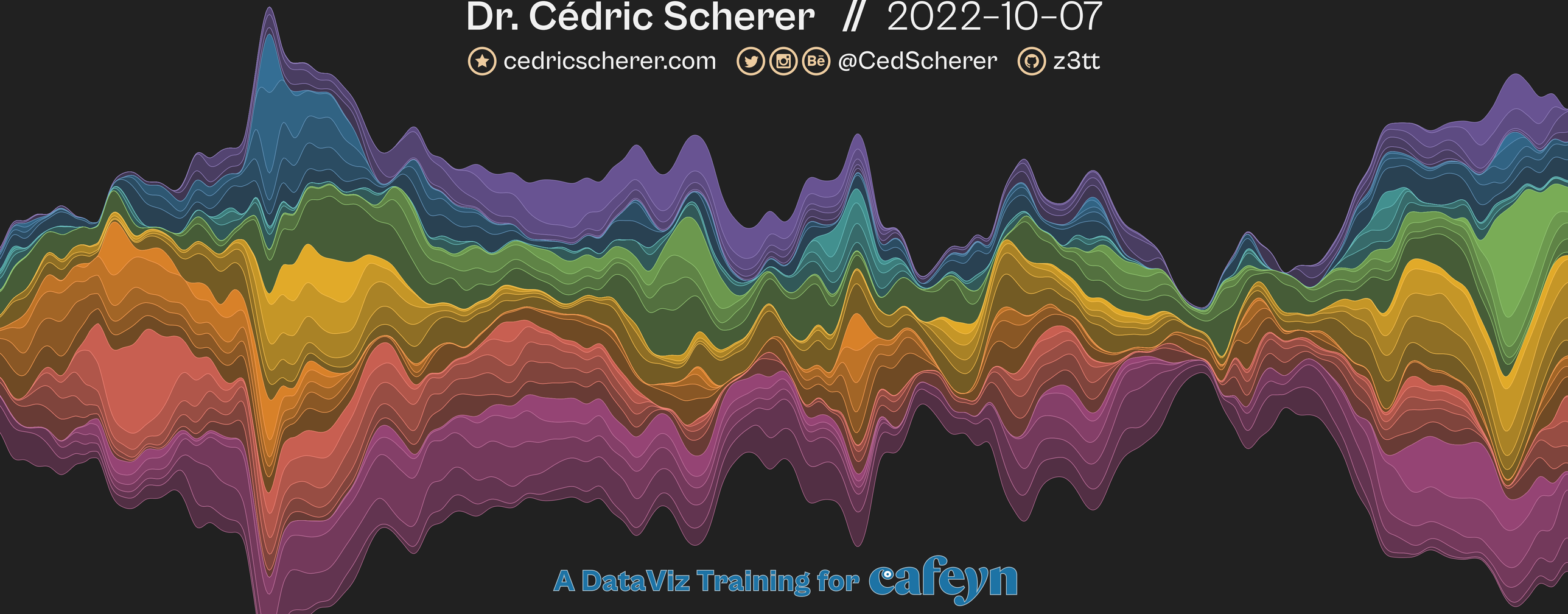
Effective Data Visualization

Design graphics that tell stories in an engaging way

Dr. Cédric Scherer // 2022-10-07

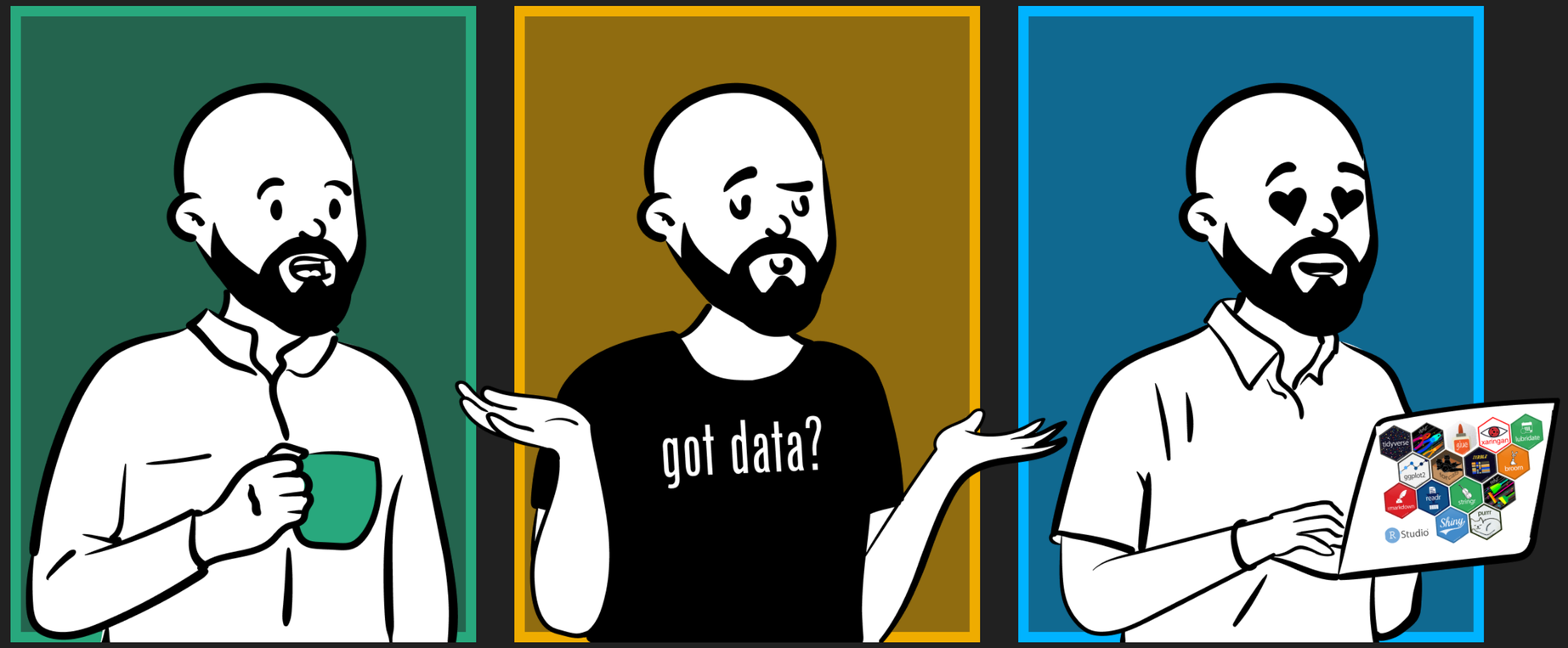
cedricscherer.com @CedScherer z3tt



A DataViz Training for cafeyn

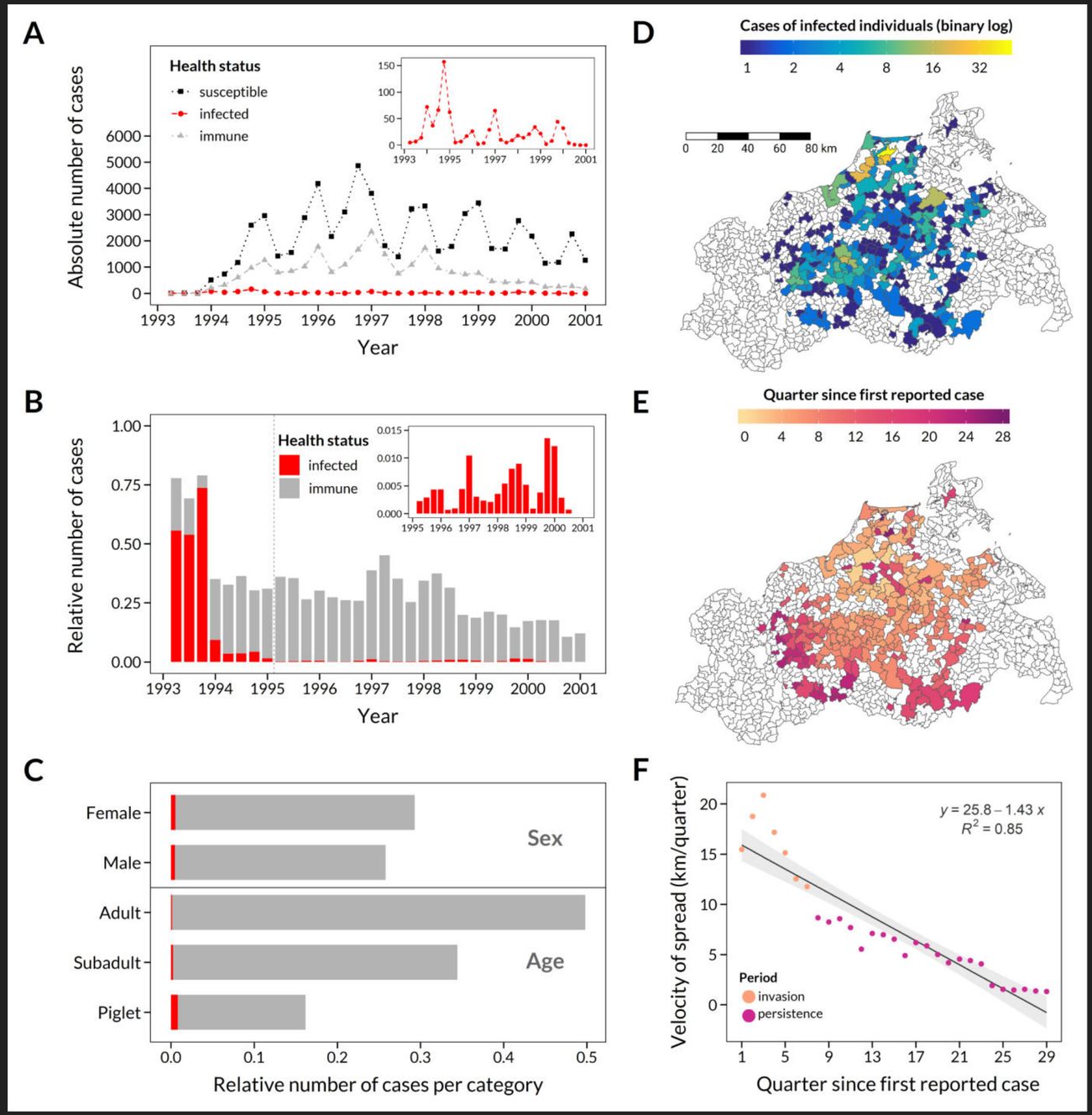
Cédric Scherer

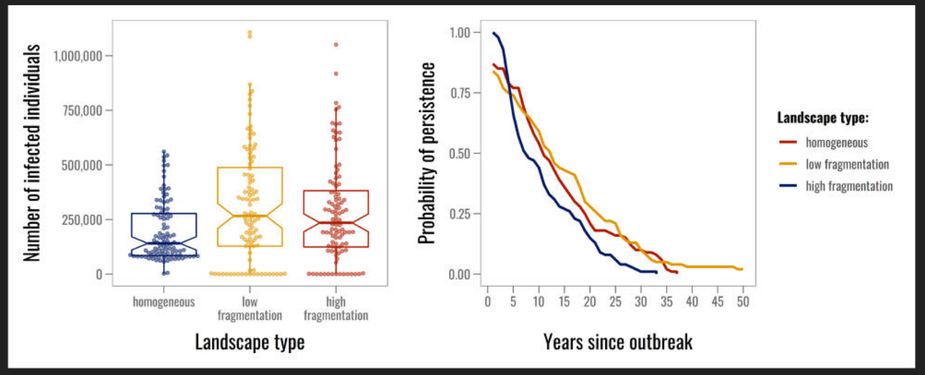
Independent Data Visualization Designer
Computational Ecologist at IZW Berlin

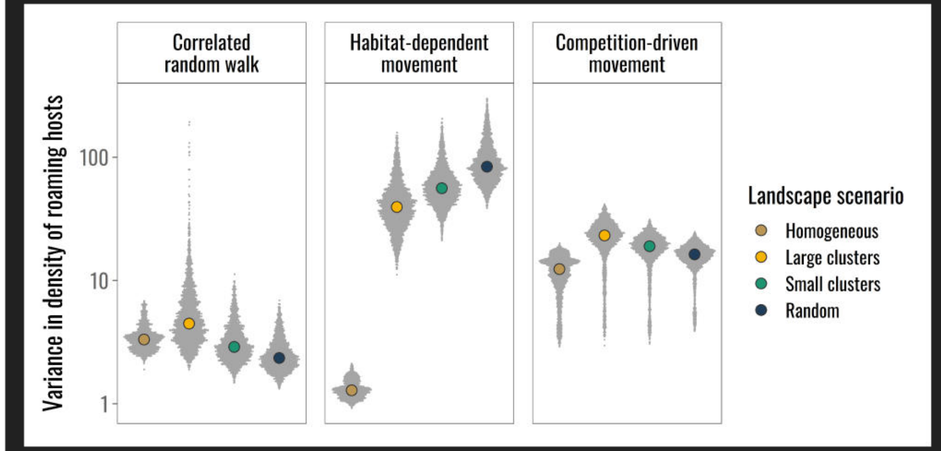


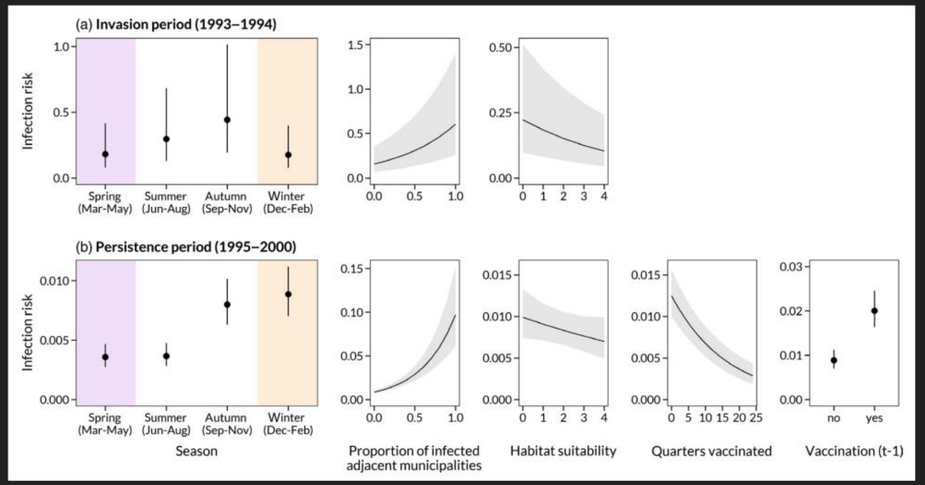
Consulting Coaching Coding

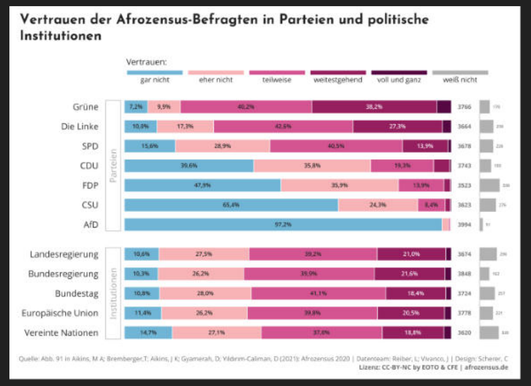
cedricscherer.com @CedScherer z3tt

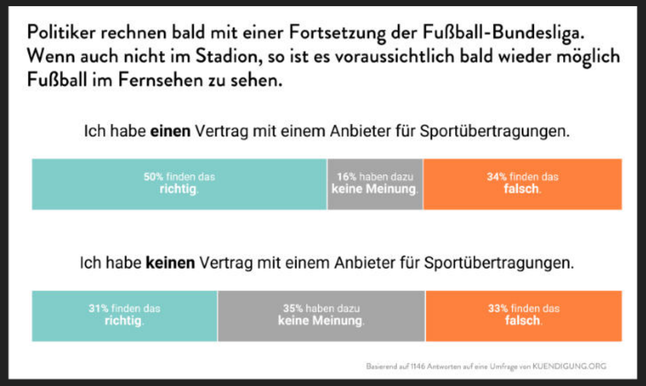

A
Health status
susceptible
infected
immune
Absolute number of cases
Year
B
Health status
infected
immune
Relative number of cases
Year
C
Female
Male
Adult
Subadult
Piglet
Sex
Age
Relative number of cases per category
D
Cases of infected individuals (binary log)
1 2 4 8 16 32
0 20 40 60 80 km
E
Quarter since first reported case
0 4 8 12 16 20 24 28
F
y = 25.8 – 1.43 x
R² = 0.85
Period
invasion
persistence
Velocity of spread (km/quarter)
Quarter since first reported case

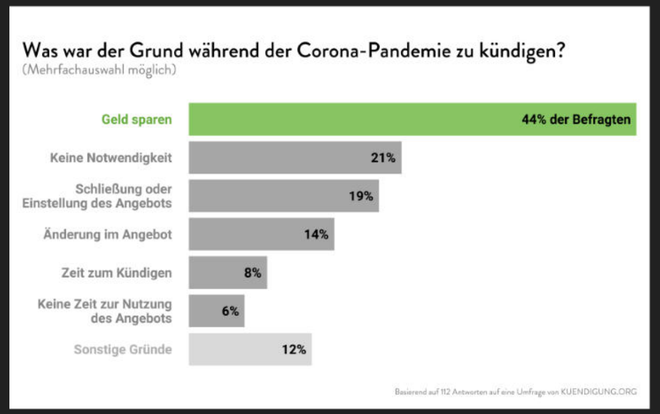

Number of infected individuals
homogeneous
low fragmentation
high fragmentation
Landscape type
Probability of persistence
Years since outbreak
Landscape type:
homogeneous
low fragmentation
high fragmentation

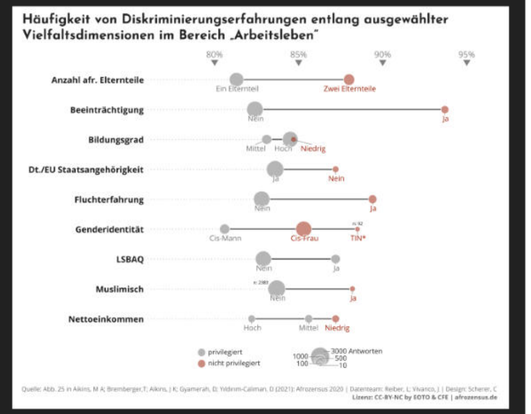

Correlated random walk
Habitat-dependent movement
Competition-driven movement
Variance in density of roaming hosts
Landscape scenario
Homogeneous
Large clusters
Small clusters
Random


(a) Invasion period (1993–1994)
(b) Persistence period (1995–2000)
Infection risk
Spring (Mar-May)
Summer (Jun-Aug)
Autumn (Sep-Nov)
Winter (Dec-Feb)
Season
Proportion of infected adjacent municipalities
Habitat suitability
Quarters vaccinated
Vaccination (t-1)
no
yes









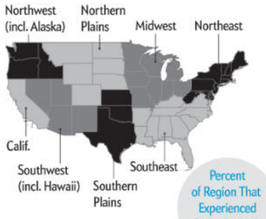
GRAPHIC SCIENCE

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

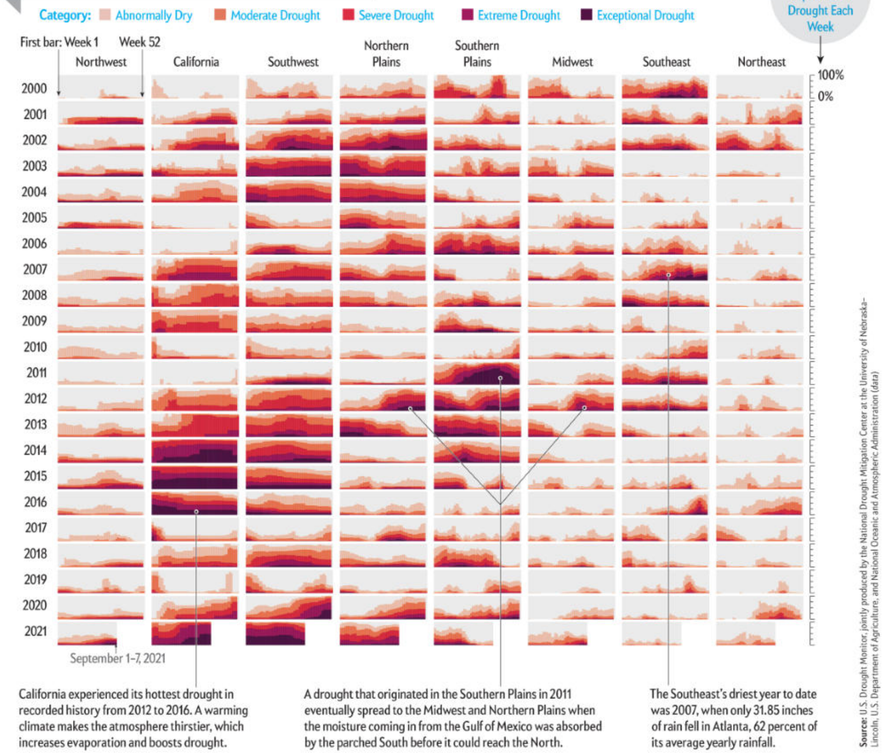
Escalating Drought

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

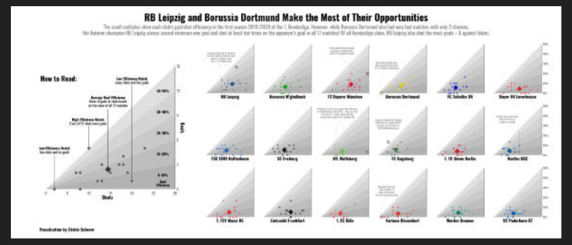


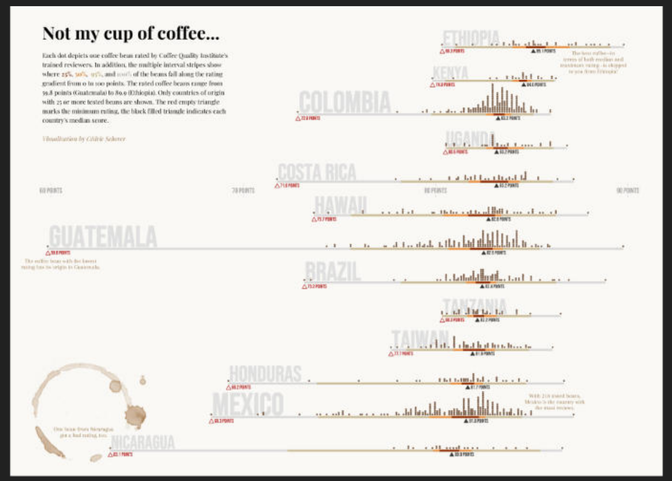
California experienced its hottest drought in recorded history from 2012 to 2016. A warming climate makes the atmosphere thirstier, which increases evaporation and boosts drought.

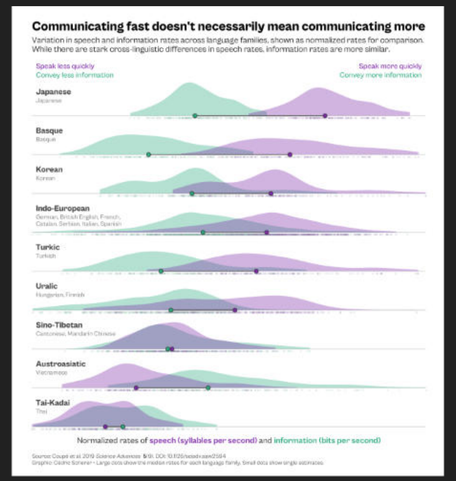
A drought that originated in the Southern Plains in 2011 eventually spread to the Midwest and Northern Plains when the moisture coming in from the Gulf of Mexico was absorbed by the parched South before it could reach the North.

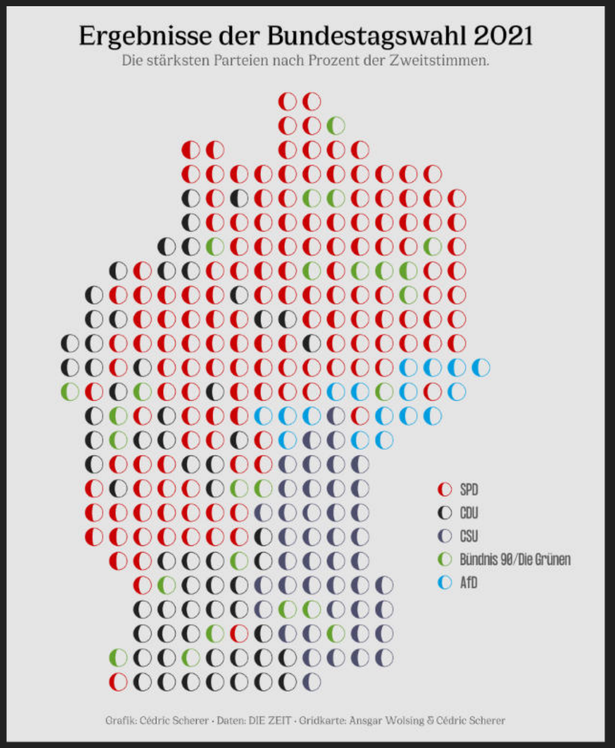
The Southeast's driest year to date was 2007, when only 31.85 inches of rain fell in Atlanta, 62 percent of its average yearly rainfall.

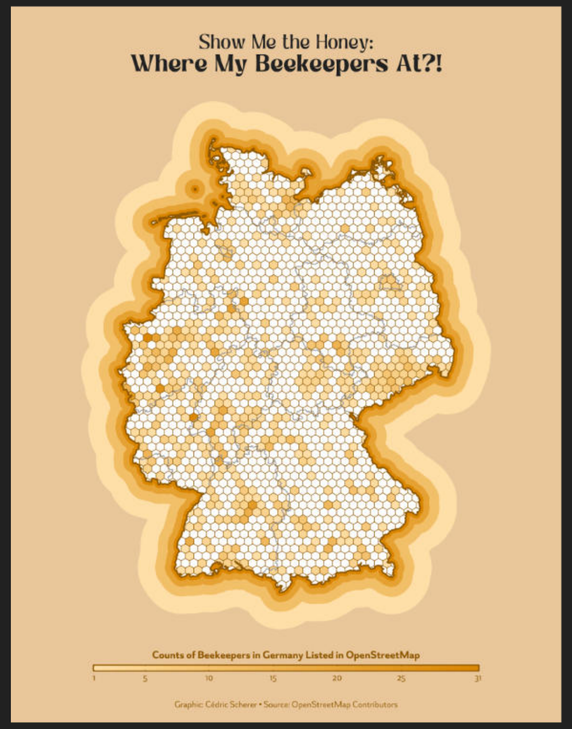
Source: U.S. Drought Monitor, jointly produced by the National Drought Mitigation Center at the University of Nebraska–Lincoln, U.S. Department of Agriculture, and National Oceanic and Atmospheric Administration (data)

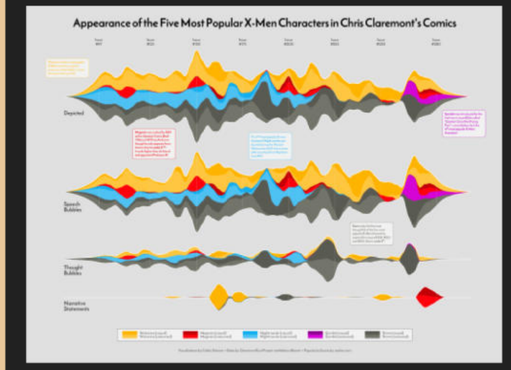


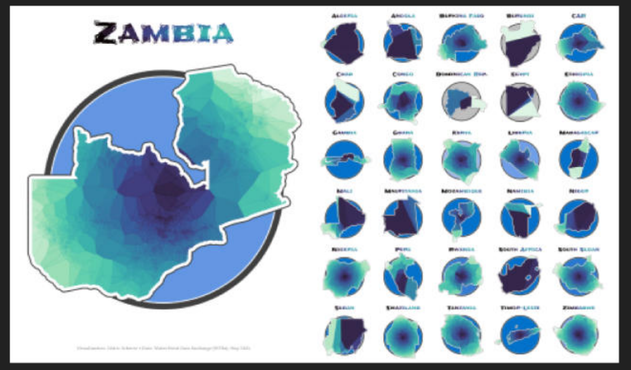


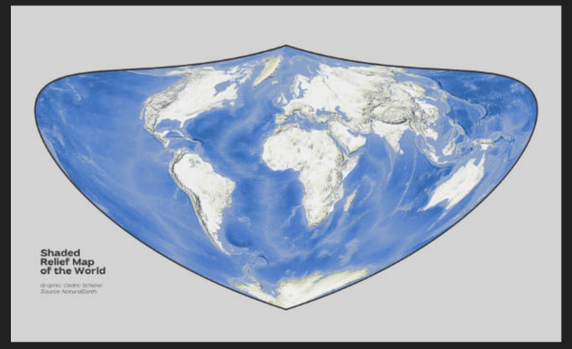


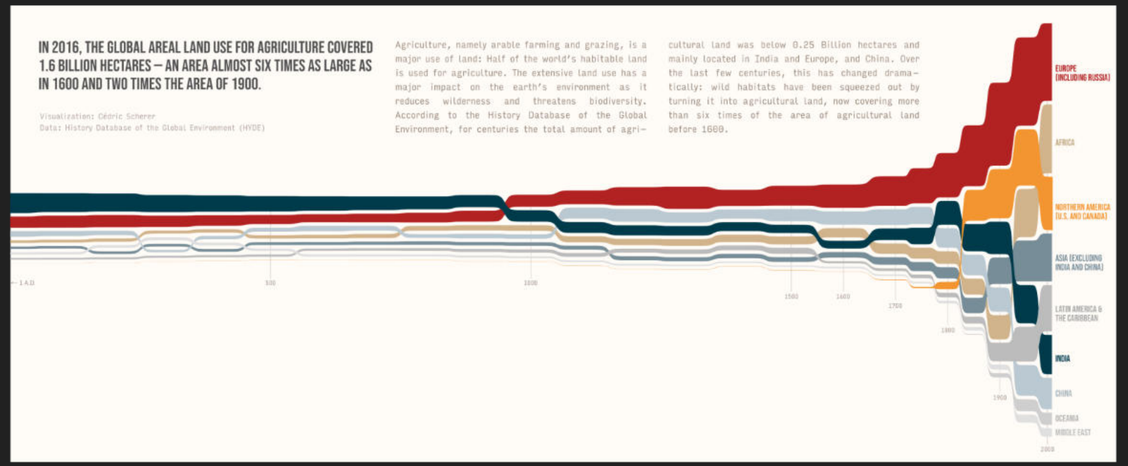


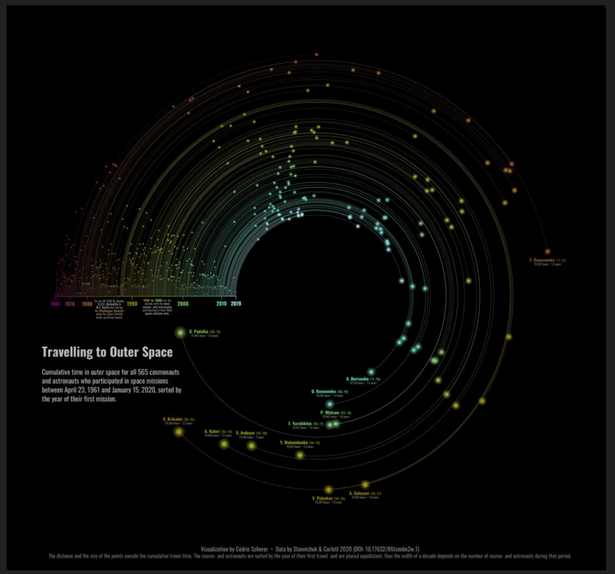












cedricscherer.com @CedScherer z3tt

Slides

cedricscherer.com/workshops/cafeyn_dataviz.pdf

Data Visualization

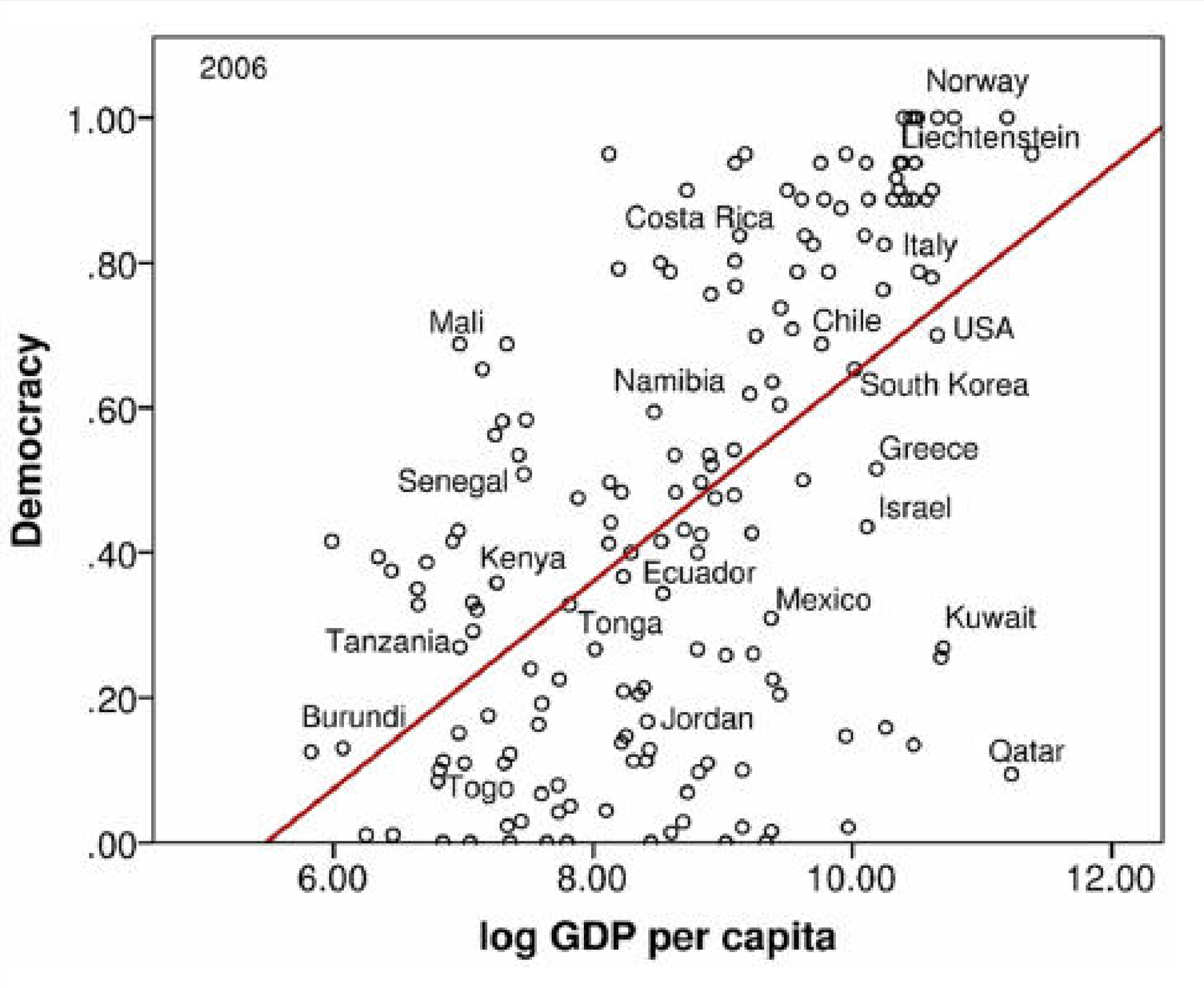
is any graphical representation
of information and data.

Data Visualization

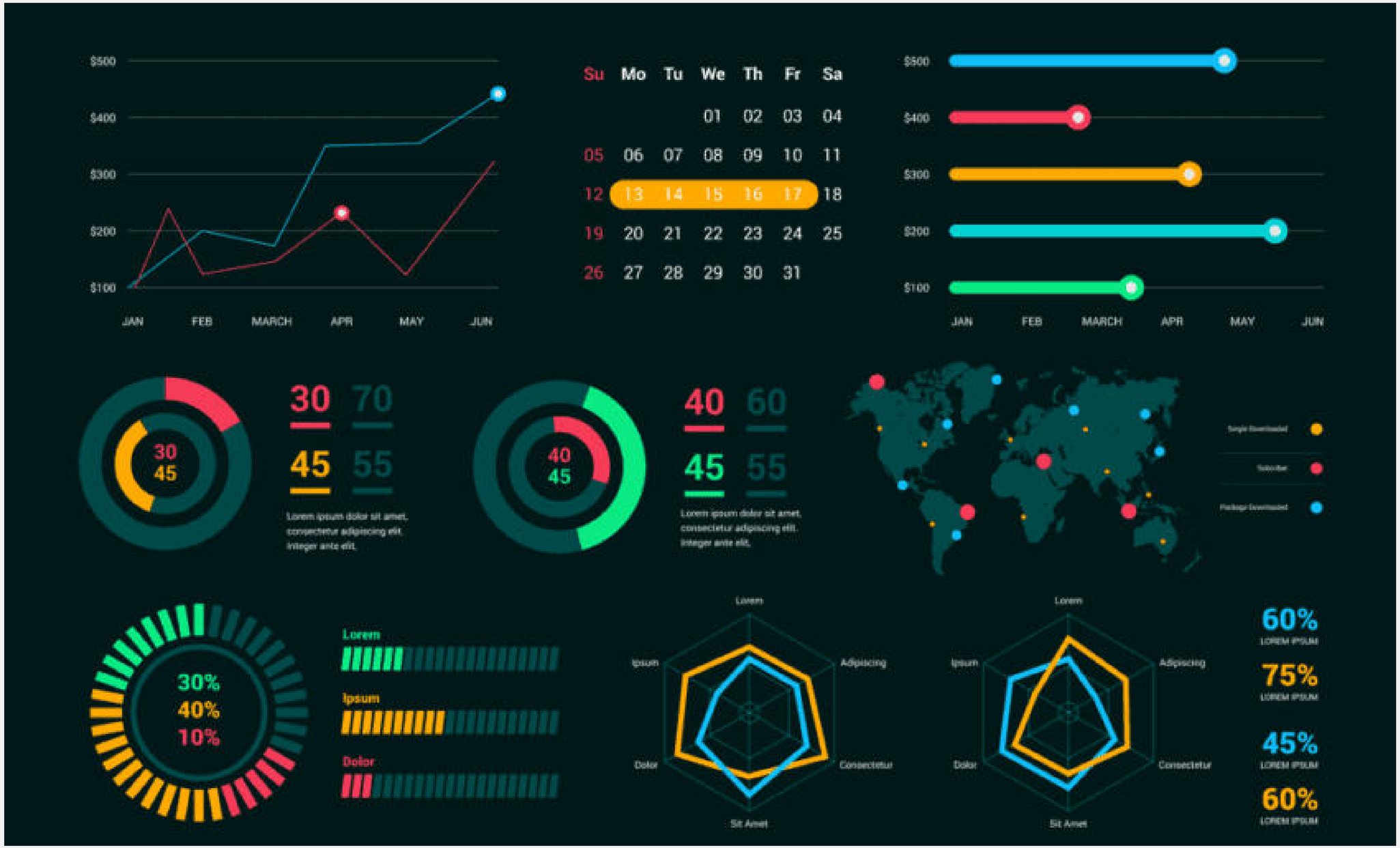
helps to amplify cognition, gain insights, discover, explain, and make decisions.

Data Visualization

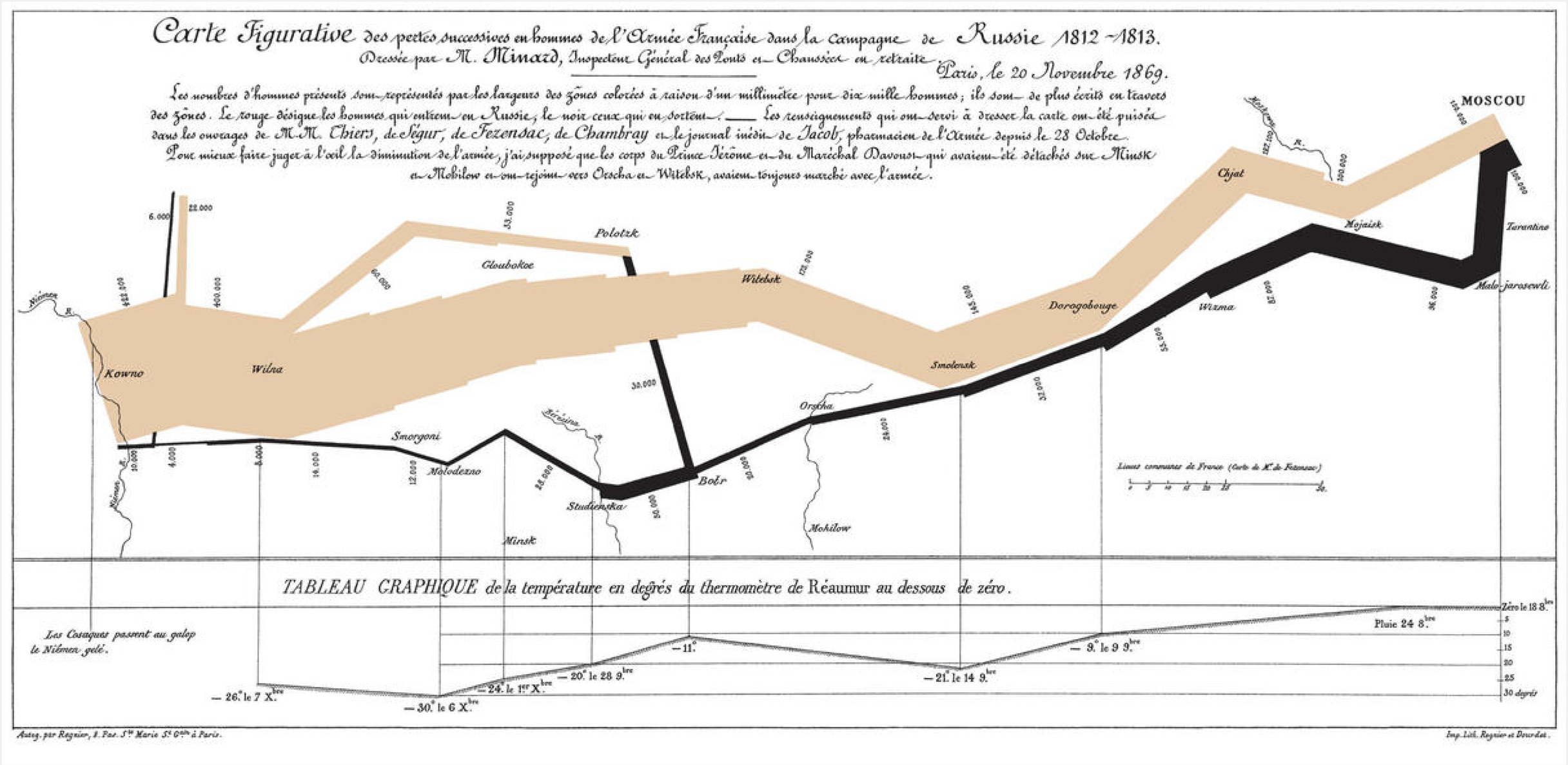
converts information into visual forms as quantifiable features.



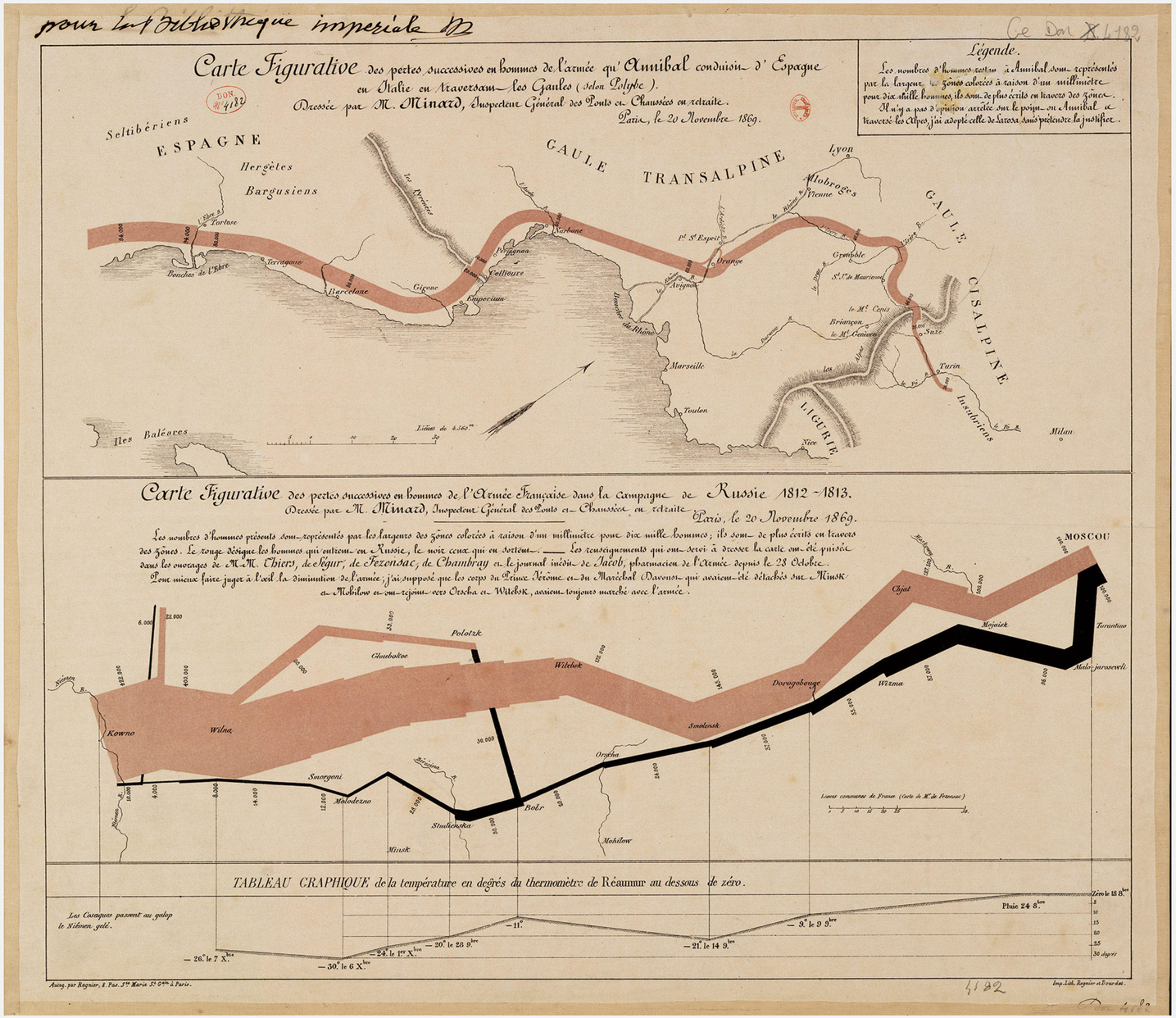
Source: Ranganathan et al. 2014



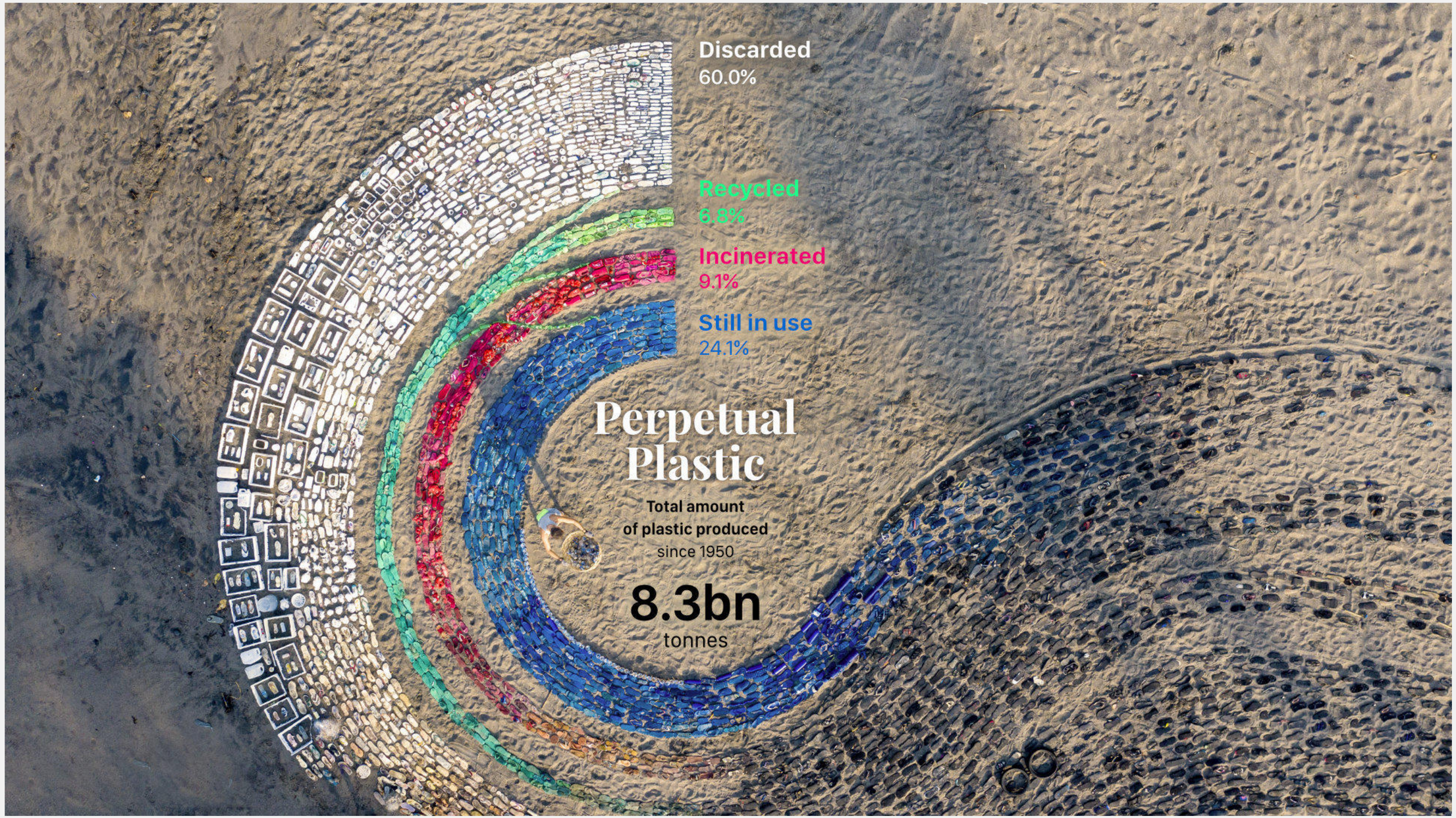
Source: datameer.com



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

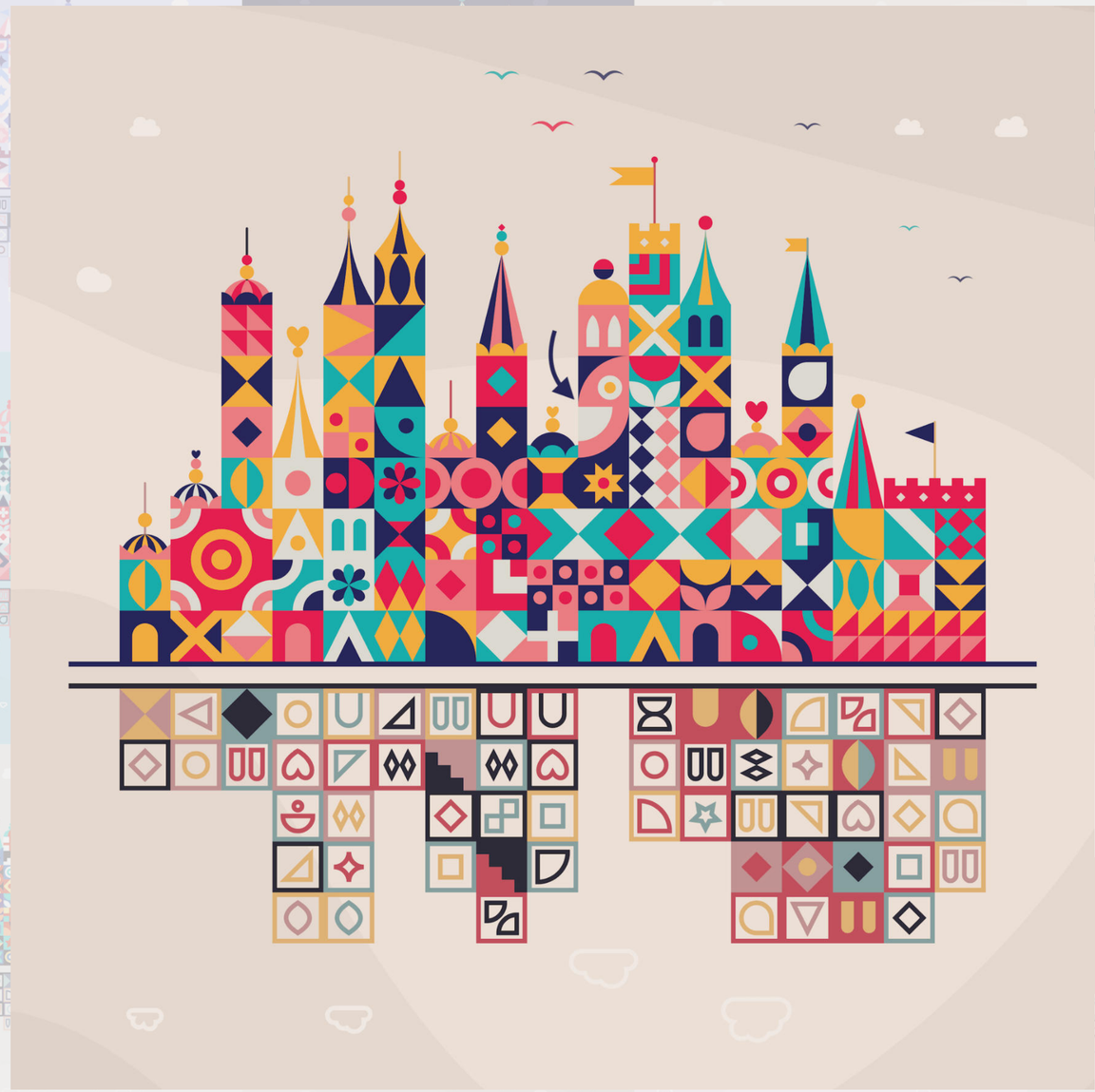


Carte figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812–1813 and Carte figurative des pertes successives en hommes de l'Armée qu'Annibal conduisit d'Espagne en Italie en traversant les Gaules (selon Polybe) by Charles Joseph Minard



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

Source: “Patchwork Kingdoms” by Nadieh Bremer

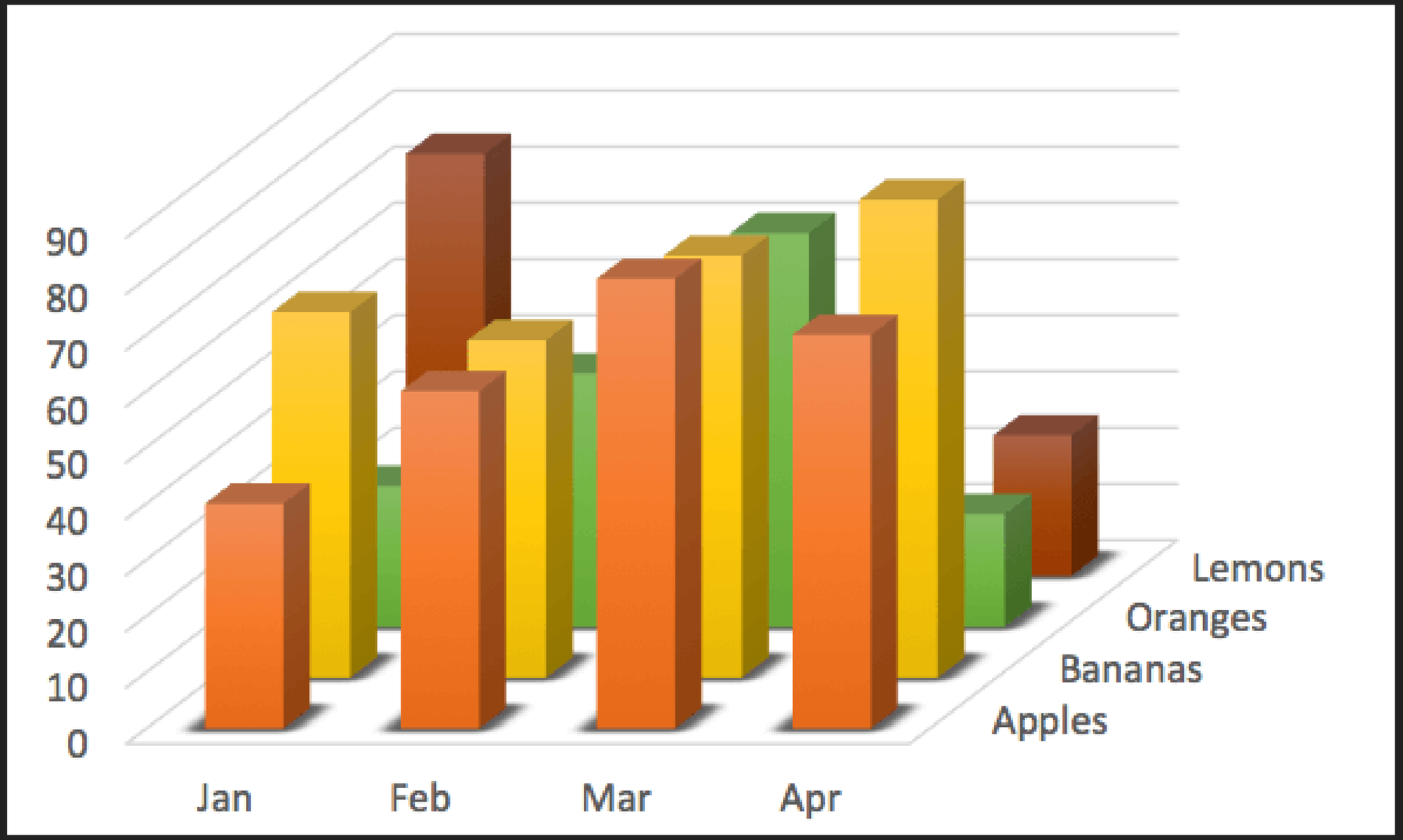


schools **connected**
to the internet

schools **not connected**
to the internet

Source: “Patchwork Kingdoms” by Nadieh Bremer

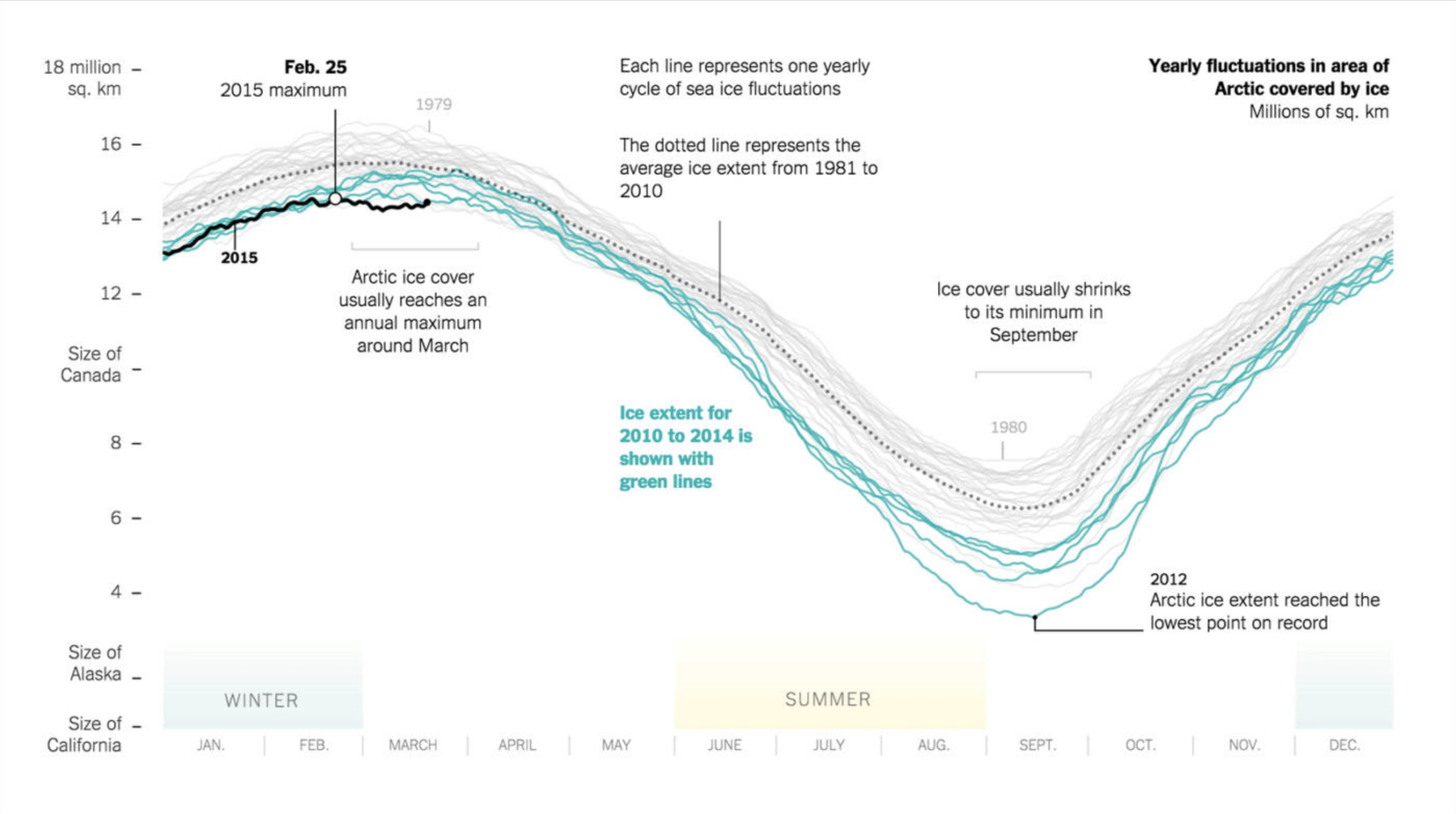
What makes it a bad data visualization?



What makes it a bad data visualization?

- 👉 **substantive** problems (bad data / story)
- 👉 **aesthetic** problems (bad design)
- 👉 **perceptual** problems (bad encoding)

What makes it a good data visualization?



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

What makes it a good data visualization?

👉 **INFORMATION** (integrity)

What makes it a good data visualization?

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

What makes it a good data visualization?

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

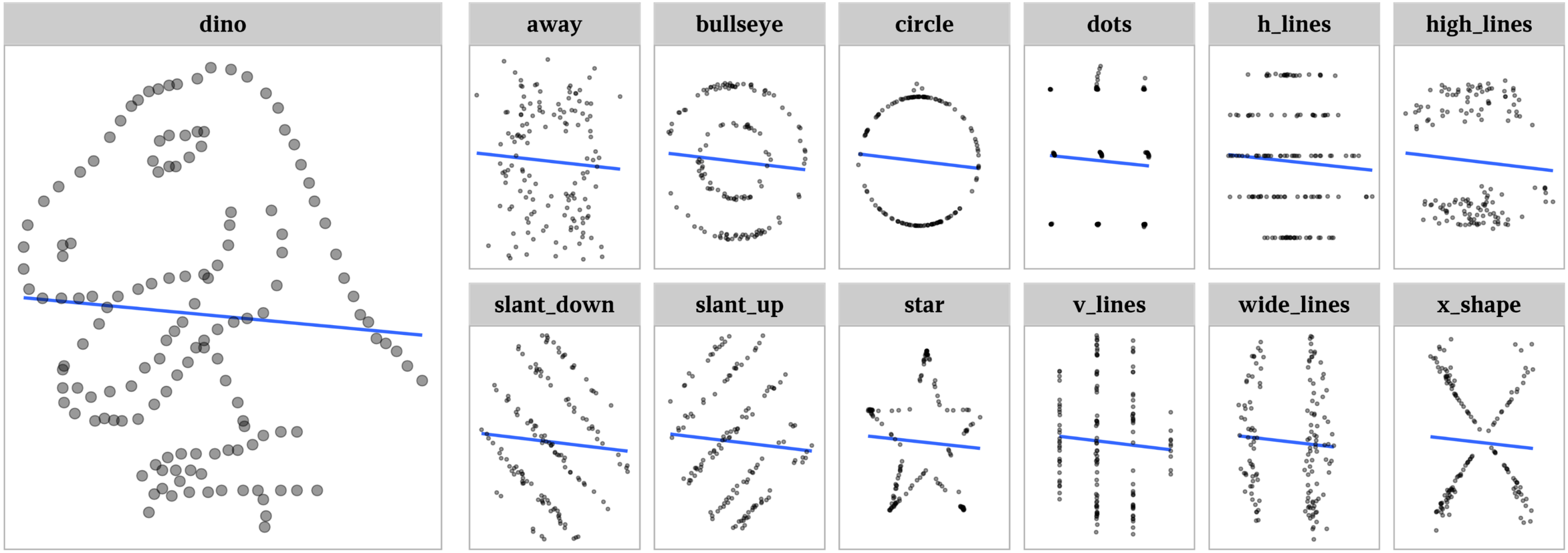
What makes it a good data visualization?

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

INFORMATION

Understand your data and be accurate

Visualize Your Data

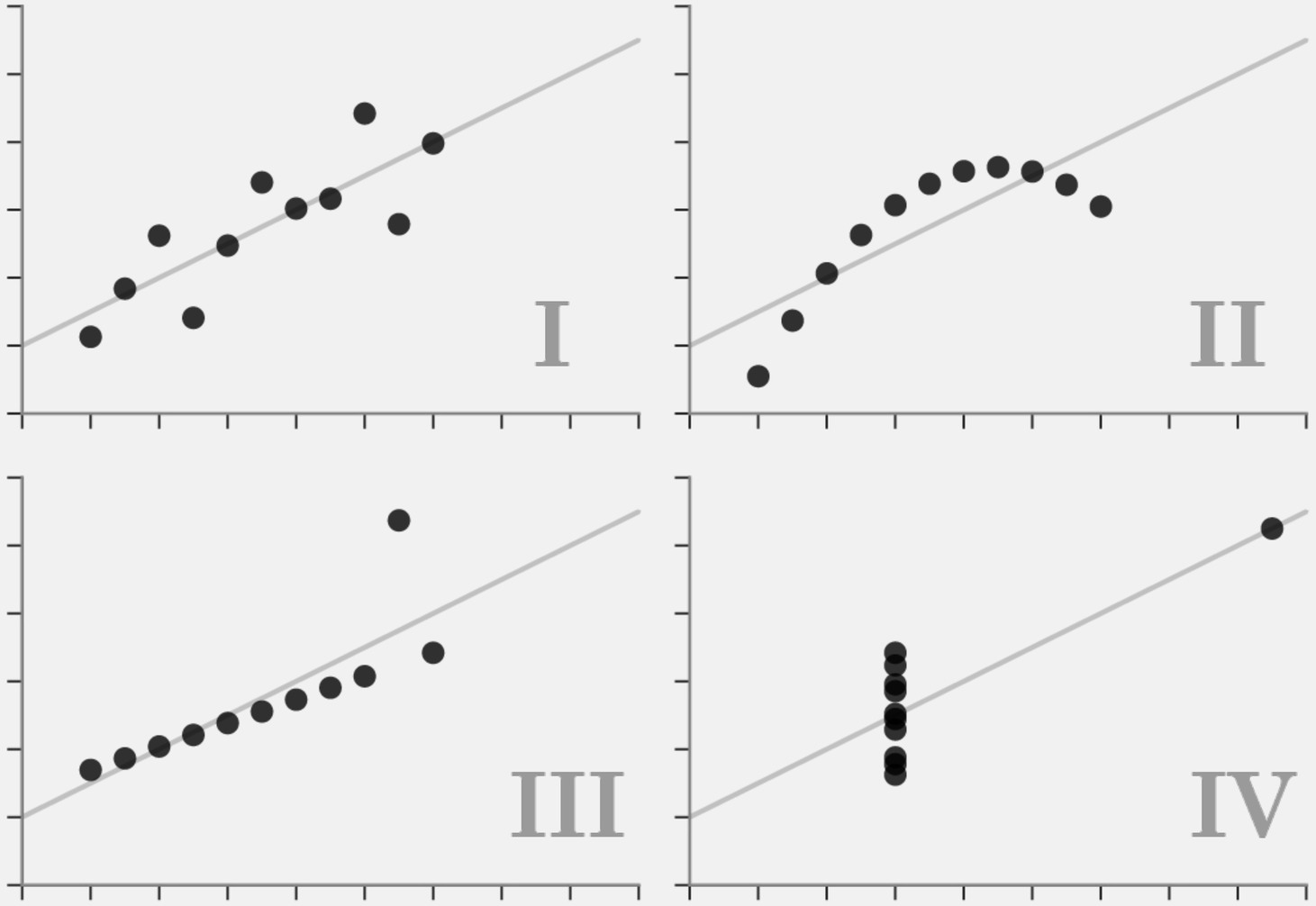


"Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing"
by Justin Matejka & George Fitzmaurice, ACM SIGCHI Conference on Human Factors in Computing Systems 2017

cedricscherer.com @CedScherer z3tt

Anscombe's Quartet

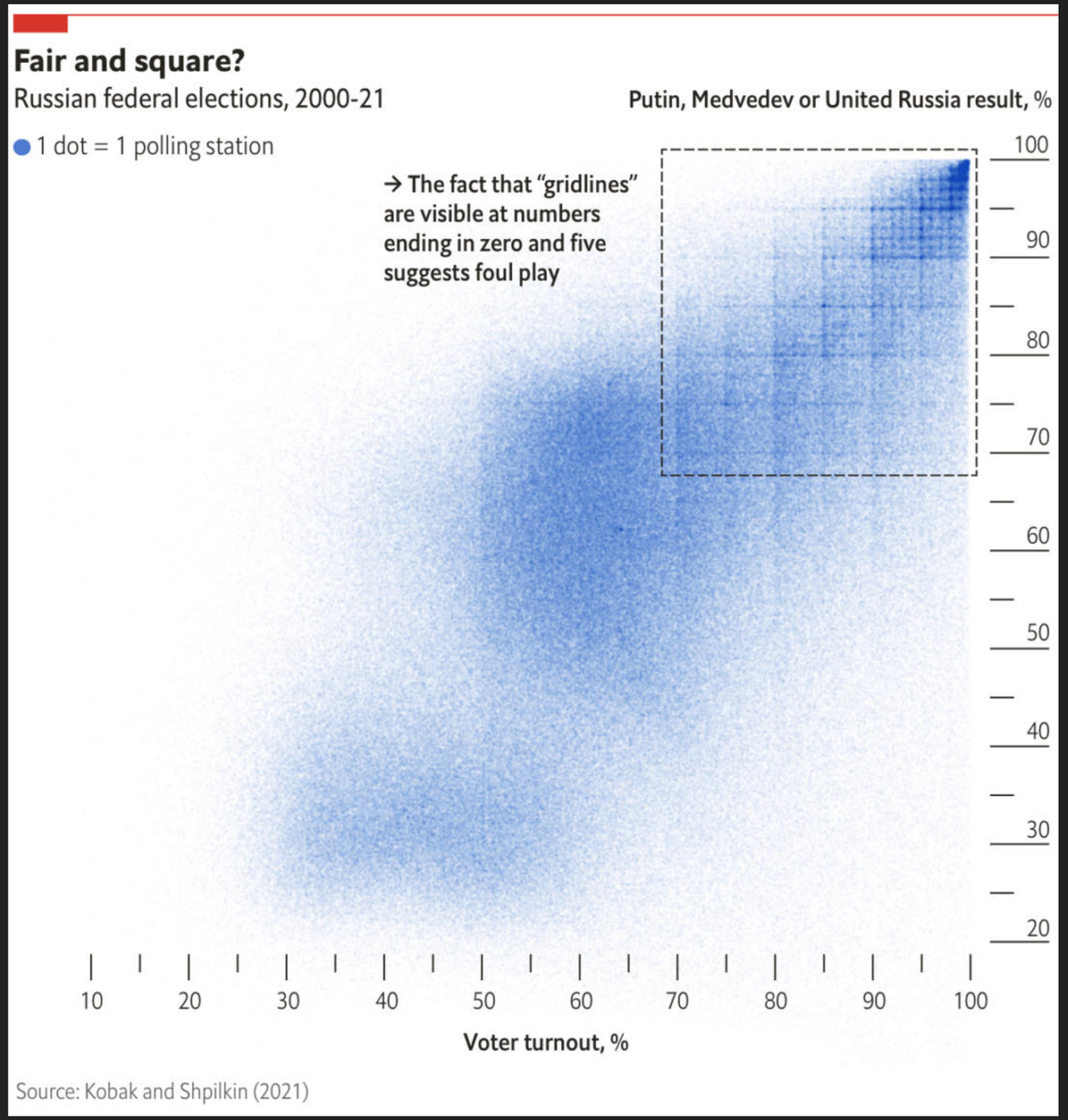
Each dataset has the same summary statistics
mean, standard deviation, and correlation
but are visually distinct.



