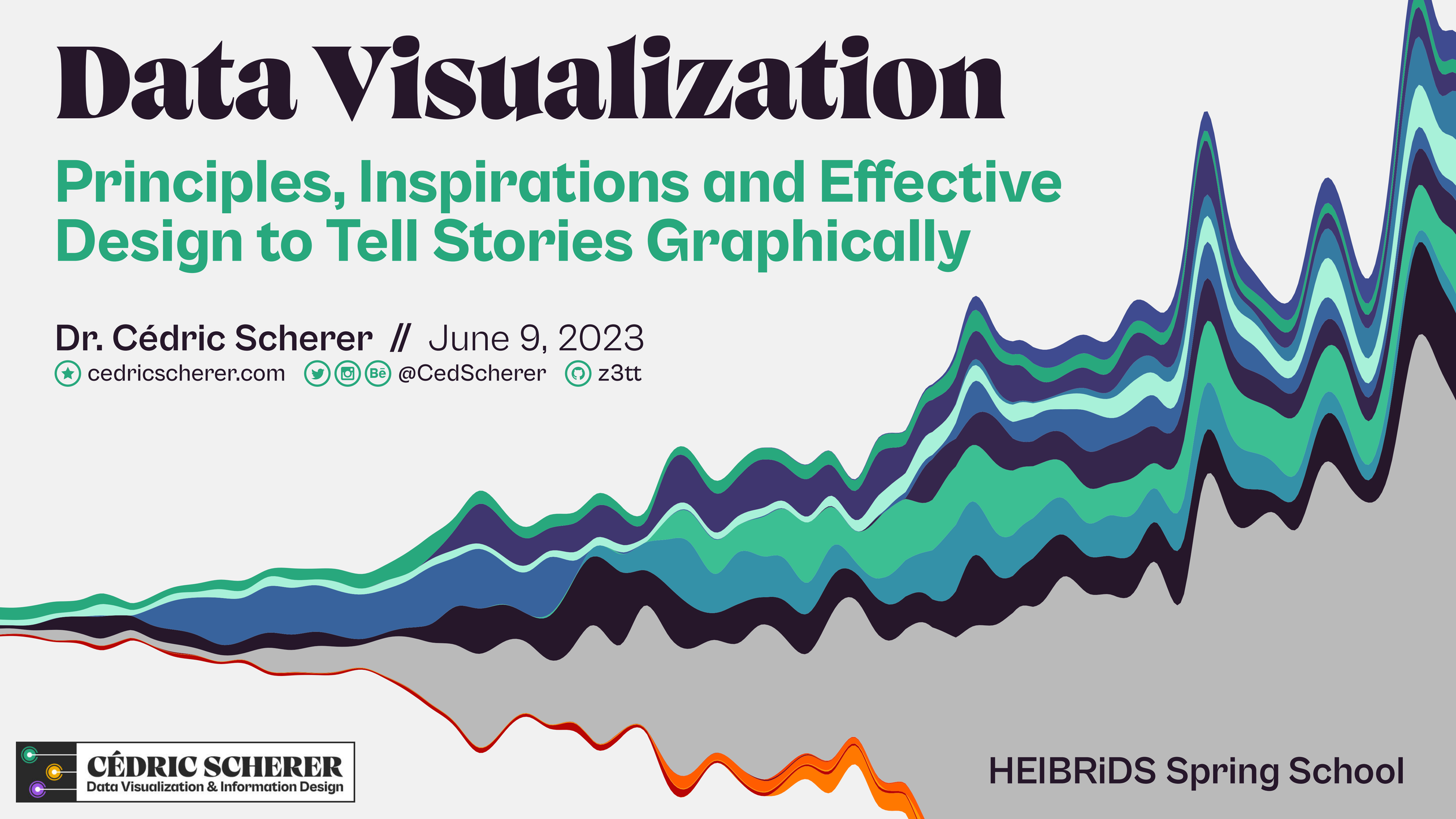
Data Visualization

Principles, Inspirations and Effective Design to Tell Stories Graphically

Dr. Cédric Scherer // June 9, 2023

cedricscherer.com @CedScherer z3tt



CÉDRIC SCHERER
Data Visualization & Information Design

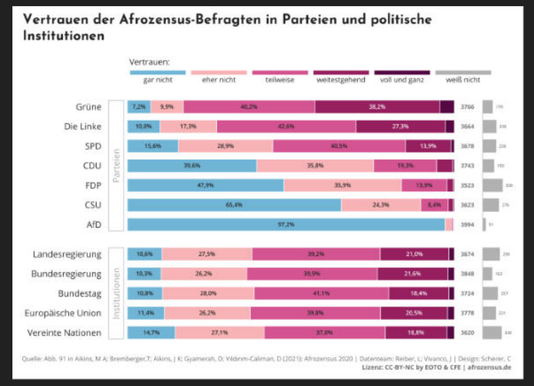
HEIBRiDS Spring School

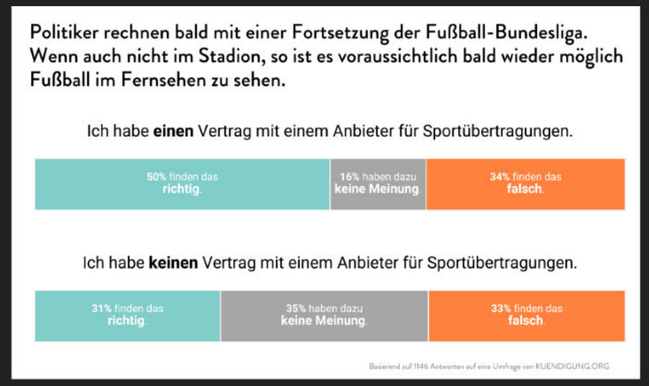
CÉDRIC SCHERER

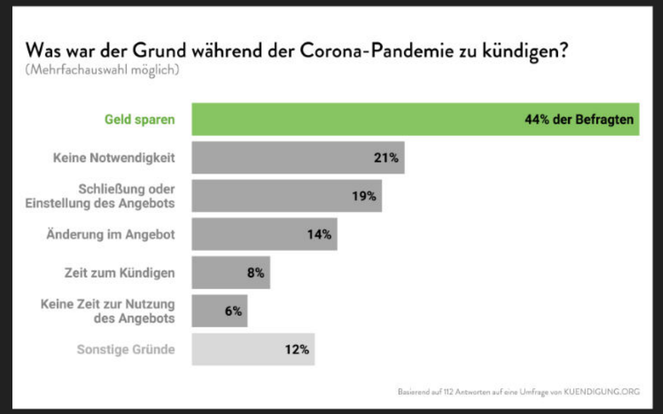
Data Visualization & Information Design

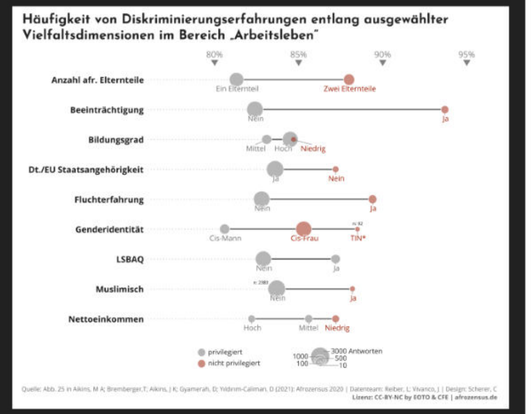
Consulting **Coaching** **Coding**

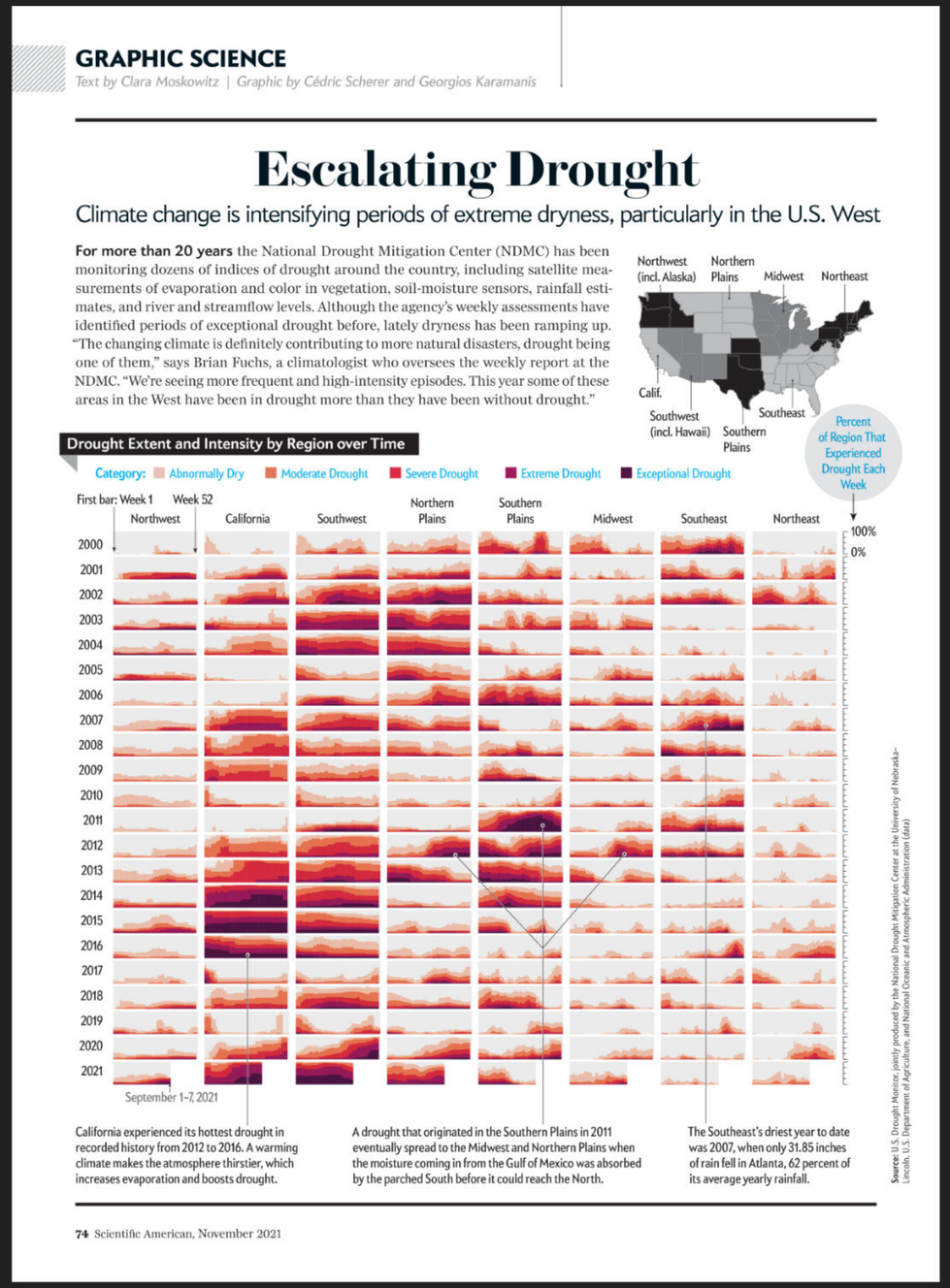
cedricscherer.com @CedScherer z3tt











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."

Northwest (incl. Alaska)
Northern Plains
Midwest
Northeast
Calif.
Southwest (incl. Hawaii)
Southern Plains
Southeast

Drought Extent and Intensity by Region over Time

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

First bar: Week 1 | Week 52

Northwest | California | Southwest | Northern Plains | Southern Plains | Midwest | Southeast | Northeast

Percent of Region That Experienced Drought Each Week: 100% / 0%

2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021

September 1–7, 2021

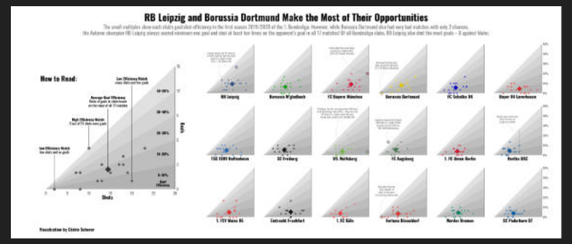
California experienced its hottest drought in recorded history from 2012 to 2016. A warming climate makes the atmosphere thirstier, which increases evaporation and boosts drought.

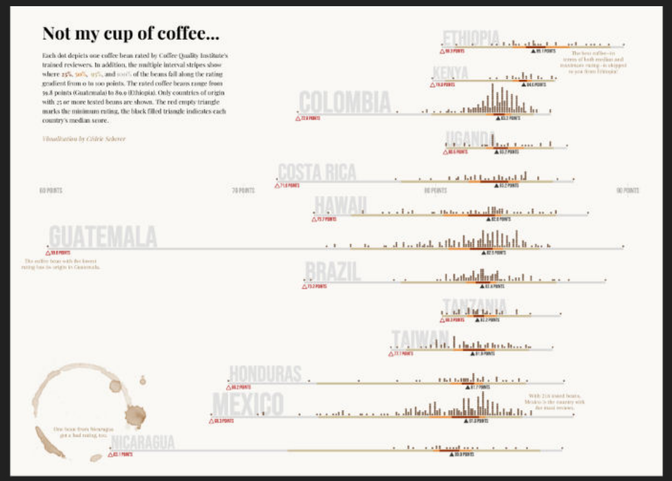
A drought that originated in the Southern Plains in 2011 eventually spread to the Midwest and Northern Plains when the moisture coming in from the Gulf of Mexico was absorbed by the parched South before it could reach the North.

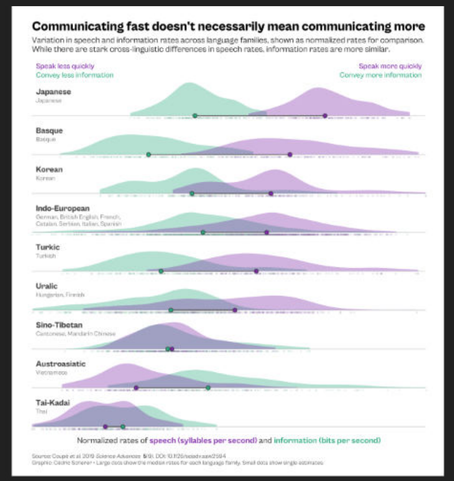
The Southeast's driest year to date was 2007, when only 31.85 inches of rain fell in Atlanta, 62 percent of its average yearly rainfall.

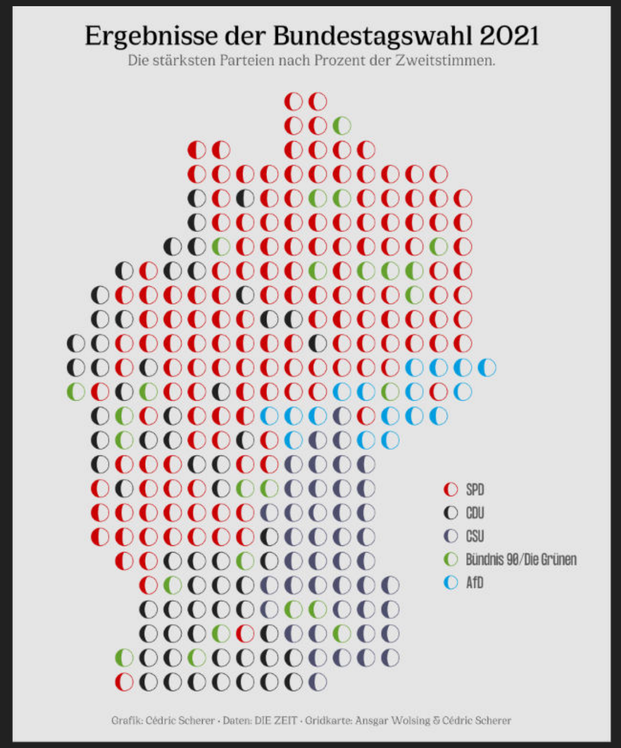
Source: U.S. Drought Monitor, jointly produced by the National Drought Mitigation Center at the University of Nebraska–Lincoln, U.S. Department of Agriculture, and National Oceanic and Atmospheric Administration (data)

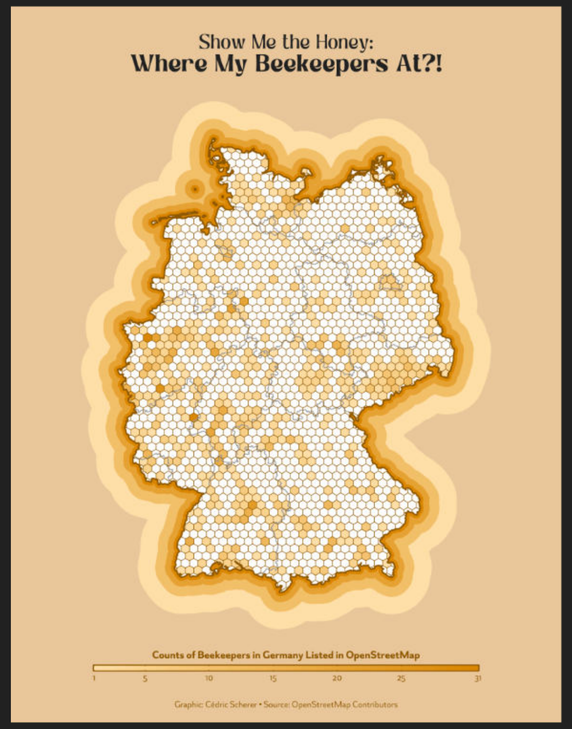
74 Scientific American, November 2021

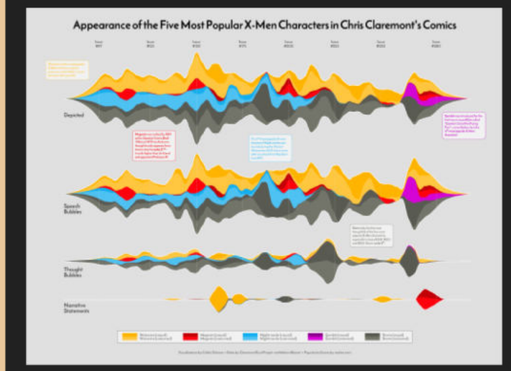

RB Leipzig and Borussia Dortmund Make the Most of Their Opportunities
How to Read:

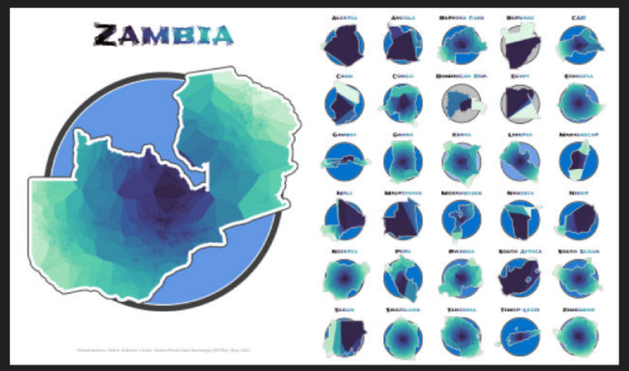

Not my cup of coffee...
ETHIOPIA
KENYA
COLOMBIA
BRAZIL
COSTA RICA
HAWAII
GUATEMALA
BRAZIL
TANZANIA
TAIWAN
HONDURAS
MEXICO
NICARAGUA

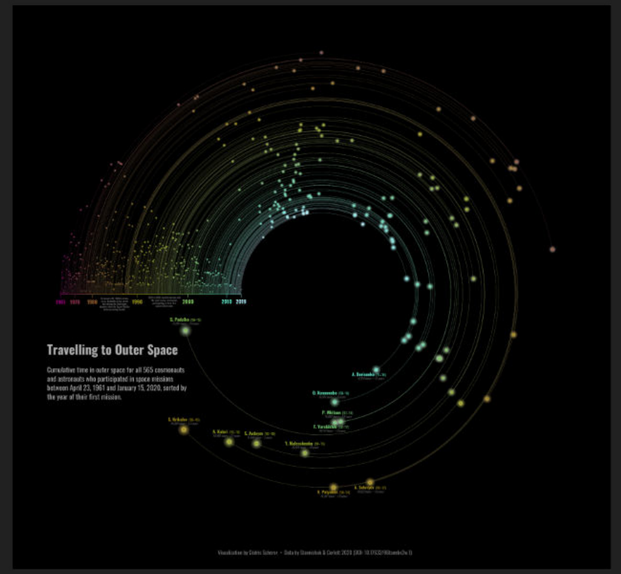

Communicating fast doesn't necessarily mean communicating more
Speak less quickly
Convey less information
Speak more quickly
Convey more information
Japanese
Basque
Korean
Indo-European
Turkic
Uralic
Sino-Tibetan
Austroasiatic
Tai-Kadai
Normalized rates of speech (syllables per second) and information (bits per second)

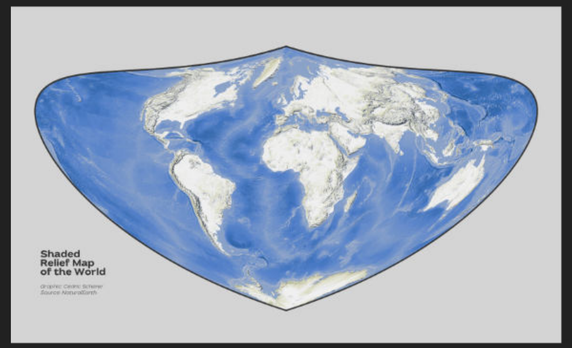

Ergebnisse der Bundestagswahl 2021
Die stärksten Parteien nach Prozent der Zweitstimmen.
SPD
CDU
CSU
Bündnis 90/Die Grünen
AfD
Grafik: Cédric Scherer · Daten: DIE ZEIT · Gridkarte: Ansgar Wolsing & Cédric Scherer

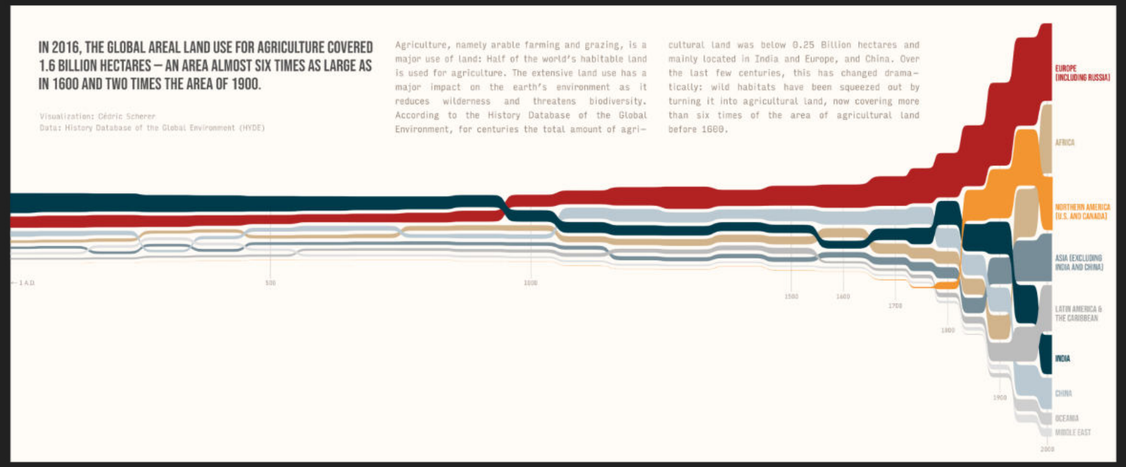

Show Me the Honey:
Where My Beekeepers At?!
Counts of Beekeepers in Germany Listed in OpenStreetMap
1
5
10
15
20
25
31
Graphic: Cédric Scherer • Source: OpenStreetMap Contributors


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


ZAMBIA


Travelling to Outer Space


Shaded
Relief Map
of the World


IN 2016, THE GLOBAL AREAL LAND USE FOR AGRICULTURE COVERED 1.6 BILLION HECTARES — AN AREA ALMOST SIX TIMES AS LARGE AS IN 1600 AND TWO TIMES THE AREA OF 1900.
Visualization: Cédric Scherer
Data: History Database of the Global Environment (HYDE)
Agriculture, namely arable farming and grazing, is a major use of land: Half of the world's habitable land is used for agriculture. The extensive land use has a major impact on the earth's environment as it reduces wilderness and threatens biodiversity. According to the History Database of the Global Environment, for centuries the total amount of agricultural land was below 0.25 Billion hectares and mainly located in India and Europe, and China. Over the last few centuries, this has changed dramatically: wild habitats have been squeezed out by turning it into agricultural land, now covering more than six times of the area of agricultural land before 1600.
EUROPE (INCLUDING RUSSIA)
AFRICA
NORTHERN AMERICA (U.S. AND CANADA)
ASIA (EXCLUDING INDIA AND CHINA)
LATIN AMERICA & THE CARIBBEAN
INDIA
CHINA
OCEANIA
MIDDLE EAST

cedricscherer.com

Blog Gallery About Me Links

CÉDRIC SCHERER

Data Visualization & Information Design

Hi, I am Cédric 👋

Data Visualization Designer, Consultant and Instructor for Engaging and Effective Graphical Storytelling.

→ Read more about me → Schedule a discovery call

The 30DayChartChallenge is Ready to Kick Off

The #30DayChartChallenge is a data visualization community challenge with the aim of creating a data visualization on a particular theme for each day in April. Read on for a behind-the-scenes look at the two previous years' challenges, then join us on social media (Twitter, Instagram, LinkedIn, and Mastodon) to participate!

Tuesday • March 28, 2023

2-Day Workshop on "Graphic Design with ggplot2" at rstudio::conf 2022

End of July, I had the honor to teach an in-person ggplot2 workshop at the rstudio::conf in Washington DC. All course resources are available on the course webpage featuring slides, hands-on R codes, recap notes, and exercises including prepared scripts and step-by-step

Always coding. Passionate about design. Worried about nature and inequality. Proud dad.

Quick Links

ggplot2 Tutorial
Evolution of a ggplot
rstudio::conf Workshop

Featured Tags

30DAYCHARTCHALLENGE
ANIMATIONS BERLIN
CHALLENGE COVID19
DATAVIZ DESIGN GGPLOT2

Data Visualization

is any graphical representation
of information and data.

Data Visualization

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

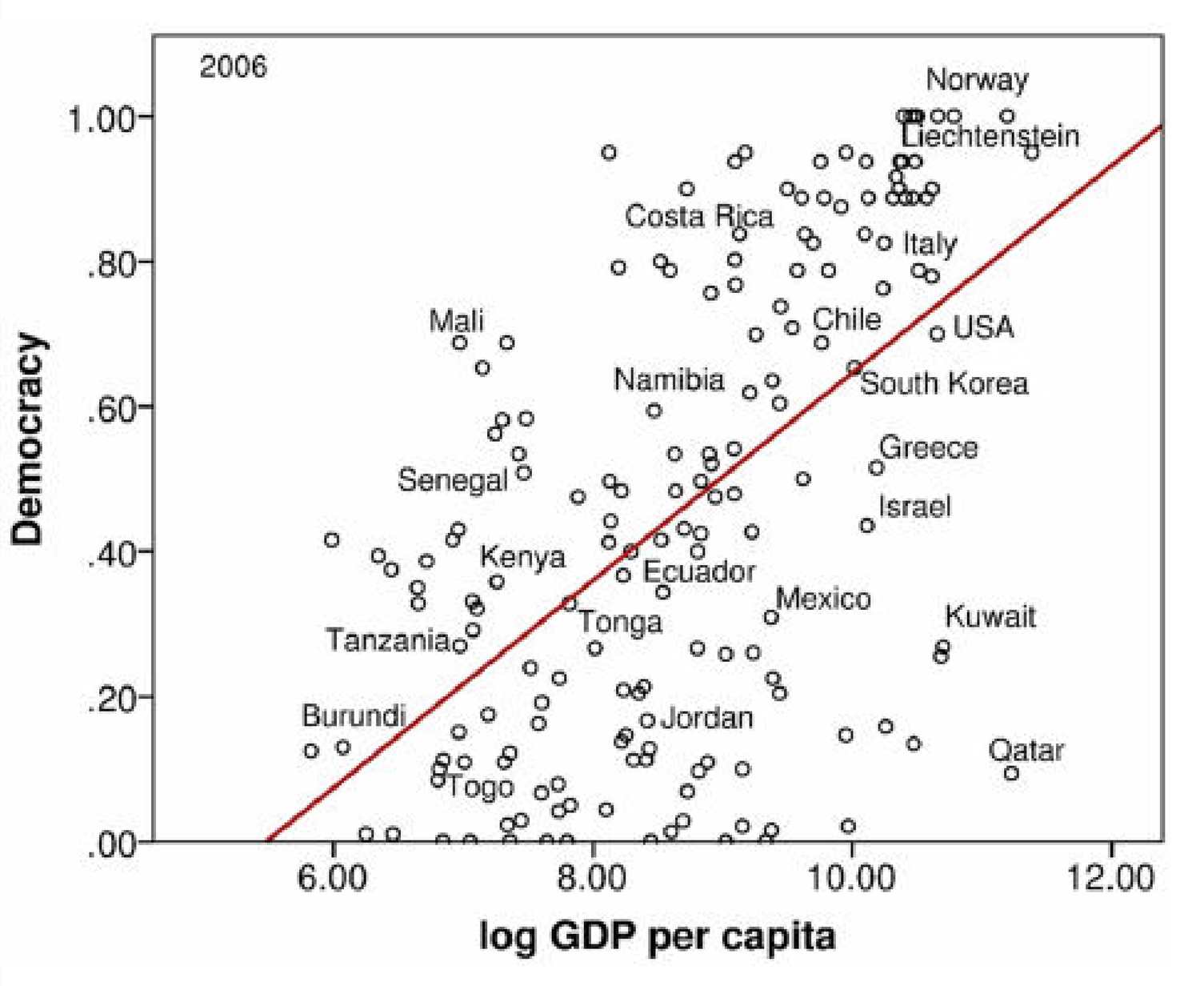
Data Visualization

converts information into visual forms as quantifiable features.

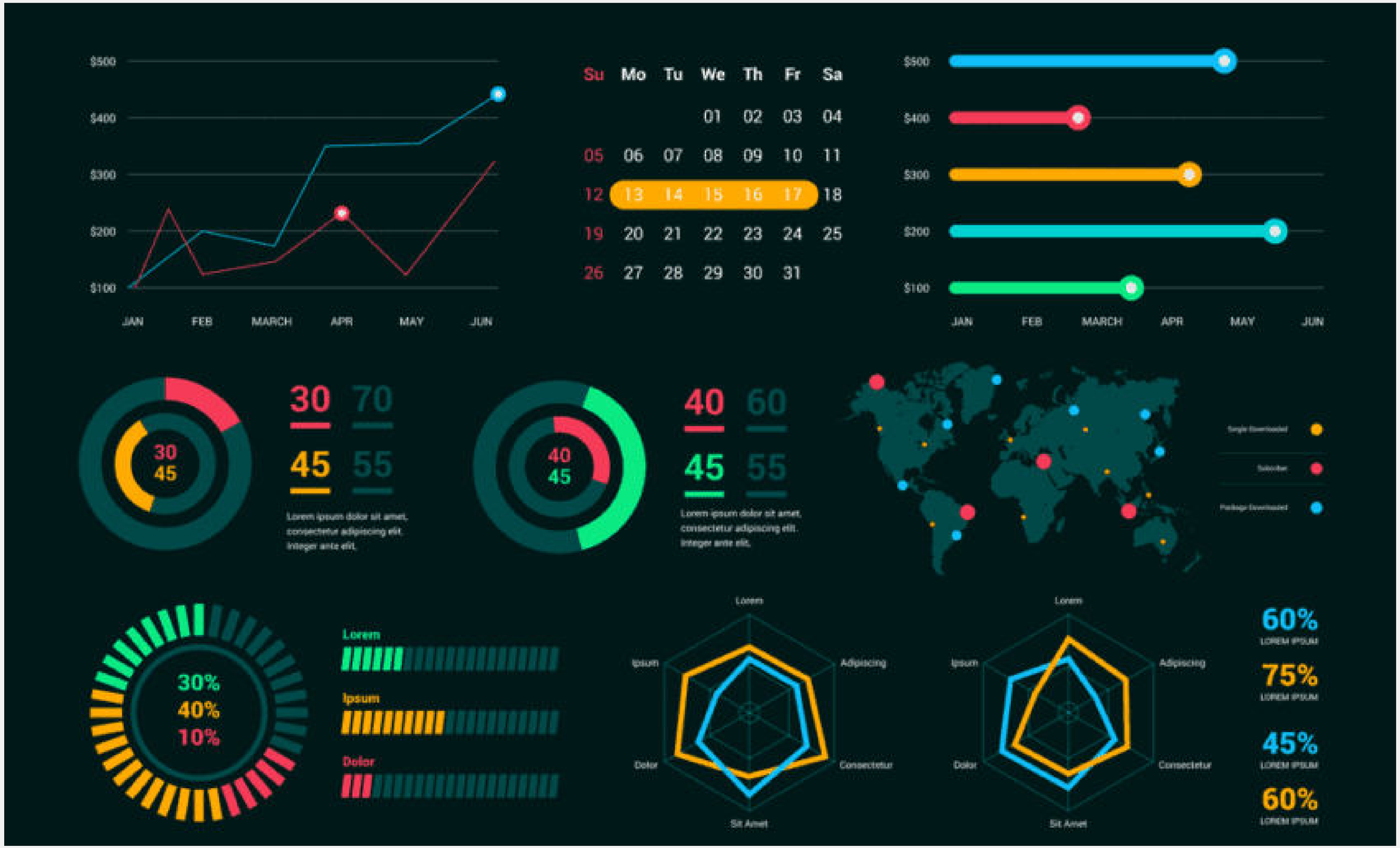
Data Visualization

is part art and part science.

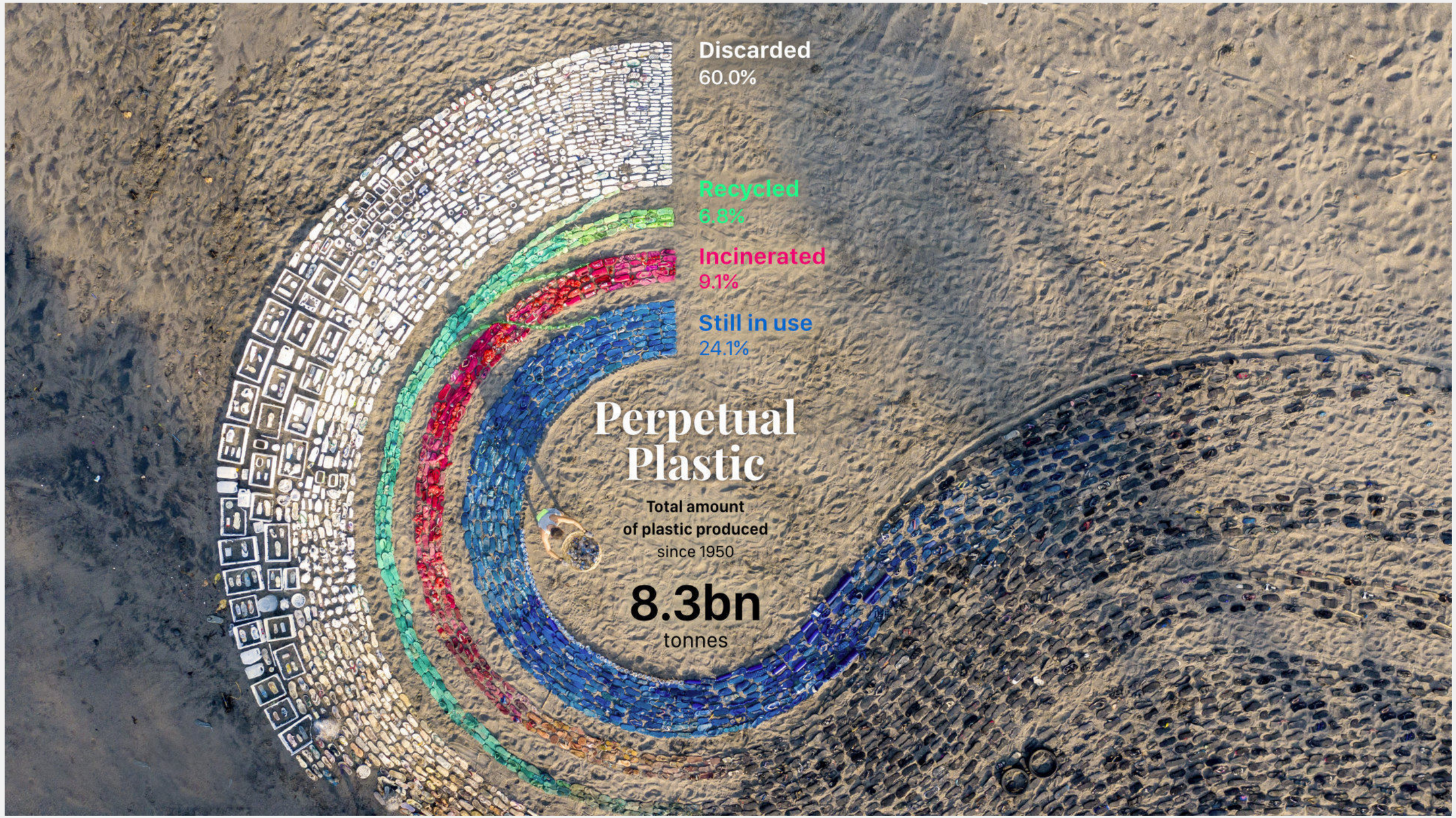
Source: eazybi.com



Source: Ranganathan et al. 2014

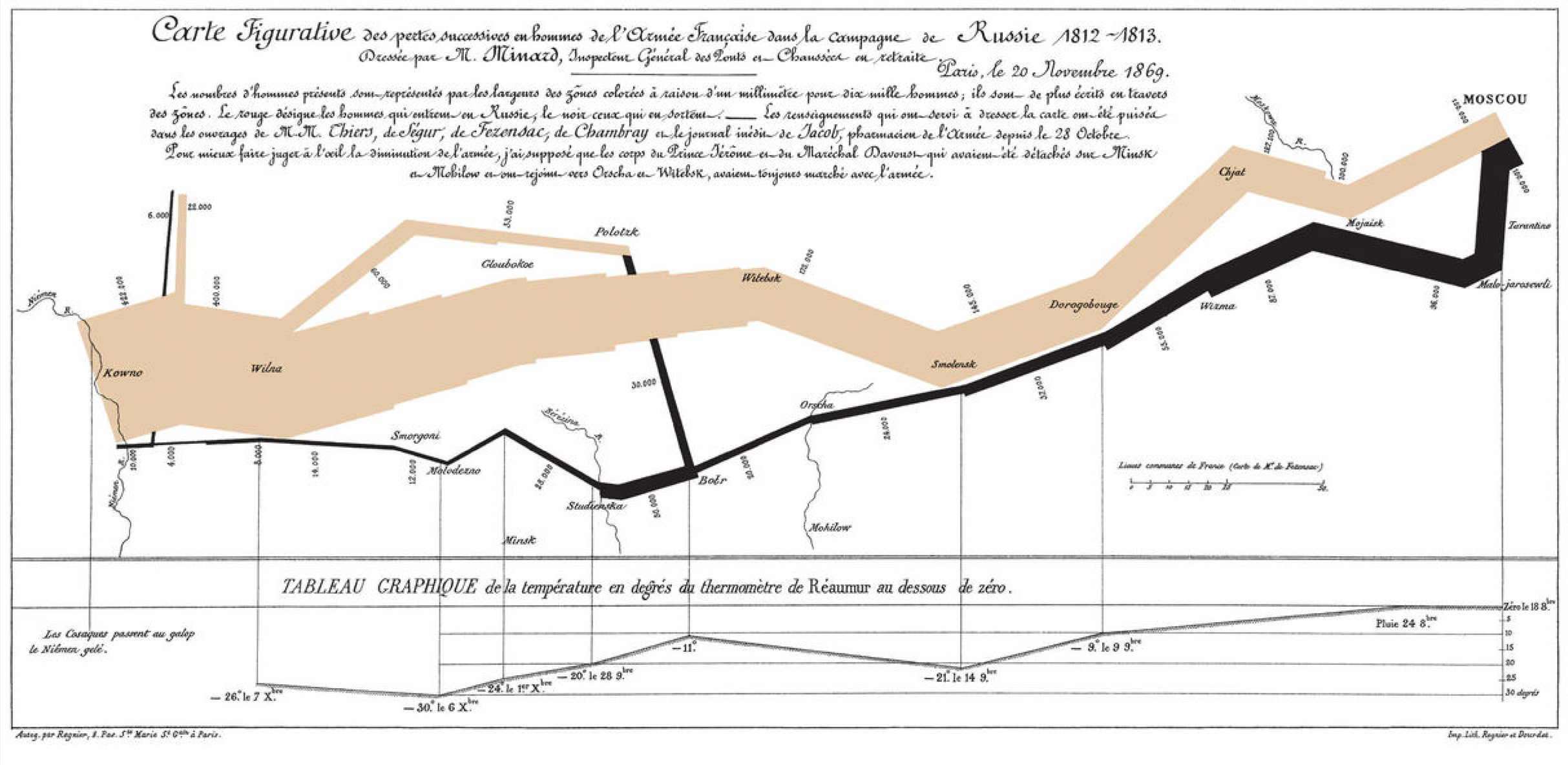


Source: datameer.com

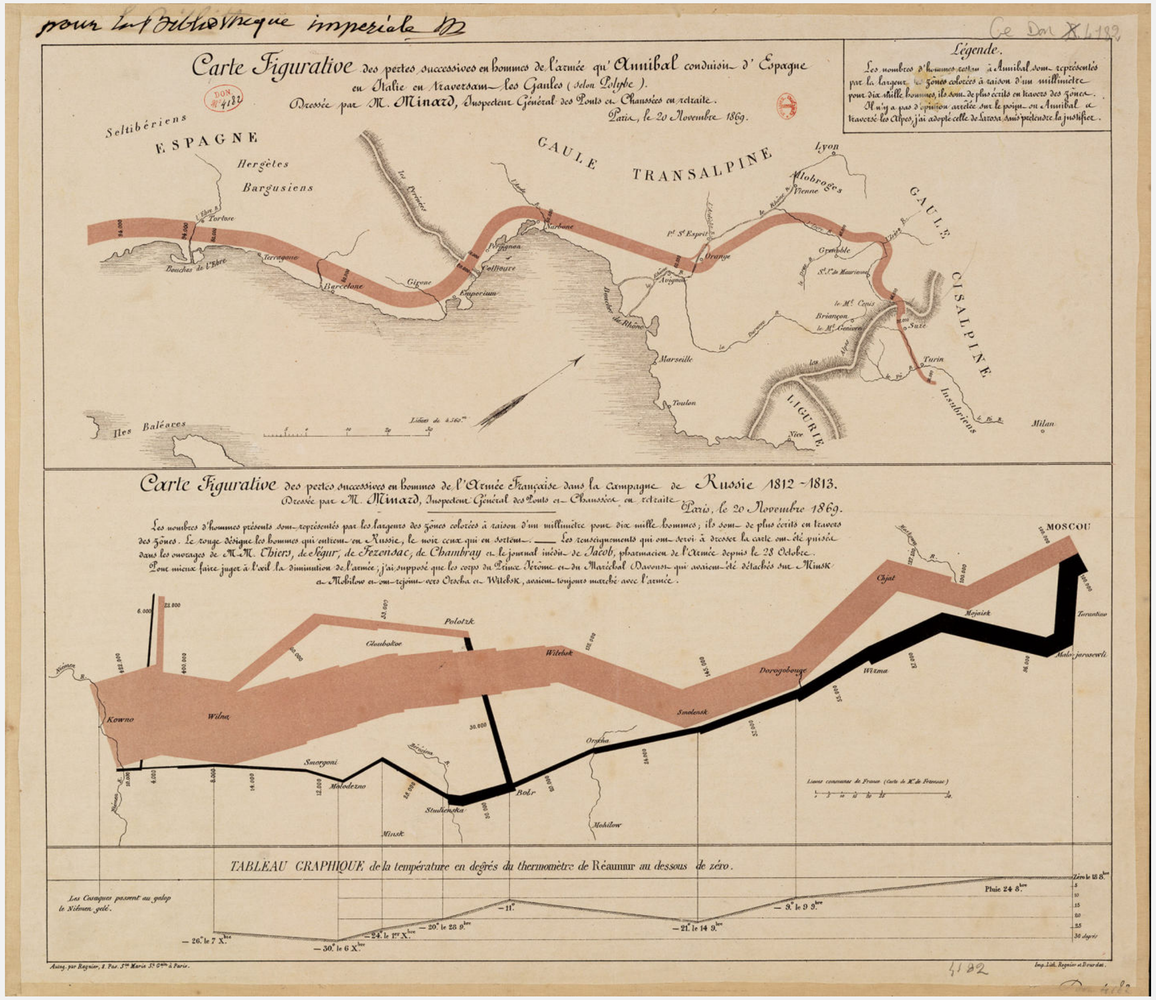


Source: "Perpetual Plastic" by Liina Klauss, Skye Morét and Moritz Stefaner

Source: *“Patchwork Kingdoms”* by Nadieh Bremer

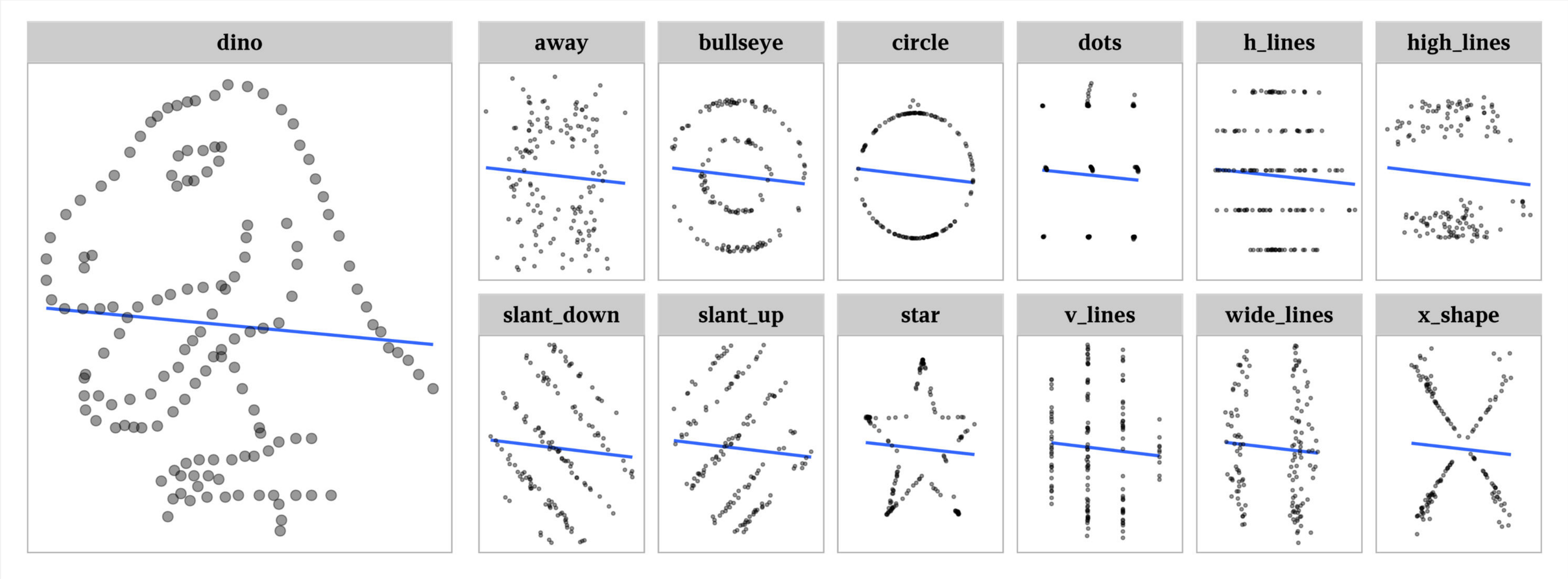


"Carte figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812–1813" by Charles Joseph Minard (1869)



"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)"* by Charles Joseph Minard (1869)

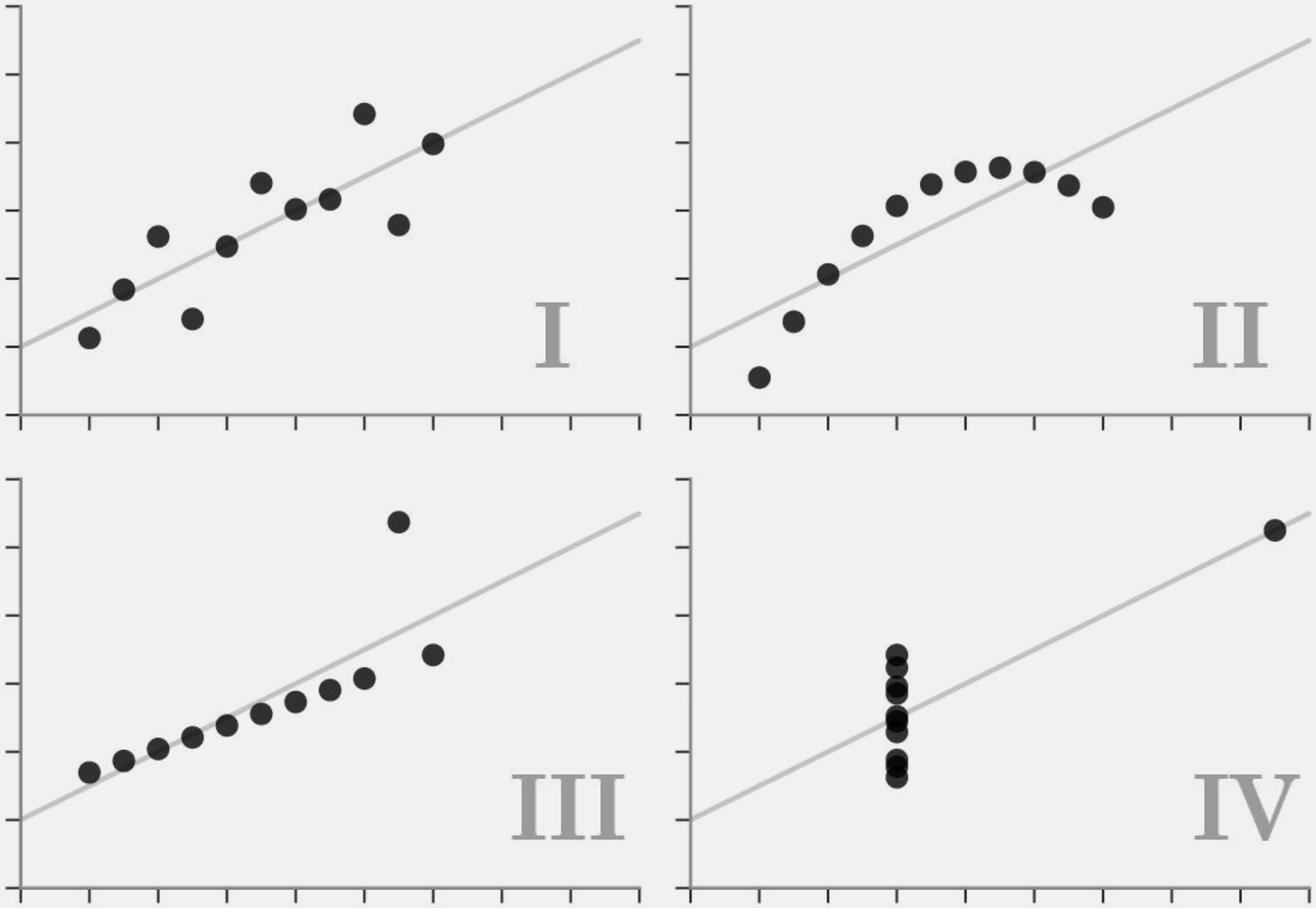
- shows the force levels of the armies of Hannibal (218 BC) and Napoleon (1812-1813), respectively
- some data visualization practioners call it (one of) **the best statistical drawings ever created**



"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

Anscombe's Quartet

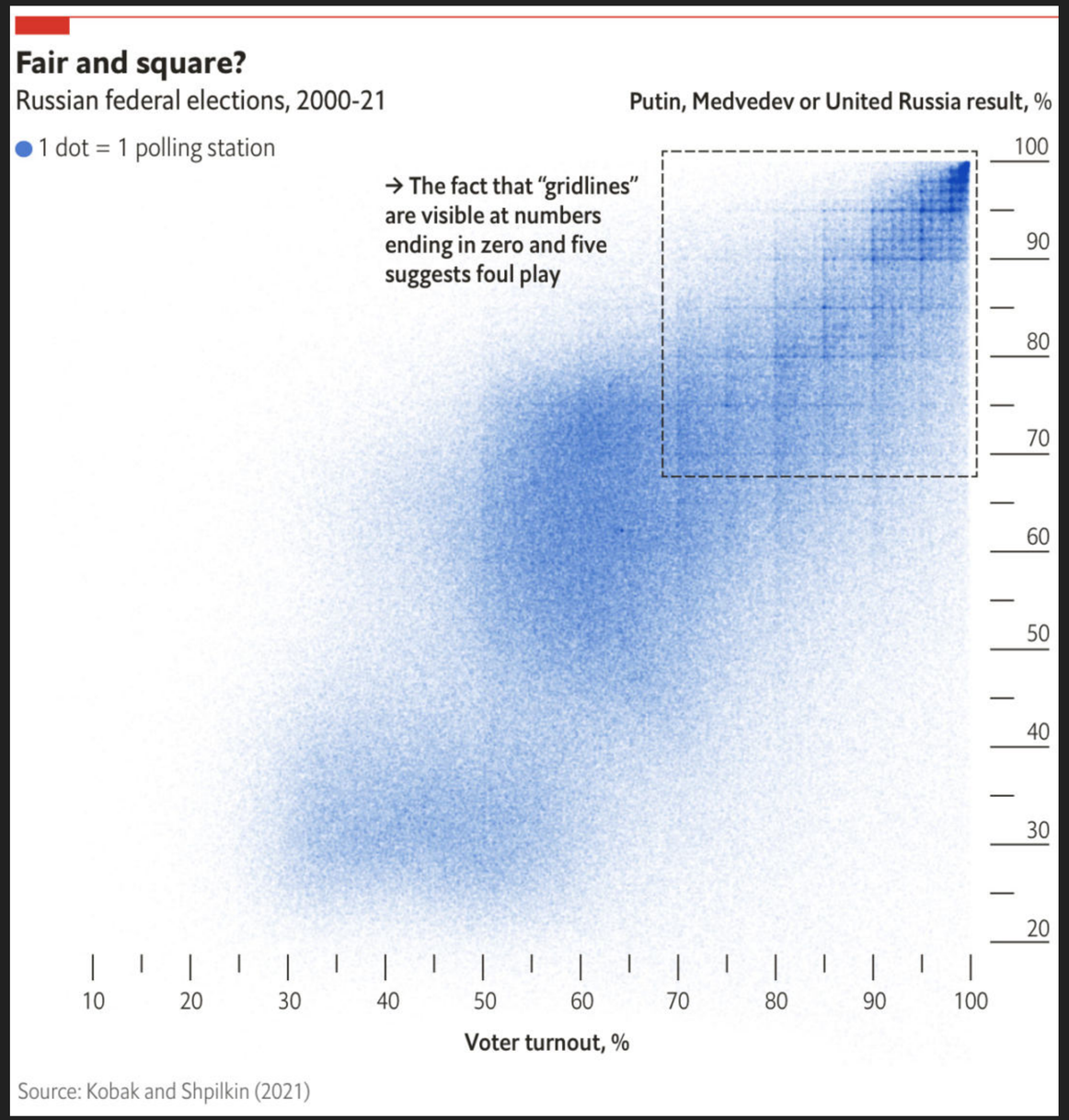
Each dataset has the same summary statistics
mean, standard deviation, and correlation
but are visually distinct.



"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

Visualize Your Data

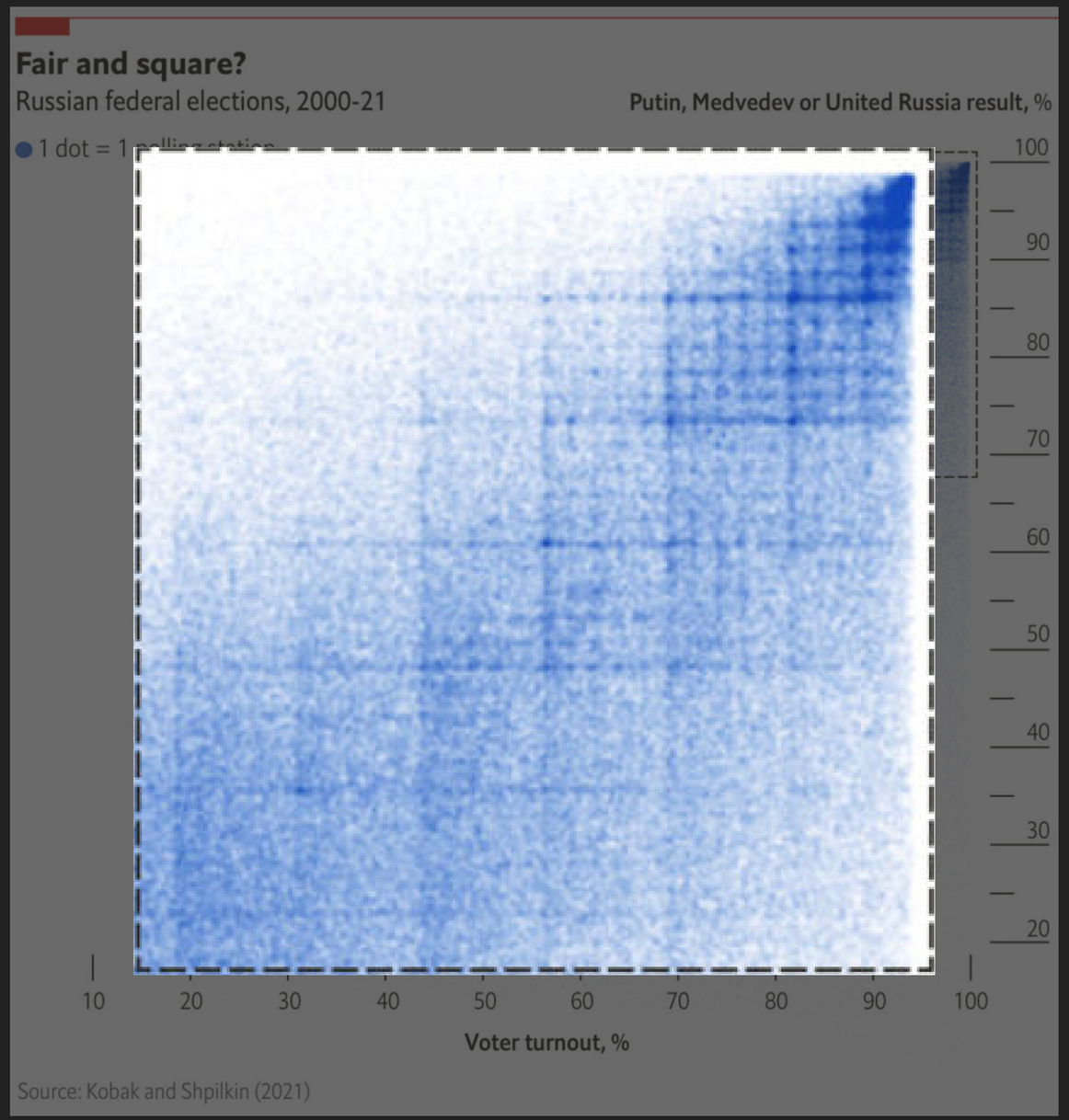
"When **Dmitry Kobak and Sergey Shpilkin** [...] analysed the results, they found that **an unusually high number of turnout and vote-share results were multiples of five** (eg, 50%, 55%, 60%), a tell-tale **sign of manipulation**."



Source: The Economist

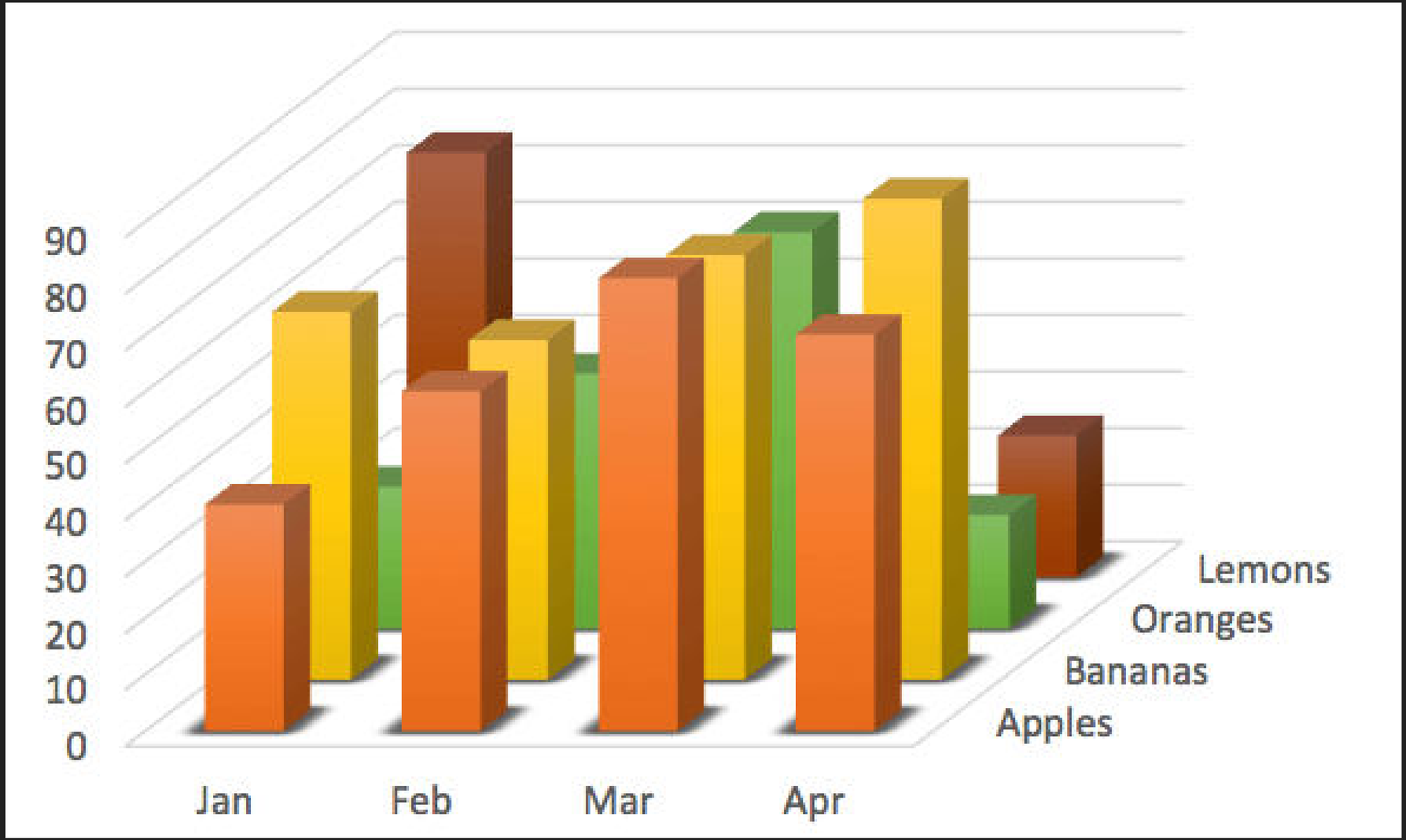
Visualize Your Data

"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**."

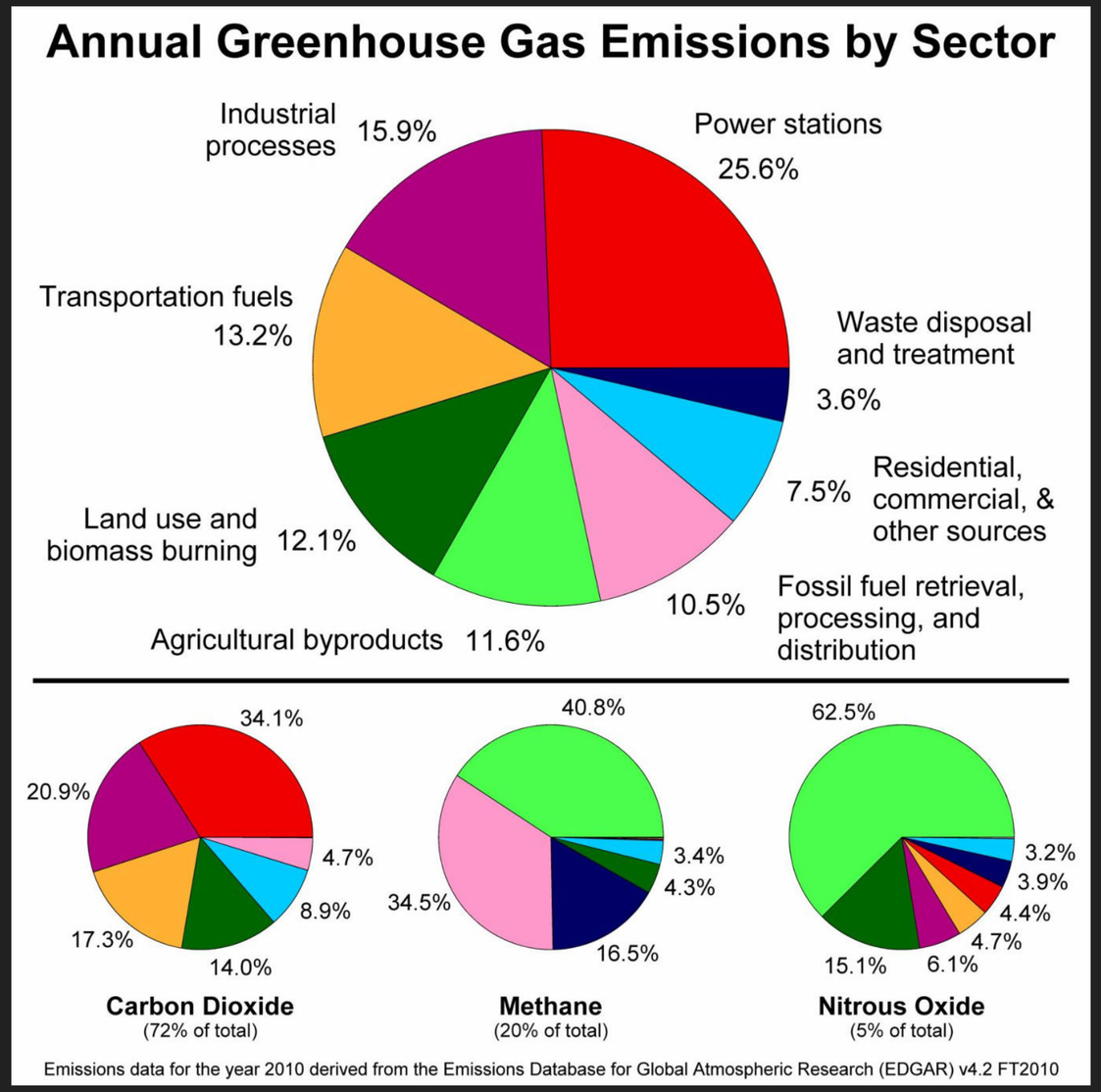


Source: The Economist

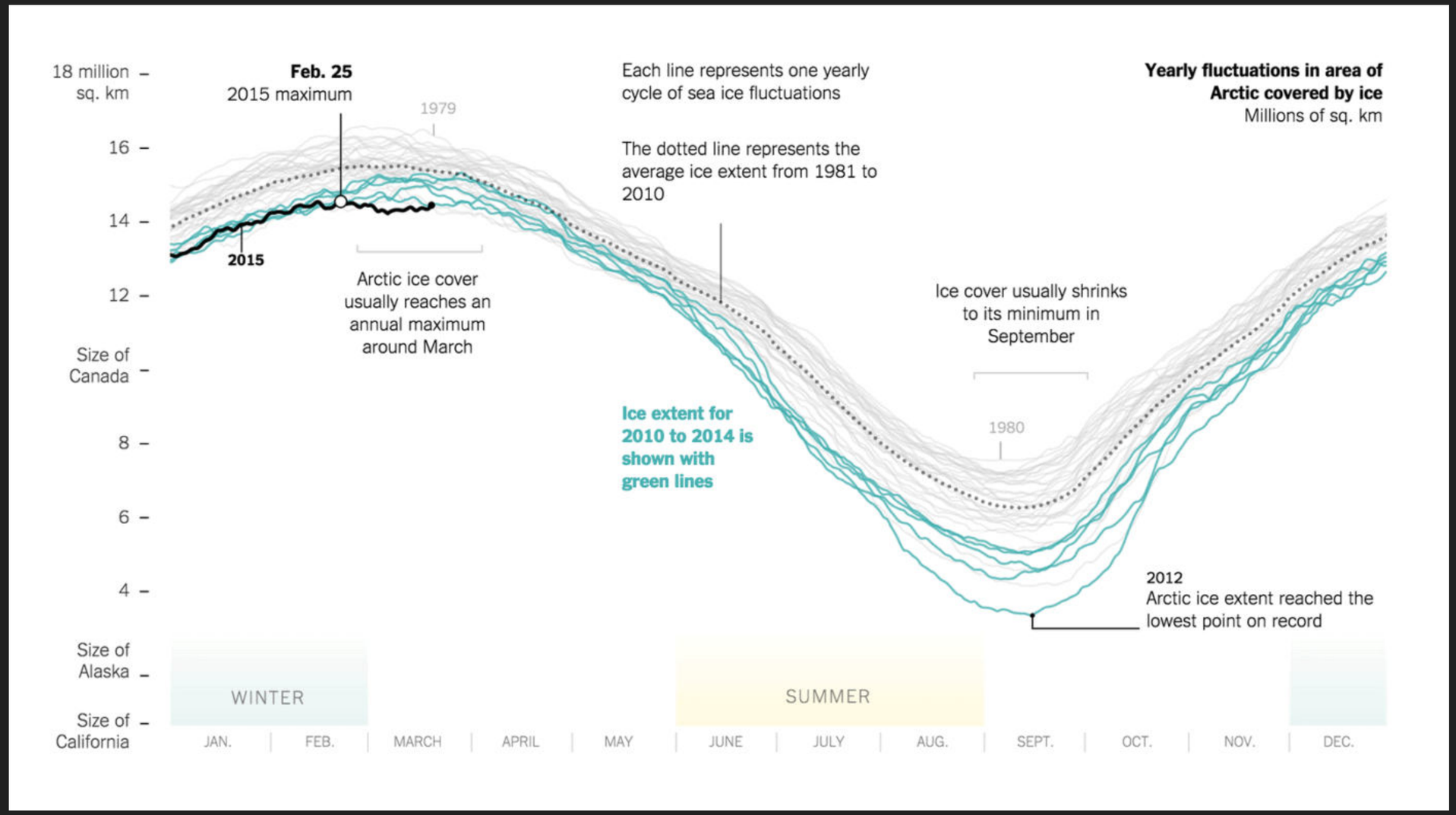
What makes it a bad data visualization?



What makes it a bad data visualization?



What makes it a good data visualization?



"Yearly Fluctuations in Area of Arctic Covered by Ice" by Derek Watkins (New York Times)

Typology of Information Graphics

based on Juuso Koponen & Jonatan Hildén, "Data Visualization Handbook" (2020), page 25

Is the **information** conceptual or measurable?

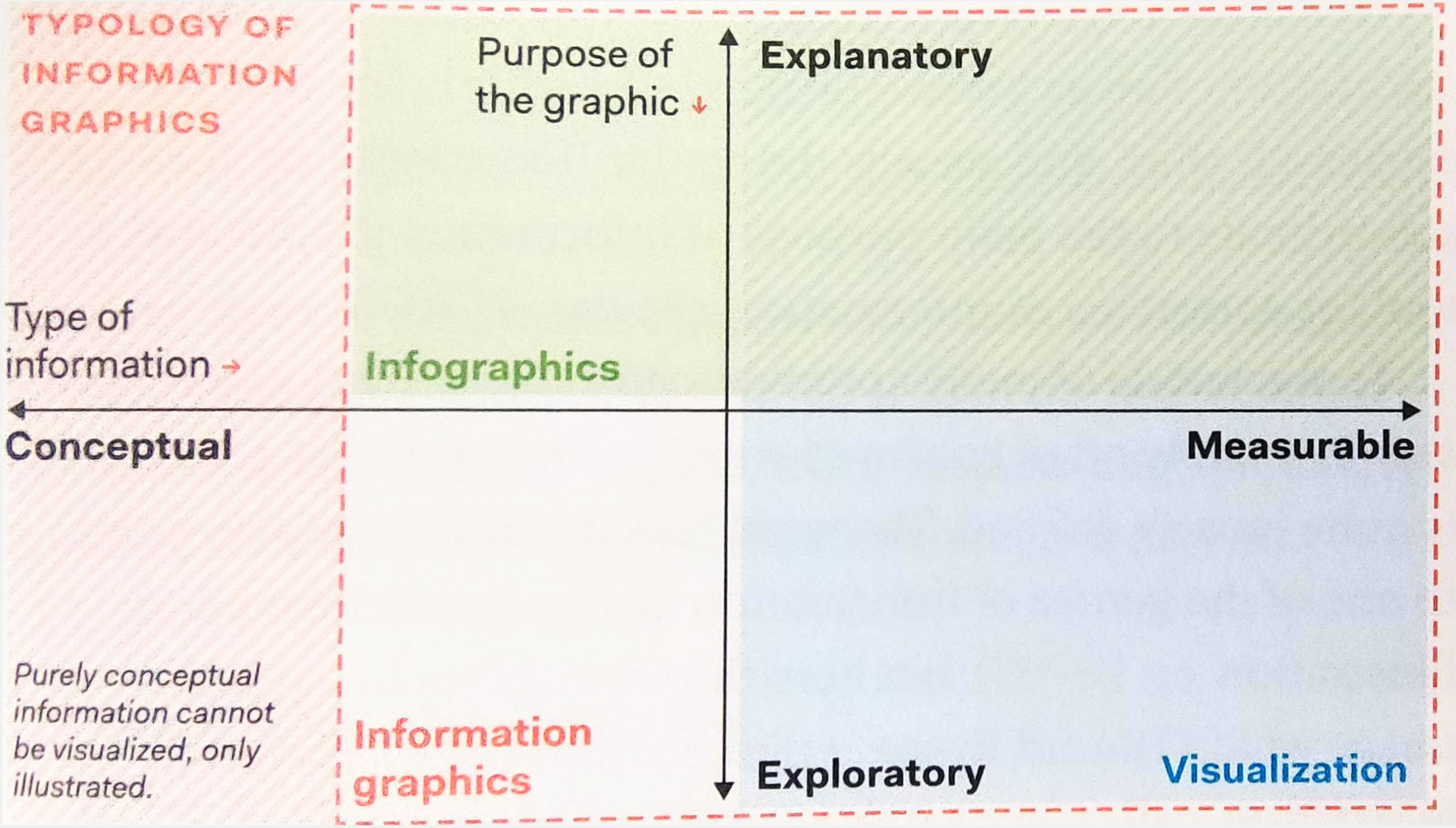
☛ **Type of information**: depict conceptual information <> convert information into visual forms

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

☛ **Purpose of the graphic**: facilite discovery <> communicate information

Typology of Information Graphics

based on Juuso Koponen & Jonatan Hildén, "Data Visualization Handbook" (2020), page 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.”**

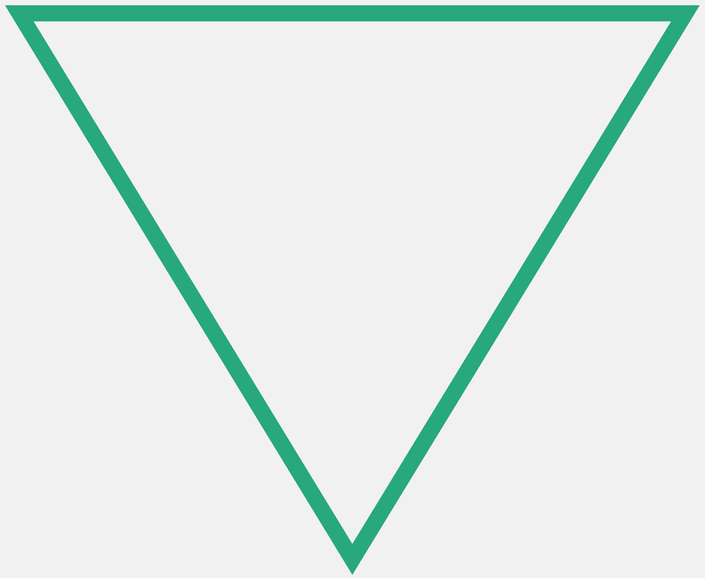
Alberto Cairo

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

What is the purpose?

Exploratory
Discovery

Explanatory
Communication



Affective
Emotion

*The "**Vertices of Visualization**" by Alberto Cairo, personal communication (modified version)*

What is the purpose?

Exploratory
Discovery

Explanatory
Communication

Affective
Emotion

Priority:
efficienct + effective

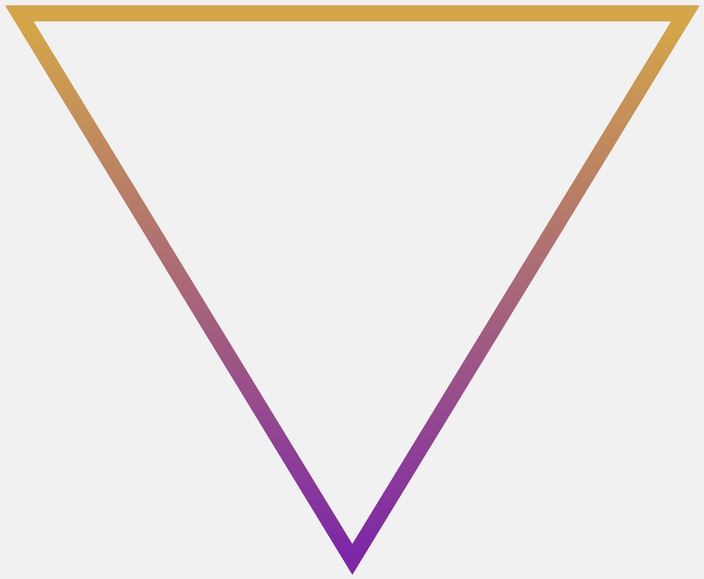
Aim:
functional

The "Vertices of Visualization" by Alberto Cairo, personal communication (modified version)

What is the purpose?

