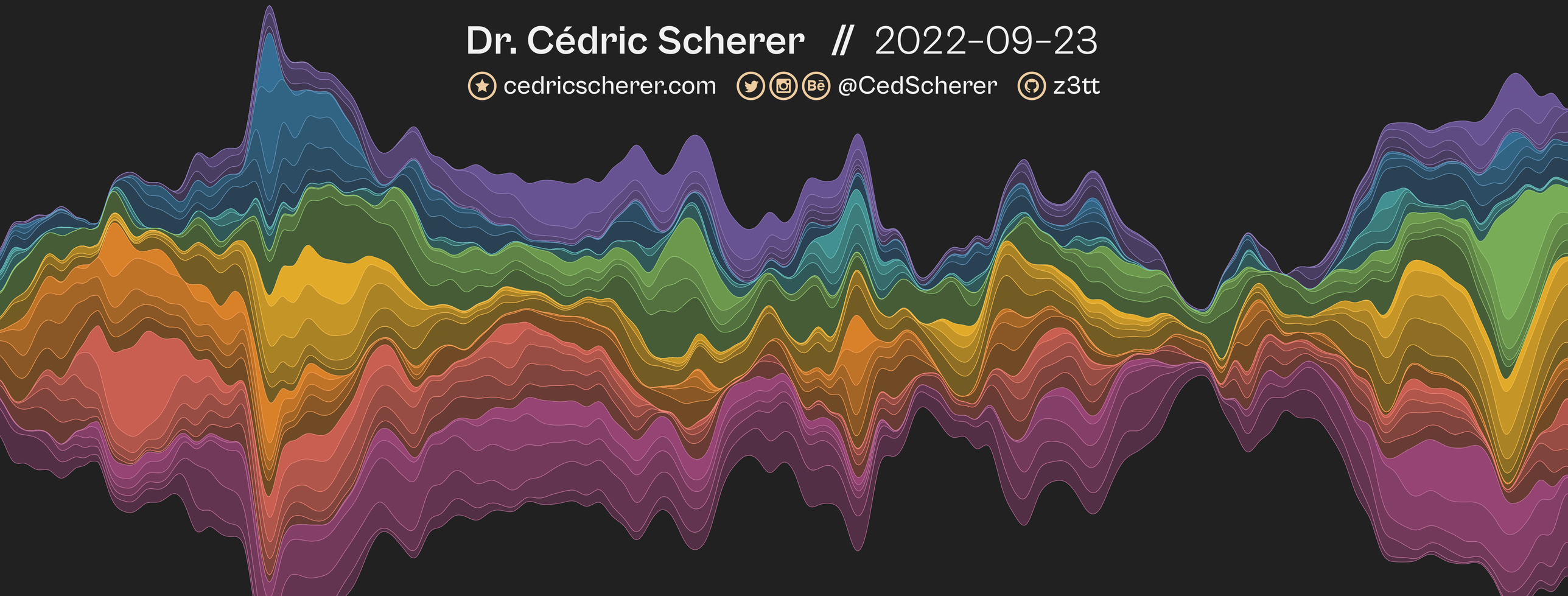
Effective Data Visualization

Design graphics that tell stories in an engaging way

Dr. Cédric Scherer // 2022-09-23

cedricscherer.com @CedScherer z3tt



Cédric Scherer

Independent Data Visualization Designer
Computational Ecologist at IZW Berlin

Consulting | Coaching | Coding

cedricscherer.com @CedScherer z3tt

Slides

cedricscherer.com/workshops/BSIO_DataViz_Slides.pdf

Data Visualization

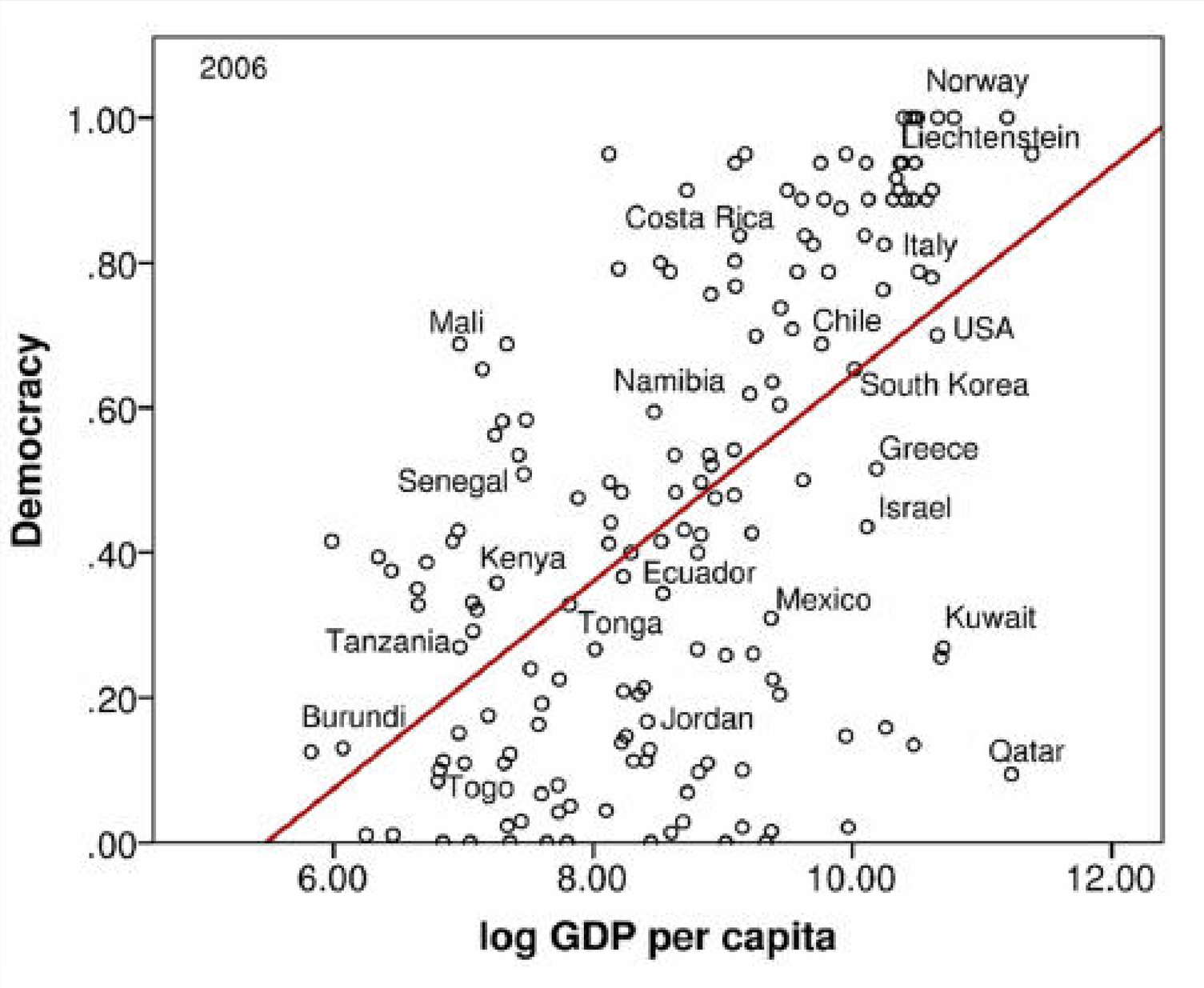
is any graphical representation
of information and data.

Data Visualization

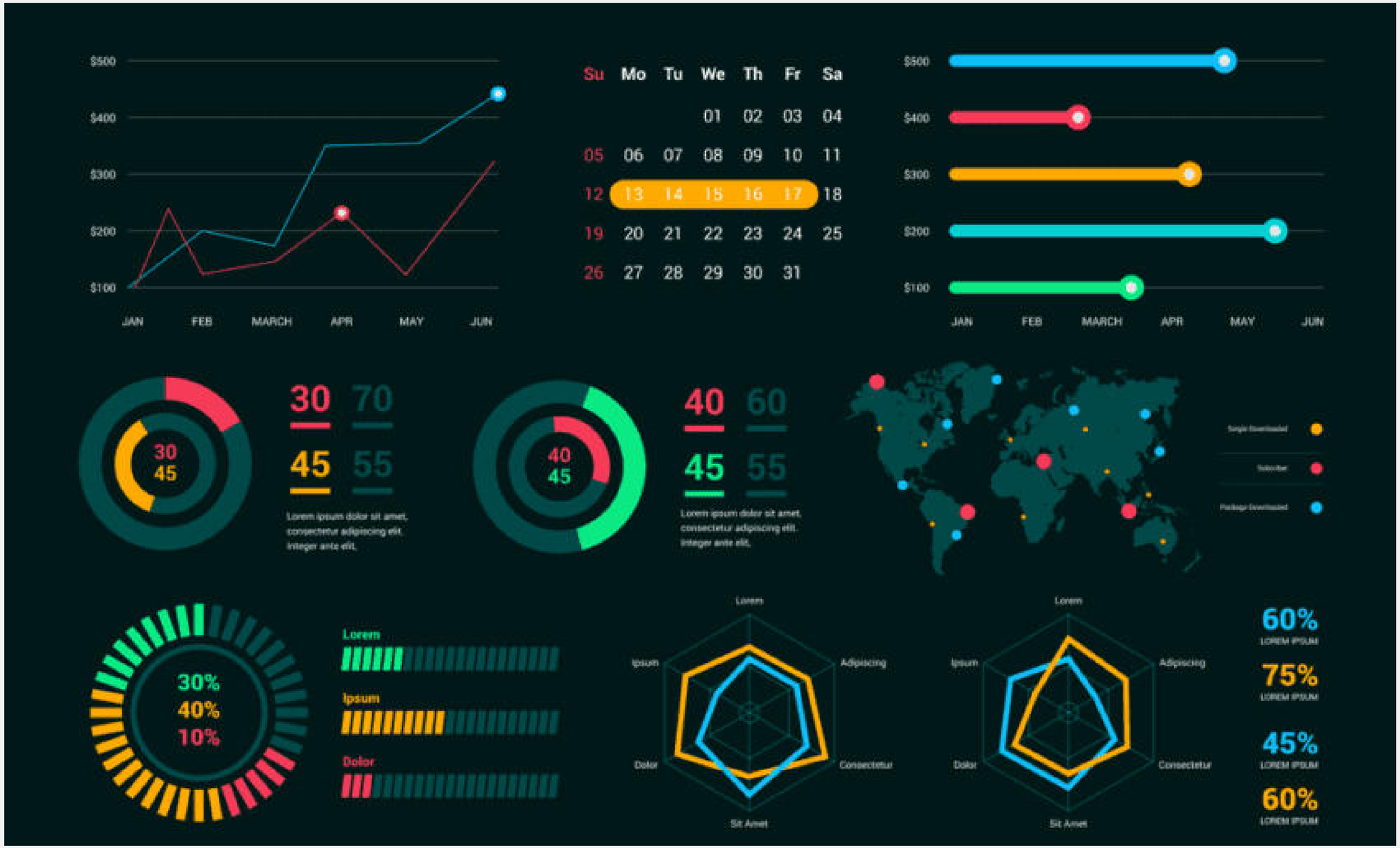
helps to amplify cognition, gain insights, discover, explain, and make decisions.

Data Visualization

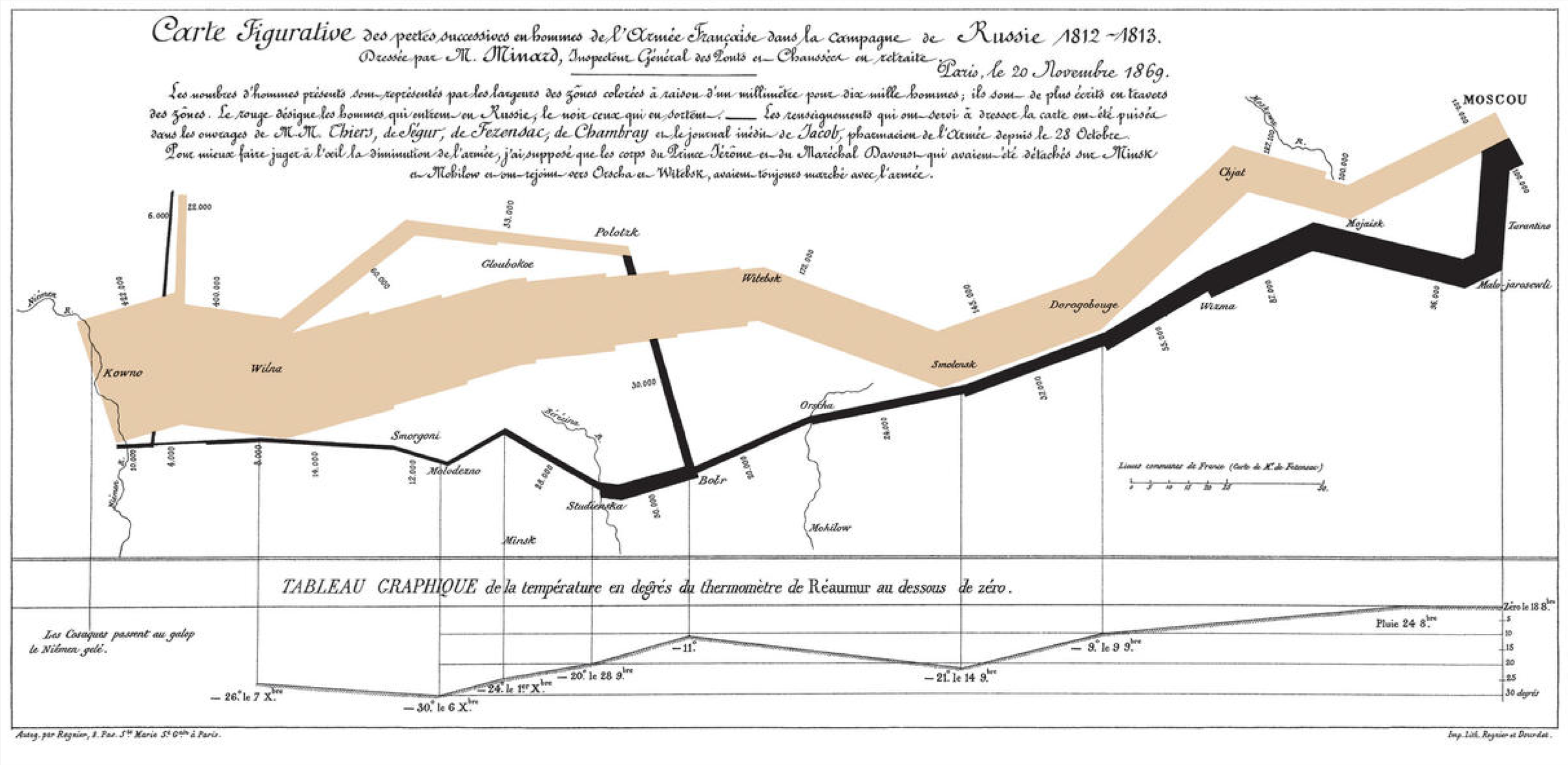
converts information into visual forms as quantifiable features.



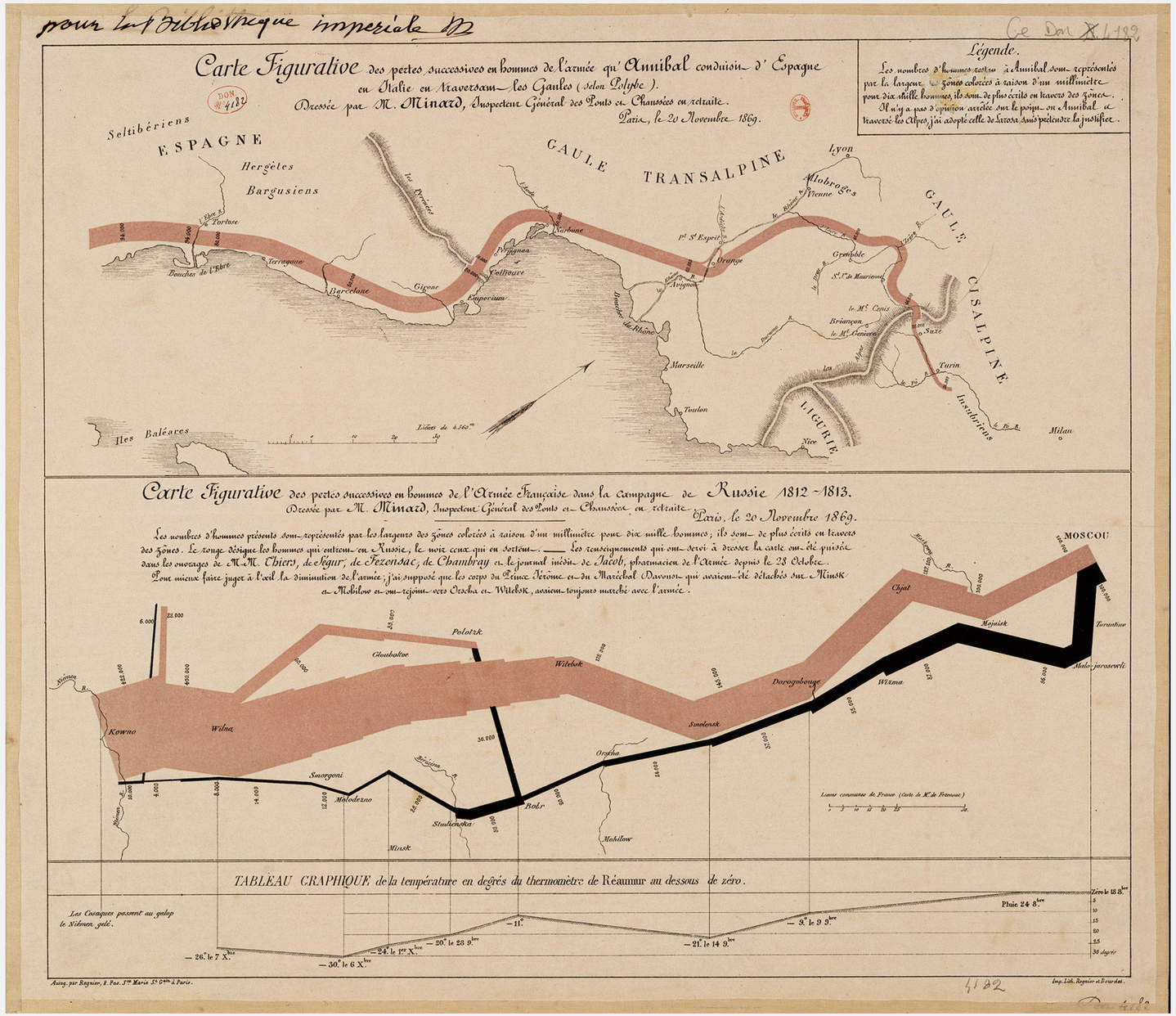
Source: Ranganathan et al. 2014



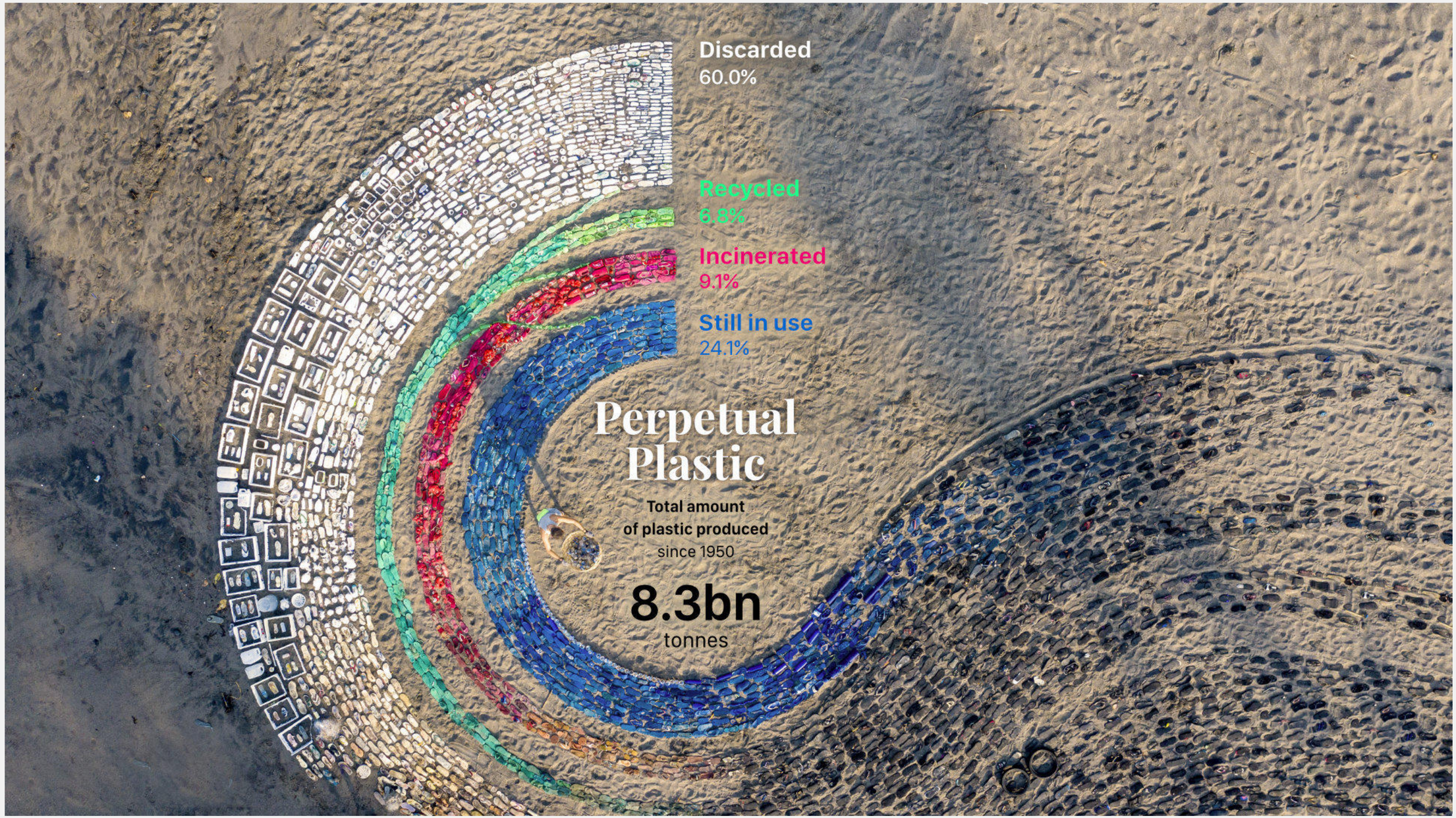
Source: datameer.com



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

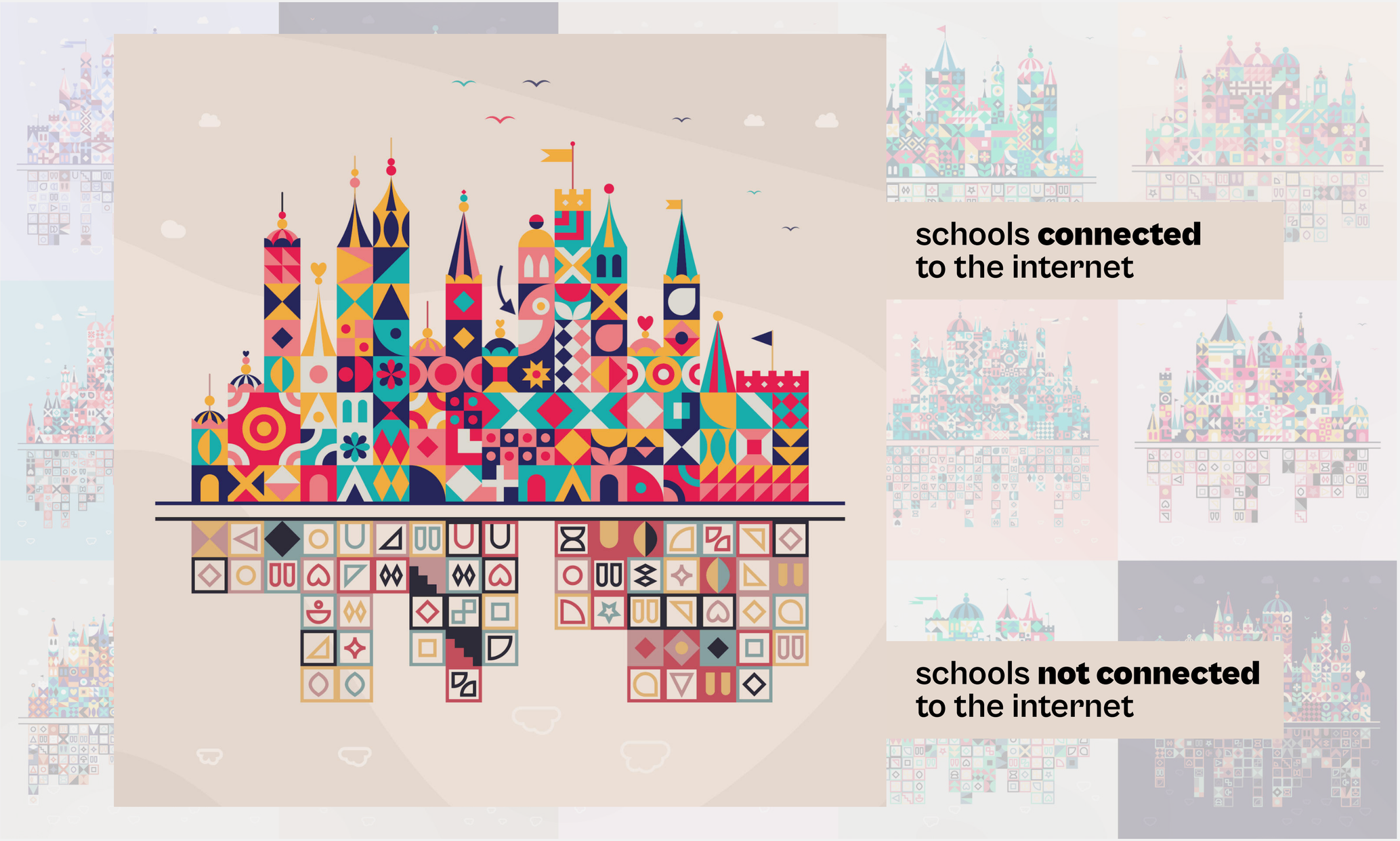


Carte figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812–1813 and Carte figurative des pertes successives en hommes de l'Armée qu'Annibal conduisit d'Espagne en Italie en traversant les Gaules (selon Polybe) by Charles Joseph Minard



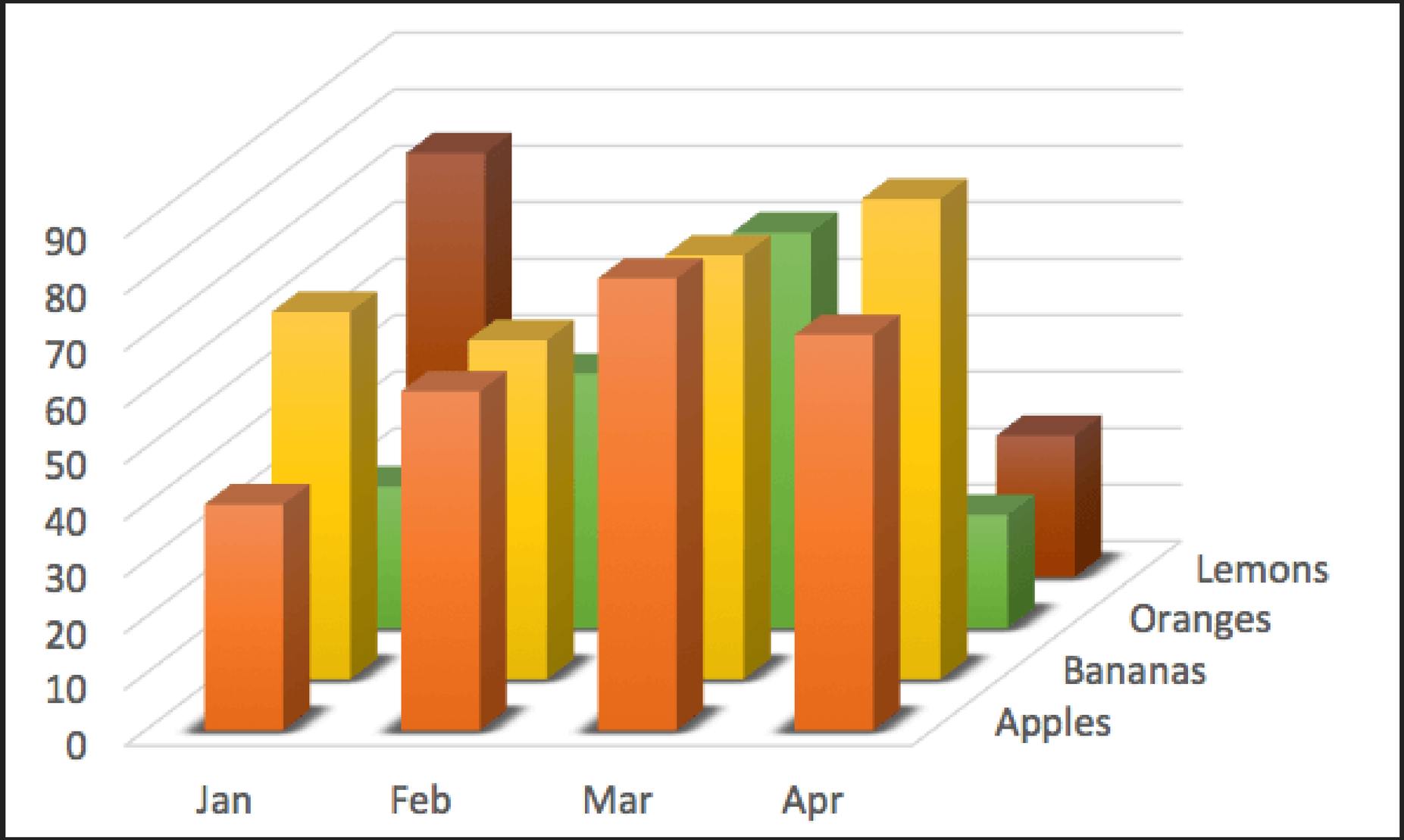
Source: "Perpetual Plastic" by Liina Klauss, Skye Morét and Moritz Stefaner

Source: “Patchwork Kingdoms” by Nadieh Bremer

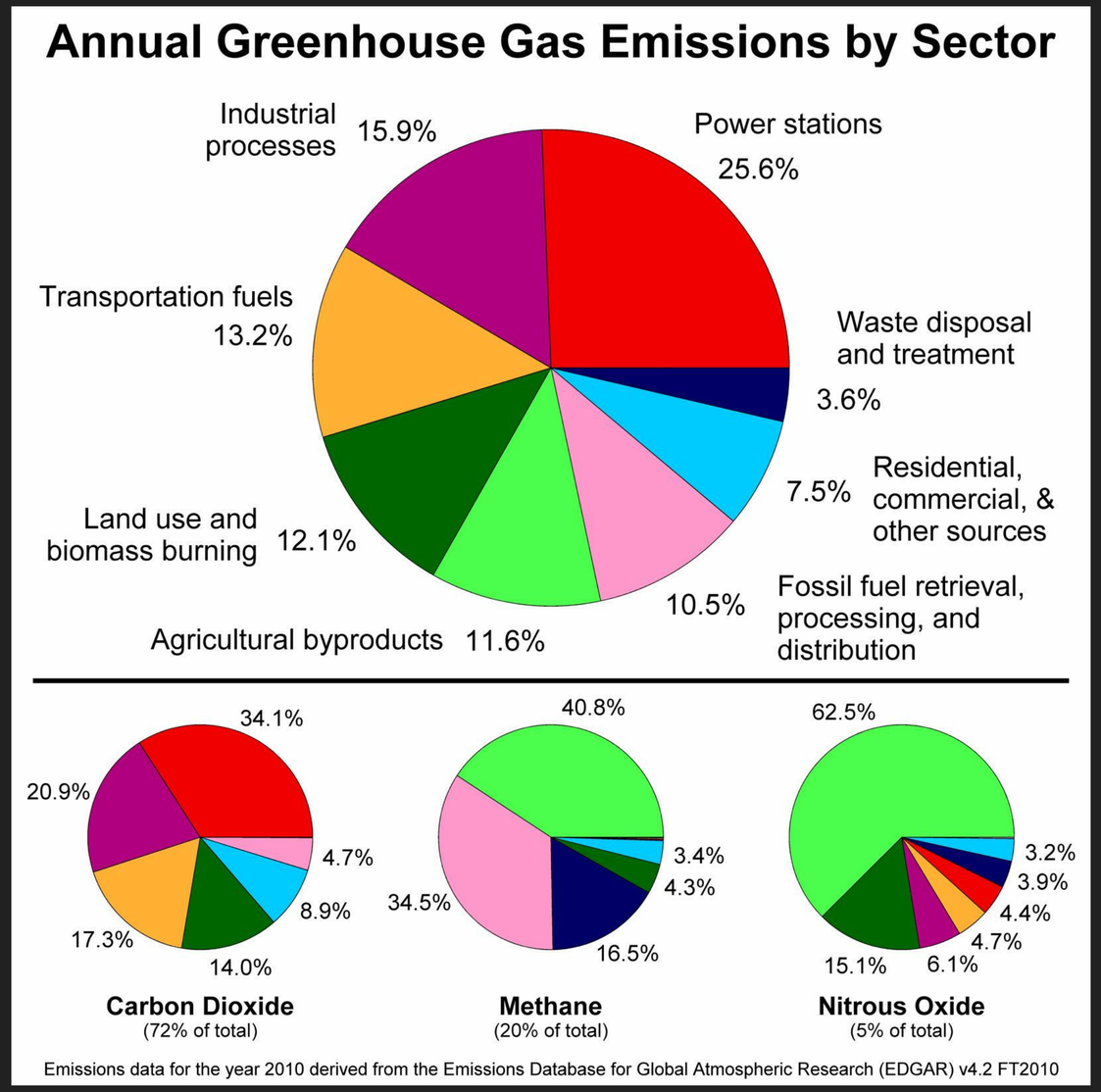


Source: "Patchwork Kingdoms" by Nadieh Bremer

What makes it a bad data visualization?



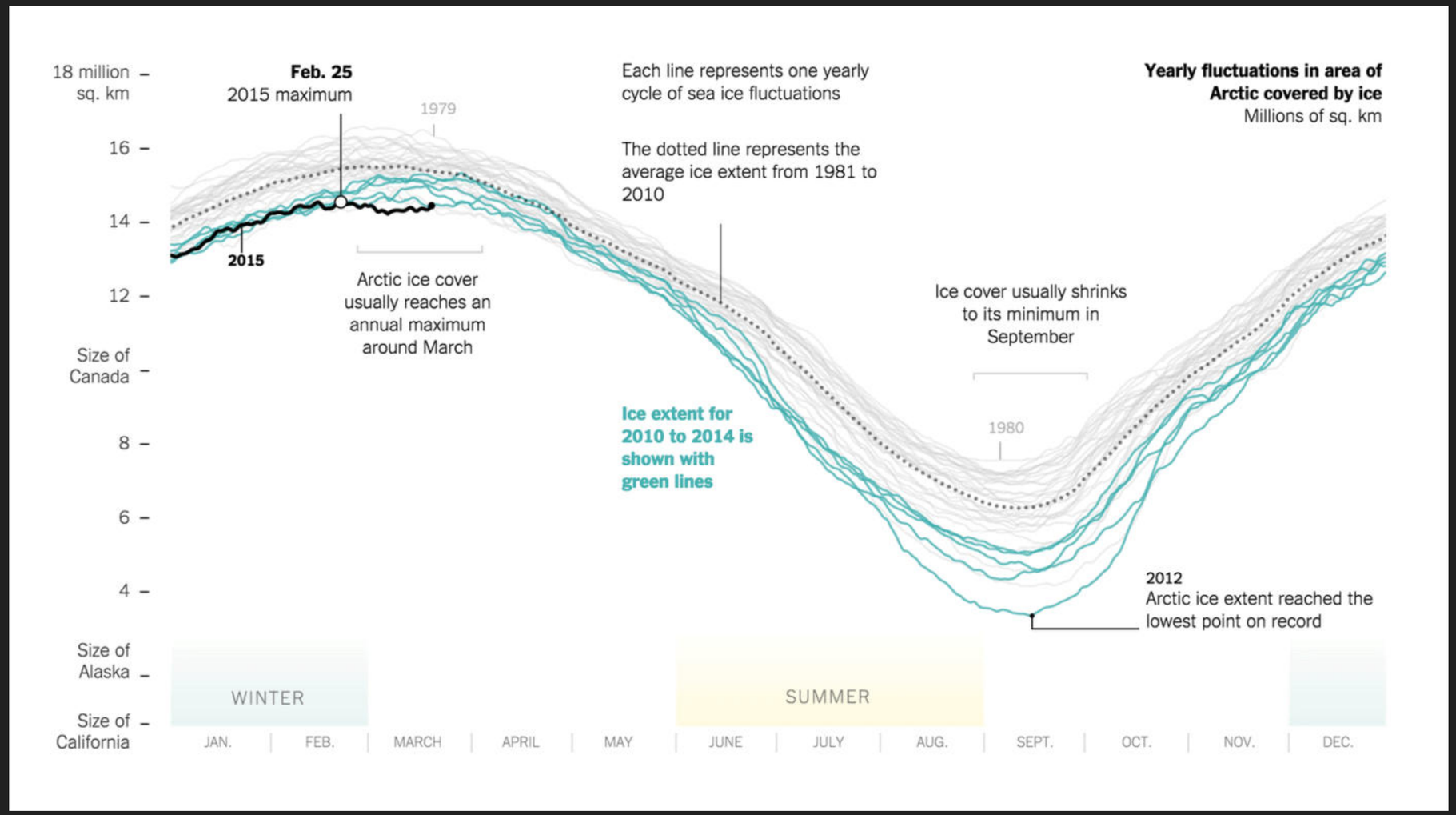
What makes it a bad data visualization?



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)

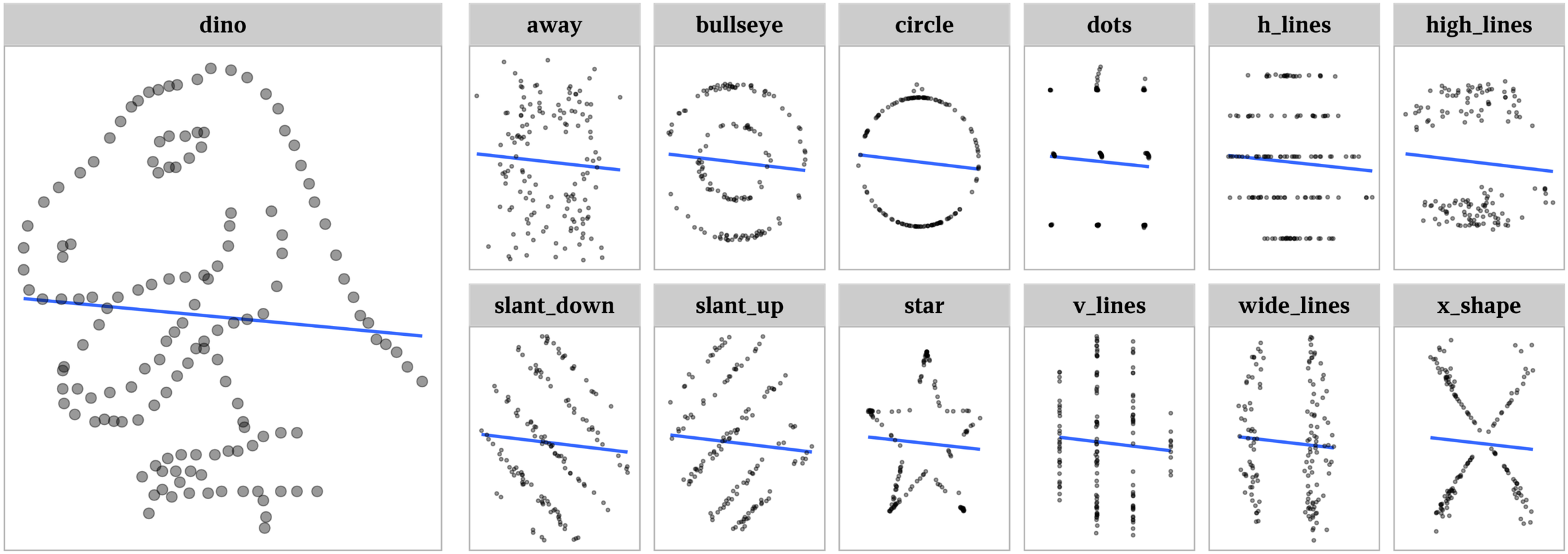
What makes it a good data visualization?

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

INFORMATION

Understand your data and be accurate

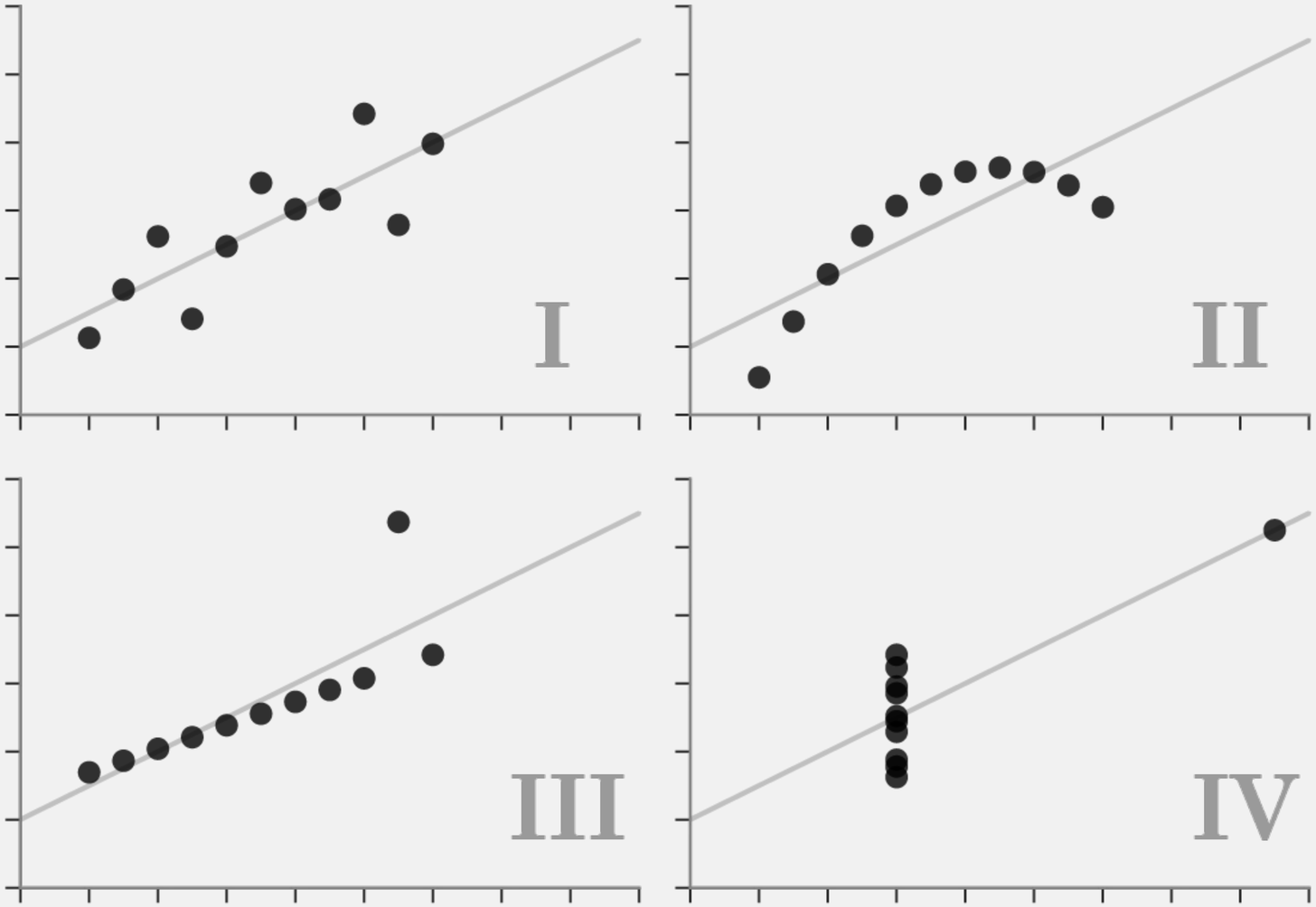
Visualize Your Data



"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

Anscombe's Quartet

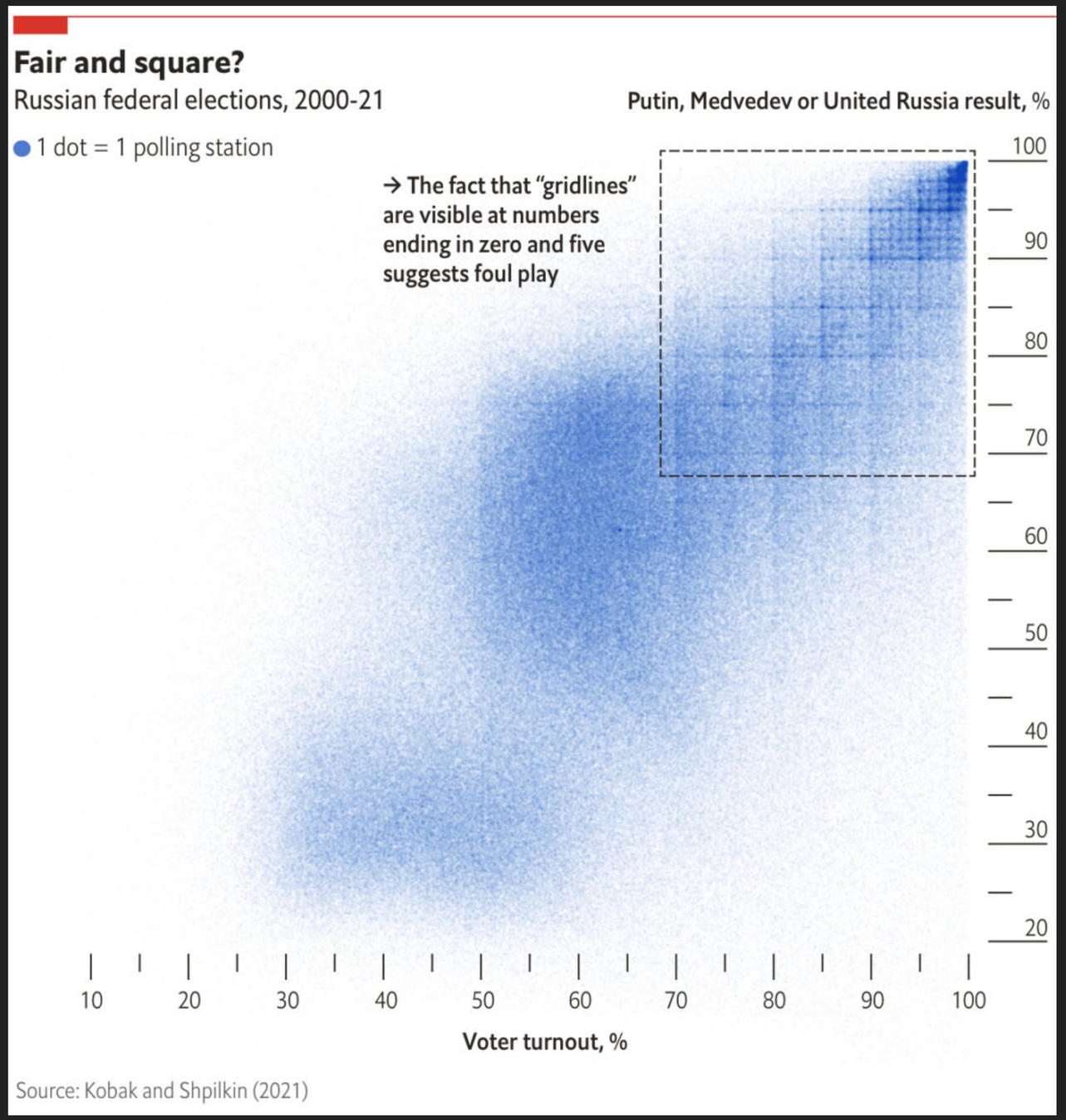
Each dataset has the same summary statistics
mean, standard deviation, and correlation
but are visually distinct.



"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

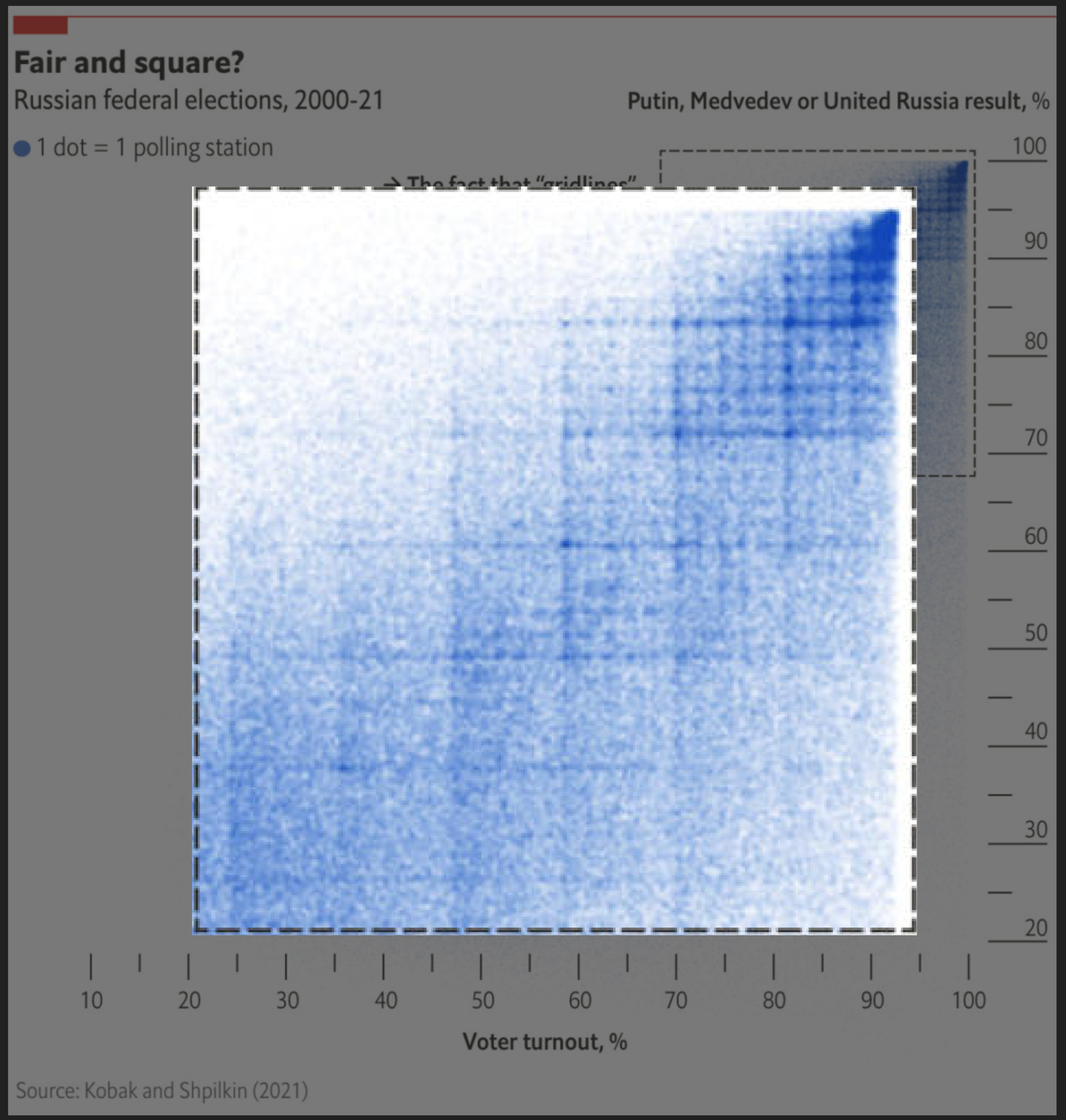
Visualize Your Data

"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**."