"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

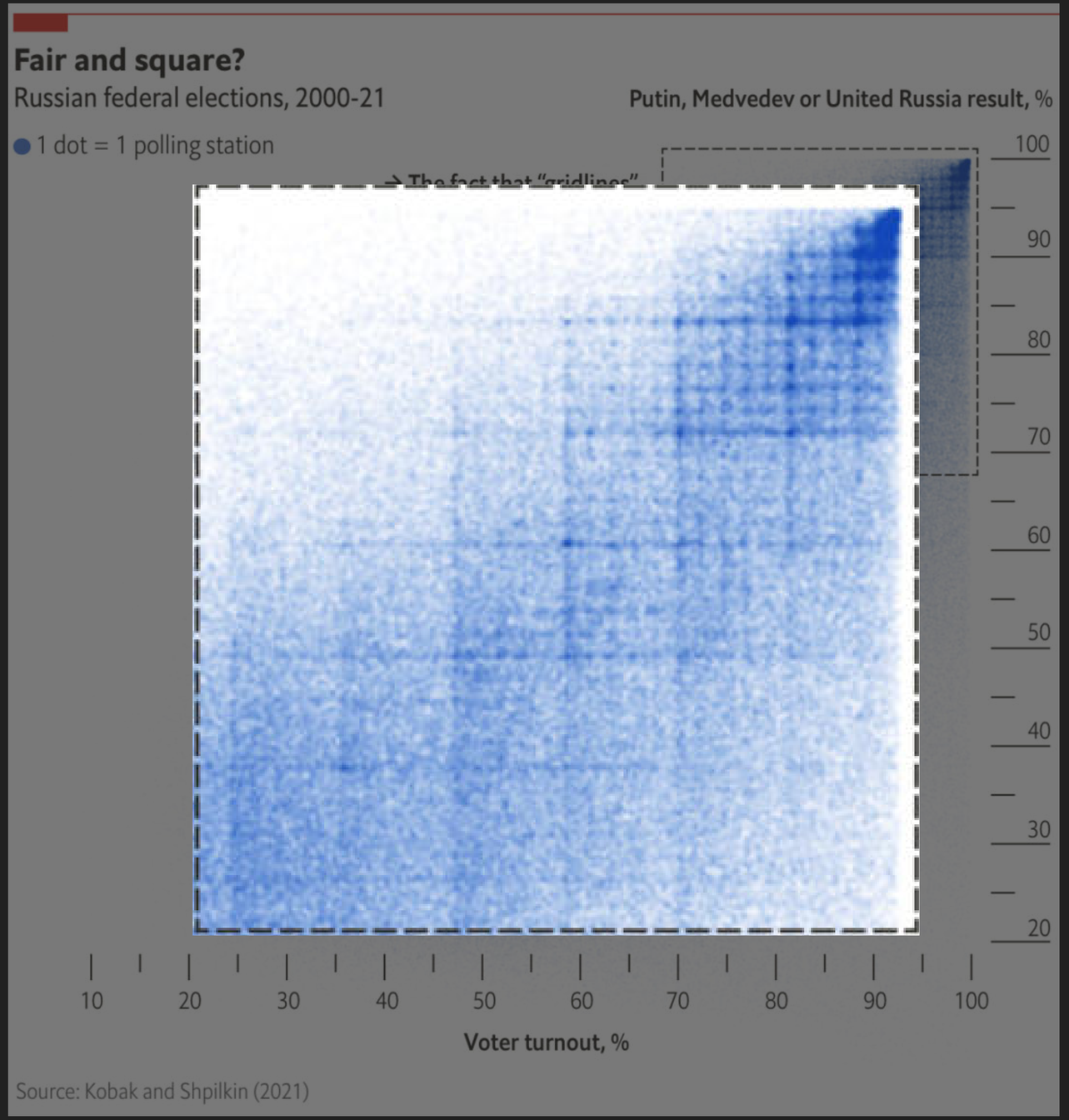
Visualize Your Data

"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**."



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**."



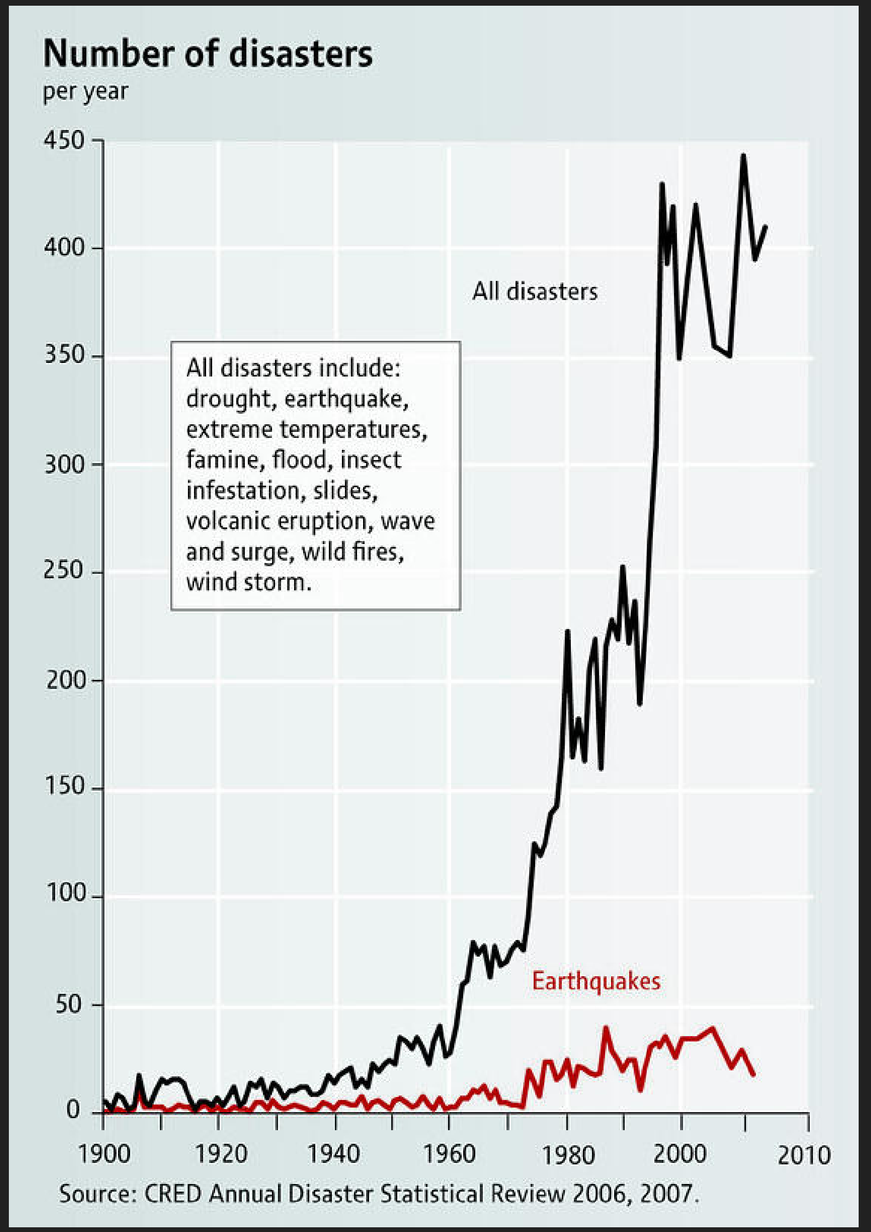
Our data is **never** a perfect reflection of the real world.

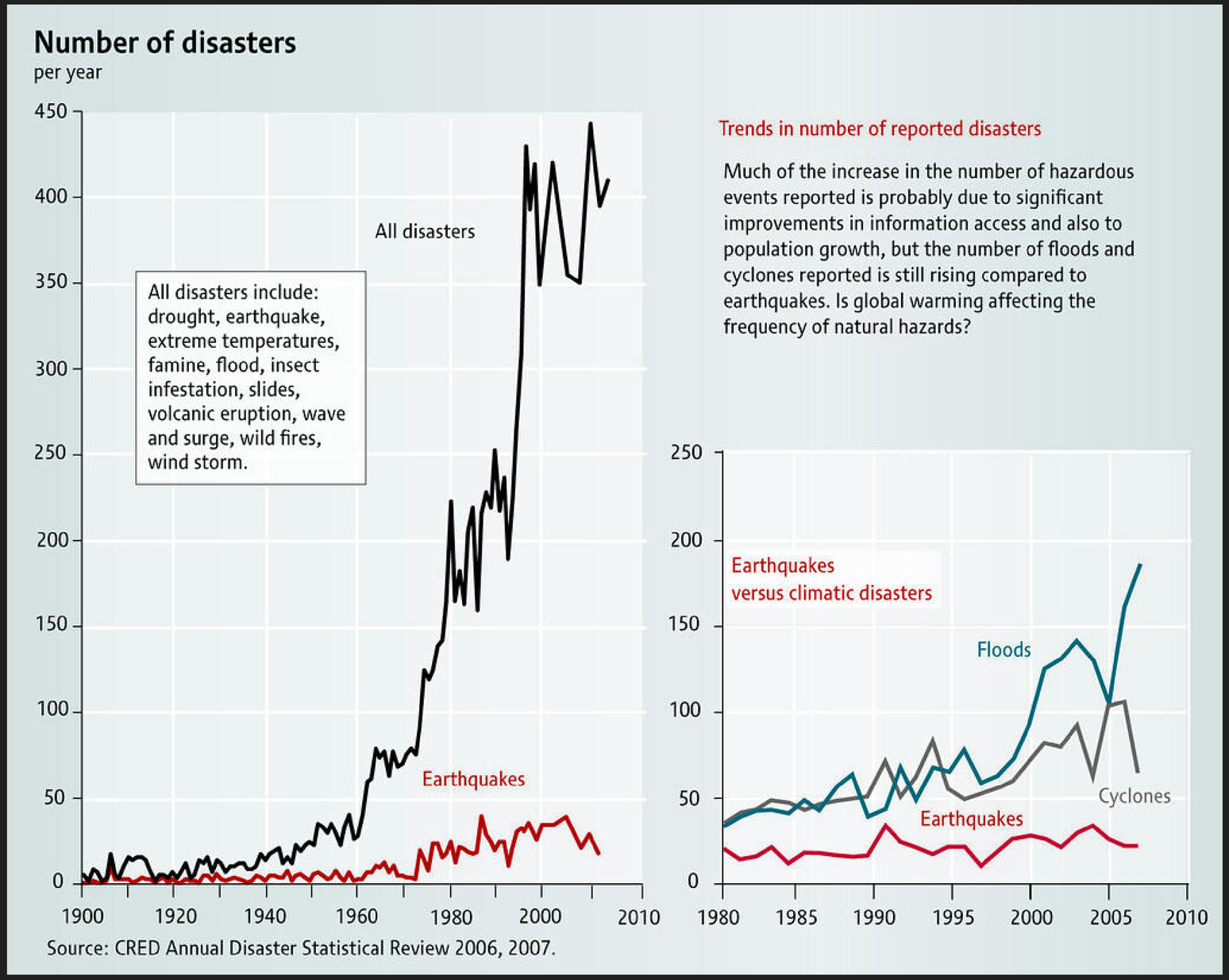
Our data is **never** a perfect reflection of the real world.

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

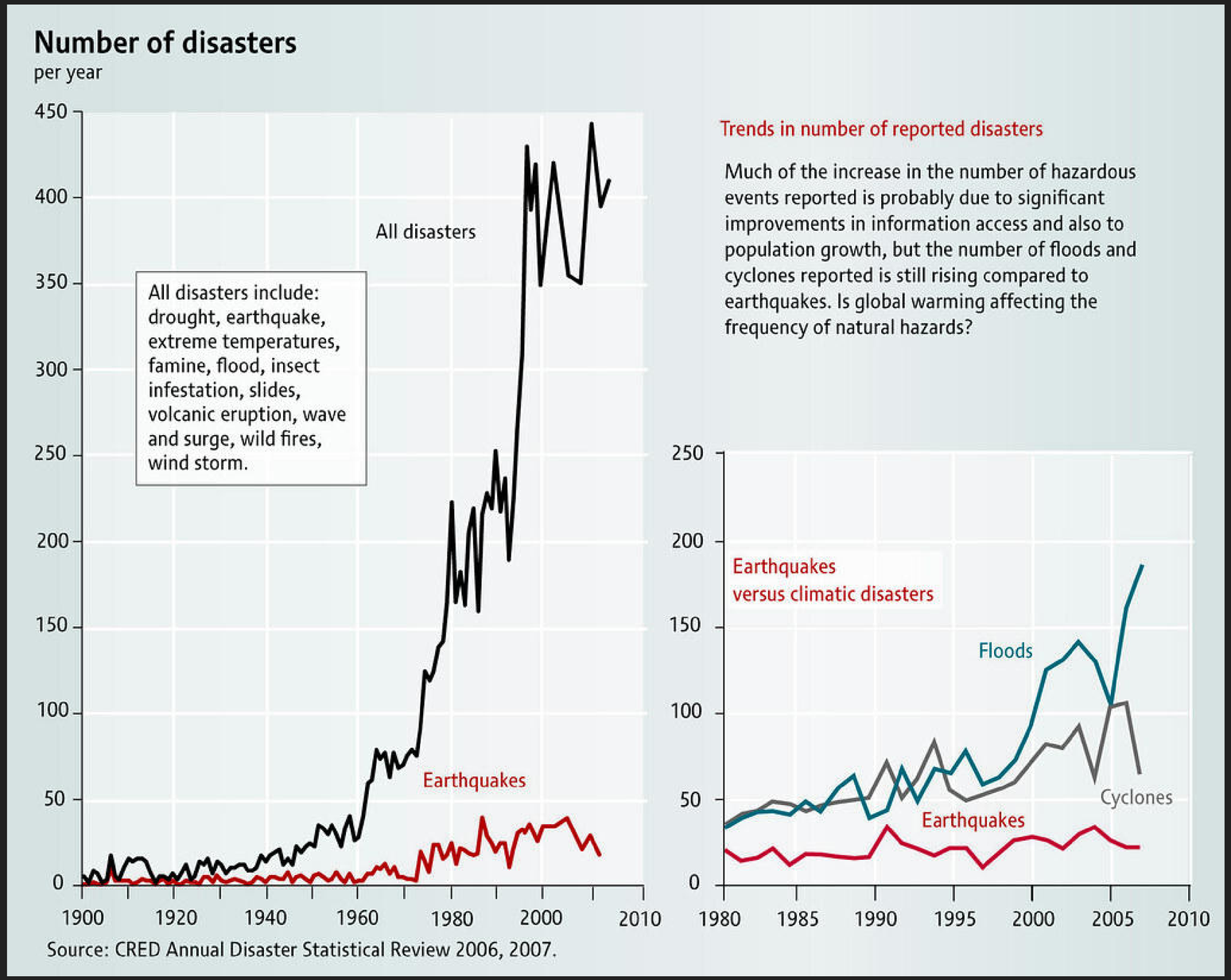
→ **collected by humans:** guesstimation, precision and errors

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

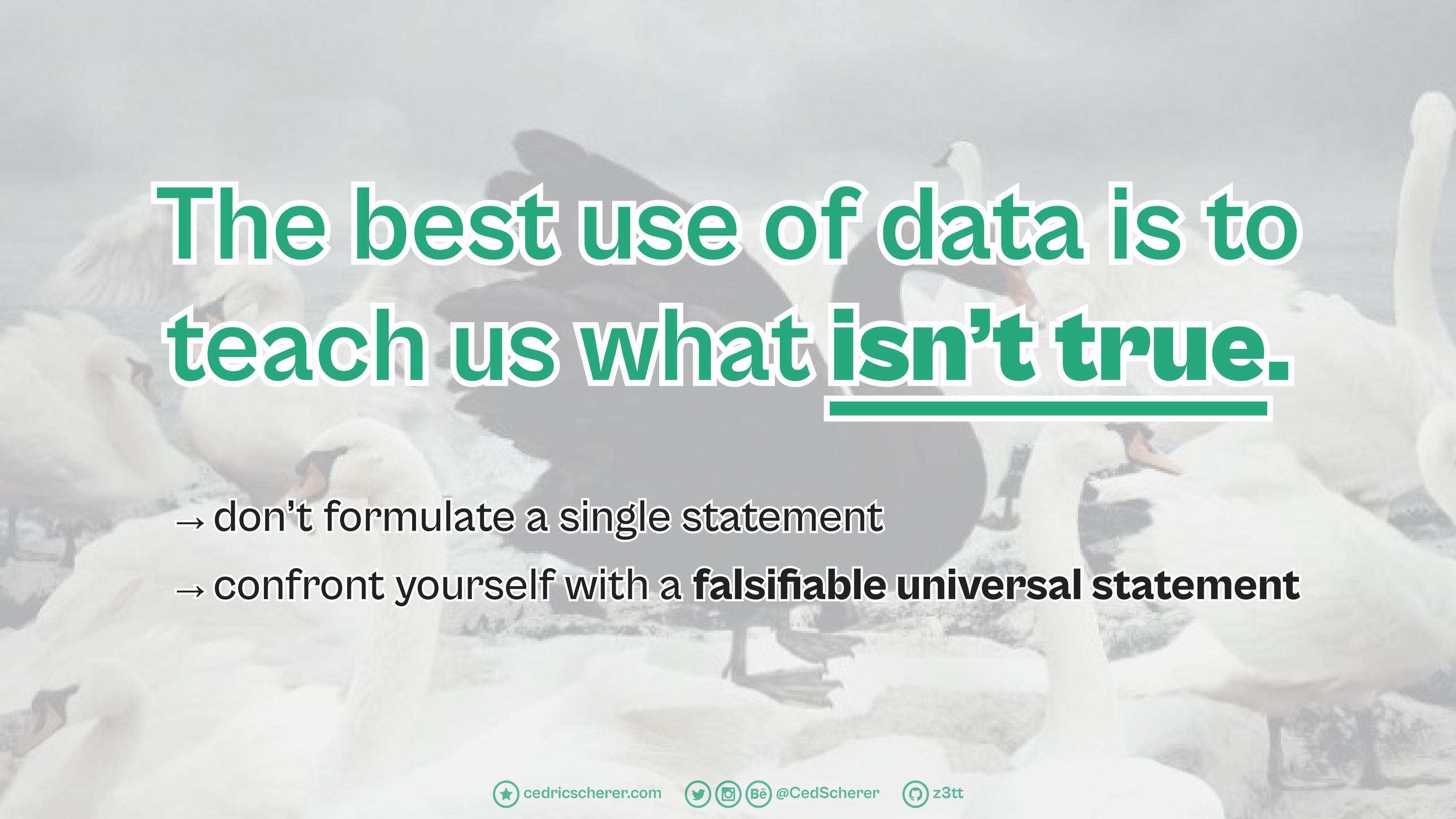

Number of disasters
per year
450
400
350
300
250
200
150
100
50
0
All disasters
All disasters include: drought, earthquake, extreme temperatures, famine, flood, insect infestation, slides, volcanic eruption, wave and surge, wild fires, wind storm.
Earthquakes
1900
1920
1940
1960
1980
2000
2010
Source: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


Number of disasters
per year
450
400
350
300
250
200
150
100
50
0
All disasters
All disasters include: drought, earthquake, extreme temperatures, famine, flood, insect infestation, slides, volcanic eruption, wave and surge, wild fires, wind storm.
Earthquakes
1900
1920
1940
1960
1980
2000
2010
Trends in number of reported disasters
Much of the increase in the number of hazardous events reported is probably due to significant improvements in information access and also to population growth, but the number of floods and cyclones reported is still rising compared to earthquakes. Is global warming affecting the frequency of natural hazards?
Earthquakes versus climatic disasters
250
200
150
100
50
0
Floods
Cyclones
Earthquakes
1980
1985
1990
1995
2000
2005
2010
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"**



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



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

→ don't formulate a single statement

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

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

→ single statement: "The swan is white."

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

STORY

Be clear about the message of your graphic

Typology of Information Graphics

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

Is the **information** conceptual or measurable?

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

Typology of Information Graphics

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

Is the **information** conceptual or measurable?

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

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

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

”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.”**

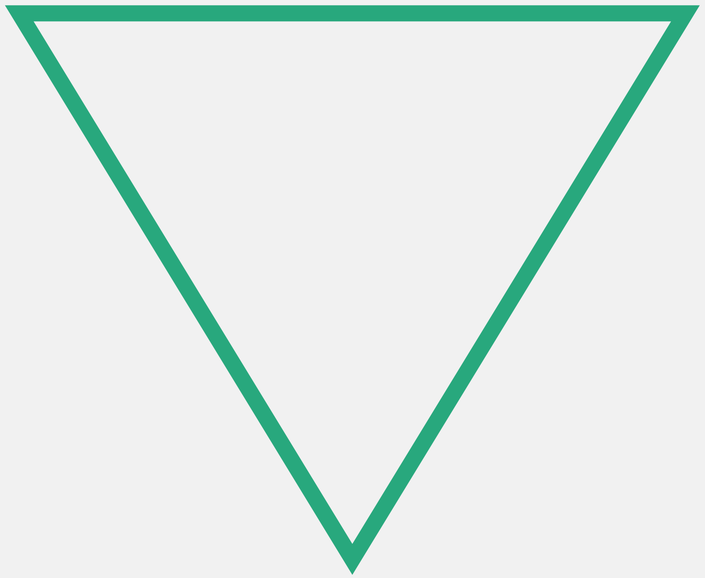
Alberto Cairo

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

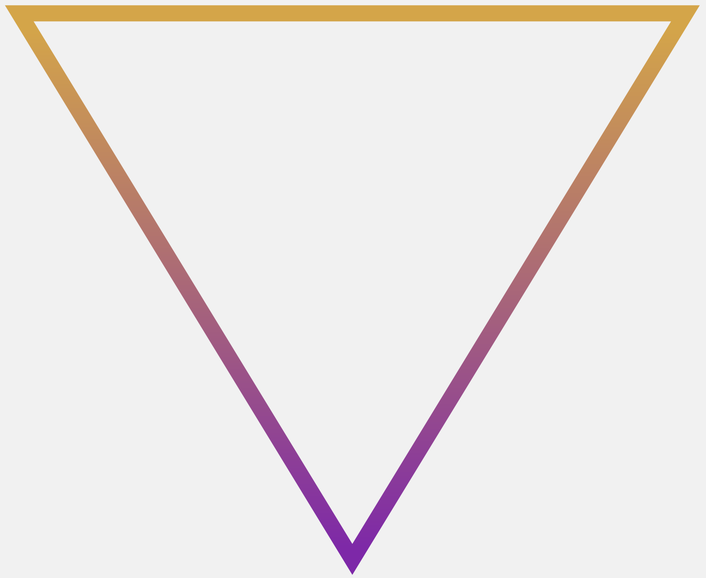
Aim:
functional

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)

Who is my audience?

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

Who is my audience?

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

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

Who is my audience?

 Which story is **interesting** for them?

 Which variables are **meaningful** to them?

 What are **relevant** details to include?

Who is my 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 my audience?

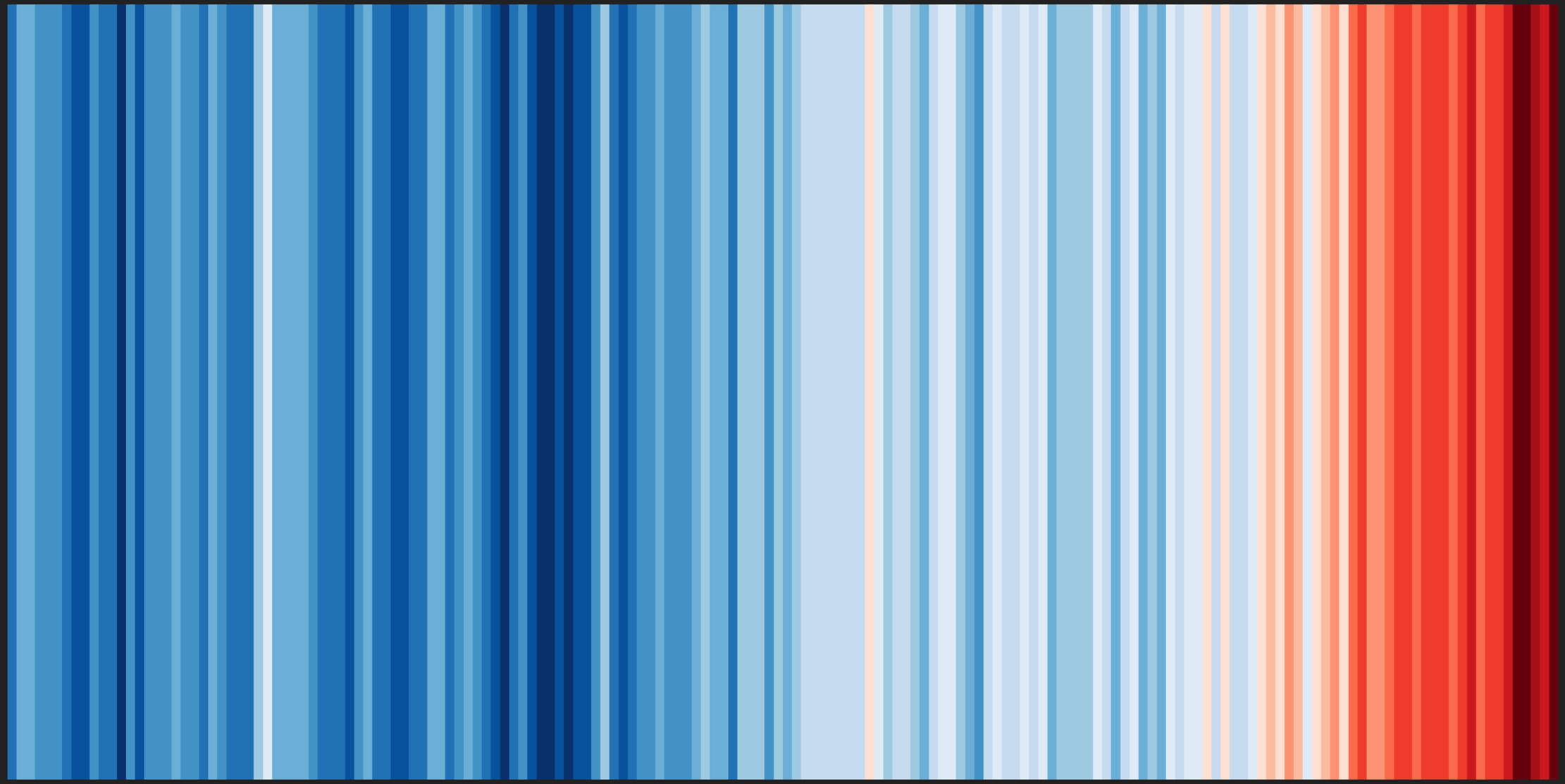
 Which story is **interesting** for them?

 Which variables are **meaningful** to them?

 What are **relevant** details to include?

 How will they **encounter** the visualization?

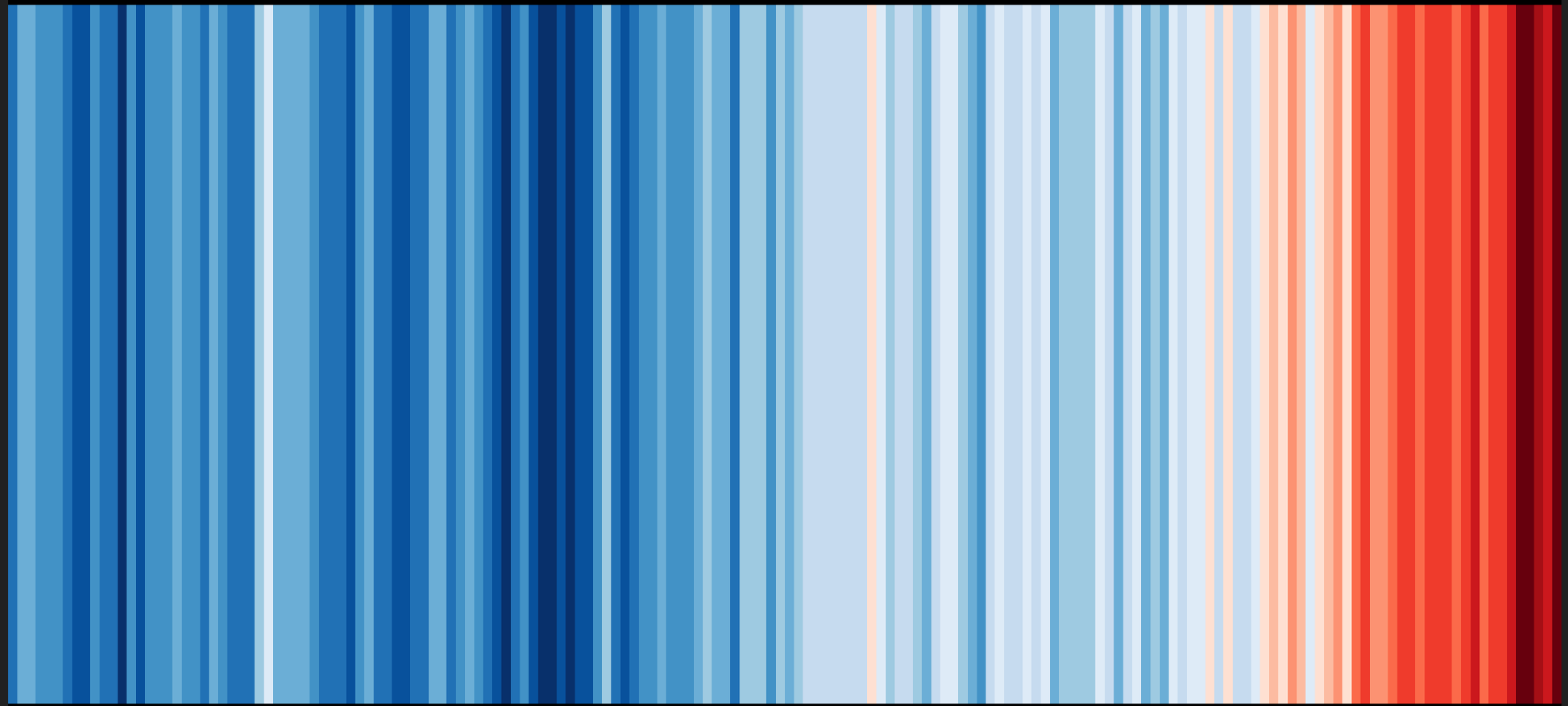
 **Do I need a visualization at all?**

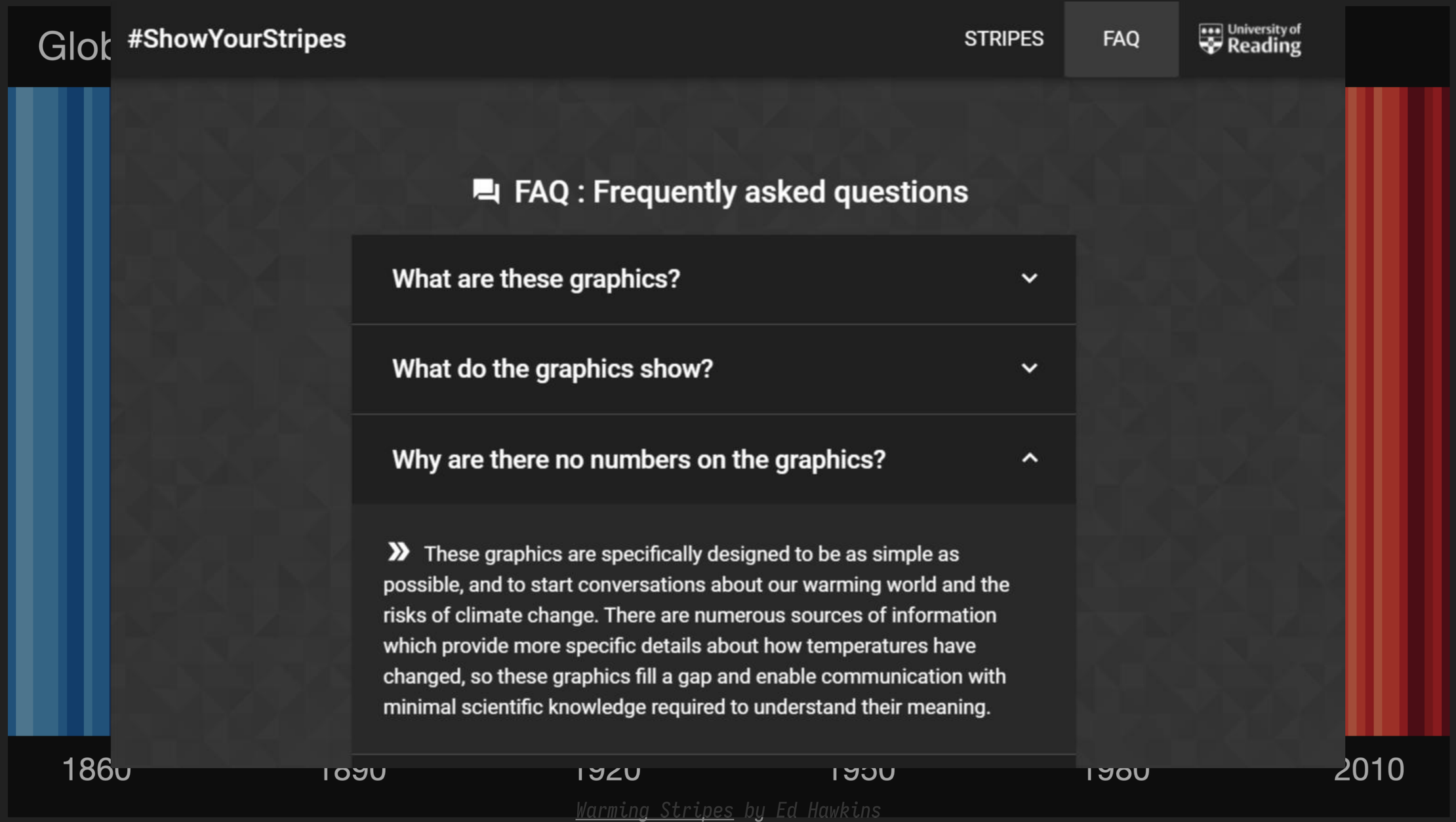


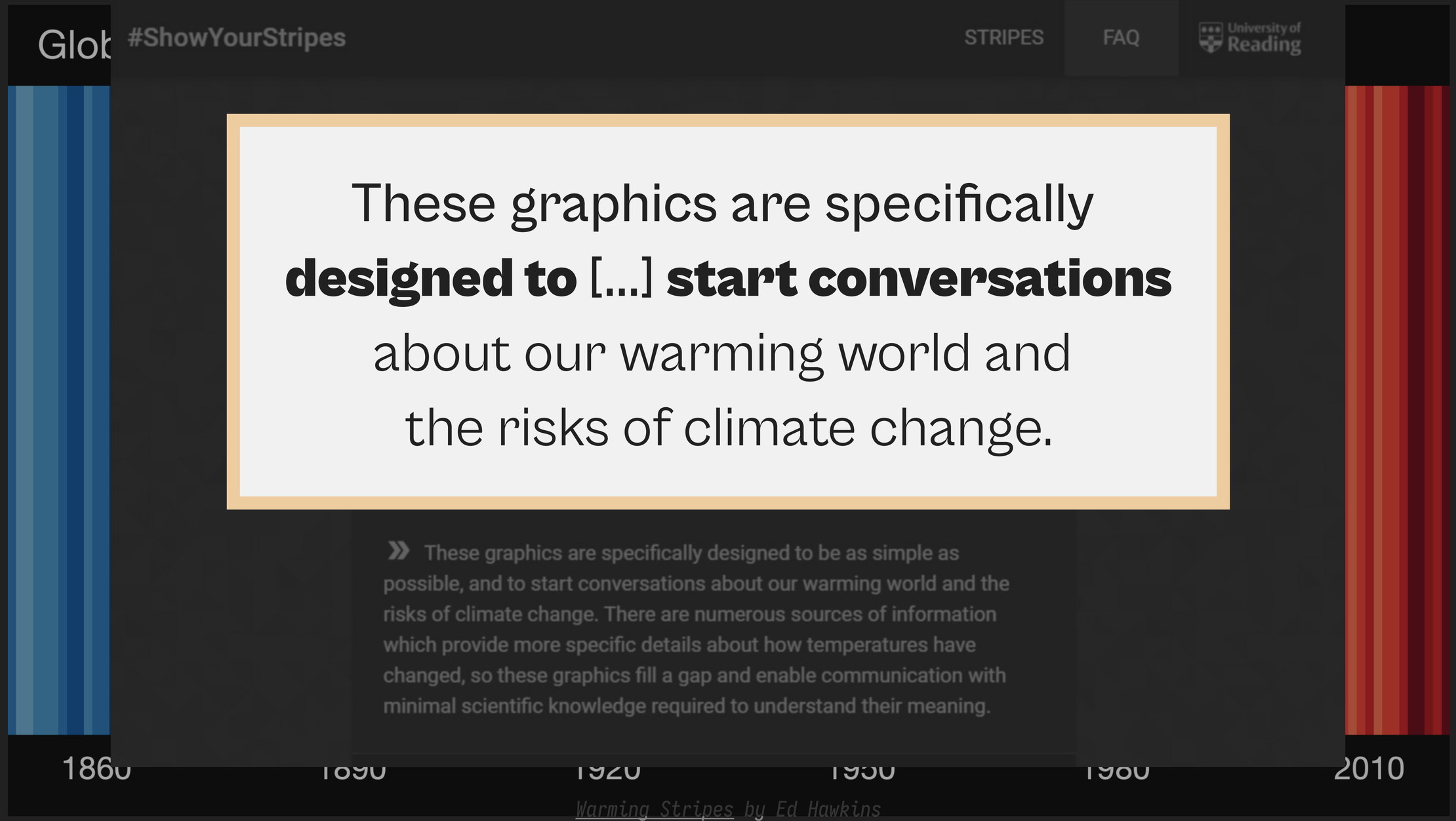
Warming Stripes by Ed Hawkins

cedricscherer.com @CedScherer z3tt

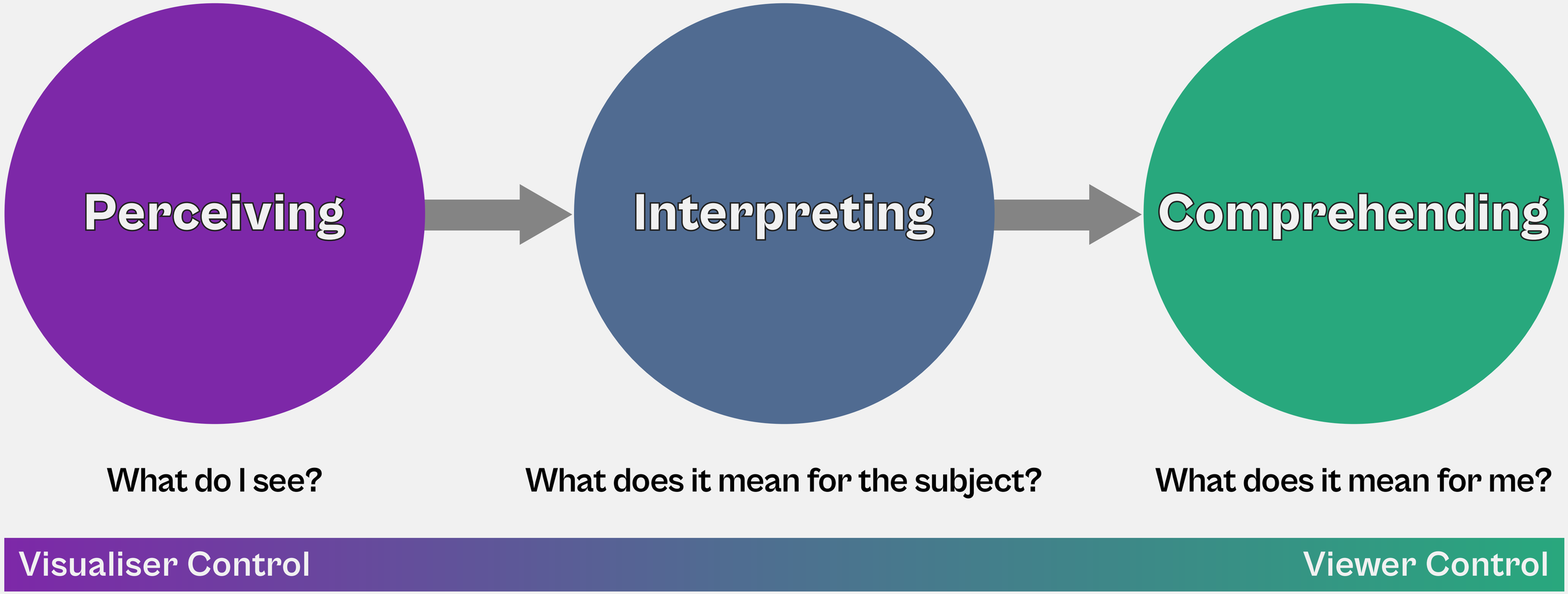
Global temperature change (1850-2019)




Glob
#ShowYourStripes
STRIPES
FAQ
University of Reading
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.
1860
1890
1920
1950
1980
2010
Warming Stripes by Ed Hawkins



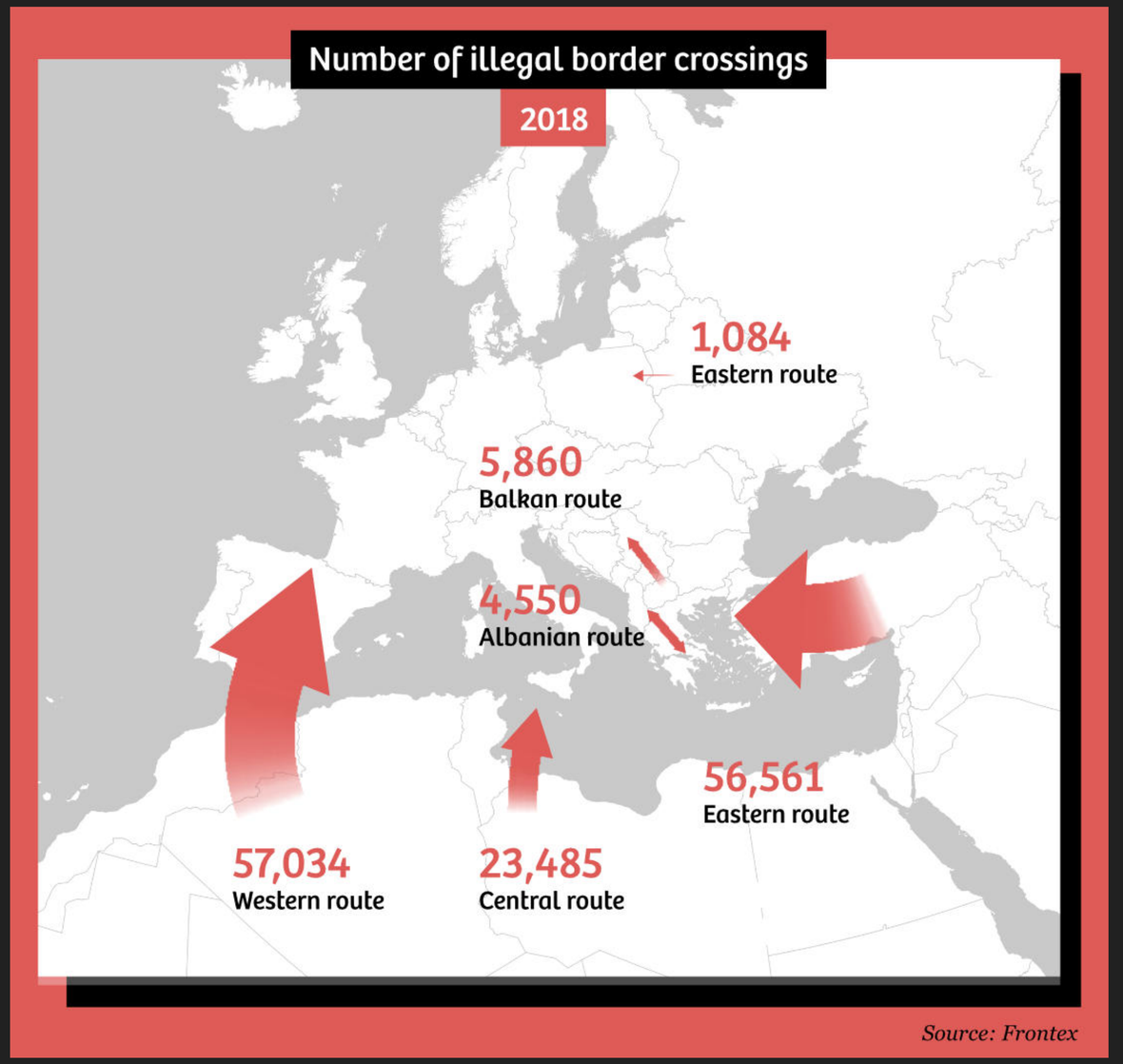
These graphics are specifically
designed to [...] **start conversations**
about our warming world and
the risks of climate change.



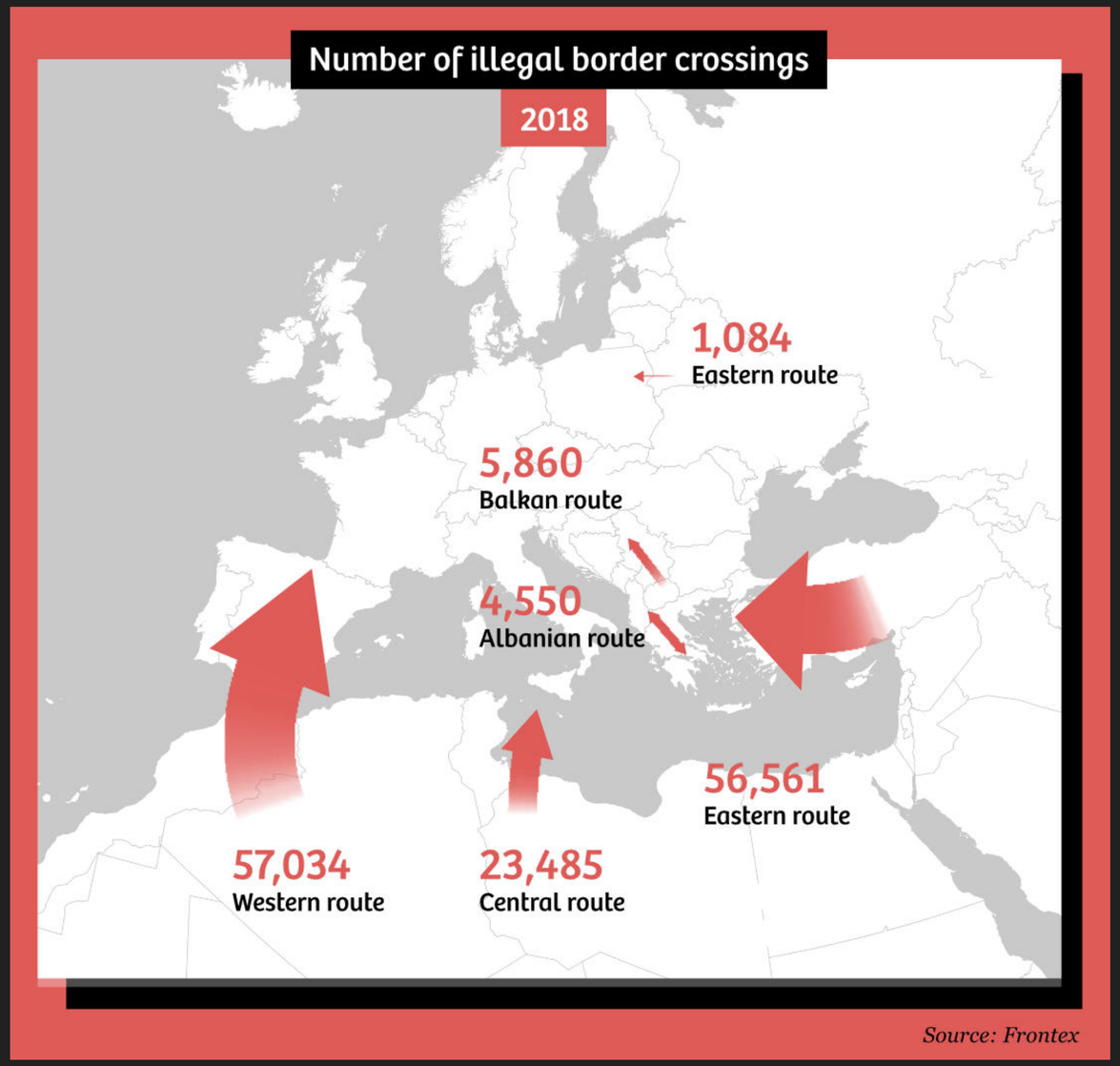
Scheme by Andy Kirk

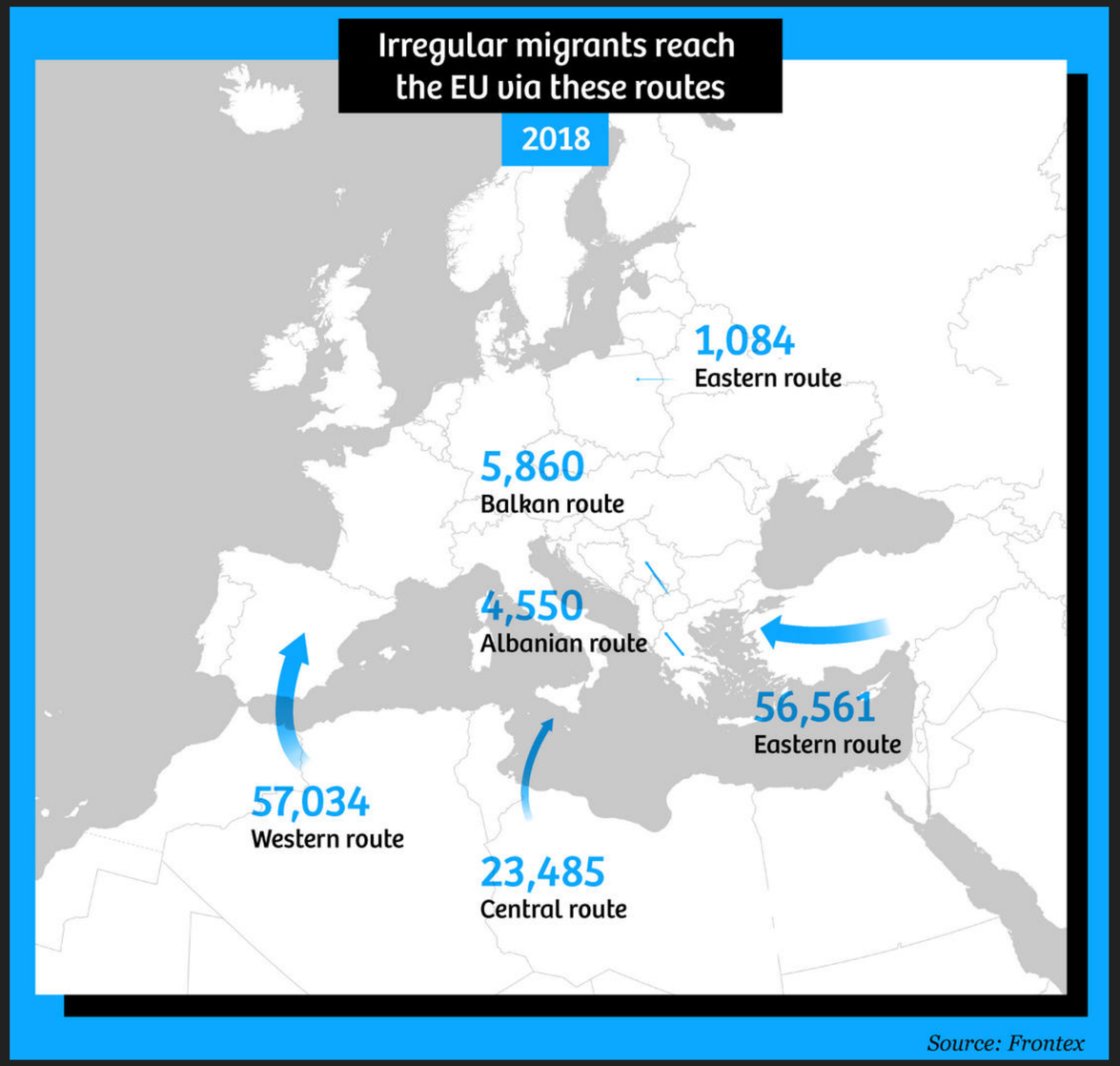
GOAL

Select charts that successfully tell your story

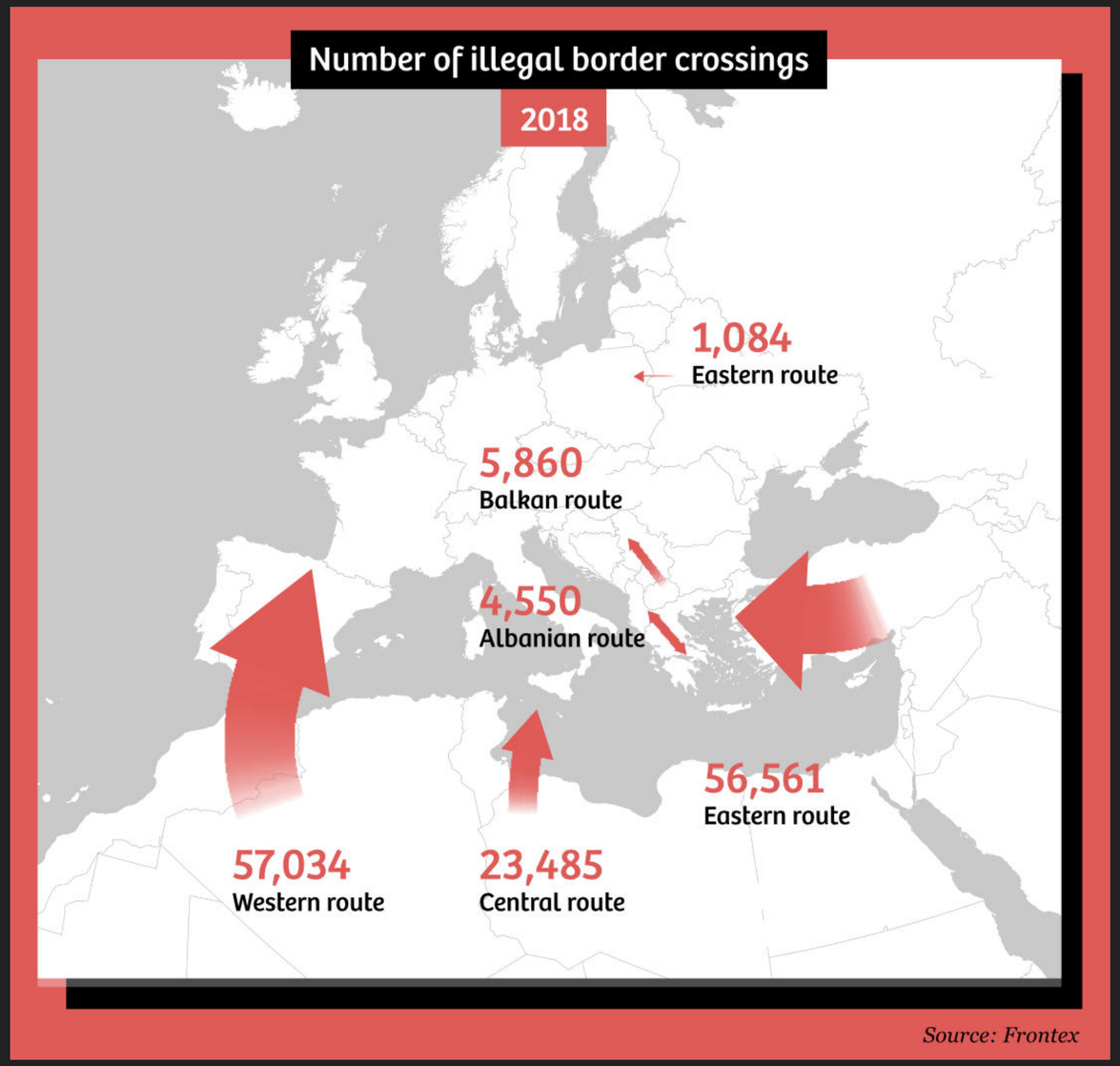


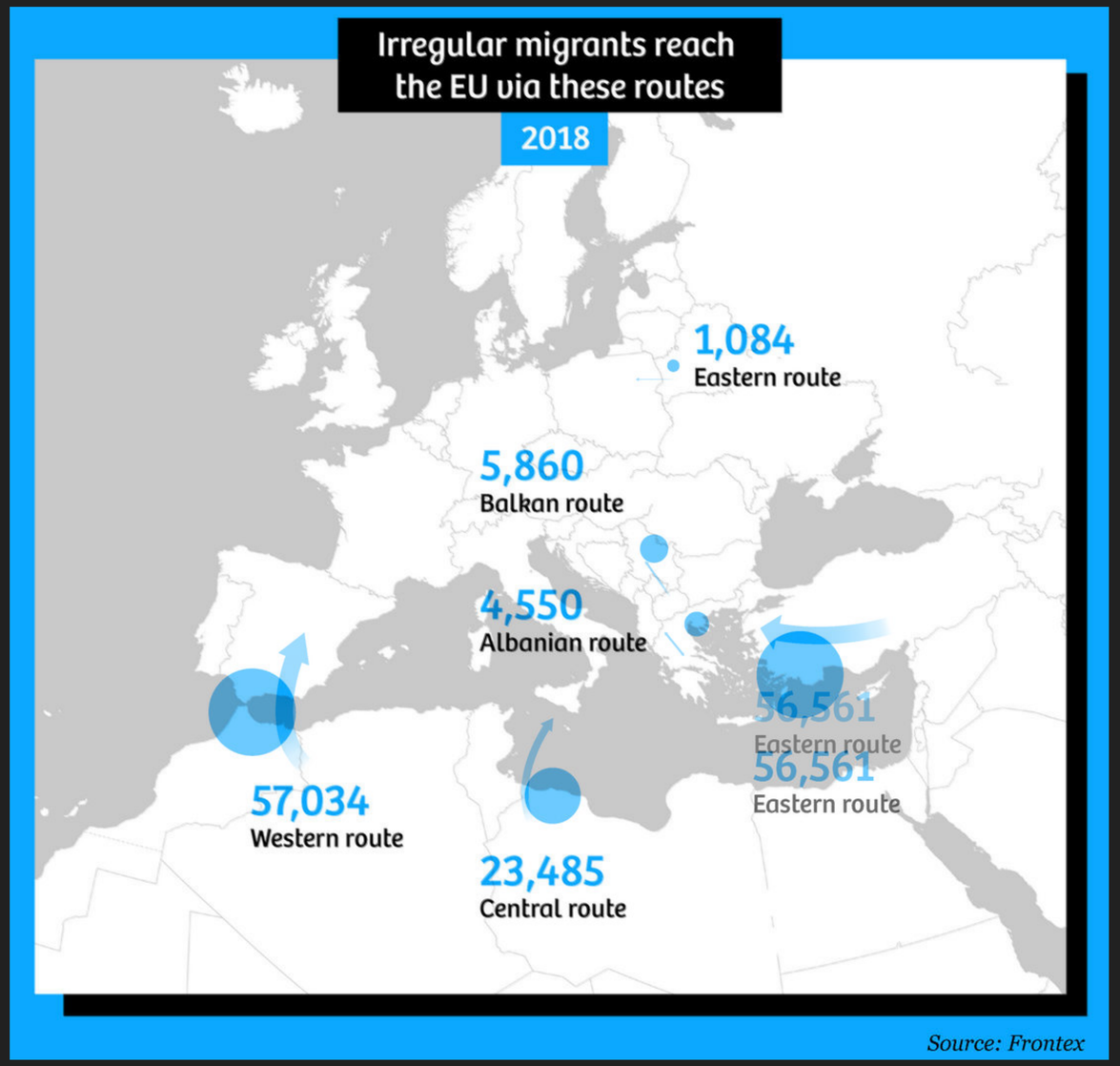
"How maps in the media make us more negative about migrants" by Maite Vermeulen, Leon de Korte & Henk van Houtum



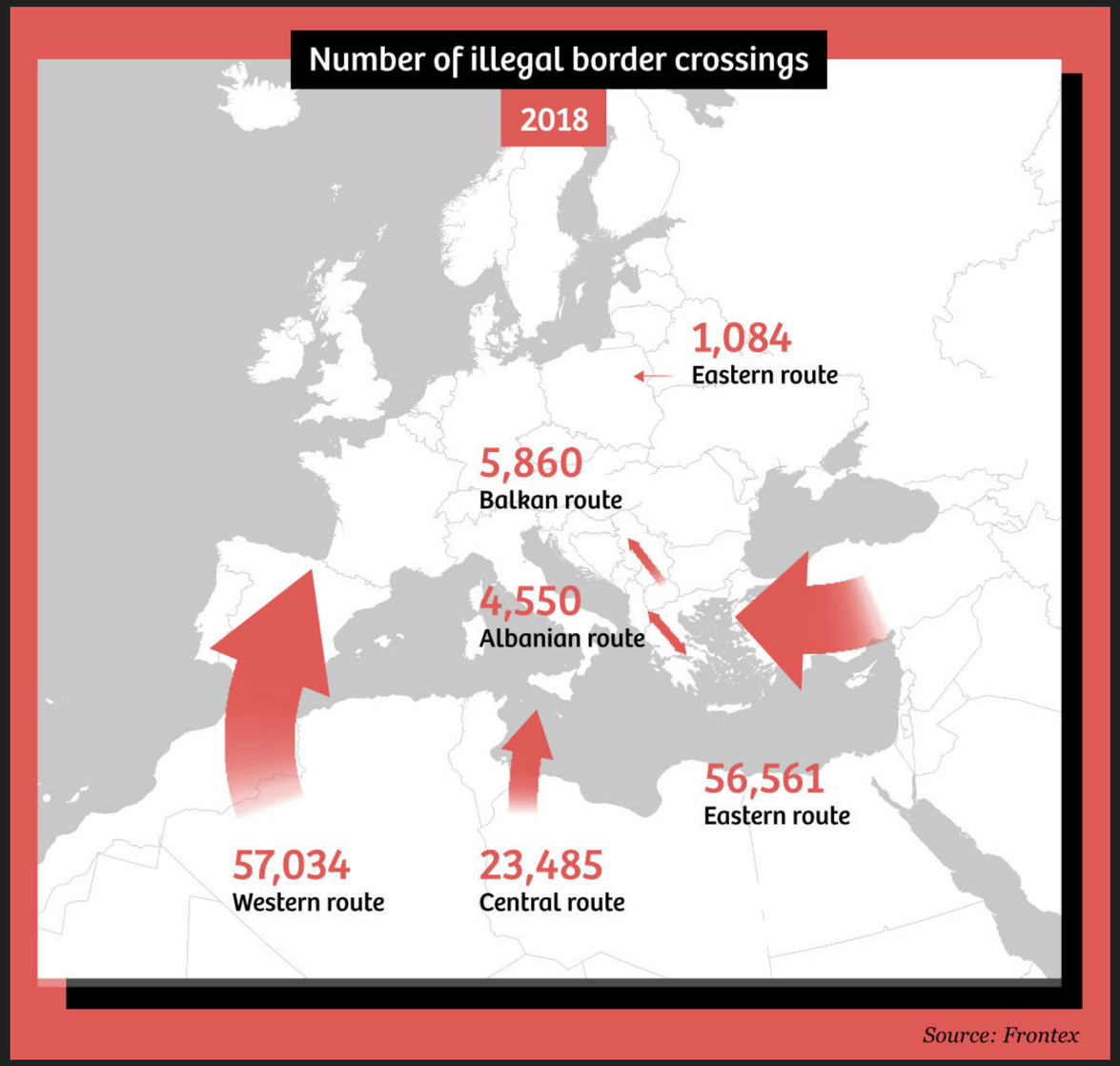


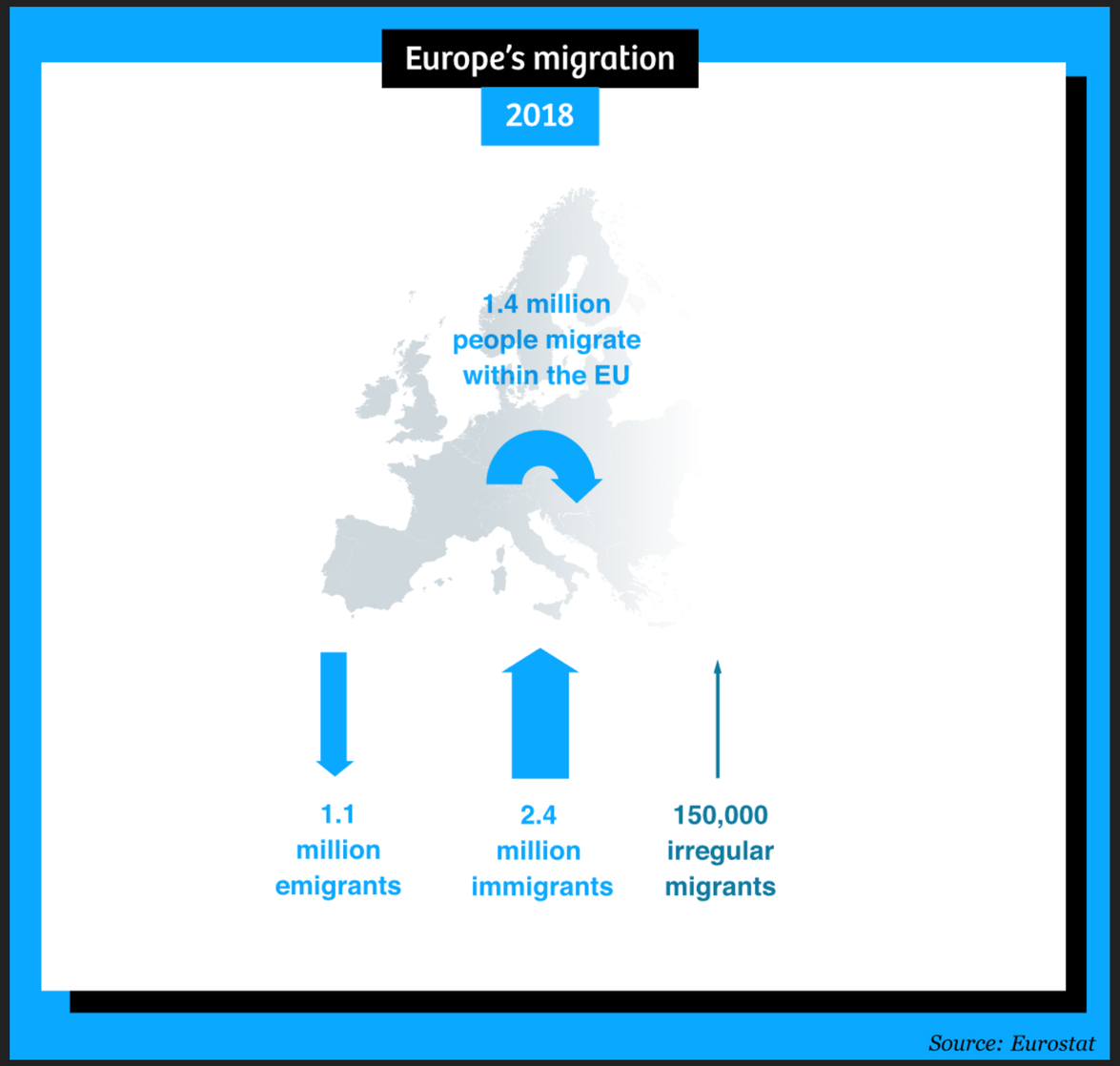
"How maps in the media make us more negative about migrants" by Maite Vermeulen, Leon de Korte & Henk van Houtum