Exploratory
Discovery

Explanatory
Communication



Affective
Emotion

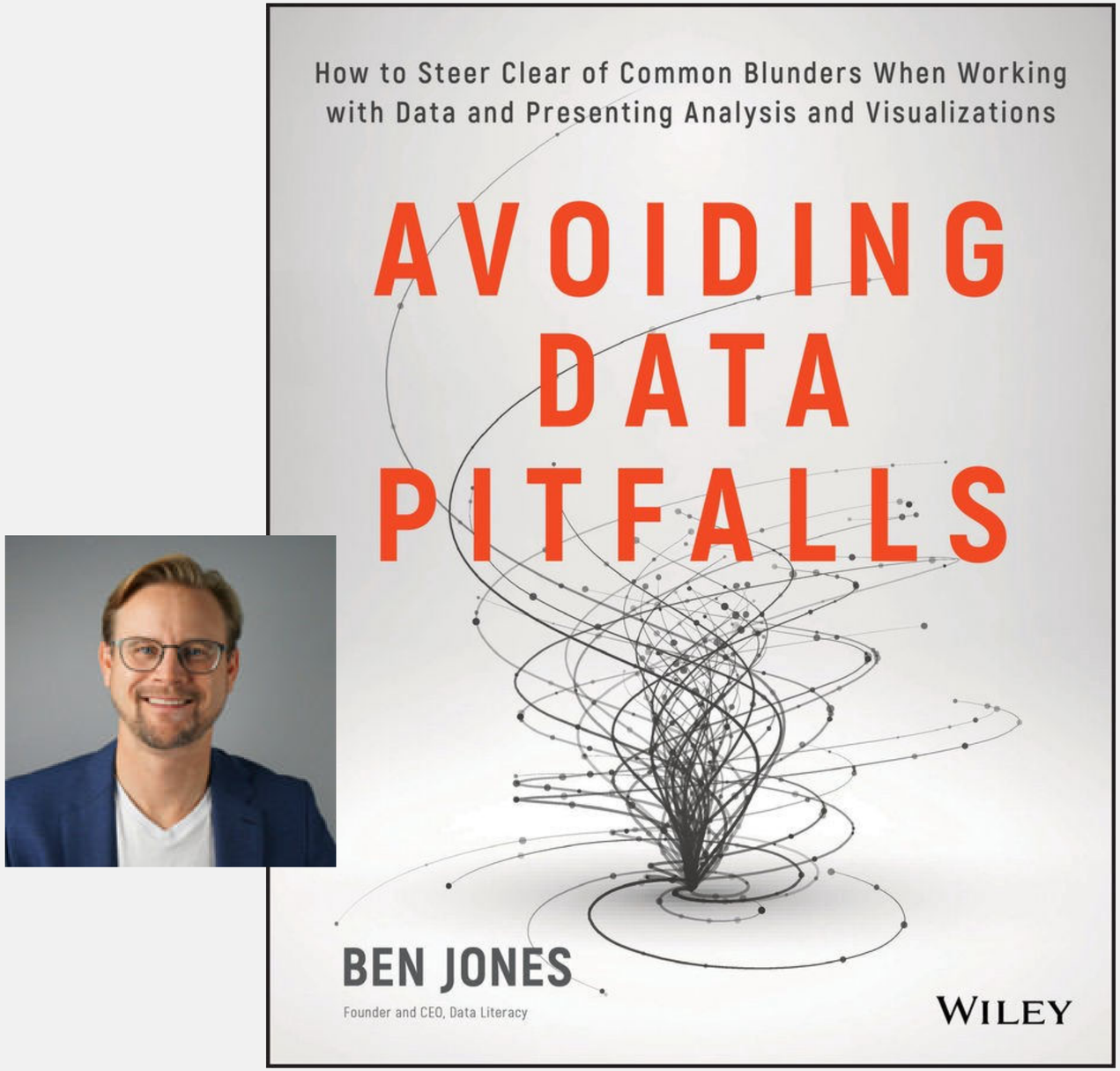
Priority:
efficienct + effective

Response:
functional

Priority:
creative + novel

Response:
emotional

The "Vertices of Visualization" by Alberto Cairo, personal communication (modified version)

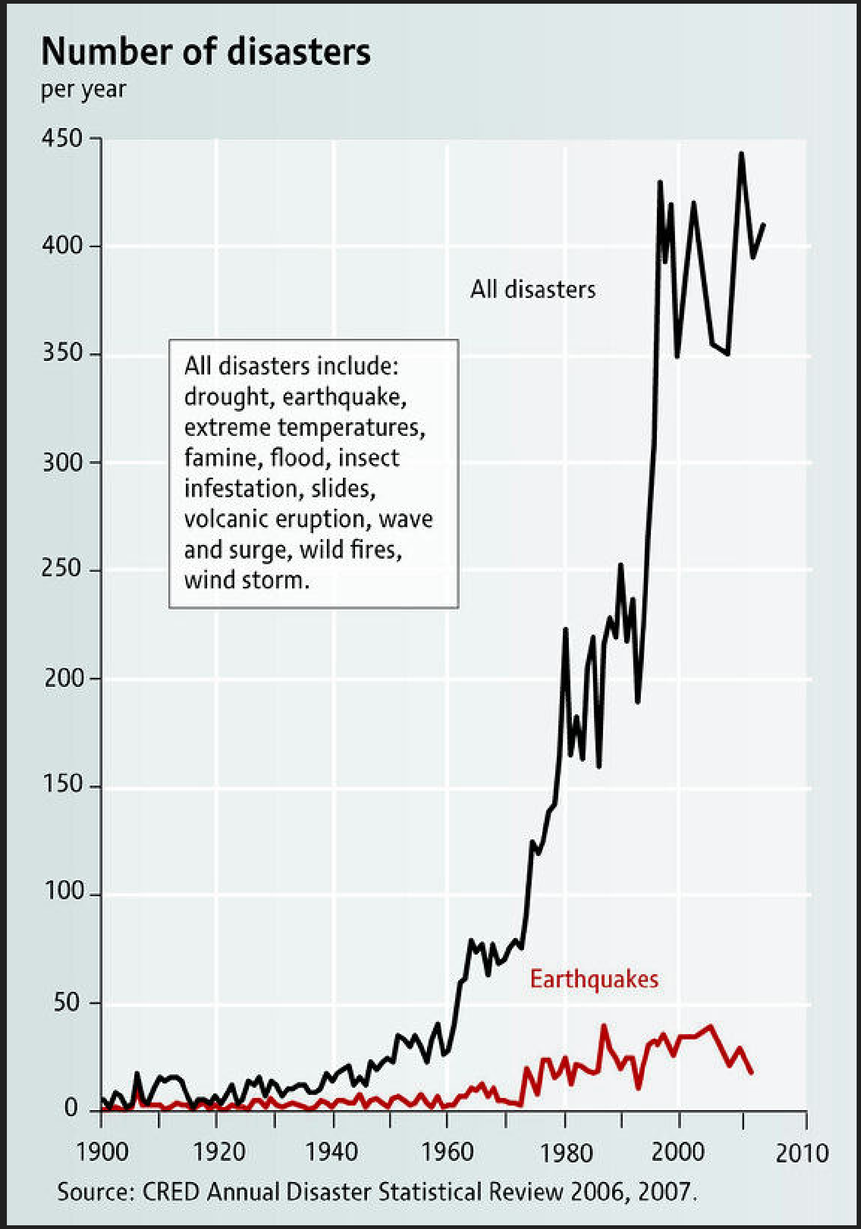

How to Steer Clear of Common Blunders When Working with Data and Presenting Analysis and Visualizations
AVOIDING DATA PITFALLS
BEN JONES
Founder and CEO, Data Literacy
WILEY

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

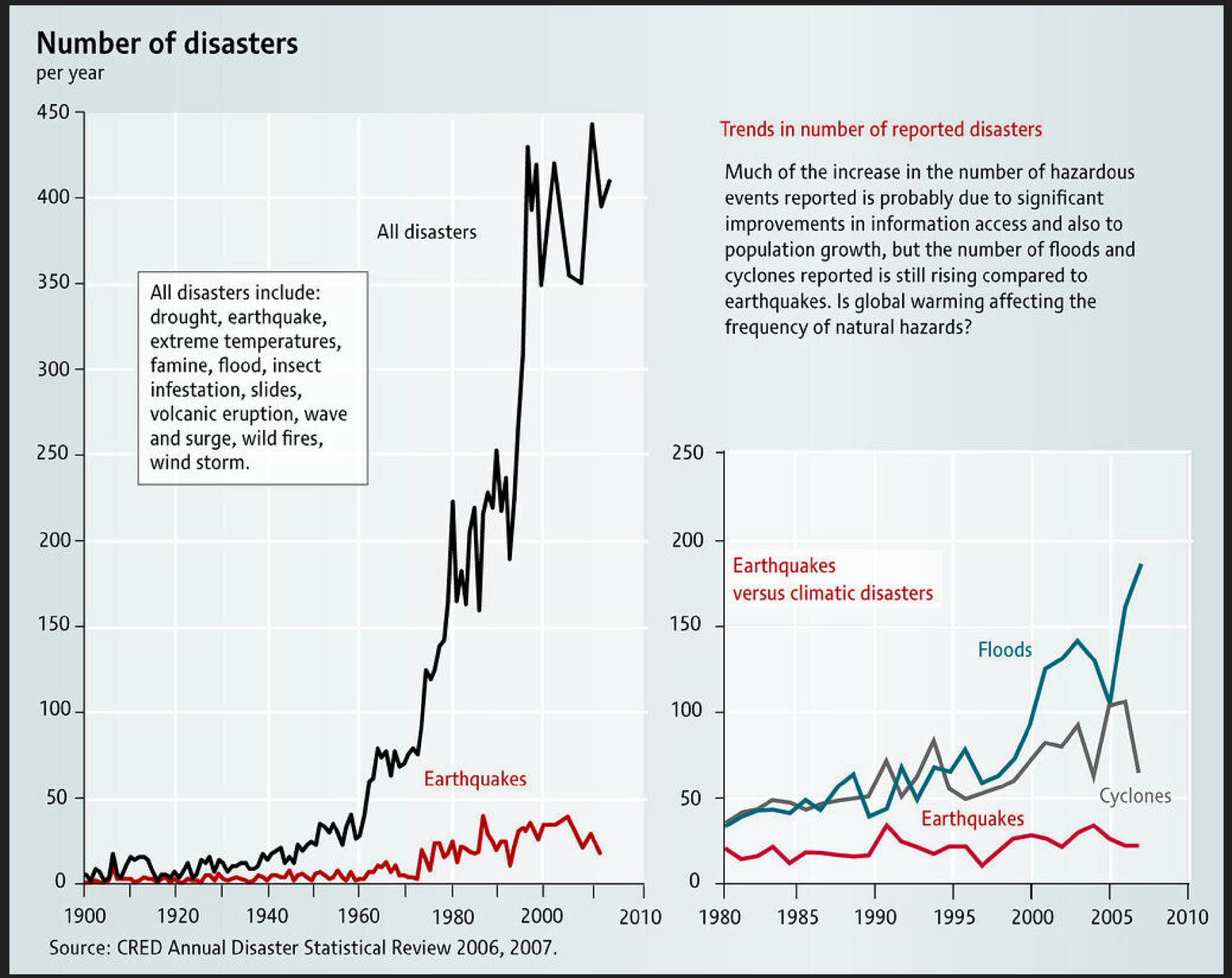
→ **only a subset:** not crime but reported crime

→ **collected by humans:** guesstimation, precision and errors

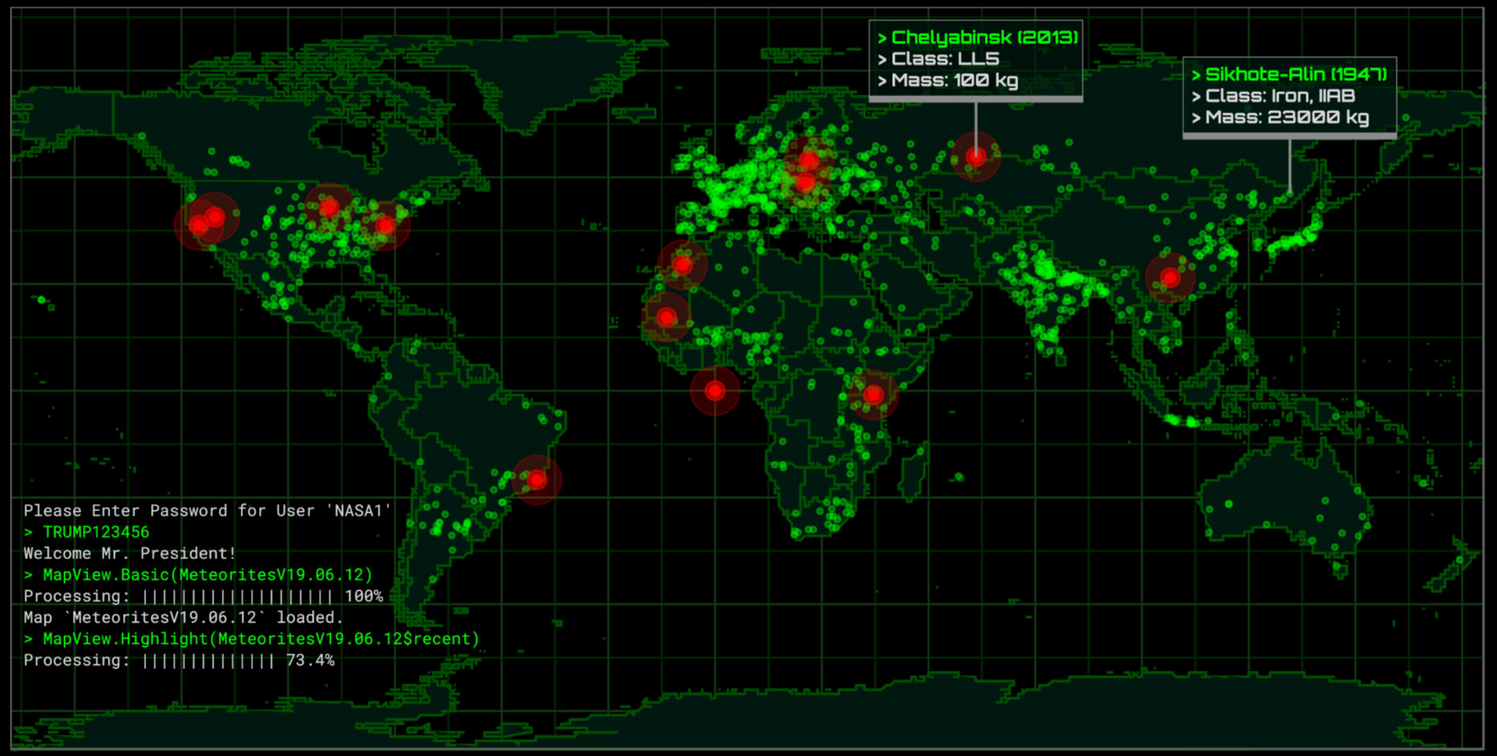
→ **collected by machines:** precisions and errors


Number of disasters
per year
450
400
350
300
250
200
150
100
50
0
1900
1920
1940
1960
1980
2000
2010
All disasters
Earthquakes
All disasters include: drought, earthquake, extreme temperatures, famine, flood, insect infestation, slides, volcanic eruption, wave and surge, wild fires, wind storm.
Source: CRED Annual Disaster Statistical Review 2006, 2007.

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



SPACE OBSERVER 3000 X V3.5.1



Provided by Cédric Scherer & National Aeronautics and Space Administration (NASA)

WARNING!
The number of fallen meteorites
is drastically decreasing!
100
90
80
70
60
50
40
30
20
10
0
1500 1550 1600 1650 1700 1750 1800 1850 1900 1950 2000 2050

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

→ don’t formulate a single statement

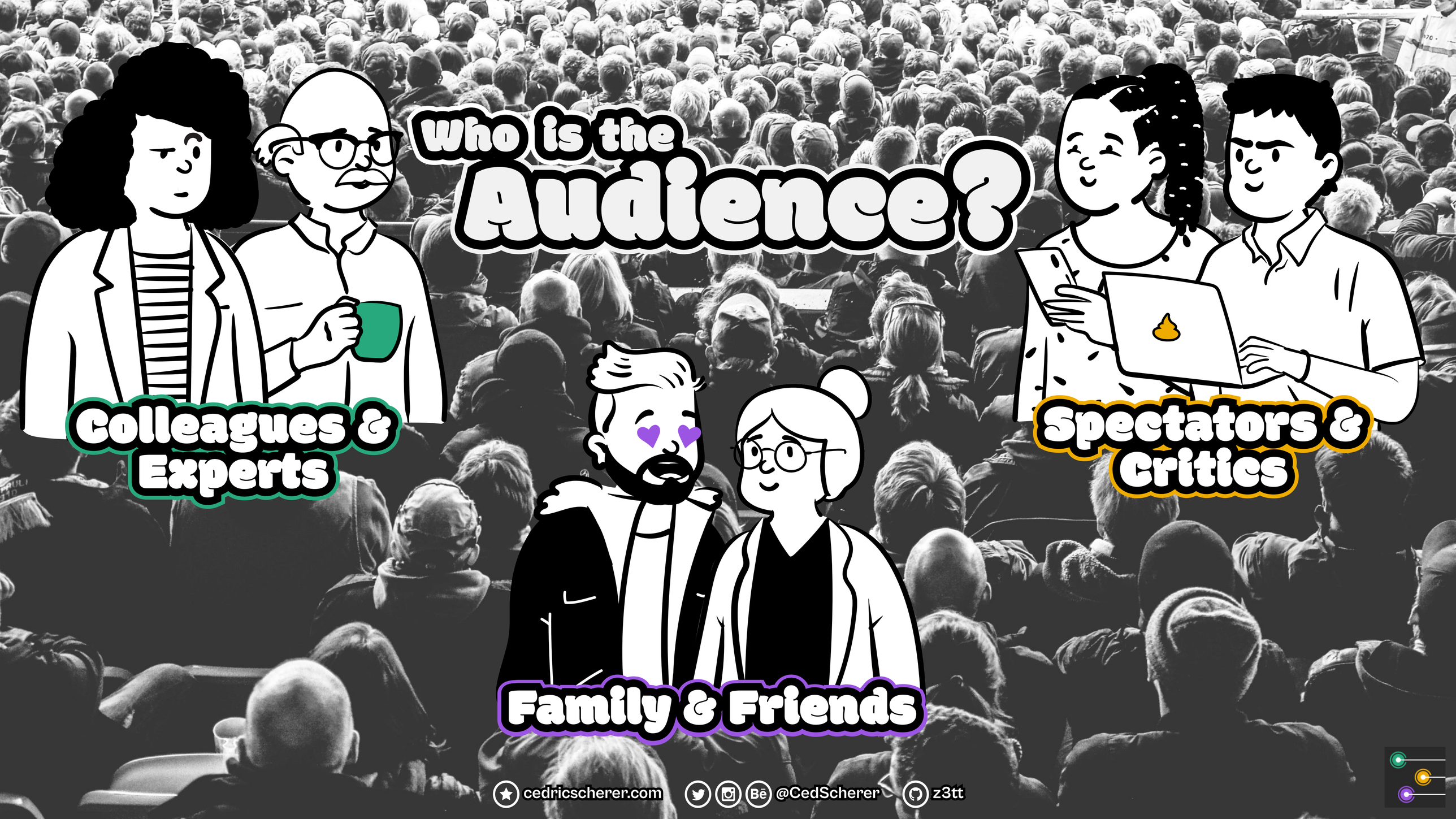
→ confront yourself with a **falsifiable universal statement**

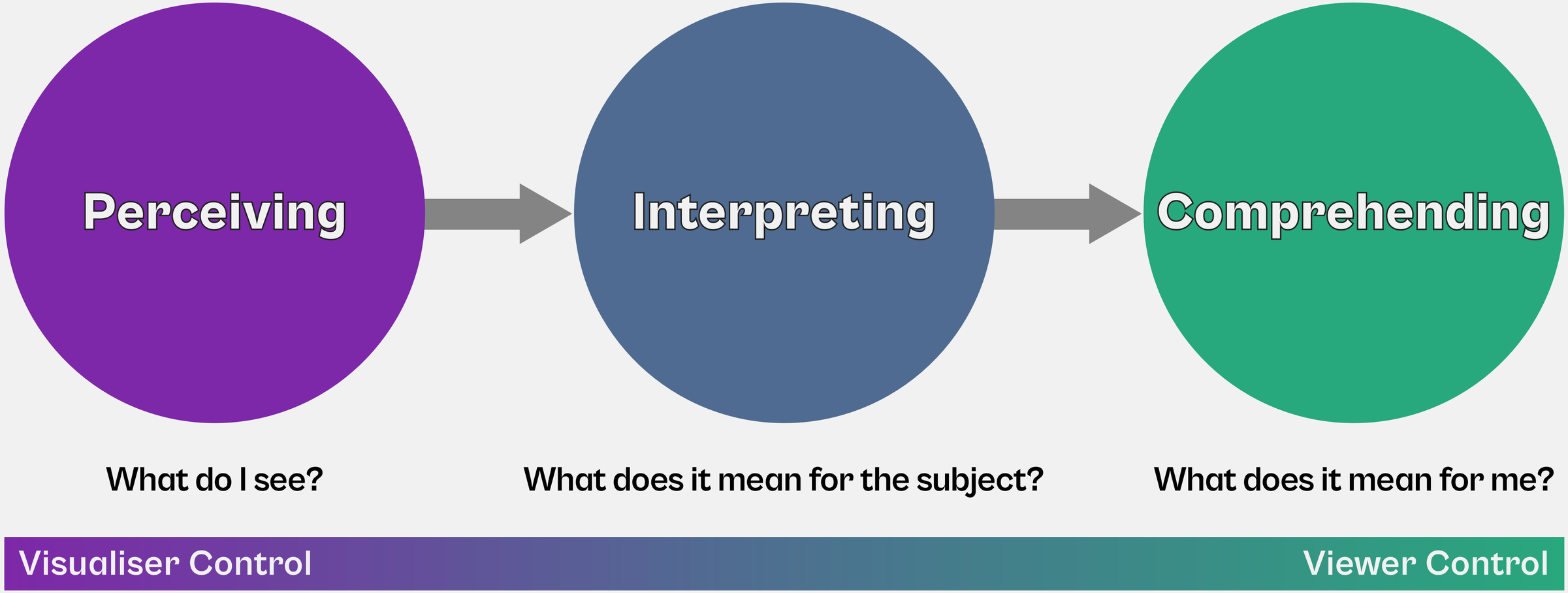
The best use of data is to teach us what **isn't true.**

→ single statement: "The swan is white."

→ **falsifiable universal statement: "All swans are white."**

the
Audience
cedricscherer.com
@CedScherer
z3tt


Who is the Audience?
Colleagues & Experts
Spectators & Critics
Family & Friends
cedricscherer.com
@CedScherer
z3tt



Scheme by Andy Kirk

Who is the Audience?

🤩 Which story is **interesting** for them?

Who is the Audience?

🤩 Which story is **interesting** for them?

🤔 Which variables are **meaningful** to them?

Who is the Audience?

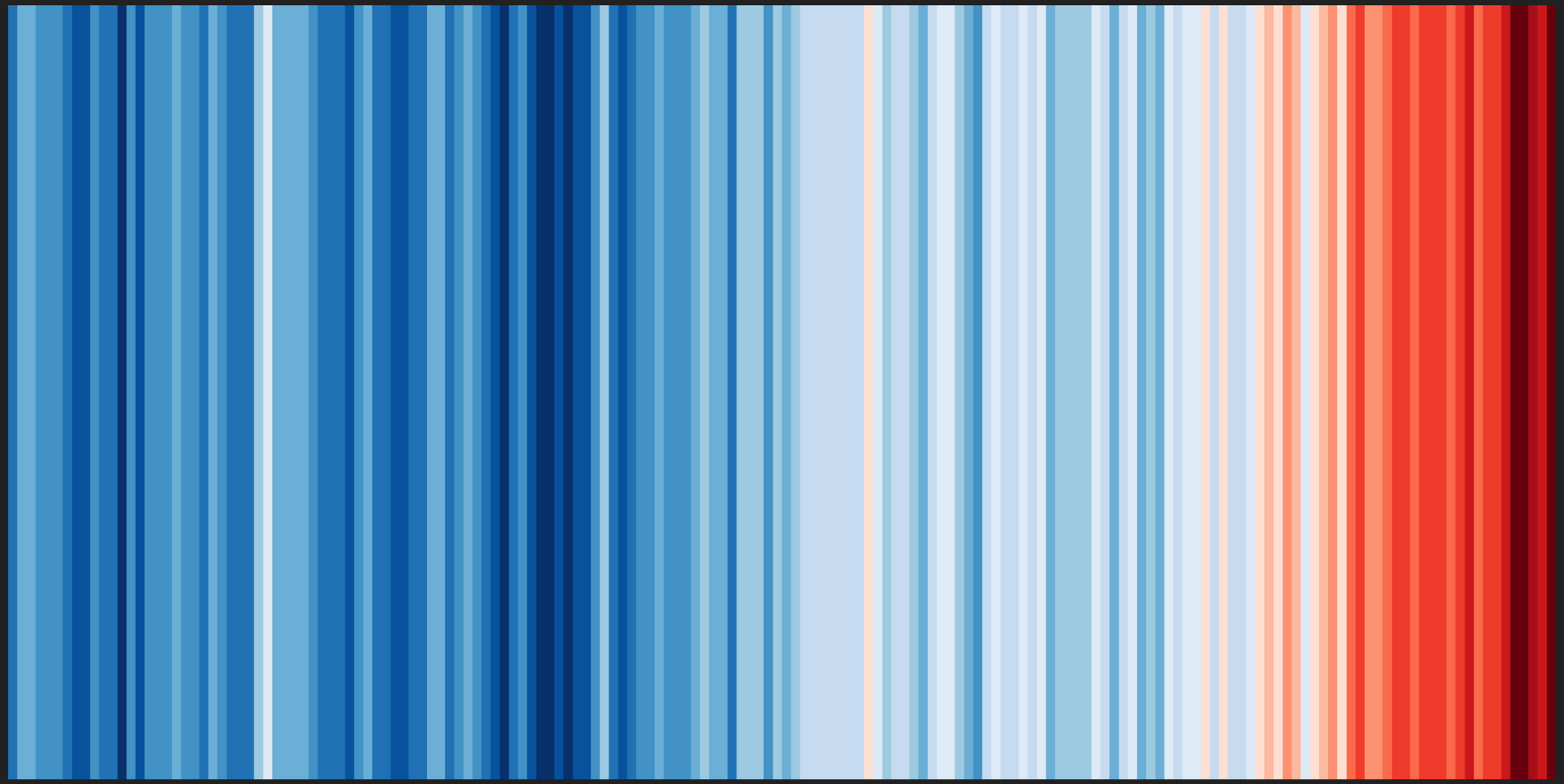
- 🤩 Which story is **interesting** for them?
- 🤔 Which variables are **meaningful** to them?
- 🤓 What are **relevant** details to include?

Who is the Audience?

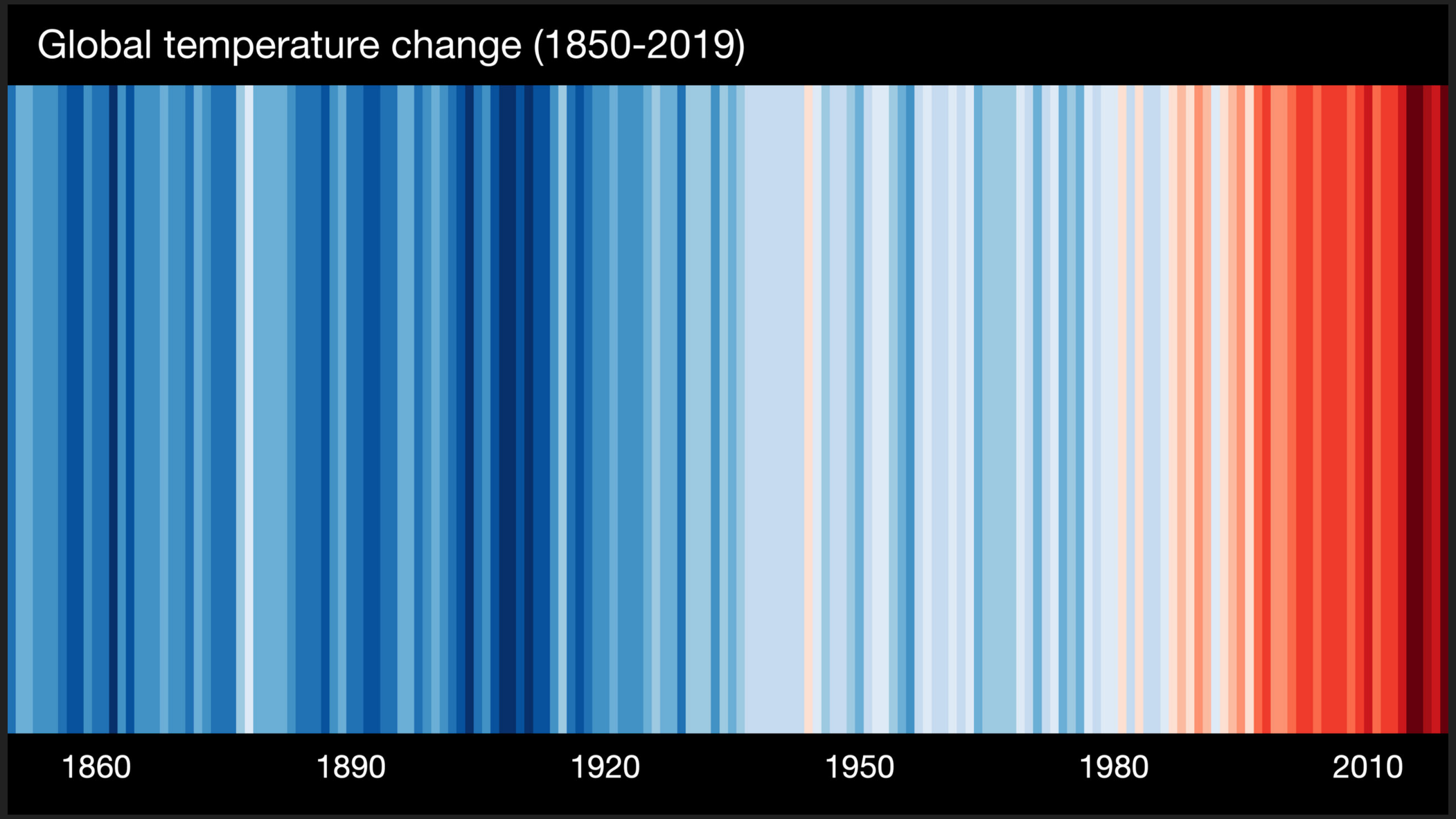
- Which story is **interesting** for them?
- Which variables are **meaningful** to them?
- What are **relevant** details to include?
- How will they **encounter** the visualization?

Who is the Audience?

- Which story is **interesting** for them?
- Which variables are **meaningful** to them?
- What are **relevant** details to include?
- How will they **encounter** the visualization?
- **Do I need a visualization at all?**



"Warming Stripes" by Ed Hawkins



"Warming Stripes" by Ed Hawkins

FAQ : Frequently asked questions

What are these graphics?

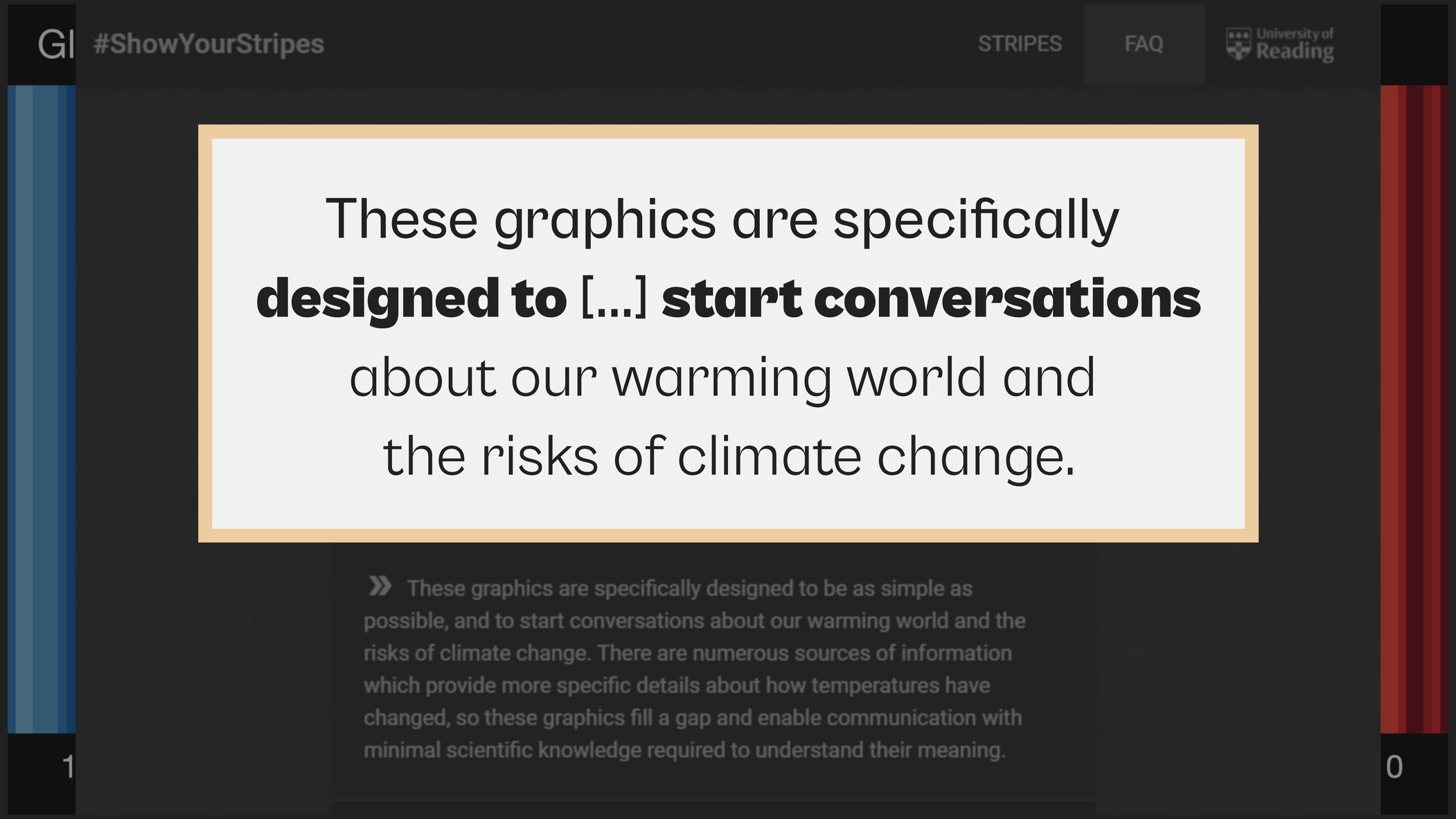
What do the graphics show?

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.

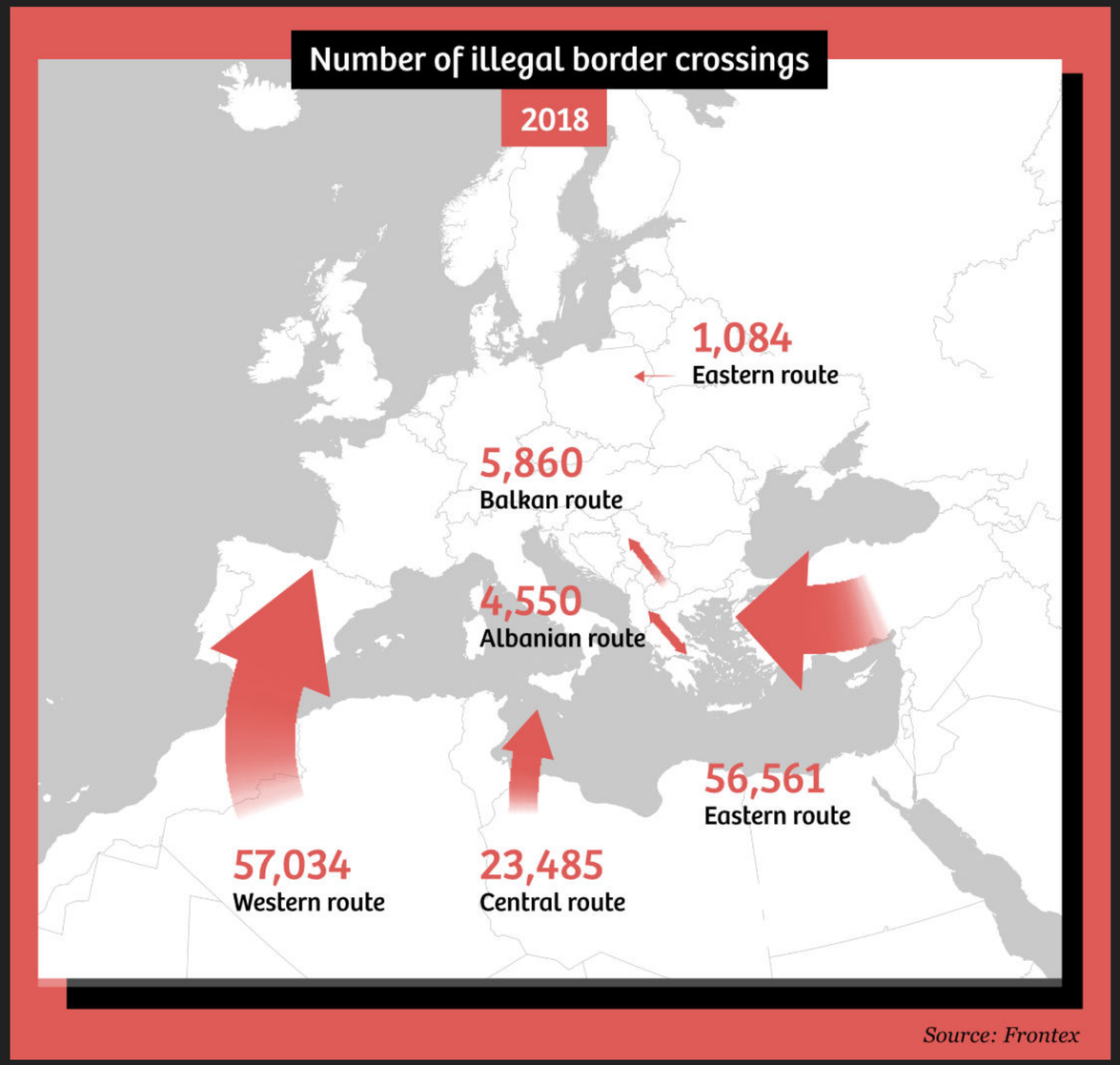
showyourstripes.info/faq

cedricscherer.com @CedScherer z3tt

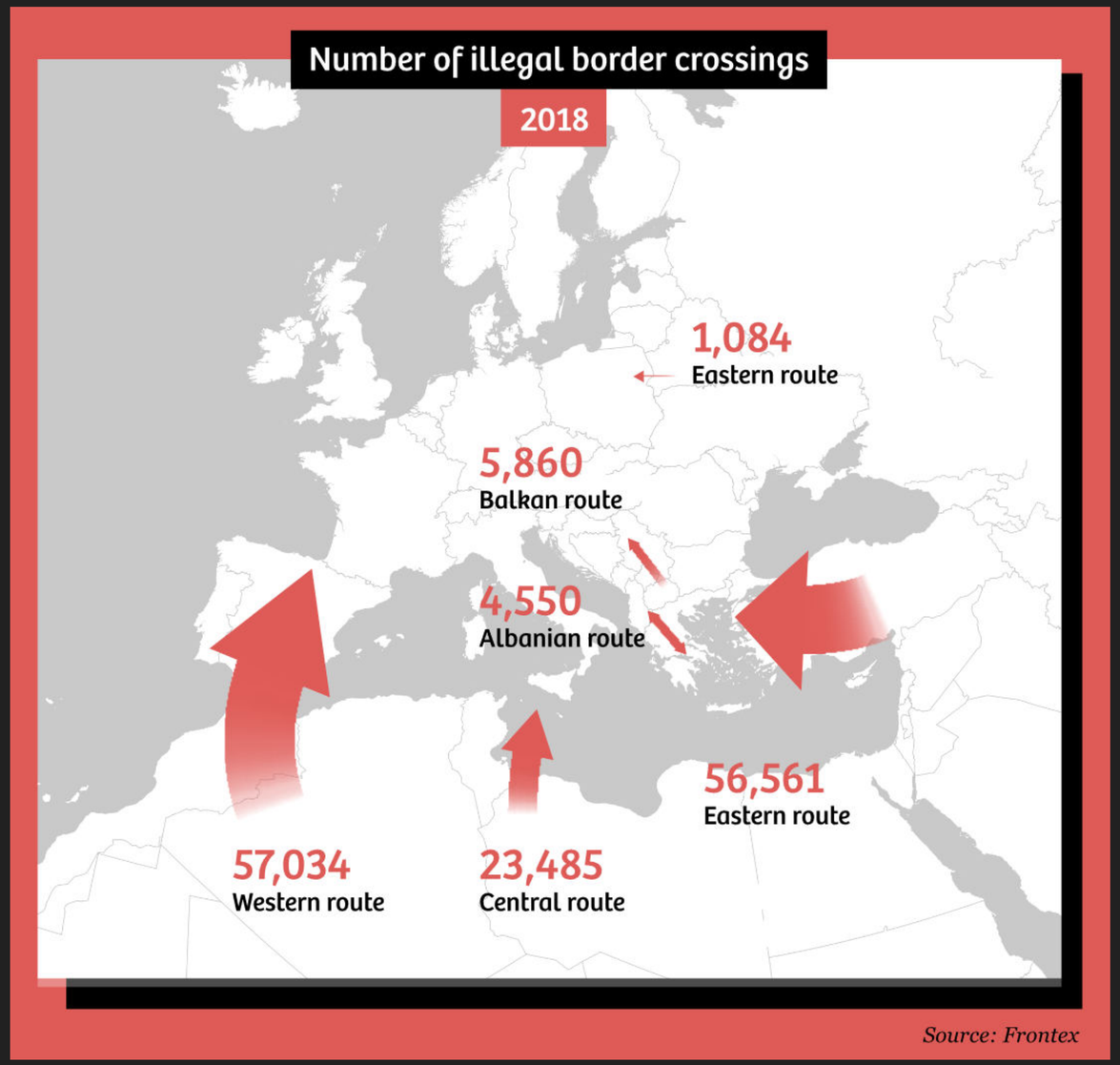


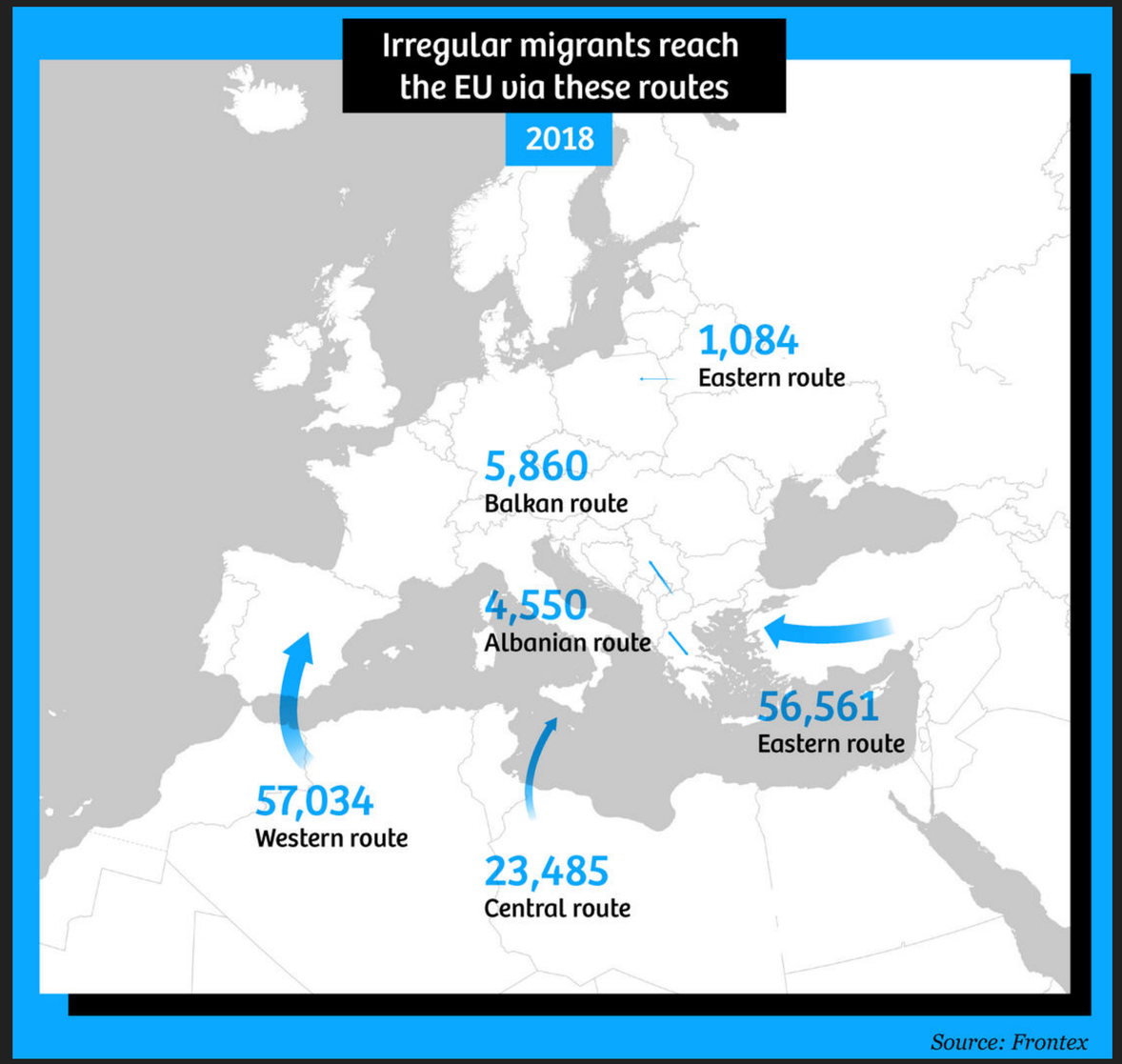
These graphics are specifically **designed to** [...] **start conversations** about our warming world and the risks of climate change.

showyourstripes.info/faq

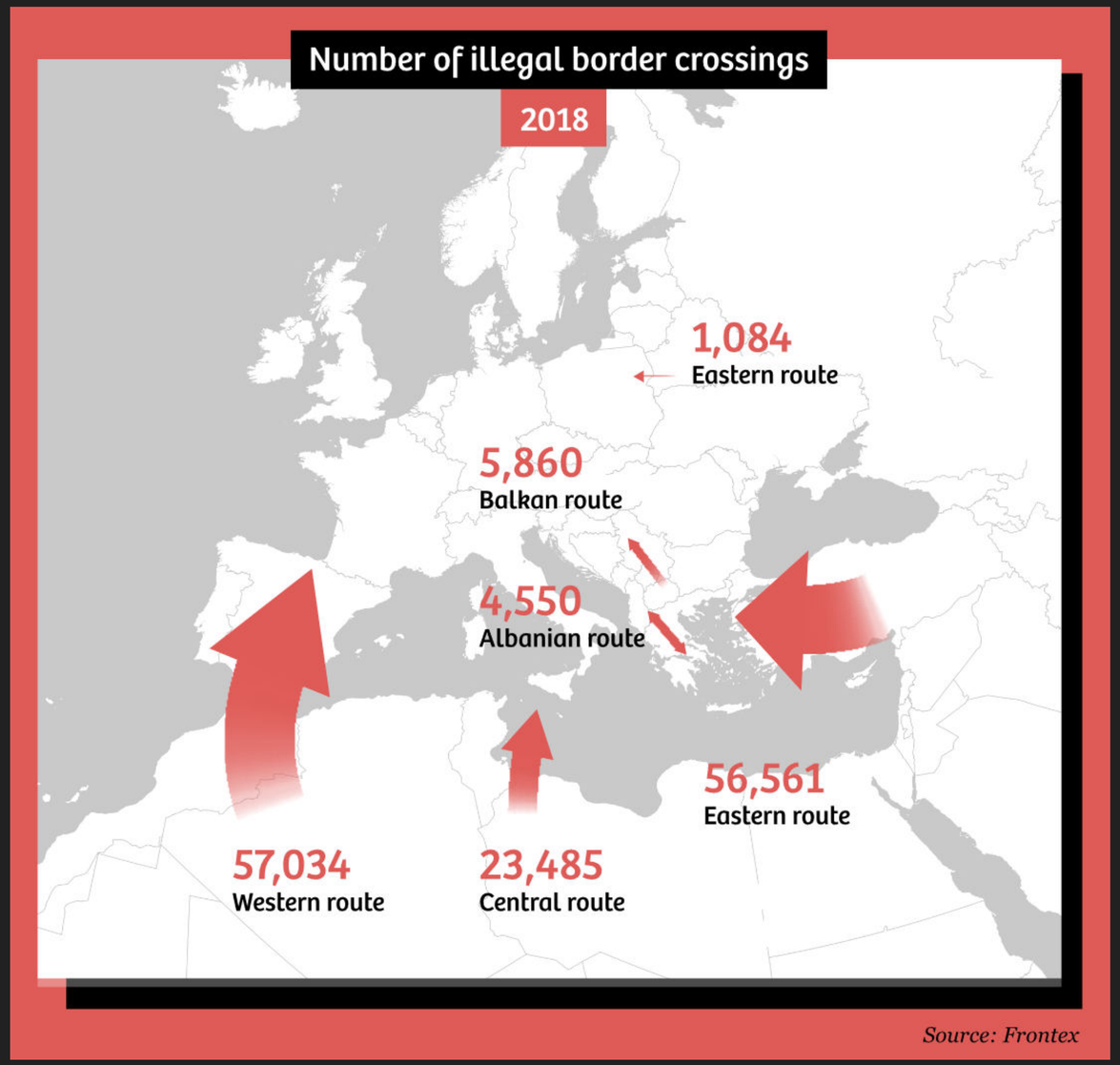


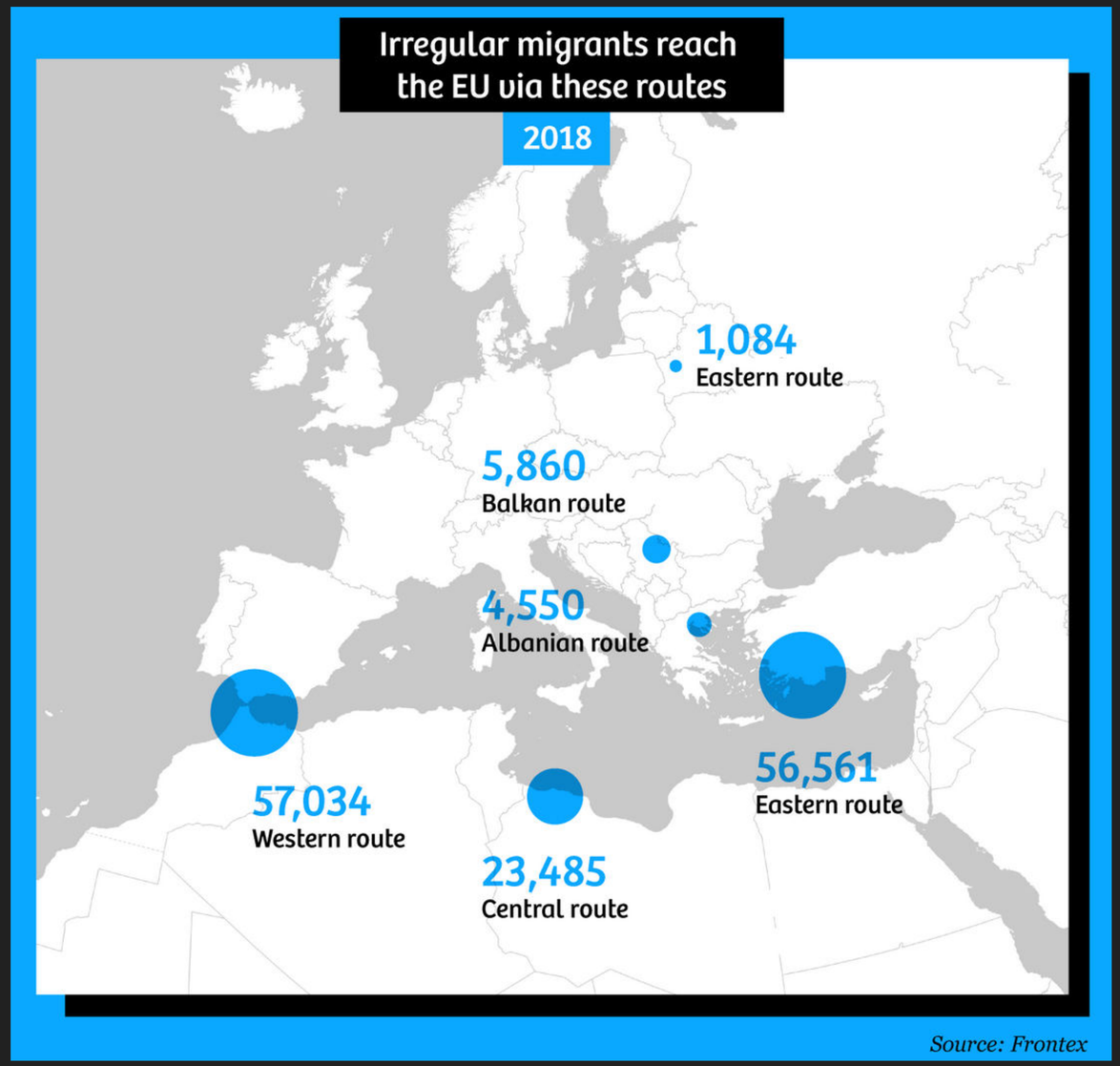
"How maps in the media make us more negative about migrants" by Maite Vermeulen, Leon de Korte & Henk van Houtum



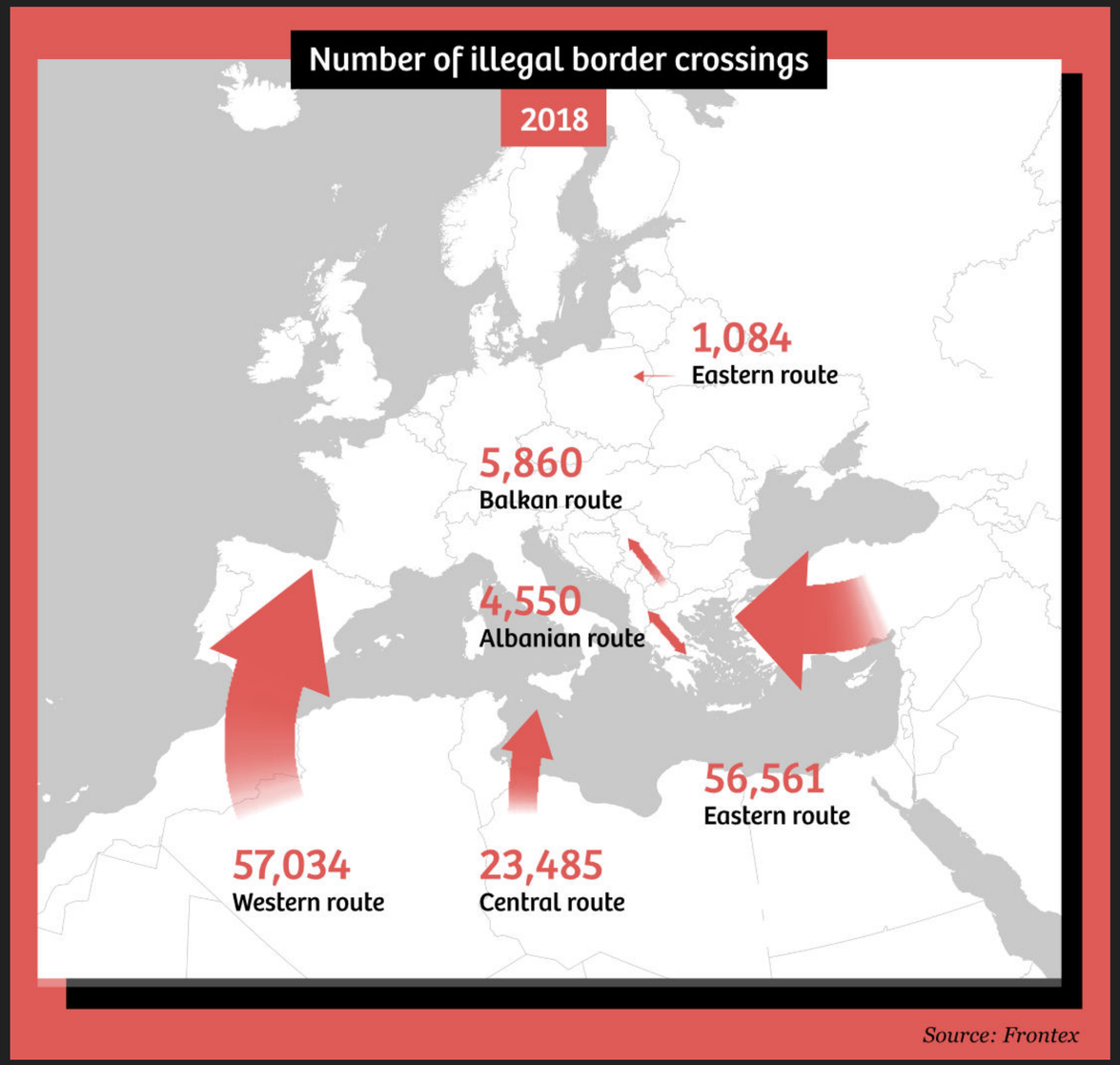


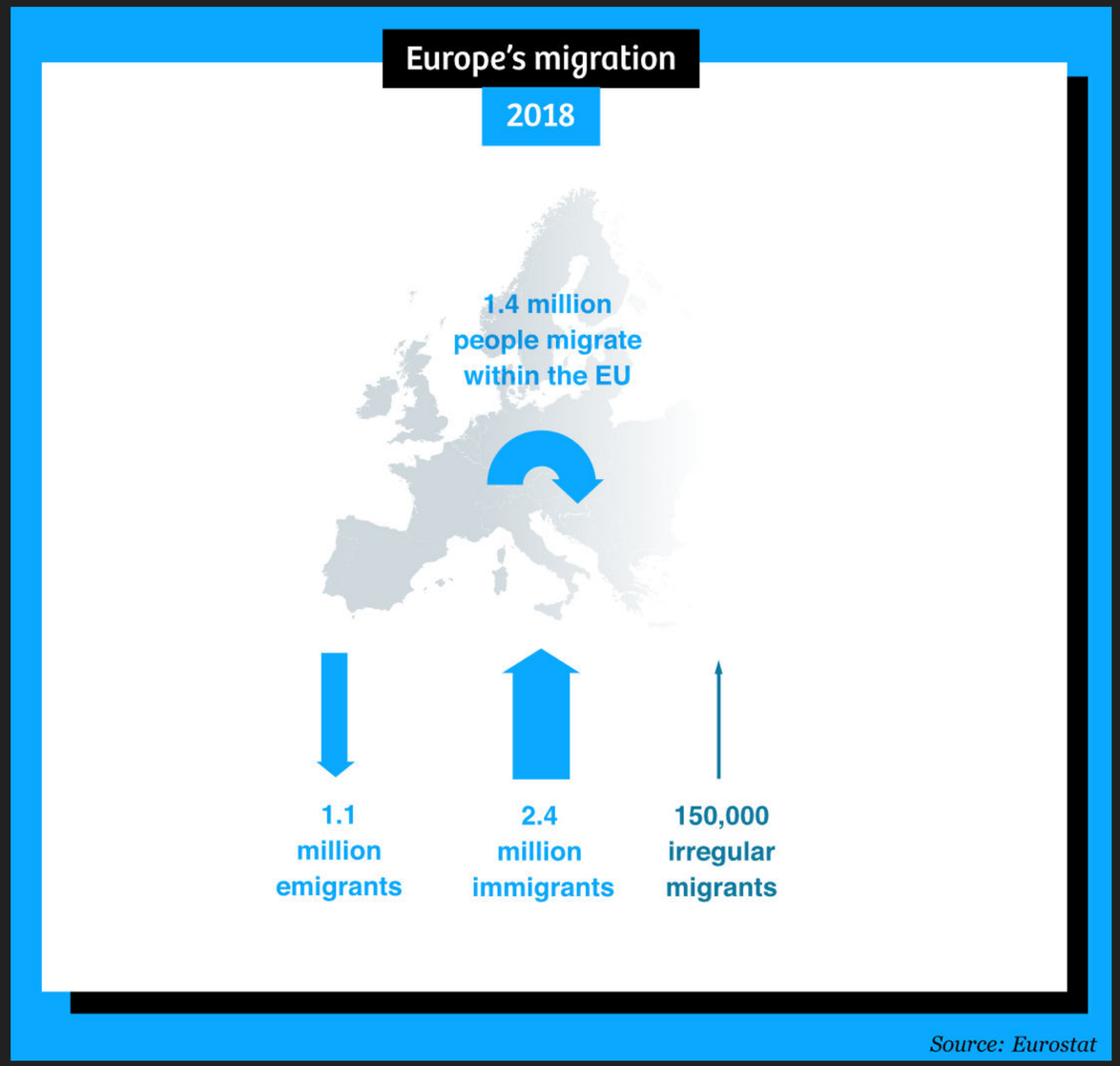
"How maps in the media make us more negative about migrants" by Maite Vermeulen, Leon de Korte & Henk van Houtum



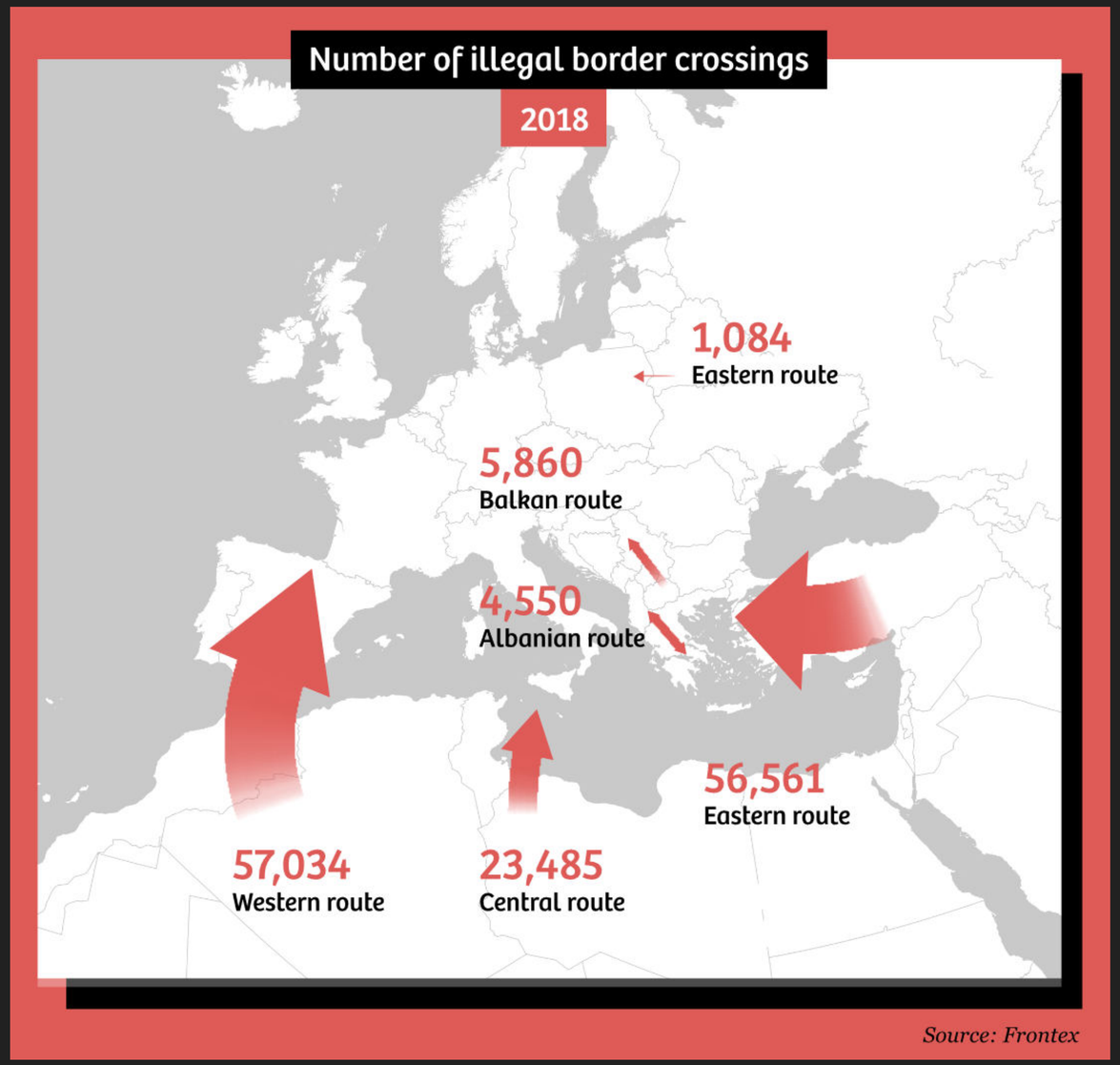


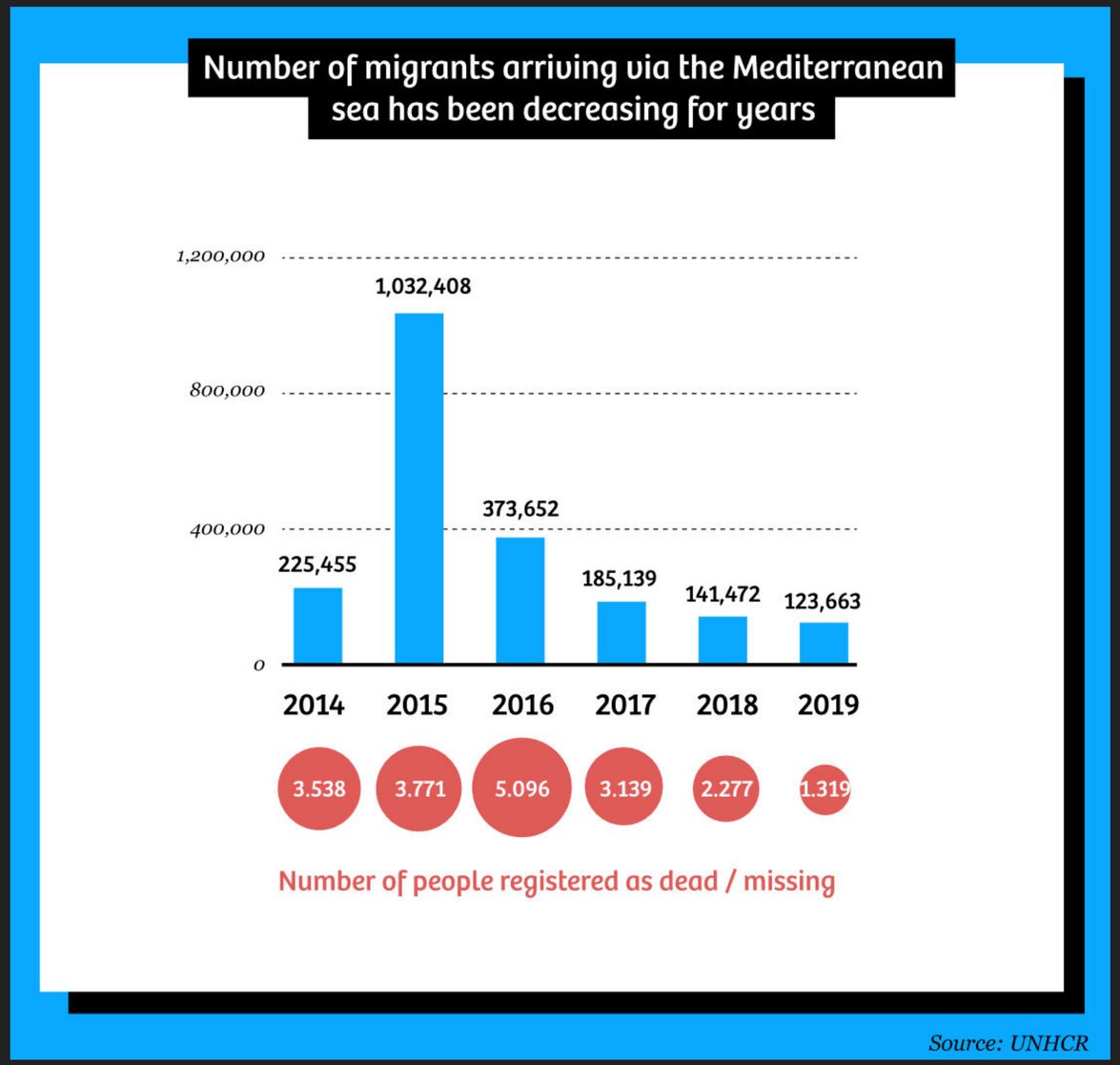
"How maps in the media make us more negative about migrants" by Maite Vermeulen, Leon de Korte & Henk van Houtum



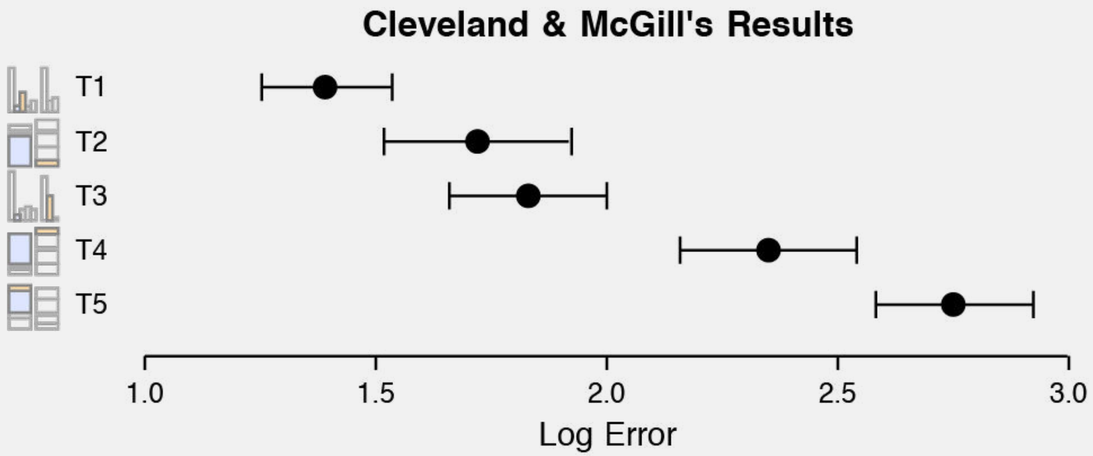


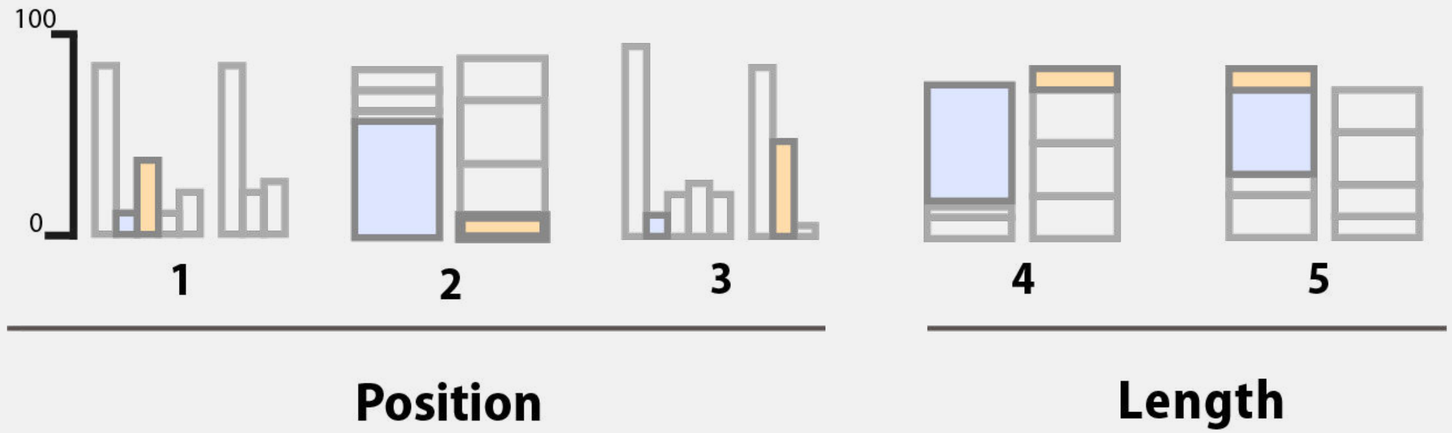
"How maps in the media make us more negative about migrants" by Maite Vermeulen, Leon de Korte & Henk van Houtum



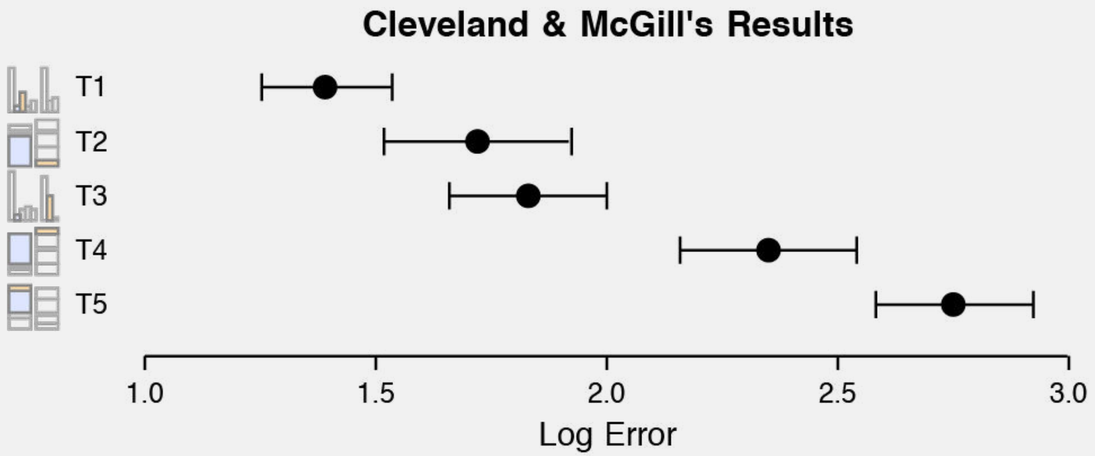


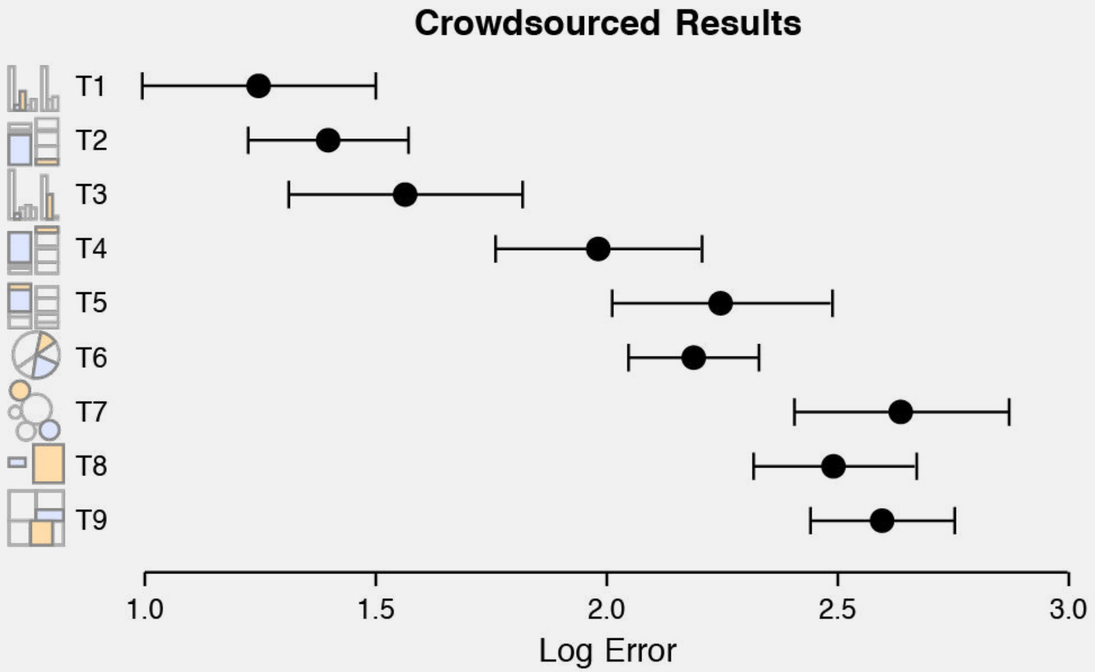
"How maps in the media make us more negative about migrants" by Maite Vermeulen, Leon de Korte & Henk van Houtum





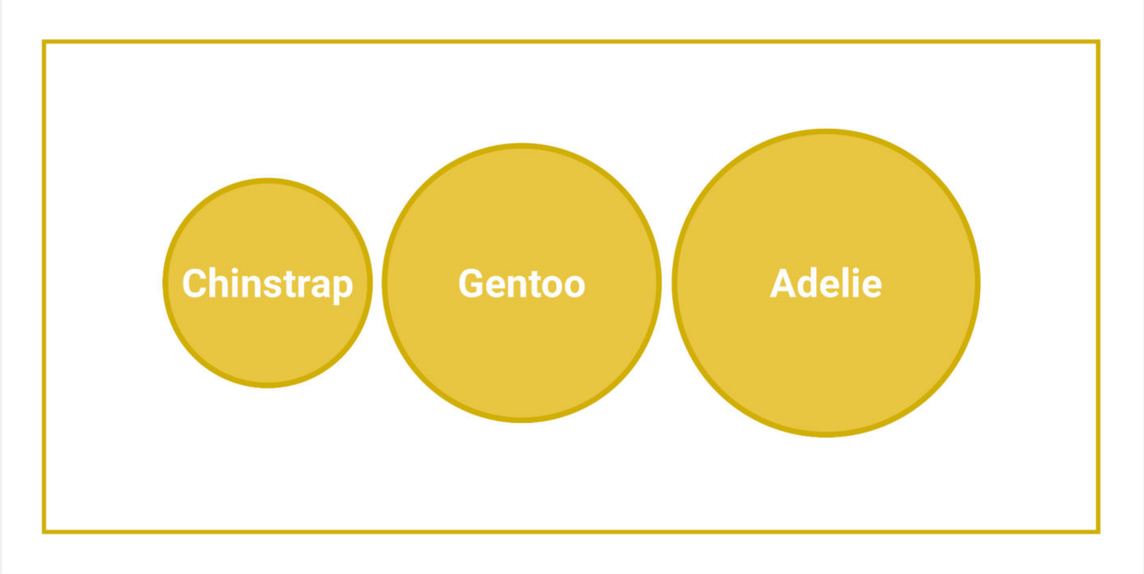
Source: Kieran Healy's "Data Visualization: A Practical Introduction"; results based on Heer and Bostock, following Cleveland and McGill



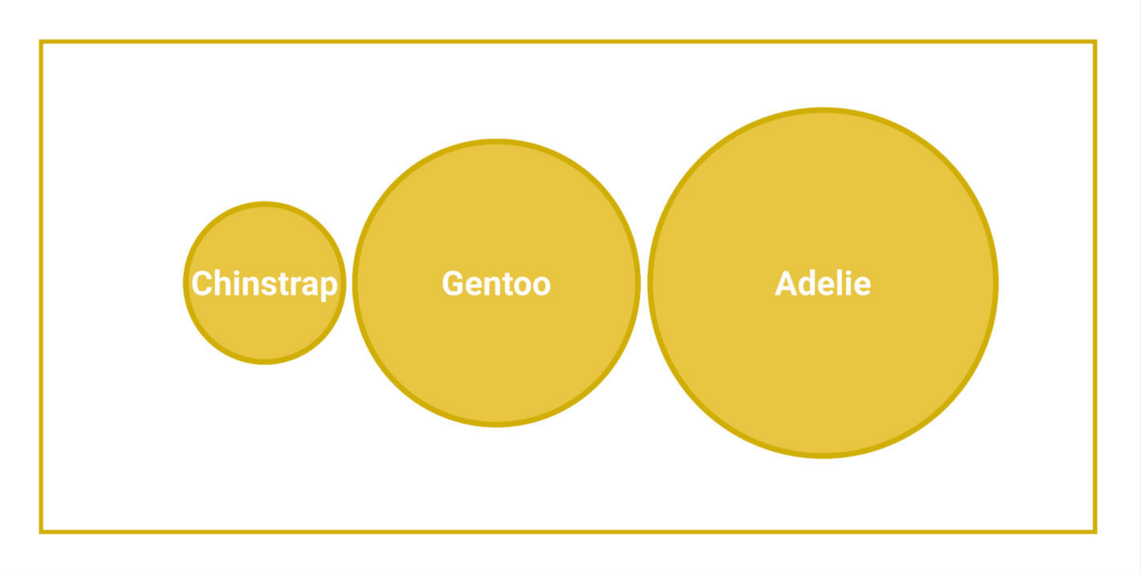


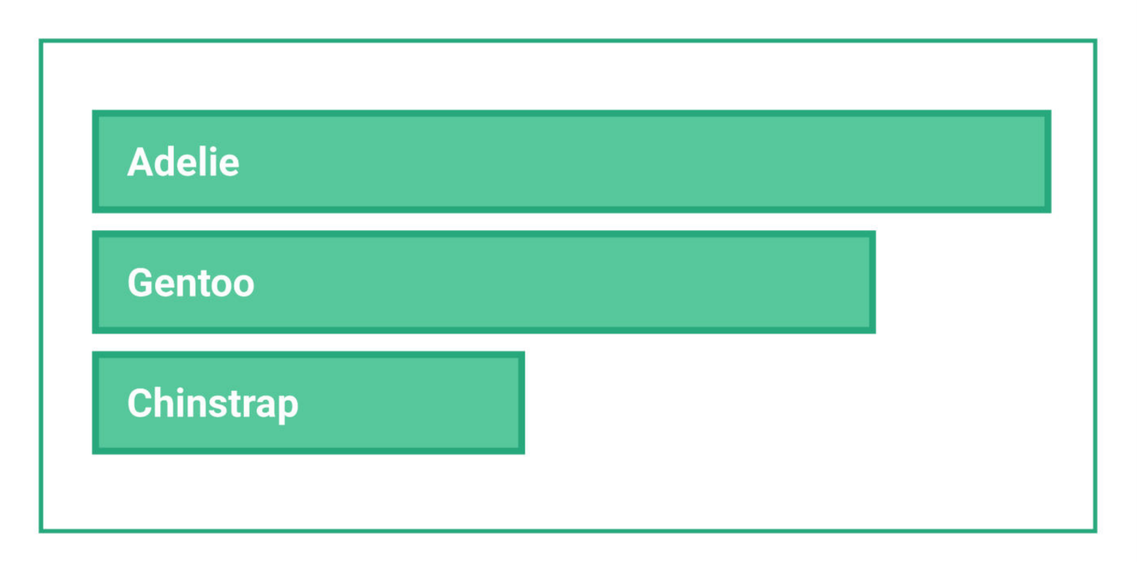
Source: Kieran Healy's "Data Visualization: A Practical Introduction"; results based on Heer and Bostock, following Cleveland and McGill

Use area.



Not radius!

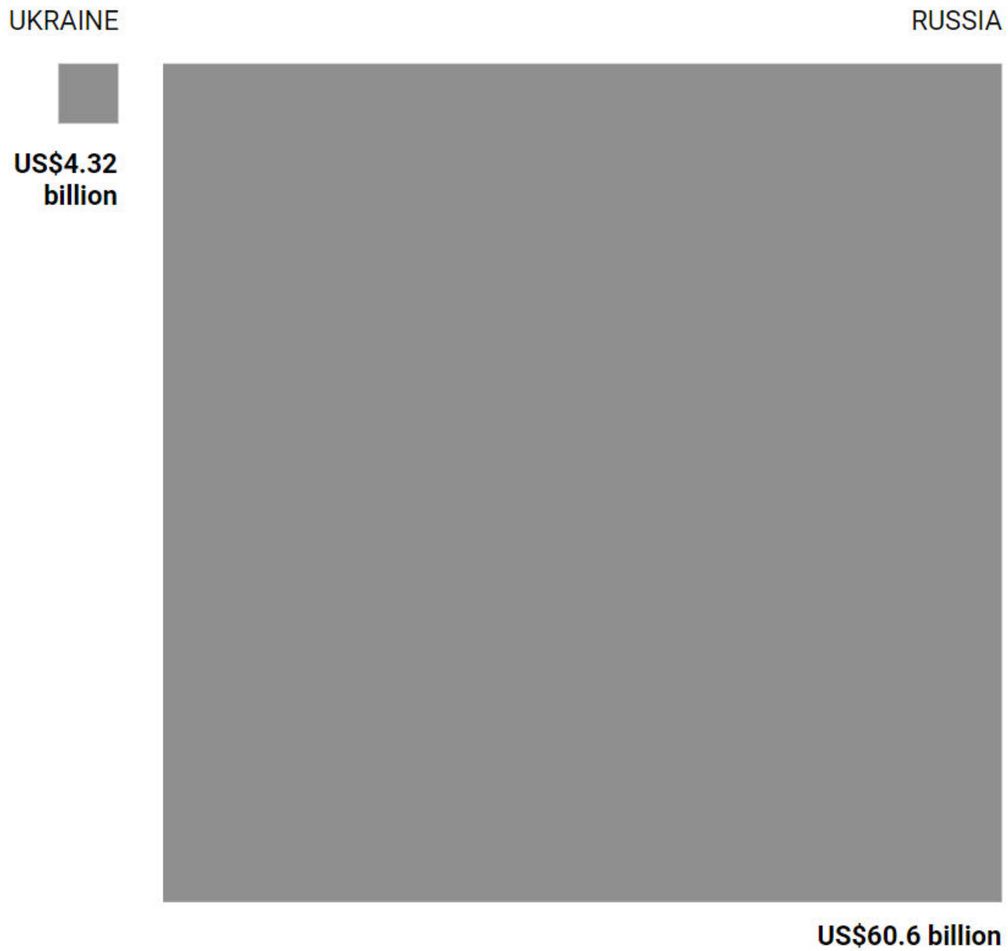




or: bars!

DEFENCE BUDGETS: RUSSIA VS UKRAINE (2020)

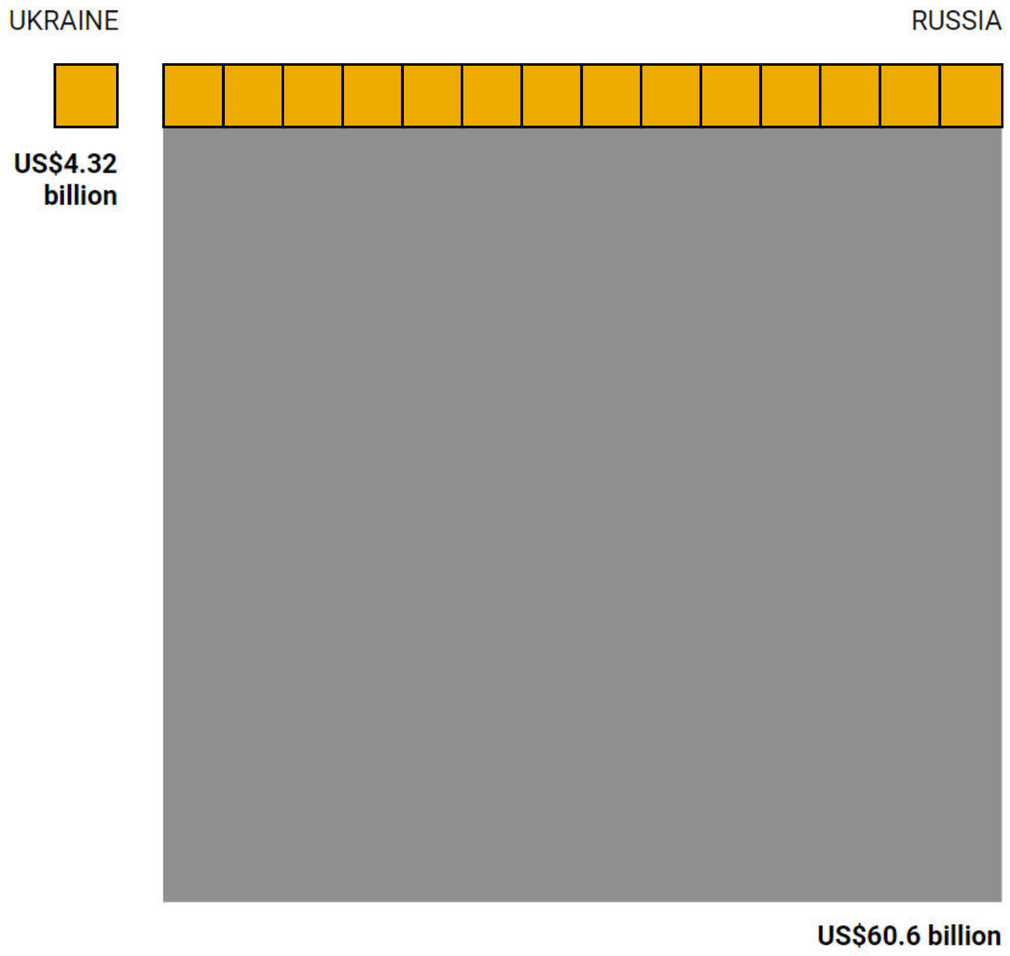
The national balance of forces is overwhelmingly in Russia's favour. Russian military spending in 2020 amounted to US$60.6 billion in 2020. Ukraine's was less than a 10th of that amount.



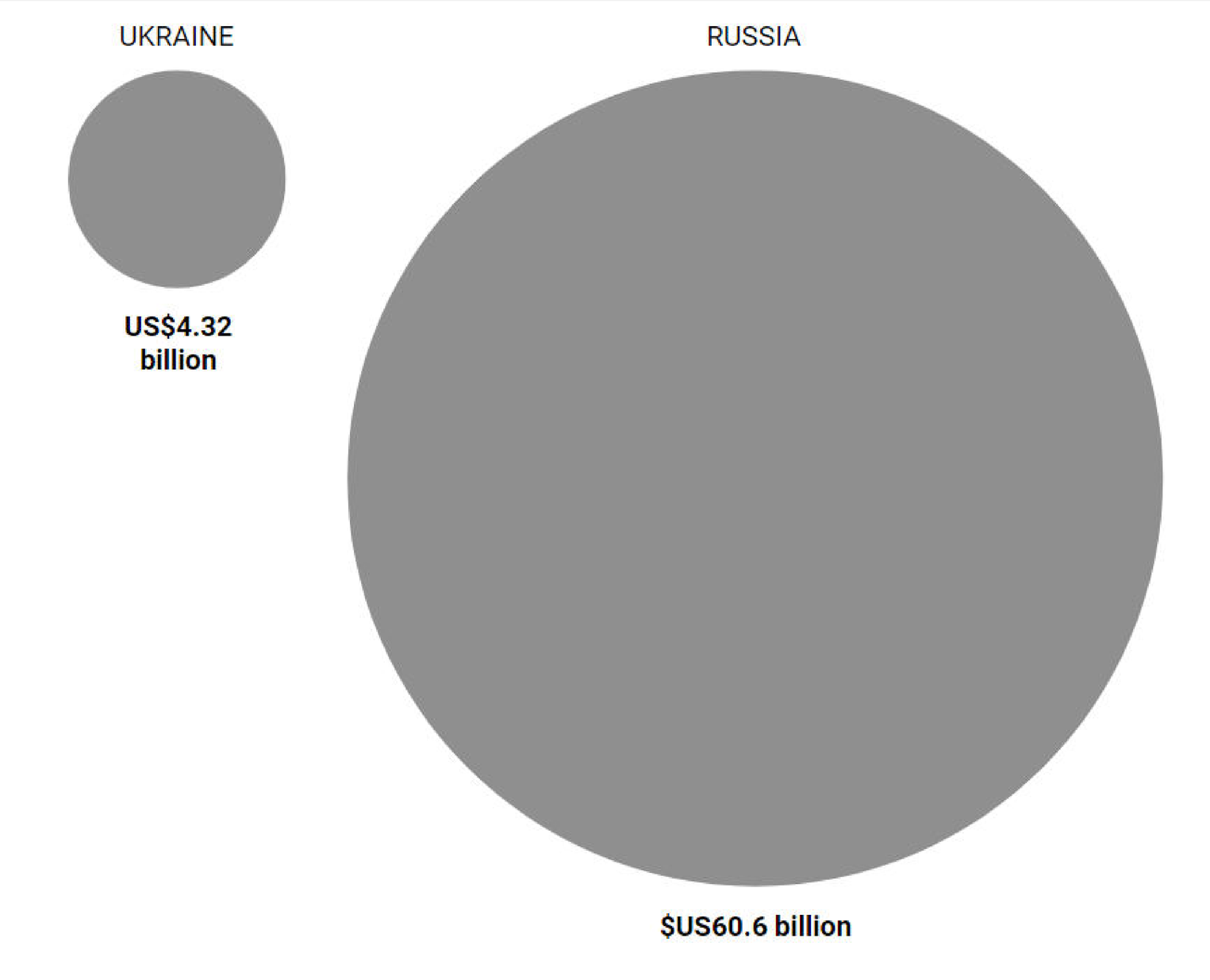
"Russia attacks Ukraine" by SCMP Graphic (South China Morning Post)

DEFENCE BUDGETS: RUSSIA VS UKRAINE (2020)

The national balance of forces is overwhelmingly in Russia's favour. Russian military spending in 2020 amounted to US$60.6 billion in 2020. Ukraine's was less than a 10th of that amount.