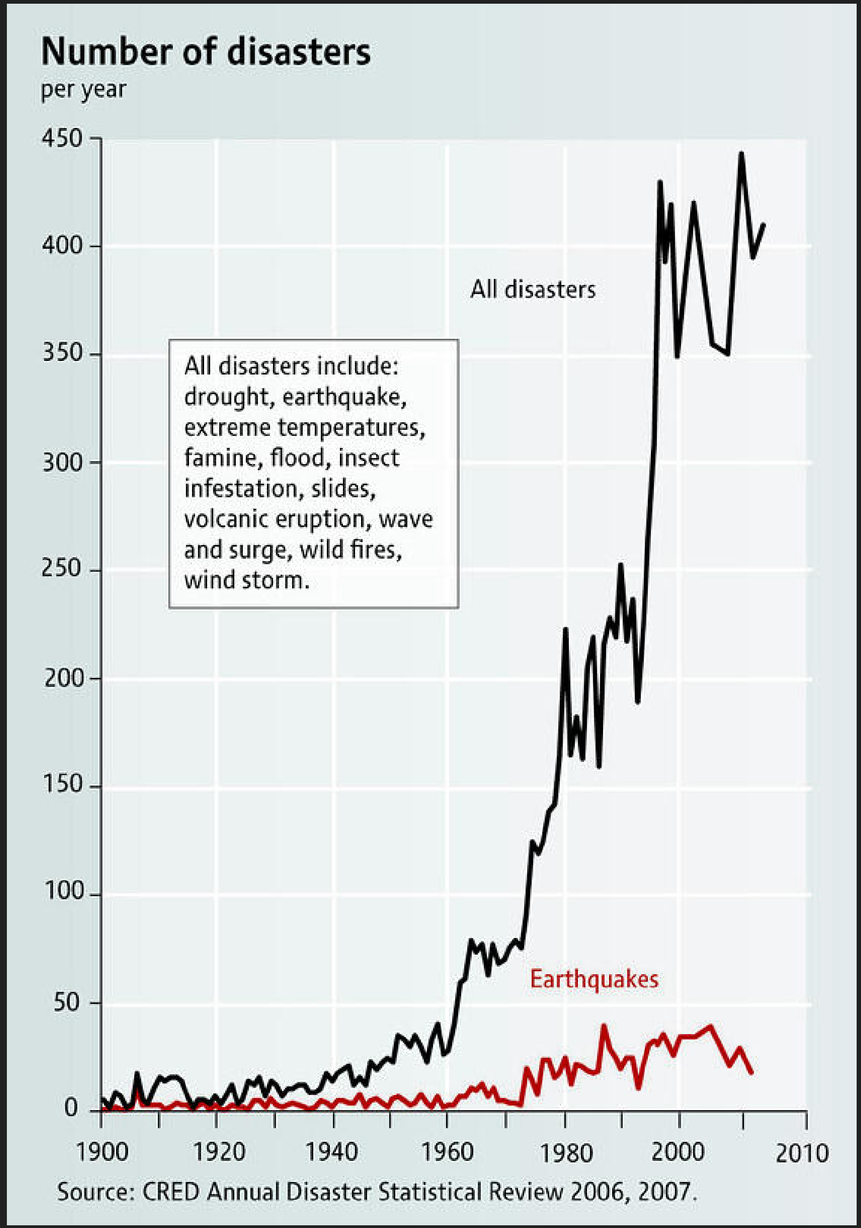
Our data is **never** a perfect reflection of the real world.

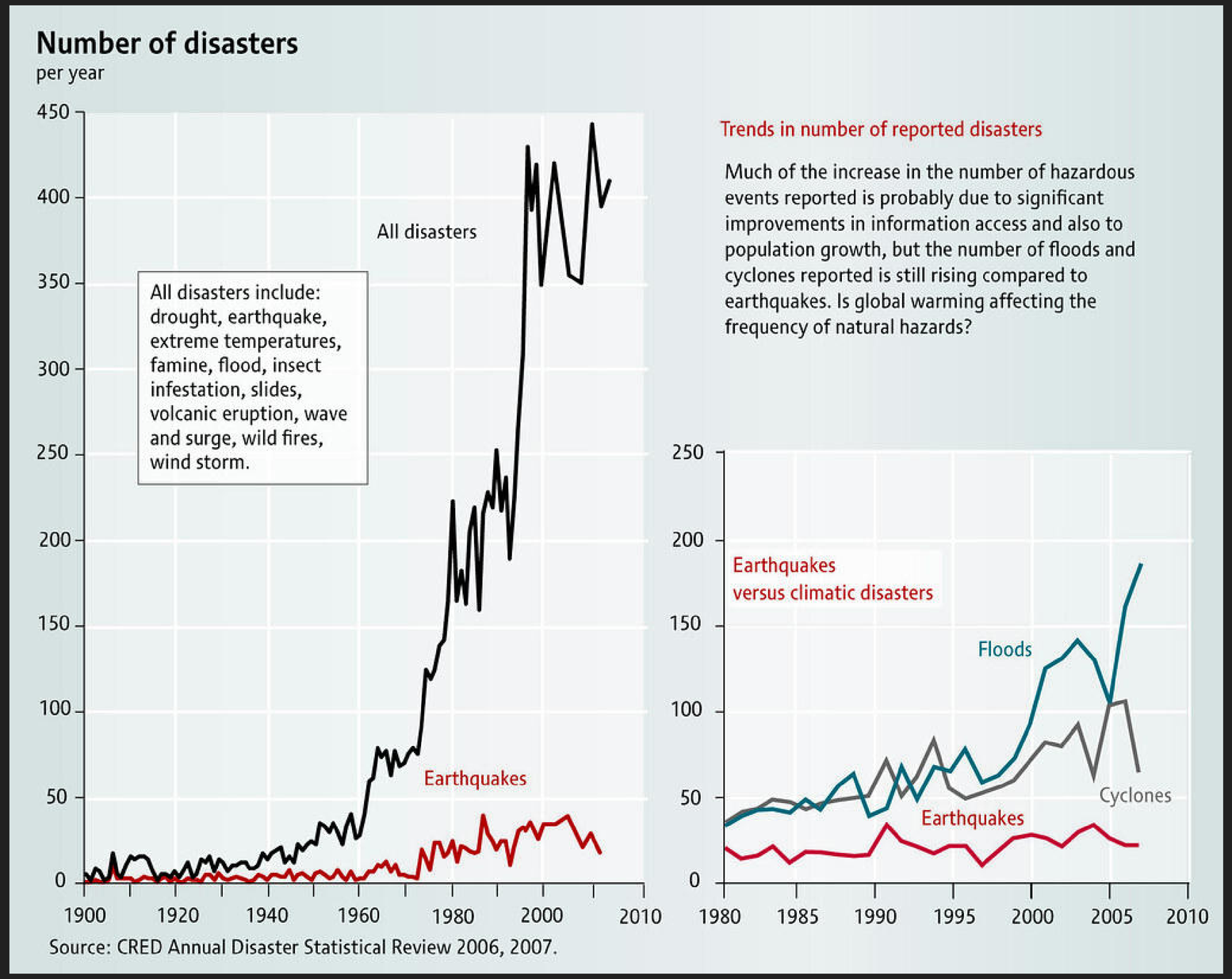
Our data is **never** a perfect reflection of the real world.

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

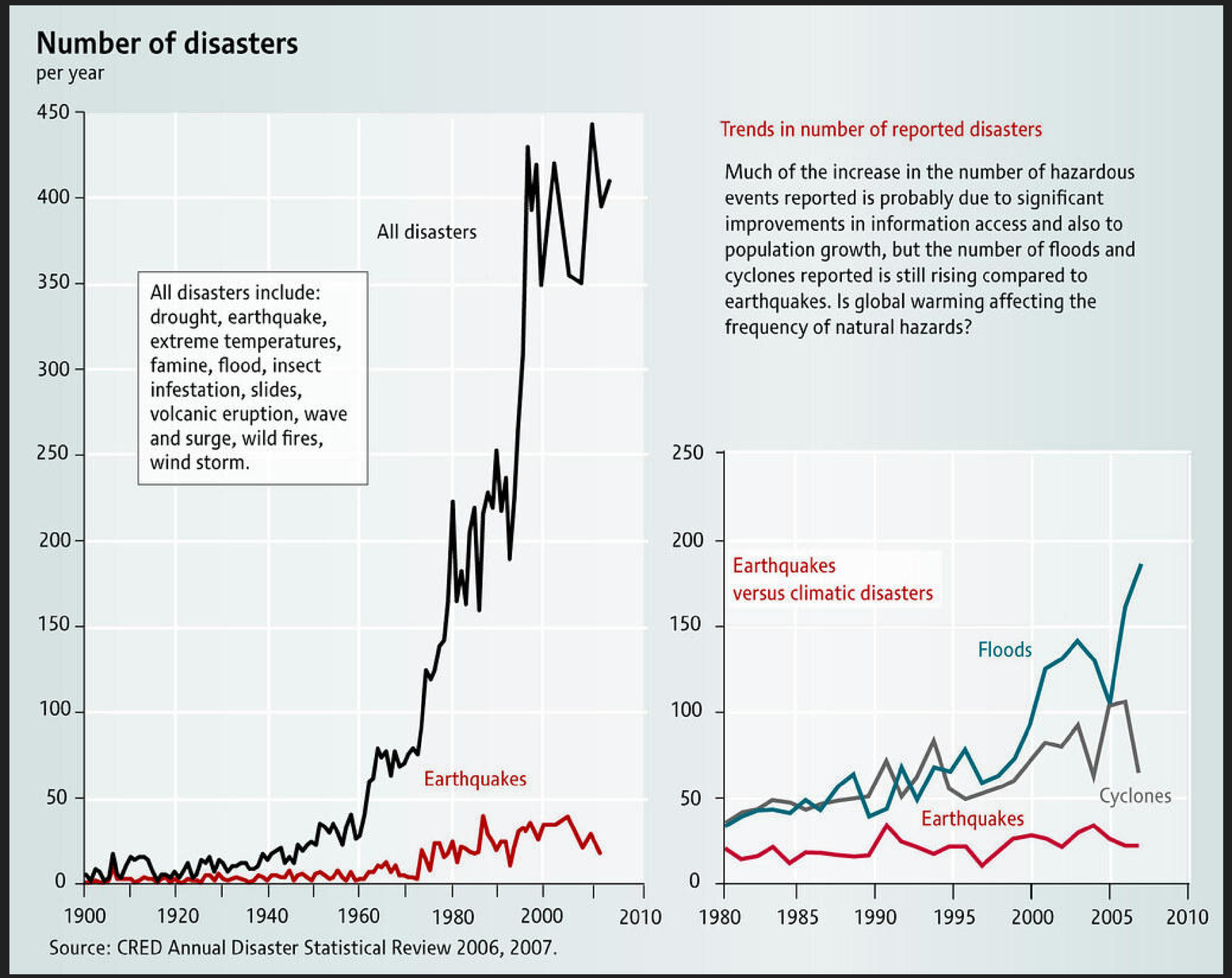
→ **collected by humans:** guesstimation, precision and errors

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


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


Number of disasters
per year
450
400
350
300
250
200
150
100
50
0
1900
1920
1940
1960
1980
2000
2010
All disasters
All disasters include: drought, earthquake, extreme temperatures, famine, flood, insect infestation, slides, volcanic eruption, wave and surge, wild fires, wind storm.
Earthquakes
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?
250
200
150
100
50
0
1980
1985
1990
1995
2000
2005
2010
Earthquakes versus climatic disasters
Floods
Cyclones
Earthquakes
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."

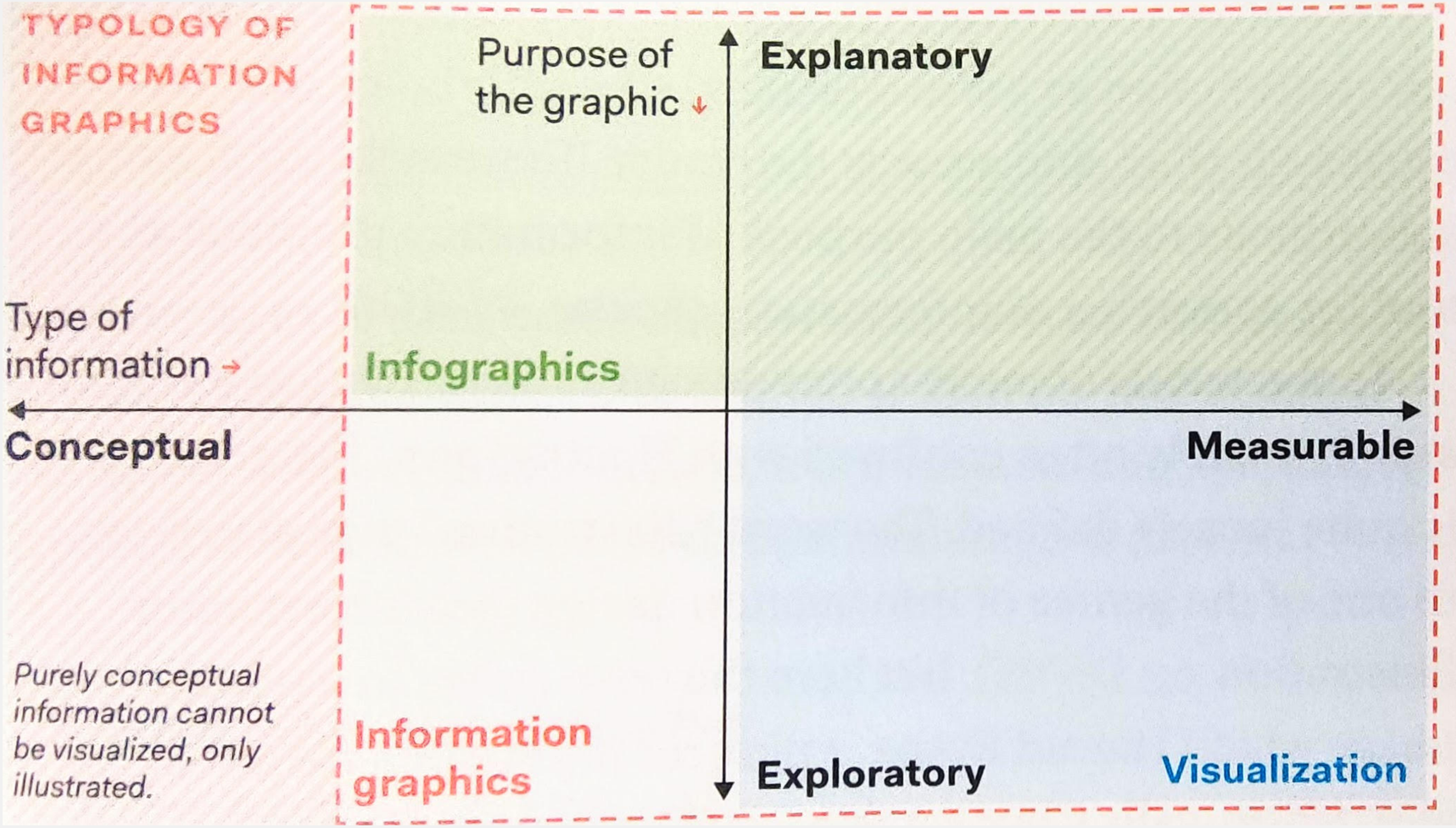
→ **falsifiable universal statement: "All swans are white."**

STORY

Be clear about the message of your graphic

Typology of Information Graphics

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



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

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

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

Moritz Stefaner

”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.”

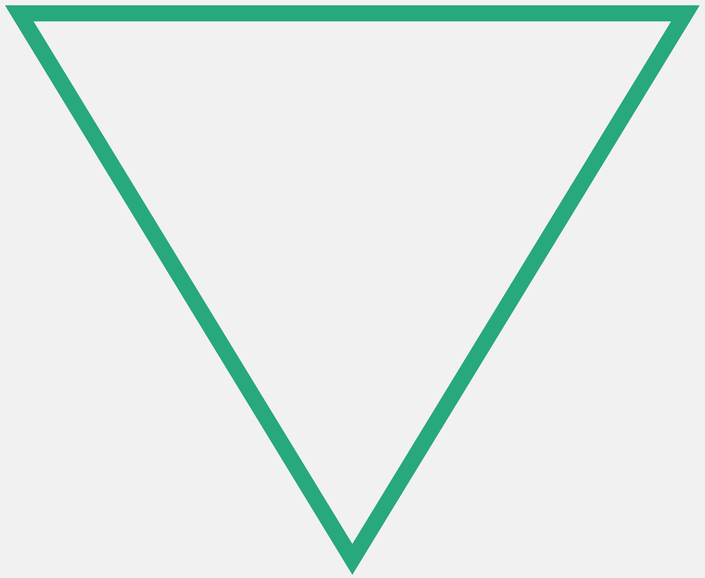
Alberto Cairo

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

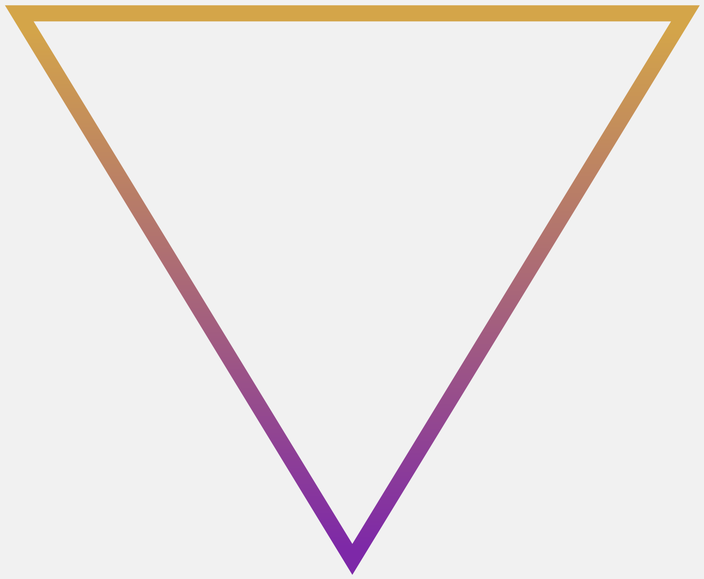
Aim:
functional

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?

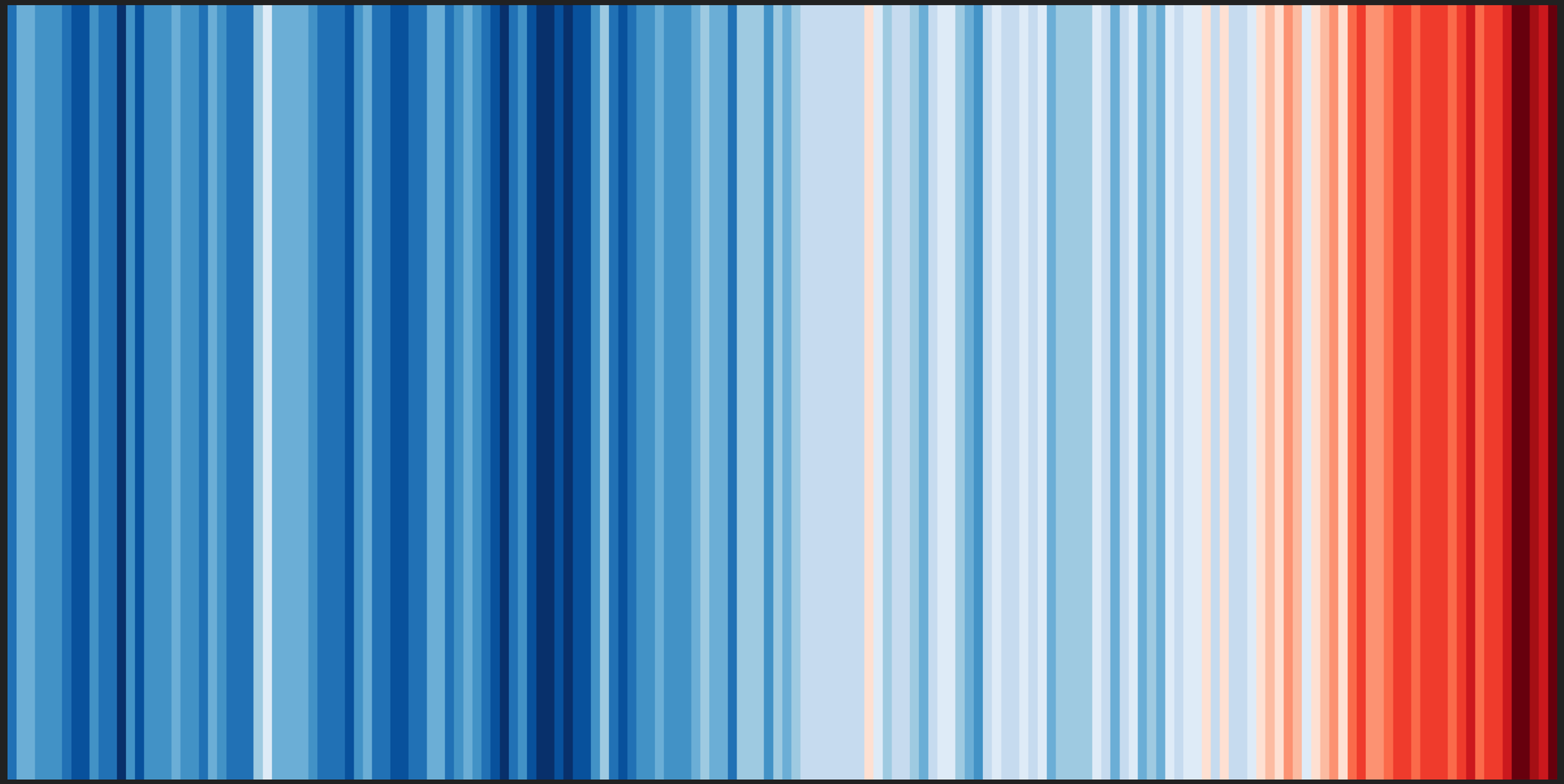
- Which story is **interesting** for them?
- Which variables are **meaningful** to them?
- What are **relevant** details to include?

Who is my audience?

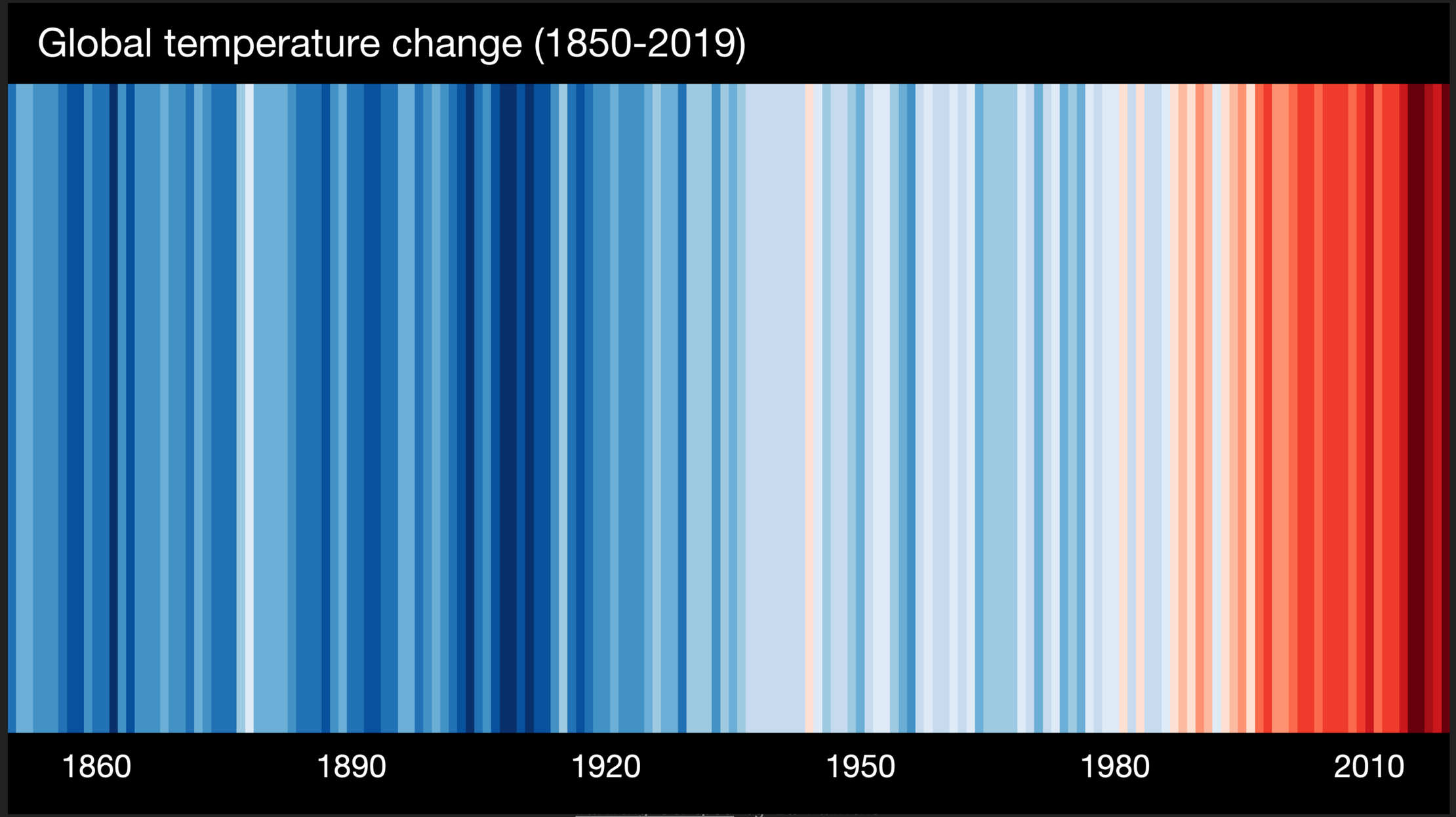
- Which story is **interesting** for them?
- Which variables are **meaningful** to them?
- What are **relevant** details to include?
- How will they **encounter** the visualization?

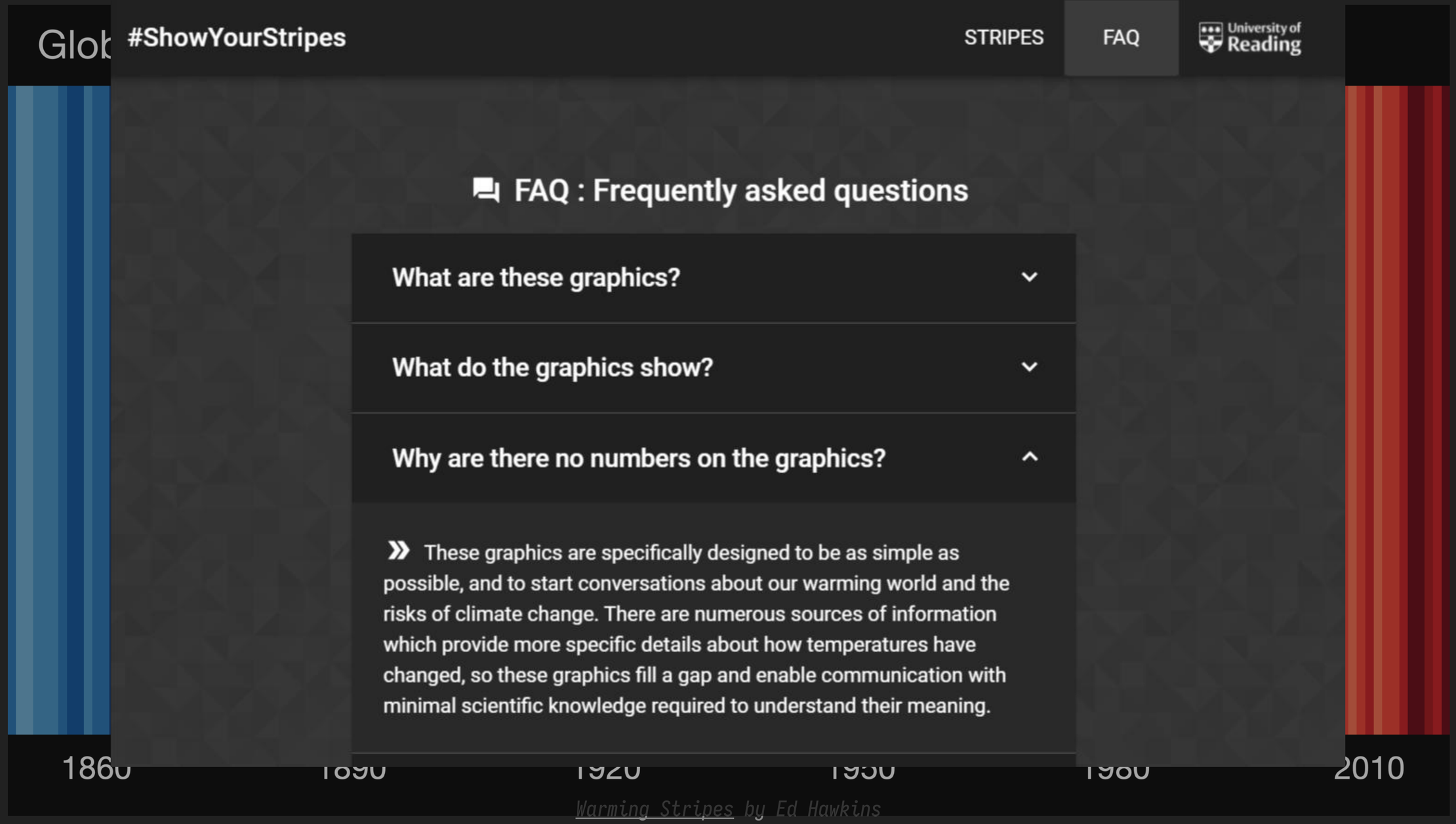
Who is my audience?

- 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?**

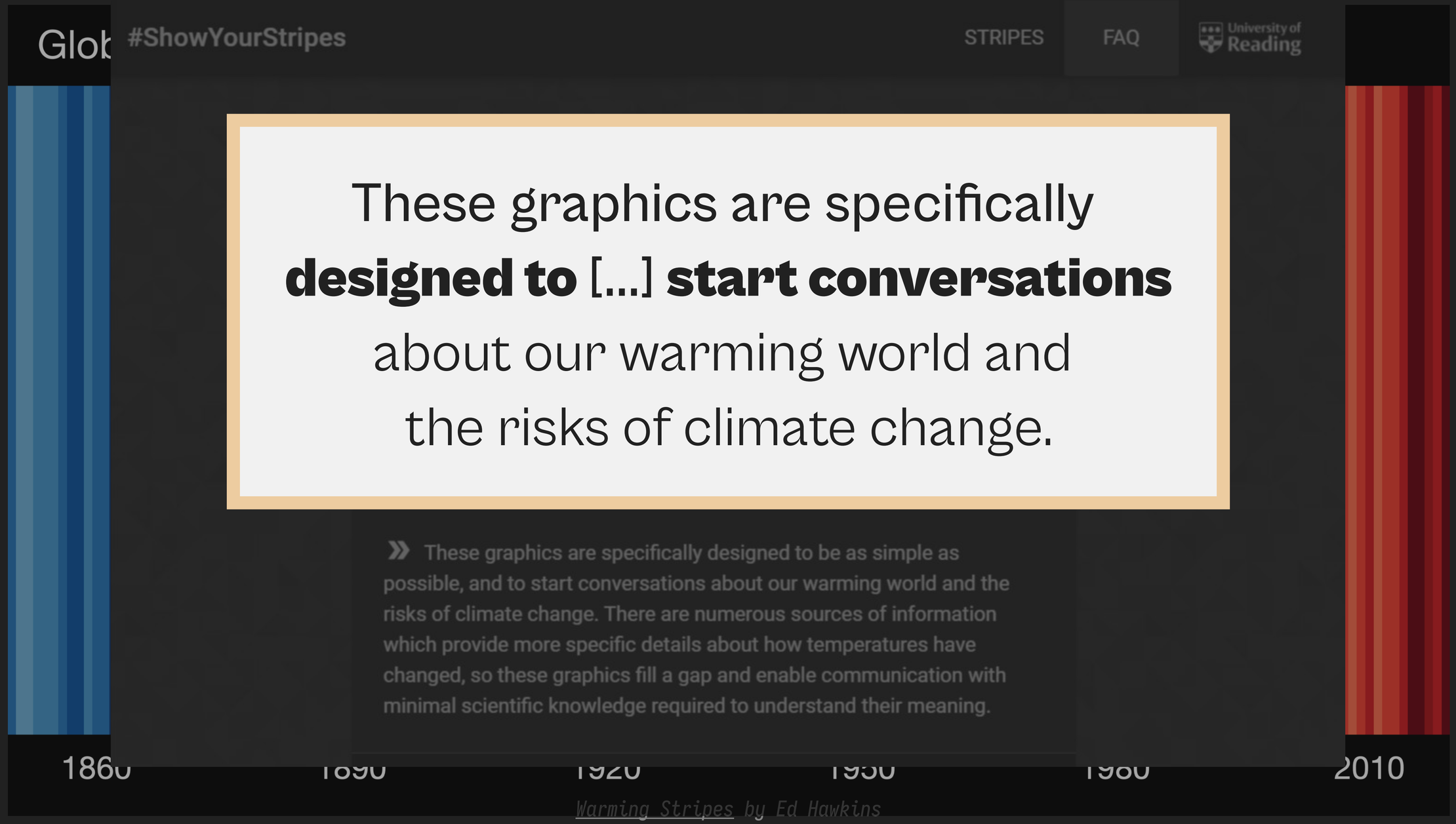


Warming Stripes by Ed Hawkins


Global temperature change (1850-2019)
1860
1890
1920
1950
1980
2010

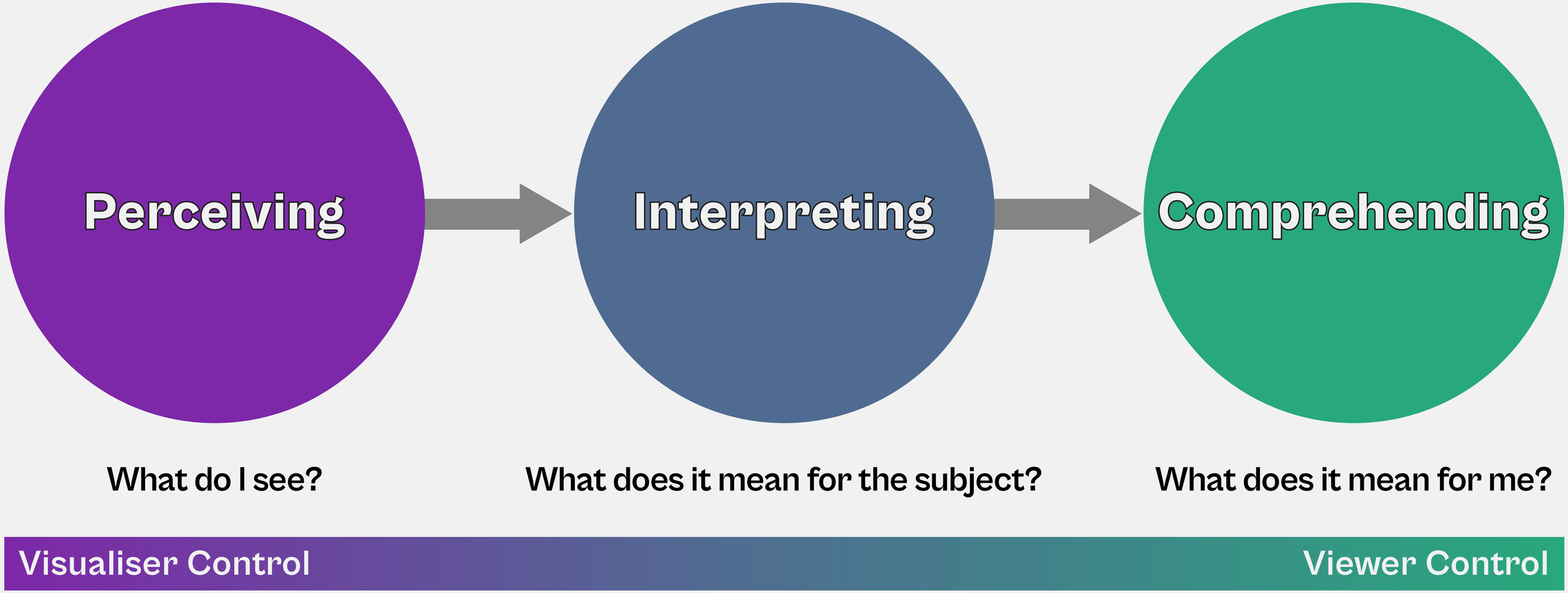


Warming Stripes by Ed Hawkins



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

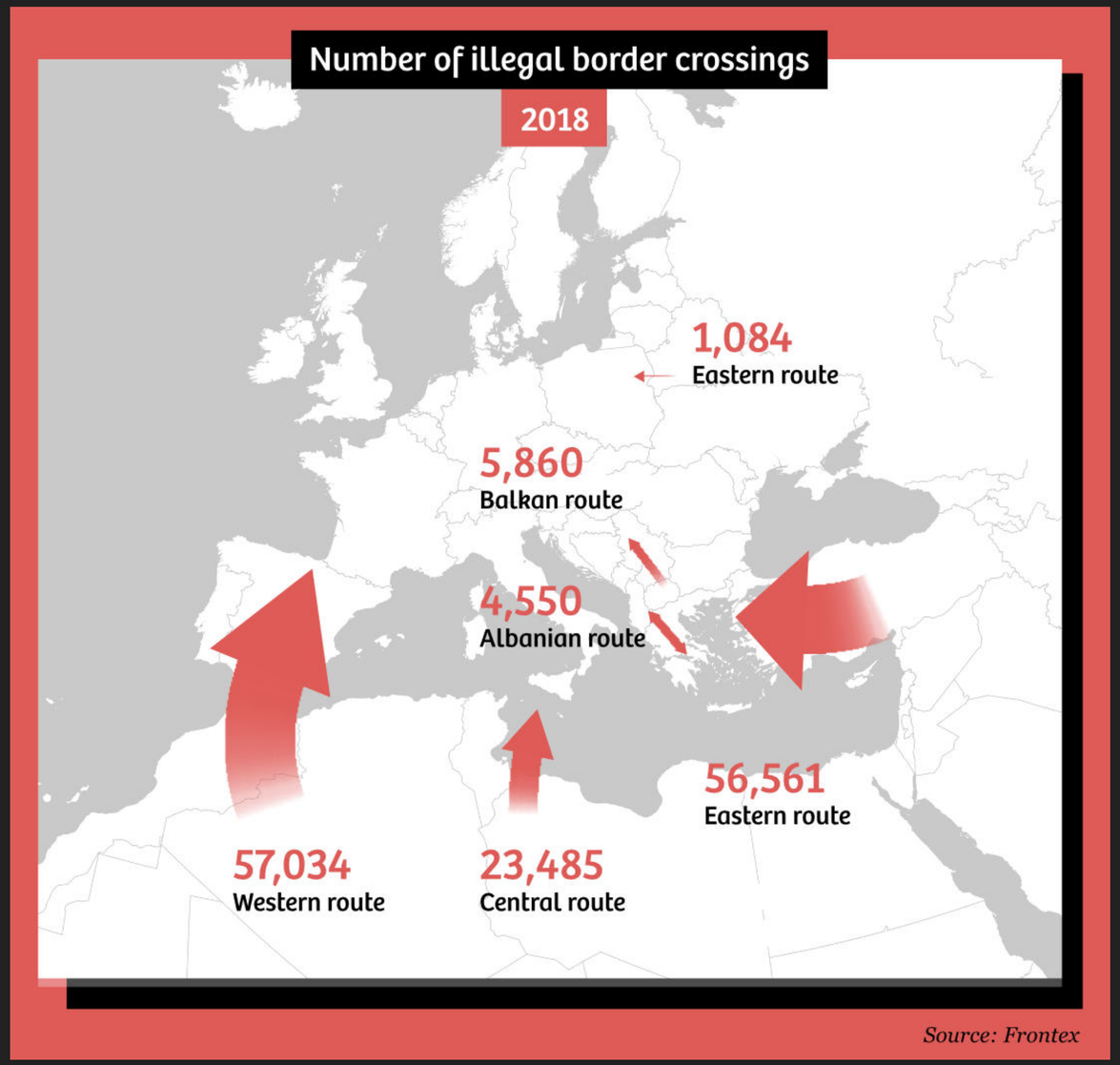
Warming Stripes by Ed Hawkins



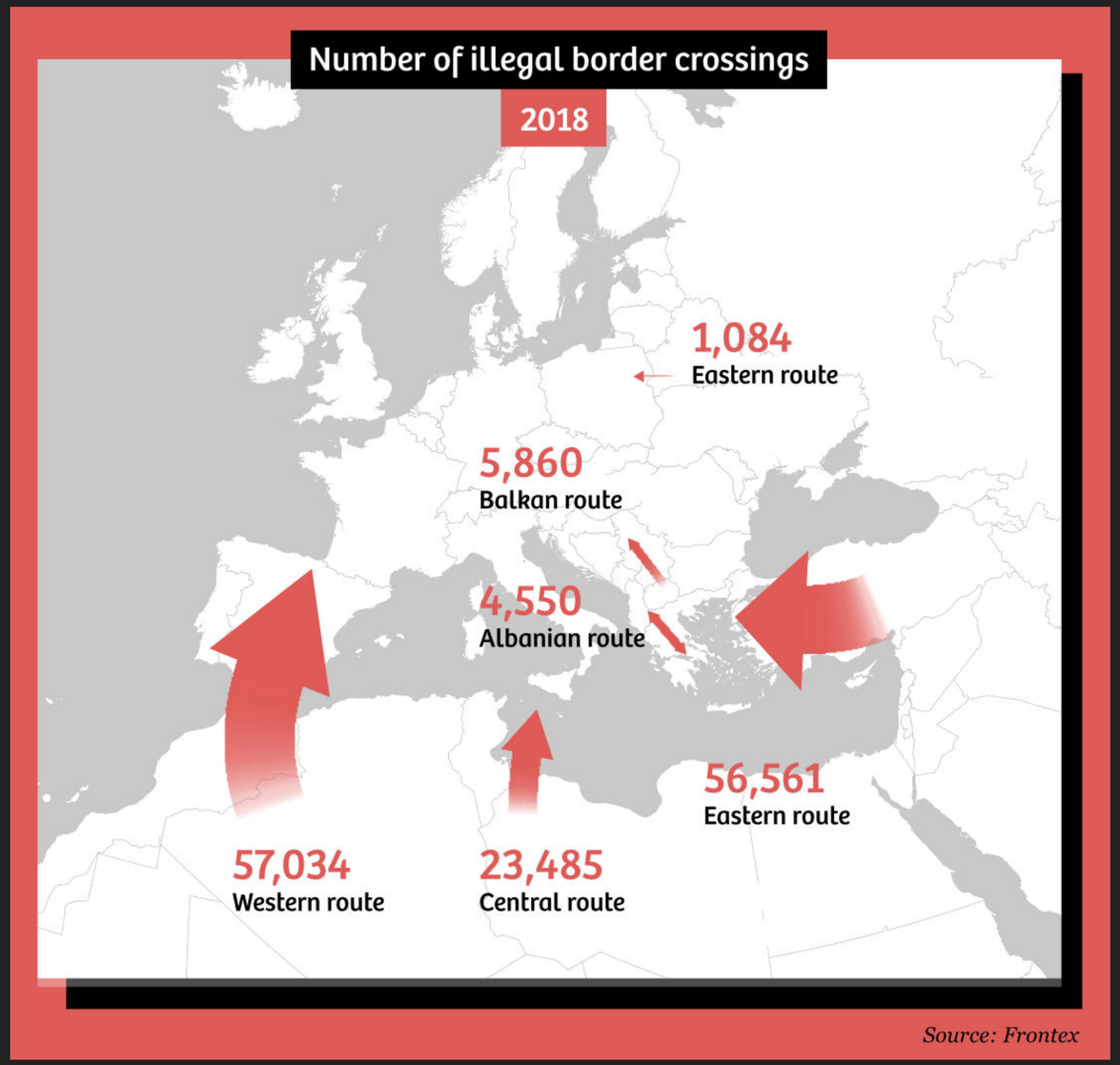
Scheme by Andy Kirk

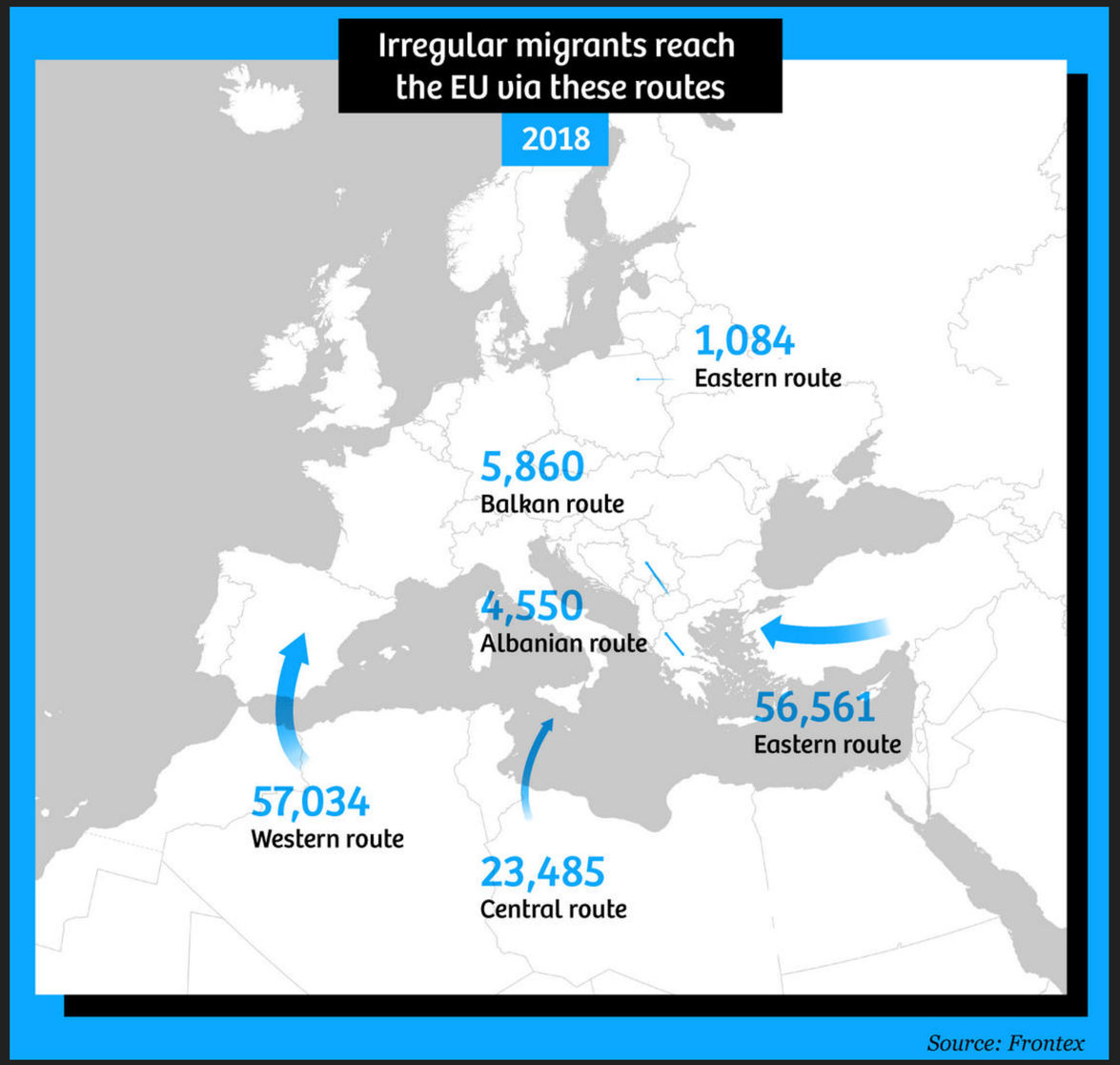
GOAL

Select charts that successfully tell your story

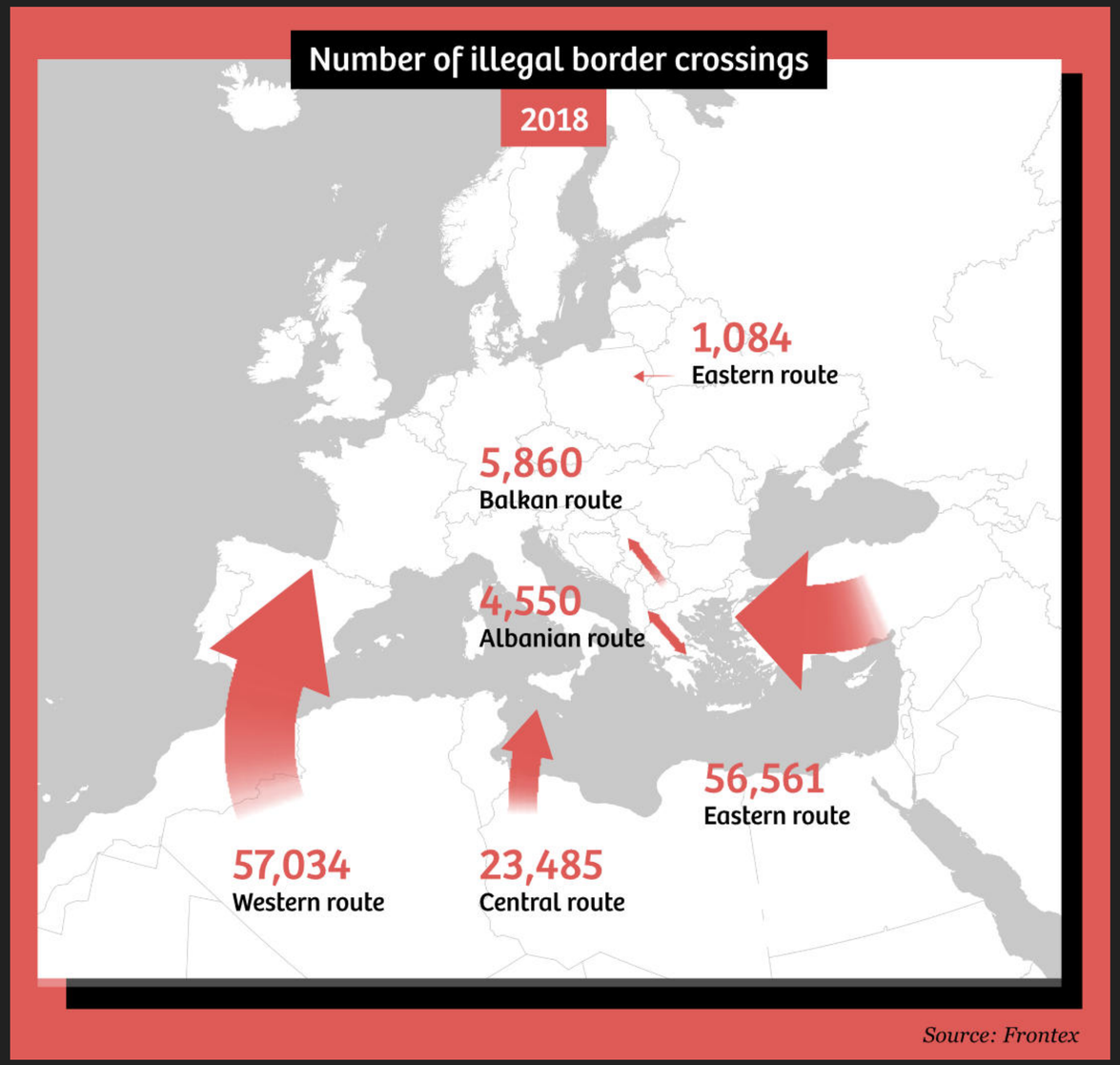


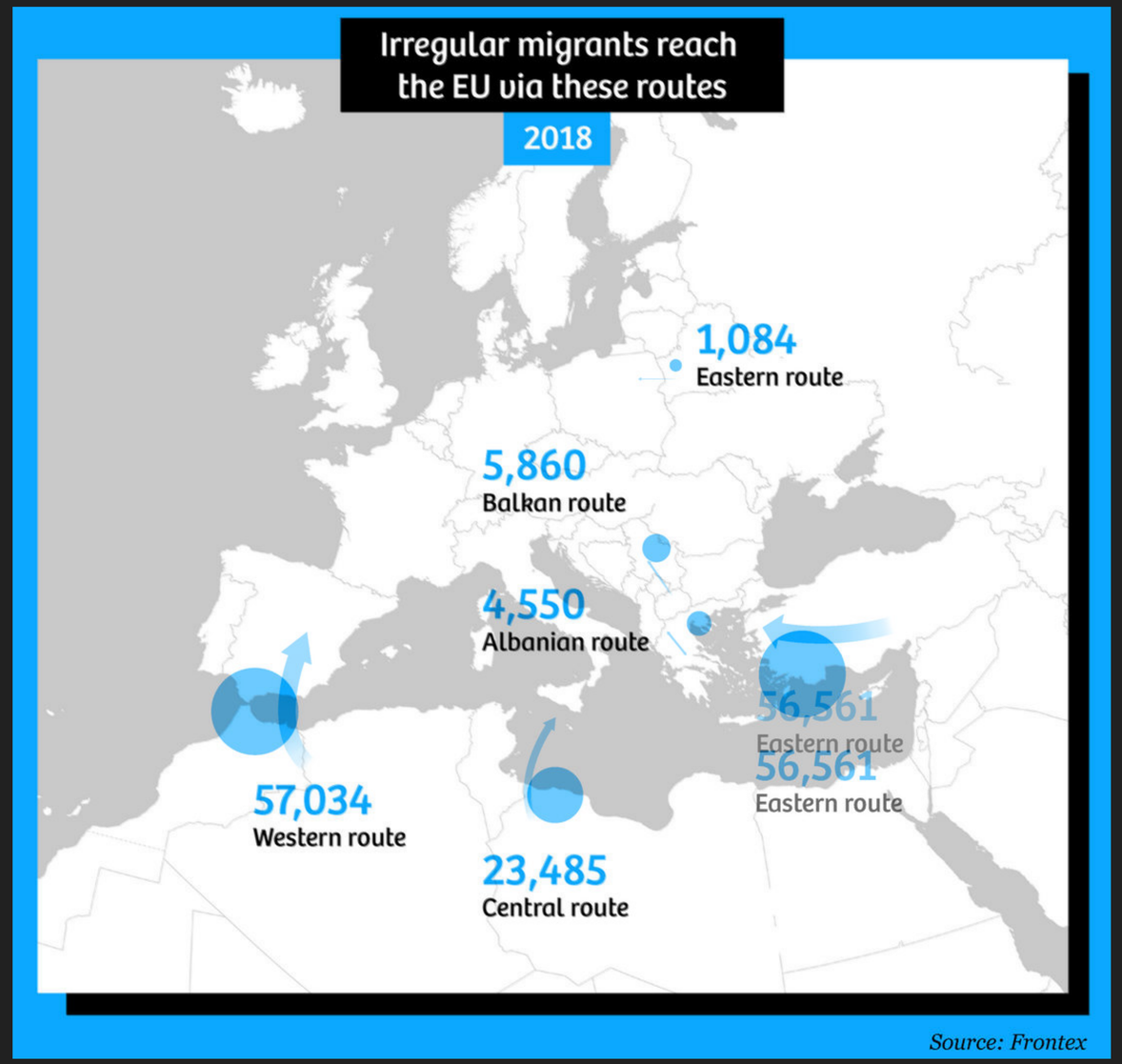
"How maps in the media make us more negative about migrants" by Maite Vermeulen, Leon de Korte & Henk van Houtum