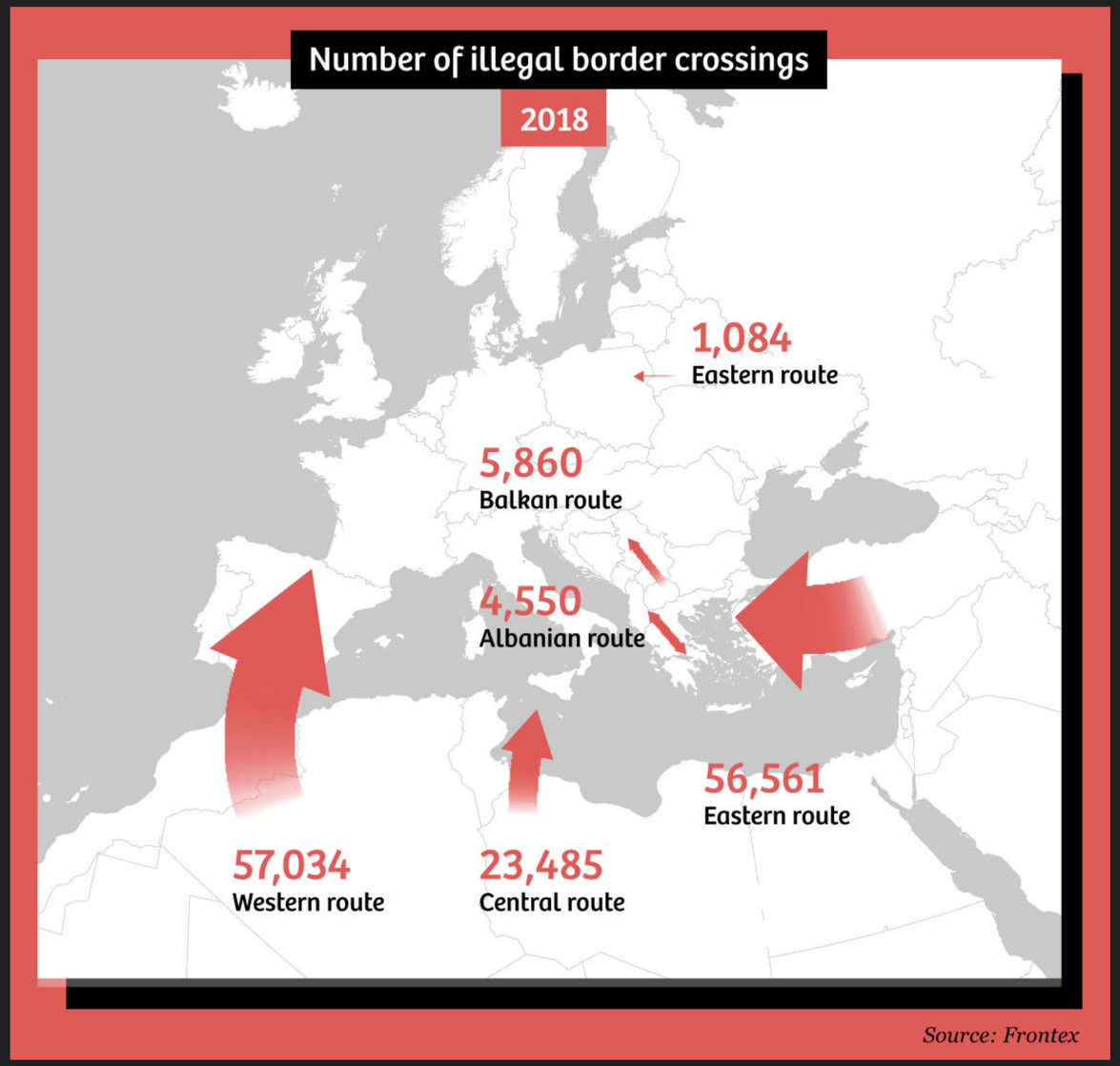


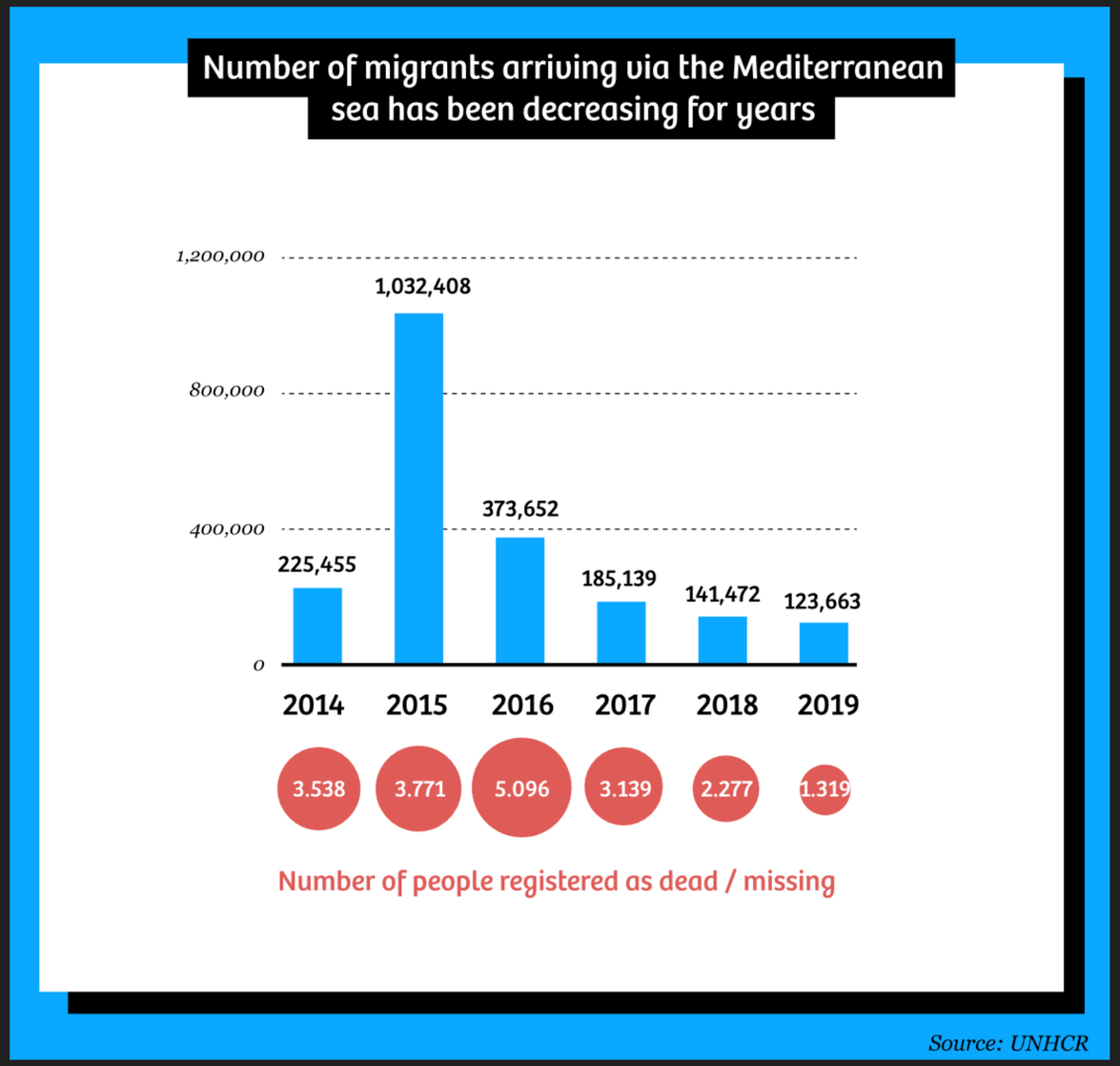
"How maps in the media make us more negative about migrants" by Maite Vermeulen, Leon de Korte & Henk van Houtum



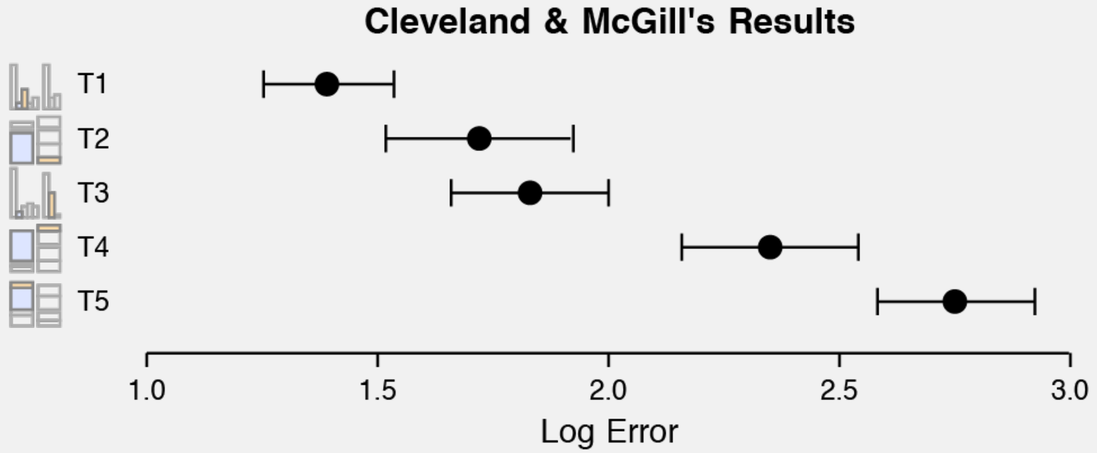


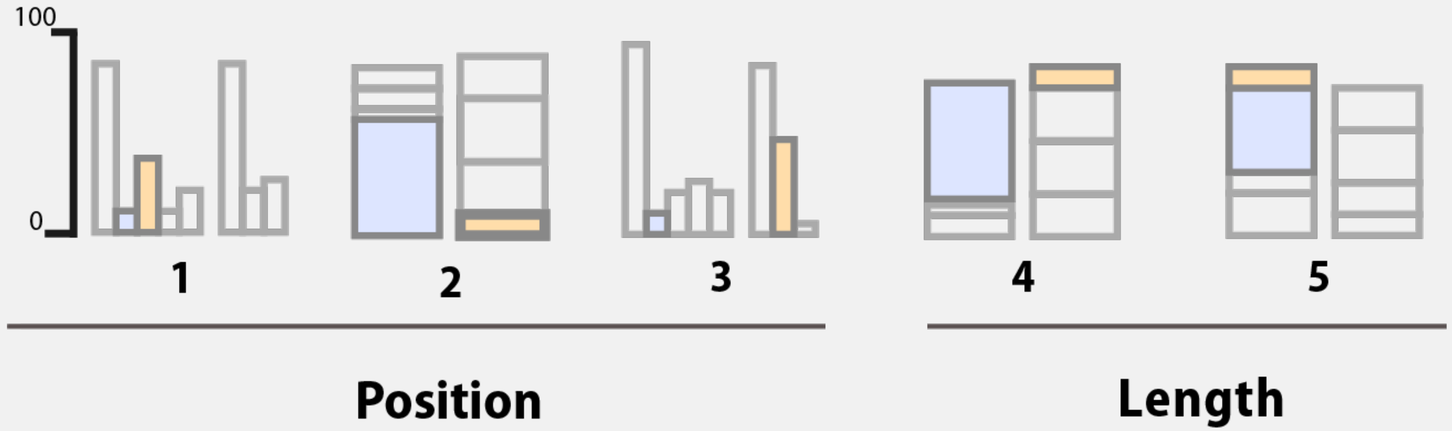
"How maps in the media make us more negative about migrants" by Maite Vermeulen, Leon de Korte & Henk van Houtum



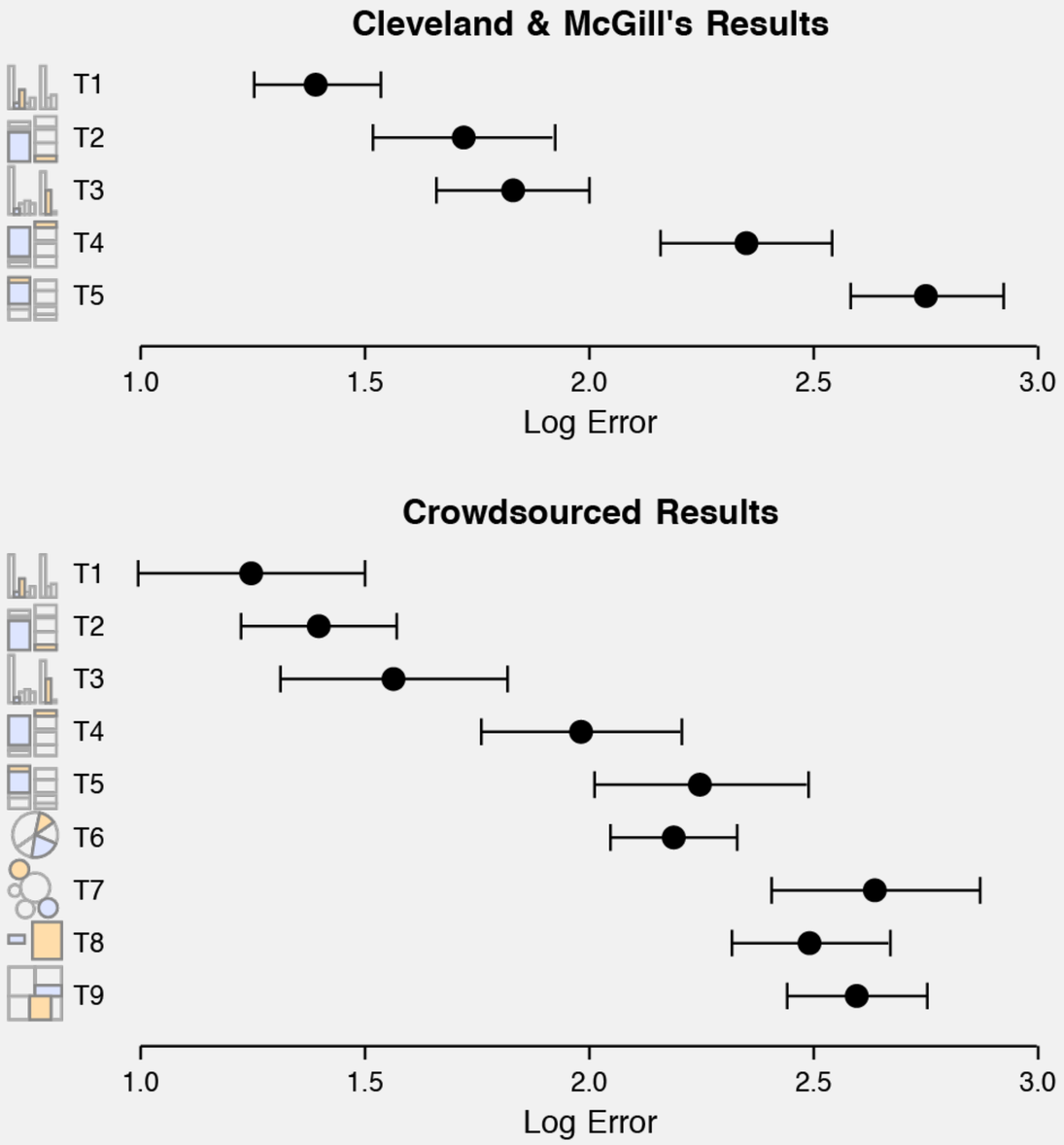


"How maps in the media make us more negative about migrants" by Maite Vermeulen, Leon de Korte & Henk van Houtum



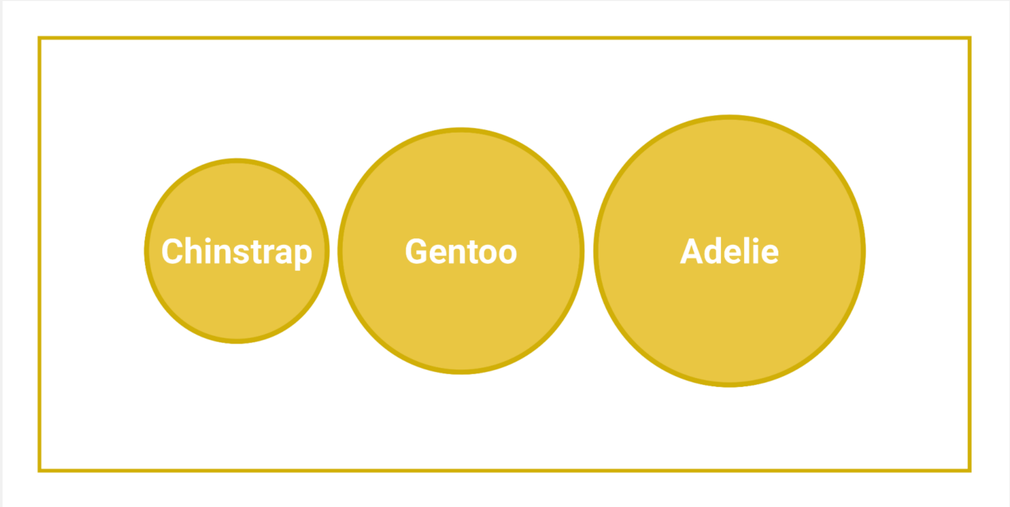


Source: Kieran Healy's "Data Visualization: A Practical Introduction"; results based on Heer and Bostock, following Cleveland and McGill

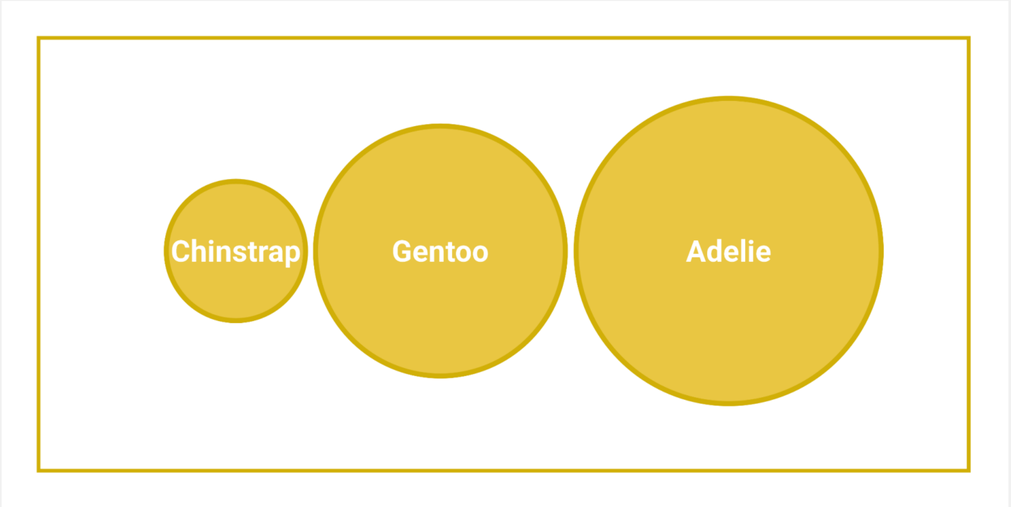


Source: Kieran Healy's "Data Visualization: A Practical Introduction"; results based on Heer and Bostock, following Cleveland and McGill

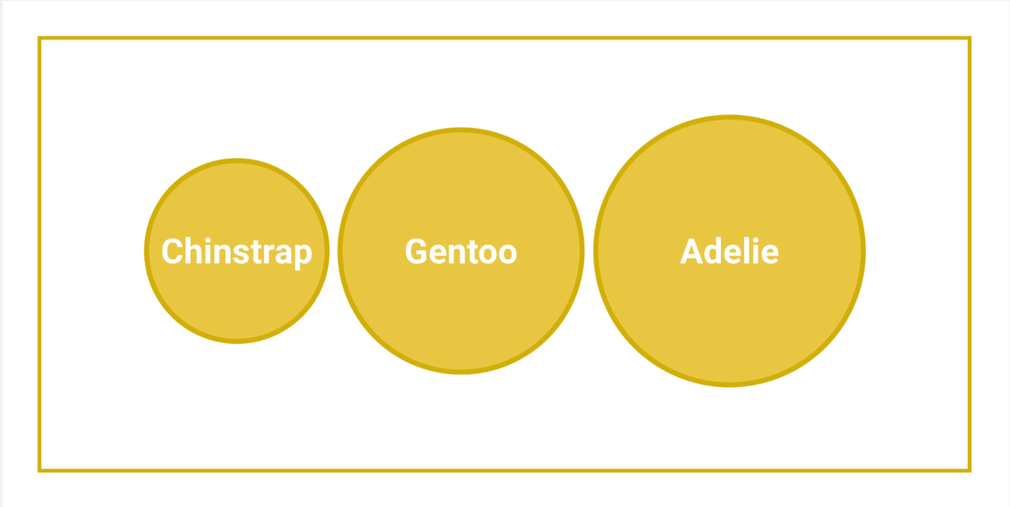
Use area.



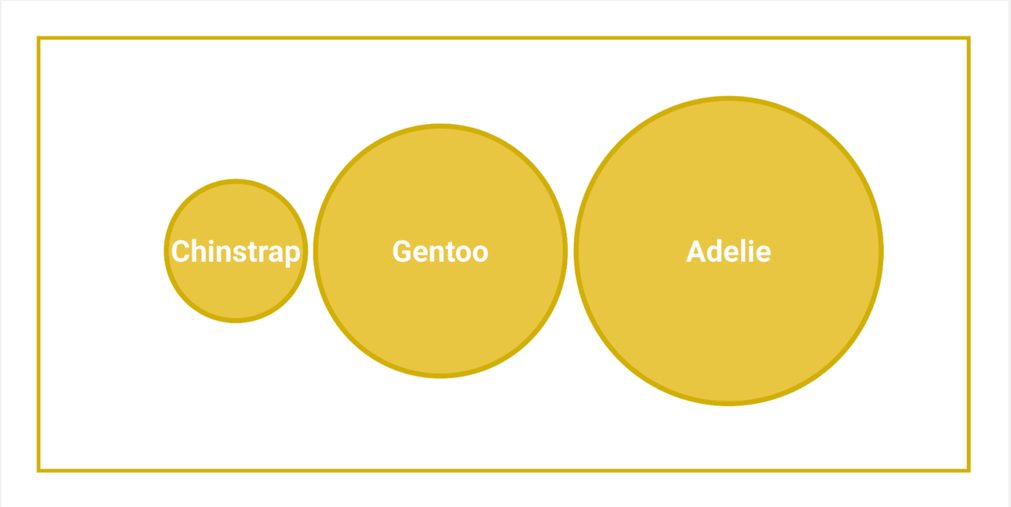
Not radius!

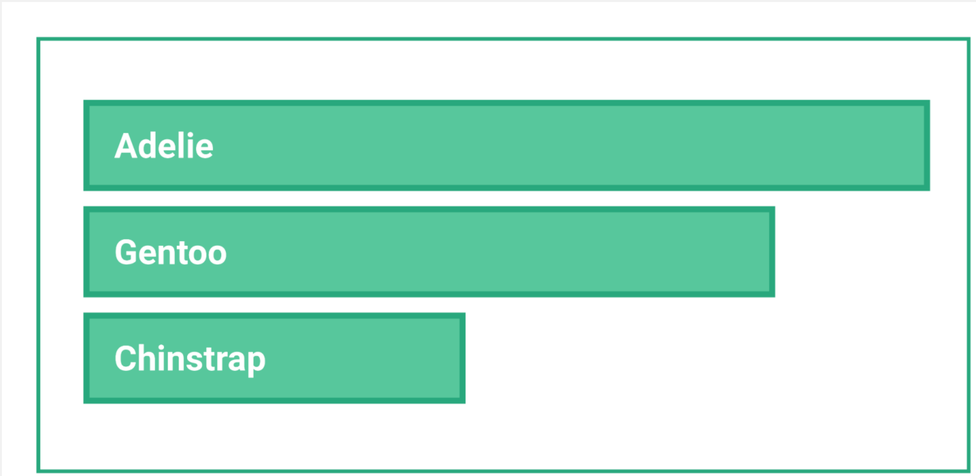


Use area.



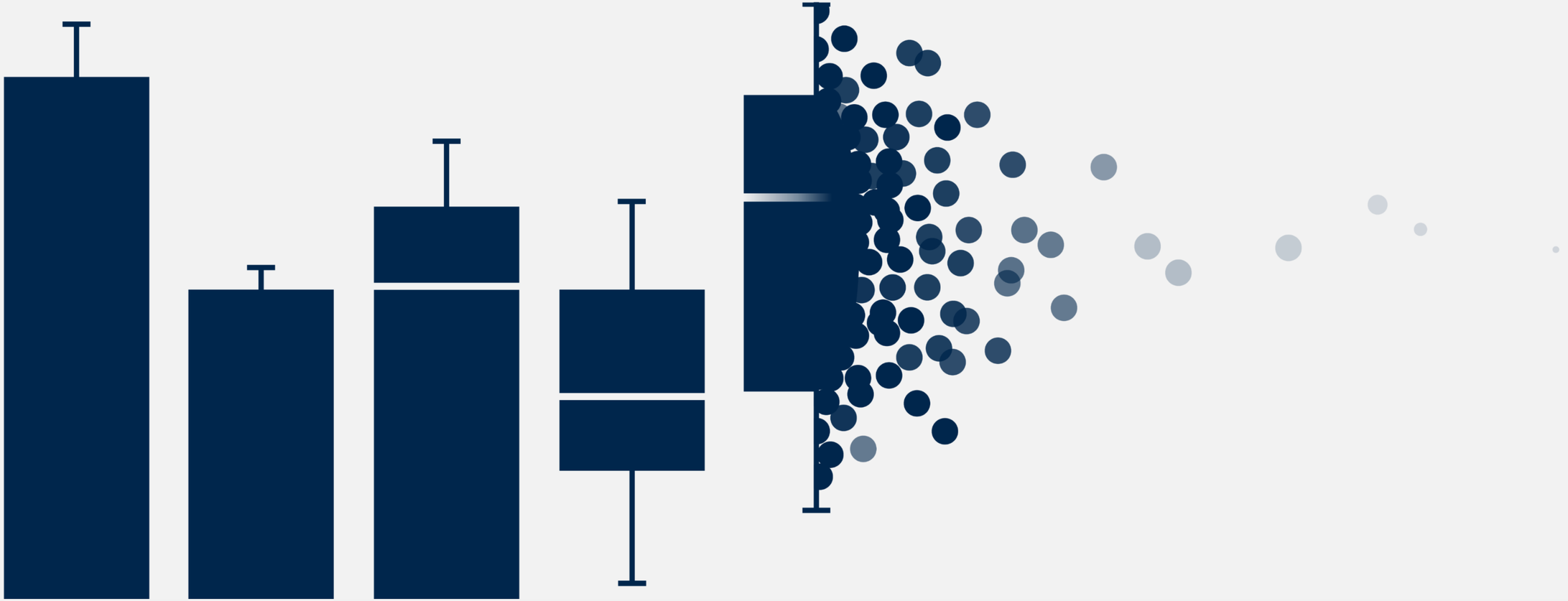
Not radius!



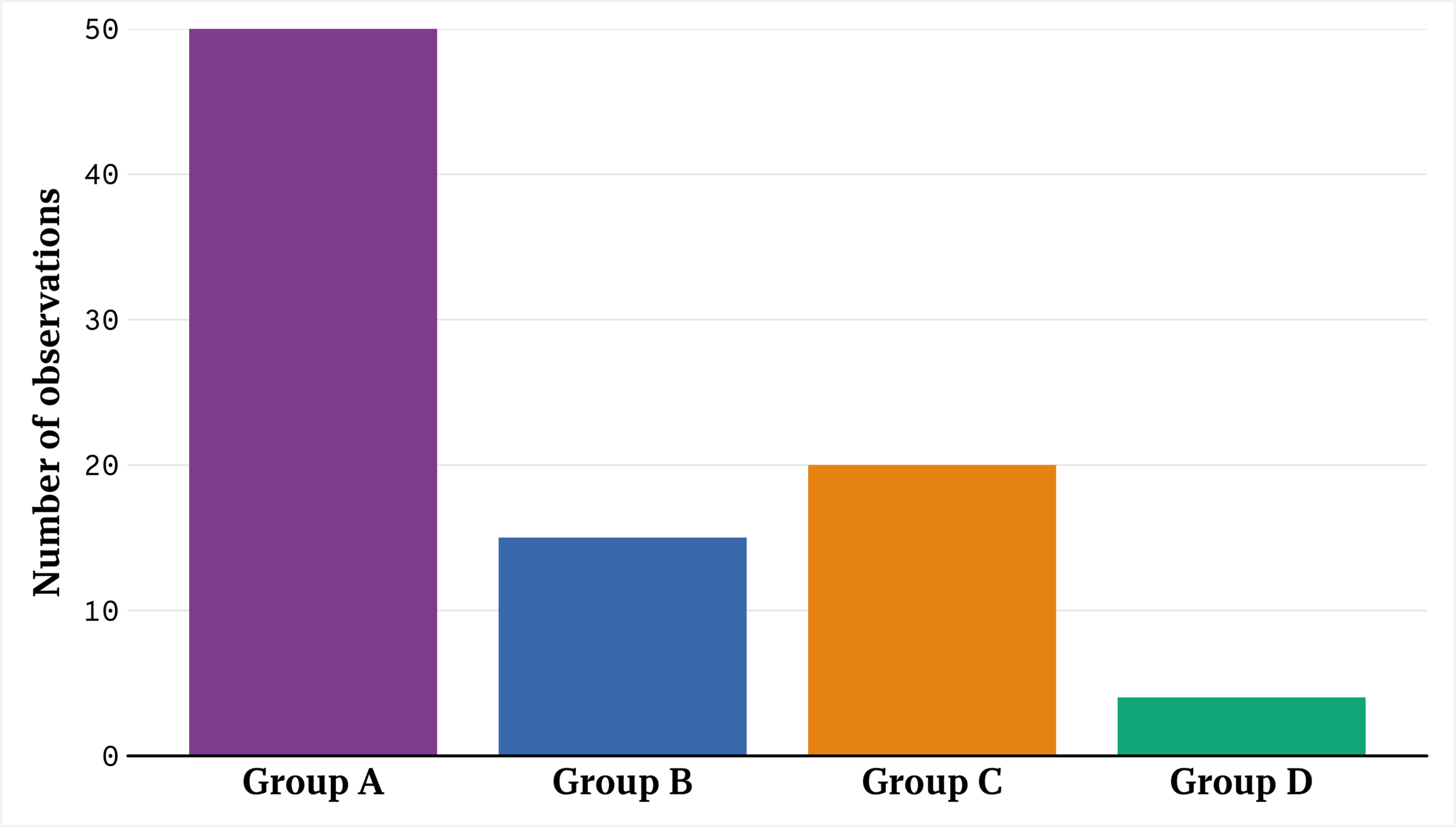


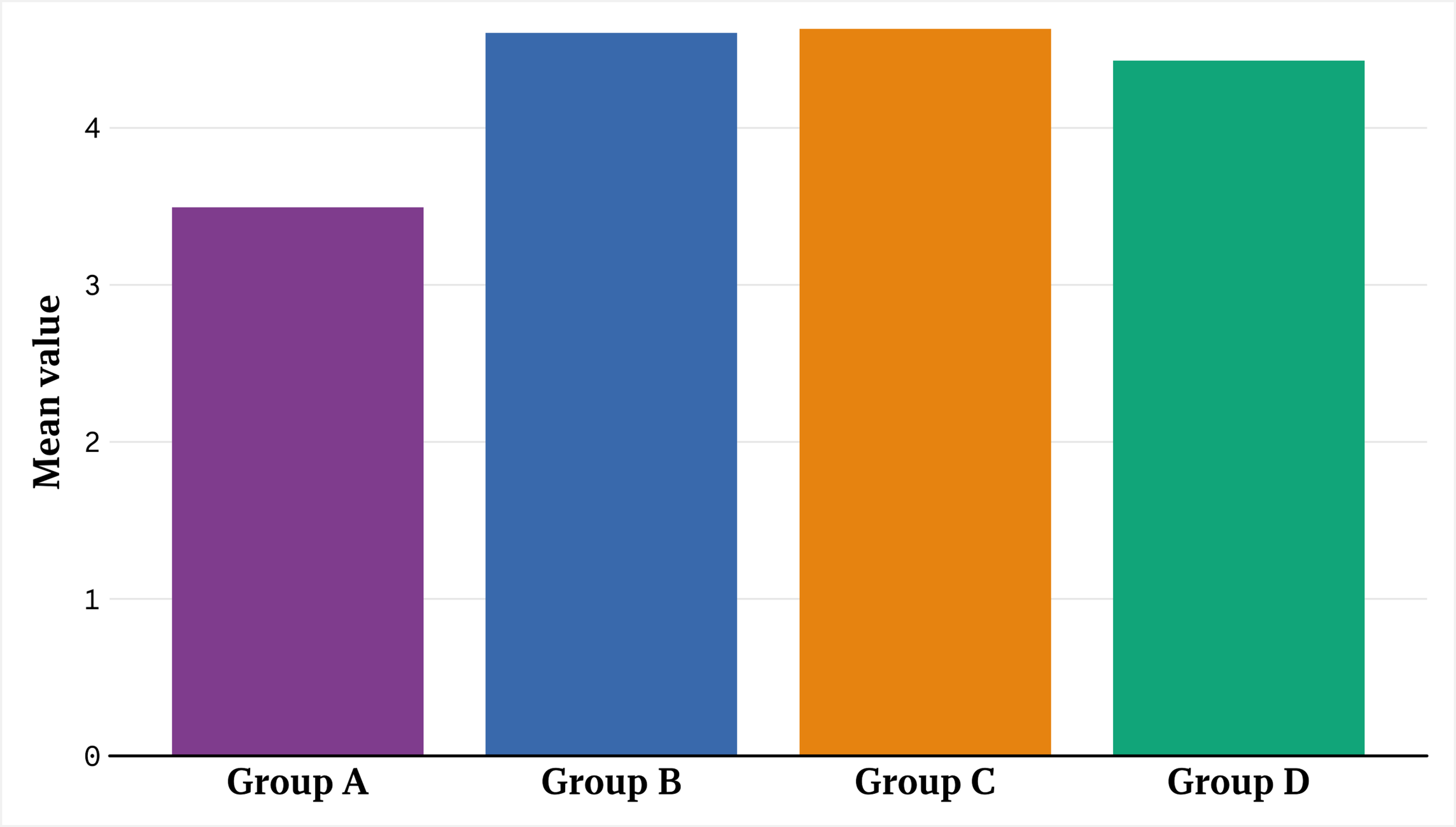
or: bars!

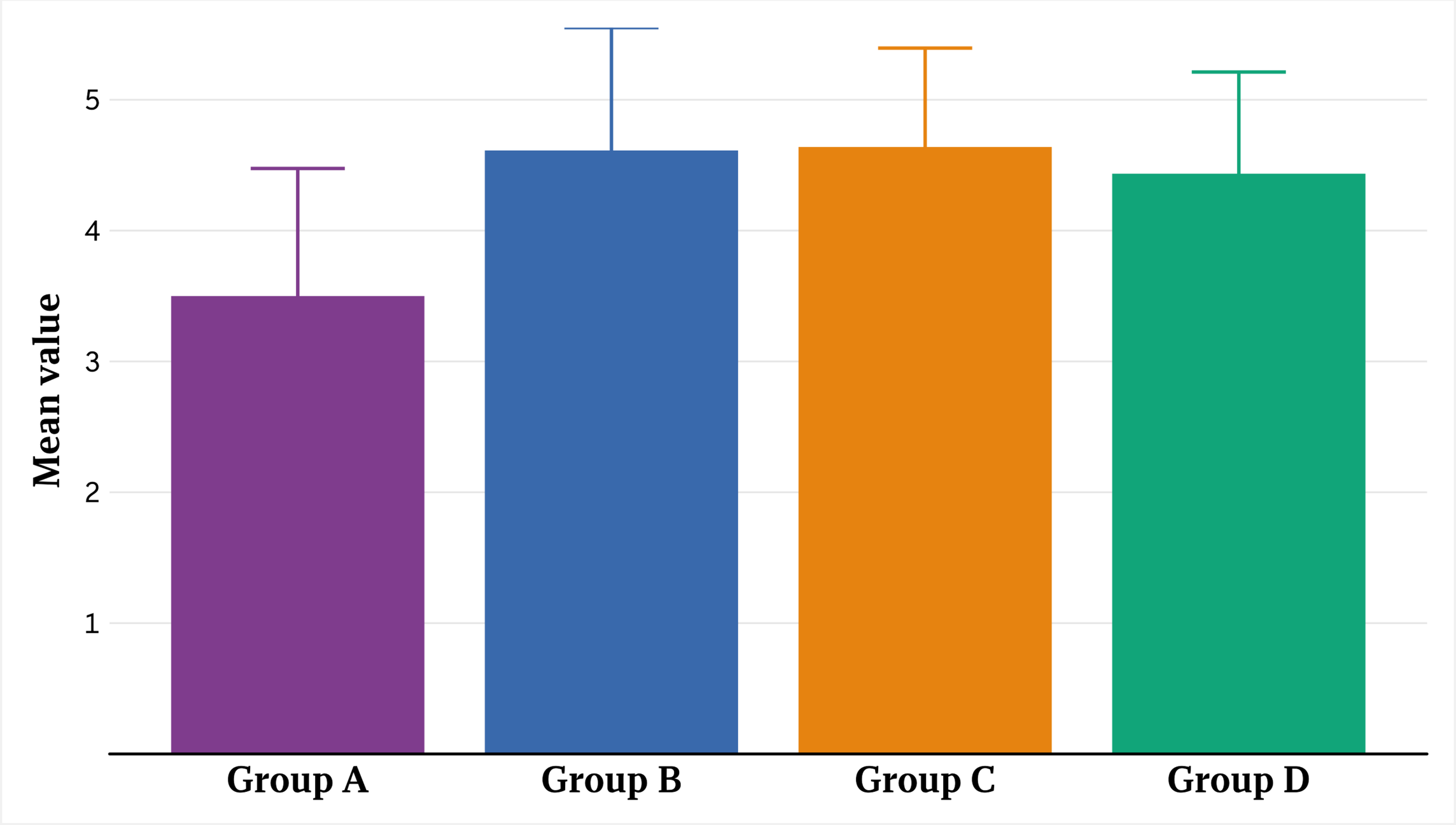
Beyond Bar and Box Plots

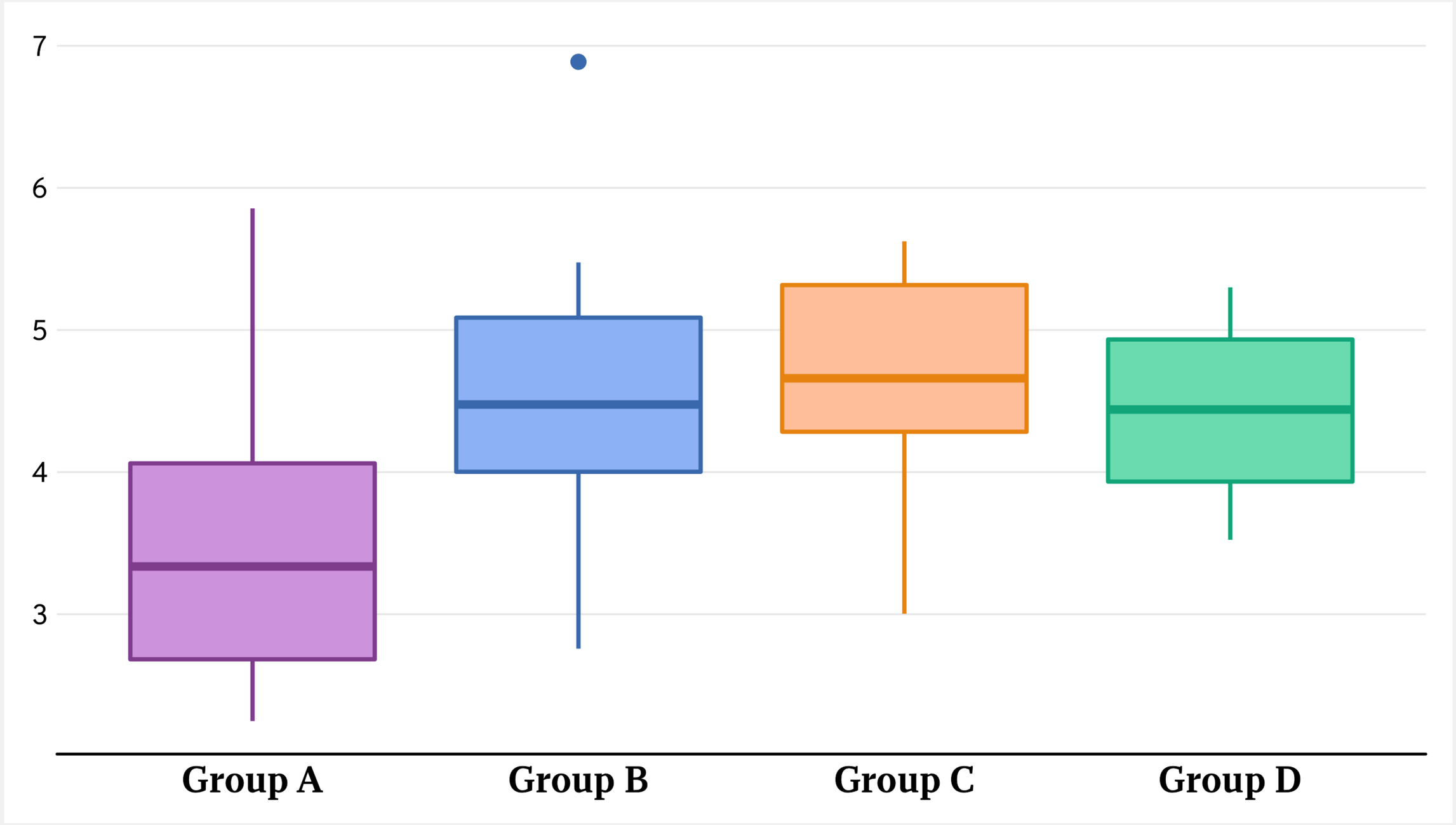


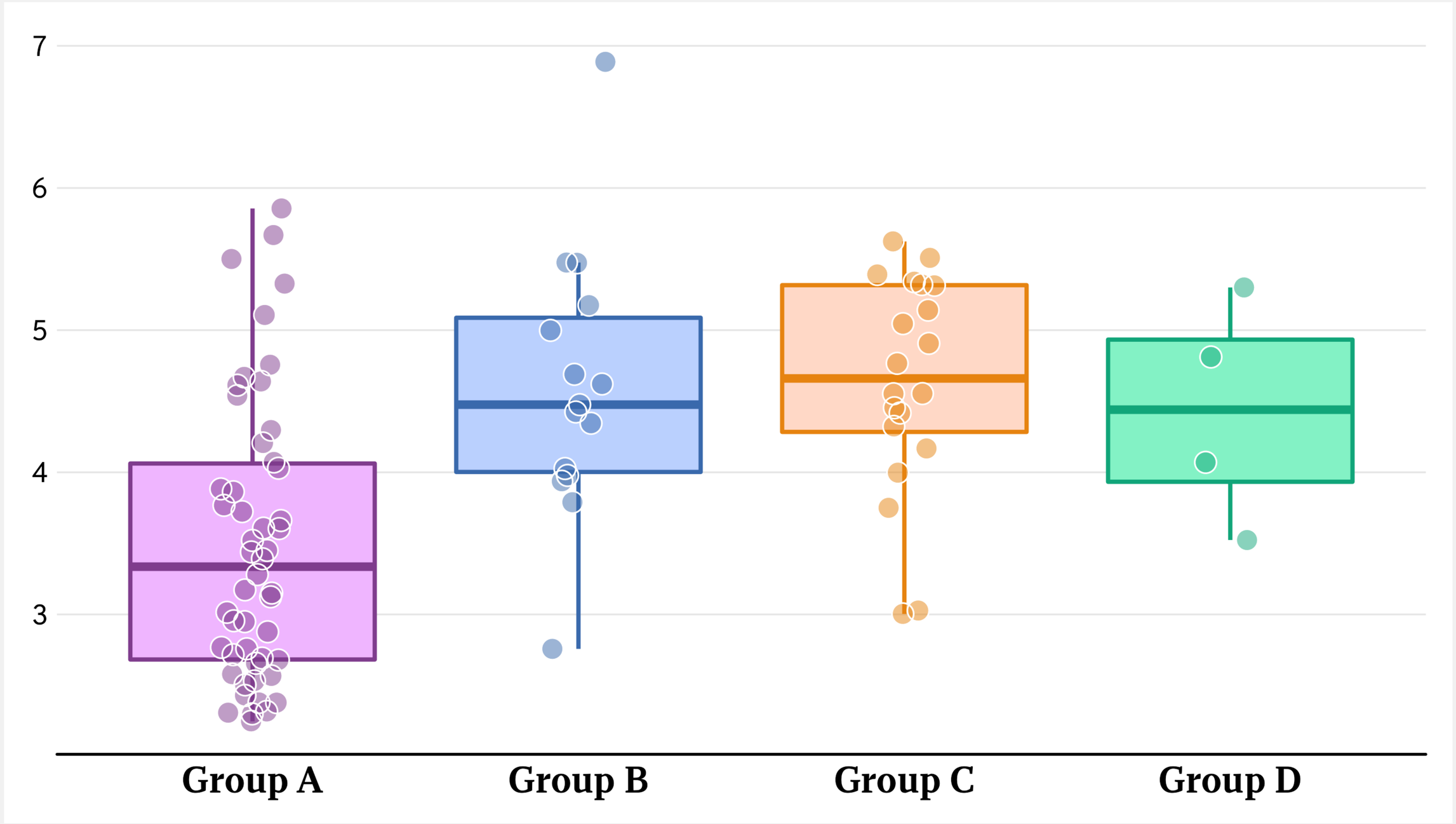
cedricscherer.com @CedScherer z3tt

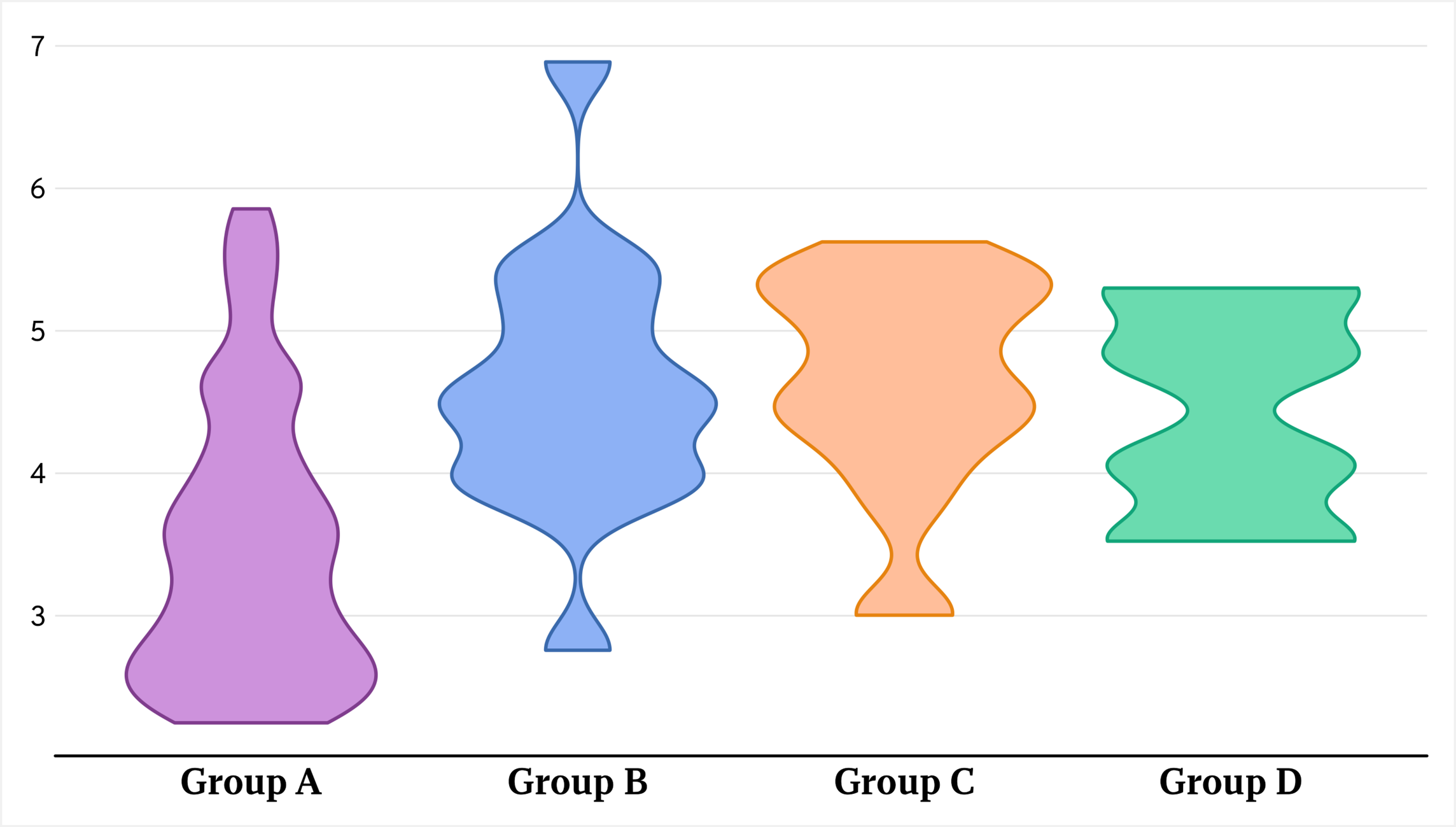

Number of observations
50
40
30
20
10
0
Group A
Group B
Group C
Group D

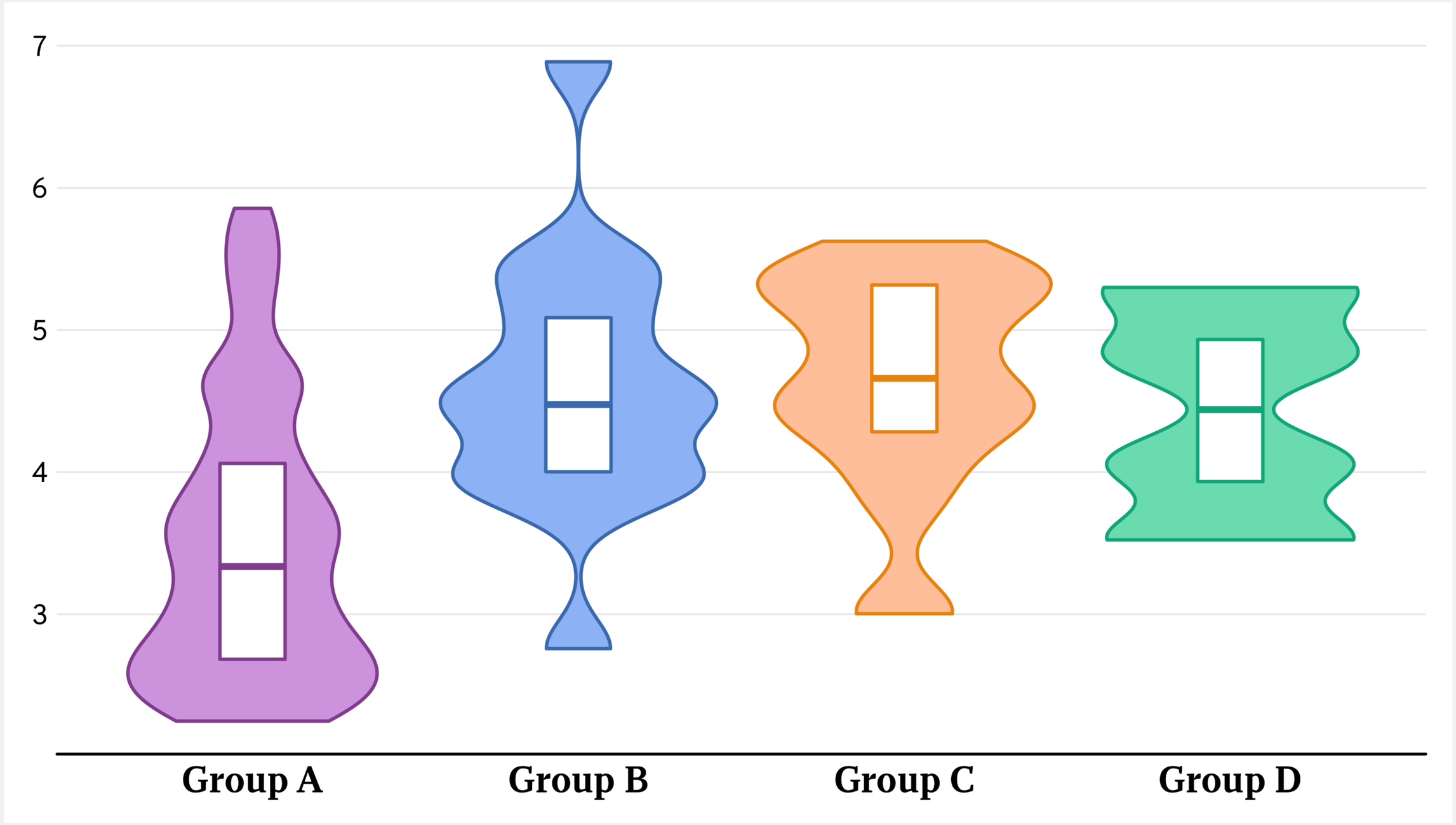

Mean value
0
1
2
3
4
Group A
Group B
Group C
Group D

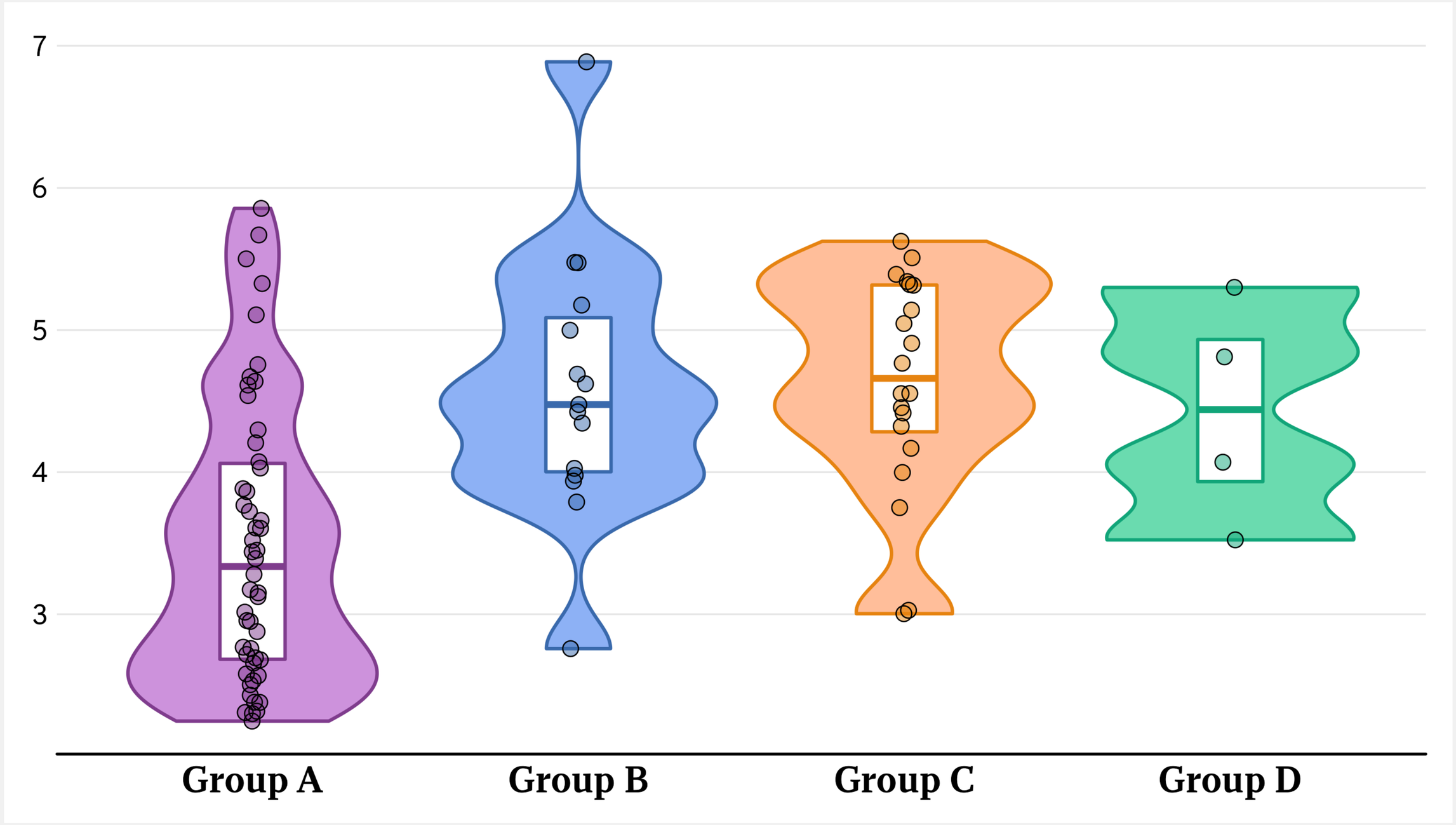

Mean value
5
4
3
2
1
Group A
Group B
Group C
Group D

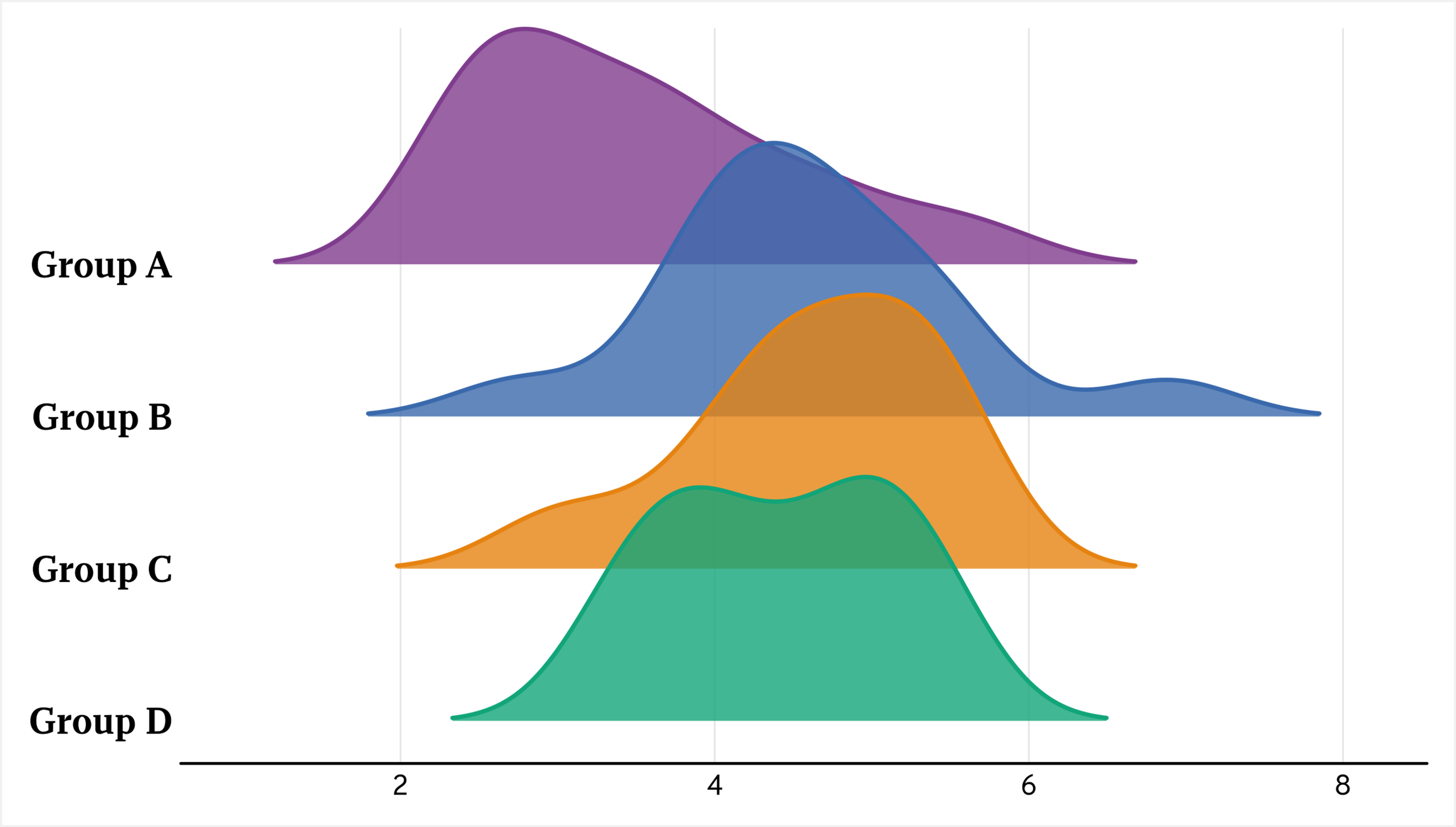

7
6
5
4
3
Group A
Group B
Group C
Group D

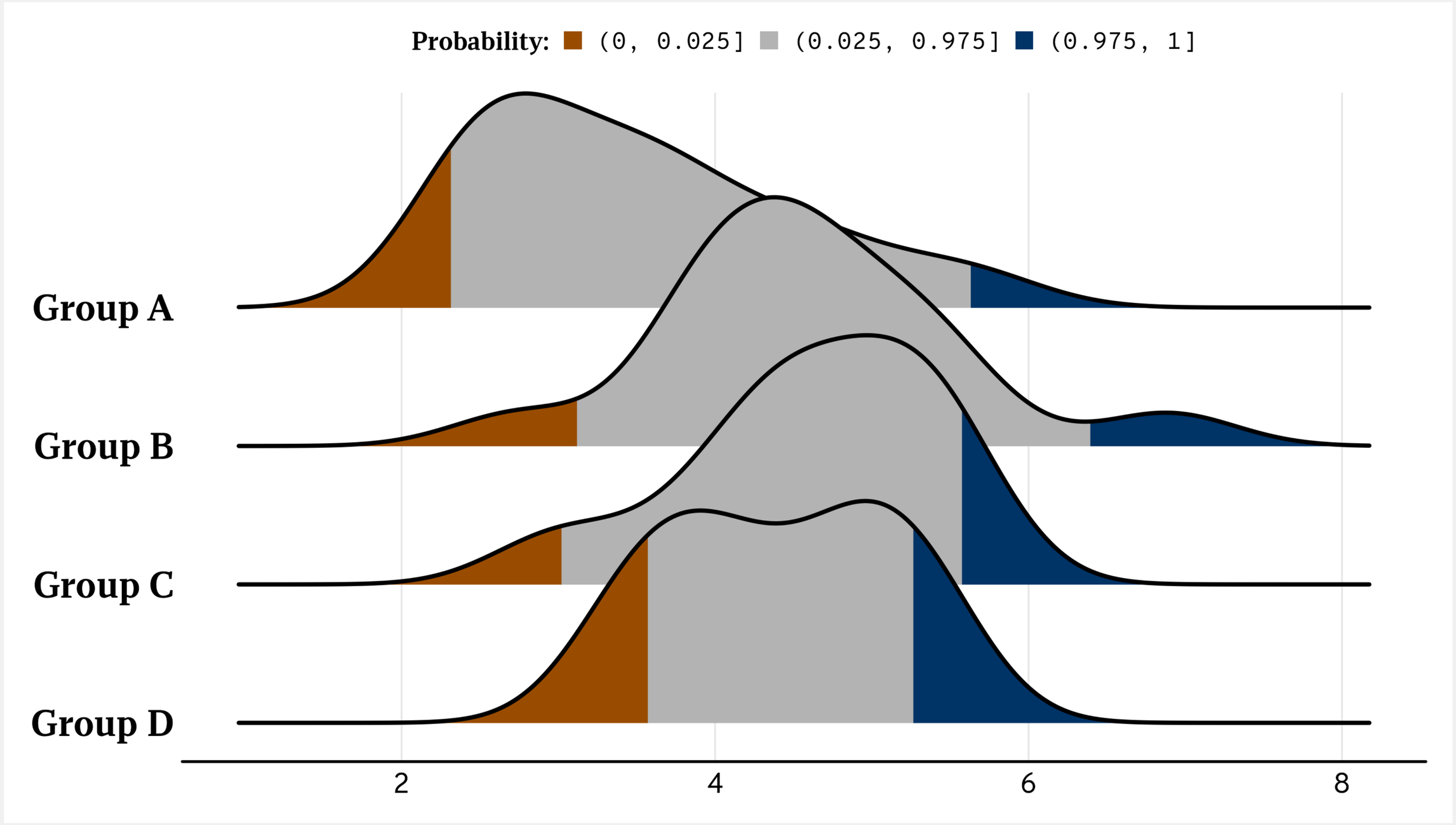

7
6
5
4
3
Group A
Group B
Group C
Group D


7
6
5
4
3
Group A
Group B
Group C
Group D


7
6
5
4
3
Group A
Group B
Group C
Group D

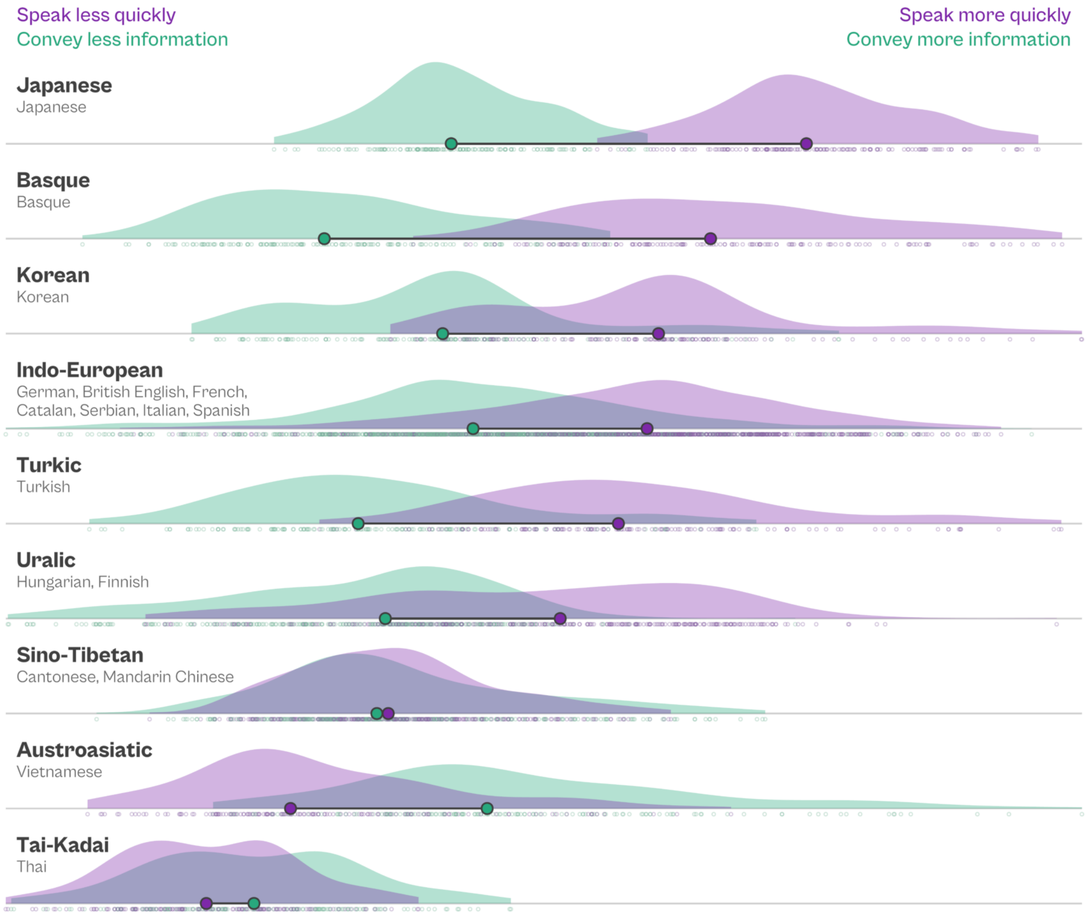

7
6
5
4
3
Group A
Group B
Group C
Group D


Group A
Group B
Group C
Group D
2
4
6
8


Probability: (0, 0.025] (0.025, 0.975] (0.975, 1]
Group A
Group B
Group C
Group D
2
4
6
8

Communicating fast doesn't necessarily mean communicating more

Variation in speech and information rates across language families, shown as normalized rates for comparison. While there are stark cross-linguistic differences in speech rates, information rates are more similar.