"Russia attacks Ukraine" by SCMP Graphic (South China Morning Post)



"Russia attacks Ukraine" by SCMP Graphic (South China Morning Post)

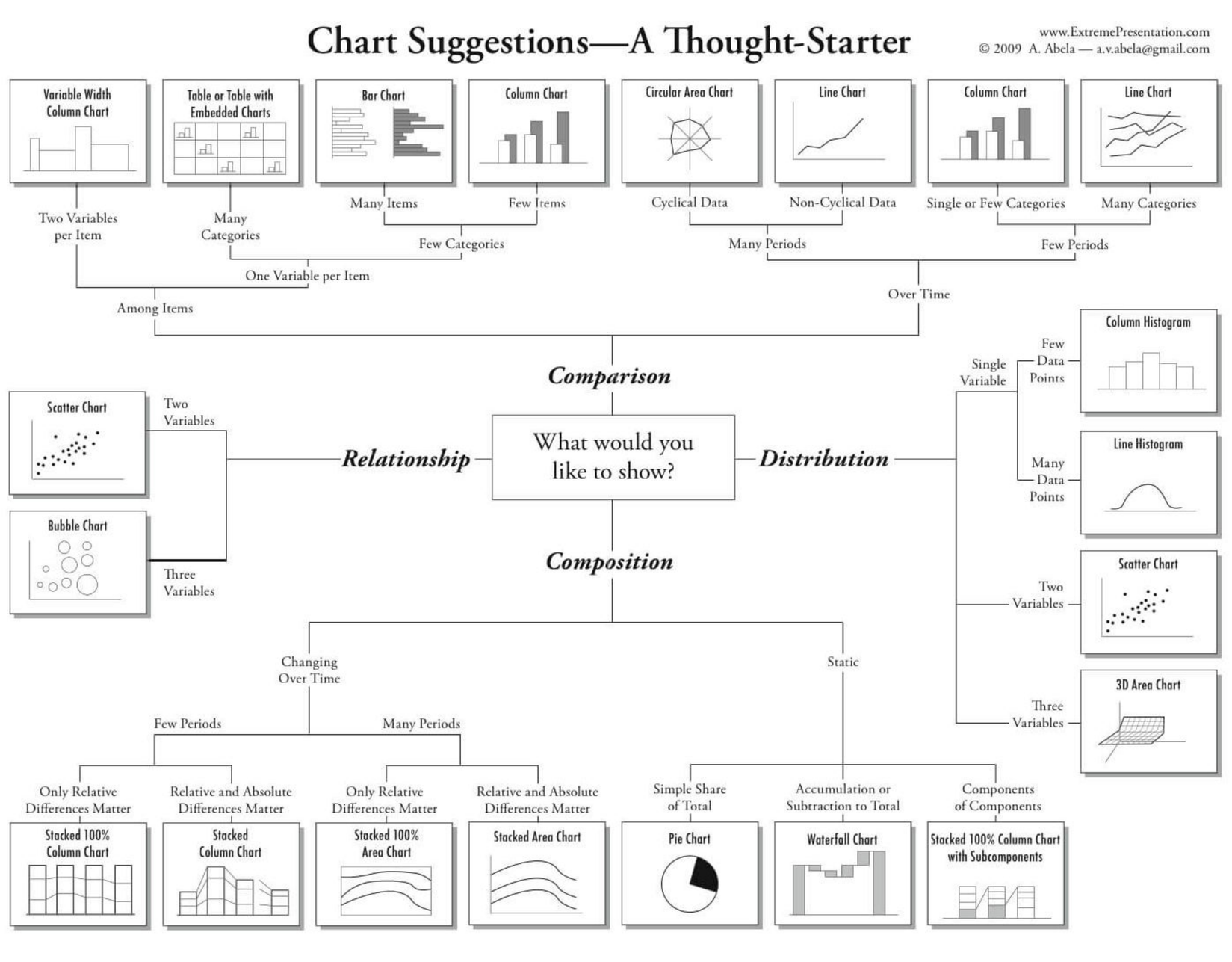
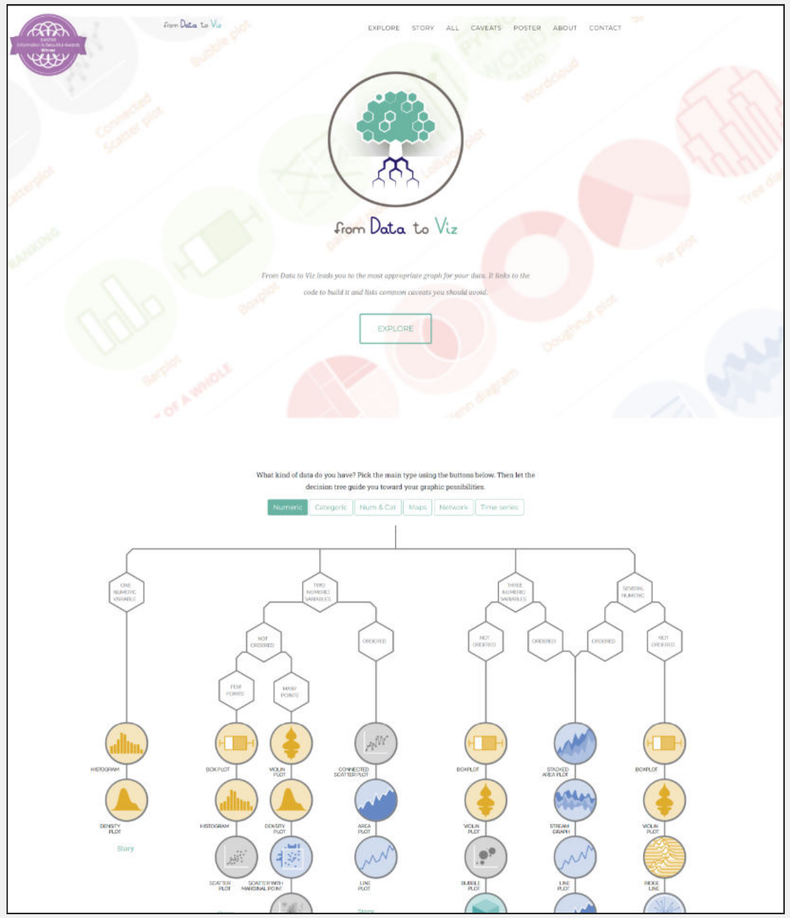


Chart Suggestions—A Thought-Starter
www.ExtremePresentation.com
© 2009 A. Abela — a.v.abela@gmail.com
Variable Width Column Chart
Table or Table with Embedded Charts
Bar Chart
Column Chart
Circular Area Chart
Line Chart
Column Chart
Line Chart
Two Variables per Item
Many Categories
Many Items
Few Items
Cyclical Data
Non-Cyclical Data
Single or Few Categories
Many Categories
Few Categories
Many Periods
Few Periods
One Variable per Item
Over Time
Among Items
Comparison
Scatter Chart
Two Variables
Relationship
What would you like to show?
Distribution
Bubble Chart
Three Variables
Composition
Single Variable
Few Data Points
Column Histogram
Many Data Points
Line Histogram
Two Variables
Scatter Chart
Three Variables
3D Area Chart
Changing Over Time
Static
Few Periods
Many Periods
Only Relative Differences Matter
Relative and Absolute Differences Matter
Only Relative Differences Matter
Relative and Absolute Differences Matter
Simple Share of Total
Accumulation or Subtraction to Total
Components of Components
Stacked 100% Column Chart
Stacked Column Chart
Stacked 100% Area Chart
Stacked Area Chart
Pie Chart
Waterfall Chart
Stacked 100% Column Chart with Subcomponents

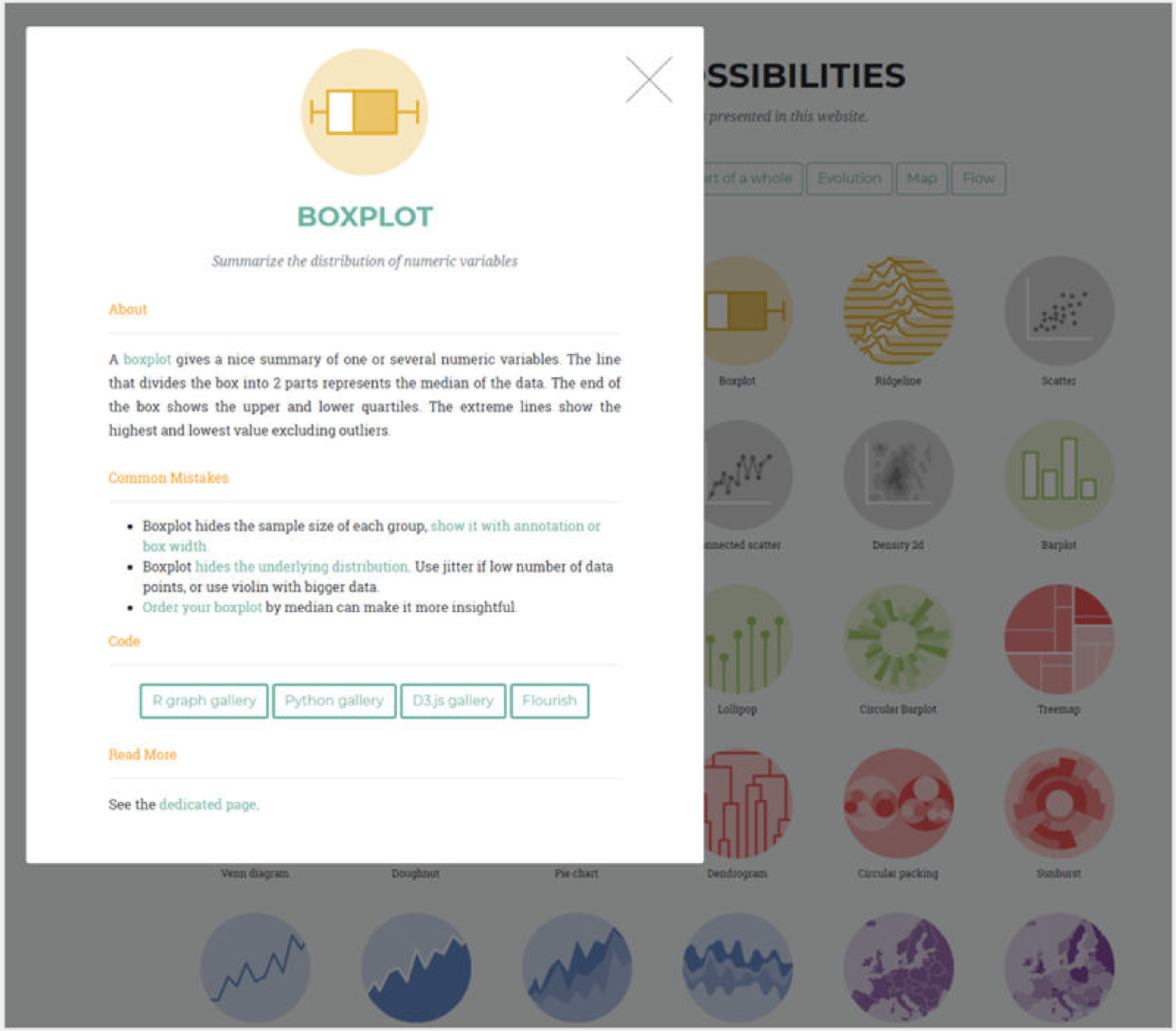


data-to-viz.com

datavizproject.com

visualizationuniverse.com

Source: data-to-viz.com



Source: data-to-viz.com

Group Exercise

Have a look at the “Chart of Charts” poster

☛ Fill one dot for each chart type you have used already to communicate your data

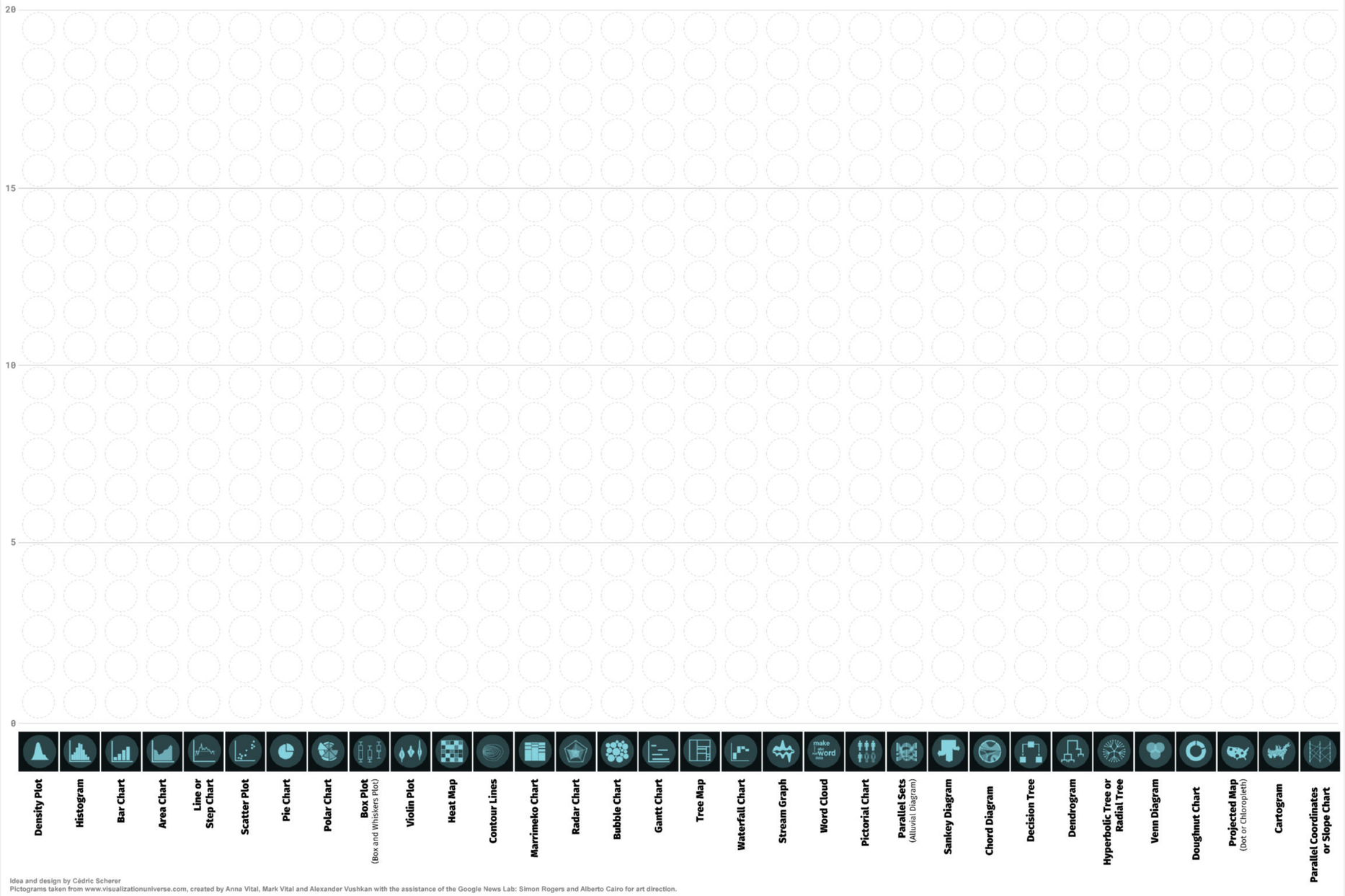
Have a look at the “Chart Choice Helpers”

☛ Explore and discuss potential chart types for future data visualizations

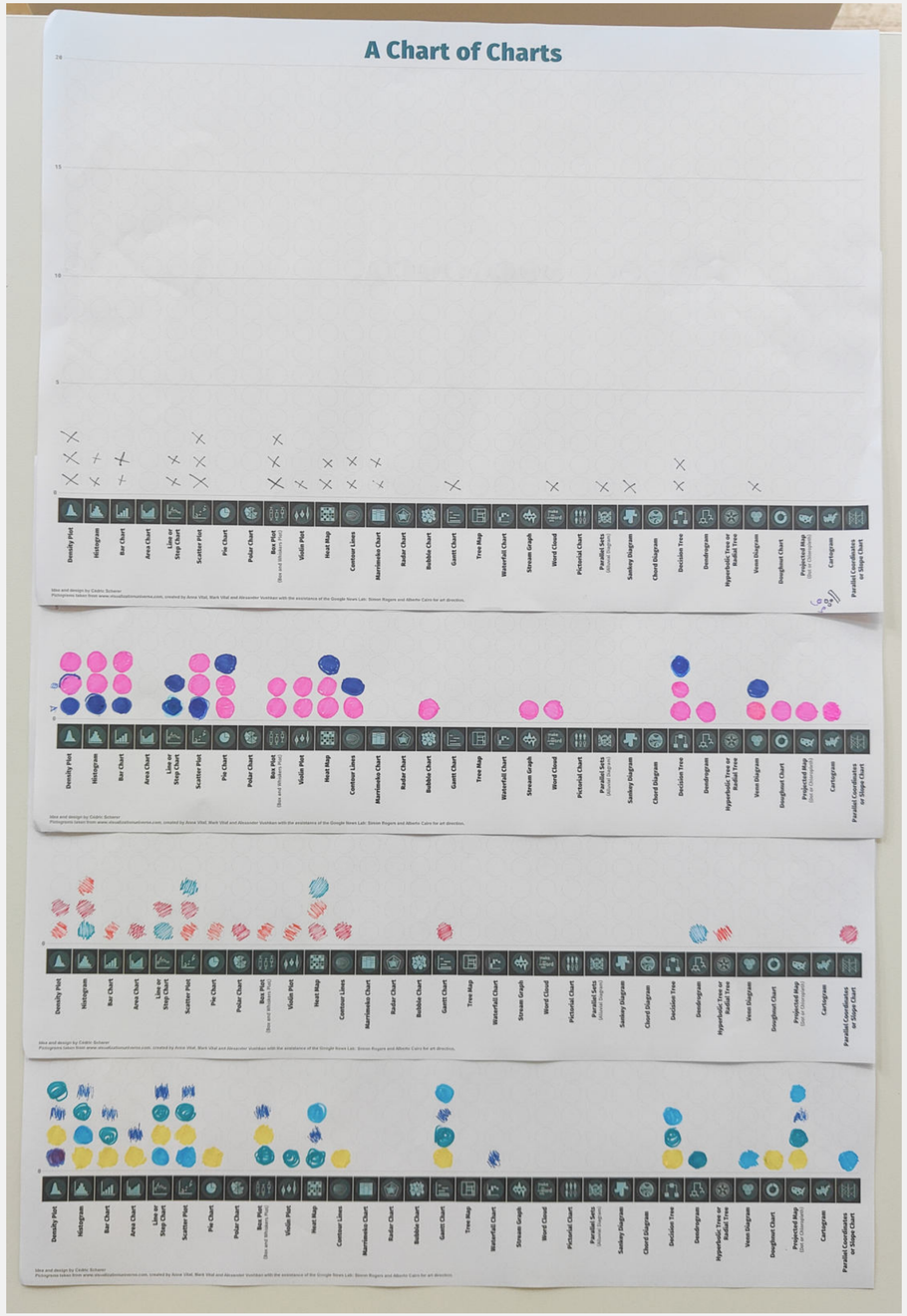
☛ Look up the chart types on the poster you didn’t know

☛ Are there chart types you find hard to read / understand?

A Chart of Charts

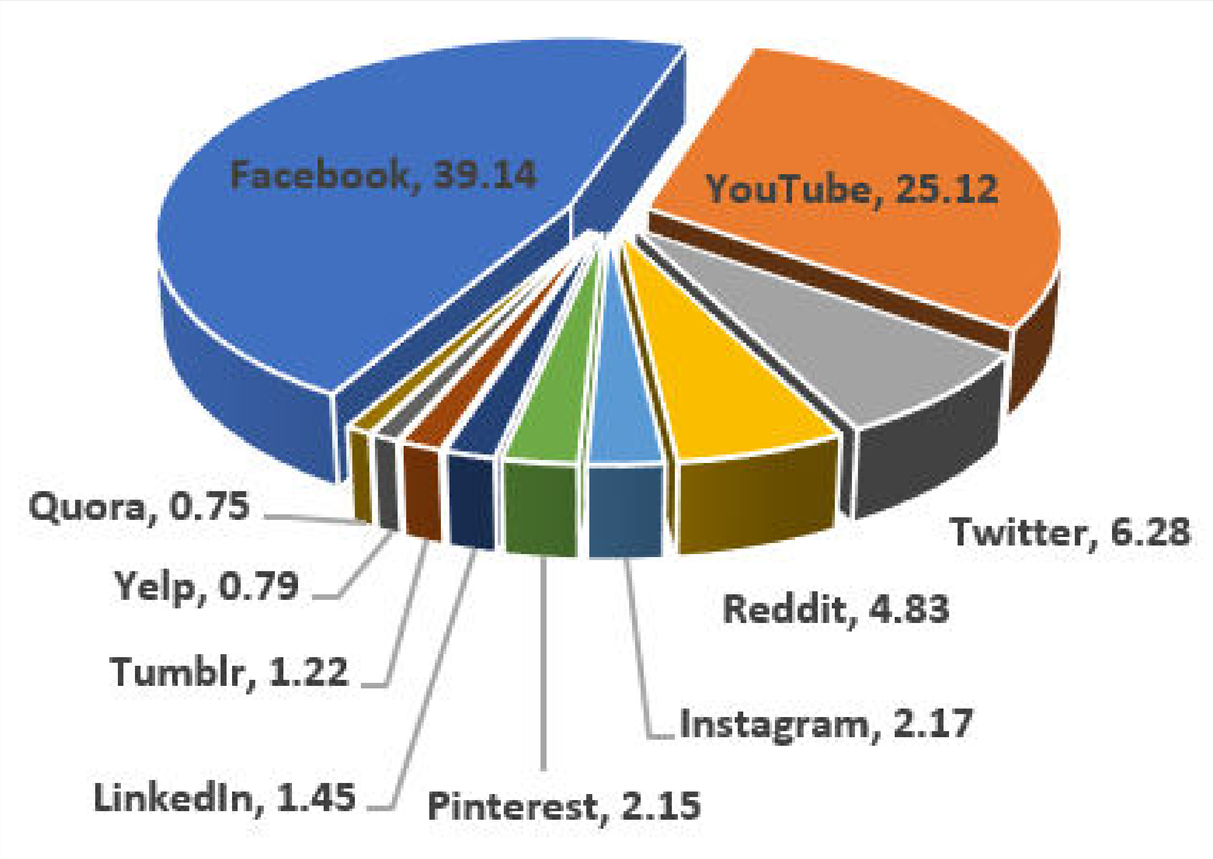


Idea and design by Cédric Scherer
Pictograms taken from www.visualizationuniverse.com, created by Anna Vital, Mark Vital and Alexander Vushkan with the assistance of the Google News Lab: Simon Rogers and Alberto Cairo for art direction.



cedricscherer.com @CedScherer z3tt

Group Exercise

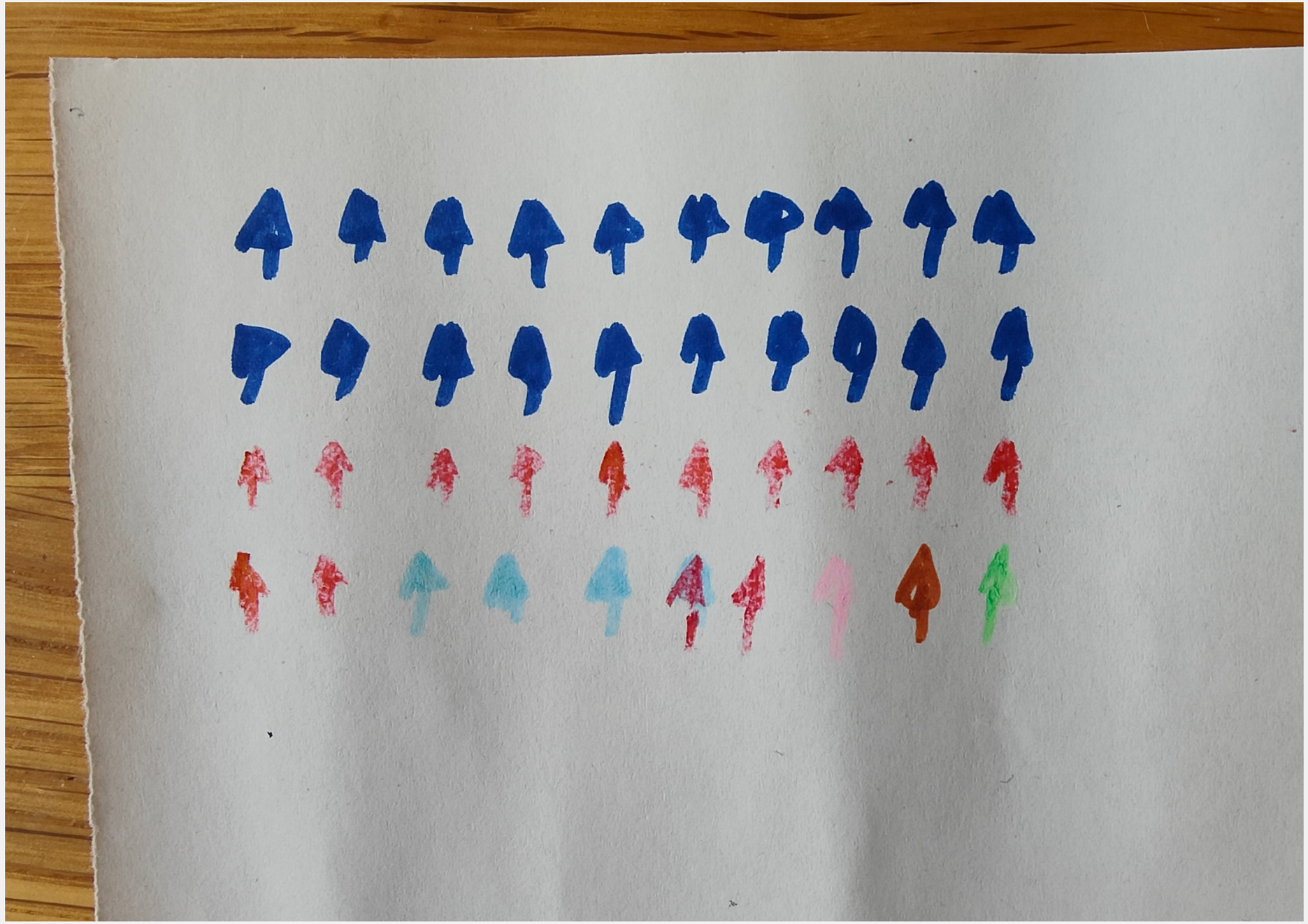


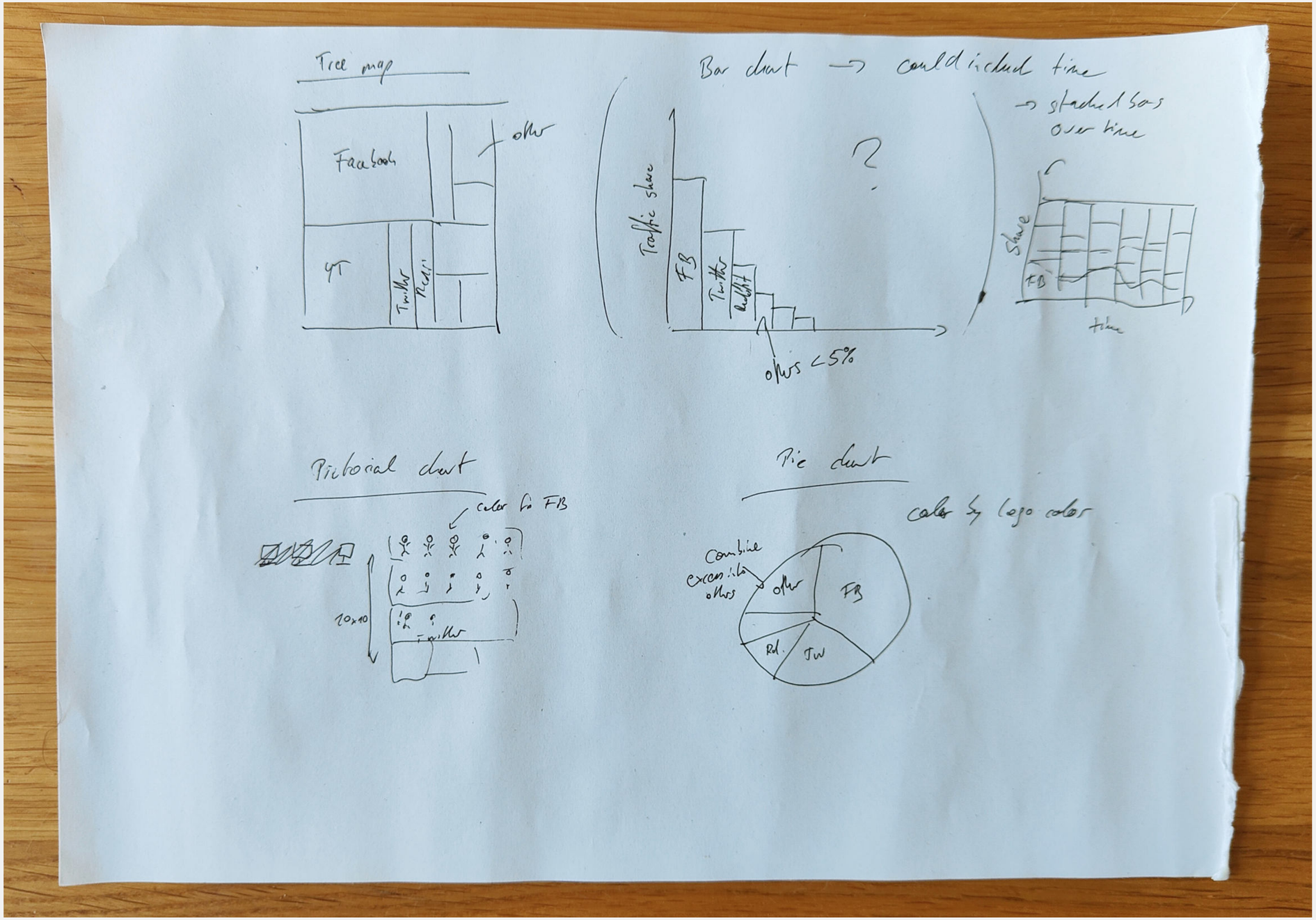
Market share of visits to social network sites in November 2017, found in the blog post on "Why you shouldn't use pie charts" by The University of Melbourne

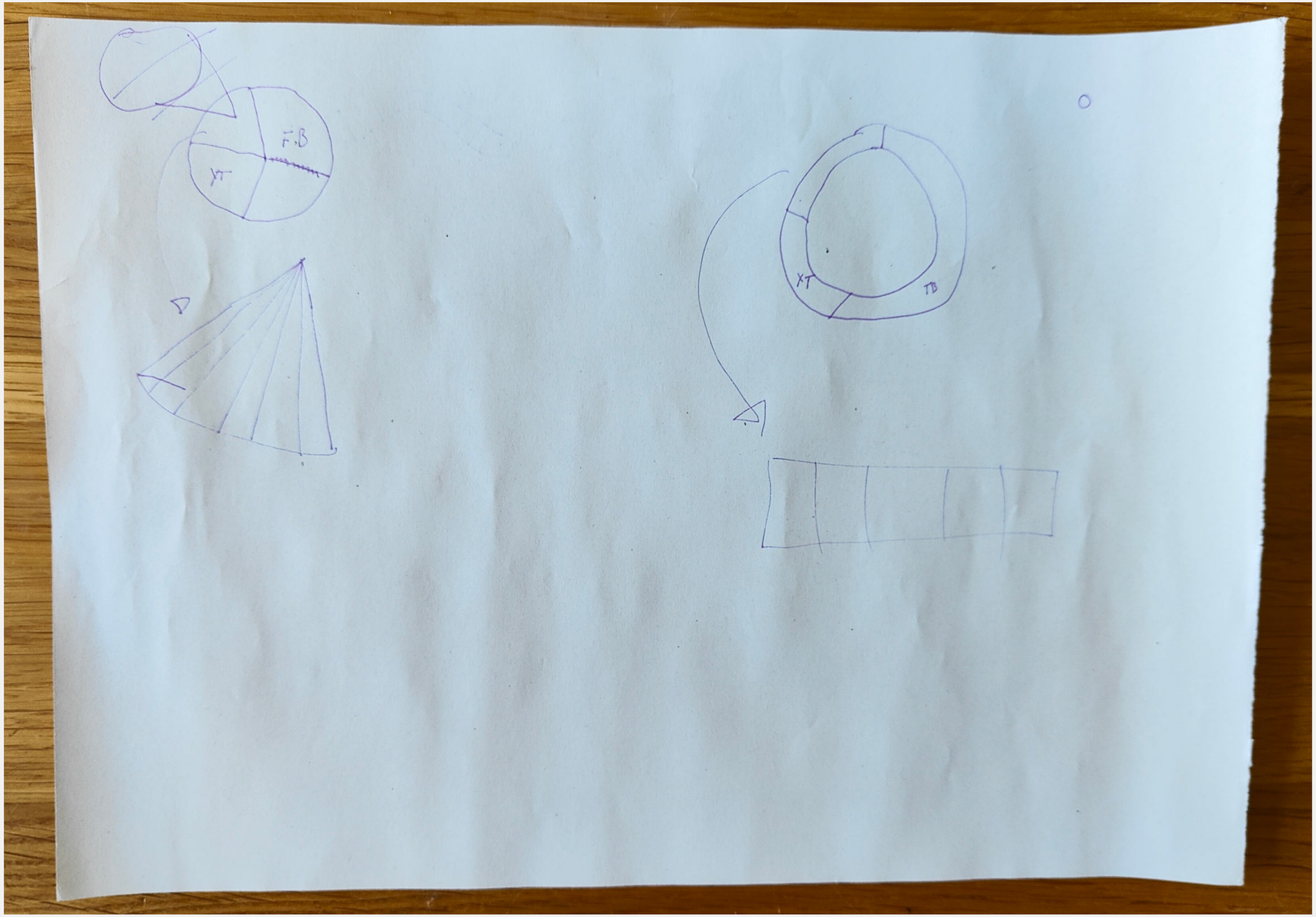
Discuss this graphic

- ☛ What is the main message?
- ☛ Does the chart type work?
- ☛ Is there a better chart type?
- ☛ **Draw an alternative version using a chart type that is more suitable!**

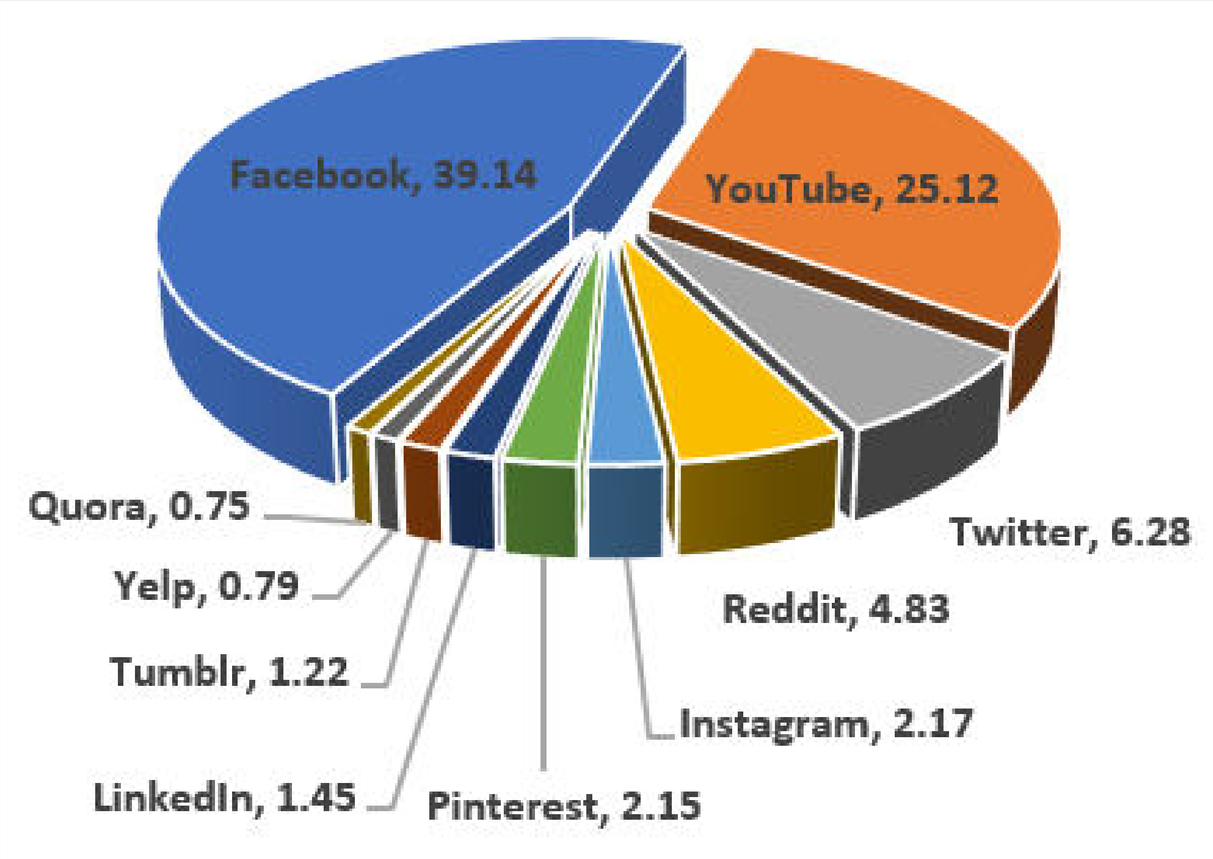
Share of visits per social media platform
Visits share
50
25
0
Facebook
Youtube
Twitter
Reddit
Instagram
Pinterest
LinkedIn
Tumblr
Yelp
Quora

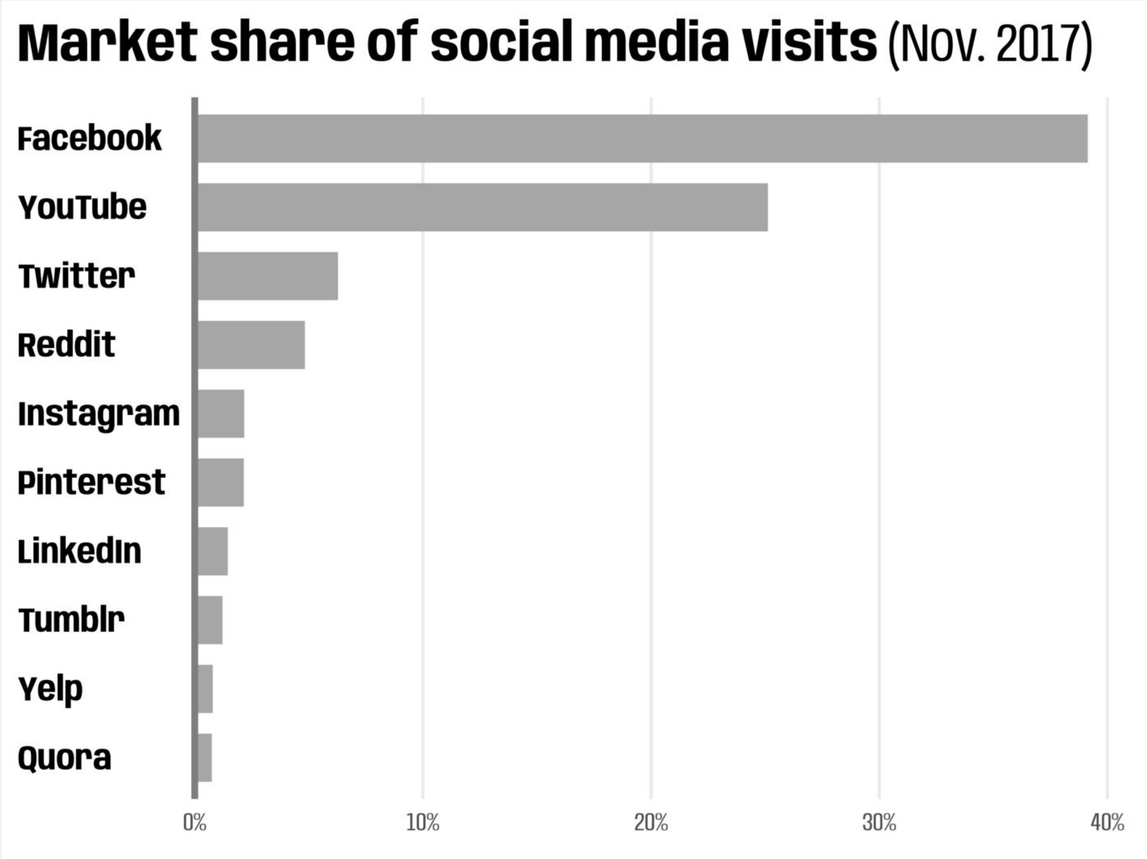



Tree map
Facebook
other
YT
Twitter
Reddit
Bar chart → could include time
Traffic share
FB
Twitter
Reddit
?
others <5%
→ stacked bars over time
Share
FB
time
Pictorial chart
color for FB
10x10
Twitter
Pie chart
color by logo color
Combine excessive others
other
FB
Rd.
TW


F.B
XT
XT
FB

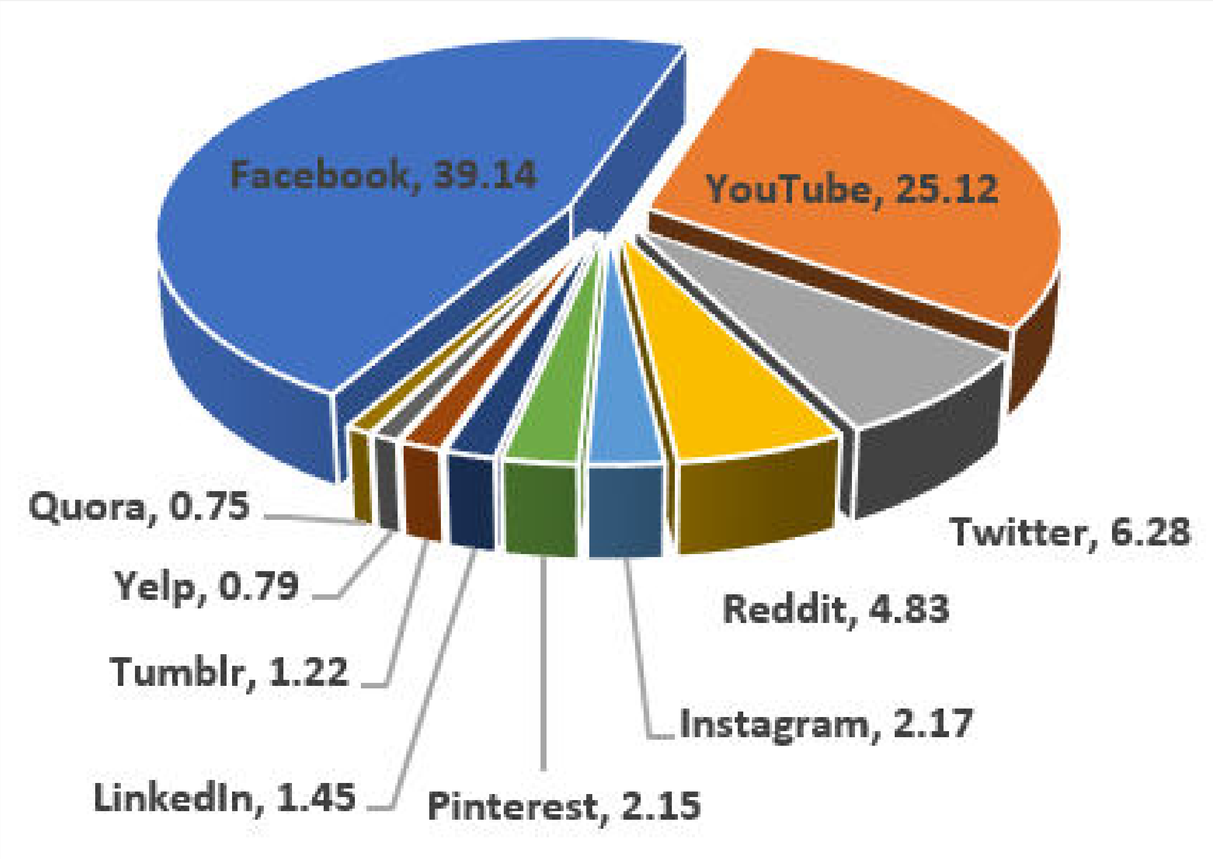
Group Exercise

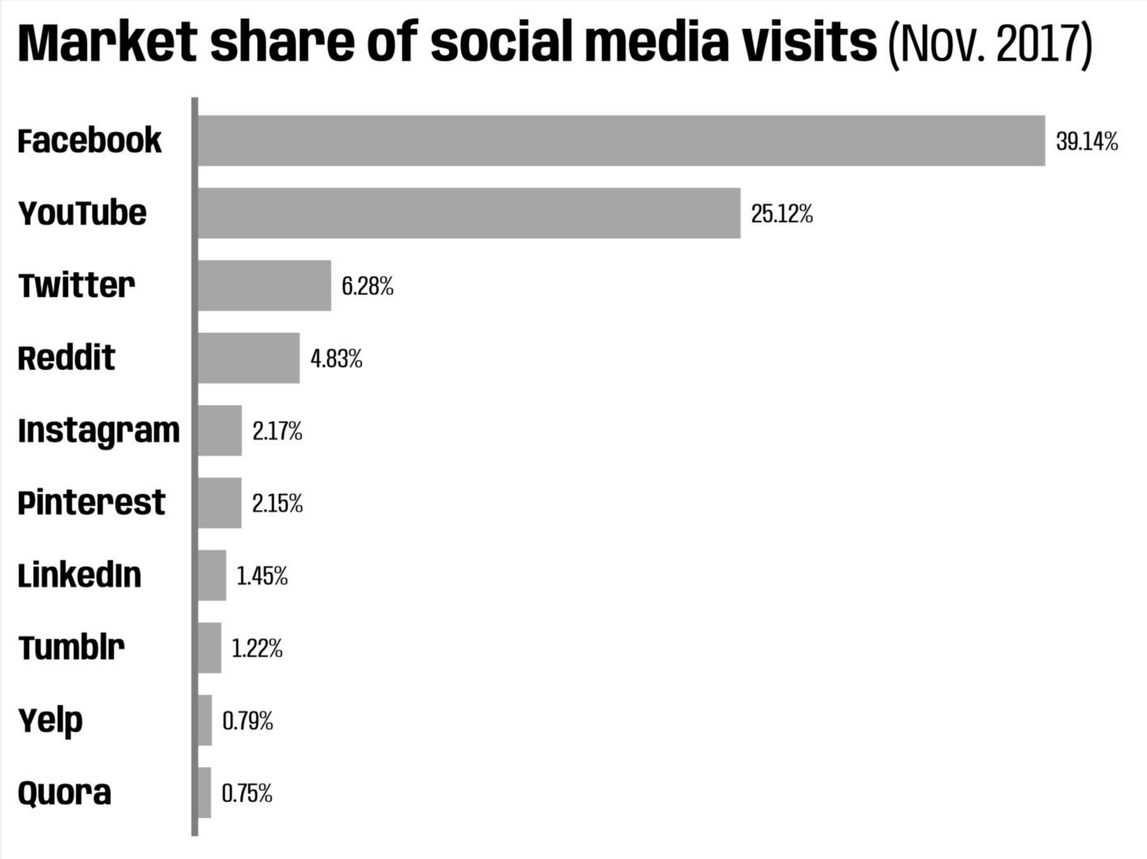




Market share of visits to social network sites in November 2017, found in the blog post on "Why you shouldn't use pie charts" by The University of Melbourne

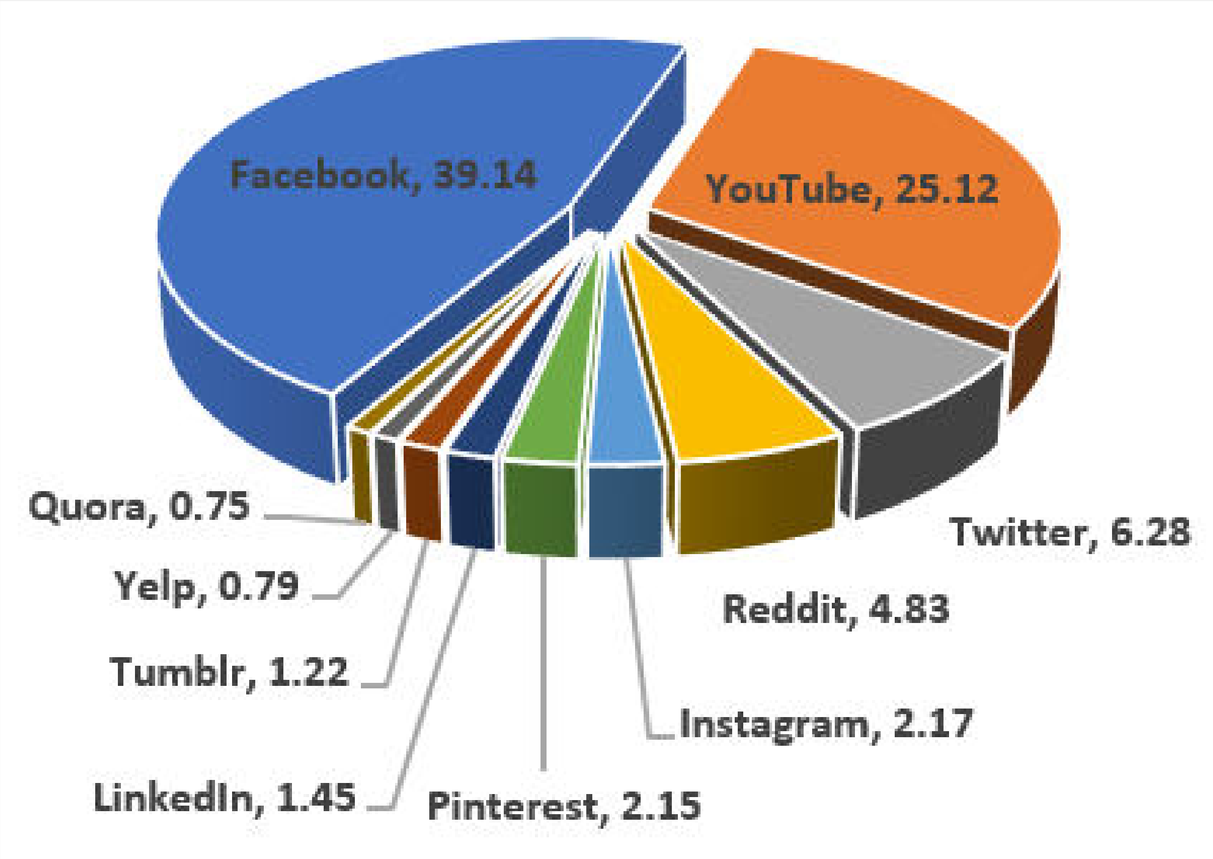
Group Exercise

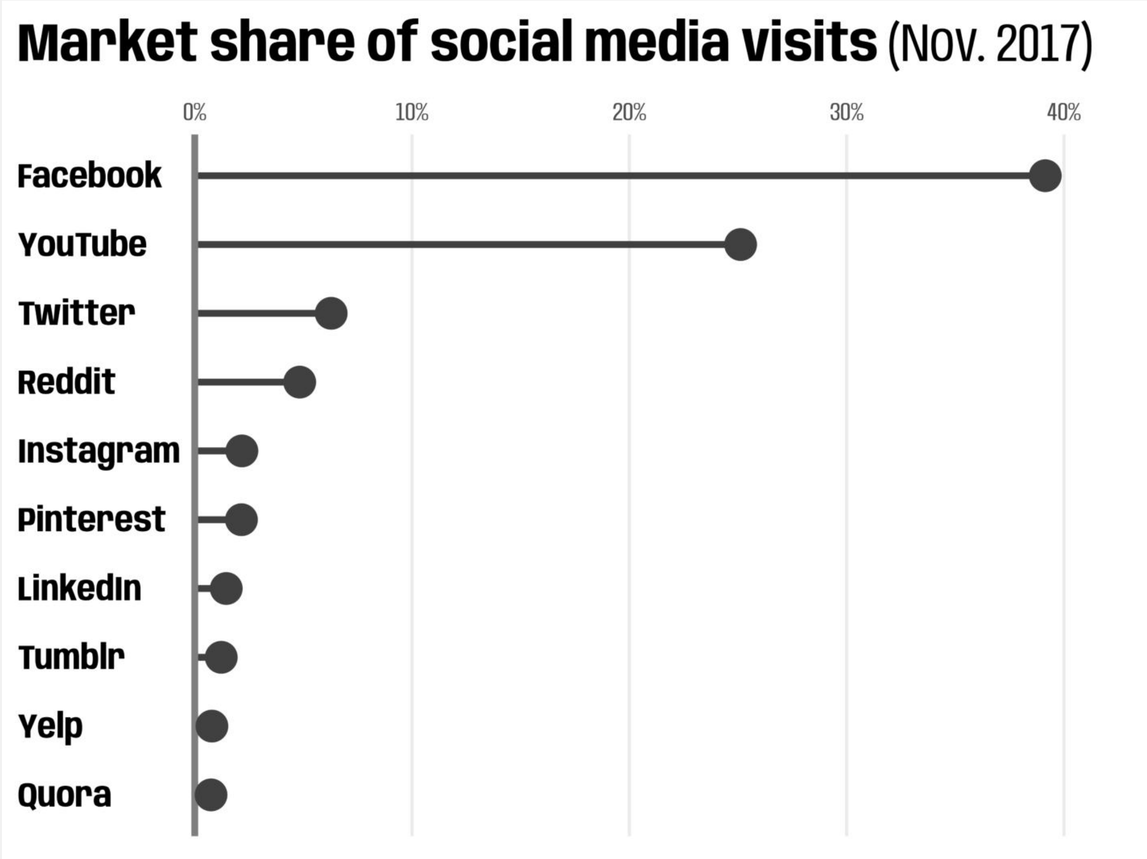




Market share of visits to social network sites in November 2017, found in the blog post on "Why you shouldn't use pie charts" by The University of Melbourne

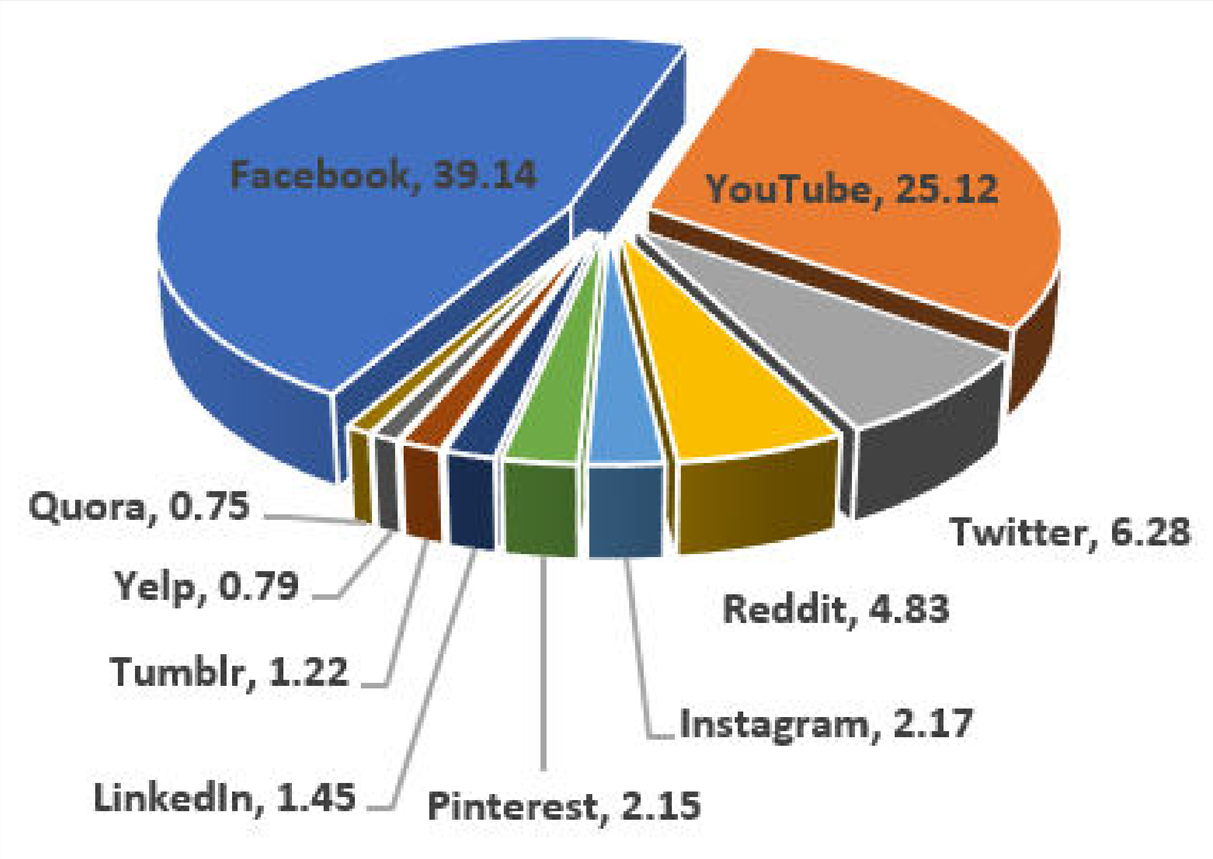
Group Exercise

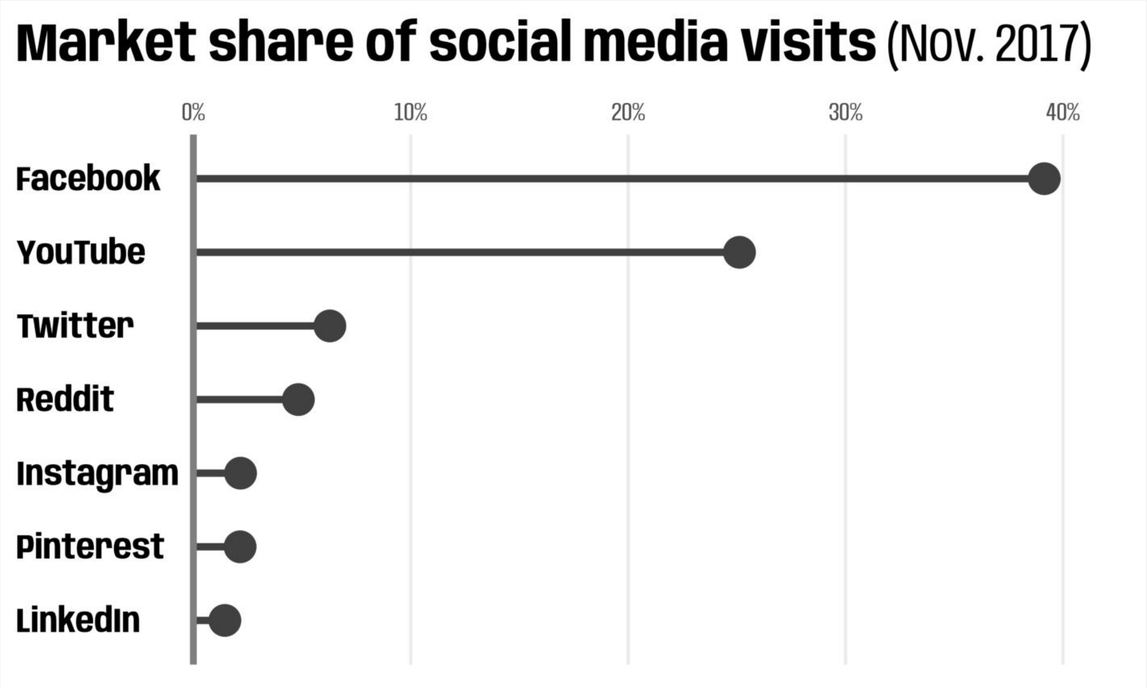




Market share of visits to social network sites in November 2017, found in the blog post on "Why you shouldn't use pie charts" by The University of Melbourne

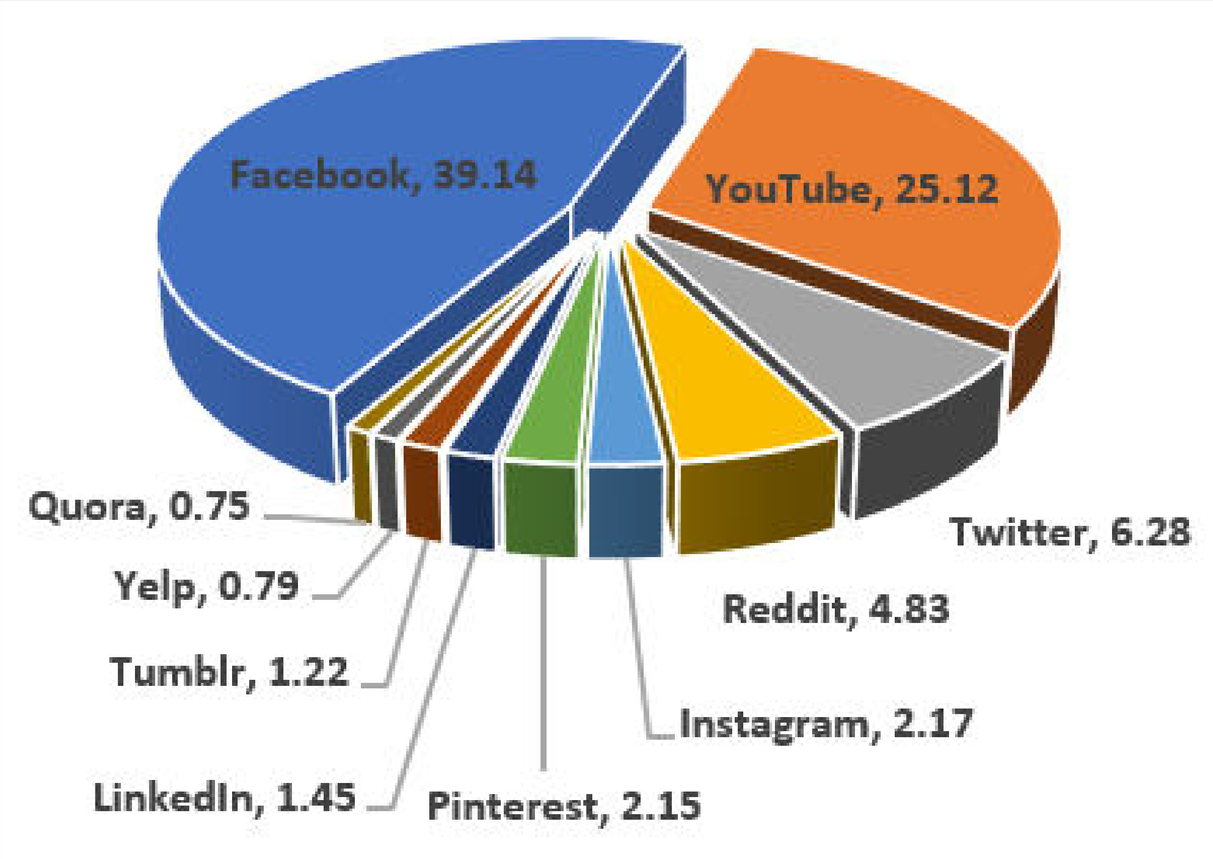
Group Exercise

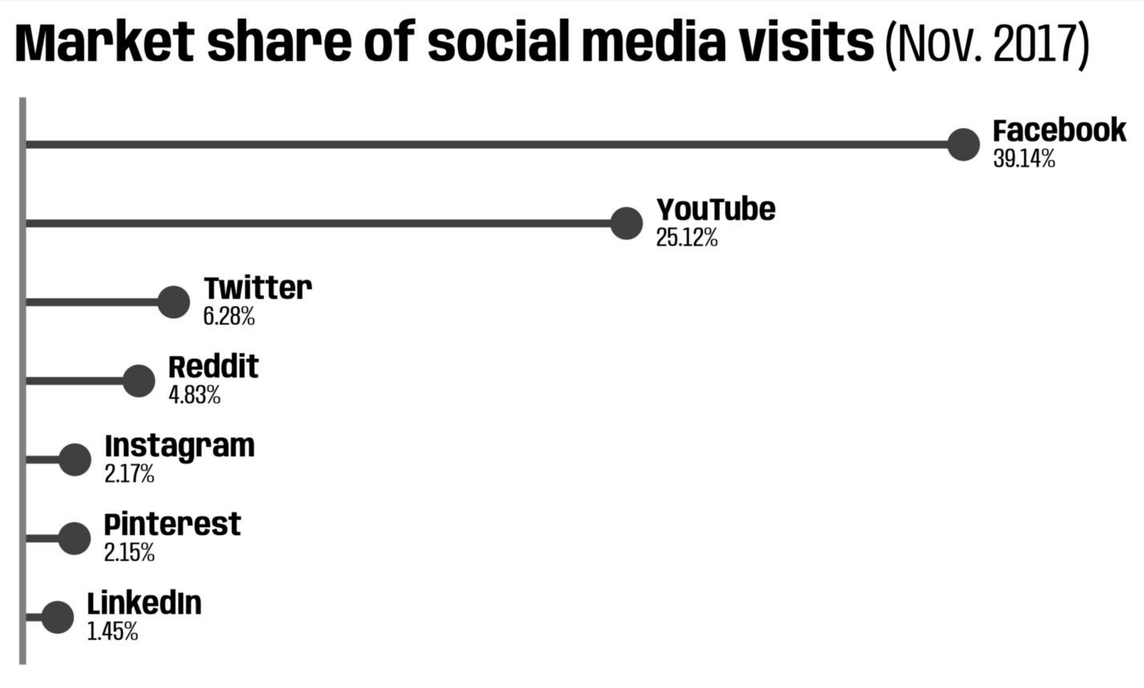




Market share of visits to social network sites in November 2017, found in the blog post on "Why you shouldn't use pie charts" by The University of Melbourne

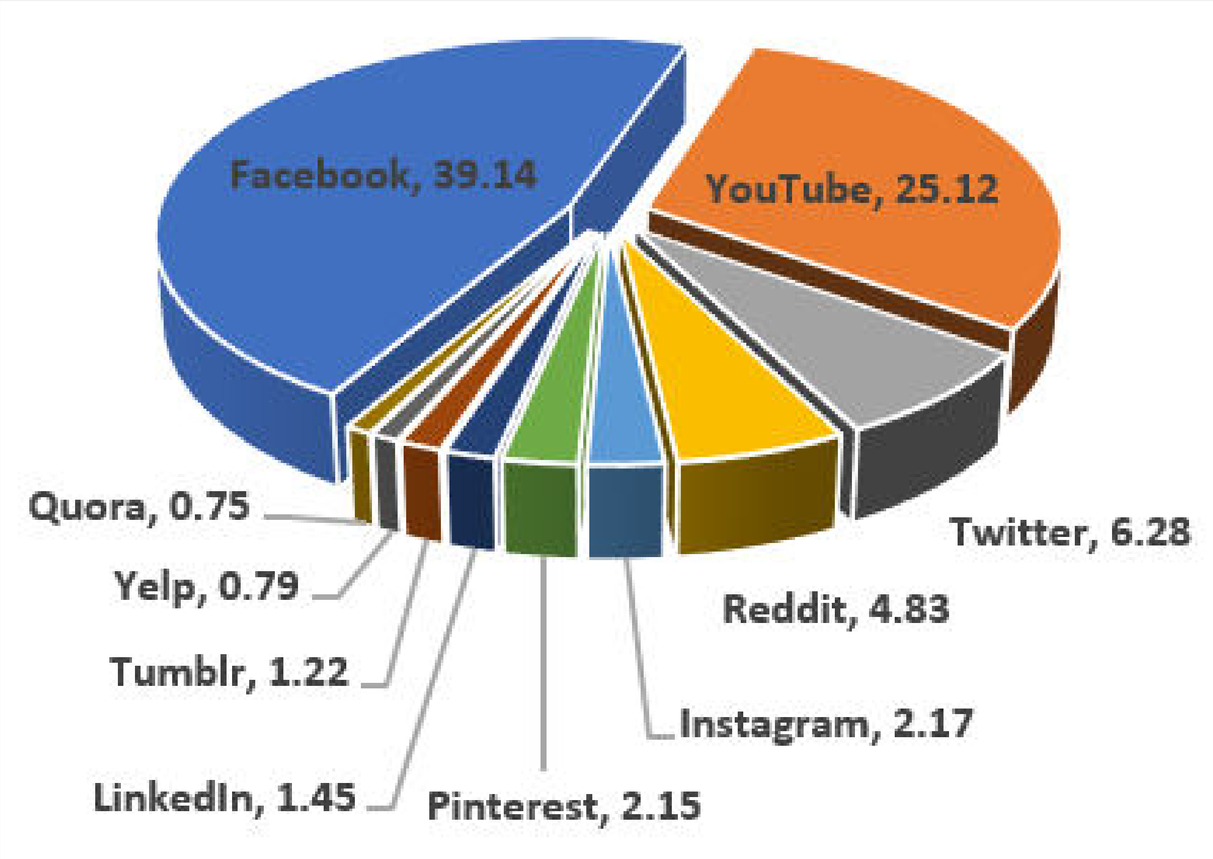
Group Exercise

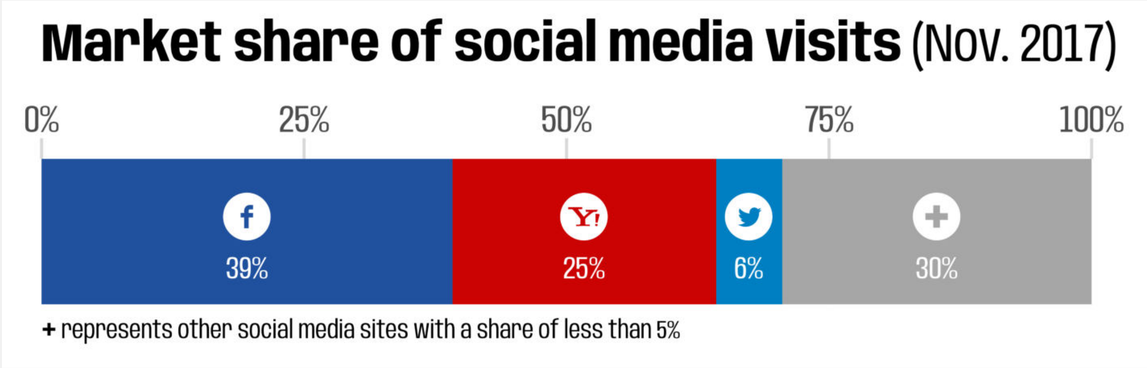




Market share of visits to social network sites in November 2017, found in the blog post on "Why you shouldn't use pie charts" by The University of Melbourne

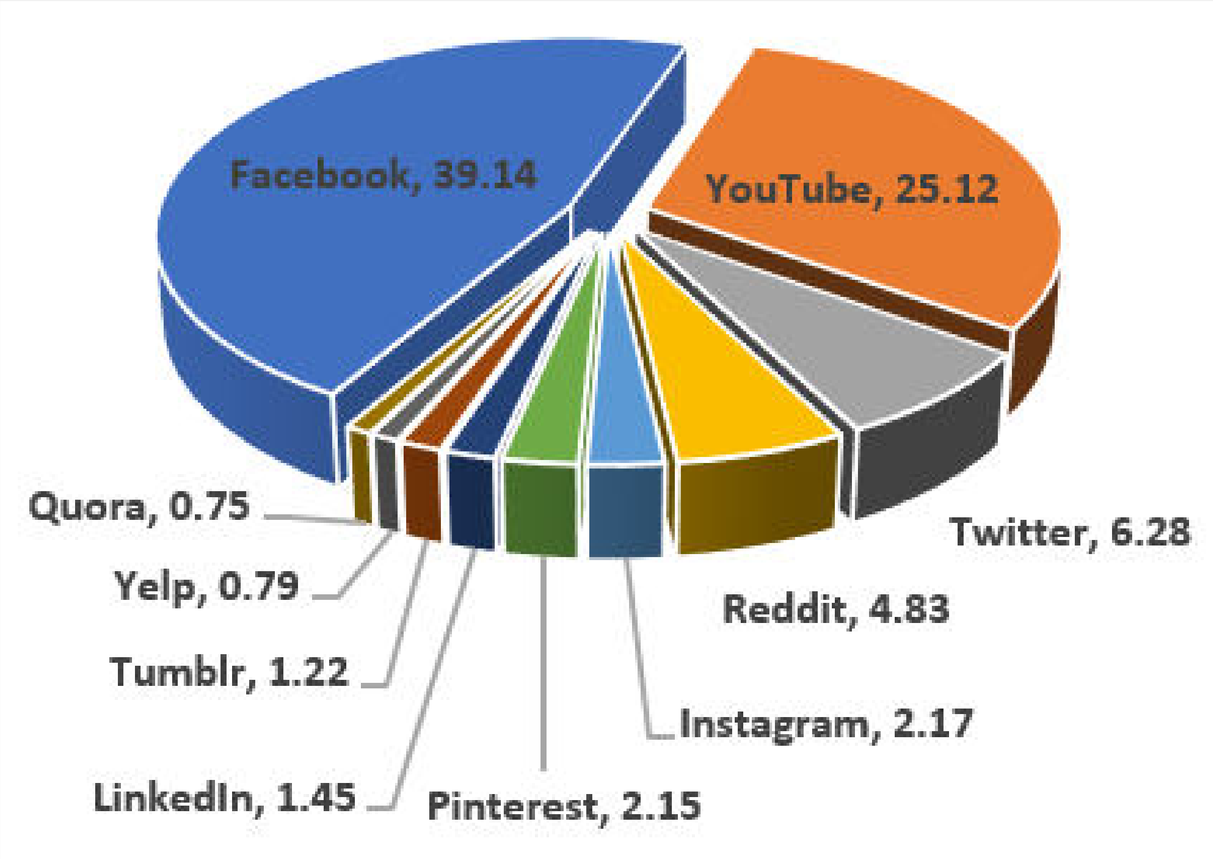
Group Exercise



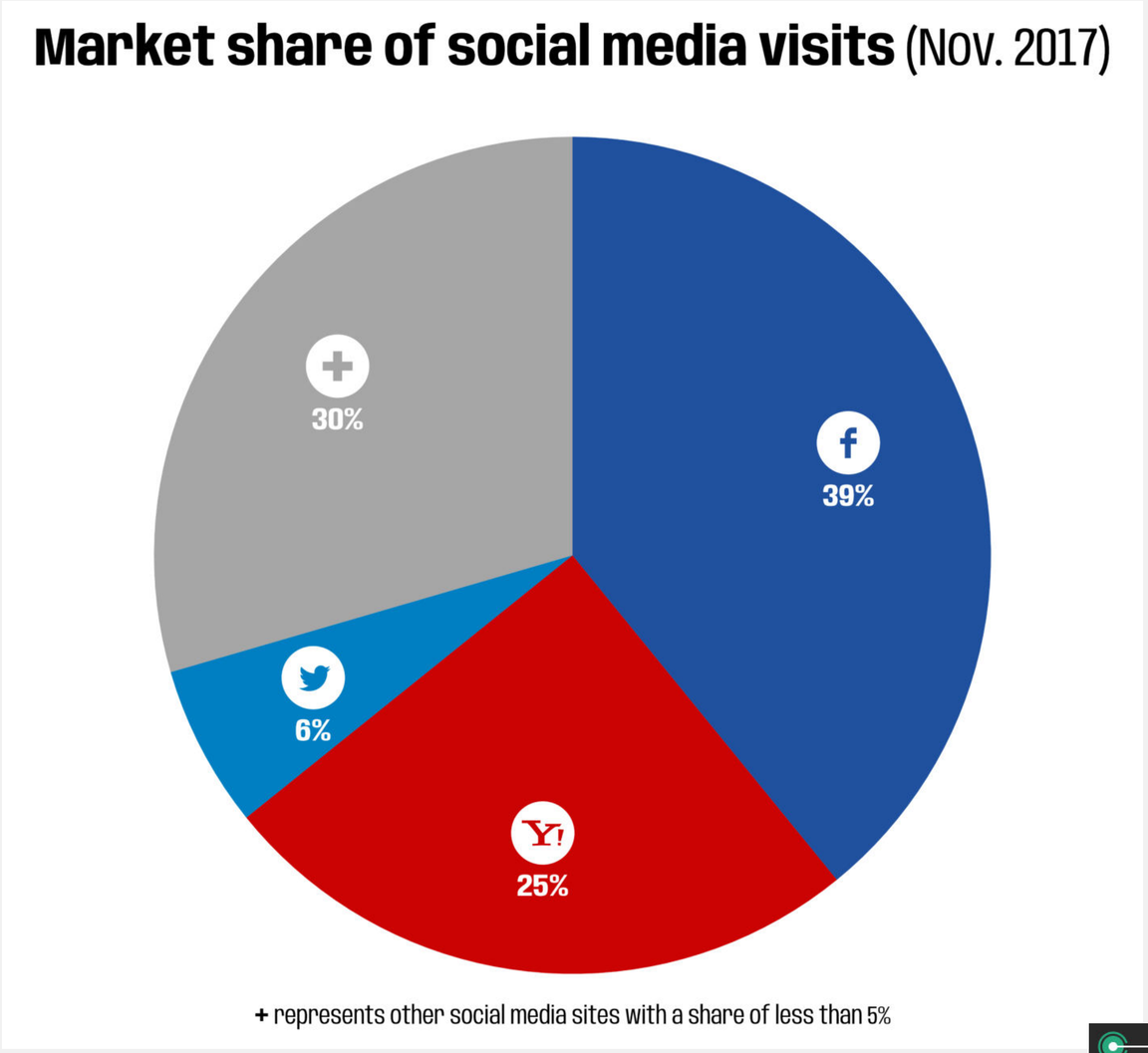


Market share of visits to social network sites in November 2017, found in the blog post on "Why you shouldn't use pie charts" by The University of Melbourne

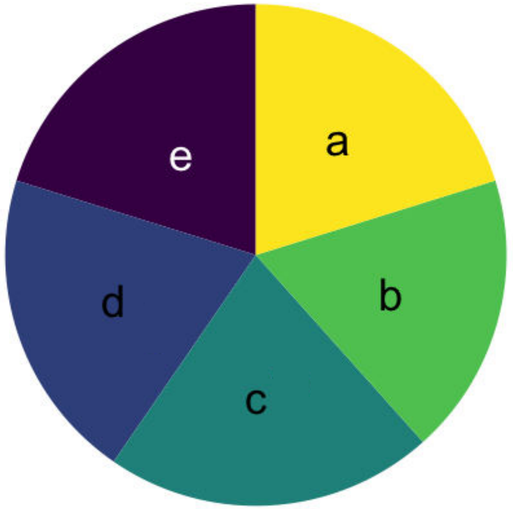
Group Exercise

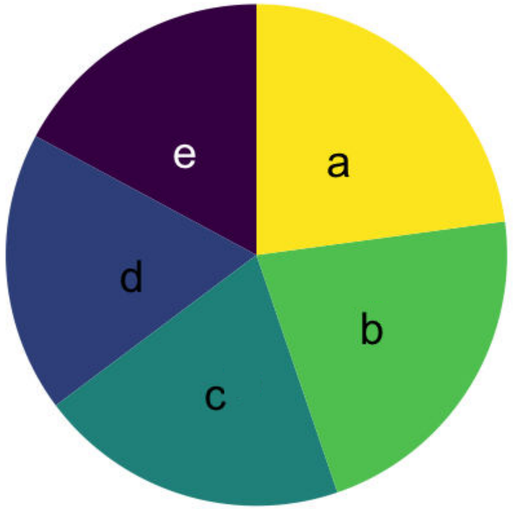


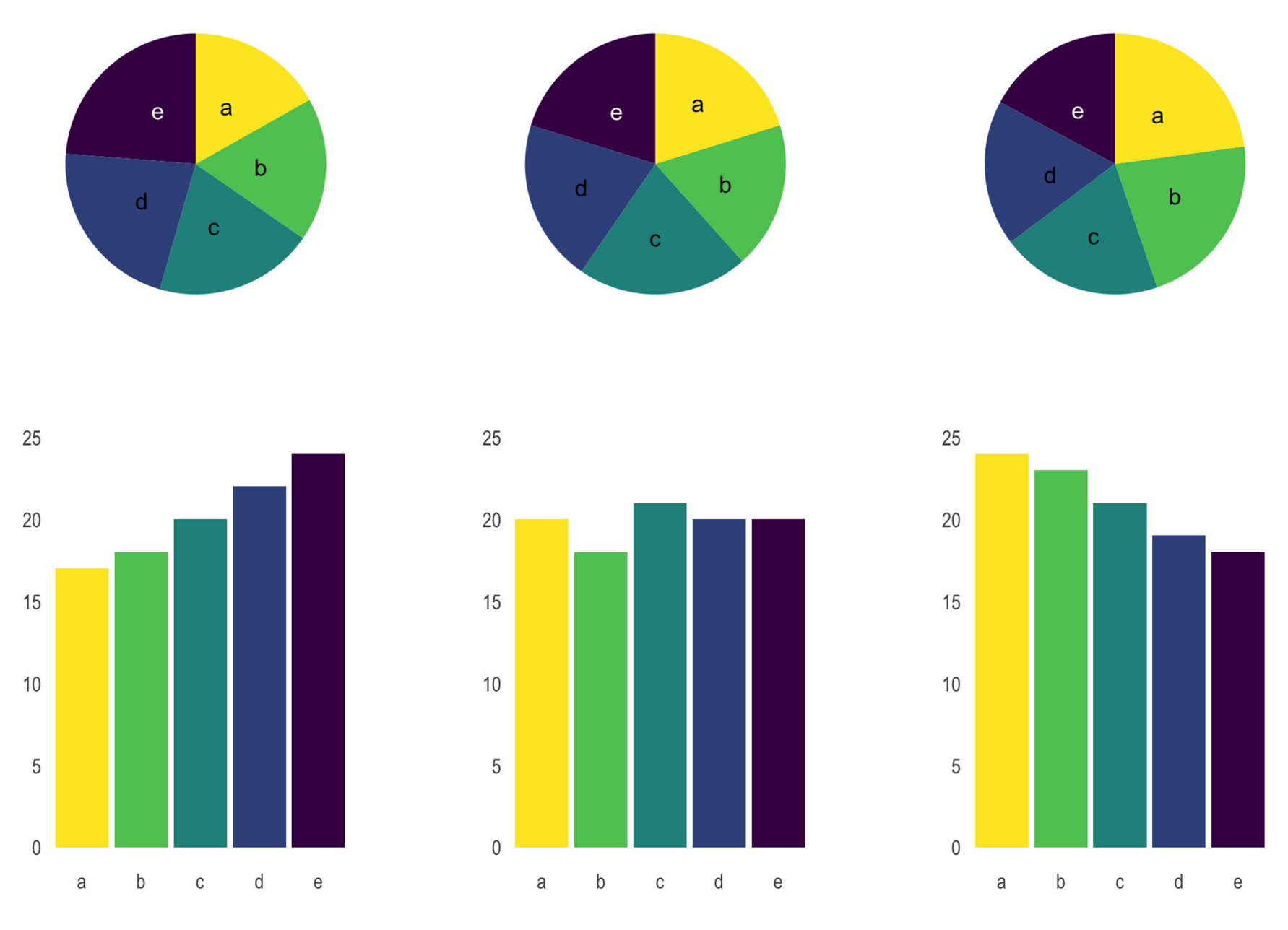
Market share of visits to social network sites in November 2017, found in the blog post on "Why you shouldn't use pie charts" by The University of Melbourne

