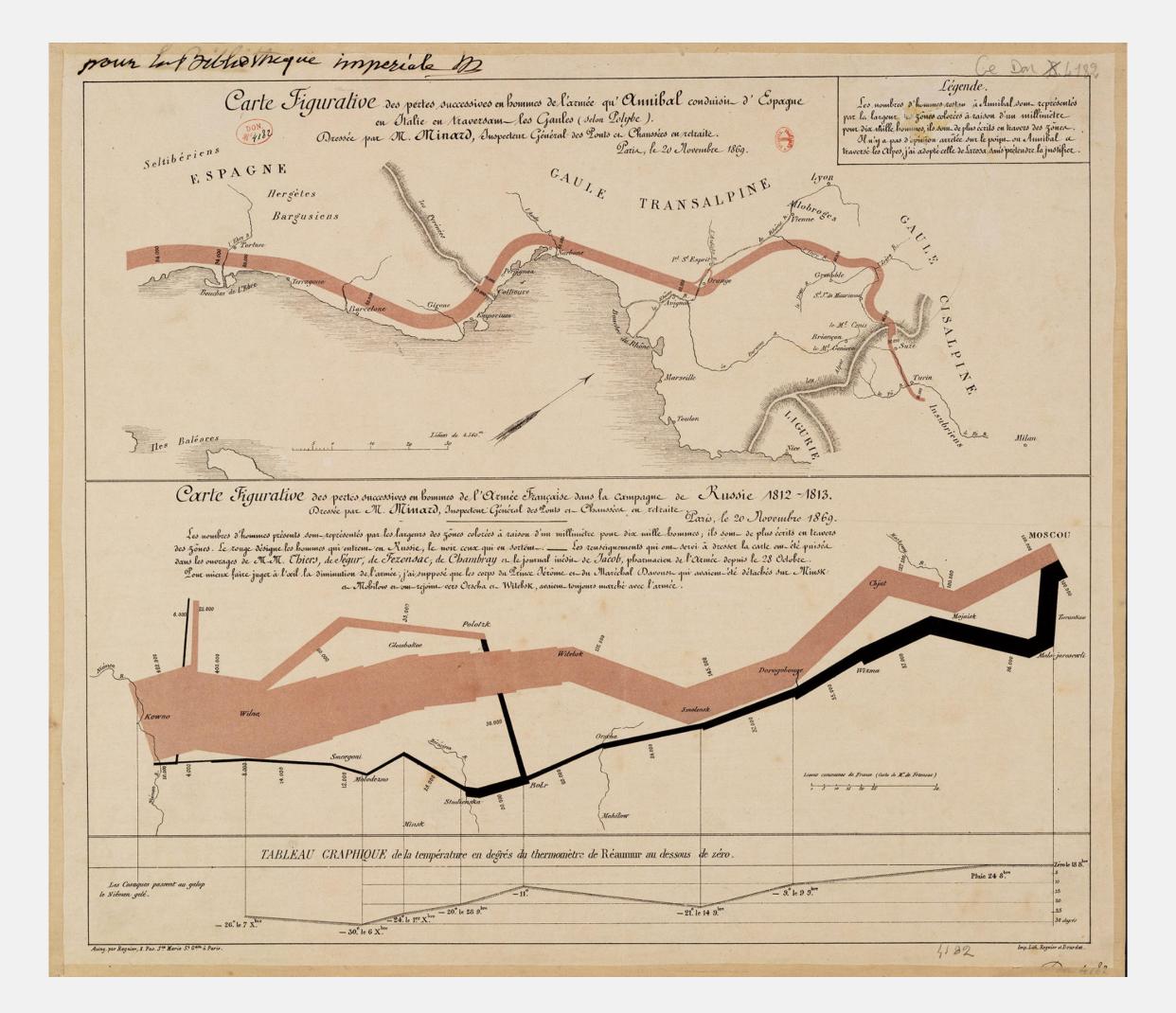




Quelle: eazybi.com



Carte figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812–1813 von Charles Joseph Minard



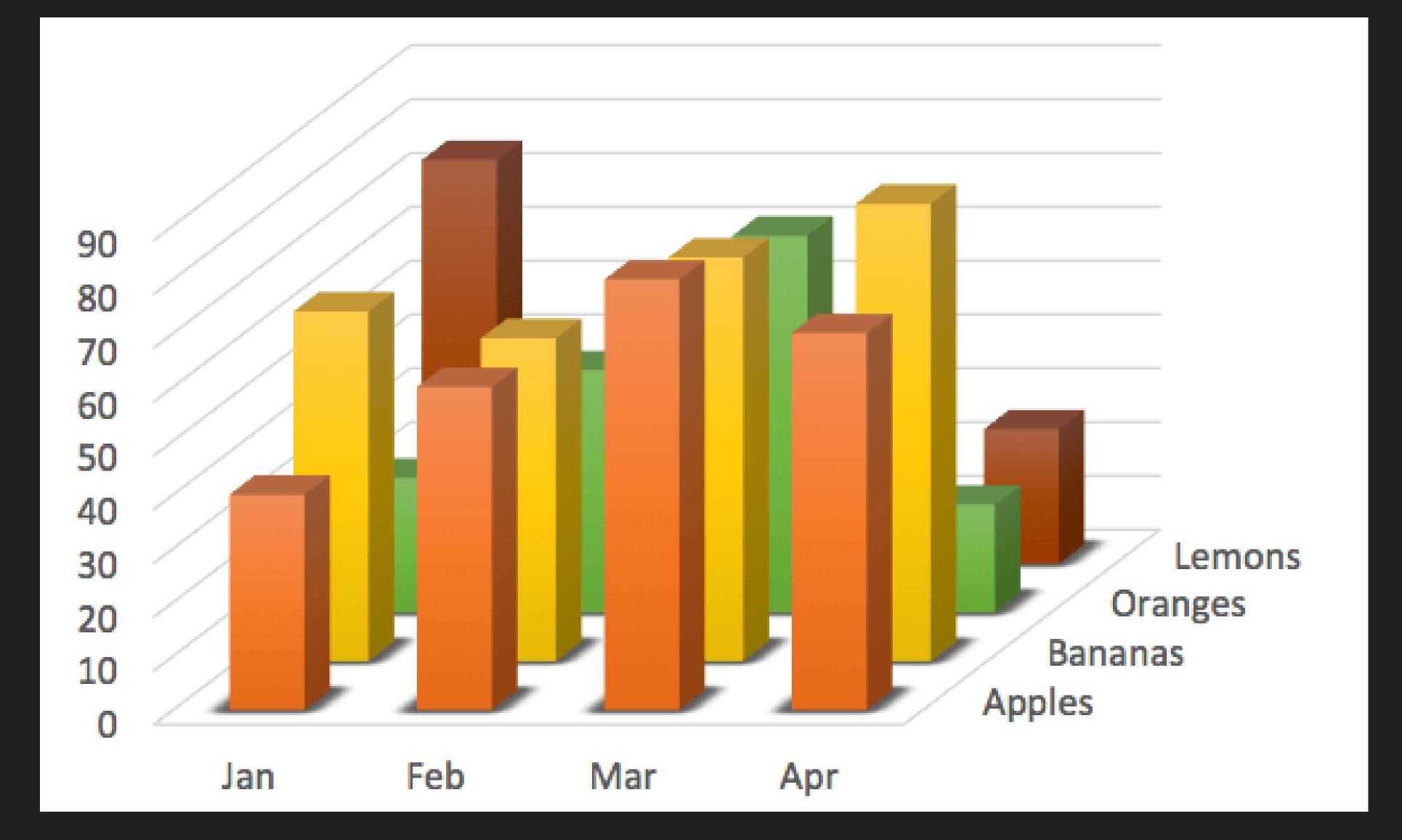
Carte figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812–1813 and Carte figurative des pertes successives en hommes de l'Armée qu'Annibal conduisit d'Espagne en Italie en traversant les Gaules (selon Polybe) von Charles Joseph Minard

- zeigt das Vorrücken der Truppen von **Hannibal** (218 v. Chr.) und **Napoleon** (1812-1813)
- · wird oft als die beste jemals gezeichnete statistische Grafik bezeichnet





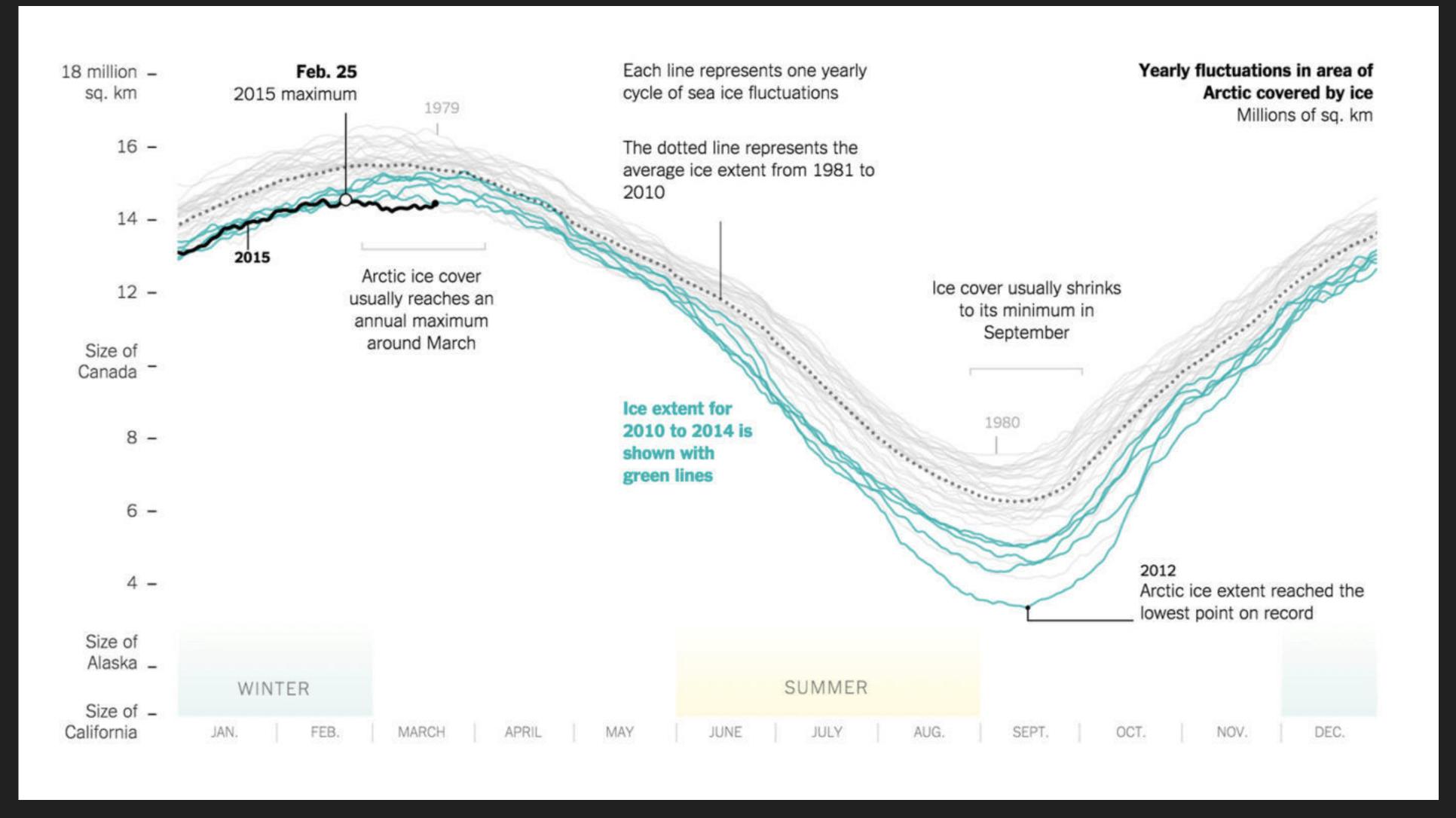
Was macht es zu einer schlechten Grafik?



Was macht es zu einer schlechten Grafik

- substantive Probleme (schlechte Daten)
- aesthetische Probleme (schlechtes Design)
- perzeptionelle Probleme (schlechte Codierung)

Was macht es zu einer guten Grafik?



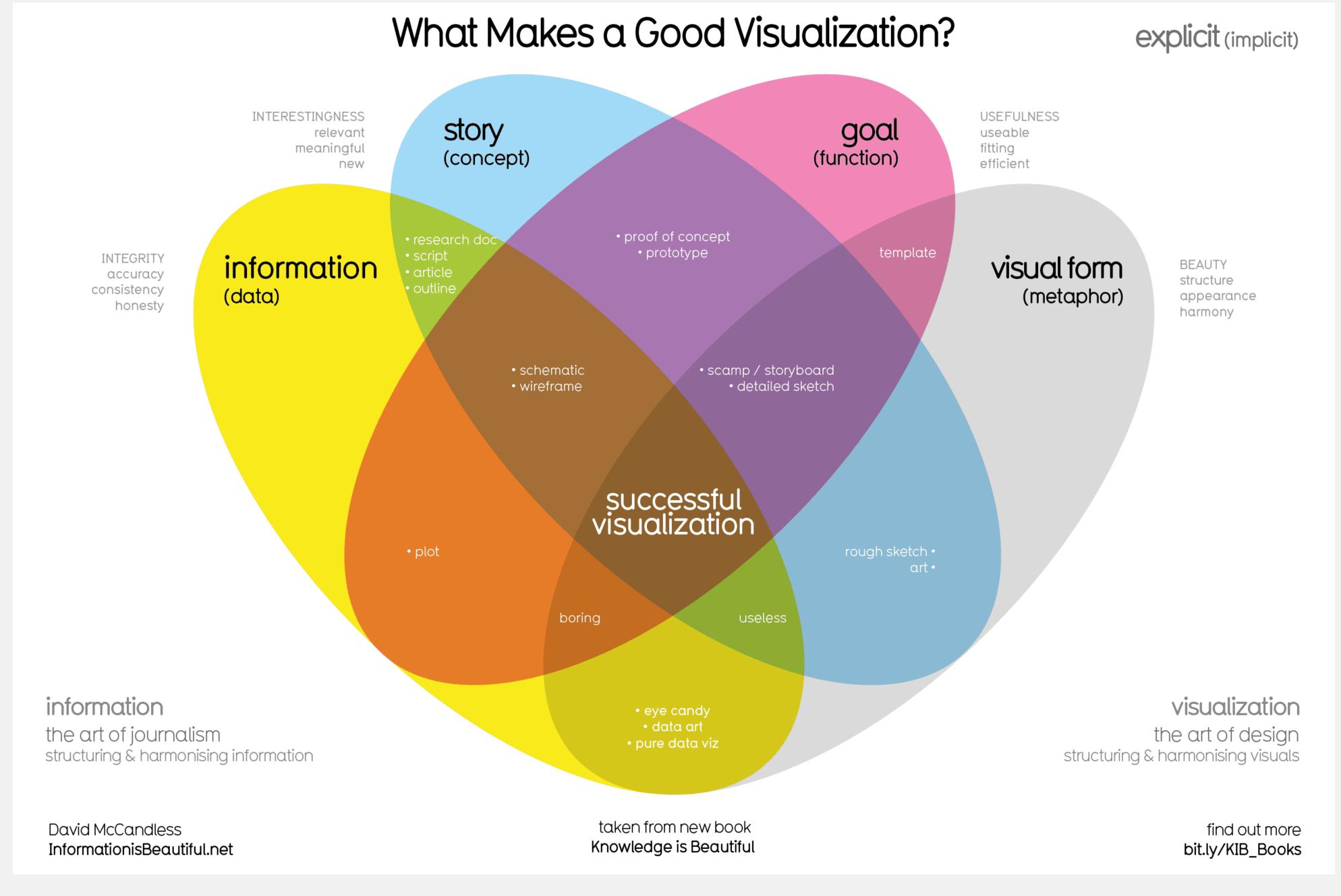






Was macht es zu einer guten Grafik

- INFORMATION (Korrektheit)
- STORY (Bedeutsamkeit)
- GOAL (Zweckmäßigkeit)
- FORM (Schönheit)







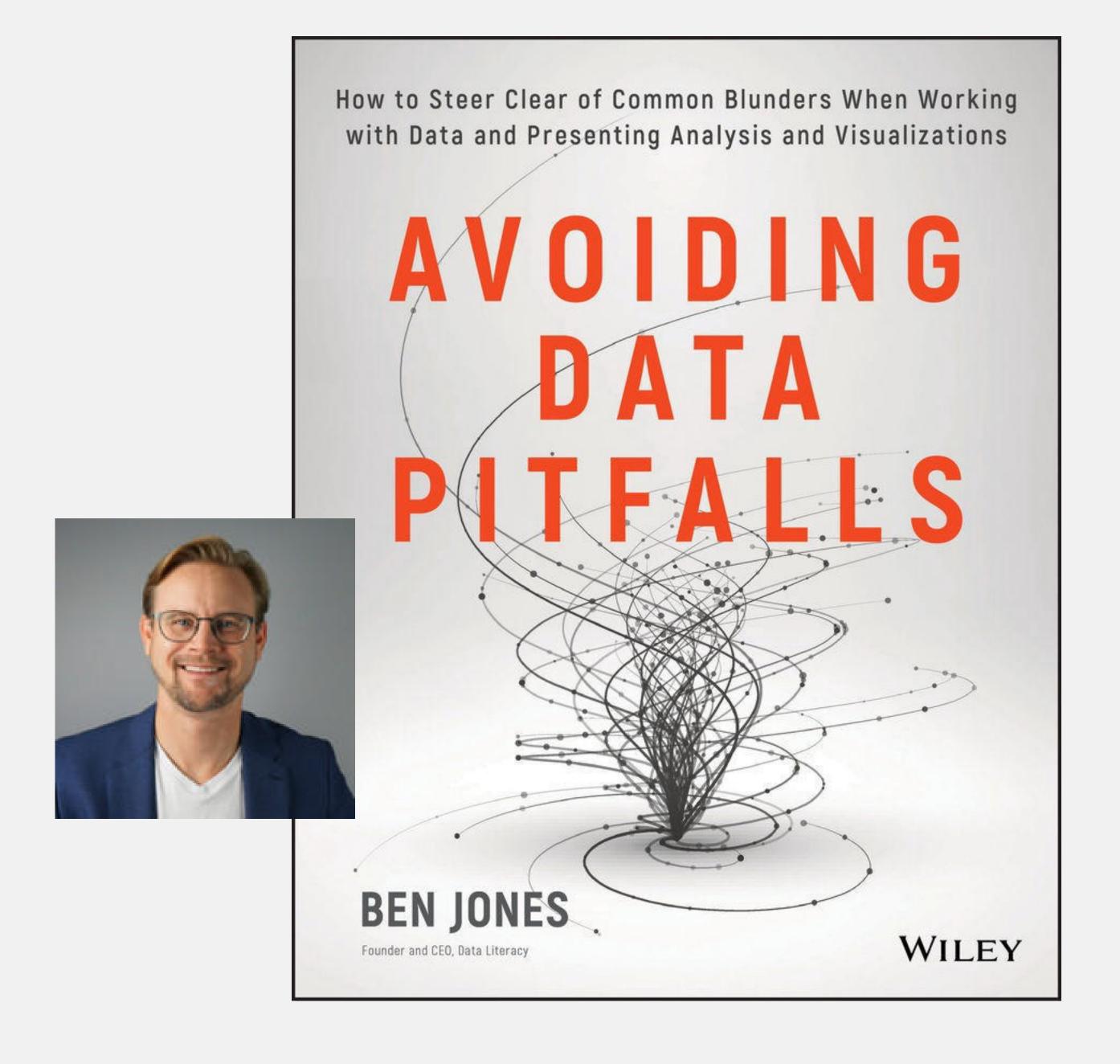


INFORMATION

Verstehe deine Daten und sei genau













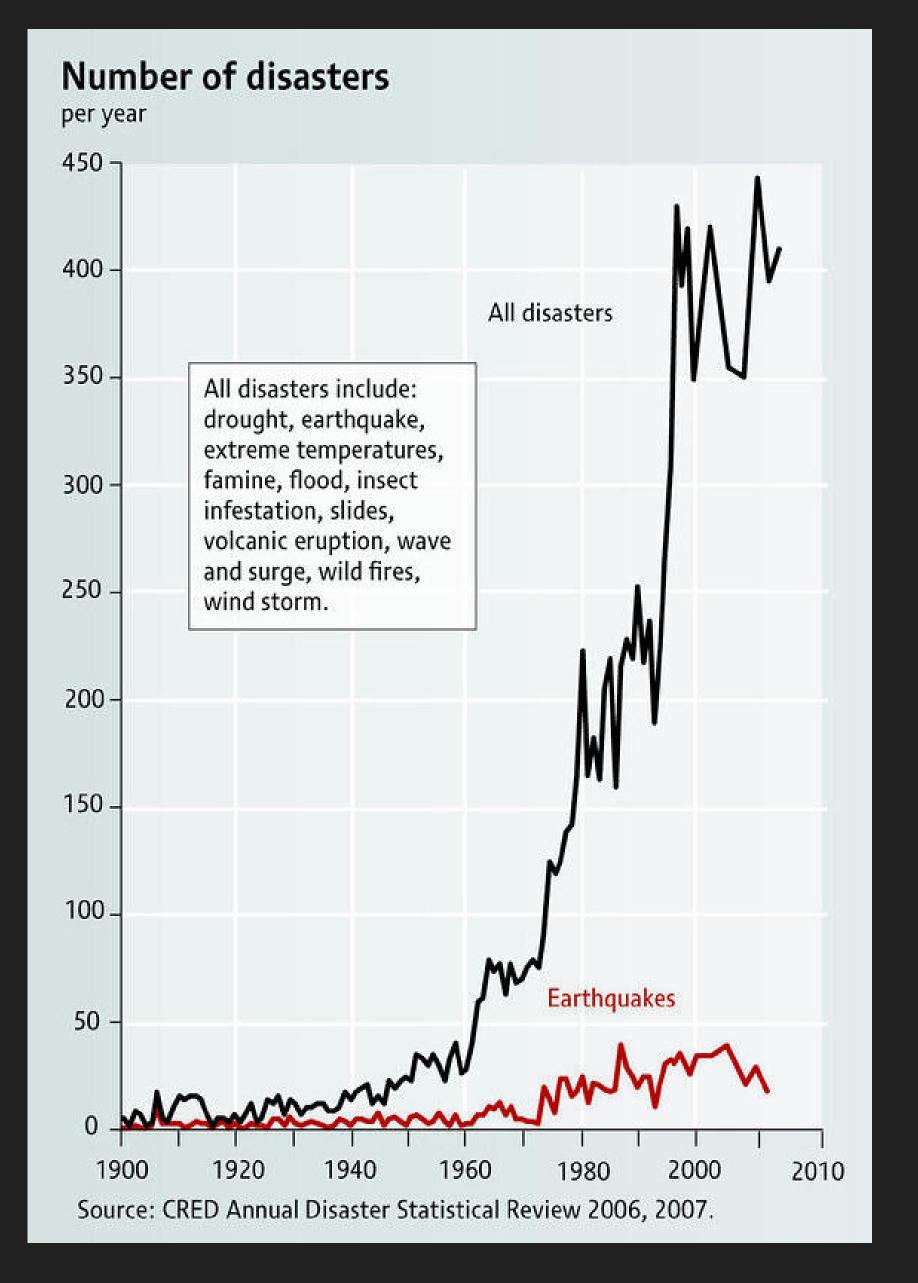
Unsere Daten sind nie ein perfektes Abbild der realen Welt.

- → nur eine Teilmenge: gemeldete Verbrechen
- → von Menschen erhoben: Schätzungen, Genauigkeit & Fehler
- → maschinell erfasst: Genauigkeiten & Fehler





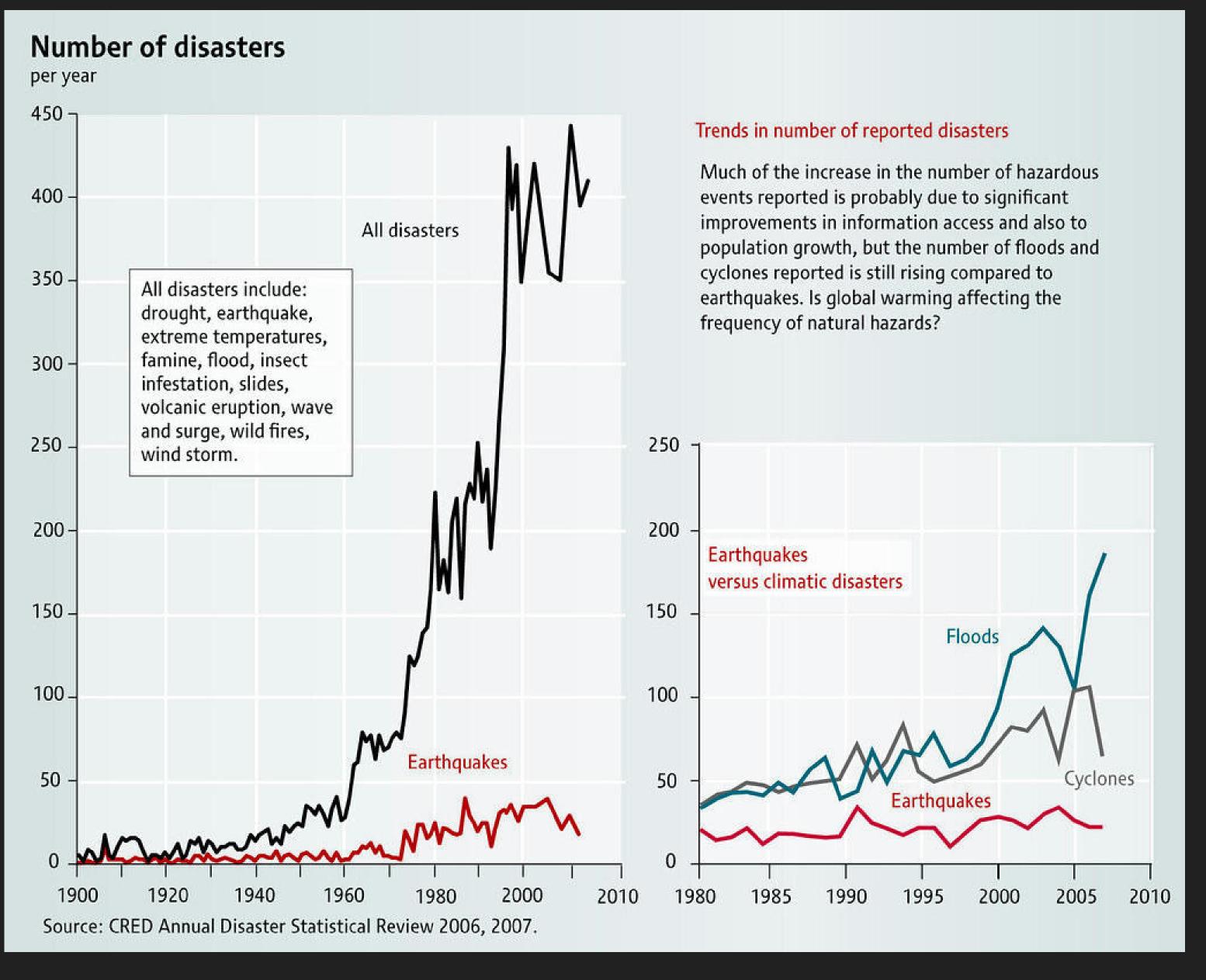








"Ein Großteil des Anstiegs der gemeldeten gefährlichen Ereignisse ist wahrscheinlich auf den erheblich verbesserten Zugang zu Informationen zurückzuführen."









Der beste Nutzen von Daten besteht darin uns zu lehren, was nicht wahr ist.





Der beste Nutzen von Daten besteht darin uns zu lehren, was nicht wahr ist.

- → keine Einzelaussagen formulieren
- → sondern falsifizierbare universelle Aussagen





Der beste Nutzen von Daten besteht darin uns zu lehren, was nicht wahr ist.

- → "Der Schwan ist weiß."
- → "Alle Schwäne sind weiß."





Sei dir über die Botschaft der Daten im Klaren





Welche Geschichte ist für sie interessant?





Welche Geschichte ist für sie interessant?

Was sind wirklich relevante Details?





Welche Geschichte ist für sie interessant?

Was sind wirklich relevante Details?

Welche Variablen sind für sie bedeutsam?







Welche Geschichte ist für sie interessant?

Was sind wirklich relevante Details?

Welche Variablen sind für sie bedeutsam?

Wie werden sie der Grafik begegnen?







Welche Geschichte ist für sie interessant?

Was sind wirklich relevante Details?

Welche Variablen sind für sie bedeutsam?

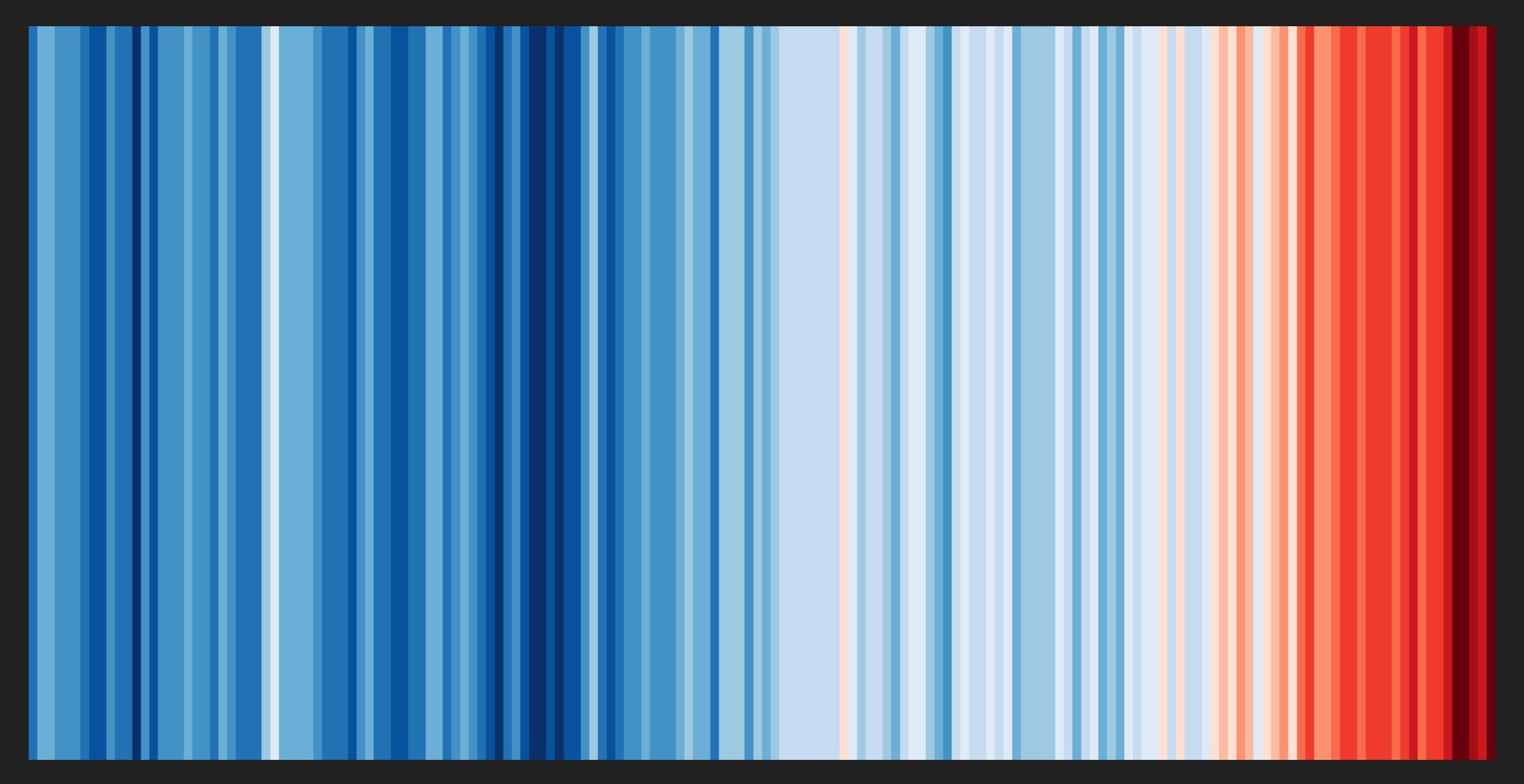
Wie werden sie der Grafik begegnen?