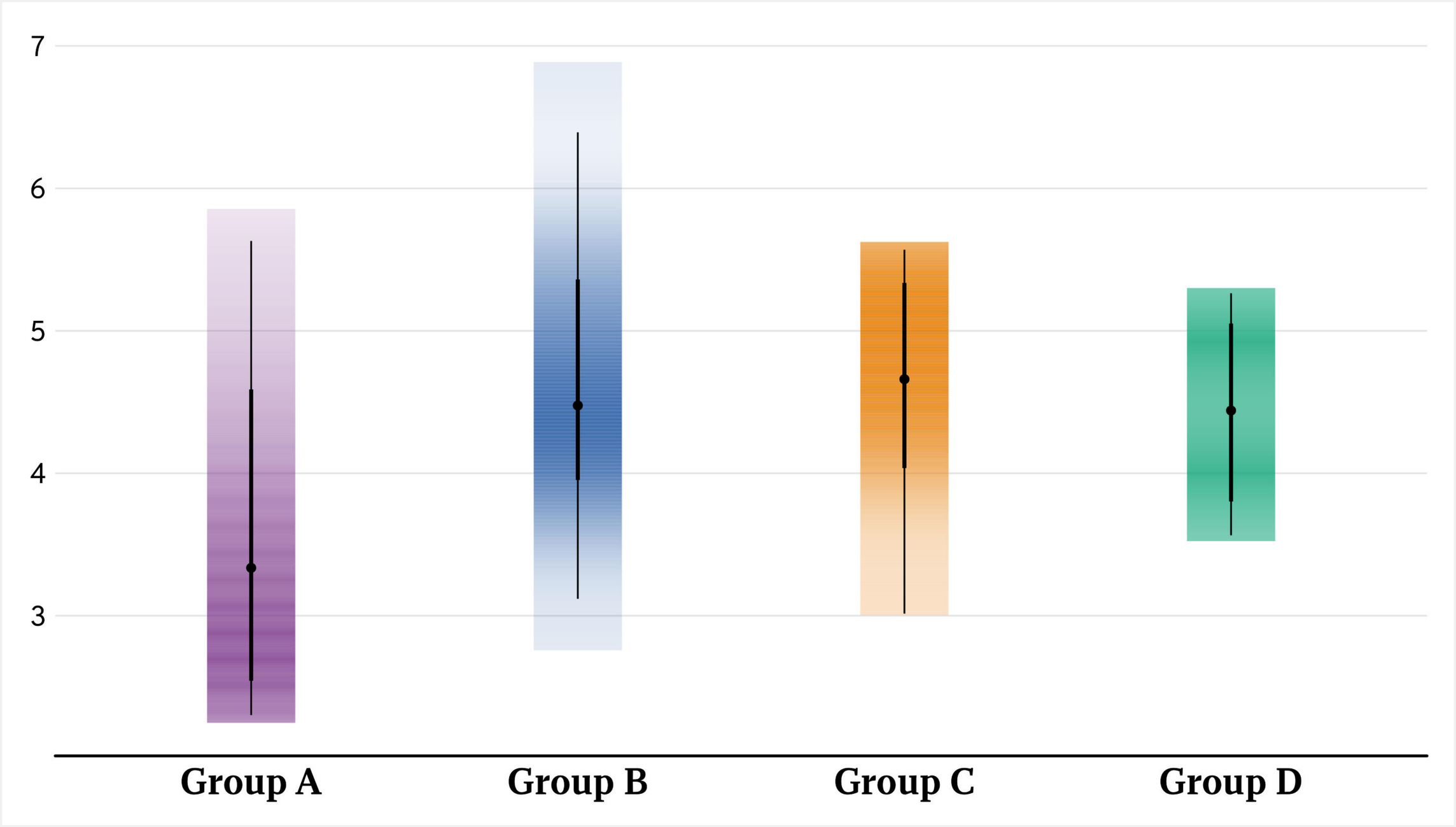


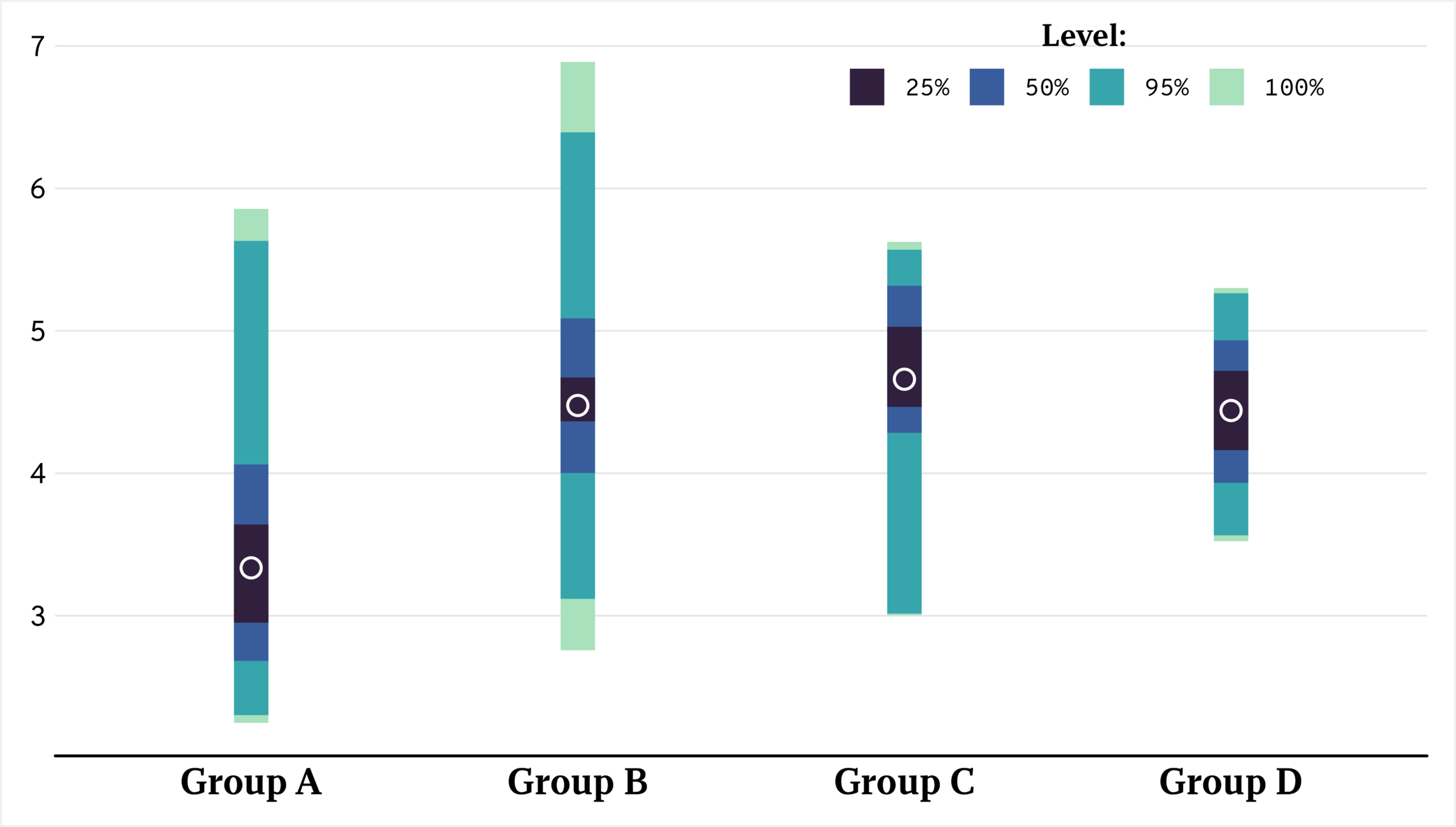
Normalized rates of **speech (syllables per second)** and **information (bits per second)**

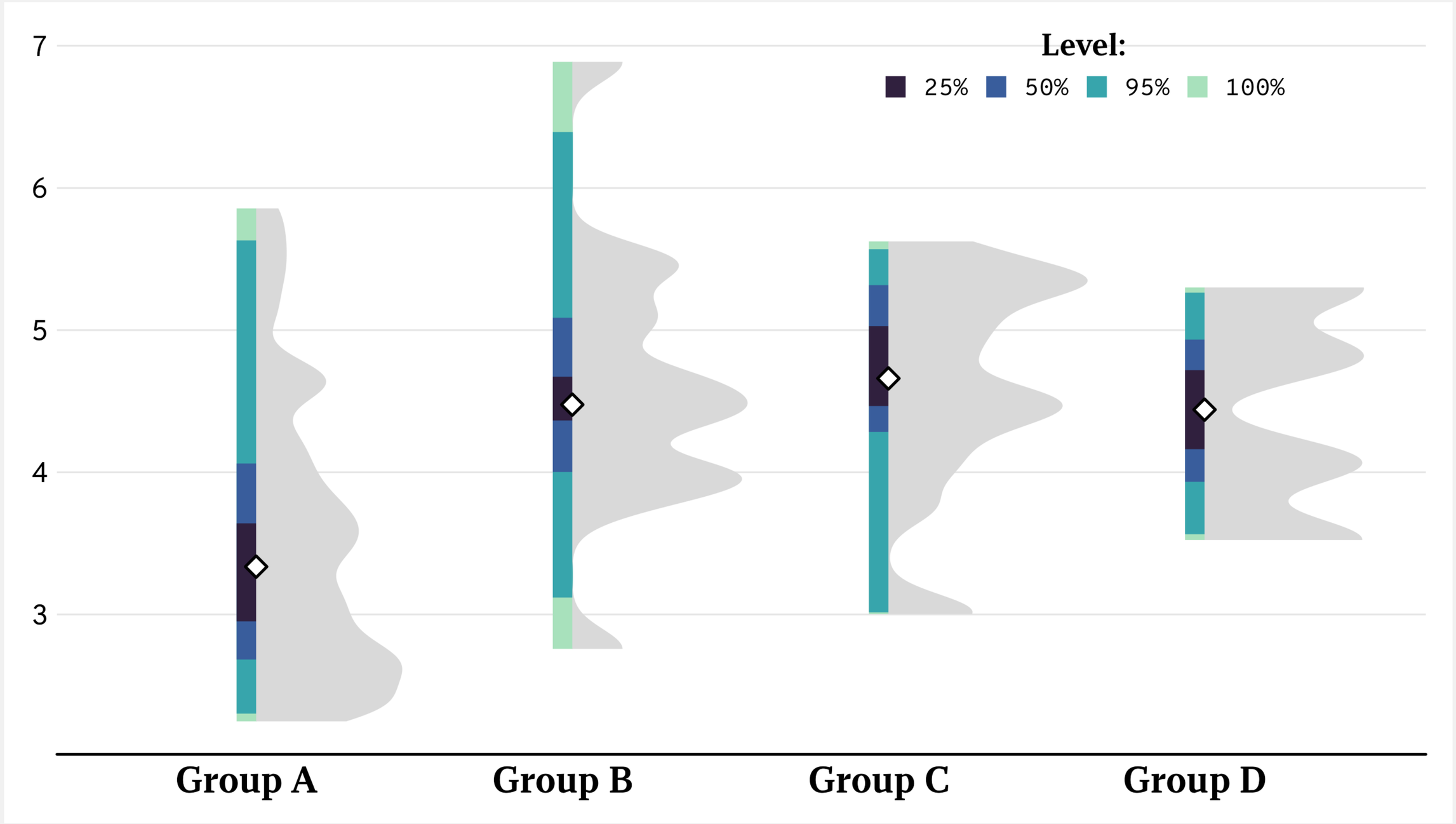
Source: Coupé *et al.* 2019 *Science Advances* **5**(9). DOI: 10.1126/sciadv.aaw2594
Graphic: Cédric Scherer • Large dots show the median rates for each language family. Small dots show single estimates.

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

cedricscherer.com @CedScherer z3tt


7
6
5
4
3
Group A
Group B
Group C
Group D

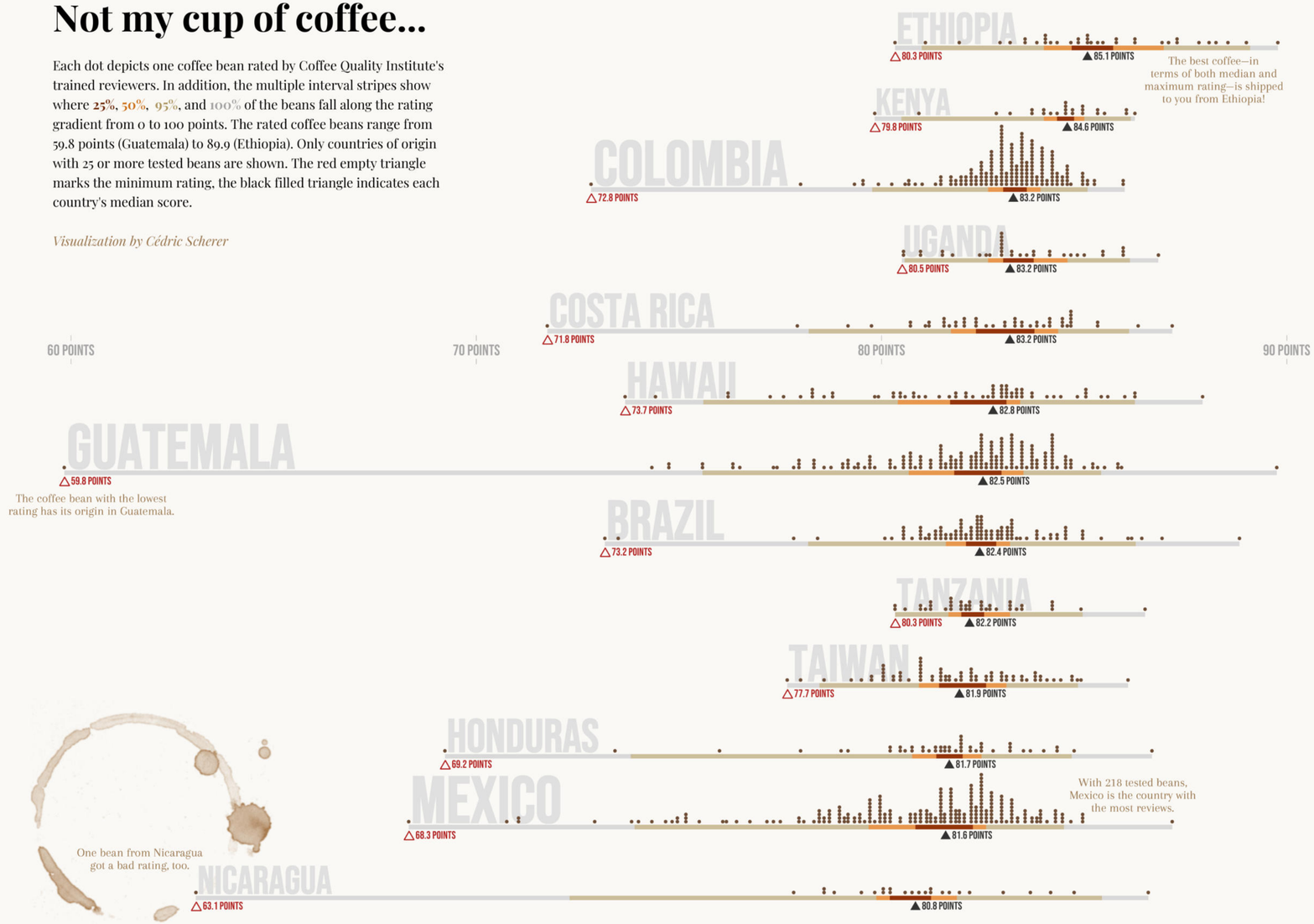

Level:
25%
50%
95%
100%
7
6
5
4
3
Group A
Group B
Group C
Group D


Level:
25%
50%
95%
100%
7
6
5
4
3
Group A
Group B
Group C
Group D

Not my cup of coffee...

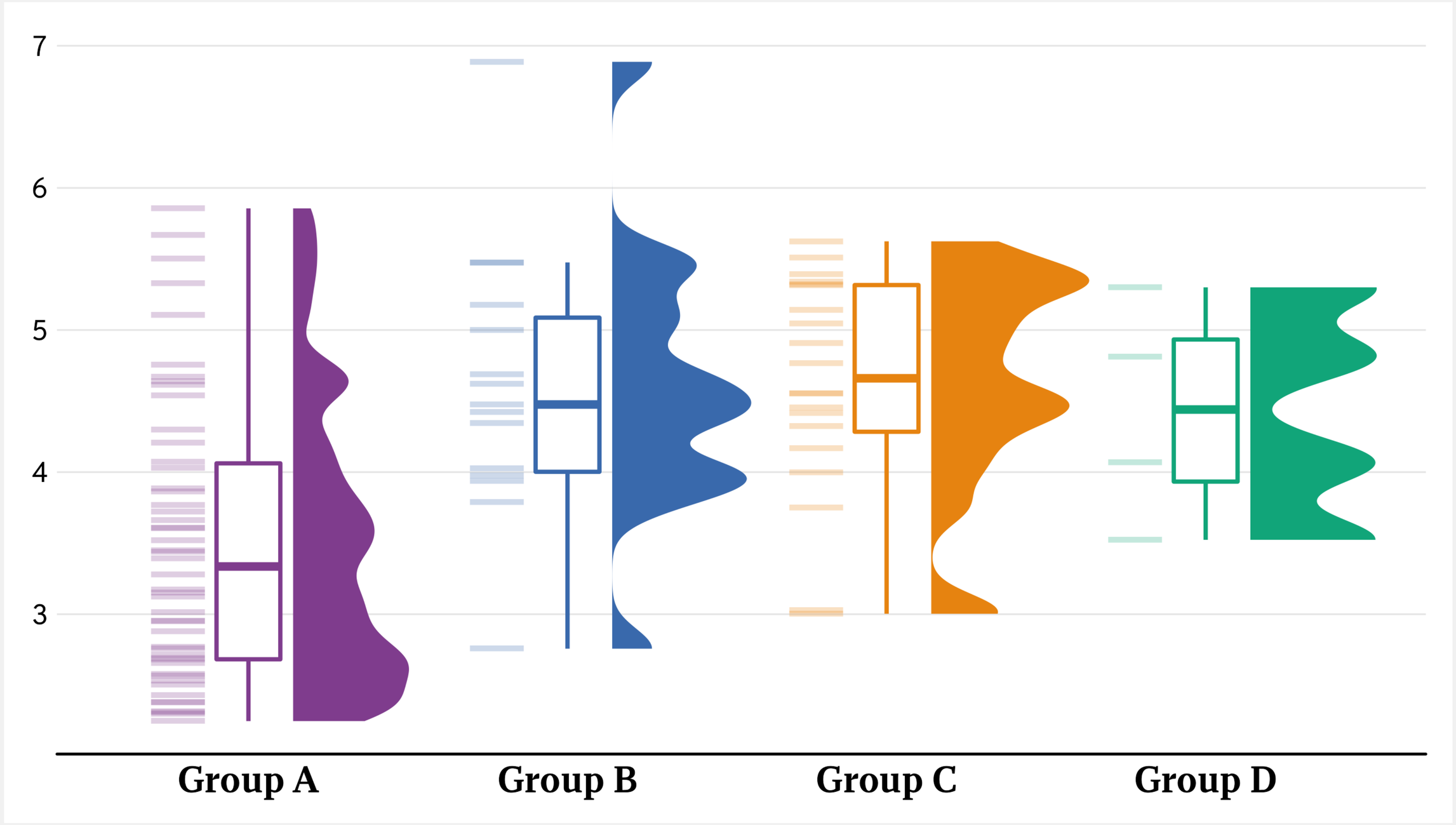
Each dot depicts one coffee bean rated by Coffee Quality Institute's trained reviewers. In addition, the multiple interval stripes show where 25%, 50%, 95%, and 100% of the beans fall along the rating gradient from 0 to 100 points. The rated coffee beans range from 59.8 points (Guatemala) to 89.9 (Ethiopia).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

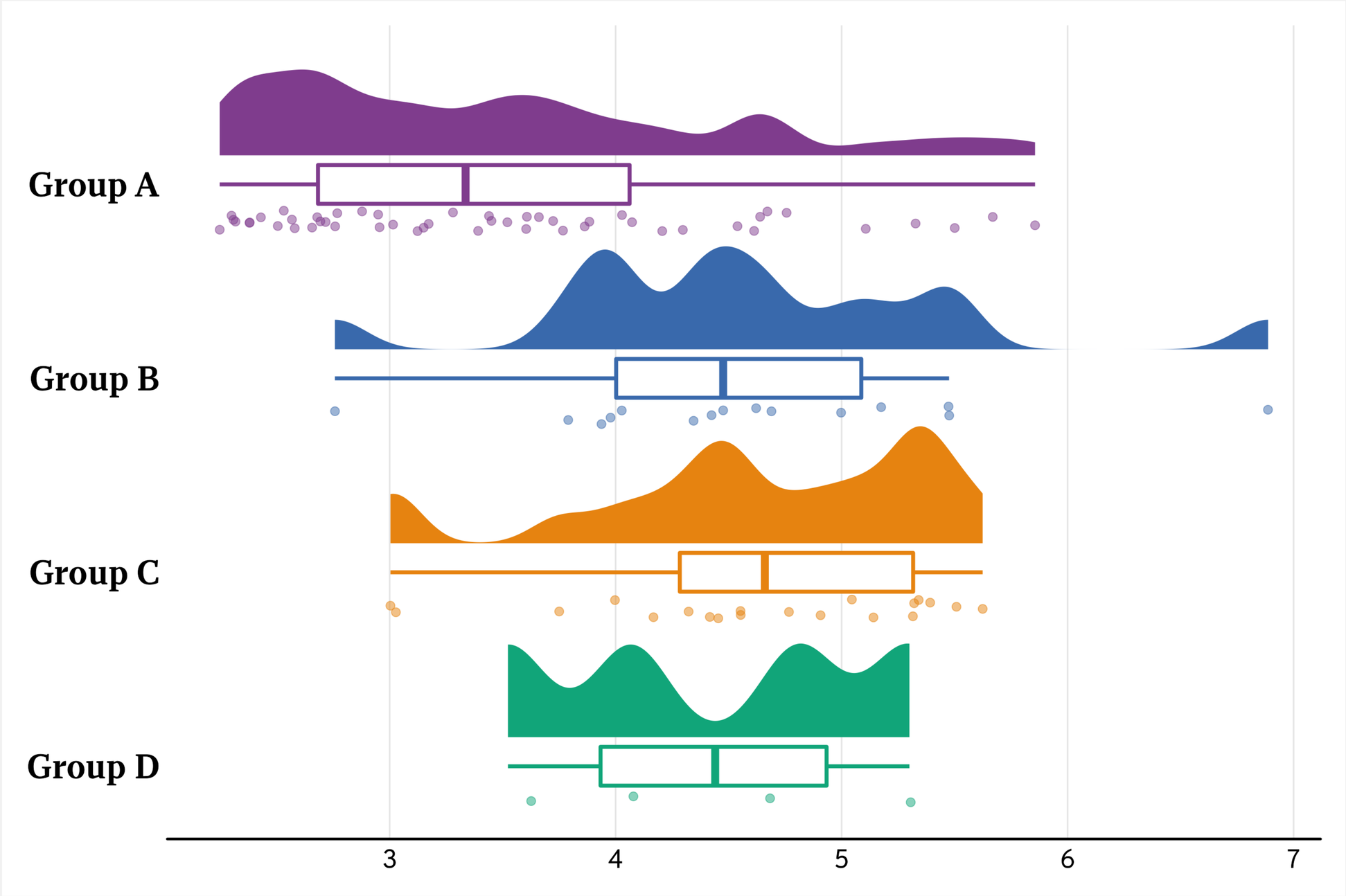
Visualization by Cédric Scherer



"Not my cup of coffee", #TidyTuesday Contribution

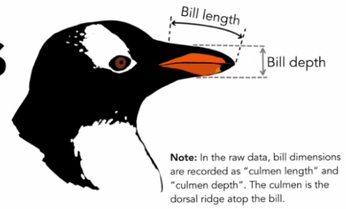
cedricscherer.com @CedScherer z3tt


7
6
5
4
3
Group A
Group B
Group C
Group D

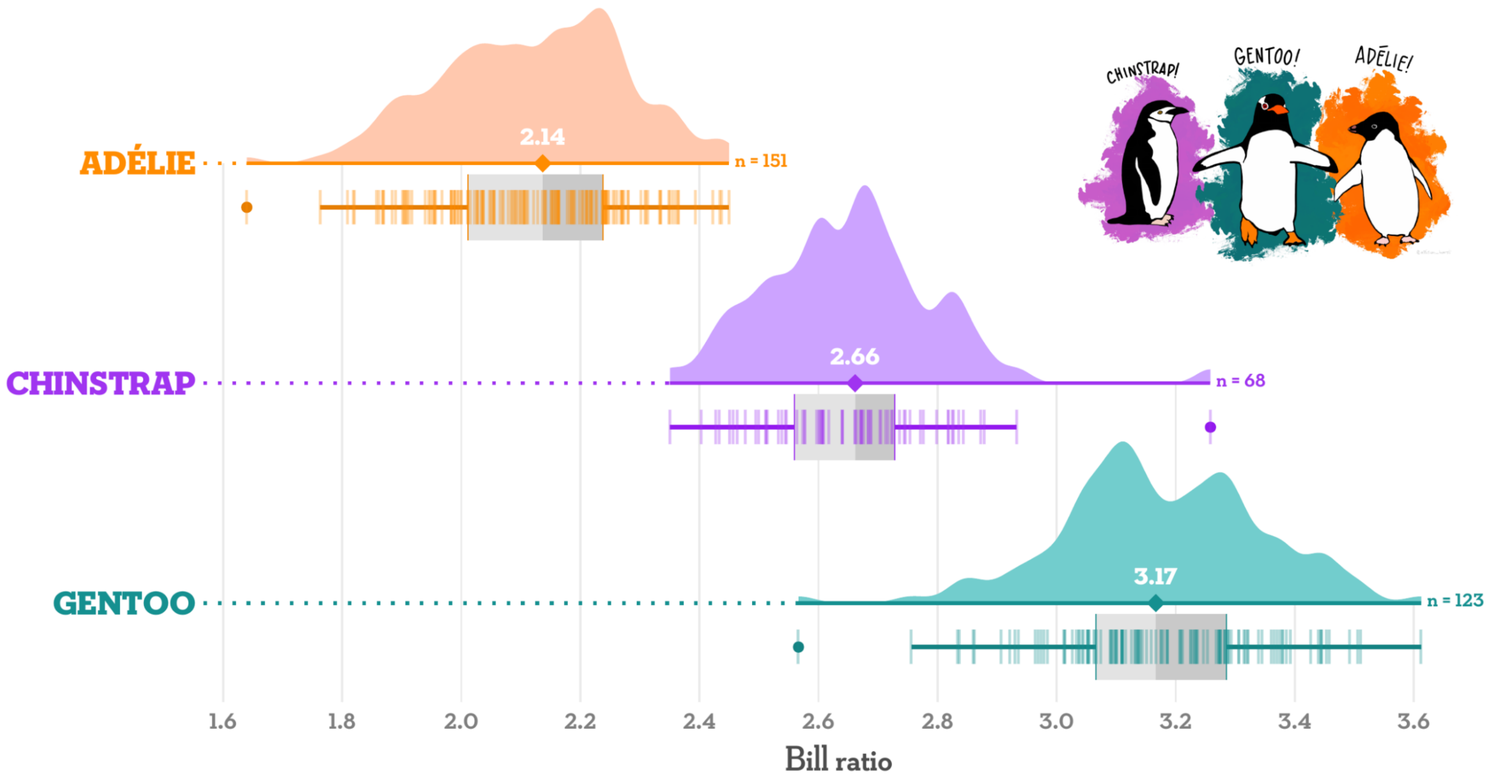

Group A
Group B
Group C
Group D
3
4
5
6
7

BILL DIMENSIONS OF BRUSH-TAILED PENGUINS

Pygoscelis adéliae (Adélie penguin) • P. antarctica (Chinstrap penguin) • P. papua (Gentoo penguin)

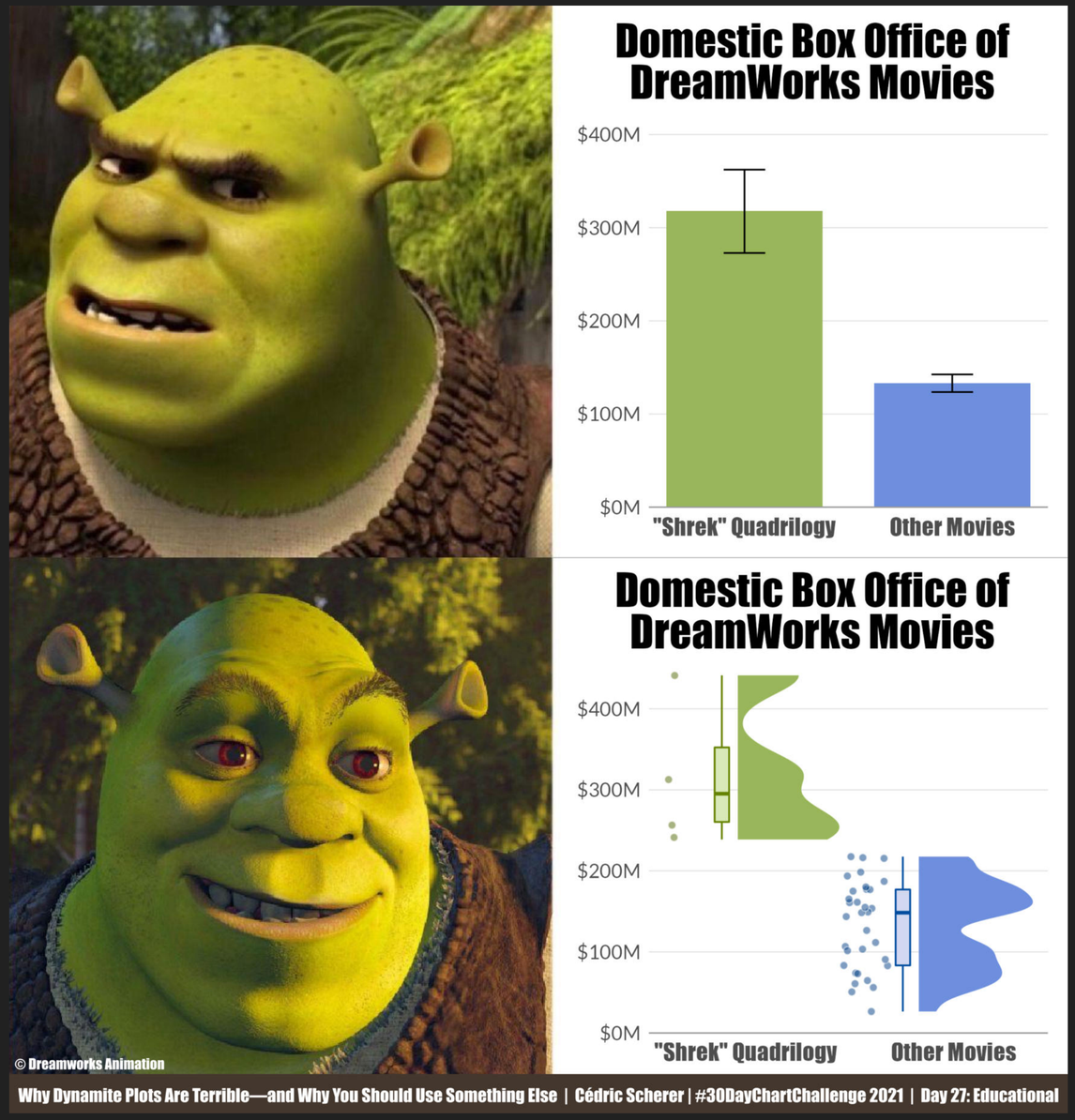


Distribution of the bill ratio, estimated as bill length divided by bill depth

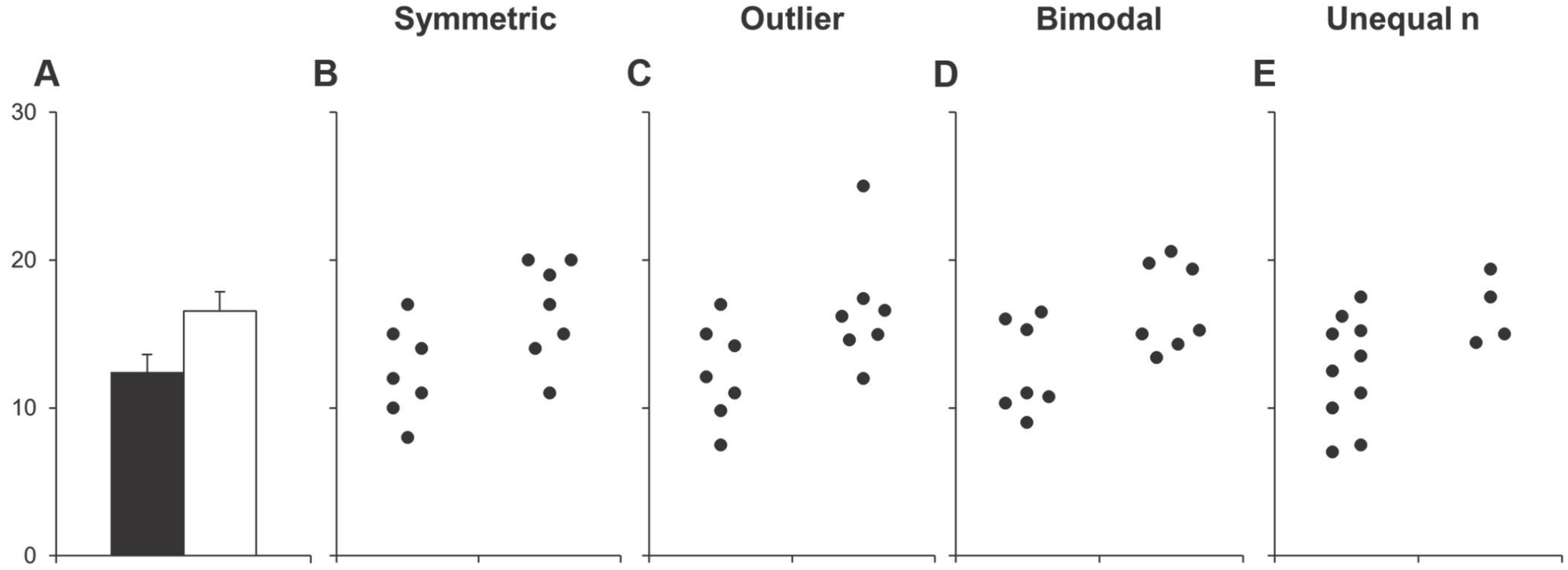


Note: In the original data, bill dimensions are recorded as "culmen length" and "culmen depth". The culmen is the dorsal (upper) ridge of a bird's bill.
Visualization: Cédric Scherer • Data: Gorman, Williams & Fraser (2014) DOI: 10.1371/journal.pone.0090081 • Illustrations: Allison Horst

"Bill dimensions of brush-tailed penguins", part of a #TidyTuesday contribution

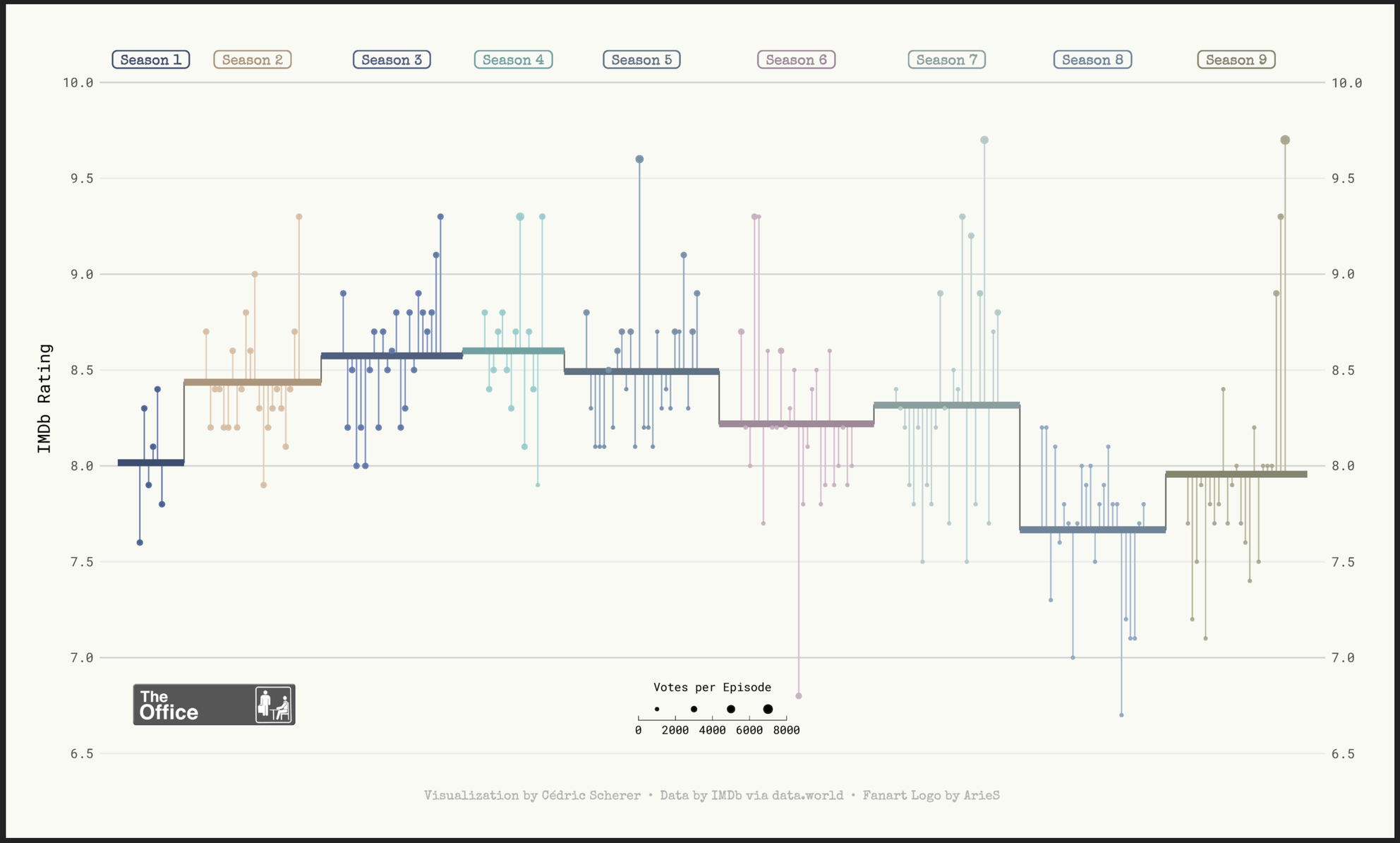


“Why dynamite plots are terrible - and why you should use something else”, #30DayChartChallenge Contribution



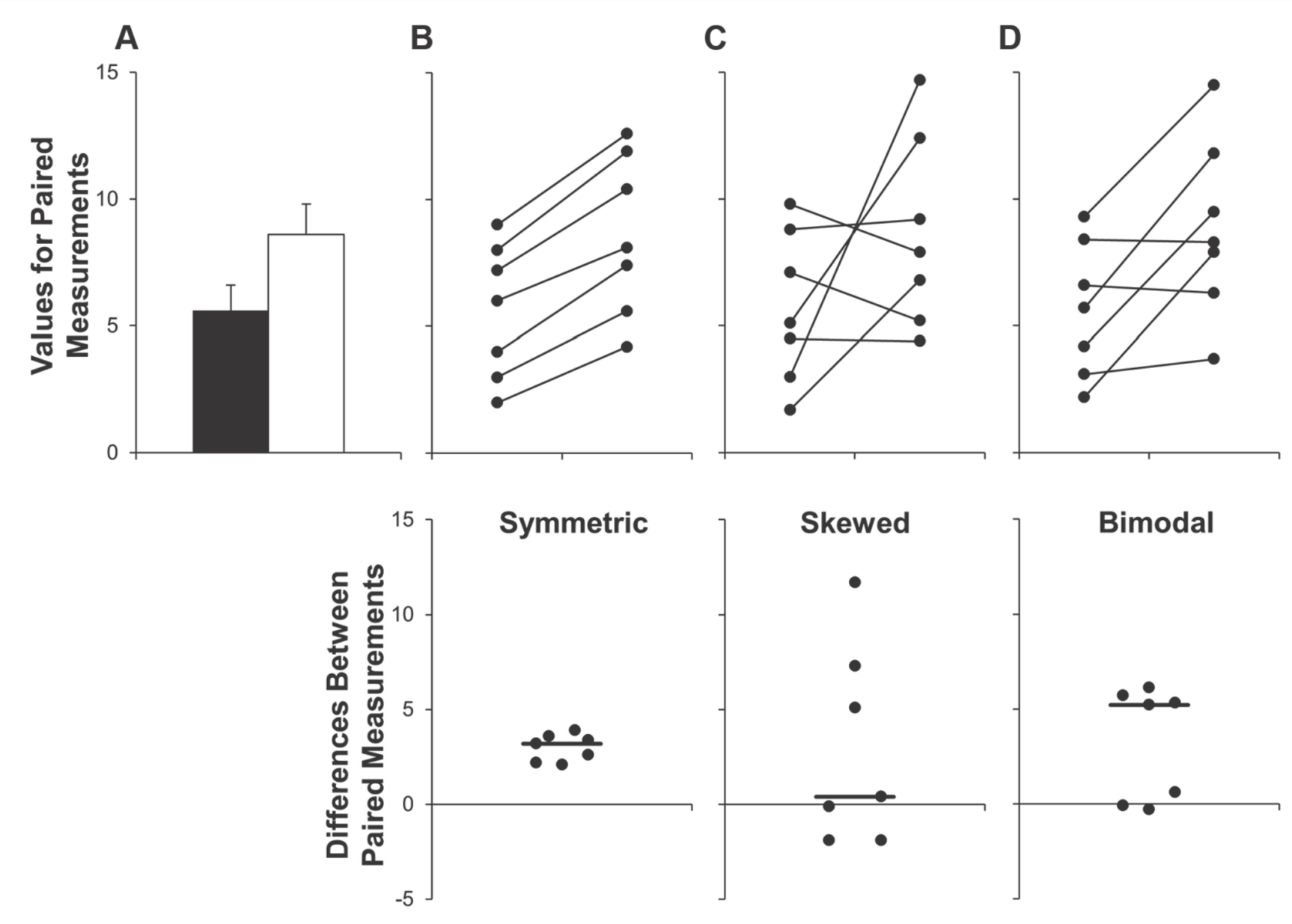
Test	p value			
T-test: Equal var.	0.035	0.050	0.026	0.063
T-test: Unequal var.	0.035	0.050	0.026	0.035
Wilcoxon	0.054	0.073	0.128	0.103

Source: Weissgerber et al. (2015) PLoS Biology

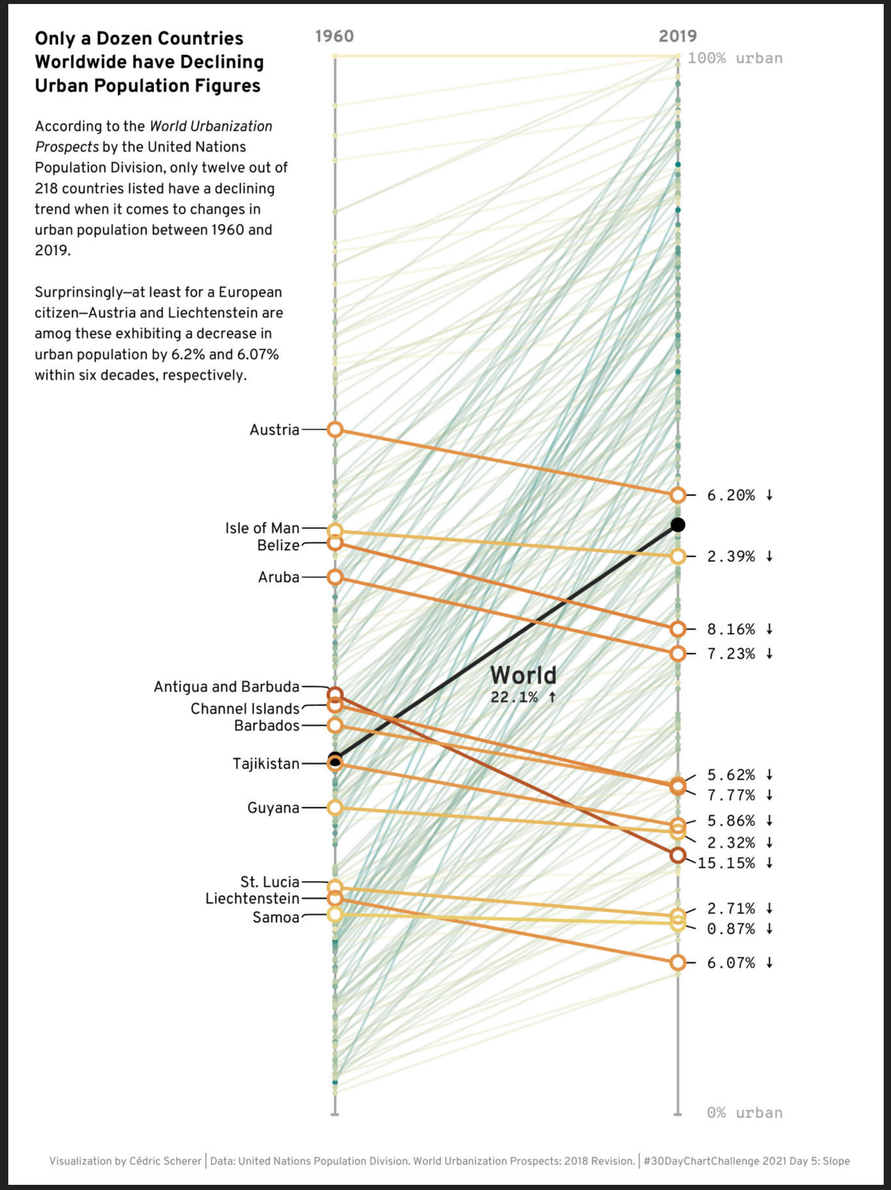


The Office IMDb Ratings, #TidyTuesday Contribution

cedricscherer.com @CedScherer z3tt



Source: Weissgerber et al. (2015) PLoS Biology

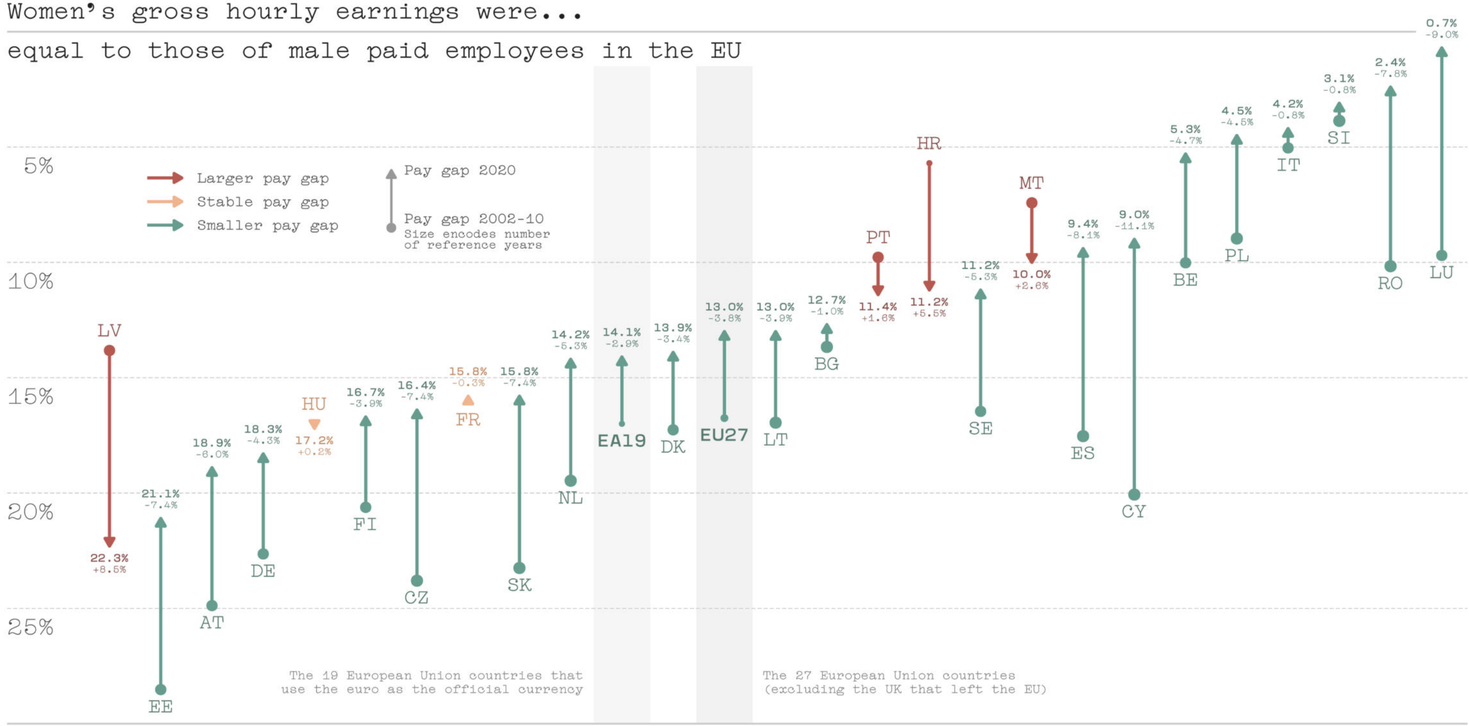


Urban Population Trends, #30DayChartChallenge Contribution

cedricscherer.com @CedScherer z3tt

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.*

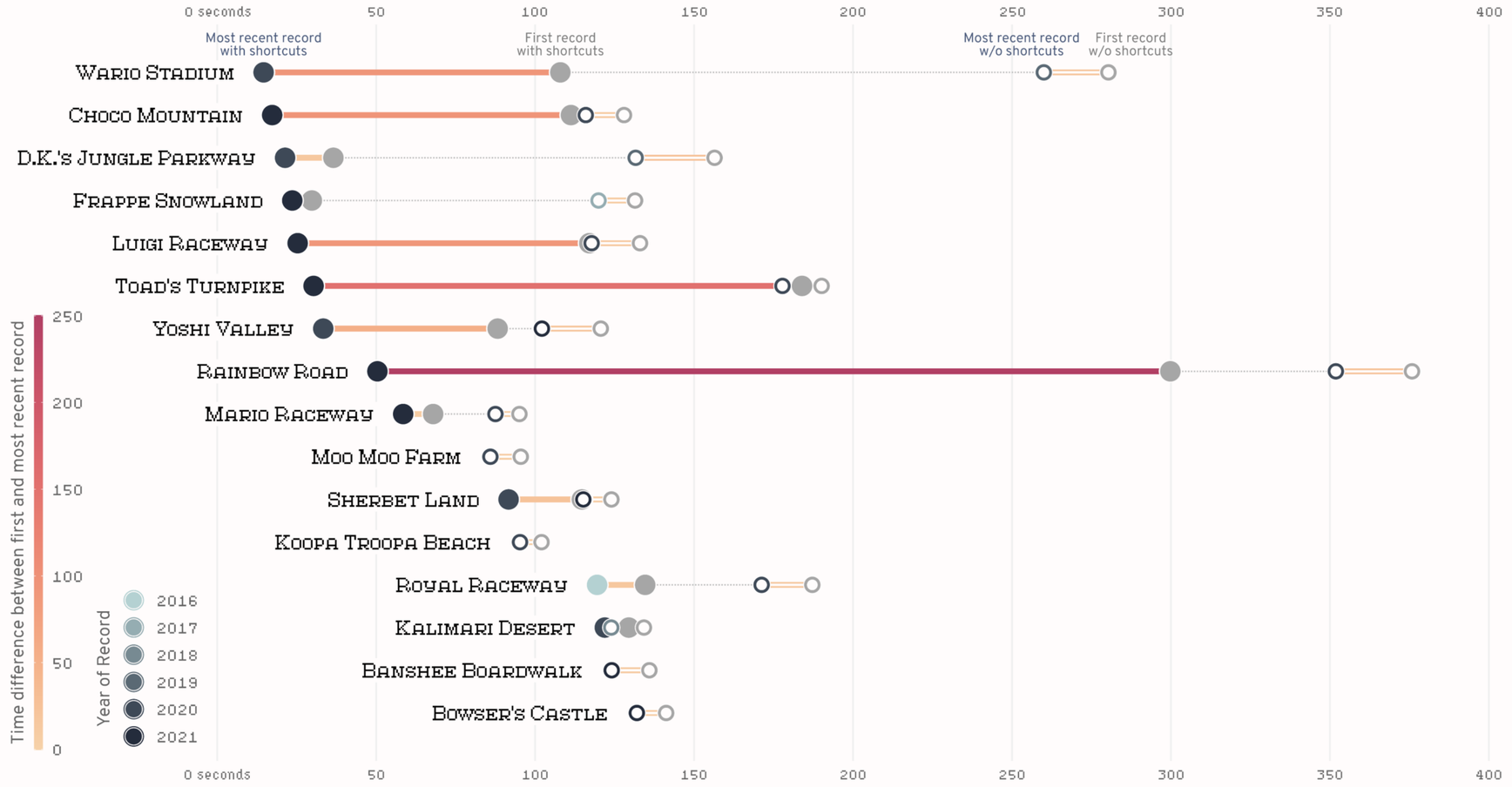
Graphic: Cédric Scherer • Data: Eurostat (SDG_05_20; no data for Greece and Ireland) • #InternationalWomensDay2022

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

cedricscherer.com @CedScherer z3tt

Let's-a-Go! You May Still Have Chances to Grab a New World Record for Mario Kart 64

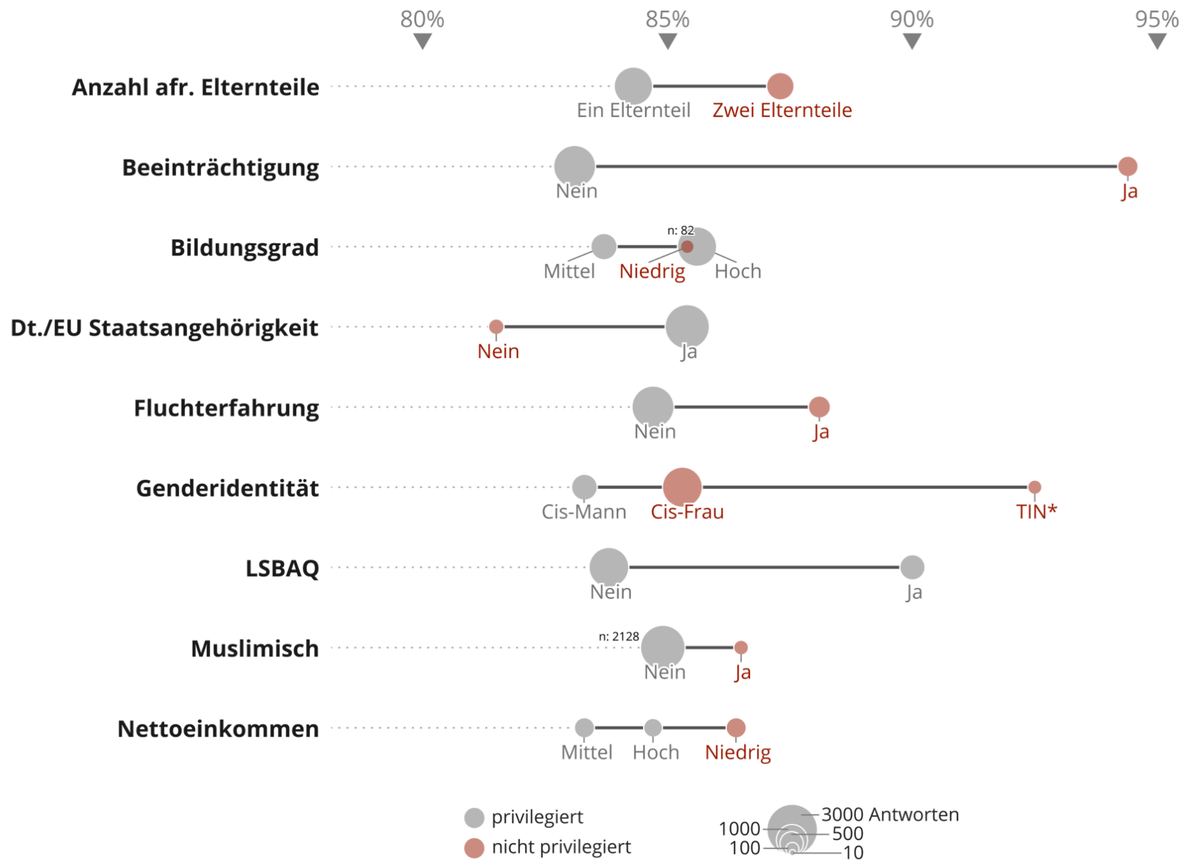
Most world records for Mario Kart 64 were achieved pretty recently (13 in 2020, 10 in 2021). On several tracks, the players considerably improved the time needed to complete three laps when they used shortcuts (*Choco Mountain*, *D.K.'s Jungle Parkway*, *Frappe Snowland*, *Luigi Raceway*, *Rainbow Road*, *Royal Raceway*, *Toad's Turnpike*, *Wario Stadium*, and *Yoshi Valley*). Actually, for three out of these tracks the previous records were more than halved since 2020 (*Luigi Raceway*, *Rainbow Road*, and *Toad's Turnpike*). Four other tracks still have no records for races with shortcuts (*Moo Moo Farm*, *Koopa Troopa Beach*, *Banshee Boardwalk*, and *Bowser's Castle*). Are there none or did nobody find them yet? Pretty unrealistic given the fact that since more than 24 years the game is played all around the world—but maybe you're able to find one and obtain a new world record?



Visualization: Cédric Scherer • Data: mkwrs.com/mk64

Source: "Afrozensus 2020" by Citizens For Europe & EOTO e.V.

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

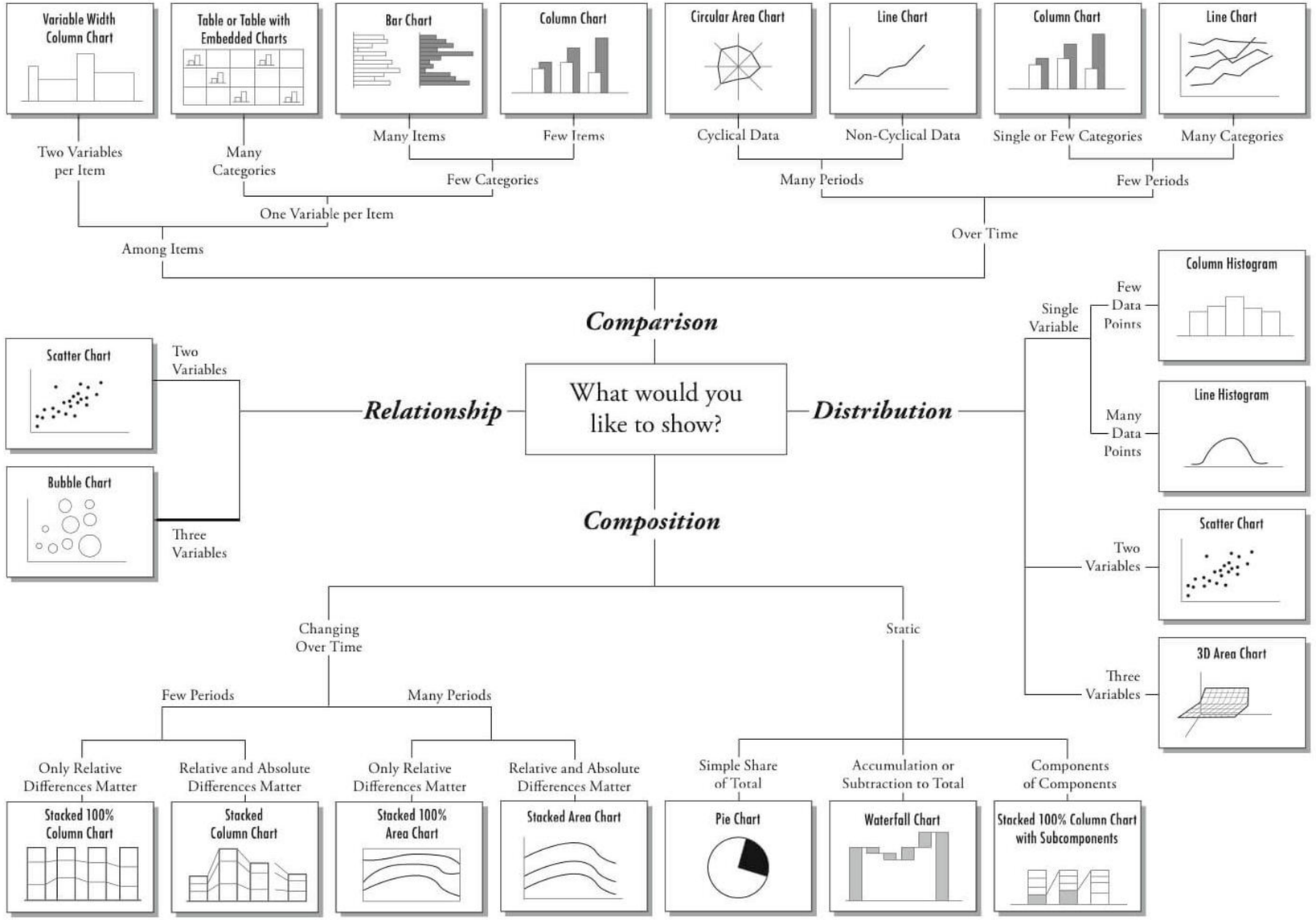


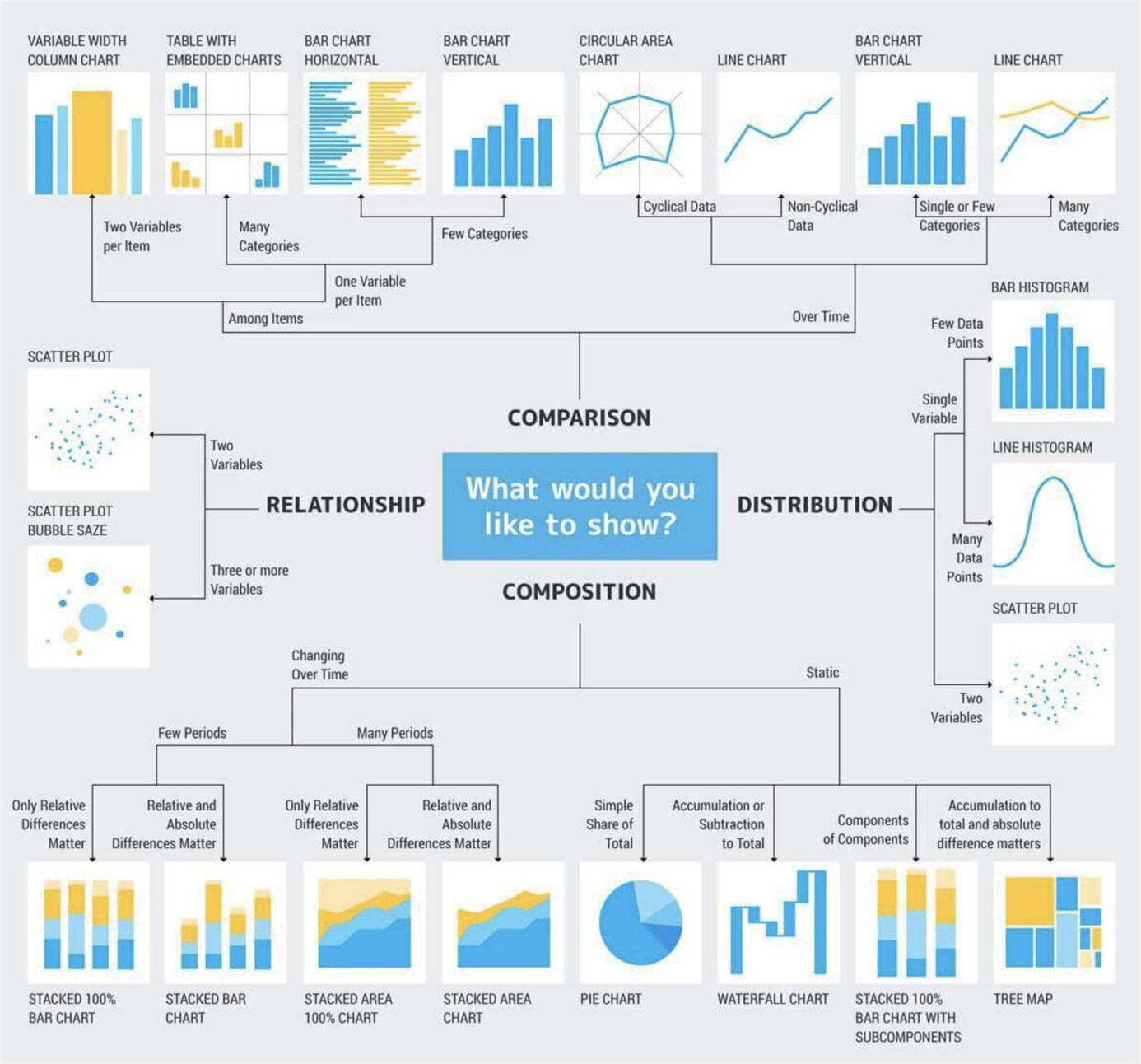
Lesebeispiel: *LSBAQ-Befragte des Afrozensus geben im Vergleich mit heterosexuellen Afrozensus-Befragten häufiger an, im Lebensbereich „Medien und Internet" in den letzten zwei Jahren Diskriminierung erlebt zu haben.*

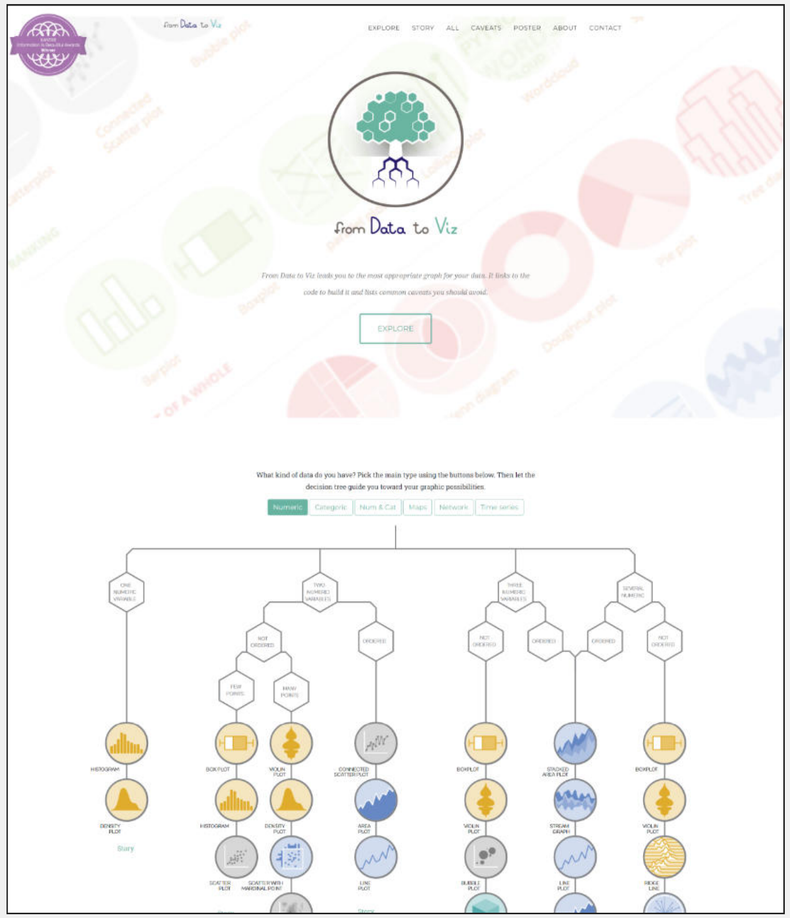
Quelle: Abb. 46 in Aikins, M A; Bremberger,T; Aikins, J K; Gyamerah, D; Yıldırım-Caliman, D (2021): Afrozensus 2020 | Datenteam: Reiber, L; Vivanco, J | Design: Scherer, C

Source: "Afrozensus 2020" by Citizens For Europe & EOTO e.V.