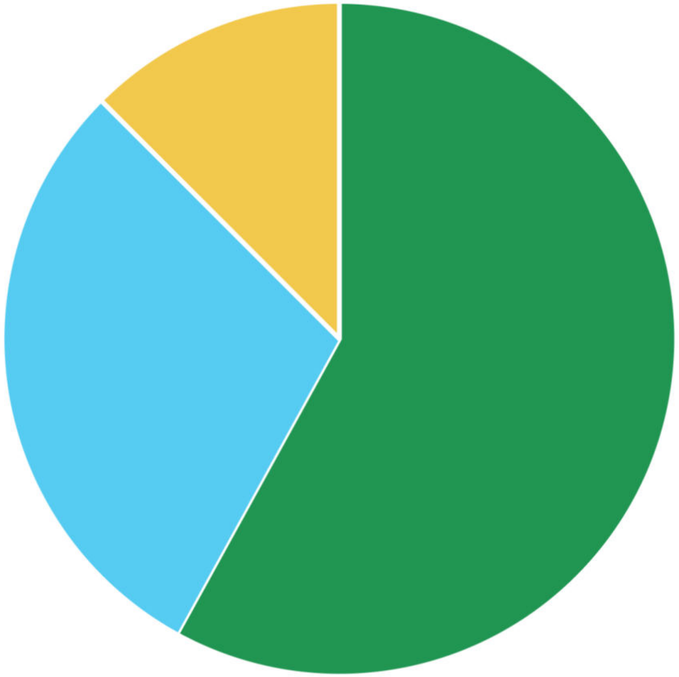


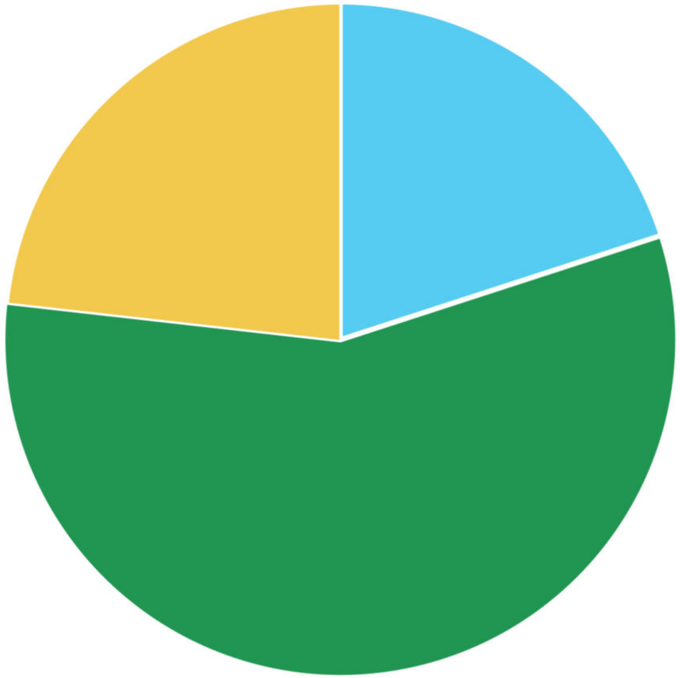
e
a
b
d
c


e
a
d
b
c

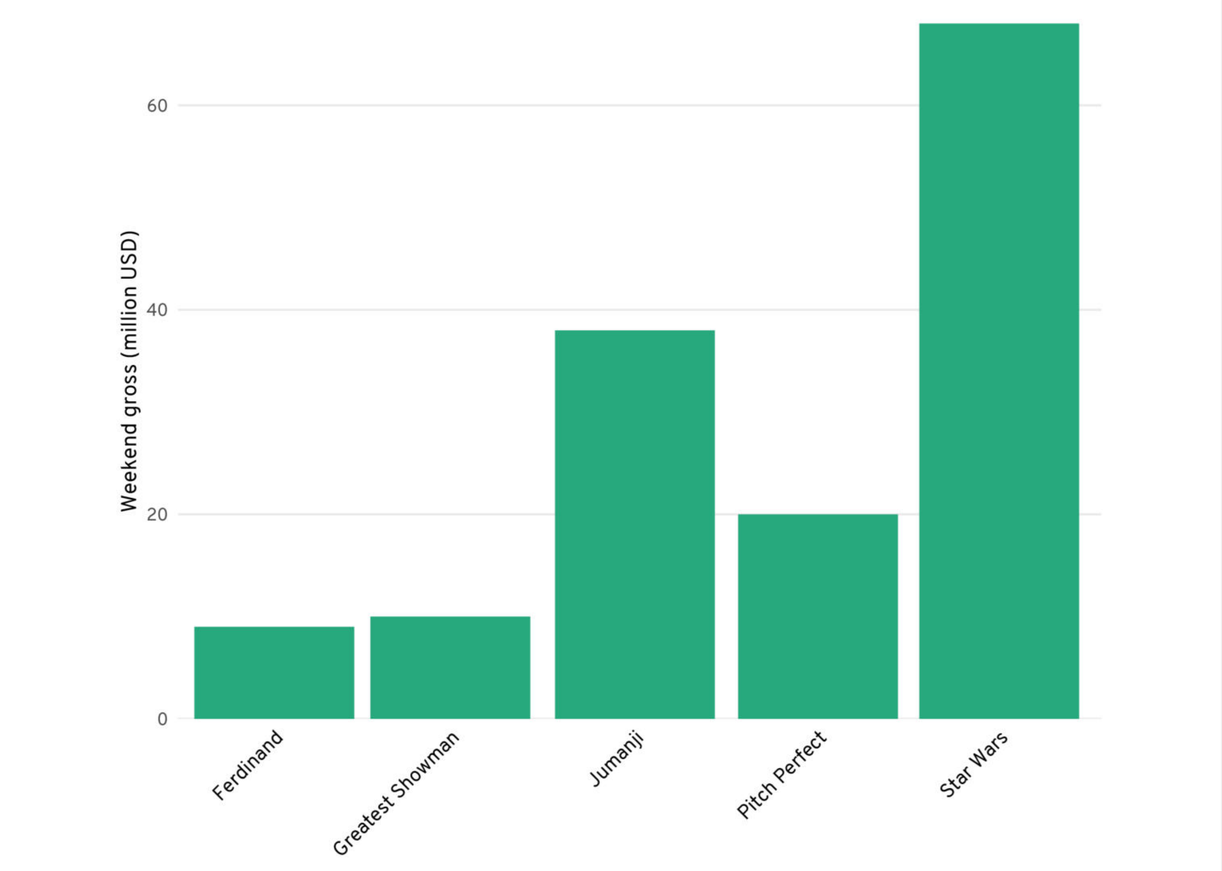

e
a
d
b
c

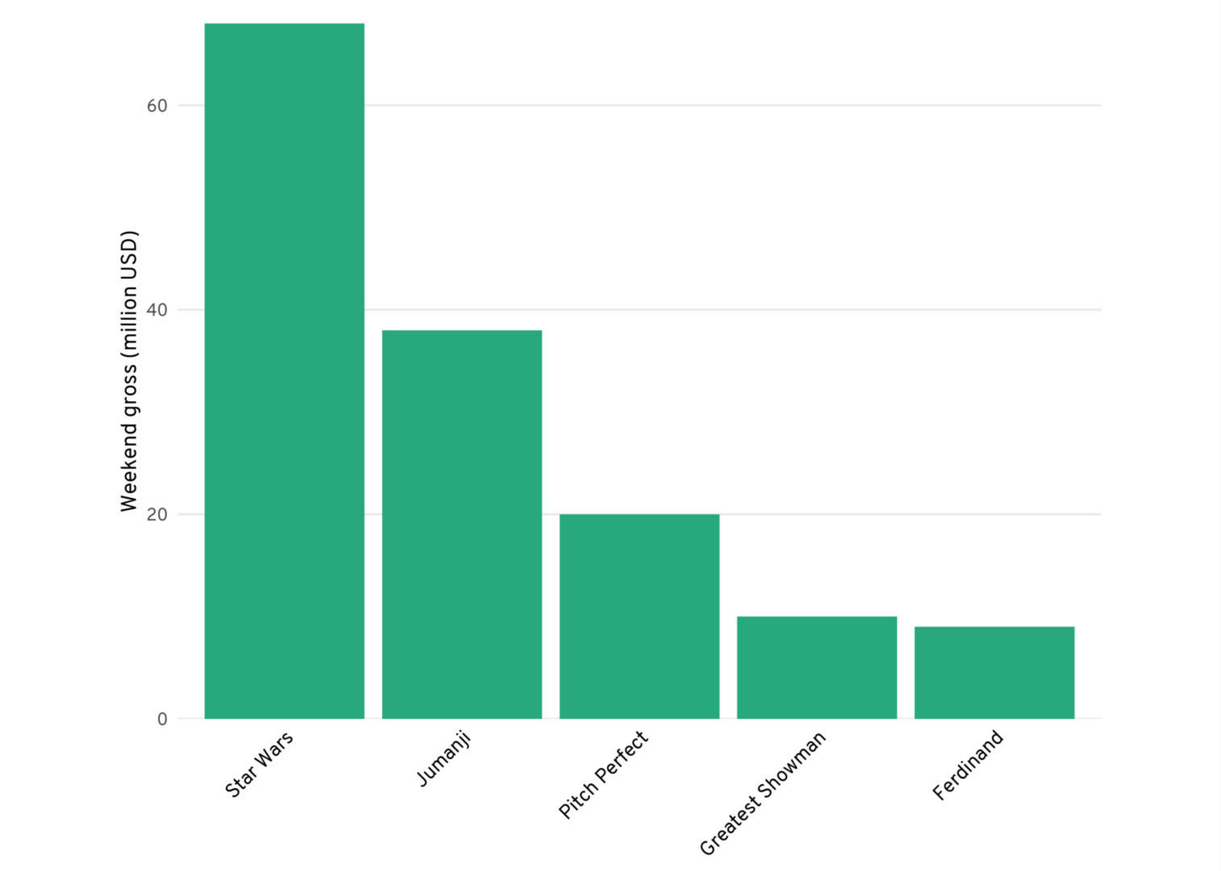

e
a
b
c
d
e
a
b
c
d
e
a
b
c
d
25
20
15
10
5
0
a
b
c
d
e
25
20
15
10
5
0
a
b
c
d
e
25
20
15
10
5
0
a
b
c
d
e



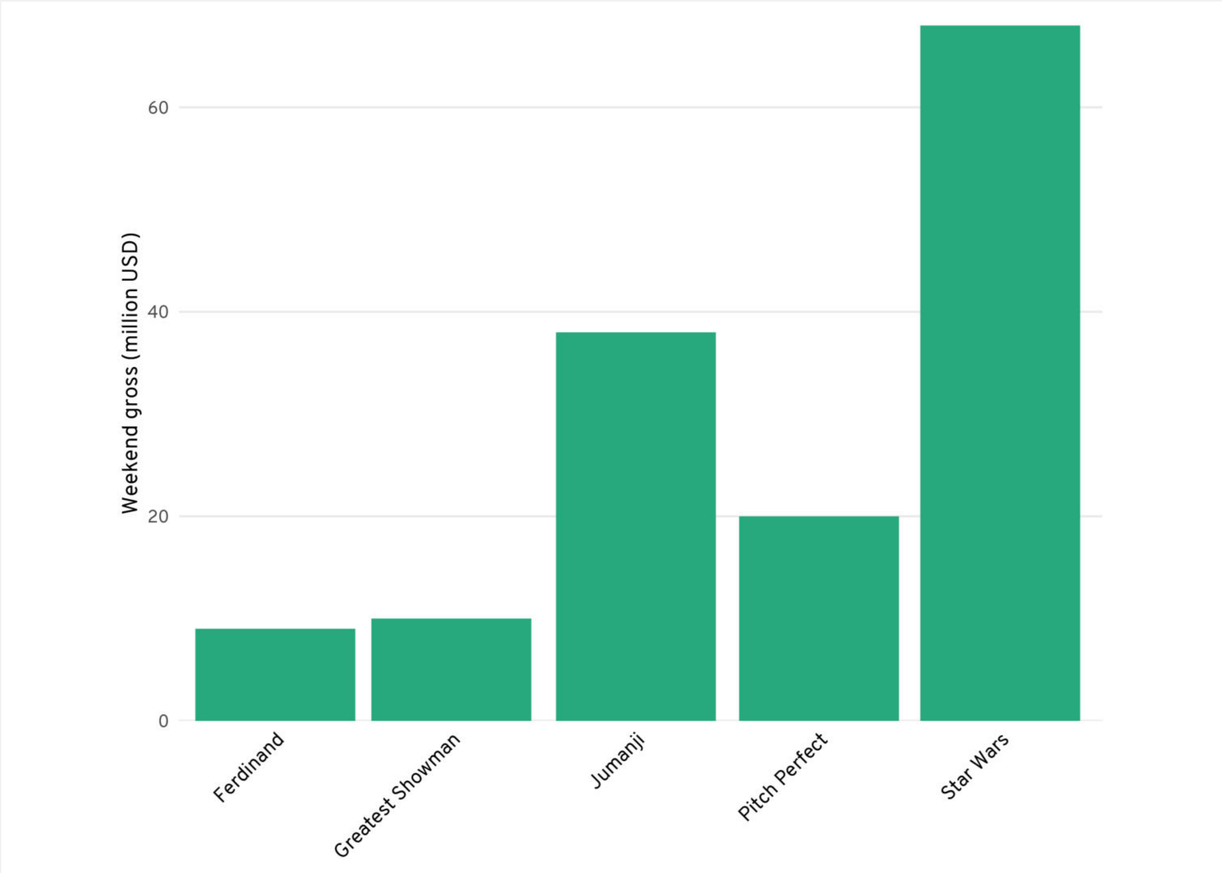


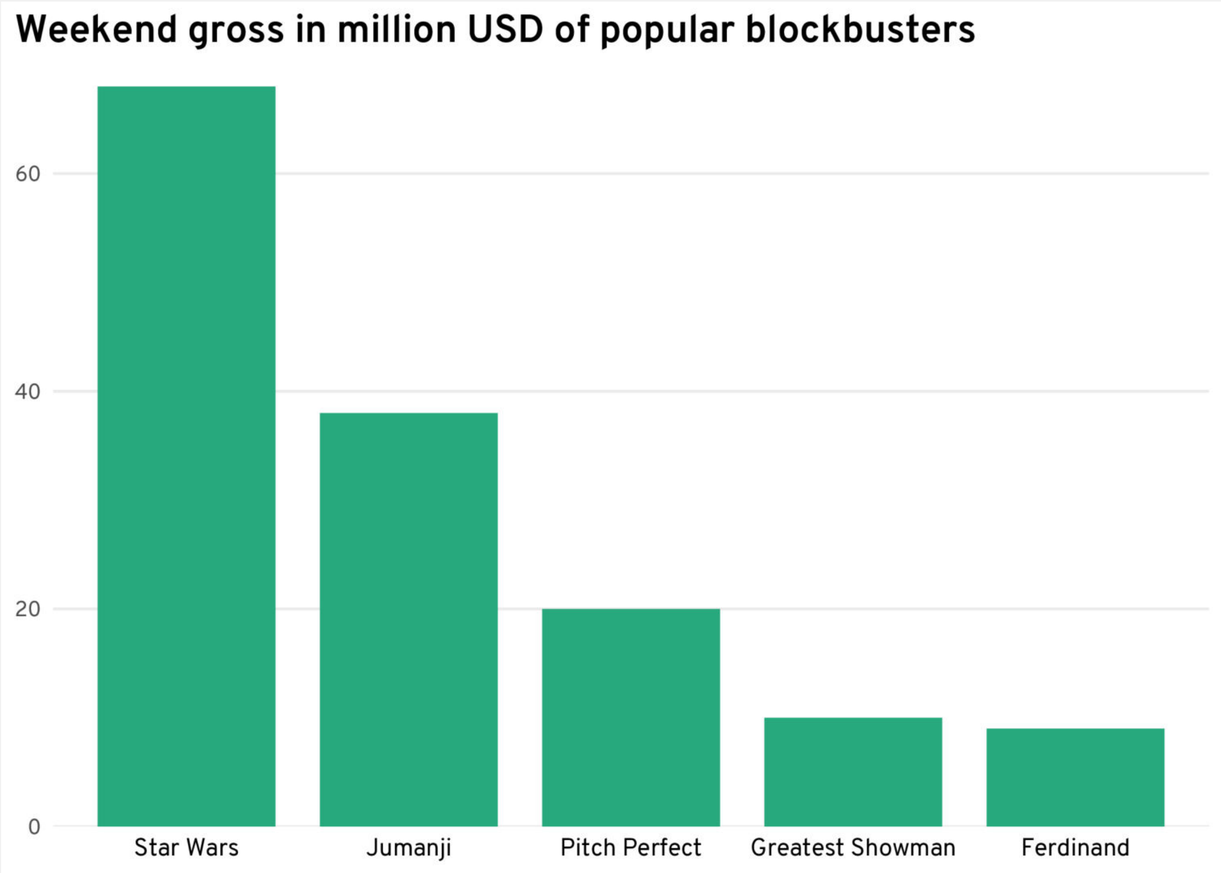
Sort Your Data



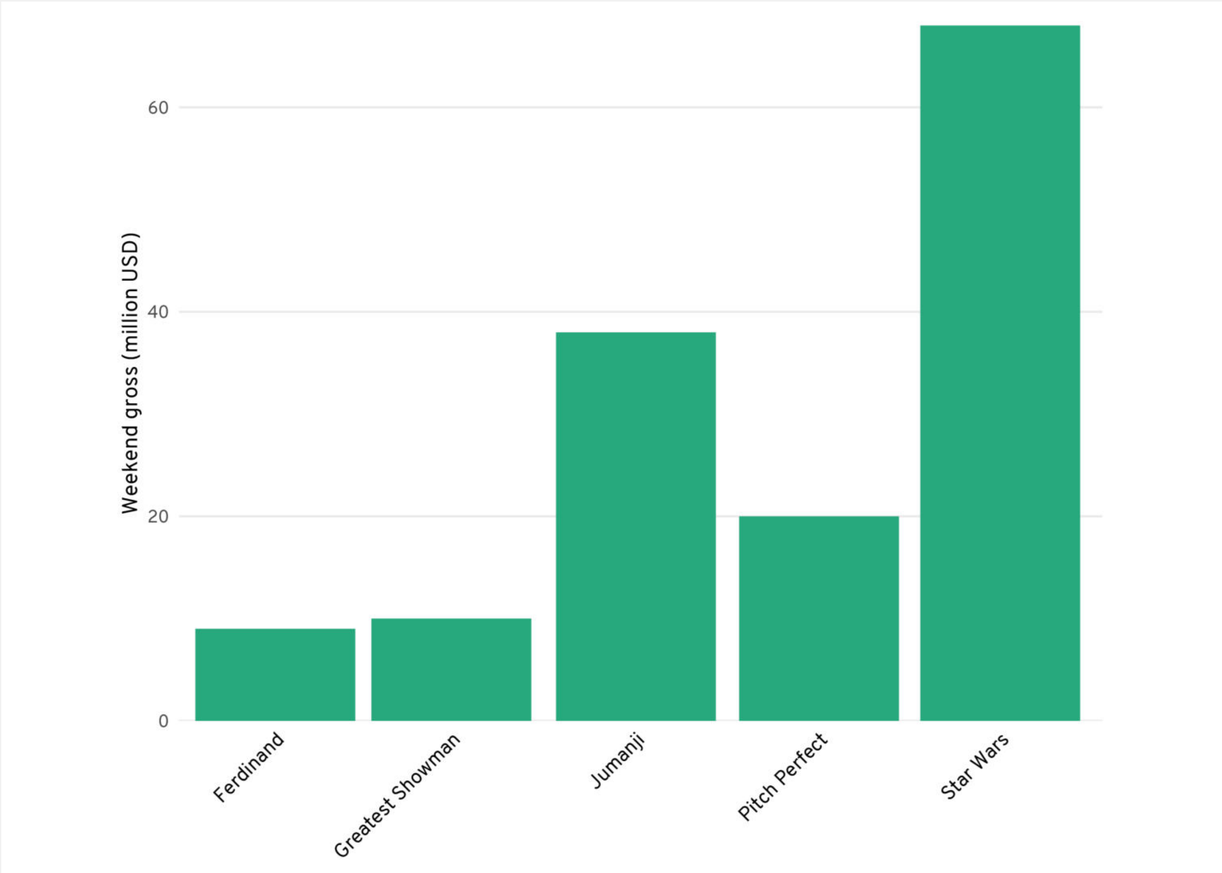


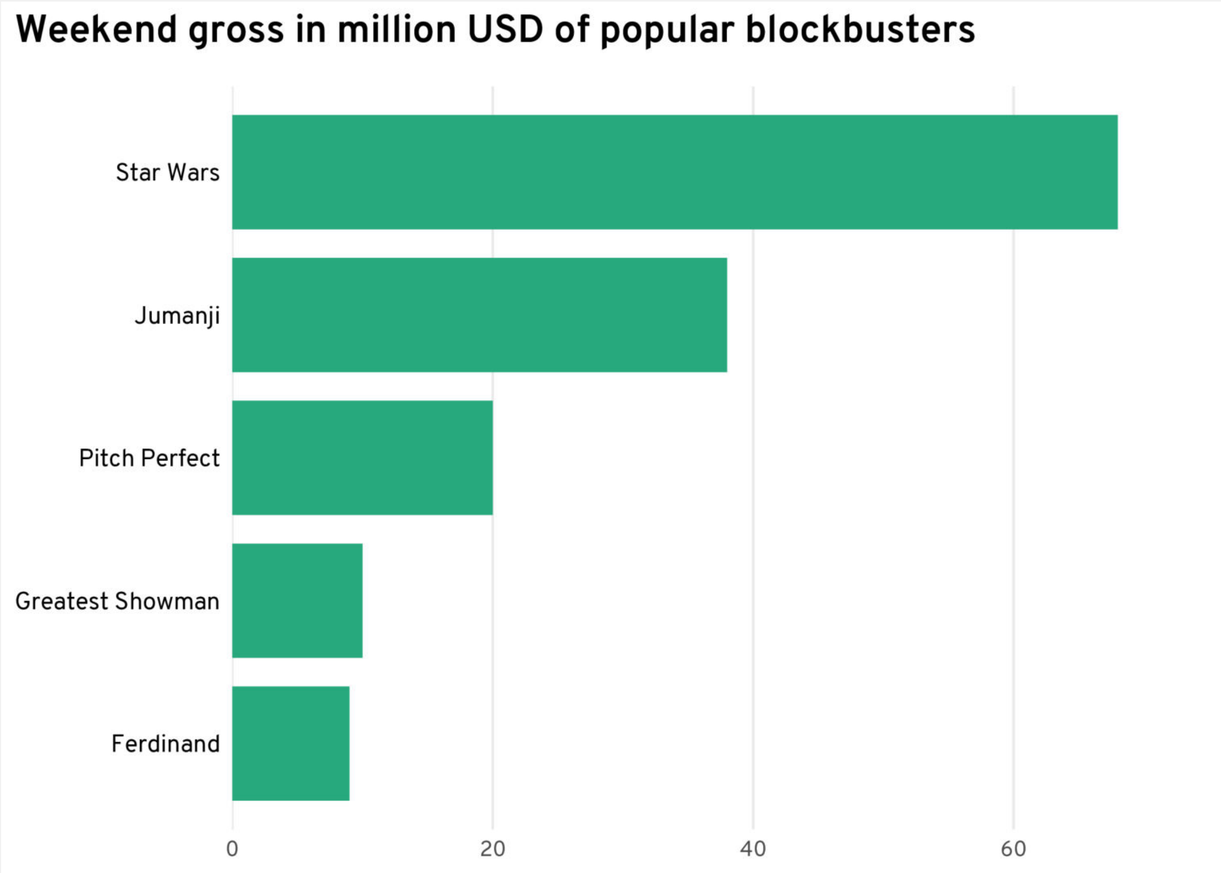
Avoid Rotated Text



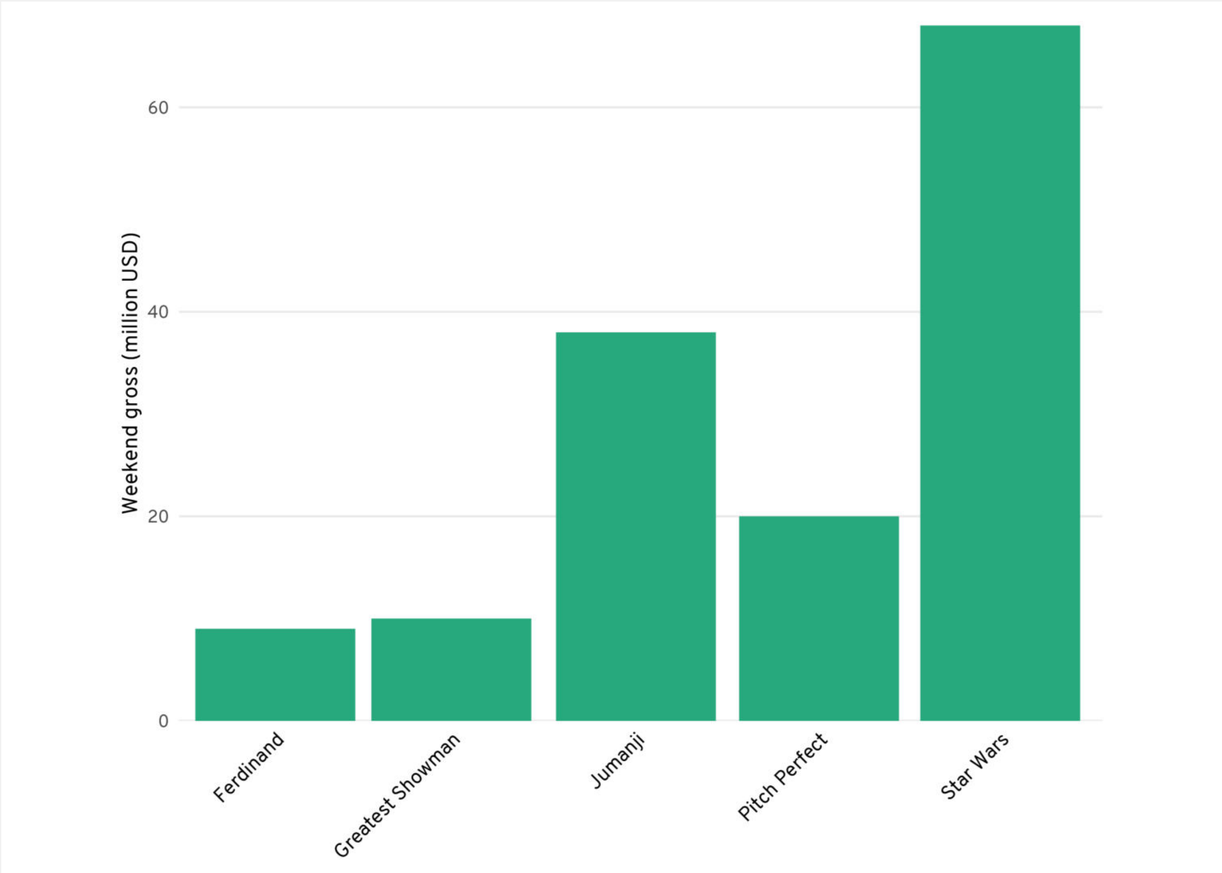


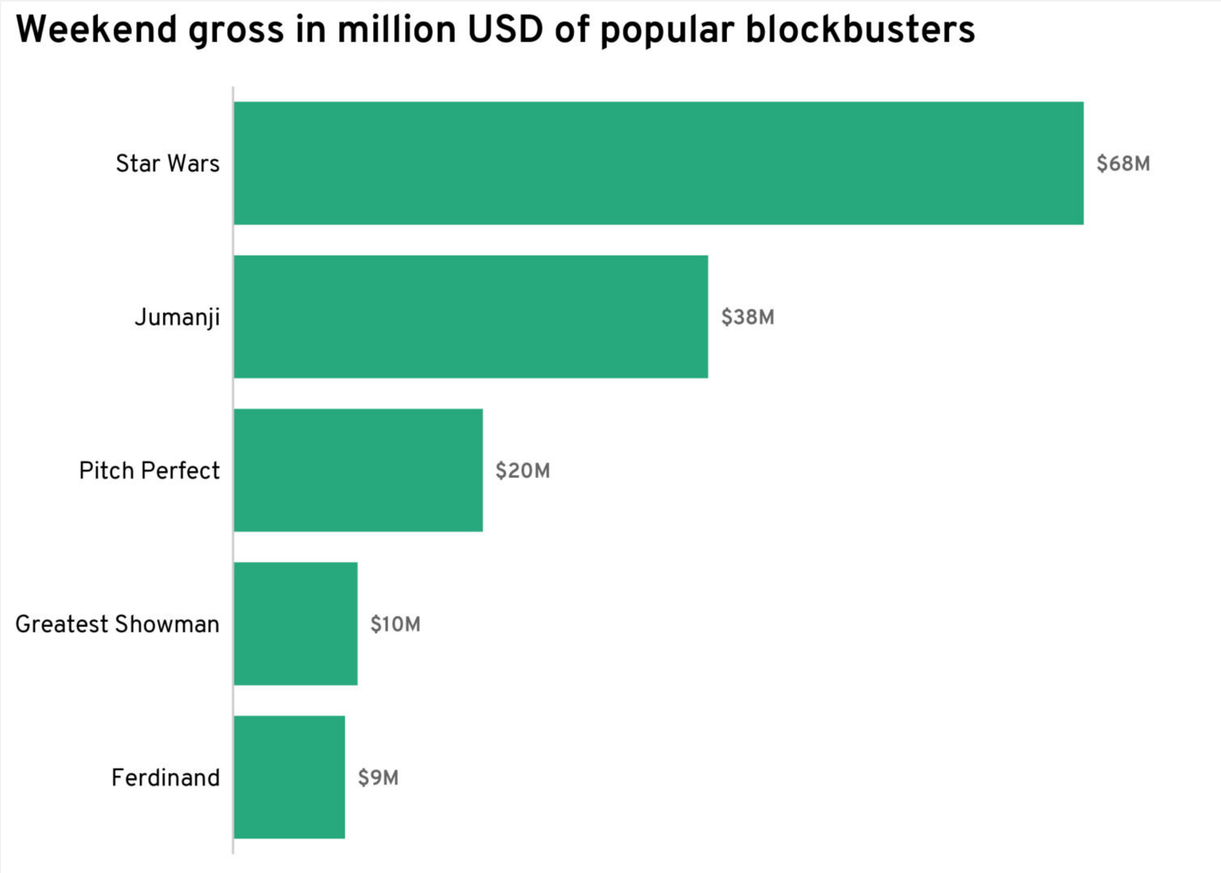
Avoid Rotated Text



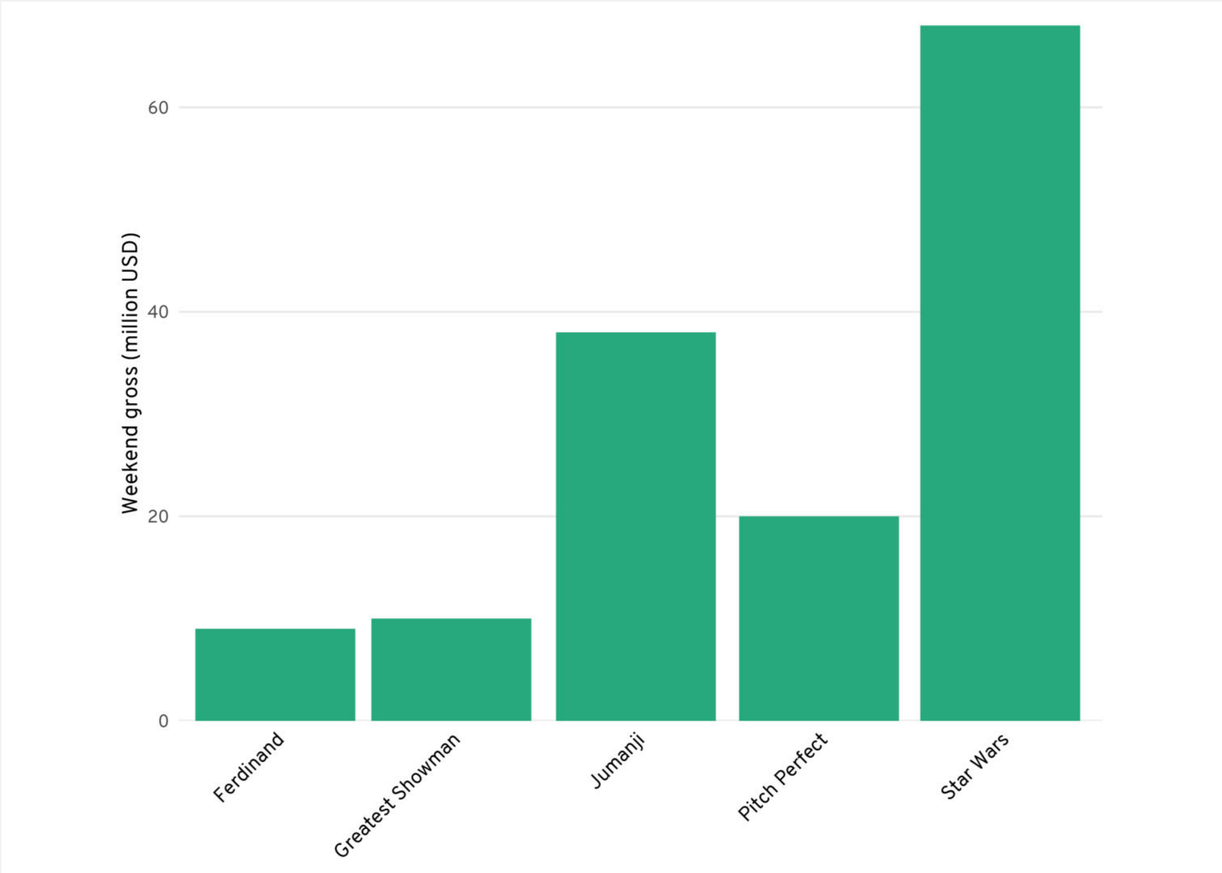


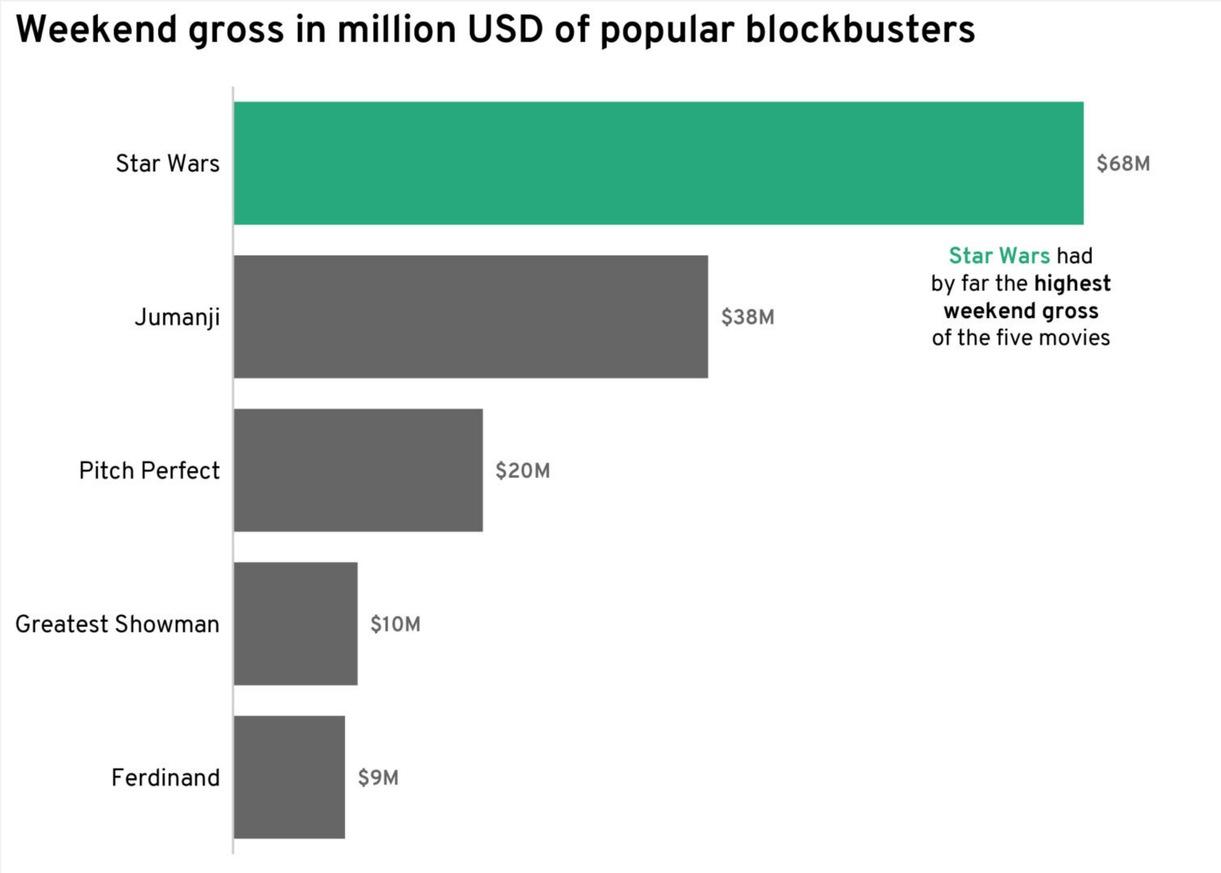
Add Direct Labels



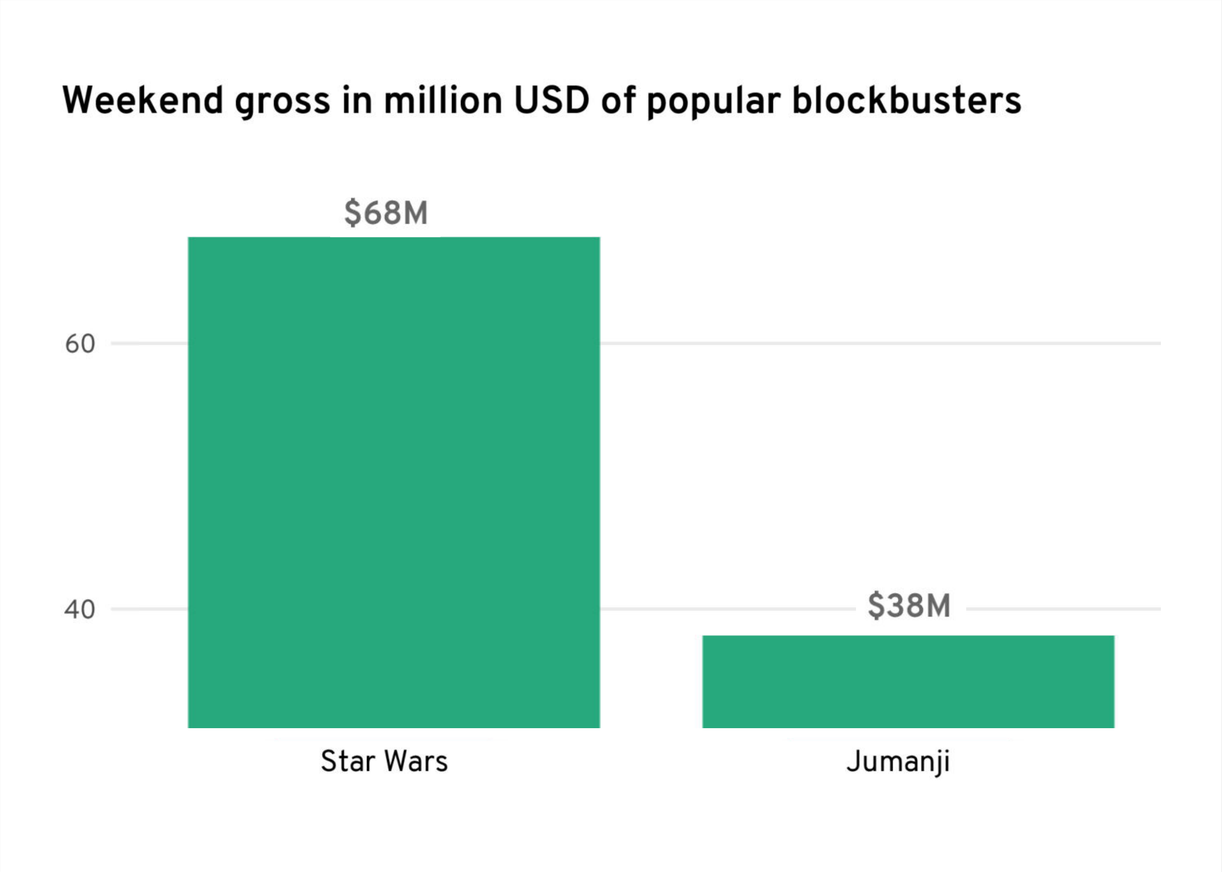


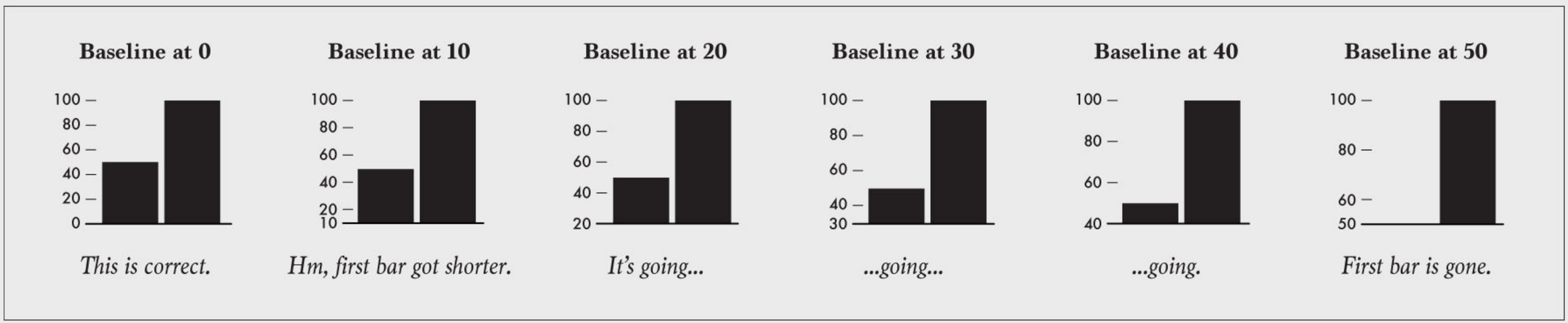
Set Highlights



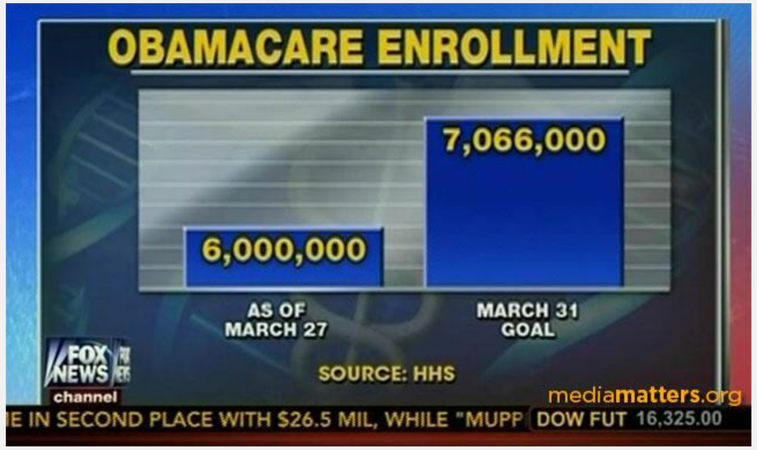


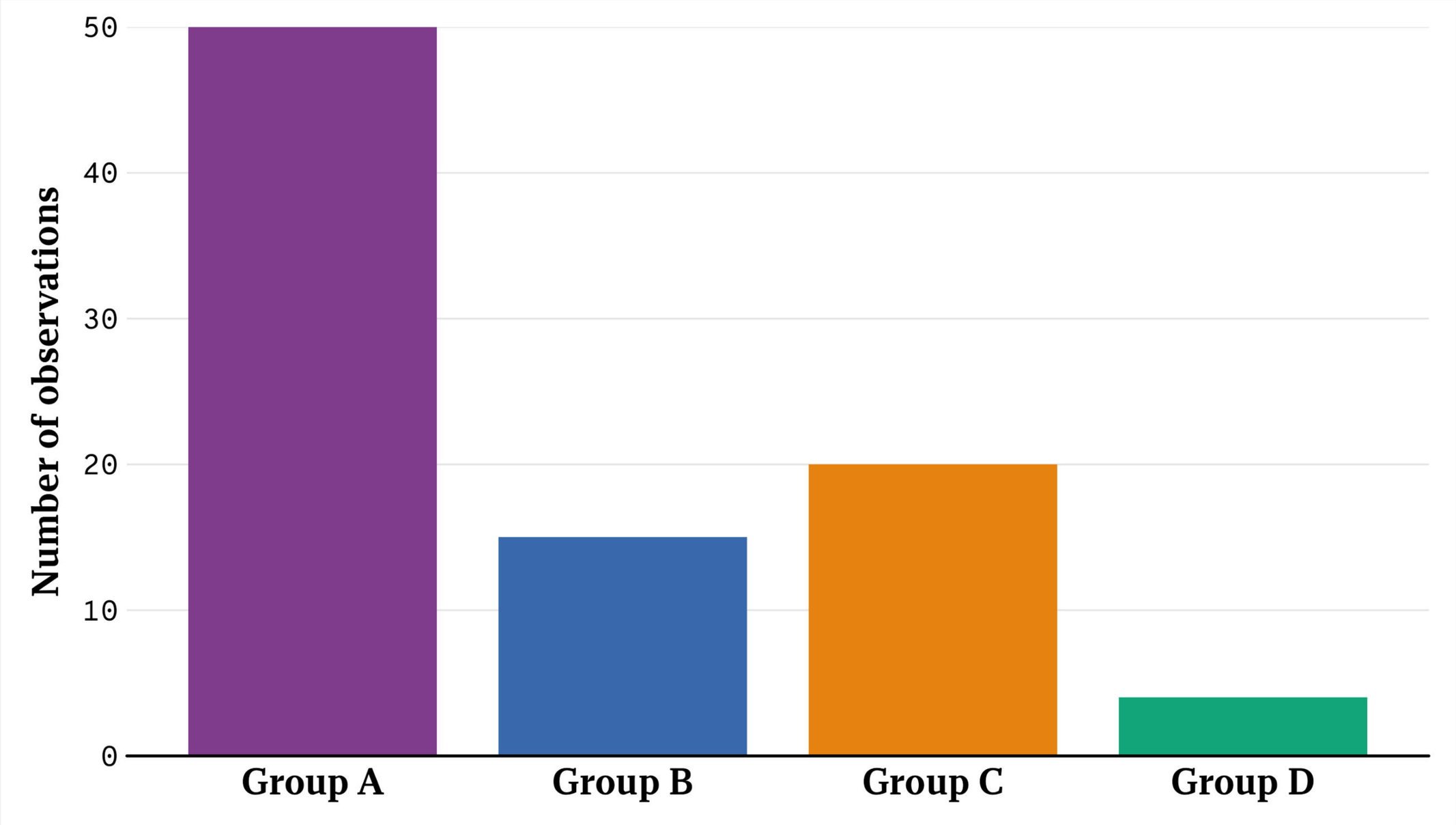
Always Start at Zero



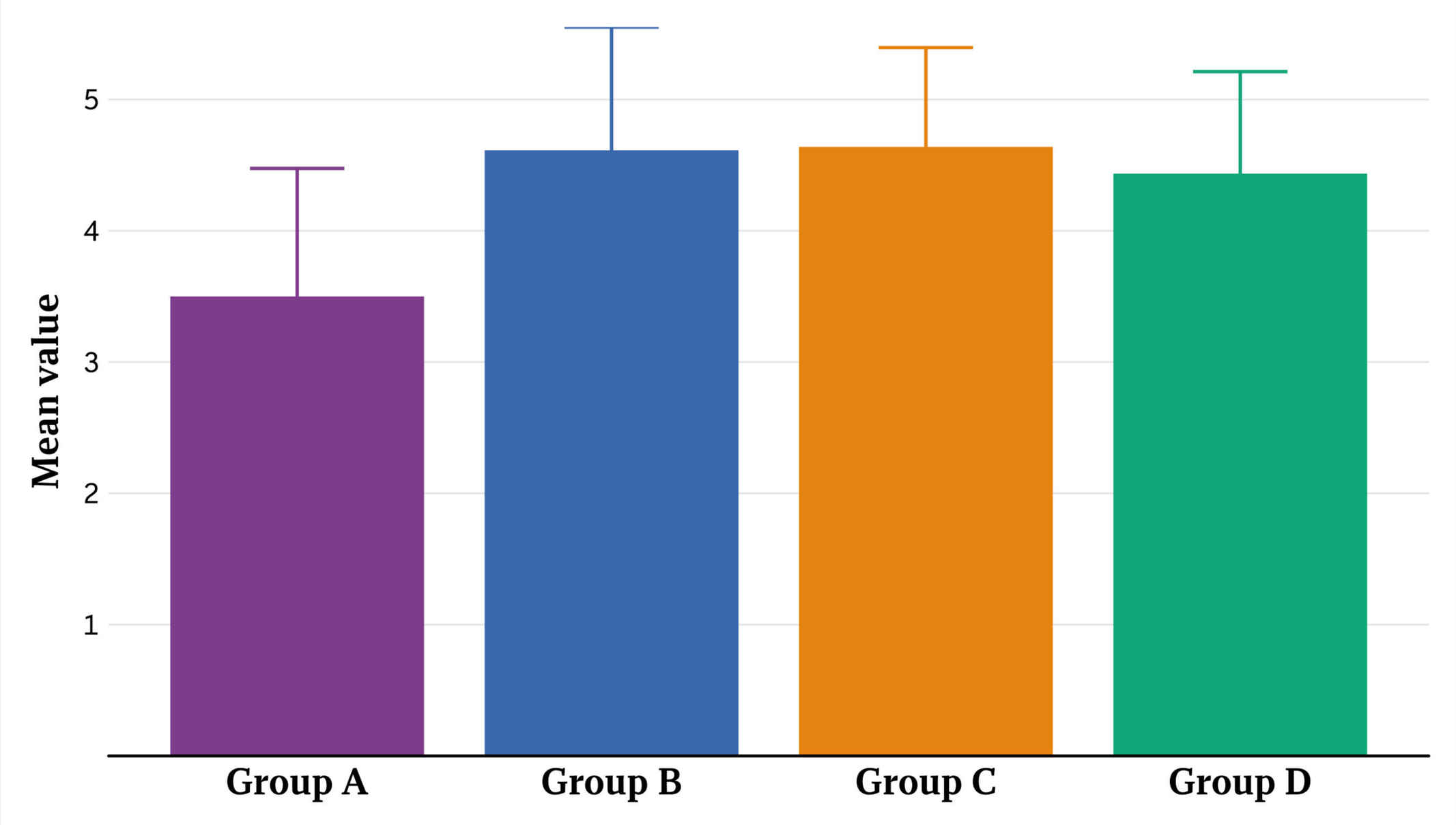


Source: Nathan Yau (flowingdata.com)





*My webinar "**Beyond Bar and Box Plots**" for USGS Data Science [Slides | Recording]*



*My webinar "**Beyond Bar and Box Plots**" for USGS Data Science [Slides | Recording]*

PLOS | BIOLOGY

PERSPECTIVE

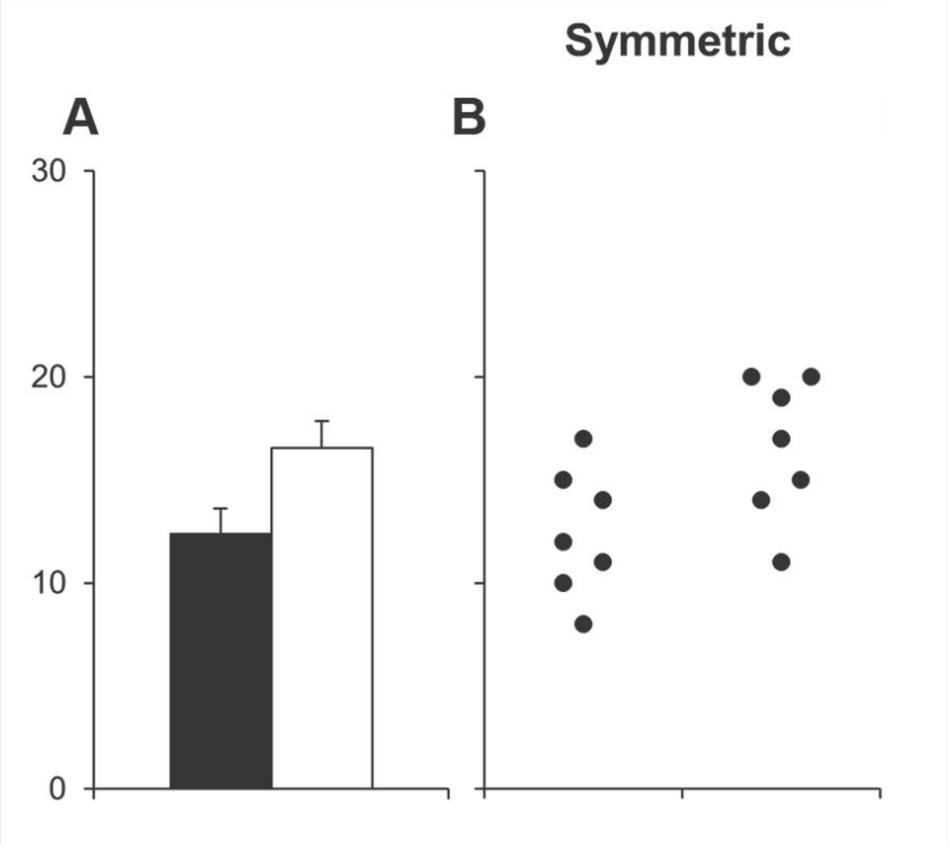
Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm

Tracey L. Weissgerber[1]*, Natasa M. Milic[1,2], Stacey J. Winham[3], Vesna D. Garovic[1]

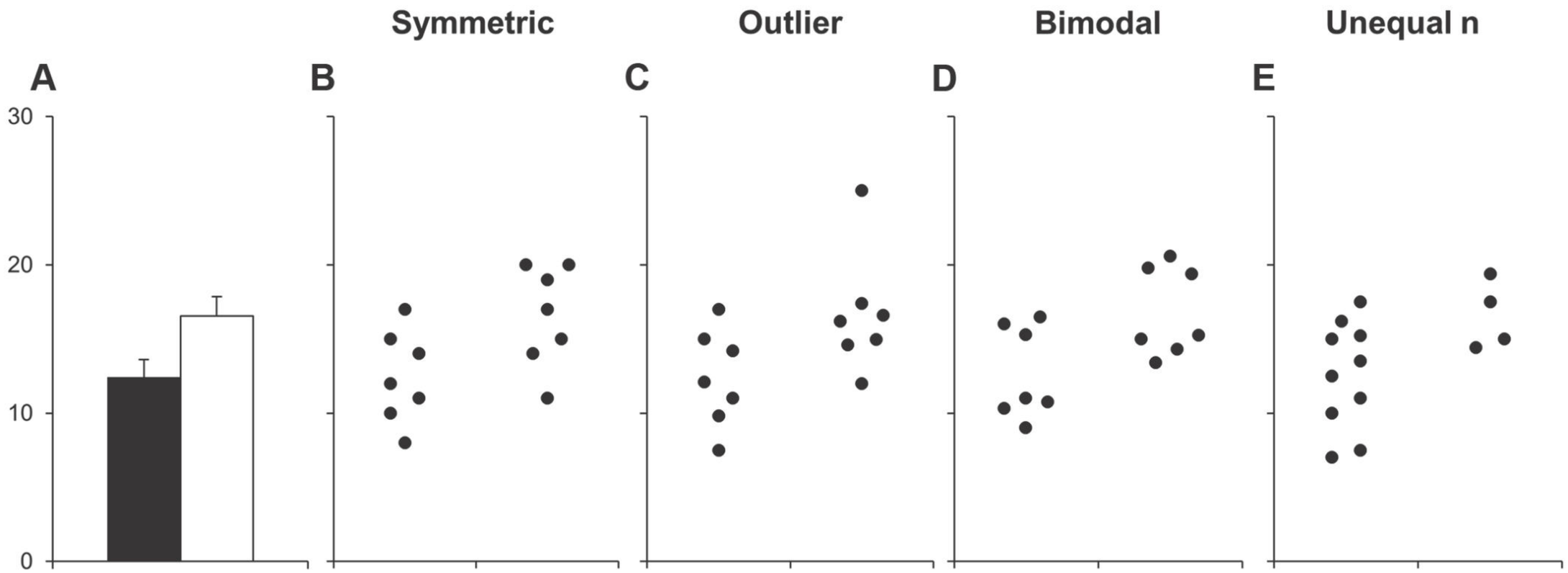
1 Division of Nephrology & Hypertension, Mayo Clinic, Rochester, Minnesota, United States of America, **2** Department of Biostatistics, Medical Faculty, University of Belgrade, Belgrade, Serbia, **3** Division of Biomedical Statistic and Informatics, Mayo Clinic, Rochester, Minnesota, United States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

* weissgerber.tracey@mayo.edu

Weissgerber et al. (2015) PLoS Biology

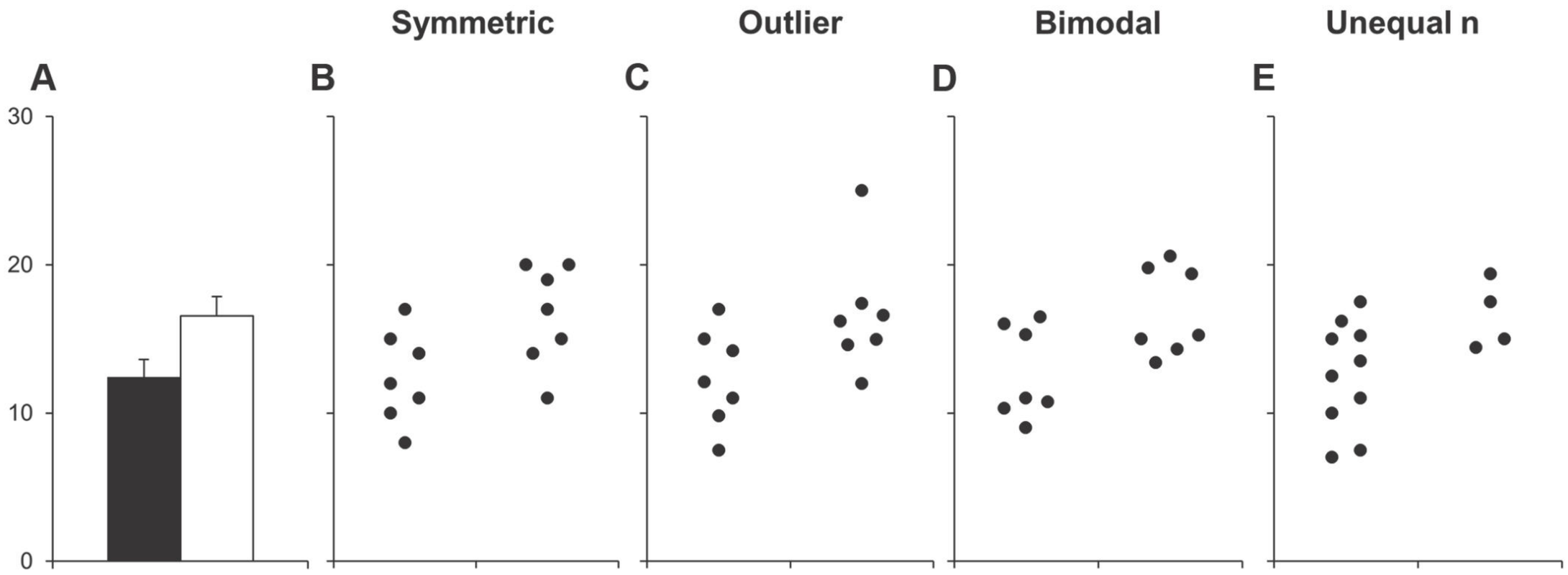


Weissgerber et al. (2015) PLoS Biology



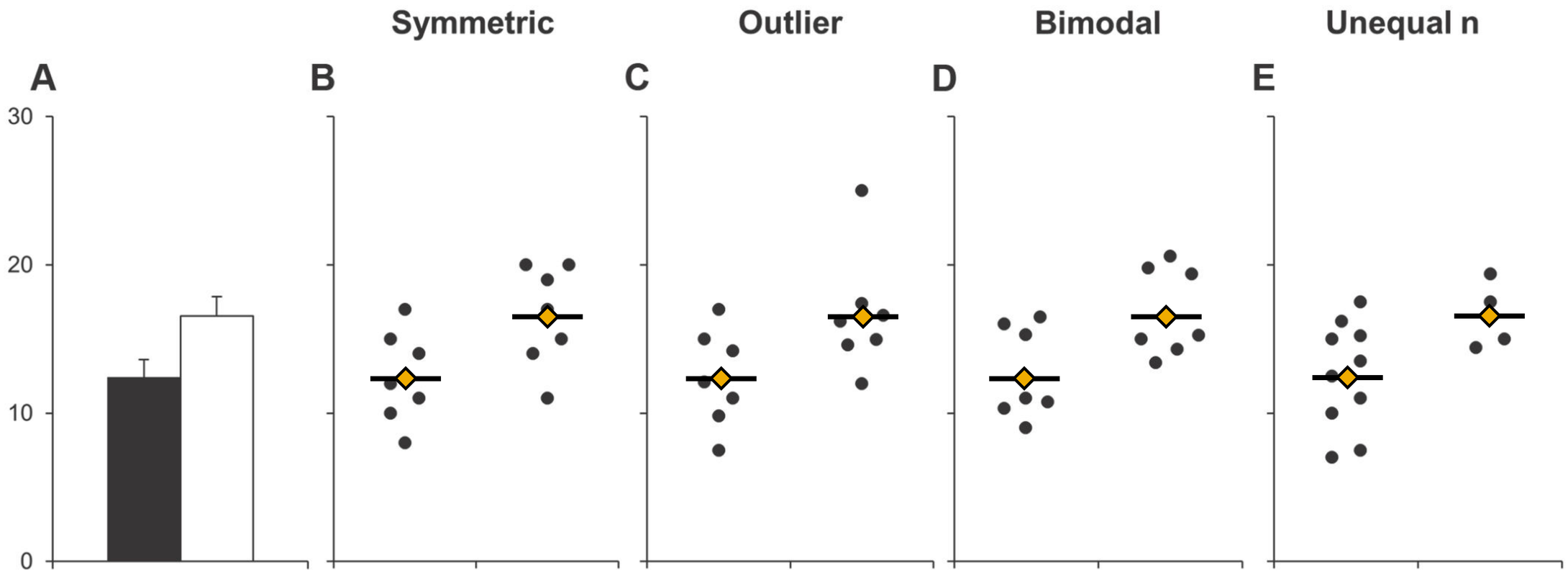
Test	p value			
T-test: Equal var.	0.035	0.050	0.026	0.063
T-test: Unequal var.	0.035	0.050	0.026	0.035
Wilcoxon	0.054	0.073	0.128	0.103

Modified from Weissgerber et al. (2015) PLoS Biology



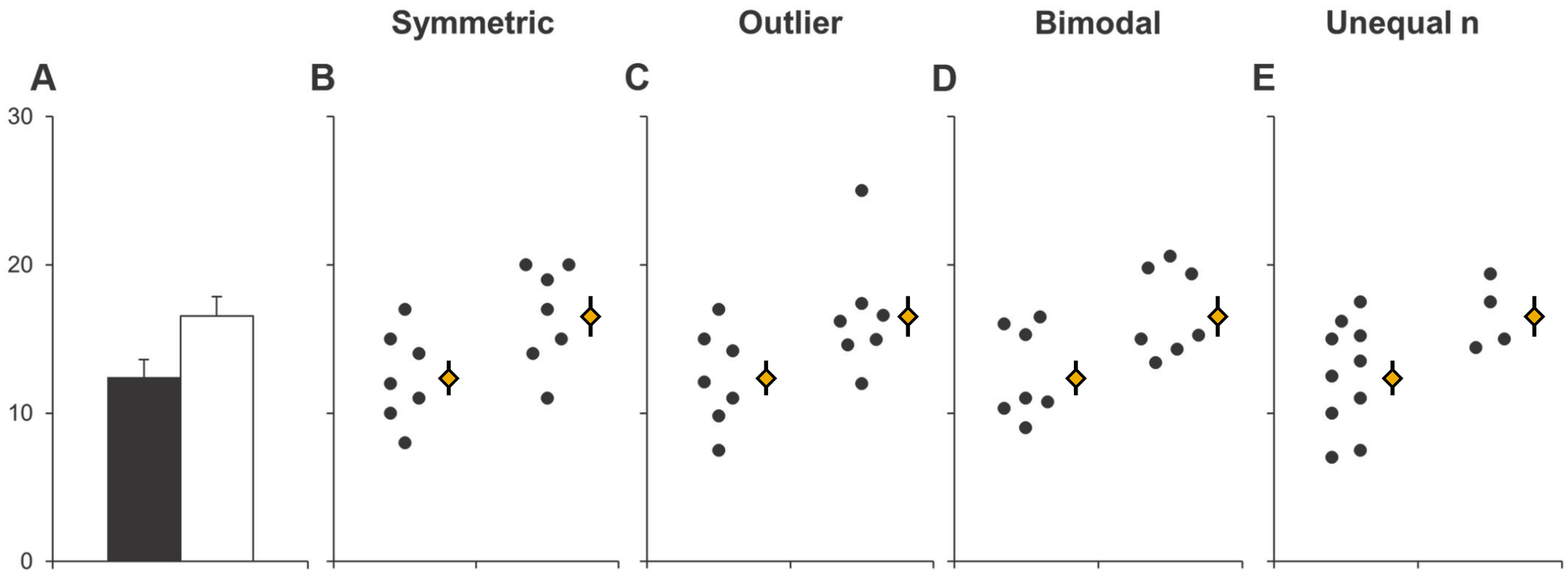
Test				
	p value			
T-test: Equal var.	0.035	0.050	0.026	0.063
T-test: Unequal var.	0.035	0.050	0.026	0.035
Wilcoxon	0.054	0.073	0.128	0.103

Modified from Weissgerber et al. (2015) PLoS Biology



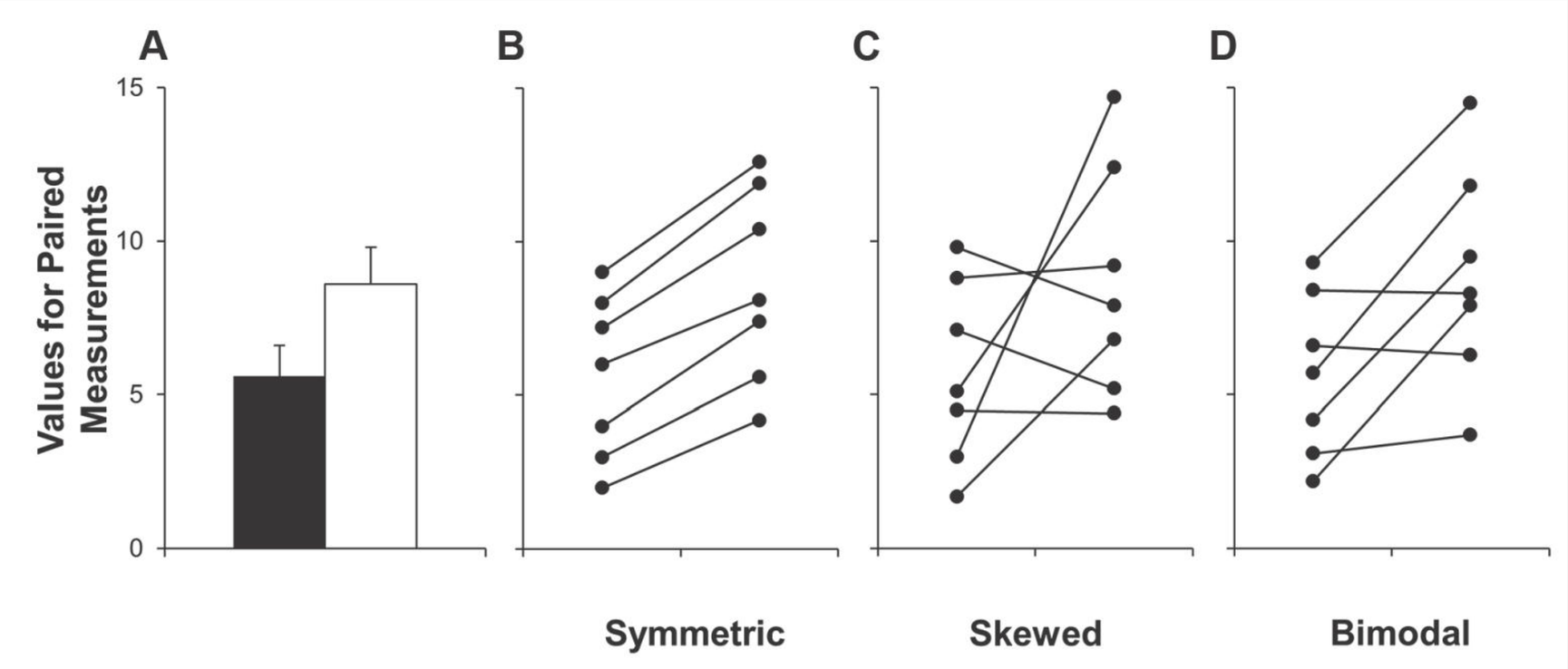
Test		p value		
T-test: Equal var.	0.035	0.050	0.026	0.063
T-test: Unequal var.	0.035	0.050	0.026	0.035
Wilcoxon	0.054	0.073	0.128	0.103

Modified from Weissgerber et al. (2015) PLoS Biology

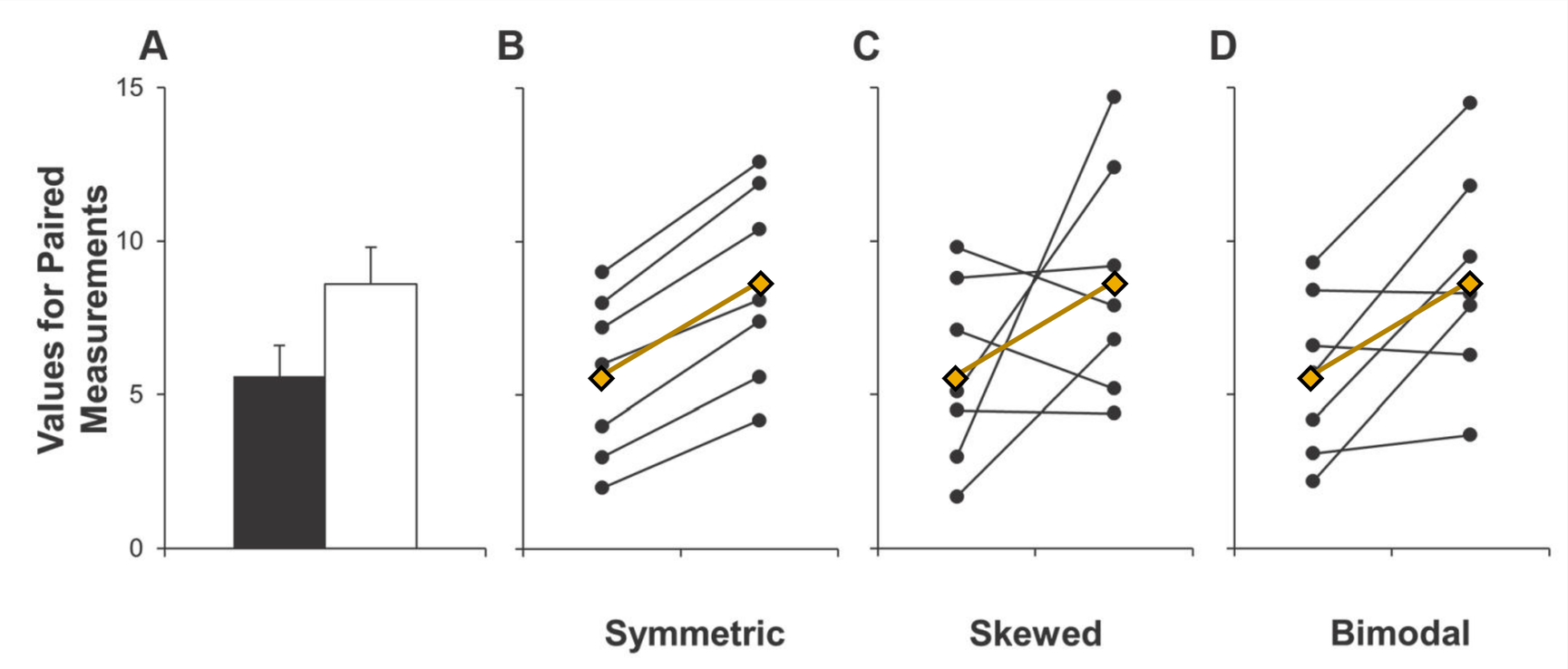


Test				
	p value			
T-test: Equal var.	0.035	0.050	0.026	0.063
T-test: Unequal var.	0.035	0.050	0.026	0.035
Wilcoxon	0.054	0.073	0.128	0.103

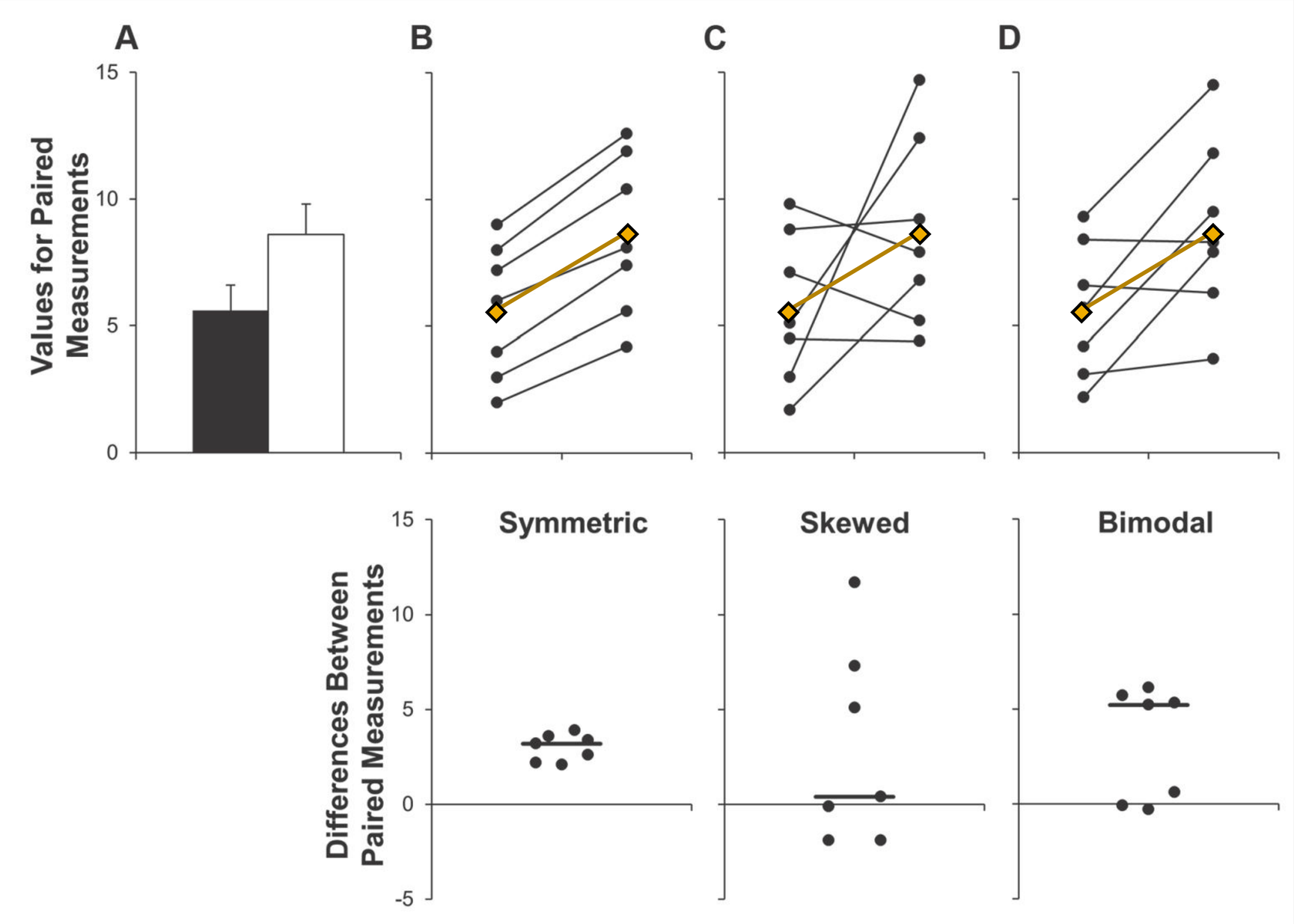
Modified from Weissgerber et al. (2015) PLoS Biology



Weissgerber et al. (2015) PLoS Biology



Modified from Weissgerber et al. (2015) PLoS Biology

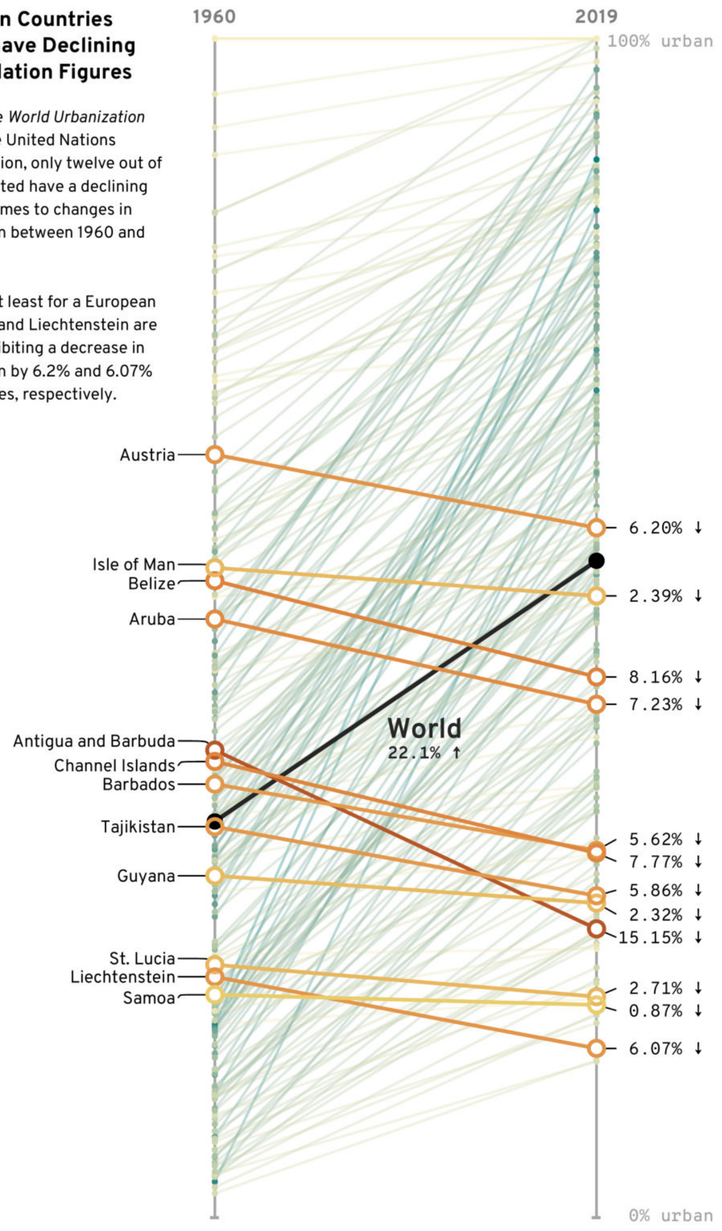


Modified from Weissgerber et al. (2015) PLoS Biology

Only a Dozen Countries Worldwide have Declining Urban Population Figures

According to the *World Urbanization Prospects* by the United Nations Population Division, only twelve out of 218 countries listed have a declining trend when it comes to changes in urban population between 1960 and 2019.

Surprinsingly—at least for a European citizen—Austria and Liechtenstein are amog these exhibiting a decrease in urban population by 6.2% and 6.07% within six decades, respectively.

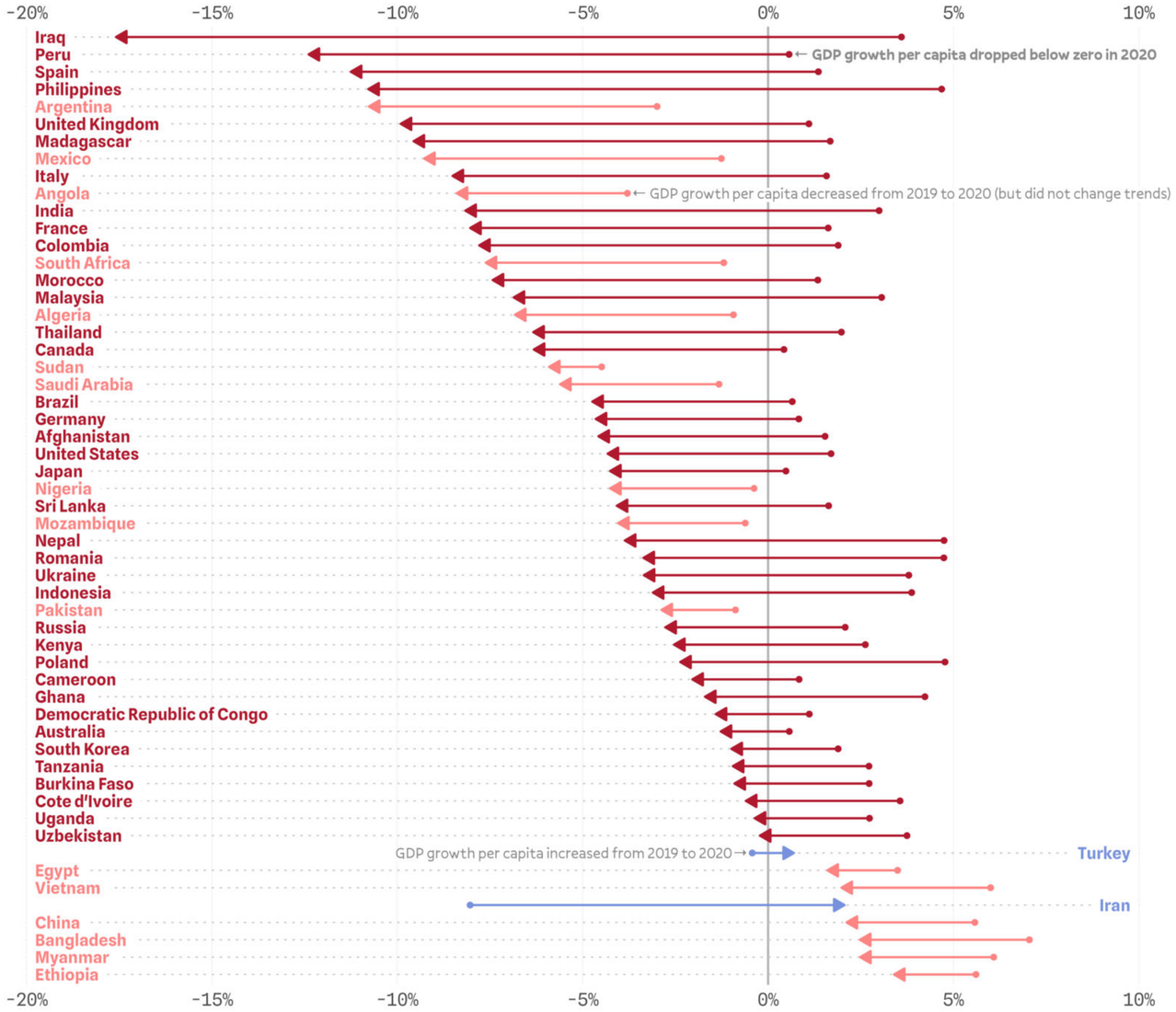


Visualization by Cédric Scherer | Data: United Nations Population Division. World Urbanization Prospects: 2018 Revision. | #30DayChartChallenge 2021 Day 5: Slope

Average income has decreased in 185 countries from 2019 to 2020, with GDP per capita growth turning negative in 72% of these cases.