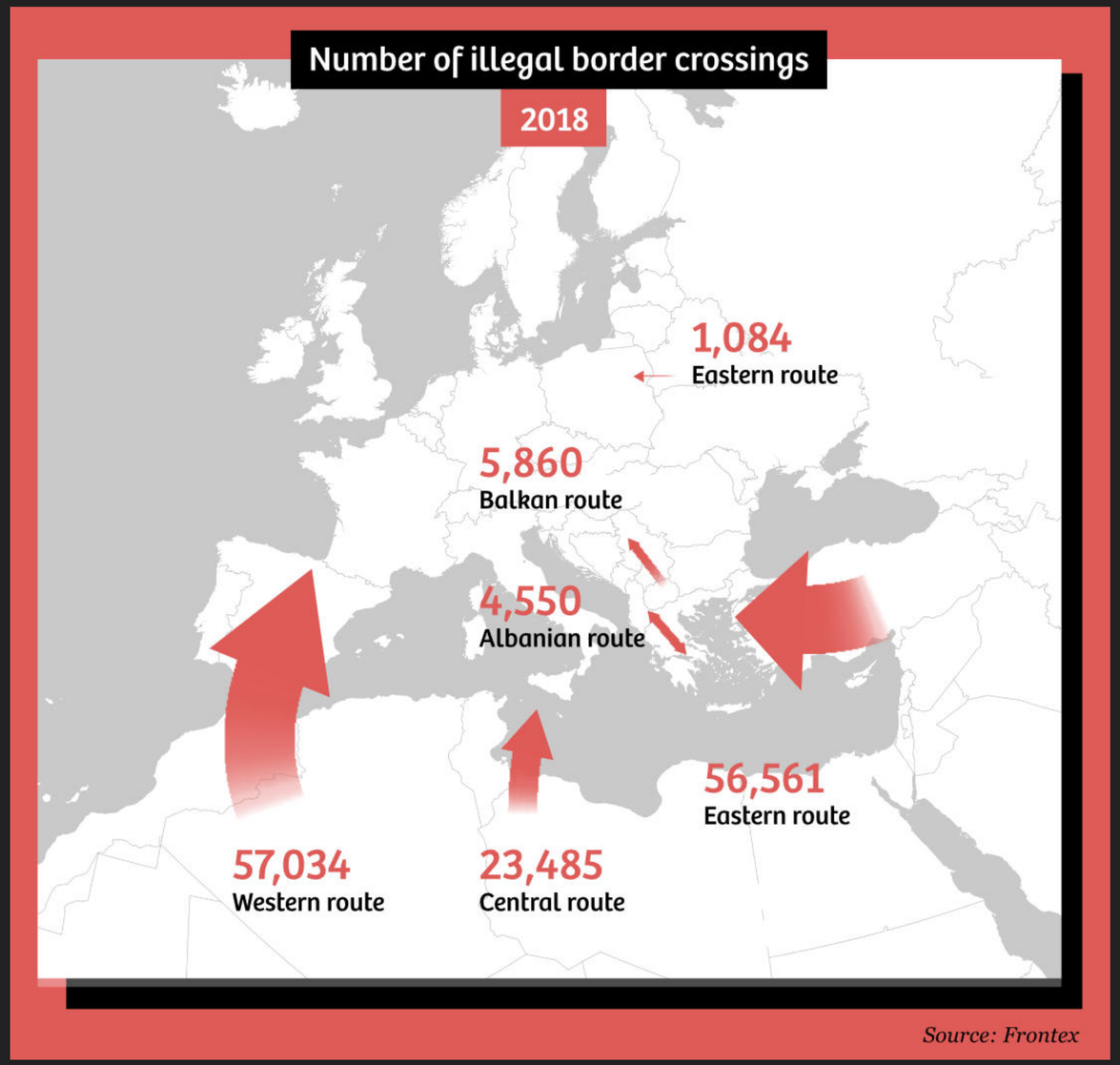


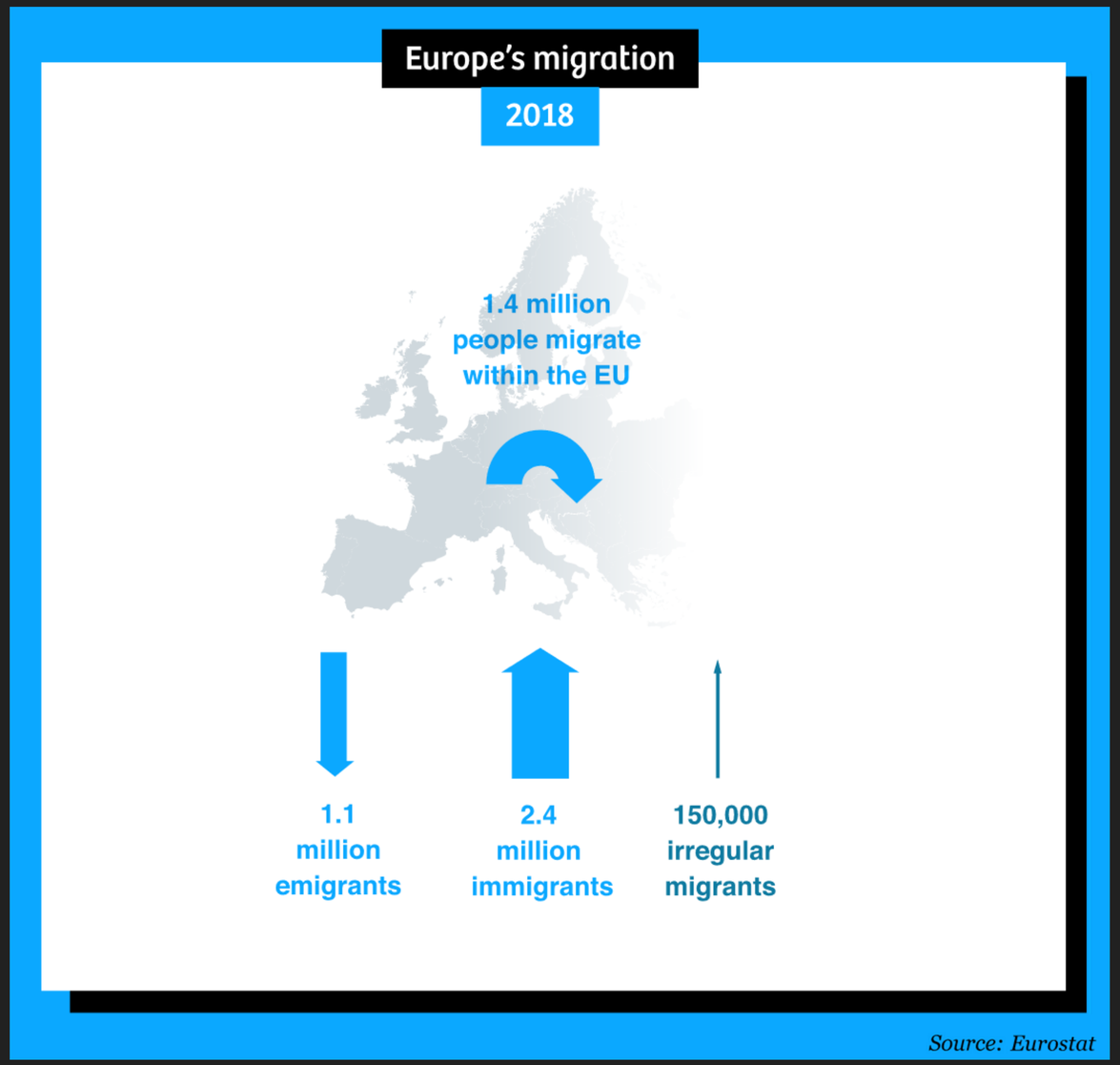
"How maps in the media make us more negative about migrants" by Maite Vermeulen, Leon de Korte & Henk van Houtum



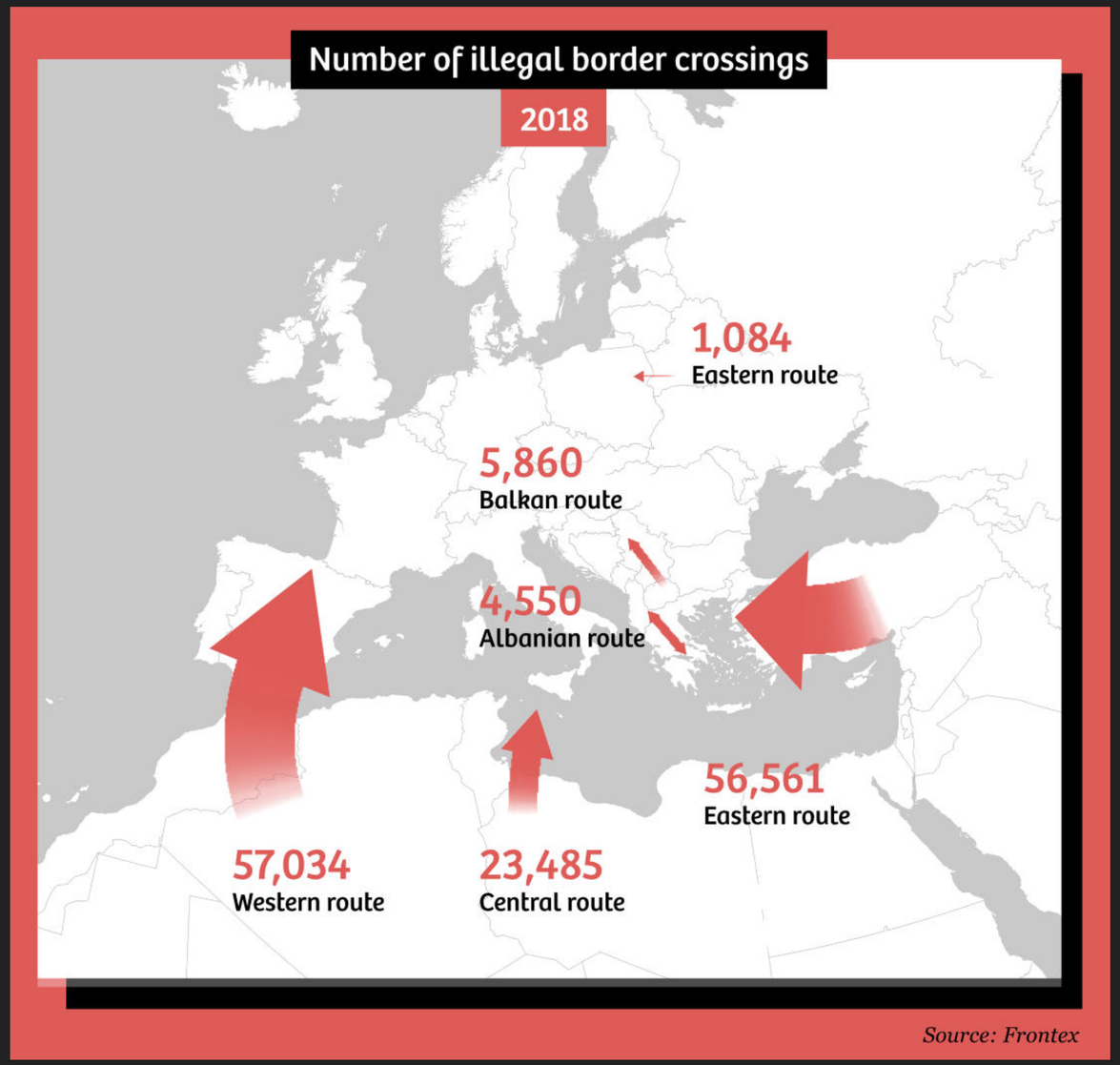


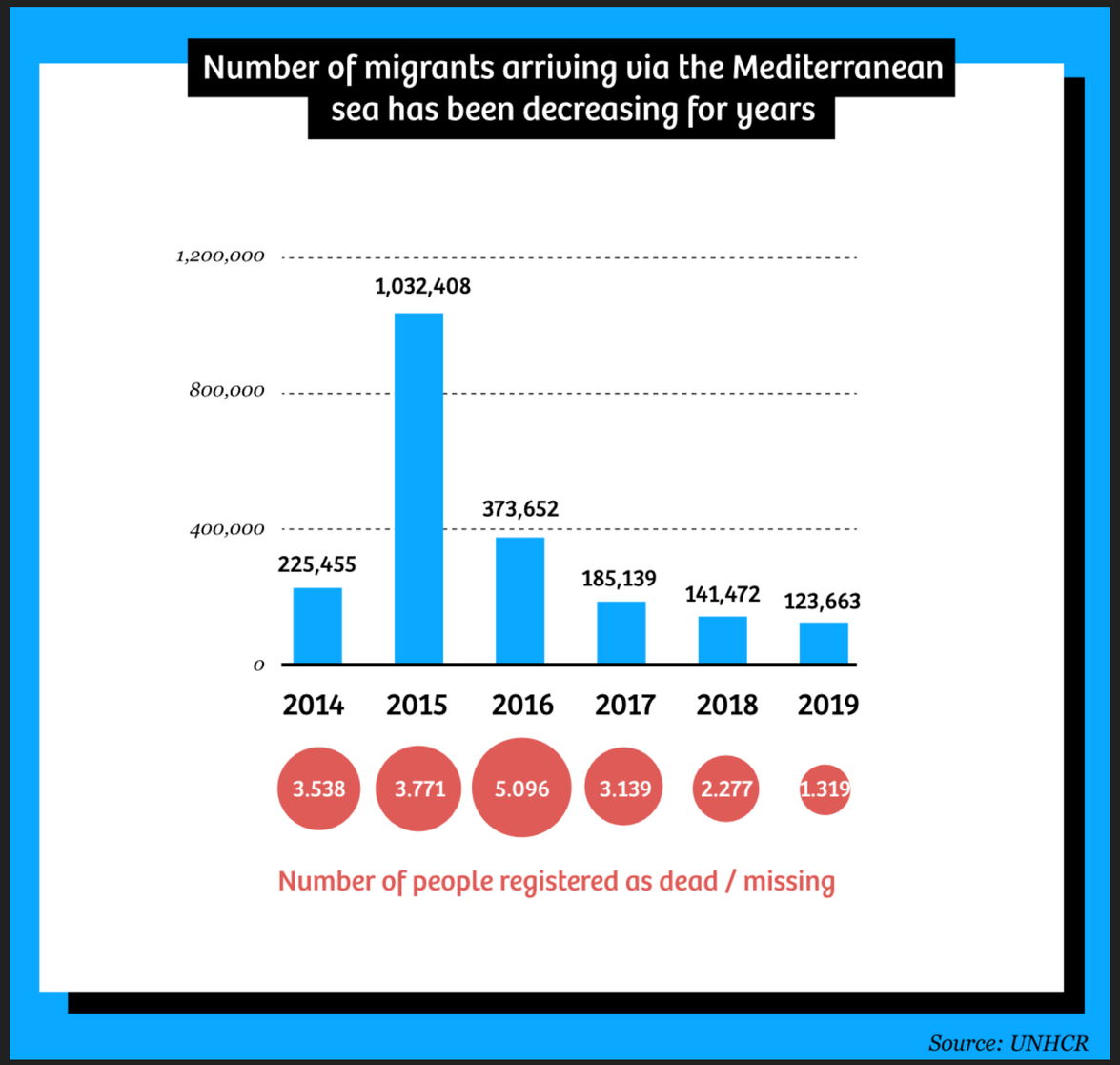
"How maps in the media make us more negative about migrants" by Maite Vermeulen, Leon de Korte & Henk van Houtum



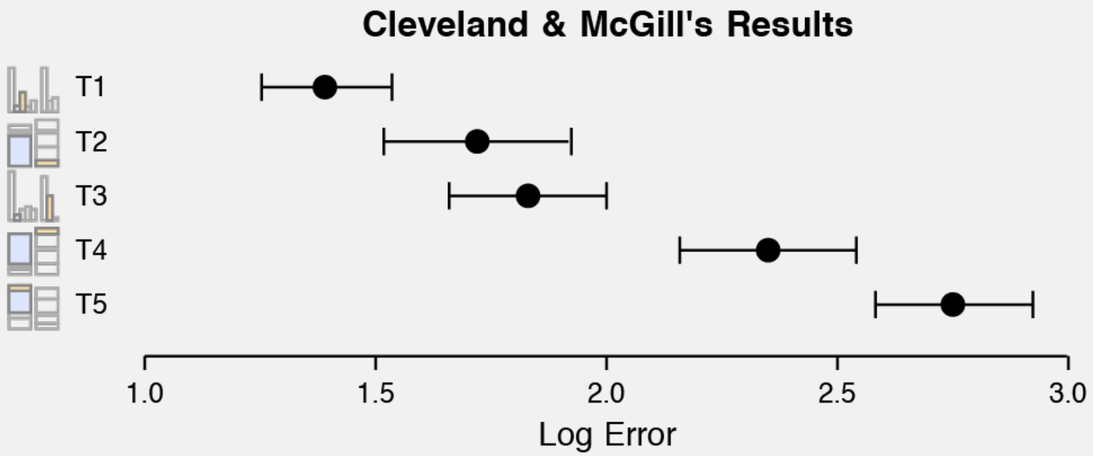


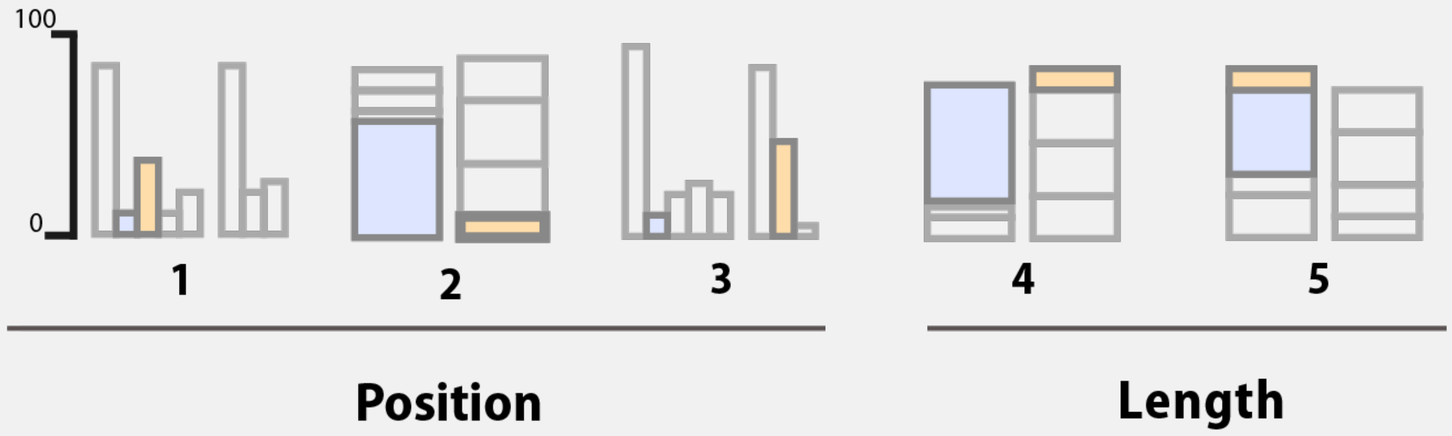
"How maps in the media make us more negative about migrants" by Maite Vermeulen, Leon de Korte & Henk van Houtum



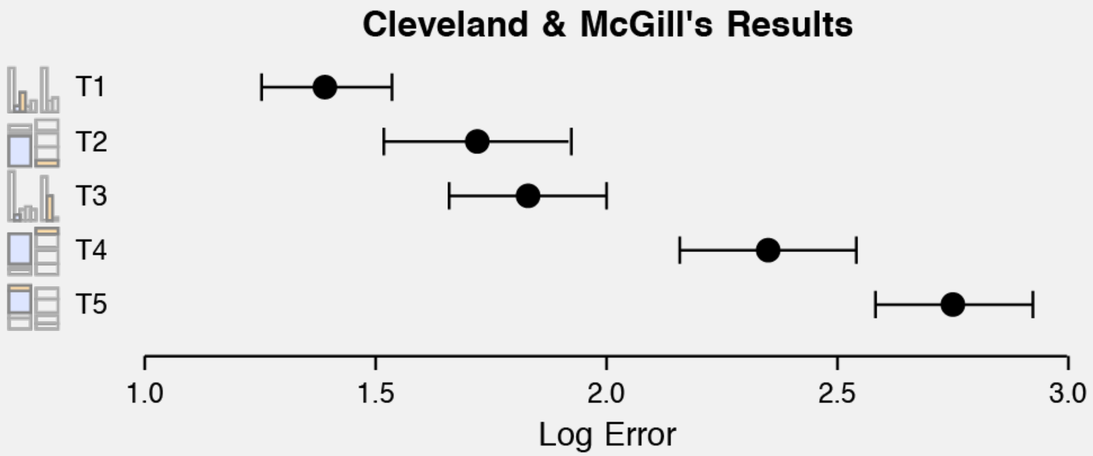


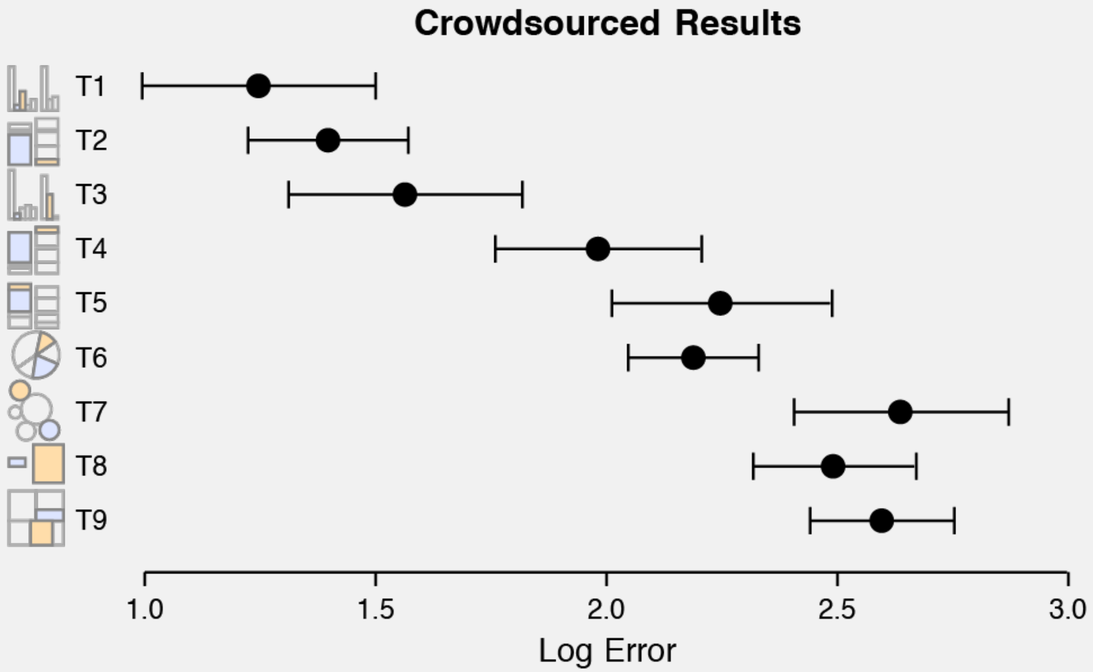
"How maps in the media make us more negative about migrants" by Maite Vermeulen, Leon de Korte & Henk van Houtum





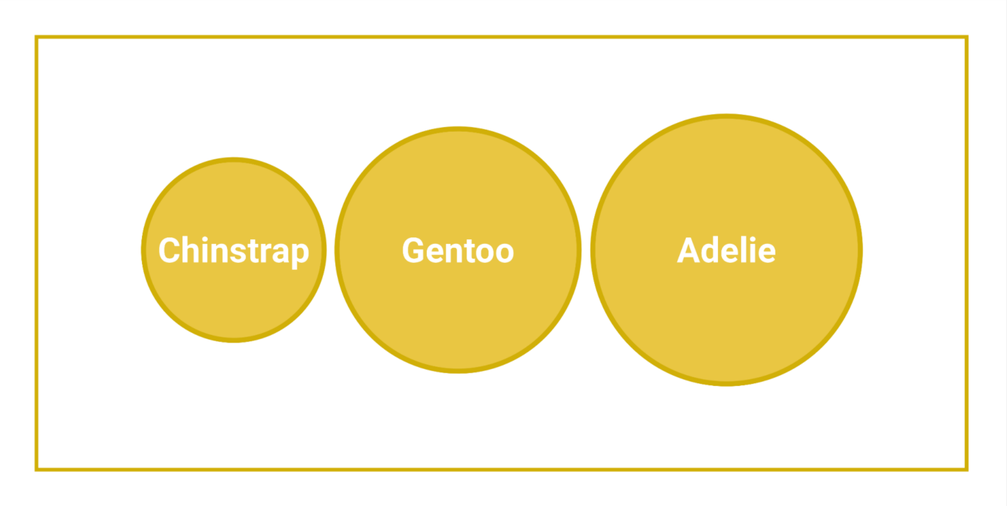
*Source: Kieran Healy's **"Data Visualization: A Practical Introduction"**; results based on Heer and Bostock, following Cleveland and McGill*



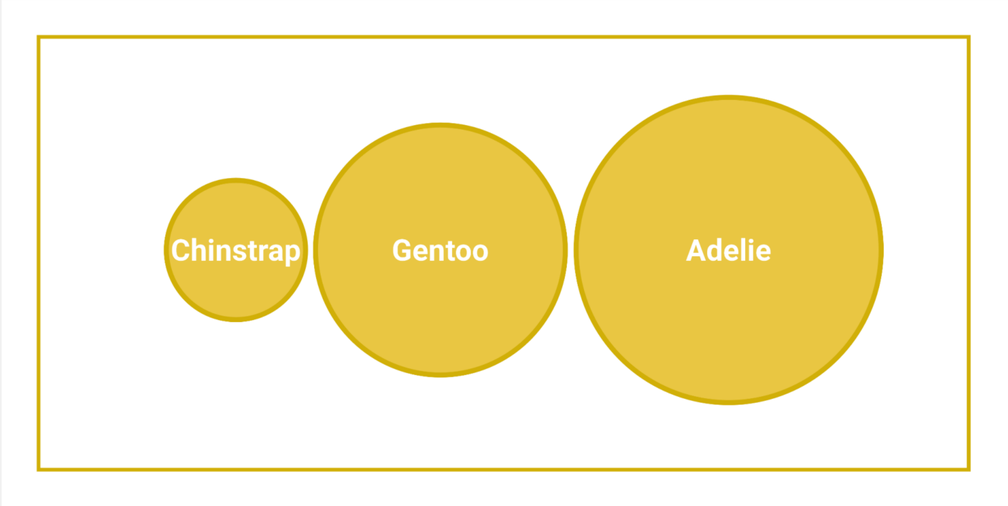


Source: Kieran Healy's "Data Visualization: A Practical Introduction"; results based on Heer and Bostock, following Cleveland and McGill

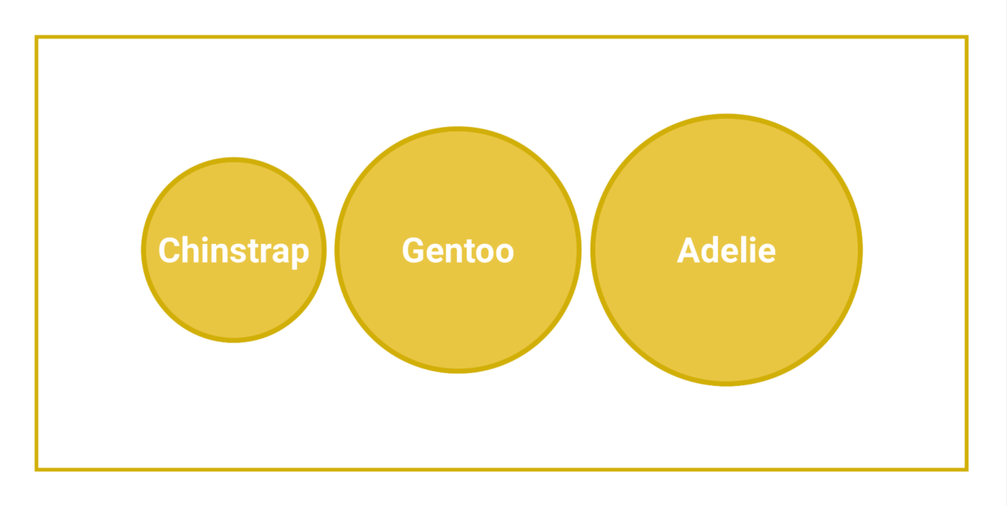
Use area.



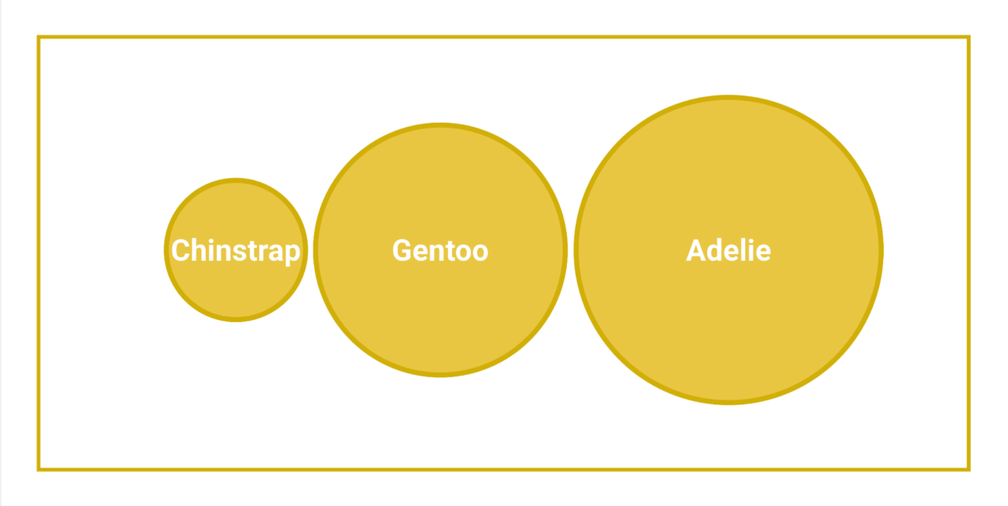
Not radius!

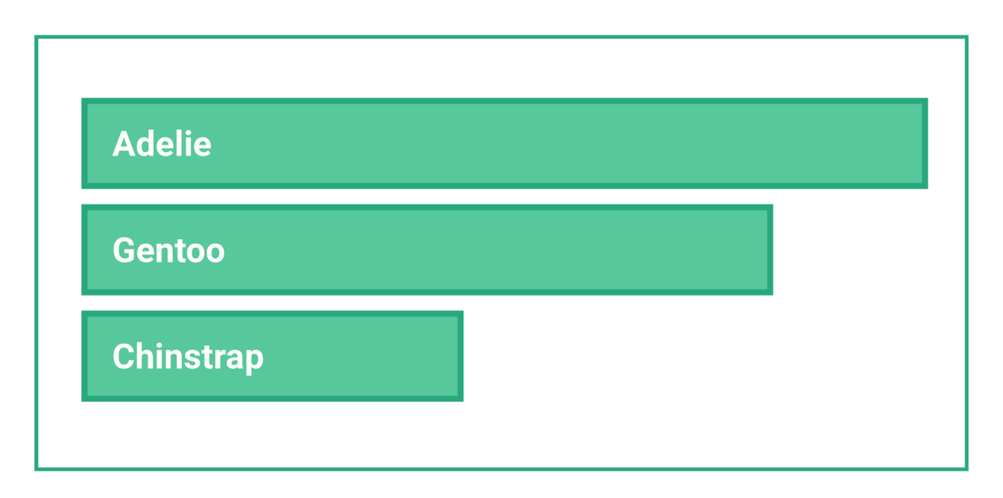


Use area.

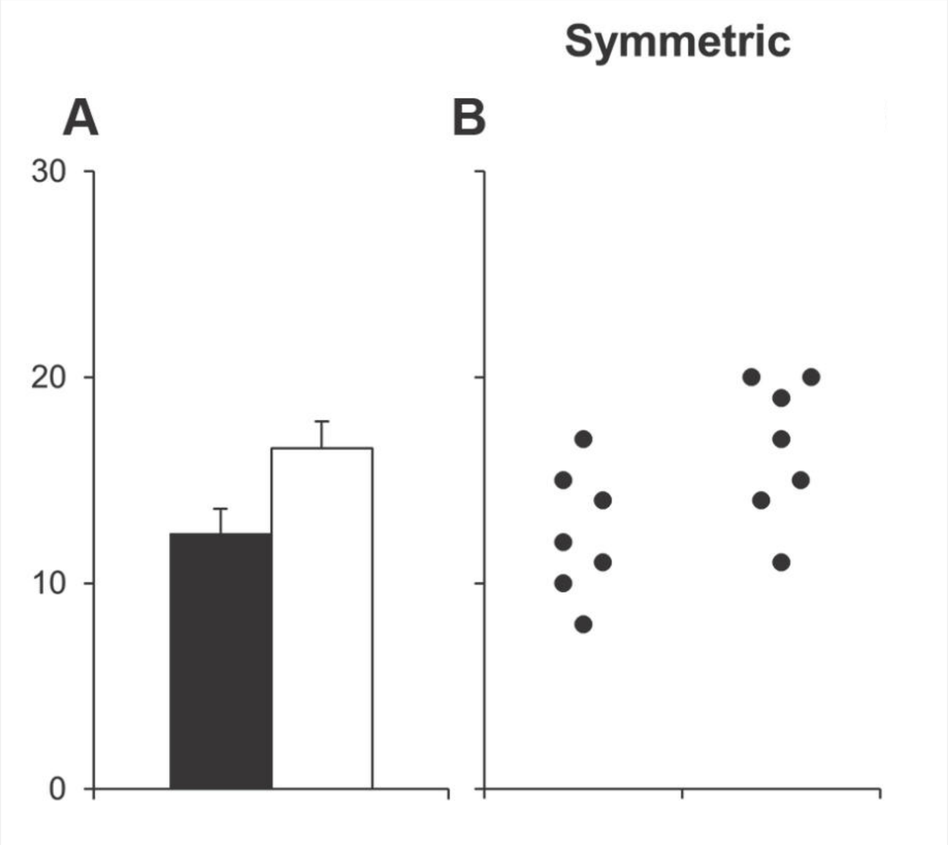


Not radius!

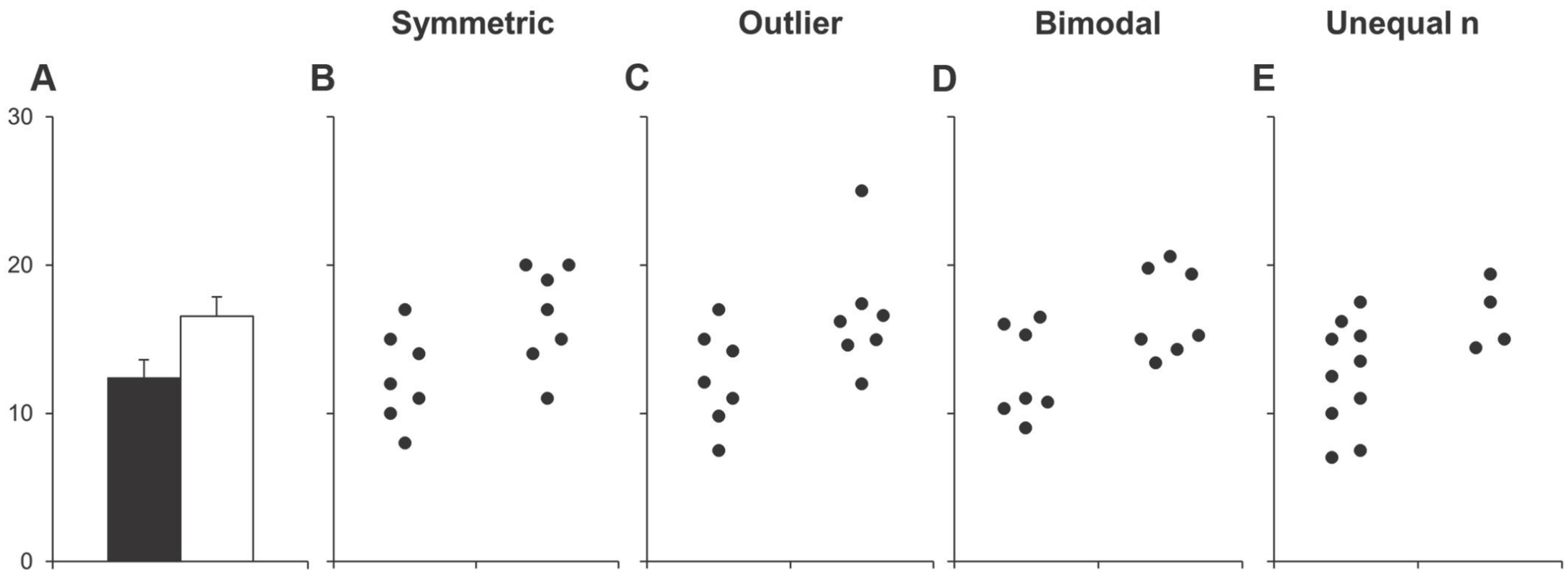




or: bars!

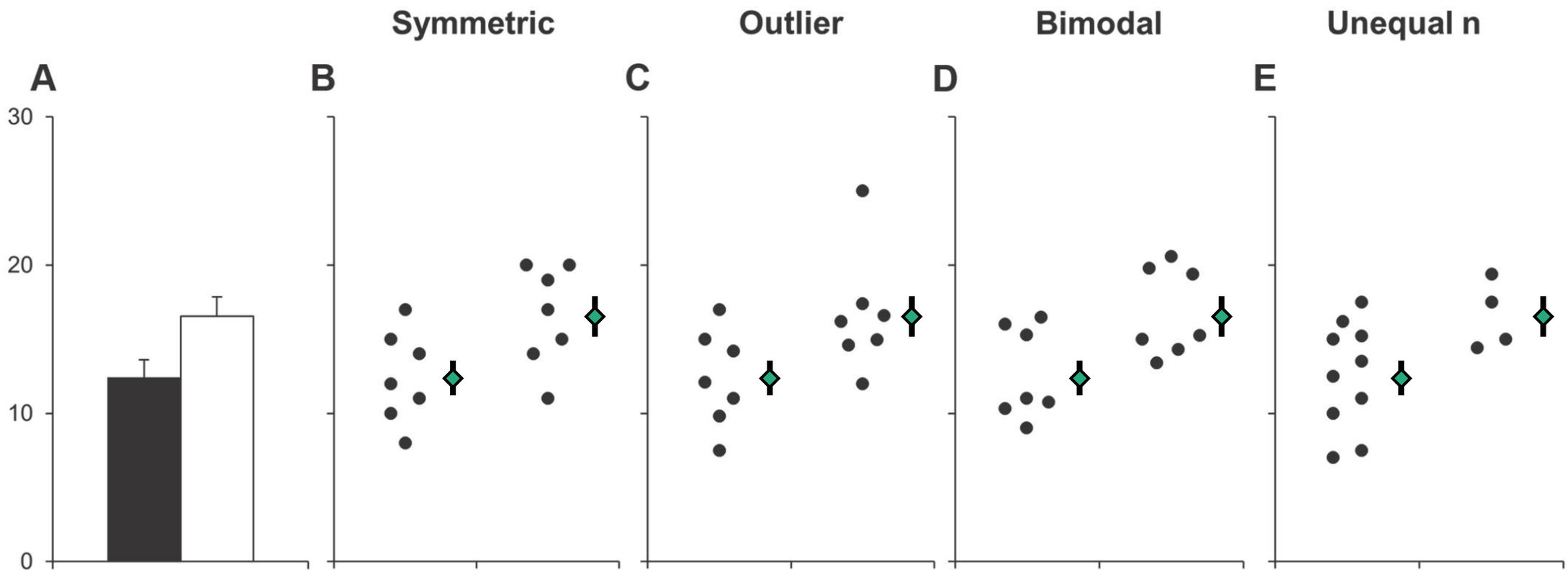


Source: Weissgerber et al. (2015) PLoS Biology



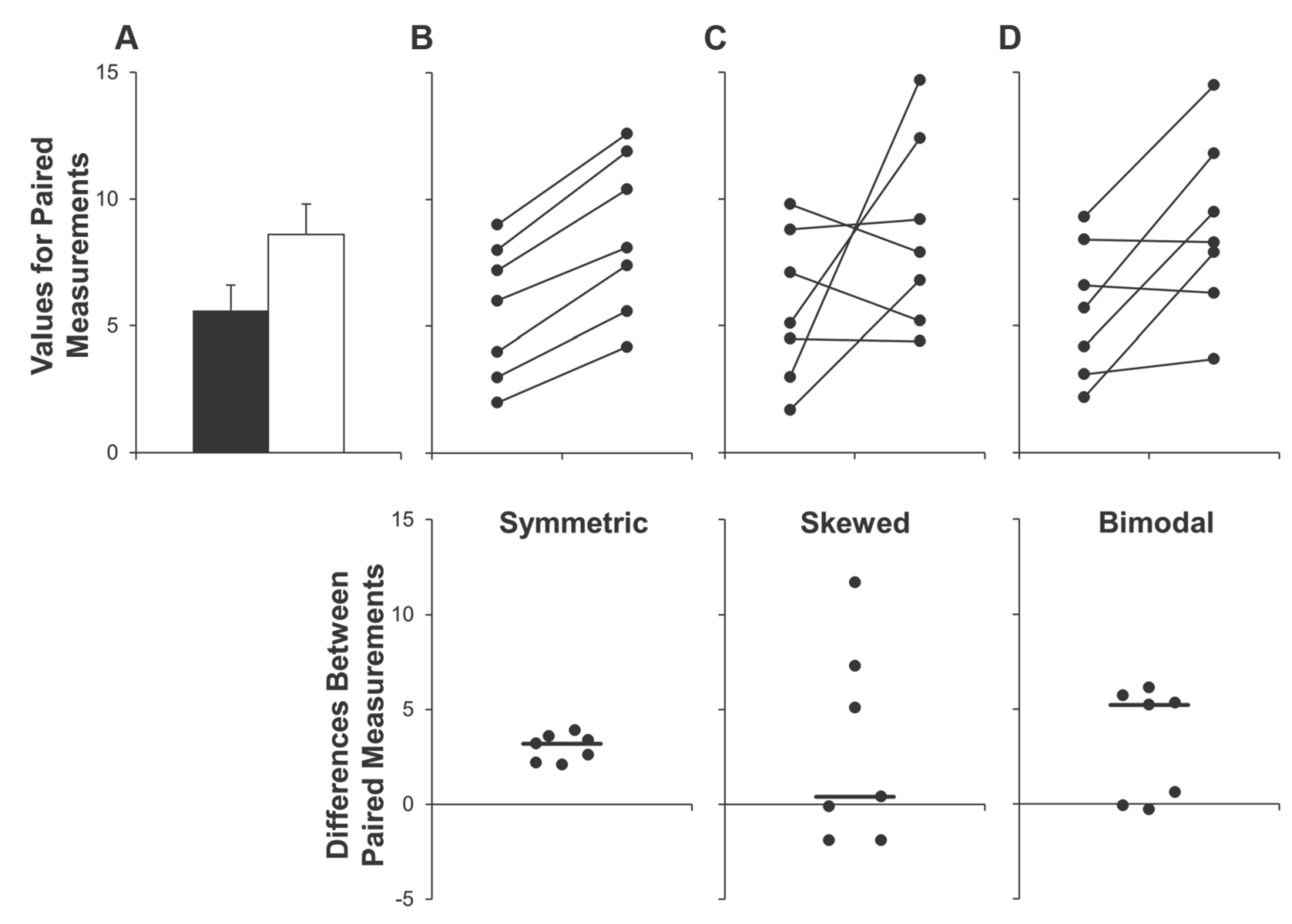
Test				
	p value			
T-test: Equal var.	0.035	0.050	0.026	0.063
T-test: Unequal var.	0.035	0.050	0.026	0.035
Wilcoxon	0.054	0.073	0.128	0.103

Source: Weissgerber et al. (2015) PLoS Biology

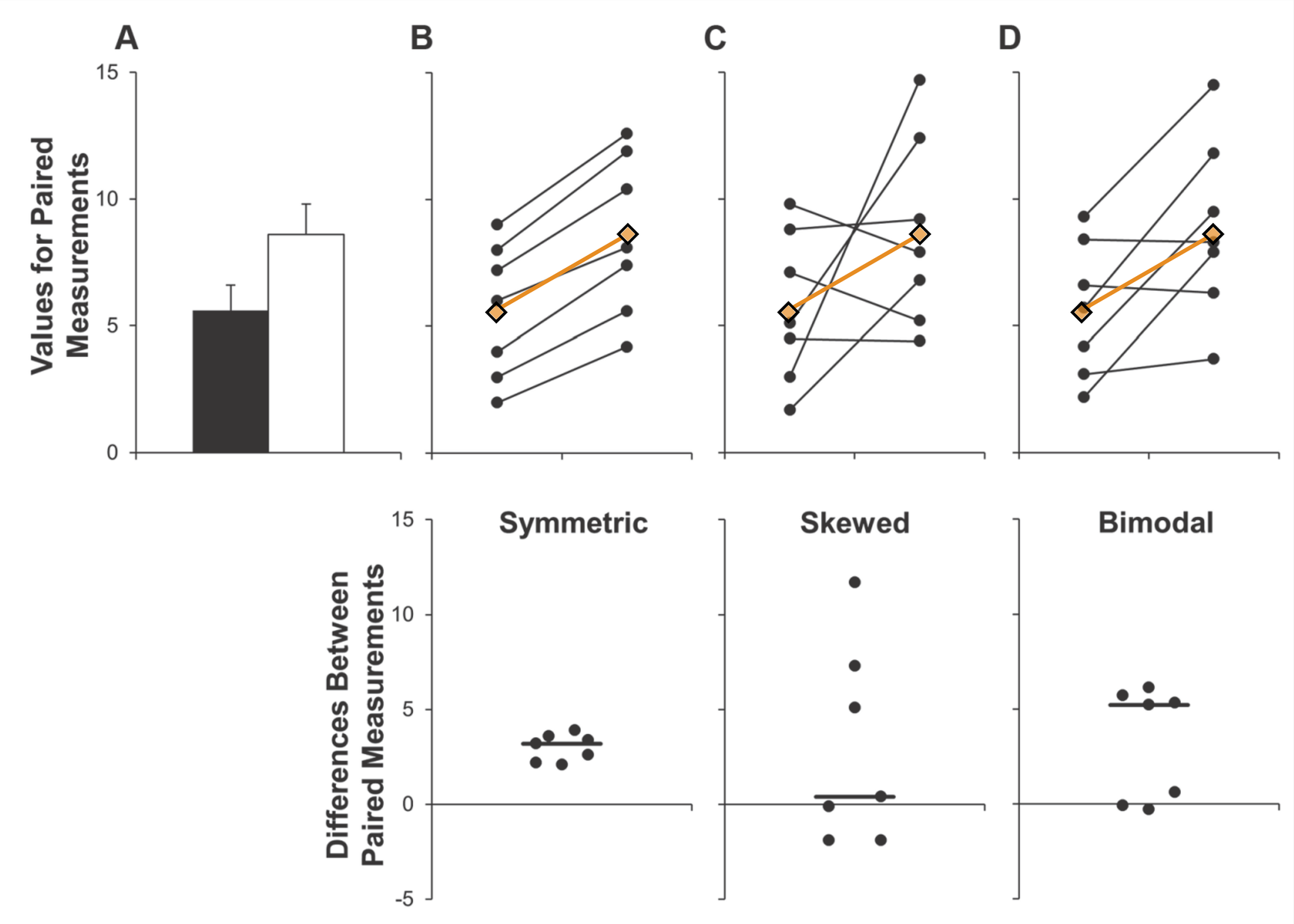


Test		p value		
T-test: Equal var.	0.035	0.050	0.026	0.063
T-test: Unequal var.	0.035	0.050	0.026	0.035
Wilcoxon	0.054	0.073	0.128	0.103

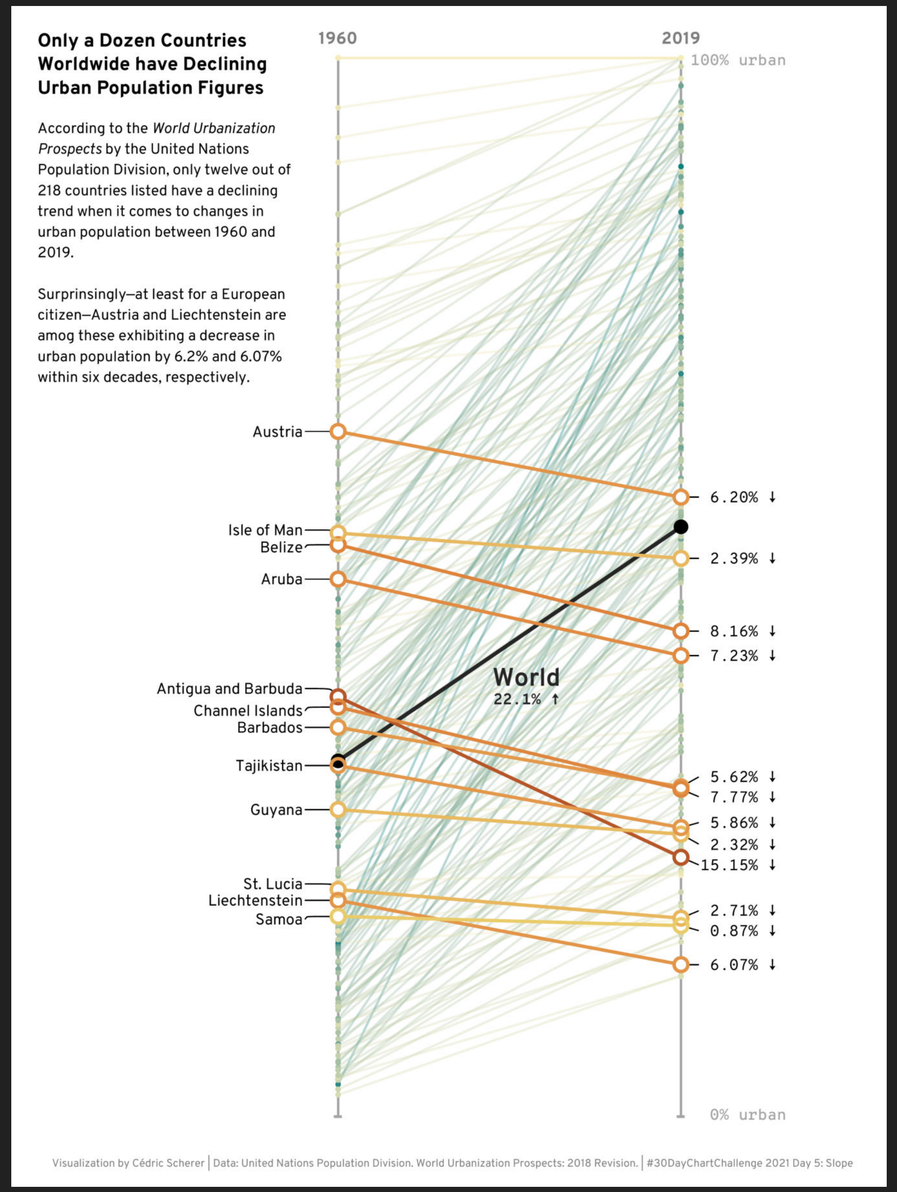
Modified from Weissgerber et al. (2015) PLoS Biology