Chart Suggestions—A Thought-Starter

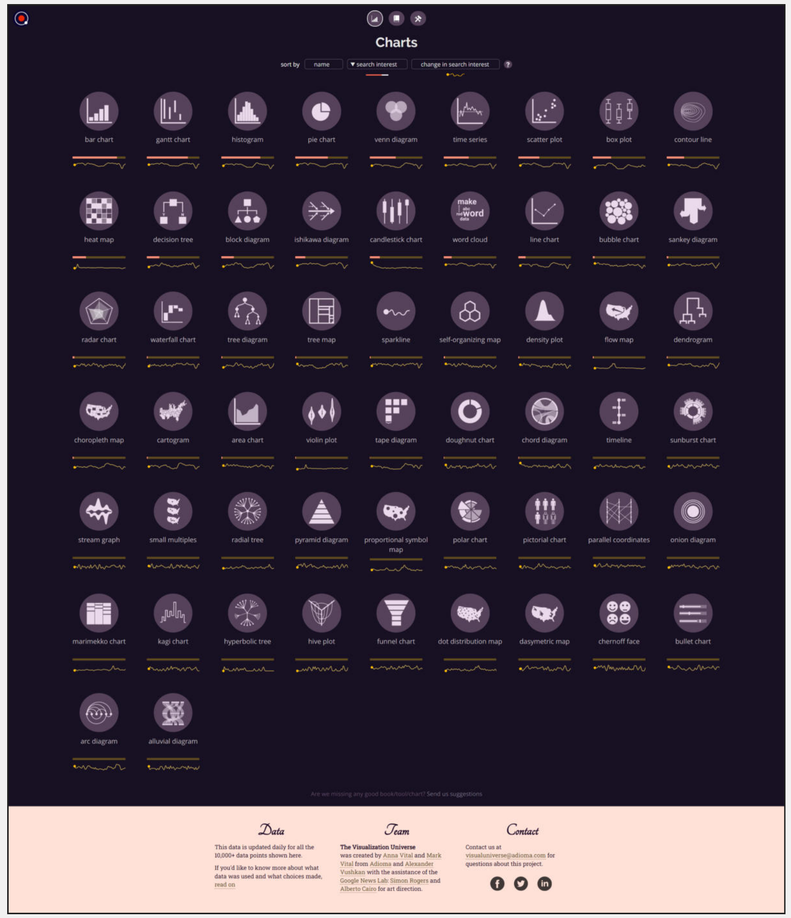



What would you like to show?
COMPARISON
RELATIONSHIP
DISTRIBUTION
COMPOSITION
VARIABLE WIDTH COLUMN CHART
TABLE WITH EMBEDDED CHARTS
BAR CHART HORIZONTAL
BAR CHART VERTICAL
CIRCULAR AREA CHART
LINE CHART
BAR CHART VERTICAL
LINE CHART
Two Variables per Item
Many Categories
Few Categories
One Variable per Item
Among Items
Cyclical Data
Non-Cyclical Data
Single or Few Categories
Many Categories
Over Time
SCATTER PLOT
Two Variables
SCATTER PLOT BUBBLE SAZE
Three or more Variables
BAR HISTOGRAM
Few Data Points
Single Variable
LINE HISTOGRAM
Many Data Points
SCATTER PLOT
Two Variables
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
Accumulation to total and absolute difference matters
STACKED 100% BAR CHART
STACKED BAR CHART
STACKED AREA 100% CHART
STACKED AREA CHART
PIE CHART
WATERFALL CHART
STACKED 100% BAR CHART WITH SUBCOMPONENTS
TREE MAP



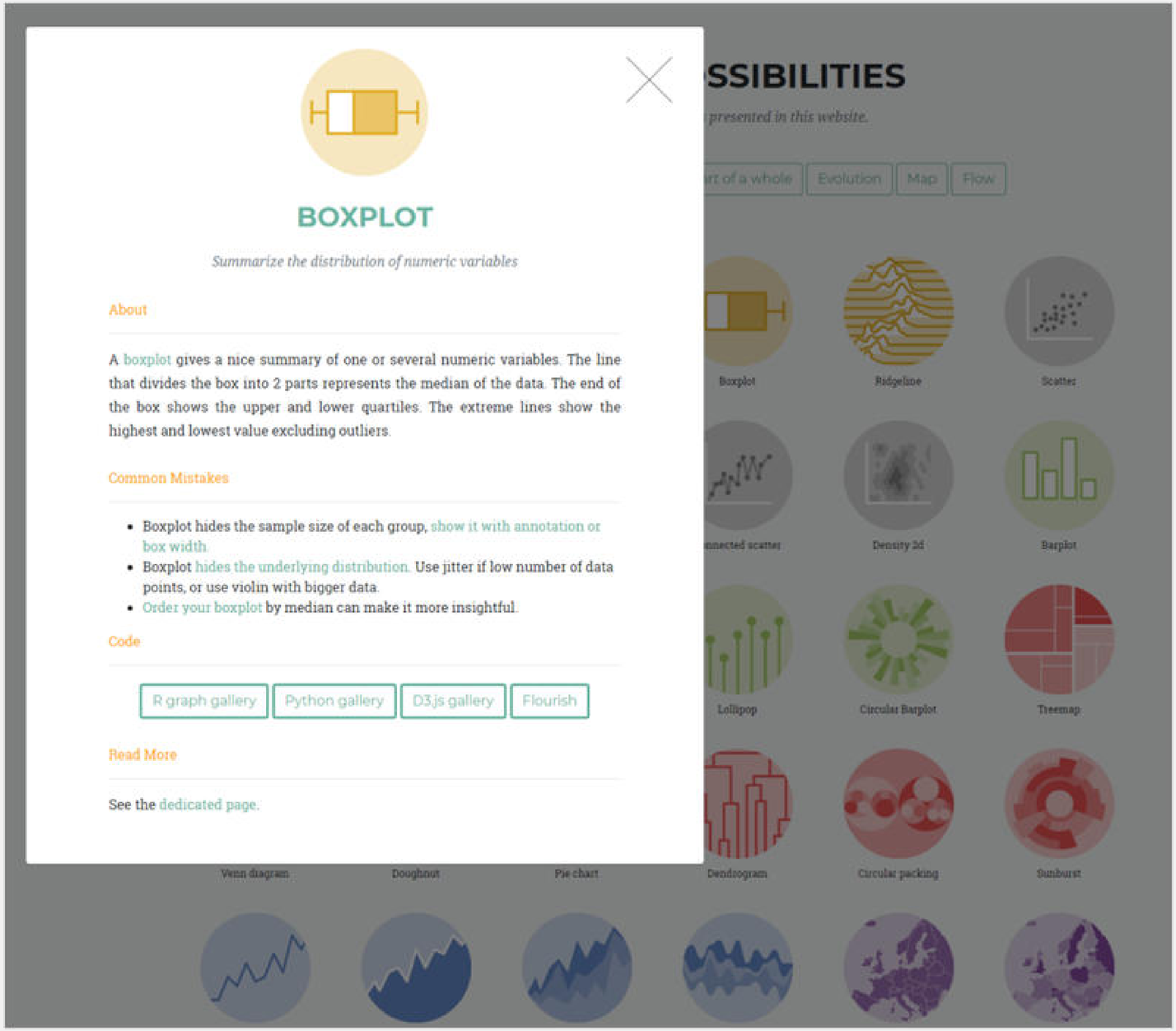
data-to-viz.com

datavizproject.com



visualizationuniverse.com

Source: data-to-viz.com

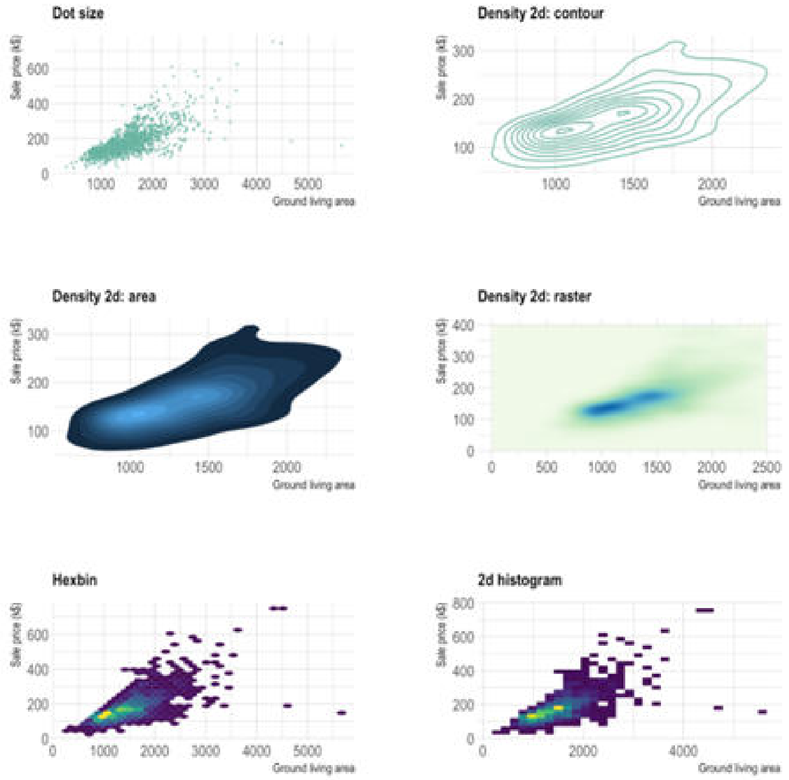


Source: data-to-viz.com

Overplotting

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

CODE

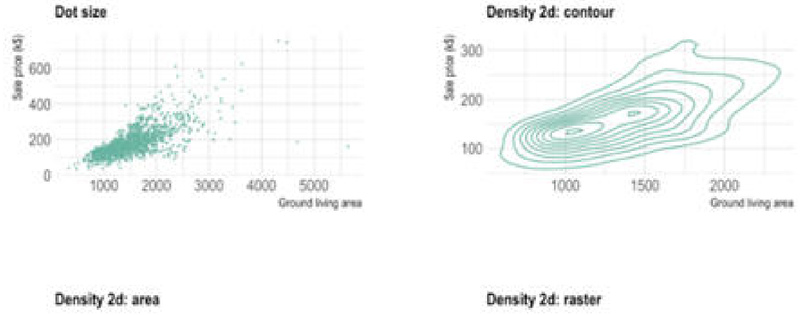


Going further

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

HIDE

```
# code for all graphics:
p <- data %>%
  ggplot( aes(x=GrLivArea, y=SalePrice/1000)) +
    theme_ipsum() +
    theme(
      plot.title = element_text(size=12)
    ) +
    ylab('Sale price (k$)') +
    xlab('Ground living area')

# Reduce dot size
p1 <- p + geom_point(color="#69b3a2", alpha=0.8, size=0.2) + ggtitle("Dot size")

# Use density estimate
p2 <- p + geom_density2d(color="#69b3a2") + ggtitle("Density 2d: contour")

# Use density estimate (area)
p3 <- p + stat_density_2d(aes(fill = ..level..), geom = "polygon") + ggtitle("Density 2d: area") + theme(legend.positio
n="none")

# With raster
p4 <- p +
  stat_density_2d(aes(fill = ..density..), geom = "raster", contour = FALSE) +
  scale_fill_distiller(palette=4, direction=1) +
  scale_x_continuous(expand = c(0, 0)) +
  scale_y_continuous(expand = c(0, 0)) +
  theme(
    legend.position='none'
  ) +
  ggtitle("Density 2d: raster") +
  xlim(0,2500) +
  ylim(0,400)

# Hexbin
p5 <- p + geom_hex() +
    scale_fill_viridis() +
    theme(legend.position="none") +
    ggtitle("Hexbin")

# 2d histogram
p6 <- p + geom_bin2d( ) +
    scale_fill_viridis( ) +
    theme(legend.position="none")  +
    ggtitle("2d histogram")

p1 + p2 + p3 + p4 + p5 + p6 + plot_layout(ncol = 2)
```



Source: data-to-viz.com

Group Exercise

Have a look at the “Chart Choice Helpers”

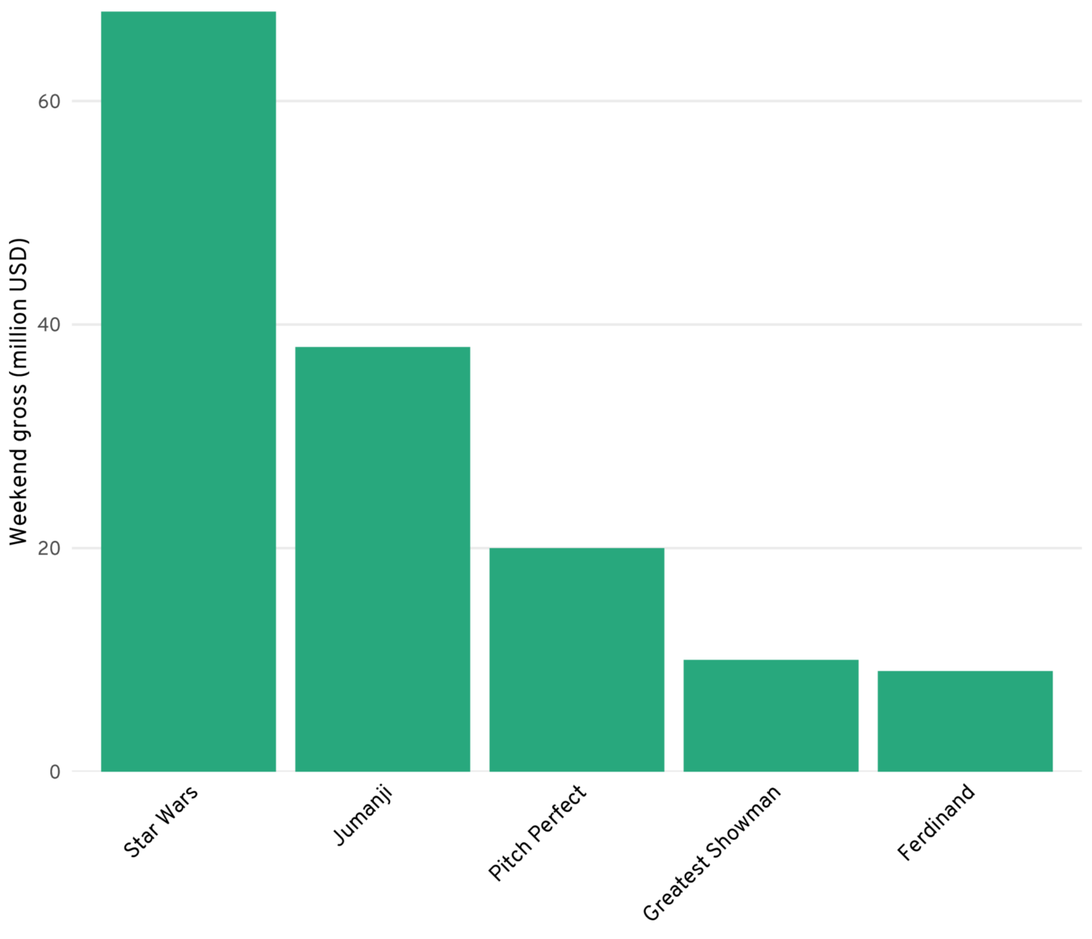
☛ Explore and discuss potential chart types for future data visualizaitons

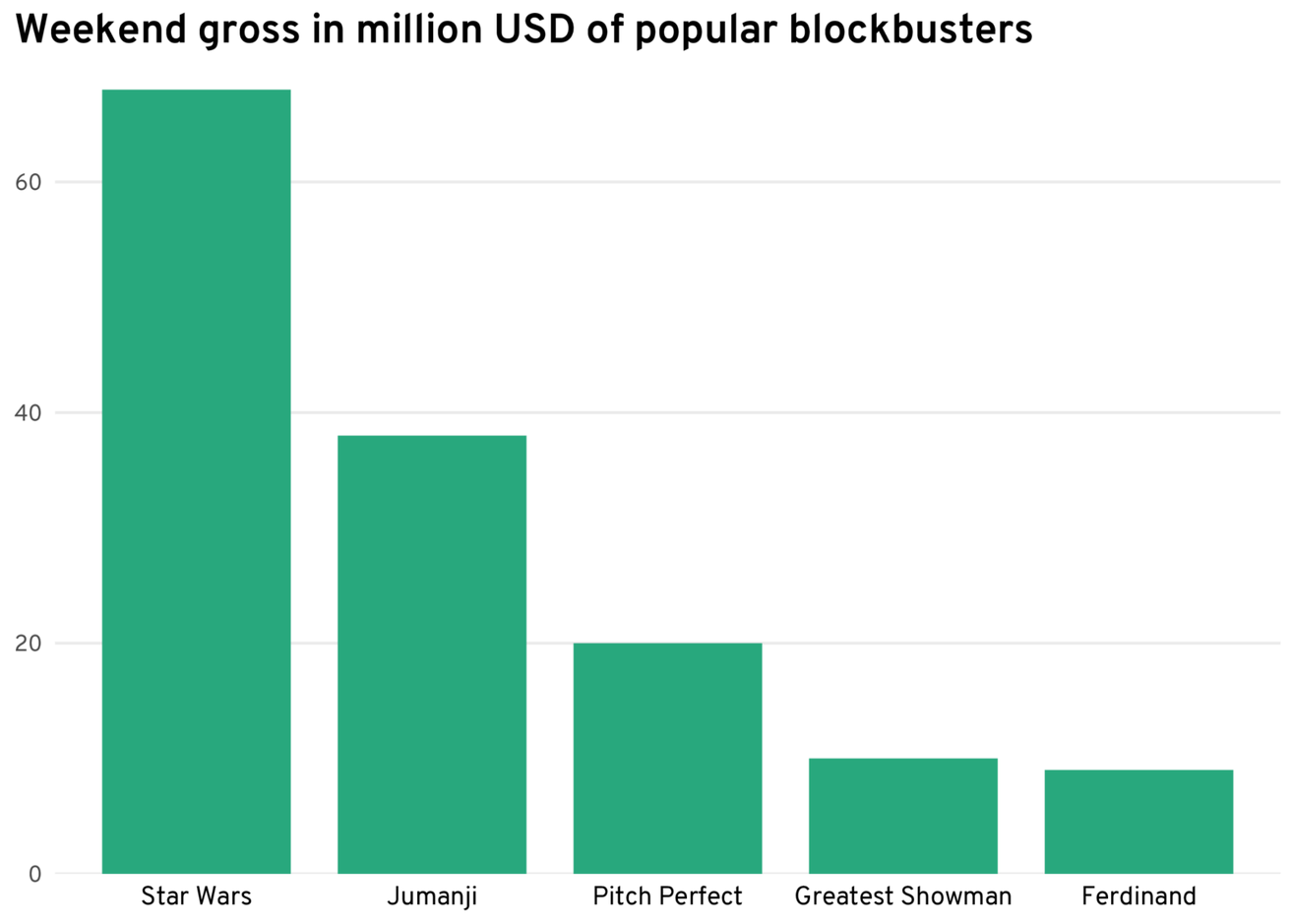
☛ Look up the chart types you don’t know yet

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

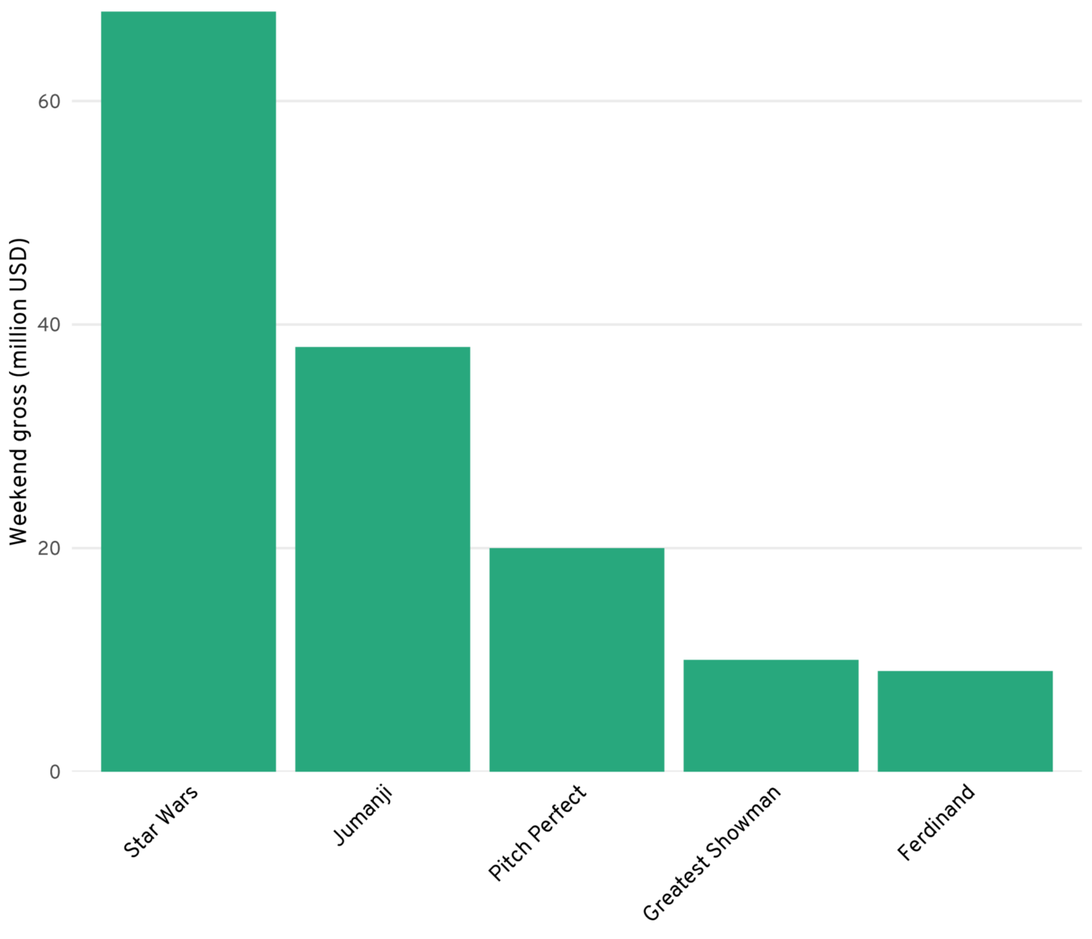
DATAVIZ ~~RULES~~ Recommendations

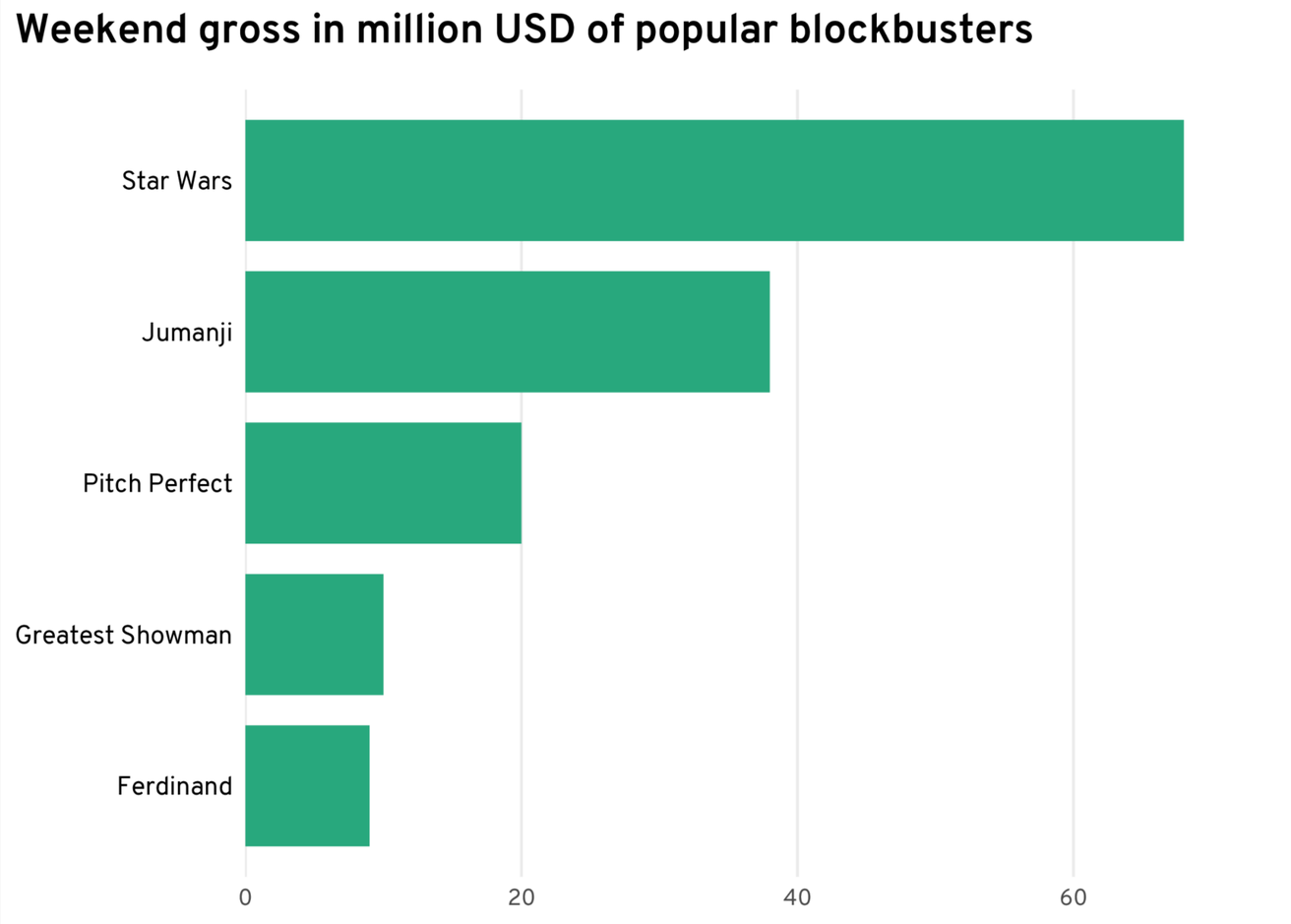
Don’t Tilt Labels



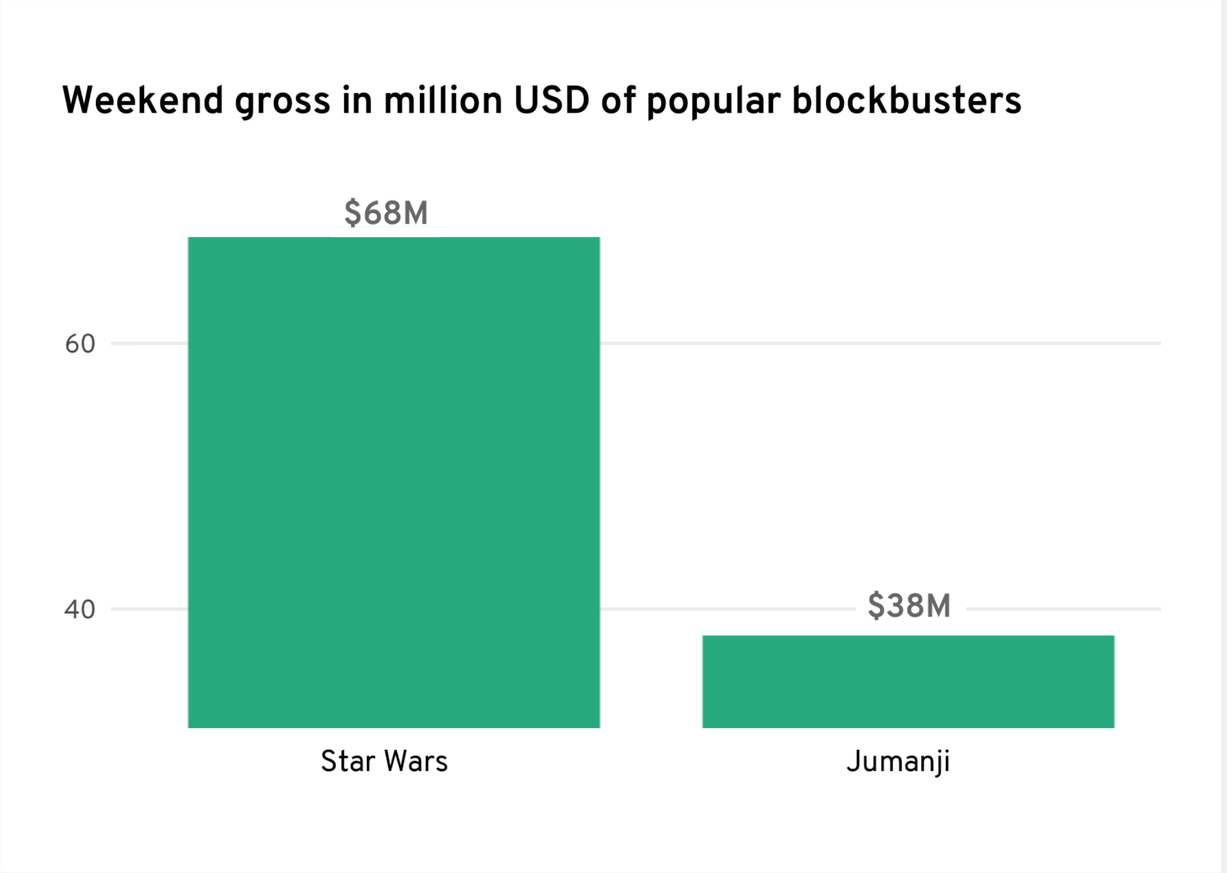


Don’t Tilt Labels

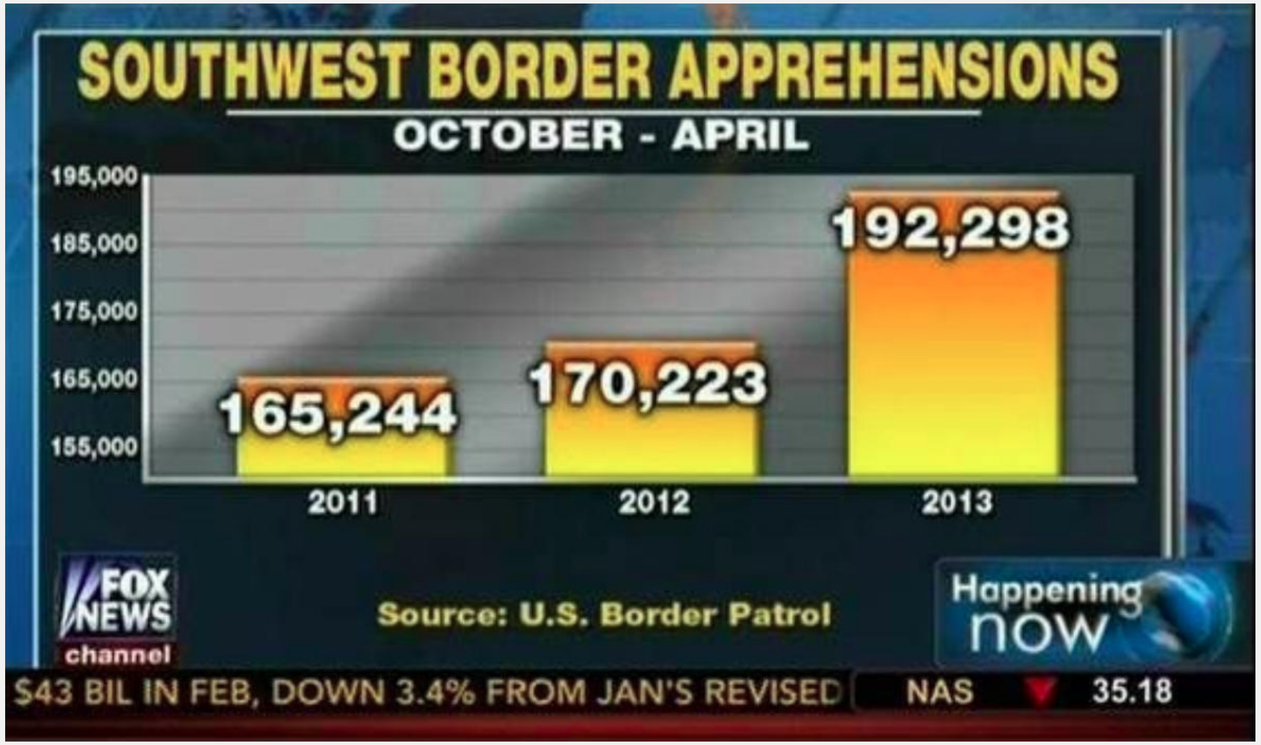




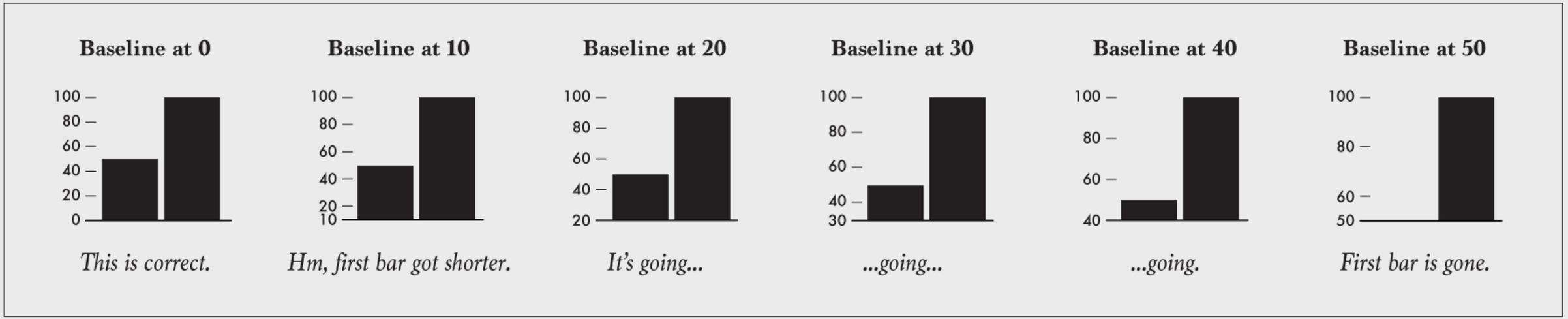
Always Stat at Zero



Always Stat at Zero

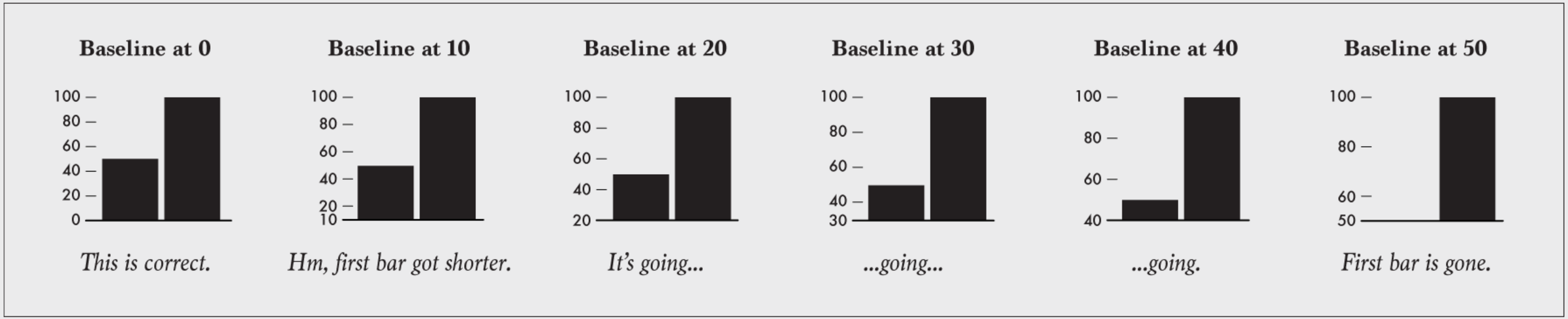


Bar Graphs Should Start at Zero



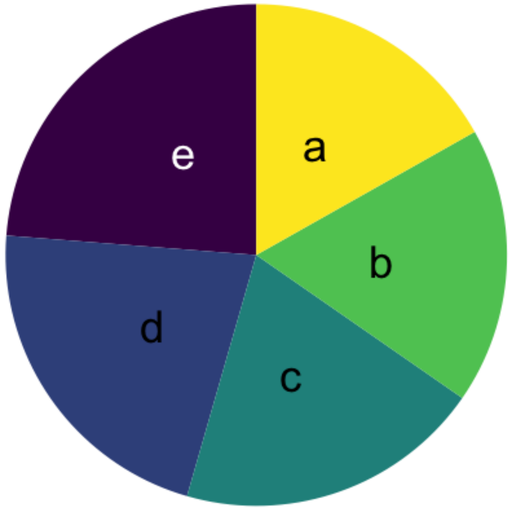
Source: Nathan Yau (flowingdata.com)

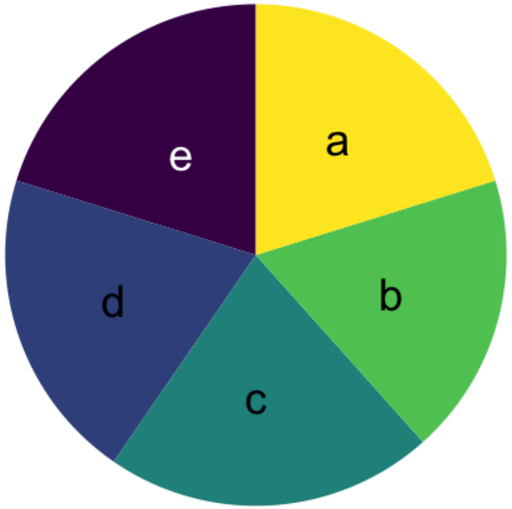
Bar Graphs Should Start at Zero

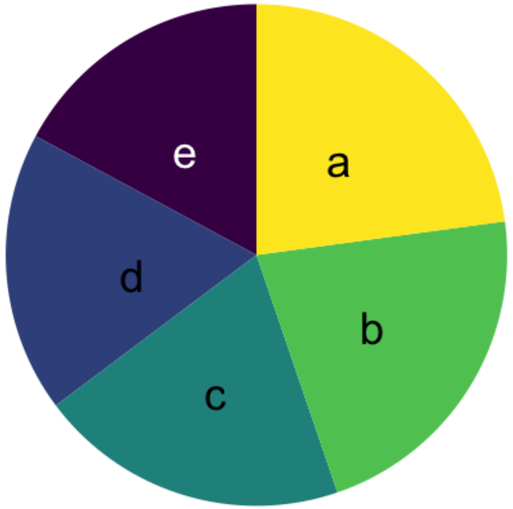


Source: Nathan Yau (flowingdata.com)

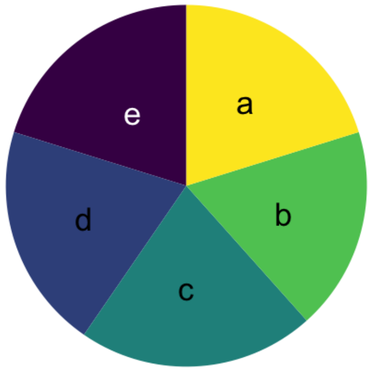
Beware of Pie Charts

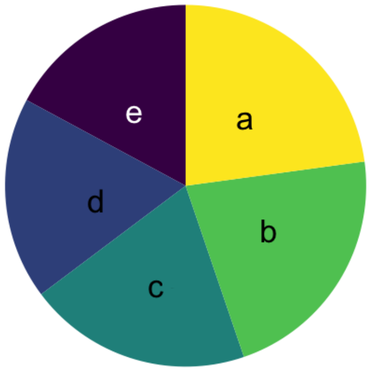


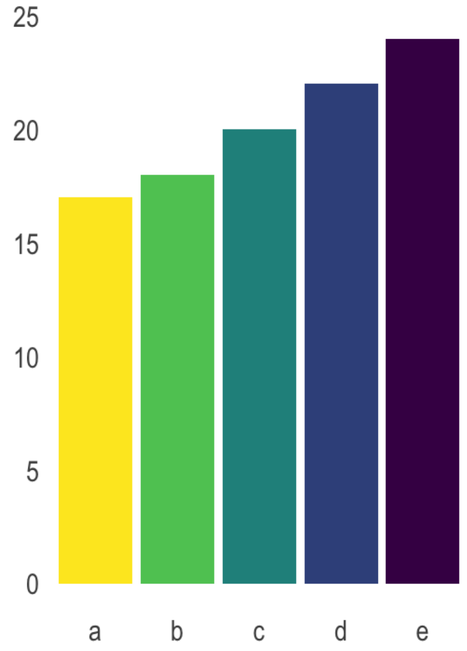


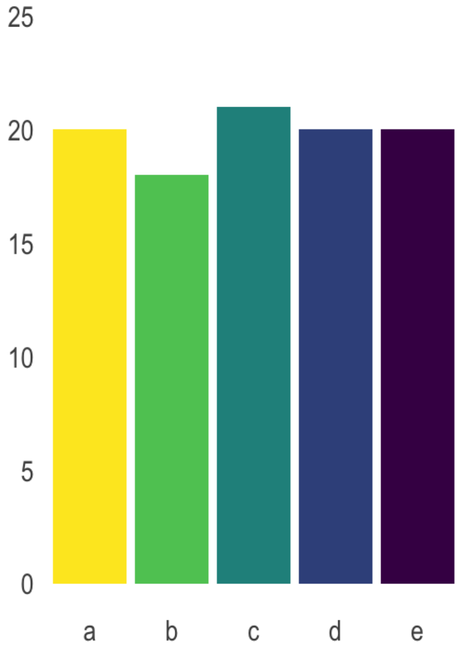


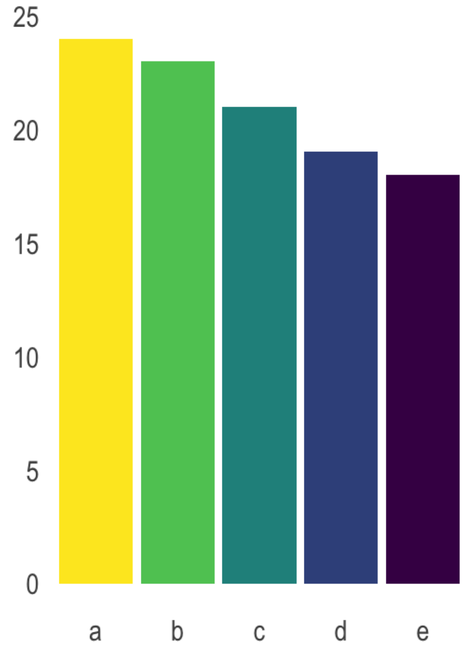

e
a
b
d
c


e
a
d
b
c

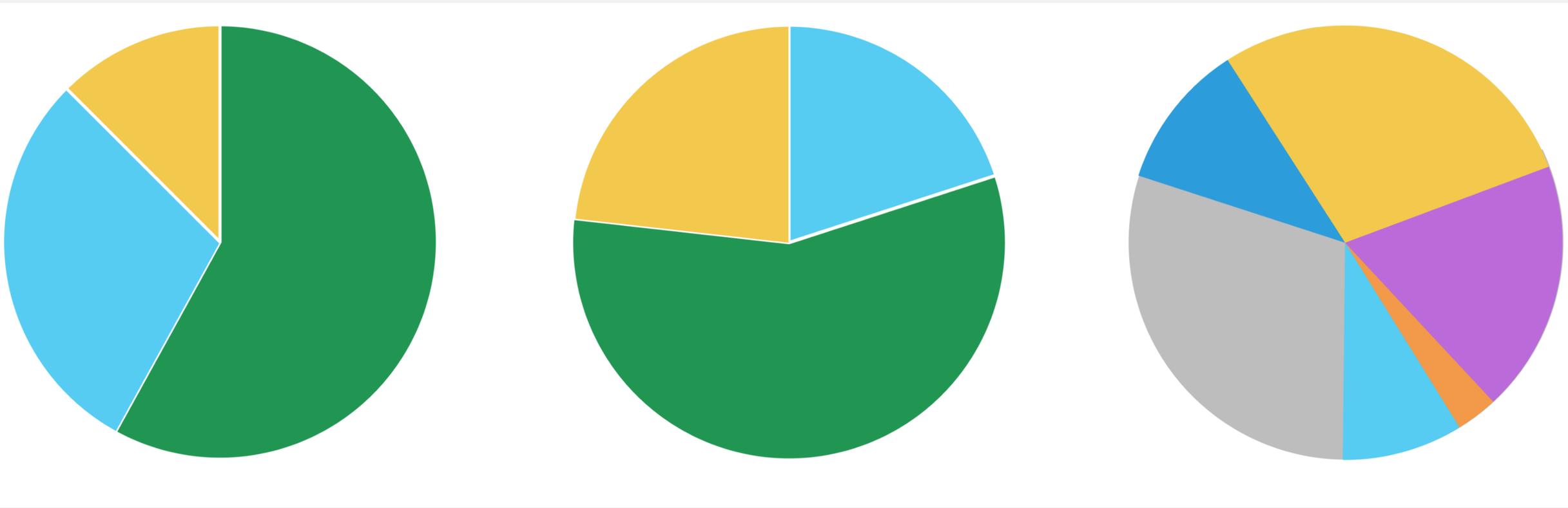

e
a
d
b
c


25
20
15
10
5
0
a
b
c
d
e


25
20
15
10
5
0
a
b
c
d
e

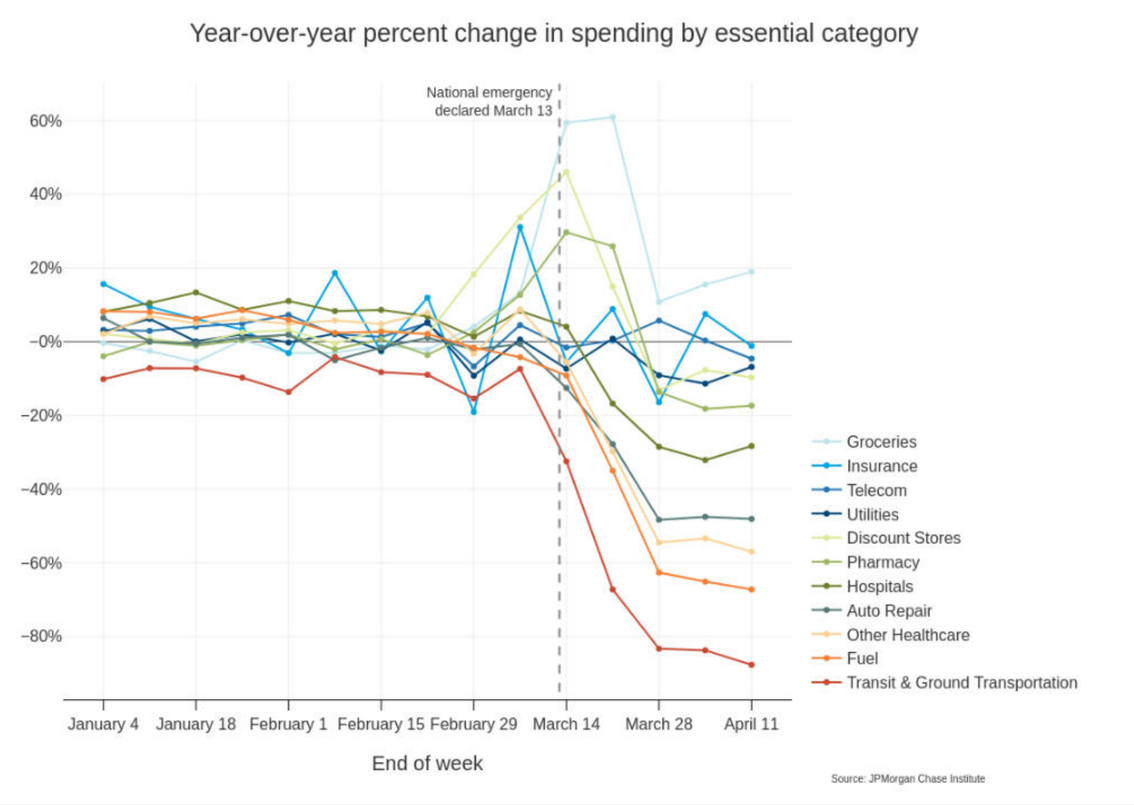

25
20
15
10
5
0
a
b
c
d
e

Beware of Pie Charts

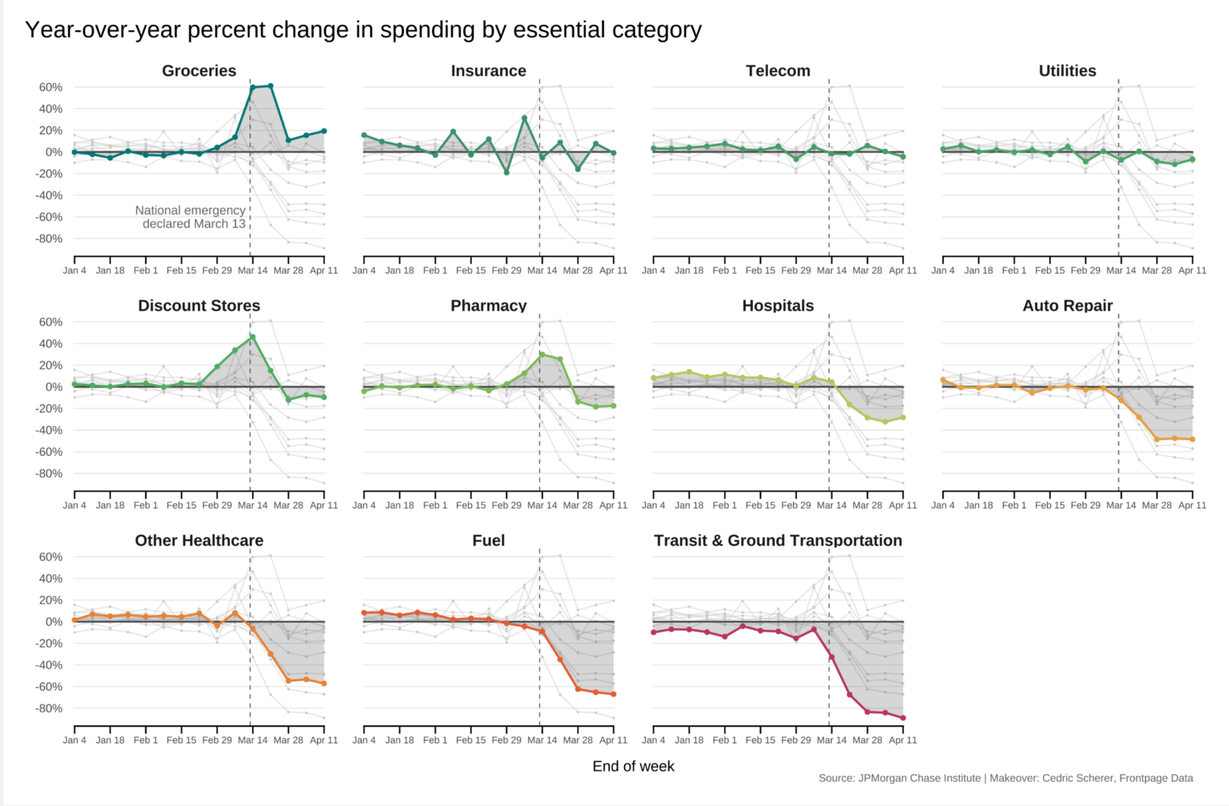


Source: "Hands-On Data Visualization" by Jack Dougherty & Ilya Ilyankou

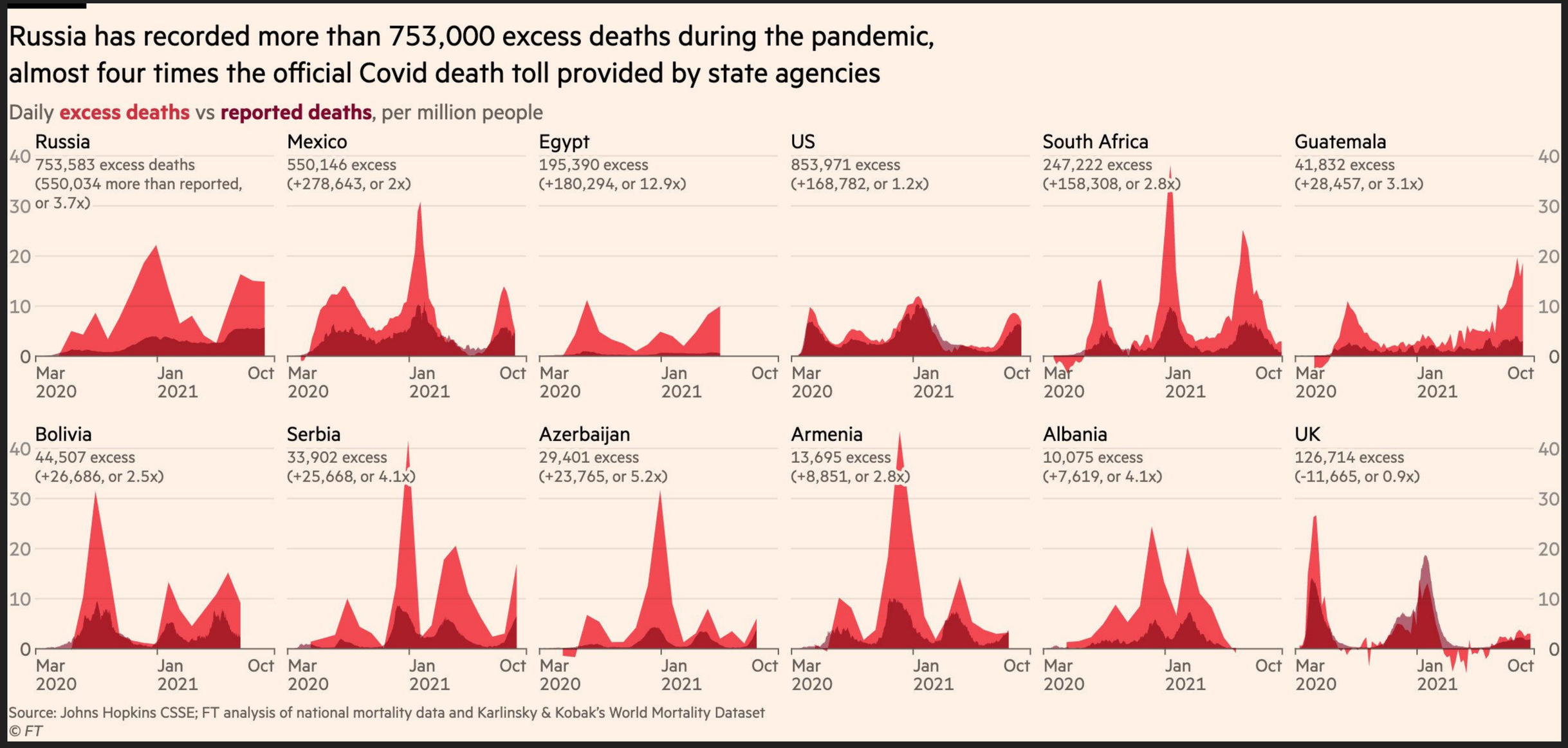
The Power of Small Multiples



Original graphic by JPMorgan Chase Institute



Makeover using small multiples



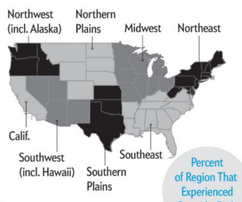
Source: John Burn-Murdoch (Financial Times)

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

Escalating Drought

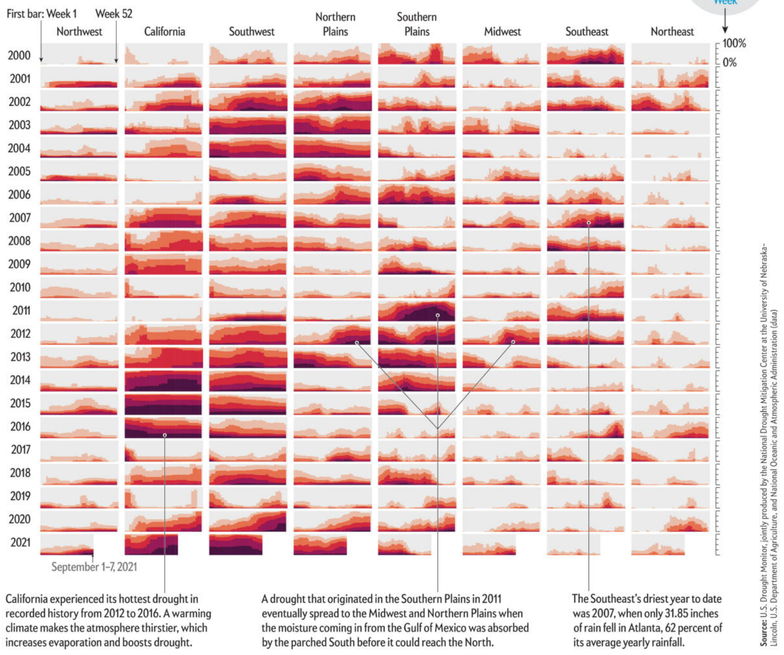
Climate change is intensifying periods of extreme dryness, particularly in the U.S. West

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

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

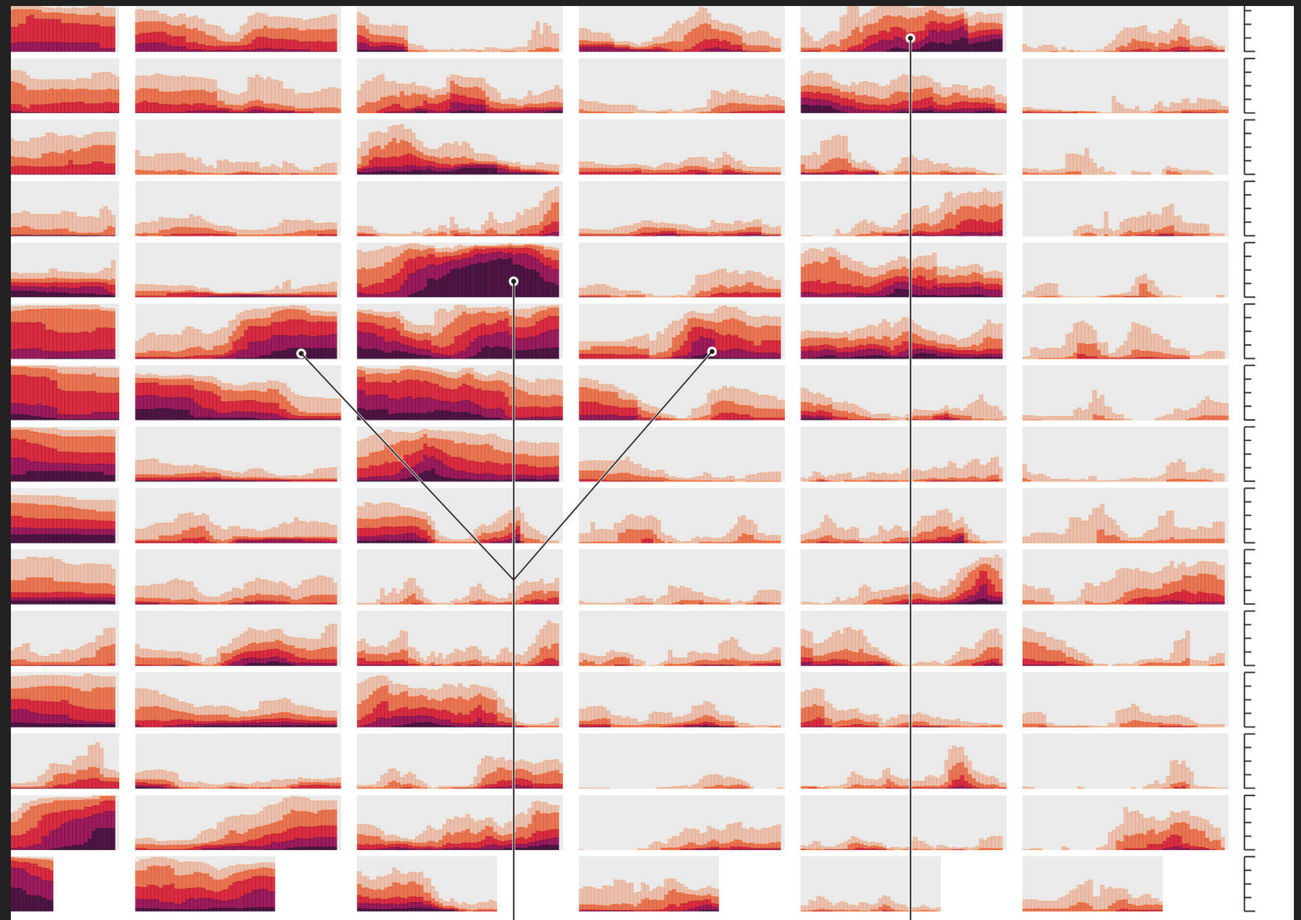


California experienced its hottest drought in recorded history from 2012 to 2016. A warming climate makes the atmosphere thirstier, which increases evaporation and boosts drought.

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

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

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



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.

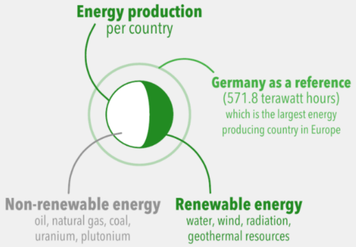
Source: "Escalating Drought", together with Georgios Karamanis for Scientific American, Issue Nov 2021

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.

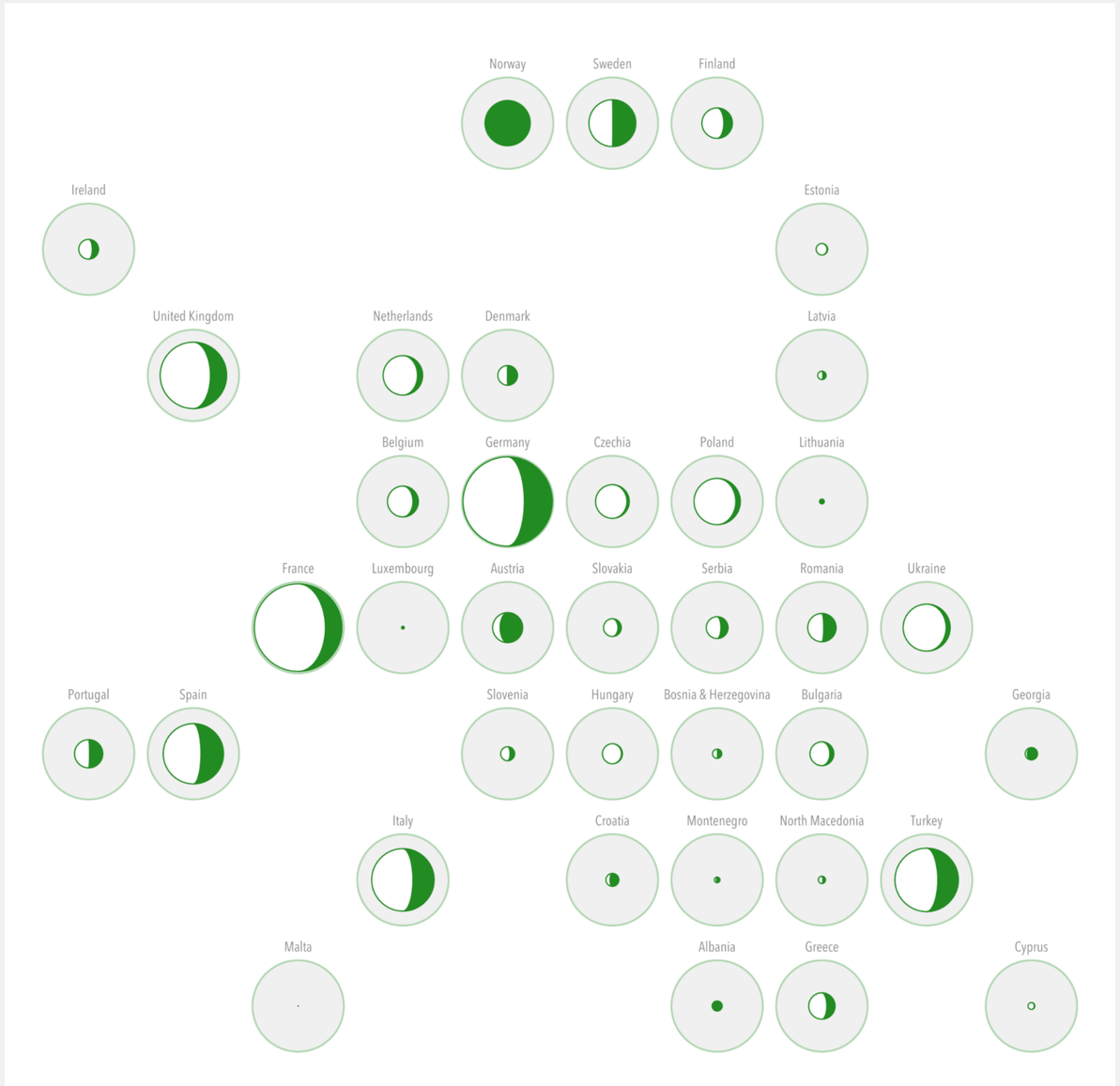
Renewable:	Nuclear:	Conventional thermal:
Germany	**France**	**Germany**
179.1 TWh	393.2 TWh	320.4 TWh
(31% of its production)	(71% of its production)	(56% of its production)

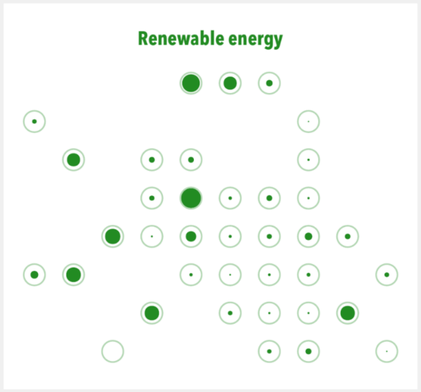
Renewable energy is energy that comes from resources that are naturally replenished such as sunlight, wind, water, and geothermal heat. Unlike fossil fuels, such as oil, natural gas and coal, or nuclear power sources such as uranium and plutonium, renewable energy regenerates naturally in a short period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

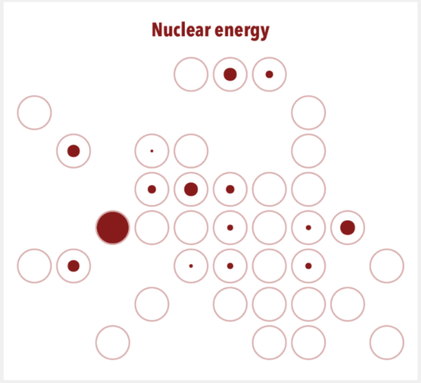


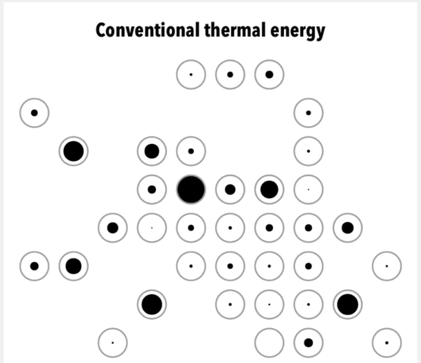
Norway had an electricity production almost entirely made up of renewable energy (98%). This makes Norway the second largest producer of this energy type in Europe. Interestingly, most of the renewable energy is produced by hydro power that take up 95% and only 3% by wind. In contrast, twelve European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



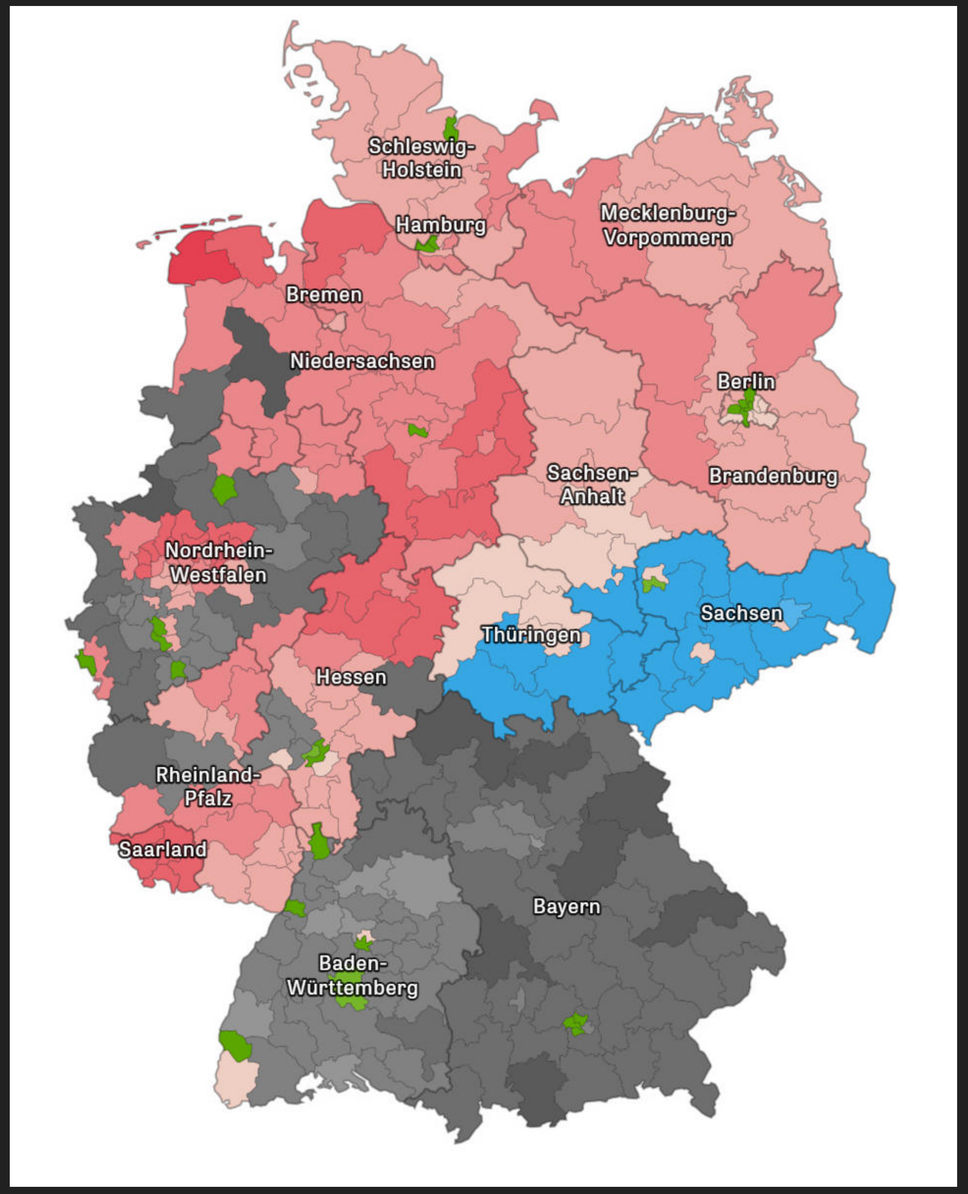


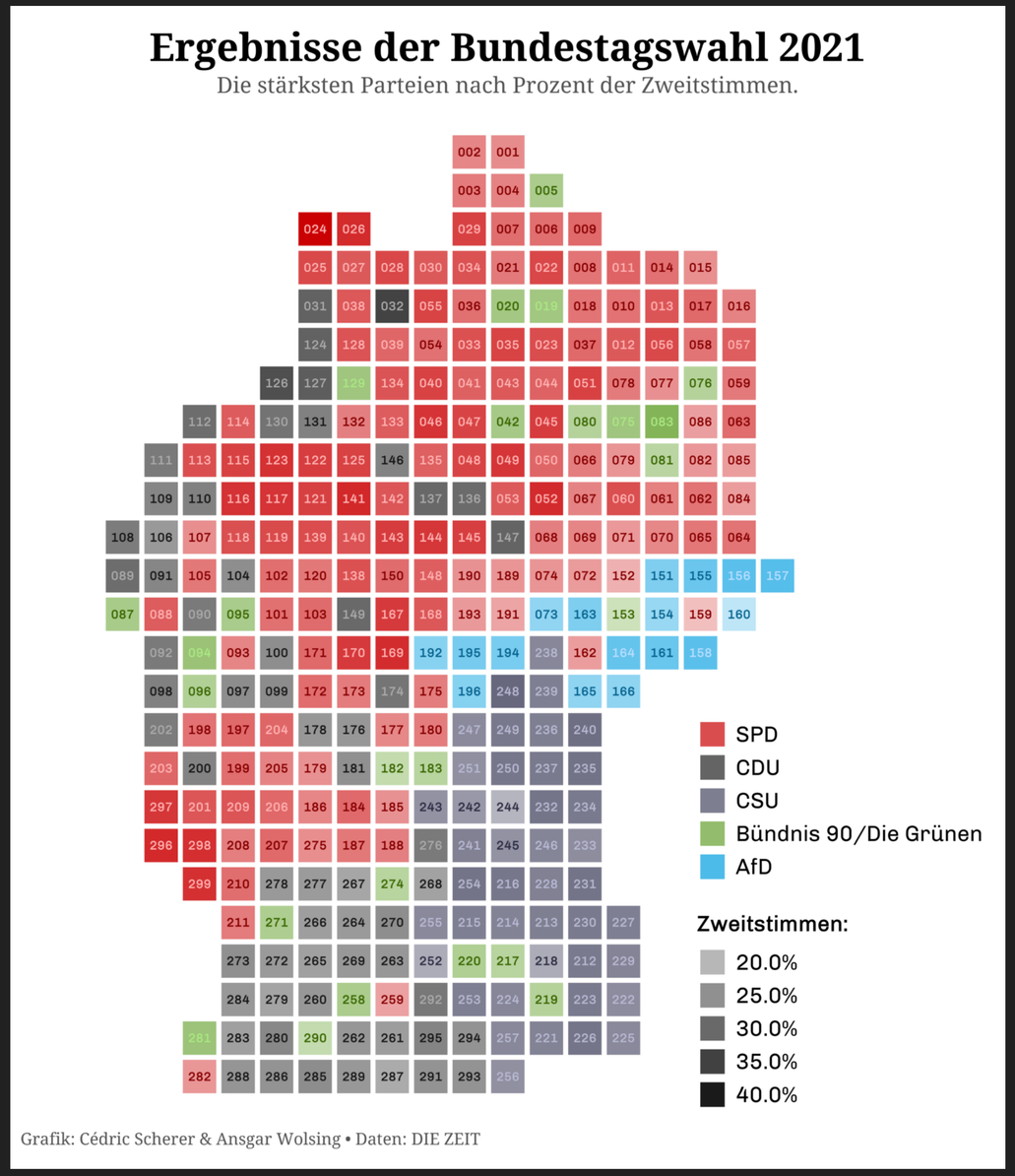




Source: "How European countries generated electricity in 2018", #TidyTuesday Contribution

cedricscherer.com @CedScherer z3tt



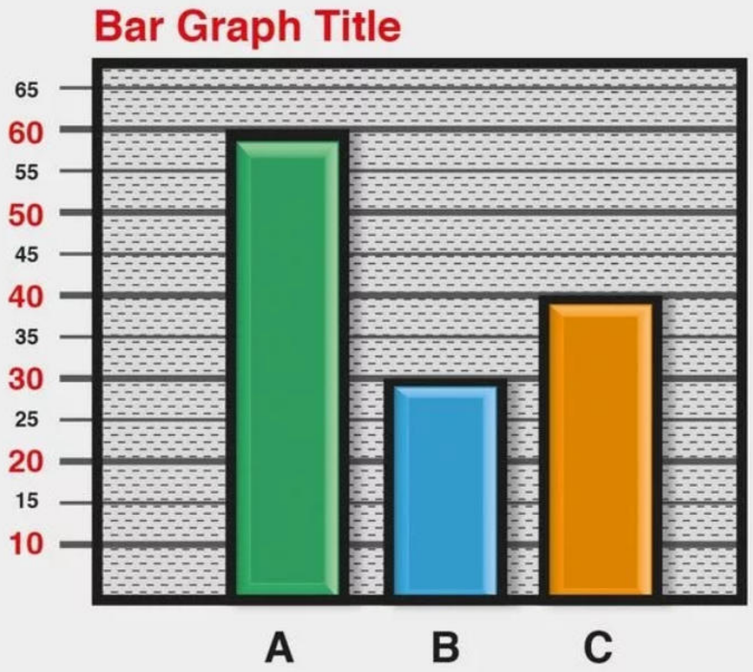


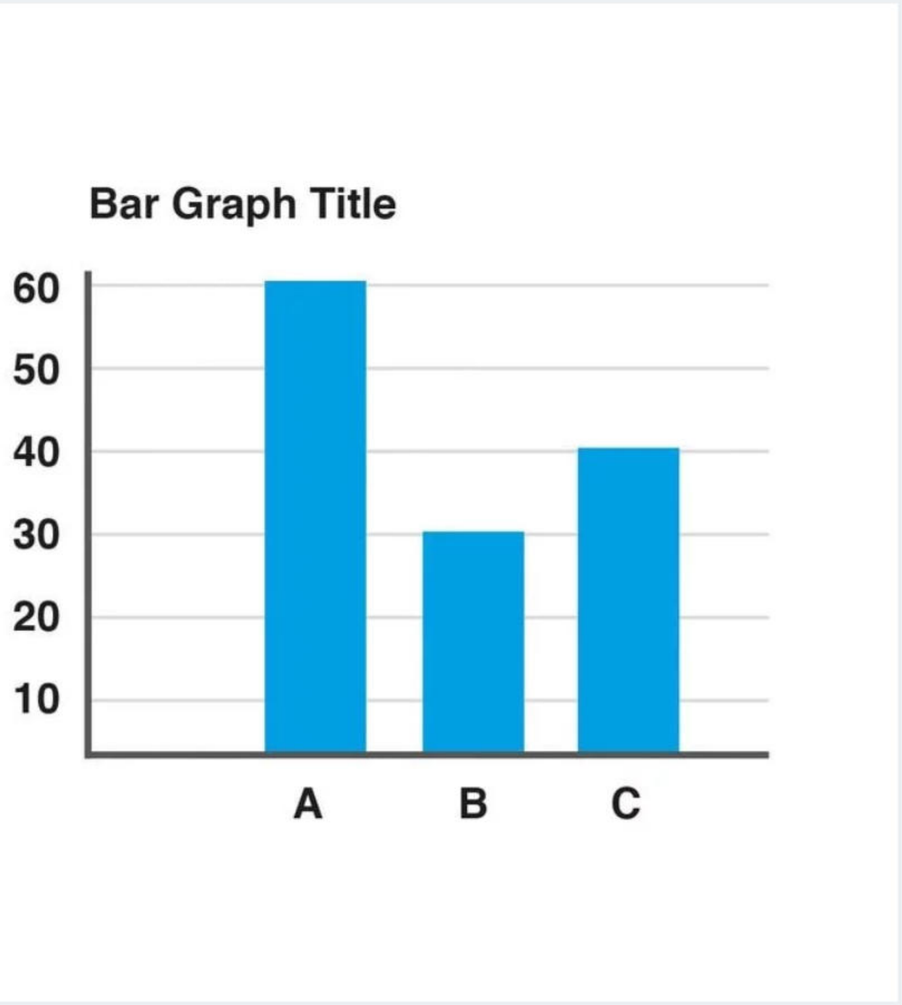
Source: Choropleth Map by Die Zeit (left) | Tile Grid Map by Cédric Scherer & Ansgar Wolsing (right)