Brauche ich überhaupt eine Visualisierung??





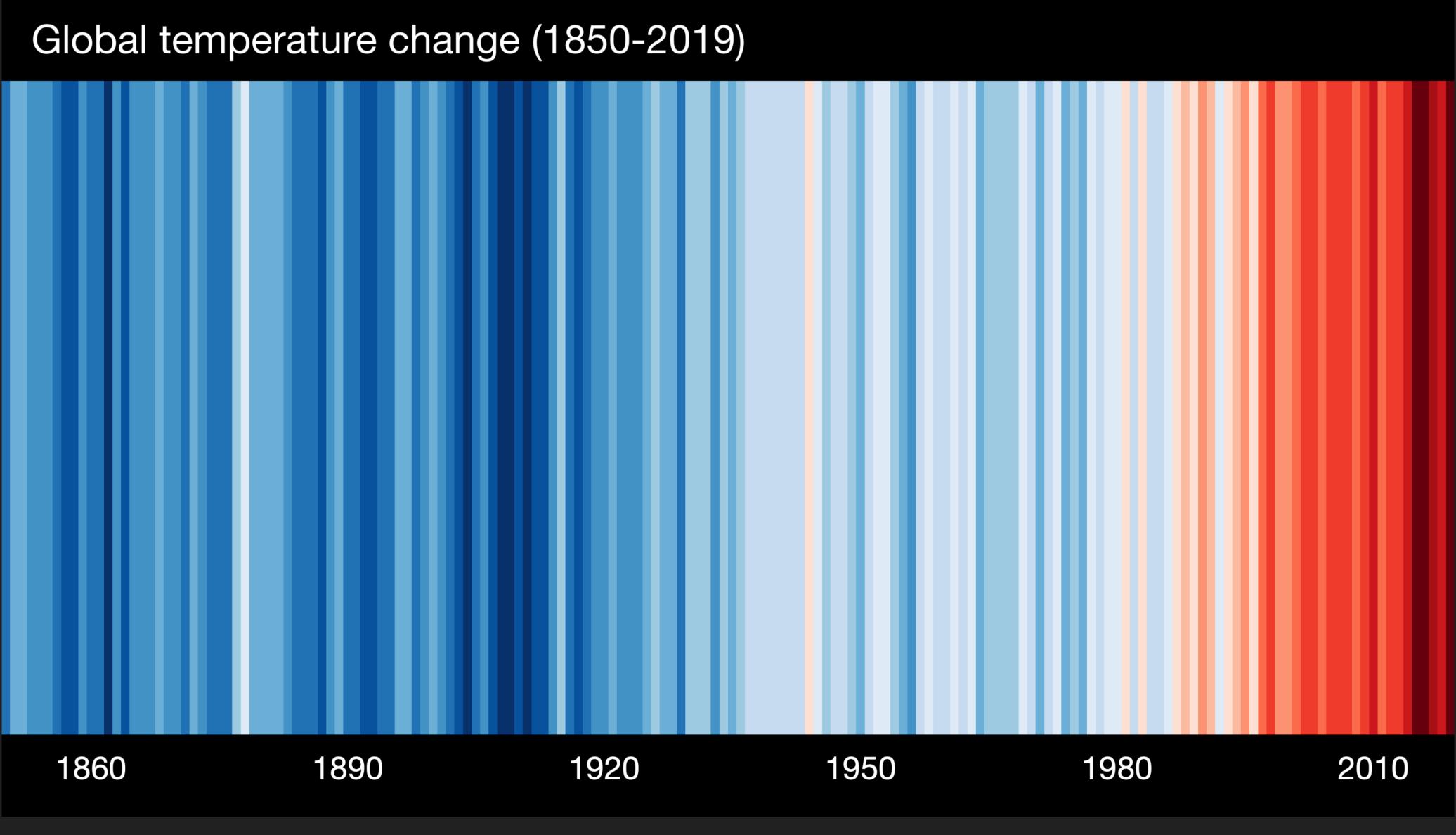




<u>Warming Stripes</u> von Ed Hawkins





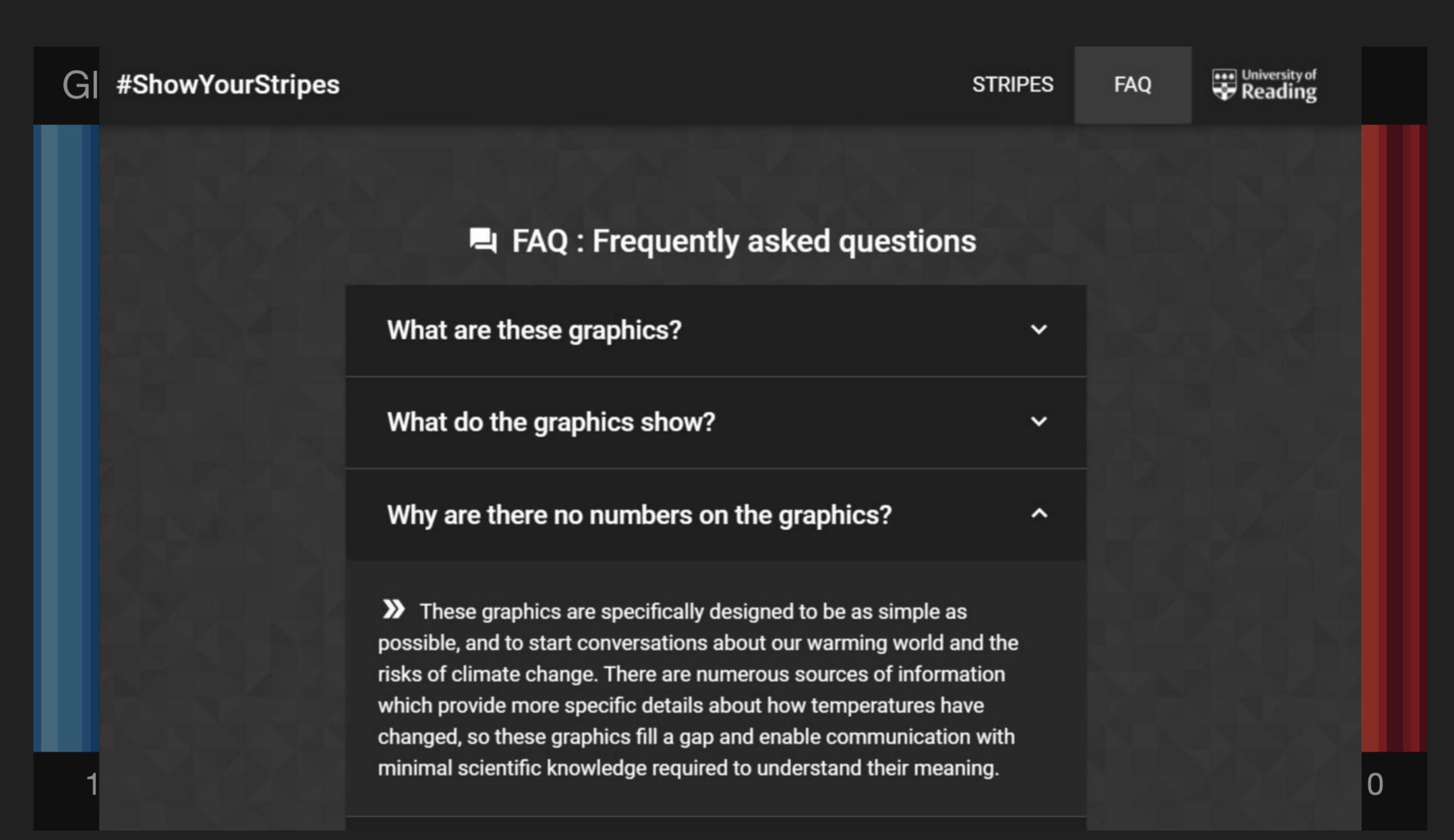


<u>Warming Stripes</u> von Ed Hawkins









showyourstripes.info/faq







Diese Grafiken sind speziell dafür konzipiert um Gespräche anzuregen über die sich erwärmende Welt und die Risiken des Klimawandels.

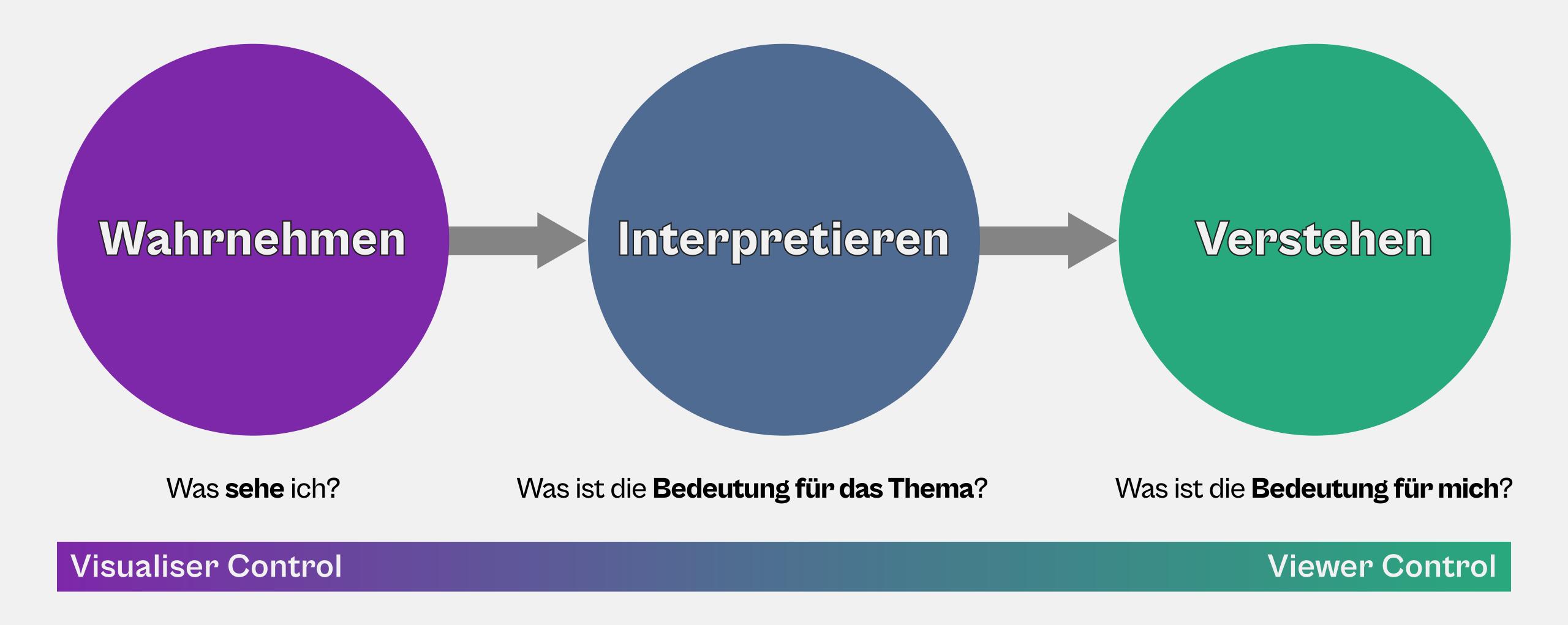
These graphics are specifically designed to be as simple as possible, and to start conversations about our warming world and the risks of climate change. There are numerous sources of information which provide more specific details about how temperatures have changed, so these graphics fill a gap and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

showyourstripes.info/faq







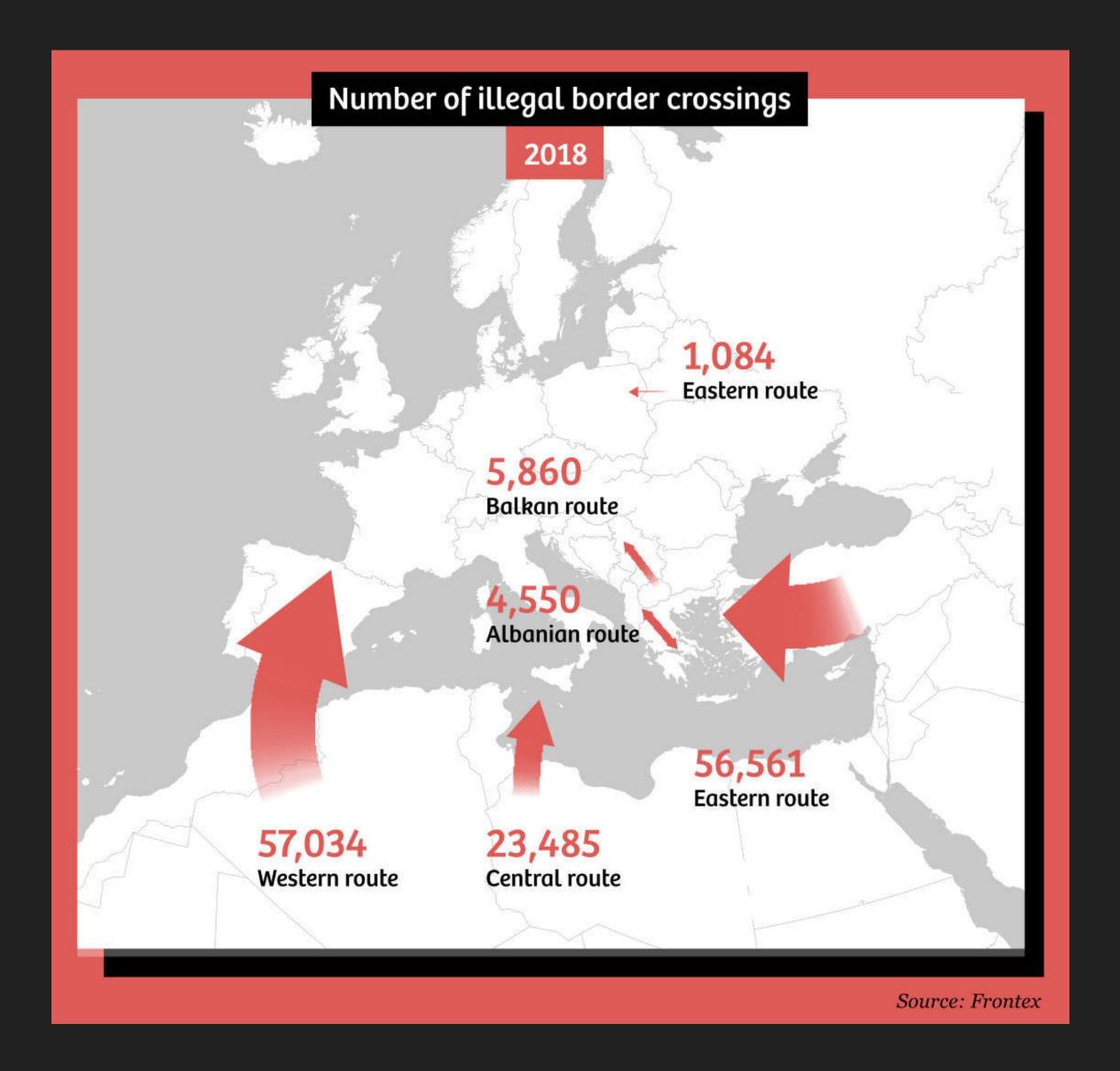


Schema von Andy Kirk



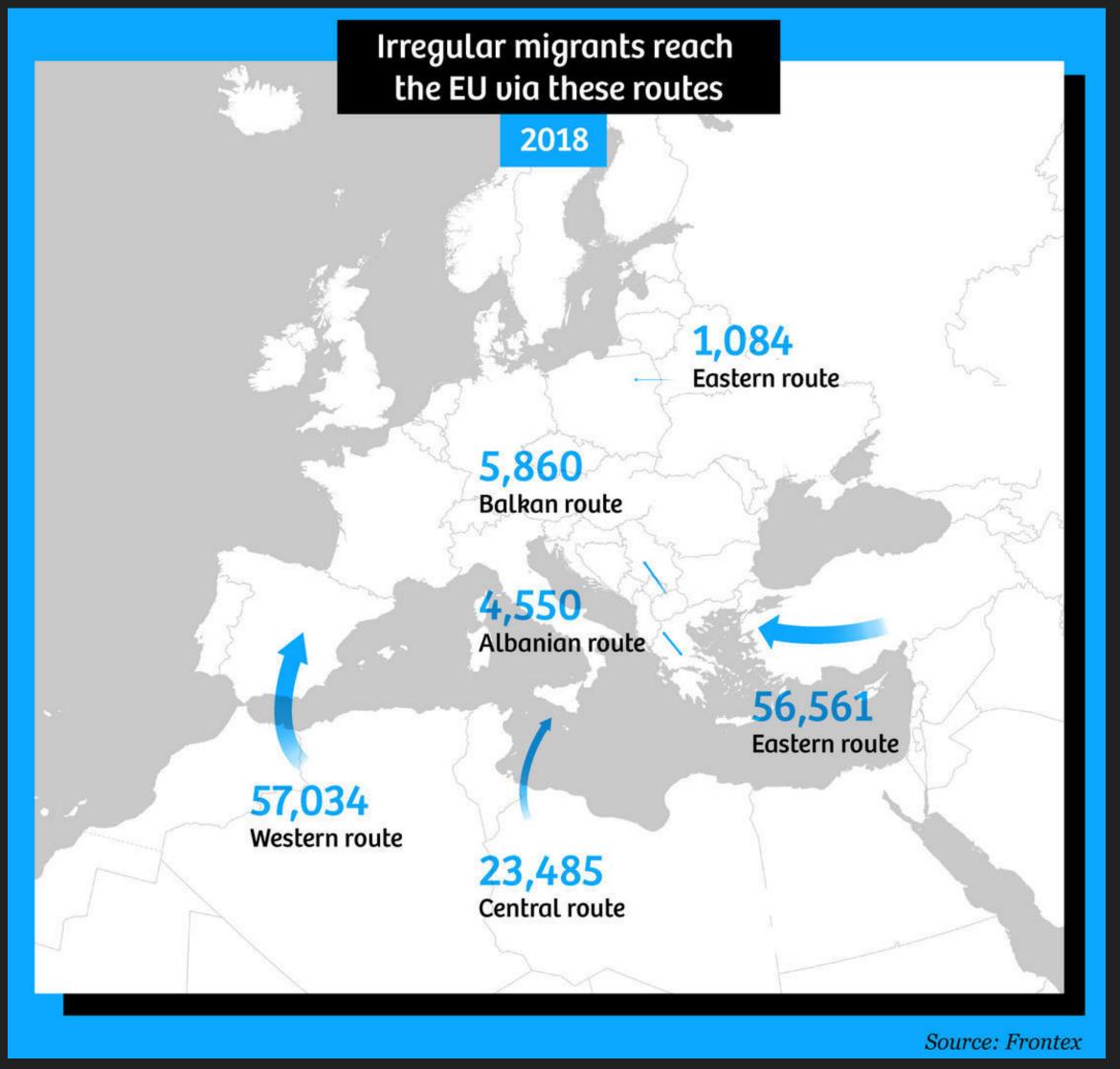






"How maps in the media make us more negative about migrants" by Maite Vermeulen, Leon de Korte & Henk van Houtum