Source: Weissgerber et al. (2015) PLoS Biology

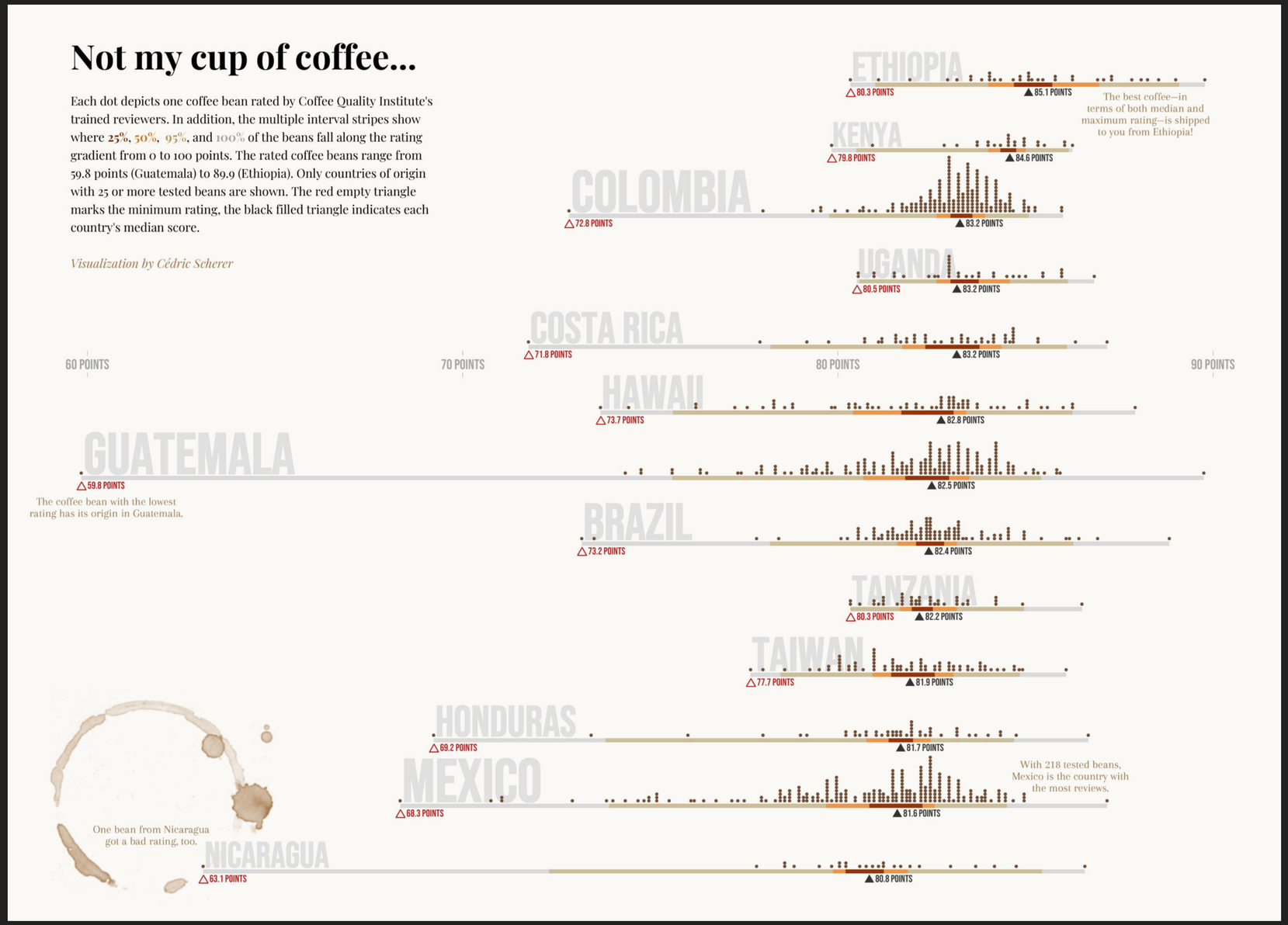


Source: Weissgerber et al. (2015) PLoS Biology



Urban Population Trends, #30DayChartChallenge Contribution

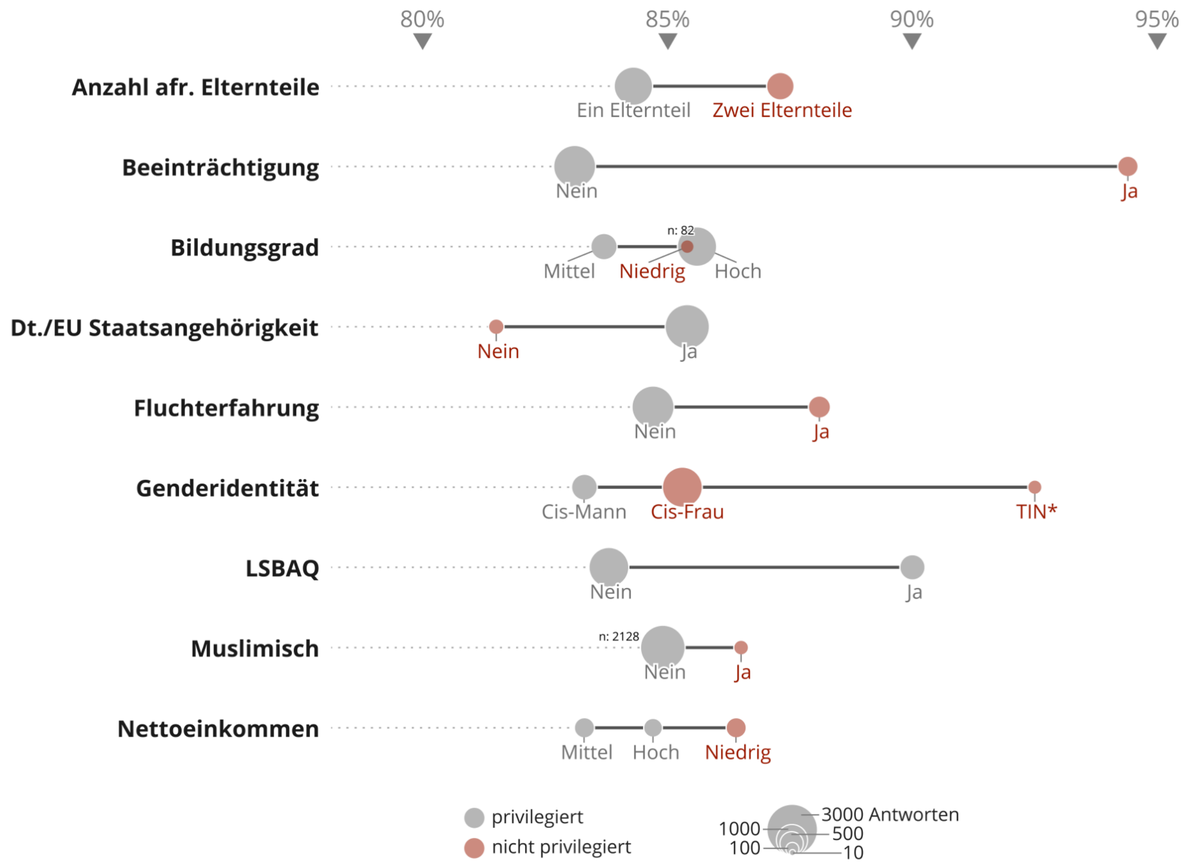
cedricscherer.com @CedScherer z3tt



"Not my cup of coffee", #TidyTuesday Contribution

cedricscherer.com @CedScherer z3tt

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



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

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

Lizenz: CC-BY-NC by EOTO & CFE | afrozensus.de

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

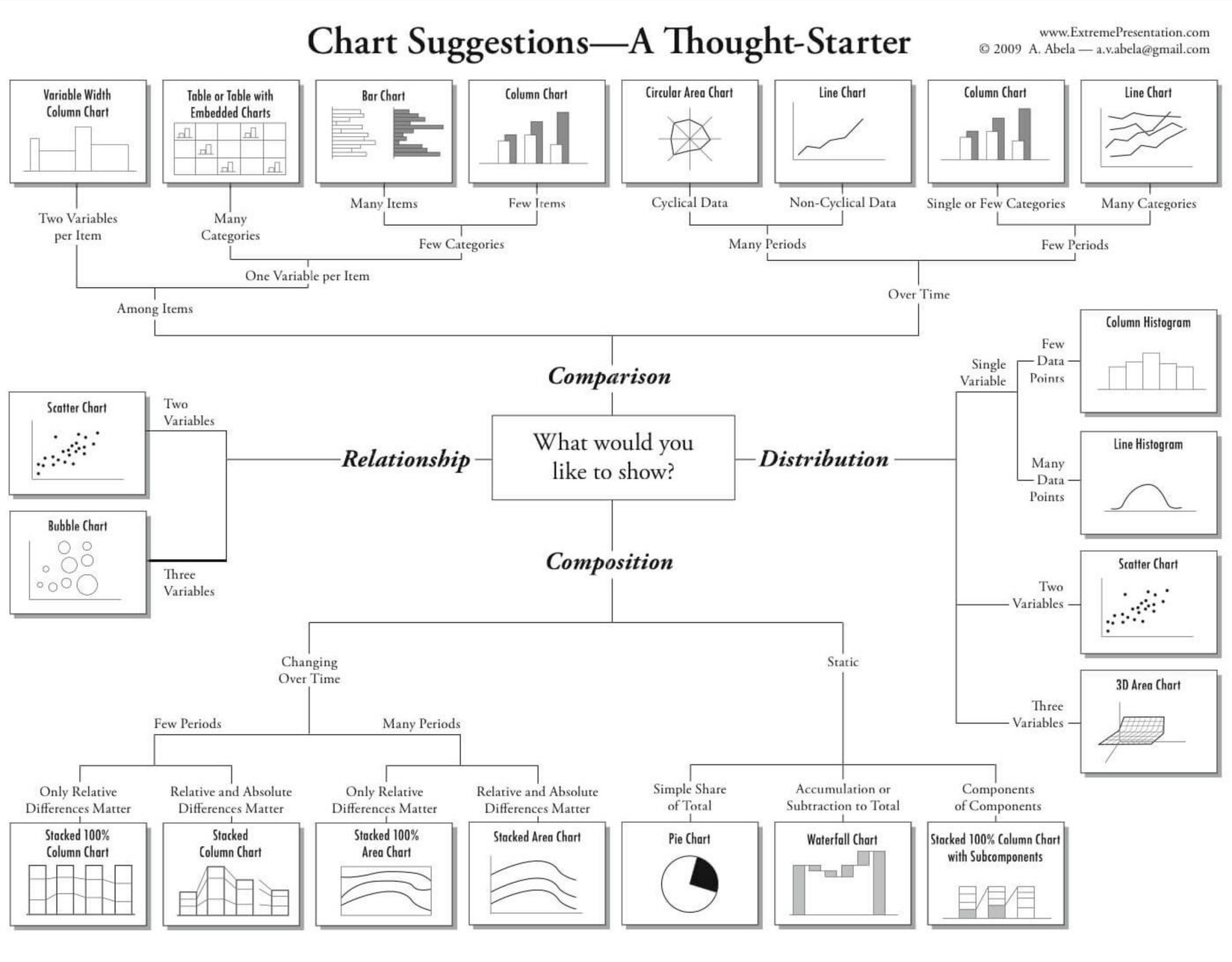
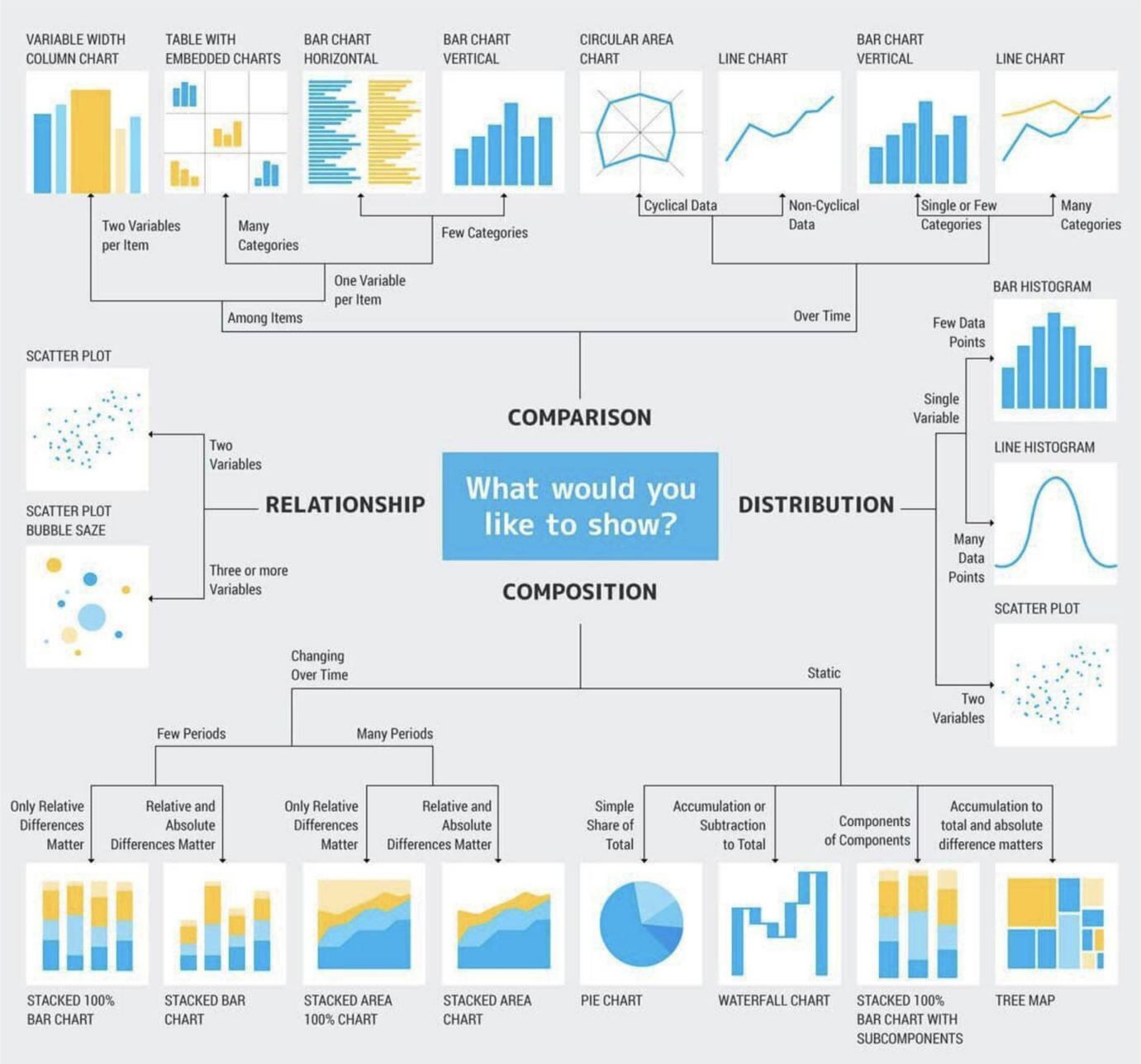
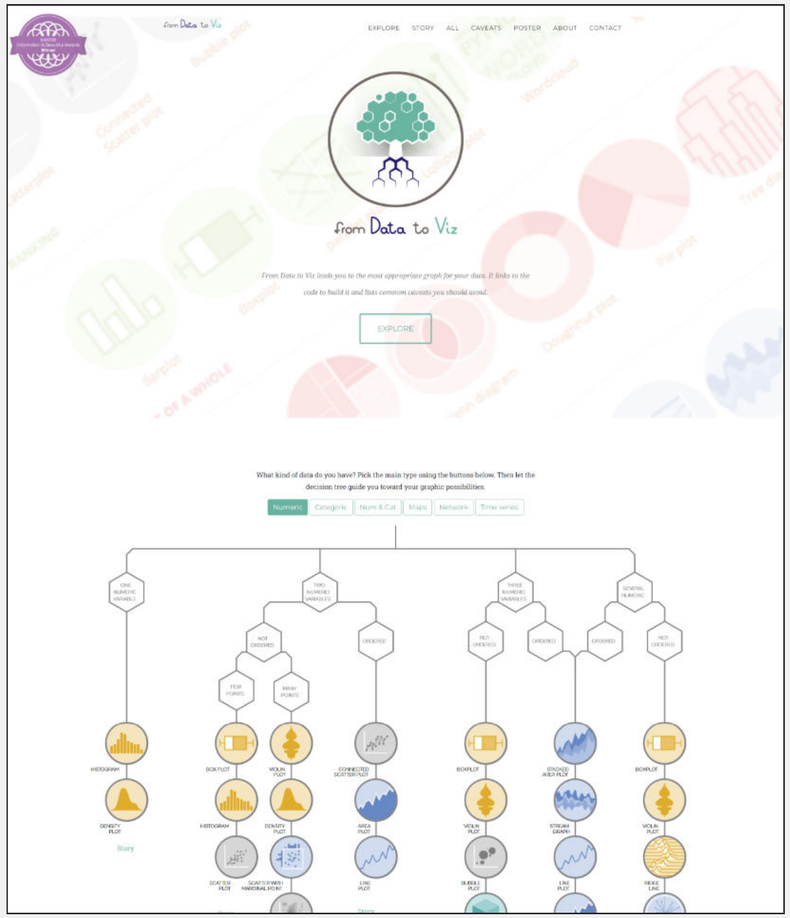


Chart Suggestions—A Thought-Starter
www.ExtremePresentation.com
© 2009 A. Abela — a.v.abela@gmail.com
Variable Width Column Chart
Table or Table with Embedded Charts
Bar Chart
Column Chart
Circular Area Chart
Line Chart
Column Chart
Line Chart
Two Variables per Item
Many Categories
Many Items
Few Items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


What would you like to show?
COMPARISON
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
RELATIONSHIP
SCATTER PLOT
SCATTER PLOT BUBBLE SAZE
Two Variables
Three or more Variables
DISTRIBUTION
BAR HISTOGRAM
Few Data Points
Single Variable
LINE HISTOGRAM
Many Data Points
SCATTER PLOT
Two Variables
COMPOSITION
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

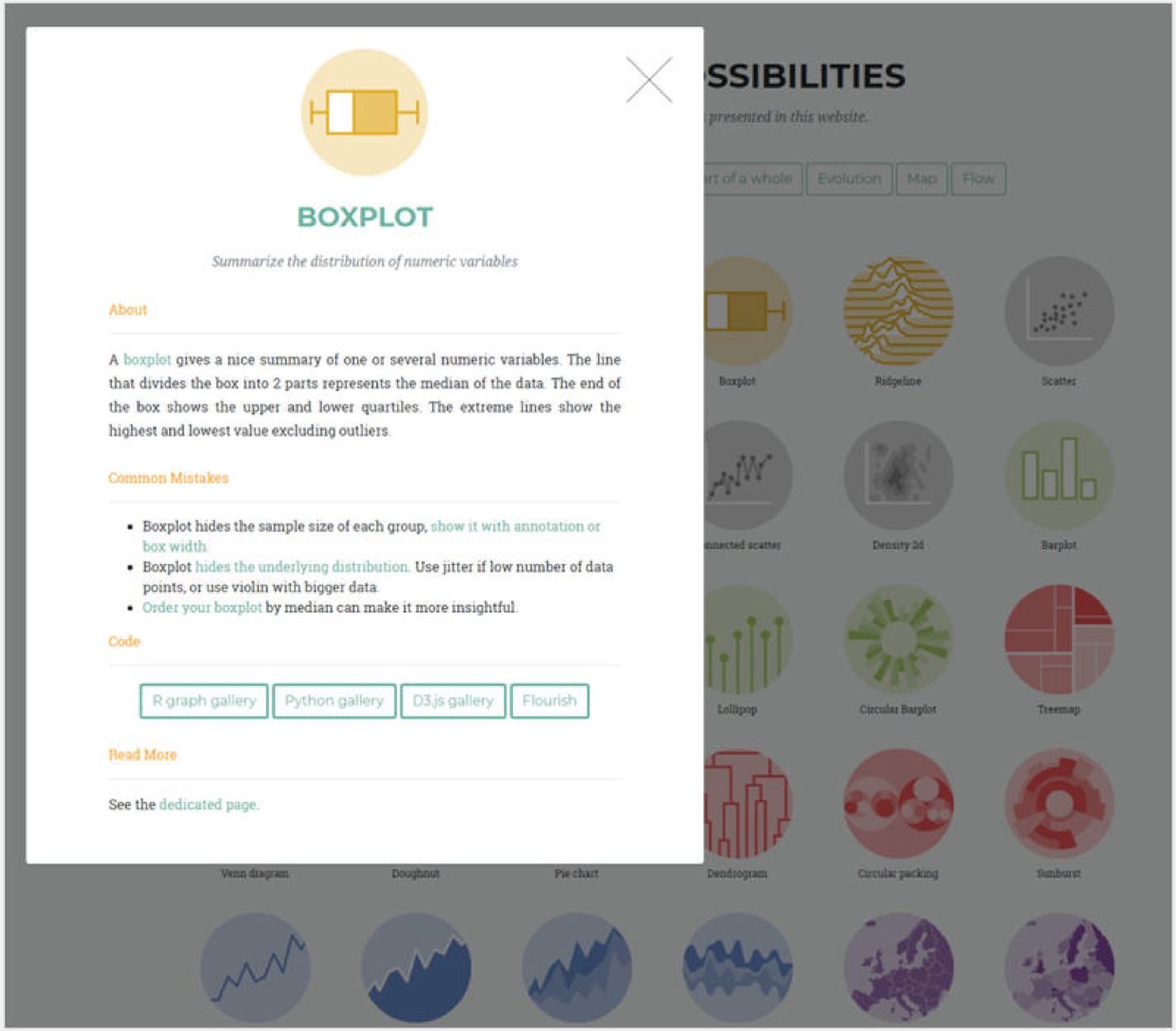


data-to-viz.com

datavizproject.com

visualizationuniverse.com

Source: data-to-viz.com

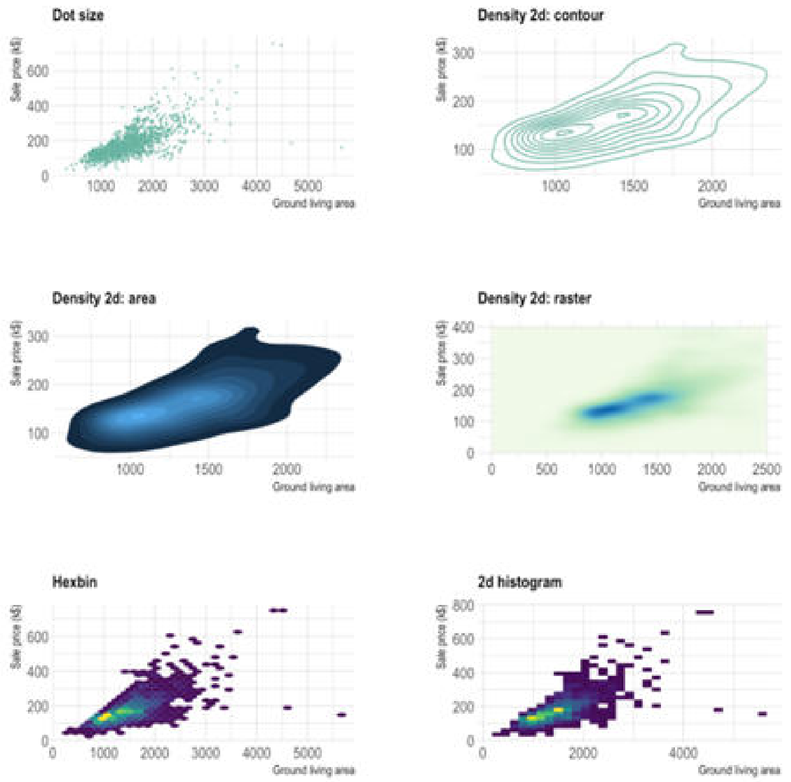


Source: data-to-viz.com

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

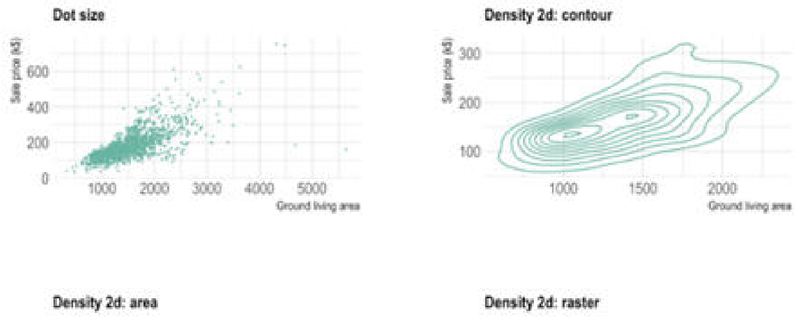


Going further

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 o Charts” poster

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

Have a look at the “Chart Choice Helpers”

☛ Explore and discuss potential chart types for future data visualizaitons

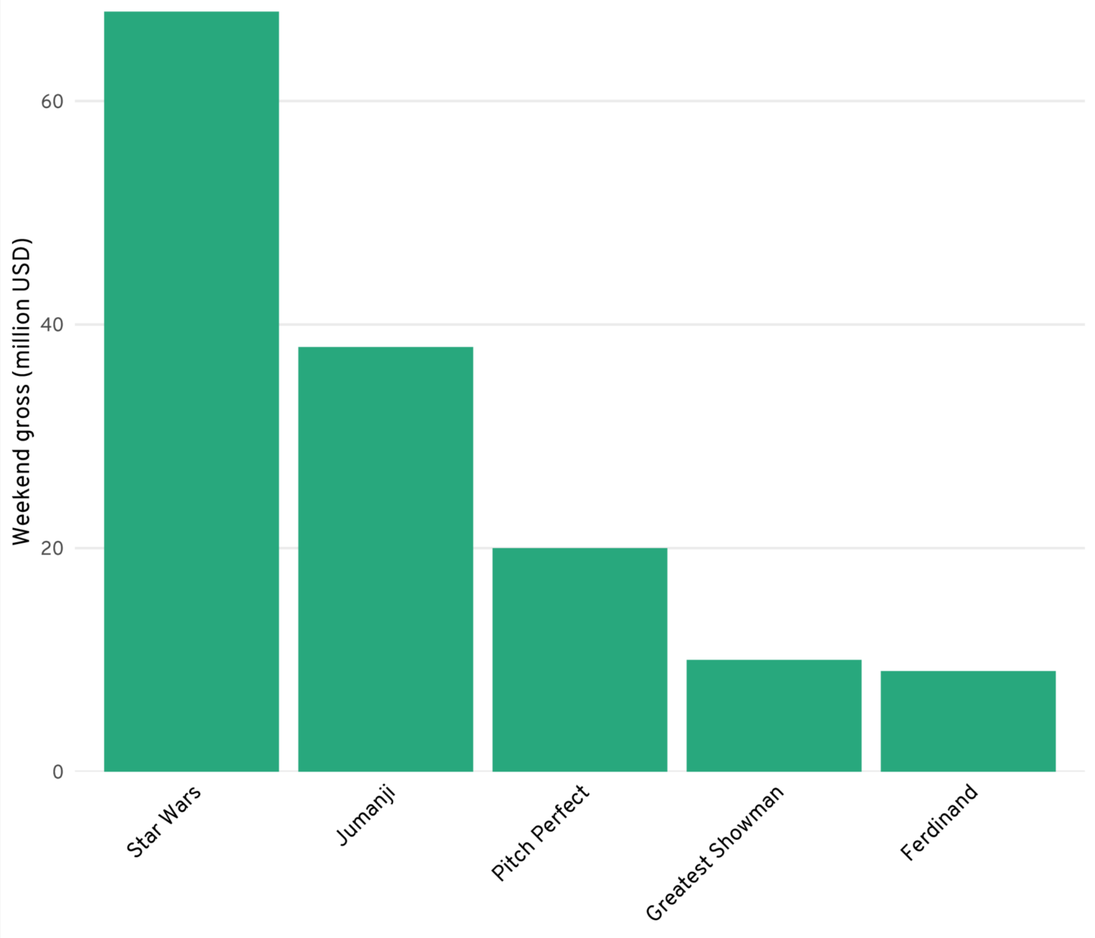
☛ Look up the chart types on the poster you didn’t know

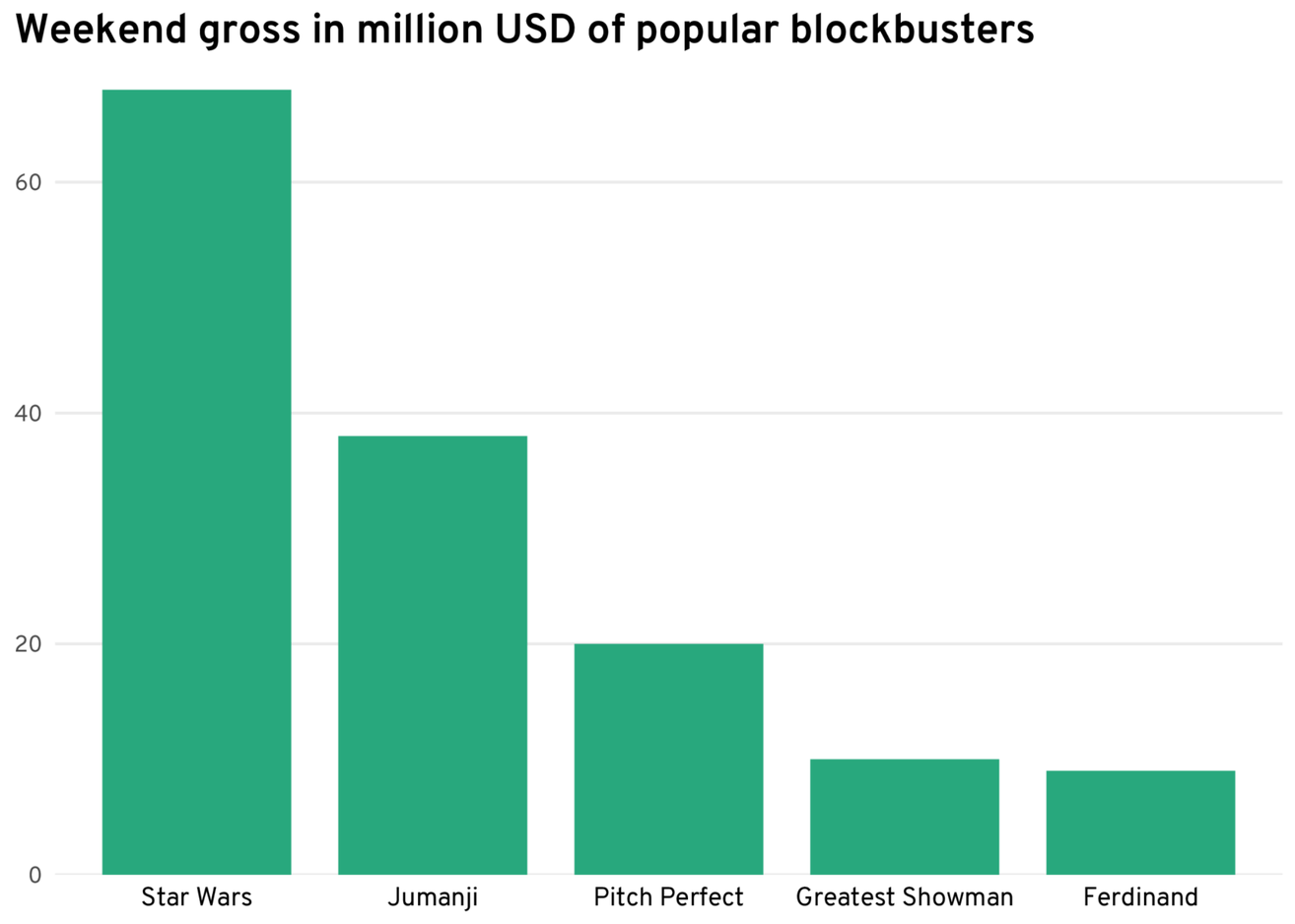
☛ Are there chart types you find hard to read / understand?

DATAVIZ RULES

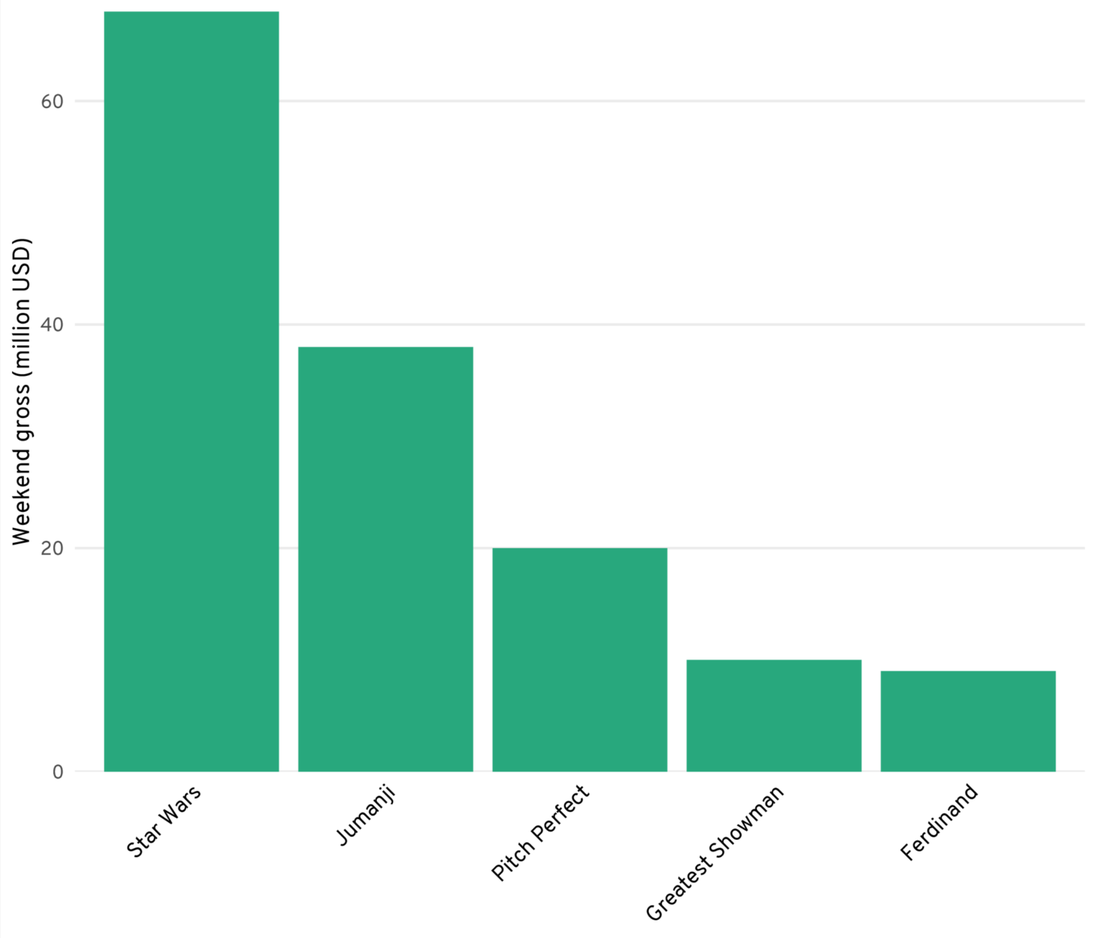
Recommendations

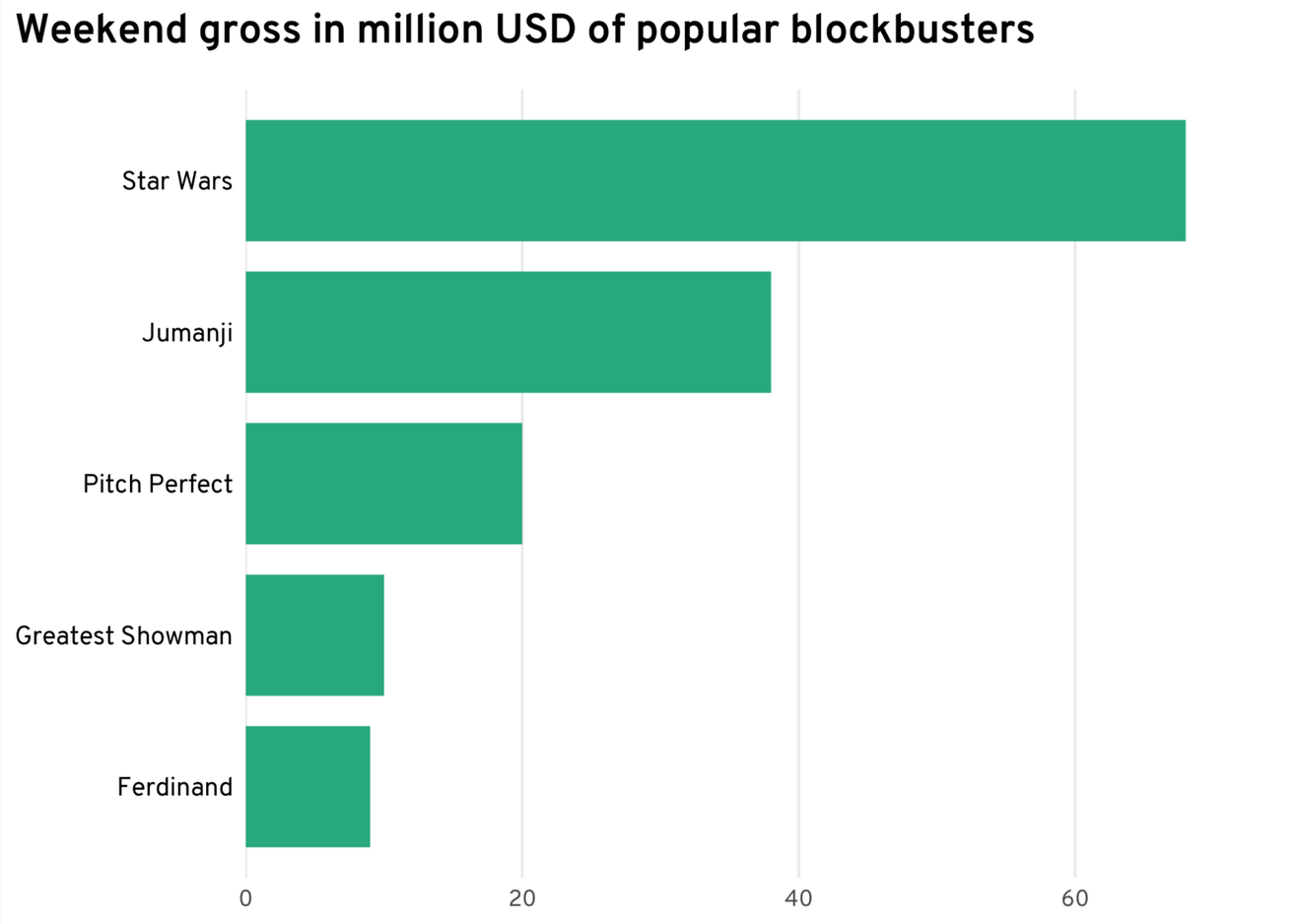
Don’t Tilt Labels



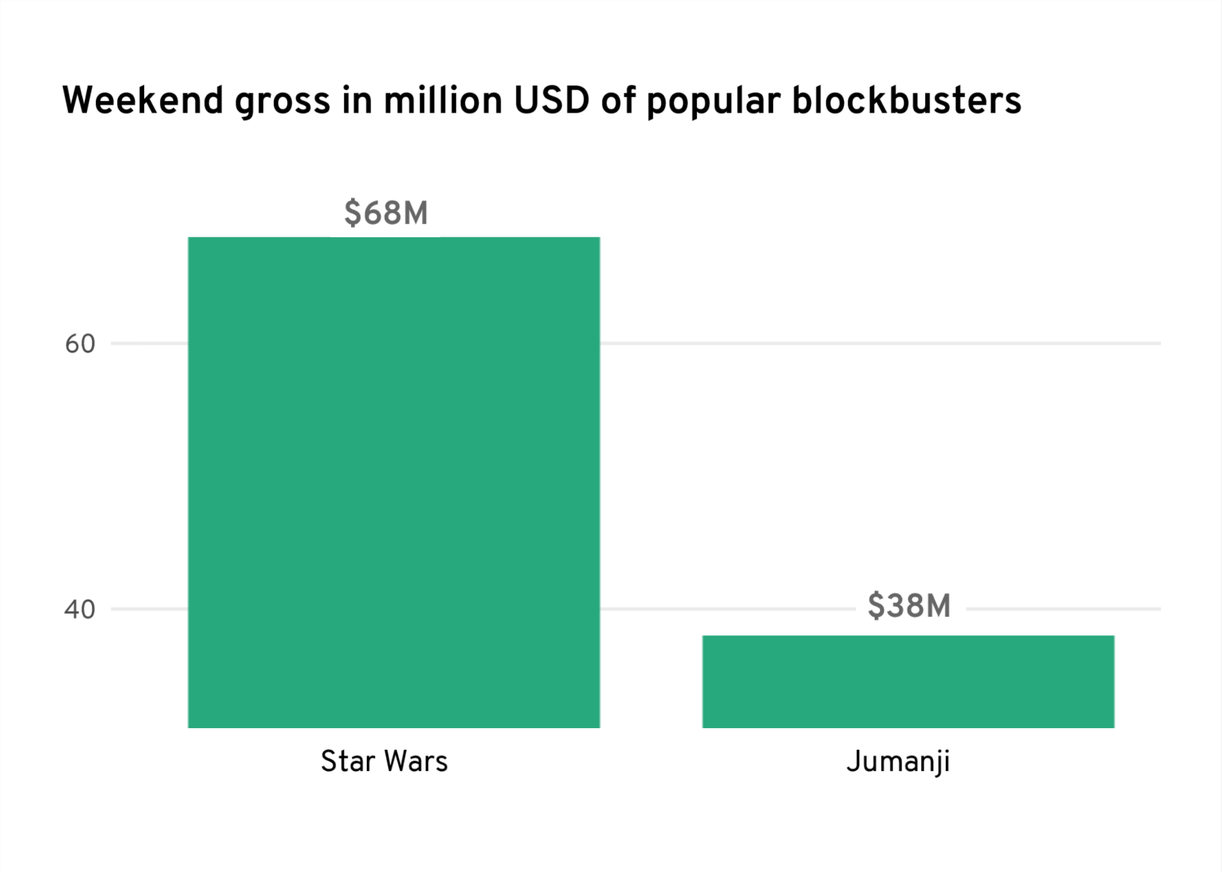


Don’t Tilt Labels

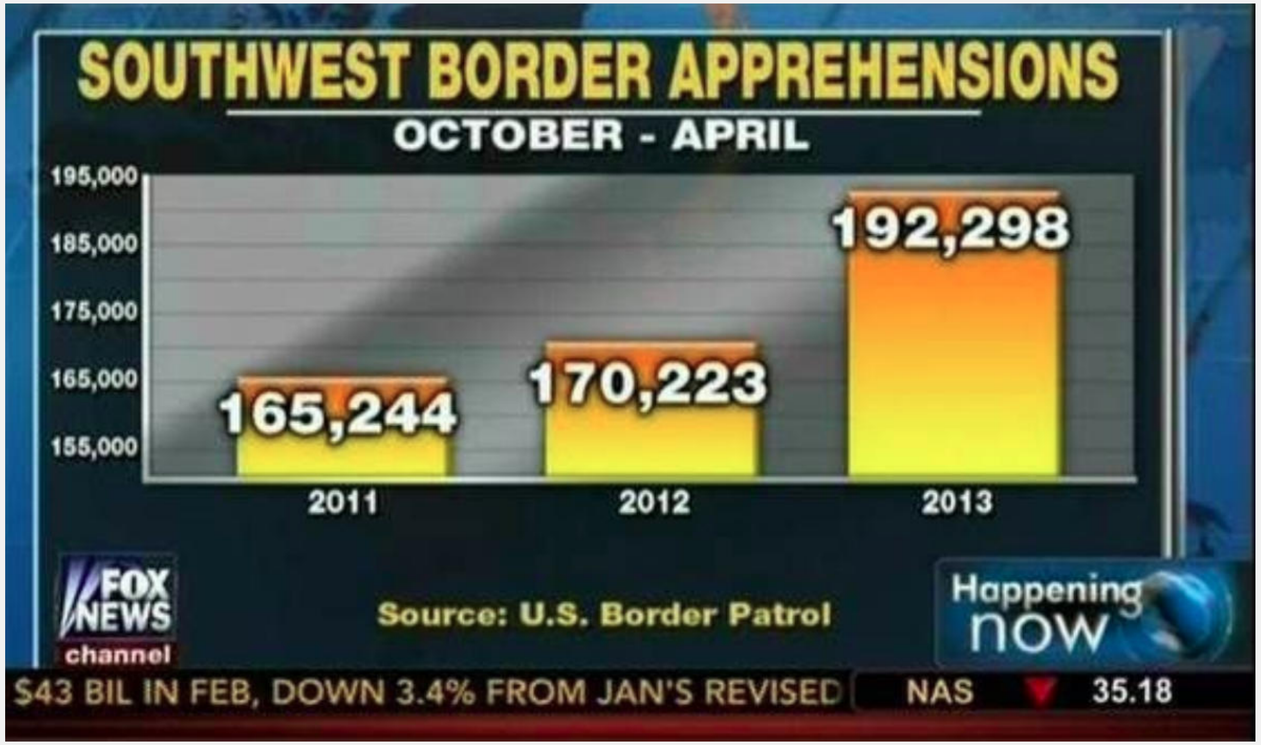




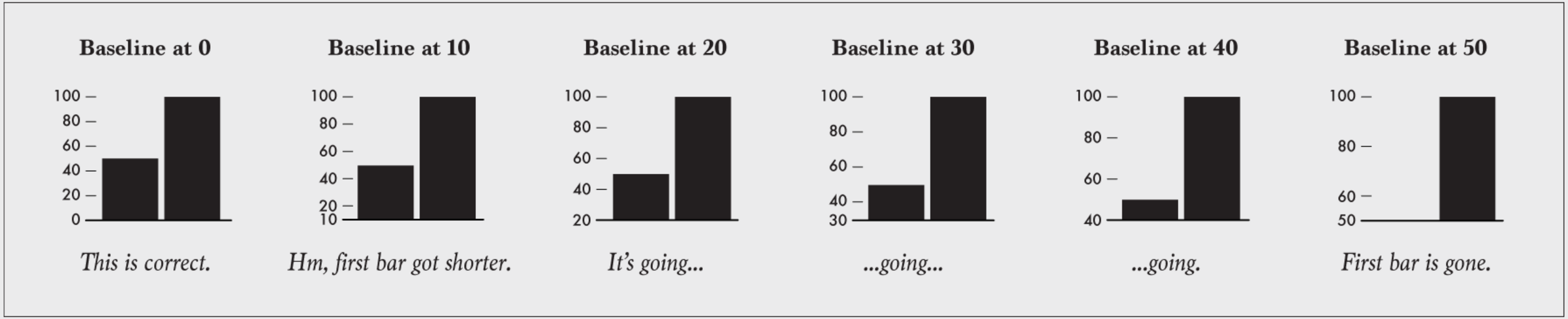
Always Stat at Zero



Always Stat at Zero

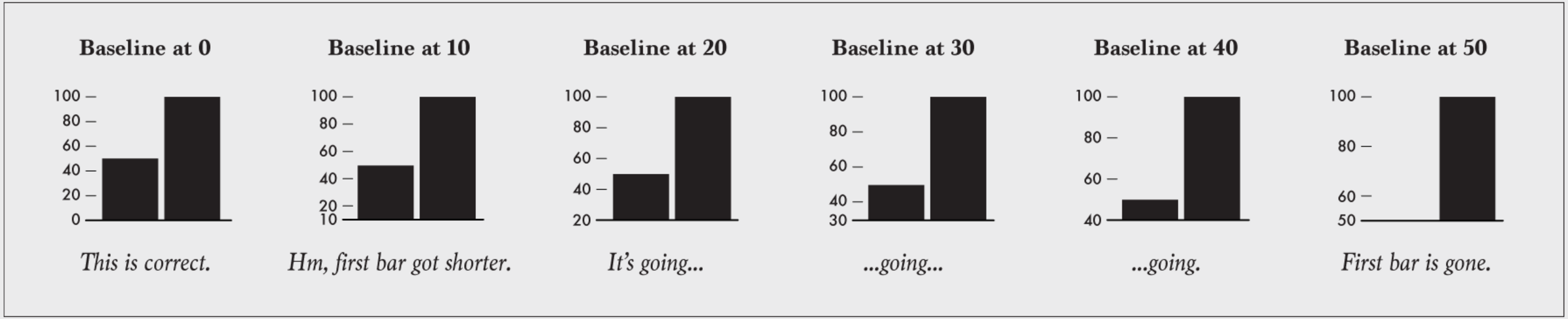


Bar Graphs Should Start at Zero



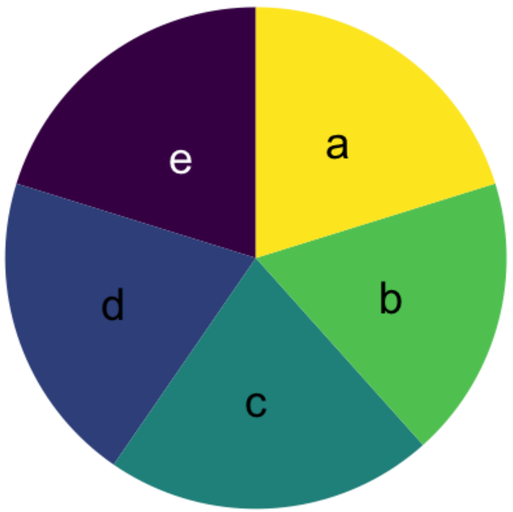
Source: Nathan Yau (flowingdata.com)

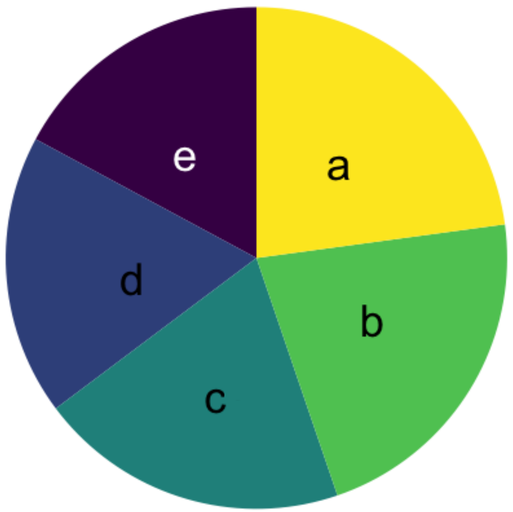
Bar Graphs Should Start at Zero

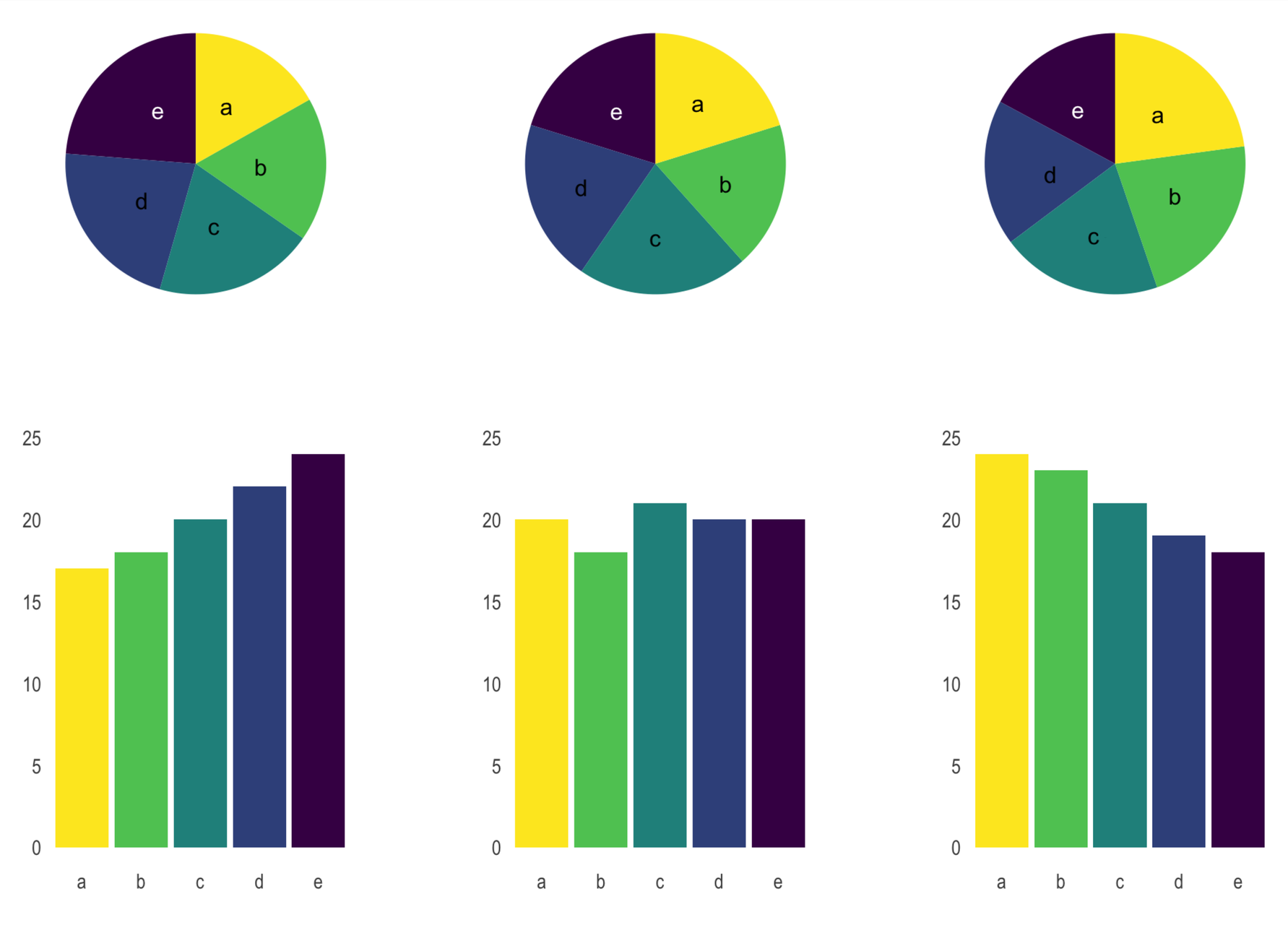


Source: Nathan Yau (flowingdata.com)