VISUAL FORM

Present information in a logical, coherent way

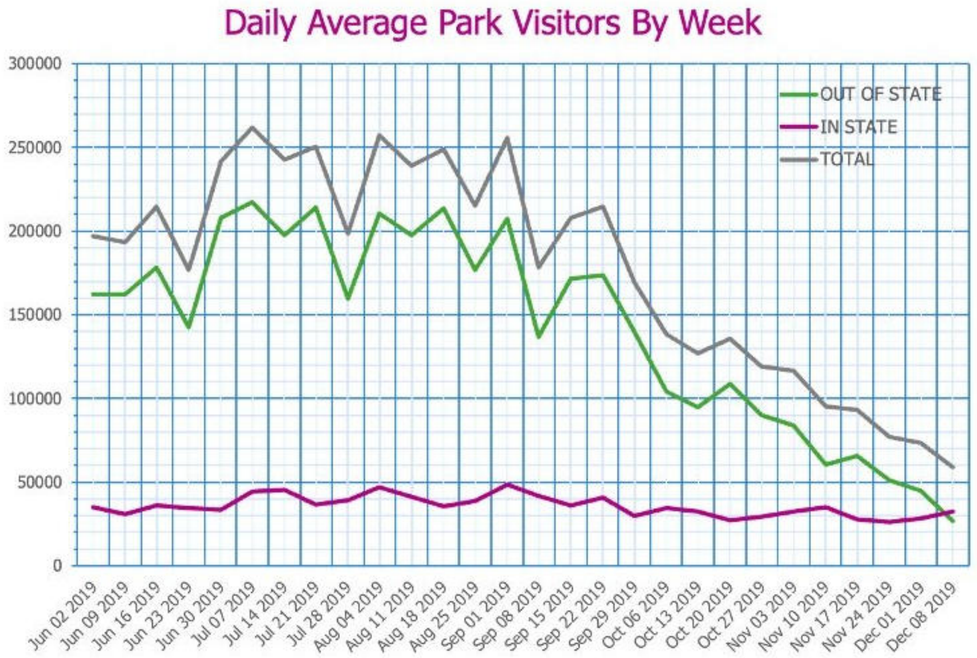
Declutter You Charts



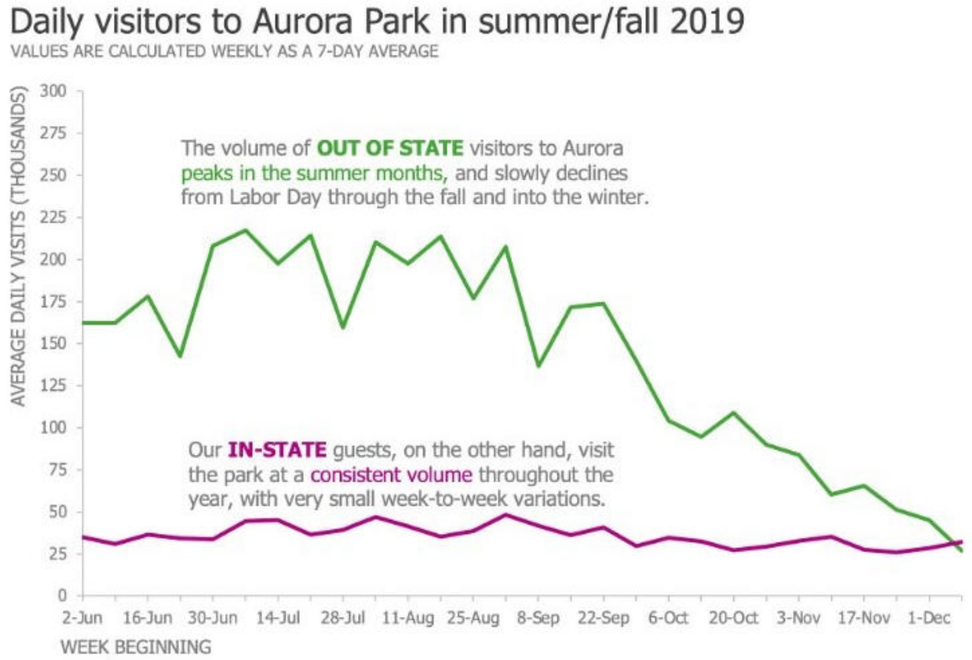


Source: canva.com

BEFORE: SPOOKY SKELETON

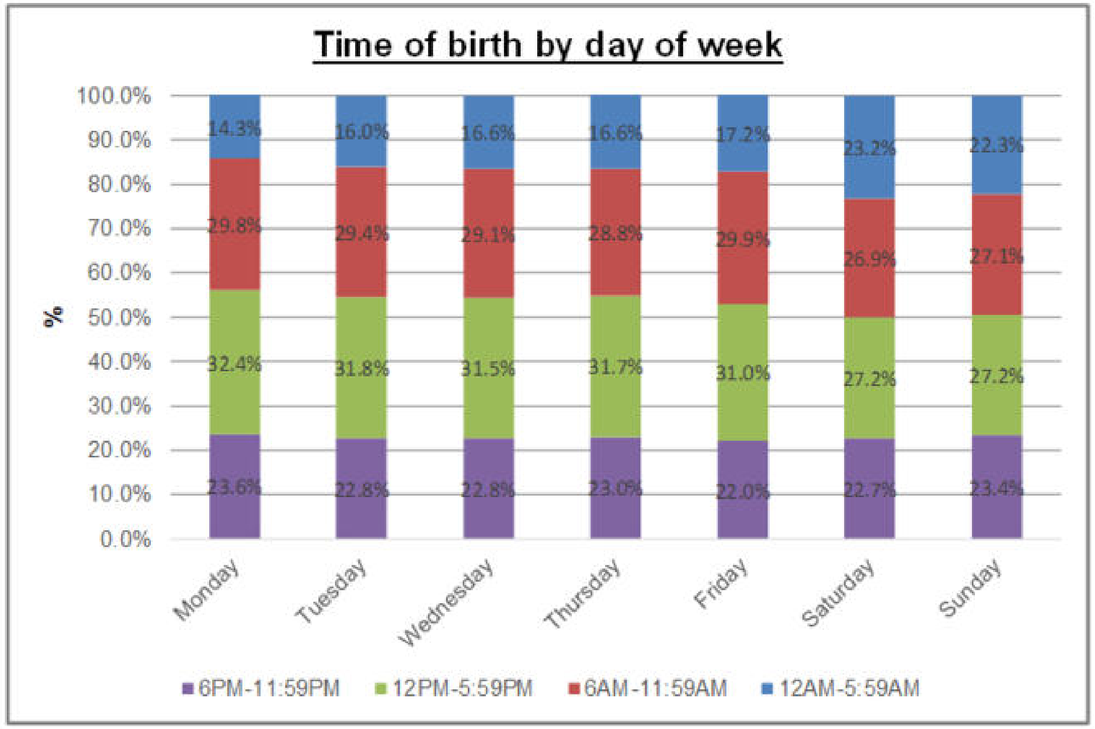


AFTER: GOOD BONES



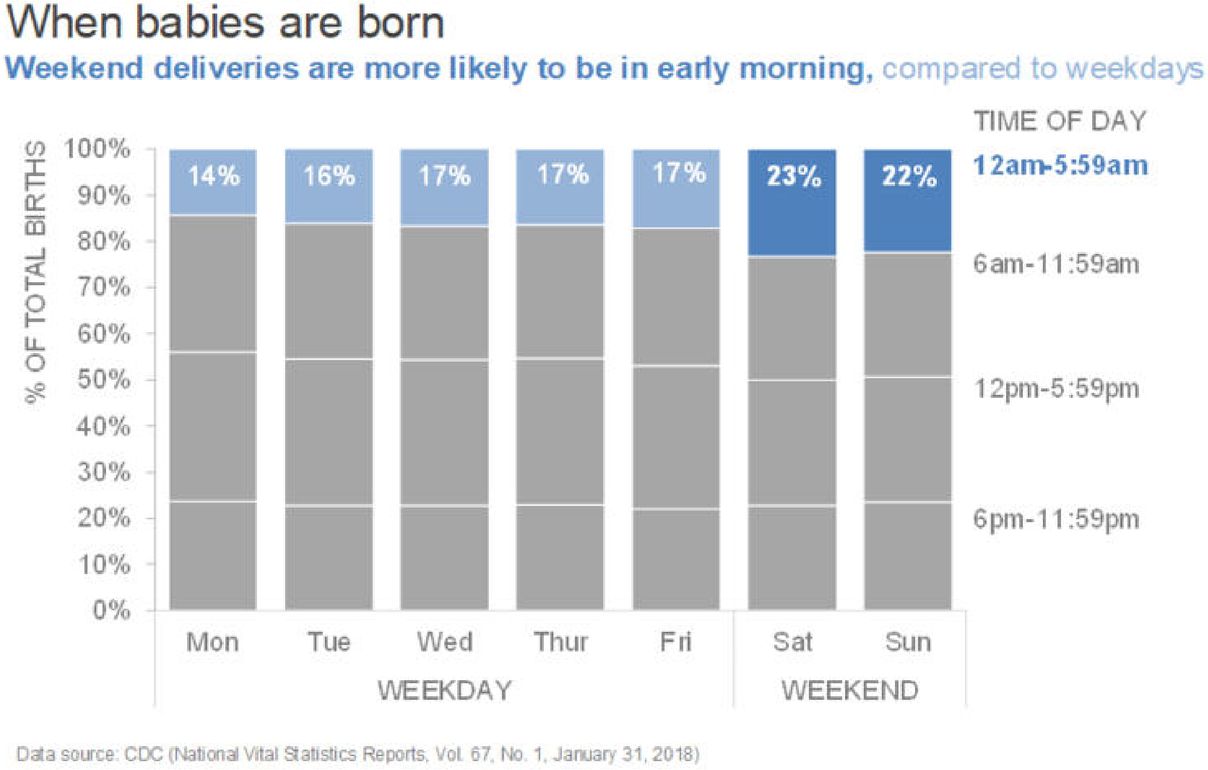
Source: "Your graph Skeleton Shouldn't Be Scray" by Mike Cisneros (Storytelling With Data)

BEFORE



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

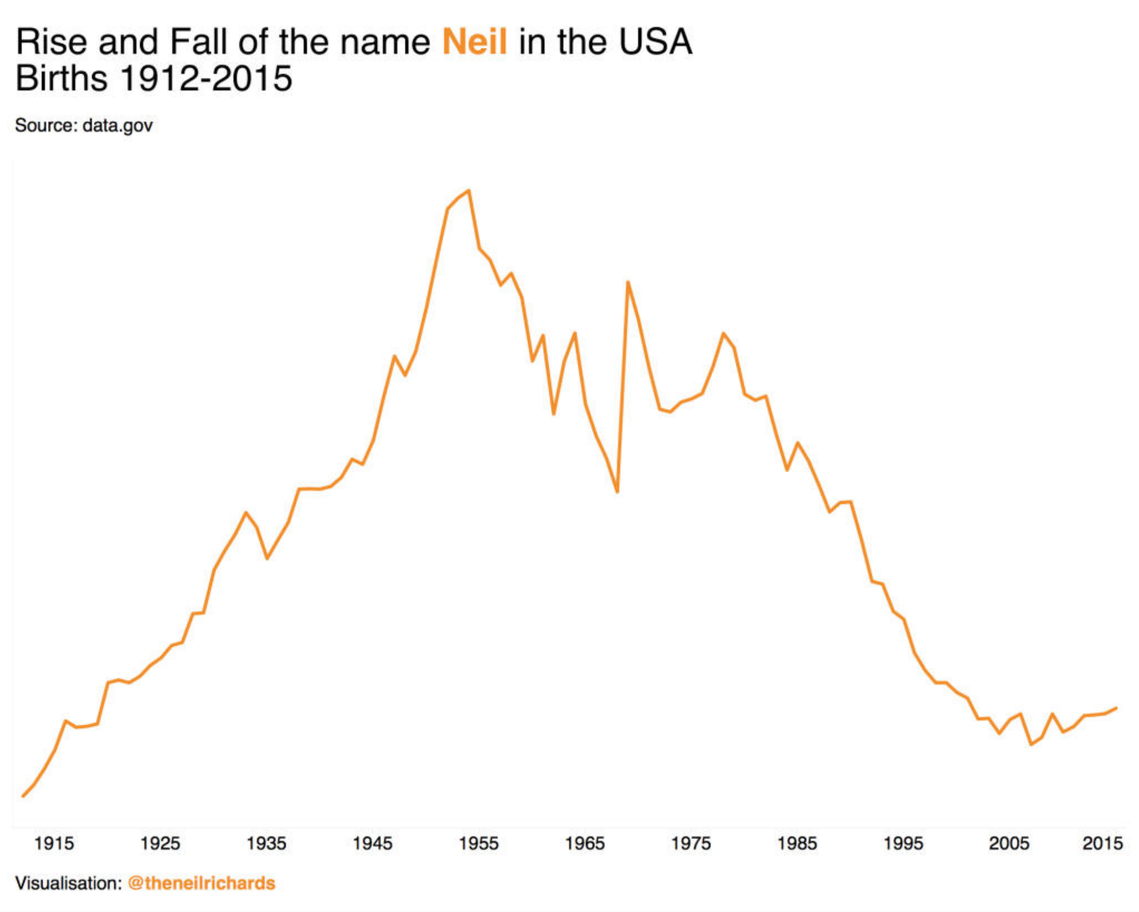
AFTER



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

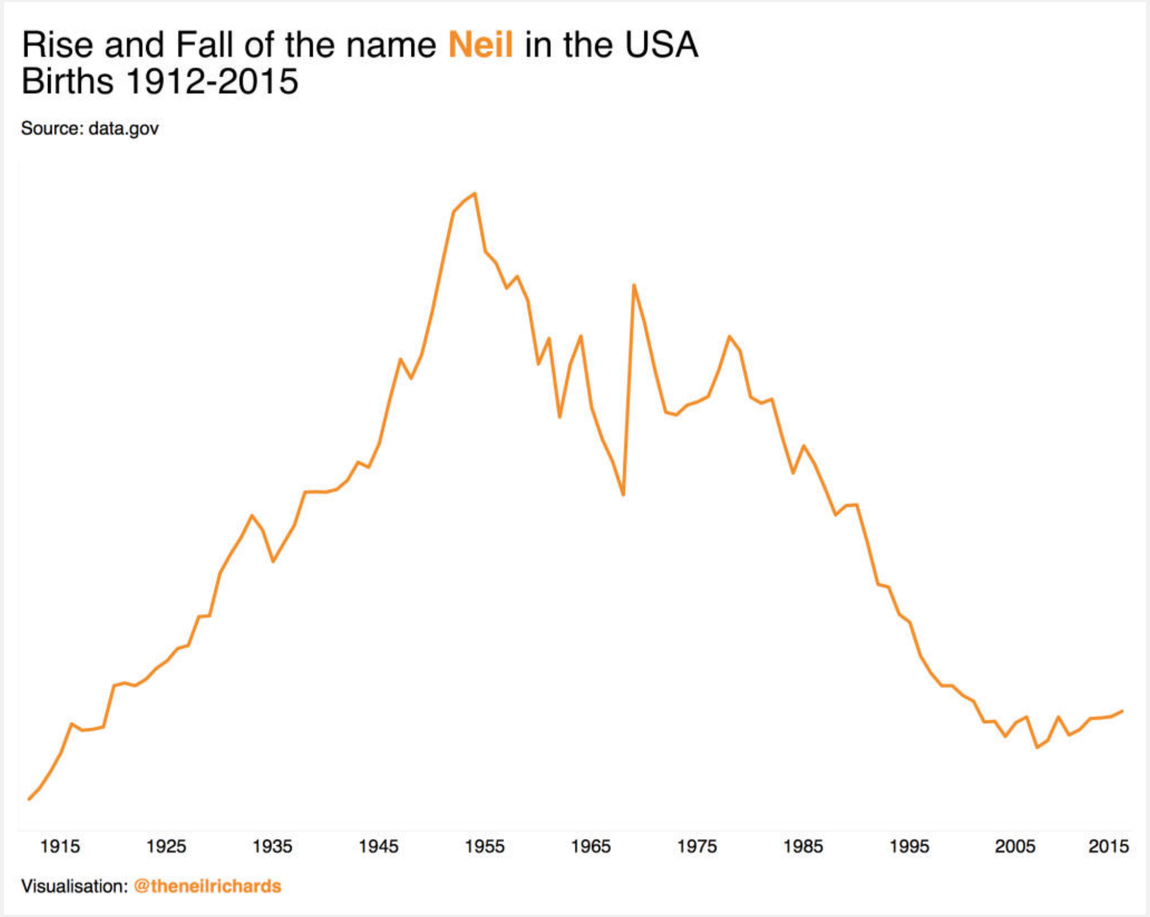
Source: "Declutter! (and Question Default Settings)" by Elizabeth Ricks (Storytelling With Data)

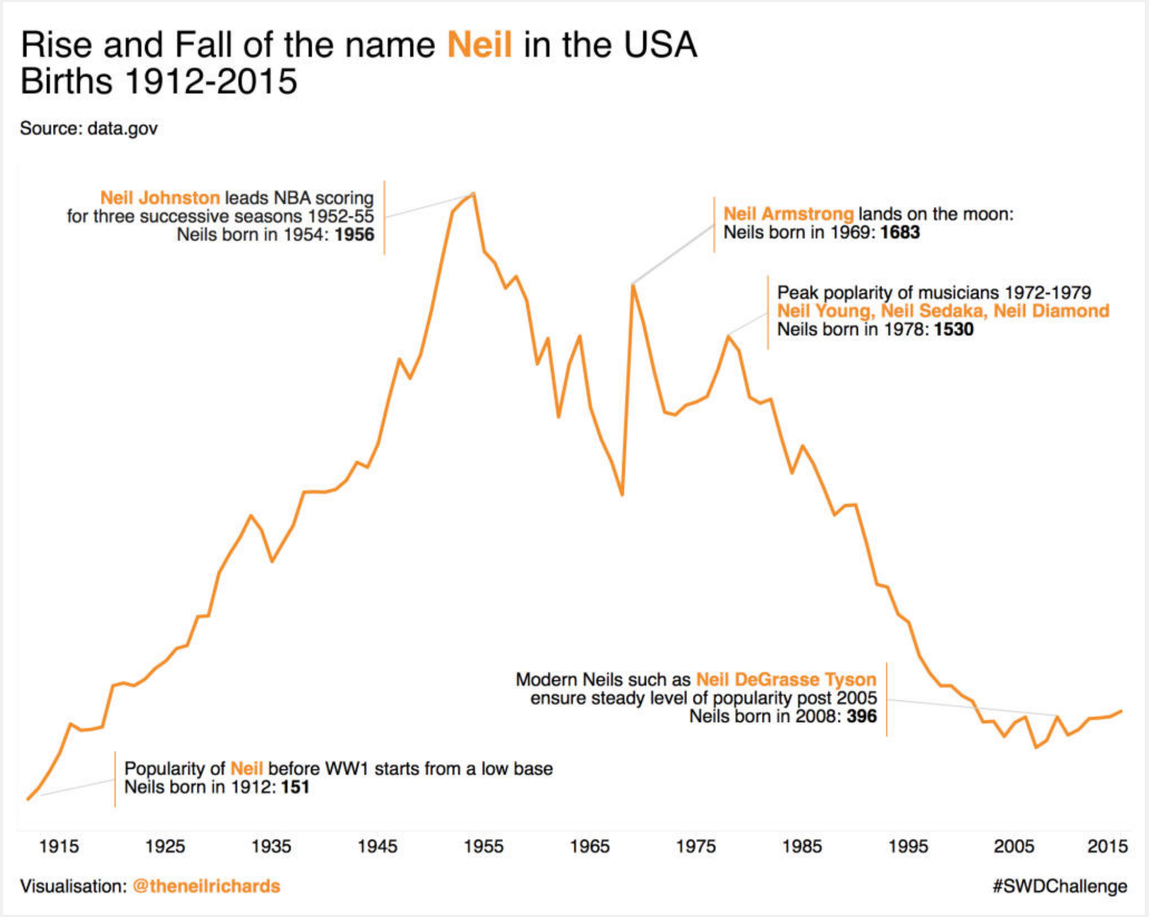
The Power of Annotations



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

The Power of Annotations

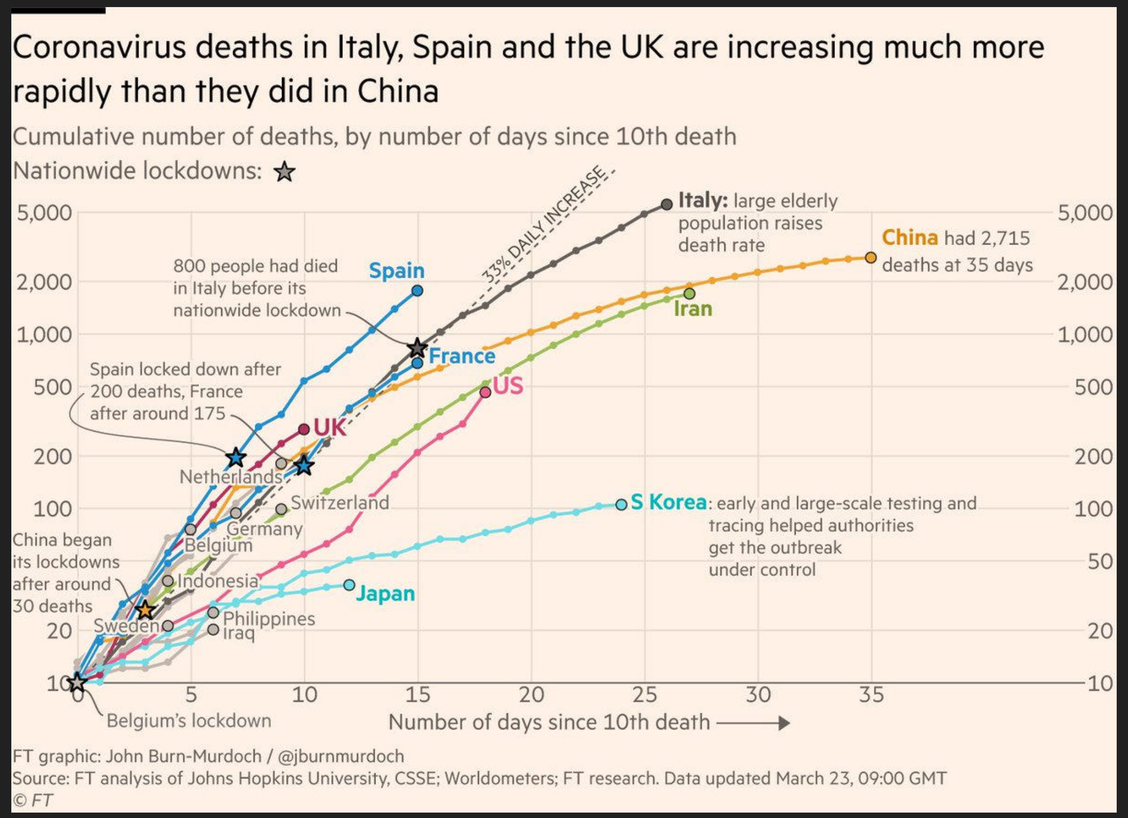


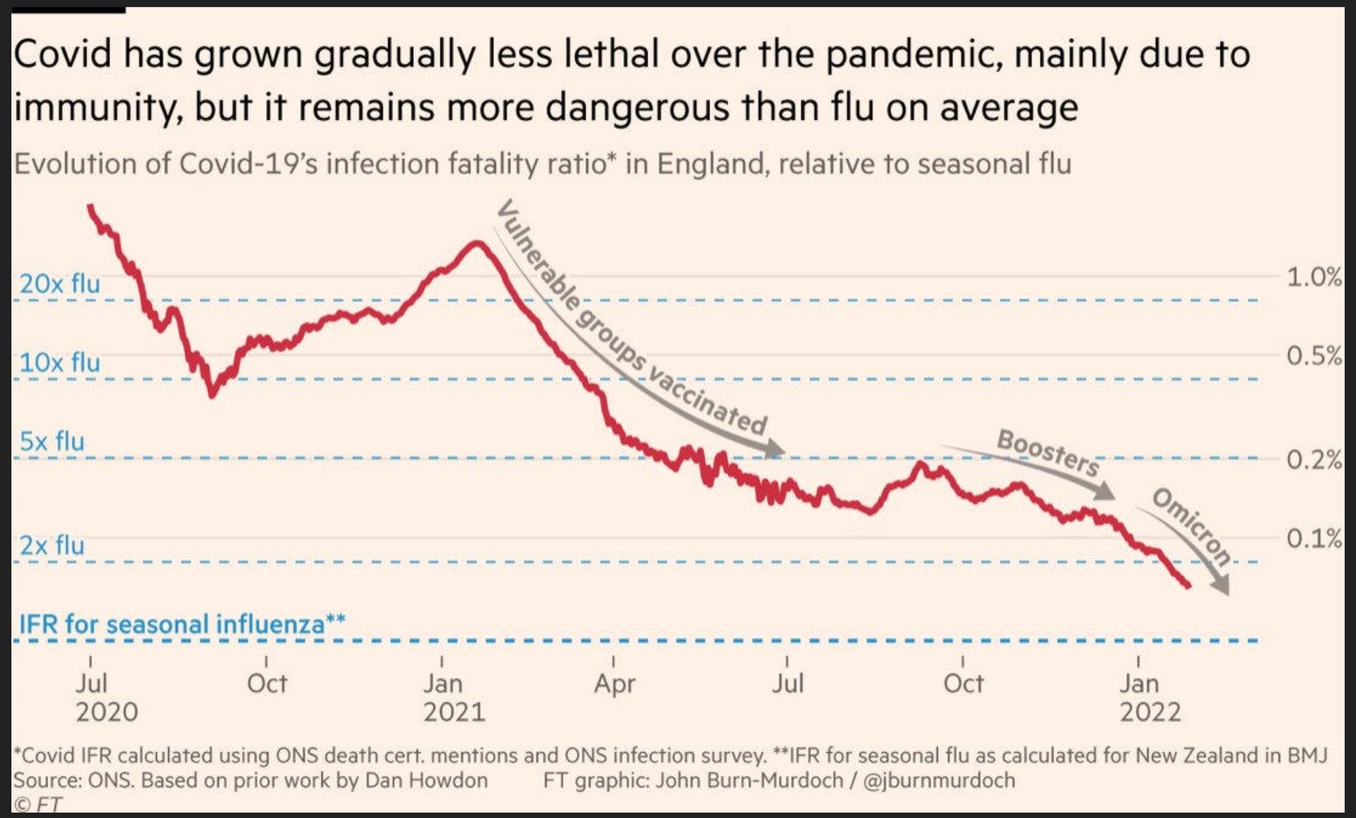


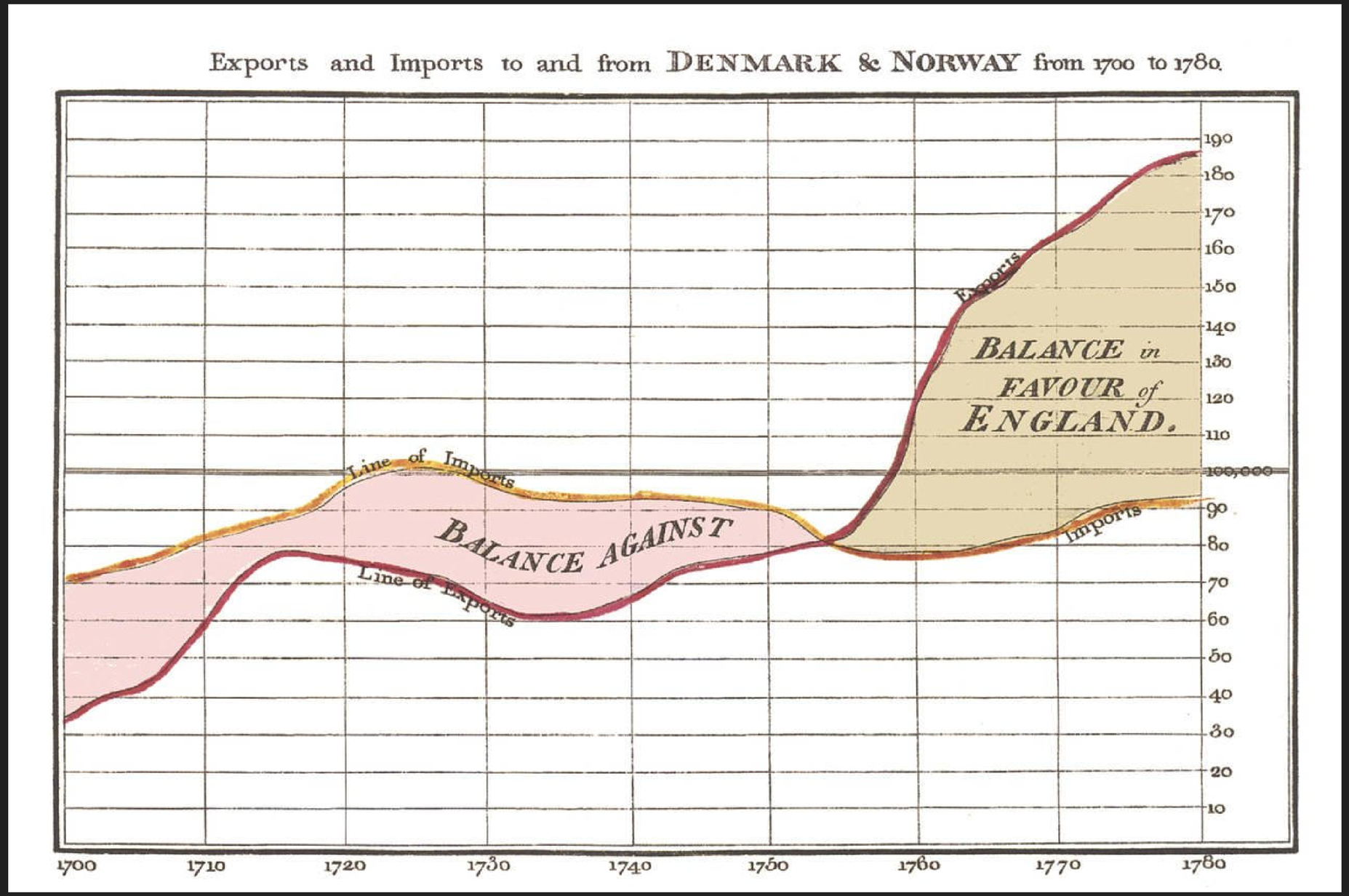
Source: "Is white space always your friend?" by Neil Richards

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

~ John Burn-Murdoch, Financial Times



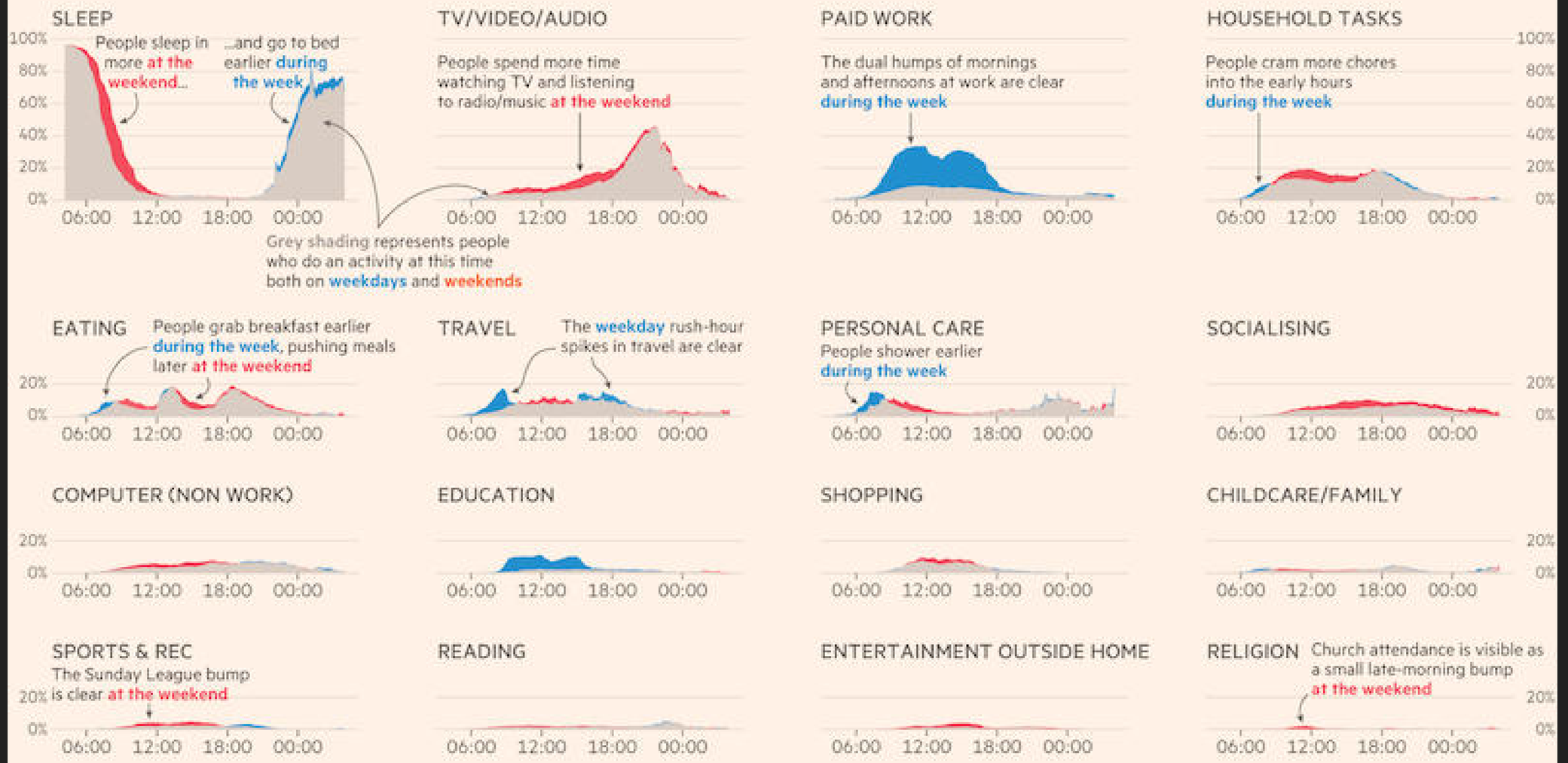




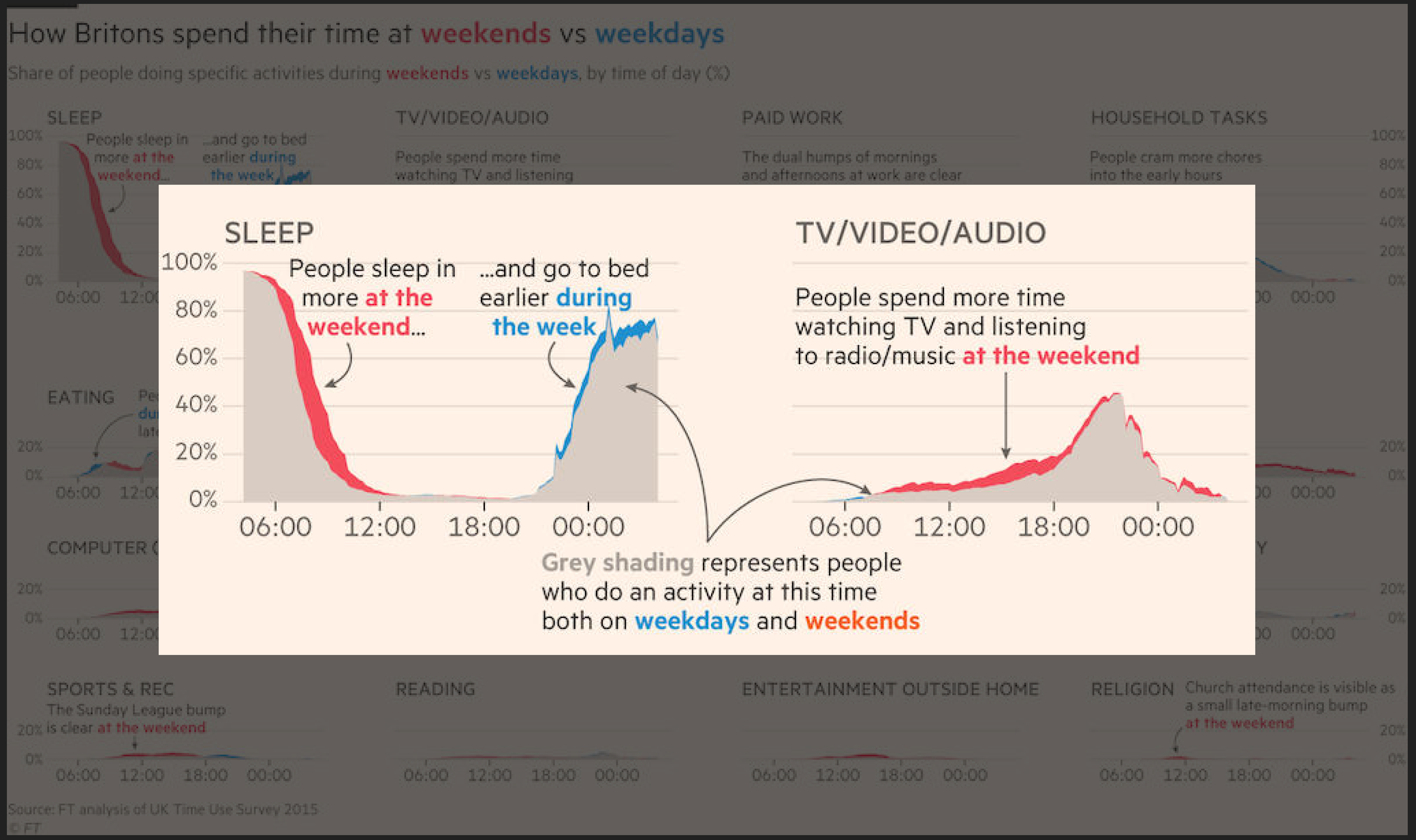
Annotated time-series chart by William Playfair from "The Commercial and Political Atlas and Statistical Breviary" (1786)

How Britons spend their time at weekends vs weekdays

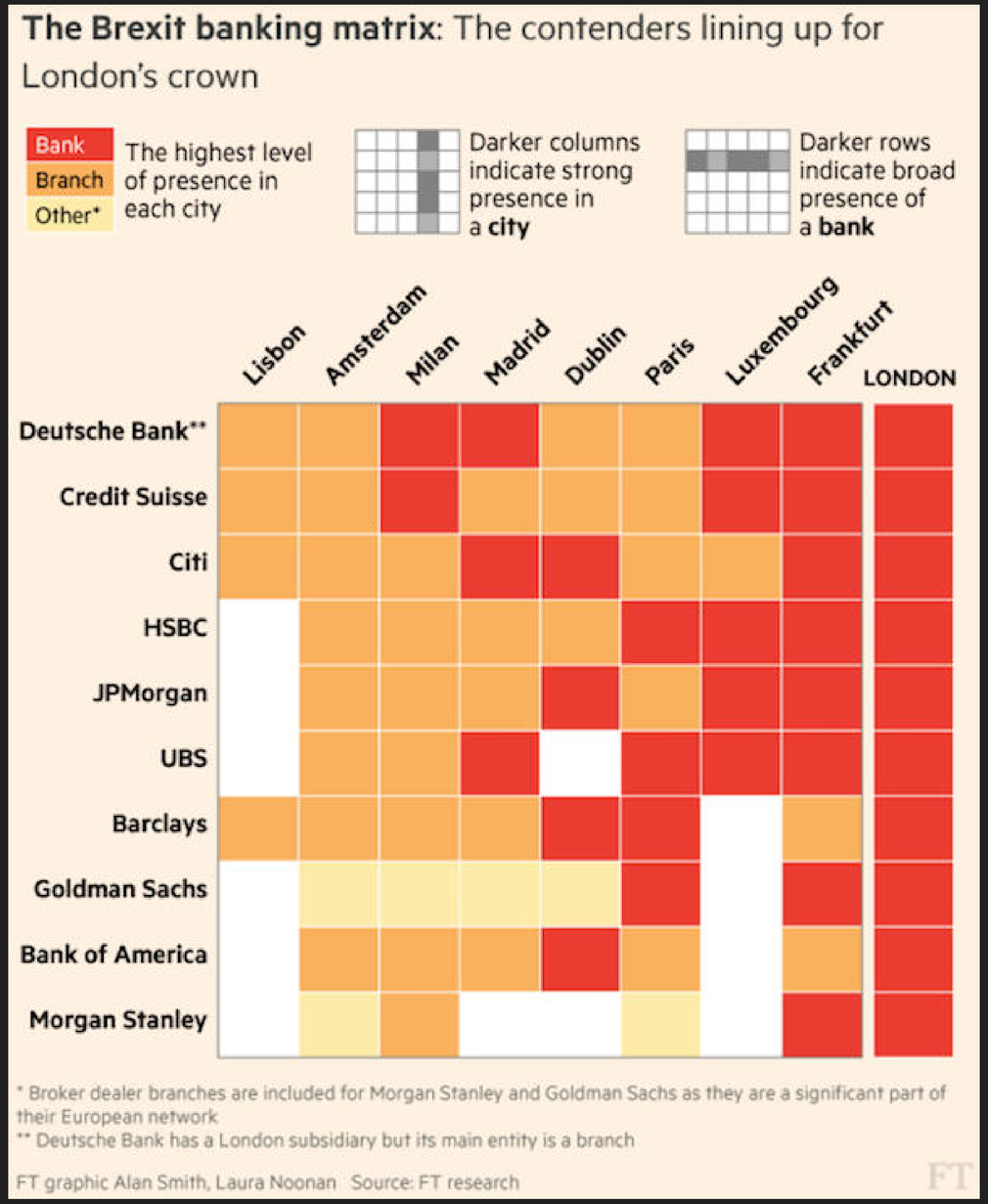
Share of people doing specific activities during weekends vs weekdays, by time of day (%)



Source: FT analysis of UK Time Use Survey 2015

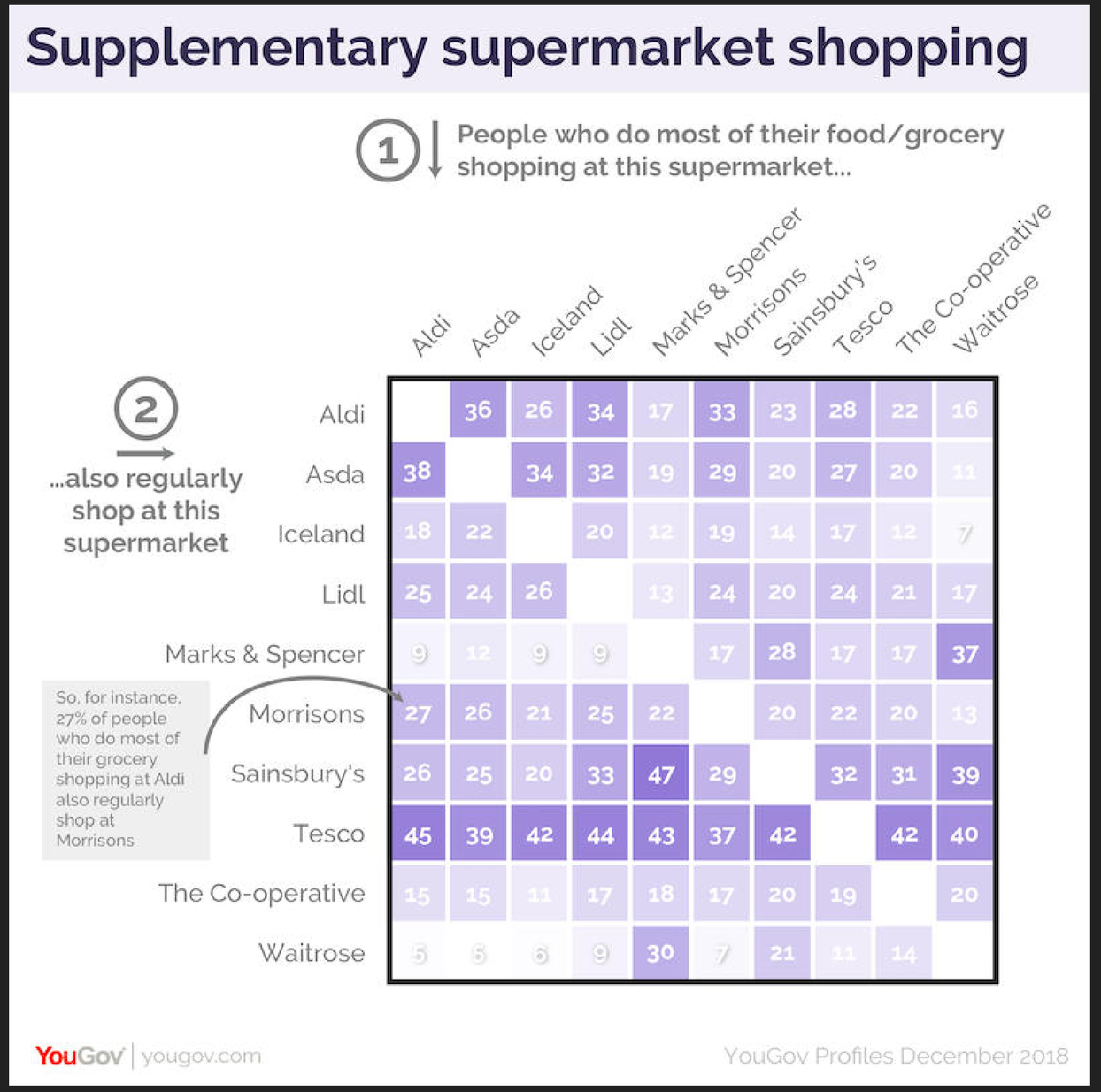


Source: "The Truth about Weekend Working" by John-Burn Murdoch (Financial Times)



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

Supplementary supermarket shopping



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

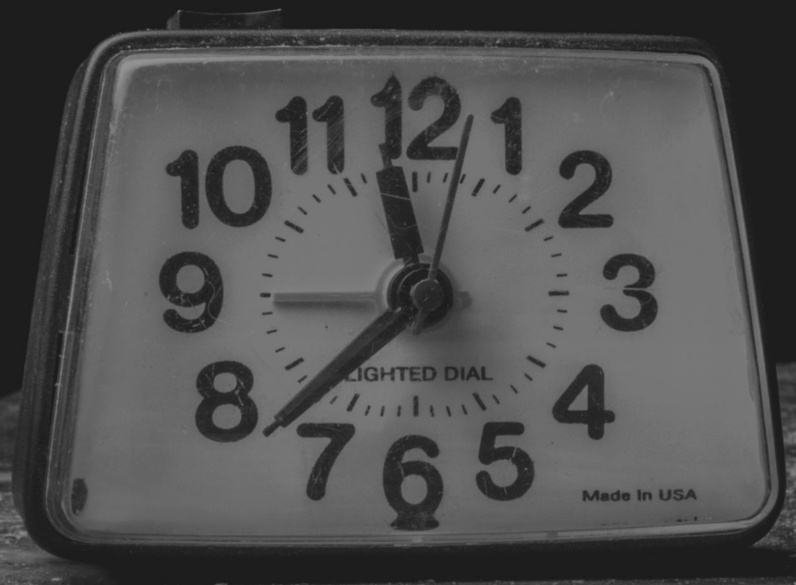

cedricscherer.com
@CedScherer
z3tt

I'll be waiting for you!

I'LL BE WAITING FOR YOU!

I'LL BE WAITING FOR YOU!

I'll be waiting for you!



Which would you prefer to wake up to?

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 **monospaced 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!**

Visualize Hierarchy

I am important!

I am important, too.

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

Hey, also look at me!

Visualize Hierarchy

I am important!

I am important, too.

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

Hey, also look at me!

Visualize Hierarchy

I am important!

I am important, too.

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

Hey, also look at me!

Keep it Simple

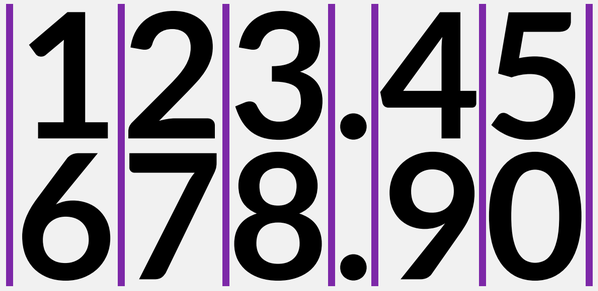
Using a lot of fonts
and **different** sizes
makes your design look
cluttered
overcomplicated
AND JUST **NOT VERY NICE.**

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*.

Proportional Numbers

Tabular Numbers



The 1Il Test

1Il Calibri

1Il Cabin

1Il Open Sans

1Il Monda

1Il Roboto

1Il Chivo

1Il Lato

1Il Fira Sans

1Il Oswald

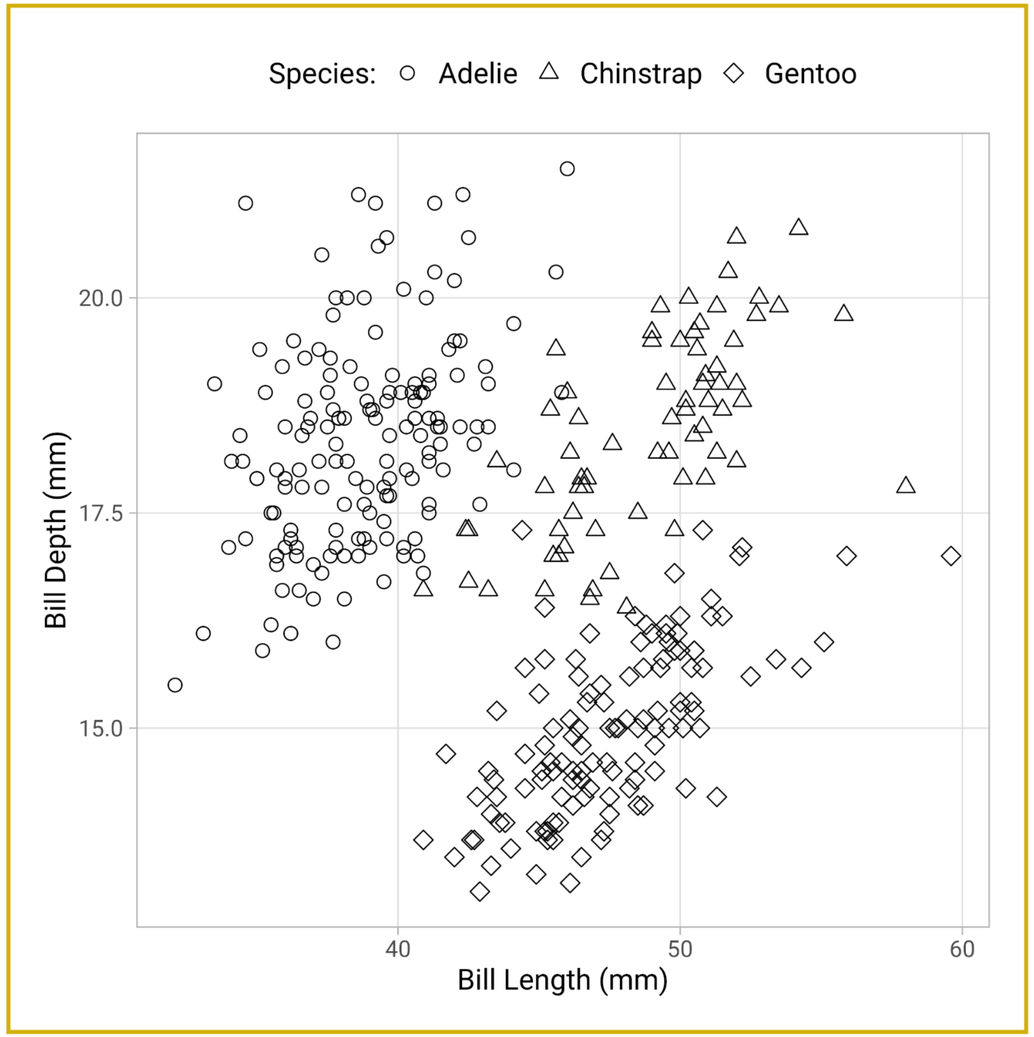
1Il Noto Sans

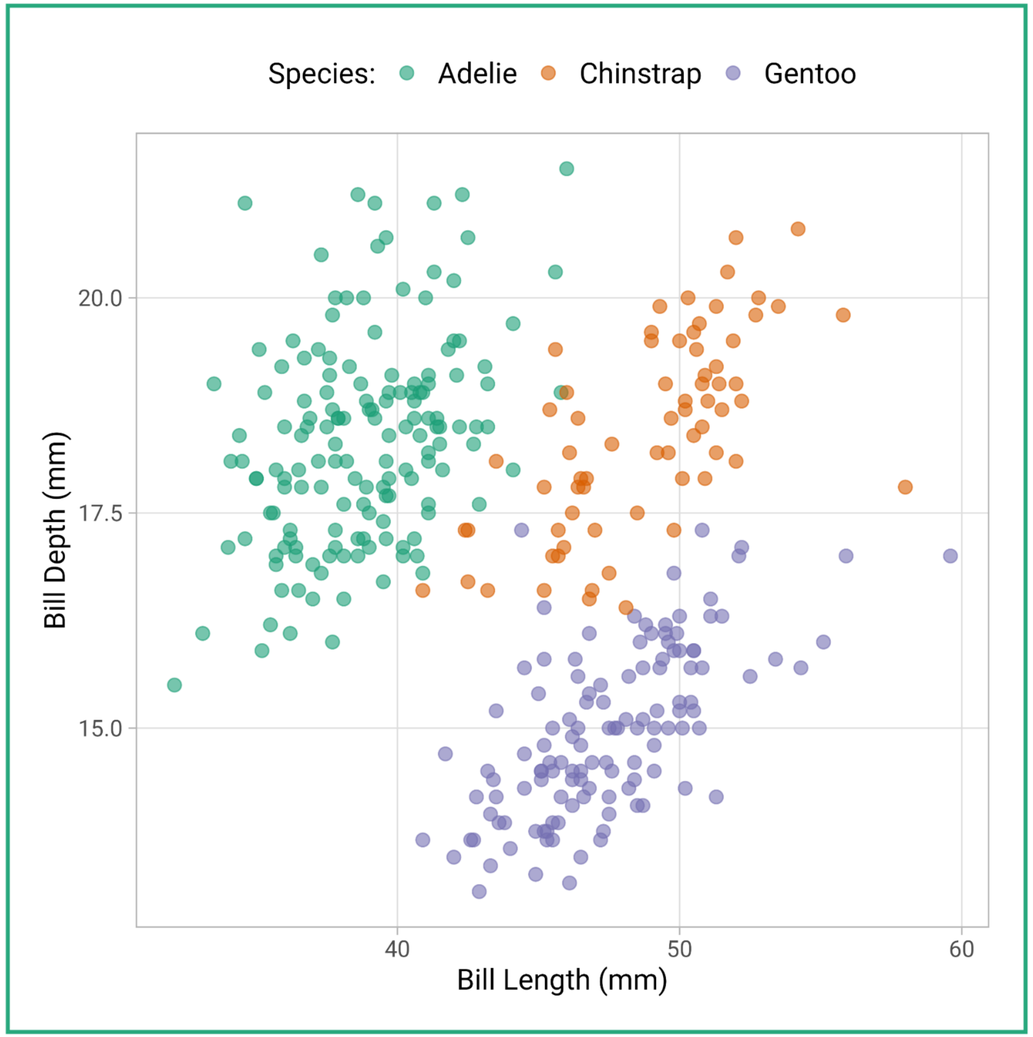
1Il Cabinet Grotesk

1Il Amulya

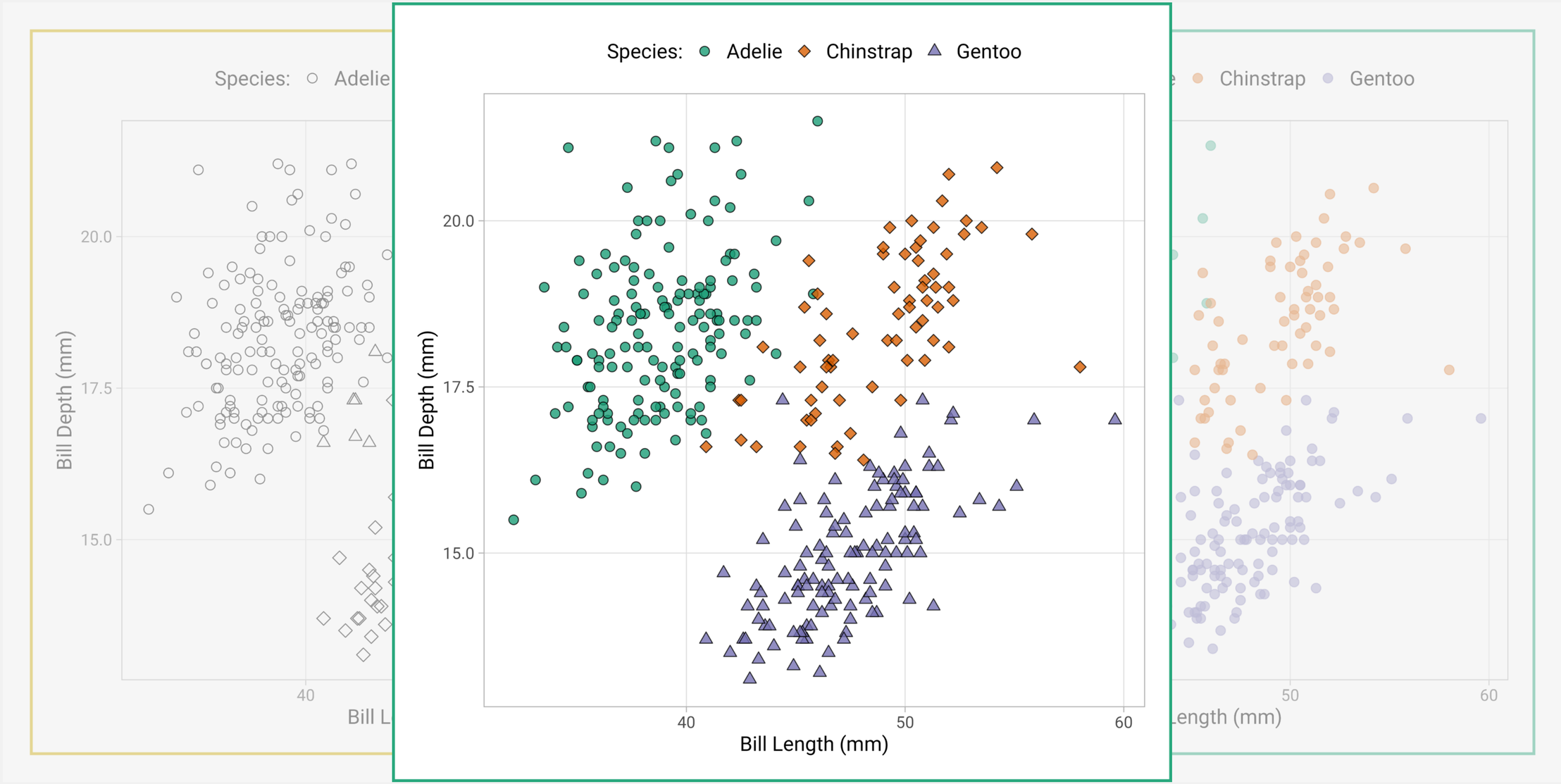

COLORS
and Pitfalls

Channels to Encode Information

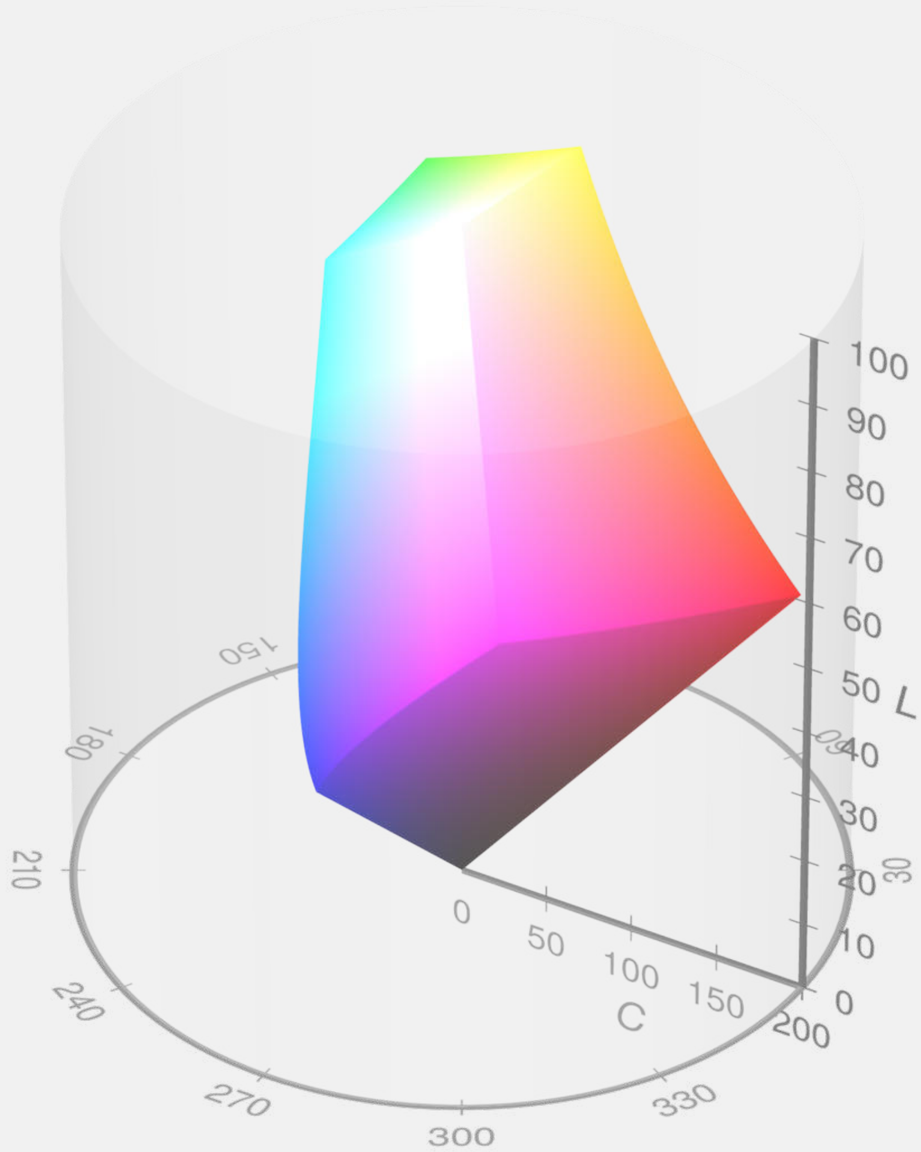




Use of Double Encoding



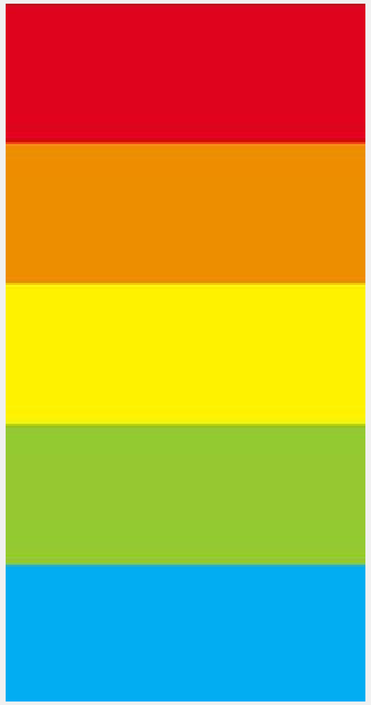
The HCL Color Space



Michael Horvath & Christoph Lipka

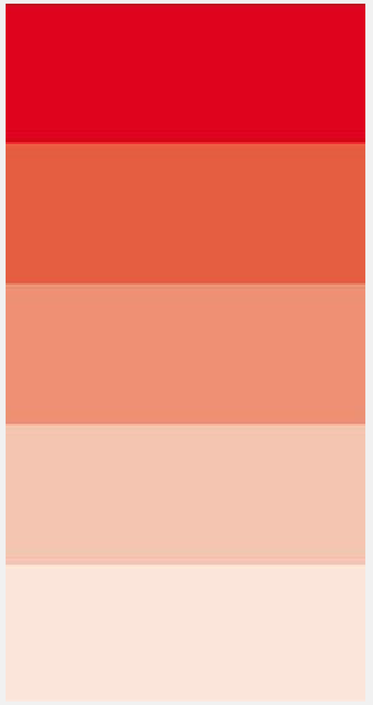
cedricscherer.com @CedScherer z3tt

Hue
color family



Chroma
colorfulness

Luminance
lightness / brightness



Modified from uxplanet.org

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

Color Palette Choice

Sequential

Palette

Desaturated

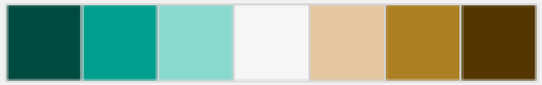
Use to encode
numerical information with order

use highest contrast for most important information

either single- or multi-hue

Diverging

Palette



Desaturated

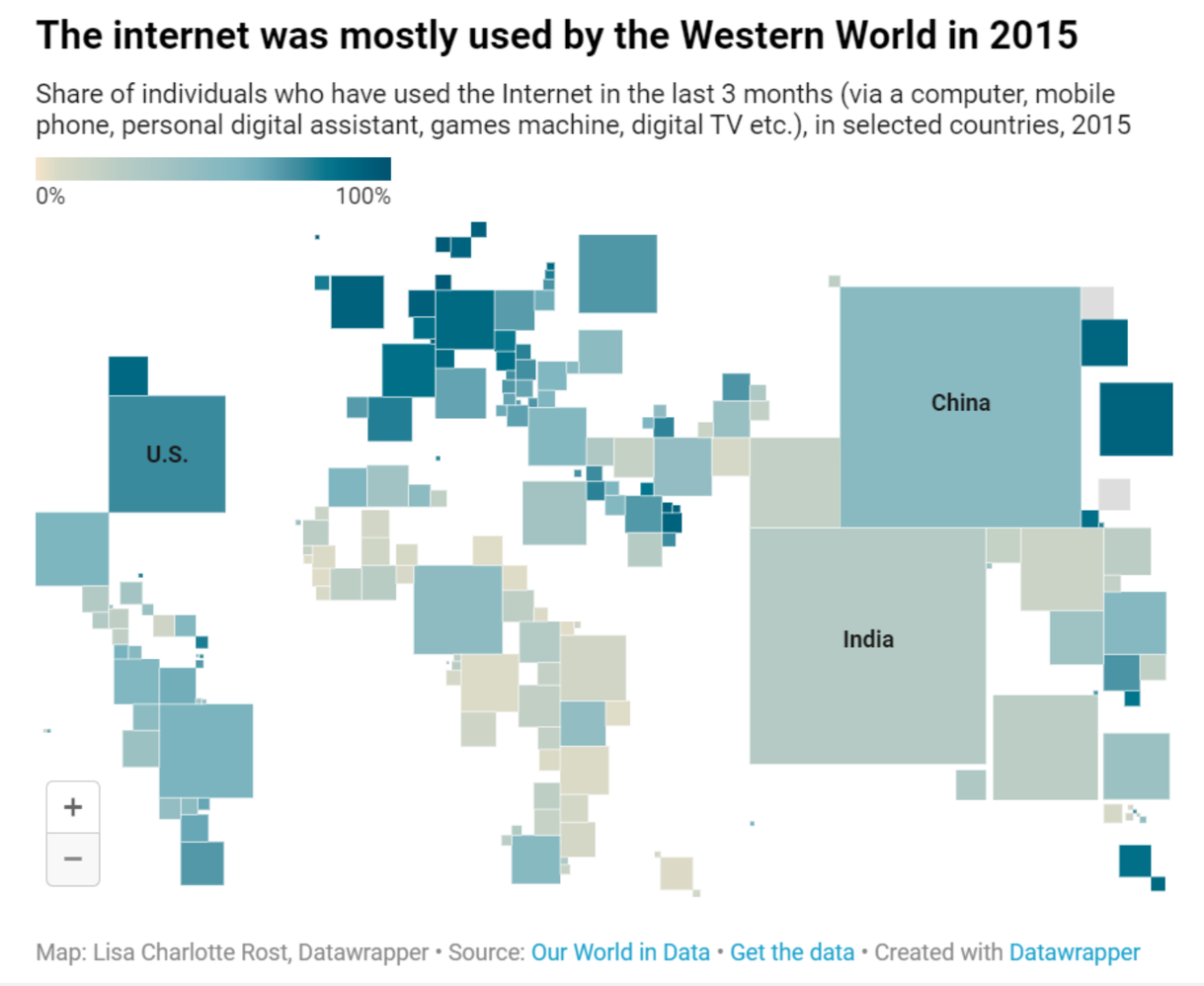
Use to encode
numerical information with critical midpoint