Graphic: Cédric Scherer • Data: World Bank and OECD via OurWorldInData

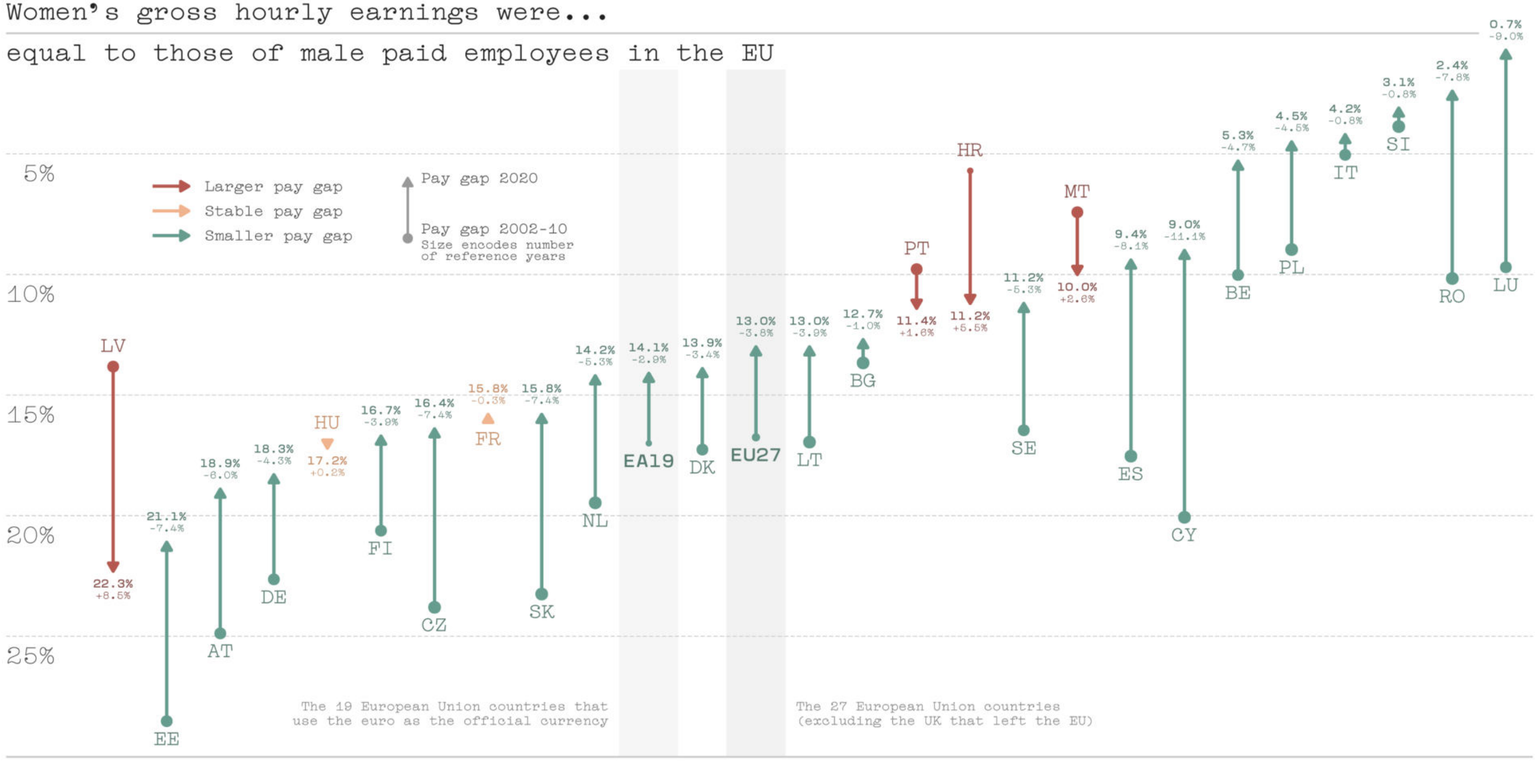
Average income has decreased in almost all countries from 2019 to 2020, with GDP per capita growth turning from positive to negative in most cases.



Graphic: Cédric Scherer • Data: World Bank and OECD via OurWorldInData • The graphic contains countries that inhabit more than 20M people.

Women's gross hourly earnings were...

equal to those of male paid employees in the EU



30% below those of male paid employees in the EU

In 2020, women's gross hourly earnings were on average 13% below those of men in the EU.

Many countries were able to decrease the pay gap between male and female employees: most notably Luxembourg with the smallest gap in 2020 of 0.7% and Romania with a decrease of 13.6% within 18 years to 2.4%, ranking second. However, in several countries the pay gap was almost stable (Hungary and France) or even increased (Latvia, Portugal, Croatia, and Malta) compared to the gap in the reference period from 2002 to 2010 for all years available via Eurostat.

As an unadjusted indicator, the gender pay gap gives an overall picture of the differences between men and women in terms of earnings and measures – a concept which is broader than discrimination in the sense of: *equal pay for work of equal value.*

Graphic: Cédric Scherer • Data: Eurostat (SDG_05_20; no data for Greece and Ireland) • #InternationalWomensDay2022

"The Pay Gap in Europe", personal project for International Women's Day 2022

cedricscherer.com @CedScherer z3tt

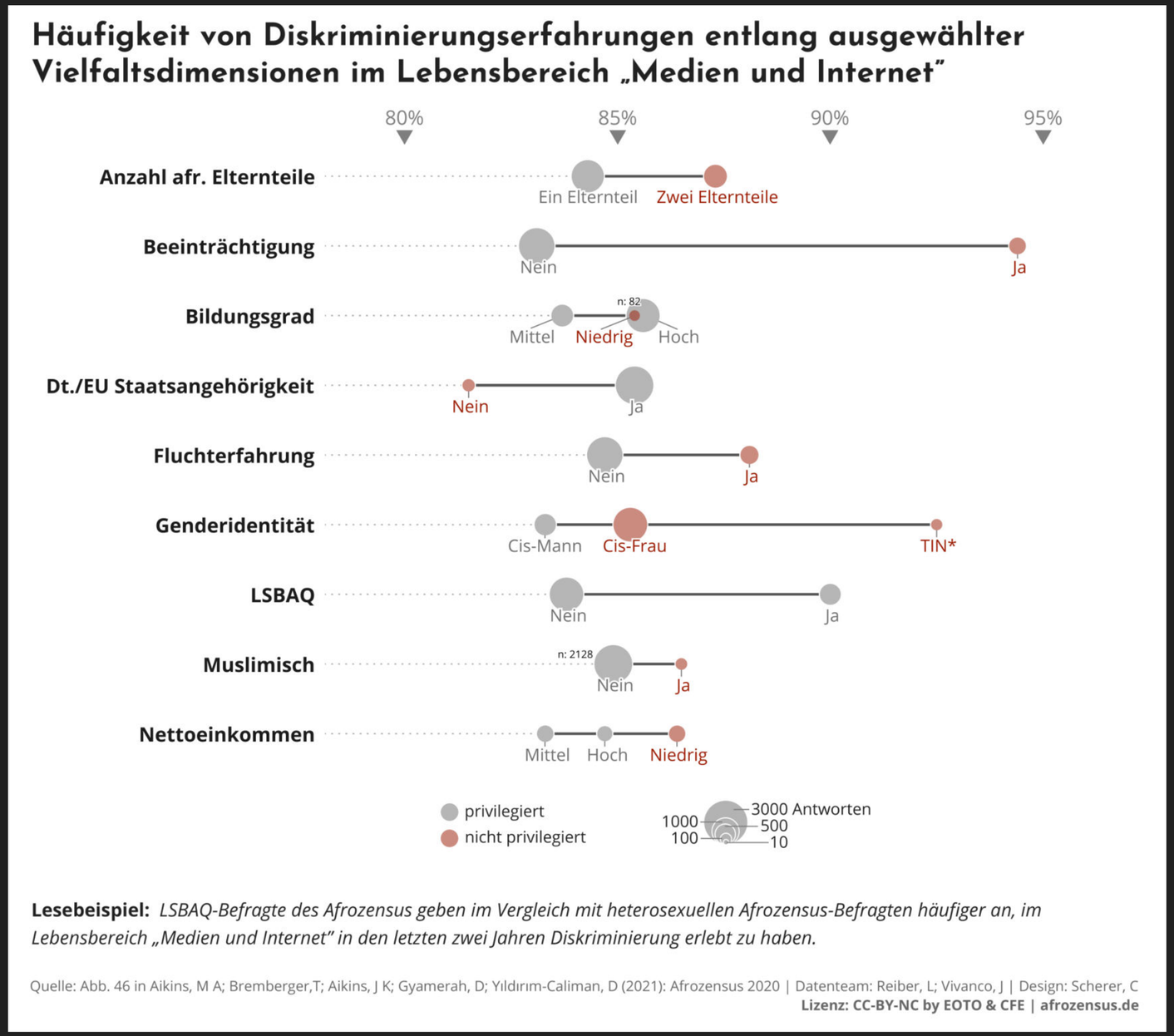
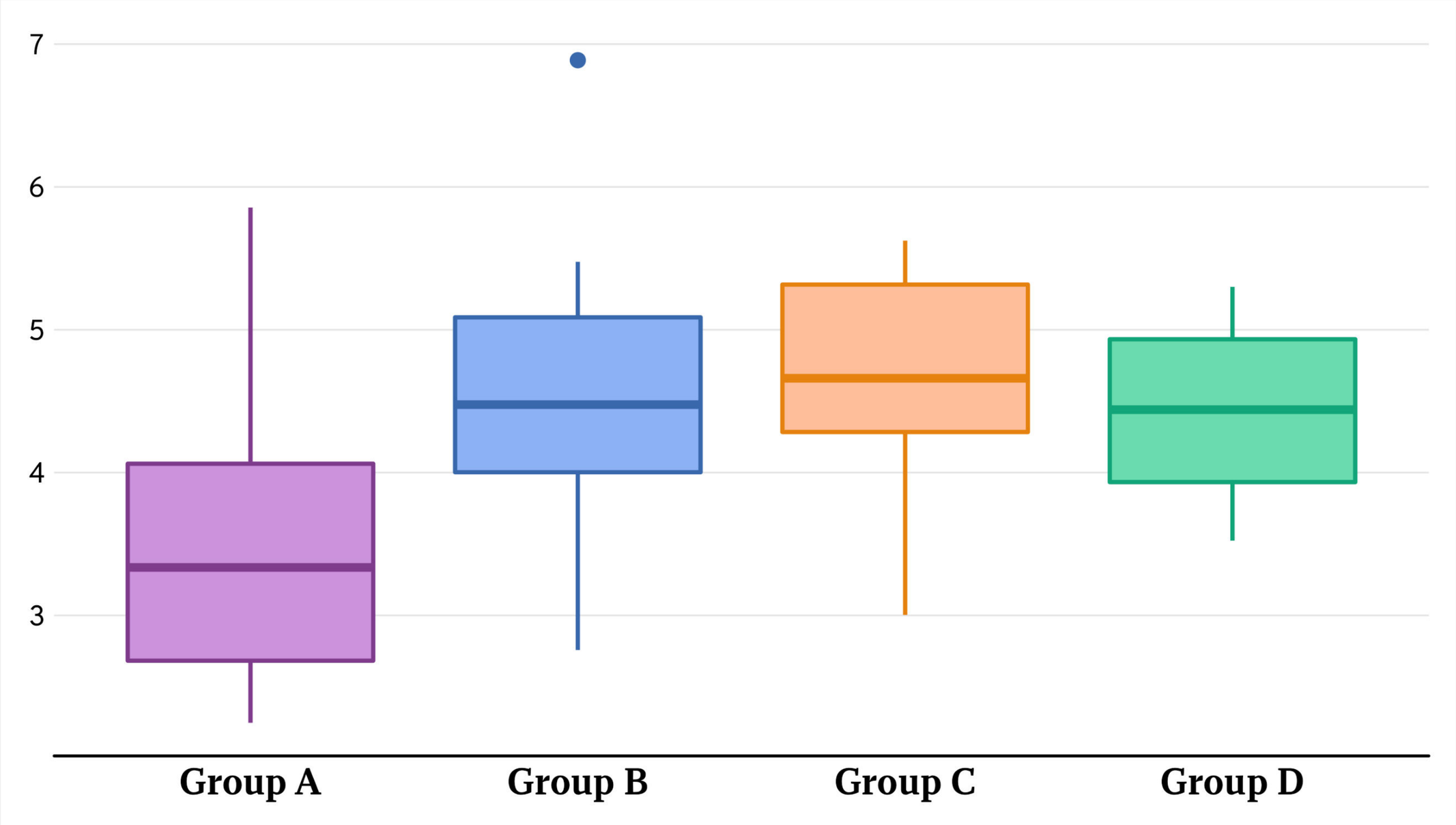
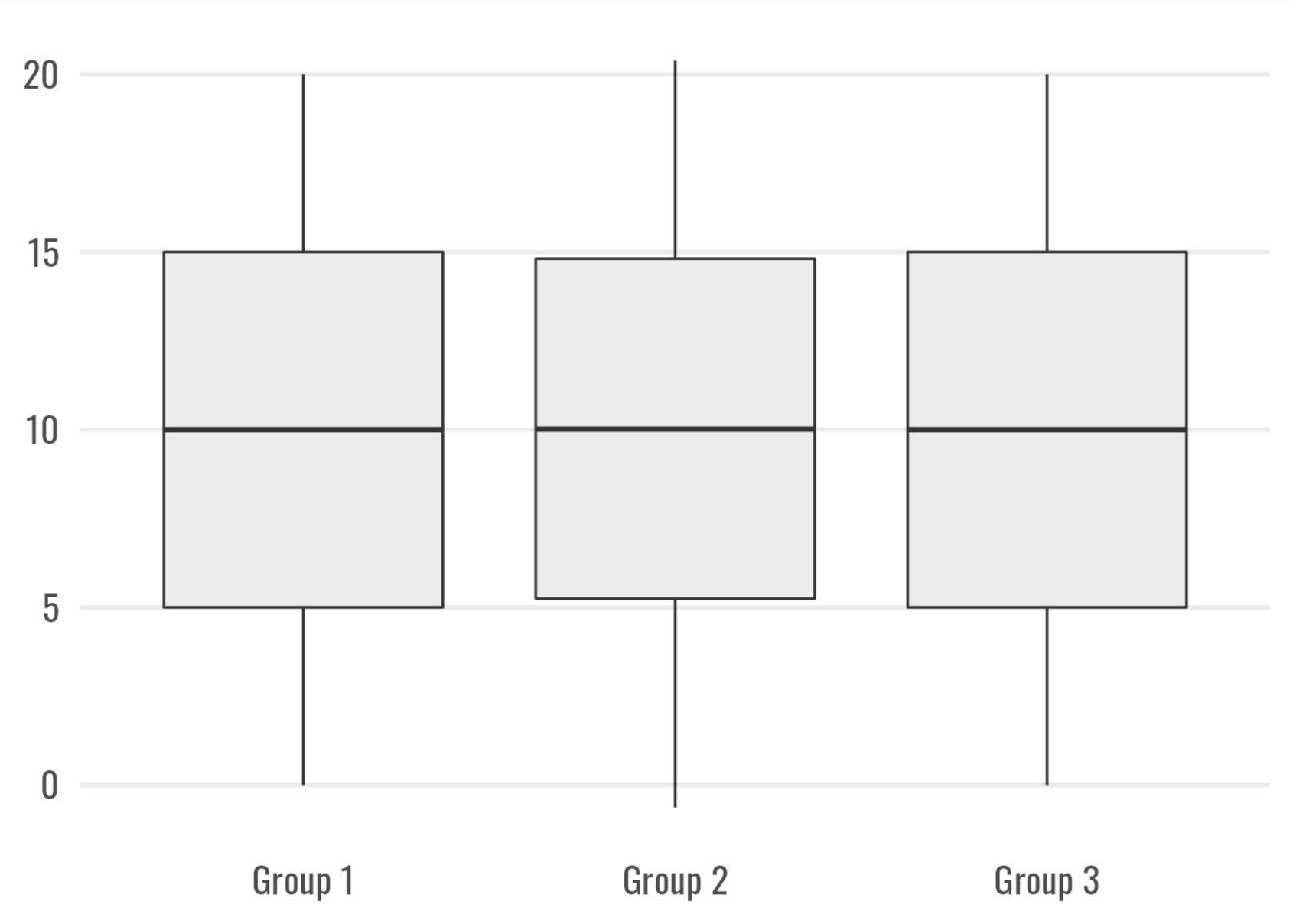


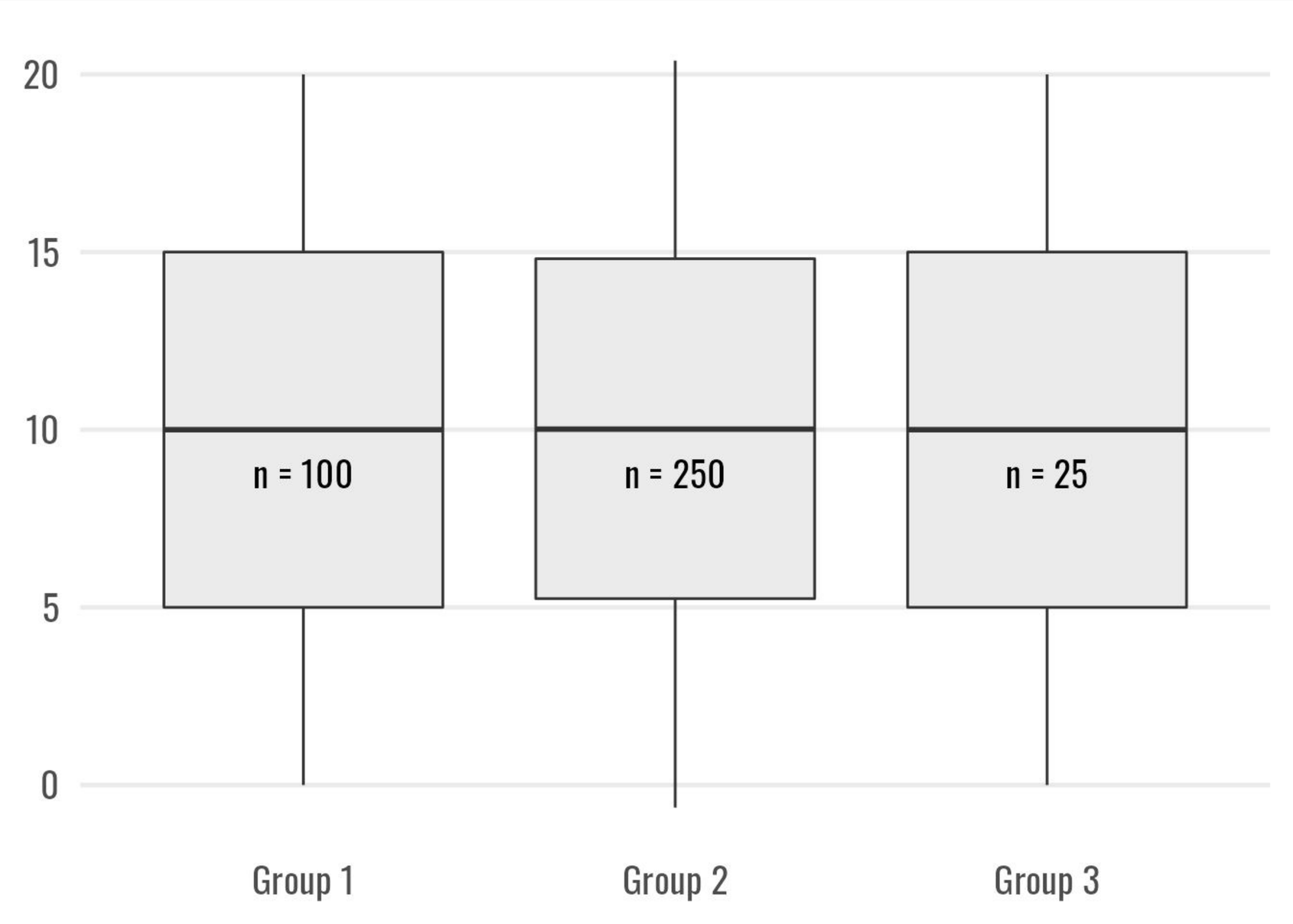
Fig. 46 "Afrozensus 2020" for Citizens For Europe & EOTO e.V.



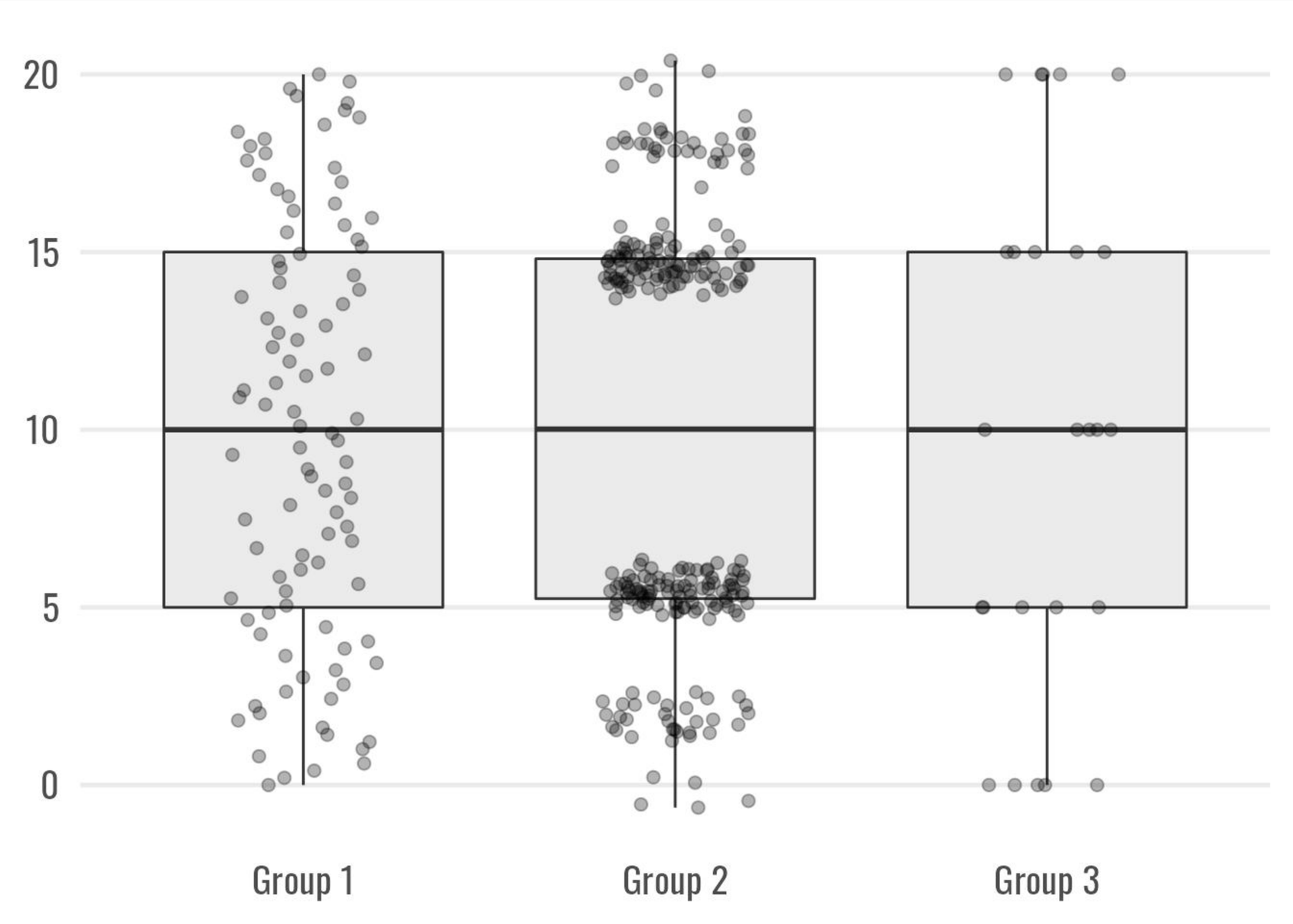
*My webinar "**Beyond Bar and Box Plots**" for USGS Data Science [Slides | Recording]*



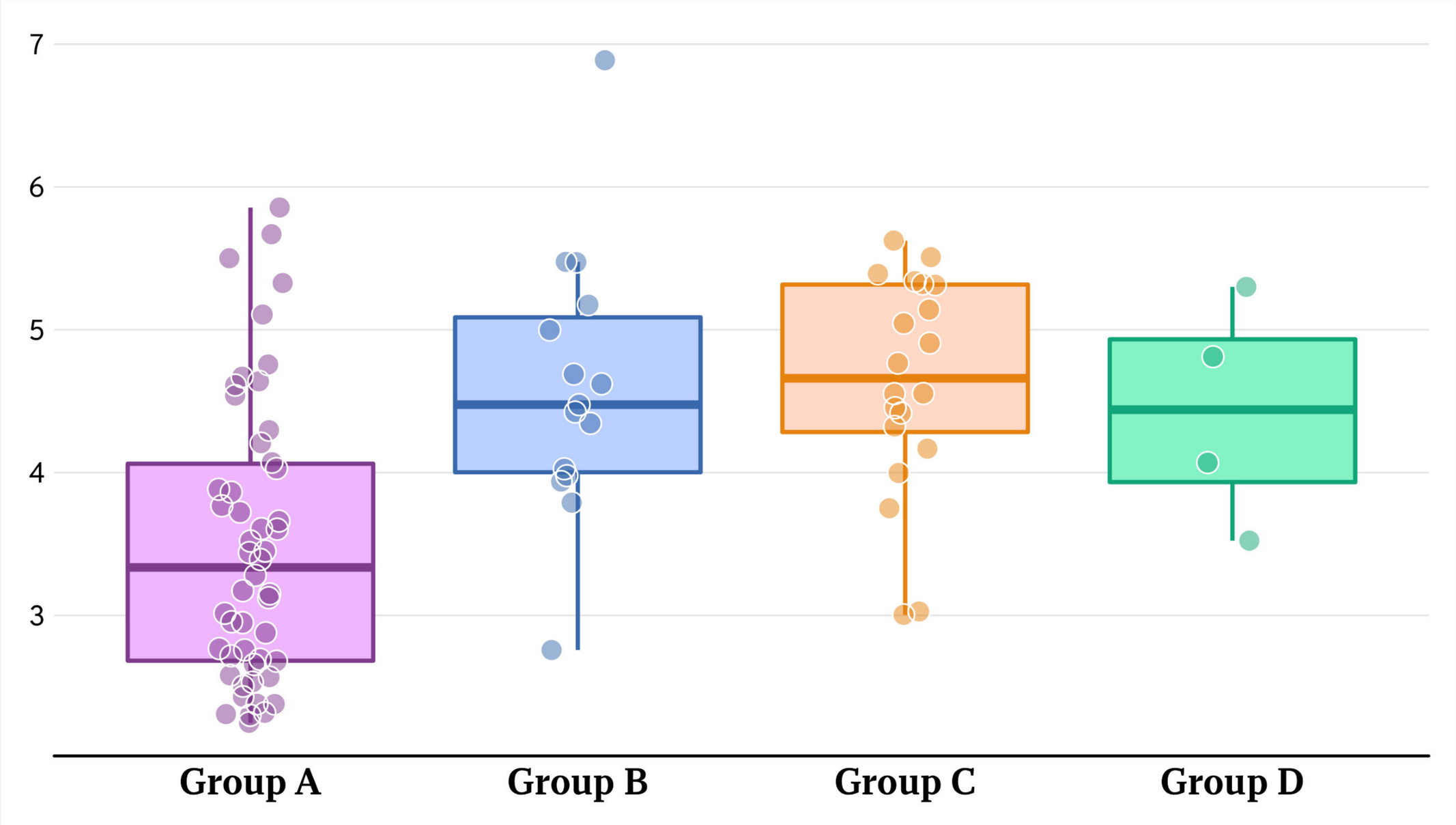
From my blog post "Visualizing Distributions with Raincloud Plots (and How to Create Them with ggplot2)"



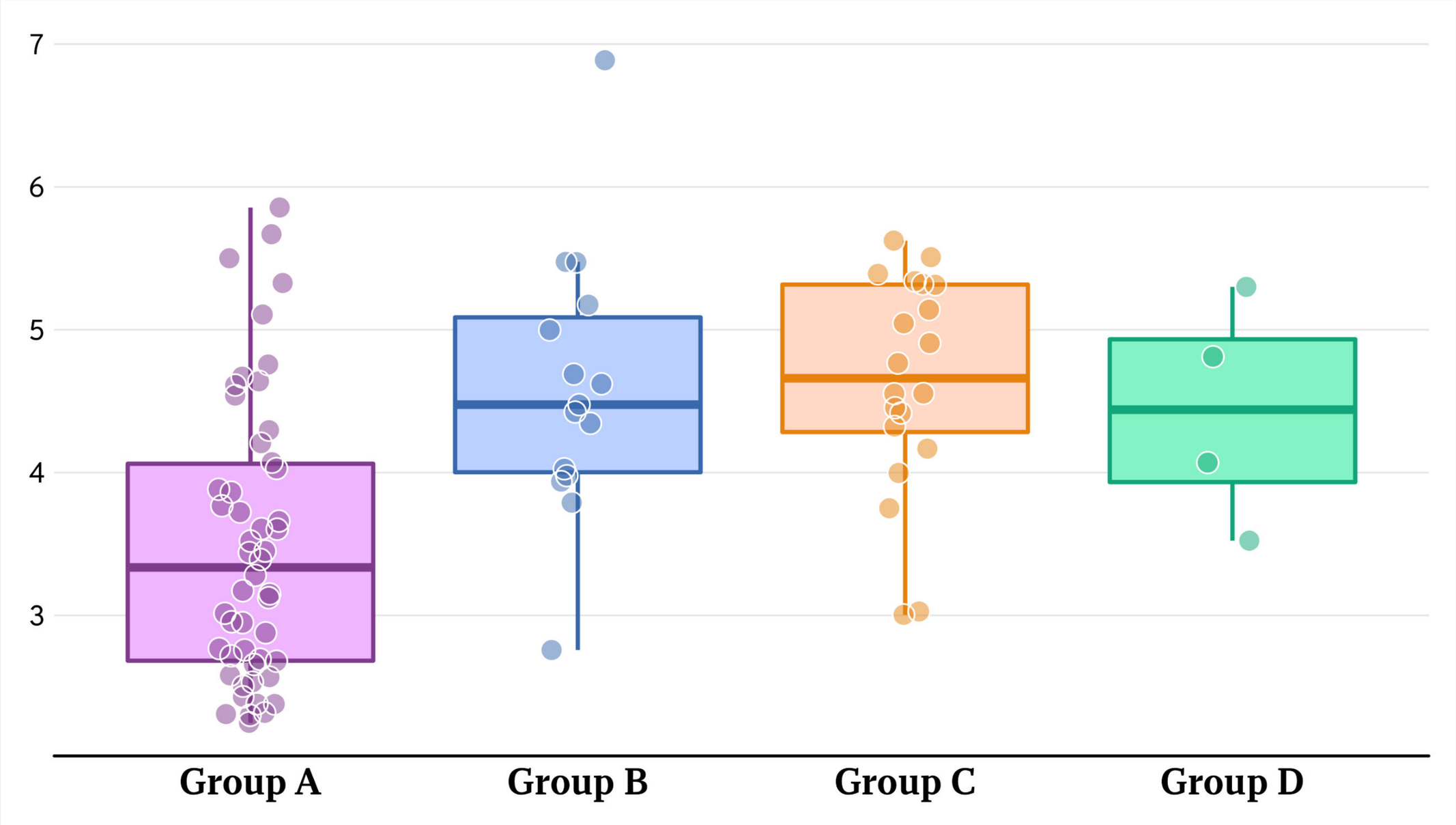
From my blog post "Visualizing Distributions with Raincloud Plots (and How to Create Them with ggplot2)"



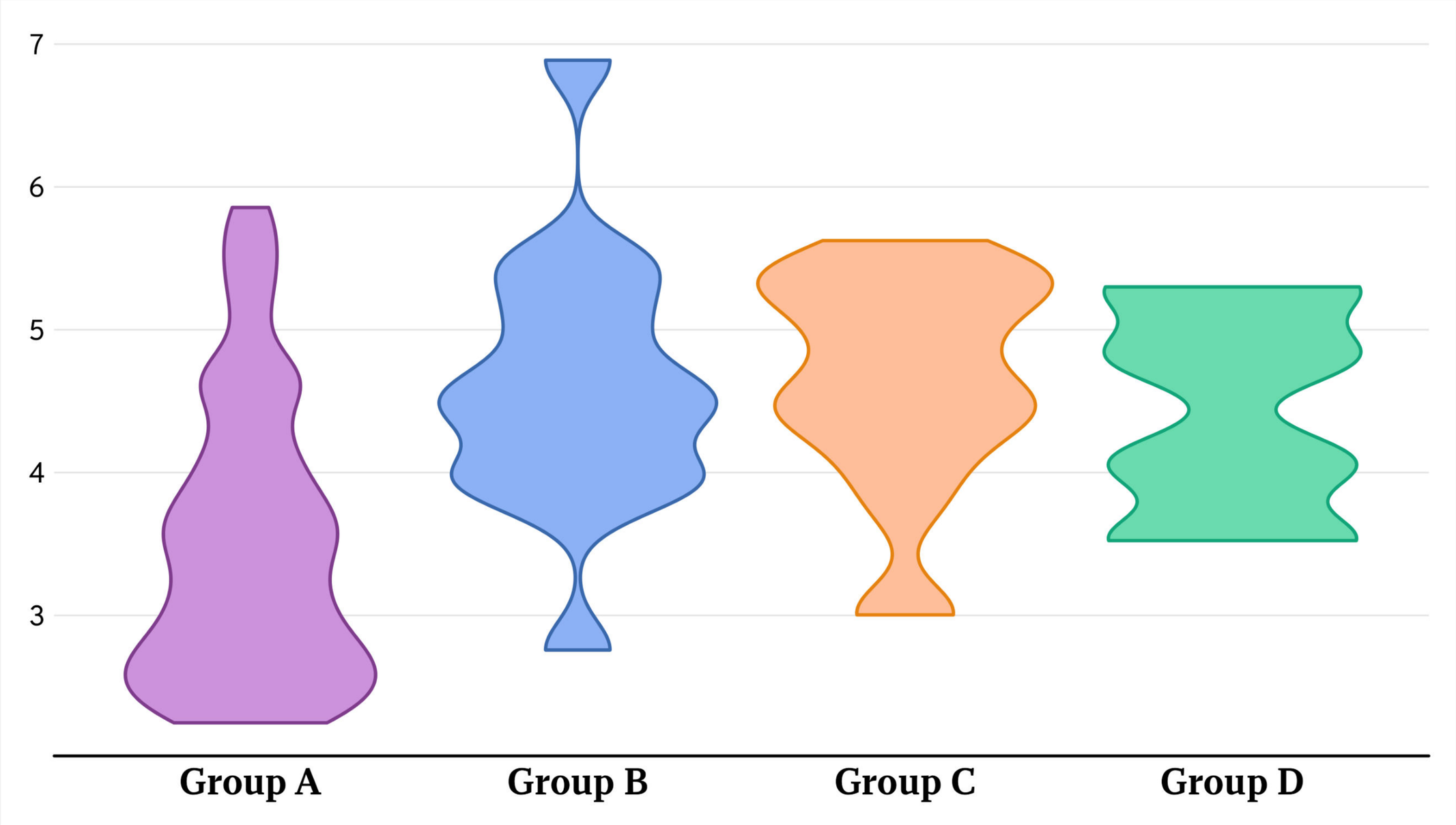
From my blog post "Visualizing Distributions with Raincloud Plots (and How to Create Them with ggplot2)"



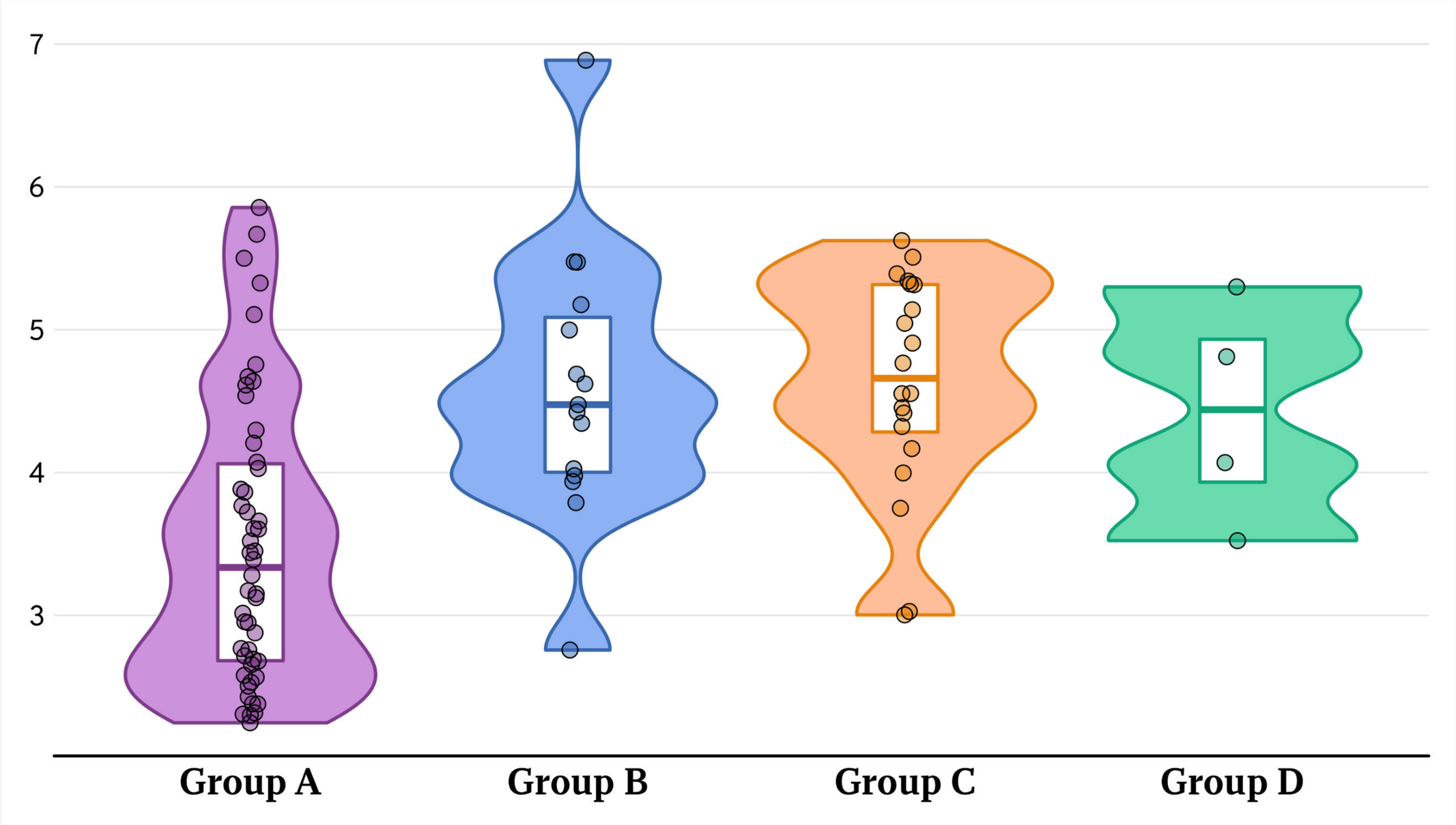
*My webinar "**Beyond Bar and Box Plots**" for USGS Data Science [Slides | Recording]*



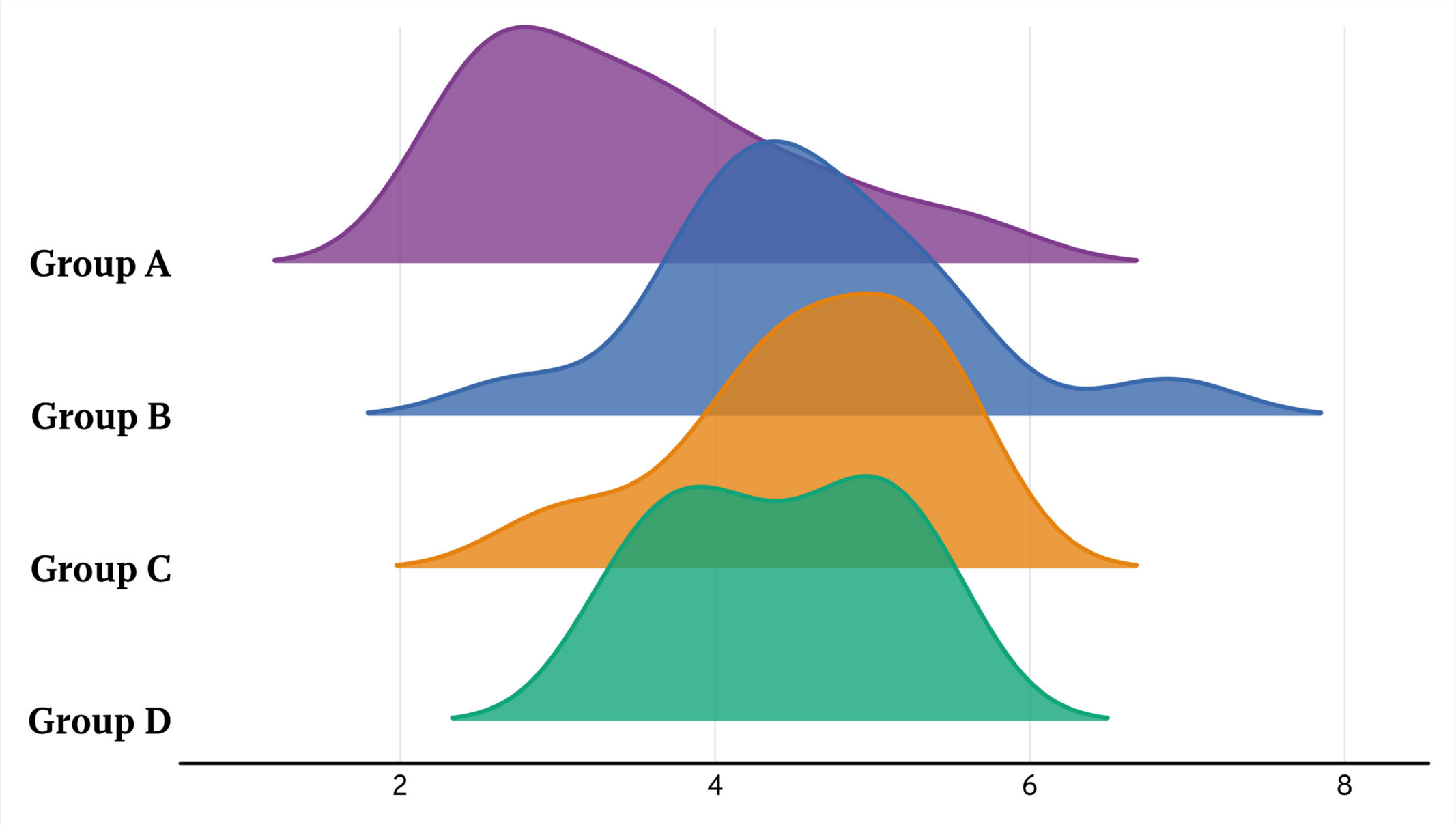
My webinar "Beyond Bar and Box Plots" for USGS Data Science [Slides | Recording]



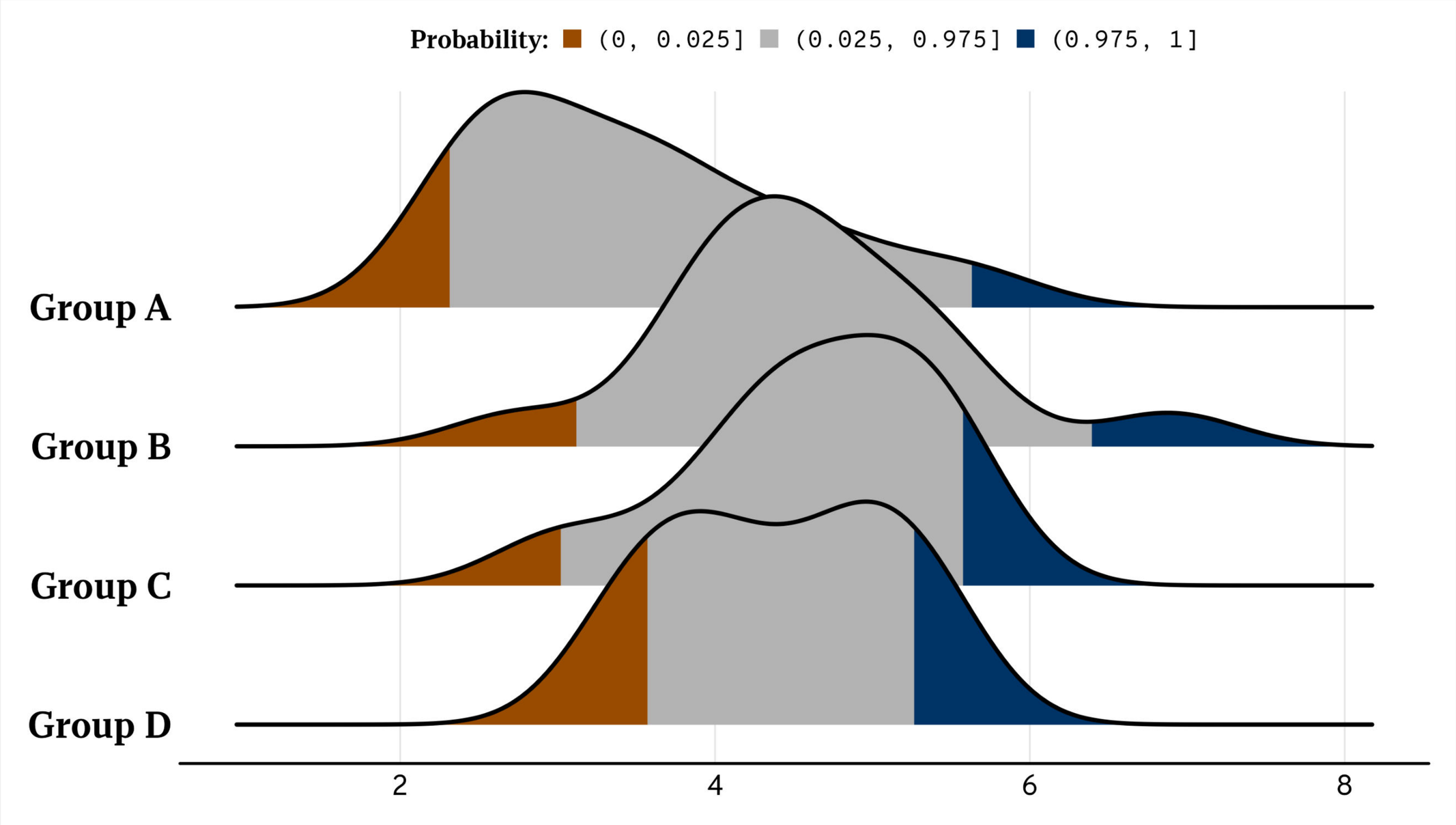
*My webinar "**Beyond Bar and Box Plots**" for USGS Data Science [Slides | Recording]*



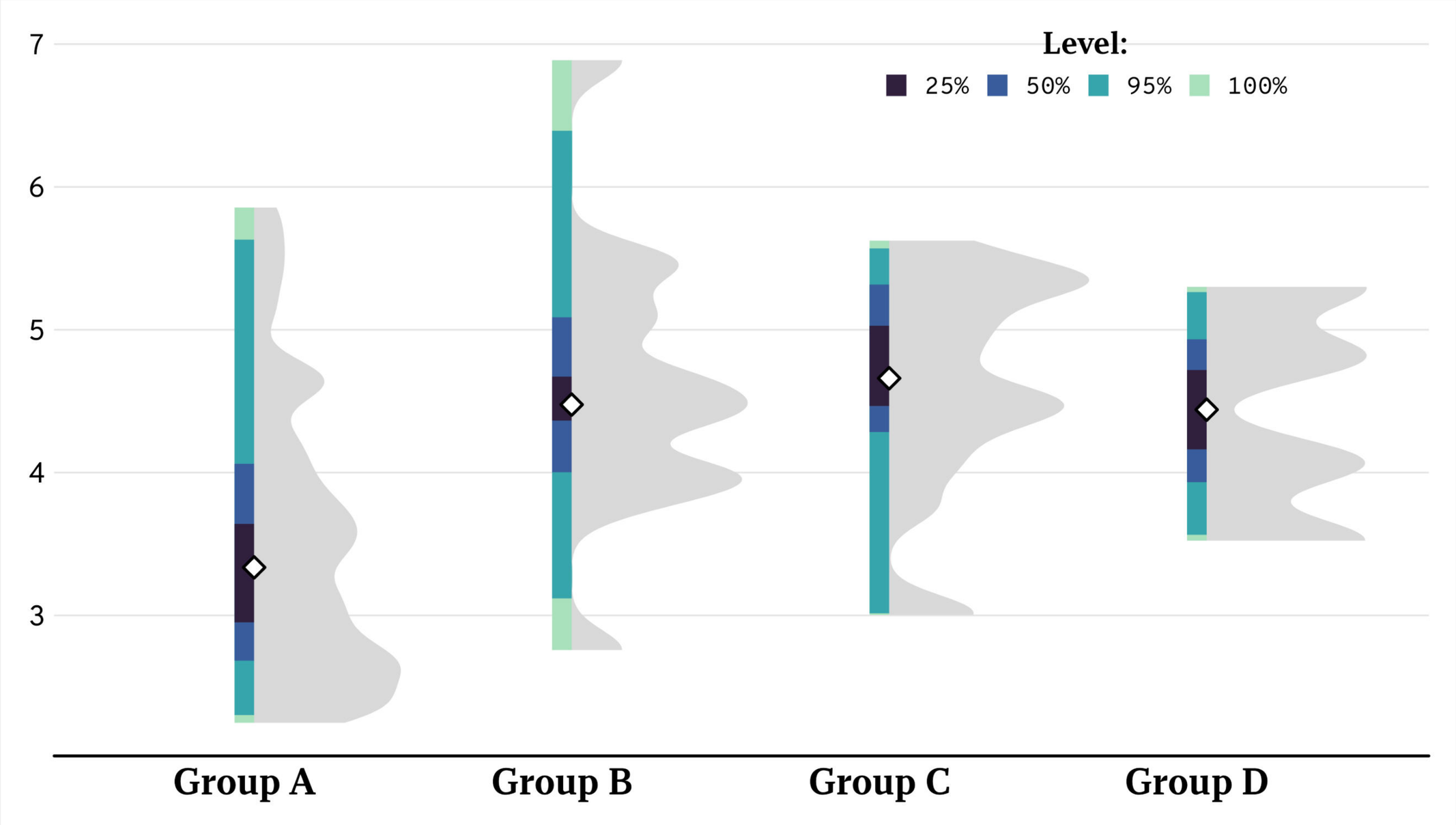
*My webinar "**Beyond Bar and Box Plots**" for USGS Data Science [Slides | Recording]*



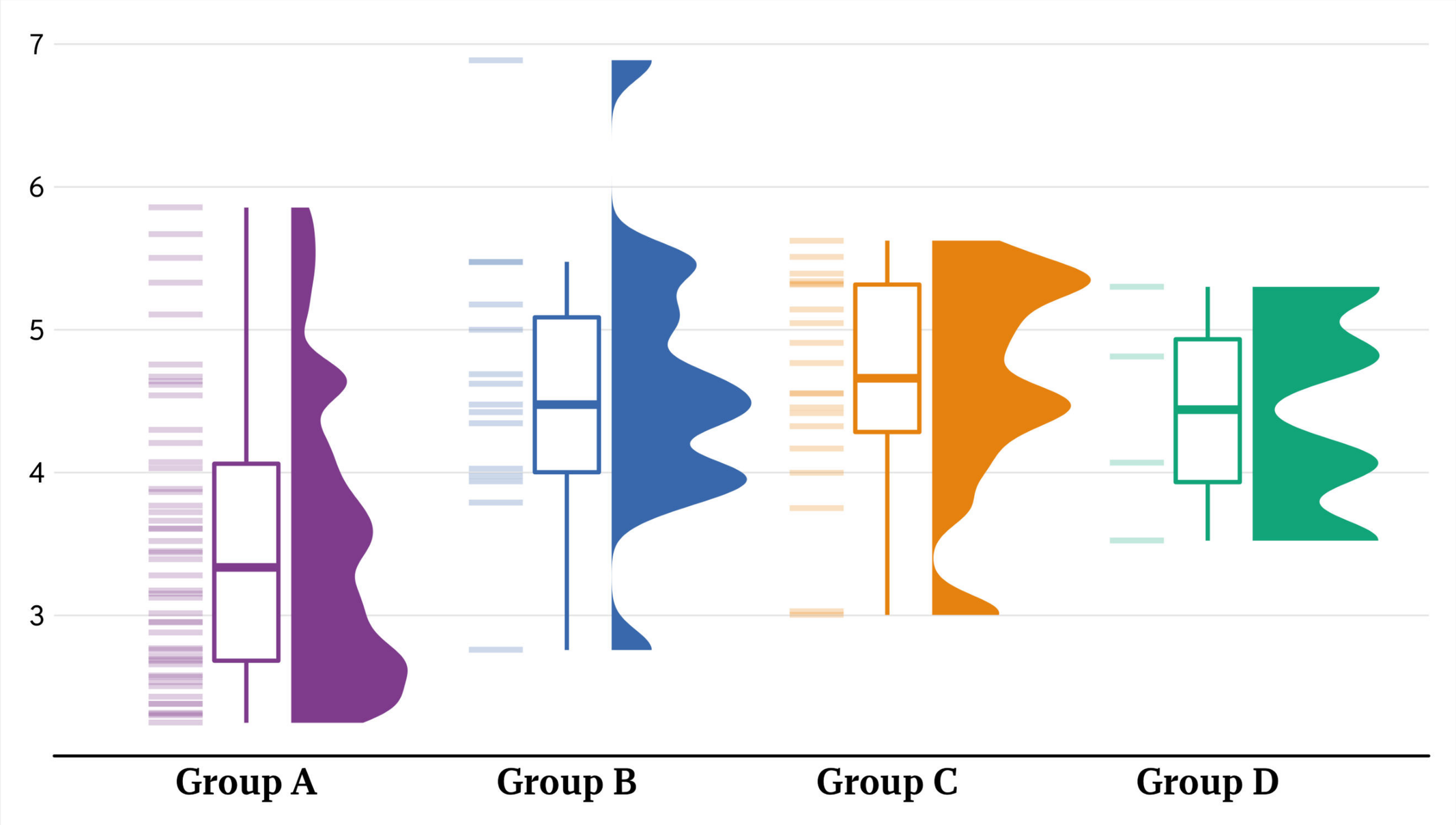
*My webinar "**Beyond Bar and Box Plots**" for USGS Data Science [Slides | Recording]*



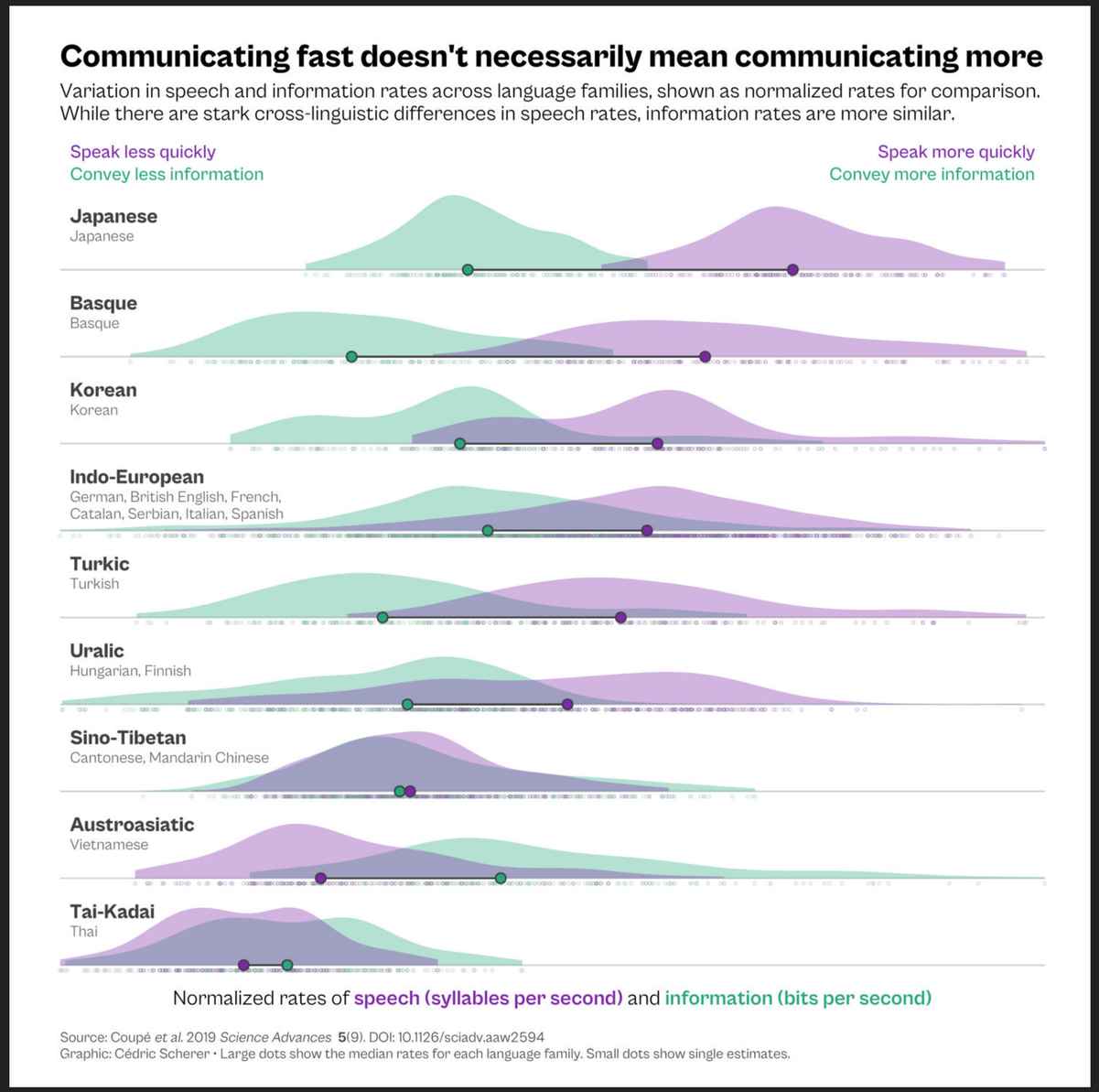
*My webinar "**Beyond Bar and Box Plots**" for USGS Data Science [Slides | Recording]*



*My webinar "**Beyond Bar and Box Plots**" for USGS Data Science [Slides | Recording]*

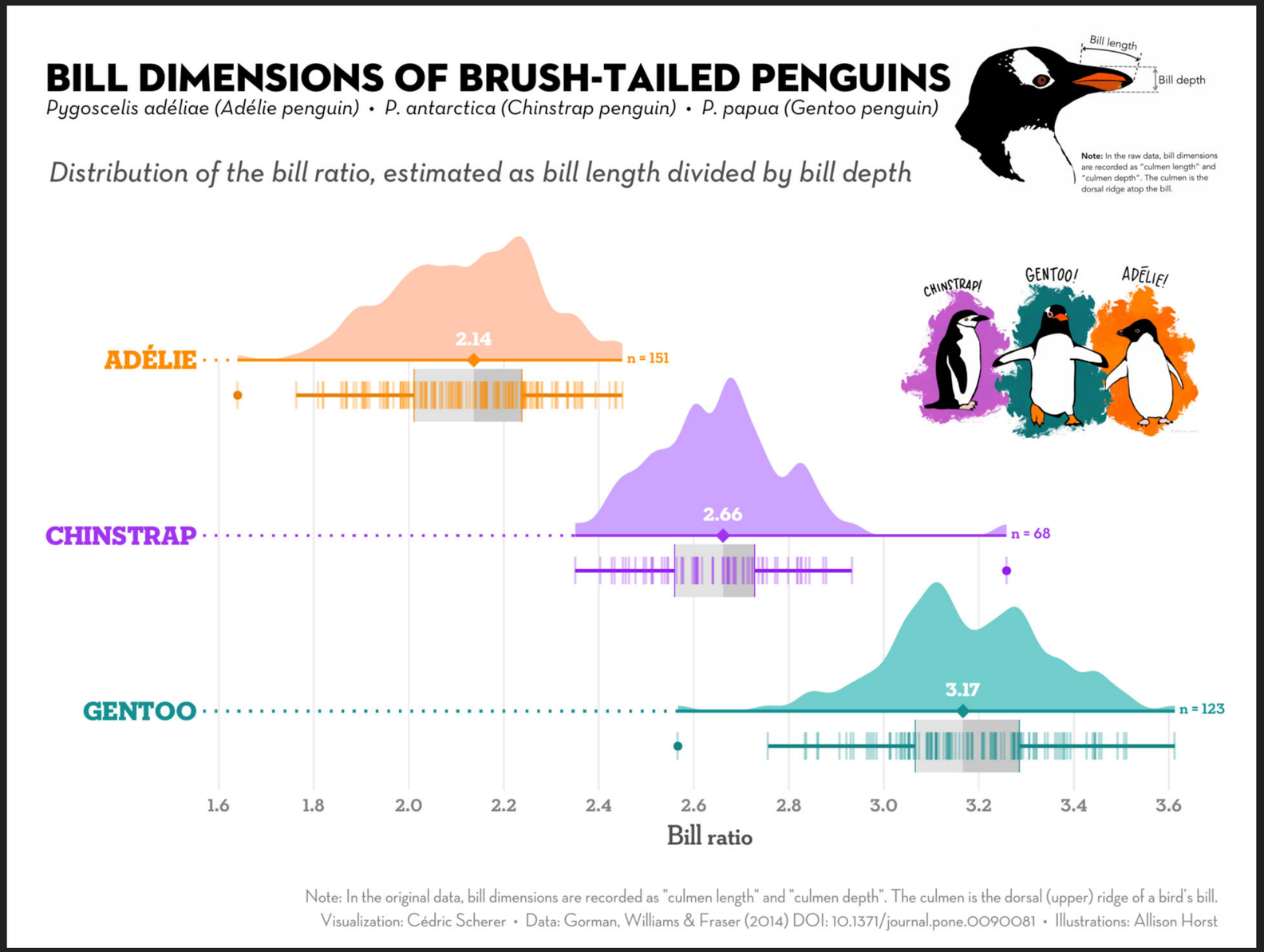


*My webinar "**Beyond Bar and Box Plots**" for USGS Data Science [Slides | Recording]*

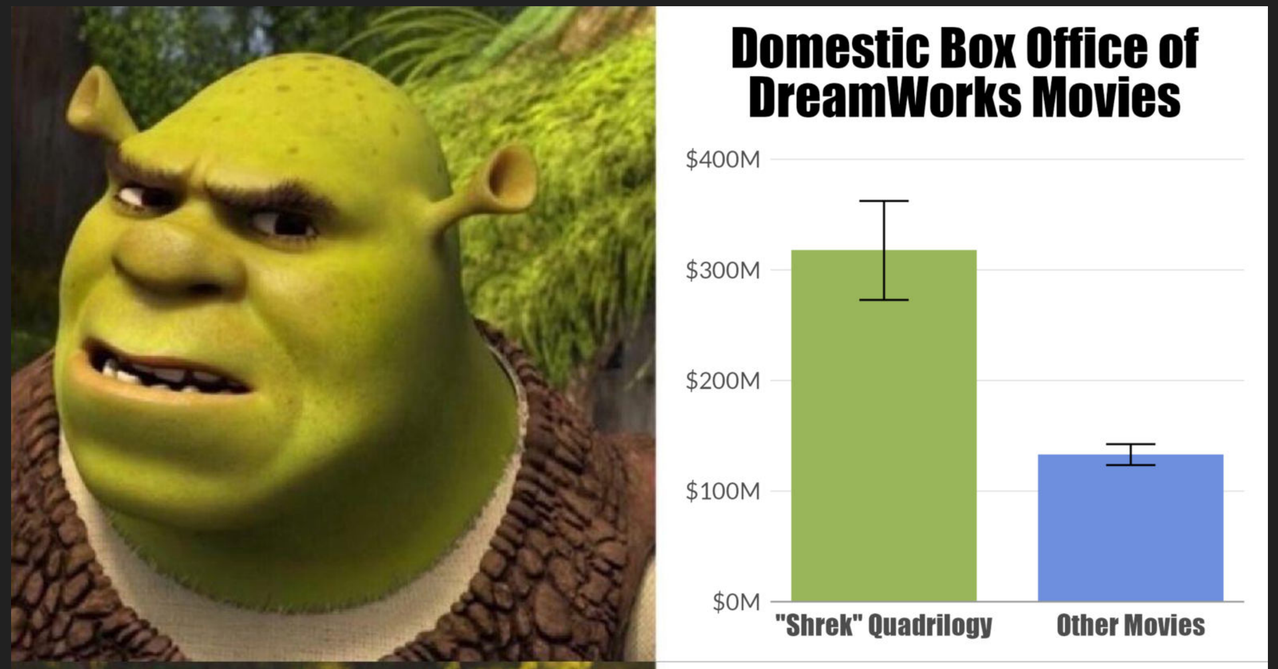


"Communicating fast doesn't necessarily mean communicating more", #30DayChartChallenge Contribution

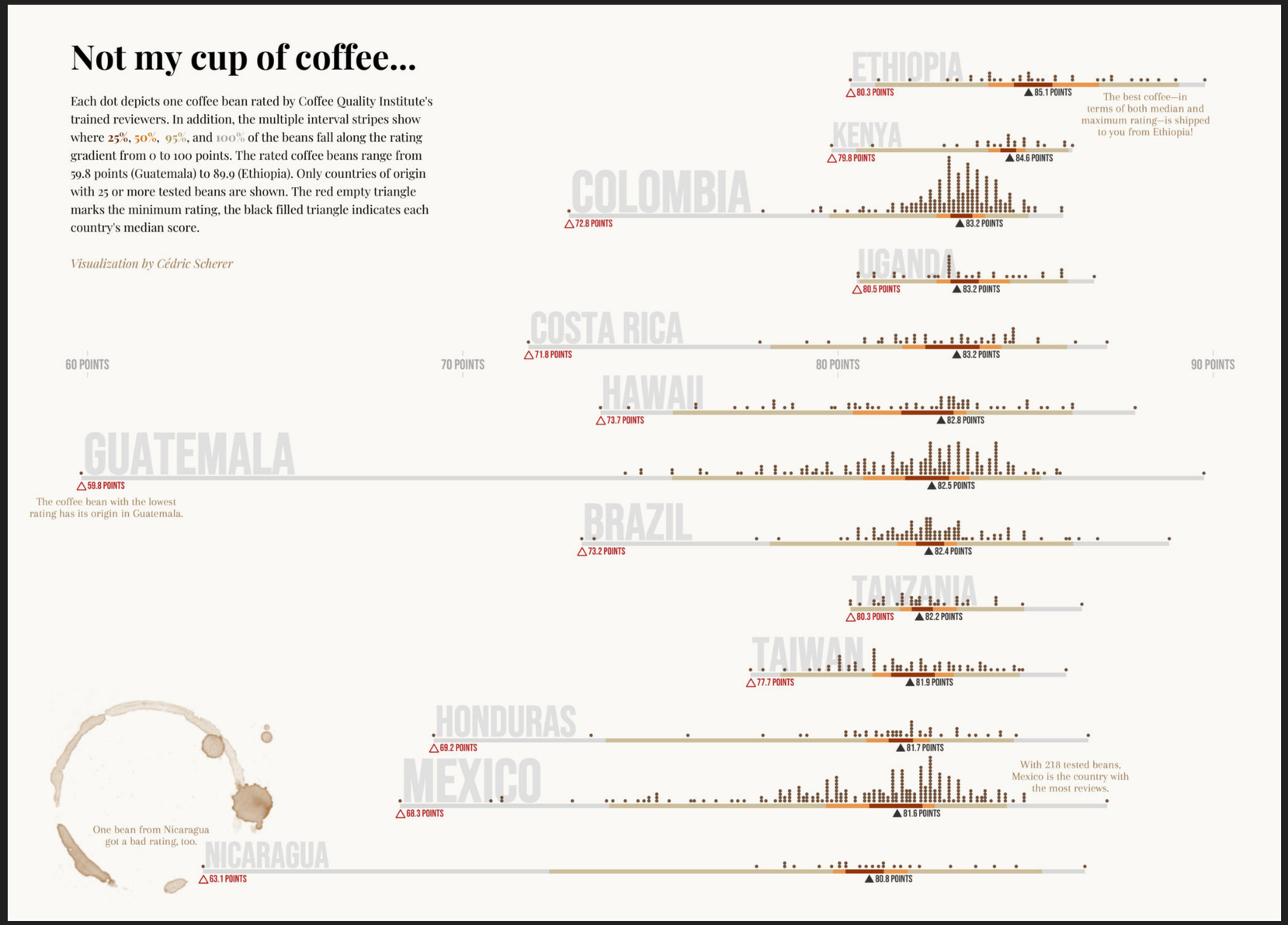
cedricscherer.com @CedScherer z3tt



From my blog post "Visualizing Distributions with Raincloud Plots (and How to Create Them with ggplot2)"



Why Dynamite Plots Are Terrible—and Why You Should Use Something Else | Cédric Scherer | #30DayChartChallenge 2021 | Day 27: Educational

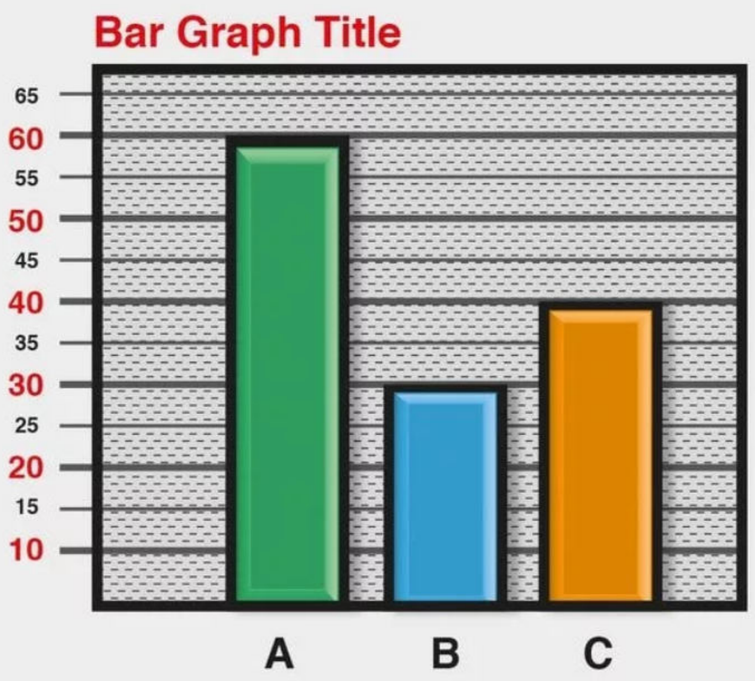


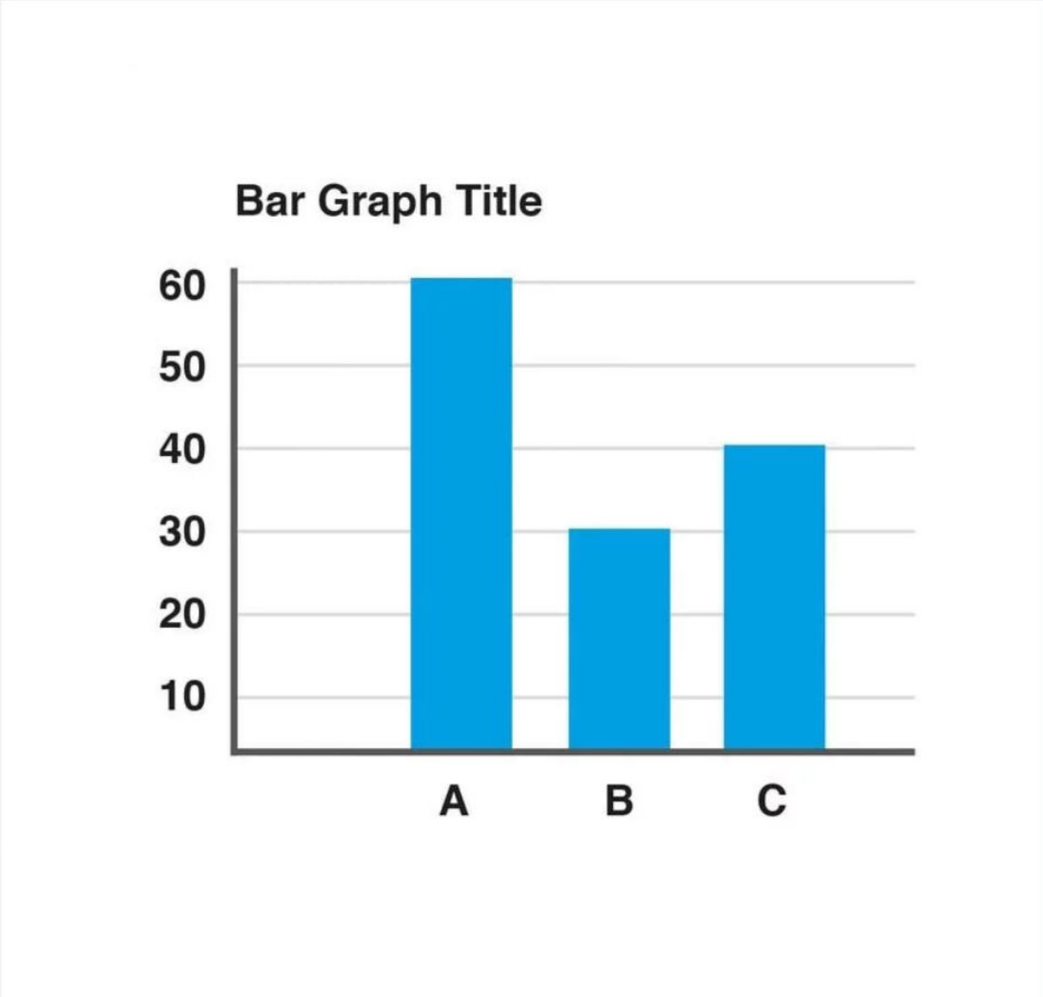
"Not my cup of coffee", #TidyTuesday Contribution

cedricscherer.com @CedScherer z3tt

ONE
WAY
ONE
WAY
Guide the
Viewer (eg)

Decluttering

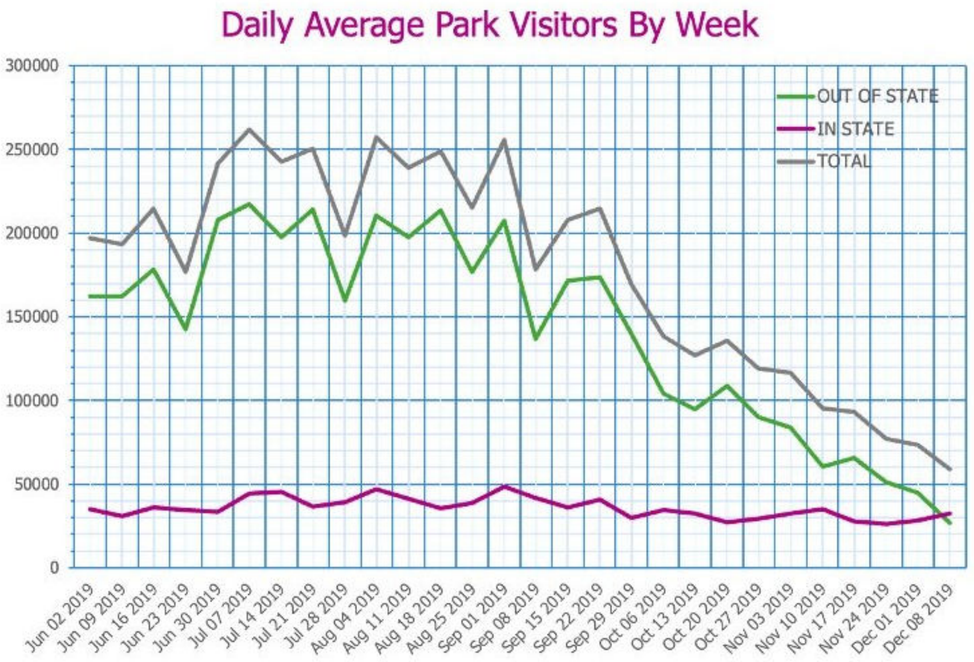




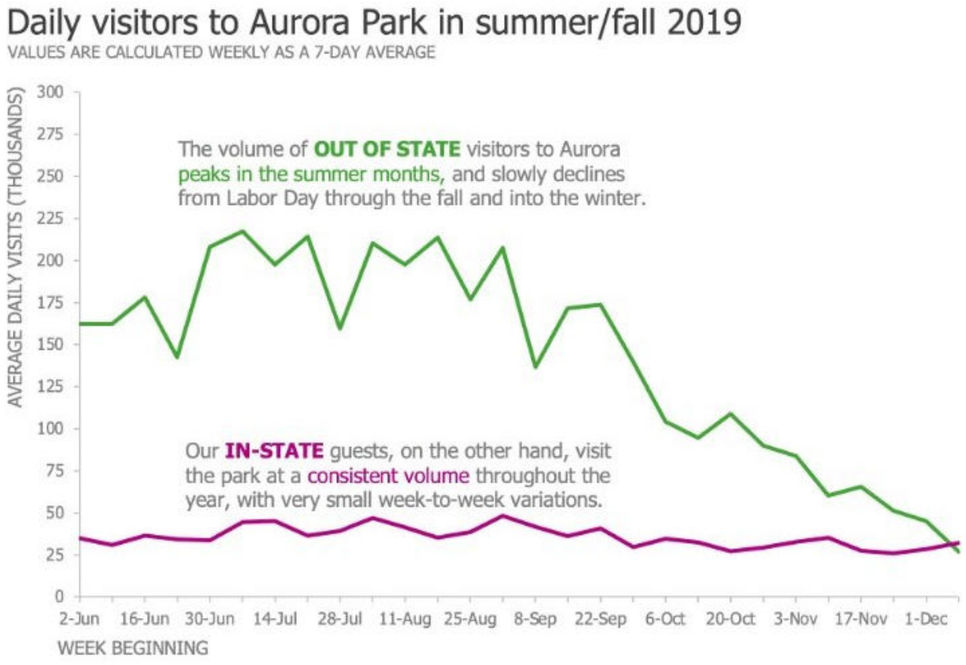
Source: canva.com

Decluttering

BEFORE: SPOOKY SKELETON



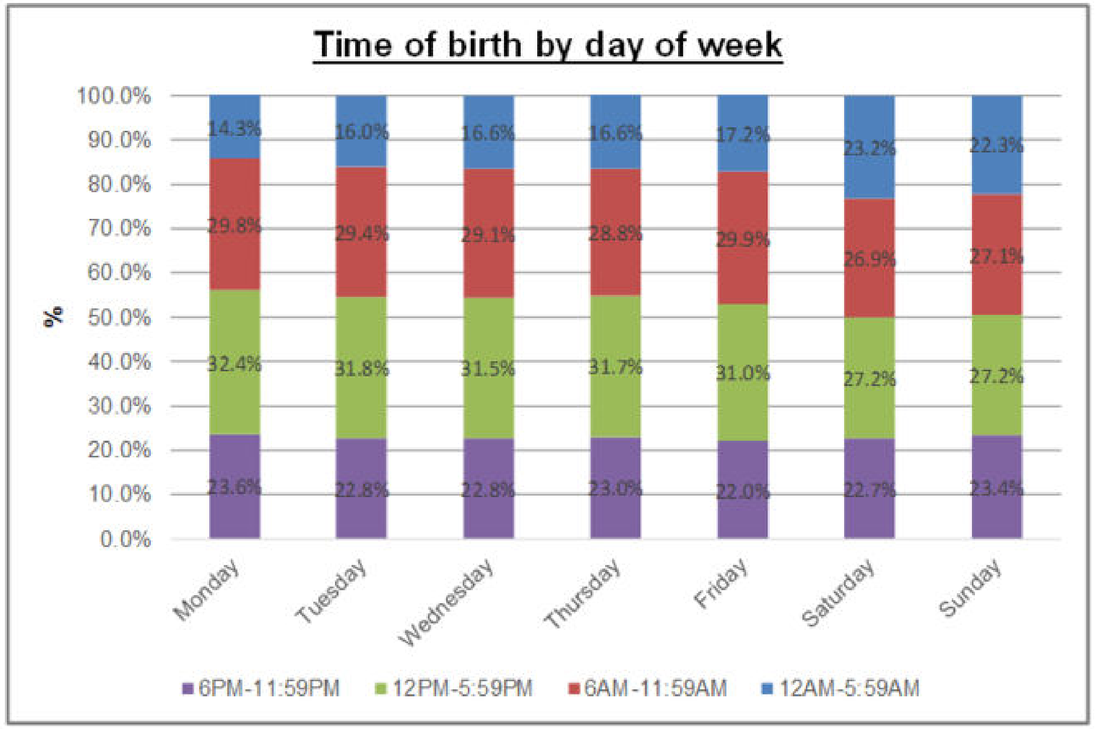
AFTER: GOOD BONES



"Your graph Skeleton Shouldn't Be Scary" by Mike Cisneros (Storytelling With Data)

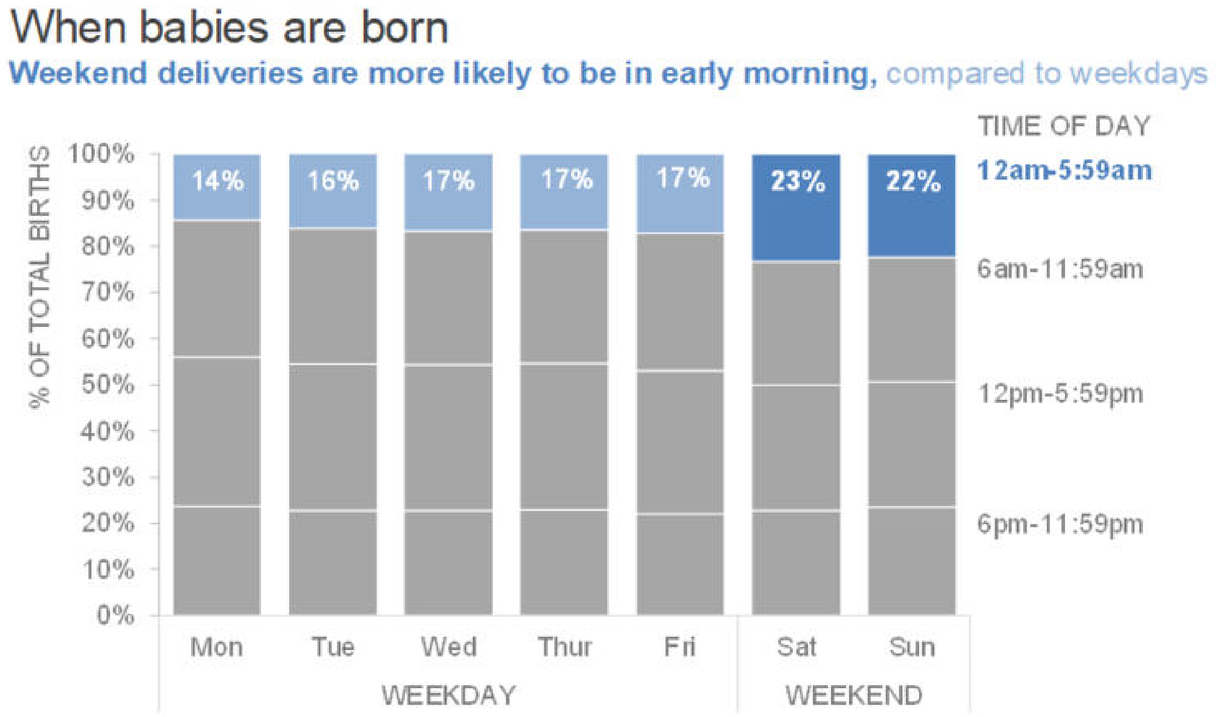
Decluttering

BEFORE



Data source: CDC (National Vital Statistics Reports, Vol. 67, No. 1, January 31, 2018)