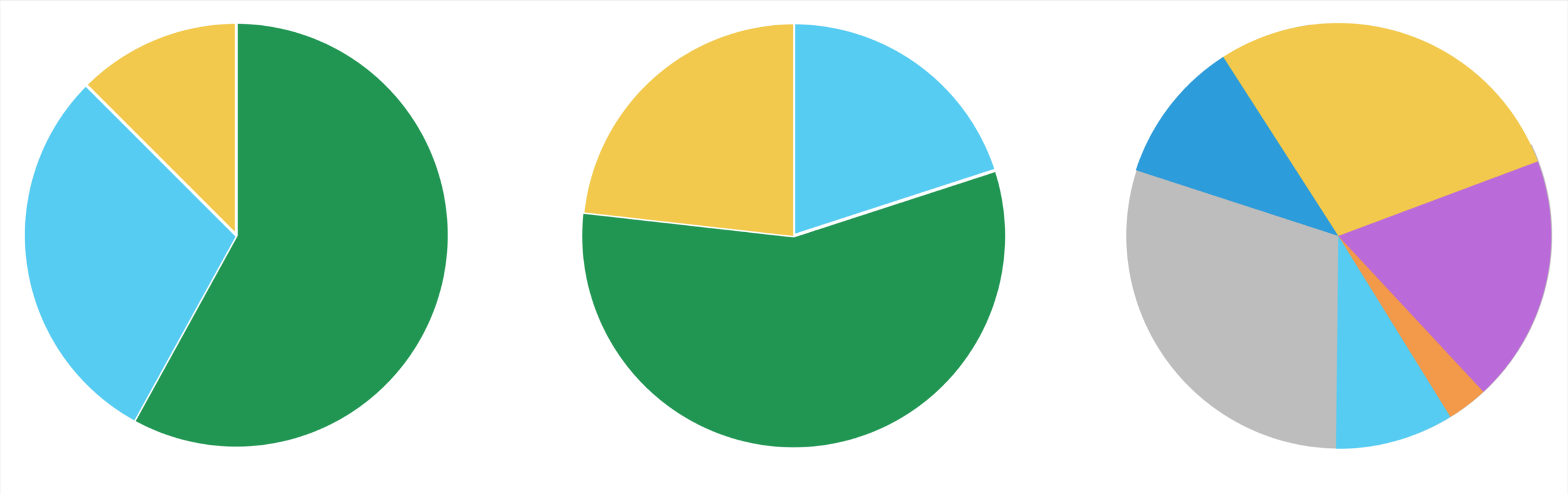
Beware of Pie Charts





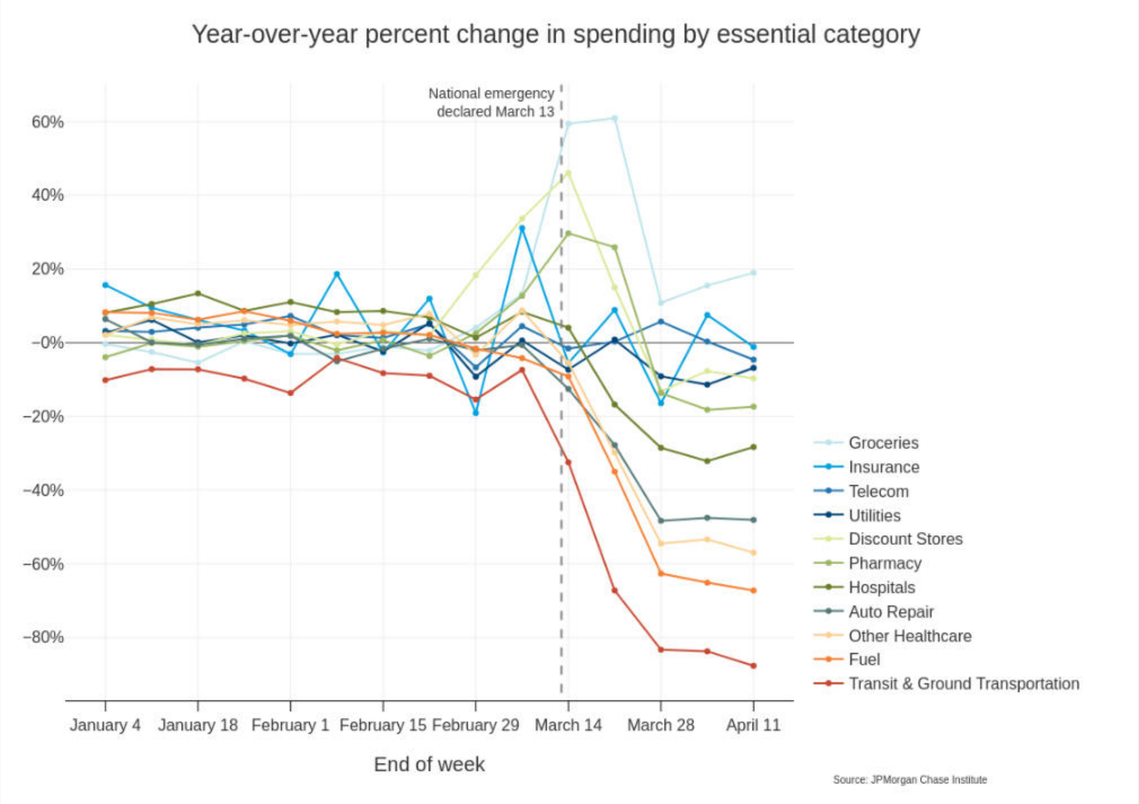

e
a
b
c
d
e
a
b
c
d
e
a
b
c
d
25
20
15
10
5
0
a
b
c
d
e
25
20
15
10
5
0
a
b
c
d
e
25
20
15
10
5
0
a
b
c
d
e

Beware of Pie Charts

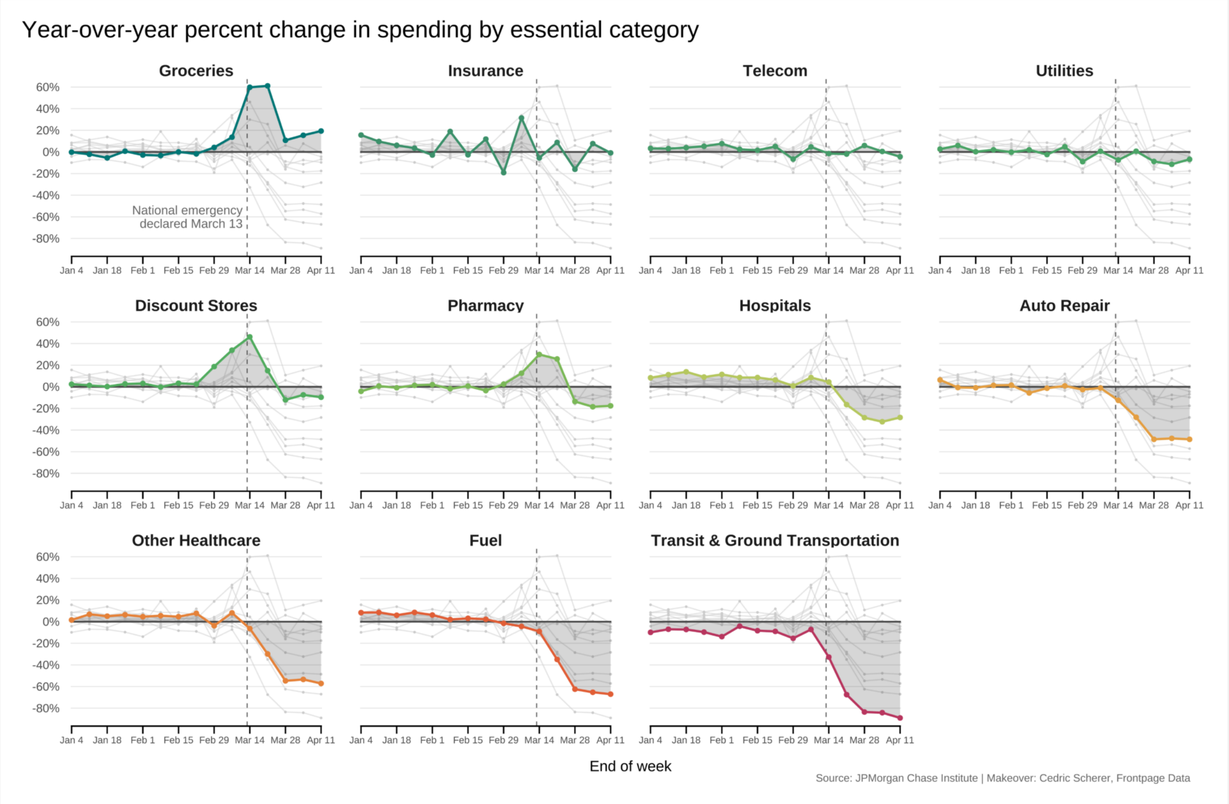


Source: "Hands-On Data Visualization" by Jack Dougherty & Ilya Ilyankou

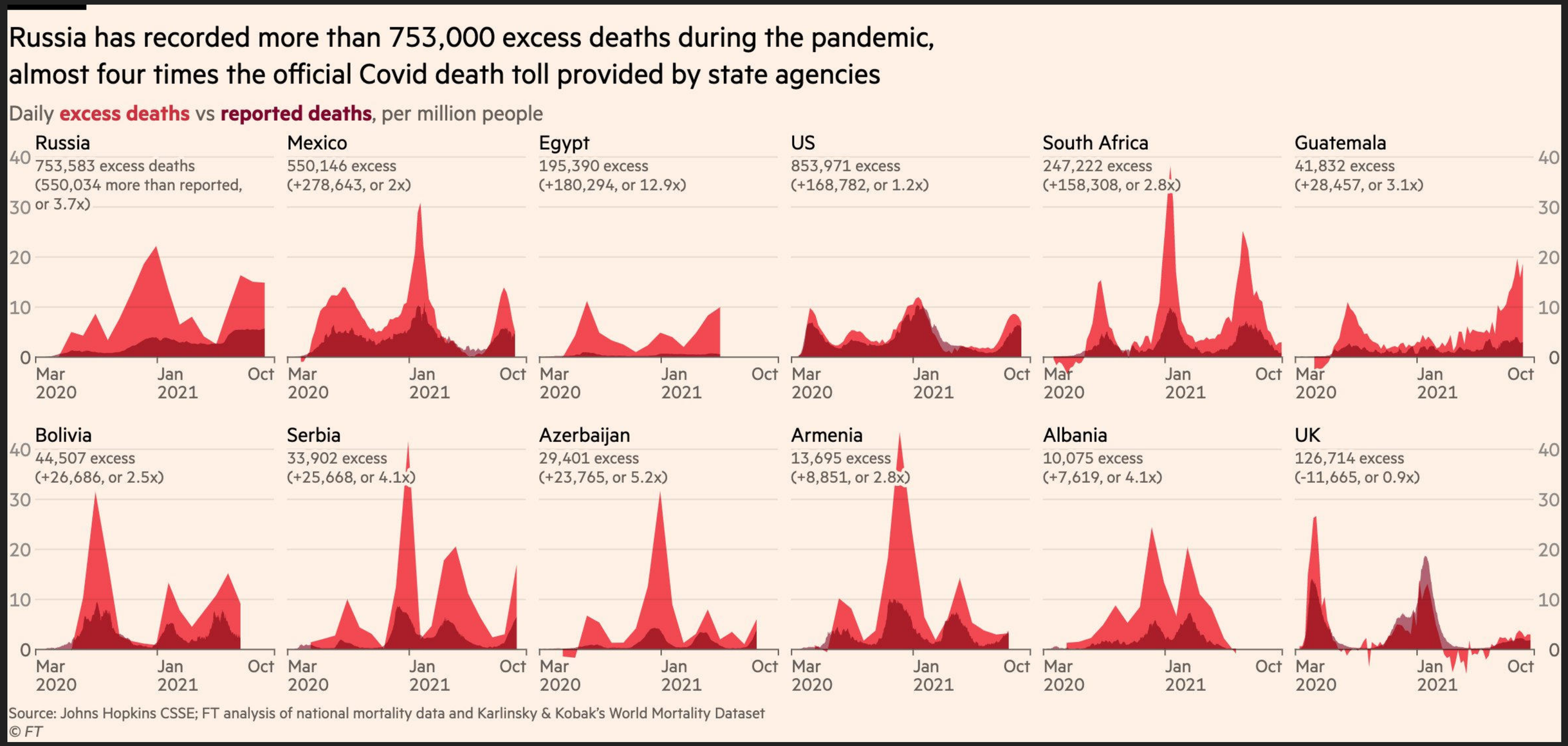
The Power of Small Multiples



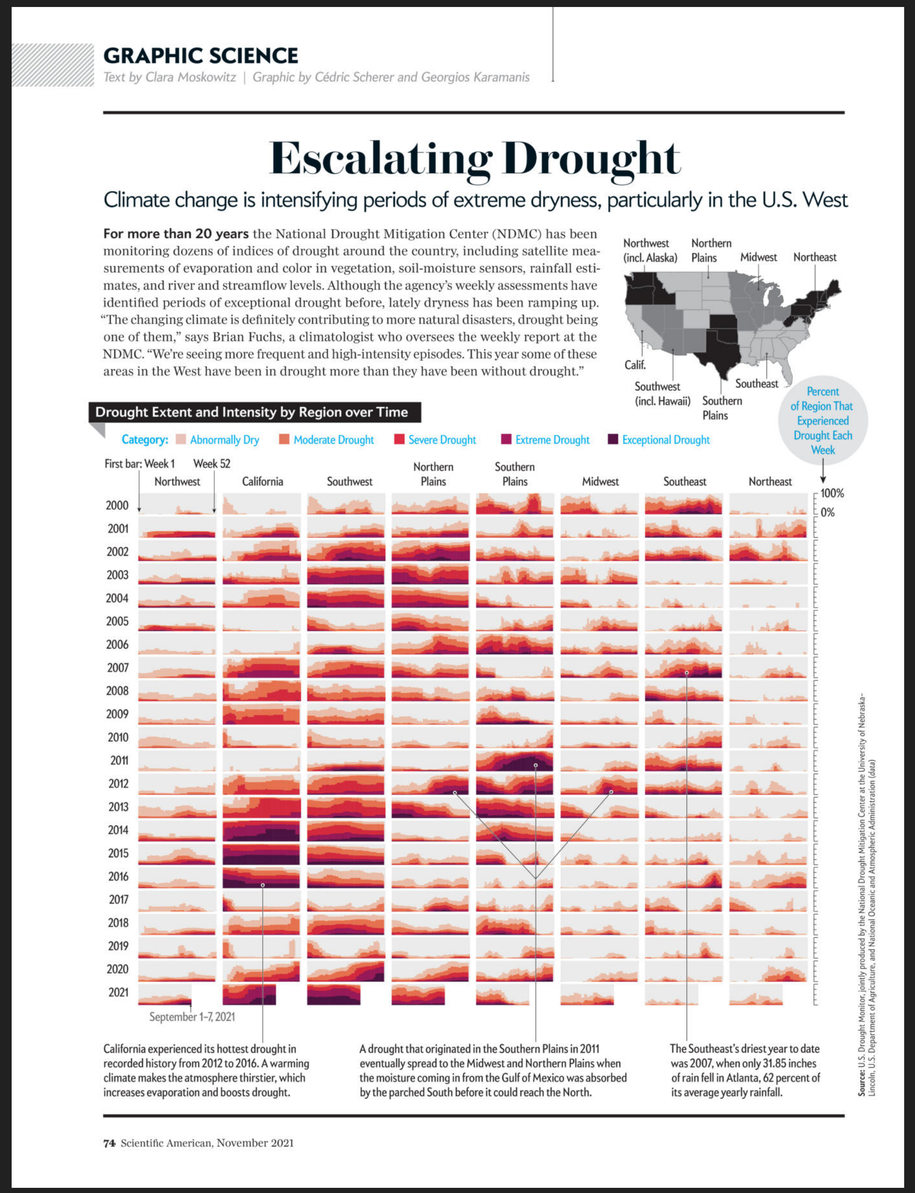
Original graphic by JPMorgan Chase Institute



Makeover using small multiples



Source: John Burn-Murdoch (Financial Times)



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."

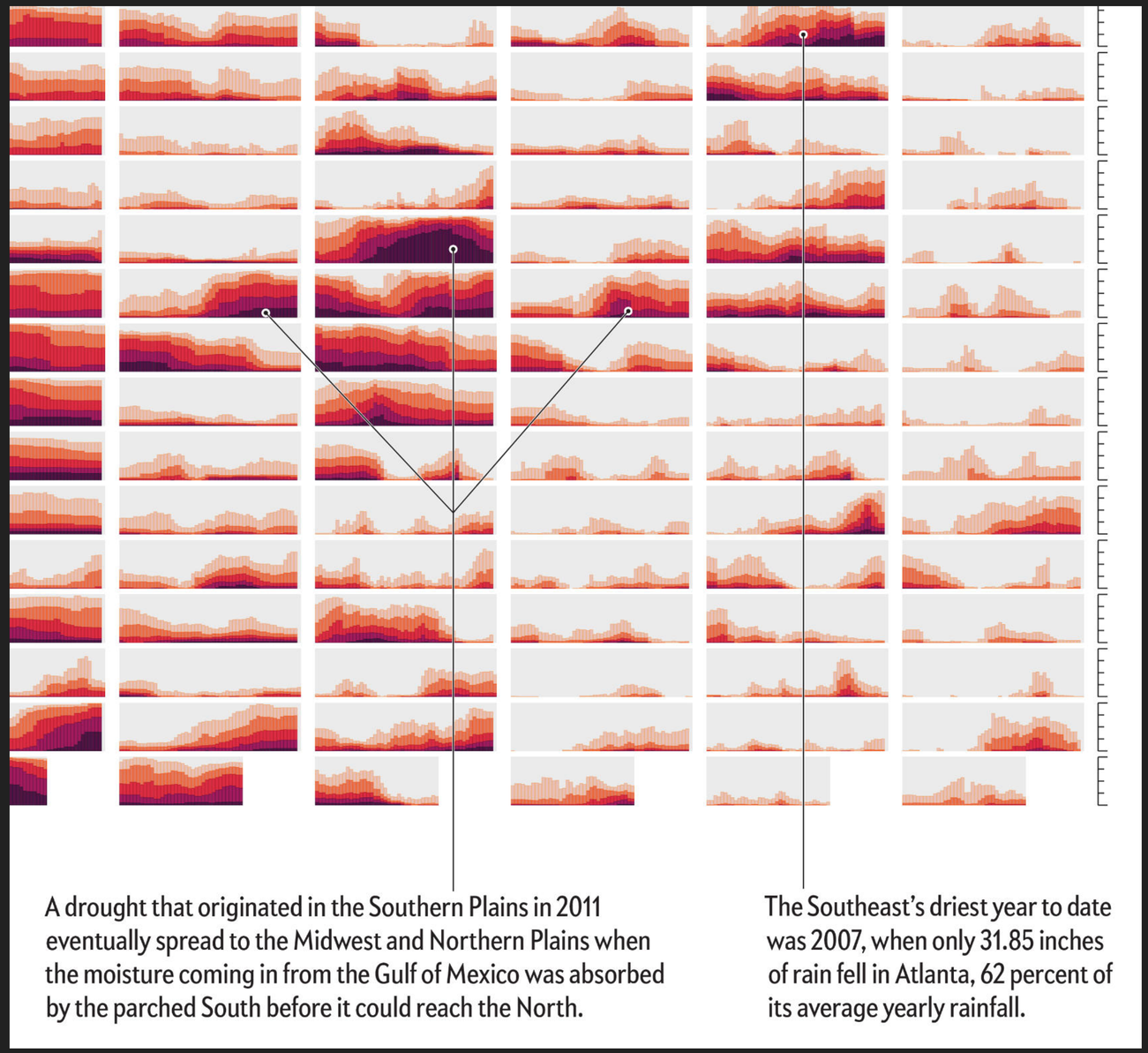
Drought Extent and Intensity by Region over Time

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)



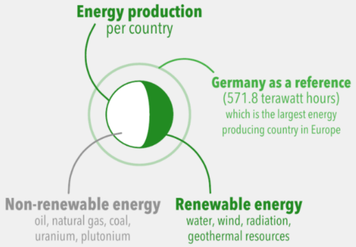
Source: "Escalating Drought", together with Georgios Karamanis for Scientific American, Issue Nov 2021

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.

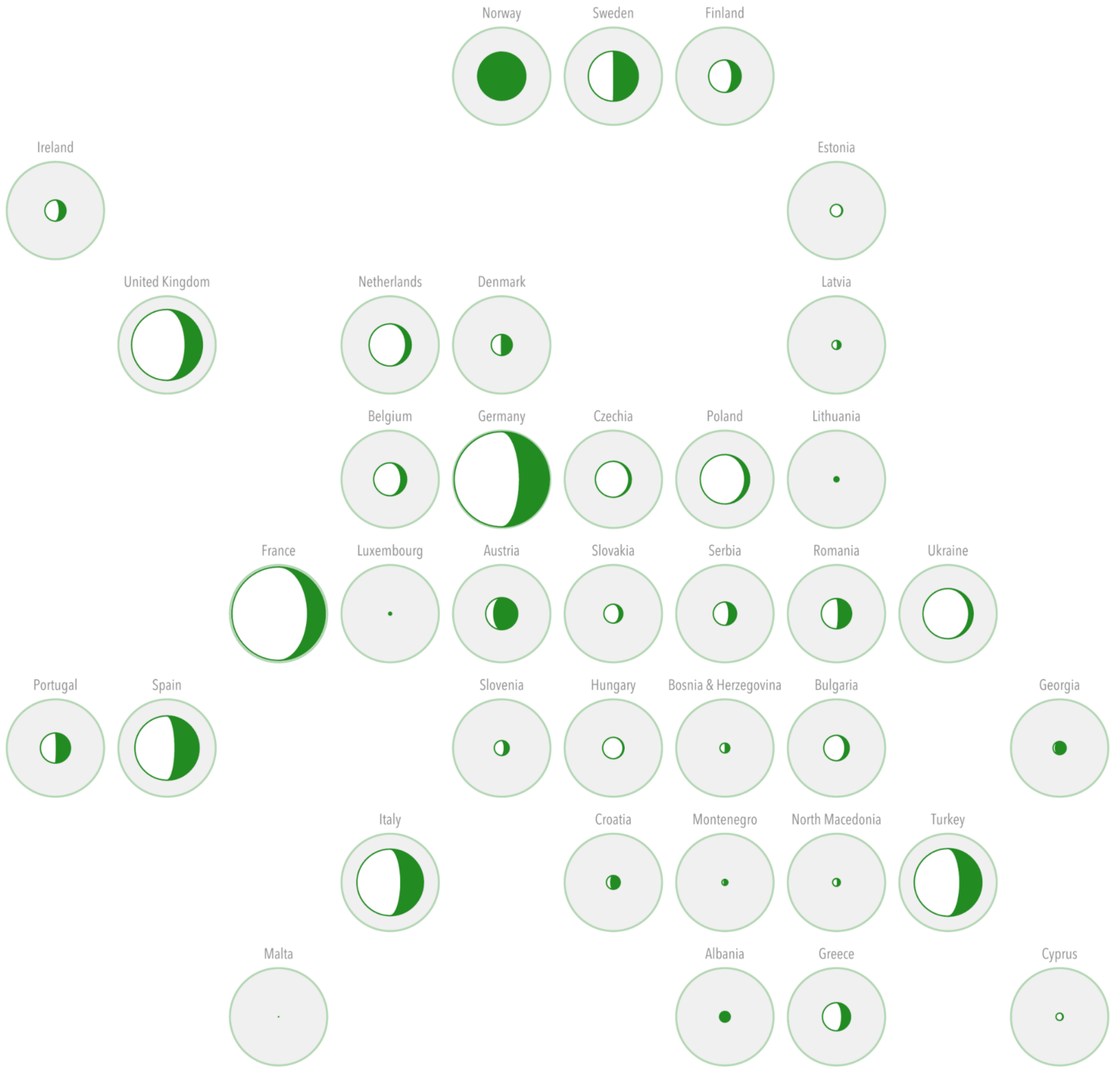
Renewable:	Nuclear:	Conventional thermal:
Germany	**France**	**Germany**
179.1 TWh	393.2 TWh	320.4 TWh
(31% of its production)	(71% of its production)	(56% of its production)

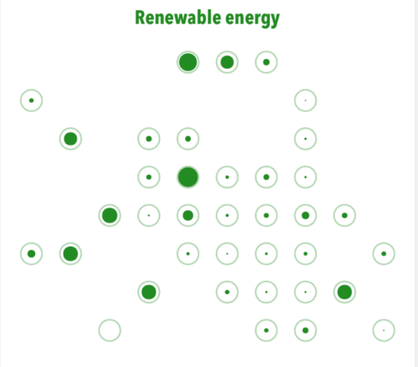
Renewable energy is energy that comes from resources that are naturally replenished such as sunlight, wind, water, and geothermal heat. Unlike fossil fuels, such as oil, natural gas and coal, or nuclear power sources such as uranium and plutonium, renewable energy regenerates naturally in a short period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

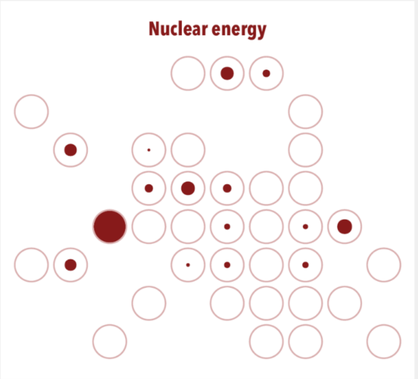


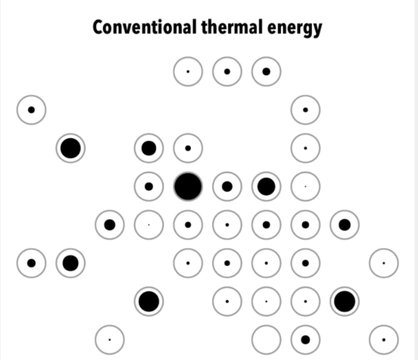
Norway had an electricity production almost entirely made up of renewable energy (98%). This makes Norway the second largest producer of this energy type in Europe. Interestingly, most of the renewable energy is produced by hydro power that take up 95% and only 3% by wind. In contrast, twelve European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

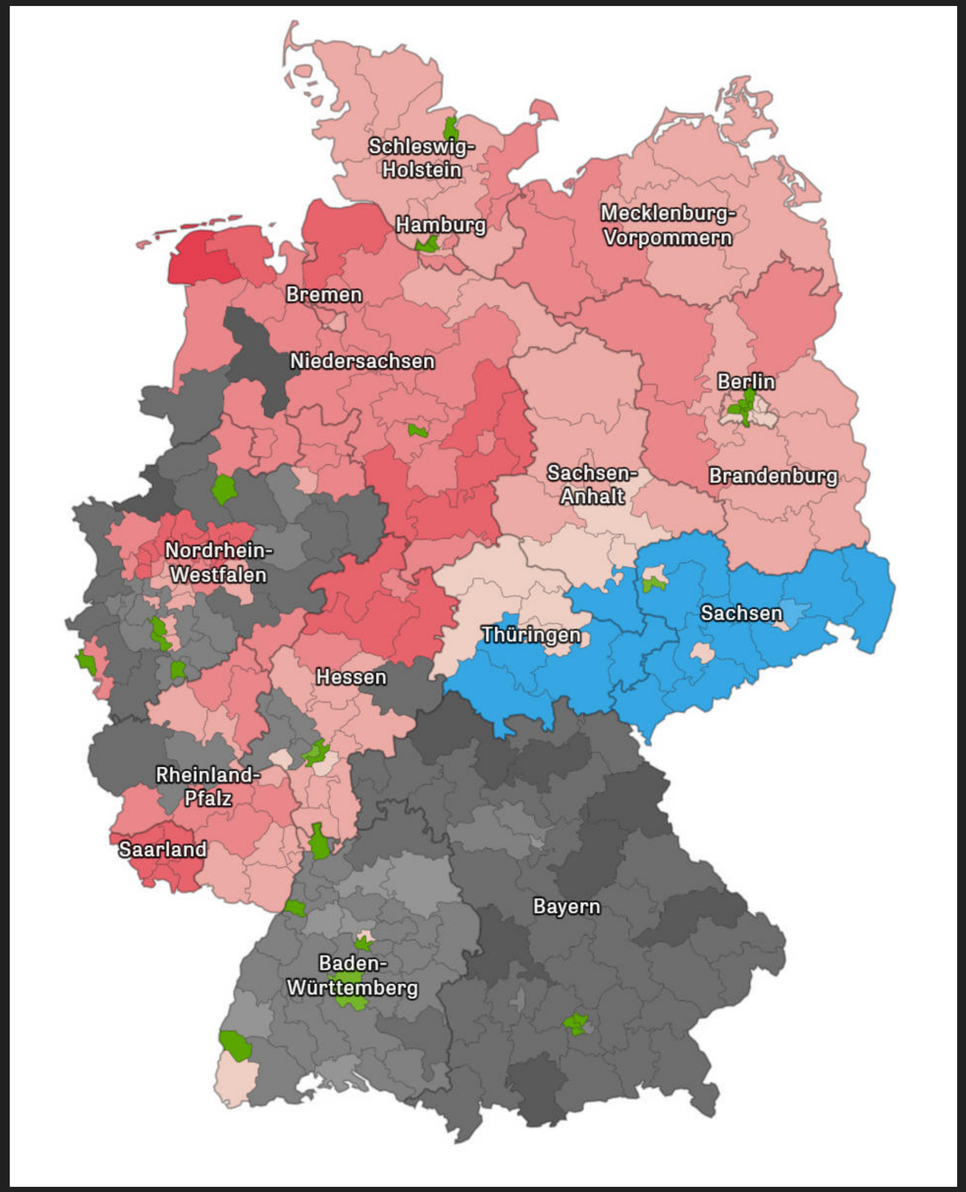


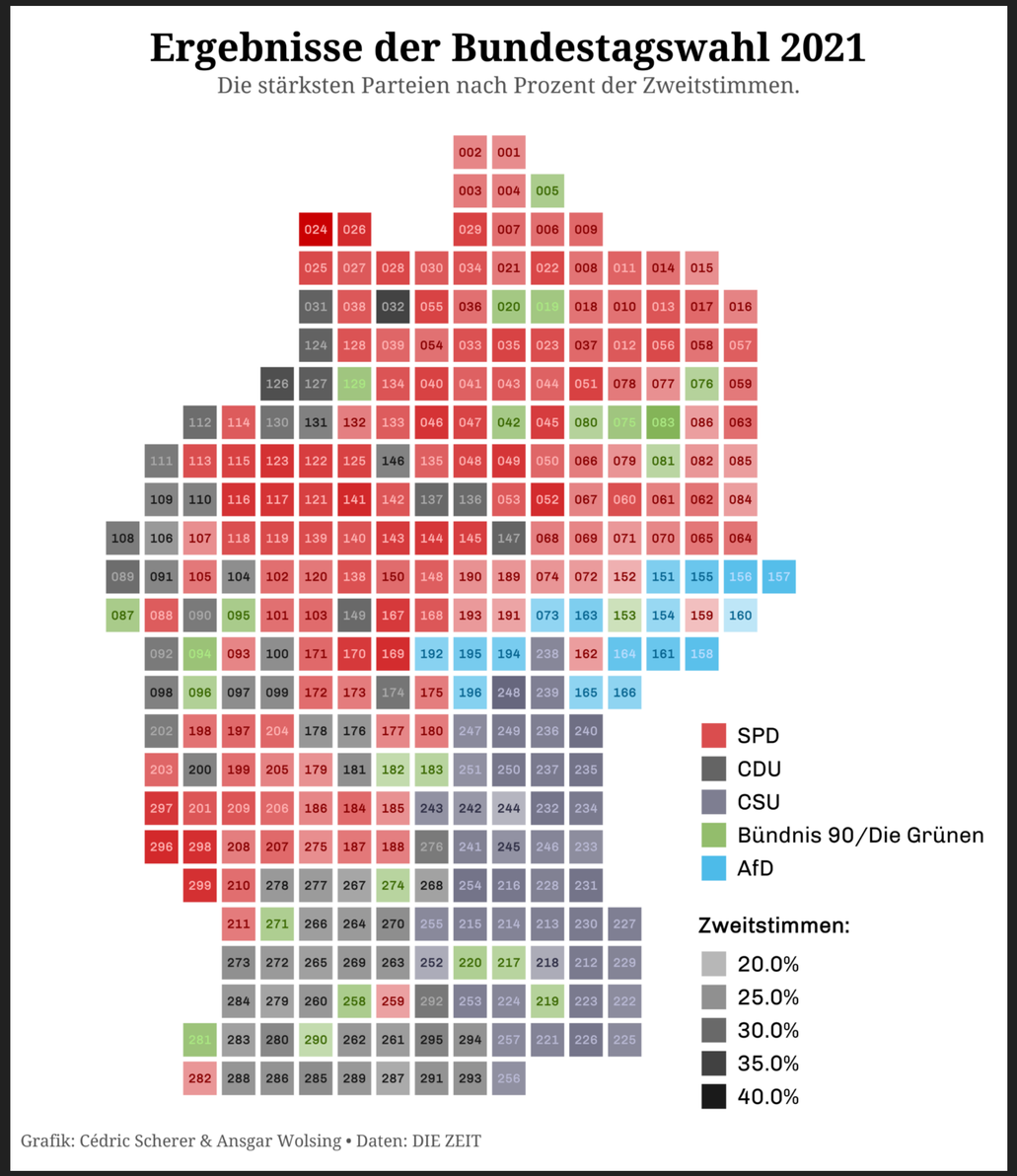






Source: "How European countries generated electricity in 2018", #TidyTuesday Contribution



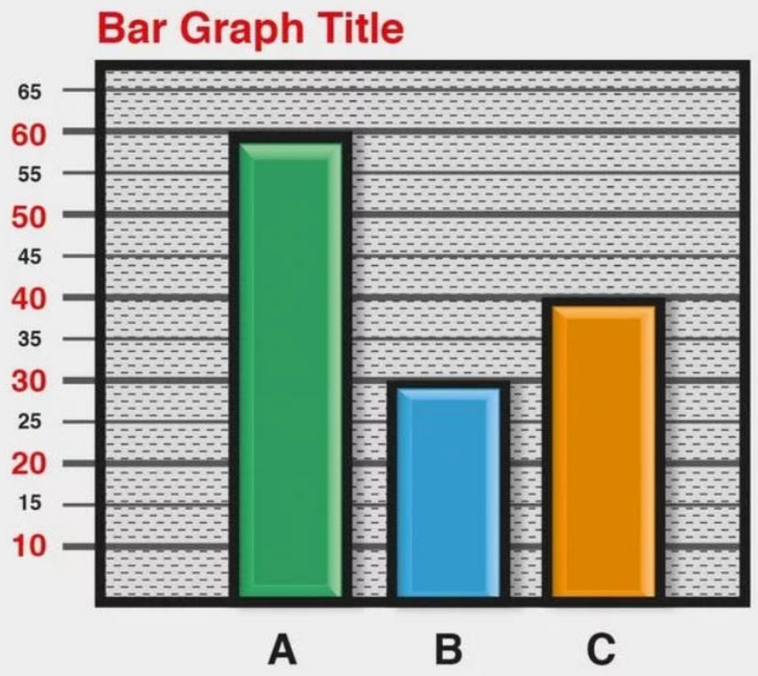


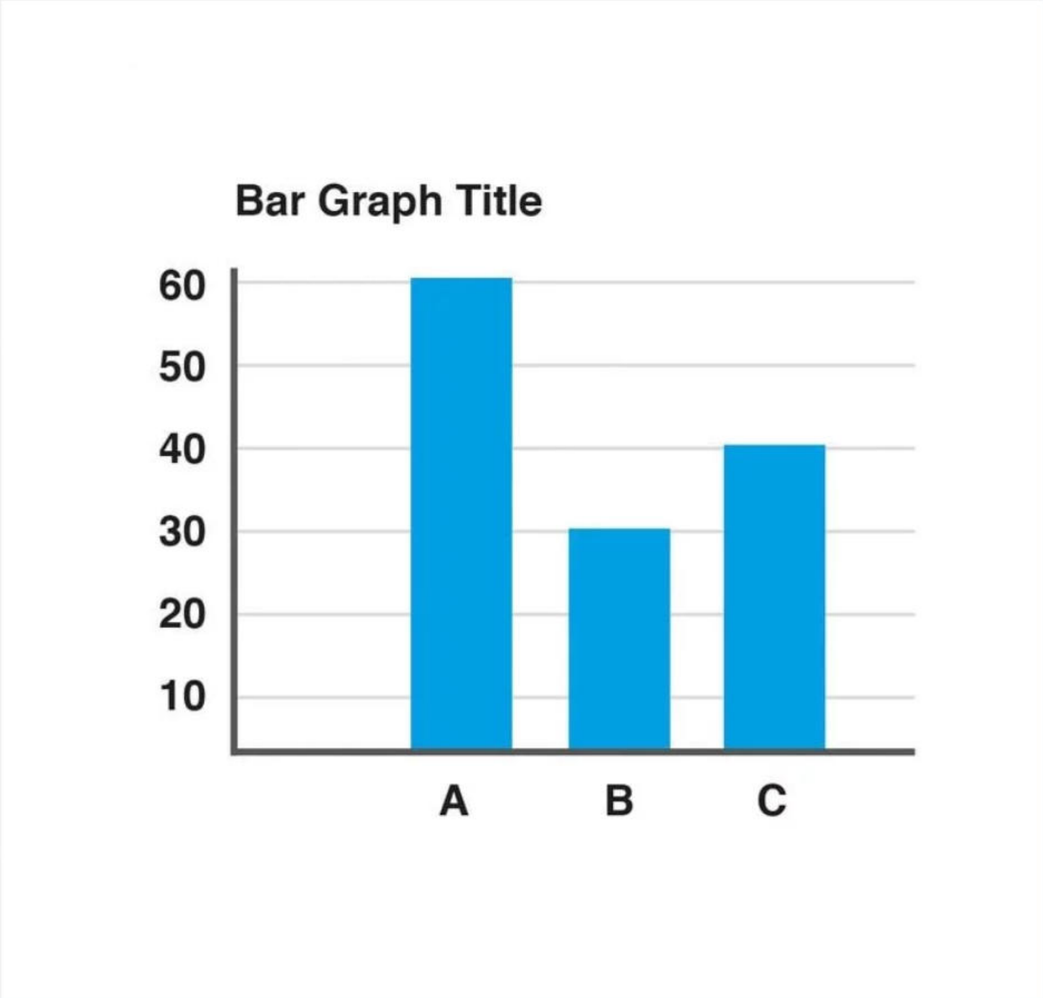
Source: Choropleth Map by Die Zeit (left) | Tile Grid Map by Cédric Scherer & Ansgar Wolsing (right)

VISUAL FORM

Present information in a logical, coherent way

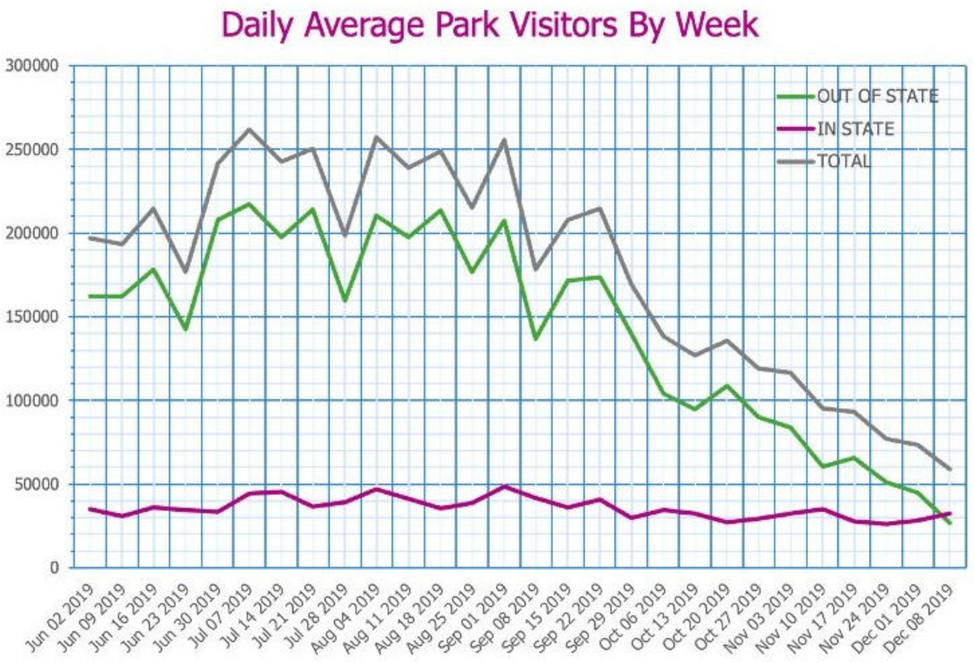
Declutter You Charts



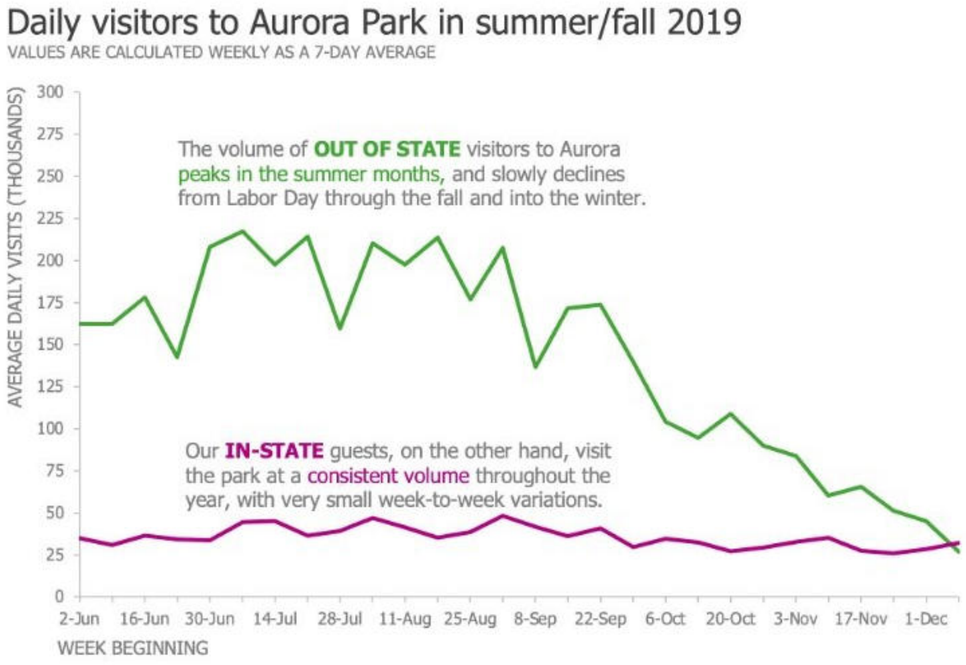


Source: canva.com

BEFORE: SPOOKY SKELETON

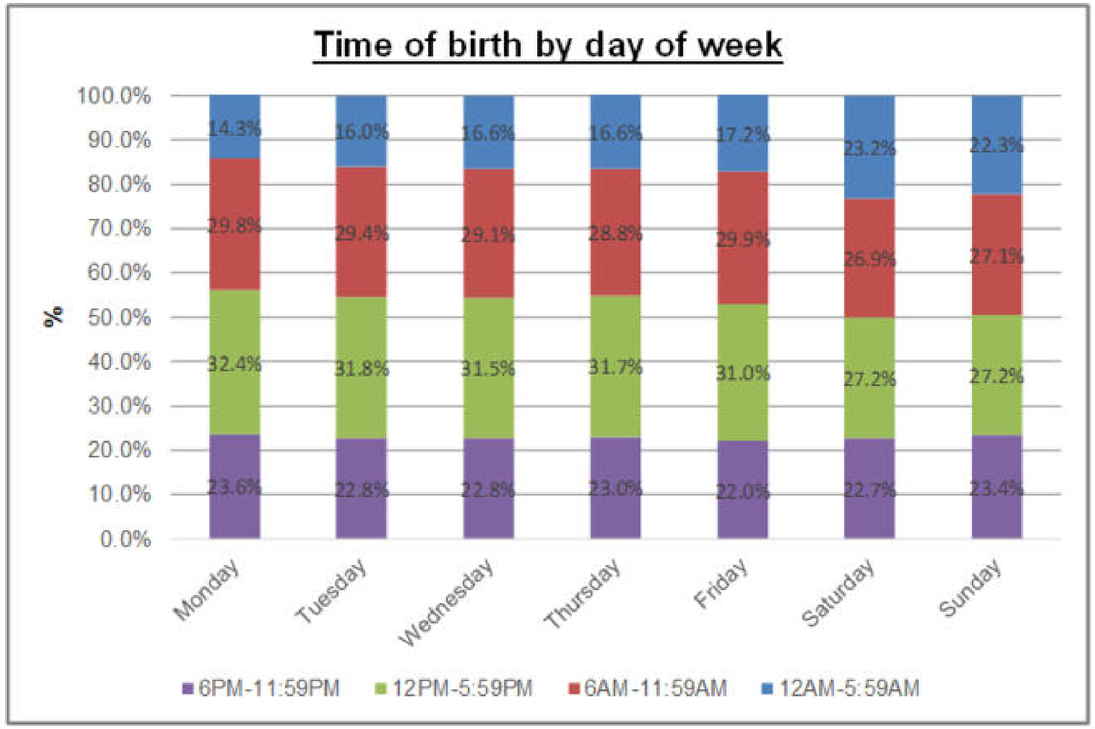


AFTER: GOOD BONES



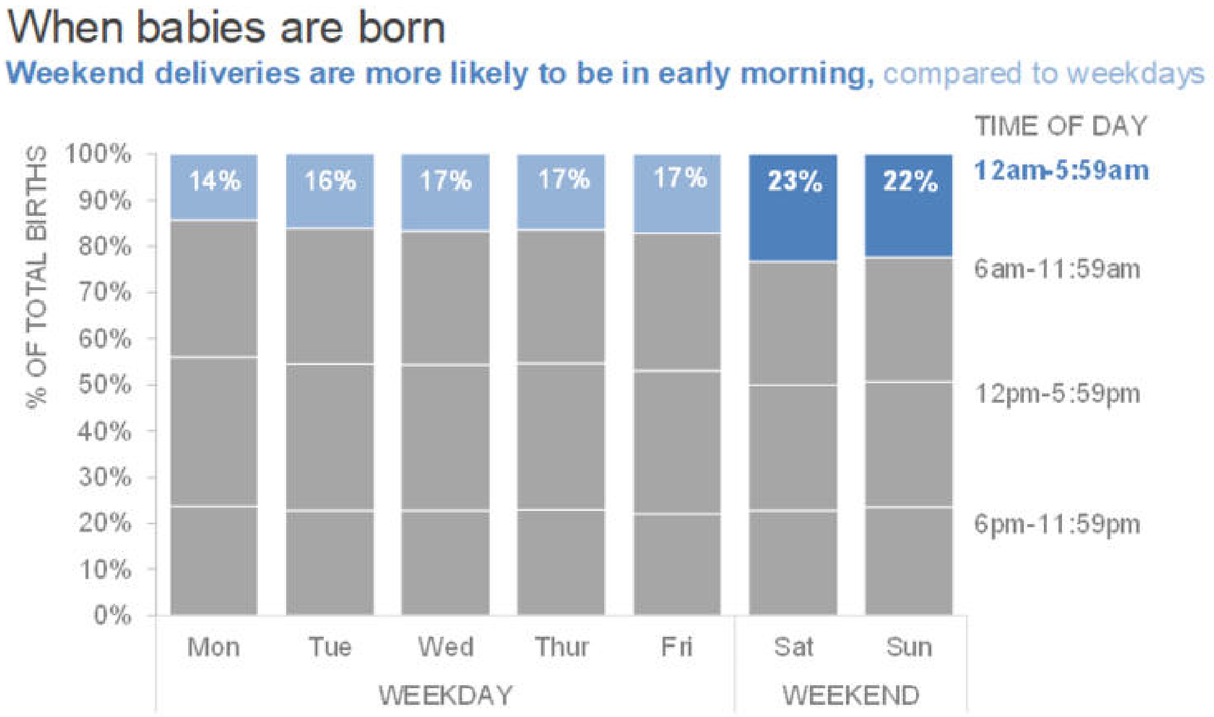
Source: "Your graph Skeleton Shouldn't Be Scray" by Mike Cisneros (Storytelling With Data)

BEFORE



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

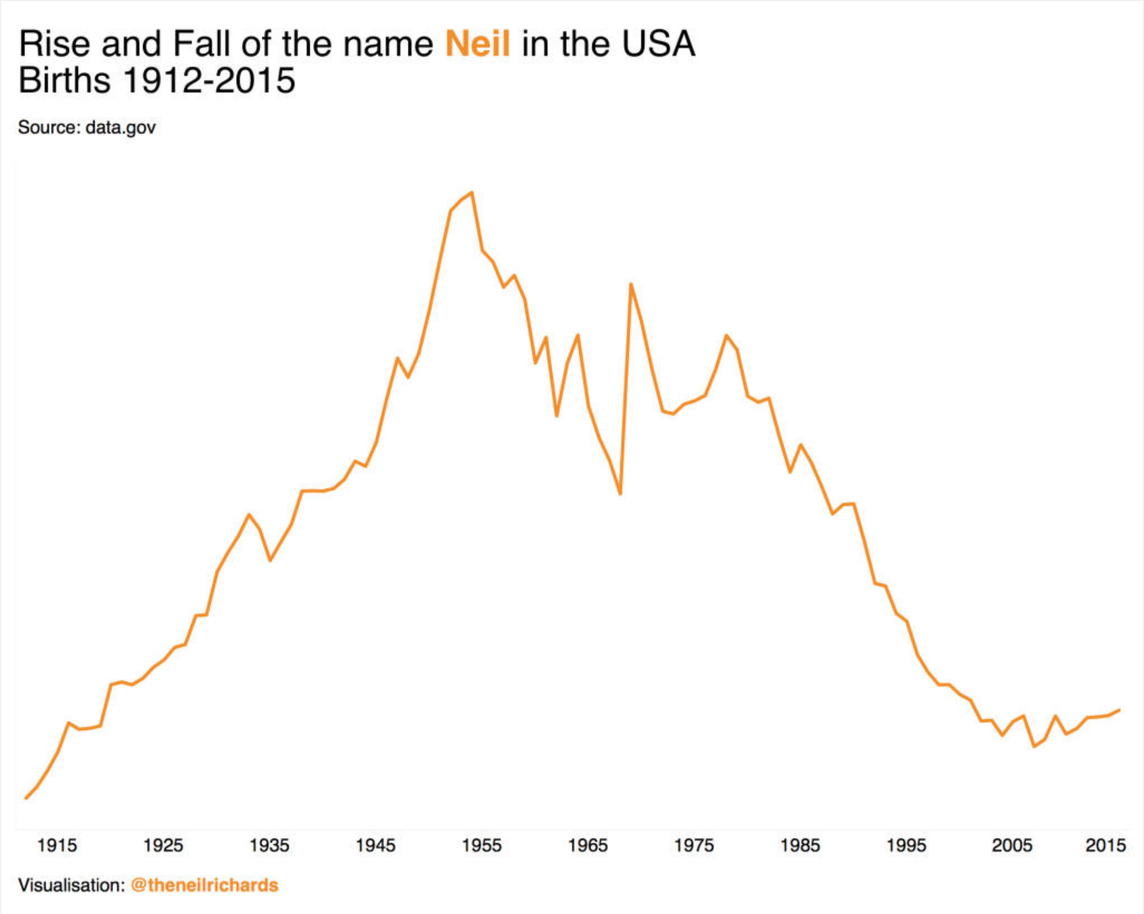
AFTER



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

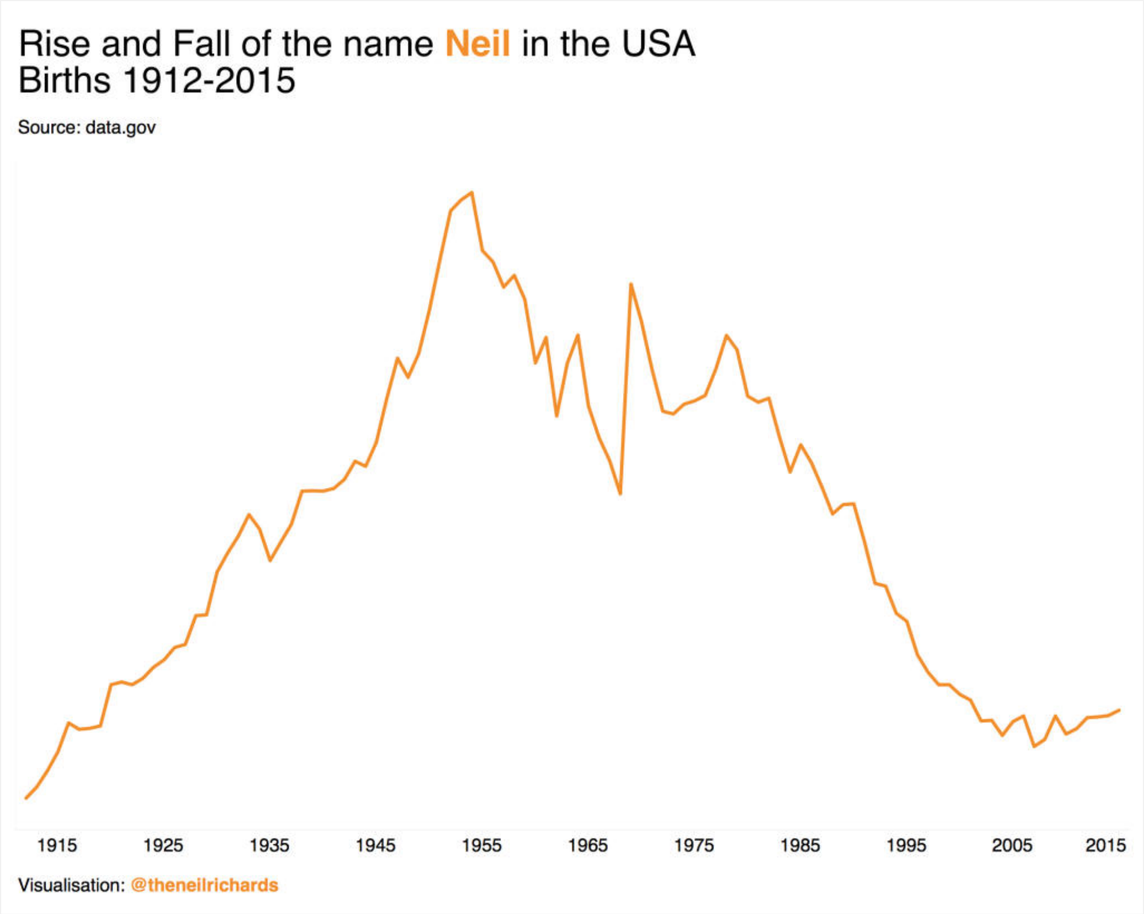
Source: "Declutter! (and Question Default Settings)" by Elizabeth Ricks (Storytelling With Data)