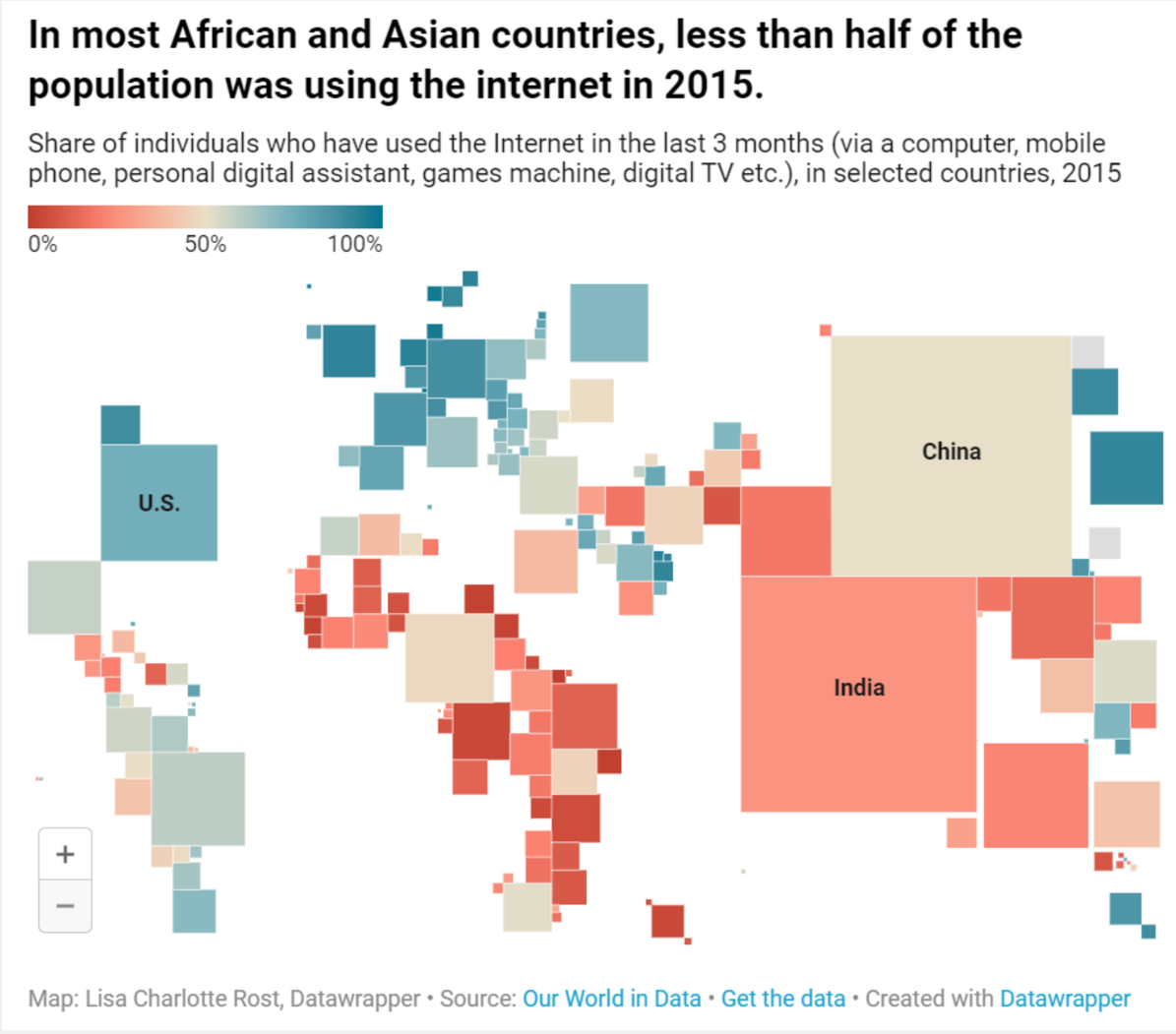
ensure a meaningful midpoint value and use balanced extremes

combination of two sequential palettes

Modified from the {colorspace} R package vignette

Sequential versus Diverging Palettes





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

Color Palette Choice

Sequential

Palette

Desaturated

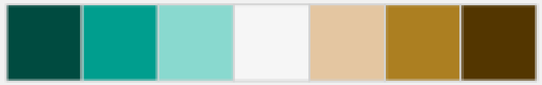
Use to encode
numerical information with order

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

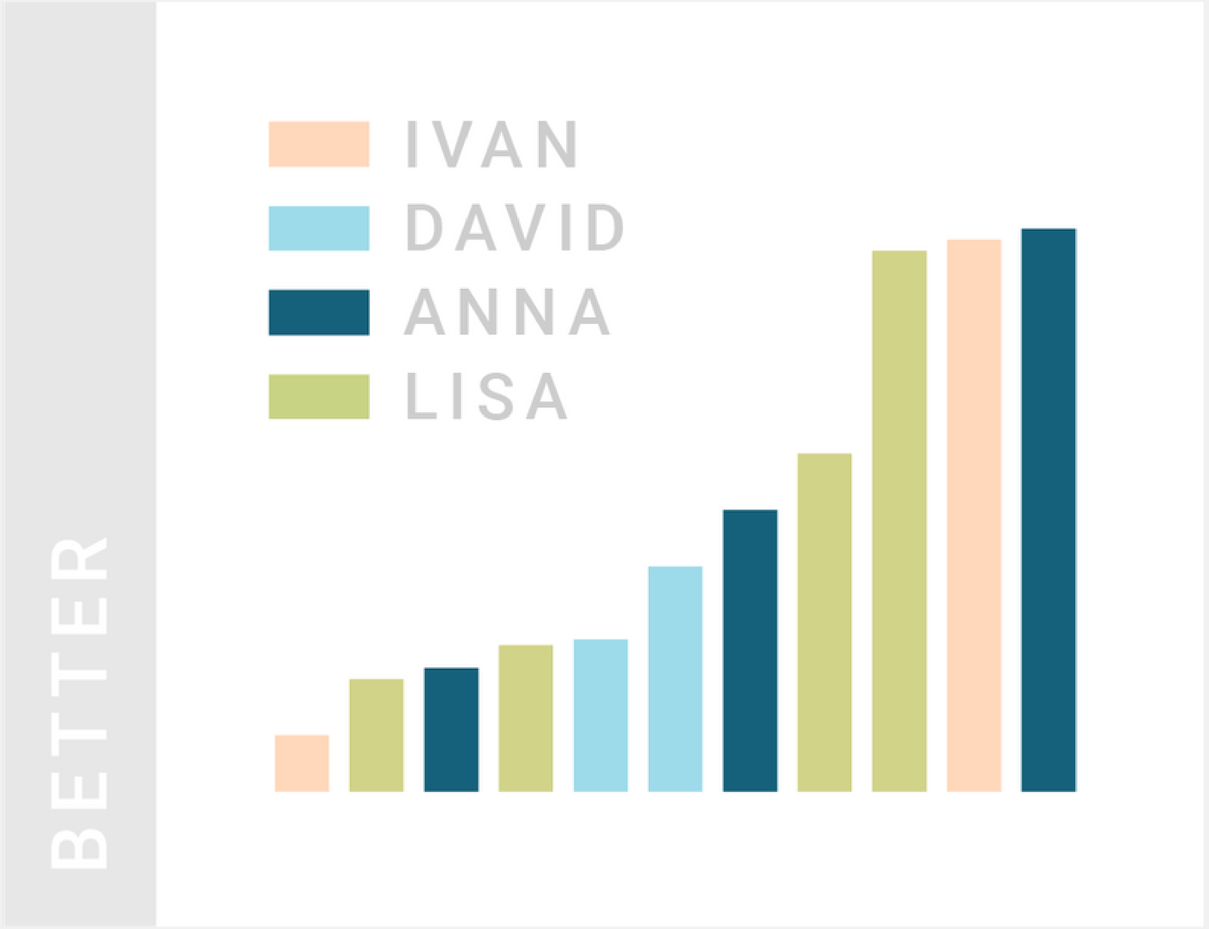
Modified from the {colorspace} R package vignette

cedricscherer.com @CedScherer z3tt

Sequential versus Qualitative Palettes

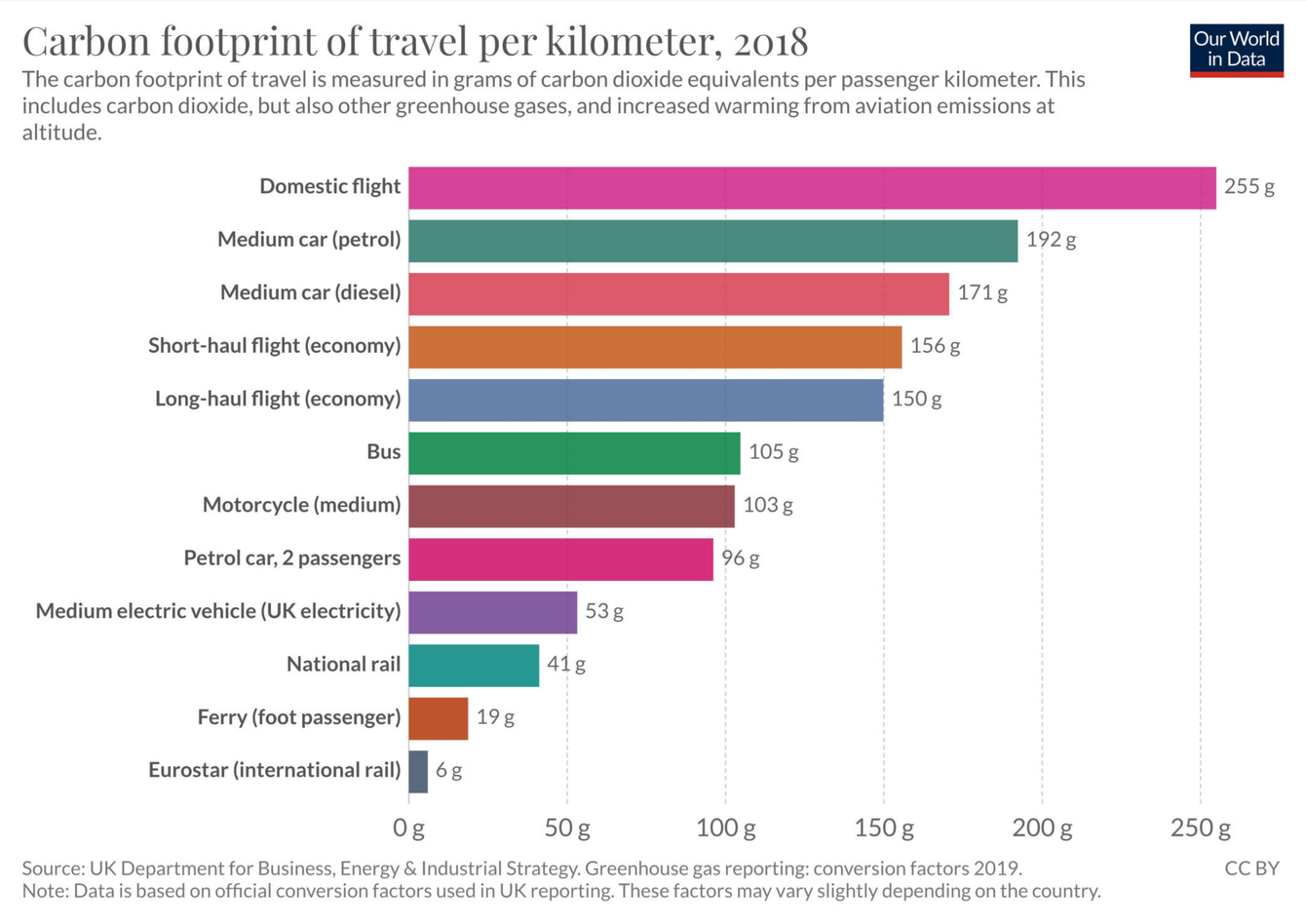
NOT IDEAL

BETTER



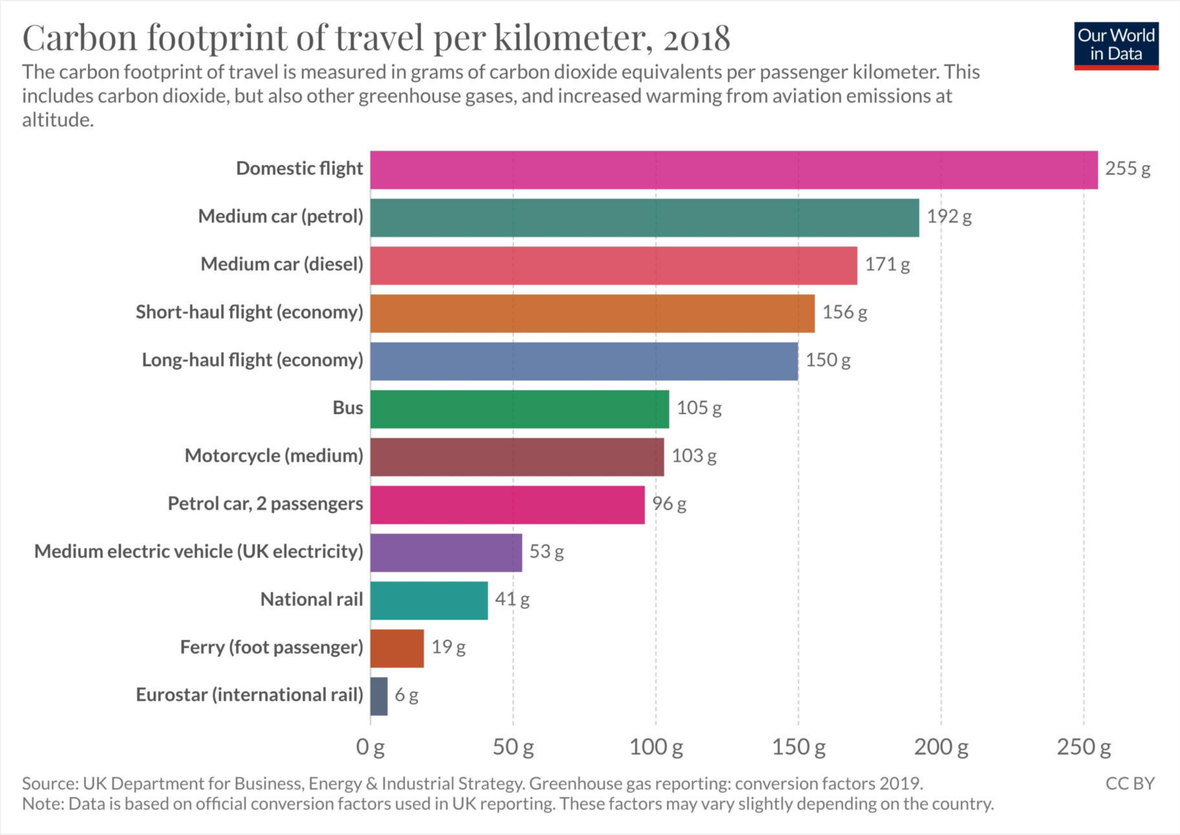
Source: "When to use sequential and when to use diverging color scales" by Lisa C. Muth / DataWrapper

Use color wisely (and correctly)

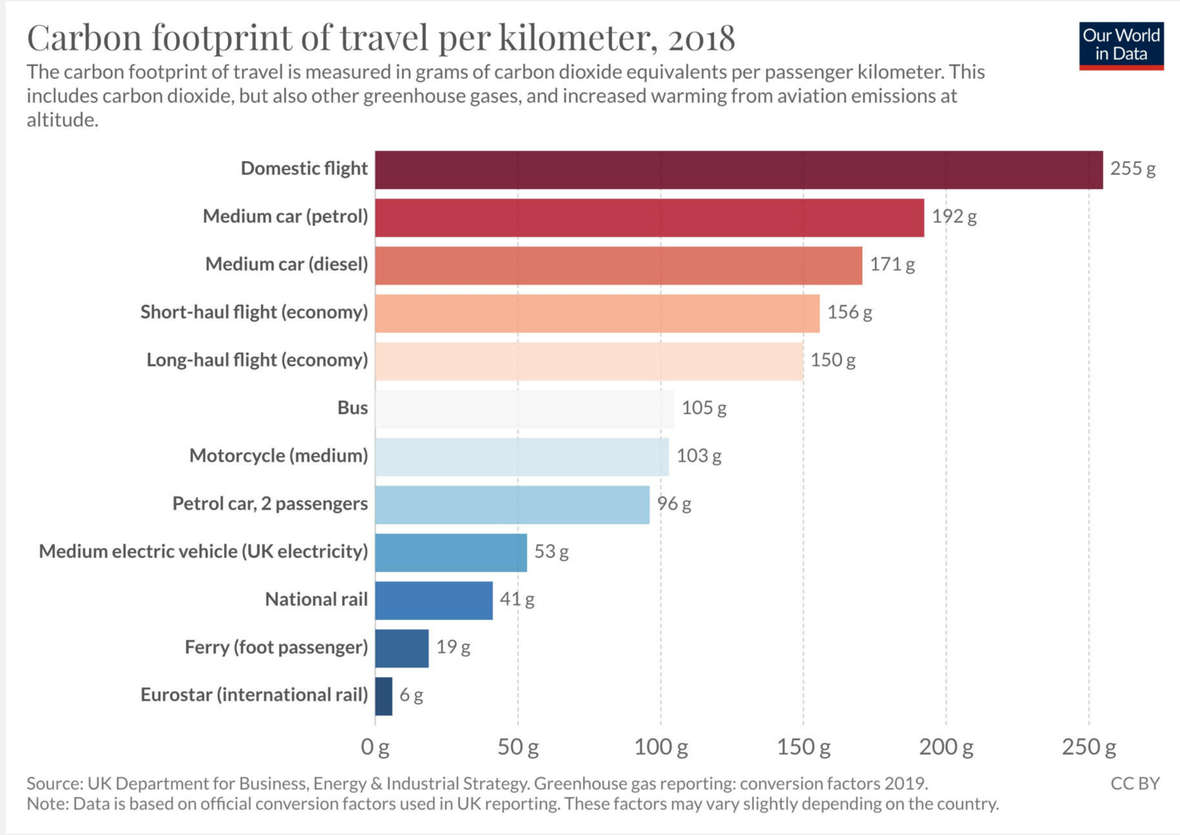


Original graphic with a random categorical palette

Use color wisely (and correctly)

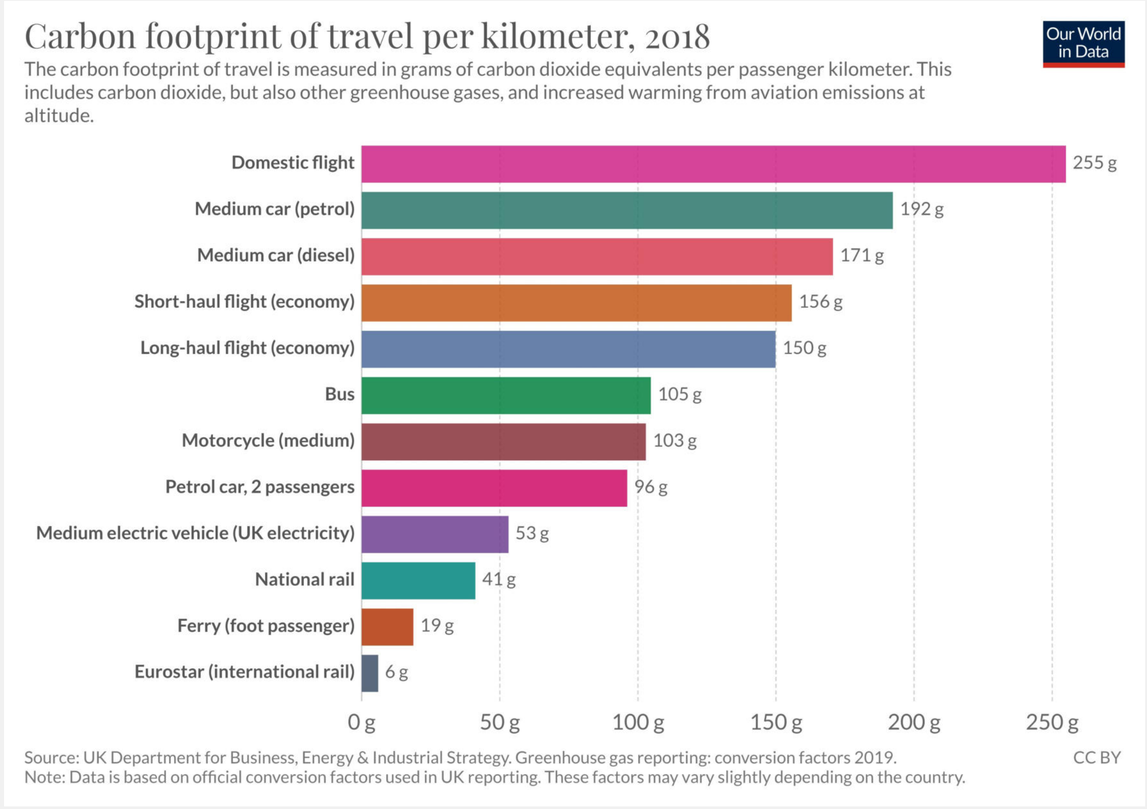


Original graphic with a random categorical palette

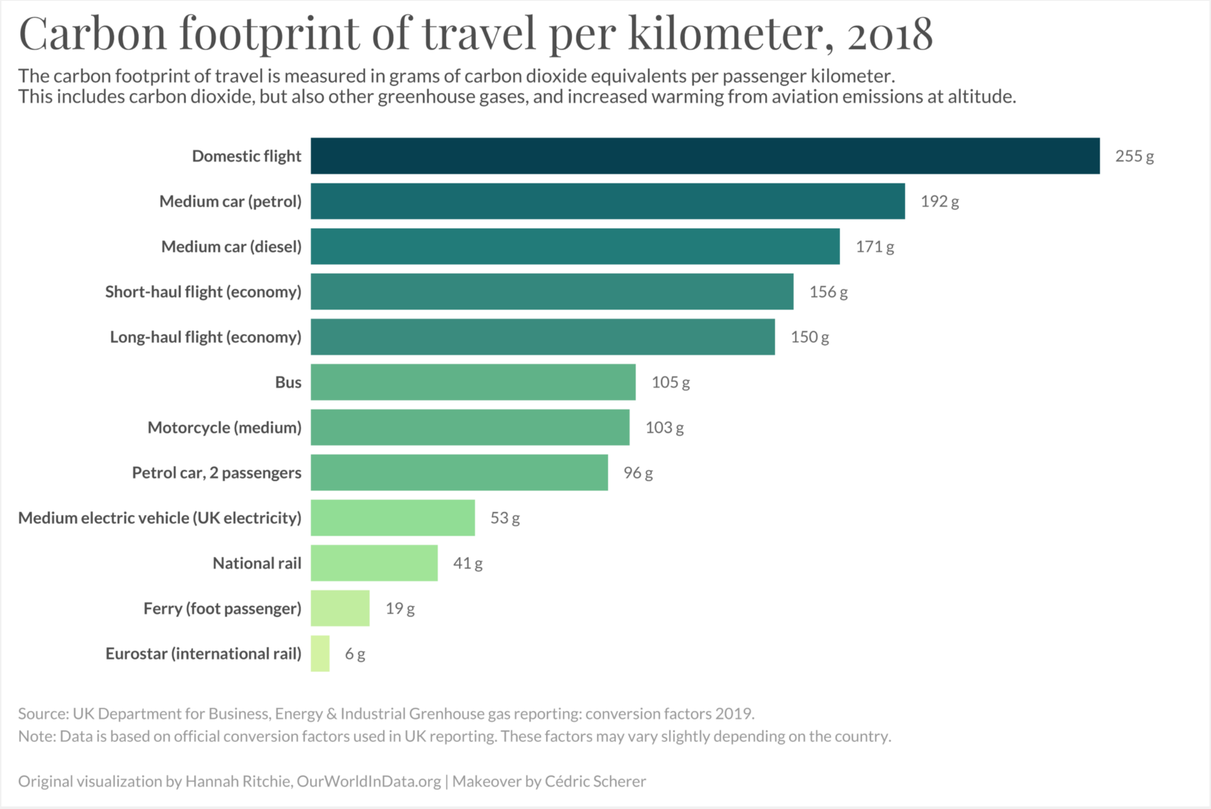


Reworked graphic using a diverging palette

Use color wisely (and correctly)



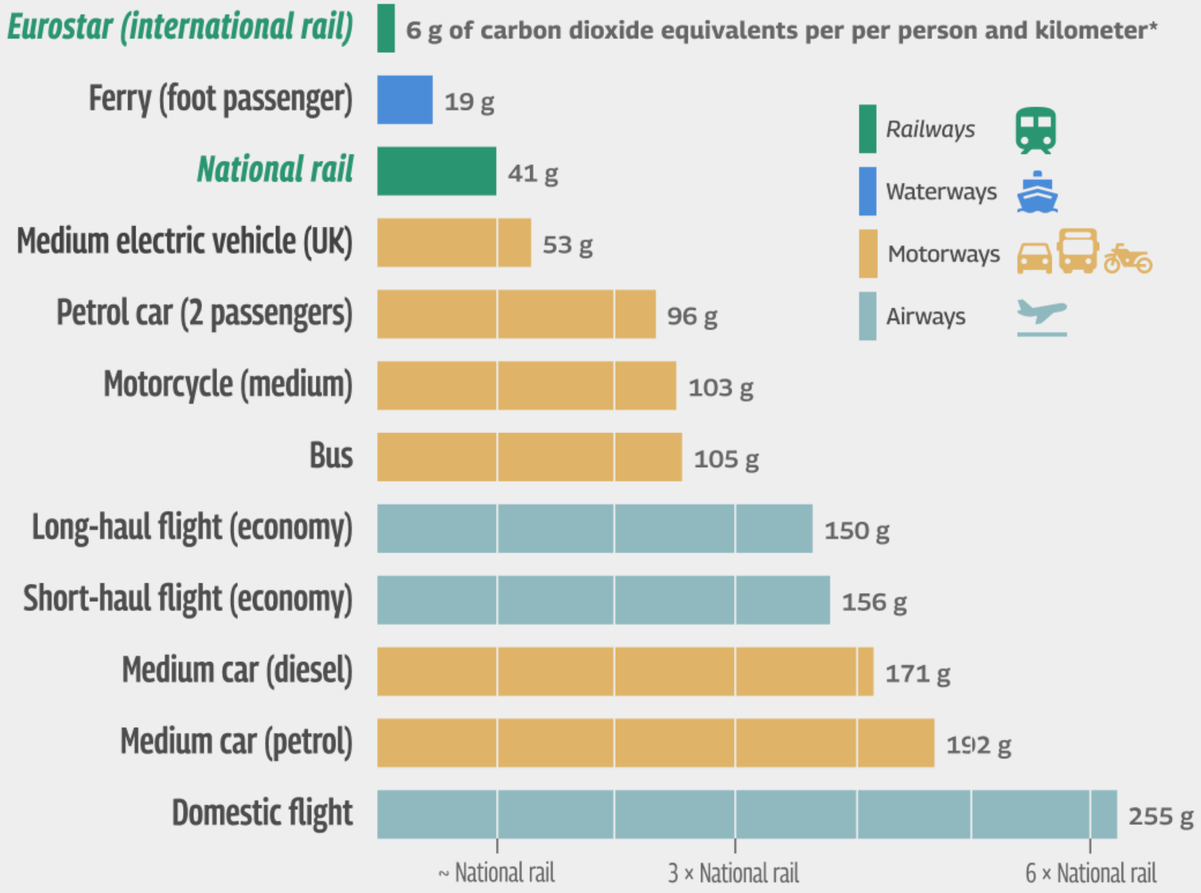
Original graphic with a random categorical palette



Makeover using a continuous palette

Reduce your Carbon Footprint: *Take the Train*

It's too far to walk or go by bike? Trains are nearly always the winning option over other moderate-to-long distance transportation modes. Taking a train instead of a car for medium-length distances would cut your emissions by ~80% and by ~84% if the train ride replaces a domestic flight.

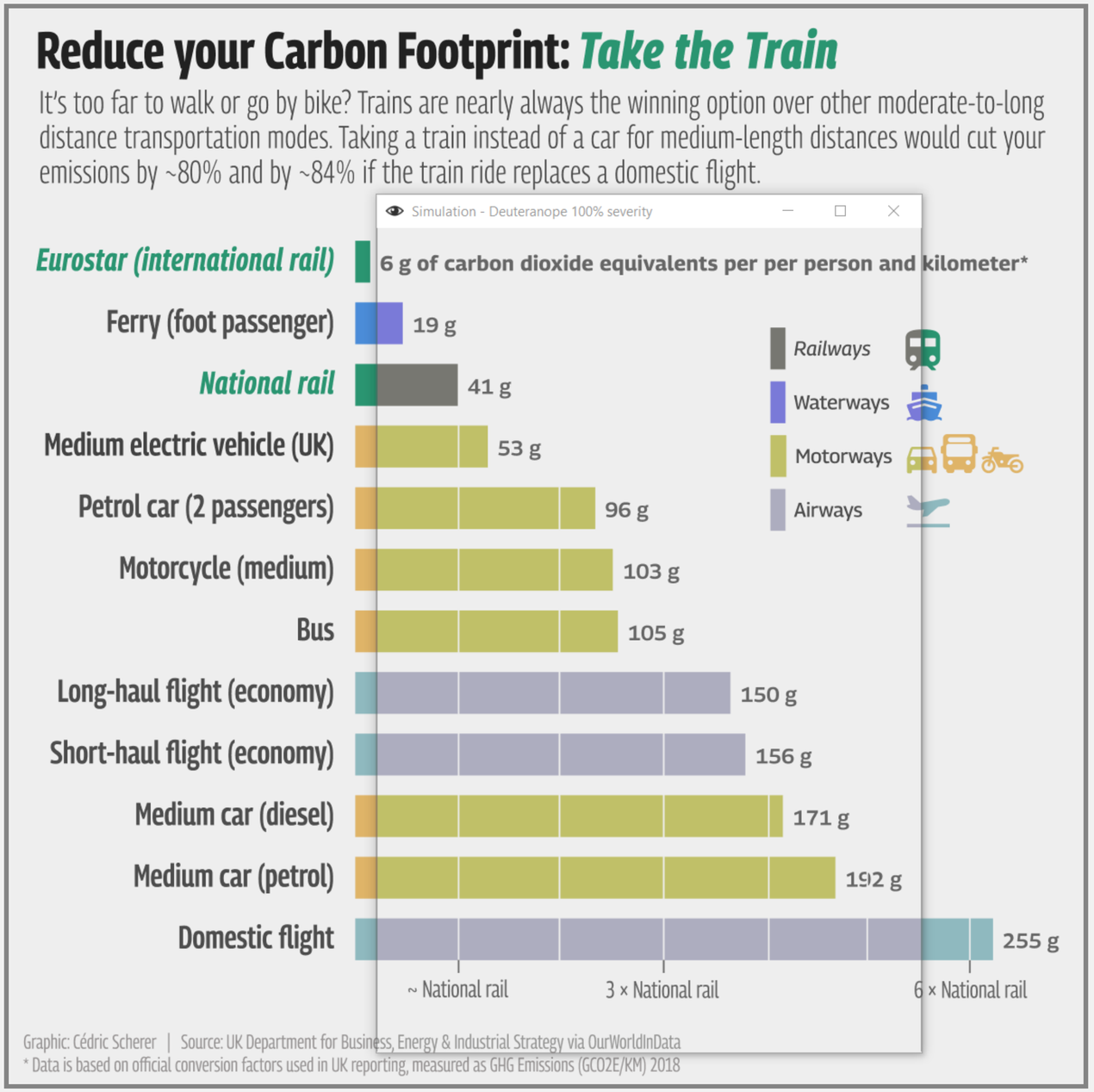


Graphic: Cédric Scherer | Source: UK Department for Business, Energy & Industrial Strategy via OurWorldInData
* Data is based on official conversion factors used in UK reporting, measured as GHG Emissions (GCO2E/KM) 2018

Reworked gaphic as contribution to the #30DayChartChallenge 2022

cedricscherer.com @CedScherer z3tt

Don’t trust color palettes — test them!

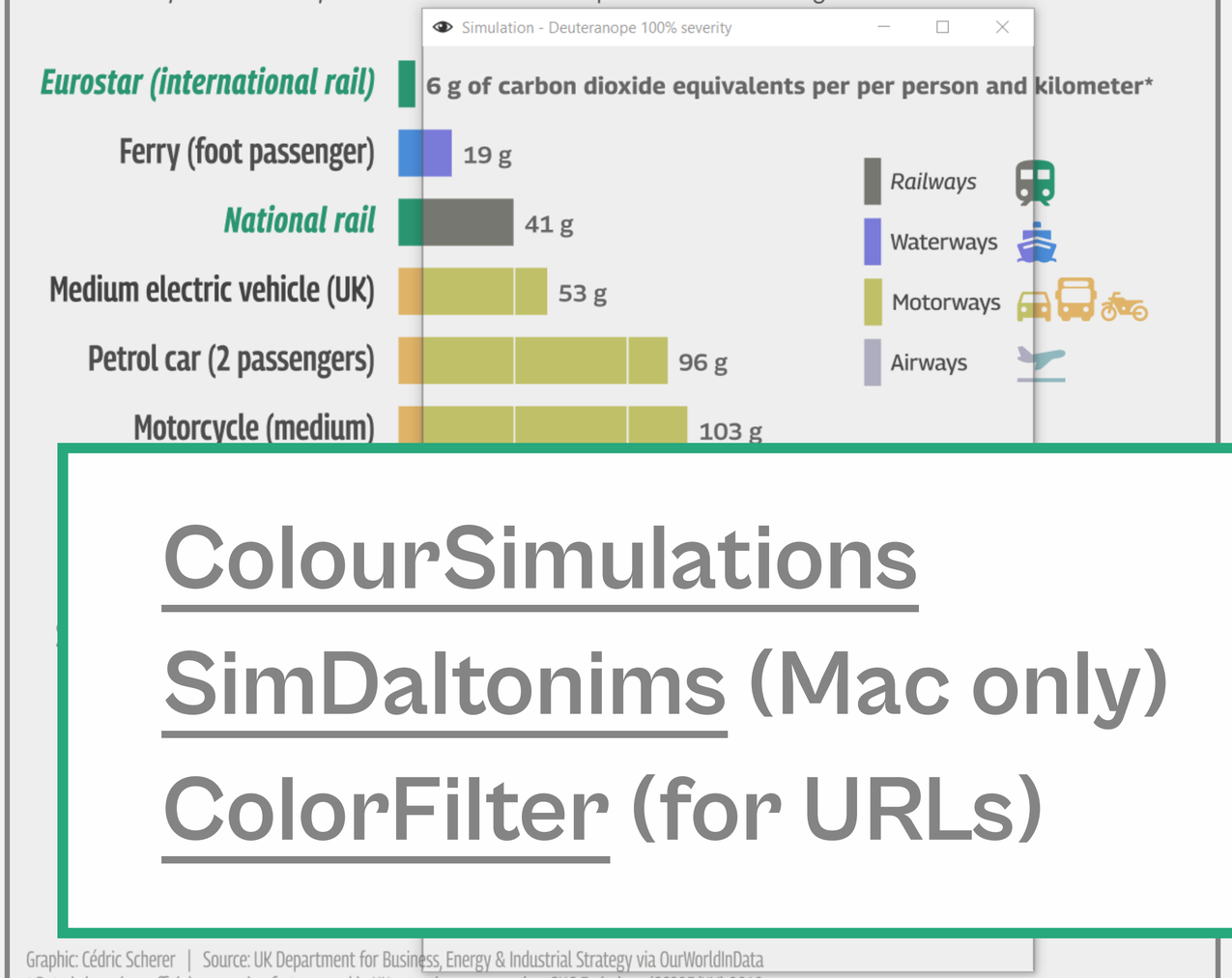


Reworked gaphic as contribution to the #30DayChartChallenge 2022

cedricscherer.com @CedScherer z3tt

Reduce your Carbon Footprint: *Take the Train*

It's too far to walk or go by bike? Trains are nearly always the winning option over other moderate-to-long distance transportation modes. Taking a train instead of a car for medium-length distances would cut your emissions by ~80% and by ~84% if the train ride replaces a domestic flight.



Graphic: Cédric Scherer | Source: UK Department for Business, Energy & Industrial Strategy via OurWorldInData
* Data is based on official conversion factors used in UK reporting, measured as GHG Emissions (GCO2E/KM) 2018

ColourSimulations

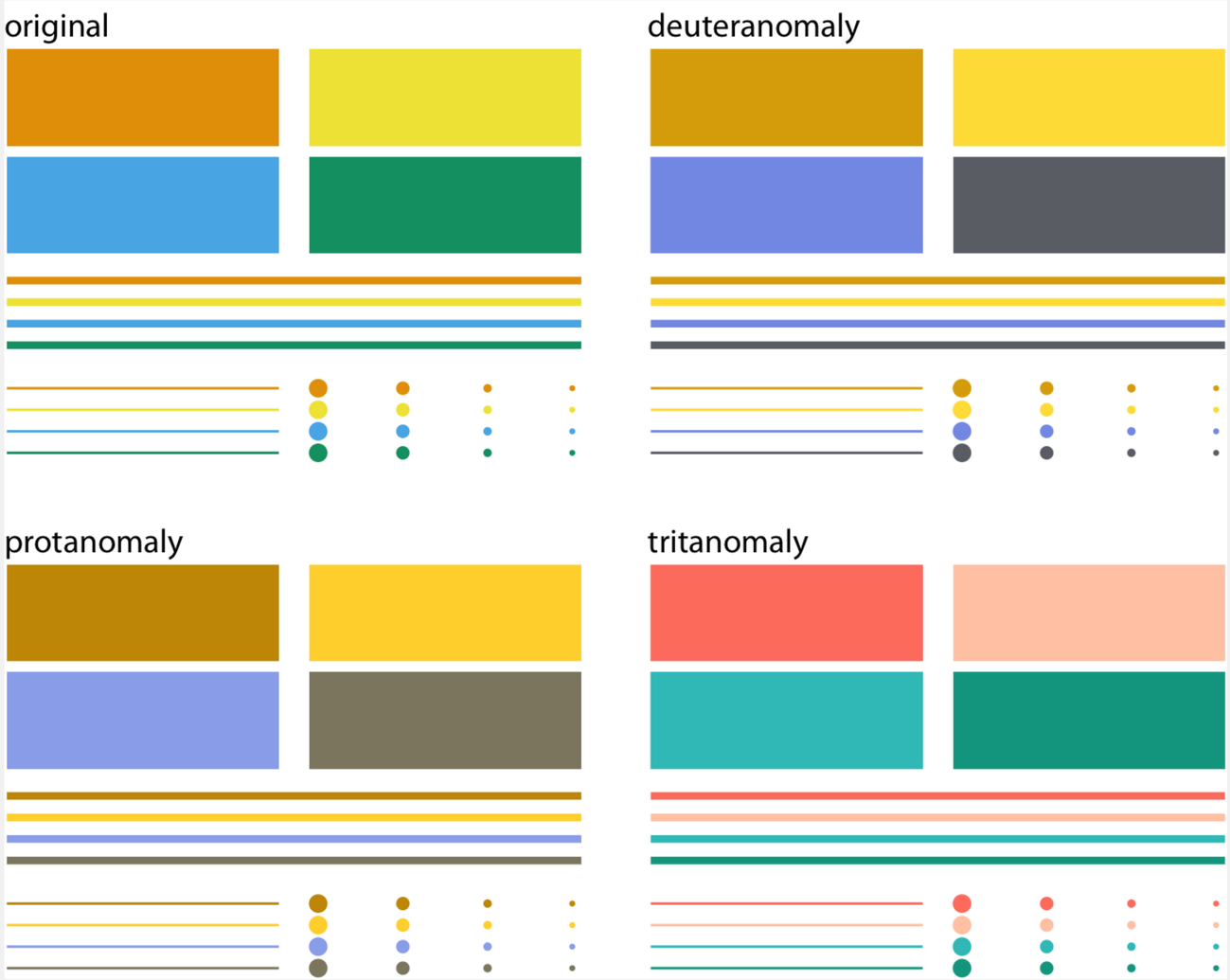
SimDaltonims (Mac only)

ColorFilter (for URLs)

Don't trust color palettes **— test them!**

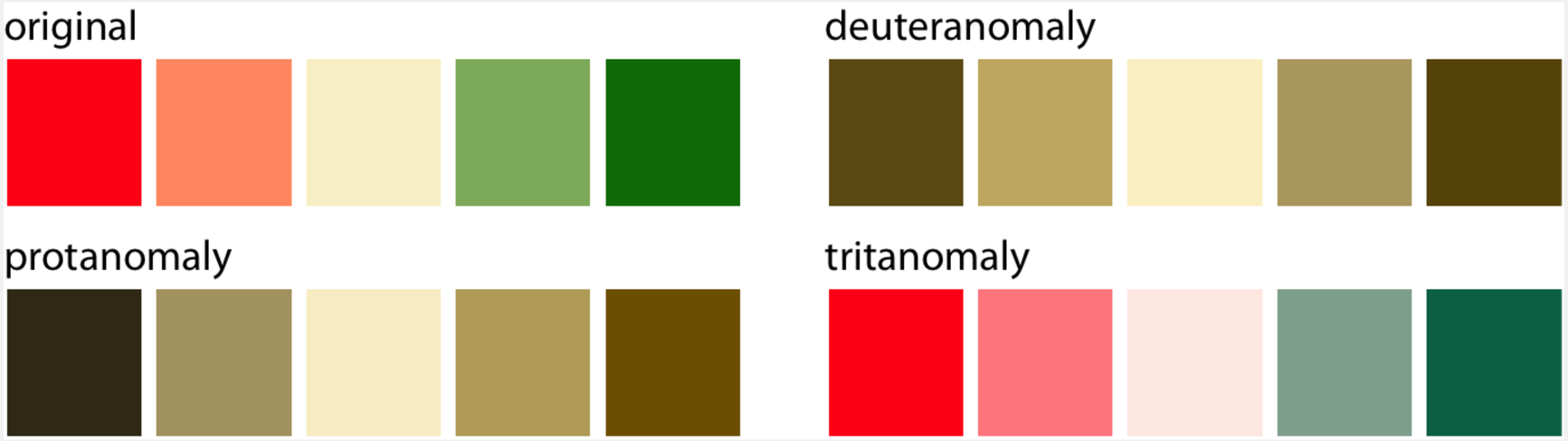
Reworked gaphic as contribution to the #30DayChartChallenge 2022

Ensure Readibility for Color-Blind Persons



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

Ensure Readibility for Color-Blind Persons



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

Ensure Readibility for Color-Blind Persons

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

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

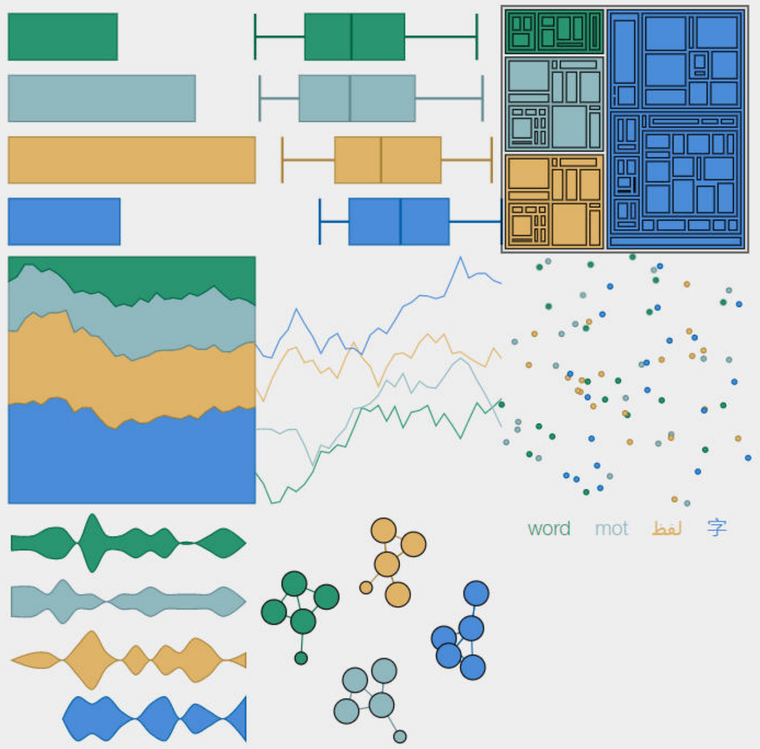
Background color: #eeeeee
Font color: #212121

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



Viz Palette displaying the "colors in action" without color deficiency

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

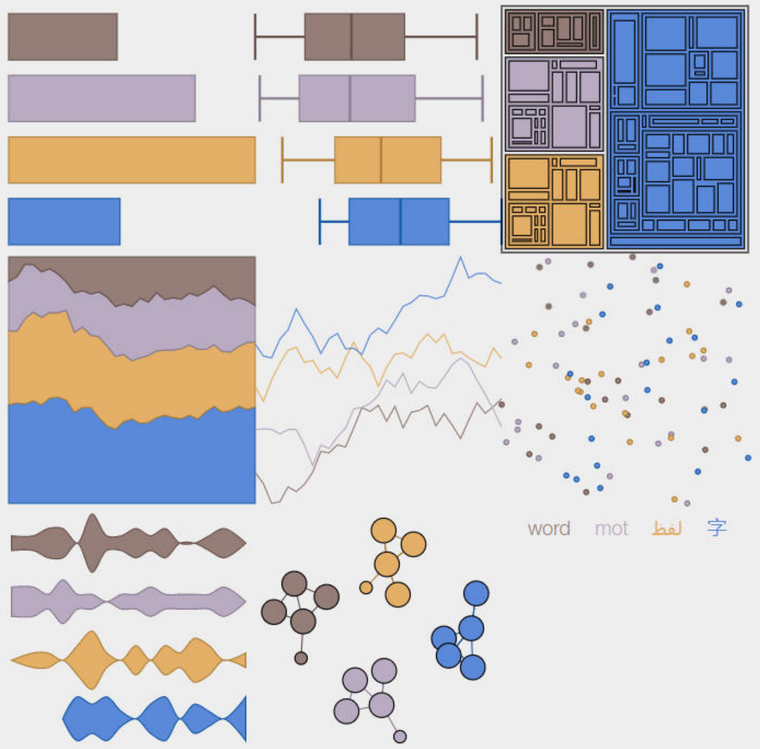
Background color: #eeeeee
Font color: #212121

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



Viz Palette displaying the "colors in action" without color deficiency

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

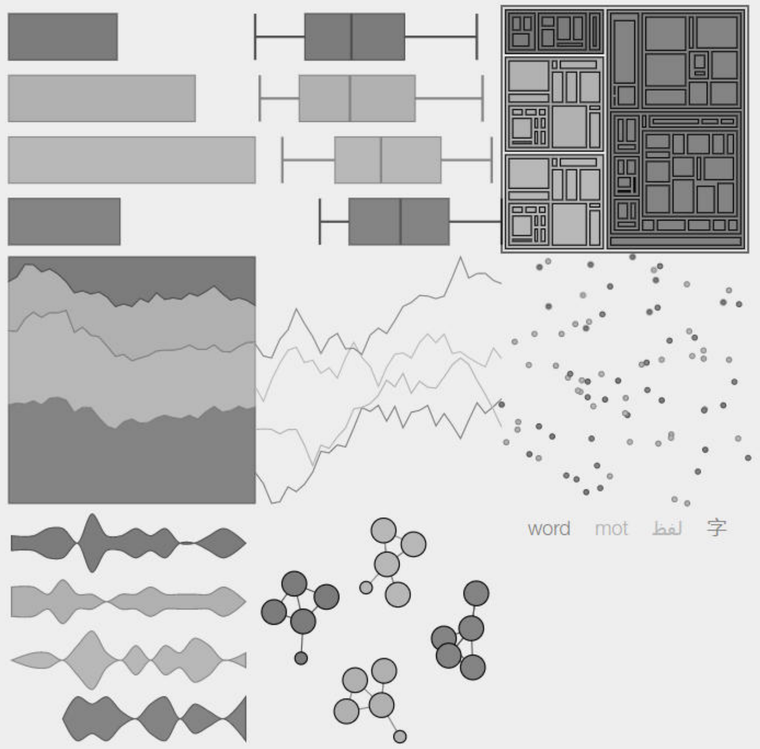
Background color: #eeeeee
Font color: #212121

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



Viz Palette displaying the "colors in action" without color deficiency

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

#hex rgb hsl

```
["#2a9571",
"#8fb9bf",
"#dfb468",
"#4b8cd8"]
```

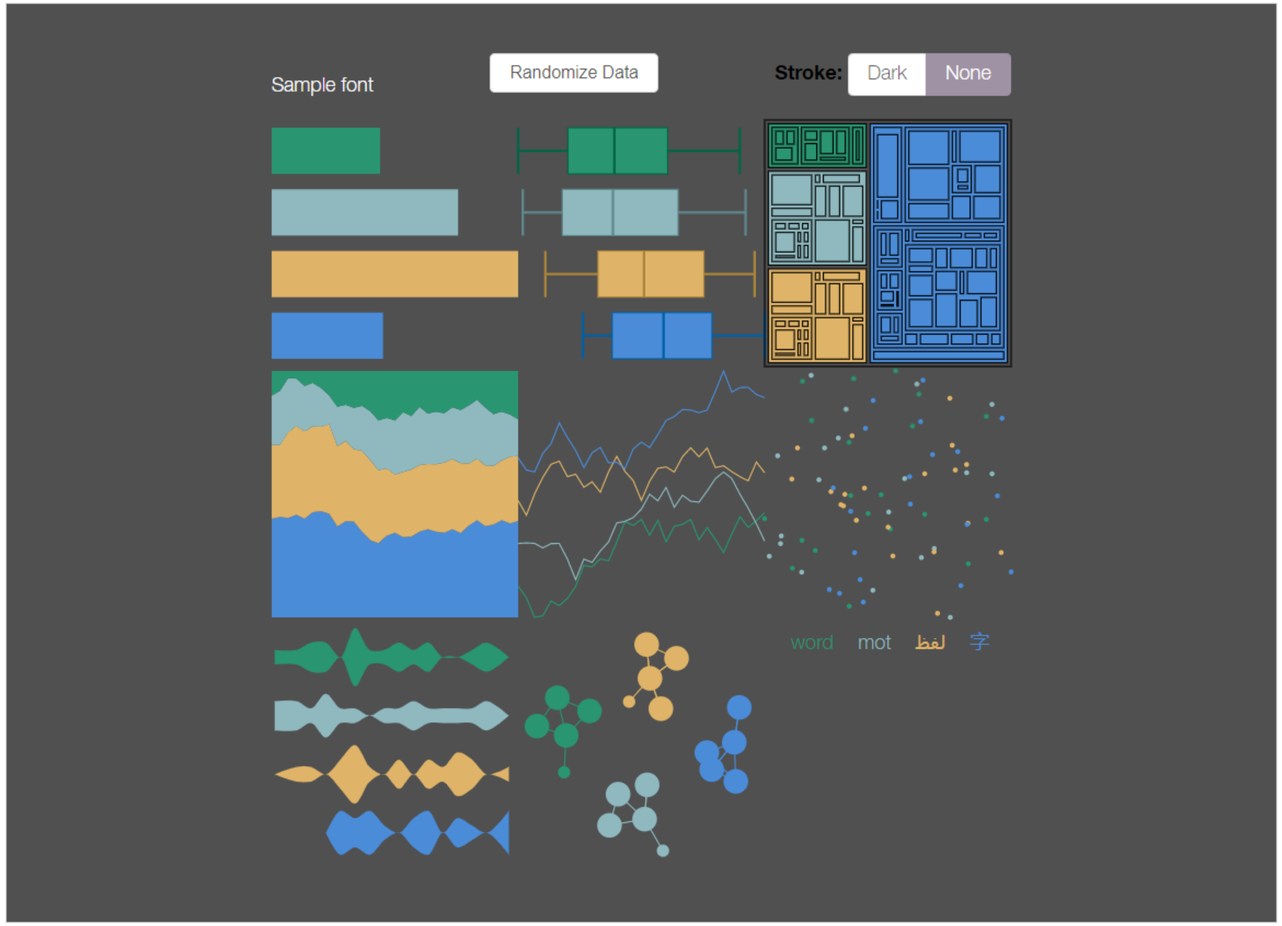
COLORS IN ACTION

Background color: #505050
Font color: #fefefe

Charts made with Semiotic

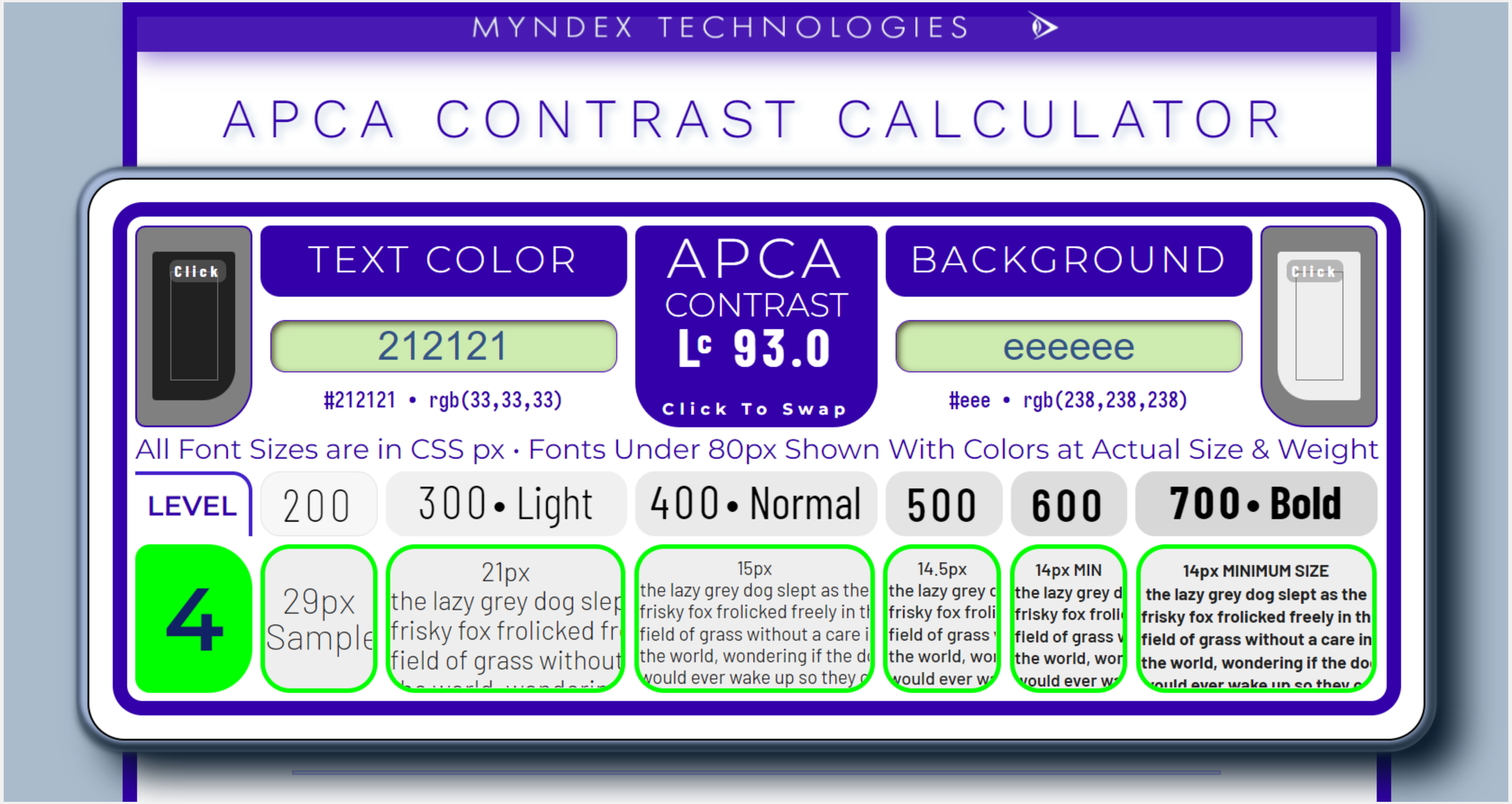
Color Population:

No Color Deficiency - 96% | Deuteranomaly - 2.7% | Protanomaly - 0.66% | Protanopia - 0.59% | Deuteranopia - 0.56% | Greyscale

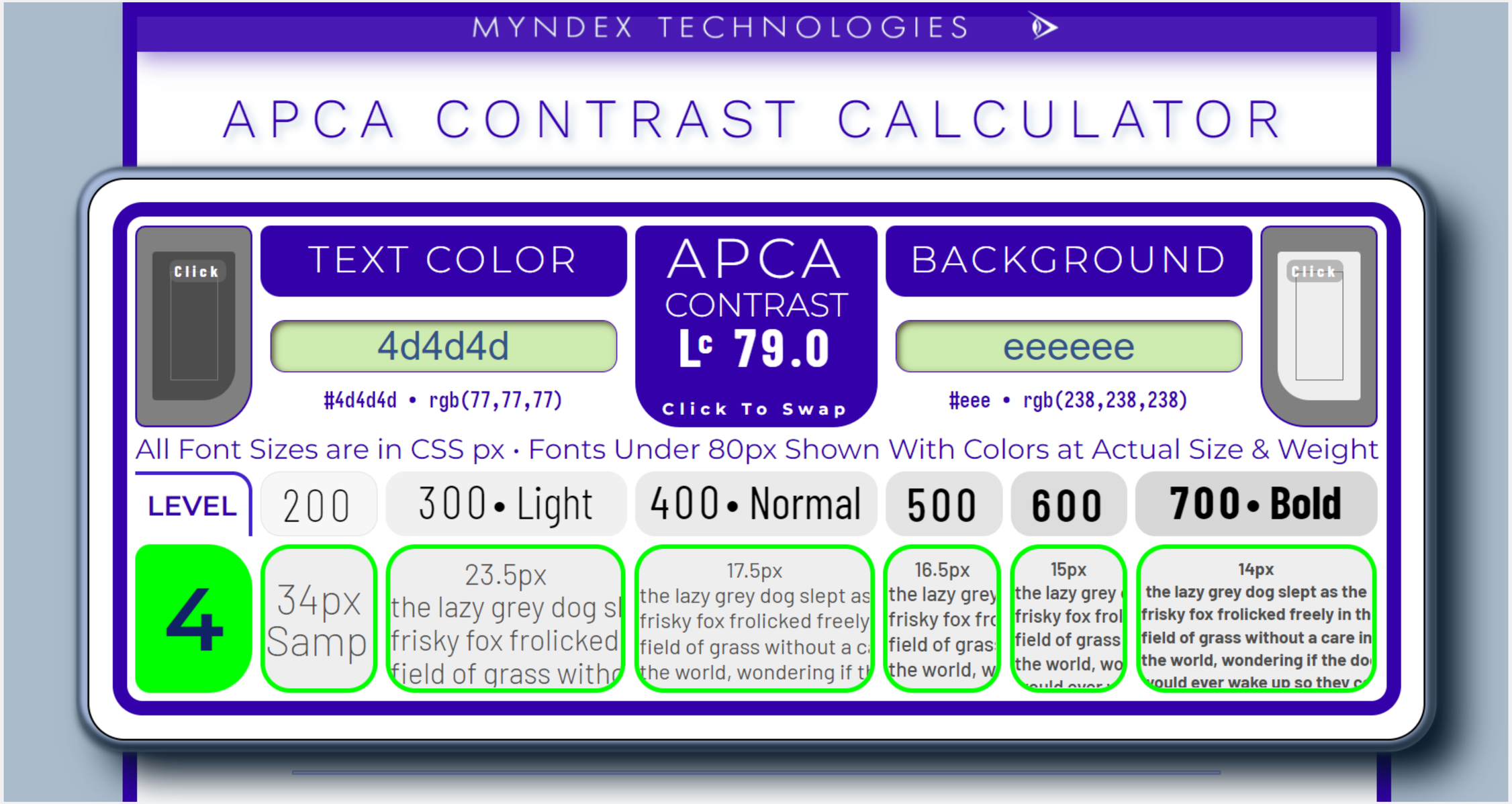


Viz Palette displaying the "colors in action" without color deficiency

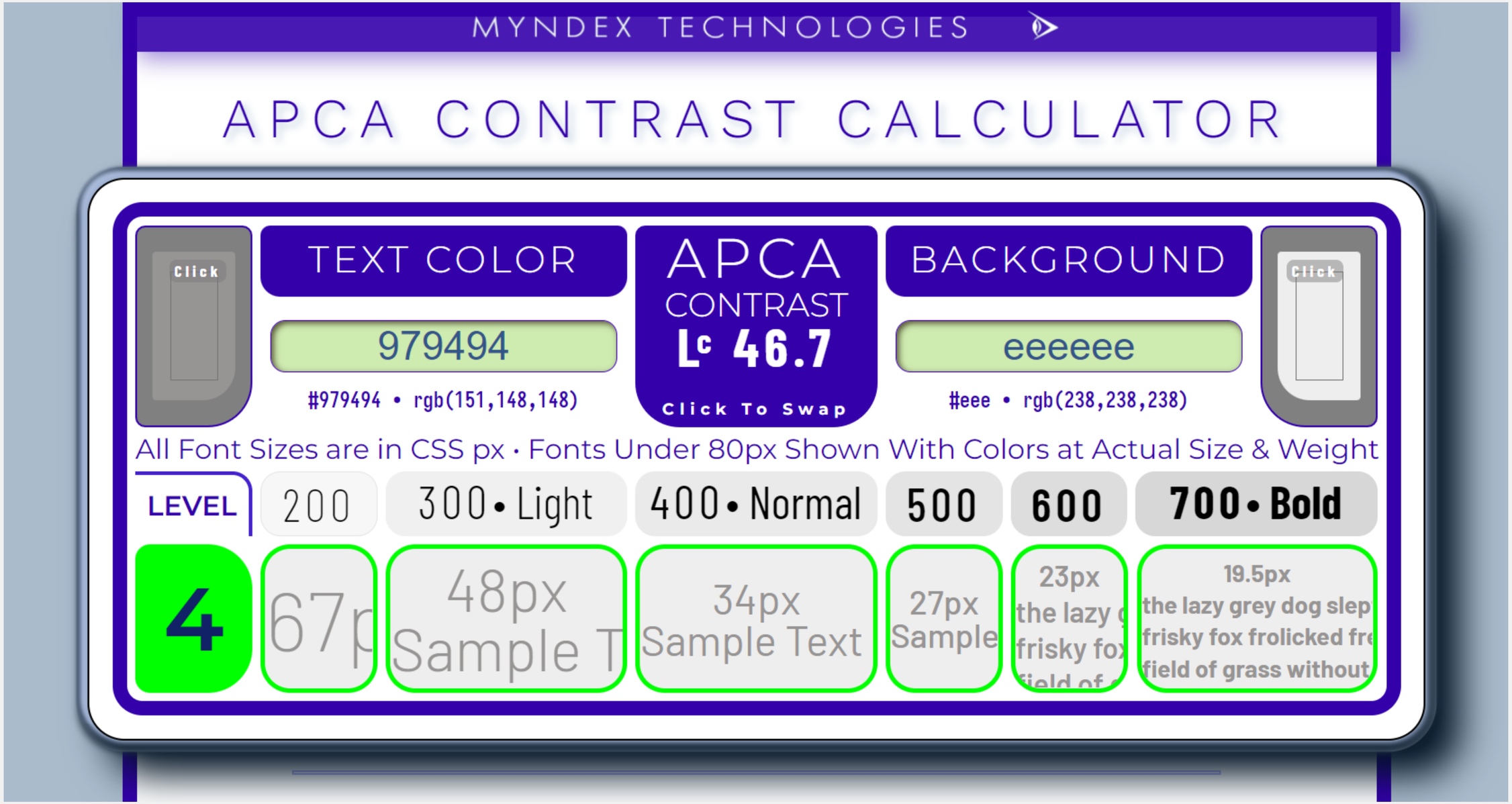
cedricscherer.com @CedScherer z3tt



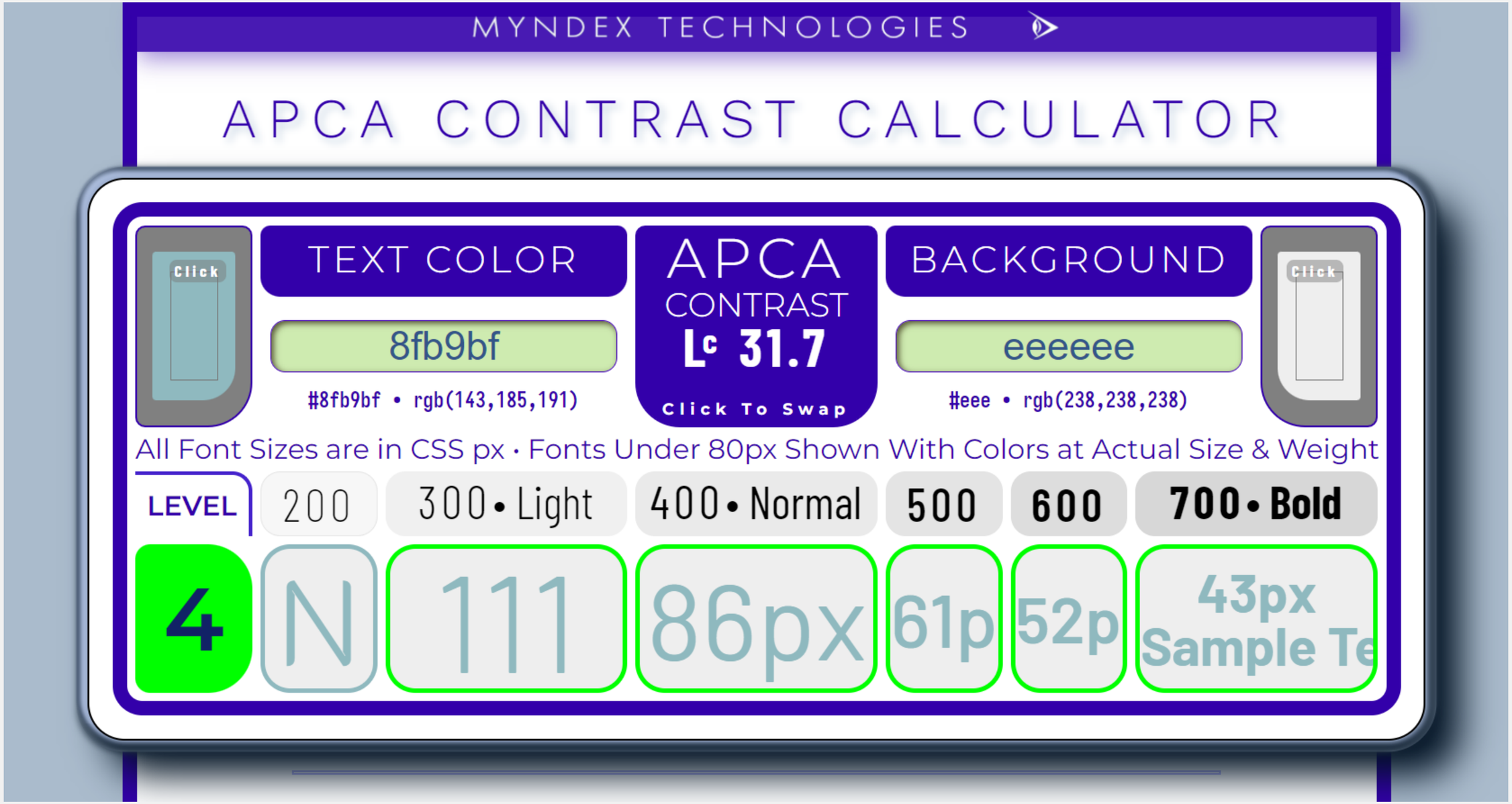
The Myndex APCA Contrast Calculator displays modern contrast ratios for various combinations of text size and font weight



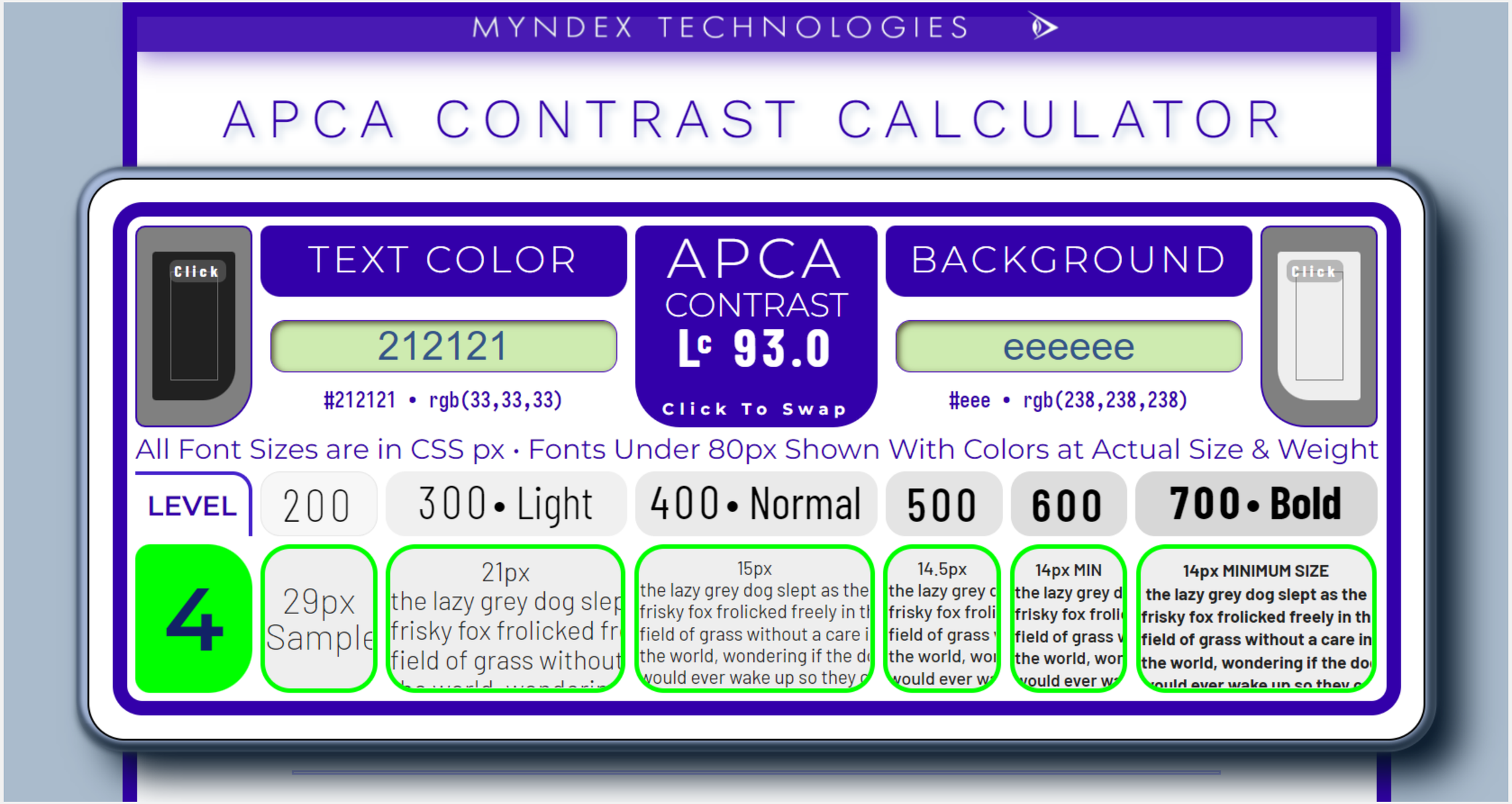
The Myndex APCA Contrast Calculator displays modern contrast ratios for various combinations of text size and font weight



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

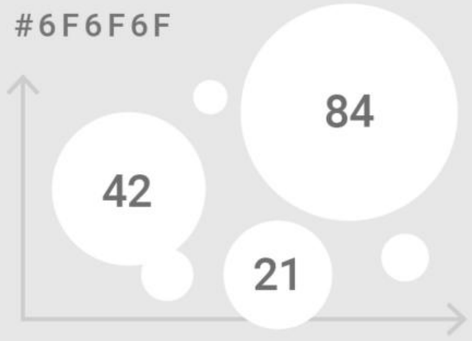


The Myndex APCA Contrast Calculator displays modern contrast ratios for various combinations of text size and font weight



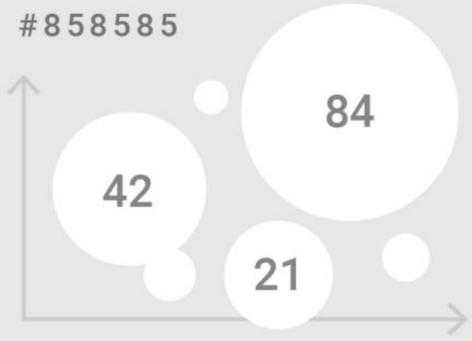
The Myndex APCA Contrast Calculator displays modern contrast ratios for various combinations of text size and font weight

#6F6F6F



WCAG 2.0: 5:1 AAA
APCA: 75% AAA

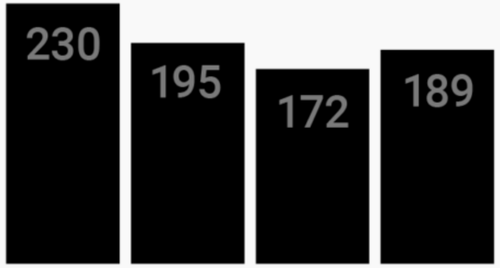
#858585



WCAG 2.0: 3.7:1 AA
APCA: 65% AAA

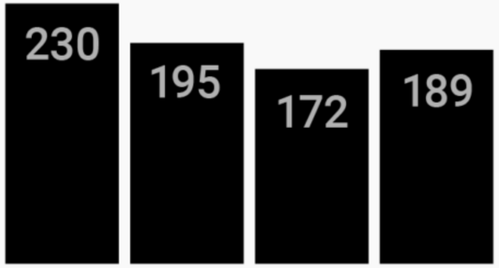
APCA allows for lighter grey on light backgrounds...