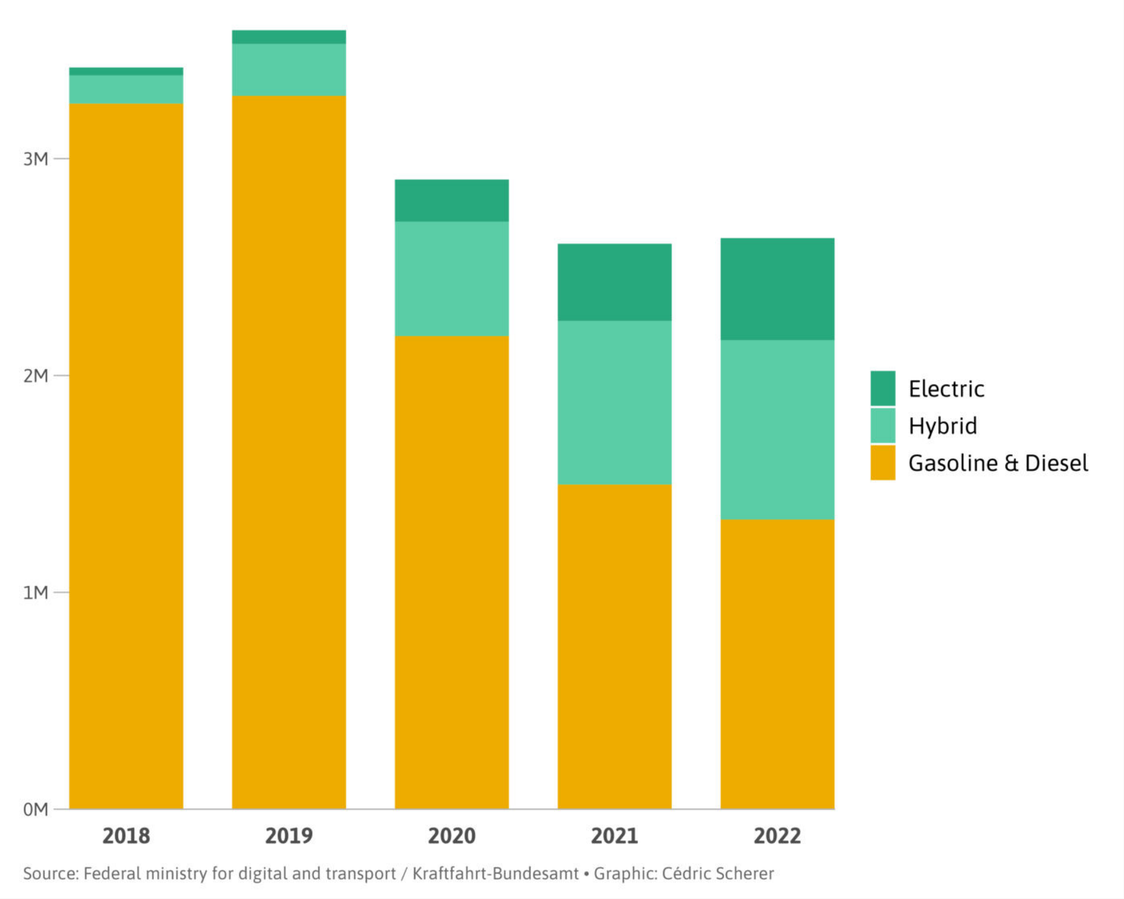
AFTER

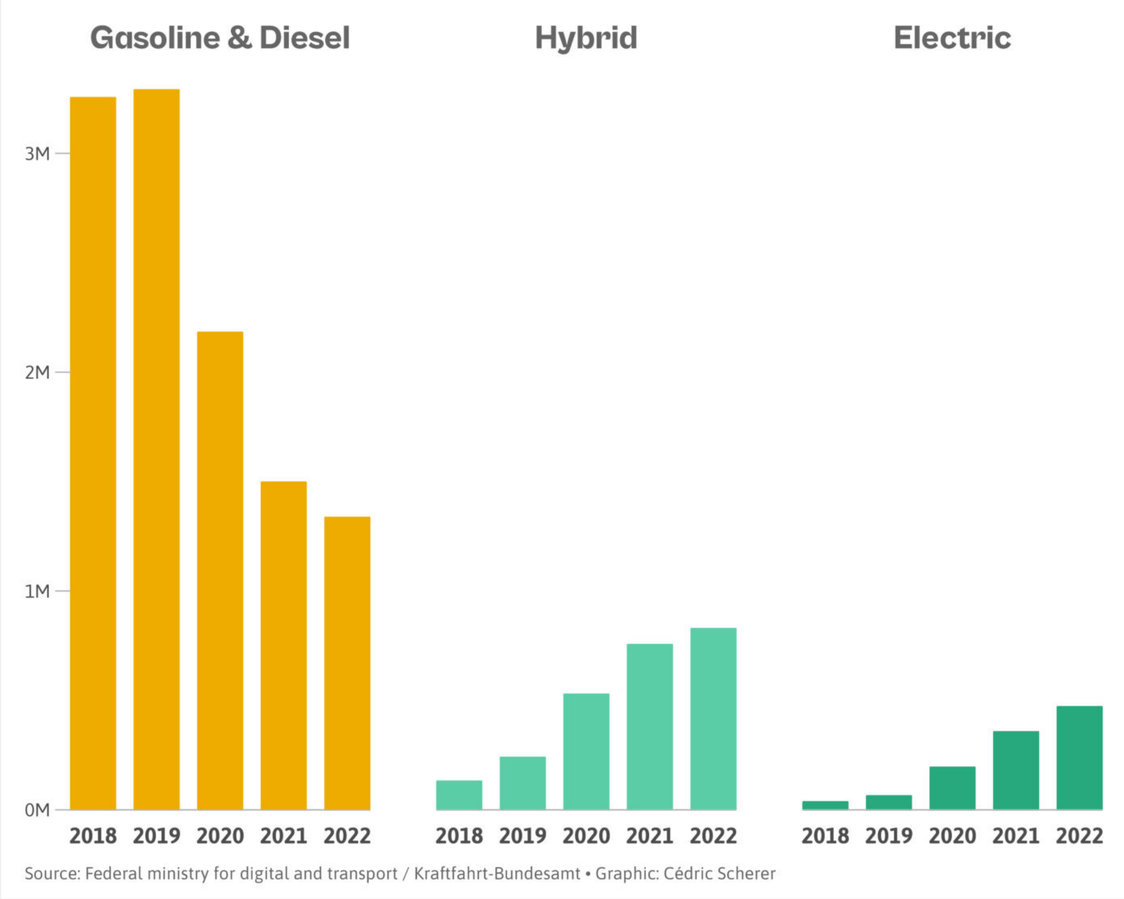


Data source: CDC (National Vital Statistics Reports, Vol. 67, No. 1, January 31, 2018)

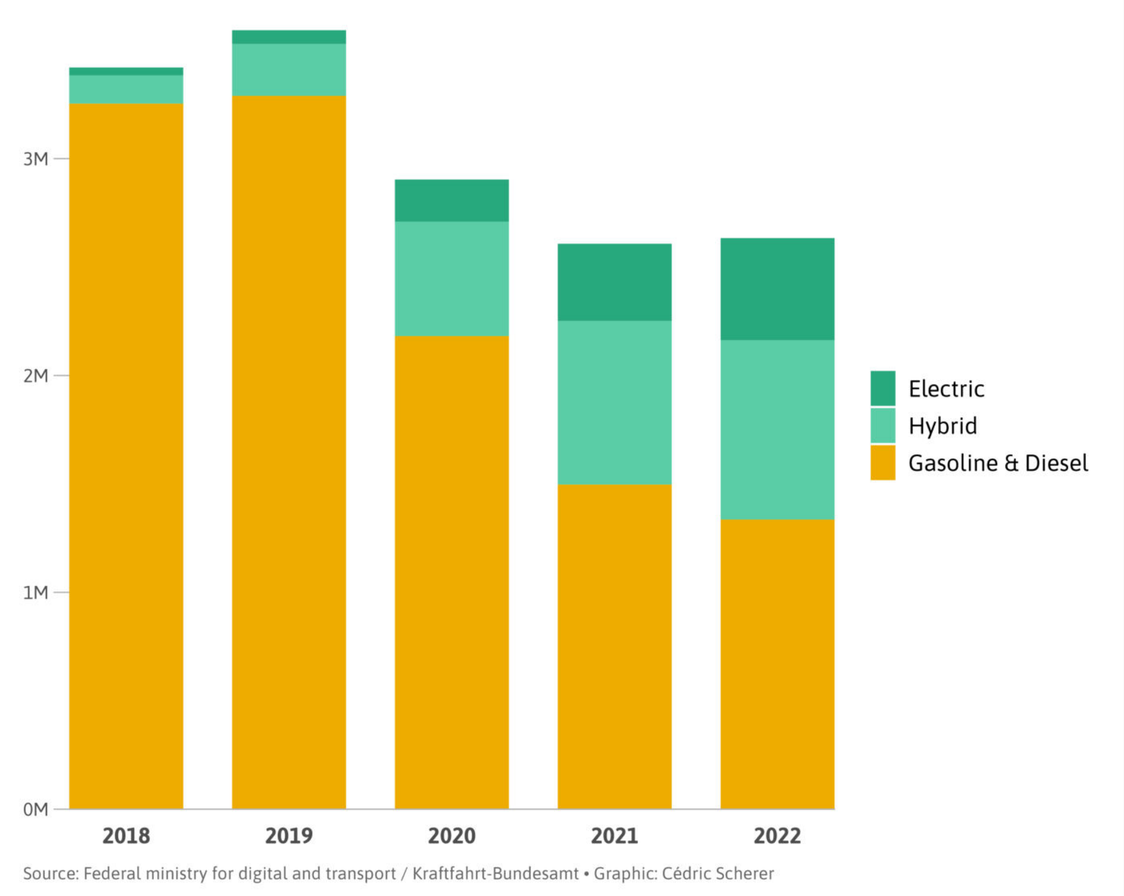
"Declutter! (and Question Default Settings)" by Elizabeth Ricks (Storytelling With Data)

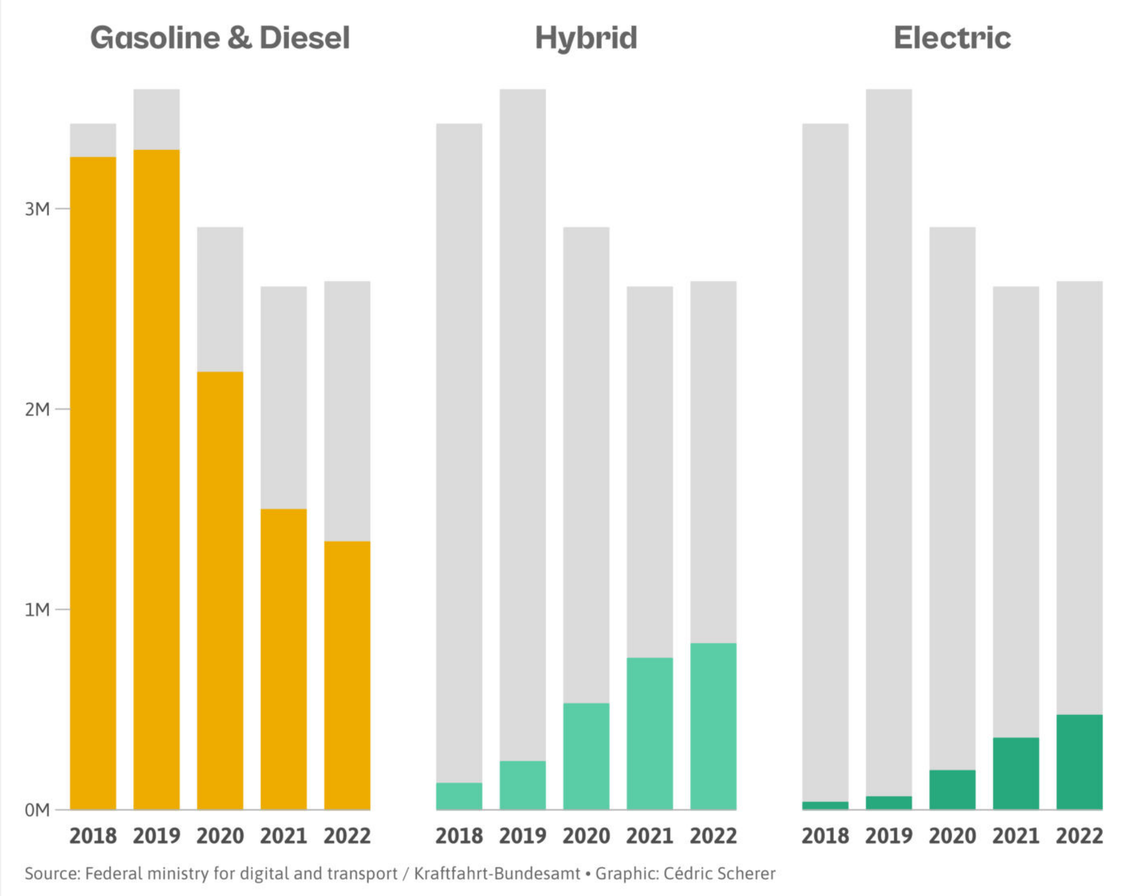
The Power of Small Multiples



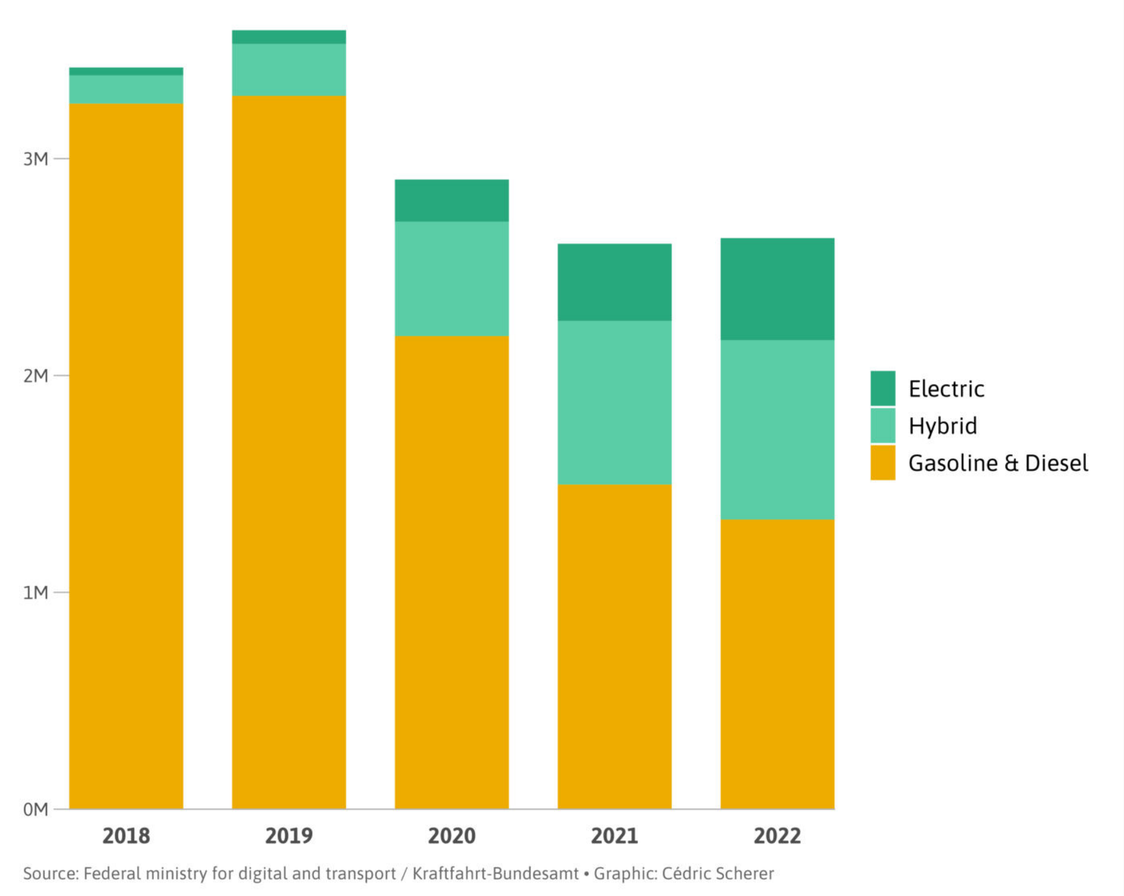


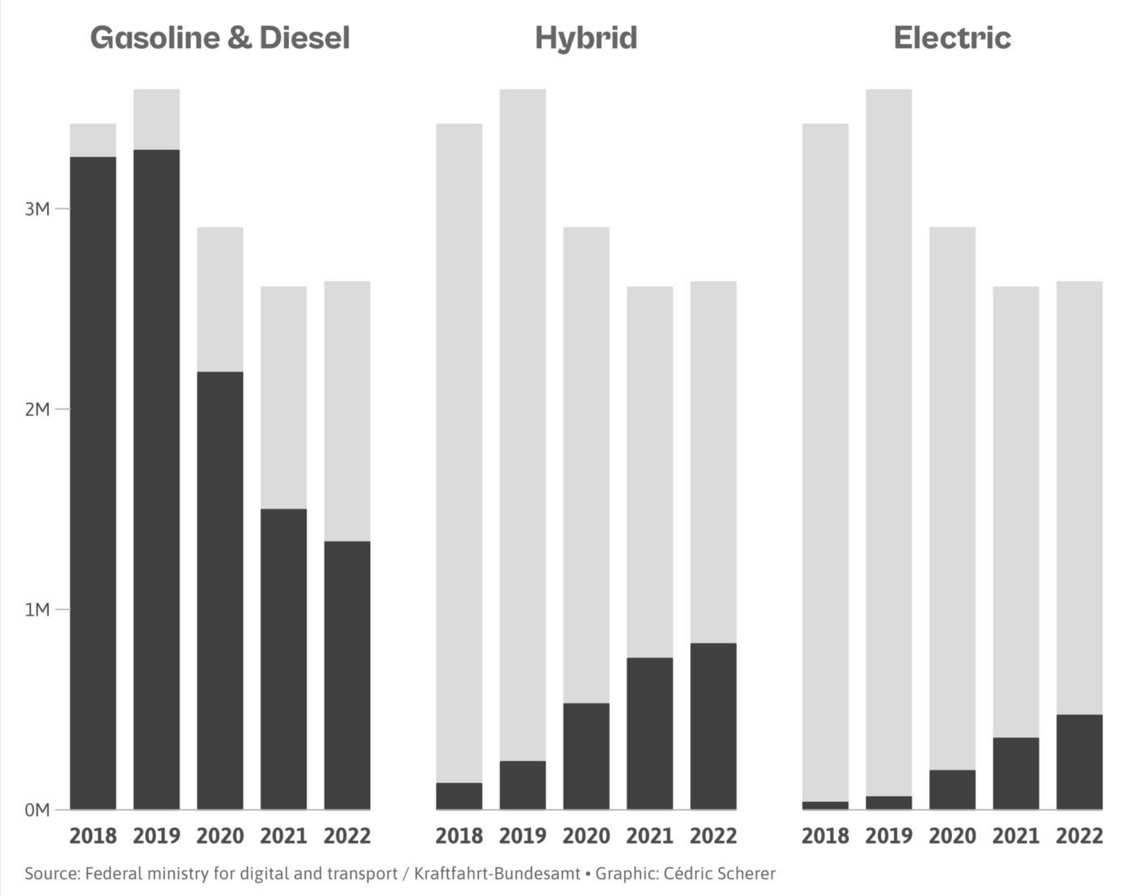
The Power of Small Multiples



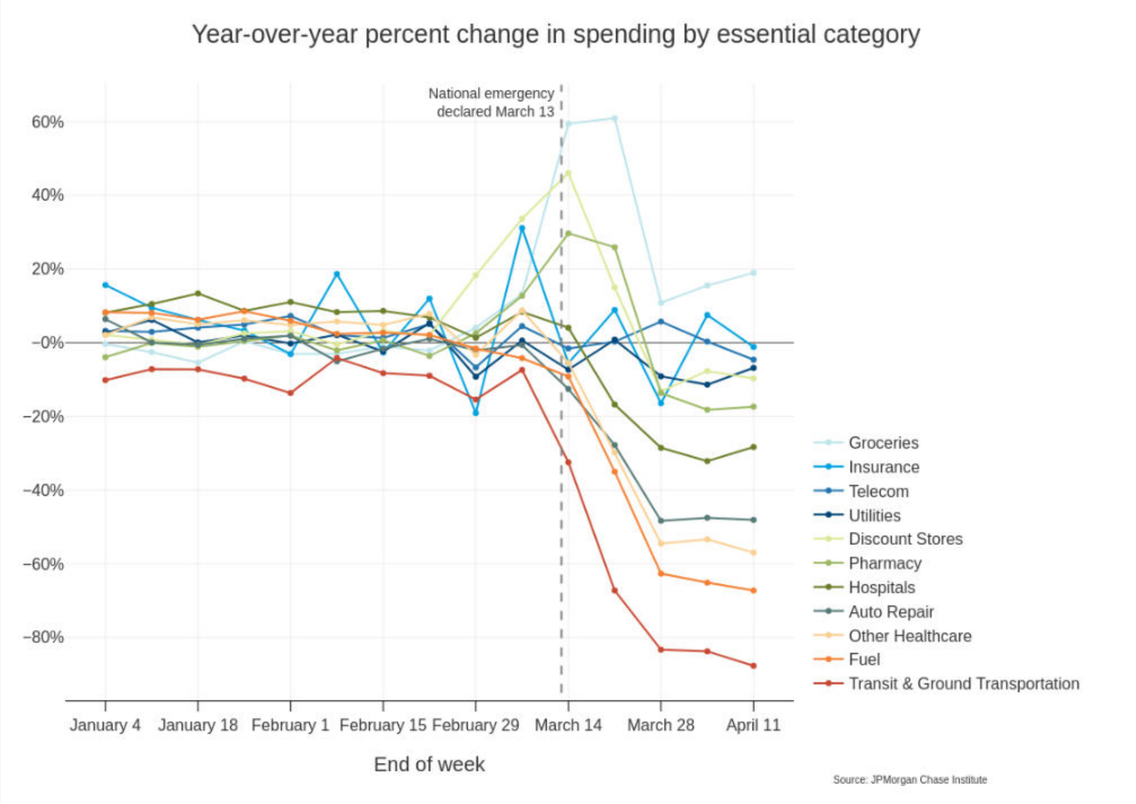


The Power of Small Multiples

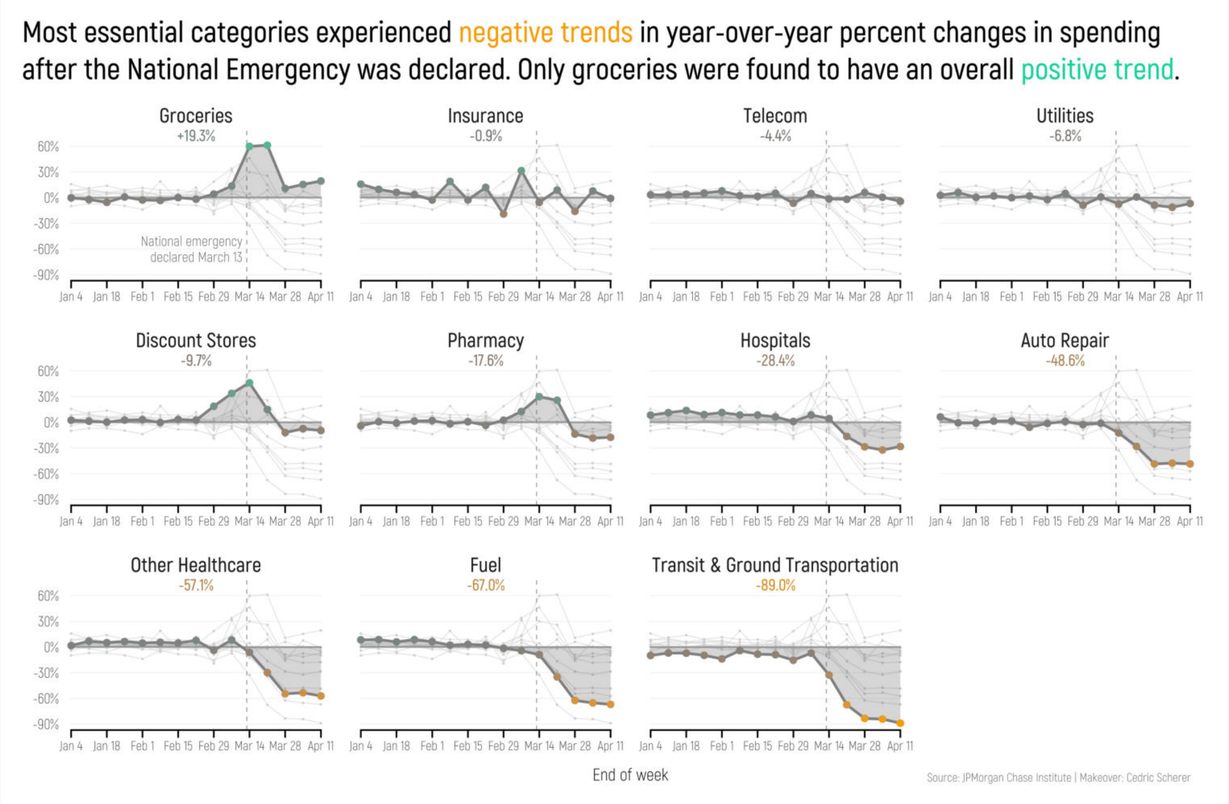




The Power of Small Multiples

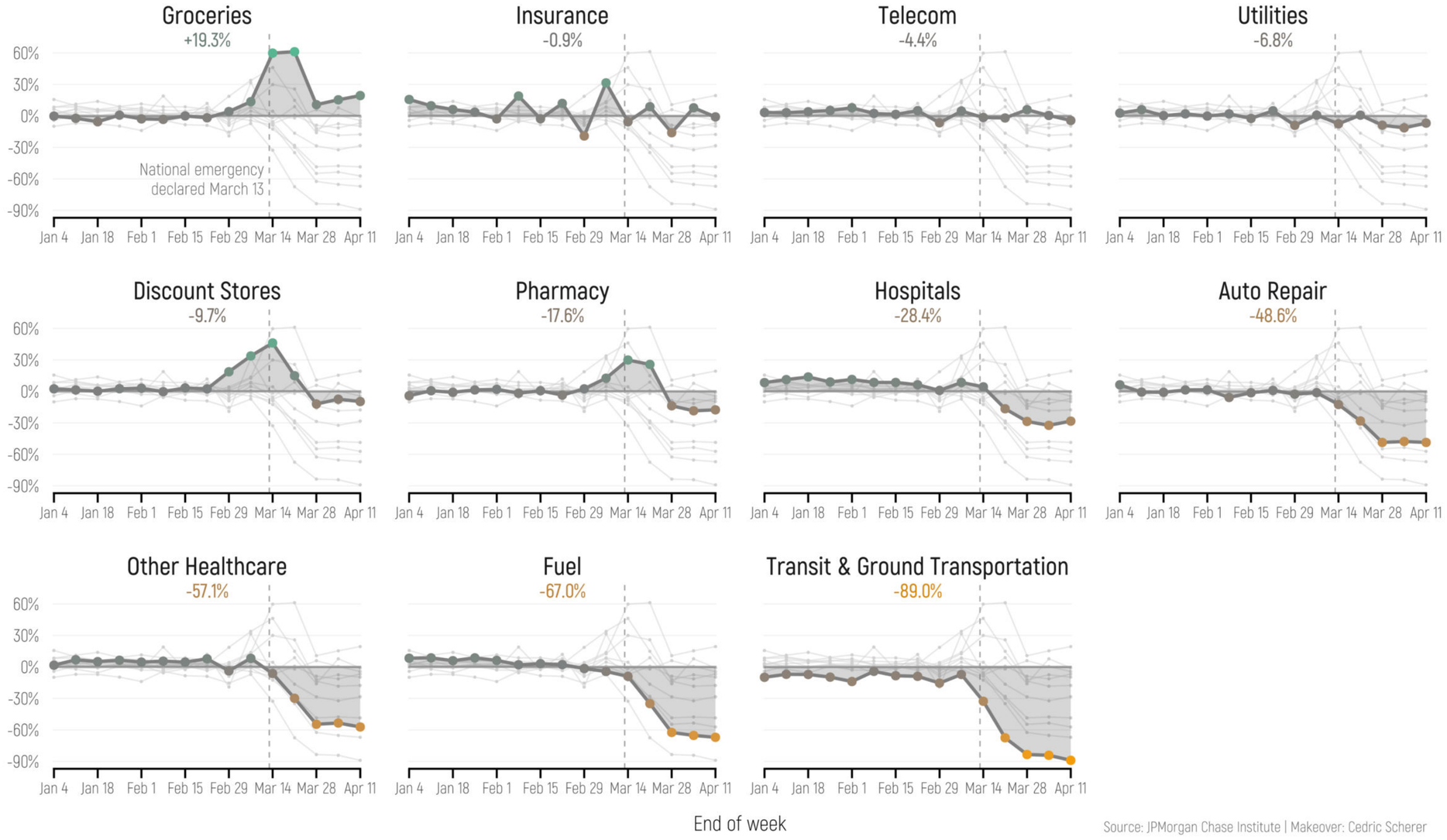


Originalgrafik vom JPMorgan Chase Institut

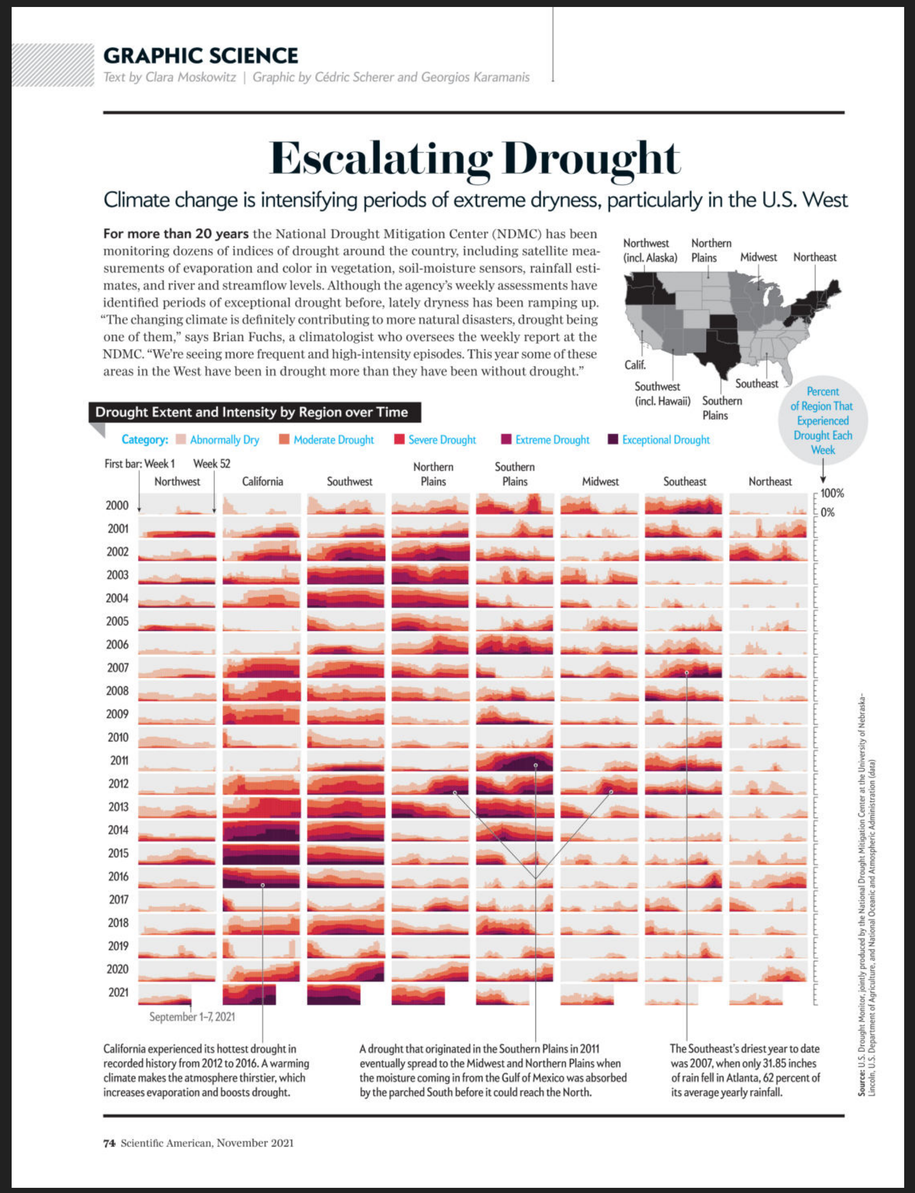


Reworked graphic using small multiples

Most essential categories experienced negative trends in year-over-year percent changes in spending after the National Emergency was declared. Only groceries were found to have an overall positive trend.



Source: JPMorgan Chase Institute | Makeover: Cedric Scherer



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."

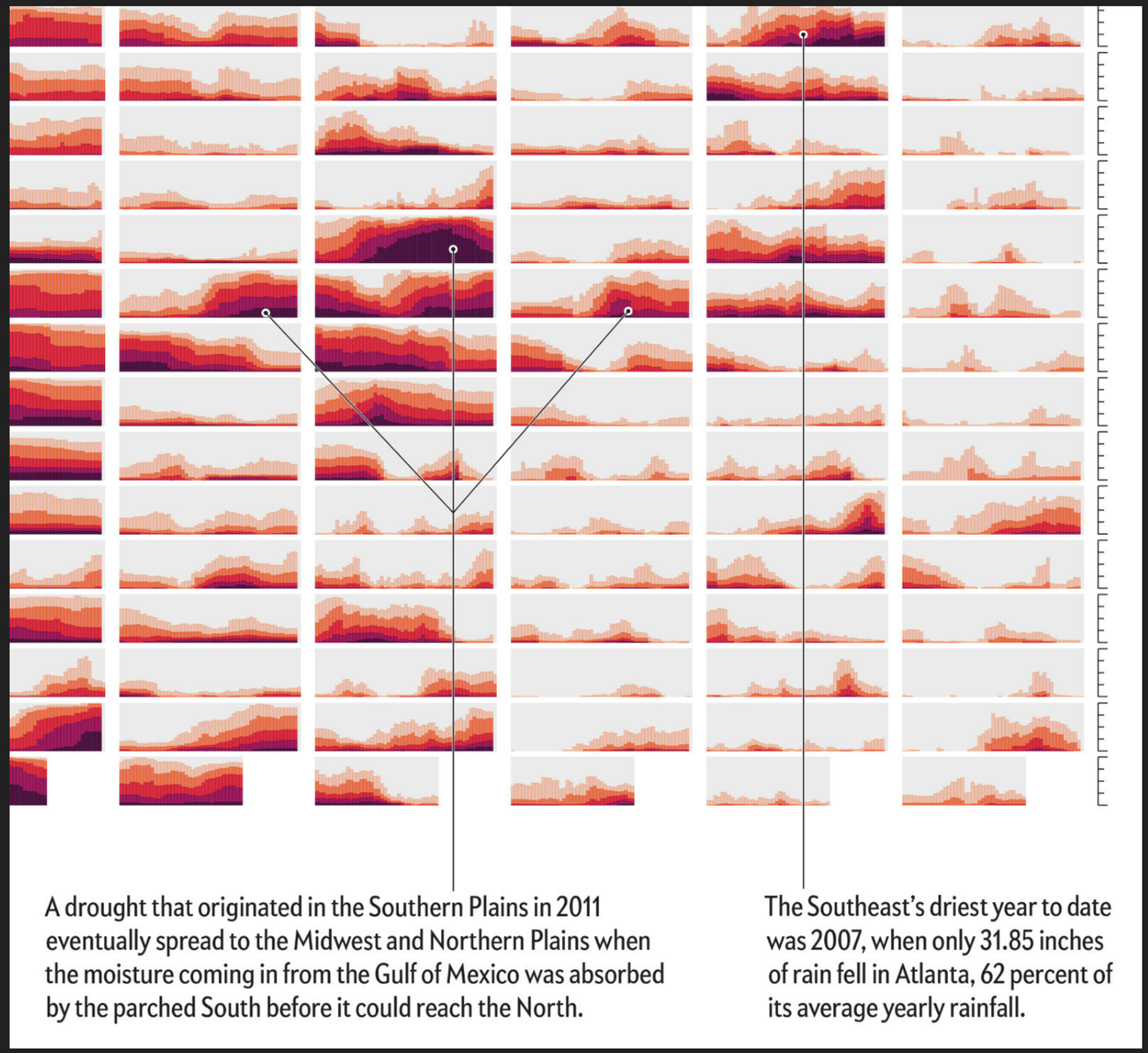
Drought Extent and Intensity by Region over Time

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.

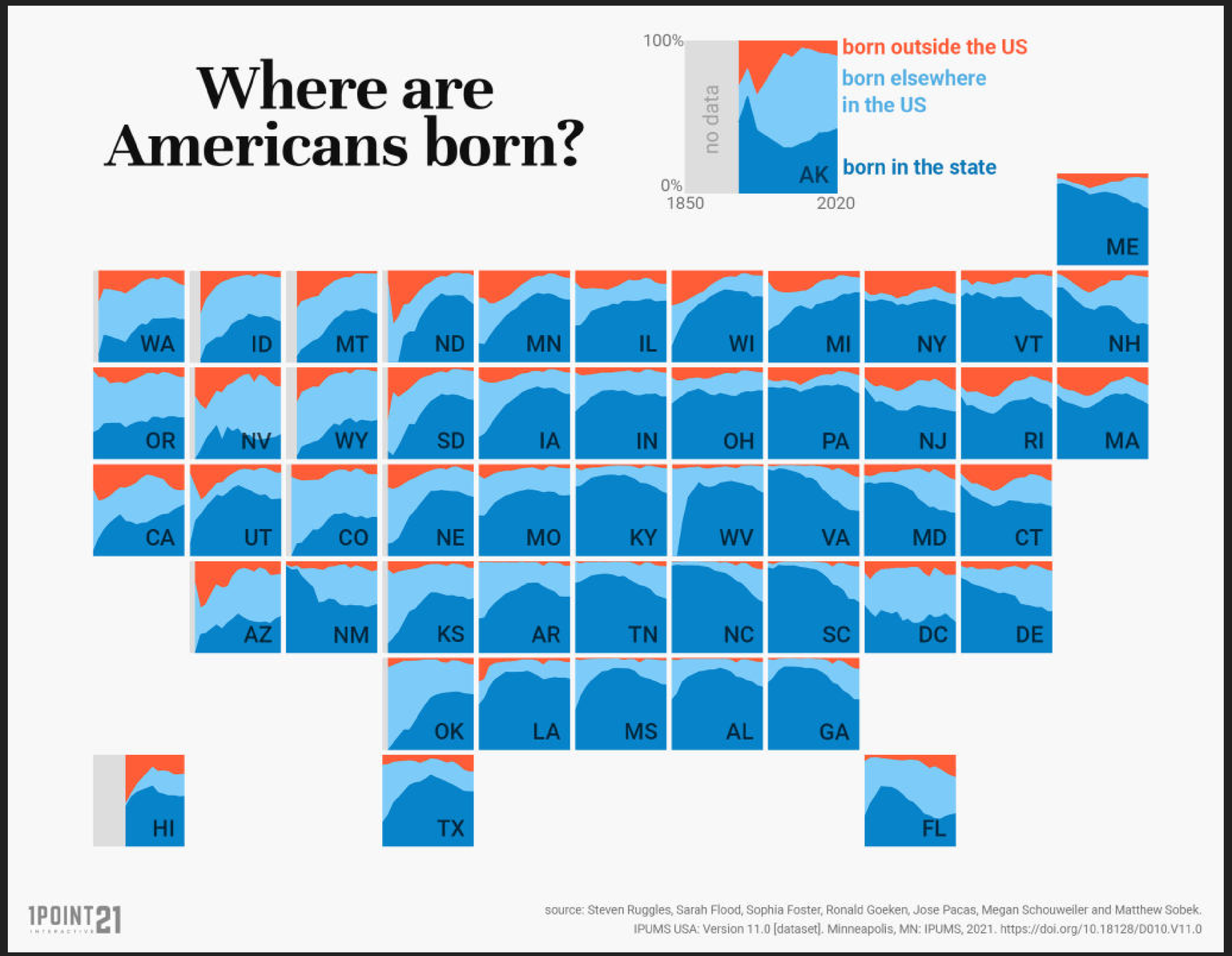
Source: U.S. Drought Monitor, jointly produced by the National Drought Mitigation Center at the University of Nebraska-Lincoln, U.S. Department of Agriculture, and National Oceanic and Atmospheric Administration (data)



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.

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



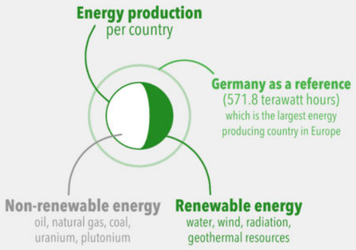
"Where are Americans born?" by @ErinDataViz

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.

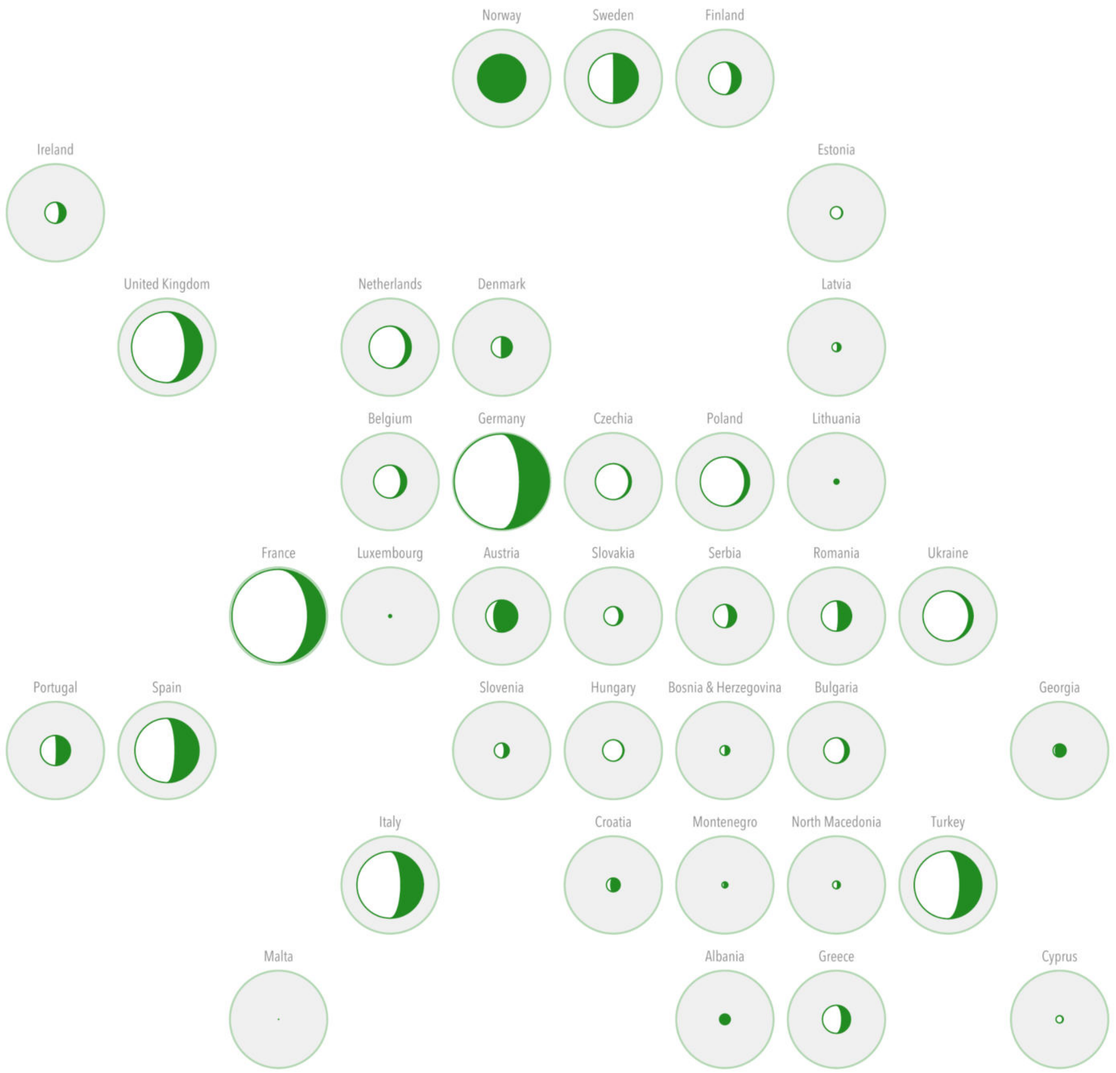
Renewable:	Nuclear:	Conventional thermal:
Germany	**France**	**Germany**
179.1 TWh	393.2 TWh	320.4 TWh
(31% of its production)	(71% of its production)	(56% of its production)

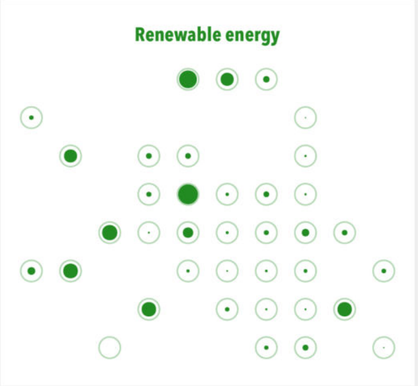
Renewable energy is energy that comes from resources that are naturally replenished such as sunlight, wind, water, and geothermal heat. Unlike fossil fuels, such as oil, natural gas and coal, or nuclear power sources such as uranium and plutonium, renewable energy regenerates naturally in a short period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

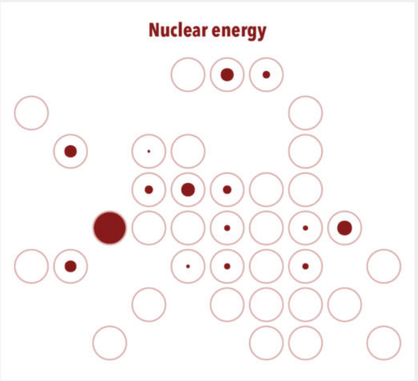


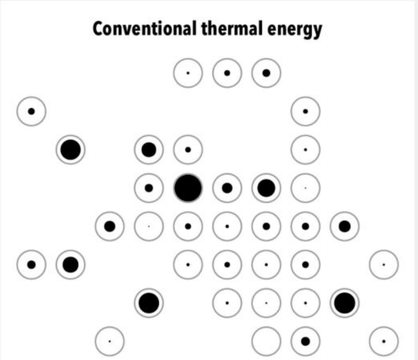
Norway had an electricity production almost entirely made up of renewable energy (98%). This makes Norway the second largest producer of this energy type in Europe. Interestingly, most of the renewable energy is produced by hydro power that take up 95% and only 3% by wind. In contrast, twelve European 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



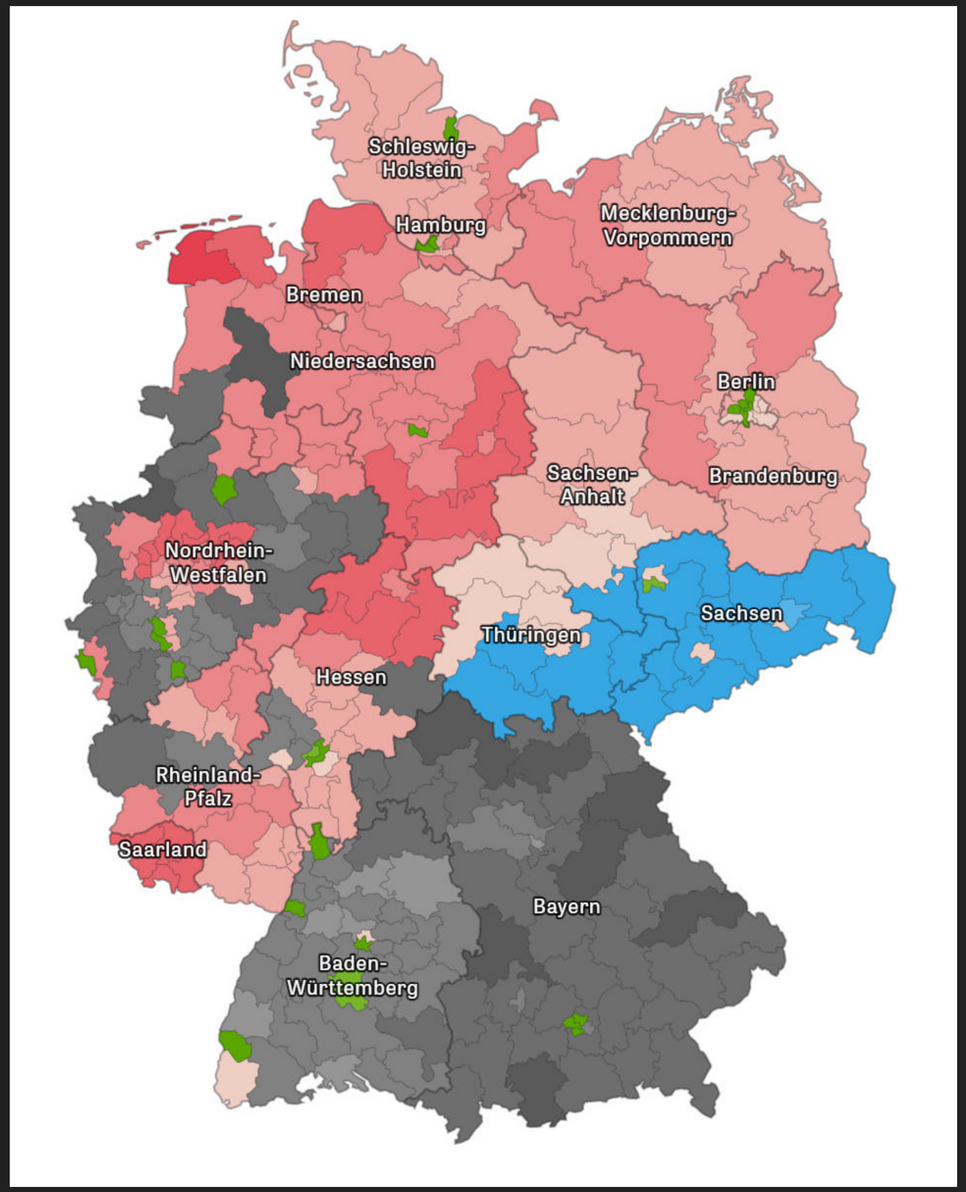


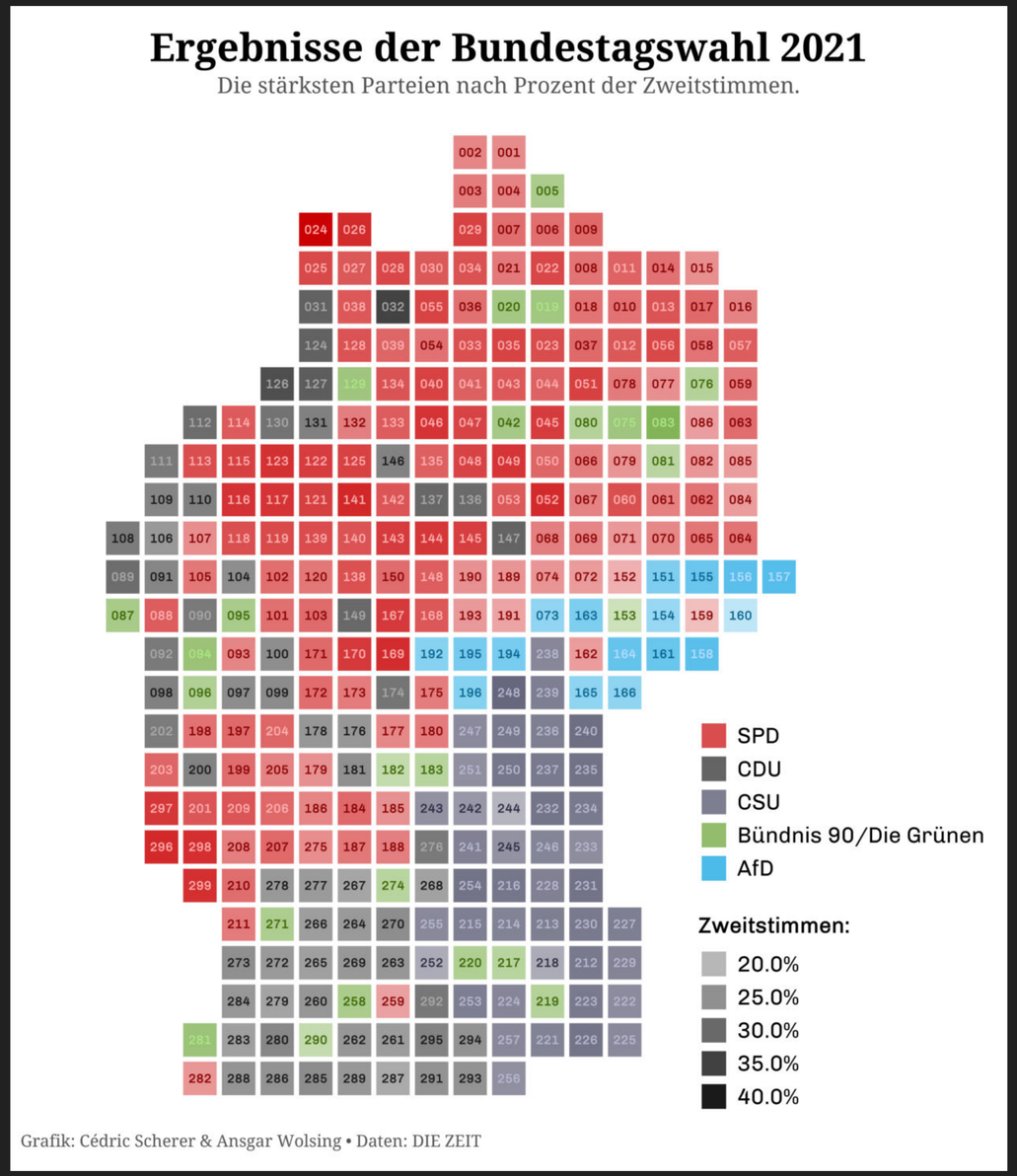




"How European countries generated electricity in 2018", #TidyTuesday Contribution

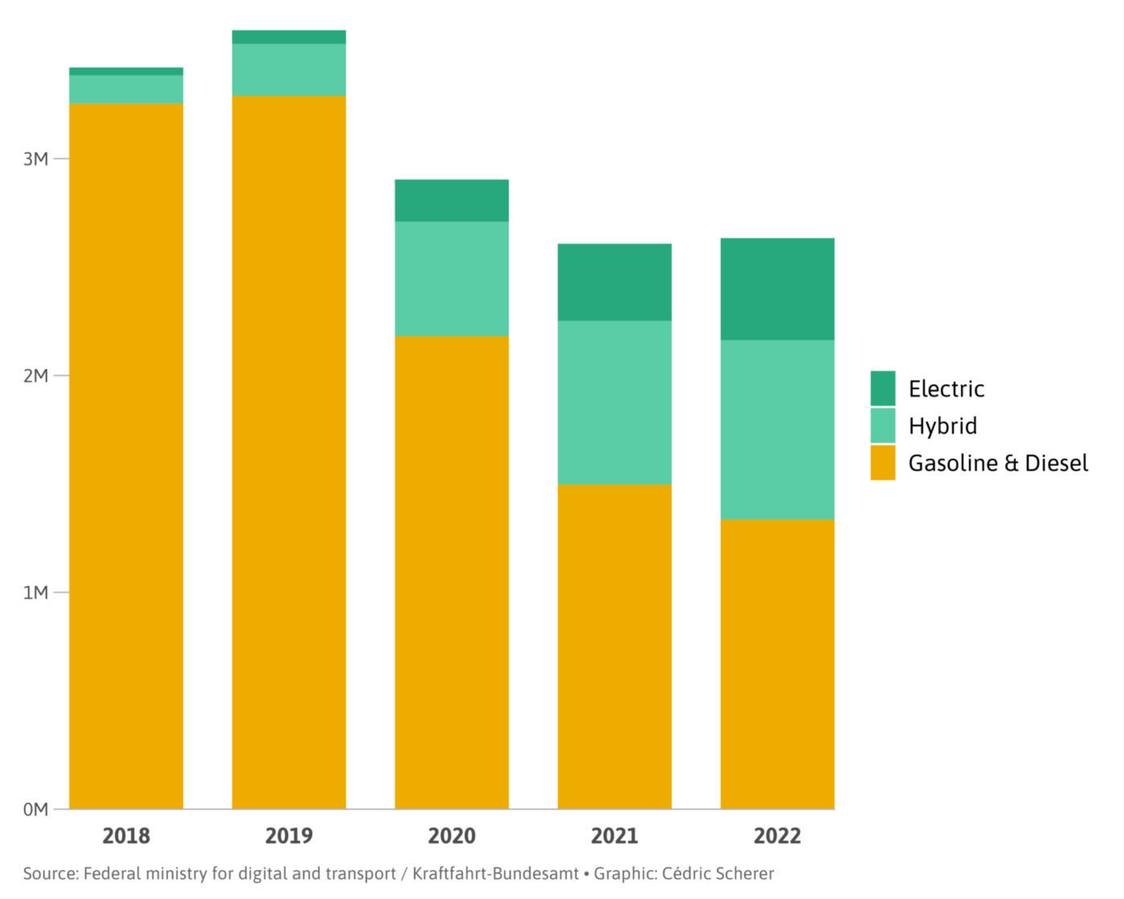
cedricscherer.com @CedScherer z3tt

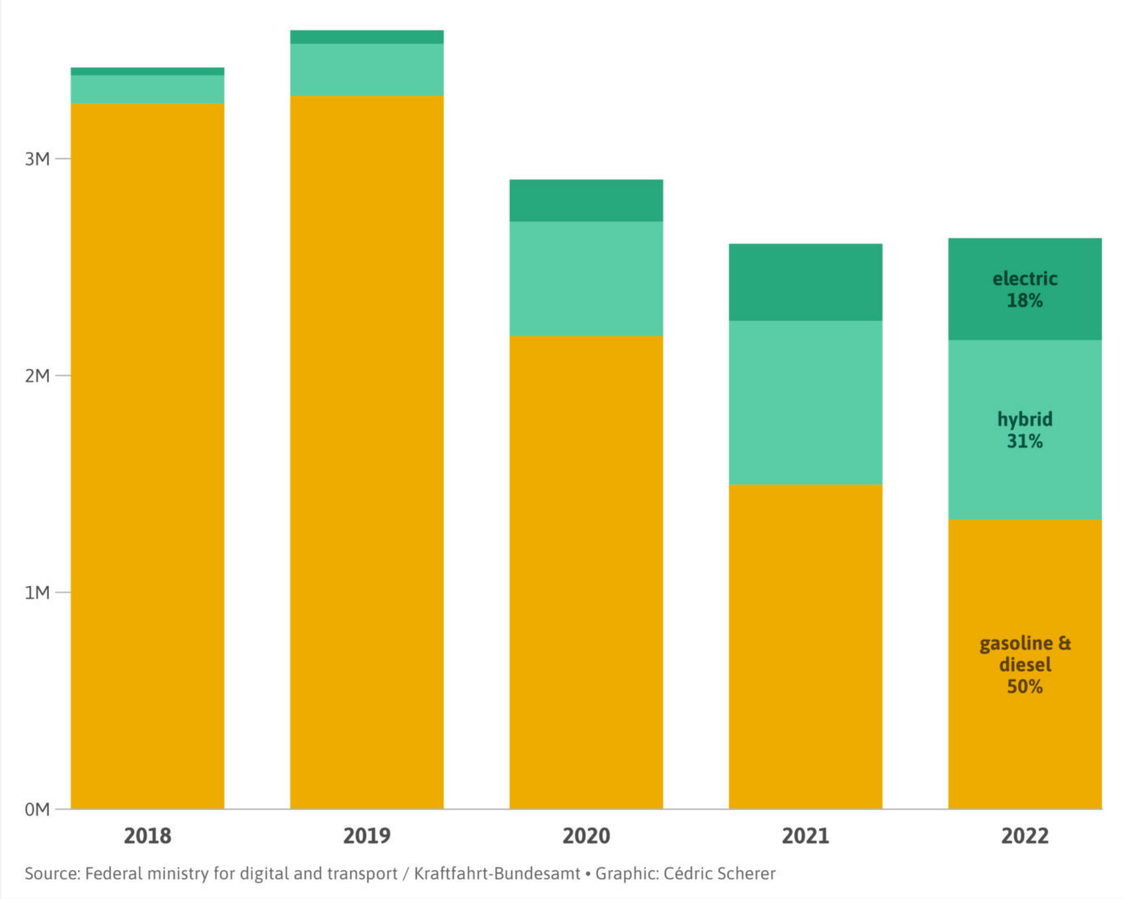




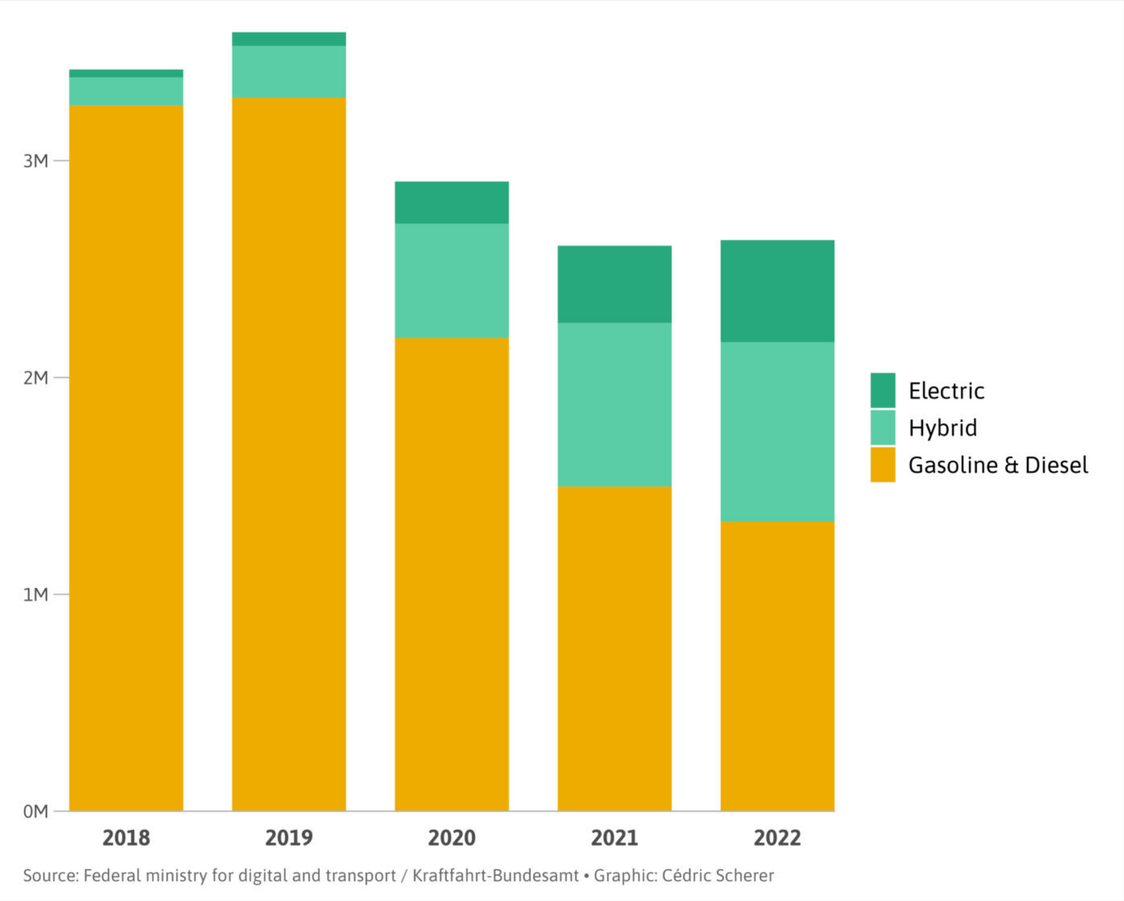
Left: choropleth map by Die Zeit | right: tile grid map by Cédric Scherer & Ansgar Wolsing

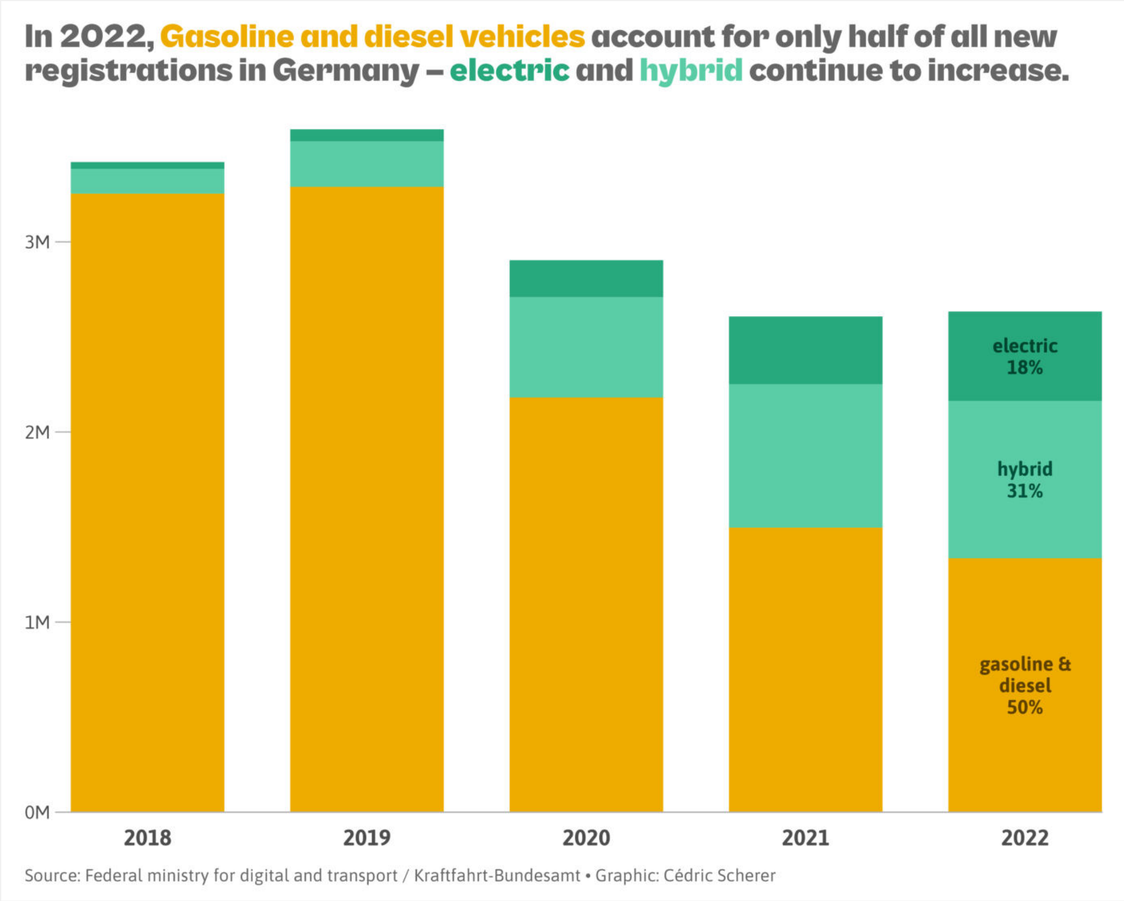
The Power of Annotations



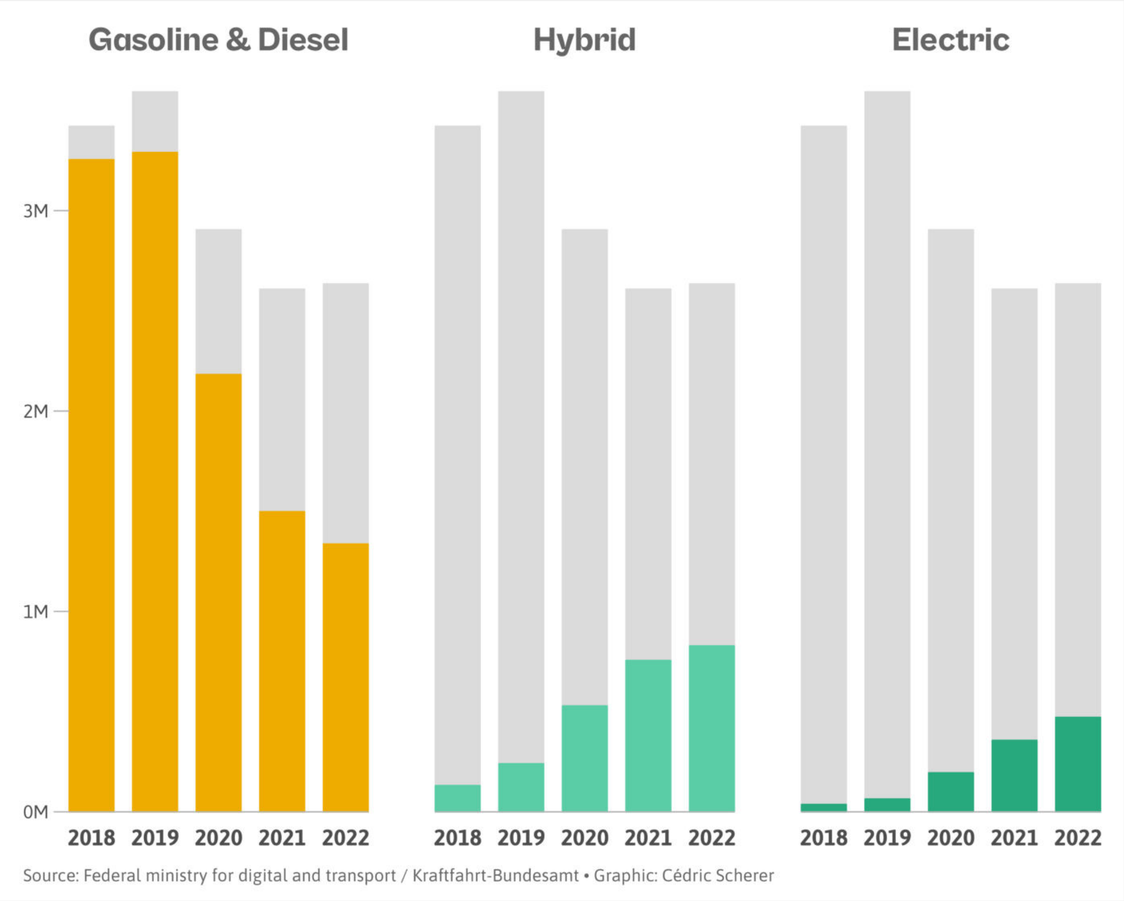


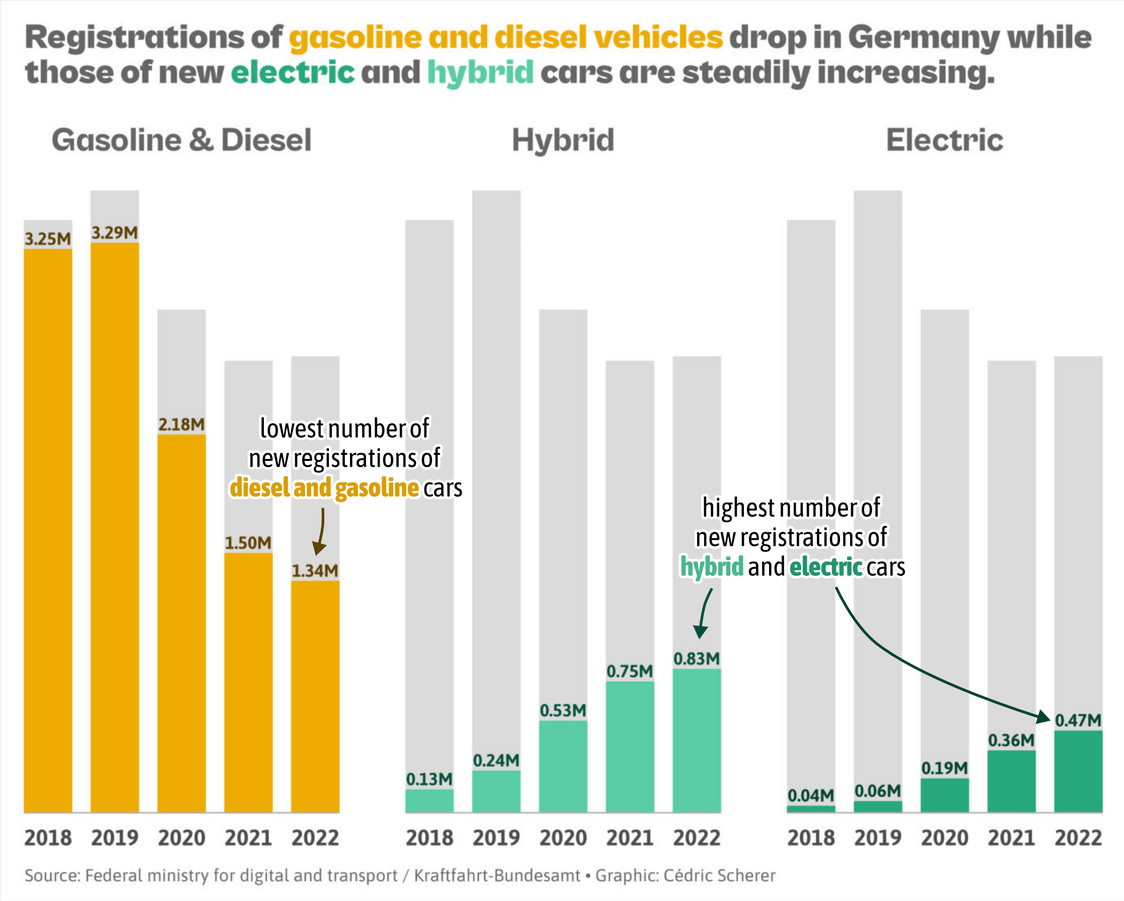
The Power of Annotations



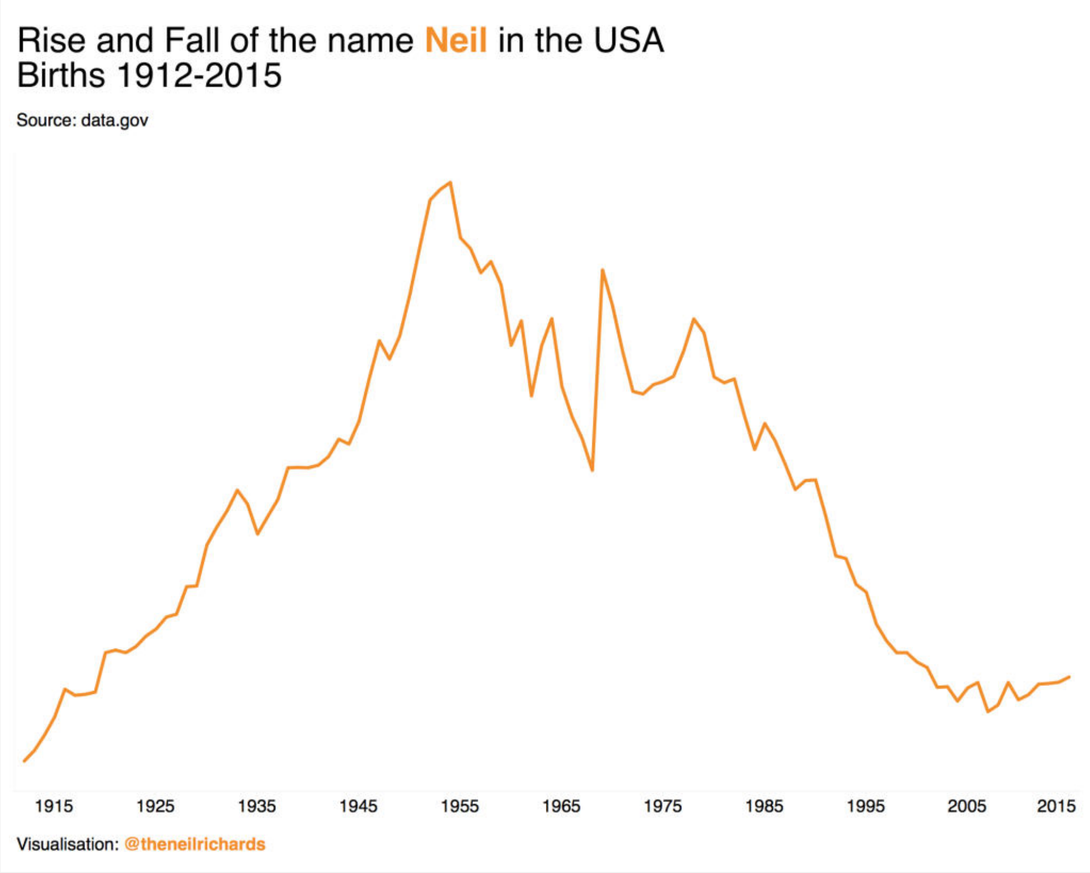


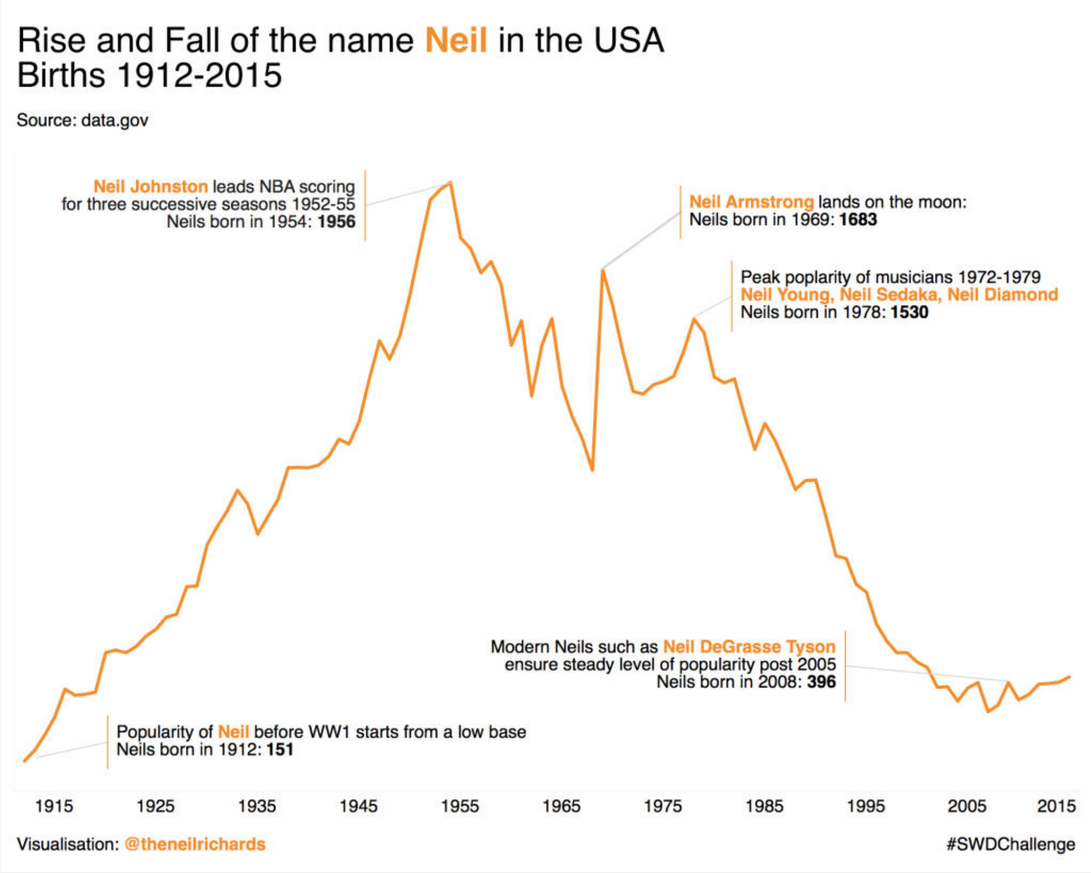
The Power of Annotations



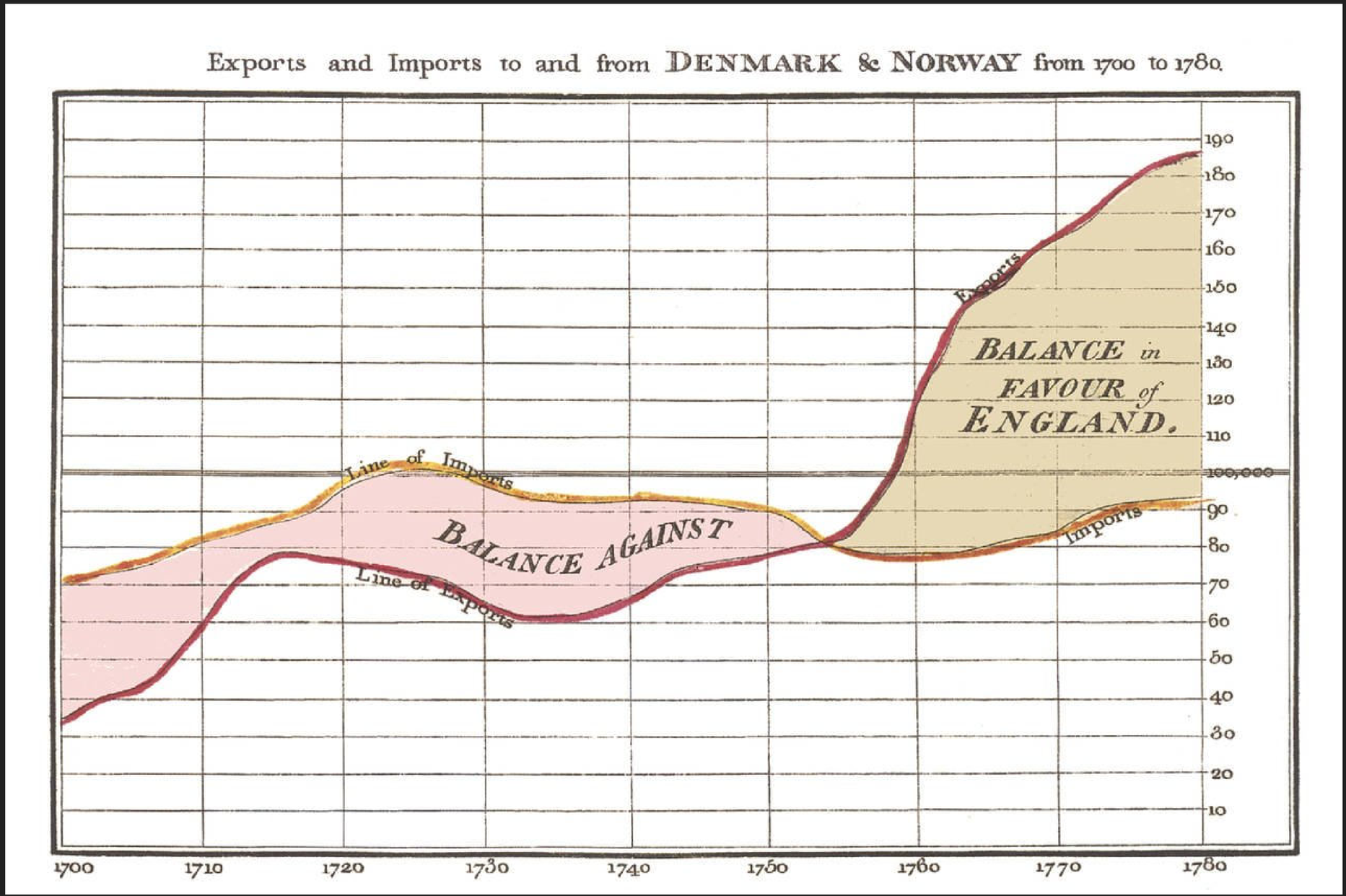


The Power of Annotations





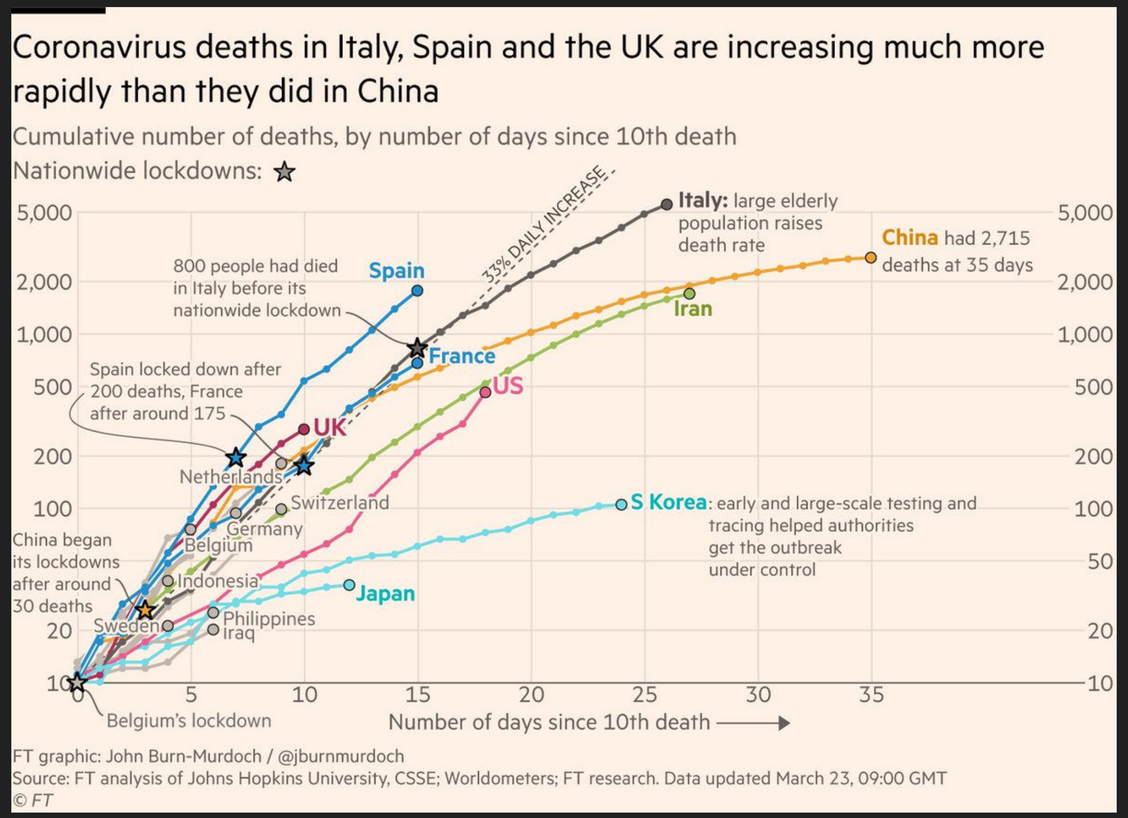
"Is white space always your friend?" by Neil Richards