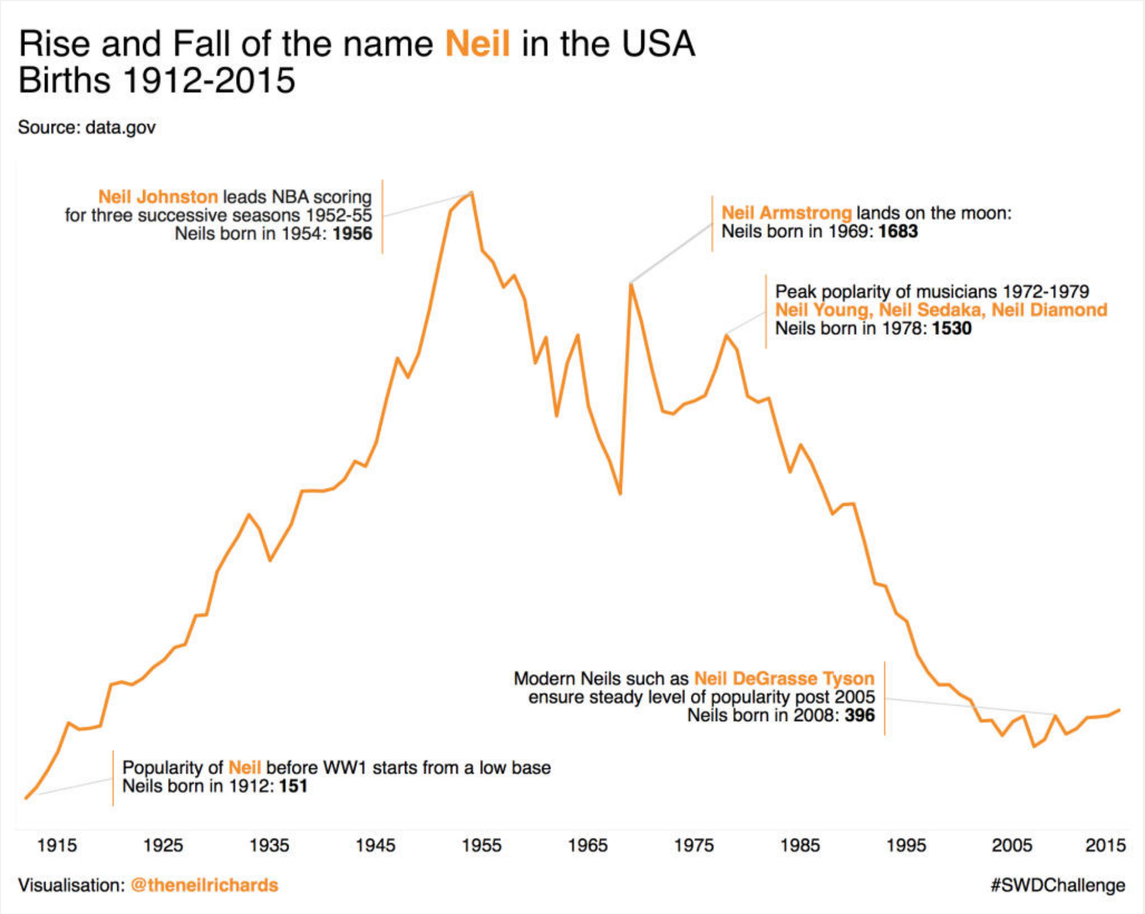
The Power of Annotations



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

The Power of Annotations

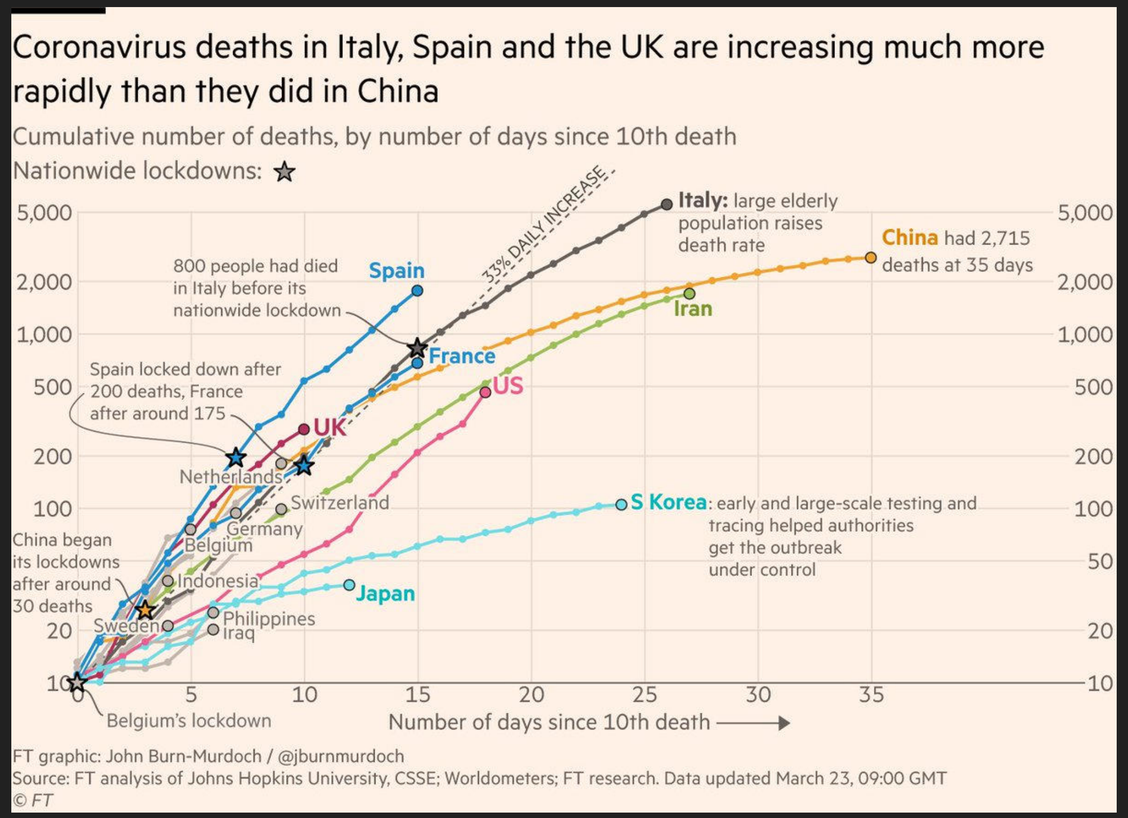


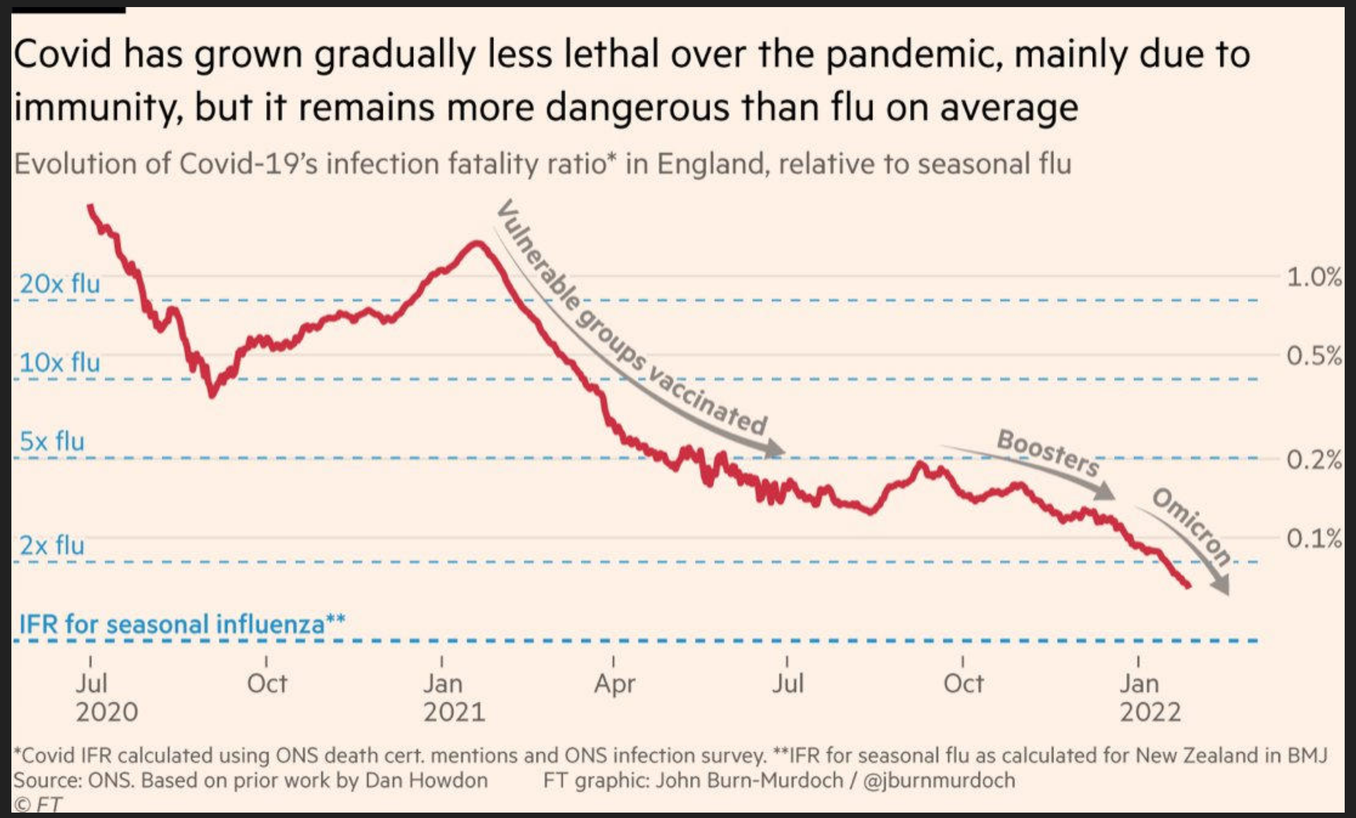


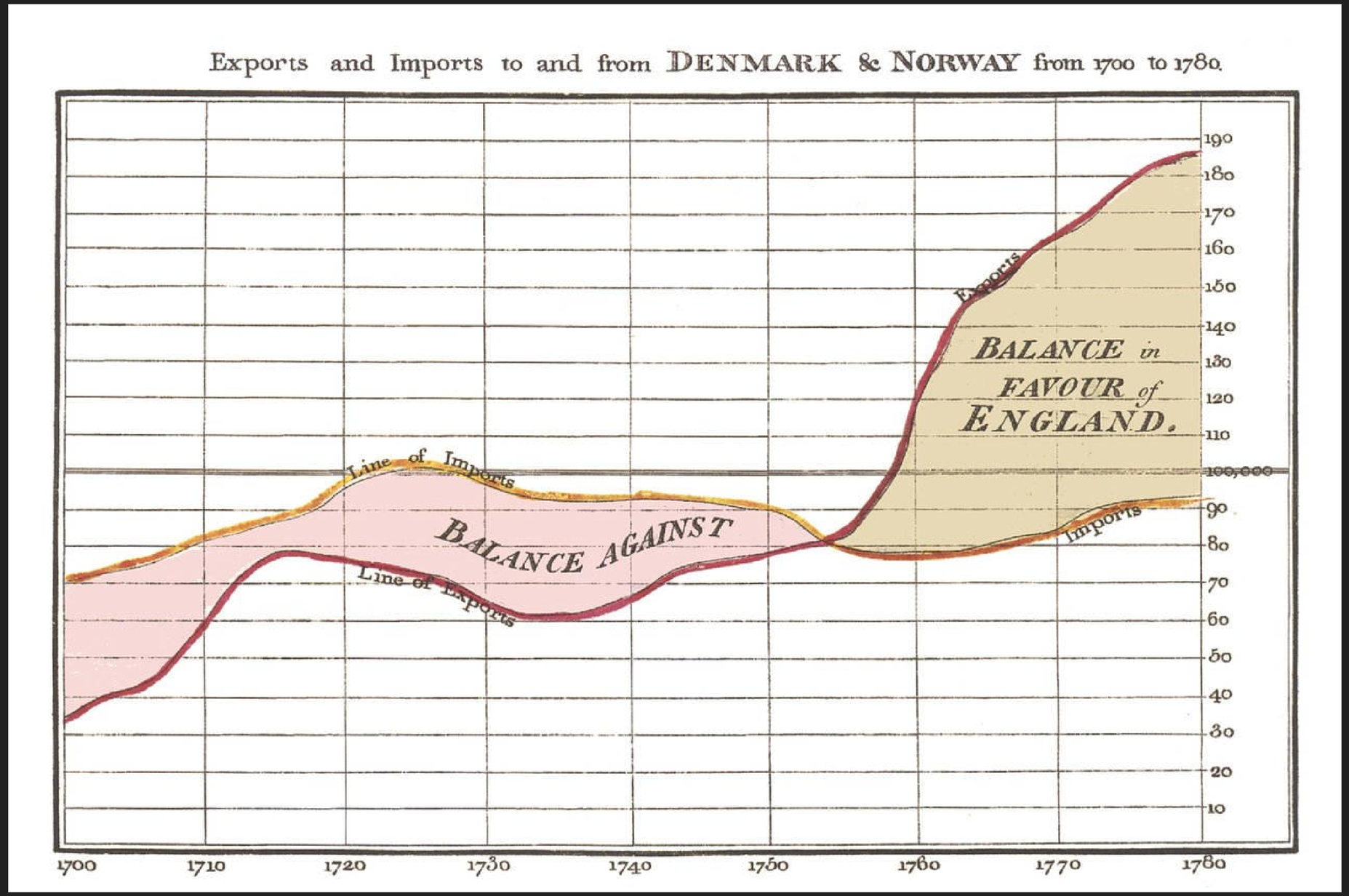
Source: "Is white space always your friend?" by Neil Richards

“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

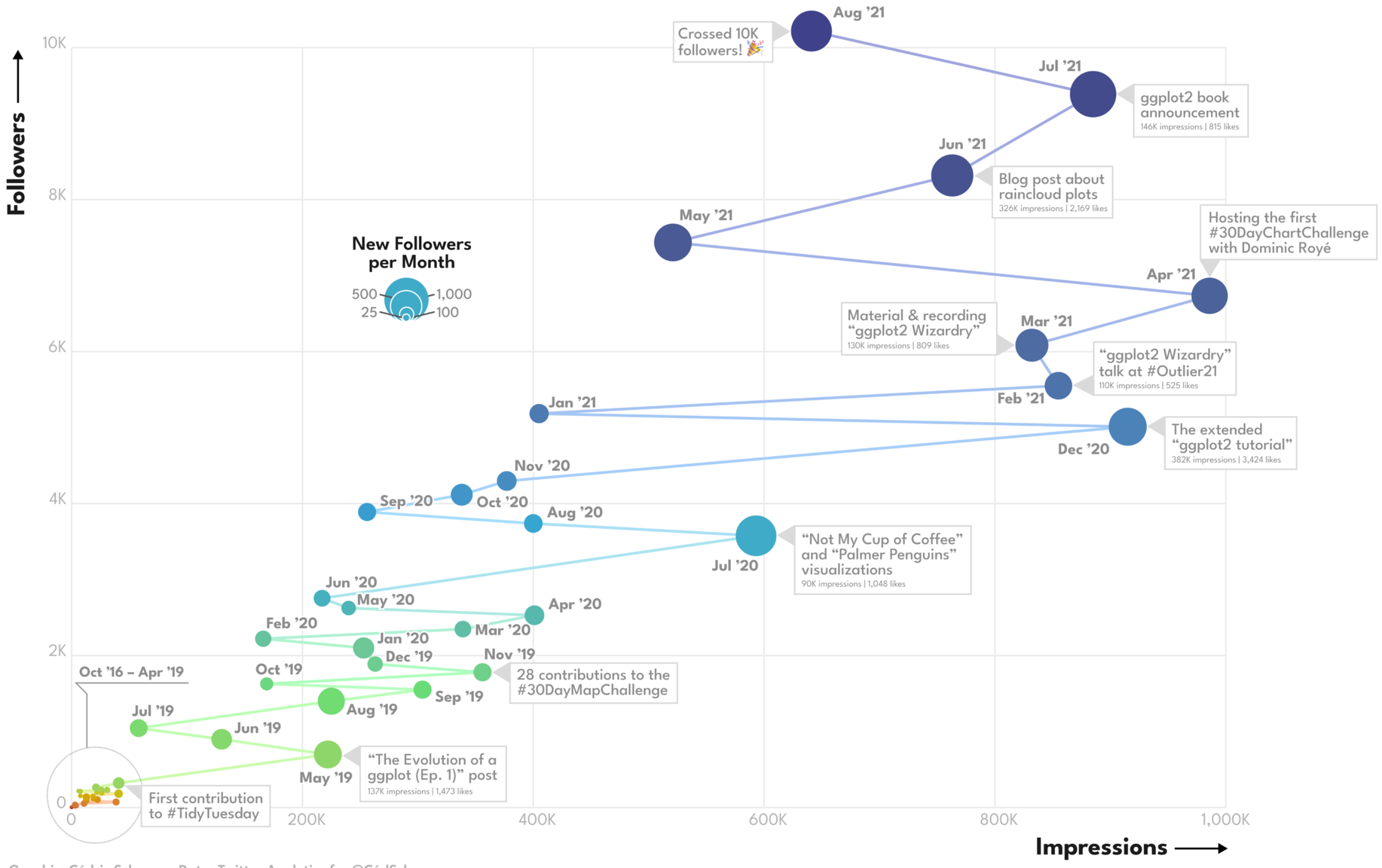




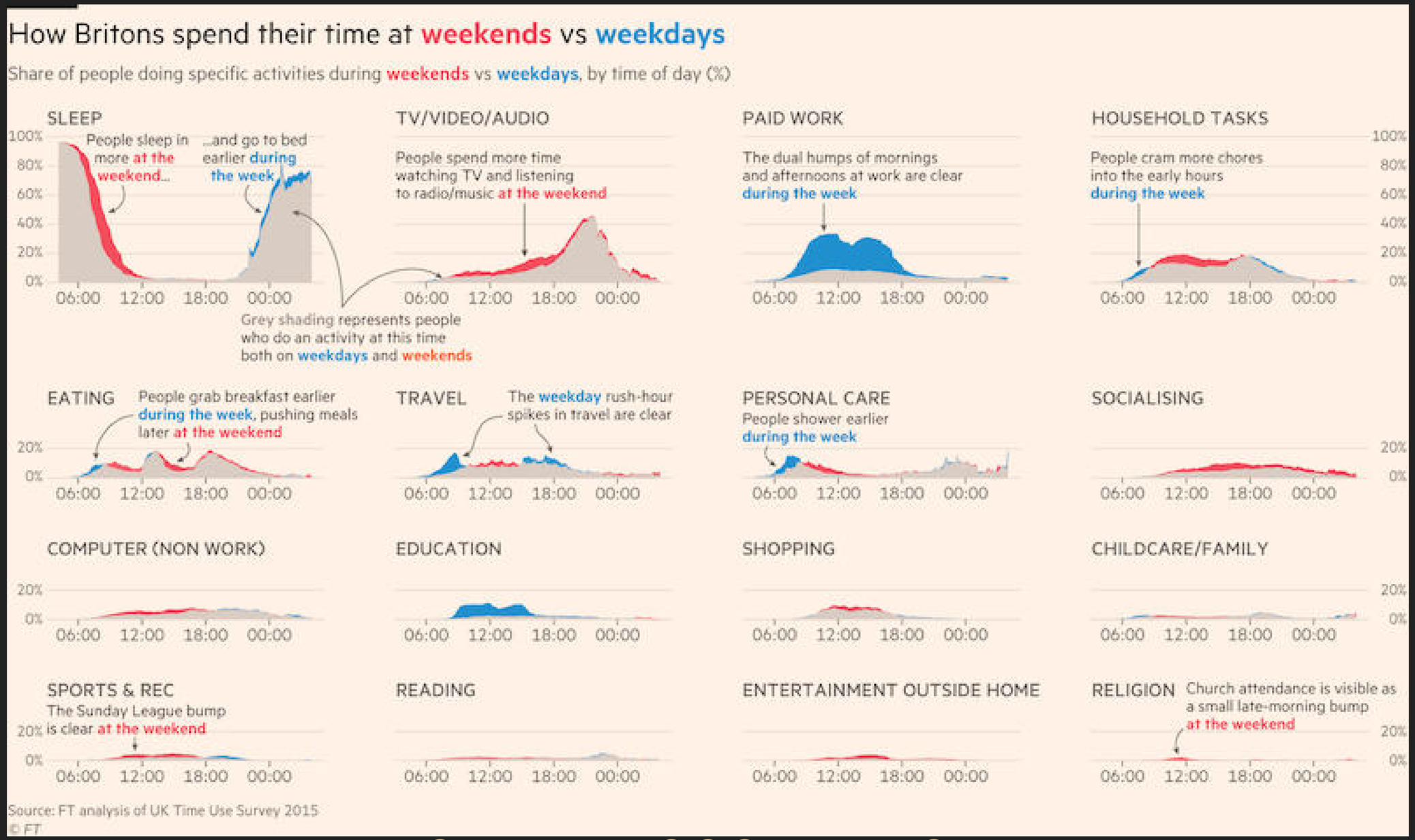


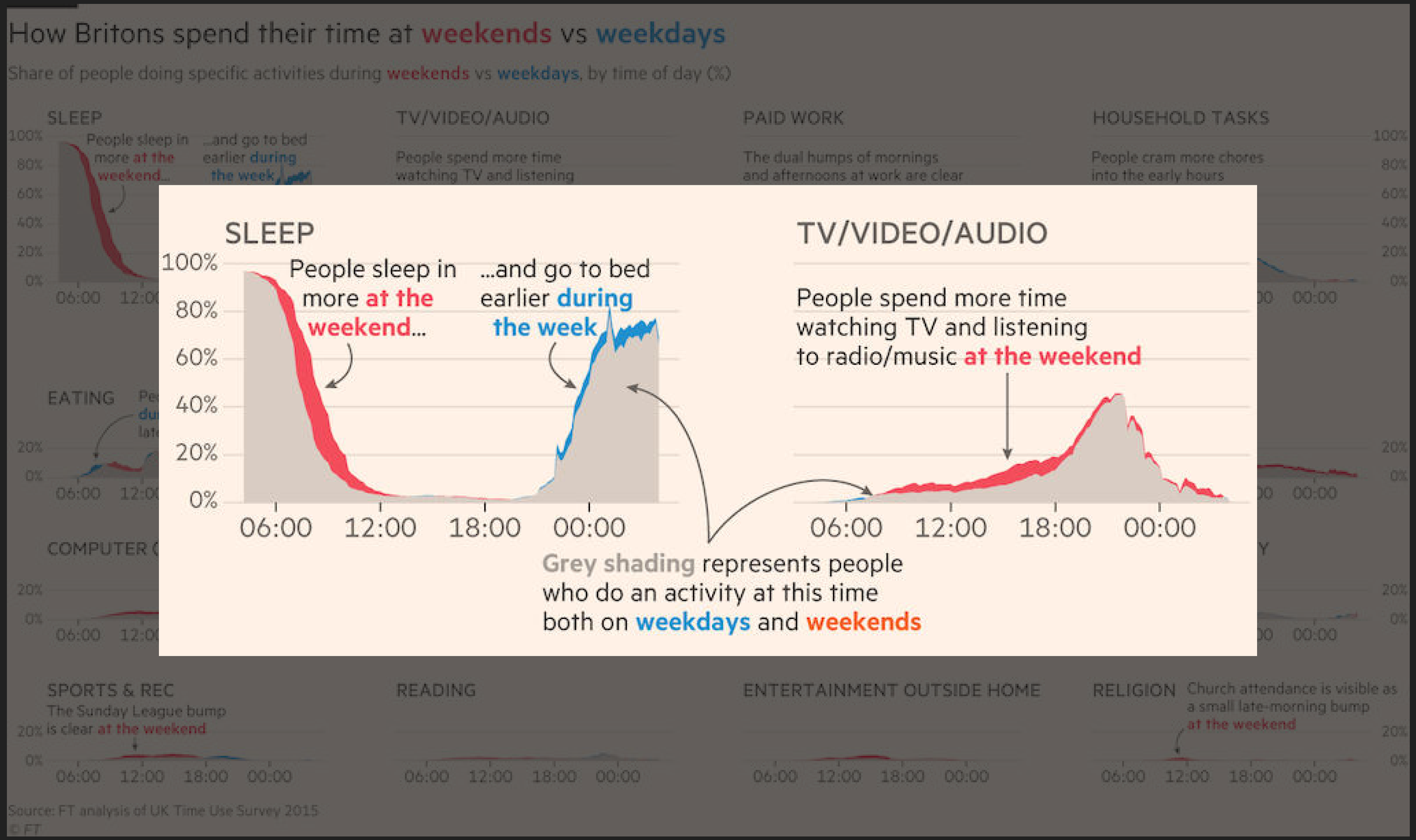
Annotated time-series chart by William Playfair from "The Commercial and Political Atlas and Statistical Breviary" (1786)

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

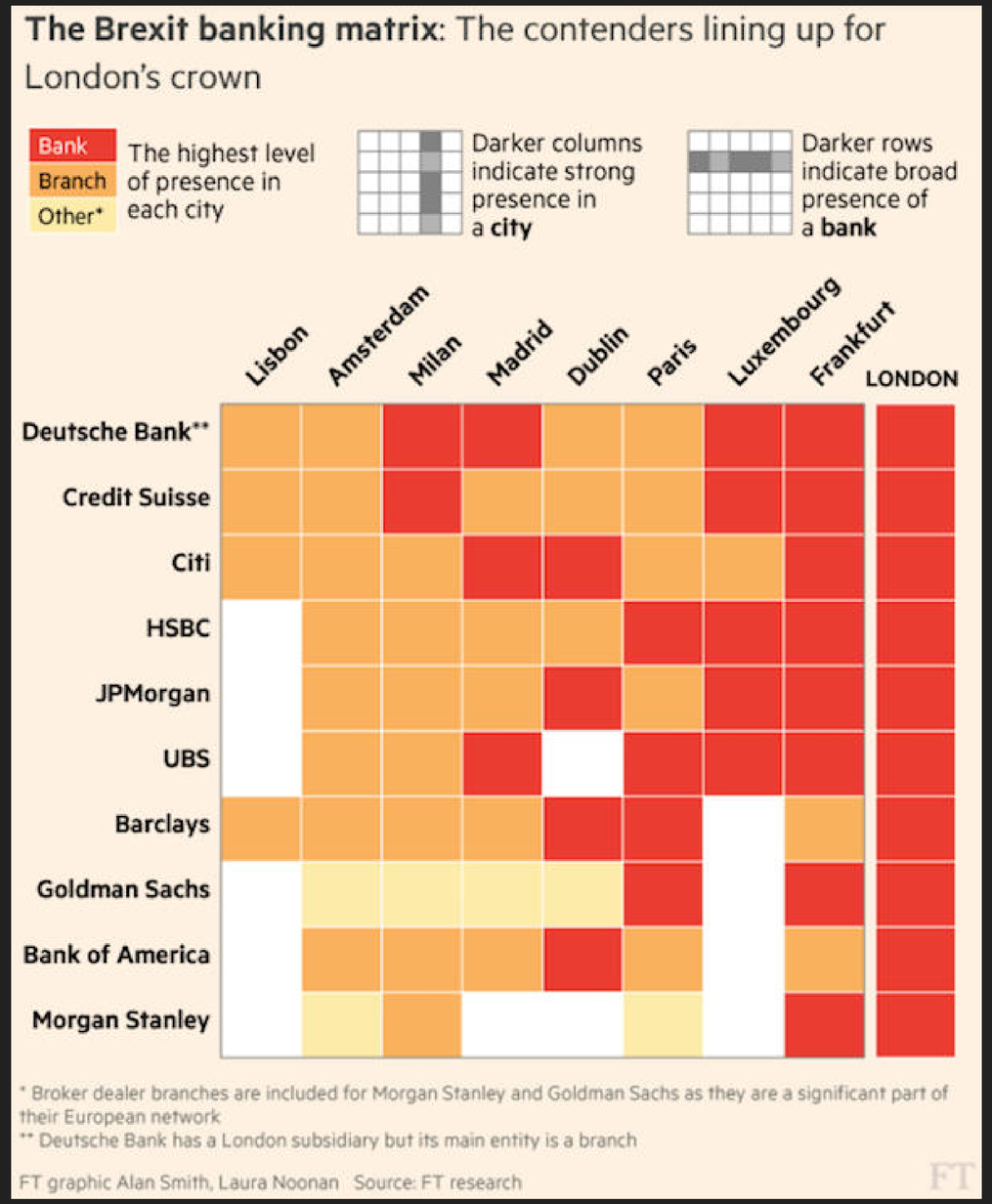


Graphic: Cédric Scherer • Data: Twitter Analytics for @CédScherer


How Britons spend their time at weekends vs weekdays
Share of people doing specific activities during weekends vs weekdays, by time of day (%)
SLEEP
People sleep in more at the weekend...
...and go to bed earlier during the week
TV/VIDEO/AUDIO
People spend more time watching TV and listening to radio/music at the weekend
PAID WORK
The dual humps of mornings and afternoons at work are clear during the week
HOUSEHOLD TASKS
People cram more chores into the early hours during the week
Grey shading represents people who do an activity at this time both on weekdays and weekends
EATING
People grab breakfast earlier during the week, pushing meals later at the weekend
TRAVEL
The weekday rush-hour spikes in travel are clear
PERSONAL CARE
People shower earlier during the week
SOCIALISING
COMPUTER (NON WORK)
EDUCATION
SHOPPING
CHILDCARE/FAMILY
SPORTS & REC
The Sunday League bump is clear at the weekend
READING
ENTERTAINMENT OUTSIDE HOME
RELIGION
Church attendance is visible as a small late-morning bump at the weekend
100%
80%
60%
40%
20%
0%
06:00
12:00
18:00
00:00
Source: FT analysis of UK Time Use Survey 2015
© FT



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

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

2 ...also regularly shop at this supermarket

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

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

YouGov | yougov.com

YouGov Profiles December 2018

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

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...

But well, look at me!

Visualize Hierarchy

I am important!

I am important, too.

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

But well, look at me!

Visualize Hierarchy

I am important!

I am important, too.

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

But well, look at me!

Visualize Hierarchy

I am important!

I am important, too.

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

But well, look at me!

Visualize Hierarchy

I am important!

I am important, too.

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

But well, 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

123.45
678.90

Tabular Numbers

123.45
678.90

The 1ll Test

1ll Calibri

1ll Open Sans

1ll Roboto

1ll Lato

1ll Oswald

1ll Cabinet Grotesk

1ll Cabin

1Il Monda

1Il Chivo

1ll Fira Sans

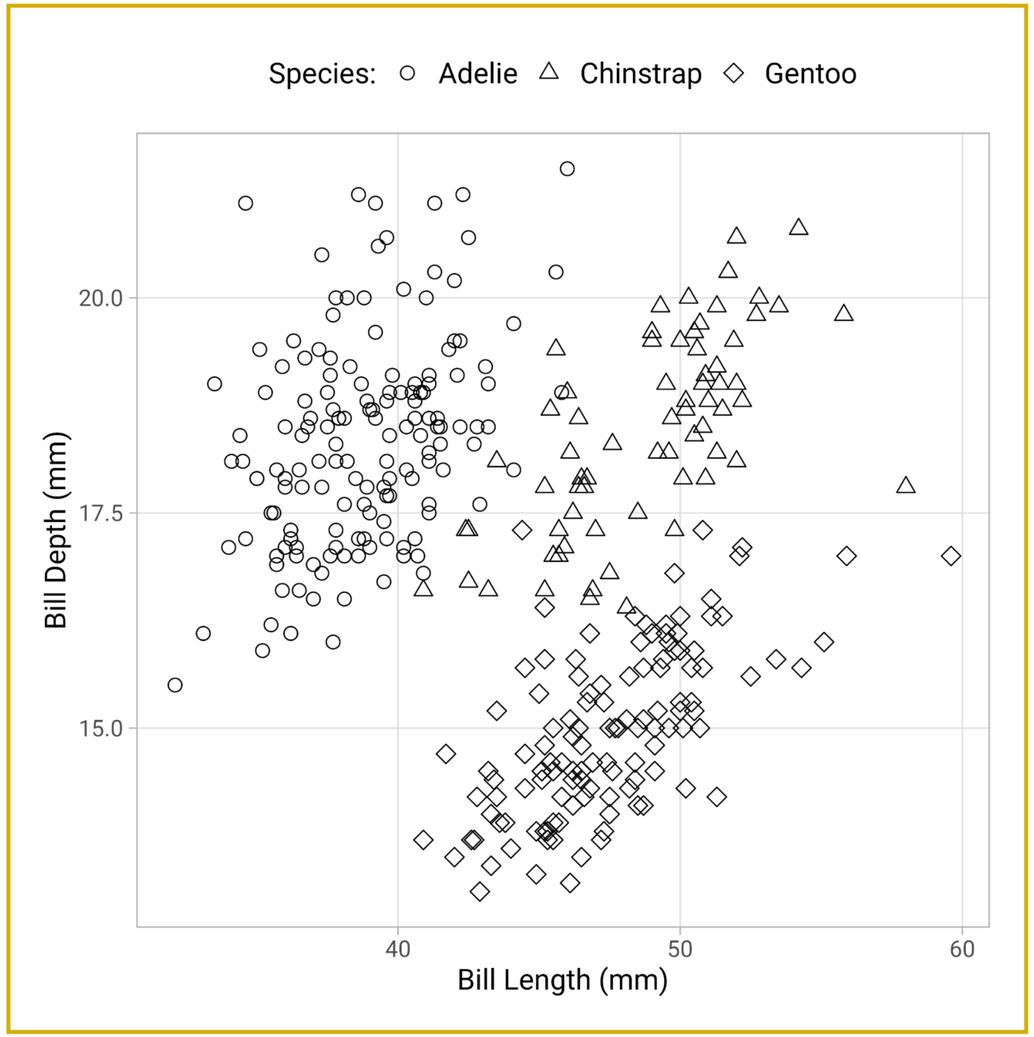
1Il Noto Sans

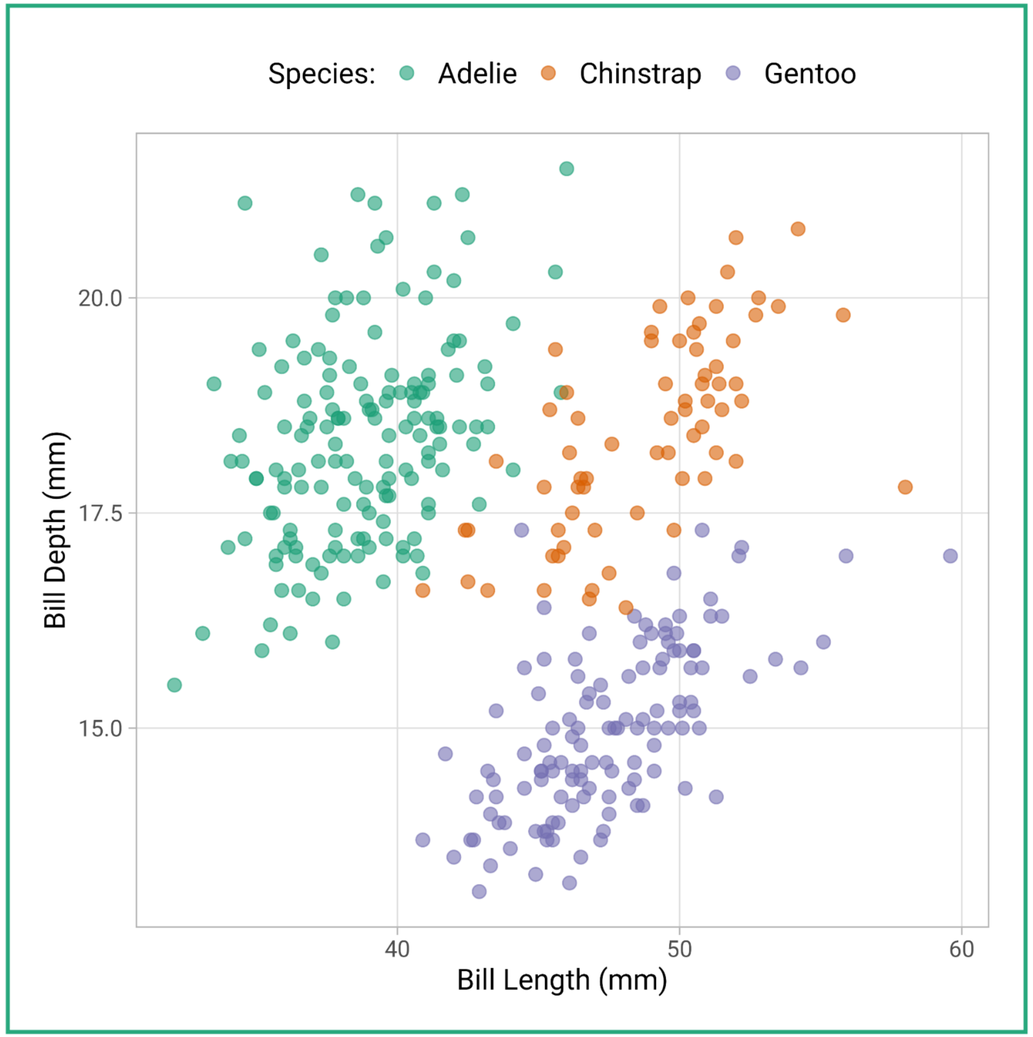
1ll Amulya

COLORS

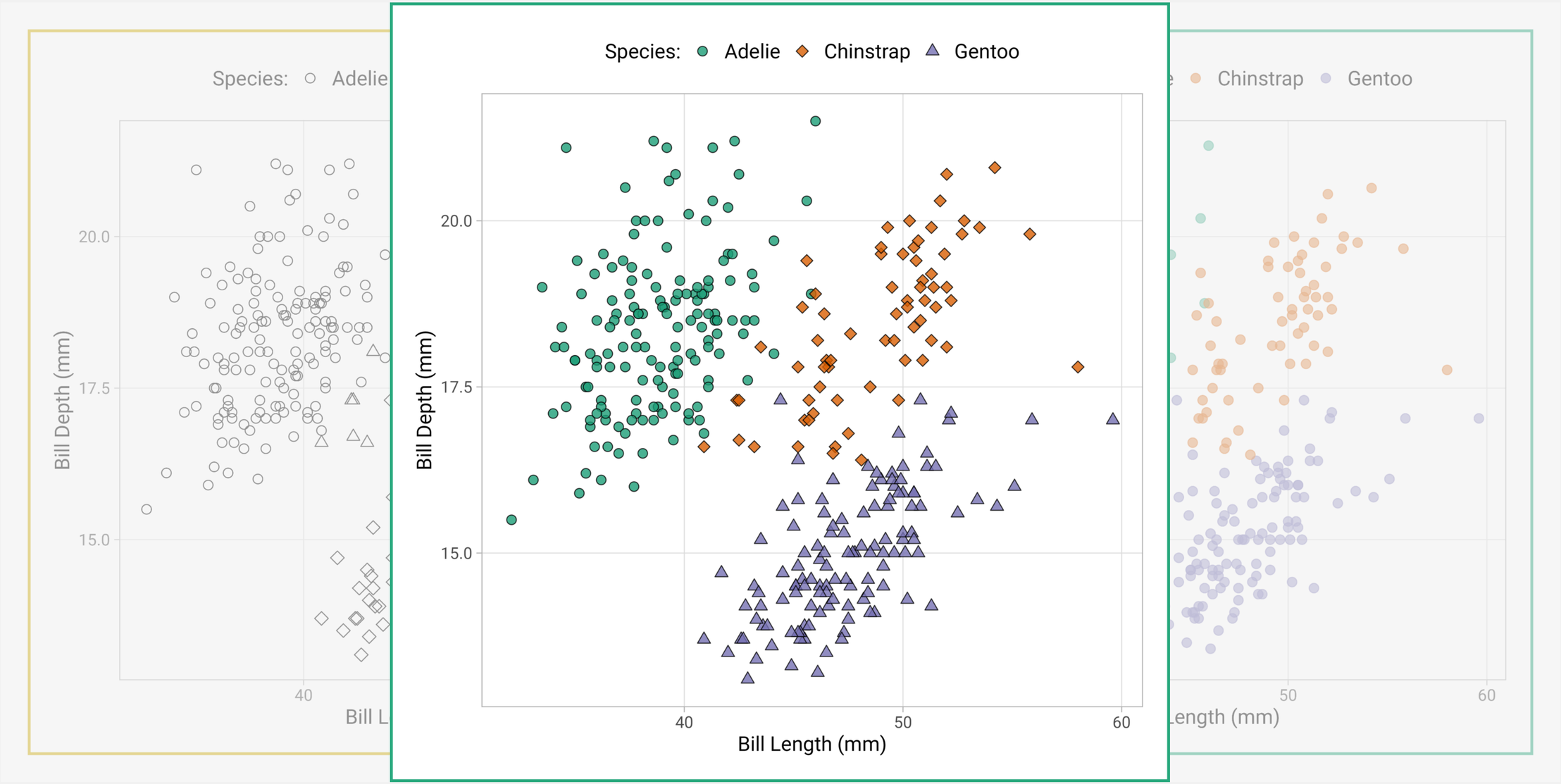
and Pitfalls

Channels to Encode Information

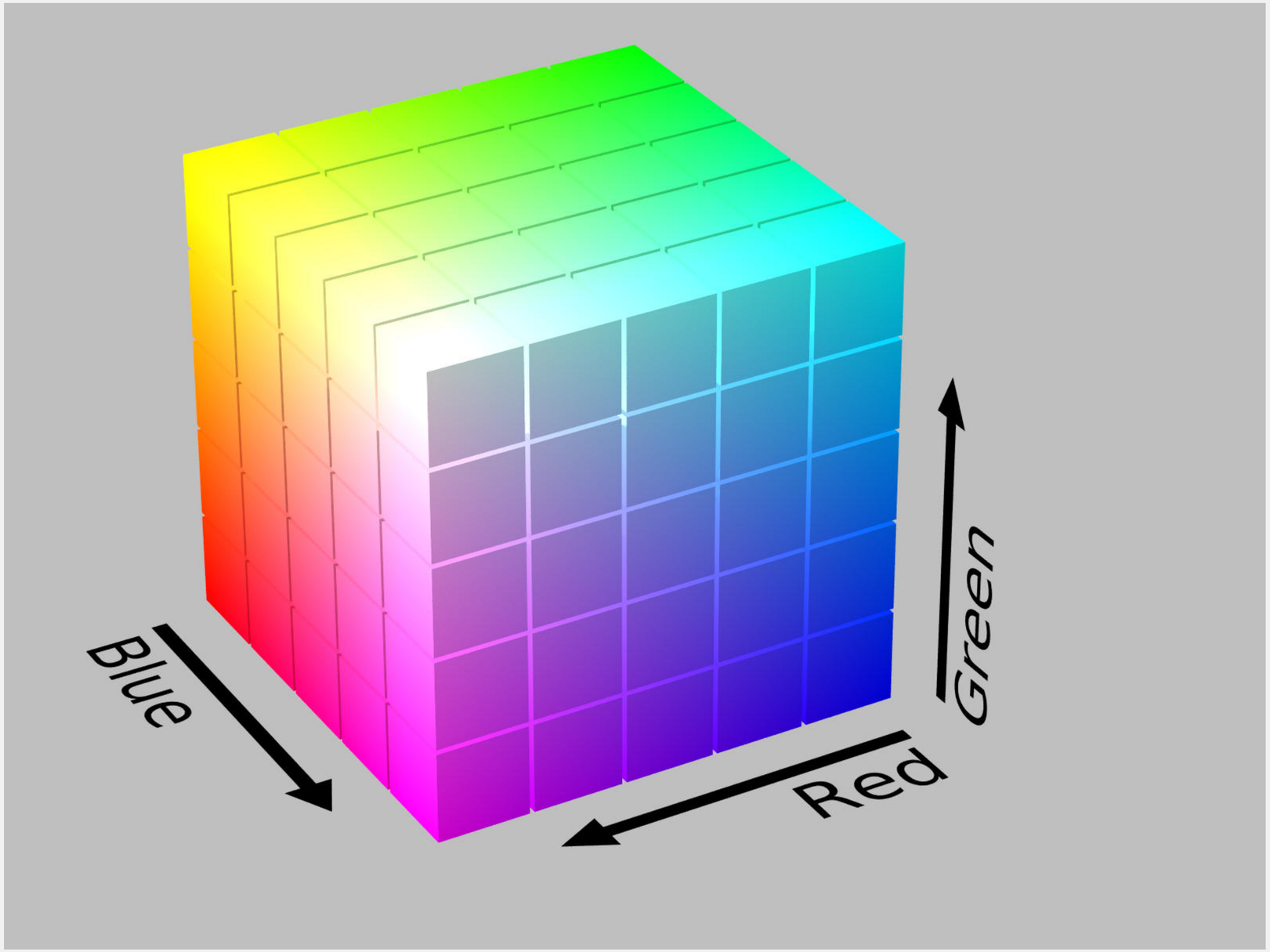




Use of Double Encoding

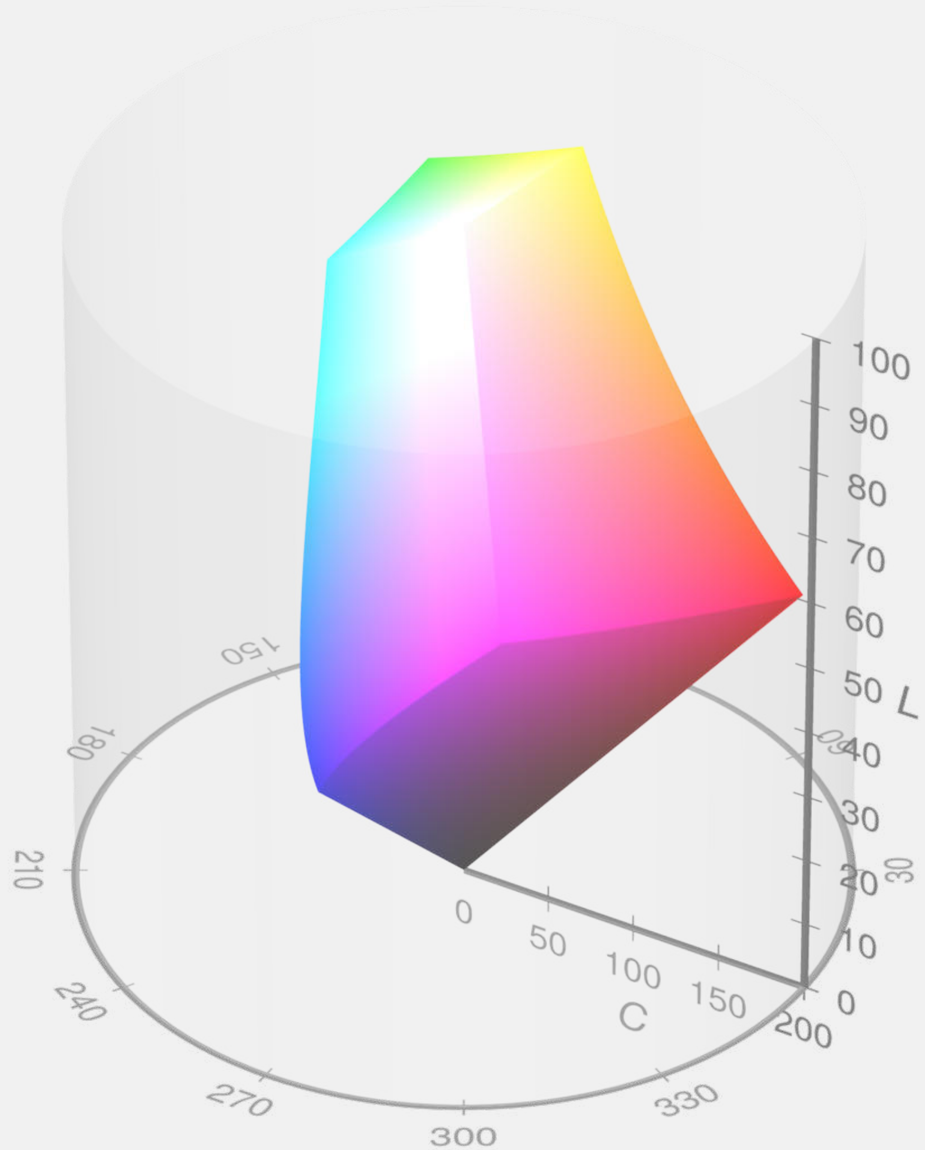


The RGB Color Space



Michael Horvath

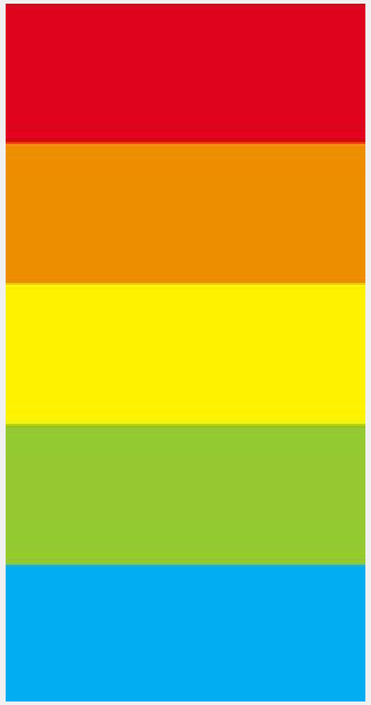
The HCL Color Space



Michael Horvath & Christoph Lipka

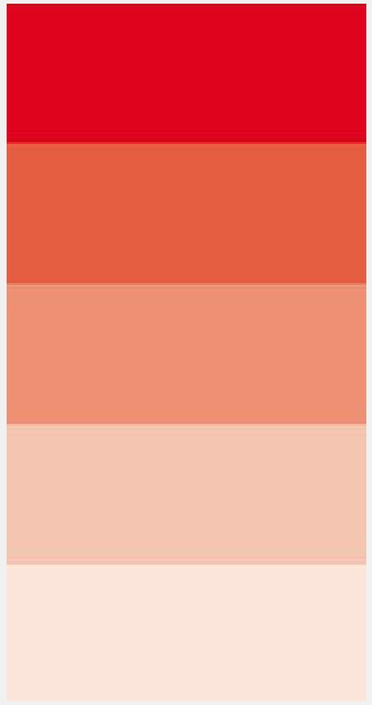
cedricscherer.com @CedScherer z3tt

Hue
color family

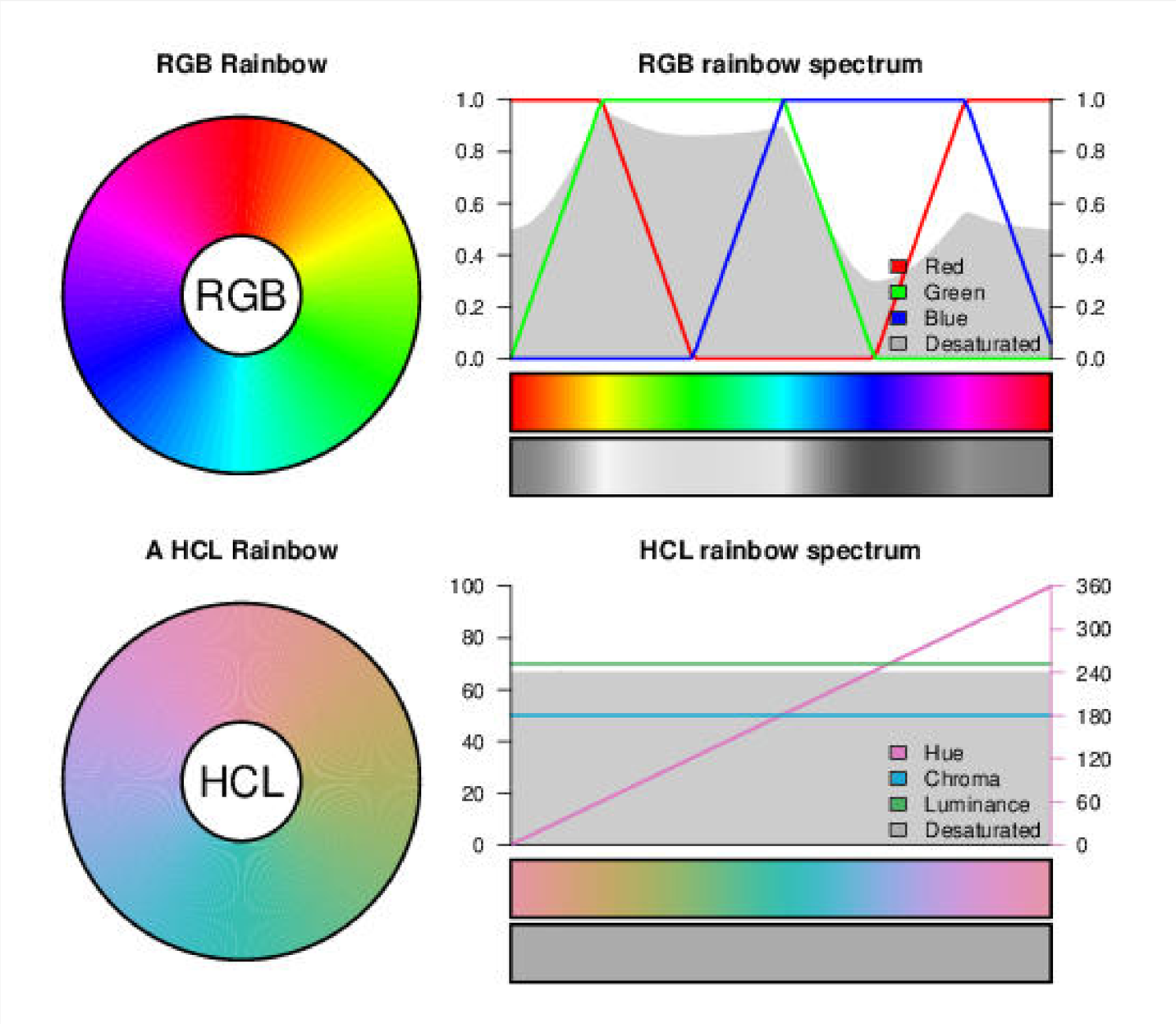


Chroma
colorfulness

Luminance
lightness / brightness



Modified from uxplanet.org



hclwizard.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

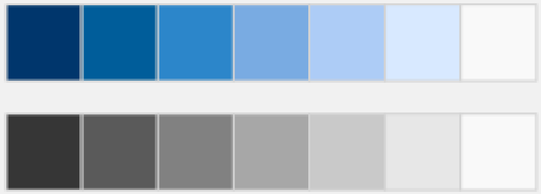
Modified from the {colorspace} R package vignette