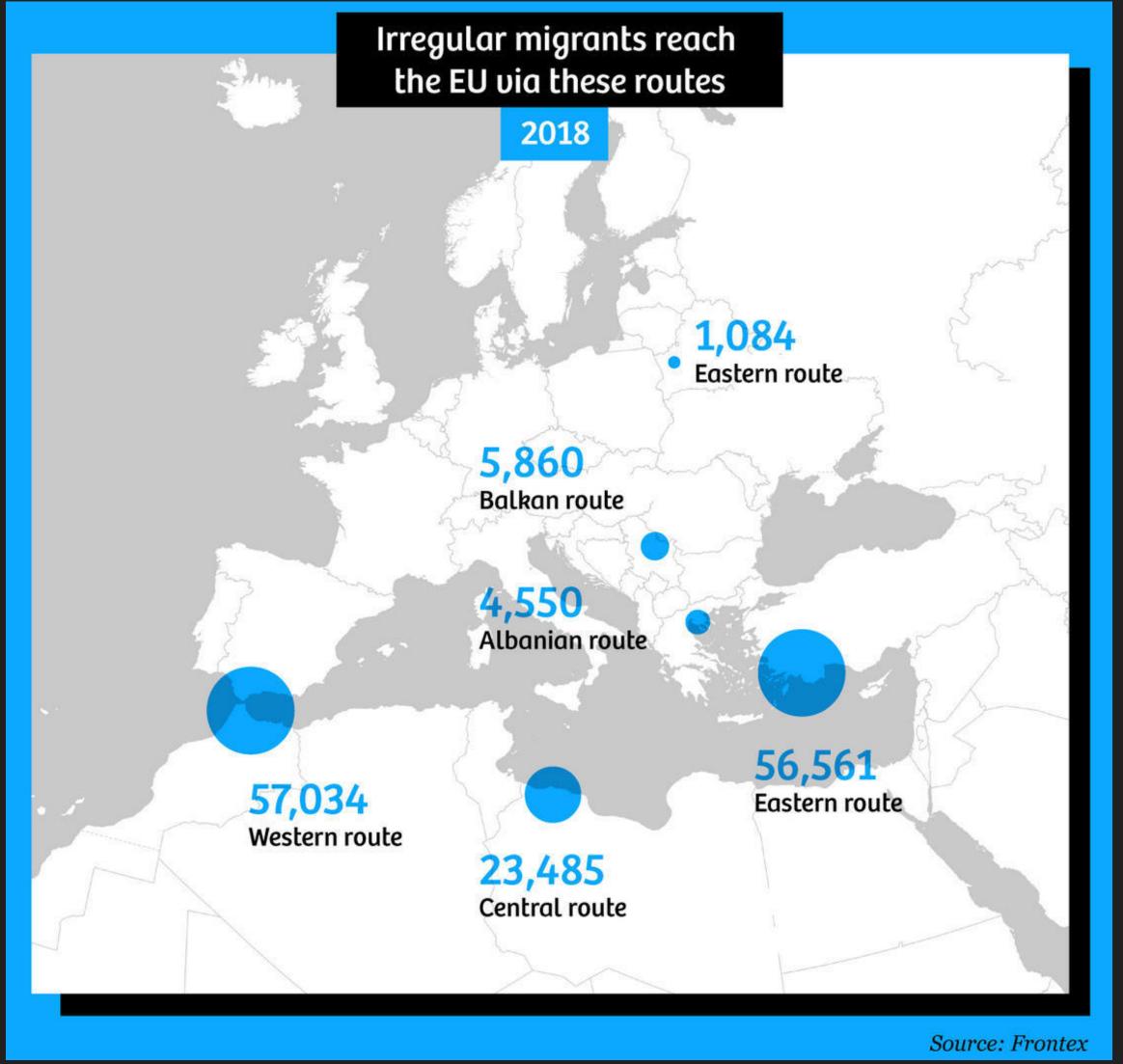






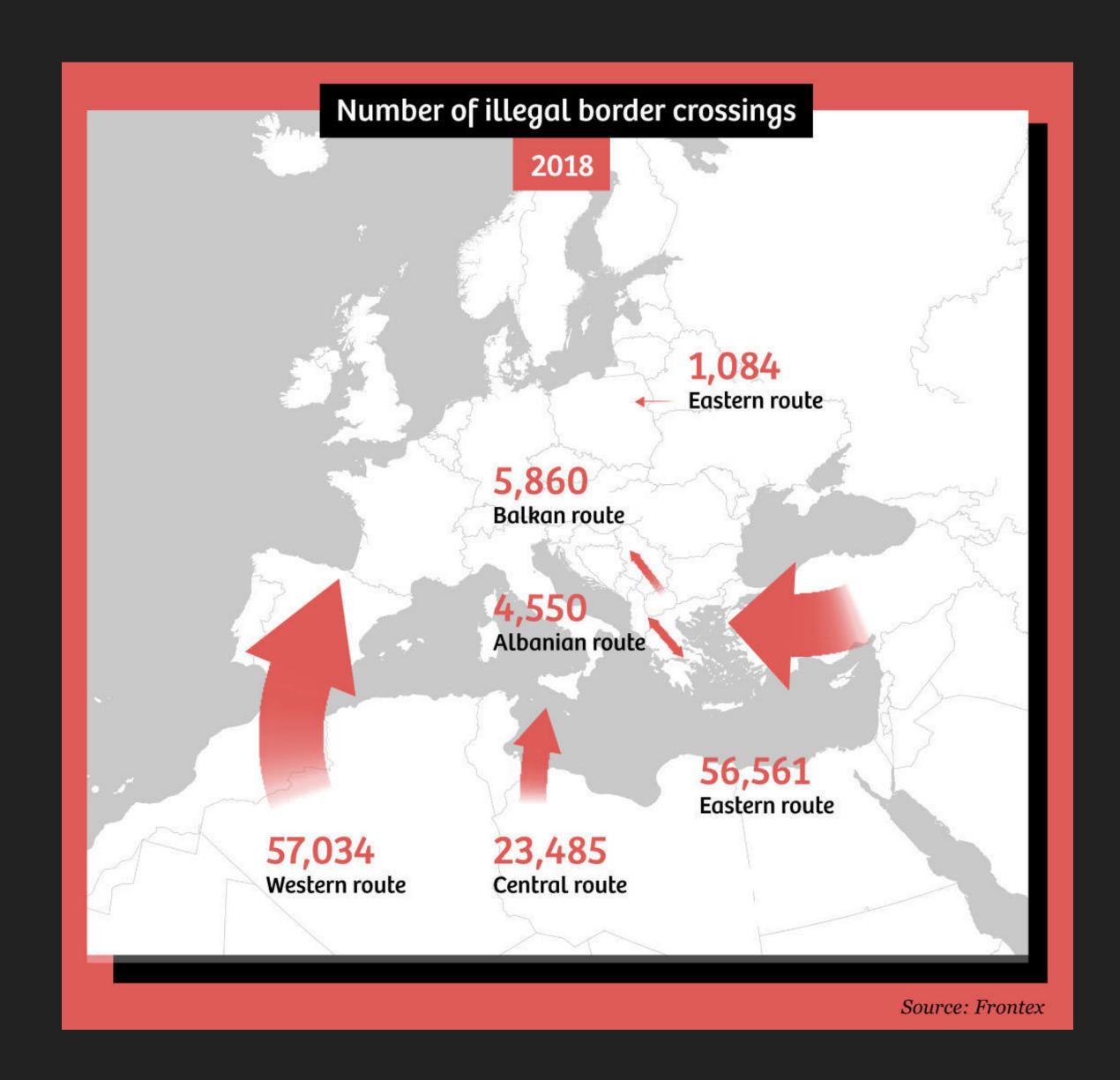


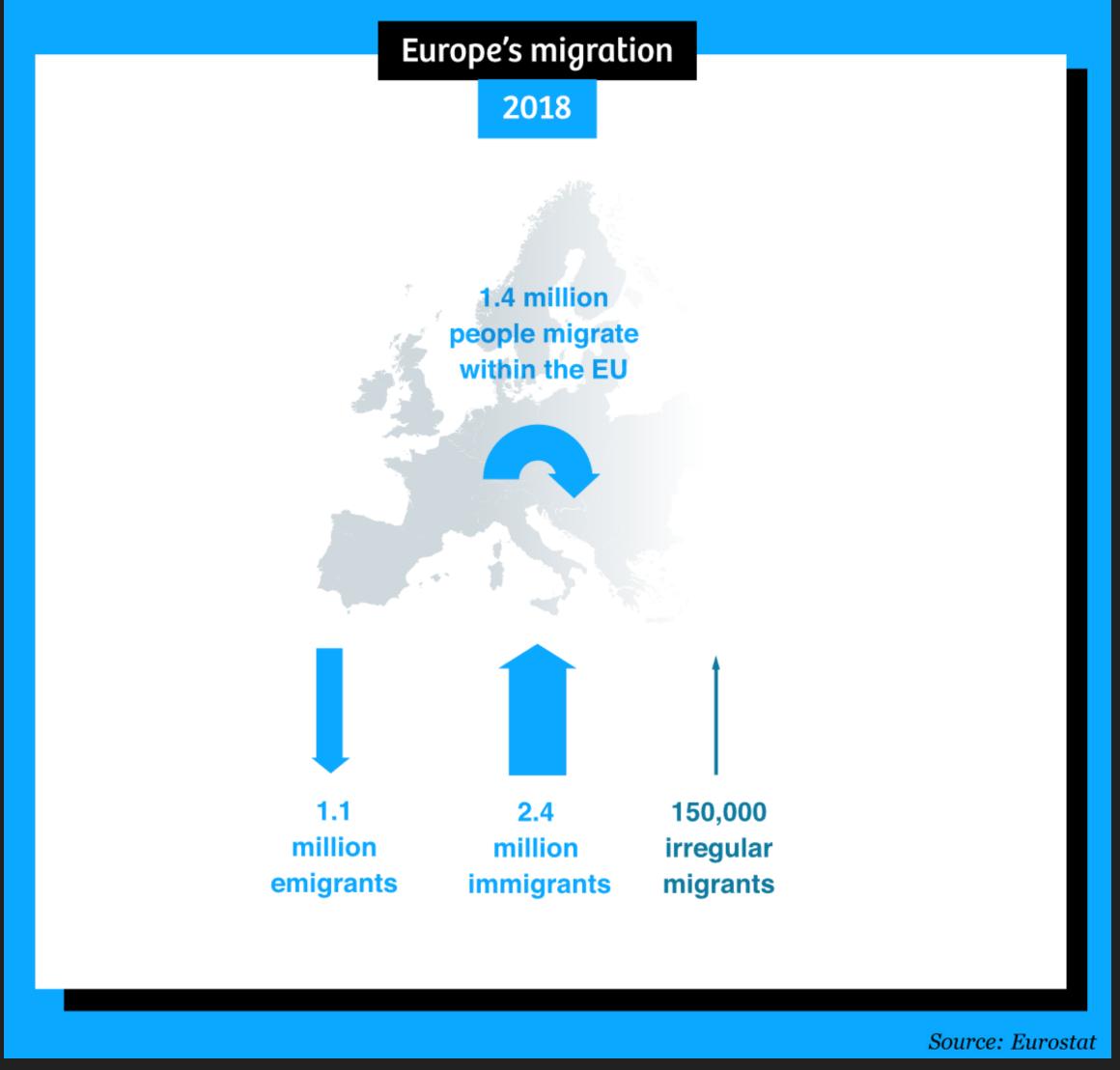






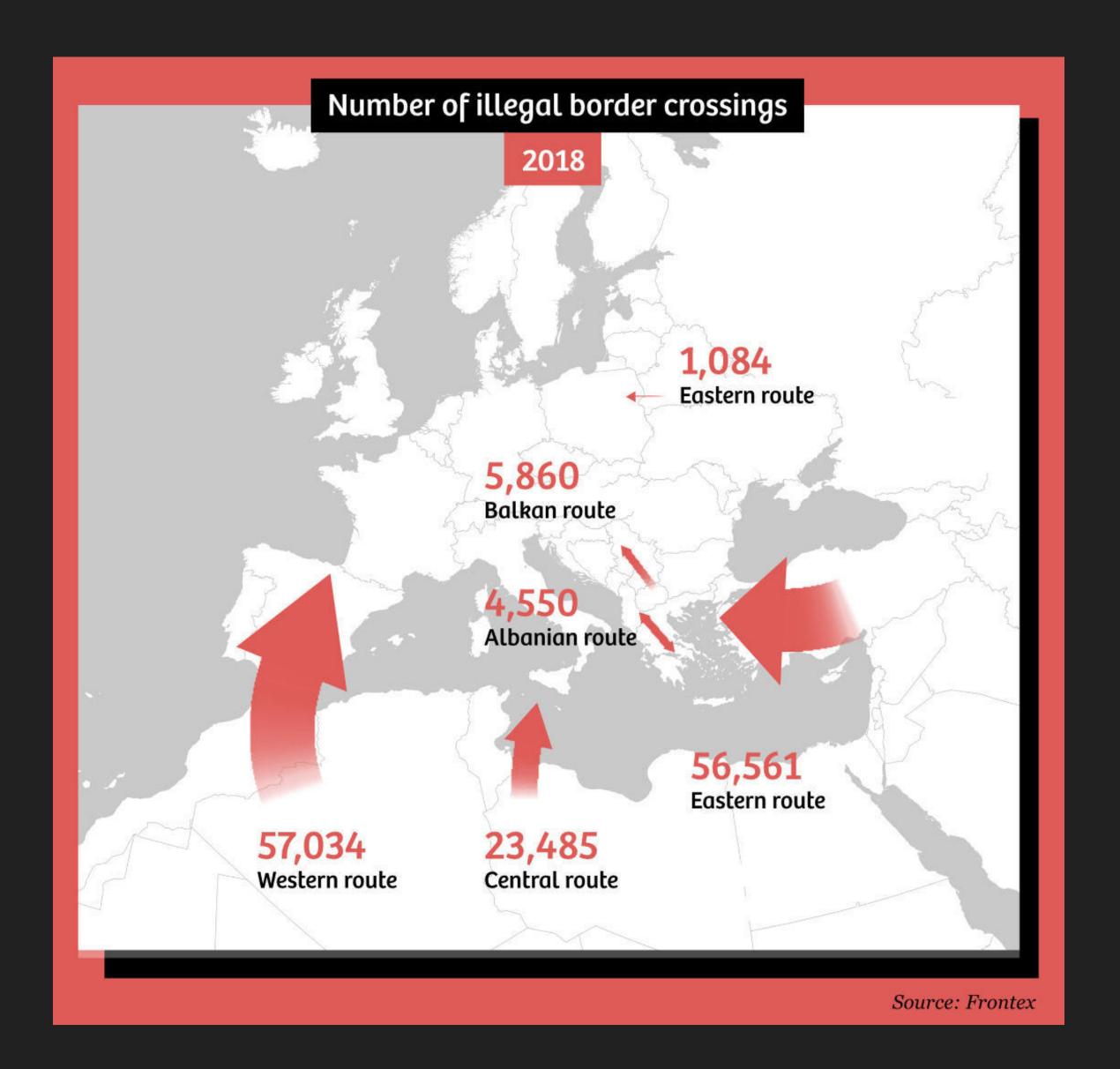


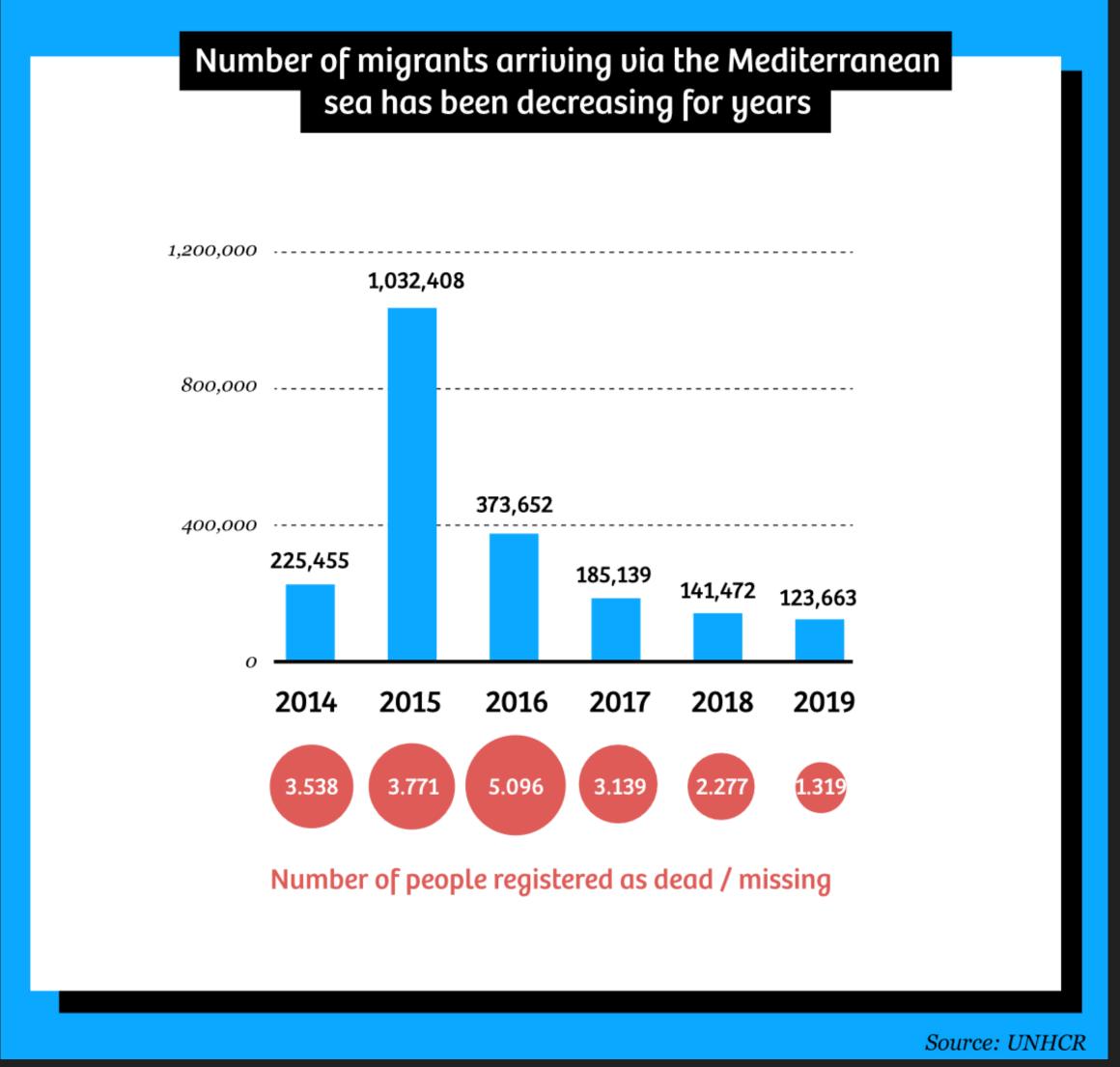
















GOAL

Wähle geeignete Grafiken, um die Geschichte zu erzählen





Typologie von Informationsgrafiken

nach Juuso Koponen & Jonatan Hildén, "Data Visualization Handbook" (2020), Seite 25

Sind die Informationen konzeptionell oder messbar?

Art der Information: Darstellung von Konzepten <> Umwandlung von Daten in visuelle Formen

Ist der **Zweck** Informationen zu erkunden oder zu erklären?

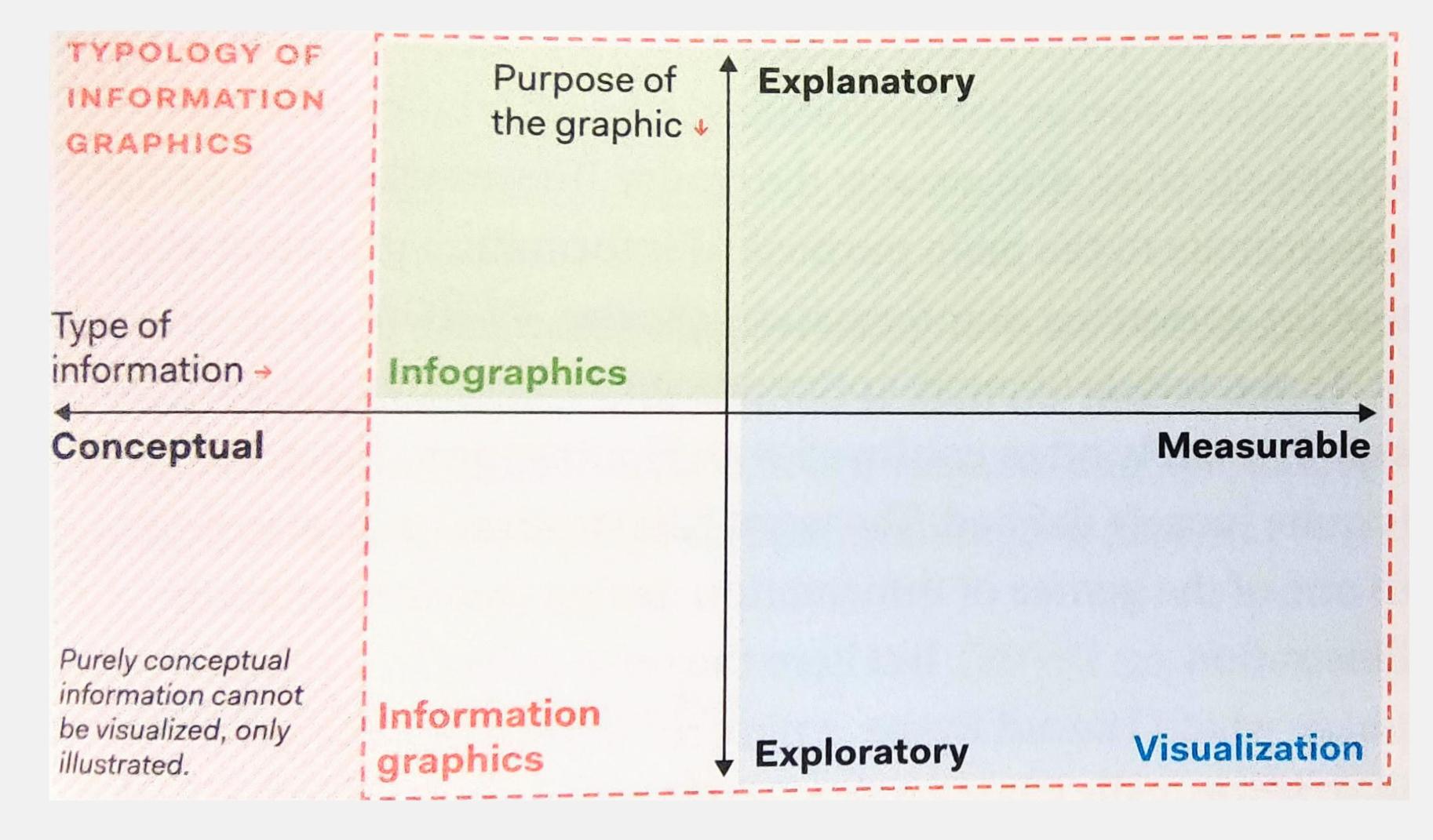
Zweck der Grafik: Erleichterung der Entdeckung <> Vermittlung von Informationen





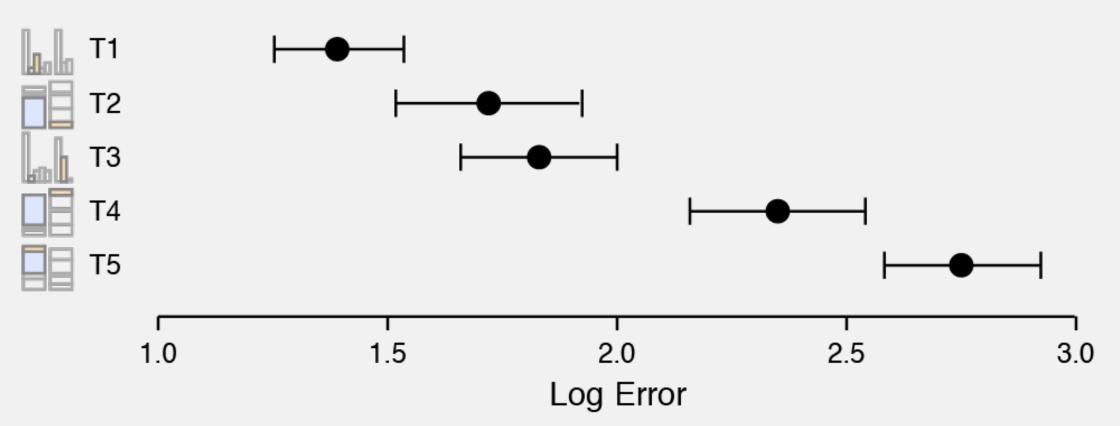
Typologie von Informationsgrafiken

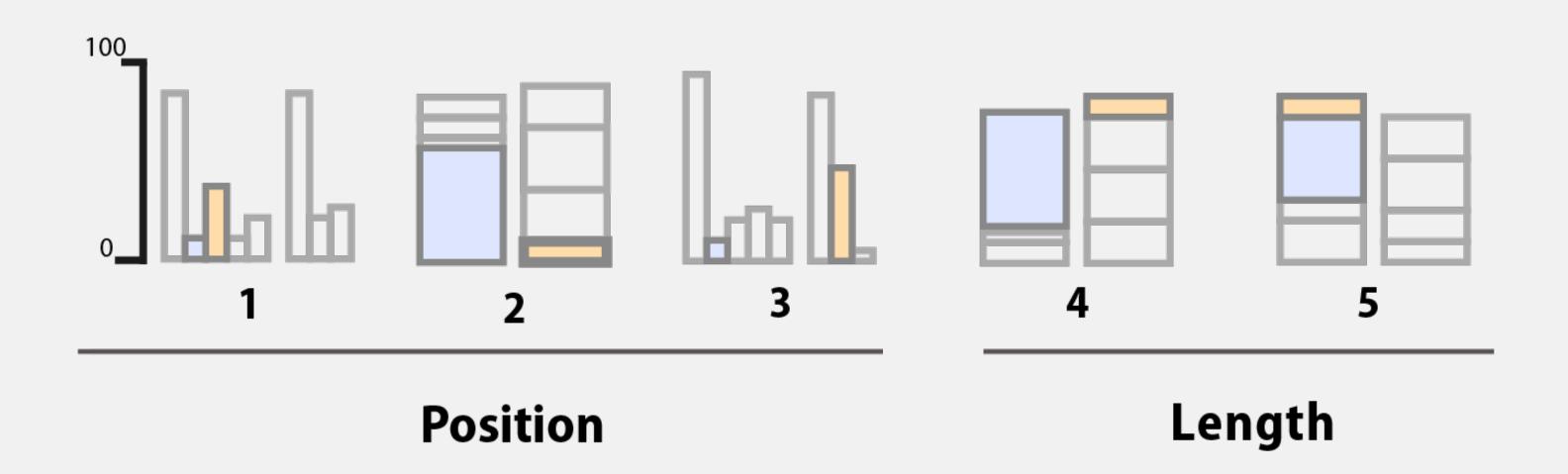
nach Juuso Koponen & Jonatan Hildén, "Data Visualization Handbook" (2020), Seite 25





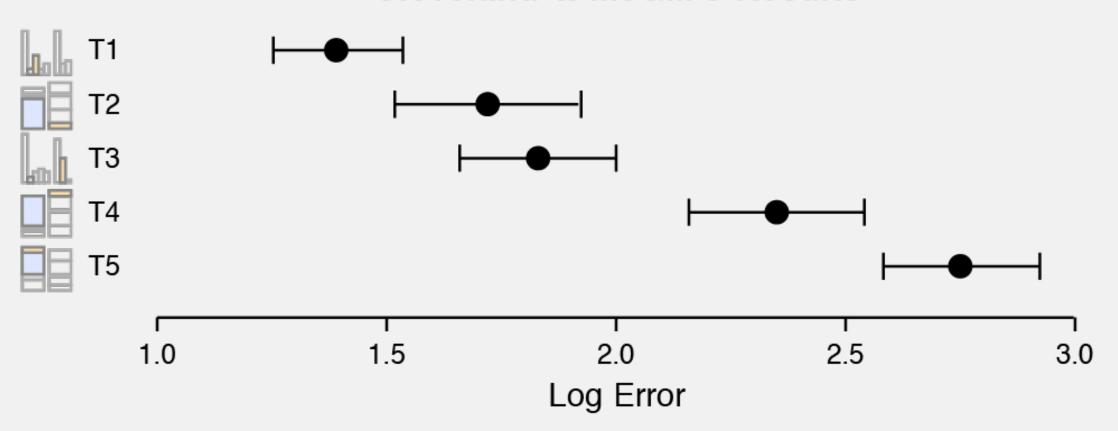
Cleveland & McGill's Results



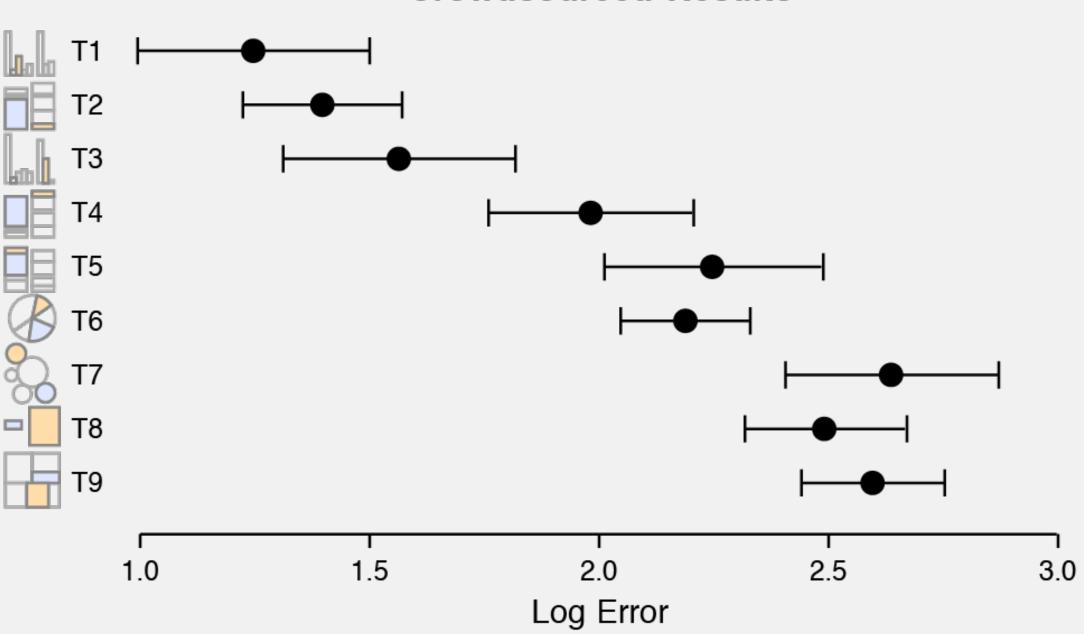


Quelle: Kieran Healys "Data Visualization: A Practical Introduction"; Ergebnisse basierend auf Heer & Bostock sowie Cleveland & McGill

Cleveland & McGill's Results



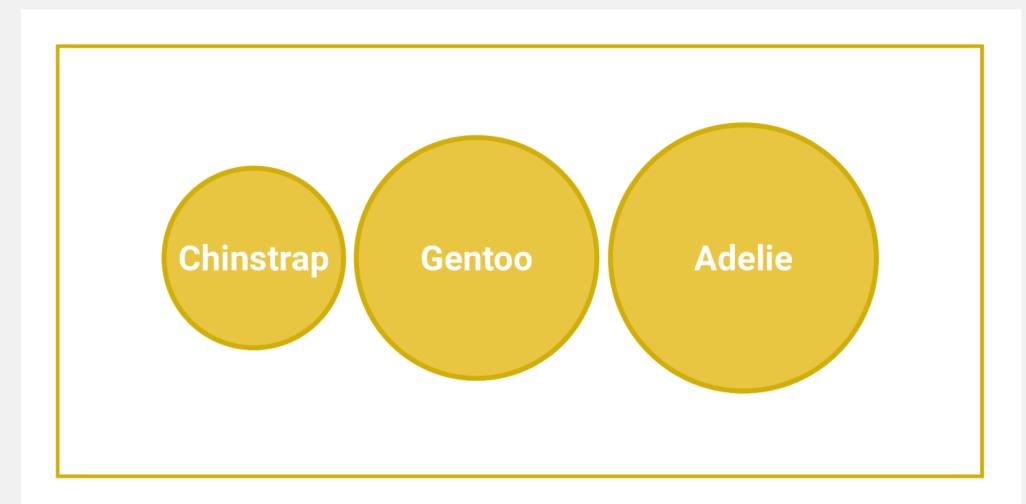
Crowdsourced Results



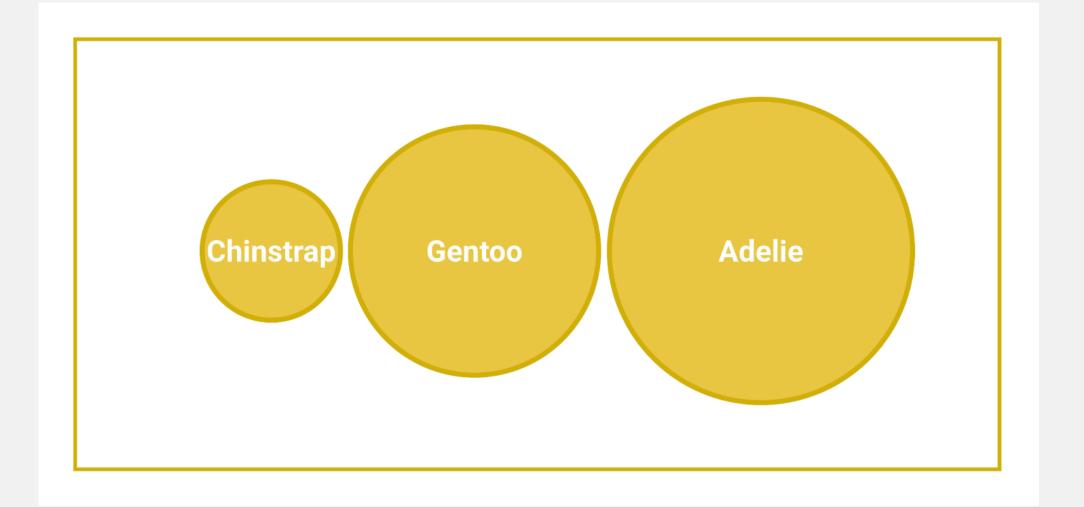
Quelle: Kieran Healys "Data Visualization: A Practical Introduction"; Ergebnisse basierend auf Heer & Bostock sowie Cleveland & McGill

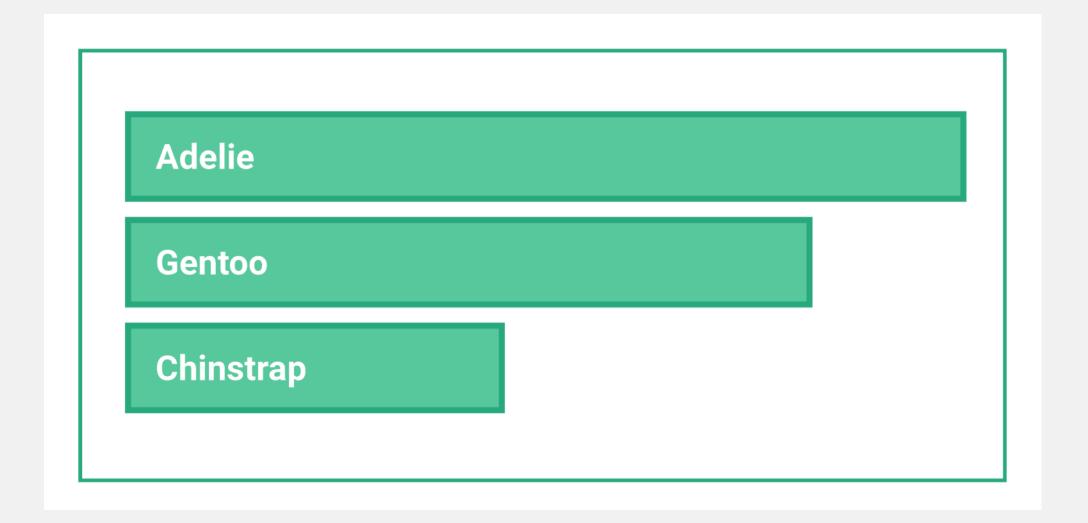
Nutze immer Fläche.





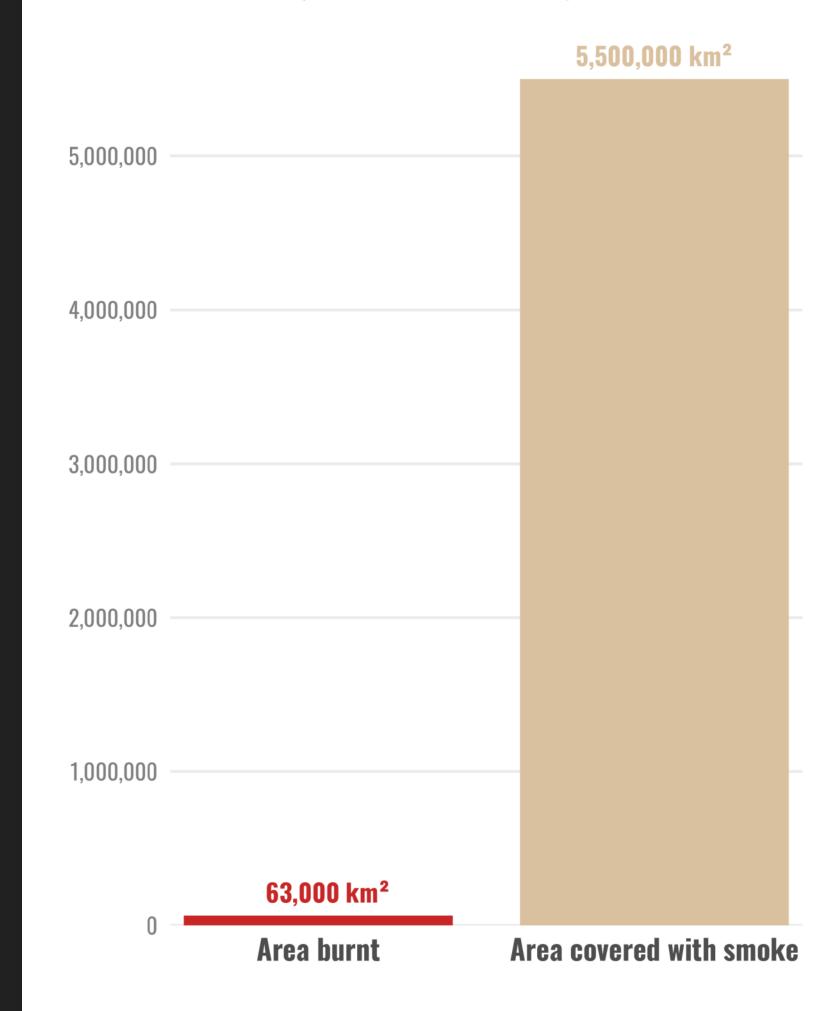
Nutze niemals Radius!