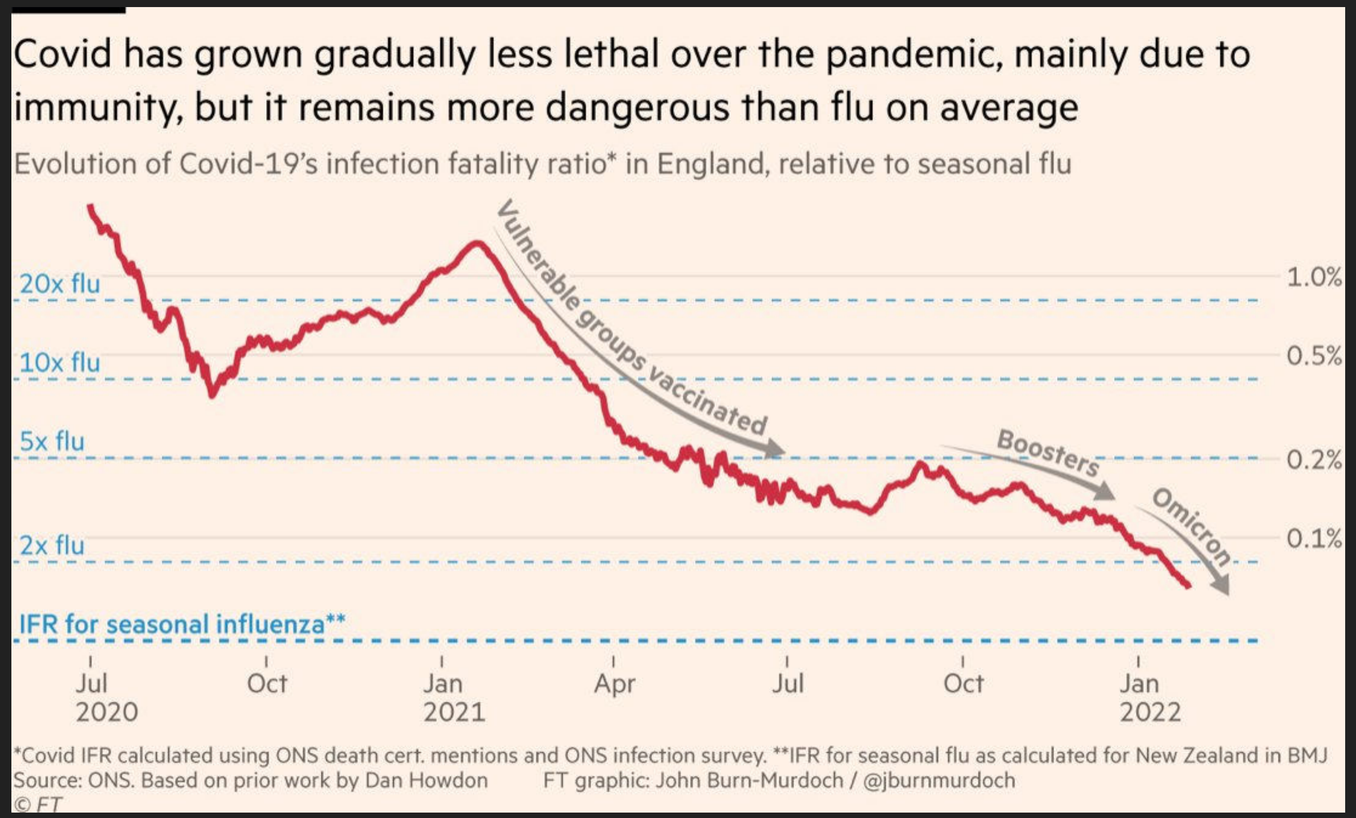


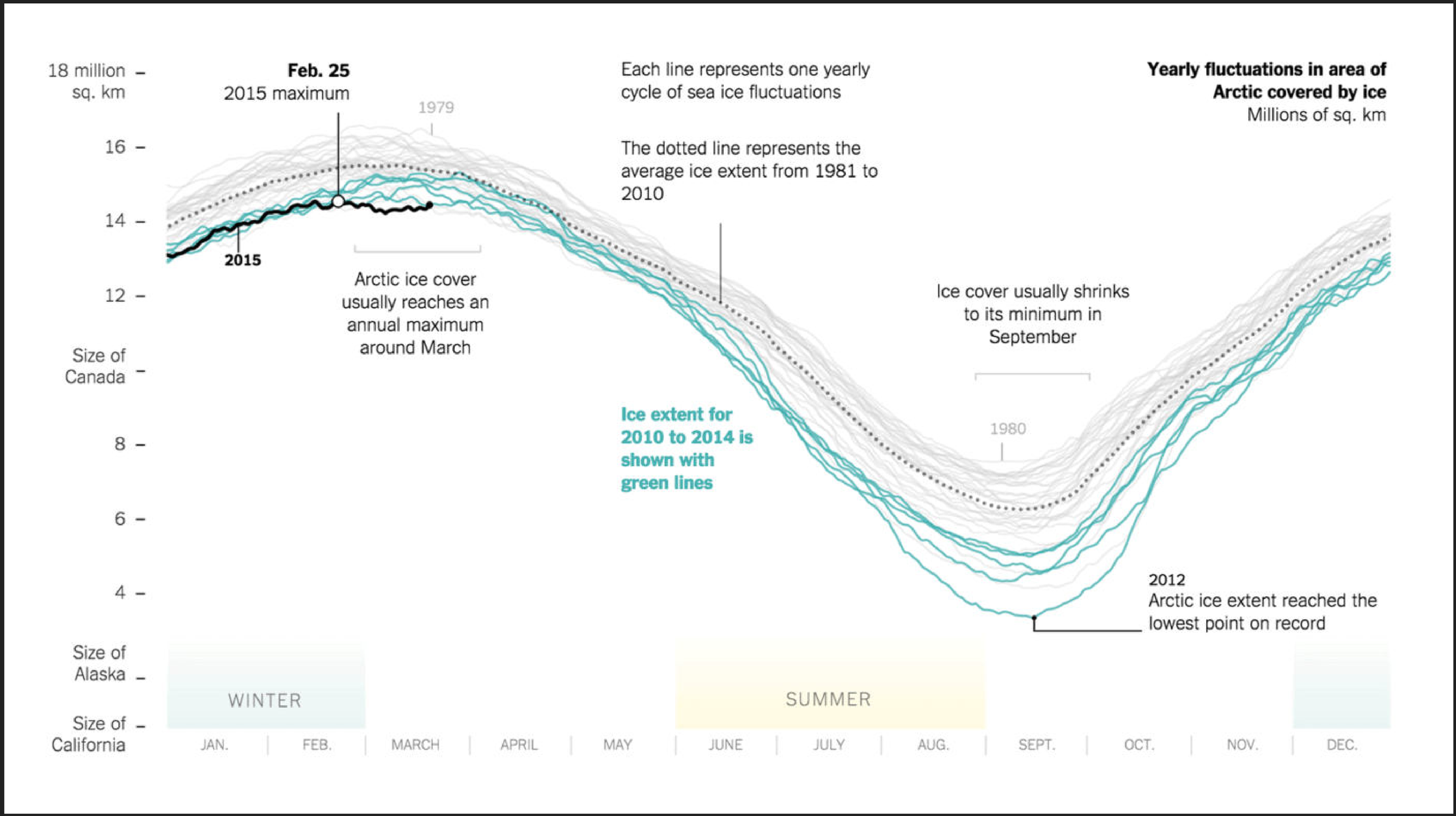
Time series with annotations by William Playfair from "The Commercial and Political Atlas and Statistical Breviary" (1786)

“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.**”

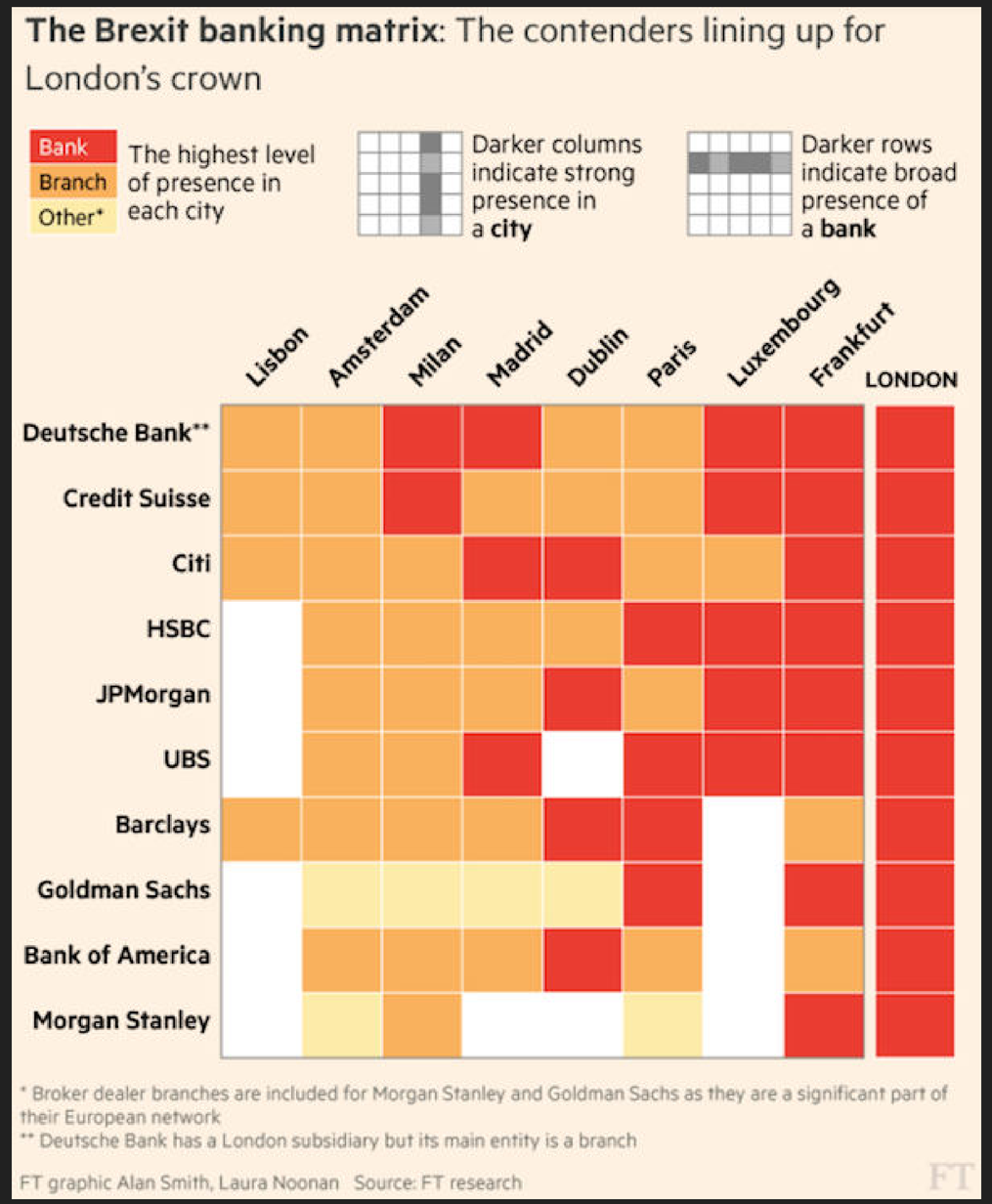
~ John Burn-Murdoch, Financial Times







"Yearly Fluctuations in Area of Arctic Covered by Ice" by Derek Watkins (New York Times)



"Frankfurt vies for UK banking jobs post-Brexit" by Alan Smith und Laura Noonan (Financial Times)

Supplementary supermarket shopping

1 People who do most of their food/grocery shopping at this supermarket...

2 ...also regularly shop at this supermarket

	Aldi	Asda	Iceland	Lidl	Marks & Spencer	Morrisons	Sainsbury's	Tesco	The Co-operative	Waitrose
Aldi		36	26	34	17	33	23	28	22	16
Asda	38		34	32	19	29	20	27	20	11
Iceland	18	22		20	12	19	14	17	12	7
Lidl	25	24	26		13	24	20	24	21	17
Marks & Spencer	9	12	9	9		17	28	17	17	37
Morrisons	27	26	21	25	22		20	22	20	13
Sainsbury's	26	25	20	33	47	29		32	31	39
Tesco	45	39	42	44	43	37	42		42	40
The Co-operative	15	15	11	17	18	17	20	19		20
Waitrose	5	5	6	9	30	7	21	11	14	

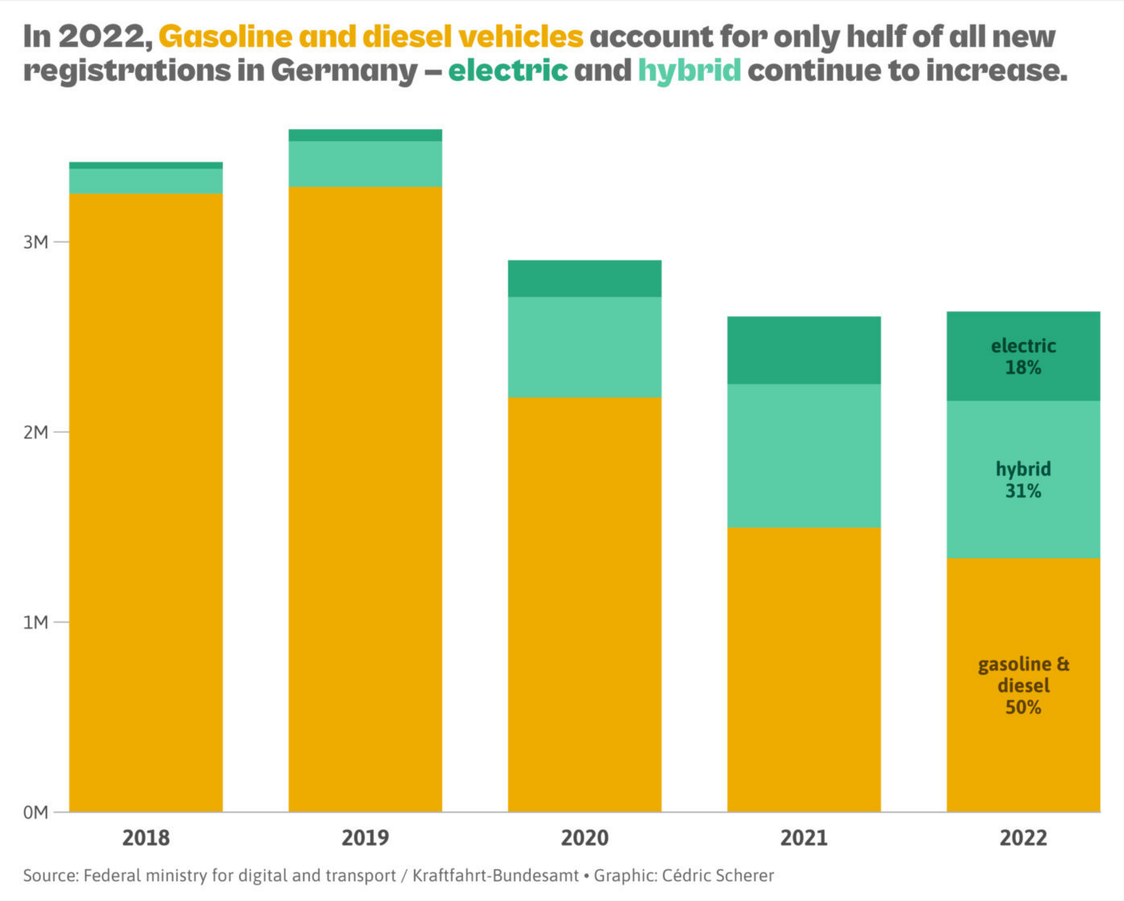
So, for instance, 27% of people who do most of their grocery shopping at Aldi also regularly shop at Morrisons

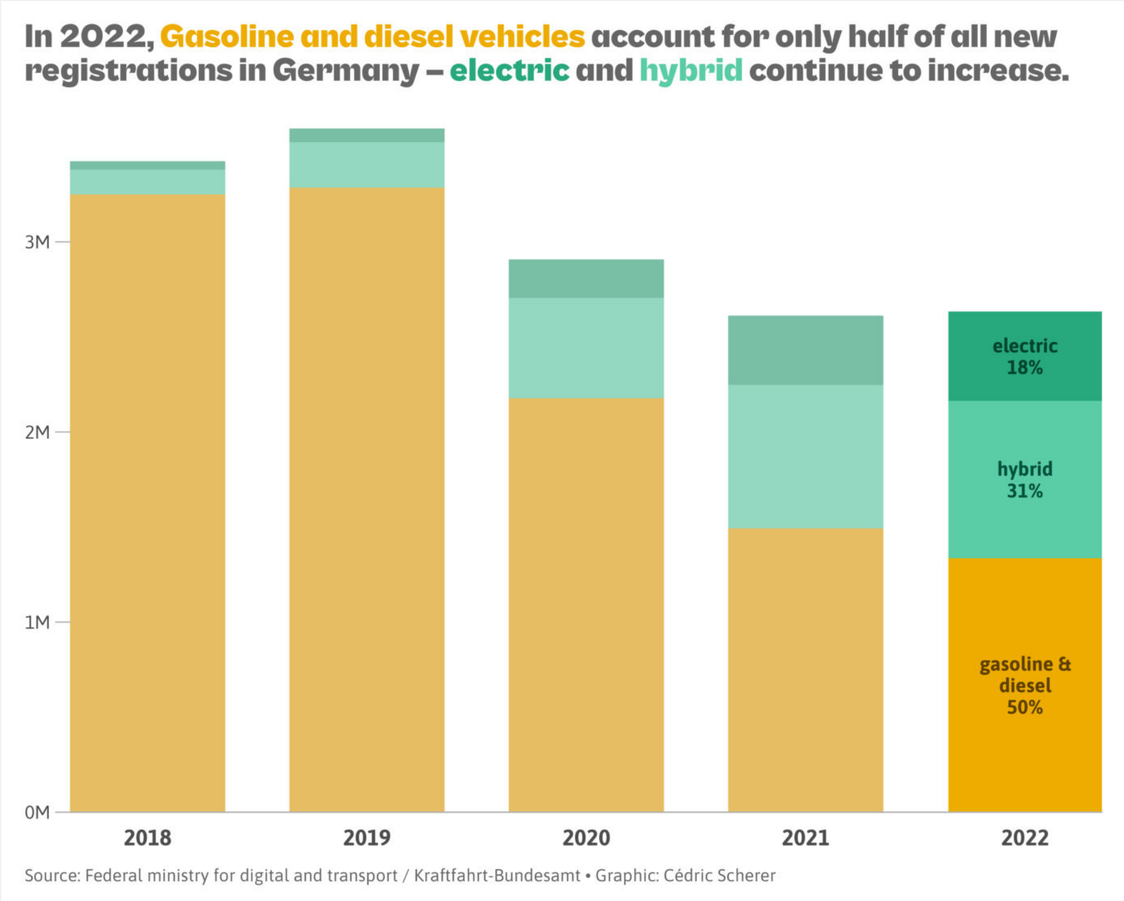
YouGov | yougov.com

YouGov Profiles December 2018

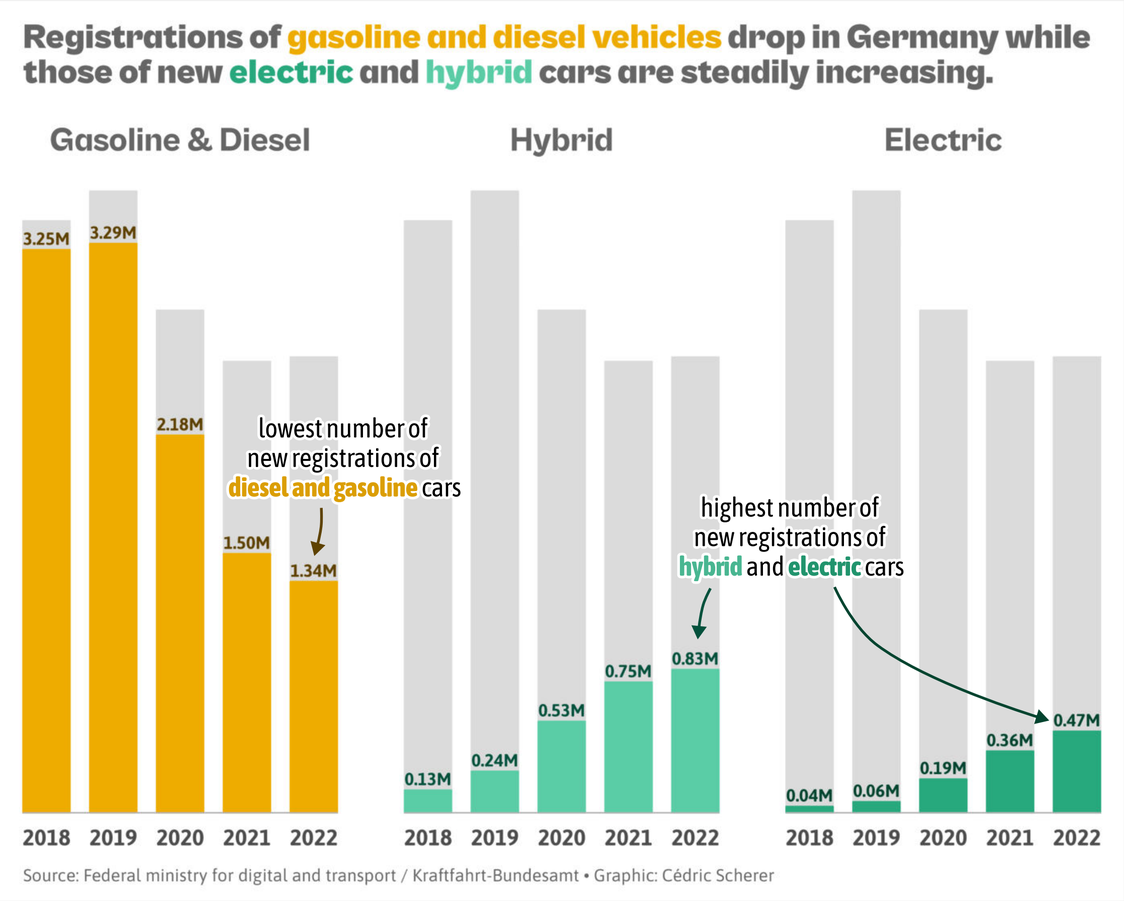
"Tesco is the nation's primary AND secondary supermarket" by Matthew Smith (YouGov)

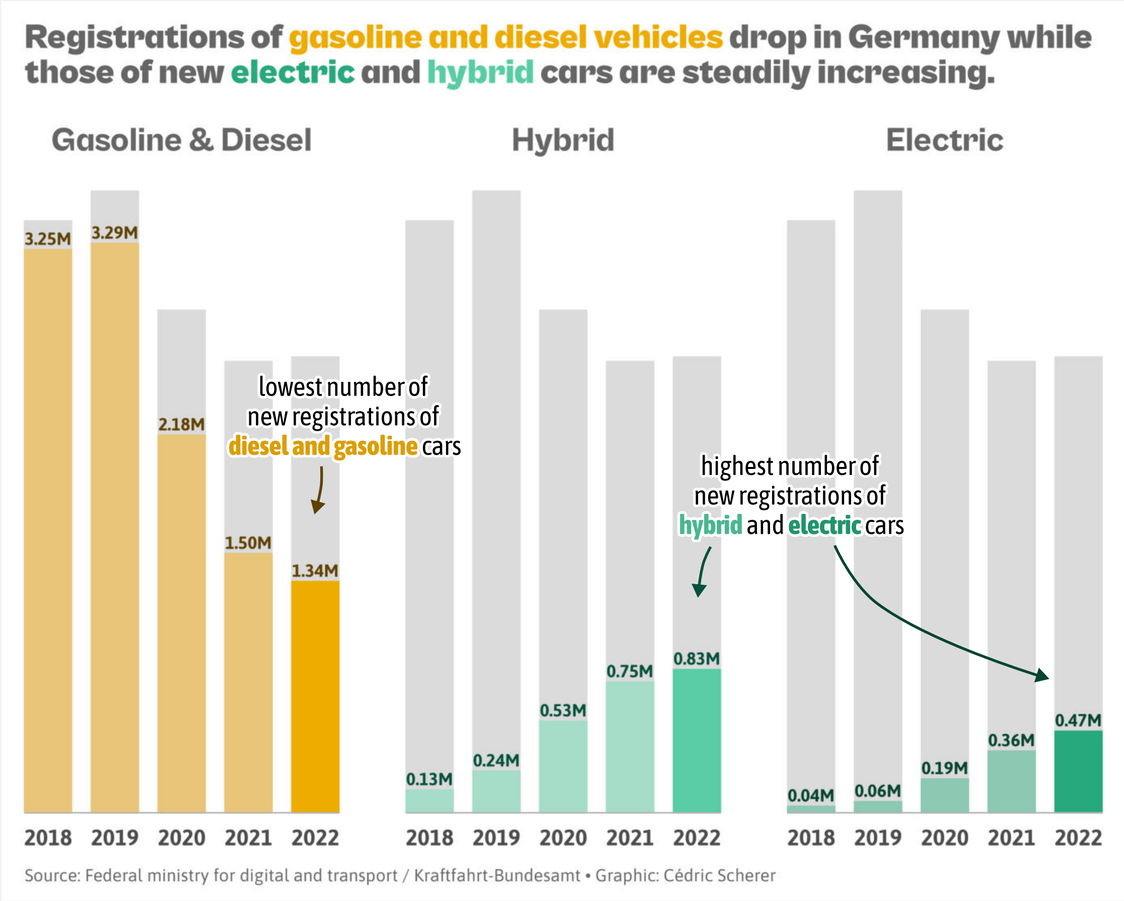
The Power of Colors



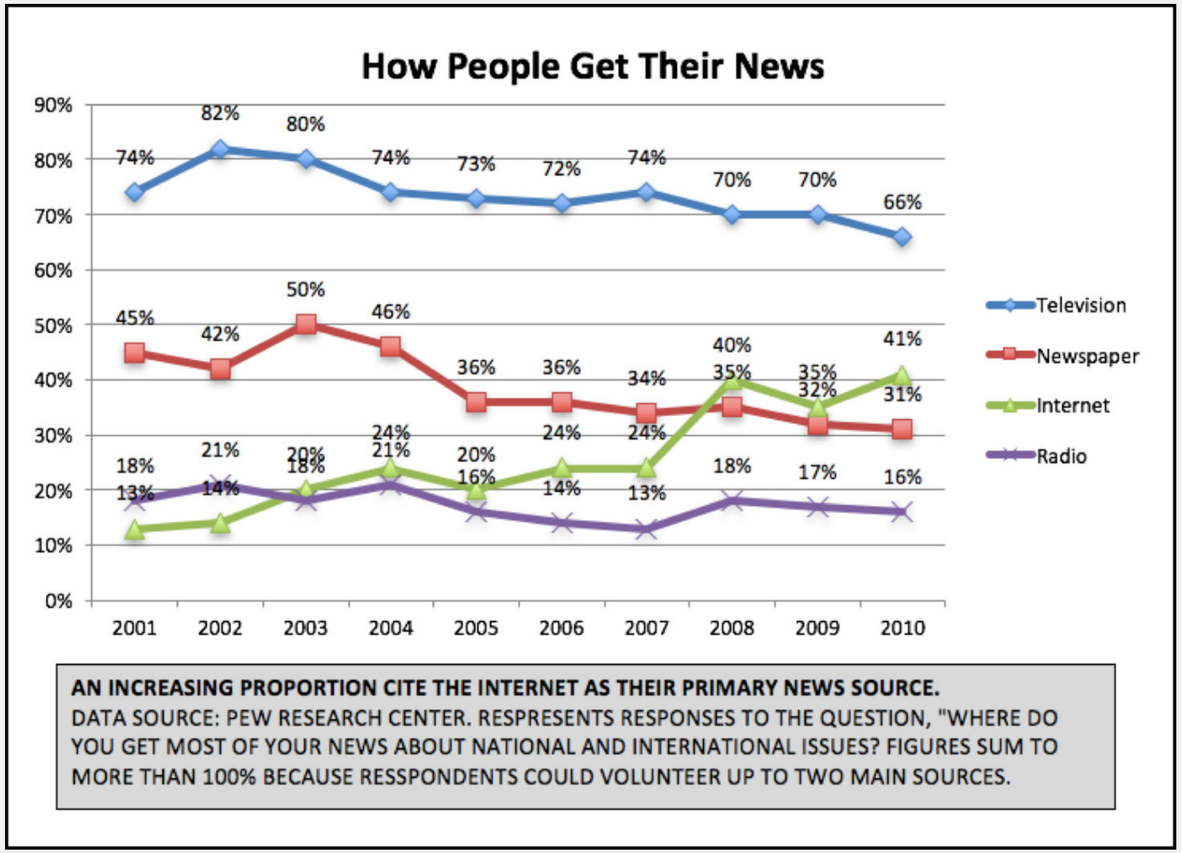


The Power of Colors





Group Exercise



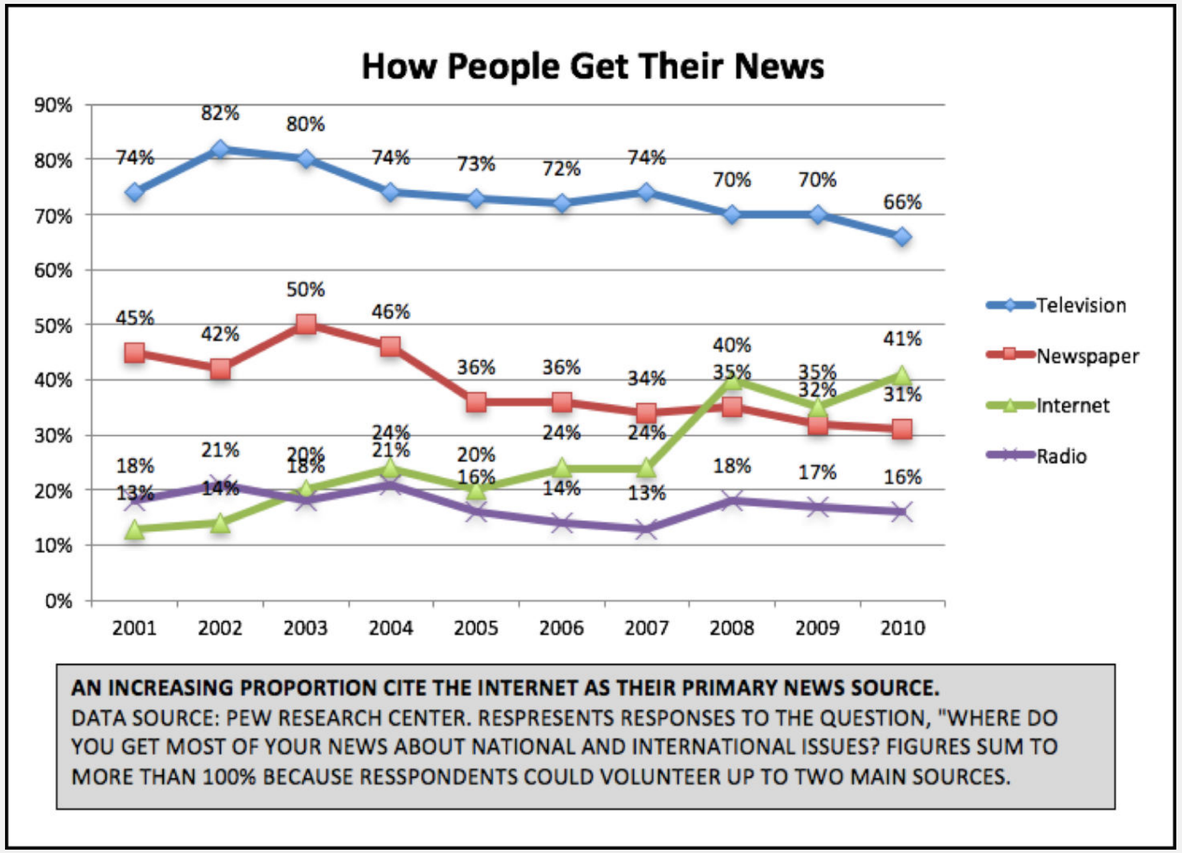
Source: StorytellingWithData

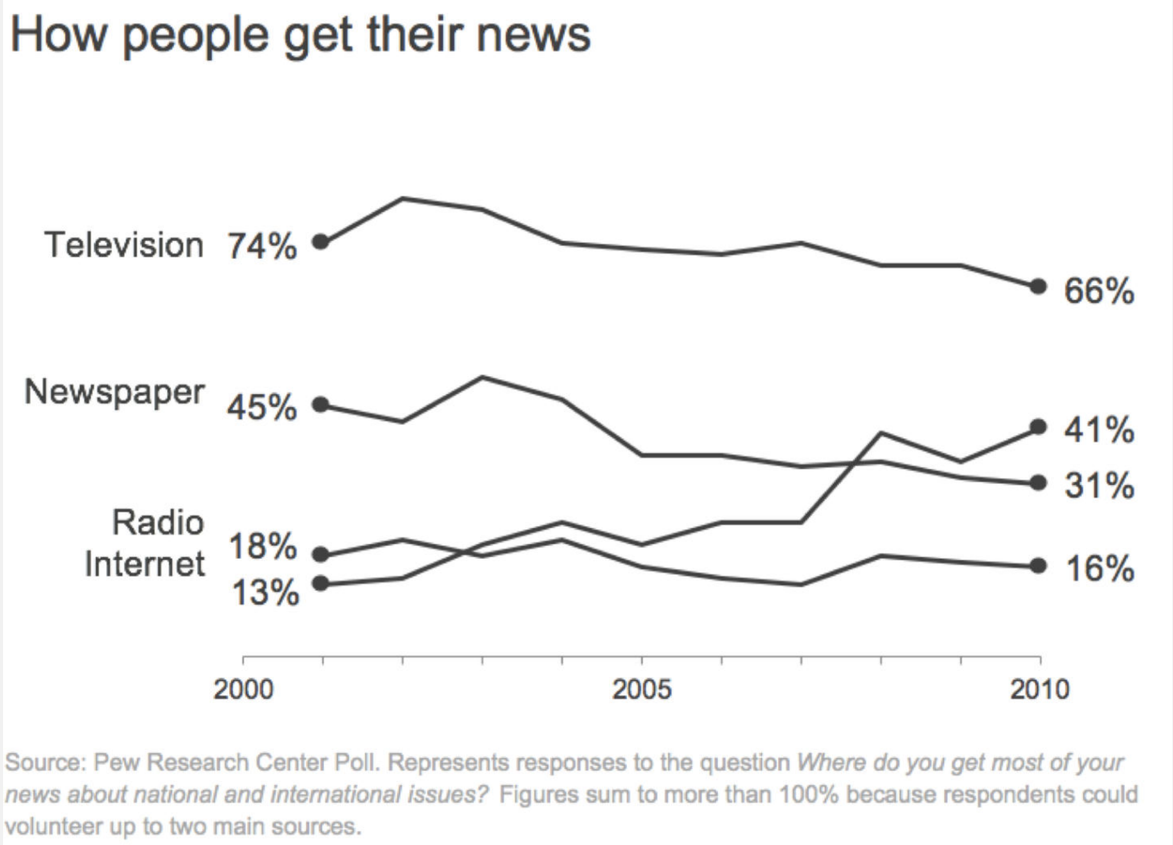
Declutter this graphic

- ☛ What is the main message?
- ☛ Is all the data relevant?
- ☛ What would you change?
- ☛ How can you highlight the **increase**?

- ☛ **Draw a decluttered version!**

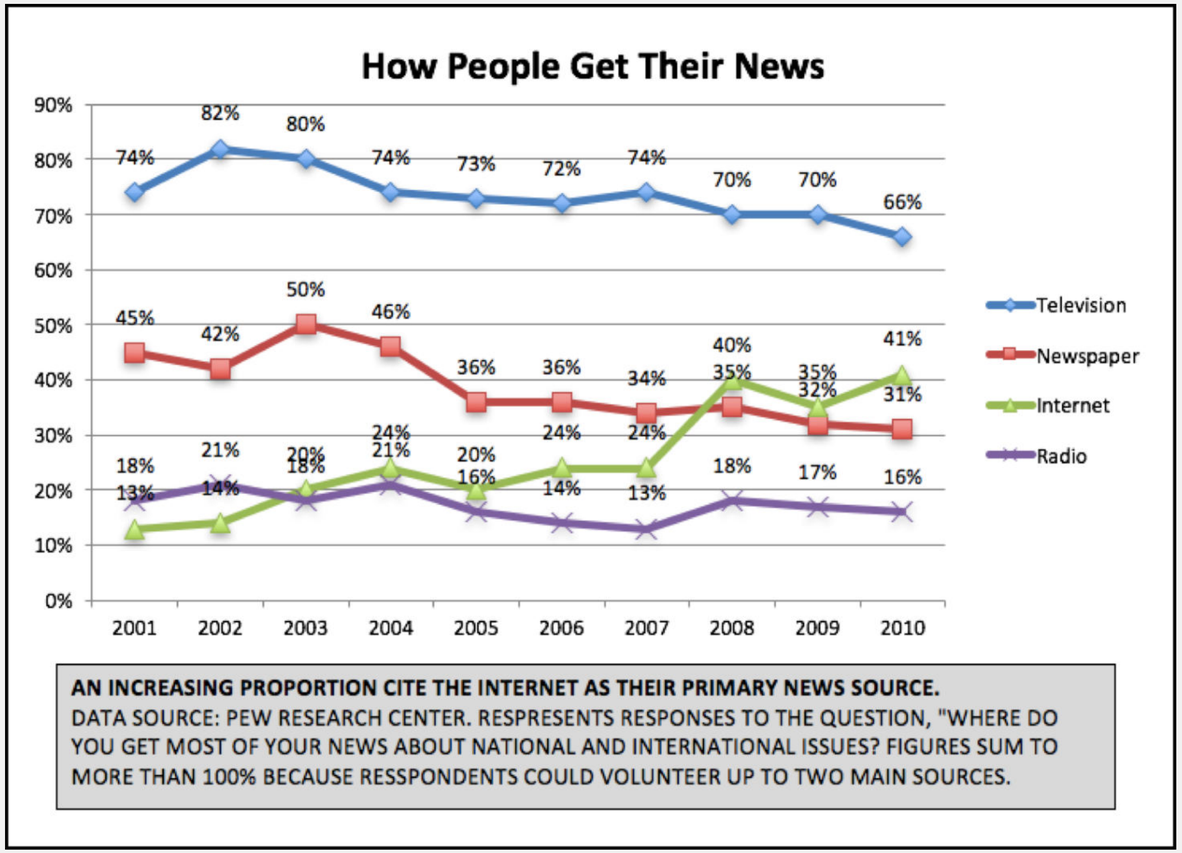
Group Exercise

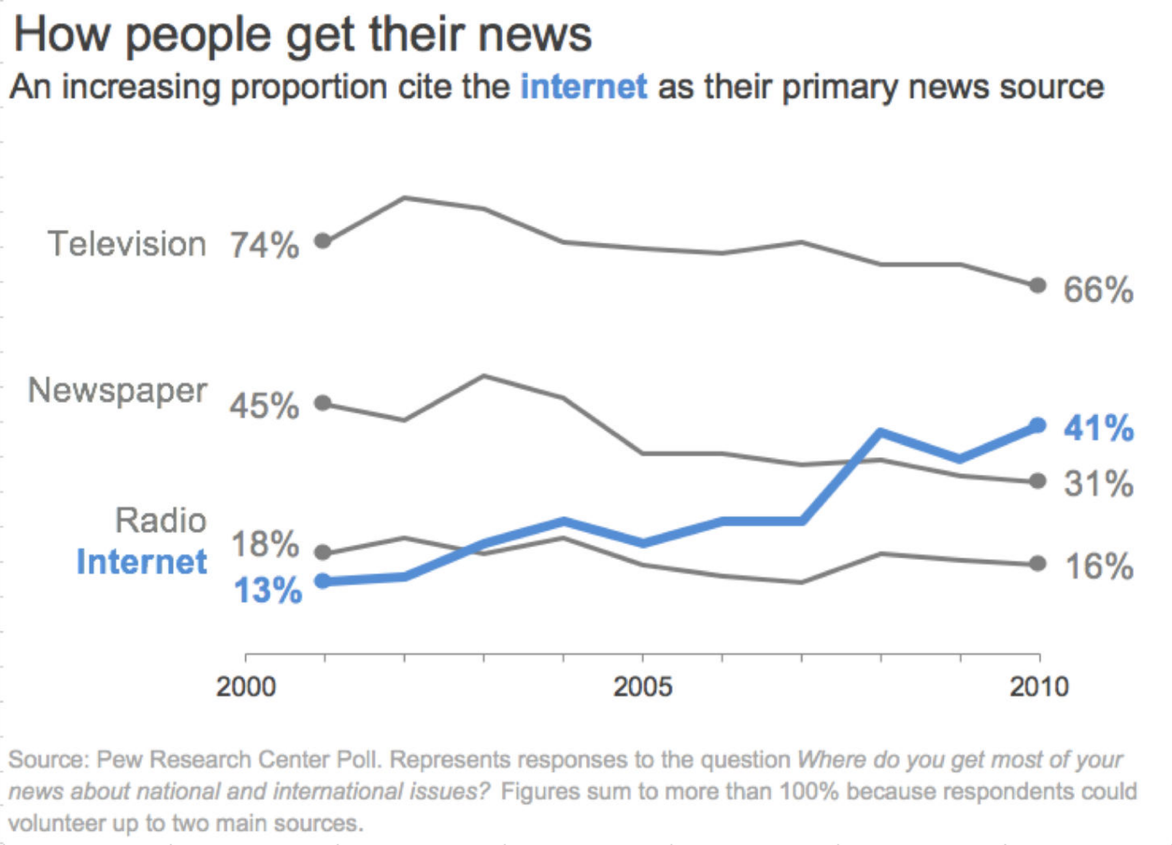




Source: StorytellingWithData

Group Exercise



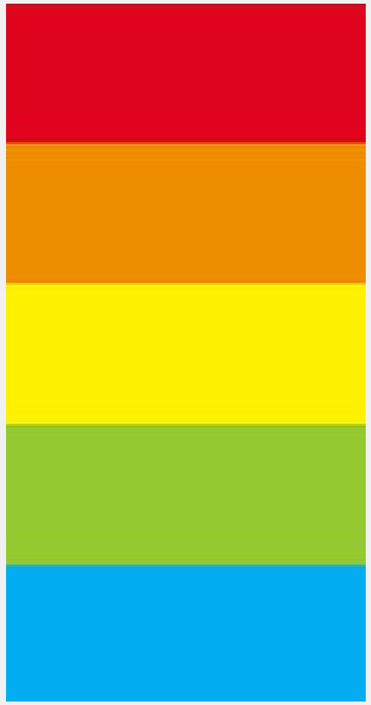


Source: StorytellingWithData

COLORS

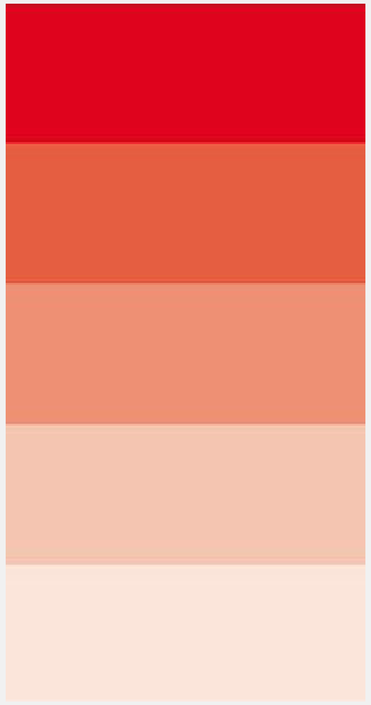
and Pitfalls

Hue
color family



Chroma
colorfulness

Luminance
lightness / brightness



Non Uniform Distances

Med Phys. 2015 Jun; 42(6): 2942–2954. Published online 2015 May 20. doi: 10.1118/1.4921125

PMCID: PMC5148121 | PMID: 26127048

Effect of color visualization and display hardware on the visual assessment of pseudocolor medical images

Silvina Zabala-Travers, Mina Choi, Wei-Chung Cheng, and Aldo Badano[a)]

IEEE Computer Graphics and Applications

Rainbow Color Map (Still) Considered Harmful

March/April 2007, pp. 14-17, vol. 27

DOI Bookmark: 10.1109/MCG.2007.46

Authors

David Borland, University of North Carolina at Chapel Hill

Russell M. Taylor II, University of North Carolina at Chapel Hill

10 March 2017

Interpretation of the rainbow color scale for quantitative medical imaging: perceptually linear color calibration (CSDF) versus DICOM GSDF

Frédérique Chesterman, Hannah Manssens, Céline Morel, Guillaume Serrell, Bastian Piepers, Tom Kimpe

Author Affiliations +

Proceedings Volume 10136, Medical Imaging 2017: Image Perception, Observer Performance, and Technology Assessment; 101360R (2017) https://doi.org/10.1117/12.2253885

Event: SPIE Medical Imaging, 2017, Orlando, Florida, United States

Education and communication

Rainbow color map distorts and misleads research in hydrology – guidance for better visualizations and science communication

Michael Stoelzle[1] and Lina Stein[2]

[1]Faculty of Environment and Natural Resources, University of Freiburg, Freiburg, Germany

[2]Department of Civil Engineering, University of Bristol, Bristol, UK

Med Phys. 2015 Jun; 42(6): 2942–2954. Published online 2015 May 20. doi: 10.1118/1.4921125

PMCID: PMC5148121 | PMID: 26127048

Effect of color visualization and display hardware on the visual assessment of pseudocolor medical images

Silvina Zabala-Travers, Mina Choi, Wei-Chung Cheng, and Aldo Badano[a)]

IEEE Computer Graphics and Applications

Rainbow Color Map (Still) Considered Harmful

March/April 2007, pp. 14-17, vol. 27

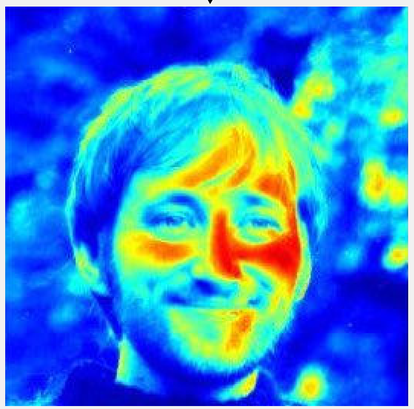
DOI Bookmark: 10.1109/MCG.2007.46

Authors

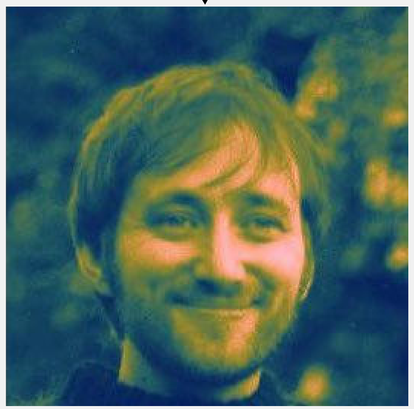
"The ad hoc manner in which color is handled and the lack of standard approaches have been associated with suboptimal and **inconsistent diagnostic decisions with a negative impact on patient treatment and prognosis**."

Zabala-Travers, Choi, Cheng & Badano 2015 Med Phys.

true-color Phil



rainbow Phil



batlow Phil

Mdofied example from Fabio Crameri

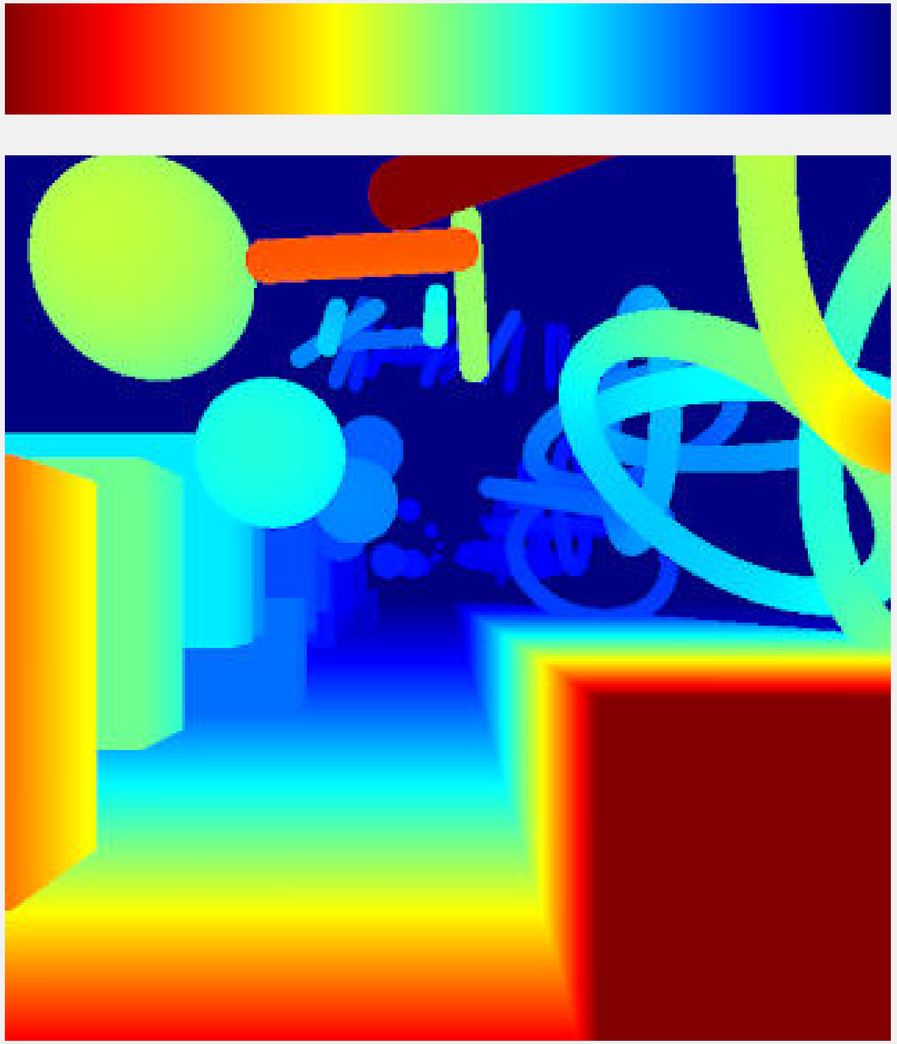
Turbo

Rainbow

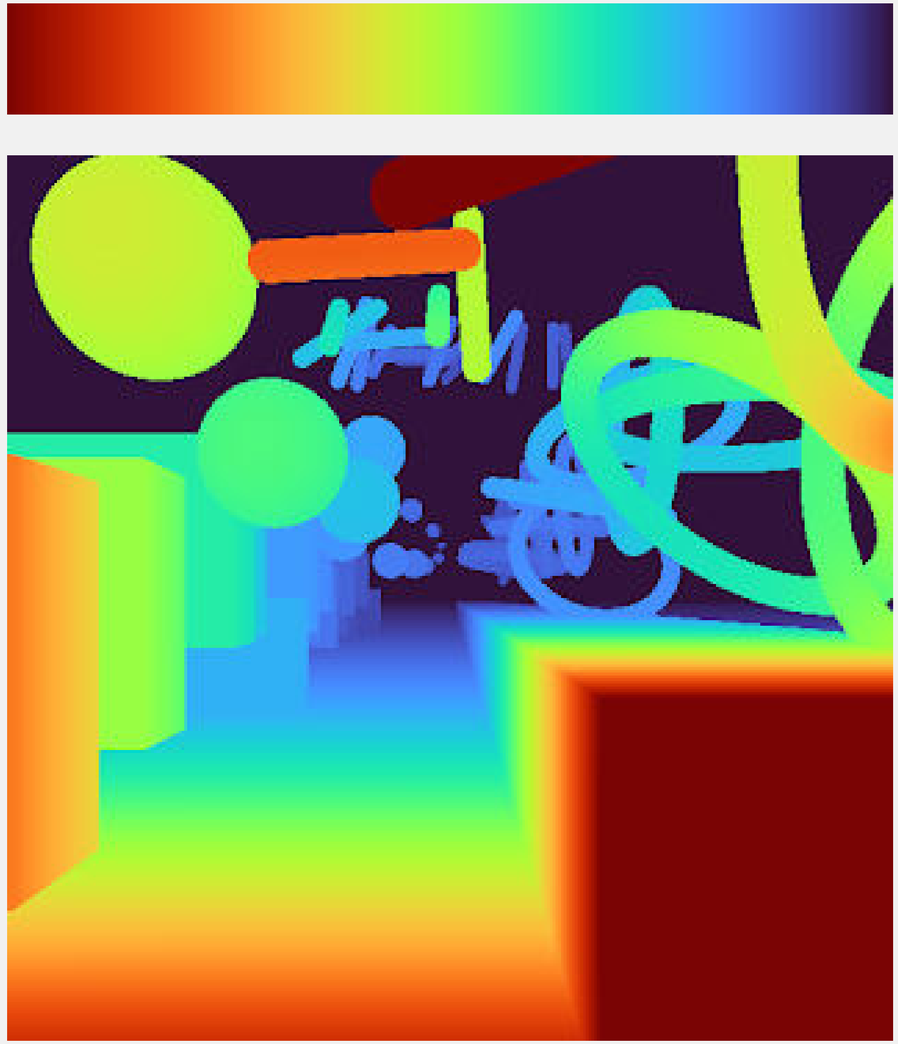
Turbo

Jet

Jet



Turbo



"Turbo, An Improved Rainbow Colormap for Visualization" by Anton Mikhailov (Google Research)

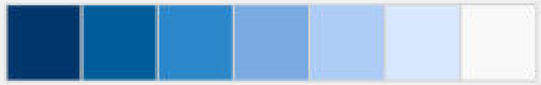
modification with examples	luminance variation	false boundaries	color vision deficiency sensitive
isoluminant rainbow (e.g. Brewer, 1997; Kindlmann et al., 2002; Rogowitz & Kalvin, 2001; Szafir et al., 2016a)	addressed		
luminance and chromaticity correction (Sisneros et al., 2016)	addressed	addressed	
adjusting color distribution to fit it to data (Wang et al., 2021)		addressed	
Turbo (Mikhailov, 2019)		addressed	addressed

Gołębiowska & Çöltekin (2022) ISPRS Journal of Photogrammetry and Remote Sensing 194

Color Palette Choice

Sequential

Palette



Desaturated

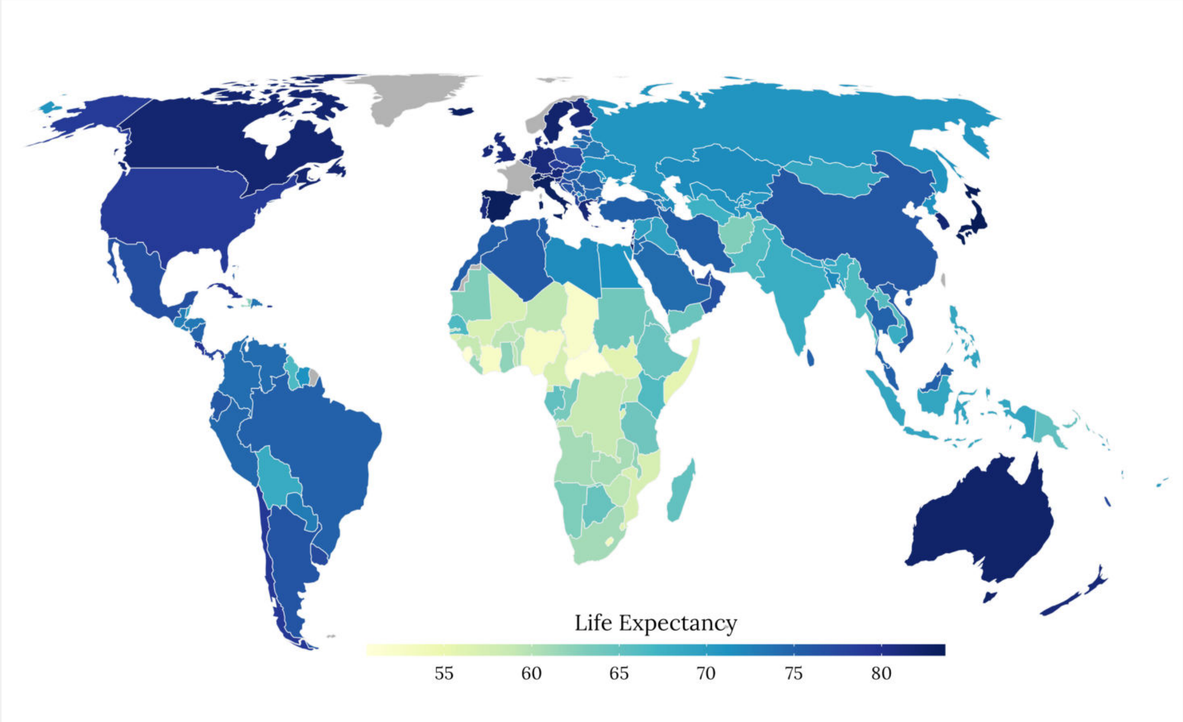
Use to encode
numerical information with order

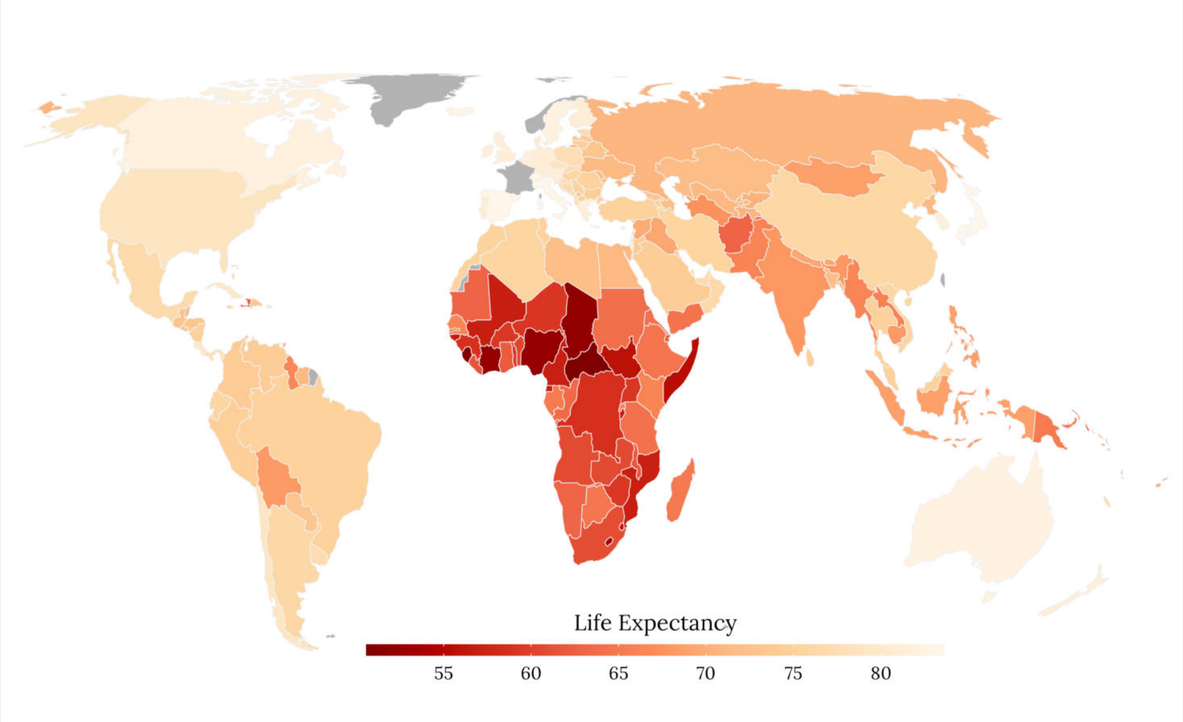
use highest contrast for most important information

either single- or multi-hue

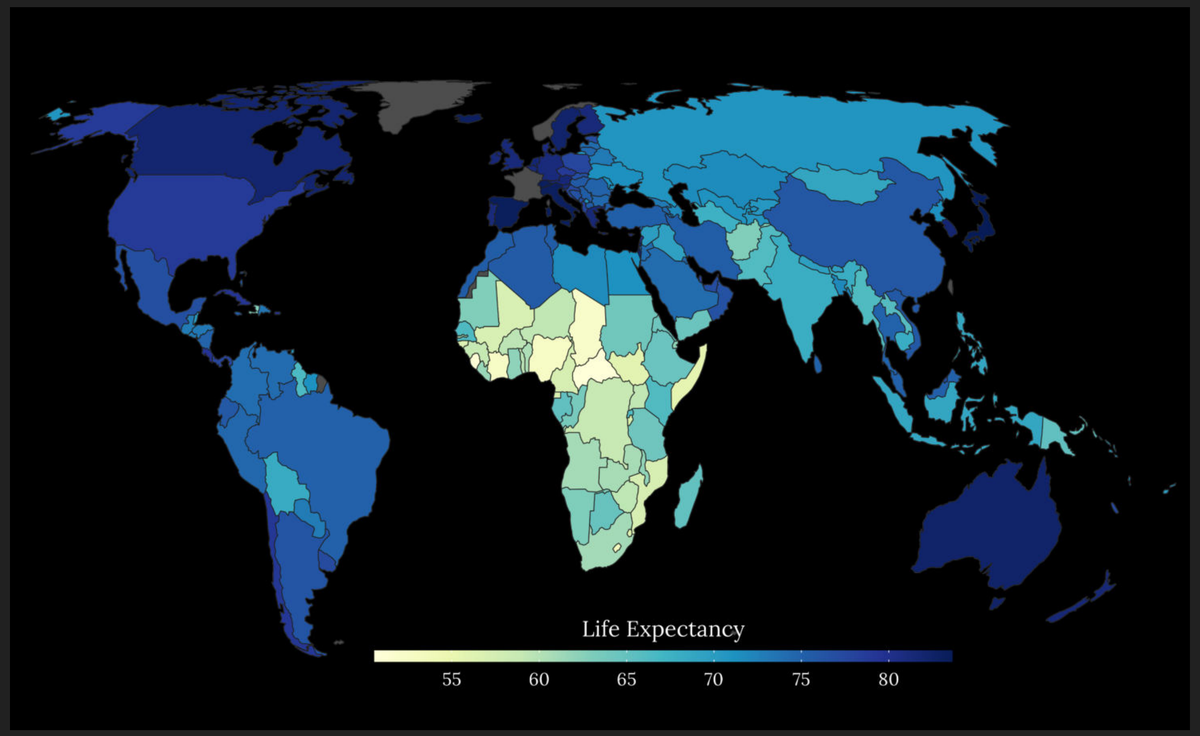
Modified from the {colorspace} R package vignette

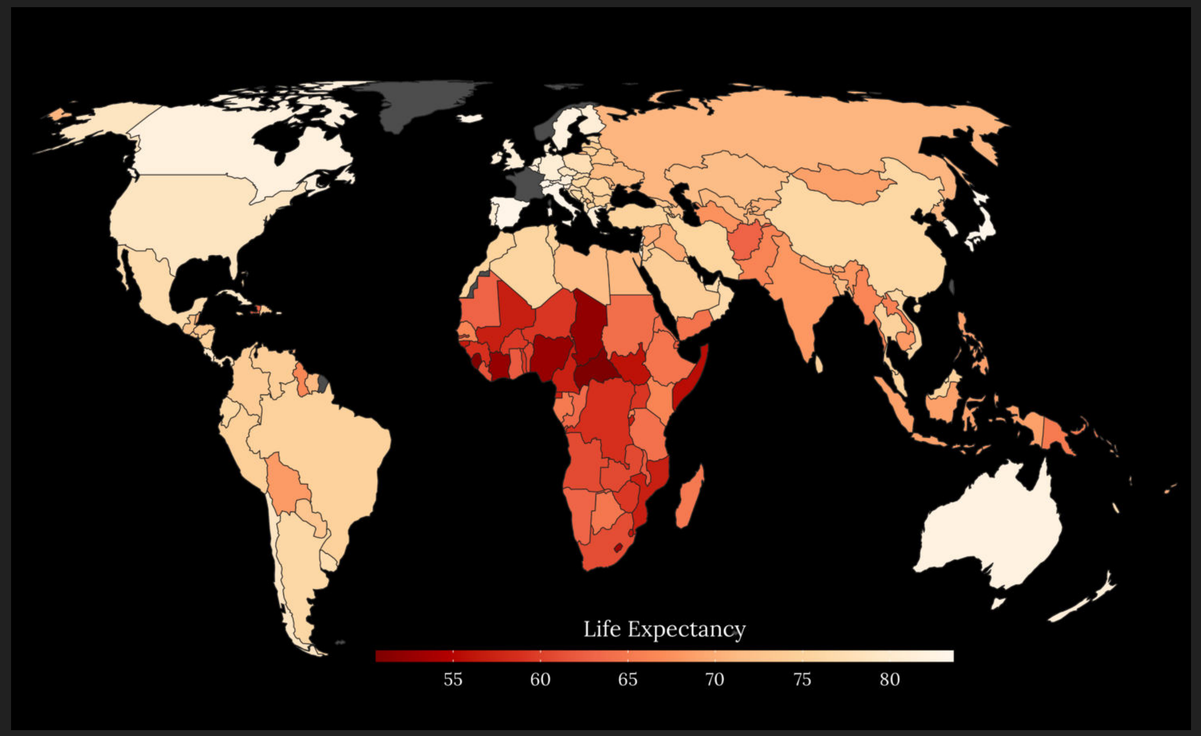
Dark Equals More?





Dark Equals More?

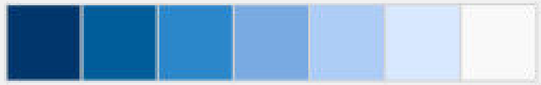




Color Palette Choice

Sequential

Palette



Desaturated

Use to encode
numerical information with order

use highest contrast for most important information

either single- or multi-hue

Diverging

Palette

Desaturated

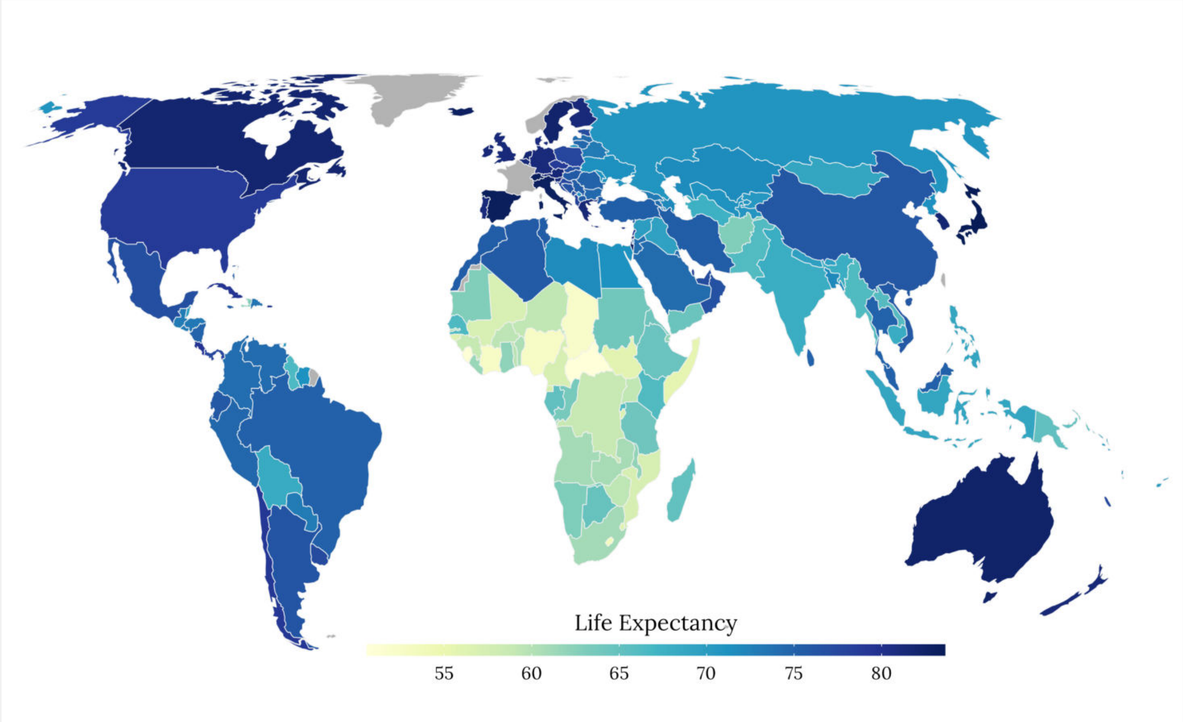
Use to encode
numerical information with critical midpoint

ensure a meaningful midpoint value and use balanced extremes

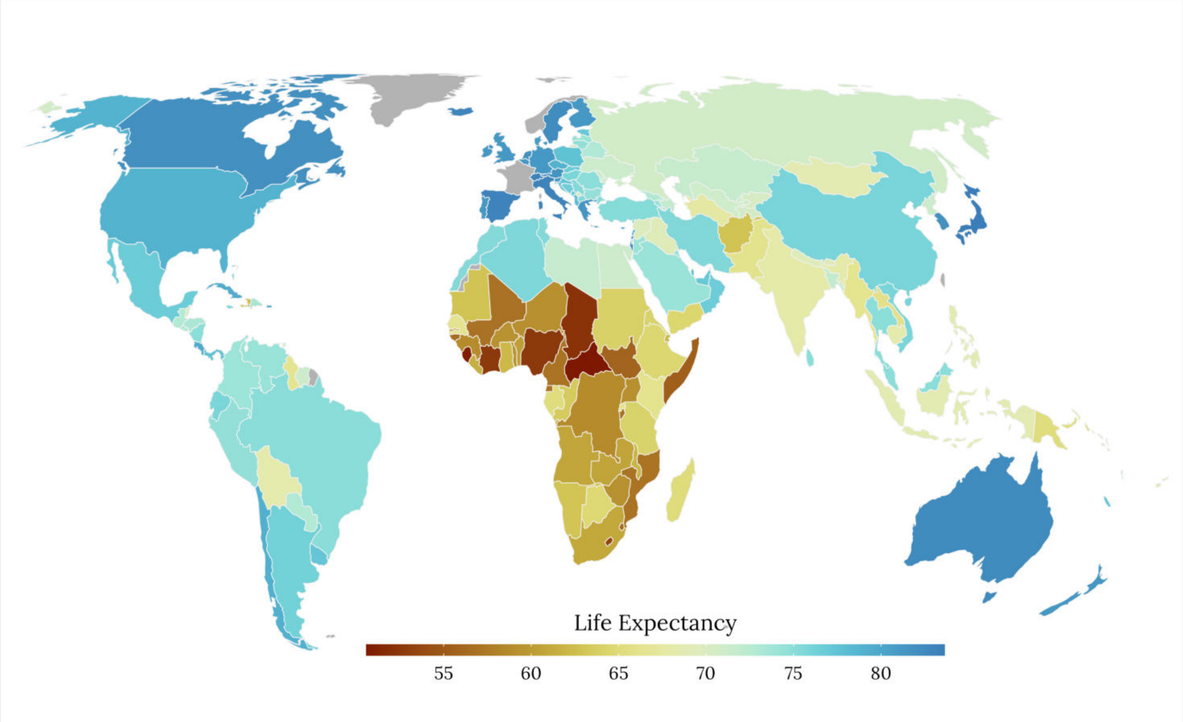
combination of two sequential palettes

Modified from the {colorspace} R package vignette

Sequential or Diverging?



multi-hue sequential



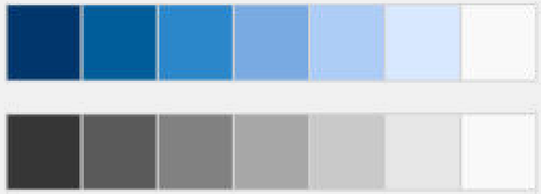
diverging

Color Palette Choice

Sequential

Palette

Desaturated



Use to encode
numerical information with order

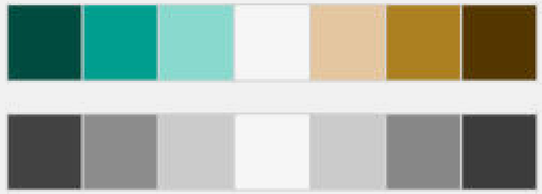
use highest contrast for most important information

either single- or multi-hue

Diverging

Palette

Desaturated



Use to encode
numerical information with critical midpoint

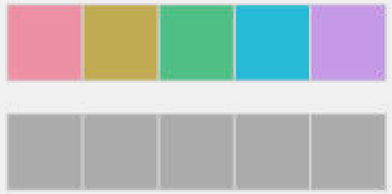
ensure a meaningful midpoint value and use balanced extremes

combination of two sequential palettes

Qualitative

Palette

Desaturated



Use to encode
categorical information

pick distinct colors with the same perceptual weight

limit categories to 6-8

Modified from the {colorspace} R package vignette

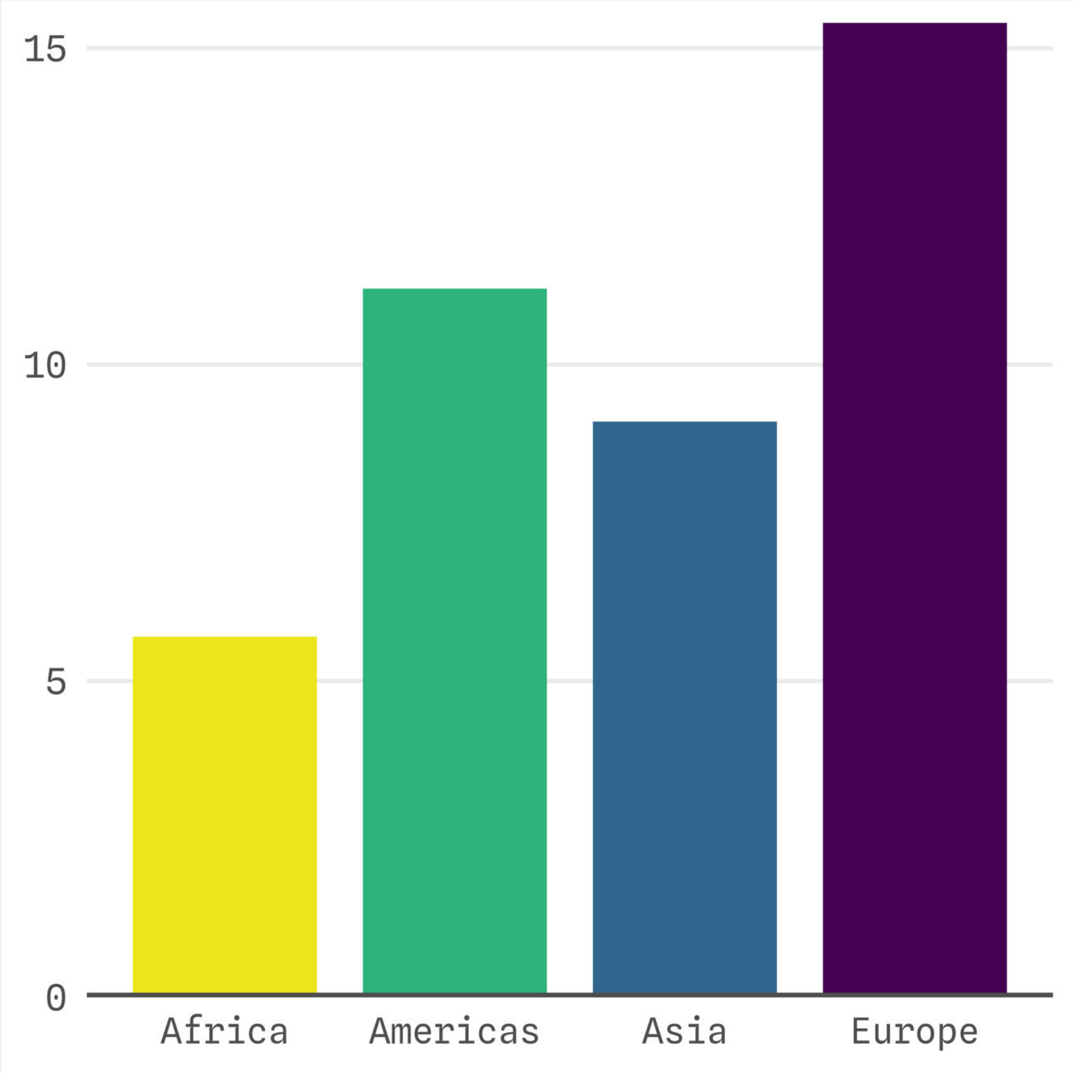
Sequential or Qualitative?

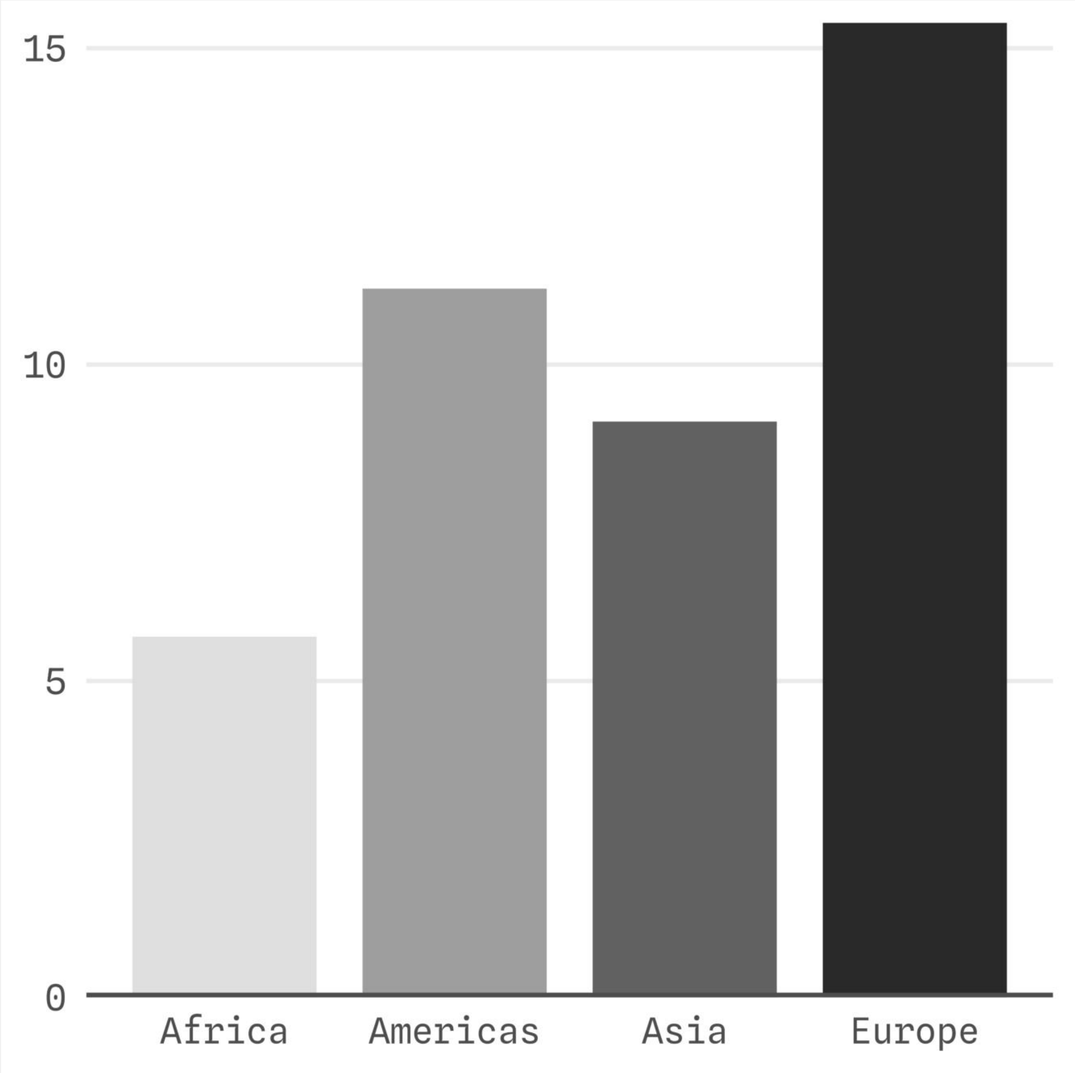
Source: "When to use sequential and when to use diverging color scales" by Lisa C. Muth / DataWrapper

Sequential or Qualitative?

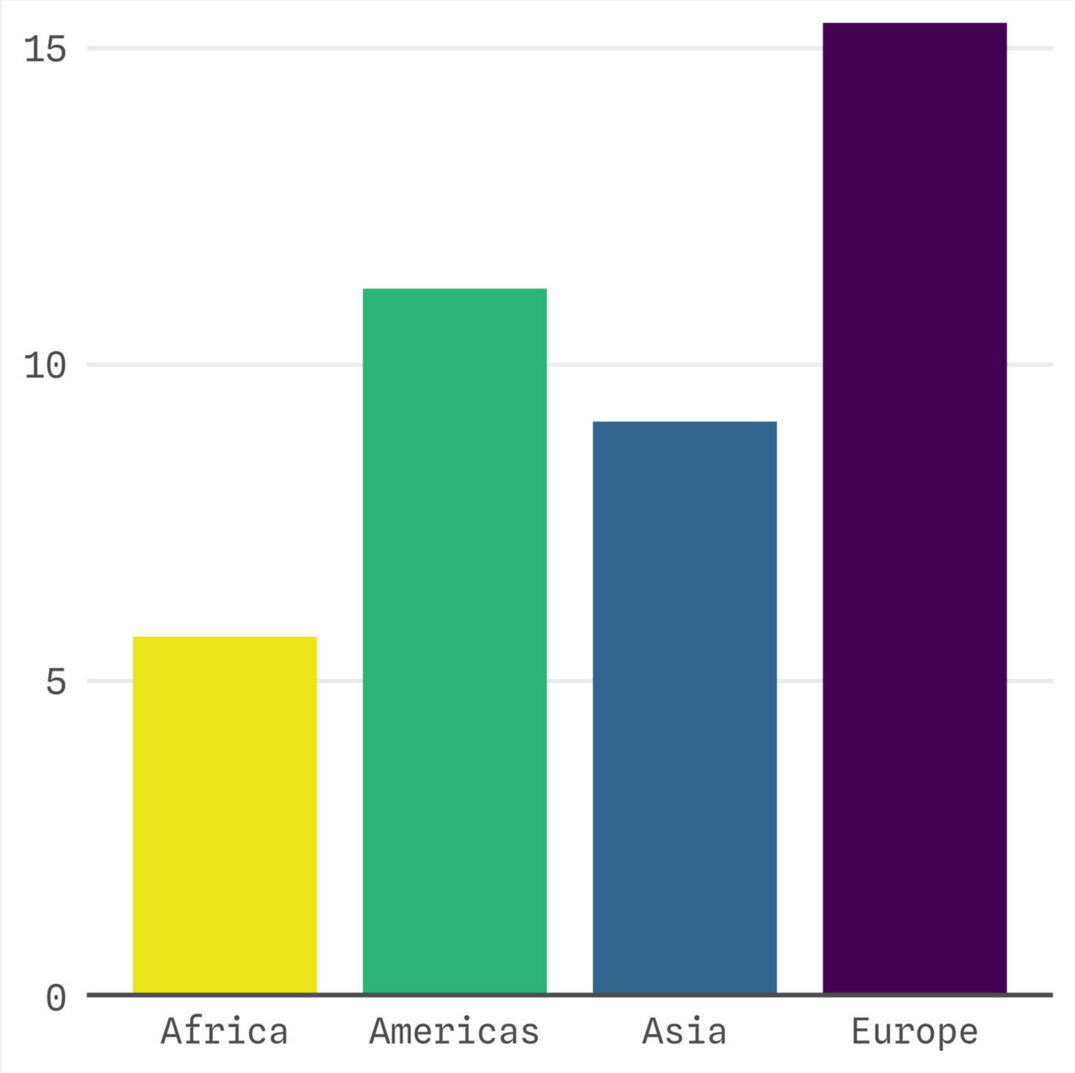
Source: "When to use sequential and when to use diverging color scales" by Lisa C. Muth / DataWrapper

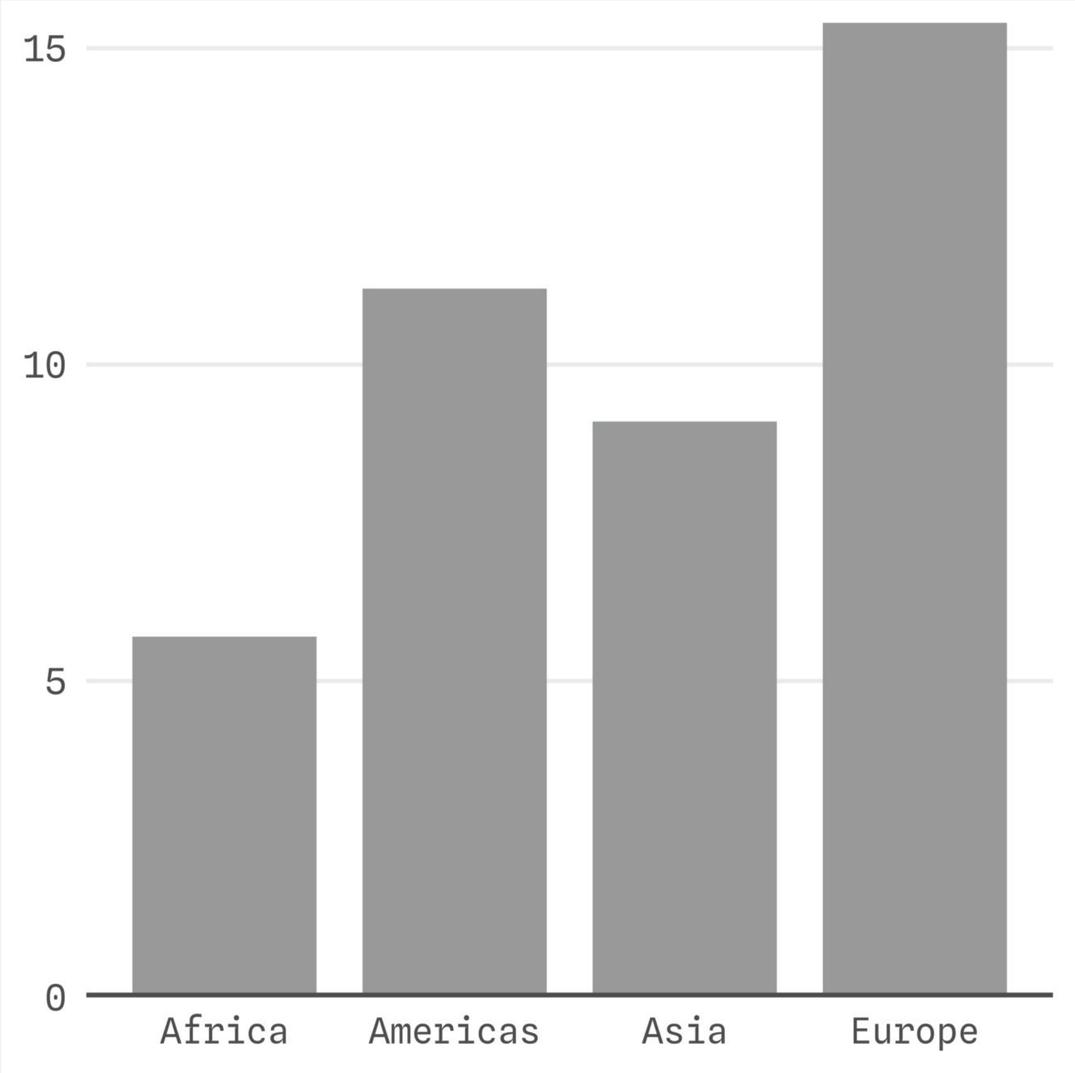
“Viridis is THE perfect color palette!”