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

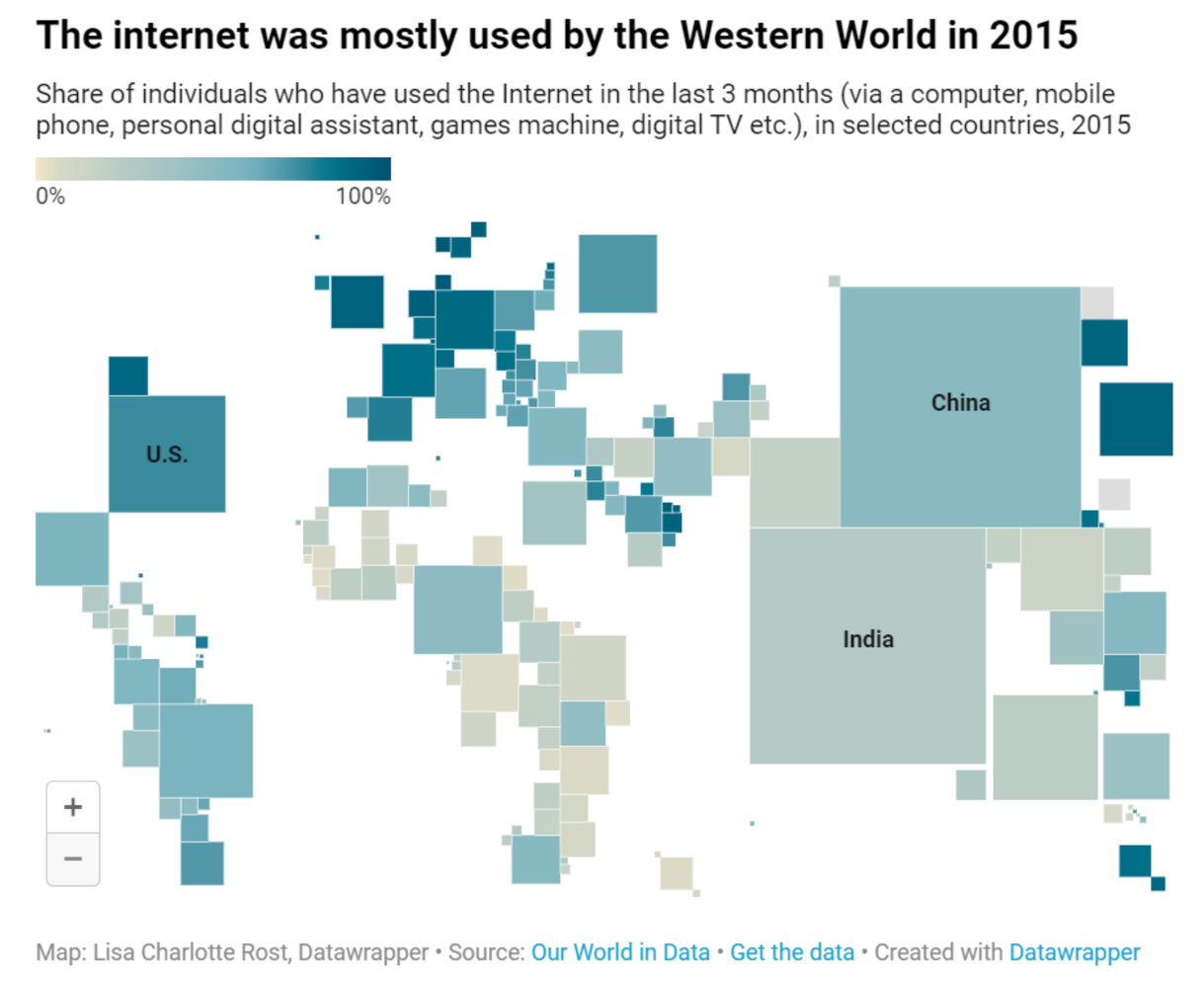
Use to encode
numerical information with critical midpoint

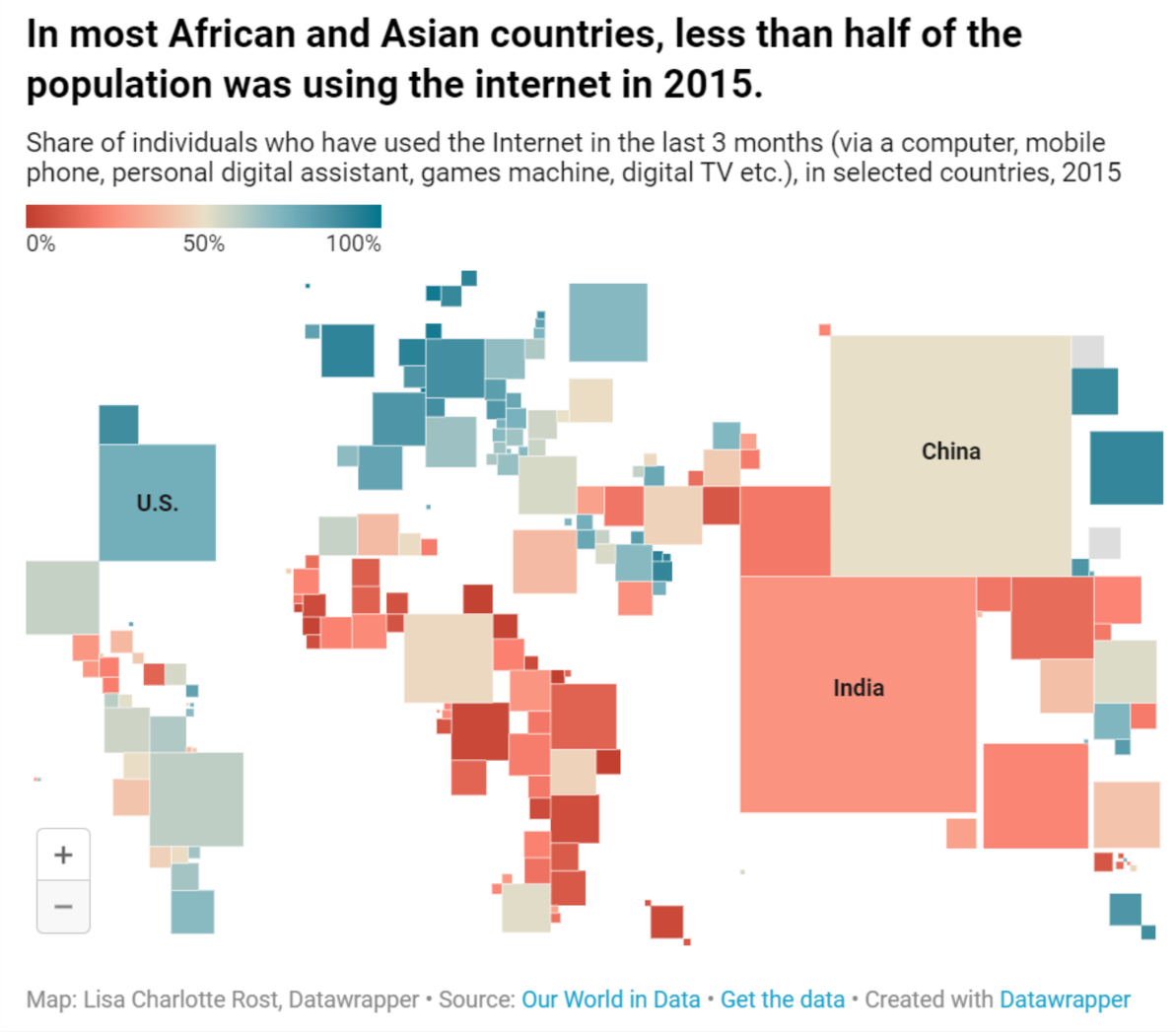
ensure a meaningful midpoint value and use balanced extremes

combination of two sequential palettes

Modified from the {colorspace} R package vignette

Sequential versus Diverging Palettes





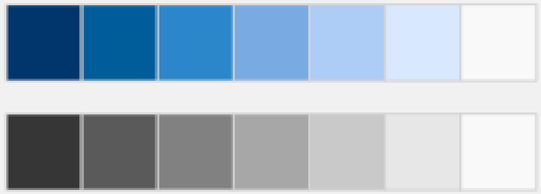
Source: "When to use sequential and when to use diverging color scales" by Lisa C. Muth / DataWrapper

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

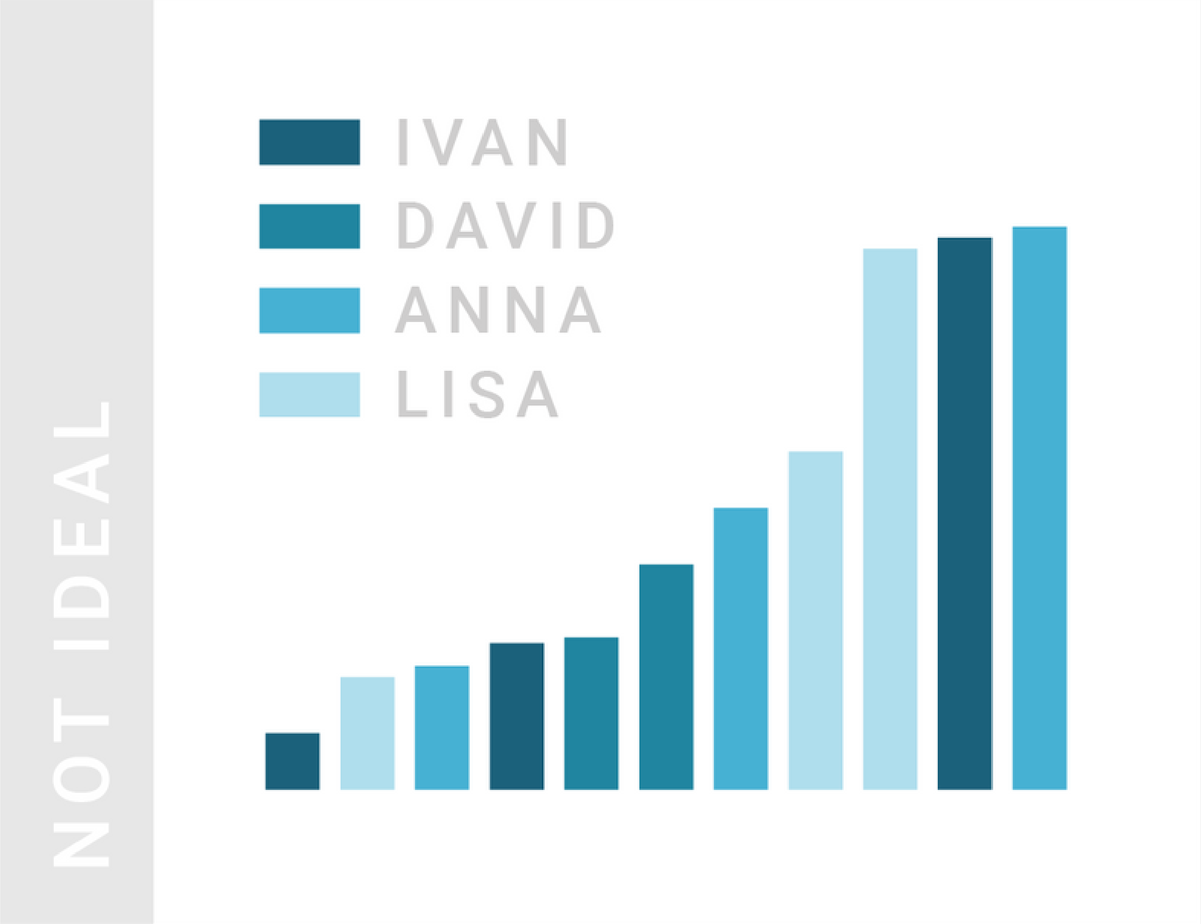
Use to encode
categorical information

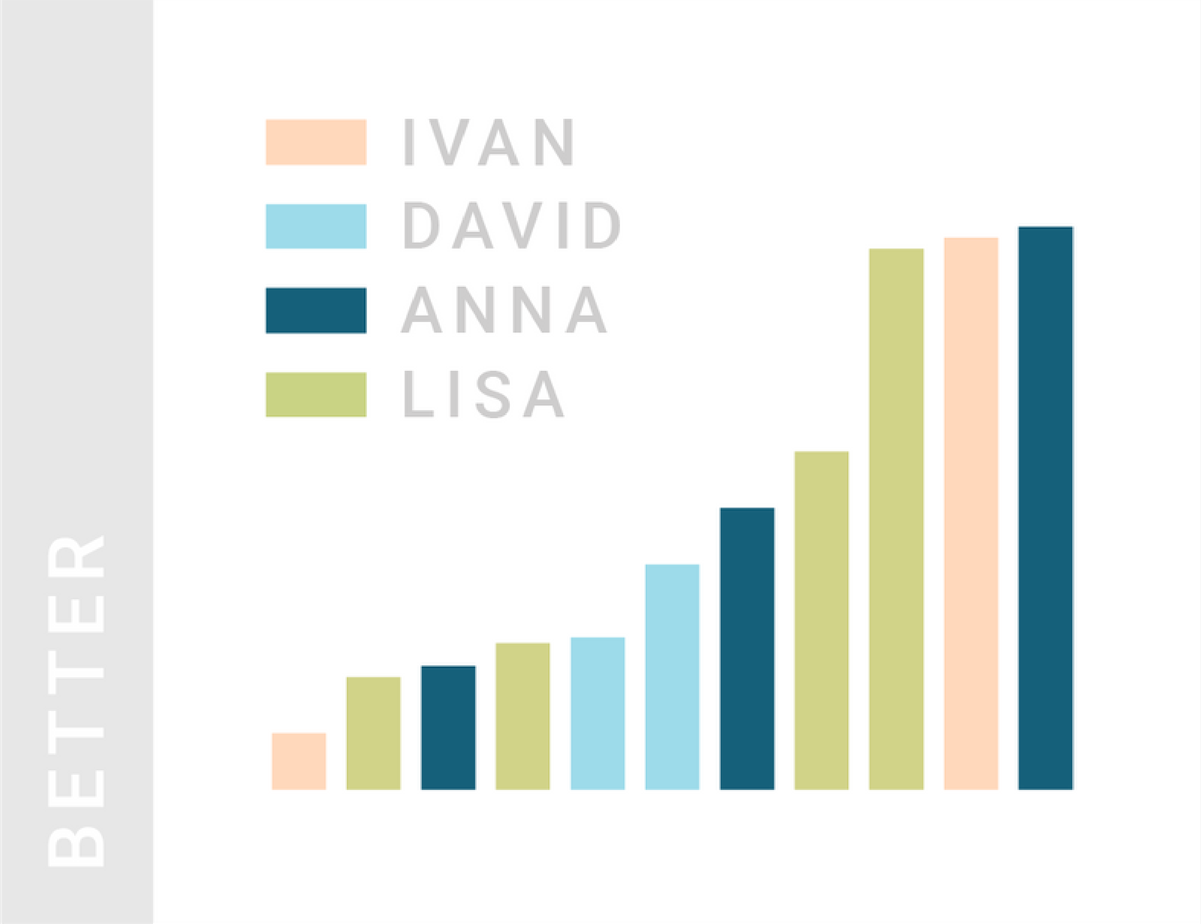
pick distinct colors with the same perceptual weight

limit categories to 6–8

Modified from the {colorspace} R package vignette

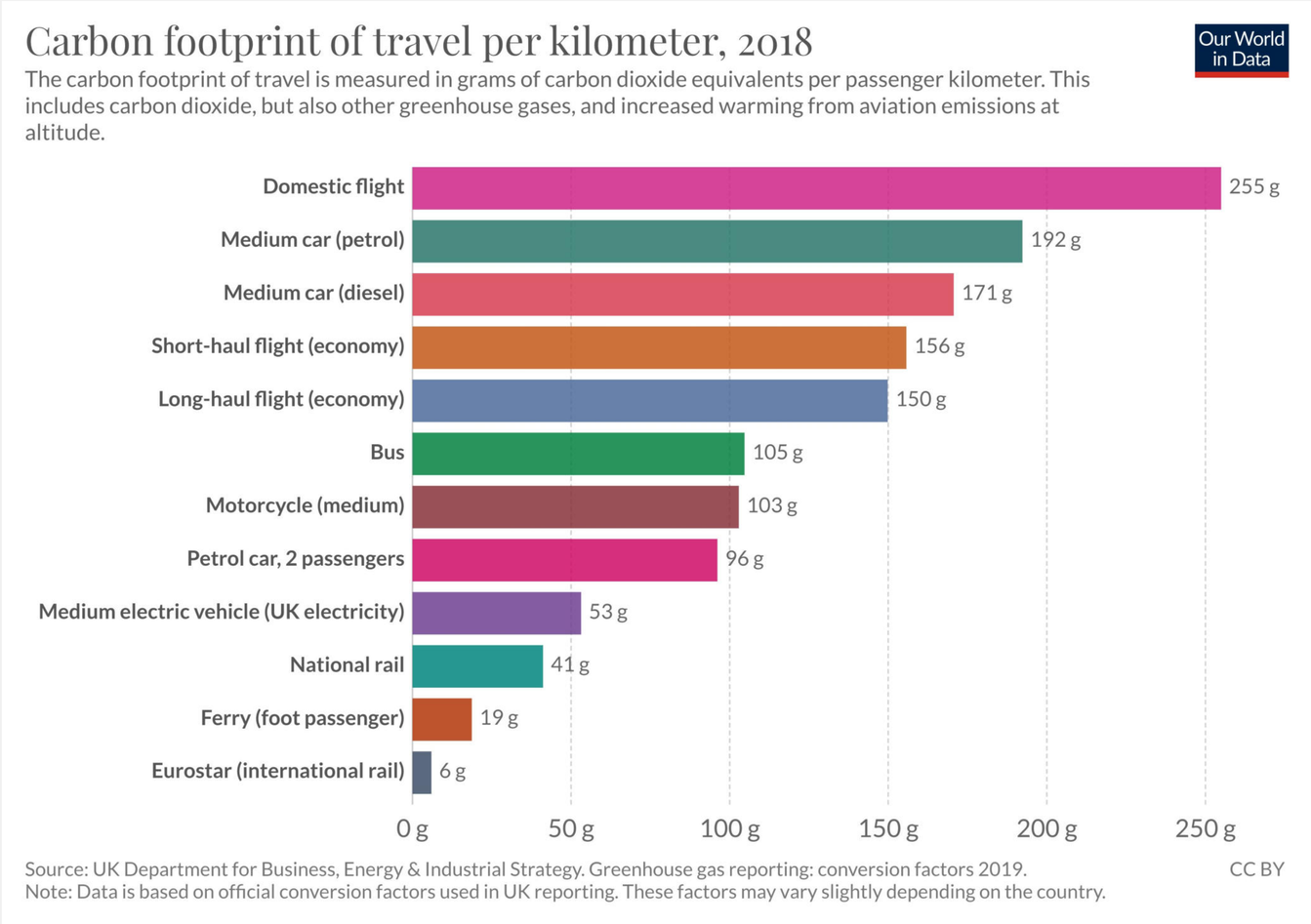
Sequential versus Qualitative Palettes





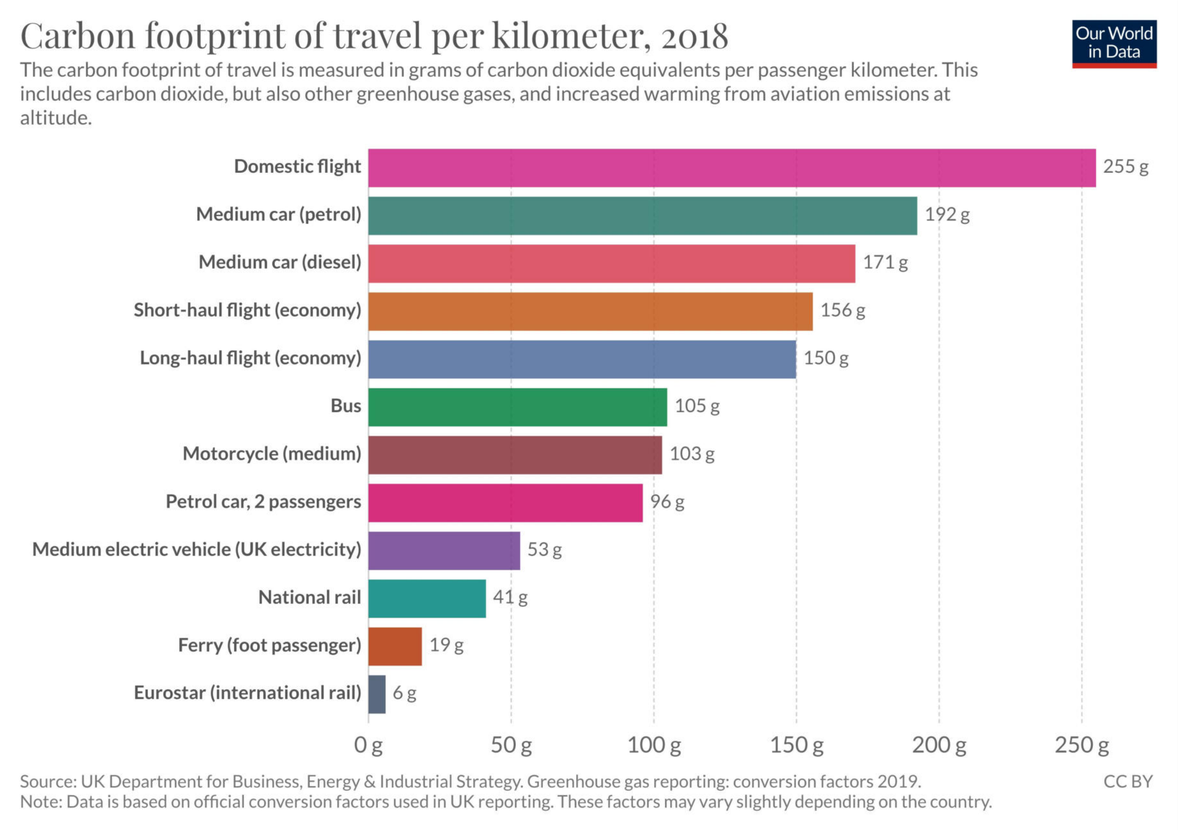
Source: "When to use sequential and when to use diverging color scales" by Lisa C. Muth / DataWrapper

Use color wisely (and correctly)

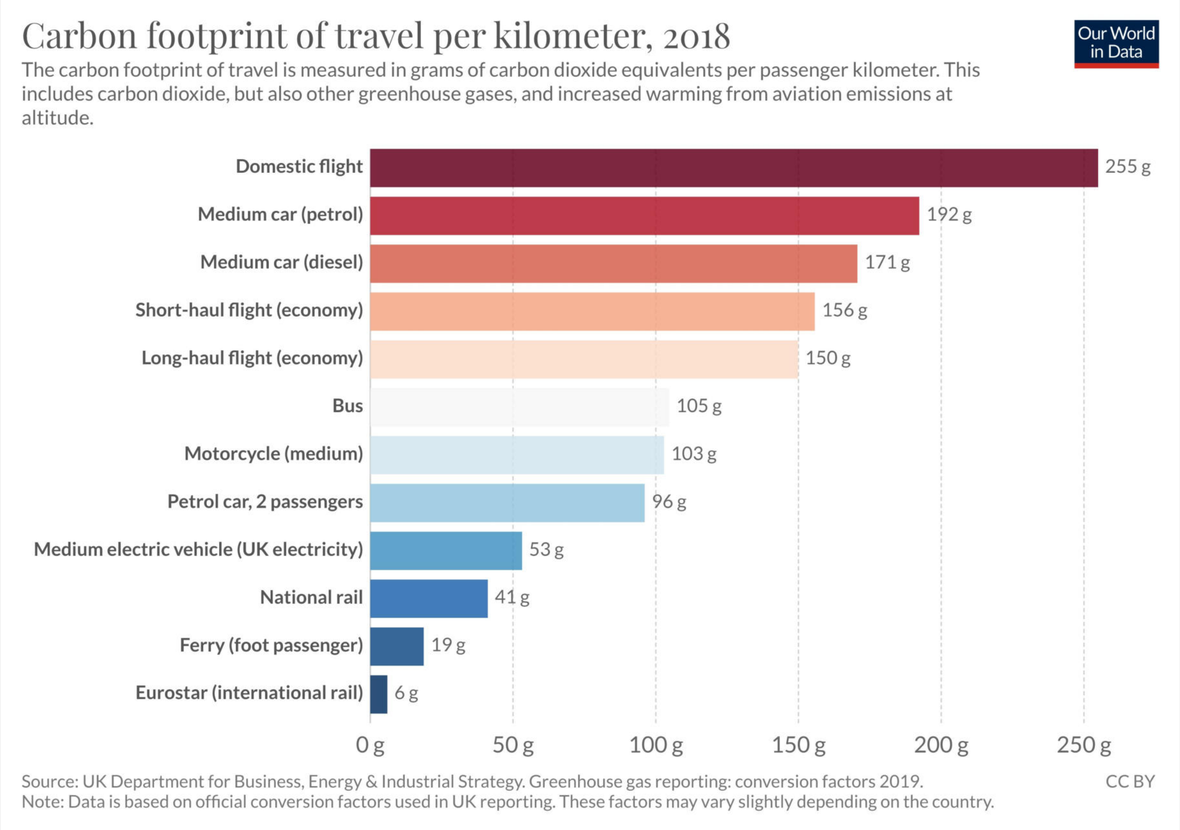


Original graphic with a random categorical palette

Use color wisely (and correctly)

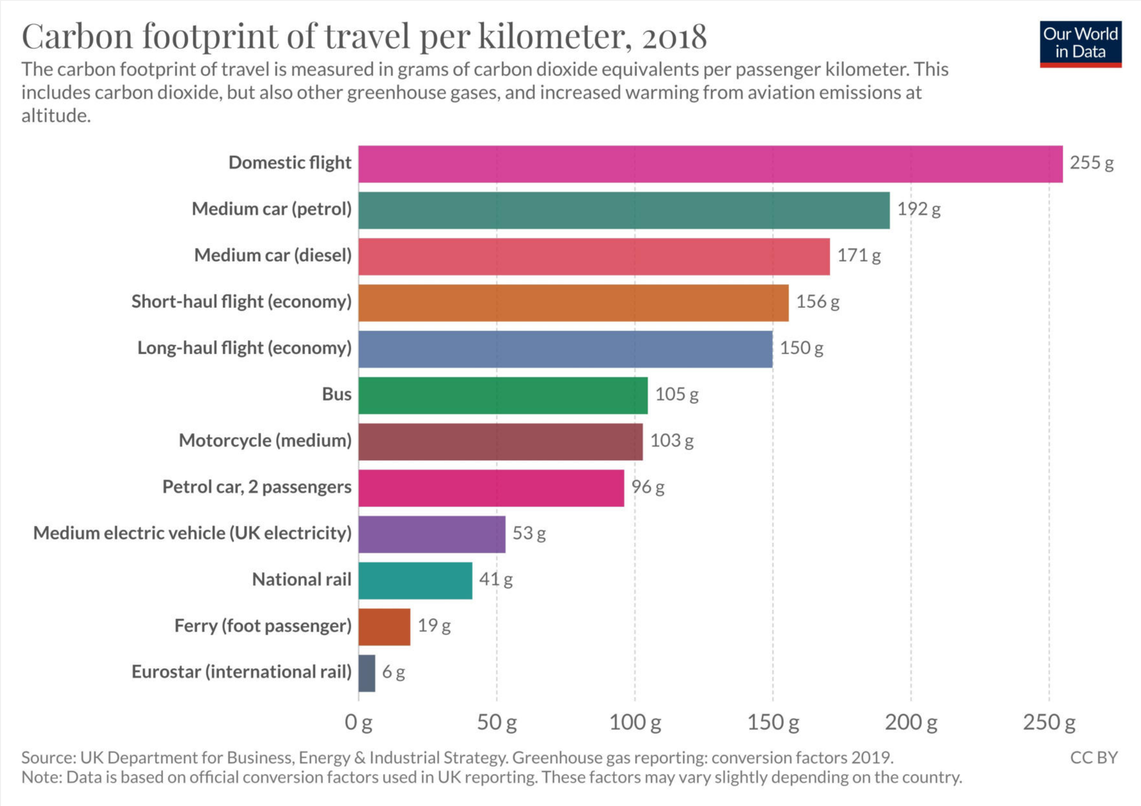


Original graphic with a random categorical palette

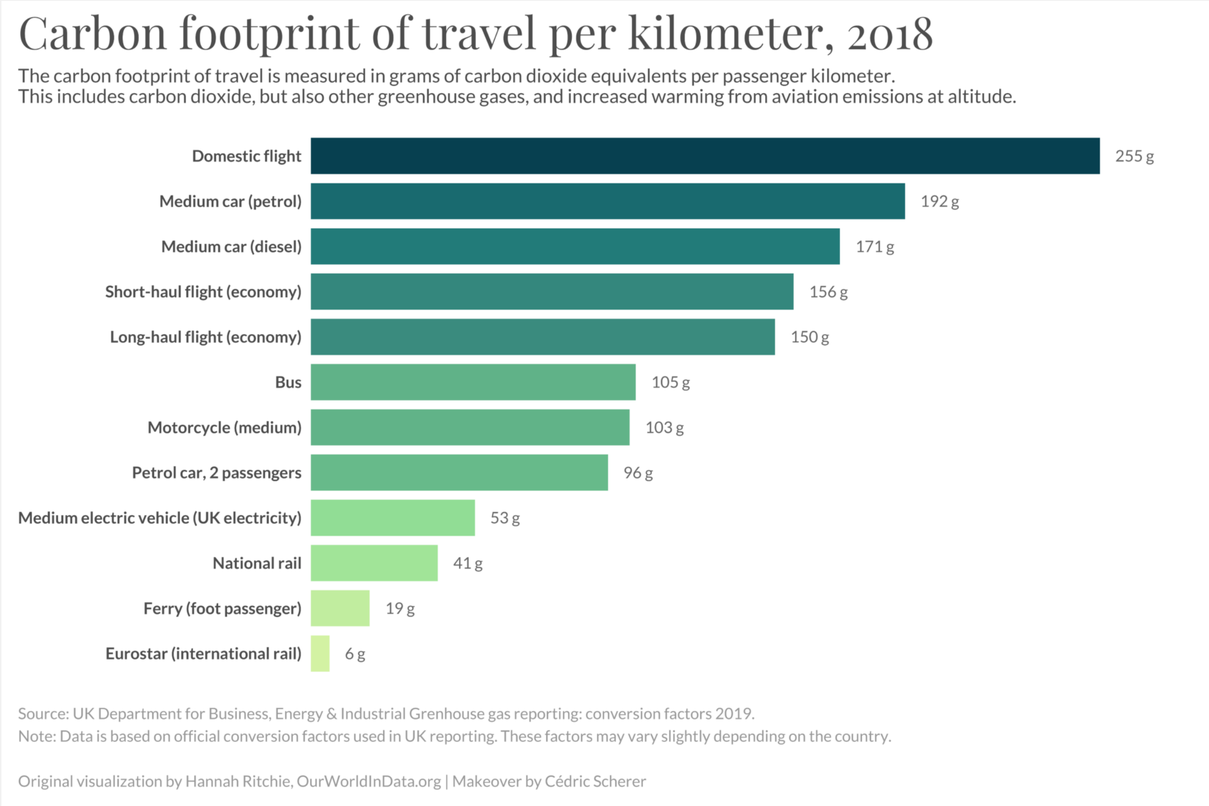


Reworked graphic using a diverging palette

Use color wisely (and correctly)



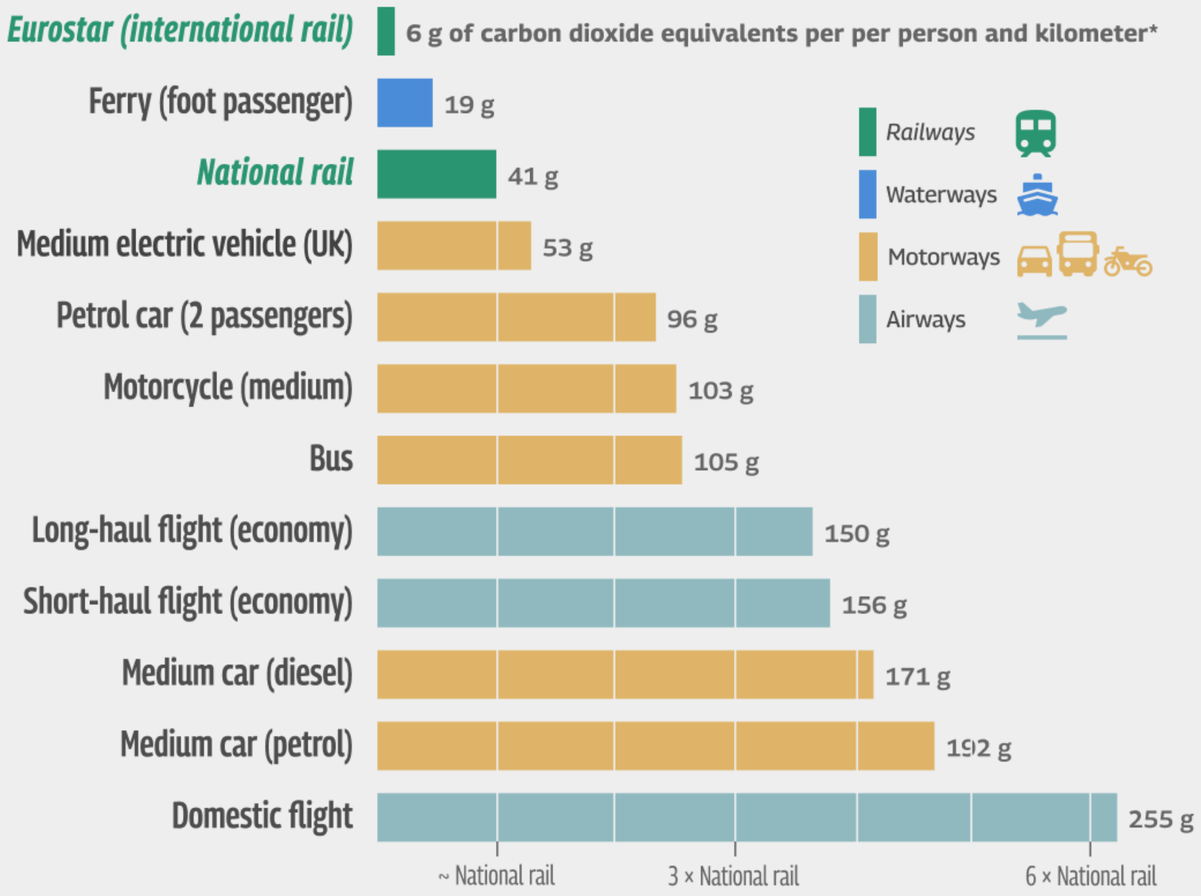
Original graphic with a random categorical palette



Makeover using a continuous palette

Reduce your Carbon Footprint: *Take the Train*

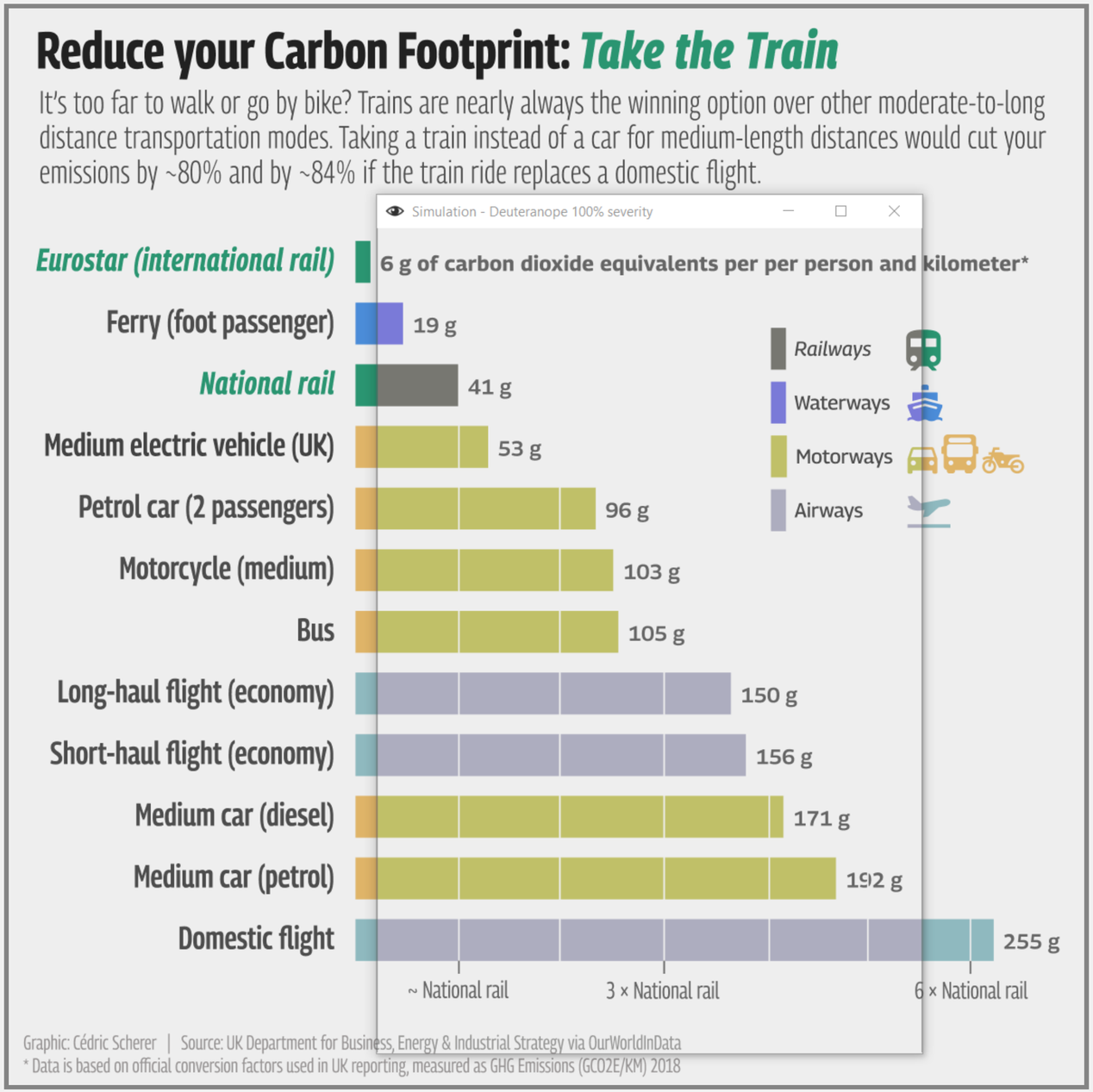
It's too far to walk or go by bike? Trains are nearly always the winning option over other moderate-to-long distance transportation modes. Taking a train instead of a car for medium-length distances would cut your emissions by ~80% and by ~84% if the train ride replaces a domestic flight.



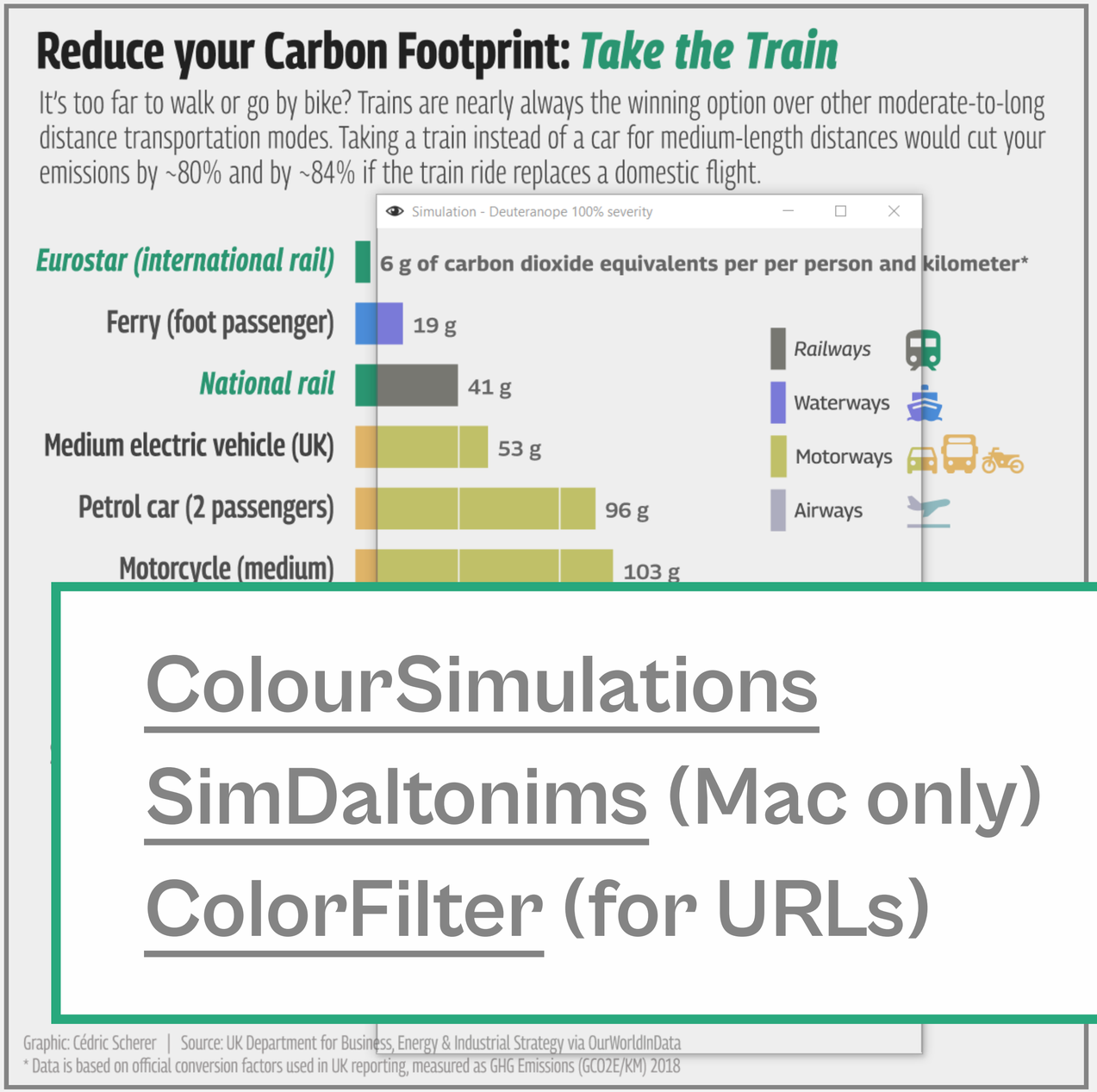
Graphic: Cédric Scherer | Source: UK Department for Business, Energy & Industrial Strategy via OurWorldInData
* Data is based on official conversion factors used in UK reporting, measured as GHG Emissions (GCO2E/KM) 2018

Reworked gaphic as contribution to the #30DayChartChallenge 2022

Don’t trust color palettes — test them!



Reworked gaphic as contribution to the #30DayChartChallenge 2022



ColourSimulations

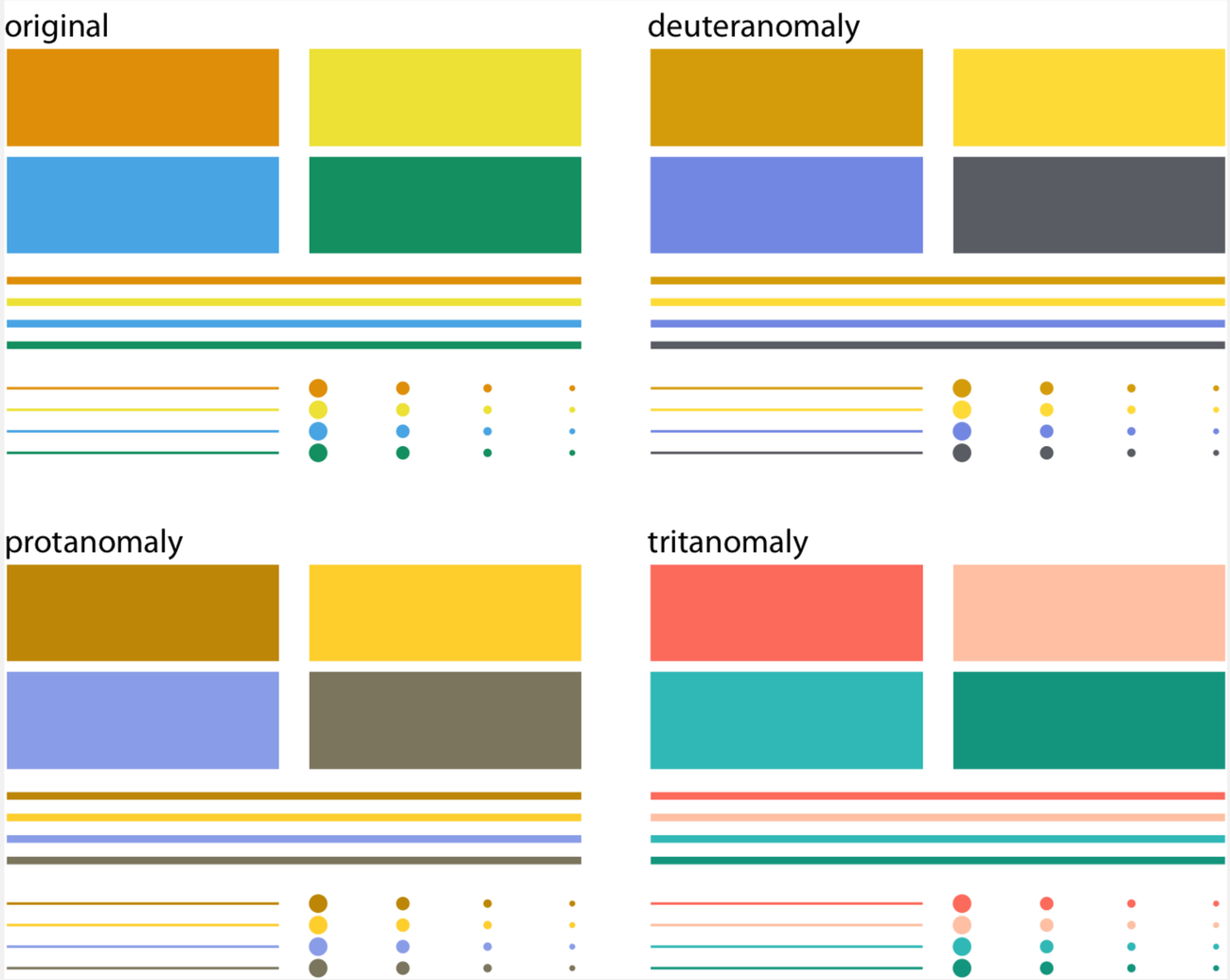
SimDaltonims (Mac only)

ColorFilter (for URLs)

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

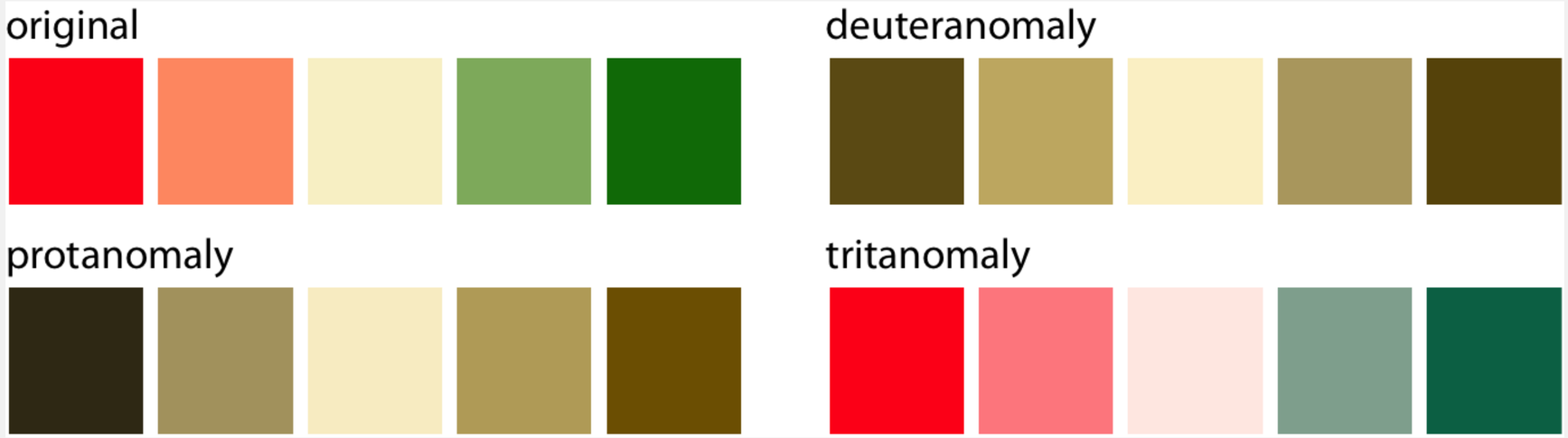
Reworked gaphic as contribution to the #30DayChartChallenge 2022

Ensure Readibility for Color-Blind Persons



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

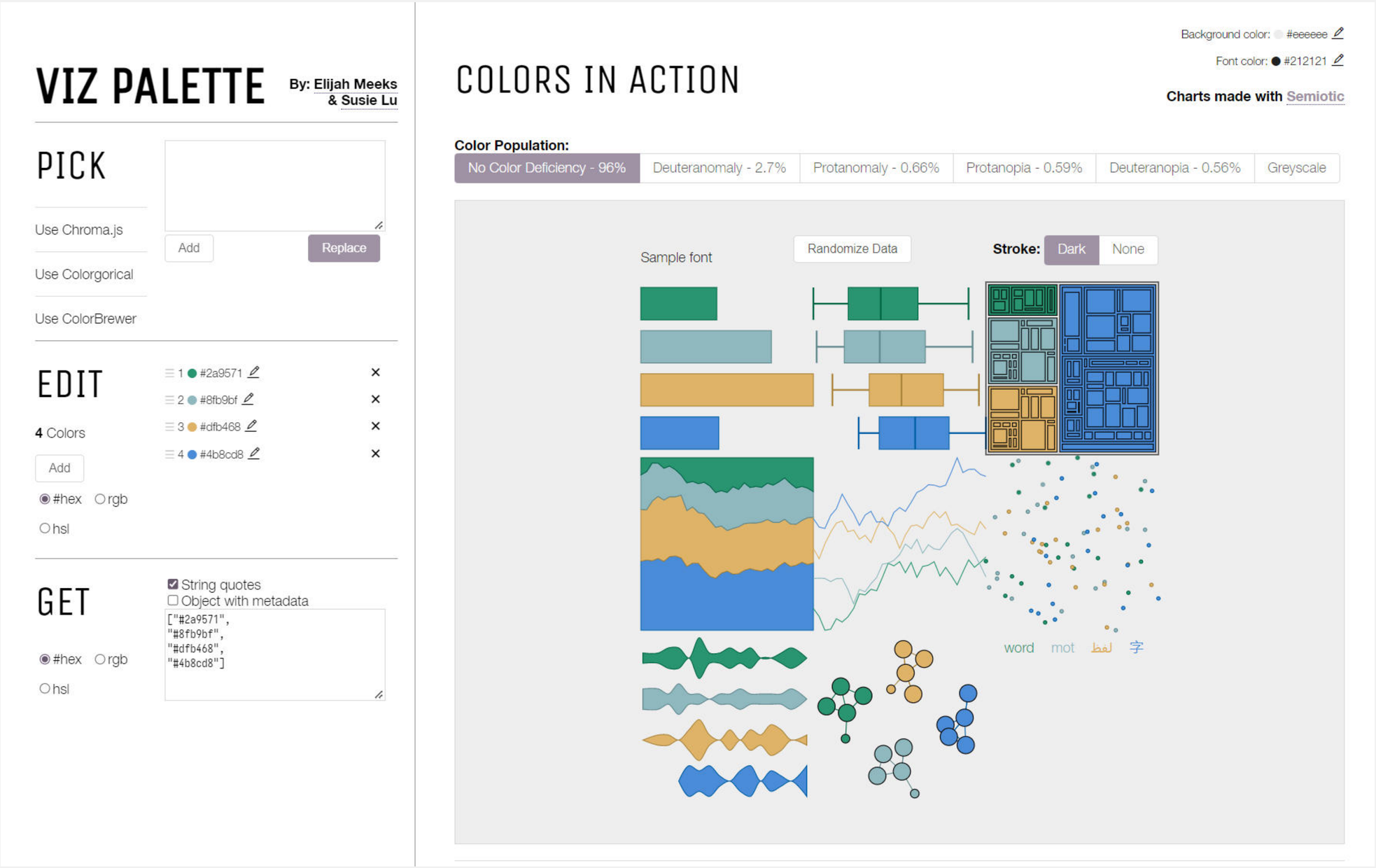
Ensure Readibility for Color-Blind Persons