Burnt land and plume of smoke caused by the Australian bushfires in 2019/20

(as of 6th of January 2020)















PERSPECTIVE

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

Tracey L. Weissgerber¹*, Natasa M. Milic^{1,2}, Stacey J. Winham³, Vesna D. Garovic¹

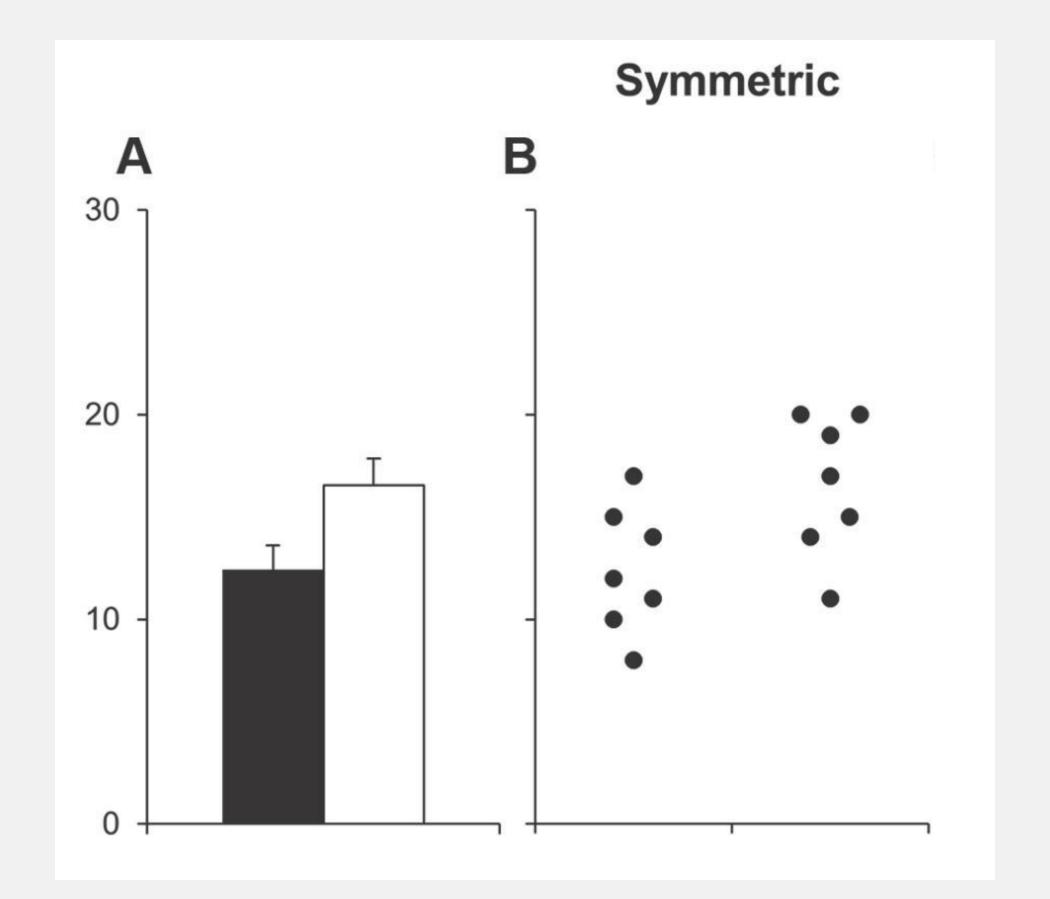
- 1 Division of Nephrology & Hypertension, Mayo Clinic, Rochester, Minnesota, United States of America,
- 2 Department of Biostatistics, Medical Faculty, University of Belgrade, Belgrade, Serbia, 3 Division of Biomedical Statistic and Informatics, Mayo Clinic, Rochester, Minnesota, United States of America
- weissgerber.tracey@mayo.edu

Weissgerber et al. (2015) PLoS Biology

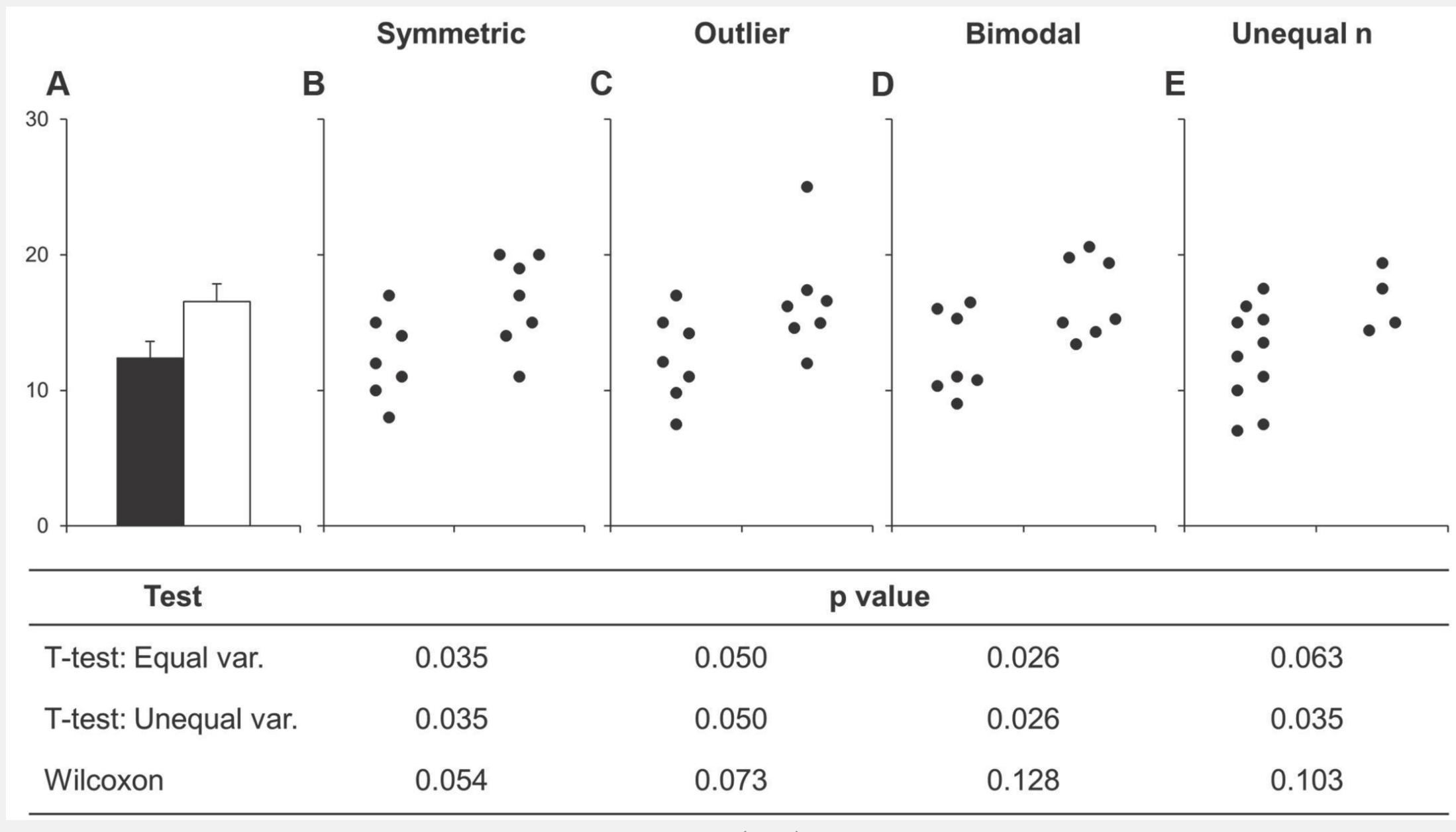








Weissgerber et al. (2015) PLoS Biology

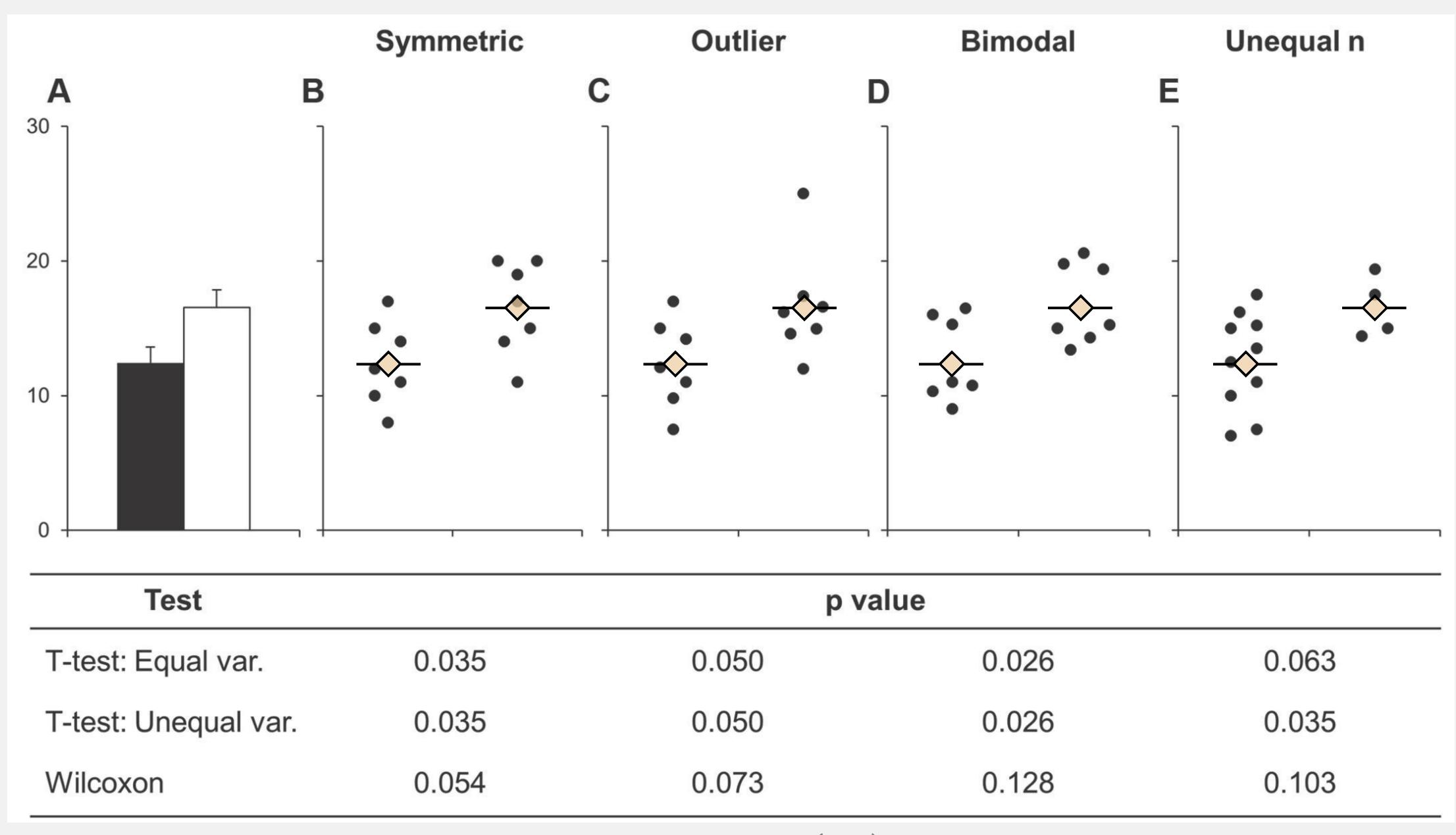


Weissgerber et al. (2015) PLoS Biology







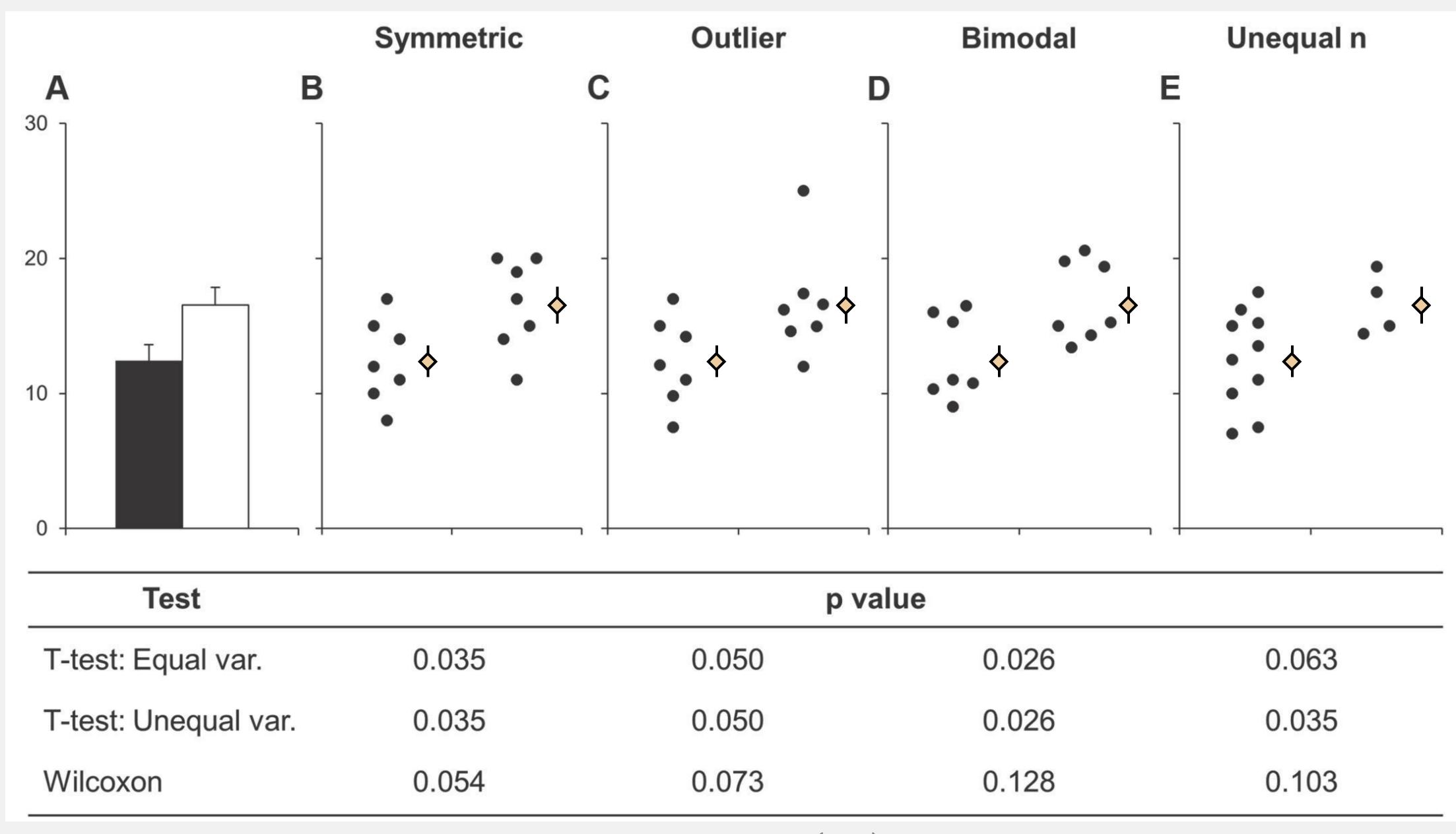


Modified from <u>Weissgerber et al. (2015) PLoS Biology</u>







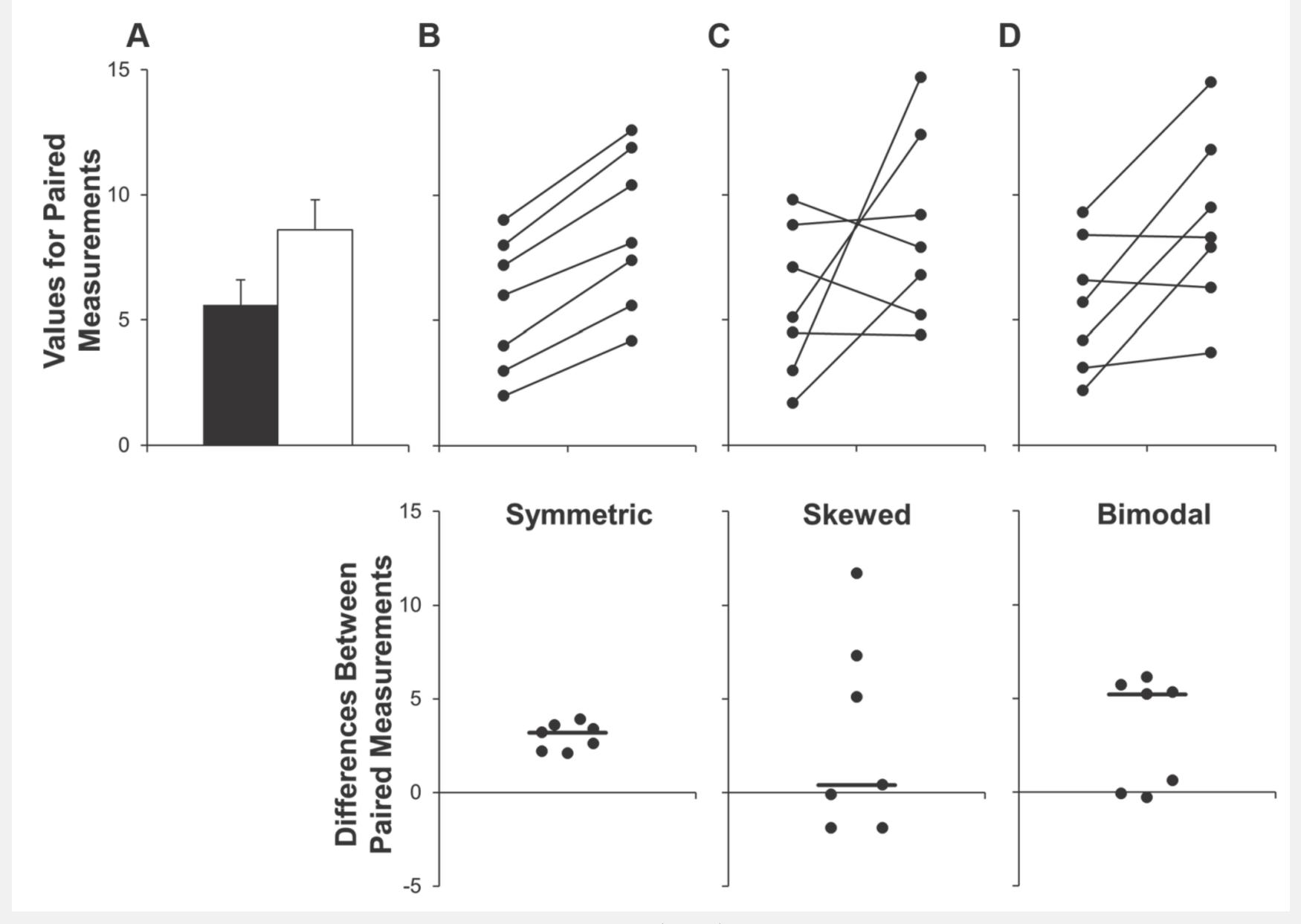


Modified from <u>Weissgerber et al. (2015) PLoS Biology</u>





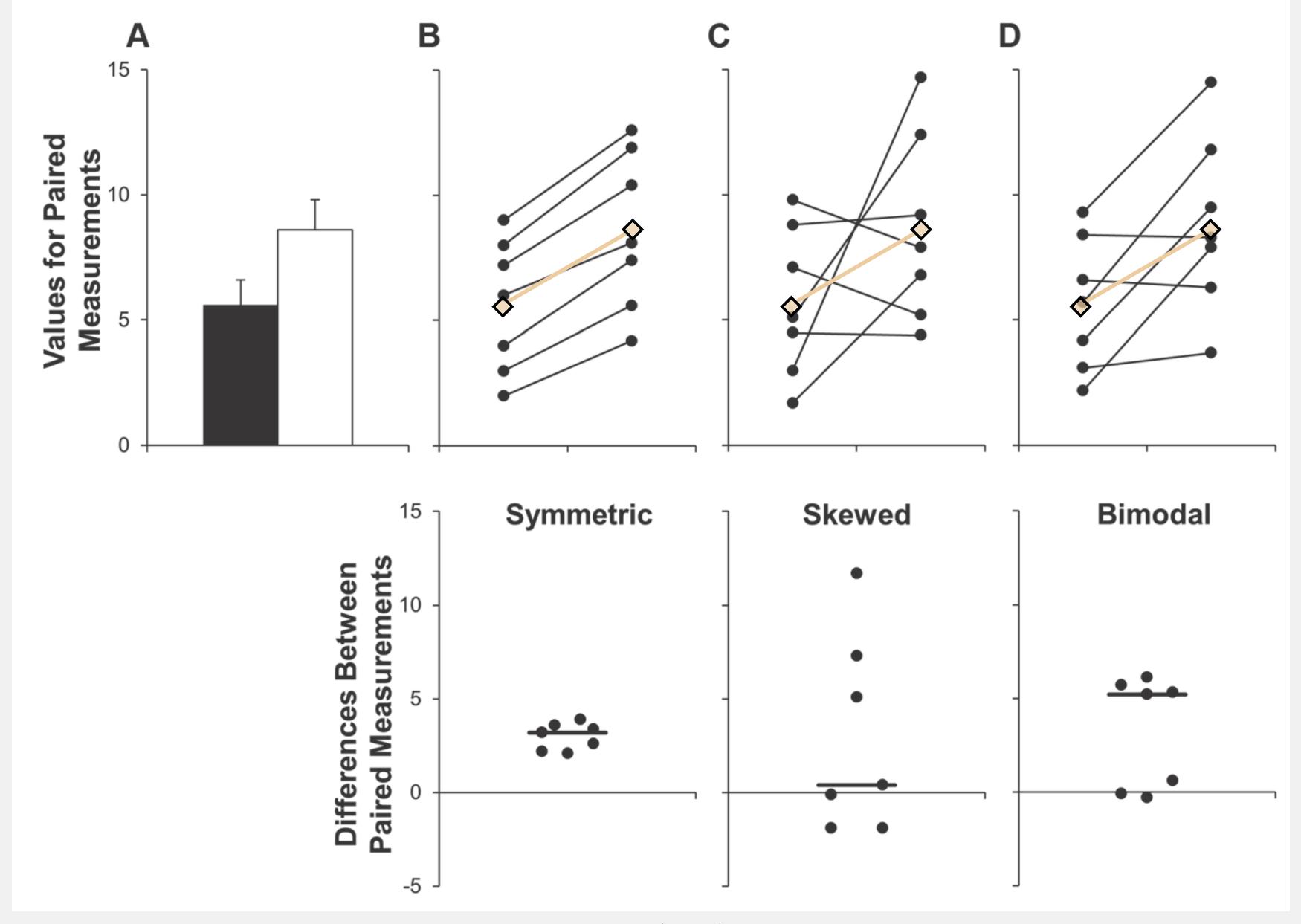




Weissgerber et al. (2015) PLoS Biology





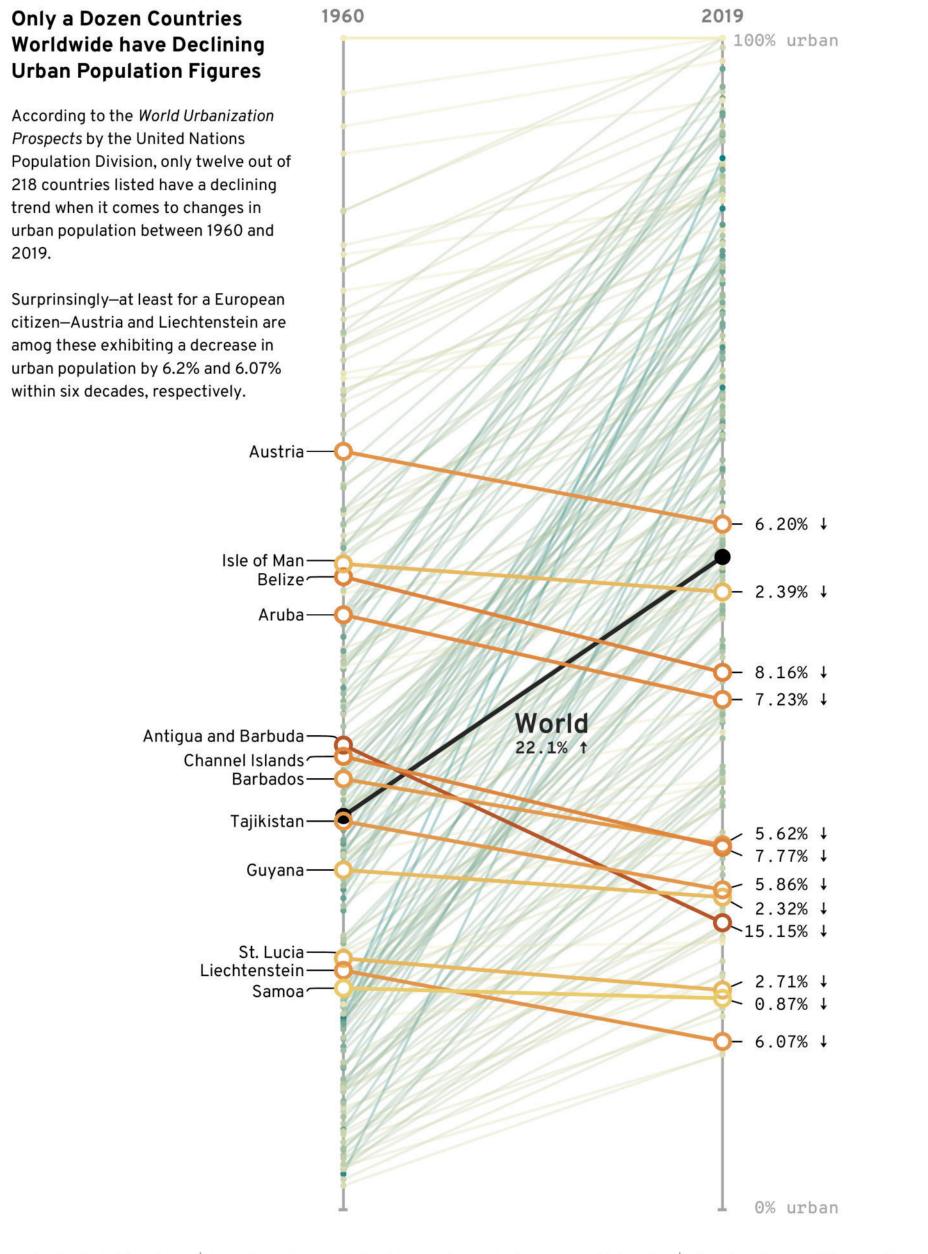


Weissgerber et al. (2015) PLoS Biology









Visualization by Cédric Scherer Data: United Nations Population Division. World Urbanization Prospects: 2018 Revision. #30DayChartChallenge 2021 Day 5: Slope

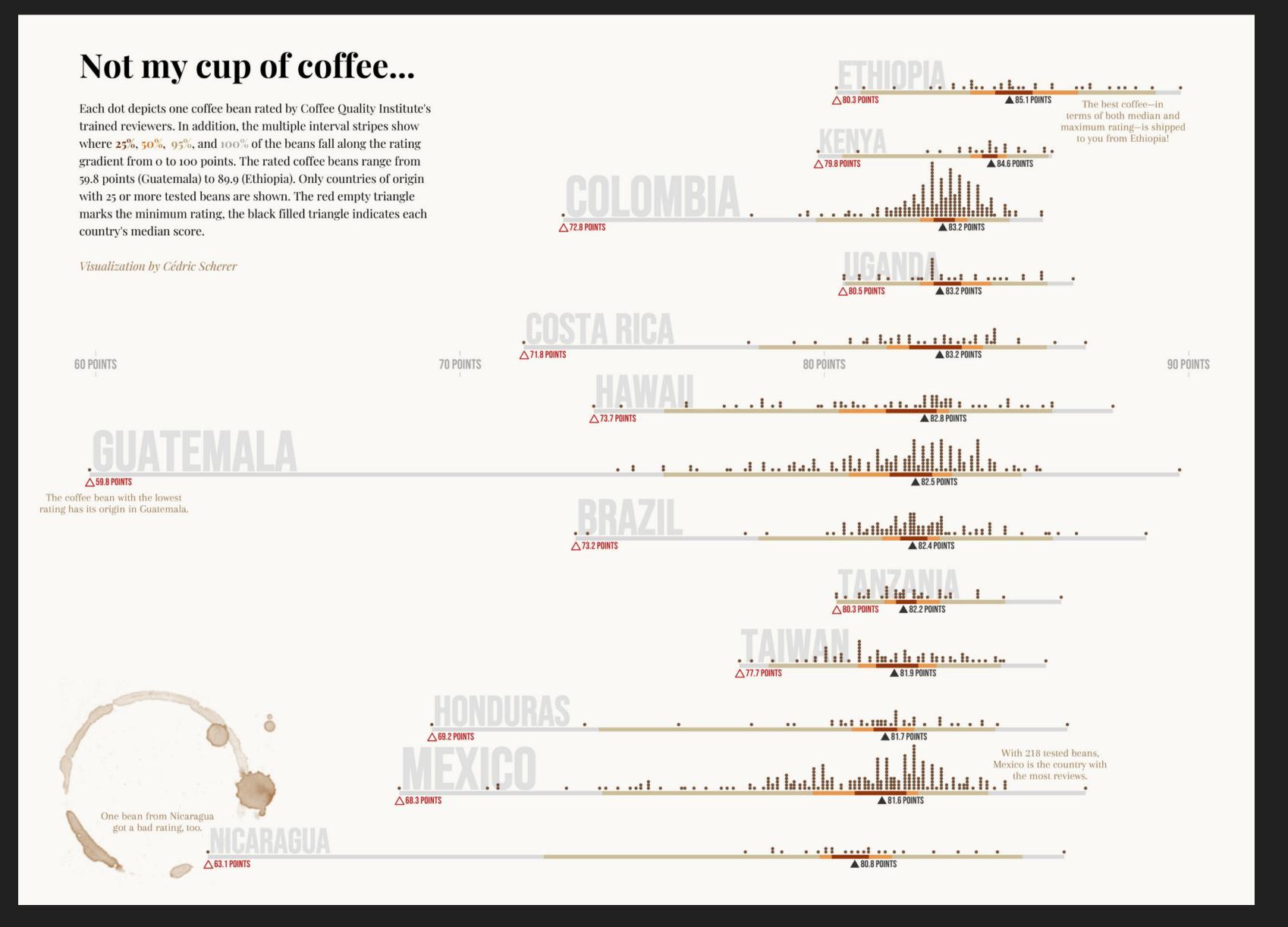










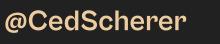


"Not my cup of coffee", #TidyTuesday Contribution

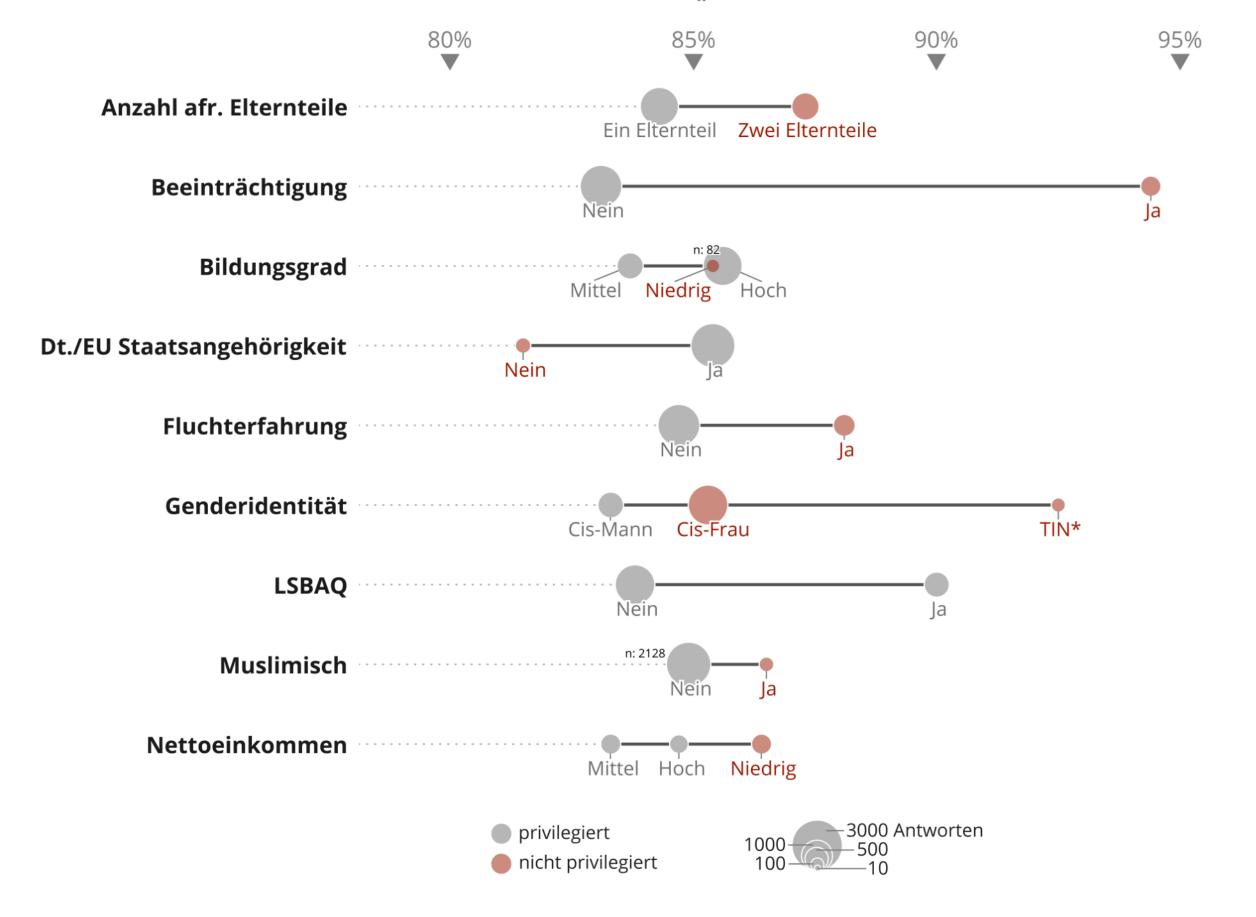








Häufigkeit von Diskriminierungserfahrungen entlang ausgewählter Vielfaltsdimensionen im Lebensbereich "Medien und Internet"



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.