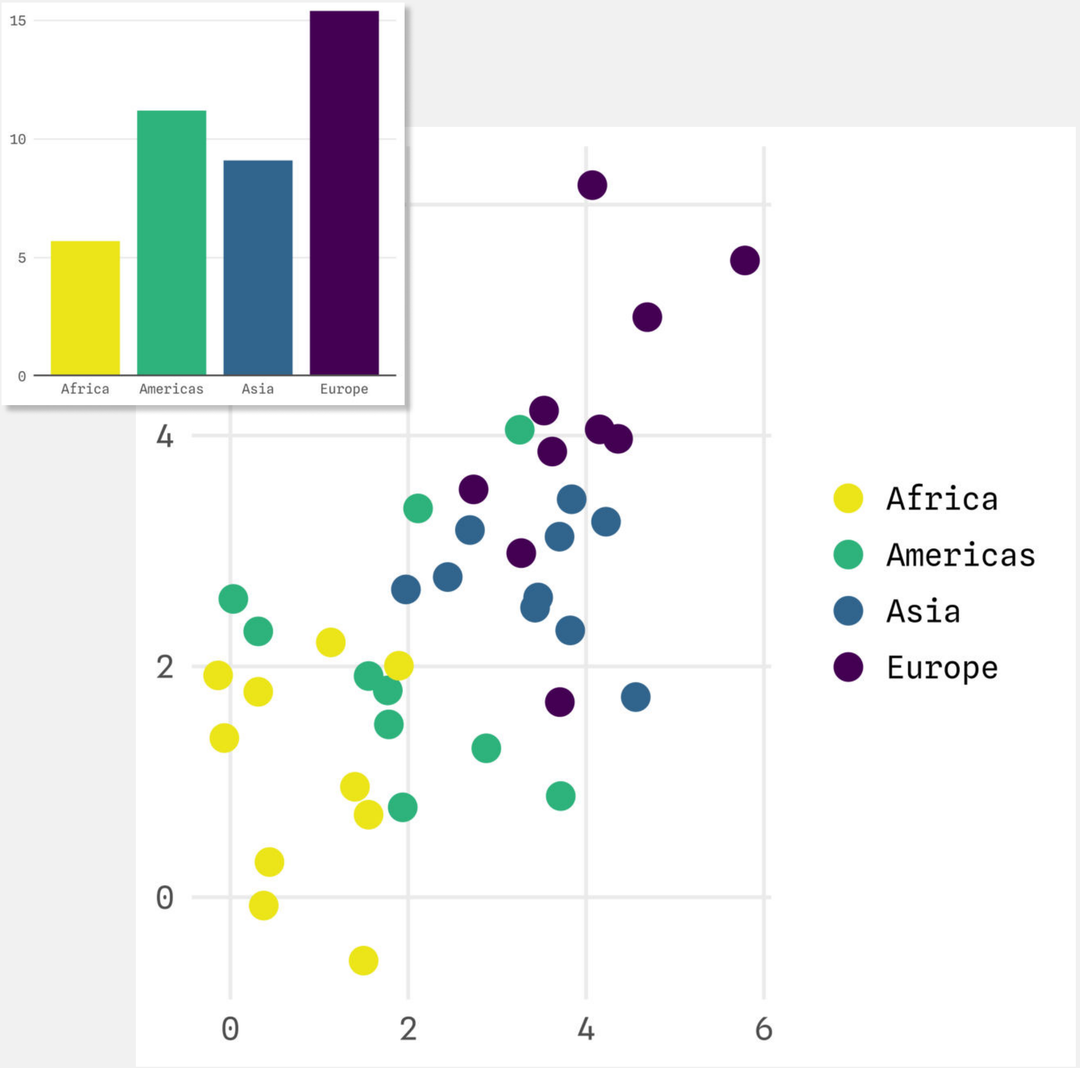


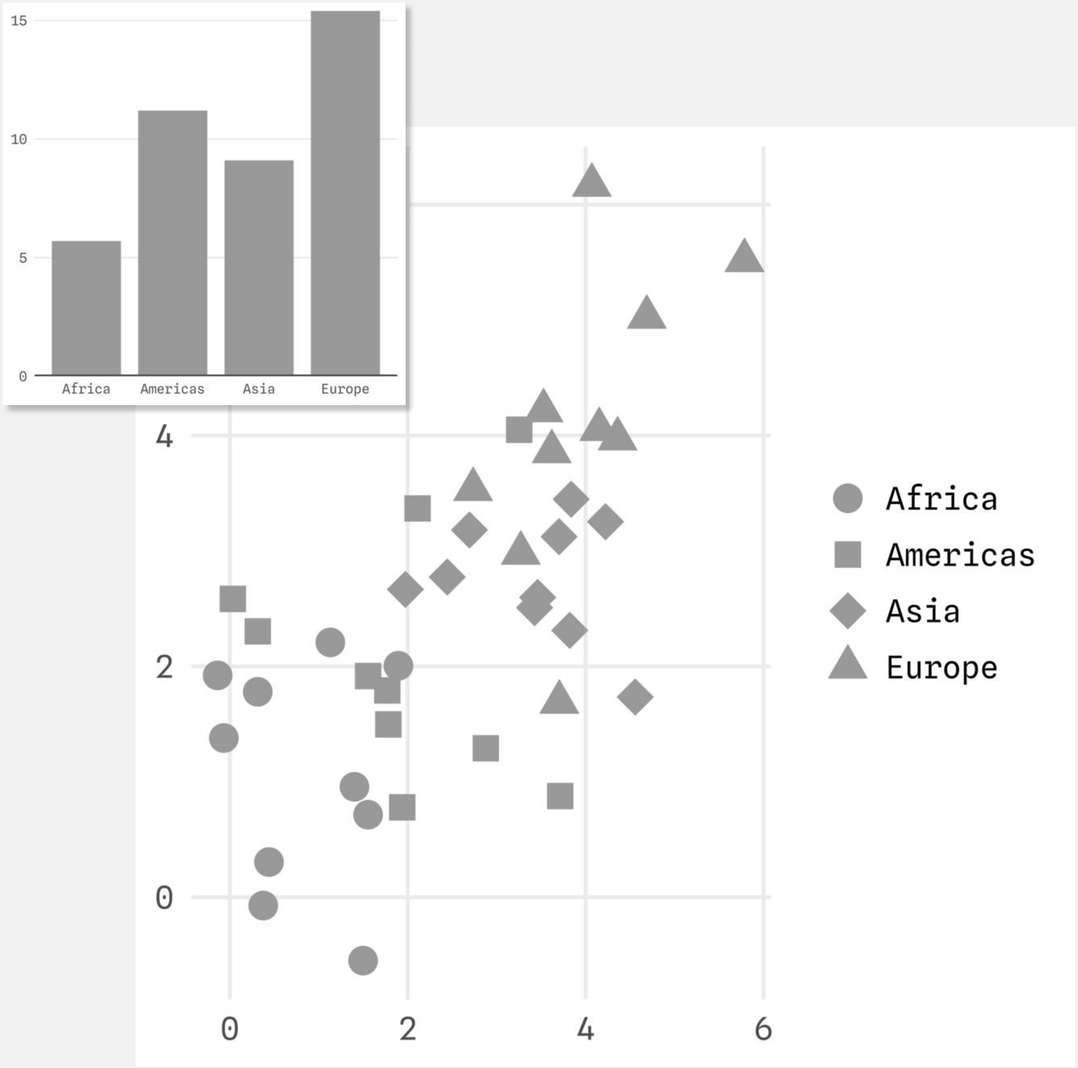
“Viridis is THE perfect color palette!”



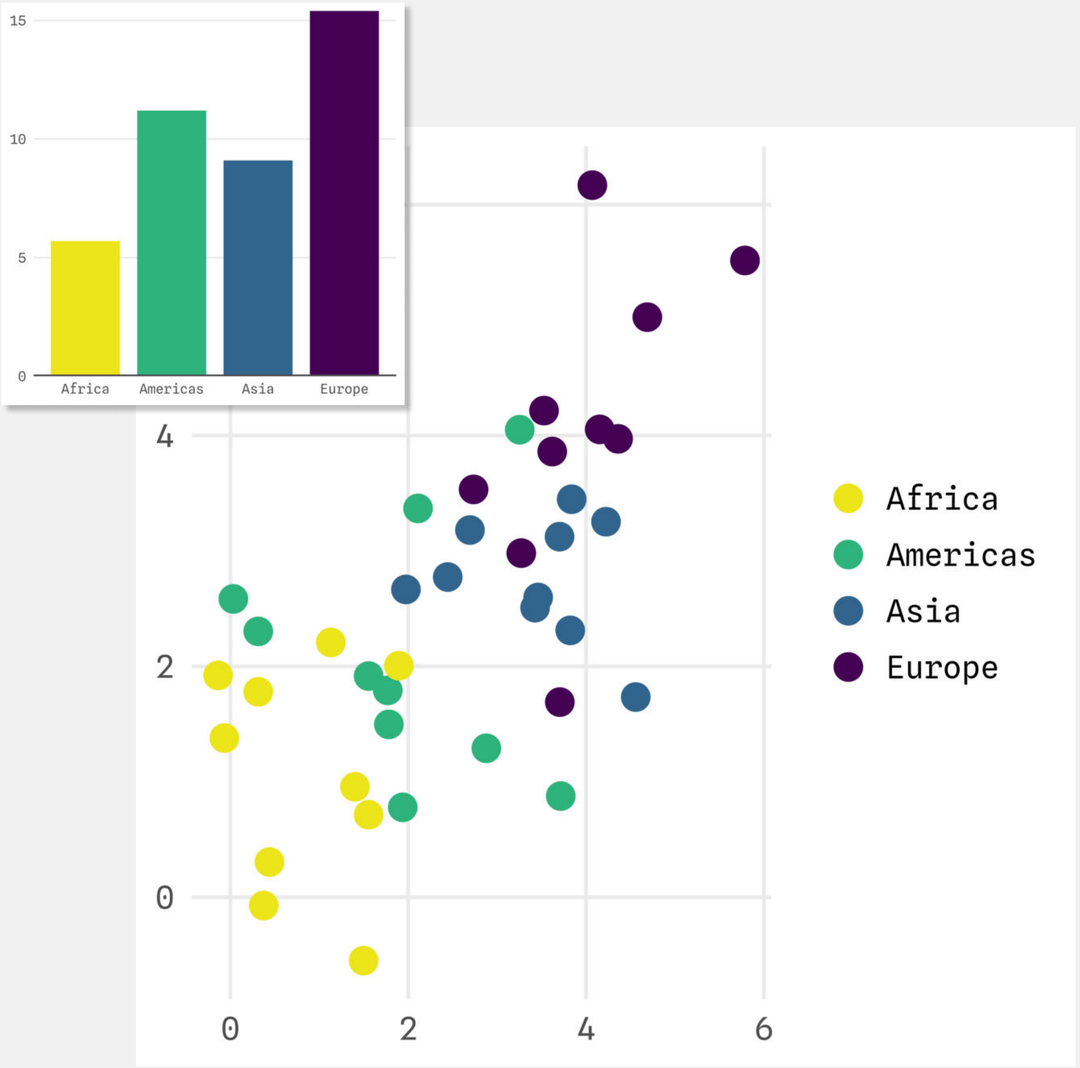


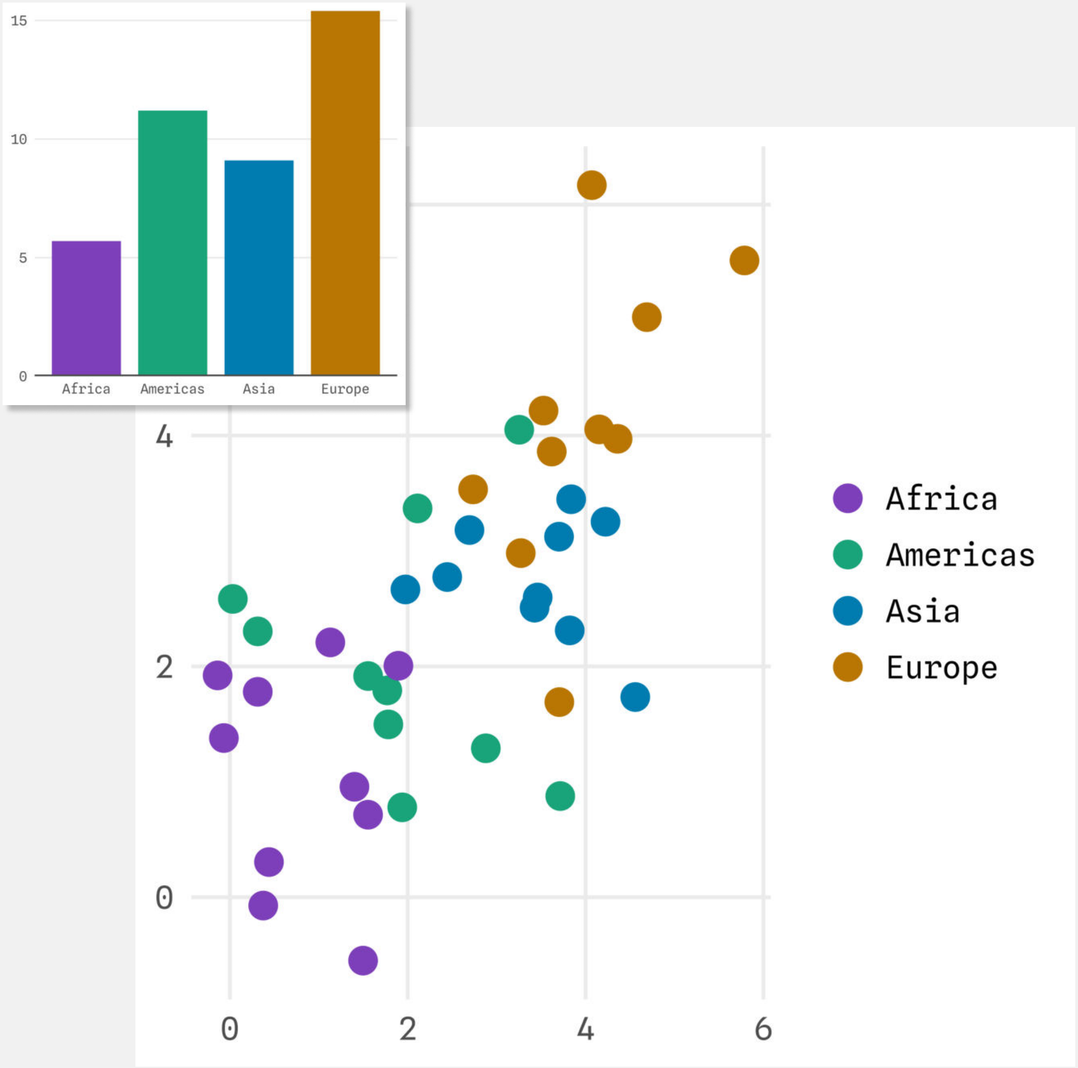
“Viridis is THE perfect color palette!”





“Viridis is THE perfect color palette!”





Use Colors Consistently

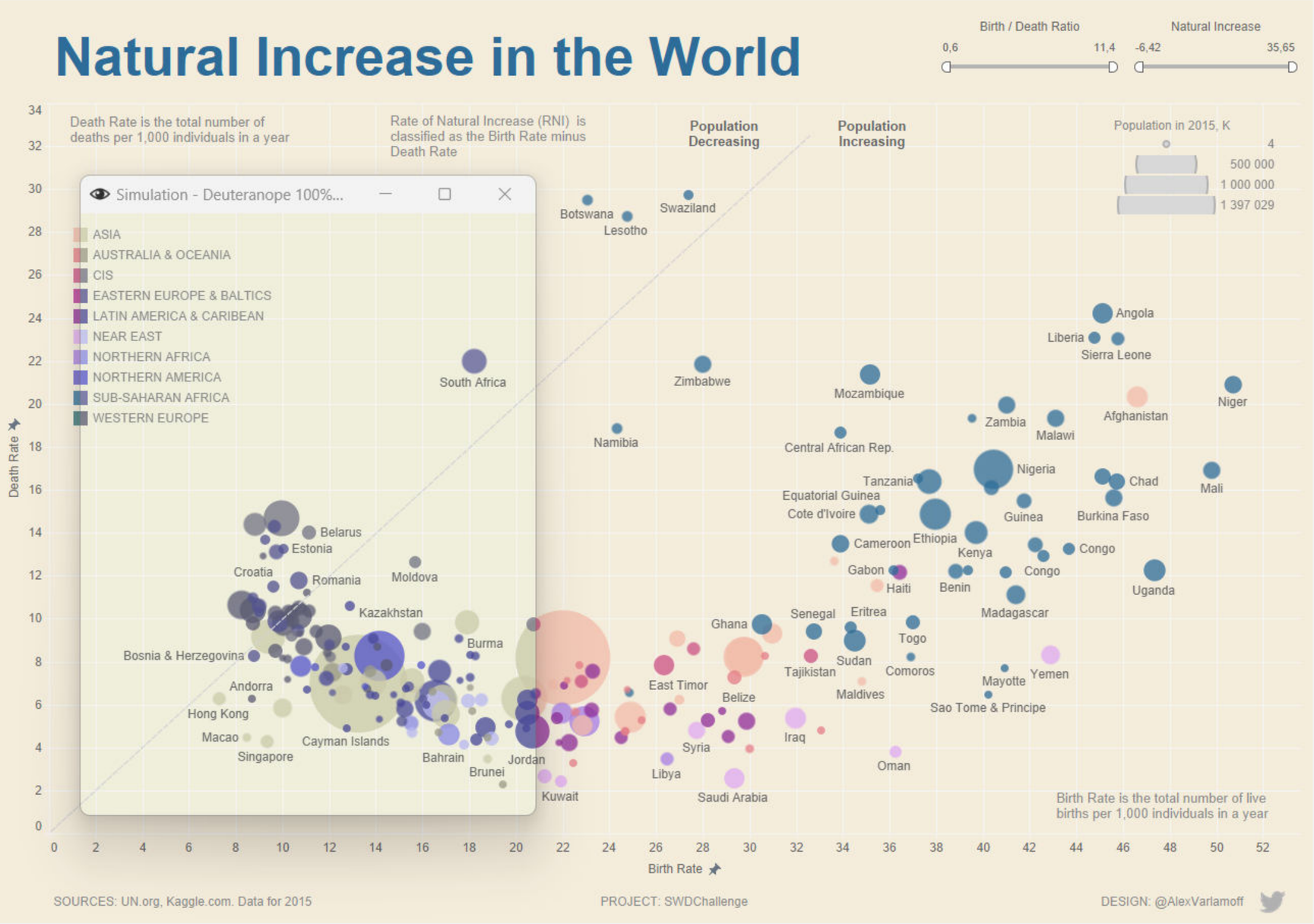
Source: "When to use sequential and when to use diverging color scales" by Lisa C. Muth / DataWrapper

Use Intuitive Colors (?)

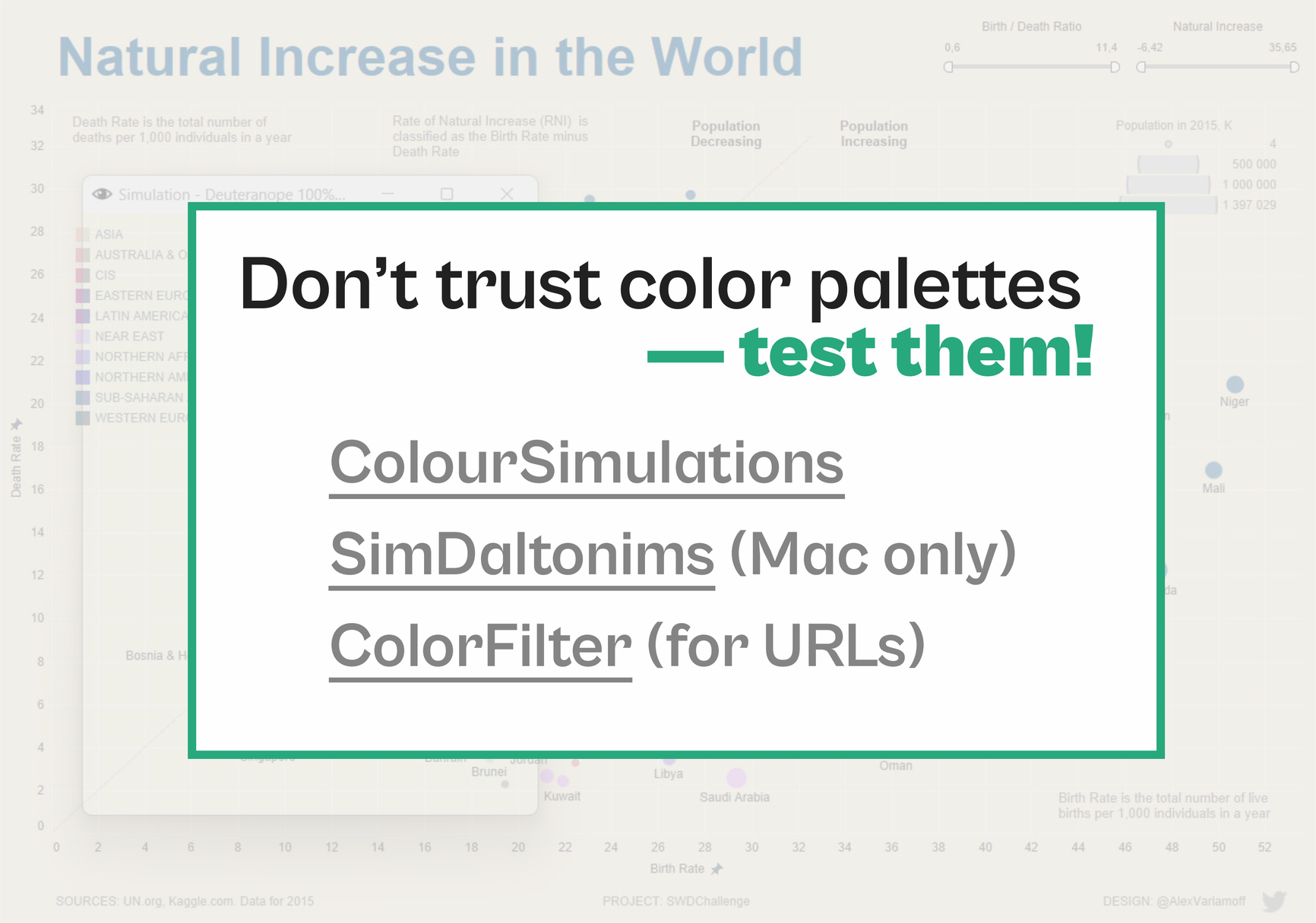
Source: "When to use sequential and when to use diverging color scales" by Lisa C. Muth / DataWrapper

Ensure Readibility for Color-Blind Persons

Source: "Fundamentals of Data Visualization" by Claus O. Wilke



Contribution to the SWDchallenge by Alex Varlamoff — tested with a color-blindness simulator



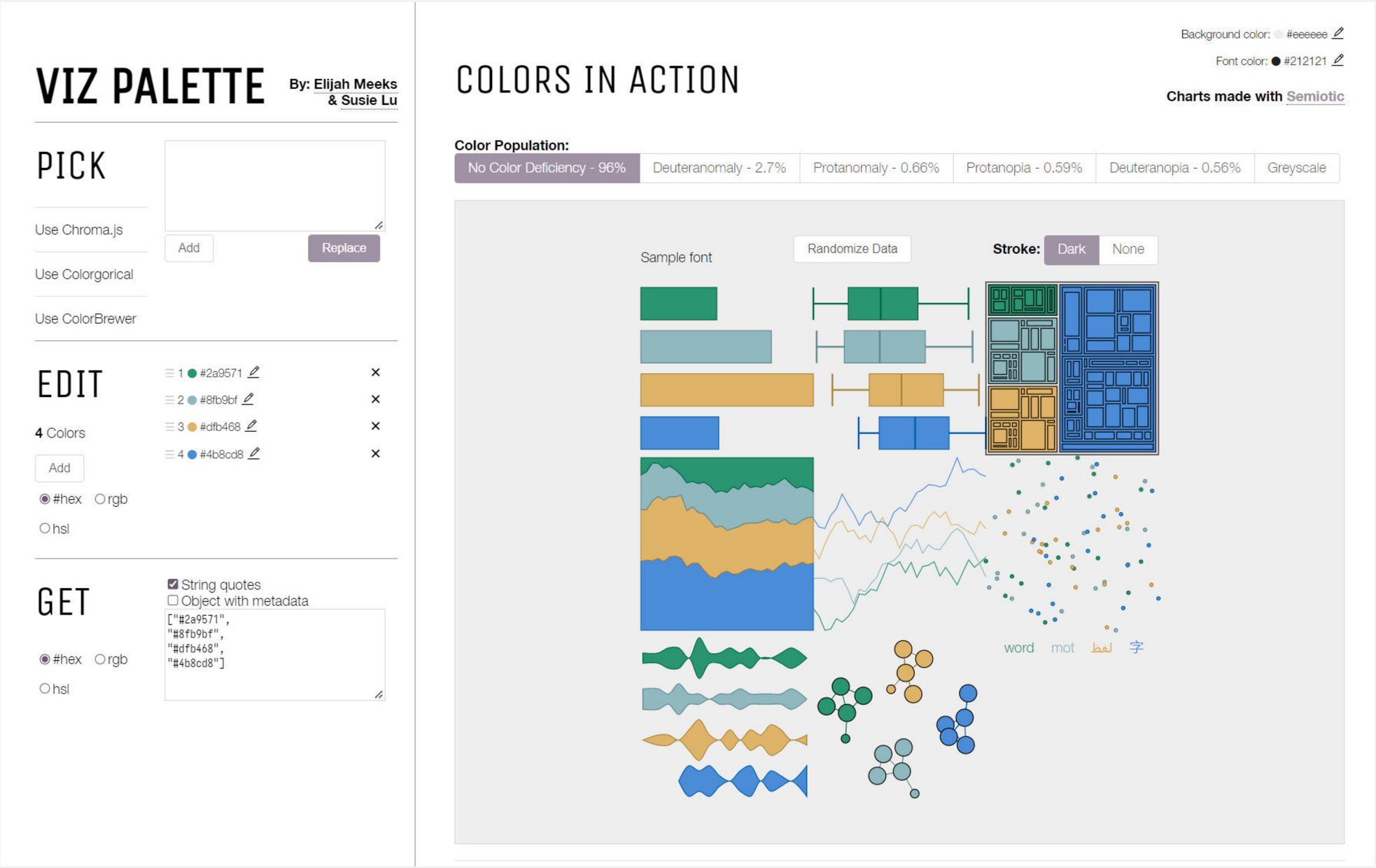
Don’t trust color palettes — test them!

ColourSimulations

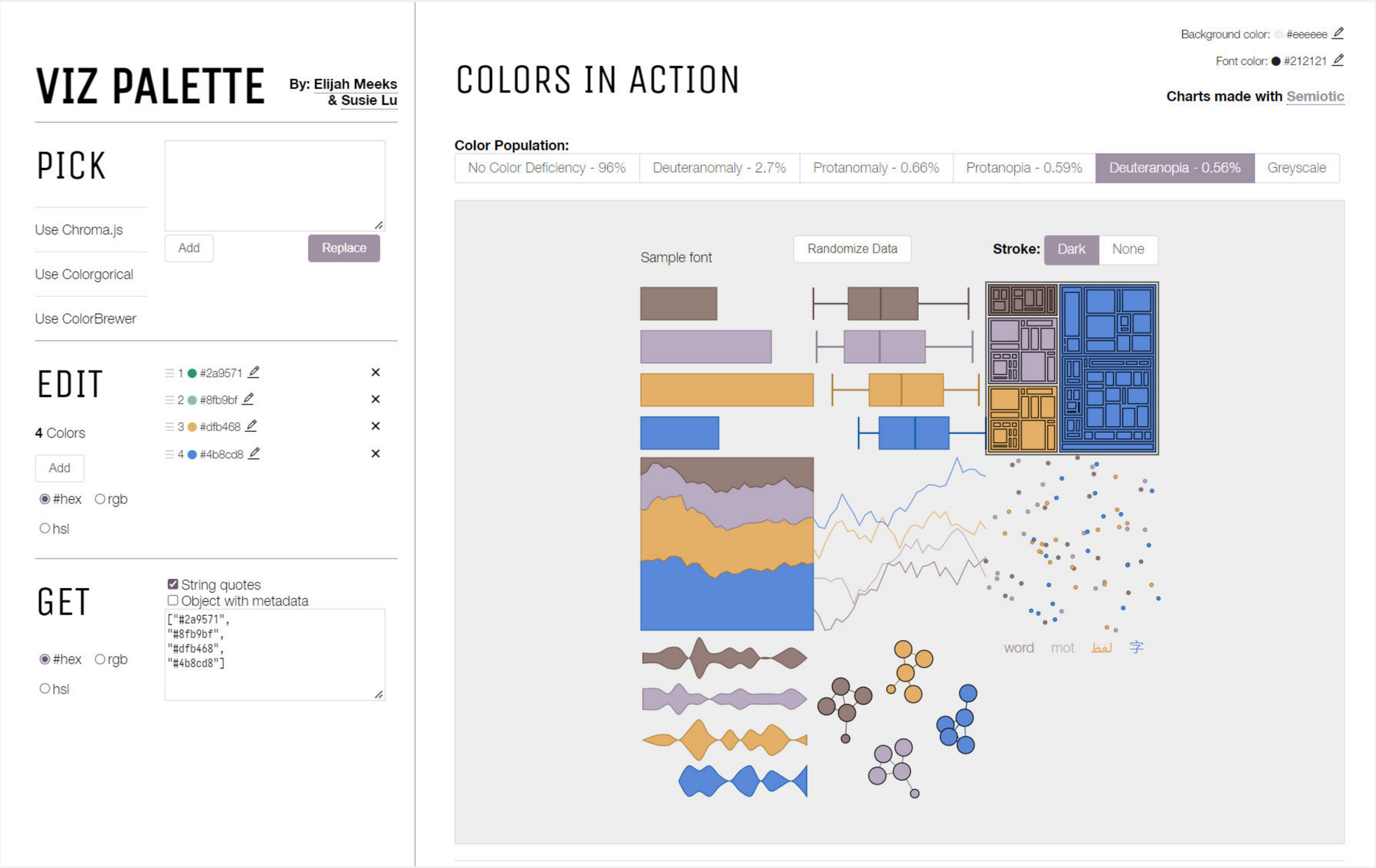
SimDaltonims (Mac only)

ColorFilter (for URLs)

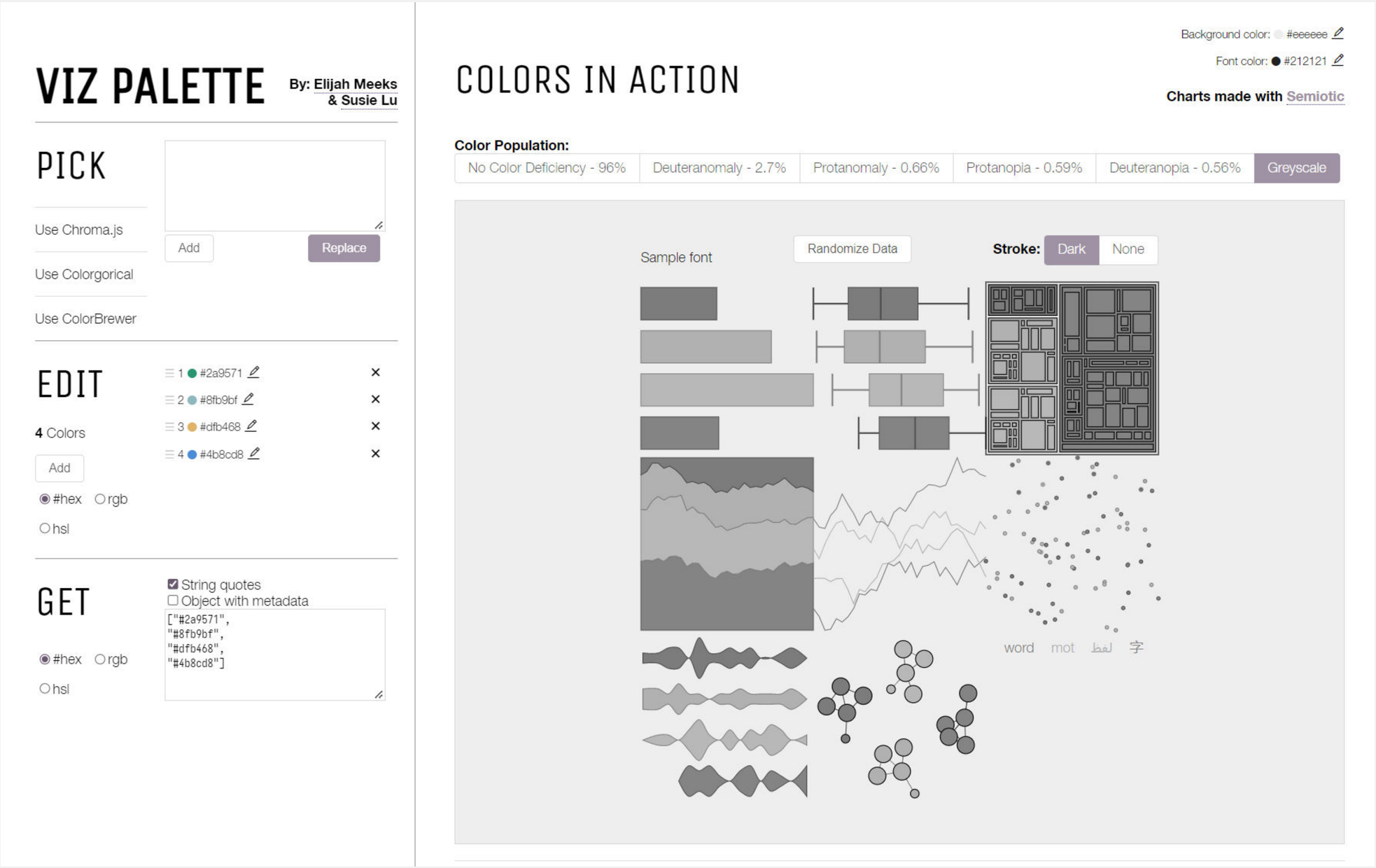
Contribution to the SWDchallenge by Alex Varlamoff — tested with a color-blindness simulator


VIZ PALETTE
By: Elijah Meeks & Susie Lu
PICK
Use Chroma.js
Use Colorgorical
Use ColorBrewer
Add
Replace
EDIT
4 Colors
1 #2a9571
2 #8fb9bf
3 #dfb468
4 #4b8cd8
Add
#hex rgb
hsl
GET
String quotes
Object with metadata
["#2a9571",
"#8fb9bf",
"#dfb468",
"#4b8cd8"]
#hex rgb
hsl
COLORS IN ACTION
Background color: #eeeeee
Font color: #212121
Charts made with Semiotic
Color Population:
No Color Deficiency - 96%
Deuteranomaly - 2.7%
Protanomaly - 0.66%
Protanopia - 0.59%
Deuteranopia - 0.56%
Greyscale
Sample font
Randomize Data
Stroke: Dark None
word mot لفظ 字

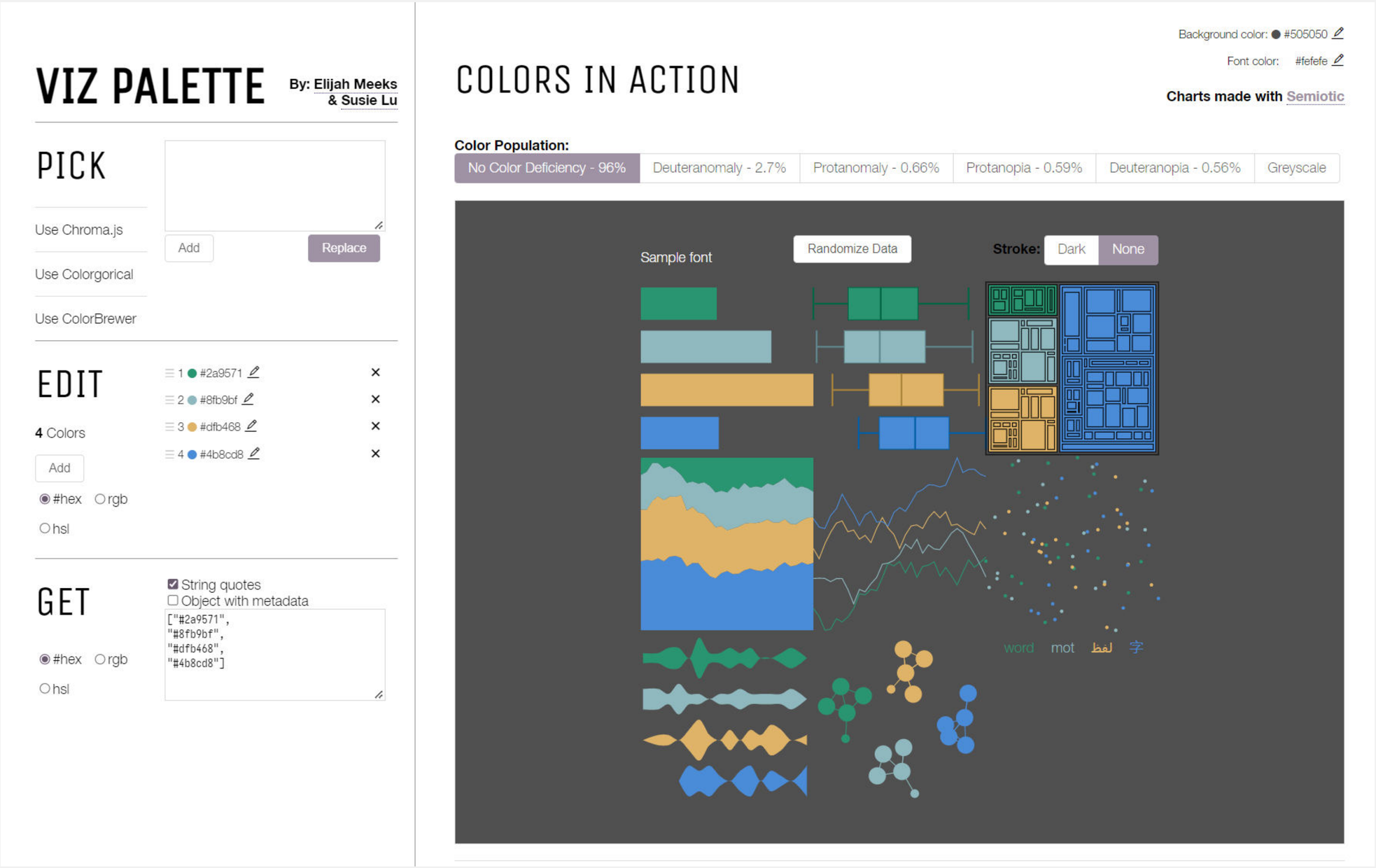
Viz Palette


VIZ PALETTE
By: Elijah Meeks & Susie Lu
PICK
Use Chroma.js
Use Colorgorical
Use ColorBrewer
Add
Replace
EDIT
4 Colors
Add
#hex
rgb
hsl
1 #2a9571
2 #8fb9bf
3 #dfb468
4 #4b8cd8
GET
String quotes
Object with metadata
["#2a9571",
"#8fb9bf",
"#dfb468",
"#4b8cd8"]
#hex
rgb
hsl
COLORS IN ACTION
Background color: #eeeeee
Font color: #212121
Charts made with Semiotic
Color Population:
No Color Deficiency - 96%
Deuteranomaly - 2.7%
Protanomaly - 0.66%
Protanopia - 0.59%
Deuteranopia - 0.56%
Greyscale
Sample font
Randomize Data
Stroke:
Dark
None
word
mot
لفظ
字

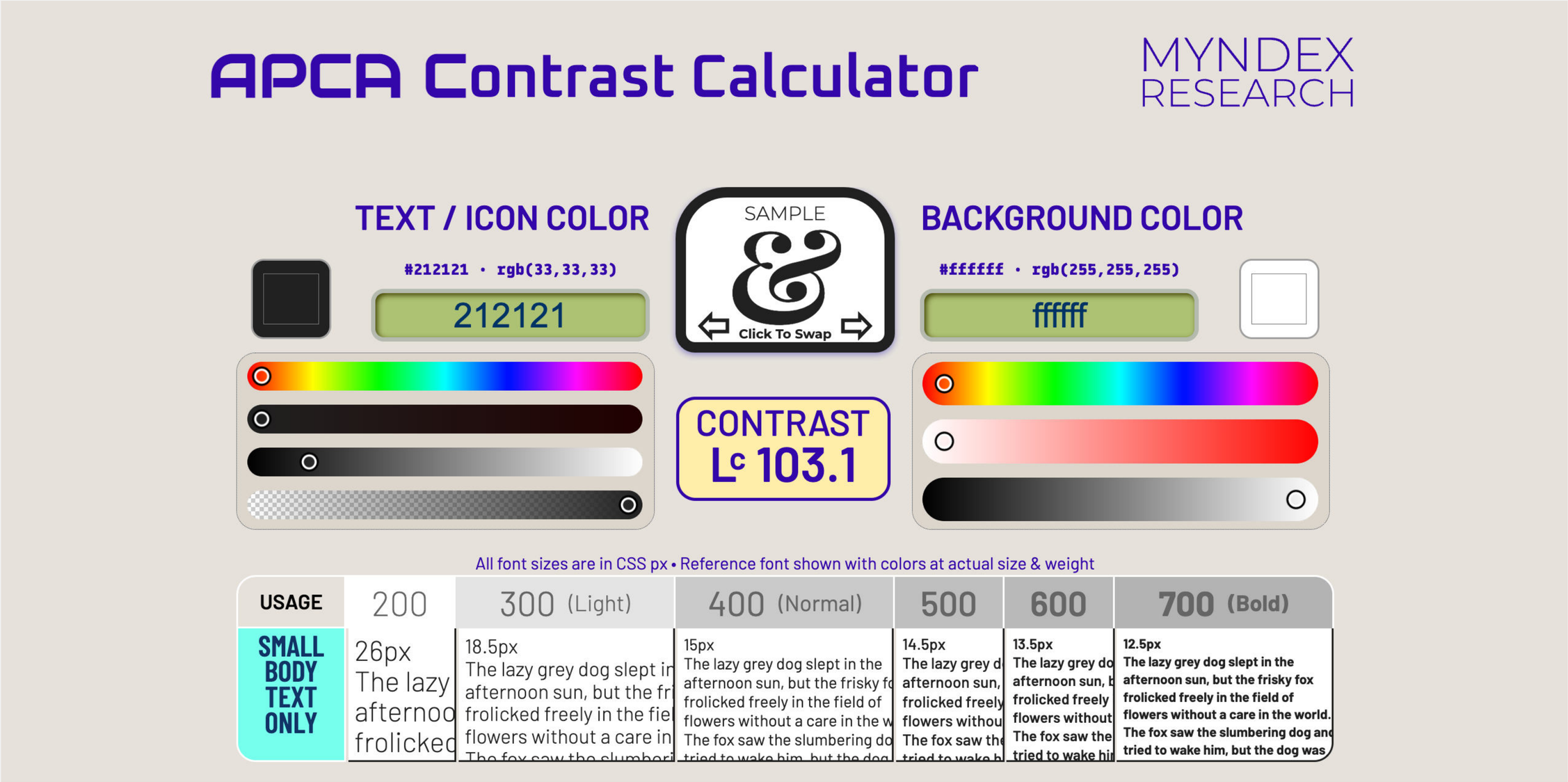
Viz Palette


VIZ PALETTE
By: Elijah Meeks & Susie Lu
PICK
Use Chroma.js
Use Colorgorical
Use ColorBrewer
Add
Replace
EDIT
4 Colors
1 #2a9571
2 #8fb9bf
3 #dfb468
4 #4b8cd8
Add
#hex rgb
hsl
GET
String quotes
Object with metadata
["#2a9571",
"#8fb9bf",
"#dfb468",
"#4b8cd8"]
#hex rgb
hsl
COLORS IN ACTION
Background color: #eeeeee
Font color: #212121
Charts made with Semiotic
Color Population:
No Color Deficiency - 96%
Deuteranomaly - 2.7%
Protanomaly - 0.66%
Protanopia - 0.59%
Deuteranopia - 0.56%
Greyscale
Sample font
Randomize Data
Stroke: Dark None
word mot لغط 字

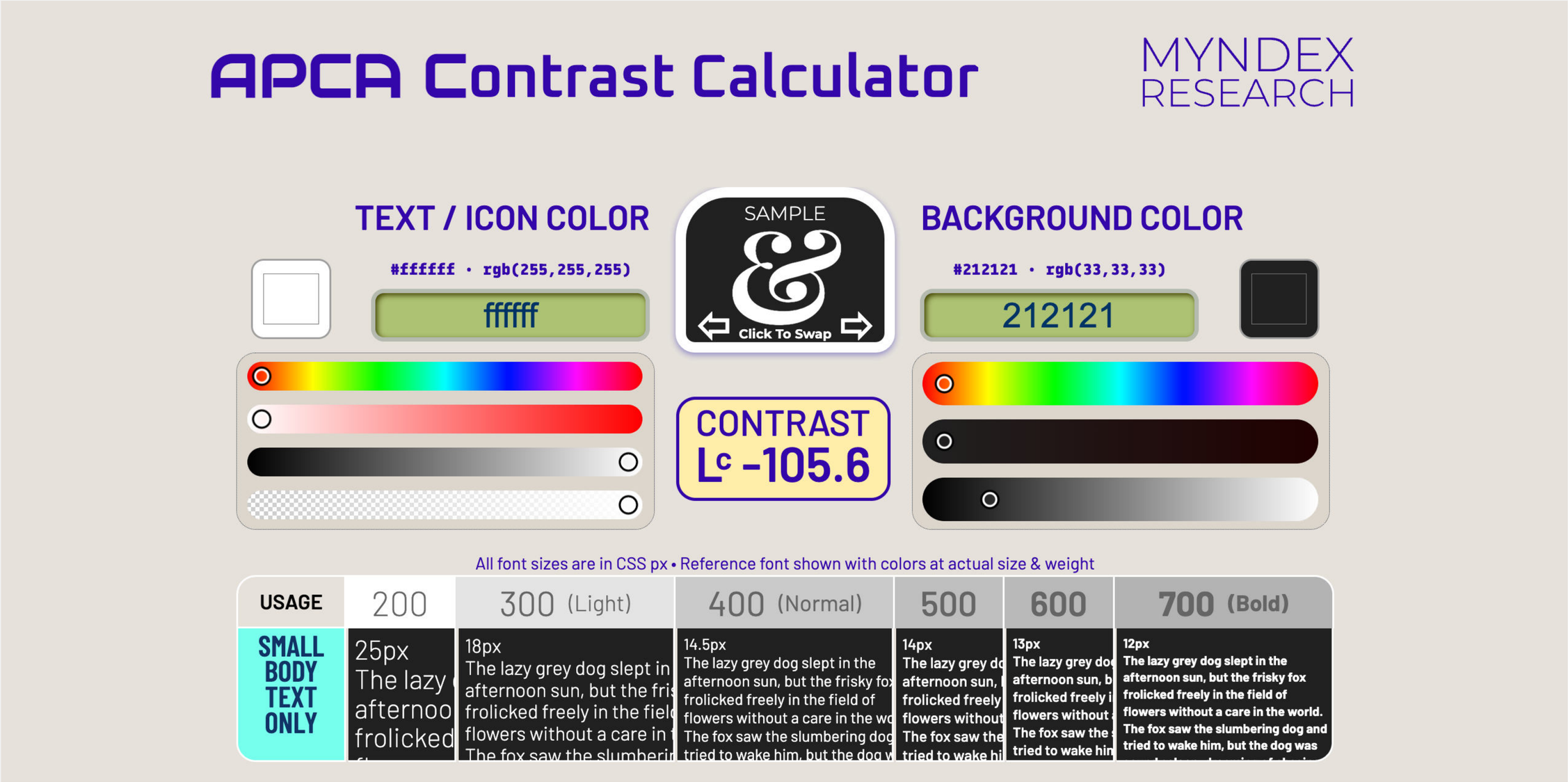
Viz Palette


VIZ PALETTE
By: Elijah Meeks & Susie Lu
PICK
Use Chroma.js
Use Colorgorical
Use ColorBrewer
Add
Replace
EDIT
4 Colors
Add
#hex rgb
hsl
1 #2a9571
2 #8fb9bf
3 #dfb468
4 #4b8cd8
GET
#hex rgb
hsl
String quotes
Object with metadata
["#2a9571",
"#8fb9bf",
"#dfb468",
"#4b8cd8"]
COLORS IN ACTION
Background color: #505050
Font color: #fefefe
Charts made with Semiotic
Color Population:
No Color Deficiency - 96%
Deuteranomaly - 2.7%
Protanomaly - 0.66%
Protanopia - 0.59%
Deuteranopia - 0.56%
Greyscale
Sample font
Randomize Data
Stroke:
Dark
None
word
mot
لفظ
字

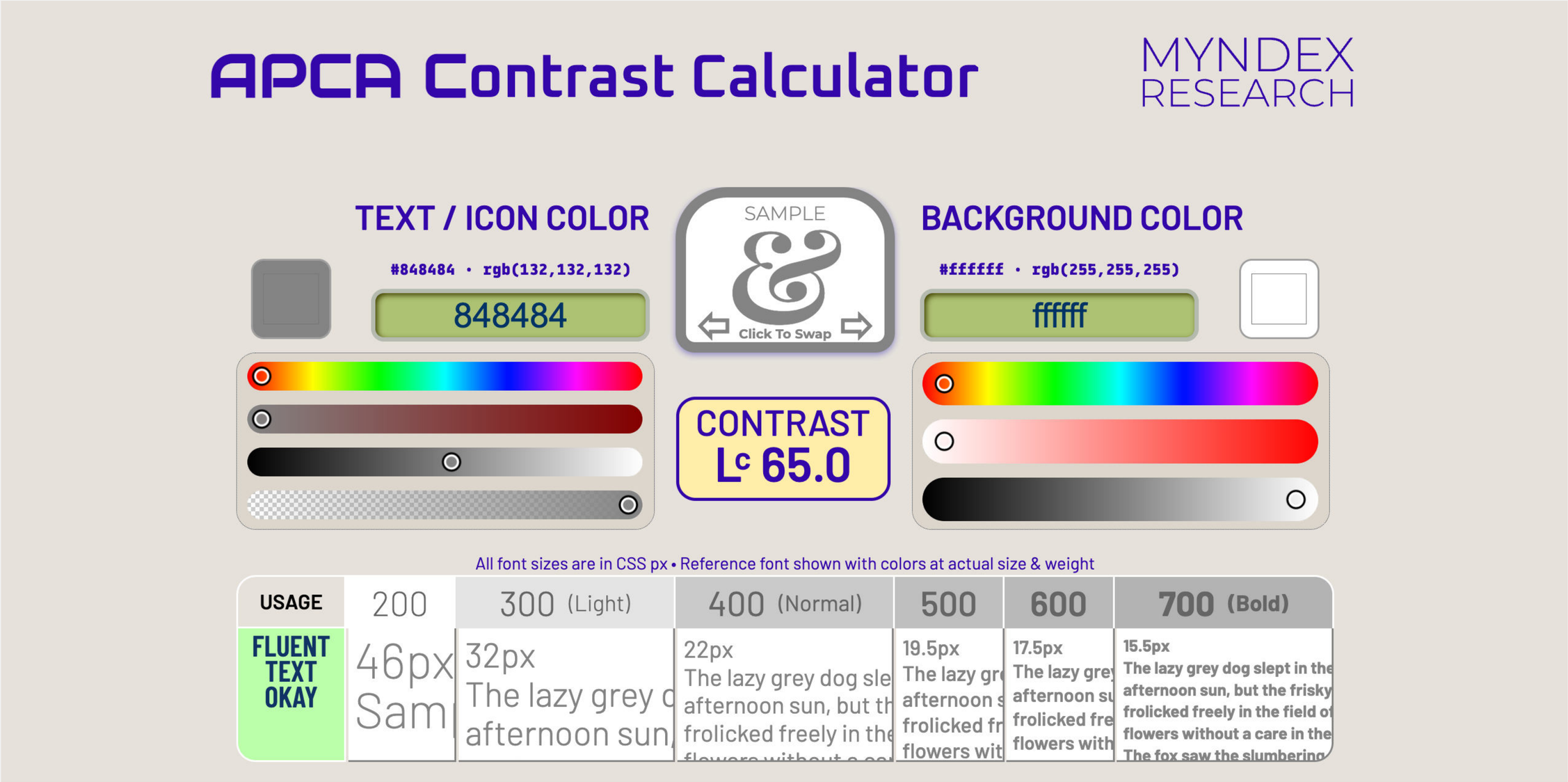
Viz Palette



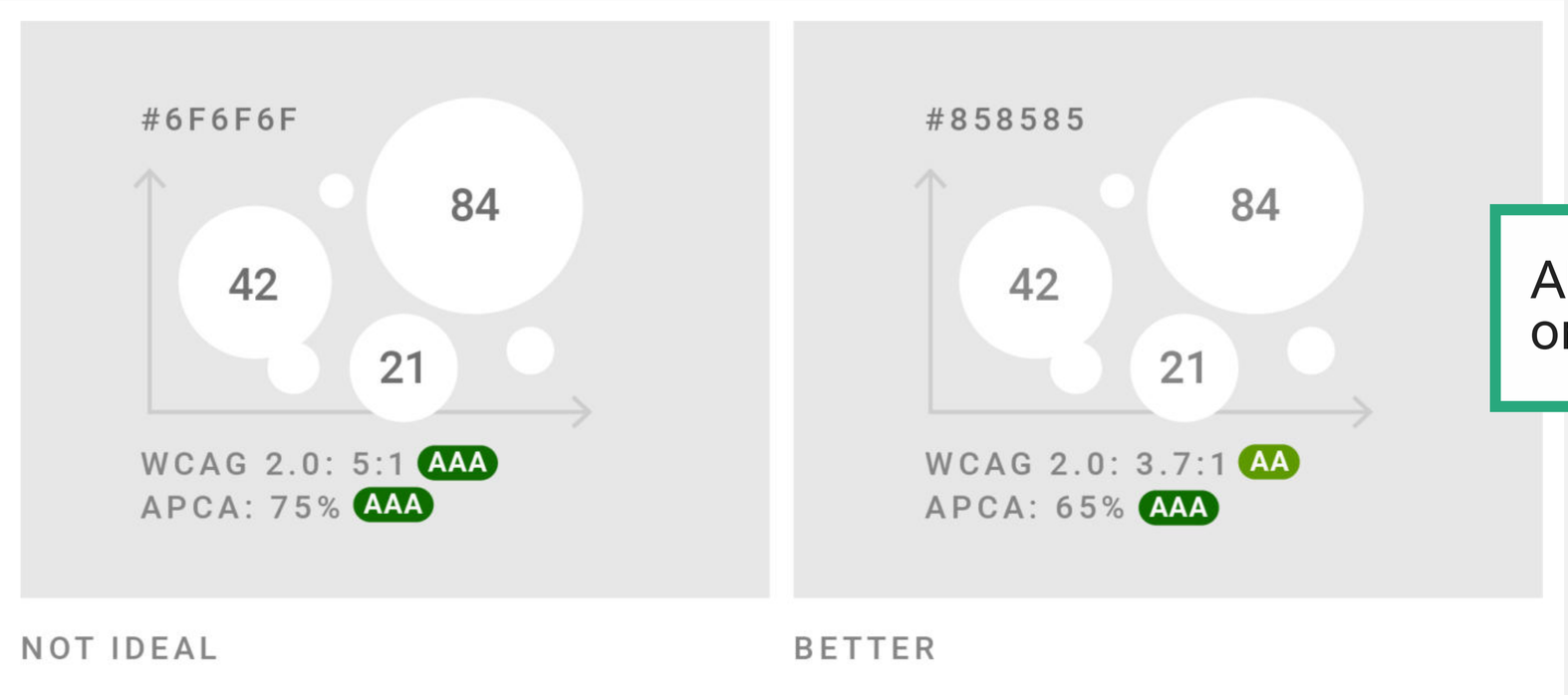
The Myndex APCA Contrast Calculator displays modern contrast ratios for various combinations of text size and font weight



The Myndex APCA Contrast Calculator displays modern contrast ratios for various combinations of text size and font weight



The Myndex APCA Contrast Calculator displays modern contrast ratios for various combinations of text size and font weight



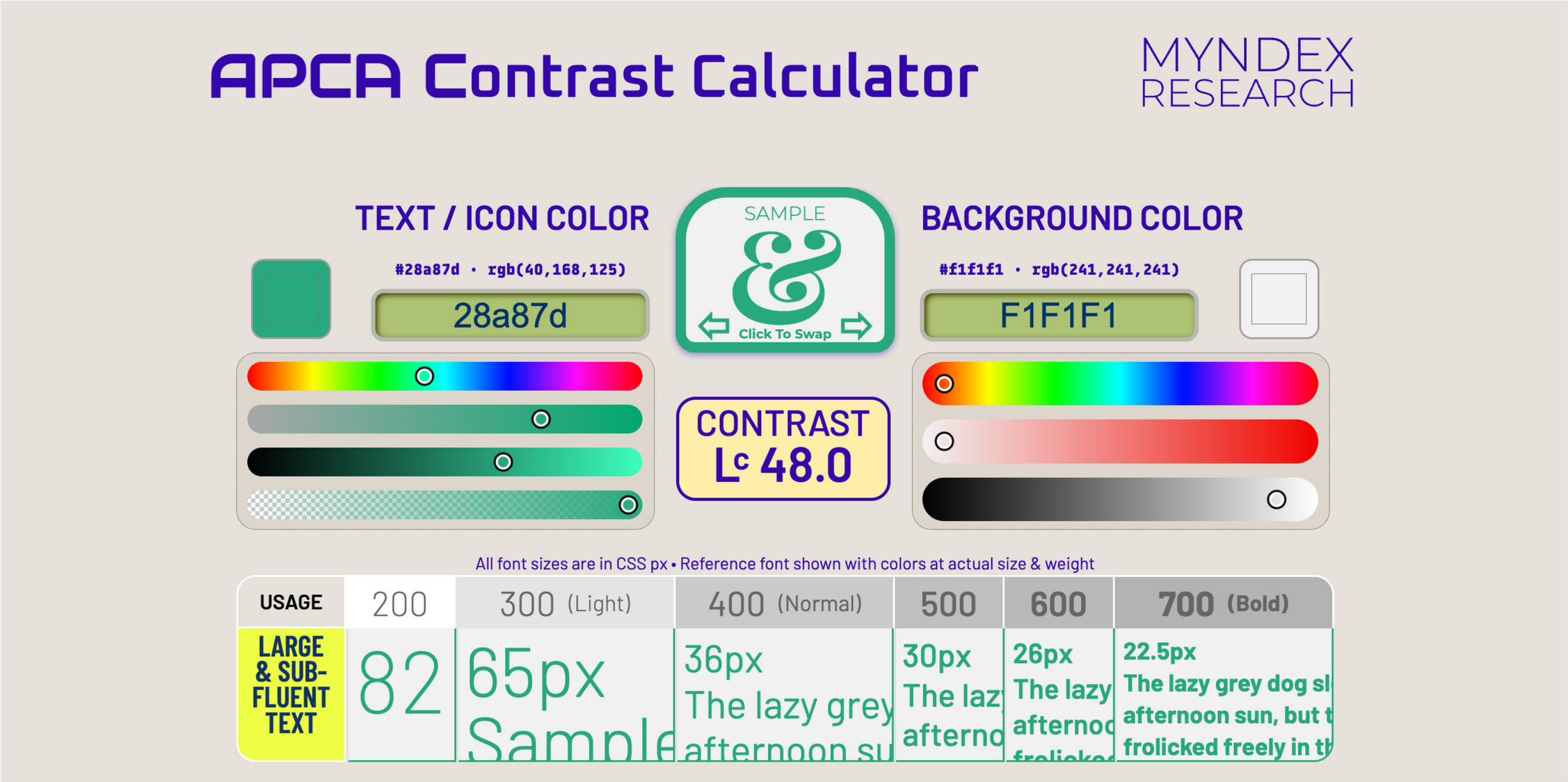
APCA allows for lighter gray on light backgrounds

"It's time for a more sophisticated color contrast check for data visualizations" by Lisa C. Muth / DataWrapper

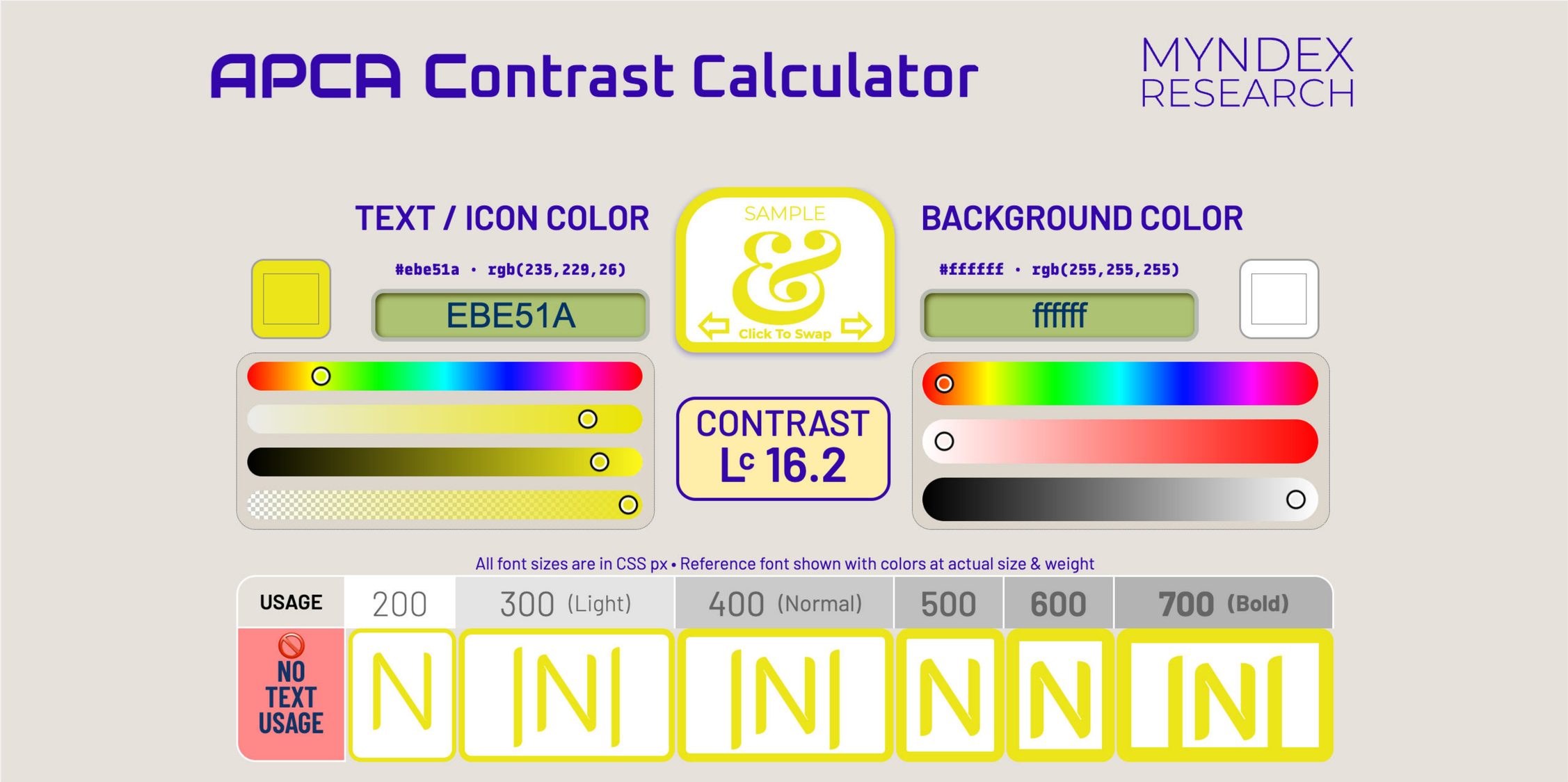
APCA allows for lighter gray on light backgrounds...

... but also requests lighter gray on dark backgrounds

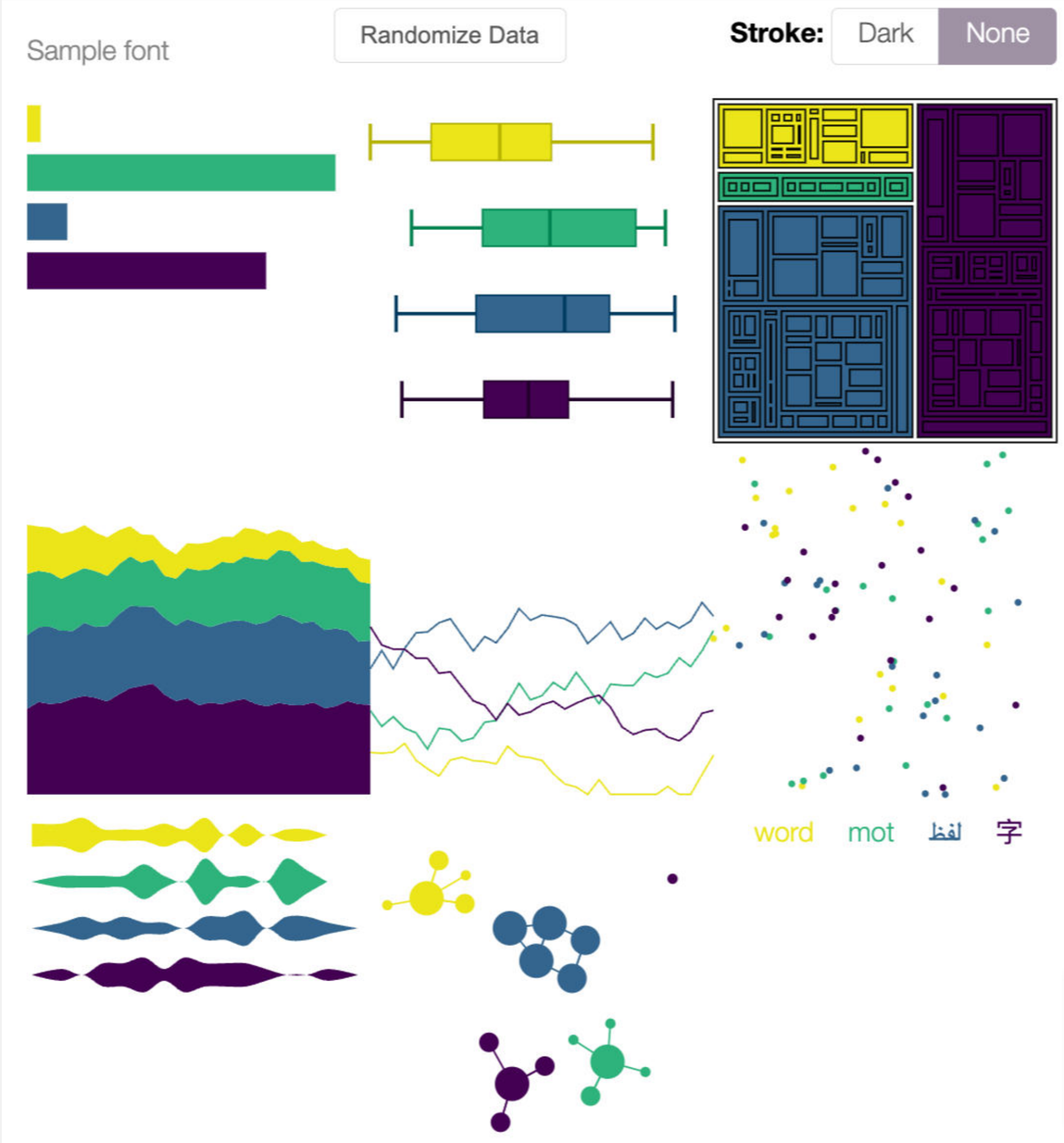
"It's time for a more sophisticated color contrast check for data visualizations" by Lisa C. Muth / DataWrapper



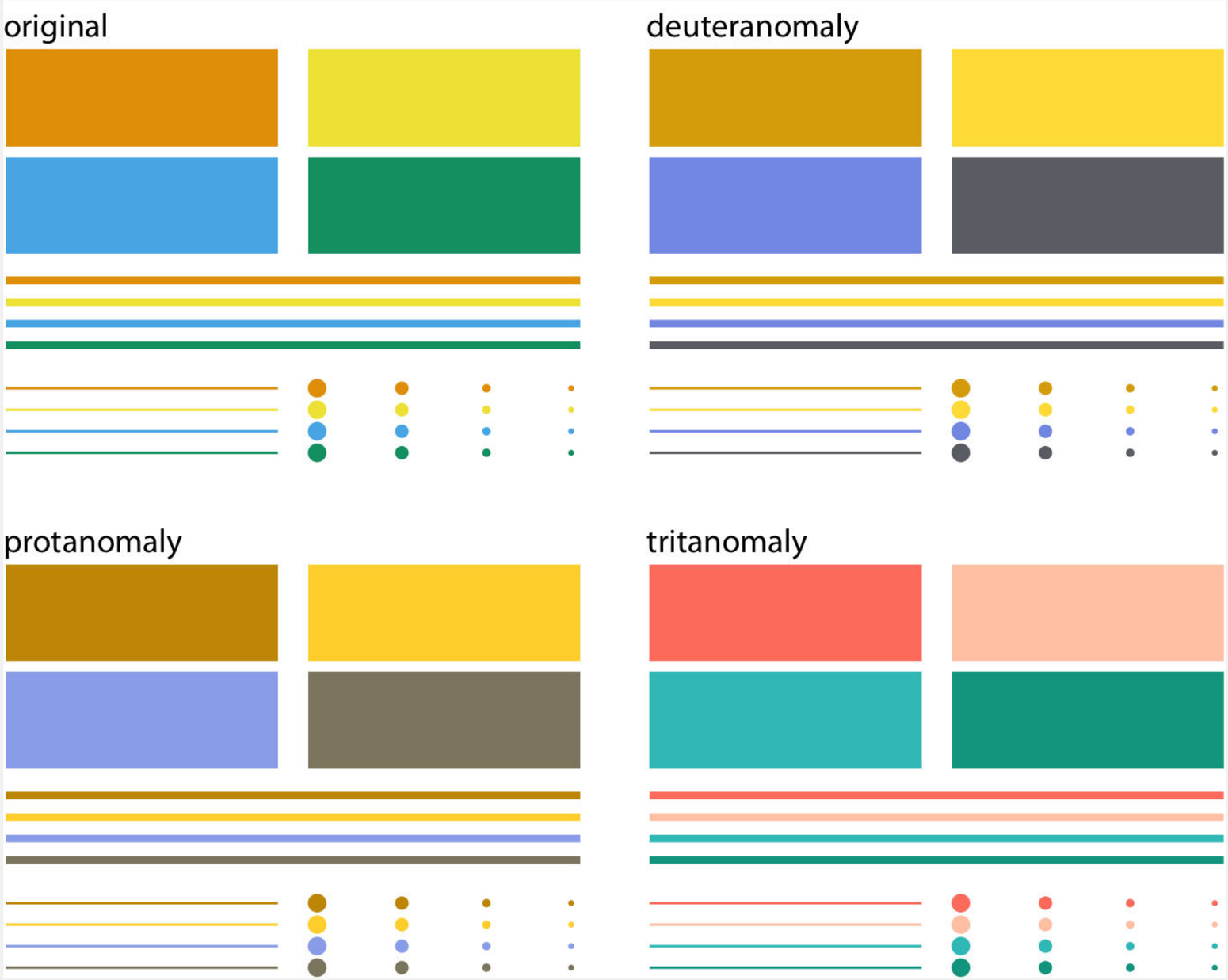
The Myndex APCA Contrast Calculator displays modern contrast ratios for various combinations of text size and font weight



The Myndex APCA Contrast Calculator displays modern contrast ratios for various combinations of text size and font weight



Viz Palette



Source: *"Fundamentals of Data Visualization"* by Claus O. Wilke

Typography
Typography
Typography
Typography

cedricscherer.com @CedScherer z3tt

I'll be waiting for you!

I'LL BE WAITING FOR YOU!

I'LL BE WAITING FOR YOU!

I'll be waiting for you!

I'll be waiting for you!

I'll be waiting for you!

I'll be waiting for you!

I'll be waiting for you!

The Choice of the Typeface

- ☛ **Context matters**: font(s) should fit the topic and audience
- ☛ Use different sizes, weights and colors to **visualize hierarchy**
- ☛ Avoid using **ALL CAPS** and too many different styles
- ☛ Use **tabular typefaces** for numbers

The Choice of the Typeface

- ☛ **Context matters**: font(s) should fit the topic and audience
- ☛ Use different sizes, weights and colors to **visualize hierarchy**
- ☛ Avoid using **ALL CAPS** and too many different styles
- ☛ Use **tabular typefaces** for numbers
- ☛ **Consistency is key!**

The 1ll Test

1ll Calibri

1ll Open Sans

1ll Roboto

1ll Lato

1ll Oswald

1ll Cabinet Grotesk

1ll Cabin

1Il Monda

1Il Chivo

1Il Fira Sans

1Il Noto Sans

1ll Amulya

Proportional Numbers

123.45
678.90

Tabular Numbers

123.45
678.90

Using a lot of fonts
and different sizes
makes your design look
cluttered
overcomplicated
AND JUST NOT VERY NICE.

But if you just use
a small selection of
typefaces, styles and sizes
you can keep your design
cleaner, clearer
and just much easier to digest.

Visualize Text Hierarchy

I am important!

I am important, too.

Oh, hi there. Thanks for reading me...

Hey, also look at me!

Visualize Text Hierarchy

I am important!

I am important, too.

Oh, hi there. Thanks for reading me...

Hey, also look at me!

Visualize Text Hierarchy

I am important!

I am important, too.

Oh, hi there. Thanks for reading me...

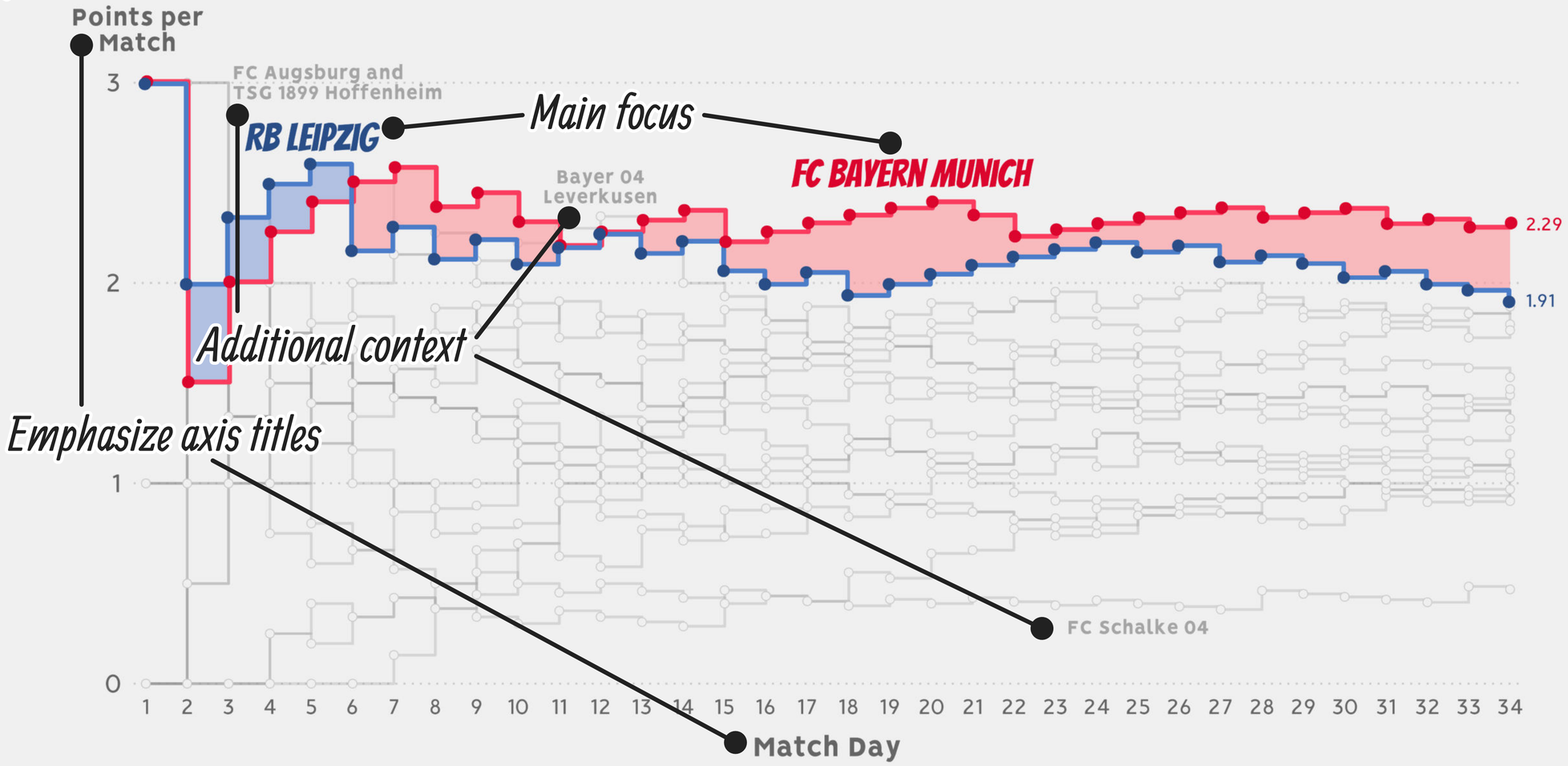
Hey, also look at me!

Bold primary summary

Bayern Munich's Road to the 31st Championship

Lighter secondary information incl. color highlights

FC Bayern Munich is the champion of the Bundesliga season 2020/21 with incredible 2.29 points per match, more than 0.3 points more than the best chasers. RB Leipzig (1.91 points per match) was getting close several times but has lost the trace again in the last matches.



Small, light credit

Visualization: Cédric Scherer

Wrap-Up

Information

Understand your data and be accurate.

Story

Be clear about the message of your visualization.

Goal

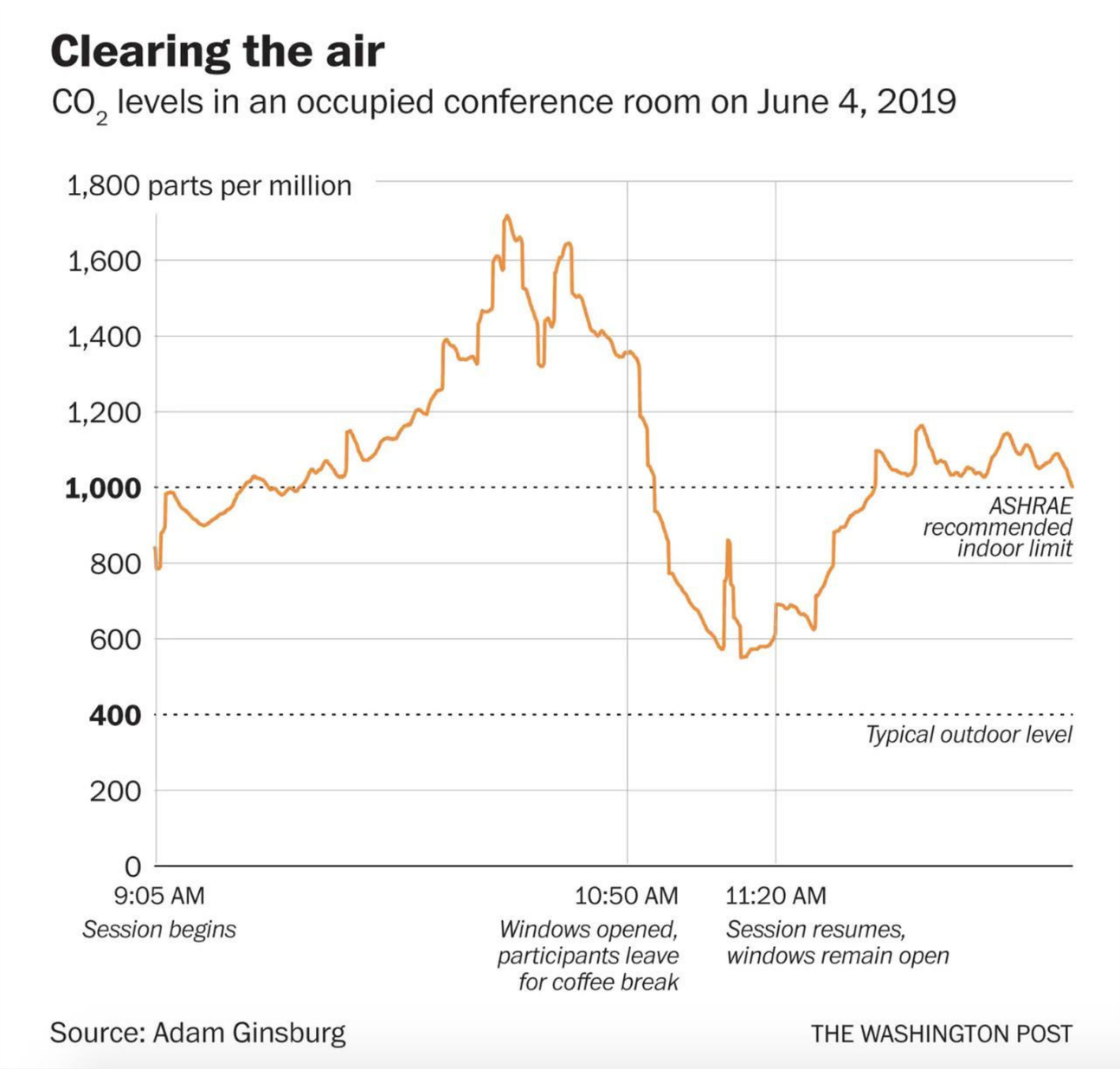
Select charts that successfully transport your story.

Visual Form

Present information in a logical, coherent way.

Guide the View(er)

- tell a story, do not “let the data speak for itself”
- use a suitable chart type that transports the main finding
- use colors correctly and wisely
- add annotations, supporting text, and visual cues
- aim to work with direct labels, not legends



"Clearing the Air" by Adam Ginsburg (Washington Post)

...with notes by Francis Gagnon (Voilà)

Survey:

survey.mdc-berlin.de/129849

Thank you!

cedricscherer.com @CedScherer z3tt

CÉDRIC SCHERER
Data Visualization & Information Design