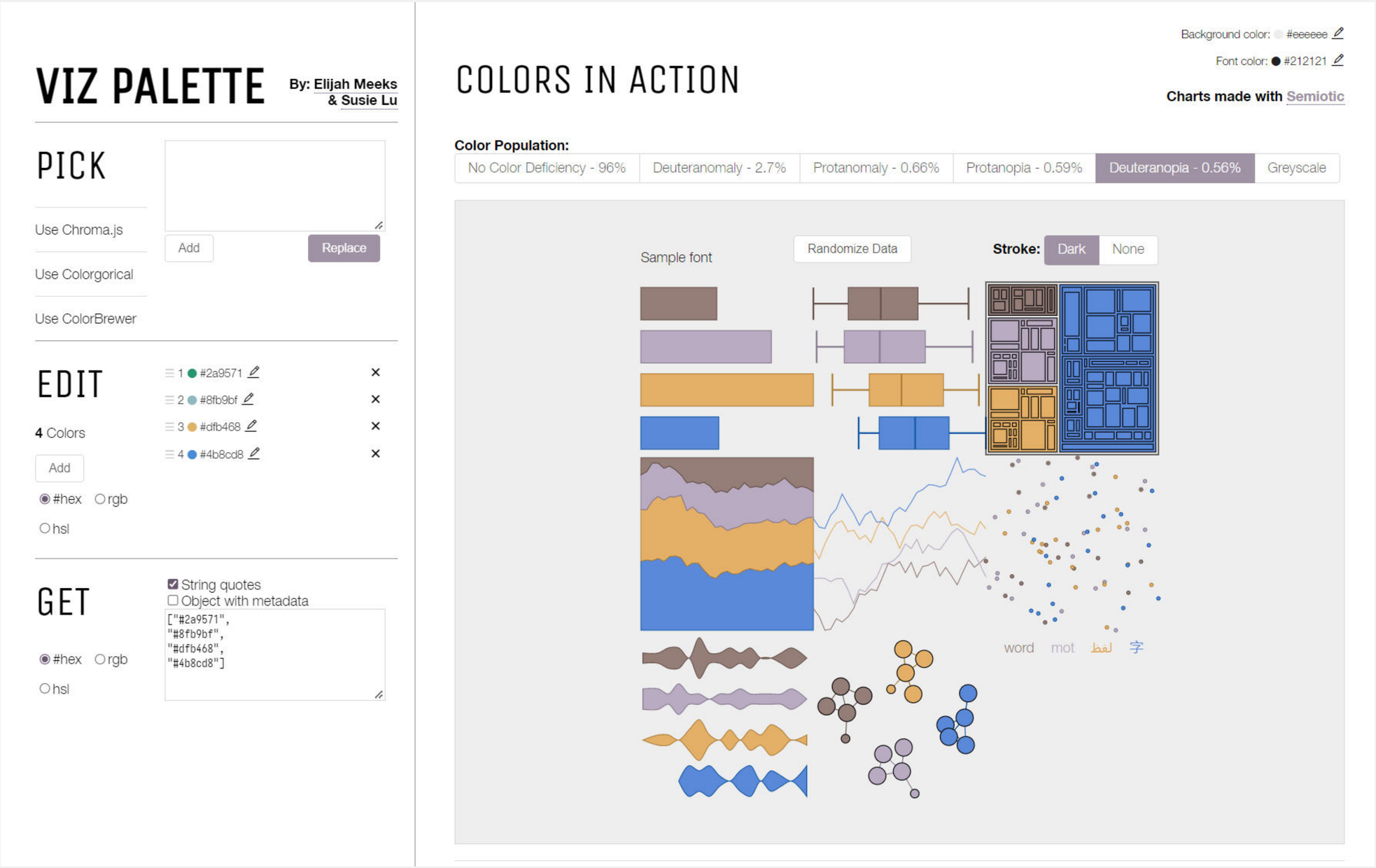
Source: “Fundamentals of Data Visualization” by Claus O. Wilke

Ensure Readibility for Color-Blind Persons

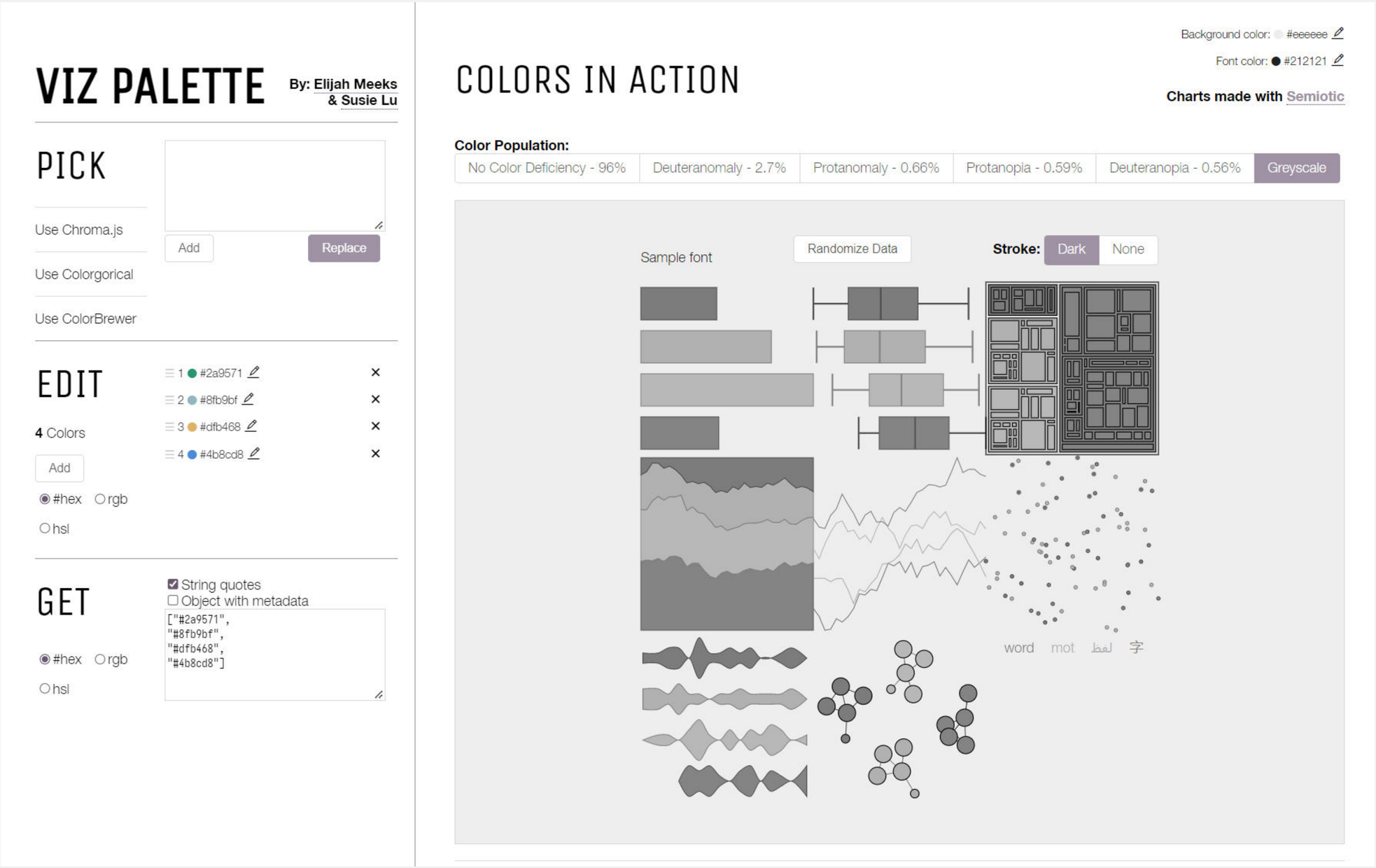
Source: "Fundamentals of Data Visualization" by Claus O. Wilke



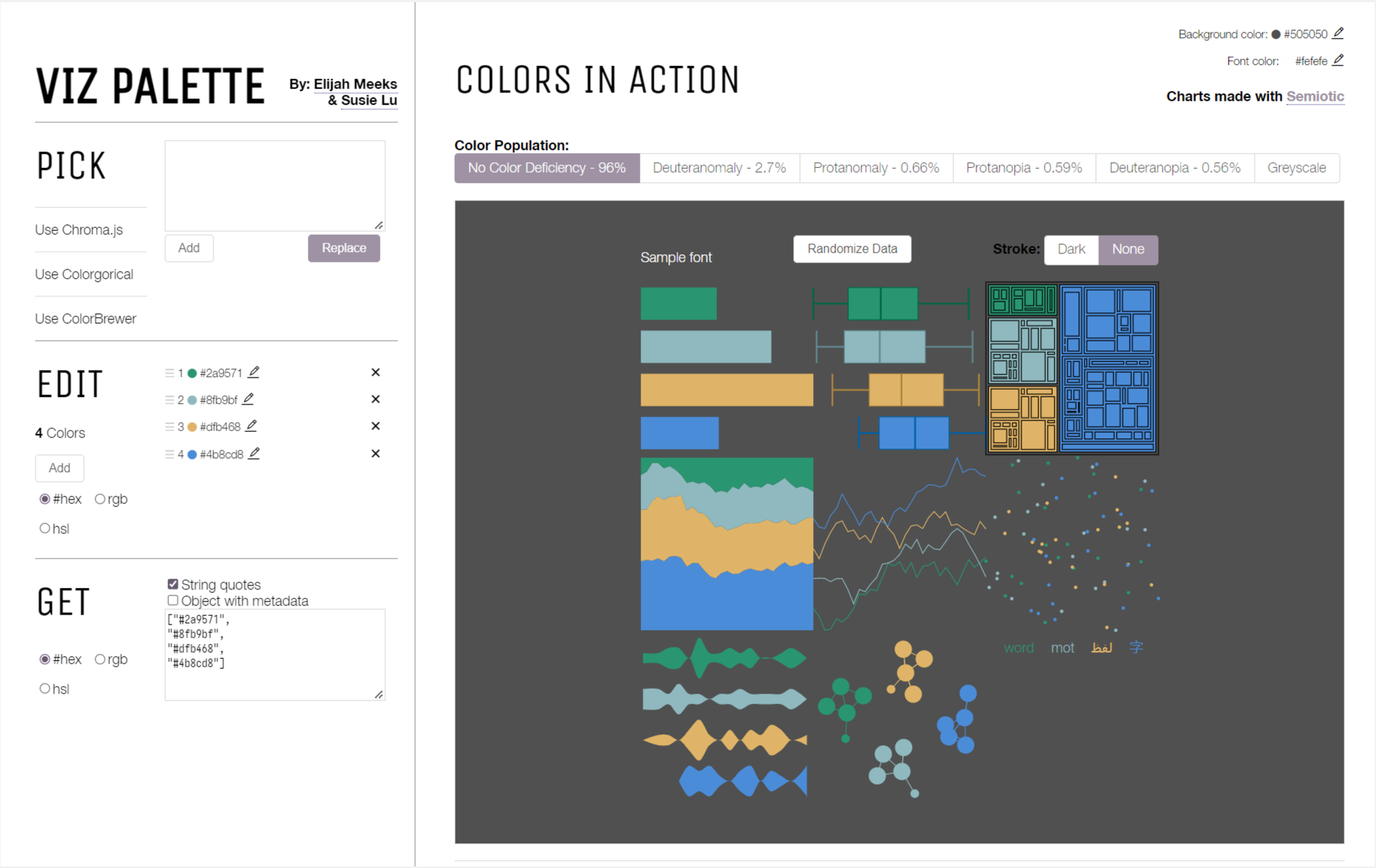
Viz Palette displaying the "colors in action" without color deficiency



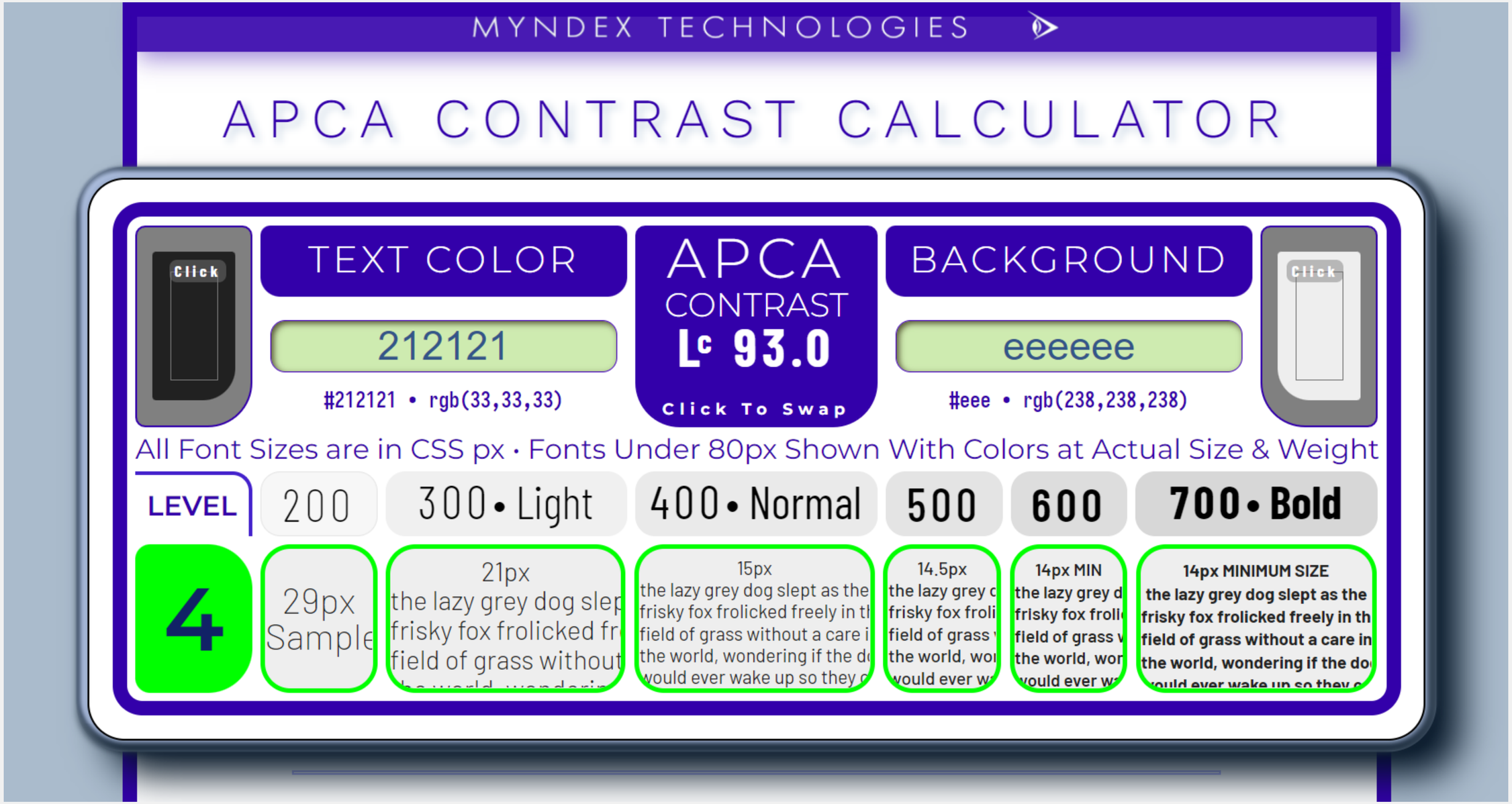
Viz Palette displaying the "colors in action" without color deficiency



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

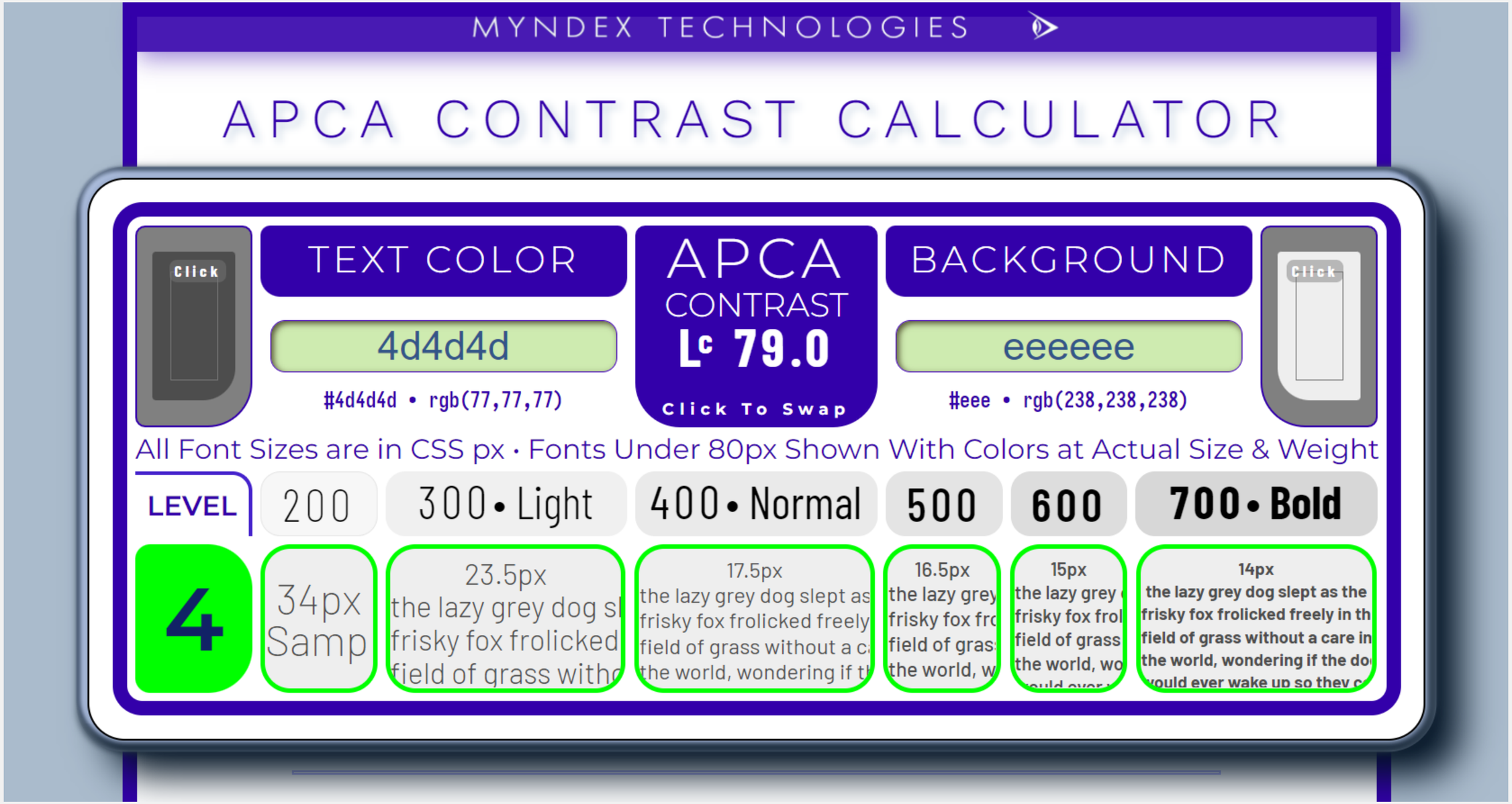


Viz Palette displaying the "colors in action" without color deficiency

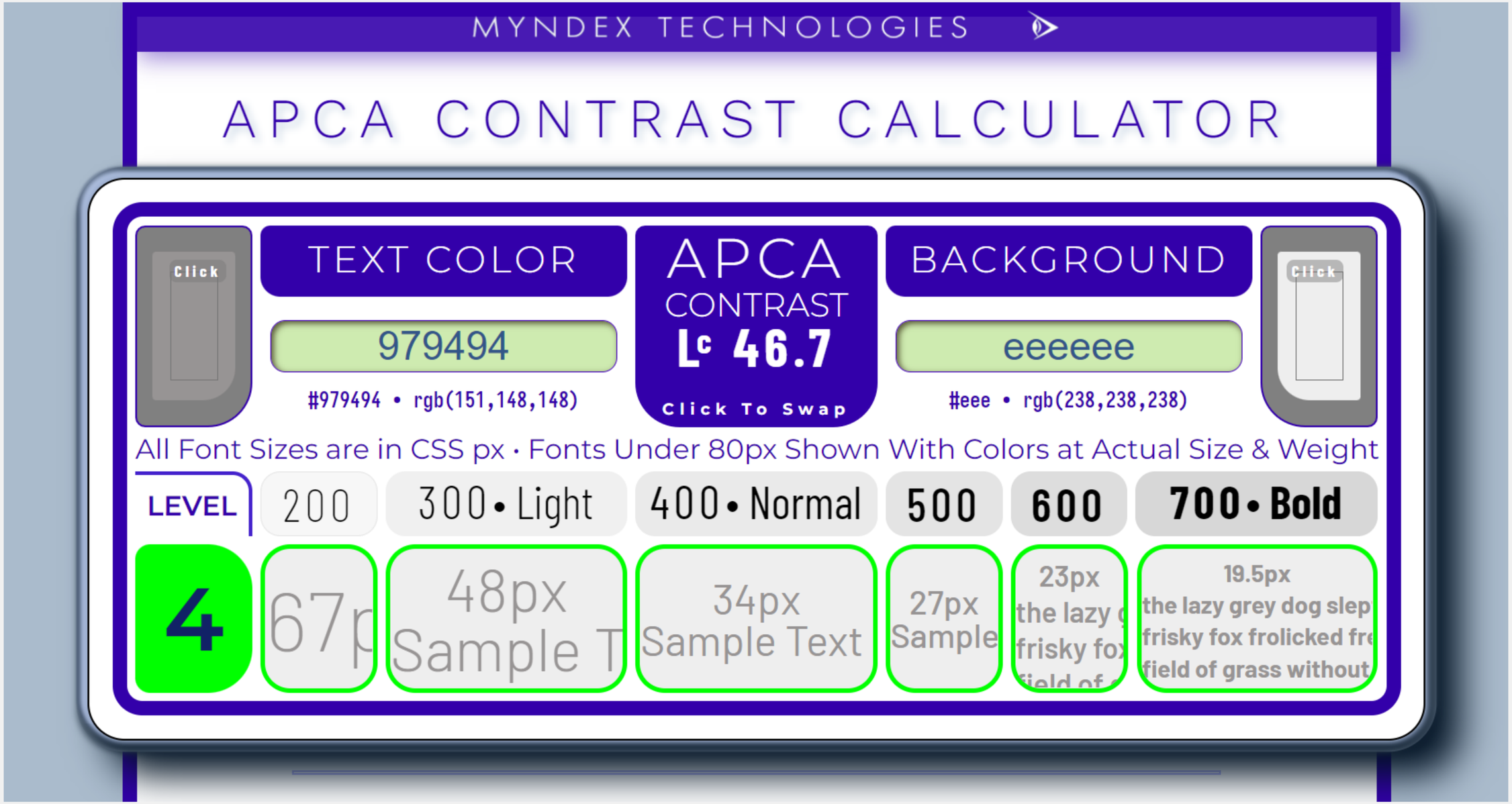


The Myndex APCA Contrast Calculator displays modern contrast ratios for various combinations of text size and font weight

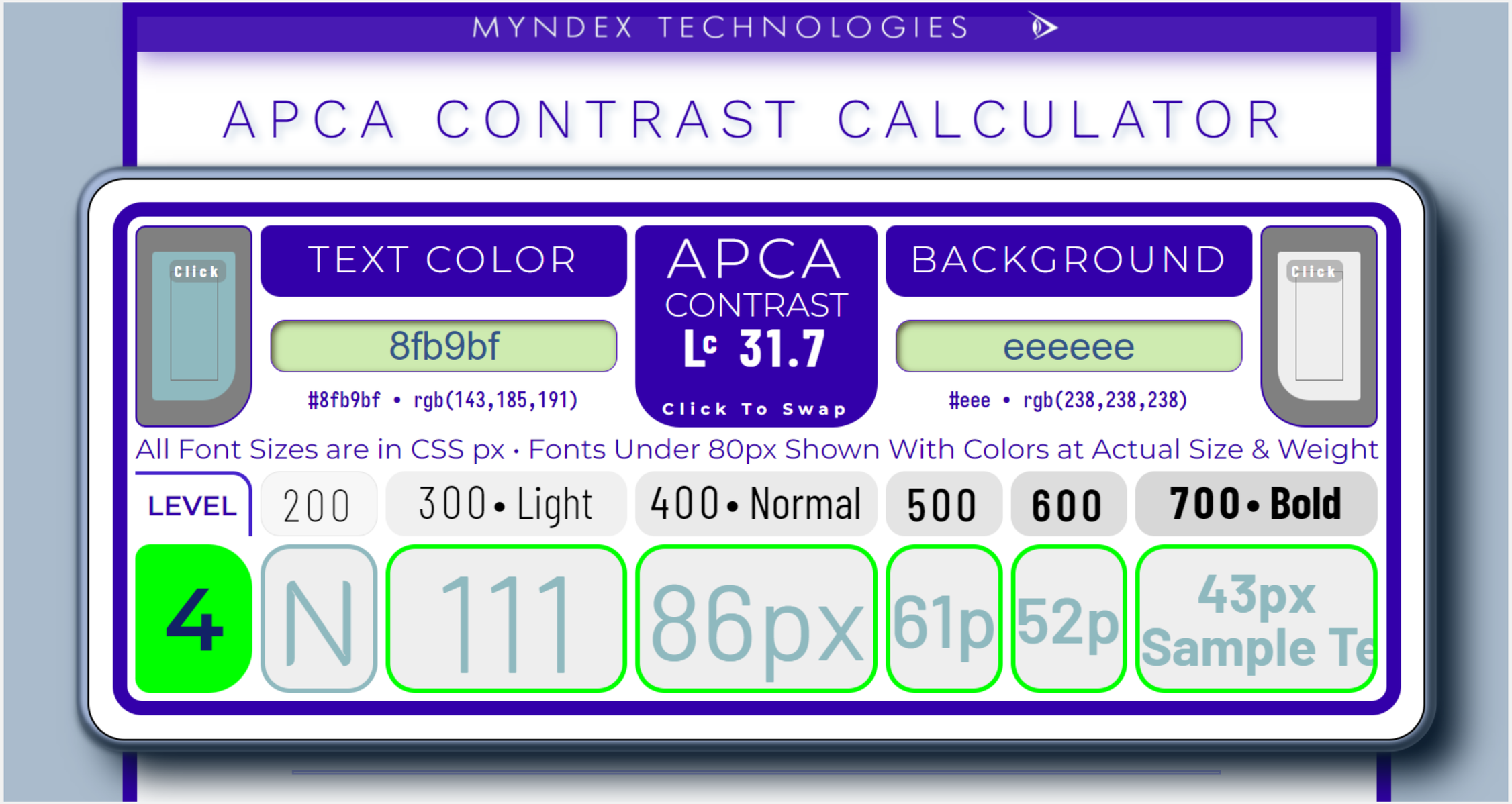
cedricscherer.com @CedScherer z3tt



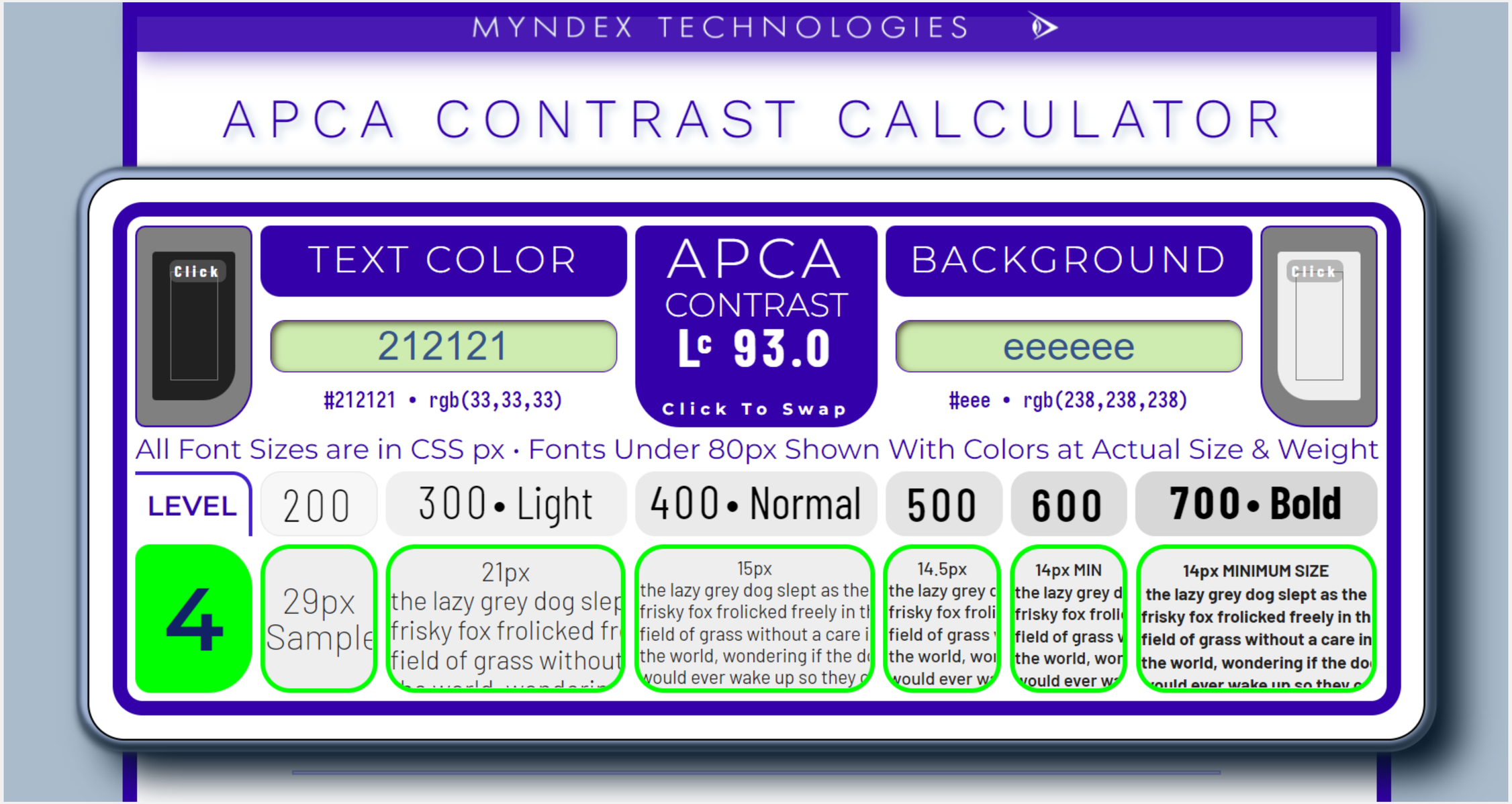
The Myndex APCA Contrast Calculator displays modern contrast ratios for various combinations of text size and font weight



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

cedricscherer.com @CedScherer z3tt

APCA allows for lighter grey on light backgrounds...

... 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

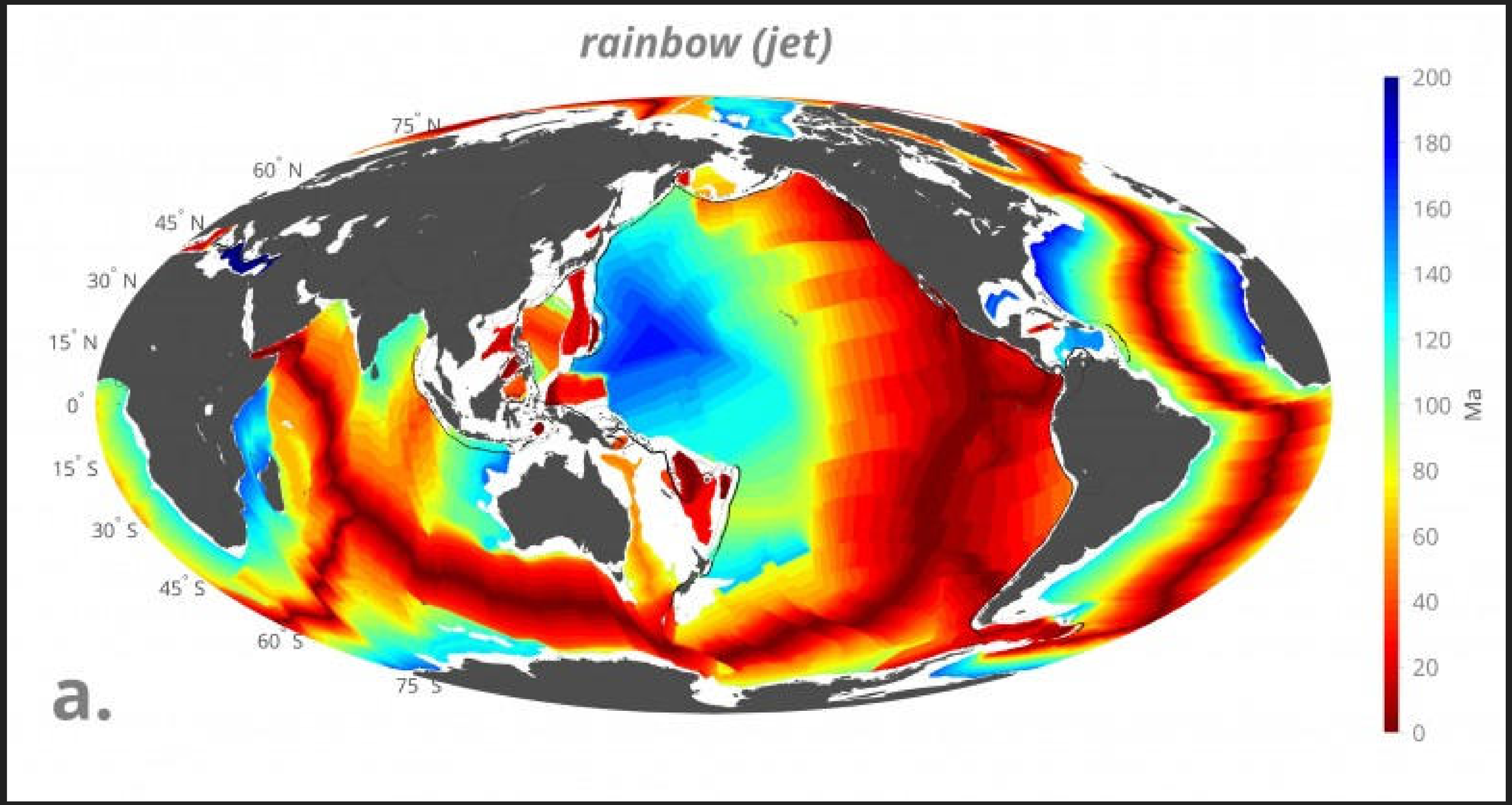
PMCID: PMC5148121 | PMID: 26127048

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)]

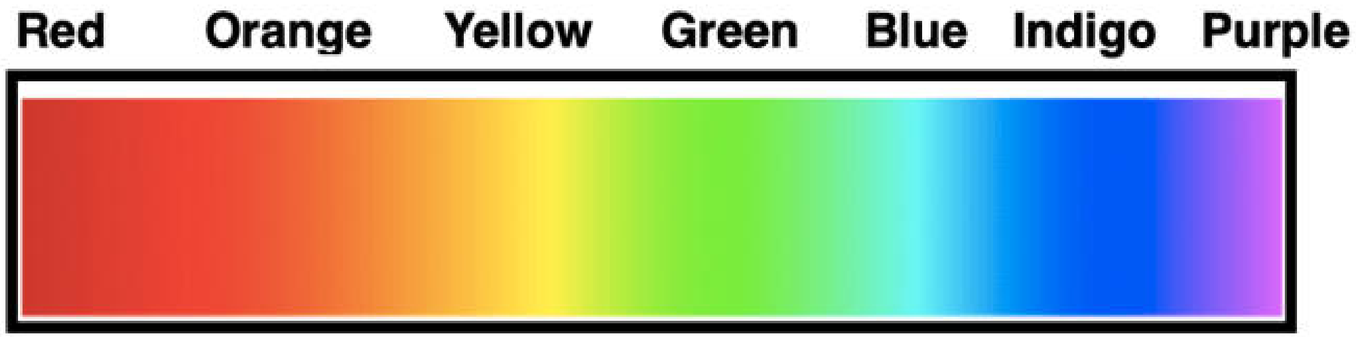
"The ad hoc manner in which color is handled and the lack of standard approaches have been associated with suboptimal and **inconsistent diagnostic decisions with a negative impact on patient treatment and prognosis**."

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

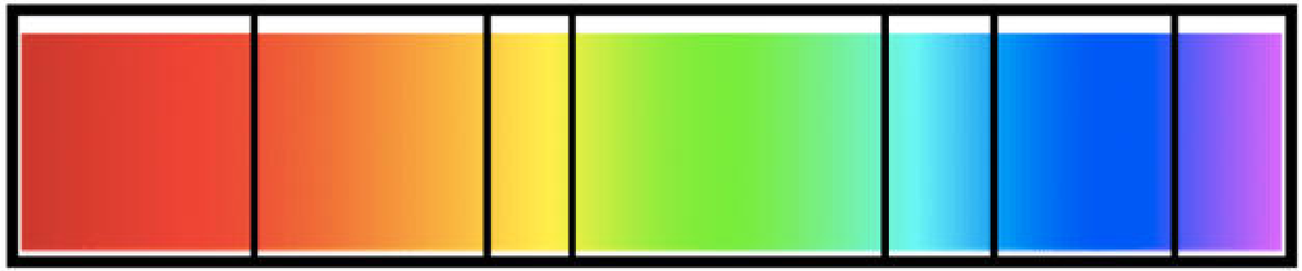


Source: “The Rainbow Colour Map (repeatedly) considered harmful”, EGU Blogs

Non-Uniform Distances between Hues in the Newton Rainbow Colormap



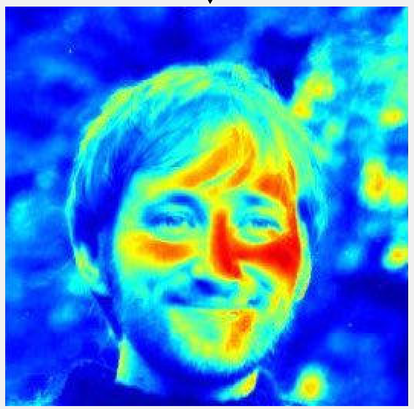
Typical Rainbow Colormap



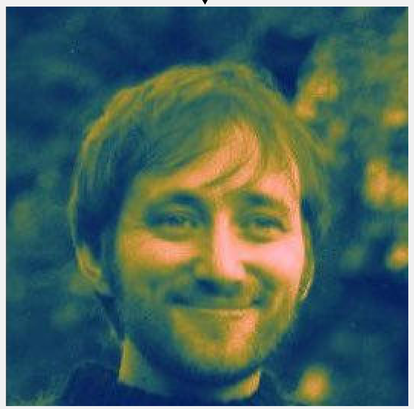
Non-Uniform Distances between Hues.

Modified from Fabio Crameri

true-color Phil



rainbow Phil



batlow Phil

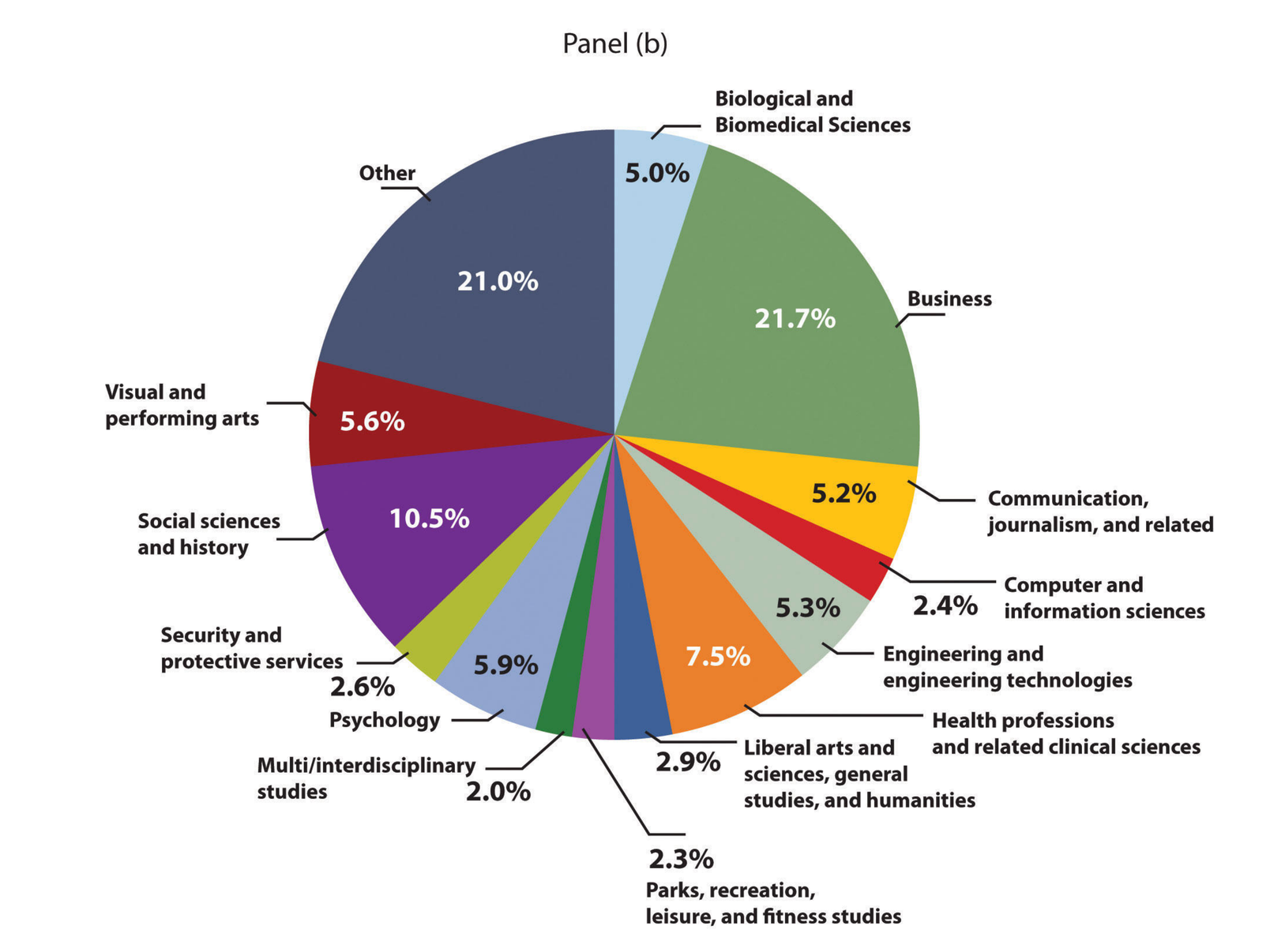
Modified from Fabio Crameri

Summary

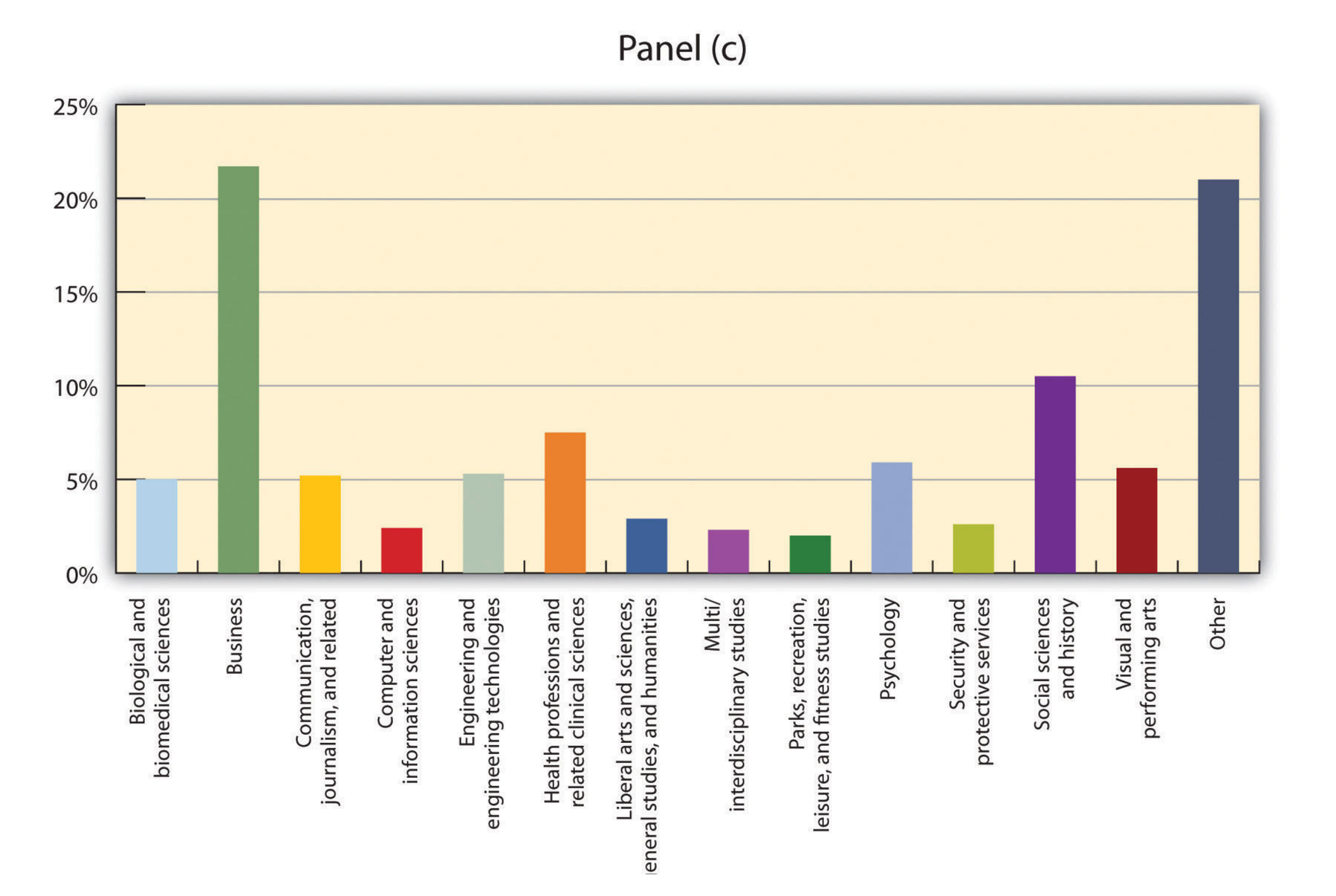
Panel (a)

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%

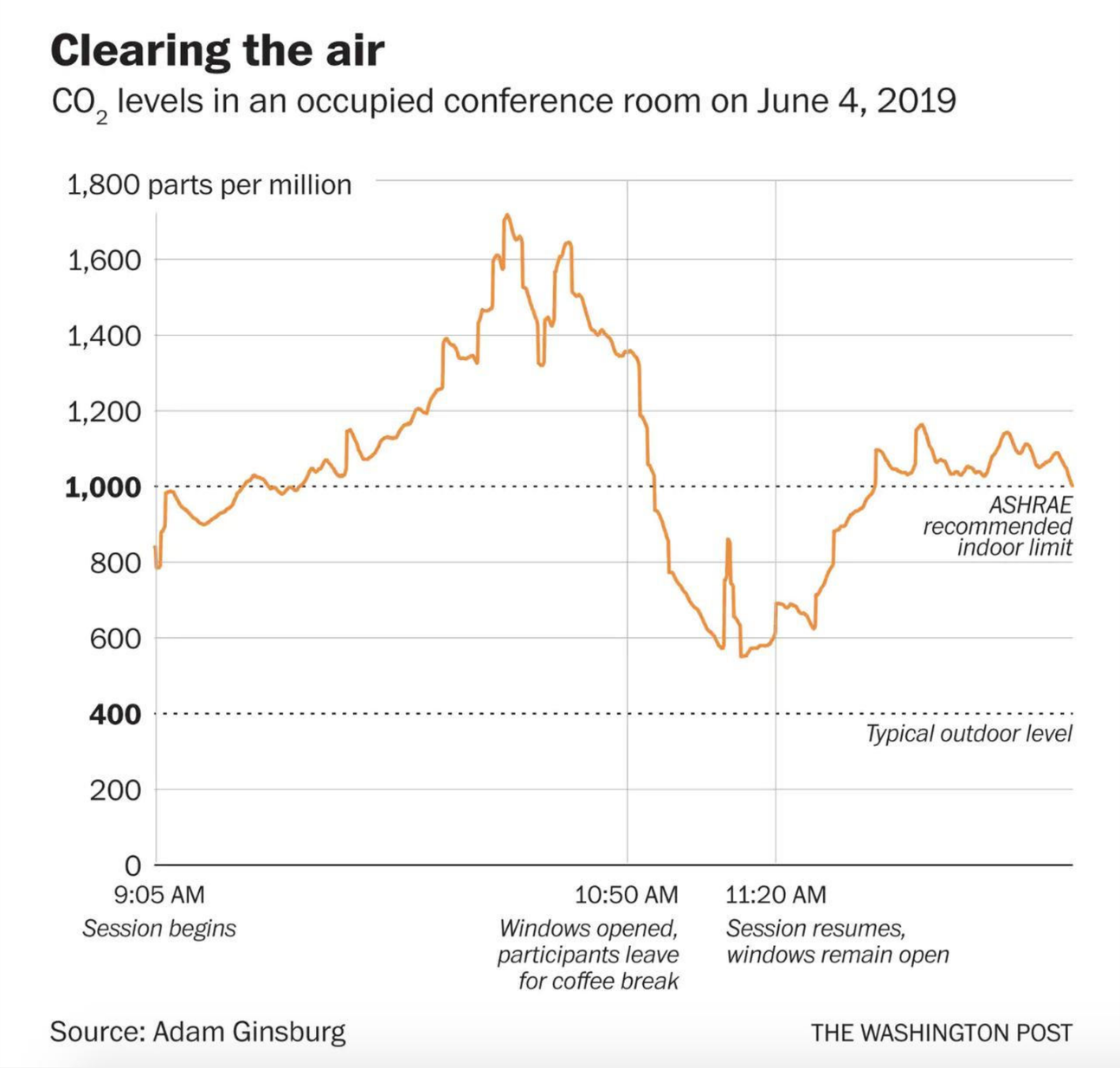
Source: saylordotorg.github.io/text_principles-of-economics-v2.0



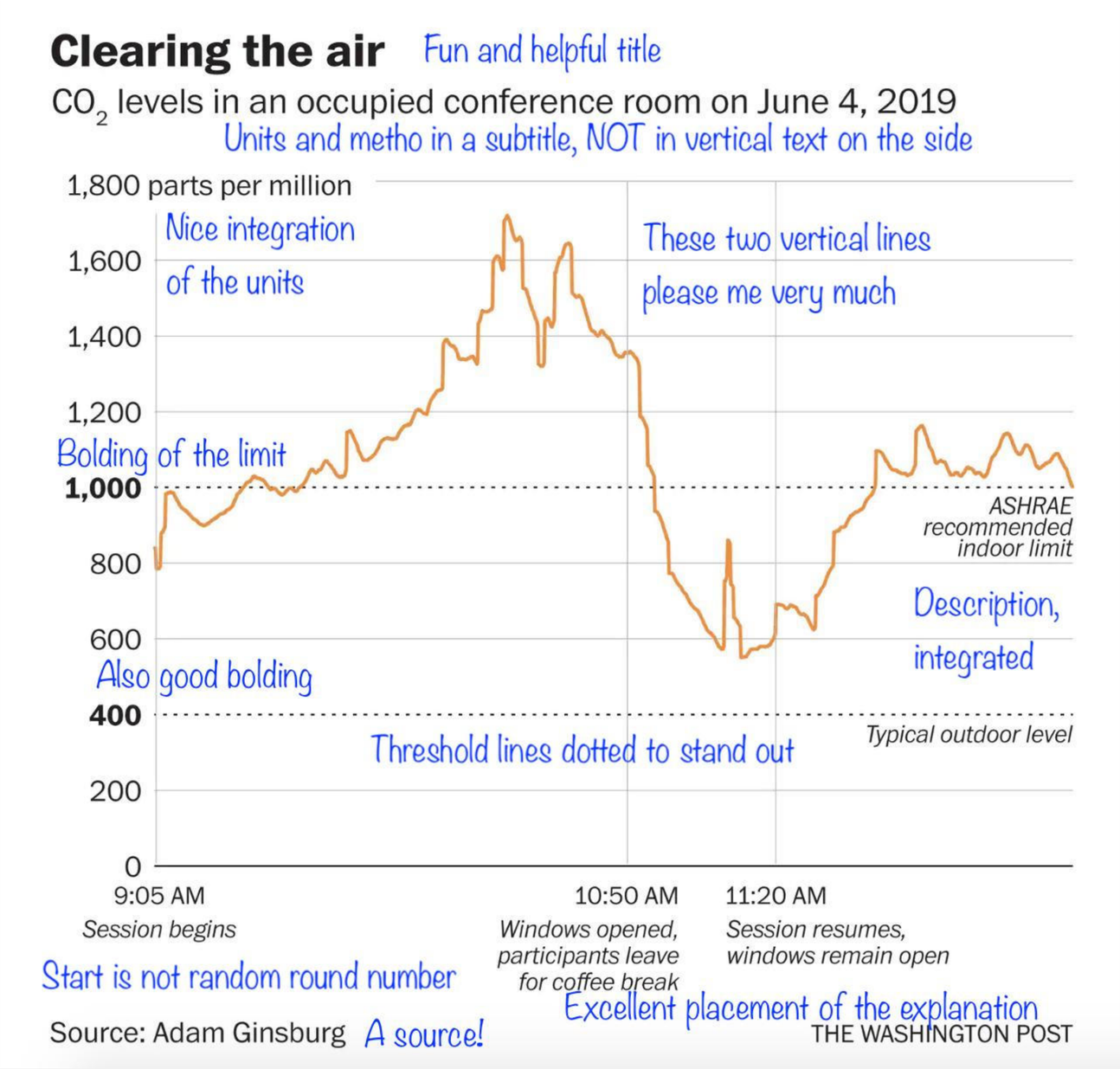
Source: saylordotorg.github.io/text_principles-of-economics-v2.0



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?

cedricscherer.com/workshops/BSIO_DataViz_Exercise.pdf

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