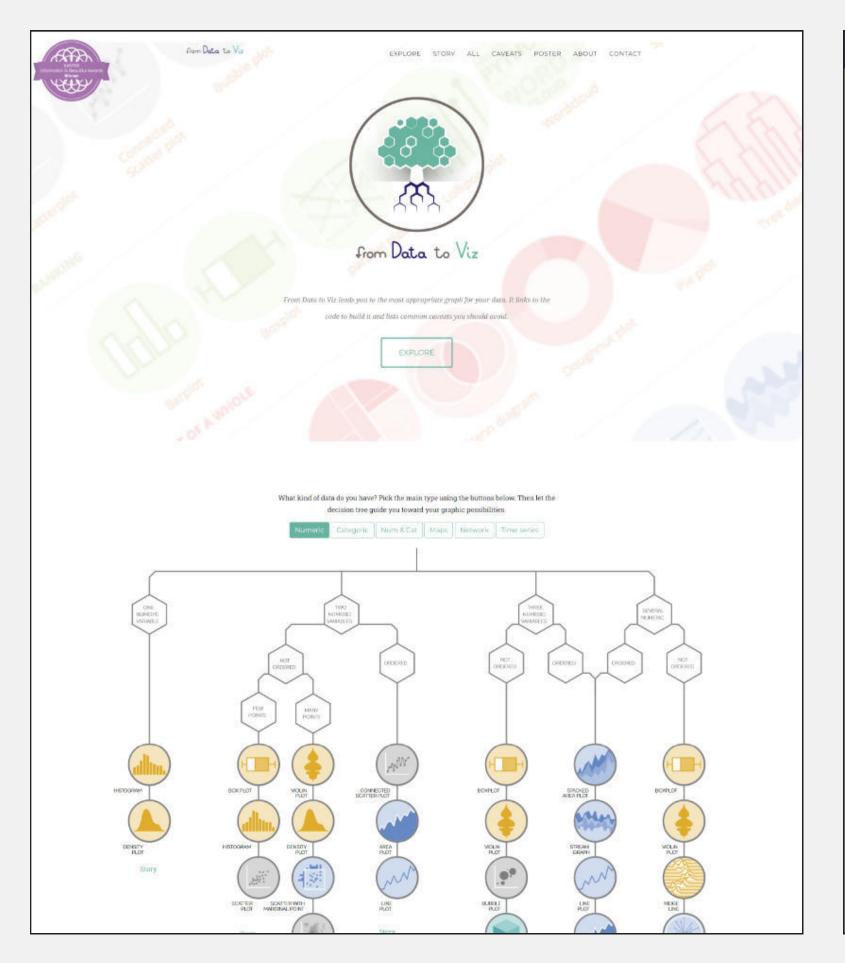
Quelle: Abb. 46 in Aikins, M A; Bremberger, T; Aikins, J K; Gyamerah, D; Yıldırım-Caliman, D (2021): Afrozensus 2020 | Datenteam: Reiber, L; Vivanco, J | Design: Scherer, C Lizenz: CC-BY-NC by EOTO & CFE | afrozensus.de

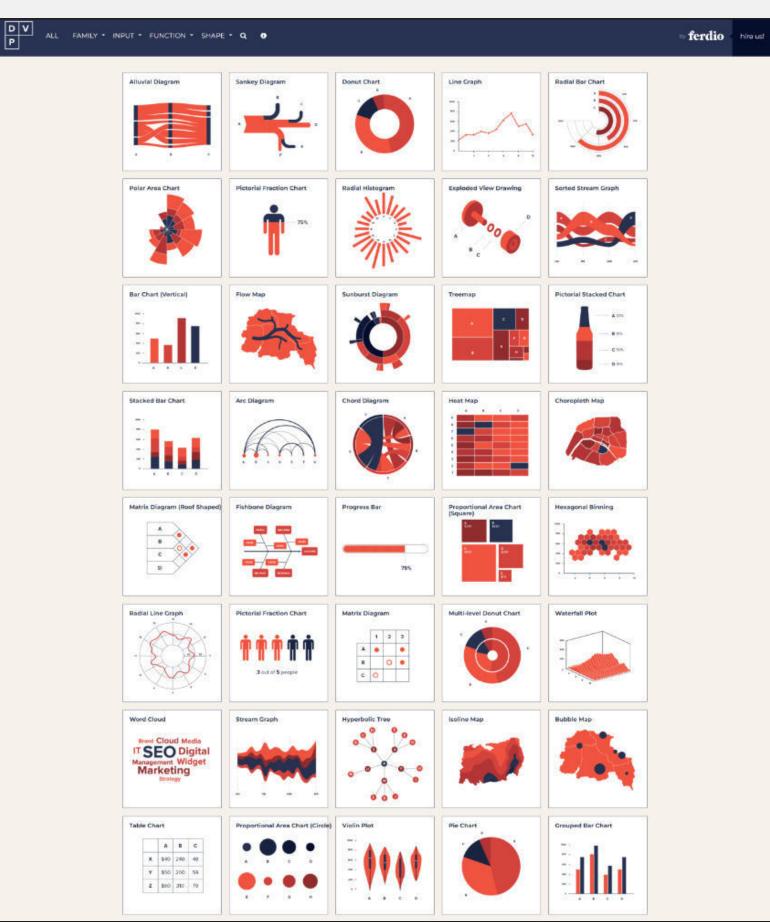
Abb. 46 "Afrozensus 2020" von Citizens For Europe & EOTO e.V.

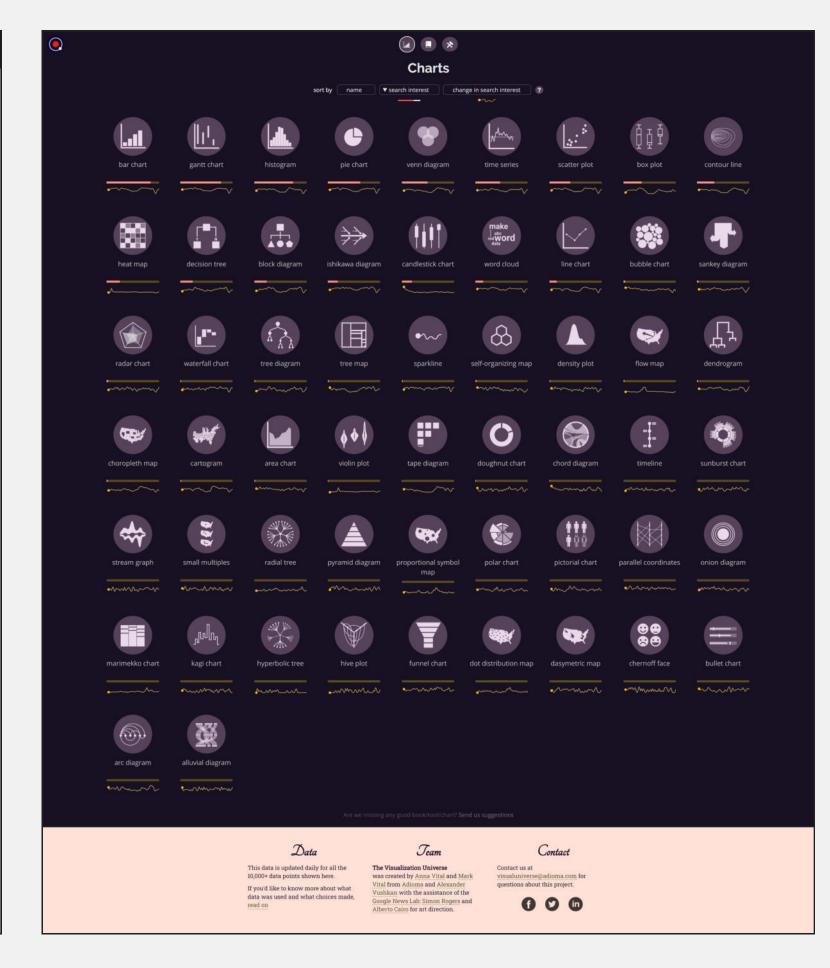












data-to-viz.com

datavizproject.com

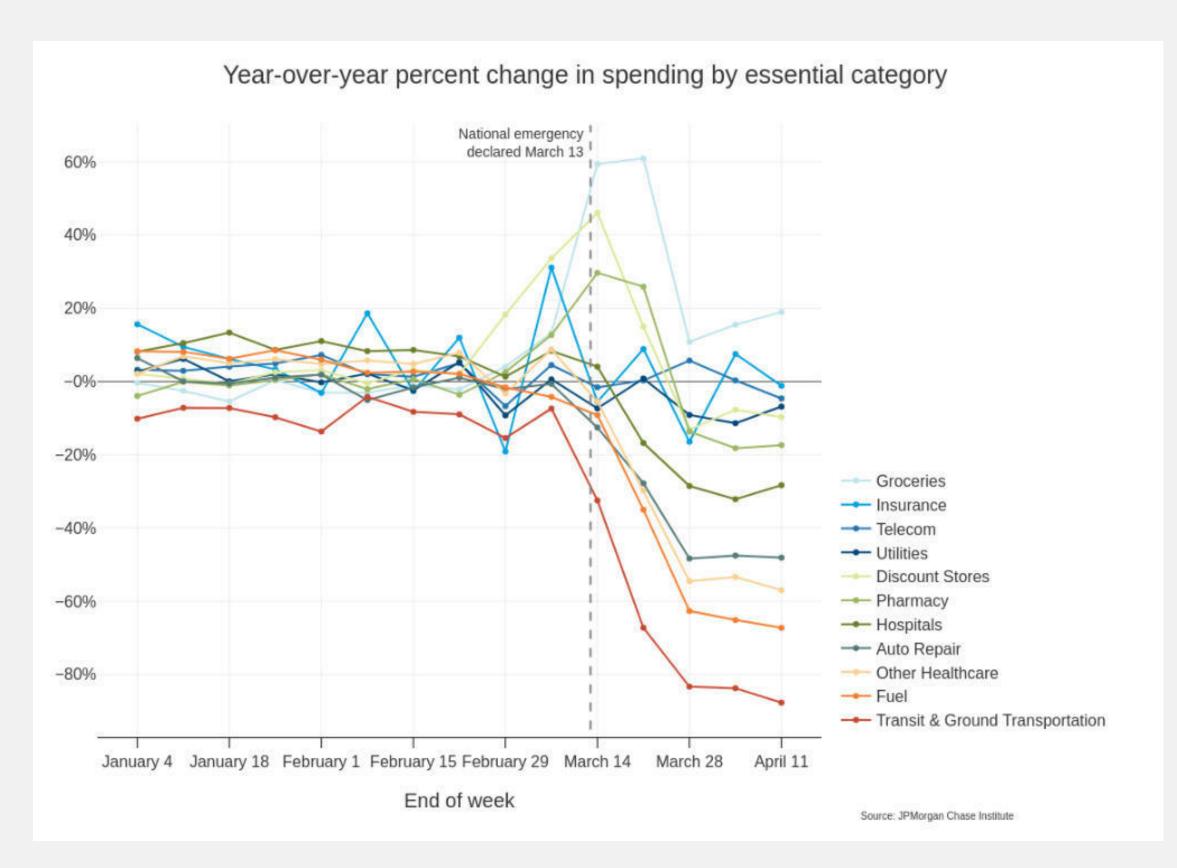
visualizationuniverse.com



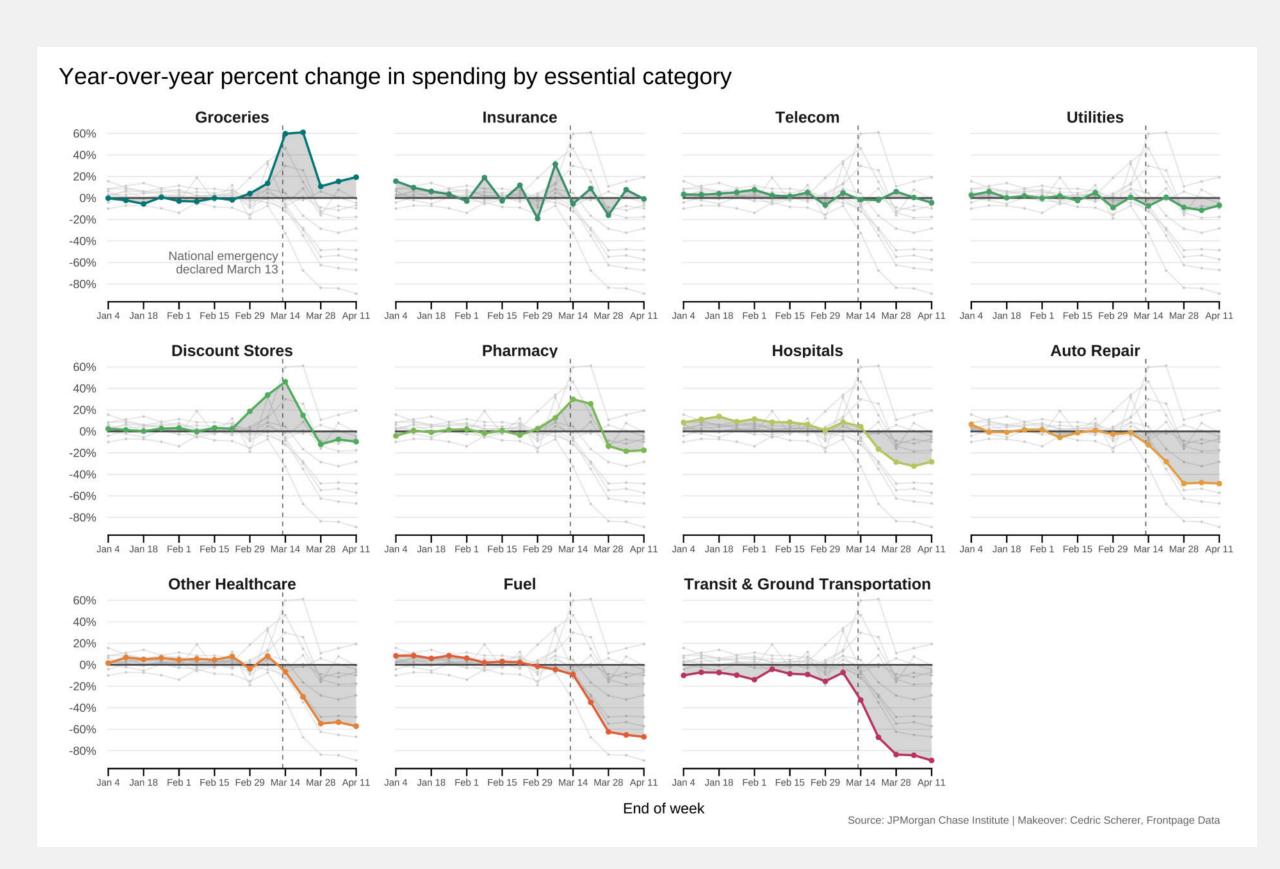




Die Wirkung von "Small Multiples"



Originalgrafik vom JPMorgan Chase Institut



Umgestaltung mit "Small Multiples"



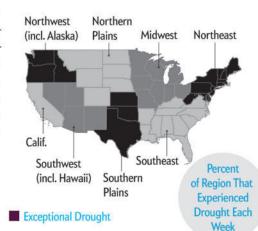


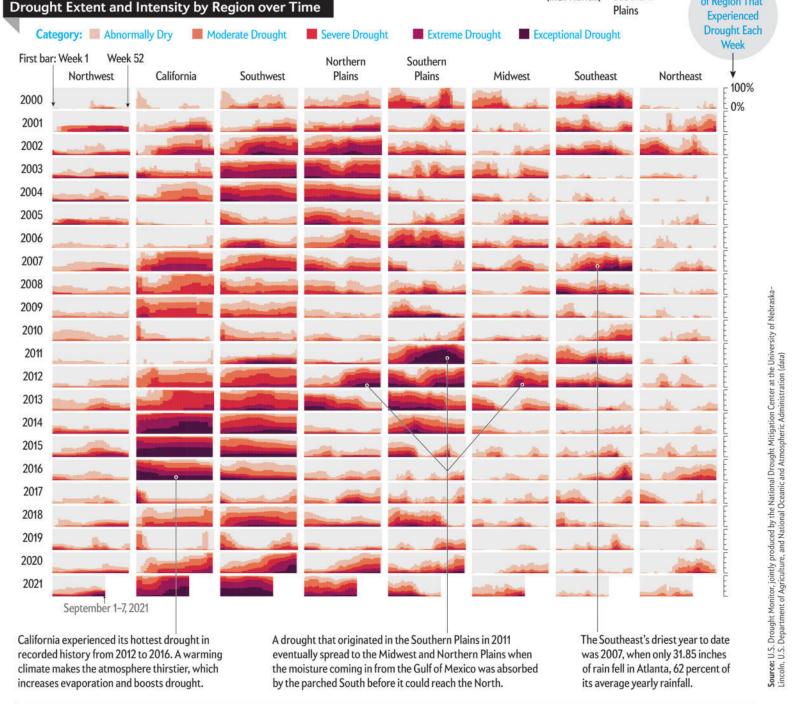


Escalating Drought

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."





74 Scientific American, November 2021



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.

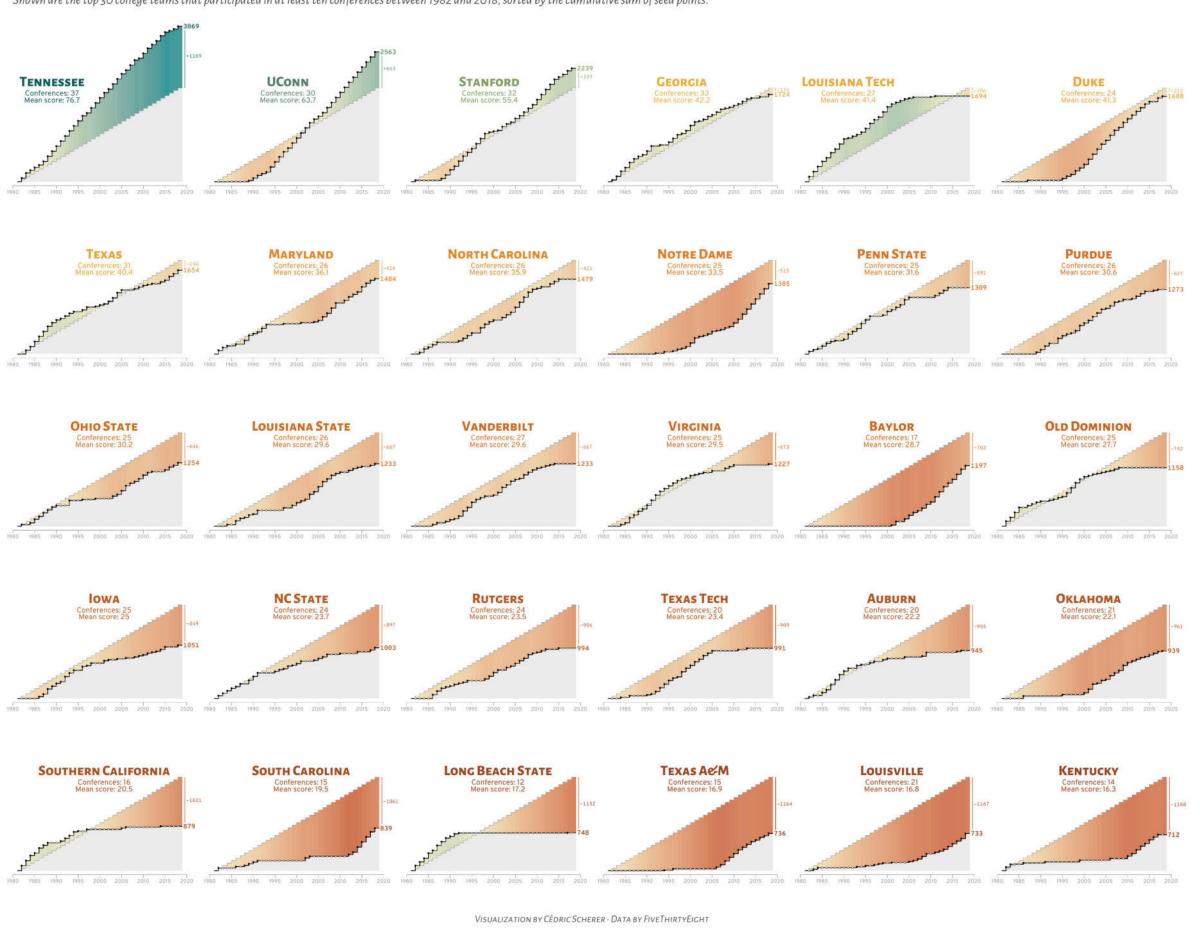
was 2007, when only 31.85 inches of rain fell in Atlanta, 62 percent of its average yearly rainfall.

"Escalating Drought", zusammen mit Georgios Karamanis für den Scientific American, Ausgabe Nov 2021

— 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.

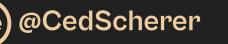


"The Rise and Fall of Women's College Basketball Dynasties", #TidyTuesdαy Contribution

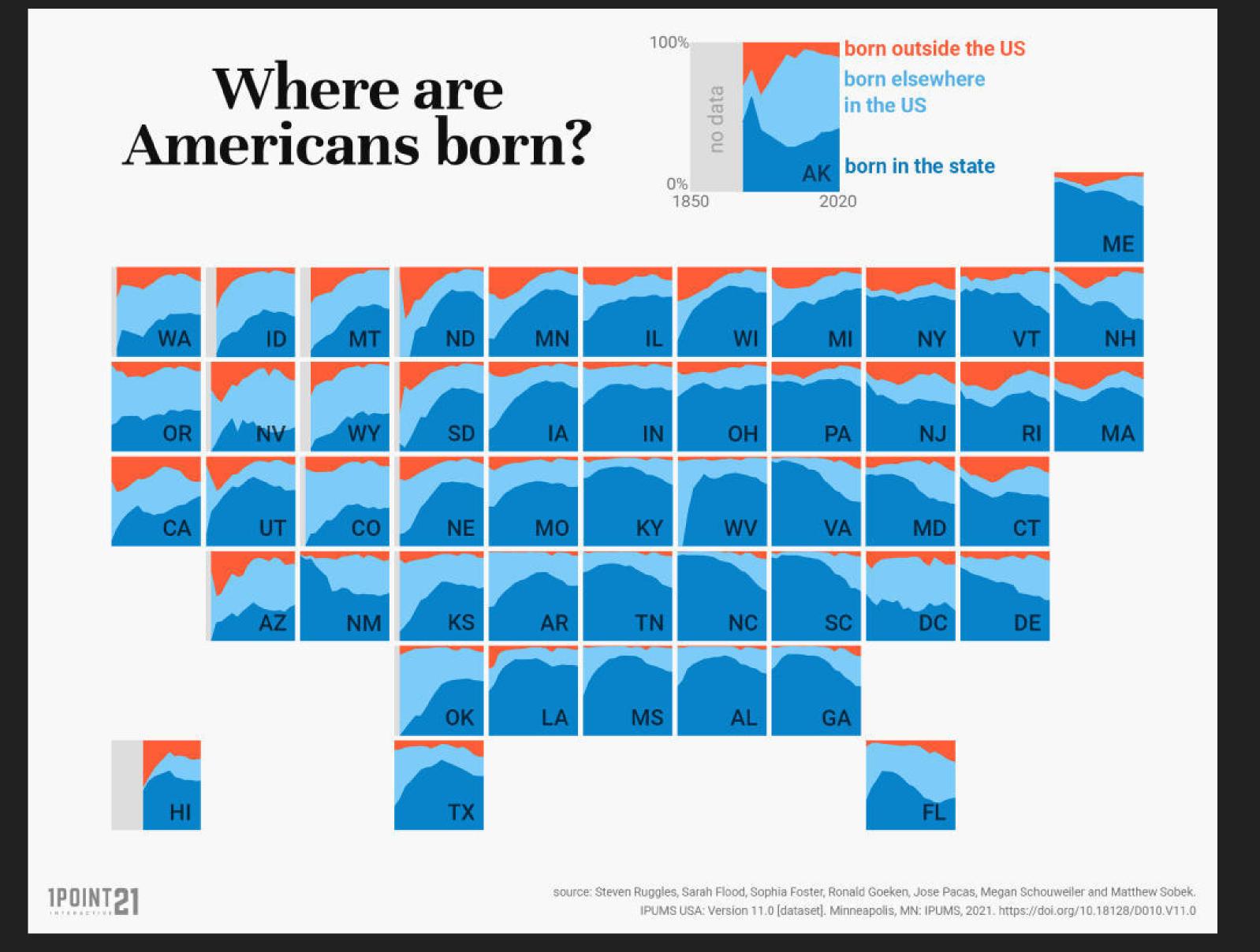










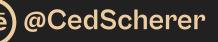


"Where are Americans born?" von @ErinDataViz









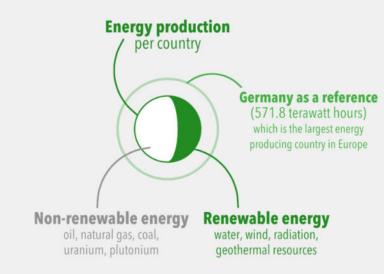


How European countries generated electricity in 2018

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.

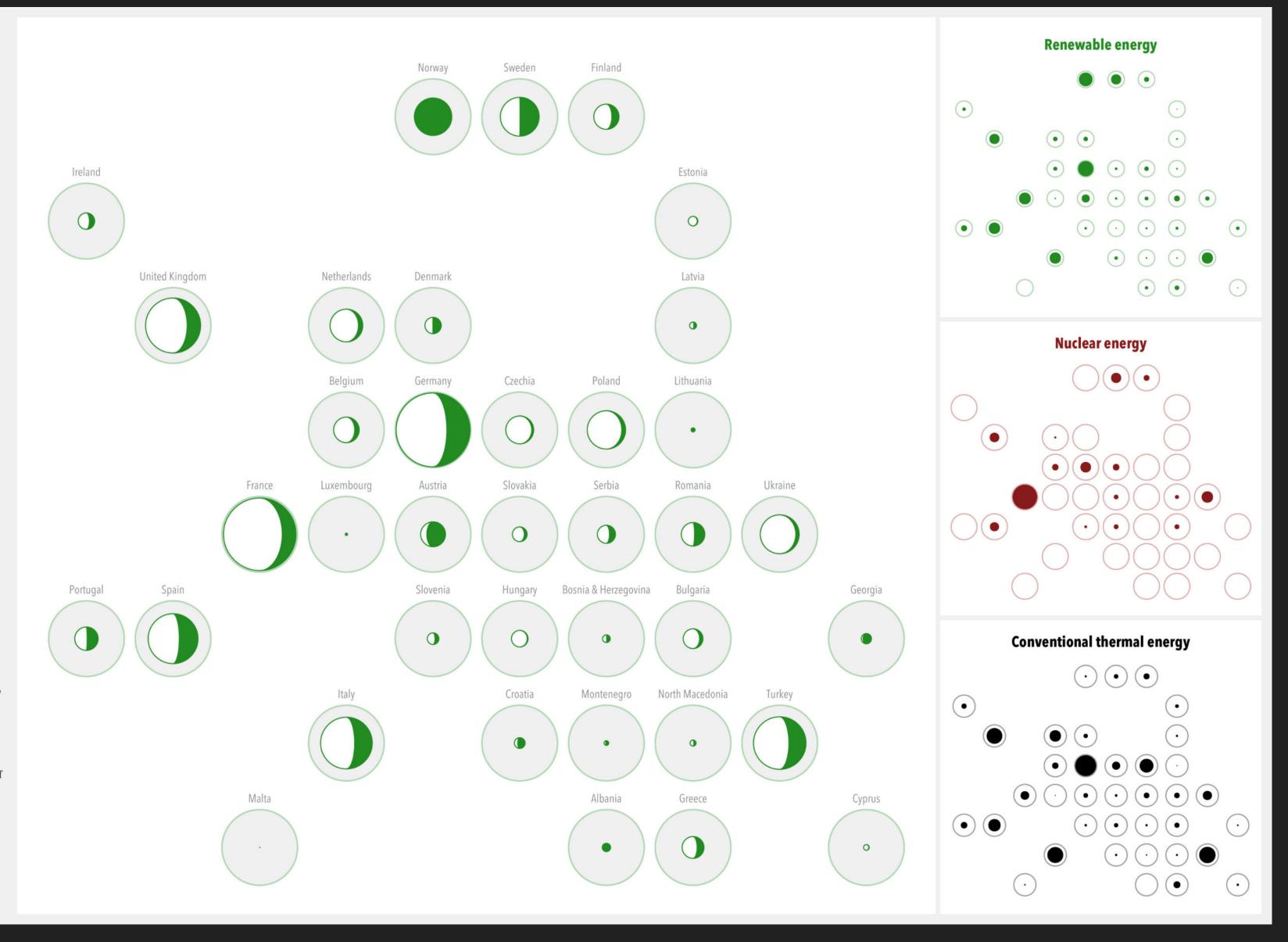
Conventional thermal: Renewable: Nuclear: Germany France Germany 179.1 TWh 320.4 TWh (31% of its production) (71% of its production) (56% of its production)

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

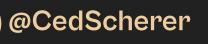


"How European countries generated electricity in 2018", #TidyTuesday Contribution