#7A7A7A



WCAG 2.0: 4.9:1 AAA
APCA: 32% ✗

#B1B1B1



WCAG 2.0: 9.8:1 AAA
APCA: 60% AAA

... but also requests lighter grey on dark backgrounds

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

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)]

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

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

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

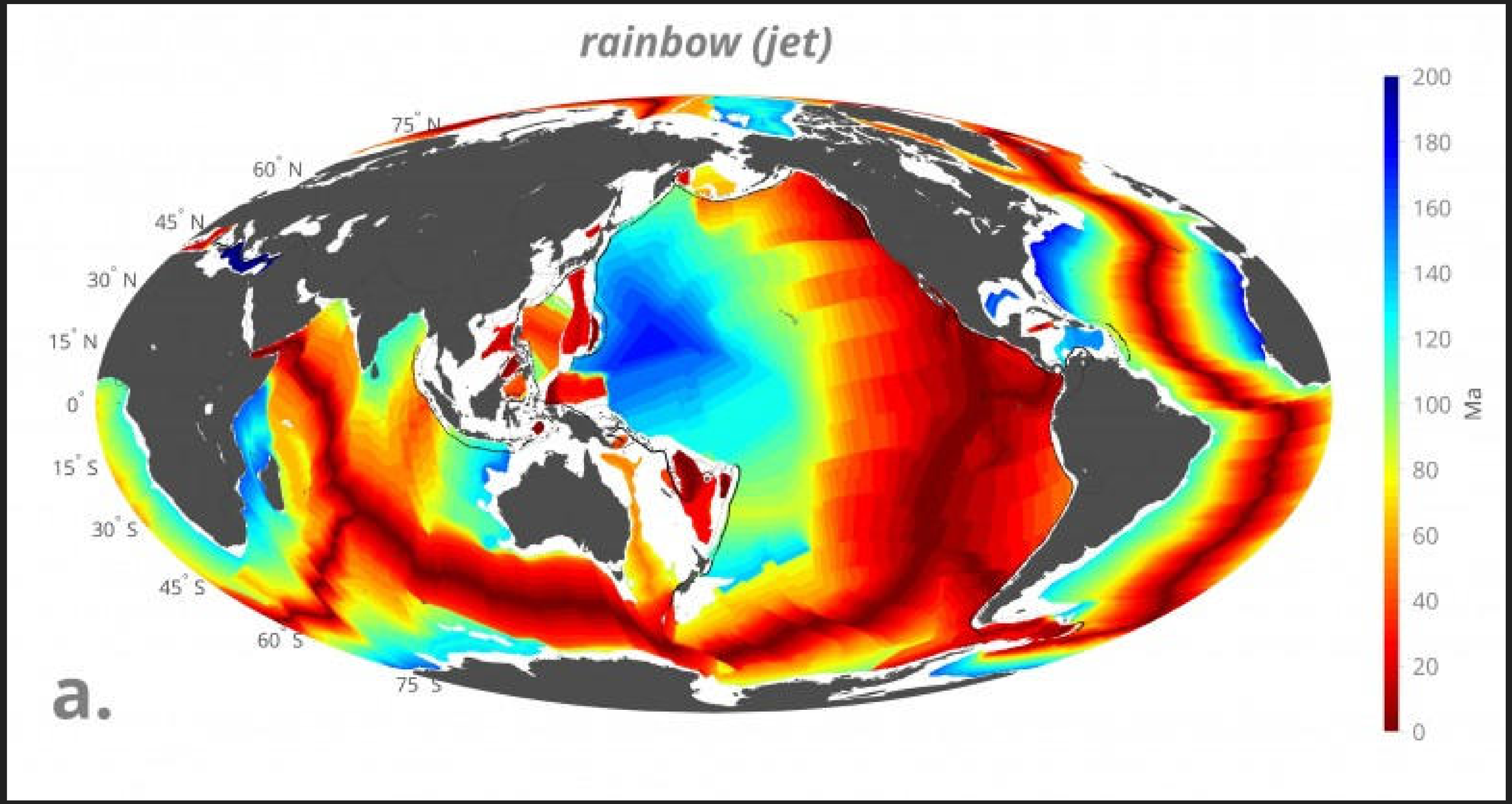
PMCID: PMC5148121 | PMID: 26127048

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)]

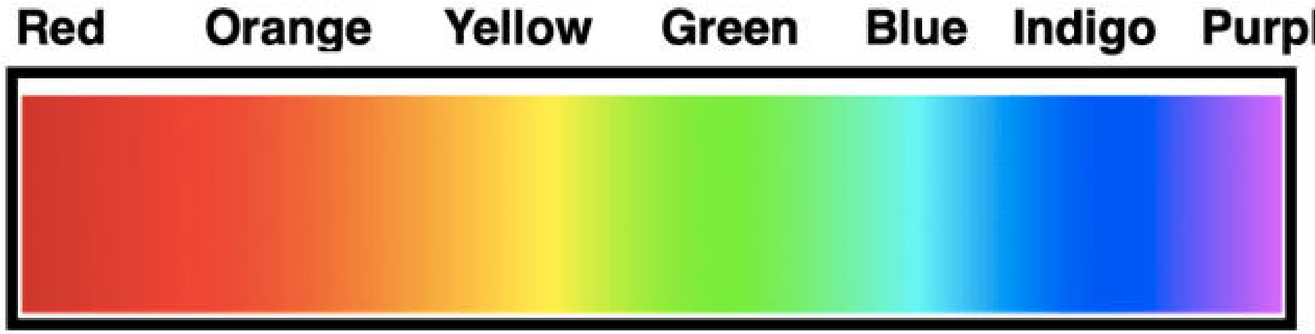
"The ad hoc manner in which color is handled and the lack of standard approaches have been associated with suboptimal and **inconsistent diagnostic decisions with a negative impact on patient treatment and prognosis**."

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

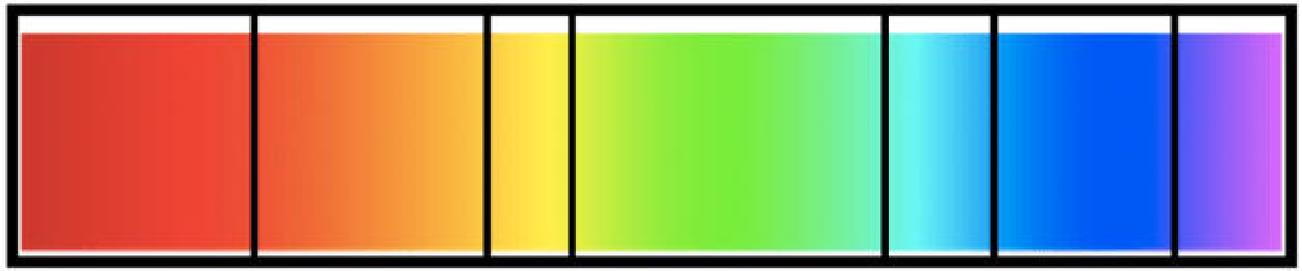


Source: "The Rainbow Colour Map (repeatedly) considered harmful", EGU Blogs

Non-Uniform Distances between Hues in the Newton Rainbow Colormap



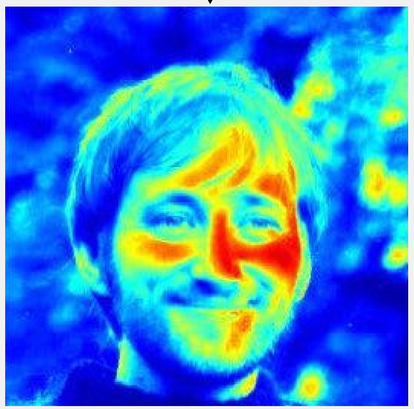
Typical Rainbow Colormap



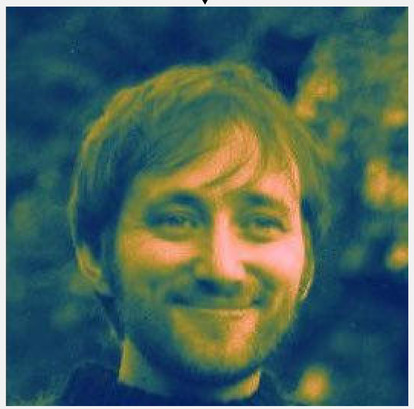
Non-Uniform Distances between Hues.

Modified from Fabio Crameri

true-color Phil



rainbow Phil



batlow Phil

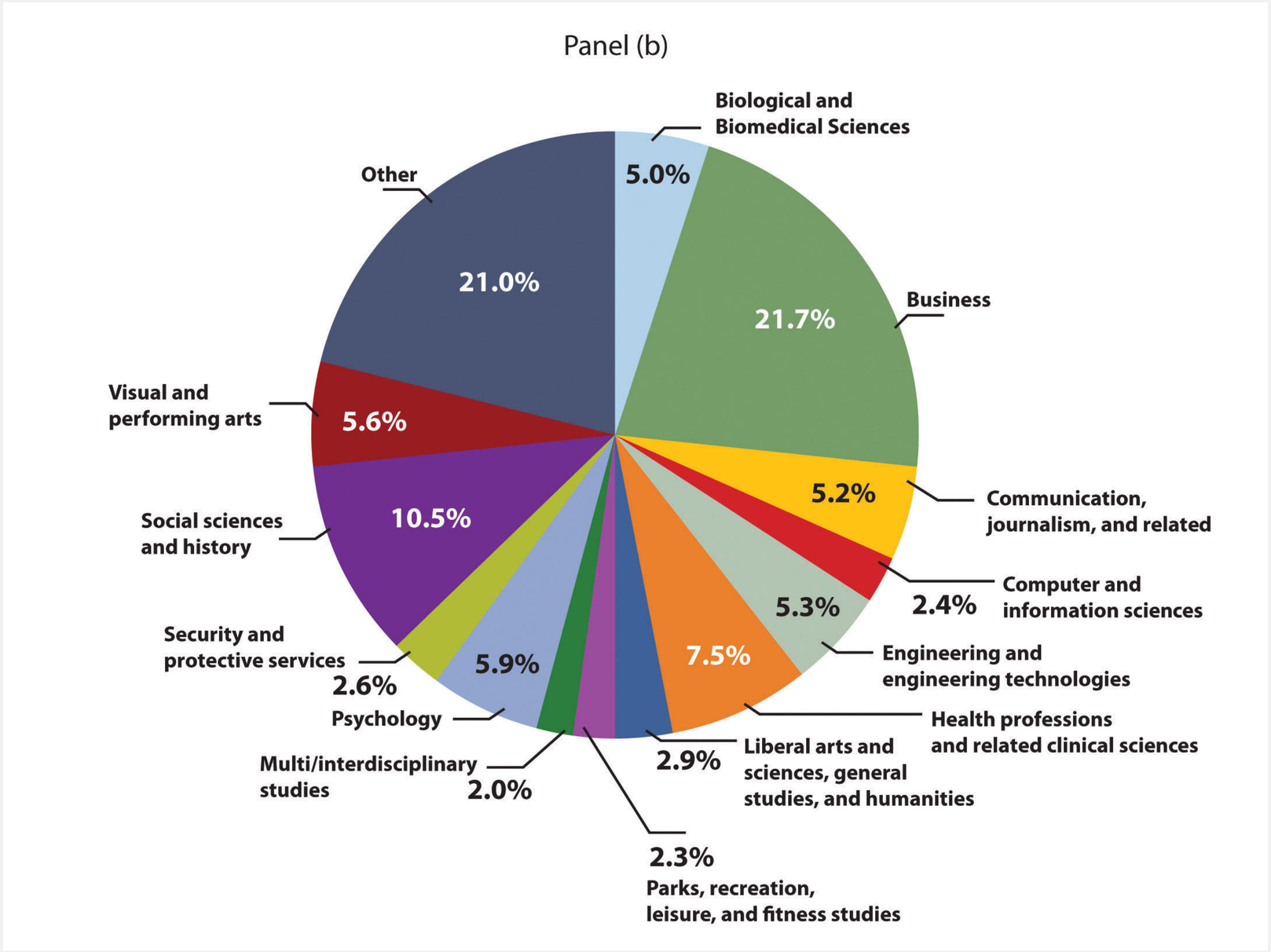
Modified from Fabio Crameri

Summary

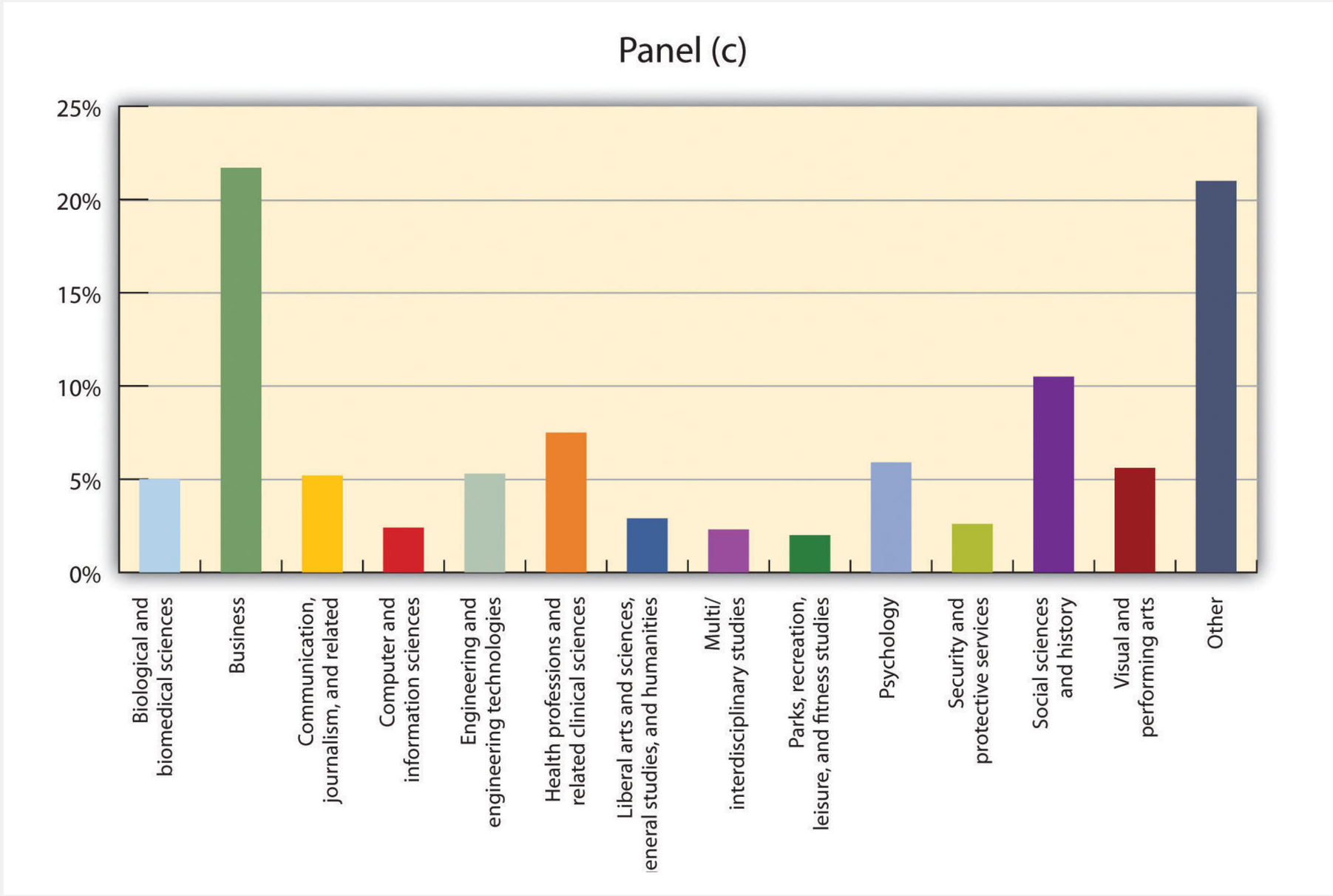
Panel (a)

Field of Study	Percentage of Bachelor's Degrees
Biological and biomedical sciences	5.0%
Business	21.7%
Communication, journalism, and related	5.2%
Engineering and engineering technologies	5.3%
Health professions and related clinical sciences	7.5%
Liberal arts and sciences, general studies, and humanities	2.9%
Psychology	5.9%
Social sciences and history	10.5%
Visual and performing arts	5.6%
Security and protective services	2.6%
Computer and information sciences	2.4%
Multi/interdisciplinary studies	2.3%
Parks, recreation, leisure, and fitness studies	2.0%
Other	21.0%

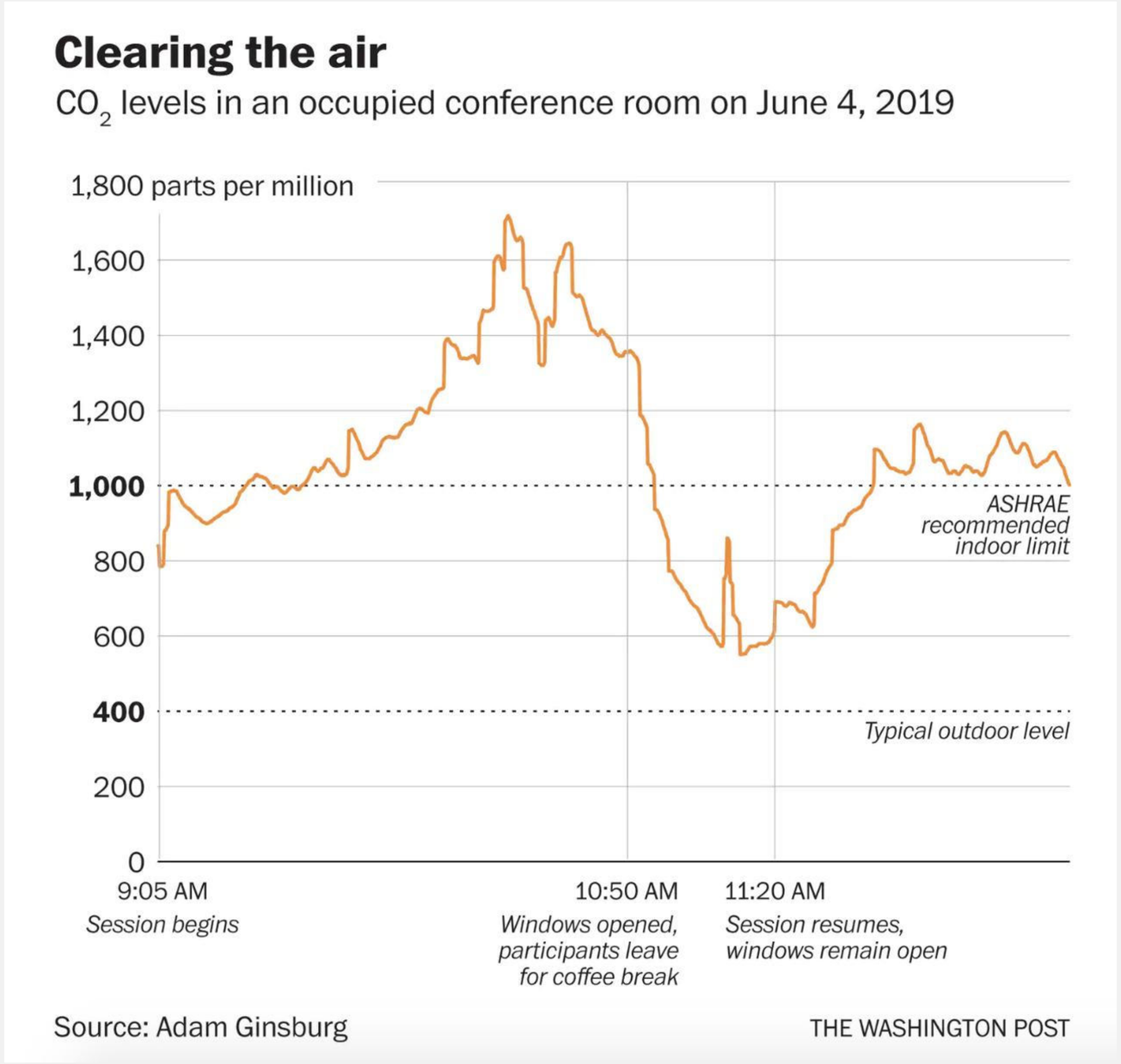
Source: saylordotorg.github.io/text_principles-of-economics-v2.0



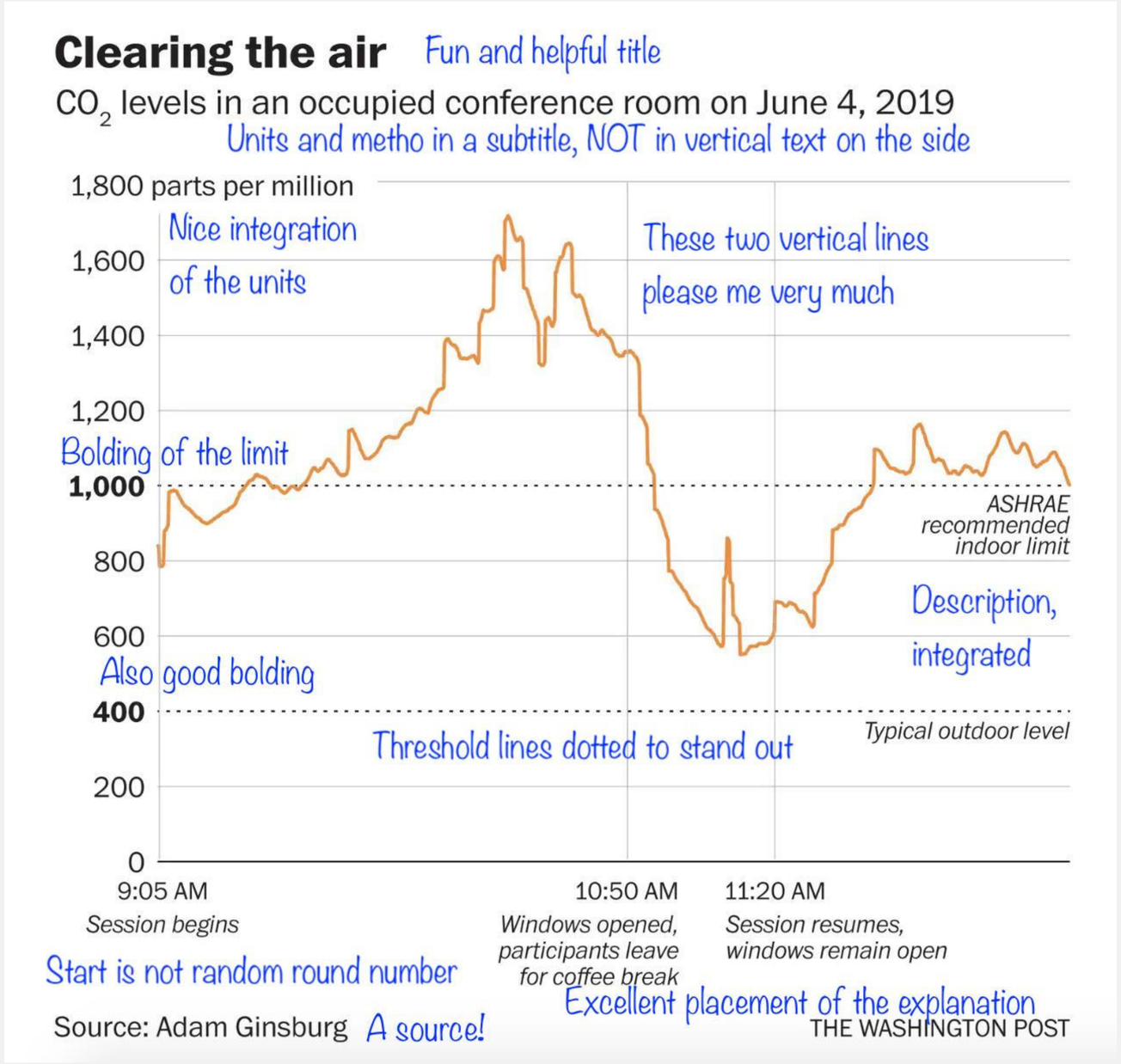
Source: saylordotorg.github.io/text_principles-of-economics-v2.0



Source: saylordotorg.github.io/text_principles-of-economics-v2.0



Source: "Clearin the Air" by Adam Ginsburg (Washington Post)



Notes by Francis Gagnon (Voilà)

Information

Understand your data and be accurate.

Story

Be clear about the message of your visualization.

Goal

Select charts that successfully transport your story.

Visual Form

Present information in a logical, coherent way.

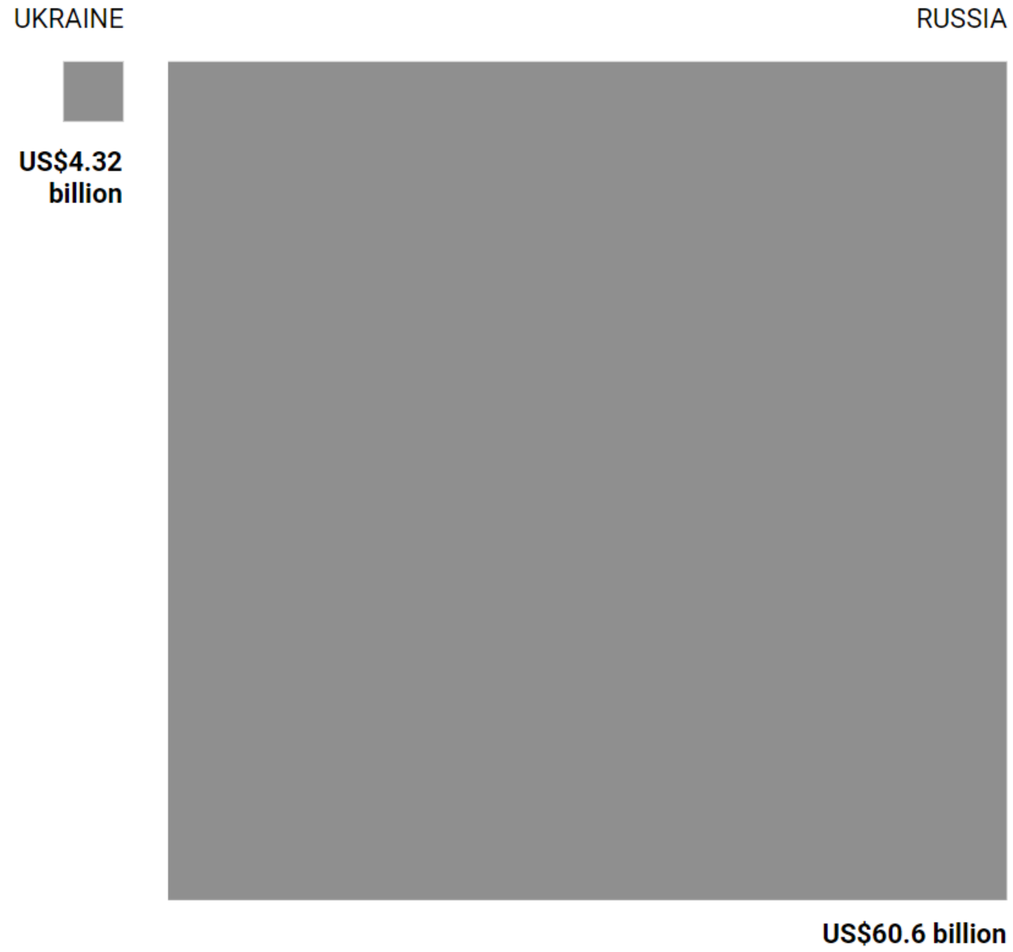
Group Exercise

Discuss the visualizations with regard to the 4 levels:

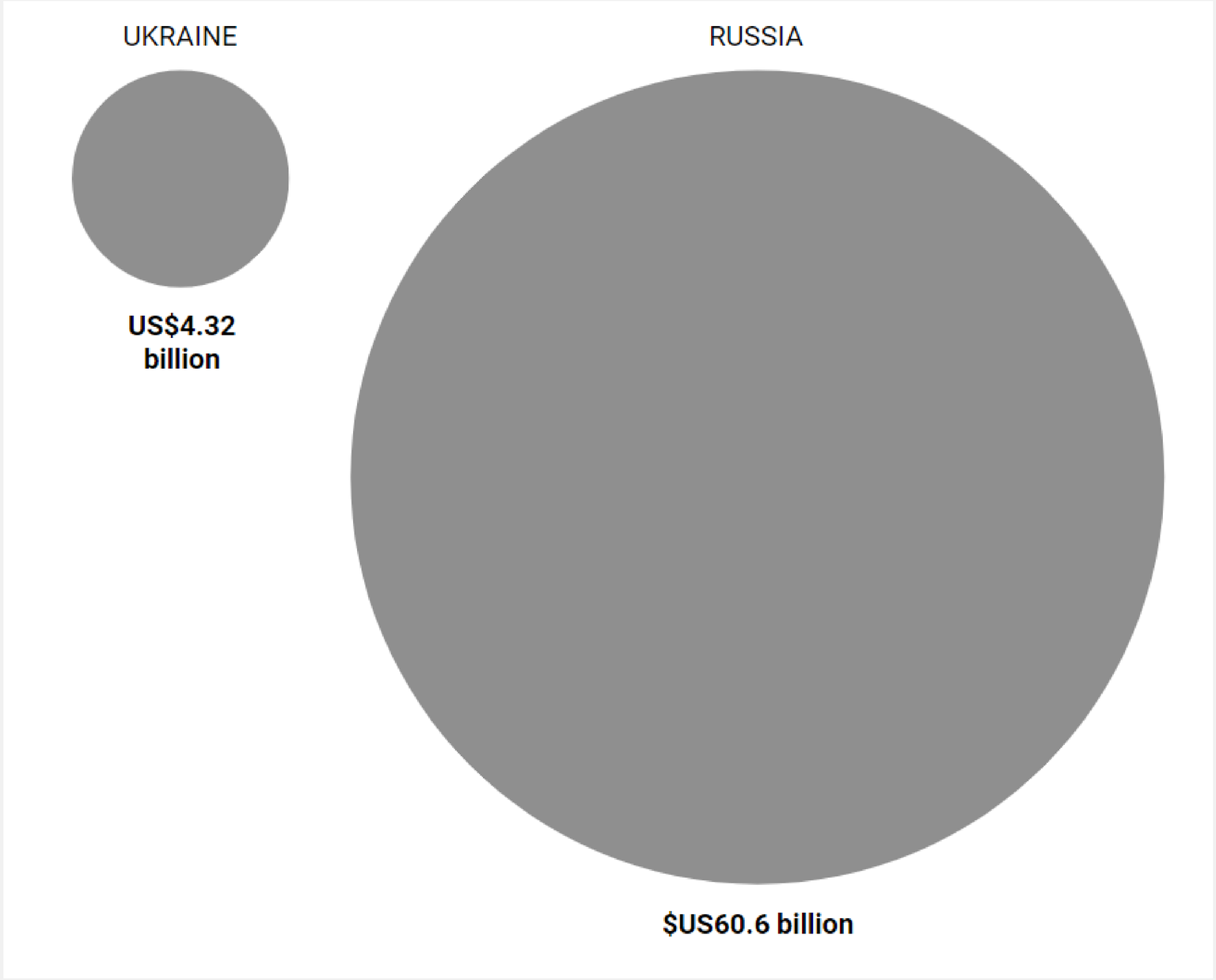
- ☛ Overall, do you think it is a good or a bad visualization?
- ☛ What are details you like?
- ☛ How could one improve the chart?
- ☛ Is there another (potentially better) way to tell the story?

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.



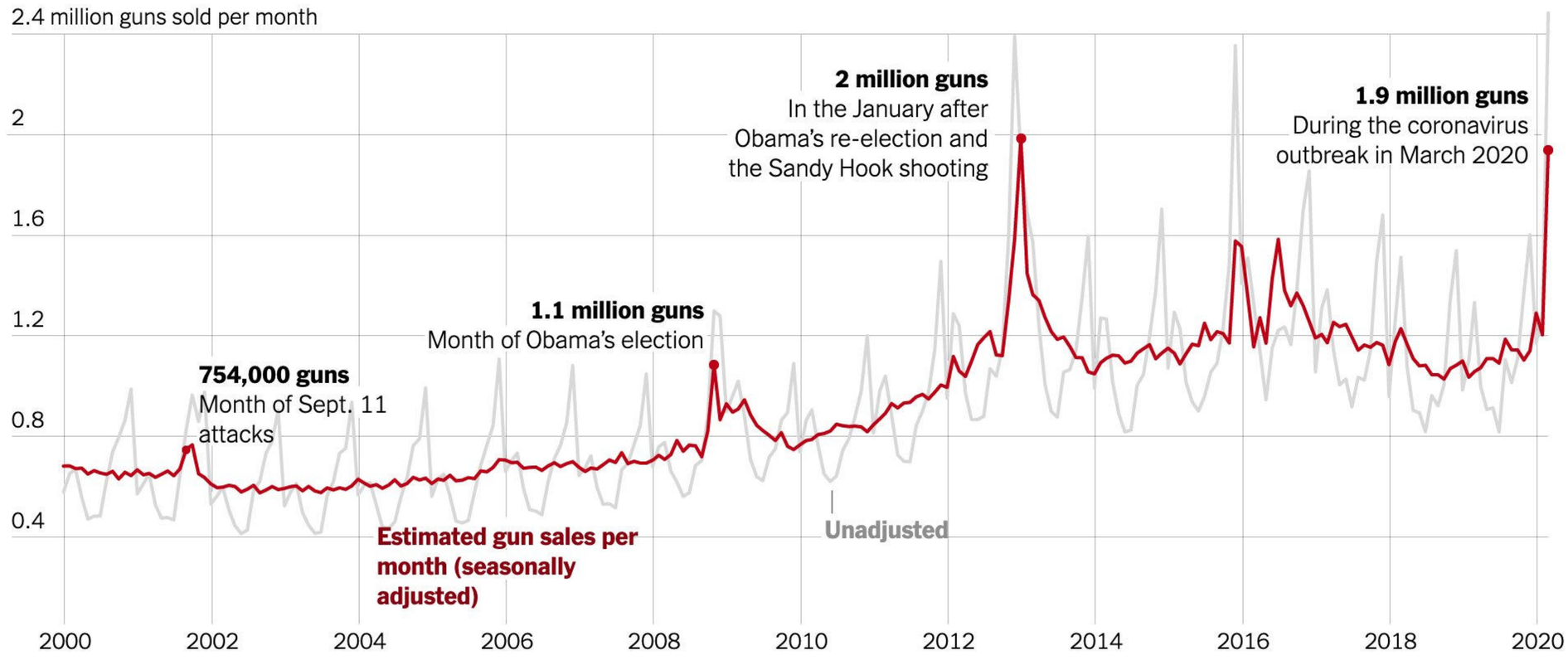
"Russia attacks Ukraine" by SCMP Graphic (South China Morning Post)



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

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

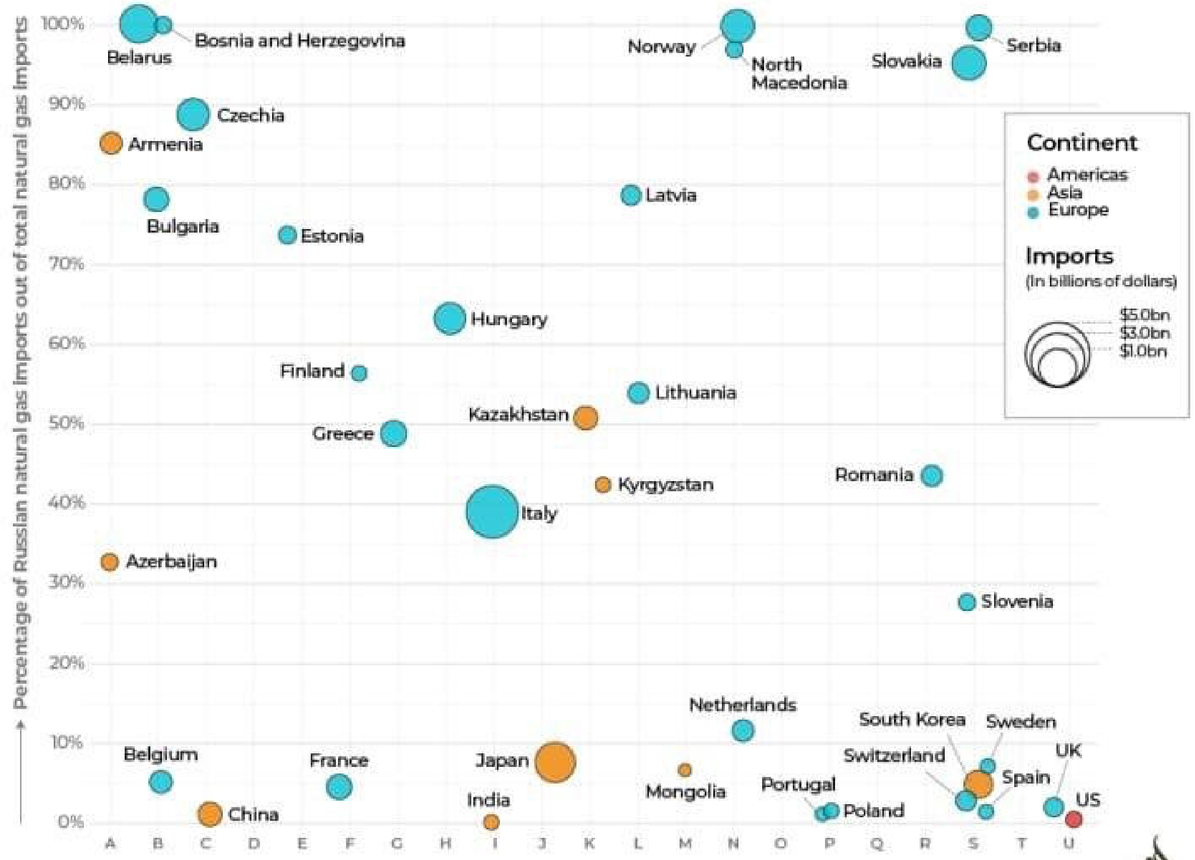
By Keith Collins and David Yaffe-Bellany April 1, 2020



ENERGY

Which countries directly import the most Russian natural gas?

Russia is the world's third highest exporter of natural gas after Australia and Qatar.
In 2019, **37 countries bought Russian gas worth $24.5bn.**



Countries not on this graphic did not import gas from Russia in 2019.
Natural gas includes gas in gaseous state and liquefied.
Source: OEC | 2019

@AJLabs AL JAZEERA

ENERGY

Which countries directly import the most Russian natural gas?

Russia is the world's third highest exporter of natural gas after Australia and Qatar.
In 2019, **37 countries bought Russian gas worth $24.5bn.**

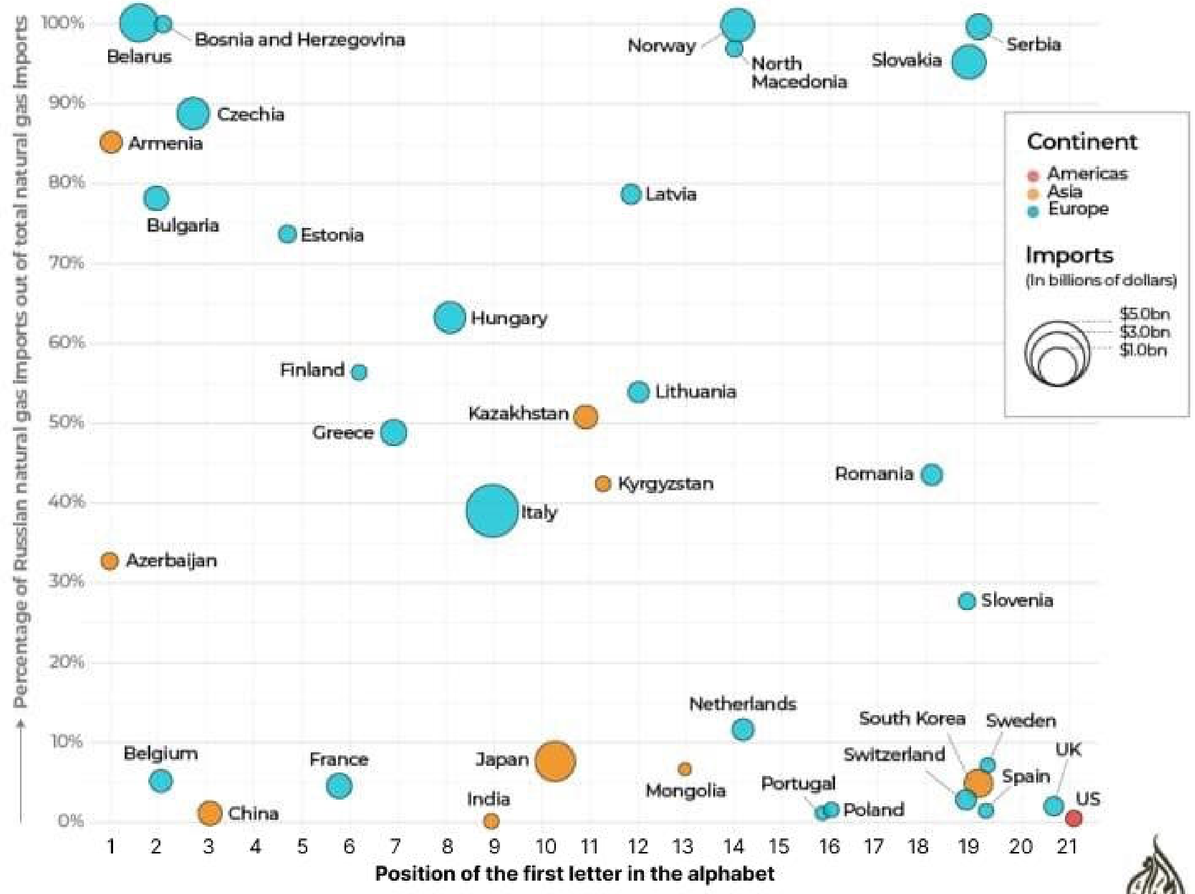
Countries not on this graphic did not import gas from Russia in 2019.
Natural gas includes gas in gaseous state and liquefied.
Source: OEC | 2019

@AJLabs AL JAZEERA

ENERGY

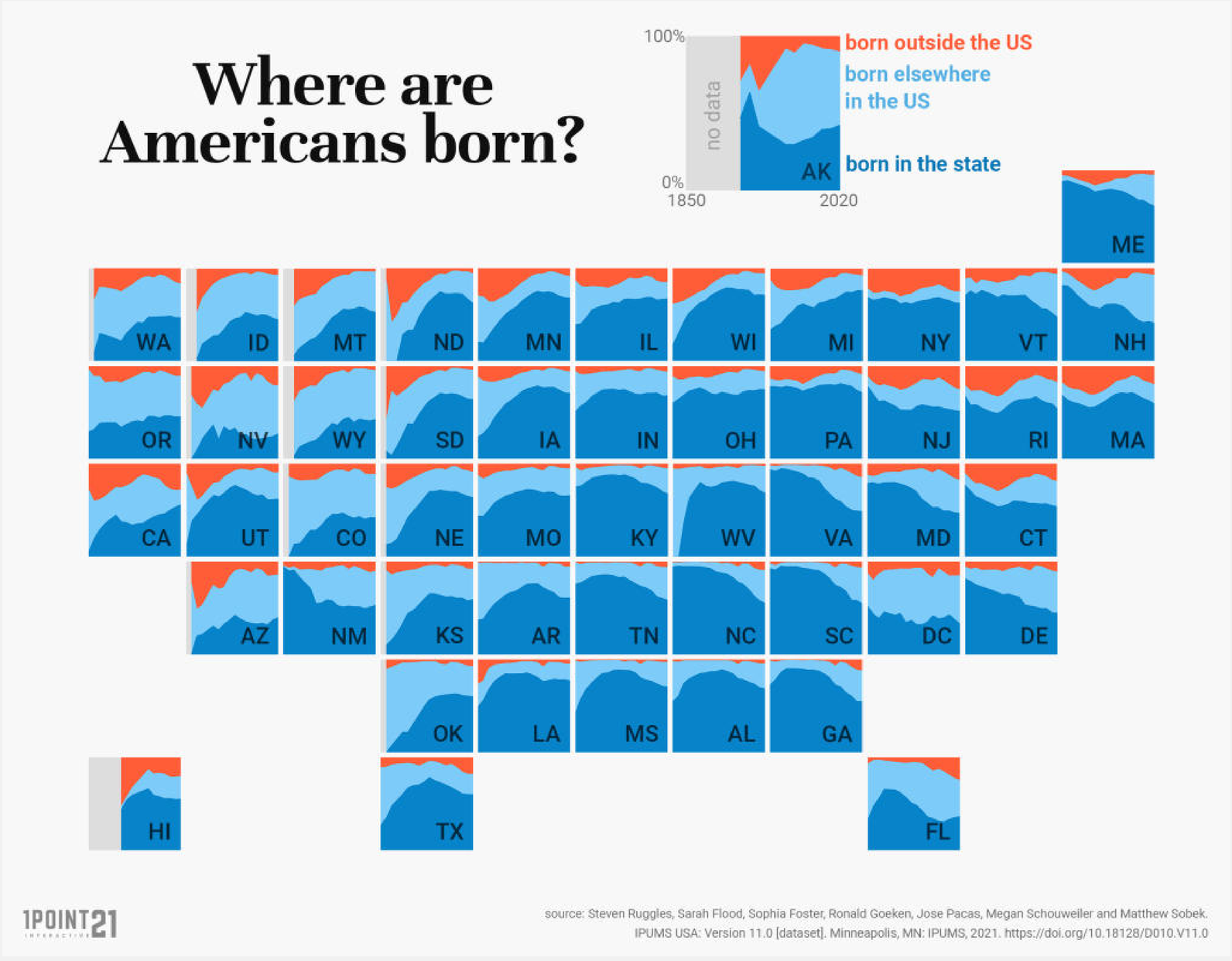
Which countries directly import the most Russian natural gas?

Russia is the world's third highest exporter of natural gas after Australia and Qatar.
In 2019, **37 countries bought Russian gas worth $24.5bn.**



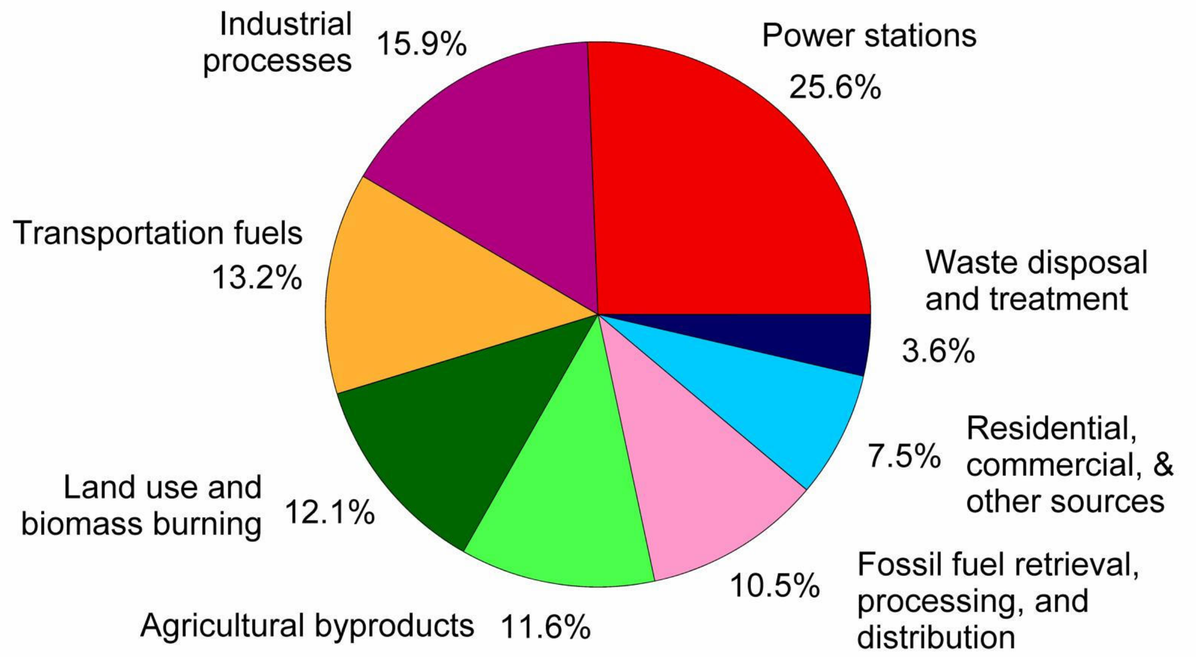
Countries not on this graphic did not import gas from Russia in 2019.
Natural gas includes gas in gaseous state and liquefied.
Source: OEC | 2019

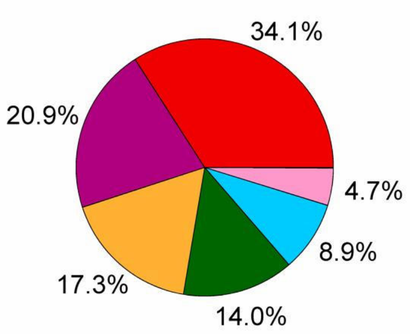
@AJLabs AL JAZEERA



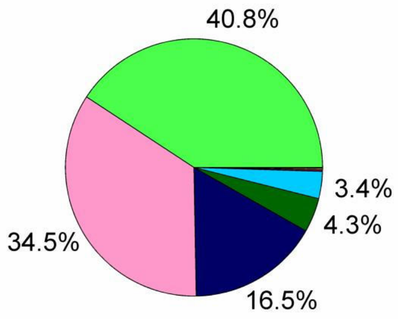
"Where are Americans born?" by @ErinDataViz

Annual Greenhouse Gas Emissions by Sector

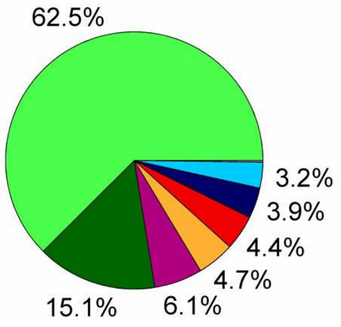




Carbon Dioxide
(72% of total)



Methane
(20% of total)



Nitrous Oxide
(5% of total)

Emissions data for the year 2010 derived from the Emissions Database for Global Atmospheric Research (EDGAR) v4.2 FT2010

"Annual greenhouse gas emissions by sector" by Robert Rohde

Driving Safety, in Fits and Starts

AMERICANS drive a staggering number of miles — close to three trillion every year, according to the government. (That is half a light-year, or 120 million trips around the world.) And although traffic accidents remain a major public safety problem, the biggest killer of people ages 5 to 34, vehicle travel is far safer than it was a few decades ago.

Several factors appear to account for the sharp decline in fatalities. Technology (like anti-lock brakes and air bags) and road behavior (like wearing seat belts and driving sober) have both improved greatly since 1950. Americans almost always drive more each year than the previous one — at least until recently, when the recession curtailed road habits. And the auto fatality rate has been decreasing since the 1960s, when cars with massive engines carried their unbuckled passengers on primarily two-lane roads.

The safety data is usually charted as deaths per miles traveled. But what happens when the metrics are teased apart, and familiar data is charted in an unfamiliar way?

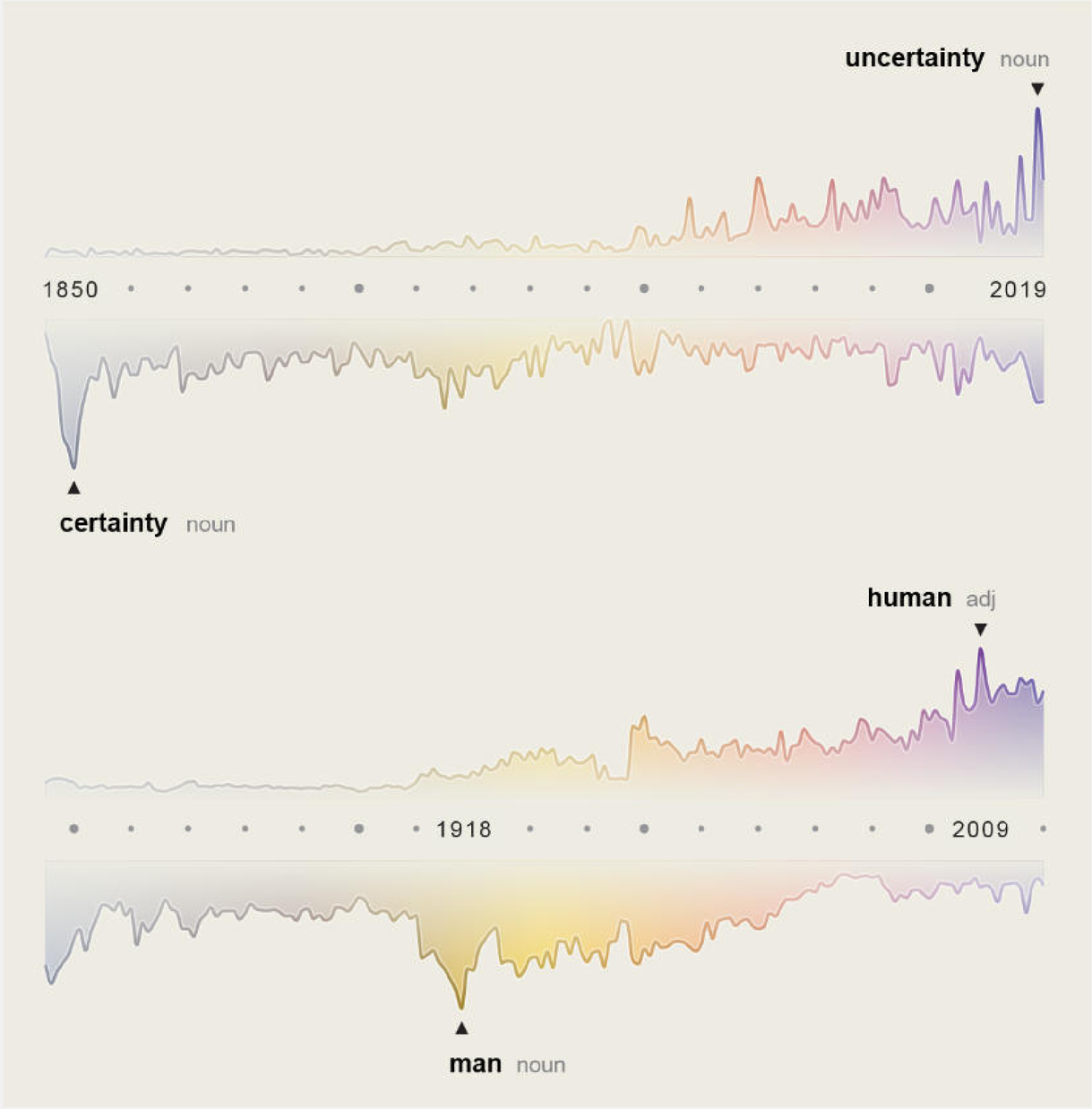
Plotting the two most important variables against each other — miles traveled versus deaths per 100,000 population — yields a pattern that looks like a plateau followed by a steep drop. It evokes the theory of punctuated equilibrium, proposed by the paleontologists Stephen Jay Gould and Niles Eldredge, which suggests that instead of continuous gradual evolution, change occurs abruptly after periods of virtual standstill.

"You see fatalities drop after a breakthrough in new technologies or behaviors, and then plateau until the next one," said David L. Strickland, administrator of the National Highway Traffic Safety Administration. "It takes time for new safety technologies to work their way into the whole fleet of cars on the road."

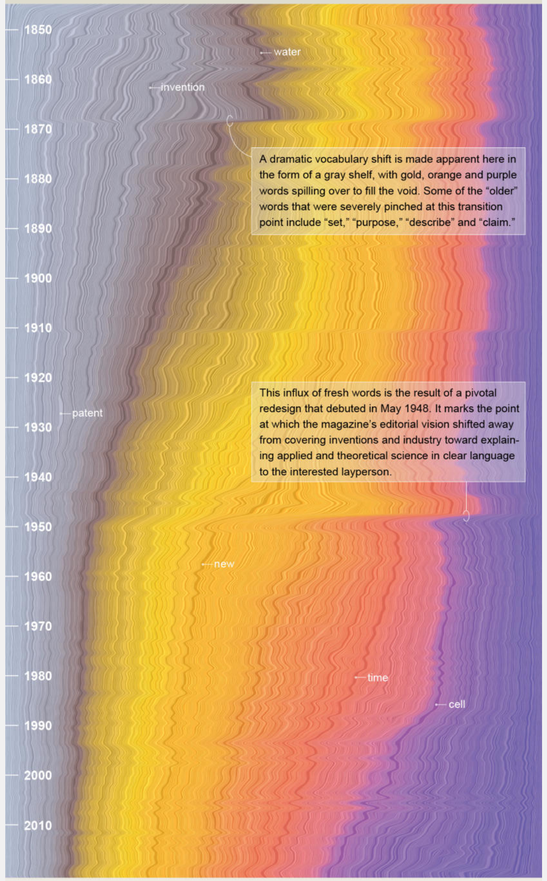
The New York Times Published: September 17, 2012 By HANNAH FAIRFIELD

Sources: National Highway Traffic Safety Administration; Federal Highway Administration

"Driving Safety, in Fits and Starts" by Hannah Fairfield (New York Times)



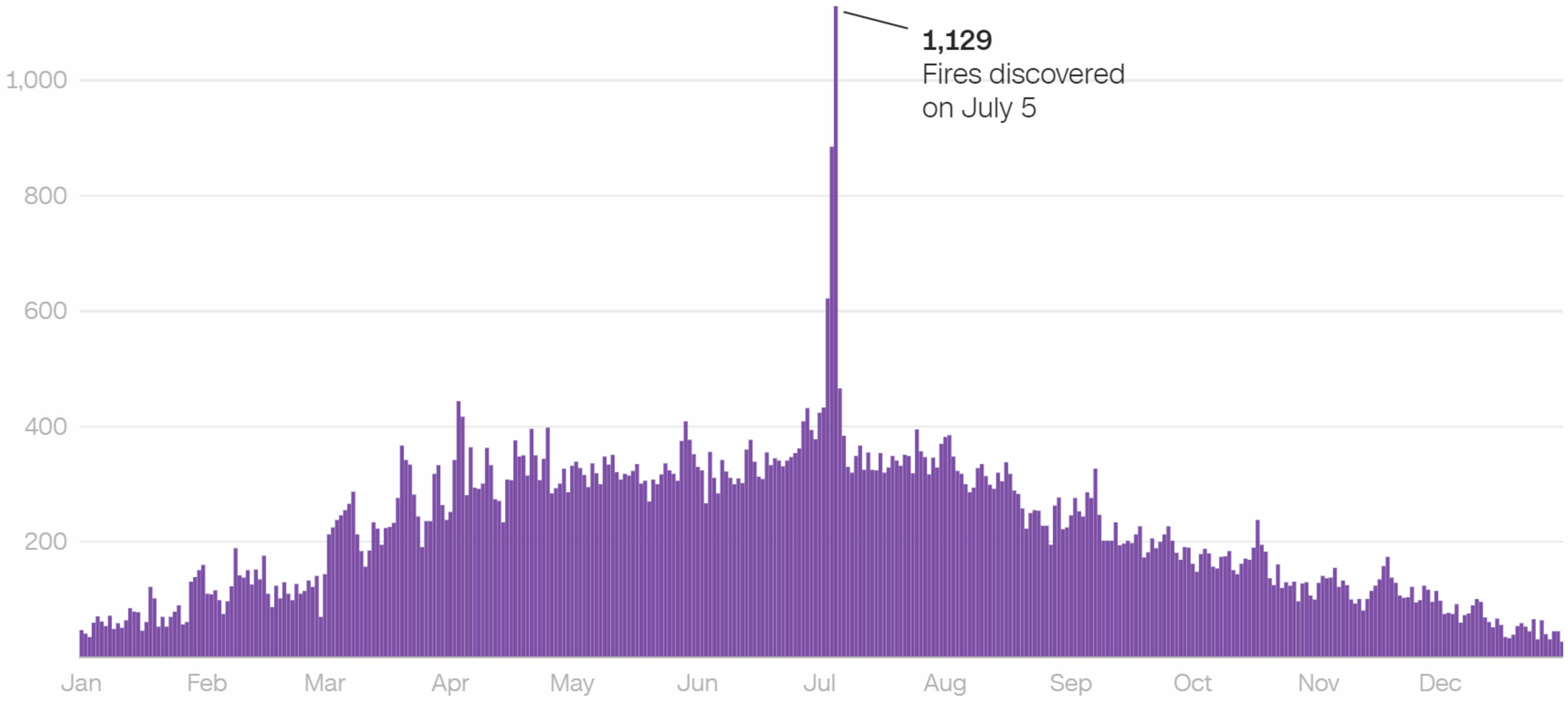
The most popular words used in the pages of *Scientific American* are displayed here by frequency, from 1845 (*top*) through 2020 (*bottom*). Before visualizing the full corpus of our archives, we culled words shorter than three letters, numbers and so-called stop words such as "then" and "or." The remaining top 1,000 words were gathered for each of the 175 years and merged across the years for a total of 4,420 prevailing words. Each layer represents one word, and the thickness of the layer corresponds to the fraction of text occupied by that word, by year. The color and horizontal position of each layer are based on the year in which the respective word's relative frequency peaked: Words routinely used in the early days of the magazine (*gray*) slowly give way to words used more often in recent years (*purple*). (The range of brightness of neighboring layers alternates for improved legibility.) The jarring visual effect of those horizontal stripes signals sudden changes in vocabulary. The three annotation bubbles here offer some historical context for both rapid shifts and consistent periods. —*J.C.*



The word landscape remained relatively steady during the postredesign balance of editor in chief Dennis Flanagan's tenure (1947–1984). Four decades of smaller but distinct ripples followed, perhaps reflecting traces of changing editorial leadership (Jonathan Piel, 1984; John Rennie, 1994; and Mariette DiChristina, 2009) and likely in part resulting from the impact of the Internet on legacy-print publications.

"The Language of Science" by Moritz Stefaner, Lorraine Daston, Jen Christiansen

Total wildfires discovered each day of the year since 2014



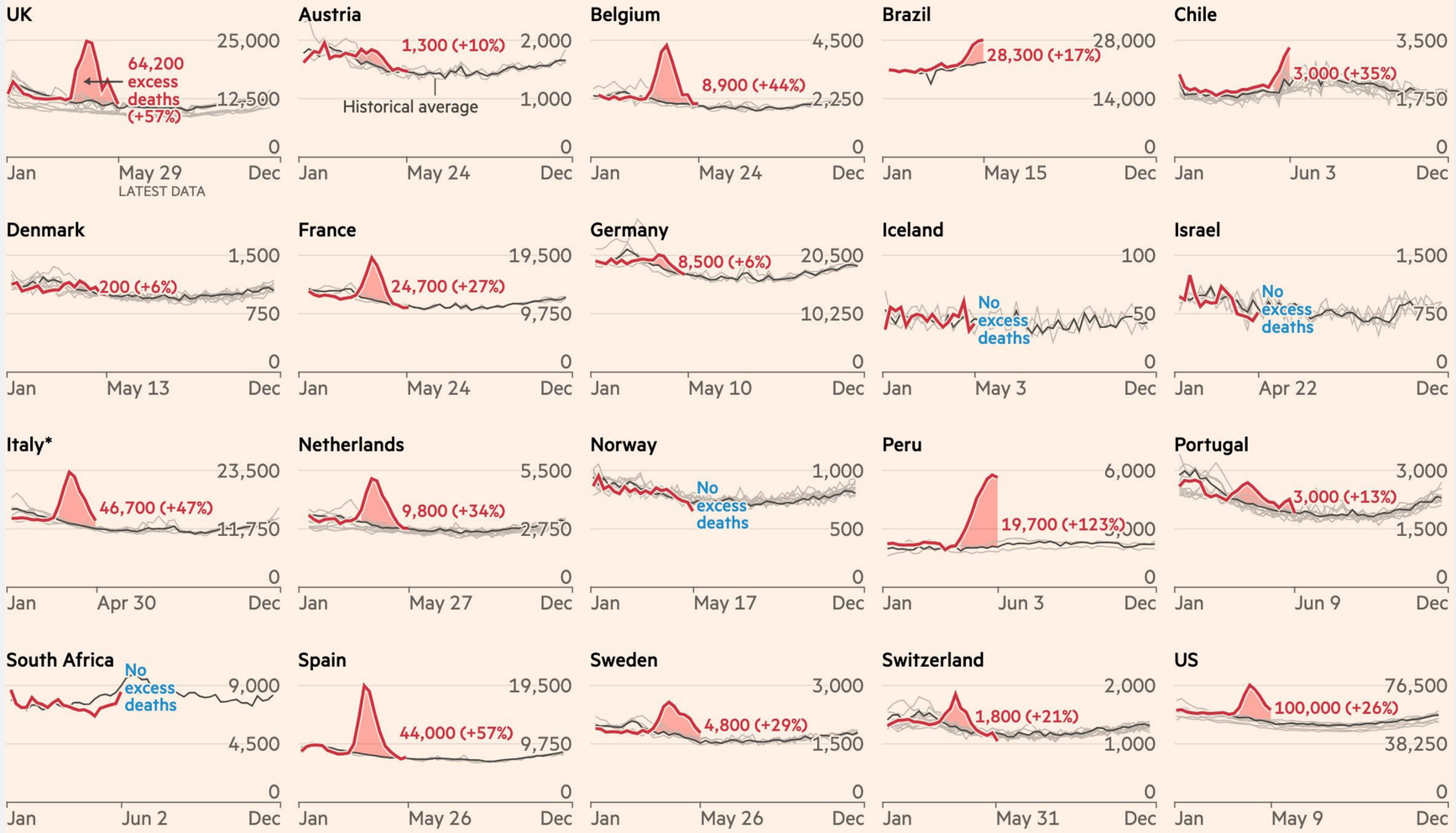
Human-caused fires, excluding prescribed fires. 2022 fires included through June 30. All incident times Eastern.

Sources: CNN analysis of data from the National Interagency Fire Center
Graphic: John Keefe, CNN

"Western cities get creative after megadrought leads some to cancel firework displays" by John Keefe (CNN)

Death rates have climbed far above historical averages in many countries that have faced Covid-19 outbreaks

Number of deaths per week from all causes, 2020 vs recent years: Shading indicates total excess deaths during outbreak



*Italian figures may not exactly match the source data as they were scraped from a PDF
Source: FT analysis of mortality data. Data updated June 10
FT graphic: John Burn-Murdoch / @jburnmurdoch

Excess Death Trajectories by John Burn-Murdoch (Financial Times)

"Reservoirs are drying up as consequences of the Western drought worsen" by Zach Levitt (Washington Post)

Wrap~Up

Design for your audience.

- Choose charts based on your goal not tradition or novelity (only).
- Make sure your visualizations are accesible for everyone (colors, readability, ALT text).
- Use visual contrast to highlight important information.
- Provide meaningful labels.

Design for your audience.

- Choose charts based on your goal not tradition or novelity (only).
- Make sure your visualizations are accesible for everyone (colors, readability, ALT text).
- Use visual contrast to highlight important information.
- Provide meaningful labels.

Be honest.

- Check your data carefully, via summary statistics and visualizations.
- Show the distribution of the raw data if possible.
- Don’t truncate bar charts, add spacing to truncated axes.

Design for your audience.

- Choose charts based on your goal not tradition or novelity (only).
- Make sure your visualizations are accesible for everyone (colors, readability, ALT text).
- Use visual contrast to highlight important information.
- Provide meaningful labels.

Be honest.

- Check your data carefully, via summary statistics and visualizations.
- Show the distribution of the raw data if possible.
- Don’t truncate bar charts, add spacing to truncated axes.

Lend a helping hand.

- Use annotations and direct labels instead / in addition to captions and legends.
- Order your data, either by value or intrinsic ranking.
- Focus on the main message and reduce data complexity.
- Reveal information step by step (if applicable).

Thank you!

cedricscherer.com @CedScherer z3tt