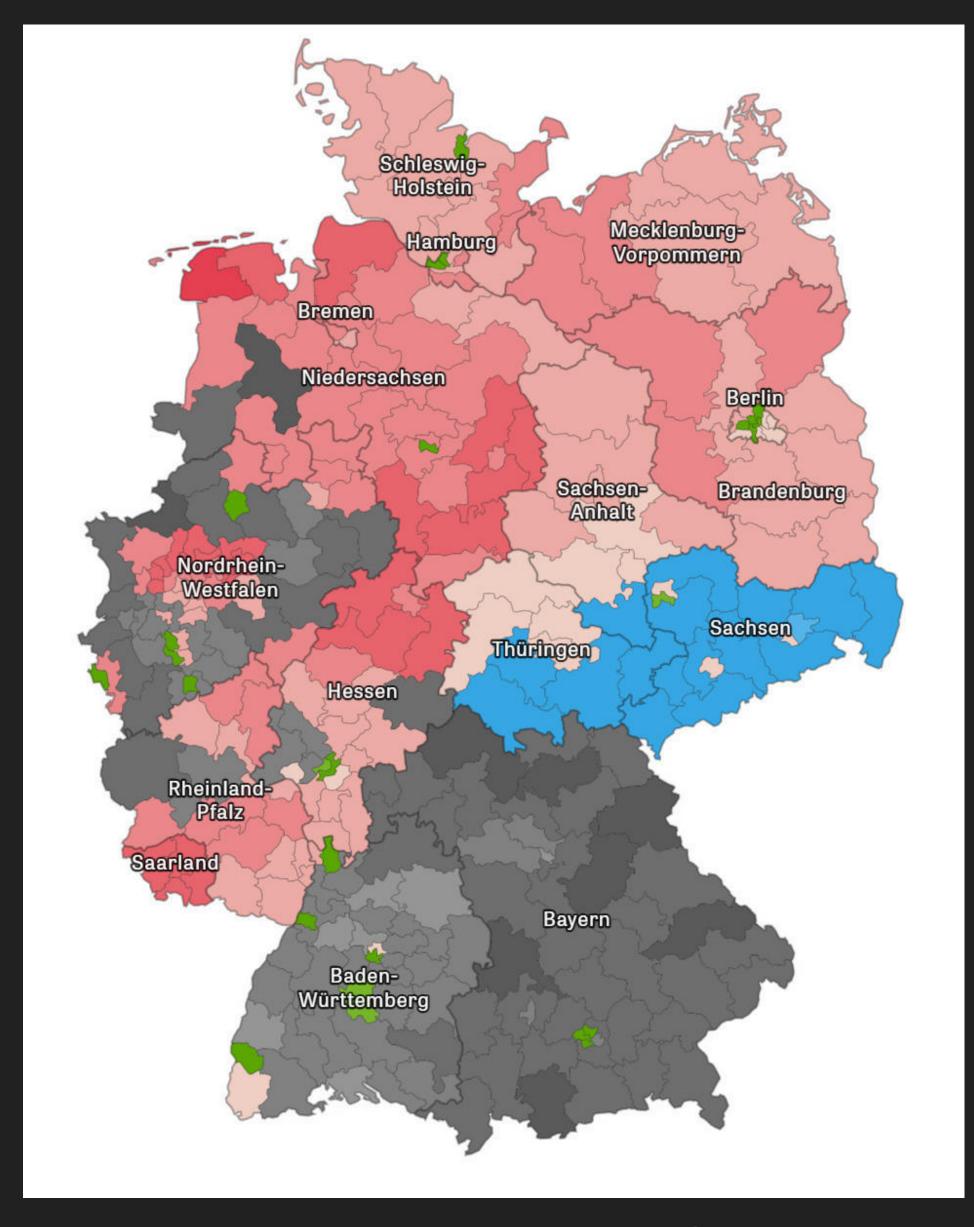


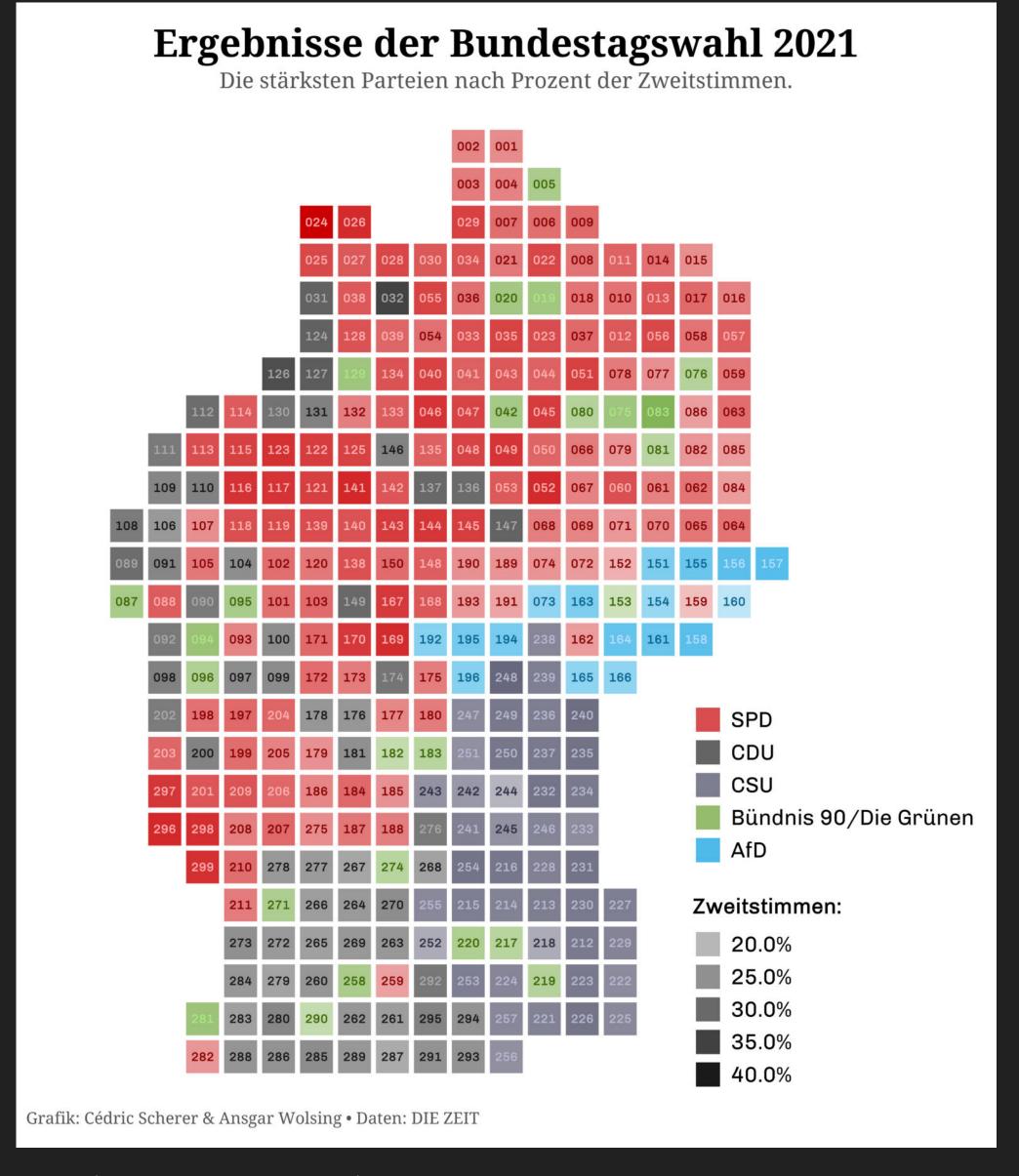












Links: Choropleth-Karte von Die Zeit | Rechts: Kachel-Karte ("Tile Grid Map") von Cédric Scherer & Ansgar Wolsing

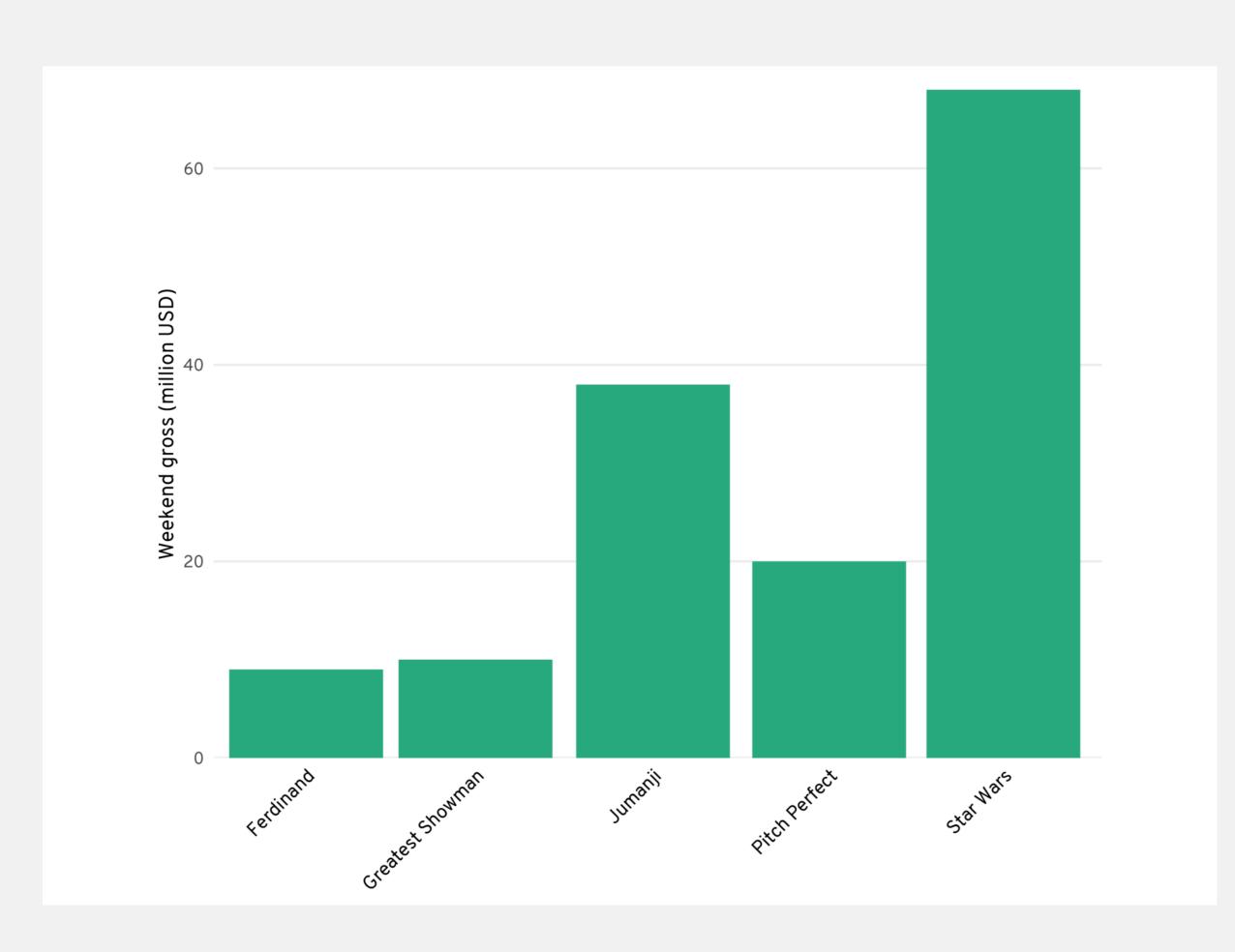
VISUAL FORM

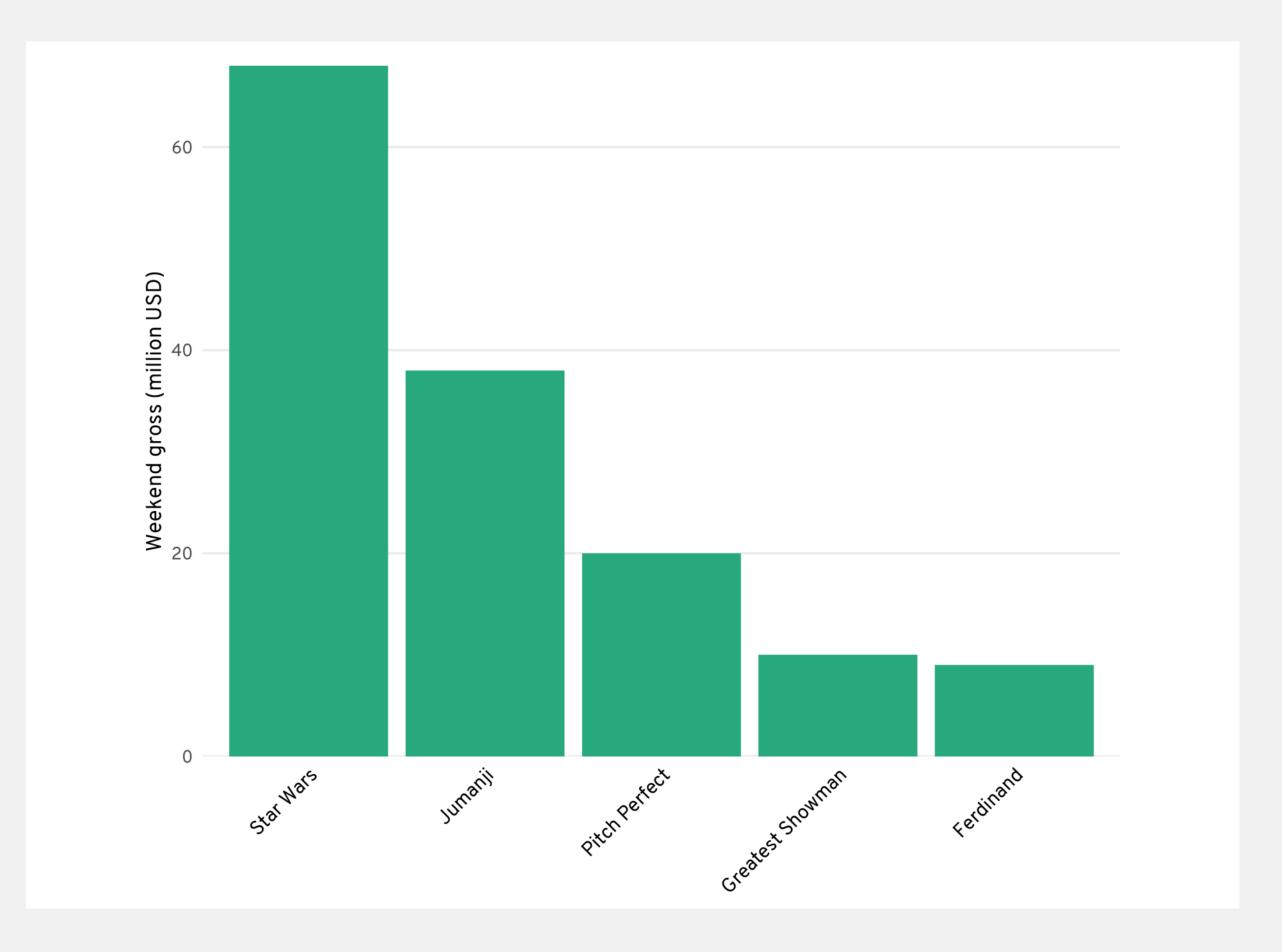
Folge Grundsätzen aus Design und Datenvisualisierung





Daten sortieren



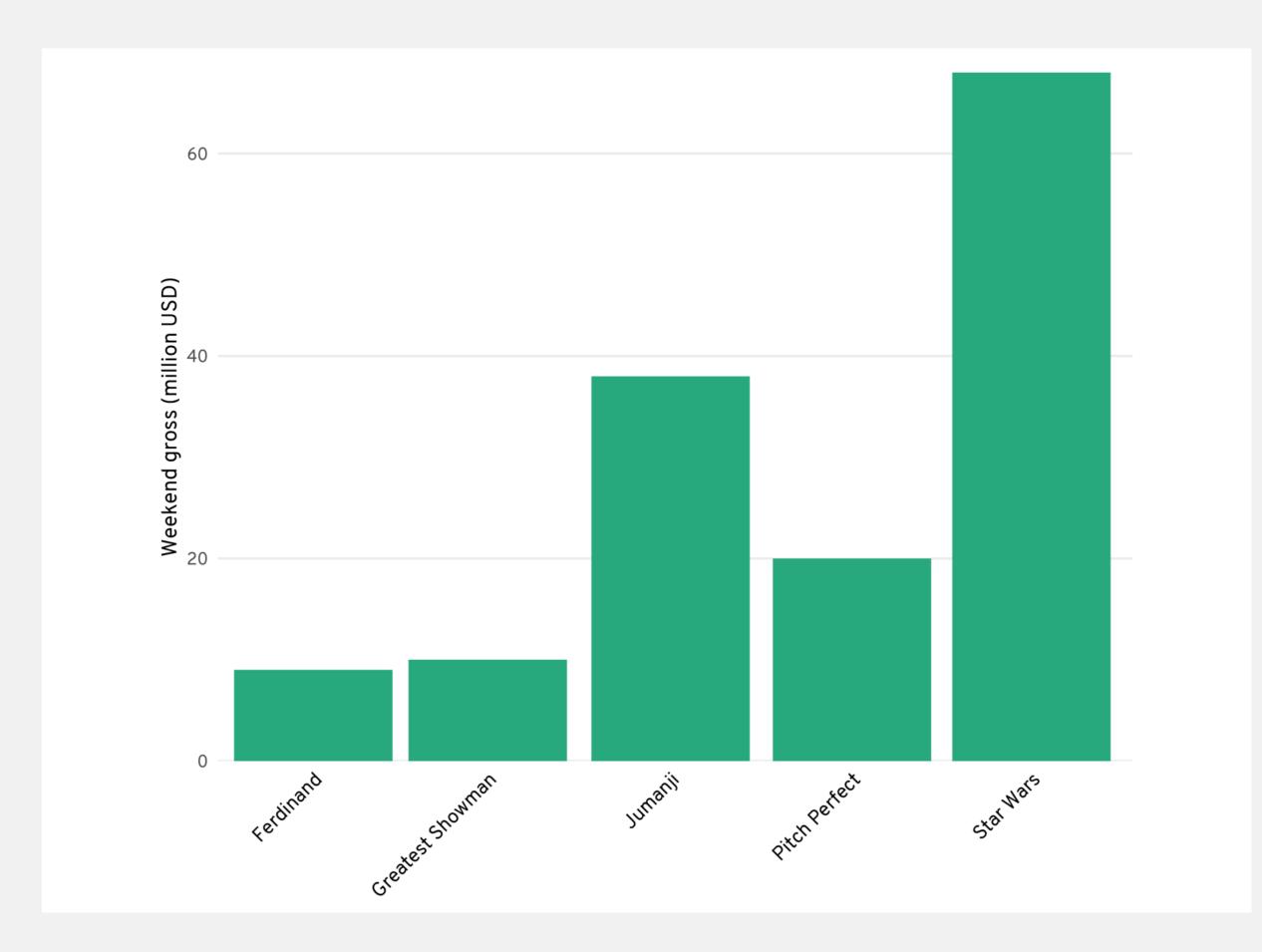


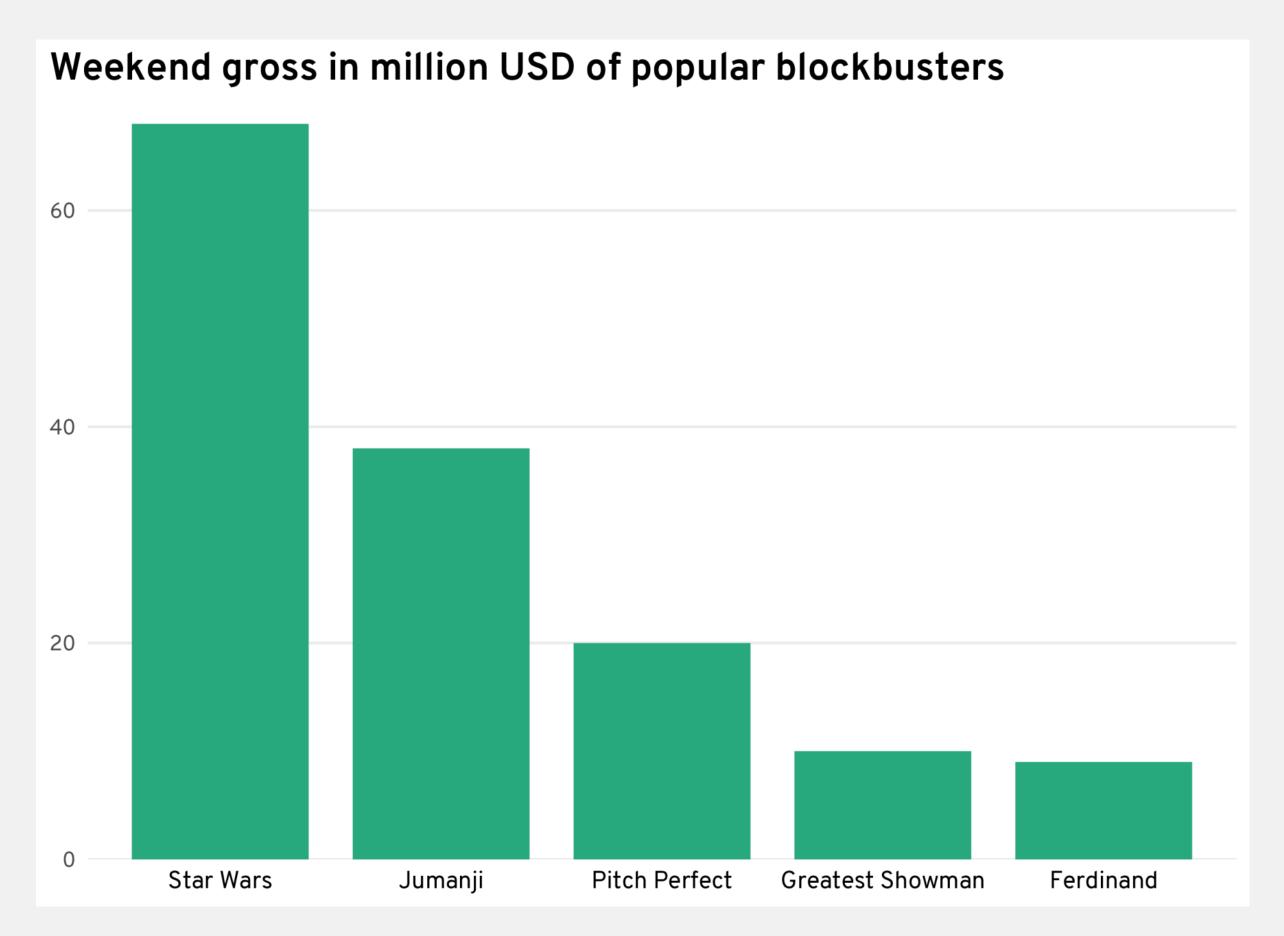






Textdrehung vermeiden



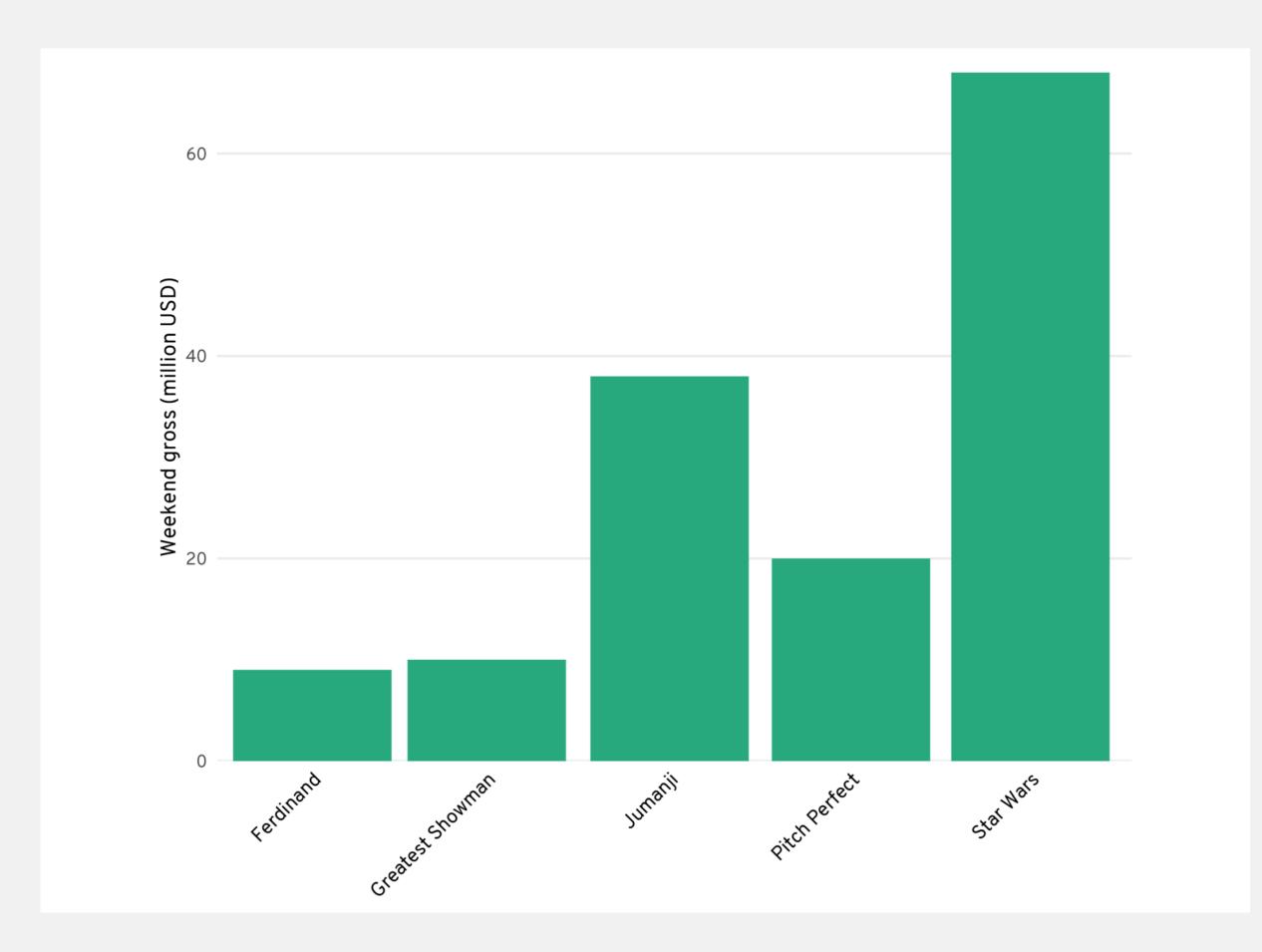


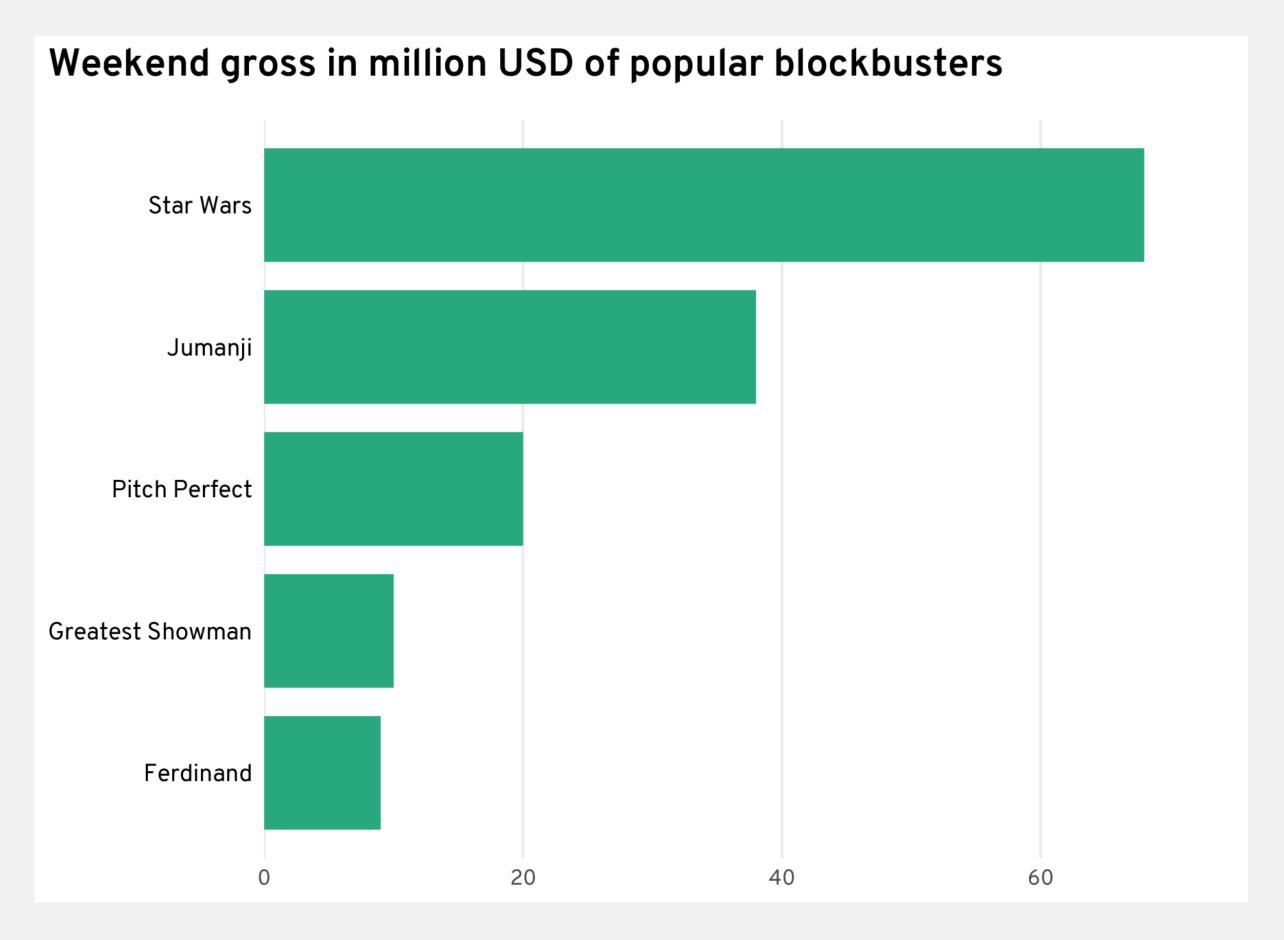






Textdrehung vermeiden



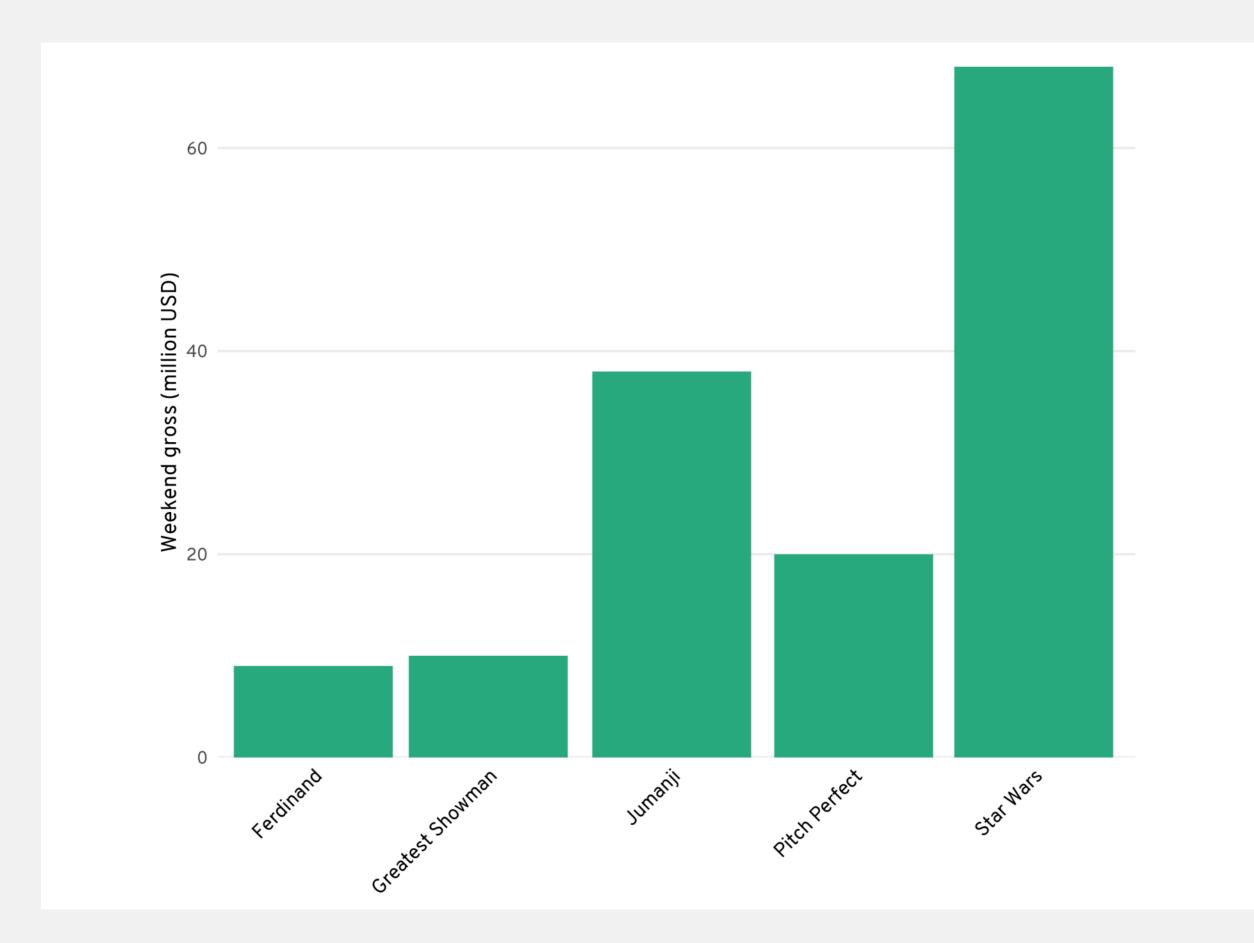


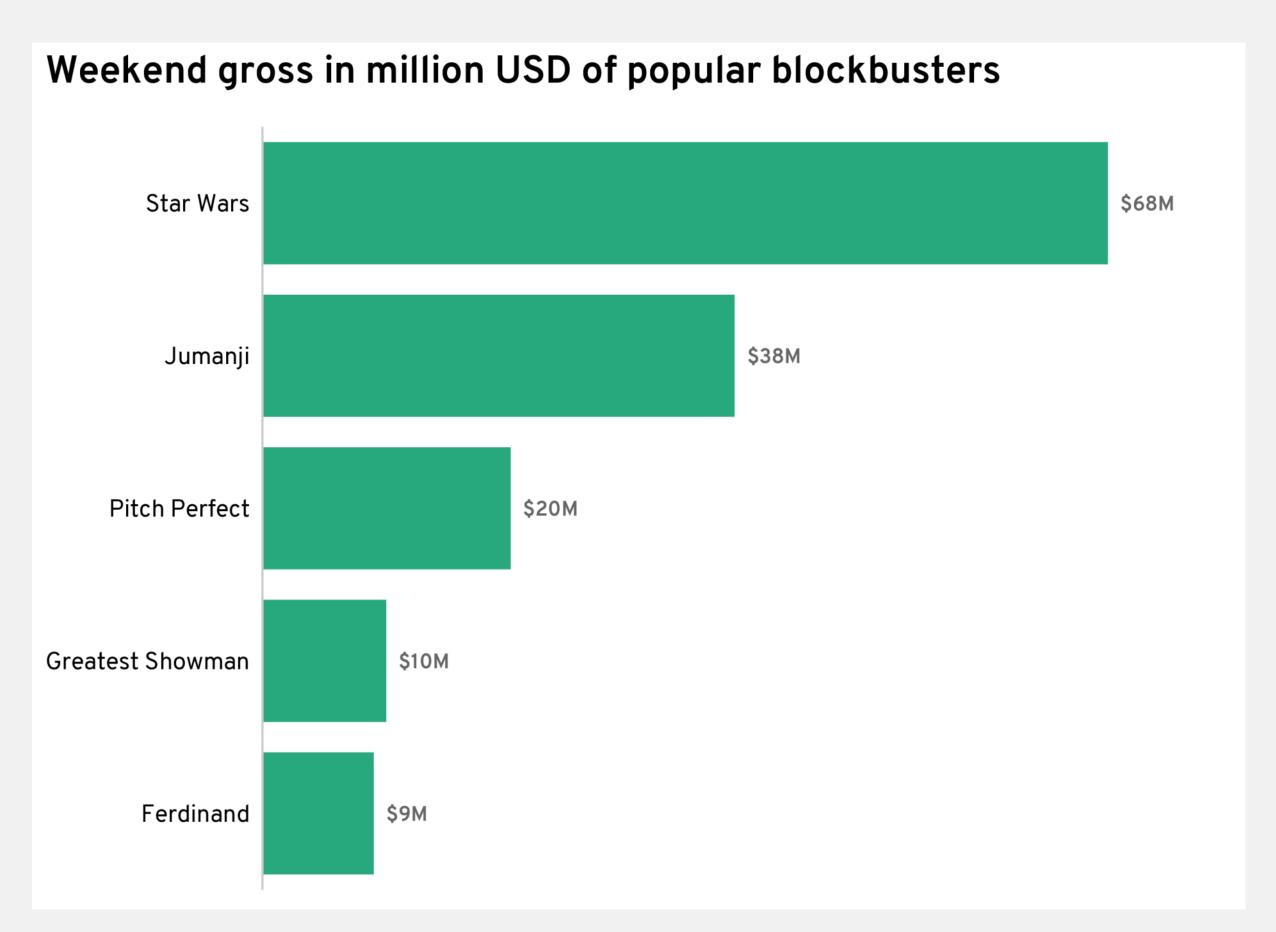






Direktbeschriftungen hinzufügen



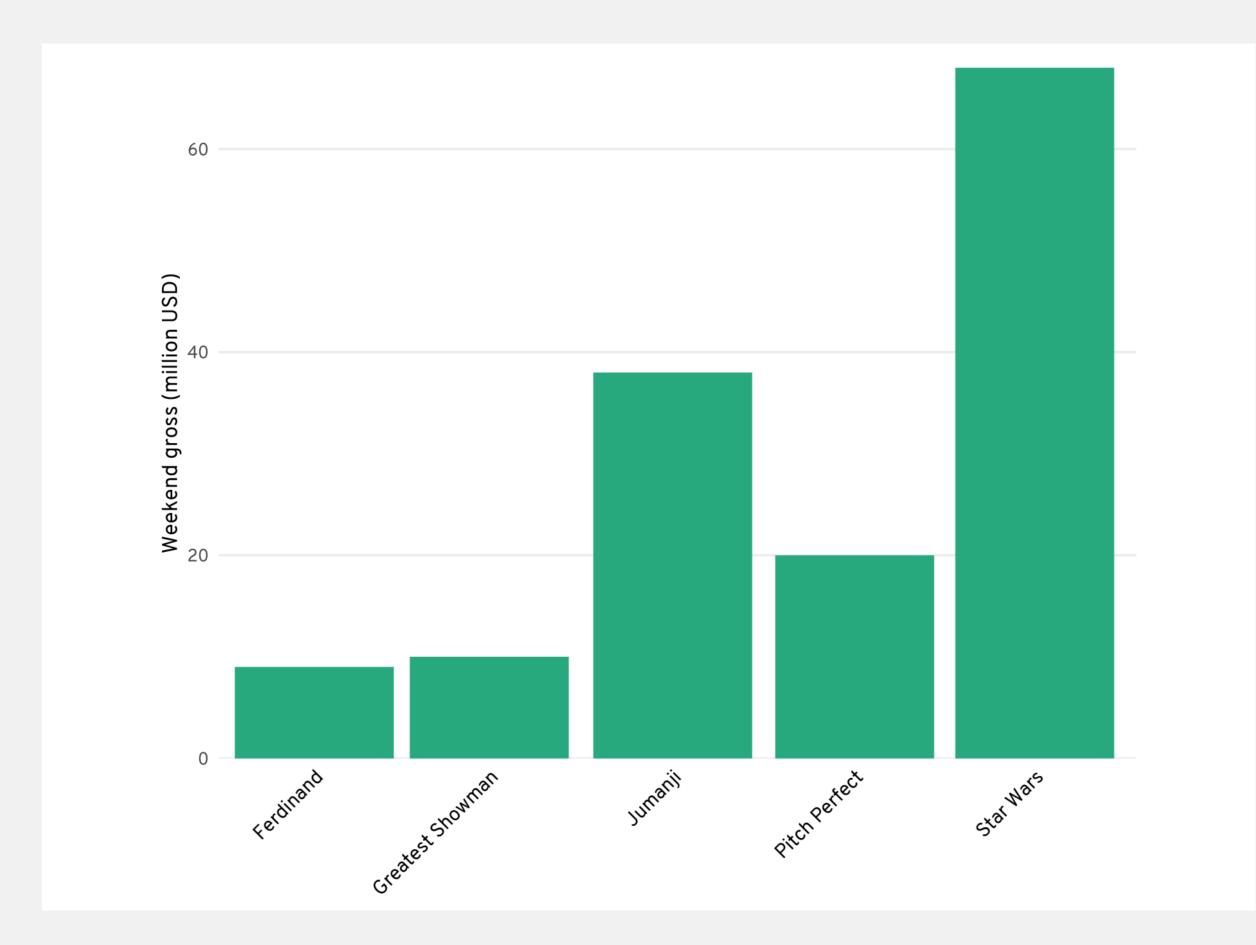


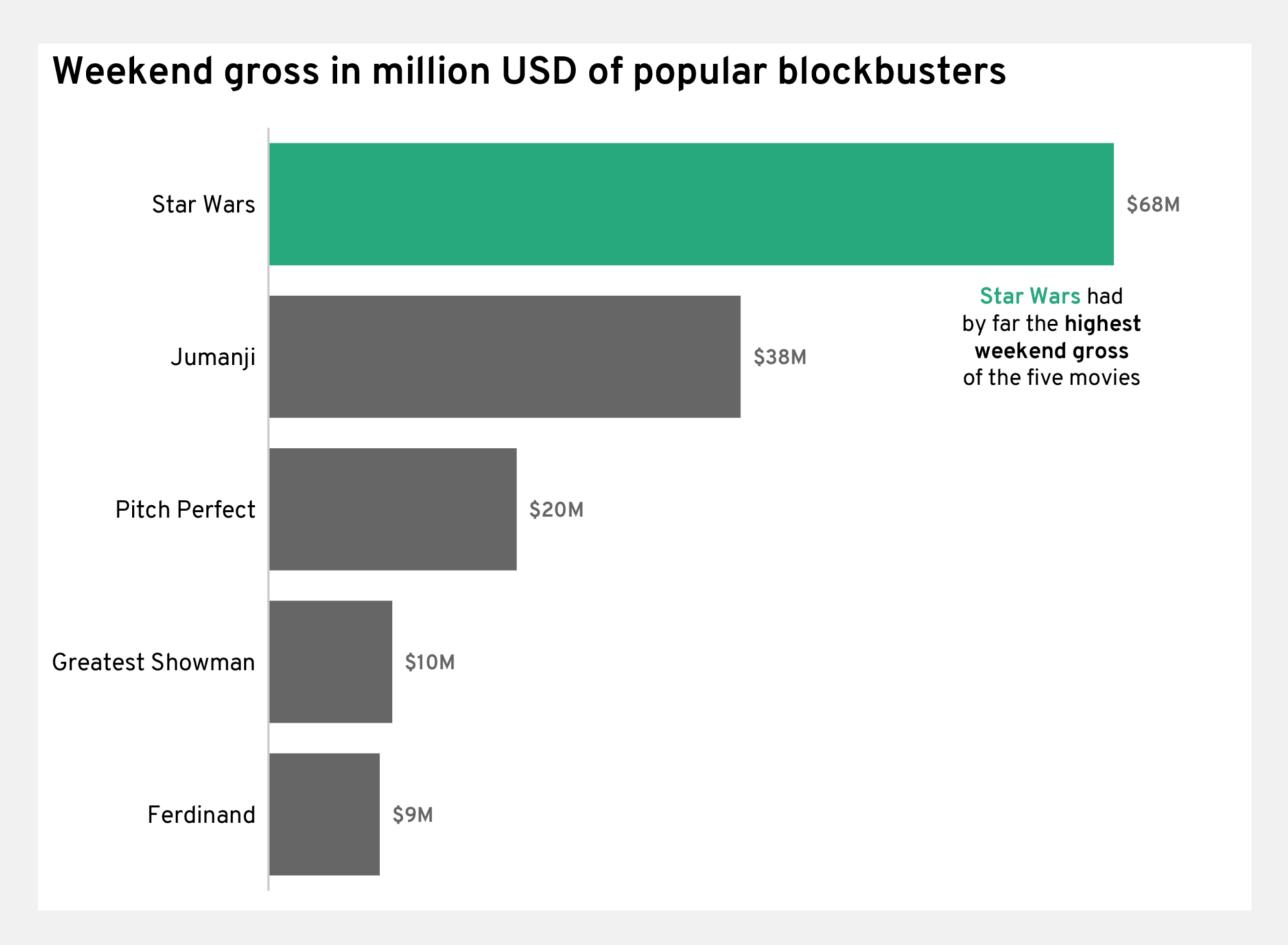






Farben + Anmerkungen nutzen



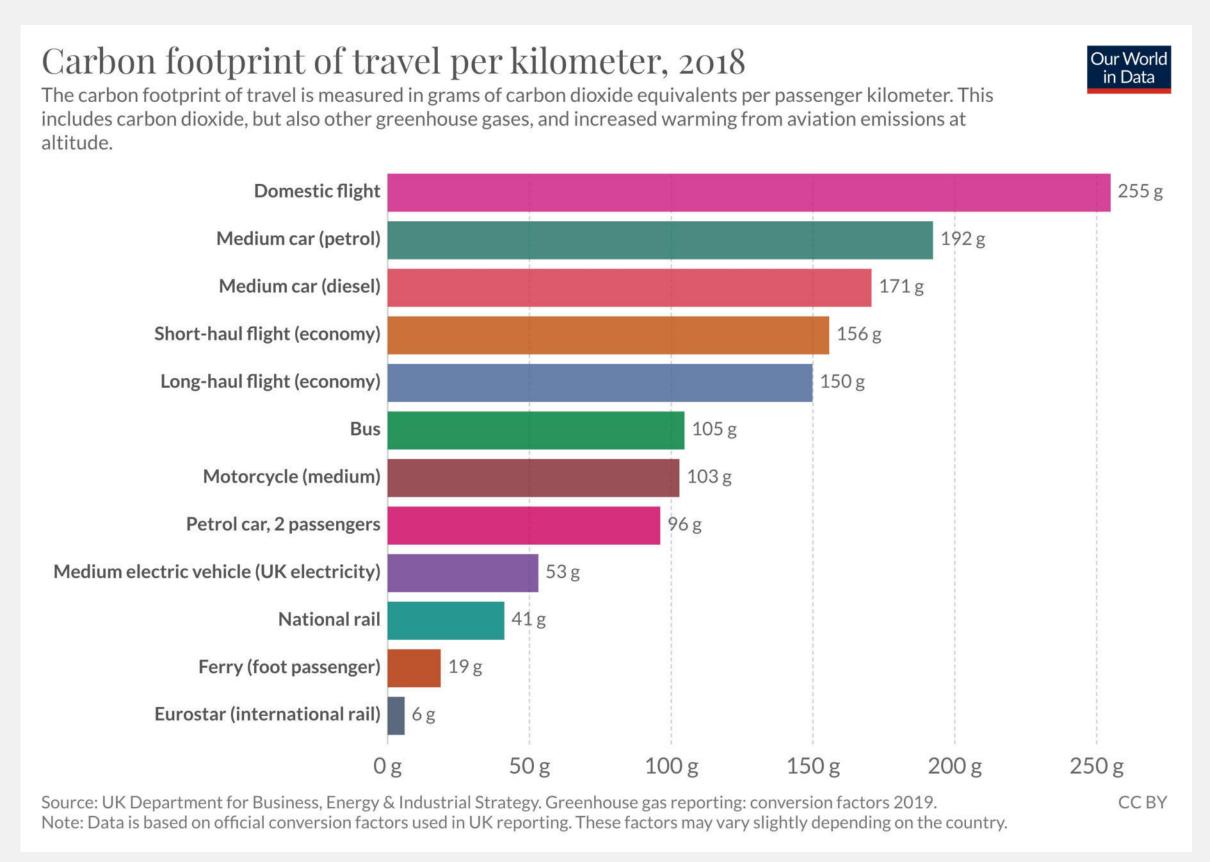




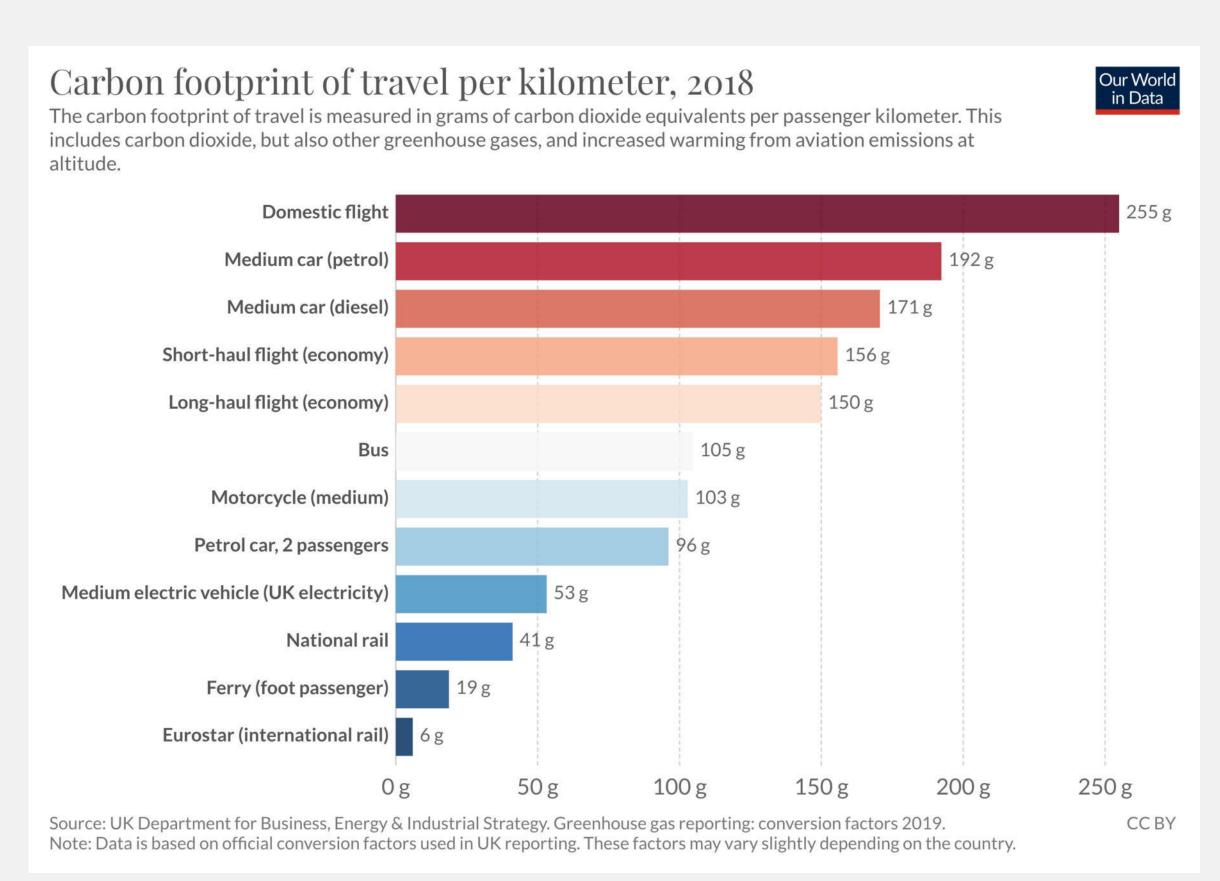




Farben (korrekt) nutzen



Originalgrafik mit einer zufälligen kategoriellen Farbpalette



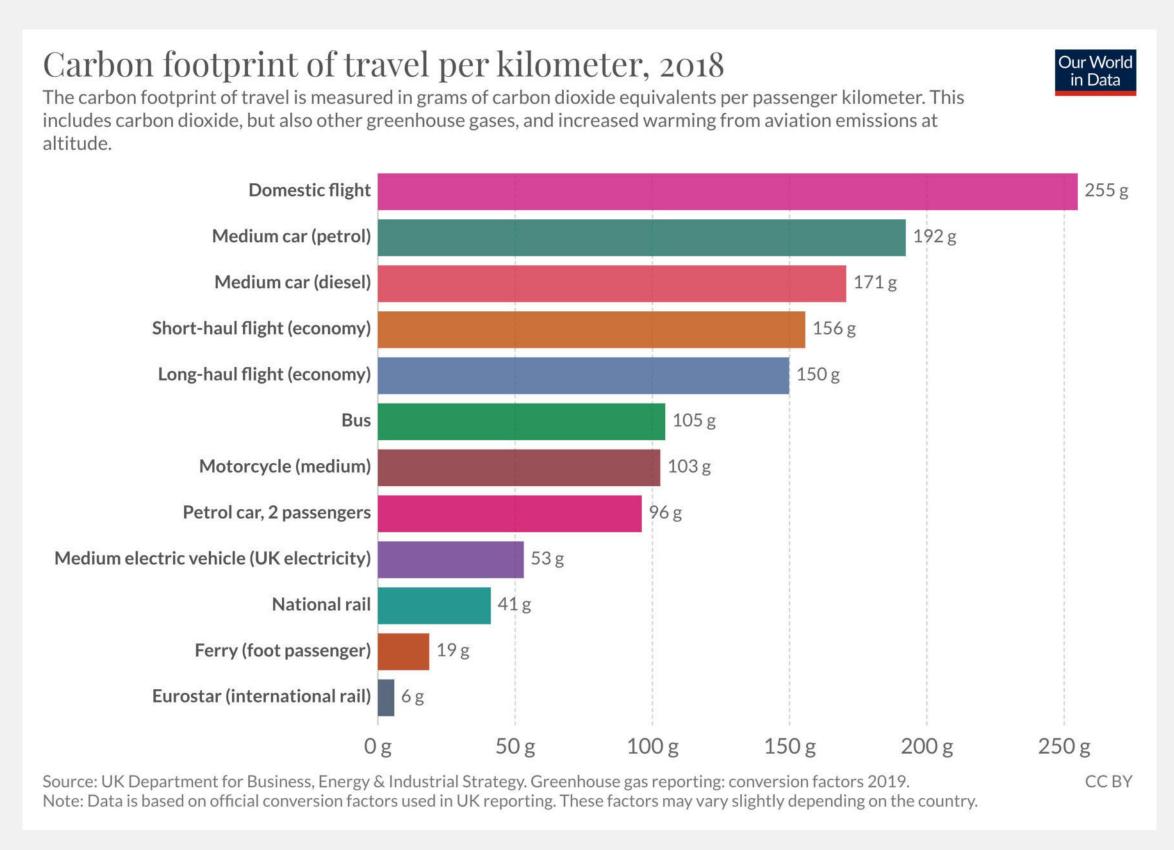
Überarbeitete Grafik mit einer divergierenden Farbpalette



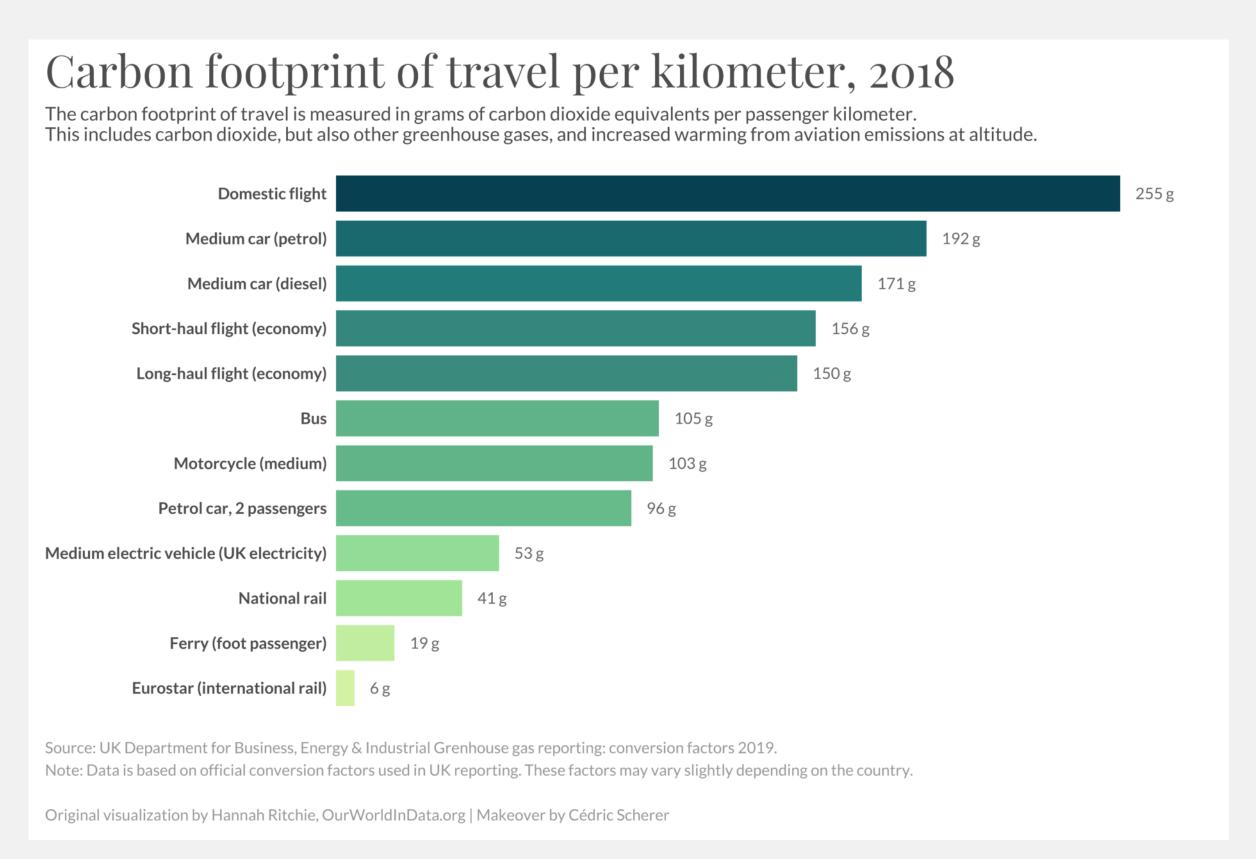




Farben (korrekt) nutzen



Original graphic with a random categorical palette



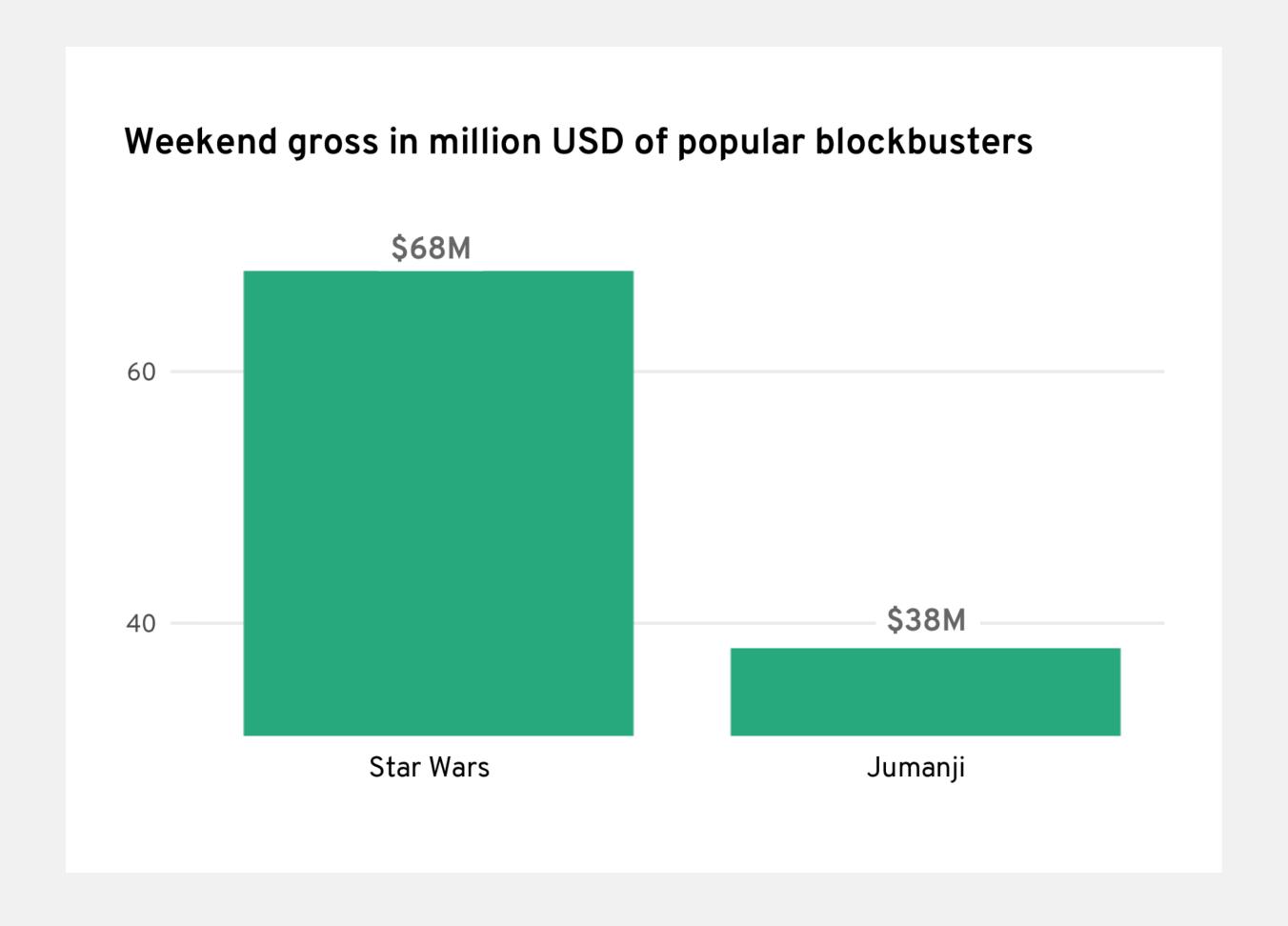
Makeover using a continuous palette







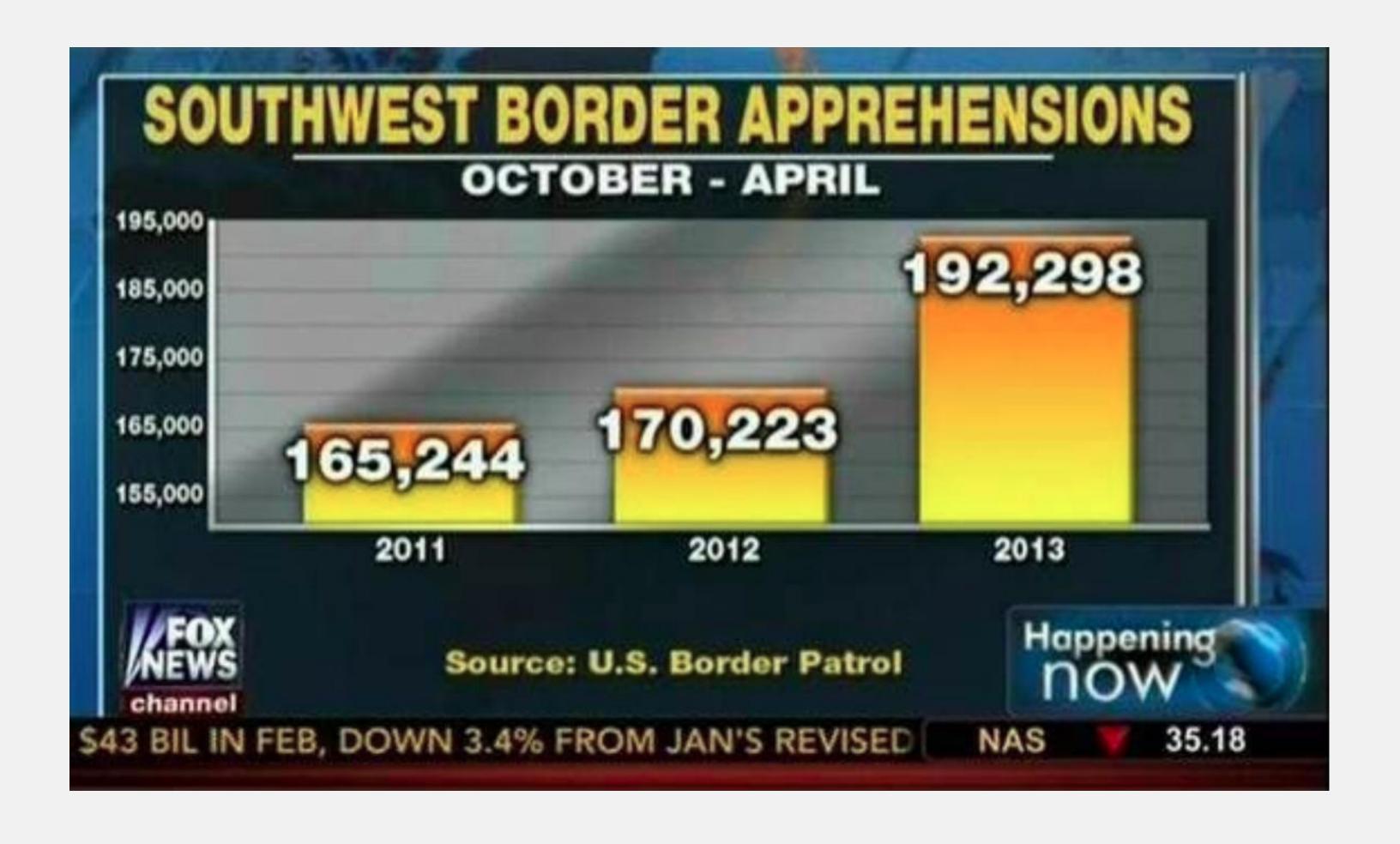
Immer bei Null starten



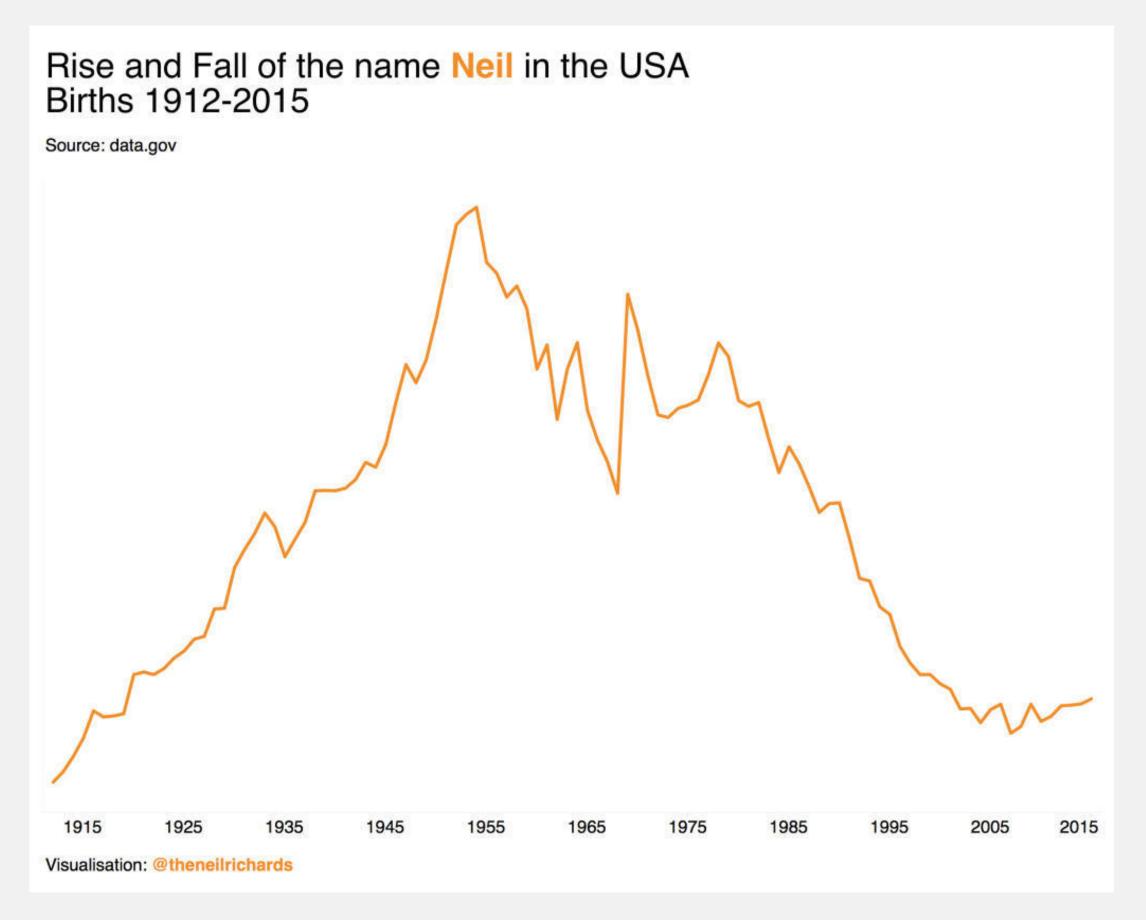


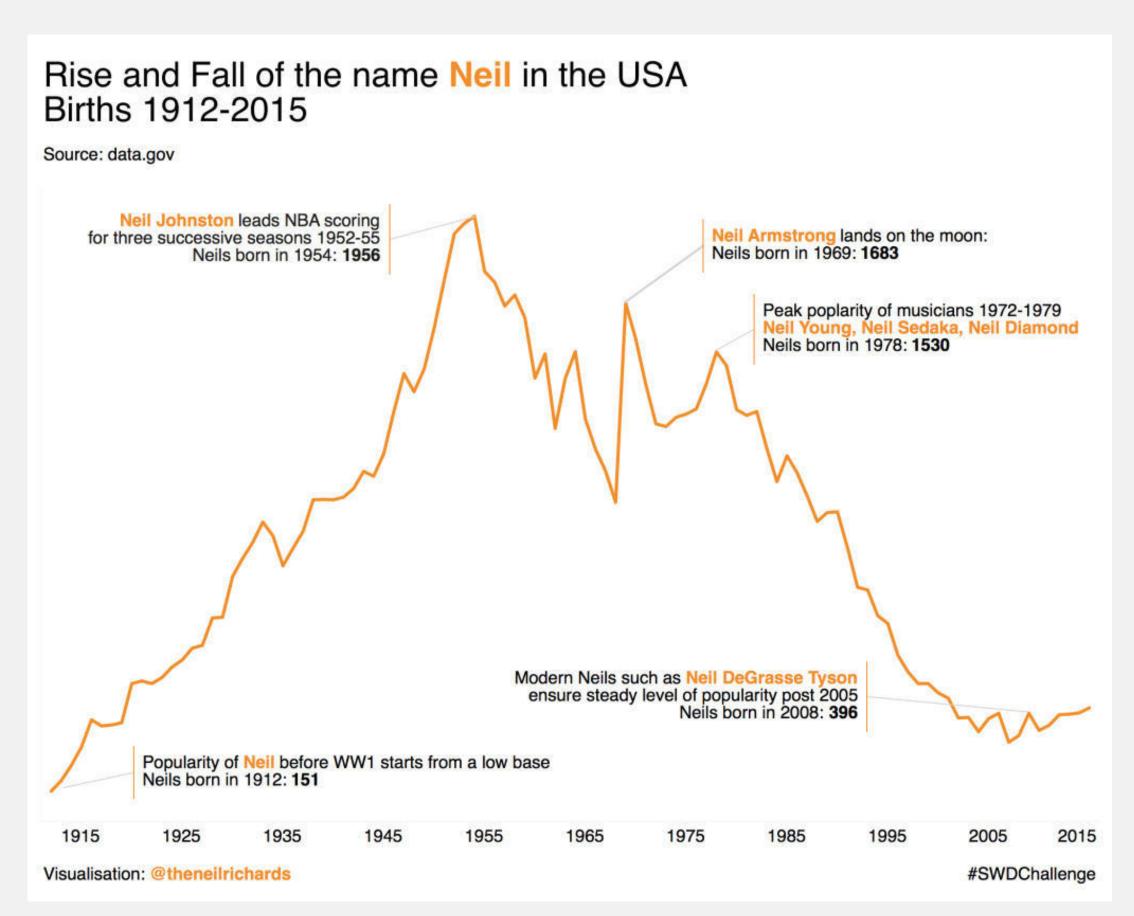


Immer bei Null starten



Die Wirkung von Beschriftungen



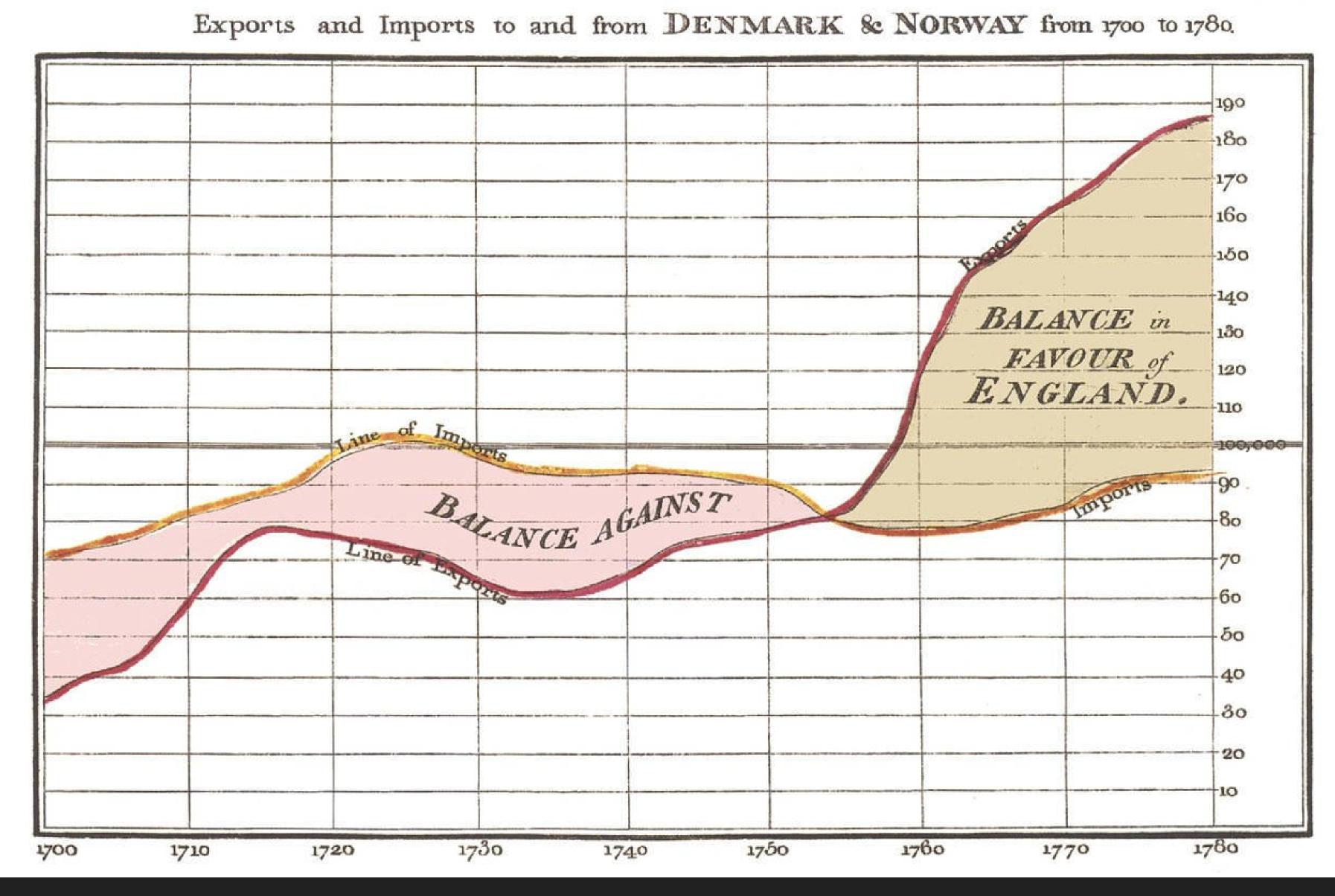


"Is white space always your friend?" von Neil Richards





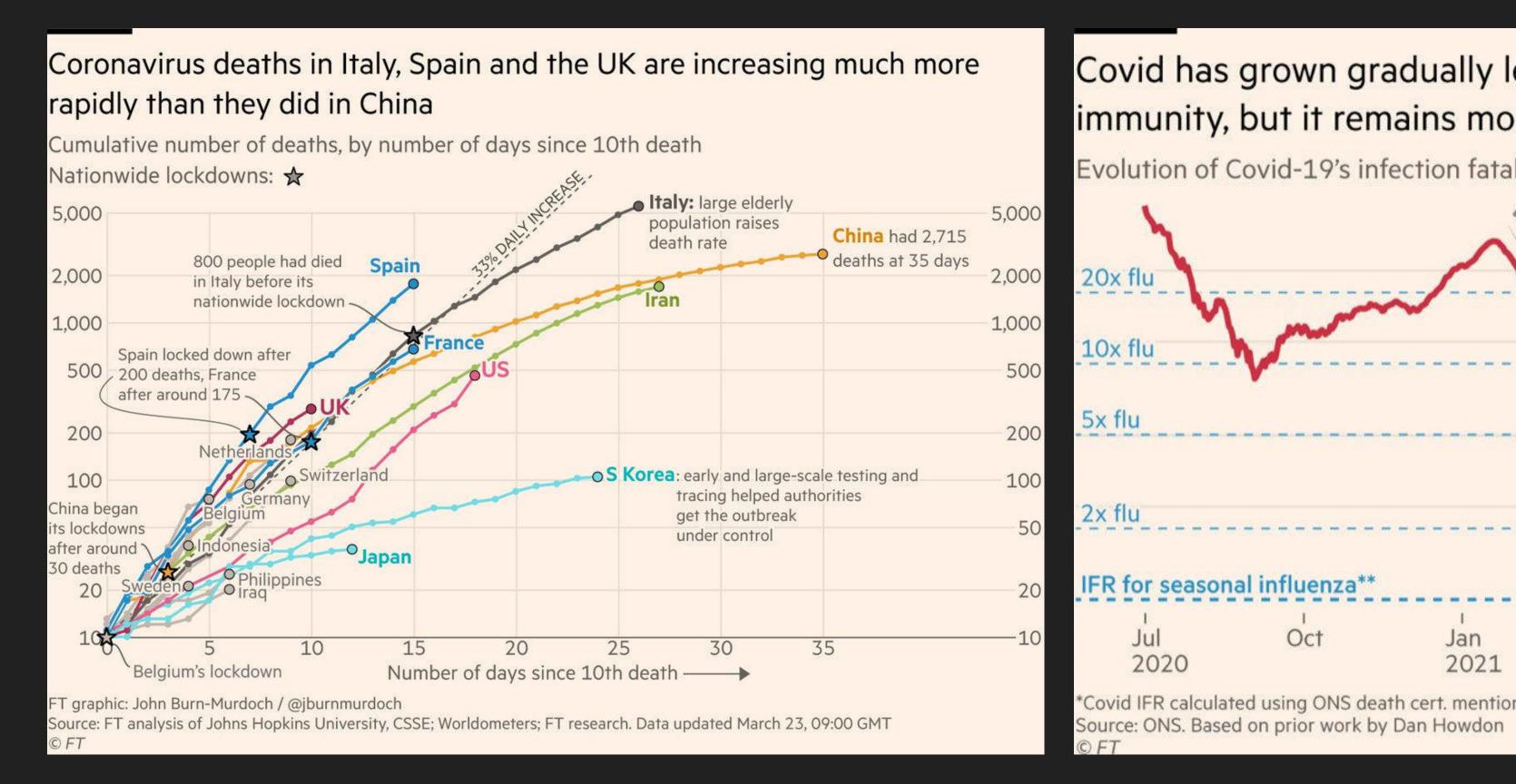




Zeitreihe mit Beschriftungen von William Playfair from "The Commercial and Political Atlas and Statistical Breviary" (1786)

"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 enourmous value for non-chart people."

~ John Burn-Murdoch, Financial Times





FT graphic: John Burn-Murdoch / @jburnmurdoch

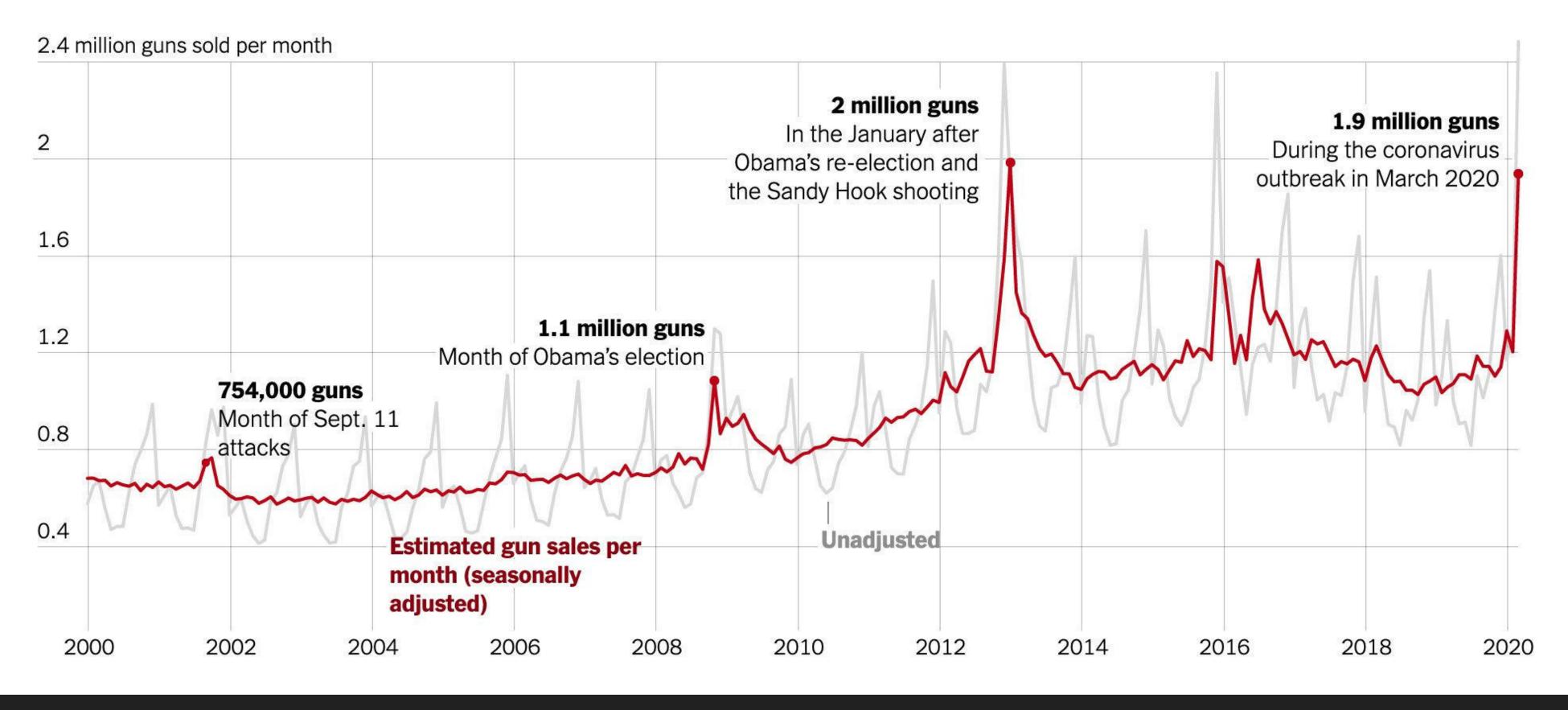






About 2 Million Guns Were Sold in the U.S. as Virus Fears Spread

By Keith Collins and David Yaffe-Bellany April 1, 2020



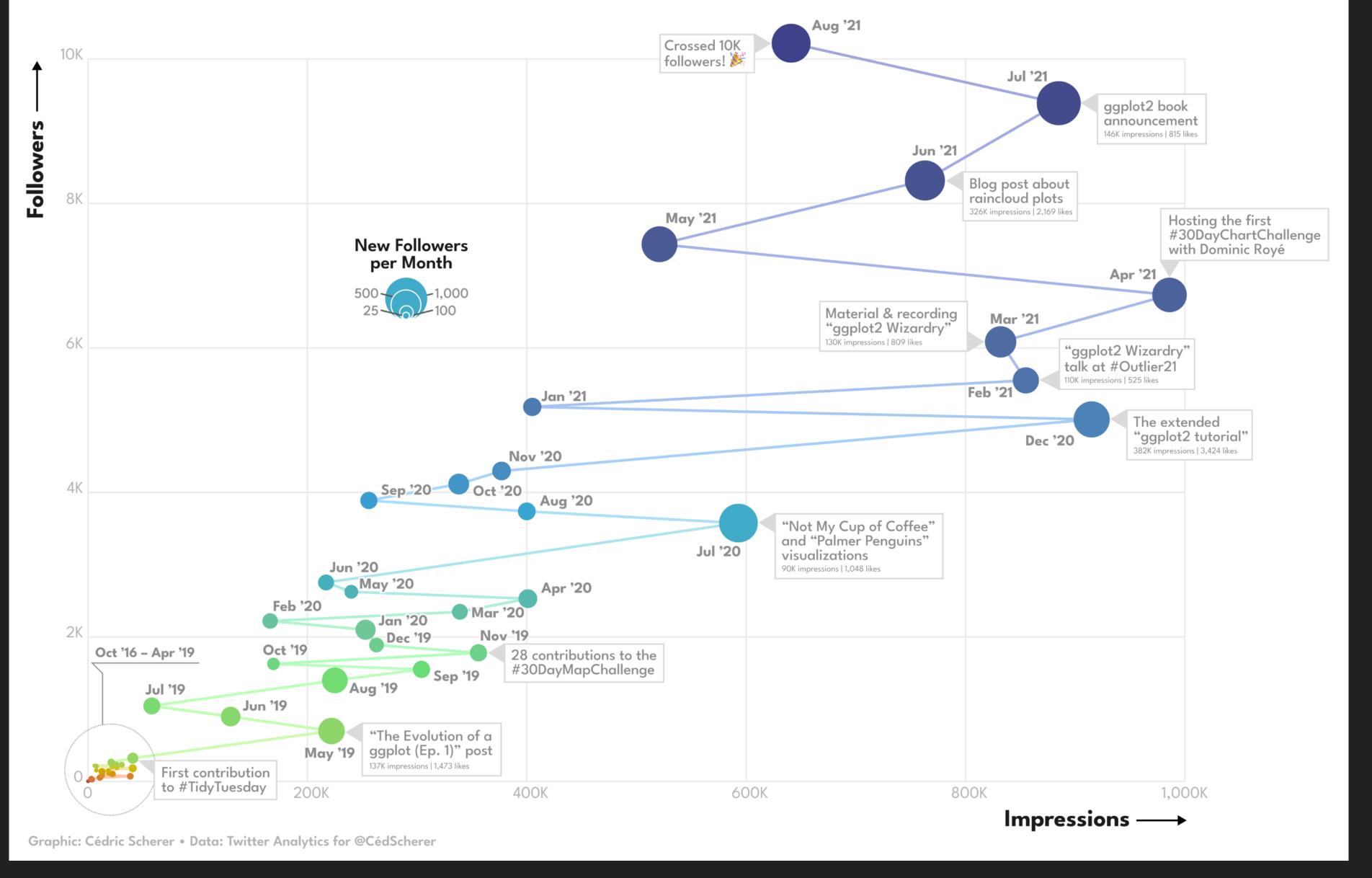
"About 2 Million Guns Were Sold in the U.S: as Virus Feats Spread" von Keith Collins and David Yaffe-Bellany (New York Times)







My Road to 10K on Twitter — Thank You All for Following!



ZUSAMMENFASSUNG





Information

Verstehe deine Daten und sei genau.

Story

Sei dir über die Botschaft der Daten im Klaren.

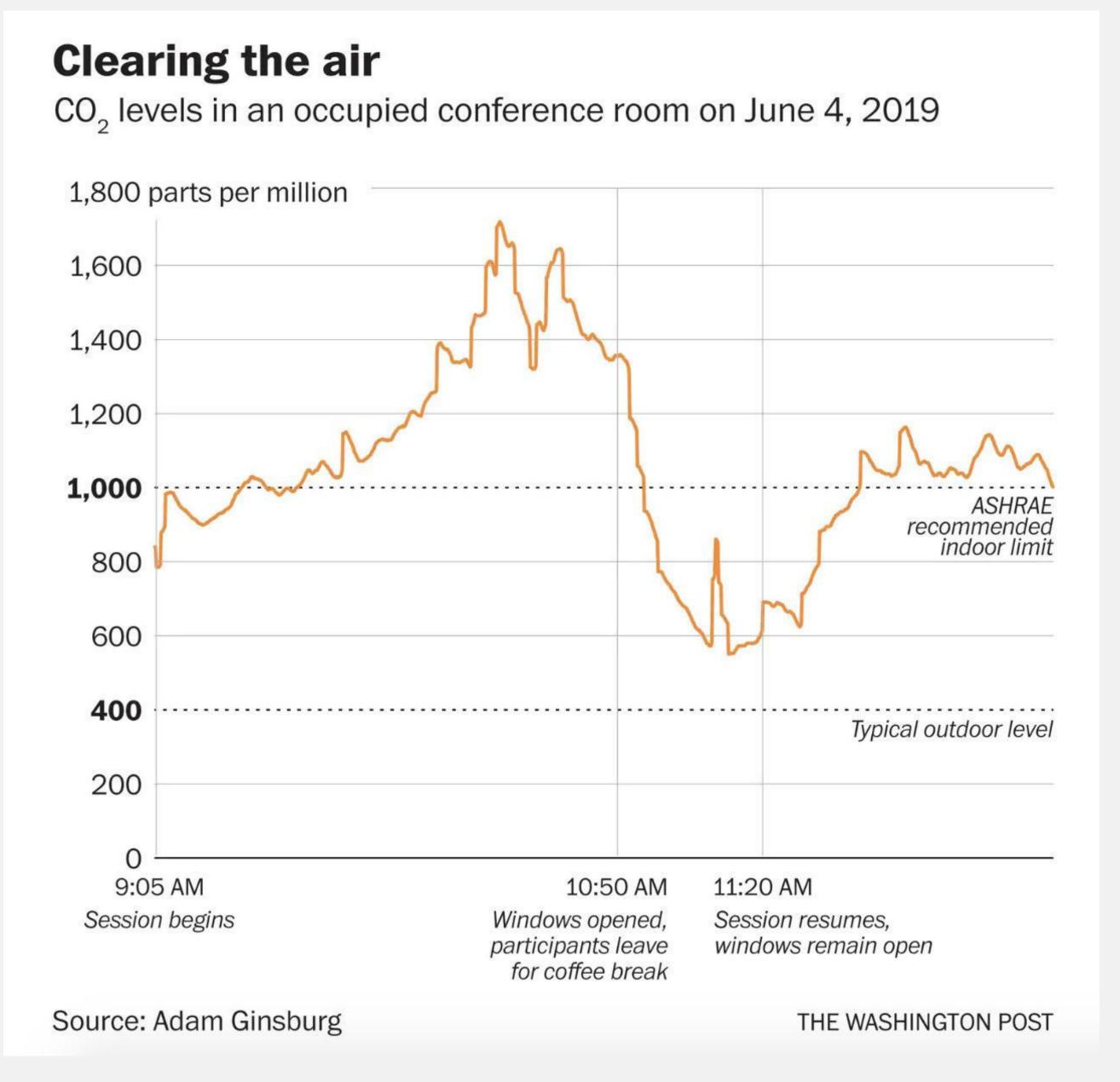
Goa

Wähle geeignete Grafiken, um die Geschichte zu erzählen.

Visual Form

Folge Grundsätzen aus Design und Datenvisualisierung.





"Clearin the Air" von Adam Ginsburg (Washington Post)









...mit <u>Gedanken von Francis Gagnon (Voilà)</u>





Gestaltung für das Zielpublikum

- · Wähle von Diagrammen nach dem Ziel aus und nicht (nur) nach Tradition oder Neuartigkeit.
- · Stelle sicher, dass die Grafiken für jeden zugänglich sind (Farben, Lesbarkeit, ALT-Text).
- · Verwende visuelle Kontraste, um wichtige Informationen hervorzuheben.
- Füge aussagekräftige Beschriftungen hinzu.





Gestaltung für das Zielpublikum

- · Wähle von Diagrammen nach dem Ziel aus und nicht (nur) nach Tradition oder Neuartigkeit.
- Stelle sicher, dass die Grafiken für jeden zugänglich sind (Farben, Lesbarkeit, ALT-Text).
- · Verwende visuelle Kontraste, um wichtige Informationen hervorzuheben.
- Füge aussagekräftige Beschriftungen hinzu.

Ehrlichkeit zählt

- · Zeige wann immer möglich die Verteilung der Rohdaten.
- Schneide Balkendiagramme nicht ab, füge abgeschnittenen Achsen Abstände hinzu.
- · Achte auf eine einheitliche Achsenskalierung (insbesondere bei "Small Multiples").





Gestaltung für das Zielpublikum

- · Wähle von Diagrammen nach dem Ziel aus und nicht (nur) nach Tradition oder Neuartigkeit.
- · Stelle sicher, dass die Grafiken für jeden zugänglich sind (Farben, Lesbarkeit, ALT-Text).
- · Verwende visuelle Kontraste, um wichtige Informationen hervorzuheben.
- · Füge aussagekräftige Beschriftungen hinzu.

Ehrlichkeit zählt

- · Zeige wann immer möglich die Verteilung der Rohdaten.
- · Schneide Balkendiagramme nicht ab, füge abgeschnittenen Achsen Abstände hinzu.
- · Achte auf eine einheitliche Achsenskalierung (insbesondere bei "Small Multiples").

Hilfestellung leisten

- · Verwende direkte Anmerkungen anstelle von/zusätzlich zu Beschriftungen und Legenden.
- Ordne Daten, entweder nach ihrem Wert oder nach ihrer intrinischen Ordnung.
- · Konzentriere dich auf die Hauptaussage und versuche die Datenkomplexität zu reduzieren.
- · Decke Informationen Schritt für Schritt auf (falls möglich).







